

AGRADECIMIENTOS

A Estela, por su apoyo incondicional, su paciencia, sus enseñanzas de la vida académica y, más importante, sus enseñanzas a nivel personal. Me gusta la científica en la que me transformé en estos 11 años juntas.

A Pepe por permitirme trabajar en este proyecto y guiarme en esto de la bioinformática. Resultó que esto terminó uniendo el inicio de mi formación en la Facultad de Ingeniería con el final de esta etapa (era la bioinformática mi vocación).

A Fernanda, sin ella mucho de esto no hubiera sido posible. Más allá de lo académico, una amiga de fierro, que siempre está para lo que se la necesite.

Al grupo de gusanos por sus aportes constantes a este proyecto. Gracias Uriel, Matías, Jimena. A las que ya tomaron otro rumbo: Vicky, Soledad, Serrana, Mariana

A la 313, los mejores compañeros, siempre apoyando en lo que se necesite y haciendo que ir al laboratorio sea un placer. Ana, Manu, Julliette, Stefani. Gracias por estar!

A todos los integrantes de la Sección Bioquímica, por todo lo compartido, por hacer que más que un lugar de trabajo, esta fuera una casa.

A mis nuevas Jefas: Vale, Ana F, Paula, Ana R (por partida doble), por confiar en que podía colaborar en sus proyectos y permitir que me tomara el tiempo para escribir la tesis.

Al "Cyber" del Departamento de Genética, Santiago, Pantera, Checha, por integrarme y hacerme sentir parte.

A las integrantes del tribunal por aceptar corregir este trabajo y por sus excelentes críticas y consejos para que la versión final sea mucho mejor (espero haberlo conseguido).

A mis amigas, por su apoyo y acompañamiento, por preguntarme siempre en que andaba la tesis, mi tercer parto académico

A Titho, mi compañero, por entender mi pasión, las ausencias de sólo pensar en la tesis y por apoyarme en este camino.

A Lu, por estar siempre a mi lado, sin importar las circunstancias, hermana y amiga incondicional

Dedico esta tesis a la memoria de mis padres, que me bancaron en todo mientras estuvieron. Gracias a ellos estoy donde estoy y soy lo que soy.

A todos, ¡GRACIAS!

A las agencias financiadoras CAP y ANII por financiar las becas que me permitieron llevar adelante este proceso

ÍNDICE GENERAL

AGRADECIMIENTOS	2
ÍNDICE GENERAL.....	3
RESUMEN	7
ABREVIATURAS	8
INTRODUCCIÓN	9
1. Las infecciones parasitarias como problema de salud	9
2. <i>Mesocetoides corti</i> como modelo de estudio de cestodos	9
2.1. Características generales de cestodos	10
2.2. Clasificación y relaciones filogenéticas	10
2.3. Ciclo de vida	11
2.4. Multiplicación asexual del tetratiridio	13
2.5. Gusano adulto y estrobilización	14
2.6. Oncósferas y primer hospedador intermediario	15
2.7. <i>M. corti</i> como modelo de estudio del desarrollo de cestodos	16
3. Genomas de platelmintos parásitos	16
4. Las células proliferantes de platelmintos.....	21
4.1. Los neoblastos de planaria	21
4.1.1. Métodos de estudio de los neoblastos	22
4.1.2. Marcadores moleculares	23
4.1.3. Heterogeneidad de los neoblastos y clasificación	24
4.2. Las células madre de trematodos	27
4.2.1. Las células tipo neoblasto del estadio adulto de <i>S. mansoni</i>	28
4.2.2. Las células tipo neoblasto en estadios larvarios de <i>S. mansoni</i>	29
4.2.3. Diferenciación de células tipo neoblasto de <i>S. mansoni</i>	30
4.2.4. Clasificación de células tipo neoblasto de <i>S. mansoni</i>	31
4.2.5. Células tipo neoblasto en <i>Fasciola hepatica</i>	34
4.3. Las células germinativas de cestodos	35
4.3.1. Células germinativas en <i>Echinococcus multilocularis</i>	36
4.3.1.1. Caracterización celular	36
4.3.1.2. Marcadores moleculares	37
4.3.1.3. Eliminación de células germinativas	38
4.3.1.4. Diferencias con otros platelmintos	39
4.3.2. Proliferación en <i>Mesocetoides corti</i>	39
4.3.2.1. Genes marcadores	42
HIPÓTESIS, OBJETIVOS Y ESTRATEGIA EXPERIMENTAL	47
Hipótesis	47
Objetivo general	47
Objetivos específicos	47
Estrategia experimental	47
MATERIALES Y MÉTODOS.....	49

1. Cultivo de <i>M. corti</i>	49
2. Irradiación	49
3. Evaluación de los efectos de la radiación sobre las células proliferantes	49
3.1. Extracción de ARN, retrotranscripción y PCR en tiempo real	49
3.2. Análisis de incorporación de análogos de timidina (EdU)	51
3.3. Análisis del ciclo celular	51
4. Genoma y anotación de <i>M. corti</i>	51
5. Secuenciación de ARN de organismos irradiados y control	52
5.1. Extracción de ARN de larvas, construcción de librerías y secuenciación	52
5.2. Mapeo de lecturas y comparación con la anotación existente	52
6. Secuenciación de ARN de células purificadas	53
6.1. Purificación de células proliferantes por citometría de flujo y extracción de ARN	53
6.2. Construcción de librerías y secuenciación	53
7. Actualización del genoma de <i>M. corti</i> utilizando <i>maker</i>	53
7.1. Evidencia utilizada	53
7.1.1. Anotación de genes ribosomales y eliminación de lecturas correspondientes a estos genes	53
7.1.2. Construcción de librería de repetidos específica de <i>M. corti</i>	54
7.1.3. Información basada en secuenciación de ARN	54
7.1.3.1. Estrategia basada en el genoma	54
7.1.3.2. Ensamblado de transcritos de novo	55
7.1.4. Información adicional utilizada como evidencia	55
7.2. Corrida de <i>maker</i>	55
7.3. Análisis de la anotación	56
7.3.1. Anotación funcional	56
7.3.2. Ortología	57
7.3.3. Clasificación de la anotación	58
7.3.4. Comparación de los modelos génicos obtenidos con la anotación disponible (WBPS9)	58
7.3.5. Evaluación de la anotación	58
7.4. Búsqueda de genes de interés	58
7.4.1. Helicasas DEAD-box	58
7.4.2. Nanos	59
8. Análisis de expresión diferencial en organismos irradiados y células proliferantes purificadas	60
8.1. Calidad de las lecturas, mapeo al genoma y expresión diferencial	60
8.2. Análisis de componentes principales (PCA) y análisis de distancias	60
8.3. Enriquecimiento de vías metabólicas y términos de ontología génica (GO)	61
8.4. Evaluación de genes conservados en células madre de platelmintos	61
8.4.1. Genes característicos de células tipo neoblasto de <i>S. mansoni</i>	61
8.4.2. Genes característicos de neoblastos y progenitores de <i>S. mediterranea</i>	61

8.4.3. Ortología.....	62
8.5. Expresión de repetidos	62
CAPÍTULO 1: Reanotación del genoma de <i>M. corti</i> utilizando datos transcriptómicos	63
1. Introducción.....	63
1.1. Anotación de genomas	63
2. Resultados.....	65
2.1. Extracción de ARN y preparación de librerías	65
2.2. Calidad de las lecturas	66
2.3. Estadísticas de mapeo al genoma	67
2.4. Número de transcritos y comparación con la anotación existente (WBPS1)	68
2.5. Estrategia de reanotación del genoma	70
2.6. Comparación de datos de reanotación con genomas disponibles	71
2.7. Análisis de nuevos genes descubiertos por <i>maker</i>	77
2.8. Análisis de nuevas isoformas.....	79
2.9. Análisis de genes de interés.	81
2.9.1. <i>Nanos</i>	81
2.9.2. <i>Helicasas DEAD-box</i>	82
3. Discusión.....	84
CAPÍTULO 2: Efectos de la radiación sobre las células proliferantes de <i>M. corti</i>	86
1. Introducción.....	86
1.1. Irradiación como herramienta para el estudio de células en proliferación.....	86
2. Resultados.....	87
2.1. Evaluación de distintas dosis de irradiación en la sobrevivencia de larvas.	87
2.2. Análisis de incorporación de análogos de timidina (EdU)	88
2.3. Análisis de ciclo celular	89
2.4. Análisis de los niveles de expresión de ARNm de genes marcadores.....	91
2.4.1. <i>Validación del ensayo de cuantificación</i>	91
2.4.2. <i>Expresión de pcna</i>	93
2.4.3. <i>Expresión de genes pumilio</i>	94
2.4.4. <i>Expresión de nanos</i>	96
2.4.5. <i>Expresión de pL10</i>	97
3. Discusión.....	97
CAPÍTULO 3: Análisis de los perfiles de expresión de células proliferantes y células diferenciadas de <i>M. corti</i>.....	102
1. Introducción.....	102
2. Resultados.....	103
2.1. Optimización de la extracción de ARN de células purificadas y construcción de librerías de secuenciación.	103
2.2. Reproducibilidad de las réplicas y comparación entre condiciones	105
2.3. Análisis de expresión diferencial.....	108
2.3.1. <i>Expresión diferencial entre organismos irradiados y control</i>	108
2.3.2. <i>Expresión diferencial entre células proliferantes y células diferenciadas</i>	110
2.3.2.1. <i>Enriquecimiento de vías y términos de ontología génica</i>	111
2.3.3. <i>Comparación de genes expresados diferencialmente en organismos irradiados y</i>	

<i>células proliferantes purificadas</i>	115
2.4. Análisis de expresión de secuencias repetidas y elementos transponibles	117
2.5. Análisis comparativo entre genes marcadores de células proliferantes en platelmintos. 118	
3. Discusión	121
CONCLUSIONES Y PERSPECTIVAS	123
REFERENCIAS	125
ANEXO	141

RESUMEN

En platelmintos de vida libre, las células somáticas diferenciadas no se dividen, siendo las responsables de la renovación celular, homeostasis y la regeneración, las células madre indiferenciadas denominadas neoblastos. En platelmintos parásitos se ha descrito un mecanismo similar, donde las células con capacidad proliferativa se denominan células tipo neoblasto o células germinativas. Estas células han sido estudiadas en varios modelos de parásitos tanto a nivel celular como molecular. En *Mesocostoides corti*, nuestro modelo de estudio, se ha realizado una caracterización celular de las células proliferantes y se han analizado algunos marcadores moleculares conservados. Además se optimizó un protocolo para su aislamiento por citometría de flujo y se comenzó a trabajar en las condiciones de cultivo.

En este trabajo se realizó la caracterización a nivel transcriptómico de las células proliferantes mediante distintos abordajes.

En primer lugar se trabajó en la evaluación de la irradiación como herramienta para el estudio de células proliferantes en *M. corti*. Para ello se utilizaron diferentes técnicas que mostraron una tendencia a la disminución de células en fase S del ciclo celular, indicando que las células proliferantes son sensibles a la radiación, como ocurre en otros platelmintos.

Se obtuvo el transcriptoma de células proliferantes y de células diferenciadas purificadas por citometría de flujo, así como también el transcriptoma de organismos irradiados (con cantidad reducida de células proliferantes) y organismos control. Esto permitió identificar un conjunto de genes que se expresa preferencialmente en las células proliferantes. Muchos de estos genes además son característicos de las células madre de otros platelmintos.

Con los datos generados se actualizó la anotación del genoma disponible, lo que permitió identificar muchos genes no anotados previamente y distintas isoformas de los modelos anotados.

Este trabajo abre un panorama prometedor para continuar con la caracterización detallada de este tipo celular en *M. corti* en particular y en cestodos en general

ABREVIATURAS

ADN: Ácido desoxirribonucleico

ADNc: ADN complementario

ARN: Ácido ribonucleico

ARNi: ARN interferente

ARNInc: ARN largo no codificante

ARNm: ARN mensajero

ARN-polyA: ARN poliadenilado

ARNr: ARN ribosomal

BrdU: 5-bromo-2-desoxiuridina

BSA: Seroalbúmina bovina

DAPI: 4',6-diamidino-2-fenilindol

DNasa: Nucleasa de ARN

DSN: *Double-Strand Specific Nuclease* o Nucleasa específica de doble cadena

EdU: 5-etinil-2'-deoxiuridina

EST: *expressed sequence tag* o marcador de secuencia expresada

Gb: gigabases

Gy: Grays

Kb: kilobases

Mb: megabases

miARN: micro-ARN

pb: pares de bases

PBS: Tampón fosfato salino

PCR: Reacción en cadena de la polimerasa

piARN: *piwi interacting*-ARN

qPCR: PCR cuantitativo o PCR en tiempo real

scRNAseq: *single cell RNA sequencing* o secuenciación de ARN de célula aislada

INTRODUCCIÓN

1. Las infecciones parasitarias como problema de salud

Los platelmintos parásitos implican una enorme carga al desarrollo humano, por su relevante impacto en la salud humana y en la producción. La cisticercosis producida por *Taenia solium* y la hidatidosis causada por especies del género *Echinococcus* son las enfermedades por cestodos más relevantes globalmente, y con una alta incidencia en América Latina (García et al., 2007; Webb y Cabada, 2017). Las regiones más pobres del planeta son también las más afectadas y por lo tanto son en general enfermedades desatendidas, lo que se agrava por ser coendémicas con otras enfermedades más estudiadas como la malaria, tuberculosis y SIDA, resultando frecuentemente en infecciones múltiples que agravan la presentación clínica y morbilidad (Hotez et al., 2008).

El impacto económico de las helmintiasis es también enorme, pues las principales especies animales productivas son infectadas por helmintos, resultando en pérdidas directas (por disminución de peso, estado general, calidad de carne, lana o leche) o indirectas (por los costos de tratamiento o pérdida de la fuerza de trabajo en aquellos países menos industrializados donde se usa tracción animal). Una característica de estas enfermedades es su cronicidad, porque los parásitos pueden permanecer en sus hospederos respectivos por muchos años, en algunos casos sin manifestaciones clínicas, y las personas y/o animales infectados pueden actuar como reservorios para la transmisión de las enfermedades. El control de las enfermedades es complejo, pues si bien existen drogas que son efectivas, éstas son costosas, pueden causar problemas de hepato-toxicidad, y en general no impiden la re-infección. A este panorama debe sumarse la creciente aparición de organismos resistentes a las drogas más frecuentemente utilizadas como el praziquantel y los benzimidazoles (albendazol y triclabendazol) (Brennan et al., 2007). Esto lleva a que sea necesario buscar nuevos mecanismos de control, y en particular a identificar nuevos posibles blancos para el diseño de vacunas o drogas. Para ello es importante conocer en profundidad los aspectos celulares y moleculares del desarrollo de estos parásitos.

En particular, las células madre constituyen un tipo celular especialmente importante para estudiar ya que son las únicas con capacidad proliferativa y dan origen todos los tejidos diferenciados del parásito. En el caso del cestodo *Echinococcus multilocularis*, se ha demostrado que los benzimidazoles utilizados en el tratamiento no son eficientes en la eliminación de estas células, ya que la isoforma de beta-tubulina más expresada en este tipo celular tiene menor afinidad por la droga (Brehm, Kronthaler, et al., 2000; citado por Brehm y Koziol, 2014; y Schubert et al., 2014). El desarrollo de drogas que tengan como blanco mecanismos esenciales para este tipo celular permitiría eliminar la capacidad proliferativa del parásito, acabando con la infección y evitando la recurrencia del parásito al finalizar el tratamiento.

2. *Mesocostoides corti* como modelo de estudio de cestodos

M. corti es un cestodo que se ha adoptado como modelo de estudio para profundizar en el conocimiento de la biología de estos organismos debido a sus múltiples ventajas para reproducir su ciclo de vida en cultivo *in vitro* desde oncósferas obtenidas *in vivo* hasta el

estadio adulto sexualmente maduro (Voge y Seidel, 1968; Barrett et al., 1982; Thompson et al., 1982; Ong y Smyth, 1986). Es posible obtener el estadio larvario en grandes cantidades mediante infección experimental de ratones. En nuestro laboratorio se han optimizado las condiciones para su cultivo desde la fase larvaria hasta el estadio adulto (Britos et al., 2000) y es utilizado desde hace un tiempo para el estudio de genes potencialmente involucrados en los eventos de su ciclo de vida que implican la invasión al hospedero y su desarrollo (Lalanne et al., 2004; Britos et al., 2007; Koziol et al., 2008, 2011; Koziol, Iriarte, et al., 2009; Koziol, Lalanne, et al., 2009; Costábile et al., 2017, 2018). Es además un cestodo que raramente infecta humanos, por lo que es más seguro para su manipulación con respecto a otros de mayor relevancia sanitaria.

2.1. Características generales de cestodos

Los cestodos son platelmintos endoparásitos hermafroditas. Todos los platelmintos presentan simetría bilateral, con el cuerpo aplanado dorso - ventralmente, pseudometamerización y tres capas embrionarias. A partir del mesodermo se originan capas musculares bien desarrolladas, incluyendo fibras longitudinales y circulares, y los órganos reproductivos. Los cestodos carecen de sistema digestivo, circulatorio, respiratorio y esquelético. Se alimentan mediante absorción a través de la pared del cuerpo. Presentan un sistema excretor del tipo protonefridial, con conductos excretores que culminan en un poro en la región posterior. Son acelomados y el espacio entre los órganos está compuesto por un parénquima poco compacto. El sistema nervioso consiste en una masa cerebral en la región anterior conectada con nervios longitudinales, los cuales atraviesan el cuerpo y se interconectan mediante comisuras transversales. Exteriormente están cubiertos por un tegumento que presenta microvellosidades especializadas denominadas microtricos (Storer et al., 1986).

2.2. Clasificación y relaciones filogenéticas

Tradicionalmente se admitían cuatro clases de Platelminotos de acuerdo a los caracteres de los estadios adultos: la clase Turbellaria, que incluye a los platelmintos de vida libre, y las clases Monogenea, Trematoda y Cestoda que incluyen a los organismos parásitos y pertenecen al clado monofilético de los Neodermados. La filogenia interna de los platelmintos ha sido objeto de varios estudios moleculares y morfológicos; sin embargo no existe un consenso acerca de la posición de los diferentes grupos. Utilizando datos de genes ortólogos, se ha mostrado que entre los platelmintos parásitos los Cestodos están más relacionados con los Trematodos que con los Monogeneos, contrariamente a la visión tradicional (Egger et al., 2015). En la Figura I 1 se muestran las relaciones filogenéticas de los organismos analizados en este trabajo.

Tradicionalmente se clasifica a *M. corti* dentro del Filo Platelminotos, dentro de la Clase Cestoda en el Orden de los Ciclofilídeos (Rausch, 1994, citado por Britos, 2000). Existen diversos estudios sobre la clasificación de cestodos, aunque no hay un acuerdo sobre las relaciones de las familias dentro del orden. En estudios filogenéticos, se ha visto que la familia Mesocestoididae aparece como una rama muy basal dentro de los Ciclofilídeos al considerar caracteres morfológicos, o incluso por fuera del orden cuando se analizan caracteres morfológicos y moleculares (Egger et al., 2015; Littlewood et al., 2015). El desconocimiento de la totalidad de su ciclo de vida dificulta una clasificación taxonómica definitiva (Loos-Frank, 1991).

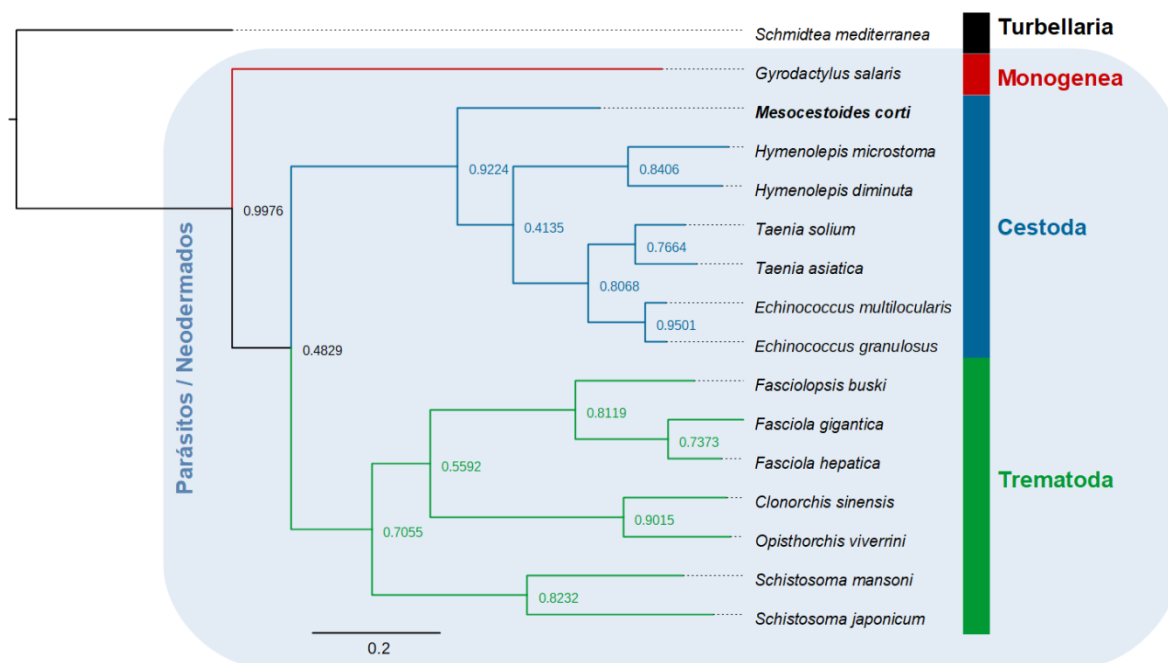


Figura I 1: Relaciones filogenéticas de los platelmintos analizados en este trabajo. De modo de representar las relaciones entre todos los organismos analizados en este trabajo, se muestra el árbol generado durante el análisis de datos del Capítulo 1 (2924 genes ortólogos de acuerdo al resultado del programa *Orthofinder*, ver punto 7.3.2 de Materiales y Métodos). La barra vertical coloreada distingue las clases Trematoda (verde), Cestoda (azul), Monogenea (rojo) y Turbellaria (negro). En negrita se señala a *M. corti*. El cuadro celeste indica especies dentro del grupo de los neodermados parásitos.

2.3. Ciclo de vida

A pesar de que hace más de 65 años que se investiga el ciclo de vida de platelmintos del género *Mesocostoides*, aún no se ha podido determinar completamente. Se ha postulado que incluye tres hospederos, siendo dos de ellos intermediarios y el tercero definitivo (Figura I 2).

En el ciclo de vida propuesto, los huevos producidos por el gusano adulto serían ingeridos por el primer hospedero intermediario, de identidad aún desconocida, que se presume podría ser un artrópodo. En éste cada oncósfera se desarrollaría a larva cisticercoide (Webster, 1949), la cual continúa su desarrollo a larva tetratiridia cuando el primer hospedero es ingerido por el segundo hospedero intermediario. Finalmente, cuando el hospedero definitivo ingiere al segundo hospedero intermediario, el tetratiridio se desarrolla a gusano adulto, sexualmente maduro, que genera los huevos para cerrar el ciclo. Si bien se pudieron reconstituir *in vitro* las etapas del ciclo de vida desde oncósferas obtenidas de infecciones naturales hasta gusanos adultos sexualmente maduros (Voge y Coulombe, 1966; Voge y Seidel, 1968; Thompson et al., 1982), el desarrollo de larvas a partir de las oncósferas no se ha observado *in vivo*. Por esta razón no se puede comparar el desarrollo *in vitro* con el que ocurre en el ciclo natural del organismo.

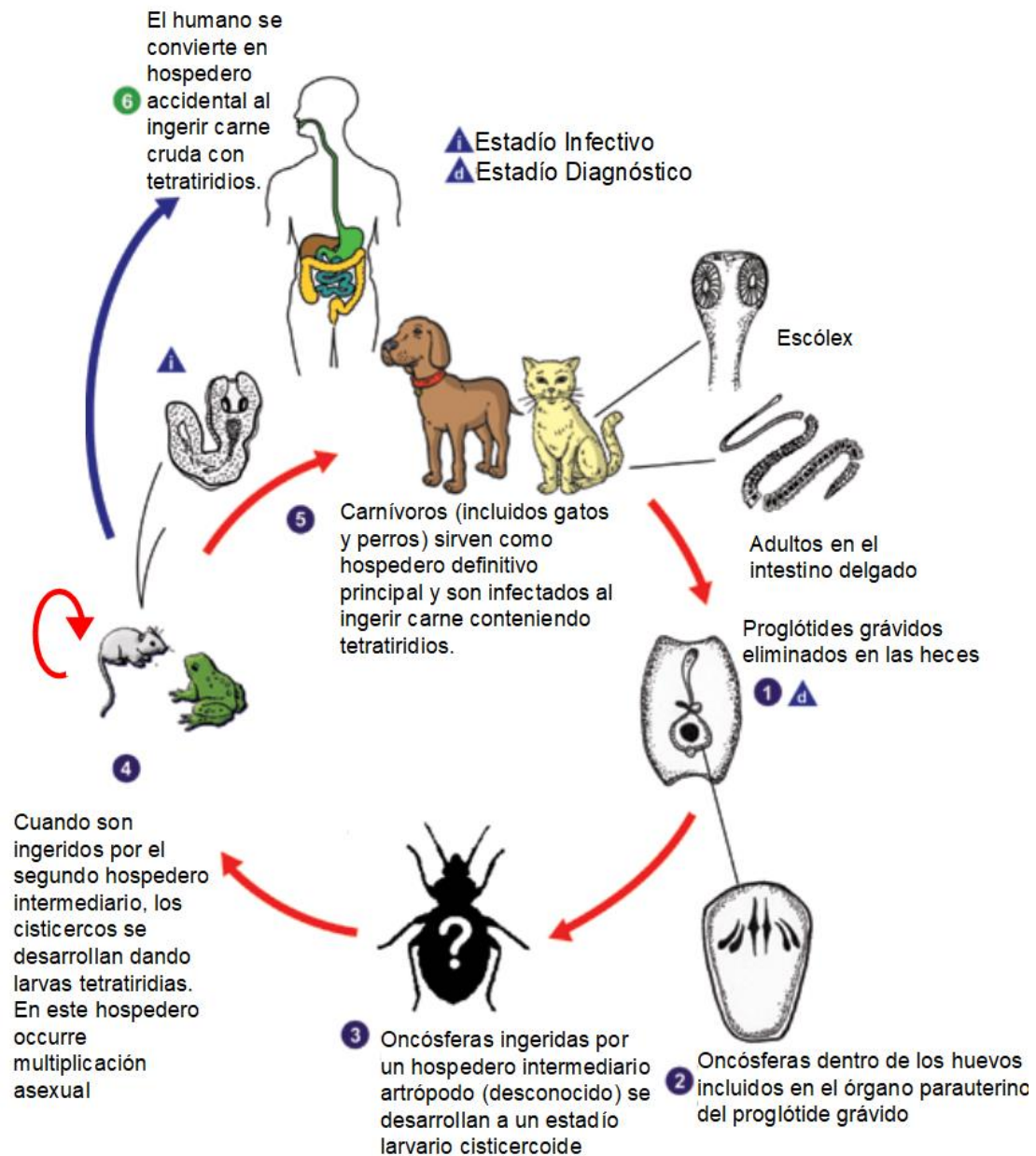


Figura 1 2: Ciclo de vida de *M. corti*. Requiere tres hospederos para completar su ciclo de vida. 1: Los proglótidos grávidos son eliminados en las heces del hospedero definitivo (principalmente carnívoros, incluyendo cánidos, félidos y mustélidos). 2: Dentro de cada proglótide grávido, cientos de oncosferas están contenidas en el órgano parauterino. 3: El primer hospedero intermediario es presuntamente un artrópodo (de identidad desconocida), que es infectado al ingerir los proglótidos grávidos o las oncosferas. En este se cree que las oncosferas se desarrollan a un segundo estadio larvario (cisticercoide o procercoide). 4: Cuando el primer hospedero intermediario es ingerido por el segundo hospedero intermediario (pequeños mamíferos, aves, reptiles y anfibios) el segundo estadio larvario se desarrolla al tercer estadio larvario infectivo (tetratiridio). En este hospedero ocurre multiplicación asexual de la larva 5: El hospedero definitivo se infecta al ingerir carne contaminada con tetratiridios. En este hospedero, el cestodo se establece en el intestino delgado, donde madura. Pueden observarse proglótidos grávidos en las heces a las dos semanas de infección. 6: Los humanos no son los hospederos definitivos usuales, pero pueden infectarse al ingerir carne poco cocida de animales infectados con tetratiridios.

Adaptado de <http://www.dpd.cdc.gov/HTML/Mesocestoidiasis.htm>.

2.4. Multiplicación asexual del tetratiridio

Los tetratiridios son los estadios larvarios que se desarrollan en el segundo hospedero intermediario. Son aplanados dorso-ventralmente y en general presentan una ligera curvatura hacia el lado proliferativo, midiendo en promedio unos 2 mm. Como todos los organismos del género *Mesocetoides* carecen de rostelo y su escólex es aplanado, presentando cuatro ventosas musculares separadas de a pares por el surco apical. Su cuerpo sólido consiste en tegumento, parénquima (con numerosos corpúsculos calcáreos y células almacenadoras de glucógeno) y músculos internos. Se observa un sistema excretor, compuesto por una vesícula excretora y cuatro canales que terminan en el poro excretor en la región posterior, y un sistema nervioso. No presenta ningún esbozo de aparato reproductor (Hess, 1980; Barrett et al., 1982). Son capaces de infectar una gran variedad de vertebrados, como anfibios, reptiles, aves o mamíferos (Eckert et al., 1969).

El estadio larvario se multiplica asexualmente en la cavidad del cuerpo y el hígado del segundo hospedero intermediario (Specht y Voge, 1965; Novak, 1972). En la Figura I 3 se muestran las etapas observadas de la multiplicación asexual. Hess (1980) realizó un estudio ultraestructural de la larva tetratiridia y su multiplicación asexual, observando que en la región central del escólex entre las ventosas y por debajo del surco apical, a la que denominó macizo apical, existe una masa celular reticular que contiene numerosos núcleos en mitosis (algo que no se observó en otras partes del organismo).

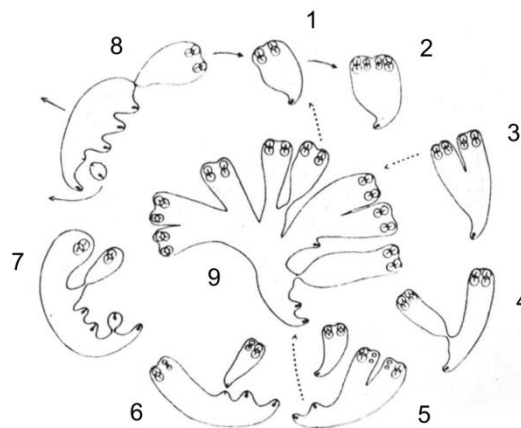


Figura I 3: Patrón de multiplicación asexual del tetratiridio de *M. corti* en roedores. 1: Tetratiridio, 2: Profundización del surco apical y aparición de esbozos de nuevas ventosas. 3: Aumento de tamaño de las ventosas y profundización del surco longitudinal. 4: Culminación de la duplicación de las estructuras nerviosas y excretoras. Estado de división avanzado. Alejamiento de la larva hija de la parental. 5: Separación de la larva hija dejando en el punto de separación sobre el lado proliferativo de la larva parental, yemas y nuevos poros excretorios asociados a una vesícula excretora funcional. 6: Separación de una nueva larva hija, siempre del lado proliferativo de la larva parental, en una posición anterior. 7: Por la aceleración del proceso, las nuevas larvas hijas y la parental presentan un único par de ventosas, desarrollándose el segundo luego de comenzada la división del escólex. 8: Generación de una larva alargada con numerosas yemas del lado proliferativo, que ocasionalmente pueden separarse formando un fragmento acéfalo que degenera. 9: Observación de fragmentos policefálicos que al introducirse en ratones dan lugar a tetratiridios simples. Tomado de Novak (1972)

2.5. Gusano adulto y estrobilización

El gusano adulto reside en el intestino de vertebrados carnívoros, como coyotes, zorros, zorrillos, gatos y perros (Specht y Voge, 1965; Eckert et al., 1969; Crosbie et al., 2000; Padgett y Boyce, 2004; Padgett et al., 2005). Se observa la coexistencia de tetratiridios con gusanos adultos en estos hospederos y además los tetratiridios son capaces de multiplicarse asexualmente (Eckert et al., 1969).

El gusano adulto es un típico cestodo segmentado, compuesto por un escólex anterior con cuatro ventosas y una estróbila formada por varios segmentos o proglótides en distintos estadios de maduración. Los segmentos anteriores son inmaduros (en general más anchos que largos) y los posteriores maduros (en general más largos que anchos). Consecuentemente hay un proceso continuo de maduración de los segmentos, que depende de la proliferación y diferenciación. Un esquema del gusano adulto se muestra en la Figura I 4.

Los tetratiridios al ingresar al hospedero definitivo se alargan y adquieren una forma vermiforme. Suelen aparecer varios segmentos simultáneamente, siendo el desarrollo más acelerado en los posteriores. En paralelo con el desarrollo de los proglótides aparecen nuevos segmentos en la región anterior. El proceso de maduración puede determinarse cuando se analizan los segmentos sucesivos (Figura I 4). Al tiempo que aparecen los segmentos, se observa el rudimento del sistema reproductor masculino (Estadio 1, Figura I 4), que se diferenciará en el saco del cirro y los testículos. El sistema reproductor femenino se desarrolla levemente después que el masculino, y comienza por la aparición del rudimento femenino (Estadio 4, Figura I 4). Cuando los testículos están maduros, el ovario aún está en desarrollo, aunque el útero se desarrolla simultáneamente. Para el momento que maduran los ovarios y las glándulas vitelinas, los testículos comienzan a degenerar. El segmento maduro (Estadios 5 y 6, Figura I 4) incluye dos canales excretores laterales, conectados por una comisura transversal, y los sistemas reproductores masculino y femenino. El sistema reproductor masculino está compuesto por los testículos ubicados lateralmente, conectados a vasos deferentes que desaparecen una vez que el testículo vació su contenido; y el saco del cirro, que contiene un cirro enrollado. El sistema reproductor femenino está compuesto por un ovario bilobular, con dos glándulas vitelinas pequeñas, el cual se conecta con el útero ubicado en la línea media. En el extremo posterior del útero existe una masa oscura de células que darán lugar al órgano parauterino, claramente visible en los segmentos más posteriores. Éste comienza a desarrollarse cuando los ovarios y glándulas parauterinas degeneran, una vez que el útero está lleno de óvulos y células vitelinas. Surge por la proliferación de las células de la región posterior del útero, diferenciándose en una cápsula fibrosa de paredes delgadas. En estos segmentos ya no se observan los canales excretores. En los segmentos grávidos (Estadio 9, Figura I 4) se observan en el órgano parauterino las oncósferas, que se identifican por los 6 ganchos y la cubierta fina protectora del huevo (Barrett et al., 1982; Thompson et al., 1982).

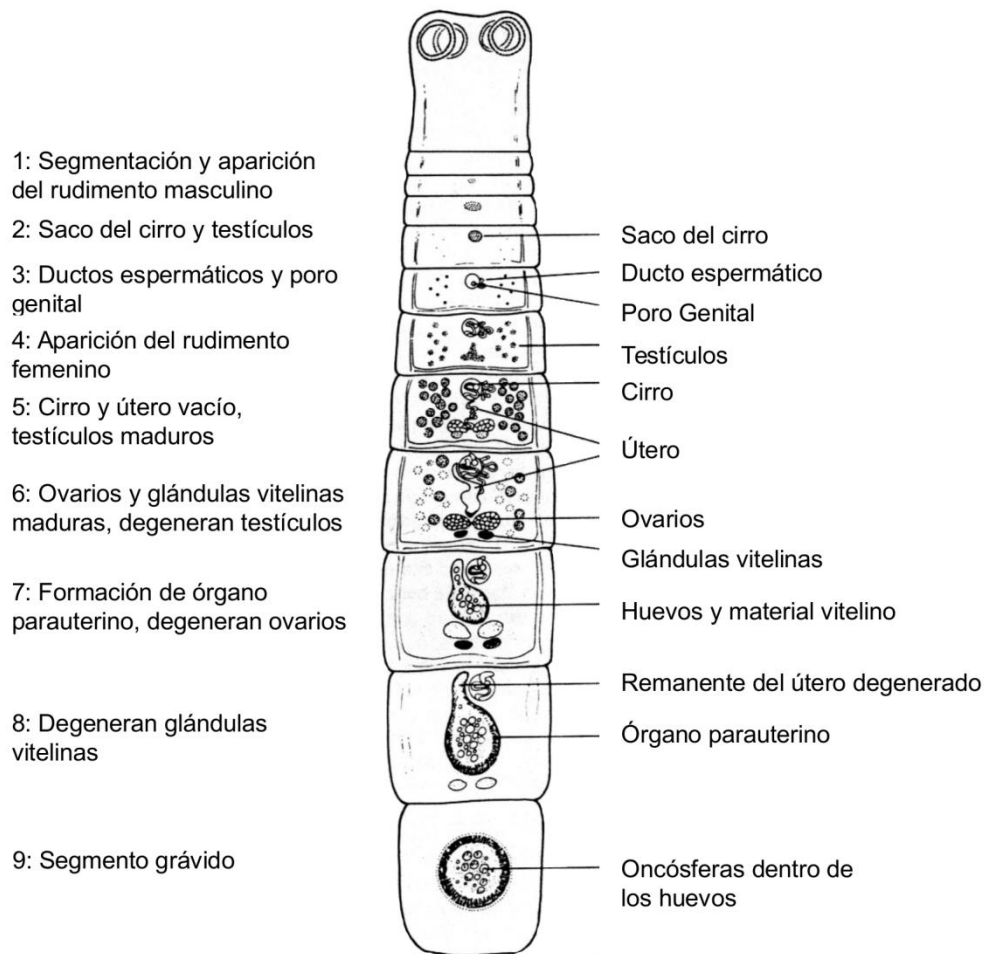


Figura I 4: Esquema del gusano adulto. Se muestra el gradiente de maduración antero-posterior a lo largo de los proglótides. Adaptado de Barret *et al.* 1982.

2.6. Oncósferas y primer hospedador intermediario

Como se mencionó anteriormente, no se conoce con certeza la identidad del primer hospedero intermediario, capaz de ser infectado por las oncósferas producidas por el gusano adulto. Soldatova (1944; citado por Loos-Frank, 1991) reportó la infección de ácaros con oncósferas, pero no obtuvo material suficiente para continuar estudiando el ciclo experimentalmente, además de que no trabajó con animales de laboratorio con los controles adecuados que permitieran descartar una infección natural previa. Webster (1949) plantea una discusión con respecto a cuál podría ser el primer hospedero intermediario, en base a la forma de alimentación del segundo hospedero intermediario, Loos-Frank (1991) retoma esta discusión. Los animales en los que se observan infecciones por tetratiridios son muy variados, incluyendo varias clases de vertebrados con diferentes hábitos. Estos autores plantean que el primer hospedero intermediario debe ser un invertebrado predador, capaz de ingerir un segmento grávido entero, o al menos el órgano parauterino, ya que cuando se liberan las oncósferas de su cubierta se desecan rápidamente. Se descartan los anélidos y moluscos en base al tipo de alimentación de los segundos hospederos intermediarios conocidos. Dado que éstos en general son herbívoros o insectívoros, es natural pensar que el primer hospedero sería un insecto predador o, aunque menos probable, un organismo

que se alimente de materia en descomposición, pudiendo ingerir con ella las oncósferas. Siguiendo este razonamiento, se descarta cualquier tipo de vertebrado.

Webster (1949) intentó infectar 36 especies de invertebrados (insectos, nematodos, moluscos, crustáceos y anélidos) con segmentos grávidos, siendo los resultados negativos en todos los casos. Loos-Frank (1991) reportó el intento de infección de varias especies de ácaros, sin obtener resultados. Finalmente, Padgett y Boyce (2004) reportaron la presencia de ADN procedente del género *Mesocostoides* en hormigas que co-habitan con hospederos definitivos infectados. Sin embargo, estas hormigas no eran capaces de infectar al segundo hospedero intermediario.

Estas observaciones llevaron a pensar que no sería necesario un primer hospedero intermediario y que las oncósferas serían capaces de infectar lo que hoy se conoce como segundo hospedero intermediario. Varios autores (Henry, 1927; Witenberg, 1934; Srivastava, 1939; Webster, 1949; citados por Loos-Frank, 1991) intentaron infectar anfibios, reptiles y roedores con huevos del parásito sin éxito. Otra opción, de la cual no existen reportes, es que las oncósferas eclosionen en las heces del hospedero intermediario y el embrión resultante sea el que infecte al que actualmente se conoce como segundo hospedero intermediario.

A pesar de no haber certeza de que exista un primer hospedero intermediario, y en caso de existir cuál es su identidad, se ha podido obtener tetratiridios totalmente desarrollados a partir de oncósferas mantenidas en cultivo *in vitro* (Voge y Seidel, 1968). Si bien sería necesario comparar las distintas etapas del desarrollo obtenidas de esta manera con el desarrollo *in vivo*, ésta es una buena aproximación para conocer el patrón de desarrollo post-embionario.

2.7. *M. corti* como modelo de estudio del desarrollo de cestodos

Como se mencionó anteriormente, se ha logrado reproducir la mayor parte del ciclo de vida *in vitro*, desde oncósferas obtenidas a partir de gusanos segmentados de infecciones naturales a tetratiridios (Voge y Seidel, 1968), la multiplicación asexual de éstos (Voge y Coulombe, 1966), y el desarrollo estrobilar y maduración partiendo del tetratiridio hasta alcanzar el estadio adulto (Barrett et al., 1982; Thompson et al., 1982). Esto, sumado a la facilidad de obtener material parasitario en grandes cantidades a partir de ratones infectados experimentalmente, constituye una de las principales ventajas para la adopción de este organismo como modelo de estudio de cestodos. Los tetratiridios utilizados actualmente fueron aislados de lagartijas (*Sceloporus occidentalis biseriatus*) infectadas naturalmente (Specht y Voge, 1965), los que fueron mantenidos por pasaje seriado en ratas y ratones de laboratorio.

3. Genomas de platelmintos parásitos

La genómica de platelmintos comenzó con la publicación de transcriptomas de diversos trematodos (ver organismos y referencias en Tabla I 1) y se amplió en 2009 con la publicación de los genomas de los trematodos *Schistosoma mansoni* y *Schistosoma japonicum*. Con la disminución de costos de las tecnologías de secuenciación masiva, comenzaron a publicarse transcriptomas y genomas de varios Platelminfos (Tabla I 1). En particular el 2013 marcó un hito para la genómica de cestodos, con la publicación de los genomas de las especies *Echinococcus multilocularis*, *Echinococcus granulosus*,

Hymenolepis microstoma y *Taenia solium* (Tsai, IJ et al., 2013; Zheng et al., 2013). Ese mismo año comenzaron a generarse los primeros borradores de varias especies de helmintos, en el marco de la iniciativa *50 Helminth Genomes* del *Wellcome Trust Sanger Institute* para lo cual aportamos ADN genómico para la secuenciación de *M. corti* (Coghlan et al., 2019). Estos genomas fueron compilados junto a los genomas ya existentes en la plataforma *WormBase ParaSite* (Howe et al., 2016), cuya primera versión se hizo pública en Octubre de 2014. Actualmente, la versión 12 de *WormBase ParaSite* incluye 152 genomas correspondientes a 127 especies de helmintos aportados por laboratorios de todo el mundo (<https://parasite.wormbase.org/index.html>). En particular existen genomas de 31 especies de platelmintos, correspondiendo a 15 genomas de cestodos, 12 de trematodos, 2 de monogeneos y 2 platelmintos de vida libre.

Los genomas de platelmintos tienen tamaños variables dependiendo de la clase. Los genomas de cestodos son los más pequeños, con un tamaño del orden de los 100Mb, seguidos de los genomas de trematodos (entre 300 Mb para organismos del genero *Schistosoma* y 1 Gb para *Fasciola hepatica*) y de turbelarios (entre 600 y 900Mb). Estos genomas presentan diferente grado de contigüidad, desde un N50 de 29kb para el monogeneo *Protopolystoma xenopodis*, que tiene un genoma de 617.3Mb, hasta un N50 de 32.1Mb para el genoma del trematodo *S. mansoni*, con un genoma de 346.1Mb (Coghlan et al., 2019).

El genoma de *E. multilocularis* es tomado como referencia de genomas de cestodos debido a su alta calidad, producto de su finalización manual con datos de mapeo óptico. El 89% de la secuencia se encuentra contenido en 9 cromosomas que presentan únicamente 23 *gaps*. Uno de estos cromosomas está completo y 13 de los 18 telómeros fueron unidos a los *scaffolds* (Tsai, IJ et al., 2013).

M. corti tiene 7 pares de cromosomas (Raghunathan y Voge, 1974) y su genoma fue secuenciado con la tecnología *Illumina*, utilizando librerías pareadas de 100pb construidas a partir de fragmentos de ADN de 500pb. El borrador del genoma tiene una cobertura de 233X y un tamaño de 117 Mb con un N50 de 65,8Kb. Se encuentra fragmentado en 7068 *contigs* y presenta 3.101 *gaps* y 907.300 bases ambiguas. Tiene un contenido de GC de 41.5% y el 10.2% constituyen secuencias repetidas (Coghlan et al., 2019). En base a este borrador del genoma, se utilizó una estrategia de anotación común para todos los genomas secuenciados en el marco de la iniciativa *50 Helminth Genomes*, en la cual se utilizó la herramienta MAKER realizando varias iteraciones para el entrenamiento de los predictores *ab initio* SNAP, Augustus y GeneMark-ES, combinado con el uso de ESTs y proteínas de organismos relacionados (por detalles, ver Berriman et al. (2018)). A partir de esta anotación, se predijeron 10.614 genes codificantes para proteínas, con una densidad de 90,58 genes por Mb. Las proteínas predichas tienen un largo promedio de 463.3 aminoácidos (Coghlan et al., 2019). Esto es comparable a la predicción de genes en *E. multilocularis*, en el que se utilizaron datos de secuenciación de ARN y genes conservados como apoyo para el refinamiento manual de los modelos. En este organismo se predijeron 10.150 genes codificantes con un largo promedio de 500,72 aminoácidos (Tsai, IJ et al., 2013).

Tabla I 1: Resumen de los genomas y transcriptomas publicados de platelmintos parásitos

Organismo	Tipo	Referencia
<i>Echinostoma paraensei</i> (T)	Transcriptoma	(Adema et al., 2000)
<i>Echinococcus granulosus</i> (C)	Transcriptoma	(Fernández et al., 2002; Parkinson et al., 2012)
<i>Schistosoma mansoni</i> (T)	Transcriptoma	(Verjovski-Almeida et al., 2003)
<i>Schistosoma japonicum</i> (T)	Transcriptoma	(Hu et al., 2003)
<i>Paragonimus westermani</i> (T)	Transcriptoma	(Kim et al., 2006)
<i>Moniezia expansa</i> (C)	Transcriptoma	(Zhao et al., 2009)
<i>Schistosoma mansoni</i> (T)	Genoma	(Berriman et al., 2009)
<i>Schistosoma japonicum</i> (T)	Genoma	(The <i>Schistosoma japonicum</i> Genome Sequencing and Functional Analysis Consortium, 2009)
<i>Taenia solium</i> (C)	Transcriptoma	(Almeida et al., 2009)
<i>Opisthorchis viverrini</i> (T)	Transcriptoma	(Young, Campbell, et al., 2010)
<i>Clonorchis sinensis</i> (T)	Transcriptoma	(Young, Campbell, et al., 2010)
<i>Fasciola hepatica</i> (T)	Transcriptoma	(Cancela et al., 2010; Young, Hall, et al., 2010)
<i>Fasciola gigantica</i> (T)	Transcriptoma	(Young et al., 2011)
<i>Clonorchis sinensis</i> (T)	Genoma	(Wang et al., 2011; Huang et al., 2013)
<i>Schistosoma haematobium</i> (T)	Genoma	(Young et al., 2012)
<i>Echinococcus multilocularis</i> (C)	Genoma	(Tsai, IJ et al., 2013)
<i>Echinococcus granulosus</i> (C)	Genoma	(Tsai, IJ et al., 2013; Zheng et al., 2013)
<i>Taenia solium</i> (C)	Genoma	(Tsai, IJ et al., 2013)
<i>Hymenolepis microstoma</i> (C)	Genoma	(Tsai, IJ et al., 2013)
<i>Fasciola hepatica</i> (T)	Genoma	(Cwiklinski et al., 2015; McNulty et al., 2017)
<i>Echinococcus canadensis</i> (C)	Genoma	(Maldonado et al., 2017)
<i>Hymenolepis microstoma</i> (C)	Transcriptoma	(Olson et al., 2018)
<i>Mesocestoides corti</i> (C)*	Genoma	(Coghlan et al., 2019)

Negrita: genomas disponibles en WormBase ParaSite. C: Cestodo, T: Trematodo. *: Publicado junto a los genomas de 50 Helminth Genomes Initiative

La comparación de los proteomas de cestodos con los de varias especies de organismos bilaterales (*Drosophila melanogaster*, *Caenorhabditis elegans* y mamíferos, entre otros) muestra una gran cantidad de ganancias y pérdidas de genes asociadas a la adaptación al parasitismo. En particular, muchas vías metabólicas se encuentran reducidas o totalmente ausentes, como la síntesis *de novo* de purinas, pirimidinas, algunos aminoácidos y lípidos (ácidos grasos y colesterol) (Figura I 5). Todos estos compuestos son tomados de los hospederos, lo que se ve reflejado en el aumento en el número de copias y los niveles de expresión de genes codificantes para distintos transportadores comparado con otros bilaterales. Además del aumento de la cantidad de transportadores, existen otros genes que presentan un mayor número de copias, como las proteínas HSP70 (*Heat Shock Protein 70*), USP (*Universal Stress Protein*), PARP (*Poly-(ADP-ribose) polymerase*), Tetraspaninas y proteínas con dominios asociados a la adhesión célula - célula, entre otros (Tsai, IJ et al., 2013; Zheng et al., 2013).

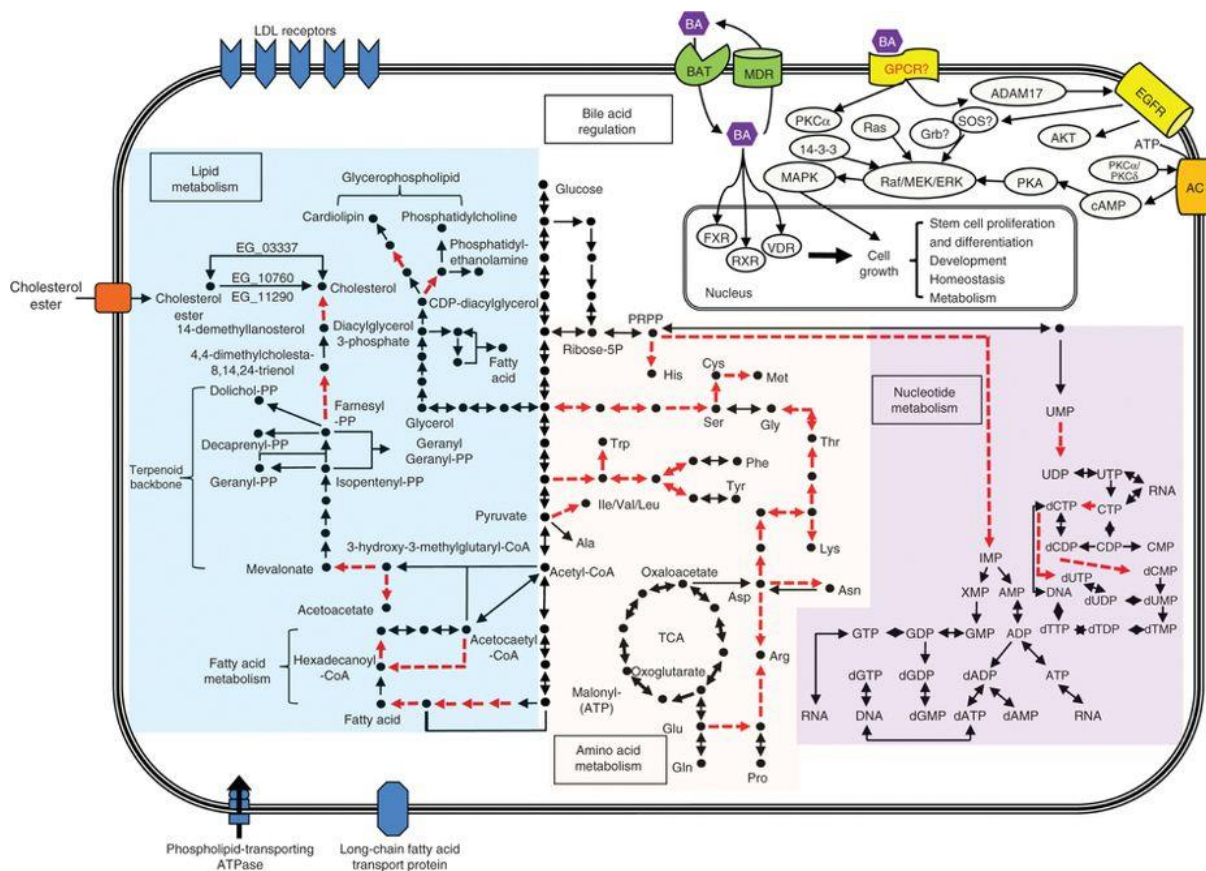


Figura I 5: Esquema de vías metabólicas y regulación por ácidos biliares en *E. granulosus*. Los distintos colores de fondo representan distintas vías que tienen pérdida de genes. Las flechas negras indican enzimas presentes en *E. granulosus* mientras que las rojas representan las enzimas ausentes. En la parte superior derecha se muestra el transporte de ácidos biliares y las vías reguladas por los mismos. Tomado de Zheng et al. (2013).

Por otro lado, también se observa la ausencia de varios genes relevantes en el desarrollo y en el funcionamiento de las células madre. La reducción en la cantidad de genes *homeobox* es común a todos los platelmintos parásitos, pero en el caso de cestodos la reducción es aún mayor. También el complemento de genes *wnt* se encuentra reducido.

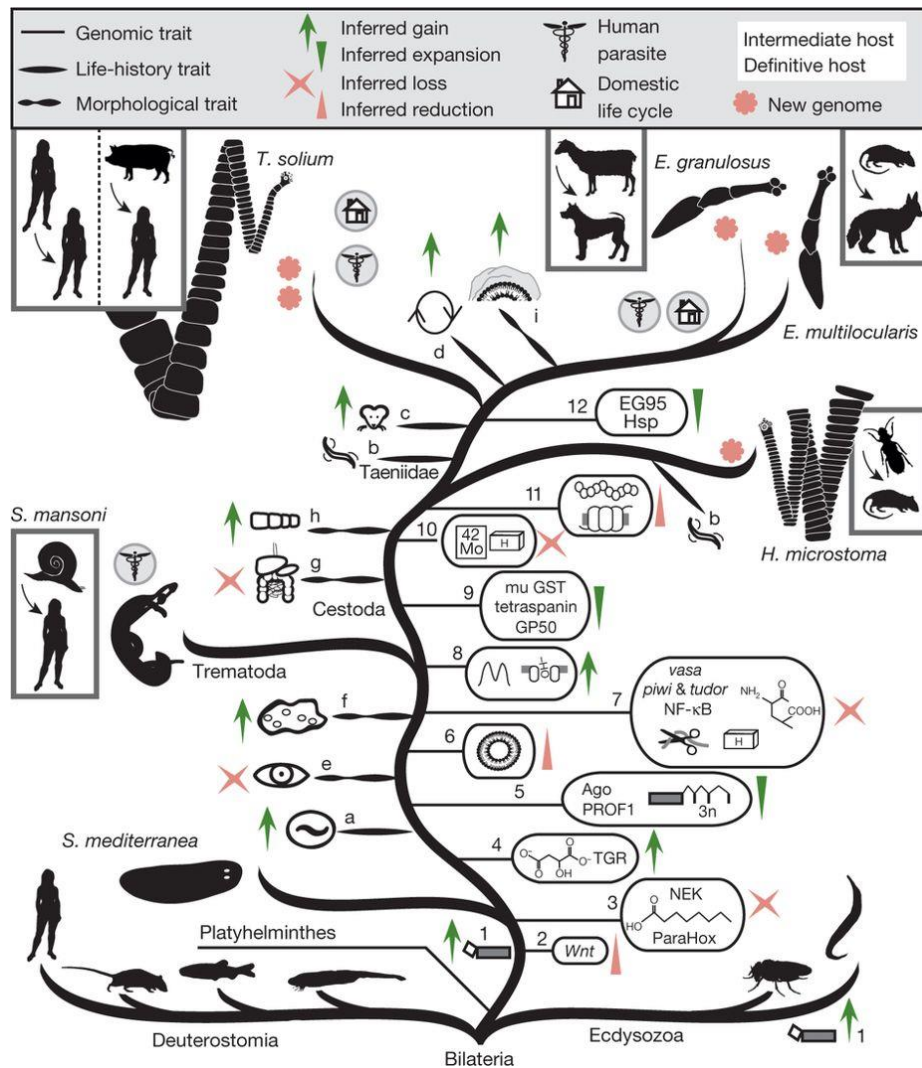


Figura I 6: Evolución del parasitismo en cestodos. Filogenia de las ramas principales de Bilaterales: Ecdysozoa, Deuterostomia y Lophotrochozoa. Las ganancias y pérdidas de características del ciclo de vida incluyen: a) evolución del endoparasitismo, b) transferencia pasiva entre hospederos, c) adquisición de hospederos intermediarios vertebrados, d) capacidad de proliferar asexualmente en el hospedero intermediario. Los rasgos morfológicos que han evolucionado incluyen: e) pérdida de ojos, f) ganancia del epitelio sincitial de los neodermados, g) pérdida de intestino, h) segmentación del cuerpo, i) aparición de la capa laminar, conteniendo apomucinas. Las pérdidas y ganancias a nivel genómico incluyen: 1) *Trans-splicing*, 2) pérdida de genes *wnt*, 3) pérdida de NEK quinasas, biosíntesis de ácidos grasos y genes *ParaHox*, 4) capacidad de metabolismo anaerobio por medio de la vía de la dismutación del malato/rodoquinona y unión de *glutaredoxina* y *tioeredoxina reductasa* en el gen *tioeredoxina glutatión reductasa* (TGR), 5) evolución de familias de genes específicas de cestodos y trematodos como proteínas Argonauta, *micro-exón genes*, y PROF1 GPCRs, 6) pérdida de genes peroxisomales, 7) pérdida de genes *vasa*, *piwi* y *tudor*, de la vía NF-κB, de 24 familias de genes *homeobox*, de proteasas metabólicas y biosíntesis de aminoácidos. Las pérdidas y ganancias específicas de cestodos incluyen 8) innovación de la distribución bimodal de intrones y mayor cantidad de transportadores de ácidos grasos, 9) expansión de glutation-treansferasas de clase Mu, Antígenos GP50 y tetraspaninas, 10) pérdida de la vía de biosíntesis de molibdopterina y de 10 familias de *homeoboxes*, 11) menos GPCRs y neurotransmisores codificados por cada protopéptido y 12) expansión de proteínas HSP y antígenos específicos de especie. Tomado de Tsai, IJ et al. (2013)

Otra característica de los genomas de platelmintos parásitos es la ausencia de genes marcadores de células madre, como *vasa*, *piwi* y *tudor*. Los trematodos y cestodos tienen dos copias de la helicasa DEAD-box *pL10* y se propone que una de estas copias reemplaza

la función de vasa. Las proteínas Piwi pertenecen a la familia de proteínas Argonauta y están implicadas en el silenciamiento de elementos transponibles mediante la vía de silenciamiento de piARNs. Estos piARNs tampoco se han encontrado en trematodos y cestodos, aunque existen otros ARNs pequeños para los cuales aún no se ha definido una función y se cree que podrían estar cumpliendo una función similar a los piARNs. Existe una subfamilia de proteínas argonauta (denominada FLAGO) que es específica de platelmintos parásitos y se postula que podría unirse a algún tipo de ARN pequeño para producir el silenciamiento de elementos transponibles (Fontenla et al., 2017). Estos genes son esenciales para la determinación del destino de las células madre, por lo que su ausencia sugiere que las vías asociadas al mantenimiento y diferenciación de las células madre en platelmintos parásitos está altamente modificada (Tsai, IJ et al., 2013; Skinner et al., 2014; Fontenla et al., 2017). En la Figura I 6 se resumen las pérdidas y ganancias de genes junto a distintas características de los platelmintos.

4. Las células proliferantes de platelmintos

Como se describió anteriormente para *M. corti* y en general para todos los cestodos, estos organismos requieren de una constante proliferación celular tanto para la renovación de sus tejidos como para la generación de proglótides en la región del cuello durante la segmentación del estadio adulto. Durante este proceso en cada segmento se forman los órganos reproductivos a medida que los proglótides van madurando. En el estadio larvario, los procesos de proliferación son necesarios durante la metamorfosis desde la oncósfera a metacestodo y en la reproducción asexual de los metacestodos, cuando existe. Esta estrategia de renovación tisular es similar a la observada en platelmintos de vida libre y trematodos.

En Platelmintos las células diferenciadas son post-mitóticas, es decir que no vuelven a entrar al ciclo celular, y existe una población de células indiferenciadas que son la fuente de nuevas células necesarias para la renovación de tejidos, crecimiento y regeneración (Peter et al., 2004; Reuter y Kreshchenko, 2004; Rink, 2013; Koziol y Brehm, 2015; Zhu y Pearson, 2016). Esta estrategia de renovación celular es distinta a la existente en otros animales conocidos donde, dependiendo del tejido, pueden coexistir células proliferantes diferenciadas con células madre específicas del tejido (con diversos niveles de compromiso para su diferenciación) y elementos celulares ya diferenciados (Peter et al., 2004).

Estas células indiferenciadas han sido muy estudiadas en planarias (platelmintos de vida libre) y en menor medida en cestodos. Más recientemente, se han realizado estudios a nivel celular y molecular en el trematodo *Schistosoma mansoni* y en los cestodos *E. multilocularis* y *M. corti*. Dependiendo de la clase de Platelmintos, las células indiferenciadas se denominan neoblastos (en platelmintos de vida libre), células tipo neoblasto (en trematodos) o células germinativas (en cestodos). En este trabajo se utilizará la terminología células proliferantes para denominar a este tipo celular en *M. corti*.

4.1. Los neoblastos de planaria

Las planarias son platelmintos de vida libre conocidos por su alta capacidad regenerativa. Son capaces de regenerar un organismo completo a partir de prácticamente cualquier región del cuerpo (Reddien y Sánchez Alvarado, 2004) y renuevan sus tejidos somáticos a

una gran velocidad (Rink, 2013). Sus principales tejidos (sistema nervioso, muscular, digestivo, entre otros) carecen de células proliferativas residentes y su renovación y crecimiento depende de neoblastos ubicados en el parénquima adyacente (Rossi et al., 2008; Rink, 2013). Estos neoblastos se localizan a lo largo de todo el cuerpo, con excepción de la faringe y la región más anterior por delante de los fotorreceptores (Figura I 7 A) y precisamente estas regiones son incapaces de regenerar nuevos organismos tras la amputación (Newmark y Sánchez Alvarado, 2000; Reddien y Sánchez Alvarado, 2004; Rink, 2013).

Los neoblastos son pequeñas células basofílicas con morfología indiferenciada (Figura I 7 B). Poseen un gran núcleo con un nucleolo prominente y escaso citoplasma (Baguñà y Romero, 1981; Peter et al., 2004; Rossi et al., 2008). A nivel ultraestructural (Figura I 7 C) presentan una gran cantidad de ribosomas libres y mitocondrias, mientras que carecen de retículo endoplasmático rugoso y aparato de Golgi (Morita et al., 1969; Hay y Coward, 1975). Presentan también gránulos ribonucleoproteicos perinucleares denominados cuerpos cromatoides. Estos son similares tanto molecular como morfológicamente a los gránulos germinales de las células germinales de varios animales (Morita et al., 1969; Hay y Coward, 1975; Auladell et al., 1993; Yoshida-Kashikawa et al., 2007; Rossi et al., 2008).

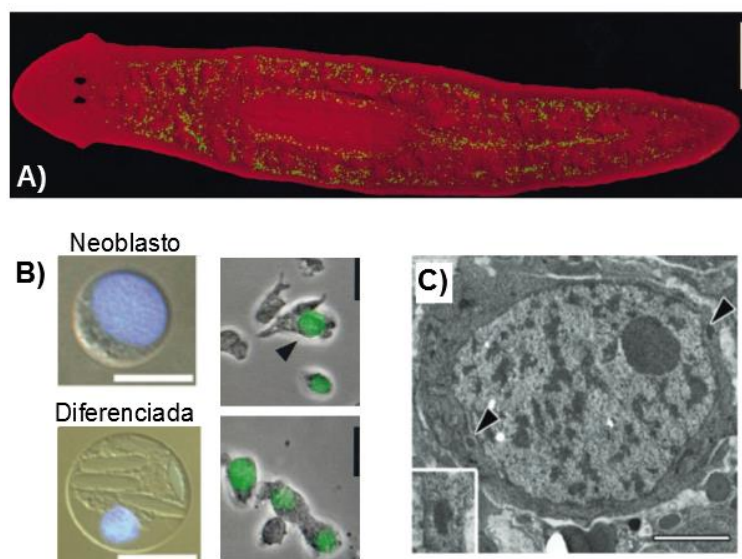


Figura I 7: Neoblastos de planaria. A) Neoblastos de *Girardia dorotocephala* marcados con BrdU (Tomada de Newmark y Sánchez Alvarado, 2000). La barra representa 450 μm . B) Morfología de neoblastos y células diferenciadas de *Schmidtea mediterranea* aisladas por citometría de flujo (izquierda, Tomada de Sánchez Alvarado, 2007) y macerados celulares incubados con BrdU (derecha, Tomada de Newmark y Sánchez Alvarado, 2000). 48 horas luego del marcado (superior) se observan células en proceso de diferenciación (flecha), mientras que 17 horas luego del marcado (inferior) solo se marcan células con morfología similar a neoblastos y se observa una célula en proceso de división. Las barras representan 10 μm . C) Ultraestructura de un neoblasto de *Dugesia japonica* (Tomada de Shibata et al., 2010). Se observa la gran relación núcleo / citoplasma y un nucleolo prominente. Se señalan con flechas los cuerpos cromatoides, uno de los cuales se muestra magnificado en el recuadro inferior. La barra representa 2 μm

4.1.1. Métodos de estudio de los neoblastos

El desarrollo y optimización de técnicas de biología celular y molecular para estos organismos ha hecho que el conocimiento sobre la biología de los neoblastos de planaria esté muy avanzado. Dentro de las técnicas utilizadas se encuentran la interferencia de ARN

(ARNi) (Sánchez Alvarado y Newmark, 1999), aislamiento de neoblastos por citometría de flujo (Hayashi et al., 2006), hibridación *in situ* de genes marcadores e incorporación de análogos de timidina (Newmark y Sánchez Alvarado, 2000; Pearson et al., 2009), transplante de células (Baguna et al., 1989; Wagner et al., 2011) y secuenciación de ARN de poblaciones de células y de células aisladas (*single-cell RNAseq*) (Hayashi et al., 2010; van Wolfswinkel et al., 2014; Wurtzel et al., 2015; Molinaro y Pearson, 2016; Fincher et al., 2018; Plass et al., 2018; Zeng et al., 2018). Al ser las únicas células con capacidad proliferativa, son sensibles a la radiaciones ionizantes (Figura I 8 A, E y F), lo que constituye una de las principales herramientas utilizadas para su estudio (Hayashi et al., 2006; Rossi et al., 2007; Eisenhoffer et al., 2008; Salvetti et al., 2009; Solana et al., 2012).

