



Tesina para optar por el grado de
Licenciado en Bioquímica

Proteínas de secreción y su rol adaptativo en el phylum Platyhelminthes: aproximación desde la genómica comparativa.

Mauricio Langleib de Souza

Orientador

Dr. Andrés Iriarte

Laboratorio de Biología Computacional
Departamento de Desarrollo Biotecnológico
Instituto de Higiene
Facultad de Medicina
Universidad de la República

Junio de 2018

*El que abandona...
¡no tiene premio!*

Patricio Rey y sus Redonditos de Ricota

Agradecimientos

A mi familia, que me bancó en el divague de querer ver la vida a través del lente de la ciencia, y que es la que siempre paga los platos rotos. En particular, a mis viejos Nancy y Marcelo, y a mi hermana Emilia: los amo.

A mis amigos y amigas, que también son familia y son simplemente lo más grande del Mundo. Es probable que el placer más grande que me dió esta carrera sea ganar armas para pasar días y noches filosofando sobre bueyes perdidos con ustedes.

A todos mis compañeros del Departamento de Desarrollo Biotecnológico, y en especial a la barra de mi querido Laboratorio de Biología Computacional. Con gente como ustedes da gusto trabajar.

A la vieja y querida Facultad de Ciencias, que es y va a seguir siendo una segunda casa para mí. Ojalá todos pudieran saber lo lindo que es entender una idea elegante en tus clases, o tirarse a la bartola en tus pastos a ver como pasa la vida entre gente querida.

A Pablo, que se tomó el trabajo de leer mi laburo. Por si fuera poco, fue la primer persona que me habló de genómica en una época en la que no sabía qué rumbo tomar.

A Andrés, por ser un gran tutor y mejor persona. Siempre me diste la oportunidad de debatir y laburar en libertad, cosa que lamentablemente no es la norma en estos días. Por eso y muchas cosas más te estoy eternamente agradecido. ¡Salú y a por esa segunda ronda!

Por último, nobleza obliga, quiero agradecer **a la ANII**. El financiamiento del proyecto "*Evolución de familias multigénicas codificantes para proteínas de secreción en el phylum Platyhelminthes*" (en el cual se enmarca mi tesina de grado) fue y es esencial. Ojalá el futuro venga acompañado de un aumento en la financiación de los proyectos que nacen de la pasión que tienen los integrantes de la hermosa comunidad científica de este país.

Índice

Agradecimientos

Resumen	1
Introducción	2
Aspectos generales del filo Platyhelminthes	2
Proteínas de secreción y su función	2
La duplicación génica: una fuente de materia prima para la evolución	3
Hipótesis de Trabajo	6
Objetivos	6
Materiales y Métodos	7
Obtención de Secuencias	7
Procesamiento de secuencias codificantes para proteínas de <i>Schmidtea mediterranea</i>	7
Predicción de secuencias codificantes para proteínas de secreción.	8
Detección de secuencias codificantes para presuntas proteínas secretadas por vía canónica de secreción.	8
Detección de secuencias codificantes para presuntas proteínas secretadas por vías alternativas de secreción.	9
Predicción de secuencias destinadas a residir en mitocondria	9
Análisis filogenético	10
Agrupamiento de secuencias en <i>clusters</i> de homólogos	10
Alineamiento de secuencias	10
Inferencia de árboles filogenéticos	10
Detección de grupos de secuencias inparálogas	11
Asignación de términos funcionales	13
Análisis de enriquecimiento	13
Análisis de distribución en isócoros de secuencias codificantes en el genoma de <i>S. mansoni</i>	14
Análisis de evolución molecular	14
Resultados	16
Obtención de clusters de homólogos	16
Mapeo de genes inparálogos a isócoros del genoma del tremátodo <i>Schistosoma mansoni</i>	21
Análisis de enriquecimiento en términos funcionales	23
Detección de genes inparálogos evolucionando en régimen de selección natural positiva	25
Discusión	29
La duplicación de genes codificantes para proteínas de secreción tendría un rol adaptativo en el filo Platyhelminthes	29
Análisis de la evolución de inparálogos a la luz de algunos modelos	30
Fuentes de error	31
Distribución de inparálogos en isócoros del genoma de <i>Schistosoma mansoni</i> : una hipótesis mecanística	33
Parásitos y organismos de vida libre muestran adaptaciones a nivel de su secretoma	34
Perspectivas	34
Conclusión	35
Bibliografía	36
Anexo	41
Correlación entre valores de FDR e indicadores N90 y N50	41
Correlación entre porcentaje del proteoma perteneciente al secretoma y porcentaje de secuencias descartadas del secretoma	43
Análisis de enriquecimiento en términos funcionales	43
Detección de genes inparálogos evolucionando en régimen de selección natural positiva	59

Resumen

Los platelmintos son invertebrados que presentan simetría biletal y cuerpo no segmentado. Dentro de este filo se encuentran tanto organismos de vida libre como parásitos de ciclos de vida complejo capaces de infectar a un amplio rango de hospederos, entre los que se encuentran animales de relevancia económica y humanos. Parte del éxito que poseen estos parásitos reside en la capacidad de modulación y evasión de la respuesta inmune del hospedero, en la cual juegan un rol preponderante proteínas de secreción que son producidas en las diferentes fases del ciclo de vida de estos organismos. Algunos de los genes codificantes para estas proteínas de secreción pertenecen a grandes familias multigénicas, las cuales han sido generadas por procesos de duplicación génica. No obstante, el rol que juega la selección natural positiva en la evolución de estas familias multigénicas no ha sido estudiado en gran profundidad. La publicación de un número considerable de genomas de especies pertenecientes a las cuatro clases de este filo permiten abordar esta problemática desde la genómica comparativa. Esta tesina de grado se centra en el estudio de duplicaciones linaje-específicas de genes codificantes para proteínas de secreción, teniendo como objetivo evaluar el rol adaptativo que poseen las mismas en los organismos del filo Platyhelminthes.

Introducción

Aspectos generales del filo Platyhelminthes

El filo Platyhelminthes (del griego, *platys* = plano + *helmins* = gusano) está compuesto por organismos metazoarios aplanados a lo largo del eje dorsoventral y con simetría bilateral (Hickman et al., 2007).

Dentro de la filogenia de los animales, los platelmintos se ubican dentro de Lophotrochozoa, un clado integrado por otros invertebrados como los anélidos (gusanos segmentados) y los moluscos (Dunn, Giribet, Edgecombe, & Hejnol, 2014). La ausencia de estructuras rígidas como un exoesqueleto o un caparazón permiten en este inmenso filo una enorme diversidad de planes de desarrollo, morfologías y tamaños (Collins, 2017).

Los platelmintos son a su vez agrupados en cuatro clases: Turbellaria, Trematoda, Monogenea y Cestoda (Hickman et al., 2007). La primera de estas está compuesta por organismos de vida libre, simbioses y parásitos. La enorme mayoría de los mismos habitan nichos marinos y de agua dulce, aunque también existen excepciones que pueblan territorios terrestres húmedos. Algunas de las especies del mismo, como *Schmidtea mediterranea* son organismos modelo ampliamente utilizados en biología molecular para el estudio en campos tan diversos como lo son la biología del desarrollo (Elliott & Sánchez Alvarado, 2013) (donde son de especial interés las células pluripotenciales) o las neurociencias (gracias a la presencia de un sistema nervioso capaz de llevar a cabo tareas relativamente complejas) (Buttarelli, Pellicano, & Pontieri, 2008).

Las otras tres clases componen el clado Neodermata, siendo las mismas compuestas en su totalidad por organismos parásitos. Los taxa son agrupados en este clado en base a la presencia de una sinapomorfia, a saber la presencia de un sincitio denominado *tegumento*, el cual recubre a estos organismos (Hickman et al., 2007). El tegumento está compuesto por múltiples proteínas, las cuales no sólo protegen al organismo parásito del sistema inmune del huésped, si no que también constituyen una interfaz a través de la cual el organismo capta nutrientes y excreta diferentes productos al medio (Dalton, Skelly, & Halton, 2004).

Los platelmintos parásitos se caracterizan por sus ciclos de vida complejos, los cuales pueden involucrar a múltiples huéspedes intermedios. El rango de huéspedes es amplio e involucra tanto a vertebrados como a invertebrados, lo que lleva a que estos parásitos se vean implicados en enfermedades de relevancia económica al afectar a organismos marinos, ganado y cánidos, entre otros. Al ser el humano un huésped intermedio en muchos de estos ciclos, enfermedades como la hidatidosis, la esquistosomiasis y la cisticercosis también tienen su origen en la infección con estos platelmintos (Muller et al., 1990).

Se presume que la presencia de tegumento no constituye una homoplasia, sino que es un verdadero carácter homólogo que denota el hecho que los organismos pertenecientes a estas tres clases poseen un único ancestro común (Littlewood, 2006). Teniendo en cuenta que los neodermata son capaces de parasitar virtualmente todo vertebrado descrito, se suele asegurar que el clado constituye la radiación adaptativa más exitosa del parasitismo (Littlewood, 2006).

Proteínas de secreción y su función

Múltiples estudios han demostrado la importancia que poseen las proteínas de secreción para el correcto desarrollo y sobrevivencia de organismos de diferentes filos (Joly, Feau, Tanguay, & Hamelin, 2009; Kaphingst, Persky, & Lachance, 2010; Karn, Clark, Nguyen, & Swanson, 2008).

Las mismas abarcan funciones sumamente diversas, incluyendo la señalización química, la degradación de materia orgánica, la regulación de procesos metabólicos y vinculados al desarrollo y la excreción de material excedente. Acorde a esto, las mismas son tanto enzimas como hormonas, toxinas y péptidos antimicrobianos (Alberts, Melrose, Perroy, & Careas, 2015).

Desde un punto de vista mecánico, dos tipos de proteínas de secreción son distinguibles: aquellas que son excretadas por la denominada *vía canónica de secreción*, y aquellas que lo hacen por vías alternativas. En el primero de los casos, las proteínas poseen en su extremo N-terminal un conjunto de aminoácidos usualmente denominado como *péptido señal*, el cual es reconocido por la proteína SRP (Akopian, Shen, Zhang, & Shan, 2013). La interacción SRP-péptido señal lleva a la finalización del proceso de traducción de estas proteínas, a al vez que asegura la traslocación del complejo ARNm-ribosoma al interior del retículo endoplasmático, organelo central en el proceso de excreción de aquellos productos destinados tanto a membranas como a el medio extracelular (Alberts, Melrose, Perroy, & Careas, 2015).

No es de extrañar que la presencia conservada de este péptido señal haya sido utilizada como marcador al momento de generar herramientas bioinformáticas destinadas a predecir la probabilidad de que una proteína sea secretada (Nielsen, Henrik, Engelbrecht, Brunak, & von Heijne, 1997). No obstante, es importante tener en cuenta que el carácter hidrofóbico de este péptido señal complejiza la tarea de detección de este tipo de proteínas debido a su similitud desde el punto de vista fisicoquímico con aquellos tramos de la estructura de proteínas transmembrana que permiten el anclaje en la bicapa lipídica de

las células. A su vez, debido a que el pasaje por el retículo endoplasmático también es una característica de aquellas proteínas destinadas a la presencia en membranas de organelos como la mitocondria, especial cuidado se debe tener en discriminar entre estos casos y aquellos en los que efectivamente el producto analizado está destinado al medio extracelular (Emanuelsson, Nielsen, Brunak, & von Heijne, 2000).

Empero todo lo dicho en las líneas anteriores, existen proteínas excretadas que utilizan vías alternativas de secreción, abarcando las mismas diversos mecanismos (*e.g.* la excreción mediante el denominado Complejo de Secreción No Convencional; ver Rabouille, C., Malhotra, V., & Nickel, W., 2012). La variedad de mecanismos implicados lleva a que no exista un único patrón conservado en las secuencias de estas proteínas, por lo cual desde un punto de vista bioinformático se suelen usar algoritmos que tengan en cuenta predictores de propiedades fisicoquímicas que puedan delatar adaptaciones al medio extracelular (ver sección Materiales y Métodos, donde se describe el uso de redes neuronales artificiales en este contexto).

El rol de estas proteínas en el desarrollo de la vida de los parásitos helmintos ha sido estudiado en profundidad, con especial énfasis en aquellos casos donde existe una fuerte respuesta del sistema inmune del hospedero. A su vez, es sabido que los parásitos de este phylum son capaces de modular la respuesta inmune de sus huéspedes a través de la acción de proteínas de secreción, logrando así una eficaz evasión de la misma (Ito, Lightowers, Rickard, & Mitchell, 1988). Algunos ejemplos de esto son la proteína EgAgB (Sánchez et al., 1991; Silva-Álvarez et al., 2015) y Eg95 (Chow, Gauci, Cowman, & Lightowers, 2001; Gauci, Merli, & Muller, 2002) del céstode *Echinococcus granulosus*, SmVal del tremátodo *Schistosoma mansoni* (Curwen, 2006; Hansell et al., 2008) o las catepsinas secretadas por el también tremátodo *Fasciola hepatica* (Cancela et al., 2008). Esta modulación de la respuesta inmune a través de interacciones específicas de las proteínas de secreción con sus respectivos blancos posee un rol central en la persistencia del parásito en su huésped.

La duplicación génica: una fuente de materia prima para la evolución

Se dice que un grupo de genes constituyen una familia multigénica cuando son homólogos y conservan una similitud de secuencia y un núcleo funcional (Dayhoff, 1978). Esta definición aplica tanto a genes homólogos existentes entre diferentes especies (*ortólogos*) como a copias originadas por un evento de duplicación en un mismo genoma (*parálogos*).

A las secuencias originadas mediante eventos de duplicación linaje-específicas se les denomina *inparálogos* (Koonin, 2005. Ver **Figura 1** donde se detalla el uso de este y otros términos). La conservación de las mismas en un linaje dado puede denotar presiones selectivas ejercidas sobre el mismo (ver Demuth y Hahn, 2009).

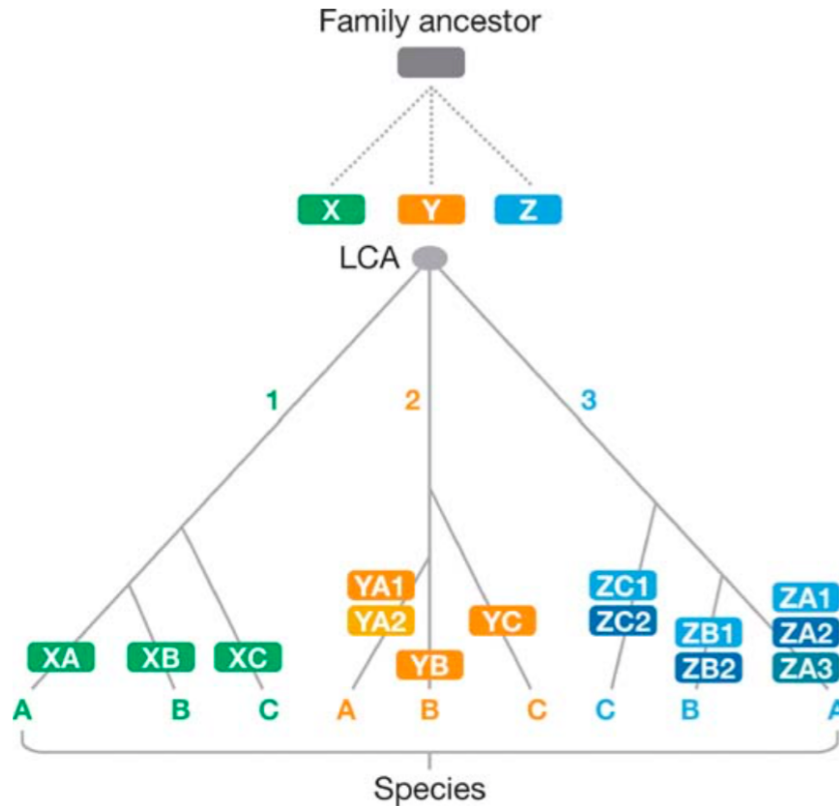


Figura 1. Relaciones de homología. Se consideran tres genes, X, Y y Z, originados a partir de una duplicación de un gen ancestral. A su vez se tienen en cuenta las especies A, B y C. Se dice que las secuencias YA1 e YA2 son inparálogos. Lo mismo ocurre con las secuencias ZA[1,2,3], ZB[1,2] o ZC[1,2], en cuanto las mismas son duplicaciones específicas de un linaje. Las secuencias XA y XB son, en cambio, ortólogas: las mismas son variantes originadas por descendencia vertical, luego de un evento de especiación. Por último, cualquiera de las combinaciones de X, Y y Z (*e.g* XA1 y ZC2) definen secuencias parálogos, ya que las mismas se originaron a través de una duplicación anterior a la especiación que dió lugar a las especies A, B y C. Tomado de Koonin (2005)

La idea de que los eventos de duplicación génica constituyen una de las principales fuentes de variación a nivel molecular en eucariotas goza de amplia aceptación. Diversos mecanismos como la recombinación no homóloga (productora de duplicaciones en *tandem*), inserción mediada por retrotransposición o la reparación en caso de ruptura de la doble hélice de ADN son fuente de este tipo de mutaciones, cuya frecuencia es de orden comparable a las mutaciones nucleotídicas más comunes (ver Lynch, 2002).

Ya en 1970 Susumu Ohno, en su libro *Evolution by Gene Duplication* (Ohno, 1970), planteaba que un evento de duplicación génica acompañado por un proceso de divergencia evolutiva, bajo un régimen de selección natural positiva, podría llevar a la creación de nuevas funciones moleculares. La lógica detrás de este postulado es sencilla: la exploración del espacio secuencial por parte de un gen se ve restringido por el impacto negativo que puede producir en el *fitness* de su organismo portador un cambio que afecte radicalmente su función; un evento de duplicación presenta en este contexto la posibilidad de relajar la restricción impuesta por este régimen de selección purificadora. Una vez ocurre el evento de duplicación, y siempre y cuando el genoma del organismo continúe codificando para una copia funcional del gen en cuestión (si no existen un efecto negativo del dosaje), una de las secuencias podrá recorrer el espacio secuencial. Generalmente esto deviene en pseudogenización, pero eventualmente se crearán nuevas funciones moleculares que posean un efecto positivo sobre el *fitness* del organismo portador.

Naturalmente, cabe preguntarse: ¿qué tan plausible es ésta hipótesis? ¿Cómo se da éste proceso, y cómo se describe en términos de régimen selectivo las diversas fases del mismo?

Múltiples modelos de evolución por duplicación génica han sido planteados para responder estas y otras interrogantes. A continuación se repasan tres modelos sencillos, con el fin de mostrar algunas de las hipótesis planteadas desde 1970 hasta la fecha: el modelo de neofuncionalización, el modelo de subfuncionalización y el modelo radiación adaptativa (Demuth & Hahn, 2009; Force et al., 1999; Francino, 2005).

i) Modelo de neofuncionalización

El modelo de neofuncionalización toma como punto de partida un gen codificante, el cual se ve duplicado. En este contexto, cabe asumir que ambas secuencias codificantes acumularán mutaciones en un proceso que se podría describir bajo un régimen de selección purificadora: se acumularán mutaciones al azar, algunas de las cuáles son sinónimas y otras no sinónimas. No obstante, la función que cumple la secuencia codificada por dichos genes debe mantenerse, de tal forma que no es posible que ambas copias exploren a la vez el espacio secuencial por un tiempo prolongado.

La duplicación generada se podrá fijar en la población, ya sea por un proceso de deriva génica como por una ventaja selectiva (producto, por ejemplo, de un efecto de dosage).

Si alguna de las copias acumula un número suficiente de cambios, mientras la otra continúa codificando una secuencia de tal forma que el *fitness* del organismo portador no se vea afectado, sucederá que aquella copia que ha acumulado un mayor número de cambios se verá ahora bajo un régimen selectivo relajado.

Con mayor frecuencia las mutaciones que acumulará llevarán a la pseudogenización. No obstante, en una fracción de secuencias puede ocurrir que la modificación en la secuencia codificada confiera una ventaja selectiva al organismo portador, debido a que se ha generado una innovación desde el punto de vista molecular. La misma estará ahora en condiciones de fijarse en la población, en un proceso que podemos describir como de selección natural positiva.

ii) Modelo de subfuncionalización

El punto de partida del modelo de subfuncionalización es una secuencia codificante donde existen al menos dos módulos funcionales diferentes. Este tipo de estructuras son particularmente comunes en eucariotas, donde este tipo de arquitectura abarca todos los taxa del Dominio. Se asume a su vez que los dominios ejercen su función de forma subóptima: debido a que los mismos interactúan a nivel molecular por pertenecer a una misma estructura, no es posible explorar los espacios secuenciales y optimizar sus estructuras sin que esto lleve a cambios pleiotrópicos.

Así, si existiera un evento de duplicación génica, las secuencias acumularán cambios de forma azarosa en ambos dominios. La pérdida de función debido a estas mutaciones en uno de los dominios de una de las copias no afectará el *fitness* del organismo portador, pero si fijará bajo un régimen de selección purificadora al dominio codificado en la otra copia génica.

Si el proceso sucede de forma simétrica para el otro dominio, se llegará a la situación en la que ambas funciones moleculares son ahora ejercidas por dos copias diferentes. Siendo este el caso, las secuencias podrán ahora eventualmente optimizar sus respectivas funciones. Esto no implica, al menos en principio, una innovación desde el punto de vista funcional.

iii) Modelo de radiación adaptativa

El modelo de radiación adaptativa (Francino et al., 2005) busca explicar, desde una perspectiva que adapta algunas nociones de ecología al ámbito molecular, dos aspectos teóricos: en primer lugar, se busca dar una explicación plausible a la fijación de las duplicaciones génicas a nivel poblacional, así como dar una descripción del régimen selectivo que rige esta fijación. En segundo lugar, se busca dar una explicación al surgimiento de innovaciones moleculares.

La premisa principal del modelo es que las innovaciones moleculares a partir de duplicaciones génicas surgen principalmente cuando ocurre la aparición de un *nicho molecular*, el cual es ocupado de forma subóptima por un gen presente en un organismo de una población.

Pongamos un ejemplo para clarificar este concepto: supongamos que un organismo posee en su genoma un gen codificante para un transportador de aminoácidos. Si la especie a la que pertenece el organismo habitara nuevas zonas geográficas, lo cual llevara a un cambio en la ingesta de los individuos, una nueva variedad de metabolitos podrán ingresar a dicho organismo. En caso de que el transportador de aminoácidos ya codificado en el genoma de la especie pueda interactuar con alguno de esos metabolitos, esta interacción se dará y dicho metabolito será capaz de ingresar a las células de este organismo. Siendo este el caso, y si esto confiere una ventaja selectiva a los organismos con esta capacidad, claramente *la duplicación génica en sí* evolucionará bajo un régimen de selección natural positiva. Una vez que se fije en la población dichas amplificaciones, se verá en un escenario plausible de que dichas secuencias evolucionen de forma similar a como se plantea en el modelo de neofuncionalización.

La ventaja principal de este modelo radica por tanto en plantear un mecanismo que da una mayor plausibilidad a las primeras etapas de la evolución hacia la generación de innovaciones moleculares, gracias a que pone el énfasis en la posibilidad de que la duplicación génica en sí sea seleccionada en alguna de las fases del proceso.

Hipótesis de Trabajo

- i)* Las duplicaciones génicas son un evento frecuente, a partir de este proceso se constituye un material sobre el cual la selección natural opera dirigiendo al evolución de las secuencias.
- ii)* El aumento en el número de copias genes codificantes para proteínas de secreción posee un rol adaptativo en el phylum Platyhelminthes.

Objetivos

- i)* Identificar proteínas de secreción que podrían tener un rol en la adaptación de estos parásitos a sus nichos específicos.
- ii)* Estudiar el papel de la selección natural en la evolución de las secuencias codificantes para proteínas de secreción o en el aumento de su número de copias en el genoma.

Materiales y Métodos

Obtención de Secuencias

Treinta secuencias de genomas correspondientes a especies de las cuatro clases que componen el *phylum* Platyhelminthes fueron obtenidas de las bases de datos WormBase ParaSite v9.0 (<http://parasite.wormbase.org/index.html>) y PlanMine v2.0beta (<http://planmine.mpi-cbg.de/planmine/begin.do>). Cuando estuvieren disponibles, también se descargó las secuencias proteicas codificadas en los mismos. El único caso en el cuál se operó de forma diferente en lo que a esto refiere fue en el caso de *S. mediterranea* (ver subsección siguiente).

Tabla 1. Lista de organismos empleados en este estudio.

Especie	Clase	Base de datos
<i>Diphyllobothrium latum</i>	Cestoda	WormBase ParaSite
<i>Echinococcus canadensis</i>	Cestoda	WormBase ParaSite
<i>Echinococcus granulosus</i>	Cestoda	WormBase ParaSite
<i>Echinococcus multilocularis</i>	Cestoda	WormBase ParaSite
<i>Hydatigera taeniaeformis</i>	Cestoda	WormBase ParaSite
<i>Hymenolepis diminuta</i>	Cestoda	WormBase ParaSite
<i>Hymenolepis microstoma</i>	Cestoda	WormBase ParaSite
<i>Hymenolepis nana</i>	Cestoda	WormBase ParaSite
<i>Mesocestoides corti</i>	Cestoda	WormBase ParaSite
<i>Schistocephalus solidus</i>	Cestoda	WormBase ParaSite
<i>Spirometra erinaceieuropaei</i>	Cestoda	WormBase ParaSite
<i>Taenia asiatica</i>	Cestoda	WormBase ParaSite
<i>Taenia saginata</i>	Cestoda	WormBase ParaSite
<i>Clonorchis sinensis</i>	Trematoda	WormBase ParaSite
<i>Echinostoma caproni</i>	Trematoda	WormBase ParaSite
<i>Fasciola hepatica</i>	Trematoda	WormBase ParaSite
<i>Opisthorchis viverrini</i>	Trematoda	WormBase ParaSite
<i>Schistosoma curassoni</i>	Trematoda	WormBase ParaSite
<i>Schistosoma haematobium</i>	Trematoda	WormBase ParaSite
<i>Schistosoma japonicum</i>	Trematoda	WormBase ParaSite
<i>Schistosoma mansoni</i>	Trematoda	WormBase ParaSite
<i>Schistosoma margrebowiei</i>	Trematoda	WormBase ParaSite
<i>Schistosoma mattheei</i>	Trematoda	WormBase ParaSite
<i>Schistosoma rodhaini</i>	Trematoda	WormBase ParaSite
<i>Trichobilharzia regenti</i>	Trematoda	WormBase ParaSite
<i>Gyrodactylus salaris</i>	Monogenea	WormBase ParaSite
<i>Protopolystoma xenopodis</i>	Monogenea	WormBase ParaSite
<i>Macrostomum lignano</i>	Turbellaria	WormBase ParaSite
<i>Schmidtea mediterranea</i>	Turbellaria	PlanMine

Procesamiento de secuencias codificantes para proteínas de *Schmidtea mediterranea*

Las secuencias codificantes para proteínas pertenecientes al genoma de *Schmidtea mediterranea* fueron obtenidas de la base de datos PlanMine v2.0beta (<http://planmine.mpi-cbg.de/planmine/begin.do>).

Las mismas corresponden a datos obtenidos por el Laboratorio encabezado por el Dr. Aziz Aboobaker, del Departamento de Zoología de la Universidad de Oxford (Inglaterra).

Estos datos transcriptómicos brindan una mayor confianza para el trabajo aquí planteado, ya que se tendrá sólo en cuenta aquellas secuencias que sean codificantes en cada genoma. Se optó por utilizar estos datos tomando en cuenta dos factores: la baja calidad de los datos disponibles para este organismo en la base de datos WormBase ParaSite v9.0 (<http://parasite.wormbase.org/index.html>), y el hecho de que las secuencias transcriptómicas obtenidas por el Dr. Aboobaker permiten asegurar que los miembros de las familias génicas estudiadas se ven expresadas en el organismo.

Los transcritos identificados fueron traducidos en los seis marcos de lectura posibles, seleccionándose como secuencia proteica

codificada aquella de mayor tamaño. Para ello se utilizó la herramienta *getorf*, del paquete bioinformático EMBOSS (Rice et al., 2000). Este tratamiento lleva a que sólo se consideren las isoformas de mayor tamaño para cada CDS considerado.

Predicción de secuencias codificantes para proteínas de secreción.

Las secuencias nucleotídicas codificantes para proteínas fueron analizadas con el fin de identificar aquel subgrupo de las mismas que definen el *secretoma* de cada especie.

Para ello, se siguió el protocolo descrito por Garg & Ranganathan (Garg & Ranganathan, 2011. Ver **Figura 2**).

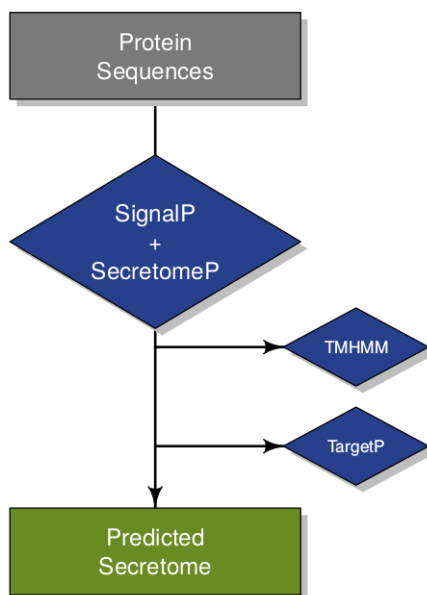


Figura 2. Flujo de trabajo para predicción de secretomas. Adaptado de Garg & Ranganathan, 2011

A continuación el lector encontrará una descripción de los programas utilizados en este protocolo. En términos generales, se identificaron secuencias codificantes para proteínas candidatas a ser secretadas por la vía canónica de secreción y por vías alternativas¹, descartándose posteriormente aquellas que presentaran indicios de poseer hélices transmembrana o de ser destinadas a residir en la mitocondria.

Detección de secuencias codificantes para presuntas proteínas secretadas por vía canónica de secreción.

SignalP v4.1 (Petersen et al., 2011) fue utilizado para detectar aquellas proteínas plausibles de ser secretadas por la vía canónica de secreción.

El programa se basa en la utilización de redes neuronales artificiales (*artificial neural networks*, en inglés), mediante las cuales es posible entrenar al mismo con un set de proteínas que oficie de control positivo (es decir, proteínas que en efecto son secretadas por esta vía), así como también un set de proteínas que oficie de control negativo. En este último caso, se utilizan proteínas tanto citosólicas y nucleares (*'SignalP-noTM networks'*) como poseedoras de hélices transmembrana (*'SignalP-TM networks'*).

Para cada posición en la secuencia analizada se producen tres indicadores, utilizados para determinar la presencia o ausencia de péptido señal en la misma (para más detalles ver Kihara, D (ed), 2017.):

i) C-score: producido por redes neuronales entrenadas para distinguir entre sitios de clivaje de cualquier otro tipo de secuencia. El mismo está definido de tal forma que aumenta de forma considerable inmediatamente después del sitio de clivaje (es decir, en el primer residuo aminoacídico perteneciente a la proteína madura).

¹Ver sección **Introducción**, donde se presentan las principales características que definen a cada uno de estos grupos.

ii) *S-score*: producido por redes neurales entrenadas para distinguir sitios internos del péptido señal de sitios pertenecientes a la proteína madura (o provenientes de proteínas sin péptido señal).

iii) *Y-score*: promedio geométrico entre el *C-score* y la pendiente del *S-score*. Permite distinguir entre múltiples sitios con un alto *C-score* asignado, tomándose como sitio de clivaje aquel donde el *S-score* es mayor.

A partir de estos indicadores es posible generar otros dos descriptores globales para cada secuencia:

i) *S promedio*: promedio de los *S-score* asociados a los sitios pertenecientes al péptido señal tentativo (*i.e.* desde la primer posición desde el extremo N-terminal de la secuencia proteica hasta el sitio anterior a aquel donde se computó el mayor *Y-score*).

ii) *D-score*: promedio ponderado del *S promedio* y el máximo *Y-score* computado. **Este es el indicador empleado para distinguir secuencias del péptido señal del resto de las secuencias.**

En este trabajo los sets de entrenamiento utilizados fueron aquellos brindados por defecto por los desarrolladores del programa. Un detalle no menor a tener en cuenta refiere a la sensibilidad de esta versión del programa respecto a versiones anteriores. Tal como reportan en la bibliografía sus autores (ver Kihara, D (ed), 2017.), SignalP v4.1 posee menor especificidad que SignalP v3.0. No obstante, es posible utilizar parámetros de *cutoff* proporcionados por los autores para emular la sensibilidad alcanzada por esta versión anterior. Con el fin de disminuir el número de falsos negativos se optó por correr el programa utilizando estos parámetros.

Por tanto, aquellas secuencias proteicas con un *D-score* mayor o igual a 0.34 (ya sean computadas con *SignalP-TM* o *SignalP-noTM networks*) fueron asumidas como candidatas a ser secretadas por vía canónica de secreción.

DetECCIÓN DE SECUENCIAS CODIFICANTES PARA PRESUNTAS PROTEÍNAS SECRETADAS POR VÍAS ALTERNATIVAS DE SECRECIÓN.

SecretomeP v1.0 (Bendtsen et al., 2004) fue utilizado para identificar aquellas proteínas plausibles de ser secretadas por vías alternativas de secreción (*i.e.* toda proteína secretada que carezca de un péptido señal en su extremo N-terminal).

En resumen, este predictor utiliza cinco redes neuronales artificiales distintas, entrenadas para predecir si la secuencia en cuestión está destinada a ser secretada.

Para ello, seis indicadores derivados de la secuencia proporcionada: número de átomos, número de residuos aminoacídicos cargados positivamente, localización celular (según lo predicho por ProP 1.0c (Duckert et al., 2004), existencia de regiones de baja complejidad (acorde a SEG (Wootton et al., 1993)) y presencia y topología de hélices transmembrana (predicho por TMHMM v2.0c (Krogh et al., 2001)). Se obtiene un valor de predicción ('NN-score') para cada una de estas redes neuronales artificiales, siendo la media de los valores calculados aquel valor que describe la probabilidad de que la proteína en cuestión sea secretada.

Una vez realizada la predicción para cada una de las proteínas presentes en los genomas utilizados, se consideraron como proteínas plausibles de ser secretadas por vías no clásicas a aquellas para las cuales se obtuvo un NN-score ≥ 0.9 (siguiendo a Cuesta-Astroz et al. 2017). Es importante recalcar que este valor es conservador, asegurando que el número de falsos positivos obtenido sea mínimo (ver **Discusión**, donde se trata la importancia de este punto).

PREDICCIÓN DE SECUENCIAS DESTINADAS A RESIDIR EN MITOCONDRIA

Las proteínas destinadas a residir en mitocondrias y cloroplastos poseen un péptido señal en su extremo N-terminal, realizando un pasaje por el retículo endoplasmático y el aparato de Golgi de la misma manera que lo hacen aquellas proteínas secretadas a través de la vía canónica de secreción. Es por tanto pertinente lograr discernir estas proteínas de aquellas que en efecto son secretadas por dicha vía.

La predicción de proteínas con destino mitocondrial fue realizada utilizando el programa TargetP v1.1 (Emanuelsson et al. 2007), el cual se basa también en el uso de redes neuronales artificiales para asignar una probabilidad de secreción a este compartimiento a las proteínas de interés. Como umbral se utilizó un NN-score ≥ 0.95 (siguiendo a Cuesta-Astroz et al. 2017), siendo entrenado el set de trabajo con aquel que brindan por defecto los creadores de dicho programa.

Análisis filogenético

Agrupamiento de secuencias en *clusters* de homólogos

Las proteínas anotadas en los genomas utilizados² fueron agrupadas en *clusters* de homólogos, a fin de establecer posteriormente relaciones evolutivas entre las mismas.

Para ello, se utilizó el programa `get_homologues` v.18042017 (Contreras-Moreira et al., 2013), utilizándose OrthoMCL (Li et al., 2003) como algoritmo para resolver las relaciones de homología. El mismo ha demostrado ser eficaz al momento de desempeñar esta tarea sobre secuencias proteicas relativas a genomas eucariotas, pudiendo lidiar con las complejidades del caso (*e.g.* presencia de múltiples eventos de duplicación, arquitectura multidominio de gran cantidad de genes, existencia de redundancia a nivel de función con consecuente similitud a nivel de estructura proteica.) (Li et al., 2003). En el caso particular de este trabajo, el establecimiento de relaciones de inparalogía resulta esencial, permitiendo este algoritmo la detección de las mismas a través de una definición clara.

En breve, el algoritmo establece relaciones de homología de la siguiente manera:

- i*) Se realizan alineamientos locales pareados entre todas las secuencias brindadas, utilizando para ello el programa BLASTP (Altschul et al., 1990).
- ii*) Sean los genomas *i* y *j*, dos secuencias serán definidas como posibles ortólogos si presentan de forma recíproca los *e-values* más significativos al momento de alinear todas las proteínas del genoma de la especie *i* con las del genoma *j*. Las mismas deberán presentar a su vez un valor de cobertura e identidad nucleotídica superiores a un valor límite establecido por el usuario.
- iii*) Para cada uno de los ortólogos definidos en el punto anterior, se definirán relaciones de inparalogía. Dos secuencias son definidas como *inparálogas* si presentan un *e-value* más significativo entre sí que con cualquier proteína de otro genoma.
- iv*) Las relaciones de ortología e inparalogía son representadas como grafos. Los nodos de este grafo representan una secuencia proteica, mientras que los arcos uniéndolos poseerán un valor proporcional al *e-value* obtenido al momento de realizarse el alineamiento con BLASTP (ver Li et al., 2003 para más detalles respecto a las normalizaciones realizadas).
- v*) El grafo planteado en el punto anterior es representado en una matriz de Markov, sobre la cual se aplica un algoritmo de clusterización (*Markov cluster algorithm*, ver van Dongen et al., 2000), el cual permite realizar agrupamientos entre los nodos teniendo en cuenta la cantidad de arcos que conectan los nodos del grafo. El algoritmo está diseñado para mantener en un cluster sólo aquellos nodos que están más conectados entre sí que con el resto de los nodos presentes en el grafo.

En este trabajo se utilizó como criterio para establecer relaciones de homología que aquellas secuencias que pertenecen a un mismo *cluster* deben presentar un **valor de cobertura mayor o igual a 75 %**, así como una **identidad nucleotídica mayor o igual a 40 %**³.

Alineamiento de secuencias

Se alinearon las secuencias proteicas de todo *cluster* de homólogos compuesto por al menos dos miembros. Para ello se utilizó el programa MUSCLE v3.8.31 (primera versión descrita en Edgar, 2004). Teniendo en cuenta que la enorme cantidad de secuencias agrupadas en estos *clusters* de homólogos, y el hecho de que la cantidad y el largo de las mismas varía según el caso, se decidió correr el programa utilizando los argumentos sugeridos para alineamientos de gran tamaño. Esta opción contempla tanto la velocidad con la que se realizan los alineamientos como la memoria RAM empleada en el caso de que las secuencias a alinear sean demasiadas o de gran tamaño.

Inferencia de árboles filogenéticos

Teniendo en cuenta que las secuencias nucleotídicas poseen un mayor grado de saturación, y por tanto conllevan una pérdida de señal si el tiempo de divergencia entre homólogos es considerable, se decidió en este trabajo utilizar secuencias proteicas al momento de inferir árboles filogenéticos.

²Es decir, las secuencias proteicas disponibles en archivos multi-FASTA en WormBase ParaSite, así como aquellas obtenidas por traducción del genoma de *S. mediterranea* (ver sección **Procesamiento de secuencias codificantes para proteínas de *Schmidtea mediterranea***).

³Ver **Discusión**, donde se plantean posibles repercusiones de plantear estos valores como umbral y no otros.

FastTree2 (Price et al., 2010) fue utilizado para inferir árboles filogenéticos a partir de los alineamientos de las secuencias proteicas pertenecientes a los *clusters* de homólogos considerados.

El programa utiliza como método de inferencia la máxima verosimilitud aproximada, utilizando como punto de partida un árbol de topología aproximada calculado a partir de una variante heurística del método de *neighbor joining* (Saitou et al., 1987). El árbol generado se refina, aumentándose la verosimilitud de forma heurística en cada caso, alternando para ello rearrreglos de la topología denominados NNIs y SPRs (ver Zheng, 2014 para una definición de los mismos).

Como modelo evolutivo se utilizó en este caso aquel postulado por JTT (Jones, Taylor & Thornton; 1992), utilizándose a su vez 20 categorías para modelar la tasa evolutiva de cada sitio considerado en el alineamiento, variando las mismas entre 0.05 y 20. Debido a la magnitud del número de secuencias consideradas, se decidió correr el programa en su modalidad rápida, lo cual implica que durante la fase de estimación de topología inicial no se busquen alternativas de mayor verosimilitud. Según simulaciones corridas por Price (Price, 2010) sobre un alineamiento 237,000 16S rRNAs, esto implica solo un pequeño efecto sobre la calidad de los árboles filogenéticos inferidos, si se mide esto en base al largo de ramas o al logaritmo de la verosimilitud de los mismos.

DetECCIÓN DE GRUPOS DE SECUENCIAS INPARÁLOGAS

Con el fin de detectar genes inparálogos en los *clusters* de homólogos definidos previamente, se realizó la siguiente aproximación:

i) Teniendo en cuenta la definición de inparalogía es posible afirmar que, dado un grupo monofilético compuesto por proteínas codificadas por genes pertenecientes al genoma de una única especie, las mismas serán codificadas por genes inparálogos⁴.

A modo de ejemplo, sea un árbol filogenético (**Figura 3**) representando las relaciones evolutivas existentes entre las proteínas *S1*, *S2*, *S3*, *S4*, *S5*, *X1*, *Y1*, y *Z1* (codificadas por genes de los genomas de las especies S, X, Y y Z respectivamente). Diremos que los genes que codifican para *S3*, *S4* y *S5* son inparálogos, ya que estas proteínas forman un grupo monofilético en dicho árbol.

ii) Con el fin de disminuir la cantidad de árboles filogenéticos a analizar en busca de grupos de inparálogos, se realiza una preselección de los mismos. Son tenidos en cuenta aquellos árboles en los cuáles se detecta al menos un grupo monofilético de inparálogos.

iii) Se analizan todas las clanes existentes en el árbol, a fin de encontrar al menos uno compuesto en exclusiva por proteínas codificadas en el genoma de un mismo organismo. Esto implica, en efecto, buscar al menos un caso de inparalogía.

