

Análisis transcriptómico de los estadios asociados al huésped
intermediario de *F. hepatica*

Lic. Mauricio Langleib de Souza

Tesis para optar por el título de Magíster en Ciencias Biológicas, subárea Genética

Tutor: Dr. José F. Tort

Co-tutor: Dr. Andrés Iriarte

Programa de Desarrollo de las Ciencias Básicas
Universidad de la República

Mayo 2022

*No te des por vencido, ni aun vencido,
no te sientas esclavo, ni aun esclavo;
trémulo de pavor, piénsate bravo,
y arremete feroz, ya mal herido.*

*Ten el tesón del clavo enmohecido
que ya viejo y ruín, vuelve a ser clavo;
no la cobarde estupidez del pavo
que amaina su plumaje al primer ruido.*

*Procede como Dios que nunca llora;
o como Lucifer, que nunca reza;
o como el robledal, cuya grandeza
necesita del agua y no la implora...*

*Que muerda y vocifere vengadora,
ya rodando en el polvo, tu cabeza!*

Pedro Bonifacio Palacios (*Almafuerte*)

Agradecimientos

Las secciones de agradecimiento suelen ser pasadas de largo por quien se dispone a leer una tesis de posgrado. Como todas, esta tesis es el resultado de un trabajo colectivo y de un proceso extenuante. Ese proceso no podría haberse sostenido de ninguna manera sin la contención, el pienso y la fuerza de la gente que me rodea. Es por eso que para mí ésta es la sección más importante de las ciento y algo de páginas que forman el texto.

En primer lugar quiero agradecer a mi *familia*. Esta tesis entra en la subárea “*Genética*”, y por lo tanto en última instancia trata sobre la herencia. Interesantemente, durante el proceso de sangrado de este trabajo reforcé mis ideas sobre el concepto de familia: prefiero la interpretación sociológica o antropológica del concepto a la biologicista. Para ponerlo en términos bien *nerds*: mi postura es que la familia se forma por descendencia vertical de material genético y/o por transferencia horizontal de afecto. El único componente imprescindible acá para que considere a alguien mi familiar es el del afecto. Por eso cuando digo que quiero agradecer a la *familia* me refiero a las personas que amo y que forman mi red directa de afectos. En especial quiero agradecer a mis padres Nancy y Marcelo, a mi hermana Emilia y toda la barra amiga-hermana que amo y que me acompañó especialmente en el proceso de escribir esta tesis: Gastoncito Rijo, Sebita, Brunito, Agus, Juli (Cho), Dani, Andy, Carmela (nuestra Lucrecia Martel), Braulio, Vignale, Márian, Fede, Facu, Gige, Alfonso y tantos otrxs. Como ya saben lxs amo, y además de decirlo entre copas quiero dejar constancia escrita en estas líneas.

A mis tutores, el *Pepe Tort* y el *Andrecito Iriarte*. Por el espacio, el respaldo y por las horas eternas de discusión que tiene arriba este trabajo colectivo. Hemos pasado por un proceso extenuante, que incluyó reuniones de cuatro horas un veintinueve de diciembre. Hemos pasado por fases de discusión “a lo tano”, y disfrutado la reconciliación y alguna que otra risa entre jale, muzzarella y cerveza. Lo más importante es que gracias a esa dialéctica y ese trabajo colectivo de tesis-antítesis-síntesis bruta llegamos a un buen trabajo, del cual sé que quedamos todos conformes y que queremos publicar para que otros puedan saber de las peripecias de este bicho mata caracolitos. Y eso, eso es lo que más importa; las calenturas quedan para hacer chistes en el futuro, como *deus* manda.

Quiero agradecer tambier al tribunal que corrigió esta tesis: lxs Dres. Estela Castillo, Luisa Berná y Pablo Smircich. En un gesto de empatía infinita se tomaron el tiempo de leer este trabajo en tiempo récord por las circunstancias que imponen nuestros amables mecenas, las agencias financiadoras. Les voy a estar eternamente agradecido por ese gesto humano, por la onda que le pusieron y por las devoluciones valiosas durante la defensa de la Maestría.

A los compañeros y compañeras de los Dptos. en los que trabajé en este tiempo. A Santi Fontenla por las charlas sobre la complejidad de trabajar en *Fasciola* (y ciencia en general), a Agustín, la *Checha*, Carlos y Fernanda por alegrar mis fugaces pero divertidos pasajes por el Depto. de Genética. Muy especialmente a toda la barra de Higiene, mi segunda casa desde la tesina de grado, por hacer amenos los días lindos y los no tanto. A todxs los compañerxs del glorioso Laboratorio de Biología Computacional y la divertida gente del

Depto. de Desarrollo Biotecnológico, donde aprendí a programar (y algo de inmunología, increíblemente). Una mención especial a las DIVI, por ayudarme en miles de cuestiones logísticas durante el último empujón y por ser unas cra perdidas.

Vaya acá también una mención especial a los gigantes Martín *Martincho* Graña y Héctor *Yuyo* Romero, referentes totales que me bancaron la cabeza en el último tramo de trabajo de esta Maestría. Estos elementos míticos de la biología local dejaron que me incorporara de forma paralela a su línea de trabajo en fusógenos celulares. La motivación que me dió entrar a ese trabajo y construir un proyecto en conjunto para lo que va a ser mi doctorado me dió un impulso no menor para culminar esta Maestría en paz y motivado. Gracias por tanto y perdón por tan poco!

Por último, nobleza obliga, quiero agradecer a las instituciones que hicieron posible este trabajo. Al PEDECIBA, el programa de posgrados que elijo orgulloso para desarrollarme a nivel profesional. Estoy convencido que tenemos que incorporar que, aunque imperfecto, es un programa que en su esencia no tiene nada que envidiarle a ningún otro programa de este globito azul llamado Tierra. Aceptar lo contrario es dejar que nos colonicen mentalmente e impedir que la ciencia local se desarrolle lo más posible. Si perdemos esa fé, se va a marchitar nuestra imaginación y se va a ralentizar nuestro impulso. Lo que es peor, vamos a estar desmereciendo la labor heroica de los que impulsaron esta herramienta en las últimas tres décadas. Sigamos sembrando y seamos agradecidos de nuestra cosecha, de la cual deberíamos estar orgullosos. El futuro es nuestro :).

Agradecer también a la ANII que financió esta Maestría, tan perfectible como fundamental para el desarrollo de nuestra ciencia en general y de este trabajo en particular.

Y para cerrar quiero agradecer explícitamente a la kazaja insurrecta, la camarada Alexandra Elbakyan. Eterna gloria a ella y su proyecto (*Sci-Hub*) por devolvernos el fruto de nuestro trabajo. Por mostrar que aunque quieran comerciar con él de la forma más inmundada, siempre va a haber alguien dispuesto a dar pelea, cantarles las cuarenta y hacer un poco de justicia. Aunque implique la marginación y el exilio físico. Para ella, la Comunista más influyente de la ciencia moderna, vaya este humilde homenaje. Estoy seguro que la historia la pondrá en el lugar que se merece como integrante indiscutida del olimpo del saber.

Sin más, a toda esta gente: ***¡GRACIAS TOTALES!***

Índice general

1	Introducción	5
1.1	Aspectos generales del filo Platyhelminthes y la clase Trematoda	5
1.2	<i>F. hepatica</i> : un tremátodo de relevancia sanitaria y económica	6
1.3	Ciclo de vida de <i>F. hepatica</i>	8
1.4	Estudios moleculares	10
1.5	Etapas larvarias de <i>F. hepatica</i>	12
1.5.1	Miracidio	12
1.5.2	Estadíos de desarrollo intracaracol	14
1.5.3	Actores moleculares clave en la interfaz huésped-parásito	16
1.5.4	Proliferación de células pluripotentes: un fenómeno fundamental en platelmintos	18
1.6	Hipótesis de trabajo	22
1.7	Objetivo general	22
1.8	Objetivos específicos	22
2	Materiales y Métodos	23
2.1	Obtención de material biológico y secuenciado	24
2.2	Procesamiento de las lecturas y control de calidad	25
2.3	Ensamblado transcriptómico para estadíos miracidio e intracaracol de <i>F. hepatica</i>	26
2.4	Análisis de expresión génica	27
2.5	Anotación funcional	28
2.6	Análisis funcional de la expresión génica	29
2.7	Análisis de expresión diferencial	30
2.8	Análisis de genómica comparativa	31
2.8.1	Obtención de secuencias, inferencia de grupos de homólogos y árboles filogénicos	31
2.8.2	Análisis comparativo de genes según sus tendencias de expresión	32
2.9	Material suplementario	32

3	Resultados y Discusión	33
3.1	Transcriptoma de miracidio y estadíos intra-caracol de <i>F. hepatica</i>	33
3.1.1	Control de calidad de secuencias	33
3.1.2	Descripción general de los transcritos ensamblados	35
3.1.3	Patrones de expresión a lo largo del ciclo de vida de <i>F. hepatica</i>	38
3.2	Análisis funcional de la expresión génica	53
3.3	Análisis de expresión diferencial	59
3.3.1	Comparativas de expresión en estadios analizados	60
3.3.2	GDEs sobreexpresados en miracidios	64
3.3.3	GDEs sobreexpresados en estadíos intra-caracol 15 d.p.i.	74
3.3.4	GDEs sobreexpresados en estadíos intra-caracol 30 d.p.i.	90
4	Discusión general, conclusiones y perspectivas	99
5	Anexo	103
5.0.1	Estudio de isoformas reportadas	109
5.0.2	Relación entre media y varianza de la expresión de un gen	109
6	Bibliografía	111

Capítulo 1

Introducción

1.1 Aspectos generales del filo Platyhelminthes y la clase Trematoda

El filo Platyhelminthes (del griego, *platys* = plano + *helmins* = gusano) está compuesto por metazoarios aplanados a lo largo del eje dorsoventral y con simetría bilateral .

Los platelmintos son clásicamente divididos en cuatro clases: Turbellaria, Trematoda, Monogenea y Cestoda (Figura 1.1).

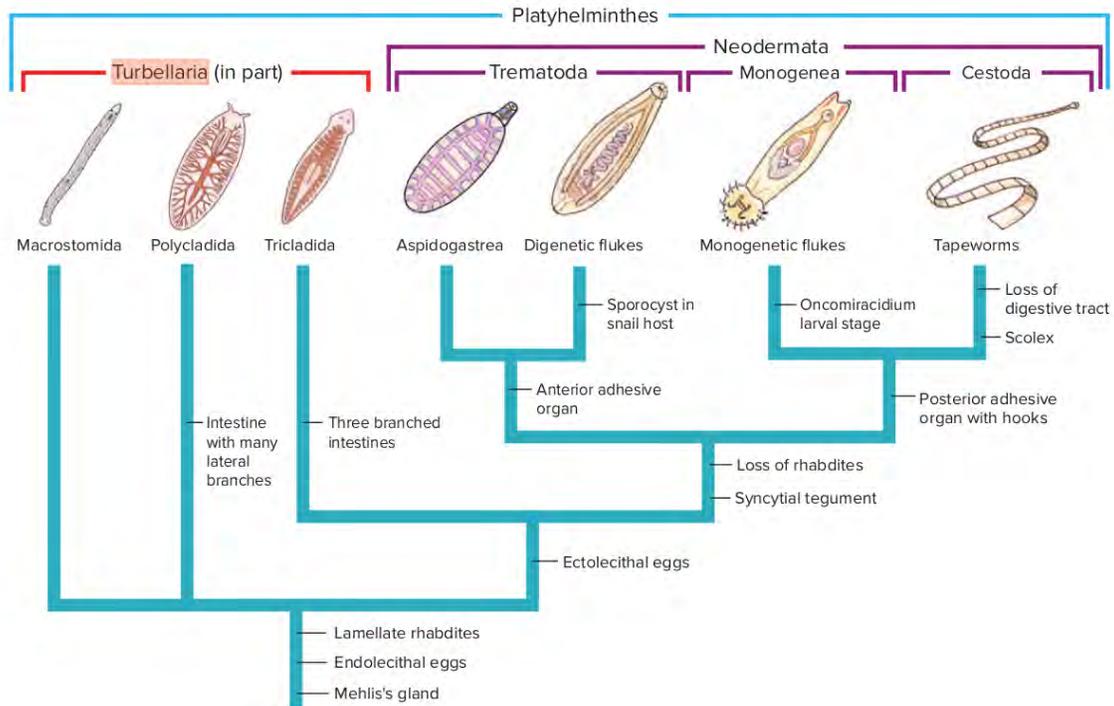


Figura 1.1: Cladograma representando las clases históricamente propuestas para el filo Platyhelminthes y las relaciones filogenéticas inferidas entre las mismas. Se marcan en la figura algunas de las sinapomorfías características de cada clado, como la presencia de un tegumento en los organismos de las clases parásitas del clado Neodermata. Notar también que se marca el pasaje por un huésped intermediario caracol como una característica conservada en tremátodos digenéticos. Figura tomada de Hickman (2021).

La primera de estas, los turbelarios, son un clado parafilético compuesto muy predominantemente por

organismos de vida libre, con algunas pocas especies parásitas. Un ejemplo típico de esta clase son las planarias como la especie modelo *Schmidtea mediterranea*. Las otras tres clases componen un clado denominado Neodermata, caracterizado por un tegumento sincitial, e integrado en su totalidad por organismos parásitos. Es generalmente aceptado que el clado se trata de la mayor radiación adaptativa de parásitos registrada al momento, con más de 15 mil especies. Se estima que todas las especies de vertebrados pueden ser infectados por algún platelminto parásito, dato que ilustra la importancia sanitaria y económica de estos organismos [1].

Los tremátodos son parásitos con ciclos de vida complejos, que involucran el pasaje obligatorio por un molusco como huésped intermediario, y un huésped definitivo vertebrado (generalmente un mamífero), por lo que se los denomina digenéticos o digenea [2]. A lo largo de su ciclo de vida estos organismos sufren cambios morfológicos y de estilo de vida marcados, alternando estadios de vida libre y parásitos. El ciclo implica la reproducción sexual en el adulto en el huésped definitivo, y un proceso de amplificación asexual con diversos estadios que ocurre en el huésped intermediario que es siempre unodos)).

Las especies más relevantes desde el punto de vista sanitario son los esquistosomátidos (*Schistosoma mansoni*, *S. japonicum*, *S. heamatobium*) que afectan a más de 200 millones de personas en todo el mundo, por lo que su estudio está avanzado. También son relevantes para la salud humana los tremátodos de transmisión alimenticia (“*foodborne trematodes*”, en inglés) de los géneros *Opisthorchis*, *Clonorchis*, *Paragonimus* y *Fasciola*, responsables de enfermedades parasitarias crónicas generalmente desatendidas.

Esta tesis de Maestría se centra en el estudio de la expresión génica en *Fasciola hepatica*, el agente principal de la fasciolosis.

1.2 *F. hepatica*: un tremátodo de relevancia sanitaria y económica

La fasciolosis causada por la infección de *Fasciola hepatica* o *Fasciola gigantica* es una enfermedad zoonótica que afecta tanto a rumiantes de alto valor económico como a humanos, por lo cual posee relevancia tanto económica como sanitaria.

Es considerada la enfermedad parasitaria de mayor rango geográfico en el mundo [3], (Figura 1.3A) generando pérdidas anuales de 3000 millones de dólares al sector agroindustrial mundial [4], Los seres humanos son huéspedes ocasionales, siendo infectados al consumir agua o vegetales contaminados. Datos fósiles muestran que ya en la prehistoria los humanos eran infectados por este tremátodo [5], y en la actualidad se estima que cerca de 17 millones de personas se encuentran infectadas y 180 millones están riesgo de ser infectadas y desarrollar esta enfermedad, por lo que ha sido considerada como una enfermedad desatendida por la Organización Mundial de la Salud [6].

Si bien Sudamérica es el continente que presenta mayor prevalencia de la enfermedad en humanos, afortunadamente en países no andinos como Uruguay los reportes de enfermedad humana son esporádicos y aislados [7]. Sin embargo, 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 [8]. 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 (Figura 1.3 B) [9]. 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).

Al igual que en otros parásitos, el estudio del ciclo de vida y de actores moleculares relevantes para el desarrollo de estos parásitos y la interacción con sus huéspedes resultan vitales para comprender la parasitosis y buscar blancos terapéuticos para el control de la enfermedad.

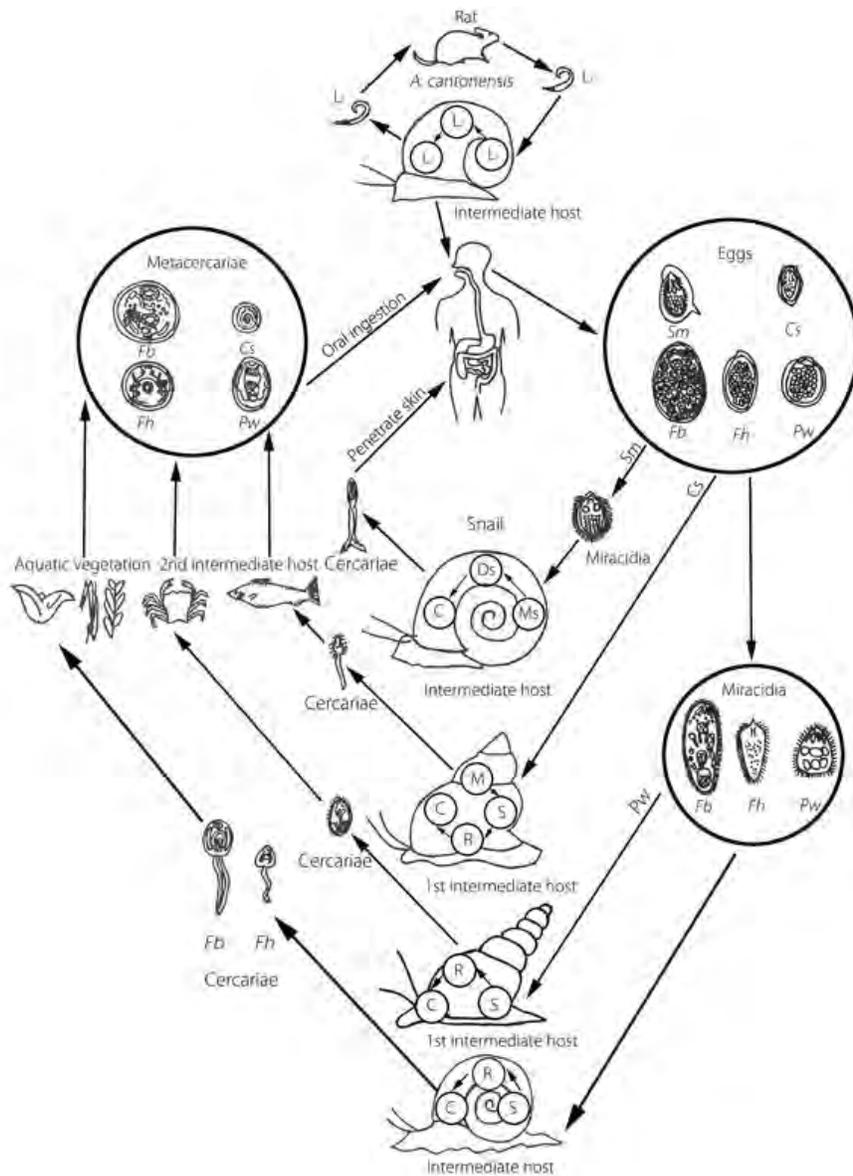


Figura 1.2: Representación del ciclo de vida de tremátodos. Se ilustra la diversidad de estadios de desarrollo en los mismos para diferentes especies de tremátodos. A remarcar, si bien diferentes tremátodos poseen un número variado de huéspedes intermediarios y huéspedes definitivos variados, todos pasan necesariamente por un estadio que infecta a un caracol donde ocurre un ciclo de amplificación asexual. Abreviaciones: Cs: *Clonorchis sinensis*; Fb: *Fasciolopsis buski*; Fh: *Fasciola hepatica*; Pw: *Paragonimus westermani*; Sm: *Schistosoma mansoni*. Tomado de Lu et al. (2018).

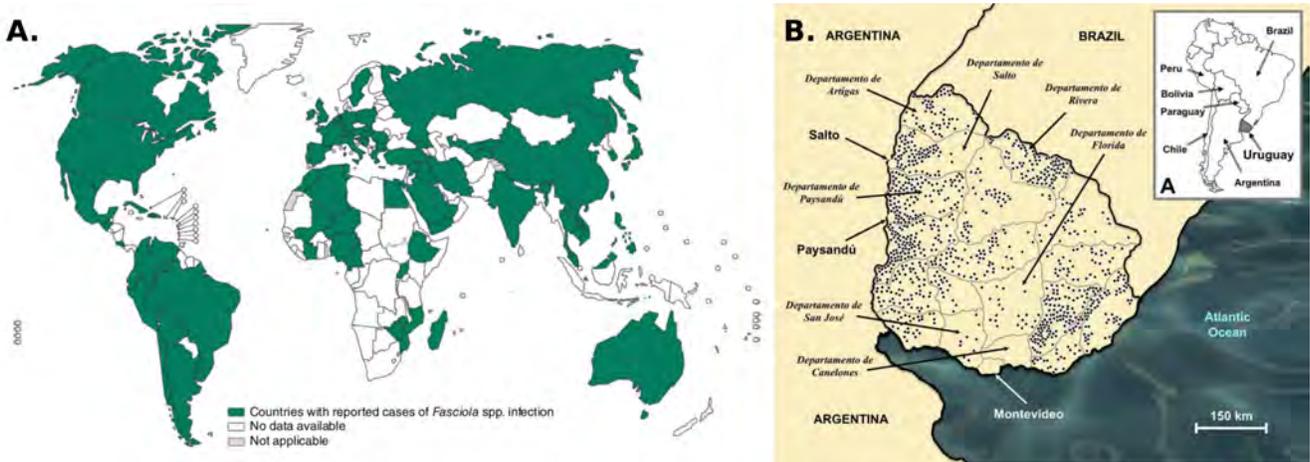


Figura 1.3: Reportes de fasciolosis a nivel global y local. A. Distribución geográfica de casos reportados de fasciolosis humana por infección con *Fasciola* spp. Tomado de Fasciolosis (2022), libro editado por John P. Dalton. B. Distribución geográfica de los casos reportados de fasciolosis en Uruguay. Modificado de Bargues et al. (2017).

1.3 Ciclo de vida de *F. hepatica*

Durante su ciclo de vida *F. hepatica* arolla variados: huevo, miracidia, esporocisto, redia, cercaria, metacercaria, juvenil y adulto (Figura 1.4).

Se puede tomar como punto de partida arbitrario para la descripción del ciclo de *F. hepatica* a los huevos, los cuales son generados en grandes cantidades (~20.000-25.000 huevos por día en oveja) a partir de la reproducción sexual de los adultos hermafroditas en el huésped mamífero (1.4). Los mismos son excretados al medio ambiente a través de heces, y consisten de un óvulo fertilizado rodeado de células vitelinas que proveen los nutrientes necesarios para el desarrollo del embrión fuera del huésped definitivo [10]. Señales externas son necesarias para iniciar el desarrollo dentro del huevo, tales como la remoción de heces y la exposición a rangos precisos de temperatura y luz, y para la eclosión del miracidio, forma larvaria ciliada que se generó dentro del huevo. Este estadio de vida libre se caracteriza por sus rápidos movimientos a través de la acción de profusas ciliadas, y la búsqueda activa de su huésped intermediario mediante quimiotaxis a compuestos secretados por caracoles del género *Lymnaea*. El metabolismo de miracidios es netamente catabólico, obteniendo energía a partir de reservas de glucógeno.

Una vez toman contacto con un posible huésped, los miracidios penetran el manto del caracol mediante la acción coordinada mecánica y enzimática, perdiendo la capa celular cubierta de ciliadas, y comenzando la transformación en esporocisto [11,12]. Este es el primer estadio de desarrollo intracaracol, y el comienzo de la fase de amplificación asexual, la cual se encuentra conservada en tremátodos [2,13]. A partir de células indiferenciadas presentes en el interior del esporocisto se generan masas germinales que evolucionan a redias. Las redias poseen una complejidad morfológica mayor (*e.g.* poseen faringe, ausente en esporocistos), y son capaces de migrar alimentándose de tejidos del caracol, mientras que en su interior se repite el proceso de proliferación asexual dando lugar a más redias hijas (que repetirán el proceso asexual) y/o cercarias. Estas son el último estadio de desarrollo que se genera al interior del caracol, y las que suelen abandonar el caracol en un período que toma de 4 a 7 semanas. Las cercarias que emergen nadan libremente en el agua por un período breve que puede durar hasta 2 hs, momento en el cual la misma se detiene y enquista sobre un vegetal, generalmente berro, transformándose en una forma resistente, las metacercarias.

La amplificación asexual intracaracol es clave para el éxito infectivo de estos parásitos, en tanto permite un aumento considerable de los números poblacionales de estos: se ha estimado que un único miracidio

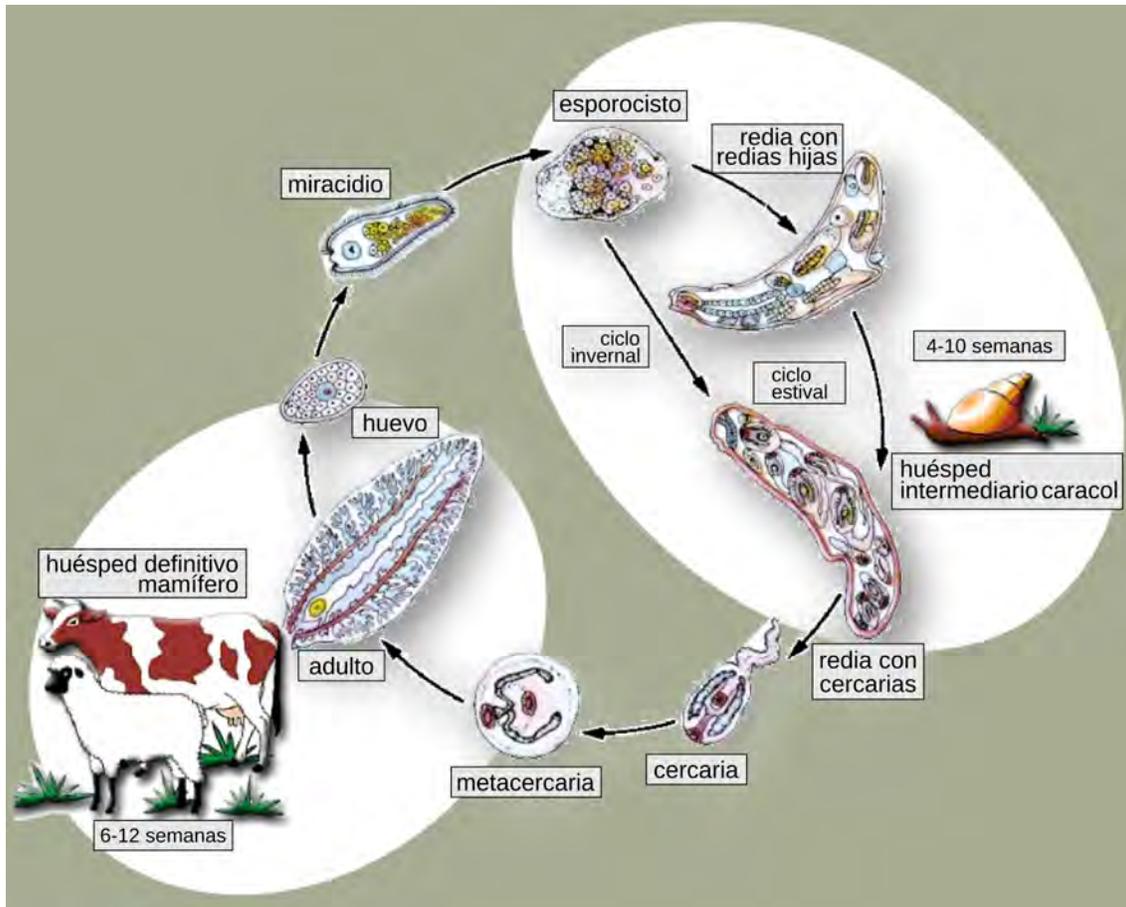


Figura 1.4: Ciclo de vida de *Fasciola hepatica*.

puede producir cerca de 4000 metacercarias.

La metacercaria está conformada por una estructura compleja, con un quiste externo y uno interno. El quiste externo actúa como barrera contra infecciones bacterianas y fúngicas, a la vez que ayuda al agarre al sustrato vegetal, la cual resulta importante para la supervivencia de la metacercaria y la infección del huésped final. El quiste interno está constituido por el tegumento de la cercaria, y como tal constituye una barrera adicional frente al medio exterior. Las metacercarias constituyen un estadio de vida libre de infección pasiva, ingresando al huésped mamífero generalmente a través de la ingesta del vegetal en el cual se adhieren, o en su defecto por ingesta de agua o material orgánico contaminado. El pasaje por el estómago activa las metacercarias, las que comienzan a desenquistarse en el duodeno estimuladas por las sales biliares. Emerge así un estadio juvenil (referida generalmente como “*NEJ*” en la literatura, sigla que hace referencia a su denominación en inglés, “*newly excysed juvenile*”), que migra activamente y en un periodo de dos horas, penetran la mucosa intestinal rompiendo sus células epiteliales, el tejido conectivo y las fibras musculares, dirigiéndose así a la cavidad peritoneal. Se traspasa así la pared intestinal pudiéndose observar en la cavidad abdominal estos juveniles migran y penetran al hígado. Allí se abren paso a través del tejido hepático por unas 5 a 6 semanas, durante las cuales sucede un crecimiento significativo de estos parásitos, y se instalan finalmente en las vías biliares. Luego de ocho semanas de infección se pueden comenzar a encontrar huevos en la bilis del huésped y posteriormente en sus heces, cerrándose así el ciclo de vida del parásito.

1.4 Estudios moleculares

La aparición de las tecnologías de secuenciación masiva de ADN, ARN y proteínas significó un salto sustancial en el estudio de parásitos helmintos en general, permitiendo comenzar a analizar los complejos ciclos biológicos desde una perspectiva transcriptómica. Nuestro modelo de estudio, *F. hepatica* no estuvo ajeno a estos avances detallándose en la Tabla 1.1 los principales estudios genómicos, transcriptómicos y proteómicos publicados desde el 2009 a la fecha relacionados a estadios de desarrollo de *F. hepatica*.

Tabla 1.1: Principales publicaciones donde se realiza secuenciación masiva de ADN, ARN y proteínas en estadios de desarrollo de *F. hepatica*.

Año	Tipo de estudio	Estadios analizados	Titulo	Autores	DOI
2009	Transcriptómica	Adultos	An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen <i>Fasciola hepatica</i> : proteins associated with invasion and infection of the mammalian host.	Robinson MW, Menon R, Donnelly SM, Dalton JP, Ranganathan S.	doi: 10.1074/mcp.M900045-MCP200.
2010	Transcriptómica	Adultos	Elucidating the transcriptome of <i>Fasciola hepatica</i> - a key to fundamental and biotechnological discoveries for a neglected parasite.	Young ND, Hall RS, Jex AR, Cantacessi C, Gasser RB.	doi: 10.1016/j.biotechadv.2009.12.003.
2010	Transcriptómica	Adultos	Survey of transcripts expressed by the invasive juvenile stage of the liver fluke <i>Fasciola hepatica</i> .	Cancela M, Ruétalo N, Dell'Oca N, da Silva E, Smircich P, Rinaldi G, Roche L, Carmona C, Alvarez-Valín F, Zaha A, Tort JF.	doi: 10.1186/1471-2164-11-227.
2015	Genómica,transc	NEJ, juvenil, adulto	The <i>Fasciola hepatica</i> genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution.	Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, Paterson S.	doi: 10.1186/s13059-015-0632-2.
2017	Genómica,transc	Huevos, adultos	Genomes of <i>Fasciola hepatica</i> from the Americas Reveal Colonization with <i>Neorickettsia</i> Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers.	McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, Fontenla S, Choi YJ, Tyagi R, Hallsworth-Pepin K, Mann VH, Kammili L, Latham PS, Dell'Oca N, Dominguez F, Carmona C, Fischer PU, Brindley PJ, Mitreva M.	doi: 10.1371/journal.pgen.1006537.
2018	Transcriptómica	Adultos	Pleiotropic alterations in gene expression in Latin American <i>Fasciola hepatica</i> isolates with different susceptibility to drugs.	Radio S, Fontenla S, Solana V, Matos Salim AC, Araújo FMG, Ortiz P, Hoban C, Miranda E, Gayo V, Pais FS, Solana H, Oliveira G, Smircich P, Tort JF.	doi: 10.1186/s13071-017-2553-2.
2018	Transcriptómica	NEJ, juveniles, Adultos	Infection by the Helminth Parasite <i>Fasciola hepatica</i> Requires Rapid Regulation of Metabolic, Virulence, and Invasive Factors to Adjust to Its Mammalian Host.	Cwiklinski K, Jewhurst H, McVeigh P, Barbour T, Maule AG, Tort J, O'Neill SM, Robinson MW, Donnelly S, Dalton JP.	doi: 10.1074/mcp.RA117.000445.
2020	Genómica,transc	Huevos, adultos	Adaptive Radiation of the Flukes of the Family Fasciolidae Inferred from Genome-Wide Comparisons of Key Species.	Choi YJ, Fontenla S, Fischer PU, Le TH, Costabile A, Blair D, Brindley PJ, Tort JF, Cabada MM, Mitreva M.	doi: 10.1093/molbev/msz204.
2021	Transcriptómica,	NEJ, juveniles, adultos	Complementary transcriptomic and proteomic analyses reveal the cellular and molecular processes that drive growth and development of <i>Fasciola hepatica</i> in the host liver.	Cwiklinski K, Robinson MW, Donnelly S, Dalton JP.	doi: 10.1186/s12864-020-07326-y.
2021	Transcriptómica	Adultos	Transcriptome assembly dataset of anthelmintic response in <i>Fasciola hepatica</i> .	Miranda-Miranda E, Cossio-Bayugar R, Aguilar-Díaz H, Narváez-Padilla V, Sachman-Ruíz B, Reynaud E.	doi: 10.1016/j.dib.2021.106808.

A la fecha existen disponibles dos ensamblados genómicos y múltiples estudios transcriptómicos y proteómicos. No obstante, estos últimos se han centrado exclusivamente en el análisis de expresión de ARN y proteínas en estadios infectivos del huésped mamífero, probablemente debido al impacto directo a nivel económico y sanitario que los mismos tienen. Sin embargo, los estadios asociados al huésped intermedio, donde ocurren procesos biológicos relevantes vinculados con el desarrollo y la amplificación asexual continúan siendo relativamente poco estudiados.

1.5 Etapas larvarias de *F. hepatica*

Respecto al grado de desarrollo de investigación sobre los estadios de desarrollo intracaracol, los trematólogos Kirill Dobrovolskij y Andrej Galaktionov planteaban hace 20 años en su libro “*The Biology and Evolution of Trematodes*” lo siguiente [2]:

“*En trabajos referentes a los ciclos de vida de una especie [de tremátodo], como regla general, se escriben a lo sumo unas pocas líneas dedicadas a esporocistos y redias. Existen relativamente pocas investigaciones relacionadas a la morfología y la biología de estos estadios en el ciclo de vida de tremátodos. No es una sorpresa que los datos referentes a las generaciones partenogénicas de tremátodos sean escasos y fragmentados. Esto se puede ilustrar en la última revisión de la filogenia de digéneos (...) construida con base en 56 caracteres morfológicos. Solo dos de estos tienen algo que ver con el miracidio, y seis con los esporocistos y las redias.*”

Si bien es cierto que actualmente se siguen dedicando pocas líneas a esta fase del ciclo de vida en revisiones y estudios de carácter general sobre tremátodos, y que es notorio que estos estadios reciben menor atención que aquellos directamente vinculados a huéspedes vertebrados, la aparición en los últimos años de varios estudios de expresión de esporocistos, redias y cercarias parece mostrar un aumento de la atención sobre los mismos. Estos trabajos refieren a los schistosomátidos, *S. mansoni* [[14]; Wang2018] y *S. japonicum* [15], el fasciólido *F. gigantica* [16] y dos especies de la familia Psilostomatidae [17].

A continuación se repasan brevemente las principales características morfológicas y de estilo de vida de estos estadios de *F. hepatica*. Luego de esta descripción se hace foco en dos aspectos clave de la biología de estos: *i*) la interacción con el sistema inmune de sus huéspedes, la principal barrera con la que se encuentran durante su desarrollo y *ii*) su capacidad de reproducción asexual, incluyendo su mecanismo subyacente y el vínculo con otros estadios de desarrollo.

1.5.1 Miracidio

El miracidio es un estadio de vida libre de alrededor de $130\mu\text{m}$ de largo, cuya embriogénesis y morfogénesis acontece en los huevos de *F. hepatica*. Frente a condiciones óptimas de temperatura, luz y humedad los huevos eclosionan y el miracidio emerge a través de una estructura denominada opérculo, presentando movimientos aleatorios de alta velocidad ($\sim 1\text{ mm/s}$). La vida media de un miracidio es de un día, tiempo durante el cual el individuo o bien infecta a un caracol del género *Lymnaea* o muere de inanición. A diferencia de estadios homólogos refieren otros clados de tremátodos (*e.g.* Hemiuroida), los miracidios de familias como Fasciolidae, Echinostomatidae y Schistosomatidae son de infección activa, presentando adaptaciones morfológicas y comportamentales que aumentan la probabilidad de infectar a un huésped caracol [2]. Un ejemplo de estas adaptaciones morfológicas es la presencia ubicua de cilias en la superficie, las cuales permiten el rápido movimiento de estos organismos (Figura 1.5). Un dato que respalda el carácter adaptativo de estas estructuras es que se encuentran en números mucho menores (o directamente ausentes) en miracidios de especies de infección pasiva.

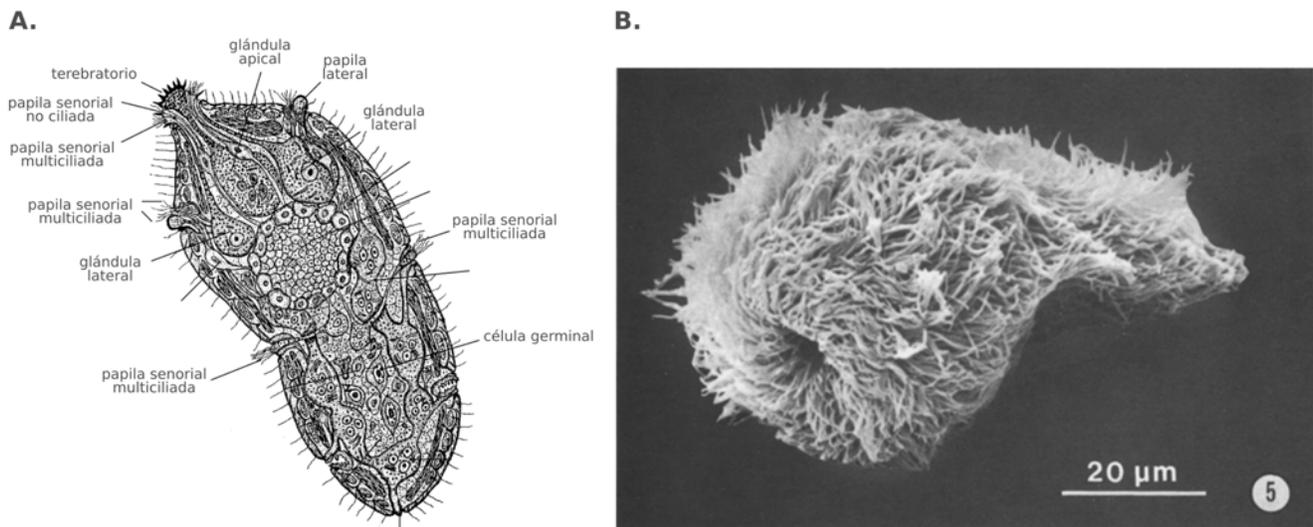


Figura 1.5: A. Representación esquemática de un miracidio, con especificación de algunas características morfológicas de relevancia. Modificado de Pan (1980). B. Imagen del miracidio de *F. hepatica* generada por microscopía electrónica de barrido. Es notable la presencia ubicua de cilias pertenecientes a las placas epiteliales, las cuales asisten el movimiento de esta etapa larvaria de vida libre. Tomado de Koie et al. (1976).

El movimiento continuo de miracidios supone un gasto energético considerable. Los miracidios poseen un metabolismo aerobio netamente catabólico. El contenido de triacilglicerolos y aminoácidos se ha constatado inalterado tanto en miracidios de *F. hepatica* como *S. mansoni*, y se ha demostrado que este estadio depende de forma casi exclusiva de sus reservas de glucógeno para la obtención de energía [18,19].

Otras estructuras de valor adaptativo presentes en el esquema de la Figura 1.5A son las múltiples papilas sensoriales (anteriores y laterales) y los dos ojos pigmentados presentes en el extremo anterior del organismo. Se presume que ambas estructuras son fotosensibles, participando de la regulación del comportamiento fototrópico que presentan los miracidios y que favorecen su movimiento hacia las zonas superficiales de las masas de agua dulce que habitan, donde suele existir una mayor densidad de caracoles del género *Lymnaea*. Se ha reportado que los miracidios de *F. hepatica* y *S. mansoni* poseen movimiento quimiotáctico, siendo atraídos por ácidos grasos de cadena corta presentes en el mucus del caracol [10]. Se ha propuesto que las papilas sensoriales también podrían ser parte de la regulación de este movimiento quimiotáctico, y resulta interesante remarcar que estas estructuras están ausentes en miracidios de infección pasiva, lo cual sugiere que también estas poseen un valor adaptativo. Se ha propuesto que en conjunto estas estructuras facilitan la infección al caracol, la cual sucedería en tres fases: *i*) tropismo hacia ambiente habitado por potenciales huéspedes (incluyendo foto- y termotropismo), *ii*) un período de movimiento aleatorio y *iii*) orientación hacia el posible huésped por quimiotaxis.

Una vez alcanzado un potencial huésped caracol, los miracidios detienen su movimiento y se adhieren a la superficie del mismo, mediante una papila apical (a veces denominada *terebratorio*) con reminiscencias a estructuras succionadoras. Se ha mostrado que esta estructura es fundamental en la adhesión e infección del caracol [11,12]. El proceso infectivo se logra tanto por fricción mecánica ejercida a partir de movimientos rotatorios, así como por la secreción de proteasas provenientes desde tres glándulas (una apical y dos laterales) que poseen la capacidad de degradar la epidermis del caracol [11,20]. A diferencia de otros tremátodos como *S. mansoni*, el proceso de transformación de miracidios a esporocistos comienza ya durante el proceso infectivo, observándose el desprendimiento de las placas epiteliales y las cilias unidas a ellas [12].

Por último, resulta importante resaltar que en su interior los miracidios poseen un conjunto de unas células no diferenciadas, denominadas células germinales (en conjunto, masa germinal). Como se verá en las siguientes secciones estas células son fundamentales para comprender el proceso de generación de redias y cercarias, los siguientes estadios que ocurren dentro del caracol. Estudios recientes han demostrado que estas células son un elemento común a lo largo del ciclo de vida de los tremátodos.

1.5.2 Estadios de desarrollo intracaracol

Ya dentro del caracol el miracidio completa la transformación, dando lugar al esporocisto. La transformación comienza con el desprendimiento de las placas epiteliales ciliadas, proceso que se acompaña por la formación del tegumento. La metamorfosis de miracidio a esporocisto se da de forma regresiva, perdiéndose múltiples estructuras: glándulas secretoras (una vez culminada la formación del tegumento), ojos, terebratorio, ganglio cerebral y algunas papilas sensoriales se pierden en el proceso (Figuras 1.6A,B). Los esporocistos de todos los tremátodos carecen de sistema digestivo, por lo cual el tegumento juega un papel esencial como interfaz para la obtención de nutrientes a partir del huésped y la secreción de desechos. Los esporocistos poseen una capa de tejido muscular que cubre su cuerpo, y en el caso de *F. hepatica* poseen capacidad migratoria, trasladándose hacia lugares favorables para su desarrollo y secretando hialuronidasa a efectos de desintegrar el tejido del huésped.

A medida que se da el proceso de desarrollo del esporocisto, una de sus características más importantes tiene lugar: la proliferación de las células germinales que se multiplican activamente, formando mórulas redondas que dan origen a las futuras redias. Su morfología sencilla y la presencia de estas masas celulares han hecho que los esporocistos (y las redias) sean informalmente llamadas “sacos germinales” en la literatura tradicional (“*germinal sacs*” en inglés), denotando la importancia de esta característica.

Cada esporocisto da lugar a varias redias, las cuales se desarrollan a la luz del esporocisto en un compartimento denominado “cámara de cría” (“*brood chamber*” en inglés). Ya en esta fase del desarrollo intracaracol se da la multiplicación del número de individuos, que en última instancia saldrán del huésped intermediario con capacidad infectiva del huésped definitivo. Si bien las redias se desarrollan de forma asincrónica, culminan su proceso de desarrollo en un tiempo más o menos similar, momento en el cual se da la ruptura del esporocisto con la concomitante liberación de las redias.

Las redias constituyen un segundo estadio larvario durante el cual acontece una expansión clonal de individuos. Resulta interesante remarcar que no todos los tremátodos pasan por el estadio de desarrollo de redia. A modo de ejemplo, esporocistos de especies de Schistosomatidae (denominados esporocistos “madre”) no generan redias, si no esporocistos “hijos”. A la fecha no existen estudios evolutivos que presenten hipótesis respecto a qué estadio de desarrollo estaba presente en tremátodos ancestrales ni al proceso evolutivo que dió lugar a esta diferencia entre sus descendientes. Las redias poseen una morfología más compleja que los esporocistos y son más grandes que estos. Presentan un sistema digestivo desarrollado, ingiriendo células del huésped de forma activa. Además presentan estructuras denominadas “brotes ambulatorios” (“*ambulatory buds*” en inglés) que facilitan su migración dentro del caracol.

Las redias poseen una morfología más compleja que los esporocistos y son más grandes que estos. Presentan un sistema digestivo desarrollado, ingiriendo células del huésped de forma activa. Además presentan estructuras denominadas “brotes ambulatorios” (“*ambulatory buds*” en inglés) que facilitan su migración dentro del caracol (Figuras 1.6C,D). Las redias también contienen en su interior masas germinales, que son las responsables de una nueva expansión clonal de individuos.

Al igual que en el caso de los esporocistos, cada redia da lugar a varias redias (denominadas “hijas”) que se desarrollan de forma asincrónica a partir de masas germinales que pasan por un proceso de embriogénesis y morfogénesis. Estas redias hijas pueden o bien dar lugar a más redias hijas (existen al menos hasta tres generaciones de redias en *F. hepatica* [10]), o dar lugar al último estadio de desarrollo intracaracol,

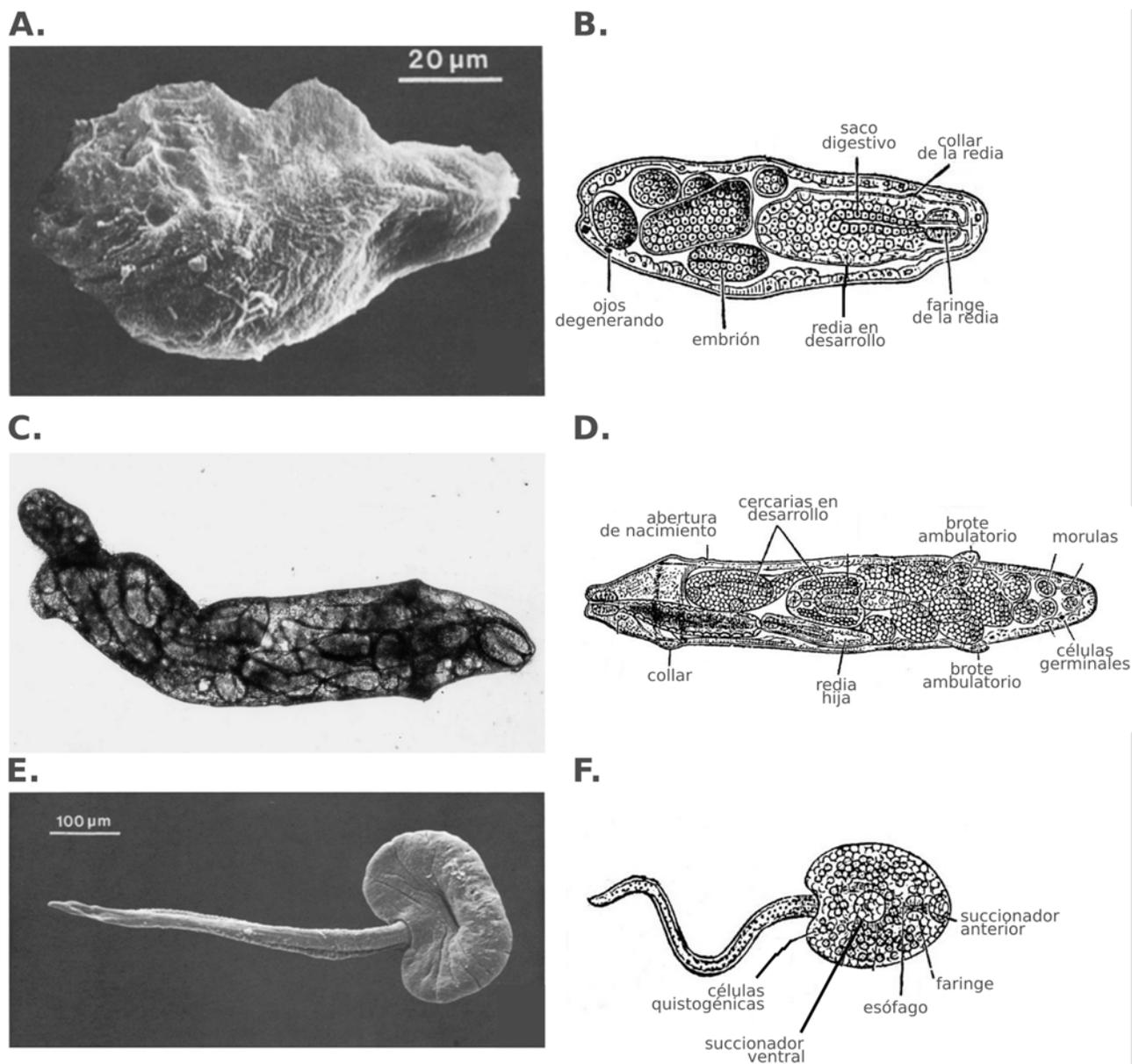


Figura 1.6: Representación esquemática e imágenes de especímenes de estadios larvarios de *F. hepatica*. A. Microscopía electrónica de barrido de un esporocisto, tomada de Koie (1976). B. Representación esquemática de un esporocisto. C. Microscopía de una redia, tomada de Augot (1997). D. Representación esquemática de una redia. E. Microscopía electrónica de barrido de una cercaria, tomada de Koie (1997). Representación esquemática de una cercaria. Todos los esquemas son tomados de la Encyclopaedia Britannica (1911).

la cercaria. La naturaleza exponencial de las sucesivas generaciones de redias son la base del aumento exponencial de individuos con capacidad infectiva. Al igual que en el caso de la transición miracidio-esporocisto o esporocisto-redia, los eventos moleculares que regulan esta transición de desarrollo no han sido establecidos a la fecha.

Las cercarias son el último estadio de desarrollo intra-caracol, pasando la última etapa de su vida como un organismo de vida libre. Su escape del huésped caracol se logra gracias a la secreción de proteasas y la acción conjunta de varias cercarias que inflige un daño al huésped que suele causar su muerte. A nivel morfológico presentan un cuerpo donde se destacan una ventosa oral y una ventral y una cola muscular que asiste a sus rápidos movimientos (Figuras 1.6E,F). A nivel metabólico este estadio posee similitudes con los miracidios, en tanto depende del catabolismo de reservas de glucógeno para sustentarse energéticamente, y posee un lapso de ~2hs para encontrar una superficie donde enquistar o morir de inanición. Una vez encontrada esta superficie (generalmente vegetal) un conjunto de células, denominadas quistogénicas, secretan componentes que formarán la estructura exterior del quiste de la metacercaria. El tegumento de la cercaria pasará a componer la estructura interna de este quiste. Con estos eventos y el desprendimiento de su cola se completa la transición de cercaria a metacercaria. Culmina así la fase de desarrollo asociada a huésped caracol del ciclo de vida de *F. hepatica*.

1.5.3 Actores moleculares clave en la interfaz huésped-parásito

Los sistemas de interacción huésped-parásito constituyen uno de los ejemplos paradigmáticos de co-evolución de especies. La interacción entre tremátodos y caracoles data de por lo menos 200 millones de años y, al igual que en el caso de los huéspedes mamíferos, se enmarca dentro de este escenario evolutivo [21]. Se configura una interacción que suele describirse como una “*carrera armamentista*” en la que se contraponen las presiones selectivas ejercidas mutuamente entre los dos componentes del sistema. Los actores moleculares que despliegan esta carrera armamentista son los componentes de la interfaz entre ambos organismos. Una visión seleccionista plantea un escenario de cambio constante sobre estos, evolucionando bajo un régimen de selección positiva [21,22]¹.

La capacidad de inmunomodulación de los productos secretados por los helmintos ha sido estudiada ampliamente en el contexto de infección a sus huéspedes mamíferos [23,24]. Sin embargo, el fenómeno también existe en la interacción con los huéspedes intermediarios. El sistema de defensa de moluscos se encuentra constituido por tres componentes principales: *i*) barreras anatómicas, *ii*) efectores humorales y *iii*) efectores celulares [25]. En la Figura 1.7A se representa el mismo en el contexto del sistema más estudiado, la interacción del trematodo *S. mansoni* y el sistema inmune del caracol *B. glabrata*, donde se ha demostrado la capacidad que poseen diversos productos secretados por los parásitos tremátodos de modular cada una de las barreras de los caracoles, facilitando el establecimiento de una infección persistente [26] (Figura 1.7 A).

Se han identificado algunos mediadores empleados por *S. mansoni* para la evasión del sistema inmune [27] (Figura 1.7 B), y mecanismos similares podrían ser utilizados por *F. hepatica* para la evasión en huéspedes intermediarios en el caso de infecciones persistentes.

Dentro de estos efectores humorales poseen un rol destacado las lectinas y moléculas de las familias FREP (con dominios de fibrinógeno) y TEP (con dominios tioéster similares a los de convertasas, algunas de las cuales son específicas de gastrópodos). Esto se encuentra bien establecido en el sistema *S. mansoni*-*B. glabrata* [26], y estudios transcriptómicos recientes muestran que *L. stagnalis* expresa genes de las familias FREP y TEP frente a infecciones de tremátodos que podrían incluir a fasciolidos [28], sugiriendo que se desplegarían procesos homólogos frente a la infección.

¹Este escenario es popularmente conocido como la “*hipótesis de la Reina Roja*”, en alusión al personaje del libro “*A través del espejo y lo que Alicia encontró allí*”. La Reina Roja postulaba que si uno pretende mantenerse en un mismo lugar debe mantenerse en constante movimiento, debido al movimiento rotatorio de la Tierra.