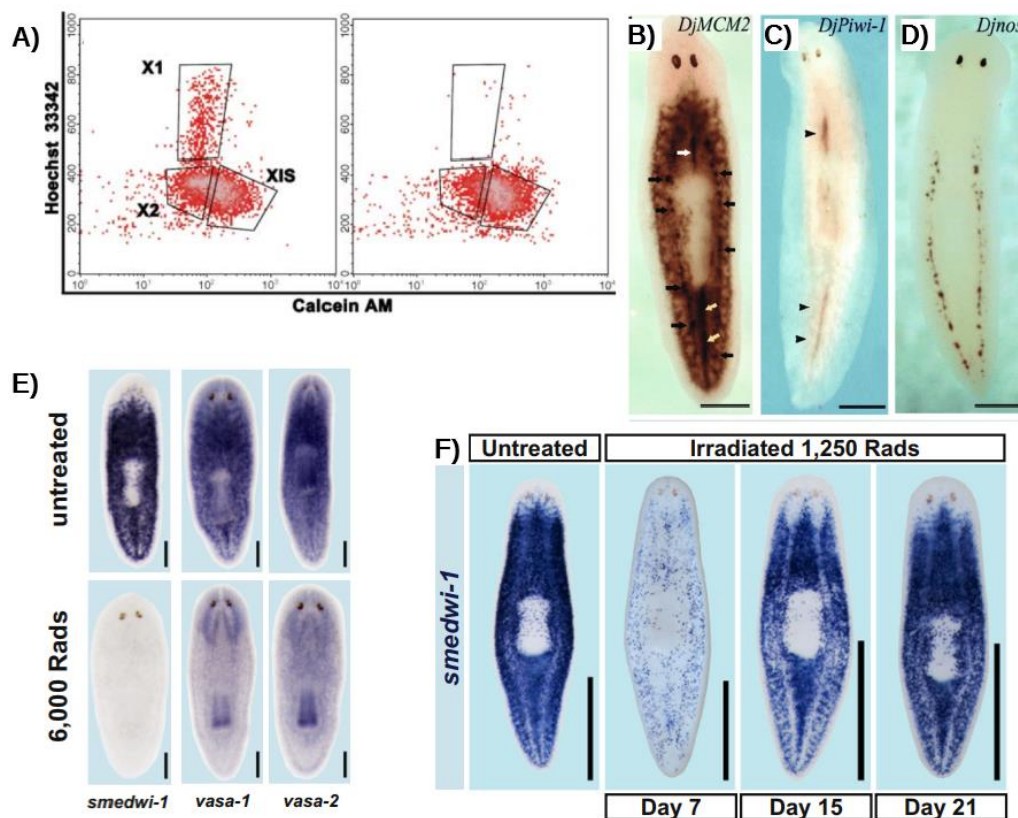


Figura I 8: Efecto de la irradiación sobre neoblastos de planaria y genes marcadores. A) Gráficos de citometría de flujo de células aisladas de planarias no tratadas (izquierda) e irradiadas (derecha). La fracción X1 que desaparece tras la irradiación corresponde a los neoblastos, la fracción X2 disminuye considerablemente y corresponde a la progenie de los neoblastos, mientras que la fracción Xins no presenta cambios y corresponde a células diferenciadas (Tomada de Hayashi et al., 2006). B a D) Localización de la expresión de algunos genes marcadores de neoblastos en *Dugesia japonica*. La barra representa 500 μ m (Tomada de Rossi et al., 2008). E) Efecto de la irradiación letal sobre la expresión del gen *piwi* y de genes *vasa* en *S. mediterranea*. La barra representa 200 μ m F) Efecto de la irradiación subletal en *S. mediterranea* y repoblación con nuevos neoblastos a lo largo del tiempo. La barra representa 500 μ m (E y F tomadas de Wagner et al., 2012).

4.1.2. Marcadores moleculares

Utilizando estas técnicas se han descrito numerosos marcadores moleculares para los neoblastos, originalmente estudiando genes candidatos aislados (Figura I 8B a D) y posteriormente mediante análisis transcriptómico de células purificadas por citometría de

flujo o de organismos a los que se le eliminaron los neoblastos por irradiación (Shibata et al., 1999; Salvetti et al., 2000, 2005; Orie et al., 2005; Reddien, 2005; Guo et al., 2006; Rossi et al., 2007, 2008; Eisenhoffer et al., 2008; Rouhana et al., 2010; Labbé et al., 2012; Önal et al., 2012; Rink, 2013).

En estos estudios se comprobó que los neoblastos expresan específicamente genes del programa de multipotencialidad germinal (GMP: *Germinal Multipotency Program*) como los genes *piwi* (Figura I 8 E y F), *vasa* (Figura I 8 E), *tudor* y *nanos* (Figura I 8 B), además de proteínas relacionadas con la regulación post-transcripcional de la expresión génica (Juliano et al., 2010; Rouhana et al., 2010). Estos genes tienen importantes funciones en el mantenimiento o diferenciación de los neoblastos. En particular, *smedwi-1*, uno de los tres genes *piwi* presentes en *S. mediterranea*, es utilizado como marcador de los neoblastos, así como sus ortólogos en otras especies de planaria. Estos genes tienen un importante rol en la diferenciación de los neoblastos y su auto-renovación (Reddien, 2005; Palakodeti et al., 2008). También se han comparado los genes marcadores de neoblastos de planaria con genes reguladores de las células madre embrionarias de mamíferos (ESC; *Embryonic Stem Cells*). Aunque no existen ortólogos claros en planaria para los principales reguladores de las ESC (*Oct4*, *Sox2* o *Nanog*), los perfiles de expresión génica en general se correlacionan muy bien. Esto indica que existe conservación de los mecanismos de regulación de la pluripotencia en mamíferos y planarias, y es posible también que sea conservado en los metazoarios (Labbé et al., 2012; Önal et al., 2012).

En los primeros estudios realizados, los neoblastos parecían constituir una población homogénea de células pluripotentes esenciales para la regeneración. Mediante experimentos de trasplante de distintas preparaciones celulares a planarias sin neoblastos, se observó que las células diferenciadas no son capaces de rescatarlas. Únicamente se logra rescatarlas al transplantar preparaciones enriquecidas en neoblastos (Baguna et al., 1989). Debido a las limitaciones de las herramientas utilizadas, estos experimentos no permitieron determinar si la totalidad de los neoblastos tienen la capacidad de repoblar las planarias irradiadas (proceso que se muestra en la Figura I 8 F), o si existen linajes con distintas potencialidades. Posteriormente, mediante estudios de trasplante de células individuales se demostró que la población de neoblastos es heterogénea, y que existe una población verdaderamente pluripotente que es capaz de repoblar las planarias irradiadas y reemplazar todos sus tejidos tras cierto tiempo (Wagner et al., 2011). A esta subpoblación se les llamó neoblastos clonogénicos (*cNeoblasts*) y constituye al menos el 5% de la totalidad de los neoblastos. Existe además evidencia de esta heterogeneidad que surge de estudios de expresión génica, en donde se observa la expresión de distintos genes marcadores en una proporción de neoblastos y no en todos (Figura I 8 B a D). Además existen células con morfología similar a la de neoblastos o con morfología intermedia entre neoblastos y células diferenciadas que son postmitóticas y expresan una combinación de marcadores que representan diferentes estados de diferenciación hacia los distintos tipos celulares (Higuchi et al., 2007; Shibata et al., 2010).

4.1.3. Heterogeneidad de los neoblastos y clasificación

El desarrollo de tecnologías de *single-cell sequencing* permitió abordar esta heterogeneidad y trazar las vías de diferenciación desde los neoblastos hacia los distintos tipos celulares diferenciados, evidenciando la existencia de neoblastos pluripotentes y de neoblastos comprometidos con los distintos linajes, que expresan distintas combinaciones de genes

marcadores. En los primeros análisis, con la caracterización de unas pocas células y una selección de genes de interés, se describieron varias clases de neoblastos comprometidos con distintos linajes. Los neoblastos ζ dan lugar al linaje epidérmico, dependen del gen *zfp-1* para mantenerse y no son requeridos para la regeneración. Los neoblastos σ son los que proliferan preferencialmente durante la regeneración y tienen una capacidad mayor, siendo precursores de distintos linajes (incluidos los neoblastos ζ). Se cree que los *cNeoblasts* están incluidos en esta clase. Los neoblastos γ pertenecen a una subclase dentro de los neoblastos σ que posiblemente sean precursores intestinales (van Wolfswinkel et al., 2014). Más adelante, se analizaron nuevos datos y se plantearon algunas modificaciones a estas relaciones entre las clases. Por ejemplo, se plantea que los neoblastos σ derivan de los *cNeoblasts*, mientras que los neoblastos γ no serían una subclase de los neoblastos σ , si no que derivarían de estos. Además se descubrieron dos clases adicionales: los neoblastos ν , ubicados en la región cefálica que parecerían ser precursores del linaje neural y un grupo adicional (Grupo 4) cuya descendencia no tiene un linaje claro (Molinari y Pearson, 2016). En la Figura I 9 se muestra un esquema representando las relaciones entre las clases de neoblastos.

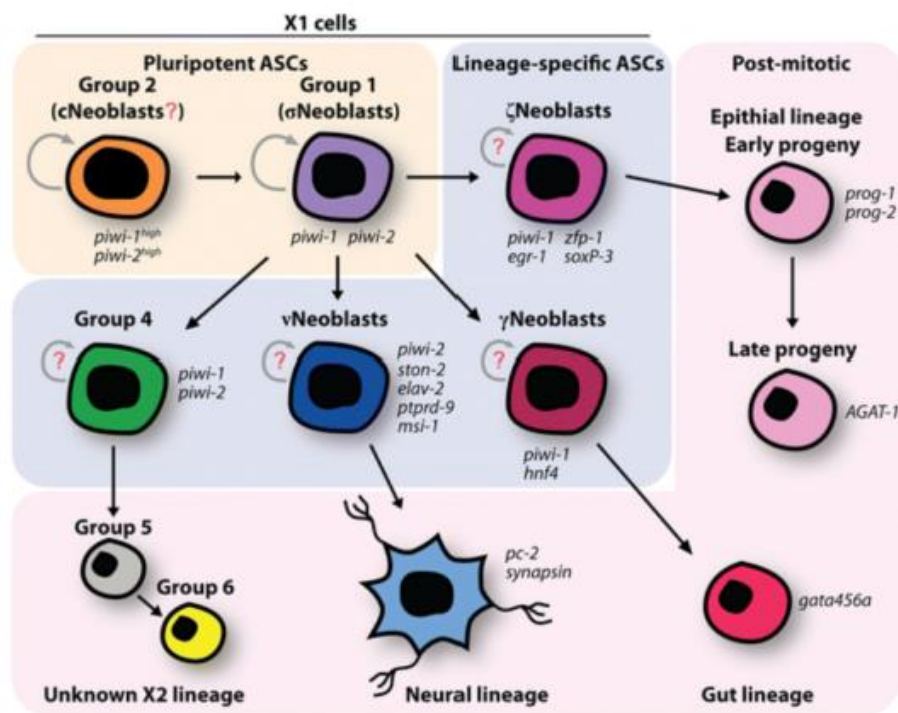


Figura I 9: Clases de neoblastos de planaria. Modelo de linajes de células de planaria en el cual los *cNeoblasts* dan lugar a los neoblastos σ pluripotentes. Estos son precursores de los neoblastos ζ , γ , ν y de una clase no identificada (grupo 4). Estos neoblastos especializados dan lugar a los distintos tejidos del adulto. Los signos de interrogación rojos indican el desconocimiento de su existencia o su capacidad de auto-renovación. Tomada de Molinari y Pearson (2016).

Más recientemente se publicaron tres estudios masivos de transcriptómica de células individuales. Uno de ellos se enfoca principalmente en los neoblastos y su descendencia, ya que parten de una población de células *piwi*⁺. El análisis de las células individuales permitió identificar 12 subtipos de neoblastos (Figura I 10 A), dentro de los cuales el grupo Nb2 se caracteriza por la expresión del marcador de superficie Tetraspanina 1 (TSPAN1). Mediante este marcador se aisló por citometría de flujo la población de células Nb2 y se probó que al transplantarlas son capaces de rescatar planarias irradiadas letalmente como lo hacen los

construir el árbol de diferenciación, se determinaron las relaciones entre los distintos grupos de neoblastos, progenitores y tejidos diferenciados (Figura I 11 B). En el centro de este árbol existe un grupo de neoblastos que se conecta con los demás grupos celulares y se considera como el grupo que contiene a los neoblastos totipotentes (Figura I 11 C). Además, se logra asociar a distintos grupos de progenitores con las clases de neoblastos previamente descritas (neoblastos γ , neoblastos σ y neoblastos ζ , Figura I 11 D)

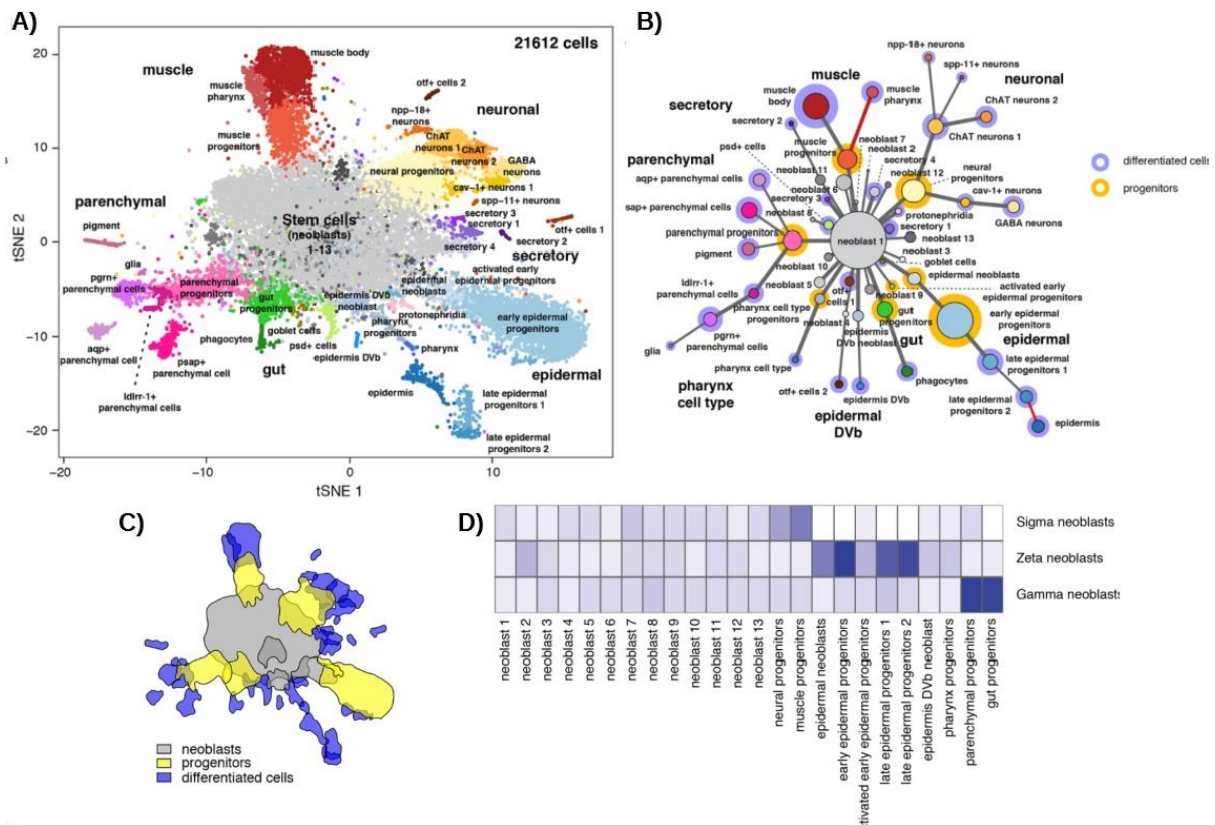


Figura I 11: Análisis transcriptómico de células individuales de planaria obtenido con la metodología Drop-Seq y relaciones de linaje entre las células. A) *Plot* de células coloreadas de acuerdo a la expresión de genes marcadores de distintos tejidos (gris: neoblastos, anaranjado: linaje neural, rojo: músculo, violeta: secretorio, azul: linaje epidérmico, rosado: protonefridio, verde: intestino y magenta: parénquima). B) Árbol de linajes mostrando las relaciones entre los distintos grupos de células. Los grupos asociados a progenitores y células diferenciadas se muestran con halos amarillos o celestes respectivamente. C) Ubicación de células diferenciadas (azul), progenitores (amarillo) y neoblastos (gris) en el *plot* de células. D) *Heatmap* mostrando los niveles de expresión de genes marcadores de las clases de neoblastos σ , ζ y γ . Blanco: menor expresión, azul: mayor expresión. Tomadas de Plass *et al.* (2018).

Desde los estudios iniciados a finales de la década de 1960 por Jaume Baguñà, con planarias residentes en un lago artificial en Barcelona, pasando por Phil Newmark como estudiante de post-doctorado en su laboratorio en la década de 1990 y el contacto de éste con Alejandro Sánchez Alvarado en el 2000, la planaria *S. mediterranea* llegó a los Estados Unidos para convertirse en uno de los modelos más importantes para el estudio de la regeneración a nivel molecular y celular (Baguñà, 2018).

4.2. Las células madre de trematodos

En el caso de trematodos, los primeros estudios moleculares sobre las células indiferenciadas fueron realizados en *Schistosoma mansoni*, estudiando este tipo celular en

el estadio adulto (Collins et al., 2013) y durante la transición desde miracidio a esporocisto (Wang et al., 2013).

4.2.1. Las células tipo neoblasto del estadio adulto de *S. mansoni*

En el adulto, se identificaron células con capacidad proliferativa distribuidas por todo el mesénquima y en los órganos reproductores de machos y hembras, con un patrón de localización y una morfología similar a los neoblastos de planarias (Figura I 12 A a C). Además, son sensibles a la radiación ya que aplicando una dosis de entre 100 y 200 Gy estas células dejan de incorporar EdU tras 48 horas y presentan 128 genes con un nivel de expresión más de dos veces menor que en organismos control (Figura I 12 D). Dentro de estos genes identifican muchas similitudes con genes asociados a neoblastos de planaria, como por ejemplo genes implicados en su mantenimiento (*p53*, *sox*, *fgfr* y *ago-2*).

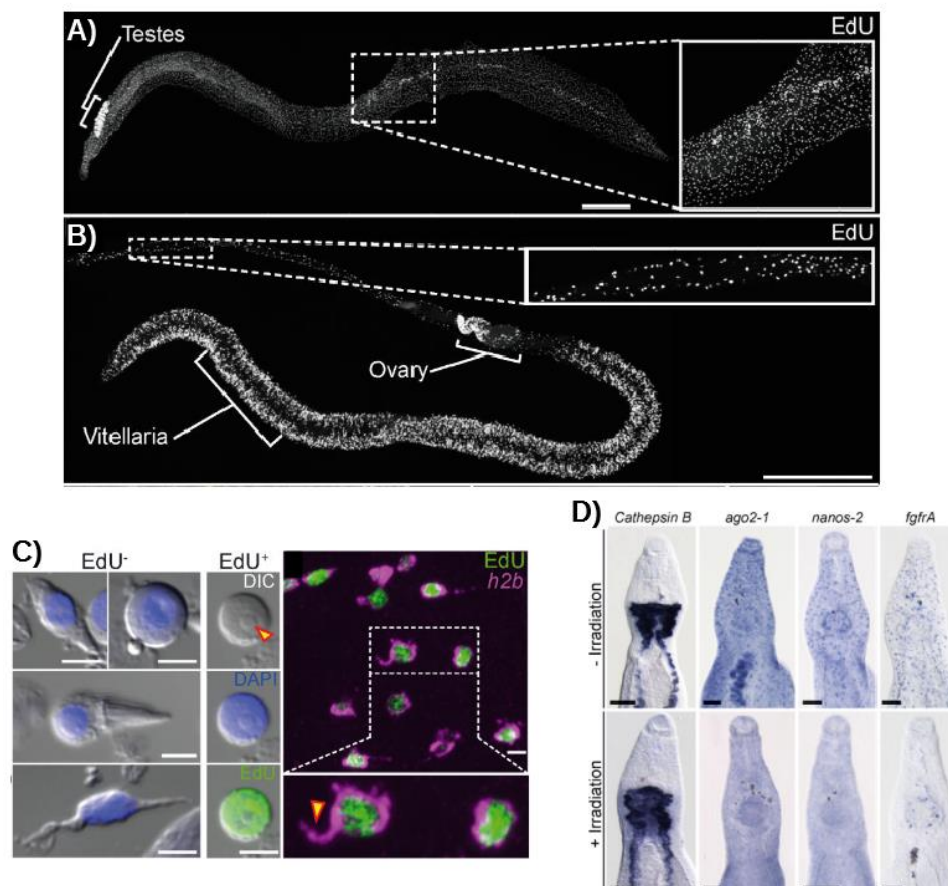


Figura I 12: Células tipo neoblasto del estadio adulto de *S. mansoni*. A y B): Células marcadas con EdU en un adulto macho (A) y hembra (B). La barra representa 500 μ m. C) Morfología de células tipo neoblasto aisladas del adulto. En verde se muestra la señal de EdU, en azul el núcleo (DAPI, izquierda, la punta de flecha señala el nucleolo) y en magenta la expresión de la histona *h2b* (izquierda, la punta de flecha señala extensiones citoplasmáticas). La barra representa 5 μ m. D) Efecto de la irradiación sobre la expresión de genes marcadores de células tipo neoblasto en el estadio adulto (*ago2-1*, *nanos-2* y *fgfrA*). La barra representa 100 μ m. Tomadas de Collins et al. (2013).

Como se mencionó anteriormente, los neoblastos de planaria expresan reguladores post-transcripcionales asociados al desarrollo de la línea germinal conservados en la mayoría de los animales (*piwi*, *tudor*, *vasa* y *nanos*). Los trematodos y los cestodos carecen en sus genomas de los genes *vasa* y *piwi*, así como también de las proteínas con dominio tudor

que interaccionan con Piwi (Tsai, IJ et al., 2013; Skinner et al., 2014). Estos genes son esenciales para el silenciamiento de los elementos transponibles en la línea germinal. Trematodos y cestodos, si bien carecen de los verdaderos ortólogos de estos genes, tienen genes relacionados que podrían estar reemplazando las funciones conservadas (dos ortólogos de *vasa/pl10* y una familia de *argonautas* específica de platelmintos). Para el cuarto gen conservado, *nanos*, sí se encuentran ortólogos en los genomas de trematodos y cestodos, y en el estadio adulto de *S. mansoni* se encuentra asociado a las células tipo neoblasto ya que su expresión disminuye en gusanos irradiados.

Los genes *SmfgfrA*, *Smago2-1* y *Smnanos-2* se expresan en células con una localización similar a la de las células tipo neoblasto y su expresión desaparece tras la irradiación (Figura I 12 D). En particular *SmfgfrA* se expresa en más del 99% de las células que incorporan EdU por lo que constituye un marcador de este tipo celular en el adulto de *S. mansoni*.

Estas células son capaces de auto-renovarse y parecen tener un proceso de división asimétrico, donde una de las células hijas es capaz de continuar proliferando mientras que la otra no. Además, son capaces de diferenciarse a diferentes tejidos, ya que en experimentos de pulso y caza, luego de 7 días las células que incorporaron EdU durante el pulso se localizaban en músculo e intestino, indicando que pueden diferenciarse a tipos celulares de distintas capas embrionarias, lo que habla de su multipotencia.

4.2.2. Las células tipo neoblasto en estadios larvarios de *S. mansoni*

En estadios larvarios, se caracterizaron histológicamente células con morfología similar a los neoblastos de planaria y a las células tipo neoblasto de los adultos (Wang et al., 2013). En la transición entre miracidio (estadio larvario de vida libre) y esporocisto madre (estadio larvario intra-caracol), se produce un aumento en la cantidad de células en proliferación 20 horas después de la transformación, siendo máxima aproximadamente a las 48 horas (Figura I 13 A). Una comparación de los niveles de expresión de genes entre miracidios y esporocistos 48 horas post-transformación permite identificar 331 genes con expresión superior en esporocistos y que presentan homología con marcadores de neoblastos de planaria. Entre ellos se destacan componentes del ciclo celular y de la maquinaria de reparación del ADN, así como factores asociados al mantenimiento de células madre y desarrollo de células germinales: *fgfr*, *sz-12*, *ezh*, *eed*, *p53*, *bruli*, *ago2* y 11 de las 37 helicasas *DEAD-box* presentes en el genoma de *S. mansoni* (entre ellas, 3 homólogos de *vasa/pl10*, denominadas *vlg-1* a *vlg-3*).

De los reguladores post-transcripcionales asociados al desarrollo de la línea germinal, *Smago2-1*, *Smvlg-3* y *Smnanos-2* aumentan su expresión en esporocistos, pero con cinéticas diferentes (Figura I 13 B): *ago2-1* aumenta constantemente durante el desarrollo temprano del esporocisto; *vlg-3* aumenta pronunciadamente tras la transformación y luego disminuye, mientras que *nanos-2* aumenta su expresión luego de 4 días de iniciada la transformación. El análisis de su localización revela que tanto *ago2* como *vlg-3* se expresan en la mayoría de las células en proliferación pero *nanos-2* es expresado en una subpoblación de estas células (Figura I 13 C). Esto sugiere la existencia de al menos dos poblaciones de células con capacidad proliferativa, caracterizadas por la presencia o ausencia de expresión de este gen. Las células que lo expresan tienen un ciclo celular más largo y replican su ADN con menor frecuencia que las células que no lo expresan (Figura I 13 D). Esta observación es consistente con el rol que se le asigna a *nanos* en células

germinales al reprimir transcritos mitóticos y alargando el ciclo celular.

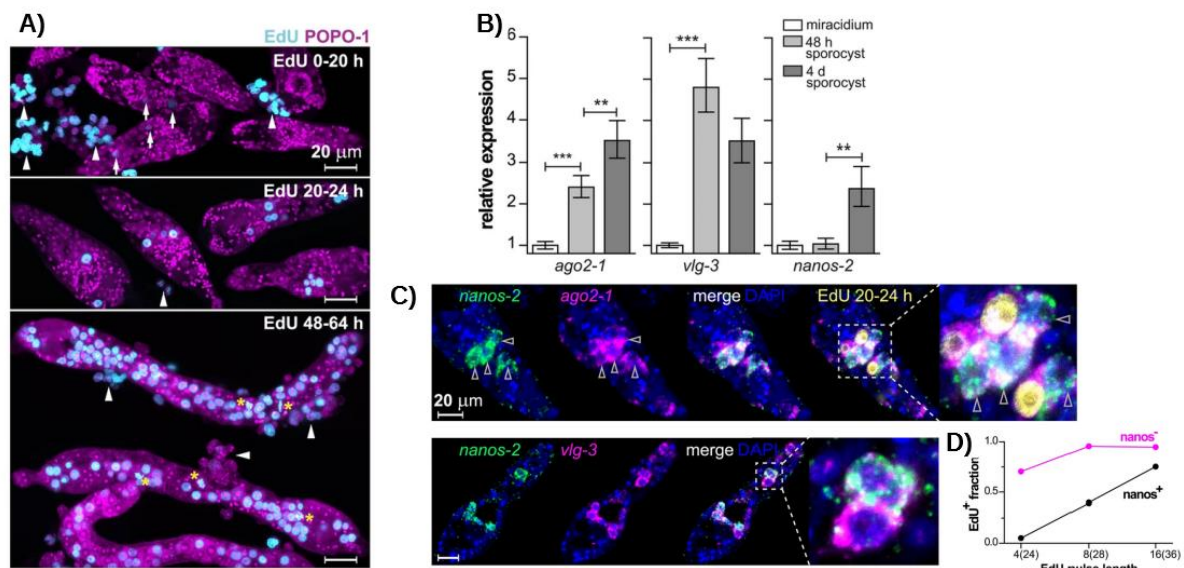


Figura I 13: Células tipo neoblasto en estadios larvares de *S. mansoni*. A) Aumento de la proliferación luego de la transformación de miracidio a esporocisto. En azul se muestra la señal de EdU y en magenta los núcleos. B) Niveles de expresión de los genes marcadores *ago2-1*, *vlg-3* y *nanos-2* durante la transformación de miracidio a esporocisto. Las barras representan la desviación estándar. ** $p < 0.01$. *** $p < 0.001$ (*t*-test). C) Co-expresión de *nanos-2* (verde) con *ago2-1* o *vlg-3* (magenta). Las puntas de flecha muestran células *nanos-2*⁺ que no incorporan EdU (amarillo). D) Cinética de incorporación de EdU en células *nanos-2*⁺ o *nanos-2*⁻. Tomadas de Wang et al. (2013).

La eliminación de *ago2-1* y *vlg-3* por interferencia de ARN reduce significativamente la cantidad de células capaces de incorporar EdU y la expresión de genes asociados al ciclo celular, indicando un rol en la regulación de la proliferación para estos genes. La eliminación de *vlg-3* produce la pérdida de todas las células tipo neoblasto, tanto las células *nanos-2*⁺ como las *nanos-2*⁻, sugiriendo que *vlg-3* está implicado en el mantenimiento de toda la población de células con capacidad proliferativa. Por el contrario, la eliminación de *ago2-1* sólo produce la pérdida de las células *nanos-2*⁻, mientras que permanecen las que sí lo expresan. Estas células sin embargo son incapaces de proliferar.

La eliminación de *fgfrA* también impide la proliferación y produce la disminución de genes asociados al ciclo celular, pero no afecta la expresión de *nanos-2*. Estos resultados son similares a lo observado en el adulto (Collins et al., 2013), indicando que *fgfrA* tiene un rol en ambos estadios en el control de la proliferación de las células tipo neoblasto.

4.2.3. Diferenciación de células tipo neoblasto de *S. mansoni*

Estudios recientes muestran que las células tipo neoblasto contribuyen a la renovación y mantenimiento del tegumento a través de células de vida corta que expresan el ARNm codificante para la Tetraspanina-2 (*tsp-2*, Smp_181530) (Collins et al., 2016). La proteína TSP-2 se encuentra asociada a la superficie del tegumento de *S. mansoni* y es estudiada como antígeno vacunal (Pearson et al., 2012). La expresión de este gen no disminuye inmediatamente luego de la irradiación, sino después de al menos 7 días, indicando que es un gen que se expresa en la progenie de las células tipo neoblasto (Collins et al., 2016). *tsp-2* no se expresa en los cuerpos celulares del tegumento sincitial sino en células muy cercanas a los mismos, por lo que las células *tsp-2*⁺ corresponderían a precursores

indiferenciados del tegumento. En experimentos de pulso y caza las células $\text{EdU}^+/\text{tsp-2}^+$ comienzan a aparecer 3 días después del pulso, indicando que estas células son postmitóticas. A partir de los 3 días comienzan a observarse células tegumentarias diferenciadas marcadas con EdU, mientras que 35 días después la marca desaparece, evidenciando un rápido recambio del tegumento. Mediante purificación por citometría de flujo de células que expresan TSP-2 en su superficie y secuenciación de su ARN se describen genes asociados al tegumento diferenciado y algunos genes que son compartidos con neoblastos de planaria. Entre ellos, existen dos genes con motivos de dedos de zinc (*zfp-1* y *zfp-1-1*) cuya interferencia disminuye considerablemente el número de células que expresan *tsp-2* y genes asociados al tegumento diferenciado. Estos genes parecen ser esenciales para la producción de células tegumentarias ya que al interferirlos disminuye la incorporación a largo plazo de nuevas células en el tegumento. Es posible que los efectos de estos genes en las células tsp-2^+ sean diferentes: *zfp-1* parecería actuar sobre las células tipo neoblasto para especificar progenitores tegumentarios tempranos tsp-2^+ , mientras que *zfp-1-1* actuaría en estos progenitores tempranos para controlar el destino de las células durante el compromiso con el linaje tegumentario (Figura I 14). Esto es similar a la organización celular de los linajes epidérmicos de planaria; a pesar de que el tegumento de *S. mansoni* y la epidermis de planarias son diferentes, los precursores que los generan son similares. En ambos casos son células que migran por el parénquima atravesando la capa muscular y la membrana basal para llegar a la superficie y además, dependen de miembros de la superfamilia de reguladores transcripcionales *zfp-1* (Wendt et al., 2018).

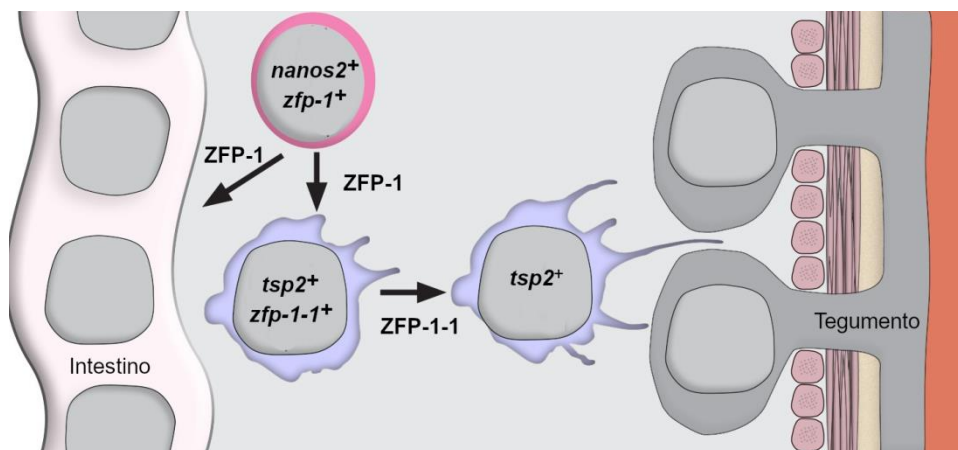


Figura I 14: Diferenciación del tegumento de *S. mansoni*. Modelo para la especificación de nuevas células tegumentarias. Las células tipo neoblasto que expresan *nanos-2* y *zfp-1* especifican células *tsp-2*⁺. Una fracción de esas células expresan *zfp-1-1*, las cuales migran para fusionarse con el tegumento. Adaptada de Wendt et al. (2018).

4.2.4. Clasificación de células tipo neoblasto de *S. mansoni*

Mediante *transcriptional profiling* se describió la heterogeneidad de las células tipo neoblasto de *S. mansoni* y se comenzaron a conocer las relaciones de este tipo celular en los estadios larvarios y el adulto (Wang et al., 2018).

Al aislar y caracterizar transcriptómicamente las células tipo neoblasto del esporocisto, se distinguen tres clases principales: células *klf*⁺/*nanos-2*⁺ denominadas κ , células *fgfrA,B*⁺ denominadas ϕ y células *nanos-2*⁺/*fgfrA,B*⁺ denominadas δ (Figura I 15 A a C). Estas células además expresan otros marcadores que contribuyen a la diferencias entre las clases: *p53* y *zfp-1* en las células δ y *hesl* en las células ϕ . Todas expresan *ago2-1* y existen

muy pocos genes específicos de cada clase, lo que indica que comparten un perfil transcriptómico común. La clase ϕ presenta la menor cantidad de genes específicos (Figura I 15 C)

Durante el desarrollo intra-caracol las proporciones de estas células van cambiando, observándose en la cercaria (estadío infectivo del hospedero mamífero) únicamente células de las clases δ y κ (Figura I 15 A), las que serían transferidas al estadío sexual que se desarrolla en el mamífero. En la schistosómula (estadío inmediatamente posterior al ingreso al hospedero mamífero) las células en proliferación son principalmente del tipo κ , lo que indicaría que esta subpoblación de células es la fuente de las células tipo neoblasto del adulto. En el estadío juvenil, estas células expresan varios transcriptos asociados a células tipo neoblasto del adulto, como *nanos-2*, *ago2-1*, *fgfrA*, *h2A*, *cyclinB* y *PCNA*, evidenciando la similitud entre las células tipo neoblasto de los juveniles y los esporocistos tanto morfológica como molecularmente. El análisis de la expresión de genes específicos de las clases κ , δ y ϕ del esporocisto, revela la existencia de dos clases de células en el juvenil, una que expresa marcadores de células δ y ϕ denominada δ' y otra denominada ϵ que expresa el gen *eled* (Figura I 15 D). Las células δ' expresan varios genes marcadores de las células tipo neoblasto del adulto (*nanos-2*, *fgfrA*, *p53*, *zfp* y *hesl*, entre otros), por lo que se propone que esta clase es la fuente de células tipo neoblasto del adulto. Por el contrario, las células ϵ presentan menores niveles de *nanos-2* y *fgfrA* y, de manera análoga a las células κ de esporocistos, no expresan *p53* ni *hesl*, por lo que se propone que las células ϵ podrían derivar de las células κ (Figura I 15 E).

El gen *eled*, marcador de las células ϵ y cuya expresión se restringe a los primordios de los órganos reproductivos, parece ser específico de trematodos, ya que no se encuentran ortólogos en los genomas de cestodos ni de planaria. Su expresión en los primordios genitales precede a la de *nanos-1* (marcador específico de la línea germinal en juveniles y adultos). Solo un subgrupo de células *eled*⁺ en los primordios genitales expresan *nanos-1*, y el número de células que los coexpresan aumenta con el tiempo de desarrollo. Esto sugiere que las células germinales podrían derivar de células ϵ presentes en el desarrollo temprano del juvenil y *eled* es el marcador identificado que se expresa más temprano en la línea germinal. Cuando se interfiere *eled* aumenta la expresión de *nanos-2* (que es expresado en células tipo neoblasto somáticas) en la zona de crecimiento posterior, donde se ubican las células ϵ , y se produce una acumulación prematura de esperma en los testes juveniles. Al interferir *nanos-1* o *nanos-2* se degeneran los testes y se pierden las células germinales diferenciadas, pero la parte de las gónadas que logra diferenciarse expresa *eled*. Esto indica que este gen inhibe la diferenciación de células germinales, mientras que los genes *nanos* son requeridos para este proceso. En este contexto, *eled* antagoniza con ambos homólogos de *nanos*: en el soma suprime la expresión de *nanos-2* en las células ϵ de la zona de crecimiento posterior, mientras que en la línea germinal inhibe la diferenciación de las células germinales.

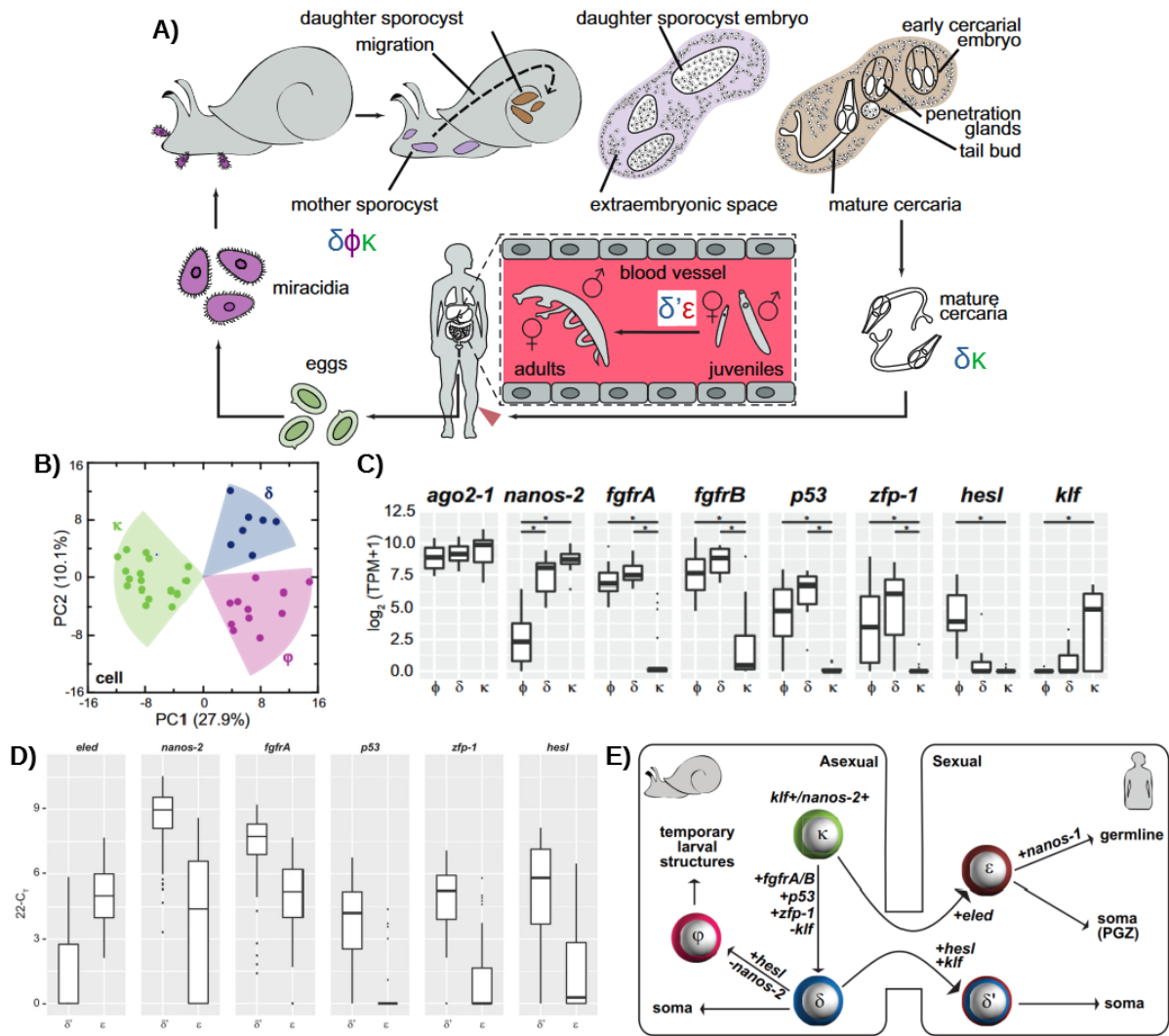


Figura I 15: Clases de células tipo neoblasto en estadios larvarios de *S. mansoni*. A) Ciclo de vida de *S. mansoni*. En cada estadio se muestran las clases de células que se identificaron (azul: células δ , verde: células κ y violeta: células ϕ en estadios intra-caracol, rojo: células ϵ y azul: células δ' en estadios intra-mamífero). B) Análisis de componentes principales del *scRNAseq* identificando las 3 clases de células tipo neoblasto en esporocistas. C) Niveles de expresión de genes marcadores que permiten distinguir las 3 clases de células tipo neoblasto en esporocistas. D) Niveles de expresión de los genes marcadores en el estadio juvenil que permite la identificación de las clases de células tipo neoblasto en estadios intra-mamífero. E) Modelo propuesto para las relaciones entre las clases de células tipo neoblasto. Las células κ , que son las que dan lugar a las demás clases, expresan *nanos-2* y *klf*. La activación de varios genes asociados al funcionamiento de células tipo neoblasto somáticas da lugar a la especificación de las células δ , que generarían los tejidos somáticos. La inactivación de *nanos-2* y la activación de *hes1* en las células δ dan lugar a la formación de las células ϕ , asociadas a estructuras transitorias de la larva. Un número reducido de células δ y κ es transmitido al estadio sexual. Tras la entrada al hospedero mamífero, la activación de *hes1* en las células δ da lugar a las células δ' que formarían los tejidos somáticos adultos. La inactivación de *nanos-2* en las células κ y la activación de *eled* generan las células ϵ . En estas células, la activación de *nanos-1* compromete a las células ϵ con la línea germinal, mientras que las que no activan *nanos* contribuyen al crecimiento somático del juvenil. Tomadas de Wang *et al.* (2018).

En base a estos resultados se propone un modelo para trazar el origen de las células tipo neoblasto en el adulto y de la línea germinal a partir de las células derivadas de la larva (Figura I 15 E). En este modelo, en el estadio larvario las células κ sirven como células

madre embrionarias, en las que la activación de varios genes asociados a la función de células tipo neoblasto en adultos lleva a la especificación de las células δ , responsables de generar los tejidos somáticos. Por otro lado, la disminución de la expresión de *nanos-2* y la activación de *hes1* llevaría a que las células κ se transformen en células ϕ , las que se encuentran restringidas a partes transitorias de la larva por lo que no serían transmitidas al estadio intra-mamífero. Unas pocas células δ en la cercaria serían las que dan lugar a las células tipo neoblasto del adulto y son las primeras en proliferar luego de la transformación a schistosómula, transformándose en células δ' . Estas serían las responsables de remodelar los tejidos de la larva mientras migra hacia la vena portal y serían necesarias para el crecimiento y maduración posterior. La cercaria también tiene un par de células κ en el lugar en el que se encontraría el primordio genital, en las cuales no se observa proliferación hasta 9 días después de la transformación. Se postula que la disminución de la expresión de *nanos-2* y la activación de *eled* en las células κ define a las células ϵ específicas de estadios intra-mamífero. La ausencia de marcadores de células δ y δ' (*fgfr*, *zfp-1*, *p53* y *hes1*) en las células κ y ϵ apoya la hipótesis de que forman un linaje separado. Durante la maduración sexual, las células ϵ gonadales activan la expresión del marcador específico de línea germinal *nanos-1* lo que llevaría a su compromiso con la línea germinal. Aún resta determinar el mecanismo por el cual las células *eled*⁺ deciden entre la generación de la línea germinal (células *nanos-1*⁺) o células somáticas posteriores (*nanos-1*). El desarrollo de técnicas de marcado de linaje permitirá confirmar esta teoría sobre las transiciones sobre los distintos tipos de células tipo neoblasto.

Estos resultados son similares a los descritos para los neoblastos de planaria en cuanto a la heterogeneidad de la población de neoblastos, con genes característicos co-regulados en cada subpoblación. En particular, los genes *fgfr*, *zfp-1* y *p53* son marcadores de linaje epitelial en planaria (van Wolfswinkel et al., 2014) y en *S. mansoni* se expresan abundantemente en las células δ y δ' , que es la población principal de células tipo neoblasto somáticas. Durante el desarrollo embrionario de planaria se describe una transición similar, inactivando genes específicos del embrión y activando genes asociados al desarrollo del adulto (Davies et al., 2017).

4.2.5. Células tipo neoblasto en *Fasciola hepatica*

Otro trematodo en que se ha comenzado a identificar células con capacidad proliferativa es *Fasciola hepatica*. En un estudio tratando de optimizar el crecimiento de juveniles *in vitro*, se describe por primera vez a las células tipo neoblasto de *F. hepatica* como las responsables del crecimiento del parásito (McCusker et al., 2016). En el juvenil, estas células tienen morfología similar a los neoblastos de planaria (redondeadas, gran núcleo, nucleolo prominente y escaso citoplasma), presentan patrones migratorios similares a los observados para las células madre de otros platelmintos y son las responsables del crecimiento observado *in vitro*. Están distribuidas en los dos tercios posteriores del juvenil, en el parénquima por debajo del tegumento y de las capas musculares, en la región posterior al ganglio cerebral. Solo se observa un aumento de la cantidad de células en proliferación en gusanos que presentan crecimiento en medio suplementado con suero de pollo contrariamente a los organismos incubados en medio no suplementado. Además el tratamiento con hidroxíurea, que bloquea específicamente la proliferación, impide el crecimiento de los juveniles aún en medio suplementado, por lo que las células tipo neoblasto serían las responsables del crecimiento observado. Estas células además son capaces de migrar, ya que en experimentos de pulso y caza con EdU las células EdU⁺

luego de 6 días se encuentran en regiones más externas del gusano e incluso en la región anterior.

En un estudio más reciente, se analizó el transcriptoma y el proteoma durante las primeras 24 horas luego del desenquiste de la metacercaria dando lugar a los juveniles (NEJ: *newly excysted juvenile*), que constituye uno de los primeros pasos de la infección al hospedero mamífero (Cwiklinski et al., 2018). Durante este proceso el parásito es capaz de aumentar la expresión de determinados genes rápidamente para dar respuesta a los cambios en el ambiente y facilitar la invasión del hospedero. Entre otros efectos, se observa una gran proliferación, lo que es consistente con el crecimiento acelerado que presenta el parásito al establecerse en su hospedero mamífero. Un análisis de genes asociados a células tipo neoblasto de *S. mansoni* muestra que los genes *argonauta 2* (*ago2.1* y *ago2.2*), *nanos* e histona *h2a* se expresan constitutivamente en el parásito en los estadios comprendidos entre NEJ y adulto pero presentan una mayor expresión en los juveniles. Por otro lado, el gen de la histona *h2b* aumenta considerablemente su expresión en el estadio juvenil, mientras que en los otros estadios analizados no se expresa a niveles considerables. Este fenómeno comienza 24 horas luego del desenquiste, lo que sugiere que el organismo se está preparando para la migración hacia el hígado donde se da su posterior crecimiento. El aumento en los niveles de expresión de marcadores de células tipo neoblasto está asociado a la proliferación celular en los juveniles, produciendo un aumento en el número de este tipo celular dos días luego del desenquiste.

4.3. Las células germinativas de cestodos

Kozioł y Castillo (2011) realizaron una revisión sobre la caracterización de la proliferación celular en cestodos. Mediante estudios ultraestructurales y de histología clásica se ha descrito a las células indiferenciadas como células redondeadas con una alta relación núcleo/citoplasma, un citoplasma altamente basofílico y electrodensito debido a la abundancia de ribosomas libres, con pocas o ninguna extensiones citoplasmáticas, con escaso o nulo retículo endoplasmático y aparato de Golgi, un gran núcleo con un nucleolo prominente y poca heterocromatina. Esta morfología es muy similar a la descrita para neoblastos de planaria, salvo por la ausencia de cuerpos cromatoides en cestodos. Existen variaciones a estas características generales, principalmente en el tamaño de la célula y la relación núcleo/citoplasma que se han propuesto como características de distintos subtipos de células germinativas, aunque en el caso de cestodos aún no hay evidencia que lo pruebe.

En base a la evidencia existente para diferentes cestodos, se observan otras características comunes además de las morfológicas. En todos los organismos estudiados las células capaces de proliferar son células indiferenciadas. La distribución en el cuello y los proglótides de los adultos también parece ser común. Las células germinativas se encuentran principalmente en la región externa del parénquima medular, cercanas a la capa muscular interna. En el caso de existir células germinativas en el parénquima cortical, éstas se encuentran restringidas a la región más interior. Durante el desarrollo del sistema reproductivo en cada proglótide se forma un primordio genital en el centro del parénquima medular que posteriormente se desarrolla generando una región compacta que no prolifera en el interior y una región exterior que continúa proliferando (Revisado por Kozioł y Castillo, 2011).

4.3.1. Células germinativas en *Echinococcus multilocularis*

En el caso de *E. multilocularis*, se caracterizaron las células germinativas en el metacestodo (estadio larvario) y en cultivos primarios de células (Koziol et al., 2014). Éstas son las únicas células capaces de proliferar y son las responsables del crecimiento del metacestodo y la regeneración. Además se analizó la expresión de varios marcadores conservados en otros platelmintos observando diferencias en los perfiles de expresión.

4.3.1.1. Caracterización celular

Durante el desarrollo larvario, el 5.9% de las células de la capa germinal incorporan el análogo de timidina EdU. Estas células se acumulan en la cápsula prolígera donde se formarán cada protoescólex y en la región posterior del mismo durante su desarrollo (Figura I 16). Una vez culminado el desarrollo, cuando el protoescólex se invagina, la cantidad de células marcadas disminuye considerablemente, mientras que cuando éste se vuelve a activar *in vitro* al imitar las condiciones de ingesta del hospedero definitivo, la proliferación aumenta nuevamente. Esto indica que durante el desarrollo, existen células capaces de proliferar, pero se mantienen en un estado quiescente o con una cinética lenta mientras el protoescólex se encuentra dentro del metacestodo.

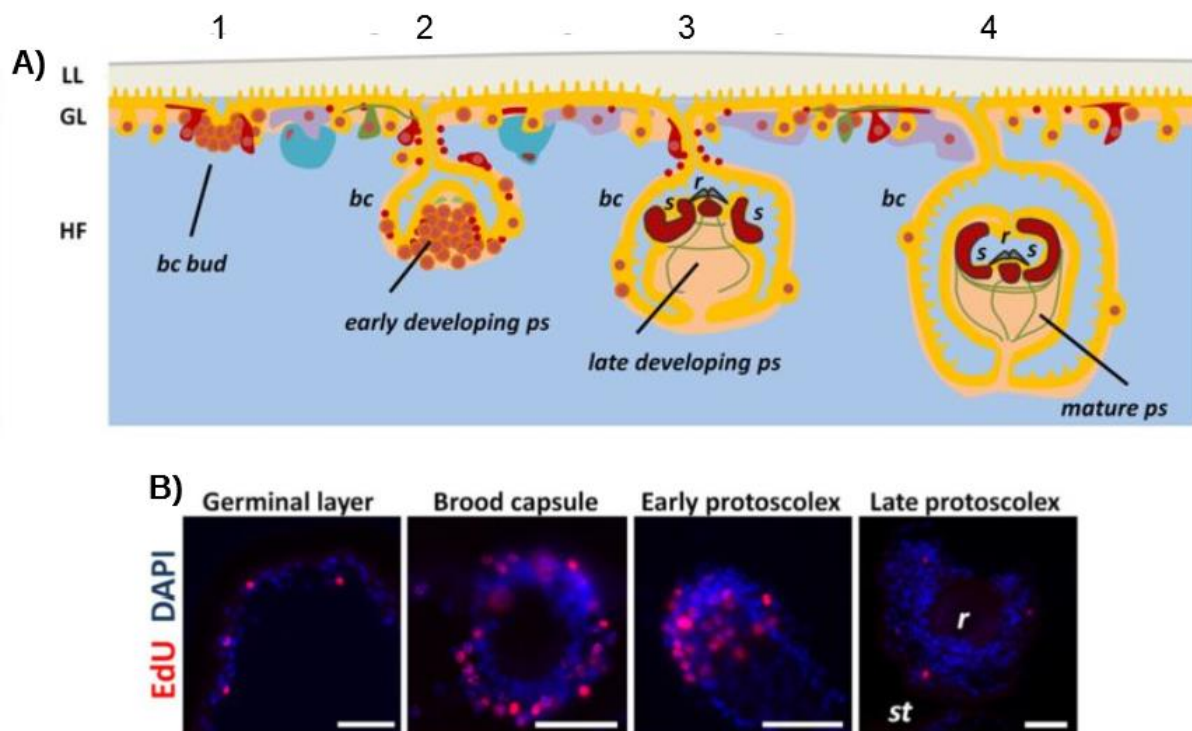


Figura I 16: Proliferación en *E. multilocularis*. A) Esquema mostrando la organización y desarrollo del metacestodo. 1: Cápsula prolígera temprana, 2: Cápsula prolígera con un protoescólex comenzando su desarrollo, 3: Cápsula prolígera con un protoescólex desarrollado, 4: Cápsula prolígera con un protoescólex invaginado. Anaranjado: tegumento, Marrón: células germinativas, Verde: células nerviosas, Rojo: fibras musculares, Violeta: células de almacenaje de lípidos o glucógeno, Celeste: corpúsculos calcáreos. bc: Cápsula prolígera, GL: capa germinal. HF: Líquido hidático, LL: Capa laminar, ps: protoescólex, r: rostellum, s: ventosa. B) Detección de EdU en diferentes estadios del desarrollo. r: rostellum, st: tallo. La barra representa 30 µm. Tomada de Koziol et al. (2014).

Las células germinativas presentan la morfología clásica de células madre de platelmintos: tamaño pequeño entre 5 y 12 μm , forma redondeada o fusiforme, con citoplasma basofílico que presenta en algunos casos extensiones saliendo de los polos y un gran núcleo con uno a tres nucleolos prominentes y heterocromatina finamente granular. Estas células son las únicas capaces de proliferar y que se observan en mitosis. Las células diferenciadas derivan de las células germinativas, ya que en experimentos con un pulso corto de EdU las células diferenciadas no incorporan la marca, pero en experimentos de pulso y caza o de marcado continuo sí se observa la marca. De todas maneras, no se puede descartar la existencia de dediferenciación o transdiferenciación.

4.3.1.2. Marcadores moleculares

En busca de marcadores moleculares para las células germinativas de *E. multilocularis* se analizaron diferentes genes homólogos a marcadores de neoblastos de planaria. El perfil de expresión de la histona *h2b* es muy similar al perfil de incorporación de EdU en la capa germinal y durante el desarrollo de cada protoescólex (78% de células *h2b*⁺ son EdU⁺ y 87% de las células EdU⁺ son *h2b*⁺), indicando que este gen es un buen marcador de células en fase S, pero no permite identificar células germinativas en otras fases del ciclo celular.

Existen dos copias del regulador post-transcripcional *nanos* que es un gen marcador de células de la línea germinal. Ambos genes se expresan en unas pocas células en la capa germinal que presentan la morfología de células germinativas y que son capaces de incorporar EdU (19% de las células son *nos1*⁺/EdU⁺). Sin embargo, la gran mayoría de las células EdU⁺ no expresan ninguno de los dos genes (más del 95% de las células EdU⁺) lo que indica que una subpoblación pequeña de células germinativas en la capa germinal expresa *nos1* y *nos2*, aunque no es claro si ambos genes se co-expresan. Durante la formación de cada protoescólex, *nos1* no es detectado, mientras que *nos2* se observa en algunas ocasiones alrededor de la cápsula prolígera y en una pequeña población de células en la base del brote del protoescólex. Además *nos2* es expresado en algunas células asociadas al desarrollo del sistema nervioso. Esto indica que las células germinativas no expresan *nanos* durante el desarrollo del protoescólex y que *nos2* parece tener un rol en el desarrollo del sistema nervioso.

Aunque los genes *piwi* no están presentes en los genomas de cestodos y trematodos, sí existen otras proteínas Argonauta, en particular un grupo que es específico de cestodos y trematodos. *E. multilocularis* tiene tres genes muy similares pertenecientes a este grupo: *em-ago2-A*, *em-ago2-B* y *em-ago2-C*, además de un pseudogen (*em-ago2-psi*). El patrón de expresión de los genes *em-ago2* es muy similar al patrón de incorporación de EdU en la capa germinal y células *ago2*⁺ se acumulan en la cápsula prolígera y en los brotes del protoescólex. En general, el número de células que expresan *ago2* es alto, llegando al 50% de las células en algunos casos, pero solamente la mitad de las células EdU⁺ son *ago2*⁺. Esto indicaría que su expresión no es específica de la población de células germinativas.

La expresión de *ago2*, *nos1* y *nos2* indica una gran heterogeneidad a nivel molecular en las células germinativas en proliferación. Por otro lado, la desacetilasa de histonas *hdac* y dos homólogos de *prohibitin* son expresados específicamente en neoblastos de planarias pero en el caso de *E. multilocularis*, ninguno de estos genes parece tener una expresión específica de células germinativas. Estos resultados indican que, si bien las células germinativas de *E. multilocularis* son similares morfológicamente entre sí, existen subpoblaciones que expresan distintos genes marcadores. Además, muchos de estos

genes, si bien tienen expresión restringida a células germinativas, presentan perfiles de expresión diferentes a los reportados para las células indiferenciadas de otros platelmintos.

Se identificó un elemento genético móvil, de la familia de los retrotransposones LTR no autónomos (*terminal-repeat retrotransposons in miniature*, TRIMs), que es específico de Taenidos y se expresa durante todo el ciclo de vida de *E. multilocularis*. Este es específicamente expresado en el 94.3% de las células en fase S del ciclo celular, indicando que se expresa en la gran mayoría de las células germinativas (Koziol et al., 2015). Más recientemente, se probó que el gen *Em-sox2* también se expresa en la mayoría de las células germinativas y es capaz de reemplazar la función de su homólogo en ratones en la inducción de células madre pluripotentes a partir de células somáticas, lo que indica que es posible que este gen sea esencial para el mantenimiento de las células germinativas de *E. multilocularis* (Cheng, Liu, Dai, et al., 2017). Estos genes parecerían ser candidatos a marcadores generales de las células germinativas de *E. multilocularis*.

Una de las quinasas tipo Polo presentes en el genoma de *E. multilocularis* (*em-plk1*) presenta una expresión restringida a las células germinativas, y el uso de un inhibidor de esta quinasa elimina estas células (Schubert et al., 2014). El factor de crecimiento epidérmico humano (EGF) activa la proliferación de las células germinativas, indicando que la vía de señalización EGFR/ERK contribuye a la regulación de este tipo celular. Inhibidores de EGFR o de MEK/ERK impiden la proliferación de las células y el crecimiento de la larva (Cheng, Liu, Li, et al., 2017). Estos genes son interesantes blancos para el desarrollo de drogas antiparasitarias ya que se impediría el crecimiento de la larva en su hospedero.

4.3.1.3. Eliminación de células germinativas

También se estudió el efecto de la irradiación y el tratamiento con hidroxurea (que es tóxico para células que están sintetizando ADN) sobre las células germinativas de *E. multilocularis*. La irradiación con una dosis comparable a la aplicada a otros platelmintos (50 - 100 Gy) solamente retarda el crecimiento de los metacestodos, pero no elimina completamente las células germinativas. Una dosis de 150 Gy produce una disminución del 22% de las células que incorporan EdU en la capa germinativa luego de 48 horas de recuperación, por lo que en el caso de este organismo una dosis que es letal en planarias no es capaz de eliminar todas las células con capacidad de proliferación. Además no se observa un aumento en la cantidad de células EdU⁺ luego de la irradiación (analizado hasta 48 días post-irradiación), así como tampoco un cambio en la supervivencia del metacestodo. Por el contrario, el tratamiento con hidroxurea 40 mM reduce en un 90% las células que incorporan EdU en cultivos primarios de células y en vesículas sin ningún protoescólex ni cápsula prolífera. Además, la regeneración de las vesículas a partir de células primarias se elimina completamente. Este efecto es más marcado en vesículas pequeñas y jóvenes, donde la eliminación alcanza entre 97.7 y 99.8%. El análisis de macerados celulares muestra que la disminución de células EdU⁺ es acompañado por una disminución de células germinativas (de 20-22% a 3-5% de todas las células), mientras que las células diferenciadas no parecen ser afectadas. Las vesículas más grandes y viejas no son capaces de restablecer la población inicial de células germinativas (al menos no luego 22 días de recuperación), pero las vesículas jóvenes y pequeñas sí lo logran, generando acumulaciones puntuales de células EdU⁺ sugiriendo un crecimiento clonal de las células germinativas que sobrevivieron. Esto indica que al menos una población de células germinativas es capaz de autorenovarse, como se observa en planarias luego de la disminución en la cantidad de neoblastos con una

dosis sub-lethal. Es interesante que una reducción menor no produce un aumento de la proliferación para recuperar la cantidad de células germinativas, lo que indicaría que un número bajo de éstas es suficiente para la renovación de tejidos en vesículas, aunque se requiere una cantidad mayor para que ocurra crecimiento (Koziol et al., 2014).

4.3.1.4. Diferencias con otros platelmintos

La principal diferencia entre las células germinativas de *E. multilocularis* y los neoblastos de planaria es la ausencia de cuerpos cromatoides y de genes esenciales para el funcionamiento de los neoblastos, como *piwi* y *vasa*. Otros genes que son específicos de neoblastos de planaria (como *hdac* y *phb1*) si bien están en el genoma de cestodos, se expresan en tipos celulares diferentes. Más allá de las diferencias en la expresión génica, también son diferentes funcionalmente ya que en platelmintos de vida libre, los neoblastos que sobreviven a la irradiación subletal rápidamente proliferan para reemplazar a los neoblastos eliminados, lo cual no ocurre en *E. multilocularis* (Koziol et al., 2014).

nanos, un marcador clásico de la línea germinal, se expresa únicamente en células germinales de planaria pero no en los neoblastos somáticos muy similares morfológicamente. En el caso de *E. multilocularis*, las células germinativas que expresan *nanos* no serían células germinales ya que no hay expresión de *nos1* durante la formación de la cápsula prolígera ni de los protoescólex, mientras que *nos2* no se expresa en la formación de la cápsula prolígera y parece estar involucrado en la formación del sistema nervioso, algo que ya se ha visto para otros organismos.

En *S. mansoni*, el gen *ago2-1* se expresa en todas las células en proliferación en el esporocisto, mientras que *nanos-2* se expresa en una sub-subpoblación (Wang et al., 2013). En el adulto ambos genes se expresan en la mayoría de las células tipo neoblasto (Collins et al., 2013), a diferencia de lo que ocurre en *E. multilocularis*. Esto refleja las diferencias en el repertorio de genes expresados por las células germinativas de cestodos y las células tipo neoblasto de trematodos (Koziol et al., 2014).

4.3.2. Proliferación en Mesocostoides corti.

Los primeros estudios sobre proliferación en *M. corti* fueron realizados por Hess en el estadio larvario (1975, 1980, 1981; citados por Koziol et al., 2010; y Koziol y Castillo, 2011). En estos trabajos se describieron células en el parénquima con características de células proliferantes que pueden diferenciarse a células musculares del parénquima pero no a células tegumentarias ni a células musculares subtegumentarias. Además se describió “el macizo apical” como un sincitio reticulado de células ubicado entre las ventosas donde ocurre proliferación celular y se detectaron mitosis. En esta región es donde se inicia la división asexual de la larva y por ello Hess propuso que estas células proliferantes son la fuente de las células durante la reproducción asexual. Smith y McKerr (2000), mediante la incorporación del análogo de timidina BrdU, marcaron células en fase S del ciclo celular en el estadio larvario. Observaron la presencia de células proliferantes distribuidas por todo el parénquima, pero no observaron una acumulación especial entre las ventosas como reportó Hess. Espinoza et al. (2007) marcaron las células en proliferación con timidina tritiada y analizaron su distribución en la larva y durante la segmentación. Existen dos momentos de alta proliferación durante el desarrollo: una inmediatamente después de la inducción de la segmentación y otra cuando comienza la elongación y segmentación. Además el perfil de expresión de la histona *h4* es consistente con la distribución de células en fase S.

