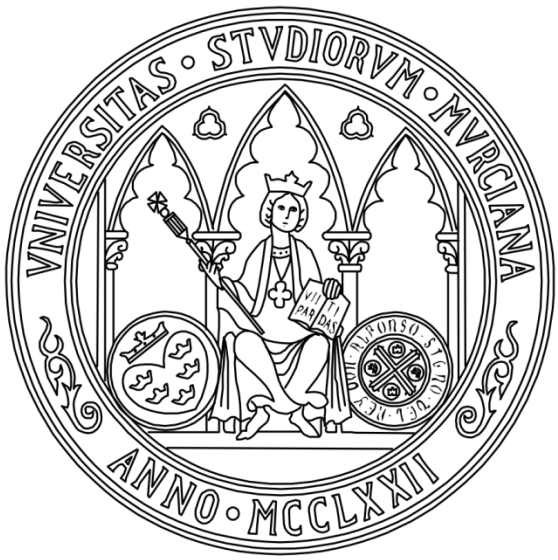


UNIVERSIDAD DE MURCIA

FACULTAD DE BIOLOGÍA

Evolutionary genomics in cicindelids:
characterisation and expression analysis
of reproductive and immune related genes

Genómica evolutiva en cicindélidos:
caracterización y análisis de expresión de genes
relacionados con la reproducción y el sistema inmune

Dña. María Juliana Rodríguez García
2015

UNIVERSIDAD DE MURCIA

Evolutionary genomics in cicindelids:

characterisation and expression analysis of reproductive and immune related genes

---

Genómica evolutiva en cicindélidos:
caracterización y análisis de expresión de genes relacionados con la reproducción y el sistema inmune

María Juliana Rodríguez García

2015

Departamento de Zoología y Antropología Física

Directores:

Dr. José Galián Albaladejo

Dr. Vilmar Machado

Contents

Resumen

Introducción

Escarabajos tigre

Los escarabajos tigre o cicindélidos forman un grupo bien definido dentro de los Adéfagos, un suborden de los coleópteros, que incluye especies no plaga de colores brillantes que son componentes importantes de los ecosistemas, siendo eslabones importantes en las cadenas alimentarias. Adephaga es el segundo suborden más grande de coleópteros, con más de 36.000 especies conocidas (Nilsson, 2001; Lorenz, 2005), incluye también los escarabajos de tierra (Carabidae), los escarabajos buceadores depredadores (Dytiscidae), los escarabajos giradores (Gyrinidae), y los escarabajos nadadores (Haliplidae), entre otras familias. Existe controversia en relación a la clasificación de los cicindélidos dentro del suborden Adephaga, habiendo sido considerados una subfamilia (Cicindelinae), una tribu (Cicindelini), o una supertribu (Cicindelitae) dentro de la familia Carabidae (Ball, 1979). Algunos estudios filogenéticos del suborden Adephaga (Shull *et al.,* 2001; Hunt *et al.,* 2007), también mostraron a los cicindélidos como un clado dentro de la familia Carabidae. Sin embargo, otros estudios más recientes demostraron la monofília de los cicindélidos, situándolos fuera de los carábidos y apoyando su estatus de familia (Cicindelidae) (Putchkov & Cassola, 2005; Maddison *et al.,* 2009; Bocak *et al.,* 2014). La historia de los cindelidos se ha caracterizado por un aumento de su diversidad, teniendo en cuenta tanto la diversidad morfológica como la taxonómica, que se inició en sus antepasados similares a carábidos (*Omus*, *Manticora*, *Megacephala*) y continúa a través de los grupos arborícolas (Collyridini) y los grupos más derivados en el género *Cicindela* adaptados a la carrera y al vuelo (Pearson & Vogler, 2001).

Para la investigación de la evolución es importante tener un contexto filogenético bien definido como herramienta para contribuir al conocimiento del proceso evolutivo, así como la caracterización funcional de los genes por métodos comparativos (Jansa *et al.,* 2003). En cicindélidos se han realizado estudios filogenéticos centrados en la identificación de las relaciones evolutivas entre diferentes grupos (Galián *et al.,* 2002) y otros enfocados a grupos específicos

(Cardoso & Vogler, 2005; Vogler *et al.,* 2005; Pons *et al.,* 2004, 2006; López-López *et al.,* 2012, 2013, 2015).

Hay varias características que los escarabajos tigre normalmente comparten: i) largas mandíbulas en forma de hoz; ii) varios tipos de dientes, dientes simples y un diente compuesto; iii) patas marchadoras relativamente delgadas y largas; iv) forma del cuerpo alargada, con los ojos y la cabeza más anchos que el tórax; y v) comportamiento de construcción de túneles en la larva (Pearson, 1988; Pearson & Vogler, 2001)

Los escarabajos tigre adultos varían en tamaño (desde 5 mm a más de 4 cm el más grande) y color (muchos son de color negro, pero algunos tienen colores brillantes como el verde, azul, violeta, rojo o amarillo). Las formas del cuerpo varían desde robusta a otras más estilizadas, y tienen alas funcionales transparentes que normalmente utilizan en vuelos cortos y bajos. Los escarabajos adultos son típicamente depredadores activos de una variedad de artrópodos, por el contrario las larvas se posicionan en la entrada de la galería para esperar a las presas. El ciclo de vida de los escarabajos tigre incluye varias etapas, la primera etapa o huevo, tres estadios de larvas, la pupa y finalmente la emergencia de los adultos; este proceso podría tener lugar entre uno a cuatro años, o incluso durar más tiempo según las especies y condiciones climatológicas (Pearson & Vogler, 2001).

Hay más de 2500 especies de cicindélidos descritos (Pearson & Cassola, 2005), que se distribuyen en todo el mundo (a excepción de Tasmania, la Antártida, y algunas islas oceánicas). Su distribución muestra puntos calientes de riqueza de especies, como el suroeste de los Estados Unidos, el suroeste de Asia, el noreste de la India y el suroeste de la Amazonia (Willis, 1972; Ali, 1978; Pearson, 1984; Pearson & Ghorpade, 1989; Cassola & Pearson, 2000). Además, ocupan una gran diversidad de hábitats en el rango de altitud desde el nivel del mar a los 3500m, incluyéndose en sus habitats praderas alpinas, desiertos, playas arenosas, pastos fangosos, cunetas o caminos forestales (Pearson, 1988; Pearson *et al.,* 1988). Sin embargo, no muestran un patrón homogéneo de distribución con relación a la latitud o la biogeografía (Pearson & Cassola, 2005). Los patrones de distribución geográfica de los escarabajos tigre son importantes para entender mejor su proceso de evolución

(Pearson & Vogler, 2001). Las especies tienden a especializarse en determinados hábitats (Pearson, 1992), y esta dependencia de hábitat está relacionada con la supervivencia de las especies y linajes (Pearson & Vogler, 2001). Los escarabajos tigre tienen varios mecanismos o adaptaciones para reducir la competencia y contribuir a mantener la riqueza de especies; la segregación estacional temporal (ciclos de vida), la segregación espacial (diferencias de hábitat o preferencias de microhábitats) (Willis, 1967) y la separación de su ciclo diario (patrones de actividad diaria), que se han estudiado previamente como las estrategias más comunes.

Los cicindélidos son organismos usados comúnmente como bioindicadores, ya que existe información relativa a su filogenia, su evolución, su biología y sus ciclos de vida. Son útiles en estudios de conservación porque su presencia o ausencia puede proporcionar información sobre la calidad, las alteraciones y etapa de sucesión de los hábitats y son considerados buenos bioindicadores de la biodiversidad general (Pearson, 1992; Pearson & Cassola, 1992; Carroll & Pearson, 1998; Rodríguez *et al.,* 1998; Cassola & Pearson, 2000).

En la península Ibérica se han reportado 22 especies de escarabajos tigre (Serrano, 2013), de los cuales 21 pertenecen a la tribu Cicindelini y uno de ellos a la tribu Megacephalini. Hay un total de cinco géneros: *Cicindela* el más abundante con nueve especies; seguido de *Cephalota* con siete especies; *Cylindera* incluyendo tres especies; *Lophyra* y *Myriochila* siendo representados por una sola especie cada uno. En la península Ibérica los cicindélidos se pueden encontrar en una diversidad de hábitats, como ríos y lagos de agua dulce, en el litoral, o saladares interiores entre otros.

Los escarabajos tigre han despertado la curiosidad de muchos científicos y aficionados y por ello han estudiado profundamente a diferentes niveles: i) morfología (Singh & Gupta, 1982); ii) taxonomía (Horn, 1915, 1926; Wiesner, 1992); iii) biología (Pearson, 1988; Sota, 1994); iv) fisiología (Spangler, 1988; Toh & Okamura, 2001); v) termorregulación (Dreisig, 1980); vi) evolución, ecología y diversidad (Pearson & Vogler, 2001); vii) evolución de cromosomas (Galián *et al.,* 2007); viii) estrategias de conservación (Pearson & Cassola, 1992); y ix) filogenia (Vogler & Pearson, 1996; Vogler & Welsh, 1997; Galián *et al.,* 2002; Vogler *et al.,* 2005; López-López *et al.,* 2012, 2013, 2015).

Sin embargo, hay poca información sobre los transcriptomas de los cicindélidos, y no se han realizado estudios para identificar y caracterizar los genes que podrían estar involucrados en el proceso evolutivo.

Transcriptomas

Hay muchos estudios cuyo objetivo es entender las bases genéticas y ambientales que intervienen en los procesos que pueden conducir a la diferenciación entre las poblaciones y la formación de nuevas especies (Ungerer e*t al.,* 2008; Ren & Siemens, 2010; Kelley *et al.,* 2012; Fan *et al.,* 2012). En este sentido, el avance técnico y científico de la genómica ha permitido: i) identificar las regiones del genoma que se ven afectadas por la selección natural; ii) entender las relaciones evolutivas entre los genes de diferentes especies; y iii) establecer con mayor precisión las relaciones filogenéticas (Mardis, 2008; Stapley, 2010).

Los análisis del transcriptoma y los marcadores de secuencias expresadas (ESTs de sus siglas en inglés "*expressed sequence tags"*) son herramientas importantes para responder a las preguntas evolutivas (Nagaraj *et al.,* 2007). A partir de los análisis de EST podemos seleccionar regiones del genoma que pueden estar sometidas al efecto de la selección, lo que los convierte en una estrategia interesante para estudiar el proceso de adaptación (Bonin, 2008). Las bibliotecas de EST se han utilizado para identificar nuevos genes mediante el mapeo y la anotación de los transcritos, utilizando para ello la información proveniente de otras especies (Liu *et al.,* 2010; Narina *et al.,* 2011; Li *et al.,* 2014). Además, la disponibilidad de una variedad de herramientas bioinformáticas junto con el análisis de expresión génica permiten la caracterización de estos genes y la inferencia de posibles funciones (Ayroles *et al.,* 2009), ya que conocer con detalle aspectos moleculares, celulares y de la función de los genes a menudo resulta esencial para comprender la variación genética y su posible impacto evolutivo (Feder & Mitchell-Olds, 2003).

La cantidad de datos generados a través de los transcriptomas, principalmente en especies modelo, han contribuido a aumentar el conocimiento de los sistemas biológicos. Sin embargo, también es importante el estudio de las especies no-modelo

para comprender la base genética de los rasgos adaptativos. Las nuevas tecnologías en genómica y transcriptómica han avanzado en esta dirección y las especies no modelos son ahora también investigadas a nivel genómico (Elmer & Meyer, 2011). El primer coleóptero cuyo genoma ha sido secuenciado es la especie modelo *Tribolium castaneum* (Richards *et al.,* 2008), perteneciente al suborden Polyphaga. Posteriormente, se han secuenciado transcriptomas de otros coleópteros, en su mayoría polífagos (Keeling *et al.,* 2013; Zhang *et al.,* 2014). La separación de este suborden con respecto al suborden Adephaga se remonta a más de 200 Ma (Hunt *et al.,* 2007). Entre los escarabajos adéfagos, se dispone del transcriptoma de *Pogonus chalceus* (Coleoptera, Carabidae) (Van Belleghem *et al.,* 2012) y algunas genotecas de ADNc de otras especies que se han utilizado en análisis sistemáticos (Theodorides *et al.,* 2002).

El análisis del transcriptoma ha demostrado ser un valioso punto de partida para estudiar las características genéticas de los organismos modelo y no modelo, y ha permitido a los investigadores obtener información sobre las secuencias y sobre los patrones de expresión de genes implicados en diferentes procesos (Mittapalli *et al.,* 2010; Poelchau *et al.,* 2011). Adicionalmente, estudios recientes han demostrado el potencial del transcriptoma para detectar los genes que participan en el proceso de adaptación (Bonin *et al.,* 2009; Renaut *et al.,* 2010). El análisis del transcriptoma se ha utilizado comúnmente para identificar genes implicados en los procesos de inmunidad y reproducción en los insectos (Attardo *et al.,* 2010; Bang *et al.,* 2015; Wei *et al.,* 2015), y varios de estos genes han sido identificados como de rápida evolución en *Drosophila* (Jagadeeshan & Singh, 2005).

Genes de rápida evolución

Los genes de rápida evolución son aquellos que codifican proteínas con una mayor tasa de sustitución de amino-ácidos entre las especies (Swanson & Vaquier, 2002). En muchos casos, esta rápida divergencia es impulsada por la evolución adaptativa, o dicho en otras palabras, se ven afectadas por selección positiva. La selección positiva es el tipo de selección que lleva a la fijación de mutaciones ventajosas (selección

direccional), mantener el polimorfismo (equilibrio de la selección) o la persistencia de un gen recientemente duplicado.

Los genes que evolucionan bajo la selección positiva son significativos desde el punto de vista evolutivo y se pueden identificar usando algoritmos que infieren que los nucleótidos no evolucionan neutralmente (Ford, 2002). Las fuerzas de selección se pueden caracterizar por la comparación de las tasas de sustitución de cambios sinónimos (dS) y no sinónimos (dN) en regiones codificantes. La relación dN/dS igual a 1 indica evolución neutral, dN/dS <1, indica selección negativa o purificadora y cuando dN/dS> 1 indica selección positiva.

Se ha hipotetizado que los genes involucrados en la reproducción y el sistema inmune experimentan selección positiva (Swanson & Vaquier, 2002; Jansa *et al.,* 2003; Bulmer, 2010). Por tanto, la caracterización, el estudio de su función y tipo de selección de estos genes podría revelar información sobre la naturaleza de la evolución adaptativa a nivel molecular.

Genes relacionados con la reproducción

La tasa de evolución de las secuencias de los genes relacionados con el sexo y la reproducción difiere de aquellos genes no relacionados. Muchos genes tanto masculinos como femeninos que participan en el proceso de reproducción sexual han evolucionado rápidamente como resultado de la evolución adaptativa (Begun *et al.,* 2000; Swanson *et al.,* 2001; Swanson & Vacquier, 2002; Zhang *et al.,* 2004; Mueller *et al.,* 2005; Haerty *et al.,* 2007). Hay estudios empíricos y teóricos que indican que la evolución rápida de las proteínas reproductivas podría ser una fuerza impulsora de la especiación (Blows, 1999; Barton, 2000).

Dentro de los genes relacionados con la reproducción, en esta tesis nos centramos específicamente en dos clases: proteínas de fluidos seminales y genes implicados en la gametogénesis, que muestran radiaciones de linajes específicos con evolución acelerada y selección positiva (Haerty *et al.,* 2007). En *D. melanogaster* una gran proporción de las proteínas del fluido seminal (SFP) se ven afectadas por la selección positiva, en comparación con los genes no relacionados, así como también se han encontrado evidencias de evolución adaptativa en varios SFPs (Aguadé, 1999; Begun

et al., 2000; Swanson *et al.,* 2001; Mueller *et al.,* 2005). Evidencias de la selección positiva también se han encontrado en los genes implicados en la gametogénesis, involucrados tanto en la espermatogénesis como en la ovogénesis (Civetta *et al.,* 2006; Bauer Dumont *et al.,* 2007).

Genes relacionados con el sistema inmunitario

Los genes del sistema inmune muestran una tasa más rápida de sustitución de nucleótidos y de aminoácidos (cambios no sinónimos), que los genes no relacionados con la inmunidad, siendo un signo de la evolución adaptativa. La detección de la selección positiva en muchos genes inmunes son una respuesta a la carrera armamentista entre huésped-patógeno, y también puede ser el resultado de las transiciones en el medio ambiente (Bulmer & Crozier, 2006; Lazzaro & Little, 2009; Bulmer, 2010). La evolución de los genes inmunes está marcada por las expansiones y contracciones de las familias de genes (Viljakainen, 2015).

La tasa de evolución difiere entre los distintos componentes del sistema inmune, siendo más común la selección positiva en los genes de reconocimiento o de señalización, que en los genes efectores que fueron analizados en la mosca de la fruta (Schlenke & Begun, 2005; Sackton *et al.,* 2007). Dentro de los genes relacionados con el sistema inmune, nos centramos específicamente en los péptidos antimicrobianos (AMP) que juegan el papel de efector en el sistema inmunológico. Los AMPs se ven afectados por diferentes tipos de selección en función de las especies que analicemos. Los estudios realizados en *Drosophila* no revelaron evidencias de evolución adaptativa en las defensinas (Sackton *et al.,* 2007; Obbard *et al.,* 2009) ni tampoco en mosquitos, que también se ven limitados evolutivamente. Sin embargo, se ha encontrado selección positiva en las defensinas de hormigas y de termitas (Bulmer & Crozier, 2004; Viljakainen & Pamilo, 2008).

*Las referencias se encuentran al final del apartado *Introduction*

Planteamiento, hipótesis y objetivos

El objetivo general de esta tesis es identificar, caracterizar y analizar en un contexto evolutivo genes relacionados con la reproducción y el sistema inmune en cicindélidos para obtener información relevante sobre su biología y comprender mejor su historia evolutiva.

Para conseguir este objetivo se describen dos genotecas de EST de *Cephalota litorea* y *Calomera littoralis,* que servirán como base para la búsqueda y caracterización de genes. En el primer capítulo se realizan análisis bioinformáticos y de expresión diferencial entre machos y hembras, a fin de identificar posibles proteínas del fluido seminal y se ha realizado una búsqueda de homología en otras especies para inferir su tipo de evolución. En el segundo capítulo se realizan análisis para caracterizar dos modificadores post traducción (ubiquitina y los pequeños modificadores similares a ubiquitinas(SUMO)), junto con análisis de selección para determinar su tipo de evolución y de expresión génica para identificar su funcionalidad. En el capítulo 3 se realizan análisis bioinformáticos para identificar características propias de los péptidos antimicrobianos, así como análisis filogenéticos para estimar la evolución y análisis de expresión génica para inferir la función de un gen posiblemente implicado en el sistema inmune de cicindélidos.

Los resultados de la tesis están estructurados en tres capítulos, los cuales se corresponden con un artículo publicado y dos enviados a revistas incluidas en el *Journal Citation Report* (JRC). Los objetivos e hipótesis de cada capítulo se detallan a continuación.

Capítulo 1. Identificación y caracterización de posibles proteínas seminales a partir de genotecas de EST de gónadas masculinas de escarabajo tigre.

El objetivo principal es identificar y caracterizar genes que codifican proteínas que son transferidas a las hembras durante el apareamiento a través del fluido seminal, así como inferir su tasa evolutiva.

Los objetivos específicos son:

-Identificar las características propias de las proteínas del fluido seminal (proteínas secretadas y con expresión diferencial entre machos y hembras) en los genes analizados.

-Caracterizar las posibles funciones de estas proteínas en base a su secuencia utilizando herramientas bioinformáticas.

-Determinar si las proteínas del fluido seminal identificadas son de rápida evolución (específicas de taxón) o de lenta evolución y por tanto conservadas en cicindélidos.

La hipótesis principal propone que las proteínas candidatas seleccionadas en *C. litorea* y *C. littoralis* por cumplir los requisitos de las proteínas secretadas, presentan expresión génica condicionada por el sexo (sesgada a machos) y por tanto son proteínas del fluido seminal. La hipótesis subyacente es que las proteínas seminales identificadas en *C. litorea* y *C. littoralis* presentan una rápida tasa de evolución y por tanto solo están presentes en un taxón o en aquellos que tengan una relación filogenética más cercana.

Capítulo 2. Caracterización y análisis de expresión de los genes UBC9 y UBS27 en gónadas en desarrollo de cicindélidos (Coleoptera: Cicindelidae)

El objetivo principal es identificar y caracterizar genes que codifican modificadores post-traducción, como es el caso de las ubiquitinas y de los pequeños modificadores similares a ubiquitinas (genes SUMO), así como analizar su posible implicación en el proceso de la gametogénesis en *Cicindela campestris*.

Los objetivos específicos son:

-Identificar y obtener la secuencia completa de estos genes para su caracterización en cicindélidos.

-Determinar el tipo de selección al que estos genes están sometidos en cicindélidos.

-Determinar su implicación en la gametogénesis realizando para ello análisis de expresión en gónadas en distintas etapas de desarrollo, tanto en hembras como en machos.

La hipótesis principal indica que las proteínas identificadas mediante herramientas bioinformáticas como ubiquitina (UBC9) y SUMO (UBS27) en *C. campestris* están involucrados en la gametogénesis (espematogéneis y oogénesis) y por tanto se espera encontrar una mayor expresión en gónadas. La hipótesis subyacente indica que los genes involucrados en la gametogénesis, y por tanto en el proceso reproductivo en cicindélidos, presentan una tasa de evolución mayor que aquellos genes no relacionados con este proceso, y por tanto estarán sometidos a selección positiva.

Capítulo 3. Identificación, caracterización estructural y análisis de expresión del gen de la defensina en el escarabajo tigre *Calomera littoralis* (Coleoptera: Cicindelidae)

El objetivo principal de este capítulo es identificar y caracterizar el gen de la defensina en los cicindélidos, analizar su implicación en el sistema inmune y estimar su historia evolutiva.

Los objetivos específicos son:

-Identificar y obtener la secuencia completa del gen defensin para su caracterización en cicindélidos.

-Determinar su posible función antimicrobiana mediante análisis bioinformáticos

-Determinar la respuesta de este gen tras la infección de individuos de *C. littoralis* con lipopolisacáridos de *Escherichia coli* realizando para ello análisis de expresión.

-Inferir su evolución a partir de la filogenia realizada con secuencias de otros insectos y en particular coleópteros.

La hipótesis principal indica que la proteína identificada mediante herramientas bioinformáticas como defensina en *C. littoralis* está involucrada en la respuesta inmune y cumple las características estructurales para ejercer función antimicrobiana. La hipótesis subyacente indica que al ser las defensinas proteínas de rápida evolución sus secuencias presentan una elevada divergencia en relación a otras especies de coleópteros. Las predicciones de esta hipótesis indican que la defensina de *C. littoralis* y las de otros coleópteros no formarán clados congruentes con sus relaciones filogenéticas.

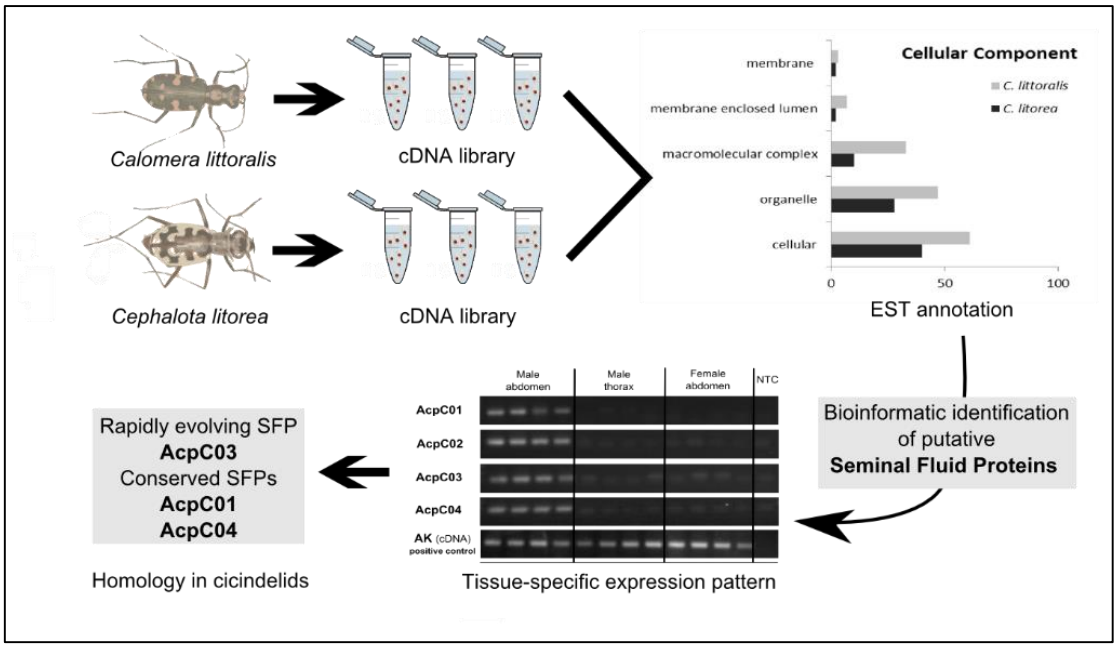
Resúmenes

Capítulo 1. Identificación y caracterización de posibles proteínas seminales a partir de genotecas de EST de gónadas masculinas de escarabajo tigre.

El estudio de las proteínas transferidas a través del semen puede proporcionar información importante para resolver cuestiones biológicas como la evolución adaptativa, el origen de nuevas especies y la riqueza de especies. El objetivo de este trabajo fue identificar proteínas del fluido seminal (SFP), que puedan contribuir al estudio del sistema reproductivo de los escarabajos tigre (cicindélidos), un grupo de más de 2500 especies distribuidas por todo el mundo y que ocupan una gran diversidad de hábitats.

Se construyeron dos genotecas de ADNc a partir de las gónadas de machos de *Calomera littoralis* y *Cephalota litorea*. Los marcadores de secuencias expresadas (ESTs) fueron analizados con herramientas bioinformáticas y se seleccionaron 14 unigenes como candidatos a SFP, se realizó una reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR) para identificar patrones de expresión específicos de tejido en machos. Se han identificado cuatro nuevas posibles SFPs de cicindélidos, para las cuales las búsquedas de similaridad no mostraron homólogos. Sin embargo, dos de las clases de proteínas (respuesta inmune y hormona) predichas con Protfun son similares a las reportadas en otros insectos. Las búsquedas de homología en otros cicindélidos mostraron una SFP específica de linaje (proteínas de rápida evolución), solo presente en las especies cercanas de *C. littoralis* y *Lophyra flexuosa* y dos SPFs conservadas que están presentes en todos las especies de escarabajo tigre analizadas.