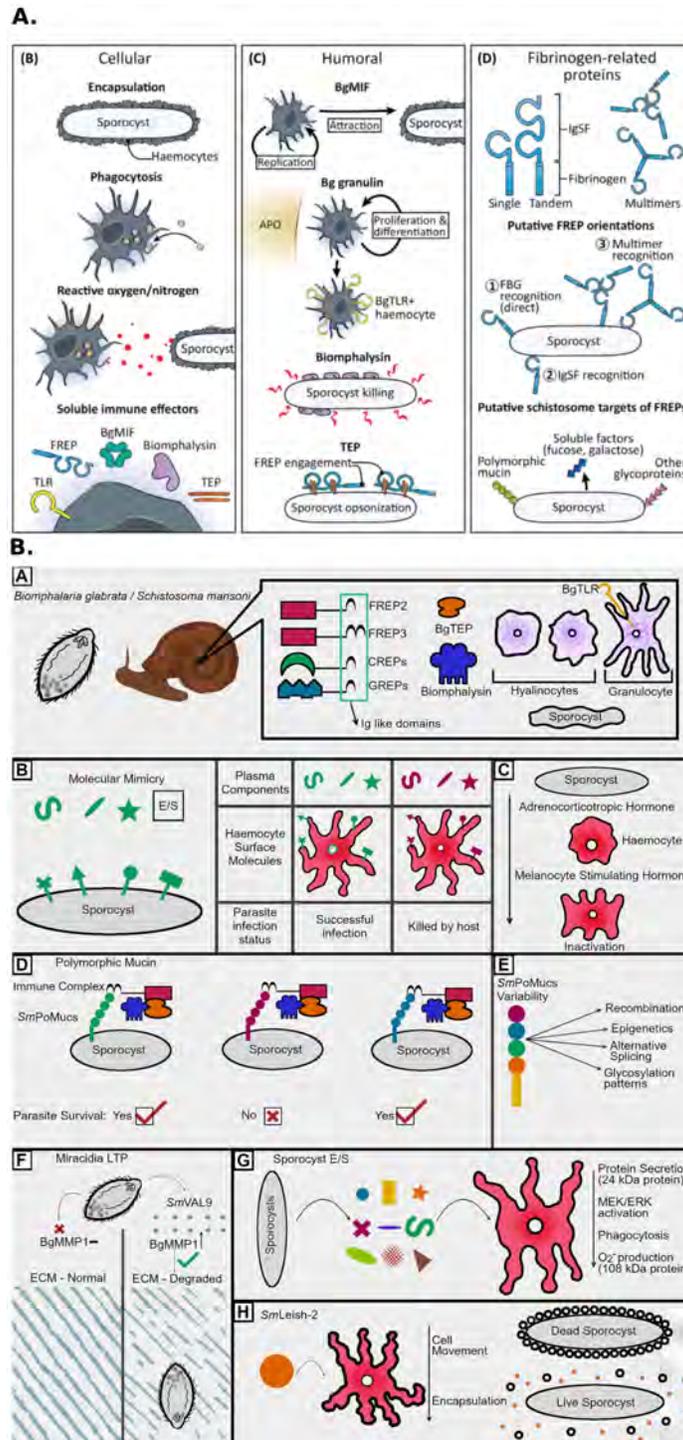


Figura 1.7: Mecanismos efectores del sistema inmune de *B. glabrata* relevantes en la respuesta a infección por *S. mansoni*. A) resumen de los mecanismos efectores principales. La interacción entre ambos organismos constituye uno de los modelos más estudiados de interacción entre un huésped caracol y un trematodo. B. Esquema indicando algunos de los principales componentes moleculares de la interacción *B. glabrata*-*S. mansoni*. Se resaltan algunos de los mecanismos de evasión del parásito, tales como la secreción de productos que inactivan a los componentes efectores del sistema inmune del caracol, o la mimetización con estructuras presentes en los hemocitos para evitar el reconocimiento. La hipervariabilidad a nivel de componentes glucídicos también es un mecanismo a resaltar, evitándose mediante el mismo el reconocimiento de efectores con dominios de tipo inmunoglobulina. Tomado de Pila et al. (2017).

De forma homóloga a lo que acontece en vertebrados, el sistema de complemento presente en invertebrados puede ser activado por la vía de las lectinas, llevando a la activación de la cascada proteolítica que inicia con la activación de la convertasa de C3 y la deposición del complejo de ataque a membrana en superficie de patógenos [25,26]. El despliegue de glicoproteínas con motivos miméticos a los usados por el huésped es uno de los mecanismos propuestos (ver Figura 1.7 B, comentado en [26]). El recambio de las glicoproteínas del tegumento ha sido reportado en Schistosomatidae, y se ha constatado que un proceso de recambio similar sucede en el pasaje de esporocistos a redias en *F. hepatica* [29–31]. No obstante, la capacidad de unión de lectinas a residuos de α -D-manosa y α -D-glucosa presentes en el tegumento de miracidios y esporocistos de *F. hepatica* ha sido reportada en la literatura [29,31], lo cual ha sido señalado como prueba de que el sistema de complemento podría ser activado. A pesar de poseer estos residuos potencialmente reconocidos por lectinas, los esporocistos logran establecer infecciones persistentes en caracoles, lo cual sugiere la existencia de otros mecanismos de evasión o resistencia a este componente humoral de la respuesta inmune innata de caracoles.

Por último, una barrera esencial de la inmunidad de los caracoles son los hemocitos, efectores celulares de la respuesta inmune en gastrópodos. Estas células efectoras son fundamentales para la eliminación de esporocistos mediante el mecanismo de encapsulación (Figura 1.7A). Durante este mecanismo de defensa los granulocitos (el morfotipo predominante descrito en la hemolinfa de caracoles a la fecha) rodean al parásito, liberando proteasas y especies reactivas de oxígeno y nitrógeno, las cuales ejercen un daño sobre los esporocistos impidiendo el establecimiento de la infección [25,32]. La importancia de los hemocitos y el mecanismo de encapsulación como respuesta a tremátodos se ve evidenciada en la comparación entre caracoles resistentes y susceptibles a infección: los primeros logran desplegar este mecanismo de forma efectiva, mientras que los tremátodos logran impedir el mecanismo en el segundo caso [25,27,33].

En resumen, el sistema inmune del huésped caracol posee capacidad de identificar y accionar mecanismos de respuesta frente a la infección por parte de tremátodos. En caracoles resistentes tal respuesta evita en las primeras 24hs el establecimiento de una infección persistente. No obstante, a lo largo de su co-evolución con sus huéspedes los tremátodos han desarrollado estrategias de evasión y modulación del sistema inmune de caracoles. Similar a lo observado en el caso de la infección de huéspedes vertebrados, proteínas de la interfaz huésped-parásito (secretadas y del tegumento) podrían poseer un rol en estos mecanismos de evasión y regulación. Si bien el sistema estudiado en mayor profundidad es el de *S. mansoni* y *B. glabrata*, estudios en *F. hepatica* muestran que estrategias similares podrían ser utilizadas también en esta especie.

La importancia de esta evasión y modulación del sistema inmune persiste durante toda la estadía en los huéspedes intermediarios, ya que durante esta acontece la ya mencionada proliferación asexual. La aparición de agentes externos al caracol genera una presión constante pues redias y cercarias también deberían ser objeto de reacción por parte de su sistema inmune.

1.5.4 Proliferación de células pluripotentes: un fenómeno fundamental en platelmintos

Los ciclos de vida de tremátodos son sumamente complejos. A lo largo de ellos se alternan formas parásitas y de vida libre, observándose distintas formas de reproducción (proliferación asexual y reproducción sexual) y cambios morfológicos drásticos con adaptaciones que aseguran la sobrevivencia en distintos ambientes y huéspedes. ¿Cómo es posible que se den estas variaciones marcadas durante la historia de vida de estos organismos a la vez que se retiene la identidad a nivel de especie?

En las últimas décadas se ha resuelto esta aparente paradoja al identificarse un linaje de células madre proliferantes transmitida entre estadíos, la cual da continuidad al material genético a lo largo del ciclo vida de varios platelmintos. Lo que es más, la diferenciación de estas células totipotentes permite el desarrollo y la homeostásis de los tejidos que conforman los diferentes estadíos de desarrollo. La especie en la cual se identificó por primera vez estas poblaciones celulares es *Schmidtea mediterranea* (generalmente conocida

como “*planaria*”). La impactante capacidad de regeneración de las planarias ha llamado la atención de biólogos durante por lo menos un siglo: ya en 1902 Morgan señaló la capacidad que posee esta especie de reconstituir su cuerpo completo a partir de un fragmento (comentado en [34]). Esta capacidad apunta hacia la existencia de una linaje celular común entre las células somáticas y la línea germinal de planarias, ya que la regeneración recompone incluso órganos sexuales y líneas germinales a partir de tejido somático. De hecho, la reproducción asexual en planarias se da por fisión y organogénesis, generando dos organismos a partir de uno.

Estudios posteriores confirmaron que una única línea celular totipotente y proliferante se encuentra presente en planarias, constituyendo ~30% de sus células. Estas células reciben el nombre de “*neoblastos*” [35]. Estudios posteriores han confirmado que la proliferación y diferenciación de neoblastos permite la especificación de las líneas somática y germinal en planarias [34]. La capacidad de autorenovación y la pluripotencia de los neoblastos se ve regulada por la expresión de un conjunto de genes ampliamente conservados en metazoarios, los cuales forman parte de lo que se ha denominado el “*programa de multipotencia de la línea germinal*” (*Germline Multipotency Program* en inglés, abreviado GMP) [36]. Las proteínas codificadas en genes componentes del GMP incluyen a reguladores post-transcripcionales de la expresión como *nanos*, *vasa* o *piwi*, remodeladores de la cromatina como integrantes del complejo Polycomb o factores de transcripción como *sox2*. La expresión localizada de estas proteínas permite regular a linajes celulares, asegurando un estado a nivel de cromatina y de expresión génica que permite mantener la no diferenciación (y así la multipotencia) de linajes multipotentes, a la vez que se activan los mecanismos para su autorenovación [36]. Los neoblastos de planaria expresan varios genes del programa GMP y son homólogos a reguladores expresados en células madre embrionarias de mamífero [34,37] o en líneas celulares proliferantes de especies de filos de metazoarios tan dispares como Porifera, Echinodermata, Ecdysozoa o Cnidaria [36].

Poblaciones proliferantes de células con morfología reminiscente a los neoblastos de planaria fueron descritos hace décadas en tremátodos, siendo estas históricamente denominadas “*células germinales*” [38]. Estas células germinales están presentes en miracidios, encontrándose masas de células con morfología similar en los estadios responsables de reproducción asexual como esporocistos o redias. La presencia de estas poblaciones reavivó la hipótesis de que células totipotentes podrían constituir un linaje celular transmitido a lo largo del ciclo de vida de tremátodos [39]. En la última década una serie de estudios realizados sobre el tremátodo *S. mansoni* confirmaron esta hipótesis. A través del marcado con EdU (análogo fluorescente de timidina que se incorpora durante la fase de síntesis de la replicación celular) fue posible identificar a las poblaciones proliferantes en miracidios, esporocistos y adultos de *S. mansoni* [14,40], habilitándose así los primeros perfiles transcriptómicos globales de estas poblaciones. Estas poblaciones celulares presentan similitudes morfológicas con los neoblastos de planaria: alta relación núcleo-citoplasma, presencia de gránulos densos a nivel electrónico de ribonucleoproteínas, nucleolo de gran desarrollo. Experimentos de “*pulso y caza*” (“*pulse-chase*”, en inglés) con EdU han permitido establecer que, también similar al aporte de homeostásis tisular observado en planarias, estas poblaciones diferencian aportando a la constitución y renovación de estructuras como el tegumento, un rasgo que sugiere adaptación de las mismas en la interfaz huésped-parásito [41]. Por último, el perfil transcriptómico de estas células madre tipo-neoblasto muestra que, además de expresar marcadores de proliferación celular como histonas *h2a*, *h2b*, *pcna* o *ciclina B*, las mismas expresan también genes del programa GMP (e.g. *vasa*, *nanos2*, *piwi*, *ago2-1*) sugiriendo otra reminiscencia clara a los neoblastos de planaria en cuanto a posibles mecanismos de regulación de su renovación y pluripotencia [14,40–42]. Un estudio de la cinética de proliferación demostró, no obstante, que estas poblaciones presentan cierta heterogeneidad en miracidios y esporocistos, asociada a la expresión de genes del programa GMP: sub-poblaciones con expresión del marcador *Sm-nanos2* presentan una proliferación más lenta que sus contrapartes *Sm-nanos2⁻* [14]. Esto sugiere que existe una diferenciación de poblaciones tipo-neoblastos en *S. mansoni* con posibles roles funcionales diversos. El reciente establecimiento de perfiles transcripcionales a resolución celular (mediante la técnica conocida como *single-cell RNA-seq* en inglés) para las poblaciones tipo-neoblasto para todos los estadios de *S. mansoni*

permitió confirmar la existencia de una heterogeneidad a nivel de estas células en correlación con los niveles de expresión de distintos genes del programa GMP entre otros. Además, también confirman que algunas de estas sub-poblaciones de células madre pluripotentes son transmitidas de un estadio a otro (Figura 1.8) [42].

En la Figura 1.8 se muestra el modelo presentado por Wang et al en dicho estudio.

En el mismo se observan las sub-poblaciones de células tipo-neoblasto de *S. mansoni* identificadas por su perfil transcripcional: células κ , δ y ϕ en los estadios larvarios asociados a huésped caracol, y células ϵ y δ' expresadas en estadios intramamífero.

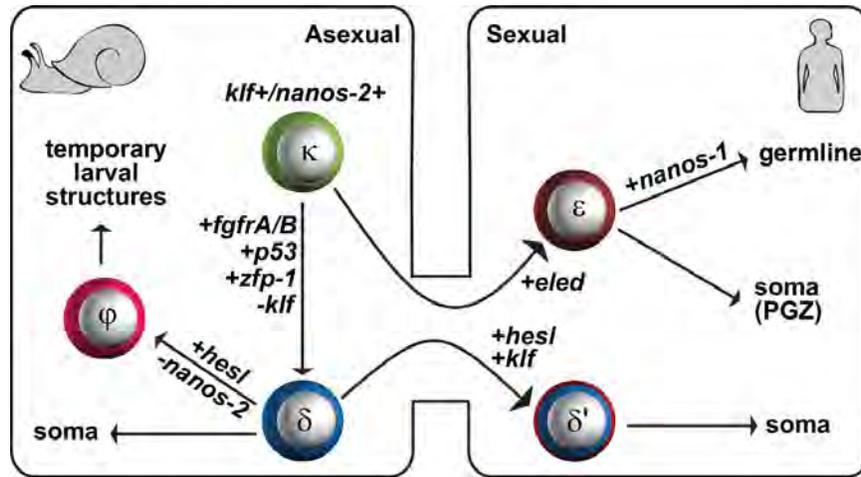


Figura 1.8: Modelo propuesto para la transmisión y diferenciación de poblaciones de células madre tipo-neoblasto a lo largo del ciclo de vida de *S. mansoni*. Se observan las sub-poblaciones de células tipo-neoblasto de *S. mansoni* identificadas por su perfil transcripcional: células kappa, delta y phi en los estadios larvarios asociados a huésped caracol, y células y expresadas en estadios intra-mamífero. Se indican algunos genes marcadores de cada tipo celular. Tomado de Wang et al. (2018).

Al igual que en el caso de planaria, el modelo plantea que células tipo-neoblasto se diferencian para dar lugar a las poblaciones somáticas que componen los tejidos de cada estadio de desarrollo (células δ y ϕ en esporocistos, y células δ' en huésped mamífero) y a la reproducción asexual en estadios larvarios (células κ) y por línea germinal (células ϵ) en adulto. Estudios posteriores de otros grupos en el mismo modelo *S. mansoni** han confirmado estas observaciones [42,43].

Siguiendo el modelo propuesto en *S. mansoni*, estudios recientes han establecido el perfil transcriptómico global de estadios larvarios asociados a caracol de varios tremátodos, observando expresión génica de genes marcadores asociados al programa GMP en *S. japonicum* [15], y *P. simillimum* y *S. pseudoglobulus*. Estos dos últimos pertenecen a la familia Psilostomatidae, del suborden Echinostomata, al igual que Fasciolidae. En el caso de *F. hepatica* se ha establecido la existencia de poblaciones tipo-neoblasto en estadio juvenil, también presentando expresión de algunos genes GMP [44].

Estos estudios sugieren que las células madre tipo-neoblasto formarían una parte integral conservada del ciclo de vida de tremátodos en general. Lo que es más, la observación de poblaciones similares en organismos como *E. granulosus* [45] extiende esto a cestodos. De ser así, fenómenos observados hace décadas en este filo tales como la regeneración en planarias, la proliferación asexual en tremátodos y cestodos y el recambio constante de tejidos frente a la respuesta de hospederos podrían haber evolucionado a partir de la adaptación de líneas celulares existentes en algún ancestro común de los platelmintos. Esto sugeriría que la aparición de los ciclos complejos de tremátodos podría tener base en esta adaptación de linajes existentes en platelmintos de vida libre [14,41,42]. Est aspectos de la biología de *F. hepatica* permanecen inexplorado a nivel transcriptómico, en tanto no se posee información asociada a los estadios

larvarios a la fecha, una brecha que este trabajo busca cubrir.

1.6 Hipótesis de trabajo

La regulación de la expresión génica constituye un mecanismo relevante para la adaptación y sobrevivencia de los estadios de desarrollo de *F. hepatica* asociados a su huésped intermediario caracol. Poseen un rol central aquellos genes codificantes para productos vinculados a procesos de desarrollo o implicados en la interfaz huésped-parásito.

1.7 Objetivo general

Contribuir al entendimiento de los procesos que permiten la adaptación y sobrevivencia de los estadios de desarrollo larvarios de *F. hepatica* asociados a su huésped intermediario caracol, mediante el ensamblado y análisis de datos transcriptómicos y análisis comparativo con datos de expresión génica de otros tremátodos.

1.8 Objetivos específicos

- **Objetivo Específico I.** Analizar datos transcriptómicos de estadios asociados a huésped intracaracol (miraciredios y estadios iacaracol 15, 21 y 30 días post-infección cercarias) de *F. hepatica* en el huésped intermedio *Lymnaea viatrix*.
- **Objetivo Específico II.** Detectar genes diferencialmente expresados entre estadios y otros estadios, analizando en detalle la expresión de genes codificantes para proteínas de la interfaz huésped-parásito y aquellos involucrados en procesos de desarrollo.
- **Objetivo Específico III.** Comparar los genes diferencialmente expresados en estadios asociados a huésped caracol en los tremátodos *F. hepatica* y *F. gigantica*, con énfasis en el estudio de proteínas con marcada expresión diferencial.

Capítulo 2

Materiales y Métodos

2.1 Obtención de material biológico y secuenciado

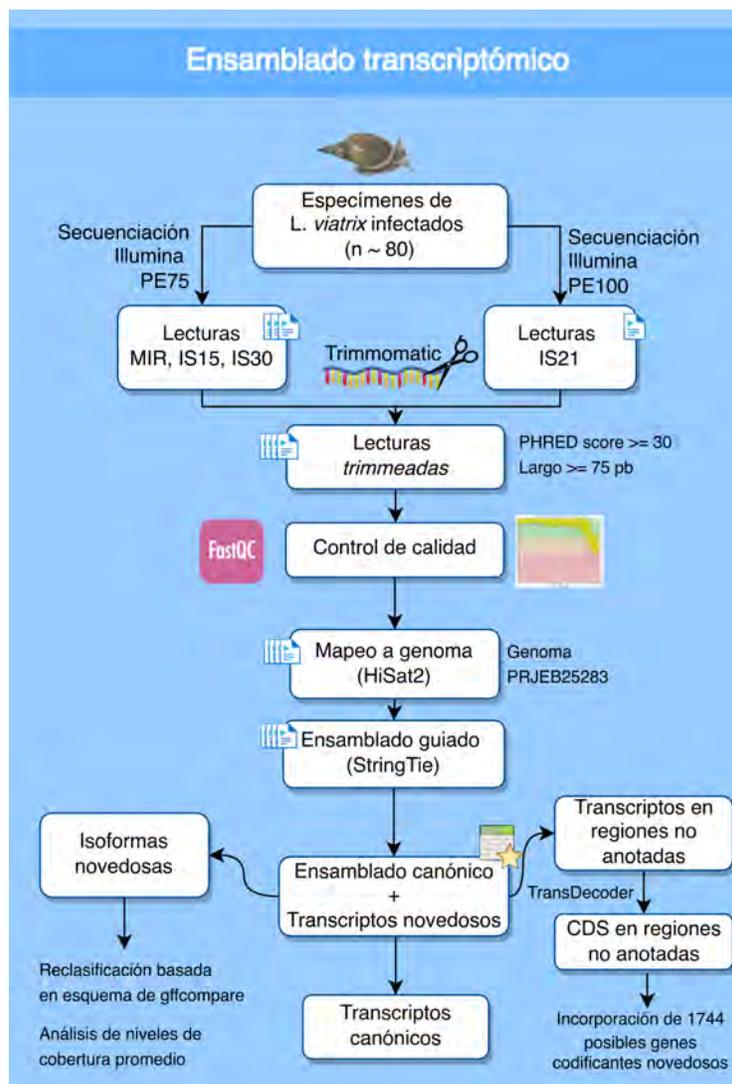


Figura 2.1: Esquema general de procesamiento de lecturas, control de calidad y ensamblado transcriptómico utilizado en este estudio. En las subsecciones 2.1, 2.2 y 2.3 se describen en detalle los métodos representados en este esquema.

Las muestras empleadas en esta tesis de Maestría fueron obtenidas a partir de infecciones experimentales de caracoles *Lymnaea* mantenidos en el laboratorio por la Br. Sabina Wlodeck. A partir de disecciones de caracoles infectados a los tiempos indicados se extrajo el material parasitario (2.2). El ARN total fue extraído por la Dra Fernanda Domínguez utilizando el Direct-Zol RNA kit. Un total de 11 muestras fueron tomadas, siguiendo el esquema indicado en la Tabla 5.1. Las muestras fueron enviadas para su secuenciación en plataforma Illumina al Centro Sanger, utilizando los kits standard (PE75) o a Macrogen utilizando el kit TruSeq Stranded mRNA LT Sample Prep Kit (PE100).

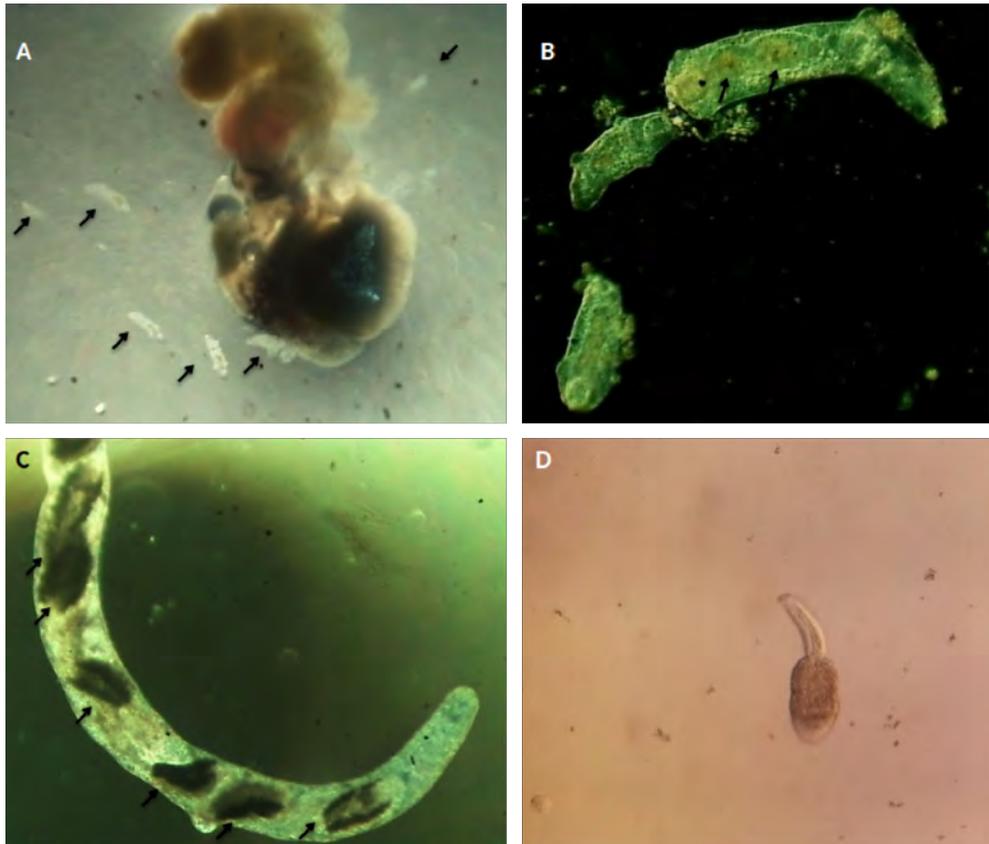


Figura 2.2: Disección de caracoles y ejemplares recuperados, incluyendo aquellos secuenciados en este estudio. A. Caracol diseccionado con 15 días infección con parásitos emergentes del caracol (indicados con flechas). B. Parásitos representativos de estadios tempranos de infección. Se observan 3 redias, identificándose la presencia de masas de células proliferativas en el interior de una de ellas (flechas) C. Redia conteniendo cercarias (flechas) en su interior, representativo de ejemplares de estadios tardíos de infección. D. Cercaria libre presente en estadios tardíos de infección.

2.2 Procesamiento de las lecturas y control de calidad

El control de calidad de las secuencias se realizó con FastQC v0.11.8 [46]. Los resultados fueron recopilados en un único reporte con MultiQC v1.9 ([47]; archivo adjunto *trimmed_samples_multiqc_report.html*). Con el objetivo de trabajar con lecturas de alta calidad, se procesaron las mismas con Trimmomatic v0.38 [48], especificándose: (i) la eliminación de adaptadores de Illumina, empleados en el proceso de secuenciación (ii) la eliminación de secuencias de inicio y finales de baja calidad (*i.e.* un PHRED-33 score menor a 30), (iii) el escaneo de la secuencia con una ventana deslizante de un largo de 4pb, cortándose la secuencia cuando la calidad promedio fuera menor a un PHRED-33 score de 30; y (iv) la eliminación de secuencias

con un largo final menor a 75 pb.

2.3 Ensamblado transcriptómico para estadíos miracidio e intracaracol de *F. hepatica*

Se obtuvo un ensamblado transcriptómico para las muestras de miracidio y estadíos intra-caracol. Existen a la fecha dos estrategias para llevar a cabo esta tarea: *i*) ensamblado transcriptómico guiado (con referencia), y *ii*) ensamblado transcriptómico *de novo*. Se optó por la primera de estas estrategias, la cual se detalla en lo que sigue de esta sub-sección.

El ensamblado transcriptómico fue realizado utilizando como referencia el genoma PRJEB25283 de *Fasciola hepatica* depositado en WormBase ParaSite. Este es un ensamblaje mejorado por lecturas largas de PacBio respecto al originalmente publicado [49]. Contamos localmente con otro ensamblaje genómico con una anotación mejorada por integración de datos genómicos, transcriptómicos y de homología [50], basado en el genoma publicado por McNulty *et al.* [51] (PRJNA179522). Se utilizó este segundo ensamblaje (con mejor anotación, pero más fragmentado) para comparación en algunos casos puntuales de interés.

Las lecturas procesadas fueron mapeadas con HiSat2 v2.0.5 [52] especificando la opción `--rna-strandness RF` para orientación de lecturas y el resto de los parámetros por defecto. Los alineamientos obtenidos en formato SAM fueron ordenados y convertidos a archivos binarios BAM con samtools v1.9 [53]. Los ensamblados transcriptómicos fueron obtenidos con StringTie v1.3.6 [54] especificando opción `--rf` para tener en cuenta orientación de las lecturas y el resto de los parámetros por defecto, a partir de los alineamientos BAM y la anotación en formato GFF del genoma de *F. hepatica*.

Los transcriptos ensamblados fueron caracterizados utilizando gffcompare v0.11.2 [55]. Dicha herramienta permite la comparación de los transcriptos inferidos con la anotación de referencia (*i.e.* la del genoma contra el que se mapea), identificándose transcriptos ya descritos, posibles isoformas novedosas y transcriptos desconocidos (ver ccb.jhu.edu/software/stringtie/gffcompare.shtml y Figura 3.2A para más detalles). Dicha herramienta permite a su vez la integración de las anotaciones para los transcriptos inferidos para todas las muestras y la anotación genómica disponible en la base de datos (*i.e.* el genoma de referencia empleado).

De especial interés resultan los transcriptos codificados en regiones intergénicas de la anotación de referencia (*i.e.* categoría “*u*” del esquema de categorización de gffcompare). Entre las mismas se incluyen posibles genes previamente no predichos y/o identificados con información transcriptómica anterior. Se anotaron secuencias codificantes (CDS; *coding DNA sequences* en inglés) para los ensamblados transcriptómicos obtenidos utilizando TransDecoder v5.5.0 (github.com/TransDecoder/TransDecoder) con parámetros por defecto. Las secuencias proteicas codificadas en transcriptos de categoría “*u*” fueron incluidas en análisis de secuencias proteicas posteriores, utilizándose en todos los casos las secuencias de mayor largo.

Aquellas lecturas que no mapearon contra el genoma de referencia de *F. hepatica* fueron re-mapeados contra el genoma de referencia de *L. stagnalis* (código BioProject PRJEB11470 en NCBI) para evaluar si los mismos provienen de contaminación del huésped hospedero.

La elección de tal genoma responde a la falta de un genoma de referencia para *L. viatrix* a la fecha, siendo *L. stagnalis* el único ensamblado genómico disponible para un caracol del género *Lymnaea*. Teniendo en cuenta la posible existencia de sustituciones debida a la distancia evolutiva entre *L. viatrix* y *L. stagnalis*, tal mapeo se realizó modificando los parámetros de penalización de alineamientos de HiSat2 según tres de los esquemas empleados por Baruzzo *et al.* [56].

2.4 Análisis de expresión génica

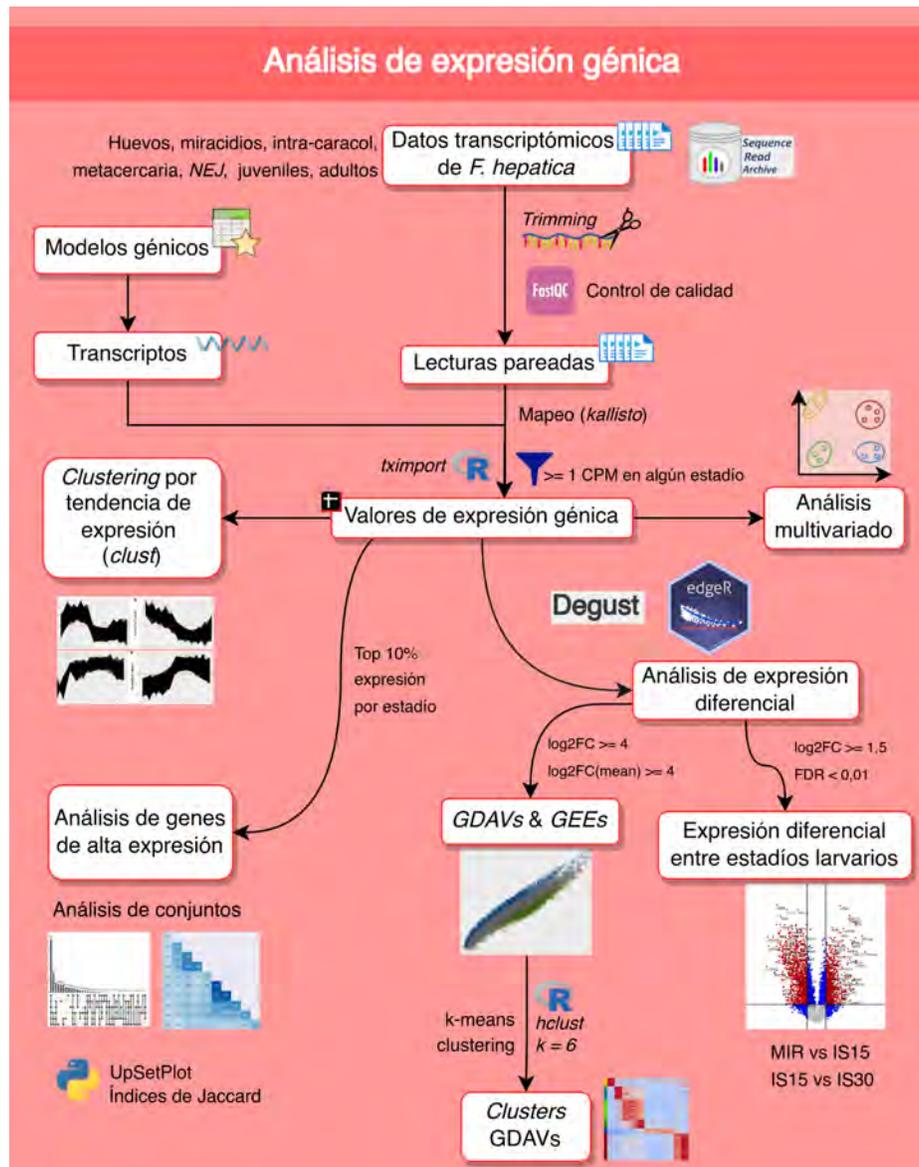


Figura 2.3: Esquema general de los análisis de la expresión génica realizado en este estudio. En las subsecciones 2.4, 2.7 y 2.8 se describen en detalle los métodos representados en este esquema.

Los niveles de expresión génica fueron analizados teniendo en cuenta la información transcriptómica disponible para el ciclo de vida de *F. hepatica* a la fecha (ver Tabla 5.1, donde se enlistan las muestras empleadas). Las lecturas pareadas de los estadios disponibles en bases de datos públicas fueron procesadas y sujetas a un control de calidad análogo al descrito en la Sección 2.2.

A efectos de considerar en el análisis de expresión génica tanto los modelos génicos disponibles en la anotación de referencia como aquellos a partir de las muestras novedosas (Sección 2.3), se decidió combinar ambas anotaciones utilizando gffcompare. El conjunto de lecturas procesadas para todas las muestras de *F. hepatica* fue entonces mapeado contra este conjunto de transcritos con *kallisto* v0.46.2 [57]. Se optó por dicha herramienta debido a su eficiencia computacional, tanto del punto de vista de rapidez de cómputo como de almacenamiento.

Los niveles de expresión génica fueron estimados a partir del conteo para cada transcripto con la librería tximport v4.0.3 de R [58], utilizando la anotación combinada descrita anteriormente. La normalización de la expresión génica en transcriptos por millón (TPM, *transcript per million* en inglés) fue hecha con dicha librería, así como en *conteos por millón* (CPM, *counts per million* en inglés) con la función cpm() de la librería edgeR v3.24.3 [59,60] con normalización por método TMM (“*Trimmed mean of M-values*”, en inglés) [61] para posterior análisis de expresión diferencial. La normalización es un paso necesario para eliminar sesgos debidos a tamaños distintos de las bibliotecas correspondientes a cada estadio o réplica.

Se consideró para análisis posteriores a todo gen que tuviera al menos 1 CPM en todas las réplicas de una misma condición biológica. Tal criterio buscó contemplar la heterogeneidad existente entre los estadios del ciclo de vida de *F. hepatica*.

Con el fin de analizar la similitud a nivel transcriptómico de los diferentes estadios estudiados se estimó la dispersión intra- e inter-muestral a nivel de la expresión génica mediante un escalado multi-dimensional (MDS; *Multidimensionalscaling* en inglés). El análisis fue realizado en el servidor Degust (degust.erc.monash.edu), el cual realiza un MDS a partir del logaritmo los conteos de cada gen, a fin de estabilizar la varianza en los valores de expresión y evitar un aporte desmedido de los genes de mayor expresión.

Como una primera aproximación al estudio de las tendencias de expresión génica a lo largo del ciclo de vida de *F. hepatica* se infirieron agrupamientos (*clusters*) de co-expresión con el programa *clust* v1.12.0 [62], utilizando valores de expresión normalizados (TPM) y parámetros por defecto y especificación de parámetro de ajuste (“*tightness*”) de los *clusters* $t = 4$.

Por último, comparaciones respecto a conjuntos de genes expresados entre diferentes estadios de *F. hepatica* fueron realizadas a partir del cálculo índice de Jaccard.

Siendo A y B dos conjuntos a comparar, el índice de Jaccard entre ambos se define como [63]

$$\text{Índice de Jaccard} = \frac{A \cap B}{A \cup B}$$

De esta forma, como casos extremos tenemos: *i*) dos conjuntos totalmente disjuntos, cuyo índice de Jaccard es 0, y *ii*) dos conjuntos iguales, cuyo índice de Jaccard corresponde a 1. Naturalmente, casos intermedios presentarán un índice de Jaccard entre 0 y 1, con valores más altos reflejando mayor similitud entre los conjuntos A y B .

En el caso de comparaciones realizadas entre estadios de desarrollo de *F. hepatica*, se realizó dicha comparación a nivel de expresión génica. Así, si se comparó el estadio de desarrollo “A” con el “B”, el conjunto A definido arriba corresponde a los genes expresados en dicho estadio de desarrollo (*i.e.* niveles de expresión ≥ 1 CPM en todas las muestras del estadio). Así, en la definición de Índice de Jaccard el término $A \cap B$ equivale al número de genes para los que se detecta expresión en ambos estadios, mientras que $A \cup B$ corresponde al número de genes para los cuales se detectó expresión, ya sea en el estadio A o en el B .

2.5 Anotación funcional

Las secuencias proteicas codificadas en el conjunto de genes analizados fueron anotadas funcionalmente con e-mapper tool v2.1.5 [64], corriendo contra la base de datos eggNOG v5.0 [65,66]. Se realizó dicha anotación especificando como nivel taxonómico “Metazoa” (parámetro *-target_taxa*), a fin de evitar traspaso de términos funcionales provenientes de taxas evolutivamente muy lejanos en donde una proteína pudiera estar vinculada a otros roles. El traspaso de anotación desde las secuencias homólogas a la secuencia anotada fue realizado bajo la modalidad *--target_orthologs all*, donde se restringe la anotación a

homólogos uno a uno con la secuencia a ser anotada. Se acota de esta forma el traspaso de información funcional de homólogos taxonómicamente distantes.

Se trasladó así la anotación funcional de las bases de datos Gene Ontology [67], KEGG [68], PFAM v34.0 [69,70] y COG [71] (Bacteria), KOG (Eukarya) [72] y arKOG (Archaea) [73]. Una segunda estrategia de anotación fue la búsqueda de dominios en la base de datos InterPro [74] utilizando la herramienta InterProScan v5 [75] corriendo con parámetros por defecto. Las anotaciones obtenidas se agregaron y se encuentran disponibles en la tabla *fhepatica_ anotacion_funcional.tsv* del *Material suplementario* a esta tesis.

Esta anotación funcional automática posee un valor en sí misma, siendo especialmente valiosa en el caso de genes pobremente estudiados y/o en posibles modelos génicos novedosos (genes que naturalmente no están funcionalmente anotados por no existir en las bases de datos actuales que contienen el genoma de *F. hepatica*). A su vez, la misma fue empleada para realizar análisis de enriquecimiento funcional de ontología génica y vías metabólicas para diferentes conjuntos de genes de interés (e.g. genes de mayor expresión en un determinado estadio, genes diferencialmente expresados entre un par de estadios).

2.6 Análisis funcional de la expresión génica

Siguiendo a Zhang *et al.* [16] se realizó una primera aproximación al análisis funcional de la expresión génica agrupando genes según su clasificación funcional y observando las tendencias de expresión de estos grupos funcionales a lo largo del ciclo de vida. Brevemente, se sumaron los valores de expresión de genes pertenecientes a la misma categoría BRITE de la base de datos KEGG, y se graficó el *z-score* de $\log(\text{CPM}+1)$ con la función *clustermap* de la librería *seaborn* de Python [76].

Análisis de enriquecimiento funcional fueron realizados para grupos de genes de interés, tales como genes de mayor expresión en un estadio, genes que presentan expresión diferencial significativa entre pares de estadios o de expresión estadio-específica. En todos los casos, la búsqueda de enriquecimiento funcional para el conjunto de genes de interés se realizó utilizando como referencia el conjunto de genes expresados en el estadio (*i.e.* aquellos que presentan > 1 CPM en todas las muestras del mismo). El uso de este conjunto de referencia permite evitar la aparición artefactual de términos enriquecidos debido al sesgo de expresión que posee cada estadio (comentado en [77]).

En el caso de enriquecimiento funcional para ontologías génicas (“*GO terms*”, en inglés), el enriquecimiento funcional fue testeado con la librería GOATOOLS de Python [78] con un test exacto de Fisher con ajuste FDR (*false discovery rate*, en inglés) de Benjamini-Hochberg para control de falsos positivos [79]. Se reportan como enriquecidos aquellos términos a nivel de significancia $\text{FDR} < 0,05$.

Los términos GO reportados como enriquecidos fueron agrupados por similitud semántica y graficados en un *heatmap*, una aproximación que ya ha sido descrita en la literatura (*e.g.* [80]). La agrupación semántica se realizó siguiendo el procedimiento descrito en un protocolo desarrollado en GOATOOLS, disponible en su repositorio de GitHub, el cual está basado en el coeficiente de similitud semántica [81] (github.com/tanghaibao/goatools/blob/main/notebooks/semantic_similarity_wang.ipynb). La representación gráfica se restringió a aquellos términos GO enriquecidos en estadios de miracidio y estadios intracaracol pertenecientes a la ontología de procesos biológicos, contemplando aquellos términos de especificidad moderada (*i.e.* nivel de ontología entre 4 y 9). Se realizó un *heatmap* con la función *clustermap* de *seaborn* [76], graficando en cada caso los valores de enriquecimiento calculados ($-\log_{10}(p_{\text{valor}})$) para la totalidad de estadios de *F. hepatica*, a efectos de comparar el nivel de conservación de los patrones observados. Por último se agregó manualmente una clasificación funcional de los términos enriquecidos a efectos de facilitar la visualización de los resultados.

Los análisis de enriquecimiento funcional para vías y módulo metabólicos de KEGG fueron llevados a

cabo con la librería `gprofiler2` v0.2.0 de R [82], utilizando la función `gost()` y $\text{FDR} < 0,05$ como valor de corte de significancia estadística.

2.7 Análisis de expresión diferencial

Los análisis de expresión diferencial fueron llevados a cabo en el servidor *Degust* ([<degust.erc.monash.edu/>](http://degust.erc.monash.edu/)) [83] que implementa herramientas de la librería `edgeR`. Se define a un gen como diferencialmente expresado (“*GDE*”) entre dos estadíos si se observa $|\log_2\text{FC}| \geq 2$ y $\text{FDR} < 0,01$. En todos los casos se realiza el análisis sobre el conjunto de genes expresado en alguna de las condiciones a comparar (*i.e.* $\text{TPM} > 1$ para todas las muestras del estadío).

Se llevaron a cabo dos análisis de expresión diferencial:

i) Análisis de genes diferencialmente expresados (GDEs) entre estadíos de desarrollo asociados al huésped caracol. Se detectaron *GDEs* entre las muestras de miracidio, estadío intracaracol 15 d.p.i. y 30 d.p.i. siguiendo la progresión del ciclo de vida del parásito. Se excluyó del análisis a las muestras de estadío intracaracol 21 d.p.i., en tanto las mismas provienen de otro *batch* de secuenciación.

Los *GDEs* detectados entre estadíos fueron representados mediante un *volcano plot* generado con la librería `EnhancedVolcano` v1.4.0 (github.com/kevinblighe/EnhancedVolcano), resaltando genes descritos como relevantes para procesos biológicos de estadíos de desarrollo de *F. hepatica* u otros parásitos. A su vez se realizó un análisis de enriquecimiento en vías KEGG para los conjuntos de *GDEs* detectados.

ii) Definición de genes diferencialmente expresados de alta variabilidad En búsqueda de perfiles de expresión estadío-específicos que pudieran brindar información acerca de genes con posible relevancia en cada estadío de desarrollo, se identificó *GDEs* con tendencia a la expresión diferencial estadío-específica. Se consideraron los niveles de expresión a lo largo de todos los estadíos de desarrollo. Aproximaciones similares han sido realizadas estudiando la expresión génica diferencial en *F. gigantica* [16] y *S. japonicum* [15]. Teniendo en cuenta la heterogeneidad del set de datos analizados se decidió aplicar un criterio restrictivo para disminuir la aparición de falsos positivos. Se define a un *GDE* como “*gen diferencialmente expresado de alta variabilidad*” (“*GDAV*”) si su nivel de expresión es al menos 16 veces superior en algún estadío respecto al promedio del resto. Los *GDAV* fueron agrupados mediante algoritmo de *k*-medias empujando la función `hclust` de la librería `stats` v3.6.1 de R. Se especificó un número de *clusters* $k = 6$, teniendo en cuenta los agrupamientos observados en el análisis de escalado multidimensional. Con fines comparativos el mismo procedimiento fue realizado para detectar grupos de *GDAV* en *F. gigantica*, empleando los datos publicados por Zhang *et al.* [16].

2.8 Análisis de genómica comparativa

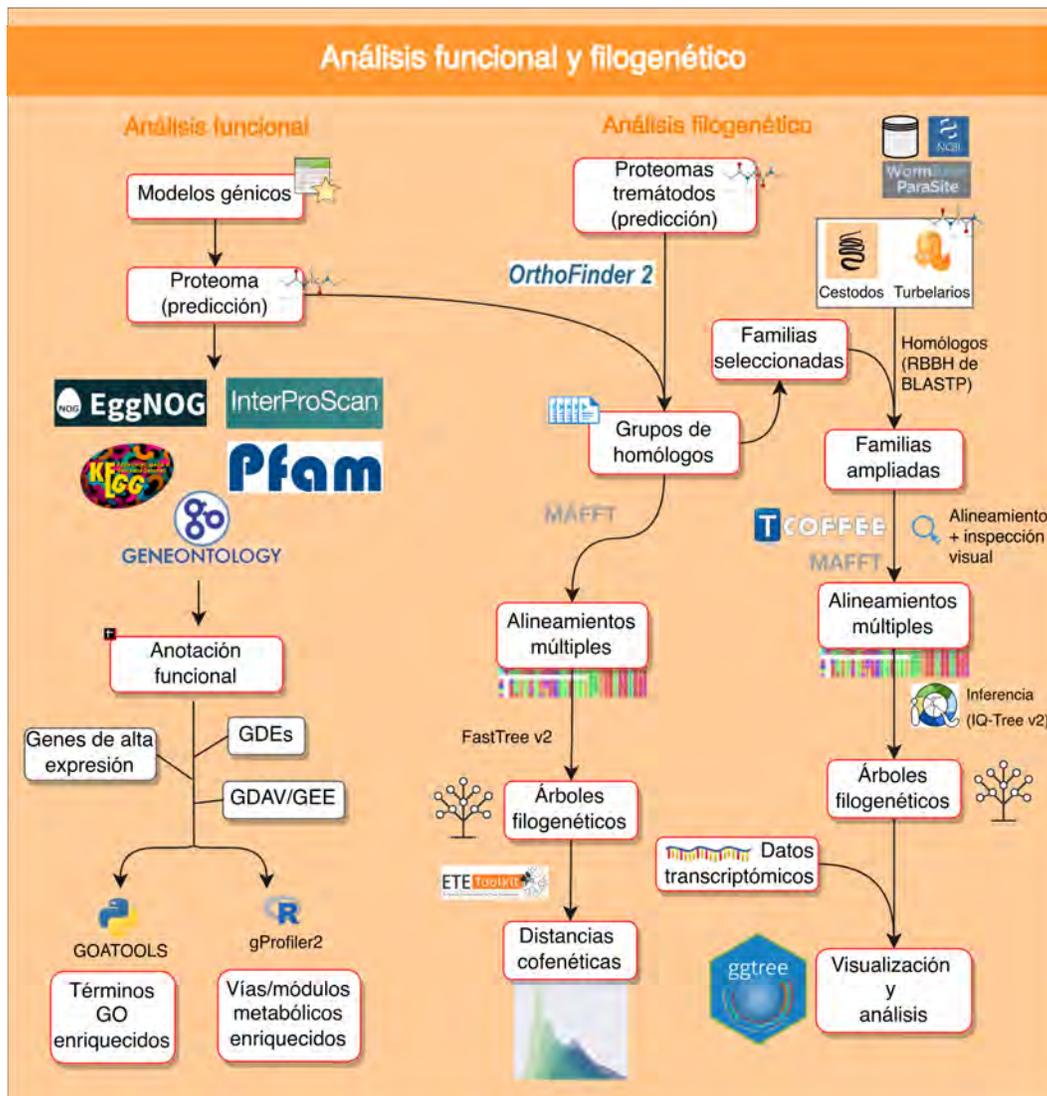


Figura 2.4: Esquema general de los análisis funcionales y filogenéticos realizados en este trabajo. En las subsecciones 2.5 y 2.6 se describen en detalle métodos referentes al análisis funcional representados en este esquema. En la subsección 2.8 se describen métodos relacionados al análisis filogenético realizado en el trabajo.

2.8.1 Obtención de secuencias, inferencia de grupos de homólogos y árboles filogenéticos

Se descargaron los proteomas predichos de los tremátodos disponibles en WormBase ParaSite v15 (parasite.wormbase.org/index.html) y del ensamblado GCA_008360955.1 de *F. buski* disponible en NCBI. Grupos de homólogos fueron inferidos utilizando OrthoFinder v2 [84] y los alineamientos múltiples de proteínas fueron realizados con MAFFT v7.429 [85] bajo la modalidad L-INS-I (recomendada para mayor exactitud). Debido a la cantidad de grupos analizados la inferencia filogenética fue realizada con FastTree v2.1.10 [86] corriendo con parámetros por defecto.

En el caso de algunas familias seleccionadas en base a literatura, se realizó un análisis más detallado. Se

incluyeron secuencias homólogas de cestodos (*E. granulosus*, PRJEB121; *E. multilocularis*, PRJEB122; *T. solium*, PRJNA170813; *M. corti*, PRJEB510; *H. microstoma*, PRJEB124) y platelmintos de vida libre (*S. mediterranea*, PRJNA379262; *M. lignano*, PRJNA371498) obtenidas de WormBase ParaSite v15. La detección de homólogos se realizó mediante alineamientos pareados de BLASTP v2.11.0+ [87], definidos como *hits* bidireccionales que presenten un e-value $\leq 1e^{-30}$. A fin de reducir la tasa de falsos positivos, solo se tomaron en cuenta aquellos alineamientos donde se detectó porcentaje de identidad aminoacídica de al menos 35% con una cobertura de secuencia de al menos 70% [88]. Se realizaron alineamientos múltiples con MAFFT, los cuales fueron evaluados corriendo T-COFFEE v12.00.7fb08c2 [89] con la opción `-score` y analizados visualmente. Secuencias con un valor de score de conservación menor a 50 fueron eliminadas, y los alineamientos fueron re-calculados. La inferencia de árboles filogenéticos fue realizada por el método de máxima verosimilitud con IQ-Tree v2.1.2 [90]. Se especificó la selección de modelos evolutivos con ModelTest [91] utilizando la opción `-m TEST`, 1000 réplicas para evaluación de largo ramas por test SH-aLRT y 1000 réplicas de los árboles por *ultrafast bootstrap* [92].

2.8.2 Análisis comparativo de genes según sus tendencias de expresión

Con el fin de realizar análisis comparativos de grupos con distintas tendencias de expresión, se definió un conjunto de genes con expresión relativamente estable a lo largo del ciclo de vida. Una medida empleada para cuantificar la variación de un gen es el *coeficiente de variación* ($CV = \frac{\text{desvío estándar}}{\text{promedio}}$) de sus valores de expresión (usado por ejemplo en [93,94]). Se definió a un gen de expresión estable (“*GEE*”) si presenta un $CV < 0,5$.

Dos tipos de análisis comparativos se realizaron entre *GDAVs* y *GEEs*:

- Conservación (*i.e.* presencia/ausencia) de cada grupo tanto para *F. hepatica* como para *F. gigantica*. Brevemente, se descargaron los grupos de homólogos conservados en Metazoa (*i.e.* presencia en $\geq 90\%$ de las especies del clado) de la base de datos OrthoDB v10 [95]. Se buscaron secuencias ortólogas empleando la librería *homologr* (github.com/drostlab/homologr/) de R, especificando método de mejor hit recíproco en alineamientos de DIAMOND [96] y especificando un valor mínimo de 40% de identidad y una cobertura de al menos 70% de la secuencia *query*.
- Divergencia a nivel de secuencia entre genes. Se compararon distancia cofenéticas para genes homólogos de *F. hepatica* y *F. gigantica*. Las distancias obtenidas para distintos pares de secuencias fueron clasificadas según involucraran a *GDAVs* de las dos especies, *GEEs* de las dos especies u otros tipos de comparaciones. La estimación de distancia cofenética se realizó empleando los árboles génicos antes descritos, y utilizando el método `.get_distance()` de la librería *ete3* v3.1.2 [97] de Python.

Por último, se analizó la concordancia a nivel de genes homólogos entre los *GDAVs* de *F. hepatica* y *F. gigantica*, utilizando como métrica el índice de Jaccard. Brevemente, se ve la proporción de grupos de homólogos que poseen un miembro detectado en ambos *GDAVs* (numerador) respecto al total de grupos de homólogos detectado en el par (denominador). Una aproximación similar fue realizada por Nesterenko *et al.* para estimar la similitud entre estadios de tremátodos tomando en cuenta si expresan genes ortólogos [17].

2.9 Material suplementario

El material suplementario a esta tesis se encuentra disponible en el siguiente repositorio de GitHub: github.com/mauriciolangeib/fhepatica_snailseq

Capítulo 3

Resultados y Discusión

3.1 Transcriptoma de miracidio y estadíos intra-caracol de *F. hepatica*

3.1.1 Control de calidad de secuencias

Antes de proceder al mapeo y análisis de la expresión génica es necesario realizar controles de calidad con el fin de asegurar que el proceso de secuenciación masiva fue exitoso. Las lecturas de secuenciación obtenidas fueron procesadas con el fin de remover de los extremos bases de mala calidad empleando la herramienta Trimmomatic (ver Sección 2.2 de *Materiales y Métodos*).

El análisis de calidad realizado con FastQC sobre las lecturas procesadas con Trimmomatic muestra la obtención de lecturas pareadas de alta calidad (ver resumen de todos los reportes en el archivo "*trimmed_samples_multiqc_report.html*" del *Material suplementario* a esta tesis). Los mismos poseen un PHRED-score de más de 30 a lo largo de todas sus bases, lo cual implica una tasa de error de 0.1% en las bases secuenciadas [98]. Se logró obtener entre 16.80 a 92 millones de lecturas pareadas de alta calidad, constituyendo desde 71.33% al 99.96% de las lecturas originales (ver Tabla 3.1). La existencia de secuencias duplicadas y sobre-representadas reportada por el programa es de esperar en un experimento de RNA-seq, donde existe heterogeneidad en los niveles de expresión de distintos genes. Lo que es más, es sabido que existen familias multigénicas amplificadas en *F. hepatica* codificantes para proteínas expresadas en estadíos previamente estudiados [49,99,100], por lo que la expresión simultánea de varios de sus miembros podría explicar en parte el fenómeno .

En resumen, el procesamiento del material secuenciado permitió la obtención de lecturas de alta calidad para las muestras de miracidio y estadíos intra-caracol.

Tabla 3.1: Estadísticas de procesamiento y mapeo de muestras de miracidio y estadíos intracaracol de *F. hepatica* secuenciadas para este estudio.

Estadío de desarrollo	Muestra	Núm. reads inicial	Núm. reads procesados	Reads procesados (%)	Tasa de mapeo (HiSat2)	Tasa de mapeo (kallisto)
Miracidia	MIR-1	7.11e7	7.11e7	99.96%	83.58%	81.10%
Miracidia	MIR-2	7.94e7	7.93e7	99.97%	88.73%	85.80%
Intra-caracol (15 d.p.i.)	IS15-1	6.92e7	6.92e7	99.96%	78.05%	76.00%
Intra-caracol (15 d.p.i.)	IS15-2	9.29e7	9.29e7	99.96%	79.47%	78.60%
Intra-caracol (21 d.p.i.)	IS21-1	2.20e7	1.68e7	76.42%	85.53%	83.20%
Intra-caracol (21 d.p.i.)	IS21-2	3.27e7	2.50e7	76.43%	86.97%	85.40%
Intra-caracol (21 d.p.i.)	IS21-3	2.39e7	1.70e7	71.33%	83.24%	82.00%
Intra-caracol (30 d.p.i.)	IS30-1	7.79e7	7.79e7	99.95%	73.82%	73.00%
Intra-caracol (30 d.p.i.)	IS30-2	7.13e7	7.13e7	99.94%	73.94%	76.50%
Intra-caracol (30 d.p.i.)	IS30-3	8.13e7	8.12e7	99.95%	80.08%	79.80%
Intra-caracol (30 d.p.i.)	IS30-4	8.58e7	8.58e7	99.96%	72.44%	72.20%

3.1.1.1 Estadísticas de mapeo

El mapeo inicial de muestras contra el ensamblado genómico PRJEB25283 de *F. hepatica* revela una tasa de mapeo que va desde 72.44% al 88.73% para la totalidad de muestras de miracidio y estadíos intra-caracol secuenciados (Tabla 3.1), lo cual resulta muy razonable¹. Las diferencias observadas se asocian al distinto origen de los *batch* de secuenciación, pero en todos los casos las tasas de mapeo están dentro de los valores usualmente publicados.

