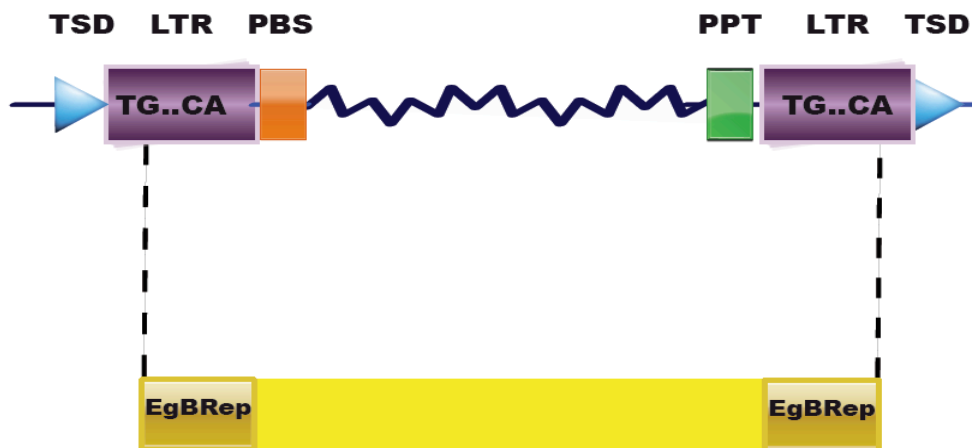


# *Identificación y caracterización de una nueva familia de elementos transponibles en platelmintos.*



***Responsable: Br. Santiago Radío.***

*Orientador: Dr. Pablo Smircich.*

*Co-orientadora: Dra. Cecilia Fernández.*



> Introducción.....	4
Aspectos generales de los Ténidos.....	4
Equinococosis y cisticercosis.....	6
Anatomía y ciclo de vida de las especies del género <i>Echinococcus</i> .....	7
Aspectos genómicos en Ténidos.....	11
Elementos genéticos móviles: historia e impacto evolutivo.....	13
Aspectos generales de los elementos transponibles.....	13
<i>TRIMs (Terminal repeat Retrotransposons In Miniature)</i> .....	15
>Materiales, métodos y estrategia:.....	21
Obtención de ESTs ( <i>Expressed sequence tags</i> ) y ensamblados genómicos.....	21
Búsqueda de <i>TRIMs</i> en <i>E. multilocularis</i> y en otros platelmintos.....	21
Identificación del Sitio de Unión al Primer (PBS) de <i>TRIM</i> .....	25
Caracterización de estructura secundaria de <i>TRIMs</i> .....	25
Distribución de <i>TRIMs</i> e identificación de sitios preferenciales de inserción.....	25
Análisis filogenéticos.....	26
Búsqueda de retrotransposones de LTRs autónomos.....	27
Construcción de Logos.....	27
Análisis de datos de expresión.....	27
Asociación de <i>TRIMs</i> con genes.....	28
>Resultados .....	29
Descubrimiento de elementos <i>TRIM</i> en cestodos ténidos.....	29
Descripción de elementos transponibles en <i>Echinococcus</i> .....	29
Búsqueda de <i>TRIMs</i> en organismos relacionados.....	31
Clasificación de los <i>ta-TRIMs</i> según su conservación.....	33
Análisis de localización en el genoma.....	34
Aspectos evolutivos de <i>ta-TRIMs</i> .....	34
Análisis filogenéticos.....	34
Un elemento autónomo candidato para la movilización de <i>ta-TRIMs</i> en <i>trans</i> .....	37
<i>Lennie</i> , el candidato autónomo.....	37
Análisis de expresión masiva de <i>ta-TRIMs</i> y características funcionales.....	40

Expresión masiva de <i>ta-TRIMs</i> y generación de transcritos novel a partir de solo-LTRs.....	40
Análisis funcional de genes relacionados a <i>ta-TRIMs</i> . .....	45
Panorama general de la expresión de <i>ta-TRIMs</i> . .....	45
Análisis de expresión de <i>ta-TRIMs</i> en células germinativas de <i>E. multilocularis</i> .....	47
>Discusión.....	50
>Bibliografía.....	54
>Material suplementario .....	63

## > Introducción

### Aspectos generales de los Ténidos.

El parasitismo es el estilo de vida más común en la tierra. Los parásitos juegan un rol fundamental en los ecosistemas, contribuyendo enormemente a la biomasa (Strona y Fattorini, 2014). Los cestodos son un grupo de gusanos planos parásitos que constituyen la clase Cestoda del *filum* Platelmino. Los miembros de esta clase se caracterizan por no poseer intestino; realizan todo el intercambio metabólico a través de una cubierta sincicial externa, el tegumento. Anteriormente, poseen un órgano especializado, el escólice, con cuatro ventosas y un rostelo con una doble corona de ganchos. El cuerpo es segmentado y cada unidad es una unidad reproductiva (proglótida).

El número de cestodos es muy difícil de estimar en parte debido a que continuamente se descubren nuevos géneros y especies (Kornyushin y Greben, 2014). El mayor grupo taxonómico dentro de los cestodos es el orden de los ciclofilídeos cuyos adultos son parásitos de anfibios, reptiles, aves y mamíferos (Figura 1). Se estima que más de 15 familias y miles de especies integran este orden. La familia *Taeniidae* está conformada por los géneros *Echinococcus* y *Taenia* (*Eucestoda: Cyclophyllidea: Taeniidae: Taenia* y *Echinococcus*), como se observa en la Figura 1. Últimamente, se han realizado esfuerzos para definir la taxonomía de la familia *Taeniidae*, sin embargo, hasta el día de hoy no hay un panorama del todo claro (Nakao et al. 2013; Knapp et al. 2011).

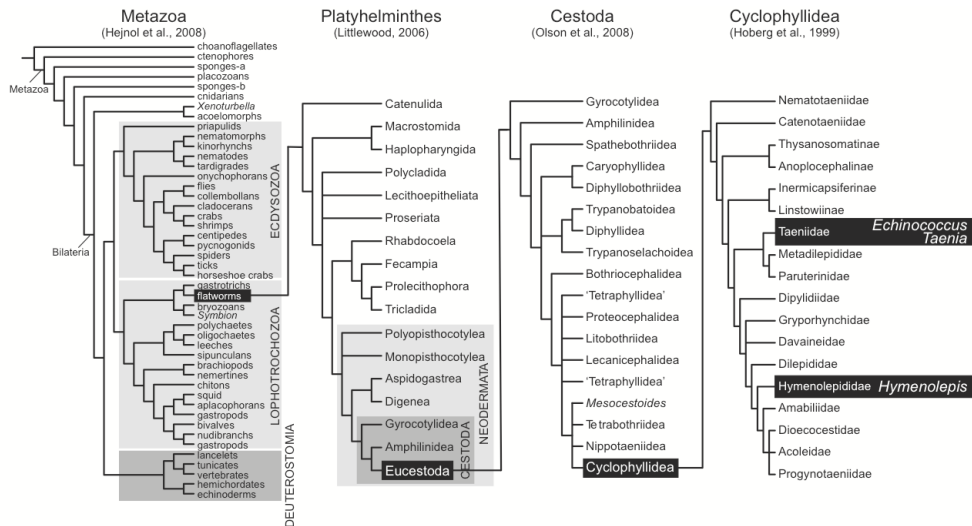


Figura 1. Historia evolutiva de los cestodos taenidos, tomada de Thompson y McManus 2002.

El género *Echinococcus* es un grupo taxonómico pequeño que contiene un núcleo de especies muy similares entre sí, y cada una de ellas tiene una relación hospedero-patógeno característica. La taxonomía de *Echinococcus* ha sido un tema controversial durante muchas décadas. Recientes estudios taxonómicos han revalorizado investigaciones pioneras en esta área ratificando especies previamente derogadas, gracias a un extensivo estudio de epidemiología molecular (Thompson y McManus 2002). Hasta mediados de la primera década del 2000, se reconocían como válidas 4 especies del género *Echinococcus* (*E. multilocularis*, *E. granulosus*, *E. oligarthus* y *E. vogeli*), pero recientemente se han descrito dos especies; *Echinococcus shiquicus*, presente en zorros de la meseta tibetana (Xiao et al. 2006) y *Echinococcus felidis*, aislado de leones de África (Hüttner et al. 2008). El número de proglótidas, la forma de los ganchos rostellares, la ubicación del poro genital y la disposición de los órganos reproductivos son ligeramente variables entre especies, sin embargo de sorprendente similitud (Nakao et al. 2013).

El género *Taenia* es un gran grupo taxonómico integrado por lo menos por 42 especies. Algunas de ellas fueron de los primeros helmintos reconocidos por los seres humanos, con registros escritos que se extienden hasta la antigüedad (Hoberg 2006). Solo *T. solium*, *T. saginata* y *T. asiatica* son capaces de infectar humanos; eso provoca que la mayor parte de los estudios se centren en ellas, principalmente en *T. solium*. La divergencia entre especies es mucho mayor que en el caso de *Echinococcus*.

### **Equinococosis y cisticercosis.**

Los Taenidos son un grupo importante de helmintos de importancia global a nivel sanitario y económico. Incluyen los parásitos responsables de la equinococosis y cisticercosis. Estas, y muchas de las enfermedades que tienden a afectar a los sectores socio-económicos más débiles, por lo general no son priorizadas a nivel regional ni global.

El término clínico equinococosis, más conocida como hidatidosis, es usado para describir un grupo de enfermedades zoonóticas causadas por la infección de la larva (metacestodo) del género *Echinococcus*. Las especies más reconocidas por su impacto, tanto médico como veterinario y económico, son *Echinococcus multilocularis* y *Echinococcus granulosus*, causantes de las enfermedades equinococosis quística (EC) y equinococosis alveolar (EA), respectivamente. *E. multilocularis* está geográficamente confinado al hemisferio norte, pero dentro de esta región tiene una amplia distribución. La EA humana es una enfermedad emergente en Europa. Estudios de la vida silvestre están detectando al parásito en nuevas áreas (Torgerson et al. 2010). Aún no está claro si el parásito está ampliando su rango de infección o si el mayor control y las nuevas técnicas de diagnóstico provocan un aumento de la tasa de detección (Torgerson et al. 2010). La mayor prevalencia de EC en hospederos animales y humanos se encuentra en zonas templadas. Incluyendo el sur de Sudamérica, el litoral del Mediterráneo, algunas zonas de Norteamérica (generalmente introducida por inmigrantes), Europa del este, Asia central, China, Australia y algunas partes de África (Moro y Schantz 2009). Está presente en más de 100 países y en todos los continentes, excluyendo la Antártida (Eckert y Deplazes 2004).

Existen diferentes cepas de *E. granulosus*, que difieren en la especificidad por sus hospederos. Numerosos estudios moleculares han definido la existencia de 10 genotipos (G1-10). Se presentan; G1 y G2 en ovinos, G3 y G5 en bovinos, G4 en equinos, G6 en camélidos, G7 y G9 en suinos, G8 y G10 en cérvidos. La noción de una serie de especies de *Echinococcus* específicas de distintos hospederos no es nueva, y muchos de los primeros parasitólogos en revisar estos aspectos habían realizado observaciones al respecto. En relación a esto, fue propuesto que muchos

de estos genotipos asciendan a la categoría de especie, incluyendo G1 (*Echinococcus granulosus sensu stricto*) y G4 (*Echinococcus equinus*) (Moro y Schantz 2009; Nakao et al. 2013). La cepa G1 es la forma más frecuente en las ciudades y es la más asociada a infecciones humanas.

Hasta la década de 1980, la cirugía era la única opción para el tratamiento de equinocosis; sin embargo, la quimioterapia con compuestos de benzimidazol y, más recientemente, el tratamiento de punción quística, aspiración, inyección de productos químicos y re-aspiración; han permitido grandes avances en el tratamiento de la infección. No obstante, la EA es muy difícil de tratar y los tratamientos no dan gran resultado. En el caso de EC se obtienen mejores resultados, ya que la cirugía es menos peligrosa. En ambos casos el tratamiento, no quirúrgico, se basa en el uso de benzimidazoles que afectan la  $\beta$ -tubulina necesaria para el ensamblaje del citoesqueleto. Sin embargo, como la  $\beta$ -tubulina del parásito es similar a la del hospedero el uso de este medicamento trae muchos efectos secundarios indeseados (Epping y Brehm 2011; Brunetti 2010).

La cisticercosis es una enfermedad causada por la presencia en los tejidos de larvas (cisticercos) de varias especies de cestodos del género *Taenia*. En particular la cisticercosis humana es causada por el metacestodo de *T. solium*. La transmisión de *T. solium* y el desarrollo de la neurocisticercosis es prevalente en muchas zonas del África sub-sahariana, centro América, norte de Sudamérica y partes del sureste asiático.

En las últimas décadas, la combinación de las pruebas modernas de diagnóstico, el uso de fármacos antiparasitarios, tratamientos anti-inflamatorios mejorados y neurocirugía mínimamente invasiva; han mejorado el pronóstico de los pacientes infectados con *T. solium*. A pesar de estos avances, la neurocisticercosis es todavía la infección neurológica por helmintos más común, siendo un importante problema de salud pública en gran parte del mundo (García et al. 2014).

#### **Anatomía y ciclo de vida de las especies del género *Echinococcus*.**

El género *Echinococcus* muestra ciertas características que lo alejan de otros géneros principales de la familia de los Ténidos. Los adultos miden unos pocos

milímetros de largo y raramente poseen más de cinco proglótidas (3-6 milímetros); mientras que los de algunas especies de la familia *Taeniidae* pueden llegar a medir varios metros y poseer miles de proglótidas. A diferencia de lo que sucede en *Taenia*, las larvas de *Echinococcus* poseen poca especificidad por sus hospederos y un enorme potencial reproductivo. *Echinococcus* requiere de dos hospederos mamíferos para completar su ciclo de vida (Figuras 2 y 3). El gusano adulto se desarrolla en el intestino del hospedero definitivo, el cual siempre es un carnívoro y por lo general un cánido. La región anterior (como en el resto de los cestodos) es el escólice, y presenta los órganos necesarios para adherirse a la pared intestinal del hospedero, como son el rostelo y las ventosas. Inmediatamente posterior al escólice, la región del cuello genera continuamente unidades reproductivas hermafroditas. Este proceso de formación, desarrollo y maduración de proglótidas se conoce como proglotización. La segmentación en cestodos se da de forma proximal y no distal como en otros organismos; además, en algunos casos, la formación de proglótidas puede no estar acompañada por la segmentación (Freeman et al. 1973). Dentro de cada proglótida ocurre la fertilización. La reproducción asexual (rara en cestodos) es la forma predominante, pero debido al carácter hermafrodita de las proglótidas, también ocurre reproducción. Se cree que el hermafroditismo combinado con la auto-inseminación es una ventaja evolutiva ya que, para pequeños gusanos, puede resultar difícil encontrar un compañero adecuado (Swiderski et al. 2007). Una vez producidos, los huevos son liberados junto a las heces del hospedero al ambiente. Se cree que los huevos liberados ya están plenamente embrionados y que son infectivos. En cambio, los huevos expulsados por los Ténidos pueden estar en diferentes etapas de desarrollo y se cree que pueden madurar, si las condiciones ambientales son las adecuadas, a huevos infectivos (Craig et al. 2003). Los huevos liberados por *Echinococcus* contienen el primer estado larvario, la oncoesfera. Una vez ingeridos por el hospedero intermediario adecuado, los huevos viables de *Echinococcus* se adhieren al intestino, y mediante la disgregación de la cubierta, liberación y activación de la oncoesfera, entran al sistema porta desde donde son distribuidos al resto del cuerpo. En el caso de *E. granulosus*, el metacestodo se puede desarrollar en casi cualquier órgano; en el caso de *E. multilocularis*, el desarrollo es casi exclusivamente en el hígado. Sin embargo, un 80% de los pacientes estudiados por Larrieu y Fried tenían un único órgano involucrado, generalmente el hígado en *E.*



*multilocularis*; el pulmón y, en menor medida, el cerebro en *E. granulosus*. (Brunetti et al. 2010). Una vez activada, la oncoesfera atraviesa una metamorfosis que dura aproximadamente 1-14 días, e involucra proliferación celular, degeneración de los ganchos, atrofia muscular, formación de una cavidad central, vesicularización, y desarrollo de las capas germinal y laminar (Heath y Lawrence 1976). En el caso de *E. granulosus*, cada oncoesfera desarrolla una única vesícula (unilocular); en cambio, el desarrollo de *E. multilocularis* es multivesicular. En este caso, en lugar de desarrollarse brotes endógenos (ver más adelante), se desarrollan brotes exógenos (Eckert et al. 1983). Estos quistes (vesículas), llenos de fluido hidatídico, presentan una capa germinal interna cubierta por un tegumento sincitial que secreta una capa acelular, elástica y rica en carbohidratos; la capa laminar. La capa germinativa está formada por diferentes tipos celulares como las células tegumentarias, musculares, indiferenciadas, etc. Las células indiferenciadas son proliferativas y son las encargadas de formar las llamadas vesículas prolíferas (*brood capsules*), las cuales se originan como pequeñas masas celulares o brotes, que se expanden y proliferan hacia la cavidad quística. Dentro de cada vesícula surgen nuevos brotes que son los encargados de formar protoescólices. Los protoescólices son la forma infectiva del hospedero definitivo (Thompson 1986; Koziol, Krohne y Brehm 2013). El ciclo de vida de *Echinococcus* se cierra cuando el hospedero definitivo, accidentalmente el humano, ingiere de manera directa o indirecta un hospedero intermediario infectado. Luego de la infección, el protoescólice se adhiere al intestino y se desarrolla a gusano adulto.