Este trabajo supone la primera caracterización de posibles SFPs en una especie adéfaga del orden Coleoptera. Los resultados servirán como fundación para futuros estudios encaminados a comprender las funciones del gen (y de las proteínas) y sus implicaciones evolutivas en este grupo de escarabajos ecológicamente relevantes.

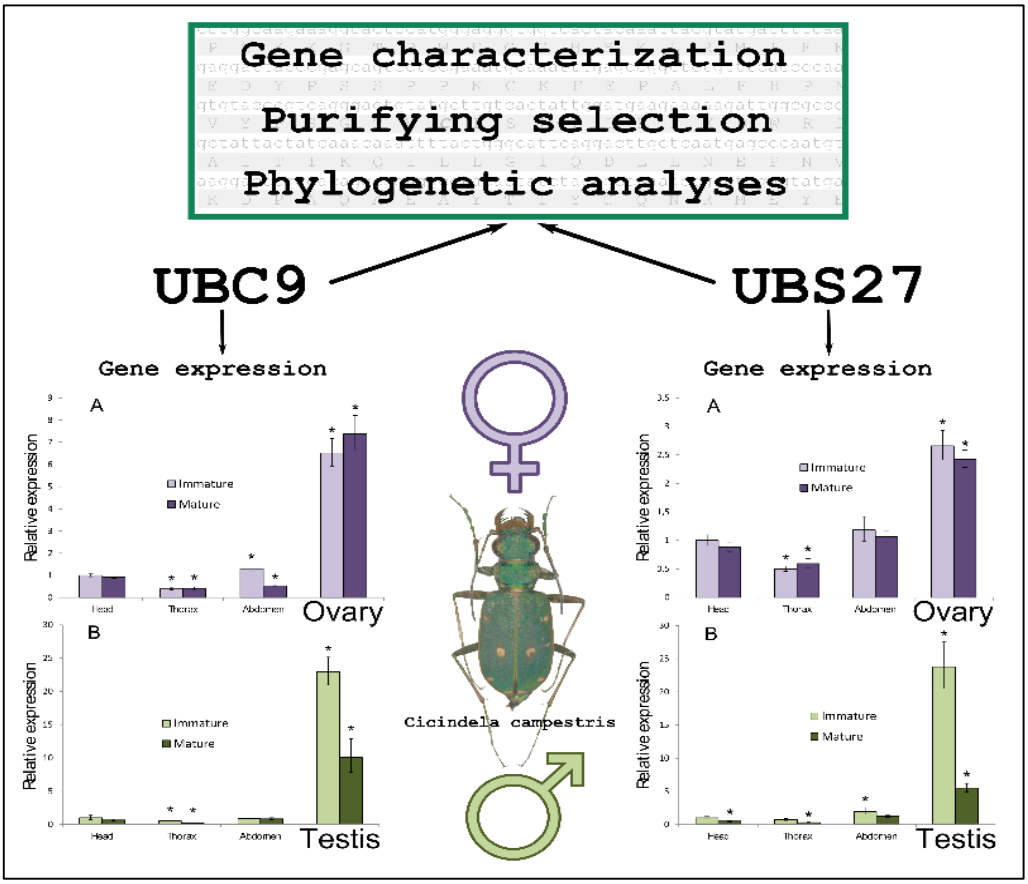


Capítulo 2. Caracterización y análisis de expresión de los genes UBC9 y UBS27 en gónadas en desarrollo de cicindélidos (Coleoptera: Cicindelidae)

Las ubiquitinas y los pequeños modificadores similares a ubiquitinas (genes SUMO) son modificadores post traducción esenciales en una variedad de procesos celulares, incluida la gametogénesis. La enzima de conjugación E2 SUMO (UBC9) y una proteína de fusión ubiquitina-ribosómica (UBS27) han sido caracterizadas en varias especies modelos, sin embargo su expresión en coleópteros no ha sido estudiada.

En este trabajo han sido caracterizados por primera vez los genes UBC9 y UBS27 en el escarabajo tigre *Cicindela campestris*. Los análisis bioinformáticos mostraron que el gen Cc-UBC9 codifica una proteína de 159 aminoácidos con un peso molecular esperado de 18,18 KDa, el gen Cc-UBS27codifica una proteína de 156 aminoácidos y un peso molecular de 17,71 KDa. Los análisis de selección realizados en varias especies de cicindelas mostraron que ambos genes estás afectados por selección purificadora. Los análisis de PCR cuantitativa a tiempo real mostraron que los genes Cc-UBC9 y Cc-UBS27 se expresan en los distintos tejidos analizados. Los mayores niveles de expresión se encontraron en ovario y testículo, y se encontraron niveles de expresión diferencial entre el estadio maduro e inmaduro de los testículos. Los

niveles de expresión de los genes Cc-UBC9 y Cc-UBS27 sugieren que estos genes juegan un papel importante en la gametogénesis en *C. campestris*. Esta información es relevante para comprender mejor el proceso reproductivo en cicindélidos así como la función de los genes de ubiquitinas y SUMO en los coleópteros.



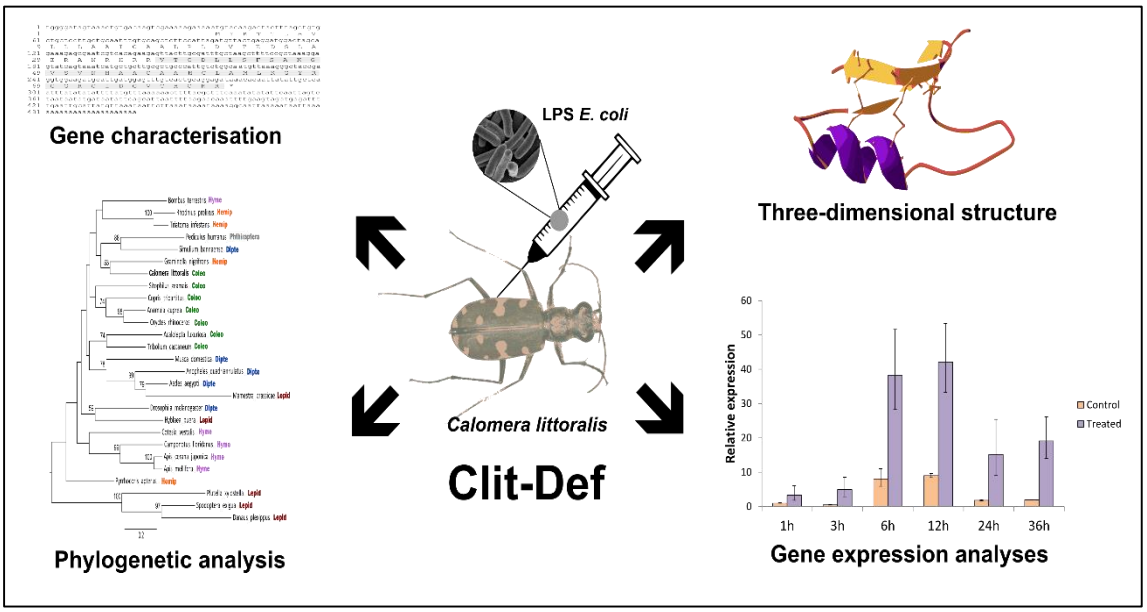
Capítulo 3. Identificación, caracterización estructural y análisis de expresión del gen de la defensina en el escarabajo tigre *Calomera littoralis* (Coleoptera: Cicindelidae)

Las defensinas son importantes péptidos antimicrobianos que juegan un papel crucial en el sistema inmune innato de insectos. Son predominantemente activas contra bacterias Gram (+), y también contra Gram (-), hongos y algunos virus. Las defensinas han sido caracterizadas en diversas especies de coleópteros, sin embargo no existe ningún reporte en el suborden Adephaga.

En este trabajo, ha sido caracterizado por primera vez un gen de la defensina en el escarabajo tigre *Calomera littoralis*, y ha sido designado como Clit-Def. Los análisis

bioinformáticos mostraron que el gen Clit-Def tiene un marco abierto de lectura de 246 pb, el cual contiene un péptido maduro de 46pb. El péptido maduro de Clit-Def tiene las características propias de la función antimicrobiana: un punto isoeléctrico catiónico de 8,94, seis residuos de cisteína que forman tres enlaces disulfuro, y la estructura típica CSαβ. Los análisis de PCR cuantitativa en tiempo real mostraron que Clit-Def se sobreexpresa en los diferentes tejidos analizados después de la infección con lipopolisacáridos de *Escherichia coli*, y también se indica que tiene un pico de expresión tras 12 h de infección. Los análisis filogenéticos mostraron una alta variabilidad en las defensinas de los coleópteros analizados.

Los patrones de expresión de Clit-Def sugieren que este gen desempeña un papel importante en el sistema humoral de *C. littoralis*. La información obtenida sobre la defensina de un cicindélido permitirá comprender mejor la estructura y la función del sistema inmune en coleopteros.



Comentarios generales y conclusiones

En esta tesis se han identificado y caracterizado por primera vez genes relacionados con la reproducción y el sistema inmune en cicindélidos, uno de los grupos de escarabajos más estudiados, de las que sin embargo la información relativa a su transcriptoma es muy limitada. A partir del análisis de genotecas de EST de distintos escarabajos tigre (*Calomera littoralis, Cephalota litorea y Cicindela campestris*) se han identificado y caracterizado cuatro proteínas del fluido seminal (Acp01-Acp04), la enzima de conjugación E2 SUMO (Cc-UBC9) y una proteína de fusión ubiquitina-ribosómica (Cc-UBS27), todos ellas involucradas en el proceso reproductivo. También se ha caracterizado una defensina (Clit-Def) involucrada en el sistema inmune. Los resultados obtenidos de la caracterización de la secuencia y su patrón de expresión se han analizado en un contexto evolutivo.

La descripción de estos genes proporciona información relevante sobre la biología de los cicindélidos, y también sirve de base para entender mejor su proceso de diversificación, lo que conduce a una mayor comprensión de sus relaciones filogenéticas y aporta indicios sobre los procesos de especiación.

Conclusiones

Capítulo 1. Identificación y caracterización de posibles proteínas seminales a partir de genotecas de EST de gónadas masculinas de escarabajo tigre.

1. De las 14 proteínas seleccionadas como candidatas cuatro de ellas cumplen los criterios de las proteínas del fluido seminal, son secretadas y se expresan fundamentalmente en machos.

2. Las búsquedas de homología con otras especies para inferir función en las cuatro proteínas del fluido seminal caracterizadas no son efectivas debido a que, o bien no se encuentra homología, o bien se obtiene homología con proteínas cuya función no ha sido anotada.

3. Las predicciones bioinformáticas de función a partir de la secuencia y estructura de las proteínas, para dos de estas proteínas corroboran que AcpC01 (sistema inmune) y AcpC03 (hormona) son proteínas seminales debido a que presentan funciones ya descritas previamente para estas proteínas.

4. AcpC03 únicamente amplificó en *C. littoralis* y *Lophyra flexuosa* especies que guardan una estrecha relación filogenética y por tanto se trataría de una proteína de rápida evolución.

5. AcpC04 amplifica en todas las especies de cicindélidos analizadas, por lo que se trataría de una proteína conservada al menos en los escarabajos tigre.

6. AcpC01 amplifica en todas las especies de cicindélidos analizadas excepto en *Cephalota maura* que es la especie filogenéticamente más alejada del resto por lo que se trataría de una proteína conservada en al menos un grupo de escarabajos tigre.

Capítulo 2. Caracterización y análisis de expresión de los genes UBC9 y UBS27 en gónadas en desarrollo de cicindélidos (Coleoptera: Cicindelidae)

7. Tras la anotación de los EST, y la posterior caracterización de su secuencia completa se concluye que el gen descrito como Cc-UBC9 en *C. campestris* codifica una enzima de conjugación E2 SUMO y el gen descrito como Cc-UBS27 codifica una proteína de fusión ubiquitina-ribosómica.

8. Las secuencias de aminoácidos y los análisis filogenéticos demuestran que ambos genes son altamente conservados tanto en cicindélidos como en el resto de coleópteros.

9. Los análisis de selección revelaron que ambos genes están sometidos a selección purificadora los que pondría de manifiesto la importancia de la función de estas proteínas.

10. Ambos genes muestran mayores niveles de expresión en gónadas en comparación con el resto de tejidos analizados, lo que indica que estos genes juegan un papel activo en la gametogénesis de cicindélidos.

11. La mayor expresión de estos genes en testículo inmaduro en relación al estadio maduro, indica que estos genes están involucrados en mayor medida en las etapas tempranas de la gametogénesis.

Capítulo 3. Identificación, caracterización estructural y análisis de expresión del gen de la defensina en el escarabajo tigre *Calomera littoralis* (Coleoptera: Cicindelidae)

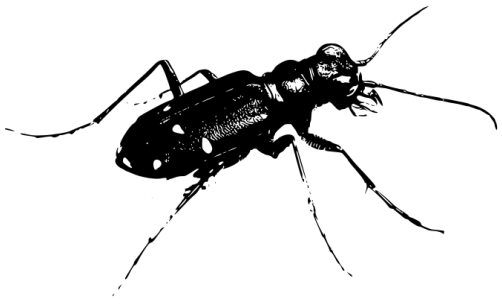
12. El gen caracterizado como Clit-Def en *Calomera littoralis* cumple las características estructurales para ser una defensina.

13. Los análisis estructurales y bioinformáticos sugieren que Clit-Def tiene capacidad antimicrobiana y por tanto está involucrada en el sistema inmune.

14. Clit-Def en *C. littoralis* se sobreexpresa en diferentes tejidos, principalmente en abdomen, tras la infección con lipopolisacáridos de *Escherichia coli*, lo que sugiere que este gen juega un papel importante en la respuesta humoral.

15. El análisis filogenético de la defensina de este trabajo junto con las de otros coleópteros obtenidas de GenBank no se agrupan en clados congruentes con la taxonomía lo que indica que se trata de proteínas de alta diversidad.

Introduction



Background

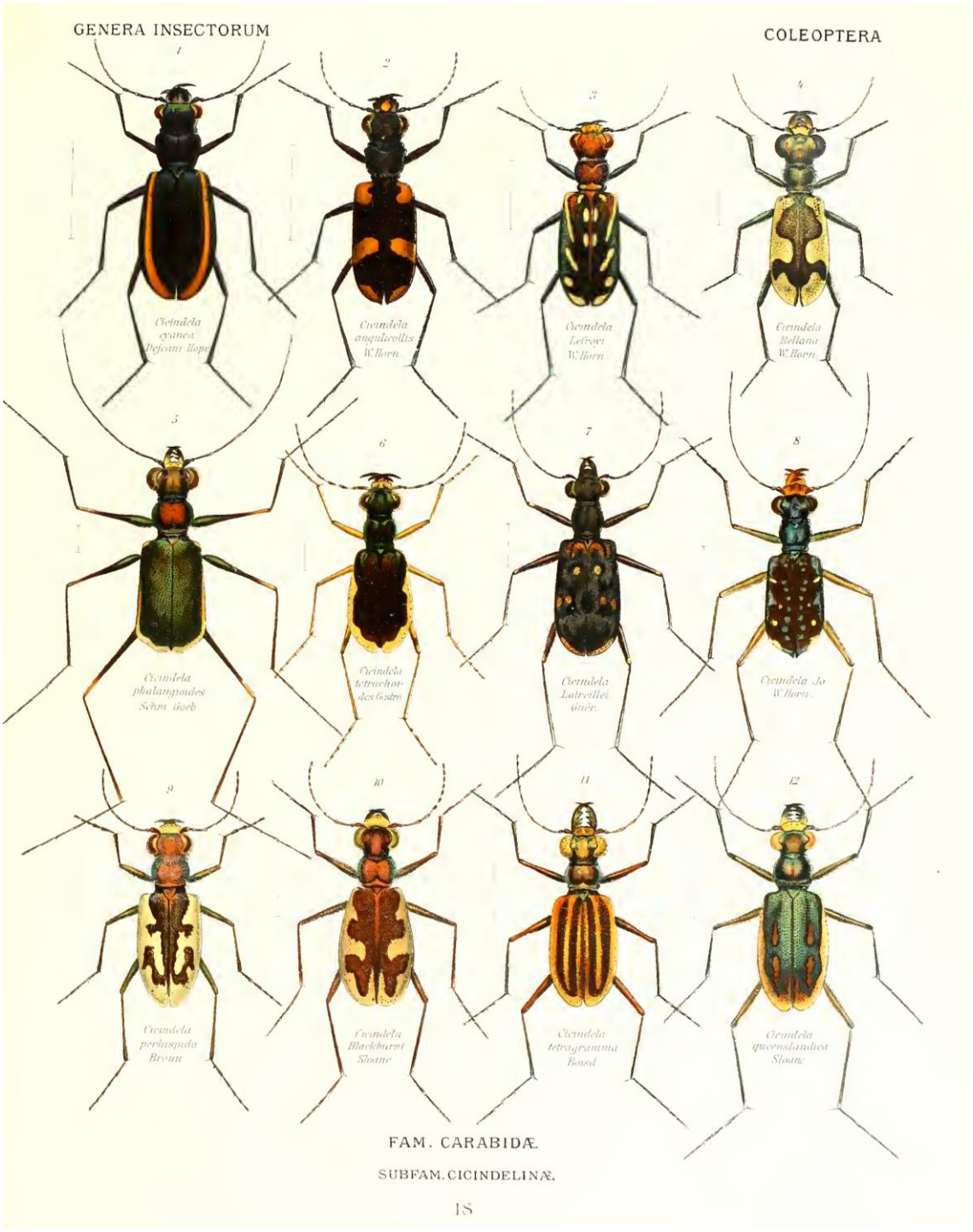
Tiger beetles

Tiger beetles or cicindelids (Cicindelidae) form a well defined group within the Adephaga, a suborder of Coleoptera, which includes conspicuous, brightly coloured, non-pest species that are significant components of ecosystems, being important links in food chains. Adephaga is the second largest suborder of Coleoptera , with over 36,000 known species (Nilsson, 2001; Lorenz, 2005), that includes tiger beetles (Cicindelidae), ground beetles (Carabidae), the predaceous diving beetles, (Dytiscidae), the whirligig beetles (Gyrinidae), the water beetles (Haliplidae), among other families. There is a controversy in relation to the classification of cicindelids, which have been considered a subfamily (Cicindelinae), a tribe (Cicindelini), or a supertribe (Cicindelitae) within the family Carabidae (Ball, 1979). Some phylogenetic works including the suborder Adephaga (Shull *et al.*, 2001; Hunt *et al.*, 2007), also showed cicindelids as a clade inside the family Carabidae. Nevertheless, more recent studies showed the monophyly of cicindelids outside of carabids and support their family status (Cicindelidae) (Putchkov & Cassola, 2005; Maddison *et al.*, 2009; Bocak *et al.,* 2014). Cicindelids history have been characterised by an increase of diversity, considering both morphological and taxonomical diversity, that begun with carabid-like ancestors (*Omus, Manticora, Megacephala*), continue with the arborean groups (Collyridini) and the most derived groups in the genus *Cicindela* adapted for running and flying (Pearson & Vogler, 2001).

In evolutionary research is important having a well-defined phylogenetic context as a tool for contributing to the knowledge of functional and evolutionary process of genes by comparative method (Jansa *et al.,* 2003). There are several phylogenetic studies focused on identifying the evolutionary relationships in Cicindelids between different groups (Galián *et al.,* 2002) and particular groups (Cardoso & Vogler, 2005; Vogler *et al.,* 2005; Pons *et al.,* 2004, 2006; López-López *et al.,* 2012, 2013, 2015).

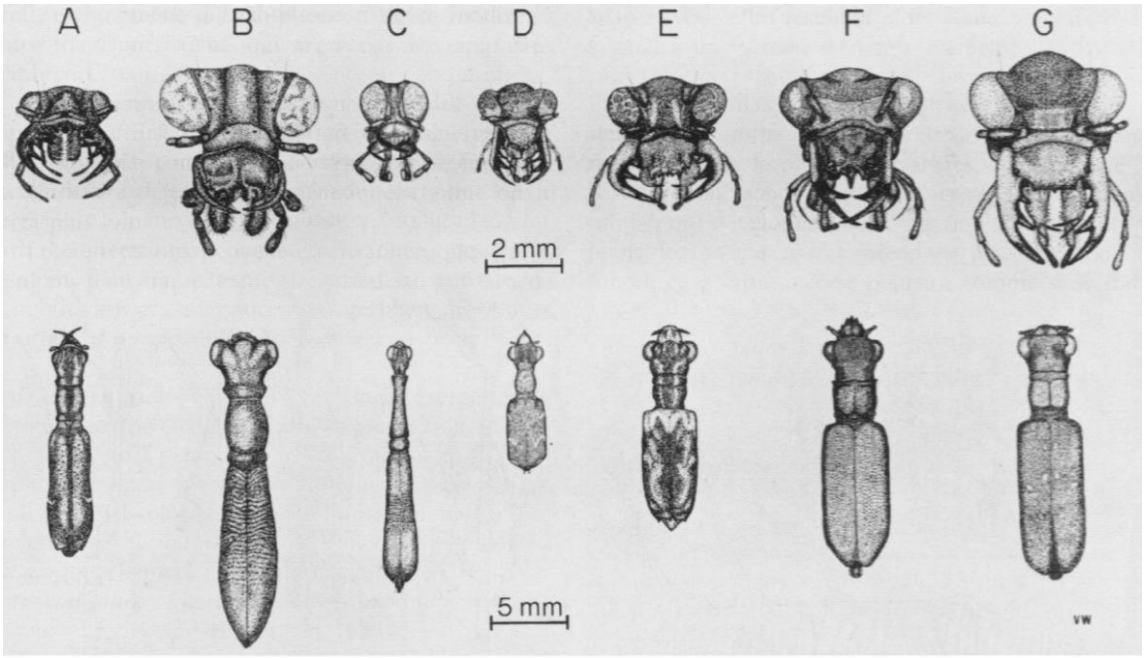
There are several characteristics that tiger beetles normally share: i) long sickle-shaped mandibles; ii) several kind of teeth, simple teeth and a compound tooth; iii) relatively thin and long cursorial legs; iv) long body form, with eyes and head that

together are wider than the thorax; and v) tunnels construction behavior in larva (Pearson, 1988; Pearson & Vogler, 2001).



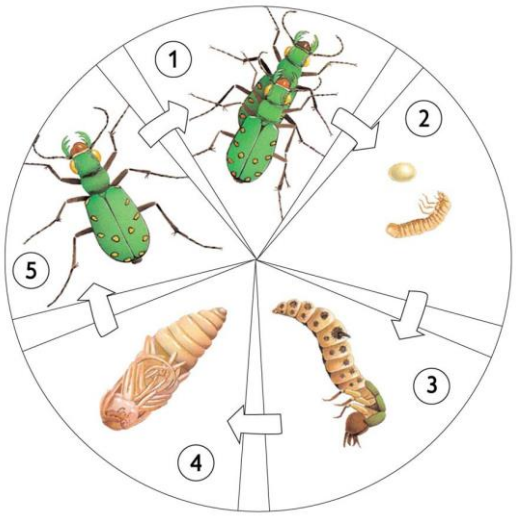
Twelve species of tiger beetles representing varios degrees of elytral color, pattern, body size and shape. From Horn (1915).

Adult tiger beetles vary in size (from 5 mm to more than 4 cm the largest one) and color (many are black, but some have brilliant colors as green, blue, violet, red, yellow). Their body shapes vary from robust to more stylish forms, and have functional transparent wings that they normally use in short and low flights. Adult beetles are typically cursorial predators of a variety of arthropods, in contrast, larvae are ambush predators that position themselves at the entrance of gallery burrows and wait for preys.



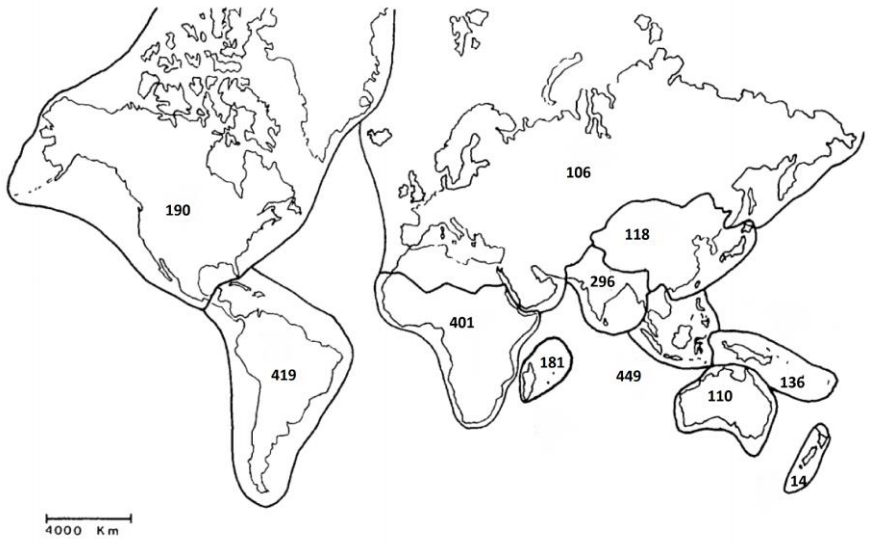
Dorsal view of the head and body from species of seven genera of tiger beetles from various biogeographical regions. A, *Ctenosoma dormer* (Neotropical); B, *Tricondyla aptera* (Australian); C, *Neocollorys constricticollis* (Oriental); D, *Pentacomia egregia* (Neotropical); E, *Therates basalis* (Oriental); F, Odontocheila cayennensis (Neotropical); G, *Cicindela xanthophila* (Ethiopian). From Pearson (1980)

The life cycle of tiger beetles includes several stages. After fertilization females lays eggs under the surface of the earth. Depending species, temperature and habitat, the larva hatches 9 to 38 days after oviposition. There are three larvae instars before the formation of the pupa, then takes place the metamorphosis and finally the adult emergence; this process could take place in one to four years, or even longer (Pearson & Vogler, 2001).



Graphical representation of tiger beetle life cycle: adult copula (1); ovo stage (2); larva (3); pupa (4) and adult (5) From Q-files. The Great Ilustrated Ecyclopedia (https://www.q-files.com/life/insects/beetles/).

More than 2500 species of cicindelid beetles have been described (Pearson & Cassola, 2005), which are distributed worldwide (except for Tasmania, Antarctica, and some Oceanic islands). Their distribution shows hot spots of species richness, such as the southwestern United States, southwestern Asia, northeastern India and southwestern Amazonia (Willis, 1972; Ali, 1978; Pearson, 1984; Pearson & Ghorpade, 1989; Cassola & Pearson, 2000). The geographical distribution of tiger beetles is important to better understand their evolution process (Pearson & Vogler, 2001).