Las lecturas que no alinearon contra este genoma fueron alineadas contra el genoma de referencia de *L. stagnalis*², a efectos de inspeccionar si los mismos resultan de contaminación con material biológico del huésped caracol. Las tasas de mapeo de las lecturas remanentes utilizando parámetros permisivos (ver detalles en Sección 2.4) fueron bajas para los estadíos miracidio e intra-caracol 15 y 30 d.p.i. (máximo observado de 12.7% para la muestra *IS15-1*), aunque resultaron superiores al 40% (llegando a 48.2% en la muestra *IS21-3*) en el caso de las muestras provenientes de estadío intra-caracol 21 d.p.i. (ver Tabla 5.7 del *Anexo*). Esto sugiere que puede existir algo de material proveniente del huésped, pero en cualquier caso las lecturas provenientes de este no fueron incluidas en los análisis posteriores.

3.1.2 Descripción general de los transcriptos ensamblados

El análisis de los datos de mapeo y ensamblado de transcriptos generó en el orden de ~30.000 transcriptos por estadío (ver Tablas 5.5 y 5.2 del *Anexo*). Dicho valor se encuentra en el orden de magnitud de “unigenes” (transcriptos agrupados) reportado por Zhang *et al.* para estadíos de desarrollo homólogos en *F. gigantica* (58,422) [16] y es compatible con la cantidad de genes codificantes anotados en el genoma de referencia (9732).

Como se observa en la Figura 3.1, al rededor del 50% de los estadíos tienen una cobertura promedio de al menos 10x. Los resultados de clasificación de transcriptos mostrados de aquí en más refieren a estos transcriptos con un nivel de cobertura mínimo considerado aceptable.

Como se ve en la Figura 3.2B se observan tendencias homogéneas para los diferentes estadíos de estudio. La mayoría (~50%) de los transcriptos reportados por StringTie pertenecen a la categoría “*Variante de splicing*”. Las isoformas preponderantes en esta categoría son pertenecientes a la categoría “*j*” (“*multi-exón con solape en al menos un sitio de empalme*”) reportada por gffcompare (ver Tabla 5.3). Un estudio de los largos de estas isoformas respecto a las anotaciones de la referencia (ver Sección 5.0.1 del *Anexo*) sugiere que en los transcriptos codificantes las diferencias se centran principalmente en la CDS; solo una fracción marginal de los transcriptos anotados en esta categoría (*i.e.* menor al 10%) poseen diferencias marcadas solo en sus extremos no codificantes, pudiendo tratarse de errores de anotación. Los resultados obtenidos sugieren la existencia de nuevas isoformas (novedosas), las cuales podrían tener relevancia funcional en el contexto en el cual se desarrollan estos estadíos. Estudios futuros con mayor cobertura y secuenciación con lecturas largas serán necesarios para confirmar la presencia de la diversidad transcripcional sugerida por este estudio.

Las isoformas clasificadas como “novedosas” constituyen el segundo grupo mayoritario en los estadíos estudiados (~20%). Dentro de las mismas son el grupo mayoritario aquellas que se encuentran en regiones intergénicas de la anotación de referencia (categoría “*u*” en la Tabla 5.3 del *Anexo*).

Las categorías con definición de intrones idéntica a la referencia constituyen alrededor del 20% de los transcriptos reportados, y otras categorías tales como extensiones de transcriptos ya reportados en la

¹Resultados para todos los estadíos en Tabla 5.4 del *Anexo*

²Al momento de escritura de esta tesis no se encuentra disponible en las bases de datos públicas un ensamblado genómico o transcriptómico para una cepa de *L. viatrix*. El único ensamblado genómico disponible para un caracol del género *Lymnaea* es el de *L. stagnalis* aquí empleado.

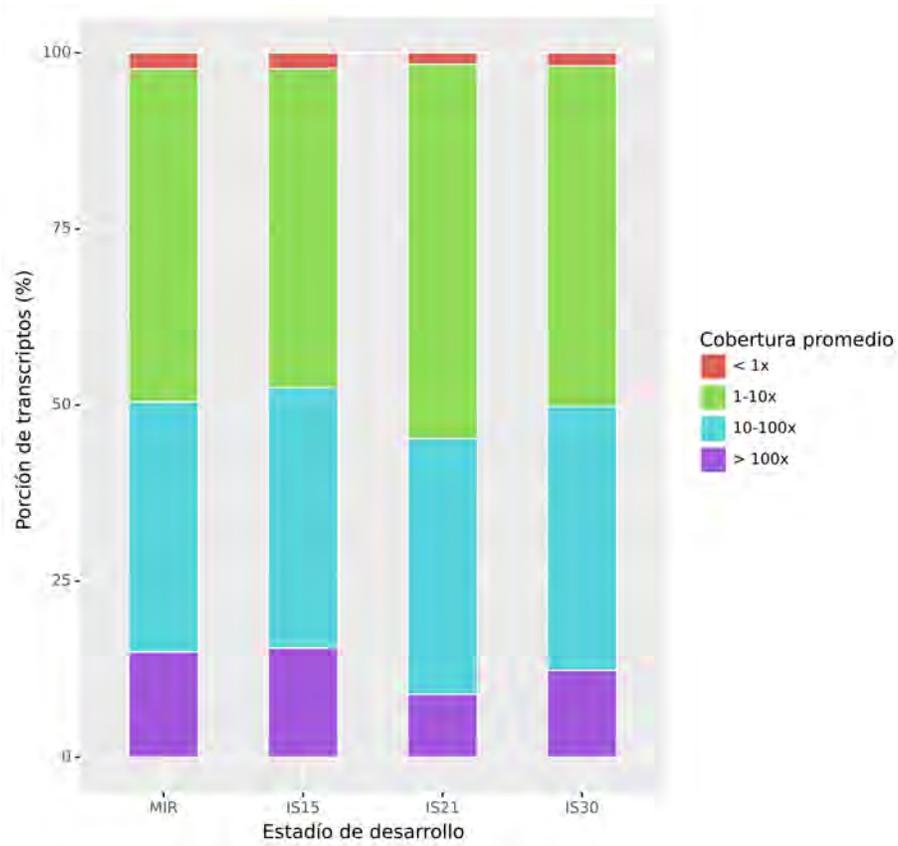


Figura 3.1: Transcritos ensamblados para muestras de miracidio y estadios intra-caracol de *F. hepatica*. Se grafica la porción de transcritos ensamblado para cada muestra a diferentes niveles de cobertura promedio reportadas por StringTie.

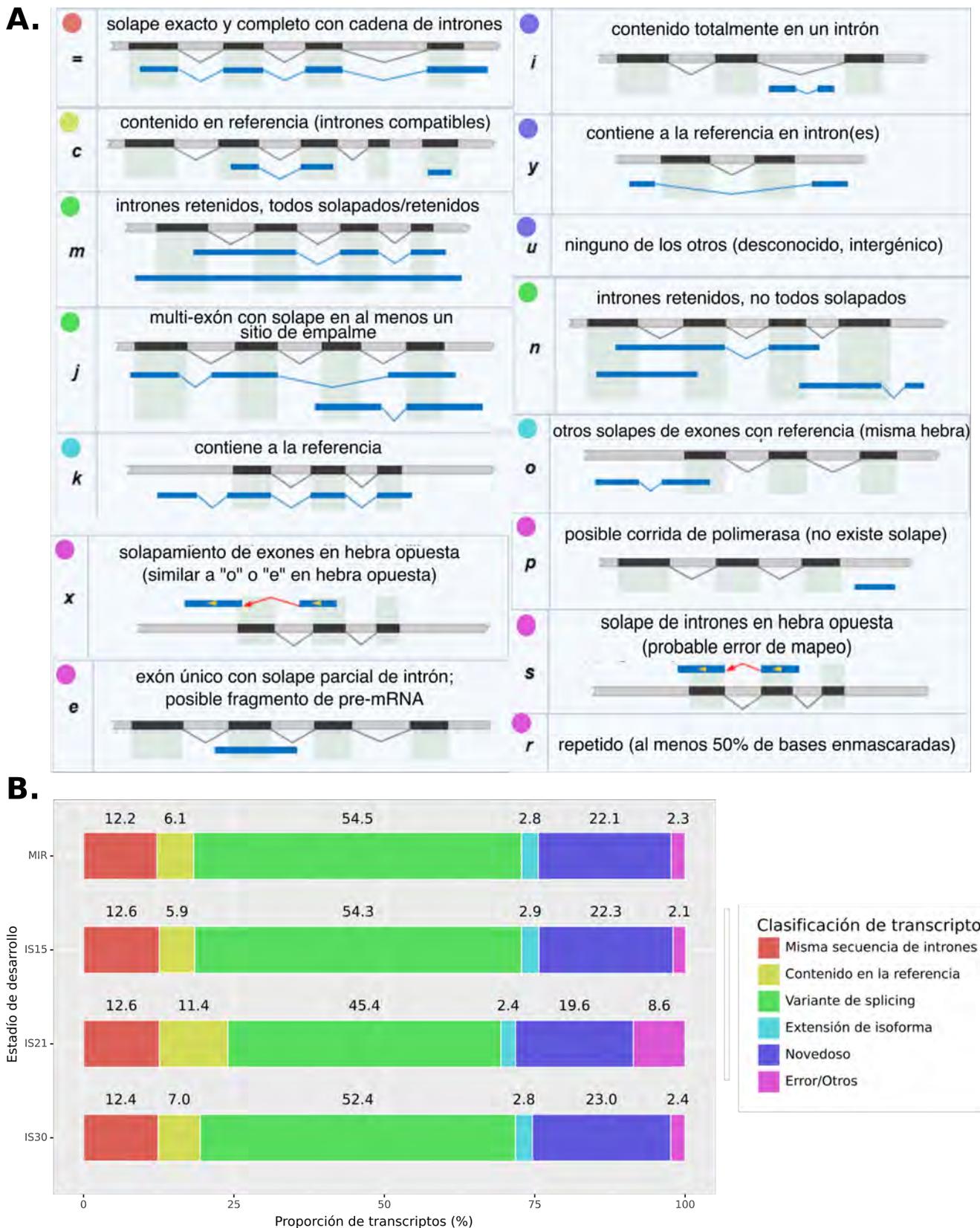


Figura 3.2: A. Sistema de clasificación de isoformas reportadas por gffcompare empleado en este trabajo. Imagen original tomada del manual de dicho programa. B. Transcritos ensamblados para muestras de miracidio y estadios intra-caracol. Se grafica la porción de transcritos ensamblado para cada muestra para aquellos transcritos con un nivel de cobertura promedio de al menos 10x reportado por StringTie.

referencia o posibles errores de mapeo y secuenciación constituyen la fracción menor en los estadios estudiados (como máximo 10% de los transcritos).

De especial interés resultan aquellos transcritos novedosos codificados en regiones intergénicas pues pueden tratarse de genes con posible expresión estadio-específica, no descritos previamente por ausencia de datos de estos estadios. Consecuentemente, seleccionamos del total reportado de transcritos de esta categoría (29.567 con cobertura promedio de al menos 10x), a los 10.947 (26.9%) que poseen una secuencia codificante para proteína (*CDS*, en inglés) detectada. Estos transcritos fueron incluidos junto al proteoma conceptual codificado en el genoma de referencia empleado de *F. hepatica* para los estudios posteriores de expresión génica. De estos modelos génicos, 1744 lograron cumplir con los criterios mínimos de expresión utilizados para definir a un gen como expresado en algún estadio estudiado, y 532 (30,5%) lograron ser anotados por el *pipeline* de anotación funcional empleado. Tal proporción sugiere que existe la posibilidad que entre aquellos *CDS* que no lograron ser anotados se encuentren secuencias pertenecientes a familias con función actualmente desconocida (con potencial a desempeñar funciones en procesos adaptativos en los estadios de estudio), o bien secuencias altamente divergentes. Estos genes fueron incorporados al conjunto de genes de la anotación canónica. Del total de 11452 genes codificantes para proteínas considerados, un 69% poseen anotación funcional de algún tipo (Tabla *fhepatica_annotacion_funcional.tsv*, *Material suplementario* a esta tesis).

3.1.3 Patrones de expresión a lo largo del ciclo de vida de *F. hepatica*

3.1.3.1 Análisis multivariado

Los estadios de desarrollo de tremátodos poseen marcadas diferencias morfológicas, fisiológicas y de estilo de vida que podrían requerir cambios en la expresión génica estadio-específicas.

Desde esta perspectiva, la identificación algorítmica de genes con perfiles de expresión estadio-específicos permite identificar actores moleculares con posibles roles funcionales de relevancia para dichos estadios. Esta aproximación posee el potencial de no estar sesgada hacia el estudio de genes frecuentemente citados en la literatura, los cuales poseen relevancia funcional/adaptativa pero eventualmente podrían constituir solo un subgrupo de los genes de interés en este aspecto. Por último, la identificación de perfiles de expresión estadio-específicos en varias especies permite eventualmente un estudio comparativo de rasgos de expresión compartidos entre estadios homólogos de distintos parásitos.

A continuación se muestran los resultados obtenidos al realizar un análisis multivariado de la expresión génica para estadios de desarrollo de *F. hepatica*, integrando a los modelos génicos pertenecientes a la anotación canónica junto a los transcritos previamente descritos.

Las primeras dos dimensiones del análisis de escalado multidimensional para las muestras de estadios de *F. hepatica* analizadas en esta tesis (ver 6.1). representan ~50% de la varianza total, (explicando ~30% y ~20% respectivamente, Figuras 3.3A y C).

El patrón de agrupamientos observado recapitula las similitudes entre los estadios de desarrollo a lo largo del ciclo de vida de *F. hepatica* (ver Sección 1.3 del *Capítulo 1*).

Se observan agrupamientos claros de muestras de: *i*) los estadios de huevo y adulto (*EGG* y *AD*), *ii*) miracidio, *iii*) los estadios intra-caracol (*IS15*, *IS21* e *IS30*), *iv*) muestras de metacercaria (*MET*) y las primeras etapas de estadio juvenil post-desenquiste (*NEJ1*, *NEJ3* y *NEJ24*), *v*) juveniles (*JUV*). Es probable que el primero de estos agrupamientos refleje la presencia de huevos en las muestras de adulto procesadas. El agregado de la tercera dimensión (Figura 3.3B, ~70% de la varianza explicada) del escalado multidimensional separa de forma clara las muestras de huevo y adulto. Las muestras de miracidio y juvenil por su parte muestran proximidad a los estadios larvarios intra-caracol en el primer caso y a las muestras de adulto y juveniles de reciente desenquiste en el último. Nuevamente, este agrupamiento refleja el ciclo

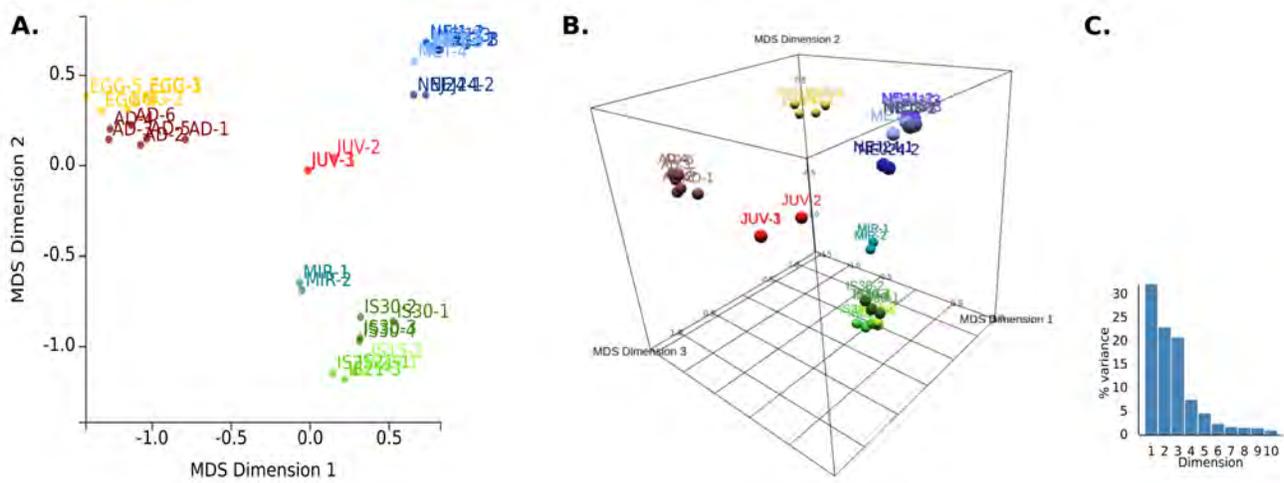


Figura 3.3: A. Análisis de escalado multidimensional a nivel de expresión génica para muestras de estadios de desarrollo de *F. hepatica* analizadas en esta tesis. Se muestran las primeras dos dimensiones principales. B. Representación visualizando los tres ejes principales C. Proporción de la varianza explicada por cada dimensión.

de vida de *F. hepatica* : los miracidios poseen en su interior masas germinales a partir de las cuales se desarrollan a los estadios larvarios intra-caracol, y los estadios juveniles de reciente desenquiste darán lugar a los juveniles que, posteriormente, desarrollarán en el adulto.

En conclusión, el agrupamiento de las muestras analizadas según su perfil de expresión génica se da acorde a lo esperado teniendo en cuenta las similitudes en el ciclo de vida del parásito. Tal agrupamiento se da aun cuando las muestras analizadas provienen de diferentes estudios, habilitando un análisis global de los patrones de expresión en los estadios analizados.

3.1.3.2 Solapamiento de patrones de expresión entre estadios de desarrollo

Los patrones observados en el análisis multivariado (en coherencia con las marcadas diferencias biológicas entre los distintos estadios) podrían explicarse a nivel transcripcional por distintos escenarios. La expresión de conjuntos de genes marcadamente diferentes podrían ser la base de estos agrupamientos, así como diferencias más o menos sutiles entre un mismo conjunto de genes expresados en los estadios estudiados. Naturalmente, entre estos dos extremos existen escenarios intermedios. Como una primera aproximación al análisis de la expresión génica durante el ciclo de vida de *F. hepatica* se decidió estudiar el grado de solapamiento en expresión génica entre los estadios de desarrollo estudiados, a efectos de cuantificar estas diferencias en la expresión génica. Además de realizar este análisis para el conjunto de genes para los que se detectó transcripción, se realizó el mismo estudio sobre el conjunto de genes de mayor expresión, en tanto algunos de estos genes pueden desempeñar funciones de relevancia específica para cada estadio (de allí su alto grado de expresión). El análisis formal de expresión diferencial, el cual permite el estudio de diferencias más sutiles entre estadios, se deja para una sección aparte, en la cual se hará foco en el estudio de las diferencias entre estadios larvarios asociados al huésped caracol (ver Sección 3.3).

En la Figura 3.4A se observa al conjunto de genes expresados a lo largo del ciclo de vida de *F. hepatica*, clasificándose a los mismos según el conjunto de estadios en los cuales se los detecta.

Más del 50% de los genes son detectados en todos los estadios analizados (categoría “*Expresión en todos los estadios*” en la figura). Interesantemente, se observa un aumento marcado para esta categoría en los estadios de huevo y adulto (*EGG* y *AD* en la Figura). Estos son los dos estadios para los cuales se observa

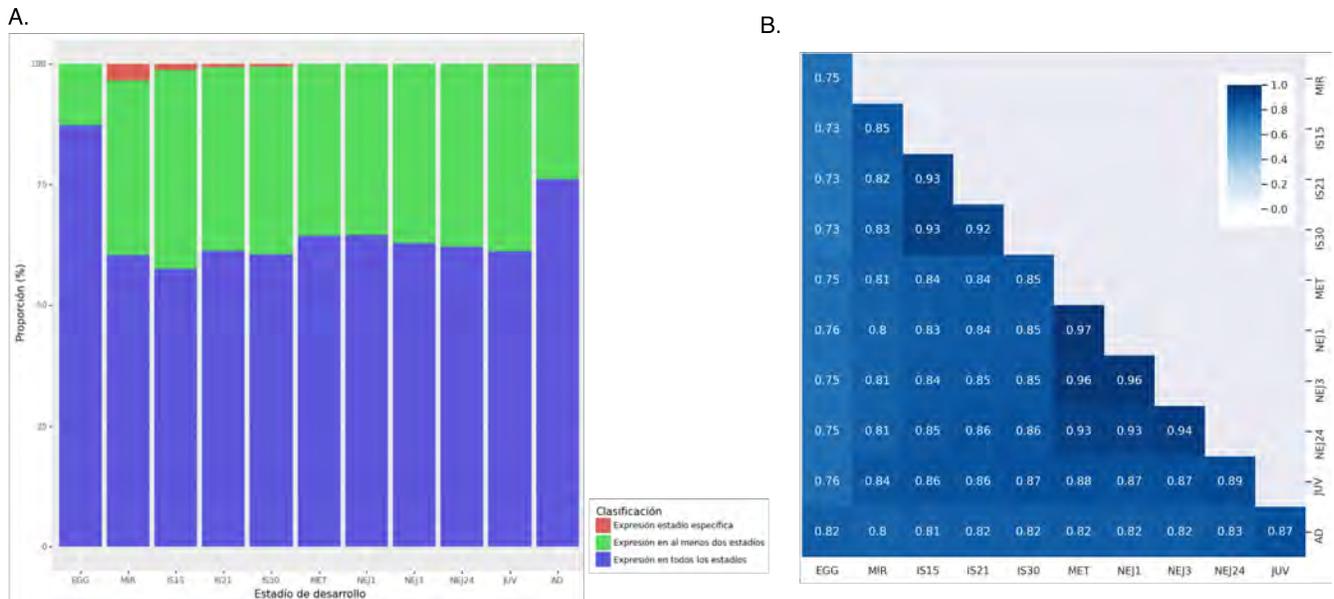


Figura 3.4: Análisis de genes detectados a lo largo de los estadios de desarrollo de *F. hepatica* con datos transcriptómicos disponibles a la fecha. A. Se indica la proporción de genes expresados en cada estadio de *F. hepatica*, clasificados según si se expresan en todos los estadios (violeta), en múltiples estadios (expresión en al menos dos estadios, verde) o poseen estadio específica (rojo). B. Análisis de proporción de genes expresados simultáneamente en pares de estadios de desarrollo de *F. hepatica*.

un menor número de genes con expresión detectada, por lo cual el fenómeno observado podría deberse a una subdetección de genes expresados pertenecientes a otras categorías (*e.g.* por falta de profundidad de secuenciación). No obstante, no es posible establecer que esta sea la causa del patrón observado con los datos disponibles.

En segundo lugar de representación se encuentra en todos los casos la categoría “*Expresión en al menos dos estadios*”. Los genes de la misma componen desde el ~10% (*EGG*) hasta el ~45% (*IS15*). No se observó una tendencia marcada en los genes de esta categoría a encontrarse en estadios de desarrollo similares, con agrupamientos similares a los observados en el análisis multivariado (datos no mostrados).

Por último, se observa un porcentaje marginal de genes con expresión estadio específica, particularmente en las muestras de miracidio y estadios intra-caracol, en los que se centra este estudio. La mayoría de estos corresponden a transcritos codificantes para proteínas ensamblados con StringTie en regiones intergénicas de la anotación de referencia, lo cual sugiere que estos podrían corresponder con genes de expresión restringida que no habían sido detectados o confirmados previamente por la ausencia de datos transcriptómicos que los sostengan.

En la Figura 3.4B se observan las comparaciones pareadas de grupos de genes expresados simultáneamente en estadios de desarrollo de *F. hepatica*. Coherente con los resultados mostrados en la Figura 3.4 A, se observa un alto solapamiento de los conjuntos de genes expresados entre estadios: como mínimo se observa un solapamiento del 73% de los conjuntos de genes expresados entre estadios, correspondiendo este valor a la comparación entre huevo y los estadios intra-caracol. A su vez, los resultados son coherentes en términos generales con los presentes en los análisis multivariados: valores más altos de solapamiento se observan entre los estadios intra-caracol (llegando a 93% de solapamiento) y entre las muestras de metacercaria y juveniles de reciente desenquiste (con máximo de 97% de solapamiento). Las muestras de miracidio poseen una mayor similitud con las muestras de estadio intaracaracol y juveniles, mientras que la expresión en huevo posee mayor similitud con la de adulto (el cual posee a su vez una alta similitud con juvenil).

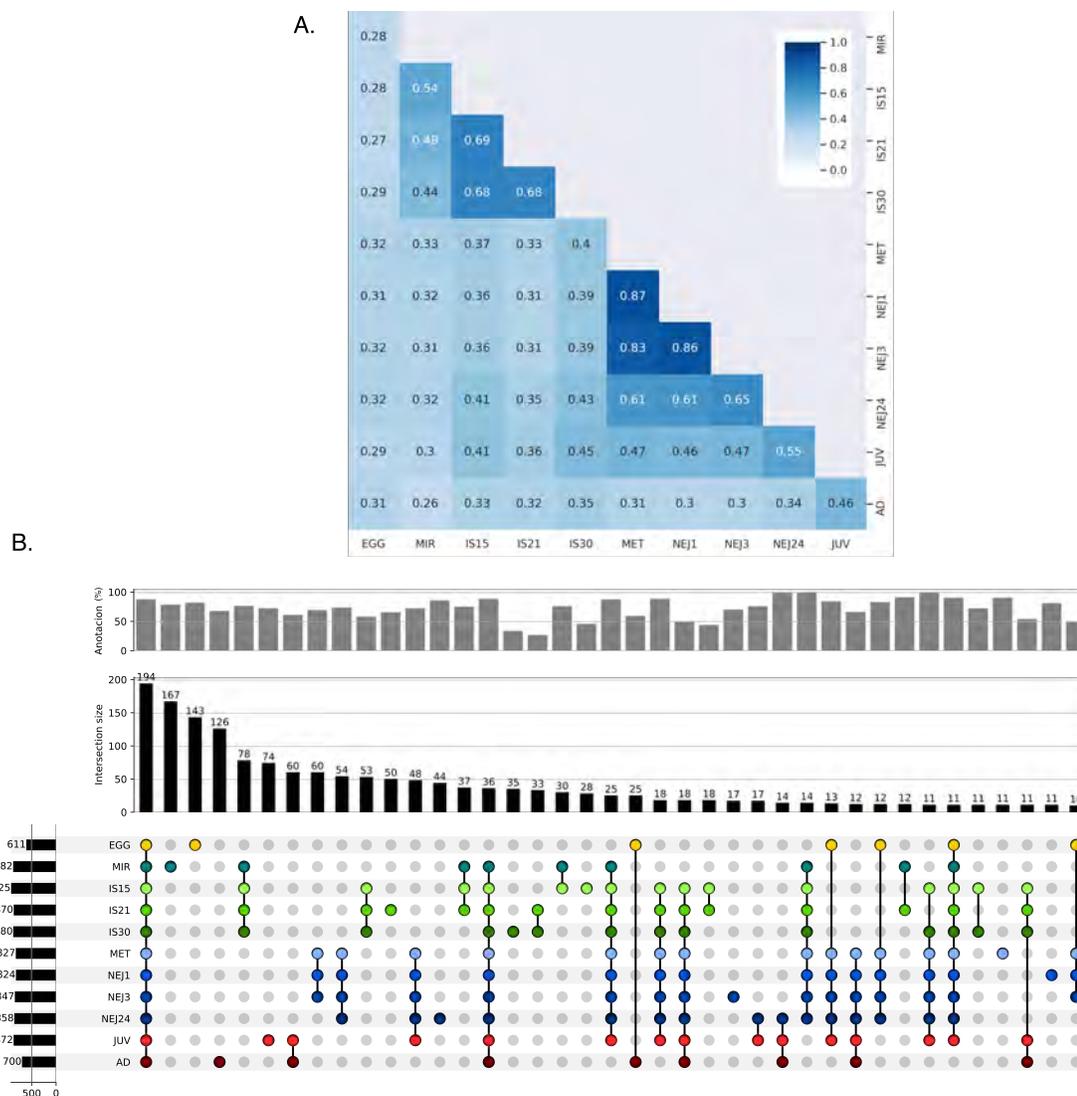


Figura 3.5: Análisis de patrones de expresión para el conjunto de genes de mayor expresión en cada estadio de desarrollo de *F. hepatica*. A. Análisis de proporción de genes en esta categoría simultáneamente en pares de estadios. B. Visualización de los grupos de intersección para estos genes para combinaciones de estadios de desarrollo. Se observa al distribución para conjuntos compuestos por al menos diez genes. Se describe a su vez el porcentaje de genes con algún tipo de anotación funcional (GO, KEGG o PFAM) en estos conjuntos.

En la Figura 3.5A se observa el resultado de restringir este análisis de expresión conjunta al 10% de genes de mayor expresión en cada estadio. Como es de esperar, se observa la misma tendencia descrita en la Figura 3.4B. A pesar de ello, los valores observados abarcan un rango sensiblemente mayor, yendo desde el 28% de solapamiento (comparación entre huevo y estadio intra-caracol 21 d.p.i.) hasta 87% como máximo (comparación entre metacercaria y juvenil 1 hora post desenquiste). En ese sentido resalta un grupo de genes con expresión compartida entre los estadios intra-caracol y otros vinculados a los primeros estadios de infección del huésped mamífero (agrupando metacercaria y estadios juveniles de reciente desenquiste). En la Figura 3.5B se observa en detalle la distribución de estos genes de alta expresión según estadios de desarrollo. Este enfoque resalta por un lado un núcleo central de genes expresados a lo largo de todo el ciclo, genes específicos de estadio y genes expresados en estadios similares relevantes del ciclo biológico, particularmente los vinculados al huésped intermediario y al definitivo. Si bien el grupo mayoritario se encuentra expresado en todos los estadios, en el resto de los grupos se observa o bien una tendencia de agrupamiento entre estadios de desarrollo similares o bien la detección de alta expresión estadio específica.

En resumen, como era esperado buena parte de los genes se expresan a lo largo de todo el ciclo, y existen distintos agrupamientos de genes con expresión restringida, y un grupo más reducido de genes de expresión estadio específica. A pesar de que en un primer de análisis no se observaron agrupamientos claros de genes expresados en simultáneo en estadios de desarrollo similares, el análisis pareado de conjuntos expresados sí refleja dicha tendencia, coherente con el análisis multivariado de la expresión. La restricción del análisis a los genes de mayor expresión en cada estadio revela más claramente que una fracción de los mismos (~30 %) se encuentra expresada en todos los estadios, mientras que aquí aparecen resaltados genes de expresión específica, los cuales podrían estar vinculados a eventos biológicos clave como lo son la interacción con el huésped intermediario o la invasión al huésped definitivo.

3.1.3.3 Clusterización de genes por su patrón de expresión

Desde una perspectiva biológica resulta claro que las tendencias de expresión asociadas a estadios no tienen por qué restringirse a genes de alta expresión. Cambios regulatorios sutiles en componentes clave de redes de interacción (*e.g.* en vías de desarrollo o metabólicas) pueden ser suficientes para establecer diferencias significativas en la expresión génica entre estadios.

Como una primera aproximación a la búsqueda de perfiles transcripcionales a lo largo del ciclo de vida de *F. hepatica* se empleó el programa *clust*, utilizado en estudios de expresión de *F. gigantica* y en el estudio de redias, cercarias y adultos de especies de la familia Psilostomatidae (orden Plagiorchiida, al igual que Fasciolidae) [16,17].

En la Figura 3.6 se observan los *clusters* de co-expresión de genes inferidos con este método.

En consonancia con los agrupamientos observados en los análisis de expresión mostrados en las Figuras 3.4 B y 3.5 se infirieron *clusters* de co-expresión con expresiones sesgadas hacia estadios intra-caracol (*clusters* A y B) o intra-mamífero (*clusters* F y G). Este resultado apoya la noción de existencia de diferencias a nivel de los programas regulatorios expresados en uno u otra huésped, los cuales podrían permitir la adaptación de los estadios de desarrollo involucrados a sus respectivos entornos. Estos perfiles pueden o no estar asociados a genes de alta expresión como los que se observaban en la Figura 3.5.

Otros perfiles de expresión son más difíciles de asociar a una condición biológica en particular (*e.g.* *clusters* C, D y E), pareciendo consistir en conjuntos de genes reprimidos en algunos estadios de desarrollo como el huevo o miracidio.

En todos los casos los *clusters* generados poseen perfiles de expresión asociados a más de un estadio de desarrollo, no pudiéndose detectar con esta metodología *clusters* de co-expresión estadio-específica. Los resultados obtenidos son similares en este aspecto a lo reportado por Zhang *et al.* en su estudio de *F. gigantica*, donde en general no se observan *clusters* de co-expresión estadio-específicos en un abordaje

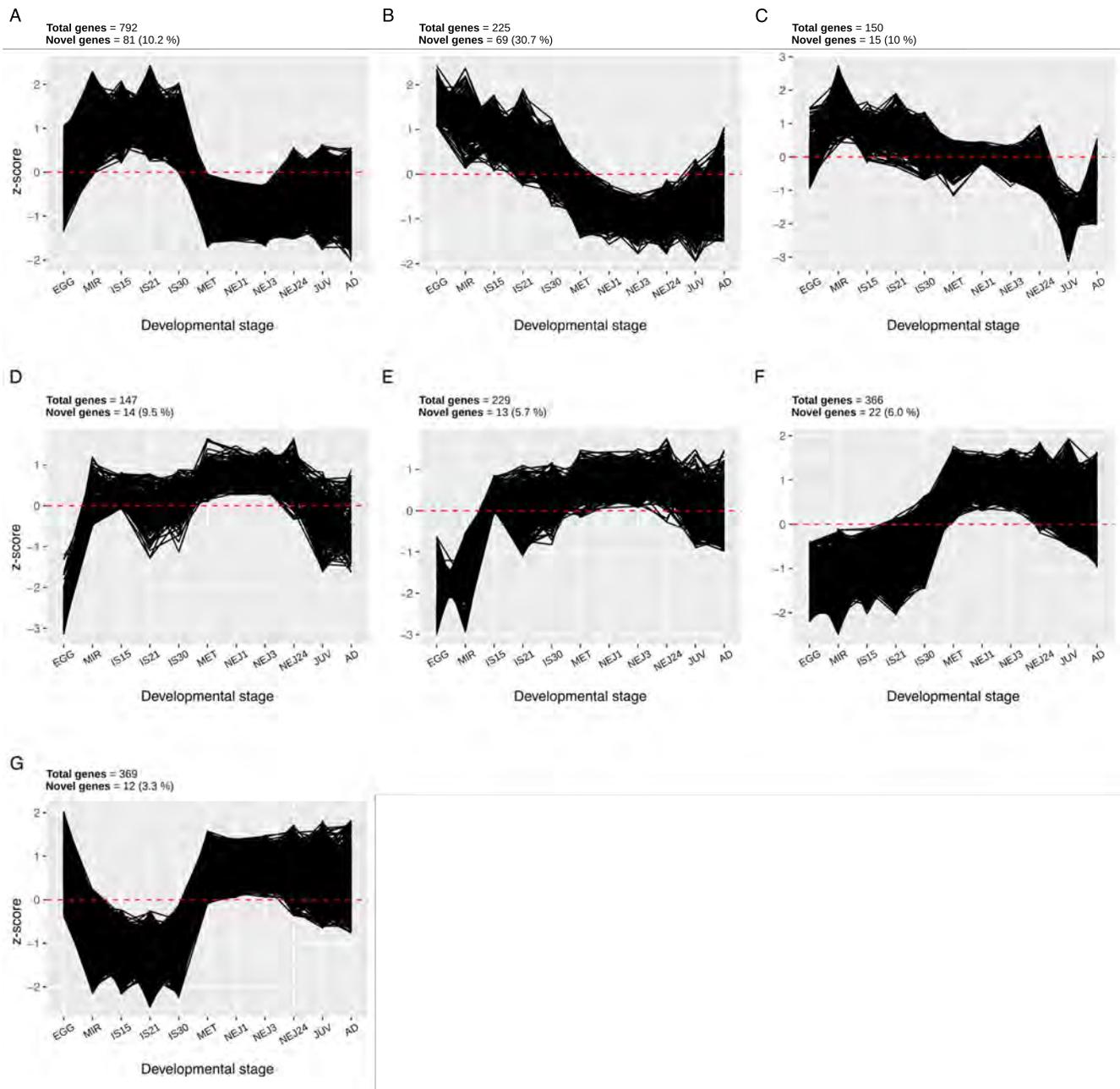


Figura 3.6: Clusters de co-expresión de genes a lo largo del ciclo de vida de *F. hepatica* inferidos con la herramienta clust. Se observan las tendencias de expresión normalizadas (z-score) para los genes pertenecientes a cada cluster. La inferencia se realizó corriendo clust. Genes con una tendencia de expresión estable en el ciclo de vida son filtrados.

directo de agrupamiento con *clust*. En otros casos, como el reportado con Nesterenko et al para especies de Psilostomatidae es posible que se haya logrado detectar *clusters* de co-expresión con mayor especificidad debido al estudio de un menor número de estadíos de desarrollo, factor que disminuye la variabilidad presente en el estudio.

3.1.3.4 Diferenciación de genes de expresión estable y de expresión altamente variable

En búsqueda de perfiles de expresión estadio-específicos que pudieran brindar información acerca de genes con posible relevancia en cada estadio de desarrollo, Zhang et al optaron por el agrupamiento del subconjunto de genes diferencialmente expresados que presentara mayores diferencias entre estadíos [16]. Se optó por una estrategia similar para la detección de dichos perfiles. En su estudio, Zhang et al. realizaron el agrupamiento sobre genes diferencialmente expresados que mostraran una diferencia de al menos 16 veces a nivel de expresión entre alguno de los pares de estadíos analizados [16]. Teniendo en cuenta la heterogeneidad del set de datos analizado de *F. hepatica* (*i. e.* muestras provenientes de diferentes *batches* de secuenciación), se decidió aplicar un criterio similar, pero restringir el análisis a un subconjunto de genes que presenten alta variabilidad a fin de intentar disminuir la aparición de falsos positivos. Se decidió restringir el análisis al conjunto de genes diferencialmente expresados (*i.e.* $|\log_2 FC| \geq 4$) y $FDR < 0,01$) que aumentan su expresión al menos 16 veces respecto al valor promedio en algún estadio. Un análisis de estos genes muestra que este es un subgrupo de aquellos genes que muestran una expresión diferencial con $|FC| \geq 16$ entre algún par de estadíos (criterio que como se mencionó antes emplearon Zhang et al. en su estudio), restringiéndose a aquellos genes que muestran cambios marcados en al menos un estadio de desarrollo. Otros criterios menos estrictos han sido empleados en la literatura. A modo de ejemplo, recientemente Cheng et al. definen un gen como de expresión estadio-específica en un estadio larvario de *S. japonicum* si muestra $FC > 5$ y $FDR < 0,05$ respecto al resto de los estadíos [15]. No obstante, los autores realizan un análisis de muestras pertenecientes a un mismo análisis, por lo que es razonable suponer que se enfrentan a un set de datos más homogéneo.

En la Figura 3.7 se observa la relación existente entre la expresión promedio de un gen y la varianza de esta magnitud (σ^2) a lo largo de los estadíos de desarrollo de *F. hepatica*.

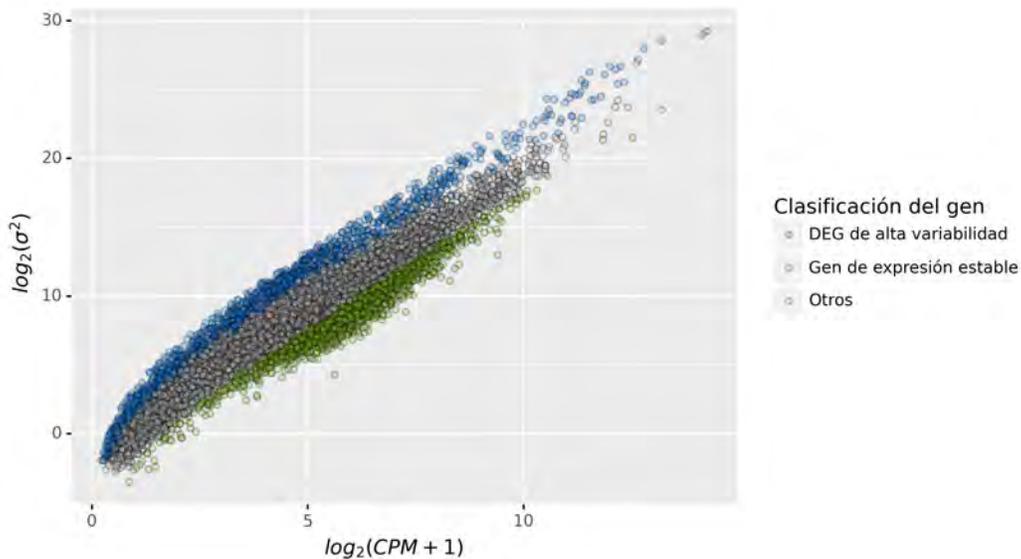


Figura 3.7: Relación entre la media de expresión de un gen y su varianza a lo largo del ciclo de vida de *F. hepatica*. Se clasifica a los genes según su patrón de expresión en genes diferencialmente expresados de alta variabilidad (GDAVs, verde), genes de expresión estable (GEEs, azul) y otros (gris).

Se observa una relación de proporcionalidad entre ambas variables, acorde a lo predicho por los modelos de distribución binomial negativa con los que modelan los valores esperados para conteos de genes herramientas como edgeR o DESeq2 (ver Sección 5.0.2 del *Anexo*).

En azul se observa el conjunto de 1784 genes muy diferencialmente expresados entre pares de condiciones que fue seleccionado para análisis posteriores, con los criterios detallados en los párrafos anteriores.

Como se observa en la Figura, los mismos se encuentran entre los genes con mayor varianza en su expresión respecto a su expresión promedio. En virtud de esto se los decidió denominar como “*genes diferencialmente expresados de alta variabilidad*” (*GDAV*). En la misma Figura se muestra a su vez al conjunto de 2431 genes con menor coeficiente de variabilidad ($CV = \frac{\text{desvío estándar}}{\text{promedio}}$; en este caso $CV < 0,5$), el cual fue utilizado como conjunto de referencia para el estudio de los *GDAV*. El mismo pasará ser denominado “*genes de expresión estable*” (*GEE*) en este contexto.

Este conjunto de genes con expresión más estable a lo largo del ciclo de vida será comparado como punto de referencia más adelante, al momento de estudiar aspectos como la conservación a nivel de secuencia o distribución filética de los *GDAV*.

Por último, resulta pertinente remarcar que la definición de estos grupos de genes resulta coherente con los análisis de expresión mostrados previamente. Esta diferenciación en la dinámica transcripcional parece consistente con la existencia de procesos biológicos necesarios para el desarrollo de todos los estadios (*e.g.* vías metabólicas esenciales), así como la existencia de genes con valor adaptativo en estadios particulares.

3.1.3.5 Patrones de expresión de los *GDAV*

Por su definición, los *GDAV* poseen tendencia a la expresión para algún conjunto de estadios. Por tanto, empleando estrategias algorítmicas simples de clasificación se los puede organizar en subconjuntos según sus tendencias de expresión.

En la Figura 3.8 se observa el resultado de agrupar los *GDAV* mediante un algoritmo de *k*-medias en seis grupos. Dicho número fue seleccionado teniendo en cuenta los patrones de agrupamiento observados en el análisis de escalado multidimensional previamente presentado (Figura 3.3).

Como se observa en la Figura 3.8, se conforman seis agrupamientos con tendencia a la expresión específica en: *i*) huevos (“*EGG*” en la Figura); *ii*) miracidio (“*MIR*”); *iii*) estadios tempranos de la infección a huésped caracol (“*IS15/21*”); *iv*) estadios tardíos de la infección a huésped caracol (“*IS30*”); *v*) metacercarias y juveniles de desenquiste temprano (“*MET/NEJ*”); y *vi*) estadio juvenil y adulto (“*JUV/AD*”).

Este patrón es consistente en líneas generales con los agrupamientos observados previamente para las muestras analizadas en el por escalado multidimensional. No obstante, el agrupamiento por algoritmo de *k*-medias no identifica un grupo de *GDAV* con tendencia a la expresión específica en el estadio juvenil; a pesar de la distinción de las muestras de este estadio y las de adulto observadas mediante escalado multidimensional, se observa la conformación de un grupo con *GDAV* con tendencia a la expresión específica en ambos estadios (con preponderancia a la mayor expresión en estadio adulto).

En contraposición, sí se conforman dos grupos con tendencia a la expresión diferencial en estadios intracaracol: uno de ellos contempla a *GDAV* con mayores niveles de expresión en las primeras semanas de la infección al huésped caracol (grupo “*IS15/21*” en la Figura 3.8, mientras que el otro agrupa a *GDAV* con mayor expresión en la etapa tardía de la infección (grupo “*IS30*”). Tal distinción entre los patrones de expresión de estos estadios es apreciable en el análisis de escalado multidimensional, aunque no en los análisis de conjuntos de genes co-expresados entre estadios (Figuras 3.4B, 3.5 y 3.6).

Los patrones de expresión observados para los *GDAV* de los agrupamientos formados muestran una especificidad parcial para los estadios, siendo coherentes también en este aspecto con los análisis de escalado multidimensional. En particular: *i*) los genes del agrupamiento “*EGG*” muestran también aumentos de

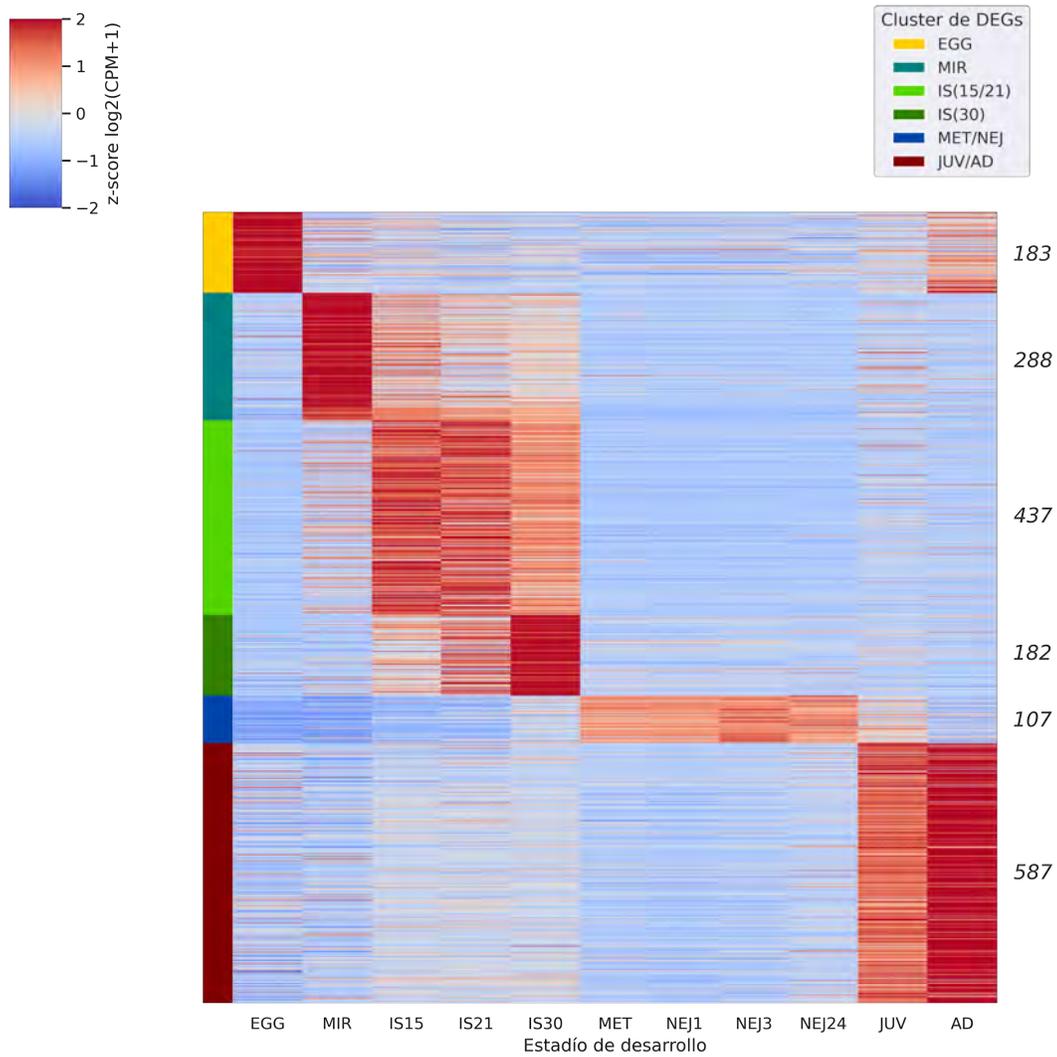


Figura 3.8: Patrones de expresión para genes pertenecientes a clusters de GDAVs de *F. hepatica*. Clusters generados por algoritmo de *k-means* ($k = 6$) para agrupar DEGs ($\log_2\text{FC} \geq 4$ y $\text{FDR} < 0.01$ en alguna comparación pareada) que muestran $\log_2\text{FC} \geq 4$ respecto al promedio de expresión en algún estadio de desarrollo.

expresión en el estadio adulto; *ii*) los genes del agrupamiento “*MIR*” muestran también aumentos de expresión en estadios intra-caracol; *iii*) los genes del agrupamiento “*IS15/21*” muestran también aumentos en la expresión en el estadio intra-caracol 30 d.p.i.; *iv*) los genes del agrupamiento “*IS30*” muestran también aumentos en la expresión en los estadios intra-caracol 15 y 21 d.p.i.

En resumen, se logró identificar agrupamientos de *GDAV* por algoritmo de *k*-medias con tendencia a la expresión diferencial en conjuntos de estadios de desarrollo de *F. hepatica*.

En la Figura 3.9C se observa la proporción de genes de *clusters* de *GDAVs* que poseen algún tipo de anotación funcional. Los porcentajes más bajos de anotación funcional se observan en los *clusters* de genes con tendencia a la expresión diferencial específica en estadios larvarios asociados a huésped caracol, los cuales se encuentran por debajo de los porcentajes de anotación obtenidos para el conjunto general de genes. En contraste, genes clasificados como *GEE* son en su mayoría genes con anotación conocida. Estos genes *GDAV* son potenciales candidatos a desempeñar funciones de relevancia específicas en los estadios intra-caracol. El hecho de que no presenten una señal clara de homología a nivel de secuencia con genes anotados funcionalmente puede o bien deberse a una especificidad filética (*i.e.* encontrarse conservados a nivel de platemintos, tremátodos o fasciolidos) que haya impedido el estudio de genes homólogos en grupos taxonómicos de mayor estudio (*e.g.* organismos modelo) o directamente codifican proteínas que participan en vías metabólicas o procesos biológicos que no se han descrito a la fecha. En cualquier caso, estudios *in silico* más pormenorizados (ver *Discusión general, conclusiones y perspectivas*) y aproximaciones experimentales serán necesarias para identificar su identidad funcional y su rol específico en el contexto de infección a caracol por parte de *F. hepatica*.

En cuanto a aquellos con algún tipo de anotación funcional, no se observa una tendencia funcional única. Un abordaje global a las categorías presentes en los *GDAV* que si lograron ser anotados funcionalmente muestra que se trata de genes vinculados a categorías variadas, incluyendo genes codificantes para proteínas estructurales, procesos metabólicos o proteasas (Tabla 3.2). No se presenta una asociación evidente asociada a los estadios en los cuales los *GDAV* se sobreexpresan y su función. Algunos de estos genes presentan expresión diferencial entre estadios intra-caracol, por lo que serán abordados más adelante. En otros casos como el de *hpvt* se observan genes de alta expresión constituyentes de vías regulatorias. Este caso puntual parece ejemplificar que algunos de los *GDAV* podrían ser genes participantes de una vía metabólica o regulatoria relevante, cuya posición en la red de interacciones pueda verse regulada específicamente en estadios intra-caracol.

Tabla 3.2: Conteo de genes *GDAVs* clasificados funcionalmente por inspección visual. Se sigue aproximadamente el esquema de clasificación KEGG BRITE.

Cluster	Clasificación	Etiqueta	Conteo
MIR	Metabolismo	Biosíntesis de lípidos	1
MIR	Metabolismo	Metabolismo de aminoácidos	1
MIR	Metabolismo	Metabolismo de carbohidratos	1
MIR	Metabolismo	Metabolismo de cofactores y vitaminas	1
MIR	Metabolismo	Metabolismo de nucleótidos	2
MIR	Metabolismo	Peptidasas e inhibidores	2
MIR	Otros	Cáncer	1
MIR	Otros	Tráfico de membranas	1
MIR	Proc. Info. Genética	Factores de transcripción	2
MIR	Proc. Info. Genética	Proteín quinasas	3
MIR	Proc. Info. Genética	Tráfico de membranas	1

MIR	Señalización y procesos celulares	Citoesqueleto	3
MIR	Señalización y procesos celulares	Cromosoma y proteínas asociadas	1
MIR	Señalización y procesos celulares	Exosoma	1
MIR	Señalización y procesos celulares	Factores de transcripción	1
MIR	Señalización y procesos celulares	Lisosoma	1
MIR	Señalización y procesos celulares	Moléculas CD	2
MIR	Señalización y procesos celulares	Moléculas de adhesión celular	2
MIR	Señalización y procesos celulares	No clasificada	3
MIR	Señalización y procesos celulares	Otros	1
MIR	Señalización y procesos celulares	Transducción de señales	1
MIR	Señalización y procesos celulares	Transportadores	1
IS15/21	Metabolismo	Metabolismo de carbohidratos	1
IS15/21	Metabolismo	Metabolismo de lípidos	3
IS15/21	Metabolismo	Metabolismo de nucleótidos	3
IS15/21	Metabolismo	No clasificada	2
IS15/21	Metabolismo	Oxidoreductasas	1
IS15/21	Metabolismo	Peptidasas e inhibidores	3
IS15/21	Metabolismo	Sistema de ubiquitina	1
IS15/21	Metabolismo	Transferasas	1
IS15/21	Proc. Info. Genética	Factores de transcripción	6
IS15/21	Proc. Info. Genética	Liasas	1
IS15/21	Proc. Info. Genética	Maquinaria de transcripción	1
IS15/21	Proc. Info. Genética	Proteín quinasas	1
IS15/21	Proc. Info. Genética	Sistema de ubiquitina	1
IS15/21	Proc. Info. Genética	Tráfico de membranas	1
IS15/21	Señalización y procesos celulares	Citoesqueleto	3
IS15/21	Señalización y procesos celulares	Cromosoma y proteínas asociadas	1
IS15/21	Señalización y procesos celulares	Moléculas CD	1
IS15/21	Señalización y procesos celulares	No clasificada	2
IS15/21	Señalización y procesos celulares	Transferasas	1
IS15/21	Señalización y procesos celulares	Transportadores	10
IS30	Metabolismo	Biosíntesis de lípidos	1
IS30	Metabolismo	Biosíntesis y metabolismo de glicanos	4
IS30	Metabolismo	Metabolismo de aminoácidos	2
IS30	Metabolismo	Metabolismo de cofactores y vitaminas	1
IS30	Metabolismo	Peptidasa e inhibidores	6
IS30	Proc. Info. Genética	Cromosomas y proteínas asociadas	1
IS30	Señalización y procesos celulares	No clasificada	2
IS30	Señalización y procesos celulares	Transportadores	5

3.1.3.6 Comparación con genes de expresión altamente variable de *F.gigantica*

Si bien como se describió el análisis de la anotación resulta poco informativo para los *clusters* de *GDAV*, es posible caracterizar algunos aspectos de los mismos a partir de la comparación de sus secuencias. En particular, se decidió estudiar tres aspectos a fin de obtener una aproximación al perfil de estos genes: *i*) divergencia a nivel de secuencia respecto a genes homólogos en *F. gigantea*, *ii*) conservación filética

de los genes en metazoarios y *iii*) existencia de *GDAV* homólogos en *F. gigantea*. A fin de comparar los resultados obtenidos para los *GDAV* con alguna referencia, se incorporó también a los *GEE* a los estudios de distancia y conservación filética. La aproximación permite tener una noción de los cambios acumulados a nivel de secuencia para este grupo de genes y su especificidad a nivel taxonómico, analizándose estos aspectos por comparativa con la contraparte de genes de expresión más estable. La existencia de tendencias de expresión similares en homólogos de *F. gigantea* permite analizar la conservación de las mismas en otro fasciólido, a la vez que provee información de expresión proveniente de un estudio independiente, en donde se esperaría observar tendencias muy similares debido a la proximidad filogenética de las especies.