Nuestro grupo de trabajo ha realizado una extensa caracterización de las células proliferantes de *M. corti*. Koziol *et al.* (2010) estudiaron estas células durante el desarrollo estrobilar *in vitro* desde larva hasta gusano segmentado joven (no completamente maduro). Para determinar la localización de células proliferantes y cómo éstas contribuyen a la formación del primordio genital, se realizó marcado con BrdU con diferentes tiempos de incubación. Las células proliferantes se encuentran distribuidas en el parénquima medular, principalmente en la periferia junto a la capa muscular interna, formando un anillo (Figura I 17 A). En el escólex presentan una distribución similar, junto a las capas musculares y entre de las ventosas. No se encuentran células en proliferación en la región del parénquima cortical (Figura I 17 B), pero las células proliferantes son capaces de migrar a esa región, alcanzando la región subtegumentaria entre 24 y 72 horas después de ser marcadas. Esto indica que la renovación y el crecimiento de estos tejidos se deben a la migración de células proliferantes desde el parénquima medular. En la larva existe un gradiente anteroposterior de células proliferantes, siendo más abundantes en la región del escólex y el cuello, donde van a originarse los nuevos segmentos durante el desarrollo estrobilar. La región posterior, que no participa en el desarrollo, posee muy pocas células proliferantes (Figura I 17 C). En particular, tampoco se encontró una abundancia mayor de células en la región del macizo apical descrita por Hess como la fuente de células subtegumentarias. De hecho, organismos decapitados son capaces de proliferar y las células proliferantes migran a la región subtegumentaria desde el parénquima medular. Estos resultados indican que el macizo apical no tiene un rol preponderante durante el desarrollo estrobilar, pero no descartan que cumpla una función importante durante la división asexual. Durante la segmentación, las células proliferantes se acumulan en el centro de cada proglótide constituyendo el primordio genital (Figura I 17 D a H), el cual crece por proliferación *in situ* de las células más externas, mientras que las células internas no proliferan. En proglótides más desarrollados, este primordio se divide en dos regiones que formarán los rudimentos masculino (región anterior) y femenino (región posterior) y aparecen también los primordios de los testes a los lados del primordio, donde se observa también una gran proliferación. Estas características observadas para las células proliferantes de *M. corti* son similares a las reportadas para otros cestodos.

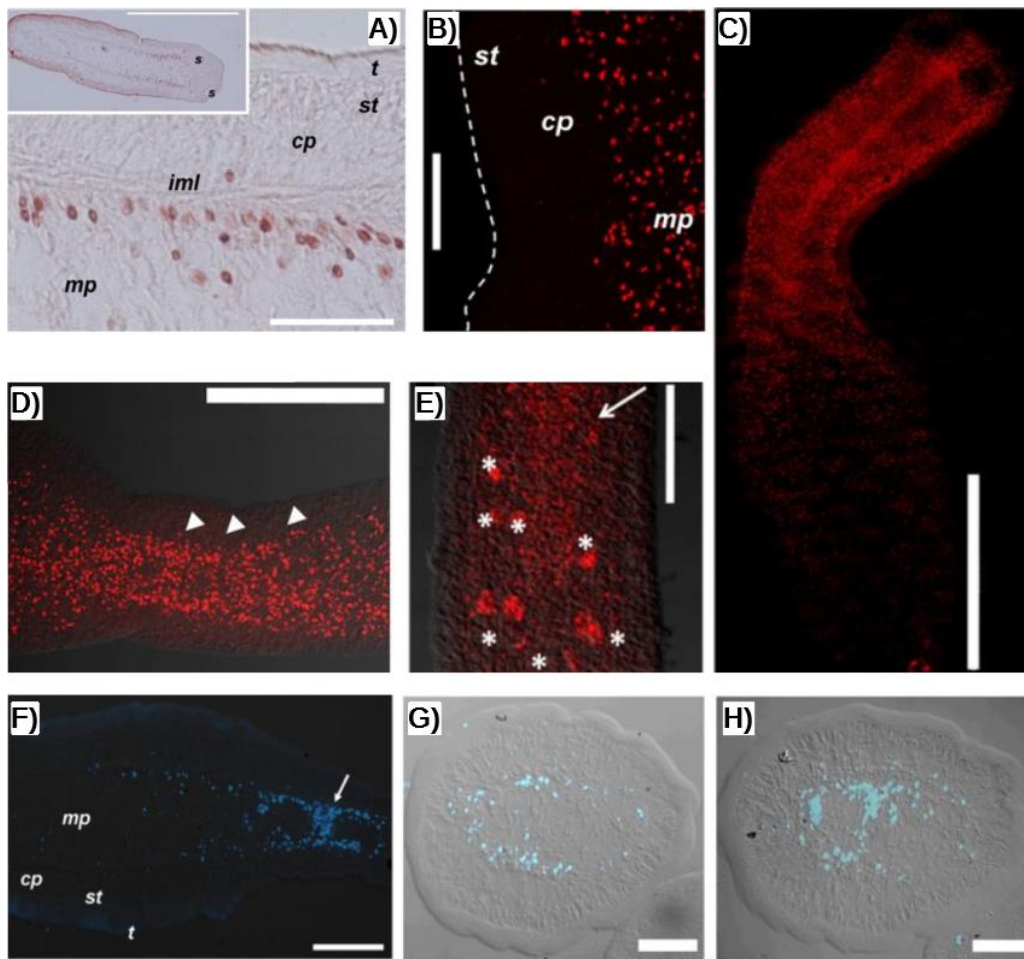


Figura I 17: Células proliferantes de *M. corti* marcadas con BrdU (pulso de 4 horas en A y de 24 horas en B a H). A) Corte sagital de la larva. B) Detalle de la región subtegumentaria. La línea punteada representa el límite del cuerpo de la larva. C) Vista general de la larva, mostrando el gradiente antero-posterior de células proliferantes. D) Segmento anterior de gusano comenzando la segmentación. Las flechas señalan acumulaciones periódicas de células marcadas. E) Segmento posterior de gusano comenzando la segmentación. La flecha señala el primordio genital y los asteriscos los primordios de los testes. F) Corte sagital del segmento más posterior de un gusano comenzando la segmentación. La flecha marca el primordio genital. G y H) Corte transversal de segmentos anteriores. Las barras representan 200 µm (vista general) y 20µm (detalle) en A, 50 µm en B, D, G y H, 1000 µm en C y 100 µm en E y F. cp: parénquima cortical, mp: parénquima medular, st: subtegumento, t: tegumento, iml: capa muscular interna. Tomadas de Koziol *et al.* (2010).

Más recientemente María Fernanda Domínguez, en su tesis de doctorado, continuó con el estudio de las células proliferantes. Se logró aislar y caracterizar células en las distintas etapas del ciclo celular mediante citometría de flujo (Domínguez *et al.*, 2014) (Figura I 18 A). La mayoría de las células se encuentran en fases G0/G1 (92%), mientras que las restantes se dividen casi en las mismas proporciones entre las fases S (3.7%) y G2/M (3.3%) (Figura I 18 C). Las células en las fases G0 y G1 del ciclo celular presentan morfología y tamaño heterogéneo, mientras que las células en fase G2 y M son homogéneas en forma y en tamaño (aproximadamente 5 µm de diámetro), observándose las características descritas para las células proliferantes de cestodos (gran núcleo y escaso citoplasma) (Figura I 18 B). Ensayos de cultivo *in vitro* de células obtenidas a partir de suspensiones celulares de *M. corti* permiten identificar células capaces de proliferar tras 20 y 72 horas de cultivo (Domínguez, 2016).

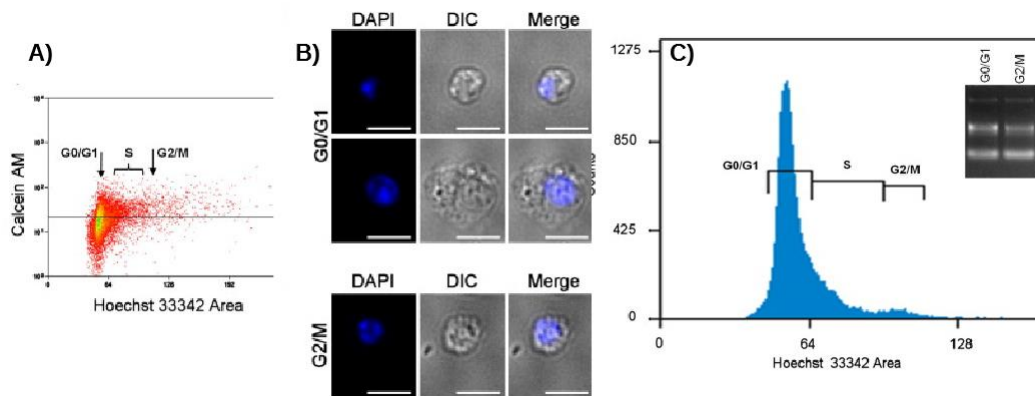


Figura I 18: Purificación de células proliferantes de *M. corti* por citometría de flujo. A) Plot de células teñidas con *Calcein AM* y *Hoechst 33342*. B) Morfología de células en las fracciones G0/G1 y G2/M. C) Histograma de células vivas teñidas con *Hoechst 33342* mostrando la cantidad de células en cada fase del ciclo celular. Se muestra el ARN aislado de las fracciones G0/G1 y G2/M. Tomadas de Domínguez et al. (2014)

4.3.2.1. Genes marcadores

Durante los últimos años nuestro grupo ha aislado varios candidatos a genes marcadores de células proliferantes de *M. corti*, en base al conocimiento de genes caracterizados en neoblastos de planaria y células madre en general. Estos genes incluyen *PCNA* (Caurla, 2015), *Nanos* (Bizzozero, 2010), *Pumilio 1*, *Pumilio 2* (Koziol et al., 2008; Koziol, 2009; Domínguez, 2016) y *pL10* (Domínguez, 2016).

4.3.2.1.1. PCNA: Proliferating Cell Nuclear Antigen

PCNA es ampliamente conocido por su rol en la replicación del ADN, como abrazadera de la ADN polimerasa que aumenta su procesividad y la coordinación de la síntesis de ambas hebras del ADN. Además cumple roles muy importantes durante la reparación del ADN, control del ciclo celular, ensamblado de la cromatina y regulación de la transcripción (Strzalka y Ziemienowicz, 2011). De acuerdo a su función es una proteína que se expresa en células en proliferación, y constituye un buen marcador de células en fase S del ciclo celular y por lo tanto de células proliferantes.

En planarias, este gen es expresado en neoblastos y es utilizado como marcador de este tipo celular (Orie et al., 2005; Salvetti et al., 2009).

4.3.2.1.2. Pumilio

Las proteínas Pumilio forman parte de la familia de proteínas PUF de reguladores post-transcripcionales que se unen al ARN mediante el dominio conservado PUM-HD. Este dominio reconoce en la región 3'UTR de su ARN blanco la secuencia UGUANAUA conocida como PBE (*PUM Binding Element*) o PRE (*PUM Response Element*). Este gen fue descubierto en *Drosophila*, donde se une a la región 3' UTR del ARNm de *hunchback* y controla la embriogénesis. El homólogo de *C. elegans* (FBF) se une de manera análoga al ARNm del gen *fem-3* y controla la gametogénesis.

La unión de proteínas Pumilio a su secuencia blanco reprime la expresión del gen al disminuir la cantidad de ARNm y de proteína. Si bien su rol más conocido es en la represión

de la traducción, existe evidencia que indica que Pumilio también puede activar la traducción de ciertos mensajeros. Además, el ARN largo no codificante (ARNInc) *NORAD*, activado frente al daño al ADN, modula la actividad de proteínas Pumilio actuando como un inhibidor competitivo. La acción de Pumilio puede regularse por la unión a distintas proteínas, de lo cual va a depender la actividad y la especificidad del blanco. En particular, en mamíferos se une a Nanos, Brain Tumor y CPEB y también se ha reportado que se une a miembros de la superfamilia de proteínas de unión al ARN DAZ/Boule, las cuales tienen un rol muy importante en la línea germinal. Pumilio también se asocia con proteínas Argonauta, el componente principal de la maquinaria de silenciamiento por miARNs. Esto, sumado a la abundancia de sitios de unión a miARN cerca de los PRE, indica que Pumilio puede estar implicado en la regulación por miARNs (Revisado por Goldstrohm et al., 2018).

Entre las principales funciones biológicas de las proteínas Pumilio se encuentran su rol en el mantenimiento de las células madre, desarrollo y crecimiento, así como también en la línea germinal y en el sistema nervioso. Están asociadas al control del destino de las células madre, siendo su principal función la de mantener su proliferación. En células madre embrionarias de mamífero son requeridas para salir de la autorrenovación, promoviendo la diferenciación al reprimir factores de transcripción asociados a la pluripotencia. También cumplen este rol en el destino de neuronas, línea germinal y células madre hematopoyéticas (HSC) y en organismos como *Drosophila*, *C. elegans* y mamíferos regulan la gametogénesis. Durante el desarrollo y funcionamiento del sistema nervioso, las proteínas Pumilio también juegan un importante rol. En mamíferos, ambas proteínas Pumilio se expresan en células madre neurales, progenitores y neuronas maduras durante el desarrollo del cerebro. En ratones, las proteínas Pumilio parecen promover la proliferación y la diferenciación neuronal, además de evitar la apoptosis (Goldstrohm et al., 2018).

En la planaria *Dugesia japonica*, el ARNm de Pumilio se expresa en los neoblastos y en menor medida en el ganglio nervioso anterior. Es esencial en este organismo para la regeneración tras la amputación y la renovación de los tejidos en condiciones de homeostasis (Salveti et al., 2005). Los platelmintos parásitos poseen dos genes *pumilio* bastante similares mientras que los de vida libre poseen una única copia, indicando que posiblemente hubo una duplicación reciente de *pumilio* en los platelmintos parásitos (Koziol et al., 2008). En *M. corti*, ambos genes presentan un perfil de expresión similar, tanto en células distribuidas por el parénquima (que podrían ser células proliferantes) como en los primordios de los testes y en el primordio genital, siendo la expresión más fuerte en la región exterior, coincidente con la localización de células proliferantes en esa zona. Esto indica un patrón de expresión en células proliferativas de todo el organismo aunque no es exclusivo de las mismas (Koziol, 2009).

4.3.2.1.3. Nanos

Nanos es un gen originalmente descubierto en *Drosophila melanogaster* con una función esencial en la determinación del eje antero - posterior, formación del abdomen y desarrollo de células germinales (Irish et al., 1989; Wang y Lehmann, 1991; Kobayashi et al., 1996; citados por De Keuckelaere et al., 2018). Es un marcador de la línea germinal, cuya expresión y función es conservada en varios metazoarios (Extavour, 2007). Además se lo vincula con distintos procesos moleculares esenciales como el ciclo celular, pluripotencia y supervivencia celular.

Posee un dominio de dedos de zinc en su extremo C-terminal, que es la única región

conservada entre los distintos organismos, y es esencial para su función al mediar su unión a aquellos ARNm que contienen el elemento de respuesta a Nanos (NRE) en su región 3' y a varias proteínas formando un complejo que inhibe la traducción. La principal proteína con la que interacciona es Pumilio y al igual que ella, Nanos interactúa con el complejo de desadenilación CCR4-NOT en todos los organismos estudiados. Existen también interacciones con otras proteínas que dependen del organismo, del parálogo de Nanos y del ARNm blanco.

Diversos ARNm presentan la secuencia blanco para la unión del complejo Nanos/Pumilio que son expresados tanto en la línea germinal como en células somáticas para controlar el ciclo celular o la apoptosis.

En *Drosophila* la presencia de la proteína Nanos materna en el polo posterior es esencial para la migración de las células polares hacia las gónadas para formar la línea germinal. Durante esta migración, la mitosis de las células polares es inhibida por represión del ARNm de la ciclina B mediada por el complejo Nanos - Pumilio. Además, en estas células Nanos suprime la apoptosis y expresión de genes somáticos al reprimir la traducción del gen pro-apoptótico *hid* (*head involution definitive*). El complejo Nanos - Pumilio también tiene un rol durante la neurogénesis y en la autorrenovación de las células madre ováricas. En fibroblastos humanos y en *Drosophila*, el complejo reprime la traducción del oncogén E2F3, que es un factor de transcripción con un rol importante en la progresión del ciclo celular y la inducción de la apoptosis. (De Keuckelaere et al., 2018).

El gen *nanos* es expresado en los cuerpos cromatoides de planaria y en las estructuras análogas de otros metazoarios. En la forma sexual de planarias de la especie *Dugesia japonica*, *nanos* se expresa en ovarios y testes, particularmente en las espermatogonias y oogonias premitóticas, mientras que en la forma asexual se expresa en grupos de células en la región dorsal, donde se encuentran los órganos reproductivos en las planarias sexuales. Estas células en la forma asexual son sensibles a la radiación, no contribuyen a la regeneración y constituyen una subpoblación de neoblastos que se cree darán lugar a la línea germinal (Sato et al., 2006). En la planaria *S. mediterránea*, *nanos* presenta un perfil de expresión idéntico en planarias sexuales y asexuales, pero además se observa expresión durante el desarrollo y la regeneración de los fotorreceptores (Handberg-Thorsager y Saló, 2007). La interferencia de *nanos* en *S. mediterranea* resulta en el impedimento del desarrollo, regeneración y mantenimiento de gónadas en la forma sexual (Wang et al., 2007). En planarias sólo se ha identificado un gen *nanos*, mientras que en trematodos y cestodos existen dos copias. En ambos casos estos genes son expresados en una subpoblación de células en proliferación. En *S. mansoni*, *nanos-1* se expresa en una subpoblación que se postula que dará lugar a la línea germinal, mientras que *nanos-2* se expresa en células tipo neoblasto somáticas (Collins et al., 2013; Wang et al., 2013, 2018). En *E. multilocularis* menos del 5% de las células germinativas expresan *nanos* en la capa germinativa del metacestodo. Durante la formación de cada protoescólex se detecta *nanos-2* en una baja cantidad de células en la base del brote y asociadas al desarrollo del sistema nervioso (Koziol et al., 2014).

4.3.2.1.4. pL10 y helicasas DEAD-box

PL10 es una helicasa de ARN dependiente de ATP perteneciente a la familia DEAD-box. Esta es la familia más grande de helicasas de ARN, con 37 miembros en humanos y 26 en *S. cerevisiae*, estando todas caracterizadas por la presencia del motivo DEAD (Asp-Glu-Ala-

Asp). Estas helicasas tienen roles muy importantes en el metabolismo del ARN, formando parte de complejos de múltiples componentes como la maquinaria de empalme del ARN o la de inicio de la traducción eucariota. Todas las helicasas DEAD-box poseen un núcleo altamente conservado que incluye los sitios de unión a ATP y al ARN, que tiene al menos 12 motivos ubicados en posiciones conservadas, siendo algunos de estos conservados en la superfamilia SF2 de helicasas y otros específicos de la familia DEAD-box. Diferentes miembros de la familia están implicados en roles muy diferentes, los que se asocian con las regiones más variables de la proteína (Linder y Jankowsky, 2011). En la

Tabla I 2 se resumen las principales funciones asociadas a estas helicasas en humanos y los miembros de la familia que las cumplen.

Tabla I 2: Función asociada a cada miembro de la familia de helicasas DEAD-box humanas

Función	Helicasa
Metabolismo de ARN mitocondrial	DDX28
Decaimiento de ARN	DDX5 y DDX6
Almacenamiento de ARN	DDX3X, DDX3Y, DDX6
Traducción	DDX2A, DDX2B, DDX3X, DDX3Y, DDX4 , DDX19
Procesamiento de miARNs (RISC)	DDX5
Decaimiento mediado por codones STOP prematuros (NMD)	DDX48
Biogénesis de nucleoproteínas pequeñas nucleares (snRNP)	DDX20
Exportación del ARN	DDX3X, DDX3Y, DDX19, DDX25, DDX39, UAP56
Biogénesis del ribosoma	DDX3X, DDX3Y, DDX5, DDX10, DDX18, DDX21, DDX24 , DDX27, DDX31, DDX47, DDX48, DDX49, DDX50, DDX51, DDX52 , DDX54, DDX55, DDX56
Transcripción	DDX5, DDX17, DDX20, DDX21
Empalme del ARN,	DDX3X, DDX3Y, DDX23, UAP56, DDX42, DDX46

Adaptada de Linder y Jankowsky (2011). En Negrita se indican las helicasas DEAD-box ausentes en genomas de cestodos según Tsai, IJ *et al.* (2013). DDX1, DDX11, DDX12, DDX41, DDX43, DDX53, DDX58, **DDX59** y DDX69 no han sido asignadas a las funciones descritas. DDX8, DDX9, DDX15, DDX16, DDX29, DDX30, DDX31 - DDX38, DDX40 y DDX57 pertenecen a la familia relacionada de DEAH helicasas.

p10 fue originalmente identificado en ratón utilizando una sonda para una secuencia del cromosoma Y. Se determinó que la secuencia estaba relacionada con otras helicasas de ARN como el factor de iniciación de la traducción murino *eIF4A* y el gen *vasa* de *Drosophila*. En ratón, se expresa en la línea germinal masculina de una manera regulada durante el desarrollo, siendo su expresión máxima durante la meiosis de la espermatogénesis (Leroy et al., 1987, 1989). Presenta dos parálogos ligados a cromosomas sexuales llamados DDX3X y DDX3Y que también están presentes en humanos (Rosner y Rinkevich, 2007; Chang y Liu, 2010). En mamíferos euterios hay varias copias de genes de la familia DDX3, algunos de ellos poseen intrones y están ubicados en cromosomas sexuales mientras que otros se encuentran en cromosomas autosómicos y no poseen intrones, indicando que se originaron por retrotransposición. Los genes autosómicos están más relacionados al gen ligado al cromosoma X (Chang y Liu, 2010).

DDX3X se expresa en la mayoría de los tejidos, mientras que DDX3Y se limita a la línea germinal masculina, estando implicado en la fertilidad. Este gen participa en varios aspectos del metabolismo del ARN, en particular regulando la expresión génica. Es capaz de regular la transcripción por la activación de ciertos promotores y la represión de otros. También actúa a nivel del empalme del ARNm, contribuye a la exportación nuclear de los ARNs y cumple un rol en la regulación de la traducción, actuando como activador o como represor dependiendo del gen traducido. Estos procesos tienen impacto en la progresión del ciclo celular, ya que se ha visto que células sin este gen son incapaces de transitar a la fase S del ciclo celular. DDX3 potencia la traducción de la ciclina E1 y por otro lado, inhibe la ciclina D causando la pausa del ciclo celular. Además en el desarrollo embrionario temprano del ratón, DDX3 regula la supervivencia y el ciclo celular (Revisado por Ariumi, 2014).

Otro miembro de la superfamilia de las helicasas DEAD-box, el gen *vasa* incluido en la familia DDX4, es un marcador de la línea germinal en varios metazoarios formando parte del programa de multipotencialidad germinal (GMP: *Germinal Multipotency Program*).

vasa es expresado en neoblastos de planaria y en órganos reproductores de planarias sexuales (Shibata et al., 1999) y del platelminto de vida libre *Macrostomum lignano* (Pfister et al., 2008). Además, este gen es necesario para la regeneración de la planaria *Dugesia japonica*, aunque su interferencia por ARN no elimina los neoblastos ni produce una disminución en el número de mitosis luego de 7 días de tratados (Rouhana et al., 2010). En la planaria *Schmidtea mediterranea*, la interferencia de *Smed-vasa-1* impide la repoblación de neoblastos somáticos luego de la irradiación y su diferenciación posterior (Wagner et al., 2012). Como se mencionó anteriormente, en genomas de trematodos y cestodos no se encuentra este gen pero presentan al menos dos genes tipo *pL10* que se propone que cumplen funciones similares a *vasa* (Skinner et al., 2012; Tsai, IJ et al., 2013).

En *M. corti*, *p10* se expresa en larvas con 4 días de cultivo, sin indicios de segmentación, con un patrón discreto de células distribuidas por todo el parénquima y en células debajo del tegumento. En gusanos segmentados, se observa expresión mayoritaria en los primordios de los testes. En macerados celulares se observa expresión en distintos tipos celulares, en particular en células con morfología similar a las células germinativas, pero no exclusivamente en estas (Domínguez, 2016). En ese trabajo sólo se conocía un fragmento del gen, correspondiente a una región conservada, por lo que no puede descartarse la existencia de hibridación cruzada con otras helicasas de la familia DEAD-box.

HIPÓTESIS, OBJETIVOS Y ESTRATEGIA EXPERIMENTAL

Hipótesis

Las células proliferantes de *M. corti* son sensibles a la radiación y tienen un perfil de expresión característico, diferente al de las células diferenciadas. Muchos de los genes marcadores de este tipo celular están conservados en las células madre de otros platelmintos.

Objetivo general

Realizar una caracterización molecular de las células proliferantes del cestodo modelo *M. corti* para identificar genes marcadores de este tipo celular.

Objetivos específicos

1. Obtener el transcriptoma de organismos irradiados y de células proliferantes purificadas
2. Identificar genes marcadores de células proliferantes de *M. corti* y determinar su conservación como marcadores en Platelminotos
3. Obtener una anotación mejorada del genoma de *M. corti* utilizando los datos de secuenciación de ARN generados.

Como actividades para alcanzar esos objetivos nos planteamos:

- 1) Determinar el efecto de la radiación y sobre las células proliferantes de *M. corti*
- 2) Determinar los perfiles de expresión génica de organismos irradiados y control
- 3) Purificar células en fases G2/S/M y fases G0/G1 del ciclo celular por citometría de flujo
- 4) Determinar los perfiles de expresión génica de organismos con una reducción de células proliferantes y de células proliferantes purificadas por citometría de flujo
- 5) Seleccionar candidatos a genes marcadores de células proliferantes de *M. corti* y genes conservados en Platelminotos.

Una actividad que surgió del análisis de los datos fue la de mejorar la anotación del genoma de *M. corti* utilizando datos de secuenciación de ARN.

Estrategia experimental

En la Figura E 1 se muestra la estrategia experimental seguida para cumplir los objetivos. En primer lugar se evaluó el efecto de distintas dosis de radiación y tiempos de recuperación sobre las células proliferantes de *M. corti*. En base a estos resultados, se seleccionó una dosis en la cual disminuye la cantidad de células proliferantes sin dañar la integridad del organismo para determinar el perfil transcripcional de organismos irradiados comparado con organismos control (Figura E 1, marcado con 1).

Además, se siguió una estrategia alternativa, consistente en la purificación de células proliferantes (en fases S/G2/M del ciclo celular) y células mayoritariamente diferenciadas

(en fases G0/G1 del ciclo celular) mediante citometría de flujo. A partir de estas células purificadas se extrajo ARN para determinar los perfiles de expresión de las células y determinar genes preferencialmente expresados en células proliferantes (Figura E 1, marcado con 2).

En base a los resultados obtenidos del análisis primario de los datos, surgió una nueva actividad, consistente en utilizar estos datos de secuenciación de ARN para mejorar la anotación existente del genoma de *M. corti* (Figura E 1, marcado con 3).

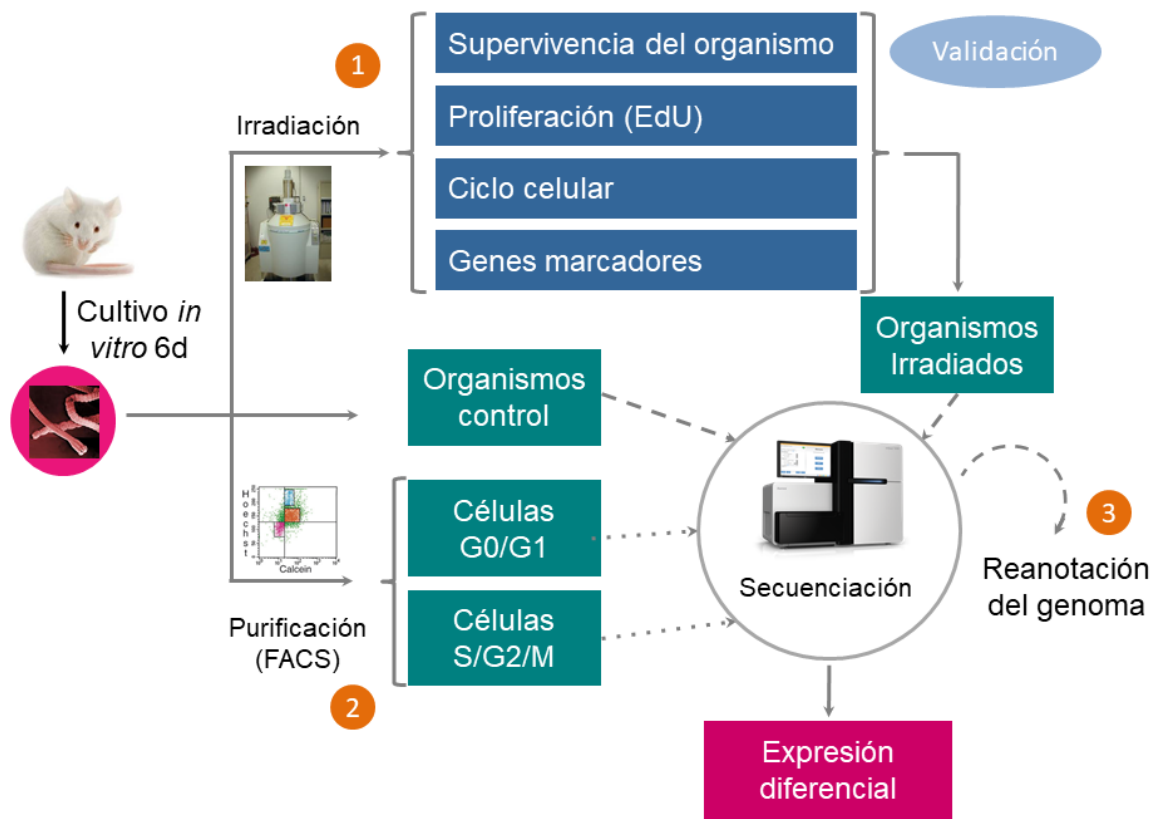


Figura E 1: Estrategia experimental. Se muestran las distintas aproximaciones utilizadas en este trabajo.

Se presentarán los resultados de la mejora en la anotación del genoma en primer lugar, ya que estos datos fueron utilizados para el análisis de la expresión diferencial. A continuación, se presentará el estudio del efecto de la irradiación sobre las células proliferantes de *M. corti* y finalmente los resultados de la caracterización transcriptómica de este tipo celular utilizando datos obtenidos a partir de organismos irradiados y control (Figura E 1, punto 1), y de células proliferantes y células diferenciadas purificadas por citometría de flujo (Figura E 1, punto 2).

MATERIALES Y MÉTODOS

1. Cultivo de *M. corti*

Los tetratiridios de *M. corti* fueron mantenidos mediante pasajes intraperitoneales sucesivos en ratones por Jenny Saldaña y colaboradores (Laboratorio de Experimentación Animal, Facultad de Química, Universidad de la República). El aislamiento de *M. corti* utilizado fue obtenido originalmente por Specht y Voge (1965).

La extracción de parásitos, el cultivo y la inducción de la estrobilización se realizaron como se ha descrito previamente (Britos et al., 2000; Koziol et al., 2010). Los tetratiridios se cultivaron durante 6 días en medio RPMI modificado, suplementado con suero fetal bovino (10%) y taurocolato de sodio (1mg/ml). Se realizó cambio de medio cada 48 o 72 horas.

2. Irradiación

Luego de 6 días en cultivo, los parásitos se lavaron con medio fresco y se introdujeron en tubos de tapa rosca estériles. Los experimentos de irradiación se realizaron en un equipo *GammaCell 220 (MDS Nordion)* disponible en el Instituto Nacional de Donación y Trasplante (Hospital de Clínicas, Universidad de la República). Se utilizaron dosis de 100 Gy, 200 Gy, 300 Gy, 500 Gy y 1000 Gy. Los parásitos control no irradiados fueron tratados en idénticas condiciones. Luego de la irradiación, los parásitos se dejaron recuperar en cultivo durante 1 o 5 días. Para cada condición se realizaron 3 réplicas.

3. Evaluación de los efectos de la radiación sobre las células proliferantes

3.1. Extracción de ARN, retrotranscripción y PCR en tiempo real

Los parásitos se lavaron 3 veces en PBS y se extrajo su ARN utilizando *Trizol (Ambion)*, siguiendo las instrucciones del fabricante. Para facilitar la resuspensión, se realizó una incubación a 65°C durante 5 minutos. El tratamiento con DNasa (*TURBO DNase, Ambion*), cuantificación (*Qubit RNA Broad Range, Invitrogen*) y retrotranscripción a partir de 500ng de ARN total con cebadores al azar (*SuperScript II, Invitrogen*), se realizaron como se describió previamente (Costábile et al., 2017). Para cada ronda de retrotranscripción se eligió una muestra al azar para realizar un control sin transcriptasa reversa.

Para los experimentos de PCR en tiempo real (qPCR) se utilizó GAPDH como control endógeno y se seleccionaron varios genes marcadores de neoblastos de planaria (PCNA, Nanos, pL10, PUM1 and PUM2) para analizar los cambios en sus niveles de expresión frente a distintas dosis de radiación. Para el diseño de los cebadores se utilizó el programa *Primer3plus* (Untergasser et al., 2007). En la Tabla M 1 se muestra la secuencia de los cebadores y el identificador del gen blanco.

Tabla M 1: Cebadores utilizados para cuantificación de expresión por qPCR

Nombre	Secuencia	Aislado por	Código WBPS ¹
qMcPCNA_Fw	AATGCCTGCATCTGAGCTTC	Cauria (2015)	MCOS_0000682501
qMcPCNA_Rv	CCTTGGCAACGGATATCACAAC		
qMcNos_Fw	ATCCTTGAGCGAACACCAAG	Bizzozero (2010)	MCOS_0000573201
qMcNos_Rv	AACAGTGCCCAATTGAGGTG		
qMcPUM1_fw	ACCACGTTGTTTCAGAAAGTGC	Koziol <i>et al.</i> (2008) EU124648*	MCOS_0000894701
qMcPUM1_Rv	TTGAATGTACGTGCCCCTTG		
qMcPUM2_Fw	ATGGCTGTAGAGTGATCCAACG	Koziol <i>et al.</i> (2008) EU124650*	MCOS_0000745201
qMcPUM2_Rv	AGGTTGTCAACGCCTTTGTG		
qMcPL10_Fw	GATGAAGCTCGCAAATTCGC	Domínguez (2016)	MCOS_0000374401
qMcPL10_Rv	ACTTCGAGGAGTTGTTTGCG		

* Aislado por Koziol *et al.* (2008), se indica el código de acceso GenBank

¹: Código del gen en la versión 9 de la anotación de *WormBase ParaSite*

Los experimentos de PCR en tiempo real se realizaron en un equipo *Step One Plus Real Time PCR system (Applied Biosystems)* con *QuantiNova SYBR Green PCR kit (Qiagen)*. Se utilizaron 700nM de cada primer y 0,25µL de reacción de retrotranscripción como molde, en un volumen final de 10µL. Se agregó ROX al *master mix* como referencia pasiva, como indica el fabricante. Se realizó un ciclado rápido (*fast-cycling conditions*) con activación de la polimerasa 2 minutos a 95°C y 40 ciclos de desnaturalización a 95°C por 5 segundos seguido de un paso a 60°C por 10 segundos para la hibridación de cebadores y polimerización. Finalizado el ciclado, se realizó la curva de desnaturalización para verificar la especificidad de la amplificación utilizando el programa por defecto del equipo. Se incluyeron, para todos los cebadores, controles sin molde (agua para detectar contaminación de reactivos y sin transcriptasa reversa para verificar ausencia de ADN genómico).

Para cada par de cebadores se determinó la eficiencia de amplificación utilizando diluciones seriadas (1/5 o 1/10, dependiendo del nivel de expresión de cada gen) de una mezcla conteniendo ADNc de cada condición analizada (Costábile *et al.*, 2017).

Se determinó el nivel de expresión relativa para cada gen utilizando el método $2^{-\Delta\Delta Ct}$ (Livak y Schmittgen, 2001) utilizando organismos no irradiados como condición de referencia y GAPDH como control endógeno. Para validar que estos genes no cambian su nivel de expresión en las condiciones utilizadas, se utilizó el método de $2^{-\Delta Ct'}$ (Livak y Schmittgen, 2001). Los análisis estadísticos se realizaron sobre los valores de $\Delta\Delta Ct$ o $\Delta Ct'$ utilizando el paquete de R *ggpubr* (v0.2, Kassambara, 2018). Se realizó el test no paramétrico de Kruskal-Wallis para la comparación general y el test de Wilcoxon para las comparaciones pareadas. Como método de ajuste del valor p se utilizó la corrección de Holm (por defecto

en el paquete `ggpubr`).

3.2. Análisis de incorporación de análogos de timidina (EdU)

Los parásitos irradiados y control se incubaron *in vitro* con 20µM del análogo de timidina EdU (5-etinil-2'-deoxiuridina, *Invitrogen*) en medio RPMI sin extracto de levadura durante las últimas 4 horas de cultivo. Se lavaron dos veces con PBS y se les removió el tegumento por incubación en agua destilada durante 3 horas. Se fijaron durante 2 horas en paraformaldehído 4% en PBS y se deshidrataron progresivamente por incubación en soluciones con concentración creciente de etanol diluido el PBS. Las muestras se almacenaron en etanol 100% a -20°C.

Para la detección de EdU, los parásitos se rehidrataron progresivamente y se permeabilizaron con Proteinasa K (20µg/mL en PBS con Triton-X100 0.1%) durante 20 minutos a temperatura ambiente. Se incubaron en PBS-BSA 3% durante 30 minutos a temperatura ambiente y se procedió al lavado y detección de acuerdo al protocolo sugerido por el fabricante del *kit Click-iT EdU Imaging (Invitrogen)*. Para la tinción nuclear, luego del revelado de EdU se incubaron los parásitos con DAPI 2µg/mL en PBS durante 15 minutos a temperatura ambiente protegidos de la luz. Finalmente, se montaron en medio de montado *SlowFade (Invitrogen)* y se fotografiaron en un microscopio de epifluorescencia Olympus IX18 (Unidad de Microscopía, Institut Pasteur de Montevideo).

3.3. Análisis del ciclo celular

Se realizaron suspensiones celulares de gusanos irradiados y sin irradiar siguiendo el protocolo descrito por Domínguez *et al.* (2014). Las células se fijaron con etanol 70% frío durante al menos 48 horas y se tiñeron con Ioduro de Propidio para el análisis del ciclo celular como describe Domínguez *et al.* (2014). El análisis estadístico se realizó como se describió en el punto 3.1 de Materiales y Métodos.

4. Genoma y anotación de *M. corti*

Se descargó el borrador del genoma de *M. corti* junto con la anotación disponible al momento del inicio del trabajo (versión WBPS1) de *WormBase ParaSite* (**¡Error! Referencia de hipervínculo no válida.**).

Para cada nueva versión publicada de *WormBase ParaSite* (WBPS2 a WBPS12), se evaluó si hubo cambios con respecto a la versión WBPS1. En la versión WBPS6 se modificaron los modelos génicos, permaneciendo incambiados hasta la versión 12. Para la Reanotación se utilizó la versión WBPS7 (ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS7/species/mesocestoides_corti/PRJEB510/), mientras que en comparaciones de la reanotación con anotaciones previas, se utilizó la anotación en su versión WBPS1 y WBPS9 (ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS9/species/mesocestoides_corti/PRJEB510/).

5. Secuenciación de ARN de organismos irradiados y control

5.1. Extracción de ARN de larvas, construcción de librerías y secuenciación

Se extrajo ARN a partir de organismos sin irradiar y organismos irradiados con una dosis de 100 Gy y 1 día de recuperación. Se utilizó el *kit miRvana* (Ambion) siguiendo las instrucciones del fabricante, utilizando un homogeneizador de tubos *eppendorf* para la disrupción de tejidos en *buffer* de lisis. Se obtuvo por separado la fracción de ARNs pequeños (<200 nucleótidos) y la de ARNs de más de 200 nucleótidos. Los ARNs pequeños no fueron utilizados en este trabajo, por lo que se continuó con la fracción de ARN mayor a 200 nucleótidos. Se trató el ARN con DNasa (*TURBO DNase*, Ambion) siguiendo las recomendaciones del fabricante y se cuantificó mediante fluorescencia utilizando el *kit Qubit RNA Broad Range* (Invitrogen).

Dos muestras por condición fueron enviadas en hielo seco al servicio de secuenciación (*Beijing Genomics Institute*, Shenzhen, China) para el análisis de calidad (*Agilent 2100 Bioanalyzer*), construcción de las librerías y secuenciación en un equipo *Illumina HiSeq 2000* con lecturas pareadas de 100 pb cada una.

Se siguieron dos estrategias para la eliminación del ARN ribosomal: tratamiento con DSN (*Duplex-Specific Nuclease*) siguiendo las recomendaciones de *Illumina* (*DSN Normalization Sample Preparation Guide, Part # 15014673 Rev. A*) o purificación de ARN poliadenilado utilizando el *kit TruSeq™ RNA Sample Preparation kit* (*Illumina*).

5.2. Mapeo de lecturas y comparación con la anotación existente

Las lecturas proporcionadas por el servicio de secuenciación corresponden a lecturas “limpias”, es decir luego de la eliminación de adaptadores, lecturas de baja calidad y lecturas conteniendo bases indeterminadas (N). Las lecturas de cada muestra se mapearon por separado al genoma de *M. corti* utilizando *tophat2* (v2.0.13, Kim et al., 2013) sin anotación de referencia y los demás parámetros por defecto. En base a los datos de mapeo se realizó el ensamblado de los transcritos de cada muestra utilizando *cufflinks* (v2.2.1, Trapnell et al., 2012) y para obtener el conjunto final de transcritos unificando todas las muestras se utilizó *cuffmerge* (v1.1.0, Trapnell et al., 2012). A partir del archivo *gtf* generado se extrajeron las secuencias de los transcritos ensamblados utilizando la herramienta *gffread* (Trapnell et al., 2012).

Para comparar los modelos génicos generados utilizando datos de secuenciación de ARN con la anotación previa del genoma (WBPS1: *WormBase ParaSite*, v1) se utilizó la herramienta *cuffcompare* (v2.2.1), que compara los dos archivos de coordenadas génicas y analiza superposición de modelos.

Además, se utilizó la herramienta *BUSCO* (v 3.0.2, Waterhouse et al., 2018) con el modo *transcripts* y el linaje *metazoa_odb9* para analizar la completitud de los transcritos obtenidos. Se realizó el mismo análisis con los transcritos de la versión WBPS1 de la anotación del genoma de *M. corti* y con los transcritos de *E. multilocularis* como referencia.

6. Secuenciación de ARN de células purificadas

6.1. Purificación de células proliferantes por citometría de flujo y extracción de ARN

Se separaron y purificaron células en fases S/G2/M y G0/G1 del ciclo celular por citometría de flujo de acuerdo a Domínguez *et al.* (2014). La cantidad de células obtenidas fue del orden de 10^5 y 10^6 por condición, respectivamente. Las células se colectaron en PBS - BSA 3% y se diluyeron hasta 10 ml con PBS antes de ser colectadas por centrifugación a 4000 rpm durante 10 minutos. El sobrenadante se centrifugó en iguales condiciones para colectar cualquier célula remanente. Las células se resuspendieron en 100 μ L de Trizol (Ambion) y se almacenaron a -80C.

Se ensayaron distintos *kits* para la extracción de ARN a partir de purificaciones individuales de células. Los *kits* utilizados son especializados para partir de poca cantidad de tejido o células (*PureLink RNA Mini kit*: 5×10^6 - 1×10^8 células (Invitrogen), *Quick RNA Miniprep*: hasta 1×10^7 células (Zymo Research), *ZR RNA Microprep*: hasta 1×10^5 células (Zymo Research) y *Direct-zolTM RNA Miniprep*: hasta 1×10^7 células (Zymo Research)).

Debido a la poca cantidad de células obtenidas por ronda de purificación y a su pequeño tamaño, se juntaron 3 tandas de purificación para cada extracción. El ARN se extrajo utilizando *Direct-zolTM RNA MiniPrep kit* (Zymo Research), siguiendo el protocolo descrito por el fabricante, incluido el tratamiento con DNasa en columna. El ARN se cuantificó utilizando *Qubit RNA Broad Range kit* (Invitrogen).

6.2. Construcción de librerías y secuenciación.

Las librerías de secuenciación se construyeron utilizando el *kit NEBNext Ultra Directional library* (New England Biolabs, Catalog N°: E7420) con el módulo de purificación de ARN poliadenilado. Se utilizaron 60 ng de ARN total para cada librería siguiendo las instrucciones del fabricante y utilizando diferentes índices para cada librería. Se enviaron a Novogene (Beijing, China) para la evaluación de la calidad de las librerías mediante *Bioanalyzer* y secuenciación en un equipo *HiSeq* (Illumina) con lecturas pareadas de 150 pares de bases.

7. Actualización del genoma de *M. corti* utilizando *maker*.

En base a la información obtenida del análisis de los datos de secuenciación de ARN del estadio larvario (ver punto 5 de Materiales y Métodos), se decidió utilizar esta información para realizar una reanotación del genoma. Para ello se utilizó el programa *maker* (v2.31.8, Cantarel *et al.*, 2008), que permite integrar toda la información disponible sobre modelos génicos de una especie o de especies relacionadas.

7.1. Evidencia utilizada

7.1.1. Anotación de genes ribosomales y eliminación de lecturas correspondientes a estos genes

De manera de filtrar lecturas correspondientes a ARN ribosomal, se realizó una anotación de genes ribosomales, ya que éstos no están predichos en la anotación disponible. Se

descargaron de la base de datos de nucleótidos del NCBI secuencias de ARN ribosomal de platelmintos (término de búsqueda: *Platyhelminthes* [Organism] AND rRNA NOT assembly NOT mRNA, acceso 23 de Octubre de 2017). Estas secuencias se filtraron manualmente para eliminar aquellas que eran codificantes para proteínas ribosomales, cromosomas completos o *contigs* muy largos y genomas mitocondriales completos. Las secuencias que pasaron el filtro fueron agrupadas por identidad utilizando el programa *cd-hit-est* (v4.6, Fu et al., 2012) con el parámetro *-c 0.95*. De esta manera se obtuvieron 3959 secuencias representativas que se utilizaron como *query* para una búsqueda por *blastn* local (v2.2.31+, -evaluate 0,0005, Camacho et al., 2009) en el genoma de *M. corti*. Para cada secuencia representativa se descargó la anotación de NCBI utilizando *Entrez Batch* y se diseñó un *script* de *python* para determinar las regiones del genoma que corresponden a cada región de los genes de ARN ribosomal (18S, ITS1, 5.8S, ITS2 o 28S).

Además se utilizaron los programas *infern* (v1.1.2, Nawrocki y Eddy, 2013) para predecir secuencias no codificantes (incluidos los ARNr) y *Rnammer* (v1.2, Lagesen et al., 2007) para predecir ARNr, en ambos casos utilizando parámetros por defecto.

Las lecturas de secuenciación de ARN fueron mapeadas a las secuencias correspondientes a genes ribosomales con *bowtie2* (v 2.2.6, Langmead y Salzberg, 2012) con los parámetros por defecto.

7.1.2. Construcción de librería de repetidos específica de *M. corti*

Para enmascarar las secuencias repetidas en el genoma, se construyó una base de datos de secuencias repetidas específica para *M. corti*. Se utilizó el programa *RepeatModeler* (v1.0.4, Smit y Hubley, 2008) siguiendo las instrucciones sugeridas por los creadores de *maker*¹. Las secuencias repetidas fueron filtradas para excluir segmentos correspondientes a proteínas altamente repetidas en el genoma (por ejemplo, la superfamilia CAP, histonas, entre otras). Para el filtrado, se utilizó la herramienta *blastx* (v2.2.28+, -evaluate 0.0001, Camacho et al., 2009) con las secuencias repetidas como blanco y las proteínas anotadas de *M. corti* disponibles en *WormBase ParaSite* (PRJEB510, v7) como base de datos. Para las proteínas anotadas que presentaron homología con las secuencias repetidas se extrajeron los dominios proteicos asignados por *WormBase ParaSite* (PRJEB510, v7). Sólo aquellas secuencias repetidas con homología a proteínas con dominios asociados a elementos transponibles (ejemplo, transcriptasa reversa, integrasa, etc) fueron mantenidos en la base de datos, mientras los que presentaron dominios asociados a proteínas “clásicas” fueron eliminados (ejemplo: dominio CAP, histonas, integrinas, etc).

7.1.3. Información basada en secuenciación de ARN

7.1.3.1. Estrategia basada en el genoma

Los archivos de mapeo generados por *tophat* para cada muestra (ver punto 5.2 de Materiales y Métodos) se utilizaron como entrada del programa *braker*, que permite el entrenamiento no supervisado de los predictores génicos *GeneMark* y *augustus* (v1, Hoff et al., 2016).

Se utilizó además como evidencia la secuencia de los transcriptos obtenidos en el punto 5.2

¹ http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction--Basic

(Materiales y Métodos).

7.1.3.2. Ensamblado de transcriptos de novo

Las lecturas que no mapearon a las secuencias de ARN ribosomal anotadas manualmente (ver punto 7.1 de Materiales y Métodos) fueron recortadas para eliminar adaptadores y bases de baja calidad utilizando `Trimmomatic` (v0.33, Bolger et al., 2014) con los parámetros `HEADCROP:8 ILLUMINACLIP:TruSeq2-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36`. Las lecturas pareadas y las desapareadas obtenidas luego del recorte se utilizaron para ensamblar transcriptos *de novo* utilizando `trinity` (v2.0.6, Haas et al., 2013) con los parámetros por defecto.

7.1.4. Información adicional utilizada como evidencia

La anotación del genoma de *M. corti* disponible en *WormBase ParaSite* (PRJEB510, v7; 10.614 genes codificantes) se utilizó como evidencia en la corrida de `maker`. Para la transformación a un formato `gff3` adecuado para `maker`, en primer lugar se convirtió a formato `gtf` con el programa `gff3_to_gtf` (GenomeTools, v1.5.9, Gremme et al., 2013) y se reconvirtió a formato `gff3` con la herramienta `genemark_gtf2gff3` incluida en la distribución de `maker`.

Se utilizó la herramienta *web* OrthoVenn (<http://www.bioinfogenome.net/OrthoVenn/>) para obtener grupos de proteínas ortólogas dentro de las predichas en *WormBase ParaSite* (v7) para los platelmintos *Echinococcus multilocularis* (PRJEB122), *Echinococcus granulosus* (PRJEB121 y PRJNA182977), *Schistosoma mansoni* (PRJEA36577), *Hymenolepis microstoma* (PRJEB124) y *Schmidtea mediterranea* (PRJNA12585).

Además, se descendieron del NCBI las 20 secuencias de ARNm de *M. corti* (descargados el 3 de Noviembre de 2016) y ESTs obtenidos por secuenciación del ARN del estadio larvario (Bizarro et al., 2005; NCBI BioSample: SAMN00175995 y SAMN00175994, Acceso: CX863392 a CX865174, descendidos el 3 de Noviembre de 2016).

7.2. Corrida de maker

Para la reanotación del genoma se utilizó la herramienta `maker` (v 2.31.8), con la información descrita anteriormente como evidencia: ESTs y ARNm disponibles en NCBI junto con los transcriptos ensamblados *de novo* y los transcriptos ensamblados utilizando mapeo al genoma (*est evidence*); la anotación disponible en *WormBase ParaSite* (*est_gff evidence*); proteínas ortólogas de platelmintos y predicciones de `augustus` generadas por `braker` (*protein evidence*) y los repetidos filtrados (`rmlib`) junto con el archivo de elementos transponibles distribuido con `maker` “`te_proteins.fa`” (*repeat_protein*) en la sección de enmascarado de repetidos. Se activó la opción de predicción de genes a partir de homología de ESTs y proteínas y el análisis con `tRNAscan` para predicción de ARN de transferencia. Se muestran los parámetros modificados en el archivo de control en la Figura M 1.

Se reportaron todas las predicciones, independientemente del valor de calidad o su largo. Estos modelos fueron posteriormente filtrados, utilizando como criterios que el tamaño mínimo de proteína sea 50 aminoácidos y el valor de confiabilidad máximo sea de 0.75 (*AED score: Annotation Edit Distance*; valor 0: predicción apoyada por la evidencia, valor 1: predicción no apoyada por la evidencia).

```

#-----Genome (these are always required)
genome= Mcorti.genome.fa #genome sequence (fasta file or fasta embedded in GFF3
file)
organism_type=eukaryotic #eukaryotic or prokaryotic. Default is eukaryotic

#-----EST Evidence (for best results provide a file for at least one)
est=
EST_Mcorti_Bizarro.fasta,mRNA_McortiNCBI.fasta,Trinity.fasta,Transcripts_sinRef2
.fa #set of ESTs or assembled mRNA-seq in fasta format
est_gff= mcorti.wbps7.gff3 #aligned ESTs or mRNA-seq from an external GFF3 file

#-----Protein Homology Evidence (for best results provide a file for at least
one)
protein= OrthoVenn_Emul.Sman.Smed.Hmic.EgCh.EgSa.fa, augustus2.aa #protein
sequence file in fasta format (i.e. from mutiple oransisms)

#-----Repeat Masking (leave values blank to skip repeat masking)
model_org=all #select a model organism for RepBase masking in RepeatMasker
rmlib= consensi.final.fa #provide an organism specific repeat library in fasta
format for RepeatMasker
repeat_protein=/etc/opt/maker/data/te_proteins.fasta #provide a fasta file of
transposable element proteins for RepeatRunner
softmask=1 #use soft-masking rather than hard-masking in BLAST (i.e. seg and
dust filtering)

#-----Gene Prediction
est2genome=1 #infer gene predictions directly from ESTs, 1 = yes, 0 = no
protein2genome=1 #infer predictions from protein homology, 1 = yes, 0 = no
trna=1 #find tRNAs with tRNAscan, 1 = yes, 0 = no
unmask=1 #also run ab-initio prediction programs on unmasked sequence, 1 = yes,
0 = no

#-----MAKER Behavior Options
AED_threshold=1 #Maximum Annotation Edit Distance allowed (bound by 0 and 1)
min_protein=0 #require at least this many amino acids in predicted proteins
alt_splice=0 #Take extra steps to try and find alternative splicing, 1 = yes, 0
= no
always_complete=0 #extra steps to force start and stop codons, 1 = yes, 0 = no
keep_preds=0 #Concordance threshold to add unsupported gene prediction (bound by
0 and 1)
correct_est_fusion=0 #limits use of ESTs in annotation to avoid fusion genes

```

Figura M 1: Parámetros relevantes utilizados en el archivo de opciones de maker. Los que no se muestran se dejaron por defecto.

7.3. Análisis de la anotación

7.3.1. Anotación funcional.

Se utilizó la herramienta InterProScan (v 5.20-59.0, Jones et al., 2014) para la asignación de dominios proteicos, con la opción de reportar los términos GO (opción `-goterm`) y los demás parámetros por defecto. Además, se buscó homología de las proteínas reportadas por maker con las proteínas curadas de la base de datos Uniprot (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz, descargada el 6 de mayo de 2017) utilizando blastp (v2.2.28+; -evaluate 1e-5, -max_target_sequences 1, Camacho et al., 2009). Esta información se agregó al gff generado por maker utilizando los scripts `ipr_update_gff` y `maker_functional_gff` disponibles con la distribución de maker. La asignación de términos KEGG se realizó con la herramienta online GhostKOALA (Kanehisa et al., 2016) (parámetros: KEGG GENES databases: `genus_prokaryotes + family_eukaryotes`)

7.3.2. Ortología.

Se identificaron grupos de ortólogos entre las proteínas anotadas de *M. corti* y las de otros platelmintos (ver Tabla M 2) con el programa *OrthoFinder* (v2.2.3, Emms y Kelly, 2015) utilizando los parámetros por defecto. Para los transcritos de *S. mediterranea* se realizó la traducción con *TransDecoder* (v4.1.0, Haas et al., 2013) con los parámetros por defecto.

Tabla M 2 Organismos y origen de las proteínas utilizadas para la búsqueda de ortólogos

Organismo (abreviación)	Versión	Fuente
<i>M. corti</i> (MM)		Este trabajo
<i>M. corti</i> (MC)	PRJEB510	<i>WormBase ParaSite</i> (v7)
<i>E. granulosus</i> (EG)	PRJEB121	<i>WormBase ParaSite</i> (v7)
<i>E. multilocularis</i> (EM)	PRJEB122	<i>WormBase ParaSite</i> (v7)
<i>H. microstoma</i> (HM)	PRJEB124	<i>WormBase ParaSite</i> (v7)
<i>H. diminuta</i> (HD)	PRJEB507	<i>WormBase ParaSite</i> (v7)
<i>T. solium</i> (TS)	PRJNA170813	<i>WormBase ParaSite</i> (v7)
<i>T. asiatica</i> (TS)	PRJNA299871	<i>WormBase ParaSite</i> (v7)
<i>C. sinensis</i> (CS)	PRJDA72781	<i>WormBase ParaSite</i> (v7)
<i>O. viverrini</i> (OV)	PRJNA222628	<i>WormBase ParaSite</i> (v7)
<i>F. hepatica</i> (FH)	v2	M. Mitreva (comunicación personal)
<i>F. gigantica</i> (FG)	v1	M. Mitreva (comunicación personal)
<i>F. buski</i> (FB)	v1	M. Mitreva (comunicación personal)
<i>S. mansoni</i> (SM)	PRJEA36577	<i>WormBase ParaSite</i> (v7)
<i>S. japonicum</i> (SJ)	PRJEA34885	<i>WormBase ParaSite</i> (v7)
<i>M. lignano</i> (ML)	PRJNA284736	<i>WormBase ParaSite</i> (v7)
<i>S. mediterranea</i> (SM)	Smed_20140614.nt dd_Smed_v6.pcf.contigs.fasta dd_Smed_v4.nuc.fasta	smedgd.stowers.org planmine.mpi-cbg.de idisk.mpi-cbg.de
<i>G. salaris</i> (GD)	Gsa-gene-models.fasta	Hahn et al (2014) http://invitro.titan.uio.no/gyrodactylus/downloads.html

7.3.3. Clasificación de la anotación.

Los modelos génicos obtenidos se clasificaron de acuerdo a la presencia de genes ortólogos en otros platelmintos y evidencia de anotación funcional (homología con proteínas curadas de UniProt, presencia de dominios proteicos, presencia de términos de ontología génica y ortología con proteínas implicadas en vías metabólicas KEGG). Se realizaron diagramas de Venn con el paquete `VennDiagrams` de R (v1.6.20, Chen, 2018)

7.3.4. Comparación de los modelos génicos obtenidos con la anotación disponible (WBPS9)

Se utilizó la herramienta `cuffcompare` (v2.2.1, Trapnell et al., 2012) para analizar la superposición de los modelos génicos generados por `maker` y la anotación WBPS9.

7.3.5. Evaluación de la anotación.

Para evaluar la completitud de los modelos génicos obtenidos, se compararon las diferentes versiones de la anotación con la herramienta `BUSCO` (v 3.0.2, Waterhouse et al., 2018) con el modo *protein* y el linaje *metazoa_odb9*.

Las estadísticas generales de la anotación de `maker` y de *WormBase ParaSite* (v1 y v9) fueron generadas con el programa `eval` (v2.2.8, Keibler y Brent, 2003).

7.4. Búsqueda de genes de interés

7.4.1. Helicasas DEAD-box

Se buscaron en la base de datos *Planmine* (Rozanski et al., 2019) aquellos transcritos de *S. mediterranea* que tuvieran un dominio *InterPro* asociado a helicasas DEAD-box. Para ello se utilizó el *template* “*Interpro protein domain name to transcript*”. Los transcritos obtenidos se tradujeron con `TransDecoder` con los parámetros `--retain_blastp_hits` y `--retain_long_orfs_length 1000`, manteniendo sólo aquellos marcos abiertos de lectura (ORFs) que presentaron homología con proteínas en la base de datos *Swiss_Prot* (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz, descargada el 6 de mayo de 2017). Para *S. mansoni* y *E. granulosus*, se buscó en la versión WBPS9 de la anotación de *WormBase ParaSite* los genes que tuvieran un dominio *InterPro* asociado a helicasas DEAD-box. Para *M. corti* se buscó en la anotación generada en este trabajo y en la versión WBPS9 de la anotación de *WormBase ParaSite* con la misma estrategia. Se incluyeron además las proteínas de ratón y humano utilizadas por Domínguez (2016) para la clasificación de las proteínas de cestodos.

Las secuencias proteicas se alinearon con `muscle` (Edgar, 2004), con los parámetros por defecto y se realizó un árbol filogenético con `fasttree` (v 2.1.7, Price et al., 2010). En base a esta filogenia se seleccionaron los genes que corresponden a helicasas DEAD-box y no a las helicasas relacionadas de la familia DEAH. Las secuencias seleccionadas se alinearon nuevamente con `muscle` y se realizó en análisis filogenético con el programa *PhyML* en su modalidad *online* (<http://www.atgc-montpellier.fr/phyml-sms/>), incluyendo la predicción del modelo evolutivo con la herramienta *SMS* (Guindon et al., 2010; Lefort et al., 2017).

7.4.2. Nanos

Se descargaron las secuencias proteicas de genes *nanos* de humano, ratón y *C. elegans* de UniProt y las de *E. multilocularis* y *S. mansoni* de WormBase ParaSite. Para planaria, se buscaron en Planmine (Rozanski et al., 2019) genes *nanos* codificadas en el transcriptoma denominado “dd_smed_v6”, encontrándose un único gen. En el caso de *M. corti* se seleccionaron los genes codificantes para proteínas Nanos en la anotación generada en este trabajo y en la anotación de WormBase ParaSite. En la Tabla M 3 se muestran los nombres de los genes y los números de acceso de cada uno.

Tabla M 3: Genes *nanos* utilizados en el análisis filogenético

Organismo (abreviación)	Gen	Código	Fuente
Humano	HsNos1	Q8WY41	Uniprot
	HsNos2	P60321	
	HsNos3	P60323	
Ratón	MmNos1	Q80WY3	
	MmNos2	P60322	
	MmNos3	P60324	
<i>C. elegans</i>	CeNosA	Q23406	WormBase ParaSite (v7)
	CeNosB	Q09597	
<i>E. multilocularis</i>	EmNos1	EmuJ_000861500	
	EmNos2	EmuJ_000606200	
<i>S. mansoni</i>	SmNos1	Smp_055740	
	SmNos2	Smp_051920	
<i>S. mediterránea</i>	SmedNos	dd_Smed_v6_10484_0_1	PlanMine (planmine.mpi-cbg.de), dd_smed_v6
<i>M. corti</i>	--	MCU_009857-RA	Este trabajo
	--	MCU_000830-RA	
	--	MCOS_293601-mRNA-1	WormBase ParaSite (v7)
	--	MCOS_573201-mRNA-1	

Las secuencias se alinearon con el programa Clustal Omega (Sievers et al., 2011) utilizando el modelo de Markov de proteínas Nanos descargado de Panther (<http://www.pantherdb.org/panther/exportHmM.jsp?acc=PTHR12887>, Familia NANOS

PROTEIN: PTHR12887). Se seleccionó la región correspondiente al dominio de dedos de zinc (el único dominio conservado entre los diferentes genes) y se determinó el modelo de sustitución de aminoácidos con la herramienta SMS (Lefort et al., 2017). El análisis filogenético se realizó con el programa PhyML (Guindon et al., 2010) dentro de la herramienta UGENE (Okonechnikov et al., 2012). Se utilizó el modelo VT y se seleccionaron las opciones de “Búsqueda de árbol” = NNI +SPR, “Optimizar topología del árbol” y como soporte de ramas el método SH-like. Las demás opciones se dejaron por defecto.

8. Análisis de expresión diferencial en organismos irradiados y células proliferantes purificadas

8.1. Calidad de las lecturas, mapeo al genoma y expresión diferencial

Se utilizaron en este análisis las lecturas obtenidas de organismos control y organismos irradiados con una dosis de 100 Gy (ver punto 5 de Materiales y Métodos), así como también las lecturas obtenidas a partir de células purificadas por citometría de flujo (ver punto 6 de Materiales y Métodos).

En primer lugar, las lecturas fueron recortadas para eliminar adaptadores y bases de baja calidad utilizando Trimmomatic (v 0.33, Bolger et al., 2014) con los parámetros HEADCROP:8 ILLUMINACLIP:Trimmomatic-0.33/adapters/TruSeq2-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36.

El mapeo al genoma y el análisis de expresión diferencial se realizaron en CLC Genomic Workbench 11.0 (<https://www.qiagenbioinformatics.com/>). Las lecturas pareadas resultantes del recorte se importaron en el programa y se mapearon al genoma de *M. corti* (WBPS1) utilizando los parámetros por defecto. Se seleccionó para el caso de las librerías de células purificadas la opción de mapear las lecturas a la hebra reversa ya que son librerías direccionales que preservan la información de la molécula de ARN. En el caso de las librerías de organismos control e irradiados se mapearon a ambas hebras. Se utilizaron los modelos génicos generados en este trabajo (punto 7 de Materiales y Métodos) como archivo de anotación del genoma, calculando los niveles de expresión (en *Transcripts Per Million*: TPM) para los genes (sin discriminar entre isoformas). Se consideraron diferencialmente expresados aquellos genes con un *False Discovery Rate* (FDR) <0.05 y un *Fold Change* mayor a 2 (expresión aumentada en muestras conteniendo células proliferativas: G2/S/M y gusanos control) o menor a -2 (expresión aumentada en células diferenciadas: G0/G1 y gusanos irradiados).

8.2. Análisis de componentes principales (PCA) y análisis de distancias

Para el análisis de la reproducibilidad de las réplicas se tomaron los valores de conteo para cada gen en cada condición y se convirtieron a valores de pseudoconteo ($\log_2 [\text{count}+1]$).

El análisis de componentes principales, se realizó con la función `prcomp` de R (v3.5.2, R Core Team, 2018) y se graficó utilizando el paquete `ggplot2` (v3.1.0, Wickham, 2016).

Para la generación del *heatmap* de las distancias entre muestras se utilizó la función `cim` del paquete de R `mixOmics` (v 6.3.2, Rohart et al., 2017) con los parámetros por defecto (`dist.method: euclidean` y `clust.method: complete linkage`).

8.3. Enriquecimiento de vías metabólicas y términos de ontología génica (GO)

Se seleccionaron las secuencias proteicas correspondientes a genes diferencialmente expresados en células proliferantes o en células diferenciadas y se analizaron en busca de vías metabólicas enriquecidas utilizando KOBAS3.0 (Xie et al., 2011) con los parámetros por defecto (versión online: http://kobas.cbi.pku.edu.cn/anno_iden.php). Se eligió como organismo de referencia al platelminto *S. mansoni* que sólo permite el análisis de vías KEGG, así como también humano, que permite el análisis de vías KEGG, PANTHER, Reactome y BioCyc. Las nubes de palabras se realizaron con el script disponible en http://www.sthda.com/upload/rquery_wordcloud.r con los parámetros por defecto.

El mismo conjunto de genes se utilizó para el análisis de enriquecimiento de términos GO utilizando el paquete de R topGO (v2.34.0, Alexa y Rahnenfuhrer, 2016). Se utilizaron los términos GO definidos por el programa InterProScan (punto 7.3.1 de Materiales y Métodos) y se siguió el procedimiento descrito en <http://avrilomics.blogspot.com/2015/07/using-topgo-to-test-for-go-term.html>.