Map of the world showing 11 biogeographical regions and the number of cicindelids species in each one. From Pearson & Vogler (2001).

Additionally, they occupy a great diversity of habitats that range from sea level to 3500 m in altitude, including alpine meadows, desert grassland, sandy beaches, muddy pastures, roadside ditches, or forest paths (Pearson, 1988; Pearson *et al.,* 1988). Nevertheless, they do not show a homogenous pattern of distribution with relation to latitude or biogeography (Pearson & Cassola, 2005). Species tend to specialise in particular habitats (Pearson, 1992), and these habitat dependence is related to the survival of species and lineages (Pearson & Vogler, 2001).

Sampled places. From the upper left to the lower: Salinas del Rasall, (Murcia, Spain); Rambla Salada (Murcia, Spain); Laguna de Pinilla (Albacete, Spain); Laguna del Arquillo (Albacete, Spain); and Laguna de Pétrola (Albacete, Spain). Photos: Alejandro López, María Juliana Rodríguez.

They have several mechanisms or adaptations to reduce competition and contribute to maintain species richness; seasonal temporal segregation (life cycles), spatial segregation (differences in habitat or microhabitats preferences) (Willis, 1967) and temporal partitioning on the diel scale (diel activity patterns), which have been previously studied as the most common strategies.

Cicindelids are organisms that are commonly used as bioindicators, thanks to their information concerning to their phylogeny, evolution, biology and life cycles. They are useful in conservation studies as their presence or absence can provide information on the quality, alterations and successional stage of habitats and they are considered to be good bioindicators of general biodiversity (Pearson, 1992; Pearson & Cassola, 1992; Carroll & Pearson, 1998; Rodríguez *et al.,* 1998; Cassola & Pearson, 2000).

In the Iberian Peninsula there are reported 22 species of tiger beetles (Serrano, 2013), of which 21 belong to the tribe Cicindelini and one to the tribe Megacephalini. There are a total of five genus: *Cicindela* is the most abundant with nine species; *Cephalota* with seven species; *Cylindera* including three species; *Lophyra* and *Myriochila* being represented by a single species each. In the Iberian Peninsula cicindelids can be found in a diversity of habitats around the geography as rivers and freshwater lakes, in the littoral or in salt marshes among others.

Tiger beetles studied. From left to right: *Cephalota litorea*; *Calomera littoralis* and *Cicindela campestris*. Photos: Pedro Requena Díaz (https://www.flickr.com/photos/alcedofoto/), Giacomo Radi (https://www.flickr.com/photos/61626776@N03/), Jürgen Mangelsdorf (https://www.flickr.com/photos/22084572@N07/)

Tiger beetles have aroused the curiosity of many scientists and amateurs and therefore they have been deeply studied at different levels: i) morphology (Singh & Gupta, 1982); ii) taxonomy (Horn, 1915, 1926; Wiesner, 1992); iii) biology (Pearson, 1988; Sota, 1994); iv) physiology (Spangler, 1988; Toh & Okamura, 2001); v) thermoregulation (Dreisig, 1980); vi) evolution, ecology and diversity (Pearson & Vogler, 2001); vii) chromosome evolution (Galián *et al.,* 2007); viii) conservation strategies (Pearson & Cassola, 1992); and ix) phylogenetics (Vogler & Pearson, 1996; Vogler & Welsh, 1997; Galián *et al.,* 2002; Vogler *et al.,* 2005; López-López *et al.*, 2012, 2013, 2015). Nevertheless, there is little information on the transcriptomes of cicindelids, and no studies have been conducted to identify and characterise genes that could be involved in the evolutionary process.

Transcriptome studies

There are many studies with the aim to understand the genetic and environmental bases involved in the processes that can lead to differentiation among populations and the formation of new species (Ungerer *et al.,* 2008; Ren & Siemens, 2010; Fan *et al.,* 2012; Kelley *et al.,* 2012). Thus, technical and scientific advance of genomics have allowed to: i) identify regions of the genome that are affected by natural selection; ii) understand the evolutionary relationships among genes of different species; and iii) establish more accurately the phylogenetic relationships (Mardis, 2008; Stapley, 2010).

Transcriptome and expressed sequence tags (ESTs) analyses are important tools for answering evolutionary questions (Nagaraj *et al.,* 2007). ESTs approaches are an interesting strategy to study the adaptation process since it uses genome regions that most likely undergo selection effect (Bonin, 2008). EST libraries have been used to identify new genes, by mapping and annotating transcripts, using information generated in projects from other species (Liu *et al.,* 2010; Narina *et al.,* 2011; Li *et al.,* 2014). In addition, the availability of a variety of bioinformatic tools together with gene expression analyses allow the characterisation of these genes and the inference

of putative functions (Ayroles *et al.,* 2009), as knowing the detail about molecular, cellular and organismal function of the genes is often essential to understand the genetic variation and its evolutionary impact (Feder & Mitchell-Olds, 2003).

The substantial amount of data generated by the study of transcriptomes, mainly in model species, have contributed to increase the knowledge of biological systems. Nevertheless, the study of non-model species is also important to understand the genetic basis of adaptive traits. Modern genomic and transcriptomic technologies have advanced in this direction and non-traditional model systems can now be investigated at genomic level (Elmer & Meyer, 2011). The first coleopteran for which the genome has been sequenced was the model species *Tribolium castaneum* (Richards *et al.,* 2008) belonging to the suborder Polyphaga. Subsequently, transcriptomes from others coleoptera, mostly polyphagan species have become available (Keeling *et al.,* 2013; Zhang *et al.,* 2014). The separation of this suborder to the suborder Adephaga dates back more than 200 Ma (Hunt *et al.,* 2007). In adephagan beetles, the transcriptome of *Pogonus chalceus* has become available (Coleoptera, Carabidae) (Van Belleghem *et al.,* 2012) and some cDNA libraries which have been used in systematic analyses (Theodorides *et al.,* 2002).

The transcriptome analysis has proven to be a valuable first step to study genetic characteristics in both model and non-model organisms and has allowed researchers obtaining sequence information and expression patterns of genes involved in different processes (Mittapalli *et al.,* 2010; Poelchau *et al.,* 2011). Recent studies have demonstrated the potential of transcriptome of different populations to detect genes involved in the adaptation process (Bonin *et al.*, 2009; Renaut *et al.,* 2010). In addition, transcriptome analysis have been commonly used to identify genes involved in immunity and reproduction processes in insects (Attardo *et al.,* 2010; Bang *et al.,* 2015; Wei *et al.,* 2015) and several of these genes have been identified as rapidly evolving in *Drosophila* (Jagadeeshan & Singh, 2005).

Rapidly evolving genes

Rapidly evolving genes are those that encode proteins with a higher rate of amino-acid substitutions between species (Swanson & Vaquier, 2002). In many cases, this rapid divergence is driven by adaptive evolution, in other words, they are affected by positive Darwinian selection. Positive selection is the type of selection leading to fixation of advantageous mutations (directional selection), maintain polymorphism (balancing selection) or the persistence of a recently duplicated gene.

Genes that appear to evolve under positive selection are evolutionary significant and can be identified using algorithms that infer which nucleotides evolve non- neutrally (Ford, 2002). The selection forces can be characterized by comparing the synonymous (dS) and non-synonymous (dN) substitution rates in coding regions. The dN/dS ratio of 1 indicates neutral evolution, dN/dS < 1, purifying selection or negative selection and dN/dS > 1 indicates positive selection.

Genes associated with reproduction and the immune response are hypothesized to have experienced long term positive selection (Swanson & Vaquier, 2002; Jansa *et al.,* 2003; Bulmer, 2010). Therefore the study of their function and evolution could reveal much about the nature of adaptive evolution at molecular level.

Sex and reproduction-related genes

The rate of sequence evolution from sex and reproduction-related (SRR) genes differs from non-reproductive genes. Many male and female genes involved in the sexual reproduction process evolved rapidly as a result of adaptive evolution (Begun *et al.,* 2000; Swanson *et al.,* 2001; Swanson & Vacquier, 2002; Zhang *et al.,* 2004; Mueller *et al.*, 2005; Haerty *et al.,* 2007). There are empirical and theoretical studies that indicate that the rapid evolution of reproductive proteins could be a driving force in speciation (Blows, 1999; Barton, 2000).

Within the reproduction-related genes in this study we focus specifically in two classes: seminal fluid proteins and genes involved in gametogenesis that show lineage-specific bursts of accelerated evolution and positive selection (Haerty *et al.,*

2007). In *D. melanogaster* a great proportion of **seminal fluid protein** (SFP) genes are affected by positive selection in comparison to non-reproductive genes and evidence for adaptive evolution has been found at several loci encoding SFPs, from an evolutionary perspective (Aguadé, 1999; Begun *et al.,* 2000; Swanson *et al.*, 2001; Mueller *et al.,* 2005). Evidence of positive selection has also been found for **genes involved in gametogenesis**, controlling both spermatogenesis and oogenesis (Civetta *et al.,* 2006; Bauer DuMont *et al.,* 2007).

Immune system-related genes

Immune genes typically show a faster rate of nucleotide and amino acid substitutions (non-synonymous) than non-immune-related genes as a sign of adaptive evolution. The detection of positive selection in many immune genes is a response to host-pathogen arm race, and may be also the result of evolutionary transitions in environment (Bulmer & Crozier, 2006; Lazzaro & Little, 2009; Bulmer, 2010). The evolution of immune genes is marked by expansions and contractions of gene families (Viljakainen, 2015).

The rate of evolution differs across the components of the immune system, being positive selection more common in recognition than signaling or effector proteins analysed in fruitflies (Schlenke & Begun, 2005; Sackton *et al.,* 2007). Within the immune related genes in this study we focus specifically in the **antimicrobial peptides (AMPs)** that plays the role of effectors in the immune system. AMPs are affected by different kind of selection across different organisms. Studies in defensins in *Drosophila* have not revealed evidence of adaptive evolution, as it happens in mosquitoes, which are also evolutionarily constrained (Sackton *et al.,* 2007; Obbard *et al.,* 2009). Nevertheless, positive selection has been found in ants and termites defensins (Bulmer & Crozier, 2004; Viljakainen & Pamilo, 2008).

Planning, objectives and hypotheses

The objective of this thesis is to identify, characterize and analyze reproductive and immune related genes in cicindelids from an evolutionary perspective to obtain relevant information on their biology to better understand their evolutionary history.

To achieve this objective, two EST libraries from *Calomera littoralis* and *Cephalota litorea* are described, which serve as a foundation for searching and characterizing genes. In the first chapter, bioinformatics and differential expression analyses between males and females are performed in order to identify putative seminal fluid proteins. In addition, homology searches in other species are performed to infer the type of evolution. In the second chapter analyses to characterize two post-translation modifiers (ubiquitin and small ubiquitin-like modifiers) are performed, as well as a selection analysis to determine the type of evolution and gene expression analyses to identify their function. In the third chapter, bioinformatic analyses are performed to identify characteristics typical of antimicrobial peptides, a phylogenetic analysis to estimate evolution and gene expression analyses to infer the function of a gene possibly involved in the immune system of cicindelids are also performed.

The results of this thesis are included into these three chapters, which correspond to a published article and two manuscripts submitted to journals included in the Journal Citation Report (JRC). The objectives and hypotheses of each chapter are detailed below.

Chapter 1. Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles.

The main objective of this study is to identify and characterise genes encoding proteins that are transferred to females during mating through seminal fluid, and infer their evolutionary rate.

The particular objectives are:

-To identify the common features of seminal fluid proteins (secreted and male-biased proteins)

-To characterise the possible functions of these proteins based on their sequence using bioinformatics tools.

-To determine if the seminal fluid proteins identified are rapidly evolving (taxon specific) or conserved in cicindelids.

The initial hypothesis proposes that selected candidate proteins in *C. littoralis* and *C. litorea* meet the requirements of secreted proteins, present male-biased gene expression and thus, are seminal fluid proteins. Another hypothesis postulates that the seminal proteins identified in *C. littoralis* and *C. litorea* have a rapid evolutionary rate and therefore they are only present in a taxon or those taxa with a close phylogenetic relationship.

Chapter 2. Characterisation and expression analysis of UBC9 and UBS27 genes in developing gonads of cicindelids (Coleoptera: Cicindelidae)

The main objective is to identify and characterize genes encoding post-translational modifiers, as ubiquitin and small ubiquitin-like modifiers (SUMO genes), and to analyze their possible involvement in the gametogenesis process in *Cicindela campestris*.

The particular objectives are:

-To identify and to obtain the complete sequence of these genes for their characterization in cicindelids.

-To determine the type of selection that these genes have undergone in cicindelids.

-To determine their involvement in the gametogenesis performing expression analysis in developing gonads in females and males of *C. campestris*.

The initial hypothesis postulates that the proteins identified using bioinformatics tools as ubiquitin (UBC9) and SUMO (UBS27) in *C. campestris* are involved in the gametogenesis (spermatogenesis and oogenesis) and therefore a high expression in gonads is expected. A subjacent hypothesis indicates that genes involved in gametogenesis, and hence in the reproductive process in cicindelids, have an

evolutionary rate higher than those genes non-relatedwith this process, and consequently are subjected to positive selection.

Chapter 3. Identification, structural characterisation and expression analysis of a defensin gene from the tiger beetle *Calomera littoralis* (Coleoptera: Cicindelidae)

The main objective of this chapter is to identify and characterise a defensin gene in cicindelids, to analyse their involvement in the immune system and to estimate their evolutionary history.

The particular objectives are:

-To identify and to obtain the complete sequence of the defensin gene Clit-Def for its characterisation in *C. littoralis*.

-To determine their likely antimicrobial function using bioinformatic analysis.

-To determine the response of this gene by expression analysis in individuals of *C. littoralis* after infection with lipopolysaccharide from *Escherichia coli*.

-To infer their evolution from the phylogeny reconstructed with sequences from other insects, particularly beetles.

The initial hypothesis proposes that the protein identified by bioinformatic tools as defensin in *C. littoralis*, is involved in the immune response and meets the structural characteristics for carrying out its antimicrobial function. Another hypothesis postulates that defensins as rapidly evolving genes their sequences show a high divergence in beetle species. The predictions of this hypothesis indicate that defensin from *C. littoralis* and from other beetle species will not form clades consistent with their phylogenetic relationships.

References

Aguadé, M. (1999) Positive selection drives the evolution of the Acp29AB accessory gland protein in *Drosophila*. *Genetics* **152**: 543-551.

Ali, H. (1978) Faunistic study of the Cicindelidae (Coleoptera) of Iraq and southwest Asia. *Coleop Bull* **32**: 1-20.

Attardo, G.M., Ribeiro, J.M., Wu, Y., Berriman, M. and Aksoy, S. (2010) Transcriptome analysis of reproductive tissue and intrauterine developmental stages of the tsetse fly (*Glossina morsitans morsitans*). *BMC Genomics* **11**: 160.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M. *et al.* (2009) Systems genetics of complex traits in Drosophila melanogaster. *Nat Genet* **41**: 299-307.

Ball, G.E. (1979) Conspectus of carabid clasisifications: History, holomorphology and higher taxa. In *Carabid Beetles: Their evolution, Natural History and Classification* (Erwin, T.L., Ball, G.E.D. and Whitehead, R., eds) pp. 63-111. Boston, London, The Hague:Junk.

Bang, K., Hwang, S., Lee, J. and Cho, S. (2015) Identification of immunity-related genes in the larvae of *Protaetia brevitarsis seulensis* (Coleoptera: Cetoniidae) by a Next-Generation sequencing-based transcriptome analysis. *J Insect Scie* **15**: 142.

Barton, N. (2000) The rapid origin of reproductive isolation. *Science* **290**: 462-463.

Bauer DuMont, V.L., Flores, H.A., Wright, M.H. and Aquadro, C.F. (2007) Recurrent positive selection at *bgcn*, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein bam. *Mol Biol Evol* **24**: 182-191.

Begun, D.J., Whitley, P., Todd, B.L., Waldrip-Dail, H.M. and Clark, A.G. (2000) Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* **156**: 1879-1888.

Blows, M.W. (1999) Evolution of the genetic covariance between male and female components of mate recognition: an experimental test. *P Roy Soc B-Biol Sci* **266:** 2169-2174.

Bocak, L., Barton, C., Crampton-Platt, A., Chesters, D., Ahrens, D. and Vogler, A.P. (2014) Building the Coleoptera tree-of-life for > 8000 species: composition of public DNA data and fit with Linnaean classification. *Syst Entomol* **39**: 97-110.

Bonin, A. (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Mol Ecol* **17**: 3583-3584.

Bonin, A., Paris, M., Tetreau, G., David, J.P. and Després, L. (2009) Candidate genes revealed by a genome scan for mosquito resistance to a bacterial insecticide: sequence and gene expression variations. *BMC Genomics* **10**: 551-563.

Bulmer, M.S (2010) Evolution of Immune Proteins in Insects. In *Encyclopedia of Life Sciences* (ELS) (John Wiley and Sons, eds) Ltd: Chichester.

Bulmer, M.S. and Crozier, R.H. (2004) Duplication and diversifying selection among termite antifungal peptides. *Mol Biol Evol* **21**: 2256-2264.

Bulmer, M.S. and Crozier, R.H. (2006) Variation in positive selection in termite GNBPs and Relish. *Mol Biol Evol* **23**: 317-326.

Cardoso, A. and Vogler, A. (2005) DNA taxonomy, phylogeny and Pleistocene diversification of the *Cicindela hybrida* species group (Coleoptera: Cicindelidae). *Mol Ecol* **14**: 3531-3546.

Carroll, S.S. and Pearson, D.L. (1998) Spatial modeling of butterfly species richness using tiger beetles (Cicindelidae) as a bioindicator taxon. *Ecol Appl* **8**: 531-543.

Cassola, F. and Pearson, D.L. (2000) Global patterns of tiger beetle species richness (Coleoptera: Cicindelidae): their use in conservation planning. *Biol Conserv* **95:** 197-208.

Civetta, A., Rajakumar, S.A., Brouwers, B. and Bacik, J.P. (2006) Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in Drosophila. *Mol Biol Evol* **23**: 655-662.

Dreisig, H. (1980) Daily activity, thermoregulation and water loss in the tiger beetle *Cicindela hybrid. Oecologia* **44**: 376-389.

Elmer, K.R. and Meyer, A. (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol Evol* **6**: 298-306.

Fan, S., Elmer, K.R. and Meyer, A. (2012) Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. *Philos T Roy Soc B: Biol Sci* **367**: 385-394.

Feder, M.E. and Mitchell-Olds, T. (2003) Evolutionary and ecological functional genomics. *Nat Rev Genet* **4**: 649-655.

Ford, M.J. (2002) Application of selective neutrality test to molecular ecology. *Mol Ecol* **11:** 1245-1262.

Galián, J., Hogan, J.E. and Vogler, A.P. (2002) The origin of multiple sex chromosomes in tiger beetles. *Mol Biol Evol* **19**: 1792-1796.

Galián, J., Proença, S.J.R. and Vogler, A.P. (2007) Evolutionary dinamics of autosomal-heterosomal rearrangements in a multiple-X chromosome system of tiger beetles (Cicindelidae). *BMC Evol Biol* **7**: 158.

Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ram, K.R., Sirot, L.K. *et al.* (2007) Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. *Genetics* **177:** 1321-1335.

Horn, W. (1915) Coleoptera Adephaga (family Carabidae, subfamily Cicindelinae) In *Genera Insectorum* (Wystman, P. and Brussels, eds) pp 1-486. Desmet-Vereneuil.

Horn, W. (1926) Pars 86: Carabidae: Cicindelinae. In *Coleopterum catalogus* (Junk and Schenkling, eds) Berlin.

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, OS., Wild, R. *et al.* (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318: 1913–1916.

Jagadeeshan, S. and Singh, R.S. (2005) Rapidly evolving genes of *Drosophila*: differing levels of selective pressure in testis, ovary, and head tissues between sibling species. *Mol Biol and Evol* ***22***: 1793-1801.

Jansa, S.A., Lundrigan, B.L. and Tucker, P.K. (2003) Tests for positive selection on immune and reproductive genes in closely related species of the murine genus Mus. *J Mol Evol* **56**: 294-307.

Keeling, C.I., Yuen, M.M., Liao, N.Y., Docking, T.R., Chan, S.K., Taylor, G.A. *et al.* (2013) Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. *Genome Biol* **14:** R27.

Kelley, J.L., Passow, C.N., Plath, M., Rodriguez, L.A., Yee, M.C. and Tobler, M. (2012) Genomic resources for a model in adaptation and speciation research: characterization of the *Poecilia mexicana* transcriptome. *BMC Genomics* **13**: 652.

Lazzaro, B.P. and Little, T.J. (2009) Immunity in a variable world. *Philos Trans R Soc Lond B Biol Sci* **364**: 15-26.

Li, H., Lang, K.L., Fu, H.B., Shen, C. P., Wan, F.H. and Chu, D. (2014) Analysis of expressed sequence tags (ESTs) from a normalized cDNA library and isolation of EST simple sequence repeats from the invasive cotton mealybug *Phenacoccus solenopsis*. *Insect Sci* **0**:1-7.

Liu, S., Lee, H., Kang, P.S., Huang, X., Yim, J.H., Lee, H.K. *et al.* (2010) Complementary DNA library construction and expressed sequence tag analysis of an Arctic moss, *Aulacomnium turgidum*. *Polar Biol* **33**: 617-626.

López-López, A., Abdul Aziz, A. and Galián, J. (2015) Molecular phylogeny and divergence time estimation of *Cosmodela* (Coleoptera: Carabidae: Cicindelinae) tiger beetle species from Southeast Asia. *Zool Scripta* **44**: 437-445.

López-López, A., Hudson, P. and Galián, J. (2012) The *blackburni/marchisona* species complex in Australian *Pseudotetracha* (Coleoptera: Carabidae: Cicindelinae: Megacephalini): evaluating molecular and karyological evidence. *J Zoolog Syst Evol Res* **50**: 177-183.

López-López, A., Hudson, P. and Galián, J. (2013) Recent origin of a chiasmatic sex trivalent in Australian *Pseudotetracha* tiger beetles. *J Zool* **289**: 262-269.

Lorenz, W. (2005) A systematic list of extant ground beetles of the world (Coleoptera "Geadephaga": Trachypachidae and Carabidae incl. Paussinae, Cicindelinae, Rhysodinae) (Lorenz W, ed) Tutzing.

Maddison, D.R., Moore, W., Baker, M.D., Ellis, T.M., Ober, K.A., Cannone, J.J. *et al.* (2009) Monophyly of terrestrial adephagan beetles as indicated by three nuclear genes (Coleoptera: Carabidae and Trachypachidae). *Zool Scripta* **38:** 43-62.

Mardis, E.R. (2008) Next-Generation DNA Sequencing Methods. *Annu Rev Genomics Hum Genet* **9**: 387-402.

Mittapalli, O., Bai, X., Mamidala, P., Rajarapu, S.P., Bonello, P. and Herms, D.A. (2010) Tissue-specific transcriptomics of the exotic invasive insect pest emerald ash borer (*Agrilus planipennis*). *PLoS One* **5**: e13708.

Mueller, J.L., Ram, K.R., McGraw, L.A., Qazi, M.B., Siggia, E. D., Clark, A. G. *et al.* (2005) Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **171**: 131-143.

Nagaraj, S.H., Gasser, R.B. and Ranganathan, S. (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform* **8:** 6-21.

Narina, S.S., Buyyarapu, R., Kottapalli, K.R., Sartie, A.M., Ali, M.I., Robert, A. *et al.* (2011) Generation and analysis of expressed sequence tags (ESTs) for marker development in yam (*Dioscorea alata* L.). *BMC Genomics* **12**:100.

Nilsson, A.N. (2001) Dytiscidae (Coleoptera). In *World Catalogue of Insects*. (Apolo books, eds) pp. 1–395. Stenstrup, Denmark.

Obbard, D.J., Welch, J.J., Kim, K.W. and Jiggins, F.M. (2009) Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet* **5**: e1000698.

Pearson, D.L. (1980). Patterns of limiting similarity in tropical forest tiger beetles (Coleoptera: Cicindelidae). *Biotropica* **1**: 195-204.

Pearson, D.L. (1984) The tiger beetles (Coleoptera: Cicindelidae) of the Tambopata Reserved Zone, Madre de Dios, Peru. *Rev Peru Entomol* **27**: 15-24.

Pearson, D.L. (1988) Biology of Tiger Beetles. *Ann Rev Entomol* **33**: 123-147.

Pearson, D.L. (1992) Tiger beetles as indicators for biodiversity patterns in Amazonia. *Res Explor* **8:** 116-117.

Pearson, D.L. and Cassola, F. (1992) World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): indicator taxon for biodiversity and conservation studies. C*onserv Biol* **6**: 376-391.

Pearson, D.L. and Cassola, F. (2005) A quantitative analysis of species descriptions of tiger beetles (Coleoptera: Cicindelidae), from 1758 to 2004, and notes about related developments in biodiversity studies. *Coleops Bull* **59**: 184-193.

Pearson, D.L. and Ghorpade, K. (1989) Geographical and ecological history of tiger beetles (Coleoptera: Cicindelidae) of the Indian subcontinent. *J Biogeogr* **16**: 333-344.

Pearson, D.L and Vogler, A.P. (2001) In *Tiger beetles: the evolution, ecology and diversity of the cicindelids*. Nueva York: Cornell University Press.

Pearson, D.L., Blum, M.S., Jones, T.H., Fales, H.M., Gonda, E. and White, B.R. (1988) Historical perspective and the interpretation of ecological patterns: defensive compounds of tiger beetles (Coleoptera: Cicindelidae). *Am Nat* **132**: 404-416.

Poelchau, M.F., Reynolds, J.A., Denlinger, D.L., Elsik, C.G. and Armbruster, P.A. (2011) A de novo transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. *BMC genomics* **12**: 619.

Pons, J., Barraclough, T.G., Gómez-Zurita, J., Cardoso, A., Duran, D.P. and Hazell, S. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol* **55**: 595-609.

Pons, J., Barraclough, T.G., Theodorides, K., Cardoso, A. and Vogler, A.P. (2004). Using exon and intron sequences of the gene Mp20 to resolve basal relationships in *Cicindela* (Coleoptera: Cicindelidae). *Syst Biol* **53**: 554-570.

Putchkov, A. and Cassola, F. (2005) Tiger Beetles deserve separate family status in suborder Adephaga (Coleoptera, Cicindelidae). *Bull Soc Entomol Fr* **110**: 281-293.