Se estimó la distancia cofenética para genes homólogos de *F. hepatica* y *F. gigantea*, a fin de comparar la divergencia a nivel de secuencia para grupos de homólogos con distintas dinámicas de expresión (ver *Materiales y Métodos*, Sección 2.8.2). Es de esperar que genes evolucionando bajo un régimen de selección negativa más estricto acumulen cambios aminoacídicos a una menor tasa que genes evolucionando bajo un régimen selectivo más relajado, neutro o bajo selección positiva.

En efecto pares de genes *GEE* homólogos tienden a mostrar las distancias cofenéticas más pequeñas observadas, mientras que pares de *GDAV* en ambas especies presentan una distribución con tendencia a valores mayores para dicha medida; el resto de las comparaciones realizadas presenta una distribución de valores de tendencia intermedia (3.9A). Este resultado muestra que los genes *GDAV* tienden a poseer una mayor tasa de cambios aminoacídicos que genes *GEE*.

Si los genes *GEE* deben su mayor estabilidad en la expresión al hecho de codificar productos con funciones esenciales a lo largo del ciclo de vida del parásito, es de esperar que la evolución de los mismos suceda bajo un régimen de selección negativa, acumulando una menor tasa de cambios que otros genes. Esta hipótesis podría explicar el hecho de que se observe que pares de estos genes acumulen cambios aminoacídicos a una tasa menor que otros pares de genes. De forma similar, el hecho de que pares de *GDAV* acumulen cambios a una mayor tasa de sustitución aminoacídica que los otros pares de genes podría reflejar la evolución bajo un régimen selectivo más relajado. Esto podría incluir en algunos casos evolución bajo régimen de selección positiva. Alternativamente, algunos de estos genes podrían estar en proceso de pseudogenización, aunque el hecho de que sean codificantes para proteínas y se vean transcriptos es evidencia parcial en contra de este escenario. En cualquier escenario, el hecho de que conjuntos de genes codificantes para proteína posean tendencias distintas en sus tasas de acumulación de cambios aminoacídicos según su dinámica transcripcional constituye un dato interesante que podría verse vinculado a la función de los productos codificados para estos genes. Resulta importante considerar que las distancias cofenéticas calculadas combinan el efecto de todos los cambios acumulados, llegando incluso a incorporar distancias dadas por duplicaciones ancestrales al evento de especiación entre ambas especies (si es que los parálogos ancestrales fueron combinados en un mismo *OrthoGroup*). Lo que es más, distintos grupos de homólogos variarán en este aspecto, en tanto no es necesario que concuerden. A modo de ejemplo, familias con una divergencia marcada a nivel de secuencia pueden presentar erosión de la señal de secuencia, siendo fraccionadas en *OrthoGroups* pequeños; las comparaciones hechas entre homólogos de estos grupos presentarían distancias cofenéticas artefactualmente pequeñas. El resultado obtenido puede ser tomado como una primera aproximación al problema, donde se observan tendencias generales. Incluso si los homólogos *GDAV* de *F. hepatica* y *F. gigantea* presentaran distancias cofenéticas pareadas más largas por una tendencia a una mayor tasa de duplicación (ya sea ancestral o linaje-específica), esta dinámica evolutiva sería interesante, en tanto podría eventualmente operar allí un régimen de selección positiva, ya sea por *i*) selección de aumento de dosaje del producto codificado en el gen, o *ii*) selección divergente operando en algún linaje particular. A su vez, la existencia de un mayor número de copias génicas podría permitir la acumulación de cambios sin pérdida funcional de la función del gen ancestral, habilitando por tanto la evolución bajo un régimen selectivo neutro o de selección negativa relajada.

La conservación taxonómica de los genes *GDAV* y *GEE* resulta un dato de relevancia para evaluar algunas de las hipótesis mencionadas hasta el momento. Una hipótesis naïve respecto a las dinámicas de

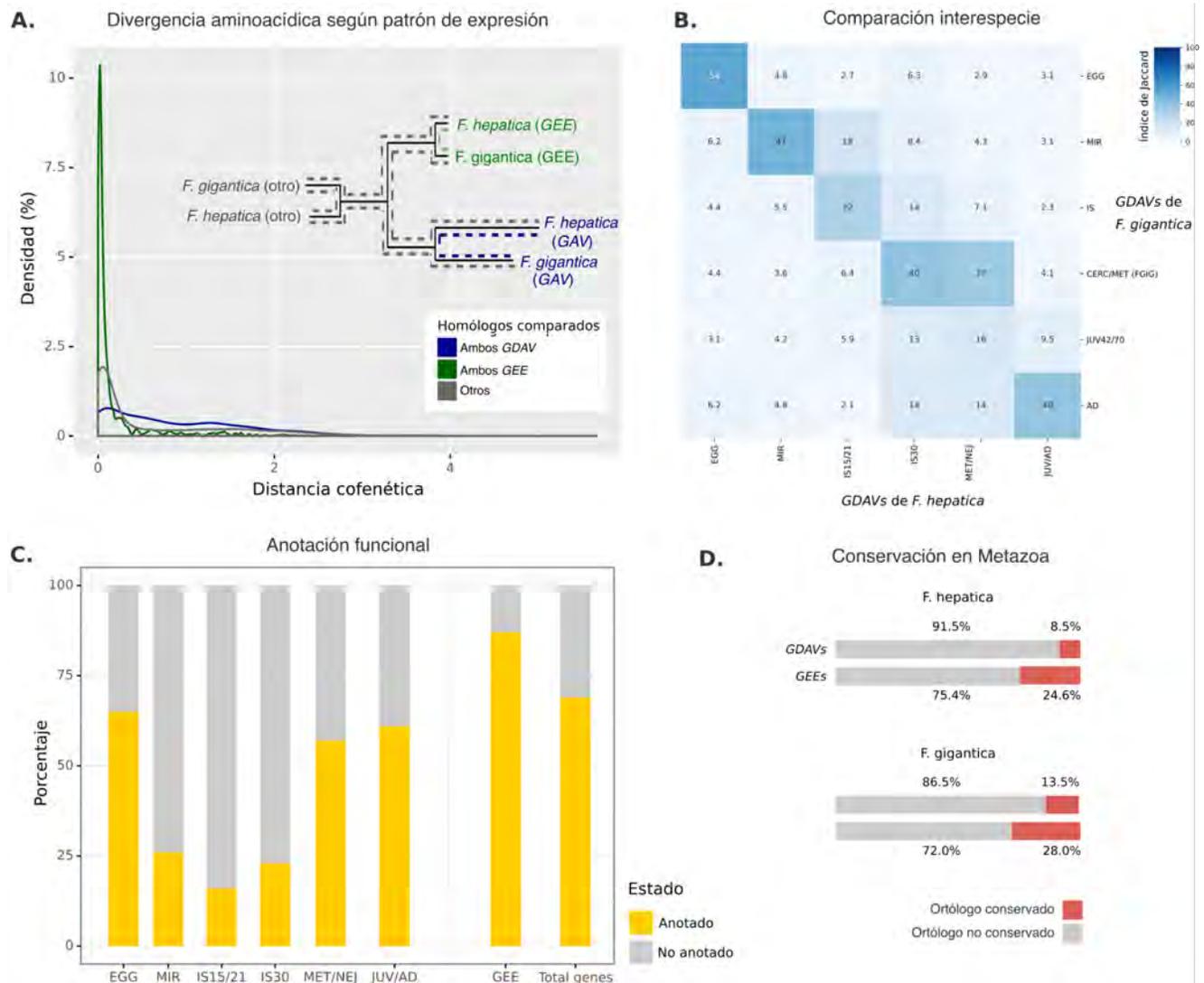


Figura 3.9: Comparación de características de genes según su perfil de expresión a lo largo del ciclo de vida de *F. hepatica*. A. Distribuciones observadas para distancia cofenética (*i.e.* largo de rama) calculada para pares de genes homólogos de *F. hepatica* y *F. gigantica*. Se clasifican las comparaciones según estas impliquen la comparación de GDAV, GEE o genes no clasificados en estas categorías. B. Solapamiento entre clusters de GDAVs de *F. hepatica* y *F. gigantica*, cuantificados según índice de Jaccard basado en grupos de homólogos inferidos con OrthoFinder v2. C. Porcentaje de genes anotados funcionalmente en *F. hepatica*. Se presentan agrupadas todas las anotaciones funcionales empleadas, mostrando los resultados para clusters de GDAVs (izquierda), y para GEEs y la totalidad de la anotación génica empleada en este estudio (derecha). D. Proporción de GDAVs y GEEs de *F. hepatica* y *F. gigantica* que son ortólogos a genes conservados en metazoarios (*i.e.* pertenecientes a grupos de homólogos presentes en 90 por ciento de los metazoarios de OrthoDB v9.0).

conservación de genes con expresión estable a lo largo del ciclo de vida de un organismo es que los mismos codifican para productos génicos involucrados en procesos esenciales para la supervivencia celular/organísmica, en tanto son necesarios sin importar el contexto específico en el que se desarrolla el organismo. Bajo esta concepción resulta natural plantear que genes con tendencia a la expresión en contextos específicos pueden codificar para productos de relevancia en estos contextos, pero cuya función suele ser más prescindible para el desarrollo celular/organísmico.

En la Figura 3.9C se observa la proporción de genes *GDAV* y de expresión estable con presunta ortología (detectada en forma de mejores *hits* recíprocos en alineamientos realizados con DIAMOND) a genes de familias multigénicas de presencia ubicua en metazoarios (*i.e.* al menos 90% de conservación filogenética en los organismos del clado en OrthoDB). En la concepción misma de esta base de datos se plantea que familias génicas constituidas por miembros codificantes para productos de relevancia para la supervivencia de organismos de un taxón dado tenderán a estar ampliamente conservadas en dicho taxón [101]³.

Como se observa en la Figura, los genes *GEE* tienden a estar más conservados en metazoarios que los genes *GDAV*, según revela el porcentaje de los mismos que poseen homólogos en este taxón.

En línea con la concepción expuesta por Waterhouse et al sobre OrthoDB [101] y las distribuciones observadas para distancias co-fenéticas entre *GDAV* y *GEE* de *F. hepatica* y *F. gigantica*, este resultado podría sugerir que la expresión estable de los *GEE* y su mayor conservación a nivel de secuencia tanto entre pares como a nivel filético se debe a que estos genes codifican para productos génicos esenciales para procesos celulares necesarios a lo largo del ciclo de vida de *F. hepatica*. En contraposición, los *GDAV* podrían presentar una expresión más sesgada y una menor conservación a nivel de secuencia y filética por codificar para genes con relevancia funcional para estadios de desarrollo específicos.

Resulta interesante que en ambos casos se observe un porcentaje considerable (~70%) de genes que no presentan homología a nivel de secuencia con miembros de familias multigénicas conservadas en Metazoa. Un escenario posible es que esto se deba a erosión de señal a nivel de secuencia, debido a la amplitud filética de los metazoarios. Si este fuera el caso, sería de esperar que los porcentajes de conservación observados para genes *GDAV* y *GEE* aumentaran conforme se restringe el nivel taxonómico para el cual se estudia conservación (*e.g.* conservación a nivel de Lophotrochozoa).

Por último se analizó la concordancia de expresión para genes *GDAV* homólogos de *F. hepatica* y *F. gigantica*.

En la Figura 3.9B se observa el resultado de comparar los agrupamientos de *GDAV* inferidos para *F. hepatica* y *F. gigantica* en base a los grupos de homólogos para los cuales se detectan miembros pertenecientes a cada conjunto.

Acorde a lo esperado, se observan valores de solapamiento menores en las comparaciones que involucran agrupamientos de *GDAV* con tendencia a la expresión en estadios no homólogos respecto a aquellas comparaciones que contemplan agrupamientos de *GDAV* con tendencia a la expresión en estadios homólogos de ambas especies. El valor mínimo de solapamiento (2.1%) se observa al comparar el agrupamiento “*IS15/21*” de *GDAV* de *F. hepatica* con el agrupamiento “*AD*” de *F. gigantica*. En cambio, la comparación de *GDAV* con mayor expresión en estadios homólogos de ambas especies reporta valores que van desde un 9.5% (agrupamientos “*JUV/AD*” y “*JUV42/70*”, de *F. hepatica* y *F. gigantica* respectivamente) a un 54% (agrupamientos “*EGG*” de ambas especies). En este aspecto resulta de interés remarcar las tendencias observadas para los agrupamientos “*IS30*” y “*JUV/AD*” de *F. hepatica*. Los primeros dos de estos poseen su máximo de solapamiento con el agrupamiento “*CERC/MET*” de *F. gigantica*, el cual contempla a *GDAV* con tendencia a la mayor expresión en los estadios de cercaria y metacercaria de esta

³Como detallan los autores, se permite un margen de 10% de ausencia taxonómica, en tanto la integridad de los ensamblados genómicos es variable según los esfuerzos de investigación dedicados a cada especie, y en tanto organismos con diferentes estilos de vida pueden prescindir de algunos productos génicos (el caso icónico son los organismos parásitos, en donde se observa reducción del genoma para vías relacionadas a productos obtenidos del huésped, entre otros).

especie. Para el caso del agrupamiento “*IS30*” dicha observación es coherente con una predominancia de *GDAV* asociados predominante con la población de cercarias en dichas muestras de estadio intra-caracol. En el caso del agrupamiento “*JUV/AD*”, la observación de un máximo de solapamiento con *GDAV* con tendencia a la mayor expresión en estadio adulto de *F. gigantica*, y no con el agrupamiento “*JUV42/70*”, con máximos de expresión en juveniles, es coherente con la tendencia a la mayor expresión en adultos que poseen los *GDEv* de este agrupamiento (ver Figura 3.8).

Los valores calculados con la metodología empleada podrían ser una subestimación del solapamiento real de genes de homólogos con tendencia a expresar en el mismo estadio de ambas especies. Aspectos relevantes a considerar en este sentido son: *i*) el efecto de fragmentación de los genomas empleados ya comentado previamente [51], el cual podría evitar que genes homólogos en efecto sean asignados a un mismo grupo de homólogos; *ii*) el hecho de que ambos estudios pertenecen a estudios independientes, por lo cual presentan protocolos de trabajo y *batches* de secuenciación distintos que podrían repercutir en los valores de expresión inferidos (afectando de forma más marcada a genes de expresión moderada por posible imposibilidad de detección); *iii*) la existencia de heterogeneidad en las poblaciones celulares secuenciadas, debido al modo de reproducción asincrónica que tienen los estadios intra-caracol, lo cual repercute en las poblaciones de ARNm secuenciadas. Los resultados comparativos reportados en el estudio de expresión génica de cuatro tremátodos en estadio de redia, cercaria y adulto resultan interesantes en este aspecto, en tanto revela una tendencia similar cuando se analizan genes ortólogos (definidos como mejores alineamientos recíprocos de BLASTP) [17]. Los porcentajes de solapamiento observados parecen razonablemente cercanos a los observados aquí (~50%), si se tiene en cuenta que se trata de organismos de diferentes géneros (aunque mismo orden, Plagiorchiida) [17]. La existencia de valores comparables en estudios independientes de características similares da un mayor sustento a la idea de que en efecto los porcentajes de solapamiento observados puedan depender de la aproximación técnica empleada. Alternativamente, podría ser que en efecto existieran diferencias regulatorias sutiles entre los estadios. Estudios de secuenciación a resolución celular (scRNAseq) futuros permitirán establecer hasta qué punto las poblaciones celulares expresan conjuntos homólogos de genes para regular los procesos biológicos llevados a cabo durante la fase parásita intra-caracol.

En resumen, se definió un conjunto de genes diferencialmente expresados con tendencias de expresión estadio-específicas (nombrados *GDAVs*). Si se los compara con genes de expresión más estable, estos *GDAVs* mismos poseer una mayor divergencia a nivel de secuencia respecto a sus homólogos en *F. gigantica* y una conservación taxonómica menor en metazoarios. Ambos resultados sugieren que estos conjuntos de genes con dinámicas transcripcionales diferentes podrían estar evolucionando bajo regímenes de selección también distintos, presumiblemente debido a las funciones celulares que poseen los productos génicos para los cuales codifican. Tal hipótesis debe seguir siendo explorada por estudios futuros.

A su vez, los resultados muestran que estadios homólogos de *F. hepatica* y *F. gigantica* tienden a expresar genes *GDAV* homólogos. De asumirse que dichos homólogos poseen una función similar en ambas especies, tal patrón sugiere una contribución de estos a procesos específicos asociados a dichos estadios de desarrollo. En casos donde dichos *GDAV* expresados sean ortólogos, el patrón sugiere la extensión de dicho valor adaptativo al ancestro común entre ambas especies. Otros tipos de relaciones de homología (*e.g.* paralogía) pueden sugerir la utilización de copias pertenecientes a una misma familia multigénica para cumplir funciones similares en cada estadio de desarrollo.

La presencia de genes homólogos en especies cercanas con tendencias de expresión similar reveladas por estudios independientes define un subconjunto de genes de interés para estudios funcionales futuros enfocados en el estudio de la interacción entre fasciúlidos y huéspedes moluscos. En este sentido, esta aproximación algorítmica puede ser de especial interés para considerar genes no anotados funcionalmente a la fecha que posean valores de expresión elevados o cambios de expresión marcados en los estadios intra-caracol.

Por otro lado, la existencia de genes *GDAV* específicos de uno u otro organismo en un estadio dado podría

tratarse de un artefacto debido a la sensibilidad de las técnicas empleadas para definición de homólogos y la detección de expresión entre otros motivos, así como también a diferencias regulatorias que deberían ser estudiadas por otros abordajes experimentales (*e.g.* scRNA-seq y ensayos de ARNi).

3.2 Análisis funcional de la expresión génica

Los resultados presentados en la sección anterior refieren a la existencia de genes con dinámicas transcripcionales variadas a lo largo del ciclo de vida de *F. hepatica*: un conjunto amplio de genes parece ser expresado en todos los estadios de desarrollo, mientras que otro conjunto de genes se encuentra expresado en dos o más estadios simultáneamente. A su vez, existen genes con una expresión relativamente estable a lo largo del ciclo de vida, mientras que otros poseen una tendencia a la expresión estadio específica.

Realizamos estudios funcionales de la expresión génica a nivel general, buscando identificar los procesos funcionales enriquecidos en los genes de mayor expresión en cada estadio. El estudio funcional de los genes con tendencia a la expresión estadio-específica en huésped intra-caracol se ve dificultado por el bajo porcentaje de anotación que poseen estos genes. No obstante, algunos de estos poseen también expresión diferencial durante las transiciones de desarrollo entre miracidios y estadios intra-caracol, por lo cual será abordado su posible rol funcional más adelante (ver Sección 3.3).

El análisis de los genes de expresión diferencial a lo largo del ciclo de vida de *F. hepatica* según categorías de KEGG BRITE (Figura 3.10), resultó en la anotación de 3053 (28.9%) de los genes analizados a una categoría BRITE, porcentaje similar al observado en la anotación génica empleada, donde ~34% de los genes considerados poseen una anotación a una vía KEGG.

Acorde a lo esperado para el estado de proliferación celular activa que acontece en los estadios larvarios durante su amplificación asexual, se observa un aumento específico de las categorías “*Transcription*”, “*Translation*” y “*Replication and repair*” en los estadios MIR, IS15 e IS21. La misma tendencia fue reportada durante el análisis de redias de *F. gigantica*, aunque en dicho trabajo no se observa la misma tendencia para las muestras de miracidio [16]. El aumento en la expresión de genes en las categorías “*Cell growth and death*”, “*Cancer: overview*”, “*Cancer: specific types*” o “*Development and regeneration*” posiblemente se deba al mismo fenómeno. Acorde a lo esperado, estas últimas categorías también tienen una tendencia al aumento de la expresión en los estadios de huevo y metacercaria, durante los cuales se lleva a cabo el desarrollo de los estadios de infección a los huéspedes molusco y mamífero respectivamente. En consonancia con este estado de proliferación, con aumento transcripcional y traduccionalmente activo, se observa también un aumento de la expresión para la categoría “*Folding, sorting and degradation*” en estos estadios. Esto es esperable, dado que la expresión de proteínas trae aparejada la necesidad de regulación del plegado, el establecimiento de localización y el recambio de las mismas.

Los procesos celulares descritos anteriormente son altamente demandantes desde un punto de vista metabólico. En consonancia con este hecho, se observa también un aumento de la expresión para las categorías “*Carbohydrate metabolism*”, “*Nucleotide metabolism*”, “*Lipid metabolism*” y “*Xenobiotics biodegradation and metabolism*” en estos estadios. Dichas categorías engloban tanto procesos catabólicos como anabólicos, por lo cual es necesario un análisis más detallado para dar una descripción de dichos aspectos del metabolismo (ver análisis de vías metabólicas y ontologías génicas enriquecidas en genes de alta expresión en estos estadios, desarrollados abajo). El caso del metabolismo de lípidos resulta interesante, ya que es generalmente aceptado que los estadios intra-mamífero del parásito no son capaces de realizar catabolismo de estas biomoléculas para obtención de energía [10], un aspecto que analizaremos más detalladamente más adelante (ver Sección 3.3.3.2.2).

Resulta interesante que genes de la categoría “*Glycan biosynthesis and metabolism*” muestren un aumento en su expresión en estadios intra-caracol y de invasión temprana a huésped mamífero, lo cual podría estar vinculado a la formación y mantenimiento del tegumento en estos estadios. Es sabido que el tegumento

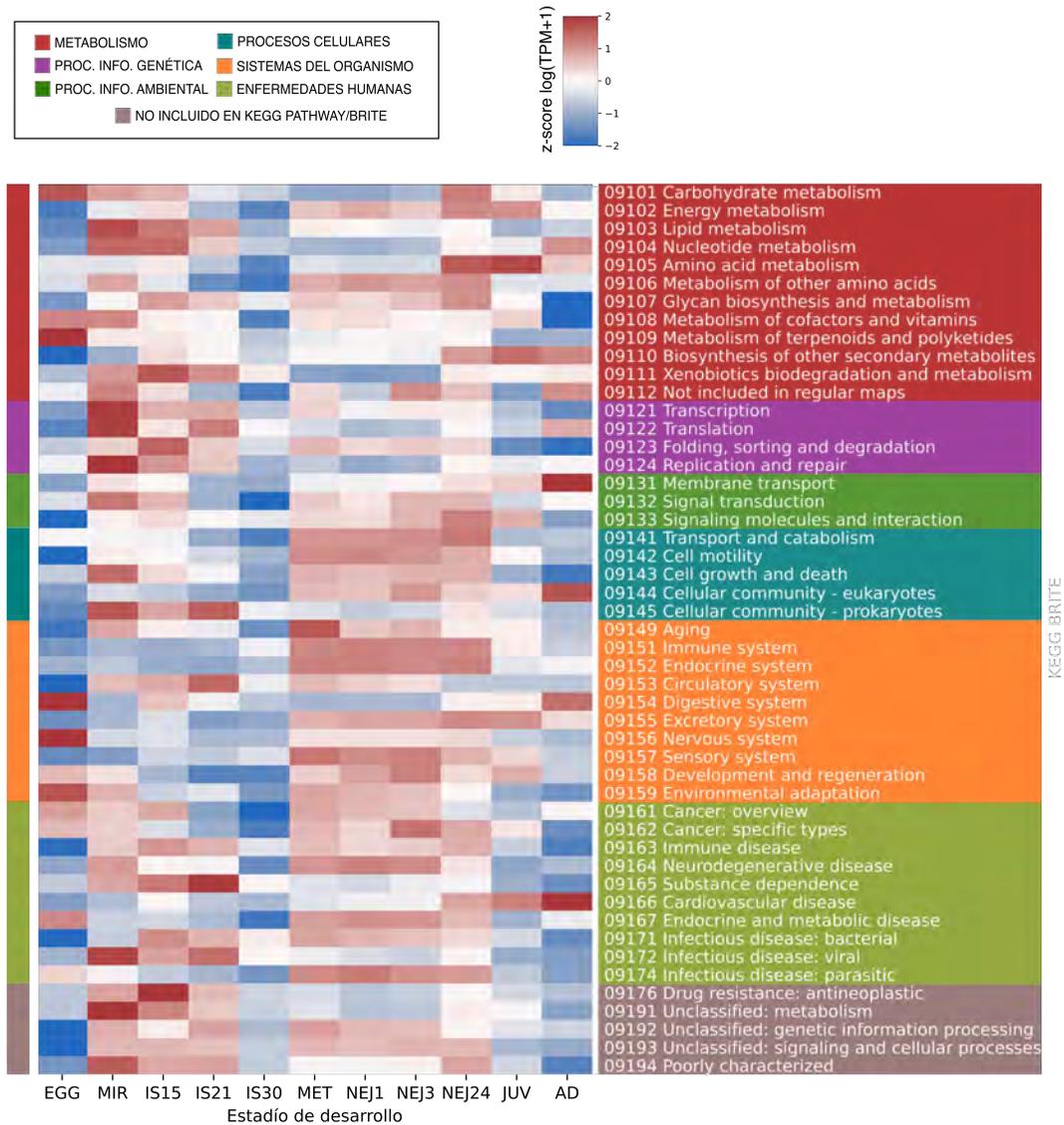


Figura 3.10: Cambio en niveles de expresión para genes agrupados según KEGG BRITeS a lo largo del ciclo de vida. Se grafica el cambio en los niveles de expresión para el total de genes en dichas categorías, mostrándose el $z\text{-score}(\log_2(\text{TPM}+1))$ en cada caso.

de *F. hepatica* presenta una rica cubierta glicoproteica con recambio activo al menos en el proceso de invasión a los mamíferos, siendo posible que un proceso similar se de en la interacción con el huésped caracol. Una tendencia similar se observa para *F. gigantica* [16].

Otros aspectos singulares de los estadios larvarios infectantes de caracol que parecen verse reflejados en las tendencias de expresión génica son

i) la adaptación al ambiente, necesaria para el estadio de vida libre miracidio (con tendencias similares observadas en huevo y metacercaria, también estadios de vida libre), observada en el término “*Environmental adaptation*” y

ii) la capacidad infectiva de los mismos, reflejada en las categorías “*Infectious disease: bacterial*” o “*Infectious disease: viral*”. Interesantemente, la categoría “*Infectious disease: parasitic*” posee su nivel de expresión más alto en estadios de infección al huésped mamífero.

Por último, KEGG BRITE incluye en su sistema de clasificación las categorías “*Unclassified: metabolism*”, “*Unclassified: genetic information processing*”, “*Unclassified: signalling and cellular processes*” y “*Poorly characterized*”, que representan genes pobremente caracterizados. Estas categorías constituyen una fracción marginal (~4%, 132/3053) de los genes analizados que poseen anotación para un KEGG BRITE, por lo que la enorme mayoría de los genes asignados a un KEGG BRITE poseen una anotación informativa. No obstante, es importante tener presente que la fracción anotada mediante KEGG BRITE es muy minoritaria (28.9%), por lo que el análisis en sí es realizable sobre una fracción acotada de los genes analizados.

Cuando se analiza el enriquecimiento en vías metabólicas KEGG para el 10% de los genes de mayor expresión en los estadios estudiados, se observa la presencia de 7 vías enriquecidas en miracidios, 6 en *IS15*, 5 en *IS21* y 7 en *IS30* (Tabla 3.3). El análisis de enriquecimiento de términos GO para este mismo set de genes de alta expresión resalta algunos grupos y procesos (Figura 3.11). Ambos análisis son concordantes, aunque el análisis para vías metabólicas reportó un número sensiblemente menor de términos enriquecidos. Es probable que esto se deba a la menor anotación que poseen las proteínas de alta expresión para estos términos funcionales, si se tiene en cuenta que solo el 53% de los genes anotados poseen anotación para algún término GO.

Tabla 3.3: Vías KEGG enriquecidas para el conjunto de genes pertenecientes al 10 por ciento más expresado en estadios larvarios de *F. hepatica*, reportadas por gProfileR.

Vía KEGG (ID)	MIR	IS15	IS21	IS30
Proteasome (map03050)	28/448; p = 1.40e-13	30/464; p = 1.24e-15	23/422; p = 5.95e-9	23/416; p = 4.19e-9
Carbon metabolism (map01200)	28/448; p = 3.42e-2	31/464; p = 4.23e-3	-	28/416; p = 8.34e-3
Citrate cycle (TCA cycle) (map00020)	15/448; p = 2.75e-3	19/464; p = 4.15e-6	15/422; p = 1.19e-3	16/416; p = 1.80e-4
RNA transport (map03013)	46/448; p = 2.34e-5	-	-	-
Spliceosome (map03040)	48/448; p = 2.92e-6	41/464; p = 5.05e-3	37/422; p = 1.36e-2	36/416; p = 2.07e-2
Oxidative phosphorylation (map00190)	40/448; p = 3.11e-10	49/464; p = 7.87e-17	47/422; p = 6.93e-17	49/416; p = 5.04e-19
Ribosome (map03010)	75/448; p = 2.43e-27	75/464; p = 2.79e-26	77/422; p = 3.25e-31	74/416; p = 8.98e-29
Protein processing in endoplasmic reticulum (map04141)	-	-	-	37/416; p = 3.08e-2

Se observa enriquecimiento funcional en términos vinculados a varios de los procesos biológicos previamente descritos, particularmente los vinculados a procesos de biosíntesis de nucleótidos, metabolismo energético, expresión génica y desarrollo, e inmunidad, en consonancia con los procesos biológicos desarrollados dentro del huésped caracol.

Varios de los términos GO enriquecidos están asociados a la biosíntesis de nucleótidos (“*nucleoside phosphate biosynthetic process*” (GO:1901293), “*ribonucleoside triphosphate metabolic process*” (GO:0009199), “[*purine/pyridine*] *nucleotide metabolic process*” (GO:0006163, GO:0019362)). Dentro de los genes de alta expresión en estadios miracidio e intra-caracol anotados con dichos términos se

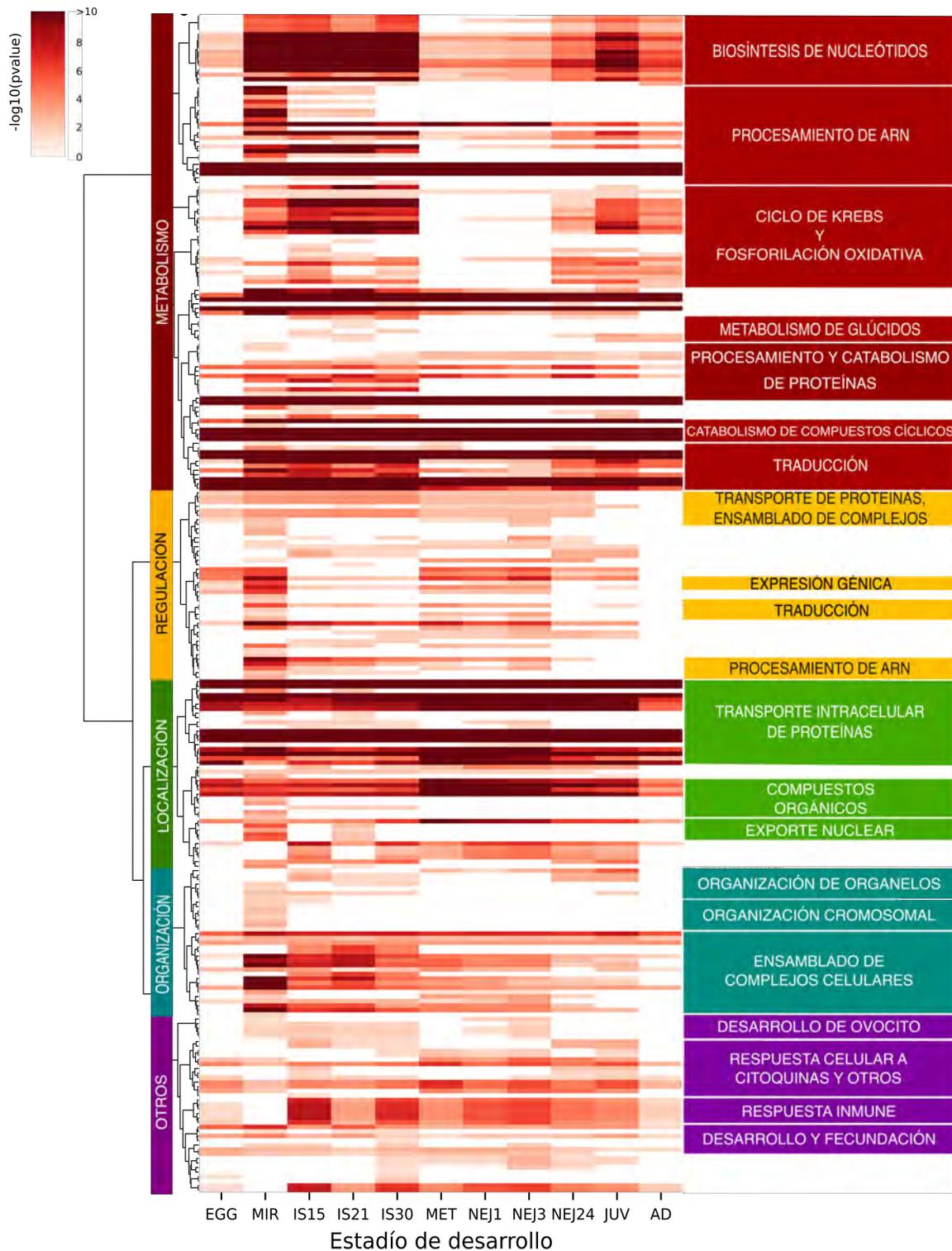


Figura 3.11: Heatmap de términos GO (ontología "biological_process", correspondiente a procesos biológicos) enriquecidos en proteínas de mayor expresión en estadios de miracidio e intra-caracol. Se muestran los valores de enriquecimiento de estos términos en el resto de estadios analizados. Para mayor claridad se muestran términos con un nivel de especificidad moderada (*i.e.* niveles de 4 a 9), los cuales fueron agrupados en base a similitud semántica y categorizados manualmente (ver *Materiales y Métodos*)

encuentran genes de la vía de salvataje de purinas, tales como *adk*, *adss*, *ada* y *hprt1*. Este último gen aumenta su expresión durante la transición miracidio-esporocisto, sugiriendo la relevancia de esta vía para el desarrollo intra-caracol de *F. hepatica* (este punto será discutido en mayor profundidad en la Sección 5.4.2.4). Algunos de los términos relacionados con la organización cromosomal, tales como “[nucleosome/chromatin]* *assembly*” (GO:0034728, GO:0031497), son coherentes con esta interpretación.

Otro grupo de genes identificados en este set codifican para proteínas con participación en vías canónicas obtención de energía mediante el catabolismo de nutrientes y subsecuente síntesis de ATP. Estos productos génicos se ven asociados también al enriquecimiento en términos vinculados al ciclo de Krebs y la fosforilación oxidativa. Entre los mismos se encuentran “*mitochondrial ATP synthesis coupled to electron transport*” (GO:0042775), junto a términos referentes a pasos de la cadena de transporte de electrones mitocondrial (“*mitochondrial electron transport, [NADH to ubiquinone/succinate to ubiquinone/ubiquinol to cytochrome c]*”, términos GO:0006120, GO:0006121, y GO:0006122, respectivamente) o el término “*oxidative phosphorylation*” (término GO:0006119). Procesos relacionados al metabolismo de intermediarios del ciclo de Krebs se reportan como enriquecidos, tales como “[*succinate/oxaloacetate/pyruvate/citrate*] *metabolic process*” (GO:0006105, GO:0006107, GO:0006090, GO:0006101).

Entre los genes de alta expresión vinculadas al ciclo de Krebs y la obtención de energía por fosforilación oxidativa se encuentran: *i*) genes codificantes para subunidades de todas las proteínas del ciclo, a excepción de la fumarasa, *ii*) subunidades de la ATP sintasa mitocondrial, *iii*) componentes de la citocromo c oxidasa (i.e. COX4, COX5), *iv*) subunidades del complejo NADH deshidrogenasa y *v*) UQCRFS1 y UQCRQ del complejo III (citocromo c reductasa). Varios de los términos antes descritos no se encuentran enriquecidos en estadios de vida libre como el huevo y la metacercaria (o las primeras etapas del desenquiste juvenil, ya dentro del huésped mamífero), siendo aparentemente este aspecto más específico de los estadios parasitarios. En efecto, aún existiendo transcripción para estos genes en estadios intra-mamífero, el consenso en la literatura apunta hacia la utilización de un metabolismo anaerobio a medida que los juveniles de reciente desenquiste se desarrollan y migran hacia ducto biliar [10,102]. Estudios desde una perspectiva translatómica y proteómica podrían ayudar a determinar hasta qué punto los transcritos de estos genes se traducen en estos estadios y si existe una regulación fina post-transcripcional o traduccional asociada a estos cambios metabólicos reportados en la literatura.

Varios de los términos enriquecidos se encuentran vinculados a la expresión génica y síntesis de proteínas, así como a la regulación de dichos procesos.

El estado transcripcionalmente activo de los estadios asociados al huésped caracol se refleja en el enriquecimiento para términos tales como “*gene expression*” (GO:0010467), “[*rRNA/mRNA/ncRNA*] *metabolic process*” (GO:0016072, GO:0016071, GO:0034660), “[*RNA 3'/mRNA/ncRNA*] *processing*” (GO:0031123, GO:0006397, GO:0034470) o “*posttranscriptional regulation of gene expression*” (GO:0010608). Nuevamente, el enriquecimiento en términos relacionados con la organización estructural del ADN tales como “*DNA geometrc change*” (GO:0032392) o “*DNA duplex unwinding*” (GO:0032508) también puede ser relacionado a este estado transcripcionalmente activo, en tanto cambios topológicos son necesarios para que se lleve a cabo la transcripción.

Términos como “*RNA splicing*” (GO:0008380) o “*regulation of RNA splicing*” (GO:0043484) también forman parte de los términos vinculados al fenómeno de transcripción, e interesantemente algunos de estos términos se encuentran enriquecidos únicamente en miracidio y estadios intracaracol. Dentro de los genes que poseen altos niveles de expresión en estos estadios vinculados al proceso de *splicing* se encuentran aquellos codificantes para componentes proteicos del *spliceosoma* tales como snRNPs (e.g. snRNPc, snRNP70) y proteínas reguladoras tales como factores auxiliares del proceso de *splicing* (SF1, U2AF1 y 2), reguladores de la familia SR (SRSF1) o proteínas de la familia hnRNPs [103]. Se encuentran también miembros del complejoEJC (“*exon-junction complex*”, en inglés) como *magoh*, *rbm8A* y *eIF4A3*, el cual no solo forma parte de la regulación del *splicing* del ARNm, sino también en la interacción con

moléculas encargadas de su transporte al núcleo y en la estabilidad del ARNm [104,105].

También poseen un rol fundamental en el transporte de ARNm al núcleo los genes *ALYREF*, *THOC3* y *DDX39A/UAP56*, miembros del conservado complejo TREX que acopla la maduración y exporte a citosol de ARNm [106]. Estos genes de alta expresión se asocian al enriquecimiento en términos GO tales como “*RNA export from nucleus*” (GO:0006405).

Ontologías génicas vinculadas a la traducción tales como “*translation*” (GO:0006412), “*translational initiation*” (GO:0006413) o “*regulation of translation*” (GO:0006417), y las asociadas al ensamblado de organelos vinculados, tales como “*ribosome biogenesis*” (GO:0042254) aparecen enriquecidas. Entre los genes de alta expresión vinculados a estos procesos se destacan: i) tRNA sintetetasas, tales como lisil-, treonin-, aspartil- o arginin-tRNA sintetetasas, la EPRS bifuncional o FARSA, entre otras; ii) subunidades mayores y menores de las proteínas ribosomales, RPL; iii) factores de iniciación; iv) factores de elongación v) factores de terminación traduccional. Algunos de los términos enriquecidos refieren al transporte intracelular que ocurre durante o después de la síntesis proteica, tales como “*protein targeting to ER*” (GO:0045047), “*protein targeting to membrane*” (GO:0006612) o “*establishment of protein localization to organelle*” (GO:0072594).

Respecto al catabolismo de proteínas, se observa la expresión de varios genes codificantes para proteínas componentes del proteasoma (*e.g.* de las familias PSMA, B, C y D) o del sistema de ubiquitinación, como UBE2C o UbcD1. La presencia de estos se refleja en el enriquecimiento en términos como “*proteasome-mediated ubiquitin-dependent protein catabolic process*” (GO:0043161). Asociadas a esta anotación se encuentran también genes codificantes para las proteínas SKP1 y RBX1, miembros del complejo SCF involucrado en la regulación del ciclo celular a través de la degradación de ciclinas [108].

Otro conjunto de los términos GO enriquecidos se encuentran relacionados a procesos inmunes y de desarrollo llevados a cabo dentro del huésped caracol.

En el primero de estos grupos se destaca la presencia de términos GO asociados a respuesta inmune, tales como “*response to cytokine*” (GO:0034097), “*interleukin-12 signaling pathway*” (GO:0035722), “*cellular response to interleukin-4*” (GO:0071353) o “[*myeloid cell/neutrophil/leukocyte*] *activation involved in immune response*” (GO:0002275, GO:0002283, GO:0002366). Una inspección de los genes de alta expresión anotados con dichos términos muestra un set diverso de genes codificantes para proteínas de la interfaz huésped-parásito como *hsp70*, *catB*, *gst*, y la tetraspanina *cd63*. Si bien se ha reportado en otros organismos un rol en el procesamiento y presentación de antígenos o la activación de células del sistema inmune por parte de dichas proteínas [109], las mismas suelen ser asociadas con procesos como la invasión de tejidos, el procesamiento de nutrientes o la inmunomodulación en el contexto de la infección de platelmintos [111–115]. A pesar de ello no se observa enriquecimiento en vías particulares de respuesta inmune en el análisis de términos KEGG.

Por último, se observa enriquecimiento en términos GO asociados a procesos de desarrollo, tales como “*liver development*”, “*oocyte axis specification*” (GO:0007309) o “*intracellular mRNA localization involved in anterior/posterior axis specification*” (GO:0060811).

Como ya se dijo anteriormente, los procesos de proliferación celular, embriogénesis y morfogénesis son de una relevancia crucial en los estadios biológicos intra-caracol ya que en su conjunto estos eventos son los que permiten la amplificación asexual en el huésped. Dentro de los genes anotados con estos términos GO se encuentran *pcna*, *llgl1* y *ezh2*. El primero de estos codifica para la abrazadera involucrada en dar alta procesividad a la ADN polimerasa δ en eucariotas, siendo un marcador asociado a la progresión del ciclo celular ya reportado en poblaciones de células proliferantes de *S. mansoni* [14,40,42]. Dicho gen es considerado un marcador de células proliferantes en general, por lo cual es esperable que se encuentre altamente expresado en estadios de desarrollo constituídos por células tipo neoblasto. La alta expresión de *llgl1* podría en cambio relacionarse con procesos de morfogénesis, en tanto se ha demostrado por silenciamiento de ARN el rol de un homólogo de este gen en el desarrollo del tegumento y la deposición de huevos en el tremátodo *S. japonicum* [116].

Un aspecto relevante para el mantenimiento de la pluripotencia de células madre tipo neoblasto es el mantenimiento de represión de genes vinculados a la diferenciación celular. En este sentido resulta interesante la alta expresión de un gen codificante para EZH2, la subunidad catalítica del complejo PRC2 que regula patrones de expresión mediante la trimetilación de histonas y consecuente represión de expresión génica [117]. Se ha demostrado, mediante su silenciamiento selectivo *in vitro* el rol de esta enzima en la regulación de la expresión de genes vinculados al desarrollo de huevos, homeostasis celular, replicación de ADN y metabolismo de ARN no codificante (principalmente ARNr y ARNt) en *S. mansoni* [118]. A la vez, consistentemente con nuestras observaciones, se ha mostrado un aumento de su expresión en la progresión de miracidios a esporocistos en *S. mansoni* [14,119].

El mantenimiento de los neoblastos de planaria requiere por su parte de los genes del complejo EJC (*magoh*, *rbm8* y *eIF4A3*) [120]. Es posible que la alta expresión de *magoh* pueda implicar un rol similar para el mantenimiento de células madre proliferantes en *F. hepatica*. Alternativamente, se ha reportado que este gen posee un rol en la especificación del eje antero posterior en el ovocito de *D. melanogaster* [121]. Es posible que un rol homólogo se desempeñe en el desarrollo intra-caracol de tremátodos, incluyendo a *F. hepatica*.

La alta expresión de genes vinculados a procesos de proliferación, mantención de la pluripotencia, embrión y morfogénesis es coherente con la necesidad de expresión de proteínas que desarrollen estos procesos durante la proliferación asexual en los estadios intra-caracol. Como es de esperar, otros genes relevantes para estos procesos pueden poseer una expresión moderada y baja, siendo su regulación durante la transición de estadios un punto de control para estos procesos. En coherencia con los resultados mencionados aquí, también se observa esta expresión diferencial en estadios intra-caracol. Se analizará este punto en detalle en la sección referente a expresión diferencial (subsección 3.3.3.1).

En resumen, un análisis funcional de los estadios de miracidio y estadios intra-caracol de *F. hepatica* revela una tendencia al aumento de la expresión en genes vinculados a la replicación celular, con un estado transcripcional y traduccional activo. Esta tendencia se observa a su vez en el conjunto de genes con mayor expresión en estos estadios, los cuales se vinculan a procesos de biosíntesis de ácidos nucleicos y proteínas, así como a la regulación de dichos procesos y de la localización intracelular de los productos celulares sintetizados. Este estado de biosíntesis es energéticamente demandante; genes codificantes para proteínas asociadas al transporte y obtención de energía a partir de nutrientes también poseen una alta expresión. Se observó también enriquecimiento funcional para categorías con posible vínculo a la interfaz huésped-parásito (como lo son términos funcionales asociados a procesos inmunes) y al desarrollo, aunque la interpretación de la lista de genes asociada no resulta trivial.

Los análisis de enriquecimiento funcional para genes de alta expresión con términos KEGG y GO son en términos generales concordantes. Por último, resulta importante recalcar que varios de los términos enriquecidos en los genes de mayor expresión muestran enriquecimiento en todos los estadios; esto probablemente se vincule con la esencialidad de los procesos celulares mencionados. Tales procesos suelen a su vez estar sobre-representados en la anotación funcional, por lo cual es de esperar que procesos más específicos vinculados a la expresión de genes menos conservados (y por tanto peor anotados) no sean revelados por tales análisis.

3.3 Análisis de expresión diferencial

En la Sección 3.1.3 se presentó de forma general los patrones de expresión conservados y con tendencia a la expresión estadio-específica a lo largo del ciclo de vida de *F. hepatica*. Esta aproximación permite discernir entre genes candidatos a codificar para productos necesarios en todos los estadios de desarrollo y aquellos genes que podrían ejercer un rol específico en un conjunto acotado de estadios de desarrollo.

Otro conjunto de genes puede ubicarse en tendencias intermedias entre estos extremos del espectro de la dinámica transcripcional. A modo de ejemplo, desde el punto de vista del desarrollo del ciclo de vida de un organismo también son relevantes aquellos genes que presentan una regulación en su expresión durante las transiciones entre estadios de desarrollo consecutivos. Estos cambios son candidatos a estar vinculados a procesos biológicos relevantes para dichos estadios, sin ser necesaria una especificidad en la expresión de estos genes. Genes vinculados a vías de desarrollo (*e.g.* embriogénesis) constituyen un ejemplo claro de esto, en tanto pueden presentar cambios de expresión periódicos a lo largo del ciclo de vida.

El estudio de cambios de expresión entre estadios contiguos se acotó al estudio de las transiciones entre muestras de estadios larvarios asociados al huésped caracol, en tanto las mismas son el objeto principal de análisis de este trabajo. Si bien se poseen muestras de estadio intra-caracol 21 d.p.i., las mismas fueron excluidas de los análisis realizados dado que pertenecen a un *batch* de secuenciación diferente al resto de las muestras, por lo cual la utilización de las mismas con criterios convencionales de análisis de expresión diferencial podría traer aparejado la aparición de una tasa considerable de falsos positivos.

3.3.1 Comparativas de expresión en estadios analizados

En las Figuras 3.12 y 3.14 se observa el conjunto de genes diferencialmente expresados (*GDEs* de ahora en más) entre miracidios y estadios caracol 15 d.p.i. y entre este último y los estadios intra-caracol 30 d.p.i., respectivamente.

De los 1545 *GDEs* detectados, 330 se encuentran sobreexpresados en miracidios y los restantes 1215 se encuentran sobreexpresados en estadio intrac-aracol 15 d.p.i.

No se detectó enriquecimiento funcional para ontologías GO o vías metabólicas KEGG en el caso de los *GDEs* sobreexpresados en miracidios. No obstante, es posible identificar dentro de este conjunto a genes reportados que han sido vinculados a los procesos infectivos, de transducción de señal y movilidad frente a estímulos externos y al metabolismo de glúcidos en miracidios y otros estadios de desarrollo de tremátodos. Algunos de estos se encuentran resaltados en la Figura 3.12 y mencionados en la Sección 3.3.2. Sí fue posible detectar enriquecimiento funcional en el caso de los genes sobreexpresados en estadios intra-caracol 15 d.p.i. durante esta transición, los cuales podrían relacionarse con procesos biológicos clave desempeñados dentro del huésped intermediario caracol (Figura 3.13).

La presencia de términos como *Protein digestion and absorption* o *Lysosome* podría asociarse a la sobreexpresión de genes codificantes para proteasas de las familias de las catepsinas y legumainas. El rol de proteínas de estas familias en los procesos de digestión de proteínas del huésped tanto para la facilitación de procesos migratorios como para la obtención de materia prima para los procesos biosintéticos desempeñados durante los estadios de desarrollo intra-caracol han sido descritos en la literatura en casos como el de *S. mansoni*, y serán discutidos en detalle más adelante (ver Secciones 3.3.3.3 y 3.3.3.2.1). Vías como *ECM-receptor interaction* y *Neuroactive ligand receptor interaction* también podrían verse relacionadas al proceso migratorio en estadios intra-caracol, en tanto la interacción con la matriz extracelular del huésped y el control neuromotor son necesarios para que se desempeñe la función locomotora. Por otro lado, la sobreexpresión de algunos de estos genes codificantes para proteasas, así como de genes codificantes para proteínas como NPC2 o HSP70/90 podría explicar también el enriquecimiento observado para el término *Antigen processing and presentation*, en tanto las mismas participan de dichos procesos en algunos vertebrados. Como ya se ha mencionado, el rol de estas proteínas en el contexto de infección de parásitos platelmintos suele estar más asociado en la literatura con el transporte de nutrientes o la respuesta al estrés oxidativo, aunque existen estudios que indagan sobre su rol en la modulación de respuesta inmune en huéspedes vertebrados.

Otro conjunto de vías hacen referencia a la adhesión celular y las estructuras celulares que aseguran la misma en tejidos. Entre estas vías se pueden mencionar *Cell adhesion molecules*, *Focal adhesion* y

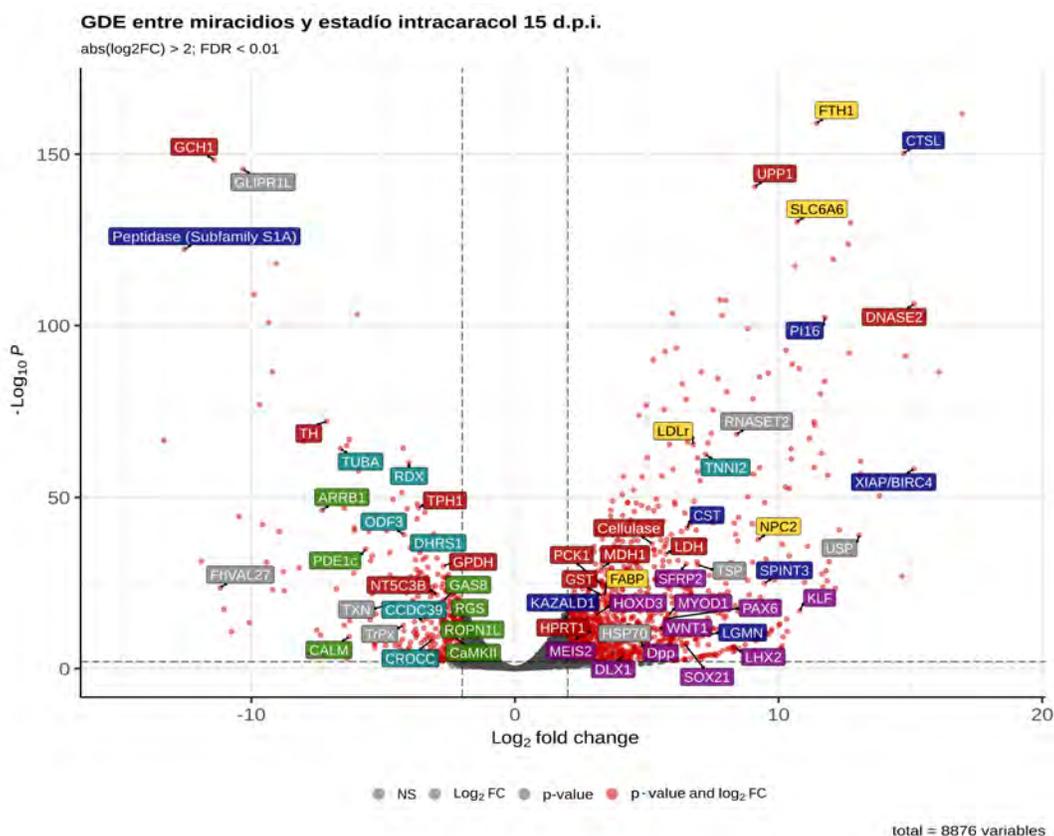


Figura 3.12: Genes diferencialmente expresados (GDEs) entre miracidios y estadios intra-caracol 15 d.p.i. reportados por edgeR. Se resalta un conjunto de GDEs para los cuales se posee información bibliográfica sobre su posible rol funcional en miracidios y esporocistos de tremátodos, coloreándose a estos según categorías funcionales (ver leyenda). Se considera a un gen como diferencialmente expresado si se reporta $|\log_2\text{FC}| > 2$ entre condiciones y $\text{FDR} < 0.01$ en edgeR, aplicándose el método de corrección de Benjamini-Hochberg para control de falsos positivos.