8.4. Evaluación de genes conservados en células madre de platelmintos

8.4.1. Genes característicos de células tipo neoblasto de *S. mansoni*

Se seleccionaron los identificadores de los genes que se expresan preferencialmente en células tipo neoblasto de los trabajos de:

- Wang et al., (2018): Genes que caracterizan las distintas clases de células tipo neoblasto identificadas. Se seleccionaron todos los genes reportados por los autores (Tabla Suplementaria 1: 154 genes con *Fold Change* entre clases mayor a 4 y valor $p < 0.01$)
- Collins et al., (2016): Genes que disminuyen sus niveles 2 días y 14 días luego la irradiación (que también disminuyen ante la interferencia de los genes *h2b* y *fgfr*). Se seleccionaron todos los genes reportados por los autores (Tabla Suplementaria 1 (hoja 1): 135 genes con un *Fold Change* < -1.34 (correspondiente al *Fold Change* del gen *pcna*), no reportan valor p por lo que se asume que son significativos)
- Wang et al., (2013): Genes conservados con marcadores de neoblastos de planaria que presentan mayor nivel de expresión en el esporocisto respecto a miracidios (Tabla Suplementaria 1C: 163 genes con un *Fold Change* > 2 (correspondiente al *Fold Change* del gen *pcna*), no reportan valor p por lo que se asume que son significativos)
- Collins et al., (2013): Genes que disminuyen su expresión 2 días luego de la irradiación (Tabla Suplementaria 2 (hoja 1): 268 genes con un *Fold Change* < -1.34 (correspondiente al *Fold Change* del gen *pcna*) y valor $p < 0.05$)

8.4.2. Genes característicos de neoblastos y progenitores de *S. mediterranea*

Se seleccionaron los identificadores de genes de *S. mediterranea* que caracterizan los distintos grupos de células identificados por *single-cell sequencing* de los siguientes trabajos:

- Fincher et al., (2018): Transcritos asociados al grupo denominado *Neoblast* (Tabla Suplementaria 2 (hoja All *Neoblast*): 2413 transcritos correspondientes a 1127 genes en el transcriptoma dd_Smed_v4)

- Plass *et al.*, (2018): Transcritos asociados a los grupos que incluyen la palabra *Neoblasts* o *Progenitors* (Tabla Suplementaria 3: 1397 transcritos correspondientes a 599 genes en el transcriptoma dd_Smed_v6)
- Zheng *et al.*, (2018): Genes asociados a los grupos de neoblastos Nb1 a Nb12 (Tabla Suplementaria 2: 118 genes en el transcriptoma smed2014)

8.4.3. Ortología

Como en los trabajos en *S. mediterranea* se utilizaron distintos transcriptomas, en primer lugar se buscaron mediante `blastn` (v2.2.31+, -num_alignments 1, -evalue 1e-5, Camacho et al., 2009) genes homólogos entre los tres transcriptomas (ver el origen de los transcriptomas en la Tabla M 2 de Materiales y Métodos). Se creó un *script* de `python` para evaluar la conservación de los genes como marcadores de neoblastos de planaria en los 3 trabajos. Un *script* similar se utilizó para evaluar la conservación de genes entre los distintos trabajos de *S. mansoni*.

Además, se buscaron genes homólogos entre *S. mediterranea* y *M. corti* y entre *S. mediterranea* y *S. mansoni* por `tblastn` (v2.2.31+, -num_alignments 1, -evalue 1e-5, Camacho et al., 2009) utilizando como *query* las secuencias proteicas de *M. corti* y *S. mansoni* (ver Tabla M 2 de Materiales y Métodos) y como base de datos cada uno de los transcriptomas de *S. mediterranea*. Se identificaron genes homólogos entre *M. corti* y *S. mansoni* por `blastp` (v2.2.28+, -evalue 1e-5 -max_target_seqs 1, Camacho et al., 2009).

Se creó un *script* de `python` para generar una tabla conteniendo el homólogo de cada gen caracterizado en células proliferantes de alguna especie y su conservación en las demás especies. Los diagramas de Venn se realizaron con el paquete `VennDiagrams` de R (v1.6.20, Chen, 2018).

8.5. Expresión de repetidos

A partir de los repetidos identificados por `maker`, se generó un archivo de anotación en formato `gtf` utilizando el *script* `makeTEgtf.pl.gz` disponible online (http://labshare.cshl.edu/shares/mhammelllab/www-data/TEToolkit/TE_GTF/). Se mapearon las lecturas correspondientes a las muestras de células purificadas utilizando el programa `hisat2` (v2.1.0, Kim et al., 2015) con parámetros por defecto y los archivos de mapeo se utilizaron para analizar los niveles de expresión de repetidos con la herramienta `TEtranscript` del paquete `TEtoolkit` (Jin et al., 2015).

CAPÍTULO 1: Reanotación del genoma de *M. corti* utilizando datos transcriptómicos

1. Introducción

1.1. Anotación de genomas

La anotación de genomas consiste en determinar las regiones codificantes para proteínas, así como también ARNs no codificantes. La anotación de las regiones codificantes suele basarse en predictores génicos, para los cuales el conocimiento previo sobre los genes de la especie de interés o especies relacionadas es de mucha utilidad para generar buenas predicciones. En este sentido el uso de datos de transcriptómica y/o proteómica mejora las predicciones (Prasad et al., 2017)

EL proceso de anotación consta de dos grandes etapas, la primera es la alineación de la evidencia (proteínas o transcritos conocidos) al genoma y predicción de nuevos genes con predictores *ab initio* en base a las características de esta evidencia. La segunda consiste en sintetizar esta información en la anotación de los genes. La etapa de predicción encuentra la región codificante más probable para un gen, mientras que en la anotación se incluye la información de la evidencia y permite determinar las regiones no traducidas y, en caso de existir, formas de *splicing* alternativo (Yandell y Ence, 2012).

El primer paso en un proyecto de anotación es la identificación y enmascaramiento de secuencias repetidas, que incluyen regiones de baja complejidad y elementos transponibles. Las regiones enmascaradas no son utilizadas en la etapa de predicción de genes, por lo que se evitan falsos positivos. La identificación de elementos transponibles es desafiante ya que presentan una conservación relativamente baja, por lo que es necesario crear una base de datos de secuencias repetidas específica para la especie. Estas bases de datos son creadas utilizando dos estrategias: por homología y *de novo*. Las búsquedas por homología buscan elementos transponibles conocidos de otros organismos, por lo que existe la posibilidad de no encontrar aquellos que son más divergentes. Las búsquedas *de novo* identifican regiones repetidas en el genoma por comparación del genoma consigo mismo, por lo tanto no distinguen si las secuencias corresponden a elementos transponibles o a familias multigénicas de proteínas, como histonas o tubulina entre otras. En este caso es necesario realizar un filtrado de estas secuencias repetidas para no enmascarar las regiones codificantes de proteínas de interés. Esta estrategia tiene la ventaja de identificar elementos transponibles más divergentes.

Para obtener una anotación de un genoma de buena calidad es necesario contar con algún tipo de evidencia para la predicción de genes. Los predictores génicos *ab initio* usan modelos matemáticos para la determinación de la estructura exón-intrón. Estos predictores tienen que ser entrenados para generar el modelo matemático adecuado al organismo, determinando ciertas características de los genes, como largo de intrones y exones y frecuencia de codones para distinguir genes de regiones intergénicas. Para el entrenamiento es necesario contar con un conjunto de genes confiables, cuya estructura génica sea correcta. En el caso de organismos no modelo, no siempre se cuenta con esta información (Yandell y Ence, 2012). Actualmente existen herramientas que utilizan evidencia de datos de secuenciación de ARN y alineamiento de proteínas o ESTs al

genoma para entrenar los predictores génicos. La herramienta *braker1* se basa en los predictores *GeneMark-ET* y *Augustus* para obtener los modelos génicos, siendo muy útil para genomas de organismos de los que existen pocos genes confiables con evidencia experimental. *GeneMark-ET* incorpora la información de mapeo al genoma de lecturas de secuenciación de ARN para su entrenamiento de manera no supervisada y hace una predicción inicial. Al usar datos de mapeo y no transcritos ensamblados se evita la introducción de errores asociados al ensamblado de los mismos. Esta predicción es utilizada para entrenar a *Augustus*, el cual genera los modelos génicos finales (Hoff et al., 2016).

Maker2 es una herramienta de anotación de genes codificantes para proteínas que permite identificar y enmascarar repetidos, alinear *ESTs* y proteínas al genoma y utilizar esta evidencia para el entrenamiento de los predictores *ab initio* *Augustus*, *SNAP* y *GeneMark*. Finalmente integra toda esta información produciendo la anotación de genes codificantes en el genoma. Este programa está diseñado para la anotación de genomas nuevos, que no poseen un conjunto de genes de referencia para la evaluación de la precisión de los modelos (mediante la sensibilidad (SN) y especificidad (SP)). En este caso, los modelos se evalúan de acuerdo al apoyo que presentan en la evidencia ingresada al programa. Se define un criterio de congruencia (C) que es conceptualmente similar a la precisión. En este caso, $C = (|in_j|/|j| + |in_j|/|i|)/2$, donde *in_j* es la cantidad de bases del modelo que son apoyados por la evidencia, mientras que *i* y *j* son el número total de bases en los modelos y en la evidencia. *Maker* reporta para cada modelo el valor de AED (*Annotation Edit Distance*), que corresponde a la incongruencia (1 - C) o la distancia entre la evidencia y el modelo. Aquellos modelos con un valor de AED igual a 0 coinciden completamente, mientras que aquellos con valor de 1 no tienen ningún tipo de apoyo en la evidencia. Además, reporta para cada transcripto un resumen de la intersección con la evidencia junto a otros parámetros de interés: largo de la región no traducida en 5' (*5'UTR*), fracción de sitios de *splicing* confirmados por alineamiento de *ESTs*, fracción de exones que se superponen con alineamiento de *ESTs*, fracción de exones que se superponen con alineamientos de *ESTs* o proteínas, fracción de sitios de *splicing* confirmados por predicciones *ab initio*, fracción de exones que se superponen con una predicción *ab initio*, número de exones en el ARNm, largo de la región no traducida en 3' (*3'UTR*) y largo de la proteína producida. Todos estos valores son útiles a la hora de evaluar los modelos generados (Cantarel et al., 2008; Holt y Yandell, 2011).

La anotación de un genoma es un proceso iterativo, en el que la aparición de nueva evidencia, como la secuenciación de transcritos de nuevos estadíos, permite verificar la calidad de la anotación existente y mejorarla agregando nuevos genes o variantes.

Con el objetivo de caracterizar las células proliferantes de *M. corti* se obtuvo el transcriptoma de organismos irradiados (con menor cantidad de células proliferantes, ver Capítulo 2) y de organismos sin irradiar. En este capítulo se presentan los resultados del análisis primario de los datos de transcriptómica y su uso para mejorar la anotación del genoma. Los datos de secuenciación de ARN mostraron una riqueza mayor de transcritos comparada con la versión de la anotación génica disponible a la fecha (versión WBPS1, con 9.056 modelos génicos). A partir de estos resultados, los datos fueron utilizados para mejorar la anotación del genoma. Las versiones de la base de datos *WormBase ParaSite* son actualizadas periódicamente cada 4 meses en promedio, pero la anotación de *M. corti*

se mantuvo incambiada hasta la versión WBPS6, donde la cantidad de modelos génicos aumentó a 10.614. En las versiones posteriores (actualmente está disponible la versión WBPS12, de Diciembre de 2018) no hubieron cambios en la cantidad de modelos.

2. Resultados

2.1. Extracción de ARN y preparación de librerías

Para la secuenciación de organismos irradiados y control realizamos una extracción con el *kit* MiRvana, que permite separar los ARNs largos de los ARNs pequeños (menores a 200 bases). En la Figura 1. 1 se muestra la calidad y cantidad de los ARNs obtenidos.

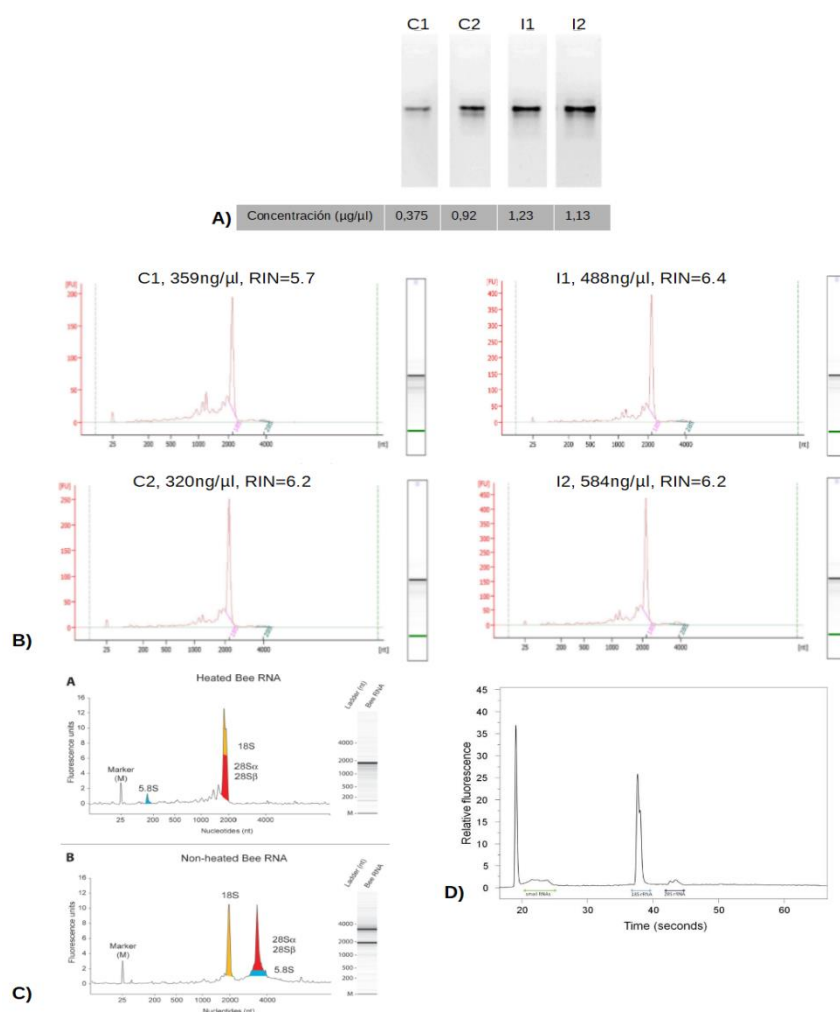


Figura 1. 1: Calidad del ARN obtenido a partir de organismos irradiados (I) y control (C). A) electroforesis en gel de agarosa al 2% de ARNs largos luego del tratamiento con DNasa. Se muestran por separado los duplicados de cada condición experimental. En el cuadro inferior se muestra la concentración en µg/µl de cada muestra. B) Estudio por *Bioanalyzer* de la calidad de las muestras de ARN. Junto al nombre de la muestra se muestra la concentración y el RIN determinados por el servicio de secuenciación. C) Efecto del tratamiento con calor (necesario previo al estudio con *Bioanalyzer*) sobre el ARN total de abeja (tomado de Winnebeck et al., 2010). D) Electroferograma de ARN total del platelminto monogéneo *Gyrodactylus salaris* (tomado de Fromm et al., 2011).

En el panel A se observa la calidad de los ARNs largos luego del tratamiento con DNasa y en el B el estudio por *Bioanalyzer* realizado por el servicio de secuenciación. Para evaluar el ARN utilizado para secuenciar se utiliza el *RNA Integrity Number* (RIN), un parámetro de

calidad que depende de la relación entre la cantidad de ARNr 18S y 28S. En el caso de mamíferos, un RIN mayor o igual a 7 indica una calidad adecuada para la secuenciación. En el caso de Platelminfos e insectos (Figura 1. 1, paneles C y D), se ha reportado que los perfiles observados en *Bioanalyzer* difieren de los esperados para células de mamíferos (Winnebeck et al., 2010; Fromm et al., 2011). Al realizar la incubación a alta temperatura previo a la corrida de *Bioanalyzer*, el ARN 28S se escinde en dos fragmentos de tamaño similar al del ARN 18S y por lo tanto se observa un único pico correspondiente al tamaño del ARNr 18S, en lugar de los dos picos esperados (ARNr 18S y ARNr 28S). Este mismo efecto puede observarse en la corrida electroforética (Figura 1. 1, Panel A) ya que, al incubar el ARN a 65° para facilitar su resuspensión, se produce la rotura del ARN 28S y se visualiza una única banda en el gel. Dados estos antecedentes, se procedió al secuenciado de las muestras a pesar de que el RIN reportado por el estudio por *Bioanalyzer* sea menor a 7.

2.2. Calidad de las lecturas

En una primera instancia, se solicitó al servicio de secuenciación la realización de un tratamiento con DSN (*Duplex-Specific Nuclease*: nucleasa específica de doble cadena) para la eliminación de ARN ribosomal evitando utilizar un método de purificación por polyA y de esta manera conservar todos los ARNs largos, incluidos aquellos ARNm y ARNlnc no poliadenilados. El análisis de estos datos reveló que el tratamiento con DSN no fue efectivo, ya que más del 86% de las lecturas mapearon a ARNr (datos no mostrados). Por este motivo se solicitó la purificación de ARN-polyA y construcción de librerías a partir de esta fracción de ARN purificada.

La secuenciación resultó en aproximadamente 100 millones de lecturas para cada muestra. El filtrado primario por calidad realizado por el servicio de secuenciación mantuvo entre 93 y 107 millones de lecturas limpias para cada muestra, representando en promedio el 94.5% de las lecturas obtenidas originalmente (Figura 1. 2).

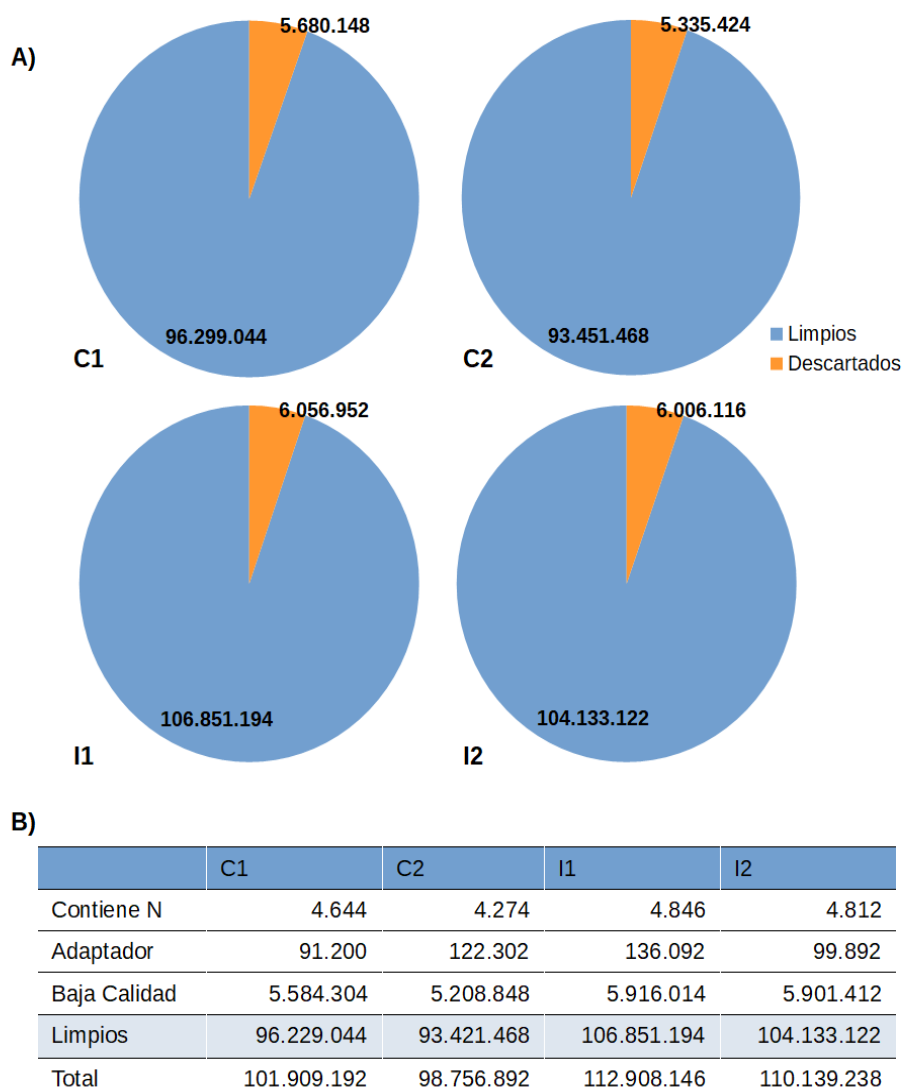


Figura 1. 2: Cantidad de lecturas obtenidas por muestra. A) Cantidad de lecturas limpias (celeste) o descartadas (anaranjado) para cada muestra secuenciada. En la categoría Descartados se muestran combinadas las lecturas que contienen “N” o adaptadores y lecturas de baja calidad. B) Tabla indicando la cantidad de lecturas obtenidas y la clasificación de las mismas. Filtrado realizado por el servicio de secuenciación. Se nos entregaron las lecturas “limpias”. Datos proporcionados por el servicio de secuenciación (BGI). C: Muestras control, I: Muestras irradiadas

2.3. Estadísticas de mapeo al genoma

Luego de comprobar la calidad de las lecturas se realizó el mapeo al genoma de *M. corti* con *tophat*, obteniendo un alto porcentaje de lecturas mapeadas al genoma de manera concordante (ambas lecturas del par mapean adecuadamente). En promedio, el 85.5% de las lecturas obtenidas para cada muestra mapean al genoma y de ellas, el 78.2% mapean de manera concordante (ver Tabla 1. 1). Este porcentaje de mapeo de datos de transcriptómica es similar al observado en estudios en diversos cestodos, que van entre un 71 y un 88% para genomas ensamblados principalmente con tecnología *Illumina* (Liu et al., 2017) y llega a un 98% en el caso de *T. multiceps*, cuyo genoma se ensambló utilizando tecnología *PacBio* y *Hi-C* (Li et al., 2018).

Tabla 1. 1: Datos de mapeo al genoma

		C1		C2		I1		I2	
		Conteo	%	Conteo	%	Conteo	%	Conteo	%
Izquierda	Total	48.149.522	100	46.735.734	100	53.425.597	100	52.066.561	100
	Mapeados	41.032.037	85,2	40.346.386	86,3	46.252.828	86,6	44.989.364	86,4
	Múltiples alineamientos	764.317	1,9	862.109	2,1	1.108.421	2,4	969.327	2,2
Derecha	Total	48.149.522	100	46.725.734	100	53.425.597	100	52.066.561	100
	Mapeados	40.465.624	84,0	39.825.067	85,2	45.626.487	85,4	44.364.498	85,2
	Múltiples alineamientos	759.272	1,9	853.409	2,1	1.097.281	2,4	954.831	2,2
Todos	Tasa de mapeo general	84,6		85,8		86,0		85,8	
	Pares alineados	37.334.807	100	36.650.825	100	42.171.565	100	40.764.054	100
	Múltiples alineamientos	689.550	1,8	778351	2,1	1.005.013	2,4	867.385	2,1
	Alineamientos discordantes	60.789	0,2	54.397	0,1	66.662	0,2	62.915	0,2
	Tasa de alineamiento concordante	77,4		78,3		78,8		78,2	

C: Muestras control, I; muestras irradiadas

2.4. Número de transcritos y comparación con la anotación existente (WBPS1)

A partir del mapeo al genoma, se ensamblaron con *cufflinks* 22.391 transcritos, correspondientes a 14.353 genes. Al comparar las coordenadas génicas de estos 22.391 transcritos con las de los 9.056 genes predichos en la anotación del genoma disponible en ese momento (versión WBPS1), se encontraron 8.837 transcritos expresados en el estadio larvario que no están predichos en la anotación WBPS1. Además, 7.680 de los transcritos expresados corresponden a nuevas isoformas de los genes previamente anotados (Figura 1. 3A). En este sentido, el uso de información transcriptómica permite mejorar la definición de modelos génicos existentes.

De los 9.056 genes predichos en la anotación WBPS1, 8.122 se expresan en el estadio larvario analizado en este trabajo. Los 934 genes restantes pueden ser genes expresados en otros estadios del ciclo de vida o en diferentes condiciones.

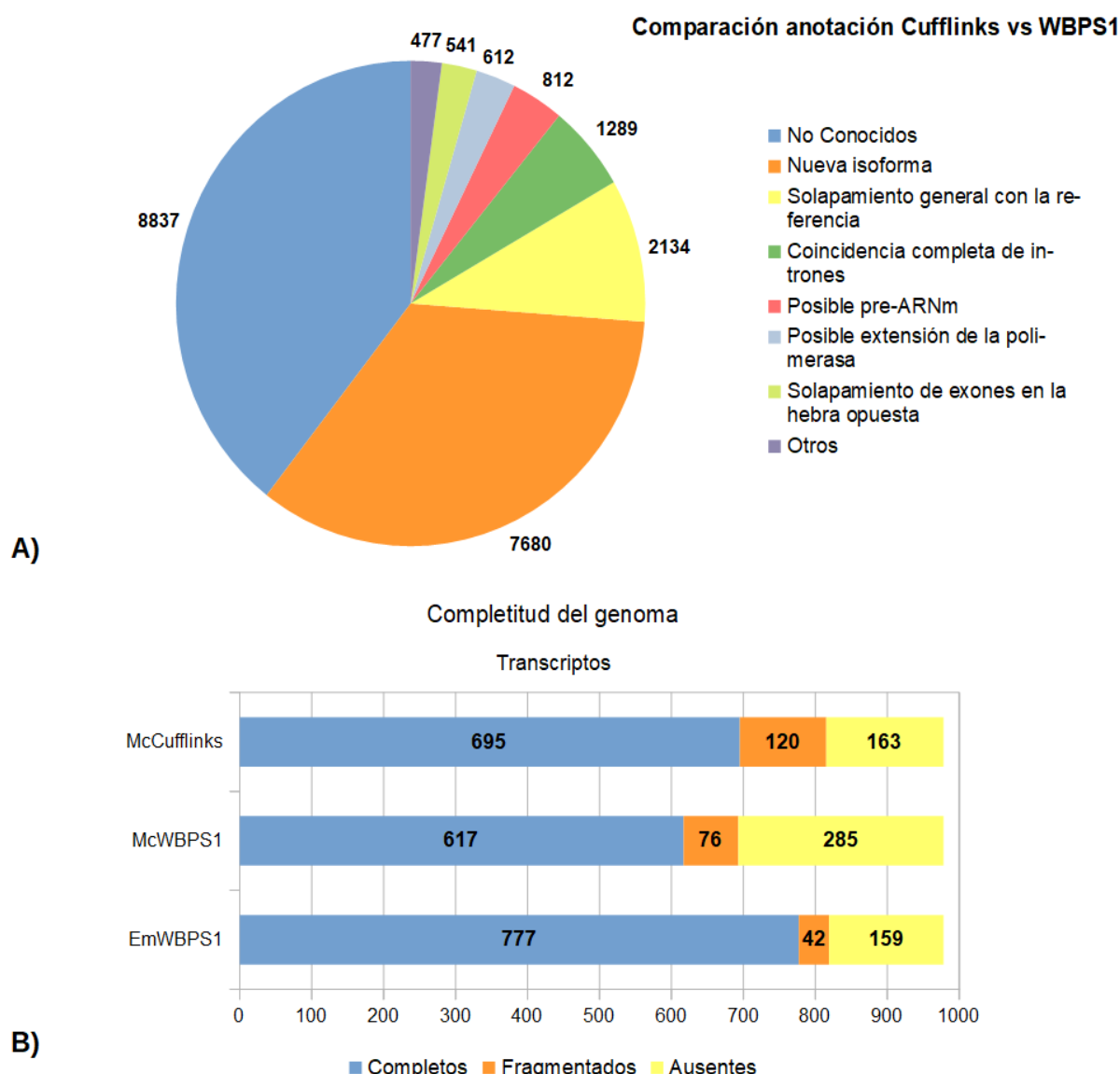


Figura 1. 3: Evaluación de los transcritos obtenidos a partir de datos de secuenciación de ARN. A) Comparación de los transcritos obtenidos a partir de datos de transcriptómica con la anotación WBPS1 mediante la herramienta *cuffcompare*. Se seleccionaron las categorías (*class code*) con mayor representación y las restantes se muestran bajo la categoría “Otros”. B) Comparación de la completitud de los modelos génicos con la herramienta *BUSCO* a nivel de transcritos. McCufflinks: genes obtenidos usando datos de secuenciación de ARN, McWBPS1: genes existentes en la anotación WBPS1. Se muestra la completitud de *E. multilocularis* (EmWBPS1) como referencia (ver el texto para más detalles).

La herramienta *BUSCO* (*Benchmarking Universal Single-Copy Ortholog*) permite determinar la completitud de una anotación o genoma, es decir, si se encuentran todos los genes que se espera para ese organismo. Toma como referencia un grupo de genes ortólogos que se espera estén presentes en un grupo taxonómico dado y reporta la cantidad de esos genes que se encuentran completos de copia única, completos duplicados, fragmentados o ausentes (Waterhouse et al., 2018). Se realizó este análisis utilizando un grupo de 978 genes ortólogos que se espera se encuentren en todos los animales multicelulares (metazoarios) (Zdobnov et al., 2017). Al analizar los transcritos generados por *cufflinks* a partir de datos de secuenciación de ARN, se observa que aumenta la cantidad de genes

completos y fragmentados en comparación a la anotación WBPS1 (Figura 1. 3B, 815 en *McCufflinks* contra 693 en *McWBPS1*), indicando que el uso de datos de secuenciación de ARN permite mejorar la anotación existente. La cantidad de genes que no están presentes en los transcritos ensamblados (*McCufflinks*, categoría “ausentes”, 163) es similar a la observada para *E. multilocularis* (159), cuyo genoma es tomado como referencia de cestodos dada su buena calidad (Tsai, IJ et al., 2013). Como se ha descrito en la introducción, estos organismos carecen de ciertas vías metabólicas, lo que se refleja en la ausencia de estos genes en el análisis de completitud.

2.5. Estrategia de reanotación del genoma

Como se mencionó en el punto anterior, la información transcriptómica permite descubrir nuevos genes no anotados en la anotación WBPS1 del genoma, así como descubrir nuevas isoformas. Por este motivo, se decidió realizar la reanotación del genoma incorporando esta evidencia. En la Figura 1. 4 se muestra un esquema de la estrategia utilizada, indicando toda la información incorporada en el análisis.

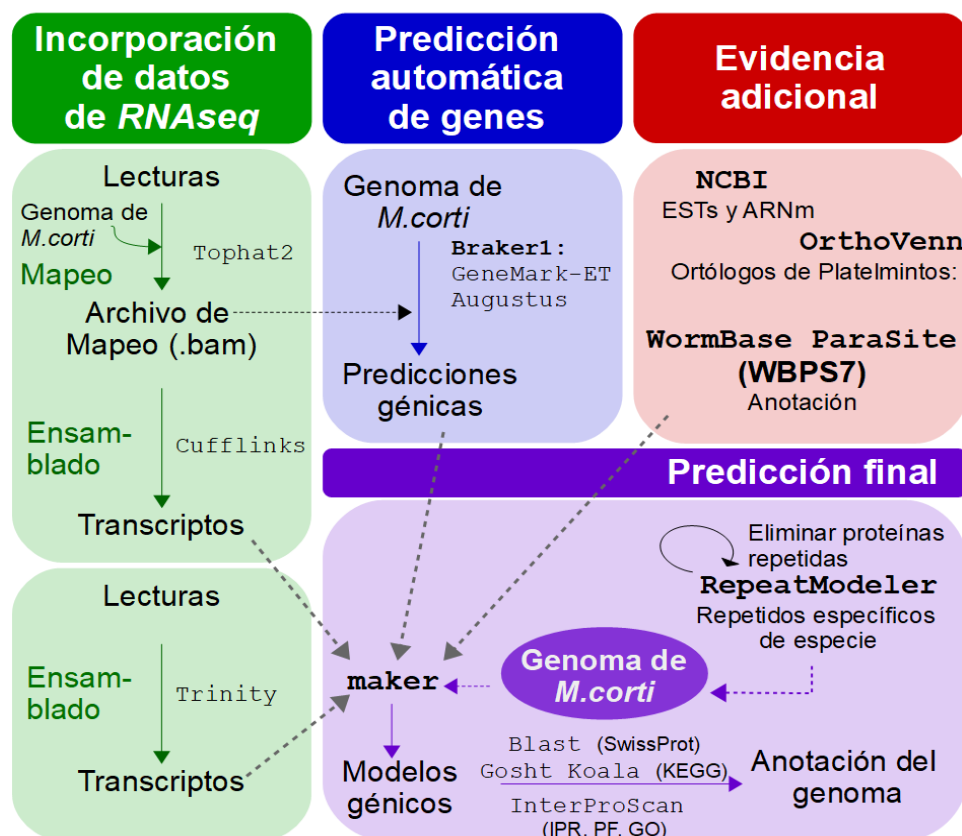


Figura 1. 4: Estrategia y datos utilizados para la mejora de la anotación del genoma de *M. corti*. Ver el texto por detalles de la información.

Para enmascarar en el genoma las secuencias repetidas, se construyó una base de datos de específica de *M. corti* en base al análisis del genoma (utilizando el programa RepeatModeler). Esta base de datos fue filtrada para eliminar regiones correspondientes a proteínas conocidas no relacionadas a elementos transponibles clásicos. De las 176 secuencias identificadas como repetidas, se eliminaron 124 ya que presentaban dominios asociados a proteínas que suelen estar repetidas en los genomas (Histonas, Tubulinas, Quinasas y GST entre otras). Las restantes secuencias fueron utilizadas junto a la base de

datos de elementos transponibles distribuida con `maker` para enmascarar el genoma como primer paso de la anotación (Figura 1. 4, recuadro violeta).

Los datos de *RNAseq* fueron incorporados como evidencia utilizando dos estrategias. Por un lado, se utilizaron los datos de mapeo obtenidos con `tophat` (ver Capítulo 1: punto 2.3) como entrada del programa `braker`. Como se mencionó en la introducción del capítulo, este programa permite predecir modelos génicos con `Augustus` y `GeneMark` sin necesidad de contar con un conjunto de genes de entrenamiento confiable. Las 13.679 proteínas predichas por `braker` se incorporaron como evidencia proteica en `maker` (Figura 1. 4, recuadro azul). Como una estrategia paralela, debido a la fragmentación del borrador del genoma de *M. corti*, se ensamblaron transcritos *de novo* con la herramienta Trinity. Esta estrategia permitió ensamblar 76.681 transcritos correspondientes a 48.118 genes. Los transcritos ensamblados *de novo* también fueron utilizados como evidencia en `maker`, así como también los 22.391 transcritos ensamblados por `cufflinks` (Figura 1. 4, recuadro verde).

De modo de integrar la mayor cantidad de evidencia (Figura 1. 4, recuadro rojo), también se incorporaron los 10.614 genes predichos en la versión WBPS7 de la anotación de *WormBase ParaSite* (la última actualización disponible en el momento del inicio de la re-anotación) y los datos de ARNm y ESTs de *M. corti* disponibles en NCBI (20 ARNm y 1783 ESTs). Finalmente, se incluyó como evidencia proteínas que presentan ortólogos en los Plelmintos *E. multilocularis* (3.522 proteínas de 10.669 totales), *E. granulosus* (3.478 de 11.319 y 3485 de 10.273; dos anotaciones diferentes en WBPS, ver Materiales y Métodos), *H. microstoma* (3.654 de 12.373), *S. mansoni* (4.107 de 11.774) y *S. mediterránea* (5.075 de 29.850).

2.6. Comparación de datos de reanotación con genomas disponibles

Durante el transcurso del trabajo, la anotación del genoma de *M. corti* disponible en *WormBase ParaSite* fue mejorando, por lo que los resultados de la reanotación fueron comparados con la última anotación disponible (WBPS9: 10.614 genes).

La reanotación produjo un total de 22.424 transcritos correspondientes a 14.913 genes (1,5 isoformas para cada gen en promedio; Tabla 1. 2). El largo de los transcritos es en promedio similar a los predichos para *E. multilocularis*, aunque se obtiene menor cantidad de exones por transcripto (Tabla 1. 2). El número de exones predicho por `maker` es similar al número predicho en la anotación de *WormBase ParaSite* (versión WBPS9).

McMaker: Anotación generada en este trabajo, McWBPS9: Versión 9 de *WormBase ParaSite* para el genoma de *M. corti*, EmWBPS9: Versión 9 de *WormBase ParaSite* para el genoma de *E. multilocularis*.

Se predijeron 2.590 genes confiables que, de acuerdo al análisis realizado, únicamente se encuentran en *M. corti* ya que no presentan ninguna evidencia de anotación funcional ni ortólogos en otros platelmintos. De ellos, 1.388 no están anotados en la anotación WBPS9. No se descarta que algunos de estos transcritos posean homología con alguna proteína que no se encuentre en la base de datos de proteínas curadas (SwissProt) o que no esté anotada en los platelmintos analizados. En promedio, un 33% de estos genes son expresados en el estadio larvario (TPM>1). En la Figura 1. 6 se muestra un histograma con los valores de expresión.

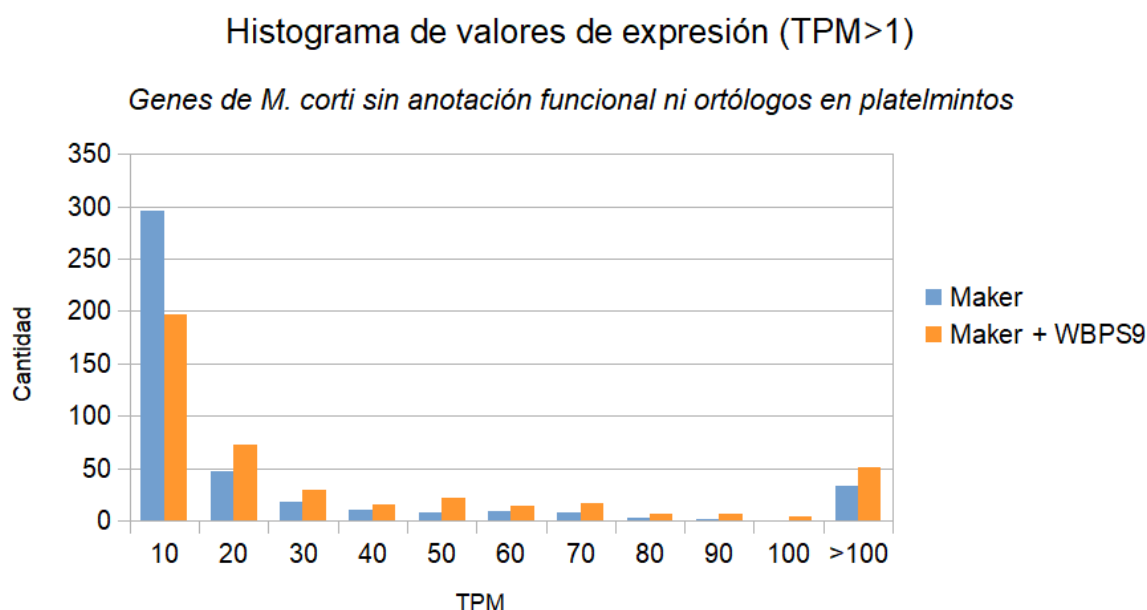


Figura 1. 6: Histograma de los valores de expresión para genes de *M. corti* sin evidencia de anotación funcional ni ortólogos en otros platelmintos. En celeste se muestran genes predichos que no se superponen con modelos de la anotación de *WormBase ParaSite* y en anaranjado los genes que se encuentran en ambas anotaciones.

Entre los 865 genes anotados únicamente en *M. corti* y expresados en el estadio larvario, un 57% tienen nivel de expresión bajo (entre 1 y 10 TPM) pero el resto tienen niveles de expresión relativamente altos. En particular uno de ellos es el tercer transcripto con mayor nivel de expresión en la larva. El hecho de que tengan evidencia de expresión, indica que estos genes no serían un artefacto de la anotación y es altamente probable que cumplan alguna función, aún por determinar.

Se analizó además la completitud de la anotación comparando los nuevos modelos génicos con el conjunto de ortólogos presentes en todos los metazoarios con la herramienta BUSCO (Figura 1. 7). Al comparar con la primera versión de la anotación de *WormBase ParaSite* (WBPS1), la reanotación mejora el contenido génico, ya que se recupera una mayor cantidad de genes completos (732 en la reanotación contra 625 en la versión WBPS1) y disminuye la cantidad de genes fragmentados. Al comparar con la última versión de la anotación de *WormBase ParaSite* (WBPS9) se obtiene una mayor cantidad de genes no anotados en el genoma (185 genes “ausentes” en la reanotación contra 136 en la versión WBPS9, habiendo 142 genes que no están presentes en el genoma de *E. multilocularis*).

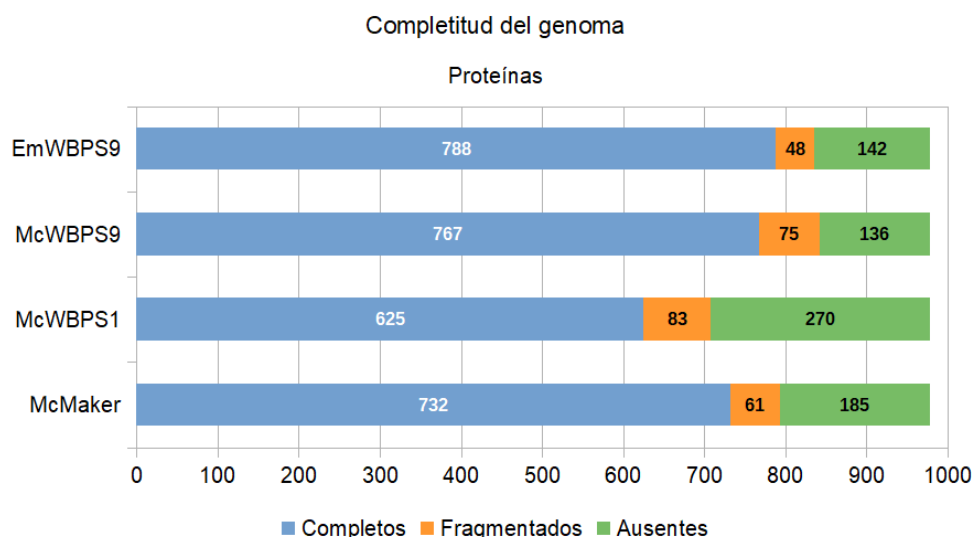


Figura 1. 7: Análisis de completitud de la anotación utilizando BUSCO. Se muestra en la categoría “Completos” los genes completos de copia única y duplicados. EmWBPS9 corresponde a la completitud del genoma de *E. multilocularis* que se muestra como referencia. McMaker corresponde a la anotación generada en este trabajo y McWBPS1/McWBPS9 corresponden a las versiones 1 y 9 de la anotación de *WormBase ParaSite* disponible para *M. corti*

Los resultados de este análisis difieren de lo esperado, ya que la anotación de *WormBase ParaSite* fue incluida como evidencia en `maker`, por lo que se espera que estos genes estén incluidos en los modelos generados. Para evaluar la causa de estas diferencias, se analizó cada uno de los 978 grupos de ortólogos usados como referencia en el análisis y se evaluó como fueron catalogados para la reanotación (McMaker) y para la versión WBPS9. Estos resultados se muestran resumidos en la Tabla 1. 3. La gran mayoría de los grupos de ortólogos (820 de 978) fueron asignados a la misma categoría en ambas anotaciones. Sin embargo, 100 grupos de ortólogos tienen menor calidad en la reanotación, con 72 genes catalogados como ausentes (que en la versión WBPS9 aparecen fragmentados o completos) y 28 genes fragmentados (que aparecen completos en la versión WBPS9). Por otro lado, se mejora la anotación para 58 genes, de los cuales 55 fueron catalogados como completos en la reanotación y fragmentados o ausentes en la versión WBPS9).

Tabla 1. 3: Cantidad de grupos de ortólogos que se mantienen igual, mejoran o empeoran en la anotación de *maker* en comparación con la anotación WBPS9

N° de grupos de ortólogos	McMaker	McWBPS9	Categoría	Total
677	C	C	Igual	820
113	M	M		
30	F	F		
28	F	C	Empeoran ¹	100
59	M	C		
13	M	F		
35	C	F	Mejoran ²	58
20	C	M		
3	F	M		

C: catalogado como completo por BUSCO, F: catalogado como fragmentado por BUSCO, M: catalogado como ausente por BUSCO

1: Grupos de ortólogos que empeoran en McMaker con respecto a McWBPS9.

2: Grupos de ortólogos que mejoran en McMaker con respecto a WBPS9.

De los 100 genes que fueron catalogados como ausentes o fragmentados en la re-anotación y como completos en la anotación WBPS9, 68 se intersectan a nivel de coordenadas génicas (determinado con la herramienta *cuffcompare*), es decir que los genes fueron anotados por *maker*, pero con estructura génica diferente a la anotación WBPS9. En la Figura 1. 8 se muestra un caso representativo.

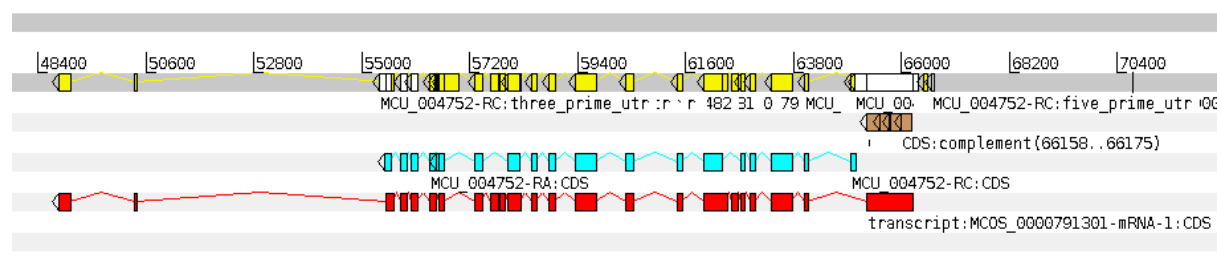


Figura 1. 8: Comparación de un modelo génico catalogado como ausente por BUSCO en la reanotación y como completo en la anotación WBPS9. En amarillo se muestran los exones codificantes, en blanco los exones correspondientes a regiones no traducidas (UTR), en celeste las isoformas anotadas por *maker* y en rojo el transcripto anotado en la anotación WBPS9. La región reconocida por el modelo oculto de Markov se muestra en marrón claro.

Como puede apreciarse, los modelos difieren únicamente en los extremos, siendo la mayor parte del gen coincidente. La principal diferencia reside en el primer exón, al cual *maker* consideró como región no traducida pero aparece como codificante en la anotación WBPS9.

Al tener una estructura génica tan parecida no se esperaría que se clasifique como completo en una anotación y ausente en otra, sino que se esperaría que aparezca como fragmentado. Analizando el modelo oculto de Markov (*HMM: Hidden Markov Model*) generado para este grupo de ortólogos, se observa que la región que reconoce coincide con el primer exón codificante anotado en la versión WBPS9 y con la región que en la reanotación está anotada como región no traducida 5'. Por este motivo es que BUSCO (al analizar proteínas) lo clasifica como ausente en la reanotación, lo que indica que si bien ésta es una herramienta útil para evaluar la generalidad de la completitud de los genomas, en algunos casos da resultados incorrectos. Hay que considerar también que no existe un grupo de ortólogos específico para platelmintos, sino que se utilizó el grupo de ortólogos de todos los metazoarios. Como se mencionó anteriormente, los platelmintos parásitos carecen de ciertos genes conservados y además de que existen genes que se encuentran únicamente en estos organismos, por lo que es necesario generar grupos de ortólogos más específicos. En la actualidad esto es posible debido a la existencia de un gran número de genomas de varias especies de platelmintos. El análisis de grupos de ortólogos realizado para la caracterización de los modelos génicos obtenidos en este trabajo, que incluye especies de Trematodos, Cestodos y Monogeneos, revela la existencia de 3143 grupos de ortólogos que presentan al menos un gen para cada organismo. Este es un insumo interesante para realizar el análisis de completitud de los genomas de estos organismos. Recientemente se ha realizado un análisis comparativo de los genomas de nematodos y platelmintos disponibles en *WormBase ParaSite*, y se observan muchas familias génicas que están presentes en parásitos pero no en *C. elegans*, y aproximadamente la mitad no tienen anotación funcional (Coghlan et al., 2019).

Una comparación a nivel de intersección de coordenadas génicas entre la reanotación y la anotación de *WormBase ParaSite* en sus versiones WBPS1 o WBPS9 se muestra en la Figura 1. 9. Puede observarse que a pesar de que la cantidad de nuevos genes (categoría “No conocidos”) disminuye al comparar con WBPS9, respecto a la comparación con WBPS1, esta sigue siendo alta (4.878 genes nuevos). Aumenta además la cantidad de isoformas, existiendo diferencias en la anotación de 12.785 isoformas.

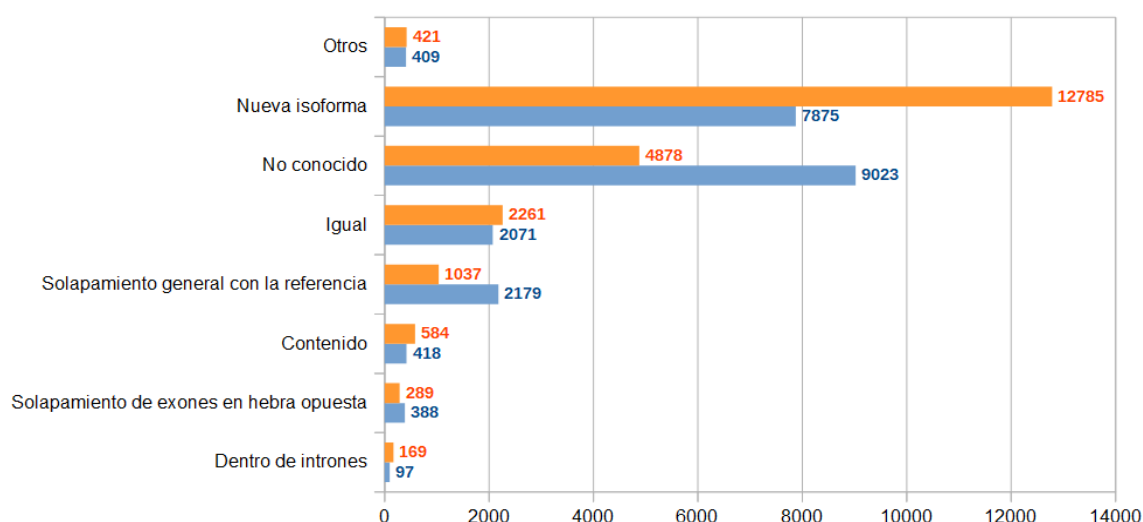


Figura 1. 9: Comparación a nivel de coordenadas génicas (*cuffcompare*) de la reanotación y dos versiones de la anotación de *WormBase ParaSite* (WBPS1 y WBPS9).

2.7. Análisis de nuevos genes descubiertos por *maker*

Al analizar las 4.878 isoformas nuevas (Figura 1. 10), un poco más de la cuarta parte tienen confiabilidad baja (ya sea, AED ≥ 0.75 o largo de proteínas < 50), aproximadamente una cuarta parte presentan evidencia de anotación (homología con proteínas de curadas (SwissProt), términos KEGG, términos GO o dominios conservados), otra cuarta parte presenta homología con genes de otros platelmintos y el resto no presentan anotación ni ortología con platelmintos.

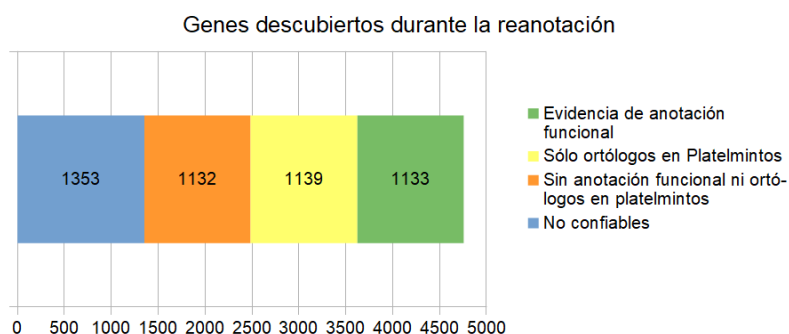


Figura 1. 10: Análisis de los 4.878 genes nuevos anotados por *maker* que no se encuentran en la versión WBPS9 de la anotación de *WormBase ParaSite* para el genoma de *M. corti*

Dentro de los 1.113 genes con evidencia de anotación funcional, se evaluó cuáles eran los dominios conservados que aparecen más representados en los genes nuevos (Figura 1. 11)

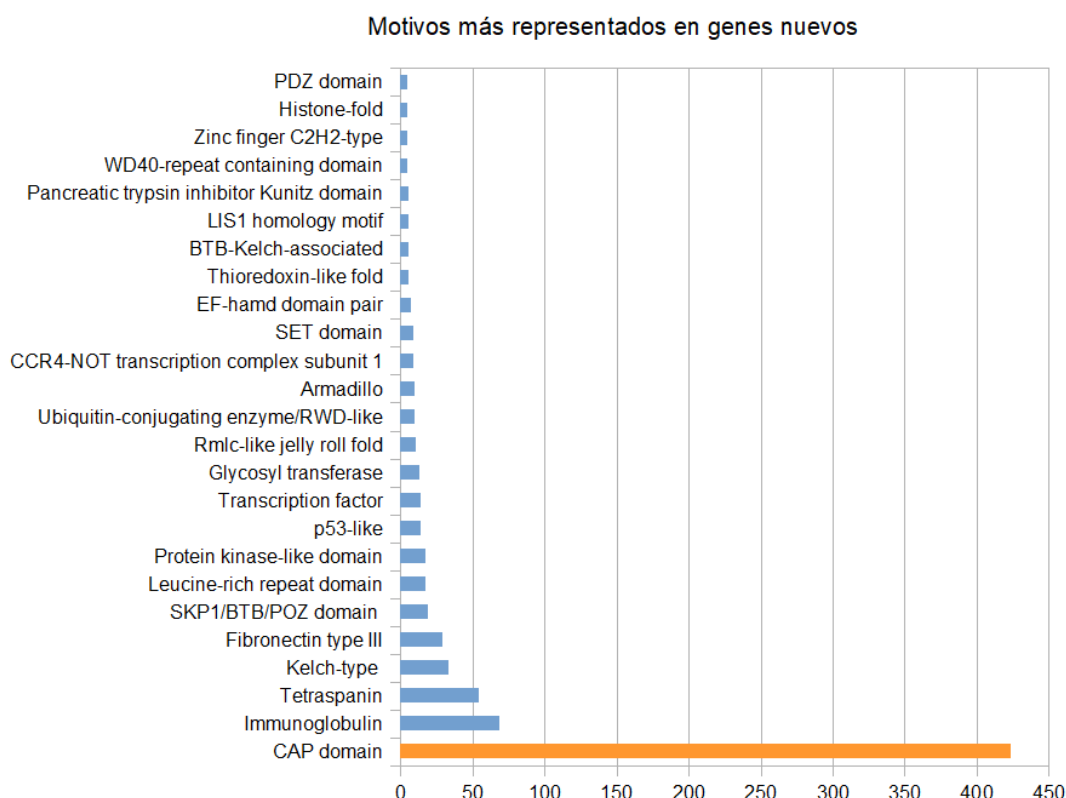


Figura 1. 11: Conteo de dominios InterPro representados más de 5 veces en los nuevos genes. En anaranjado se muestran los dominios correspondientes a proteínas de la superfamilia SCP/TAPs.

Se destaca la cantidad de dominios pertenecientes a la superfamilia SCP/TAPS (424) obtenidos en la anotación de `maker` que no estaban anotados en la versión WBPS9 de *WormBase ParaSite*. Estas proteínas se encuentran en todos los organismos pero no se conoce su función específica. Están implicadas en diversos procesos, como la reproducción en el caso de mamíferos, respuesta a patógenos en plantas y se las ha implicado en la modulación del sistema inmune del hospedero en el caso de nematodos (Gibbs et al., 2008; Cantacessi y Gasser, 2012). En *S. mansoni*, distintos miembros de la superfamilia se expresan preferencialmente en diferentes estadios del ciclo de vida (Chalmers et al., 2008) y se ha reportado que son capaces de unir esteroides (Kelleher et al., 2014). Recientemente reportamos la expansión de este grupo de proteínas en el genoma de *M. corti* (Costábile et al., 2018; el artículo completo se encuentra en el anexo final). Mediante búsquedas con la herramienta `blast` y curado manual se encontraron 444 secuencias que contienen este dominio conservado, de las cuales 271 son completas y 173 son posibles pseudogenes o proteínas parciales debidas a la fragmentación del genoma. En la anotación WBPS9 existen únicamente 54 genes con este dominio, mientras que la anotación generada en este trabajo recupera 398 de las 444 proteínas reportadas, y 38 proteínas adicionales. En otros cestodos se encuentran 48 proteínas pertenecientes a la superfamilia en promedio, de las cuales 26 están completas. El análisis filogenético de esas proteínas reveló una expansión masiva en *M. corti*, en donde un clado monofilético contiene el 71 % de las proteínas completas (Figura 1. 12, clado marcado en rojo)

Varios clados dentro de esta gran expansión están sujetos a presión selectiva y una de las proteínas que se encuentra dentro de estos clados es secretada al medio y está implicada en la evasión del sistema inmune. Esta proteína inhibe la activación de células dendríticas, suprimiendo la respuesta Th1 observada en la etapa inicial de la invasión de cestodos (Vendelova et al., 2016, 2015).

Por otra parte, otro miembro de la superfamilia se expresa diferencialmente durante el desarrollo estrobilar, disminuyendo a medida que avanza la segmentación. En el estadio larvario presenta un patrón de expresión en el sistema neuroendócrino, mientras que en el gusano segmentado se expresa en los primordios genitales (Costábile et al., 2017).

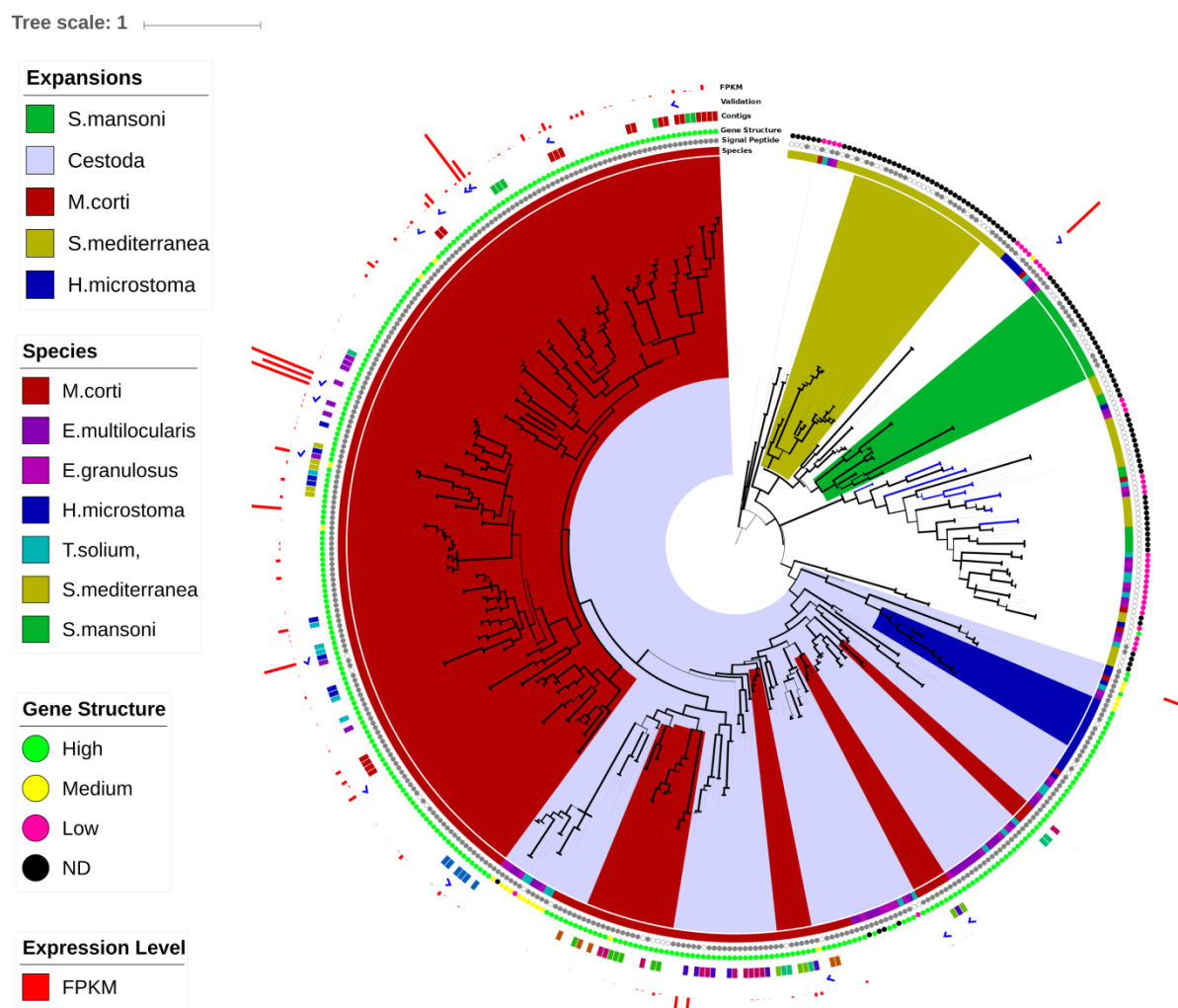


Figura 1. 12: Análisis filogenético por máxima similitud de los dominios CAP completos de platelmintos. La primera fila de cuadrados por fuera del árbol indica la especie descrita en la leyenda (*Species*). La segunda fila indica presencia o ausencia de péptido señal de secreción. La tercera fila indica la conservación de la estructura génica (*Gene Structure*). La cuarta fila indica agrupamiento de genes en el mismo *contig*. Las V azules indican genes validados experimentalmente por RT-PCR. Las barras rojas de la fila exterior indican los niveles de expresión (en FPKM). Los clados monofiléticos en donde se observan expansiones especie-específicas se marcan con los colores descriptos en la leyenda (*Expansions*). Tomada de Costábile et al. (2018).

La diversidad de genes de la superfamilia CAP y sus perfiles de expresión dinámicos indica que estos genes estarían implicados en diversas funciones en la biología de *M. corti*, por lo que profundizar en su estudio es de gran importancia.

2.8. Análisis de nuevas isoformas

La reanotación incluye varias isoformas nuevas respecto a la anotación WBPS9, entre las que se puede observar una mayor cantidad de regiones no traducidas y de mayor largo (20.004 regiones no traducidas 5' y 14.344 regiones no traducidas 3', ver Tabla 1. 4), con un aumento en el número de exones. La anotación de estas regiones tiene importancia ya que permite predecir regiones regulatorias, por ejemplo sitios blanco de unión a miARNs o a proteínas de regulación post-transcripcional.

Tabla 1. 4: Métricas referentes a la cantidad de exones y regiones no traducidas (UTRs) entre las distintas anotaciones de *M. corti*

		McMaker	McWBPS9
Exones	Conteo	77.264	62.450
	Largo promedio	215,68	236,61
3' UTR	Conteo	14.344	143
	Largo promedio	316,8	165,76
5'UTR	Conteo	20.004	1.631
	Largo promedio	194,89	40,64

McMaker: Anotación generada en este trabajo, McWBPS9: Versión 9 de la anotación de *WormBase ParaSite* para el genoma de *M. corti*.

La anotación de nuevas isoformas además puede evidenciar la existencia de *splicing* alternativo. Los genes de tropomiosina (TPM) representan un claro ejemplo de *splicing* alternativo de exones mutuamente excluyentes. En platelmintos existen dos genes TPM, cada uno de los cuales presenta dos isoformas, una de alto peso molecular (TPM-HMW que incluye los exones 1a, 2, 3, 4b, 5a, 6b, 7a, 8a y 9b) y otra de bajo peso molecular (TPM-LMW, que incluye los exones 1b, 3, 4a, 5b, 6a, 7b, 8b y 9a). En *M. corti* se han aislado secuencias parciales para estas isoformas y se ha verificado que se expresan en el tejido muscular, haciendo estos genes un marcador molecular de este tejido (Koziol et al., 2011). En la Figura 1. 13 se muestran las isoformas generadas por *maker*, junto a la anotación en la versión WBPS9 y las isoformas parciales aisladas experimentalmente.

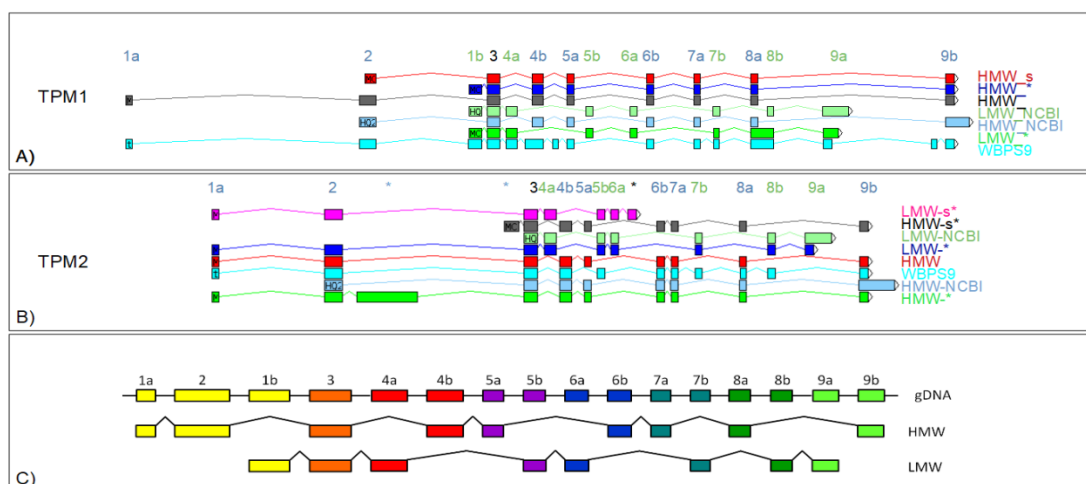


Figura 1. 13: Isoformas de tropomiosina. A) Tropomiosina 1, B) Tropomiosina 2, C) estructura génica de genes de tropomiosina y de las isoformas generadas (Tomado de Koziol et al. (2011)). En A y B los colores rojo, azul, gris, verde y magenta corresponden a isoformas anotadas en *maker*; en cyan se muestra la isoforma anotada en WBPS9, en celeste y verde agua se muestran los exones correspondientes a las isoformas de alto y bajo peso molecular aisladas experimentalmente (N° acceso NCBI: HQ230017 - HQ230020 para TPM1-LMW, TPM1-HMW, TPM2-LMW, TPM2-HMW)

respectivamente). La numeración de los exones se presenta del mismo color que las isoformas en las que se encuentran, excepto el exón 3 que es común a ambas isoformas. Los * en B) indican exones no reportados previamente o conflictivos.