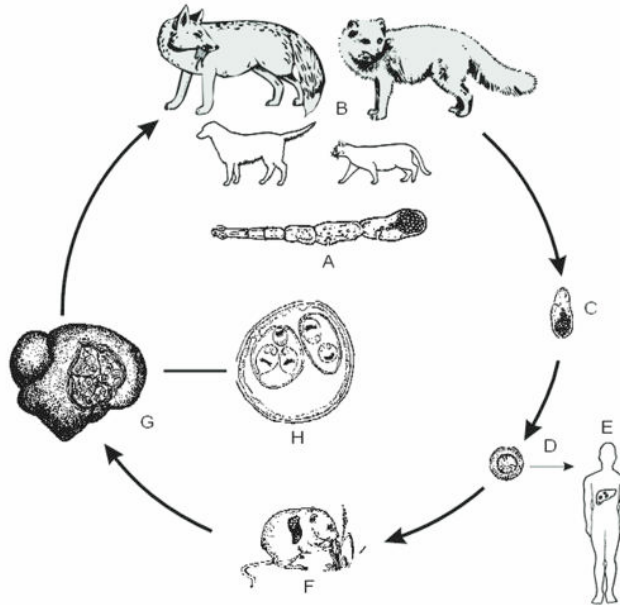


Figura 2. Ciclo de vida de *E. multilocularis*. (A) Parásitos adultos. (B) Zorros (izquierda, zorro Rojo; derecha, zorro Ártico) como principales hospederos definitivos; perros, otros cánidos, y gatos pueden verse involucrados en el ciclo. (C) Proglótida con huevos. (D) Huevos con oncoesferas. (E) Infección de humanos. (F) Roedor infectado con metacestodos. (G) Hígado de roedor con metacestodos. (H) Metacestodo con protoescólices. Tomado de <http://cmr.asm.org/>

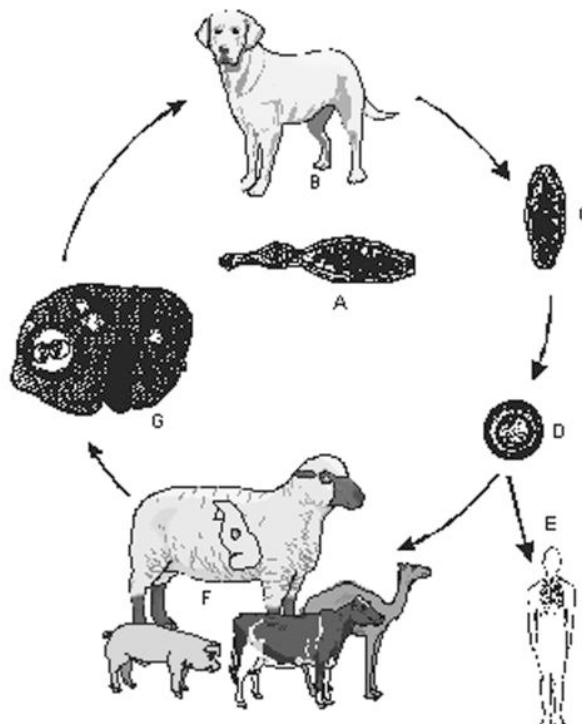


Figura 3. Ciclo de vida de *E. granulosus* (cepa común de oveja). (A) Parásitos Adultos. (B) Perros doméstico como principales hospederos definitivos; cánidos salvajes (dingo, hiena, etc.) pueden verse involucrados en el ciclo. (C) Proglótida con huevos. (D) Huevos con oncoesferas. (E) Infección de humanos. (F) Oveja como principal hospedero intermediario; otros ungulados, poco significativos como hospederos. (G) Hígado de oveja con metacestodos. Tomado de <http://cmr.asm.org/>

### Aspectos genómicos en Ténidos.

El primer genoma en ser secuenciado fue el de un parásito, un bacteriófago llamado *phiX174* (Sanger et al. 1977). Este hecho marcó una nueva era en la biología, provocando un cambio de enfoque. Los biólogos moleculares pasaron de estudiar genes individuales a estudiar el sistema entero. Si bien este hecho fue revolucionario, hubo que esperar más de dos décadas para ver el primer genoma bacteriano completo. El año 2001 marcó un hecho histórico para la humanidad, ya que en ese año se presentaron los primeros genomas humanos. A finales del 2002, se publicó el primer genoma de un parásito humano, *Plasmodium falciparum* (Gardner et al. 2002). Con la introducción de las tecnologías de secuenciamiento de nueva generación a mediados del 2006, se logró pasar a una nueva etapa en la era genómica. Estos avances tecnológicos permitieron que los análisis genómicos se trasladaran de las grandes instalaciones a los laboratorios individuales (Forrester y Hall 2014). Hoy en día, las tecnologías de secuenciamiento siguen avanzando a un ritmo inesperado y por lo tanto se espera que el número de genomas secuenciados se incremente rápidamente. Adicionalmente, los estudios de expresión génica utilizando RNA-Seq (secuenciamiento del transcriptoma) proveen a los investigadores la posibilidad de visualizar la expresión del RNA en forma cuantitativa.

Particularmente, en cestodos el secuenciamiento de genomas completos comenzó en el 2004 e incluye actualmente a *E. multilocularis*, *E. granulosus*, *T. solium* e *Hymenolepis microstoma* (Tabla 1). Los genomas de los cestodos son pequeños en tamaño, alrededor de 110 Mb comparado con los 363 Mb en *Schistosoma* (Digenea) (Berriman et al. 2009), 700 Mb en *Schmidtea* (planaria, Tricladida) (Robb et al. 2008) y 330-1100 Mb en *Macrostomum* (Macrostomida) (Olson et al 2012) (ver Figura 1).

<i>Especie</i>	<i>Tamaño Genómico (Mb)</i>	<i>% G/C</i>	<i>Nro. cromosomas</i>
<i>E. multilocularis</i>	113	42	2N = 18
<i>E. granulosus</i>	106	42	2N = 18
<i>T. solium</i>	117	43	2N = 18?
<i>H. microstoma</i>	147	35	12

Tabla 1. Características genómicas de los cestodos. Adaptado de Tsai et al 2013.

Se ha propuesto que esta diferencia se debe principalmente a que los cestodos poseen menos elementos genéticos móviles que los trematodos o las planarias (Olson et al. 2012; Tsai et al. 2013; Copeland CS et al. 2003; Copeland CS et al. 2005). El análisis de los genomas y transcriptomas completos de cuatro cestodos presentado por Tsai et al. en 2013 (*E. multilocularis*, aislados Europeos; *E. granulosus sensu stricto*, de Uruguay; *T. solium*, de México; *H. microstoma*, cepa Nottingham) y el genoma de un aislado chino de *E. granulosus sensu stricto* realizado por Zheng et al. (Zheng et al. 2013), revelaron extensas adaptaciones al parasitismo. Entre ellas, encontramos la pérdida de muchas vías importantes en la síntesis *de novo* de aminoácidos, nucleótidos, ácidos grasos y colesterol; los cuales tienen que ser tomados del hospedero. Gran parte de esta pérdida génica se puede asociar a la ausencia de intestino (Olson et al. 2012). Además, componentes claves en el desarrollo, como las familias génicas *Hox* y *Wtn*, se encuentran reducidos en número. Sin embargo, una de las conclusiones más importantes para esta tesis a la cual llegaron ambos estudios fue que, así como los trematodos (Wang et al. 2013), los cestodos emplearían un sistema de células madres altamente modificado con respecto al de los platelmintos de vida libre, lo cual fue confirmado por estudios posteriores (ver más adelante y Koziol et al 2014).

Los cestodos, así como muchos otros invertebrados, poseen una línea germinal discontinua, la cual es generada por células madres luego del desarrollo embrionario (Extavour y Akam 2003; Juliano, Swartz y Wessel 2010). En planaria y otros platelmintos de vida libre, estas células madres somáticas, multipotentes son denominadas neoblastos y son la base celular para las capacidades regenerativas, incluida la de formar *de novo* la línea germinal (Sato et al. 2006; Wang et al. 2007;

Juliano, Wang y Lin 2011; Rink 2013). Los platelmintos de vida libre expresan homólogos de la proteína *piwi* en sus neoblastos. En cambio, los ortólogos de *piwi* se han perdido en la línea de los Neodermata (incluye a Digenea y Cestoda; ver Figura 1). Sin embargo, en *S. mansoni* se ha identificado un gen *argonauta* (*sm-ago2-1*), altamente expresado en neoblastos, y se ha propuesto que puede realizar funciones similares a los PIWIs en otros organismos (Wang, Collins y Newmark 2013). Un ortólogo de este gen fue encontrado en *E. multilocularis* (Tsai et al. 2013), pero se expresa solo en una fracción de sus neoblastos, por lo que no parece ser candidato a sustituir el rol de los PIWIs (Koziol et al. 2014).

En cestodos, los neoblastos son denominados generalmente células germinales (Koziol y Castillo 2011). Estas células han sido recientemente caracterizadas mediante técnicas histológicas, como células redondas u ovaladas, presentando un citoplasma fuertemente basófilo, grandes núcleos y prominente nucléolo (Koziol et al. 2014). Koziol et al. demostraron que las únicas células capaces de proliferar en el metacestodo de *E. multilocularis* son las células germinales, que conducen probablemente a la continua proliferación de vesículas. Demostraron además que las células germinales no son una única población celular, sino que existen subpoblaciones dentro de ellas. En los estudios de Brehm et al. y en los transcriptomas publicados, se encontraron tres isoformas de  $\beta$ -tubulina; siendo la Tub-2 el isotipo predominante en los neoblastos (Tsai et al. 2013; Brehm et al. 2000; Skinner et al. 2014). Debido a la diferencia en la eficacia de la quimioterapia con benzimidazol sobre los diferentes tipos celulares de *Echinococcus*, Schubert et al. (Schubert et al. 2014), concluyeron que las células germinativas son insensibles a la quimioterapia porque expresan exclusivamente una  $\beta$ -tubulina con limitada afinidad al benzimidazol. Debido a esto, es de suma importancia que las drogas desarrolladas para el tratamiento de equinococosis afecten también a las células madres.

## **Elementos genéticos móviles: historia e impacto evolutivo.**

### **Aspectos generales de los elementos transponibles.**

A finales de la década de 1940, Barbara McClintock se encontraba investigando el comportamiento de los cromosomas en maíz y los efectos de la

deleciones y rearrreglos. En 1945, el mismo año en que fue nombrada Presidenta de la Sociedad de Genética de América (Genetics Society of America), McClintock caracterizó cuatro inusuales mutaciones inestables (McClintock 1946). Sus ideas fueron tomadas con cautela por la comunidad científica, ya que la noción de que ciertos elementos genéticos no fueran estables en sus posiciones en los cromosomas era antidogmática contrastando con la muy expandida idea de la estabilidad de la información genética hereditaria. Más adelante, en las cercanías de 1970, mutaciones inestables en los operones de galactosa y lactosa de *Escherichia coli* fueron caracterizadas como inserciones de una secuencia particular, encontrada en múltiples localizaciones a través del genoma (Shapiro et al. 1969). Sin embargo, la década de 1970 fue la década decisiva para que la comunidad científica empezara a considerar a estos elementos genéticos móviles (EGMs). En particular, fue muy importante la investigación en *Drosophila* del fenómeno disgenesia híbrida, en la cual cruces entre líneas específicas de *D. melanogaster* conducen a varios cambios genéticos incluyendo la esterilidad e incrementos en las tasas de mutación y recombinación (Picard et al. 1978; Biémont et al. 2010). Estos efectos fueron asociados a la movilización de ciertos elementos genéticos; elementos P (transposones de ADN) y elementos I (retrotransposones). Este descubrimiento despertó un gran interés en la comunidad científica no solo porque estos elementos eran capaces de moverse por el genoma, sino porque los elementos P e I invadieron todas las poblaciones de *D. melanogaster* en un espacio de 50 años (Anxolabéhère et al. 1988; Bélmont 2010). Los EGMs incluyen plásmidos, virus endógenos y elementos transponibles (ETs).

Mucho tiempo transcurrió de esas pioneras investigaciones, y durante tres décadas los ETs fueron concebidos como ADN parasitario. Los transposones tienen la habilidad de copiarse y moverse a nuevas posiciones dentro del genoma y, como consecuencia de eso, se pueden acumular. Su acumulación y movilidad hacen que puedan promover redistribuciones del genoma de manera directa o indirecta. Tienen la habilidad de iniciar eventos recombinacionales dada sus secuencias con alta homología dispersas en todo el genoma, en el mismo o diferente cromosoma. Los ETs pueden, a su vez, generar variaciones en la expresión génica, dependiendo su efecto de la localización específica del sitio en el que se integran. Luego de haber sido encontrados, estos elementos fueron considerados como “ADN basura” y se los

consideró ADN egoísta, en el sentido de ser una especie de parásito celular que no trae consigo ningún beneficio. Sin embargo, hoy se acepta que pueden tener varios beneficios para el genoma en el cual se encuentran. Por ejemplo, servir como mecanismo de reparación de roturas de la doble hebra o generar isoformas de genes gracias a que portan, por lo general, su propio promotor. Por otra parte, varios genes contienen en sus exones secuencias derivadas de elementos móviles. Es el caso de proteínas importantes como por ejemplo la telomerasa y proteínas que intervienen en la recombinación VDJ de las inmunoglobulinas. Por lo tanto, en términos evolutivos, tienden a ser beneficiosos ya que, a la larga, generan mayor variabilidad; pero, a corto plazo, pueden producir desregulaciones que podrían ser letales a nivel celular. Debido a los efectos deletéreos de los EGMs, el hospedero ha desarrollado mecanismos para suprimir su expansión y expresión. Esto es de particular importancia en la línea germinal de los animales, ya que la expansión de los EGMs puede llevar a su acumulación en la próxima generación. La vía PIWI constituye un mecanismo conservado en metazoarios para silenciar EGMs en la línea germinal (Juliano, Wang, y Lin 2011). PIWI es un miembro de la familia de las proteínas Argonautas que están involucradas en el silenciamiento génico mediante el uso de pequeños ARNs, en particular piARNs. Junto a los piARNs, PIWI forma un complejo piARN/PIWI que puede silenciar a los EGMs a nivel epigenético o mediante regulaciones post-transcripcionales de la estabilidad del ARN (Hock y Meister 2008). Como hemos mencionado anteriormente, este mecanismo no se encuentra presente en cestodos, y no parece haber un claro candidato que medie el silenciamiento de los EGMs; sin embargo, la carga de EGMs en los genomas de cestodos ténidos es baja en comparación a otros cestodos por lo que debe existir un mecanismo de supresión de EGMs (Tsai et al. 2013; Skinner et al. 2014).

En la actualidad los EGMs de todo tipo están en el foco de interés por su rol en la evolución molecular y su potencial como herramientas biotecnológicas.

***TRIMs (Terminal repeat Retrotransposons In Miniature).***

Existen diferentes clasificaciones de los ETs, basadas ya sea en propiedades estructurales, en mecanismos de transposición u otros. Recientemente Whicker propuso clasificar a los ETs en diferentes órdenes jerárquicos que simulen la clasificación de los organismos (Whicker 2007). Este sistema de clasificación de

ETs incluye los niveles de clase, subclase, orden, superfamilia, familia y subfamilia. La clase divide a los elementos transponibles en cuanto a presencia (clase I) o ausencia (clase II) de intermediario de transposición de ARN. La subclase está dividida entre los ETs que se copian a sí mismos para la inserción y los ETs que dejan su sitio dador para integrarse en cualquier otro lugar. En consecuencia, esta separación refleja el número de hebras de ADN que se deben cortar en el sitio donador. Las superfamilias dentro de un orden comparten la estrategia general de replicación pero se distinguen por aspectos como la estructura de dominios proteicos o no codificantes, presencia y tamaño del sitio blanco de duplicación (TSD; por su nombre en inglés: *target site duplication*). No hay conservación de secuencia a nivel de ADN y hay limitada similitud a nivel de proteínas. Las familias están definidas por el grado de conservación de secuencia de ADN. Las subfamilias se pueden definir en base a datos filogenéticos. El nivel más bajo de anotación es la inserción, y se refiere a un elemento específico.

En la clase II, como ya se ha mencionado, están los ETs que se mueven mediante transposición sin intermediario de ARN. Existen tres mecanismos predominantes mediante los cuales se da esta transposición:

1) Transposición replicativa; un transposón donante se duplica y el ADN intermedio se integra al sitio receptor.

2) Transposición no replicativa; el elemento se escinde, y se integra en otra región, sin necesidad de duplicarse.

3) Transposición conservativa; igual que la anterior, pero no se genera una ruptura luego de la salida del transposón. Ambos mecanismos requieren enzimas (transposasas), generalmente codificadas en dichos elementos o, si estos elementos están truncados, pueden utilizar las enzimas de otros transposones funcionales.

La secuencia donde se insertan este tipo de transposones tiene un consenso muy laxo y por lo tanto son capaces de moverse a casi cualquier lugar del genoma. Sin embargo, en una proporción muy alta de los casos se ve que la inserción es en una región cercana a la ubicación del transposon original. La mayoría de los elementos móviles en bacterias son transposones de ADN, a diferencia de los eucariotas que presentan mayormente retrotransposones.



Por otra parte, en la clase I, están los elementos transponibles que se mueven a través de un intermediario de ARN (rtnARN), los cuales se denominan retrotransposones. Este tipo de transposición es exclusivo de los organismos eucariotas. Para que se dé, el ARN sintetizado a partir de la secuencia de ADN del retrotransposón debe ser copiado nuevamente a ADN antes de su inserción en el genoma. Este proceso requiere una enzima que debe ser codificada por el retrotransposón (llamada retrotranscriptasa), ya que no es necesaria para el flujo “normal” de la información en la célula. Existen muchas variantes de retrotransposones, pero nos centraremos en los retrotransposones de LTR.

Los retrotransposones LTR (*long terminal repeat*) son un grupo principal de EGMs de Clase I; se diferencian del resto de los órdenes de su clase por poseer repetidos terminales largos directos en sus extremos, de ahí su nombre. Están presente en casi todos los eucariotas, y han sido muy estudiados por su relevancia evolutiva y, más recientemente, por su importancia biotecnológica (Boeke y Stoye 1997; Havecker, Gao y Voytas 2004; Wicker et al. 2007). Los LTR están estrechamente relacionados con los retrovirus, pero se diferencian de éstos por no poseer secuencias codificantes para la proteína *env* necesaria para la encapsidación viral. Se especula que los retrovirus pueden haber evolucionado a partir de un elemento LTR que adquirió la capacidad de sintetizar la envoltura viral al incorporar un gen celular. Los LTR flanquean secuencias codificantes para componentes estructurales (Gag) y enzimáticos (proteasa (PR), transcriptasa reversa (RT), RNAsa H (RH), e integrasa (IN)). Todos los componentes enzimáticos son sintetizados inicialmente como una poliproteína (Pol), la cual posteriormente es procesada en sus componentes individuales por acción de la PR. Una característica importante de los repetidos terminales es que cumplen funciones de promotor y terminador de la transcripción. EL 5' LTR funciona como un promotor de la ARN polimerasa II y el 3' LTR es el sitio en el cual termina la transcripción del ARN, el cual es posteriormente escindido y poliadenilado. Vale aclarar que ambas secuencias son idénticas. El ARNrtn, que carece de la región más 5' del 5' LTR (región U3) y la región más 3' del 3' LTR (U5), es exportado al citoplasma donde es traducido. En el citoplasma, el ARNrtn se asocia con Gag para formar la partícula tipo viral (VLP) en la cual ocurre la transcripción reversa. Este es un mecanismo complejo que genera un retrotransposon de ADNc de doble cadena y con LTR

completos (Telesnitsky y Goff 1997; Havecker, Gao y Voytas 2004; Figura 4). Existen otras secuencias que son fundamentales para el proceso de la transcripción reversa. El sitio de unión del *primer* (PBS, *primer binding site*) y el tracto de polipurinas (PPT, *polypurine tract*). El PBS se ubica casi inmediatamente después del 5' LTR, puede medir entre 8 y 19 bases y sirve como sitio de unión de un ARN de transferencia (ARNt). El ARNt actúa como cebador para el inicio de la síntesis de la cadena negativa de un retrotransposón de ADN (ADNrtn). Por otra parte, el PPT, que se ubica en la región *upstream* antes del comienzo del 3' LTR, tiene como función iniciar la síntesis de la cadena positiva del ADNrtn (Mak y Kleiman 1997; Telesnitsky y Goff 1997). El ADNrtn es luego importado al núcleo, donde es integrado al genoma mediante la acción de la IN. Para esto, la IN reconoce secuencias al final de los LTR (8 - 20 pb), siendo fundamental para el funcionamiento de la enzima que los LTR finalicen en 5' -CA-3' (Hindmarsh y Leis 1999; Zhou et al. 2001). La IN cataliza la unión de los repetidos terminales del ADNrtn a un corte (escalonado) en el ADN genómico. Los *gaps* que se generan en este proceso son reparados por la maquinaria celular, lo cual resulta en la duplicación del sitio de corte. Eso provoca que los LTR se encuentren flanqueados por secuencias duplicadas denominadas TSD (*target site duplication*; Hindmarsh y Leis 1999; Ballandras-Colas et al. 2013).