Dicho análisis fue realizado empleando las librerías *ape* (Paradis et al., 2004), *phangorn* (Schliep et al., 2010) y *phytools* (Revell et al., 2012) del entorno de programación R (R Development Core Team, 2016).

Se ilustrará la estrategia utilizando nuevamente a modo de ejemplo la filogenia de la **Figura 3**.

⁴Es importante tener en cuenta que es razonable afirmar esto gracias a que previamente se han establecido criterios de cobertura e identidad aminoacídica mínimos al momento de determinar los *clusters* de homólogos considerados.

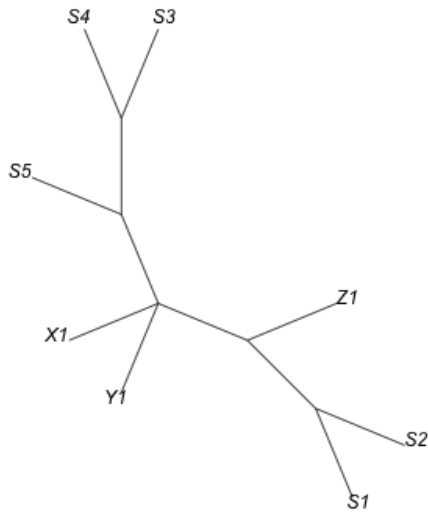


Figura 3. Árbol filogenético hipotético. Se consideran secuencias codificantes para proteínas en los genomas de las especies X , Y , Z y S , denotándose con diferentes números a diferentes secuencias codificantes.

En primer lugar se realiza una preselección de aquellos árboles que posean al menos un grupo monofilético para la especie de interés. Esto se logra realizando las siguientes operaciones sobre cada árbol provisto al script:

i) Se clasifican las proteínas pertenecientes a la filogenia según la especie del genoma que las codifica. Así, para la filogenia ilustrada en este caso, se obtienen los vectores \mathbf{X} (compuesto por $X1$), \mathbf{Y} (compuesto por $Y1$), \mathbf{Z} (compuesto por $Z1$) y \mathbf{S} (compuesto por $S1, S2, S3, S4$ y $S5$).

ii) Se establecen en cada caso todas las combinaciones posibles para dichos elementos. En este caso sólo los elementos de \mathbf{S} tendrán combinaciones a considerar. Se tienen en cuenta las combinaciones de elementos tomados de $n = 2, 3, \dots, 5$.

iii) Utilizando la función *is.monophyletic* (del paquete *ape*) se evalúa si al menos una de estas combinaciones constituye un grupo monofilético. En caso afirmativo se guarda el árbol con un código que lo identifica para proseguir con el análisis. En este caso cualquiera de los grupos $(S1, S2)$ o $(S3, S4)$ serán evaluados como monofiléticos, culminando el proceso.

Una vez concluida esta primera fase, se pasa al análisis exhaustivo de los grupos monofiléticos presentes, con el fin de realizar su extracción.

Esto se logra a través del siguiente conjunto de operaciones:

i) Se obtienen una matriz que representa todos los clanes presentes en la filogenia, utilizando la función *getClans* (del paquete *phangorn*).

En el caso considerado esto generará los clanes $(S3)$, $(S3, S4)$, $(S3, S4, S5)$, $(S3, S4, S5, X1)$, $(S3, S4, S5, X1, Y1)$, \dots , $(S3, S4, S5, X1, Y1, Z1, S1, S2)$.

ii) Se guardan aquellos clanes compuestos en su totalidad por proteínas codificadas en el genoma de la especie de interés.

En este caso, si la especie de interés es S , se considerarán en lo que sigue del análisis los clanes $(S3)$, $(S3, S4)$, $(S3, S4, S5)$, $(S1)$ y $(S1, S2)$.

iii) Se colapsan aquellos clanes anidados. Esto se logra reemplazando aquellos clanes con intersección no vacía por el clan de mayor tamaño.

En este caso los clanes $(S3)$, $(S3, S4)$ y $(S3, S4, S5)$ comparten el elemento $S3$, por lo cual sólo se considerará el clan $(S3, S4, S5)$.

El final de este proceso es, en este caso, la obtención de los clanes ($S3, S4, S5$) y ($S1, S2$) que son, en efecto, los grupos monofiléticos compuestos en su totalidad por proteínas codificadas en el genoma de interés. Lo que es más, estos son grupos de posibles inparálogos.

En los casos en los que existe anotación de las isoformas codificadas por un gen, se pasó a considerar únicamente aquella isoforma de mayor tamaño. Esta consideración resulta esencial al momento de realizar esta aproximación para la detección de inparálogos. De no realizarse este paso previo, se incurre en la posibilidad de que un conjunto de isoformas sea erróneamente asignada como un conjunto de inparálogos.

Asignación de términos funcionales

Una ontología implica el establecimiento de relaciones claras entre un conjunto definido de entidades. Así, el denominado *Gene Ontology project* busca clarificar y jerarquizar la información de diferentes bases de datos referentes a procesos biológicos. Tres categorías (o *gene ontology terms*; GO terms de aquí en adelante) son contempladas con el fin de organizar la información existente referente al rol funcional de un gen dado según esta aproximación (ver Ashburner et al., 2000): aquella referente al **proceso biológico** en el que participa, definido según la función biológica objetivo a la que aporta la actividad de este gen; la **función molecular del mismo**, definida según la actividad bioquímica del mismo; y el **componente celular** donde este gen se encuentra activo en el espacio celular.

Dada una secuencia proteica, en este trabajo se asignó a la misma un conjunto de GO terms utilizando el programa eggNOG mapper (Huerta-Cepas et al., 2017).

De forma breve, la asignación de términos ontológicos se realiza según esta aproximación de la siguiente manera:

i) Utilizando o bien HMMER 3 (Eddy et al., 2011) o bien DIAMOND (Buchfinck et al., 2015), se establece qué proteína de la base de datos eggNOG (<http://eggnogdb.embl.de/#/app/home>, Powell et al., 2012) posee mayor similitud con la secuencia de interés.

La definición del método utilizado debe tomar en cuenta principalmente la cantidad de secuencias a las cuales se desea asignar GO terms, ya que si bien una búsqueda utilizando HMMER 3 puede resultar más sensible también implica un mayor tiempo de computo. En este trabajo se decidió utilizar DIAMOND en esta fase, debido al gran número de secuencias consideradas. El método, de forma similar al que utiliza el popular programa BLAST (Altschul et al., 1990), alinea en primera instancia k-meros (llamados informalmente *seeds*; “semillas”) derivados de la secuencia de interés a k-meros derivados de las secuencias contenidas en la base de datos de eggNOG. En caso de que el alineamiento supere un cierto puntaje, se pasa a la extensión del mismo. La rapidez del mismo, respecto a la alineación por BLAST reside en la indexación en k-meros no sólo de las secuencias presentes en la base de datos sino también de aquellas secuencias de interés utilizadas para el alineamiento (ver Huson et al., 2015 para más detalles respecto al funcionamiento de este doble indexado y sus ventajas).

ii) Aquella secuencia de la base de datos eggNOG que presenta un *e-value* más significativo es empleada para buscar secuencias ortólogas en organismos cercanos. Para ello, se utiliza información extraída de árboles filogenéticos curados manualmente, disponibles también en eggNOG. Se evita en estos casos el reporte de secuencias inparálogas que pudieran tener un conjunto radicalmente diferente de GO terms.

iii) Todos los descriptores funcionales descritos para estas secuencias ortólogas son asignados a la secuencia de interés. El usuario puede restringir el nivel de proximidad filogenética mínimo requerido al momento de realizar esta transferencia de términos (evitando así la anotación con términos de especies muy lejanas a nivel evolutivo), o incluso el restringir esta acción a la secuencia ortóloga más cercana. Por defecto son transferidos **GO terms**, **rutas KEGG** anotadas (Kanehisa et al., 2002) y **categorías funcionales de COGs** (Tatusov et al., 2000).

Análisis de enriquecimiento

Los análisis de enriquecimiento en esta tesina fueron realizados mediante un test de Fisher. Los *p-values* obtenidos en cada caso fueron ajustados mediante el método FDR (del inglés, *false discovery rate*), a fin de considerar los efectos de testeo múltiple.

Se realizaron análisis de enriquecimiento para proteínas de secreción dentro de los inparálogos detectados mediante el análisis filogenético ya descrito en esta sección. Esta aproximación es válida porque *a priori* uno puede afirmar que dichas categorías son independientes: no existe una asociación mecánica entre la propiedad de una secuencia proteica de ser secretada y su carácter de ser codificada por un gen que ha sido duplicado.

Los análisis de enriquecimiento en términos funcionales tanto para proteínas codificadas por genes inparálogos, como para los subgrupos de las mismas que codifican para proteínas de secreción o para proteínas de secreción en donde se postula un fenómeno adaptativo, fueron realizados utilizando el paquete topGO (Rahnenfuehrer, 2016) del entorno de programación R.

A fin de obtener una visión global de los términos funcionales en los que se encuentra enriquecido el conjunto de secuencias analizado, se mapeó los GO Terms obtenidos contra un conjunto de GO Slims genérico (obtenido de <http://www.geneontology.org/page/go-slim-and-subset-guide>). Un GO Slim se define como un subconjunto de GO terms de carácter general, el cual está ideado para permitir una visualización tan fina de un conjunto de términos funcionales como se desee. Para realizar este mapeo se utilizó el programa Map2Slims, del paquete OWITools (<https://github.com/owlcollab/owltools>).

Análisis de distribución en isocóros de secuencias codificantes en el genoma de *S. mansoni*

Para analizar si existe correlación entre el hecho de que un gen sea inparólogo o codificante para una proteína de secreción y su pertenencia a estructuras del genoma caracterizadas por un determinado contenido en GC (es decir, a un determinado isocóro), se utilizó información proporcionada por el Dr. Guillermo Lamolle (previamente reportada en Lamolle et al., 2016).

Genes con dichas características pertenecientes al genoma de *S. mansoni* fueron mapeados a estas estructuras de tipo isocóro, realizándose posteriormente un análisis de enriquecimiento (ver las subsecciones Asignación de términos funcionales y Análisis de enriquecimiento).

Los autores de dicho trabajo describieron estas estructuras similares a isocóros mediante el siguiente proceso:

i) Se filtraron CDS que no presentaran codones de iniciación/stop, así como aquellos que presentaran cambios internos en el marco de lectura. También se excluyó del análisis aquellos con una longitud menor a 100 pb.

ii) Para todo análisis, así como para la definición de los isocóros, se utilizaron ventanas deslizantes de 100kb.

iii) Se clasificó toda zona según su contenido GC como perteneciente a los siguientes isocóros: **A** ($GC\% < 33.5$), **B** ($33.5 < GC\% < 36.5$) y **C** ($GC\% > 36.5$). Los mismos son arbitrarios, y representan un compromiso entre la definición de tres isocóros y la cantidad de ventanas deslizantes pertenecientes a cada región según dicha definición. Los autores reportaron obtener resultados similares variando dichas definiciones.

Análisis de evolución molecular

Para poner a prueba la hipótesis según la cual la selección natural positiva opera dirigiendo la evolución de las secuencias pertenecientes a los grupos de inparálogos previamente caracterizados, se utilizó el programa CodeML (Yang, 2000) para estimar el indicador omega (relación entre la tasa de sustituciones no sinónimas y sinónimas; dN/dS). Este indicador permite distinguir si las secuencias analizadas evolucionan bajo un régimen de selección purificadora ($dN/dS < 1$), bajo un régimen neutro ($dN/dS = 1$) o bajo selección natural positiva ($dN/dS > 1$).

Según la hipótesis central de este trabajo (*i.e.* existen presiones selectivas dadas por la interacción huésped-patógeno, las cuales llevan a que secuencias duplicadas codificantes para proteínas de secreción evolucionen bajo un régimen de selección natural positiva) no es verosímil asumir que existe un mismo régimen selectivo para todos los aminoácidos de una proteína. Por tanto, es necesario asumir lo que se denomina un *modelo de sitios*, según el cuál se asumen diferentes modelos de dN/dS al momento de modelar la evolución de las secuencias codificantes estudiadas. Siguiendo la línea de razonamiento establecida por la hipótesis central de este trabajo, este modelo permite por ejemplo establecer un $dN/dS > 1$ para una fracción de los codones considerados (los cuales representan aquellos sitios que codifican para sitios seleccionados para evadir la respuesta inmune del hospedero) y una fracción de sitios con un régimen neutro o purificador (siendo estas categorías representantes de la fracción de sitios que no se ven afectados por la respuesta inmune del hospedero, o que cumplen un rol esencial para la proteína codificada y por tanto no pueden variar sin afectar el *fitness* del organismo).

En este trabajo no se utilizó un modelo de ramas, según el cual diferentes linajes poseen un diferente régimen evolutivo. La decisión se basa en asumir que un conjunto de secuencias inparálogas se encuentran bajo un régimen selectivo similar (guiado por las presiones selectivas que se ejercen sobre el organismo), asunción que no necesariamente es cierta y puede generar artefactos (ver Discusión para más detalles).

¿Cómo se detectan entonces los sitios sobre los cuales es factible que opere la selección natural positiva? Como describen Yang y Bielawski (Yang & Bielawski, 2000) el proceso consta de dos grandes etapas. En primer lugar, es necesario testear si de hecho es necesaria la presencia de sitios con un dN/dS para observar el conjunto de datos analizado. Esto se logra contrastando un modelo donde dicha asunción es tomada contra otro donde se sumen únicamente sitios con dN/dS menor o igual a 1. La prueba estadística de esta noción se realiza a través de lo que se denomina LRT (por sus siglas inglés, *Likelihood Ratio Test*). Una vez se considera plausible que existan sitios donde opere la selección natural positiva, es posible realizar cálculos de probabilidad posterior (utilizando programas que se basan en la estadística bayesiana, como el programa *BEB* del paquete CodeML) para detectar aquellos sitios del conjunto de proteínas analizadas donde existe una gran probabilidad de que suceda el fenómeno.

Los modelos de codones utilizados en este trabajo son todos basados en cálculos de verosimilitud, en donde se asigna a una determinada fracción de sitios un dN/dS estimado a partir de los datos. Las fracciones a asignar son basados en distribuciones estadísticas según el modelo. (El lector puede referirse a Yang & Bielawski para una descripción más detallada.)

A continuación una breve descripción de los modelos empleados

i) Modelo M0.

Bajo este modelo todas las secuencias evolucionan bajo un único omega, siendo el mismo estimado a partir de un promedio de la tasa dN/dS para todos los codones existentes en la secuencia.

ii) Modelo M1. El modelo asume dos clases de sitios: uno bajo un régimen neutro ($dN/dS = 1$) y otro bajo régimen de selección purificadora ($dN/dS < 1$).

iii) Modelo M2. Se agrega al modelo M1 una clase de sitios bajo selección natural positiva ($dN/dS > 1$), donde dN/dS es estimado a partir de los datos observados en base a cálculos de máxima verosimilitud.

En este caso, si el Modelo M2 es más verosímil que el Modelo M1 (en base a un test LRT), y el dN/dS estimado es mayor a 1, entonces es posible afirmar que existen sitios que evolucionan bajo selección natural positiva.

iv) Modelo M7 De forma análoga a los casos de M1 y M2, es posible plantear que las categorías de dN/dS empleadas provengan de una distribución estadística como puede ser una distribución beta. En efecto, el modelo M7 se basa en esta noción, permitiéndose que omega asuma valores menores o iguales a uno, tomados de distribuciones estadísticas con este perfil.

v) Modelo M8

Siguiendo la línea argumental del caso anterior, el modelo M8 posee las mismas características que el M7, pero se permite una categoría extra donde se extraen valores de omega mayores a uno, también tomados de una distribución beta.

Acorde a lo recomendado por Yang (Yang, 2000), se compararon los modelos M2 vs M1 y M8 vs M7 a fin de evaluar si la utilización de categorías de selección natural positiva es estadísticamente significativa para los datos analizados (según el criterio de LTR antes mencionado). Considerando lo que se formula en dicho manual, se toman en cuenta los casos donde el uso del modelo M2 resulta adecuado, ya que esta condición es más restrictiva que aquella impuesta por el modelo M8 (debido a que el último emplea un mayor número de parámetros).

Posteriormente se realizó una búsqueda por similitud de las secuencias para las que se observó una evolución molecular en régimen de selección natural positiva a fin de identificar posibles secuencias homólogas a las mismas. Para ello, se empleó el software DIAMOND (Buchfink et al., 2015), realizándose la búsqueda contra la base de datos RefSeq del NCBI (<https://www.ncbi.nlm.nih.gov/refseq/>. O'Leary et al., 2016.). De forma coherente a lo realizado al momento de definir grupos de homólogos, se requiere un mínimo de 75 % de identidad aminoacídica y 40 % de cobertura entre secuencias para ser definidas como posibles homólogos.

Resultados

Obtención de clusters de homólogos

Se establecieron relaciones de homología entre las proteínas anotadas codificadas en los genomas de treinta platelmintos. Para ello se utilizó el programa *get_homologues*, fijándose como límites inferiores de identidad y cobertura el 75 % y 40 %, respectivamente (ver sección Materiales y Métodos para más detalles sobre los algoritmos empleados por dicho programa). De las especies analizadas, 15 (50 %) pertenecen a la clase Cestoda y 11 (36,7 %) a la clase Trematoda, siendo estas dos las dos clases más ampliamente representadas en el conjunto de genomas disponibles a la fecha. Las otras dos clases que componen el filo, Monogenea y Turbellaria, son representadas por dos especies, lo cual equivale a un 6,7 % del total en cada caso. En total, este estudio utilizó 586852 secuencias proteicas, las cuales están anotadas según el protocolo de WormBase ParaSite, exceptuando aquellas predichas para el caso de *S. mediterranea*, donde se partió de las secuencias codificantes disponibles para un transcriptoma disponible en la base de datos PlanMine (ver sección Materiales y Métodos). De estas secuencias, 224423 (38,2 %) son codificadas por secuencias pertenecientes a genomas de céstodos, 225008 están codificadas en genomas de tremátodos, y unas 53342 (9,1 %) y 84079 (14,3 %) por organismos de las clases Monogenea y Turbellaria, respectivamente. Como se puede apreciar, la distribución de secuencias analizadas en las diferentes clases refleja a *grosso modo* el esquema de distribución observable para las especies, aunque es importante notar que las secuencias disponibles para las proteínas codificadas por organismos clasificados en la clase Turbellaria en parte debe su cantidad al hecho de que uno de los organismos, *M. lignano*, es tetraploide (ver sección Discusión para más detalles sobre las implicancias de este hecho).

A partir de estas secuencias proteicas se infirieron 361306 clusters de homólogos, de los cuales 66626 (18,0 %) son conformados por secuencias codificadas por los genomas de más de una especie. Los demás clusters son, en cambio, están integrados por secuencias *huérfanas*: según el método empleado y los límites establecidos para identidad y cobertura, las mismas no muestran relaciones de homología con ninguna otra secuencia.

En base a aquellos grupos de homólogos con al menos dos secuencias proteicas pertenecientes al mismo genoma se infirieron los árboles filogenéticos sobre los que posteriormente se basó el resto del trabajo realizado. En total, **29997** árboles fueron inferidos utilizando el programa FastTree 2 (ver sección Materiales y Métodos para más detalles). Es de notar todo árbol de tres secuencias donde dos pertenecieran a una misma especie fue dejado de lado, para evitar interpretar inparalogía donde no se pudiera asegurar tal (el lector puede referirse a la sección Discusión para un análisis de dicha decisión y sus implicancias en este trabajo).

La **Tabla 1** muestra en detalle el número de árboles filogenéticos obtenidos donde se encontró representación de cada especie. A su vez se muestra el número de grupos inparálogos obtenidos en cada caso, así como la cantidad de secuencias proteicas dentro de los mismos, el porcentaje que representan dentro de la totalidad de secuencias proteicas de la especie utilizadas en este trabajo y el número promedio de secuencias inparálogas por grupo de inparálogos. Dentro de las tendencias extremas se pueden remarcar los casos de las especies *H. microstoma* (11,41 %), *S. solidus* (13,72 %), *S. erinaceieuropaei* (22,25 %), *T. asiatica* (10,10 %), *F. hepatica* (11,89 %), *S. mansoni* (11,24 %) y *M. lignano* (32,32 %), para las cuáles el porcentaje de secuencias inparálogas inferido representa más del 10 % de las secuencias proteicas anotadas para sus respectivos genomas, mientras que en los casos de *S. haematobium* (0,80 %), *S. japonicum* (0,35 %) y *S. matthei* (0,77 %) se observó la tendencia opuesta, con menos de un 1 % de las secuencias siendo catalogadas como inparálogas.

Tabla 1. Conjunto de genes identificados como inparálogos.

Especie	Núm. de árboles donde está presente	Numero de arboles con inparalogos	Numero de grupos de inparalogos	Numero de inparalogos	Numero total de genes	Porcentaje de inparalogos	Promedio de inparalogos por grupo
<i>Diphyllobothrium latum</i>	2183	266	257	600	19966	3,01 %	2,3
<i>Echinococcus canadensis</i>	3950	175	173	627	11448	5,48 %	3,6
<i>Echinococcus granulosus</i>	5045	77	75	152	10273	1,48 %	2,0
<i>Echinococcus multilocularis</i>	5180	218	213	559	10669	5,24 %	2,6
<i>Hydatigera taeniaeformis</i>	5503	83	80	171	11649	1,45 %	2,1
<i>Hymenolepis diminuta</i>	4236	88	78	203	11271	1,80 %	2,6
<i>Hymenolepis microstoma</i>	4695	557	557	1403	12373	11,33 %	2,5
<i>Hymenolepis nana</i>	4264	258	254	761	13777	5,52 %	3,0
<i>Mesocestoides corti</i>	4094	139	134	359	10614	3,38 %	2,7
<i>Schistocephalus solidus</i>	3492	549	541	2343	20228	11,58 %	4,3
<i>Spirometra erinaceieuropaei</i>	5190	3876	3871	8777	39557	22,19 %	2,3
<i>Taenia asiatica</i>	4769	545	546	1312	13322	9,85 %	2,4
<i>Taenia saginata</i>	4654	395	395	923	13161	7,01 %	2,3
<i>Taenia solium</i>	5452	88	86	188	12481	1,51 %	2,2
<i>Clonorchis sinensis</i>	2116	142	143	525	13634	3,85 %	3,7
<i>Echinostoma caproni</i>	1994	357	355	1339	18607	7,80 %	3,8
<i>Fasciola hepatica</i>	2475	484	480	1592	33454	4,76 %	3,3
<i>Opisthorchis viverrini</i>	2164	147	144	375	16356	2,29 %	2,6
<i>Schistosoma curassoni</i>	2017	126	129	311	23546	1,32 %	2,4
<i>Schistosoma haematobium</i>	2633	234	241	822	13073	6,29 %	3,4
<i>Schistosoma japonicum</i>	2112	130	132	430	12738	3,38 %	3,3
<i>Schistosoma mansoni</i>	2862	563	556	1323	11774	11,24 %	2,4
<i>Schistosoma margrebowiei</i>	2888	588	612	2114	26189	8,07 %	3,5
<i>Schistosoma matheei</i>	1962	80	79	176	22997	0,77 %	2,2
<i>Schistosoma rodhaini</i>	2221	238	237	932	24089	3,87 %	3,9
<i>Trichobilharzia regenti</i>	1556	190	189	532	22185	2,40 %	2,8
<i>Gyrodactilus salaris</i>	1809	138	130	316	15436	2,05 %	2,4
<i>Protopolystoma xenopodis</i>	1270	377	367	1177	37906	3,11 %	3,2
<i>Macrostomum lignano</i>	12335	12000	11985	36818	60534	60,82 %	3,1
<i>Schmidtea mediterranea</i>	3665	1890	1885	4538	23545	19,27 %	2,4

Los colores azul, rojo, verde y violeta indican la pertenencia de la especie a las clases Cestoda, Trematoda, Monogenea y Turbellaria (respectivamente).

En la **Tabla 2** se observan los resultados obtenidos al analizar las secuencias proteicas de interés con los programas SignalP, SecretomeP, TMHMM y TargetP, a fin de detectar proteínas plausibles de ser secretadas (ver sección Materiales y Métodos para mayor información sobre los fundamentos que justifican esta aproximación).

El porcentaje de proteínas pertenecientes a los *secretomas* (conjunto de proteínas secretadas por un organismo) deducido por la metodología empleada oscila entre el 8,16% (*G. salaris*) y el 16,16% (*P. xenopodis*) de los respectivos proteomas anotados. Interesantemente, ambos organismos pertenecen a la clase Monogenea.

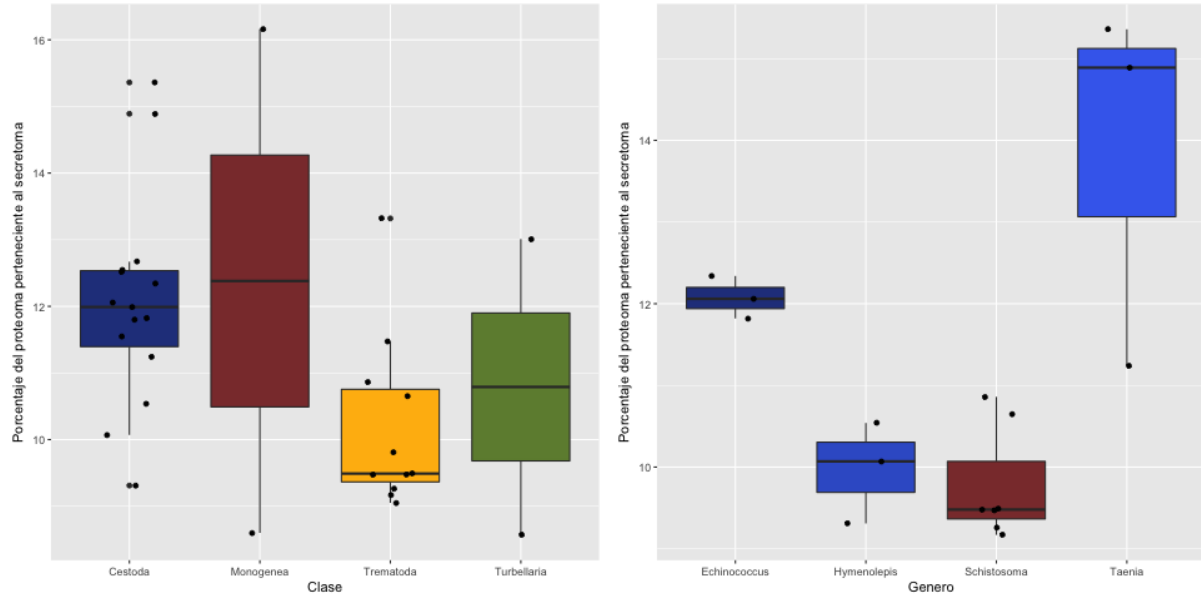
En todos los casos el número de secuencias predichas como candidatas a ser secretadas por la vía clásica (detectadas por el programa SignalP) fue mayor a aquel donde la vía de secreción es no-canónica (detectados por SecretomeP), existiendo a su vez proteínas que fueron detectadas por ambos programas (ver sección Discusión, donde se mencionan casos que ponderan la confiabilidad de la predicción para ambos casos). Otra tendencia general observada refiere al número de secuencias descartadas por presencia de hélices transmembrana o por señales para su transporte a mitocondria: en todos los casos el número de secuencias descartadas por el primer motivo superó ampliamente a las otras, existiendo casos donde ni siquiera se detectaron secuencias plausibles de ser destinadas a este organelo (*i.e.* *E. canadiensis*, *T. solium*, *S. curassoni*, *S. haematobium* y *S. mansoni*). Cabe destacar, para una clara lectura de los datos, que también fueron detectados en estos casos proteínas destinadas a mitocondria donde se predice a su vez la presencia de hélices transmembrana, pudiendo ser este el caso de proteínas destinadas a alguna de las membranas de este organelo.

En la **Tabla 2** se observa a su vez, en la novena columna, qué porcentaje de secuencias proteicas se descartó, dentro de las identificadas como candidatas a pertenecer al secretoma, debido a que se predijo que las mismas poseían hélices transmembrana o tenían residencia mitocondrial. Como se observa, se obtuvo un rango de entre 18,96% (*P. xenopodis*) hasta 53,51% (*S. mediterranea*). Para testear si aquellos genomas que poseen una mayor relación secretoma/proteoma poseen a su vez un menor número de secuencias descartadas mediante la metodología empleada (lo cual podría llevar a malas interpretaciones de los resultados obtenidos, ya que implicaría que existe un sesgo en la capacidad de predicción para los diferentes organismos), se realizó una correlación entre ambas variables. La **Figura S3** (ver Anexo) muestra dicho gráfico, donde en el eje de las abscisas se encuentran el porcentaje de proteínas del proteoma pertenecientes al secretoma para cada organismo, en función del porcentaje de proteínas que se conservó una vez depurado cada secretoma de secuencias con hélices transmembrana o residencia mitocondrial. Se observa una pobre correlación ($R^2 = 0.314$), de lo cual se infiere que este factor no parece influir en gran medida en las deducciones realizadas.

Es interesante destacar que aquellos organismos más cercanos a nivel filogenético como lo son las especies pertenecientes a los géneros Echinococcus, Hymenolepis, Taenia y Schistosoma muestran un comportamiento similar en cuanto al porcentaje de su proteoma predicho como integran del secretoma de la especie.

A su vez, la **Figura 4** muestra el resultado de graficar en forma de *boxplots* el porcentaje de proteínas integrante del secretoma para las especies de cada clase. Como ya se mencionó, Monogenea resultó la clase con más amplio rango en este aspecto, aunque lamentablemente a la fecha existen pocos genomas de la especie secuenciados, lo cual impide ver de forma cabal la tendencia de los organismos de esta clase en forma fidedigna. Si bien no parecen existir grandes diferencias entre las diferentes clases en cuanto a la propiedad graficada, vale decir que los tremátodos poseen la media más pequeña de todo el conjunto estudiado.

La **Figura 5** muestra el resultado de graficar las mismas propiedades para aquellos géneros donde se posee más de una especie. Como se puede ver no existen grandes discrepancias intragénero.



Figuras 4 y 5. Porcentaje del proteoma perteneciente al secretoma según clases y géneros. Se muestra el resultado de graficar el porcentaje del proteoma que integra el secretoma si se agrupa estos porcentajes según clase (Figura 4) o según pertenencia al mismo género (Figura 5). Se consideró para ello las especies pertenecientes a los géneros *Echinococcus*, *Hymenolepis*, *Schistosoma* y *Taenia*. En todos los casos el color azul hace referencia a la clase Cestoda, el rojo a la clase Trematoda, el amarillo a la clase Monogenea y el verde a la clase Turbellaria

Tabla 2. Predicción del secretoma para las especies utilizadas en este trabajo.

Especie	Secreción canónica (SignalP)	Secreción alternativa (SecretomeP)	Secreción	Proteínas transmembrana (TMHMM)	Proteínas mitocondriales (TargetP)	Secretoma	Porcentaje	Porcentaje descartado
<i>Diphyllobothrium latum</i>	2418	1258	3403	1096	0	2307	11,55 %	32,21 %
<i>Echinococcus canadensis</i>	2230	468	2563	1182	0	1381	12,06 %	46,12 %
<i>Echinococcus granulosus</i>	2107	400	2362	1146	2	1214	11,82 %	48,52 %
<i>Echinococcus multilocularis</i>	2219	392	2466	1169	2	1295	12,34 %	47,49 %
<i>Hydatigera taeniaeformis</i>	1973	506	2336	961	0	1375	11,80 %	41,14 %
<i>Hymenolepis diminuta</i>	1702	401	1983	932	2	1049	9,31 %	47,10 %
<i>Hymenolepis microstoma</i>	1975	499	2346	1039	3	1304	10,54 %	44,42 %
<i>Hymenolepis nana</i>	1997	616	2468	1075	5	1388	10,07 %	43,76 %
<i>Mesocestoides corti</i>	1906	439	2231	954	4	1273	11,99 %	42,94 %
<i>Schistocephalus solidus</i>	2927	1090	3796	1233	0	2563	12,67 %	32,48 %
<i>Spirometra erinaceieuropaei</i>	4383	2937	6695	1732	0	4963	12,55 %	25,87 %
<i>Taenia asiatica</i>	2972	668	3429	1439	6	1984	14,89 %	42,14 %
<i>Taenia saginata</i>	2928	721	3441	1416	4	2021	15,36 %	41,27 %
<i>Taenia solium</i>	2125	502	2474	1071	0	1403	11,24 %	43,29 %
<i>Clonorchis sinensis</i>	2486	487	2823	1116	0	1707	12,52 %	39,53 %
<i>Echinostoma caproni</i>	2612	967	3400	1264	2	2134	11,47 %	37,24 %
<i>Fasciola hepatica</i>	3365	1241	4285	1255	4	3026	9,05 %	29,38 %
<i>Opistorchis viverrini</i>	2598	858	3284	1106	0	2178	13,32 %	33,68 %
<i>Schistosoma curassoni</i>	2231	1295	3338	1106	0	2232	9,48 %	33,13 %
<i>Schistosoma haematobium</i>	1870	490	2260	1020	0	1240	9,49 %	45,13 %
<i>Schistosoma japonicum</i>	1858	468	2217	861	0	1356	10,65 %	38,84 %
<i>Schistosoma mansoni</i>	1821	349	2078	988	0	1090	9,26 %	47,55 %
<i>Schistosoma margrebowiei</i>	2295	1442	3575	1172	2	2401	9,17 %	32,84 %
<i>Schistosoma mattheei</i>	2079	1278	3186	1008	1	2177	9,47 %	31,67 %
<i>Schistosoma rodhaini</i>	2548	1485	3819	1202	1	2616	10,86 %	31,50 %
<i>Trichobilharzia regenti</i>	2068	1196	3102	925	1	2176	9,81 %	29,85 %
<i>Gyrodactylus salaris</i>	1667	672	2198	960	0	1328	8,60 %	43,68 %
<i>Protopolystoma xenopodis</i>	4583	3712	7560	1425	8	6127	16,16 %	18,96 %
<i>Macrostomum lignano</i>	14217	1926	15401	7485	39	7877	13,01 %	48,85 %
<i>Schmditea mediterranea</i>	3667	674	4341	2322	1	2018	8,57 %	53,51 %

Se muestra, para todos los genomas empleados, el número de secuencias asignadas como secretadas por la vía canónica (es decir, que superaron el umbral establecido para NN-score de SignalP), por vía alterantiva (idem para SecretomeP), así como la cantidad de secuencias que integran ambas vías. Se muestra a su vez la cantidad de secuencias pertenecientes a ambas vías que corresponden a proteínas transmembrana (identificadas por TMHMM) o a integrar mitocondrias (identificado con TargetP). Por último, se muestra el porcentaje del proteoma perteneciente al secretoma, y la cantidad de secuencias descartadas dentro de las secuencias secretadas por ser o bien proteínas transmembrana o por ser proteínas integrantes de mitocondrias.

En la **Tabla 3** se observan los resultados obtenidos para el análisis de enriquecimiento en proteínas de secreción efectuado sobre los grupos de inparálogos detectados para cada especie.

Tabla 3. Análisis de enriquecimiento en proteínas de secreción para los inparálogos de cada genoma analizado.

Especie	<i>p-value</i>	FDR
<i>Diphyllobothrium latum</i> *	6.2e-05	1.2e-04
<i>Echinococcus canadensis</i>	0.13	0.16
<i>Echinococcus granulosus</i>	5.8e-02	7.5e-02
<i>Echinococcus multilocularis</i> *	1.3e-13	4.0e-12
<i>Hydatigera taeniaeformis</i> *	8.5e-10	3.7e-09
<i>Hymenolepis diminuta</i>	0.43	0.44
<i>Hymenolepis microstoma</i> *	2.2e-05	5.2e-05
<i>Hymenolepis nana</i> *	5.8e-05	1.2e-04
<i>Mesocestoides corti</i> *	4.0e-04	7.1e-04
<i>Schistocephalus solidus</i>	0.16	0.18
<i>Spirometra erinaceieuropaei</i> *	9.0e-07	2.7e-06
<i>Taenia asiatica</i> *	3.9e-10	2.3e-09
<i>Taenia saginata</i> *	1.5e-10	1.5e-09
<i>Taenia solium</i> *	2.9e-03	4.6e-03
<i>Clonorchis sinensis</i>	0.30	0.33
<i>Echinostoma caproni</i> *	7.6e-07	2.5e-06
<i>Fasciola hepatica</i> *	5.7e-03	8.2e-03
<i>Opistorchis viverrini</i> *	6.9e-08	2.6e-07
<i>Schistosoma curassoni</i> *	0.30	0.373
<i>Schistosoma haematobium</i> *	2.8e-05	6.1e-05
<i>Schistosoma japonicum</i> *	5.3e-11	8.0e-10
<i>Schistosoma mansoni</i> *	1.1e-05	3.0e-05
<i>Schistosoma margrebowiei</i> *	1.6e-03	2.7e-03
<i>Schistosoma mattheei</i>	1	1
<i>Schistosoma rodhaini</i> *	1.4e-05	3.5e-05
<i>Trichobilharzia regenti</i>	0.11	0.13
<i>Gyrodactylus salaris</i> *	3.2e-10	2.3e-09
<i>Protopolystoma xenopodis</i> *	3.5e-03	5.2e-03
<i>Macrostomum lignano</i>	4.9e-02	6.7e-02
<i>Schmidtea mediterranea</i> *	6.7e-10	3.4e-09

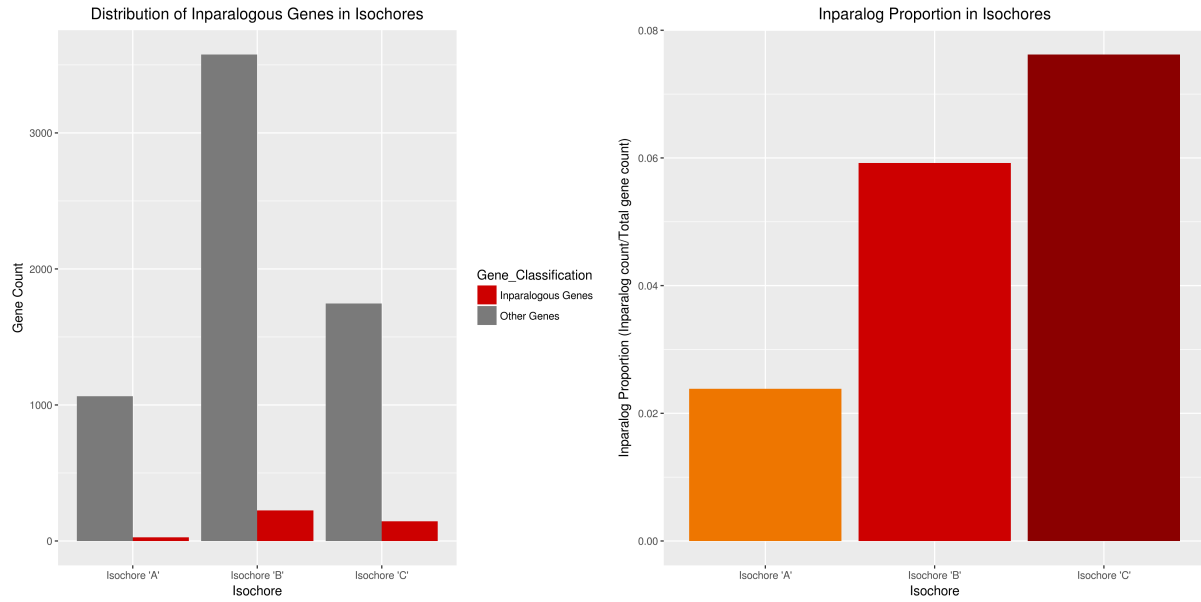
Los colores azul, rojo, verde y violeta indican la pertenencia de la especie a las clases Cestoda, Trematoda, Monogenea y Turbellaria (respectivamente). Con asteriscos se resaltan los resultados estadísticamente significativos (*FDR < 0.05*).

Mapeo de genes inparálogos a isocóros del genoma del tremátodo *Schistosoma mansoni*

Las **Figuras 6 y 7** muestran el resultado de asignar genes catalogados como inparálogos e inparálogos codificantes para proteínas pertenecientes al secretoma a bandas con un contenido en GC (*i.e.* estructuras de tipo isocóro) característico pertenecientes al genoma del tremátodo *S. mansoni* (trabajo previo de Lamolle et. al; ver sección Materiales y Métodos para mayor información). Se muestran a su vez la proporción de bases de estos genes en relación al tamaño de estas estructuras genómicas.