Renaut, S., Nolte, A. and Bernatchez, L. (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol Ecol* **19**:115-131.

Renn, S.C. and Siemens, D.H. (2010) Ecological genomics - changing perspectives on Darwin's basic concerns. *Mol Ecol* **19**: 3025-3030.

Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W. *et al.* (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949-955.

Rodriguez, J.P., Pearson, D.L. and Barrera, R. (1998) A test for the adequacy of bioindicator taxa: are tiger beetles (Coleoptera: Cicindelidae) appropriate indicators for monitoring the degradation of tropical forests in Venezuela? *Biol Conserv* **83**: 69-76.

Sackton, TB., Lazzaro, B.P., Schlenke, T.A., Evans, J.D., Hultmark, D. and Clark, A.G. (2007) Dynamic evolution of the innate immune system in *Drosophila*. *Nat Genet* **39**: 1461-1468.

Schlenke, T.A. and Begun, D.J. (2005) Linkage disequilibrium and recent selection at three immunity receptor loci in *Drosophila simulans*. *Genetics* **169**: 2013-2022.

Serrano, J. (2013) *New catalogue of the Carabidae of the Iberian Peninsula (coleoptera)*. (Serrano J, ed). Universidad de Murcia: servicio de publicaciones.

Shull, V.L., Vogler, A.P., Baker, M.D., Maddison, D.R. and Hammond, P.M. (2001) Sequence alignment of 18S ribosomal RNA and the basal relationships of adephagan beetles: evidence for monophyly of aquatic families and the placement of Trachypachidae. *Syst Biol* **50**: 945-969.

Singh, T. and Gupta, S. (1982) Morphology and histology of the mandibular gland in *Cicindela sexpunctata* Fabr. (Coleoptera: Cicindeliadae). *Uttar pradesh J Zool* **2**: 14-18.

Sota, T. (1994) Variation of carabid life cycles along climatic gradients: An adaptive perspective for life-history evolution under adverse conditions. In *Insect Life-Cycle Polymorphism: Theory, Evolution; and Ecological Consequences for Seasonal and Diapause Control* (Danks HV, ed) pp 91-112. Dordrecht, Netherlands: Kluwer Academic Publishers.

Spangler, H.G. (1988) Hearing in tiger beetles (Cicindelidae). *Physiol Entomol* **13**: 447-452.

Stapley, J., Reger, J., Feulner, P. G., Smadja, C., Galindo, J., Ekblom, R. *et al.* (2010) Adaptation genomics: the next generation. *Trends Ecol Evol* **25**: 705-712.

Swanson, W.J. and Vacquier, V.D. (2002) The rapid evolution of reproductive proteins. *Nat Rev Genet* **3**: 137-144.

Swanson, W.J., Clark, A.G., Waldrip-Dail, H.M., Wolfner, M.F. and Aquadro, C.F. (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila. P Natl Acad USA* **98**: 7375-7379.

Theodorides, K., De Riva, A., Gómez-Zurita, J., Foster, P.G. and Vogler, A.P. (2002) Comparison of EST libraries from seven beetle species: towards a framework for phylogenomics of the Coleoptera. *Insect Mol Biol* **11:** 467-475.

Toh, Y. and Okamura, J.Y. (2001) Behavioural responses of the tiger beetle larva to moving objects: role of binocular and monocular vision. *J Exp Biol* **204**: 615-625.

Ungerer, M.C., Johnson, L.C. and Herman, M.A. (2008) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity* **100**:178-183.

Van Belleghem, S.M., Roelofs, D., Van Houdt, J. and Hendrickx, F. (2012) De novo transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle *Pogonus chalceus* (Coleoptera, Carabidae). *PloS One* **7**: e42605.

Viljakainen, L. (2015) Evolutionary genetics of insect innate immunity. *Brief Funct Genomics*: elv002.

Viljakainen, L. and Pamilo, P. (2008) Selection on an antimicrobial peptide defensin in ants. *J Mol Evol* **67**: 643-652.

Vogler, A.P. and Pearson, D.L. (1996) A molecular phylogeny of the tiger beetles (Cicindelidae): congruence of mitochondrial and nuclear rDNA data sets. *Mol Phylogenet Evol* **6**: 321-338.

Vogler, A.P. and Welsh, A. (1997) Phylogeny of North American Cicindela Tiger Beetles Inferred from Multiple Mitochondrial DNA Sequences. *Mol Phylogenet Evol* **8**: 225-235.

Vogler, A.P, Cardoso, A. and Barraclough, T.G. (2005) Exploring rate variation among and within sites in a densely sample tree: species level phylogenetics and North American tiger beetles (genus *Cicindela*). *Sist Biol* **54**: 4-20.

Wei, D., Tian, C.B., Liu, S.H., Wang, T., Smagghe, G., Jia, F.X. *et al.* (2015) Transcriptome analysis have been used to identify genes for peptides and proteins involved in immunity and reproduction from male accessory glands and ejaculatory duct of *Bactrocera dorsalis*. *Peptides:* doi:10.1016/j.peptides.2015.08.007

Wiesner, J. (1992) Verzerichnis der Sabdlaufkäfer der Welt. *Checklist of the tiger beetles of the world.* (Verlag Erna Bauer, eds) Keltern, Germany.

Willis, H.L. (1967) Bionomics and zoogeography of tiger beetles of saline habitats in the central United States (Coleoptera: Cicindelidae). *Univ Kans Sci Bull* **47**: 145-313.

Willis, H.L. (1972) Species density of North American Cicindela. *Cicindela* **4**: 29-34.Zhang, W., Song, W., Zhang, Z., Wang, H., Yang, M., Guo, R. *et al.* (2014) Transcriptome analysis of *Dastarcus helophoroides* (Coleoptera: Bothrideridae) using Illumina HiSeq Sequencing. *PLoS One* **9**: e10067.

Zhang, Z., Hambuch, T.M. and Parsch, J. (2004) Molecular evolution of sex-biased genes in Drosophila. *Mol Biol Evol* **21**: 2130-2139.

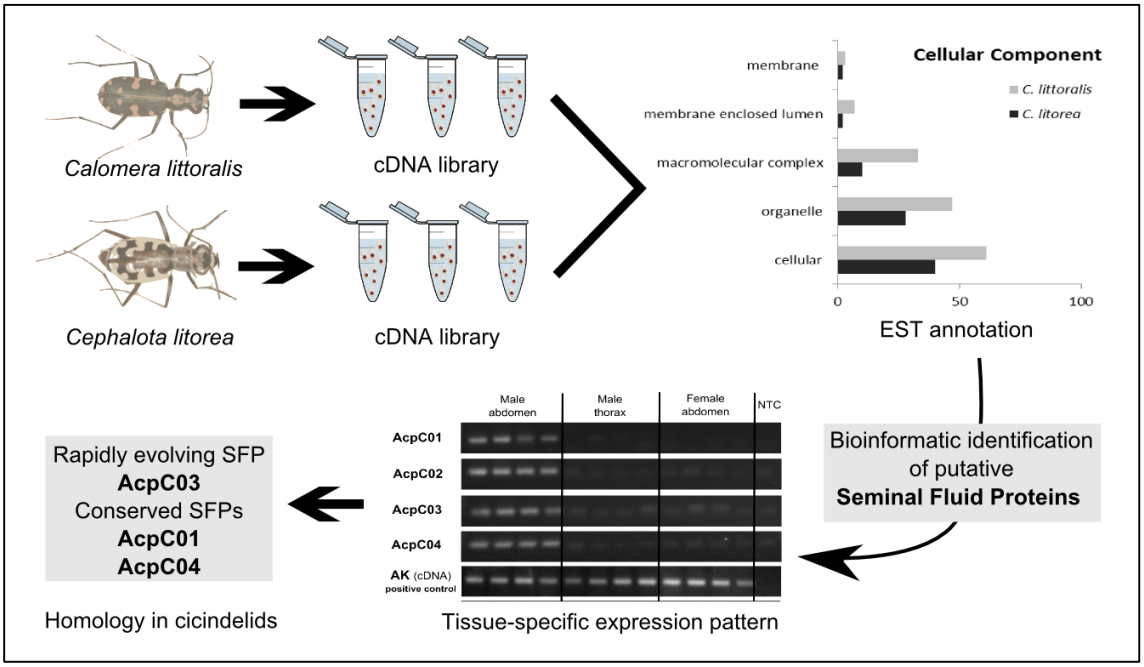
Chapter 1

Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles

Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles

María Juliana Rodríguez-García[1*], **Vilmar Machado**[1] and **José Galián**[1]

[1] *University of Murcia, Department of Zoology and Physical Anthropology, Faculty of Veterinary, Campus Mare Nostrum, E-30100 Murcia, Spain.*



Article published in BMC Genomics:

Rodríguez-García, M.J., Machado, V. and Galián, J. (2015) Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles. *BMC Genomics* **16:** 391. DOI: 10.1186/s12864-015-1619-9.

Abstract

Introduction. The study of proteins transferred through semen can provide important information for biological questions such as adaptive evolution, the origin of new species and species richness. The objective of this study was to identify seminal fluid proteins (SFPs) that may contribute to the study of the reproductive system of tiger beetles (cicindelids), a group of more than 2500 species distributed worldwide that occupy a great diversity of habitats.

Results. Two cDNA libraries were constructed from the male gonads of *Calomera littoralis* and *Cephalota litorea*. Expressed sequence tags (ESTs) were analysed by bioinformatics approaches and 14 unigenes were selected as candidate SFPs, which were submitted to Reverse Transcription Polymerase Chain Reaction (RT-PCR) to identify patterns of tissue-specific expression. We have identified four novel putative SFPs of cicindelids, of which similarity searches did not show homologues with known function. However, two of the protein classes (immune response and hormone) predicted by Protfun are similar to SFPs reported in other insects. Searches for homology in other cicindelids showed one lineage specific SFPs (rapidly evolving proteins), only present in the closely related species *C. littoralis* and *Lophyra flexuosa* and two conserved SFP present in other tiger beetles species tested.

Conclusions: This work represents the first characterisation of putative SFPs in adephagan species of the order Coleoptera. The results will serve as a foundation for further studies aimed to understand gene (and protein) functions and their evolutionary implications in this group of ecologically relevant beetles.

Key words: seminal fluid proteins, adephagan, tiger beetles, expressed sequences tags, gene expression

Introduction

Tiger beetles or cicindelids belong to the Adephaga, a suborder of Coleoptera that includes conspicuous, brightly coloured, non-pest species that are significant components of ecosystems, being important links in food chains. Cicindelids are organisms that are commonly used as bioindicators, as their presence or absence can provide information on the quality, alterations and successional stage of habitats, and are considered to be good bioindicators of general biodiversity (Pearson, 1992; Pearson & Cassola, 1992; Carroll & Pearson, 1998; Rodríguez *et al.,* 1998; Cassola & Pearson, 2000). There have been more than 2500 species of cicindelid beetles described (Pearson & Cassola, 2005), which are distributed worldwide (except for Tasmania, Antarctica, and some Oceanic islands) and occupy a great diversity of habitats (alpine meadows, dessert grassland, among others) (Pearson, 1988; Pearson *et al.,* 1988). Nevertheless, they do not show a homogenous pattern of distribution with relation to latitude or biogeography (Pearson & Cassola, 2005) and species tend to specialise in particular habitats (Pearson, 1992). They have several mechanisms or adaptations to reduce competition and contribute to maintain species richness; seasonal temporal segregation (life cycles), spatial segregation (differences in habitat or microhabitats preferences) (Willis, 1967) and temporal partitioning on the diel scale (diel activity patterns) have been previously studied as the most common strategies.

Tiger beetles have been deeply studied at different levels: i) morphology (Singh & Gupta, 1982); ii) taxonomy (Horn, 1915, 1926; Wiesner, 1992); iii) biology (Pearson, 1988; Sota, 1994); iv) physiology (Spangler, 1988; Toh & Okamura, 2001); v) thermoregulation (Dreisig, 1980); vi) evolution, ecology and diversity (Pearson & Vogler, 2001); vii) chromosome evolution (Galián *et al.*, 2007); and conservation strategies (Pearson & Cassola, 1992). Nevertheless, the reproductive biology has not been studied, on physiological or molecular grounds, and little information on the transcriptome and gene expression related to physiological processes is available. The transcriptome analysis is an important tool to help identifying putative function of genes, translating the sequence of nucleotides in a sequence of amino acids, which is more likely to be conserved (Haas *et al.,* 2002; Nagaraj *et al.*, 2007). In addition, the

availability of a variety of bioinformatics tools allows the characterisation of these genes (Ayroles *et al.,* 2009).

Some studies show that proteins with high expression levels in male reproductive tissues and with characteristics that meet the criteria of extracellular secretion, are good candidates to be considered as seminal fluid proteins (SFPs) (Davies & Chapman, 2006; Walters & Harrison, 2008, 2010; Almeida & DeSalle, 2009; Scolari *et al.,* 2012). The seminal fluid of insects contains sperm and a complex mixture of proteins, inorganic solutes, carbohydrates and lipids that are transferred to females during mating via the spermatophore. These proteins, which are produced in male gonads (testes, vas deferens and accessory glands), are important in the reproduction process by inducing physiological and behavioural changes in females, reducing responsiveness to other males, increasing the ovulation and egg laying rates, and altering feeding activity and also immune response (Haerty *et al.,* 2007; Findlay *et al.,* 2008; Avila *et al.,* 2011).

The study of proteins transferred through the semen provides information for important biological questions such as the origin of new species and the origin of new molecules involved in sperm competition and coevolution between males and females (McGraw *et al*., 2004; Fiumera *et al*., 2005; Wigby & Chapman, 2005; Findlay *et al*., 2008; Wigby *et al*., 2009; Fedorka *et al*., 2011; LaFlamme *et al*., 2012). Seminal fluid proteins have two characteristics that according to several theoretical models might lead to speciation. i) These proteins are related with sexual selection and sexual conflict; and ii) the rapid evolutionary rate of these proteins may also contribute to the evolution of reproductive barriers between populations. Furthermore, there are experimental evidences indicating a correlation between features that undergo the action of sexual selection and the speciation process (Ritchie, 2007; Snoock *et al.,* 2009; Castillo & Moyle, 2014).

In addition, several studies have demonstrated that many of these SFPs have similar characteristics to those found in the taxonomically restricted genes (TRGs), such as high evolutionary rate and low similarity between closely related species (Findlay *et al.,* 2008, 2009; Walters & Harrison, 2010; Marshall *et al*., 2011). According to Avila *et al.* (2011) the analysis of SFPs provides insight into the evolutionary patterns of

reproductive traits. Therefore, a better understanding of cicindelid reproductive molecules and their actions provides opportunities to reveal functionally conserved mechanism in cicindelids reproduction (highly conserved SFPs), as well as mechanisms involved in the reproductive isolation between species (lineage-specific SFPs) as a subset of seminal proteins is among the most rapidly evolving proteins (Mueller *et al.,* 2004; Clark & Swanson, 2005; Andres *et al.,* 2006; Clark *et al.,* 2006).

To date, several SFPs have been described in insects, such as flies and mosquitoes (Diptera), field crickets (Orthoptera), honeybee (Hymenoptera), moths and butterflies (Lepidoptera) and in the beetle genus *Tribolium* (Andres *et al.,* 2006, 2008; Collins *et al*., 2006; Sirot *et al*., 2008; Almeida & DeSalle, 2009; Baer *et al*., 2009; Walters & Harrison, 2010; South *et al*., 2011). Apart from the Polyphagan genus *Tribolium*, which is considered a model organism, no other species of beetles have been analysed for these proteins.

The aim of this study was to identify and characterise genes encoding proteins that are transferred to females during mating through seminal fluids in cicindelids to contribute to the knowledge of the nature and function of insect SFPs and particularly in this ecologically important group of adephagan beetles. To identify these proteins, EST libraries from gonads and accessory glands of male *Calomera littoralis* and *Cephalota litorea* were made and bioinformatically analysed. Proteins selected as candidate SFPs were submitted to gene expression analysis in female and male tissues. The complete sequence of genes showing differential expression patterns was obtained and the prediction of the function was inferred either by comparing with other insects or based on the sequence properties. Searches for homology of these putative SFP were performed in other cicidelid species. This work represents the first characterisation of SFP in adephagan beetles.

Material and methods

Library construction and EST assembly

Tiger beetles

Two cDNA libraries were constructed from male reproductive tissues (testes and accessory glands) of two *Calomera littoralis* males and a *Cephalota litorea* male. The developmental stage of the testes was that of sexually mature males with the final half of the testes full of spermatozoa (pearl white colour) and the anterior part with active meiosis (transparent white colour). Reproductive tissues were extracted and preserved in RNAlater (Qiagen, Crawley, UK) and stored at -20 ºC prior to extraction.

cDNA libraries

According to the manufacturer's instructions, total RNA was precipitated using the RNeasy Protect Mini Kit (Qiagen, Crawley, UK). The RNA sample concentrations were quantified by spectrometry. mRNA was purified using the Oligotex mRNA mini kit (Qiagen, Crawley, UK). The BD SMART PCR cDNA Synthesis kit (BD Biosciences) was used for cDNA libraries construction. The first strand was synthesised using BD PowerScript Reverse Transcriptase, the SMART IIA Oligonucleotide and the CDS IIA primer provided in the kit. The double- stranded cDNA was synthesised by PCR and purified using Micropure-EZ (Millipore). The cDNA products were analysed using agarose gel electrophoresis to determine their quality before cloning. cDNA was ligated into a TOPO vector and was transformed by TOPO TA cloning (Invitrogen) using TOP10 chemically competent *E. coli* cells. Libraries were plated on LB medium and grown at 37 ºC overnight. Colonies were manually picked up for PCR amplification with M13 and T7 universal vector primers and subsequent sequencing with poly-T primer on an ABI Prism 3700 sequencer (Applied Biosystems).

Annotation – gene ontology

The Seqman module of DNAStar (ver. Madison, WI) was used to remove the vector sequence, trim ends using program defaults and assemble sequences. Assembling parameters were: 80% for minimal match percentage, 100 for minimal sequence length, 0 for gap penalty, and 0. 7 for gap length penalty.

Assembled sequences (contigs and singletons) were subjected to a similarity search for assigned putative protein functions using BlastX of the Blast2GO v2.5.0

software (Conesa *et al.*, 2005) with 10^{-3} for the cut-off E-value. BlastX reference was used from the non-redundant protein database of GenBank in the National Center for Biotechnology Information. Gene ontology enrichment analysis was performed with Blast2GO mapping to determine protein functions in biological processes.

Identification of putative SFP

Indirect strategies based on bioinformatics tools were previously employed in insects to identify putative SFP unigenes (Findlay *et al.*, 2009; Walters & Harrison, 2010; Sonenshine *et al.*, 2011; Scolari *et al.*, 2012). First, open reading frames (ORF) of each unigene generated in Orfinder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) were selected, and SignalP 4.1 software (Nordahl Petersen *et al.*, 2011) was used to identify a predicted signal peptide. Additionally, integral prediction of protein location was analysed by ProtComp v9 software (http://www.softberry.com) and TMHMM tool in InterProScan (Zdobnov & Apweiler, 2001) was used to determine the presence of a membrane helix. Those candidate genes with similar predicted functions to known SFPs (extracellular and/or with membrane destination, as the signal peptide can sometimes be recognised as a membrane helix) were also selected.

Second, patterns of tissue-specific expression were examined for each SFP candidate via reverse transcription PCR (RT-PCR). PCR primers were designed with PrimerExpress 3.0 (Applied Biosystems) (Table 1). Total RNA was isolated from the male abdomen, male thorax (pronotum) and female abdomen, in three males and three females of *C. littoralis* and *C. litorea*. RNA was extracted using the RNeasy Protect Mini Kit (Qiagen) following the manufacturer's protocol. Around 1,2 µg of each RNA extraction were treated with TURBO DNA free (Ambion, Life Technologies) and reverse-transcribed using the QuantiTect Reverse Transcription kit (Qiagen). One µl of a ten-fold dilution of cDNA was used as template for a 12,5 µl RT-PCR experiments. PCR was performed using the following cycling parameters: one cycle of 2 minutes at 96 ºC, 35 cycles of 30 seconds at 96 ºC, 30 seconds at 60 ºC and 1 minute at 72 ºC, and a final extension of 10 minutes at 72 ºC. The arginine kinase gene (AK) was used as a positive control using in this case cDNA and also RNA to discard genomic DNA amplification. PCR amplicons were electrophoresed on agarose gel with RedSafe™ (INTRON Biotechnologies, Korea).

SFP characterisation

The SMARTer™ RACE cDNA Amplification Kit (Clontech Laboratories, Inc. Kyoto, Japan) was used to obtain the full length cDNA sequences. Total RNA from one *C. littoralis* male and one *C. litorea* male was used to obtain the first strand of 3' and 5' RACE Ready CDNA. Gene-specific primers for 5' and 3' RACE were designed (Table 1). The synthesis of first strand cDNA was performed following PCR conditions, as indicated the manufacturer's instructions. The amplification products were sequenced in triplicate by SAI at the University of Murcia (Spain) using an ABI Prism 3130 Sequencer (Applied Biosystems) and assembled by GENEIOUS v5 (Drummond *et al.,* 2010) to obtain the complete cDNA sequence.

Table 1. RACE-PCR and RT-PRC primers for the four putative seminal fluid proteins and amplicon size of each one.

	Primer sequence		Amplicon size
AcpC01	**RACE_F1**	5'-GTATTCCATTGTGTCCACCACCTCCGG-3'	**128**
	RACE_R1	5'-TGGTGGACAAGGTGGACAACATGGAAC -3'	
	RT_F1	5'-TTGCCCTCCATGTGCAGTAC-3'	**139**
	RT_R1	5'-TGGCTTCTGTGGCTCAAATTT-3'	
AcpC02	**RACE_F2**	5'- TGAGGAACCAGCCGCACAAGTAAAGAC-3'	**191**
	RACE_R2	5'-AGACCGACTCTGCAGTTTTTGTCTCGG-3'	
	RT_F2	5'-AGGAACCAGCCGCACAAGTA-3'	**110**
	RT_R2	5'-CTCCTTGTTGGGTGGTGCAT-3'	
AcpC03	**RACE_F3**	5'-TCATAACGATGATTCTGCCGCTCGTGG-3'	**176**
	RACE_R3	5'-GACACTCGAGATGCCTACAGTCCGGTA-3'	
	RT_F3	5'-ATGCTGTGCTGCTTGTGCAT-3'	**100**
	RT_R3	5'-GGACAACAGGCCGGAAATG-3'	
AcpC04	**RACE_F4**	5'-ACCAGTTTGTGATTGTCCGCCGTTACG-3'	**300**
	RACE_R4	5'-GTGTAACTGAACGCACGGGAAATAGCC-3'	
	RT_F4	5'-GCTATTTCCCGTGCGTTCAG-3'	**100**
	RT_R4	5'-CGGAGATCTCGTCTGCGTTT-3'	

The complete ORFs derivate from the full length sequences were generated in Orfinder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and were used as queries in BlastX to search for homologues in other species with an e-value cut-off of 10^{-4} and identities >30%. We also searched for similarity between our putative SFPs and other known insect proteins using Flybase (St. Pierre *et al.*, 2014). Additionally, protein domains were searched again using InterProScan (Jones *et al.*, 2014) and sequences were submitted to the PHYRE protein fold recognition server (Bennett-Lovsey *et al.*, 2008) to generate the protein structure. Sequences were submitted to ProtFun 2.2 Server (http://www.cbs.dtu.dk/services/ProtFun/) for analysis of the GO ontology to predict function based on amino acid sequence-derived input features (physical/chemical and functional biological properties) such as predicted protein secondary structure, transmembrane helices, subcellular localisation and post-transcriptional modifications (Jensen *et al.*, 2003).

Homology in cicindelid species

To determine the presence/absence of these putative SFPs in other species, total RNA was extracted from the abdomen and thorax of two males and the abdomen of two females of the following species: *Lophyra flexuosa*, *Cephalota maura*, *Cephalota deserticoloides* and *Cylindera trisignata*. RNA was extracted using the RNeasy Protect Mini Kit (Qiagen) following the manufacturer's instructions. For each tissue type around 1,2 μg of extracted RNA was treated with TURBO DNA free (Ambion, Life Technologies) to remove DNA contamination, and reverse-transcribed using the QuantiTect Reverse Transcription kit (Qiagen). RACE-PCR and RT-PCR primers that had been previously designed (Table 1) were tested in all of the available samples under the same conditions previously described for each PCR reaction. The arginine kinase gene (AK) was used as a positive control using cDNA and RNA.

Results and discussion

Library construction and EST assembly

Two separate cDNA libraries were constructed from RNA extracted from reproductive tissues (testes and accessory glands) of two males *C. littoralis* and one

male of *C. litorea*. A total of 1144 EST sequences were generated; 568 clones were sequenced from the *C. littoralis* library and 576 clones were sequenced from *C. litorea*. These sequences were trimmed for the removal of vector sequences, and sequences <100bp were excluded by Seqman (DNAstar, Inc. Madison, WI). EST sequences with high quality were previously deposited in GenBank under the following accession numbers: *C. litorea* (CV156657: CV157115) and *C. littoralis* (CV157116: CV157483). The high quality sequences of *C. littoralis* were assembled to 101 contigs (two or more sequences) and 84 singletons (single sequence), and *C. litorea* ESTs were assembled to 154 contigs and 58 singletons. Each cDNA library had a minimum average inset size ranging from 114 to 1245bp in *C. littoralis*, and from 139 to 1246 in *C. litorea* library. The maximum number of ESTs that formed each contig was 15 ESTs in *C. littoralis* and 18 ESTs in *C. litorea* (Table 2).

Table 2: Summary of EST analyses from *C. littoralis* and *C. litorea* male gonad cDNA libraries.

	C. littoralis	***C. litorea***
Number of ESTs generated	568	576
Number of High quality ESTs	368	459
Number of contigs	101	154
Number of singletons	84	58
Average length of contigs	523bp	628bp
Number of EST range in the contig	2-15	2-18

Annotation – gene ontology

Blast2GO software showed that 82 contigs of *C. litorea* and 72 contigs of *C. littoralis* had no blast hits against the non-redundant protein database at NCBI. The annotation of the 64 ESTs of *C. litorea* and 75 ESTs from *C. littoralis* were designated by database search algorithms BlastX for proteins in the NCBI web server (Table 3). Additionally, gene ontology annotations of all contigs were performed using Blast2GO.

Table 3. Summary of unigenes from *C. littoralis* and *C. litorea* male gonad analyses and annotated by Blast2go.