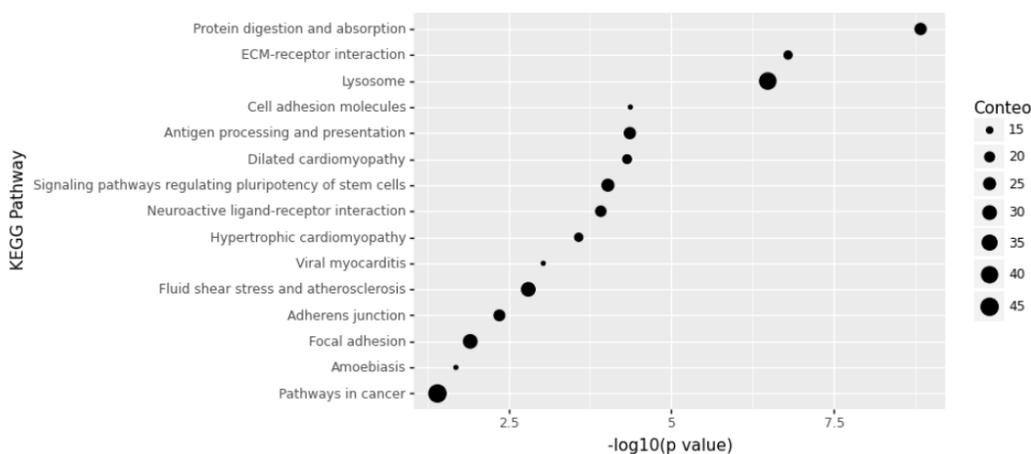


Figura 3.13: Vías metabólicas KEGG reportadas por gprofiler2 como enriquecidas en genes sobreexpresados en estadio intra-caracol 15 d.p.i. respecto a miracidios. Se reporta las vías funcionales enriquecidas en estos genes empleando como referencia al conjunto de genes expresados en el estadio. Los resultados son ordenados en el eje de las abscisas según su significancia estadística, y la cantidad de genes sobreexpresados anotados como pertenecientes a una vía es proporcional al radio del círculo que representa a dicha vía.

Adherens junction. Las mismas podrían estar vinculadas con los procesos de morfogénesis de la superficie parasitaria y a la interacción con la matriz del huésped durante el proceso migratorio.

Por último, aunque no menos importante, cabe resaltar el enriquecimiento para la vía *Signaling pathways regulating pluripotency of stem cells*. El enriquecimiento en esta vía se ve vinculado a la sobreexpresión de genes asociados al mantenimiento de la pluripotencia y otros procesos de desarrollo, como lo son los ligandos de la familia Wnt y receptores de la familia Frizzled. La sobreexpresión de genes vinculados a esta vía regulatoria es esperable, en tanto como se mencionó en la introducción a esta tesis la transición de miracidios a estadios intra-caracol inicia el proceso de proliferación asexual y morfogénesis basada en la presencia de células tipo neoblastos en otros tremátodos (*i.e.* *S. mansoni*). En la Sección 3.3.3.1 se discute en mayor detalle la sobreexpresión de miembros de estas familias.

En resumen, el análisis funcional resulta una primera aproximación informativa respecto a algunos de los procesos que podrían operar durante la transición de miracidios a estadios intra-caracol 15 d.p.i., observándose enriquecimiento para vías vinculadas a procesos biológicos de relevancia según la literatura en transiciones similares entre estadios de otros tremátodos. La inspección de los GDEs más allá de este análisis es también relevante, en tanto la anotación funcional puede verse limitada o estar vinculada a los procesos más estudiados en organismos modelo. Esto podría explicar el enriquecimiento en algunos términos como “*Amoebiasis*”, “*Fluid shear stress and atherosclerosis*”, “*Viral myocarditis*” o “*Dilated cardiomyopathy*”, las cuales son funciones que podrían asociarse a estos genes en otros organismos

En el caso de la transición entre estadios intra-caracol 15 y 30 d.p.i. (Figura 3.14), la cantidad de genes reportados como expresados diferencialmente en la transición es menor al caso anterior. Tal resultado es esperable, en tanto los estadios de desarrollo intra-caracol presentan alta similitud a nivel transcripcional entre sí (resultados presentados a lo largo de la Sección 3.1).

Similar a lo observado en los genes sobreexpresados en miracidios, no se observa enriquecimiento funcional para los genes sobreexpresados en estadio intra-caracol 15 d.p.i.. Esto parece razonable si se tiene en cuenta la cantidad de genes sobreexpresados en este estadio. Al igual que los genes que sobreexpresan en este estadio respecto a miracidios, muchos de estos genes están vinculados a vías de desarrollo, por lo que serán mencionados en conjunto a estos.

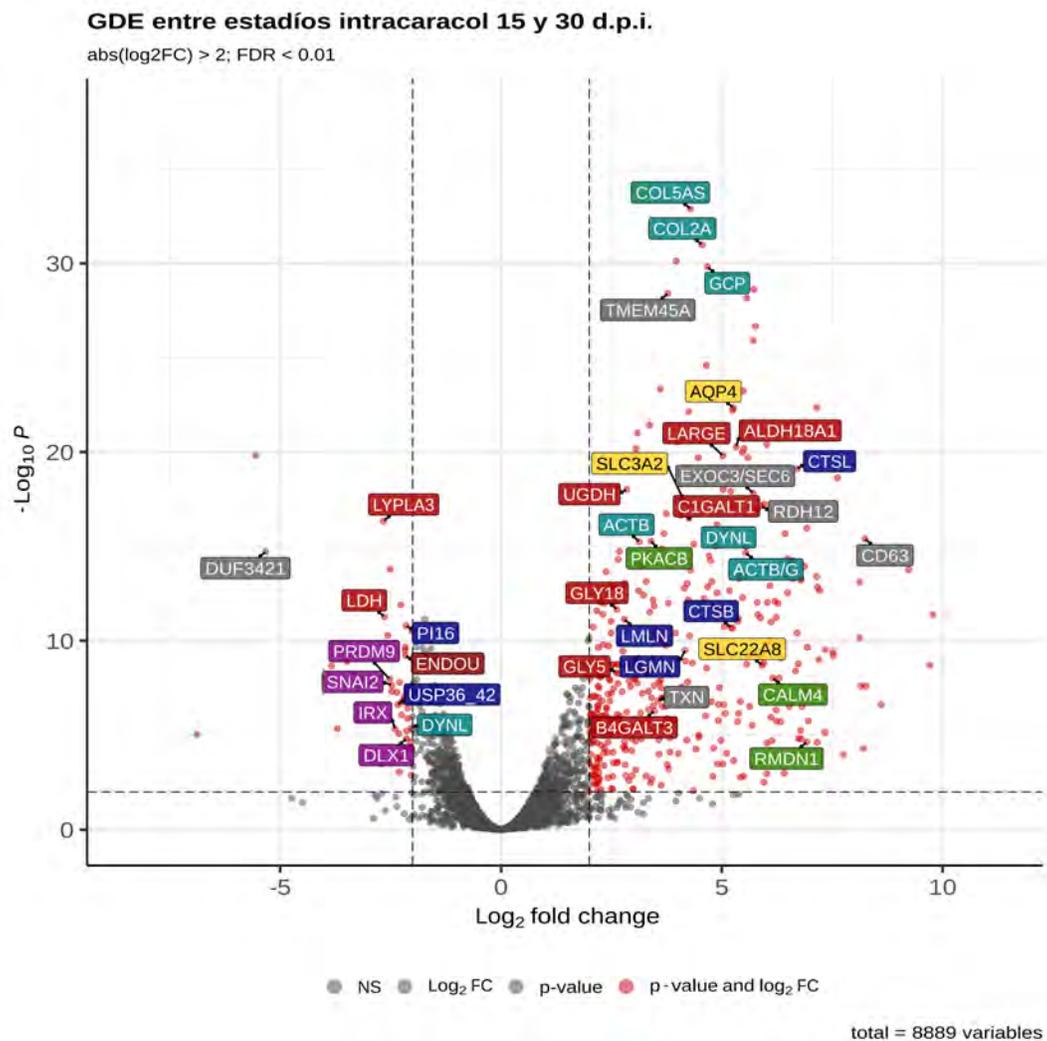


Figura 3.14: Genes diferencialmente expresados (GDEs) entre estadios intra-caracol 15 y 30 d.p.i. reportados por edgeR. Se resalta un conjunto de GDEs para los cuales se posee información bibliográfica sobre su posible rol funcional en miracidios y esporocistos de tremátodos, coloreándose a estos según categorías funcionales (ver leyenda). Se considera a un gen como diferencialmente expresado si presenta $|\log_2\text{FC}| > 2$ entre condiciones y $\text{FDR} < 0,01$, aplicándose el método de corrección de Bejamini-Hochberg para control de falsos positivos.

Una revisión de los genes asociados a estos términos revela nuevamente una presencia ubícua de proteasas de las familias de las catepsinas B y L, así como legumaínas (Figura 3.15).

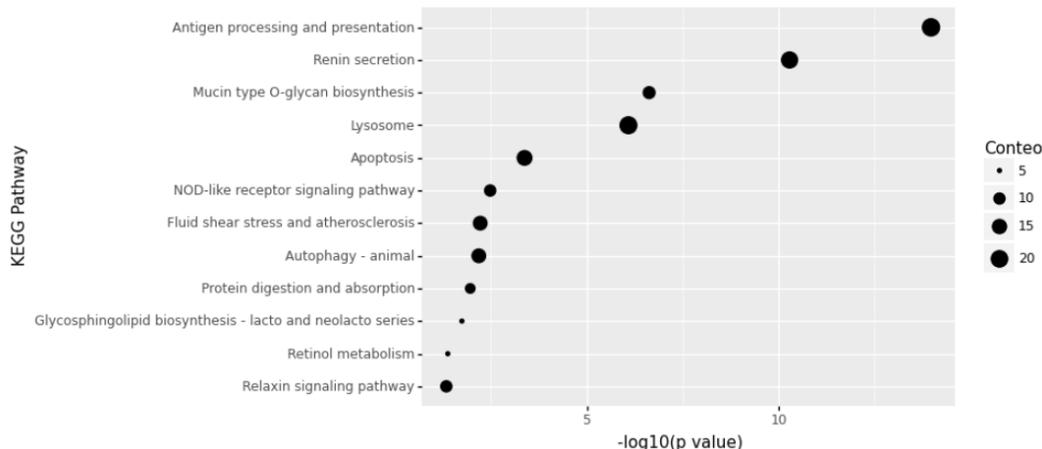


Figura 3.15: Vías metabólicas KEGG reportadas por gprofiler2 como enriquecidas en genes sobreexpresados en estadio intra-caracol 30 d.p.i. respecto a 15 d.p.i.. Se reporta las vías funcionales enriquecidas en estos genes respecto al conjunto de genes expresados en el estadio, empleado como conjunto de referencia para el estudio de enriquecimiento. Los resultados son ordenados en el eje de las abscisas según su significancia estadística. La cantidad de genes sobreexpresados anotados como pertenecientes a una vía es proporcional al radio del círculo que representa a dicha vía.

Interesantemente, miembros de estas familias se encuentran también sobreexpresadas en estadio intracaraco 15 d.p.i. respecto a miracidio. La presencia de las mismas se ve representada en el enriquecimiento en los términos “*Antigen processing and presentation*”, “*Lysosome*”, “*Apoptosis*”, “*NOD-like receptor signaling pathway*”, “*Autophagy - animal*” y “*Protein digestion and absorption*”. Nuevamente, el rol relativo de estos genes a dichos procesos en el contexto de la infección a huésped caracol debería ser investigado en mayor detalle con estudios funcionales. No obstante, en la literatura es casi nula la mención de contribución de catepsinas de tremátodos a procesos como la señalización por medio de receptores NOD, en tanto su aporte a otros procesos parece ser más parsimonioso. Interesantemente se observa enriquecimiento en vías de biosíntesis de compuestos glucídicos como “*Mucin type O-glycan biosynthesis*” o “*Glycosphingolipid biosynthesis - lacto and neolacto series*”. Como se discute en la Sección 3.3.4.4 la sobreexpresión de genes vinculados a la biosíntesis de mucinas resulta interesante, en tanto en los estadios de cercaria se comienzan a secretar las glicoproteínas tegumentarias que en última instancia constituyen la capa interior del quiste de metacercarias de *F. hepatica*. Por último, el enriquecimiento en la vía “*Relaxin signaling pathway*” resulta también complejo, siendo probable que se deba nuevamente a roles de genes sobreexpresados en otros organismos.

3.3.2 GDEs sobreexpresados en miracidios

Como se describió en la introducción de este trabajo (ver Sección 1.5) el miracidio constituye un estadio larvario de vida libre, el cual infecta de forma activa al huésped intermediario caracol para dar lugar a los estadios de desarrollo intra-caracol. La capacidad motora y el funcionamiento de un metabolismo que sustenten energéticamente la misma son por tanto esenciales en este estadio. Otro aspecto esencial es la puesta en marcha de mecanismos que faciliten la infección activa del huésped caracol.

3.3.2.1 Proteínas ciliares y flagelares

Una de las diferencias morfológicas más notables entre miracidios y esporocistos es la presencia ubicua de una cubierta de células ciliadas, que proporciona la capacidad natatoria que deben desplegar en busca de un huésped caracol. Las células ciliadas se pierden durante el proceso de invasión a huésped caracol, en lo que se ha postulado constituye el comienzo del proceso de transformación de miracidio a esporocisto [12]. Las ciliias móviles son organelos altamente conservados en eucariotas, con una estructura arquetípica (“9+2”) de nueve pares de microtúbulos exteriores dispuestos en pares y un par central que brinda movilidad [122], estructuración que ha sido constatada en las ciliias de *S. mansoni* por microscopía electrónica [123].

Coherente con las diferencias morfológicas antes mencionadas, GDEs codificantes para proteínas ciliares se encuentran sobreexpresados en miracidios, como lo son un gen codificante para una α -tubulina (TUBA) (que podría estar vinculado a estructuras del citoesqueleto) y para CROCC (también conocida como “rootletin”), proteína que brinda apoyo estructural a ciliias móviles de *Mus musculus* y *Drosophila melanogaster* al localizarse en la raíz cercana a los cuerpos basales de estas [124,125].

Otros GDEs que codifican para proteínas pertenecientes a componentes estructurales y regulatorios de la función ciliar son *ccdc39*, *rps4* y *gas8*. Se ha demostrado que FAP59 (homólogo de CCDC39) y FAP172 poseen un rol en el ensamblado de las unidades regulares de 96nm que componen el axonema de *Chlamydomonas reinhardtii* (organismo modelo para el estudio de ciliias eucariotas). Este rol de “regla molecular” tendría como base las interacciones que poseen FAP59/FAP172 con otros componentes del axonema, reclutándolos durante la ciliogénesis [126]. RPS4 y GAS8 son proteínas pertenecientes a estas estructuras. La primera de estas es parte de las proteínas que componen el radio de rayos (“radial spoke”, en inglés), estructura que une al par central de microtúbulos con los pares exteriores. GAS8 en cambio es parte del *Nexin-Dynein Regulatory Complex* (N-DRC), el cual conecta pares adyacentes de microtúbulos exteriores. Ambas estructuras poseen roles regulatorios, al permitir el transporte de señales desde los pares centrales a los microtúbulos exteriores, permitiendo la activación alternada de dineínas que resulta en el movimiento ciliar [122]. El homólogo de GAS8 en *C. reinhardtii* (DRC4) se encuentra ausente en las ciliias de especímenes *knockout* para FAP59, coherente con el rol que posee este último en el reclutamiento del primero [126].

Otros genes identificados (i.e. SPA17 y ODF3) codifican para proteínas asociadas a la estructura flagelar de esperma de metazoarios (i.e. SPA17 y ODF3). Si bien estas proteínas han sido fundamentalmente estudiadas en estructuras flagelares, están presentes también en ciliias móviles de metazoarios como *M. musculus* en el caso de SPA17, o del anélido *Platynereis dumerilii* para ODF3, donde ha sido vinculada a la ciliogénesis [127]. La expresión de estos genes podría por tanto indicar un rol estructural en ciliias primarias o móviles de miracidios.

Recientemente se ha reportado la sobreexpresión de genes codificantes para proteínas ciliares (i.e. proteínas CFAP) en miracidios de *S. japonicum* [15], dejando en evidencia de que esta podría ser una característica de los miracidios en general. El fenómeno de sobreexpresión de estos genes en miracidios es consistente con una mayor presencia de estas estructuras en los miracidios y la necesaria regulación del movimiento ciliar en el miracidios, en respuesta a señales ambientales y/o endógenas que faciliten el movimiento necesario para la localización del próximo huésped.

3.3.2.2 Transducción celular mediada por calcio

Cabe destacar como GDEs a genes codificantes para proteínas que median procesos de señalización intracelular en respuesta a cambios en la concentración de Ca^{2+} , tales como la subunidad reguladora de la quinasa PKA (PKA-R), la subunidad catalítica de la fosfodiesterasa PDE1 (PDE1c), una proteína tipo calmodulina (CaM) y una quinasa dependiente de este mediador, CaMKII.

El calcio constituye un mediador universal de la señalización celular involucrado en procesos celulares variados como la regulación del crecimiento y el ciclo celular, la apoptosis, la contracción muscular y la movilidad. Resulta por tanto difícil restringir el rol funcional de estos en miracidios; a pesar de esto algunos estudios experimentales dilucidan parte del rol de estas proteínas en miracidios y esporocistos de tremátodos.

Los trabajos de Matsuyama *et al.* [128] muestran que el movimiento ciliar estaría mediado por señalización de Ca^{2+} mediada por cAMP en *S. mansoni*, pudiéndose estimular o inhibir utilizando activadores de proteínas participantes en esta vía de señalización tales como la adenilato ciclasa, PDE1 y PKA. Este modelo de regulación del movimiento ciliar es coherente con el comportamiento motor diferencial frente a soluciones de diferentes ratios $[\text{Ca}^{2+}]/[\text{Mg}^{2+}]$ [129], la inhibición reversible de la eclosión de huevos y nado de miracidios en medios de alta salinidad [Samuelson1984] y la dependencia del movimiento ciliar no sólo de las condiciones de osmolaridad sino también de la concentración de ATP [123]. Consistentemente, la exposición de miracidios de *F. hepatica* a agentes quelantes como el EDTA resultan en su inmovilización.

Es de esperar que la señalización mediada por Ca^{2+} y cAMP posea un rol homólogo en la regulación del movimiento de miracidios de *F. hepatica*, en tanto se ha reportado que estos agentes se ven implicados en la contracción muscular de las miofibrillas de este organismo [10]. Modelos similares han sido reportados para platelmintos de otras clases, como el cestode *Hymenolepis microstoma* [130].

Además de PKA y PDE1, otras proteínas vinculadas a la señalización mediada por Ca^{2+} están posiblemente implicadas en la regulación del movimiento ciliar de tremátodos. A modo de ejemplo, la eclosión de huevos de *S. mansoni* no es solo dependiente de la concentración de Ca^{2+} , sino que la presencia de antagonistas de calmodulina (CaM) inhibe este proceso [131–133]. Resulta por tanto interesante que al menos un GDE sobreexpresado en miracidios codifique para esta proteína. La calmodulina es un mediador de la acción celular de Ca^{2+} en procesos tan variados como la transcripción, la contracción muscular, la reorganización del citoesqueleto o la muerte celular. Genes codificantes para CaM o proteínas tipo CaM han sido reportados y caracterizados en *S. mansoni* y *F. hepatica* [134–136]. Lo que es más, mediante el uso de anticuerpos monoclonales se ha localizado a CaM en cilias, placas epidérmicas y papilas sensoriales no ciliadas del miracidio de *S. mansoni* [136], así como en las células vitelinas de huevos y en el parénquima de juveniles de *F. hepatica* [135,137]. McCammik *et al.* demostraron que la supresión de la expresión génica de todas las copias detectadas codificantes para calmodulina por técnica de RNAi resulta en la aparición de juveniles con fenotipos de hipermovilidad y crecimiento disminuido, recapitulándose dichos fenotipos con antagonistas de CaM [137].

La expresión diferencial de genes reguladores de la contracción muscular y movilidad ciliar en el miracidio de *S. mansoni* sugiere que procesos homólogos pueden mediar estas funciones también en el miracidio de *F. hepatica*. Este estadio de infección activa se caracteriza por su profuso movimiento, el cual permite eventualmente la infección del huésped intermedio; es de esperar por tanto que algunos genes vinculados a movilidad se encuentren expresados diferencialmente respecto a esporocistos, un estadio con movilidad un poco más restringida.

A pesar de esto, es claro que la regulación mediada por calcio puede estar involucrada también en otros procesos. A modo de ejemplo, se ha reportado que el proceso de transformación de miracidios a esporocistos es dependiente de la concentración de este ión en *S. mansoni* [138]. Es posible que en *F. hepatica* también se utilicen cambios en la concentración de Ca^{2+} como señal ambiental que regule procesos de desarrollo en las primeras etapas de invasión a huésped caracol (*e.g.* vía CaMKII por activación de vía Wnt/ Ca^{2+} , ver Figura 3.16 y mención más adelante en Sección 3.3.3.1 a la vía Wnt/ Ca^{2+}).

El estudio de los genes involucrados en vías de señalización mediados por calcio que poseen expresión en estadios larvarios de *F. hepatica* muestra que estos sistemas podrían estar involucrados en la señalización no solo de la contracción muscular, sino también en procesos como la proliferación celular, procesos metabólicos o el funcionamiento mitocondrial. Estudios experimentales son necesarios para dilucidar los procesos en los que participa este sistema de señalización en miracidios de *F. hepatica*.

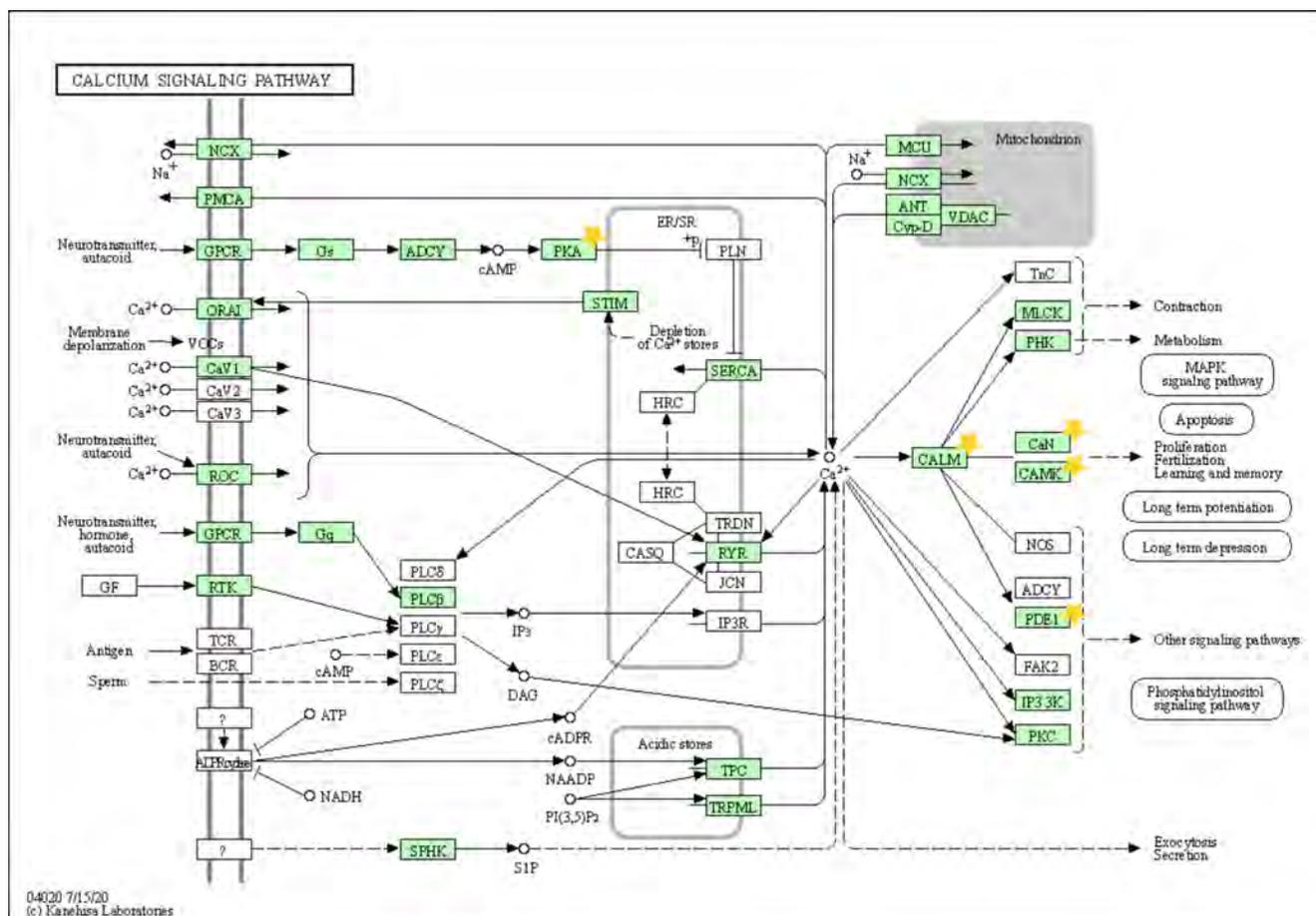


Figura 3.16: Vías de señalización mediada por calcio de la base de datos KEGG (Calcium signaling pathway; código map04020). Se anota en verde aquellos genes para los que se detecta expresión en algún estadio larvario asociado al huésped caracol (i.e. expresión ≥ 1 CPM en todas las muestras). Esquema construido con la opción Reconstruct de la herramienta KEGG mapper (<https://www.genome.jp/kegg/mapper/reconstruct.html>)

3.3.2.3 Neurotransmisores

La posible relevancia de expresión diferencial de señalizadores del control motor de miracidios se ve a su vez sugerida en la expresión diferencial de genes codificantes para enzimas vinculadas a la biosíntesis de los neurotransmisores serotonina y dopamina, tales como triptófano-hidroxilasa (TPH1) y tirosin hidroxilasa (TH). Se observa también expresión diferencial para GTP ciclohidrolasa I (GCTH1) necesaria para la síntesis de tetrahidrobiopterina, cofactor necesario para la biosíntesis de ambos neurotransmisores. La otra enzima que participa de estas vías (DAAC) se expresa en miracidios, aunque no aparece como sobreexpresada, abriendo así la posibilidad de que este neurotransmisor se esté sintetizando de forma endógena en miracidios.

La capacidad que posee la serotonina como estimulante de la contracción muscular en adultos de *F. hepatica* fue demostrada tempranamente [139]. Esta capacidad de contracción explicaría la quimiotaxis hacia la serotonina de miracidios de *S. mansoni*, al aumentar su tasa de giro por contracción muscular [138]. Estudios más recientes en *S. mansoni*, muestran que la serotonina estimula de forma selectiva a un receptor GPCR (Sm5HTR) y estimula la producción de cAMP. Por técnicas inmunocitológicas demuestran la expresión abundante de este receptor y de células serotoninérgicas en el sistema nervioso de schistosomas y adultos de *S. mansoni*, existiendo una pérdida significativa de la capacidad motora en estos estadios cuando se impide la expresión del receptor por técnica de RNAi [140]. La síntesis de serotonina endógena podría por tanto también formar parte de los mecanismos reguladores del movimiento de miracidios de *F. hepatica*.

La serotonina parecería jugar un rol conservado en la contracción muscular de tremátodos, como sugiere su presencia en inervaciones sensoriales de varias especies de esta clase. Estudios de histoquímica muestran la presencia de este neurotransmisor en las enervaciones sensoriales de *S. mansoni* [141] y *O. felineus* [142], así como en las ventosas orales y ventrales de adultos de *F. hepatica* [143], adultos de *S. mansoni* [140] y metacercarias de *O. felineus* [144]. Un estudio reciente extiende dicha presencia a adultos, cercarias y metacercarias de especies de amplio rango taxonómico de tremátodos por técnicas de inmunocitoquímica y microscopía láser confocal [145].

Algunos de los órganos descritos en el miracidio de *F. hepatica* podrían ser blanco de regulación neuromuscular por vía de estos neurotransmisores. Además de las ciliias (cuyo movimiento ya ha sido discutido), los miracidios poseen una papila apical (a veces denominada *terebratorium*) con reminiscencias a estructuras succionadoras que participa en la adhesión e infección del caracol tanto en *F. hepatica* [11,12], como en *F. gigantica* [146] y *S. mansoni* [147]. El proceso es fuertemente dependiente de la contracción muscular [2]. No existen a la fecha reportes de enervaciones neuronales en estas estructuras del estadio, pero el hecho de que genes implicados en la biosíntesis de serotonina posean expresión diferencial en este estadio sugiere a ésta como posible regulador de la contracción muscular en este estadio de *F. hepatica*.

Por último, resulta interesante destacar que en sus trabajos Mansour demostró también que la captación de glucosa, la degradación de glucógeno y la glucólisis es estimulada en extractos de tejido de *F. hepatica* preincubados con serotonina [148,149]. Se postula que esta acción es mediada por el cAMP como mensajero secundario. La capacidad de la serotonina en regular la fosforilación (y consecuente activación) de la fosfofructoquinasa en adultos de *F. hepatica* fue demostrada *in vivo* por Kamemoto et al. [150], en coherencia con lo postulado por Mansour.

Nuevamente, este hecho remarca que un mismo conjunto de moléculas puede participar de vías de señalización variados y solapantes; factores como el tejido donde se expresan o la temporalidad probablemente influyen sobre qué vía actúan en cada caso.

3.3.2.4 Metabolismo

Otros de los GDEs relacionados a procesos metabólicos y de transporte de nutrientes codifican para las proteínas 5'-nucleotidasa (NT5C3B) y glicerol-3-deshidrogenasa (GPDH1), cuyo rol en estadios larvarios de *F. hepatica* y *S. mansoni* ha sido discutido en la literatura.

Los miracidios de *F. hepatica* dependen de forma casi exclusiva de la obtención de energía a partir del catabolismo de sus reservas de glucógeno [18], por lo cual la existencia de transportadores de glucosa en tejidos de alta actividad podría resultar esencial en este estadio. La expresión de la proteína NT5C3B ha sido constatada en tejidos de miracidios [151] y esporocistos [152] de *F. hepatica* por medio de detección de actividad enzimática en tejidos parasitarios. En miracidios se atribuye a esta proteína un posible rol en el transporte de glúcidos.

Los patrones de expresión reportados permiten, no obstante, hipotetizar también sobre posibles roles vinculados a las altas demandas energéticas que suceden durante y luego de la transición miracidio-esporocisto. NT5C3B se encuentra fundamentalmente expresada en el arquenterón, protonefridas y ganglio nervioso de los miracidios [151], mientras que las reservas de glucógeno movilizadas durante el movimiento del miracidio se alojan en las placas epiteliales [2]. En esporocistos la expresión de NT5C3B ocurre fundamentalmente en membranas celulares y en el citosol de células tegumentales, parenquimales y masas germinales; por lo tanto, es plausible que en este estadio supla de nutrientes a las masas germinales demandantes de energía para desarrollar su proceso de replicación. Además de un rol en el transporte de glucosa, se ha postulado un rol en el transporte de nucleótidos a las masas germinales en este estadio [152]. Resulta por tanto interesante considerar la hipótesis de que la sobreexpresión de mRNA codificante para NT5C3B en miracidios constituya una adaptación que asegure la rápida traducción de la proteína inmediatamente después de iniciada la transformación a esporocisto, en tanto la reproducción asexual requiere de la obtención de purinas del huésped.

Respecto a la expresión diferencial de GPDH, al menos dos hipótesis pueden ser consideradas. En primer lugar, una de las isozimas de GPDH participa en el transporte de electrones del citosol a la matriz mitocondrial durante la cadena respiratoria (la denominada “*lanzadera del glicerol*” [153]). Por lo tanto, la expresión diferencial de GPDH podría reflejar una mayor necesidad de síntesis de proteínas vinculadas a la cadena respiratoria por las exigencias metabólicas de los miracidios, o la presencia de un mayor número de mitocondrias. Alternativamente, se ha sugerido que esta expresión podría estar vinculada a posibles funciones desempeñadas por esporocistos en las fases tempranas de la infección a caracol (y por lo tanto requiriendo síntesis previa a entrada al huésped). Esto se vincula al rol de GPDH en procesos catabólicos (posibilidad de metabolizar glicerol por vía de fermentación) como biosintéticos (utilización de intermediarios en la síntesis de triacilgliceroles). El contenido de triacilgliceroles ha sido reportado invariante en miracidios de *F. hepatica* [18] y *S. mansoni* [19], lo cual sugiere que las funciones metabólicas de GPDH podrían ser utilizadas en esporocistos. Esta hipótesis ha sido postulada por Khayath et al., que confirmaron por qRT-PCR la expresión diferencial del gen en miracidios y esporocistos de *S. mansoni* respecto a otros estadios, constatándose una mayor expresión en miracidios [154]. Nuestros resultados sugieren que un escenario homólogo podría existir en el caso de miracidios y esporocistos de *F. hepatica*.

Cabe recordar que todos los estadios larvarios de *F. hepatica* muestran expresión alta de genes vinculados a la obtención de energía, posiblemente debido a las necesidades metabólicas elevadas que los caracterizan. Los GDEs detectados pueden por tanto estar reflejando aspectos regulatorios sutiles vinculados a diferencias de la fisiología de los estadios estudiados, como se presentó en el caso de NT5C3B y GPDH.

3.3.2.5 Respuesta a estrés oxidativo

Se constata la expresión diferencial en miracidios de genes codificantes para una tioredoxina (Trx) y una peroxirredoxina (TrPx). Ambas enzimas son parte del sistema molecular que poseen platelmintos para

mantener la homeostasis redox. Su capacidad reductora es empleada por estos organismos en procesos variados, como el estado redox de enzimas durante su ciclo catalítico o la de detoxificación de xenobióticos al reducir compuestos hidrofóbicos, aumentando su solubilidad [155].

De particular interés es la capacidad de reducción de especies reactivas de oxígeno (“ROS”, en inglés) que poseen estas enzimas. Se podrían hipotetizar dos posibles funciones asociadas a una sobreexpresión de genes codificantes para proteínas de este sistema en miracidios. En primer lugar, se ha sugerido que este sistema estaría relacionado al control endógeno de ROS generados durante la intensa actividad metabólica aerobia [156].

Por otro lado, como se detalló en la Sección 1.5.3, la producción de ROS y RNS constituye un mecanismo de respuesta molecular fundamental por parte de los hemocitos del caracol: se ha mostrado que el encapsulamiento y la eliminación de esporocistos mediada por estas especies químicas en un plazo de 24-48hs constituyen la principal respuesta inmune temprana en especímenes de *B. glabrata* resistentes a *S. mansoni* [26,157,158].

En este contexto se podría hipotetizar sobre el valor adaptativo de la síntesis de transcriptos codificantes para proteínas con capacidad de detoxificar especies reactivas de oxígeno y/o nitrógeno previo al ingreso a huésped caracol, la cual podría aumentar las probabilidades del establecimiento de una infección persistente. Ya se ha postulado que proteínas de esta familia podrían participar de la modulación que ejerce *F. hepatica* sobre la respuesta inmune innata de huésped mamífero [159], por lo que resulta razonable considerar funciones similares en el caso de huésped caracol. En *S. mansoni* se ha mostrado por técnica de inmunohistoquímica que existe expresión de copias de TrPx en la papila apical de miracidios [160]. A su vez, existe un aumento de la expresión de estos genes inducible por H_2O_2 con expresión en tegumento [160], hechos coherentes con la hipótesis de que podrían jugar un rol en la interfaz huésped-parásito. El hecho de que en el análisis realizado en esta tesis se observe sobreexpresión de TrPx en miracidios y no en esporocistos de *F. hepatica* podría ser atribuible a la existencia de diferentes dinámicas de interacción entre el sistema inmune de huésped y los parásitos al comienzo de la infección (como en los estudios citados, que se limitan a las primeras 9h de transformación *in vitro* de esporocistos) y en infecciones ya establecidas (15 días post-infección, en nuestro estudio).

La expresión de otros genes en miracidios también podría tratarse de una adaptación que permita una respuesta rápida a ROS. Se ha reportado que un aumento de expresión de GPDH podría jugar este rol en el céstodo *E. granulosus*, en tanto la producción de glicerol y su expresión ayudarían en la osmoregulación frente al aumento de permeabilidad que produce un ambiente rico en H_2O_2 [161]; un rol similar podría existir en el caso de *F. hepatica*.

La observación de GDEs en miracidios podría por tanto tratarse de una adaptación con función a las primeras etapas de la infección a huésped caracol.

Es importante señalar que esta copia TrPx posee una expresión más bien baja en miracidios (i.e. menos de 20 CPM), coherente con una expresión restringida pero sugiriendo que no se trata de un producto de expresión preponderante. Este hecho remarca la necesidad de llevar a cabo investigación experimental que brinde información sobre su función en miracidios y etapa temprana de desarrollo de esporocistos.

3.3.2.6 Expresión larvaria de miembros de la familia SCP/TAPS

Miembros de las proteínas de la superfamilia SCP/TAPS (también conocida como CAP) han sido descritas en varios parásitos helmintos [162] y en especies de todas las clases de platelmintos [163]. El estudio en estos últimos se ha centrado principalmente en miembros de la familia VAL (“Venom Allergen Like”, en inglés). En el caso de tremátodos, la familia SmVAL de *S. mansoni* es la que más estudios presenta en la literatura. Se ha reportado regulación temporal de la transcripción de sus miembros [164,165] así

como regulación espacial, con miembros expresados de forma específica en tegumento, esófago, ventosas ventrales y orales o glándulas acetabulares de cercarias [165–168].

Un estudio inicial en varias familias de tremátodos identificó miembros de la familia VAL en fasciólidos [169], aunque el mismo se vió acotado al estudio de expresión en estadíos intramamífero debido a la escasez de información transcriptómica. Debido a la falta de información experimental, tampoco son formuladas hipótesis funcionales en dicho estudio. En *F. hepatica* se ha detectado expresión de genes codificantes para proteínas VAL en juveniles en etapa infectiva [170], y mediante el uso de *arrays* de lectinas ha podido establecerse la localización tegumental de *FhVAL11* en el tegumento parasitario [171].

La integración de nuevos datos transcriptómicos muestra que existen copias codificantes para VALs con tendencia a la expresión predominante en estadíos intra-caracol tanto en *F. hepatica* como en *F. gigantica*, principalmente en miracidios (Figura 3.17).

El árbol filogenético inferido para la familia muestra un conjunto de amplificaciones génicas independientes tanto en tremátodos (familias Schistosomatidae y Fasciolidae) como en platelmintos de vida libre como *S. mediterranea*. Estas amplificaciones podrían sugerir un valor adaptativo para la sobrevivencia de los organismos portadores en los diferentes linajes, en tanto el patrón observado puede surgir por un aumento de la retención de copias obtenidas por un proceso de duplicación génica neutro. Cabe destacar que amplificaciones génicas de familias de SCP/TAPS ya han sido observadas en otros platelmintos parásitos como el céstode *M. corti* [172]

El patrón de agrupamientos observado sugiere la presencia de al menos cuatro secuencias parálogas en el ancestro común a las especies pertenecientes al suborden Echinostomata (*E. caproni*, *F. hepatica*, *F. gigantica* y *F. buski*). Dos de estos clados están compuestos por secuencias de copia única en las especies donde están presentes, con una alta concordancia en los patrones de expresión observados en *F. hepatica* y *F. gigantica* (3 y 4 de VALs de Echinostomata, en la Figura 3.17). Estos grupos de ortólogos tienden a expresar en miracidios y adultos, respectivamente. Algo similar sucede con el Grupo 2 de VALs de echinostomátidos, compuesto por dos secuencias codificadas en el genoma de *F. gigantica* y una en el de *F. hepatica*. Estas secuencias tienden a expresar en estadíos intra-caracol tardío o en metacercarias. En contraste, el clado restante de secuencias de echinostomátidos (Grupo 1) es una expansión génica que posee patrones de expresión más variados (tendencia a la expresión en miracidios, estadíos intra-caracol y adultos). Se podría hipotetizar que el aumento del número de copias durante el proceso de duplicación que dió lugar a este clado tenga vínculo con la mayor diversidad observada en los patrones de expresión. En este escenario hipotético el aumento del número de copias genera una redundancia funcional que relaja las presiones selectivas sobre las copias génicas, habilitando el proceso de divergencia funcional.

Si bien no se posee en la actualidad información funcional para estos genes en *F. hepatica*, si existen algunos estudios publicados referentes a sus homólogos codificados en el genoma de *S. mansoni*, las cuales integran el Grupo 1 de la familia multigénica SmVAL. En base a sus patrones de expresión y reactividad inmunológica cruzada, este subclado se puede dividir a su vez en los subgrupos A-D [165]; integran el grupo de homólogos estudiados aquí secuencias de los Grupos 1A (*sm-val5*, 15, 27 y 28) y B (*sm-val1*, 2, 14 y 21). Aunque todavía escasa, la creciente información experimental muestra que los miembros del grupo 1 de SmVALs poseen relevancia funcional en los estadíos intra-caracol de *S. mansoni*. La proteína SmVAL9 (miembro del grupo 1A) ha sido identificada en estudios de proteómica como parte de los factores secretados durante la transición miracidio-esporocisto en *S. mansoni* [173], y se ha caracterizado a su expresión localizada en poblaciones parenquimales, tegumentarias y masas germinales de miracidios y esporocistos [174]. La proteína es capaz de inducir una sobreexpresión de factores de remodelado de la matriz extracelular tanto en líneas celulares *Bge* de *B. glabrata* como en macrófagos derivados de médula ósea de *M. musculus*, por lo cual Yoshino et al sugieren que podría favorecer el proceso de infección por parte de miracidios y la migración de esporocistos, pudiendo operar también en la infección temprana a mamífero [174]. Respecto al grupo 1B, es interesante destacar que *SmVAL22* ha sido reportada en poblaciones proliferantes de esporocistos *S. mansoni* [14,40], aunque al igual que lo que sucede con otros

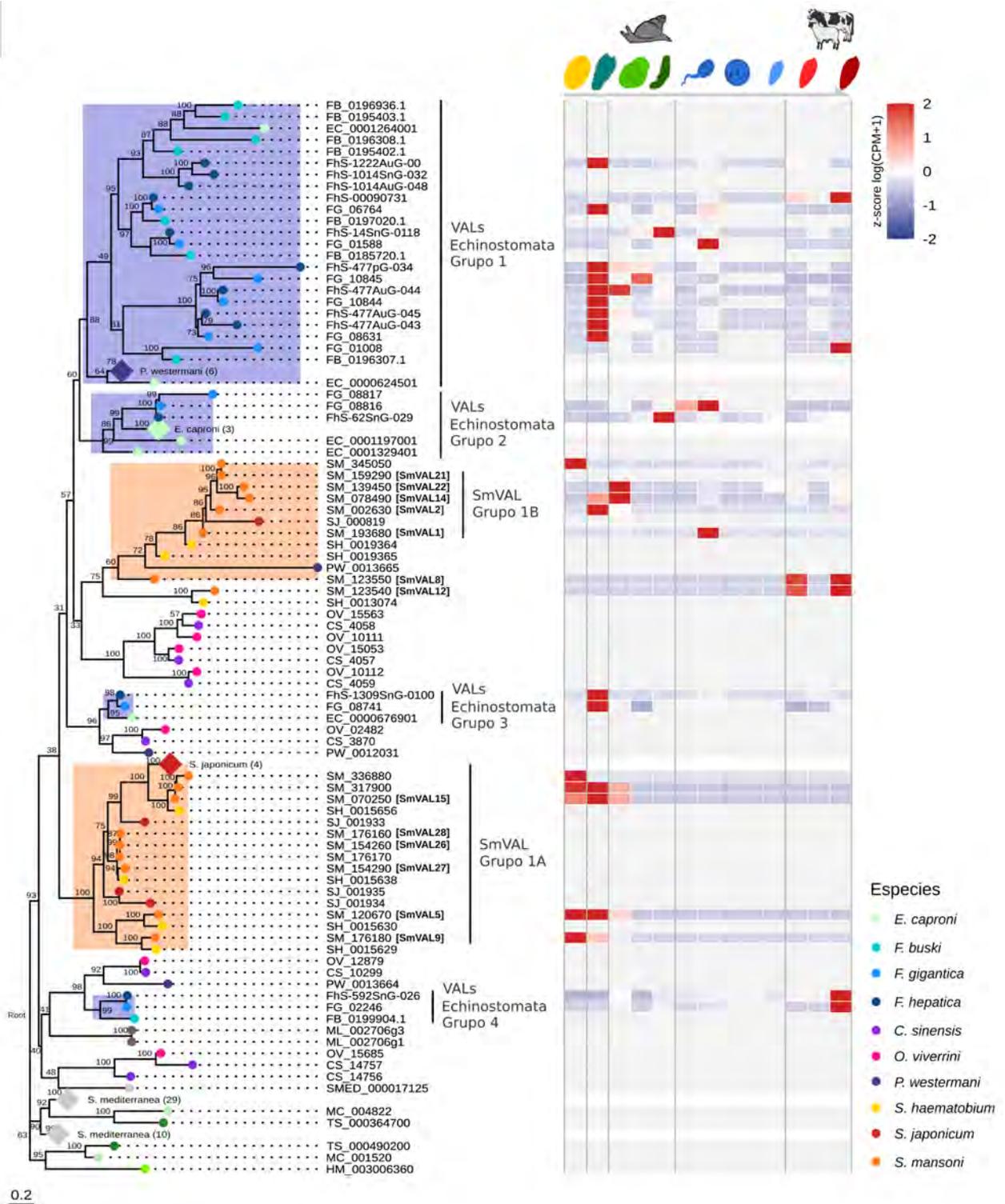


Figura 3.17: Árbol filogenético para miembros del grupo de homólogos OG0000035, con miembros homólogos a proteínas de tremátodos de la familia SCP/TAPS. Se muestran las tendencias de expresión (z-score) de cada gen para los miembros codificados en los genomas de *F. hepatica* y *F. gigantica*. El árbol filogenético fue inferido por método de máxima verosimilitud empleando el programa IQTree v2 sobre alineamientos de proteínas, realizando la selección del modelo evolutivo a emplear con el programa ModelTest y calculando el soporte de rama para clados mediante método de bootstrap bajo la modalidad de ultrafast-bootstrap. Arbol enraizado tomando como grupo externo un clado compuesto exclusivamente por secuencias codificadas en genomas de cestodos y turbelarios.

miembros de la familia SmVAL expresados diferencialmente en miracidios de *S. mansoni*, a la fecha no existe información funcional disponible [165]. Cabe destacar que la posible relevancia funcional de los integrantes del Grupo 1 de la familia SmVAL en estadios asociados a caracol excede a los grupos 1A y B. A modo de ejemplo, SmVAL4 y 18 (grupo 1C) poseerían capacidad de unión a lípidos y activación de plasminógeno, respectivamente [167,175]. Este último caso revela que los patrones de expresión en estadios larvarios pueden estar asociados incluso a la infección del siguiente hospedero: SmVAL18 se expresa en masas germinales de schistosomulas, localizándose luego en las glándulas acetabulares de cercarias y siendo secretada en las primeras horas del desarrollo de schistosomulas en el huésped mamífero, donde contribuye a la activación de plasminógeno. Resulta controversial plantear una función homóloga en el caso de huéspedes caracoles, en tanto estos poseen un sistema coagulatorio diferente al de mamíferos y cuyos mecanismos no han sido casi estudiados en gastrópodos⁴. Un escenario posible es que las proteínas codificadas en estos transcritos participen de cascadas líticas, participando de mecanismos regulatorios como los que activan al plasminógeno. Alternativamente, cabe la posibilidad de que de forma similar la expresión de miembros de esta familia en miracidios y esporocistos corresponda a funciones ejercidas en estadios de desarrollo posteriores.

Por todo lo expresado anteriormente la expresión diferencial de miembros de la familia SCP/TAPS en miracidios y estadios intra-caracol de *F. hepatica* resulta de interés en tanto puede poseer un rol funcional en la interfaz huésped-parásito y/o en procesos regulatorios del desarrollo intra-caracol, constituyendo candidatos interesantes para estudios experimentales futuros que permitan su caracterización.

3.3.2.7 Proteasas vinculadas al proceso de invasión del huésped caracol

Los miracidios de *F. hepatica* ingresan a sus huéspedes caracoles a través de un proceso de invasión activa, el cual se ha descrito requeriría de la secreción de proteasas a través de la papila apical para ser exitoso [11,12]. Recientemente Kianifard et al mostraron la actividad serin proteasas en un lisado obtenido de miracidios de *F. hepatica*, hipotetizando sobre el posible rol de estas proteínas en el proceso de invasión a caracol [20]. Coherente con dicho resultado, uno de los GDEs sobreexpresados en miracidios codifica para una proteína anotada como posible peptidasa de la familia S1A de serin proteasas. El mismo es a su vez un gen de alta expresión detectado de forma específica en estadios intra-caracol (ver Sección 3.1.3). Integrantes de la familia de serin proteasas S1A ya han sido descritas como relevantes en el contexto de invasión a huésped intermediario y definitivo por parte de tremátodos. El ejemplo arquetípico es una familia amplificada de proteínas con actividad elastasa, que son preponderantes en las secreciones de la cercaria de *S. mansoni* durante el proceso de invasión al huésped mamífero [176]. Algunos miembros de esta familia han sido postulados como activos y relevantes durante el proceso de salida del huésped caracol por parte de cercarias [177], demostrando la capacidad que poseen estas serin proteasas de degradar tejido del huésped caracol *B. glabrata*. En dicho estudio no se detectó, no obstante, expresión de miembros de la familia en miracidios, por lo que el proceso de invasión de huésped caracol ocurre presuntamente gracias a la expresión de otras proteasas (al menos en *S. mansoni*).

La serin proteasa diferencialmente expresada en miracidios de *F. hepatica* no es homóloga a las elastasas de cercaria (datos no mostrados), y representa otra familia de serin proteasas tipo S1A que parece ser específica de los estadios larvarios asociados a caracol. Resulta interesante que en *Gyrodactylus salaris*, un Monogenea ectoparásito de peces, también aparece amplificada una familia de serin-proteasas [178]. La presencia de este tipo de proteasas suele asociarse así a procesos de invasión que implican el tegumento externo del huésped (piel en el caso de las elastasas de cercarias, la piel o branquias de peces, el tegumento del pie o manto del caracol). A partir de nuestros datos y estudios transcriptómicos similares en otros platelmintos es posible estudiar más en detalle la contribución de las serin proteasas en los procesos de invasión al huésped y establecimiento de la infección.

⁴Ver Sección 3.3.3.3 donde se toca este punto en más detalle

3.3.3 GDEs sobreexpresados en estadíos intra-caracol 15 d.p.i.

El proceso de amplificación asexual que acontece durante el desarrollo intra-caracol de esporocistos y redias supone la sucesión de los procesos de replicación celular, embriogénesis y morfogénesis de estos estadíos. Como es de esperar, dicha secuencia supone un desafío metabólico, en tanto se requiere no solo de la obtención de energía para llevarla a cabo sino también de los nutrientes empleados en los procesos biosintéticos de las macromoléculas que constituyen cada organismo generado. Similar a lo observado en el caso de la infección de huéspedes mamíferos, durante la co-evolución con caracoles los tremátodos han adquirido mecanismos moleculares para parasitar estos recursos de sus huéspedes. Los procesos de desarrollo, la obtención de nutrientes y el escape exitoso a la respuesta inmune de caracoles son por tanto procesos clave para la sobrevivencia de estadíos intra-caracol de *F. hepatica*.

3.3.3.1 Genes vinculados a desarrollo

Como se describió en la *Introducción* a esta tesis (ver Sección 1.5.4, se ha detectado en platelmintos la presencia de poblaciones celulares totipotentes (llamadas según el caso células madre, germinales, o neoblastos). A partir de la diferenciación de las mismas se generan las líneas germinales y somáticas de organismos de este filo. En los estadíos larvarios asociadas a huésped caracol, la reproducción asexual se da inicialmente por proliferación de estas células, las cuales se diferencian durante el proceso de embriogénesis para generar los estadíos de desarrollo intra-caracol. Genes del programa de desarrollo GMP regulan la proliferación y la multipotencia de las células germinales de varios tremátodos, por lo cual resulta posible que algunos de los genes asociados a los términos funcionales KEGG anteriormente nombrados estén asociados, entre otros, a estos genes.

En líneas generales la mayoría de los marcadores de células germinales estudiados poseen su pico de expresión en miracidios o estadíos intra-caracol, acorde a lo esperado dada la presencia y replicación de estas células en estos estadíos de *F. hepatica* (Figuras 3.18 y 3.19).

Dentro de estos genes se encuentran marcadores de la replicación celular y sus procesos biosintéticos asociados (*pcna*, *cyclinB*, *ef-tu*, histonas *h2a* y *h2b*), reguladores de la proliferación de células madre (*egr*, *efgr*, *fgfrA*, *fgfrB*, *phb*), reguladores post-transcripcionales asociados a la multipotencia de células madre conservados en metazoarios (genes *vasa-like*, *ago2-1*) y genes codificantes para integrantes de complejos reguladores de la cromatina que aseguran la estabilidad de un estado indiferenciado de estas poblaciones celulares (*thoc*, *sz12*, *ezh*, *sedt8*, *chd*).

Este resultado muestra que las células germinales de estos estadíos de *F. hepatica* podrían estar reguladas en parte por genes conservados del programa GMP al igual que otros tremátodos. Acorde a lo esperado, las tendencias observadas son a su vez en general concordantes entre estadíos homólogos de *F. hepatica* y *F. gigantica*, indicando que los mecanismos regulatorios son conservados. Esto es esperable, en tanto como se ha dicho la conservación regulatoria de genes del programa GMP no sólo abarca a especies de tremátodos, céstodos y platelmintos de vida libre, si no a células pluripotentes de metazoarios en general. Resulta difícil atribuir una causa única a los casos en los que se observa discrepancia con las tendencias reportadas en la literatura. Esto podría tratarse de diferencias a nivel regulatorio, pero resulta evidente que al tratarse este de un perfil transcriptómico global es de esperar que los resultados observados sean altamente dependientes de las proporciones existentes entre las poblaciones celulares analizadas. Por lo tanto, variaciones sutiles de expresión de genes reguladores en células tipo-neoblasto o aquellos con picos de expresión muy acotados en el tiempo pueden haber visto afectada su capacidad de detección. Como escenario alternativo, células proliferantes de otros estadíos podrían estar expresando estos marcadores en mayor abundancia, incluso por haberse secuenciado un momento particular del estadío donde esto sucede. Argumentos análogos son empleables para las discrepancias observadas entre *F. hepatica* y *F. gigantica*. Estudios transcriptómicos a resolución celular deberán realizarse para dilucidar este tipo de

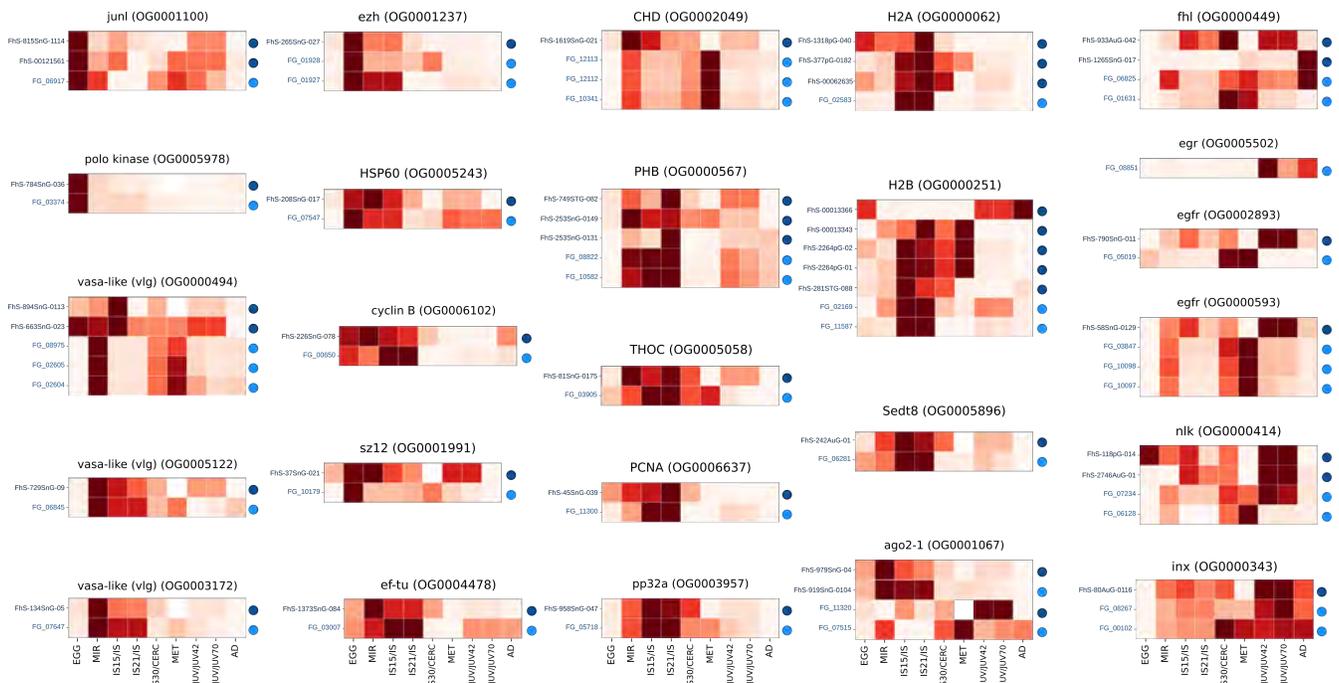


Figura 3.18: Tendencias de expresión para genes marcadores asociados a poblaciones celulares tipo-neoblasto estudiados en estadios larvarios de *S. mansoni*, publicados en Wang *et al.* (2013). Se reportan genes homólogos (i.e. integrantes de un mismo OrthoGroup) a los reportados por estos autores en su trabajo en las especies *F. hepatica* y *F. gigantica*. En cada caso se muestra la expresión normalizada para cada gen, donde la máxima expresión equivale a un valor de 1 y la mínima a un valor de 0. Los datos de expresión de *F. gigantica* corresponden a los publicados por Zhang *et al.* (2019).