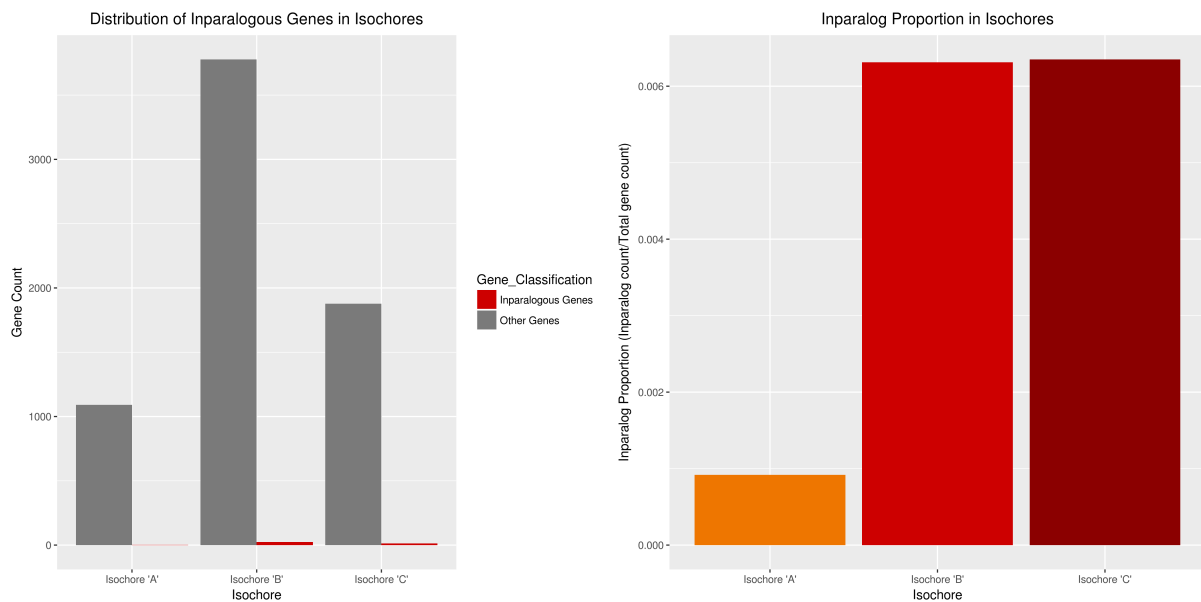
En particular, se asignaron dichos genes a los **isocóros A** (GC% < 33.5), **B** (33.5 < GC% < 36.5) y **C** (GC% > 36.5).

En lo que refiere al análisis de los genes inparálogos, un total de 26 genes (con un total de 107439 pb) fueron mapeados al isocóro A (de un total de 1091 genes codificados en 1983651 pb), 225 genes (codificados en 739218) fueron asignados al isocóro B (de un total de 3801 genes codificados en 6531129 pb) y 144 genes fueron asignados al isocóro C (de 1890 genes que abarcan 3078381 pb).



Figuras 6 y 7. Mapeo de genes inparálogos a isocóros en *S. mansoni*. Se muestra el conteo de genes inparálogos perteneciente a los isocóros A ($GC\% < 33.5$), B ($33.5 < GC\% < 36.5$) y C ($GC\% > 36.5$) (izquierda) y la proporción de bases que ocupan dichos genes respecto al total de bases codificantes en cada isocóro (derecha).

Para la el subgrupo de estos inparálogos que codifican para proteínas del secretoma, se logró asignar tan sólo 1 gen (de 387 pb) al isocóro A, 24 genes (con un total de 73167 pb) al isocóro B y 12 genes inparálogos (codificados en 44748) al isocóro C (Figuras 8 y 9).



Figuras 8 y 9. Mapeo de genes inparálogos codificantes para proteínas de secreción a isocóros en *S. mansoni*. Se muestra el conteo de genes perteneciente a los isocóros A ($GC\% < 33.5$), B ($33.5 < GC\% < 36.5$) y C ($GC\% > 36.5$) (izquierda) y la proporción de bases que ocupan dichos genes respecto al total de bases codificantes en cada isocóro (derecha).

Análisis de enriquecimiento en términos funcionales

El conjunto de secuencias inparálogas estudiadas para cada clase del phylum se ve enriquecido en funciones moleculares similares (**Tabla 4**). En el caso de las clases netamente parásitas, se observa un enriquecimiento en funciones moleculares vinculadas a la obtención de nutrientes y especies químicas de relevancia (como lo son la unión a lípidos e iones) que en principio estos organismos parasitan de su huésped. Se observa también, en todas las clases, un enriquecimiento en términos vinculados al intercambio de productos con el medio (*i.e.* transporte transmembrana). La consideración de los términos ontológicos para los cuales se ve enriquecido cada uno de los organismos estudiados (ver Anexo) muestra a su vez casos donde es posible ver enriquecimiento en términos vinculados a la unión de glúcidos, alcoholes y aminoácidos, además de los anteriores. Se observa a su vez un enriquecimiento en funciones moleculares vinculadas a actividad enzimática variada, con posible vinculación a procesos de remodelación del medio extracelular, la maduración de proteínas y la interacción con biomoléculas del medio/huésped (*e.g.* actividad peptidasa, transferasa e hidrolasa). También es posible vincular a algunos de los términos con procesos de señalización celular (actividad quinasa y fosfatasa) y de modulación de la expresión génica y de traducción (unión a ARN y ARNm, unión a histonas y actividad metiltransferasa). Algunos de los organismos donde se observan un enriquecimiento en este tipo de funciones moleculares reportan un enriquecimiento en términos ontológicos de procesos biológicos vinculados a la diferenciación celular y el desarrollo, como lo son “pigment cell differentiation”, “digestive tract development” o “pharynx development” (ver Anexo). No obstante, los términos ontológicos vinculados a procesos biológicos deben ser tomados con recaudo, ya que funciones moleculares similares son ejercidas en diferentes contextos, por lo cual es posible vincular un mismo GO term referente a una función molecular con procesos biológicos que no poseen relevancia en un contexto dado.

Tabla 4. Funciones moleculares enriquecidas en los conjuntos de inparálogos analizados.

Cestoda		Porcentaje mapeado a GO Slims: 70,4%
lipid binding (GO:0008289)	transferase activity, transferring glycosyl groups (GO:0016757)	
peptidase activity (GO:0008233)	oxidoreductase activity (GO:0016491)	
lyase activity (GO:0016829)	ion binding (GO:0043167)	
transferase activity, transferring acyl groups (GO:0016746)	nuclease activity (GO:0004518)	
kinase activity (GO:0016301)	hydrolase activity, acting on glycosyl bonds (GO:0016798)	
transmembrane transporter activity (GO:0022857)	translation factor activity, RNA binding (GO:0008135)	
structural constituent of ribosome (GO:0003735)	structural molecule activity (GO:0005198)	
transferase activity, transferring alkyl or aryl (other than methyl) groups (GO:0016765)	ligase activity (GO:0016874)	
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds (GO:0016810)		
Trematoda		Porcentaje mapeado a GO Slims: 82,6%
transferase activity, transferring glycosyl groups (GO:0016757)	ATPase activity (GO:0016887)	
transmembrane transporter activity (GO:0022857)	ion binding (GO:0043167)	
peptidase activity (GO:0008233)	lipid binding (GO:0008289)	
structural molecule activity (GO:0005198)	kinase activity (GO:0016301)	
histone binding (GO:0042393)	cytoskeletal protein binding (GO:0008092)	
oxidoreductase activity (GO:0016491)	isomerase activity (GO:0016853)	
hydrolase activity, acting on glycosyl bonds (GO:0016798)	enzyme binding (GO:0019899)	
lyase activity (GO:0016829)		
Monogenea		Porcentaje mapeado a GO Slims: 100,0%
nuclease activity (GO:0004518)	transferase activity, transferring glycosyl groups (GO:0016757)	
transmembrane transporter activity (GO:0022857)	structural molecule activity (GO:0005198)	
cytoskeletal protein binding (GO:0008092)	translation factor activity, RNA binding (GO:0008135)	
peptidase activity (GO:0008233)	kinase activity (GO:0016301)	
oxidoreductase activity (GO:0016491)		
Turbellaria		Porcentaje mapeado a GO Slims: 75,0%
mRNA binding (GO:0003729)	ATPase activity (GO:0016887)	
transmembrane transporter activity (GO:0022857)	structural constituent of ribosome (GO:0003735)	
methyltransferase activity (GO:0008168)	oxidoreductase activity (GO:0016491)	
transferase activity, transferring glycosyl groups (GO:0016757)	phosphatase activity (GO:0016791)	
hydrolase activity, acting on glycosyl bonds (GO:0016798)		

Se separaron los datos para cada clase (Cestoda, Trematoda, Monogenea y Turbellaria). Se muestran los GO Slims (GO terms de carácter general, ver Materiales y Métodos) a los cuales mapean los GO terms en los que se vió enriquecido cada organismo de estas clases. Se consideraron p-values < 0.005. Se muestra a su vez el porcentaje de términos que se logró mapear a GO Slims. En Anexo el lector encontrará el detalle de todos los términos ontológicos para los que se encontró enriquecimiento, considerándose estos para cada genoma.

La tendencia observada se mantiene cuando se restringe el análisis a aquellas secuencias de inparálogos que se predice son secretadas (Tabla 5), y a aquellas que además es posible estén evolucionando bajo un régimen de selección natural positiva (Tabla 6). Acorde a lo esperado, a medida que se restringe el conjunto de secuencias a aquellas que son secretadas se observa como tendencia general un enriquecimiento en términos de localización extracelular o de vesículas de secreción (ver Anexo, donde se muestra el enriquecimiento para términos ontológicos específicos en cada especie). No obstante, es importante recalcar que en muchos casos se observa enriquecimiento en compartimientos intracelulares también, existiendo organismos para los cuales directamente no se observan términos ontológicos de localización celular que refieran al medio externo (ver sección Discusión para un detalle de posibles explicaciones para este hecho).

En la sección Anexo el lector encontrará en detalle el resultado de los análisis de enriquecimiento en términos ontológicos para todos los organismos estudiados, considerándose los términos vinculados a procesos biológicos, función molecular y componente celular.

Tabla 5. Funciones moleculares enriquecidas en el subconjunto de inparálogos secretados	
Cestoda	Porcentaje mapeado a GO Slims: 65,9%
peptidase activity (GO:0008233)	transferase activity, transferring glycosyl groups (GO:0016757)
enzyme binding (GO:0019899)	ion binding (GO:0043167)
DNA binding (GO:0003677)	nuclease activity (GO:0004518)
lyase activity (GO:0016829)	nucleotidyltransferase activity (GO:0016779)
structural molecule activity (GO:0005198)	phosphatase activity (GO:0016791)
transmembrane transporter activity (GO:0022857)	ligase activity (GO:0016874)
Trematoda	Porcentaje mapeado a GO Slims: 77,8%
translation factor activity, RNA binding (GO:0008135)	cytoskeletal protein binding (GO:0008092)
ATPase activity (GO:0016887)	ion binding (GO:0043167)
peptidase activity (GO:0008233)	lipid binding (GO:0008289)
lyase activity (GO:0016829)	ligase activity (GO:0016874)
transmembrane transporter activity (GO:0022857)	kinase activity (GO:0016301)
oxidoreductase activity (GO:0016491)	hydrolase activity, acting on glycosyl bonds (GO:0016798)
isomerase activity (GO:0016853)	phosphatase activity (GO:0016791)
structural molecule activity (GO:0005198)	methyltransferase activity (GO:0008168)
Monogenea	Porcentaje mapeado a GO Slims: 71,4%
nuclease activity (GO:0004518)	ubiquitin-like protein binding (GO:0032182)
ion binding (GO:0043167)	peptidase activity (GO:0008233)
rRNA binding (GO:0019843)	
Turbellaria	Porcentaje mapeado a GO Slims: 60,0%
peptidase activity (GO:0008233)	oxidoreductase activity (GO:0016491)

Tabla 6. Funciones moleculares enriquecidas en el subconjunto de inparálogos secretados evolucionando bajo selección positiva.	
Cestoda	Porcentaje mapeado a GO Slims: 69,2%
peptidase activity (GO:0008233)	transferase activity, transferring glycosyl groups (GO:0016757)
ion binding (GO:0043167)	DNA binding (GO:0003677)
nuclease activity (GO:0004518)	lyase activity (GO:0016829)
enzyme binding (GO:0019899)	transmembrane transporter activity (GO:0022857)
DNA-binding transcription factor activity (GO:0003700)	ligase activity (GO:0016874)
Trematoda	Porcentaje mapeado a GO Slims: 85,7%
transferase activity, transferring glycosyl groups (GO:0016757)	structural molecule activity (GO:0005198)
peptidase activity (GO:0008233)	lipid binding (GO:0008289)
nuclease activity (GO:0004518)	oxidoreductase activity (GO:0016491)
transmembrane transporter activity (GO:0022857)	phosphatase activity (GO:0016791)
ion binding (GO:0043167)	lyase activity (GO:0016829)
ATPase activity (GO:0016887)	cytoskeletal protein binding (GO:0008092)
Monogenea	Porcentaje mapeado a GO Slims: 50,0%
peptidase activity (GO:0008233)	
Turbellaria	Porcentaje mapeado a GO Slims: 66,7%
transferase activity, transferring glycosyl groups (GO:0016757)	peptidase activity (GO:0008233)
oxidoreductase activity (GO:0016491)	enzyme binding (GO:0019899)

Detección de genes inparálogos evolucionando en régimen de selección natural positiva

Un total de 3775 grupos de inparálogos fueron anotados como candidatos a estar evolucionando bajo un régimen de selección natural positiva. Los mismos están conformados en su conjunto por 13030 secuencias codificantes. De las mismas, entre el 3,6 % y el 58,3 % son predichas como proteínas de secreción, aunque estos números extremos corresponden a organismos donde un bajo número de grupos de inparálogos fueron detectados como seleccionados positivamente. Descartando estos casos no se observan grandes discrepancias en el porcentaje de las secuencias codificantes que pertenecen al secretoma, oscilando este número entre el 15 % y el 25 % aproximadamente. Cabe destacar que en múltiples casos, al intentar identificar qué codifican estas secuencias, se encuentra o bien que las mismas se encuentran anotadas como proteínas hipotéticas, o bien no se obtiene similitud considerable con ninguna secuencia anotada en el conjunto RefSeq de la base de datos del NCBI. (**Tabla 7**).

Tabla 7. Conjunto de inparálogos con posible evolución bajo selección positiva en los genomas estudiados.

Especie	Núm. de grupos de inparálogos	Núm. de genes	Porcentaje de genes secretados	Numero de genes con los que se encontro homologia	Porcentaje mapeado a proteínas hipotéticas
<i>Diphyllobothrium latum</i>	48	117	20(17,1%)	1	0
<i>Echinococcus canadensis</i>	11	47	7(14,9%)	12	8
<i>Echinococcus granulosus</i>	1	2	0(0,0%)	2	2
<i>Echinococcus multilocularis</i>	22	90	23(25,6%)	73	18
<i>Hydatigera taeniaeformis</i>	4	12	7(58,3%)	0	0
<i>Hymenolepis diminuta</i>	18	64	5(7,8%)	1	0
<i>Hymenolepis microstoma</i>	54	179	43(24,0%)	163	101
<i>Hymenolepis nana</i>	71	236	28(11,9%)	20	9
<i>Mesocestoides corti</i>	32	91	19(20,9%)	1	0
<i>Schistocephalus solidus</i>	137	540	71(13,1%)	5	0
<i>Spirometra erinaceieuropaei</i>	349	845	101(12,0%)	19	1
<i>Taenia asiatica</i>	28	71	20(28,2%)	7	1
<i>Taenia saginata</i>	49	129	20(15,5%)	6	0
<i>Taenia solium</i>	11	27	7(25,9%)	3	0
<i>Clonorchis sinensis</i>	26	160	12(7,5%)	155	102
<i>Echinostoma caproni</i>	59	198	16(8,1%)	2	0
<i>Fasciola hepatica</i>	58	206	35(17,0%)	16	5
<i>Opistorchis viverrini</i>	30	111	34(30,6%)	111	111
<i>Schistosoma curussoni</i>	13	29	1(3,4%)	6	2
<i>Schistosoma haematobium</i>	9	32	8(25,0%)	29	24
<i>Schistosoma japonicum</i>	40	169	30(17,8%)	56	20
<i>Schistosoma mansoni</i>	22	66	10(15,2%)	62	27
<i>Schistosoma margrebowiei</i>	86	375	36(9,6%)	38	21
<i>Schistosoma mattheei</i>	11	28	1(3,6%)	6	5
<i>Schistosoma rodhaini</i>	25	99	11(11,1%)	14	8
<i>Trichobilharzia regenti</i>	27	118	9(7,6%)	2	0
<i>Gyrodactylus salaris</i>	14	35	10(28,6%)	0	0
<i>Protopolystoma xenopodis</i>	59	222	38(17,1%)	1	0
<i>Macrostomum lignano</i>	2390	8556	999(11,7%)	3580	3573
<i>Schnditea mediterranea</i>	71	176	21(11,9%)	5	0

Dentro de las secuencias que si exhiben una similitud considerable con secuencias de esta base de datos (*i.e.* una identidad de al menos 75 %, con una cobertura de al menos 40 %), se observa un conjunto variado de proteínas, entre las que se incluyen posibles ortólogos (**Tabla 8**). Cabe destacarse la presencia de múltiples proteínas reportadas como inmunogénicas o inmunomoduladoras (diagnostic antigen gp50, cytoplasmic antigen 1, Tetraspanin-5), vinculadas a transposones (gag pol polyprotein, Mariner Mos1 transposase, retrovirus-related Pol polyprotein from type-1 retrotransposable element 3 R2). Se observa también la presencia de proteínas de unión o de transporte (Sodium channel protein para, ferritin-like protein, solute carrier family 33, Bile salt export pump), así como también de proteasas reportadas como vinculadas a procesos de invasión de huésped (Venom allergen 5, secreted cathepsin L2, invadolysin (M08 family)). Por último, es también resaltante la presencia de proteínas vinculadas al desarrollo celular y morfológico (homeobox protein, axis inhibition protein A). Se observa a su vez un ortólogo de una subunidad de la proteína SRP. Tal resultado puede estar vinculado a la coevolución de esta subunidad con secuencias codificantes para proteínas de secreción, hipótesis que implicaría un estudio particular de dicho caso.

Tabla 8. Secuencias homólogas a un subconjunto de inparálogos que evolucionan bajo selección positiva.

Anotación del gen	Núm. de genes
<i>E. canadensis</i>	
EUB54891.1 Multidrug and toxin extrusion protein [Echinococcus granulosus]	8
<i>E. multilocularis</i>	
CDI96652.2 diagnostic antigen gp50 [Echinococcus multilocularis]	6
CDI96674.1 cytoplasmic antigen 1 [Echinococcus multilocularis]	9
CDI96790.2 rRNA promoter binding protein [Echinococcus multilocularis]	4
CDS35425.1 homeobox protein [Echinococcus multilocularis]	4
CDS41803.1 fibronectin type III domain-containing protein [Echinococcus multilocularis]	4
CDI96497.1 lactate dehydrogenase protein [Echinococcus multilocularis]	5
CDS36262.1 Tetraspanin-5 [Echinococcus multilocularis]	6
<i>H. microstoma</i>	
CDS28967.1 diagnostic antigen gp50 [Hymenolepis microstoma]	2
CDS31987.1 transposase [Hymenolepis microstoma]	2
CDS29319.1 gag pol polyprotein [Hymenolepis microstoma]	2
CDS25455.1 histone lysine N methyltransferase SETMAR [Hymenolepis microstoma], CDS26545.1 Mariner Mos1 transposase [Hymenolepis microstoma]	2
CDS29388.1 proprotein convertase subtilisin:kexin type 5 [Hymenolepis microstoma]	2
CDS27007.1 signal recognition particle receptor subunit [Hymenolepis microstoma]	2
<i>H. nana</i>	
CDS26143.1 beta 13 n galactosyltransferase [Hymenolepis microstoma]	6
CDS33104.1 reverse transcriptase [Hymenolepis microstoma]	3
CDS33626.1 solute carrier family 5 [Hymenolepis microstoma]	2
CDS26502.2 sodium and chloride dependent glycine [Hymenolepis microstoma]	2
CDS29949.1 solute carrier family 33 [Hymenolepis microstoma]	4
<i>M. corti</i>	
CDS19936.1 kinesin light chain 4 [Echinococcus granulosus]	2
<i>S. solidus</i>	
XP_005107471.1 PREDICTED: craniofacial development protein 2-like [Aplysia californica]	2
AFM74212.1 ferritin-like protein [Spirometra erinaceieuropaei]	3
CDS24615.1 myosin heavy chain [Echinococcus granulosus/E. multilocularis]	2
<i>S. erinaceieuropaei</i>	
EUB64307.1 Cell cycle checkpoint protein RAD1 [Echinococcus granulosus]	2
EUB64148.1 Sodium channel protein para [Echinococcus granulosus]	4
AFM74210.1 tenascin-C [Spirometra erinaceieuropaei]	4
CDS33789.1 filamin [Hymenolepis microstoma]	2
CDS41468.1 phosphoglycerate mutase [Echinococcus multilocularis]	2
EUB60796.1 AP-1 complex subunit gamma-1 [Echinococcus granulosus]	4
XP_008294557.1 PREDICTED: WD repeat-containing protein 5-like [Stegastes partitus]	2
<i>T. asiatica</i>	
AAW88556.1 oncosphere protein Tso31a [Taenia solium]	2
CDS24815.1 lon protease mitochondrial [Echinococcus granulosus]	3
CDS35735.1 neutral amino acid transporter A [Echinococcus multilocularis]	3
EUB54936.1 Arginase-2 [Echinococcus granulosus]	2
EUB57861.1 Heat shock cognate protein [Echinococcus granulosus]	2
<i>T. saginata</i>	
EUB57021.1 Heat shock cognate protein [Echinococcus granulosus]	3
CUT99846.1 histone H2A [Echinococcus multilocularis]	2
CDS24809.1 histone H2B, partial [Echinococcus granulosus]	2
<i>T. solium</i>	
EUB57161.1 Puromycin-sensitive aminopeptidase [Echinococcus granulosus]	2
EUB55282.1 Estradiol 17 beta-dehydrogenase [Echinococcus granulosus]	2
<i>C. sinensis</i>	

GAA51502.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis]	3
GAA47942.1 ATP-binding cassette transporter [Clonorchis sinensis]	2
GAA39798.2 amine GPCR [Clonorchis sinensis]	2
GAA48255.1 protein disulfide-isomerase, partial [Clonorchis sinensis]	7
<i>F. hepatica</i>	
PIS77885.1 papain family cysteine protease [Fasciola hepatica]	2
PIS79113.1 transporter, major facilitator family protein [Fasciola hepatica]	2
ABQ95351.1 secreted cathepsin L2 [Fasciola hepatica]	4
PIS83512.1 putative chaperone protein DnaK [Fasciola hepatica]	2
AAF88069.1 amoebapore-like protein [Fasciola hepatica]	2
PIS79273.1 epididymal secretory protein E1 family protein [Fasciola hepatica]	3
<i>S. curassoni</i>	
XP_012801036.1 Cathepsin B-like cysteine proteinase [Schistosoma haematobium]	2
<i>S. haematobium</i>	
XP_012792511.1 Venom allergen 5 [Schistosoma haematobium]	2
XP_012792756.1 Ca(2+)/calmodulin-responsive adenylate cyclase [Schistosoma haematobium]	2
<i>S. japonicum</i>	
CAX73628.1 6-glucosidase [Schistosoma japonicum]	2
CAX74030.1 Glutamic acid-rich protein precursor [Schistosoma japonicum]	4
CAX74194.1 TonB box, N-terminal, domain-containing protein [Schistosoma japonicum]	2
CAX74232.1 Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 [Schistosoma japonicum]	2
<i>S. mansoni</i>	
XP_018652502.1 putative alpha-amylase [Schistosoma mansoni]	2
XP_018645754.1 rna recognition motif (rrm) domain containing protein	4
XP_018647223.1 egg protein CP391S-like [Schistosoma mansoni]	5
XP_018653704.1 putative tropomyosin [Schistosoma mansoni]	6
XP_018646757.1 MEG-3 (Grail) family [Schistosoma mansoni]	2
XP_018648790.1 putative fgf receptor activating protein [Schistosoma mansoni]	2
XP_018646414.1 invadolysin (M08 family) [Schistosoma mansoni]	2
<i>S. margrebowiei</i>	
XP_012801036.1 Cathepsin B-like cysteine proteinase [Schistosoma haematobium]	2
ADG95879.1 alkaline phosphatase [Schistosoma mansoni]	4
XP_012797173.1 Protein CASP [Schistosoma haematobium]	3
XP_012799362.1 Bile salt export pump [Schistosoma haematobium]	3
CAX69930.1 LanC-like protein 1 [Schistosoma japonicum]	2
<i>S. rodhaini</i>	
XP_012801017.1 Titin [Schistosoma haematobium]	2
<i>T. regenti</i>	
XP_012796759.1 Spectrin alpha chain [Schistosoma haematobium]	3
<i>M. lignano</i>	
XP_013395914.1 homeobox protein prophet of Pit-1 [Lingula anatina]	4
ALD16261.1 Peptidyl-prolyl cis-trans isomerase FKBP6 [Macrostomum lignano]	4
XP_021373861.1 ELAV-like protein 3 isoform X3 [Mizuhopecten yessoensis]	3
PFX20388.1 Tubulin alpha-1D chain [Stylophora pistillata]	5
<i>S. mediterranea</i>	
ALJ75660.1 ASH2 [Schmidtea mediterranea]	3
AHD24739.1 IFI6-like protein, partial [Schmidtea mediterranea]	2
ADW27177.1 axis inhibition protein A [Schmidtea mediterranea]	2
AKN21403.1 slc2a-9, partial [Schmidtea mediterranea]	2
AFJ24785.1 vault protein inter alpha trypsin-1	6

No obstante, otras secuencias muestran posible homología con proteínas cuyo rol es no trivial en el contexto de estudio (*e.g.* Estradiol 17 beta-dehydrogenase, kinesin light chain 4). Un estudio detallado de estos casos es por tanto necesario para dilucidar si se tratan en efecto de secuencias codificantes para proteínas con un rol adaptativo.

Discusión

La duplicación de genes codificantes para proteínas de secreción tendría un rol adaptativo en el filo Platyhelminthes

En este trabajo se observó, tanto en genomas de parásitos como en el de organismos de vida libre, que una enorme cantidad de grupos de inparálogos se ven enriquecidos en genes codificantes para proteínas de secreción.

Este hecho puede tener base en la amplia gama de funciones desempeñadas por las proteínas pertenecientes a esta clase: las mismas son componentes clave no sólo de la interacción huésped-patógeno (como ya se destacó en la sección Introducción de esta tesis), sino también en procesos biológicos comunes a todo organismo como lo son el desarrollo, procesos digestivos, inmunológicos y de señalización celular, entre otros.

Es de esperar que resultados similares se obtengan si se analiza, por ejemplo, el enriquecimiento en inparálogos codificantes para proteínas transmembrana, las cuales también han sido reportadas como fundamentales en estos procesos en todos los dominios de la vida.

La existencia de duplicaciones génicas ha sido vinculada en múltiples trabajos a procesos adaptativos, argumentándose que estos genes codifican para proteínas con alta relevancia funcional. No obstante, discernir si en efecto la selección natural positiva opera sobre estas duplicaciones linaje-específicas requiere no sólo un análisis funcional (ya sea experimental o de enriquecimiento de estos genes en términos funcionales), si no también del análisis de evolución de las secuencias en sí. Esto implica la evaluación de indicadores como lo son la tasa de sustitución no-sinónima respecto a la de sustituciones sinónimas en dichas secuencias (dN/dS).

En este trabajo tanto los resultados obtenidos para el enriquecimiento en términos ontológicos como de análisis de evolución molecular para los diferentes grupos de secuencias inparálogas codificantes para proteínas de secreción apoyan la idea de que en efecto la evolución de dichas secuencias se ve moldeada al menos en parte por la selección natural positiva.

En términos específicos, para los grupos de genes codificantes inparálogos de los genomas estudiados se observa:

i) Un enriquecimiento en términos ontológicos vinculados a procesos de relevancia tanto en organismos de vida libre (como pueden ser procesos de adquisición de nutrientes, metabólicos y de adhesión a sustratos) como en parásitos (como pueden ser de regulación de la respuesta inmune de huésped).

La principal fortaleza de realizar un análisis de enriquecimiento en términos ontológicos en las secuencias estudiadas reside en que el mismo tiene en cuenta incluso aquellas proteínas para las que no es posible inferir ningún tipo de homología respecto a proteínas anotadas actualmente en bases de datos como RefSeq.

Tal y cómo se ve en la subsección *Detección de genes inparálogos evolucionando en régimen de selección natural positiva* (de la sección Resultados), el conjunto de proteínas para las que se observa un proceso adaptativo se ve enriquecida en términos asociados a funciones moleculares de unión o transporte de metabolitos, nutrientes e iones (entre los cuales se encuentran alcoholes, glúcidos, ácidos grasos y amino ácidos, así como zinc y otros iones de relevancia bioquímica).

Esto parece lógico si se considera que todos los organismos parásitos deben obtener muchos de estos nutrientes de sus huéspedes al haber perdido la capacidad de codificar múltiples proteínas de la maquinaria biosintética. Un ejemplo concreto es el de céstodos, para los cuales ya no existen genes codificantes para enzimas pertenecientes a rutas funcionales de biosíntesis de ácidos grasos. Este fenómeno no es exclusivo de parásitos, ya que en organismos de vida libre si bien se conservan rutas de biosíntesis también existe la necesidad de adquirir y modificar nutrientes obtenidos del medio ambiente.

La presencia de términos funcionales asociados a la modificación de glúcidos también resulta coherente, si se tiene en cuenta que los entornos extracelulares que recubren las mucosas de huéspedes, parásitos y organismos en general se ven cubiertas por glicoproteínas esenciales en procesos de adhesión, reconocimiento celular y ejecución de funciones variadas.

De especial interés resulta la presencia de términos asociados a actividad proteolítica (ya sea a través de peptidasas de tipo cisteín o serin-proteasa) y reguladores e inhibidores enzimáticos. Como ya se ha dicho, se ha demostrado que este tipo de proteínas juegan un rol en la interfaz que separa a estos organismos de su medio externo. La relevancia de las mismas es aún mayor en el caso de los parásitos, en donde se ha reportado que las mismas modulan la actividad inmunológica de sus huéspedes como estrategia adaptativa de sobrevida.

Siguiendo esta línea de razonamiento, la presencia de términos vinculados a procesos biológicos tales como “producción de anticuerpos en tejidos mucosos” resulta también interesante, aunque se deben tratar los mismos con mayor cuidado que aquellos términos que refieren funciones moleculares, ya que en muchos casos las proteínas asociadas a este tipo de términos funcionales son variadas, e incluso debe ser contemplado el uso de “calificadores” que modifican el sentido de la descripción brindada por el término ontológico (en el ejemplo citado, es posible que esté asociado el calificador “inhibidor”, alterando completamente la interpretación).

No obstante, y como demuestran Demuth y Hahn (Demuth & Hahn, 2009), el análisis de estos términos funcionales se ve afectado por errores incluso a nivel del alineamiento de secuencias empleadas en el análisis. Los autores advierten en dicho trabajo que, debido a la enorme cantidad de falsos positivos y fluctuaciones en los resultados obtenidos al momento de detectar selección natural positiva, el análisis de enriquecimiento funcional posterior que se suele realizar debe ser tomado con cautela.

ii) La existencia de una tasa dN/dS > 1 en familias de genes codificantes, los cuales presentan posible homología a genes codificantes para proteínas ya reportadas como fundamentales para la persistencia de organismos parásitos en su huésped, así como para proteínas que suscitan una fuerte reacción por parte del sistema inmune de estos.

En efecto, muchas de las proteínas codificadas en genes inparálogos para las que se postula un efecto adaptativo presentan similitud de secuencia con proteínas ya reportadas como importantes en la interacción huésped-parásito y para las cuales se observa una respuesta fuerte por parte del sistema inmune de los hospederos. Entre las mismas se puede nombrar a la **tetraspanina** (Hu et al. 2015), al **antígeno GP50** (Huang et al. 2016; Huang et al. 2017; Ahn et al. 2017), el **antígeno p40** (Nene et al. 1986) presente en huevos de *S. mansoni*, **SmVal** (Yoshino et al., 2014), la **proteína de la oncósfera Tso31a** (Mayta et al. 2007) presente en *T. solium* o las proteasas **papaina** (Lowther et al. 2009) y **catepsina L** (Collins et al. 2004), entre otros.

Un estudio de estos casos permitirá distinguir si en efecto lo observado se trata a que ortólogos de estas proteínas en diferentes organismos juegan papeles adaptativos en la fisiología de estos parásitos.

iii) La tendencia de genes inparálogos dentro del genoma de *S. mansoni* a ubicarse dentro de isocóros de alto contenido GC, los cuales están asociados a regiones con una alta tasa de replicación.

Respecto a este punto ver la subsección *Distribución de inparálogos en isocoros del genoma de Schistosoma mansoni: una hipótesis mecanística*, donde se aborda el tema en más detalle.

Análisis de la evolución de inparálogos a la luz de algunos modelos

Es posible distinguir dos escenarios respecto al carácter adaptativo de la duplicación génica en sí misma: por un lado, modelos clásicos como el de neofuncionalización y subfuncionalización asumen el aumento en un número de copias como un hecho neutral en sí mismo, el cual no confiere una ventaja adaptativa al organismo que lo posee. No obstante, también es posible plantear el caso en el cual la duplicación génica en sí misma es blanco de selección (ver sección Introducción para más detalles).

En los modelos de neofuncionalización y subfuncionalización la elevada tasa de proteínas de secreción se explicaría gracias a la ventaja selectiva que confieren dichas copias al organismo. Así, la fijación de las copias en la población se da con base en el fenómeno de deriva génica, con posterior retención de las copias que posean ya sea o bien una nueva función o bien dividiendo funciones ancestrales entre ambas copias, aumentando la efectividad de las mismas.

Para ejemplificar el primero de estos casos, podemos suponer una duplicación génica de una proteasa involucrada en el proceso de invasión de un platelminto parásito; si se fijara de forma neutra esta duplicación y un conjunto de mutaciones llevara a una nueva proteasa que logre invadir otro tipo de tejidos, la ventaja selectiva proporcionada por la capacidad de invadir nuevos nichos (*i.e.* otros hospederos) llevaría a la fijación de esta mutación en la población, incluso quizá llevando al comienzo de un proceso que culmine en la especiación.

De la misma forma, se ha postulado a la subfuncionalización como un proceso que permite generar una diversidad proteica tal que se pueda modular la expresión espacial o temporal de una determinada función ancestral. Esta idea resulta interesante cuando se piensa en la cantidad de procesos a los cuales se enfrenta un parásito a lo largo de su vida, donde se recorren múltiples tejidos a lo largo del desarrollo (pensar, por ejemplo, en el ciclo de vida del tremátodo *Fasciola hepatica*).

Si bien en ambos casos se puede pensar que las condiciones impuestas para la fijación de las copias (y por tanto para el establecimiento del enriquecimiento observado) son demasiado conservadoras, lo cierto es que Lynch et al (Lynch et al. 2002) han demostrado que bajo tal escenario lo que se requiere es que se cumpla la condición $\frac{1}{N_g} < \frac{s^2}{8}$ (donde N_g es el número efectivo de genes por locus y s el coeficiente de selección). Tal condición podría cumplirse en parásitos como los platelmintos, donde existen múltiples instancias a lo largo del ciclo vital donde el número de organismos es superlativo.

No obstante, resulta interesante considerar como plausible el escenario planteado por Francino (Francino, 2005), donde el aumento de número de copias en sí resulta en primera instancia en una ventaja adaptativa para el organismo en cuestión.

Si se observa la lista de términos ontológicos para los cuales se existe un enriquecimiento, se notará que una fracción considerable de los mismos podrían aplicar a esta perspectiva. Así, genes codificantes para catabolizadores (e.g. “glycolipid catabolic process”, “membrane lipid catabolic process”,), transportadores (“transepithelial transport”), proteínas de respuesta a estímulos externos (“response to vitamin D”), de síntesis de material extracelular (“proteoglycan biosynthetic process”), o proteasas posean en primera instancia un valor adaptativo gracias al aumento de dosaje, el cual en algunas ocasiones puede llevar a una mayor eficiencia para de los procesos considerados.

Así por ejemplo, en una primera instancia, si se posee un gen codificante para un transportador de un nutriente presente en el medio interno del hospedero, el cual no posee una eficiencia óptima, un aumento en el número de copias puede llevar en primera instancia a una mayor obtención del nutriente, lo cual brinda al organismo considerado una ventaja selectiva. Posteriormente, la selección natural positiva puede actuar sobre las copias. Si en este proceso se generan variantes de alta eficiencia, las mismas se conservarán y es de esperar que exista un proceso de pérdida de los genes codificantes para las copias subóptimas.

De ser cierto en algunos casos este escenario, se estaría observando la primer fase de este proceso (es decir, el aumento en número de copias con posterior acción de selección natural positiva).

A su vez, el aumento de dosaje no sólo se puede observar por la mantención de un aumento en el número de copias, sino también por procesos selectivos que actúen sobre secuencias reguladoras y lleven a un aumento en la expresión de la proteína codificada, tal y como reportaron Raymond y colaboradores (Raymond et al., 1998) para el caso del insecto *Culex pipens*, donde se documentó de forma completa tanto la amplificación en el número de copias como la selección positiva operando a nivel de secuencias reguladoras para la expresión de genes que confieren resistencia a insecticidas. Un análisis del rol que juega la selección natural positiva en la evolución de secuencias reguladoras para estas proteínas es por tanto pertinente.

Fuentes de error

Pese a todo lo descrito en la subsección anterior, múltiples factores deben ser tenidos en cuenta a fin de considerar explicaciones alternativas que den cuenta de los resultados observados.

Cuatro principales fuentes de error deben, en principio, ser tenidas en cuenta:

i) Calidad de los genomas analizados, ya sea en cuanto al ensamblado de los mismos como a la anotación.

Varios son los posibles inconvenientes asociados a la utilización de genomas de pobre calidad. En primer lugar, existe la posibilidad de que una baja cobertura del mismo lleve al colapso de *reads* pertenecientes a dos copias diferentes en un único gen. Siendo este el escenario se estaría, no obstante, frente a una subestimación del número de inparálogos, lo cual en principio no debería afectar a los cálculos de enriquecimiento realizados. De hecho, el efecto neto es el de ir contra la hipótesis planteada, ya que se estarían detectando menos inparálogos de secreción que los existentes de cumplirse la premisa.

No obstante, también debe considerarse que la existencia de *contigs* con genes partidos puede llevar a una asignación errónea de homología, debido al carácter de modularidad que posee una gran cantidad de las proteínas codificadas por el genoma de organismos eucariotas. Así, es posible que se asigne como miembros de una misma familia multigénica a dos proteínas que en efecto no son homólogas debido a la presencia de un dominio con similaridad de secuencia y a la falta del resto de los dominios codificados por el gen presente en el organismo.

Del mismo modo debe considerarse la presencia de genes en *contigs* partidos. Si el largo de la secuencia resultante de tal fragmentación lleva a que no se satisfagan los valores mínimos de cobertura establecidos, esto puede llevar a que el gen en cuestión no se anote en ninguna familia multigénica. Nuevamente, esto llevaría a una subestimación del número de secuencias inparálogas.

De todos modos, en principio no existe un motivo para presumir que dentro del conjunto de proteínas truncadas exista una prevalencia de motivos asociados a proteínas de secreción; esto lleva a pensar que se verían afectadas de igual forma secuencias codificantes para proteínas secretadas y proteínas no secretadas, quedando los análisis de enriquecimiento realizados libres de sesgo. Es posible evaluar esta noción de forma sencilla, eliminando previamente aquellos *contigs* donde se detecte que las proteínas anotadas se encuentran sobre el comienzo o fin de los mismos y volviendo a realizar dichos análisis.

Respecto a la anotación de genomas, dos problemas revisten interés en el marco de este trabajo: en primer lugar es posible que exista una subanotación de secuencias, lo cual colabore a no detectar eventos adaptativos de relevancia para este estudio. Nuevamente, este escenario sólo plantea una subestimación del efecto analizado por la hipótesis central del trabajo, por lo cual en última instancia la gravedad de la misma es relativa.

El segundo escenario posible es que los indicadores de fenómenos adaptativos detectados sean un artefacto de una anotación sobrerrepresentada por secuencias de interés. Así, por ejemplo, la detección de un enriquecimiento en términos ontológicos vinculados a la respuesta inmune, el metabolismo o procesos de desarrollo se podrían explicar por falta de anotación de otros términos funcionales al momento.

Dos indicadores parecen, no obstante, aportar en este sentido. En primer lugar, y como muestra la **Figura S1 y S2** (ver Anexo), no parece haber asociación alguna entre el FDR estimado para el enriquecimiento en inparálogos de secreción y algunos indicadores de calidad de genoma como son el N50 o N90. Así, tanto genomas de ensamblado robusto y alto curado manual (e.g. *S. mansoni*) parecen poseer la misma tendencia que aquellos más precarios.

La falta de curación manual dificulta a su vez la anotación de aquellos genes inparálogos para los que se detectó evolución en régimen de selección natural positiva. En muchos casos, sólo se logran identificar ortólogos anotados en genomas curados manualmente (e.g. *Schistosoma mansoni*), y si bien los umbrales mínimos establecidos para identidad y cobertura permiten asegurar con cierta seguridad que en efecto se trata de genes homólogos, no es trivial asegurar que estas secuencias cumplan el mismo rol en diferentes organismos. Por ejemplo, una cathepsina puede jugar un rol en el proceso de invasión de *F. hepatica*, pero no se puede asegurar que en otros tremátodos (o incluso en especies de otras clases) cumpla la misma función o se enfrente a presiones selectivas comparables. Esta incertidumbre se intensifica a medida que se relajan los umbrales de identidad y cobertura, pudiendo existir alineamientos entre secuencias debido a la presencia de dominios proteicos similares, los cuales pueden o no residir en estas proteínas debido a homología, a la vez que tampoco es claro que cumplan funciones comparables en diferentes organismos (e.g. la presencia de posibles elastasas en *S. mediterranea* con similitud a secuencias de genes codificantes para proteínas de invasión de *S. mansoni*).

Respecto a la influencia de una subanotación de términos funcionales, el problema ya fue planteado en múltiples trabajos (ver Gaudet et al. 2017 para una revisión de este y otros problemas asociados), y en efecto parece ser un problema intrínseco a la técnica que persistirá hasta que se posea una anotación extensiva y manual de una enorme cantidad de genomas en algún futuro.

En última instancia este tipo de análisis solo poseen un valor confirmatorio de cierta tendencia observada por otros métodos (e.g. análisis de evolución molecular u observaciones experimentales).

ii) Muestreo filogenético dispar entre las clases del filo.