Se puede observar que en este caso *maker* es capaz de recuperar las dos isoformas reportadas para cada gen, además de algunas isoformas adicionales. En cambio, la anotación de *WormBase ParaSite* predice correctamente la mayoría de los exones, pero no distingue las isoformas generadas por *splicing* alternativo. En el caso de TPM1 se predice correctamente una isoforma de alto peso molecular (gris) mientras que la isoforma de bajo peso molecular (verde) incluyó ambos exones 8 y el intrón en el medio de ambos. Además se predicen dos isoformas de alto peso molecular, una que carece del exón 1a (rojo: HMW-s), pero coincide en el resto de los exones y otra que comienza en el exón 1b, correspondiente a las isoformas de bajo peso molecular (azul: HMW-*). Para TPM2 se predijo correctamente una isoforma de alto peso molecular (rojo: HMW) y dos isoformas que incluyen un exón distinto al esperado antes del exón 3. Una de estas isoformas (verde: HMW-*) incluye un exón más largo que los demás entre los exones 2 y 3, algo que se observó para algunos trematodos (Koziol et al., 2011). La otra isoforma comienza en un exón corto no utilizado por ninguna otra isoforma, que podría corresponderse con el exón 1b de acuerdo a su tamaño y ubicación. Por el contrario, las isoformas de bajo peso molecular predichas todas incluyen los exones 1a y 2 característicos de las isoformas de alto peso molecular (magenta: LMW-s* y azul: LMW-*), siendo una de ellas más corta en el extremo C-terminal. Si estas isoformas son verificadas experimentalmente, la diversidad de isoformas generadas por *splicing* alternativo de genes tropomiosina sería mayor a la descrita previamente.

2.9. Análisis de genes de interés.

2.9.1. *Nanos*

Como se mencionó anteriormente, en platelmintos parásitos existen dos genes *nanos*, mientras que en planaria existe sólo uno. *nanos* en planaria es expresado en la línea germinal, al igual que *Sm-nos1* del trematodo *S. mansoni*. Por otro lado, *Sm-nos2* es un marcador de una subpoblación de neoblastos (ver Introducción General, punto 4.3.2.1.3). En la reanotación se recuperaron ambos homólogos de *nanos*, al igual que en la anotación de *WormBase ParaSite*. Se evaluaron las relaciones filogenéticas para determinar la ortología con otros genes *nanos* de platelmintos. En la Figura 1. 14 se muestra el resultado y puede observarse que ambos genes *nanos* de cestodos agrupan con el gen *SmedNos* de planaria y con *Sm-nos1* de *S. mansoni*, mientras que *Sm-nos2* queda por fuera del grupo. Esto podría explicar por qué ambos genes *nanos* se expresan en muy pocas células en el metacestodo de *E. multilocularis* (Koziol et al., 2014). Si como es el caso de *Sm-nos1* y *SmedNos* ambos genes *nanos* de cestodos son marcadores de la línea germinal, la proporción de células que expresan estos genes en los estadios larvarios se espera que sea nula o baja. Esta baja expresión se observa también mediante RT-qPCR en el caso de *M. corti* (ver Capítulo 2). Esto indicaría además que los cestodos carecen de ortólogos de *Sm-nos2*, por lo que esta es una diferencia entre las células proliferativas de cestodos y trematodos (Ver Introducción).

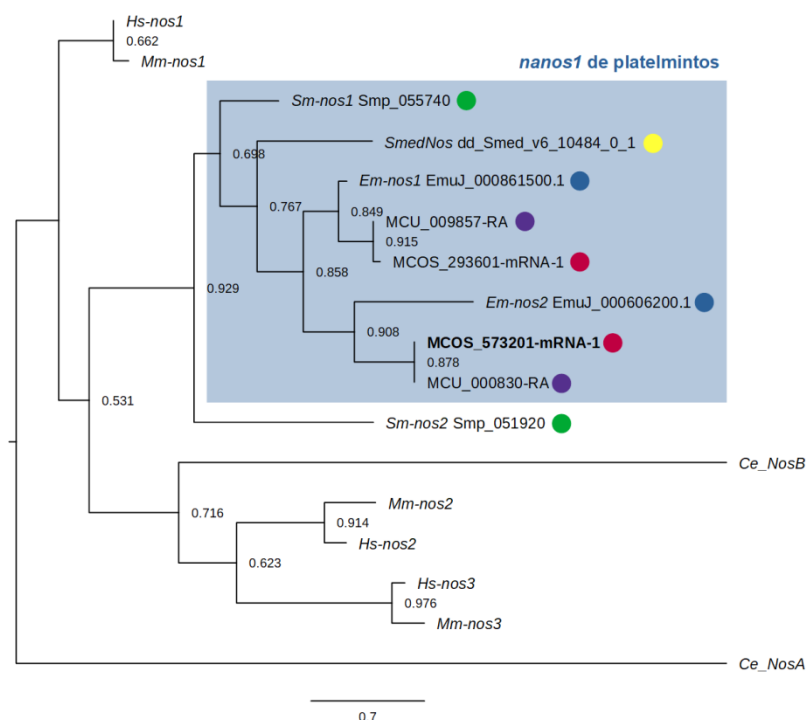


Figura 1. 14: Relaciones filogenéticas de proteínas Nanos de Platelminotos. Se incluyeron en el análisis los dos genes de *M. corti* anotados por *maker* (prefijo “MCU”, violeta) y los dos anotados en la versión WBPS9 de la anotación de *WormBase ParaSite* (Prefijo “MCOS”, rojo), los dos genes de *E. multilocularis* (azul) y de *S. mansoni* (verde), y el gen de *S. mediterránea* (amarillo). Como referencia se incluyeron los genes de *C. elegans*, ratón y humano. Con fondo celeste se señala el grupo monofilético que incluye los genes nanos de cestodos.

2.9.2. Helicasas DEAD-box

La familia de las helicasas DEAD-box es interesante por la ausencia de genes vasa (DDX4) en los genomas de platelmintos parásitos y por el rol de los demás miembros en procesos implicados en el metabolismo del ARN (ver Introducción General, punto 4.3.2.1.4). Se evaluaron las relaciones filogenéticas de los genes anotados con dominio DEAD-box en platelmintos parásitos, incluyendo los miembros representativos de humano y ratón para su clasificación (Figura 1. 15).

En este análisis puede observarse que la reanotación recupera todos los genes pertenecientes a la familia que están presentes en los platelmintos parásitos, y además con el largo suficiente para poder detectar el dominio DEAD-box. En algunos casos se encontraron además varias isoformas (DDX42, DDX46, DDX3_pL10, entre otros). No se encontraron miembros de platelmintos para los clados DDX52 y DDX59 como ya se ha reportado (Tsai, IJ et al., 2013). Estos autores reportan que DDX28 no está presente en los genomas de platelmintos parásitos pero sí está presente en planaria, igual al resultado obtenido en este análisis. Por otro lado, DDX24 reportado como ausente en cestodos por estos autores, sí está presente en *E. multilocularis* y *M. corti*. Nuevamente, las mejoras en la anotación del genoma permiten conocer más sobre la biología del organismo ya que la ausencia de DDX24 llamó la atención por estar implicado en la biogénesis del ribosoma. Otro miembro que parece estar ausente en platelmintos es uno de los miembros del clado DDX19/25, ya que en platelmintos solo se encuentra un gen para cada especie en lugar de dos. Además se observa una duplicación en cestodos del gen DDX6 y se observan al

menos dos duplicaciones en todos los platelmintos dentro del caldo que incluye a DDX2 (eIF4).

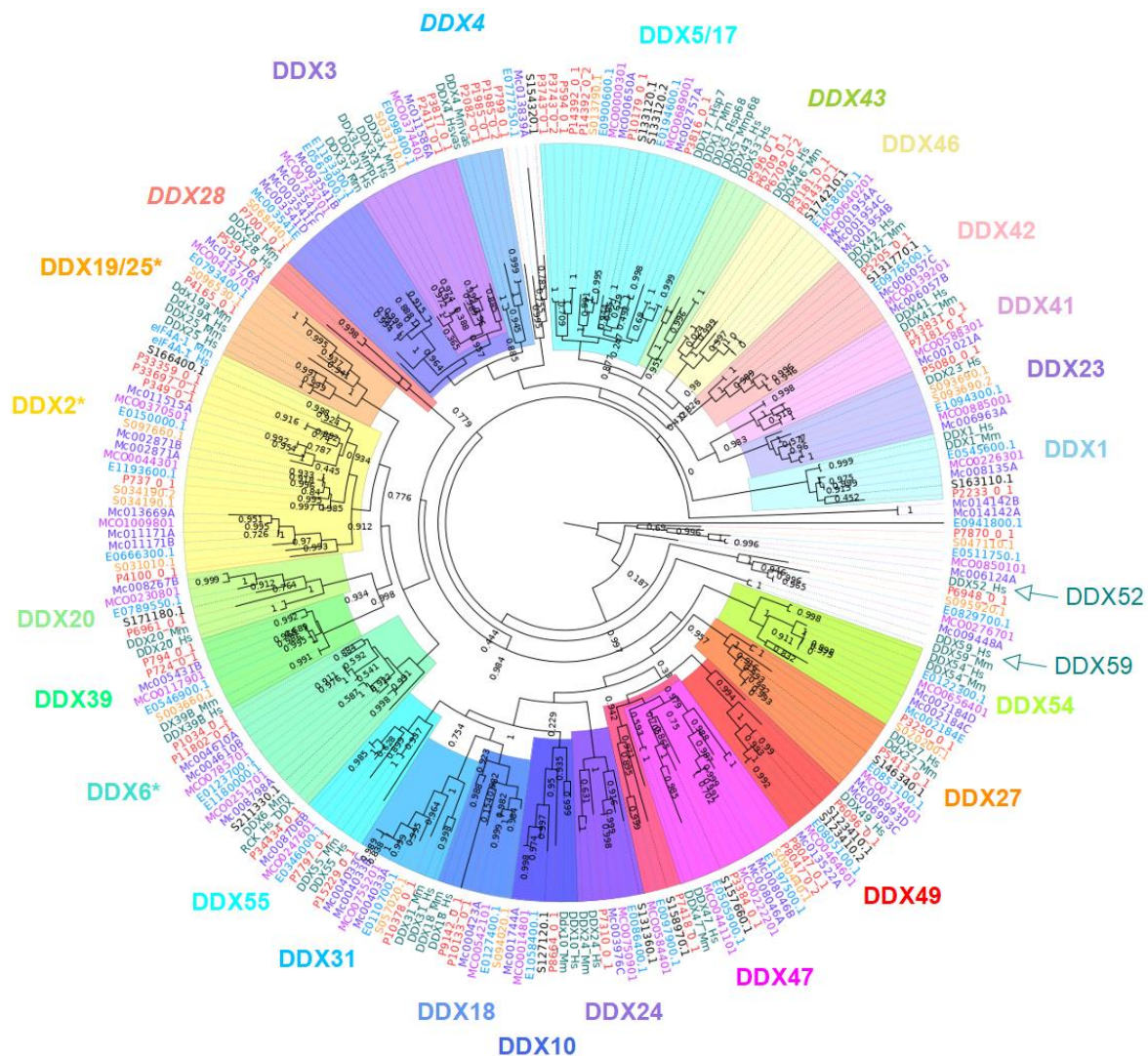


Figura 1. 15: Filogenia de proteínas de la familia DDX de helicasas de ARN. Se incluyeron secuencias proteicas con dominio helicasa DEAD-box de humano y ratón (prefijo DDX, color de hoja verde), *E. multilocularis* (prefijo E, color de hoja celeste), *S. mediterranea* (prefijo P, color de hoja rojo), *S. mansoni* (prefijo S, color de hoja anaranjado) y *M. corti* (reanotación: prefijo Mc, color de hoja violeta, la letra al final del nombre indica la isoforma; anotación WBPS9: prefijo MCO, color de hoja rosado). Se muestran con distintos colores (sombreado en el árbol y nombre del gen) los clados monofiléticos que contienen cada uno de los miembros de la familia. En *itálica* se muestran los genes ausentes en los genomas de platelmintos parásitos (DDX4-vasa, DDX43, DDX28) y con una flecha los ausentes en todos los platelmintos (DDX52 y DDX59). Se indican con un asterisco (*) los genes que presentan algún patrón a destacar.

Para el caso de *vasa* (DDX4) y *pL10* (DDX3) se observa la ausencia de *vasa* y la presencia de una duplicación de *pL10* en platelmintos. El clado que incluye DDX3 de vertebrados (sombreado de color violeta claro) presenta dos miembros en planaria y uno para cada platelminto parásito (en el caso de *M. corti* este fue el gen analizado por qPCR en el Capítulo 2, mientras que el gen de *S. mansoni* corresponde al gen *Sm-vlg1*). El clado relacionado que incluye la duplicación presenta dos genes para *E. multilocularis* y un gen para *S. mansoni* (*Sm-vlg3*) y *M. corti* (la anotación con *maker* produjo 5 isoformas distintas).

Se observa además un clado por fuera del que incluye a *p10* y *vasa* que presenta únicamente miembros de platelmintos parásitos (en *S. mansoni* corresponde al gen *Sm-vlg2*).

3. Discusión

En este trabajo, mediante el uso de datos de secuenciación de ARN se reanotó el genoma de *M. corti*, mejorando la anotación disponible en *WormBase ParaSite*. La estrategia de anotación de *WormBase ParaSite* para la mayoría de los genomas disponibles utiliza el programa *maker*, integrando evidencia de predicciones *ab initio*, la anotación de *C. elegans* y la anotación de genomas, ESTs, ARNm y proteínas de helmintos relacionados (https://parasite.wormbase.org/Mesocetoides_corti_prjeb510/Info/Index/). La estrategia utilizada en este trabajo para la reanotación es similar, pero se incluyeron datos de secuenciación de ARN, lo que permitió mejorar los modelos génicos, produciendo nuevas isoformas y regiones no traducidas, además de encontrar 4.878 genes que no estaban anotados en la versión WBPS9. Esta anotación mejorada será enviada a *WormBase ParaSite* para que esté disponible en próximas versiones de la base de datos.

Se destaca que el 80% de los genes predichos con calidad y largo adecuado tienen algún apoyo de homología que sustente su existencia. De ellos, el 78% (8.287 genes) tienen alguna evidencia de anotación funcional, mientras que los restantes, si bien sin anotación funcional, presentan ortólogos en otros platelmintos. Dentro de los genes nuevos, se encontraron muchos pertenecientes a familias multigénicas que no estaban previamente anotados, como es el caso de miembros de la superfamilia SCP/TAP descrita anteriormente. También se anotaron nuevos miembros de la familia de las Tetraspaninas, proteínas de membrana que cumplen importantes funciones en todos los organismos (Hemler, 2005) y distintos miembros de esta familia son marcadores de neoblastos pluripotentes en la planaria *S. mediterranea* (Zeng et al., 2018) o precursores de células tegumentarias en *S. mansoni* (Collins et al., 2016).

Existen 2.590 genes que solo se encuentran anotados en *M. corti*, de los cuales 854 presentan evidencia de expresión. En la Tabla 1. 5 se muestra la cantidad de genes que únicamente presentan ortólogos en Platelmintos o se encuentran hasta el momento únicamente anotados en *M. corti*. Dentro de los genes únicamente anotados en *M. corti*, una tercera parte son expresados, mientras que el resto presentan valores de TPM menores a 1. Esta tendencia se invierte al considerar las isoformas con ortólogos en platelmintos pero sin evidencia de anotación funcional, lo que le da a estas predicciones aún mayor confiabilidad.

Tabla 1. 5: Genes sin evidencia de anotación funcional

	Expresados	Hipotéticos	Total
Anotados sólo en <i>M. corti</i>	854 (33%)	1736 (67%)	2590
Sólo ortólogos en Platelmintos	1393 (59%)	967 (41%)	2360

Estos genes anotados únicamente en platelmintos son interesantes desde el punto de vista del diseño de estrategias de control parasitario, ya que al no estar presentes en el hospedero, su tratamiento en principio no tendría efectos secundarios. Por este motivo es importante intentar determinar su función, o al menos si son genes esenciales para la

biología del parásito, ya sea en su metabolismo o mediando los procesos de infección o evasión de la respuesta inmune.

El uso de datos de secuenciación de ARN para la reanotación permitió predecir isoformas de *splicing* alternativo, como es el caso de los genes Tropomiosina, para los cuales existe evidencia experimental del uso de exones mutuamente excluyentes (Koziol et al., 2011). El *splicing* alternativo es una estrategia utilizada por los organismos parásitos para evadir la respuesta inmune de su hospedero (Hull y Dlamini, 2014). En particular, en *S. mansoni* se ha probado que algunos miembros de la superfamilia SCP/TAPS presentan isoformas de *splicing* alternativo, las cuales son expresadas diferencialmente en distintos momentos de su ciclo de vida (Chalmers et al., 2008). También se ha observado la existencia de *splicing* alternativo en algunos miembros de esta superfamilia, entre otros genes en los cestodos *E. granulosus* y *E. multilocularis* (Liu et al., 2017). Los resultados obtenidos en este trabajo indican que hay 3.989 genes que presentan más de una isoforma, abriendo la posibilidad a estudiar en profundidad este proceso en *M. corti*. Otro fenómeno que es posible analizar con los datos obtenidos es la existencia de *trans-splicing*, mecanismo por el cual la secuencia denominada *spliced-leader* proveniente de otra región del genoma se une al extremo 5' de un ARNm, y fue reportado en otros platelmintos (Davis et al., 1995; Brehm, Jensen, et al., 2000; Zayas et al., 2005). En este caso, es necesario utilizar transcritos ensamblados *de novo*, para identificar la región 5' no contigua al resto del gen en el genoma. Los transcritos ensamblados *de novo*, pueden también permitir inferir posibles uniones entre los *contigs* o *scaffolds* del genoma. Si un mismo transcripto presenta homología con los extremos de dos *contigs* / *scaffolds* diferentes puede en principio indicar que son contiguos, permitiendo intentar validarlo experimentalmente mediante PCR y secuenciación del fragmento amplificado.

Este trabajo se enfocó en la anotación de genes codificantes para proteínas, pero los resultados obtenidos permiten además estudiar ARNs largos no codificantes. Recientemente se utilizaron datos de secuenciación de ARN para la predicción de ARNinc y ARNs antisentido, mediante herramientas que están disponibles públicamente que pueden ser utilizadas en nuestro modelo de estudio (Vasconcelos et al., 2017).

Los datos generados y los resultados obtenidos permitieron mejorar los modelos génicos existentes, importante para la determinación de la expresión diferencial en células proliferantes de *M. corti*, pero además abren nuevas posibilidades para el estudio de mecanismos de regulación de la expresión génica en este parásito.

CAPÍTULO 2: Efectos de la radiación sobre las células proliferantes de *M. corti*

1. Introducción

Para el estudio de células proliferantes platelmintos a nivel transcriptómico suelen utilizarse dos estrategias. Una es la purificación de células en proliferación por citometría de flujo a partir de suspensiones celulares del organismo. Esta técnica está optimizada para *M. corti* (Domínguez et al., 2014), aunque se obtiene un bajo rendimiento en la cantidad de ARN purificado. Un abordaje alternativo es la irradiación de los organismos para eliminar las células en proliferación. Esta es una técnica más sencilla que la purificación de células por citometría de flujo y es posible obtener cantidades adecuadas de ARN necesarias para la secuenciación, por lo que nos planteamos comenzar con este abordaje para el estudio de las células proliferantes de *M. corti*.

1.1. Irradiación como herramienta para el estudio de células en proliferación.

La irradiación con rayos X o γ es una herramienta utilizada frecuentemente para el estudio de la población de células en proliferación de diversos platelmintos, ya que éstas células desaparecen o al menos disminuyen notoriamente luego de la irradiación. Las dosis utilizadas van desde los 10 hasta los 200 Gy, dependiendo del organismo (Tabla 2. 1). En planarias, una dosis de 60 Gy es letal, ya que produce una eliminación total de neoblastos por lo que el organismo no es capaz de regenerar sus tejidos. En el caso de platelmintos parásitos, la irradiación para el estudio de las células en proliferación se ha utilizado en un número reducido de casos. En trematodos, se utilizó una dosis de 200 Gy para el estudio de células tipo neoblasto en el estadio adulto de *S. mansoni* (Collins et al., 2013). En este estudio identificaron 128 genes característicos de neoblastos de planaria cuya expresión disminuye en organismos irradiados. A partir de ese reporte, otros trabajos han utilizado esta herramienta para el estudio de células tipo neoblastos en *S. mansoni* (Collins et al., 2016; Wendt et al., 2018). En el caso de cestodos, en *E. multilocularis* dosis comparables a la dosis letal utilizada en planaria (entre 50 y 100 Gy) no eliminan las células germinativas, aunque se observa un retraso en el crecimiento y la proliferación del metacestodo luego de 14 semanas, junto a un cambio de forma y reducción de la movilidad de los protoescólex. Además se observan alteraciones ultraestructurales a nivel de la capa laminar en el corto y largo plazo. En particular se observa una capa germinal menos compacta y con pérdida de los contactos célula - célula (Pohle et al., 2011). Más recientemente, se mostró que 2 días luego de aplicar una dosis de 150 Gy la cantidad de células capaces de incorporar EdU disminuye en un 22%, por lo que no las eliminan totalmente y esto no influye en la supervivencia de vesículas (Koziol et al., 2014). En metacestodos de *E. granulosus* se observan alteraciones similares, y además un aumento de la apoptosis dependiente de la dosis (Alam-Eldin y Badawy, 2015).

En suma, en todos los platelmintos estudiados, la irradiación al menos disminuye la cantidad de células en proliferación (Tabla 2. 1), y la dosis necesaria parece ser dependiente de la clase. Estos antecedentes indican que la irradiación para disminuir la cantidad de células proliferantes es una estrategia posible para el estudio de este tipo celular en *M. corti*

Tabla 2. 1: Efecto de distintas dosis de radiación en diferentes platelmintos

Organismo	Dosis (Gy)	Efecto	Referencia
<i>E. multilocularis</i>	50 - 150	Reducción menor en el N° células germinativas (22%). Crecimiento retrasado de vesículas. Alteraciones ultraestructurales	(Pohle et al., 2011; Koziol et al., 2014)
<i>E. granulosus</i>	15, 30, 60	Aumento de apoptosis. Alteraciones ultraestructurales	(Alam-Eldin y Badawy, 2015)
<i>S. mansoni</i>	100 - 250	Eliminación de células tipo neoblasto y su progenie	(Collins et al., 2013, 2016; Wang et al., 2018; Wendt et al., 2018)
<i>S. mediterranea</i>	>60	Letal, eliminación completa de neoblastos, no existe repoblación posterior	(Eisenhoffer et al., 2008; Wagner et al., 2012)
<i>S. mediterranea</i>	10 - 12.5	Subletal, eliminación de la mayoría de los neoblastos, pero existe repoblación posterior	(Wagner et al., 2012)
<i>Dugesia japonica</i>	>15	Letal, eliminación completa de los neoblastos	(Salveti et al., 2009)
<i>Macrostomum lignano</i>	10, 20, 40 y 80	Disminución de mitosis y células en fase S	(Pfister et al., 2007)

Con el objetivo de utilizar esta técnica para el estudio de las células proliferantes en *M. corti*, se comenzó trabajando en la optimización de las condiciones de irradiación para eliminarlas o disminuir su cantidad. Se ensayaron distintas dosis de radiación (100 Gy, 200 Gy, 300 Gy, 500 Gy y 1000 Gy) y se analizaron dos tiempos de recuperación luego de la irradiación (1 día y 5 días). Para cada condición, se analizó el efecto general de la radiación en base a la observación de los organismos en cultivo y se determinó su efecto sobre la cantidad de células proliferantes mediante incorporación de EdU, análisis de ciclo celular y RT-qPCR de genes que son marcadores de neoblastos de planaria.

2. Resultados

2.1. Evaluación de distintas dosis de irradiación en la sobrevivencia de larvas.

Se irradiaron las larvas con diferentes dosis (100, 200, 300, 500 y 1000 Gy) y se analizó la sobrevivencia de los organismos en cultivo *in vitro*. Inmediatamente después de la irradiación, todos los organismos se encuentran en buenas condiciones, sin pérdida de movilidad ni

daños morfológicos apreciables (Figura 2. 1, primera fila de imágenes), lo que permite deducir que la manipulación y transporte no causan deterioro por sí mismos. Transcurrido un día de recuperación (Figura 2. 1, segunda fila de imágenes), los organismos irradiados con dosis más altas (mayor o igual a 300 Gy) presentan claros signos de deterioro, se observan daños en el tegumento que producen hinchazón en algunas regiones del organismo. Los organismos irradiados con dosis más bajas no presentan anomalías morfológicas apreciables.

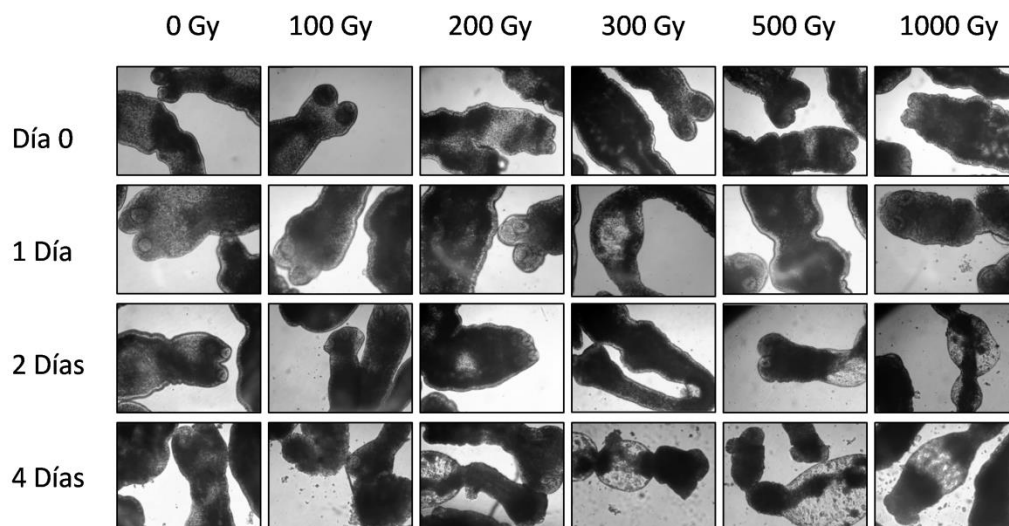


Figura 2. 1: Sobrevivencia de organismos en cultivo tras la irradiación. Tetratiridios irradiados con distintas dosis (indicado en la parte superior) mantenidos en cultivo a distintos tiempos de recuperación (indicado a la izquierda de la imagen).

Tras dos días de recuperación (Figura 2. 1, tercera fila de imágenes), los organismos irradiados con altas dosis continúan deteriorándose, con importante ruptura del tegumento y aparición de fragmentos de organismos probablemente debido a la pérdida de integridad del tegumento. Los irradiados con 200 Gy comienzan a presentar signos de deterioro. Finalmente, a los 4 días de recuperación, todos los organismos irradiados con altas dosis presentan una disminución de movilidad apreciable y la cantidad de fragmentos liberados aumenta. De todas maneras, mantienen sus movimientos característicos a pesar de su deterioro indicando que siguen con vida. Los organismos irradiados con 100 y 200 Gy presentan signos de deterioro, aunque menor que los irradiados con altas dosis, e incluso los organismos control presentan deterioro del tegumento (Figura 2. 1, cuarta fila de imágenes).

2.2. Análisis de incorporación de análogos de timidina (EdU)

Otra estrategia elegida para evaluar el efecto de la radiación fue el estudio de la incorporación del análogo de timidina EdU, que marca células en fase S del ciclo celular. Se evaluaron organismos con 1 día de recuperación y utilizando las dosis 0 Gy, 100 Gy, 200 Gy y 300 Gy. Los resultados (Figura 2. 2) muestran que la incorporación de EdU se ve disminuida con la irradiación, siendo la incorporación menor para dosis más grandes.

La señal en el cuerpo (estróbila), en la región posterior al escólex es prácticamente eliminada para todas las dosis con respecto al control sin irradiar. En la región del escólex, para la dosis de 100 Gy la señal disminuye aunque no llega a desaparecer y para dosis

mayores la señal tiene una intensidad similar a la observada en el control negativo de revelado (organismos no irradiados cultivados sin EdU).

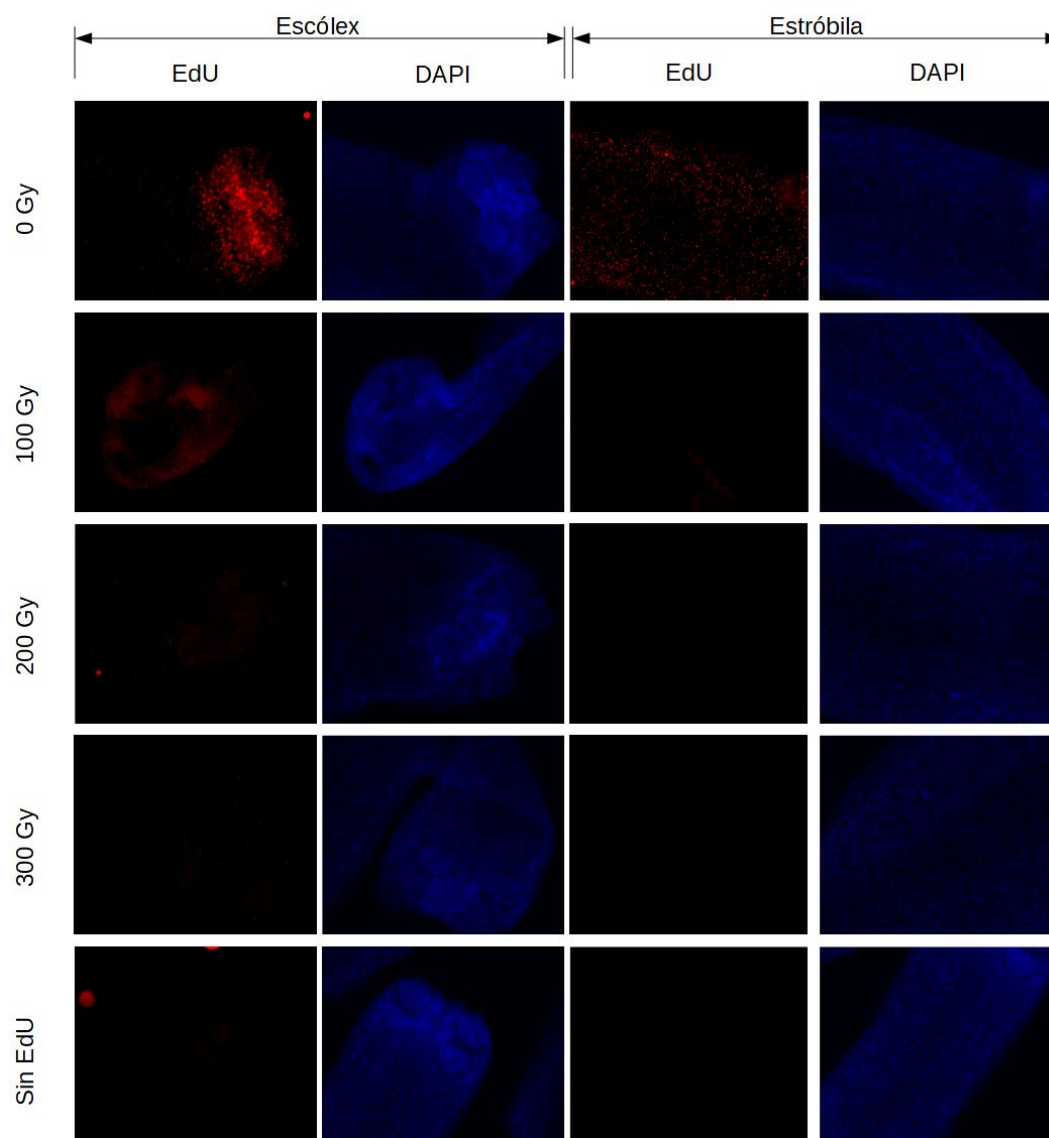


Figura 2. 2: Cambios en la cantidad de células en fase S del ciclo celular tras la irradiación. Células marcadas con EdU en organismos tratados con distintas dosis de radiación (indicado a la izquierda) tras 1 día de recuperación. Se muestra la región del escólex (izquierda) y la región de la estróbila posterior al escólex (derecha). En rojo se muestra la señal del EdU (proliferación) y en azul los núcleos (tinción con DAPI)

2.3. Análisis de ciclo celular

Otro abordaje utilizado para analizar el efecto de la radiación sobre *M. corti* fue el análisis del ciclo celular. Si la radiación afecta las células proliferantes, es de esperar que la proliferación celular se vea alterada, ya sea porque el daño causado en el ADN produce una detención del ciclo celular en espera de la reparación del mismo, o porque ante el daño de al menos una población de célula proliferantes, las que se mantienen funcionales aumenten su tasa de proliferación para contrarrestar el efecto.

Para este estudio se analizaron dosis de 100, 500 y 1000 Gy y dos tiempos de recuperación (1 y 5 días). Se decidió evaluar un tiempo de recuperación mayor para analizar si se

observan diferencias en la proliferación dependiendo del tiempo transcurrido tras la irradiación. Este análisis fue realizado con la colaboración de la Dra. María Fernanda Domínguez, que ha optimizado la técnica de citometría de flujo para la caracterización de células proliferativas de *M. corti*. Se realizaron 3 réplicas de cada condición con organismos obtenidos de un mismo ratón y procesados en paralelo. En la Figura 2. 3 se muestra la reproducibilidad de la técnica, ya que al menos dos de las 3 réplicas realizadas presentan el mismo perfil. Las réplicas que no mostraron picos definidos en el análisis, evidenciando un problema en la tinción o el procesamiento de la muestra, fueron eliminadas.

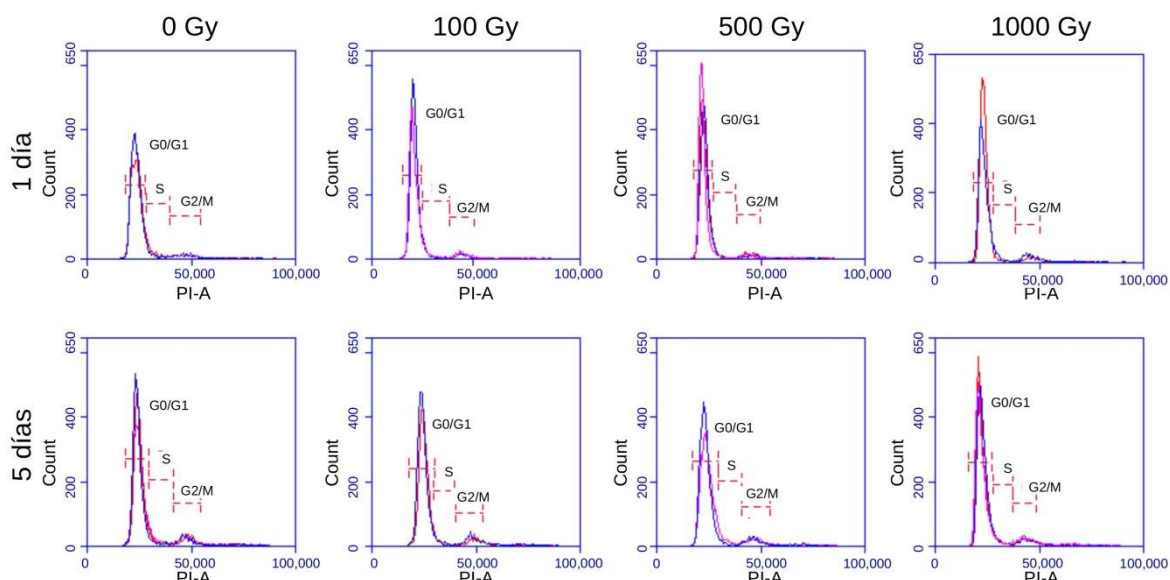


Figura 2. 3: Efecto de la radiación sobre el ciclo celular. Gráficos de ciclo celular (Conteo de células (*Count*) versus área de Ioduro de Propidio (PI-A)). Se muestra en cada panel la superposición de las réplicas para distintas dosis (indicado en la parte superior) y días de recuperación (indicado a la izquierda). Los distintos colores indican las diferentes réplicas ($n=3$ para 0 Gy y 100 Gy con 5 días de recuperación y $n=2$ para el resto).

Luego de un día de recuperación, la fase S del ciclo celular es la que presenta un cambio más marcado en la cantidad de células (Figura 2. 4). En este caso, la población celular que se encuentra en etapa de replicación del ADN disminuye para todas las dosis con respecto al control sin irradiar. Las células en fase de división (G2/M) no presentan grandes cambios. Concordante con esto, hay un aumento de células en fase G0/G1 para los organismos irradiados, indicando que estas células se quedan en una fase de quiescencia (G0) o se detienen en la etapa G1 del ciclo, disminuyendo de esta manera la proliferación celular.

En el caso de organismos con 5 días de recuperación, las diferencias en la fase S no son tan marcadas si bien hay una leve disminución respecto al control sin irradiar (Figura 2. 4). En este caso, además hay una pequeña disminución de las células en fase G2/M. Es importante destacar también el leve aumento en la cantidad de células en la población “Otro”, que implica células por fuera de los *gates* seleccionados para cada fase del ciclo. Estas podrían ser células apoptóticas que suelen formar un pico subG0 en los gráficos de ciclo celular. Este pico no se observa en los gráficos de ciclo celular (Figura 2. 3) por lo que es una hipótesis a probar.

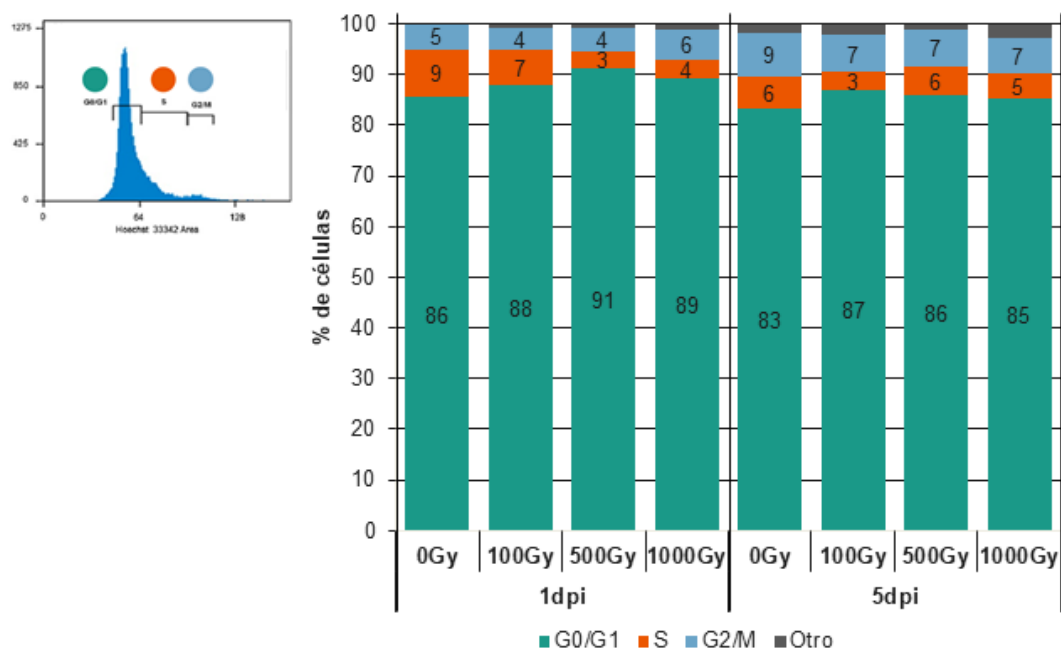


Figura 2. 4: Efecto de la radiación sobre ciclo celular. Gráfico de barras apilado mostrando el porcentaje promedio de células en cada fase del ciclo celular ($n=2$ o 3 , dependiendo de la muestra, ver Figura 2. 3). En la categoría “Otro” se muestran células detectadas por fuera de las fases del ciclo celular. A la izquierda se muestra un gráfico característico de ciclo celular, donde se representan las distintas fases del ciclo con círculos del mismo color que las barras en el gráfico de la derecha.

El análisis estadístico (test de Kruskal-Wallis, valor de significancia del 95%) no muestra que estas diferencias sean significativas, pero el bajo número de réplicas (2 o 3 dependiendo de la muestra) disminuye la potencia del test. Dada la tendencia observada es interesante analizar el número de réplicas necesarias para obtener una potencia adecuada, de manera de evaluar la significancia de los resultados obtenidos.

2.4. Análisis de los niveles de expresión de ARNm de genes marcadores

Se seleccionaron los genes candidatos a marcadores de células proliferativas *nanos*, *pcna*, *pum1*, *pum2* y *pL10* para evaluar cómo cambian sus niveles de expresión frente a la irradiación en *M. corti*.

2.4.1. Validación del ensayo de cuantificación

En primer lugar se evaluó la eficiencia de amplificación (Figura 2. 5A) y la especificidad de los cebadores analizando las curvas de desnaturalización realizadas luego de la amplificación (Figura 2. 5B). Para todos los genes, la eficiencia de amplificación fue muy cercana al 100% en las condiciones ensayadas. Además, de acuerdo al análisis de las curvas de desnaturalización, los productos de amplificación son específicos, existiendo un único pico para cada gen.

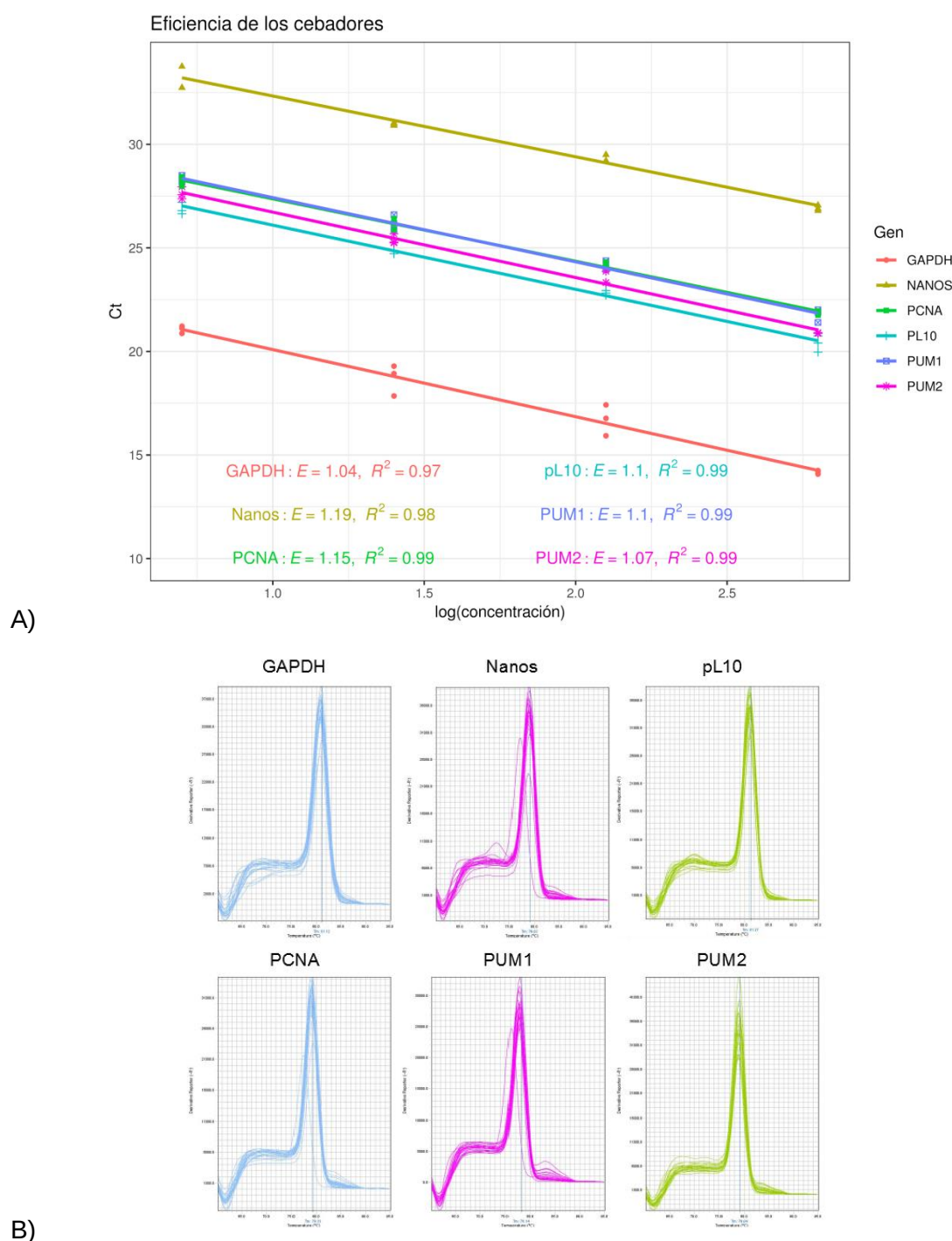


Figura 2. 5: Eficiencia y especificidad de los cebadores. A) Ciclo umbral (Ct) en función del logaritmo de la concentración para cada gen. Utilizando el mismo color de la curva se muestra el valor de eficiencia obtenido (E) y el coeficiente de correlación (R²). B) Curvas de desnaturalización para los distintos genes analizados.

Se eligió *gapdh* como gen normalizador endógeno y se evaluó si presentaba variación en su nivel de expresión en las condiciones ensayadas. Para ello se utilizó el método $2^{-\Delta Ct}$ (Livak y Schmittgen, 2001), siendo $\Delta Ct = Ct_{GEN, dosisX} - Ct_{GEN, dosis0}$. Se evaluó también la variación de los genes que se desea cuantificar (Figura 2. 6). Si un gen no cambia su nivel de expresión entre las distintas condiciones, se espera que el valor de $2^{-\Delta Ct}$ sea 1.

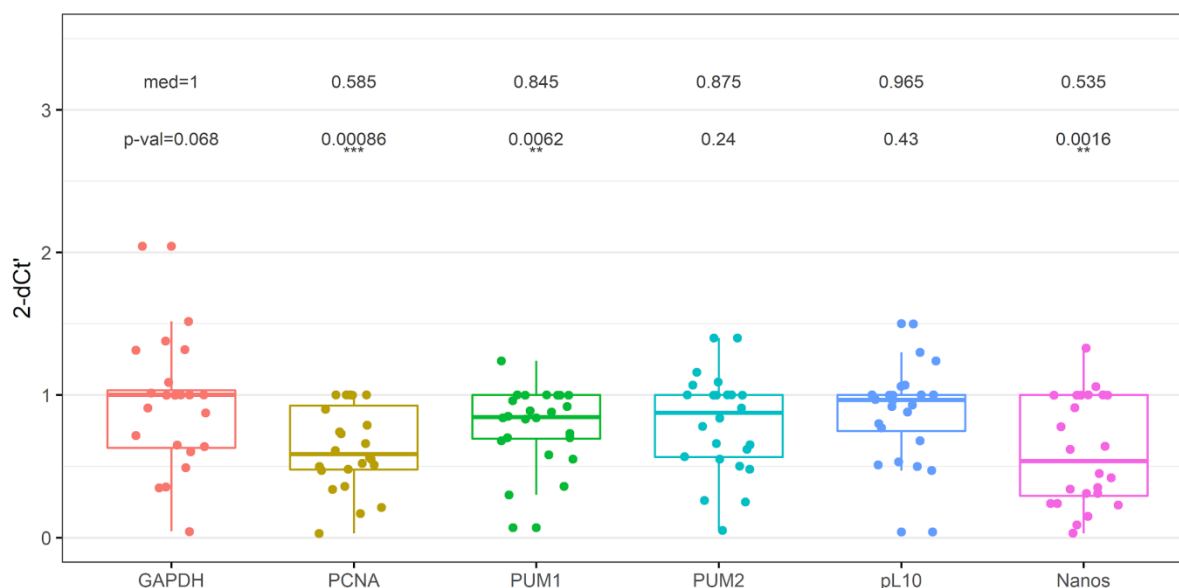


Figura 2. 6: Análisis de la variación de los genes de interés y *gapdh* para las distintas dosis de radiación estudiadas. Se muestra el valor de $2^{-\Delta Ct'}$ ($\Delta Ct' = Ct_{GEN, dosisX} - Ct_{GEN, dosis0}$) para cada gen calculado para cada dosis respecto a los organismos sin irradiar. Encima de cada gráfico se muestra el valor de la mediana y el valor p (test: Kruskal-Wallis, n=24 para cada gen). ***: valor p < 0.001, **: valor p < 0.01

Como se observa en la Figura 2. 6, *gapdh* presenta una dispersión de los datos un poco mayor a la de los demás genes, con dos valores alejados en el extremo superior. De todas maneras, la mediana del valor de $2^{-\Delta Ct'}$ es 1 (rango intercuartil: 0,631 – 1,03) y la diferencia entre los valores para las distintas condiciones no es significativo con un 95% de confianza (Kruskal-Wallis, p=0,068). *pcna* tiene una mediana de 0,585 (rango intercuartil: 0,478 – 0,925) y la diferencia entre los valores es significativa (Kruskal-Wallis, p=0,00086), por lo que se infiere que presenta cambios en los niveles de expresión en al menos una de las condiciones. Algo similar ocurre con *pum1* (mediana: 0,845, rango intercuartil: 0,695 - 1, Kruskal-Wallis, p=0,0062) y con *nanos* (mediana: 0,535, rango intercuartil: 0,292 - 1, Kruskal-Wallis, p=0,0062). Por otro lado, los genes *pL10* (mediana: 0,965, rango intercuartil: 0,748 - 1, Kruskal-Wallis, p=0,43) y *pum2* (mediana: 0,875, rango intercuartil: 0,565 - 1, Kruskal-Wallis, p=0,24) no parecen tener valores alterados por el efecto de la irradiación, de manera similar a lo que ocurre con *gapdh*.

2.4.2. Expresión de *pcna*

Se analizaron los cambios del nivel de expresión de los genes de interés de acuerdo a la dosis de radiación y los días de recuperación. En la Figura 2. 7 se muestran los resultados obtenidos para el gen *pcna*. Analizando globalmente el efecto de la irradiación, sin considerar el tiempo de recuperación (Figura 2. 7A, agrupado 1 día y 5 días de recuperación), el nivel de expresión del gen disminuye de manera significativa para todas las dosis respecto al control sin irradiar (Wilcoxon, valor p 0,0028 para 100 Gy y 500 Gy y 0,049 para 1000 Gy). Cuando se analizan los datos considerando los días de recuperación (Figura 2. 7B), *pcna* disminuye su nivel de expresión tras 1 día de recuperación para todas las dosis analizadas (100, 500 y 1000 Gy, Kruskal-Wallis: 0,023). En este caso, dado que únicamente realizamos 3 réplicas por condición y día de recuperación, las comparaciones

pareadas en todos los casos no dan significativas. De todas maneras, el resultado del test global indica que al menos para una de las dosis el nivel de expresión es diferente a las demás. De acuerdo a los valores de la mediana, se puede inferir que la radiación disminuye los niveles de expresión de *pcna*, implicando un descenso en la proliferación. Tras 5 días de recuperación, los niveles aumentan para las dosis más bajas, y la diferencia global ya no es significativa (Kruskal-Wallis: 0,12).

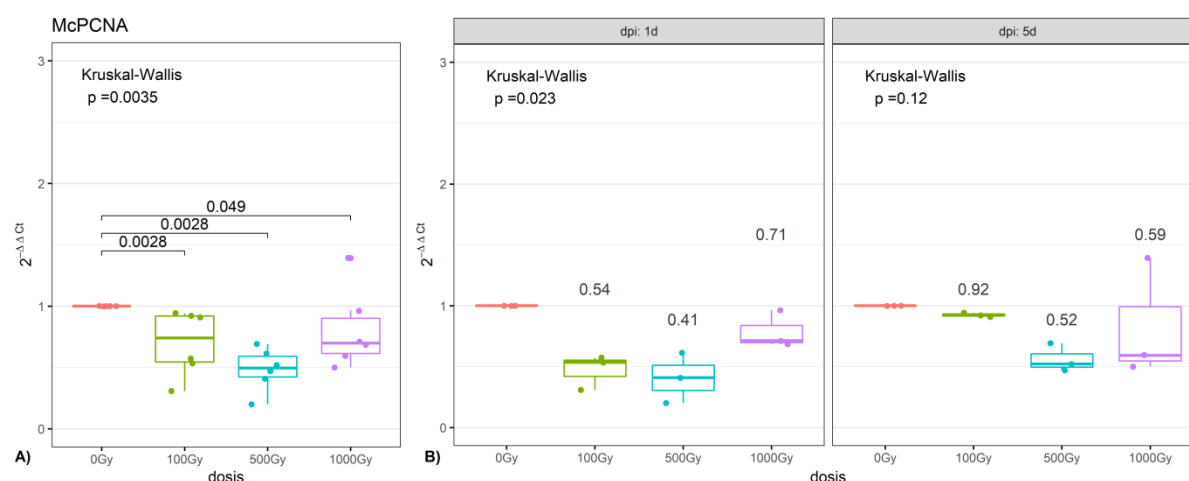


Figura 2. 7: Cambios en los niveles de expresión de *pcna* en respuesta a diferentes dosis de radiación. El análisis se realizó con el método $2^{-\Delta\Delta C_t}$. En todos los casos se muestra el valor p obtenido con el test Kruskal-Wallis. A) *Fold change* respecto al control sin irradiar (0Gy), agrupando ambos tiempos de recuperación. Los números representan el valor p no corregido del test pareado de Wilcoxon. B) *Fold change* respecto al control sin irradiar (0Gy), luego de 1 día de recuperación y 5 días de recuperación (n=3 para cada dosis). El valor de la mediana se presenta sobre cada gráfico.

2.4.3. Expresión de genes *pumilio*

En el caso de los genes *pumilio* (Figura 2. 8 y Figura 2. 9) el efecto de la irradiación sobre su expresión es dependiente de la dosis, no habiendo una tendencia común. Para ambos genes en el análisis global, no teniendo en cuenta el tiempo de recuperación, hay una leve disminución cuando se irradia con una dosis de 100 Gy, mientras que con una dosis de 500 Gy se mantendría en un nivel similar al control y con una dosis de 1000 Gy se observa un aumento con respecto al control sin irradiar.

En el caso de *pum1*, la diferencia global con el control sin irradiar (Figura 2. 8A) es significativa en al menos una condición (Kruskal-Wallis: 0,014), siendo la diferencia para la dosis de 100 Gy la más significativa (Kruskal-Wallis: 0,0028). Al discriminar por días de recuperación (Figura 2. 8B) se observa la misma tendencia, siendo más marcada la diferencia en la expresión entre las distintas dosis luego de un día de recuperación comparado con cinco días de recuperación. Estas diferencias sin embargo no son significativas (1 día: Kruskal-Wallis: 0,089; 5 días: Kruskal-Wallis: 0,24).

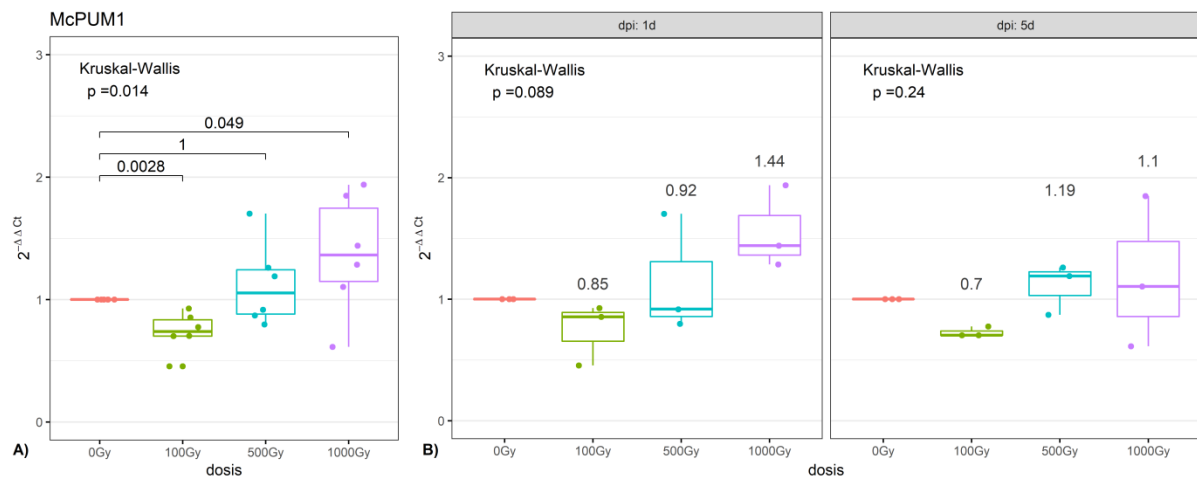


Figura 2. 8: Cambios en los niveles de expresión de *pum1* en respuesta a diferentes dosis de radiación. El análisis se realizó con el método $2^{-\Delta\Delta C_t}$. En todos los casos se muestra el valor p obtenido con el test Kruskal-Wallis. A) *Fold change* respecto al control sin irradiar (0Gy), agrupando ambos tiempos de recuperación. Los números representan el valor p no corregido del test pareado de Wilcoxon. B) *Fold change* respecto al control sin irradiar (0Gy), luego de 1 día de recuperación y 5 días de recuperación (n=3 para cada dosis). El valor de la mediana se presenta sobre cada gráfico.

En el caso de *pum2*, el análisis sin considerar los días de recuperación muestra que en al menos una de las condiciones la diferencia del nivel de expresión con el control sin irradiar es significativa (Kruskal-Wallis: 0,0092), siendo la comparación pareada significativa para la dosis de 100 Gy (Kruskal-Wallis: 0,0028). Este comportamiento es similar al observado para *pum1*. Para dosis mayores, la diferencia con el control no es significativa, por lo que no hay cambios en el nivel de expresión de este gen (Figura 2. 9A). Al discriminar por días de recuperación (Figura 2. 9B), luego de 1 día, las diferencias no son significativas (Kruskal-Wallis: 0,18), siendo el nivel de expresión similar al control para el caso de las dosis mayores y para la dosis de 100 Gy levemente inferior. Sin embargo, luego de 5 días de recuperación, la diferencia global del nivel de expresión es significativa (Kruskal-Wallis: 0,042), indicando que al menos una de las condiciones es diferente al control. El valor de expresión para la dosis de 100 Gy es aproximadamente la mitad del observado en el control, mientras que para las dosis más altas el nivel es similar.

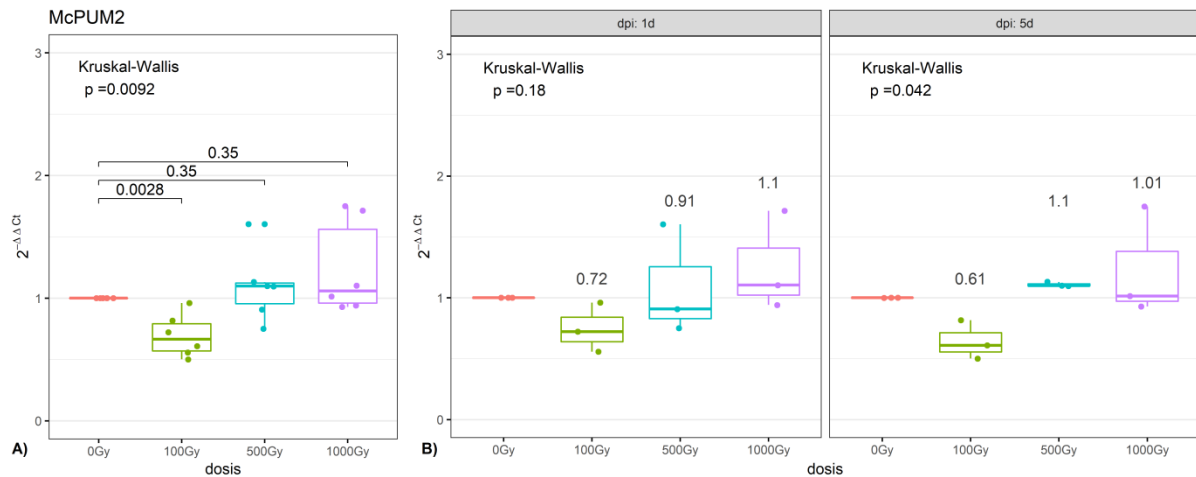


Figura 2. 9: Cambios en los niveles de expresión de *pum2* en respuesta a diferentes dosis de radiación. El análisis se realizó con el método $2^{-\Delta\Delta Ct}$. En todos los casos se muestra el valor p obtenido con el test Kruskal-Wallis. A) *Fold change* respecto al control sin irradiar (0Gy), agrupando ambos tiempos de recuperación. Los números representan el valor p no corregido del test pareado de Wilcoxon. B) *Fold change* respecto al control sin irradiar (0Gy), luego de 1 día de recuperación y 5 días de recuperación (n=3 para cada dosis). El valor de la mediana se presenta sobre cada gráfico.

2.4.4. Expresión de nanos

El gen *nanos*, que es el principal interactor de proteínas Pumilio en los organismos estudiados, presenta cambios significativos en sus niveles de expresión como respuesta a la radiación (Figura 2. 10), habiendo una reducción a la mitad del nivel de expresión. En el análisis global se observa que para todas las dosis analizadas hay una disminución significativa respecto al control (Figura 2. 10A; Kruskal-Wallis: 0,049 para 100 Gy y 0,0028 para 500 y 1000 Gy). Esta tendencia se mantiene cuando se discrimina por los días de recuperación (Figura 2. 10B, Kruskal-Wallis 0,047 y 0,036 para 1 y 5 días de recuperación respectivamente).

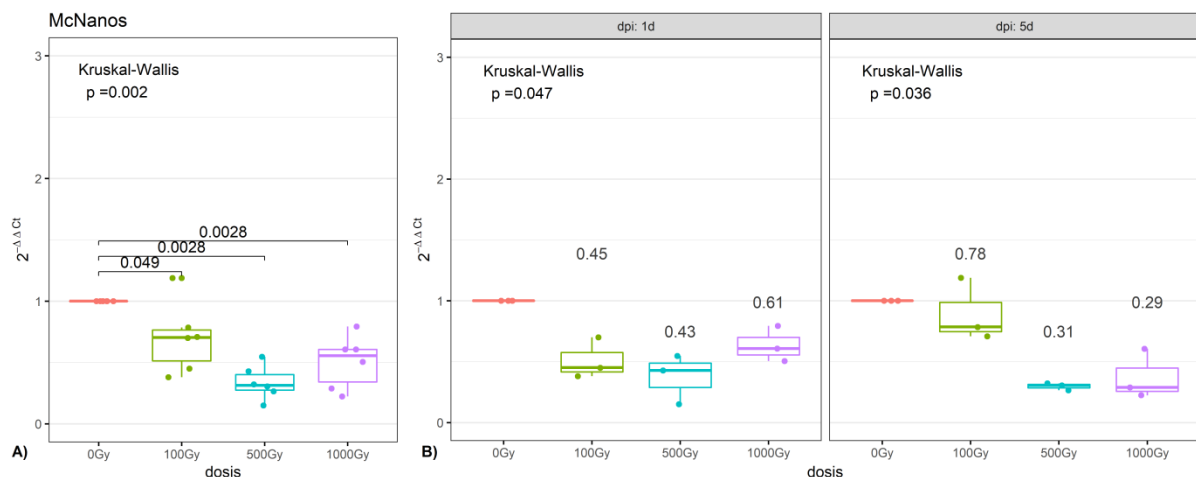


Figura 2. 10: Cambios en los niveles de expresión de *nanos* en respuesta a diferentes dosis de radiación. El análisis se realizó con el método $2^{-\Delta\Delta Ct}$. En todos los casos se muestra el valor p obtenido con el test Kruskal-Wallis. A) *Fold change* respecto al control sin irradiar (0Gy), agrupando ambos tiempos de recuperación. Los números representan el valor p no corregido del test pareado de Wilcoxon. B) *Fold change* respecto al control sin irradiar (0Gy), luego de 1 día de recuperación y 5 días de recuperación (n=3 para cada dosis). El valor de la mediana se presenta sobre cada gráfico.

2.4.5. Expresión de *pL10*

Por último, el gen *pL10* no muestra cambios significativos en el nivel de expresión globalmente (Figura 2. 11A) ni discriminando por día de recuperación (Figura 2. 11B). De todas maneras se observa una tendencia hacia la disminución respecto al control para las dosis más bajas luego de un día de recuperación.

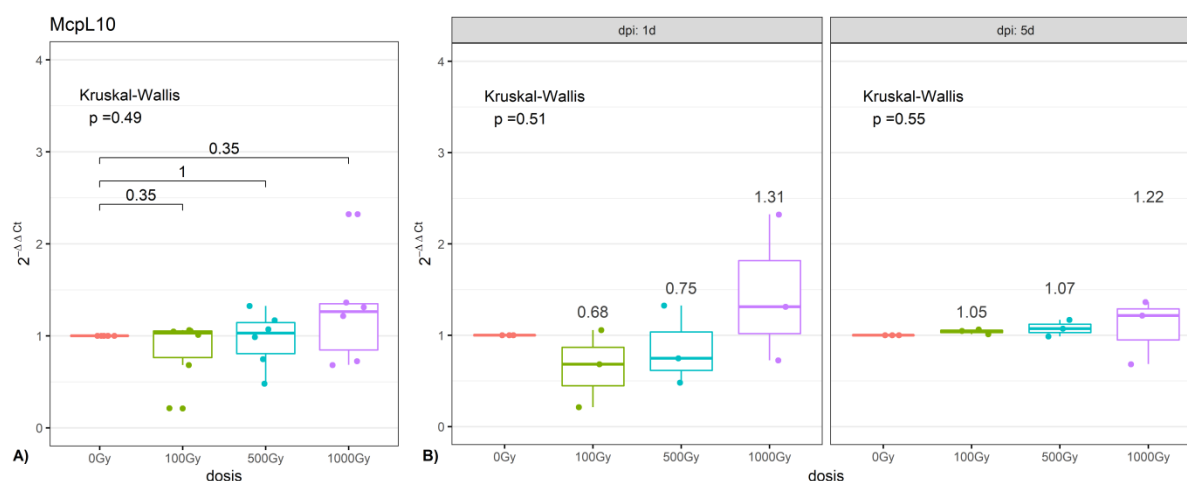


Figura 2. 11: Cambios en los niveles de expresión de *pL10* en respuesta a diferentes dosis de radiación. El análisis se realizó con el método $2^{-\Delta\Delta C_t}$. En todos los casos se muestra el valor p obtenido con el test Kruskal-Wallis. A) *Fold change* respecto al control sin irradiar (0Gy), agrupando ambos tiempos de recuperación. Los números representan el valor p no corregido del test pareado de Wilcoxon. B) *Fold change* respecto al control sin irradiar (0Gy), luego de 1 día de recuperación y 5 días de recuperación (n=3 para cada dosis). El valor de la mediana se presenta sobre cada gráfico.

3. Discusión

Los diferentes experimentos realizados para evaluar el efecto de la radiación sobre las células proliferantes de *M. corti*, muestran una tendencia hacia la disminución de las células en fase S del ciclo celular (incorporación de EdU, análisis del ciclo celular y disminución de niveles de expresión de *pcna*). Esto puede implicar que al tener su ADN dañado por la radiación las células proliferantes se detienen en la fase G0 del ciclo celular, esperando por la reparación del ADN, o que son eliminadas como se ha reportado para otros organismos.

La morfología de los organismos no se encuentra mayormente afectada para las dosis bajas, pero para dosis mayores los organismos se deterioran con el paso del tiempo, indicando que el efecto de la radiación es mayor. Para dosis de 200 Gy o mayores los organismos dejan de sintetizar ADN, ya que la incorporación de EdU desaparece. Esto también se ve reflejado por la disminución de células en fase S observada en el análisis de ciclo celular.