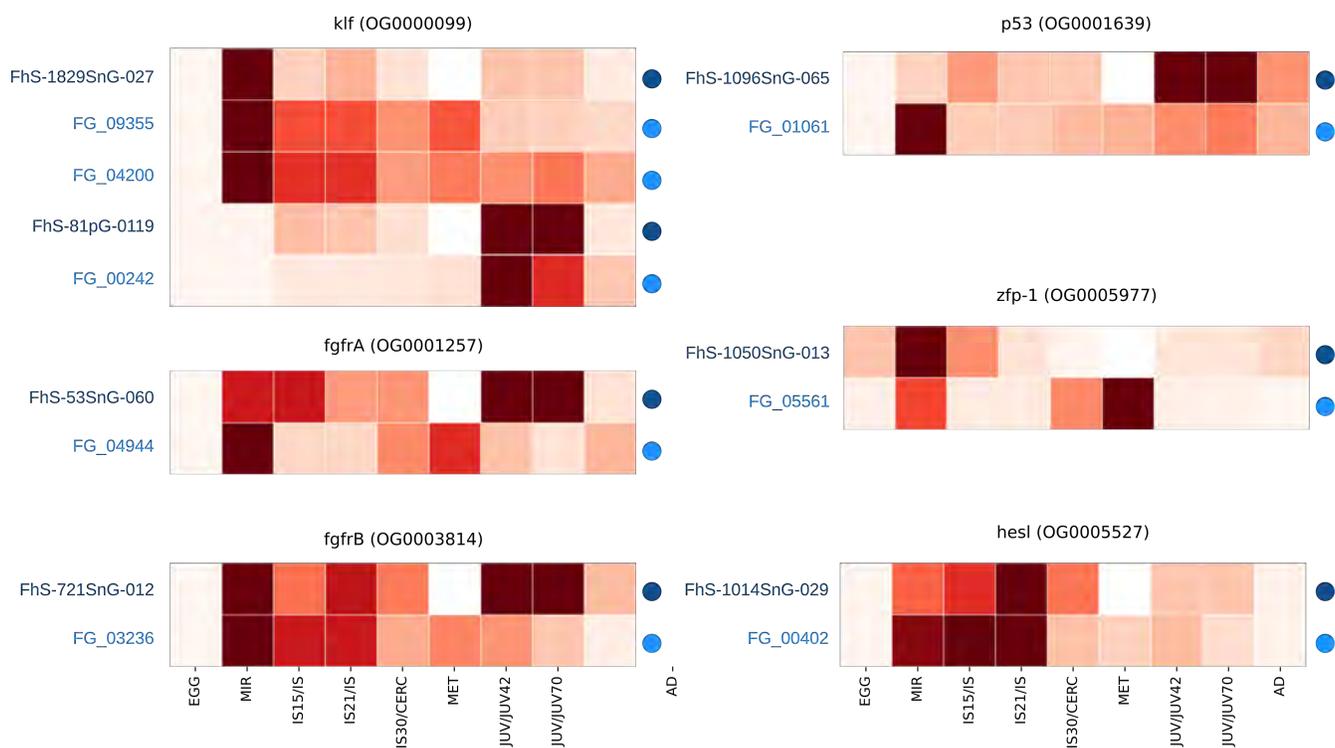


Figura 3.19: Tendencias de expresión para marcadores celulares estudiados por Wang et al (2018) a lo largo del ciclo de vida de *F. hepatica* y *F. gigantica*. Este conjunto de marcadores son empleados por los autores para distinguir a subpoblaciones de células germinales presentes a lo largo de vida de *S. mansoni* (ver Introducción de esta tesis).

cuestiones. Estudios morfológicos y de expresión han mostrado la presencia de poblaciones celulares tipo-neoblasto en estadios intramamífero de *F. hepatica* [44], y tendencias de expresión coherente para algunos de los genes del programa GMP (e.g. *ago2-1*, *nanos*) ya han sido reportados para estadios juveniles de *F. hepatica* [179]. No obstante, los resultados mostrados constituyen el primer indicio de que el programa de regulación GMP podría operar en las células germinales de *F. hepatica* durante su desarrollo intra-caracol, presuntamente de forma homóloga a lo observado para otros platelmintos.

Como se dijo anteriormente, durante el desarrollo intra-caracol de *F. hepatica* acontecen los procesos de embriogénesis y morfogénesis que permiten el aumento exponencial de individuos con capacidad infectiva del huésped definitivo. Los procesos de desarrollo de redias y cercarias implican claramente la necesidad de establecer polaridad en ejes corporales, por lo que no es sorprendente que dentro de los GDEs sobreexpresados en IS15 respecto a miracidios se encuentran varios genes codificantes para proteínas relevantes en procesos de desarrollo y morfogénesis (algunos resaltados en violeta en Figura 3.12 y marcados en la Figura 3.20). En particular, es notable la presencia de genes involucrados en las vías de señalización Wnt/ β -catenina y BMP, así como la presencia de genes pertenecientes a las familia multigénicas Hox y Fox.

El rol de las vías de señalización Wnt y BMP en el establecimiento de los ejes anteroposterior (vía Wnt/ β -catenina) y dorsoventral (vía BMP) se encuentra conservado en la mayoría de los bilaterios [180,181]. Genes codificantes para miembros de estas vías se encuentran conservados en platelmintos, siendo extensamente estudiados sus roles en el establecimiento de ejes durante el desarrollo y regeneración de tejidos en embriones y adultos de planaria (i.e. organismos modelo *S. mediterranea* y *S. polychroa*). Su papel en el desarrollo en especies parasitas comienza a investigarse en cestodos, y en mucho menor medida trematodos [182].

En la vía Wnt canónica ligandos de la familia Wnt (e.g. WNT1) se unen a receptores Frizzled y LRP5/6, causando el reclutamiento de la proteína Dishevelled e inhibiendo así al complejo de degradación de β -catenina (que incluye a Axin-1, GSK-3 y APC) (Figura 3.20). Al estabilizarse la expresión de β -catenina esta transloca al núcleo y, junto a factores de transcripción de la familia TCF/LEF, regula la expresión de factores de transcripción [180,181]. Se ha confirmado experimentalmente el rol de esta vía en el establecimiento del eje anteroposterior en *S. mediterranea*, en donde gradientes de expresión de antagonistas de las familias sFRP (homólogo de *Frizzled* secretado) y Notum a lo largo del eje jugarían también un rol en el control de la actividad transcripcional [182]. La expresión diferencial de los genes *wnt1*, *notum*, *fdz4*, *fdz5*, *sfrp2*, *axin1* y *tcf15* en IS15 podría ser por tanto indicio de la activación de la vía canónica de Wnt en estadios intra-caracol 15 d.p.i.. A su vez, genes pertenecientes a las familias involucradas en esta vía se ven expresados en estadios intra-caracol de *F. hepatica*, salvo en dos excepciones: GBP y LRP5/6. Si bien se ha descrito que la proteína GBP puede ser dispensable para la activación de la vía canónica de Wnt en mamíferos [183], el receptor LRP5/6 suele ser descrito como esencial para la activación de la vía [181] con confirmación experimental para el caso de *S. mediterranea* [184]. Una búsqueda de homólogos en trematodos para una copia de LRP5/6 de *S. mediterranea* (código NCBI DAD52840.1) no revela ninguna secuencia candidata. Por lo tanto, o bien existe algún homólogo altamente divergente, como se ha descrito para otros genes de la vía en cestodos (i.e. GSK-3 y axina [185]), o bien este gen descrito generalmente como esencial para la vía canónica de Wnt se ha perdido en los trematodos para los cuales se tiene información genómica a la fecha. Como se muestra en la Figura 3.20, algunos de los componentes de la vía Wnt/PCP tampoco parecen expresarse significativamente en los estadios estudiados, mientras que todos los componentes de la vía Wnt/ Ca^{2+} poseen algún homólogo expresado en estadio intra-caracol 15 d.p.i.. Estudios experimentales son necesarios para dilucidar los aportes relativos de cada una de las vías y sus funciones *in vivo*.

Roles paralelos a la activación de la vía canónica de Wnt han sido estudiados para miembros de la familia en *S. mediterranea*. Un ejemplo lo constituye el rol postulado para WntP-2 y los receptores PTK-7 y ndl-3 (de la familia FGFR) durante la regeneración (y posible establecimiento) de la identidad del tronco a lo

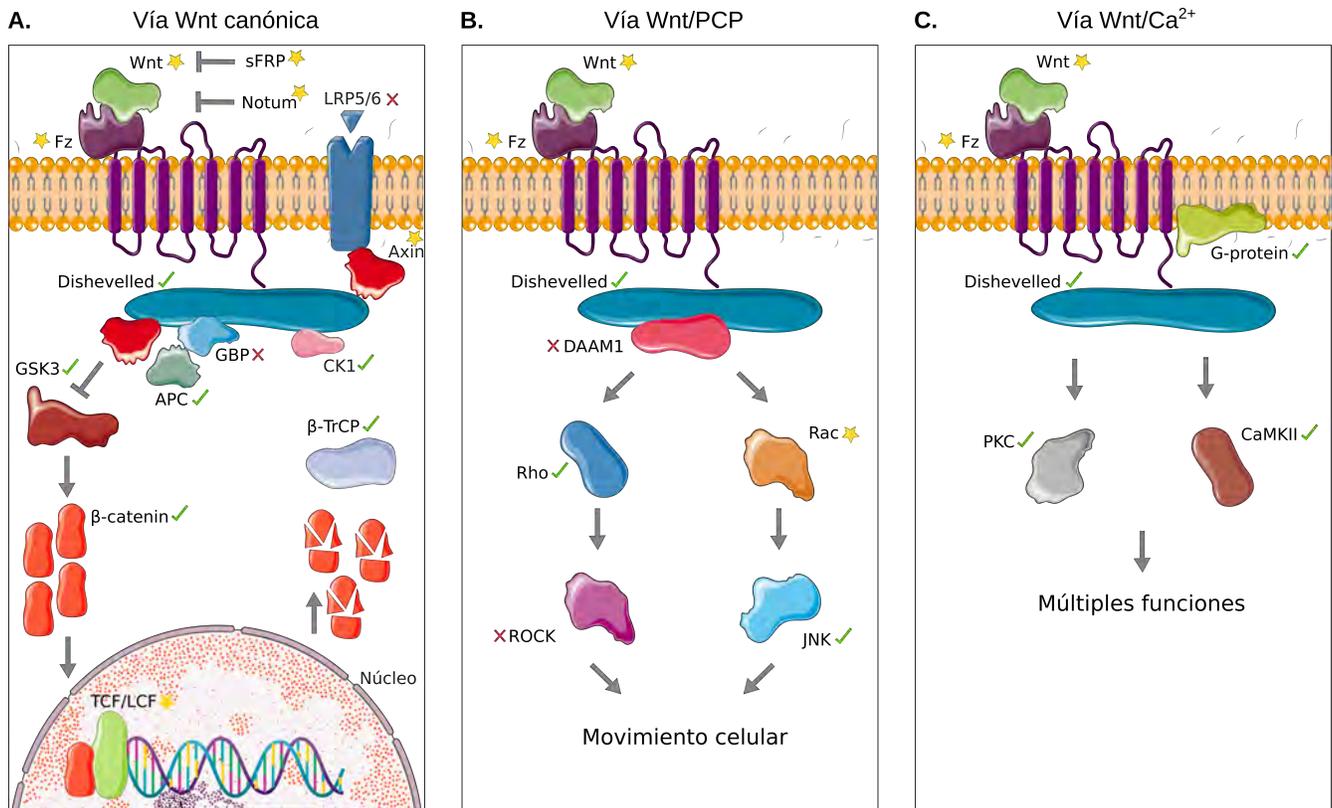


Figura 3.20: Componentes principales de las vías Wnt usualmente nombradas en la literatura. A. Vía Wnt canónica, también conocida como Wnt/beta-catenina. B. Vía Wnt/PCP (del inglés, polaridad celular plana) C. Vía Wnt/calcio, regulada por calcio. Modificado de Ruggiero (2015), imagen creada con bioicons (bioicons.com).

largo del eje anteroposterior [186]. Un rol homólogo podría existir durante los procesos de desarrollo de redias y cercarias de *F. hepatica*, en tanto se observa también sobreexpresión de PTK7 y de un receptor de la familia FGFRL con posible homología a *ndl* de *D. japonica*.

Se observa dentro de los GDEs en IS15 también a posibles homólogos de WNT5 y SLIT, que cumplen un papel en el establecimiento del eje medio-lateral en *S. mediterranea* [182]. Por último, vale la pena mencionar que se ha vinculado a Wnt en el proceso de renovación de neoblastos de planaria [187]. Estas evidencias apuntan a considerar la hipótesis de que participe también de la regulación del proceso de reproducción asexual de esporocistos y redias. Esta hipótesis es contrastable mediante el uso de técnicas como RNAi en *F. hepatica* o *S. mansoni*. El establecimiento del eje dorsoventral depende también del gradiente de expresión de morfógenos. En este caso, juegan un rol central la proteína BMP4⁵, las proteínas Noggin y miembros de la familia Noggin-like (*ngl*) en *S. mediterranea* [182,188,189].

Además de BMP4 y Noggin, dentro de los GDEs sobreexpresados en IS15 relevantes para esta vía se encuentra un miembro de la familia Tolloid/BMP1 (encargado del clivaje del precursor de BMP4 [190]) y SMAD6, que participa del proceso de regulación de la vía al inhibir la acción de BMP4 [188]. Por último, un estudio reciente ha identificado a dos de estos GDEs (*prdm1-1* y *kal*) como marcadores de poblaciones celulares de ejes dorsal y ventral activados en respuesta a BMP secretadas durante el proceso de regeneración en planaria [191].

Si bien el establecimiento del eje dorsoventral en tremátodos se encuentra poco estudiado, existen estudios sobre algunas de las familias génicas implicadas. BMP4 pertenece a la superfamilia TGF- β , compuesta por las subfamilias de ligandos tipo TGF- β y BMP [181].

En los tremátodos los estudios de homólogos de la familia BMP se ha centrado principalmente en miembros de la familia inhibina/activina. Los mismos han sido detectados en *S. japonicum*, *S. mansoni* (SmInAct; [192]) y *F. hepatica* (FhTLM; [193]), donde podrían cumplir roles tanto en el desarrollo de los organismos, como en la modulación del sistema inmune de huéspedes mamíferos. Su papel en el desarrollo del eje dorsoventral se encuentra a la fecha inexplorado.

Por último, la expresión de factores de transcripción de las familias Hox y Fox podría vincularse a la especificación de los ejes anteroposterior, dorsoventral y mediolateral por activación de programas transcripcionales dependientes de los gradientes de los morfógenos antes mencionados [181]. Algunos de estos genes se encuentran también diferencialmente expresados entre IS15 e IS30 (Figura 3.14), mostrando la relevancia que poseen dichos programas transcripcionales en las etapas más tempranas de la infección a huésped caracol.

En resumen, la expresión diferencial simultánea de genes asociados a vías de especificación de los ejes corporales conservados en metazoarios (y ya estudiados en el género *Schmidtea*), y de factores de transcripción asociados a procesos de morfogénesis sugiere que estas vías se encontrarían también conservadas en *F. hepatica*, y es consistente con los procesos de desarrollo y morfogénesis asociados a la generación de redias y cercarias. Debido a la falta de poblaciones sincrónicas y al hecho de poseer información transcritómica global, no es posible determinar precisamente en qué estadios operan dichas vías (esporocistos, redias), y sobre qué poblaciones celulares se encontrarían los presuntos gradientes de expresión de existir. Son necesarios otros estudios más detallados, que permitan definir el patrón espacio-temporal de expresión de estos genes asociados a los procesos de desarrollo.

3.3.3.2 Metabolismo y obtención de nutrientes en estadios intra-caracol

Los estadios intra-caracol se caracterizan por una alta demanda energética asociada al proceso de reproducción asexual y sus procesos de biosíntesis y morfogénesis asociados. Como se vio en la Sección 3.2

⁵Homólogo de la proteína Dpp (*decapentaplegic*) de *D. melanogaster*, nombre con el cual se encuentra anotado el GDE de *F. hepatica* en WormBase ParaSite.

procesos vinculados se ven enriquecidos entre los genes de alta expresión en estos estadíos. Es por tanto esperable que genes codificantes para productos involucrados en la obtención y metabolismo de nutrientes se vean sobreexpresados en estadíos intra-caracol 15 d.p.i. respecto a miracidios de *F. hepatica*.

3.3.3.2.1 Metabolismo de proteínas El enriquecimiento funcional en las vías metabólicas “*Protein digestion and nutrient absorption*” y “*Lyosome*” es coherente con la ya mencionada demanda biosintética existente durante el desarrollo intra-caracol de *F. hepatica*. Se ha señalado que la hemolinfa de huéspedes caracoles de tremátodos posee poca disponibilidad de aminoácidos, por lo cual los estadíos parásitos de estos organismos dependen de la degradación de proteínas existentes en la misma y posterior absorción de los aminoácidos resultantes a través de su tegumento para poder satisfacer esta demanda biosintética [176,194–197].

La degradación de proteínas del huésped tendría así un rol preponderantemente anabólico. Esta asunción se ve respaldada por la poca capacidad de obtención de energía a partir de aminoácidos que exhiben los esporocistos de *S. mansoni* [154]. Algunos de los GDEs sobreexpresados en estadíos IS15, como las catepsinas B y L (Figura 3.12), podrían vincularse a estos procesos de obtención de nutrientes.

Las catepsinas L constituyen el producto preponderante de secreciones de los estadíos intramamífero de *F. hepatica*, con capacidad proteolítica de hemoglobina, fibrinógeno y colágeno en esos hospederos [111,198]. Varios estudios muestran una progresión temporal en la expresión de miembros de la familia en el proceso coincidente con el proceso de invasión y establecimiento [99,199–201]. La observación de expresión diferencial en la transición miracidio-esporocisto (Figura 3.21) con abundante expresión en estos estadíos sugiere fuertemente que miembros de estas familias juegan un papel homólogo en la invasión y establecimiento en el huésped intermediario.

Ya se ha descrito previamente que esta familia multigénica posee amplificaciones específicas de las familias Schistosomatidae, Opisthorchiidae y Fasciolidae [50,51]. Se ha atribuido un rol adaptativo a estas expansiones, y en el caso de Fasciolidae, la integración de datos transcriptómicos y proteómicos al análisis filogenético da soporte a la idea de divergencia funcional, con expresión diferencial en diferentes estadíos de desarrollo de *F. hepatica* [51]. Los datos generados en este estudio muestra que algunos miembros de esta familia (particularmente los subgrupos CL8/9 y CL10), para los que no se poseía información transcriptómica o proteómica poseen expresión diferencial en estadíos larvarios asociados al huésped intermediario.

El hecho que estas secuencias se agrupen según su patrón de expresión tanto en *F. hepatica* como en *F. gigantica*, es sugerente, siguiendo el criterio de parsimonia, de la existencia de al menos cuatro secuencias en el ancestro común de las especies del suborden Echinostomata (*F. hepatica*, *F. gigantica*, *F. buski* y *E. caproni*), con tendencias de expresión diferentes (huevo, estadíos intra-caracol y estadíos invasivos y adultos en el huésped mamífero). La expansión de estas secuencias ancestrales y la conservación del patrón de expresión podría explicar en rasgos generales el patrón observado. La duplicación y divergencia funcional es un mecanismo evolutivo establecido, y ha sido estudiado en múltiples familias codificadas para proteínas secretadas [101,202–204].

Algunos autores plantean el proceso evolutivo de estas familias en dos fases: una primera instancia la retención de copias génicas que provee algún tipo de ventaja selectiva (*e.g.* aumento de dosaje del producto codificado [202]), y una segunda etapa donde la redundancia funcional entre las copias permite la divergencia funcional sin influir en el *fitness* de los organismos portadores [205,206]. Estudios de los sitios activos de los genes expresados en los estadíos asociados al mamífero muestran diferencias funcionales asociadas a cambios en unas pocas posiciones que resultan en diferentes preferencias de sustrato. Una observación preliminar muestra cambios en estos sitios conservados en los genes expresados en los estadíos intra-caracol, sugerente de que también aquí pueden existir especificidades diferentes.

Interesantemente, la integración de datos transcriptómicos de estadíos larvarios al patrón filogenético

inferido para la familia multigénica de legumaínas (aspartilendopeptidasas reguladoras de las catepsinas L [176]) es coherente con la hipótesis de un rol funcional para la expresión estadio específica de catepsinas L durante el desarrollo intra-caracol del parásito. El patrón observado es altamente reminiscente al caso previo. Nuevamente se observa en este caso una descendencia común de secuencias con tendencias de expresión similares, aunque a diferencia de las catepsinas L el patrón observado sugiere la presencia de dos copias en el ancestro común de las especies del suborden Echinostomata (Figura 3.22).

El hecho de que en ambas familias multigénicas se observe divergencia funcional a nivel de la expresión génica, donde algunos miembros tienden a expresar en estadios de desarrollo intra-caracol y otros en estadios de desarrollo intramamífero podría sugerir una coevolución entre copias de ambas familias si se tiene en cuenta que las legumaínas regulan la actividad de catepsinas L.

La hemoglobina constituye el principal componente proteico de la hemolinfa de huéspedes caracol de tremátodos (*i.e.* ~70-80% en *B. glabrata* [195]), siendo razonable que lo mismo acontezca en caracoles del género *Lymnaea*. De ser acertada tal asunción, la regulación de la expresión de copias de legumaínas y catepsinas L permitiría la digestión de hemoglobina presente en la hemolinfa de *L. viatrix*, facilitando los aminoácidos necesarios para los procesos de biosíntesis del desarrollo larvario, en un escenario similar al propuesto para *S. mansoni* [196]⁶.

Es conocido el hecho que los caracoles infectados con *F. hepatica* suelen ser de mayor tamaño, y tienen una capacidad reducida de producir huevos. Esto se ha asociado con la migración de las redias y la castración del caracol por la ingesta directa del ovotestis [207]. Estas proteasas podrían entonces también participar en los procesos asociados a la migración y alimentación de las redias.

La capacidad hemolítica de la catepsina L se ha postulado poseería un doble rol adaptativo: los parásitos helmintos han perdido la capacidad de sintetizar grupos hemo *de novo* (ver “*Parasitic worms and heme*” en [208]), por lo cual se ha postulado que podrían obtenerlos de la hemoglobina haciéndoles posible sintetizar globinas, citocromo P450 y otras proteínas con estos grupos. El aumento de expresión de copias para ferritina resulta coherente con este escenario, en tanto permitiría lidiar el efecto tóxico de los ROS potencialmente generados por reacciones de Fe^{3+} [208]. (invertir) rol que ya ha sido postulado para ferritinas de otros tremátodos y una copia sobreexpresada en huevos de *F. hepatica* [209]. Si bien se asume que la obtención de hierro se obtiene a partir de la hemoglobina obtenida del huésped [196,208], no es posible descartar que la ferritina juegue también un rol en el transporte de hierro, un cofactor necesario para el desarrollo larvario intra-caracol.

La digestión de proteínas de huésped es generalmente asumida como extracelular, aunque se ha postulado que podría tener una última fase de digestión endolisosomal con acción de LAPs [176]. Por último, la expresión diferencial y abundante de quimotripsinógeno parece sugiere que otras proteasas previamente no estudiadas podrían jugar un rol en la obtención de nutrientes a partir del huésped caracol en *F. hepatica*.

Por último, cabe destacar que si bien en la literatura se suele asumir que la degradación de aminoácidos para la obtención de energía sería un proceso secundario en estadios intra-caracol, se ha señalado como excepción el caso de la glutamina, aminoácido que puede proveer intermediarios al ciclo de Krebs vía enzima PEPCK [154]. La síntesis de dicho aminoácido se ve estimulada en células *Bge* de *B. glabrata* en co-cultivo con esporocistos de *S. mansoni* [210]. El gen codificante para PEPCK es también un GDE sobreexpresado en IS15 de *F. hepatica*, por lo cual se podría hipotetizar que podría emplearse este aminoácido como fuente de energía *S. mansoni*.

En resumen, durante la transición miracidio-esporocisto se observa un aumento en la expresión de copias génicas codificantes de proteasas de las familias multigénicas de *i*) quimotripsinogeno; *ii*) leucilaminopeptidasas (LAPs); *iii*) legumaínas y *iv*) catepsinas B y L, proteínas que podrían asistir en los procesos de invasión, establecimiento, y en la obtención de nutrientes del huésped.

⁶Resulta interesante resaltar que dicho estudio se realiza a 20 d.p.i. de los caracoles de *B. glabrata*, haciendo que la comparación de resultados sea al menos razonable.

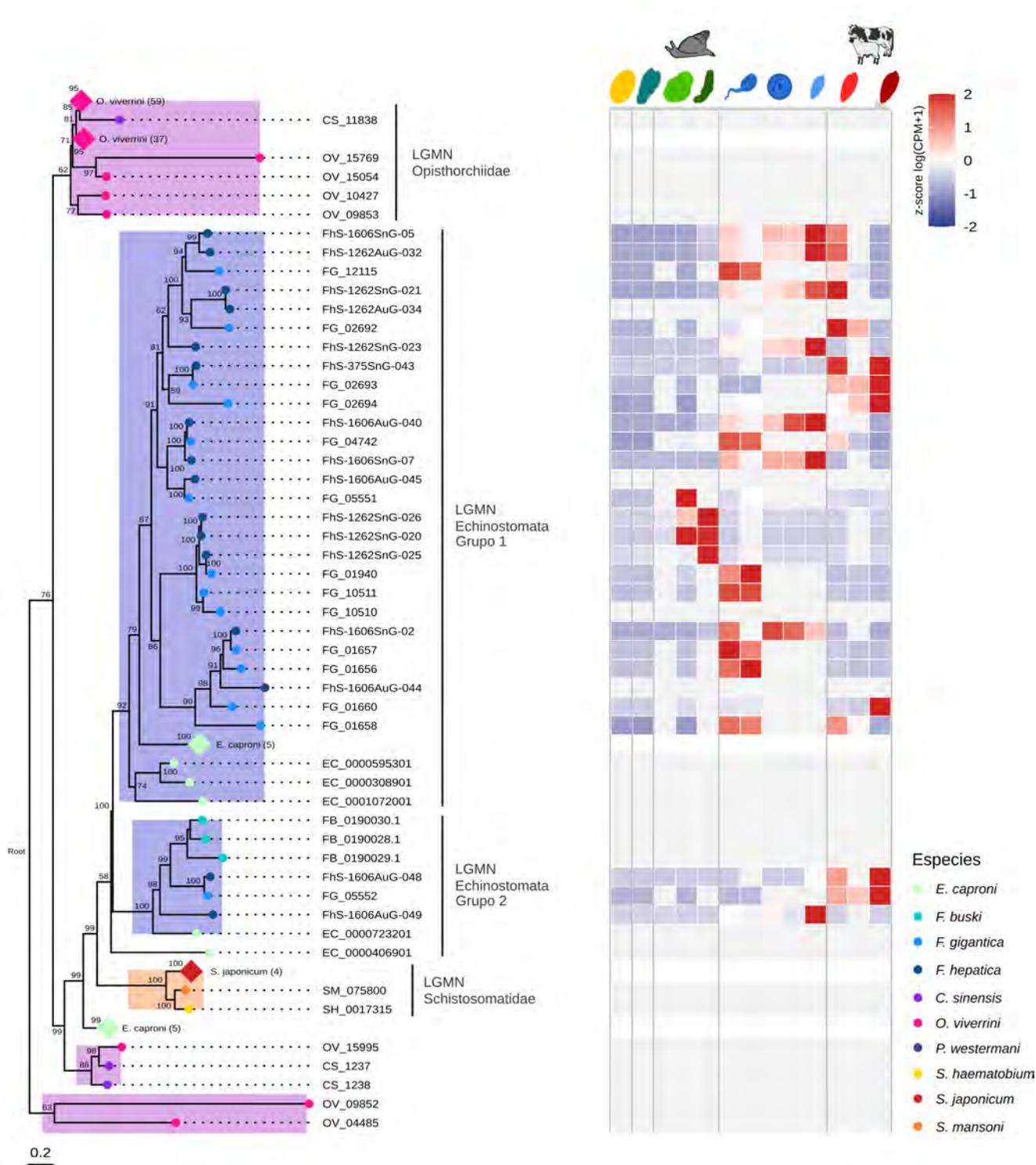


Figura 3.22: Árbol filogenético inferido para miembros de la familia multigénica de legumaínas y tendencias de expresión a lo largo del ciclo de vida para miembros codificados en los genomas de *F. hepatica* y *F. gigantica*. El árbol fue enraizado en su punto medio empleando la librería ape de R.

3.3.3.2.2 Metabolismo de lípidos La detección de expresión diferencial de genes codificantes para transportadores lipídicos (FABP, LDLr y NPC2) así como lipasas y fosfolipasas durante la transición de miracidios a estadíos intra-caracol 15 d.p.i. (Figura 3.12) sugiere que durante el desarrollo intra-caracol de *F. hepatica* también se parasitan estos nutrientes. Es comunmente aceptado que esto ocurre durante el desarrollo dentro del huésped mamífero [10,211], por lo que este es un posible rasgo común de los estadíos parasitarios.

La obtención de energía a partir del catabolismo de lípidos en fasciólidos es en cambio un tema controversial en la literatura. Algunos autores han planteado que dicho proceso no tendría lugar en adultos de *F. hepatica* [10]. A pesar de esta concepción, se ha determinado relativamente poco de este aspecto en estadíos asociados al huésped caracol.

Estudios experimentales parecen respaldar que tal proceso podría ocurrir en esta etapa del ciclo de vida. Humiczewska y Rajski midieron la cantidad de lípidos en glándulas digestivas de individuos de *L. truncatula* infectados con *F. hepatica* mediante técnicas histológicas y citohistoquímicas, reportando que ~50% del contenido de los mismos se encuentra ya depletado a los 20 días de infección⁷. Los autores atribuyen tal movilización a la depleción de glúcidos que genera la infección del parásito, obteniendo el caracol energías de sus reservas lipídicas. No obstante, los autores notan a la vez que tal depleción se da con concomitante presencia de lípidos en tejidos metabólicamente activos de esporocistos, redias y cercarias (*e.g.* masas germinales, faringe, ventosas [212]. Estudios posteriores constataron actividad elevada de esterazas y lipasas en tejidos de estos estadíos durante el proceso infectivo [213]⁸.

Estudios genómicos recientes han corroborado la presencia de los genes pertenecientes a las vías anabólicas y catabólicas de lípidos en *F. hepatica* [51]. En este aspecto resulta notable que la totalidad de los genes codificantes para proteínas implicadas en la beta-oxidación de ácidos grasos se encuentren expresados en los estadíos larvarios del parásito (Figura 3.23)

Si bien estos genes no se encuentran diferencialmente expresados durante la transición miracidio-esporocisto, es claro que se detecta expresión de ARNm para los mismos en estos estadíos. Este resultado es coherente con la posibilidad de que en efecto los estadíos intra-caracol de *F. hepatica* sean capaces de obtener lípidos del huésped con fines tanto biosintéticos como de obtención de energía. De ser correcta esta noción, se constataría lo propuesto por estudios genómicos de McNulty *et al*, donde se postula que *F. hepatica* (y probablemente los tremátodos transmitidos por vía digestiva en general) poseen capacidades metabólicas menos restringidas que las especies de Schistosomatidae [51]. Estudios recientes han mostrado la expresión de genes relacionados a la biosíntesis de lípidos en tremátodos de la familia Psilostomatidae (orden Plagiorchiida, al igual que la familia Fasciolidae) durante la infección a huésped caracol [17], aportando información independiente sobre la capacidad de desempeñar estos procesos por parte de otros tremátodos.

3.3.3.2.3 Metabolismo de glúcidos El metabolismo de los esporocistos de *S. mansoni* es descrito como anaerobio facultativo, siendo la principal vía de generación de energía el catabolismo de glúcidos a partir del ciclo de Krebs [194]. Este estadío posee capacidad fermentadora en condiciones anóxicas *in vitro*, observándose un efecto Pasteur con aumento del consumo de glucosa externa y la aparición de lactato (aparentemente por fermentación homoláctica) y succinato entre los productos excretados. En cultivo *in vitro* este estadío utiliza como fuente principal de energía sus reservas de glucógeno, aunque luego de 24 hs de incubación se observa una depleción considerable de las mismas y un aumento del consumo de glucosa externa [194].

Se ha propuesto por tanto que los esporocistos de tremátodos podrían parasitar *in vivo* las reservas de

⁷Tal cantidad llega a ~80% entre los 40 y los 60 días

⁸Lamentablemente este artículo se encuentra publicado únicamente en su idioma original (polaco), por lo cual sólo se dispone de la traducción del resumen del mismo. Las líneas anteriores fueron escritas con base en el mismo.

glucógeno de sus huéspedes. El estudio de caracoles *G. trivolvis* infectados con larvas del tremátodo *G. pennsylvanensis* parece apoyar esta hipótesis. En dicho sistema se constata una disminución en las reservas de glucógeno y un aumento de glucosa en células hepáticas del huésped, junto a la concomitante aparición de depósitos de glucosa en la superficie externa y tegumento de esporocistos [214]. Tal resultado sugeriría la síntesis de glucógeno en el parásito a partir de la glucosa obtenida de huésped (*e.g.* para la generación de reservas utilizadas en el siguiente estadio de vida libre), así como la obtención de energía a partir de glucosa excedente.

La observación de expresión diferencial de copias codificantes para glucógeno fosforilasa (PYG) y gluco-genina (GYG) en la transición miracidio y estadios intra-caracol 15 d.p.i. es coherente con esta noción (Figura 3.12). A esto se suma la expresión elevada de genes pertenecientes a las vías de fermentación de glúcidos y ciclo de Krebs, discutida en la Sección 3.2. A su vez se constata también la expresión diferencial de una copia codificante para LDH, habilitando también en *F. hepatica* la posibilidad de obtención de energía en condiciones hipóxicas por fermentación homoláctica. Este resultado es esperable, en tanto los caracoles del género *Lymnaea* habitan ambientes donde pueden existir fluctuaciones en los niveles de tensión de oxígeno [215]; se ha propuesto que esto podría llevar *in vivo* a un entorno hipóxico para las larvas de tremátodos [194].

Se suele proponer a la dismutación del malato como la vía metabólica de preferencia para adultos de *F. hepatica* en las condiciones de desarrollo hipóxicas del intestino de huéspedes mamíferos [10]. Los estudios de metabolismo en condiciones de anoxia *in vitro* para esporocistos de *S. mansoni* constatan no obstante como únicos metabolitos secretados al lactato y el succinato, pero no acetato (producto preponderante de la vía del malato) para este tremátodo [194]. Si bien ninguna de las enzimas distintivas de esta vía (*e.g.* enzima málica) se encuentran diferencialmente expresadas en el pasaje de miracidios a esporocistos en *F. hepatica*, ambas se encuentran altamente expresadas en estos estadios, por lo cual no es posible descartar que esta vía también opere en el huésped caracol para este parásito.

3.3.3.2.4 Metabolismo de nucleótidos Un aspecto singular del metabolismo de tremátodos es la imposibilidad de realizar síntesis *de novo* de purinas. Esta incapacidad fue señalada por primera vez por Senft *et al.*, al demostrar que adultos de *S. mansoni* no sintetizan los anillos nitrogenados de purinas por la vía clásica a partir de los carbonos de metabolitos como glicina y formato, entre otros [216]. La expansión clonal de esporocistos y redias supone un esfuerzo mayor en la síntesis de ADN y ARN y en la utilización de intermediarios de energía que también involucran purinas, por lo cual el metabolismo nucleotídico es particularmente relevante para estos estadios. Es notoria la tendencia a la expresión en estadios larvarios asociados a huésped caracol que poseen todos los genes codificantes para enzimas participantes de esta vía metabólica (Figura 3.24).

Esta tendencia parece reflejar una necesidad exacerbada de obtención de nucleótidos para la realización los procesos biosintéticos mencionados anteriormente. Como se mencionó en la Sección 3.2 varios de estos genes se encuentran entre el 10% con mayor expresión en los estadios larvarios asociados a caracol (*i.e.* *hprt1*, *adk*, *adss*, *ada*), coherente con esta interpretación de los resultados. En la misma línea, genes codificantes para transportadores de purinas se encuentran también dentro de los que tienen mayor expresión en estos estadios (SLC43A3, con homólogos que actúan como transportadores de purinas en mamíferos [217]), mientras que otros también aumentan su expresión en la transición miracidio-esporocisto de *S. mansoni* [14].

Resulta interesante observar en algunos genes involucrados en la vía de salvataje de purinas la presencia de parálogos con tendencia a la expresión en estadios intramamífero (*i.e.* genes codificantes para enzimas PNP, AdK, y ADSL). Se podría hipotetizar que esta divergencia a nivel de expresión podría significar algún tipo de adaptación regulatoria a diferencias a nivel del metabolismo de purinas de los huéspedes moluscos y mamíferos. Futuros estudios experimentales son necesarios para determinar las diferencias funcionales existentes entre estas copias.

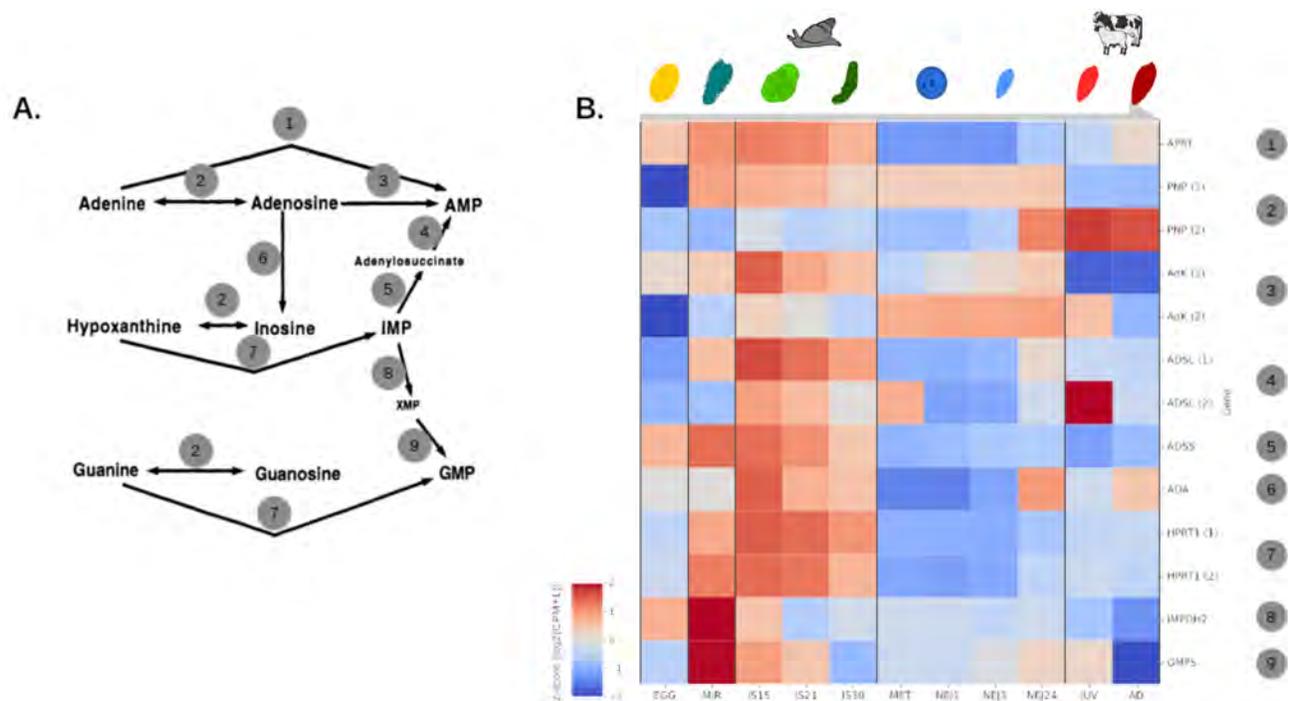


Figura 3.24: Vías de salvataje de purinas relevantes para el desarrollo de tremátodos en sus estadios parasitarios. A. Vía de salvataje de purinas propuesta para *S. mansoni*. Imagen tomada de Berens y Muller (1995). B. Tendencias de expresión de genes involucrados en el salvataje de purinas a lo largo del ciclo de vida de *F. hepatica*.

Además de los estudios detallados en adultos y schistosomulas de *S. mansoni*, en la literatura existen pocos estudios referentes al metabolismo de purinas en tremátodos. Trembacz et al. reportaron la expresión de enzimas involucradas en el salvataje de purinas en tremátodos de tres familias distintas, lo cual nuevamente sugiere que es una característica metabólica conservada en la clase Trematoda [218]. En coherencia con los resultados obtenidos en esta tesis de maestría, la comparación de actividad enzimática entre esporocistos y cercarias (ya en vida libre) de *Nocotocotylus ephemera* muestra una mayor actividad en el estadio de desarrollo intra-caracol [218]. Estudios experimentales deberán ser realizados para dilucidar la relevancia de esta vía metabólica *in vivo* para el desarrollo intra-caracol de *F. hepatica*.

3.3.3.3 Migración e interacción con tejido de huésped

La expresión diferencial de receptores con actividad neuromuscular de las familias FRFamida y 5-HT₁ (receptor de serotonina) podrían estar vinculadas a la actividad motora migratoria, en tanto se ha probado la capacidad de estimulación del movimiento que poseen los neuropéptidos y la serotonina en esporocistos de tremátodos [219]. A su vez, algunas de las proteasas con expresión diferencial en estos estadios podrían estar involucradas en esta migración. Se ha asociado a las catepsinas L y a las catepsinas B en particular a la degradación de las proteínas de la matriz extracelular que acompaña a la migración dentro del huésped mamífero, debido a su capacidad de degradación del colágeno [176,198,201]. La expresión diferencial de copias de estos genes en el estadio intra-caracol podría cumplir un rol homólogo. Una función similar podría vincularse a la expresión de una copia codificante para ADAMTS3, proteína de la familia ADAMTS (“*A Disintegrin and Metalloproteinase with Thrombospondin motifs*”) con capacidad de clivaje de pro-colágeno [220]. Alternativamente esta metaloproteinasa podría estar relacionada con la síntesis de colágeno necesaria para los tejidos de esporocistos y/o redias.

Por último, resulta interesante considerar que tanto la actividad hematófaga de algunos tremátodos como su movimiento en los tejidos de sus huéspedes ha sido asociado con la necesidad de inhibición de procesos como la coagulación, que podría impedir la movilidad de los parásitos. La capacidad de digestión de fibrinógeno que posee la catepsina L de *F. hepatica* constituye un ejemplo, y en el mismo sentido se ha probado que copias tegumentales de calpaínas de *S. mansoni* pueden clivar a la fibronectina [221], vinculada a la coagulación en mamíferos. En el mismo sentido, inhibidores de serin-proteasas como las serpinas e inhibidores de la familia Kazal han sido reportados como reguladores de serin-proteasas de la cascada proteolítica del proceso coagulatorio. La necesidad de lidiar con esto acontece también en el huésped caracol, pero debe ser tenido en cuenta que los sistemas coagulatorios de vertebrados e invertebrados no son homólogos. La investigación de mecanismos coagulatorios en los últimos se ha limitado principalmente a artrópodos [222], por lo cual a la fecha es difícil evaluar si los productos secretados de tremátodos podrían regular dicho sistema. A pesar de esto, si bien estos mecanismos descritos en invertebrados son diversos, algunos de los más estudiados implican también cascadas de serin proteasas. Recientemente se ha descrito que en el caracol *Achatina fulica* la coagulación mediada por detección de LPS utiliza este mecanismo regulatorio [223], abriendo la posibilidad de especulación sobre si mecanismos similares podrían existir en respuesta a tremátodos en caracoles limneos.

3.3.3.4 Respuesta a sistema inmune del huésped caracol

Como se detalló en la Sección 1.5.3 la evasión del sistema inmune constituye un elemento fundamental para el establecimiento de la infección de *F. hepatica* en sus huéspedes. Si bien a la fecha no se han determinado los actores moleculares expresados por *F. hepatica* implicados en la evasión del sistema inmune de *L. viatrix*, estudios sobre la resistencia a infección de *F. hepatica* de algunas poblaciones naturales de otros caracoles limneos (*i.e.* *Pseudosuccinea columella* [33,224]), la respuesta a nivel de expresión de *L. stagnalis* frente a infecciones de tremátodos [28] y del sistema *S. mansoni*-*B. glabrata* [26,27,32] permiten realizar hipótesis razonables a la luz de los resultados obtenidos en este trabajo.

Como se mencionó en la introducción de estas tesis, pruebas experimentales parecen mostrar que existe un recambio de los motivos glucídicos expuesto en la superficie tegumentaria durante el desarrollo intra-caracol de *F. hepatica* [29–31]. A pesar de esto, estos motivos son detectados por lectinas, por lo cual mecanismos adicionales podrían estar jugando un rol en la evasión exitosa de la respuesta inmune del huésped. Estudios recientes han demostrado que los NEJ de *F. hepatica* son capaces de inhibir de forma específica la activación de la vía de las lectinas al expresar dos proteínas de la familia de las serpinas (denominadas *FhSrp-1* y *Fh-Srp2*), las cuales impiden la acción de las proteínas MASP (*MBL-associated serine proteases*) evitando así el clivaje de C3b a C4b y la deposición de este, primer paso de la cascada proteolítica que eventualmente lleva a la deposición del complejo de ataque a membrana [225]. Por lo tanto, el aumento de expresión de serpinas en la transición miracidio-esporocisto (Figura 3.12) sugiere fuertemente que las etapas larvarias intra-caracol de *F. hepatica* podrían emplear un mecanismo homólogo en la evasión de la respuesta humoral innata del huésped caracol.

La expresión de otros inhibidores de serin-proteasas, como miembros de la familia SPINT3 de inhibidores tipo Kunitz o los inhibidores de la familia Kazal, podrían también participar de la regulación de este fenómeno. Incluso componentes estructurales, como la paramiosina, también han sido señalados como relevantes para el impedimento de la unión a lectinas al tegumento de tremátodos [226].

Los efectores humorales de la inmunidad innata de caracoles no se reducen solo a activadores del sistema de complemento, si no que incluyen también a miembros de las familias FREP y TEP de particular importancia en gastrópodos, con algunos de los de sus miembros como BgFREP3 y BgTEP con capacidad de opsonización en *B. glabrata* [26,27]. La expresión diferencial de FREPs en algunos caracoles de *L. stagnalis* infectados con tremátodos de la familia Opisthorchiidae y la capacidad de codificación para TEPs en esta especie sugieren fuertemente que los mismos podrían cumplir funciones homólogas en este caracol [28].

La expresión diferencial de catepsinas L es sumamente sugerente en este contexto, en tanto las mismas poseen capacidad de proteólisis de fibrinógeno [176]. Las proteínas de la familia FREP (y de las familias relacionadas CREP y GREP de rol aparentemente similar) pueden interactuar tanto con hemocitos como con el tegumento de esporocistos a partir de dominios tipo fibrinógeno [26], por lo cual esta capacidad de proteólisis selectiva de los mismos sugiere un posible rol directo de catepsinas L para el bloqueo de la respuesta humoral del caracol, análogo al que se ha descrito en la evasión de respuesta del huésped mamífero donde degrada inmunoglobulinas [227]. Si bien estos efectos han sido observados estudios in vitro con catepsinas producidas por los estadios asociados al huésped mamífero, es plausible que mecanismos similares estén en juego utilizando la batería de catepsinas expresadas en los estadios intra-caracol. Se ha reportado que otras proteasas como las calpaínas también podrían estar involucradas en procesos de respuesta similares en *S. mansoni* [27], por lo cual resulta interesante la sobreexpresión de genes codificantes para proteínas de esta familia en estadios intra-caracol 15 d.p.i..

La expresión diferencial de cistatinas podría poseer un rol funcional en este contexto. Los miembros de esta familia son inhibidores de cisteín-proteasas que se han aislado en complejos con catepsinas L de *F. hepatica*, por lo cual se ha sugerido un rol regulatorio sobre las mismas [228,229]. El mismo rol podría cumplirse en el contexto de la infección a huésped molusco. No obstante, BgTEP es también una cisteín-proteasa [230], por lo cual la expresión de cistatinas podría jugar en este contexto también un rol de evasión de este actor del sistema inmune humoral del caracol. Resulta también interesante que en nemátodos las stefinas (miembros de la superfamilia de cistatinas) module las respuestas de citoquinas y la presentación a antígenos, induciendo un ambiente anti-inflamatorio (comentado en [228]).

Como se mencionó en la introducción a esta tesis, la expresión de especies reactivas de oxígeno y nitrógeno (“ROS” y “RNS”, respectivamente, en inglés) es fundamental durante la respuesta inmune de caracoles frente a tremátodos. La expresión diferencial de copias codificantes para proteínas antioxidantes ya descritas como secretadas y relevantes en la respuesta a ROS en huésped mamífero resulta por lo tanto interesante, en tanto sugiere un mecanismo de respuesta frente a la respuesta celular del sistema inmu-

ne del huésped. Esto incluye miembros de la familia GST y la familia de las tioredoxinas entre otras. También es de destacar la expresión diferencial de HSP70 y de copias codificantes para retinol deshidrogenasa (RDH13, de la familia SDR de deshidrogenasa). Miembros de estas familias han sido identificadas como relevantes en el proceso de detoxificación de ROS y especies reactivas de carbonilo en el céstodo *E. granulosus* [161].

Es interesante que se encuentra también diferencialmente expresada una copia codificante para peroxidasa, identificada como diferencialmente expresada por hemocitos de *L. stagnalis* en contexto de infección [28]. Una caracterización funcional de la copia de peroxidasa expresada por esporocistos de *F. hepatica* es necesaria para poder comprender este resultado anti-intuitivo y determinar si el parásito también utiliza la producción de ROS de forma hostil hacia los hemocitos u otras células del caracol.

La secreción de proteasas por parte de hemocitos también ha sido descrita en la literatura. Interesantemente, en el caso de *B. glabrata* esto incluiría la secreción de catepsinas y elastasas (familia de serin proteasas) [32]. Por lo tanto, es notable que en la transición miracidio-esporocisto se liberen los inhibidores de proteasas ya mencionados, los cuales pertenecen a por lo menos dos categorías funcionales (*i.e.* inhibidores de serin-proteasas y de cisteín-proteasas). Otros casos similares, como la presencia de copias codificantes para miembros de la familia PI16 (de la superfamilia CAP) o la ribonucleasa T2 (RNASET2), sugieren que baterías de proteasas e inhibidores son utilizados tanto por el parásito como el huésped como mediadores de la interacción. Futuros estudios serán necesarios para dilucidar la acción de estas proteínas en el contexto de infección a huésped caracol, como ya ha sido sugerido previamente [15,231].

Por último, cabe mencionar que copias de algunos de los genes anteriormente nombrados (*i.e.* tioredoxinas, GST, inhibidores tipo Kunitz, e incluso FABP) han mostrado roles en la regulación de la expresión de citoquinas en huésped mamífero, pudiendo así influir en el reclutamiento y maduramiento de efectores celulares de la inmunidad [227]. Queda por estudiar si roles homólogos son ejercidos en el contexto de la infección a huésped caracol, donde se ha detectado la expresión de citoquinas homólogas a las del sistema inmune mamífero que cumplen roles similares en hemocitos (*e.g.* MIF) [32].

3.3.4 GDEs sobreexpresados en estadíos intra-caracol 30 d.p.i.

Los estadíos intra-caracol 30 d.p.i. son en general una mezcla compleja de redias y cercarias, con predominio de las últimas. Al igual que en otros estadíos de desarrollo intra-caracol, la migración constituye un proceso esencial ejercido en estas etapas; la regulación del movimiento y la secreción de proteasas que permitan la degradación de tejidos y matriz extracelular son por tanto aspectos característicos de estos. La secreción de proteasas permite a su vez la salida del huésped caracol por parte de las cercarias, el último de los estadíos que se desarrolla en este organismo. Por último, una característica a remarcar de estos estadíos es la formación del tegumento, que en última instancia constituirá la capa interna del quiste de la metacercaria, y será el precursor del tegumento sincitial característico del adulto. A continuación se presentan algunos GDEs sobreexpresados en estadíos intra-caracol 30 d.p.i. vinculados a estos procesos, muchos de los cuales son pertenecientes a familias multigénicas nombradas en secciones anteriores.

3.3.4.1 Proteasas, inhibidores de proteasas y otras proteínas secretadas

Algunas de las vías metabólicas de la base de datos KEGG que se ven enriquecidas entre los GDEs sobreexpresados en *IS30* respecto a *IS15* son las mismas reportadas para los GDEs sobreexpresados en *IS15* respecto a miracidios (*e.g.* vías “*Protein digestion and absorption*”, “*Antigen processing and presentation*”, “*Lysosome*”). Esto se debe al aumento de expresión para un número considerable de proteasas secretadas, principalmente de las familias de legumainas, y catepsinas B y L, algunos de los miembros también sobreexpresados en la transición miracidio-*IS15* (Figura 3.12). La sobreexpresión de un gen codificante para EXOC3/SEC6, miembro del sistema intracelular que permite la secreción, es coherente con el aumento

de necesidad de secreción activa que poseen las cercarias en desarrollo. Una inspección visual de los genes sobreexpresados anotados para las vías metabólicas sobrerrepresentadas de difícil interpretación, tales como “Apoptosis”, “Autophagy - animal”, “Fluid shear stress and atherosclerosis”, “Renin secretion” o “Relaxin signaling pathway” señala que las mismas se ven asociadas principalmente a estos mismos genes; es probable por tanto que la sobrerrepresentación de las mismas no refleje aspectos particulares de la transición entre estadios de desarrollo intra-caracol, si no que se deba a un artefacto producto de la variedad de procesos a los que se ven asociados estos genes en metazoarios.

La sobreexpresión de miembros de estas familias en la transición de esporocistos a cercarias ha sido reportada también en tremátodos como *F. gigantica* y *S. japonicum* [15,16]. Esto es de esperar, en tanto las proteasas constituyen el componente principal de las secreciones de cercarias de tremátodos [176]. La sobreexpresión de las mismas en cercarias sugiere un rol para miembros de estas familias en las últimas etapas proceso de migración intra-caracol y probablemente también en la salida de este huésped. La observación de coexpresión con miembros de la familia de legumainas es coherente con el rol regulador de la función de catepsinas postulado para miembros de esta familia. Resulta interesante remarcar son distintas isoformas dentro de las mismas familias las sobreexpresadas en estadios intra-caracol temprano y tardío (Figuras 5.10 y 5.11).