Quizá uno de los mayores problemas de este estudio sea la ausencia de genomas disponibles para un mayor número de representantes de las clases Monogenea y Turbellaria.

Es de esperar que conforme se ensamblen y anoten los genomas de diversos organismo de estas clases el número de inparálogos inferido por el método aquí utilizado disminuya drásticamente, simplemente debido a la posibilidad de establecer relaciones de ortología entre parálogos anteriores al evento de especiación entre la especie actualmente secuenciada y aquellas que todavía no

poseen su genoma disponible. Así, el ensamblado y anotación de especies cercanas a *M. lignano* y *S. mediterranea* puede llevar a una disminución del número de inparálogos detectados, simplemente porque en efecto estos no eran duplicaciones linaje-específicas: la falta de anotación de las mismas en especies antes no consideradas que sí poseen estas duplicaciones llevaría en este caso al artefacto de detectarlas como duplicaciones específicas del linaje considerado.

Esta idea está respaldada por el número de inparálogos detectados en este trabajo: en efecto, aquellos organismos con genomas de alta calidad y que poseen organismos cercanos a nivel filogenético secuenciado (*e.g.* *E. granulosus* o *S. mansoni*) poseen un menor número de inparálogos que aquellos de clases con una menor representación. Una forma de testear esta noción sería graficando el número de inparálogos detectado respecto a la distancia filogenética existente entre los organismos considerados.

iii) Definición de familias multigénicas y predicción del secretoma de cada especie.

La inferencia de homología a partir de secuencias nucleotídicas o aminoácídicas es un problema complejo ampliamente discutido en la literatura (Rost, 1999; Koonin, 2005; Pearson, 2013).

En concreto, debido a la redundancia del código genético múltiples mutaciones silenciosas pueden ocurrir a secuencias homólogas sin que esto se traduzca en un cambio a nivel aminoácídico. Esto se traduce en una conservación de la secuencia aminoácídica codificada aún cuando exista divergencia entre las secuencias nucleotídicas homólogas consideradas, lo cual se traduce hasta cierto punto en una conservación de la estructura y la función de las proteínas codificadas (ver Abagyan & Batalov, 1997; Rost, 2002; Tian & Skolnick, 2003; para una discusión de hasta qué punto esta idea es cierta). Siguiendo esta línea argumental, es posible que la identidad existente entre dos secuencias disminuya aún cuando el *fitness* de él o los organismos considerados no se vea afectado por este proceso.

Lo que es más, Rost (1999) ha discutido hasta qué punto es posible afectar la identidad existente entre dos secuencias proteicas que conservan una estructura en común. Sin ahondar en detalles, el rango definido al rededor del 20% y el 35% de identidad (*i.e.* la zona popularmente denominada *twilight zone*) entre dos secuencias aminoácídicas (con un largo en el orden de las centenas) se caracteriza por un estallido de falsos positivos en lo que ha inferencia de homología se refiere, siendo imposible inferir homología basándose simplemente en las secuencias aminoácídicas codificadas.

En este trabajo, no obstante, se impuso como mínimo de identidad entre las secuencias codificadas por los miembros de una familia multigénica un 40%. Si bien es cierto que en principio este umbral permite alejarse de la zona de incertidumbre previamente nombrada, lo cierto es que existe una dependencia de los valores de identidad a emplear respecto al largo del alineamiento: a menor largo de secuencias, lógicamente se requiere de mayor porcentaje de identidad para inferir homología de forma razonable. Afortunadamente, OrthoMCL se basa en el uso de alineamientos realizados a través de BLAST, prefijándose de antemano un máximo *e-value* a existir entre las secuencias consideradas para poder formar parte de una misma familia multigénica (recordar que el *e-value* posee una relación directa con el largo de la secuencia a emplear).

Dicho esto, no obstante, lo cierto es que la definición de familias multigénicas a partir de secuencias aminoácídicas posee la desventaja de no poseer la resolución suficiente para detectar la enorme mayoría de secuencias que poseen una misma estructura (a partir de la cual es posible inferir homología por todo lo antes dicho). Esto impone un límite al número de familias de las que se parte para el análisis, el cual lógicamente se podría ampliar si se utilizara información estructural para aumentar el número de familias a considerar.

Resulta interesante respecto a este punto también nombrar el trabajo realizado por Demuth y Hahn con familias multigénicas codificadas en seis genomas de mamíferos (Demuth & Hahn, 2009). En el mismo los autores demuestran que, en su caso, si bien el uso de diferentes umbrales en aquellos parámetros que definen a operacionalmente a las familias multigénicas repercuten en el número absoluto de secuencias que pertenece a cada una de ellas, tendencias estadísticas referentes a la expansión y contracción de las mismas a lo largo de su historia evolutiva no se ven afectadas. Si bien esto se cumple estrictamente para el tratamiento utilizado por los autores y las familias multigénicas de los genomas empleados, lo cierto es que es razonable esperar cierta extrapolación de este concepto a fenómenos relacionados con el nacimiento y muerte de familias génicas (como puede ser el enriquecimiento testeado en este trabajo para un tipo de proteínas en particular) y para otros genomas eucariotas.

En cuanto a la predicción de proteínas de secreción utilizadas en este trabajo, el porcentaje del genoma de cada especie que codifica para proteínas del secretoma es comparable al obtenido por Cuesta-Astroz (Cuesta-Astroz et al., 2017). Siendo que el autor trabaja en buena medida con los mismos genomas que los empleados en esta tesina, lo observado confirma que a grandes rasgos no existen pruebas que lleven a pensar en una mala implementación de la técnica de predicción en cuestión. Descartado este tipo de errores, parece lógico suponer que no debe existir *a priori* una tendencia inherente a la técnica a detectar de forma sesgada a secuencias inparálogas como codificantes para proteínas de secreción. Considerando que también es razonable esperar independencia entre ambas variables (léase, el hecho de que una proteína sea secretada no posee base molecular en que su secuencia codificante posea inparálogos), se está en condiciones de afirmar que el test de enriquecimiento planteado en esta tesina parece ser en sus bases conceptuales confiable.

iv) Error en la inferencia de selección natural positiva

En lo que respecta a la detección de grupos de inparálogos evolucionando bajo un régimen de selección natural positiva, al menos dos fuentes de error son planteables: en primer lugar se debe considerar la pertinencia del modelo evolutivo asumido (en este trabajo se optó por lo que se denomina un *modelo de sitios específicos* (ver Yang & Nielsen, 2002); en segundo lugar se debe considerar el efecto causado por inferencia errónea de sitios homólogos, debida ésta a errores al momento de alinear las secuencias analizadas.

Respecto al primer punto es pertinente resaltar que el modelo de sitios-específicos aquí empleado tiene como asunción central que las secuencias analizadas se encuentran bajo un mismo régimen selectivo (*i.e.* todas las ramas se comportan de igual manera). Si bien puede ser razonable asumir esta condición para secuencias inparálogas con poca divergencia, ya que en principio todas las copias en esta situación codifican para proteínas que afectan al fenotipo de forma similar, lo cierto es que en caso de que las secuencias consideradas evolucionen bajo un régimen de selección natural positiva es de esperar que, si se consideran secuencias con suficiente tiempo de divergencia, algunas de las ramas sobre las que opera la selección tengan un comportamiento diferencial (por ejemplo,

respecto a otras ramas de la familia que se encuentren bajo un proceso de pseudogenización). Una solución para este problema puede ser la utilización de un modelo de *ramas y sitios*, donde ambas categorías son estimadas a través de métodos de máxima verosimilitud. El programa HyPhy, del paquete DataMonkey (Kosakovsky et al., 2005), por ejemplo, es un buen candidato a ser utilizado para este propósito, gracias a la automatización que permite el hecho de que en el mismo se testea que cada una de las ramas existentes en la topología del árbol en cuestión se encuentren evolucionando a tasas diferenciales (por el contrario, en otros paquetes como PAML el usuario debe definir la rama en cuestión, lo cual torna inviable el uso de la herramienta para grandes cantidades de familias multigénicas).

Un caso diferente surge si se toma en cuenta la posibilidad de error ya no debida a el modelo de análisis planteado, si no a desperfectos técnicos surgidos de los alineamientos empleados al momento de realizar la estimación de indicadores como dN/dS. Ya Markova-Raina & Petrov (Markova-Raina & Petrov, 2011) demostraron utilizando genes homólogos codificados en 12 genomas de *Drosophila melanogaster* que la detección de selección natural positiva operando sobre la evolución de dichas familias es fuertemente dependiente del programa utilizado para inferir los alineamientos sobre los que se realiza el análisis de evolución molecular. Lo que es más, los autores demuestran que existe amplia discordancia entre el número y la identidad de las secuencias para las que se detecta el fenómeno, así como para los sitios inferidos. El fenómeno incluye un altísimo porcentaje de falsos positivos (producto de la inferencia errónea de la tasa de sustitución no sinónima), muchos de los cuáles se vinculan a regiones codificantes cercanas a la región límite entre exones e intrones. Si bien la restricción a especies filogenéticamente cercanas disminuye la discordancia entre los diferentes alineadores, el fenómeno sigue existiendo y también acarrea el problema de la alta tasa de falsos positivos. La utilización de GBLOCKS con el fin de eliminar regiones cuestionables de los distintos alineamientos de forma previa al cálculo del estimador dN/dS, no obstante, se espera disminuya al menos en parte estas problemáticas.

Por último, es importante considerar que una forma sensata de reducir el número de falsos positivos al momento de detección de casos donde opera la selección natural positiva en la evolución de las secuencias es la restricción del análisis a casos donde el número de cambios sinónimos observado sea moderado. En efecto, la consideración de tiempos evolutivos demasiado pequeños (reflejados en un valor de dS pequeño) puede llevar a un valor de dN/dS cuestionable si se considera que el azar puede influir en que este valor se encuentre sobreestimado: a valores de tasa de sustitución extremadamente chicos, la aparición por azar de unas pocas sustituciones no sinónimas (incluso por efecto de un proceso de pseudogenización) lleva a un dN/dS superior a 1, incluso cuando esta tendencia no se observaría si se consideraran las mismas secuencias con suficiente tiempo de divergencia, ya sea por estabilización del valor de dS observado o por la pseudogenización total de las secuencias. A su vez, si se consideran tiempos de divergencia demasiado grandes (*i.e.* dS altos) se corre el riesgo de que se haya alcanzado la saturación: bajo este escenario, múltiples cambios sinónimos ocurrieron en un mismo sitio, sin que todos estos cambios aporten a la estimación del valor de dS empleado. Esto lleva, igual que en el caso de utilización de secuencias poco divergentes, a una sobreestimación del valor de dN/dS, con el riesgo de asumir un proceso de evolución en régimen de selección natural positiva errado.

Distribución de inparálogos en isócoros del genoma de *Schistosoma mansoni*: una hipótesis mecanística

Como se dijo en la sección Materiales y Métodos, Lamolle y colaboradores (Lamolle et al., 2016) describieron estructuras de tipo isocóro en el trematodo *S. mansoni*. La distribución obtenida tanto para los inparálogos detectados en este trabajo, así como para la fracción de los mismos que se detectó podría estar evolucionando bajo un régimen de selección positiva puede explicarse teniendo en cuenta algunas de las correlaciones ya estudiadas entre contenido GC de isócoros y la distribución de genes y sus tasas de expresión y recombinación.

La presencia de un mayor número de genes por base en isocóros de alto contenido GC ya ha sido descrita por diversos autores (ver Bernardi, 2000), así como la tendencia a una mayor expresión de los mismos y la existencia de un estado de menor condensación de la cromatina. Resultado directo de estas observaciones es la mayor tasa de recombinación existente entre los genes ubicados en estos isocóros de alto contenido GC (Bernardi, 2000).

La tendencia de inparálogos a prevalecer en zonas de mayor contenido GC puede ser considerado como un hecho coherente con la presencia de un mayor número de genes en estas zonas, así como con la tendencia de los mismos a poseer una mayor tasa de recombinación (y por tanto una mayor posibilidad de generar copias en tándem). No obstante, si se compara la **Figura 9** con la **Figura 7** de la sección Resultados, se notará que existe en la segunda una distribución incluso más sesgada hacia isocóros de alto contenido en GC que en la primera.

Este último hecho puede estar vinculado con la mayor tasa de expresión que poseen los genes en estas zonas: si en efecto se genera un mayor número de inparálogos, tanto por la mayor presencia de genes en los isocóros de mayor contenido en GC, así como la mayor tasa de recombinación asociado a los mismos, también es de esperar que la selección natural positiva opere preferencialmente sobre aquellos genes que poseen una mayor tendencia a expresarse, debido al efecto que tendrán en el *fitness* del organismo portador. En otras palabras, no sólo se generará un menor número de secuencias inparálogas en isocóros con menor número de genes y con menores tasas de recombinación: es de esperar que si los mismos se expresan en menores tasas, pseudogenicen a un ritmo mayor, en un proceso evolutivo de carácter principalmente neutro.

No obstante, los números de genes mapeados son sumamente bajos, lo cual obstaculiza realizar un análisis estadístico de este fenómeno. La búsqueda de estructuras de tipo isocóro en otros platelmintos puede por tanto revertir interés para saldar esta discusión, permitiendo a su vez evaluar el fenómeno desde un punto de vista más global en el filo.

Parásitos y organismos de vida libre muestran adaptaciones a nivel de su secretoma

A diferencia del resto de los organismos, *S. mediterranea* y *M. lignano* son organismos de vida libre. Resulta interesante por tanto contrastar los resultados obtenidos para los mismos respecto a los de organismos parásitos: ¿qué diferencias adaptativas muestran estos organismos de vida libre respecto a los parásitos, en cuanto a proteínas de secreción y su evolución se refiere?

Sólo en *S. mediterranea* se observa un proteínas de secreción entre los inparálogos detectados, a la vez que en ambos organismos fue posible detectar casos de grupos de los mismos evolucionando en un proceso de selección natural positiva. No obstante, resulta interesante hacer hincapié en los perfiles funcionales que arrojan ambos (reflejado en los GO terms que se destacan como enriquecidos). En el caso de *M. lignano* se resaltan algunas funciones vinculadas al metabolismo de lípidos (*e.g.* “lipid catabolic process”, y “phosphatidylglycerol metabolic process”), de regulación de procesos del desarrollo (“regulation of BMP signaling pathway”, “peptide hormone processing”), así como funciones relacionadas con la adhesión a sustrato (“substrate adhesion-dependent cell spreading”), una función ya descrita en proteínas de secreción. Algo más críptico resulta aquel proceso biológico que arrojó un valor más significativo entre los inparálogos de secreción para los que se detectó un fenómeno adaptativo, vinculado a la maquinaria de incorporación de selenocisteína.

Un análisis de los términos vinculados a la función molecular muestra, no obstante, que la proteólisis parece ser la función más enriquecida entre los inparálogos en cuestión. Este resultado es similar al obtenido en muchos organismos parásitos, aunque presumiblemente aquí podría tener un rol vinculado a la obtención de nutrientes.

Siendo *M. lignano* un organismo con presunta tetraploidía (Zadesenets et al., 2016), los resultados obtenidos no deben ser interpretados sin tener en cuenta que las copias pertenecientes a distintos cromosomas del organismo poseen segregación independiente (es decir, sufrirán procesos de deriva génica con posible independencia), a la vez que se verán afectados por los diferentes procesos que pueden haber moldeado la arquitectura de estos cromosomas.

En lo que refiere a *S. mediterranea*, el resultado puede ser considerado por lo menos inesperado: procesos biológicos vinculados a la interacción entre especies y al desarrollo de interacciones simbióticas se ven enriquecidos entre los inparálogos analizados, junto a términos vinculados al desarrollo. Estos resultados deben ser tomados con cautela, ya que también se observa que los términos vinculados a componentes celulares enriquecidos corresponden al aparato de Golgi.

Desde un punto de vista molecular, se observa un enriquecimiento en funciones vinculadas al procesamiento de glúcidos, lo cual como ya se mencionó en esta tesis está en consonancia quizá con el rol que poseen múltiples glicoproteínas y glúcidos en cuanto a interacción entre organismos y modelado del medio extracelular. Se observan también términos ontológicos vinculados a proteólisis y transferasas, también discutidos en esta sección.

El enriquecimiento en proteínas de secreción entre los inparálogos de este organismo, así como el análisis ya mencionado, parecen ser coherentes con los resultados obtenidos apoyan lo reportado por Bocchinfuso y colaboradores (Bocchinfuso et al. 2012), donde un estudio proteómico arrojó la enorme similitud existente entre los componentes proteicos secretados por este organismo y aquellos de parásitos *S. mansoni*. En efecto, algunos de los grupos de inparálogos plausibles de encontrarse evolucionando en un proceso de selección natural positiva en este organismo presentan una posible homología con proteínas como SmVal (una elastasa ampliamente estudiada en el trematodo), aunque esta posible homología se observa sólo si se relajan los valores umbrales de identidad (a un 50 % mínimo), lo cual sugiere o bien la presencia de dominios similares o bien una relación de homología remota.

Si bien no es posible hacer aseveraciones de ningún tipo sin realizar un estudio más a fondo, lo cierto es que resulta interesante plantearse la plausibilidad de que muchas de las proteínas secretadas con un rol en la invasión y mantención de parásitos y sus huéspedes posean un ancestro común con proteínas secretadas en organismos de vida libre. Si este es el caso, se estaría ante un proceso de cooptación de las mismas en el marco de la evolución de dichos parásitos.

Perspectivas

Para finalizar esta sección resulta pertinente plantear al menos dos análisis a realizar para profundizar en los resultados obtenidos y obtener un mejor panorama de la relevancia de los fenómenos descritos en cuanto al proceso adaptativo de estos organismos se refiere.

En primer lugar, un análisis estadístico de los fenómenos de expansión y contracción de familias génicas (tal y como realizan Demuth & Hahn en su trabajo de 2009) es pertinente para observar no sólo si este enfoque confirma lo planteado en este trabajo, sino también para testear hasta qué punto los fenómenos de contracción diferencial de familias multigénicas también puede poseer un rol adaptativo en la fisiología de estos organismos: es evidente que en organismos parásitos no sólo la ganancia de funciones moleculares a través de expansiones génicas, si no también la pérdida diferencial de otras como resultado del parasitismo aportan a la comprensión de la diversidad de mecanismos adaptativos presentes en estos organismos.

Por otro lado, estudios transcriptómicos podrían permitir distinguir no sólo falsos positivos (no es de esperar que la selección natural opere sobre genes no expresados), sino que también puede brindar un panorama respecto a en qué puntos del ciclo de vida y desarrollo de estos organismos se ven implicados las familias génicas estudiadas.

Conclusión

En esta tesina de grado se buscó indagar en el rol adaptativo que podrían poseer duplicaciones génicas linaje-específicas codificantes para proteínas de secreción en una treintena de platelmintos. La observación de que este conjunto de duplicaciones linaje-específicas se ve enriquecida en secuencias codificantes para proteínas de secreción en la enorme mayoría de los organismos estudiados respalda esta noción. Lo que es más, un análisis funcional de las mismas muestra un enriquecimiento en funciones moleculares vinculadas a procesos ya descritos previamente como esenciales para la sobrevivencia de estos organismos. A su vez, se logró detectar un conjunto de secuencias que podrían estar evolucionando bajo un régimen de selección natural positiva, entre las cuales se encuentran tanto proteínas anotadas y con función clara en estos organismos como proteínas cuyo rol todavía no es claro.

Los resultados obtenidos sugieren por tanto que en efecto la duplicación génica de secuencias codificantes para proteínas de secreción constituye una fuente de materia prima para la evolución de funciones moleculares que cumplen un rol adaptativo para los organismos de este filo. Un estudio detallado de algunas de las familias multigénicas estudiadas en esta tesina, así como de otros factores que aportan a la adaptación de estos organismos a su medio, puede contribuir a comprender de forma más cabal la biología de los mismos.

Bibliografia

- Abagyan, R. A., & Batalov, S. (1997).** Do aligned sequences share the same fold? *Journal of Molecular Biology*, 273(1), 355–368. <https://doi.org/10.1006/jmbi.1997.1287>
- Ahn, C.-S., Kim, J.-G., Han, X., Kang, I., & Kong, Y. (2017).** Comparison of *Echinococcus multilocularis* and *Echinococcus granulosus* hydatid fluid proteome provides molecular strategies for specialized host-parasite interactions. *Oncotarget*, 8(57), 97009–97024. <https://doi.org/10.18632/oncotarget.20761>
- Akopian, D., Shen, K., Zhang, X., & Shan, S. (2013).** Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annual Review of Biochemistry*, 82(1), 693–721. <https://doi.org/10.1146/annurev-biochem-072711-164732>
- Alberts, B., Melorose, J., Perroy, R., & Careas, S. (2015).** *Molecular Biology of the Cell. Statewide Agricultural Land Use Baseline 2015 (Vol. 1)*. <https://doi.org/10.1017/CBO9781107415324.004>
- Alexa, A., & Rahnenfuhrer, J. (2016).** topGO.
- Altschul, S. (1990).** Basic Local Alignment Search Tool. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1990.9999>
- Alvite, G., & Esteves, A. (2012).** Lipid binding proteins from parasitic platyhelminthes. *Frontiers in Physiology*, 3 SEP(September), 1–7. <https://doi.org/10.3389/fphys.2012.00363>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000).** Gene ontology: Tool for the unification of biology. *Nature Genetics*. <https://doi.org/10.1038/75556>
- Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., & Brunak, S. (2004).** Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering, Design and Selection*. <https://doi.org/10.1093/protein/gzh037>
- Bernardi, G. (2000).** Isochores and the evolutionary genomics of vertebrates. *Gene*. [https://doi.org/10.1016/S0378-1119\(99\)00485-0](https://doi.org/10.1016/S0378-1119(99)00485-0)
- Bernardi, G. (2004).** Structural and Evolutionary Genomics: Natural Selection in Genome Evolution. *New Comprehensive Biochemistry (Vol. 37)*. <https://doi.org/10.1016/S0167730604370079>
- Bocchinfuso, D. G., Taylor, P., Ross, E., Ignatchenko, A., Ignatchenko, V., Kislinger, T., . . . Moran, M. F. (2012).** Proteomic Profiling of the Planarian *Schmidtea mediterranea* and Its Mucous Reveals Similarities with Human Secretions and Those Predicted for Parasitic Flatworms. *Molecular & Cellular Proteomics*, 11(9), 681–691. <https://doi.org/10.1074/mcp.M112.019026>
- Buchfink, B., Xie, C., & Huson, D. H. (2015).** Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. <https://doi.org/10.1038/nmeth.3176>
- Buttarelli, F. R., Pellicano, C., & Pontieri, F. E. (2008).** Neuropharmacology and behavior in planarians: Translations to mammals. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*. <https://doi.org/10.1016/j.cbpc.2008.01.009>
- Cancela, M., Acosta, D., Rinaldi, G., Silva, E., Duran, R., Roche, L., Tort, J. F. (2008).** A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie*, 90(10), 1461–1475. <https://doi.org/10.1016/j.biochi.2008.04.020>
- Chow, C., Gauci, C. G., Cowman, A. F., Lightowlers, M. W. (2001).** A gene family expressing a host-protective antigen of *Echinococcus granulosus*. *Molecular and Biochemical Parasitology*, 118(1), 83–88. [https://doi.org/10.1016/S0166-6851\(01\)00373-5](https://doi.org/10.1016/S0166-6851(01)00373-5)
- Collins, J. J. (2017).** Platyhelminthes. *Current Biology*, 27(7), R252–R256. <https://doi.org/10.1016/j.cub.2017.02.016>
- Contreras-Moreira, B., & Vinuesa, P. (2013).** GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology*. <https://doi.org/10.1128/AEM.02411-13>
- Cuesta-Astroz, Y., Oliveira, F. S. de, Nahum, L. A., & Oliveira, G. (2017).** Helminth secretomes reflect different lifestyles and parasitized hosts. *International Journal for Parasitology*. <https://doi.org/10.1016/j.ijpara.2017.01.007>
- Cuesta-Astroz, Y., Oliveira, F. S. de, Nahum, L. A., & Oliveira, G. (2017).** Helminth secretomes reflect different lifestyles and parasitized hosts. *International Journal for Parasitology*. <https://doi.org/10.1016/j.ijpara.2017.01.007>
- Curwen, R. S. (2006).** Identification of Novel Proteases and Immunomodulators in the Secretions of Schistosome Cercariae That Facilitate Host Entry. *Molecular Cellular Proteomics*, 5(5), 835–844. <https://doi.org/10.1074/mcp.M500313-MCP200>
- Dayhoff MO. (1976).** The origin and evolution of protein superfamilies. *Fed Proc* 35: 2132–2138. PMID: 181273
- Dalton, J. P., Neill, S. O., Stack, C., Collins, P., Walshe, A., Sekiya, M., . . . Donnelly, S. M. (2003).** *Fasciola hepatica* cathepsin L-like proteases: Biology, function, and potential in the development of first generation liver fluke vaccines. In *International Journal for Parasitology*. [https://doi.org/10.1016/S0020-7519\(03\)00171-1](https://doi.org/10.1016/S0020-7519(03)00171-1)
- Dalton, J. P., Skelly, P., & Halton, D. W. (2004).** Role of the tegument and gut in nutrient uptake by parasitic platyhelminths. *Canadian Journal of Zoology*. <https://doi.org/10.1139/z03-213>
- Delpont, 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*. <https://doi.org/10.1093/bioinformatics/btq429>

- Demuth, J. P., & Hahn, M. W. (2009).** The life and death of gene families. *BioEssays*, 31(1), 29–39. <https://doi.org/10.1002/bies.080085>
- Duckert, P., Brunak, S., & Blom, N. (2004).** Prediction of proprotein convertase cleavage sites. *Protein Engineering, Design and Selection*. <https://doi.org/10.1093/protein/gzh013>
- Dunn, C. W., Giribet, G., Edgecombe, G. D., & Hejnol, A. (2014).** Animal Phylogeny and Its Evolutionary Implications. *Annual Review of Ecology, Evolution, and Systematics*, 45(1), 371–395. <https://doi.org/10.1146/annurev-ecolsys-120213-091627>
- Edgar, R. C. (2004).** MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkh340>
- Elliott, S. A., & Sánchez Alvarado, A. (2013).** The history and enduring contributions of planarians to the study of animal regeneration. *Wiley Interdisciplinary Reviews: Developmental Biology*. <https://doi.org/10.1002/wdev.82>
- Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007).** Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2(4), 953–971. <https://doi.org/10.1038/nprot.2007.131>
- Emanuelsson, O., Nielsen, H., Brunak, S., & von Heijne, G. (2000).** Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.2000.3903>
- Esteves, A., Portillo, V., & Ehrlich, R. (2003).** Genomic structure and expression of a gene coding for a new fatty acid binding protein from *Echinococcus granulosus*. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1631(1), 26–34. [https://doi.org/10.1016/S1388-1981\(02\)00321-9](https://doi.org/10.1016/S1388-1981(02)00321-9)
- Finn, R. D., Clements, J., & Eddy, S. R. (2011).** HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkr367>
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999).** Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. <https://doi.org/10.1016/1011175>
- Francino, M. P. (2005).** An adaptive radiation model for the origin of new gene functions. *Nature Genetics*. <https://doi.org/10.1038/ng1579>
- Garg, G., & Ranganathan, S. (2011).** In silico secretome analysis approach for next generation sequencing transcriptomic data. In 10th Int. Conference on Bioinformatics - 1st ISCB Asia Joint Conference 2011, InCoB 2011/ISCB-Asia 2011: Computational Biology - Proceedings from Asia Pacific Bioinformatics Network (APBioNet). <https://doi.org/10.1186/1471-2164-12-S3-S14>
- Gauci, C., Merli, M., Muller. (2002).** Molecular Cloning of a Vaccine Antigen against Infection with the Larval Stage of *Echinococcus multilocularis* Molecular Cloning of a Vaccine Antigen against Infection with the Larval Stage of *Echinococcus multilocularis*. *Infection and Immunity*, 70(7), 3969–3972. <https://doi.org/10.1128/IAI.70.7.3969>
- Gaudet, P., & Dessimoz, C. (2017).** Gene ontology: Pitfalls, biases, and remedies. In *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-3743-1_14
- Gene Ontology Consortium. (2004).** The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkh036>
- Hansell, E., Braschi, S., Medzihradsky, K. F., Sa jid, M., Debnath, M., Ingram, J., . . . McKerrow, J. H. (2008).** Proteomic analysis of skin invasion by blood fluke larvae. *PLoS Neglected Tropical Diseases*, 2(7). <https://doi.org/10.1371/journal.pntd.0000262>
- Hickman, C. P., Roberts, L. S., Keen, S. L., Larson, A., L'Anson, H., & Elsenhour, D. J. (2007).** Integrated Principles of Zoology. *Integrated Principles of Zoology*, 14, 936.
- Hu, D., Song, X., Xie, Y., Zhong, X., Wang, N., Zheng, Y., . . . Yang, G. (2015).** Molecular insights into a tetraspanin in the hydatid tapeworm *Echinococcus granulosus*. *Parasites and Vectors*, 8(1), 1–10. <https://doi.org/10.1186/s13071-015-0926-y>
- Huang, F., Dang, Z., Suzuki, Y., Horiuchi, T., Yagi, K., Kouguchi, H., . . . Oku, Y. (2016).** Analysis on Gene Expression Profile in Oncospheres and Early Stage Metacestodes from *Echinococcus multilocularis*. *PLoS Neglected Tropical Diseases*, 10(4), 1–22. <https://doi.org/10.1371/journal.pntd.0004634>
- Huang, F., Dang, Z., Zhang, H., Yagi, K., Kim, K., Joseph, M., . . . Oku, Y. (2017).** Comparative study on secretome and transmembrane of immature and mature metacestodes of *Echinococcus multilocularis*. *Veterinary Parasitology*, 245(January), 153–159. <https://doi.org/10.1016/j.vetpar.2017.05.006>
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. (2017).** Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper.
- Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., & Bork, P. (2008).** eggNOG: Automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research*, 36(SUPPL. 1), 250–254. <https://doi.org/10.1093/nar/gkm796>

- Joly, D. L., Feau, N., Tanguay, P., & Hamelin, R. C. (2009).** Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). *BMC Genomics*. Retrieved from <http://www.biomedcentral.com/content/pdf/1471-2164-11-422.pdf> Cnpapers://74e8b4ef-c919-4470-93db-e267bd9e4a22/Paper/p9454
- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992).** The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/8.3.275>
- Kanehisa, M. (2002).** The KEGG databases at GenomeNet. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/30.1.42>
- Kaphingst, K. A., Persky, S., & Lachance, C. (2010).** NIH Public Access, 14(4), 384–399. <https://doi.org/10.1080/10810730902873927>. Testing
- Karn, R. C., Clark, N. L., Nguyen, E. D., & Swanson, W. J. (2008).** Adaptive evolution in rodent seminal vesicle secretion proteins. *Molecular Biology and Evolution*, 25(11), 2301–2310. <https://doi.org/10.1093/molbev/msn182>
- Kihara, D. (Ed.) . (2017).** Protein Function Prediction. *Methods in Molecular Biology*. <https://doi.org/10.1007/978-1-4939-7015-5>
- Koonin, E. V. (2005).** Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics*. <https://doi.org/10.1146/annurev.genet.39.073003.114725>
- Krogh, a, Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001).** Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.2000.4315>
- Lamolle, G., Protasio, A. V., Iriarte, A., Jara, E., Simón, D., & Musto, H. (2016).** An Isochore-Like Structure in the Genome of the Flatworm *Schistosoma mansoni*. *Genome Biology and Evolution*. <https://doi.org/10.1093/gbe/evw170>
- Lamolle, G., Protasio, A. V., Iriarte, A., Jara, E., Simón, D., & Musto, H. (2016).** An Isochore-Like Structure in the Genome of the Flatworm *Schistosoma mansoni*. *Genome Biology and Evolution*. <https://doi.org/10.1093/gbe/evw170>
- Li, L., Stoekert, C. J. J., & Roos, D. S. (2003).** OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes – Li et al. 13 (9): 2178 – *Genome Research*. *Genome Research*, 13(9), 2178–2189. <https://doi.org/10.1101/gr.1224503.candidates>
- Littlewood DTJ. The evolution of parasitism in flatworms (2006)..** In: Maule AG, Marks NJ, editors. *Parasitic flatworms: molecular biology, biochemistry, immunology and physiology* , Wallingford (United Kingdom) Cabi Publishing-C a B Int(pg. 1-36)
- Lowther, J., Robinson, M. W., Donnelly, S. M., Xu, W., Stack, C. M., Matthews, J. M., & Dalton, J. P. (2009).** The importance of pH in regulating the function of the *Fasciola hepatica* cathepsin L1 cysteine protease. *PLoS Neglected Tropical Diseases*. <https://doi.org/10.1371/journal.pntd.0000369>
- Lynch, M., & Conery, J. S. (2003).** The evolutionary demography of duplicate genes. *Journal of Structural and Functional Genomics*, 3(1–4), 35–44. <https://doi.org/10.1023/A:1022696612931>
- Malhotra 2013 Rabouille, C., Malhotra, V., & Nickel, W. (2012).** Diversity in unconventional protein secretion. *Journal of Cell Science*. <https://doi.org/10.1242/jcs.103630>
- Markova-Raina, P., & Petrov, D. (2011).** High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Research*. <https://doi.org/10.1101/gr.115949.110>
- Martinez, J., Perez-Serrano, J., Bernadina, W. E., & Rodriguez-Caabeiro, F. (1999).** Echinococcus granulosus: In vitro effects of ivermectin and praziquantel on hsp60 and hsp70 levels. *Experimental Parasitology*, 93(3), 171–180. <https://doi.org/10.1006/expr.1999.4449>
- Martinez, J., Perez-Serrano, J., Bodega, G., Casado, N., & Rodriguez-Caabeiro, F. (1999).** Heat shock proteins HSP70 and HSP60 in *Echinococcus granulosus* protoscolices. *Folia Parasitologica*, 46(1), 76–78.
- Mayta, H., Hancock, K., Levine, M. Z., Gilman, R. H., Farfán, M. J., Verastegui, M., . . . Tsang, V. C. W. (2007).** Characterization of a novel *Taenia solium* oncosphere antigen. *Molecular and Biochemical Parasitology*. <https://doi.org/10.1016/j.molbiopara.2007.07.017>
- McCarthy, E., Stack, C., Donnelly, S. M., Doyle, S., Mann, V. H., Brindley, P. J., . . . Dalton, J. P. (2004).** Leucine aminopeptidase of the human blood flukes, *Schistosoma mansoni* and *Schistosoma japonicum*. *International Journal for Parasitology*, 34(6), 703–714. <https://doi.org/10.1016/j.ijpara.2004.01.008>
- Morales, M. E., Kalinna, B. H., Heyers, O., Mann, V. H., Schulmeister, A., Copeland, C. S., . . . Brindley, P. J. (2004).** Genomic organization of the *Schistosoma mansoni* aspartic protease gene, a platyhelminth orthologue of mammalian lysosomal cathepsin D. *Gene*. <https://doi.org/10.1016/j.gene.2004.05.017>
- Nene, V., Dunne, D. W., Johnson, K. S., Taylor, D. W., & Cordingley, J. S. (1986).** Sequence and expression of a major egg antigen from *Schistosoma mansoni*. Homologies to heat shock proteins and alpha-crystallins. *Molecular and Biochemical Parasitology*, 21(2), 179–188. [https://doi.org/10.1016/0166-6851\(86\)90021-6](https://doi.org/10.1016/0166-6851(86)90021-6)
- Nielsen, Henrik, Engelbrecht, J., Brunak, S., & von Heijne, G. (1997).** Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*, 10(1), 1–6. <https://doi.org/10.1142/S0129065797000537>
- Paradis, E., Strimmer, K., Claude, J., Noel, Y., & Bolker, B. (2003).** Analyses of Phylogenetics and Evolution: the ape Package. *Bioinformatics*, 20(2), 289–290. <https://doi.org/10.1093/bioinformatics/btg412>

- Pearson, W. R. (2013).** An introduction to sequence similarity (“homology”) searching. *Current Protocols in Bioinformatics*. <https://doi.org/10.1002/0471250953.bi0301s42>
- Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011).** SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*. <https://doi.org/10.1038/nmeth.1701>
- Porfido, J. L., Alvite, G., Silva, V., Kennedy, M. W., Esteves, A., & Corsico, B. (2012).** Direct Interaction between EgFABP1, a Fatty Acid Binding Protein from *Echinococcus granulosus*, and Phospholipid
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010).** FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0009490>
- Quezada, L. A. L., & McKerrow, J. H. (2011).** Schistosome serine protease inhibitors: Parasite defense or homeostasis? *Anais Da Academia Brasileira de Ciencias*. <https://doi.org/10.1590/S0001-37652011000200025>
- R Development Core Team. (2016).** R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna Austria. <https://doi.org/10.1038/sj.hdy.6800737>
- Raymond, M., Chevillon, C., Guillemaud, T., Lenormand, T., & Pasteur, N. (1998).** An overview of the evolution of overproduced esterases in the mosquito *Culex pipiens*. *Philosophical Transactions of the*
- Revell, L. J. (2016).** Package ‘phytools.’ R Topics Documented.
- Rice, P., Longden, L., & Bleasby, A. (2000).** EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)
- Rinaldi, G., Morales, M. E., Alrefaei, Y. N., Cancela, M., Castillo, E., Dalton, J. P., . . . Brindley, P. J. (2009).** RNA interference targeting leucine aminopeptidase blocks hatching of *Schistosoma mansoni* eggs. *Molecular and Biochemical Parasitology*. <https://doi.org/10.1016/j.molbiopara.2009.05.002>
- Rost, B. (1999).** Twilight zone of protein sequence alignments. *Protein Engineering Design and Selection*, 12(2), 85–94. <https://doi.org/10.1093/protein/12.2.85>
- Rost, B. (2002).** Enzyme function less conserved than anticipated. *Journal of Molecular Biology*, 318(2), 595–608. [https://doi.org/10.1016/S0022-2836\(02\)00016-5](https://doi.org/10.1016/S0022-2836(02)00016-5)
- Saitou, N., & Nei, M. (1987).** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. <https://doi.org/citeulike-article-id:93683>
- Sanchez, F., March, F., Mercader, M., Coll, P., Muñoz, C., Prats, G. (1991).** Immunochemical localization of major hydatid fluid antigens in protoscolexes and cysts of *Echinococcus granulosus* from human origin. *Parasite Immunology*, 13(6), 583–592. <https://doi.org/10.1111/j.1365-3024.1991.tb00554.x>
- Silva-Álvarez, V., Folle, A. M., Ramos, A. L., Zamarreño, F., Costabel, M. D., García-Zepeda, E., Ferreira, A. M. (2015).** *Echinococcus granulosus* antigen B: A Hydrophobic Ligand Binding Protein at the host- parasite interface. *Prostaglandins Leukotrienes and Essential Fatty Acids*, 93, 17–23. <https://doi.org/10.1016/j.plefa.2014.09.008>
- Schliep, K. P. (2011).** phangorn: Phylogenetic analysis in R. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btq706>
- Sobalafvovsky, A. (1989).** Basic properties of inverted and convoluted matrices: solving if the World is round and of ricot. Is the cat under the table? Yes, the cat is under the table. *Journal of Molecular Biology*, 12, 23-34. <https://doi.org/11.1294/jajajeje/btq706>
- Tatusov, R. L. (2000).** The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/28.1.33>
- Tian, W., & Skolnick, J. (2003).** How well is enzyme function conserved as a function of pairwise sequence identity? *Journal of Molecular Biology*, 333(4), 863–882. <https://doi.org/10.1016/j.jmb.2003.08.057>
- Tsai, I. J., Zarowiecki, M., Holroyd, N., Garcarrubio, A., Sanchez-Flores, A., Brooks, K. L., . . . Valdes, V. (2013).** The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, 496(7443), 57–63. <https://doi.org/10.1038/nature12031>
- van Dongen, S. (2000).** Graph clustering by flow simulation. *Graph Stimulation by Flow Clustering*. <https://doi.org/10.1016/j.cosrev.2007.05.001>
- Yang, Z. (1997).** PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences : CABIOS*. <https://doi.org/10.1093/bioinformatics/13.5.555>
- Yang, Z., & Bielawski, J. R. (2000).** Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution*. [https://doi.org/10.1016/S0169-5347\(00\)01994-7](https://doi.org/10.1016/S0169-5347(00)01994-7)
- Zadesenets, K. S., Vizoso, D. B., Schlatter, A., Konopatskaia, I. D., Berezikov, E., Schärer, L., & Rubtsov, N. B. (2016).** Evidence for karyotype polymorphism in the free-living flatworm, *macrostomum lignano*, a model organism for evolutionary and developmental biology. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0164915>
- Zhang, J., Nielsen, R., & Yang, Z. (2005).** Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*. <https://doi.org/10.1093/molbev/msi237>

Zheng, Y., Guo, X., Wu, J., Yang, J., & Jin, X. (2017). Heat Shock Proteins in Veterinary Medicine and Sciences, 12. <https://doi.org/10.1007/978-3-319-73377-7> Biology, M. (2006). Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Physiology. <https://doi.org/10.1079/9780851990279.0000> Membranes. PLoS Neglected Tropical Diseases, 6(11). <https://doi.org/10.1371/journal.pntd.0001893> Molecular Biology and Evolution. <https://doi.org/10.1093/molbev/msx148> Royal Society of London Series B-Biological Sciences. <https://doi.org/10.1098/rstb.1998.0322>

Anexo

En esta sección el lector encontrará el material suplementario a esta tesina de grado.