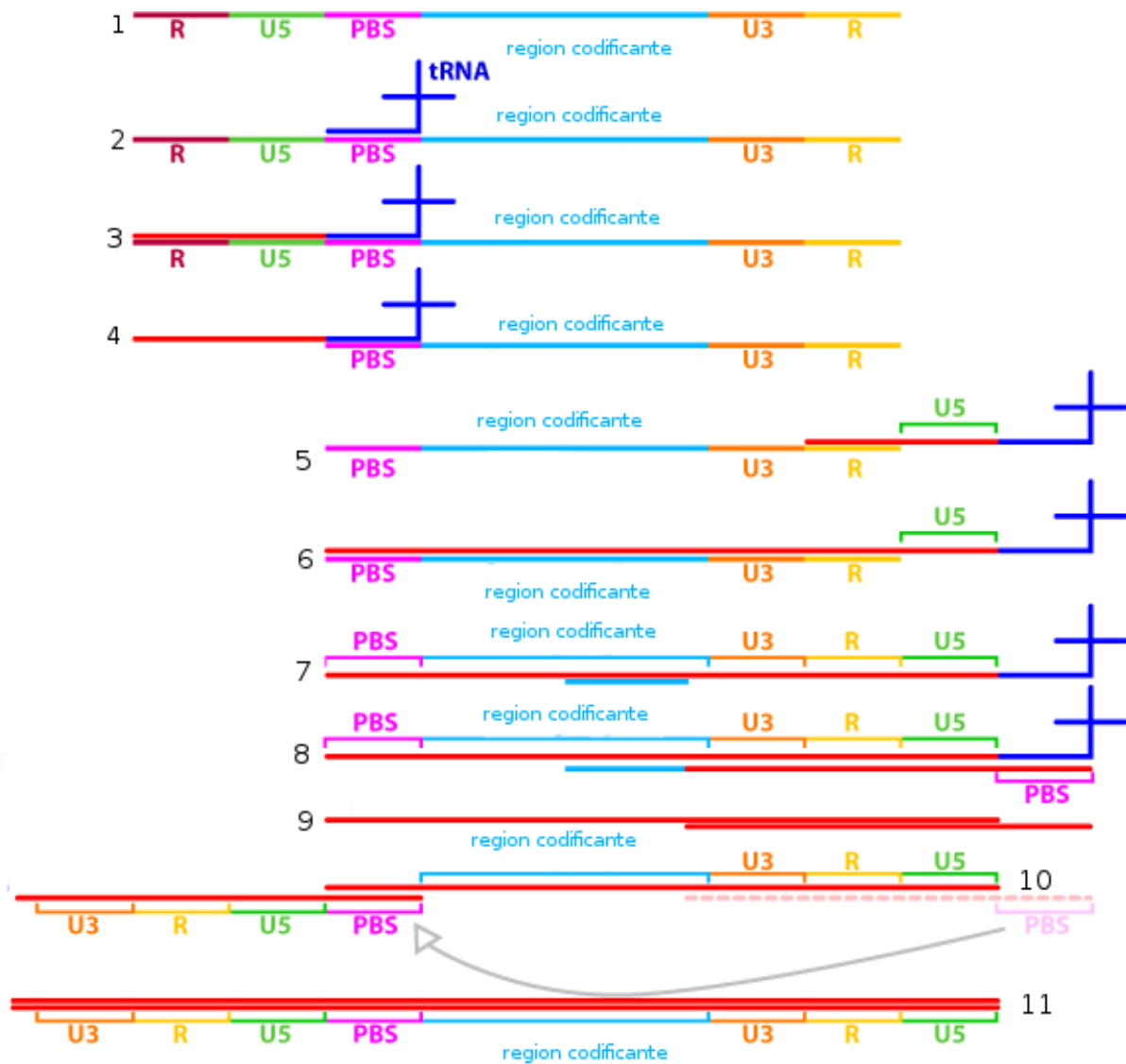


Figura 4. Mecanismo general de retrotranscripción en retrotransposones de LTR. 1 - ARN transcrito a partir de copia original del retrotransposón. Notar ausencia de 5' U3 y 3' U5. 2 - Unión del ARN a la secuencia PBS del intermediario de ARN. 3 - Extensión de molécula de ADN a partir del ARN. 4 - Degradación del ARN en el híbrido ADN-ARN, excluyendo la secuencia PBS. 5 - Primer salto del ADN. Hibridación con la región "R" restante. 6 - ADN asociado al ARN es extendido a partir de su extremo 3'. 7 - Degradación de la mayor parte del híbrido. 8 - Síntesis de la molécula de ADN hacia el extremo 3'. 9 - Degradación total del ARN incluido el ARNt. 10 - Segundo salto del ADN. Hibridación con la secuencia PBS. 11 - Se completa la síntesis de ADN hacia el extremo 3'. Adaptada de studentreader.com.

Además de estos retrotransposones funcionales, se han descrito otros elementos que carecen de todas o casi todas sus secuencias codificantes, pero que aún pueden ser movilizados por la maquinaria funcional de elementos autónomos. Hasta la fecha existen solo dos tipos reportados de retrotransposones de LTR no

autónomos: los largos derivados de retrotransposones (LARDs, *large retrotransposon derivatives*) y los retrotransposones en miniatura con repetidos terminales (*TRIMs*). Ambos tipos poseen las secuencias no codificantes fundamentales para transcribirse, LTRs, PBS y PPT. Difieren en el tamaño: mientras que los LARDs poseen LTRs de 4.4-kb y dominios centrales de 3.5 kb, los *TRIMs* poseen tamaños menores a 1000 pb con LTRs de entre 100 y 350 pb. LARDs y *TRIMs* fueron originalmente descritos en plantas (Witte et al. 2001; Jiang, Jordan y Wessler 2002; Kalendar et al. 2008), y solo recientemente se presentó el primer ejemplo de un grupo de *TRIMs* en metazoarios (hormiga cosechadora roja, *Pogonomyrmex barbatus*; (Zhou y Cahan 2012)). Los *TRIMs* están involucrados activamente en la remodelación de los genomas de plantas, afectando promotores, regiones codificantes, y estructuras de exones e intrones. Algunas características únicas de los *TRIMs*, como su pequeño tamaño, inserción preferencial en regiones eucromáticas y número alto de copias, han hecho que sean prometedores marcadores moleculares. Se presume que los *TRIMs* de plantas y hormigas sean el resultado de la evolución reductiva convergente de elementos autónomos, manteniendo únicamente las secuencias mínimas para completar eficientemente el ciclo de retrotransposición (Schulman 2012).

En esta tesina reportamos el descubrimiento de un nuevo grupo de *TRIMs* que son específicos de los cestodos ténidos. Esta familia de *TRIMs* se ha expandido y divergido durante la evolución de los ténidos. Aunque en *Echinococcus* estos elementos parecen estar inactivos, debido a su capacidad de generar alteraciones en los patrones de expresión génica en los sectores donde se insertan (ya sea interrumpiendo regiones reguladoras de genes o secuencias génicas; transportando promotores y potenciadores en su secuencia; o alterando la expresión de genes cercanos a los lugares dónde se insertan), proponemos que estos elementos pueden haber tenido gran influencia en el rearrreglo del genoma y transcriptoma del genoma del hospedero taenido. A su vez, en colaboración con el Dr. Koziol, reportamos también cómo una fracción de estos *TRIMs* han escapado al silenciamiento y son expresados masivamente en las células germinales de *E. multilocularis*, constituyendo hoy en día el mejor marcador molecular para este tipo celular.

### >Materiales, métodos y estrategia:

Los *script in house* que se utilizaron en este trabajo fueron realizados en diversos lenguajes de programación. Si bien la mayor parte de ellos están escritos en Python, también se programaron scripts en Bash y R. Los scripts se encuentran disponibles en el siguiente link: [Scripts](#).

### Obtención de ESTs (*Expressed sequence tags*) y ensamblados genómicos.

A partir del banco de datos GeneDB (<http://www.genedb.org/>) se descargaron las secuencias genómicas y sus correspondientes anotaciones de *Echinococcus multilocularis*, *Echinococcus granulosus*, *Taenia Solium*, *Hymenolepis microstoma* (Tsai et al. 2013), y *Schistosoma mansoni* (Berriman et al. 2009). A su vez utilizamos los genomas en borrador de varios organismos: *Taenia asiatica*, *Taenia taeniaeformis*, *Mesocestoides corti* y *Diphyllobothrium latum*, los cuales fueron generados por el grupo de genómica de parásitos de la Wellcome Trust en el contexto de la iniciativa de los 50 genomas de helmintos. Estos ensamblajes genómicos están disponibles en <ftp://ftp.sanger.ac.uk/pub/project/pathogens/HGI>. El genoma en borrador de la planaria *Schmidtea mediterranea* fue obtenido de SmedGD, [smedgd.neuro.utah.edu](http://smedgd.neuro.utah.edu) (Robb et al. 2008). Los ESTs de las especies *Echinococcus* fueron descargados de PartiGeneDB ([www.partigenedb.org/](http://www.partigenedb.org/)), GeneDB y del *Echinococcus* Full-Length cDNA project ([http://fullmal.hgc.jp/index\\_em\\_ajax.html](http://fullmal.hgc.jp/index_em_ajax.html)).

### Búsqueda de TRIMs en *E. multilocualris* y en otros platelmintos.

El alineamiento de nucleótidos o aminoácidos es una comparación de estructuras primarias que tiene como finalidad encontrar regiones homólogas que puedan ayudarnos a comprender sus relaciones biológicas, ya sean estructurales, evolutivas, y/o funcionales. Dentro del alineamiento de secuencias podemos definir claramente dos tipos, los globales y los locales. Los alineamientos locales, a diferencia de los globales, buscan encontrar similitud en regiones de las secuencias sin importar las secuencias adyacentes. BLAST (*Basic Local Alignment Search Tool*; Camacho et al. 2008) es el programa que elegimos para mapear las secuencias de

ESTs a los genomas correspondientes. A lo largo de esta tesina, utilizamos varias versiones del programa, que ejecutamos localmente.

Inicialmente, utilizamos el *software* BLASTN para buscar regiones homólogas a los ARN largos no codificantes (ARNInc), previamente descritos por Parkinson et al. en el transcriptoma de *Echinococcus granulosus* (*Cluster A*), en los ESTs de la especie *E. multilocularis* (Parkinson et al. 2012). El programa fue ejecutado utilizando un valor *e* de  $1E^{-10}$  y el resto de los parámetros por defecto. Los ESTs que presentaron homología con elementos del *Cluster A* fueron tomados como *query* en el algoritmo utilizado por BLASTN, para la búsqueda de *TRIMs* en el genoma de *E. multilocularis*. Debido a la característica heurística del algoritmo, y a la naturaleza repetitiva de los elementos, en cada ejecución esperábamos encontrarnos con posibles *TRIMs* o fragmentos de éstos no identificados previamente. Luego de cada búsqueda, el resultado obtenido se comparó con los de las ejecuciones anteriores, de tal forma de seleccionar los mejores *hits* (mayor *score*) en el total de las búsquedas. Esto se logró gracias a la utilización de un *script in house* cuyo fin es chequear si los resultados de las ejecuciones se solapan y, de ser así, quedarse con el *hit* de mayor *score*; siempre y cuando el tamaño del *hit* sea mayor a 150 pares de base. La estrategia anterior dio como resultado una gran variedad de fragmentos tipo *TRIMs*. Por lo tanto y con el objetivo de identificar *TRIMs* completos decidimos buscar regiones del genoma en donde se hallaran dos LTRs próximos. Para esto realizamos un BLASTN (mismos parámetros) sobre el genoma de *E. multilocularis* tomando como *query* los fragmentos de ESTs, obtenidos previamente, que presentaron similitud con repetidos EgBrep (LTR) descritos por Marin et al (Marin et al. 1993). A partir de los resultados obtenidos, se generó una lista de posibles LTRs la cual fue filtrada, utilizando un *script in house*, con el fin de obtener solo los que tuvieran un par de LTR en un rango de distancia de 500 hasta 1200 bases. Logrando así que el tamaño final del elemento concordara con lo descrito por Parkinson et al (Parkinson et al. 2012). La lista de LTRs cercanos se comparó con la lista de fragmentos de *TRIMs*, ya descrita, resultando en la generación de un primer set de *TRIMs* (*TRIMs A*, Figura 5), los cuales fueron curados manualmente.

Con el fin de mejorar y expandir aún más los resultados previamente obtenidos construimos modelos ocultos de Markov (HMM) a partir de las secuencias de *TRIMs* curadas manualmente. Se utilizó el software hmmer2.3, con parámetros

por defecto ([www.hmmmer.org](http://www.hmmmer.org)). Los términos de HMM-search fueron optimizados para el genoma y la calidad de modelos de HMM. Las predicciones de HMM (*TRIMs* B, ver Figura 5) no fueron el final del análisis, sino el punto de inicio de un cuidadoso curado manual de todas las secuencias tipo *TRIMs* identificadas con ambas estrategias. Los resultados fueron visualizados en el *software* Artemis (Carver et al. 2012), a partir del cual se realizó la anotación manual, identificando en cada uno de los fragmentos genómicos con similitud a ESTs del *Cluster A* los componentes característicos de los *TRIMs*. La estrategia se representa en la Figura 5.

El curado manual es la mejor manera de identificar y caracterizar los retrotransposones que pertenecen a familias de *TRIMs*. Este paso laborioso es indispensable para el correcto análisis de este tipo de elementos.

Debido al origen común de los elementos *TRIMs* de los ténidos, es posible utilizar las secuencias manualmente curadas para *E. multilocularis* como *query* en la búsqueda de secuencias homólogas tipo *TRIMs* en el resto de las especies mencionadas. La búsqueda se realizó a través de BLASTN, utilizando los parámetros previamente mencionados (Figura 5). Una lista de elementos completos de cada especie fue extraída y manualmente curada con el fin de usarlos en diferentes análisis a lo largo del trabajo (material suplementario 1). A su vez, se generó una lista independiente con todas aquellas secuencias mayores a 800 pares de bases a fin de usarlas en estudios filogenéticos.

Además de lo descrito anteriormente, se utilizó el LTR FINDER (Zao y Hao 2007) que es un *software* específico de búsqueda de transposones para identificar y anotar *TRIMs*. Se utilizaron los parámetros por defecto con las siguientes excepciones: el tamaño mínimo esperado para LTRs fue de 150 pb y la distancia mínima entre ellos fue 500 pb. Se utilizó el <sup>Leu</sup>ARNt como *query* para la identificación de PBS. Las secuencias encontradas fueron chequeadas para confirmar la presencia de *TRIMs*. La mayor parte de las secuencias identificadas con esta herramienta correspondieron a falsos positivos; el resto habían sido identificadas y curadas como *TRIMs* con las estrategias descritas previamente.

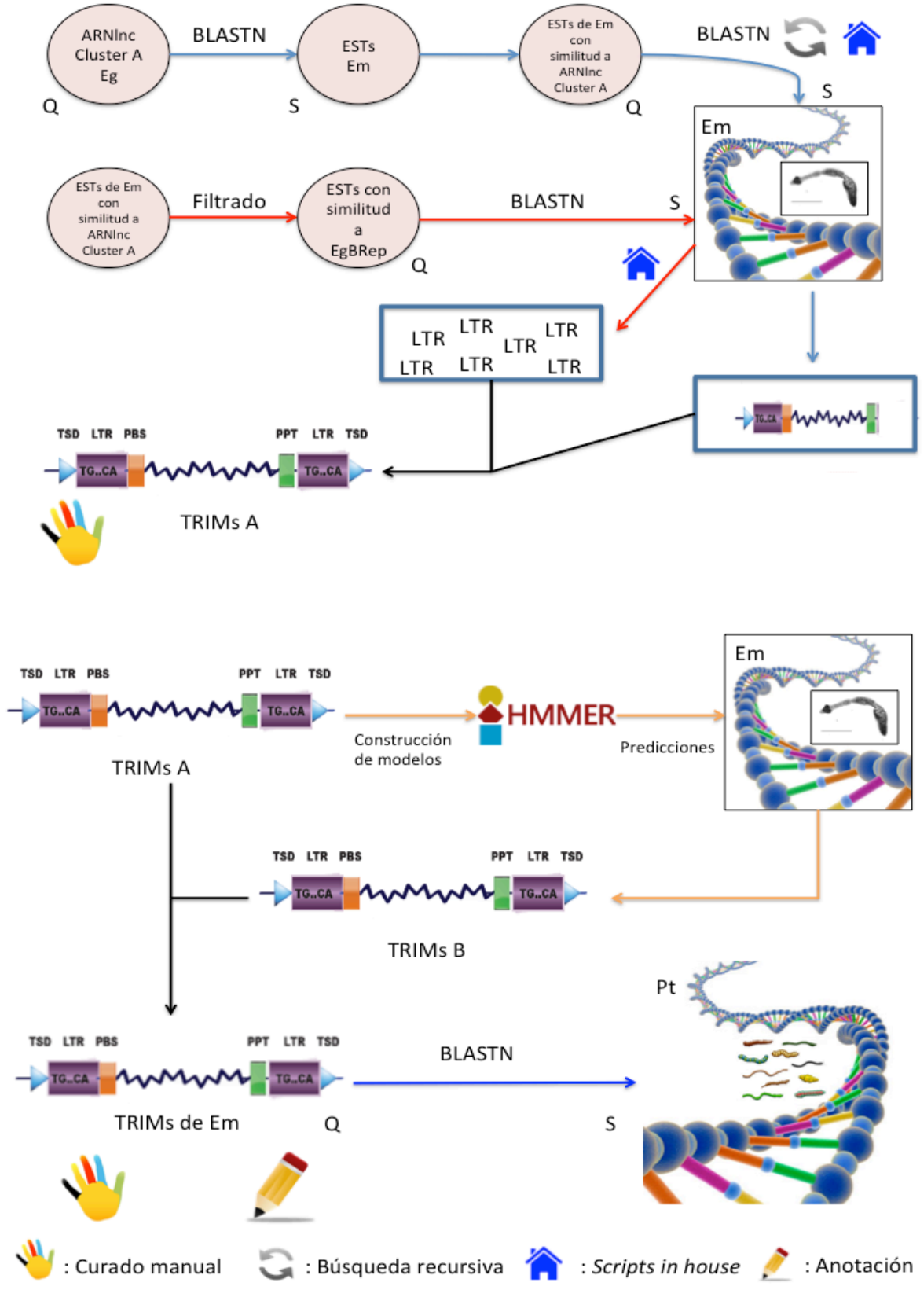


Figura 5. Estrategia de búsqueda de TRIMs en *E. multilocularis* y otros platelmintos. La estrategia se explica con detalle en el texto. Q: query; S: subject; Em: *E. multilocularis*; Pt: platelmintos.



### **Identificación del Sitio de Unión al Primer (PBS) de *TRIM*.**

Debido a que el mecanismo de retrotransposición implica la utilización de un ARNt con secuencia complementaria al PBS, generamos una lista de genes de ARNt de *E. multilocularis* y *E. granulosus* a través del uso del tRNAscan-SE (Lowe y Eddy, 1997). Esto nos permitió identificar cuál de ellos es complementario al PBS de los *TRIMs* identificados en el trabajo.

### **Caracterización de estructura secundaria de *TRIMs*.**

La estructura secundaria del transcrito fue predicha a través de la utilización del *software* WAR (Torarinsson, 2008). Este *software* se basa en encontrar la estructura secundaria consenso entre una variedad de *softwares* que predicen estructuras secundarias. La estructura consenso ideal es aquella de menor energía libre, por lo que el programa se optimiza seleccionando aquellos programas que mejor se adapten a la secuencia dada por el usuario. Para identificar la estructura secundaria en las regiones transcritas de copias completas de *TRIMs* de *E. granulosus* utilizamos los siguientes programas: CMfinder (Yao et al. 2006); FoldalignM (Torarinsson et al. 2007), ClustalW + RNAalifold (Thompson et al. 1994 y Bernhart et al. 2008); LaRA (Bauer et al. 2007) y RNASampler (Xu et al. 2007).