Se observa también la expresión diferencial de miembros de familias de inhibidores de proteasas tales como serpinas e inhibidores de la familia tipo Kunitz (Figura 3.25).

Dentro de los clados observados para los genes integrantes de este grupo de homólogos se observan tendencias de expresión variadas, incluyendo tendencias a la expresión en adultos, huevos y estadios de desarrollo intra-caracol. Se vislumbran dos clados independientes que presentan tendencia a la expresión en estadios intra-caracol de *F. hepatica*. En ambos grupos se observa en la mayoría de los casos que esta expresión es mayor en estadios intra-caracol tardíos (IS30) de *F. hepatica* o en cercarias y metacercarias de *F. gigantica*. Resulta interesante que genes codificantes para proteínas tipo Kunitz o de la familia PI16 presenten mayor expresión en estadios intra-caracol tempranos, mientras miembros del grupo de homólogos OG000045 presenten esta tendencia a la expresión más bien tardía dentro de este huésped. Esto podría reflejar diferentes fenómenos, incluyendo respuestas diferenciales frente a diferentes entornos dentro del caracol (debido a migración en tejidos) o a cambios en las respuestas inmunes de huésped a medida que avanza el tiempo de infección. Estudios experimentales son necesarios para dilucidar entre estos y otros escenarios, pero el hecho de que patrones similares se observen en estas dos especies del género *Fasciola* resulta sugerente de la relevancia funcional de esta familia multigénica.

En otros casos se sobreexpresan algunos genes vinculados a familias que no presentan expresión diferencial en otras etapas del desarrollo, como por ejemplo una copia de leishmanolisina en IS30. Las leishmanolisinas son metaloproteasas que han sido reportadas como inmunomoduladoras de relevancia para la sobrevida en huésped caracol en *S. mansoni* [232]. Estudios proteómicos revelan que constituye ~85% de las proteínas secretadas por cercarias libres de *S. japonicum*, por lo cual también se le ha adjudicado un rol a invasión a huésped en este tremátodo [233]. Si bien no se ha descrito una posible función en la interacción de fasciólidos con huéspedes caracoles, es de esperar que pueda jugar papeles similares a otras proteasas en esta interacción.

La sobrerrepresentación de la vía metabólica “Retinol metabolism” se debe a la expresión diferencial de copias codificantes para retinol deshidrogenasa 12 y 13 (RDH12/13), de la familia SDR de reductasas y deshidrogenasas de compuestos de cadena corta. Se observa expresión diferencial para otros miembros de la familia, como la aldheído deshidrogenasa (ALDH18A1) y estradiol-17 β -deshidrogenasa (HSD17B12). Este último gen se ha asociado a funciones de detoxificación de compuestos con carbonilos reactivos (generados en presencia de ROS) en el cestode *E. granulosus* [161]. Se observa la expresión diferencial también de otros componentes del sistema de respuesta a estrés oxidativo, como tioredoxina, superóxido dismutasa, HSP70 o fosfolípido-hiperperóxido glutatión peroxidasa (GPX4). Nuevamente, algunos de miembros de estas familias también son sobreexpresados en la transición miracidio-esporocisto. La sobreexpresión de

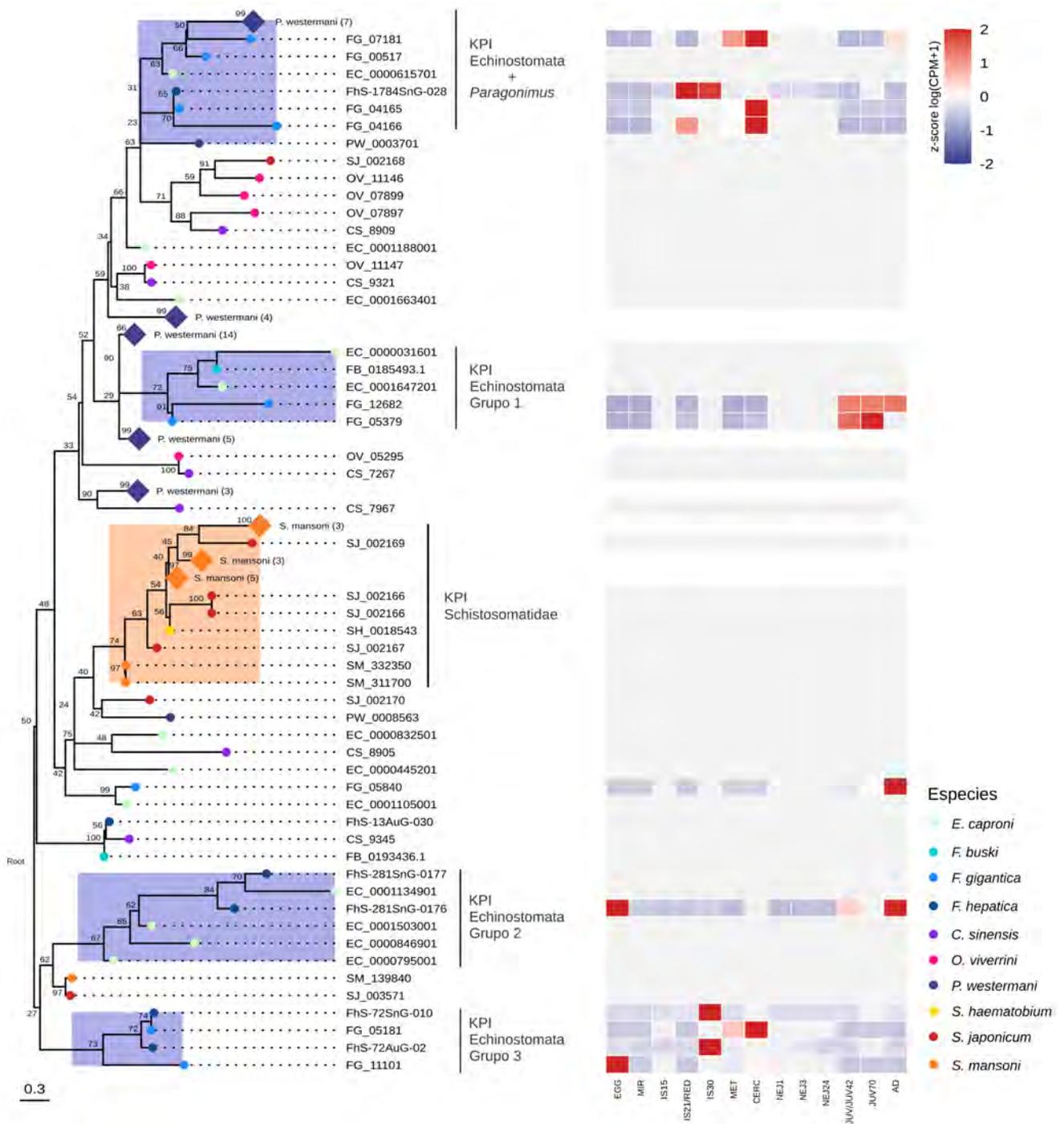


Figura 3.25: Arbol génico para miembros del grupo de homólogos OG0000045, con miembros homólogos a proteínas de la familia de inhibidores de serin-proteasas tipo Kunitz. Se muestran las tendencias de expresión (z-score) de cada gen para los miembros codificados en los genomas de *F. hepatica* y *F. gigantica*. El árbol filogenético fue inferido por método de máxima verosimilitud empleando el programa IQTree v2 sobre alineamientos de proteínas, realizando la selección del modelo evolutivo a emplear con el programa ModelTest y calculando el soporte de rama para clados mediante método de bootstrap bajo la modalidad de ultrafast-bootstrap. Arbol enraizado en su punto medio empleando la librería ape de R.

estos genes podría estar vinculada nuevamente a la respuesta al sistema inmune de caracol, o podría alternativamente funcionar para mantener la homeostásis redox durante la fase de vida libre de la cercaria de *F. hepatica*, caracterizada por un metabolismo sumamente activo. Resulta interesante resaltar también que tioredoxinas o superóxido dismutasas poseen alta expresión en metacercarias de *F. hepatica* y *F. gigantica* [16,179], atribuyéndoseles en este caso un rol protector frente al estrés oxidativo ambiental. La sobreexpresión de ARNm codificante para genes con estas funciones en cercarias podría también estar vinculado a este fenómeno.

3.3.4.2 Obtención de nutrientes del huésped

Algunos de los GDEs sobreexpresados en cercarias se encuentran relacionados a la obtención de nutrientes a partir del huésped. Se observa nuevamente expresión diferencial para transportadores de lípidos, en este caso posibles copias codificantes para FABP3 y FABP7. Trabajos histológicos han descrito la presencia de gotas de lípidos en cercarias de *F. hepatica*, resaltando que se encuentran concentradas en tejidos metabólicamente activos como el tegumento o los ventosas orales. Se menciona a su vez que se ha constatado la presencia de enzimas activas vinculadas al metabolismo de ácidos grasos (*i.e.* deshidrogenasas de ácido β -hidroxibutírico y acetil-CoA) en este estadio [213]. Nuevamente, la expresión diferencial encontrada es coherente con parte de la literatura en este aspecto, apuntando hacia la obtención y uso de lípidos tanto para la biosíntesis como para la obtención de energía en todos los estadios intra-caracol.

Dentro de los GDEs con mayor cambio en su expresión se encuentra un gen codificante para una posible aquaporina (AQP4). Se ha descrito que existe regulación temporal en la expresión de miembros de esta familia multigenica en los estadios intramamífero de *F. hepatica* [179]. Asimismo, se ha reportado la expresión diferencial de aquaporinas particulares de los estadios de vida libre y parásitos de *F. gigantica* [16]. En las especies de la familia Psilostomatidae (orden Plagiorchiida), donde se observan copias con expresión preferencial en cercarias [17]. En todos los casos se ha sugerido que la expresión diferencial de miembros de la familia permite la regulación osmótica a los diferentes estadios a medida que cambian de ambiente. Esta hipótesis resulta razonable para las cercarias, que emergen del caracol al medio externo. Alternativamente, miembros de la subfamilia de aquagliceroporinas permiten también el intercambio de pequeños solutos como el glicerol, por lo que podrían favorecer la adquisición de nutrientes del huésped. Otros genes codificantes para transportadores de solutos (SLC3A2, SLC22A8) podrían facilitar procesos similares.

3.3.4.3 Proteínas con rol estructural y regulación del movimiento

La expresión diferencial de genes codificantes para proteínas estructurales como β - y γ -actina (ACTB, ACTB/G), cadenas livianas de dineína (DYNL), miofilina, proteínas de unión a actina (gelsolina) o reguladores de la dinámica de microtúbulos (RMDN1) podrían vincularse nuevamente a la regulación del movimiento, necesario para la emisión de las cercarias. En este aspecto muestran cierto paralelismo con la expresión diferencial observada en miracidios respecto a IS15, en tanto también se observa expresión diferencial de PDE1c, calmodulina-4 (CALM4) y PKA, también posiblemente vinculadas a la expresión de este fenómeno. Alternativamente, la secreción activa tanto de proteasas como de vesículas con componentes formadores del tegumento podría requerir un aumento de la expresión de estas proteínas estructurales y de la regulación del movimiento de estas vesículas. Coherente con este aspecto se observa la expresión diferencial de componentes estructurales de vesículas extracelulares (EXOC3/SEC6, SEC22B). Estas pueden estar asociadas a la acumulación en vesículas de elementos necesarios ya sea para la formación del quiste, o a posteriori, para las primeras interacciones con el huésped mamífero una vez que el parásito emerge del quiste.

Con relación estructural a la formación del tegumento se encuentran también dos de los GDEs con mayor cambio, los cuales son codificantes para isoformas de colágeno (COL2A y COL5AS). Estudios estructurales de glicocalix de cercarias muestran que el mismo está compuesto por fibrillas [10]. La ausencia de estas estructuras en esporocistos y redias podría por tanto explicar en parte la expresión diferencial para dichos genes.

3.3.4.4 Biosíntesis de mucinas y otros componentes del tegumento

El enriquecimiento observado para las vías metabólicas “*Mucin-type O-glycan biosynthesis*” y “*Glycosphingolipid biosynthesis - lacto and neolacto series*” condice con la expresión diferencial de genes participantes de la vía de síntesis de O-glicoproteínas (también conocidas como mucinas). La síntesis de las estructuras glucídicas de los O-glicanos acontece en el aparato de Golgi, gracias al accionar secuencial de glicosiltransferasas que actúan sobre una estructura denominada antígeno Tn que se forma al comienzo del proceso. Estudios genómicos muestran, en efecto, que las vías de biosíntesis de estos polisacáridos se encuentran completas en el genoma de *F. hepatica* [234], y se ha descrito la formación de mucinas en NEJ de *F. hepatica* [235].

Los genes participantes en la síntesis de las estructuras núcleo de O-glicanos de tipo I y tipo II se encuentran diferencialmente expresadas en estadios intra-caracol 30 d.p.i. respecto a 15 d.p.i. (Figura 3.26), acorde al rol que cumplen los mucopolisacáridos en el desarrollo de la cercaria y, posteriormente, metacercarias.

En cercarias se ha hipotetizado un rol protector para los mucopolisacáridos presentes en el tegumento, los cuales también podrían facilitar el proceso de adhesión a superficies que sucede durante el enquistamiento. En metacercarias el tegumento también posee un rol protector, constituyendo a su vez la interfaz de interacción con el huésped mamífero hasta el desenquistamiento [2,10]. En especies de la familia Plagiorchiidae tanto las glándulas de secreción encargadas de proveer los mucopolisacáridos del tegumento de cercarias (que eventualmente conformará la capa externa del quiste la metacercaria), como la glándula quistogénica encargada de secretar el material que conforma la capa interna del quiste son formadas durante la morfogénesis de cercarias [2], por lo cual es esperable un aumento de expresión de estos genes a mayores tiempos de infección de huésped (con concomitante aumento de la proporción poblacional de cercarias).

Acorde a lo predicho por McVeigh et al en su estudio genómico de las capacidades de síntesis de glicanos de *F. hepatica*, los estadios larvarios expresan también genes necesarios para la síntesis de N-glicanos (Figura 3.27), aunque ninguno de estos se encuentra diferencialmente expresado en estadios intra-caracol 30 d.p.i.. Se ha reportado la abundancia de proteínas N-glicosiladas tanto secretadas como presentes en ventosas y espinas y en la superficie del tegumento de *F. hepatica* [171,234], por lo que constituyen una primera línea de interacción en la compleja interfaz huésped parásito.

Las estructuras base de O-glicanos y N-glicanos sintetizadas por las vías representadas en las Figuras 3.26 y 3.27 son modificadas mediante la acción de enzimas que permiten la adición de glúcidos, aumentando así la complejidad de O y N-glicanos. Algunos de los genes implicados en este proceso de ramificación también se encuentran expresados en estadios larvarios asociados al huésped caracol, y en algunos casos se encuentran diferencialmente expresados en estos estadios respecto a 15 d.p.i. (Figura 3.28).

Acorde a estudios previos, se observa expresión en estadios larvarios de *F. hepatica* para un conjunto acotado de los genes participantes en estas vías [234], incluyendo además otros genes de la vía que no habían sido detectados en los estudios previos. Futuros estudios serán necesarios para dilucidar el papel de estos genes en la síntesis del repertorio de N- y O-glicanos característico.

En conclusión, la detección de expresión de genes codificantes para enzimas participantes en la síntesis de N- y O-glicanos, así como la detección de expresión diferencial entre IS15 e IS30 para genes como

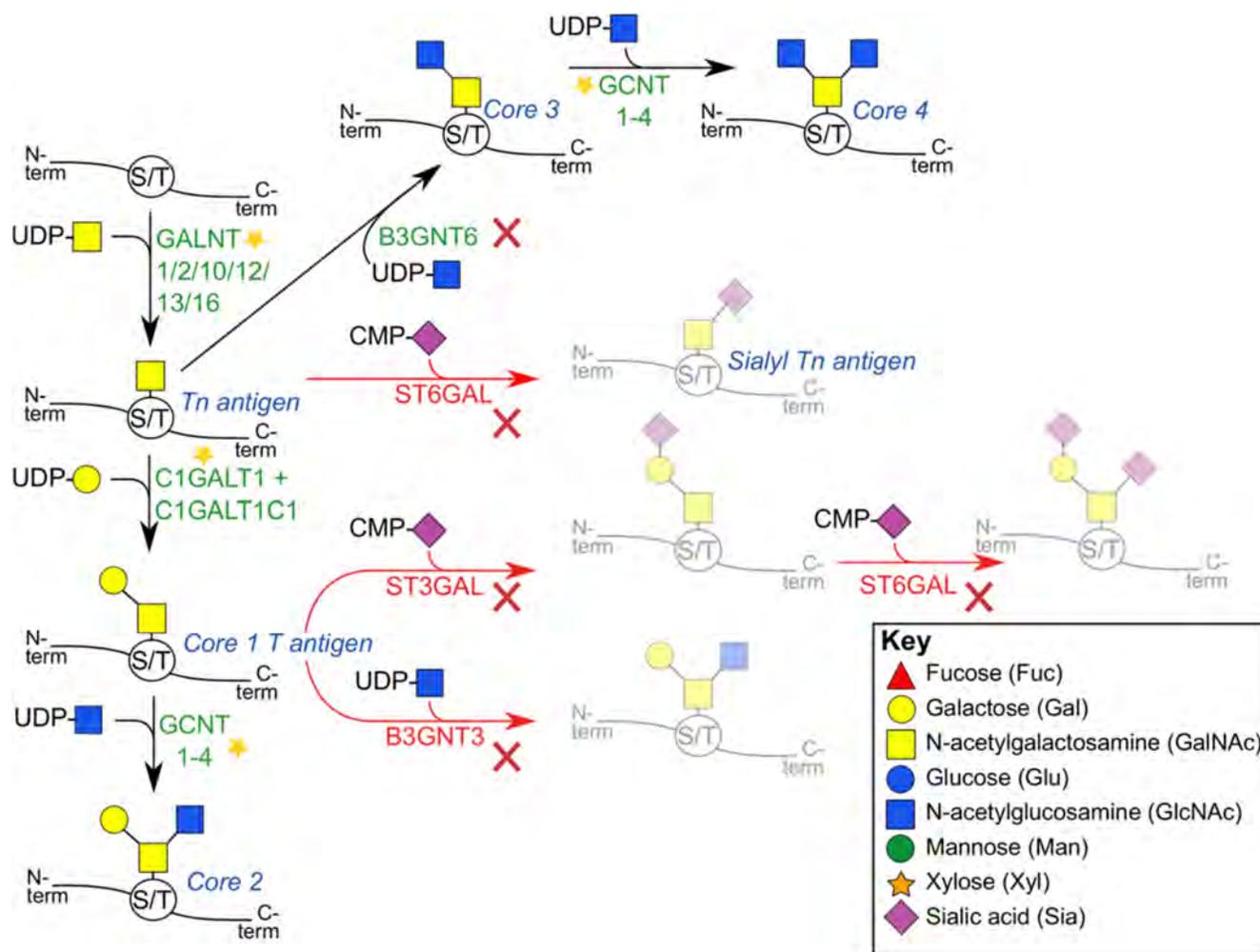


Figura 3.26: Vía de síntesis de las estructuras base de O-glicanos en *F. hepatica*. Se indica la presencia/ausencia de expresión en estadios larvarios del parásito asociados a huésped caracol (tick verde y cruces rojas), así como GDEs entre IS15 e IS30 (estrellas). Estructuras no sintetizables según estudios genómicos de McVeigh et al son mostradas en colores tenues. Figura modificado de McVeigh et al (2018).

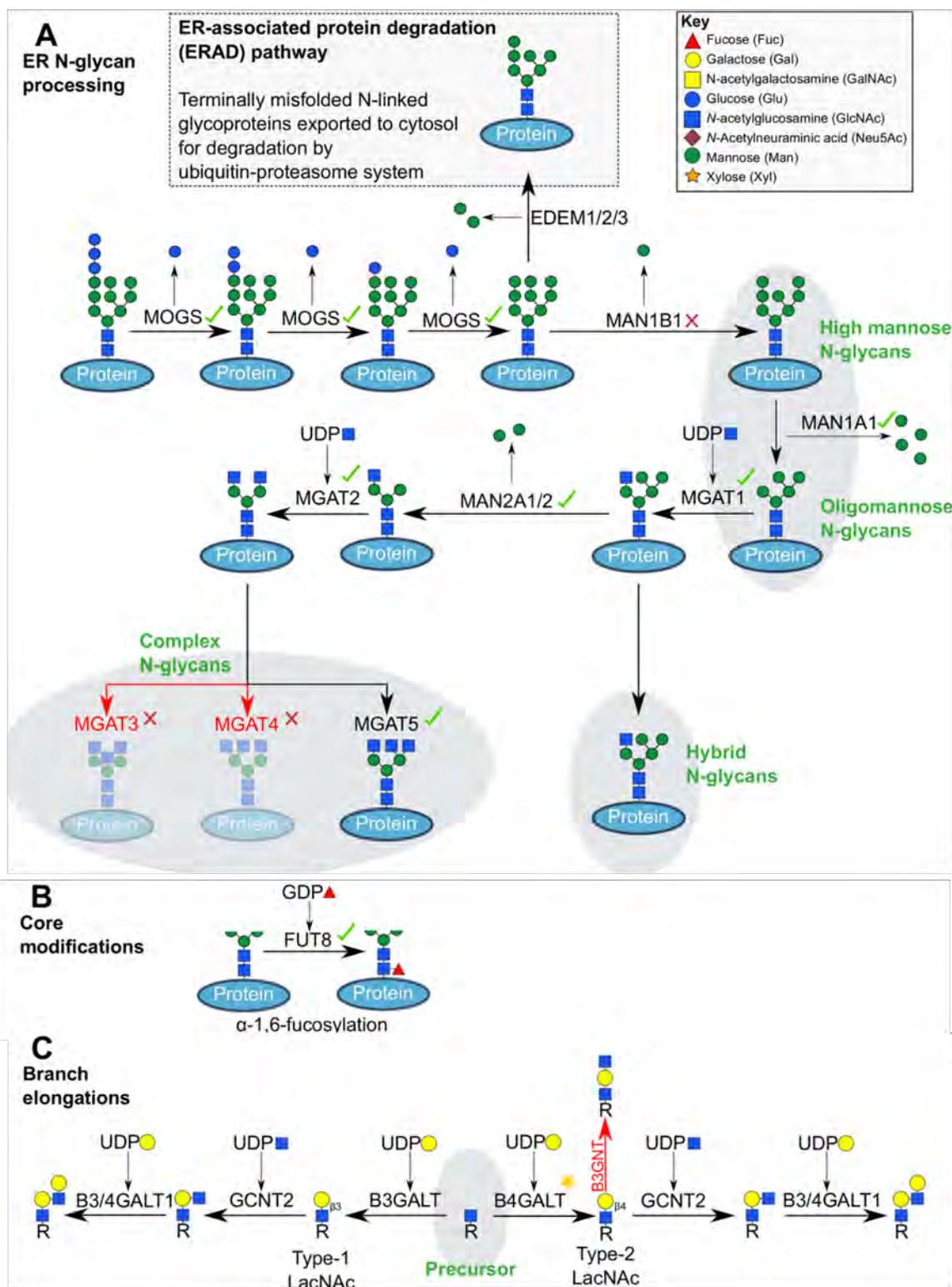


Figura 3.28: Vías de procesamiento de N-glicanos y ramificación de N- y O-glicanos en *F. hepatica*. Se indica la presencia/ausencia de expresión en estadios larvarios del parásito asociados a huésped caracol (tick verde y cruces rojas), así como GDEs entre IS15 e IS30 (estrellas). A. Procesamiento de N-glicanos por glicosidasas y glicosiltransferasas en retículo endoplásmico, el cual da lugar a la formación de estructuras de mayor complejidad. Flechas rojas indican ausencia de ortólogos en *F. hepatica*, según estudio genómico de McVeigh et al. (2018) B. Vía de fucosilación de estructura base de N-glicanos por acción de la fucosiltransferasa FUT8 C. Extensión de las ramas de N- y O-glicanos por acción de un conjunto acotado de glicosiltransferasas presentes en el genoma de *F. hepatica*. Figura modificado de McVeigh et al. (2018).

C1GALT1, CGNT1, GALNT, B4GALT2 (participantes de estas vías) indica que a el aumento de proporción poblacional de cercarias corresponde con un aumento de la expresión de genes vinculados a la biosíntesis de este material secretado con funciones para el desarrollo de cercarias y metacercarias.

Capítulo 4

Discusión general, conclusiones y perspectivas

Esta tesis constituye la primera aproximación transcriptómica al estudio de los estadios larvarios de *F. hepatica* asociados al huésped intermediario, los caracoles del género *Lymnaea*. Este aporte brinda información valiosa respecto a esta etapa poco estudiada del ciclo de vida del parásito.

Al inicio de este estudio se procesó de forma homogénea las lecturas empleadas y se verificó la calidad de las secuencias resultantes. Esta estrategia asegura que se cumpla con criterios mínimos de calidad de forma independiente a la procedencia de cada muestra, algo importante si se tiene en cuenta que parte de los análisis realizados involucran datos provenientes de estudios independientes.

De este análisis se concluye que la secuenciación de las muestras biológicas extraídas de caracoles fue exitosa, resultando en lecturas de alta calidad. Coherente con esto, todas las muestras presentaron una tasa de mapeo superior al 70% contra un ensamblado genómico de referencia de *F. hepatica*. Se determinó a su vez que la mayoría de las lecturas que no mapearon contra este ensamblado no provienen del huésped caracol. Es posible que se trate de otras fuentes de contaminación (*e.g.* alimento ingerido por el huésped, material humano) o que sean transcriptos codificados en regiones mal ensambladas de este genoma de referencia. Es de esperar que entre las mismas se encuentren transcriptos expresados por *F. hepatica*, siendo estos datos procedentes de estadios intra-caracol no contemplados en esta tesis. Un estudio de los mismos en el futuro puede brindar información adicional relevante.

Mediante el uso de métodos estandarizados de mapeo y ensamblado se identificaron posibles isoformas novedosas, se ensamblaron transcriptos codificados en regiones genómicas no anotadas y se determinaron los niveles de expresión génica. Los resultados obtenidos sugieren que el *splicing* alternativo podría ser un mecanismo empleado por *F. hepatica* para aumentar la diversidad transcripcional. Datos recientes apuntan a que lo mismo podría ocurrir en estadios intra-caracol de *S. japonicum* [15]. El análisis de la cobertura promedio para las isoformas reportadas impide llevar a cabo un análisis detallado, pero muestra que estos estadios expresan isoformas novedosas. Estudios a mayor profundidad de secuenciación o con técnicas de secuenciado a resolución celular (*single cell RNAseq*) serían necesarios para aportar información al respecto. La secuenciación con tecnología de lecturas largas (*i.e.* secuenciación con PacBio y/o Oxford Nanopore) también podría aportar información valiosa en este aspecto (si se realiza a suficiente cobertura). Eventualmente aproximaciones de este estilo podrían incluso permitir determinar si existe expresión diferencial de isoformas con un posible rol funcional. Parte de esta diversidad incluye a transcriptos codificados en regiones génicas previamente no anotadas, ayudando a mejorar el estado de anotación de este ensamblado genómico. Más aún, el perfil de expresión estadio-específico de estos modelos génicos los hace interesantes en sí, y el hecho de que algunos de ellos codifiquen para proteínas en *F. gigantica* que presentan el mismo patrón de expresión es coherente con un posible rol funcional

de estos. Como perspectiva de trabajo es posible utilizar algunos refinadores de predicción y anotación genómica como BRAKER2 y MAKER, a efectos de refinar estos modelos génicos, pudiendo pasar por una fase de refinado manual.

Más allá del estudio de la diversidad transcripcional expresada por un organismo en un momento dado, una de las aplicaciones más atractivas y utilizadas de la transcriptómica es la búsqueda de genes con expresión relevante por su rol funcional y/o adaptativo para el desarrollo del organismo estudiado en su medio. En esta línea el análisis se centró principalmente en la identificación de genes con tendencia a la expresión específica (*GDAVs*) en algún estadio intra-caracol y genes diferencialmente expresados entre estos estadios, en el entendido de que son claros candidatos a jugar roles relevantes en la biología de estos estadios.

El primero de estos abordajes se logró mediante la integración de los nuevos datos de expresión obtenidos para esta tesis a los ya disponibles para otros estadios de desarrollo de *F. hepatica*. De los análisis multivariados se concluye que los estadios asociados a huésped caracol y mamífero poseen diferencias en sus patrones de expresión. Acorde a lo esperado estadios de desarrollo consecutivos en el tiempo tienden a ser similares entre sí. Es de remarcar que esto se da incluso cuando en el análisis se integraron muestras provenientes de estudios independientes, lo cual sugiere que la variabilidad biológica de las muestras es menor a su variabilidad técnica. Para intentar minimizar la tasa de falsos positivos se optó por considerar cambios de expresión marcados (*i.e.* $FC > 16$ entre estadios y respecto a la media de expresión) al momento de definir un conjunto de *GDAVs*. En el caso de los análisis de expresión diferencial entre estadios intracaracol, los mismos fueron realizados sobre conjuntos de muestras de un mismo *batch* de secuenciación.

El agrupamiento de estadios biológicamente similares se observa repetidamente cuando se analizan los genes de mayor expresión en cada estadio, se infieren *clusters* de genes con patrón de expresión similar o se realizan agrupamientos de los genes diferencialmente expresados con mayor proporción de cambio entre estadios. Esto implica que los estadios larvarios asociados a huésped caracol comparten un perfil transcriptómico distinto al resto de los estadios. El caso más marcado son los genes *GDAV*. Estos genes son los que poseen mayor variabilidad en su expresión a lo largo del ciclo de vida de *F. hepatica*. Antagónicamente existen genes (*GEEs*) que se encuentran en el extremo más estable de expresión a lo largo del ciclo.

Los análisis de secuencia y estimación de tasas de divergencia aminoacídica realizados sugieren que los genes *GDAV* poseen algunas características que los hacen un grupo particular respecto a la media de los genes codificantes con homólogos detectados, y en particular respecto a los *GEEs*. Cuando se compara estos grupos de genes en *F. hepatica* y *F. gigantica*, los *GDAV* tienden a presentar tasas de sustitución mayores a la media, mientras que los *GEE* poseen tasas menores. Esto puede sugerir la evolución bajo regímenes selectivos diferentes, aunque esta conclusión debe tomarse con cautela ya que diferencias en las tendencias de duplicación podrían también explicar el fenómeno. Al mismo tiempo, los *GDAV* tienden a presentar una conservación taxonómica menor a la de los *GEE* tanto en *F. hepatica* como en *F. gigantica*. A esta observación se suma el hecho de que los primeros poseen un estado de anotación funcional menor a la media, mientras que los últimos se encuentran más anotados que esta. En conjunto y teniendo en cuenta su dinámica transcripcional, esto sugiere que los *GDAV* pueden estar cumpliendo funciones específicas de cada estadio y menos conservadas evolutivamente, mientras que los *GEE* podrían estar codificando para productos con funciones moleculares requeridas en un mayor número de contextos y más conservadas evolutivamente. Una perspectiva de trabajo interesante es ahondar en el estudio de los genes *GDAV* con el fin de obtener más indicios sobre sus posibles roles, ya sea mediante enfoques *in silico* como eventualmente experimentales en un conjunto acotado de estos genes. Si bien el enfoque clásico de transferencia de anotación por inferencia de homología a través del alineamiento de secuencias resultó poco fructífero, algunos métodos de análisis de secuencia podrían ser explorados. En particular, se podrían realizar búsquedas a partir de perfiles de secuencias generados a partir de los *GDAVs* y homólogos.

Este enfoque puede presentar como obstáculo la acotada distribución taxonómica de algunos *GDAV*, la cual puede llevar a la obtención de alineamientos con poca variabilidad de secuencia, empobreciendo la capacidad de detección de homología remota. De forma complementaria, se pueden plantear al menos dos estrategias adicionales para el estudio a nivel de secuencias: utilización de bases de datos de interacción de proteínas (*e.g.* STRING database) y análisis de co-evolución de familias génicas. La información de interacción entre homólogos de los *GDAV* y otros productos génicos puede resultar en una inferencia de posible función, en tanto proteínas que interaccionan suelen participar en vías funcionales conjuntas. Este vínculo funcional lleva muchas veces a la co-evolución de secuencias, debido a la restricción evolutiva que impone la interacción funcional, por lo cual estudios filogenómicos podrían aportar también una línea de evidencia paralela en este sentido. Todas estas alternativas implican en última instancia la inferencia de homología basada en el alineamiento de secuencias. Este enfoque presenta limitación en casos donde se da el fenómeno denominado “erosión de la señal de secuencia”. En escenarios donde el tiempo o las tasas de divergencia son demasiado grandes, las identidades aminoacídicas disminuyen a un punto en el cual se da un aumento significativo de falsos positivos en la detección de homología [88]. Como alternativa, un modelo de estructura tridimensional confiable permite buscar similitudes estructurales (señal de homología), indetectables incluso por los más sensibles métodos basados en secuencia [236]. En este contexto, la revolucionaria aparición de métodos de predicción estructural de calidad sin precedentes como AlphaFold [237] está aumentando significativamente la disponibilidad de predicciones estructurales confiables. Esto trae aparejado un aumento de la capacidad de detectar homología, dado que la estructura se preserva más que la secuencia [238], lo cual se podría usar para ahondar en el estudio de *GDAVs* o genes diferencialmente expresados que presenten grandes cambios entre transiciones de estadíos. A la fecha se encuentran disponibles para uso público las predicciones de estructura terciaria para 48 proteomas (“*foldomas*”) de los tres dominios de la vida, incluyendo el de *S. mansoni* [239]. Este número se ha duplicado en menos de un año, y el aumento de estudios que emplean esta metodología crece a un ritmo acelerado [240], por lo cual es de esperar que los *foldomas* de otros tremátodos y cestodos se encuentren disponibles en el futuro próximo. De ser así se abriría la promisoriosa posibilidad de realizar alineamientos estructurales masivos contra estos y otros *foldomas* sin necesidad de computar previamente las predicciones estructurales (las cuales poseen a la fecha un costo computacional relativamente alto). Aproximaciones similares de alineamiento estructural masivo ya han sido implementadas en trabajos recientes con resultados prometedores (*e.g.* [241]). Resulta una perspectiva tentadora aplicar este enfoque al estudio de tremátodos donde no existen trabajos publicados a la fecha y donde existe una abundancia de secuencias candidatas a ser relevantes sin anotación funcional como las mencionadas anteriormente.

El estudio de los genes diferencialmente expresados entre estadíos intra-caracol también se vería beneficiado por esta aproximación. Dentro de estos genes se encuentran varios grupos funcionales ya reportados como relevantes en la biología de tremátodos, principalmente en *S. mansoni* y *F. gigantica*, sugiriendo que podrían jugar roles homólogos en *F. hepatica*.

Entre los mismos se encuentran genes codificantes para proteínas involucradas en la ejecución y coordinación del aparato motor de miracidios (señalización mediada por calcio, biosíntesis de transmisores), proteasas con posible rol en la infección del caracol, migración y digestión de proteínas dentro del mismo, o proteínas con rol antioxidativo con posible función en respuesta a sistema inmune. En algunos casos, como el de catepsinas L o legumainas, un análisis filogenético muestra que diferentes clados de secuencias tienden a expresar a lo largo del ciclo de vida. Nuestro análisis muestra que diferentes copias génicas se expresan en estadíos intra-caracol e intra-mamífero, presumiblemente cumpliendo un rol funcional específico en cada ambiente. Patrones similares se observan en familias multigénicas para las que se posee poca caracterización funcional. Una mayor caracterización de las mismas podría ayudar por tanto a comprender aún más la adaptación del parásito a su huésped caracol.

También resulta interesante la sobreexpresión en estadio intracaracol de grupos vinculados al metabolismo y el desarrollo morfológico, dos elementos fundamentales e interconectados para la supervivencia y proliferación asexual dentro de caracoles. El análisis realizado muestra que durante su pasaje por el huésped

caracol los parásitos poseen un metabolismo activo, coherente con su exacerbada replicación. A remarcar, las vías de salvataje de purinas tienden a expresarse en estos estadios, consistente con la noción de que este recurso debe ser obtenido de los huéspedes. Otro aspecto a remarcar del análisis realizado es que muestra que vías metabólicas asociadas a la utilización de lípidos se ven expresadas en este estadio, en contraste con los estadios intra-mamífero donde este recurso no es utilizado. Trabajos recientes en otros tremátodos son consistentes con este hallazgo, apuntando a que este aspecto es particular de los estadios intra-caracol. Respecto a los procesos de reproducción asexual y morfogénesis, los análisis realizados sugieren que genes de vías ampliamente conservadas en metazoarios como el programa regulador de la multipotencia y vías de determinación de los ejes antero-posterior y dorso-ventral sobreexpresan en estos estadios. Resultados similares han sido hallados en platelmintos de vida libre y parásitos, siendo estas vías regulatorias conservadas en el filo.

Uno de los aspectos a mejorar del enfoque tomado es que el análisis se vio fuertemente influenciado por lo ya reportado en la literatura. En este sentido deben tenerse en cuenta las limitaciones que presenta intrínsecamente la secuenciación transcriptómica global. Los resultados observados son en última instancia una ponderación del estado transcripcional de poblaciones heterogéneas de células presentes en diferentes proporciones. Esto lleva a que la interpretación de los datos tenga su mayor sustento en la presunción de homología con resultados experimentales reportados para el organismo estudiado u organismos filogenéticamente cercanos. Por otro lado, la aplicación de técnicas menos sesgadas a lo reportado en la literatura tiene un límite en el estudio de parásitos platelmintos ya que estos son organismos con genomas pobremente curados y con algunas vías metabólicas con funcionamiento atípico. Esto hace que algunos análisis *in silico* realizables en organismos modelo sean imposibles de trasladar, dificultándose en parte una aproximación no sesgada a los datos. El análisis de estos datos se puede enriquecer con las mismas perspectivas de trabajo que en los *GDAV* cuando se considera el caso de genes sin anotación funcional. En cuanto a algunas de las hipótesis planteadas, a medida que se realicen estudios transcriptómicos a resolución celular de los estadios intra-caracol en *F. hepatica* será posible obtener un panorama más claro respecto al grado de homología funcional con lo observado en otros organismos. Un caso paradigmático de este proceso en otro tremátodo es el de *S. mansoni*, donde algunas de las hipótesis planteadas en estudios de transcriptómica global fueron contrastadas y enriquecidas con la aparición de datos de *scRNA-seq*. La aplicación de técnicas de biología molecular como la interferencia de ARN también puede aportar información valiosa sobre el rol funcional de algunos de estos genes en *F. hepatica*, siendo una verificación necesaria y complementaria al planteado por las técnicas de secuenciación masiva.

Por último, una perspectiva de trabajo necesaria es continuar integrando datos provenientes de otros tremátodos como *S. mansoni* o *S. japonicum* a efecto de seguir acotando la lista de candidatos promisorios para estudio experimental. Esto permitirá contemplar tanto genes ya contemplados en la literatura como aquellos de los que no se posee información a la fecha. A la vez, permite definir de forma clara grupos sobre los que realizar análisis más pormenorizados, similares en espíritu a los que se esbozaron en esta tesis en el estudio de catepsinas, legumainas, SCP/TAPS e inhibidores de proteasas tipo Kunitz. Eventualmente esto podría incluir la anotación manual de copias génicas, la aproximación a la dinámica evolutiva a nivel molecular mediante el uso de estimadores clásicos como el dN/dS y el estudio de variantes en dominios de interés.

En conclusión, en este trabajo se aportó información relevante para comprender múltiples aspectos de la biología de los estadios larvarios de *F. hepatica* asociados a huésped caracol. El abordaje incluyó la caracterización de su diversidad transcripcional, la identificación de genes diferencialmente expresados entre estos estadios y la comparación de estos y el resto de los estadios del ciclo de vida del parásito. Se hipotetizó sobre el posible rol funcional de estos agentes moleculares a la luz de bibliografía de este y otros tremátodos, y se estudiaron los patrones de expresión génica para una selección de familias integrando herramientas de análisis filogenético y los datos transcriptómicos. Se logró por tanto cumplir con los objetivos generales y específicos planteados para esta maestría, planteándose a su vez posibles líneas de trabajo futuras para seguir ahondando en el estudio de estos estadios de desarrollo.

Capítulo 5

Anexo

Tabla 5.1: Datos transcriptómicos de *F. hepatica* empleadas en este estudio. Se marca con asteriscos a las muestras secuenciadas para este estudio.

Estadío de desarrollo	Código de secuenciación	Nombre de muestra
Huevo	ERR576955	EGG-1
Huevo	ERR577158	EGG-2
Huevo	ERR576954	EGG-3
Huevo	SRR2038743	EGG-4
Huevo	SRR2038787	EGG-5
Miracidia*	32377_7#1	MIR-1
Miracidia*	32377_7#8	MIR-2
Intracaracol (15 d.p.i.)*	32377_7#2	IS15-1
Intracaracol (15 d.p.i.)*	32377_7#3	IS15-2
Intracaracol (21 d.p.i.)*	F1	IS21-1
Intracaracol (21 d.p.i.)*	F4	IS21-2
Intracaracol (21 d.p.i.)*	F5	IS21-3
Intracaracol (30 d.p.i.)*	32377_7#4	IS30-1
Intracaracol (30 d.p.i.)*	32377_7#5	IS30-2
Intracaracol (30 d.p.i.)*	32377_7#6	IS30-3
Intracaracol (30 d.p.i.)*	32377_7#7	IS30-4
Metacercaria	ERR576958	MET-1
Metacercaria	ERR576959	MET-2
Metacercaria	ERR576965	MET-3
Metacercaria	ERR577156	MET-4
Metacercaria	ERR576964	MET-5
Juvenil desenquistado (1hr)	ERR576960	NEJ1-1
Juvenil desenquistado (1hr)	ERR576961	NEJ1-2
Juvenil desenquistado (1hr)	ERR576966	NEJ1-3
Juvenil desenquistado (3hr)	ERR576962	NEJ3-1
Juvenil desenquistado (3hr)	ERR576963	NEJ3-2
Juvenil desenquistado (3hr)	ERR576969	NEJ3-3
Juvenil desenquistado (24hr)	ERR576967	NEJ24-1
Juvenil desenquistado (24hr)	ERR576968	NEJ24-2
Juvenil	ERR576956	JUV-1
Juvenil	ERR577159	JUV-2
Juvenil	ERR576957	JUV-3
Adulto	ERR577157	AD-1
Adulto	ERR577160	AD-2
Adulto	SRR2038730	AD-3
Adulto	SRR2038894	AD-4
Adulto	SRR2038734	AD-5
Adulto	SRR2038745	AD-6

Tabla 5.2: Número de transcriptos ensamblados por StringTie a diferentes niveles mínimos de cobertura promedio.

Estadío	Núm. genes detectados	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5	Muestra 6
EGG	7664	11241	10996	11179.0	8083.0	7096.0	-
MIR	9041	16247	16222	-	-	-	-
IS15	9743	17038	14266	-	-	-	-
IS21	9555	15917	11628	13522.0	-	-	-
IS30	9798	15499	11466	13048.0	13609.0	-	-
MET	8783	15025	15018	16120.0	15317.0	15172.0	-
NEJ1	8654	13462	13518	15716.0	-	-	-
NEJ3	8806	14486	14544	15378.0	-	-	-
NEJ24	8877	13452	15762	-	-	-	-
JUV	9520	19562	15533	19638.0	-	-	-
AD	8835	15438	11966	9157.0	8558.0	10204.0	9204.0

Tabla 5.3: Número de transcriptos reportados con una cobertura promedio mínima de 10x. Se desglosa los mismos según su clasificación en el esquema empleado por gffcompare.

Estadío de desarrollo	Código de muestra	Nombre de muestra	=	c	e	i	j	k	m	n	o	p	s	u	x	y
Intracaracol (15 d.p.i.)*	32377_7#2	IS15-1	1910	904	84	386	7306	66	249	486	354	30	2	2922	181	66
Intracaracol (15 d.p.i.)*	32377_7#3	IS15-2	1971	915	100	400	7932	72	263	513	391	32	2	3011	215	91
Intracaracol (21 d.p.i.)*	F1	IS21-1	1362	974	112	252	4267	16	160	231	218	33	11	1464	536	39
Intracaracol (21 d.p.i.)*	F4	IS21-2	1255	1524	150	353	4601	13	184	338	274	58	16	2092	957	50
Intracaracol (21 d.p.i.)*	F5	IS21-3	1287	1044	114	250	3925	13	150	249	216	30	13	1557	640	36
Intracaracol (30 d.p.i.)*	32377_7#4	IS30-1	1903	952	106	414	7351	77	262	517	364	33	5	3068	215	50
Intracaracol (30 d.p.i.)*	32377_7#5	IS30-2	1701	961	101	381	6264	45	225	387	332	23	3	2417	204	59
Intracaracol (30 d.p.i.)*	32377_7#6	IS30-3	1813	1060	114	417	7070	59	242	484	354	32	4	3016	220	62
Intracaracol (30 d.p.i.)*	32377_7#7	IS30-4	1796	1078	103	412	6879	57	239	484	351	33	3	2987	214	61
Miracidia*	32377_7#1	MIR-1	1668	980	107	331	6868	68	197	446	305	39	2	2415	169	48
Miracidia*	32377_7#8	MIR-2	1857	776	91	370	7451	102	245	547	346	33	1	3159	213	64

Tabla 5.4: Estadísticas de procesamiento y mapeo de todas las muestras de *F. hepatica* utilizadas para este estudio.

Estadio de desarrollo	Código de secuenciación	Nombre de muestra	Pares de lecturas	Pares de lecturas trimeadas	Pares de lecturas trimeadas (%)	Tasa de alineamiento (HiSat2)	Tasa de alineamiento (kallisto)
Huevo	ERR576955	EGG-1	37008210	36147568	97.67446	-	92.9
Huevo	ERR577158	EGG-2	42363722	41942946	99.00675	-	92.7
Huevo	ERR576954	EGG-3	37187237	36418253	97.93213	-	92.9
Huevo	SRR2038743	EGG-4	22580803	21302048	94.33698	-	88.1
Huevo	SRR2038787	EGG-5	34704635	13238458	38.14608	-	78.7
Miracidia*	32377_7#1	MIR-1	35600000	35500000	99.71910	83.58	81.1
Miracidia*	32377_7#8	MIR-2	39700000	39700000	100.00000	88.73	85.8
Intracaracol (15 d.p.i.)*	32377_7#2	IS15-1	34600000	34600000	100.00000	78.05	76.0
Intracaracol (15 d.p.i.)*	32377_7#3	IS15-2	46500000	46400000	99.78495	79.47	78.6
Intracaracol (21 d.p.i.)*	F1	IS21-1	11000000	8420000	76.54545	85.53	83.2
Intracaracol (21 d.p.i.)*	F4	IS21-2	16300000	12500000	76.68712	86.97	85.4
Intracaracol (21 d.p.i.)*	F5	IS21-3	11900000	8520000	71.59664	83.24	82.0
Intracaracol (30 d.p.i.)*	32377_7#4	IS30-1	39000000	38900000	99.74359	73.82	73.0
Intracaracol (30 d.p.i.)*	32377_7#5	IS30-2	35700000	35600000	99.71989	73.94	76.5
Intracaracol (30 d.p.i.)*	32377_7#6	IS30-3	40600000	40600000	100.00000	80.08	79.8
Intracaracol (30 d.p.i.)*	32377_7#7	IS30-4	42900000	42900000	100.00000	72.44	72.2
Metacercaria	ERR576958	MET-1	44258893	43527925	98.34843	-	84.2
Metacercaria	ERR576959	MET-2	43919803	43097533	98.12779	-	84.2
Metacercaria	ERR576965	MET-3	66662513	65594946	98.39855	-	85.5
Metacercaria	ERR577156	MET-4	42444892	42068155	99.11241	-	86.6
Metacercaria	ERR576964	MET-5	66495596	65384808	98.32953	-	82.1
Juvenil desenquistado (1hr)	ERR576960	NEJ1-1	50270024	49499961	98.46815	-	85.5
Juvenil desenquistado (1hr)	ERR576961	NEJ1-2	50106428	49205360	98.20169	-	85.5
Juvenil desenquistado (1hr)	ERR576966	NEJ1-3	83454099	81801450	98.01969	-	88.8
Juvenil desenquistado (3hr)	ERR576962	NEJ3-1	45746545	45007770	98.38507	-	85.3
Juvenil desenquistado (3hr)	ERR576963	NEJ3-2	45540680	44685303	98.12173	-	85.3
Juvenil desenquistado (3hr)	ERR576969	NEJ3-3	72247393	71058460	98.35436	-	88.5
Juvenil desenquistado (24hr)	ERR576967	NEJ24-1	63496634	58870827	92.71488	-	89.1
Juvenil desenquistado (24hr)	ERR576968	NEJ24-2	83491240	82171313	98.41908	-	87.7
Juvenil	ERR576956	JUV-1	43048853	42371311	98.42611	-	65.1
Juvenil	ERR577159	JUV-2	59062081	58422363	98.91687	-	82.0
Juvenil	ERR576957	JUV-3	42792167	42020563	98.19686	-	65.0
Adulto	ERR577157	AD-1	47117937	46654845	99.01716	-	78.5
Adulto	ERR577160	AD-2	41650959	41176685	98.86131	-	77.9
Adulto	SRR2038730	AD-3	19931404	18631101	93.47611	-	73.4
Adulto	SRR2038894	AD-4	65734911	19265255	29.30749	-	59.6
Adulto	SRR2038734	AD-5	24876643	23483311	94.39904	-	73.8
Adulto	SRR2038745	AD-6	23961106	10364902	43.25719	-	46.7

Tabla 5.5: Número de transcritos ensamblados con StringTie para muestras de *F. hepatica* secuenciadas para este estudio. Se discrimina según cobertura promedio mínima de los mismos.

Estadio de desarrollo	Código de muestra	Nombre de muestra	Núm. total de transcritos	Núm. transcritos (coverage $\geq 1x$)	Núm. transcritos (coverage $\geq 10x$)	Núm. transcritos (coverage $\geq 100x$)
Miracidia	32377_7#1	MIR-1	27191	26697	13643	4024
Miracidia	32377_7#8	MIR-2	30118	29301	15255	4538
Intracaracol (15 d.p.i.)	32377_7#2	IS15-1	28970	28290	14946	4065
Intracaracol (15 d.p.i.)	32377_7#3	IS15-2	29841	29187	15908	5042
Intracaracol (21 d.p.i.)	F1	IS21-1	21774	21376	9675	1725
Intracaracol (21 d.p.i.)	F4	IS21-2	25469	25120	11865	2558
Intracaracol (21 d.p.i.)	F5	IS21-3	21435	21072	9524	1852
Intracaracol (30 d.p.i.)	32377_7#4	IS30-1	30473	29860	15317	3772
Intracaracol (30 d.p.i.)	32377_7#5	IS30-2	26863	26331	13103	3170
Intracaracol (30 d.p.i.)	32377_7#6	IS30-3	29925	29381	14947	3840
Intracaracol (30 d.p.i.)	32377_7#7	IS30-4	29284	28803	14697	3592

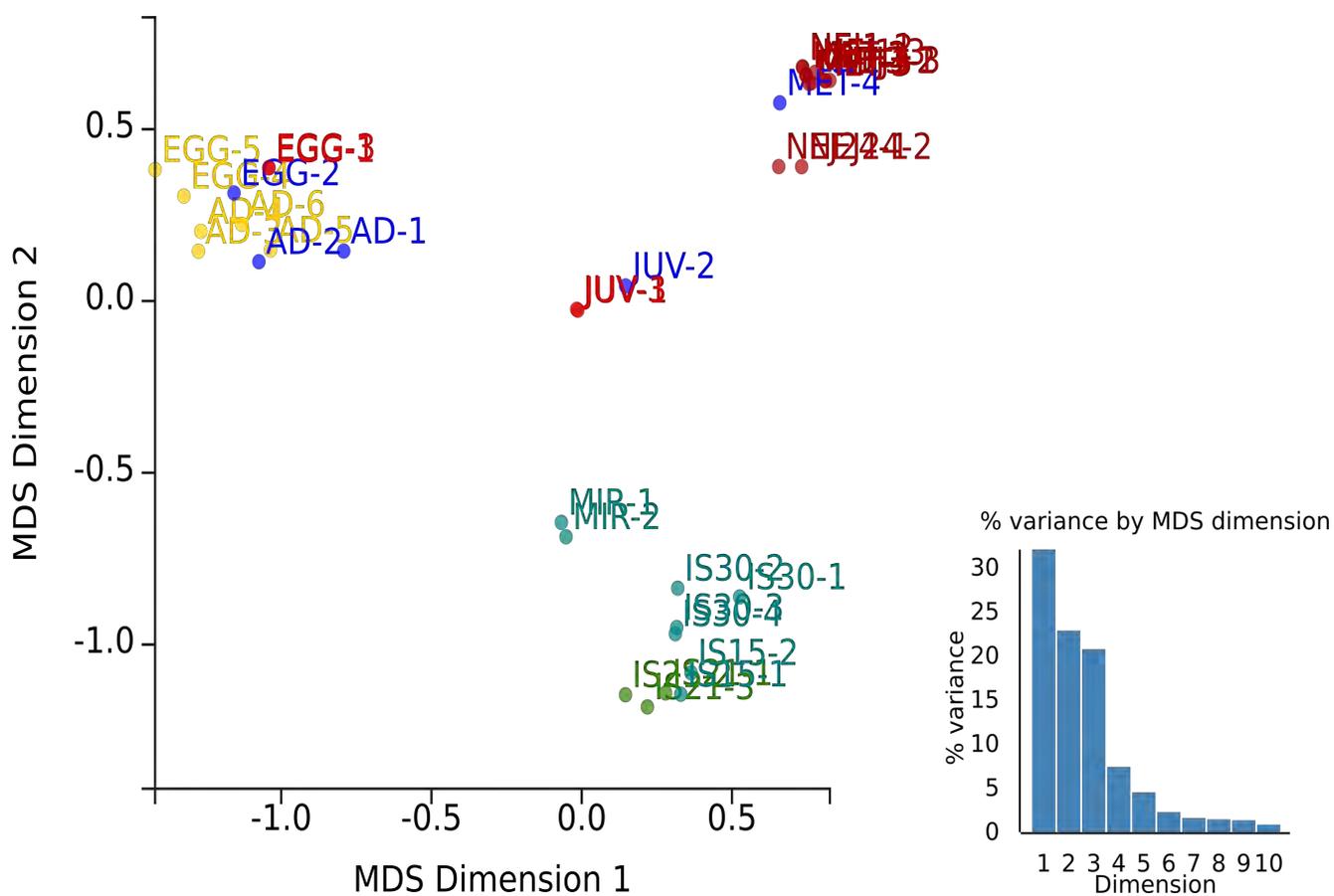


Figura 5.1: Análisis de escalado multidimensional a nivel de expresión génica para muestras de estadios de desarrollo de *F. hepatica* analizadas en esta tesis. Se muestran los tres dimensiones principales, y se colorea a las mismas según estudio de secuenciación.

Tabla 5.6: Porcentaje para categorías de mapeo según esquema de clasificación de HISAT2 para pares de reads de muestras de miracidia y estadios intracaracol mapeados contra el genoma de *F. hepatica*.