Cuando se comenzó este trabajo, aún no existía el genoma de *M. corti*, por lo que nos enfocamos en genes de *M. corti* aislados en nuestro laboratorio que son homólogos a genes marcadores de neoplastos en planarias.

pcna además de estar implicado en la replicación del ADN, y por lo tanto asociado a procesos de proliferación celular, cumple un rol durante la reparación del daño al ADN. El

incremento de la expresión observado en organismos irradiados con una dosis de 100 Gy luego de 5 días de recuperación podría deberse a un aumento en la proliferación o en la reparación del ADN. Evaluar genes específicos de la vía de reparación de ADN permitirá probar cuál de los mecanismos es el que se desencadena en respuesta a la radiación.

Los genes *pumilio* tienen un rol importante en el mantenimiento de las células madre y además son expresados en el sistema nervioso. En planaria, la irradiación produce una disminución de la expresión de *DjPum* asociada a neoblastos pero se mantiene su expresión en el sistema nervioso (Salveti et al., 2005). En base a los resultados obtenidos en *M. corti* parecería haber un comportamiento similar; si bien disminuye la expresión de ambos genes para la dosis de 100 Gy, al aplicar dosis más altas el nivel de expresión de *pum2* se mantiene, mientras que la expresión de *pum1* aumenta con una dosis de 1000 Gy y se mantiene con la de 500 Gy. Esto podría indicar que *pum1* se asocia al mantenimiento de células proliferantes, y que para las dosis más altas, cuando el daño causado posiblemente sea mayor, el organismo responde con un incremento en la expresión de este gen buscando mantener estas células.

En trematodos y cestodos existen dos genes *nanos*. El análisis filogenético de estos genes (ver Capítulo 1) indica que el gen *nanos* que se analizó en este caso es ortólogo a *Sman-nanos1*, que es marcador de la línea germinal, al igual que *SmedNos*. En base a esto podemos postular que éste gen *nanos* en *M. corti* está implicado en la línea germinal y parecería que estas células son las más afectadas por la radiación, lo que aseguraría que ante daños del ADN éste no sea transferido a través de los gametos a la descendencia. Además, *nanos* es el gen que presenta menor nivel de expresión de los genes analizados, ya que tiene valores de Ct más altos (ver Figura 2. 5A) consistente con que en el estadio larvario los órganos reproductivos aún no están desarrollados y comienzan a observarse acumulaciones de células que darán lugar al primordio genital a medida que el gusano se desarrolla. De ser correcta esta hipótesis, con 7 días de cultivo ya existiría una expresión baja de este marcador de la línea germinal, indicando que una población de células proliferantes ya fue destinada a su formación. Cabe destacar que este resultado es opuesto al observado para los genes *pumilio*, genes cuyos productos proteicos interaccionan con Nanos en los organismos estudiados. Nanos media su represión interactuando con Pumilio, pero se ha postulado que Pumilio podría actuar independiente de Nanos. Es posible entonces que los genes *pumilio* aumenten su expresión para mantener las células proliferantes o para cumplir su función en el sistema nervioso independiente de Nanos.

En *M. corti*, *pl10* se expresa en células con morfología proliferativa, pero también se observa señal por hibridación *in situ* sobre células con morfología diferenciada, mientras que en otros organismos es un marcador de la línea germinal (ver Punto 4.3.2.1.4 de la Introducción general). En particular en *S. mansoni* los tres genes *vlg* (*vasa-like genes*) estudiados se expresan en el ovario (Skinner et al., 2012). En el estadio larvario analizado en este trabajo los órganos reproductores no se encuentran aún desarrollados, por lo que la expresión observada se debería a la expresión de *pL10* en células proliferativas somáticas, en precursores de la línea germinal, como es el caso de *nanos*, o en células diferenciadas. Si esta expresión se da en células diferenciadas los resultados son consistentes con lo observado en otros organismos, en los que la radiación no afecta a las células diferenciadas. Si por el contrario, la expresión se da en células proliferativas, estas formarían una población que no es afectada por la radiación. De acuerdo a estos resultados, este gen no es un marcador adecuado para el análisis de los efectos de la

irradiación en el estadio larvario.

Es interesante que todos los genes analizados al compararlos en conjunto (Figura 2. 12), presentan un comportamiento similar frente a la radiación. Un día luego de la irradiación, disminuyen su expresión para la dosis menor (100 Gy). Se esperaría que para dosis mayores se mantuviera esta disminución, sin embargo para 500 Gy aumenta o disminuye levemente respecto a 100 Gy y para la dosis de 1000 Gy aumenta para todos los genes. Para saber si este aumento se debe a un incremento en la proliferación se podría realizar marcado con EdU o BrdU para contar la cantidad de células marcadas, idealmente a diferentes tiempos post-irradiación. Cuando se comparan los niveles de expresión luego de 5 días, los comportamientos son más dispares. Ambos genes *pumilio* se comportan de la misma manera: para la dosis de 100 Gy disminuyen aún más que luego de 1 día; aumentan para la dosis de 500 Gy pero disminuyen para la dosis de 1000 Gy (en todos los casos comparando con 1 día de recuperación), siendo *pum1* el que más aumenta para 1000 Gy. *nanos* tiene un comportamiento opuesto: se recupera para la dosis de 100 Gy pero cae para las dosis mayores. Saber si Pumilio es capaz de interactuar con Nanos en estas condiciones podría explicar este comportamiento opuesto. Es posible que Pumilio esté inactivando a Nanos promoviendo la degradación de su ARNm, por lo que evaluar los niveles de proteína Nanos y analizar por *Northern blot* el ARNm nos podría indicar si esta hipótesis es válida en caso de observar degradación del ARNm y la consecuente disminución de la cantidad de proteína o de su largo. Disponemos de anticuerpos policlonales para ambas proteínas Pumilio y estamos trabajando para la producción de Nanos recombinante para generar un anticuerpo. Evaluar qué ocurre con el otro gen *nanos* de *M. corti* también contribuirá a comprender estos temas.

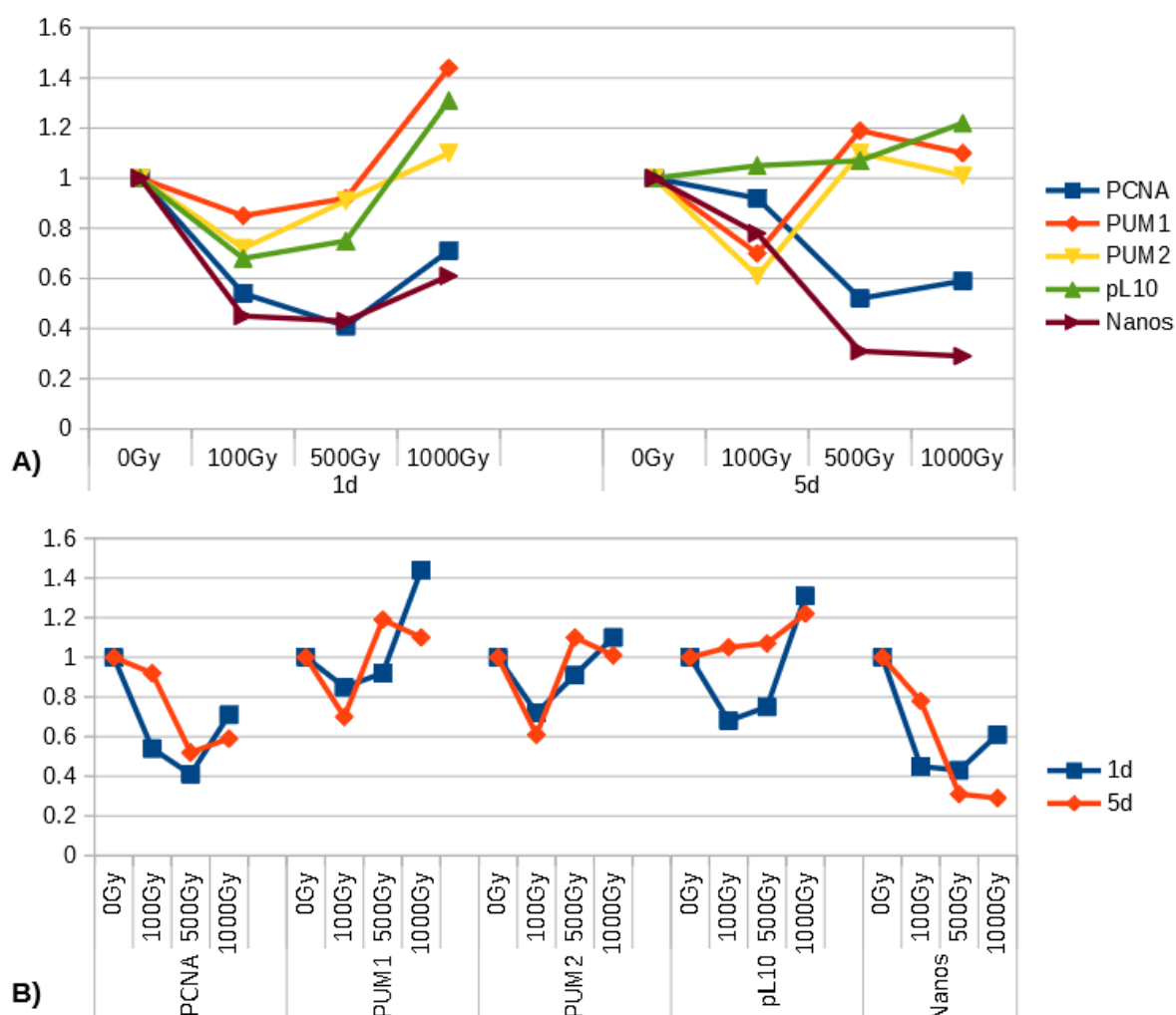


Figura 2. 12: Comparación de los niveles de expresión de los genes estudiados. A) Comportamiento de cada gen en las 3 dosis ensayadas y los distintos días de recuperación, B) Comparación del nivel de expresión de acuerdo a los días de recuperación para cada gen en las distintas dosis

En este trabajo nos propusimos realizar un análisis exploratorio, por lo que se evaluaron distintas dosis de radiación y tiempos de recuperación (en total 8 muestras diferentes). Esto llevó a que la cantidad de réplicas (3) que se pudieron realizar para procesar en paralelo no fue suficiente para obtener la potencia necesaria en los test estadísticos. La variabilidad observada al trabajar con réplicas biológicas obtenidas de distintos ratones y procesadas en distintos momentos fue muy amplia, por lo que hubiera sido deseable trabajar con más réplicas. En base a estos resultados, se plantea como perspectiva seleccionar los genes que presentan resultados más prometedores y evaluar el efecto de la radiación en un mayor número de réplicas. Una alternativa que se podría evaluar en *M. corti* es el uso de hidroxiurea para eliminar las células que se encuentran en fase S del ciclo celular, como se realizó exitosamente en *E. multilocularis*.

Dado que posteriormente al análisis de estos datos se hizo público el genoma de *M. corti* y se cuenta con modelos génicos de calidad (ver Capítulo 1), es interesante analizar la expresión de otros genes frente a la radiación. Algunos candidatos relevantes pueden ser

genes implicados específicamente en la reparación del ADN o proteínas reguladoras del ciclo celular para evaluar si hay un detenimiento en alguna de las etapas del ciclo. Analizar por hibridación *in situ* la expresión de estos genes marcadores, junto a la incorporación de análogos de timidina permitirá evaluar si estos se expresan específicamente en células proliferantes y si estas células desaparecen tras la irradiación o sólo dejan de expresar los genes marcadores. Por otro lado, experimentos de pulso y caza con análogos de timidina a distintos tiempos de recuperación tras la irradiación permitirán analizar la dinámica de las células madre y si son capaces de repoblar los organismos irradiados.

Una dosis de 1000 Gy es más de 10 veces mayor que la dosis letal utilizada en planarias y 5 veces mayor que la dosis aplicada en el adulto de *S. mansoni* para eliminar las células tipo neoblasto (Collins et al., 2013). En el caso de *M. corti*, si bien disminuye la cantidad de células, parecería que las células proliferantes son más resistentes a la radiación que las de otros platelmintos. Es posible que existan diferencias a nivel de la expresión génica de las células madre, o en estructuras corporales que puedan estar protegiendo a las células proliferantes lo que las hace aparentemente más resistentes a la radiación. Dentro de estas estructuras, el tegumento, común a trematodos y cestodos, puede estar implicado en la diferencia con planarias, que tienen un epitelio simple recubriendo su cuerpo. Por otro lado, los corpúsculos calcáreos son muy abundantes en el estadio larvario de cestodos y pueden estar absorbiendo parte de la radiación debido a la presencia de minerales de alto peso molecular, como ocurre con el calcio de los huesos frente a los rayos X.

El objetivo de este estudio fue evaluar los efectos de la radiación sobre las células proliferantes y así determinar una dosis de radiación adecuada para realizar el análisis transcriptómico de este tipo celular. Si bien para todas las dosis se observa un efecto, para las más altas hay alteraciones morfológicas notorias lo cual puede enmascarar el análisis de expresión diferencial tras la disminución en la cantidad de células proliferantes. Por este motivo seleccionamos la dosis más baja (100 Gy) de manera de afectar este tipo celular, pero no producir un cambio drástico en todo el organismo como respuesta a la radiación.

CAPÍTULO 3: Análisis de los perfiles de expresión de células proliferantes y células diferenciadas de *M. corti*.

1. Introducción

Los primeros estudios transcriptómicos de células con capacidad proliferativa en platelmintos se realizaron en planarias. En estos primeros trabajos se compararon los genes que no se expresan en planarias irradiadas (reducidas en neoblastos) pero que sí se expresan en el control sin irradiar (Eisenhoffer et al., 2008; Rossi et al., 2008). Este abordaje tiene el problema de que la irradiación además de eliminar los neoblastos de las planarias, produce un gran número de células apoptóticas y causa efectos indirectos relacionados a las respuestas al estrés. Por el contrario, el transcriptoma de neoblastos y células diferenciadas purificadas por citometría de flujo permite conocer específicamente los genes expresados por cada población celular (Labbé et al., 2012; Önal et al., 2012). Debido a que los neoblastos son las únicas células con capacidad proliferativa, los genes expresados preferentemente por estas células están implicados en la replicación y regulación del ciclo celular. Además expresan genes reguladores de mecanismos de desarrollo en otros animales, reguladores epigenéticos, factores de transcripción, y componentes de vías de señalización, incluyendo receptores de FGF (*Fibroblast Growth Factor*) (Ogawa et al., 2002; Rouhana et al., 2010; Aboobaker, 2011; Wagner et al., 2012; Rink, 2013). Además, expresan específicamente componentes de Programa de Multipotencia Germinal (GMP: *Germline Multipotency Program*), y de manera más general, proteínas relacionadas a regulación de expresión génica (Juliano et al., 2010; Rouhana et al., 2010). Mediante estudios de interferencia de ARN se ha podido relacionar a estos genes con el mantenimiento, proliferación y diferenciación de neoblastos (Eisenhoffer et al., 2008). Varios genes *piwi* son expresados en neoblastos de planaria, y *smewi-1* de *S. mediterranea* y su ortólogo en otras especies es el marcador molecular de neoblastos más utilizado. Aunque no es muy claro el fenotipo observado por interferencia de ARN para *smewi-1*, otros genes *piwi* (*smewi-2* y *smewi-3*) han mostrado ser necesarios para la diferenciación y auto renovación de los neoblastos (Reddien, 2005; Palakodeti et al., 2008).

El estudio de este tipo celular en platelmintos parásitos es más reciente y existen pocos trabajos en los que se analice expresión génica en células en proliferación, existiendo estudios transcriptómicos únicamente para el trematodo *S. mansoni* (Collins et al., 2013, 2016; Wang et al., 2013; Wendt et al., 2018). Los estudios realizados, basados principalmente en la estrategia de irradiación, describen un conjunto de genes marcadores de células tipo neoblasto y su progenie (ver la Introducción general por más detalles). No existen datos publicados hasta el momento de estudios transcriptómicos de células proliferantes de cestodos.

Existen unos pocos genes marcadores de células tipo neoblasto de *S. mansoni* validados mediante interferencia de ARN y/o expresión del ARNm en células que incorporan análogos de timidina (hibridación *in situ* y marcado con EdU o BrdU) (Ver Tabla 3. 1 por genes y referencias). Muchos de estos marcadores de *S. mansoni* son homólogos a marcadores de neoblastos de planaria.

Tabla 3. 1: Genes caracterizados en células tipo neoblasto de *S. mansoni*

Gen	Descripción	Fenotipo ARNi	Referencia	Identificador	Ortólogo en <i>M.corti</i> *
<i>p53</i>	Supresor de tumores	nd	(Collins et al., 2013; Wendt et al., 2018)	Smp_139530	MCU_008396
<i>sox</i>	Factor de transcripción	nd	(Collins et al., 2013)	Smp_076600	MCU_004703/ MCU_008125/ MCU_002410
<i>fgfrA</i>	Receptor de factor de crecimiento de fibroblasto	Impide la proliferación, Disminución de células en proliferación <i>nanos-2</i> ⁻	(Collins et al., 2013; Wang et al., 2013)	Smp_175590	MCU_013052
<i>ago-2</i>	Desarrollo línea germinal	Disminución de células en proliferación <i>nanos-2</i> ⁻	(Collins et al., 2013; Wang et al., 2013)	Smp_179320	MCU_008922/ MCU_011167/ MCU_011738
<i>nanos-2</i>	Desarrollo línea germinal	Degeneración de los testis y ausencia de células germinales diferenciadas	(Collins et al., 2013; Wang et al., 2013, 2018)	Smp_051920	Ninguno
<i>nanos-1</i>	Desarrollo línea germinal	Degeneración de los testis y ausencia de células germinales diferenciadas	(Wang y Collins, 2016; Wang et al., 2018)	Smp_055740	MCU_000830/ MCU_009857
<i>vlg-3</i>	Desarrollo línea germinal	Disminución de todas las células en proliferación	(Wang et al., 2013)	Smp_068440	MCU_003541
<i>h2a</i>	Histona	nd	(Wang et al., 2013)	Smp_086860	MCU_000483
<i>h2b</i>	Histona	nd	(Wang et al., 2013)	Smp_108390/ Smp_121380	MCU_003226
<i>brca1</i>	Proteína de susceptibilidad al cáncer de mama humano	Necesario para el mantenimiento de neoblastos	(Wendt et al., 2018)	Smp_137090	MCU_006277
<i>bard1</i>	Proteína con dominio RING asociado a BRCA1	Necesario para el mantenimiento de neoblastos	(Wendt et al., 2018)	Smp_137090	MCU_011829
<i>fgfr1/4</i>	Receptor de factor de crecimiento de fibroblasto	Necesario para el mantenimiento de neoblastos	(Wendt et al., 2018)	Smp_157300	MCU_006790/ MCU_008419/ MCU_008421
<i>zfp-1</i>	Proteína con dedos de zinc	Disminuye 50% las células <i>tsp2</i> ⁺	(Wendt et al., 2018)	Smp_145470	MCU_004294/ MCU_005537
<i>zfp-1-1</i>	Proteína con dedos de zinc	Disminuye 50% las células <i>tsp2</i> ⁺	(Wendt et al., 2018)	Smp_049580	MCU_002177
<i>eled</i>	Antagonista de <i>nanos</i>	Aumenta expresión de <i>nanos-2</i> en PGZ	(Wang et al., 2018)	Smp_041540	Ninguno

* Ortólogo determinado por análisis de ortología y *blastp* entre las proteínas de *S. mansoni* y las proteínas resultantes de la reanotación descrita en el Capítulo 1 para *M. corti*.

2. Resultados

2.1. Optimización de la extracción de ARN de células purificadas y construcción de librerías de secuenciación.

Para la obtención de células proliferantes y células diferenciadas, se purificaron células por

citometría de flujo de acuerdo a su contenido de ADN: Las células en fases G0/G1 del ciclo celular contienen una cantidad 2C de ADN y es una población que está enriquecida en células diferenciadas. Por otro lado, las células en fases S/G2/M, contienen 4C de ADN ya que han duplicado su ADN y se encuentran en proliferación. La purificación de células en el caso de *M. corti* produce del orden de 1×10^6 células en fases G0/G1 y 1×10^5 células en fases S/G2/M (ver Tabla 3. 2).

Para la extracción de ARN a partir de estas poblaciones celulares se ensayaron distintos kits de extracción especializados para partir de poca cantidad de tejido o células. Las extracciones a partir de un único experimento de purificación dieron cantidades de ARN muy bajas, menores al límite de detección del kit de cuantificación utilizado (*Qubit Broad Range kit*, valor mínimo detectado 0.05 ng/μl). Se decidió entonces juntar tres experimentos de purificación para realizar la extracción con una mayor cantidad de células. El kit *DirectZol RNA Miniprep* fue el que dio mejores rendimientos en estas condiciones. Para los *pooles* de tres purificaciones de células en fase S/G2/M se obtuvo en promedio 125 ng de ARN total y para los de células en fase G0/G1 se obtuvo en promedio 1μg de ARN total, consistente con la mayor cantidad de células obtenidas (Tabla 3. 2).

La cantidad de ARN requerido por los servicios de secuenciación para la construcción de librerías es entre 1 μg y 10 μg. Dada la baja cantidad de ARN total obtenido, se decidió purificar el ARN poliadenilado y construir las librerías en nuestro laboratorio utilizando un kit que permite utilizar tan solo 100 ng de ARN total como mínimo (*NEBNext Ultra Directional library kit*). De acuerdo a los rendimientos de ARN obtenidos para las células en fase S/G2/M se utilizaron 60 ng de ARN total de cada muestra para la construcción de las librerías, sin realizar un control previo de la calidad del ARN.

Tabla 3. 2: Cantidad de células y ARN obtenido en experimentos de purificación de células

Exp ^[a]	Muestra ^[b]	G0/G1			S/G2/M		
		Nº de células (x10 ⁶)	Total células (x10 ⁶)	ng ARN	Nº de células (x10 ⁶)	Total células (x10 ⁶)	ng ARN
1	0	1,03	3,51	279	0,17	0,56	44,8
2		1,36			0,23		
3		1,12			0,16		
4	1	1,68	5,7	705,6	0,18	0,66	128,8
5		1,99			0,29		
6		2,03			0,19		
7	2	1,19	4,43	988	0,21	0,66	180
8		1,81			0,2		
9		1,43			0,24		
10	3	1,46	5,49	1279,2	0,22	0,64	71,2
11		2,4			0,23		
12		1,63			0,2		

[a]: Número de purificación de células. [b]: Número de muestra utilizada en la secuenciación

Las librerías fueron enviadas al servicio de secuenciación para el análisis de su calidad por *Bioanalyzer* y posterior secuenciación. El análisis de la calidad mostró la presencia de ADN del tamaño correspondiente a los fragmentos purificados para la misma (Figura 3. 1, línea roja), así como también un pico de bajo peso molecular correspondiente a adaptadores sin ligar (Figura 3. 1, flecha). A pesar de tener adaptadores libres, y debido a la baja concentración de las librerías, se procedió a secuenciar las librerías sin la eliminación de estos adaptadores.

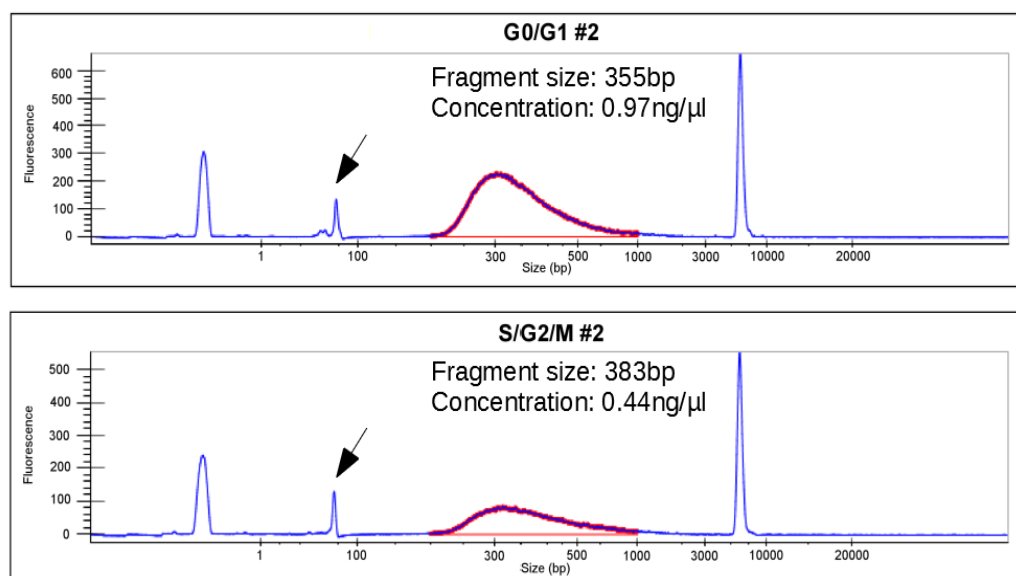


Figura 3. 1: Análisis de calidad de las librerías construidas a partir de ARN de células purificadas. Se muestra el resultado de una librería para cada población celular. Las demás muestras presentan resultados similares. Flecha: Pico correspondiente a adaptadores

Se obtuvieron entre 39 y 85 millones de lecturas crudas para las librerías (Tabla 3. 3), de las que se descartaron entre un 16 y un 20 % por tener baja calidad o corresponder a adaptadores. Un 13 a 18% adicional fue descartado por mapear a secuencias ribosomales. Se utilizó entonces para el análisis entre 28 y 57 millones de lecturas dependiendo de la muestra (Tabla 3. 3).

Tabla 3. 3: Calidad de las lecturas obtenidas

Muestra ^[a]	Original	Recorte			Mapean a ribosomales	
		Pareados	Desapareados	Descartados	Pareados	Desapareados
D1	85.423.212	68.104.862	8.499.629	8.818.721	10.794.326	1.701.922
D2	69.314.016	58.090.134	5.505.568	5.715.314	8.818.038	956.930
D3	68.475.398	54.955.204	6.675.963	6.844.231	11.065.424	157.692
P1	65.788.918	54.968.196	5.283.915	5.536.807	7.873.366	1.018.159
P2	54.387.642	44.963.832	4.636.925	4.786.885	7.568.670	936.268
P3	39.619.702	33.149.320	3.164.899	3.305.483	4.657.168	584.688

[a]: D#: Muestras correspondientes a células diferenciadas en fase G0/G1, P#: Muestras correspondientes a células proliferantes en fase S/G2/M

Estos resultados indican que la estrategia elegida para la generación de bibliotecas a partir de cantidades muy pequeñas de ARN fue exitosa.

Llamaremos células proliferantes (P) a las células en las fases S/G2/M y células diferenciadas (D) a las células en fases G0/G1. En esta población se encuentran también células proliferantes que en el momento de la purificación transitaban la fase G1 del ciclo, pero se espera que estas constituyan una minoría dentro de la población total.

2.2. Reproducibilidad de las réplicas y comparación entre condiciones

Se realizó un análisis de componentes principales (PCA: *Principal Component Analysis*) para analizar el agrupamiento de las muestras de células purificadas en distintos estados de proliferación y las muestras de organismos irradiados y control (descriptas en los puntos 2.1

y 2.2 del Capítulo 1).

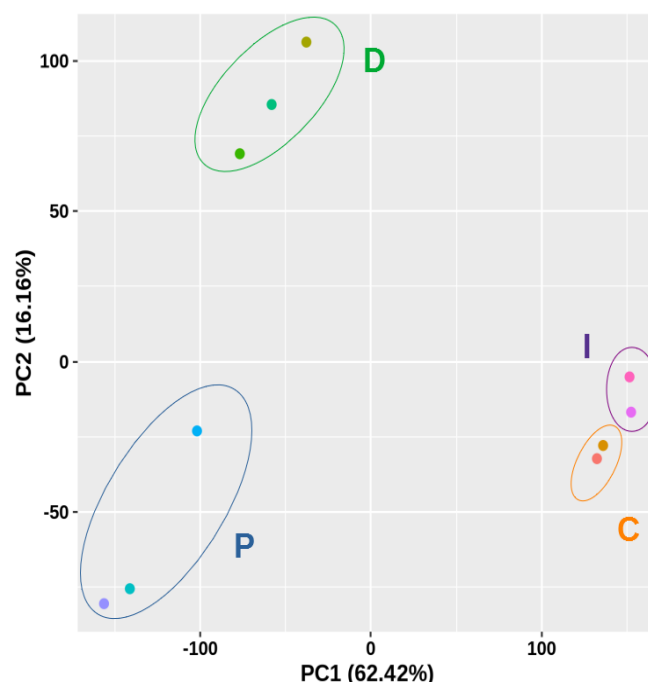


Figura 3. 2: Análisis de componentes principales de las muestras secuenciadas. D: Células diferenciadas (en fases G0/G1), P: Células proliferantes (en fases S/G2/M), I: Organismos irradiados, C: Organismos control.

El resultado del análisis (Figura 3. 2) muestra que las poblaciones de células en distintos estados de proliferación (G0/G1 y S/G2/M) se diferencian claramente, siendo el segundo componente principal el que mejor las separa. En el caso de los organismos irradiados y control, presentan una diferencia mucho menor en el componente 2, quedando las muestras de organismos irradiados hacia arriba (al igual que las muestras de G0/G1) y las de organismos control hacia abajo (al igual que las muestras S/G2/M). En este sentido el componente 2 indicaría la presencia o ausencia de células proliferativas. Cabe destacar que las diferencias entre organismos irradiados y control son menores que la variación que existe dentro de cada grupo de células purificadas. El componente principal 1, que presenta el 62% de la varianza separa claramente las muestras correspondientes a células purificadas de los organismos enteros.

Para analizar las similitudes entre las réplicas y cuál es su comportamiento con respecto a las demás muestras se realizó un *heatmap* de las distancias entre ellas (Figura 3. 3).

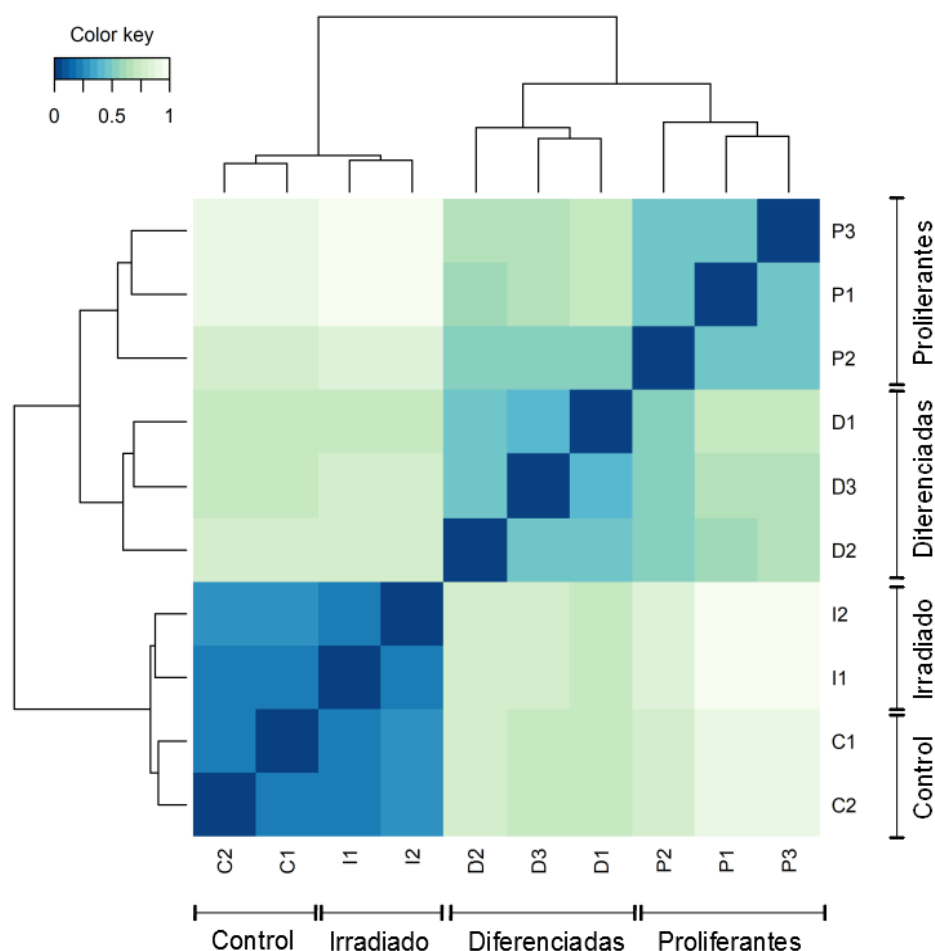


Figura 3. 3: Heatmap de las distancias entre las muestras secuenciadas. Se muestra en la parte superior la escala de colores (colores azules muestras más cercanas y colores verde claro muestras más alejadas). D: Células diferenciadas (en fases G0/G1), P: Células proliferantes (en fases S/G2/M), I: Organismos irradiados, C: Organismos control.

En este análisis vuelve a observarse la similitud existente entre los organismos irradiados y control. Si bien llegan a agruparse de acuerdo al tratamiento, las diferencias entre estas muestras son mucho menores a las observadas entre réplicas de células purificadas. Incluso, la réplica número 1 de los organismos irradiados tiene casi la misma distancia con los organismos control que con la otra réplica de irradiados.

Según los resultados de análisis del ciclo celular en organismos control (ver Capítulo 2), aproximadamente un 15% de las células se encuentran en proliferación (S/G2/M), con un 9% específicamente en fase S. Debido a la baja proporción de células que serían eliminadas por radiación, es de esperar que los perfiles de expresión sean más similares entre organismos irradiados y control que entre células purificadas en distintos estadios del ciclo celular.

Con respecto a las células purificadas, las distancias son consistentes con lo esperado, teniendo las células en fase G0/G1 menor distancia entre sí que con las células en fase S/G2M y viceversa. Este resultado además confirma que la técnica de purificación de células optimizada anteriormente (Domínguez et al., 2014) efectivamente permite separar las fracciones celulares de interés.

2.3. Análisis de expresión diferencial

2.3.1. Expresión diferencial entre organismos irradiados y control

El análisis comparativo de los transcritos en tetratiridios irradiados y control muestra un grupo reducido de genes expresados diferencialmente. Hay 7 genes que aumentan su expresión en organismos irradiados (Tabla 3. 4). Entre ellos se destacan un gen perteneciente a la superfamilia SCP/TAPS (MCU_010595, *Peptidase Inhibitor 16*) y un gen únicamente anotado en *M. corti* (MCU_009707), que además es el que presenta mayor nivel de expresión. Además 3 genes (sombreados en gris) no presentan evidencia de anotación funcional, pero se encuentran ortólogos en otros platelmintos.

Tabla 3. 4: Genes que presentan mayor expresión en organismos irradiados

Gen	TPM C	TPM I	FC	FDR	Anotación	Brite
MCU_001290	0,10	3,47	-33,51	0,00	Ortólogos en Platelminos excepto en Trematodos	
MCU_001683	0,18	1,67	-9,36	0,00	Ortólogos en Cestodos excepto en <i>Echinococcus</i>	
MCU_014823	2,68	2,62	-2,05	0,02	<i>flot2a: Flotillin-2a (Danio rerio)</i>	G/S
MCU_010595	17,73	35,35	-2,00	0,05	<i>PI16: Peptidase inhibitor 16 (Homo sapiens)</i>	
MCU_001950	14,43	28,45	-1,98	0,00	Ortólogos en Cestodos	
MCU_002677	591,53	1.158,76	-1,97	0,03	<i>TGFB1I1: Transforming growth factor beta-1-induced transcript 1 protein (Bos taurus)</i>	
MCU_009707	3.393,31	6.520,01	-1,93	0,03	Sólo en <i>M. corti</i>	

TMP: *transcripts per million mapped reads*, FC: *Fold change*: Con células proliferantes(C) / Sin células proliferantes (I). FDR: *False discovery rate*. C: Organismos control; I: Organismos irradiados. Categorías de Brite: G: Procesamiento de la información genética; S: Señalización y procesos celulares

Por otro lado, se encontraron 35 genes que disminuyen su expresión por la irradiación (Tabla 3. 5), los cuales podrían ser genes originalmente expresados por las células proliferantes.

Un análisis de las vías en las que estos genes están implicados (columna Brite en Tabla 3. 5) revela que en su mayoría de los genes que presentan menor nivel de expresión en organismos irradiados están relacionados con procesamiento de la información genética (replicación, ciclo celular o factores de transcripción). Estos procesos son los esperados para genes expresados en células en proliferación y que se ha visto que están enriquecidos en neoblastos de planaria (Eisenhoffer et al., 2008; Galloni, 2012; Labbé et al., 2012). Se destaca también que varios de los genes estudiados en células tipo neoblasto de *S. mansoni* (ver Tabla 3. 1 en la Introducción del Capítulo 3) presentan un nivel de expresión menor en organismos irradiados (*p53*, *sox*, *brca1*, *bard1* y un ortólogo de *zfp-1*; marcados en negrita en la Tabla 3. 5).

Tabla 3. 5: Genes que presentan menor expresión en organismos irradiados

Gen	TPM C	TPM I	FC	FDR	Anotación	Brite
MCU_004025	87,96	50,38	1,73	0,03	KPNA2: Importin subunit alpha-1 (<i>Homo sapiens</i>)	G
MCU_001682	34,14	18,61	1,82	0,01	ARHGAP19: Rho GTPase-activating protein 19 (<i>Gallus gallus</i>)	
MCU_009305	1.242,61	660,17	1,86	0,02	tuba: Tubulin alpha chain (<i>Xenopus laevis</i>)	G/S
MCU_006062	7,31	3,85	1,88	0,03	mcm4-b: DNA replication licensing factor mcm4-B (<i>Xenopus laevis</i>)	G
MCU_010994	45,53	23,22	1,94	0,02	Sólo en <i>M. corti</i>	
MCU_002881	29,89	15,01	1,97	0,02	CDC7: Cell division control protein 7 (<i>Saccharomyces cerevisiae</i>)	G
MCU_008029	37,77	18,90	1,98	0,00	IPR017877:Myb-like domain	
MCU_004024	14,41	7,21	1,98	0,00	SLC6A5S; solute carrier family 6 (neurotransmitter transporter, amino acid) member 5/7/9/14	S
MCU_007061	28,88	14,30	2,00	0,00	Ppt1: Palmitoyl-protein thioesterase 1 (<i>Mus musculus</i>)	M
MCU_008396	7,87	3,83	2,04	0,01	tp53: Cellular tumor antigen p53 (<i>Xiphophorus maculatus</i>)	
MCU_008486	6,16	2,94	2,07	0,01	Rif1: Telomere-associated protein RIF1 (<i>Mus musculus</i>)	G
MCU_006812	13,89	6,62	2,08	0,00	mcm5-a: DNA replication licensing factor mcm5-A (<i>Xenopus laevis</i>)	G
MCU_006326	8,69	3,82	2,25	0,00	zmcm3: Zygotic DNA replication licensing factor mcm3 (<i>Xenopus tropicalis</i>)	G
MCU_012956	22,29	9,75	2,27	0,00	Ortólogos en Platelminios endoparásitos	
MCU_004703	4,42	1,94	2,27	0,05	SOX8: Transcription factor SOX-8 (<i>Gallus gallus</i>)	G
MCU_003804	1.359,40	586,62	2,30	0,00	Sólo en <i>M. corti</i>	
MCU_010783	3,85	1,66	2,30	0,00	ASPM, ASP; abnormal spindle-like microcephaly-associated protein	G
MCU_005027	16,21	6,98	2,30	0,04	IPR001478:PDZ domain	
MCU_012015	7,94	3,34	2,36	0,02	Ortólogos en Platelminios	
MCU_009565	451,41	187,45	2,39	0,00	Histone H1-beta late embryonic (<i>Strongylocentrotus purpuratus</i>)	G
MCU_005442	1,93	0,79	2,41	0,03	Sólo en <i>M. corti</i>	
MCU_000354	9,48	3,80	2,47	0,00	nfil3: Nuclear factor interleukin-3-regulated protein (<i>Danio rerio</i>)	
MCU_011829	6,21	2,43	2,53	0,03	Bard1: BRCA1-associated RING domain protein 1 (<i>Mus musculus</i>)	G
MCU_009941	9,06	3,46	2,60	0,00	CIT: Citron Rho-interacting kinase (<i>Homo sapiens</i>)	M
MCU_005342	6,76	2,56	2,60	0,00	scra: Anillin (<i>Drosophila melanogaster</i>)	S
MCU_001783	4,57	1,70	2,67	0,03	IPR008967:p53-like transcription factor, DNA-binding@IPR012346:p53/RUNT-type transcription factor	
MCU_008160	8,23	2,96	2,76	0,00	IPR007588:Zinc finger, FLYWCH-type	
MCU_006277	1,73	0,60	2,88	0,02	BRCA1: Breast cancer type 1 susceptibility protein homolog (<i>Bos taurus</i>)	G
MCU_003048	8,18	2,71	2,97	0,05	Ortólogos en Platelminios parásitos	
MCU_012071	3,98	1,23	3,20	0,03	unc-9: Innexin unc-9 (<i>Caenorhabditis elegans</i>)	S
MCU_010057	9,11	2,50	3,63	0,00	Ortólogos en Platelminios parásitos	
MCU_009865	0,93	0,24	3,89	0,03	Awh: LIM/homeobox protein Awh (<i>Drosophila melanogaster</i>)	
MCU_005537	3,17	0,45	6,96	0,00	ZNF2: Zinc finger protein 2 (<i>Pongo abelii</i>)	
MCU_005864	2,64	0,29	9,03	0,04	CCDC130: Coiled-coil domain-containing protein 130 (<i>Bos taurus</i>), IPR007590:CWC16 protein	G
MCU_005413	1,31	0,06	20,61	0,05	fezf2: Fez family zinc finger protein 2 (<i>Danio rerio</i>)	

TPM: transcripts per million mapped reads, FC: Fold change: Con células proliferantes (C) / Sin células proliferantes (I). FDR: False discovery rate. C: Organismos control; I: Organismos irradiados. Categorías de Brite: G: Procesamiento de la información genética; S: Señalización y procesos celulares, M: Metabolismo, G/S: G y S

2.3.2. Expresión diferencial entre células proliferantes y células diferenciadas

Al analizar las células purificadas se obtienen 2.206 genes con una diferencia significativa en el nivel de expresión en células proliferantes respecto a células diferenciadas ($FDR < 0,05$). De ellos, 1.640 tienen un nivel de cambio mayor a dos veces ($Fold\ Change > 2$ o $Fold\ Change < -2$), con 535 genes que aumentan su expresión en células proliferantes y 1.105 que disminuyen. En la Figura 3. 4 se presenta un *volcano plot* mostrando esta variación.

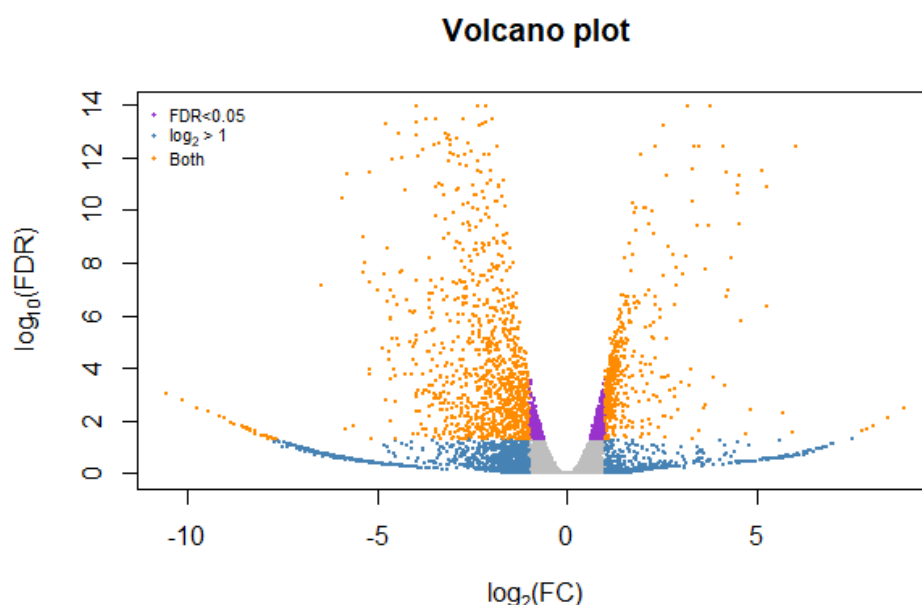


Figura 3. 4: Volcano plot comparando células en fase S/G2/M vs G0/G1. Los puntos violetas corresponden a genes con una diferencia significativa, aunque baja (cambio menor a 2 veces). Los puntos azules corresponden a genes que presentan una diferencia mayor a 2 veces en su nivel de expresión pero esta diferencia no es significativa. Los puntos anaranjados son los que cambian su nivel de expresión más de 2 veces y son significativos.

De los 2.206 genes expresados diferencialmente, existen 50 genes que presentan un nivel de expresión 16 veces mayor ($\log_2(FC) > 4$) en células proliferantes, mientras que 101 presentan la misma variación en células no proliferantes (Figura 3. 5). Utilizando un criterio menos estricto, 159 genes presentan un nivel de expresión 4 veces mayor ($\log_2(FC) > 2$) en células proliferantes mientras que 596 presentan un nivel 4 veces mayor en células diferenciadas (Figura 3. 5).

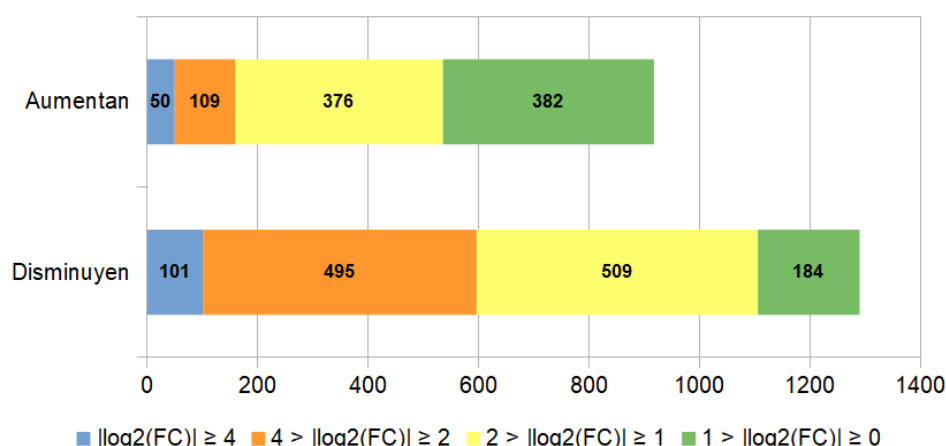


Figura 3. 5: Discriminación de niveles de cambio entre los genes que presentan expresión diferencial significativa. Aumentan: expresión superior en células proliferantes, Disminuyen: expresión superior en células no proliferantes.

2.3.2.1. Enriquecimiento de vías y términos de ontología génica

Se analizó el enriquecimiento de vías metabólicas entre los genes diferencialmente expresados utilizando la herramienta online KOBAS (Xie et al., 2011). Dentro de las opciones del análisis, esta herramienta permite utilizar a *S. mansoni* como organismo de referencia, pero únicamente se pueden analizar las vías KEGG. Al seleccionar al humano como organismo de referencia es posible analizar una mayor cantidad de vías (*Panther*, GO, KEGG y *Reactome*). Los resultados sobre las vías KEGG en ambos organismos son coincidentes (no se muestra), por lo que se seleccionó como referencia al humano.

Se obtuvieron 538 vías de la base de datos *Reactome* enriquecidas en células diferenciadas (valor p corregido <0.01) y 470 vías enriquecidas en células proliferantes. Las 10 vías más enriquecidas para cada población celular se muestran en la Tabla 3.6. Las vías predominantes en células proliferantes están asociadas a ciclo celular, metabolismo de ribosomas, replicación y reparación del ADN. El enriquecimiento en genes asociados a ribosomas explica la basofilia observada en este tipo celular. Por el contrario, consistente con la hipótesis de que las células diferenciadas no se dividen y permanecen en las fases G0/G1 del ciclo celular, los genes que aumentan su expresión en esta población están asociados a vías de células diferenciadas, como ser células nerviosas y células musculares.

Tabla 3. 6: Diez vías más significativas asociadas a genes que aumentan su expresión en células proliferantes y células diferenciadas. (KOBAS 3.0; base de datos Reactome)

<i>Path</i>	<i>ID</i>	<i>Input number</i>	<i>Background number</i>	<i>P-Value</i>	<i>Corrected P-Value</i>	
<i>Gene Expression</i>	R-HSA-74160	445	1.719	7,86E-252	1,25E-249	Proliferantes
<i>Cell Cycle</i>	R-HSA-1640170	224	607	1,18E-146	1,03E-144	
<i>Metabolism of proteins</i>	R-HSA-392499	284	1.378	2,00E-131	1,54E-129	
<i>Cell Cycle, Mitotic</i>	R-HSA-69278	184	503	4,35E-119	2,84E-117	
<i>rRNA processing in the nucleus and cytosol</i>	R-HSA-8868773	136	184	1,43E-118	9,23E-117	
<i>rRNA processing</i>	R-HSA-72312	137	194	1,62E-117	1,03E-115	
<i>Major pathway of rRNA processing in the nucleolus and cytosol</i>	R-HSA-6791226	130	173	5,96E-114	3,73E-112	
<i>L13a-mediated translational silencing of Ceruloplasmin expression</i>	R-HSA-156827	110	112	4,08E-105	2,31E-103	
<i>Translation</i>	R-HSA-72766	119	157	1,46E-104	8,18E-103	
<i>Cap-dependent Translation Initiation</i>	R-HSA-72737	110	120	6,65E-103	3,66E-101	
<i>Neuronal System</i>	R-HSA-112316	177	339	8,98E-108	8,72E-106	Diferenciadas
<i>Transmission across Chemical Synapses</i>	R-HSA-112315	120	208	1,03E-76	5,18E-75	
<i>Transmembrane transport of small molecules</i>	R-HSA-382551	172	657	8,84E-67	3,53E-65	
<i>Ion channel transport</i>	R-HSA-983712	98	211	4,24E-56	1,28E-54	
<i>Signal Transduction</i>	R-HSA-162582	294	2.448	3,48E-47	8,40E-46	
<i>Muscle contraction</i>	R-HSA-397014	86	204	1,05E-46	2,53E-45	
<i>Stimuli-sensing channels</i>	R-HSA-2672351	66	108	8,57E-44	1,95E-42	
<i>Axon guidance</i>	R-HSA-422475	126	549	9,76E-44	2,19E-42	
<i>Disease</i>	R-HSA-1643685	75	902	2,66E-06	9,22E-06	
<i>Developmental Biology</i>	R-HSA-1266738	154	841	9,75E-43	2,12E-41	

Tabla 3. 7: Enriquecimiento de términos de ontología génica en células proliferantes

	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Biological Process	GO:0006412	Translation	155	100	20	<1e-30
	GO:0015986	ATP synthesis coupled proton transport	12	10	2	0
	GO:0006457	protein folding	43	19	6	0
	GO:0006270	DNA replication initiation	7	6	1	0
	GO:0042254	ribosome biogenesis	29	20	4	0
	GO:0006414	translational elongation	10	8	1	0
	GO:0006364	rRNA processing	16	10	2	0
	GO:0006413	translational initiation	13	7	2	0
	GO:0006269	DNA replication, synthesis of RNA primer	3	3	0	0
	GO:0006260	DNA replication	41	22	5	0
Molecular Function	GO:0003735	structural constituent of ribosome	96	71	11	< 1e-30
	GO:0003743	translation initiation factor activity	19	14	2	0
	GO:0005524	ATP binding	585	106	68	0
	GO:0003723	RNA binding	154	60	18	0
	GO:0003678	DNA helicase activity	19	10	2	0
	GO:0046933	proton-transporting ATP synthase activit...	6	6	1	0
	GO:0051082	unfolded protein binding	24	12	3	0
	GO:0003677	DNA binding	367	56	43	0
	GO:0003676	nucleic acid binding	828	168	97	0
	GO:0000166	nucleotide binding	947	157	110	0
Cellular Compartment	GO:0005840	Ribosome	98	71	13	< 1e-30
	GO:0005852	eukaryotic translation initiation factor...	9	9	1	0
	GO:0015934	large ribosomal subunit	6	6	1	0
	GO:0042555	MCM complex	6	6	1	0
	GO:0005737	Cytoplasm	382	125	50	0
	GO:0015935	small ribosomal subunit	10	6	1	0
	GO:0045261	proton-transporting ATP synthase complex	4	4	1	0
	GO:0005622	Intracellular	956	210	124	0
	GO:0005741	mitochondrial outer membrane	7	5	1	0
	GO:0005730	Nucleolus	4	4	1	0

Por el contrario, en las células diferenciadas (Tabla 3. 8) se encuentran enriquecidos términos asociados al transporte de distintos elementos, lo cual es consistente con funciones asociadas a células diferenciadas

Tabla 3. 8: Enriquecimiento de términos de ontología génica en células diferenciadas

	GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
Biological Process	GO:0007156	homophilic cell adhesion via plasma memb..	49	32	7	0
	GO:0007186	G-protein coupled receptor signaling pat...	81	44	12	0
	GO:0006811	ion transport	206	72	30	0
	GO:0006813	potassium ion transport	50	24	7	0
	GO:0070588	calcium ion transmembrane transport	15	10	2	0
	GO:0007165	signal transduction	342	106	49	0
	GO:0007154	cell communication	355	115	51	0
	GO:0006814	sodium ion transport	29	10	4	0
	GO:0055085	transmembrane transport	223	52	32	0
Molecular Function	GO:0007155	cell adhesion	64	38	9	0
	GO:0005509	calcium ion binding	187	64	27	0
	GO:0004930	G-protein coupled receptor activity	65	35	9	0
	GO:0005515	protein binding	1.598	279	233	0
	GO:0005230	extracellular ligand-gated ion channel a...	34	18	5	0
	GO:0005249	voltage-gated potassium channel activity	25	12	4	0
	GO:0005216	ion channel activity	113	60	16	0
	GO:0005234	extracellularly glutamate-gated ion chan...	8	6	1	0
	GO:0004970	ionotropic glutamate receptor activity	13	7	2	0
	GO:0005262	calcium channel activity	12	9	2	0
Cellular Compartment	GO:0004697	protein kinase C activity	3	3	0	0
	GO:0016020	Membrane	911	231	155	0
	GO:0005886	plasma membrane	101	55	17	0
	GO:0016021	integral component of membrane	448	109	76	0
	GO:0005856	Cytoskeleton	160	27	27	0
	GO:0005615	extracellular space	4	4	1	0
	GO:0005581	collagen trimer	3	3	1	0
	GO:0005891	voltage-gated calcium channel complex	6	4	1	0
	GO:0005921	gap junction	14	6	2	0
	GO:0008076	voltage-gated potassium channel complex	18	7	3	0
	GO:0019898	extrinsic component of membrane	10	6	2	0

2.3.3. Comparación de genes expresados diferencialmente en organismos irradiados y células proliferantes purificadas

Se compararon los resultados de expresión diferencial entre los dos análisis en busca de genes que presenten el mismo comportamiento. En total, 25 de los 42 genes que presentan expresión diferencial en organismos irradiados respecto al control también se expresan diferencialmente en células purificadas (Tabla 3. 9). De ellos, 3 tienen el comportamiento opuesto al esperado y los 22 restantes se comportan de manera concordante. Entre estos, solamente un gen disminuye en células en proliferación (*Transforming growth factor beta-1-induced transcript 1 protein*), mientras que los demás aumentan. Entre los que aumentan, existen 2 que poseen ortólogos en platelmintos pero sin anotación funcional y uno anotados sólo en *M. corti*. Los genes restantes en su mayoría están implicados en la replicación o asociados al cromosoma. Esto indica que la irradiación, aunque tiene un efecto menor, altera de alguna manera los procesos de proliferación.

Tabla 3. 9: Genes que disminuyen su expresión en organismos irradiados y aumentan en células proliferantes

Gen	TPM D	TPM P	FC	FDR	TPM C	TPM I	FC	FDR	Anotación	Brite
MCU_004025	48,66	471,82	12,22	0,00	87,96	50,38	1,73	0,03	<i>KPNA2: Importin subunit alpha-1 (Homo sapiens)</i>	G
MCU_001682	124,62	292,93	2,98	0,00	34,14	18,61	1,82	0,01	<i>ARHGAP19: Rho GTPase-activating protein 19 (Gallus gallus)</i>	
MCU_009305	632,05	4.103,94	8,24	0,00	1.242,61	660,17	1,86	0,02	<i>tuba: Tubulin alpha chain (Xenopus laevis)</i>	G/S
MCU_006062	12,23	150,03	15,53	0,00	7,31	3,85	1,88	0,03	<i>mcm4-b: DNA replication licensing factor mcm4-B (Xenopus laevis)</i>	G
MCU_002881	13,24	138,50	13,44	0,00	29,89	15,01	1,97	0,02	<i>CDC7: Cell division control protein 7 (Saccharomyces cerevisiae)</i>	G
MCU_004024	61,71	195,75	4,03	0,00	14,41	7,21	1,98	0,00	<i>SLC6A5S; solute carrier family 6 (neurotransmitter transporter, amino acid) member 5/7/9/14</i>	S
MCU_007061	247,19	443,86	2,26	0,00	28,88	14,30	2,00	0,00	<i>Ppt1: Palmitoyl-protein thioesterase 1 (Mus musculus)</i>	M
MCU_008486	7,85	41,11	6,62	0,00	6,16	2,94	2,07	0,01	<i>Rif1: Telomere-associated protein RIF1 (Mus musculus)</i>	G
MCU_006812	7,14	141,28	25,24	0,00	13,89	6,62	2,08	0,00	<i>mcm5-a: DNA replication licensing factor mcm5-A (Xenopus laevis)</i>	G
MCU_006326	9,36	140,83	18,95	0,00	8,69	3,82	2,25	0,00	<i>zmcm3: Zygotic DNA replication licensing factor mcm3 (Xenopus tropicalis)</i>	G
MCU_012956	19,89	121,16	7,73	0,00	22,29	9,75	2,27	0,00	Ortólogos en Platelminos endoparásitos	
MCU_004703	48,82	97,47	2,54	0,00	4,42	1,94	2,27	0,05	SOX8: Transcription factor SOX-8 (Gallus gallus)	G
MCU_010783	3,77	33,70	11,42	0,00	3,85	1,66	2,30	0,00	<i>ASPM, ASP; abnormal spindle-like microcephaly-associated protein</i>	G
MCU_005027	33,53	89,25	3,38	0,00	16,21	6,98	2,30	0,04	<i>IPR001478:PDZ domain</i>	
MCU_012015	2,45	50,17	25,62	0,00	7,94	3,34	2,36	0,02	Ortólogos en Platelminos	
MCU_009565	52,24	281,98	6,98	0,00	451,41	187,45	2,39	0,00	Histone H1-beta late embryonic (Strongylocentrotus purpuratus)	G
MCU_005442	2,77	55,05	24,81	0,00	1,93	0,79	2,41	0,03	Sólo en <i>M. corti</i>	
MCU_011829	1,58	46,68	38,01	0,00	6,21	2,43	2,53	0,03	Bard1: BRCA1-associated RING domain protein 1 (Mus musculus)	G
MCU_009941	4,41	50,18	14,40	0,00	9,06	3,46	2,60	0,00	<i>CIT: Citron Rho-interacting kinase (Homo sapiens)</i>	M
MCU_005342	2,40	61,90	32,13	0,00	6,76	2,56	2,60	0,00	<i>scra: Anillin (Drosophila melanogaster)</i>	S
MCU_006277	0,45	14,53	39,25	0,00	1,73	0,60	2,88	0,02	BRCA1: Breast cancer type 1 susceptibility protein homolog (Bos taurus)	G

TPM: *transcripts per million mapped reads*, FC: *Fold change*: Con células proliferantes (C o P) / Sin células proliferantes (I o D). FDR: *False discovery rate*. C: Organismos control; I: Organismos irradiados; P: células proliferantes (fases S/G2/M); D: células diferenciadas (fases G0/G1). Categorías de Brite: G: Procesamiento de la información genética; S: Señalización y procesos celulares, M: Metabolismo, G/S: G y S

Existen 17 genes que se expresan diferencialmente en organismos irradiados respecto al control sin irradiar que no cambian su nivel de expresión en células purificadas, los cuales podrían ser genes vinculados a la respuesta a la radiación. En particular, hay 4 que aumentan en organismos irradiados, de los cuales uno no presenta ortólogos en otros

platelmintos y otro únicamente presenta ortólogos en otros cestodos. Es interesante profundizar en el estudio de estos genes para evaluar si cumplen un rol en la resistencia a la radiación que presentan están los cestodos y *M. corti* en particular.

2.4. Análisis de expresión de secuencias repetidas y elementos transponibles

Como se mencionó anteriormente, las células germinativas de *E. multilocularis* expresan altos niveles de un retrotransposón en miniatura específico de taenidos (taTRIM). Por este motivo se evaluó si en el caso de *M. corti* existe algún elemento transponible que esté siendo expresado y en particular si se expresa diferencialmente en células proliferantes.

Se extrajeron los elementos transponibles anotados por `maker` (primer paso del proceso de anotación descrito en el Capítulo 1) y se evaluaron sus niveles de expresión. Se obtuvieron 14 elementos con expresión diferencial (Tabla 3. 10), de las cuales 6 se expresan en mayor cantidad en células diferenciadas ($\log_2FC < 0$) mientras que los restantes lo hacen en células proliferantes.

Tabla 3. 10: Secuencias enmascaradas con expresión diferencial en células proliferativas y diferenciadas

Name	Family	Log2 (FC)	padj	
TOM_I-int (1)	LTR-Gypsy	-7,09	0,0038	Diferenciadas
Gypsy-22-I_SP(1)	LTR-Gypsy	-6,68	0,0188	
L2-16_NVe (1)	LINE-L2	-5,61	0,0015	
Copia3-I_CR (3)	LTR-Copia	-3,10	0,0387	
TpSIREcon (2)	LTR-Copia	-3,07	0,0175	
Copia-11_BG-I (2)	LTR-Copia	-2,62	0,0087	
EnSpm-7_CCri (4)	DNA-CMC-EnSpm	1,74	0,0054	Proliferantes
Penelope-1_HRo (1)	LINE-Penelope	2,31	0,0000	
rnd-5_family-743 (26)	Unknown	2,80	0,0000	
Jockey-5_CQ (2)	LINE-I-Jockey	3,14	0,0000	
Gypsy-3_MN-I (1)	LTR-Gypsy	3,61	0,0042	
LTR-30_OS-LTR (1)	LTR-Gypsy	4,60	0,0000	
EnSpm-1N1_DDi (1)	DNA-CMC-EnSpm	5,11	0,0000	
MuDR-15_OS (1)	DNA-MULE-MuDR	5,33	0,0015	

Se muestran sombreados los repetidos asociados a retrotransposones

Debido a que Cestodos y Trematodos carecen de muchos de los genes asociados al silenciamiento de elementos transponibles en la línea germinal, se evaluó además si los niveles de expresión de secuencias repetidas y elementos transponibles son diferentes entre células proliferantes y células diferenciadas. Para ello, se compararon los valores de cpm (*count per million*) de los que se expresan en al menos una condición (Figura 3. 7).

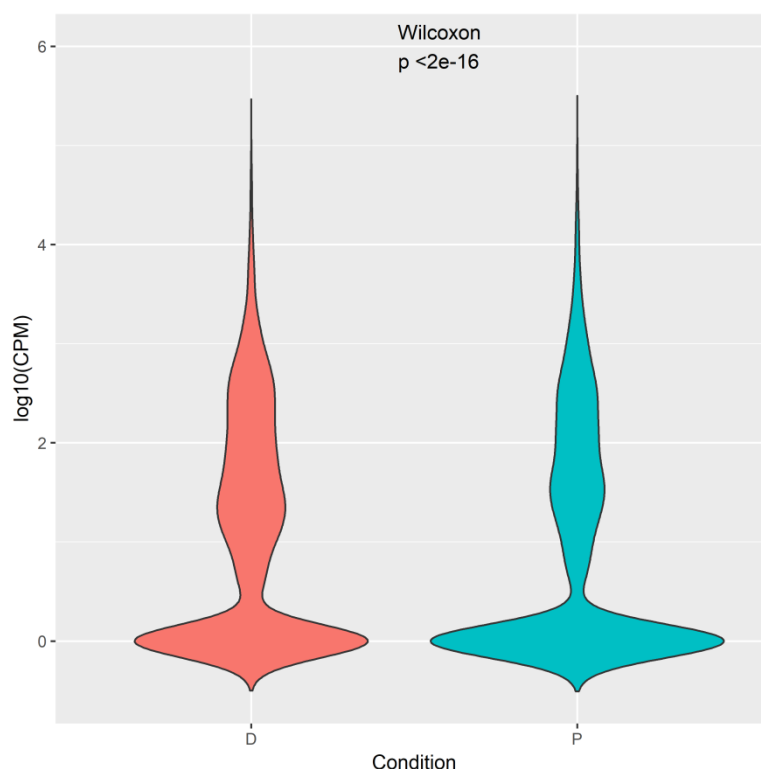


Figura 3. 7: Violin plots mostrando los niveles de expresión de repetidos en células diferenciadas (D) y células proliferantes (P).

De este análisis se concluye que a nivel general, las células en proliferación expresan menor cantidad de repetidos ya que hay una mayor cantidad de repetidos con valores de expresión bajos. La mediana de los valores de cpm para células diferenciadas es de 8.47, mientras que para las células proliferantes el valor es 0. Esta diferencia es significativa (test de Wilcoxon, valor $p < 2e-16$), lo que indicaría que si bien los platelmintos parásitos no tienen las vías clásicas de silenciamiento de los elementos transponibles (mediadas por piARNs), al menos en *M. corti* parece existir un mecanismo que impide su expresión.

2.5. Análisis comparativo entre genes marcadores de células proliferantes en platelmintos.

Una vez identificados los genes que presentan un mayor nivel de expresión en células proliferantes de *M. corti*, se compararon los mismos con genes descritos para células en proliferación de otros platelmintos.

En el caso de *S. mediterranea*, se utilizaron los genes que caracterizan a las subpoblaciones de neoblastos y de su progenie en estudios de transcriptómica de células aisladas (Fincher et al., 2018; Plass et al., 2018; Zeng et al., 2018). Los enfoques utilizados en estos trabajos y la cantidad de células caracterizadas son diferentes (Ver Introducción General, punto 4.1.3). A esto, se suma el hecho de que en cada trabajo se utilizó un transcriptoma de referencia diferente. Por este motivo, para las comparaciones entre los trabajos se comenzó buscando homología entre los distintos transcriptomas mediante `blastn`, y a partir de esta homología se identificaron las coincidencias entre los estudios (Figura 3. 8A). Solo 7 genes son identificados como característicos de neoblastos en los 3

estudios publicados. El trabajo de Zheng *et al.* (2018), al estar enfocado en la población de neoblastos, identifica un número de genes reducido, por lo que es esperable que la coincidencia con los otros dos trabajos sea baja. Al comparar los trabajos de Fincher *et al.* (2018) y Plass *et al.* (2018), que utilizan versiones diferentes del mismo transcriptoma, la intersección de genes característicos de neoblastos es mayor (248 genes).