### **Distribución de *TRIMs* e identificación de sitios preferenciales de inserción.**

Se utilizó un *script in house* para analizar la distribución de *TRIMs* en el genoma de *E. multilocularis*. En el análisis se utilizaron todos los fragmentos de *TRIMs* identificados en *E. multilocularis*. Con el fin de evaluar la distribución de los *TRIMs*, el *script* recorre la secuencia nucleotídica de los *scaffolds* genómicos a través de la generación de ventanas deslizantes no solapantes de 10, 50 o 100 kb; dependiendo de los parámetros utilizados. Posteriormente, se comprueba a través de un test de *Student* si el número de *TRIMs* en una ventana dada de un *scaffold* presenta una diferencia significativa al resto de las ventanas del correspondiente *scaffold*. También se compara el número de elementos presentes *scaffold vs scaffold* (normalizados por tamaño). Aquellos casos que presentaron una diferencia significativa fueron analizados manualmente.

Se realizó otro *script* para indagar si a nivel de composición de bases existen sitios preferenciales de inserción. El *script* obtiene las secuencias lindantes (10, 50, 100 pares de bases) a los *TRIMs* completos de *E. multilocularis*, calcula su composición en GC y las compara con la composición en GC del mismo número de secuencias genómicas obtenidas aleatoriamente en regiones intergénicas. La significación estadística se prueba a través de test *t* de *Student*.

### **Análisis filogenéticos.**

Utilizando el software ClustalW (Thompson et al. 1994) se realizó un alineamiento de todos los elementos completos o fragmentos mayores a 800 pares de bases de *TRIM* de todas las especies en las cuales se identificaron, eliminándose la región correspondiente al 5' LTR (ya que no es independiente de la región 3' LTR). La historia evolutiva fue inferida utilizando el método de máxima verosimilitud basado en el modelo de Tamura-Nei. El árbol consenso fue inferido a través de 500 réplicas. Aquellas ramas reproducidas en menos de un 50% de los *bootstrap* fueron colapsadas (Felsenstein 1993). El árbol inicial para la búsqueda heurística fue inferido automáticamente aplicando los algoritmos Neighbor-Join y BioNJ a una matriz de distancia obtenida mediante el método Maximum Composite Likelihood (MCL). El análisis incluyó 130 secuencias nucleotídicas representando a 6 especies (*E. multilocularis*, *E. granulosus*, *T. saginata*, *T. asiatica*, *T. taeniaeformis* y *T. solium*). Los análisis filogenéticos de máxima verosimilitud fueron realizados con el software MEGA 5.0 (Tamura et al. 2011).

### **Comparación de *loci* de *TRIMs* entre especies de Ténidos.**

Para evidenciar si un determinado *TRIM* fue insertado antes o después de la divergencia de especies, se obtuvo de cada elemento completo la secuencia lindante 2 kb *upstream* y *downstream* del 5' y 3' respectivamente. Posteriormente, se analizó manualmente la presencia o no de un elemento dado en el genoma del otro. La estrategia se utilizó para las especies de *Echinococcus* y para *T. solium* y *T. asiatica*.

### **Búsqueda de retrotransposones de LTRs autónomos.**

La presencia de retrotransposones de LTRs en *E. multilocularis* fue indagada mediante la búsqueda por TBLASTN de secuencias aminoacídicas correspondientes a proteínas *Pol* de *Drosophila melanogaster* y *Schistosoma mansoni* (Laha et al. 2001; DeMarco et al 2004; Laha et al. 2004). No pudimos identificar ninguna copia intacta de proteínas tipo *Pol*. Sin embargo, entre las secuencias con mayores fragmentos de *Pol*, identificamos copias de retrotransposones LTRs previamente no identificados. A partir de estos elementos se realizó una estrategia análoga a la búsqueda de *TRIMs*, la cual resultó en la identificación manual de LTRs, PBS, y PPTs en todos los retrotransposones de LTR que presentaron características tipo *TRIM* en los genomas de las especies estudiadas.

### **Construcción de Logos.**

Para examinar la composición nucleotídica de las secuencias características de *TRIMs* y de su par autónomo (LTRs, PBS, PPT y TSD), analizamos copias completas de estos elementos de *E. granulosus* y *E. multilocularis* y copias completas de *TRIMs* de *T. solium*. Las similitudes y diferencias dentro de alineamientos múltiples de secuencias se representaron mediante un logo generado con WebLogo 3.4 (Crooks et al. 2004).

### **Análisis de datos de expresión.**

Con el objetivo de indagar en el comportamiento transcripcional de los elementos tipo *TRIMs*, ESTs de *E. multilocularis* con similitud a *TRIMs* fueron mapeados al genoma de este parásito a través de BLASTN (valor de  $e$  de  $1E-10$ ). Se decidió eliminar todos los *hits* con menos de un 99% de identidad al genoma o menores a 50 pb, así como también aquellos que mapearon en más de un *loci* con >99% de identidad. Si bien estos requerimientos disminuyen el número de posibles elementos transcritos, dan mayor confiabilidad a los elementos seleccionados. Los *loci* resultantes fueron analizados manualmente para determinar si eran copias completas de *TRIMs*, fragmentos o derivados de LTRs, si la transcripción se iniciaba

dentro del LTR y si existía acople transcripcional con regiones intergénicas o regiones codificantes de genes vecinos. Un análisis idéntico se realizó para *E. granulosus*, excepto que se utilizó como límite 95% de identidad, debido a la menor calidad del genoma ensamblado.

Por otro lado, y también para determinar la actividad transcripcional de *TRIMs*, los datos de transcriptómica obtenidos por Tsai et al. (Tsai et al. 2013) fueron mapeados al genoma de referencia utilizando Tophat 2.0.6 (Trapnell et al. 2012). Los parámetros utilizados fueron -r 300 -mate-std-dev 100 -i 10 -l 4000 -g 40 -a 4, seteados utilizando el conocimiento previo del genoma y la librería de RNA-Seq. Luego de eliminar lecturas (*reads*) duplicadas, *reads* multimapeados, y *reads* con una calidad de mapeo menor a 30; los *reads* por elemento fueron calculados utilizando featureCount del paquete Rsubreads (Liao et al. 2014). El filtrado de lecturas fue realizado por SAMtools v.0.1.19+ (Li et al. 2009) y *scripts* locales (esto último fue realizado por Magdalena Zarowiecki en el Instituto Sanger).

#### **Asociación de *TRIMs* con genes.**

Mediante un *script in house* se obtuvo una lista de genes que se ubican a una distancia máxima de 500 pares de bases *downstream* del extremo 3' de cada LTR aislado. Para comprobar si los genes asociados a los *TRIMs* poseían relaciones funcionales se realizaron estudios de sobrerrepresentación de términos de ontología génica mediante la herramienta BLAST2GO (Conesa et al. 2005).

## >Resultados

### Descubrimiento de elementos *TRIM* en cestodos ténidos.

#### Descripción de elementos transponibles en *Echinococcus*.

El análisis del transcriptoma de *E. granulosus* mostró que los ESTs más abundantes en varios estadios larvarios, corresponden a ARNinc de aproximadamente 900 pares de bases, los cuales también fueron identificados en *E. multilocularis* y *T. solium*. Estas secuencias fueron ensambladas en contigs que se agruparon en dos clusters (A y B) de acuerdo a su similitud. Estructuralmente, estos elementos presentan dos repetidos directos en los extremos y una región no codificante central (Parkinson et al. 2012). Los repetidos terminales presentaron homología con el elemento repetido EgBRep descrito previamente en *E. granulosus* (Marin et al. 1993). Mediante el mapeo de estos ESTs de *E. multilocularis* y *E. granulosus* a las secuencias genómicas recientemente publicadas (Tsai et al. 2013), hemos encontrado que estos ARNinc son transcritos de secuencias genómicas que presentan todas las características de elementos transponibles; estructura, distribución genómica, abundancia, etc (ver la sección de Materiales, métodos y estrategias para una descripción detallada de la búsqueda de estos elementos en el genoma; Figura 5). En particular, por su pequeño tamaño, la falta de regiones codificantes y los repetidos terminales, pensamos que las secuencias genómicas de las cuales surgen los transcritos previamente mencionados son retrotransposones no autónomos tipo *TRIMs*. Como los retrotransposones tipo *TRIMs* tienen características conservadas (Zhou y Cahan 2012), nos propusimos observar si esas características estaban presentes en nuestras secuencias. Un análisis manual nos permitió observar que nuestras secuencias (Figura 6) poseen dos LTRs idénticos de aproximadamente 198 pares de bases que comienzan en 5' TG y terminan en 3' CA. Estos dirigen, en los retrotransposones, el inicio de la transcripción (5' LTR), su terminación y poliadenilación (3' LTR). También presentan un tracto polipurínico (PPT), de 15 pares de bases, inmediatamente adyacente al 3' LTR que colabora con el mecanismo de transcripción reversa; así como un sitio de unión al *primer* (PBS), complementario a un <sup>LEU</sup>ARNt y adyacente al 5' LTR, en donde comienza la transcripción reversa. Entre ambos LTRs, hay aproximadamente 590 pares de

bases que carecen de cualquier tipo de marco abierto de lectura mayor a 120 codones y, mediante BLASTX (parámetros por defecto), no muestran ninguna similitud con ninguna proteína conocida. Dado que la región central no presenta conservación a nivel primario pero sí a nivel secundario, estudiamos si contiene la señal de empaquetamiento (packaging signal, PSI), la cual es vital para el empaquetamiento de los retrovirus de ARN con proteínas Gag (Wicker et al. 2007), y se cree que es importante para el mecanismo de retrotransposición. Mediante el uso de software WAR, vimos que existiría esta estructura secundaria conservada, lo cual apoyaría la hipótesis de conservación de la señal PSI (material suplementario 2). Las secuencias poseen también repetidos directos cortos (4 a 6 pares de bases) que flanquean el exterior de los LTRs y son generados en el proceso de integración (TSD). Todas estas características indican fuertemente que los transcritos surgen de regiones genómicas con características de retrotransposones no autónomos que pueden ser movilizados por un elemento autónomo en *trans*.

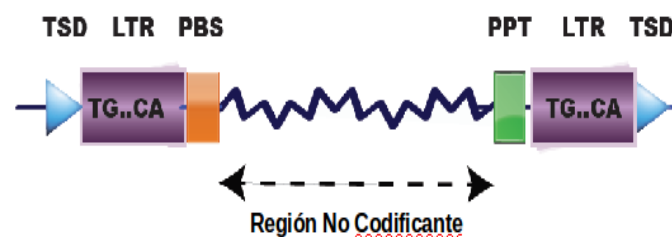


Figura 6. Estructura de un retrotransposón genérico de la familia *TRIM*. PBS, sitio de unión al *primer*; PPT, sitio polipurínico; TSD, sitio blanco de duplicación.

A partir de ESTs *full length* de ambas especies de *Echinococcus*, y de datos de RNA-Seq (confirmados por experimentos de 3' RACE en *E. multilocularis*; ver [Apéndice](#); Koziol, Radío et al. 2015) se pudo identificar claramente el sitio de inicio de la transcripción dentro del 5' LTR y el sitio de poliadenilación y terminación dentro del 3' LTR (Figura 6 y 7). La poliadenilación de *TRIMs* de diferentes loci ocurre dentro de una ventana de 6 pb en el extremo del 3' LTR, en el cual no se puede encontrar ninguna señal canónica de poliadenilación (AATAAA). Sin embargo, hay una región rica en A/T que contiene una secuencia de poliadenilación previamente propuesta para especies de *Echinococcus*; AATATA (Konrad et al. 2003; Koziol et al. 2009) localizada 12 a 15 pares de base *upstream* del sitio de escisión y

poliadenilación. De esta manera, el LTR puede ser dividido en regiones U3 y U5 (Figura 7).

Los resultados descritos anteriormente nos llevaron a concluir que los transcriptos analizados surgen de retrotransposones no autónomos. En particular, de una familia perteneciente a la clase II de ETs, denominada *TRIMs*. Esta familia está ampliamente distribuida en plantas y se ha descrito recientemente en metazoarios (hormiga cosechadora roja, *Pogonomyrmex barbatus* (Zhou y Cahan 2012)). Los *TRIMs* son un tipo de retrotransposones de LTR no autónomos; esta es la segunda vez que se evidencia su existencia en metazoarios y la primera que se los describe en parásitos.

#### **Búsqueda de *TRIMs* en organismos relacionados.**

Con el fin de entender un poco más de la biología de los *TRIMs* decidimos buscarlos en otras especies. Los buscamos en genomas borrador de otros cestodos ténidos (*Taenia solium*, *Taenia asiatica*, *Taenia taeniaeformis* y *Taenia saginata*), así como en especies relacionadas como son otros cestodos no ténidos y otros platelmintos (*Hymenolepis microstoma*, *Mesocestoides corti*, *Diphyllobothrium latum*, *Schistosoma mansoni*, *Fasciola hepatica* y *Schmidtea mediterranea*). Pudimos identificar *TRIMs* en todas las especies de ténidos, los cuales presentan al menos un 50% de identidad en los alineamientos (Figura 7). Además de compartir un 50% de identidad comparten características claves de los retrotransposones no autónomos de la familia *TRIMs* como son una alta homología en la secuencia y/o posición de LTRs, PBS y PPTs (Figura 6 y 7). Por lo general, fue posible identificar TSD flanqueando estos elementos en todos los genomas en los que fueron identificados (material suplementario 3-5). Por otro lado, no pudimos identificar *TRIMs* en grupos no ténidos ya sea en los genomas borrador o en el GenBank. Debido a que estos *TRIMs* únicamente están presentes en los principales linajes de los cestodos ténidos (Nakao et al. 2013), podemos inferir que surgieron en la base del árbol filogenético de los ténidos y que son específicos de esta familia. Por esta razón, hemos denominado a esta familia de *TRIMs*, *ta-TRIMs*.

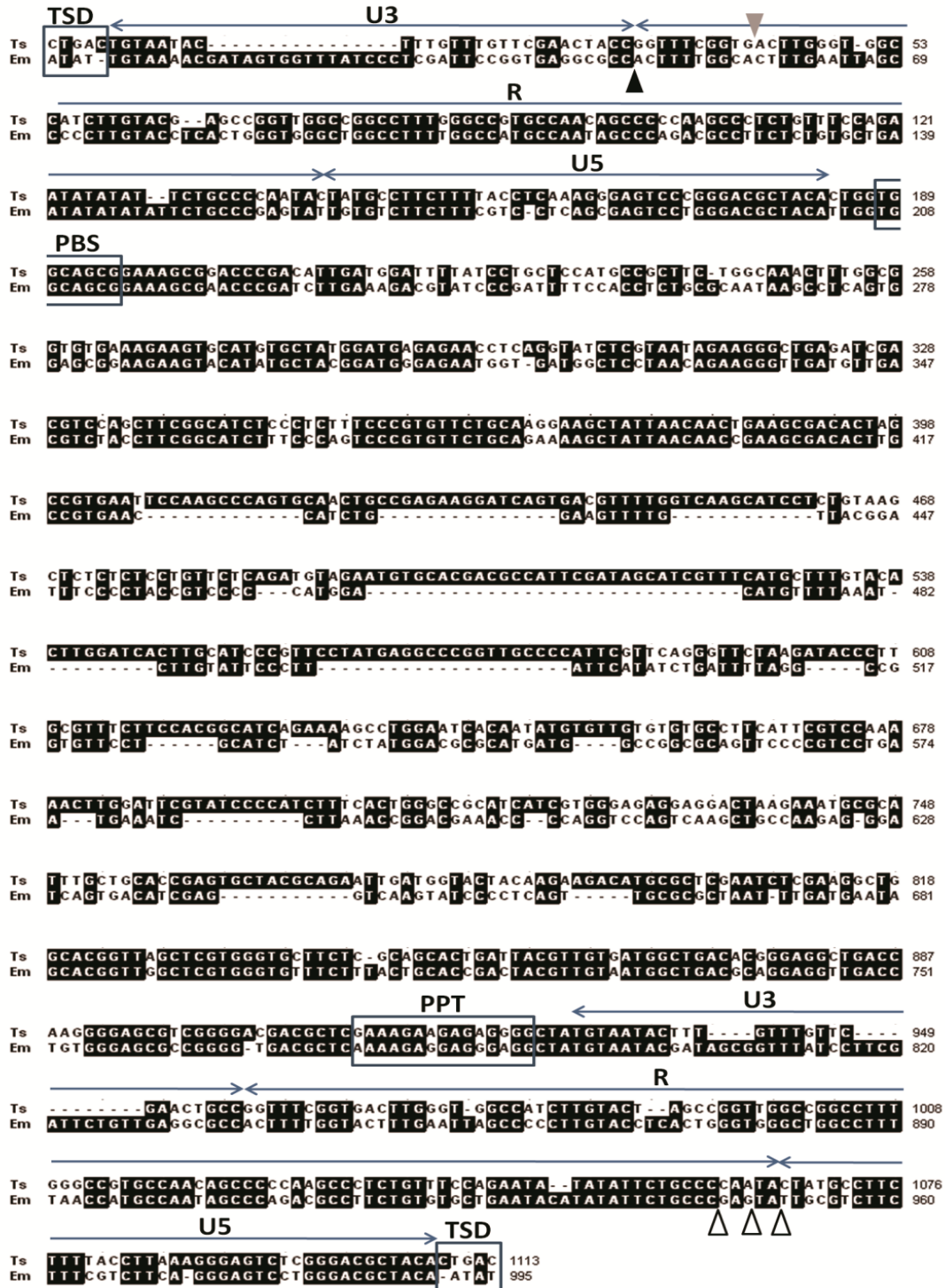


Figura 7. Alineamiento de *ta-TRIMs* de *Echinococcus multilocularis* (Em) y *Taenia solium* (Ts). El límite de las regiones U3, R y U5 de los LTRs está indicado arriba del alineamiento. Las cabezas de las flechas en el 5' LTR muestran el inicio de la transcripción en elementos pertenecientes a Em (negro), y Ts (gris), determinados a partir de ESTs *full length*. Las cabezas de flecha blancas en el 3' LTR muestran el extremo 3' de los transcritos *ta-TRIMs* determinados a partir de ESTs *full length* y experimentos de 3' RACE. PBS, sitio de unión al primer; PPT, sitio polipurínico; TSD, sitio blanco de duplicación.



### Clasificación de los *ta-TRIMs* según su conservación.

Con el objetivo de realizar una descripción más detallada de los *ta-TRIMs*, decidimos clasificarlos en elementos perfectos (dos LTRs que comienzan en 5' TG y terminan en 3'CA, PBS y PPT sin *mismatches*), copia completa (*full length*, contienen dos LTRs reconocibles, PBS y/o PPT (pueden contener *mismatches*)) y elementos parciales. Existen 19 elementos perfectos y por lo menos 123 elementos *full length* en *E. multilocularis*. Si bien el mecanismo de transposición de estos elementos no autónomos no se comprende del todo, es de suponer que los elementos clasificados como perfectos tienen todas las características necesarias como para ser movilizados. En *E. granulosus*, se obtuvieron 103 elementos de copia completa siendo 25 de ellos perfectos. La menor cantidad de elementos tipo *ta-TRIMs* encontrados en *E. granulosus* puede deberse en parte a la peor calidad del ensamblado. Adicionalmente, 1939 *loci* presentaron homología mediante BLASTN y búsquedas mediante motivos HMM a las regiones LTR de *E. multilocularis* mientras que fueron 1183 lo hicieron en el genoma de *E. granulosus*.