	C. littoralis	*C. litorea*
Annotated contigs	75	64
No annotated contigs	19	38
No mapping contigs	19	28
No blast hits contigs	72	82
No blast contigs	0	0
Total	185	212

Annotated sequences were classified according to their gene ontology (GO) into three categories: biological process, molecular function and cellular component (Fig 1, 2, and 3).

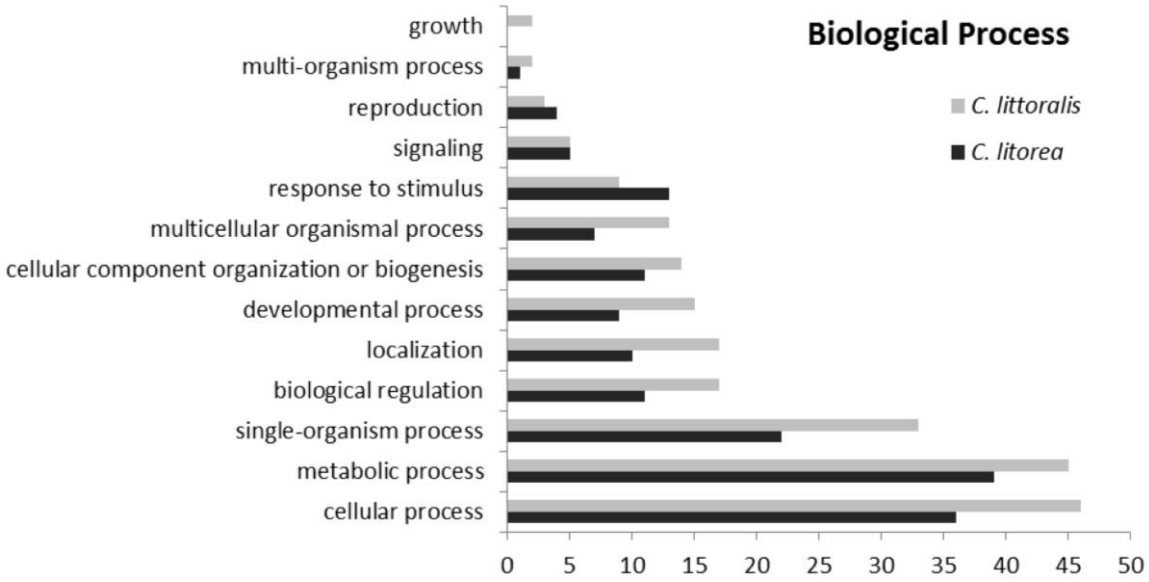


Figure 1. GO term distribution in biological processes for *C. littoralis* and *C. litorea* gonad unigenes. Percentages are in proportion to the total biological process GO annotations.

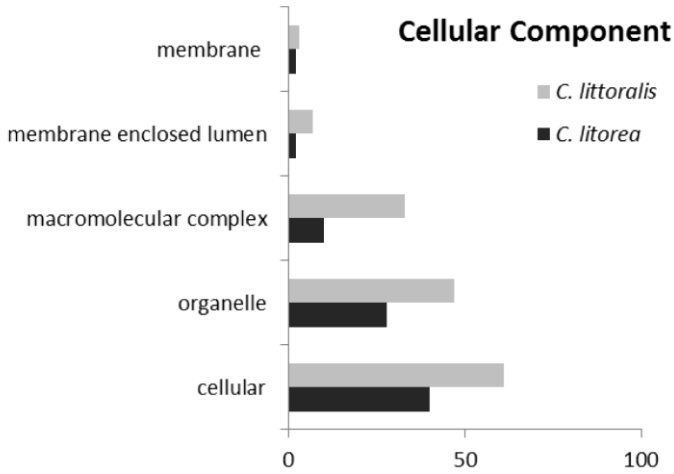


Figure 2. GO term distributions in the molecular functions for *C. littoralis* and *C. litorea* gonad unigenes. Percentages are in proportion to the total molecular function GO annotations.

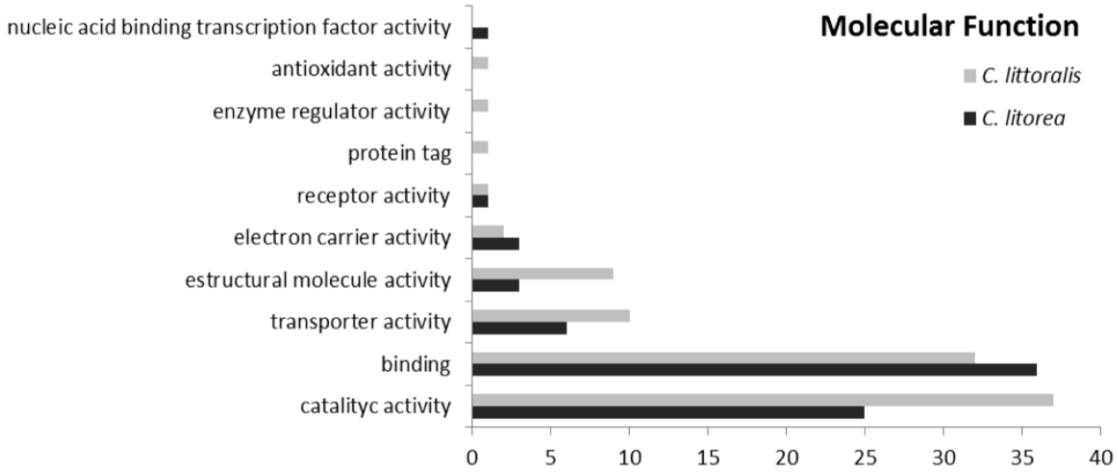


Figure 3. GO term distributions in the cellular component for *C. littoralis* and *C. litorea* gonad unigenes. Percentages are in proportion to the total cellular component GO annotations**.**

The proportions of genes associated with the different categories were highly similar among the two libraries; it is important to note that a sequence could be included in different categories and be associated to multiple GO. Within the category "Cellular Component", the subcategories "cellular" and "organelle" were the most abundant in both libraries. However, the "extracellular" subcategory, where putative SFPs should be included, was not present in the analysis performed, and no sequences were annotated as SFPs. This could be due to both the low number of ESTs

obtained in the cicindelid libraries and to the small number of coleopteran libraries available for comparisons.

Identification of putative SFPs

Nine genes putatively encoding SFPs were identified in *C. littoralis* and five in *C. litorea* by indirect approaches. Candidate genes were selected after detection of one or more of the following characteristics: i) the presence of a computationally predicted signal peptide inferred via Signal P 4.1 software (Nielsen *et al.*, 1997; Bendtsen *et al.*, 2004); ii) its localization as extracellular and/or with plasma membrane destination inferred via ProtCom; and iii) the recognition of a membrane helix inferred via TMHMM (InterproScan).

Of the 14 candidates, it was only possible to design useful RT-PCR primers for 12 (*C. littoralis*: eigth and *C. litorea*: four) (Table 4).

Table 4. Summary of the characteristics present in the candidate gene

	Contig	Cellular location	Signal peptide	Membrane helix
C. littoralis	13_59 (AcpC01)	Extracellular	No	Yes
	31_59	Plasma membrane	No	Yes
	50_59 (AcpC02)	Plasma membrane	No	Yes
	95_59	Extracellular	No	No
	129_59 (AcpC03)	Extracellular	Yes	Yes
	139_59	Plasma membrane	No	Yes
	161_59	Plasma membrane	No	No
	171-59	Extracellular	No	No
	173_59	Plasma membrane	No	No
	46_58	Plasma membrane	No	Yes
C. litorea	70_58 (AcpC04)	Extracellular	No	No
	126_58	Plasma membrane	No	Yes
	101_58	Extracellular	Yes	No
	204_58	Extracellular	No	No

Tissue-specific expression patterns were obtained in the 12 available candidates. RT-PCR revealed strong amplification from the male abdomen but weak or not expression in female abdomens and male thoraxes in four of the genes (*C. littoralis*: AcpC01, AcpC02 and AcpC03 and *C. litorea*: AcpC04). The positive control gene (arginine kinase) amplified in all tissues (Fig 4). The identification of an extracellular component and the tissue-specific patterns of expression found in these four genes suggest that they encode seminal fluid proteins. The other eight candidate genes with an identified extracellular component did not show any differential expression; some studies have demonstrated that not all SFPs have robust expression in male reproductive glands (Findlay *et al.,* 2008, 2009), which could explain the low number of putative SFPs characterised in the present survey.

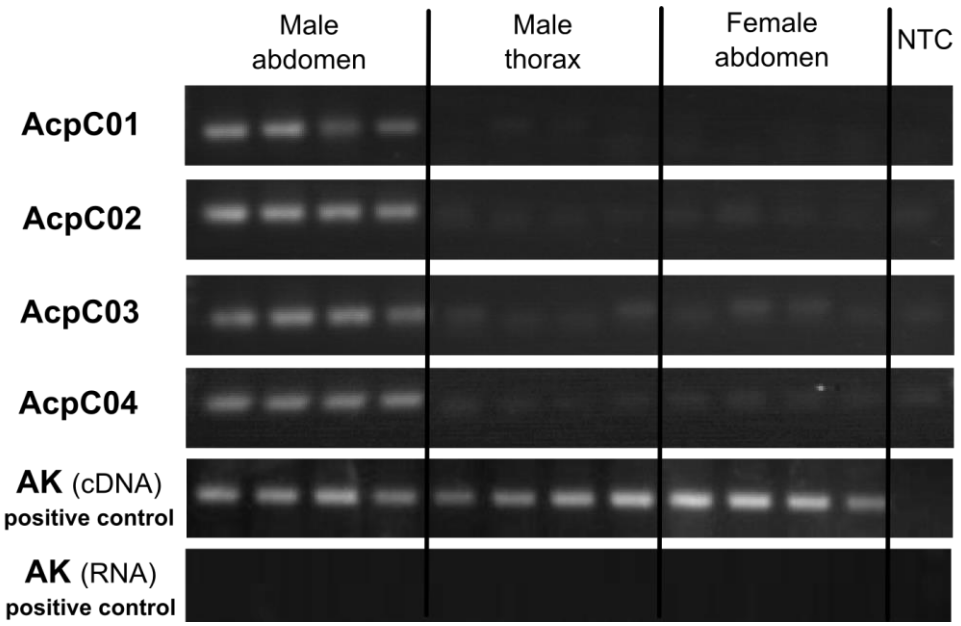


Figure 4. Results from RT-PCR, showing the expected pattern of expression in the different tissues for the four putative SFPs. Arginine kinase gene was used as positive control and was amplified from cDNA and RNA (free of DNAse) in all samples. NTC indicates negative template control.

SFP Characterisation

Full length sequences and complete Open Reading Frames (ORFs) were obtained via Rapid amplification of cDNA ends PCR (RACE-PCR) for the four putative SFPs. Sizes ranged from 66 to 218 amino acids, which is in agreement with ACPs

characterised in *Drosophila* (Chapman & Davies, 2004 and references therein). Sequences are available in GenBank (KP164546–KP164549). Two of these proteins (AcpC01 and AcpC02) did not show any significant BlastX similarity (E-value > 10^{-4}) against GenBank proteins, and appear to be novel proteins (Table 5). However, the protein AcpC01 yielded a similarity close to 30% with two seminal fluid proteins (HACP050: *Heliconius hereto* and CSSFP014: *Chilo suppressalis*). This similarity, although low, may be a sign of the high evolutionary rates documented for some of these SFPs (Clark *et al.*, 1995; Swanson *et al.,* 2001; Swanson & Vaquier, 2002; Clark & Swanson, 2005; Andres *et al.*, 2006). The proteins AcpC03 and AcpC04 showed significant similarity to sequences from *Drosophila yakuba* and *D. mojavensis,* respectively, which have not yet been characterised. Searches against several other insect genomes, run in Flybase, yielded the same results.

Table 5. Summary of the characterisation of cicindelids seminal fluid proteins. The homologies and different functions assigned by the different software packages used are indicated

GEN	amino acids size	CDS	GenBank hit	Flybase	PHYRE	Interpro	Protfun (gen ontology category)
AcpC01	66	Complete	No	No	NA	NA	Immune response
AcpC02	218	Complete	No	No	NA	NA	Growth factor
AcpC03	105	Complete	*Drosophila yakuba* XP_002100905.1 (e-value 7e-4, 30% similarity)	Yes/ Dyak/GE17317-PA	NA	Svwc domine	Hormone
AcpC04	206	Complete	*Drosophila mojavensis* XP_002007890.1 (e-value 8e-4, 30% similarity)	Yes Dmoj /GI12127-PA	NA	NA	Transcription regulation

Additionally, we tried to determine protein structure (3D fold) using PHYRE protein fold recognition metaserver and protein domains using InterProScan software. PHYRE did not yield consistent results for any of the genes tested. This approach, which proves the annotation in the tertiary structure of the proteins, was useful to annotate *Drosophila* SFPs (Mueller *et al.*, 2004), suggesting that candidate cicindelid SFPs do not meet the criteria found in *Drosophila*, due to differences in the structure and/or function.

In the AcpC03 gene, a Single domain von Willebrand factor type C (SVWC) was detected via InterProScan. SVWC family proteins, which are largely present in arthropods, normally contain ten cysteines, and are thought to respond to environmental challenges, such as bacterial infection and nutritional status (Zinke *et al.*, 2002; Sheldon *et al.*, 2007). Several studies have pointed out that SFPs may be involved in the immune response, as mating processes can transfer numerous pathogens into the female tract, jeopardising the reproductive success. Several SFPs analysed in *D. melanogaster* seem to have direct antimicrobial activity, protecting the male and subsequently the female reproductive tracts and even eggs against bacterial infection (Samakovlis *et al.*, 1991; Lung e*t al.*, 2001) and/or stimulation of antimicrobial gene expression levels (Mueller *et al.*, 2007). Other putative SFPs have been identified in other Diptera, such as *A. aegypti* and *A. gambiae* (Rogers *et al.*, 2008; Sirot *et al.*, 2008), which are related to immune response. In Coleoptera, South *et al.* (2011) identified a putative SFP in *Tribolium* which is a predicted prophenoloxidase, an important component of the innate immune response in Arthropoda. However, Protfun identifies AcpC03 as a hormone. This result could be in accordance with protein classes that are found in seminal proteins in different animals. Wolfner (2009) stated in her work that 40% of accessory gland proteins appear to be peptide hormones or prohormones, and in *Drosophila melanogaster* ACP26Aa SFP was found to have similarity with califin C, a hormone from *Aplysia californica* (Heifeltz *et al.*, 2000) which is involved in the egg-laying process (Rothman *et al.*, 1986). This could be an example of how a function assignment based on the sequence and structure similarity (InterproScan) could actually be different from a function assignment based not only on the structure but also on the physical/chemical and functional biological properties (Protfun). In other words, a conserved structure of a protein does not ensure a conserved function (Jensen *et al.,* 2003).

Finally, analysis with Protfun identified the gene AcpC01 to be an immune response protein. The other analyses did not assign a function to this protein based on similarity searches. However, Protfun analysis based on amino acid-derived input features did identify a function for this protein. This could be explained when considering the AcpC01gene as a novel putative SFP class that is either present only

in tiger beetles (taxonomically restricted gene) or has not yet been characterised in other insects. The protein AcpC04 was identified as a transcription regulation factor by Protfun and also has similarities in Flybase and GenBank with a non-characterised protein in *D. mojavensis*. Although transcription regulation factors are not included within the described protein classes of SFPs across animals (Wolfner, 2009), AcpC04 meets the requirements to be considered an SFP; therefore, further genetic studies may corroborate the biological function of this protein in tiger beetles. A similar consideration can be made in relation to the protein AcpC02, which according to the ontogenetic categories, is considered to be a growth factor by Protfun, although in this case, no homology has been found in the databases.

Homology in cicindelid species

RT-PCR and RACE-PCR primers were used in an attempt to amplify homologous sequences in the available cicindelid species (*C. litorea, C. littoralis, Lophyra flexuosa*, *Cephalota maura*, *Cephalota deserticoloides* and *Cylindera trisignata*). The AcpC04 gene, found originally in *C. litorea,* yielded homologous sequences in male abdomens of all of the analysed species (Fig 5).

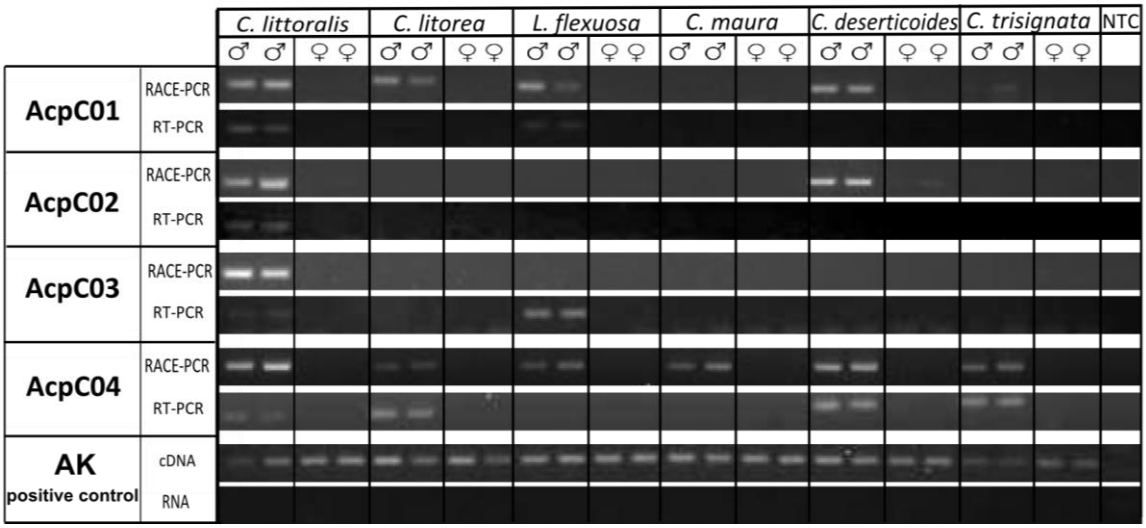


Figure 5. **Results of the SFP homology searches**. RT-PCR and RACE-PCR primers were used to amplify the four putative SFPs in six cicindelid species. Arginine kinase gene was used as positive control and was amplified from cDNA and RNA (free of DNAse) in all samples. NTC indicates negative template control**.**

This result is not surprising considering that although SFP are considered to have high evolutionary rate and low similarity between closely related species (Findlay *et al.*, 2008, 2009; Walters & Harrison, 2010; Marshall *et al.*, 2011), not all SFPs evolve rapidly and some loci are conserved between divergent taxa (Walters & Harrison, 2010). In this line of evidence, AcpC01 showed clear amplification in *C. litorea, C. littoralis, L. flexuosa* and *C. deserticoloides,* although weak signal in *C. trisignata* and no signal at all in *C. maura*. This result is coherent with the phylogenetic relationships obtained using cytochrome oxidase I gene (COI) (unpublished data) with the cicindelid species under study (Fig 6). In fact *C. maura* is the most distantly related species.

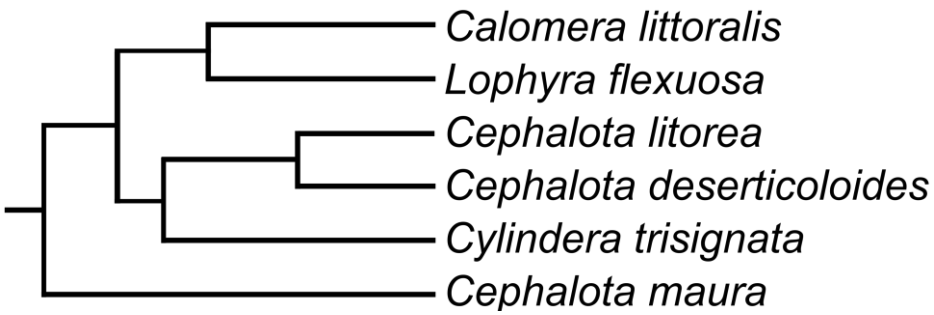


Figure 6. Phylogenetic tree based on COI sequences (unpublished data) from the cicindelid species analysed.

However, AcpC03 showed amplification only in male abdomens of *L. flexuosa* (Fig 5). The phylogenetic tree (Fig 6) showed that *C. littoralis* and *L. flexuosa* are closely related species. This close relationship could explain why these putative SFPs obtained from *C. littoralis* were also found in *L. flexuosa.* The lack of amplification in the other cicindelid species tested might be interpreted as a consequence of the rapid evolution that is generally considered for SFPs (Clark *et al.*, 1995, 2006; Swanson *et al.*, 2001, Swanson & Vacquier, 2002; Clark & Swanson, 2005; Andres *et al.*, 2006; Panhuis *et al.*, 2006). AcpC02 showed amplification in *C. littoralis* and *C. deserticoloides* with the RACE-PCR primers, but using the RT-PCR primers only *C. littoralis* gave positive results. However the detection of this gene in these two species suggest that it might be present in the most recent common ancestor of these two species, although with the primer pairs used was not detected.

Conclusions

Two cDNA libraries were constructed from gonads of *C. littoralis* and *C. litorea* as a foundation to understanding the male reproductive system. A total of 568 and 576 ESTs were sequenced and analysed, assembled in 185 and 212 unigenes, respectively. Also, 75 and 64 contigs were annotated via Blast2go and no SFPs were found. We have identified 14 putative SFPs by bioinformatics analysis and found that four of them met the criteria of tissue-specific expression patterns, which led to them being considered as putative SFP. Functional annotation was difficult due to the fact that the four SFPs either do not show homology via similarity searches or reassemble with sequences whose function is unknown; only Interpro assigns a function in the immune response for AcpC03, as occurs with others insect SFPs. The predicted assignment of function via Protfun for AcpC01 and AcpC03 was immune response and hormone, respectively; these protein functions are included within the conserved protein classes of SFPs that have already been reported in several insect species. Searches for homology in other cicindelids showed that AcpC03 is only present in *C. littoralis* and *Lophyra flexuosa*, species which have a close phylogenetic relationship. SFPs are among the most rapidly evolving proteins and therefore this new putative SPF might represent a linage-specific SFP involved in reproductive isolation between species. In contrast, AcpC04 is present in all cicindelid species analysed here, and could be an example of a highly conserved SFP, at least in tiger beetles. The same can be said about AcpC01, with is present in all species tested except for the most distantly related species *Cephalota maura*. This work represents the first identification of putative SFPs in tiger beetles that are one of the best studied worldwide distributed non-pest insects (more than 2500 species described), are important in ecosystems as predators and are commonly used as bioindicators. The identification of cicindelid SFPs (both the rapidly evolving and the highly conserved) could represent a significant approach for understanding the male reproductive system in cicindelids and furthermore the species richness and adaptive evolution in adephagan beetles.

Acknowledgments

We greatly thank Alfried P. Vogler, Alessandra de Riva and Kosmas Theodorides for contributing library construction and sequencing. Thanks are due to PRS Proofreading Service for language editing of the paper. This work has been funded by project CGL2008-03628 of the Ministerio de Ciencia e Innovación, Spain.

References

Almeida, F.C. and DeSalle, R. (2009) Orthology, function and evolution of accessory gland proteins in the *Drosophila repleta* group. *Genetics* **181**: 235-245.

Andres, J.A., Maroja, L.S., Bogdanowicz, S.M., Swanson, W.J. and Harrison, R.G. (2006) Molecular evolution of seminal proteins in field crickets. *Mol Biol Evol* **23**: 1574-1584.

Andres, J.A., Maroja, L.S. and Harrison, R.G. (2008) Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. *Proc R Soc Lond B Biol Sci* **275**: 1975-198.

Avila, F.W., Sirot, L.K., LaFlamme, B.A., Rubinstein, C.D. and Wolfner, M.F. (2011) Insect Seminal Fluid Proteins: Identification and Function. *Annu Rev Entomol* **56**: 21-40.

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M. *et al.* (2009) Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet* **41**: 299-307.

Baer, B., Heazlewood, J.L., Taylor, N.L., Eubel, H. and Millar, A.H. (2009) The seminal fluid proteome of the honeybee *Apis mellifera*. *Proteomics* **9**: 2085-2097.

Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: signalP 3.0. *J Mol Biol* **340**: 783-795.

Bennett-Lovsey, R.M., Herbert, A.D., Sternberg, M.J.E. and Kelley, L.A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* **70**: 611-625.

Carroll, S.S. and Pearson, D.L. (1998) Spatial modeling of butterfly species richness using tiger beetles (Cicindelidae) as a bioindicator taxon. *Ecol Appl* **8**: 531-543.

Cassola, F. and Pearson, D.L. (2000) Global patterns of tiger beetle species richness (Coleoptera: Cicindelidae): their use in conservation planning. *Biol Conserv* **95:** 197-208.

Castillo, D.M. and Moyle, L.C. (2014) Intraspecific sperm competition genes enforce post-mating species barriers in *Drosophila*. *P Roy Soc B-Biol Sci* **281**: 2014-2050.

Chapman, T. and Davies, S.J. (2004) Functions and analysis of the seminal fluid proteins of male Drosophila *melanogaster* fruit flies. *Peptides* **25**: 1477-1490.

Clark, N.L. and Swanson, W.J. (2005) Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet* **1**: e35.

Clark, N.L., Aagaard, J.E. and Swanson, W.J. (2006) Evolution of reproductive proteins from animals and plants. *Reproduction* **131**: 11-22.

Clark, A.G., Aguade, M., Prout, T., Harshman, L.G. and Langley, C.H. (1995) Variation in sperm displacement and its association with accessory-gland protein loci in *Drosophila melanogaster*. *Genetics* **139**: 189-201.Collins, A.M., Caperna, T.J., Williams, V., Garrett, W.M. and Evans, J.D. (2006) Proteomic analyses of male contributions to honey bee sperm storage and mating. *Insect Mol Biol* **15**: 541-549.

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

Davies, S.J. and Chapman, T. (2006) Identification of genes expressed in the accessory glands of male Mediterranean fruit flies (*Ceratitis capitata*). *Insect Biochem Mol Biol* **36**: 846-856.

Dreisig, H. (1980) Daily activity, thermoregulation and water loss in the tiger beetle *Cicindela hybrid*. *Oecologia* **44**: 376-389.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C. *et al.* (2010) Geneious v5.5 <http://www.geneious.com>.

Fedorka, K.M., Winterhalter, W.E. and Ware, B. (2011) Perceived sperm competition intensity influences seminal fluid protein production prior to courtship and mating. *Evolution* **65**: 584-590.

Findlay, G.D., MacCoss, M.J. and Swanson, W.J. (2009) Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Res* **19**: 886-896.