ID	Estadio de desarrollo	Sample	Nombre de muestra	Aigned concordantly zero times	Aigned concordantly one time	Aigned concordantly multiple times	Overall alignment rate
MIR	Miracidia	32377_7#1	MIR-1	21.34	75.30	3.35	83.58
MIR	Miracidia	32377_7#8	MIR-2	16.77	80.02	3.21	88.73
IS15	Intracaracol (15 d.p.i.)	32377_7#2	IS15-1	27.15	69.25	3.60	78.05
IS15	Intracaracol (15 d.p.i.)	32377_7#3	IS15-2	26.37	69.97	3.66	79.47
IS21	Intracaracol (21 d.p.i.)	F1	IS21-1	17.31	75.68	7.01	85.53
IS21	Intracaracol (21 d.p.i.)	F4	IS21-2	16.93	78.23	4.84	86.97
IS21	Intracaracol (21 d.p.i.)	F5	IS21-3	19.52	75.35	5.13	83.24
IS30	Intracaracol (30 d.p.i.)	32377_7#4	IS30-1	31.81	59.61	8.58	73.82
IS30	Intracaracol (30 d.p.i.)	32377_7#5	IS30-2	33.21	53.06	13.73	73.94
IS30	Intracaracol (30 d.p.i.)	32377_7#6	IS30-3	25.13	61.55	13.32	80.08
IS30	Intracaracol (30 d.p.i.)	32377_7#7	IS30-4	32.63	53.93	13.44	72.44

Tabla 5.7: Porcentaje para categorías de mapeo según esquema de clasificación de HISAT2 para pares de reads de muestras de miracidia y estadios intracaracol mapeados contra genoma del caracol *L. stagnalis*. Se alinean reads que no alinearon contra el genoma de *F. hepatica*.

ID	Estadio de desarrollo	Sample	Nombre de muestra	Aigned concordantly zero times	Aigned concordantly one time	Aigned concordantly multiple times	Overall alignment rate
MIR	Miracidia	32377_7#1	MIR-1	99.35	0.63	0.02	3.30
MIR	Miracidia	32377_7#8	MIR-2	99.19	0.79	0.02	3.80
IS15	Intracaracol (15 d.p.i.)	32377_7#2	IS15-1	93.53	6.09	0.39	12.73
IS15	Intracaracol (15 d.p.i.)	32377_7#3	IS15-2	98.90	1.05	0.06	4.13
IS21	Intracaracol (21 d.p.i.)	F1	IS21-1	76.77	22.95	0.28	48.86
IS21	Intracaracol (21 d.p.i.)	F4	IS21-2	79.37	20.59	0.04	42.90
IS21	Intracaracol (21 d.p.i.)	F5	IS21-3	75.30	24.52	0.18	48.23
IS30	Intracaracol (30 d.p.i.)	32377_7#4	IS30-1	98.45	1.44	0.11	4.64
IS30	Intracaracol (30 d.p.i.)	32377_7#5	IS30-2	98.18	1.70	0.12	5.46
IS30	Intracaracol (30 d.p.i.)	32377_7#6	IS30-3	98.72	1.21	0.07	4.30
IS30	Intracaracol (30 d.p.i.)	32377_7#7	IS30-4	98.70	1.22	0.09	3.71

5.0.1 Estudio de isoformas reportadas

Del análisis de isoformas reportadas surge que una porción considerable (~50%) de los transcritos son clasificados por gffcompare como isoformas de la categoría “j”. Esta categoría engloba a transcritos con más de un exón que solapan en al menos un sitio de empalme con las isoformas canónicas de la anotación empleada.

Dentro de la definición de esta categoría se encuentran aquellos transcritos con variaciones en los extremos 5’ y 3’ UTR de los transcritos ensamblados. Es de notar que la población de transcritos secuenciada tenderá a variar en estas regiones, ya que las mismas son proclives a la degradación enzimática y al decaimiento de la calidad de secuenciado. Transcritos que solo varíen en esta región serán considerados isoformas de tipo “j” por el algoritmo de comparación de gffcompare, llevando a una sobreestimación de las isoformas biológicamente relevantes presentes en la población de transcritos secuenciados.

Como una primera aproximación para evaluar si esta es una explicación razonable al número observado de isoformas catalogadas en esta categoría, se analizó la variación de largos en los extremos 5’ y 3’ UTR de los transcritos ensamblados respecto a los transcritos canónicos con los que corresponden. Como punto de corte arbitrario, se observó el porcentaje de transcritos de esta categoría que presentan diferencias de, como máximo, 60pb en sus extremos 5’ o 3’ respecto al modelo génico de referencia. A continuación se muestra la distribución de diferencias en ambos extremos

El porcentaje de transcritos que cumple con la premisa de diferir en, a lo sumo, 60 pb en ambos extremos es de 9,2%. La amplia mayoría de las isoformas “j” poseen variaciones más marcadas o bien en sus extremos 5’ o bien en los 3’, así como en sus regiones codificantes, sugiriendo que la anotación artefactual de isoformas “j” debido a variaciones menores en los extremos de los transcritos es un fenómeno minoritario dentro de la diversidad transcripcional observada.

5.0.2 Relación entre media y varianza de la expresión de un gen

En el manual de edgeR [242] se desarrolla el siguiente modelo para la relación entre media y varianza de la expresión de un gen:

Los perfiles de RNA-seq están formados por n muestras. Sea π_{gi} la fracción de todos los fragmentos de ADNc en la i -ésima muestra que se originan del gen g . Si con G se denota el número total de genes, se tiene entonces que $\sum_{g=1}^G \pi_{gi} = 1$ para cada muestra. Sea $\sqrt{\phi_g}$ es el coeficiente de variación (desviación estándar σ dividida por la media μ_{gi}) de π_{gi} entre las replicas i . Denotamos al total de reads mapeados para la muestra i con N_i y al número que mapean al g -ésimo gen con y_{gi} . Se tiene que

$$E(y_{gi}) = \mu_{gi} = N_i \pi_{gi}$$

Si se asume que los conteos y_{gi} siguen una distribución de Poisson para corridas repetidas de secuenciación para una misma muestra de ARN, es sabido que una fórmula conocida para la varianza de una distribución mixta es

$$\text{var}(y_{gi}) = E_{\pi}[\text{var}(y|\pi)] + \text{var}_{\pi}[E(y|\pi)] = \mu_{gi} + \phi_g \mu_{gi}^2$$

(...)

El modelo implica una relación cuadrática positiva entre ambas variables, acorde a lo observado al graficar ambas variables.

Capítulo 6

Bibliografía

- 1 Hickman C. Integrated principles of zoology. 2020
- 2 Galaktionov KV, Dobrovolskij AA. The biology and evolution of trematodes : An essay on the biology, morphology, life cycles, transmission, and evolution of digenetic trematodes. Kluwer Academic Publishers; 2003.
- 3 Nyindo M, Lukambagire AH. Fascioliasis: An ongoing zoonotic trematode infection. *BioMed Research International*. 2015;2015. DOI: 10.1155/2015/786195
- 4 Cwiklinski K, O'Neill SM, Donnelly S, Dalton JP. A prospective view of animal and human fasciolosis. *Parasite Immunology*. 2016 Sep;38(9):558.
- 5 Dittmar K, Teegen WR. The presence of *fasciola hepatica* (liver-fluke) in humans and cattle from a 4,500 year old archaeological site in the saale-unstrut valley, germany. *Memorias do Instituto Oswaldo Cruz*. 2003;98 Suppl 1(SUPPL. 1):141–3.
- 6 Organization WH. Neglected tropical diseases: Fascioliasis Available from: <https://www.who.int/news-room/questions-and-answers/item/q-a-on-fascioliasis>
- 7 Bargues MD, Gayo V, Sanchis J, Artigas P, Khoubbane M, Birriel S, et al. DNA multigene characterization of *fasciola hepatica* and *lymnaea neotropica* and its fascioliasis transmission capacity in uruguay, with historical correlation, human report review and infection risk analysis. *PLoS Neglected Tropical Diseases*. 2017 Feb;11(2). DOI: 10.1371/JOURNAL.PNTD.0005352
- 8 Carmona C, Tort JF. Fasciolosis in south america: Epidemiology and control challenges. *Journal of helminthology*. 2017 Mar;91(2):99–109.
- 9 Costa RA da, Corbellini LG, Castro-Janer E, Riet-Correa F. 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*. 2019;49(11). DOI: 10.1016/j.ijpara.2019.06.005
- 10 Dalton JP. Fasciolosis. CAB International; 2022.
- 11 Dawes B. Penetration of the liver-fluke, *fasciola hepatica* into the snail, *limnaea truncatula*. *Nature* 1959 184:4695. 1959;184(4695):1334–5.
- 12 Kjøie M, Christensen NØ, Nansen P. Stereoscan studies of eggs, free-swimming and penetrating miracidia and early sporocysts of *fasciola hepatica*. *Zeitschrift fur Parasitenkunde (Berlin, Germany)*. 1976 Jan;51(1):79–90.
- 13 Lu XT, Gu QY, Limpanont Y, Song LG, Wu ZD, Okanurak K, et al. Snail-borne parasitic diseases: An update on global epidemiological distribution, transmission interruption and control methods. *Infectious Diseases of Poverty*. 2018 Apr;7(1):1–16.

- 14 Wang B, Collins JJ, Newmark PA. Functional genomic characterization of neoblast-like stem cells in larval schistosoma mansoni. *eLife*. 2013 Jul;2013(2). DOI: 10.7554/ELIFE.00768
- 15 Cheng S, Zhu B, Luo F, Lin X, Sun C, You Y, et al. Comparative transcriptome profiles of schistosoma japonicum larval stages: Implications for parasite biology and host invasion. *PLOS Neglected Tropical Diseases*. 2022 Jan;16(1):e0009889.
- 16 Zhang XX, Cwiklinski K, Hu RS, Zheng WB, Sheng ZA, Zhang FK, et al. Complex and dynamic transcriptional changes allow the helminth fasciola gigantica to adjust to its intermediate snail and definitive mammalian hosts. *BMC Genomics*. 2019 DOI: 10.1186/s12864-019-6103-5
- 17 Nesterenko MA, Starunov VV, Shchenkov SV, Maslova AR, Denisova SA, Granovich AI, et al. Molecular signatures of the rediae, cercariae and adult stages in the complex life cycles of parasitic flatworms (digenea: Psilostomatidae). *Parasites and Vectors*. 2020 Dec;13(1):1–21.
- 18 Boyunaga H, Schmitz MG, Brouwers JF, Hellemond JJV, Tielens AG. Fasciola hepatica miracidia are dependent on respiration and endogenous glycogen degradation for their energy generation. *Parasitology*. 2001;122(Pt 2):169–73.
- 19 Tielens AG, Pas FA van de, Heuvel JM van den, Bergh SG van den. The aerobic energy metabolism of schistosoma mansoni miracidia. *Molecular and biochemical parasitology*. 1991;46(1):181–4.
- 20 Kianifard L, Yakhchali M, Imani M. Serine proteases activity in miracidia of fasciola hepatica and effects of chemical and herbal inhibitors. *Veterinary Research Forum*. 2021;12(3):291.
- 21 Loker ES, Hofkin BV. *Parasitology: A conceptual approach*. 2015.
- 22 Brockhurst MA, Chapman T, King KC, Mank JE, Paterson S, Hurst GD. Running with the red queen: The role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences*. 2014 Dec;281(1797). DOI: 10.1098/RSPB.2014.1382
- 23 Curwen RS, Ashton PD, Sundaralingam S, Wilson RA. Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Molecular & cellular proteomics : MCP*. 2006 May;5(5):835–44.
- 24 Hewitson JP, Grainger JR, Maizels RM. Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. *Molecular and Biochemical Parasitology*. 2009 DOI: 10.1016/j.molbiopara.2009.04.008
- 25 Al-Khalaifah H, Al-Nasser A. Immune response of molluscs. *Molluscs*. 2018 Nov DOI: 10.5772/INTECHOPEN.81778
- 26 Pila EA, Li H, Hambrook JR, Wu X, Hanington PC. Schistosomiasis from a snail's perspective: Advances in snail immunity. *Trends in parasitology*. 2017 Nov;33(11):845–57.
- 27 Hambrook JR, Hanington PC. Immune evasion strategies of schistosomes. *Frontiers in Immunology*. 2021 Feb;11:3820.
- 28 Seppälä O, Walser JC, Cereghetti T, Seppälä K, Salo T, Adema CM. Transcriptome profiling of lymnaea stagnalis (gastropoda) for ecoimmunological research. *BMC Genomics*. 2021 Dec;22(1):1–16.
- 29 Georgieva K, Yoneva A, Popov I, Mizinska-Boevska Y, Stoitsova SR. Lectin-binding properties of the surface of fasciola hepatica sporocysts. 2005.
- 30 Georgieva K, Yoneva A, Mizinska-Boevska Y. Lectin binding characteristics of fasciola hepatica rediae. *Comptes Rendus de L'Academie Bulgare des Sciences*. 2007 Feb;60:315–8.
- 31 Georgieva K, Georgieva S, Mizinska Y, Stoitsova SR. Fasciola hepatica miracidia: Lectin binding and stimulation of in vitro miracidium-to-sporocyst transformation. *Acta parasitologica*. 2012 Mar;57(1):46–52.

- 32 Coustau C, Gourbal B, Duval D, Yoshino TP, Adema CM, Mitta G. Advances in gastropod immunity from the study of the interaction between the snail *Biomphalaria glabrata* and its parasites: A review of research progress over the last decade. *Fish & shellfish immunology*. 2015 Sep;46(1):5–16.
- 33 Alba A, Vázquez AA, Sánchez J, Gourbal B. Immunological resistance of *Pseudosuccinea columella* snails from Cuba to *Fasciola hepatica* (Trematoda) infection: What we know and where we go on comparative molecular and mechanistic immunobiology, ecology and evolution. *Frontiers in Immunology*. 2022 Jan;0:28.
- 34 Newmark PA, Wang Y, Chong T. Germ cell specification and regeneration in planarians. *Cold Spring Harbor symposia on quantitative biology*. 2008;73:573–81.
- 35 Baguna J, Saló E, Auladell C. Regeneration and pattern formation in planarians. III. That neoblasts are totipotent stem cells and the cells. *Development*. 1989 Sep;107(1):77–86.
- 36 Juliano CE, Swartz SZ, Wessel GM. A conserved germline multipotency program. *Development*. 2010 Dec;137(24):4113–26.
- 37 Önal P, Grün D, Adamidi C, Rybak A, Solana J, Mastrobuoni G, et al. Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *EMBO Journal*. 2012 Jun;31(12):2755–69.
- 38 Pan SC tung. The fine structure of the miracidium of *Schistosoma mansoni*. *Journal of Invertebrate Pathology*. 1980 Nov;36(3):307–72.
- 39 Whitfield PJ, Evans NA. Parthenogenesis and asexual multiplication among parasitic platyhelminths. *Parasitology*. 1983;86 (Pt 4)(4):121–60.
- 40 III JJC, Wang B, Lambrus BG, Tharp ME, Iyer H, Newmark PA. Adult somatic stem cells in the human parasite *Schistosoma mansoni*. *Nature*. 2013 DOI: 10.1038/nature11924
- 41 Collins JJ, Wendt GR, Iyer H, Newmark PA. Stem cell progeny contribute to the schistosome host-parasite interface. *eLife*. 2016 DOI: 10.7554/eLife.12473
- 42 Wang D, Korhonen PK, Gasser RB, Young ND. Improved genomic resources and new bioinformatic workflow for the carcinogenic parasite *Clonorchis sinensis*: Biotechnological implications. *Biotechnology Advances*. 2018 Jul;36(4):894–904.
- 43 Li P, Sarfati DN, Xue Y, Yu X, Tarashansky AJ, Quake SR, et al. Single-cell analysis of *Schistosoma mansoni* identifies a conserved genetic program controlling germline stem cell fate. *Nature Communications* 2021 12:1. 2021 Jan;12(1):1–12.
- 44 McCusker P, McVeigh P, Rathinasamy V, Toet H, McCammick E, O'Connor A, et al. Stimulating neoblast-like cell proliferation in juvenile *Fasciola hepatica* supports growth and progression towards the adult phenotype in vitro. *PLOS Neglected Tropical Diseases*. 2016 Sep;10(9):e0004994.
- 45 Koziol U, Rauschendorfer T, Rodríguez LZ, Krohne G, Brehm K. The unique stem cell system of the immortal larva of the human parasite *Echinococcus multilocularis*. *EvoDevo*. 2014 Mar;5(1):1–23.
- 46 Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. FastQC: A quality control tool for high throughput sequence data. 2010 Jan
- 47 Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016 Jun;32(19):3047–8.
- 48 Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Apr;30(15):2114–20.

- 49 Cwiklinski K, Dalton JP, Dufresne PJ, Course JL, Williams DJL, 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 Biology*. 2015 Apr;16(1). DOI: 10.1186/s13059-015-0632-2
- 50 Choi YJ, Fontenla S, Fischer PU, Le TH, Costábile A, Blair D, et al. Adaptive radiation of the flukes of the family fasciolidae inferred from genome-wide comparisons of key species. *Molecular Biology and Evolution*. 2020 Jan;37(1):84.
- 51 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 Genetics*. 2017 DOI: 10.1371/journal.pgen.1006537
- 52 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nature Biotechnology*. 2019 Aug;37(8):907–15.
- 53 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. *Bioinformatics*. 2009 Aug;25(16):2078–9.
- 54 Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from rna-seq reads. *Nature Biotechnology*. 2015 Feb;33(3):290–5.
- 55 Pertea G, Pertea M. GFF utilities: GffRead and gffcompare. *F1000Research*. 2020 Apr;9:304.
- 56 Baruzzo G, Hayer KE, Kim EJ, Camillo BD, Fitzgerald GA, Grant GR. Simulation-based comprehensive benchmarking of rna-seq aligners. *Nature Methods* 2016 14:2. 2016 Dec;14(2):135–9.
- 57 Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic rna-seq quantification. *Nature Biotechnology*. 2016 Apr;34(5):525–7.
- 58 Sonesson C, Love MI, Robinson MD. Differential analyses for rna-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*. 2015 Dec;4:1521.
- 59 Robinson MD, McCarthy DJ, Smyth GK. EdgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
- 60 McCarthy, J. D, Chen, Yunshun, Smyth, K. G. Differential expression analysis of multifactor rna-seq experiments with respect to biological variation. *Nucleic Acids Research*. 2012;40(10):4288–97.
- 61 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of rna-seq data. *Genome Biology* 2010 11:3. 2010 Mar;11(3):1–9.
- 62 Abu-Jamous B, Kelly S. Clust: Automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome biology*. 2018 Oct;19(1). DOI: 10.1186/S13059-018-1536-8
- 63 Jaccard P. THE distribution of the flora in the alpine zone.1. *New Phytologist*. 1912 Feb;11(2):37–50.
- 64 Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Mering CV, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*. 2017 DOI: 10.1093/molbev/msx148
- 65 Jensen LJ, Julien P, Kuhn M, Mering C von, Muller J, Doerks T, et al. EggNOG: Automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research*. 2008 DOI: 10.1093/nar/gkm796
- 66 Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*. 2016 DOI: 10.1093/nar/gkv1248
- 67 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. *Nature Genetics*. 2000 DOI: 10.1038/75556

- 68 Kanehisa M. The kegg databases at genomenet. *Nucleic Acids Research*. 2002 DOI: 10.1093/nar/30.1.42
- 69 Sonnhammer EL, Eddy SR, Durbin R. Pfam: A comprehensive database of protein domain families based on seed alignments
- 70 El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The pfam protein families database in 2019. *Nucleic Acids Research*. 2019 Jan;47(D1):D427–32.
- 71 Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the cog database. *Nucleic Acids Research*. 2015 Jan;43(Database issue):D261.
- 72 Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The cog database: An updated version includes eukaryotes. *BMC Bioinformatics*. 2003 Sep;4:41.
- 73 Makarova KS, Wolf YI, Koonin EV. Archaeal clusters of orthologous genes (arCOGs): An update and application for analysis of shared features between thermococcales, methanococcales, and methanobacteriales. *Life*. 2015 Mar;5(1):818.
- 74 Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The interpro protein families and domains database: 20 years on. *Nucleic Acids Research*. 2021 Jan;49(D1):D344–54.
- 75 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*. 2014 DOI: 10.1093/bioinformatics/btu031
- 76 Waskom ML. Seaborn: Statistical data visualization. *Journal of Open Source Software*. 2021 Apr;6(60):3021.
- 77 Simillion C, Liechti R, Lischer HE, Ioannidis V, Bruggmann R. Avoiding the pitfalls of gene set enrichment analysis with setrank. *BMC Bioinformatics*. 2017 Mar;18(1):1–14.
- 78 Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztröcy AW, Naldi A, et al. GOATOOLS: A python library for gene ontology analyses. *Scientific Reports*. 2018 DOI: 10.1038/s41598-018-28948-z
- 79 Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995 Jan;57(1):289–300.
- 80 Brionne A, Juanchich A, Hennequet-Antier C. ViSEAGO: A bioconductor package for clustering biological functions using gene ontology and semantic similarity. *BioData Mining*. 2019 Aug;12(1):1–13.
- 81 Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of go terms. *Bioinformatics*. 2007 May;23(10):1274–81.
- 82 Peterson H, Kolberg L, Raudvere U, Kuzmin I, Vilo J. Gprofiler2 – an r package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* 2020 9:709. 2020 Nov;9:709.
- 83 Powell D. An interactive web-tool for rna-seq analysis. 2019 DOI: 10.5281/ZENODO.3501067
- 84 Emms DM, Kelly S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*. 2019 Nov;20(1):1–14.
- 85 Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform
- 86 Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010 Mar;5(3). DOI: 10.1371/JOURNAL.PONE.0009490
- 87 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990;215(3):403–10.

- 88 Rost B. Twilight zone of protein sequence alignments. *Protein Engineering, Design and Selection*. 1999 Feb;12(2):85–94.
- 89 J. Notredame H C., Notredame C, Higgins DG, Heringa J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*. 2000 DOI: 10.1006/jmbi.2000.4042
- 90 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Haeseler A von, et al. IQ-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*. 2020 May;37(5):1530–4.
- 91 Kalyaanamoorthy S, Minh BQ, Wong TK, Haeseler AV, Jermin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature methods*. 2017 May;14(6):587–9.
- 92 Hoang DT, Chernomor O, Haeseler AV, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*. 2018 DOI: 10.1093/molbev/msx281
- 93 Hellemans J, Mortier G, Paepe AD, Speleman F, Vandesompele J. QBase relative quantification framework and software for management and automated analysis of real-time quantitative pcr data. *Genome Biology*. 2008 Feb;8(2):1–14.
- 94 Stanton KA, Edger PP, Puzey JR, Kinser T, Cheng P, Vernon DM, et al. A whole-transcriptome approach to evaluating reference genes for quantitative gene expression studies: A case study in *mimulus*. *G3: Genes, Genomes, Genetics*. 2017 Apr;7(4):1085–95.
- 95 Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, et al. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*. 2019 Jan;47(D1):D807–11.
- 96 Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using diamond. *Nature Methods* 2021 18:4. 2021 Apr;18(4):366–8.
- 97 Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*. 2016 DOI: 10.1093/molbev/msw046
- 98 Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*. 1998 Mar;8(3):186–94.
- 99 Cwiklinski K, Jewhurst H, McVeigh P, Barbour T, Maule AG, Tort J, et al. 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*. 2018 Apr;17(4):792–809.
- 100 Cwiklinski K, Robinson MW, Donnelly S, Dalton JP. Complementary transcriptomic and proteomic analyses reveal the cellular and molecular processes that drive growth and development of *fasciola hepatica* in the host liver. *BMC Genomics*. 2021 Dec;22(1):1–16.
- 101 Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. OrthoDB: The hierarchical catalog of eukaryotic orthologs in 2011. *Nucleic Acids Research*. 2011 Jan;39(Database issue):D283.
- 102 Tielens AG, Heuvel JM van den, Bergh SG van den. The energy metabolism of *fasciola hepatica* during its development in the final host. *Molecular and biochemical parasitology*. 1984;13(3):301–7.
- 103 Matera AG, Wang Z. A day in the life of the spliceosome. *Nature reviews Molecular cell biology*. 2014 Feb;15(2):108.
- 104 Hir HL, Saulière J, Wang Z. The exon junction complex as a node of post-transcriptional networks. *Nature Reviews Molecular Cell Biology* 2015 17:1. 2015 Dec;17(1):41–54.
- 105 Sobalaflovsky A. Inverted and convoluted matrices strike back: Bluebeard against the lethal love. The world is still round and of ricot. *Journal of Mocklecular Evolution*. 1992;23(45):223–332.

- 106 Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, Reed R. The protein aly links pre-messenger-rna splicing to nuclear export in metazoans. *Nature* 2000 407:6802. 2000 Sep;407(6802):401–5.
- 107 Stäßer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, et al. TREX is a conserved complex coupling transcription with messenger rna export. *Nature*. 2002 Apr;417(6886):304–8.
- 108 Dang F, Nie L, Wei W. Ubiquitin signaling in cell cycle control and tumorigenesis. *Cell Death & Differentiation* 2020 28:2. 2020 Oct;28(2):427–38.
- 109 Pols MS, Klumperman J. Trafficking and function of the tetraspanin cd63. *Experimental cell research*. 2009 May;315(9):1584–92.
- 110 Nanut MP, Sabotič J, Jewett A, Kos J. Cysteine cathepsins as regulators of the cytotoxicity of nk and t cells. *Frontiers in Immunology*. 2014;5(DEC):616.
- 111 Tort J, Brindley PJ, Knox D, Wolfe KH, Dalton JP. Proteinases and associated genes of parasitic helminths. *Advances in parasitology*. 1999;43:161–266.
- 112 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 Neglected Tropical Diseases*. 2011 Apr;5(4):e1012.
- 113 Aguayo V, Fernandez BNV, Rodríguez-Valentín M, Ruiz-Jiménez C, Ramos-Benítez MJ, Méndez LB, et al. *Fasciola hepatica* gst downregulates nf- B pathway effectors and inflammatory cytokines while promoting survival in a mouse septic shock model. *Scientific Reports* 2019 9:1. 2019 Feb;9(1):1–16.
- 114 Cervi L, Rossi G, Masih DT. Potential role for excretory–secretory forms of glutathione-s-transferase (gst) in *fasciola hepatica*. *Parasitology*. 1999;119(6):627–33.
- 115 Liu F, Cui S-J, Hu W, Feng Z, Wang Z-Q, Han Z-G, et al. Excretory/secretory proteome of the adult developmental stage of human blood fluke, *schistosoma japonicum** s. 2009 DOI: 10.1074/mcp.M800538-MCP200
- 116 Cao Y, Shi Y, Qiao H, Yang Y, Liu J, Shi Y, et al. Distribution of lethal giant larvae (lgl) protein in the tegument and negative impact of siRNA-based gene silencing on worm surface structure and egg hatching in *schistosoma japonicum*. *Parasitology Research* 2013 113:1. 2013 Oct;113(1):1–9.
- 117 Holoch D, Margueron R. Mechanisms regulating prec2 recruitment and enzymatic activity. *Trends in biochemical sciences*. 2017 Jul;42(7):531–42.
- 118 Pereira AS, Amaral MS, Vasconcelos EJ, Pires DS, Asif H, daSilva LF, et al. Inhibition of histone methyltransferase ezh2 in *schistosoma mansoni* in vitro by gsk343 reduces egg laying and decreases the expression of genes implicated in dna replication and noncoding rna metabolism. *PLoS neglected tropical diseases*. 2018 Oct;12(10). DOI: 10.1371/JOURNAL.PNTD.0006873
- 119 Roquis D, Taudt A, Geyer KK, Padalino G, Hoffmann KF, Holroyd N, et al. Histone methylation changes are required for life cycle progression in the human parasite *schistosoma mansoni*. *PLOS Pathogens*. 2018 May;14(5):e1007066.
- 120 Kimball C, Powers K, Dustin J, Poirier V, Pellettieri J. The exon junction complex is required for stem and progenitor cell maintenance in planarians. *Developmental biology*. 2020 Jan;457(1):119–27.
- 121 Newmark PA, Boswell RE. The mago nashi locus encodes an essential product required for germ plasm assembly in *drosophila*. *Development*. 1994 May;120(5):1303–13.
- 122 Osinka A, Poprzeczko M, Zielinska MM, Fabczak H, Joachimiak E, Wloga D. Ciliary proteins: Filling the gaps. Recent advances in deciphering the protein composition of motile ciliary complexes. *Cells* 2019, Vol 8, Page 730. 2019 Jul;8(7):730.

- 123 Takahashi H, Mitsui Y, Awazawa T, Fujimaki Y, Aoki Y. Control of ciliary activities of schistosoma mansoni miracidia using triton-extracted parasites. *Journal of Parasitology*. 1995;81(5):747–52.
- 124 Yang J, Liu X, Yue G, Adamian M, Bulgakov O, Li T. Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *The Journal of cell biology*. 2002 Nov;159(3):431–40.
- 125 Styczynska-Soczka K, Jarman AP. The drosophila homologue of rootletin is required for mechanosensory function and ciliary rootlet formation in chordotonal sensory neurons. *Cilia*. 2015 Jul;4(SUPPLEMENT 1). DOI: 10.1186/S13630-015-0018-9
- 126 Oda T, Yanagisawa H, Kamiya R, Kikkawa M. A molecular ruler determines the repeat length in eukaryotic cilia and flagella. *Science (New York, NY)*. 2014 Nov;346(6211):857–60.
- 127 Bastin B. Evolution of tektin and odf3 family genes and the role of gene duplication in the specialization of motile ciliary structures in the polychaete platynereis dumerilii. 2018 Available from: <https://dr.lib.iastate.edu/entities/publication/b52d2557-86b4-4df5-8424-92df640ae9fc>
- 128 Matsuyama H, Takahashi H, Watanabe K, Fujimaki Y, Aoki Y. The involvement of cyclic adenosine monophosphate in the control of schistosome miracidium cilia. *The Journal of parasitology*. 2004 Feb;90(1):8–14.
- 129 Sponholtz GM, Short RB. Schistosoma mansoni miracidia: Stimulation by calcium and magnesium. *Journal of Parasitology*. 1976;62(1):155–7.
- 130 Preza M, Guarnaschelli I, Castillo E, Koziol U. Inhibitors of protein kinases a and c impair the motility of oncospheres of the model tapeworm hymenolepis microstoma. *Molecular and Biochemical Parasitology*. 2021 Nov;246:111423.
- 131 Katsumata T, Shimada M, Sato K, Aoki Y. Possible involvement of calcium ions in the hatching of schistosoma mansoni eggs in water. *Journal of Parasitology*. 1988;74(6):1040–1.
- 132 Katsumata T, Kohno S, Yamaguchi K, Hara K, Aoki Y. Hatching of schistosoma mansoni eggs is a ca^{2+} /calmodulin-dependent process. *Parasitology research*. 1989 Jan;76(1):90–1.
- 133 Kawamoto F, Shozawa A, Kumada N, Kojima K. Possible roles of cAMP and ca^{2+} in the regulation of miracidial transformation in schistosoma mansoni. *Parasitology research*. 1989 May;75(5):368–74.
- 134 Russell SL, McFerran NV, Hoey EM, Trudgett A, Timson DJ. Characterisation of two calmodulin-like proteins from the liver fluke, fasciola hepatica. *Biological Chemistry*. 2007 Jun;388(6):593–9.
- 135 Russell SL, McFerran NV, Moore CM, Tsang Y, Glass P, Hoey EM, et al. A novel calmodulin-like protein from the liver fluke, fasciola hepatica. *Biochimie*. 2012 Nov;94(11):2398–406.
- 136 Taft AS, Yoshino TP. CLONING and functional characterization of two calmodulin genes during larval development in the parasitic flatworm schistosoma mansoni. *The Journal of parasitology*. 2011 Feb;97(1):72.
- 137 McCammick EM, McVeigh P, McCusker P, Timson DJ, Morphey RM, Brophy PM, et al. Calmodulin disruption impacts growth and motility in juvenile liver fluke. *Parasites and Vectors*. 2016 Jan;9(1):1–13.
- 138 Samuelson JC, Quinn JJ, Caulfield JP. Hatching, chemokinesis, and transformation of miracidia of schistosoma mansoni. *Journal of Parasitology*. 1984;70(3):321–31.
- 139 Mansour TE. The effect of lysergic acid diethylamide, 5-hydroxytryptamine, and related compounds on the liver fluke, fasciola hepatica. *British journal of pharmacology and chemotherapy*. 1957 Dec;12(4):406–9.
- 140 Patocka N, Sharma N, Rashid M, Ribeiro P. Serotonin signaling in schistosoma mansoni: A serotonin-activated g protein-coupled receptor controls parasite movement. *PLOS Pathogens*. 2014 Jan;10(1):e1003878.

- 141 Gustafsson MK. Immunocytochemical demonstration of neuropeptides and serotonin in the nervous systems of adult schistosoma mansoni. *Parasitology research*. 1987 Mar;74(2):168–74.
- 142 Kreshchenko N, Terenina N, Ermakov A. Serotonin signalling in flatworms: An immunocytochemical localisation of 5-HT 7 type of serotonin receptors in *opisthorchis felinus* and *hymenolepis diminuta*. *Biomolecules*. 2021 Aug;11(8). DOI: 10.3390/BIOM11081212
- 143 Fairweather I, Maule AG, Mitchell SH, Johnston CF, Halton DW. Immunocytochemical demonstration of 5-hydroxytryptamine (serotonin) in the nervous system of the liver fluke, *Fasciola hepatica* (trematoda, digenea). *Parasitology Research* 1987 73:3. 1987 May;73(3):255–8.
- 144 Terenina NB, Kreshchenko ND, Mochalova NV, Nefedova D, Voropaeva EL, Movsesyan SO, et al. The new data on the serotonin and fmrfamide localization in the nervous system of *opisthorchis felinus metacercaria*. *Acta Parasitologica* 2020 65:2. 2020 Jan;65(2):361–74.
- 145 Terenina NB, Kreshchenko ND, Mochalova NB, Movsesyan SO. Serotonin and neuropeptide fmrfamide in the attachment organs of trematodes. *Helminthologia*. 2018 Sep;55(3):185–94.
- 146 Dawes B. Penetration of *fasciola gigantica* cobbold, 1856 into snail hosts. *Nature* 1960 185:4705. 1960;185(4705):51–3.
- 147 Wajdi N. Penetration by the miracidia of *s. Mansoni* into the snail host**. *Journal of Helminthology*. 1966;40(1-2):235–44.
- 148 Mansour TE. THE effect of serotonin and related compounds on the carbohydrate metabolism of the liver fluke, *fasciola hepatica*. *Journal of Pharmacology and Experimental Therapeutics*. 1959;126(3).
- 149 Mansour TE. Serotonin receptors in parasitic worms. *Advances in Parasitology*. 1985 Jan;23(C):1–36.
- 150 Kamemoto ES, Lan L, Mansour TE. In vivo regulation of phosphorylation level and activity of phosphofructokinase by serotonin in *fasciola hepatica*. *Archives of Biochemistry and Biophysics*. 1989 Jun;271(2):553–9.
- 151 Humiczewska M. Specific and non-specific phosphatases in the miracidium of *fasciola hepatica* I. *Folia Histochemica et Cytochemica*. 1975 Jan;13(3-4):231–6.
- 152 Humiczewska M. Some specific and non-specific phosphatases of the sporocyst of *fasciola hepatica*. II. Enzymes associated with the membrane transport. *Folia Parasitologica*. 2002;49(3):221–6.
- 153 Nelson D, Cox M. *Lehninger principles of biochemistry*. Macmillan Learning; 2021.
- 154 Khayath N, Mithieux G, Zitoun C, Coustau C, Vicogne J, Tielens AG, et al. Glyceroneogenesis: An unexpected metabolic pathway for glutamine in *schistosoma mansoni* sporocysts. *Molecular and biochemical parasitology*. 2006;147(2):145–53.
- 155 Salinas G, Selkirk ME, Chalar C, Maizels RM, Fernández C. Linked thioredoxin-glutathione systems in platyhelminths. *Trends in parasitology*. 2004 Jul;20(7):340–6.
- 156 Williams DL, Bonilla M, Gladyshev VN, Salinas G. Thioredoxin glutathione reductase-dependent redox networks in platyhelminth parasites. *Antioxidants & Redox Signaling*. 2013 Sep;19(7):735.
- 157 Bayne CJ, Hahn UK, Bender RC. Mechanisms of molluscan host resistance and of parasite strategies for survival. *Parasitology*. 2001;123 Suppl(SUPPL.). DOI: 10.1017/S0031182001008137
- 158 Zelck UE, Janowsky BV. Antioxidant enzymes in intramolluscan *schistosoma mansoni* and ros-induced changes in expression. *Parasitology*. 2004 May;128(5):493–501.
- 159 Dorey A, Cwiklinski K, Rooney J, Verissimo CDM, Corrales JL, Jewhurst H, et al. Autonomous non antioxidant roles for *fasciola hepatica* secreted thioredoxin-1 and peroxiredoxin-1. *Frontiers in Cellular and Infection Microbiology*. 2021 May;11. DOI: 10.3389/FCIMB.2021.667272/PDF

- 160 Vermeire JJ, Yoshino TP. Antioxidant gene expression and function in in vitro-developing schistosoma mansoni mother sporocysts: Possible role in self-protection. *Parasitology*. 2007 Sep;134(10):1369–78.
- 161 Cancela M, Paes JA, Moura H, Barr JR, Zaha A, Ferreira HB. Unraveling oxidative stress response in the cestode parasite echinococcus granulosus. *Scientific Reports*. 2019 Dec;9(1). DOI: 10.1038/S41598-019-52456-3
- 162 Wilbers RH, Schneiter R, Holterman MH, Drurey C, Smant G, Asojo OA, et al. Secreted venom allergen-like proteins of helminths: Conserved modulators of host responses in animals and plants. *PLOS Pathogens*. 2018 Oct;14(10):e1007300.
- 163 Chalmers IW, Hoffmann KF. Platyhelminth venom allergen-like (val) proteins: Revealing structural diversity, class-specific features and biological associations across the phylum. *Parasitology*. 2012 Sep;139(10):1231.
- 164 Chalmers IW, McArdle AJ, Coulson RM, Wagner MA, Schmid R, Hirai H, et al. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the schistosoma mansoni venom allergen-like (smval) gene family. *BMC Genomics*. 2008 Feb;9:89.
- 165 Farias LP, Chalmers IW, Perally S, Rofatto HK, Jackson CJ, Brown M, et al. Schistosoma mansoni venom allergen-like proteins: Phylogenetic relationships, stage-specific transcription and tissue localization as predictors of immunological cross-reactivity. *International journal for parasitology*. 2019 Jul;49(8):593–9.
- 166 Fernandes RS, Barbosa TC, Barbosa MMF, Miyasato PA, Nakano E, Leite LCC, et al. Stage and tissue expression patterns of schistosoma mansoni venom allergen-like proteins smval 4, 13, 16 and 24. *Parasites and Vectors*. 2017 May;10(1):1–13.
- 167 Fernandes RS, Fernandes LG, Godoy AS de, Miyasato PA, Nakano E, Farias LP, et al. Schistosoma mansoni venom allergen-like protein 18 (smval18) is a plasminogen-binding protein secreted during the early stages of mammalian-host infection. *Molecular and biochemical parasitology*. 2018 Apr;221:23–31.
- 168 Perally S, Geyer KK, Farani PS, Chalmers IW, Fernandez-Fuentes N, Maskell DR, et al. Schistosoma mansoni venom allergen-like protein 6 (smval6) maintains tegumental barrier function. *International Journal for Parasitology*. 2021 Mar;51(4):251–61.
- 169 Cantacessi C, Hofmann A, Young ND, Broder U, Hall RS, Loukas A, et al. Insights into scp/taps proteins of liver flukes based on large-scale bioinformatic analyses of sequence datasets. *PLOS ONE*. 2012 Feb;7(2):e31164.
- 170 Cancela M, Ruétalo N, Dell’Oca N, Silva E da, Smircich P, Rinaldi G, et al. Survey of transcripts expressed by the invasive juvenile stage of the liver fluke fasciola hepatica. *BMC Genomics*. 2010 Apr;11(1):227.
- 171 Ravida A, Cwiklinski K, Aldridge AM, Clarke P, Thompson R, Gerlach JQ, et al. Fasciola hepatica surface tegument: Glycoproteins at the interface of parasite and host. *Molecular & cellular proteomics : MCP*. 2016 Oct;15(10):3139–53.
- 172 Costábile A, Koziol U, Tort JF, Iriarte A, Castillo E. Expansion of cap superfamily proteins in the genome of mesocestoides corti: An extreme case of a general bilaterian trend. *Gene Reports*. 2018 DOI: 10.1016/j.genrep.2018.03.010
- 173 Wu XJ, Sabat G, Brown JF, Zhang M, Taft A, Peterson N, et al. Proteomic analysis of schistosoma mansoni proteins released during in vitro miracidium-to-sporocyst transformation. *Molecular and biochemical parasitology*. 2009 Mar;164(1):32.
- 174 Yoshino TP, Brown M, Wu XJ, Jackson CJ, Ocádiz-Ruiz R, Chalmers IW, et al. Excreted/secreted schistosoma mansoni venom allergen-like 9 (smval9) modulates host extracellular matrix remodelling gene expression. *International journal for parasitology*. 2014;44(8):551–63.

- 175 Kelleher A, Darwiche R, Rezende WC, Farias LP, Leite LC, Schneiter R, et al. Schistosoma mansoni venom allergen-like protein 4 (smval4) is a novel lipid-binding scp/taps protein that lacks the prototypical cap motifs. *Acta Crystallographica Section D: Biological Crystallography*. 2014;70(8):2186–96.
- 176 Dalton JP, Caffrey CR, Sajid M, Stack C, Donnelly S, Loukas A, et al. Proteases in trematode biology. *Parasitic flatworms: molecular biology, biochemistry, immunology and physiology*. 2006;348–68.
- 177 Ingram JR, Rafi SB, Eroy-Reveles AA, Ray M, Lambeth L, Hsieh I, et al. Investigation of the proteolytic functions of an expanded cercarial elastase gene family in schistosoma mansoni. *PLoS Neglected Tropical Diseases*. 2012 Apr;6(4). DOI: 10.1371/JOURNAL.PNTD.0001589
- 178 Hahn C, Fromm B, Bachmann L. Comparative genomics of flatworms (platyhelminthes) reveals shared genomic features of ecto- and endoparasitic neodermata. *Genome biology and evolution*. 2014;6(5):1105–17.
- 179 Cwiklinski K, Jewhurst H, McVeigh P, Barbour T, Maule AG, Tort J, et al. 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*. 2018 Apr;17(4):792–809.
- 180 Petersen CP, Reddien PW. Wnt signaling and the polarity of the primary body axis. *Cell*. 2009 Dec;139(6):1056–68.
- 181 Gilbert S. *Developmental biology*. Sinauer Associates, Inc; 2016.
- 182 Wanninger A. *Evolutionary developmental biology of invertebrates*. Springer; 2015.
- 183 Amerongen RV, Nawijn M, Franca-Koh J, Zevenhoven J, Gulden HVD, Jonkers J, et al. Frat is dispensable for canonical wnt signaling in mammals. *Genes & development*. 2005 Feb;19(4):425–30.
- 184 Pascual-Carreras E, Sureda-Gómez M, Barrull-Mascaró R, Jordà N, Gelabert M, Coronel-Córdoba P, et al. WNT-frizzled-lrp5/6 signaling mediates posterior fate and proliferation during planarian regeneration. *Genes*. 2021 Jan;12(1):1–14.
- 185 Montagne J, Preza M, Castillo E, Brehm K, Koziol U. Divergent axin and gsk-3 paralogs in the beta-catenin destruction complexes of tapeworms. *Development genes and evolution*. 2019 Jul;229(4):89–102.
- 186 Lander R, Petersen CP. Wnt, ptk7, and fgfr1 expression gradients control trunk positional identity in planarian regeneration. *eLife*. 2016 Apr;5(April 2016). DOI: 10.7554/ELIFE.12850
- 187 Nusse R. Wnt signaling and stem cell control. *Cell Research* 2008 18:5. 2008 Apr;18(5):523–7.
- 188 Molina MD, Saló E, Cebrià F. The bmp pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Developmental Biology*. 2007 Nov;311(1):79–94.
- 189 Molina MD, Saló E, Cebrià F. Organizing the dv axis during planarian regeneration. *Communicative & Integrative Biology*. 2011 Jul;4(4):498.
- 190 Hinck AP, Mueller TD, Springer TA. Structural biology and evolution of the tgf- family. *Cold Spring Harbor Perspectives in Biology*. 2016;8(12). DOI: 10.1101/CSHPERSPECT.A022103
- 191 Wurtzel O, Oderberg IM, Reddien PW. Planarian epidermal stem cells respond to positional cues to promote cell-type diversity. *Developmental cell*. 2017 Mar;40(5):491–504.e5.
- 192 Freitas TC, Jung E, Pearce EJ. TGF-beta signaling controls embryo development in the parasitic flatworm schistosoma mansoni. *PLoS pathogens*. 2007;3(4):489–97.
- 193 Japa O, Hodgkinson JE, Emes RD, Flynn RJ. TGF- superfamily members from the helminth fasciola hepatica show intrinsic effects on viability and development. *Veterinary Research*. 2015 Mar;46(1):1–11.
- 194 Tielens AG, Horemans AM, Dunnewijk R, Meer P van der, Bergh SG van den. The facultative anaerobic energy metabolism of schistosoma mansoni sporocysts. *Molecular and biochemical parasitology*. 1992;56(1):49–57.

- 195 Lee FO, Cheng TC. *Schistosoma mansoni*: Alterations in total protein and hemoglobin in the hemolymph of infected *Biomphalaria glabrata*. *Experimental parasitology*. 1972;31(2):203–16.
- 196 Lee FO, Cheng TC. Incorporation of ⁵⁹Fe in the snail *Biomphalaria glabrata* parasitized by *Schistosoma mansoni*. *The Journal of parasitology*. 1972;58(3):481–8.
- 197 Tielens A. Metabolism. *Fasciolosis*. 1999;277–305.
- 198 Berasaín P, Goñi F, McGonigle S, Dowd A, Dalton JP, Frangione B, et al. Proteinases secreted by *Fasciola hepatica* degrade extracellular matrix and basement membrane components. *Journal of Parasitology*. 1997 Feb;83(1):1–5.
- 199 Cancela M, Acosta D, Rinaldi G, Silva E, Durán R, Roche L, et al. A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie*. 2008 DOI: 10.1016/j.biochi.2008.04.020
- 200 Robinson MW, Dalton JP, Donnelly S. Helminth pathogen cathepsin proteases: It's a family affair. *Trends in Biochemical Sciences*. 2008 DOI: 10.1016/j.tibs.2008.09.001
- 201 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. *Molecular and Biochemical Parasitology*. 2009 Sep;167(1):41–7.
- 202 Kondrashov F a, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. *Genome Biol* 3: RESEARCH0008. 2002;3(2):1–9.
- 203 Gevers D, Vandepoele K, Simillion C, Peer YVD. Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends in Microbiology*. 2004;12. DOI: 10.1016/j.tim.2004.02.007
- 204 Preza M, Calvelo J, Langleib M, Hoffmann F, Castillo E, Koziol U, et al. Stage-specific transcriptomic analysis of the model cestode *Hymenolepis microstoma*. *Genomics*. 2021;113(2). DOI: 10.1016/j.ygeno.2021.01.005
- 205 Nei M. Gene duplication and nucleotide substitution in evolution. *Nature*. 1969;221(5175):40–2.
- 206 Innan H, Kondrashov F. The evolution of gene duplications: Classifying and distinguishing between models. *Nature Reviews Genetics* 2010 11:2. 2010 Jan;11(2):97–108.
- 207 Wilson RA, Denison J. The parasitic castration and gigantism of *Lymnaea truncatula* infected with the larval stages of *Fasciola hepatica*. *Zeitschrift für Parasitenkunde Parasitology Research*. 1980;61(2). DOI: 10.1007/BF00925458
- 208 Hamza I, Dailey HA. One ring to rule them all: Trafficking of heme and heme synthesis intermediates in the metazoans. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2012 Sep;1823(9):1617–32.
- 209 Cabán-Hernández K, Gaudier JF, Espino AM. Characterization and differential expression of a ferritin protein from *Fasciola hepatica*. *Molecular and Biochemical Parasitology*. 2012 Mar;182(1-2):54.
- 210 Coustau C, Mitta G, Dissous C, Guillou F, Galinier R, Allienne JF, et al. *Schistosoma mansoni* and *Echinostoma caproni* excretory–secretory products differentially affect gene expression in *Biomphalaria glabrata* embryonic cells. *Parasitology*. 2003 Dec;127(6):533–42.
- 211 Cwiklinski K, Robinson MW, Donnelly S, Dalton JP. Complementary transcriptomic and proteomic analyses reveal the cellular and molecular processes that drive growth and development of *Fasciola hepatica* in the host liver. *BMC Genomics*. 2021 Dec;22(1):1–16.
- 212 Humiczewska M, Rajska K. Lipids in the host-parasite system: Digestive gland of *Lymnaea truncatula* infected with the developmental stages of *Fasciola hepatica*. *Acta Parasitologica*. 2005;50(3).
- 213 Humiczewska M, Rajska K. Some enzymes associated with lipid catabolism in the *Fasciola hepatica* larvae–*Galba truncatula* system. I. Esterases and lipases. *Wiad Parazytol*. 2005;51(3):233–8.

- 214 Cheng TC, Snyder RW. Studies on host-parasite relationships between larval trematodes and their hosts. IV. A histochemical determination of glucose and its role in the metabolism of molluscan host and parasite. *Transactions of the American Microscopical Society*. 1963 Jul;82(3):343.
- 215 Hanley RW, Ultsch GR. Ambient oxygen tension, metabolic rate, and habitat selection in freshwater snails. *Archiv für Hydrobiologie*. 1999 Jan;144(2):195–214.
- 216 Senft AW, Miech RP, Brown PR, Senft DG. Purine metabolism in schistosoma mansoni. *International journal for parasitology*. 1972;2(2):249–60.
- 217 Furukawa J, Inoue K, Maeda J, Yasujima T, Ohta K, Kanai Y, et al. Functional identification of slc43a3 as an equilibrative nucleobase transporter involved in purine salvage in mammals. *Scientific Reports* 2015 5:1. 2015 Oct;5(1):1–11.
- 218 Trembacz H, Jezewska MM. Adenosine phosphorylase and other enzymes of purine salvage in pulmonata snails and their trematoda parasites. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*. 1994 Jan;107(1):135–9.
- 219 Boyle JP, Zaide JV, Yoshino TP. Schistosoma mansoni: Effects of serotonin and serotonin receptor antagonists on motility and length of primary sporocysts in vitro. *Experimental Parasitology*. 2000 Apr;94(4):217–26.
- 220 Kelwick R, Desanlis I, Wheeler GN, Edwards DR. The adamts (a disintegrin and metalloproteinase with thrombospondin motifs) family. *Genome Biology*. 2015 May;16(1):113.
- 221 Wang Q, Da'dara AA, Skelly PJ. The human blood parasite schistosoma mansoni expresses extracellular tegumental calpains that cleave the blood clotting protein fibronectin. *Scientific Reports*. 2017 Dec;7(1). DOI: 10.1038/S41598-017-13141-5
- 222 Cerenius L, Söderhäll K. Coagulation in invertebrates. *Journal of innate immunity*. 2011;3(1):3–8.
- 223 Lima MG, C. Augusto R de, Pinheiro J, Thiengo SC. Physiology and immunity of the invasive giant african snail, achatina (lissachatina) fulica, intermediate host of angiostrongylus cantonensis. *Developmental and comparative immunology*. 2020 Apr;105. DOI: 10.1016/J.DCI.2019.103579
- 224 Alba A, Vázquez AA, Sánchez J, Duval D, Hernández HM, Sabourin E, et al. Fasciola hepatica-pseudosuccinea columella interaction: Effect of increasing parasite doses, successive exposures and geographical origin on the infection outcome of susceptible and naturally-resistant snails from cuba. *Parasites & vectors*. 2018 Oct;11(1). DOI: 10.1186/S13071-018-3155-3
- 225 Verissimo CDM, Jewhurst HL, Dobó J, Gál P, Dalton JP, Cwiklinski K. Fasciola hepatica is refractory to complement killing by preventing attachment of mannose binding lectin (mbl) and inhibiting mbl-associated serine proteases (masps) with serpins. *PLOS Pathogens*. 2022 Jan;18(1):e1010226.
- 226 Deng J, Gold D, LoVerde PT, Fishelson Z. Mapping of the complement c9 binding domain in paramyosin of the blood fluke schistosoma mansoni. *International journal for parasitology*. 2007 Jan;37(1):67–75.
- 227 Ryan S, Shiels J, Taggart CC, Dalton JP, Weldon S. Fasciola hepatica-derived molecules as regulators of the host immune response. *Frontiers in Immunology*. 2020 Sep;11:2182.
- 228 Knox DP. Proteinase inhibitors and helminth parasite infection. *Parasite Immunology*. 2007 Feb;29(2):57–71.
- 229 Cancela M, Corvo I, Silva ED, Teichmann A, Roche L, Díaz A, et al. Functional characterization of single-domain cystatin-like cysteine proteinase inhibitors expressed by the trematode fasciola hepatica. *Parasitology*. 2017 Nov;144(13):1695–707.
- 230 Portet A, Galinier R, Pinaud S, Portela J, Nowacki F, Gourbal B, et al. BgTEP: An antiprotease involved in innate immune sensing in biomphalaria glabrata. *Frontiers in Immunology*. 2018 May;9(MAY). DOI: 10.3389/FIMMU.2018.01206/PDF

- 231 Maggio LSD, Tirloni L, Pinto AF, Diedrich JK, Yates JR, Benavides U, et al. Across intra-mammalian stages of the liver fluke *fasciola hepatica*: A proteomic study. *Scientific Reports* 2016 6:1. 2016 Sep;6(1):1–14.
- 232 Hambrook JR, Kaboré AL, Pila EA, Hanington PC. A metalloprotease produced by larval *schistosoma mansoni* facilitates infection establishment and maintenance in the snail host by interfering with immune cell function. *PLOS Pathogens*. 2018 Oct;14(10):e1007393.
- 233 Liu M, Ju C, Du XF, Shen HM, Wang JP, Li J, et al. Proteomic analysis on cercariae and schistosomula in reference to potential proteases involved in host invasion of *schistosoma japonicum* larvae. *Journal of Proteome Research*. 2015 Nov;14(11):4623–34.
- 234 McVeigh P, Cwiklinski K, Garcia-Campos A, Mulcahy G, O’Neill SM, Maule AG, et al. In silico analyses of protein glycosylating genes in the helminth *fasciola hepatica* (liver fluke) predict protein-linked glycan simplicity and reveal temporally-dynamic expression profiles. *Scientific Reports* 2018 8:1. 2018 Aug;8(1):1–15.
- 235 CANCELA M, SANTOS GB, CARMONA C, FERREIRA HB, TORT JF, ZAHA A. *Fasciola hepatica* mucin-encoding gene: Expression, variability and its potential relevance in host-parasite relationship. *Parasitology*. 2015 Oct;142(14). DOI: 10.1017/S0031182015001134
- 236 Holm L, Sander C. Mapping the protein universe
- 237 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with alphafold. *Nature* 2021 596:7873. 2021 Jul;596(7873):583–9.
- 238 Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics*. 2009 Nov;77(3):499–508.
- 239 Nature Research. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021 DOI: 10.1038/S41586-021-03828-1
- 240 Callaway E. What’s next for alphafold and the ai protein-folding revolution. *Nature*. 2022 Apr;604(7905):234–8.
- 241 Bayly-Jones C, Whisstock JC. Mining folded proteomes in the era of accurate structure prediction. *PLOS Computational Biology*. 2022;18(3). DOI: 10.1371/journal.pcbi.1009930
- 242 Chen Y, McCarthy D, Ritchie M, Robinson M, Smyth G, Hall E. EdgeR: Differential analysis of sequence read count data user’s guide
- 243 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 Jul;5(1). DOI: 10.1038/ncomms5378
- 244 Genome Sequencing *S japonicum*, Consortium FA. The *schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature*. 2009 Jul;460(7253):345–51.
- 245 Stroehlein AJ, Korhonen PK, Chong TM, Lim YL, Chan KG, Webster B, et al. High-quality *schistosoma haematobium* genome achieved by single-molecule and long-range sequencing. *GigaScience*. 2019 Sep;8(9). DOI: 10.1093/gigascience/giz108
- 246 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 Neglected Tropical Diseases*. 2012 Jan;6(1):e1455.