Muchos de los *ta-TRIMs* parciales que hemos identificado, son secuencias de LTRs aisladas a las que denominamos solo-LTR. Estos elementos son un derivado común de los retrotransposones de LTR y se piensa que se originan por recombinación no conservativa entre LTR de un mismo elemento. Consistente con este origen es el hecho que muchos LTRs se encuentran flanqueados por TSDs. En el genoma encontramos una enorme cantidad de LTRs flanqueados por TSD, lo que puede indicar un mecanismo de defensa del genoma hospedero. Un ejemplo previamente conocido de solo-LTR es el elemento EgBRep descrito en *E. granulosus*, el cual se encuentra embebido en una secuencia repetida y carece de TSDs reconocibles (Marin et al. 1993).

Se encontraron resultados similares en el genoma borrador de *T. solium*. Sin embargo, el número total de copias es difícil de estimar debido a la naturaleza fragmentada del ensamblaje genómico disponible (material suplementario 5). Al menos 827 *loci* mostraron similitud significativa con *ta-TRIMs*, entre los que se incluyen 24 elementos *full length*, de los cuales 5 son perfectos. En los borradores provisionarios no publicados de *T. asiatica* y *T. taeniaeformis* solo hicimos un análisis superficial. En *T. asiatica* pudimos encontrar 3 elementos perfectos, sin embargo en

*T. taeniaeformis* no pudimos encontrar elementos perfectos, pero sí algunos elementos casi completos que resultaron ser divergentes y poseían modificaciones en lugares claves (sugiriendo que son pseudo-elementos).

Aunque estos resultados pueden cambiar a medida que se disponga de nuevas versiones del genoma, en el caso de *E. multilocularis* creemos que los resultados no van a variar mucho ya que hoy en día se cuenta con una versión bastante curada (Tsai et al. 2013).

#### **Análisis de localización en el genoma.**

Nos preguntamos si la distribución de los *ta-TRIMs* tiene algún tipo de sesgo particular, ya sea en cuanto a la distribución en el genoma, o con respecto a preferencia en la composición de bases en el sitio de inserción. Por un lado, mediante análisis estadísticos (test de *t Student*) verificamos que ningún *scaffold* tiene significativamente más *ta-TRIMs* que otro. Concordantemente, no encontramos una concentración fuera del rango normal de *ta-TRIMs* en ninguna región de los genomas (material suplementario 6), con la única excepción de una región del *scaffold* cinco que, como producto de una duplicación genómica, presenta una acumulación de *ta-TRIMs*. Por otra parte, estudios previos han mostrado que los sitios de inserción presentan un enriquecimiento local en A y T (Zhou y Cahan 2012). Para comprobar si este es el caso en nuestros datos se analizaron los sitios de inserción en busca de algún sesgo composicional. Los resultados obtenidos indican que los sitios preferenciales de inserción no son necesariamente ricos en AT, en contraposición con otros estudios de este tipo (Zhou y Cahan 2012). Por lo tanto, concluimos que los *ta-TRIMs* de *E. multilocularis* se encuentran distribuidos en todos los *scaffolds* de manera uniforme sin presentar sitios preferenciales de inserción.

#### **Aspectos evolutivos de *ta-TRIMs*.**

##### **Análisis filogenéticos.**

Para comprobar si los *ta-TRIMs* se encontraban distribuidos en otros organismos no ténidos, se analizó su presencia en numerosos platelmintos (ver

materiales, métodos y estrategia), pudiendo identificarlos únicamente en cestodos ténidos. Debido a la similitud en secuencia que presentan podemos concluir que los *ta-TRIMs* son una familia de *TRIMs* que surgieron en un antepasado de todos los cestodos ténidos.

Los *ta-TRIMs* muestran una alta similitud entre *E. granulosus* y *E. multilocularis* y así también entre *T. solium* y *T. asiatica* (en ambos casos > 90% para las copias más similares), sin embargo si comparamos especies de *Echinococcus* con *Taenia* la similitud mostrada es mucho menor (aproximadamente 50%). Los análisis filogenéticos de los *ta-TRIMs* de todas las especies analizadas muestran que los elementos de las especies de *Echinococcus* forman un clado muy soportado, así como lo hacen las especies *T. asiatica* y *T. solium*, mientras que *T. taeniaeformis* forma otro clado independiente. Además la mayor parte de los elementos de *T. asiatica* forman un clado monofilético (Figura 8A). Esta topología sugiere, como primera hipótesis, una expansión masiva independiente en cada uno de los principales linajes de ténidos. Sin embargo, una posibilidad que se tiene que tener en cuenta para explicar los resultados de la filogenia es la posible conversión génica entre las copias de *ta-TRIMs* en cada una de las especies. Esta posibilidad fue analizada por el grupo utilizando el software GENECONV (Sawyer 1989; Sawyer 1999). Los resultados muestran un bajo nivel de conversión génica por lo que nos definimos por la primera hipótesis.

La separación entre *E. multilocularis* y *E. granulosus* fue un evento bastante reciente que ocurrió entre hace 2.5 y 9.2 millones de años (Knapp et al. 2011). Buscamos evidencia de integración de *ta-TRIMs* después de que ocurriera este fenómeno. Para esto, comparamos secuencias flanqueantes de elementos de copia completa de *E. multilocularis* con la región homóloga de *E. granulosus* (Figura 8B). En tres de las siete comparaciones encontramos fuerte evidencia de integración luego de la divergencia de ambos clados de *Echinococcus*, ya que el *ta-TRIM* correspondiente está ausente en la región equivalente de *E. granulosus*, pero sí se puede encontrar la secuencia TSD perfecta o con un *mismatch* (Figura 8B y material suplementario 7). Utilizando la misma estrategia, buscamos eventos de integración luego de la separación de *T. solium* y *T. asiática*; para cinco regiones que poseen alta homología se encontraron tres eventos de integración luego de la divergencia de estas especies (material suplementario 7). Además, utilizamos una estrategia

similar, pero menos restrictiva, para analizar con mayor profundidad lo sucedido entre *E. multilocularis* y *E. granulosus* (buscando homología en secuencias flanqueantes de 300 pares de bases, sin tener en cuenta la presencia de TSD). De esta manera, identificamos 12 elementos que se insertaron luego de la separación en *E. multilocularis*, y dos que se insertaron únicamente en *E. granulosus*. A su vez, identificamos varios casos de recombinación no conservativa ocurrida luego de la divergencia de ambas especies (se observan solo-LTRs en uno de los genomas y en el otro una copia completa *ta-TRIM*; material suplementario 8).

Como resultado del mecanismo de transcripción reversa de los retrotransposones LTR, ambos LTRs al momento de la inserción son idénticos (Telesnitsky y Goff 1997). Por lo tanto, si consideramos este mecanismo, la identidad nucleotídica entre el 5' y 3' LTR de cada *ta-TRIM* puede ser utilizada para estimar el tiempo de integración (Kijima e Innan 2010). Por esta razón se analizó la divergencia entre 5' y 3' LTRs de copias completas de *ta-TRIMs*. Específicamente, se estudió esta situación en *E. multilocularis*, *T. solium*, *T. taeniaeformis* (Figura 8C). También se analizó la situación en *E. granulosus*; se obtuvo un resultado muy similar al obtenido en *E. multilocularis*. En *E. multilocularis* la divergencia entre valores de LTR fue de 6.1% - 22.9% (n = 21), mientras que en *T. solium* y *T. taeniaeformis* la divergencia fue de 2.2% - 14.9% (n = 25) y 19% - 27% (n = 3), respectivamente. Los resultados obtenidos, asumiendo una tasa de sustitución igual para todos los linajes, implican que las ondas de retrotransposición han ocurrido en diferentes tiempos para cada uno de los linajes (con la excepción del caso de *E. multilocularis* y *E. granulosus*). Para determinar si la diferencia entre la distribución de los valores de divergencia es significativamente diferente se aplicó el test U de Mann Whitney, con la corrección de Bonferroni ( $p < 0.02$ ).

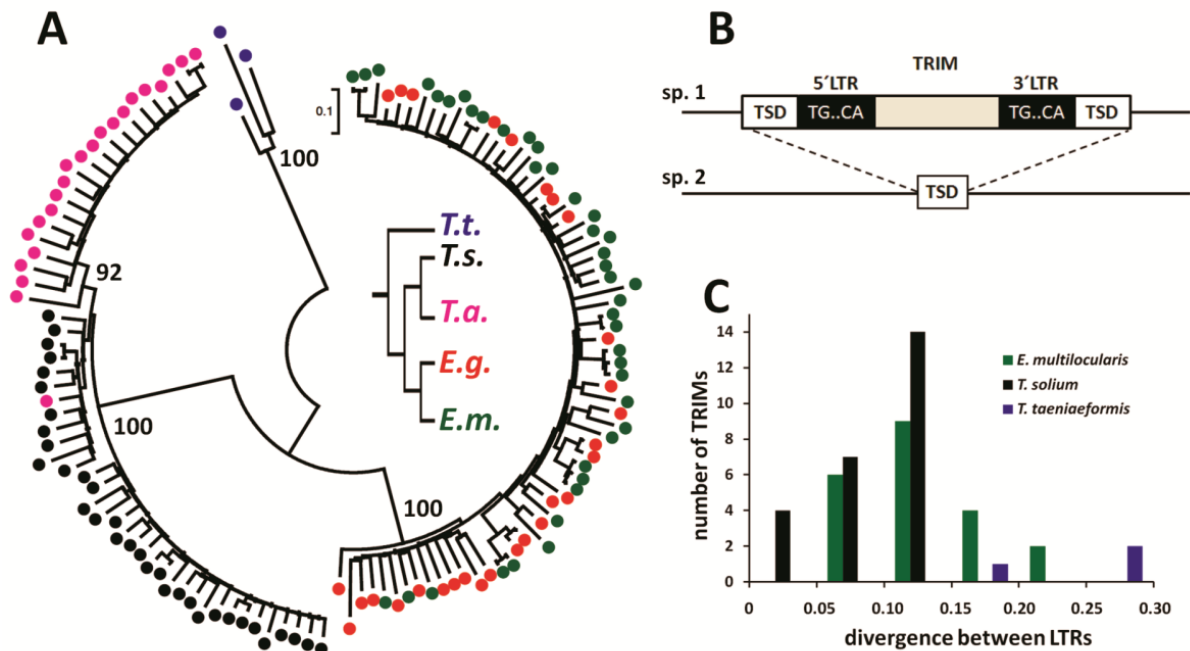


Figura 8. Evolución e inserción de *ta-TRIMs*. A. Árbol filogenético de *ta-TRIMs* de especies de ténidos, inferido por análisis de máxima verosimilitud (modelo Kimura 2-parámetros con distribución gamma para los sitios). Los valores de *bootstrap* (500 réplicas) están indicados cerca de cada nodo. La figura dentro de la filogenia muestra el árbol de ténidos obtenido a partir de los datos de Nakao et al. (Nakao et al. 2013), utilizando el mismo esquema de color que para la filogenia de *ta-TRIMs*. Eg, *E. granulosus*; Em, *E. multilocularis*; Ta, *T. asiatica*; Ts, *T. solium*; Tt, *T. taeniaeformis*. B. Diagrama explicando la identificación de sitios de inserción entre especies cercanas (sp.1 y sp.2). C. Histograma mostrando la divergencia entre 5' y 3' LTRs de *ta-TRIMs* de tres especies de ténidos.

### Un elemento autónomo candidato para la movilización de *ta-TRIMs* en *trans*.

#### *Lennie*, el candidato autónomo.

Hasta ahora, según nuestro conocimiento, ningún retrotransposón ha sido propuesto como responsable de la movilización de *TRIMs* (Schulman 2012). Esto es probablemente debido al largo período que estos elementos han permanecidos inactivados y, por lo tanto, a la baja similitud entre ellos. Con el objetivo de encontrar un elemento autónomo candidato para la movilización de *ta-TRIMs* en *trans*, nos propusimos encontrar retrotransposones autónomos candidatos en *E. multilocularis* (ver Materiales, métodos y estrategia). Encontramos una familia de retrotransposones LTR, que hemos denominado *lennie*, los cuales poseen características que sugieren que pudieron haber cumplido el rol de movilizar a los elementos *TRIMs* de forma autónoma (Figura 9). Existen regiones de estructura primaria dentro de un retrotransposón tipo LTR no autónomo que son claves para

que el elemento no autónomo pueda ser movilizado en *trans* por la maquinaria de retrotransposición de su par autónomo. Particularmente, las primeras 8 bases del LTR, las cuales son cruciales para la interacción con la IN (Zhou et al. 2001), son idénticas entre los *ta-TRIMs* y *lennies*, permitiendo la correcta interacción de la IN de los *lennies* con los elementos *ta-TRIMs*. Además, debido a que los ARNt afines se empaquetan específicamente en VLPs durante el ciclo de retrotransposición (Boeke y Stoye 1997; Mak y Kleiman 1997), cualquier elemento autónomo que pretenda movilizar a los *ta-TRIMs* debe poseer el mismo PBS. *lennie* tiene un PBS de 8 pares de bases que es complementario al mismo <sup>Leu</sup>ARNt que los *ta-TRIMs* y además se encuentra posicionado a la misma distancia del 5' LTR. A su vez otra característica que deben compartir ambos elementos es el PPT. El PPT se conserva en cuanto al número de nucleótidos y a su posición relativa al 3' LTR, además de poseer una secuencia muy similar, como se evidencia en la Figura 9.

Además del uso de las primeras 8 bases del LTR, la longitud del TSD generado en el ciclo de retrotransposición es característica de cada grupo de IN (Wu et al. 2005; Ballandras-Colas et al. 2013). La mayor parte de los TSDs analizados manualmente provenientes de *ta-TRIMs* y sus derivados solo-LTRs de *E. multilocularis* son de 4 pb de largo, mientras que los restantes son de 5 pb (52/57 casos analizados poseen 4 pb). Este resultado es consistente con los obtenidos para los *lennies* y sus derivados solo-LTRs en donde 63 de 65 TSDs analizados presentaron una secuencia de 4 pb, mientras que 2 presentaron 5 pb. La proporción de TSD de 4 y 5 pb no mostró diferencias significativas entre *lennies* y *ta-TRIMs* (test exacto de Fischer;  $p > 0.1$ ). A pesar de que pudimos observar expresión de los elementos *lennies* a partir de ESTs y RNA-Seq, (solo dentro de la región LTR), no encontramos copias intactas de *lennies* en *E. multilocularis* (todos poseen al menos un corrimiento del marco abierto de lectura dentro de pol). Esto sugiere que, si bien los *lennies* son elementos transcripcionalmente activos, es poco probable que puedan cumplir un ciclo de retrotransposición. Si los *lennies* efectivamente fueran los responsables de la movilización de *ta-TRIMs* en *trans*, entonces la ausencia de elementos *lennies* intactos podría explicar la ausencia de eventos de retrotransposición de *ta-TRIMs* en *E. multilocularis*. Es de destacar que se obtuvieron resultados similares al analizar *E. granulosus*.

Debido a la calidad de borrador de los genomas de las especies de *Taenia* no pudimos determinar la existencia de análogos de *lennies* (tenemos fuertes evidencias para considerar que no son los mismos), ni el estado transcripcional de los *ta-TRIMs*. Sin embargo, en todas las especies de *Taenia* analizadas, los TSDs de los *ta-TRIMs* fueron de 5 pb, más específicamente encontramos 8/8 eventos en *T. solium* y 2/2 eventos en *T. taeniaeformis*. La proporción de TSDs de 4 nucleótidos y 5 nucleótidos de largo es significativamente diferente entre *T. solium* y *E. multilocularis* (test exacto de Fischer;  $p < 1 \times 10^{-6}$ ). Este resultado tiene una importante consecuencia, y es que parece improbable que los Ténidos tengan como elemento autónomo movilizador de *ta-TRIMs* al elemento *lennie* que hemos descrito en el caso de las especies *Echinococcus*, debido, a que como hemos mencionado previamente, la longitud del TSD es característica del grupo IN (Wu et al. 2005; Ballandras-Colas et al. 2013).

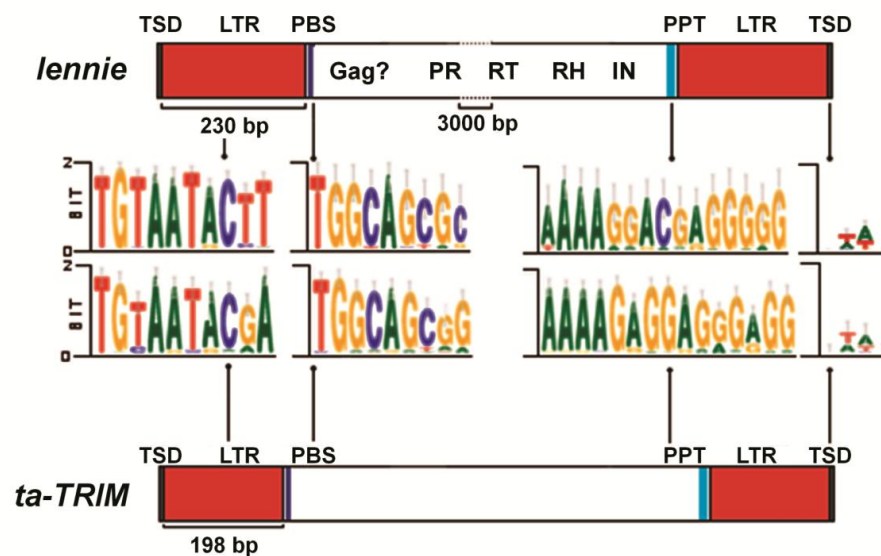


Figura 9. Diagrama en escala de elementos *ta-TRIM* y *lennie* de *E. multilocularis*, y comparación directa de logos de secuencias de características seleccionadas de ambos elementos (de izquierda a derecha, 5' LTR, PBS, PPT y TSD). Los logos fueron creados a partir del alineamiento de secuencias de elementos *ta-TRIMs* o *lennie* con WebLogo 3.4 (Crooks et al. 2004). Además de los dominios PR, RT, RH e IN, existe en *lennie* un marco abierto de lectura *upstream* a PR que puede codificar para Gag pero no muestra similitud con ninguna secuencia fuera de los platelmintos.

## **Análisis de expresión masiva de *ta-TRIMs* y características funcionales.**

### **Expresión masiva de *ta-TRIMs* y generación de transcritos novel a partir de solo-LTRs.**

Los elementos que hemos denominado *ta-TRIMs* fueron descritos originalmente en *E. granulosus* debido a su masiva expresión en ciclos claves de la vida del parásito. Aproximadamente un 10% de todos los ESTs en la librerías generadas (*oligo-capped*) pertenecieron a elementos *ta-TRIMs*. Para comprobar que estos elementos son altamente expresados en otras especies de ténidos, hemos analizado una pequeña porción de ESTs de metacestodos de *E. multilocularis* y hemos encontrado que también en esta especie los *ta-TRIMs* son altamente expresados (1.8% mostró similitud a *ta-TRIMs* por BLASTN (evalue:  $1 \times 10^{-5}$ ; n= 4195 ESTs). A su vez hemos analizado también *T. solium*, en el cual hemos encontrado una alta proporción de ESTs con similitud a *ta-TRIMs*, (1,4 %, n= 74730 ESTs).