Findlay, G.D., Yi, X.H., MacCoss, M.J. and Swanson, W.J. (2008) Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* **6**: 1417-1426.

Fiumera, A.C., Dumont, B.L. and Clark, A.G. (2005) Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins. *Genetics* **169**: 243-257.

Galián, J., Proença, S.J.R. and Vogler, A.P. (2007) Evolutionary dinamics of autosomal-heterosomal rearrangements in a multiple-X chromosome system of tiger beetles (Cicindelidae). *BMC Evol Biol* **7**: 158.

Haas, B.J., Volfovsky, N., Town, C.D., Troukhan, M., Alexandrov, N., Feldmann, K.A. *et al.* (2002) Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biol* **3**: re-search0029.

Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ravi Ram, K., Sirot, L.K. *et al.* (2007) Evolution in the fast lane: rapidly evolving sex-related genes in drosophila. *Genetics* **177**: 1321-1335.

Heifetz, Y., Oliver, L., Frongillo, E.A. and Wolfner, M.F. (2000) The Drosophila seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. *Curr Biol* **10**: 99-102.

Horn, W. (1915) Coleoptera Adephaga (family Carabidae, subfamily Cicindelinae) In *Genera Insectorum* (Wystman, P. and Brussels, eds) pp 1-486. Desmet-Vereneuil.

Horn, W. (1926) Pars 86: Carabidae: Cicindelinae. In *Coleopterum catalogus* (Junk and Schenkling, eds) Berlin.

Jensen, L.J., Ussery, D.W. and Brunak, S. (2003) Functionality of system components: conservation of protein function in protein feature space. *Genome Res* **13:** 2444-2449.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

LaFlamme, B.A., Ram, K.R. and Wolfner, M.F. (2012) The *Drosophila melanogaster* seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. *PLoS Genet* **8**: e1002435.

Lung, O., Kuo, L. and Wofner, M.F. (2001) Drosophila males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. *J Insect Physiol* **47**: 617-622.

Marshall, J.L., Huestis, D.L., Garcia, C., Hiromasa, Y., Wheeler, S., Noh, S. *et al.* (2011) Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by post-mating pre-zygotic phenotypes. *Mol Biol Evol* **28**: 423-435.

McGraw, L.A., Gibson, G., Clark, A.G. and Wolfner, M.F. (2004) Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. *Curr Biol* **14**: 1509-1514.

Mueller, J.L., Page, J.L. and Wolfner, M. (2007) An ectopic expression screen reveals the protective and toxic effects of *Drosophila* seminal fluid proteins. *Genetics* **175**: 777-783.

Mueller, J.L., Ripoll, D.R., Aquadro, C.F. and Wolfner, M.F. (2004) Comparative structural modelling and inference of conserved protein classes in *Drosophila* seminal fluid. *P Natl Acad USA* **101**: 13542-13547.

Nagaraj, S.H., Gasser, R.B. and Ranganathan, S. (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform* **8**: 6-21.

Nielsen, H., Engelbrecht, J., Brunak, S. and vonHeijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**: 1-6.

Nordahl Petersen, T., Brunak, S., vonHeijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785-786.

Panhuis, T.M., Clark, N.L. and Swanson, W.J. (2006) Rapid evolution of reproductive proteins in abalone and *Drosophila*. *Philos Trans R Soc B* **361**: 261-268.

Pearson, D.L. (1988) Biology of Tiger Beetles. *Ann Rev Entomol* **33**: 123-147.

Pearson, D.L. (1992) Tiger beetles as indicators for biodiversity patterns in Amazonia. *Res Explor* **8:** 116-117.

Pearson, D.L. and Cassola, F. (1992) World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): indicator taxon for biodiversity and conservation studies. C*onserv Biol* **6**: 376-391.

Pearson, D.L. and Cassola, F. (2005) A quantitative analysis of species descriptions of tiger beetles (Coleoptera: Cicindelidae), from 1758 to 2004, and notes about related developments in biodiversity studies. *Coleops Bull* **59**: 184-193.

Pearson, D.L and Vogler, A.P. (2001) In *Tiger beetles: the evolution, ecology and diversity of the cicindelids*. Nueva York: Cornell University Press.

Pearson, D.L., Blum, M.S., Jones, T.H., Fales, H.M., Gonda, E. and White, B.R. (1988) Historical perspective and the interpretation of ecological patterns: defensive compounds of tiger beetles (Coleoptera: Cicindelidae). *Am Nat*

132: 404-416.Ritchie, M.G. (2007) Sexual selection and speciation. *Annu Rev Ecol Evol Syst* **38**: 79-102.

Rodriguez, J.P., Pearson, D.L. and Barrera, R. (1998) A test for the adequacy of bioindicator taxa: are tiger beetles (Coleoptera: Cicindelidae) appropriate indicators for monitoring the degradation of tropical forests in Venezuela? *Biol Conserv* **83**: 69-76.

Rogers, D.W., Whitten, M.M.A., Thailayil, J., Soichot, J., Levashina, E.A. and Catteruccia, F. (2008) Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. *P Natl Acad USA* **105**: 19390-19395.

Rothman, B.S., Hawke, D.H., Brown, R.O., Lee, T.D., Dehghan, A.A., Shively, J.E. *et al.* (1986) Isolation and primary structure of the califins, three biologically active egg-laying hormone-like peptides from the atrial gland of *Aplysia californica*. *J Biol Chem* **261**: 1616-1623.

Samakovlis, C., Kylsten, P., Kimbrell, D.A., Engstrom, A. and Hultmark, A. (1991) The andropin gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster*. *EMBO J* **10**: 163-169.

Scolari, F., Gomulski, L.M., Ribeiro, J.M.C., Siciliano, P., Meraldi, A., Falchetto, M. *et al.* (2012) Transcriptional profiles of mating-responsive genes from testes and male accessory glands of the Mediterranean fruit fly, *Ceratitis capitata*. *PLoS One* **7**: e46812.

Sheldon, T.J., Miguel-Aliaga, I., Gould, A.P., Taylor, W.R. and Conklin, D. (2007) A novel family of single VWC-domain proteins in invertebrates. *FEBS Lett* **581**: 5268-5274.

Singh, T. and Gupta, S. (1982) Morphology and histology of the mandibular gland in *Cicindela sexpunctata* Fabr.(Coleoptera: Cicindeliadae). *Uttar pradesh J Zool* **2**: 14-18.

Sirot, L.K., Poulson, R.L., McKenna, M.C., Girnary, H., Wolfner, M.F. and Harrington, L.C. (2008) Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for

control of female feeding and reproduction. *Insect Biochem Mol Biol* **38**: 176-189.

Snook, R.R., Chapman, T., Moore, P.J., Wedell, N. and Crudgington, H.S. (2009) Interactions between the sexes: new perspectives on sexual selection and reproductive isolation. *Evol Ecol* **23**: 71-91.

Sonenshine, D.E., Bissinger, B.W., Egekwu, N., Donohue, K.V., Khalil, S.M. and Roe, R.M. (2011) First transcriptome of the testis-vas deferens-male accessory gland and proteome of the spermatophore from *Dermacentor variabilis* (Acari: Ixodidae). *PLoS One* **6**: e24711.

Sota, T. (1994) Variation of carabid life cycles along climatic gradients: An adaptive perspective for life-history evolution under adverse conditions. In *Insect Life-Cycle Polymorphism: Theory, Evolution; and Ecological Consequences for Seasonal and Diapause Control* (Danks HV ed) pp 91-112. Dordrecht, Netherlands: Kluwer Academic Publishers.

South, A., Sirot, L.K. and Lewis, S.M. (2011) Identification of predicted seminal fluid proteins in *Tribolium castaneum*. *Insect Mol Biol* **20**: 447-456.

Spangler, H.G. (1988) Hearing in tiger beetles (Cicindelidae). *Physiol Entomol* **13**: 447-452.

St. Pierre, S.E., Ponting, L., Stefancsik, R., McQuilton, P., the FlyBaseConsortium (2014) FlyBase 102 - advanced approaches to interrogating FlyBase. *Nucleic Acids Res* **42**: D780-D788.

Swanson, W.J. and Vacquier, V.D. (2002) The rapid evolution of reproductive proteins. *Nat Rev Genet* **3**: 137-144.

Swanson, W.J., Clark, A.G., Waldrip-Dail, H.M., Wolfner, M.F. and Aquadro, C.F. (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *P Natl Acad USA* **98**: 7375-7379.

Toh, Y. and Okamura, J.Y. (2001) Behavioural responses of the tiger beetle larva to moving objects: role of binocular and monocular vision. *J Exp Biol* **204**: 615-625.

Walters, J.R. and Harrison, R.G. (2008) EST analysis of male accessory glands from *Heliconius* butterflies with divergent mating systems. *BMC Genomics* **9**:592.

Walters, J.R. and Harrison, R.G. (2010) Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. *Mol Biol Evol* **27**: 2000-2013.

Wiesner, J. (1992) Verzerichnis der Sabdlaufkäfer der Welt. *Checklist of the tiger beetles of the world.* (Verlag Erna Bauer, eds) Keltern, Germany.

Wigby, S. and Chapman, T. (2005) Sex peptide causes mating costs in female *Drosophila melanogaster*. *Curr Biol* **15**: 316-321.

Wigby, S., Sorot, L.K., Linkalater, J.R., Buehner, N., Calboli, F.C.F., Bretman, A. *et al.* (2009) Seminal fluid protein allocation and male reproductive success. *Curr Biol* **19**: 751-757.

Willis, H.L. (1967) Bionomics and zoogeography of tiger beetles of saline habitats in the central United States (Coleoptera: Cicindelidae). *Univ Kans Sci Bull* **47**: 145-313.

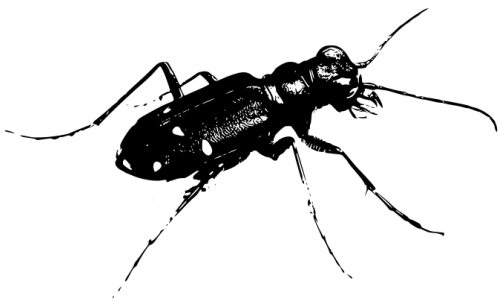
Wolfner, M.F. (2009) Battle and Ballet: molecular interactions between the sexes in *Drosophila*. *J Hered* **100**: 399-410.

Zdobnov, E.M. and Apweiler, R. (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatic* **17**: 847-8.

Zinke, I., Schutz, C.S., Katzenberger, J.D., Bauer, M. and Pankratz, M.J. (2002) Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. *EMBO J*. **21**: 6162-6173.

Chapter 2

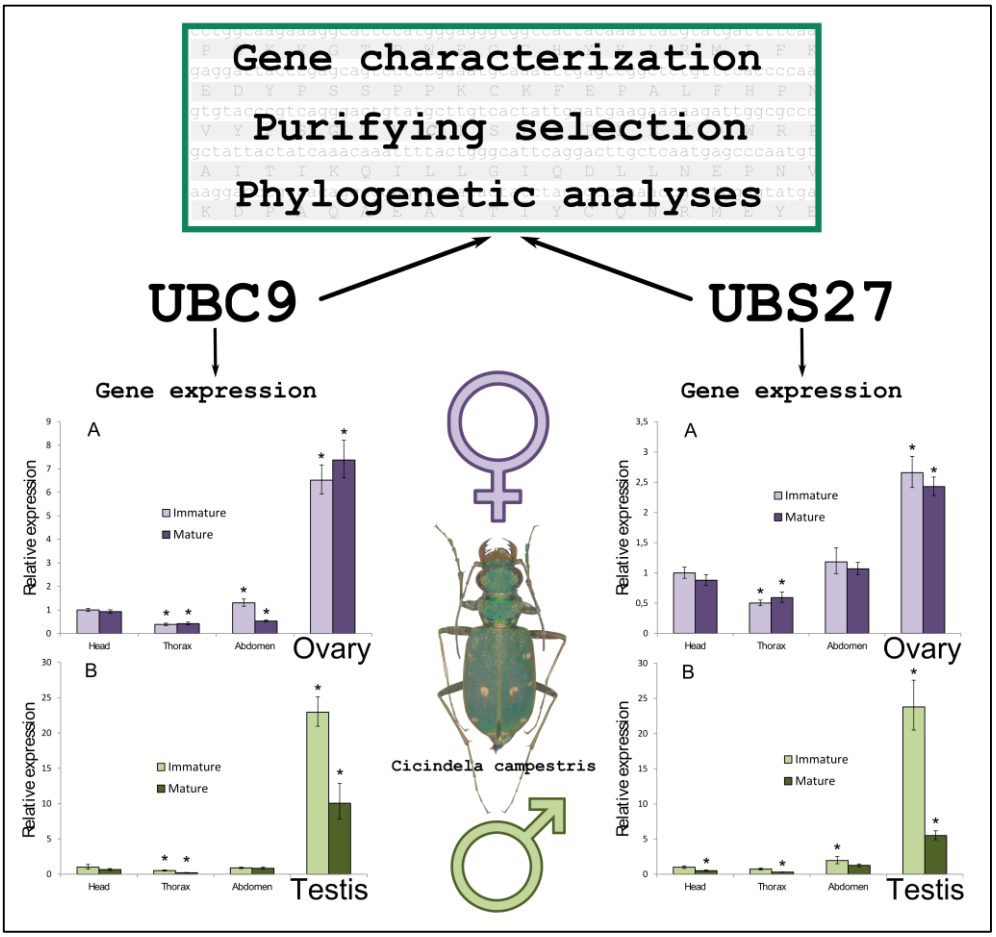
Characterisation and expression analysis of UBC9 and UBS27 genes in developing gonads of cicindelids (Coleoptera: Cicindelidae)



Characterisation and expression analysis of UBC9 and UBS27 genes in developing gonads of cicindelids (Coleoptera: Cicindelidae)

María Juliana Rodríguez-García[1*], **Andrés García-Reina**[1], **Vilmar Machado**[1] and **José Galián**[1]

[1] *University of Murcia, Department of Zoology and Physical Anthropology, Faculty of Veterinary, Campus Mare Nostrum, E-30100 Murcia, Spain.*



Abstract

Introduction. Ubiquitin and small ubiquitin-like modifiers (SUMO) are post-translational modifiers essential in a variety of cellular processes, including gametogenesis. SUMO-conjugating enzyme (UBC9) and the ubiquitin ribosomal fusion protein UBS27 have been characterised in several model species. However, their expression in coleopteran remains unstudied.

Results. In this study, UBC9 and UBS27 genes have been characterised in the tiger beetle *Cicindela campestris* for the first time. Bioinformatic analysis showed that the Cc-UBC9 gene encoded a 159 amino acid protein with a predicted molecular weight of 18.18 kDa, and the Cc-UBS27 gene encoded a 156 amino acid protein with a predicted molecular weight of 17.71 kDa. Selection analyses carried out in several cicindelid species revealed that both genes were affected by purifying selection. Real time quantitative PCR analysis demonstrated that Cc-UBC9 and Cc-UBS27 were expressed in different tissues. The highest expression was found in the ovary and testis, and there were differential expression levels between immature and mature stages of testis development.

Conclusion. The expression patterns of Cc-UBC9 and Cc-UBS27 suggest that these genes play important roles in gametogenesis in *C. campestris*. This information is relevant to better understand the reproductive process in cicindelids and the function of ubiquitin and small ubiquitin-related modifier genes in the Coleoptera.

Key words: Cc-UBC9, Cc-UBS27, Coleoptera, cicindelids, gene expression, purifying selection, gametogenesis.

Introduction

Protein post-translational modifiers are crucial in a variety of cellular processes (Walsh *et al.,* 2005). Ubiquitin is an essential highly conserved protein present in all tissues, consisting of 76 amino acids, with only four amino acid substitutions among plants, animals and yeast (Goldstein *et al.,* 1975; Ozkaynak *et al.,* 1987; Jentsch *et al.,* 1991). Two classes of genes encode ubiquitins, polyubiquitins and ubiquitin fusion genes (Finley *et al.,* 1989). Polyubiquitin gene encodes a precursor protein expressed in a head-to-tail form with several identical tandem repeat units. The two ubiquitin fusion repeats (ubl40 and ubl27) are fused with ribosomal proteins L40 and S27, respectively, at the C-terminus (Baker & Board, 1991; Mezquita *et al.,* 1997; Nenoi *et al.,* 2000). After releasing from its precursor, ubiquitin requires the concatenated action of three enzymes. First, ubiquitin conjugation enzyme (E1) activates the C-terminal Gly residue of ubiquitin in an ATP-dependent process. Second, ubiquitin is transferred to a ubiquitin conjugation enzyme (E2) and finally, in the third step, ubiquitin is linked to a target protein catalysed by a ubiquitin protein ligase enzyme (E3) (Haas *et al.,* 1982; Hershko & Ciechanover, 1998). The ubiquitin system plays important roles in the regulation of cellular processes such as cell division, apoptosis, and in conditions of cell stress. They are also involved in protein degradation, ribosomal synthesis and the immune response, and recent studies have demonstrated the role of ubiquitin proteins in reproductive processes (reviewed in Bebington *et al.,* 2001; Pickart & Eddins, 2004).

Additionally, small ubiquitin–like modifiers (SUMO) are also conjugated to proteins. The amino acid similarity between sequences of ubiquitin and SUMO is only approximately 18%, but their protein structure is conserved (Bayer *et al.,* 1998). Furthermore, SUMOylation, similar to ubiquitination, requires three enzymes, an ATP-dependent activating enzymes (E1) and a conjugating enzyme (E2, ubc9), and in some cases, requires specific SUMO ligases (E3). Therefore, SUMO proteins share similar mechanisms with ubiquitin, but there are some differences in their function. SUMO plays an important role in the regulation of transcription, nuclear transport, cell cycle, DNA replication and repair, and unlike ubiquitination, it does not intervene in degradation (Gill, 2004; Johnson, 2004).

Genes that participate in both processes are involved in reproductive function in animals. Ubiquitins are important in gametogenesis (Koken *et al.*, 1996; Roest *et al.,* 1996; Takagi *et al.*, 1997), especially in spermatogenesis, where E1 and E2 enzymes are implicated (Mitchell *et al.,* 1991; Roest *et al.,* 1996; Shen *et al.,* 2009), as well as ubiquitin ribosomal fusion proteins (Wang *et al.,* 2012a). Furthermore, different studies suggest that SUMOylation also plays an important role in gametogenesis. Expression studies of the UBC9 gene in different organisms showed that it is involved in embryogenesis, gametogenesis and sex modification (Zhang *et al.,* 2010; Wang *et al.,* 2012b; Hu & Chen, 2013).

Many genes involved in the sexual reproduction process evolved rapidly, often as a result of adaptive evolution (Swanson & Vacquier, 2002). Nevertheless, not all of the reproductive proteins evolved rapidly in certain linages (Metz *et al.,* 1998), and there are genes involved in spermatogenesis, as is the case of the t-complex polypeptide 1 gene (Tcp-1), which is subjected to purifying selection in mice (Jansa *et al.,* 2003). The ubiquitin gene family sequences are also subject to strong purifying selection (Nei *et al.,* 2000), which may also be the case for the ubiquitin conjugating enzyme family (E2) (Ying *et al.,* 2009). The study of reproductive genes that do not evolve rapidly could provide valuable information about why others have such high variability (Swanson & Vacquier, 2002).

Ubiquitin proteins have been identified and characterised through the use of genomic or expressed sequence tags (ESTs) in many insects, such as *Spodoptera litura* (Li *et al.,* 2003), *Blatella germanica* (Yu *et al.,* 2004), *Bombyx mandarina* (Chen *et al.,* 2007) and *Musca domestica* (Jin *et al.,* 2008). Nevertheless, there are few studies available in Coleoptera on the function of ubiquitin; only one paper is specifically related to their reproductive function (Yang *et al.,* 2009) in a Polyphagan beetle species. In relation to SUMO, it is difficult to find studies in insects other than *Drosophila,* where the SUMOylation roles are well studied (reviewed in Smith *et al.,* 2012), and there is still no information on the characterisation of these proteins in Coleoptera. Therefore, this study represents the first effort to characterise these important proteins in Adephagan beetles.

Cicindelids (tiger beetles), which belong to the suborder Adephaga (Coleoptera), are non-pest species important in ecosystems as predators and are commonly used as bioindicators (Pearson, 1988; Pearson & Cassola, 1992). There are more than 2500 species of cicindelids (Pearson & Cassola, 2005) that occupy a wide range of habitats around the world (Pearson, 1988; Pearson *et al.,* 1988; Cassola & Pearson, 2000). The study of ubiquitins and ubiquitin-like genes in cicindelids could be useful for a better understanding of their genome, their reproductive biology, and the species richness of the group.

In this work we have performed the following: (1) obtained the full length of ubiquitin ribosomal S27 and the SUMO conjugating enzyme E2 (UBC9) from *C. campestris*; (2) carried out phylogenetic analysis of UBS27 and UBC9 genes; (3) amplified a region of both genes in seven species of cicindelids and implemented selection analysis; and (4) investigated the expression pattern of both genes at different developmental stages and in different tissues.

Material and methods

ESTs analysis and gene identification

To identify ubiquitin or ubiquitin-like genes, EST databases of *Calomera littoralis, Cephalota litorea* (Rodríguez-García *et al.,* 2015) and *Cicindela campestris* (Theodorides *et al.,* 2002) were explored. Libraries were assembled and analysed as we previously described (Rodríguez-García *et al.,* 2015). Contigs were annotated via Blast2go v2.5.0 software (Conesa *et al.,* 2005), and nine genes predicted to be ubiquitin or ubiquitin-like genes were identified in the libraries. Two genes annotated as the ubiquitin conjugating enzyme E2 from the *C. littoralis* database and ribosomal ubiquitin from the *C. campestris* library were selected for further analyses.

Sample preparation

Cicindela campestris adults were collected from Laguna del Arquillo (Albacete, Spain). The head, thorax, abdomen and gonads were harvested and submerged in RNAlater (Qiagen, Crawley, UK) and stored at -20 ºC until RNA extraction.

The developmental stages of the gonads were discerned based on the descriptions of Paarmann (1976) for the carabid beetle *Pogonus chalceus* as follows: i) immature females, gonads containing oocytes from the undifferentiated stage to oocytes in the previtellogenesis stage; ii) mature females, containing oocytes in the vitellogenesis stage to mature eggs; iii) immature males, with undifferentiated accessory glands and very small testis that display a transparent white colour; and iv) mature males, testes full of spermatozoa with a characteristic pearl white colour.

Rapid amplification of cDNA ends (RACE)

Total RNA for the RACE reaction was extracted from an adult of *C. campestris* using TRIzol reagent (Life Technologies) following the manufacturer's protocol, and isolated RNA was treated using a TURBO DNA-free Kit (Ambion, Life Technologies) to remove DNA contamination.

Table 1. Primer sequence used for RACE-PCR amplification (RACE), phylogenetic and selection analysis in cicindelids (RC) and real time-quantitative PCR (RT-qPCR) of the Cc-UBC9 and Cc-UBS27 genes. Arginine kinase primers used as endogenous controls in the PCR are also described.

Gene	Primer sequence
Cc-UBS27	**RACE_F1** 5'-TTCGCCTGGGCATTCACGTCTCAAT-3'
	RACE_R1 5'-AAGGAGTCGACTCTGCACTTGGTG-3'
	RC_F1 5' -CCTGGGCATTCACGTCTCAA-3'
	RC_R1 5' -TCGTGAAGACTTTGACGGGT-3'
	RT-qPCR_F1 5'-CGAAGTCGCAGCACCAAGT-3'
	RT-qPCR_R1-5'-CCTGACCAGCAGCGACTGAT-3'
Cc-UBC9	**RACE_F1** 5'-CTGCAGCACGATTAGCCGAGGAACT-3'
	RACE_R1 5'-GACGGGTACACGTTGGGGTGAAACA-3'
	RC_F1 5' -ACACGAAATAATTCTCGCGCC-3'
	RC_R1 5'-TCCTGATGAGACATGACGCG-3'
	RT-qPCR_F1 5'-CGATTAGCCGAGGAACGTAAA-3'
	RT-qPCR_R1 5'-TCCCATGGAGTGCCTTTCTT-3'
Arg kin	**F**-5'-CTCGTGTGGTGCAACGAAGA-3'
	R-5'-GGTGGCTGAACGGGACTCT-3'

The 5´ and 3´ RACE were produced using the ClontechSMARTer RACE cDNA amplification Kit (Takara Bio) following the manufacturer's instructions. Gene-specific primers for 5' and 3' RACE were designed using as reference the original ESTs (Table 1). The synthesis of first cDNA strand was performed following PCR conditions: 30 cycles at 94 ºC for 30 s, 67 ºC for 30 s and 72 ºC for 3 min. PCR products were sequenced in triplicate by SAI at the University of Murcia, Spain to obtain the complete cDNA sequences.

Bioinformatic analyses

Sequences produced by RACE of both genes were edited and assembled by Geneious version 7 (http://www.geneious.com, Kearse *et al.,* 2012). The cDNA sequences were submitted to GenBank (KT877172 and KT877173). The open reading frames (ORF) were generated in ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The protein molecular weight (Mw) and isoelectric point (pI) of both genes were predicted using Compute pI/Mw (http://web.expasy.org/compute_pi/). Protein domains were searched using InterProScan (Jones *et al*., 2014). Homology searches were performed with the BLASTP program at the NCBI.

Phylogeny and selection analysis

To characterise these putative ubiquitins in other cicindelid species, the total RNA was extracted from four individuals of the following species: *Cicindela campestris*, *Calomera littoralis*, *Cephalota litorea*, *Cephalota maura*, *Cephalota deserticoloides, Lophyra flexuosa* and *Cylindera trisignata.* Total RNA was extracted using TRIzol reagent (Life Technologies) following the manufacturer's protocol. Extracted RNA was treated using a TURBO DNA-free Kit (Ambion, Life Technologies) to remove DNA contamination, and reverse-transcribed using the PrimeScript™RT Reagent Kit (Takara). Specific primers were designed to amplify a 443 bp region from UBC9 and 453 bp region from UBS27 (Table 1).

Amplification used the following PCR conditions: 2 min at 94 °C, 40 cycles at 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min and a final extension at 72 °C for 10 min. The samples were sequenced by Macrogen Europe (Amsterdam, The Netherlands)

using a 3730XL DNA sequencer (Applied Biosystems, Foster City, USA). The sequences were deposited in GenBank with accession numbers: KT877174-KT877216.