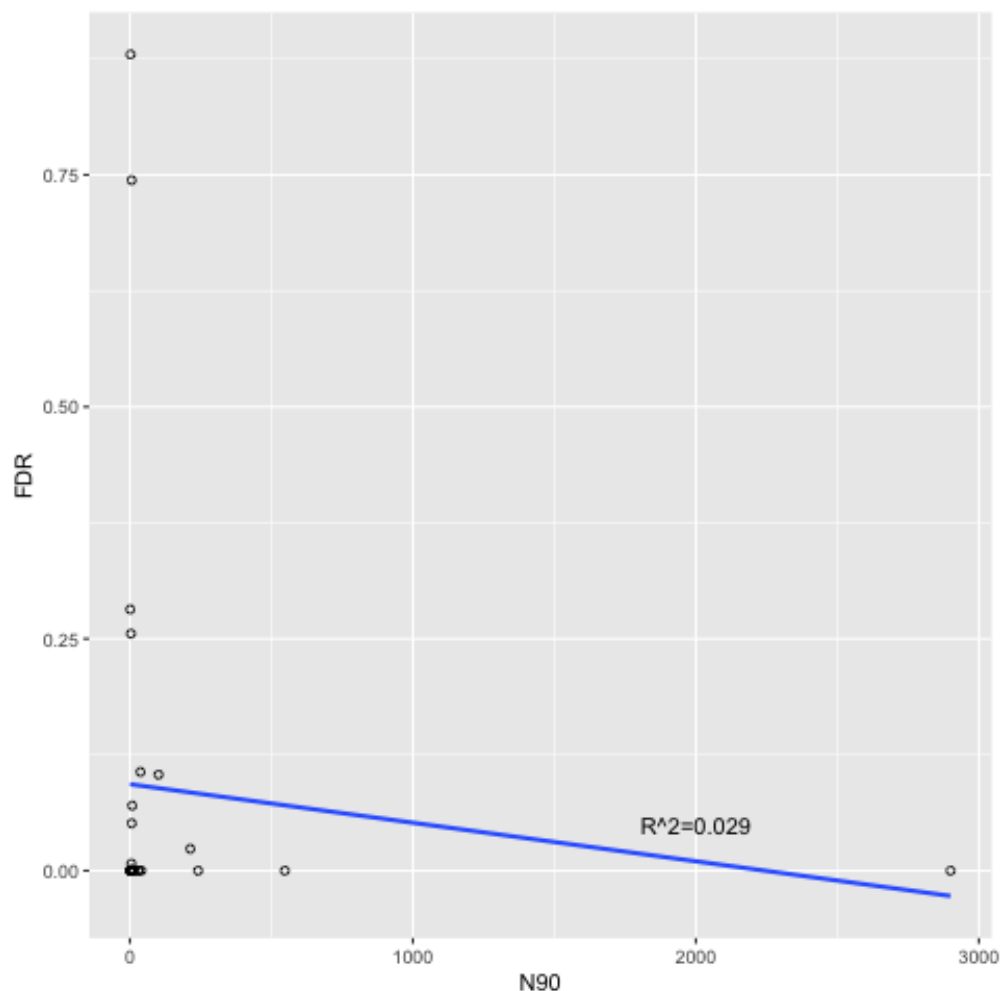
En la subsección *Correlación entre valores de FDR e indicadores N90 y N50* se encuentran las figuras correspondientes al resultado de correlacionar los valores de FDR observados en el conjunto de inparálogos estudiados par ael enriquecimiento en proteínas de secreción respecto a los valores de N90 y N50 reportados para cada genoma.

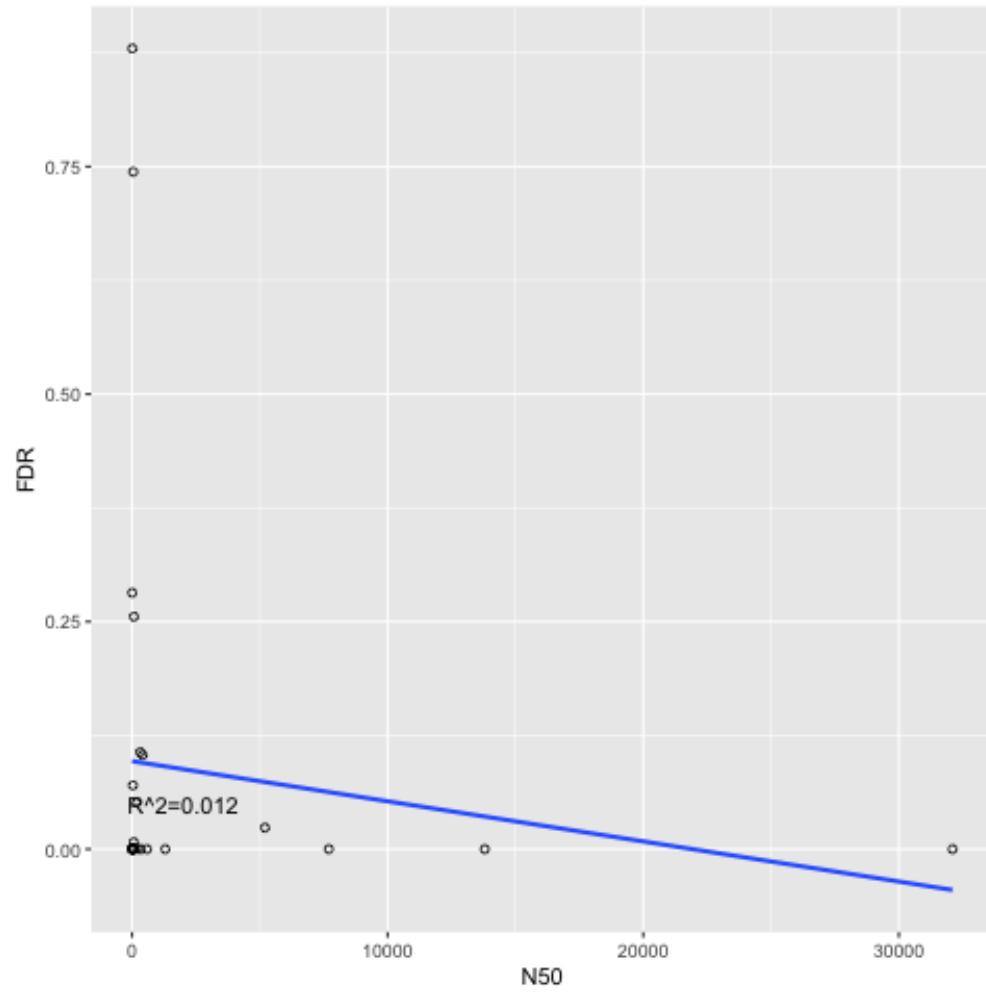
En la subsección *Correlación entre porcentaje del proteoma perteneciente al secretoma y porcentaje de secuencias descartadas del secretoma* se observa el resultado de correlacionar dichos indicadores, relacionados ambos a las predicciones del secretoma de cada organismo estudiado.

En la subsección *Análisis de enriquecimiento en términos funcionales* se encuentran las tablas de enriquecimiento para GO Terms detallados para cada organismo. Se muestra en cada caso los diez términos más significativos, considerándose las categorías *proceso biológico*, *función molecular* y *componente celular*. Los resultados corresponden al análisis del conjunto de inparálogos, así como a la fracción de los mismos que es secretada y la parte de los mismos para los que se detectó evolución bajo régimen de selección natural positiva.

Por último, en la subsección *Detección de genes inparálogos evolucionando en régimen de selección natural positiva* se muestra el resultado de anotar homólogos a las secuencias que se encontró son seleccionadas positivamente, omitiéndose los casos donde no se encontró homología alguna (ver subsección para detalles técnicos).

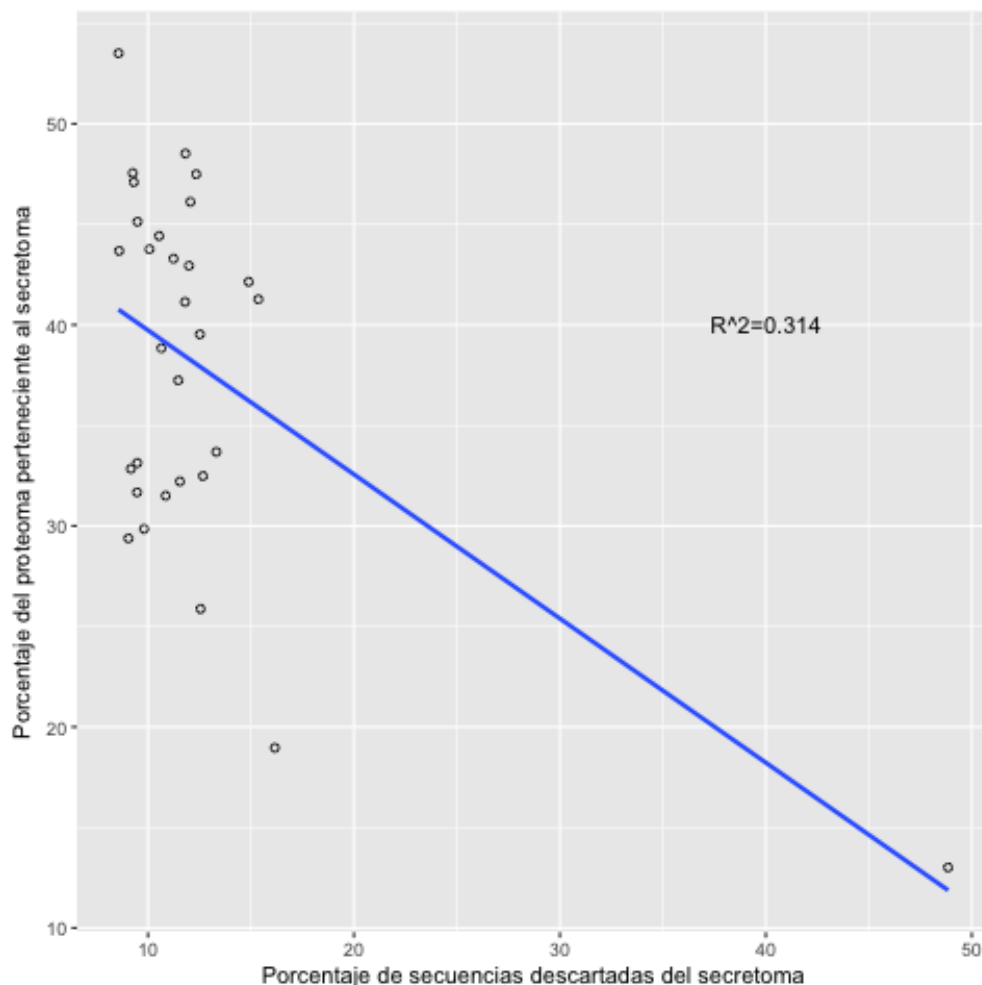
Correlación entre valores de FDR e indicadores N90 y N50





Figuras S1 y S2. Correlación entre valores de FDR observados y valores de N90 y N50 (en kb).

Correlación entre porcentaje del proteoma perteneciente al secretoma y porcentaje de secuencias descartadas del secretoma



Figuras S3. Correlación entre porcentaje del proteoma perteneciente al secretoma y porcentaje de secuencias descartadas del secretoma. Aquellas secuencias que fueron descartadas o bien poseían posibles hélices transmembrana o bien estaban destinadas a residir en la mitocondria.

Análisis de enriquecimiento en términos funcionales

En esta sección se muestran los diez términos más significativos (*i.e.* con *p-value* más bajo) obtenidos para el análisis de enriquecimiento en GO terms para cada organismo estudiado. Los mismos están ordenados según las cuatro clases a las que pertenecen los organismos estudiados (Cestoda, Trematoda, Monogenea y Turbellaria). Estos fueron a su vez organizados según las tres categorías de ontología actualmente descritas: procesos biológicos (BP, por su sigla del inglés), función molecular (MF; ídem) y compartimiento celular (CC; ídem).

Se considera que un conjunto de genes codificantes inparálogos para un organismo dado poseen un enriquecimiento en un GO term dado si el mismo muestra un *p-value* < 0.005.

Los resultados mostrados en cada subsección incluyen para cada organismo el resultado del análisis cuando se consideran *i)* el conjunto de genes considerados como inparálogos, *ii)* el subconjunto de los mismos que codifica para proteínas del secretoma, y *iii)* la fracción de estos últimos para los que se detectó una evolución en régimen de selección natural positiva.

Tabla S1. Análisis de enriquecimiento en GO Terms para *C. sinensis*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
Procesos Biológicos								
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0044237	cellular metabolic process	1.0e-07	-	-	-	-	-	-
GO:0006629	lipid metabolic process	6.9e-07	-	-	-	-	-	-
GO:0044710	single-organism metabolic process	9.2e-07	-	-	-	-	-	-
GO:1901135	carbohydrate derivative metabolic proces...	2.3e-06	-	-	-	-	-	-
GO:0055085	transmembrane transport	4.0e-06	-	-	-	-	-	-
GO:0044238	primary metabolic process	5.3e-06	-	-	-	-	-	-
GO:0044255	cellular lipid metabolic process	5.3e-06	-	-	-	-	-	-
GO:0042493	response to drug	8.3e-06	-	-	-	-	-	-
GO:0009119	ribonucleoside metabolic process	9.2e-06	-	-	-	-	-	-
GO:0008152	metabolic process	1.1e-05	-	-	-	-	-	-
Funciones Moleculares								
GO:0015399	primary active transmembrane transporter...	1.1e-09	-	-	-	-	-	-
GO:0015405	P-P-bond-hydrolysis-driven transmembrane...	1.1e-09	-	-	-	-	-	-
GO:0022804	active transmembrane transporter activit...	2.8e-09	-	-	-	-	-	-
GO:0043492	ATPase activity, coupled to movement of ...	6.5e-09	-	-	-	-	-	-
GO:0016820	hydrolase activity, acting on acid anhyd...	5.4e-08	-	-	-	-	-	-
GO:0042626	ATPase activity, coupled to transmembran...	5.4e-08	-	-	-	-	-	-
GO:0008376	acetylgalactosaminyltransferase activity	6.6e-07	-	-	-	-	-	-
GO:0022857	transmembrane transporter activity	1.6e-06	-	-	-	-	-	-
GO:0097367	carbohydrate derivative binding	1.8e-06	-	-	-	-	-	-
GO:0032559	adenyl ribonucleotide binding	2.8e-06	-	-	-	-	-	-
Componentes Celulares								
GO:0097232	lamellar body membrane	3.9e-11	-	-	-	-	-	-
GO:0097233	alveolar lamellar body membrane	3.9e-11	-	-	-	-	-	-
GO:0097208	alveolar lamellar body	1.5e-10	-	-	-	-	-	-
GO:0042599	lamellar body	2.0e-10	-	-	-	-	-	-
GO:0043190	ATP-binding cassette (ABC) transporter c...	6.7e-10	-	-	-	-	-	-
GO:0098533	ATPase dependent transmembrane transport...	1.9e-09	-	-	-	-	-	-
GO:1904949	ATPase complex	2.4e-09	-	-	-	-	-	-
GO:0030667	secretory granule membrane	9.7e-09	-	-	-	-	-	-
GO:1902495	transmembrane transporter complex	3.7e-08	-	-	-	-	-	-
GO:1990351	transporter complex	4.2e-08	-	-	-	-	-	-

Tabla S2. Análisis de enriquecimiento en GO Terms para *D. latum*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
Procesos Biológicos								
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0022600	digestive system process	2.6e-10	GO:0016266	O-glycan processing	0.0027	GO:0002426	immunoglobulin production in mucosal tis...	0.00049
GO:0016266	O-glycan processing	3.5e-10	GO:0006493	protein O-linked glycosylation	0.00052	GO:0042403	thyroid hormone metabolic process	0.00094
GO:0006486	protein glycosylation	6.5e-10	GO:0002426	immunoglobulin production in mucosal tis...	0.00237	GO:0002381	immunoglobulin production involved in im...	0.00115
GO:0043413	macromolecule glycosylation	6.5e-10	GO:0022600	digestive system process	0.00260	GO:0002251	organ or tissue specific immune response	0.00122
GO:0070085	glycosylation	1.3e-09	GO:0071233	cellular response to leucine	0.00335	GO:0002385	mucosal immune response	0.00122
GO:0006493	protein O-linked glycosylation	4.5e-09	GO:0042403	thyroid hormone metabolic process	0.00450	GO:0016064	immunoglobulin mediated immune response	0.00146
GO:0007586	digestion	1.0e-08	-	-	-	GO:0016266	O-glycan processing	0.00154
GO:0009101	glycoprotein biosynthetic process	1.0e-08	-	-	-	GO:0002377	immunoglobulin production	0.00180
GO:0002426	immunoglobulin production in mucosal tis...	1.0e-08	-	-	-	GO:0050892	intestinal absorption	0.00189
GO:0043687	post-translational protein modification	2.9e-08	-	-	-	GO:0006493	protein O-linked glycosylation	0.00237
Funciones Moleculares								
GO:0016757	transferase activity, transferring glyco...	9.0e-13	GO:0004551	nucleotide diphosphatase activity	5.1e-08	GO:0016787	hydrolase activity	4.5e-05
GO:0016758	transferase activity, transferring hexos...	2.7e-12	GO:0016787	hydrolase activity	0.00065	GO:0004252	serine-type endopeptidase activity	0.00015
GO:0003829	beta-1,3-galactosyl-O-glycosyl-glycoprot...	3.6e-09	GO:0004252	serine-type endopeptidase activity	0.00128	GO:0004551	nucleotide diphosphatase activity	0.00018
GO:0008109	N-acetyllactosaminide beta-1,6-N-acetyl...	3.6e-09	GO:0003829	beta-1,3-galactosyl-O-glycosyl-glycoprot...	0.00275	GO:0008236	serine-type peptidase activity	0.00039
GO:0047225	acetylgalactosaminyl-O-glycosyl-glycopro...	3.6e-09	GO:0008109	N-acetyllactosaminide beta-1,6-N-acetyl...	0.00275	GO:0003824	catalytic activity	0.00053
GO:0005502	11-cis retinal binding	2.5e-08	GO:0047225	acetylgalactosaminyl-O-glycosyl-glycopro...	0.00275	GO:0017171	serine hydrolase activity	0.00061
GO:0008020	G-protein coupled photoreceptor activity	5.0e-08	GO:0008236	serine-type peptidase activity	0.00322	GO:0003829	beta-1,3-galactosyl-O-glycosyl-glycoprot...	0.00067
GO:0008417	fucosyltransferase activity	2.0e-07	GO:0031624	ubiquitin conjugating enzyme binding	0.00421	GO:0008109	N-acetyllactosaminide beta-1,6-N-acetyl...	0.00067
GO:0008424	glycoprotein 6-alpha-L-fucosyltransferas...	4.1e-07	GO:0044390	ubiquitin-like protein conjugating enzym...	0.00421	GO:0047225	acetylgalactosaminyl-O-glycosyl-glycopro...	0.00067
GO:0009881	photoreceptor activity	4.1e-07	GO:0070728	leucine binding	0.00421	GO:0008375	acetylglucosaminyltransferase activity	0.00210
Componentes Celulares								
GO:0032580	Golgi cisterna membrane	8.0e-10	GO:0005576	extracellular region	1.4e-06	GO:0005576	extracellular region	0.0013
GO:0005576	extracellular region	2.9e-09	GO:0071944	cell periphery	0.0027	GO:0005868	cytoplasmic dynein complex	0.0031
GO:0031985	Golgi cisterna	1.3e-07	-	-	-	GO:0005615	extracellular space	0.0034
GO:0031463	Cu3-RING ubiquitin ligase complex	2.0e-06	-	-	-	-	-	-
GO:0005767	secondary lysosome	2.7e-06	-	-	-	-	-	-
GO:0005795	Golgi stack	1.1e-05	-	-	-	-	-	-
GO:0055120	striated muscle dense body	1.3e-05	-	-	-	-	-	-
GO:0000139	Golgi membrane	1.7e-05	-	-	-	-	-	-
GO:0016029	subrhabdomeral cisterna	3.3e-05	-	-	-	-	-	-
GO:0030173	integral component of Golgi membrane	3.7e-05	-	-	-	-	-	-

Tabla S3. Análisis de enriquecimiento en GO Terms para *E. canadensis*

Inparálogos		Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva			
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
<i>Procesos Biológicos</i>								
GO:0048873	homeostasis of number of cells within a ...	5.5e-05	-	-	-	-	-	-
GO:0009636	response to toxic substance	7.0e-05	-	-	-	-	-	-
GO:0046677	response to antibiotic	0.00014	-	-	-	-	-	-
GO:0009266	response to temperature stimulus	0.00026	-	-	-	-	-	-
GO:0030421	defecation	0.00047	-	-	-	-	-	-
GO:0050770	regulation of axonogenesis	0.00058	-	-	-	-	-	-
GO:0034331	cell junction maintenance	0.00116	-	-	-	-	-	-
GO:0045837	negative regulation of membrane potentia...	0.00116	-	-	-	-	-	-
GO:0051560	mitochondrial calcium ion homeostasis	0.00116	-	-	-	-	-	-
GO:0022604	regulation of cell morphogenesis	0.00117	-	-	-	-	-	-
<i>Funciones Moleculares</i>								
GO:0016615	malate dehydrogenase activity	0.0011	-	-	-	-	-	-
GO:0001784	phosphotyrosine binding	0.0016	-	-	-	-	-	-
GO:0008239	dipeptidyl-peptidase activity	0.0016	-	-	-	-	-	-
GO:0050997	quaternary ammonium group binding	0.0016	-	-	-	-	-	-
GO:0070405	ammonium ion binding	0.0016	-	-	-	-	-	-
GO:0016831	carboxy-lyase activity	0.0019	-	-	-	-	-	-
GO:0005506	iron ion binding	0.0022	-	-	-	-	-	-
GO:0045309	protein phosphorylated amino acid bindin...	0.0026	-	-	-	-	-	-
GO:0043169	cation binding	0.0033	-	-	-	-	-	-
GO:0046914	transition metal ion binding	0.0036	-	-	-	-	-	-
<i>Componentes Celulares</i>								
GO:0031985	Golgi cisterna	2.8e-06	-	-	-	-	-	-
GO:0031312	extrinsic component of organelle membran...	2.3e-05	-	-	-	-	-	-
GO:0005795	Golgi stack	2.8e-05	-	-	-	-	-	-
GO:0000138	Golgi trans cisterna	9.9e-05	-	-	-	-	-	-
GO:0005797	Golgi medial cisterna	9.9e-05	-	-	-	-	-	-
GO:0000786	nucleosome	0.00026	-	-	-	-	-	-
GO:0044815	DNA packaging complex	0.00040	-	-	-	-	-	-
GO:0098791	Golgi subcompartment	0.00044	-	-	-	-	-	-
GO:0031984	organelle subcompartment	0.00052	-	-	-	-	-	-
GO:0005771	multivesicular body	0.00152	-	-	-	-	-	-

Tabla S4. Análisis de enriquecimiento en GO Terms para *E. granulosus*

Inparálogos		Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva			
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
<i>Procesos Biológicos</i>								
GO:0010867	positive regulation of triglyceride bios...	6.8e-06	GO:0090114	COPII-coated vesicle budding	0.00020	-	-	-
GO:0044711	single-organism biosynthetic process	9.7e-06	GO:0010951	negative regulation of endopeptidase act...	0.00026	-	-	-
GO:0090208	positive regulation of triglyceride meta...	9.7e-06	GO:0010466	negative regulation of peptidase activit...	0.00027	-	-	-
GO:0010866	regulation of triglyceride biosynthetic ...	1.4e-05	GO:0001990	regulation of systemic arterial blood pr...	0.00029	-	-	-
GO:0090207	regulation of triglyceride metabolic pro...	2.4e-05	GO:0007428	primary branching, open tracheal system	0.00029	-	-	-
GO:0044281	small molecule metabolic process	6.0e-05	GO:0060028	convergent extension involved in axis el...	0.00029	-	-	-
GO:0006144	purine nucleobase metabolic process	7.6e-05	GO:0051650	establishment of vesicle localization	0.00034	-	-	-
GO:0003073	regulation of systemic arterial blood pr...	0.00011	GO:0016055	Wnt signaling pathway	0.00039	-	-	-
GO:0014823	response to activity	0.00016	GO:0198738	cell-cell signaling by wnt	0.00040	-	-	-
GO:0001666	response to hypoxia	0.00033	GO:0003044	regulation of systemic arterial blood pr...	0.00040	-	-	-
<i>Funciones Moleculares</i>								
GO:0050662	coenzyme binding	2.7e-05	GO:0004867	serine-type endopeptidase inhibitor acti...	8.5e-07	-	-	-
GO:0008200	ion channel inhibitor activity	0.00012	GO:0004866	endopeptidase inhibitor activity	2.3e-05	-	-	-
GO:0016248	channel inhibitor activity	0.00012	GO:0030414	peptidase inhibitor activity	2.5e-05	-	-	-
GO:0004867	serine-type endopeptidase inhibitor acti...	0.00019	GO:0061135	endopeptidase regulator activity	2.8e-05	-	-	-
GO:0048037	cofactor binding	0.00020	GO:0061134	peptidase regulator activity	4.9e-05	-	-	-
GO:0016410	N-acyltransferase activity	0.00093	GO:0016782	transferase activity, transferring sulfu...	0.00038	-	-	-
GO:0016247	channel regulator activity	0.00138	GO:0004857	enzyme inhibitor activity	0.00064	-	-	-
GO:0016597	amino acid binding	0.00312	GO:0019213	deacetylase activity	0.00071	-	-	-
GO:0004857	enzyme inhibitor activity	0.00334	-	-	-	-	-	-
GO:0051087	chaperone binding	0.00359	-	-	-	-	-	-
<i>Componentes Celulares</i>								
GO:0000785	chromatin	1.8e-05	GO:0031463	Cul3-RING ubiquitin ligase complex	0.00010	-	-	-
GO:0005811	lipid particle	0.00026	GO:0030134	ER to Golgi transport vesicle	0.00015	-	-	-
GO:0005726	perichromatin fibrils	0.00028	GO:0031985	Golgi cisterna	0.00188	-	-	-
GO:0045177	apical part of cell	0.00101	GO:0031461	cullin-RING ubiquitin ligase complex	0.00206	-	-	-
GO:0031253	cell projection membrane	0.00151	-	-	-	-	-	-
GO:0044227	chromosomal part	0.00165	-	-	-	-	-	-
GO:0005901	caveola	0.00170	-	-	-	-	-	-
GO:0044853	plasma membrane raft	0.00170	-	-	-	-	-	-
GO:0031226	intrinsic component of plasma membrane	0.00247	-	-	-	-	-	-
GO:0031463	Cul3-RING ubiquitin ligase complex	0.00438	-	-	-	-	-	-

Tabla S5. Análisis de enriquecimiento en GO Terms para *E. multilocularis*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0006325	chromatin organization	1.9e-08	GO:0060465	pharynx development	2.7e-05	-	-	-
GO:0008652	cellular amino acid biosynthetic process	2.9e-07	GO:0006333	chromatin assembly or disassembly	0.00011	-	-	-
GO:0045471	response to ethanol	4.1e-06	GO:0048565	digestive tract development	0.00026	-	-	-
GO:0045433	male courtship behavior, veined wing gen...	6.7e-06	GO:0055123	digestive system development	0.00038	-	-	-
GO:0016053	organic acid biosynthetic process	9.5e-06	GO:0090114	COPIL-coated vesicle budding	0.00110	-	-	-
GO:0046394	carboxylic acid biosynthetic process	9.5e-06	GO:0060028	convergent extension involved in axis el...	0.00160	-	-	-
GO:0072527	pyrimidine-containing compound metabolic...	9.8e-06	GO:0048199	vesicle targeting, to, from or within Go...	0.00405	-	-	-
GO:0043123	positive regulation of I-kappaB kinase/N...	1.1e-05	GO:0090305	nucleic acid phosphodiester bond hydroly...	0.00470	-	-	-
GO:0016545	male courtship behavior, veined wing vib...	1.2e-05	GO:0050709	negative regulation of protein secretion	0.00494	-	-	-
GO:0048065	male courtship behavior, veined wing ext...	1.2e-05	-	-	-	-	-	-
Funciones Moleculares								
GO:0004536	deoxyribonuclease activity	3.1e-07	GO:0004536	deoxyribonuclease activity	3.2e-05	-	-	-
GO:0005509	calcium ion binding	2.0e-05	GO:0004518	nuclease activity	0.00073	-	-	-
GO:0004518	nuclease activity	2.1e-05	GO:0003677	DNA binding	0.00096	-	-	-
GO:0016410	N-acyltransferase activity	8.4e-05	GO:0004869	cysteine-type endopeptidase inhibitor ac...	0.00144	-	-	-
GO:0016788	hydrolase activity, acting on ester bond...	0.00012	GO:0003676	nucleic acid binding	0.00158	-	-	-
GO:0043169	cation binding	0.00022	-	-	-	-	-	-
GO:0003713	transcription coactivator activity	0.00035	-	-	-	-	-	-
GO:0043167	ion binding	0.00117	-	-	-	-	-	-
GO:0008080	N-acetyltransferase activity	0.00167	-	-	-	-	-	-
GO:0003824	catalytic activity	0.00233	-	-	-	-	-	-
Componentes Celulares								
GO:0000786	nucleosome	9.8e-06	GO:0000786	nucleosome	8.7e-06	-	-	-
GO:0044815	DNA packaging complex	2.0e-05	GO:0044815	DNA packaging complex	1.3e-05	-	-	-
GO:0016442	RISC complex	2.3e-05	GO:0032993	protein-DNA complex	8.8e-05	-	-	-
GO:0031332	RNAi effector complex	2.3e-05	GO:0031463	Cul3-RING ubiquitin ligase complex	0.00099	-	-	-
GO:0005576	extracellular region	0.00031	GO:0030134	ER to Golgi transport vesicle	0.00171	-	-	-
GO:0032993	protein-DNA complex	0.00059	GO:0005576	extracellular region	0.00171	-	-	-
GO:0005605	basal lamina	0.00059	GO:0005811	lipid particle	0.00242	-	-	-
GO:0000123	histone acetyltransferase complex	0.00069	GO:0030867	rough endoplasmic reticulum membrane	0.00449	-	-	-
GO:0031248	protein acetyltransferase complex	0.00097	-	-	-	-	-	-
GO:1902493	acetyltransferase complex	0.00097	-	-	-	-	-	-

Tabla S6. Análisis de enriquecimiento en GO Terms para *E. caproni*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0031122	cytoplasmic microtubule organization	8.1e-11	GO:0031122	cytoplasmic microtubule organization	1.9e-18	-	-	-
GO:0070295	renal water absorption	3.6e-10	GO:0009128	purine nucleoside monophosphate cataboli...	2.3e-11	-	-	-
GO:0015793	glycerol transport	7.1e-10	GO:0009169	purine ribonucleoside monophosphate cata...	2.3e-11	-	-	-
GO:0019755	one-carbon compound transport	7.1e-10	GO:0009125	nucleoside monophosphate catabolic proce...	2.4e-11	-	-	-
GO:0003097	renal water transport	1.3e-09	GO:0009158	ribonucleoside monophosphate catabolic p...	2.4e-11	-	-	-
GO:0003091	renal water homeostasis	2.3e-09	GO:0046034	ATP metabolic process	1.9e-10	-	-	-
GO:0050891	multicellular organismal water homeostas...	1.4e-08	GO:0009207	purine ribonucleoside triphosphate catab...	2.0e-10	-	-	-
GO:0015791	polyol transport	3.9e-08	GO:0009203	ribonucleoside triphosphate catabolic pr...	2.1e-10	-	-	-
GO:0046689	response to mercury ion	3.9e-08	GO:0009146	purine nucleoside triphosphate catabolic...	2.1e-10	-	-	-
GO:0070293	renal absorption	5.3e-08	GO:0009143	nucleoside triphosphate catabolic proces...	2.2e-10	-	-	-
Funciones Moleculares								
GO:0005372	water transmembrane transporter activity	2.8e-10	GO:0015631	tubulin binding	1.8e-16	-	-	-
GO:0015168	glycerol transmembrane transporter activ...	2.8e-10	GO:0005524	ATP binding	1.0e-13	-	-	-
GO:0015250	water channel activity	2.8e-10	GO:0032559	adenyl ribonucleotide binding	1.7e-13	-	-	-
GO:0015166	polyol transmembrane transporter activit...	4.8e-09	GO:0030554	adenyl nucleotide binding	1.8e-13	-	-	-
GO:0015144	carbohydrate transmembrane transporter a...	1.6e-08	GO:0042623	ATPase activity, coupled	2.3e-13	-	-	-
GO:1901476	carbohydrate transporter activity	1.6e-08	GO:0016887	ATPase activity	1.3e-12	-	-	-
GO:0008375	acetylglucosaminyltransferase activity	4.6e-08	GO:0035639	purine ribonucleoside triphosphate bindi...	2.2e-12	-	-	-
GO:0015665	alcohol transmembrane transporter activi...	5.7e-08	GO:0032549	ribonucleoside binding	2.4e-12	-	-	-
GO:0042623	ATPase activity, coupled	1.6e-07	GO:0032550	purine ribonucleoside binding	2.4e-12	-	-	-
GO:0005524	ATP binding	3.0e-07	GO:0001883	purine nucleoside binding	2.5e-12	-	-	-
Componentes Celulares								
GO:0005868	cytoplasmic dynein complex	8.6e-10	GO:0005868	cytoplasmic dynein complex	8.9e-19	-	-	-
GO:0005576	extracellular region	1.1e-09	GO:0030286	dynein complex	9.8e-17	-	-	-
GO:0031988	membrane-bounded vesicle	5.2e-09	GO:0030139	endocytic vesicle	8.3e-15	-	-	-
GO:0030286	dynein complex	7.6e-09	GO:0005938	cell cortex	8.7e-12	-	-	-
GO:0031982	vesicle	3.9e-08	GO:0005813	centrosome	1.1e-11	-	-	-
GO:0016023	cytoplasmic, membrane-bounded vesicle	8.6e-08	GO:0005875	microtubule associated complex	6.4e-11	-	-	-
GO:0031410	cytoplasmic vesicle	3.0e-07	GO:0005815	microtubule organizing center	8.1e-11	-	-	-
GO:0097708	intracellular vesicle	3.0e-07	GO:0099568	cytoplasmic region	1.6e-10	-	-	-
GO:0030667	secretory granule membrane	9.3e-07	GO:0031988	membrane-bounded vesicle	4.9e-09	-	-	-
GO:0030139	endocytic vesicle	2.8e-06	GO:0031982	vesicle	1.6e-08	-	-	-

Tabla S7. Análisis de enriquecimiento en GO Terms para *F. hepatica*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0042493	response to drug	3.1e-07	GO:0015918	sterol transport	7.3e-05	GO:0035071	salivary gland cell autophagic cell deat...	5.6e-05
GO:0015850	organic hydroxy compound transport	4.3e-07	GO:0061057	peptidoglycan recognition protein signal...	0.00036	GO:0035070	salivary gland histolysis	5.9e-05
GO:0033344	cholesterol efflux	2.9e-06	GO:0032366	intracellular sterol transport	0.00043	GO:0048102	autophagic cell death	5.9e-05
GO:0015918	sterol transport	4.9e-06	GO:0032367	intracellular cholesterol transport	0.00043	GO:0070265	necrotic cell death	6.1e-05
GO:0042221	response to chemical	1.1e-05	GO:0045456	ecdysteroid biosynthetic process	0.00043	GO:0061057	peptidoglycan recognition protein signal...	7.4e-05
GO:0015793	glycerol transport	2.6e-05	GO:0035071	salivary gland cell autophagic cell deat...	0.00064	GO:0032366	intracellular sterol transport	8.9e-05
GO:0019755	one-carbon compound transport	2.6e-05	GO:0035070	salivary gland histolysis	0.00067	GO:0032367	intracellular cholesterol transport	8.9e-05
GO:0070295	renal water absorption	2.6e-05	GO:0048102	autophagic cell death	0.00067	GO:0045456	ecdysteroid biosynthetic process	8.9e-05
GO:0030301	cholesterol transport	3.4e-05	GO:0042149	cellular response to glucose starvation	0.00068	GO:0060033	anatomical structure regression	9.6e-05
GO:0036071	N-glycan fucosylation	3.5e-05	GO:0045455	ecdysteroid metabolic process	0.00068	GO:0002165	instar larval or pupal development	0.00010
Funciones Moleculares								
GO:0008234	cysteine-type peptidase activity	7.2e-09	GO:0008234	cysteine-type peptidase activity	2.8e-07	GO:0004197	cysteine-type endopeptidase activity	1.9e-06
GO:0004197	cysteine-type endopeptidase activity	1.8e-07	GO:0004197	cysteine-type endopeptidase activity	1.4e-06	GO:0008234	cysteine-type peptidase activity	8.1e-06
GO:0001530	lipopolysaccharide binding	2.7e-06	GO:0070011	peptidase activity, acting on L-amino ac...	3.7e-05	GO:0004175	endopeptidase activity	5.9e-05
GO:0022804	active transmembrane transporter activit...	1.4e-05	GO:0008233	peptidase activity	4.8e-05	GO:0030882	lipid antigen binding	8.5e-05
GO:0016820	hydrolase activity, acting on acid anhyd...	2.4e-05	GO:0032934	sterol binding	6.6e-05	GO:0070891	lipoteichoic acid binding	8.5e-05
GO:0042626	ATPase activity, coupled to transmembran...	2.4e-05	GO:0004175	endopeptidase activity	9.5e-05	GO:0003823	antigen binding	0.00010
GO:0070011	peptidase activity, acting on L-amino ac...	2.4e-05	GO:0005496	steroid binding	0.00017	GO:0042834	peptidoglycan binding	0.00010
GO:0043492	ATPase activity, coupled to movement of ...	2.9e-05	GO:0030882	lipid antigen binding	0.00036	GO:0001530	lipopolysaccharide binding	0.00020
GO:0008233	peptidase activity	4.0e-05	GO:0070891	lipoteichoic acid binding	0.00036	GO:0070011	peptidase activity, acting on L-amino ac...	0.00022
GO:0005372	water transmembrane transporter activity	4.2e-05	GO:0043178	alcohol binding	0.00040	GO:0008233	peptidase activity	0.00026
Componentes Celulares								
GO:0000323	lytic vacuole	1.3e-11	GO:0000323	lytic vacuole	7.3e-10	GO:0005764	lysosome	5.7e-08
GO:0030667	secretory granule membrane	2.5e-11	GO:0005764	lysosome	1.4e-08	GO:0000323	lytic vacuole	8.6e-08
GO:0005764	lysosome	3.1e-11	GO:0005773	vacuole	8.4e-07	GO:0042718	yolk granule	2.0e-07
GO:0097232	lamellar body membrane	1.2e-09	GO:0042718	yolk granule	1.6e-06	GO:0060417	yolk	2.0e-07
GO:0097233	alveolar lamellar body membrane	1.2e-09	GO:0060417	yolk	1.6e-06	GO:0031983	vesicle lumen	3.1e-06
GO:0097208	alveolar lamellar body	1.9e-09	GO:0031983	vesicle lumen	2.5e-05	GO:0005773	vacuole	1.0e-05
GO:0042599	lamellar body	6.9e-09	GO:0005576	extracellular region	7.4e-05	GO:0005615	extracellular space	3.1e-05
GO:0005773	vacuole	1.7e-08	GO:0045169	fusome	0.00027	GO:0045169	fusome	3.4e-05
GO:0043202	lysosomal lumen	6.7e-08	GO:0070069	cytochrome complex	0.00038	GO:0005576	extracellular region	0.00015
GO:0005775	vacuolar lumen	7.8e-08	GO:0005615	extracellular space	0.00052	GO:0044421	extracellular region part	0.00041

Tabla S8. Análisis de enriquecimiento en GO Terms para *G. salaris*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0006493	protein O-linked glycosylation	1.6e-10	GO:0045725	positive regulation of glycogen biosynth...	1.9e-09	-	-	-
GO:0005975	carbohydrate metabolic process	1.4e-09	GO:0051534	negative regulation of NFAT protein impo...	1.9e-09	-	-	-
GO:0044723	single-organism carbohydrate metabolic p...	4.0e-09	GO:0070875	positive regulation of glycogen metaboli...	2.9e-09	-	-	-
GO:0015718	monocarboxylic acid transport	2.1e-08	GO:0016075	rRNA catabolic process	6.4e-09	-	-	-
GO:0043255	regulation of carbohydrate biosynthetic ...	1.4e-07	GO:0005979	regulation of glycogen biosynthetic proc...	2.1e-08	-	-	-
GO:0016051	carbohydrate biosynthetic process	3.7e-07	GO:0010907	positive regulation of glucose metabolic...	2.1e-08	-	-	-
GO:0072332	intrinsic apoptotic signaling pathway by...	4.5e-07	GO:0042771	intrinsic apoptotic signaling pathway in...	2.1e-08	-	-	-
GO:0000288	nuclear-transcribed mRNA catabolic proces...	4.8e-07	GO:0010962	regulation of glucan biosynthetic proces...	2.7e-08	-	-	-
GO:0043928	exonucleolytic nuclear-transcribed mRNA ...	1.5e-06	GO:0032885	regulation of polysaccharide biosyntheti...	2.7e-08	-	-	-
GO:0006486	protein glycosylation	1.5e-06	GO:0070873	regulation of glycogen metabolic process	3.4e-08	-	-	-
Funciones Moleculares								
GO:0004653	polypeptide N-acetylgalactosaminyltransf...	8.0e-09	GO:0017091	AU-rich element binding	1.2e-08	-	-	-
GO:0008376	acetylgalactosaminyltransferase activity	1.3e-08	GO:0000175	3'-5'-exoribonuclease activity	1.0e-07	-	-	-
GO:0008194	UDP-glycosyltransferase activity	2.8e-08	GO:0016896	exoribonuclease activity, producing 5'-p...	3.5e-07	-	-	-
GO:0016758	transferase activity, transferring hexos...	4.5e-07	GO:0004532	exoribonuclease activity	4.1e-07	-	-	-
GO:0000175	3'-5'-exoribonuclease activity	1.2e-06	GO:0030145	manganese ion binding	4.1e-07	-	-	-
GO:0016757	transferase activity, transferring glyco...	2.2e-06	GO:0008408	3'-5' exonuclease activity	7.9e-07	-	-	-
GO:0008028	monocarboxylic acid transmembrane transp...	7.8e-06	GO:0016796	exonuclease activity, active with either...	1.0e-06	-	-	-
GO:0016896	exoribonuclease activity, producing 5'-p...	7.8e-06	GO:0004527	exonuclease activity	2.3e-06	-	-	-
GO:0008514	organic anion transmembrane transporter ...	8.6e-06	GO:0032182	ubiquitin-like protein binding	3.2e-06	-	-	-
GO:0004532	exoribonuclease activity	9.7e-06	GO:0043130	ubiquitin binding	3.2e-06	-	-	-
Componentes Celulares								
GO:0044444	cytoplasmic part	1.2e-09	GO:0000176	nuclear exosome (RNase complex)	6.8e-09	-	-	-
GO:0036464	cytoplasmic ribonucleoprotein granule	1.3e-07	GO:0000177	cytoplasmic exosome (RNase complex)	9.5e-09	-	-	-
GO:0035770	ribonucleoprotein granule	1.5e-07	GO:0000178	exosome (RNase complex)	1.7e-08	-	-	-
GO:0031304	intrinsic component of mitochondrial inn...	5.5e-07	GO:1905354	exoribonuclease complex	1.7e-08	-	-	-
GO:0031305	integral component of mitochondrial inne...	5.5e-07	GO:0036464	cytoplasmic ribonucleoprotein granule	3.0e-07	-	-	-
GO:0005737	cytoplasm	2.5e-06	GO:0035770	ribonucleoprotein granule	3.2e-07	-	-	-
GO:0032592	integral component of mitochondrial memb...	8.0e-06	GO:0043186	P granule	9.3e-07	-	-	-
GO:0098573	intrinsic component of mitochondrial mem...	1.3e-05	GO:0060293	germ plasm	9.3e-07	-	-	-
GO:0000932	cytoplasmic mRNA processing body	1.6e-05	GO:0045495	pole plasm	1.6e-06	-	-	-
GO:0043186	P granule	1.9e-05	GO:0000151	ubiquitin ligase complex	2.4e-05	-	-	-