Con el objetivo de caracterizar los *ta-TRIMs* transcripcionalmente activos decidimos mapearlos al genoma. Mediante el uso de diversas librerías de ESTs de *E. multilocularis* hemos buscado transcritos tipo *ta-TRIMs* mediante BLASTN y mapeado al genoma de *E. multilocularis* bajo estrictos requerimientos (ver Materiales, métodos y estrategia). Hemos podido identificar 73 diferentes *loci* con similitud a *ta-TRIMs* y que poseen a su vez un alto nivel de actividad transcripcional. Es de destacar que los 73 *loci* fueron curados manualmente. Dentro de los 73 *loci* identificados, en 63 la transcripción surge dentro del LTR (Figura 10 y 11, material suplementario 9); siendo un fuerte indicador de que los transcritos producidos en estas regiones son generados por la actividad promotor del LTR y no consecuencia de la transcripción de genes *upstream*.

Por otra parte, hemos podido identificar 25 *loci* correspondientes a *ta-TRIMs* de copia completa, pero que en todos los casos mostraron severas sustituciones y deleciones en sitios importantes, pero que presentan dos LTRs claramente identificables (Figura 10A). También hemos podido identificar eventos de *cis-splicing* dentro de algunos de los *ta-TRIMs* identificados (fue confirmado por el grupo mediante la realización de 3' RACE), (Figura 10A). Estos resultados evidencian un nivel de control de la expresión de los *ta-TRIMs*, aunque por el momento no hemos



analizado las consecuencias de estos fenómenos de empalme alternativo. Por otra parte, no hemos detectado evidencia de transcripción en los *ta-TRIMs* de copia completa *downstream* de la región U5 del 3' LTR, lo que nos confirma que esta región efectivamente actúa como terminador de la transcripción. El resto de los *loci* analizados (38 restantes), son solo-LTR, los cuales fueron catalogados de acuerdo al comportamiento transcripcional en tres tipos principales. En una clase catalogamos a los solo-LTR que generan transcritos abortivos, en los cuales el sitio de inicio es el límite U3 y el sitio de finalización es el límite U5 del mismo LTR. Debido al origen de los solo-LTRs, podemos decir que la ausencia de el resto de los componentes del *ta-TRIM*, provoca en gran parte de los casos la ausencia de transcritos más allá del límite U5 del LTR 5' (Figura 10B panel 3). En estudios de retrovirus mutantes se ha observado el mismo fenómeno, el cual es consecuencia de la inhibición de la poliadenilación a nivel 5' LTR (Schrom et al. 2013). Nosotros proponemos que en el caso de los *ta-TRIMs* ocurre el mismo fenómeno. La segunda clase que hemos identificado dentro de este subtipo de ESTs, son solo-LTRs que a pesar de no presentar el resto de las características de los *ta-TRIMs*, si presentan transcripción en la regiones vecinas corriente abajo. En la mayoría de los casos se encontró evidencia de cis splicing alternativo, lo que como ya hemos mencionado indica un nivel de regulación, de estos novel ARNInc (Figura 10B, panel 1 y 2). Es de destacar que por el momento no hemos realizado la caracterización funcional de estos novel ARNInc. Por último e integrando la tercera clase, se encuentran los solo-LTR cuyo producto de transcripción se haya fusionado con la región 5' de genes corriente abajo. Esto provoca la generación de niveles de control en el genoma de *Ténidos* previamente no identificados (Figura 11). Los casos analizados están conservados a lo largo de los cestodos, y la secuencia aminoacídica predicha indica fuertemente que son genes que codifican para proteínas (mediante RT-PCR el grupo confirmó que 3 de las 4 fusiones estudiadas realmente ocurría; la cuarta fue inconclusa debido a la presencia de amplificación inespecífica).

Posteriormente nos propusimos estudiar algunos de los *loci* previamente mencionados con el fin de evidenciar diferencias evolutivas entre especies *Echinococcus* cercanas. Tres de los *loci* investigados que producen fusiones con genes divergen entre *E. multilocularis* y *E. granulosus*. En el *locus* 64 de *E. multilocularis* se evidencia fusión génica a partir de LTR, sin embargo al analizar la

región correspondiente en *E. granulosus* encontramos una copia completa de *ta-TRIMs*, por lo cual no se genera fusión génica (Figura 11G). En los otros dos casos estudiados no hemos identificado *ta-TRIMs* ni solo-LTR (loci 8 y 21), sin embargo, el resto de la secuencia mantiene homología lo cual es un fuerte indicador que la inserción y reducción de los *ta-TRIMs* correspondientes se produjo luego de la divergencia de las especies (Figura 11F). El ortólogo del *locus 21* en *E. granulosus* es un gen asociado a spliced leader resultado de *trans splicing* (Brehm et al. 2000). En el *locus 21* de *E. multilocularis* el transcripto iniciado en el solo-LTR es empalmado en *cis*, en exactamente el mismo sitio que lo hace el *splice leader* en *E. granulosus*. La presencia del solo-LTR en *E. multilocularis* sugiere un modo de competencia con el *spliced leader* o una forma de compensar la ausencia de este fenómeno (Figura 11C). A nuestro entender es la primera vez que se evidencia este tipo de fenómeno. Por otra parte hemos podido confirmar la existencia de un solo-LTR el cual es empalmado en *cis* por la secuencia del *spliced leader* (Parkinson et al. 2012), sin embargo, no hemos podido identificar (BLASTN, e valor de 1E-10) la presencia de la secuencia pre-solo-LTRs (35 pb), en otro lugar del genoma.

Finalmente, hemos realizado un estrategia análoga a la anterior pero con el objetivo de estudiar las fusiones génicas desde la óptica de *E. granulosus* (ver Materiales, métodos y estrategia). Dentro de los ESTs analizados hemos podido encontrar un ejemplo de fusión transcripcional entre un LTR y un gen (material suplementario 10). Dicha fusión no se evidencia en *E. multilocularis* debido a la ausencia del LTR.

En resumen, hemos observado que en todos los genomas analizados (*E. granulosus*, *E. multilocualris* y *T. solium*), algunos *ta-TRIMs* y derivados solo-LTRs son masivamente expresados. En *E. multilocualris*, hemos presentados fuerte evidencia de la generación de ARNinc e isoformas alternativas de genes no codificantes a partir de solo-LTR. Además hemos demostrado que algunos *ta-TRIMs* fueron generados luego de la separación de las especies de *Echinococcus*.

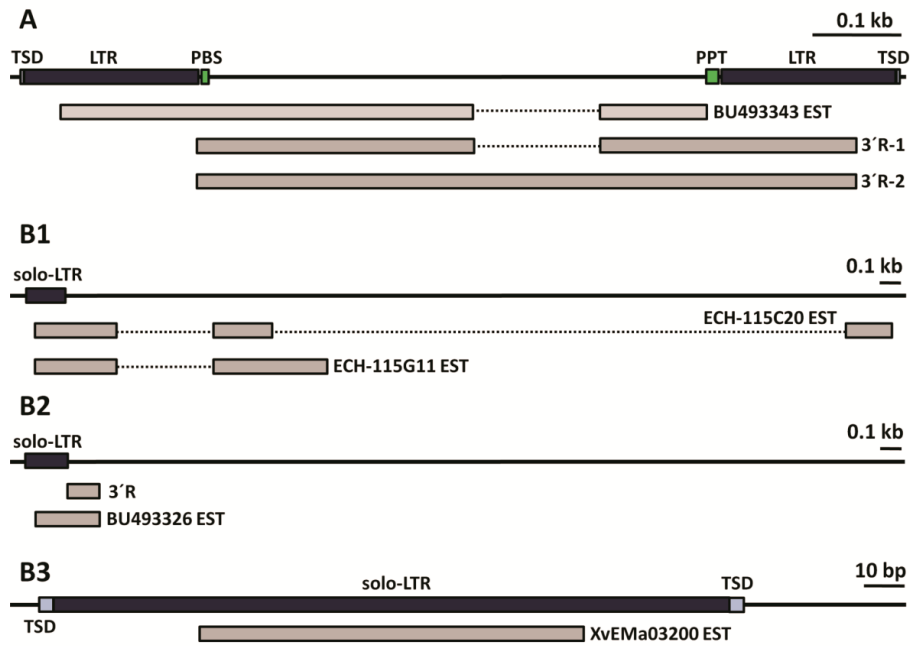


Figura 10. Transcripción de *ta-TRIMs* y elementos derivados en *E. multilocularis*. Las líneas continuas indican el *loci* genómico, los datos de EST y 3' RACE (3' R) se muestran debajo, con la regiones mapeadas dibujadas como rectángulos grises y los intrones como líneas punteadas. A. Ejemplo de un *full length ta-TRIM*. B1-B3, ejemplos de solo-LTRs iniciando la transcripción de diversos ARN no codificantes.

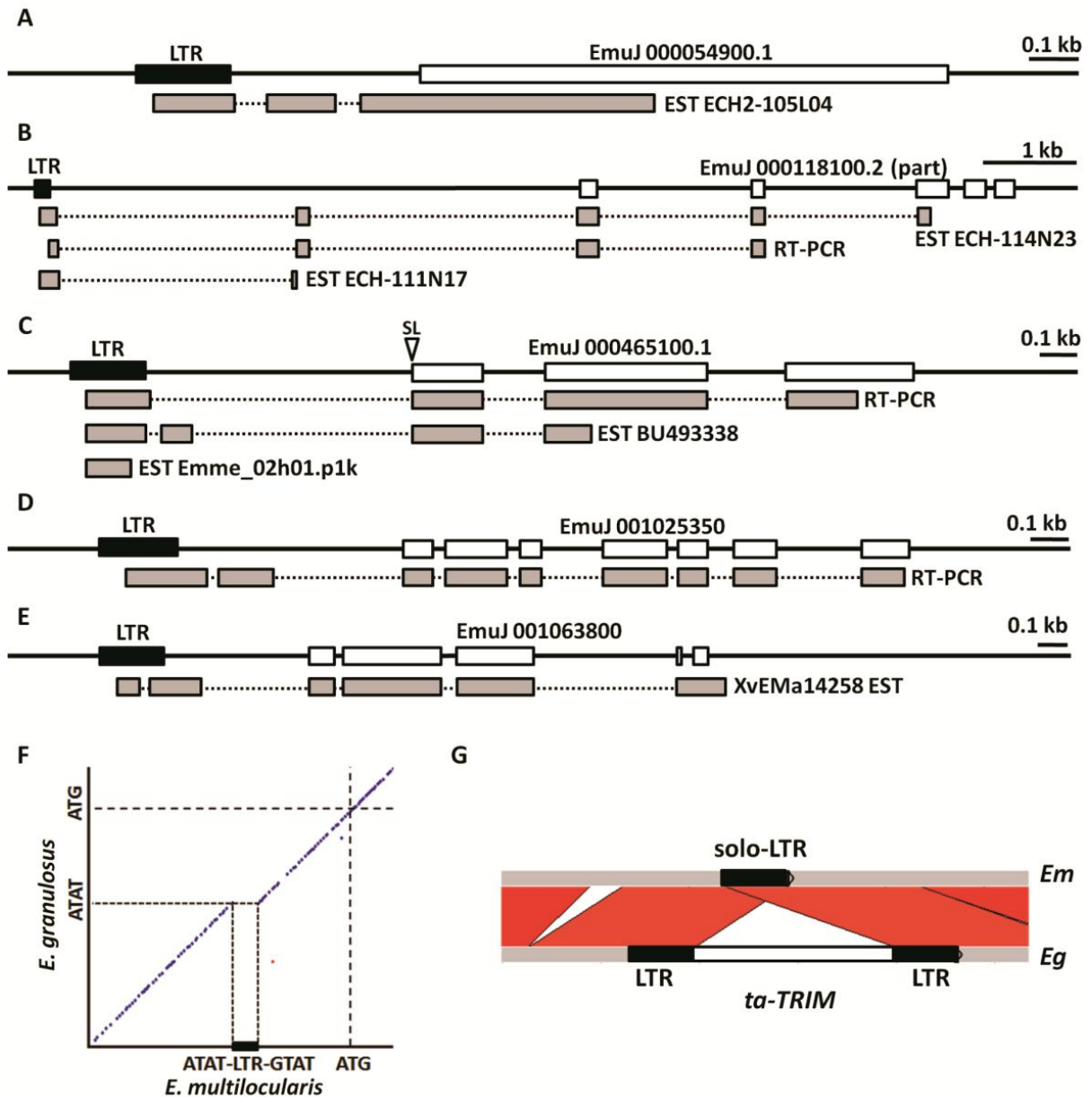


Figura 11. Fusiones transcripcionales de solo-LTRs y genes codificantes. Los dibujos están hechos como en la Figura 10. La flecha blanca ("SL") en el panel C indica la posición del sitio aceptor de *trans-splicing* en el ortólogo de *E. granulosus*. A. Locus 2. B. Locus 8. C. Locus 21. D. Locus 60. E. Locus 64. F. Dot plot (identidad dentro de una ventana de 10 pb se muestran como punto) de una región *upstream* (2 kb) del codón de inicio (ATG) de EmuJ\_000465100 (Locus 21: eje horizontal) y la región ortóloga de *E. granulosus*. El solo-LTR (rectángulo) está ausente en *E. granulosus*, y una secuencia similar (5'-ATAT-3') se observa en esta posición igualmente que en el TSD del solo-LTR de *E. multilocularis*. G. Comparación de una región *upstream* del locus 64 en *E. multilocularis* (Em), donde un solo-LTR es encontrado, con la región ortóloga de *E. granulosus* (Eg), la cual contiene un *full length ta-TRIM*. Las barras diagonales indican *hits* (BLASTN). Los dibujos fueron generados con WebACT (Abbott et al. 2005).

### **Análisis funcional de genes relacionados a *ta-TRIMs*.**

Con el objetivo de estudiar los posibles roles regulatorios de los *ta-TRIMs* decidimos investigar la existencia de correlación funcional entre los genes que poseen, en sentido *upstream*, un solo-LTRs a menos de 500 pares de bases (la distancia fue seleccionada de acuerdo a las fusiones confirmadas mediante experimentos de RT PCR). La ontología génica resultante indicó, tanto para *E. multilocularis* como *E. granulosus*, que no existen grupos de genes con características funcionales similares que fueran positivamente seleccionados para poseer en su extremo *upstream* un derivado solo-LTR. Es importante notar que estos resultados son preliminares ya que para obtener la lista de genes a analizar, no distinguimos entre aquellos que poseyeran solo-LTR activos o solo-LTR inactivos transcripcionalmente. Sin embargo, es importante destacar que algunos de los genes en los cuales encontramos fusión transcripcional poseen funciones biológicas de vital importancia, como puede ser la subunidad del exosoma RRP43 (degradación de ARNm específicos ricos en AU) (Anderson et al. 2006); CHMP5 involucrado en la formación de vesículas (Shim et al. 2005), así como proteínas hipotéticas conservadas sin función molecular aparente.

### **Panorama general de la expresión de *ta-TRIMs*.**

Con el fin de identificar cuantos *ta-TRIMs* están activos y como varía la actividad transcripcional de estos en función del ciclo de vida del parásito; hemos analizado datos de RNA-Seq, previamente publicados, de los estadios larvarios y adultos de *E. multilocularis* (Tsai et al. 2013). El *dataset* con el cual contamos consiste en experimentos de RNA-Seq de: adultos grávidos y pre-grávidos, metacestodo (sin protoescoléx), protoescoleces acitvos (tratados con pepsina y bajo pH) y no activos y cultivos primarios de células en proceso de regeneración larvaria (se utilizan como aproximado de células larvarias en las primeras etapas de desarrollo (Olson et al. 2012)).

Utilizando parámetros conservativos hemos determinado el nivel de expresión de cada copia de *ta-TRIM*. Los *ta-TRIMs* se expresan de forma notoriamente desigual. Mientras que algunos se expresan (y en algunos casos masivamente), la mayoría no se expresa (Figura 12A y material suplementarios 11, 12 y 13). Un 11%

de todos los *ta-TRIMs* generan un 90% de los *reads* mapeados de forma única. La actividad transcripcional de cada copia de *ta-TRIMs* parece ser heredada de estadio a estadio ya que no se observa variación (silenciamiento o activación) de *ta-TRIMs* interestadio. Es decir, que hay un alto nivel de correlación en el nivel de expresión en las diferentes librerías que componen el *dataset* utilizado por nosotros (Figura 12B y material suplementario 11, 12 y 13). Es importante señalar que los datos de RNA-Seq provienen de aislados de diversos lineajes de *E. multilocularis*, por lo que el nivel de expresión de cada copia de *ta-TRIM* se mantiene entre los diferentes lineajes analizados de *E. multilocularis*. Los *ta-TRIMs* transcripcionalmente activos se distribuyen de forma aproximadamente ubicua en el genoma (material suplementario 11). Los *ta-TRIMs* transcripcionalmente activos determinados por RNA-Seq correlacionan fuertemente con los determinados mediante datos de ESTs (test de correlación de Pearson,  $p < 0,00001$ ), confirmando que la variación observada entre *ta-TRIMs*, no es resultado de un artefacto del algoritmo utilizado por el mapeador, o consecuencia de los parámetros que hemos seleccionado. Es importante notar que en nuestros resultados fuimos lo más conservadores posibles para intentar, en la medida de lo posible, no introducir falsos positivos.

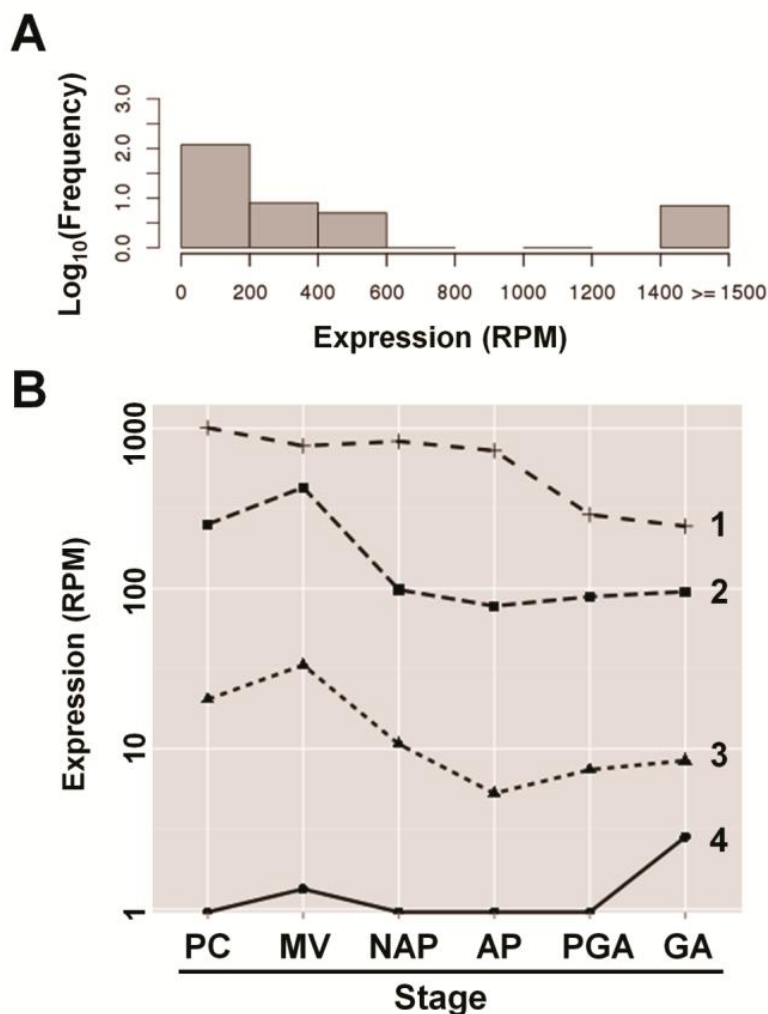


Figura 12. Análisis de RNA-Seq de *ta-TRIMs* en *E. multilocularis*. A. Histograma mostrando la distribución de niveles de expresión de *full-length ta-TRIMs* individuales (promedio de *reads* por *dataset*, normalizado por 106 lecturas de mapeo único (RPM)). B. Expresión de elementos representativos (en escala Log) a través de los *datasets* (PC: cultivo primario; MV, vesícula metacestoda; NAP, protoescólex no activado; AP, protoescólex activado; PGA, adultos pre - grávidos ; GA, adultos grávidos). Para la selección del elemento representativo, todos los *ta-TRIMs* fueron divididos en cuatro *bins* de acuerdo a su respectivo valor de RPM (1-4), y el *ta-TRIM* con el valor de mediana de RPM en cada *bin* fue seleccionado y graficado.