Homology searches were performed with the BLAST program at NCBI. Multiple sequence alignment was performed with MEGA v6 (Tamura *et al.,* 2013). GARD (genetic algorithm for recombination detection) software (Kosakovsky Pond *et al.,* 2006) implemented on the Datamonkey web server (Kosakovsky Pond & Frost, 2005a) was used to search for recombination evidence. A phylogenetic tree was constructed using the neighbour-joining method in MEGA v6 software (Tamura *et al.,* 2013). Bootstrap values were calculated using 1000 replicates.

Several methods were used to test for selection of the genes under study. Only a consensus sequence for each species of cicindelids was used. Selection is detected by comparing the number of nonsynonymous (dN) and synonymous (dS) substitutions per site in a given gene. Purifying selection is inferred when dS is greater than dN. A Z-test (codon based) for purifying selection was performed by MEGA v6 (Tamura *et al.,* 2013). The program was run under the modified Nei–Gojobori method (Zhang *et al.,* 1998), covering the overall average with 1000 bootstraps and treating missing data with pairwise deletions. To detect sites under selection, genes were also analysed using HyPhy software implemented on the Datamonkey web server (Kosakovsky Pond & Frost, 2005a). Two classic ML methods were used, the Single Likelihood Ancestor Counting (SLAC) model, and the Fixed Effect Likelihood (FEL) model (Kosakovsky Pond & Frost, 2005b). Additionally, we used the Fast Unbiased Bayesian Approximation (FUBAR) method that introduces an ultra-fast Markov chain Monte Carlo (MCMC) allowing the visualization of Bayesian inference for each site (Murrell *et al.,* 2013). The cut-off values of the p-values were <0.1 for SLAC and FEL and a posterior probability >0.90 for FUBAR.

Real-time quantitative PCR

The expression of the ubiquitin genes was investigated using real-time quantitative PCR (RT-qPCR) analysis. Gene expression levels were examined in the following tissues in adult individuals of *C. campestris*: head, thorax, abdomen and gonads (testis or ovaries) from males and females at mature, and immature developmental stages.

Equal amounts of total RNA from three different individuals were pooled. The RNA extraction and cDNA synthesis protocol were described above. The specific primers for RT-qPCR were designed based on gene sequences using Primer Express 3 software (Applied Biosystems®) (Table 1). Gene efficiency was first validated with the quantification–standard curve method using a serially diluted pool of cDNA. The gene expression experiment was carried out using the comparative C_T method (Livak & Schmittgen, 2001) with arginine kinase as the endogenous control and the immature head sample as the reference. Experiments were carried out on a StepOnePlus™ Real-Time PCR System using SYBR® Green (Applied Biosystems®). The PCR conditions were as follows: one cycle at 95 °C for 2 min and 40 cycles at 95 °C for 15 s and 60 °C for 30 s. Each sample had three technical replicates, and the experiment was repeated three times with independent RNA samples to assure the repeatability of the results. All data were analysed with the program 7500 software v2.0.5 (Applied Biosystems®).

Results

Characterisation of *C. campestris* UBC9 and UBS27

The full-length Cc-UBC9 cDNA from *C. campestris* was 650 bp long (GenBank number: KT877172). The ORF was predicted to encode a 159 amino acid protein, with residues 7 to 157, forming an UBCc domain, and containing the active Cys93 residue indispensable for SUMOylation (Fig 1a). The predicted molecular weight was 18.18 KDa and the isoelectric point was 8.57. The putative protein showed a high degree of identity with SUMO-conjugating enzyme E2 (Ubc9) of other insects. Cc-UBC9 shares 92% identity with predicted SUMO proteins of *Blattella germanica* (CE043914.1), *Apis mellifera* (XP0065664777.1), and 91% with *Tribolium castaneum* (XP967918.1).

The full-length Cc-UBS27 cDNA, isolated from *C. campestris*, was 670 bp long (GenBank number: KT877173). The ORF was predicted to encode a 156 amino acid protein, with residues 1 to 76 forming the de ubiquitin domain, and 102 to 147 forming the ribosomal S27 domain (Fig 1b). The predicted molecular weight was

17.71 kDa, and the isoelectric point was 9.91. The putative protein showed a high degree of identity with ubiquitin ribosomal protein S27 of other insects. Cc-UBS27 shares 97% identity with *Carabus granulatus* (CAH04347), 96% identity with *Tribulim castaneum* (XP969023.1) and 95% with *Agriotes lineatus* (CAJ01876.1), among others.

A:Cc-UBC9

```
  1 cttggaaaataactgtccatagtgaagcatacatcatttacacgaaataattctcgcgcc
 61 aattatcttttaatttatgtttactgttagttaaaagttgcagatatgtcgggtattgct
  1                                                 M  S  G  I  A
121 gcagcacgattagccgaggaacgtaaagcctggagaaaagaccacccttcggatttgtt
  6  A  A  R  L  A  E  E  R  K  A  W  R  K  D  H  P  F  G  F  V
181 gcaaggccggccaaaaacgcggatggctccttaaaccttatgaactgggaatgctgcatt
 26  A  R  P  A  K  N  A  D  G  S  L  N  L  M  N  W  E  C  C  I
241 cctggcaagaaaggcactccatgggagggtggtcactacaaattacgtatgattttcaaa
 46  P  G  K  K  G  T  P  W  E  G  G  H  Y  K  L  R  M  I  F  K
301 gaggattacccgagcagtcctccgaaatgcaaatttgagccggctctgtttcaccccaac
 66  E  D  Y  P  S  S  P  P  K  C  K  F  E  P  A  L  F  H  P  N
361 gtgtacccgtcagggactgtatgcttgtcactattggatgaagaaaaagattggcgcccc
 86  V  Y  P  S  G  T  V  C  L  S  L  L  D  E  E  K  D  W  R  P
421 gctattactatcaaacaaattttactgggcattcaggacttgctcaatgagcccaatgtc
106  A  I  T  I  K  Q  I  L  L  G  I  Q  D  L  L  N  E  P  N  V
481 aaggacccggcacaggcggaggcctacactatctactgccaaaaccgaatggagtatgaa
126  K  D  P  A  Q  A  E  A  Y  T  I  Y  C  Q  N  R  M  E  Y  E
541 aaacgcgtgcgggcccaggcacgcgtcatgtctcatcaggagtaagatggattcgcttca
146  K  R  V  R  A  Q  A  R  V  M  S  H  Q  E  *
601 atacttccccaagcttactttactgaaataacgcccttgaaaaaa 645
```

B:Cc-UBS27

```
  1 ttttttttttttttttttttttttttcagtgtgacatggtggtgtaaaaggtaaataa
 61 atcaacgtcatcatgcagattttcgtgaagactttgacgggtaagaccatcacccttgag
  1              M  Q  I  F  V  K  T  L  T  G  K  T  I  T  L  E
121 gttgaaccgtctgataccatcgaaaatgtgaaggcgaagatccaggataaggagggaatc
 17  V  E  P  S  D  T  I  E  N  V  K  A  K  I  Q  D  K  E  G  I
181 ccacctgaccagcagcgactgattttcgccggcaagcagttggaggatggtcgcaccctc
 37  P  P  D  Q  Q  R  L  I  F  A  G  K  Q  L  E  D  G  R  T  L
241 tccgactataacatccagaaggagtcgactctgcacttggtgctgcgacttcgcggtggc
 57  S  D  Y  N  I  Q  K  E  S  T  L  H  L  V  L  R  L  R  G  G
301 gctaagaagcgcaagaagaagaattattccacacctaagaagatcaagcacaagaagaag
 77  A  K  K  R  K  K  K  N  Y  S  T  P  K  K  I  K  H  K  K  K
361 aaggttaagctagctgtgcttaaattttataaggtggacgaaaacggtaaaatccataga
 97  K  V  K  L  A  V  L  K  F  Y  K  V  D  E  N  G  K  I  H  R
421 ttgagacgtgaatgcccaggcgaacagtgcggtgcaggagtgttcatggccgctatggaa
117  L  R  R  E  C  P  G  E  Q  C  G  A  G  V  F  M  A  A  M  E
481 gacaggcactactgtggaaaatgcggctacacactagtcttctcgaaacccgaagacaag
137  D  R  H  Y  C  G  K  C  G  Y  T  L  V  F  S  K  P  E  D  K
541 taactctgtcgctgacattctaacttttttttataaatttactaaattaggatgtggaaa
157  *
601 aatataagtaaaacaaactgaaacgttcacacacacacacgtgaaacacacactaaca
661 taaatattt 669
```

Figure 1. Complete cDNA sequences and deduced amino acid sequences from *C. campestris* of **A)** Cc-UBC9 (GenBank accession number: KT877172). The UBCc domain of UBC9 is shown with grey highlighting, the conserved active site Cys93 is shown in bold and italics. **B)** Cc-UBS27 (GenBank accession number: KT877173). The ubiquitin sequence and the S27 domain of UBS7 is shown with grey highlighting, and the diglycine motif is in bold.

Phylogenetic and evolutionary analysis

In UBC9 a sequence of 443 bp were obtained in the seven species tested, showing high within-species conservation and having 76 variable positions among species that were actually synonymous substitutions. In UBS27, a sequence of 453 bp were obtained in six species tested with the exception of *Cylindera trisignata,* where no positive amplification could be obtained. The sequences presented 46 variable positions among species, corresponding to synonymous substitutions.

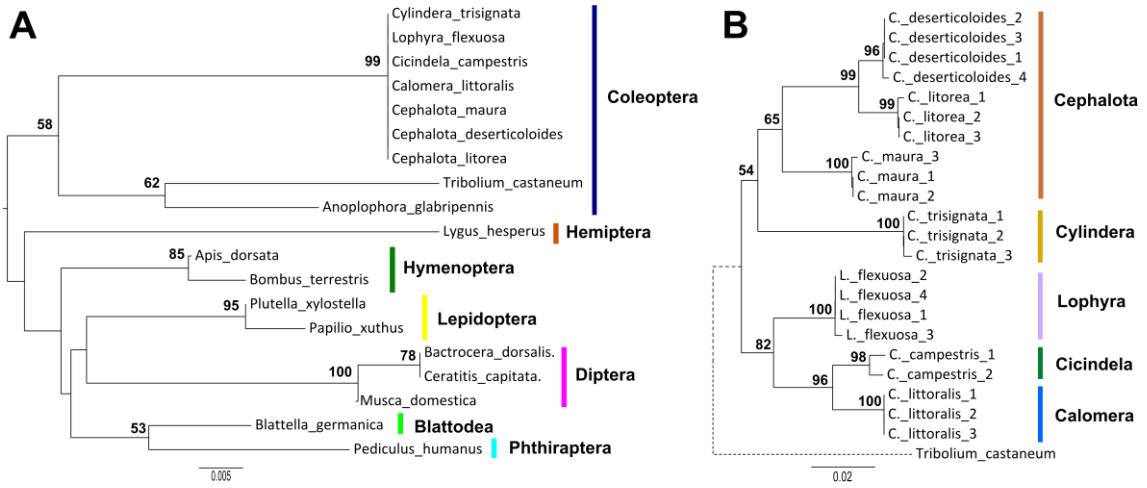


Figure 2. A) Neighbour-joining phylogenetic tree of UBC9 in representatives of the insect orders. The numbers in the nodes indicate the percentage of bootstrap support. The bar indicates the genetic distance. The GenBank IDs are as follows: *Tribolium castaneum* (XP_967918), *Anoplophora glabripennis* (XP_JAB64749), *Lygus Hesperus* (JAG34846), *Apis dorsata* (XP_006614317), *Bombus terrestris* (XP_003399340), *Plutella xylostella* (XP_011554291), *Papilio xuthus* (XP_013164502), *Bactrocera dorsalis* (JAC54005), *Ceratitis capitata* (XP_004537820), *Musca domestica* (XP_005181569), *Blattella germanica* (CEO43914), and *Pediculus humanus* (XP_002426763). **B)** Neighbour-joining phylogenetic tree based on cicindelids nucleotide sequences of UBC9. GenBank ID: KT877174- KT877195.

The GARD software results showed no significant recombination in any of the two genes analysed. To investigate the phylogenetic relationships of UBC9 and UBS27 genes within cicindelids and with other insect homologues, an unrooted neighbour-joining tree was constructed using protein sequences of several insect orders. The UBC9 phylogenetic tree showed that all of the cicindelids clustered together and were located in a different branch than the other coleopteran species analysed (*Tribolium castaneum* and *Anoplophora glabripennis*) that belonged to the suborder

Polyphaga (Fig 2). In the UBS27 phylogenetic tree, all the coleopteran sequences are clustered in the same branch, showing that the differences in the amino acid sequences for this gene are minimal within the Coleoptera (Fig 3). Additionally, two phylogenetic neighbour-joining trees were obtained using the nucleotide sequences of the cicindelids, grouping together sequences from the same species (Fig 2b, 3b), and the results agree with the relationships obtained with other markers such as the mitochondrial gene cox-1 (López-López & Galián, 2010).

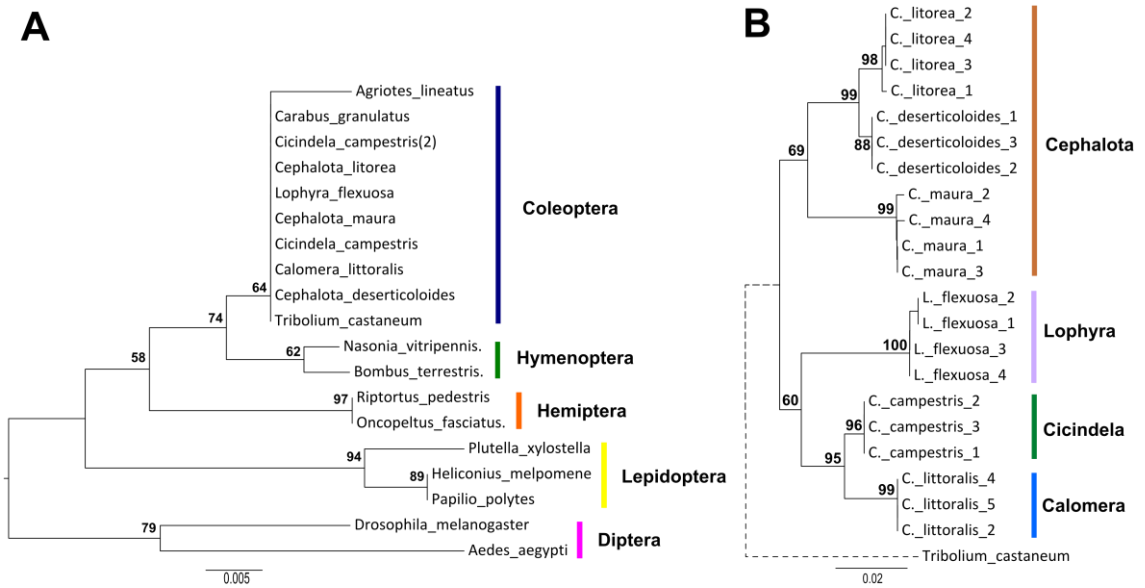


Figure 3. A) Neighbour-joining phylogenetic tree of UBS27 in representatives of the insect orders. The numbers in the nodes indicate the percentage of bootstrap support. The bar indicates the genetic distance. GenBank IDs are as follows: *Agriotes lineatus* (CAJ01876), *Cicindela campestris* (2) (CAJ01877), *Carabus granulatus* (CAH04347), *Tribolium castaneum* (XP_969023), *Nasonia vitripennis* (XP_003400429), *Bombus terrestris* (XP_008216204), *Riptortus pedestris* (BAN20166), *Oncopeltus fasciatus* (ABN54483), *Plutella xylostella* (AGC11947), *Heliconius melpomene* (AEL28836), *Papilio polytes* (XP_013139400), *Drosophila melanogaster* (NP_476778), and *Aedes aegypti* (AA579344). **B)** Neighbour-joining phylogenetic tree based on cicindelids nucleotide sequences of UBS27. GenBank ID: KT877196- KT877216.

Consensus partial protein-coding sequences for each species of the cicindelids were used to determine the selection of UBC9 and UBS27 genes. Seven species were analysed (*Cicindela campestris*, *Calomera littoralis*, *Cephalota litorea*, *Cephalota maura*, *Cephalota deserticoloides, Lophyra flexuosa* and *Cylindera trisignata*) for UBC9, and six were analysed for UBS27 (except for *C. trisignata*). Purifying selection was conducted under the modified Nei–Gojobori model, treating the data as the overall average and

using the bootstrap resampling method, which resulted in values equal to zero in both genes, thus rejecting the null hypothesis (H_0: dS=dN) and accepting the alternative hypothesis of purifying selection (H_A: dN<dS). The result of the statistic test (dS-dN) was 7.227 for UBS27 and 10.825 for UBC9.

Additionally the analysis for detecting sites under selection also revealed negative or purifying selection. For UBC9, the SLAC analysis revealed no positively selected and 9 negatively selected sites ($p<0.1$). The FEL analysis identified 43 sites influenced by negative selection ($p<0.1$), and FUBAR identified 31 purifying selected sites ($pp>0.9$). Similarly, for UBS27, the SLAC analysis revealed no positively selected and 2 negatively selected sites ($p<0.1$). The FEL identified 25 sites influenced by negative selection ($p<0.1$), and FUBAR identified 15 purifying selected sites ($pp>0.9$).

Tissue distribution and gene expression pattern of Cc-UBC9 and Cc-UBS27 at immature and mature developmental stages.

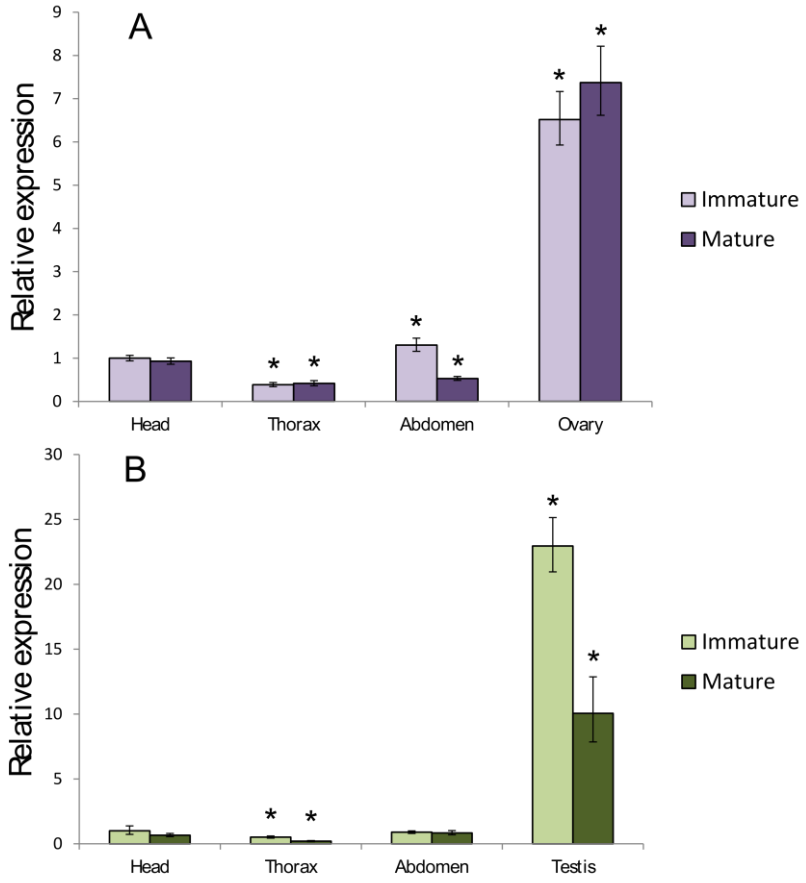


Figure 4. Relative expression of UBC9 normalized to Arginine kinase endogenous in the head, thorax and gonads of *C. campestris*. **A)** Female **B)** Male

The expression level of Cc-UBC9 and Cc-UBS27 was examined by real-time quantitative PCR (RT-qPCR) in different tissues, including the head, thorax, abdomen and gonads of males and females of *C. campestris* at different developmental stages. The RT-qPCR results showed that the expression pattern of both genes was similar (Fig 4 and 5). UBS27 and UBC9 were detected in all the tissues examined, and the gonads showed the highest relative expression, which was higher in the testes than the ovaries. The expression levels changed significantly between different developmental stages of the testis, as the uppermost expression of Cc-UBC9 and Cc-UBS27 were found in the immature stages. For the ovaries, no significant differences were found between the immature and mature stages.

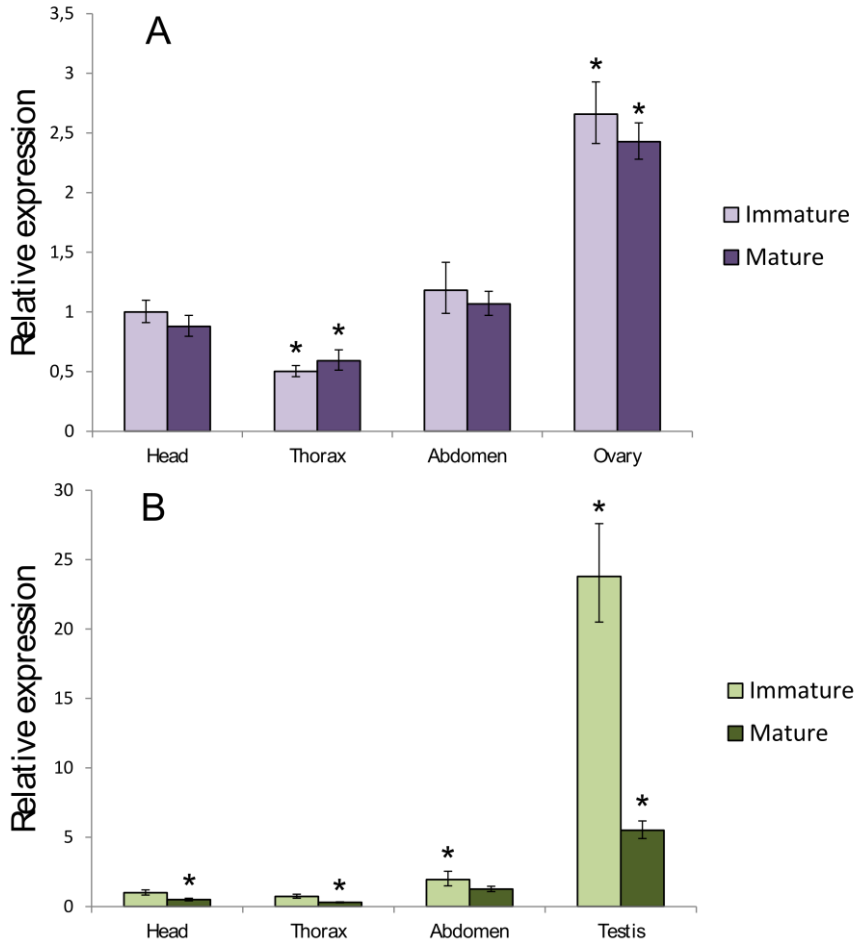


Figure 5. Relative expression of UBS27 normalized to Arginine kinase endogenous in the head, thorax and gonads of *C. campestris*. **A)** Female **B)** Male

Discussion

Considering all of the evidence, the bioinformatic analyses suggest that Cc-UBS27 is a ubiquitin ribosomal fusion protein S27 and that Cc-UBC9 is a SUMO-conjugating enzyme E2. In the present study the full-length cDNA sequence of the SUMO-UBE2 (UBC9) gene from *C. campestris* was isolated. The Cc-UBC9 gene belongs to the ubiquitin conjugating enzyme (E2s) family and possesses a highly conserved ubiquitin-conjugating catalytic (UBCc) domain, which contains an invariant cysteine residue (Cys 93) for thioester formation (Jentsch, 1992). Additionally, we isolated the full-length cDNA sequence of the UBS27 gene from *C. campestris*. In UBS27, the UB domain acts as a chaperone, protecting s27 from degradation, and a diglycine is the recognition motif for a specific endopeptidase (Amerik & Hochstrasser, 2004).

Ubiquitin and its relatives, the small ubiquitin–like modifiers (SUMO), are conjugated to proteins, modifying their properties, being essential for cell cycle regulation. Several genes that belong to ubiquitin or SUMO pathways are involved in gametogenesis among other functions (Shen *et al.,* 2009; Zhang *et al.,* 2010; Wang *et al.,* 2012a, b). Ubiquitin and SUMO genes have been characterised in human and several model species, including *Drosophila* (Smith *et al.,* 2012). However, there is a lack of knowledge on the functioning of these genes in Coleoptera. This study represents the first time that ubiquitin (UBS27) and SUMO (UBC9) genes have been identified and characterised and their expression profiles have been analysed in the developing gonads of an adephagan beetle species.

The amino acid sequences of both genes and the phylogenetic analysis showed that these genes are highly conserved. According to these results, selection analysis revealed that the UBS27 and UBC9 genes are affected by purifying selection in cicindelids. Our results on UBS27 are consistent with what Nei *et al.* (2000) described for the ubiquitin gene family. Likewise, our results on UBC9 agree with the results of Ying *et al.* (2009) for the ubiquitin-conjugating enzymes. Purifying selection acts to remove genetic variations when mutations have effects on fitness, a finding that supports the importance of these proteins. Although this type of selection is conservative, the information obtained from the study of these genes in several cicindelid species may be relevant to better understand the diversification and

speciation processes, as genes that do not evolve rapidly are also important to understand the evolutionary process according to Swanson & Vaquier (2002).

Results of the expression analyses showed the same pattern for Cc-UBC9 and Cc-UBS27 genes; they were detected in all the tissues, with the gonads (testis and ovary) showing the highest expression level. In the gametogenesis process, oogenesis and spermatogenesis are actively driven by mitosis and meiosis. Studies in other organisms show that SUMO and ubiquitin genes are active during meiosis and mitosis. UBC9 regulates meiosis and mitosis processes (Watanabe *et al.,* 1996; Sakaguchi *et al.,* 2007), and this gene is reported to be involved in embryogenesis, gametogenesis and sex modifications in the fish *Cynoglossus semilaevis* (Hu & Chen, 2013), and it is also involved in adult *Drosophila* gonad development (Hashiyama *et al.,* 2009). Similarly UBS27 is involved in gametogenesis in the crab *Eriocheir sinensis* (Wang *et al.,* 2012a).

Previous studies of both genes in *E. sinensis* showed variable expression across the developmental stages in the testis and ovary. The highest expression corresponds to the period of fast growth in the earlier stages of gonad development, when the testes are full of spermatocytes and the ovaries are full of oocytes (Wang *et al.,* 2012a, b). Our results also showed differential expression of the UBC9 and UBS27 genes between mature and immature testis in *C. campestris*. We established that testis were in the immature stage when the tubular testis have active spermatogonial mitoses and meiosis, so that it is full of spermatocytes. The lower expression of these genes in the mature testis corresponds with a tubular testis full of spermatozoa, which is also in agreement with the results of Wang *et al.,* (2012a, b). In contrast, in the ovaries, no significant differences in gene expression between immature and mature stages were observed in both genes, as reported previously in crabs by Wang *et al.* (2012a, b). We distinguished immature from mature ovaries by the absence or presence of visible eggs in the ovarioles. Additional experiments using ovaries that were even less developed may help to discern whether in cicindelids there is also a maximum expression that corresponds to the oocyte stage, as observed in crabs.