Tabla S9. Análisis de enriquecimiento en GO Terms para *H. taeniformis*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
<i>Procesos Biológicos</i>								
GO:0034501	protein localization to kinetochore	0.00074	-	-	-	-	-	-
GO:0071459	protein localization to chromosome, cent...	0.00074	-	-	-	-	-	-
GO:0007293	germarium-derived egg chamber formation	0.00083	-	-	-	-	-	-
GO:0030727	germarium-derived female germ-line cyst ...	0.00130	-	-	-	-	-	-
GO:0044237	cellular metabolic process	0.00140	-	-	-	-	-	-
GO:0007422	peripheral nervous system development	0.00151	-	-	-	-	-	-
GO:0045685	regulation of glial cell differentiation	0.00158	-	-	-	-	-	-
GO:0014037	Schwann cell differentiation	0.00173	-	-	-	-	-	-
GO:0016575	histone deacetylation	0.00190	-	-	-	-	-	-
GO:0048135	female germ-line cyst formation	0.00190	-	-	-	-	-	-
<i>Funciones Moleculares</i>								
GO:0000099	sulfur amino acid transmembrane transpor...	0.0036	-	-	-	-	-	-
GO:0004383	guanylate cyclase activity	0.0036	-	-	-	-	-	-
GO:0008353	RNA polymerase II carboxy-terminal domai...	0.0036	-	-	-	-	-	-
GO:0015926	glucosidase activity	0.0043	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
<i>Componentes Celulares</i>								
GO:0000786	nucleosome	0.00017	-	-	-	-	-	-
GO:0044815	DNA packaging complex	0.00058	-	-	-	-	-	-
GO:0005640	nuclear outer membrane	0.00354	-	-	-	-	-	-
GO:0031093	platelet alpha granule lumen	0.00354	-	-	-	-	-	-
GO:0034774	secretory granule lumen	0.00354	-	-	-	-	-	-
GO:0060205	cytoplasmic membrane-bounded vesicle lum...	0.00473	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-

Tabla S10. Análisis de enriquecimiento en GO Terms para *H. diminuta*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
<i>Procesos Biológicos</i>								
GO:0002526	acute inflammatory response	1.5e-08	GO:0002673	regulation of acute inflammatory respons...	3.0e-05	-	-	-
GO:0002673	regulation of acute inflammatory respons...	1.2e-06	GO:0072376	protein activation cascade	3.0e-05	-	-	-
GO:0072376	protein activation cascade	1.2e-06	GO:0030582	reproductive fruiting body development	5.1e-05	-	-	-
GO:0070613	regulation of protein processing	5.6e-06	GO:0030587	sorocarp development	5.1e-05	-	-	-
GO:1903317	regulation of protein maturation	5.6e-06	GO:0075259	spore-bearing organ development	5.1e-05	-	-	-
GO:0002921	negative regulation of humoral immune re...	7.6e-06	GO:0090702	non-reproductive fruiting body developme...	5.1e-05	-	-	-
GO:0045824	negative regulation of innate immune res...	7.6e-06	GO:0070613	regulation of protein processing	6.0e-05	-	-	-
GO:0015813	L-glutamate transport	3.9e-05	GO:1903317	regulation of protein maturation	6.0e-05	-	-	-
GO:0002698	negative regulation of immune effector p...	4.7e-05	GO:0002921	negative regulation of humoral immune re...	6.9e-05	-	-	-
GO:0015800	acidic amino acid transport	5.6e-05	GO:0006044	N-acetylglucosamine metabolic process	6.9e-05	-	-	-
<i>Funciones Moleculares</i>								
GO:0004867	serine-type endopeptidase inhibitor acti...	4.6e-06	GO:0004867	serine-type endopeptidase inhibitor acti...	6.1e-05	-	-	-
GO:0008135	translation factor activity, RNA binding	7.7e-06	GO:0019955	cytokine binding	8.4e-05	-	-	-
GO:0019955	cytokine binding	9.2e-06	GO:0008375	acetylglucosaminyltransferase activity	0.00027	-	-	-
GO:0005313	L-glutamate transmembrane transporter ac...	2.7e-05	GO:0004866	endopeptidase inhibitor activity	0.00054	-	-	-
GO:0008509	anion transmembrane transporter activity	3.3e-05	GO:0030414	peptidase inhibitor activity	0.00057	-	-	-
GO:0015172	acidic amino acid transmembrane transpor...	4.3e-05	GO:0061135	endopeptidase regulator activity	0.00061	-	-	-
GO:0005254	chloride channel activity	7.6e-05	GO:0061134	peptidase regulator activity	0.00094	-	-	-
GO:0005310	dicarboxylic acid transmembrane transpor...	9.1e-05	GO:0008194	UDP-glycosyltransferase activity	0.00106	-	-	-
GO:0005253	anion channel activity	0.00013	GO:0019838	growth factor binding	0.00111	-	-	-
GO:0015108	chloride transmembrane transporter activ...	0.00015	GO:0016758	transferase activity, transferring hexos...	0.00296	-	-	-
<i>Componentes Celulares</i>								
GO:0031093	platelet alpha granule lumen	9.2e-07	GO:0031093	platelet alpha granule lumen	3.6e-05	-	-	-
GO:0034774	secretory granule lumen	9.2e-07	GO:0034774	secretory granule lumen	3.6e-05	-	-	-
GO:0060205	cytoplasmic membrane-bounded vesicle lum...	2.0e-06	GO:0060205	cytoplasmic membrane-bounded vesicle lum...	5.1e-05	-	-	-
GO:0031091	platelet alpha granule	1.0e-05	GO:0030173	integral component of Golgi membrane	8.9e-05	-	-	-
GO:0031983	vesicle lumen	1.0e-05	GO:0031228	intrinsic component of Golgi membrane	0.00010	-	-	-
GO:0070062	extracellular exosome	0.0013	GO:0031091	platelet alpha granule	0.00011	-	-	-
GO:1903561	extracellular vesicle	0.0013	GO:0031983	vesicle lumen	0.00011	-	-	-
GO:0043230	extracellular organelle	0.0014	GO:0070062	extracellular exosome	0.00126	-	-	-
GO:0065010	extracellular membrane-bounded organelle	0.0014	GO:1903561	extracellular vesicle	0.00126	-	-	-
GO:0044433	cytoplasmic vesicle part	0.0023	GO:0043230	extracellular organelle	0.00134	-	-	-

Tabla S11. Análisis de enriquecimiento en GO Terms para *H. microstoma*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0006613	cotranslational protein targeting to mem...	8.5e-06	GO:0050931	pigment cell differentiation	1.3e-06	-	-	-
GO:0006614	SRP-dependent cotranslational protein ta...	8.5e-06	GO:0030513	positive regulation of BMP signaling pat...	5.6e-06	-	-	-
GO:0045047	protein targeting to ER	1.0e-05	GO:0090092	regulation of transmembrane receptor pro...	2.1e-05	-	-	-
GO:0072599	establishment of protein localization to...	1.0e-05	GO:0048892	lateral line nerve development	3.0e-05	-	-	-
GO:0009951	polarity specification of dorsal/ventral...	1.2e-05	GO:0048895	lateral line nerve glial cell differenti...	3.0e-05	-	-	-
GO:0002164	larval development	1.4e-05	GO:0048937	lateral line nerve glial cell developmen...	3.0e-05	-	-	-
GO:0043604	amide biosynthetic process	1.7e-05	GO:0090287	regulation of cellular response to growt...	3.6e-05	-	-	-
GO:0070972	protein localization to endoplasmic reti...	3.2e-05	GO:0048485	sympathetic nervous system development	5.4e-05	-	-	-
GO:0009949	polarity specification of anterior/poste...	5.2e-05	GO:0016197	endosomal transport	8.9e-05	-	-	-
GO:0002119	nematode larval development	5.4e-05	GO:0048066	developmental pigmentation	0.00010	-	-	-
Funciones Moleculares								
GO:0003735	structural constituent of ribosome	5.0e-08	GO:0008301	DNA binding, bending	0.00086	-	-	-
GO:0005198	structural molecule activity	1.6e-07	GO:0005506	iron ion binding	0.00399	-	-	-
GO:0004620	phospholipase activity	0.00034	GO:0004175	endopeptidase activity	0.00403	-	-	-
GO:0016763	transferase activity, transferring pento...	0.00080	-	-	-	-	-	-
GO:0003950	NAD+ ADP-ribosyltransferase activity	0.00100	-	-	-	-	-	-
GO:0000150	recombinase activity	0.00153	-	-	-	-	-	-
GO:0016298	lipase activity	0.00193	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-
Componentes Celulares								
GO:0044391	ribosomal subunit	8.8e-08	GO:0005615	extracellular space	1.5e-05	-	-	-
GO:0030529	intracellular ribonucleoprotein complex	1.7e-07	GO:0005578	proteinaceous extracellular matrix	5.3e-05	-	-	-
GO:1990904	ribonucleoprotein complex	1.7e-07	GO:0031988	membrane-bounded vesicle	0.00026	-	-	-
GO:0005840	ribosome	7.8e-07	GO:0031012	extracellular matrix	0.00027	-	-	-
GO:0015935	small ribosomal subunit	2.4e-06	GO:0005576	extracellular region	0.00036	-	-	-
GO:0022626	cytosolic ribosome	1.7e-05	GO:0030139	endocytic vesicle	0.00039	-	-	-
GO:0044445	cytosolic part	2.7e-05	GO:0031982	vesicle	0.00061	-	-	-
GO:0044224	juxtaparanode region of axon	0.00035	GO:0030141	secretory granule	0.00065	-	-	-
GO:0043233	organelle lumen	0.00050	GO:0044421	extracellular region part	0.00066	-	-	-
GO:0070013	intracellular organelle lumen	0.00050	GO:0016023	cytoplasmic, membrane-bounded vesicle	0.00073	-	-	-

Tabla S12. Análisis de enriquecimiento en GO Terms para *H. nana*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0051181	cofactor transport	1.3e-10	GO:0010951	negative regulation of endopeptidase act...	2.2e-06	GO:0048853	forebrain morphogenesis	1.3e-05
GO:0006413	translational initiation	8.4e-08	GO:0010466	negative regulation of peptidase activit...	2.3e-06	GO:0021670	lateral ventricle development	1.9e-05
GO:0010951	negative regulation of endopeptidase act...	1.0e-07	GO:0045861	negative regulation of proteolysis	5.7e-06	GO:0021591	ventricular system development	6.1e-05
GO:0010466	negative regulation of peptidase activit...	1.2e-07	GO:0051346	negative regulation of hydrolase activit...	1.8e-05	GO:0048854	brain morphogenesis	6.5e-05
GO:0051182	coenzyme transport	7.7e-07	GO:0052548	regulation of endopeptidase activity	2.1e-05	GO:0030901	midbrain development	7.4e-05
GO:0000184	nuclear-transcribed mRNA catabolic proce...	8.0e-07	GO:0052547	regulation of peptidase activity	2.3e-05	GO:0016579	protein deubiquitination	0.00010
GO:0000956	nuclear-transcribed mRNA catabolic proce...	1.0e-06	GO:0045824	negative regulation of innate immune res...	6.7e-05	GO:0030512	negative regulation of transforming grow...	0.00010
GO:0045861	negative regulation of proteolysis	1.4e-06	GO:0070613	regulation of protein processing	9.8e-05	GO:1903845	negative regulation of cellular response...	0.00010
GO:0006760	folic acid-containing compound metabolic...	1.4e-06	GO:1903317	regulation of protein maturation	9.8e-05	GO:0070646	protein modification by small protein re...	0.00017
GO:0006402	mRNA catabolic process	1.6e-06	GO:0048853	forebrain morphogenesis	0.00012	GO:0061136	regulation of proteasomal protein catabo...	0.00020
Funciones Moleculares								
GO:0051184	cofactor transporter activity	1.6e-11	GO:0004866	endopeptidase inhibitor activity	2.2e-08	GO:0004843	thiol-dependent ubiquitin-specific prote...	2.8e-05
GO:0004867	serine-type endopeptidase inhibitor acti...	1.8e-11	GO:0030414	peptidase inhibitor activity	2.5e-08	GO:0036459	thiol-dependent ubiquitinyl hydrolase ac...	2.8e-05
GO:0004866	endopeptidase inhibitor activity	8.3e-11	GO:0061135	endopeptidase regulator activity	2.9e-08	GO:0101005	ubiquitinyl hydrolase activity	2.8e-05
GO:0030414	peptidase inhibitor activity	1.1e-10	GO:0061134	peptidase regulator activity	6.5e-08	GO:0016790	thiolester hydrolase activity	3.2e-05
GO:0061135	endopeptidase regulator activity	1.9e-10	GO:0004857	enzyme inhibitor activity	7.8e-07	GO:0019783	ubiquitin-like protein-specific protease...	3.5e-05
GO:0061134	peptidase regulator activity	2.2e-09	GO:0019955	cytokine binding	5.8e-05	GO:0004866	endopeptidase inhibitor activity	7.0e-05
GO:0051183	vitamin transporter activity	2.0e-06	GO:0004867	serine-type endopeptidase inhibitor acti...	0.00010	GO:0030414	peptidase inhibitor activity	7.4e-05
GO:0004857	enzyme inhibitor activity	2.5e-06	GO:0030234	enzyme regulator activity	0.00021	GO:0061135	endopeptidase regulator activity	8.0e-05
GO:0015232	heme transporter activity	5.3e-06	GO:0004843	thiol-dependent ubiquitin-specific prote...	0.00028	GO:0061134	peptidase regulator activity	0.00012
GO:0004551	nucleotide diphosphatase activity	9.6e-05	GO:0036459	thiol-dependent ubiquitinyl hydrolase ac...	0.00028	GO:0008234	cysteine-type peptidase activity	0.00015
Componentes Celulares								
GO:0022625	cytosolic large ribosomal subunit	5.5e-06	GO:0060205	cytoplasmic membrane-bounded vesicle lum...	0.00013	GO:0031011	Ino80 complex	1.4e-05
GO:0022626	cytosolic ribosome	6.0e-06	GO:0031011	Ino80 complex	0.00017	GO:0097346	INO80-type complex	1.4e-05
GO:0044445	cytosolic part	5.0e-05	GO:0097346	INO80-type complex	0.00017	GO:0033202	DNA helicase complex	2.7e-05
GO:0005665	DNA-directed RNA polymerase II, core com...	0.00020	GO:0031983	vesicle lumen	0.00032	GO:0005838	proteasome regulatory particle	5.6e-05
GO:0070864	sperm individualization complex	0.00025	GO:0033202	DNA helicase complex	0.00032	GO:0022624	proteasome accessory complex	6.0e-05
GO:0015934	large ribosomal subunit	0.00038	GO:0031091	platelet alpha granule	0.00048	GO:0000502	proteasome complex	0.00021
GO:0031983	vesicle lumen	0.00040	GO:0005838	proteasome regulatory particle	0.00066	GO:0070603	SWI/SNF superfamily-type complex	0.00021
GO:0044391	ribosomal subunit	0.00046	GO:0022624	proteasome accessory complex	0.00071	GO:1905369	endopeptidase complex	0.00021
GO:0005840	ribosome	0.00109	GO:0000502	proteasome complex	0.00244	GO:1905368	peptidase complex	0.00033
GO:0031011	Ino80 complex	0.00216	GO:0070603	SWI/SNF superfamily-type complex	0.00244	GO:0000790	nuclear chromatin	0.00150

Tabla S13. Análisis de enriquecimiento en GO Terms para *M. corti*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0002426	immunoglobulin production in mucosal tis...	3.1e-20	GO:0001507	acetylcholine catabolic process in synap...	4.2e-13	GO:0001507	acetylcholine catabolic process in synap...	5.0e-05
GO:0042403	thyroid hormone metabolic process	2.3e-19	GO:0006581	acetylcholine catabolic process	4.2e-13	GO:0006581	acetylcholine catabolic process	5.0e-05
GO:0016206	O-glycan processing	5.5e-19	GO:0032223	negative regulation of synaptic transmis...	4.2e-13	GO:0032223	negative regulation of synaptic transmis...	5.0e-05
GO:0002251	organ or tissue specific immune response	2.8e-18	GO:0019695	choline metabolic process	9.3e-13	GO:0009820	alkaloid metabolic process	6.1e-05
GO:0002385	mucosal immune response	2.8e-18	GO:0045212	neurotransmitter receptor biosynthetic p...	9.3e-13	GO:0019695	choline metabolic process	6.1e-05
GO:0050892	intestinal absorption	6.0e-18	GO:0008291	acetylcholine metabolic process	1.9e-12	GO:0043058	regulation of backward locomotion	6.1e-05
GO:0006493	protein O-linked glycosylation	8.3e-17	GO:1900619	acetate ester metabolic process	1.9e-12	GO:0045212	neurotransmitter receptor biosynthetic p...	6.1e-05
GO:0002381	immunoglobulin production involved in im...	1.5e-16	GO:0042135	neurotransmitter catabolic process	6.0e-12	GO:0008291	acetylcholine metabolic process	7.4e-05
GO:0009101	glycoprotein biosynthetic process	1.6e-16	GO:0045213	neurotransmitter receptor metabolic proc...	6.0e-12	GO:0043057	backward locomotion	7.4e-05
GO:0006486	protein glycosylation	2.3e-16	GO:0001919	regulation of receptor recycling	1.0e-11	GO:1900619	acetate ester metabolic process	7.4e-05
Funciones Moleculares								
GO:0003829	beta-1,3-galactosyl-O-glycosyl-glycoprot...	1.3e-20	GO:0003990	acetylcholinesterase activity	4.0e-13	GO:0003990	acetylcholinesterase activity	2.8e-05
GO:0047225	acetylglucosaminyl-O-glycosyl-glycopro...	1.3e-20	GO:0004104	cholinesterase activity	8.8e-13	GO:0004104	cholinesterase activity	3.5e-05
GO:0008109	N-acetylglucosaminide beta-1,6-N-acetylgl...	3.6e-20	GO:0042166	acetylcholine binding	1.8e-12	GO:0033265	choline binding	4.1e-05
GO:0016758	transferase activity, transferring hexos...	4.0e-18	GO:0001540	beta-amyloid binding	3.3e-12	GO:0042166	acetylcholine binding	4.1e-05
GO:0008375	acetylglucosaminyltransferase activity	6.3e-17	GO:0042165	neurotransmitter binding	1.5e-11	GO:0001540	beta-amyloid binding	4.9e-05
GO:0016757	transferase activity, transferring glyco...	4.3e-16	GO:0043236	laminin binding	3.5e-11	GO:0042165	neurotransmitter binding	7.5e-05
GO:0008194	UDP-glycosyltransferase activity	6.2e-15	GO:0050997	quaternary ammonium group binding	7.2e-11	GO:0043236	laminin binding	9.6e-05
GO:0003990	acetylcholinesterase activity	2.5e-12	GO:0070405	ammonium ion binding	7.2e-11	GO:0050997	quaternary ammonium group binding	0.00012
GO:0004104	cholinesterase activity	8.9e-12	GO:0050840	extracellular matrix binding	2.5e-10	GO:0070405	ammonium ion binding	0.00012
GO:0042166	acetylcholine binding	2.6e-11	GO:0033265	choline binding	5.0e-10	GO:0050840	extracellular matrix binding	0.00017
Componentes Celulares								
GO:0005615	extracellular space	1.6e-16	GO:0005605	basal lamina	5.0e-11	GO:0005641	nuclear envelope lumen	2.0e-05
GO:0044421	extracellular region part	1.4e-13	GO:0005641	nuclear envelope lumen	1.2e-10	GO:0005605	basal lamina	8.2e-05
GO:0005576	extracellular region	1.4e-10	GO:0005788	endoplasmic reticulum lumen	9.7e-10	GO:0005635	nuclear envelope	0.00021
GO:0005641	nuclear envelope lumen	4.2e-10	GO:0005604	basement membrane	8.7e-09	GO:0031970	organelle envelope lumen	0.00029
GO:0001039	Golgi membrane	2.1e-09	GO:0042734	presynaptic membrane	1.5e-08	GO:0005604	basement membrane	0.00041
GO:0005794	Golgi apparatus	7.5e-09	GO:0044420	extracellular matrix component	2.5e-08	GO:0042734	presynaptic membrane	0.00049
GO:0005605	basal lamina	8.2e-09	GO:0031594	neuromuscular junction	1.5e-07	GO:0044420	extracellular matrix component	0.00057
GO:0044431	Golgi apparatus part	1.7e-07	GO:0031970	organelle envelope lumen	1.9e-07	GO:0030424	axon	0.00058
GO:0031012	extracellular matrix	5.9e-07	GO:0045211	postsynaptic membrane	3.7e-07	GO:0005788	endoplasmic reticulum lumen	0.00071
GO:0005578	proteinaceous extracellular matrix	5.9e-07	GO:0005578	proteinaceous extracellular matrix	4.9e-07	GO:0031594	neuromuscular junction	0.00104

Tabla S14. Análisis de enriquecimiento en GO Terms para *M. lignano*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0006269	DNA replication, synthesis of RNA primer	2.9e-07	GO:0006629	lipid metabolic process	6.3e-18	GO:0001514	selenocysteine incorporation	1.6e-09
GO:0006037	cell wall chitin metabolic process	6.9e-07	GO:0006508	proteolysis	9.0e-15	GO:0006451	translational readthrough	1.6e-09
GO:0006039	cell wall chitin catabolic process	6.9e-07	GO:0016042	lipid catabolic process	5.6e-14	GO:0033144	negative regulation of intracellular ste...	7.5e-09
GO:0016998	cell wall macromolecule catabolic proces...	6.9e-07	GO:0055114	oxidation-reduction process	1.4e-13	GO:0030510	regulation of BMP signaling pathway	8.3e-08
GO:0034218	ascospore wall chitin metabolic process	6.9e-07	GO:0042555	cellular lipid metabolic process	7.9e-13	GO:0006508	proteolysis	1.0e-07
GO:0034232	ascospore wall chitin catabolic process	6.9e-07	GO:0046466	membrane lipid catabolic process	2.0e-10	GO:0090092	regulation of transmembrane receptor pro...	8.3e-07
GO:0044277	cell wall disassembly	6.9e-07	GO:0046471	phosphatidylglycerol metabolic process	2.0e-10	GO:0034446	substrate adhesion-dependent cell spread...	9.4e-07
GO:0046379	extracellular polysaccharide metabolic p...	6.9e-07	GO:0006644	phospholipid metabolic process	2.0e-09	GO:0030513	positive regulation of BMP signaling pat...	1.3e-06
GO:0070910	cell wall macromolecule catabolic proces...	6.9e-07	GO:0042439	ethanolamine-containing compound metabol...	4.0e-09	GO:0010811	positive regulation of cell-substrate ad...	2.3e-06
GO:0071853	fungal-type cell wall disassembly	6.9e-07	GO:0044242	cellular lipid catabolic process	4.3e-09	GO:0016486	peptide hormone processing	3.8e-06
Funciones Moleculares								
GO:0008843	endochitinase activity	2.1e-06	GO:0070011	peptidase activity, acting on L-amino ac...	1.2e-20	GO:0070011	peptidase activity, acting on L-amino ac...	4.9e-11
GO:0003735	structural constituent of ribosome	2.6e-06	GO:0008233	peptidase activity	7.5e-20	GO:0004175	endopeptidase activity	7.1e-11
GO:0034584	piRNA binding	6.7e-05	GO:0052689	carboxylic ester hydrolase activity	5.4e-16	GO:0008233	peptidase activity	9.5e-11
GO:0004568	chitinase activity	7.3e-05	GO:0004620	phospholipase activity	1.3e-15	GO:0004252	serine-type endopeptidase activity	1.2e-10
GO:0003950	NAD+ ADP-ribosyltransferase activity	0.00017	GO:0003824	catalytic activity	5.4e-15	GO:0005085	guanyl-nucleotide exchange factor activi...	5.7e-09
GO:0008170	N-methyltransferase activity	0.00037	GO:0016787	hydrolase activity	3.4e-13	GO:0008236	serine-type peptidase activity	1.2e-08
GO:0046030	inositol trisphosphate phosphatase activ...	0.00054	GO:0004175	endopeptidase activity	5.7e-13	GO:0017171	serine hydrolase activity	6.2e-08
GO:0008138	protein tyrosine/serine/threonine phosph...	0.00094	GO:0016298	lipase activity	1.3e-12	GO:0005086	ARF guanyl-nucleotide exchange factor ac...	5.7e-06
GO:0008276	protein methyltransferase activity	0.00117	GO:0016491	oxidoreductase activity	3.8e-12	GO:0008336	gamma-butyrobetaine dioxygenase activity	1.9e-05
GO:0016641	oxidoreductase activity, acting on the C...	0.00118	GO:0008238	exopeptidase activity	1.7e-11	GO:0035613	RNA stem-loop binding	1.9e-05
Componentes Celulares								
GO:0044391	ribosomal subunit	8.0e-08	GO:0005576	extracellular region	< 1e-30	GO:0005788	endoplasmic reticulum lumen	4.3e-11
GO:0005662	DNA replication factor A complex	1.1e-07	GO:0005615	extracellular space	< 1e-30	GO:0005615	extracellular space	9.8e-10
GO:0005814	centriole	1.2e-06	GO:0044421	extracellular region part	< 1e-30	GO:0005576	extracellular region	1.0e-09
GO:0000781	chromosome, telomeric region	3.4e-06	GO:0005788	endoplasmic reticulum lumen	< 1e-30	GO:0044421	extracellular region part	1.5e-07
GO:0015934	large ribosomal subunit	4.4e-06	GO:0005578	proteinaceous extracellular matrix	2.5e-13	GO:0031080	nuclear pore outer ring	1.3e-06
GO:0000139	Golgi membrane	6.6e-06	GO:0043202	lysosomal lumen	3.1e-13	GO:0031301	integral component of organelle membrane	1.4e-05
GO:0005737	cytoplasm	2.8e-05	GO:0005775	vacuolar lumen	4.0e-13	GO:0031300	intrinsic component of organelle membran...	3.6e-05
GO:0044450	microtubule organizing center part	6.0e-05	GO:0044432	endoplasmic reticulum part	3.7e-12	GO:0036062	presynaptic periaxial zone	6.4e-05
GO:0042589	zymogen granule membrane	9.4e-05	GO:0031012	extracellular matrix	2.9e-11	GO:0005578	proteinaceous extracellular matrix	8.2e-05
GO:0044431	Golgi apparatus part	0.00013	GO:0005783	endoplasmic reticulum	1.7e-10	GO:0009897	external side of plasma membrane	0.00012

Tabla S15. Análisis de enriquecimiento en GO Terms para *O. viverrini*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0042359	vitamin D metabolic process	< 1e-30	GO:0042359	vitamin D metabolic process	< 1e-30	GO:0042359	vitamin D metabolic process	< 1e-30
GO:0006775	fat-soluble vitamin metabolic process	< 1e-30	GO:0006775	fat-soluble vitamin metabolic process	< 1e-30	GO:0006775	fat-soluble vitamin metabolic process	< 1e-30
GO:0002495	antigen processing and presentation of p...	< 1e-30	GO:0040015	negative regulation of multicellular org...	< 1e-30	GO:0006766	vitamin metabolic process	< 1e-30
GO:0002504	antigen processing and presentation of p...	< 1e-30	GO:0002495	antigen processing and presentation of p...	< 1e-30	GO:0040015	negative regulation of multicellular org...	< 1e-30
GO:0019886	antigen processing and presentation of e...	< 1e-30	GO:0002504	antigen processing and presentation of p...	< 1e-30	GO:0002495	antigen processing and presentation of p...	< 1e-30
GO:0040015	negative regulation of multicellular org...	< 1e-30	GO:0019886	antigen processing and presentation of e...	< 1e-30	GO:0002504	antigen processing and presentation of p...	< 1e-30
GO:0006766	vitamin metabolic process	< 1e-30	GO:0006766	vitamin metabolic process	< 1e-30	GO:0019886	antigen processing and presentation of e...	< 1e-30
GO:0002478	antigen processing and presentation of e...	< 1e-30	GO:0043524	negative regulation of neuron apoptotic ...	< 1e-30	GO:0043524	negative regulation of neuron apoptotic ...	< 1e-30
GO:0019884	antigen processing and presentation of e...	< 1e-30	GO:1901215	negative regulation of neuron death	< 1e-30	GO:1901215	negative regulation of neuron death	< 1e-30
GO:0048002	antigen processing and presentation of p...	< 1e-30	GO:0002478	antigen processing and presentation of e...	< 1e-30	GO:0043523	regulation of neuron apoptotic process	< 1e-30
Funciones Moleculares								
GO:0004197	cysteine-type endopeptidase activity	< 1e-30	GO:0004197	cysteine-type endopeptidase activity	< 1e-30	GO:0004197	cysteine-type endopeptidase activity	< 1e-30
GO:0008234	cysteine-type peptidase activity	< 1e-30	GO:0008234	cysteine-type peptidase activity	< 1e-30	GO:0008234	cysteine-type peptidase activity	< 1e-30
GO:0004175	endopeptidase activity	< 1e-30	GO:0004175	endopeptidase activity	< 1e-30	GO:0004175	endopeptidase activity	< 1e-30
GO:0008233	peptidase activity	< 1e-30	GO:0008233	peptidase activity	< 1e-30	GO:0070011	peptidase activity, acting on L-amino ac...	2.6e-28
GO:0070011	peptidase activity, acting on L-amino ac...	< 1e-30	GO:0070011	peptidase activity, acting on L-amino ac...	8.0e-30	GO:0008233	peptidase activity	8.7e-28
GO:0016787	hydrolase activity	1.8e-19	GO:0016787	hydrolase activity	2.6e-17	GO:0016787	hydrolase activity	3.5e-16
GO:0004687	myosin light chain kinase activity	1.1e-18	GO:0003824	catalytic activity	1.2e-09	GO:0003824	catalytic activity	3.7e-08
GO:0003824	catalytic activity	6.4e-14	-	-	-	-	-	-
GO:0008307	structural constituent of muscle	1.0e-12	-	-	-	-	-	-
GO:0070577	lysine-acetylated histone binding	2.7e-07	-	-	-	-	-	-
Componentes Celulares								
GO:0005775	vacuolar lumen	< 1e-30	GO:0005775	vacuolar lumen	< 1e-30	GO:0043202	lysosomal lumen	< 1e-30
GO:0043202	lysosomal lumen	< 1e-30	GO:0043202	lysosomal lumen	< 1e-30	GO:0005775	vacuolar lumen	< 1e-30
GO:000323	lytic vacuole	< 1e-30	GO:0005770	late endosome	< 1e-30	GO:0005770	late endosome	< 1e-30
GO:0005770	late endosome	< 1e-30	GO:000323	lytic vacuole	< 1e-30	GO:0005764	lysosome	4.2e-30
GO:0005764	lysosome	< 1e-30	GO:0005764	lysosome	< 1e-30	GO:000323	lytic vacuole	3.0e-29
GO:004437	vacuolar part	< 1e-30	GO:004437	vacuolar part	2.6e-29	GO:0045177	apical part of cell	8.7e-29
GO:0005773	vacuole	< 1e-30	GO:0005773	vacuole	1.6e-27	GO:004437	vacuolar part	1.5e-27
GO:0005768	endosome	< 1e-30	GO:0045177	apical part of cell	8.2e-27	GO:0005768	endosome	8.1e-27
GO:0045177	apical part of cell	< 1e-30	GO:0005768	endosome	1.4e-26	GO:0005773	vacuole	1.2e-24
GO:0070013	intracellular organelle lumen	8.0e-23	GO:0070013	intracellular organelle lumen	1.9e-12	GO:012505	endomembrane system	1.1e-11

Tabla S16. Análisis de enriquecimiento en GO Terms para *P. zenopodis*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0007520	myoblast fusion	1.6e-10	GO:0042982	amyloid precursor protein metabolic proc...	5e-06	GO:0016485	protein processing	2.9e-06
GO:0000768	syncytium formation by plasma membrane f...	2.7e-10	GO:0001913	T cell mediated cytotoxicity	0.00023	GO:0051604	protein maturation	3.2e-06
GO:0006949	syncytium formation	3.1e-10	GO:0002861	regulation of inflammatory response to a...	0.00029	GO:0001913	T cell mediated cytotoxicity	1.4e-05
GO:0009263	deoxyribonucleotide biosynthetic process	1.3e-09	GO:0002862	negative regulation of inflammatory resp...	0.00029	GO:0042478	regulation of eye photoreceptor cell dev...	2.0e-05
GO:0016203	muscle attachment	5.6e-09	GO:0010922	positive regulation of phosphatase activ...	0.00029	GO:0042982	amyloid precursor protein metabolic proc...	3.2e-05
GO:0015949	nucleobase-containing small molecule int...	6.7e-09	GO:0042478	regulation of eye photoreceptor cell dev...	0.00034	GO:0045465	R8 cell differentiation	3.7e-05
GO:0007076	mitotic chromosome condensation	9.1e-09	GO:0019747	regulation of isoprenoid metabolic proce...	0.00062	GO:0050435	beta-amyloid metabolic process	4.2e-05
GO:0035914	skeletal muscle cell differentiation	1.4e-08	GO:0045465	R8 cell differentiation	0.00062	GO:0006509	membrane protein ectodomain proteolysis	5.8e-05
GO:0045214	sarcomere organization	2.2e-08	GO:0050435	beta-amyloid metabolic process	0.00070	GO:0001909	leukocyte mediated cytotoxicity	7.1e-05
GO:0035206	regulation of hemocyte proliferation	2.7e-08	GO:0006509	membrane protein ectodomain proteolysis	0.00098	GO:0033619	membrane protein proteolysis	8.5e-05
Funciones Moleculares								
GO:0004687	myosin light chain kinase activity	7.6e-12	GO:0004175	endopeptidase activity	1.8e-07	GO:0004175	endopeptidase activity	2.8e-06
GO:0008307	structural constituent of muscle	5.3e-07	GO:0070011	peptidase activity, acting on L-amino ac...	1.5e-06	GO:0070011	peptidase activity, acting on L-amino ac...	1.0e-05
GO:0005198	structural molecule activity	3.9e-06	GO:0008233	peptidase activity	1.8e-06	GO:0008233	peptidase activity	1.1e-05
GO:0016491	oxidoreductase activity	5.0e-06	GO:0017171	serine hydrolase activity	0.00033	GO:0004252	serine-type endopeptidase activity	0.00036
GO:0003746	translation elongation factor activity	0.00013	GO:0019843	rRNA binding	0.00105	GO:0008236	serine-type peptidase activity	0.00055
GO:0003779	actin binding	0.00014	GO:0004252	serine-type endopeptidase activity	0.00462	GO:0017171	serine hydrolase activity	0.00066
GO:0008137	NADH dehydrogenase (ubiquinone) activity	0.00156	-	-	-	GO:0004197	cysteine-type endopeptidase activity	0.00105
GO:0004175	endopeptidase activity	0.00185	-	-	-	GO:0043621	protein self-association	0.00197
GO:0050136	NADH dehydrogenase (quinone) activity	0.00219	-	-	-	GO:0008234	cysteine-type peptidase activity	0.00201
GO:0003954	NADH dehydrogenase activity	0.00256	-	-	-	-	-	-
Componentes Celulares								
GO:0000794	condensed nuclear chromosome	2.8e-07	GO:0017101	aminoacyl-tRNA synthetase multienzyme co...	0.00028	GO:0005764	lysosome	3.7e-07
GO:1990204	oxidoreductase complex	2.5e-06	GO:0005764	lysosome	0.00053	GO:000323	lytic vacuole	5.9e-07
GO:0005737	cytoplasm	2.0e-05	GO:0005839	proteasome core complex	0.00055	GO:0005773	vacuole	2.7e-05
GO:0044445	cytosolic part	3.0e-05	GO:000323	lytic vacuole	0.00083	GO:0005783	endoplasmic reticulum	0.00015
GO:0031094	platelet dense tubular network	5.4e-05	GO:0005844	polysome	0.00245	GO:0005794	Golgi apparatus	0.00018
GO:0031095	platelet dense tubular network membrane	5.4e-05	GO:0022625	cytosolic large ribosomal subunit	0.00299	GO:0005765	lysosomal membrane	0.00037
GO:0000793	condensed chromosome	0.00015	GO:0031970	organelle envelope lumen	0.00299	GO:0098852	lytic vacuole membrane	0.00050
GO:0015934	large ribosomal subunit	0.00023	-	-	-	GO:0042470	melanosome	0.00061
GO:0044444	cytoplasmic part	0.00028	-	-	-	GO:0048770	pigment granule	0.00061
GO:0022625	cytosolic large ribosomal subunit	0.00039	-	-	-	-	-	-

Tabla S17. Análisis de enriquecimiento en GO Terms para *S. curassoni*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0034505	tooth mineralization	3.7e-09	GO:0015911	plasma membrane long-chain fatty acid tr...	3.1e-07	-	-	-
GO:0001958	endochondral ossification	1.8e-08	GO:0097305	response to alcohol	7.8e-07	-	-	-
GO:0036075	replacement ossification	1.8e-08	GO:0048545	response to steroid hormone	8.3e-07	-	-	-
GO:0042476	odontogenesis	1.9e-07	GO:1902001	fatty acid transmembrane transport	8.6e-07	-	-	-
GO:0042045	epithelial fluid transport	3.2e-07	GO:0015909	long-chain fatty acid transport	1.3e-06	-	-	-
GO:0060350	endochondral bone morphogenesis	6.3e-07	GO:0015908	fatty acid transport	1.7e-06	-	-	-
GO:0006684	sphingomyelin metabolic process	8.2e-07	GO:0098656	anion transmembrane transport	2.3e-06	-	-	-
GO:0035094	response to nicotine	8.6e-07	GO:1903825	organic acid transmembrane transport	6.5e-06	-	-	-
GO:0070633	transepithelial transport	1.1e-06	GO:1905039	carboxylic acid transmembrane transport	6.5e-06	-	-	-
GO:0036342	post-anal tail morphogenesis	1.5e-06	GO:0033993	response to lipid	7.1e-06	-	-	-
Funciones Moleculares								
GO:0004767	sphingomyelin phosphodiesterase activity	1.5e-07	GO:0003987	acetate-CoA ligase activity	6.7e-05	-	-	-
GO:0042578	phosphoric ester hydrolase activity	4.9e-05	GO:0016405	CoA-ligase activity	9.9e-05	-	-	-
GO:0004620	phospholipase activity	0.00019	GO:0016878	acid-thiol ligase activity	9.9e-05	-	-	-
GO:0008479	queuine tRNA-ribosyltransferase activity	0.00024	GO:0004467	long-chain fatty acid-CoA ligase activit...	0.00017	-	-	-
GO:0016298	lipase activity	0.00035	GO:0015645	fatty acid ligase activity	0.00017	-	-	-
GO:0016757	transferase activity, transferring glyco...	0.00080	GO:0016877	ligase activity, forming carbon-sulfur b...	0.00068	-	-	-
GO:0008081	phosphoric diester hydrolase activity	0.00083	GO:0004409	homoaconitate hydratase activity	0.00103	-	-	-
GO:0016763	transferase activity, transferring pento...	0.00091	GO:0015197	peptide transporter activity	0.00103	-	-	-
GO:0000287	magnesium ion binding	0.00097	GO:0015198	oligopeptide transporter activity	0.00103	-	-	-
GO:0008424	glycoprotein 6-alpha-L-fucosyltransferas...	0.00124	GO:0015431	glutathione S-conjugate-exporting ATPase...	0.00103	-	-	-
Componentes Celulares								
GO:0005615	extracellular space	2.0e-09	GO:0046658	anchored component of plasma membrane	0.00012	-	-	-
GO:0046658	anchored component of plasma membrane	2.4e-08	GO:0031225	anchored component of membrane	0.00023	-	-	-
GO:0044421	extracellular region part	1.1e-07	GO:0005778	peroxisomal membrane	0.00051	-	-	-
GO:0031225	anchored component of membrane	1.3e-07	GO:0031903	microbody membrane	0.00051	-	-	-
GO:0043230	extracellular organelle	7.2e-07	GO:0044438	microbody part	0.00084	-	-	-
GO:0065010	extracellular membrane-bounded organelle	7.2e-07	GO:0044439	peroxisomal part	0.00084	-	-	-
GO:0044420	extracellular matrix component	8.7e-07	GO:0046581	intercellular canalculus	0.00105	-	-	-
GO:0005576	extracellular region	3.3e-06	GO:0005886	plasma membrane	0.00163	-	-	-
GO:0031012	extracellular matrix	6.1e-06	GO:0005777	peroxisome	0.00252	-	-	-
GO:0031226	intrinsic component of plasma membrane	8.5e-06	GO:0042579	microbody	0.00252	-	-	-