En el caso de *S. mansoni* se utilizaron como genes específicos de células tipo neoblasto los identificados en el adulto por Collins *et al.* (2013) mediante irradiación, los que presentan mayor expresión en esporocistos respecto a miracidios reportados por Wang *et al.* (2013) y los genes que caracterizan a las 3 subpoblaciones de células tipo neoblasto en esporocistos determinados por transcriptómica de células aisladas (Wang *et al.*, 2018, por más detalles de los trabajos ver la Introducción general). Nuevamente, la coincidencia entre los tres trabajos es baja (2 genes coinciden entre los 3 estudios, Figura 3. 8B) pero hay que tener en cuenta que se siguieron diferentes estrategias utilizando distintos estadios del desarrollo y caracterizando números variables de genes.

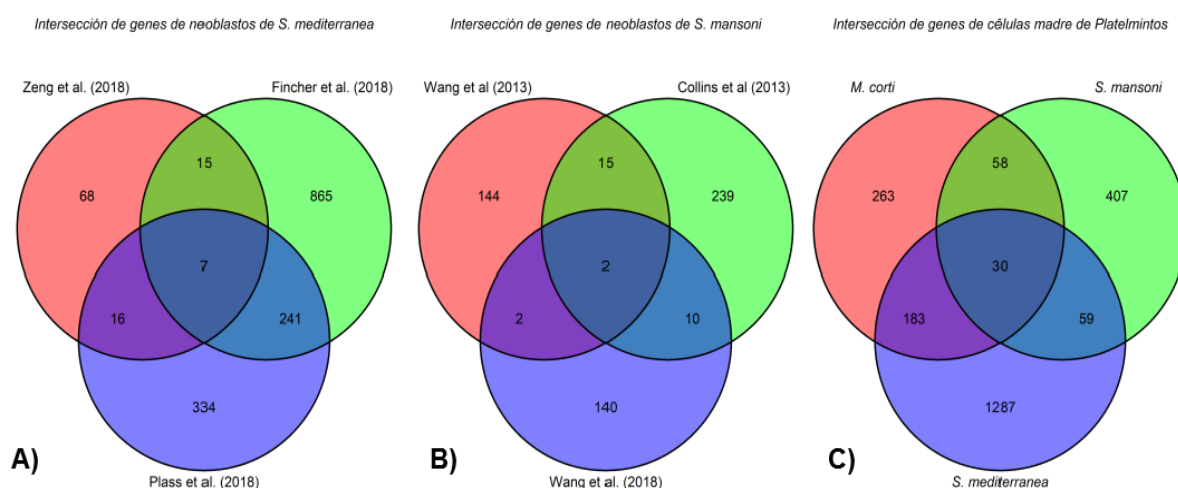


Figura 3. 8: Comparación de trabajos en que se reportan niveles de expresión en células en proliferación en platelmintos. A) Comparación de trabajos en *S. mediterranea* por *scRNAseq*. B) Comparación de trabajos en *S. mansoni* en estadio adulto y esporocisto. C) Comparación entre ambos platelmintos y *M. corti*

Para analizar la conservación de genes marcadores de células proliferantes entre las distintas clases de platelmintos un gen se consideró conservado si aparece en al menos uno de los reportes para *S. mansoni* o al menos en uno de *S. mediterranea* y con un aumento significativo en el nivel de expresión en células proliferantes de al menos 2 veces en *M. corti* (Figura 3. 8C). Existen 30 genes que son característicos de células en proliferación en los tres organismos; *M. corti* comparte con *S. mansoni* 88 genes marcadores y con *S. mediterranea* 213 genes. Todos los organismos presentan marcadores que son diferencialmente expresados únicamente en esa especie. En particular, *M. corti* posee 263 genes expresados diferencialmente en células proliferantes que no son marcadores de este tipo celular en los otros platelmintos con información transcriptómica disponible. A medida que se realicen nuevos estudios sobre este tipo celular en otros platelmintos el conjunto de genes conservados marcadores de células proliferativas de una clase de platelmintos podrá definirse con mayor precisión.

Tabla 3. 11: Anotación de genes de *M. corti* marcadores de células proliferantes conservados

Gene	Blast Uniprot	P/D		C/I	
		FC	FDR	FC	FDR
MCU_000011	<i>Prim2: DNA primase large subunit (Rattus norvegicus)</i>	89.68	0	1.87	0.9998
MCU_006812	<i>mcm5-a: DNA replication licensing factor mcm5-A (Xenopus laevis)</i>	25.24	0	2.08	0.0043
MCU_006326	<i>zmcm3: Zygotic DNA replication licensing factor mcm3 (Xenopus tropicalis)</i>	18.95	0	2.25	0.0022
MCU_013245	<i>cenpa: Histone H3-like centromeric protein A (Danio rerio)</i>	16.23	0	1.98	0.0758
MCU_006062	<i>mcm4-b: DNA replication licensing factor mcm4-B (Xenopus laevis)</i>	15.53	0	1.88	0.0269
MCU_011475	<i>Plk1: Serine/threonine-protein kinase PLK1 (Mus musculus)</i>	12.09	0	1.53	0.3602
MCU_008499	<i>mcm7-a: DNA replication licensing factor mcm7-A (Xenopus laevis)</i>	11.88	0	1.66	0.1302
MCU_002449	<i>mars: Guanylate kinase-associated protein mars (Drosophila melanogaster)</i>	10.93	0	1.76	0.8109
MCU_008653	<i>zmcm6-b: Zygotic DNA replication licensing factor mcm6-B (Xenopus laevis)</i>	10.59	0	1.24	0.9998
MCU_000285	<i>CLSPN: Claspin (Homo sapiens)</i>	9.54	0	1.82	0.2736
MCU_002238	<i>Rrm1: Ribonucleoside-diphosphate reductase large subunit (Mus musculus)</i>	8.77	0	1.29	0.9998
MCU_011129	<i>Orthologous in Platyhelminthes</i>	6.39	0	1.58	0.6393
MCU_001692	<i>mel-32: Serine hydroxymethyltransferase (Caenorhabditis briggsae)</i>	5.09	0	1.55	0.9998
MCU_002646	<i>PCNA: Proliferating cell nuclear antigen (Drosophila melanogaster)</i>	5.01	0	1.08	0.9998
MCU_012332	<i>CCNB2: G2/mitotic-specific cyclin-B2 (Mesocricetus auratus)</i>	4.16	0	1.14	0.9998
MCU_005471	<i>cht18: Chromosome transmission fidelity protein 18 homolog (Xenopus laevis)</i>	3.73	0	1.04	0.9998
MCU_009290	<i>C1qbp: Complement component 1 Q subcomponent-binding protein%2C mitochondrial (Rattus norvegicus)</i>	3.51	0	1.29	0.9998
MCU_007922	<i>ESPL1: Separin (Homo sapiens)</i>	3.33	0	2.43	0.0874
MCU_003859	<i>POLR1C: DNA-directed RNA polymerases I and III subunit RPAC1 (Bos taurus)</i>	3.31	0	1.23	0.9998
MCU_002398	<i>TYMS: Thymidylate synthase (Homo sapiens)</i>	3.1	0	1.19	0.9998
MCU_004693	<i>ahcy-a: Adenosylhomocysteinase A (Xenopus laevis)</i>	2.94	0	-1.03	0.9998
MCU_012048	<i>MED36B: Probable mediator of RNA polymerase II transcription subunit 36b (Arabidopsis thaliana)</i>	2.89	0	1.22	0.9998
MCU_004294	<i>GFI1B: Zinc finger protein Gfi-1b (Gallus gallus)</i>	2.76	0.0001	1.91	0.2361
MCU_004703	<i>SOX8: Transcription factor SOX-8 (Gallus gallus)</i>	2.54	0.0002	2.27	0.0486
MCU_010780	<i>mcm2: DNA replication licensing factor mcm2 (Xenopus laevis)</i>	2.35	0.0001	1.43	0.8086
MCU_012902	<i>TUBA1C: Tubulin alpha-1C chain (Homo sapiens)</i>	2.34	0	-1.03	0.9998
MCU_010786	<i>CCT8: T-complex protein 1 subunit theta (Gallus gallus)</i>	2.27	0	1.04	0.9998
MCU_009121	<i>FABP1: Fatty acid-binding protein homolog 1 (Echinococcus granulosus)</i>	2.12	0.0034	-1.19	0.9998
MCU_000457	<i>POLA1: DNA polymerase alpha catalytic subunit (Homo sapiens)</i>	2.09	0.0036	1.37	0.9998
MCU_007048	<i>tomboy40: Mitochondrial import receptor subunit TOM40 homolog 2 (Drosophila melanogaster)</i>	2.02	0.0419	1.07	0.9998

Se muestran sombreados los que se expresan en células proliferantes y disminuyen en irradiados, P: proliferantes, D: Diferenciadas, C: Control, I: Irradiados

En la Tabla 3. 11 se muestra la anotación y los niveles de expresión en *M. corti* de los 30 genes que coinciden entre los 3 organismos. Entre ellos hay muchos genes relacionados con replicación y reparación de ADN. En particular *pcna* aparece como marcador compartido y presenta un aumento de 5 veces en su nivel de expresión en células proliferantes. Esto indica que tal como se planteó en el Capítulo 2, este gen es un buen

candidato a marcador de células proliferantes. Por otro lado, *nanos*, que disminuye su expresión en las larvas de *M. corti* irradiadas (Capítulo 2), no aparece entre los genes comunes, a pesar de que se encuentra validado como un gen marcador en diversas especies de platelmintos (ver Introducción General para más detalles). Buscando puntualmente dentro de los datos de expresión, los dos genes *nanos* anotados presentan valores de TPM mayores en células proliferantes, aunque esta diferencia no es significativa (*nanos1*: 3.52 vs 5.34, $p=0.8931$ y *nanos2*: 0 vs 1.65, $p=0.1441$). Además, de los tres genes con dominio PUM-HD anotados, uno de ellos presenta un aumento significativo en células proliferantes (45.47 vs 72.99, $p=0.002$), mientras que los otros dos (estudiados en el Capítulo 2), presentan mayor expresión en células no proliferantes, aunque las diferencias no son significativas.

3. Discusión

Los resultados obtenidos en la expresión diferencial de genes presentan la tendencia esperada, con genes asociados a replicación de ADN y control del ciclo celular enriquecidos en las células proliferantes. Además se observan muchos genes implicados en la traducción y la biogénesis de ribosomas, lo cual es consistente con la basofilia observada para este tipo celular de otros platelmintos.

En el caso de *M. corti* la estrategia de irradiación para el estudio de células proliferantes no parece ser tan efectiva como lo es en planaria o en *S. mansoni*, y es similar a lo que ocurre en *E. multilocularis*. De todas maneras se observan unos pocos genes que disminuyen su expresión tras la irradiación, y la mayoría de ellos coincide con los expresados en células en fase S/G2/M, indicando que existe un efecto a pesar de ser bajo. Como se mencionó en el Capítulo 2, es interesante evaluar dosis mayores y más tiempo de recuperación para estudiar el efecto de la radiación sobre este tipo celular. Esta técnica es mucho más sencilla que la purificación por citometría de flujo y, de encontrar condiciones adecuadas, permitiría evaluar la expresión de genes marcadores de células proliferantes en otros estadios del desarrollo, en particular durante el desarrollo estrobilar y la formación de estructuras reproductivas.

La poca cantidad de genes diferencialmente expresados en organismos irradiados puede deberse a varios motivos. En primer lugar, hay que tener en cuenta que la cantidad de células proliferantes es alrededor de un 15% del total de células del organismo, por lo que la diferencia entre un gusano irradiado y uno no irradiado no es muy grande. En segundo lugar, la dosis aplicada puede haber sido muy baja para detectar cambios significativos. Se eligió la dosis de 100 Gy (ver Capítulo 2) ya que es una dosis que permitía observar cambios en la proliferación de los organismos pero sin daños morfológicos importantes. Los datos obtenidos permiten corroborar esto, ya que si se hubiera causado mucho daño general, se esperaría observar un aumento de genes asociados a procesos de muerte celular o a algún tipo de estrés. Tampoco se obtuvieron resultados que mostraran expresión aumentada de genes asociados a reparación o daño en el ADN. Otro posible motivo podría ser que el tiempo de recuperación ensayado fuera muy corto, por lo que el organismo no tuvo tiempo suficiente para detectar el daño causado en el ADN por la radiación y activar los mecanismos necesarios para repararlo.

La purificación de células proliferantes por citometría de flujo es la técnica ideal para la caracterización de este tipo celular. Sin embargo, es una técnica delicada y que lleva una

gran cantidad de tiempo y esfuerzo para la obtención de cantidad de ARN suficiente para la secuenciación. A pesar de esto, se logró extraer ARN de células purificadas de calidad adecuada para construir librerías de secuenciación eficientemente. Estos resultados son los primeros obtenidos para células purificadas de Cestodos por lo que representan un insumo importante para la caracterización de este tipo celular. El análisis de las réplicas reveló que esta además es una técnica reproducible, ya que hay una buena correlación para las muestras de cada condición.

La comparación de genes característicos de este tipo celular en platelmintos reveló que la cantidad de genes compartidos entre las tres clases es relativamente baja, incluso para un mismo organismo en distintos estudios y estadíos. Esto coincide con otros estudios en donde se muestra que en los platelmintos estudiados si bien las células proliferantes tienen una morfología similar, a nivel molecular existen diferencias importantes.

Estos resultados son un punto de inicio para el estudio en profundidad de genes candidatos a marcadores de células proliferantes. En el caso de *M. corti* tenemos optimizadas técnicas para estudiar la expresión temporal y espacial de genes de interés, lo que ligado a la detección de marcadores de distintos tejidos y células en fase S del ciclo celular permitirá confirmar si estos genes se expresan efectivamente en células proliferantes y si lo hacen en todas o en subpoblaciones particulares. Por el momento no existen reportes exitosos de interferencia por ARN en *M. corti*, pero nuestro grupo ha estado trabajando en su optimización. Esta herramienta es esencial para evaluar el rol que tienen los genes candidatos en la biología de las células madre.

CONCLUSIONES Y PERSPECTIVAS

En este trabajo, mediante distintos abordajes, se estudiaron las células proliferantes de *M. corti* a nivel transcriptómico y se evaluó su sensibilidad a la radiación. Además se utilizaron los datos generados para mejorar la anotación disponible del genoma de *M. corti*.

El análisis primario de los datos de secuenciación de organismos irradiados y control mostró una cantidad de transcriptos no predichos en la versión de la anotación del genoma disponible al inicio de este trabajo. Por este motivo nos propusimos utilizar estos datos para mejorar la anotación existente. Esta re-anotación recupera la mayoría de los genes predichos anteriormente, además de numerosos genes nuevos. Al basarse en información transcriptómica es capaz de predecir la existencia de isoformas de *splicing* alternativo no predichas en la anotación de *WormBase ParaSite*, así como también regiones no traducidas para una gran proporción de los genes.

Un 35% de las predicciones confiables no tienen evidencia de anotación funcional. Entre estos genes desconocidos un 48% tienen ortólogos en otros platelmintos y más del 60% presentan evidencia de expresión en el estadio larvario. El hecho de que genes sin evidencia de anotación funcional se expresen a niveles tan altos destaca la importancia de mejorar las anotaciones de los genomas de platelmintos y de estudiar estos genes para tratar de determinar su función. Evaluar los tejidos y los estadios del ciclo de vida donde se expresan puede permitir inferir posibles roles para estos genes. En los organismos en que funcionan las herramientas de interferencia mediada por ARN (ARNi), como en cultivos celulares de células primarias de *E. multilocularis* (Spiliotis et al., 2010) y *H. microstoma* (Pouchkina-Stantcheva et al., 2013), es posible analizar el fenotipo generado tras la disminución o eliminación de su expresión. En el caso de *M. corti*, hemos trabajado en las condiciones para la entrada de distintas moléculas reporteras buscando optimizar condiciones de *delivery* de ARN interferente, pero aún no existen protocolos publicados que permitan su utilización para el silenciamiento de genes (Domínguez, 2016).

Las células proliferantes de *M. corti* si bien son sensibles a la radiación, de acuerdo a la hipótesis planteada, parecen ser más resistentes que las de trematodos, y en organismos neodermados parecen serlo aún más que las de planarias. Es posible que el tegumento, común a cestodos y trematodos, sea el responsable de esta resistencia.

Es interesante estudiar la causa de esta resistencia a nivel molecular, estudiando genes asociados al control del ciclo celular y a la reparación del ADN y evaluar el comportamiento de las células proliferantes con el tiempo de recuperación. Mediante estudios de genes candidatos y de doble marcado metabólico con EdU y BrdU es posible determinar si las células que proliferaban antes de la irradiación son capaces de seguir proliferando, si son eliminadas o si entran en estado quiescente, activando mecanismos de reparación de daño al ADN. Además, la detección por hibridación *in situ* de genes marcadores de tejidos diferenciados permitiría evaluar el efecto a largo plazo de la irradiación sobre estos tejidos.

El análisis de expresión diferencial, tanto de genes que disminuyen su expresión en organismos irradiados como de genes enriquecidos en células proliferantes purificadas, es consistente con lo esperado. Existen 917 genes que presentan mayor expresión en células proliferantes purificadas. La mayoría de estos genes están asociados a la replicación del ADN, estructura cromosómica y control del ciclo celular. Además, hay muchos genes

asociados a biogénesis de ribosomas lo que es consistente con la basofilia característica de este tipo celular.

En los platelmintos en que se ha estudiado este tipo celular se ha visto que constituyen una población homogénea a nivel morfológico, pero heterogénea a nivel molecular, existiendo subtipos de células proliferantes con distinto grado de compromiso con los tejidos diferenciados. Es interesante evaluar cuáles de los genes identificados en *M. corti* son expresados por toda la población de neoblastos y cuáles se expresan en una subpoblación. Para el análisis de unos pocos genes este estudio puede realizarse mediante selección de genes candidatos y detección por hibridación *in situ* junto al marcado metabólico con EdU para identificar las células proliferantes. Nuevamente, el desarrollo de herramientas de genómica funcional es de vital importancia para evaluar la esencialidad de los genes identificados mediante análisis de expresión diferencial.

Es interesante también analizar qué ocurre con las células proliferantes a lo largo del desarrollo estrobilar, mientras el gusano genera nuevos proglótides y desarrolla los órganos reproductivos. En estadios más desarrollados es posible caracterizar las células de la línea germinal, además de las células proliferantes somáticas, lo que permitirá establecer las relaciones entre ambos tipos celulares. La multiplicación asexual del estadio larvario observada tanto *in vitro* como en los ratones infectados es otro punto donde la proliferación es máxima que también resulta interesante para profundizar.

Los resultados obtenidos son un punto de inicio para el estudio más profundo de las células proliferantes en *M. corti*. Estos resultados indican que si bien existen genes marcadores comunes a las células madre de los organismos analizados, una gran cantidad de genes parecerían ser específicos de cada organismo, en contra de la hipótesis planteada en este trabajo. A medida que surjan estudios transcriptómicos en otros platelmintos se podrá definir el repertorio de genes conservados necesarios para el funcionamiento de este tipo celular y a su vez definir las diferencias y similitudes con las células madre de sus hospederos.

Al ser las únicas células con capacidad proliferativa, conocer genes esenciales para su funcionamiento y mantenimiento permitirá desarrollar nuevas estrategias de control, dirigidas a matar al parásito al impedir el mantenimiento de sus tejidos y evitar el desarrollo del estadio adulto generador de huevos que cierra el ciclo.

REFERENCIAS

- Aboobaker, A.A. 2011. Planarian stem cells: a simple paradigm for regeneration. *Trends in Cell Biology* 21: 304-311.
- Adema, C.M., Léonard, P., DeJong, R., Day, H., Edwards, D., Burgett, G., Hertel, L., Loker, E. 2000. Analysis of messages expressed by *Echinostoma paraense* miracida and sporocysts, obtained by random EST sequencing. *J Parasitol* 86: 60-65.
- Alam-Eldin, Y.H., Badawy, A.F. 2015. Destructive effect of gamma irradiation on *Echinococcus granulosus* metacestodes. *Parasitology Research* 114: 3145-3150.
- Alexa, A., Rahnenfuhrer, J. 2016. *topGO: Enrichment Analysis for Gene Ontology*.
- Almeida, C.R., Stoco, P.H., Wagner, G., Sincero, T.C., Rotava, G., Bayer-Santos, E., Rodrigues, J.B. et al. 2009. Transcriptome analysis of *Taenia solium* cysticerci using Open Reading Frame ESTs (ORESTES). *Parasites & Vectors* 2: 35.
- Ariumi, Y. 2014. Multiple functions of DDX3 RNA helicase in gene regulation, tumorigenesis, and viral infection. *Frontiers in Genetics* 5: 423.
- Auladell, C., Garcia-Valero, J., Baguñà, J. 1993. Ultrastructural localization of RNA in the chromatoid bodies of undifferentiated cells (neoblasts) in planarians by the RNase-gold complex technique. *Journal of Morphology* 216: 319-326.
- Baguna, J., Salo, E., Auladell, C. 1989. Regeneration and pattern formation in planarians. III. that neoblasts are totipotent stem cells and the source of blastema cells. *Development* 107: 77-86.
- Baguñà, J. 2018. Planarian regeneration between 1960s and 1990s: From skilful baffled ancestors to bold integrative descendants. A personal account. *Seminars in Cell & Developmental Biology* 87: 3-12.
- Baguñà, J., Romero, R. 1981. Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians *Dugesia mediterranea* and *Dugesia tigrina*. *Hidrobiología* 84: 181-194.
- Barrett, N.J., Smyth, J.D., Ong, S.J. 1982. Spontaneous sexual differentiation of *Mesocostoides corti* tetrathyridia in vitro. *International Journal for Parasitology* 12: 315-322.
- Berriman, M., Haas, B.J., LoVerde, P.T., Wilson, R.A., Dillon, G.P., Cerqueira, G.C., Mashiyama, S.T. et al. 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460: 352-358.
- Berriman, M., Stanley, E., Coghlan, A. 2018. A MAKER pipeline for prediction of protein-coding genes in parasitic worm genomes. *Protocol Exchange* v1: .
- Bizarro, C.V., Bengtson, M.H., Ricachenevsky, F.K., Zaha, A., Sogayar, M.C., Ferreira, H.B. 2005. Differentially expressed sequences from a cestode parasite reveals conserved developmental genes in platyhelminthes. *Molecular and Biochemical Parasitology* 144: 114-118.
- Bizzozero, R. 2010. *Aporte al conocimiento de la biología de los platelmintos parásitos que*

- contribuyan al diseño de estrategias de diagnóstico y tratamientos. Tesis de Grado. Licenciatura en Ciencias Biológicas. Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.*
- Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.
- Brehm, K., Jensen, K., Frosch, M. 2000. mRNA Trans-splicing in the Human Parasitic Cestode *Echinococcus multilocularis*. *Journal of Biological Chemistry* 275: 38311-38318.
- Brehm, K., Koziol, U. 2014. On the importance of targeting parasite stem cells in anti-echinococcosis drug development. *Parasite* 21: 72.
- Brehm, K., Kronthaler, K., Jura, H., Frosch, M. 2000. Cloning and characterization of β -tubulin genes from *Echinococcus multilocularis*. *Molecular and Biochemical Parasitology* 107: 297-302.
- Brennan, G.P., Fairweather, I., Trudgett, A., Hoey, E., McCoy, McConville, M., Meaney, M. et al. 2007. Understanding triclabendazole resistance. *Experimental and Molecular Pathology* 82: 104-109.
- Britos, L. 2000. *Estudios celulares y moleculares del desarrollo estrobilar de Mesocostoides corti*. Tesis de Maestría en Ciencias Biológicas - PEDECIBA. Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.
- Britos, L., Domínguez, L., Ehrlich, R., Marín, M. 2000. Effect of praziquantel on the strobilar development of *Mesocostoides corti* *in vitro*. *Journal of Helminthology* 74: 295-299.
- Britos, L., Lalanne, A.I., Castillo, E., Cota, G., Señorale, M., Marín, M. 2007. *Mesocostoides corti* (syn. *vogae*, cestoda): Characterization of genes encoding cysteine-rich secreted proteins (CRISP). *Experimental Parasitology* 116: 95-102.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.
- Cancela, M., Ruétalo, N., Dell'Oca, N., da Silva, E., Smircich, P., Rinaldi, G., Roche, L. et al. 2010. Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*. *BMC Genomics* 11: 227.
- Cantacessi, C., Gasser, R.B. 2012. SCP/TAPS proteins in helminths – Where to from now? *Molecular and Cellular Probes* 26: 54-59.
- Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C. et al. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18: 188-196.
- Caurla, G. 2015. *Caracterización de células proliferantes en platelmintos parásitos y estudio de un posible marcador molecular*. Maestría en Ciencias Biológicas - PEDECIBA. Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.
- Chalmers, I.W., McArdle, A.J., Coulson, R.M., Wagner, M.A., Schmid, R., Hirai, H., Hoffmann, K.F. 2008. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC Genomics* 9: 89.

-
- Chang, T.-C., Liu, W.-S. 2010. The molecular evolution of PL10 homologs. *BMC Evolutionary Biology* 10: 127.
- Chen, H. 2018. *VennDiagram: Generate High-Resolution Venn and Euler Plots*.
- Cheng, Z., Liu, F., Dai, M., Wu, J., Li, X., Guo, X., Tian, H. et al. 2017. Identification of EmSOX2, a member of the Sox family of transcription factors, as a potential regulator of *Echinococcus multilocularis* germinative cells. *International Journal for Parasitology* 47: 625-632.
- Cheng, Z., Liu, F., Li, X., Dai, M., Wu, J., Guo, X., Tian, H. et al. 2017. EGF-mediated EGFR/ERK signaling pathway promotes germinative cell proliferation in *Echinococcus multilocularis* that contributes to larval growth and development Brehm, K. (ed.),. *PLOS Neglected Tropical Diseases* 11: e0005418.
- Coghlan, A., Tyagi, R., Cotton, J.A., Holroyd, N., Rosa, B.A., Tsai, I.J., Laetsch, D.R. et al. 2019. Comparative genomics of the major parasitic worms. *Nature Genetics* 51: 163-174.
- Collins, J.J., Wang, B., Lambrus, B.G., Tharp, M.E., Iyer, H., Newmark, P.A. 2013. Adult somatic stem cells in the human parasite *Schistosoma mansoni*. *Nature* 494: 476-479.
- Collins, J.J., Wendt, G.R., Iyer, H., Newmark, P.A. 2016. Stem cell progeny contribute to the schistosome host-parasite interface. *eLife* 5: .
- Costábile, A., Koziol, U., Tort, J.F., Iriarte, A., Castillo, E. 2018. Expansion of cap superfamily proteins in the genome of *Mesocostoides corti*: An extreme case of a general bilaterian trend. *Gene Reports* 11: 110-120.
- Costábile, A., Marín, M., Castillo, E. 2017. Spatio-temporal expression of *Mesocostoides corti* McVAL2 during strobilar development. *Experimental Parasitology* 181: 30-39.
- Crosbie, P.R., Nadler, S.A., Platzer, E.G., Kerner, C., Mariaux, J., Boyce, W.M. 2000. Molecular systematics of *Mesocostoides* spp. (Cestoda: Mesocostoididae) from domestic dogs (*Canis familiaris*) and coyotes (*Canis latrans*). *The Journal of Parasitology* 82: 350-357.
- Cwiklinski, K., Dalton, J.P., Dufresne, P.J., La Course, J., Williams, D.J., Hodgkinson, J., Paterson, S. 2015. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. *Genome Biology* 16: 71.
- Cwiklinski, K., Jewhurst, H., McVeigh, P., Barbour, T., Maule, A.G., Tort, J., O'Neill, S.M. et al. 2018. Infection by the Helminth Parasite *Fasciola hepatica* Requires Rapid Regulation of Metabolic, Virulence, and Invasive Factors to Adjust to Its Mammalian Host. *Molecular & Cellular Proteomics* 17: 792-809.
- Davies, E.L., Lei, K., Seidel, C.W., Kroesen, A.E., McKinney, S.A., Guo, L., Robb, S.M. et al. 2017. Embryonic origin of adult stem cells required for tissue homeostasis and regeneration. *eLife* 6: e21052.
- Davis, R.E., Hardwick, C., Tavernier, P., Hodgson, S., Singh, H. 1995. RNA Trans-splicing in Flatworms: ANALYSIS OF TRANS-SPLICED mRNAs AND GENES IN THE HUMAN PARASITE, *SCHISTOSOMA MANSONI*. *Journal of Biological Chemistry* 270: 21813-
-

21819.

- De Keuckelaere, E., Hulpiau, P., Saeys, Y., Berx, G., van Roy, F. 2018. *Nanos* genes and their role in development and beyond. *Cellular and molecular life sciences* 75: 1929-1946.
- Domínguez, M.F. 2016. *Aislamiento, cultivo y caracterización de células proliferativas de M. corti*. Tesis de Doctorado en Ciencias Biológicas, PEDECIBA Biología. Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.
- Domínguez, M.F., Koziol, U., Porro, V., Costábile, A., Estrade, S., Tort, J., Bollati-Fogolin, M., Castillo, E. 2014. A new approach for the characterization of proliferative cells in cestodes. *Experimental Parasitology* 138: 25-29.
- Eckert, J., von Brand, T., Voge, M. 1969. Asexual multiplication of *Mesocostoides corti* (Cestoda) in the Intestine of dogs and skunks. *The Journal of Parasitology* 55: 241-249.
- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792-1797.
- Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., Škunca, N. et al. 2015. A Transcriptomic-Phylogenomic Analysis of the Evolutionary Relationships of Flatworms. *Current Biology* 25: 1347-1353.
- Eisenhoffer, G.T., Kang, H., Sánchez Alvarado, A. 2008. Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell* 3: 327-339.
- Emms, D.M., Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 157.
- Espinoza, I., Gomez, C.R., Galindo, M., Galanti, N. 2007. Developmental expression pattern of histone H4 gene associated to DNA synthesis in the endoparasitic platyhelminth *Mesocostoides corti*. *Gene* 386: 35-41.
- Extavour, C.G.M. 2007. Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. *Integrative and Comparative Biology* 47: 770-785.
- Fernández, C., Gregory, W.F., Loke, P., Maizels, R.M. 2002. Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and trans-spliced transcripts and a high level of predicted signal peptide sequences. *Molecular and Biochemical Parasitology* 122: 171-180.
- Fincher, C.T., Wurtzel, O., Hoog, T. de, Kravarik, K.M., Reddien, P.W. 2018. Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* 360: eaq1736.
- Fontenla, S., Rinaldi, G., Smircich, P., Tort, J.F. 2017. Conservation and diversification of small RNA pathways within flatworms. *BMC Evolutionary Biology* 17: 215.
- Fromm, B., Harris, P.D., Bachmann, L. 2011. MicroRNA preparations from individual monogenean *Gyrodactylus salaris* - a comparison of six commercially available totalRNA extraction kits. *BMC Research Notes* 4: 217.

-
- Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150-3152.
- Galloni, M. 2012. Global irradiation effects, stem cell genes and rare transcripts in the planarian transcriptome. *The International Journal of Developmental Biology* 56: 103-116.
- Garcia, H.H., Moro, P.L., Schantz, P.M. 2007. Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis. *Current Opinion in Infectious Diseases* 20: 489-494.
- Gibbs, G.M., Roelants, K., O'Bryan, M.K. 2008. The CAP Superfamily: Cysteine-Rich Secretory Proteins, Antigen 5, and Pathogenesis-Related 1 Proteins—Roles in Reproduction, Cancer, and Immune Defense. *Endocrine Reviews* 29: 865-897.
- Goldstrohm, A.C., Hall, T.M.T., McKenney, K.M. 2018. Post-transcriptional Regulatory Functions of Mammalian Pumilio Proteins. *Trends in genetics* 34: 972-990.
- Gremme, G., Steinbiss, S., Kurtz, S. 2013. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 10: 645-656.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59: 307-321.
- Guo, T., Peters, A.H.F.M., Newmark, P.A. 2006. A bruno-like Gene Is Required for Stem Cell Maintenance in Planarians. *Developmental Cell* 11: 159-169.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B. et al. 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature protocols* 8: 1494-1512.
- Hahn, C., Fromm, B., Bachmann, L. 2014. Comparative Genomics of Flatworms (Platyhelminthes) Reveals Shared Genomic Features of Ecto- and Endoparasitic Neodermata. *Genome Biology and Evolution* 6: 1105-1117.
- Handberg-Thorsager, M., Saló, E. 2007. The planarian *nanos*-like gene *Smednos* is expressed in germline and eye precursor cells during development and regeneration. *Development Genes and Evolution* 217: 403-411.
- Hay, E.D., Coward, S.J. 1975. Fine structure studies on the planarian, *Dugesia*: I. Nature of the “neoblast” and other cell types in noninjured worms. *Journal of Ultrastructure Research* 50: 1-21.
- Hayashi, T., Asami, M., Higuchi, S., Shibata, N., Agata, K. 2006. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Development, Growth and Differentiation* 48: 371-380.
- Hayashi, T., Shibata, N., Okumura, R., Kudome, T., Nishimura, O., Tarui, H., Agata, K. 2010. Single-cell gene profiling of planarian stem cells using fluorescent activated cell sorting and its “index sorting” function for stem cell research: Single-cell gene profiling of stem cell. *Development, Growth & Differentiation* 52: 131-144.
- Hemler, M.E. 2005. Tetraspanin functions and associated microdomains. *Nature Reviews*
-

-
- Molecular Cell Biology* 6: 801.
- Henry, A. 1927. Tétrathyridium et Mesocestoides. *Bull Soc Cent Ned Vet* 80: 147-152.
- Hess, E. 1975. Apropos of the asexual multiplication of the tetrathyridia larva of *Mesocestoides corti* Hoeppli, 1925 (Cestoda: cyclophyllidea). (Preliminary note). *Acta Tropica* 32: 290-295.
- Hess, E. 1981. Ultrastructural study of the tetrathyridium of *Mesocestoides corti* Hoeppli, 1925 (Cestoda): pool of germinative cells and suckers. *Rev Suisse Zool* 88: 661-674.
- Hess, E. 1980. Ultrastructural study of the tetrathyridium of *Mesocestoides corti* Hoeppli, 1925: Tegument and parenchyma. *Zeitschrift für Parasitenkunde Parasitology Research* 61: 135-159.
- Higuchi, S., Hayashi, T., Hori, I., Shibata, N., Sakamoto, H., Agata, K. 2007. Characterization and categorization of fluorescence activated cell sorted planarian stem cells by ultrastructural analysis: Heterogeneity of planarian stem cells. *Development, Growth & Differentiation* 49: 571-581.
- Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., Stanke, M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32: 767-769.
- Holt, C., Yandell, M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12: 491.
- Hotez, P.J., Brindley, P.J., Bethony, J.M., King, C.H., Pearce, E.J., Jacobson, J. 2008. Helminth infections: the great neglected tropical diseases. *Journal of Clinical Investigation* 118: 1311-1321.
- Howe, K.L., Bolt, B.J., Cain, S., Chan, J., Chen, W.J., Davis, P., Done, J. et al. 2016. WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Research* 44: D774-D780.
- Hu, W., Yan, Q., Shen, D.-K., Liu, F., Zhu, Z.-D., Song, H.-D., Xu, X.-R. et al. 2003. Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource. *Nature Genetics* 35: 139-147.
- Huang, Y., Chen, W., Wang, X., Liu, H., Chen, Y., Guo, L., Luo, F. et al. 2013. The Carcinogenic Liver Fluke, *Clonorchis sinensis*: New Assembly, Reannotation and Analysis of the Genome and Characterization of Tissue Transcriptomes Ralph, S. A. (ed.),. *PLoS ONE* 8: e54732.
- Hull, R., Dlamini, Z. 2014. The role played by alternative splicing in antigenic variability in human endo-parasites. *Parasites & Vectors* 7: 53.
- Irish, V., Lehmann, R., Akam, M. 1989. The *Drosophila* posterior-group gene nanos functions by repressing *hunchback* activity. *Nature* 338: 646-648.
- Jin, Y., Tam, O.H., Paniagua, E., Hammell, M. 2015. TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* 31: 3593-3599.
-

-
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H. et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236-1240.
- Juliano, C.E., Swartz, S.Z., Wessel, G.M. 2010. A conserved germline multipotency program. *Development* 137: 4113-4126.
- Kanehisa, M., Sato, Y., Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* 428: 726-731.
- Kassambara, A. 2018. *ggpubr: «ggplot2» Based Publication Ready Plots*.
- Keibler, E., Brent, M.R. 2003. Eval: A software package for analysis of genome annotations. *BMC Bioinformatics* 4: 50.
- Kelleher, A., Darwiche, R., Rezende, W.C., Farias, L.P., Leite, L.C.C., Schneider, R., Asojo, O.A. 2014. *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* 70: 2186-2196.
- Kim, D., Langmead, B., Salzberg, S.L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12: 357-360.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 14: R36.
- Kim, T.-S., de Guzman, J.V., Kong, H.-H., Chung, D.-I. 2006. Comparison of gene representation between diploid and triploid *Paragonimus westermani* by Expressed Sequence Tag analyses. *Journal of Parasitology* 92: 803-816.
- Kobayashi, S., Yamada, M., Asaoka, M., Kitamura, T. 1996. Essential role of the posterior morphogen nanos for germline development in *Drosophila*. *Nature* 380: 708-711.
- Koziol, U. 2009. *Caracterización de células proliferantes en Mesocestoides corti (Cestoda), y de genes pumilio como posibles marcadores moleculares de las mismas*. Maestría en Ciencias Biológicas - PEDECIBA. Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay.
- Koziol, U., Brehm, K. 2015. Recent advances in *Echinococcus* genomics and stem cell research. *Veterinary Parasitology* 213: 92-102.
- Koziol, U., Castillo, E. 2011. Cell proliferation and differentiation in cestodes. En *Research in Helminths*, pp. 121-138. Transworld Research Network, Kervala, India.
- Koziol, U., Costábile, A., Domínguez, M.F., Iriarte, A., Alvite, G., Kun, A., Castillo, E. 2011. Developmental expression of high molecular weight tropomyosin isoforms in *Mesocestoides corti*. *Molecular and Biochemical Parasitology* 175: 181-191.
- Koziol, U., Domínguez, M.F., Marín, M., Kun, A., Castillo, E. 2010. Stem cell proliferation during in vitro development of the model cestode *Mesocestoides corti* from larva to adult worm. *Frontiers in Zoology* 7: 22.
- Koziol, U., Iriarte, A., Castillo, E., Soto, J., Bello, G., Cajarville, A., Roche, L., Marín, M.
-

-
2009. Characterization of a putative hsp70 pseudogene transcribed in protoscoleces and adult worms of *Echinococcus granulosus*. *Gene* 443: 1-11.
- Kozioł, U., Lalanne, A.I., Castillo, E. 2009. Hox Genes in the Parasitic Platyhelminthes *Mesocostoides corti*, *Echinococcus multilocularis*, and *Schistosoma mansoni*: Evidence for a Reduced Hox Complement. *Biochemical Genetics* 47: 100-116.
- Kozioł, U., Marín, M., Castillo, E. 2008. Pumilio genes from the Platyhelminthes. *Development Genes and Evolution* 218: 47-53.
- Kozioł, U., Radio, S., Smircich, P., Zarowiecki, M., Fernández, C., Brehm, K. 2015. A Novel Terminal-Repeat Retrotransposon in Miniature (TRIM) Is Massively Expressed in *Echinococcus multilocularis* Stem Cells. *Genome Biology and Evolution* 7: 2136-2153.
- Kozioł, U., Rauschendorfer, T., Rodríguez, L.Z., Krohne, G., Brehm, K. 2014. The unique stem cell system of the immortal larva of the human parasite *Echinococcus multilocularis*. *EvoDevo* 5: 10.
- Labbé, R.M., Irimia, M., Currie, K.W., Lin, A., Zhu, S.J., Brown, D.D.R., Ross, E.J. et al. 2012. A Comparative Transcriptomic Analysis Reveals Conserved Features of Stem Cell Pluripotency in Planarians and Mammals. *Stem cells (Dayton, Ohio)* 30: 1734-1745.
- Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., Ussery, D.W. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35: 3100-3108.
- Lalanne, A.I., Britos, L., Ehrlich, R., Castillo, E. 2004. *Mesocostoides corti*: a LIM-homeobox gene upregulated during strobilar development. *Experimental Parasitology* 108: 169-175.
- Langmead, B., Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-359.
- Lefort, V., Longueville, J.-E., Gascuel, O. 2017. SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution* 34: 2422-2424.
- Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D., Fellous, M. 1989. The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. *Cell* 57: 549-559.
- Leroy, P., Seboun', E., Mattei, M.G., Fellous, M., Bishop, C.E. 1987. Testis-specific transcripts detected by a human Y-DNA-derived probe. *Development* 101 (Suppl): 177-183.
- Li, W., Liu, B., Yang, Y., Ren, Y., Wang, S., Liu, C., Zhang, N. et al. 2018. The genome of tapeworm *Taenia multiceps* sheds light on understanding parasitic mechanism and control of coenurosis disease. *DNA research* 25: 499-510.
- Linder, P., Jankowsky, E. 2011. From unwinding to clamping — the DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology* 12: 505-516.
- Littlewood, D.T.J., Bray, R.A., Waeschenbach, A. 2015. Phylogenetic patterns of diversity in cestodes and trematodes. En Morand, S., Krasnov, B. R., Littlewood, D. T. J. E.
-

-
- (eds.), *Parasite Diversity and Diversification: Evolutionary Ecology Meets Phylogenetics*, pp. 304–319. Cambridge University Press.
- Liu, S., Zhou, X., Hao, L., Piao, X., Hou, N., Chen, Q. 2017. Genome-Wide Transcriptome Analysis Reveals Extensive Alternative Splicing Events in the Protoscoleces of *Echinococcus granulosus* and *Echinococcus multilocularis*. *Frontiers in Microbiology* 8: 929.
- Livak, K.J., Schmittgen, T.D. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta CT$ Method. *Methods* 25: 402-408.
- Loos-Frank, B. 1991. One or two intermediate hosts in the life cycle of *Mesocestoides* (Cyclophyllidae, Mesocestoididae)? *Parasitology Research* 77: 726-728.
- Maldonado, L.L., Assis, J., Araújo, F.M.G., Salim, A.C.M., Macchiaroli, N., Cucher, M., Camicia, F. et al. 2017. The *Echinococcus canadensis* (G7) genome: a key knowledge of parasitic platyhelminth human diseases. *BMC Genomics* 18: 204.
- McCusker, P., McVeigh, P., Rathinasamy, V., Toet, H., McCammick, E., O'Connor, A., Marks, N.J. et al. 2016. Stimulating Neoblast-Like Cell Proliferation in Juvenile *Fasciola hepatica* Supports Growth and Progression towards the Adult Phenotype In Vitro. *PLOS Neglected Tropical Diseases* 10: e0004994.
- McNulty, S.N., Tort, J.F., Rinaldi, G., Fischer, K., Rosa, B.A., Smircich, P., Fontenla, S. et al. 2017. Genomes of *Fasciola hepatica* from the Americas Reveal Colonization with *Neorickettsia* Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers Stukenbrock, E. H. (ed.), *PLOS Genetics* 13: e1006537.
- Molinaro, A.M., Pearson, B.J. 2016. In silico lineage tracing through single cell transcriptomics identifies a neural stem cell population in planarians. *Genome Biology* 17: 87.
- Morita, M., Best, J.B., Noel, J. 1969. Electron microscopic studies of planarian regeneration. I. Fine structure of neoblasts in *Dugesia dorotocephala*. *Journal of Ultrastructure Research* 27: 7-23.
- Nawrocki, E.P., Eddy, S.R. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29: 2933-2935.
- Newmark, P.A., Sánchez Alvarado, A. 2000. Bromodeoxyuridine Specifically Labels the Regenerative Stem Cells of Planarians. *Developmental Biology* 220: 142-153.
- Novak, M. 1972. Quantitative studies on the growth and multiplication of tetrathyridia of *Mesocestoides corti* Hoeppli, 1925 (Cestoda: Cyclophyllidae) in rodents. *Canadian Journal of Zoology* 50: 1189-1196.
- Ogawa, K., Kobayashi, C., Hayashi, T., Orii, H., Watanabe, K., Agata, K. 2002. Planarian fibroblast growth factor receptor homologs expressed in stem cells and cephalic ganglions. *Development, Growth and Differentiation* 44: 191-204.
- Okonechnikov, K., Golosova, O., Fursov, M. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28: 1166-1167.
- Olson, P.D., Zarowiecki, M., James, K., Baillie, A., Bartl, G., Burchell, P., Chellappoo, A. et al. 2018. Genome-wide transcriptome profiling and spatial expression analyses
-

- identify signals and switches of development in tapeworms. *EvoDevo* 9: 21.
- Önal, P., Grün, D., Adamidi, C., Rybak, A., Solana, J., Mastrobuoni, G., Wang, Y. et al. 2012. Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells: Conserved determinants of planarian pluripotency. *The EMBO Journal* 31: 2755-2769.
- Ong, S.J., Smyth, J.D. 1986. Effects of some culture factors on sexual differentiation of *Mesocestoides corti* grown from tetrathyridia *in vitro*. *International Journal for Parasitology* 16: 361-368.
- Orii, H., Sakurai, T., Watanabe, K. 2005. Distribution of the stem cells (neoblasts) in the planarian *Dugesia japonica*. 215: 143-157.
- Padgett, K.A., Boyce, W.M. 2004. Life-history studies on two molecular strains of *Mesocestoides* (Cestoda: Mesocestoididae): Identification of sylvatic hosts and infectivity of immature life stages. *Journal of Parasitology* 90: 108-113.
- Padgett, K.A., Nadler, S.A., Munson, L., Sacks, B., Boyce, W.M. 2005. Systematics of *Mesocestoides* (Cestoda: Mesocestoididae): Evaluation of molecular and morphological variation among isolates. *Journal of Parasitology* 91: 1435-1443.
- Palakodeti, D., Smielewska, M., Lu, Y.-C., Yeo, G.W., Graveley, B.R. 2008. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA* 14: 1174-1186.
- Parkinson, J., Wasmuth, J.D., Salinas, G., Bizarro, C.V., Sanford, C., Berriman, M., Ferreira, H.B. et al. 2012. A Transcriptomic Analysis of *Echinococcus granulosus* Larval Stages: Implications for Parasite Biology and Host Adaptation. *PLoS Neglected Tropical Diseases* 6: e1897.
- Pearson, B.J., Eisenhoffer, G.T., Gurley, K.A., Rink, J.C., Miller, D.E., Sánchez Alvarado, A. 2009. Formaldehyde-based whole-mount *in situ* hybridization method for planarians. *Developmental Dynamics* 238: 443-450.
- Pearson, M.S., Pickering, D.A., McSorley, H.J., Bethony, J.M., Tribolet, L., Dougall, A.M., Hotez, P.J., Loukas, A. 2012. Enhanced protective efficacy of a chimeric form of the schistosomiasis vaccine antigen Sm-TSP-2. *PLoS neglected tropical diseases* 6: e1564.
- Peter, R., Gschwentner, R., Schürmann, W., Rieger, R.M., Ladurner, P. 2004. The significance of stem cells in free-living flatworms: one common source for all cells in the adult. *Journal of Applied Biomedicine* 2: 21-35.
- Pfister, D., De Mulder, K., Hartenstein, V., Kualess, G., Borgonie, G., Marx, F., Morris, J., Ladurner, P. 2008. Flatworm stem cells and the germ line: Developmental and evolutionary implications of macvsa expression in *Macrostomum lignano*. *Developmental Biology* 319: 146-159.
- Pfister, D., De Mulder, K., Philipp, I., Kualess, G., Hrouda, M., Eichberger, P., Borgonie, G. et al. 2007. The exceptional stem cell system of *Macrostomum lignano*: Screening for gene expression and studying cell proliferation by hydroxyurea treatment and irradiation. *Frontiers in Zoology* 4: 9.
- Plass, M., Solana, J., Wolf, F.A., Ayoub, S., Misios, A., Glažar, P., Obermayer, B. et al.

-
2018. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* 360: eaaq1723.
- Pohle, S., Ernst, R., MacKenzie, C., Spicher, M., Romig, T., Hemphill, A., Gripp, S. 2011. *Echinococcus multilocularis*: The impact of ionizing radiation on metacestodes. *Experimental Parasitology* 127: 127-134.
- Pouchkina-Stantcheva, N.N., Cunningham, L.J., Hrčková, G., Olson, P.D. 2013. RNA-mediated gene suppression and in vitro culture in *Hymenolepis microstoma*. *International Journal for Parasitology* 43: 641-646.
- Prasad, T.S.K., Mohanty, A.K., Kumar, M., Sreenivasamurthy, S.K., Dey, G., Nirujogi, R.S., Pinto, S.M. et al. 2017. Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. *Genome Research* 27: 133-144.
- Price, M.N., Dehal, P.S., Arkin, A.P. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* 5: e9490.
- R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raghunathan, L., Voge, M. 1974. Chromosome analysis of *Mesocostoides corti* (Cestoda). *The Journal of Parasitology* 60: 558.
- Rausch, R. 1994. Family Mesocostoididae. En Khalil, L., Jones, A., Bray, R. (eds.), *Keys to the Cestode Parasites of Vertebrates*, pp. 309-314. CAB International, Wallingford, U.K.
- Reddien, P.W. 2005. SMEDWI-2 Is a PIWI-Like Protein That Regulates Planarian Stem Cells. *Science* 310: 1327-1330.
- Reddien, P.W., Sánchez Alvarado, A. 2004. Fundamentals of Planarian Regeneration. *Annual Review of Cell and Developmental Biology* 20: 725-757.
- Reuter, M., Kreshchenko, N. 2004. Flatworm asexual multiplication implicates stem cells and regeneration. *Canadian Journal of Zoology* 82: 334-356.
- Rink, J.C. 2013. Stem cell systems and regeneration in planaria. *Development Genes and Evolution* 223: 67-84.
- Rohart, F., Gautier, B., Singh, A., Lê Cao, K.-A. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology* 13: e1005752.
- Rosner, A., Rinkevich, B. 2007. The DDX3 subfamily of the DEAD box helicases: divergent roles as unveiled by studying different organisms and in vitro assays. *Current Medicinal Chemistry* 14: 2517-2525.
- Rossi, L., Salvetti, A., Batistoni, R., Deri, P., Gremigni, V. 2008. Molecular and Cellular Basis of Regeneration and Tissue Repair: Planarians, a tale of stem cells. *Cellular and Molecular Life Sciences* 65: 16-23.
- Rossi, L., Salvetti, A., Marincola, F.M., Lena, A., Deri, P., Mannini, L., Batistoni, R. et al. 2007. Deciphering the molecular machinery of stem cells: a look at the neoblast gene expression profile. *Genome Biology* 8: R62.
-

-
- Rouhana, L., Shibata, N., Nishimura, O., Agata, K. 2010. Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell maintenance. *Developmental Biology* 341: 429-443.
- Rozanski, A., Moon, H., Brandl, H., Martín-Durán, J.M., Grohme, M.A., Hüttner, K., Bartscherer, K. et al. 2019. PlanMine 3.0—improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Research* 47: D812-D820.
- Salveti, A., Rossi, L., Bonuccelli, L., Lena, A., Pugliesi, C., Rainaldi, G., Evangelista, M., Gremigni, V. 2009. Adult stem cell plasticity: Neoblast repopulation in non-lethally irradiated planarians. *Developmental Biology* 328: 305-314.
- Salveti, A., Rossi, L., Deri, P., Batistoni, R. 2000. An MCM2-related gene is expressed in proliferating cells of intact and regenerating planarians. *Developmental Dynamics* 218: 603-614.
- Salveti, A., Rossi, L., Lena, A., Batistoni, R., Deri, P., Rainaldi, G., Locci, M. et al. 2005. *DjPum*, a homologue of *Drosophila Pumilio*, is essential to planarian stem cell maintenance. *Development* 132: 1863-1874.
- Sánchez Alvarado, A. 2007. Stem cells and the Planarian *Schmidtea mediterranea*. *Comptes Rendus Biologies* 330: 498-503.
- Sánchez Alvarado, A., Newmark, P.A. 1999. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proceedings of the National Academy of Sciences* 96: 5049-5054.
- Sato, K., Shibata, N., Orii, H., Amikura, R., Sakurai, T., Agata, K., Kobayashi, S., Watanabe, K. 2006. Identification and origin of the germline stem cells as revealed by the expression of *nanos*-related gene in planarians. *Development, Growth & Differentiation* 48: 615-628.
- Schubert, A., Koziol, U., Cailliau, K., Vanderstraete, M., Dissous, C., Brehm, K. 2014. Targeting *Echinococcus multilocularis* stem cells by inhibition of the Polo-like kinase EmPlk1. *PLoS neglected tropical diseases* 8: e2870.
- Shibata, N., Rouhana, L., Agata, K. 2010. Cellular and molecular dissection of pluripotent adult somatic stem cells in planarians. *Development, Growth & Differentiation* 52: 27-41.
- Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K., Agata, K. 1999. Expression of *vasa(vas)*-Related Genes in Germline Cells and Totipotent Somatic Stem Cells of Planarians. *Developmental Biology* 206: 73-87.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R. et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 539.
- Skinner, D.E., Rainaldi, G., Koziol, U., Brehm, K., Brindley, P.J. 2014. How might flukes and tapeworms maintain genome integrity without a canonical piRNA pathway? *Trends in Parasitology* 30: 123-129.
- Skinner, D.E., Rainaldi, G., Suttiprapa, S., Mann, V.H., Smircich, P., Cogswell, A.A., Williams, D.L., Brindley, P.J. 2012. Vasa-like DEAD-Box RNA helicases of *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* 6: e1686.
-

- Smit, A., Hubley, R. 2008. *RepeatModeler Open-1.0*.
- Smith, A.G., McKerr, G. 2000. Tritiated thymidine ([³H]-TdR) and immunocytochemical tracing of cellular fate within the asexually dividing cestode *Mesocostoides vogae* (syn. *M. corti*). *Parasitology* 121 (Pt 1): 105-110.
- Solana, J., Kao, D., Mihaylova, Y., Jaber-Hijazi, F., Malla, S., Wilson, R., Aboobaker, A. 2012. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNA-seq, RNA interference and irradiation approach. *Genome Biology* 13: R19.
- Soldatova, A.P. 1944. A contribution to the study of the development cycle in the cestode *Mesocostoides lineatus* (Goeze, 1782), parasitic of carnivorous mammals. *Doklady Akademii nauk SSSR* 45: 310-312.
- Specht, D., Voge, M. 1965. Asexual Multiplication of *Mesocostoides* Tetrathyridia in Laboratory Animals. *The Journal of Parasitology* 51: 268.
- Spiliotis, M., Mizukami, C., Oku, Y., Kiss, F., Brehm, K., Gottstein, B. 2010. *Echinococcus multilocularis* primary cells: Improved isolation, small-scale cultivation and RNA interference. *Molecular and Biochemical Parasitology* 174: 83-87.
- Srivastava, H.D. 1939. A study of the life-history of a common tapeworm, *Mesocostoides lineatus*, of indian dogs and cats. *Indian Journal of Veterinary Science* 9: 187-190.
- Storer, T., Usinger, R., Stebbins, R., Nybakken, J., Gifre, E., Fontes, M. 1986. Platelminetos. En *Zoología General*, pp. 412-430. Omega S.A., Barcelona, España.
- Strzalka, W., Ziemienowicz, A. 2011. Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation. *Annals of Botany* 107: 1127-1140.
- The *Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium. 2009. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* 460: 345-351.
- Thompson, R.C.A., Jue Sue, L.P., Buckley, S.J. 1982. In vitro development of the strobilar stage of *Mesocostoides corti*. *International Journal for Parasitology* 12: 303-314.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H. et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7: 562-578.
- Tsai, IJ, Zarowiecki, M., Holroyd, N., Garcarrubio, A., Sanchez-Flores, A., Brooks, K.L., Tracey, A. et al. 2013. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496: 57-63.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35: W71-W74.
- van Wolfswinkel, J.C., Wagner, D.E., Reddien, P.W. 2014. Single-Cell Analysis Reveals Functionally Distinct Classes within the Planarian Stem Cell Compartment. *Cell Stem Cell* 15: 326-339.
- Vasconcelos, E.J.R., daSilva, L.F., Pires, D.S., Lavezzo, G.M., Pereira, A.S.A., Amaral,

-
- M.S., Verjovski-Almeida, S. 2017. The *Schistosoma mansoni* genome encodes thousands of long non-coding RNAs predicted to be functional at different parasite life-cycle stages. *Scientific Reports* 7: 10508.
- Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimarães, P.E.M., Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P. et al. 2003. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics* 35: 148-157.
- Voge, M., Coulombe, L.S. 1966. Growth and asexual multiplication *in vitro* of *Mesocostoides* tetrathyridia. *The American Journal of Tropical Medicine and Hygiene* 15: 902-907.
- Voge, M., Seidel, J.S. 1968. Continuous growth *in vitro* of *Mesocostoides* (Cestoda) from oncosphere to fully developed tetrathyridium. *The Journal of Parasitology* 54: 269.
- Wagner, D.E., Ho, J.J., Reddien, P.W. 2012. Genetic Regulators of a Pluripotent Adult Stem Cell System in Planarians Identified by RNAi and Clonal Analysis. *Cell Stem Cell* 10: 299-311.
- Wagner, D.E., Wang, I.E., Reddien, P.W. 2011. Clonogenic Neoblasts Are Pluripotent Adult Stem Cells That Underlie Planarian Regeneration. *Science* 332: 811-816.
- Wang, B., Collins, J.J., Newmark, P.A. 2013. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. *eLife* 2: e00768.
- Wang, B., Lee, J., Li, P., Saberi, A., Yang, H., Liu, C., Zhao, M., Newmark, P.A. 2018. Stem cell heterogeneity drives the parasitic life cycle of *Schistosoma mansoni*. *eLife* 7: e35449.
- Wang, C., Lehmann, R. 1991. Nanos is the localized posterior determinant in *Drosophila*. *Cell* 66: 637-647.
- Wang, J., Collins, J.J. 2016. Identification of new markers for the *Schistosoma mansoni* vitelline lineage. *International Journal for Parasitology* 46: 405-410.
- Wang, X., Chen, W., Huang, Y., Sun, J., Men, J., Liu, H., Luo, F. et al. 2011. The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. *Genome Biology* 12: R107.
- Wang, Y., Zayas, R.M., Guo, T., Newmark, P.A. 2007. *nanos* function is essential for development and regeneration of planarian germ cells. *Proceedings of the National Academy of Sciences* 104: 5901-5906.
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* 35: 543-548.
- Webb, C., Cabada, M.M. 2017. Intestinal cestodes. *Current Opinion in Infectious Diseases* 30: 504-510.
- Webster, J.D. 1949. Fragmentary studies on the life history of the Cestode *Mesocostoides latus*. *The Journal of Parasitology* 35: 83.
- Wendt, G.R., Collins, J.N., Pei, J., Pearson, M.S., Bennett, H.M., Loukas, A., Berriman, M. et al. 2018. Flatworm-specific transcriptional regulators promote the specification of
-

- hr/>
- tegumental progenitors in *Schistosoma mansoni*. *eLife* 7: e33221.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Winnebeck, E.C., Millar, C.D., Warman, G.R. 2010. Why Does Insect RNA Look Degraded? *Journal of Insect Science* 10: 159.
- Witenberg, G. 1934. Studies on the cestode genus *Mesocetoides*. *Arch Zool Ital* 20: 467-509.
- Wurtzel, O., Cote, L.E., Poirier, A., Satija, R., Regev, A., Reddien, P.W. 2015. A Generic and Cell-Type-Specific Wound Response Precedes Regeneration in Planarians. *Developmental Cell* 35: 632-645.
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L. et al. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research* 39: W316-322.
- Yandell, M., Ence, D. 2012. A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics* 13: 329-342.
- Yoshida-Kashikawa, M., Shibata, N., Takechi, K., Agata, K. 2007. DjCBC-1, a conserved DEAD box RNA helicase of the RCK/p54/Me31B family, is a component of RNA-protein complexes in planarian stem cells and neurons. *Developmental Dynamics* 236: 3436-3450.
- Young, N.D., Campbell, B.E., Hall, R.S., Jex, A.R., Cantacessi, C., Laha, T., Sohn, W.-M. et al. 2010. Unlocking the Transcriptomes of Two Carcinogenic Parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. *PLoS Neglected Tropical Diseases* 4: e719.
- Young, N.D., Hall, R.S., Jex, A.R., Cantacessi, C., Gasser, R.B. 2010. Elucidating the transcriptome of *Fasciola hepatica* — A key to fundamental and biotechnological discoveries for a neglected parasite. *Biotechnology Advances* 28: 222-231.
- Young, N.D., Jex, A.R., Cantacessi, C., Hall, R.S., Campbell, B.E., Spithill, T.W., Tangkawattana, S. et al. 2011. A Portrait of the Transcriptome of the Neglected Trematode, *Fasciola gigantica*—Biological and Biotechnological Implications. *PLoS Neglected Tropical Diseases* 5: e1004.
- Young, N.D., Jex, A.R., Li, B., Liu, S., Yang, L., Xiong, Z., Li, Y. et al. 2012. Whole-genome sequence of *Schistosoma haematobium*. *Nature Genetics* 44: 221-225.
- Zayas, R.M., Bold, T.D., Newmark, P.A. 2005. Spliced-Leader trans-Splicing in Freshwater Planarians. *Molecular Biology and Evolution* 22: 2048-2054.
- Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simão, F.A., Ioannidis, P., Seppey, M. et al. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research* 45: D744-D749.
- Zeng, A., Li, H., Guo, L., Gao, X., McKinney, S., Wang, Y., Yu, Z. et al. 2018. Prospectively Isolated Tetraspanin+ Neoblasts Are Adult Pluripotent Stem Cells Underlying Planaria Regeneration. *Cell* 173: 1593-1608.e20.
- Zhao, W.J., Zhang, H., Bo, X., Li, Y., Fu, X. 2009. Generation and analysis of expressed

- sequence tags from a cDNA library of *Moniezia expansa*. *Molecular and Biochemical Parasitology* 164: 80-85.
- Zheng, H., Zhang, W., Zhang, L., Zhang, Z., Li, J., Lu, G., Zhu, Y. et al. 2013. The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nature Genetics* 45: 1168-1175.
- Zhu, S.J., Pearson, B.J. 2016. (Neo)blast from the past: new insights into planarian stem cell lineages. *Current Opinion in Genetics & Development* 40: 74-80.



Contents lists available at ScienceDirect

Gene Reports

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

Expansion of cap superfamily proteins in the genome of *Mesocostoides corti*: An extreme case of a general bilaterian trend



Alicia Costáble^a, Uriel Koziol^b, José F. Tort^c, Andrés Iriarte^{d,*}, Estela Castillo^{a,*}

^a Sección Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de la República, Igúá 4225, CP 11400 Montevideo, Uruguay

^b Sección Biología Celular, Facultad de Ciencias, Universidad de la República, Igúá 4225, CP 11400 Montevideo, Uruguay

^c Departamento de Genética, Facultad de Medicina, Universidad de la República, Gral. Flores 2125, CP 11800 Montevideo, Uruguay

^d Laboratorio Biología Computacional, Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Universidad de la República, Av. Alfredo Navarro 3051, Montevideo CP 11600, Uruguay

ARTICLE INFO

Keywords:

CAP domain
SCP domain
Birth and dead evolution
Platyhelminthes

ABSTRACT

The CAP superfamily is a diverse group of proteins that are involved in different biological processes, yet their molecular functions are still incompletely understood. The α - β sandwich structure of the CAP domain is characteristic of this superfamily and several different domains may be found together with it. They are generally secreted proteins and in helminths many are secreted to the environment, and are related to the host-parasite interaction. In this work we mined cestode genomic data for members of this superfamily. Whereas in average 26 members with complete CAP domains were found in most cestodes, in *Mesocostoides corti* we strikingly found 271 members with complete domains, most of which show evidence of expression. We also found other truncated domains and putative pseudogenes. Interestingly, most of these genes were found in a monophyletic clade within a cestode-specific group of CAP domain containing proteins, and each cestode species has also developed independent duplications of these proteins. This pattern of extensive independent duplications can also be found in other parasitic and free-living flatworms, as well as in other metazoan phyla. Within the *M. corti* specific expansion, several sub-clades of these proteins showed evidence of evolution under positive selection. Our results suggest that the CAP domain containing proteins of animals evolve through a “birth and death” mechanism, and that different environmental pressures may drive this evolution in different species. In the case of helminth parasites, this could be related to the interaction between the parasite and the host, including mechanisms to evade and modulate the host immune system.

1. Introduction

The CAP superfamily includes a diversity of proteins containing the CAP conserved domain, previously named SCP/TAPS (Pfam accession number PF00188). Members of this superfamily are broadly distributed along phyla, being present in organisms as distant as bacteria (Yeats et al., 2003), plants (van Loon and van Strien, 1999), cnidarians (Hwang et al., 2007) and mammals (Gibbs et al., 2008).

The original name given to the domain comes from the first members characterized: SCP (*Sperm Coating Protein*), Tpx-1 (*Testis Specific*

Protein-1), Ag 5 (*Antigen 5*); PR-1 (*Pathogenesis-related-1*) and Sc7/14. SCP (Brooks et al., 1986) and Tpx-1 (Mizuki et al., 1992) proteins belong to the CRISP family of cysteine – rich secreted proteins, involved in mammal reproduction. Ag5 proteins are present in insect secretions, being the principal allergen in wasp venoms (Fang et al., 1988), while PR-1 proteins are expressed by plants under stress conditions or pathogen invasion (Cornelissen et al., 1986). Sc7 and Sc14 proteins are involved in fructiferous body formation in the fungus *Schizophyllum commune* (Schuren et al., 1993). Recently, the name of the superfamily was changed to CAP (from the acronym for CRISP, Ag5 and PR-1),

Abbreviations: VAL, venom allergen-like; CRISP, cysteine-rich secretory protein; Ag5, Antigen 5; PR1, Pathogen-Related 1; SCP, Sperm Coating Protein; Tpx1, Testis Specific Protein 1; Sc7/Sc14, *Schizophyllum commune* gene 7/14 family; GLIPIR, Glioma Pathogenesis-Related; GLIPIR1L, Glioma pathogenesis-related 1 like; PRY, pathogenesis-related yeast; ASP, Activation-associated Secreted Proteins or *Ancylostoma* Secreted Protein; GAPR1, Golgi-Associated Plant pathogenesis-Related 1; Ac-NIF, *Ancylostoma caninum* Neutrophil Inhibitory Factor; Ac-HPI, *Ancylostoma caninum* hookworm platelet inhibitor; LTR, Long Terminal Repeat; LINE, Long Interspersed Nucleotide Elements; SINE, Short Interspersed Nucleotide Elements; TE, transposable element; tRNA, transfer Ribonucleic Acid; gDNA, genomic Desoxyribonucleic Acid; snRNA, Small Nuclear Ribonucleic Acid; rRNA, ribosomal Ribonucleic Acid; EST, Expressed Sequence Tag; RACE, Rapid Amplification of cDNA Ends; fpkm, Fragments Per Kilobase Million; Th1/2, T-helper 1 or 2; IFN, Interferon; IL, interleukin; ES, excretory-secretory; LRT, likelihood ratio test; Pr., probability; SH-like, Shimodaira-Hasegawa-like

* Corresponding authors.

E-mail addresses: airiarte@higiene.edu.uy (A. Iriarte), castillo@fcien.edu.uy (E. Castillo).