#### Análisis de expresión de *ta-TRIMs* en células germinativas de *E. multilocularis*

Los datos y conclusiones desarrollados en esta sección fueron, en su totalidad, obtenidos por Uriel Koziol, sin embargo debido a la importancia de los *ta-TRIMs* en ellos, he decido incluir los aspectos más importantes de estos resultados. Para más información sobre el tema referirse al apéndice ([Apéndice](#); Koziol, Radío et al. 2015).

Al analizar a nivel celular la expresión de *ta-TRIMs* durante el desarrollo del metacestodo larvario, Koziol determinó que las únicas células que poseen *ta-TRIMs*

activos en *E. multilocularis* son las células germinales (las únicas proliferativas en larvas) (Figuras 13A y 13H). En concordancia con lo anterior la distribución serotípica de las células germinal co-localiza durante todo el proceso mediante el cual se desarrolla el protoescólex a partir de la capa germinativa (Figuras 13B-G).

En resumen, se puede concluir que durante el desarrollo del metacestodo larvario (Figura 14) la señal co-localiza fuerte y específicamente con las células germinativas. Estos resultados tienen un gran impacto biotecnológico aplicado debido que hasta el momento son el mejor marcador de células madres en *E. multilocularis*, y posiblemente en el resto de los cestodos ténidos.

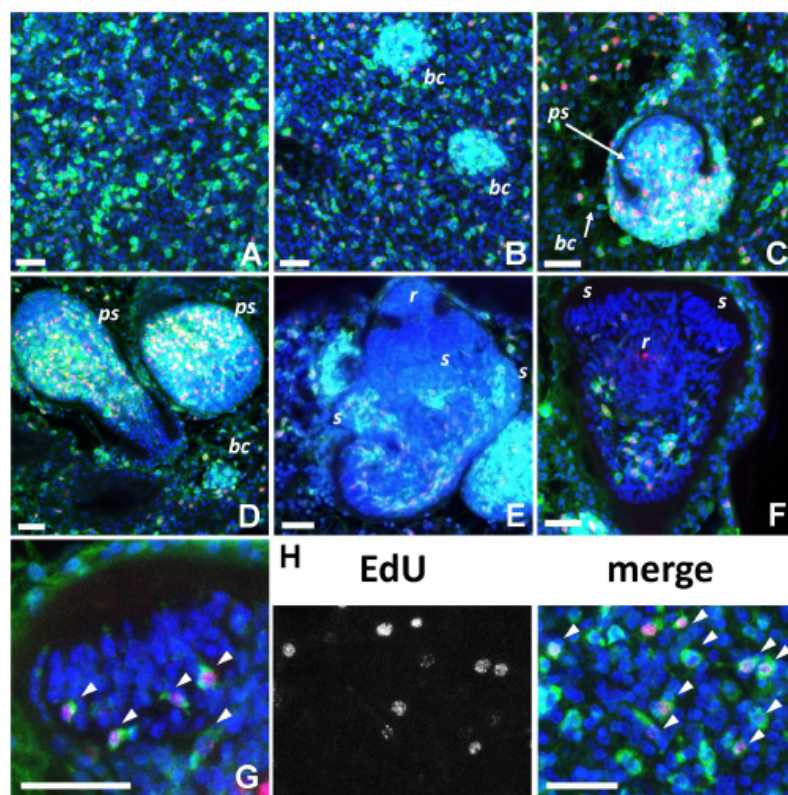


Figura 13. Análisis WMISH de la expresión de *ta-TRIMs* durante el desarrollo larvario de *E. multilocularis*. En todos los paneles, los *ta-TRIMs* se muestran en verde, DAPI (núcleos) en azul y en rojo se observa las células marcadas con EdU (proliferantes). A. Capa germinativa. B. Formación temprana de las cápsulas de cría (bc) a partir de la capa germinal. C-F. Diferentes estadios del desarrollo del protoescólex (ver Figura 14). r, rostellum (la señal rojo en el rostellum proviene de la autofluorescencia de los ganchos); s, ventosa. G. Detalles de la ventosa del protoescólex observada en (F). Las cabezas de flecha blanca señalan células EdU+ *ta-TRIM*+ en la base del desarrollo de la ventosa. H. Detalles de la capa germinativa. Las cabezas de flecha blanca señalan células EdU+ *ta-TRIM*+. Barras, 25  $\mu$ m.



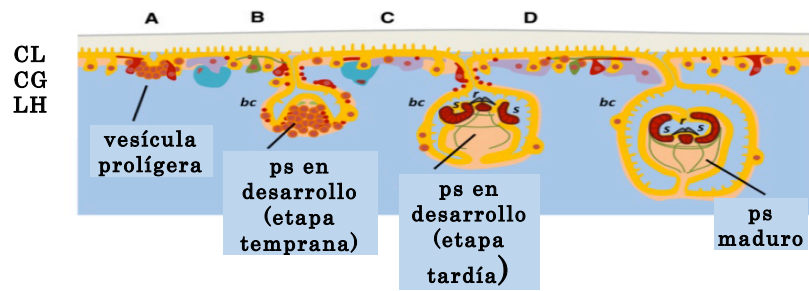


Figura 14. Organización general y desarrollo del metacestodo de *E. multilocularis*. CL – capa laminar; CG – capa germinativa; LH – líquido hidático. bc – vesícula prolígera; ps – protoescoléx; s – ventosa; r – rostelo. Tomado y adaptado de Koziol et al. (Koziol et al. 2014).

## >Discusión

Los *TRIMs* son un tipo de retrotransposones de LTR no autónomos los cuales han sido descritos únicamente en plantas (Witte et al 2001; Kalendar et al. 2008) y en hormigas (Zhou y Cahan 2012). En el presente, introducimos una nueva familia de *TRIMs*, la segunda en metazoarios y la primera en parásitos (cestodos ténidos).

Los *ta-TRIMs* de hormigas, plantas y cestodos no presentan ningún grado de similitud en su secuencia y son, como se ha propuesto, resultado de la evolución reductiva convergente de elementos autónomos que en busca de escapar a la fuerte selección negativa que opera sobre ellos, se redujeron a sus componentes mínimos (Schulman 2012). Nosotros creemos que es un mecanismo ampliamente utilizado, pero debido a que los métodos comúnmente utilizados para identificar retrotransposones no son los más adecuados (se especializan en retrotransposones autónomos), no han sido identificados en más genomas, aunque creemos que esta situación va a ir cambiando a la medida que se publiquen más trabajos como el que presentamos en esta tesina. En nuestro caso, poseemos datos de expresión que indican que estos elementos son masivamente expresados, sin embargo no sabemos (ya que no hay datos disponibles), si esto es la regla o la excepción. Sería muy interesante desarrollar un *software* especializado en detectar estos elementos transponibles.

Los datos filogenéticos presentados apoyan fuertemente la existencia de clados de *ta-TRIMs* que son específicos de linaje, contrariamente a lo que acontece en las especies de hormigas, en las cuales los autores sugieren la existencia de transferencia horizontal. Nuestra hipótesis se sustenta por la diferencia en la antigüedad de la inserción de los *ta-TRIMs* entre los principales linajes y además por la existencia de inserciones que ocurrieron luego de la separación de especies (Figura 8A, 8C y material suplementario 7 y 8).

La teoría de la evolución reductiva convergente necesita, para poder ser eficaz la existencia de un par autónomo que pueda movilizar en *trans* al par no autónomo correspondiente. Como ya hemos mencionado, para poder ser el responsable de la movilización del elemento no autónomo, el par autónomo debe poseer alta similitud de secuencia con el par no autónomo correspondiente. Si bien existen algunos pocos casos documentados de existencia de pares de alta

homología (LARD *Dasheng* y el retrotransposón de arroz RIRE-2) (Jiang et al. 2002), hasta ahora, con excepción de nuestro trabajo, ningún par se ha propuesto para familias de *TRIMs*. La alta similitud de secuencia se debe mantener en posiciones claves específicas (PBS, PPT y comienzo y final de LTRs) las cuales interaccionan con proteínas codificadas por genes que presenta el par autónomo como la IN y RT. Además para que se produzca el correcto empaquetamiento de los ribonucleótidos tipo *TRIM* en la partícula viral (VLP) es necesario la conservación de la estructura secundaria (Wicker et al. 2007). Nuestros resultados, basados en lo expresado anteriormente, indican que existe una alta probabilidad de que los *lennies* sean los elementos capaces de movilizar a los *ta-TRIMs* en *trans*. La ausencia de *lennies* intactos identificables puede explicar la falta de eventos de retrotransposición reciente, si bien hemos detectado *ta-TRIMs* en *E. multilocularis* con más de un 99% de identidad, son posiblemente consecuencia de eventos de duplicación genómica y no retrotransposición. Schulman ha propuesto también que la existencia de un grupo no autónomo sumamente eficaz lleve a la extinción del grupo autónomo correspondiente. La utilización de la maquinaria transcripcional por parte del par no autónomo hará que el ciclo de retrotransposición del par autónomo disminuya, con lo cual se generarán menos copias y las antiguas acumularán mutaciones deletéreas llevando a la extinción del grupo (Schulman 2012). Puede ser posible, que este fenómeno ocurriera en los genomas de cestodos y los propios *ta-TRIMs* llevarán a la extinción de los *lennies*. En nuestro caso, el efecto se potencia debido a que los *ta-TRIMs* son masivamente expresados. En concordancia con el efecto producido por la expresión de los *ta-TRIMs* no podemos descartar que la expresión masiva de estos elementos específicamente en las células germinales (únicas proliferativas) sea un mecanismo para limitar la expansión de los *lennies*. En la especie *Taenia*, hemos encontrado un tamaño diferente de TSDs lo cual indica que estas especies utilizan una clase de retrotransposón autónomo (no identificado) diferente al que utilizan las especie de *Echinococcus*; el tamaño del TSD es un indicativo de la IN utilizada (Zhou et al. 2001). La presencia de un factor movilizador específico para *Echinococcus* y otro (al menos uno) para *Taenia*, indican que durante la evolución y divergencia de Ténidos ocurrió un cambio en el factor movilizador, que de acuerdo a nuestros datos, debió haber ocurrido al comienzo de la separación de las especies.

Existe numerosa evidencia de que los retrotransposones participan, y de forma influyente, en la evolución de los genomas. Ya sea, mediante la expansión de tamaño genómico o mediante la regulación de genes codificantes (Peatson et al. 2004; Batut et al. 2013). En nuestro caso, la evidencia de EST (y RNA-Seq en *E. multilocularis*) sugiere que en todas las especies de ténidos, los *ta-TRIMs* son transcriptos a muy alto nivel. Si bien es posible que los *ta-TRIMs* no tengan un efecto dramático en el tamaño genómico, si influyen en el paisaje transcripcional de los cestodos ténidos. En *E. multilocularis* los derivados solo-LTRs tienen un marcado impacto en la expresión de secuencias *downstream*, resultando en la formación de nuevos ARNinc o alterando la isoforma de genes codificantes (Figuras 10 y 11). Aún no hemos investigado el posible rol de estos ARNinc. Además debido a que todos los genes que presentan fusión con los solo-LTR son (o al menos existe la posibilidad) regulados por el mismo promotor, se puede regular la expresión coordinada de este grupo de genes. En la última década más y más evidencia se ha acumulado del rol de los ARNinc en los genomas. Si bien los ARNinc generados a partir de derivados solo-LTRs pueden ser simple consecuencia del mecanismo transcripcional, a lo largo de la evolución muchos pudieron haber adaptado un rol funcional, como puede ser la regulación de genes en *cis* o en *trans* (Merceret et al. 2009).

Los numerosos ejemplos de *ta-TRIMs* (y sus derivados solo-LTRs) que difieren (presencia/ausencia) entre especies de *Echinococcus* puede generar importantes diferencias en la regulación génica entre ambas especies. Debido a su expresión específica en células germinales, es posible que los *ta-TRIMs* hayan contribuido a las diferencias encontradas en la morfología y el desarrollo de los cestodos *Echinococcus*. Si bien como hemos destacado con anterioridad, alguno de los genes que presentan fusión son claves para el organismo, necesitaríamos nuevos datos (análisis transcriptómicos de células germinales) para poder confirmar más fusiones y analizar en mayor profundidad la relación de los *ta-TRIMs* y sus derivados solo-LTRs con los genes. Así podremos entender mejor las diferencias que existen entre estas dos especies estrechamente relacionadas.

El silenciamiento de los transposones en los cestodos ténidos siempre ha sido un tema intrigante, ya que no cuentan con los mecanismos clásicos de silenciamientos (vía PIWI), y sin embargo presentan un proporción destacablemente

menor de transposones que el resto de los cestodos (Skinner et al. 2014; Tsai et al. 2013). Es posible que la propia existencia de elementos *ta-TRIMs* altamente expresados sea uno de los mecanismos de represión utilizados por estos organismos con el objetivo primario de silenciar los pares autónomos.

El descubrimiento de que los *ta-TRIMs* son marcadores específicos de células germinales puede marcar una revolución en el estudio de las células madres de estos organismos.

## >Bibliografía

Abbott JC, Aanensen DM, Rutherford K, Butcher S, Spratt BG. 2005. WebACT--an online companion for the Artemis comparison tool. *Bioinformatics* 21:3665- 3666.

Anderson JR, Mukherjee D, Muthukumaraswamy K, Moraes KC, Wilusz CJ, Wilusz J. 2006. Sequence-specific RNA binding mediated by the RNase PH domain of components of the exosome. *RNA* 12:1810-1816.

Anxolabéhère D, Kidwell MG, Perique G. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. *Mol Biol Evol* 5(3): 252-269.

Ballandras-Colas A, Naraharisetty H, Li X, Serrao E, Engelman A. 2013. Biochemical characterization of novel retroviral integrase proteins. *PLoS One* 8:e76638.

Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. 2013. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res* 23:169-180.

Bauer M, Klau GW, Reinert K. 2007. Accurate multiple sequence-structure alignment of RNA sequences using combinatorial optimization. *BMC Bioinformatics* 8(1), 271.

Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9(1), 474.

Berriman MB, et al. 2009. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 460:352-358.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* 186:1085-1093.

Boeke JD, Stoye JP. 1997. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In: JM Coffin, SH Hughes, and HE Varmus, editors. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

Brehm K, Jensen K, Frosch M. 2000. mRNA *trans*-splicing in the human parasitic cestode *Echinococcus multilocularis*. J Biol Chem 275:38311-38318.

Brunetti E, Kern P, Vuitton DA. 2010. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. Acta Trop 114(1):1-16.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2008. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28(4):464-9

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18): 3674-3676.

Copeland CS, Brindley PJ, Heyers O, Michael SF, Johnston DA, Williams D L, Kalinna BH. 2003. Boudicca, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. J Virol, 77(11): 6153-6166.

Copeland CS, Mann VH, Morales ME, Kalinna BH, Brindley PJ. 2005. The Sinbad retrotransposon from the genome of the human blood fluke, *Schistosoma mansoni*, and the distribution of related Pao-like elements. BMC Evol Biol 5(1), 20.

Craig PS, Rogan MT, Campos-Ponce M. 2003. Echinococcosis: disease, detection and transmission. Parasitology 127 Suppl:S5-20.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188-1190.

DeMarco R, et al. 2004. Saci-1,-2, and-3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. J Virol 78:2967-2978.

Eckert J, Deplazes P. 2004. Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. Clin Microbiol Rev 17(1):107-135.

Eckert J, Thompson RCA, Mehlhorn H. 1983. Proliferation and metastases formation of larval *Echinococcus multilocularis*. *Z Parasitenkd* 69(6): 737-748.

Epping K, Brehm K. 2011. *Echinococcus multilocularis*: Molecular characterization of EmSmadE, a novel BR-Smad involved in TGF- $\beta$  and BMP signaling. *Exp Parasitol* 129(2): 85-94.

Extavour CG, Akam M. 2003. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130:5869-5884.

Felsenstein J, Kishin H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst Biol* 193-200.

Forrester SJ, Hall N. 2014. The revolution of whole genome sequencing to study parasites. *Mol Biochem Parasitol* 195(2): 77-81.

Freeman RS, Fallis AM, Shea M, Maberley AL, Walters J. 1973. Intraocular *Taenia crassiceps* (Cestoda): Part II. The parasite. *Am J Trop Med Hyg* 22(4): 493-495.

Garcia HH, Rodriguez S, Friedland JS. 2014. Immunology of *Taenia solium* taeniasis and human cysticercosis. *Parasite Immunol* 36(8): 388-396.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman R W, ... Barrell B. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419(6906): 498-511.

Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. *Genome Biol* 5:225.

Heath DD, Lawrence SB. 1976. *Echinococcus granulosus*: development in vitro from oncosphere to immature hydatid cyst. *Parasitology* 73(3): 417-423.

Hindmarsh P, Leis J. 1999. Retroviral DNA integration. *Microbiol Mol Biol Rev.* 63:836-843

Hoberg EP. 2006. Phylogeny of *Taenia*: species definitions and origins of human parasites. *Parasitol Int* 55:S23-S30.

Hock J, Meister G. 2008. The Argonaute protein family. *Genome Biol* 9:210.



Hüttner M, Nakao M, Wassermann T, Siefert L, Boomker JD, Dinkel A, Sako Y, Mackenstedt U, Romig T, Ito A. 2008. Genetic characterization and phylogenetic position of *Echinococcus felidis* (Cestoda: Taeniidae) from the African lion. *Int J Parasitol* 38:861-868.

Jiang N, Jordan IK, Wessler SR. 2002. Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. *Plant Physiol* 130:1697-1705.

Juliano CE, Swartz SZ, Wessel GM. 2010. A conserved germline multipotency program. *Development* 137:4113-4126.

Juliano C, Wang J, Lin H. 2011. Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annu Rev Genet* 45:447- 469.

Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O, Schulman AH. 2008. Cassandra retrotransposons carry independently transcribed 5S RNA. *Proc Natl Acad Sci USA* 105:5833-5838.

Kijima TE, Innan H. 2010. On the estimation of the insertion time of LTR retrotransposable elements. *Mol Biol Evol* 27:896-904.

Knapp J, Nakao M, Yanagida T, Okamoto M, Saarma U, Lavikainen A, Ito A. 2011. Phylogenetic relationships within *Echinococcus* and *Taenia* tapeworms (Cestoda: Taeniidae): an inference from nuclear protein-coding genes. *Mol Phylogenet Evol* 61:628-638.