Tabla S18. Análisis de enriquecimiento en GO Terms para *S. haematobium*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0010880	regulation of release of sequestered cal...	2.7e-10	GO:0009988	cell-cell recognition	1.2e-05	-	-	-
GO:0010881	regulation of cardiac muscle contraction...	2.7e-10	GO:0008037	cell recognition	2.2e-05	-	-	-
GO:0014808	release of sequestered calcium ion into ...	2.7e-10	GO:0007339	binding of sperm to zona pellucida	4.3e-05	-	-	-
GO:1903514	calcium ion transport from endoplasmic r...	2.7e-10	GO:0035036	sperm-egg recognition	4.3e-05	-	-	-
GO:0070296	sarcoplasmic reticulum calcium ion trans...	5.3e-10	GO:0018395	peptidyl-lysine hydroxylation to 5-hydro...	0.00037	-	-	-
GO:0010882	regulation of cardiac muscle contraction...	9.8e-10	GO:0070077	histone arginine demethylation	0.00037	-	-	-
GO:0090314	positive regulation of protein targeting...	8.8e-08	GO:0070078	histone H3-R2 demethylation	0.00037	-	-	-
GO:0009988	cell-cell recognition	1.0e-07	GO:0070079	histone H4-R3 demethylation	0.00037	-	-	-
GO:1903522	regulation of blood circulation	1.4e-07	GO:0070782	phosphatidylserine exposure on apoptotic...	0.00037	-	-	-
GO:0043268	positive regulation of potassium ion tra...	2.8e-07	GO:0017121	phospholipid scrambling	0.00045	-	-	-
Funciones Moleculares								
GO:0072341	modified amino acid binding	4.0e-08	GO:0033746	histone demethylase activity (H3-R2 spec...	0.00011	-	-	-
GO:0001786	phosphatidylserine binding	1.3e-07	GO:0033749	histone demethylase activity (H4-R3 spec...	0.00011	-	-	-
GO:0045296	cadherin binding	1.3e-07	GO:0070815	peptidyl-lysine 5-dioxygenase activity	0.00016	-	-	-
GO:0005154	epidermal growth factor receptor binding	2.7e-07	GO:0005109	frizzled binding	0.00025	-	-	-
GO:0044325	ion channel binding	6.8e-07	GO:0017147	Wnt-protein binding	0.00037	-	-	-
GO:0016597	amino acid binding	6.0e-06	GO:0004714	transmembrane receptor protein tyrosine ...	0.00046	-	-	-
GO:0030507	spectrin binding	1.0e-05	GO:0016706	oxidoreductase activity, acting on paire...	0.00046	-	-	-
GO:0070851	growth factor receptor binding	1.0e-05	GO:0032451	demethylase activity	0.00055	-	-	-
GO:0015459	potassium channel regulator activity	1.3e-05	GO:0032452	histone demethylase activity	0.00055	-	-	-
GO:0005102	receptor binding	3.0e-05	GO:0051213	dioxygenase activity	0.00061	-	-	-
Componentes Celulares								
GO:0033268	node of Ranvier	1.8e-10	GO:0036126	sperm flagellum	2.8e-05	-	-	-
GO:0098794	postsynapse	2.7e-08	GO:0001669	acrosomal vesicle	0.00022	-	-	-
GO:0044304	main axon	2.7e-08	GO:0097223	sperm part	0.00077	-	-	-
GO:0098797	plasma membrane protein complex	4.3e-08	GO:0030134	ER to Golgi transport vesicle	0.00105	-	-	-
GO:0044425	membrane part	5.6e-08	GO:0005771	multivesicular body	0.00128	-	-	-
GO:0030425	dendrite	5.6e-08	GO:0032809	neuronal cell body membrane	0.00336	-	-	-
GO:0005886	plasma membrane	2.2e-07	GO:0044298	cell body membrane	0.00356	-	-	-
GO:0045211	postsynaptic membrane	2.5e-07	-	-	-	-	-	-
GO:0012505	endomembrane system	5.3e-07	-	-	-	-	-	-
GO:0016021	integral component of membrane	9.4e-07	-	-	-	-	-	-

Tabla S19. Análisis de enriquecimiento en GO Terms para *S. japonicum*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0006818	hydrogen transport	3.4e-25	GO:0009409	response to cold	1.6e-12	GO:0006123	mitochondrial electron transport, cytoch...	2.8e-14
GO:0015992	proton transport	3.4e-25	GO:0006123	mitochondrial electron transport, cytoch...	1.4e-11	GO:0009409	response to cold	2.7e-13
GO:0009409	response to cold	1.4e-21	GO:0007595	lactation	3.2e-10	GO:0007595	lactation	6.6e-13
GO:0006123	mitochondrial electron transport, cytoch...	4.4e-18	GO:0032941	secretion by tissue	8.6e-10	GO:0032941	secretion by tissue	1.8e-12
GO:0007595	lactation	1.9e-17	GO:0007589	body fluid secretion	2.0e-09	GO:0007589	body fluid secretion	4.3e-12
GO:0032941	secretion by tissue	2.9e-16	GO:0042775	mitochondrial ATP synthesis coupled elec...	3.4e-09	GO:0042775	mitochondrial ATP synthesis coupled elec...	7.4e-12
GO:0042773	ATP synthesis coupled electron transport	8.9e-16	GO:0042773	ATP synthesis coupled electron transport	3.7e-09	GO:0042773	ATP synthesis coupled electron transport	8.0e-12
GO:0007589	body fluid secretion	3.0e-15	GO:0006818	hydrogen transport	4.4e-09	GO:0006119	oxidative phosphorylation	2.7e-11
GO:0022904	respiratory electron transport chain	9.8e-15	GO:0015992	proton transport	4.4e-09	GO:0022904	respiratory electron transport chain	5.0e-11
GO:0042775	mitochondrial ATP synthesis coupled elec...	1.2e-14	GO:0006119	oxidative phosphorylation	1.2e-08	GO:0022900	electron transport chain	6.9e-11
Funciones Moleculares								
GO:0004129	cytochrome-c oxidase activity	1.1e-30	GO:0004129	cytochrome-c oxidase activity	6.4e-11	GO:0004129	cytochrome-c oxidase activity	2.6e-13
GO:0015002	heme-copper terminal oxidase activity	1.1e-30	GO:0015002	heme-copper terminal oxidase activity	6.4e-11	GO:0015002	heme-copper terminal oxidase activity	2.6e-13
GO:0016676	oxidoreductase activity, acting on a hem...	1.1e-30	GO:0016676	oxidoreductase activity, acting on a hem...	6.4e-11	GO:0016676	oxidoreductase activity, acting on a hem...	2.6e-13
GO:0016675	oxidoreductase activity, acting on a hem...	2.1e-30	GO:0016675	oxidoreductase activity, acting on a hem...	7.4e-11	GO:0016675	oxidoreductase activity, acting on a hem...	2.9e-13
GO:0009055	electron carrier activity	1.2e-23	GO:0009055	electron carrier activity	3.2e-09	GO:0009055	electron carrier activity	1.3e-11
GO:0015078	hydrogen ion transmembrane transporter a...	1.3e-22	GO:0015078	hydrogen ion transmembrane transporter a...	1.3e-08	GO:0015078	hydrogen ion transmembrane transporter a...	5.5e-11
GO:0016491	oxidoreductase activity	6.5e-18	GO:0015077	monovalent inorganic cation transmembran...	5.0e-06	GO:0015077	monovalent inorganic cation transmembran...	2.5e-08
GO:0015077	monovalent inorganic cation transmembran...	6.7e-13	GO:0016491	oxidoreductase activity	1.1e-05	GO:0016491	oxidoreductase activity	5.9e-08
GO:0003756	protein disulfide isomerase activity	2.6e-11	GO:0022890	inorganic cation transmembrane transport...	6.2e-05	GO:0022890	inorganic cation transmembrane transport...	3.6e-07
GO:0016862	intramolecular oxidoreductase activity, ...	2.6e-11	GO:0005215	transporter activity	0.00015	GO:0008324	cation transmembrane transporter activit...	1.3e-06
Componentes Celulares								
GO:0005751	mitochondrial respiratory chain complex ...	7.6e-21	GO:0005751	mitochondrial respiratory chain complex ...	1.8e-12	GO:0005751	mitochondrial respiratory chain complex ...	3.5e-15
GO:0045277	respiratory chain complex IV	7.6e-21	GO:0045277	respiratory chain complex IV	1.8e-12	GO:0045277	respiratory chain complex IV	3.5e-15
GO:0098803	respiratory chain complex	3.4e-20	GO:0070069	cytochrome complex	1.0e-11	GO:0070069	cytochrome complex	2.1e-14
GO:0070469	respiratory chain	1.9e-19	GO:0098803	respiratory chain complex	1.4e-09	GO:0098803	respiratory chain complex	2.9e-12
GO:0070069	cytochrome complex	8.7e-19	GO:0005746	mitochondrial respiratory chain	2.1e-09	GO:0005746	mitochondrial respiratory chain	4.6e-12
GO:0005746	mitochondrial respiratory chain	3.3e-18	GO:0070469	respiratory chain	2.3e-09	GO:0070469	respiratory chain	5.0e-12
GO:0098800	inner mitochondrial membrane protein com...	4.9e-17	GO:0098800	inner mitochondrial membrane protein com...	1.3e-08	GO:0098800	inner mitochondrial membrane protein com...	3.0e-11
GO:0098798	mitochondrial protein complex	3.4e-15	GO:0098798	mitochondrial protein complex	5.4e-08	GO:0098798	mitochondrial protein complex	1.2e-10
GO:0005743	mitochondrial inner membrane	2.6e-14	GO:0044455	mitochondrial membrane part	1.2e-07	GO:0044455	mitochondrial membrane part	2.8e-10
GO:0044455	mitochondrial membrane part	3.9e-14	GO:0005743	mitochondrial inner membrane	1.4e-06	GO:0005743	mitochondrial inner membrane	5.0e-08

Tabla S20. Análisis de enriquecimiento en GO Terms para *S. mansoni*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0034609	spicule insertion	1.1e-10	GO:0060361	flight	9.9e-06	-	-	-
GO:0060361	flight	2.3e-10	GO:0019377	glycolipid catabolic process	2.7e-05	-	-	-
GO:0031529	ruffle organization	7.9e-10	GO:0046479	glycosphingolipid catabolic process	2.7e-05	-	-	-
GO:0001993	regulation of systemic arterial blood pr...	1.1e-09	GO:0072347	response to anesthetic	9.8e-05	-	-	-
GO:0001996	positive regulation of heart rate by epi...	1.1e-09	GO:0090141	positive regulation of mitochondrial fis...	9.8e-05	-	-	-
GO:0003065	positive regulation of heart rate by epi...	1.1e-09	GO:0035070	salivary gland histolysis	0.00011	-	-	-
GO:0003321	positive regulation of blood pressure by...	1.1e-09	GO:0035071	salivary gland cell autophagic cell deat...	0.00011	-	-	-
GO:0030049	muscle filament sliding	1.7e-09	GO:0044242	cellular lipid catabolic process	0.00012	-	-	-
GO:0033275	actin-myosin filament sliding	3.9e-09	GO:0030149	sphingolipid catabolic process	0.00012	-	-	-
GO:0003062	regulation of heart rate by chemical sig...	4.0e-09	GO:0046466	membrane lipid catabolic process	0.00012	-	-	-
Funciones Moleculares								
GO:0008307	structural constituent of muscle	4.6e-09	GO:0008234	cysteine-type peptidase activity	7.7e-07	-	-	-
GO:0005515	protein binding	2.8e-07	GO:0047876	endoglycosylceramidase activity	4.0e-06	-	-	-
GO:0051015	actin filament binding	6.1e-07	GO:0070011	peptidase activity, acting on L-amino ac...	6.9e-05	-	-	-
GO:0019899	enzyme binding	9.3e-07	GO:0004197	cysteine-type endopeptidase activity	8.2e-05	-	-	-
GO:0008092	cytoskeletal protein binding	1.9e-06	GO:0016787	hydrolase activity	9.7e-05	-	-	-
GO:0003779	actin binding	2.0e-06	GO:0004553	hydrolase activity, hydrolyzing O-glycos...	0.00010	-	-	-
GO:0005200	structural constituent of cytoskeleton	9.5e-06	GO:0008233	peptidase activity	0.00012	-	-	-
GO:0004622	lysophospholipase activity	1.2e-05	GO:0016853	isomerase activity	0.00024	-	-	-
GO:0005488	binding	2.4e-05	GO:0016798	hydrolase activity, acting on glycosyl b...	0.00033	-	-	-
GO:0047876	endoglycosylceramidase activity	0.00013	GO:0016866	intramolecular transferase activity	0.00076	-	-	-
Componentes Celulares								
GO:0036379	myofilament	5.1e-12	GO:0043202	lysosomal lumen	9.1e-06	-	-	-
GO:0044422	organelle part	9.8e-11	GO:0005775	vacuolar lumen	1.1e-05	-	-	-
GO:0005862	muscle thin filament tropomyosin	5.2e-10	GO:0005764	lysosome	0.00014	-	-	-
GO:0032059	bleb	5.2e-10	GO:0000323	lytic vacuole	0.00026	-	-	-
GO:0005865	striated muscle thin filament	5.5e-09	GO:0005813	centrosome	0.00063	-	-	-
GO:0044446	intracellular organelle part	5.8e-09	GO:0005703	polytene chromosome puff	0.00066	-	-	-
GO:0005703	polytene chromosome puff	3.1e-08	GO:0042470	melanosome	0.00105	-	-	-
GO:0070865	investment cone	1.8e-07	GO:0048770	pigment granule	0.00105	-	-	-
GO:0043226	organelle	2.0e-06	GO:0005741	mitochondrial outer membrane	0.00203	-	-	-
GO:0001725	stress fiber	2.2e-06	GO:0005815	microtubule organizing center	0.00224	-	-	-

Tabla S21. Análisis de enriquecimiento en GO Terms para *S. rodhaini*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
Procesos Biológicos								
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0010877	lipid transport involved in lipid storag...	< 1e-30	GO:0010877	lipid transport involved in lipid storag...	< 1e-30	GO:0010877	lipid transport involved in lipid storag...	0.00025
GO:0034775	glutathione transmembrane transport	< 1e-30	GO:0034775	glutathione transmembrane transport	< 1e-30	GO:0034775	glutathione transmembrane transport	0.00025
GO:0035443	tripeptide transmembrane transport	< 1e-30	GO:0035443	tripeptide transmembrane transport	< 1e-30	GO:0035443	tripeptide transmembrane transport	0.00025
GO:0035672	oligopeptide transmembrane transport	< 1e-30	GO:0035672	oligopeptide transmembrane transport	< 1e-30	GO:0035672	oligopeptide transmembrane transport	0.00025
GO:0043215	daunorubicin transport	< 1e-30	GO:0043215	daunorubicin transport	< 1e-30	GO:0043215	daunorubicin transport	0.00025
GO:0046618	drug export	< 1e-30	GO:0046618	drug export	< 1e-30	GO:0046618	drug export	0.00025
GO:0070327	thyroid hormone transport	< 1e-30	GO:0070327	thyroid hormone transport	< 1e-30	GO:0070327	thyroid hormone transport	0.00025
GO:1901656	glycoside transport	< 1e-30	GO:1901656	glycoside transport	< 1e-30	GO:1901656	glycoside transport	0.00025
GO:0034635	glutathione transport	< 1e-30	GO:0034635	glutathione transport	< 1e-30	GO:0034635	glutathione transport	0.00025
GO:0042939	tripeptide transport	< 1e-30	GO:0042939	tripeptide transport	< 1e-30	GO:0042939	tripeptide transport	0.00025
Funciones Moleculares								
GO:0004409	homoaconitate hydratase activity	< 1e-30	GO:0004409	homoaconitate hydratase activity	< 1e-30	GO:0004409	homoaconitate hydratase activity	0.00039
GO:0015197	peptide transporter activity	< 1e-30	GO:0015197	peptide transporter activity	< 1e-30	GO:0015197	peptide transporter activity	0.00039
GO:0015198	oligopeptide transporter activity	< 1e-30	GO:0015198	oligopeptide transporter activity	< 1e-30	GO:0015198	oligopeptide transporter activity	0.00039
GO:0015431	glutathione S-conjugate-exporting ATPase...	< 1e-30	GO:0015431	glutathione S-conjugate-exporting ATPase...	< 1e-30	GO:0015431	glutathione S-conjugate-exporting ATPase...	0.00039
GO:0015432	bile acid-exporting ATPase activity	< 1e-30	GO:0015432	bile acid-exporting ATPase activity	< 1e-30	GO:0015432	bile acid-exporting ATPase activity	0.00039
GO:0034634	glutathione transmembrane transporter ac...	< 1e-30	GO:0034634	glutathione transmembrane transporter ac...	< 1e-30	GO:0034634	glutathione transmembrane transporter ac...	0.00039
GO:0035673	oligopeptide transmembrane transporter a...	< 1e-30	GO:0035673	oligopeptide transmembrane transporter a...	< 1e-30	GO:0035673	oligopeptide transmembrane transporter a...	0.00039
GO:0042937	tripeptide transporter activity	< 1e-30	GO:0042937	tripeptide transporter activity	< 1e-30	GO:0042937	tripeptide transporter activity	0.00039
GO:0071997	glutathione S-conjugate-transporting ATP...	< 1e-30	GO:0071997	glutathione S-conjugate-transporting ATP...	< 1e-30	GO:0071997	glutathione S-conjugate-transporting ATP...	0.00039
GO:1904680	peptide transmembrane transporter activi...	< 1e-30	GO:1904680	peptide transmembrane transporter activi...	< 1e-30	GO:1904680	peptide transmembrane transporter activi...	0.00039
Componentes Celulares								
GO:0046581	intercellular canaliculus	< 1e-30	GO:0046581	intercellular canaliculus	< 1e-30	GO:0046581	intercellular canaliculus	0.00024
GO:0016328	lateral plasma membrane	< 1e-30	GO:0016328	lateral plasma membrane	< 1e-30	GO:0016328	lateral plasma membrane	0.00033
GO:0030139	endocytic vesicle	< 1e-30	GO:0030139	endocytic vesicle	< 1e-30	GO:0030139	endocytic vesicle	0.00058
GO:0016323	basolateral plasma membrane	< 1e-30	GO:0016323	basolateral plasma membrane	< 1e-30	GO:0016323	basolateral plasma membrane	0.00064
GO:0016324	apical plasma membrane	< 1e-30	GO:0016324	apical plasma membrane	< 1e-30	GO:0016324	apical plasma membrane	0.00073
GO:0045177	apical part of cell	< 1e-30	GO:0005911	cell-cell junction	< 1e-30	GO:0005911	cell-cell junction	0.00110
GO:0005911	cell-cell junction	< 1e-30	GO:0045177	apical part of cell	< 1e-30	GO:0045177	apical part of cell	0.00112
GO:0030054	cell junction	< 1e-30	GO:0030054	cell junction	< 1e-30	GO:0030054	cell junction	0.00191
GO:0031226	intrinsic component of plasma membrane	< 1e-30	GO:0005887	integral component of plasma membrane	< 1e-30	GO:0005887	integral component of plasma membrane	0.00222
GO:0005887	integral component of plasma membrane	< 1e-30	GO:0031226	intrinsic component of plasma membrane	< 1e-30	GO:0031226	intrinsic component of plasma membrane	0.00241

Tabla S22. Análisis de enriquecimiento en GO Terms para *S. solidus*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
Procesos Biológicos								
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0036071	N-glycan fucosylation	2.1e-07	-	-	-	-	-	-
GO:0036065	fucosylation	1.4e-06	-	-	-	-	-	-
GO:0001504	neurotransmitter uptake	2.2e-06	-	-	-	-	-	-
GO:0098657	import into cell	2.2e-06	-	-	-	-	-	-
GO:0035051	cardiocyte differentiation	6.2e-06	-	-	-	-	-	-
GO:0060923	cardiac muscle cell fate commitment	1.3e-05	-	-	-	-	-	-
GO:0048871	multicellular organismal homeostasis	2.7e-05	-	-	-	-	-	-
GO:0055007	cardiac muscle cell differentiation	3.2e-05	-	-	-	-	-	-
GO:0002026	regulation of the force of heart contrac...	4.1e-05	-	-	-	-	-	-
GO:0050891	multicellular organismal water homeostas...	4.1e-05	-	-	-	-	-	-
Funciones Moleculares								
GO:0008424	glycoprotein 6-alpha-L-fucosyltransferas...	1.2e-07	-	-	-	-	-	-
GO:0046702	galactoside 6-L-fucosyltransferase activ...	1.2e-07	-	-	-	-	-	-
GO:0046921	alpha-(1->6)-fucosyltransferase activity	1.2e-07	-	-	-	-	-	-
GO:0008417	fucosyltransferase activity	6.6e-07	-	-	-	-	-	-
GO:0022804	active transmembrane transporter activit...	1.7e-05	-	-	-	-	-	-
GO:0072349	modified amino acid transmembrane transp...	4.5e-05	-	-	-	-	-	-
GO:0008504	monoamine transmembrane transporter acti...	0.00014	-	-	-	-	-	-
GO:0035639	purine ribonucleoside triphosphate bindi...	0.00026	-	-	-	-	-	-
GO:1901618	organic hydroxy compound transmembrane t...	0.00026	-	-	-	-	-	-
GO:0000166	nucleotide binding	0.00033	-	-	-	-	-	-
Componentes Celulares								
GO:0016021	integral component of membrane	1.5e-06	-	-	-	-	-	-
GO:0031224	intrinsic component of membrane	1.9e-06	-	-	-	-	-	-
GO:0031300	intrinsic component of organelle membran...	2.4e-06	-	-	-	-	-	-
GO:0030173	integral component of Golgi membrane	4.3e-06	-	-	-	-	-	-
GO:0031228	intrinsic component of Golgi membrane	5.0e-06	-	-	-	-	-	-
GO:0098533	ATPase dependent transmembrane transport...	3.5e-05	-	-	-	-	-	-
GO:1904949	ATPase complex	5.4e-05	-	-	-	-	-	-
GO:0032580	Golgi cisterna membrane	0.00013	-	-	-	-	-	-
GO:0098588	bounding membrane of organelle	0.00044	-	-	-	-	-	-
GO:0000139	Golgi membrane	0.00067	-	-	-	-	-	-

Tabla S23. Análisis de enriquecimiento en GO Terms para *S. matthei*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0007632	visual behavior	2.0e-11	GO:0007634	optokinetic behavior	0.0022	-	-	-
GO:0044704	single-organism reproductive behavior	6.6e-11	GO:0022617	extracellular matrix disassembly	0.0023	-	-	-
GO:0007634	optokinetic behavior	2.6e-10	GO:0014002	astrocyte development	0.0023	-	-	-
GO:0022617	extracellular matrix disassembly	3.2e-10	GO:0031581	hemidesmosome assembly	0.0024	-	-	-
GO:0014002	astrocyte development	3.3e-10	GO:0050908	detection of light stimulus involved in ...	0.0025	-	-	-
GO:001654	eye development	3.7e-10	GO:0050962	detection of light stimulus involved in ...	0.0025	-	-	-
GO:0031581	hemidesmosome assembly	4.2e-10	GO:0034446	substrate adhesion-dependent cell spread...	0.0025	-	-	-
GO:0050908	detection of light stimulus involved in ...	5.1e-10	GO:0048708	astrocyte differentiation	0.0026	-	-	-
GO:0050962	detection of light stimulus involved in ...	5.1e-10	GO:0009584	detection of visible light	0.0029	-	-	-
GO:0060041	retina development in camera-type eye	5.3e-10	GO:0007044	cell-substrate junction assembly	0.0029	-	-	-
Funciones Moleculares								
GO:0043208	glycosphingolipid binding	9.4e-12	GO:0043208	glycosphingolipid binding	0.0029	-	-	-
GO:0051861	glycolipid binding	1.2e-11	GO:0051861	glycolipid binding	0.0030	-	-	-
GO:0046625	sphingolipid binding	1.2e-11	GO:0046625	sphingolipid binding	0.0031	-	-	-
GO:0005201	extracellular matrix structural constitu...	3.3e-11	GO:0005201	extracellular matrix structural constitu...	0.0036	-	-	-
GO:0008289	lipid binding	1.2e-10	-	-	-	-	-	-
GO:0005198	structural molecule activity	1.2e-08	-	-	-	-	-	-
GO:0097367	carbohydrate derivative binding	1.1e-05	-	-	-	-	-	-
GO:0016597	amino acid binding	0.00017	-	-	-	-	-	-
GO:0070815	peptidyl-lysine 5-dioxygenase activity	0.00032	-	-	-	-	-	-
GO:0031406	carboxylic acid binding	0.00131	-	-	-	-	-	-
Componentes Celulares								
GO:0005606	laminin-1 complex	5.4e-12	GO:0005606	laminin-1 complex	0.0021	-	-	-
GO:0043256	laminin complex	5.4e-12	GO:0043256	laminin complex	0.0021	-	-	-
GO:0043259	laminin-10 complex	5.4e-12	GO:0043259	laminin-10 complex	0.0021	-	-	-
GO:0043260	laminin-11 complex	5.4e-12	GO:0043260	laminin-11 complex	0.0021	-	-	-
GO:0005605	basal lamina	9.2e-12	GO:0005605	basal lamina	0.0023	-	-	-
GO:0005604	basement membrane	2.6e-11	GO:0005604	basement membrane	0.0027	-	-	-
GO:0044420	extracellular matrix component	3.8e-11	GO:0044420	extracellular matrix component	0.0029	-	-	-
GO:0070062	extracellular exosome	4.1e-11	GO:0070062	extracellular exosome	0.0029	-	-	-
GO:1903561	extracellular vesicle	4.1e-11	GO:1903561	extracellular vesicle	0.0029	-	-	-
GO:0043230	extracellular organelle	4.3e-11	GO:0043230	extracellular organelle	0.0029	-	-	-

Tabla S24. Análisis de enriquecimiento en GO Terms para *S. margrebowiei*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0007634	optokinetic behavior	< 1e-30	GO:0034505	tooth mineralization	6.8e-09	GO:0034505	tooth mineralization	2.7e-11
GO:0014002	astrocyte development	< 1e-30	GO:0001958	endochondral ossification	2.8e-08	GO:0001958	endochondral ossification	1.1e-10
GO:0022617	extracellular matrix disassembly	< 1e-30	GO:0036075	replacement ossification	2.8e-08	GO:0036075	replacement ossification	1.1e-10
GO:0031581	hemidesmosome assembly	< 1e-30	GO:0042045	epithelial fluid transport	2.1e-07	GO:0042045	epithelial fluid transport	8.7e-10
GO:0048708	astrocyte differentiation	< 1e-30	GO:0070633	transepithelial transport	5.5e-07	GO:0070633	transepithelial transport	2.2e-09
GO:0050908	detection of light stimulus involved in ...	< 1e-30	GO:0035094	response to nicotine	7.2e-07	GO:0035094	response to nicotine	2.9e-09
GO:0050962	detection of light stimulus involved in ...	< 1e-30	GO:0036342	post-anal tail morphogenesis	7.2e-07	GO:0036342	post-anal tail morphogenesis	2.9e-09
GO:0034446	substrate adhesion-dependent cell spread...	< 1e-30	GO:0046677	response to antibiotic	8.2e-07	GO:0046677	response to antibiotic	3.3e-09
GO:0009582	detection of abiotic stimulus	< 1e-30	GO:0060350	endochondral bone morphogenesis	1.0e-06	GO:0060350	endochondral bone morphogenesis	4.3e-09
GO:0007044	cell-substrate junction assembly	< 1e-30	GO:0060349	bone morphogenesis	1.8e-06	GO:0060349	bone morphogenesis	7.5e-09
Funciones Moleculares								
GO:0043208	glycosphingolipid binding	< 1e-30	GO:0008270	zinc ion binding	3.2e-06	GO:0008270	zinc ion binding	5.2e-10
GO:0046625	sphingolipid binding	< 1e-30	GO:0046914	transition metal ion binding	2.0e-05	GO:0046914	transition metal ion binding	3.7e-09
GO:0051861	glycolipid binding	< 1e-30	GO:0000287	magnesium ion binding	0.00021	GO:0000287	magnesium ion binding	1.5e-06
GO:0008289	lipid binding	< 1e-30	GO:0050750	low-density lipoprotein particle recepto...	0.00024	GO:0046872	metal ion binding	3.6e-06
GO:0005201	extracellular matrix structural constitu...	< 1e-30	GO:0070325	lipoprotein particle receptor binding	0.00024	GO:0043169	cation binding	5.7e-06
GO:0005198	structural molecule activity	< 1e-30	GO:0046872	metal ion binding	0.00032	GO:0016791	phosphatase activity	1.1e-05
GO:0097367	carbohydrate derivative binding	< 1e-30	GO:0043169	cation binding	0.00054	GO:0050750	low-density lipoprotein particle recepto...	2.4e-05
GO:0005488	binding	1.1e-07	GO:0032266	phosphatidylinositol-3-phosphate binding	0.00073	GO:0070325	lipoprotein particle receptor binding	2.4e-05
GO:0008479	queuine tRNA-ribosyltransferase activity	2.0e-07	GO:0070273	phosphatidylinositol-4-phosphate binding	0.00111	GO:0042578	phosphoric ester hydrolase activity	3.1e-05
GO:0016763	transferase activity, transferring pento...	6.1e-05	GO:0016791	phosphatase activity	0.00141	GO:0032266	phosphatidylinositol-3-phosphate binding	7.4e-05
Componentes Celulares								
GO:0043230	extracellular organelle	< 1e-30	GO:0046658	anchored component of plasma membrane	1.4e-08	GO:0046658	anchored component of plasma membrane	5.5e-11
GO:0065010	extracellular membrane-bounded organelle	< 1e-30	GO:0005576	extracellular region	4.5e-08	GO:0031225	anchored component of membrane	2.3e-10
GO:0044420	extracellular matrix component	< 1e-30	GO:0031225	anchored component of membrane	5.8e-08	GO:0009897	external side of plasma membrane	3.9e-07
GO:0005606	laminin-1 complex	< 1e-30	GO:0031012	extracellular matrix	3.0e-06	GO:0031226	intrinsic component of plasma membrane	9.3e-07
GO:0043256	laminin complex	< 1e-30	GO:0043230	extracellular organelle	7.5e-06	GO:0098552	side of membrane	7.2e-06
GO:0043259	laminin-10 complex	< 1e-30	GO:0065010	extracellular membrane-bounded organelle	7.5e-06	GO:0044459	plasma membrane part	3.9e-05
GO:0043260	laminin-11 complex	< 1e-30	GO:0044420	extracellular matrix component	8.0e-06	GO:0031224	intrinsic component of membrane	4.2e-05
GO:0005605	basal lamina	< 1e-30	GO:0044421	extracellular region part	1.6e-05	GO:0030016	myofibril	8.9e-05
GO:0070062	extracellular exosome	< 1e-30	GO:0005615	extracellular space	2.4e-05	GO:0009986	cell surface	9.1e-05
GO:1903561	extracellular vesicle	< 1e-30	GO:0009897	external side of plasma membrane	8.6e-05	GO:0043292	contractile fiber	0.00014

Tabla S25. Análisis de enriquecimiento en GO Terms para *S. mediterranea*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0008380	RNA splicing	6.5e-11	GO:002244	hematopoietic progenitor cell differenti...	1.0e-08	GO:0044115	development of symbiont involved in inte...	1.2e-08
GO:0009615	response to virus	3.2e-10	GO:0050886	endocrine process	1.6e-07	GO:0075015	formation of infection structure on or n...	1.2e-08
GO:0003777	RNA splicing, via transesterification re...	9.2e-10	GO:0042789	mRNA transcription from RNA polymerase I...	1.8e-07	GO:0044111	development involved in symbiotic intera...	1.6e-08
GO:000398	mRNA splicing, via spliceosome	9.2e-10	GO:0009299	mRNA transcription	9.2e-07	GO:0044403	symbiosis, encompassing mutualism throug...	3.0e-05
GO:000375	RNA splicing, via transesterification re...	1.1e-09	GO:0032876	negative regulation of DNA endoreplica...	9.5e-07	GO:0044419	interspecies interaction between organis...	3.1e-05
GO:0048024	regulation of mRNA splicing, via spliceo...	2.5e-09	GO:0032609	interferon-gamma production	1.3e-06	-	-	-
GO:0043900	regulation of multi-organism process	1.4e-08	GO:0040036	regulation of fibroblast growth factor r...	1.4e-06	-	-	-
GO:0006631	fatty acid metabolic process	3.4e-08	GO:0002076	osteoblast development	1.6e-06	-	-	-
GO:0044711	single-organism biosynthetic process	3.4e-08	GO:0032875	regulation of DNA endoreplication	1.6e-06	-	-	-
GO:0043484	regulation of RNA splicing	3.8e-08	GO:0006672	ceramide metabolic process	1.6e-06	-	-	-
Funciones Moleculares								
GO:0003723	RNA binding	4.7e-06	GO:0004175	endopeptidase activity	2.3e-11	GO:0048531	beta-1,3-galactosyltransferase activity	1.4e-08
GO:0038024	cargo receptor activity	1.8e-05	GO:0038024	cargo receptor activity	5.1e-09	GO:0035250	UDP-galactosyltransferase activity	2.2e-08
GO:0003729	mRNA binding	2.1e-05	GO:0070011	peptidase activity, acting on L-amino ac...	5.3e-09	GO:0008378	galactosyltransferase activity	4.8e-08
GO:0044822	poly(A) RNA binding	2.1e-05	GO:0008233	peptidase activity	1.0e-08	GO:0008376	acetylgalactosaminyltransferase activity	1.3e-07
GO:0052689	carboxylic ester hydrolase activity	4.2e-05	GO:0008237	metallopeptidase activity	1.1e-08	GO:0008194	UDP-glycosyltransferase activity	4.3e-06
GO:0042625	ATPase coupled ion transmembrane transpo...	7.7e-05	GO:0052689	carboxylic ester hydrolase activity	2.9e-07	GO:0016758	transferase activity, transferring hexos...	1.9e-05
GO:0019829	cation-transporting ATPase activity	8.5e-05	GO:0004222	metalloendopeptidase activity	3.3e-07	GO:0016757	transferase activity, transferring glyco...	3.1e-05
GO:0015662	ATPase activity, coupled to transmembran...	0.00011	GO:0016298	lipase activity	1.1e-06	-	-	-
GO:0001664	G-protein coupled receptor binding	0.00012	GO:0016787	hydrolase activity	5.3e-06	-	-	-
GO:0022853	active ion transmembrane transporter act...	0.00014	GO:0030228	lipoprotein particle receptor activity	1.9e-05	-	-	-
Componentes Celulares								
GO:0071011	pre-catalytic spliceosome	2.0e-09	GO:0005576	extracellular region	1.9e-26	GO:0000139	Golgi membrane	0.00071
GO:0005681	spliceosomal complex	1.1e-08	GO:0044421	extracellular region part	3.8e-10	GO:0044431	Golgi apparatus part	0.00169
GO:0005774	vacuolar membrane	2.1e-07	GO:0005764	lysosome	4.8e-08	-	-	-
GO:0016939	kinesin II complex	3.8e-07	GO:0000323	lytic vacuole	4.8e-08	-	-	-
GO:0030993	axonemal heterotrimeric kinesin-II compl...	3.8e-07	GO:0005773	vacuole	9.4e-08	-	-	-
GO:0098805	whole membrane	4.1e-07	GO:0005615	extracellular space	2.0e-07	-	-	-
GO:0071013	catalytic step 2 spliceosome	4.7e-07	GO:0005578	proteinaceous extracellular matrix	4.0e-06	-	-	-
GO:0005773	vacuole	7.2e-07	GO:0016528	sarcoplasm	9.5e-06	-	-	-
GO:0044377	vacuolar part	3.0e-06	GO:0031012	extracellular matrix	3.6e-05	-	-	-
GO:000323	lytic vacuole	6.9e-06	GO:0016529	sarcoplasmic reticulum	0.00013	-	-	-

Tabla S26. Análisis de enriquecimiento en GO Terms para *S. erinacei*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0006486	protein glycosylation	4.1e-07	GO:0035072	ecdysone-mediated induction of salivary ...	2.1e-06	GO:0003096	renal sodium ion transport	3.0e-06
GO:0043413	macromolecule glycosylation	4.1e-07	GO:0035078	induction of programmed cell death by ec...	2.1e-06	GO:0071233	cellular response to leucine	1.9e-05
GO:0016266	O-glycan processing	6.7e-07	GO:0035081	induction of programmed cell death by ho...	6.2e-06	GO:0043201	response to leucine	3.7e-05
GO:0070085	glycosylation	1.3e-06	GO:0001945	lymph vessel development	6.7e-06	GO:0007141	male meiosis I	6.8e-05
GO:0048259	regulation of receptor-mediated endocyto...	3.3e-06	GO:0001946	lymphangiogenesis	6.7e-06	GO:0032007	negative regulation of TOR signaling	6.8e-05
GO:0042403	thyroid hormone metabolic process	6.5e-06	GO:0036303	lymph vessel morphogenesis	6.7e-06	GO:0033522	histone H2A ubiquitination	8.3e-05
GO:002260	lymphocyte homeostasis	1.1e-05	GO:0035071	salivary gland cell autophagic cell deat...	7.1e-06	GO:0071230	cellular response to amino acid stimulus	0.00018
GO:0009101	glycoprotein biosynthetic process	1.2e-05	GO:0035070	salivary gland histolysis	7.9e-06	GO:0001553	lutinization	0.00025
GO:0009100	glycoprotein metabolic process	1.7e-05	GO:0048102	autophagic cell death	9.7e-06	GO:0016574	histone ubiquitination	0.00025
GO:0097062	dendritic spine maintenance	2.2e-05	GO:0035924	cellular response to vascular endothelia...	2.2e-05	GO:0031293	membrane protein intracellular domain pr...	0.00038
Funciones Moleculares								
GO:0018392	glycoprotein 3-alpha-L-fucosyltransferas...	1.6e-11	GO:0005451	monovalent cation:proton antiporter acti...	6.3e-06	GO:0005451	monovalent cation:proton antiporter acti...	5.8e-07
GO:0046920	alpha-(1->3)-fucosyltransferase activity	1.6e-11	GO:0022821	potassium ion antiporter activity	1.5e-05	GO:0022821	potassium ion antiporter activity	9.6e-07
GO:0008508	bile acid:sodium symporter activity	1.2e-06	GO:0005201	extracellular matrix structural constitu...	3.0e-05	GO:0015299	solute:proton antiporter activity	1.8e-06
GO:0015370	solute:sodium symporter activity	6.3e-05	GO:0015299	solute:proton antiporter activity	4.4e-05	GO:0031624	ubiquitin conjugating enzyme binding	9.5e-06
GO:0003829	beta-1,3-galactosyl-O-glycosyl-glycoprot...	0.00011	GO:0031624	ubiquitin conjugating enzyme binding	0.00067	GO:0044390	ubiquitin-like protein conjugating enzym...	9.5e-06
GO:0008109	N-acetyllactosaminide beta-1,6-N-acetylgl...	0.00011	GO:0044390	ubiquitin-like protein conjugating enzym...	0.00067	GO:0015491	cation:cation antiporter activity	1.4e-05
GO:0047225	acetylgalactosaminyl-O-glycosyl-glycopro...	0.00011	GO:0019203	carbohydrate phosphatase activity	0.00069	GO:0015298	solute:cation antiporter activity	1.7e-05
GO:0017171	serine hydrolase activity	0.00011	GO:0050308	sugar-phosphatase activity	0.00069	GO:0070728	leucine binding	1.7e-05
GO:0008375	acetylglucosaminyltransferase activity	0.00011	GO:0070567	cytidyltransferase activity	0.00087	GO:0099516	ion antiporter activity	5.3e-05
GO:0008236	serine-type peptidase activity	0.00012	GO:0015491	cation:cation antiporter activity	0.00123	GO:0015297	antiporter activity	8.2e-05
Componentes Celulares								
GO:0001138	Golgi trans cisterna	3.2e-08	GO:0070578	RISC-loading complex	0.0010	GO:0005903	brush border	0.00083
GO:0005797	Golgi medial cisterna	5.0e-08	GO:0005751	mitochondrial respiratory chain complex ...	0.0015	GO:0000502	proteasome complex	0.00108
GO:0000139	Golgi membrane	0.00011	GO:0045277	respiratory chain complex IV	0.0015	GO:1905369	endopeptidase complex	0.00108
GO:0005581	collagen trimer	0.00025	GO:0044798	nuclear transcription factor complex	0.0021	GO:0031011	INO80 complex	0.00170
GO:0070069	cytochrome complex	0.00040	GO:0016234	inclusion body	0.0022	GO:0097346	INO80-type complex	0.00170
GO:0005795	Golgi stack	0.00065	GO:0000502	proteasome complex	0.0023	GO:0098862	cluster of actin-based cell projections	0.00196
GO:0005751	mitochondrial respiratory chain complex ...	0.00091	GO:1905369	endopeptidase complex	0.0023	GO:1905368	peptidase complex	0.00204
GO:0045277	respiratory chain complex IV	0.00091	GO:1902555	endoribonuclease complex	0.0025	GO:0033202	DNA helicase complex	0.00292
GO:0032580	Golgi cisterna membrane	0.00101	GO:0009575	RNA polymerase II transcription factor c...	0.0046	-	-	-
GO:0005921	gap junction	0.00119	-	-	-	-	-	-