<https://doi.org/10.1016/j.genrep.2018.03.010>

Received 10 October 2017; Received in revised form 26 February 2018; Accepted 9 March 2018

Available online 10 March 2018

2452-0144/ © 2018 Published by Elsevier Inc.

because it is more representative of the families belonging to the superfamily and not of specific members (Gibbs et al., 2008; Roberts et al., 2007, among others). The tridimensional structure of several proteins was solved, showing that the domain adopts a α - β - α sandwich fold, stabilized with internal hydrogen bonds and a variable number of disulfide bonds. This disulfide bonds are present in the members that contain an N-terminal signal peptide for secretion, explaining their extracytosolic stability. GAPR-1, a human protein lacking the secretion signal peptide, does not contain disulfide bonds, consistent with its cytosolic location (Gibbs et al., 2008).

In helminths, most of the studies about proteins belonging to this superfamily were done in nematodes. These proteins are commonly known as ASP (acronym for *Activation – associated Secreted Protein* or *Ancylostoma – Secreted Protein*) because their expression is highly up-regulated during the transition from the free – living stage to the parasitic stage in the dog worm *Ancylostoma caninum* (Datu et al., 2008).

In platyhelminths, Chalmers and Hoffmann (2012) and Cantacessi et al. (2012) reviewed current knowledge about these proteins. In trematodes, bioinformatics analysis showed the presence of 11 to 41 proteins containing the CAP domain in the genomes of different species of trematodes (Cantacessi et al., 2012). Later, Kelleher et al. (2014) reported the first crystal structure of a platyhelminth CAP domain containing protein (SmVAL4). This protein lacks the characteristic motifs, but it adopts the tridimensional fold observed by other prototypical members. They also proved that it can bind sterols like cholesterol, as PRY members of *S. cerevisiae* do (Choudhary and Schneider, 2012), and that the protein is able to bind palmitate, as Tablysin-15 of the horsefly *Tabanus yao* does (Xu et al., 2012). Experimental studies were carried out mainly for *S. mansoni* members, showing that several of them are differentially expressed across the different life-stages of the parasite (Chalmers et al., 2008; Rofatto et al., 2012). Some of them are exclusively expressed in the mammalian or the molluscan stages, so they may be implicated in the invasion and establishment of the parasite in their host. For example, SmVAL9 participates in tissue reorganization and extracellular matrix remodeling during intra-mammalian egg translocation, miracidia infection and intra-molluscan sporocyst development/migration (Yoshino et al., 2014). Sj-VAL 1 is expressed in cercariae and eggs of *S. japonicum*, and elicits an antibody response in mice where IgG1 is the predominant isotype (Chen et al., 2010).

In cestodes, fewer members were identified in bioinformatic searches of EST databases: two in *Taenia solium*, and only one in *Taenia asiatica*, *Taenia saginata*, *Moniezia expansa* and *Echinococcus multilocularis* (Chalmers and Hoffmann, 2012). We characterized two *Echinococcus granulosus* VAL proteins, which are expressed predominantly in the rostellum of the larval stage (Silvarrey et al., 2016). In a recent study, Liu et al. (2017) identified 22 VAL genes in *E. granulosus* and 19 in *E. multilocularis*. Among these, 9 genes are expressed in the protoscolex of both species, with one presenting alternative splicing in *E. multilocularis* and two in *E. granulosus*. Free living flatworms also contain proteins of this superfamily, indicating that some of these proteins have a function not related to parasitism (Chalmers and Hoffmann, 2012).

Mesocostoides corti is a model used to study cestode biology, as it is easily maintained in the laboratory (Barret et al., 1982; Britos et al., 2000; Ong and Smyth, 1986; Thompson et al., 1982; Voge and Seidel, 1968). We previously identified four transcripts coding for CAP domain containing proteins, with a putative secretion signal peptide (Britos et al., 2007). We thoroughly characterized expression of one of them (McVAL2, previously known as McCRISP2) during strobilar development, showing that it is dynamically and stage-specific expressed. We also identified 46 genes containing this domain in the automatic genome annotation, indicating that these proteins may have different roles in parasite biology (Costabile et al., 2017).

In this work we identified and manually annotated CAP domain

containing proteins in *M. corti* and characterized the phylogenetic relationships with other cestodes, as well as with other metazoans. We show that a massive expansion of these genes has occurred in *M. corti*, and many of these genes show clear evidence of expression. Furthermore, we show that similar independent expansions have occurred recurrently in different bilaterian groups. The largest clade-specific expansions were found in different parasitic helminths, suggesting that they could be related to the interaction between these parasites and their hosts.

2. Methods

2.1. Searches for CAP domain containing proteins

2.1.1. Cestode predicted proteins

The full length McCRISP2 transcript (*M. corti*, GenBank accession number: AY671940) was previously isolated in our lab (Britos et al., 2007). Using this sequence as query, tblastx searches were done, looking for homologous sequences in *E. multilocularis* draft genome (version 1 accessed on September 2010: <ftp://ftp.sanger.ac.uk/pub/project/pathogens/Echinococcus/multilocularis/genome/ARCHIVE/version1.EMU.fas.27-7-2010.gz>). Since at the beginning of this work, *M. corti* genome was not available, we followed a 3' RACE strategy (according to Kozioł et al. (2011)) using degenerated primers for a N-terminal conserved motif (Supplementary Table 1) to isolate new members of this family in *M. corti*. A pool of DNase-treated RNAs obtained from worms in different stages of strobilar development was used for retrotranscription. PCR products were gel purified and cloned in pGEM-T-Easy vector (Promega). Several clones were selected for plasmid extraction and sequencing (Macrogen Korea or Institut Pasteur de Montevideo).

Once published cestode genomes became available, we used nine *E. multilocularis* newly obtained sequences (see above), published *M. corti* sequences and new *M. corti* isolated sequences (see above) as queries for local tblastx (Camacho et al., 2009) searches in the genomes of *E. multilocularis*, *Echinococcus granulosus*, *Taenia solium*, *Hymenolepis microstoma* and *Mesocostoides corti* (see Supplementary Table 2 for details). Blast hits were manually annotated using Artemis genome browser (Rutherford et al., 2000), correcting splice sites and completing partial hits by visual inspection. Automatic annotations of the genomes were used as well, selecting genes containing the CAP domain (InterPro code: IPR014044). In order to avoid biases in the initial blast results, more divergent sequences were selected and used as query for new tblastx searches in the genomes. Newly identified sequences were included in the analysis. The NCBI protein database was interrogated in order to look for additional proteins annotated containing this domain (Cestoda + “SCP_euk”) not identified in our analysis.

2.1.2. Other bilaterian predicted proteins

Annotated protein sequences were downloaded for *Schmidtea mediterranea*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Lottia gigantea*, *Helobdella robusta*, *Capitella teleta*, *Necator americanus* and *Schistosoma mansoni* (See Supplementary Table 2 for details). CAP domain containing proteins were identified following two approaches: i) local blastp (Camacho et al., 2009) searches using recently annotated cestode VAL proteins (see section 2.1.1) as queries, and ii) domain searches in the annotated proteins using InterPro or Pfam codes for CAP domain (IPR014044 or PF-18088, respectively). Sequences identified by either method were included in the analysis.

For classification purposes, only one representative member of each of the reported families in mouse genome was used, as described in Silvarrey et al. (2016).

2.2. Phylogenetic analysis

The CAP domain in each sequence was identified using the batch-CD

search at NCBI (Marchler-Bauer et al., 2015) and a custom Python script was used to keep only the sequence corresponding to the complete domain. Proteins containing two CAP domains were split and the suffix “_N” or “_C” was added to the sequence identifier depending on its location (N-terminal or C-terminal). Incomplete domains were not included in further analyses.

Selected sequences were aligned by means of MAFFT program (Katoh and Standley, 2013), using the E-INSI method with default parameters. For the alignment of *Platyhelminthes* sequences, 299 sites were used with 239 being variable (80%). For the alignment of Bilateral sequences, 566 sites were used with 363 variable sites (64%). ModelGenerator version 0.85 (Keane et al., 2006) was used to find the most appropriate amino acid model of evolution. Phylogenetic tree was inferred using the Maximum likelihood method by means of Phyml version 3.0 (Guindon et al., 2010). The default SH-like test was used to evaluate branch support. Trees were visualized and annotated using the iTOL web server (Letunic and Bork, 2016).

A phylogeny including truncated sequences and pseudogenes was also built. In this case only *M. corti* sequences with two or more exons were aligned with MAFFT and phylogenetically analyzed with Phyml as described above.

2.3. Sequence analysis

Translated sequences of complete genes, truncated genes, and pseudogenes were aligned as described in section 2.2. Sites containing gaps in > 90% of the sequences were removed. Position Entropy of this alignment was calculated using Bioedit (Hall, 1999) and a sliding window of nine sites was used for smoothing the plot. Sequence logo was generated using WebLogo 3.5.0 (Crooks et al., 2004), coloring by amino acid chemistry and plotting probability. Signal Peptide prediction was made using SignalP 4.0 (Petersen et al., 2011) with the default parameters.

For gene structure conservation analysis, exon number and length, splice sites, intronic phase and motif conservation were considered.

Sequences with more than two exons of the CAP superfamily in *M. corti* were aligned as described in Section 2.2. The retrieved aligned protein sequences were back-translated to the known DNA sequences by means of the tranalign program implemented in the EMBOS package (Rice et al., 2000). Amino acid and nucleotide distance matrices were calculated using MEGA-CC (Kumar et al., 2012) with the p-distance method and pairwise deletion for gap treatment.

2.4. Transcript expression validation in *M. corti*

To determine if the mRNAs corresponding to these putative proteins were expressed, we used unpublished transcriptomic data of the larval stage of *M. corti*. Reads were mapped to the draft genome using Tophat2 (Kim et al., 2013) with default parameters and without a reference annotation file. Annotation file of *M. corti val* genes was compared to the annotation of the draft genome made using maker software (Yandell and Ence, 2012)(unpublished results) using cuffcompare (Trapnell et al., 2012, 2010) with default parameters. The combined annotation file was used as input to cuffdiff (Trapnell et al., 2013, 2012), which was run with default parameters to obtain fpkm values of each *M. corti val* gene. Transcripts with fpkm value greater or equal to 1 were considered expressed.

For experimental validation, total RNA was extracted from larvae (2 days of culture) and segmented worms (10 days of culture) using Trizol following manufacturer instructions. 5 µg of RNA was DNase treated (DNase RQ1 - Promega) following manufacturer instructions. Specific primers for 9 genes were designed (Supplementary Table 1), spanning in all cases at least one intron in order to detect any possible amplification originating from contaminating gDNA. The primers were used for RT-PCR. cDNA was synthesized using Superscript II to retro-transcribe 300 ng of RNA with 500 ng of oligo-dT. For PCR, Top Taq

polymerase (Bioron) was used following manufacturer instructions. Cycling conditions included 32 cycles of: denaturation (95 °C, 30 s), annealing (53 °C, 30 s) and polymerization (72 °C, 60 s). Amplicons were analyzed by standard agarose gel electrophoresis.

2.5. Test for selection, dN/dS

In order to test whether positive selection drives the sequence evolution of genes that comprise the recent expansion in *M. corti*, a dN/dS test was applied. The ratio of non-synonymous changes per non-synonymous sites (dN) over synonymous changes per synonymous sites (dS) is used to identify sites or lineages that evolve under negative, neutral or positive selection regimes.

For computational reasons, the 230 putative functional *M. corti* genes analyzed within the cestode-specific clade were arbitrarily split in eleven monophyletic groups that were independently analyzed. Each cluster comprised between 10 and 28 genes as described in Supplementary Fig. 1. Protein sequences within each cluster were re-aligned and subsequently back translated to the known DNA sequence as described in section 2.3. Gaps were completely removed. The phylogram for each cluster was derived from the main phylogeny as subtree and used as user-defined tree in all tests.

The selection test was done using site and branch models by means of the Codeml program, included in the PAML4.8 package (Yang, 2007). The site models allow the dN/dS ratio to vary among codons (amino acid positions in the protein) according to a predefined model, that is nearly neutral (M1a), positive selection (M2a), M7 (beta distribution without positively selective sites) and M8 (beta distribution with positive selective sites) (Yang et al., 2000). Comparison for testing positive selection between models M1a and M2a and between M7 and M8 were done using the likelihood ratio test (Supplementary Table 3).

In the branch models no variation among sites is allowed, the substitution rates are inferred from the entire alignment, while dN/dS ratios in each branch of the phylogeny are fitted independently. The “free-ratios” model, independent ratio for each branch, was tested and compared to the one-ratio model.

The GA-Branch method, which is a branch model based method implemented in the HyPhy package, was also tested for each subtree in the datamonkey web server (Delpont et al., 2010). This method does not need a priori selection of branches to test or to recursively test one branch at a time, but rather mines the data for good-fitting models (for details see Kosakovsky Pond and Frost, 2005).

2.6. Prediction of three-dimensional protein structure

One sequence of each cluster analyzed for selection was subjected to protein modeling in the Swiss-Model server (Biasini et al., 2014). Human Glioma pathogenesis-related protein was selected as template, based in the quality of the generated models (3q2r.1.A, Asojo et al., 2011). Models were visualized with Swiss-Pdb viewer (Guex and Peitsch, 1997), coloring amino acids under selection according to their properties.

2.7. Repeat classification and frequency analysis

Repeats were extracted from WormBase parasite annotation file (WBPS7 version) and filtered to exclude tRNAs, rRNAs, snRNAs, low complexity regions and simple repeats. Remaining repeats were submitted to TE-class (Abrusán et al., 2009) to assign a transposable element category (DNA transposon, LTR, LINE, SINE or general Retrotransposon). Repeats were further filtered using a custom python script to exclude those that overlap with annotated genes and our new annotation of *val* genes.

A custom python script was used to get the number of TE elements located within 1000 base pairs upstream and downstream of 444 *val* genes (complete, truncated and pseudogenes included) and, for

comparison, of randomly chosen 444 non-*val* genes with genomic lengths between 1000 and 1400 base pairs. The number of bases upstream and downstream of each gene was set to 1000 in order to avoid gene overlapping, given the short intergenic sequences found in *M. corti*.

Mann-Whitney test was used to determine if the number of repeats around *val* and non-*val* genes is significantly different.

3. Results

3.1. *M. corti* has a highly duplicated CAP superfamily compared to other cestodes

Given the availability of published genomes for *E. multilocularis*, *E. granulosus*, *T. solium* and *H. microstoma* (Tsai et al., 2013), we searched for genes containing the conserved CAP domain by looking for the domain in the annotated proteins. Since genome annotation is still partial and fragmentary, a tblastx search was performed in order to identify wrongly annotated and/or unannotated sequences. On average 48 genes containing the CAP domain were found in these organisms, 26 of which were classified as complete genes (Table 1 and Supplementary Table 4). Most of them were included in the genome annotation, but several annotated genes were truncated. Missing exons were identified by tblastx and allowed us to complete these gene models (see details in Table 1). We also found several truncated genes, which may be linked to genome fragmentation due to partial assembly.

Unexpectedly, 444 putative genes coding for VAL proteins were found in the *M. corti* draft genome. A manual curation of these genes was done. Information from blast hits was integrated with automatic genome annotation, looking for conserved motifs that characterize the protein superfamily (See Supplementary Table 5 and Table 1). We found 271 complete protein coding genes, 101 truncated, mostly because they were too close to contig limits, and 74 pseudogenes, which have internal missing exons or in frame stop codons that should not have resulted from errors in the assembly (summarized in Table 1). Of the 444 *val* genes, only 44 were annotated in the WormBase Parasite database, 11 of those being complete gene models. Manual curation allowed us to complete 25 of the truncated gene models (Table 1). The estimation of pairwise nucleotide distances among all VAL coding sequences (Fig. 1A) shows that the mean distance is 0.522 ± 0.055 . Only 8 pairs of sequences (not shown) have a *p*-distance lower than 0.1, indicating that the high number of *val* genes is not the result of artifacts in the genome assembly. This result implies that the extraordinary gene expansion in the CAP superfamily may have occurred along the lineage leading to *M. corti* after its divergence from cyclophyllidean cestodes (see also the phylogenetic analyses in Section 3.2).

When analyzing sequence alignments the motifs CAP3 and CAP4, firstly described by Gibbs et al. (2008), resulted only moderately

conserved in *M. corti* (indicated with green and cyan boxes in Fig. 1B, respectively). On the other hand, a new motif, here named “PPA”, appeared as highly conserved (bordeaux box in Fig. 1B) in *M. corti* and other organisms including mouse (not shown). As expected, the motifs CRISP1 (Prosite code: PS01009) and CRISP2 (Prosite code: PS01010) that characterize the protein superfamily were highly conserved, as well as the cysteines and the Pro-Tyr dipeptide that separates the CAP domain from the cysteine rich “hinge region”. Furthermore, several individual amino acids are highly conserved, surrounded by more divergent amino acids (Val in position 39, between CAP3 and PPA; Tyr-Ser in positions 51–52, right before first exon junction; Glu in position 56; Gly in positions 137 and 176, the former next to a Cysteine and the latter between Pro-Tyr dipeptide and the hinge region).

Most of the coding sequences analyzed have a conserved gene structure similar to that of the mouse GLIPR1L gene. They contain four exons of relatively conserved length and conserved intron phases and positions. In detail, intron 1 is located three residues after the PPA motif, intron 2 is located in the middle of CRISP1 motif and intron 3 is located in the middle of CRISP2 motif (Fig. 1B and C). On the contrary, intron length is not conserved (Fig. 1C), showing greater variation than exonic length. This further supports the idea that these are different genes and not artifacts from genome assembly.

Repeat content flanking *val* and non-*val* genes was analyzed in order to determine if there is an enrichment of repeat regions in *val* genes that could be involved in duplication mechanism as described in *S. mansoni* by Philippsen et al. (2015). Despite finding several repeats within 1000 bp up and downstream *val* genes, the number is not significantly different from the number of repeats flanking non-*val* genes (*U* test, *p*-value > 0.05). The analysis of different categories of transposable elements was also analyzed, and did not show significant differences between *val* and non-*val* genes (not shown).

We also searched CAP superfamily members in *S. mansoni* and *S. mediterranea* proteomes in order to compare with other *Platyhelminthes*. Using blastp and CAP domain searches we found 65 complete domains in *S. mediterranea* and 27 complete domains in *S. mansoni*. Several truncated genes were also found in both organisms (see Table 2 and Supplementary Table 6 for details).

3.2. Phylogenetic analyses of CAP domain containing proteins

3.2.1. Most duplicated *M. corti val* genes are species-specific

Fig. 2 shows the phylogenetic analysis of the amino acid sequences corresponding to complete CAP domains of proteins from *Platyhelminthes*. Planarian and *S. mansoni* genes were included together with cestode genes in the analysis. Numerous monophyletic clusters of species- or lineage-specific genes can be identified in the phylogeny, suggesting that several independent expansions have occurred in flatworms. Note that almost all species or groups of species have

Table 1
Number of CAP domain containing proteins in cestodes.

	Automated Genome annotation ^a	Complete ^b in Genome annotation	Partial ^c in Genome annotation	Partial in Genome Annotation extended in this work. Now complete ^d	Novel Complete ^b identified in this work	Novel Partial ^c identified in this work	Pseudogenes identified in this work ^{e,f}	Total Complete	Total Sequences
<i>M. corti</i>	44	11	33	25	235	93	74 ^f	271	444
<i>E. granulosus</i>	17	14	3	0	7	3	4	21	31
<i>E. multilocularis</i>	30	21	9	4	11	28	8	36	77
<i>H. microstoma</i>	24	18	4	3	6	5	8 ^f	27	41
<i>T. solium</i>	32	16	16	5	0	8	3	21	43

^a Number of sequences containing CAP domain (IPR014044) in the annotation provided by WormBaseParasite (version 1).

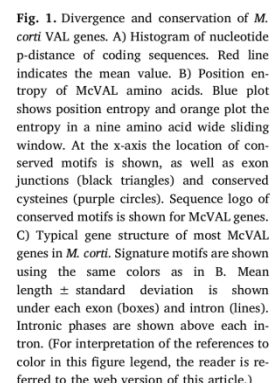
^b Number of complete sequences; at least four exon and Complete CAP domain.

^c Partial sequences, in which at least one terminal exon is missing.

^d Partial sequences in genome annotation (internal or terminal missing exons) completed in this work.

^e Sequences containing in frame stop codons or missing intermediate exons that could not be found based on homology.

^f : When indicated, two genes were partial in genome annotation but extended to pseudogenes.



presents an independently expanded monophyletic group of *val* genes (light green highlight, Fig. 2, 28 of 70 domains), from a subfamily of genes different to that of parasitic worms.

In order to analyze if independent expansions of val genes have also occurred in other bilaterian lineages, we performed blastp and domain searches in organisms with complete genomes, followed by phylogenetic analysis of these sequences in combination with all the previously analyzed flatworm sequences. [Table 2](#) shows a summary of the number of CAP domain containing proteins with both searching strategies (see Supplementary Table 6 for details).

The phylogenetic analyses revealed the presence of 3 major groups within the family (Fig. 3). In one group (green highlight) there are members of each organism analyzed. The only member of mouse proteins present in this group is Mm.GliPr2. In the second group (yellow highlight) there are only nematode members, with an extensive duplication in the parasitic nematode *N. americanus*. The third group (red

Table 2
Number of CAP domain containing proteins in other organisms.

	Domain only ^a		Blast only ^b		Both ^c		Total ^f	Used ^g
	Incomplete domain ^d	Complete domain ^e	Incomplete domain ^d	Complete domain ^e	Incomplete domain ^d	Complete domain ^e		
<i>C. elegans</i>	3	4	0	0	1	34	42	38
<i>C. teleta</i>	0	0	1	0	0	18	19	18
<i>D. melanogaster</i>	0	26	1	1	0	21	49	48
<i>H. robusta</i>	0	0	2	0	3	16	21	16
<i>L. gigantea</i>	0	0	0	0	1	16	17	16
<i>N. americanus</i>	3	43	56	3	6	33	144	79
<i>S. mansoni</i>	11	2	0	0	0	25	38	27
<i>S. mediterranea</i>	1	0	0	2	9	63	75	65

^a Number of sequences found only by domain searches (not blast).
^b Number of sequences found only by blast searches (not domain).
^c Number of sequences found only by both strategies.
^d Number of sequences containing incomplete CAP domain (NCBI CD-seach) and not used in the phylogenetic analysis.
^e Number of sequences containing complete CAP domain (NCBI CD-seach) and used in the phylogenetic analysis.
^f Total number of sequences containing the CAP domain (both complete and incomplete).
^g Number of sequences used in phylogenetic analysis (only complete domains).

highlight) includes several lineage-specific duplications in flatworms and other organisms. That is *M. corti*, *H. microstoma*, *D. melanogaster*, *S. mediterranea*, *S. mansoni*, the mollusk *Lottia gigantea* and the annelids *Helobdella robusta* and *Capitella teleta*. Phylogenetic pattern showed that cestodes selectively duplicated this group of genes, being *M. corti* the genome with more genes in the family, around 400, meaning 10 times or more than in other cestodes. Therefore, lineage-specific expansion of particular clades of CAP domain containing proteins has been a recurrent event in the evolution of bilaterian genomes, and is particularly conspicuous in the case of *M. corti*.

3.3. Several Mc-val duplicated genes are expressed in larvae and segmented worms

The very large number of genes encoding VAL proteins in *M. corti* raises the question of whether these are actually expressed. From transcriptomic data of the tetrathyridium larval stage of *M. corti*, we observed that approximately 50% of these genes are expressed (fpm > 1), although to different extents (Fig. 4A). Different mapping parameters, such as maximum mismatches allowed, did not change the frequency distribution (not shown). Of course, many of the other duplicated genes may be expressed in other developmental stages, but there is no current transcriptomic data available for other *M. corti* developmental stages. To overcome this caveat, we analyzed available

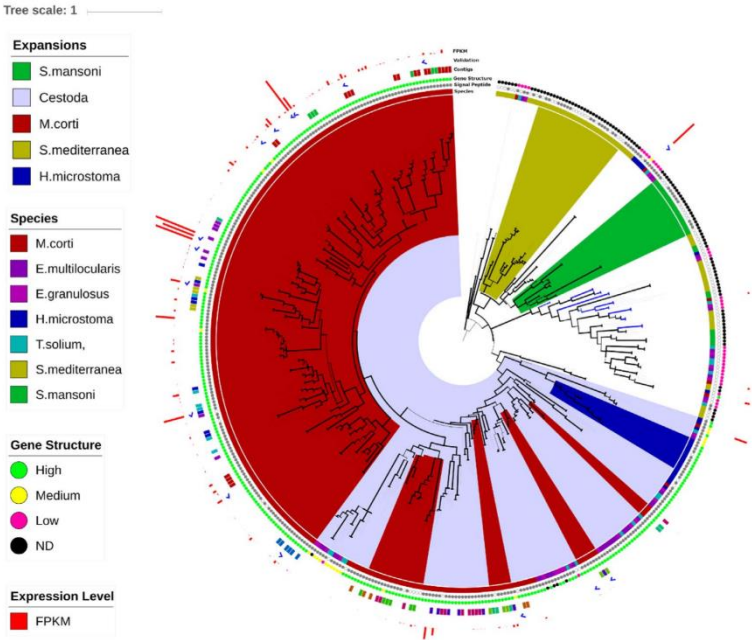


Fig. 2. Maximum likelihood phylogenetic analysis of Platyhelminth CAP domain amino acid sequences. The first row of squares indicates species described in the legend. The second row indicates presence (filled gray circle) or absence (unfilled gray circle) of N-terminal signal peptide. Third row indicates gene structure conservation described in the legend (according to Fig. 1C criteria). Fourth row of squares indicate *M. corti* scaffold clustering (same color, same scaffold). Check marks indicate genes verified by RT-PCR. The outer bar plots indicate expression level (fpm values). Highlighting marks monophyletic clades for *M. corti* (red), cestoda (light blue), *H. microstoma* (blue), *S. mediterranea* (light brown) and *S. mansoni* (green). The thickness of the lines of the tree is proportional to nodal support values (from 1 pixel (support 0) to 5 pixels (support 1)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

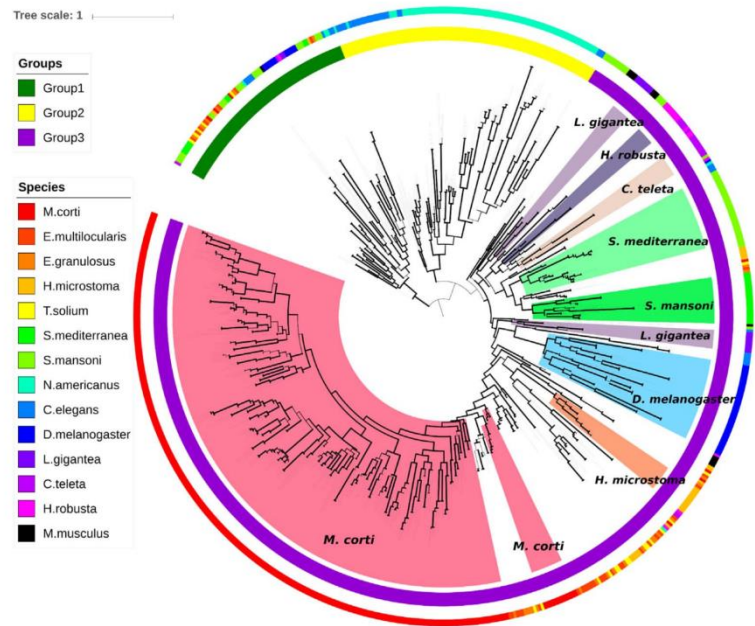


Fig. 3. Maximum likelihood phylogenetic analysis of Bilaterian CAP domain amino acids. Colored squares represent the species described in the legend. Groups of monophyletic duplications are highlighted. The thickness of the lines of the tree is proportional to nodal support values (from 1 pixel (support 0) to 5 pixels (support 1)).

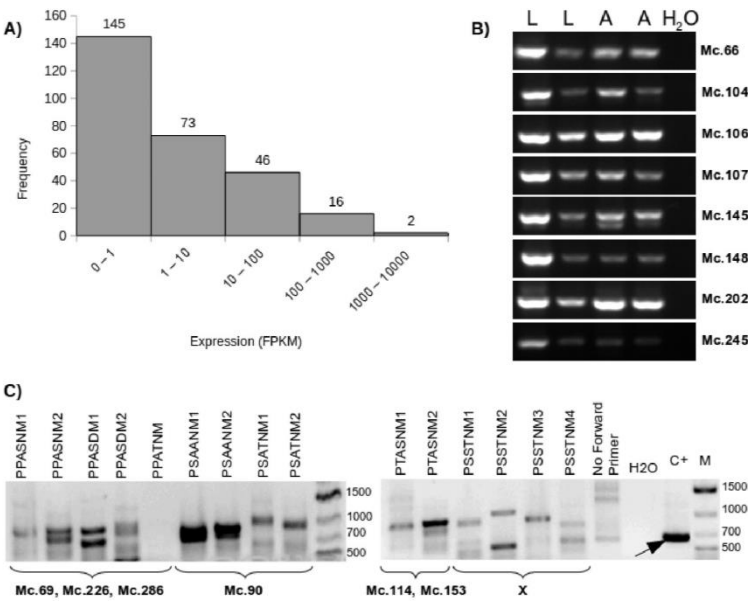


Fig. 4. Expression of McVAL genes. A) Histogram of fpkm values of larval transcriptomic data. B) Duplicates of RT-PCR experiments of selected McVAL genes in adults (A) and larvae (L). H₂O indicates NTC control. C) 3'RACE of McVAL genes using degenerated primers (indicated above each lane; Supplementary Table 1) for the most N-terminal conserved motif (named PPA in this work). VAL genes shown below the gel were identified by cloning and sequencing of pooled PCR products. The arrow indicates the band of approximately 700 bp obtained in the control using a plasmid containing the coding sequence of McVAL2 and a reverse specific primer.

transcriptomic data of other cestodes. The result suggests that these genes are indeed differentially expressed in different stages of development (Supplementary Table 7).

We also confirmed the expression of 14 *M. corti* genes in tetra-tyridia and/or in vitro cultured adult worms by RT-PCR using specific primers and 3'RACE experiments with degenerated primers for the new conserved motif named PPA (Fig. 4B and C). Taken together these results indicate that most of the complete *val* genes present in the *M. corti* specific expansion are expressed, which at the same time suggest that they are putative functional and certainly not pseudogenes.

However, some *M. corti val* genes were classified as pseudogenes because of the presence of in frame stop codons or missing internal exons. Only six of these 74 putative pseudogenes present transcriptomic evidence of expression in the larvae (not shown), and transcriptomic data of other developmental stages may help to the final annotation of these and other genes. Also, some genes were classified as truncated genes because of the missing of terminal exons. In the phylogenetic analysis of all *M. corti val* genes and pseudogenes, truncated sequences and pseudogenes are distributed rather uniformly in the tree (Supplementary Fig. 3), indicating that some genes in all duplicated clades tend to lose their function.

3.4. Positive natural selection drives the sequence evolution of some members of the CAP superfamily

Site and branch models of dN/dS test were used to determine if positive selection is acting on the evolution and divergence of the main monophyletic expansion of *M. corti val* genes. In the case of site model tests, the dN/dS ratio is estimated for each site and averaged among all branches in the tree, while in the branch model tests; the dN/dS ratio is averaged among all sites and allowed to vary among branches.

Site-model tests showed evidence for natural selection in 6 out of the 11 subtrees that comprise the main *M. corti* specific expansion (Supplementary Fig. 1), $LRT_{(M1a \text{ vs. } M2a)} p < 0.05$. At least one and up to 16 codon positions in each monophyletic subtree were identified to be evolving under positive selection ($Pr. > 50\%$, Table 3). In some cases, the same positions were identified among different subtrees. The impact of the amino acid changes in the structure was studied, but no clear link was found among the position under selection and the protein function. We did note that the general modeled structure of the conserved domain under study is similar, and most sites under selection reside in the surface of the protein (Supplementary Fig. 4).

The branch-model tests implemented in PAML showed that there is variation in the dN/dS test among branches for the majority of subtrees. The free-ratio model fit significantly better for subtrees 2, 3, 4, 5 and 11 (data not shown), indicating that in these cases the purely neutral model is significantly less probable than the model that includes variation among branches. On the other hand, results of the GA-branch tests indicated that a model including positive selection fit significantly better than the neutral model (a model that allows only neutral, nearly neutral and negative selected lineages) for 7 subtrees (Supplementary Fig. 5). Taken together, these results imply that positive selection may be driving the evolution at least in some lineages after duplication. Note that variation in the dN/dS ratio among branches may indicate an effect of positive selection restricted to a few sites.

In comparison, the same dN/dS sequence analyses were applied to the *Drosophila* CAP monophyletic cluster. Results indicate that purifying selection most probably drives the evolution of *Drosophila* CAPs protein coding genes. No site nor branch within this cluster were identified as evolving under positive selection (data not shown).

4. Discussion

Multigene families have different evolutionary patterns with varied evolutionary stabilities of their members. In the case of the CAP superfamily of proteins, our results strongly indicate a “birth and death”

Table 3
Sites under selection (Bayes Empirical Bayes (BEB) analysis) - M2a.

Subtree	Ref. Seq.	Position in the alignment	AA reference	Pr ($\omega > 1$)	Posterior mean	$\pm$ SE
1	Mc.2	44	D	0.999	5.372	2.109
		45	S	0.838	2.353	0.693
		59	D	0.502	1.773	0.802
		77	G	0.881	2.423	0.638
		84	Q	0.991	2.612	0.430
3	Mc.54	99	V	0.625	1.987	0.815
		109	E	0.905	2.471	0.609
		112	N	0.680	2.078	0.795
		113	G	0.798	2.285	0.733
		114	R	0.853	2.378	0.675
4	Mc.37	25	V	0.902	2.658	0.743
		65	V	0.998	2.864	0.563
		89	K	0.989	2.845	0.584
6	Mc.0	29	I	0.826	1.863	0.632
		42	L	0.594	1.562	0.613
		43	S	0.867	1.915	0.623
		45	T	0.707	1.700	0.631
		46	L	0.609	1.586	0.621
		50	L	0.565	1.526	0.601
		64	N	0.861	1.911	0.628
		75	A	0.792	1.824	0.640
		13	T	0.666	1.471	0.476
		17	N	0.519	1.314	0.376
8	Mc.8	25	V	0.510	1.311	0.379
		28	V	0.756	1.561	0.496
		30	L	0.774	1.586	0.508
		40	N	0.606	1.405	0.440
		46	L	0.586	1.388	0.433
		47	S	0.749	1.559	0.501
		48	K	0.906	1.730	0.527
		50	I	0.631	1.440	0.466
		51	Y	0.533	1.336	0.400
		52	A	0.801	1.619	0.521
11	Mc.68	55	S	0.631	1.434	0.458
		57	V	0.899	1.732	0.548
		58	E	0.553	1.343	0.393
		70	I	0.738	1.547	0.499
		1	M	0.878	3.118	0.883
		4	K	0.635	2.509	1.190
		23	N	0.872	3.103	0.897
		50	S	0.582	2.374	1.209
		59	G	0.997	3.419	0.450
		60	G	0.967	3.345	0.603
		63	E	0.999	3.424	0.435
		65	S	0.994	3.412	0.468
		66	G	0.879	3.124	0.881
		69	S	0.984	3.388	0.523
		73	E	0.551	2.290	1.206
		74	P	0.940	3.278	0.709
		75	L	0.963	3.333	0.622
		76	G	1.000	3.425	0.432
		77	K	0.503	2.176	1.207
		86	S	0.848	3.041	0.947

The positions in the alignments identified as evolving under positive selection (M2a model) are indicated for each subtree. A reference sequence (Ref. Seq.) and the amino acid in the position in the reference sequence (AA reference) are indicated. Pr ($\omega > 1$) indicates the posterior probability of each codon site of being under positive selection. This was calculated by means of the Bayes Empirical Bayes (BEB) method as recommended by PAML manual. Probability values higher than 99% and 95% were considered significant and are indicated in bold and italics, respectively. For each site in each subtree the estimated dN/dS (Posterior mean) and the Standard error (SE) is also shown.

model of evolution. According to this model new genes are created by frequent gene duplication, but only some duplicated genes are retained in the genome, whereas others are inactivated or deleted (Nei and Rooney, 2005). Therefore, different species tend to have different expansions of particular clades of each multigene family, which were selected for retention in the genome. Selective pressure from changing environmental variables has been suggested as the basis for the differential retention and loss of different family members (Nei and

Rooney, 2005). For example, this model has been found to describe the evolution of gene families that are involved with the immune system of vertebrates and in olfactory receptors in mammals (reviewed by Nei and Rooney, 2005). In the case of CAP domain containing proteins, some lineages (e.g. Group 1 in Fig. 3) are phylogenetically stable, that is, characterized by fewer duplications and with representatives in all major taxonomic groups, whereas other clades (e.g. Groups 2 and 3 in Fig. 3) show extreme variability in gene number, including independent duplications in different organisms. This indicates that not all members of the CAP superfamily are subject to the same selective pressures, and is reminiscent to the evolutionary pattern found for cytochrome P450 genes in vertebrates, where genes involved in the processing of endogenous ligands are phylogenetically stable, whereas genes involved in the processing of xenobiotics show high instability and independent duplications (Thomas, 2007). In *M. corti*, where we found the largest expansion of genes coding for CAP domain containing proteins, many of the expanded clades may have been subject to positive natural selection, and pseudogenes have been generated from most clades. These patterns suggest that there has been a constant process of gene duplication and selective retention of individual genes within each clade.

The results strongly suggest that positive selection may also be operating at sequence level, driving sequence evolution after duplication, on specific protein positions or clades. The relevance of the observed amino acid substitution among duplicated genes is not clear, however the fact that we are only considering the CAP conserved domain, which is the functional core of the family, may imply a considerable effect on protein function. Note that the functional core of a conserved gene is expected to evolve under strong negative selection, keeping amino acid substitutions rates at low or null levels.

Clearly, the phylogenetic patterns observed indicate that similar processes have independently occurred in different animals, both parasitic and free-living. Most CAP domain containing proteins contain the putative signal peptides and are presumably secreted, thus, it is likely that different external selective pressures have shaped the CAP superfamily gene complement in different animals. Most residues under positive selection are in the protein surface and they may interact with other molecules, thus varying their targets and effects. For example, in free-living animals, different CAP domain containing proteins have functions as components of venoms and in fertilization, two classic processes where selective pressures are likely to change constantly resulting in positive selection (Fry et al., 2003; Swanson et al., 2003; Swanson and Vacquier, 2002; Vacquier and Swanson, 2011). In the case of parasites, the pattern of “birth and death” evolution of particular *val* genes is suggestive of roles in the interaction between parasite and host. In nematodes, these genes are associated to immune modulation. The first protein identified containing the CAP domain, Ac-NIF from *Ancylostoma caninum*, can bind to integrins, blocking the adhesion of activated neutrophils to vascular endothelial cells (Moyle et al., 1994). Moreover, Ac-HPI, from the same worm, inhibits platelet aggregation (Del Valle et al., 2003) and Na-ASP2 from *Necator americanus* can recruit neutrophils in vitro and in vivo (Bower et al., 2008). Na-ASP2 also helps in the migration of the parasite through host tissues, facilitating the establishment of the larvae (Bethony et al., 2005; Goud et al., 2005). These proteins are generally secreted and glycosylated, and it is believed that may act as immune modulators, allowing the parasite to evade the host immune response during invasion (Hawdon and Hotez, 1996). Several of these nematode ASPs were used in vaccine testing, evaluating humoral and cellular immune response or the effect on challenging infections after immunization (Anand et al., 2007; Bethony et al., 2005, 2008; Diemert et al., 2012; Geldhof et al., 2008; Goud et al., 2005; MacDonald et al., 2004; Sen et al., 2000; Vlaminck et al., 2015; Zhan et al., 2012).

Parasite survival depends on the release of excretion – secretion antigens (ES) with immunomodulatory properties (Hewitson et al., 2009; McSorley et al., 2013; Pan et al., 2014). Immune response to helminth infections generally initiate with a pro-inflammatory Th1

response, gradually replaced by a Th2 response that allows host and parasite coexistence (Terrazas et al., 1998). Recently, Vendelova et al. (2015) analyzed mice immune response to *M. vogae* (syn. *M. corti*) infection. They showed that *M. corti* larvae effectively induce an early Th2 response, with no apparent previous Th1 response as seen in other cestodes. There are low levels of IFN γ and lymphocyte proliferation is suppressed 2 weeks post-infection. IL10 and IL4 are produced as early as 7 days post-infection. These events are earlier than in other cestodes (2 months for lymphocyte suppression and 3 weeks for IL10 and IL4 production) indicating that *M. corti* immune modulation is different to other cestodes. Later, Vendelova et al. (2016) proved that glycoproteins in ES products are responsible for immune evasion in *M. corti*, showing that they inhibit dendritic cell activation and consequent production of IL12, which in turn suppress Th1 response. Among the ES products they identified two VAL proteins (MCOS.000079201-mRNA, MCOS.0000907301). The first is the previously reported McCRISP3 (Britos et al., 2007), which is located within the big *M. corti* expansion (precisely in subtree 8 which is under selective pressure). The second is in a non-active fraction (Mc202, located in subtree 10 which is not under positive selective pressure). This is the first study that shows the presence of a VAL protein in cestode ES products. Previous reports did not find these proteins in cestode secretions (Bieñ et al., 2016; Virginio et al., 2012; Wang et al., 2015). As technology improves and genome information becomes available, low expressed or previously unknown genes can be detected. With our new improved manual annotation of VAL proteins, future proteomic studies may identify more VAL proteins in ES products, even in fractions with immunomodulatory activity. This in turn will help to better understand the roles of these intriguing proteins. *M. corti* is different to most cestodes since the tetrathyridia larvae can infect a wide variety of intermediate hosts (Grosbie et al., 2000; Hanson and Widmer, 1985; Specht and Voge, 1965). It is tempting to speculate that the large assortment of VAL proteins in this parasite could be related to this wide diversity of intermediate hosts.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.genrep.2018.03.010>.

Conflict of interest disclosure

The authors declare no conflict of interest associated with this study.

Acknowledgments

This work was supported by grants from: CSIC (Comisión Sectorial de Investigación Científica – Universidad de la República, Uruguay) and PEDECIBA (Uruguay). AC was recipient of a fellowship from ANII (Agencia Nacional de Investigación e Innovación, Uruguay).

References

- Abrusán, G., Grundmann, N., Demester, L., Makalowski, W., 2009. TEclass - a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25, 1329–1330. <http://dx.doi.org/10.1093/bioinformatics/btp084>.
- Anand, S.B., Gnanasekar, M., Thangadurai, M., Prabhu, P.R., Kaliraj, P., Ramaswamy, K., 2007. Immune response studies with *Wuchereria bancrofti* vespil allergen homologue (WbVAH) in human lymphatic filariasis. *DNA Seq.* 19, 981–988. <http://dx.doi.org/10.1007/s00436-007-0571-2>.
- Asojo, O. a, Koski, R. a, Bonafé, N., 2011. Structural studies of human glioma pathogenesis-related protein 1. *Acta Crystallogr. D Biol. Crystallogr.* 67, 847–855. <http://dx.doi.org/10.1107/S0907444911028198>.
- Barret, N.J., Smyth, J.D., Ong, S.J., 1982. Spontaneous sexual differentiation of *Mesocostoides corti* tetrathyridia in vitro. *Int. J. Parasitol.* 12, 315–322. [http://dx.doi.org/10.1016/0020-7519\(82\)90034-0](http://dx.doi.org/10.1016/0020-7519(82)90034-0).
- Bethony, J., Loukas, A., Smout, M., Brooker, S., Mendez, S., Plieskatt, J., Goud, G., Bottazzi, M.E., Zhan, B., Wang, Y., Williamson, A., Lustigman, S., Correa-Oliveira, R., Xiao, S., Hotez, P.J., 2005. Antibodies against a secreted protein from hookworm larvae reduce the intensity of hookworm infection in humans and vaccinated laboratory animals. *FASEB J.* 19, 1743–1745. <http://dx.doi.org/10.1096/fj.05-3936ffe>.
- Bethony, J.M., Simon, G., Diemert, D.J., Parenti, D., Desrosiers, A., Schuck, S., Fujiwara, R., Santiago, H., Hotez, P.J., 2008. Randomized, placebo-controlled, double-blind

- trial of the Na-ASP-2 hookworm vaccine in unexposed adults. *Vaccine* 26, 2408–2417. <http://dx.doi.org/10.1016/j.vaccine.2008.02.049>.
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, 252–258. <http://dx.doi.org/10.1093/nar/gku340>.
- Bieñ, J., Salamatin, R., Sulima, A., Savijoki, K., Conn, D.B., Näreaho, A., Młocicki, D., 2016. Mass spectrometry analysis of the excretory-secretory (E-S) products of the model cestode. *Acta Paediatr.* 61, 429–442. <http://dx.doi.org/10.1515/ap-2016-0058>.
- Bower, M.A., Constant, S.L., Mendez, S., 2008. *Necator americanus*: the Na-ASP-2 protein secreted by the infective larvae induces neutrophil recruitment in vivo and in vitro. *Exp. Parasitol.* 118, 569–575. <http://dx.doi.org/10.1016/j.exppara.2007.11.014>.
- Britos, L., Domínguez, L., Ehrlich, R., Marín, M., 2000. Effect of praziquantel on the strobilar development of *Mesocostoides corti* in vitro. *J. Helminthol.* 74, 295–299.
- Britos, L., Lalanne, A.I., Castillo, E., Cota, G., Schorale, M., Marín, M., 2007. *Mesocostoides corti* (syn. *Vogae*, Cestoda): characterization of genes encoding cysteine-rich secreted proteins (CRISP). *Exp. Parasitol.* 116, 95–102. <http://dx.doi.org/10.1016/j.exppara.2006.11.008>.
- Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P., Tiver, K.K., 1986. Molecular cloning of the cDNA for androgen-dependent sperm-coating glycoproteins secreted by the rat epididymis. *Eur. J. Biochem.* 161, 13–18. <http://dx.doi.org/10.1111/j.1432-1033.1986.tb0118.x>.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., Zhang, Z., Schäffer, A., Miller, W., Madden, T., Lipman, D., Koonin, E., Altschul, S., Schäffer, A., Wolf, Y., Ponting, C., Koonin, E., Aravind, L., Altschul, S., Schäffer, A., Aravind, L., Madden, T., Shavirin, S., Spouge, J., Wolf, Y., Koonin, E., Altschul, S., Waterston, R., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., Madden, T., Kent, W., Cameron, M., Williams, H., Cannane, A., States, D., Gish, W., Altschul, S., Morgulis, A., Gertz, E., Schäffer, A., Agarwala, R., Wootton, J., Federhen, S., Morgulis, A., Gertz, E., Schäffer, A., Agarwala, R., Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T., Agarwala, R., Schäffer, A., 2009. BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. <http://dx.doi.org/10.1186/1471-2105-10-421>.
- Cantacessi, C., Hofmann, A., Young, N.D., Broder, U., Hall, R.S., Loukas, A., Gasser, R.B., 2012. Insights into SCP/TAPS proteins of liver flukes based on large-scale bioinformatic analyses of sequence datasets. *PLoS One* 7, e31164. <http://dx.doi.org/10.1371/journal.pone.0031164>.
- Chalmers, I.W., Hoffmann, K.F., 2012. Platyhelminth Venom Allergen-Like (VAL) proteins: revealing structural diversity, class-specific features and biological associations across the phylum. *Parasitology* 139, 1231–1245. <http://dx.doi.org/10.1017/S0033182012000704>.
- Chalmers, I.W., McArdle, A.J., Coulson, R.M., Wagner, M. a, Schmid, R., Hirai, H., Hoffmann, K.F., 2008. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC Genomics* 9, 89. <http://dx.doi.org/10.1186/1471-2164-9-89>.
- Chen, J., Hu, X., He, S., Wang, L., Hu, D., Wang, X., Zheng, M., Yang, Y., Liang, C., Xu, J., Yu, X., 2010. Expression and immune response analysis of *Schistosoma japonicum* VAL-1, a homologue of vespid venom allergens. *Parasitol. Res.* 106, 1413–1418. <http://dx.doi.org/10.1007/s00436-010-1817-y>.
- Choudhary, V., Schneider, R., 2012. Pathogen-Related Yeast (PRY) proteins and members of the CAP superfamily are secreted sterol-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16882–16887. <http://dx.doi.org/10.1073/pnas.1209086109>.
- Cornelissen, B.J., Hooft van Huijsduijn, R.A.M., Van Loon, L.C., Bol, J.F., 1986. Molecular characterization of messenger RNAs for “pathogenesis related” proteins 1a, 1b and 1c, induced by TMV infection of tobacco. *EMBO J.* 5, 37–40.
- Costabile, A., Marín, M., Castillo, E., 2017. Spatio-temporal expression of *Mesocostoides corti* McVAL2 during strobilar development. *Exp. Parasitol.* 181, 30–39. <http://dx.doi.org/10.1016/j.exppara.2017.07.005>.
- Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. <http://dx.doi.org/10.1101/gr.849004>.
- Crosbie, P.R., Nadler, S.A., Platzer, E.G., Kerner, C., Mariaux, J., Boyce, W.M., 2000. Molecular systematics of *Mesocostoides* spp. (Cestoda: Mesocostoididae) from domestic dogs (*Canis familiaris*) and coyotes (*Canis latrans*). *J. Parasitol.* 86, 350–357.
- Datu, B.J.D., Gasser, R.B., Nagaraj, S.H., Ong, E.K., O'Donoghue, P., McInnes, R., Ranganathan, S., Loukas, A., 2008. Transcriptional changes in the hookworm, *Ancylostoma caninum*, during the transition from a free-living to a parasitic larva. *PLoS Negl. Trop. Dis.* 2, e130. <http://dx.doi.org/10.1371/journal.pntd.0000130>.
- Delport, W., Poon, A.F.Y., Frost, S.D.W., Kosakovsky Pond, S.L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26, 2455–2457. <http://dx.doi.org/10.1093/bioinformatics/btq429>.
- Diemert, D.J., Pinto, A.G., Freire, J., Jariwala, A., Santiago, H., Hamilton, R.G., Perigo, M.V., Loukas, A., Tribotet, L., Mulvenna, J., Correa-Oliveira, R., Hotez, P.J., Bethony, J.M., 2012. Generalized urticaria induced by the Na-ASP-2 hookworm vaccine: implications for the development of vaccines against helminths. *J. Allergy Clin. Immunol.* 130, 169–176. <http://dx.doi.org/10.1016/j.jaci.2012.04.027>.
- Fang, K.S., Vitale, M., Fehner, P., King, T.P., 1988. cDNA cloning and primary structure of a white-face hornet venom allergen. *Proc. Natl. Acad. Sci. U. S. A.* 85, 895–899. <http://dx.doi.org/10.1073/pnas.85.3.895>.
- Fry, B.G., Wüster, W., Kini, R.M., Brusich, V., Khan, A., Venkataraman, D., Rooney, A.P., 2003. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J. Mol. Evol.* 57, 110–129. <http://dx.doi.org/10.1007/s00239-003-2461-2>.
- Geldhof, P., Meyvis, Y., Vercruysee, J., Claerebout, E., 2008. Vaccine testing of a recombinant activation-associated secreted protein (ASP1) from *Ostertagia ostertagi*. *Parasite Immunol.* 30, 57–60. <http://dx.doi.org/10.1111/j.1365-3024.2007.01001.x>.
- Gibbs, G.M., Roelants, K., O'Bryan, M.K., 2008. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. *Endocr. Rev.* 29, 865–897. <http://dx.doi.org/10.1210/er.2008-0032>.
- Goud, G.N., Bottazzi, M.E., Zhan, B., Mendez, S., Deumic, V., Plieskatt, J., Liu, S., Wang, Y., Bueno, L., Fujiwara, R., Samuel, A., Ahn, S.Y., Solanki, M., Asajo, O.A., Wang, J., Bethony, J.M., Loukas, A., Roy, M., Hotez, P.J., Goud, G.N., Bottazzi, M.E., Ahn, S.Y., 2005. Expression of the *Necator americanus* hookworm larval antigen Na-ASP-2 in *Pichia pastoris* and purification of the recombinant protein for use in human clinical trials. *Vaccine* 23, 4754–4764. <http://dx.doi.org/10.1016/j.vaccine.2005.04.040>.
- Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-Pdb viewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723. <http://dx.doi.org/10.1002/elps.1150181505>.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. <http://dx.doi.org/10.1093/sysbio/syq010>.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hanson, G.B., Widmer, E.A., 1985. Asexual multiplication of tetrathyridia of *Mesocostoides corti* in *Crotalus viridis viridis*. *J. Wildl. Dis.* 21, 20–24.
- Hawdon, J.M., Hotez, P.J., 1996. Hookworm: developmental biology of the infectious process. *Curr. Opin. Genet. Dev.* 6, 618–623. [http://dx.doi.org/10.1016/S0959-437X\(96\)80092-X](http://dx.doi.org/10.1016/S0959-437X(96)80092-X).
- Hewitson, J.P., Grainger, J.R., Maizels, R.M., 2009. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. *Mol. Biochem. Parasitol.* 167, 1–11. <http://dx.doi.org/10.1016/j.molbiopara.2009.04.008>.
- Hwang, J.S., Ohyanagi, H., Hayakawa, S., Osato, N., Nishimiyama-Fujisawa, C., Ikey, K., David, C.N., Fujisawa, T., Gojobori, T., 2007. The evolutionary emergence of cell type-specific genes inferred from the gene expression analysis of Hydra. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14735–14740. <http://dx.doi.org/10.1073/pnas.0703331104>.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <http://dx.doi.org/10.1093/molbev/mst010>.
- Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., McInerney, J.O., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6, 29. <http://dx.doi.org/10.1186/1471-2148-6-29>.
- Kelleher, A., Darwiche, R., Rezende, W.C., Farias, L.P., Leite, L.C.C., Schneider, R., Asajo, O.A., 2014. *Schistosoma mansoni* venom allergen-like protein 4 (SmVAL4) is a novel lipid-binding SCP/TAPS protein that lacks the prototypical CAP motifs. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* 70, 2186–2196. <http://dx.doi.org/10.1107/S1399004714013315>.
- Kim, D., Pertege, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. <http://dx.doi.org/10.1186/gb-2013-14-4-r36>.
- Kosakovsky Pond, S.L., Frost, S.D.W., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. <http://dx.doi.org/10.1093/molbev/msi105>.
- Kozioł, U., Costabile, A., Domínguez, M.F., Iriarte, A., Alvite, G., Kun, A., Castillo, E., 2011. Developmental expression of high molecular weight tropomyosin isoforms in *Mesocostoides corti*. *Mol. Biochem. Parasitol.* 175, 181–191. <http://dx.doi.org/10.1016/j.molbiopara.2010.11.009>.
- Kumar, S., Stecher, G., Peterson, D., Tamura, K., 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* 28, 2685–2686. <http://dx.doi.org/10.1093/bioinformatics/bts507>.
- Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44 (W1), W242–W245. <http://dx.doi.org/10.1093/nar/gkw290>.
- Liu, S., Zhou, X., Hao, L., Piao, X., Hou, N., Chen, Q., 2017. Genome-wide transcriptome analysis reveals extensive alternative splicing events in the *Proteoscolex* of *Echinococcus granulosus* and *Echinococcus multilocularis*. *Front. Microbiol.* 8, 1–14. <http://dx.doi.org/10.3389/fmicb.2017.00929>.
- van Loon, L.C., van Strien, E.A., 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol. Mol. Plant Pathol.* 55, 85–97. <http://dx.doi.org/10.1006/pmpp.1999.0213>.
- MacDonald, A.J., Tawe, W., Leon, O., Cao, L., Liu, J., Oksov, Y., Abraham, D., Lustigman, S., 2004. Ov-ASP-1, the *Onchocerca volvulus* homologue of the activation associated secreted protein family is immunostimulatory and can induce protective anti-larval immunity. *Parasite Immunol.* 26, 53–62. <http://dx.doi.org/10.1111/j.0141-9838.2004.00685.x>.
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–6. <http://dx.doi.org/10.1093/nar/gku1221>.
- McSorley, H.J., Hewitson, J.P., Maizels, R.M., 2013. Immunomodulation by helminth parasites: defining mechanisms and mediators. *Int. J. Parasitol.* 43, 301–310. <http://dx.doi.org/10.1016/j.ijpara.2012.11.011>.
- Mizuki, N., Sarapata, D.E., Garcia-Sanz, J.A., Kasahara, M., 1992. The mouse male germ cell-specific gene Tpx-1: molecular structure, mode of expression in spermatogenesis, and sequence similarity to two non-mammalian genes. *Mamm. Genome* 3, 274–280.

- <http://dx.doi.org/10.1007/BF00292155>.
- Moyle, M., Foster, D.L., McGrath, D.E., Brown, S.M., Laroche, Y., De Meutter, J., Stanssens, P., Bogowitz, C. a, Fried, V. a, Ely, J. a, 1994. A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. *J. Biol. Chem.* 269, 10008–10015.
- Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39, 121–152. <http://dx.doi.org/10.1146/annurev.genet.39.073003.112240>.
- Ong, S.J., Smyth, J.D., 1986. Effects of some culture factors on sexual differentiation of *Mesostoides corti* grown from tetrathyridia in vitro. *Int. J. Parasitol.* 16, 361–368.
- Pan, W., Shen, Y., Han, X., Wang, Y.Y., Liu, H., Jiang, Y., Zhang, Y., Wang, Y.Y., Xu, Y., Cao, J., 2014. Transcriptome profiles of the protoscoleces of *Echinococcus granulosus* reveal that excretory-secretory products are essential to metabolic adaptation. *PLoS Negl. Trop. Dis.* 8, 1–15. <http://dx.doi.org/10.1371/journal.pntd.0003392>.
- Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. <http://dx.doi.org/10.1038/nmeth.1701>.
- Philippson, G.S., Wilson, R.A., DeMarco, R., 2015. Accelerated evolution of schistosome genes coding for proteins located at the host-parasite interface. *Genome Biol. Evol.* 7, 431–443. <http://dx.doi.org/10.1093/gbe/evu287>.
- Rice, P., Longden, I., Bleasby, A., Harper, R., Etzold, T., Argos, P., Stein, D.L., Altschul, S.F., Rice, P., et al., 2000. EMBOS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. [http://dx.doi.org/10.1016/S0168-9525\(00\)00204-2](http://dx.doi.org/10.1016/S0168-9525(00)00204-2).
- Roberts, K.P., Johnston, D.S., Nolan, M.A., Wooters, J.L., Waxmonsky, N.C., Piehl, L.B., Ensrud-Bowlin, K.M., Hamilton, D.W., 2007. Structure and function of epididymal protein cysteine-rich secretory protein-1. *Asian J. Androl.* 9, 508–514. <http://dx.doi.org/10.1111/j.1745-7262.2007.00318.x>.
- Rofatto, H.K., Parker-Manuel, S.J., Barbosa, T.C., Tararam, C. a, Alan Wilson, R., Leite, L.C.C., Farias, L.P., 2012. Tissue expression patterns of *Schistosoma mansoni* venom allergen-like proteins 6 and 7. *Int. J. Parasitol.* 42, 613–620. <http://dx.doi.org/10.1016/j.ijpara.2012.04.008>.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Ede Rajandream, M.-A., Barrell, B., 2000. Artemis: sequence visualization and annotation. *Bioinform. Appl.* 16, 944–945.
- Schuren, F.H., Asgeirsdóttir, S. a, Kothe, E.M., Scheer, J.M., Wessels, J.G., 1993. The Sc7/Sc14 gene family of *Schizophyllum commune* codes for extracellular proteins specifically expressed during fruit-body formation. *J. Gen. Microbiol.* 139, 2083–2090. <http://dx.doi.org/10.1099/00221287-139-9-2083>.
- Sen, L., Ghosh, K., Bin, Z., Qiang, S., Thompson, M.G., Hawdon, J.M., Koski, R. a, Shuhua, X., Hotez, P.J., 2000. Hookworm burden reductions in BALB/c mice vaccinated with recombinant *Ancylostoma* secreted proteins (ASPs) from *Ancylostoma duodenale*, *Ancylostoma caninum* and *Necator americanus*. *Vaccine* 18, 1096–1102. [http://dx.doi.org/10.1016/S0264-410X\(99\)00371-0](http://dx.doi.org/10.1016/S0264-410X(99)00371-0).
- Silvarrey, M.C., Echeverría, S., Costabile, A., Castillo, E., Paulino, M., Esteves, A., 2016. Identification of novel CAP superfamily protein members of *Echinococcus granulosus* protoscoleces. *Acta Trop.* 158, 59–67. <http://dx.doi.org/10.1016/j.actatropica.2016.02.011>.
- Specht, D., Voge, M., 1965. Asexual multiplication of *Mesostoides tetrathyridia* in laboratory animals. *J. Parasitol.* 51, 268–272.
- Swanson, W.J., Vacquier, V.D., 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3, 137–144. <http://dx.doi.org/10.1038/nrg733>.
- Swanson, W.J., Nielsen, R., Yang, Q., 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* 20, 18–20.
- Terrazas, L.I., Bojalil, R., Govezensky, T., Larralde, C., 1998. Shift from an early protective TH1-type immune response to a late permissive TH2-type response in murine *Cysticercosis* (*Taenia crassiceps*). *J. Parasitol.* 84, 74. <http://dx.doi.org/10.2307/3284533>.
- Thomas, J.H., 2007. Rapid birth-death evolution specific to xenobiotic cytochrome P450 genes in vertebrates. *PLoS Genet.* 3, 720–728. <http://dx.doi.org/10.1371/journal.pgen.0030067>.
- Thompson, R.C., Jue Sue, L.P., Buckley, S.J., 1982. In vitro development of the strobilar stage of *Mesostoides corti*. *Int. J. Parasitol.* 12, 303–314. [http://dx.doi.org/10.1016/0020-7519\(82\)90033-9](http://dx.doi.org/10.1016/0020-7519(82)90033-9).
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28. <http://dx.doi.org/10.1038/nbt.1621>.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. <http://dx.doi.org/10.1038/nprot.2012.016>.
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53. <http://dx.doi.org/10.1038/nbt.2450>.
- Tsai, L.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K.L., Tracey, A., Bobes, R.J., Frangos, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H.M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T. a, Deplazes, P., Estrada, K., Fernández, C., Holland, P.W.H., Hou, J., Hu, S., Huckvale, T., Hung, S.S., Kamenetzky, L., Keane, J. a, Kiss, F., Koziol, U., Lambert, O., Liu, K., Luo, X., Luo, Y., Macchiaroli, N., Nichol, S., Paps, J., Parkinson, J., Pouchkina-Stancheva, N., Riddiford, N., Rosenzvit, M., Salinas, G., Wasmuth, J.D., Zamanian, M., Zheng, Y., Cai, X., Soberón, X., Olson, P.D., Lalette, J.P., Brehm, K., Berriman, M., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496, 57–63. <http://dx.doi.org/10.1038/nature12031>.
- Vacquier, V.D., Swanson, W.J., 2011. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. *Cold Spring Harb. Perspect. Biol.* 3, a002931. <http://dx.doi.org/10.1101/cshperspect.a002931>.
- Del Valle, A., Jones, B.F., Harrison, L.M., Chadderdon, R.C., Cappello, M., 2003. Isolation and molecular cloning of a secreted hookworm platelet inhibitor from adult *Ancylostoma caninum*. *Mol. Biochem. Parasitol.* 129, 167–177. [http://dx.doi.org/10.1016/S0166-6851\(03\)00121-X](http://dx.doi.org/10.1016/S0166-6851(03)00121-X) [pii].
- Vendelova, E., Lutz, M.B., Hrková, G., 2015. Immunity and immune modulation elicited by the larval cestode *Mesostoides vogei* and its products. *Parasite Immunol.* 37, 493–504. <http://dx.doi.org/10.1111/pim.12216>.
- Vendelova, E., Camargo de Lima, J., Lorenzato, K.R., Monteiro, K.M., Mueller, T., Veepaschit, J., Grimm, C., Brehm, K., Hrková, G., Lutz, M.B., Ferreira, H.B., Nono, J.K., 2016. Proteomic analysis of excretory-secretory products of *Mesostoides corti* metacystodes reveals potential suppressors of dendritic cell functions. *PLoS Negl. Trop. Dis.* 10, 1–27. <http://dx.doi.org/10.1371/journal.pntd.0005061>.
- Virginio, V.G., Monteiro, K.M., Drumond, F., de Carvalho, M.O., Vargas, D.M., Zaha, A., Ferreira, H.B., 2012. Excretory/secretory products from in vitro-cultured *Echinococcus granulosus* protoscoleces. *Mol. Biochem. Parasitol.* 183, 15–22. <http://dx.doi.org/10.1016/j.molbiopara.2012.01.001>.
- Vlaminck, J., Borloo, J., Vercruyssen, J., Geldhof, P., Claerebout, E., 2015. Vaccination of calves against *Cooperia oncophora* with a double-domain activation-associated secreted protein reduces parasite egg output and pasture contamination. *Int. J. Parasitol.* 45, 209–213. <http://dx.doi.org/10.1016/j.ijpara.2014.11.001>.
- Voge, M., Seidel, J.S., 1968. Continuous growth in vitro of *Mesostoides* (Cestoda) from oncosphere to fully developed tetrathyridium. *J. Parasitol.* 54, 269–271.
- Wang, Y., Xiao, D., Shen, Y., Han, X., Zhao, F., Li, X., Wu, W., Zhou, H., Zhang, J., Cao, J., 2015. Proteomic analysis of the excretory/secretory products and antigenic proteins of *Echinococcus granulosus* adult worms from infected dogs. *BMC Vet. Res.* 11, 119–125. <http://dx.doi.org/10.1186/s12917-015-0423-8>.
- Xu, X., Francischetti, I.M.B., Lai, R., Ribeiro, J.M.C., Andersen, J.F., 2012. Structure of protein having inhibitory disintegrin and leukotriene scavenging functions contained in single domain. *J. Biol. Chem.* 287, 10967–10976. <http://dx.doi.org/10.1074/jbc.M112.340471>.
- Yandell, M., Ence, D., 2012. A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13, 329–342. <http://dx.doi.org/10.1038/nrg3174>.
- Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. <http://dx.doi.org/10.1093/molbev/msm088>.
- Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.-M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a003981>.
- Yeats, C., Bentley, S., Bateman, A., 2003. New knowledge from old: in silico discovery of novel protein domains in *Streptomyces coelicolor*. *BMC Microbiol.* 3, 3. <http://dx.doi.org/10.1186/1471-2180-3-3>.
- Yoshino, T.P., Brown, M., Wu, X.J., Jackson, C.J., Ocádiz-Ruiz, R., Chalmers, I.W., Kolb, M., Hokke, C.H., Hoffmann, K.F., 2014. Excreted/secreted *Schistosoma mansoni* venom allergen-like 9 (SmVAL9) modulates host extracellular matrix remodelling gene expression. *Int. J. Parasitol.* 44, 551–563. <http://dx.doi.org/10.1016/j.ijpara.2014.04.002>.
- Zhan, B., Santiago, H., Keegan, B., Gillespie, P., Xue, J., Bethony, J., de Oliveira, L.M., Jiang, D., Diemert, D., Xiao, S.-H., Jones, K., Feng, X., Hotez, P.J., Bottazzi, M.E., 2012. Fusion of Na-ASP-2 with human immunoglobulin Fc_γ abrogates histamine release from basophils sensitized with anti-Na-ASP-2 IgE. *Parasite Immunol.* 34, 404–411. <http://dx.doi.org/10.1111/j.1365-3024.2012.01371.x>.