Kornyushin V, Greben O. 2014. A new genus and species of cestodes (Cyclophyllidea, Gryporhynchidae) from Ciconiiform birds. *Vestnik Zoologii*, 48:483-494

Koziol U, Marin M, Castillo E. 2008. Pumilio genes from the Platyhelminthes. *Dev Genes Evol* 218:47-53.

Koziol U, Iriarte A, Castillo E, Soto J, Bello G, Cajarville A, Roche L, Marin M. 2009. Characterization of a putative hsp70 pseudogene transcribed in protoscoleces and adult worms of *Echinococcus granulosus*. *Gene* 443:1-11

Koziol U, Krohne G, Brehm K. 2013. Anatomy and development of the larval nervous system in *Echinococcus multilocularis*. *Front Zool* 10:24.

Koziol U, Rauschendorfer T, Zanon Rodriguez L, Krohne G, Brehm K. 2014. The unique stem cell system of the immortal larva of the human parasite *Echinococcus multilocularis*. *Evodevo* 5:10.

Laha T, Loukas A, Verity CK, McManus DP, Brindley PJ. 2001. Gulliver, a long terminal repeat retrotransposon from the genome of the oriental blood fluke *Schistosoma japonicum*. *Gene* 264:59-68.

Laha T, Loukas A, Smyth DJ, Copeland CS, Brindley PJ. 2004. The fugitive LTR retrotransposon from the genome of the human blood fluke, *Schistosoma mansoni*. *Int J Parasitol* 34:1365-1375.

Li H, et al. 2009. The Sequence Alignment/Map format and SAM tools. *Bioinformatics* 25:2078-2079.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923-930.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955-964.

Mak J, Kleiman L. 1997. Primer tRNAs for reverse transcription. *J Virol* 71:8087-8095.

Marin M, Garat B, Pettersson U, Ehrlich R. 1993. Isolation and characterization of a middle repetitive DNA element from *Echinococcus granulosus*. *Mol Biochem Parasitol* 59:335-338.

McClintock B. 1946. Maize genetics. *Carnegie Inst Washington Year Book*, 45:176-186.

Moro P, Schantz PM. 2009. Echinococcosis: a review. *Int J Inf Dis* 13:125-133.

Nakao M, Lavikainen A, Iwaki T, Haukisalmi V, Konyaev S, Oku Y, Okamoto M, Ito A. 2013. Molecular phylogeny of the genus *Taenia* (Cestoda: Taeniidae):

proposals for the resurrection of *Hydatigera* Lamarck, 1816 and the creation of a new genus *Versteria*. *Int J Parasitol* 43:427-437.

Olson PD, Zarowiecki M, Kiss F, Brehm K. 2012. Cestode genomics - progress and prospects for advancing basic and applied aspects of flatworm biology. *Parasite Immunol* 34:130-150.

Parkinson J, et al. 2012. A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. *PLoS Negl Trop Dis* 6:e1897.

Peaston AE et al. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 7:597-606.

Picard G. 1978. Non mendelian female sterility in *Drosophila melanogaster*, further data on chromosomal contamination. *Mol Gen Genet MGG* 164(3):235-247.

Rink JC. 2013. Stem cell systems and regeneration in planaria. *Dev Genes Evol* 223:67- 84.

Robb SM, Ross E, Sanchez Alvarado A. 2008. SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res* 36:D599-606.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467.

Sato K, Shibata N, Orii H, Amikura R, Sakurai T, Agata K, Kobayashi S, Watanabe K. 2006. Identification and origin of the germline stem cells as revealed by the expression of nanos-related gene in planarians. *Dev Growth Differ* 48:615- 628

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol* 6:526- 538.

Sawyer SA. 1999. GENECONV: A computer package for the statistical detection of gene conversion <http://www.math.wustl.edu/~sawyer/>.

Shim JH, et al. 2006. CHMP5 is essential for late endosome function and downregulation of receptor signaling during mouse embryogenesis. *J Cell Biol* 172:1045-1056.

Schubert A, Koziol U, Cailliau K, Vanderstraete M, Dissous C, Brehm K. 2014. Targeting *Echinococcus multilocularis* stem cells by inhibition of the polo-like kinase EmPlk1. PLoS Negl Trop Dis 8(6):e2870.

Schulman AH. 2012. Hitching a Ride: Nonautonomous Retrotransposons and Parasitism as a Lifestyle. Top Curr Gen 24:71-88.

Shapiro JA, Adhya SL. 1969. The galactose operon of *E. coli* K-12. II. A deletion analysis of operon structure and polarity. Genetics 62(2):249.

Skinner DE, Rinaldi G, Koziol U, Brehm K, Brindley PJ. 2014. How might flukes and tapeworms maintain genome integrity without a canonical piRNA pathway? Trends Parasitol 30:123-129.

Strona G, Fattorini S. 2014. Parasitic worms: how many really? Int J Parasit 44 (5): 269-272

Swiderski RE, Nishimura DY, Mullins RF, Olvera MA, Ross JL, Huang J, Stone EM, Sheffield VC. 2007. Gene expression analysis of photoreceptor cell loss in *bbs4*-knockout mice reveals an early stress gene response and photoreceptor cell damage. Invest Ophthalmol Vis Sci, 48(7), 3329.

Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739.

Telesnitsky A, Goff SP. 1997. Reverse transcriptase and the generation of retroviral DNA. In: JM Coffin, SH Hughes, and HE Varmus, eds. Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

Thompson, RCA. 1986. Biology and systematics of *Echinococcus*. In: RCA Thompson, ed. The Biology of *Echinococcus* and hydatid disease. George Allen & Unwin. p. 5-43.

Thompson RCA, McManus DP 2002. Towards a taxonomic revision of the genus *Echinococcus*, Trends Parasitol 18(10):452-457

Torarinsson E, Havgaard JH, Gorodkin J. 2007. Multiple structural alignment and clustering of RNA sequences. Bioinformatics 23(8):926-32.

Torarinsson E, Lindgreen S. 2008. WAR: Webserver for aligning structural RNAs. Nucleic Acids Res 36:79-84.

Torgerson PR, Keller K, Magnotta M, Ragland N. (2010). The global burden of alveolar echinococcosis. PLoS Negl Trop Dis 4:722.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3):562-78.

Tsai IJ, et al. 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496:57-63.

Wang B, Collins JJ 3rd, Newmark PA. 2013. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. Elife 2:e00768.

Wang Y, Zayas RM, Guo T, Newmark PA. 2007. nanos function is essential for development and regeneration of planarian germ cells. Proc Natl Acad Sci U S A 104:5901-5906

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973-982.

Witte CP, Le QH, Bureau T, Kumar A. 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98:13778-13783.

Wu X, Li Y, Crise B, Burgess SM, Munroe DJ. 2005. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. J Virol 79:5211-5214.

Xiao, Qiu J, Nakao M, Li T, Yang W, Chen X, Schantz PM, Craig PS, Ito A. 2006. *Echinococcus shiquicus*, a new species from the Qinghai–Tibet plateau region of China: Discovery and epidemiological implications. *Parasitol Int*, 55:S233-S236.

Xu X, Ji Y, Stormo GD. 2007. RNA Sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment. *Bioinformatics* 23:1883-1891.

Yao Z, Weinberg Z, Ruzzo WL. 2006. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* 22(4): 445-452.

Zhao XU, Hao W. 2007. LTR FINDER USER MANUAL version 1.0. 2.

Zheng H, et al. 2013. The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat Genet* 45:1168-1175.

Zhou H, Rainey GJ, Wong SK, Coffin JM. 2001. Substrate sequence selection by retroviral integrase. *J Virol* 75:1359-1370.

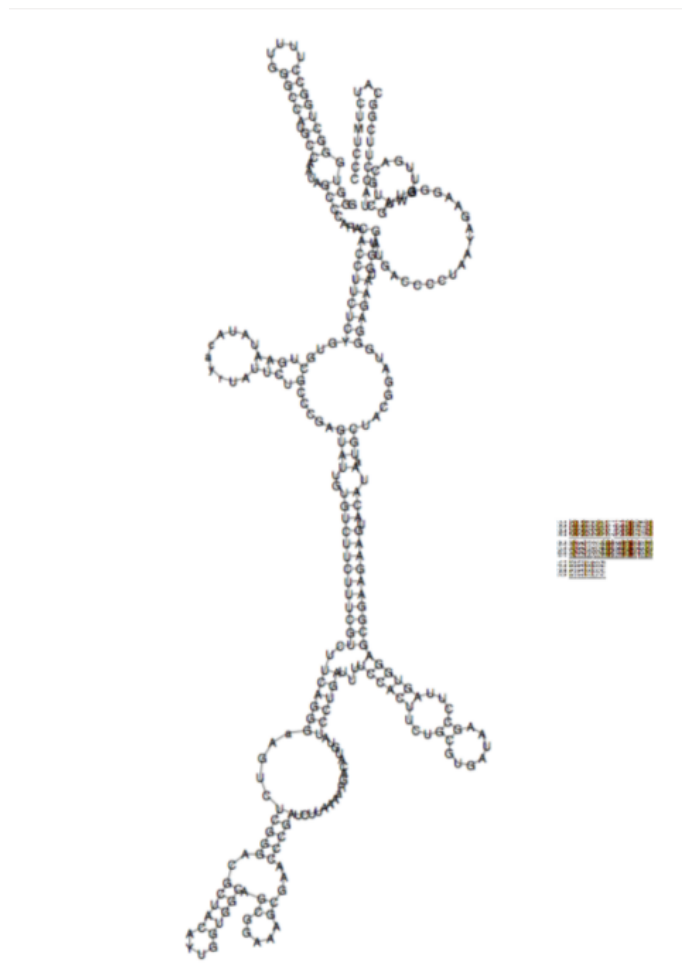
Zhou Y, Cahan SH. 2012. A novel family of terminal-repeat retrotransposons in miniature (TRIM) in the genome of the red harvester ant, *Pogonomyrmex barbatus*. *PLoS One* 7:e53401.

## >Material suplementario

El material suplementario se encuentra disponible en: [material suplementario](#).

1 – A. Archivo Fasta de elementos *ta-TRIMs* y *lennies* de *E. multilocularis*, *E. granulosus* y *T. solium*. B. Tabla de estructura de los *ta-TRIMs* en *E. multilocularis*. C. Tabla de estructura de los *ta-TRIMs* en *E. granulosus*.

2 – Estructura secundaria predicha para *ta-TRIMs* de *Echinococcus*. La estructura secundaria fue generada mediante el uso del servidor WAR. Para más información referirse a Materiales, métodos y estrategia.

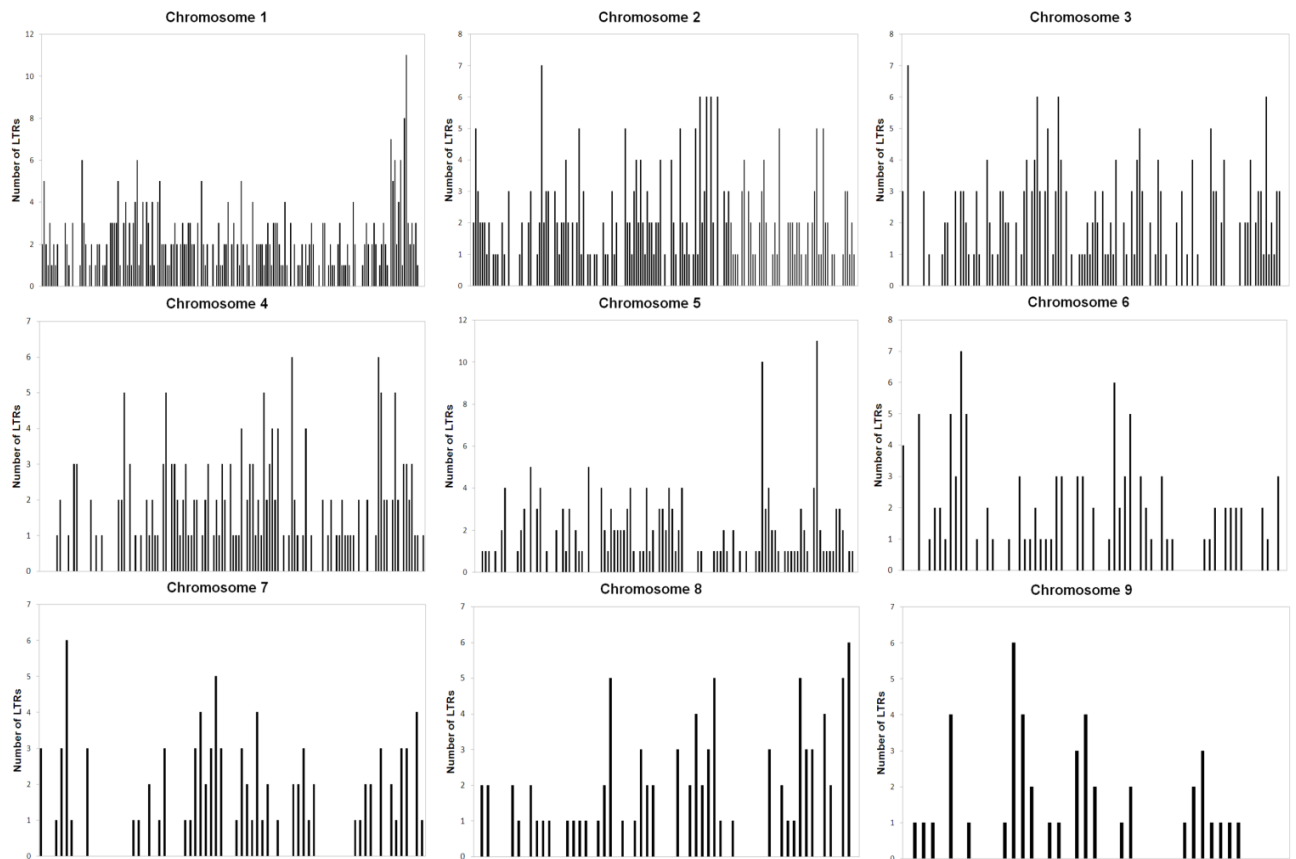


3 - Archivos GFF y tablas de *ta-TRIMs* y *lennies* de *E. multilocularis*. Los archivos gff están destinados a ser usados con el genoma de *E. multilocularis* (versión 4) de [www.genedb.org](http://www.genedb.org)

4 - Archivos GFF y tablas de *ta-TRIMs* y *lennies* de *E. granulosus*. Los archivos gff están destinados a ser usados con el genoma de *E. granulosus* (versión 3) de [www.genedb.org](http://www.genedb.org)

5 – Archivos GFF de *ta-TRIMs* de *T. solium*. Los archivos gff están destinados a ser usados con el genoma de *T. solium* de [www.genedb.org](http://www.genedb.org)

6 – Distribución de *ta-TRIMs* en el genoma de *E. multilocularis*. La figura muestra gráficamente la distribución de *ta-TRIMs* dentro de los cromosomas de *E. multilocularis*. El número de secuencias que muestran similitud a la región LTR de *ta-TRIMs* de *E. multilocularis*, obtenidos a partir de la combinación de Blast y análisis HMM, fue contado y graficado para ventanas no deslizantes de 100 kb en cada cromosoma. El largo total de cada cromosoma (1 – 9) es 20116480, 17597483, 14622996, 13762453, 11859843, 7212196, 7437390, 6203120 and 4267854 pb, respectivamente.



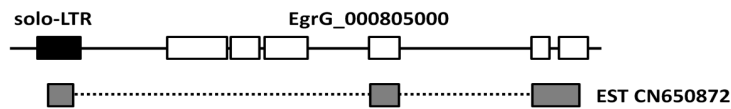
7 – Ejemplos de integración de *ta-TRIMs* luego de la divergencia de las especies *Echinococcus* y *Taenia*.

8 – Tabla de inserción de *ta-TRIMs* en *Echinococcus*. La tabla fue realizada con una estrategia menos restrictiva que los ejemplos descritos en el material suplementario 7.

9 - Lista de *loci* de *E. multilocularis* con similitud a *ta-TRIMs* y con evidencia de actividad transcripcional a partir de análisis de ESTs y 3' RACE.

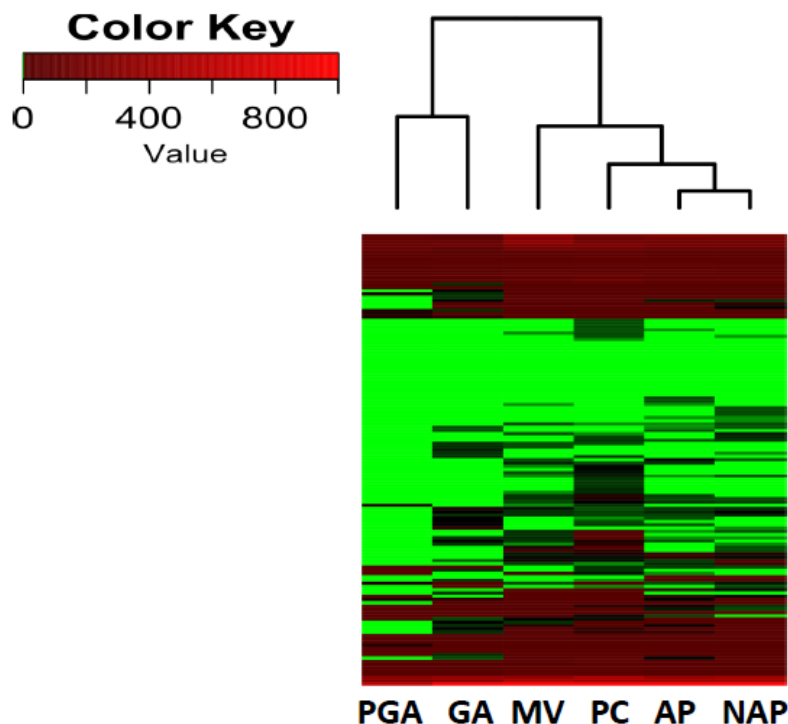
10 – Fusión transcripcional de un LTR con una región *downstream* de un gen en *E. granulosus*. Un EST (CN650872) indica una isoforma alternativa para el gen EgrG\_00080500, el cual comienza la transcripción a partir de un solo-LTR.





11 – Tabla de RNA-Seq análisis para *ta-TRIMs* individuales de *E. multilocularis*. El número de lecturas para cada elemento fue normalizado por el número total de lecturas con mapeo único para cada *dataset*. No se realizó normalización por largo de secuencia debido a que la proporción de cada elemento que puede ser mapeada de forma única varía.

12 - *Heat-map* mostrando la expresión de *full length ta-TRIMs* (conteo por millón de lecturas) en diferentes estadios del desarrollo de *E. multilocularis* (utilizando datos obtenidos a partir de material suplementario 11).



13 – Evidencia de expresión de *ta-TRIMs* parciales. Histograma mostrando la distribución de los niveles de expresión de *ta-TRIMs* parciales de *E.* (promedio de lecturas por *dataset*, normalizado por el número total de lecturas de mapeo único (RPM)). B. Expresión de solo-LTRs representativos a través del *dataset* (PC: cultivo primario; MV, vesícula metacestoda; NAP, protoescólex no activado; AP, protoescólex activado; PGA, adultos pre - grávidos ; GA, adultos grávidos). Para la selección del solo-LTR representativo, todos los solo-LTRs fueron divididos en cuatro *bins* de acuerdo a su respectivo valor de RPM (1-4), y el solo-LTR con el valor de mediana de RPM en cada *bin* fue seleccionado y graficado.