Tabla S27. Análisis de enriquecimiento en GO Terms para *T. asiatica*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0006865	amino acid transport	1.0e-07	GO:0000959	mitochondrial RNA metabolic process	7.7e-09	-	-	-
GO:0006835	dicarboxylic acid transport	2.0e-06	GO:0006352	DNA-templated transcription, initiation	2.1e-05	-	-	-
GO:0070670	response to interleukin-4	2.6e-06	GO:0002673	regulation of acute inflammatory respons...	6.2e-05	-	-	-
GO:0015849	organic acid transport	3.2e-06	GO:0007597	blood coagulation, intrinsic pathway	6.2e-05	-	-	-
GO:0071705	nitrogen compound transport	1.8e-05	GO:0072376	protein activation cascade	6.2e-05	-	-	-
GO:0019388	galactose catabolic process	3.8e-05	GO:0072378	blood coagulation, fibrin clot formation	6.2e-05	-	-	-
GO:0040032	post-embryonic body morphogenesis	3.8e-05	GO:0051457	maintenance of protein location in nucle...	8.5e-05	-	-	-
GO:0048148	behavioral response to cocaine	3.8e-05	GO:0070613	regulation of protein processing	0.00011	-	-	-
GO:0006012	galactose metabolic process	6.6e-05	GO:1903317	regulation of protein maturation	0.00011	-	-	-
GO:0032964	collagen biosynthetic process	8.8e-05	GO:0002921	negative regulation of humoral immune re...	0.00028	-	-	-
Funciones Moleculares								
GO:0000099	sulfur amino acid transmembrane transpor...	9.3e-06	GO:0003899	DNA-directed RNA polymerase activity	2.1e-08	-	-	-
GO:0015179	L-amino acid transmembrane transporter a...	3.6e-05	GO:0034062	RNA polymerase activity	2.5e-08	-	-	-
GO:1901682	sulfur compound transmembrane transporte...	5.0e-05	GO:0016779	nucleotidyltransferase activity	4.3e-06	-	-	-
GO:0005246	calcium channel regulator activity	7.4e-05	GO:0004867	serine-type endopeptidase inhibitor acti...	0.00027	-	-	-
GO:0005313	L-glutamate transmembrane transporter ac...	0.00012	GO:0019955	cytokine binding	0.00038	-	-	-
GO:0004364	glutathione transferase activity	0.00034	GO:0016772	transferase activity, transferring phosph...	0.00124	-	-	-
GO:0004402	histone acetyltransferase activity	0.00040	GO:0004536	deoxyribonuclease activity	0.00131	-	-	-
GO:0061733	peptide-lysine-N-acetyltransferase activ...	0.00040	GO:0004866	endopeptidase inhibitor activity	0.00257	-	-	-
GO:0072349	modified amino acid transmembrane transp...	0.00054	GO:0030414	peptidase inhibitor activity	0.00280	-	-	-
GO:0004053	arginase activity	0.00056	GO:0008239	dipeptidyl-peptidase activity	0.00285	-	-	-
Componentes Celulares								
GO:0005615	extracellular space	6.3e-06	GO:0009295	nucleoid	1.1e-07	-	-	-
GO:0009295	nucleoid	2.4e-05	GO:0042645	mitochondrial nucleoid	1.1e-07	-	-	-
GO:0042645	mitochondrial nucleoid	2.4e-05	GO:0005576	extracellular region	1.6e-05	-	-	-
GO:0030014	CCR4-NOT complex	2.4e-05	GO:0005615	extracellular space	3.7e-05	-	-	-
GO:0031224	intrinsic component of membrane	4.4e-05	GO:0044421	extracellular region part	0.00012	-	-	-
GO:0016021	integral component of membrane	8.8e-05	GO:0031093	platelet alpha granule lumen	0.00014	-	-	-
GO:0000138	Golgi trans cisterna	0.00011	GO:0034774	secretory granule lumen	0.00014	-	-	-
GO:0005797	Golgi medial cisterna	0.00011	GO:0000784	nuclear chromosome, telomeric region	0.00017	-	-	-
GO:0044421	extracellular region part	0.00014	GO:0060205	cytoplasmic membrane-bounded vesicle lum...	0.00021	-	-	-
GO:0005801	cis-Golgi network	0.00015	GO:0030141	secretory granule	0.00022	-	-	-

Tabla S28. Análisis de enriquecimiento en GO Terms para *T. solium*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0045676	regulation of R7 cell differentiation	1.4e-08	GO:2001244	positive regulation of intrinsic apoptot...	1.0e-11	GO:2001244	positive regulation of intrinsic apoptot...	2.8e-07
GO:2001244	positive regulation of intrinsic apoptot...	3.9e-08	GO:2001233	regulation of apoptotic signaling pathwa...	4.0e-11	GO:0031648	protein destabilization	7.8e-07
GO:0042787	protein ubiquitination involved in ubiq...	5.0e-08	GO:0031648	protein destabilization	6.5e-11	GO:0045676	regulation of R7 cell differentiation	8.8e-07
GO:0097193	intrinsic apoptotic signaling pathway	9.5e-08	GO:0045676	regulation of R7 cell differentiation	8.0e-11	GO:2001242	regulation of intrinsic apoptotic signal...	1.4e-06
GO:0007465	R7 cell fate commitment	1.1e-07	GO:2001242	regulation of intrinsic apoptotic signal...	1.7e-10	GO:0007465	R7 cell fate commitment	2.0e-06
GO:0046532	regulation of photoreceptor cell differe...	1.6e-07	GO:0007465	R7 cell fate commitment	3.4e-10	GO:0046532	regulation of photoreceptor cell differe...	2.4e-06
GO:0097190	apoptotic signaling pathway	2.5e-07	GO:0046532	regulation of photoreceptor cell differe...	4.6e-10	GO:2001235	positive regulation of apoptotic signali...	2.6e-06
GO:0007141	male meiosis I	3.3e-07	GO:2001235	positive regulation of apoptotic signali...	5.3e-10	GO:0007141	male meiosis I	1.0e-05
GO:0031648	protein destabilization	3.5e-07	GO:0007141	male meiosis I	5.5e-09	GO:0001752	compound eye photoreceptor fate commitme...	1.1e-05
GO:0006508	proteolysis	4.0e-07	GO:0001752	compound eye photoreceptor fate commitme...	6.0e-09	GO:0042706	eye photoreceptor cell fate commitment	1.2e-05
Funciones Moleculares								
GO:0043621	protein self-association	3.8e-09	GO:0004842	ubiquitin-protein transferase activity	6.5e-09	GO:0043621	protein self-association	3.9e-05
GO:0016879	ligase activity, forming carbon-nitrogen...	7.0e-06	GO:0019787	ubiquitin-like protein transferase activ...	9.6e-09	GO:0004842	ubiquitin-protein transferase activity	0.00016
GO:0004872	receptor activity	1.5e-05	GO:0016881	acid-amino acid ligase activity	2.0e-08	GO:0008270	zinc ion binding	0.00017
GO:0004842	ubiquitin-protein transferase activity	3.2e-05	GO:0016879	ligase activity, forming carbon-nitrogen...	3.2e-08	GO:0019787	ubiquitin-like protein transferase activ...	0.00019
GO:0016874	ligase activity	3.3e-05	GO:0043621	protein self-association	5.0e-08	GO:0016881	acid-amino acid ligase activity	0.00027
GO:0004888	transmembrane signaling receptor activit...	4.1e-05	GO:0016874	ligase activity	2.4e-07	GO:0016879	ligase activity, forming carbon-nitrogen...	0.00034
GO:0019787	ubiquitin-like protein transferase activ...	5.4e-05	GO:0008270	zinc ion binding	5.9e-07	GO:0046914	transition metal ion binding	0.00047
GO:0046914	transition metal ion binding	0.00011	GO:0046914	transition metal ion binding	3.3e-06	GO:0016874	ligase activity	0.00092
GO:0016881	acid-amino acid ligase activity	0.00013	GO:0046872	metal ion binding	0.00013	GO:0046872	metal ion binding	0.00414
GO:0038023	signaling receptor activity	0.00013	GO:0043169	cation binding	0.00019	-	-	-
Componentes Celulares								
GO:0030877	beta-catenin destruction complex	6.3e-08	GO:0030877	beta-catenin destruction complex	6.8e-12	GO:0030877	beta-catenin destruction complex	1.5e-07
GO:0000151	ubiquitin ligase complex	2.4e-06	GO:0005769	early endosome	3.1e-10	GO:0000151	ubiquitin ligase complex	2.4e-05
GO:0005769	early endosome	0.00023	GO:0000151	ubiquitin ligase complex	4.0e-08	GO:0005769	early endosome	2.8e-05
GO:0005640	nuclear outer membrane	0.00023	GO:0005768	endosome	1.2e-07	GO:0005768	endosome	0.00052
GO:0014802	terminal cisterna	0.00080	GO:0005773	vacuole	1.3e-06	GO:1990234	transferase complex	0.00109
GO:0001741	XY body	0.00089	GO:1990234	transferase complex	2.3e-05	GO:0005773	vacuole	0.00171
GO:0014701	junctional sarcoplasmic reticulum membra...	0.00108	GO:1902494	catalytic complex	0.00033	-	-	-
GO:0030314	junctional membrane complex	0.00108	GO:0005829	cytosol	0.00122	-	-	-
GO:0098827	endoplasmic reticulum subcompartment	0.00118	GO:0012505	endomembrane system	0.00140	-	-	-
GO:0030317	sarcoplasmic reticulum membrane	0.00182	-	-	-	-	-	-

Tabla S29. Análisis de enriquecimiento en GO Terms para *T. saginata*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0051054	positive regulation of DNA metabolic pro...	1.1e-11	GO:0070613	regulation of protein processing	0.00053	-	-	-
GO:0000080	mitotic G1 phase	1.1e-11	GO:1903317	regulation of protein maturation	0.00053	-	-	-
GO:0051318	G1 phase	1.1e-11	GO:0007319	negative regulation of oskar mRNA transl...	0.00106	-	-	-
GO:0010389	regulation of G2/M transition of mitotic...	7.6e-11	GO:0016485	protein processing	0.00112	-	-	-
GO:1902749	regulation of cell cycle G2/M phase tran...	7.6e-11	GO:0051604	protein maturation	0.00121	-	-	-
GO:0006323	DNA packaging	7.7e-10	GO:0002526	acute inflammatory response	0.00123	-	-	-
GO:0030261	chromosome condensation	1.3e-09	GO:0002921	negative regulation of humoral immune re...	0.00123	-	-	-
GO:0031146	SCF-dependent proteasomal ubiquitin-depe...	1.4e-09	GO:0046011	regulation of oskar mRNA translation	0.00123	-	-	-
GO:0035518	histone H2A monoubiquitination	1.4e-09	GO:0045824	negative regulation of innate immune res...	0.00140	-	-	-
GO:0007095	mitotic G2 DNA damage checkpoint	1.7e-09	GO:0002252	immune effector process	0.00166	-	-	-
Funciones Moleculares								
GO:0004674	protein serine/threonine kinase activity	3.6e-06	GO:0008239	dipeptidyl-peptidase activity	5e-06	-	-	-
GO:0005524	ATP binding	5.3e-05	GO:0008236	serine-type peptidase activity	0.00026	-	-	-
GO:0004672	protein kinase activity	8.2e-05	GO:0017171	serine hydrolase activity	0.00028	-	-	-
GO:0032559	adenyl ribonucleotide binding	8.7e-05	GO:0004177	aminopeptidase activity	0.00075	-	-	-
GO:0030554	adenyl nucleotide binding	9.7e-05	GO:0004867	serine-type endopeptidase inhibitor acti...	0.00099	-	-	-
GO:0008375	acetylglucosaminyltransferase activity	0.00014	GO:0008238	exopeptidase activity	0.00110	-	-	-
GO:0016773	phosphotransferase activity, alcohol gro...	0.00019	GO:0019955	cytokine binding	0.00151	-	-	-
GO:0008353	RNA polymerase II carboxy-terminal domai...	0.00024	GO:0004536	deoxyribonuclease activity	0.00432	-	-	-
GO:0097159	organic cyclic compound binding	0.00028	-	-	-	-	-	-
GO:0035173	histone kinase activity	0.00055	-	-	-	-	-	-
Componentes Celulares								
GO:0019005	SCF ubiquitin ligase complex	4.3e-09	GO:0031093	platelet alpha granule lumen	0.00054	-	-	-
GO:0005875	microtubule associated complex	7.9e-07	GO:0034774	secretory granule lumen	0.00054	-	-	-
GO:0031461	cullin-RING ubiquitin ligase complex	5.4e-06	GO:0060205	cytoplasmic membrane-bounded vesicle lum...	0.00071	-	-	-
GO:0001673	male germ cell nucleus	1.6e-05	GO:0031091	platelet alpha granule	0.00090	-	-	-
GO:0043073	germ cell nucleus	3.4e-05	GO:0031983	vesicle lumen	0.00123	-	-	-
GO:0005819	spindle	7.7e-05	GO:0005576	extracellular region	0.00418	-	-	-
GO:0000786	nucleosome	0.00015	-	-	-	-	-	-
GO:0044815	DNA packaging complex	0.00028	-	-	-	-	-	-
GO:0005576	extracellular region	0.00032	-	-	-	-	-	-
GO:0000151	ubiquitin ligase complex	0.00041	-	-	-	-	-	-

Tabla S30. Análisis de enriquecimiento en GO Terms para *T. regeneti*

Inparálogos			Inparálogos secretados			Inparálogos secretados con evolucionando bajo selección positiva		
GO ID	Termino	p-value	GO ID	Termino	p-value	GO ID	Termino	p-value
GO:0019377	glycolipid catabolic process	9.6e-18	GO:0001913	T cell mediated cytotoxicity	0.00020	-	-	-
GO:0046479	glycosphingolipid catabolic process	9.6e-18	GO:0045676	regulation of R7 cell differentiation	0.00020	-	-	-
GO:0001573	ganglioside metabolic process	3.1e-16	GO:0070734	histone H3-K27 methylation	0.00023	-	-	-
GO:0006689	ganglioside catabolic process	3.1e-16	GO:0045500	sevenless signaling pathway	0.00045	-	-	-
GO:0030207	chondroitin sulfate catabolic process	3.1e-16	GO:0007465	R7 cell fate commitment	0.00097	-	-	-
GO:0030214	hyaluronan catabolic process	3.1e-16	GO:0006479	protein methylation	0.00101	-	-	-
GO:0030149	sphingolipid catabolic process	3.5e-16	GO:0008213	protein alkylation	0.00101	-	-	-
GO:0046466	membrane lipid catabolic process	3.5e-16	GO:0010454	negative regulation of cell fate commitm...	0.00101	-	-	-
GO:0007341	penetration of zona pellucida	4.5e-16	GO:0001909	leukocyte mediated cytotoxicity	0.00143	-	-	-
GO:0009313	oligosaccharide catabolic process	6.4e-16	GO:0046532	regulation of photoreceptor cell differe...	0.00143	-	-	-
Funciones Moleculares								
GO:0016231	beta-N-acetylglucosaminidase activity	2.8e-16	GO:0008234	cysteine-type peptidase activity	0.00062	-	-	-
GO:0004563	beta-N-acetylhexosaminidase activity	6.0e-16	GO:0042054	histone methyltransferase activity	0.00276	-	-	-
GO:0004553	hydrolase activity, hydrolyzing O-glycos...	7.4e-15	GO:0004252	serine-type endopeptidase activity	0.00293	-	-	-
GO:0015929	hexosaminidase activity	8.1e-15	GO:0003824	catalytic activity	0.00338	-	-	-
GO:0016798	hydrolase activity, acting on glycosyl b...	4.8e-14	-	-	-	-	-	-
GO:0030246	carbohydrate binding	5.3e-09	-	-	-	-	-	-
GO:0046982	protein heterodimerization activity	8.0e-09	-	-	-	-	-	-
GO:0046983	protein dimerization activity	3.1e-06	-	-	-	-	-	-
GO:0042803	protein homodimerization activity	2.2e-05	-	-	-	-	-	-
GO:0019855	calcium channel inhibitor activity	2.7e-05	-	-	-	-	-	-
Componentes Celulares								
GO:0043202	lysosomal lumen	4.3e-12	GO:0035098	ESC/E(Z) complex	0.00038	-	-	-
GO:0005775	vacuolar lumen	5.8e-12	GO:0005764	lysosome	0.00136	-	-	-
GO:0000323	lytic vacuole	1.4e-11	GO:0031519	PcG protein complex	0.00137	-	-	-
GO:0005764	lysosome	2.4e-11	GO:0000786	nucleosome	0.00166	-	-	-
GO:0097223	sperm part	2.6e-11	GO:0000323	lytic vacuole	0.00185	-	-	-
GO:0001669	acrosomal vesicle	2.9e-11	GO:0045120	pronucleus	0.00258	-	-	-
GO:0005816	spindle pole body	5.2e-07	GO:0044815	DNA packaging complex	0.00450	-	-	-
GO:0019908	nuclear cyclin-dependent protein kinase ...	8.3e-06	-	-	-	-	-	-
GO:0030141	secretory granule	1.4e-05	-	-	-	-	-	-
GO:0035098	ESC/E(Z) complex	2.7e-05	-	-	-	-	-	-

Detección de genes inparálogos evolucionando en régimen de selección natural positiva

A continuación se muestran aquellas familias multigénicas donde se detectó selección natural positiva. Agrupadas por especie, se muestran a su vez los genes tentativos a los cuales se asocia cada miembro de dichas familias multigénicas (tomando como referencia aquella secuencia con la cual se obtuvo un menor *e-value* al momento de realizarse una búsqueda con DIAMOND blastp contra la base de datos RefSeq del NCBI).

<i>Tabla S31. Secuencias homólogas a inparálogos evolucionando en régimen de selección natural positiva</i>	
Anotación del gen	Núm. de genes
<i>D. latum</i>	
AFX73008.1 nucleotide pyrophosphatase/phosphodiesterase 5 Spirometra erinaceieuropaei [Spirometra erinaceieuropaei],	2
<i>E. canadensis</i>	
EUB62537.1 Lysosomal acid phosphatase [Echinococcus granulosus],	2
CDS36796.2 PKD [Echinococcus multilocularis],	10
EUB54891.1 Multidrug and toxin extrusion protein [Echinococcus granulosus],	8
CDS43304.1 ras GTP ase activating protein with iq motif [Echinococcus multilocularis],	11
<i>E. granulosus</i>	
<i>E. multilocularis</i>	
CUT98835.1 tyrosine protein kinase csk [Echinococcus multilocularis], CUT98836.1 tyrosine protein kinase csk [Echinococcus multilocularis],	2
CDI96652.2 diagnostic antigen gp50 [Echinococcus multilocularis], CDS35444.2 diagnostic antigen gp50 [Echinococcus multilocularis], CDS35474.1 diagnostic antigen gp50 [Echinococcus multilocularis], CDS35485.2 diagnostic antigen gp50 [Echinococcus multilocularis],	6
CDI96674.1 cytoplasmic antigen 1 [Echinococcus multilocularis], CDI96789.1 cytoplasmic antigen 1 [Echinococcus multilocularis], CDS35450.1 cytoplasmic antigen 1 [Echinococcus multilocularis], CDS36116.1 cytoplasmic antigen 1 [Echinococcus multilocularis], CDS38255.1 cytoplasmic antigen 1 [Echinococcus multilocularis],	9
CDI96766.2 non capsid protein NS 1 [Echinococcus multilocularis], CDI96769.1 non capsid protein NS 1 [Echinococcus multilocularis], CDS36689.1 non capsid protein NS 1 [Echinococcus multilocularis], CDS36692.1 non capsid protein NS 1 [Echinococcus multilocularis], CDS36816.1 non capsid protein NS 1 [Echinococcus multilocularis],	10
CDI96777.1 diagnostic antigen gp50 [Echinococcus multilocularis], CDS37836.1 diagnostic antigen gp50 [Echinococcus multilocularis],	2
CDI96790.2 rRNA promoter binding protein [Echinococcus multilocularis], CDS36813.1 rRNA promoter binding protein [Echinococcus multilocularis], CUT98442.1 rRNA promoter binding protein [Echinococcus multilocularis], CUT98473.1 rRNA promoter binding protein [Echinococcus multilocularis],	4
CDS36561.1 deoxyribonuclease 1 2 [Echinococcus multilocularis], CDS36839.1 deoxyribonuclease [Echinococcus multilocularis], CUT98673.1 deoxyribonuclease [Echinococcus multilocularis], CUT99718.1 deoxyribonuclease [Echinococcus multilocularis],	8
CDI97116.1 target of myb1 related tom1 [Echinococcus multilocularis], CDI97118.1 tom1 protein 2 [Echinococcus multilocularis],	2
CDS35425.1 homeobox protein [Echinococcus multilocularis], CDS35427.1 homeobox protein [Echinococcus multilocularis], CDS38246.1 homeobox protein [Echinococcus multilocularis],	4
CDS35670.1 diagnostic antigen gp50 [Echinococcus multilocularis], CDS36232.1 diagnostic antigen gp50 [Echinococcus multilocularis], CDS37866.1 diagnostic antigen gp50 [Echinococcus multilocularis],	3
CDS41803.1 fibronectin type III domain-containing protein [Echinococcus multilocularis], CDS41808.1 fibronectin type III domain-containing protein [Echinococcus multilocularis],	4
CDI96497.1 lactate dehydrogenase protein [Echinococcus multilocularis], CUT99509.1 lactate dehydrogenase a [Echinococcus multilocularis],	5
CDS41461.1 eukaryotic translation initiation factor 2c [Echinococcus multilocularis], CDS41463.2 eukaryotic translation initiation factor 2c [Echinococcus multilocularis],	2
CDS36262.1 Tetraspanin-5 [Echinococcus multilocularis], CDS43044.1 tetraspanin [Echinococcus multilocularis], CDS43046.2 tetraspanin [Echinococcus multilocularis],	6

CUT98688.1 Genomic scaffold, pathogen EMU scaffold 006738 [Echinococcus multilocularis],	4
CUT98692.1 Genomic scaffold, pathogen EMU scaffold 006738 [Echinococcus multilocularis],	
CUT98693.1 Genomic scaffold, pathogen EMU scaffold 006738 [Echinococcus multilocularis],	
CUT98476.1 senescence associated protein [Echinococcus multilocularis], CUT99816.1 senescence associated protein [Echinococcus multilocularis],	2
<i>H. taeniformis</i>	
<i>H. diminuta</i>	
CDS29449.1 ATP dependent RNA helicase Ddx1 [Hymenolepis microstoma],	2
<i>H. microstoma</i>	
CDS28967.1 diagnostic antigen gp50 [Hymenolepis microstoma], CDS33088.1 diagnostic antigen gp50 [Hymenolepis microstoma],	2
CDS28993.1 transposase [Hymenolepis microstoma], CDS31077.1 Mariner Mos1 transposase [Hymenolepis microstoma],	4
CDS31987.1 transposase [Hymenolepis microstoma], CDS33064.1 transposase [Hymenolepis microstoma],	2
CDS29319.1 gag pol polyprotein [Hymenolepis microstoma], CDS33976.1 gag pol polyprotein [Hymenolepis microstoma],	2
CDS25455.1 histone lysine N methyltransferase SETMAR [Hymenolepis microstoma], CDS26545.1 Mariner Mos1 transposase [Hymenolepis microstoma],	2
CDS29686.1 coiled coil domain containing protein 162 [Hymenolepis microstoma], CDS29687.2 coiled coil domain containing protein 162 [Hymenolepis microstoma],	2
CDS26281.1 gag pol polyprotein [Hymenolepis microstoma], CDS30805.1 gag pol polyprotein [Hymenolepis microstoma],	4
CDS33898.1 contactin,neuroglian,septate junction protein [Hymenolepis microstoma], CDS35273.1 contactin [Hymenolepis microstoma],	4
CDS28404.1 gag pol polyprotein [Hymenolepis microstoma], CDS28982.2 gag pol polyprotein [Hymenolepis microstoma], CDS29500.2 gag pol polyprotein [Hymenolepis microstoma],	3
CDS29388.1 proprotein convertase subtilisin:kexin type 5 [Hymenolepis microstoma], CUU98362.1 proprotein convertase subtilisin:kexin type 5 [Hymenolepis microstoma],	2
CDS27007.1 signal recognition particle receptor subunit [Hymenolepis microstoma], CUU00465.1 signal recognition particle receptor subunit [Hymenolepis microstoma],	2
CDS31297.2 lipopolysaccharide choline [Hymenolepis microstoma], CDS35025.1 lipopolysaccharide choline [Hymenolepis microstoma],	2
<i>H. nana</i>	
CDS26143.1 beta 13 n galactosyltransferase [Hymenolepis microstoma],	6
CDS30077.2 tyrosine protein phosphatase non receptor type [Hymenolepis microstoma],	2
CDS33104.1 reverse transcriptase [Hymenolepis microstoma],	3
CDS26967.1 aldo keto reductase family 1 member B4 [Hymenolepis microstoma], CDS26970.2 aldo keto reductase family 1 member B4 [Hymenolepis microstoma],	3
CDS33626.1 solute carrier family 5 [Hymenolepis microstoma],	2
CDS34077.1 Transcription initiation factor TFIID subunit [Hymenolepis microstoma],	3
CDS26502.2 sodium and chloride dependent glycine [Hymenolepis microstoma],	2
CDS29949.1 solute carrier family 33 [Hymenolepis microstoma],	4
CDS29684.2 ubiquitinyl hydrolase UCH37 (C12 family) [Hymenolepis microstoma],	4
CDS34661.1 zinc finger protein [Hymenolepis microstoma],	2
<i>M. corti</i>	
CDS19936.1 kinesin light chain 4 [Echinococcus granulosus],	2
<i>S. solidus</i>	
XP_005107471.1 PREDICTED: craniofacial development protein 2-like [Aplysia californica],	2
AFM74212.1 ferritin-like protein [Spirometra erinaceieuropaei],	3
CDS22215.1 Na:K ATPase alpha subunit [Echinococcus granulosus],	2
CDS24615.1 myosin heavy chain [Echinococcus granulosus], CDS41550.1 myosin heavy chain [Echinococcus multilocularis],	2
<i>S. erinaceieuropaei</i>	
EUB64307.1 Cell cycle checkpoint protein RAD1 [Echinococcus granulosus],	2
CDS22002.1 rho-associated protein kinase [Echinococcus granulosus],	4

EUB64148.1 Sodium channel protein para [Echinococcus granulosus],	4
CDS21471.1 uridine cytidine kinase 1 1 [Echinococcus granulosus],	5
AGC52721.1 cytoplasmic dynein light chain 1 [Spirometra erinaceieuropaei],	4
AFM74210.1 tenascin-C [Spirometra erinaceieuropaei],	4
CDS18277.1 zinc finger SWIM domain containing protein [Echinococcus granulosus],	4
KRX96828.1 Lysine-tRNA ligase, partial [Trichinella pseudospiralis], KRY73560.1 Lysine-tRNA ligase, partial [Trichinella pseudospiralis],	2
CDS33789.1 filamin [Hymenolepis microstoma],	2
CDS29492.1 H transporting ATP synthase O subunit [Hymenolepis microstoma],	2
CDI97505.1 eukaryotic translation initiation factor 3 [Echinococcus multilocularis], EUB60942.1 Eukaryotic translation initiation factor 3 subunit C [Echinococcus granulosus],	2
CDS41468.1 phosphoglycerate mutase [Echinococcus multilocularis],	2
EUB60796.1 AP-1 complex subunit gamma-1 [Echinococcus granulosus],	4
XP_008294557.1 PREDICTED: WD repeat-containing protein 5-like [Stegastes partitus],	2
XP_013792856.1 protein will die slowly, partial [Limulus polyphemus],	
GAA53583.1 NAD(P) transhydrogenase [Clonorchis sinensis],	2
<hr/> <i>T. asiatica</i> <hr/>	
AAW88556.1 oncosphere protein Tso31a [Taenia solium],	2
CDS24815.1 lon protease mitochondrial [Echinococcus granulosus],	3
CDS35735.1 neutral amino acid transporter A [Echinococcus multilocularis], CDS35736.1 neutral amino acid transporter A [Echinococcus multilocularis],	3
EUB54936.1 Arginase-2 [Echinococcus granulosus],	2
EUB57861.1 Heat shock cognate protein [Echinococcus granulosus],	2
<hr/> <i>T. saginata</i> <hr/>	
EUB57021.1 Heat shock cognate protein [Echinococcus granulosus],	3
CUT99846.1 histone H2A [Echinococcus multilocularis],	2
CDS24809.1 histone H2B, partial [Echinococcus granulosus],	2
CUT98920.1 solute carrier family 13 [Echinococcus multilocularis],	2
CDS25215.1 RNA directed DNA polymerase reverse transcriptase, partial [Echinococcus granulosus],	4
CDS36977.2 RNA directed DNA polymerase reverse transcriptase [Echinococcus multilocularis],	
<hr/> <i>T. solium</i> <hr/>	
CDS22284.1 puromycin sensitive aminopeptidase [Echinococcus granulosus], EUB57161.1 Puromycin-sensitive aminopeptidase [Echinococcus granulosus],	2
EUB55282.1 Estradiol 17 beta-dehydrogenase [Echinococcus granulosus],	2
<hr/> <i>C. sinensis</i> <hr/>	
GAA48447.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA50112.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA50414.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA50969.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA50970.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2, partial [Clonorchis sinensis], GAA52884.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA53017.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA53893.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA54331.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA54348.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA54719.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA55304.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA56100.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	15
GAA51502.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis], GAA54168.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	3

GAA47508.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA47605.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA47632.1 endonuclease-reverse transcriptase [Clonorchis sinensis],	
GAA47769.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA48256.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA49300.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA49760.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA50165.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA50884.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA51039.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA51148.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA51569.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA52320.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA54607.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA54740.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA55175.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA55570.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA55666.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA55705.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA55723.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA56046.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA56809.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA49796.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	2
GAA50583.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA47942.1 ATP-binding cassette transporter [Clonorchis sinensis],	
GAA52187.1 ATP-binding cassette transporter [Clonorchis sinensis],	2
GAA39798.2 amine GPCR [Clonorchis sinensis],	
GAA54720.1 amine GPCR [Clonorchis sinensis],	2
GAA48255.1 protein disulfide-isomerase, partial [Clonorchis sinensis],	
GAA52646.1 protein disulfide-isomerase [Clonorchis sinensis],	7
GAA56061.1 protein disulfide-isomerase, partial [Clonorchis sinensis],	
GAA57118.1 protein disulfide-isomerase, partial [Clonorchis sinensis],	
GAA57403.1 protein disulfide-isomerase, partial [Clonorchis sinensis],	
GAA48074.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	4
GAA50579.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
GAA52062.1 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2 [Clonorchis sinensis],	
<hr/> <i>E. caproni</i> <hr/>	
GAA49374.1 dynein heavy chain 5 axonemal [Clonorchis sinensis],	
XP_018646894.1 ciliary outer arm dynein beta heavy chain-related [Schistosoma mansoni],	2
<hr/> <i>F. hepatica</i> <hr/>	
PIS77885.1 papain family cysteine protease [Fasciola hepatica],	2
PIS92198.1 putative Cys-rich domain protein [Fasciola hepatica],	
PIS92199.1 putative Cys-rich domain protein [Fasciola hepatica],	2
PIS79113.1 transporter, major facilitator family protein [Fasciola hepatica],	2
ABQ95351.1 secreted cathepsin L2 [Fasciola hepatica],	4
PIS83512.1 putative chaperone protein DnaK [Fasciola hepatica],	
PIS83513.1 putative chaperone protein DnaK [Fasciola hepatica],	2
AAF88069.1 amoebapore-like protein [Fasciola hepatica],	
AAX68445.1 SAP-1 [Fasciola gigantica],	2
PIS79273.1 epididymal secretory protein E1 family protein [Fasciola hepatica],	3
<hr/> <i>O. viverrini</i> <hr/>	
<i>S. curassoni</i> <hr/>	
AAX30396.2 SJCHGC03271 protein, partial [Schistosoma japonicum],	2
XP_012798356.1 Protein asteroid 1 [Schistosoma haematobium],	2

XP_012801036.1 Cathepsin B-like cysteine proteinase [Schistosoma haematobium], XP_012801037.1 Cathepsin B-like cysteine proteinase, partial [Schistosoma haematobium],	2
<i>S. haematobium</i>	
XP_012792511.1 Venom allergen 5 [Schistosoma haematobium], XP_012792521.1 GLIPR1-like protein 1 [Schistosoma haematobium],	2
AAW25617.1 SJCHGC02887 protein [Schistosoma japonicum],	2
XP_012792756.1 Ca(2+)/calmodulin-responsive adenylate cyclase [Schistosoma haematobium],	2
XP_012792765.1 Ca(2+)/calmodulin-responsive adenylate cyclase [Schistosoma haematobium],	
<i>S. japonicum</i>	
CAX83709.1 Gag-Pol polyprotein [Schistosoma japonicum],	8
AAW24674.1 unknown [Schistosoma japonicum], AAW24710.1 unknown [Schistosoma japonicum],	3
CAX70257.1 Cathepsin B-like cysteine proteinase precursor [Schistosoma japonicum],	
CAX73628.1 6-glucosidase [Schistosoma japonicum], CAX73930.1 Alpha-glucosidase [Schistosoma japonicum],	2
AAW27851.1 SJCHGC02198 protein [Schistosoma japonicum], CAX83705.1 Gag-Pol polyprotein [Schistosoma japonicum],	8
CAX74030.1 Glutamic acid-rich protein precursor [Schistosoma japonicum],	4
AAX31115.1 SJCHGC09830 protein [Schistosoma japonicum],	2
AAW27851.1 SJCHGC02198 protein [Schistosoma japonicum],	3
AAW25344.1 SJCHGC06760 protein [Schistosoma japonicum],	4
ABY55276.1 cytochrome oxidase subunit I, partial (mitochondrion) [Schistosoma japonicum],	11
AAW26418.1 SJCHGC09590 protein [Schistosoma japonicum],	2
AAW26800.1 SJCHGC02019 protein [Schistosoma japonicum], AAW27905.1 SJCHGC01171 protein [Schistosoma japonicum],	3
ADG47005.1 cytochrome c oxidase subunit II (mitochondrion) [Schistosoma japonicum],	16
AAW25871.1 SJCHGC00146 protein [Schistosoma japonicum], CBX24373.1 TPA: transposase [Schistosoma japonicum],	3
AAP05886.1 SJCHGC02868 protein [Schistosoma japonicum], CAX70512.1 Cathepsin L-like proteinase precursor [Schistosoma japonicum],	3
CAX83694.1 Gap-Pol polyprotein [Schistosoma japonicum],	2
AAX26430.2 SJCHGC06497 protein, partial [Schistosoma japonicum],	3
CAX74194.1 TonB box, N-terminal, domain-containing protein [Schistosoma japonicum],	2
AAW24817.1 SJCHGC05252 protein [Schistosoma japonicum],	2
AAX26603.2 SJCHGC06460 protein, partial [Schistosoma japonicum], CAX70180.1 Annexin A3 (Annexin III) [Schistosoma japonicum],	2
CAX74232.1 Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 [Schistosoma japonicum],	2
XP_018651153.1 LOW QUALITY PROTEIN: KRAB-A domain-containing protein [Schistosoma mansoni],	5
CAX83708.1 Gag-Pol polyprotein [Schistosoma japonicum],	5
<i>S. mansoni</i>	
XP_018655431.1 putative e1b-55kD-associated protein [Schistosoma mansoni], XP_018655432.1 putative e1b-55kD-associated protein [Schistosoma mansoni], XP_018655433.1 putative e1b-55kD-associated protein [Schistosoma mansoni], XP_018655434.1 putative e1b-55kD-associated protein [Schistosoma mansoni],	6
XP_018652502.1 putative alpha-amylase [Schistosoma mansoni], XP_018652504.1 putative alpha-amylase [Schistosoma mansoni],	2
XP_018644735.1 unnamed protein product [Schistosoma mansoni], XP_018644754.1 unnamed protein product [Schistosoma mansoni], XP_018645596.1 unnamed protein product [Schistosoma mansoni],	7
XP_018654725.1 uncharacterized protein Smp_203620 [Schistosoma mansoni], XP_018654815.1 uncharacterized protein Smp_203660 [Schistosoma mansoni], XP_018655443.1 uncharacterized protein Smp_203930 [Schistosoma mansoni],	
XP_018645754.1 rna recognition motif (rrm) domain containing protein, putative [Schistosoma mansoni], XP_018645755.1 rna recognition motif (rrm) domain containing protein, putative [Schistosoma mansoni], XP_018645756.1 rna recognition motif (rrm) domain containing protein, putative [Schistosoma mansoni],	4

XP_018647223.1 egg protein CP391S-like [Schistosoma mansoni], XP_018652514.1 egg protein CP391S-like [Schistosoma mansoni],	5
XP_018653704.1 putative tropomyosin [Schistosoma mansoni], XP_018653705.1 putative tropomyosin [Schistosoma mansoni], XP_018653706.1 putative tropomyosin [Schistosoma mansoni],	6
XP_018646757.1 MEG-3 (Grail) family [Schistosoma mansoni], XP_018646759.1 MEG-3 (Grail) family [Schistosoma mansoni],	2
XP_018648790.1 putative fgf receptor activating protein [Schistosoma mansoni], XP_018652024.1 putative fgf receptor activating protein [Schistosoma mansoni],	2
XP_018646414.1 invadolysin (M08 family) [Schistosoma mansoni], XP_018646417.1 leishmanolysin-2 (M08 family) [Schistosoma mansoni],	2
XP_018647435.1 unnamed protein product [Schistosoma mansoni], XP_018647437.1 unnamed protein product [Schistosoma mansoni], XP_018647438.1 unnamed protein product [Schistosoma mansoni],	4
XP_018653702.1 putative tropomyosin [Schistosoma mansoni], XP_018653708.1 putative tropomyosin [Schistosoma mansoni], XP_018653710.1 putative tropomyosin [Schistosoma mansoni],	6
<hr/> <i>S. margrebowiei</i> <hr/>	
XP_012801036.1 Cathepsin B-like cysteine proteinase [Schistosoma haematobium],	2
CAX83710.1 endonuclease-reverse transcriptase [Schistosoma japonicum],	2
CAX83710.1 endonuclease-reverse transcriptase [Schistosoma japonicum],	2
XP_018649464.1 uncharacterized protein Smp_200200 [Schistosoma mansoni],	13
XP_018649464.1 uncharacterized protein Smp_200200 [Schistosoma mansoni],	3
ADG95879.1 alkaline phosphatase [Schistosoma mansoni], XP_018652099.1 putative alkaline phosphatase [Schistosoma mansoni],	4
CAX83710.1 endonuclease-reverse transcriptase [Schistosoma japonicum],	12
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
CAX83710.1 endonuclease-reverse transcriptase [Schistosoma japonicum], XP_018648250.1 putative prokaryotic DNA topoisomerase [Schistosoma mansoni],	2
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	12
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
CAX83710.1 endonuclease-reverse transcriptase [Schistosoma japonicum],	2
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	12
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	3
AA556756.1 reverse transcriptase SR3-right, partial [Schistosoma mansoni],	3
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
CAX83710.1 endonuclease-reverse transcriptase [Schistosoma japonicum],	3
XP_012797173.1 Protein CASP [Schistosoma haematobium],	3
CAX83710.1 endonuclease-reverse transcriptase [Schistosoma japonicum],	3
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
XP_012799362.1 Bile salt export pump [Schistosoma haematobium],	3
XP_018648250.1 putative prokaryotic DNA topoisomerase [Schistosoma mansoni],	2
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	4
CAJ00236.1 TPA: endonuclease-reverse transcriptase [Schistosoma mansoni],	2
XP_018650238.1 putative fatty-acid amide hydrolase [Schistosoma mansoni],	2
CAX69930.1 LanC-like protein 1 [Schistosoma japonicum], CAX74215.1 LanC-like protein 1 [Schistosoma japonicum],	2
XP_018652890.1 putative alpha-galactosidase/alpha-n-acetylgalactosaminidase [Schistosoma mansoni],	2
<hr/> <i>S. matthei</i> <hr/>	
XP_012791939.1 Phosphatidylinositol 4-kinase alpha [Schistosoma haematobium],	2
<hr/> <i>S. rodhaini</i> <hr/>	
XP_018652895.1 putative alpha-galactosidase/alpha-n-acetylgalactosaminidase [Schistosoma mansoni],	3
XP_018645443.1 DNA-directed RNA polymerase I, putative [Schistosoma mansoni],	3
XP_012801017.1 Titin [Schistosoma haematobium],	2
<hr/> <i>T. regenti</i> <hr/>	

XP_012796759.1 Spectrin alpha chain [Schistosoma haematobium], XP_018648992.1 putative spectrin beta chain, brain 4 (Spectrin, non-erythroid beta chain 4) (Beta-V spectrin) (BSPECV) [Schistosoma mansoni],	3
<i>G. salaris</i>	
<i>P. xenopodis</i>	
ODM93751.1 Eukaryotic translation initiation factor 3 subunit D [Orchesella cincta],	2
<i>M. lignano</i>	
XP_013395914.1 homeobox protein prophet of Pit-1 [Lingula anatina],	4
ALD16261.1 Peptidyl-prolyl cis-trans isomerase FKBP6 [Macrostomum lignano],	4
XP_013919212.1 PREDICTED: class E basic helix-loop-helix protein 23 [Thamnophis sirtalis],	4
XP_015787693.1 PREDICTED: uncharacterized membrane protein DDB_G0293934-like [Tetranychus urticae],	
XP_021373861.1 ELAV-like protein 3 isoform X3 [Mizuhopecten yessoensis],	3
PFX20388.1 Tubulin alpha-1D chain [Stylophora pistillata],	5
<i>S. mediterranea</i>	
ALJ75660.1 ASH2 [Schmidtea mediterranea],	3
AHD24739.1 IFI6-like protein, partial [Schmidtea mediterranea],	2
ADW27177.1 axis inhibition protein A [Schmidtea mediterranea],	2
AKN21403.1 slc2a-9, partial [Schmidtea mediterranea],	2
AFJ24785.1 vault protein inter alpha trypsin-1, partial [Schmidtea mediterranea],	6

Se considera que dos secuencias son homólogas si muestran una identidad de al menos 75%, una cobertura de al menos 40% y un *e-value* de a lo sumo 1e-5. Se evitan aquellos casos donde la homología se da con secuencias anotadas como *hypothetical protein*.