In conclusion, our selection analysis showed that UBC9 and UBS27 genes are affected by purifying selection in cicindelids; both genes showed higher expression in

gonads in comparison with other body tissues, which suggests an active role in gonad development processes. This finding also agrees with changes in gene expression during gonad development in males (results on ovaries are not conclusive). This is the first report of the expression of UBC9 and UBS27 genes in beetles. The characterisation of these genes provides valuable information on the reproductive processes in cicindelids and are useful for elucidating the function of ubiquitins and ubiquitin-like genes in the Coleoptera.

Acknowledgements

This work has been funded by project CGL2008-03628 of the Ministerio de Ciencia e Innovación, Spain. Thanks are due to Willey Editing Services for language editing of the paper.

References

Amerik, A.Y. and Hochstrasser, M. (2004) Mechanism and function of deubiquitinating enzymes. *BBA-Mol Cell Res* **1695**: 189-207.

Baker, R.T. and Board, P.G. (1991) The human ubiquitin-52 amino acid fusion protein gene shares several structural features with mammalian ribosomal protein genes. *Nucleic Acids Res* **19**: 1035-1040.

Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R. *et al.* (1998) Structure determination of the small ubiquitin-related modifier SUMO-1. *J Mol Biol* **280**: 275-286.

Bebington, C., Doherty, F.J. and Fleming, S. D. (2001) The possible biological and reproductive functions of ubiquitin. *Hum Reprod Update* **7**: 102-111.

Cassola, F. and Pearson, D.L. (2000) Global patterns of tiger beetle species richness (Coleoptera: Cicindelidae): their use in conservation planning**.** *Biol Conserv* **95:** 197-208.

Chen, Z.K., Zhou, J., Bao, L.J., Xu, J., Shen, W.D. and Lu, X.P. (2007) Cloning and expression of ubiquitin gene of *Bombyx mandarina*. *Canye Kexue* **33**: 297-301.

Conesa, A., Gotz, S., Garcia-Gomez, J., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674-3676.

Finley, D., Bartel, B. and Varshavsky, A. (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitins facilitates ribosome biogenesis. *Nature* **338**: 394-401.

Gill, G. (2004) SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes Dev* **18**: 2046-2059.

Goldstein, G., Scheid, M., Hammerling, U., Shlesinger, D.H., Niall, H.D. and Boyse, E.A. (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *P Natl Acad USA* **72**: 11-15.

Haas, A.L., Warms, J.V.B., Hershko, A. and Rose, I.A. (1982) Ubiquitin activating enzyme: mechanism and role in protein-ubiquitin conjugation. *J Biol Chem* **257:** 2543-2548.

Hashiyama, K., Shigenobu, S. and Kobayashi, S. (2009) Expression of genes involved in sumoylation in the *Drosophila* germline. *Gene Expr Patterns* **9**: 50-53.

Hershko, A. and Ciechanover, A. (1998) The ubiquitin system. *Annu Rev Biochem* **67**: 425-479.

Hu, Q. and Chen, S. (2013) Cloning, genomic structure and expression analysis of ubc9 in the course of development in the half-smooth tongue sole (*Cynoglossus semilaevis*). *Comp Biochem Physiol B: Biochem Mol Biol* **165:** 181-188.

Jansa, S.A., Lundrigan, B.L. and Tucker, P.K. (2003) Tests for positive selection on immune and reproductive genes in closely related species of the murine Genus *Mus*. *J Mol Evol* **56**: 294-307.

Jentsch, S. (1992) The ubiquitin-conjugation system. *Annu Rev Genet* **26**: 179-207.

Jentsch, S., Seufert, W. and Hauser, H. P. (1991) Genetic analysis of the ubiquitin system. *Biochim Biophys Acta* **1098:** 268-277.Jin, F.L., Xu, X.X., Zhang, W.Q. and Ren, S.X. (2008) Cloning and prokaryotic expression of the cDNA sequence encoding ubiquitin from *Musca domestica*. *Acta Entomol Sin* **51**: 473-479.

Johnson, E. S. (2004) Protein modification by SUMO. *Ann Rev Biochem* **73**: 355-382.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.

Koken, M.H., Hoogerhrugge, J.W., Jasper, I., de Wit, J., Willemsen, R., Roest H.P. *et al.* (1996) Expression of the ubiquitin-conjugating DNA repair enzymes HHR6A and B suggests a role in spermatogenesis and chromatin modification. *Dev Biol* **173**: 119-132.

Kosakovsky Pond, S.L. and Frost, S.D. (2005a) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**: 2531-2533.

Kosakovsky Pond, S.L. and Frost, S.D. (2005b) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**: 1208-1222.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. and Frost, S.D. (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**: 3096-3098.

Li, Z.F., Gong, Y.X., Shi, Y.X., Pan, L.J. and Pang, Y. (2003) Cloning and expression of *Spodoptera litura* ubiquitin gene. *Insect Sci* **10**: 27-34.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2<- ΔΔCT< Method. *Methods* **25**: 402-408.

López-López, A. and Galián, J. (2010) Análisis filogenético de los Cicindelini ibéricos (Coleoptera; Carabidae: Cicindelinae). *An Biol* **32**: 76-86.

Metz, E.C., Gómez Gutierrez, G. and Vacquie, V. D. (1998) Mitochondrial DNA and binding gene sequence evolution among allopatric species of the sea urchin genus. *Mol Vol Evol* **15:** 185-195.

Mezquita, J., Pau, M. and Mezquita, C. (1997) Characterization and expression of two chicken cDNAs encoding ubiquitin fused to ribosomal proteins of 52 and 80 amino acids. *Gene* **195:** 313-319.

Mitchell, M. J., Woods, D. R., Tucker, P.K., Opp, J.S. and Bishop, C.E. (1991) Homology of a candidate spermatogenic gene from the mouse Y chromosome to the ubiquitin activating enzyme E1. *Nature* **354**: 483-486.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L. *et al.* (2013) FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection. *Mol Biol Evol* **30**: 1196-1205.

Nei, M., Rogozin, I.B. and Piontkivska, H. (2000) Puryfing selection and birth-and-death evolution in the ubiquitin gene family. *P Natl Acad USA* **97:** 10865-10871.

Nenoi, M., Ichimura, S. and Mita, K. (2000) Interspecific comparison in the frequency of concerted evolution ay the polyubiquitin gene locus. *J Mol Evol* **51:** 161-165.

Ozkaynak, E., Finley, D., Solomon, M.J. and Varshavsky, A. (1987) The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J* **6:** 1429-1439.

Paarmann, W. (1976) the annual periodicity of the polivoltine ground beetle *Pogonus chalceus* Marsh. (Col., Carabidae) and its control by environmental factors. *Zool Anz Jena* **196**: 150-160.

Pearson, D.L. (1988) Biology of Tiger Beetles**.** *Ann Rev Entomol* **33**: 123-147.

Pearson, D.L. and Cassola, F. (1992) World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): indicator taxon for biodiversity and conservation studies. C*onserv Biol* **6**: 376-391.

Pearson, D.L. and Cassola, F. (2005) A quantitative analysis of species descriptions of tiger beetles (Coleoptera: Cicindelidae), from 1758 to 2004, and notes about related developments in biodiversity studies. *Coleops Bull* **59**: 184-193.

Pearson, D.L., Blum, M.S., Jones, T.H., Fales, H.M., Gonda, E. and White, B.R. (1988) Historical perspective and the interpretation of ecological patterns: defensive compounds of tiger beetles (Coleoptera: Cicindelidae). *Am Nat* **132**: 404-416.

Pickart, C.M. and Eddins, M.J. (2004) Ubiquitin: structures, function, mechanism. *Biochism Biophys Acta* **1695**: 55-72.

Rodríguez-García, M.J., Machado, V. and Galián, J. (2015) Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles. *BMC Genomics* **16:** 391.

Roest, H.P., van Klaveren, J., de Wit, J., van Gurp, C.G., Koken, M.H.M., Vermey, M. *et al.* (1996) Inactivation of the HR6B ubiquitin-conjugating DNA repair enzyme in mice causes a defect in spermatogenesis associated with chromatin modification. *Cell* **86**: 799-810.

Sakaguchi, K., Koshiyama, A. and Iwabata, K. (2007) Meiosis and small ubiquitin-related modifier (SUMO) -conjugating enzyme, Ubc9. *Febs Journal* **274:** 3519-3531.

Shen, B., Zhang , Z., Wang, Y., Wang, G., Chen, Y., Lin, P. *et al.* (2009) Differential expression of ubiquitin-conjugating enzyme E2r in the developing ovary and testis of penaeid shrimp *Marsupenaeus japonicus. Mol Biol Rep* **36:** 1149-1157.

Smith, M., Turki-Judeh, W. and Courey, A. J. (2012) SUMOylation in *Drosophila* development. *Biomolecules* **2**: 331-349.

Swanson, W.J. and Vacquier, V.D. (2002) The rapid evolution of reproductive proteins. *Nat Rev Genet* **3:** 137-144.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729.

Takagi Sawada, M., Kyozuka, K., Morinaga, C., Izumi, K. and Sawada, H. (1997) The proteasome is an essential mediator of the activation of pre-MPF during starfish oocyte maturation. *Biochem Bioph Res Commun* **236**: 40-43.

Theodorides, K., De Riva, A., Gómez-Zurita, J., Foster, P.G. and Vogler, A.P. (2002) Comparison of EST libraries from seven beetle species: towards a framework for phylogenomics of the Coleoptera. *Insect Mol Biol* **11**: 467-47.

Walsh, C.T., Garneau-Tsodikova, S. and Gatto, G.J. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Edit* **44**: 7342-7372.

Wang, Q., Chen, L., Wang, Y., Li, W., He, L. and Jiang, H. (2012a) Expression characteristics of two ubiquitin/ribosomal fusion protein genes in the developing testis, accessory gonad and ovary of Chinese mitten crab, *Eriocheir sinensis*. *Mol Biol Rep* **39**: 6683-6692.

Wang, Q., Wang, Y., Chen, L., He, L., Li, W. and Jiang, H. (2012b) Expression characteristics of the SUMOylation genes SUMO-1 and Ubc9 in the developing testis and ovary of Chinese mitten crab, *Eriocheir sinensis*. *Gene* **501**: 135-143.

Watanabe, T.K., Fujiwara, T., Kawai, A., Shimizu, F., Takami, S., Hirano, H. *et al.* (1996) Cloning, expression, and mapping of UBE2I, a novel gene encoding a human homologue of yeast ubiquitin-conjugating enzymes which are critical for regulating the cell cycle. *Cytogenet Genome Res* **72**: 86-89.

Yang, P., Zhou, W.W., Zhang, Q., Cheng, J., Zhu, Z. and Way, M.O. (2009) Differential gene expression profiling in the developed ovaries between the parthenogenetic and bisexual female rice water weevils, *Lissorhoptrus oryzophilus* Kuschel (Coleoptera: Curculionidae). *Chinese Sci Bull* **54**: 3822-3829.

Ying, M., Zhan, Z., Wang, W. and Chen, D. (2009) Origin and evolution of ubiquitin-conjugating enzymes from *Guillardia theta* nucleomorph to hominoid. *Gene* **447**: 72-85.

Yu, H., Jin, F.L., Xu, X.X., Yang, X., Gu, D.X. and Zhang, W.Q. (2004) Cloning and sequence analysis of an ubiquitin gene of *Blattella germanica*. *Acta Entomol Sin* **47**: 522-525.

Zhang, F., Chen, L., Wu, P., Zhao, W., Li, E. and Qin, J. (2010) cDNA cloning and expression of Ubc9 in the developing embryo and ovary of oriental river prawn, *Macrobrachium nipponense*. *Comp Biochem Physiol B: Biochem Mol Biol* **155**: 288-293.

Zhang, J., Rosenberg, H.F. and Nei, M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *P Natl Acad USA* **95**: 3708-3713.

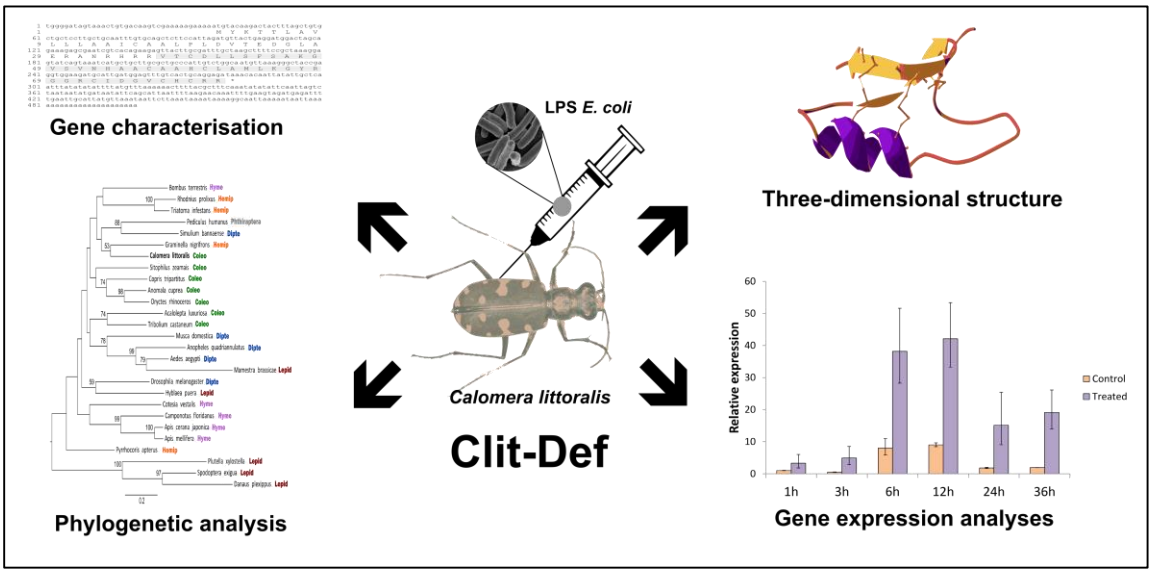
Chapter 3

Identification, structural characterisation and expression analysis of a defensin gene from the tiger beetle *Calomera littoralis* (Coleoptera: Cicindelidae)

Identification, structural characterisation and expression analysis of a defensin gene from the tiger beetle *Calomera littoralis* (Coleoptera: Cicindelidae)

María Juliana Rodríguez-García[1*], **Andrés García-Reina**[1], **Vilmar Machado**[1] and **José Galián**[1]

[1] *University of Murcia, Department of Zoology and Physical Anthropology, Faculty of Veterinary, Campus Mare Nostrum, E-30100 Murcia, Spain.*



Abstract

Introduction Defensins are important antimicrobial peptides (AMP), which play a crucial role in the innate immune systems of invertebrates. They are predominantly active against Gram (+) bacteria, but also against Gram (-) bacteria, fungi and some viruses. Defensins have been characterised in several coleopteran species, nevertheless there is no report of any defensin genes in the suborder Adephaga.

Results. In this study, a defensin gene has been characterised in the tiger beetle *Calomera littoralis* for the first time and was designated as Clit-Def. Bioinformatic analysis showed that the Clit-Def gene has an open reading frame of 246 bp that contains a 46 bp mature peptide. The Clit-Def mature peptide has the features that have been reported to be involved in the antimicrobial function: a predicted cationic isoelectric point of 8.94, six cysteine residues that form three disulphide bonds, and the typical CSαβ structural fold. Real time quantitative PCR analysis showed that Clit-Def were upregulated in the different tissues analysed after infection with lipopolysaccharide of *Escherichia coli*, and also indicated that has a expression peak at 12 h post infection. The phylogenetic analyses showed a high variability in the coleopteran defensins analysed.

Conclusions. The expression patterns of Clit-Def suggest that this gene plays important roles in the humoral system in *C. littoralis*. This information of a cicindelid species will allow to better understand the structure and function of the immune system in Coleoptera.

Key words: defensin, AMP, cicindelids, adephagan beetles, RT-qPCR

Introduction

Insects lack an adaptive immunity system and so their innate immune system is particularly important. Its efficiency is thought to be one of the biology attributes that would explain the evolutionary success of insects (Bulet & Stöcklin, 2005). Their immune system is comprised of cellular response (phagocytosis and encapsulation of invading microorganisms) and humoral immune response (melanization and synthesis of antimicrobial peptides (AMPs) (Hoffmann, 1997). AMPs play a crucial role in the rapid elimination of invading pathogens and are classified into three groups based on their amino acid sequence and structural features: (i) linear peptides which form an α-helix and lack cysteine residues; (ii) peptides with proline and/or glycine residues over-represented; and (iii) cysteine rich peptides (Bulet *et al.,* 1999). Insect defensins belong to the third group. They are small, cationic peptides, with six conserved cysteine residues that are linked in three pairs of disulfide bridges (Cys^{I}-Cys^{IV}, Cys^{II}-Cys^{V} and Cys^{III}-Cys^{VI}) (Ganz, 2003; Hazlett & Wu, 2011; Yi *et al.,* 2014). They have a consensus CSαβ motif that is the active site of the domine for antimicrobial activity (Bulet & Stöcklin, 2005). Defensins are predominantly active against Gram (+) bacteria, but also against Gram (-) bacteria, fungi and viruses (Bulet & Stöcklin, 2005; Pasupuleti *et al.,* 2012). Their action mechanisms are membrane disruption (Gao & Zhu, 2012), cell wall synthesis inhibition (Schneider *et al.,* 2010) and interactions with intracellular targets (Zhu *et al.,* 2012). This broad spectrum of anti-microbial activity make defensins excellent candidates to develop novel antimicrobial agents (Landon *et al.,* 2008), and therefore contribute to solve the problem of resistance to drugs in bacteria (Pasupileti *et al.,* 2012).

Immune genes evolve faster than non-immune–related genes (Lazzaro & Little, 2009). The evolution is marked by expansions and contractions of gene families and the host-pathogen arm race that is driven by positive selection (Viljakainen, 2015). AMPs showed a rapid diversification and species specificity (Vilcinkas, 2013), and particularly, defensins has been affected by positive selection in ants and termites (Bulmer & Crozier, 2004; Viljakainen & Pamilo, 2008).

Insect defensins were first reported in the Diptera *Sarcophaga peregrine* (Matsuyama & Natori, 1988) and *Phormia terranovae* (Lambert *et al.,* 1989), and since then they have

been mainly described in other species of the order Diptera, and also Hemiptera and Hymenoptera among others. Nevertheless, the characterization of defensin in the order Coleoptera, the most diverse order of insects, is limited to the suborder Polyphaga, and they have been isolated in the following families: Tenebrionidae (Rajamuthiah *et al.,* 2015) Cerambycidae (Ueda *et al.*, 2005), Curculionidae (Anselme *et al.,* 2008) and Scarabaeidae (Ishibashi *et al.,* 1999; Sagisaka *et al.,* 1999; Hwang *et al.,* 2009).

Cicindelids (tiger beetles) belong to the suborder Adephaga (Coleoptera), the second largest after the Polyphaga, and includes non-pest predatory species that are significant components of ecosystems, being important in the trophic chains. Tiger beetles comprise more than 2500 species (Pearson & Cassola, 2005), which are worldwide distributed (except for Tasmania, Antarctica, and some Oceanic islands), showing hot spots of species richness, such as the southwestern United States, southwestern Asia, northeastern India and southwestern Amazonia (Willis, 1972; Ali, 1978; Pearson, 1984; Pearson & Ghorpade, 1989; Cassola & Pearson, 2000). They occupy a great diversity of habitats (alpine meadows, dessert grassland, and tropical rain forest among others) in which species tend to specialise (Pearson, 1988, 1992; Pearson *et al.,* 1988). The study of defensin genes in cicindelids will be useful to better understand defense mechanisms in the Coleoptera, and will serve as a foundation for further evolutionary studies of their immune system.

In this work we have performed the following: i) characterised a defensin gene from *Calomera littorialis* (Clit-Def), including the full length cDNA sequence and the predicted three dimensional structure; ii) carried out phylogenetic analysis; and iii) investigated the expression profile after immunization with *E. coli* in different tissues and at different time post infections via RT-qPCR.

Material and methods

ESTs analysis and defensin identification

EST databases of *Calomera littoralis, Cephalota litorea* (Rodríguez-García *et al.,* 2015) and *Cicindela campestris* (Theodorides *et al.,* 2002) were explored to identify immune

related genes, Libraries were assembled and analyzed as we previously described (Rodríguez-García *et al.,* 2015). Blast searches revealed a defensin gene from *C. littoralis* that was selected for further analyses.

Insect

Calomera littoralis adults were collected from Laguna de Pétrola (Albacete, Spain). Tiger beetles were injected into the methathoracic coxa via the intersegmental membrane using a sterile micro-syringe. Treated group was injected with 1µl of 1mg/ml lipopolysaccharide (LPS) from *Escherichia coli* (0111:B4, Cat. No.: l2630, Sigma, Taufkirchen, Germany) diluted in insect saline solution. Control group was injected with 1µl of insect saline solution. Head, thorax, and abdomen were harvested at 1, 3, 6, 12, 24 and 36 h post-infection (p.i), submerged in RNA later (Quiagen), and stored at -20 ºC until RNA extraction.

Rapid amplification of cDNA ends (RACE)

Total RNA for the RACE reaction was extracted from an adult of *C. littoralis* using TRIzol reagent (Life Technologies) following the manufacturer's protocol and extracted RNA was treated using a TURBO DNA free Kit (Ambion, Life Technologies) to remove DNA contamination. The 5´ and 3´ RACE libraries were produced using the Clontech SMARTer RACE cDNA amplification Kit (Takara Bio) following the manufacturer's instructions. Gene-specific primers for 5' and 3' RACE were designed using the originals ESTs as reference to construct them (Table 1).

Table 1. Primers sequences used for RACE-PCR amplification (RACE) and Real time-PCR (RT-qPCR) in *Calomera littoralis*. Arginine kinase primers used as endogenous in real time are also described.

Gene	Primer sequence
Clit-Def	**RACE_F** 5'- ACTTTAGCTGTGCTGCTCCTTGCTG-3'
	RACE_R 5'- AATGCATCTTCCACCTCGGTAGCCC -3'
	RT-qPCR_F-5'- CTGCTCCTTGCTGCAATTTG -3'
	RT-qPCR_R-5'- CTGTGACGATTCGCTCTTTCTG -3'
Arg kin	**F**-5'-CTCGTGTGGTGCAACGAAGA-3'
	R-5'-GGTGGCTGAACGGGACTCT-3'

The synthesis of first strand cDNA was performed following PCR conditions: 30 cycles at 94 ºC for 30 s, 67 ºC for 30 s and 72 ºC for 3 min. PCR products were sequenced in triplicate by SAI at the University of Murcia, Spain to obtain the complete cDNA sequence.

Bioinformatic analyses

Sequences produced by RACE were edited and assembled by Geneious version 7 (http://www.geneious.com, Kearse *et al.,* 2012). The cDNA sequence was submitted to GenBank (accession number: KT897462). The open reading frames (ORF) was generated in ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The protein molecular weight (Mw) and isolectric point (pI) were predicted using Compute pI/Mw (http://web.expasy.org/compute_pi/). Homology search was performed with the BLASTP program at the NCBI. SignalP 4.1 software (Nordahl Petersen *et al.*, 2011) was used to identify a predicted signal peptide, and InterProScan (Jones *et al.*, 2014) was used to search protein domains.

Three-dimensional structure prediction

HHpred server (http://toolkit.tuebingen.mpg.de/hhpred) (Söding *et al.,* 2005) was used to find the best model templates according to the alignment and sequence identity. The three best templates were used together in Modeller (Sali *et al.*, 1995) to construct a molecular model for Clit-Def. The final model was evaluated trough ProSAII (Wiederstein & Sippl, 2007), in order to obtain a Z-score to discard errors in the three dimensional structure. Structure visualization was done in Swiss PDBviewer (Guex & Peitsch, 1997).

Antimicrobial activity prediction

Antimicrobial activity was predicted using several online servers, AntiBP2 server (http://www.imtech.res.in/raghava/antibp2/) was used to predict antibacterial activity, AVPpred (http://crdd.osdd.net/servers/avppred/) was used to predict antiviral activity, and iAMP-2l server (http://www.jci-bioinfo.cn/iAMP-2L) was used to detect the functional classes of AMPs.

Phylogenetic analysis

Homology search was performed with the BlastP program at the NCBI to obtain defensin homologues in different insect orders. Amino acid sequence were retrieved from GenBank and multiple sequence alignment was performed with MEGA v6 (Tamura *et al.,* 2013). A phylogenetic tree was constructed using the neighbor-joining method in MEGA v6 software (Tamura *et al.,* 2013). Bootstrap values were calculated on 1000 replicates.

Real- time quantitative PCR

The expression of the Clit-Def gene was investigated using real-time quantitative PCR (RT-qPCR) analysis. Gene expression levels were examined in abdomen at different times p.i. with lipopolysaccharides of *E. coli* (1, 3, 6, 12, 24 and 36 h). Gene expression levels were also examined in the head, thorax and abdomen on *C. littoralis* individuals after 12 hours p.i. Equal aliquots of total RNA from three different individuals were pooled. The RNA extraction and cDNA synthesis protocols were described above. The specific primers for RT-qPCR were designed based on gene sequences using Primer Express 3 software (Applied Biosystems®) (Table 1). Primers from arginine kinase were used as endogenous controls. Gene efficiencies were first validated with the quantification standard curve method using a serially diluted pool of cDNA. The gene expression experiment was carried out using the Comparative C_T method (Livak & Schmittgen, 2001) considering arginine kinase as endogenous gene. Control sample at 1h, and control head were used as reference samples for time and tissues experiments, respectively. No template and no reverse transcription were used as negative control. Experiments were carried out on a StepOnePlus™ Real-Time PCR System using SYBR® Green (Applied Biosystems®). PCR conditions were as follows: one cycle at 95 °C for 2 min and 40 cycles at 95 °C for 15 s and 60 °C for 30 s. Each sample had three replicates and the experiment was repeated three times with independent RNA samples to assure repeatability of the results. All data were analyzed with the program 7500 software v2.0.5 (Applied Biosystems®), performing a Student's t test to obtain the 95% confidence intervals.

Results

The full length of a defensin Clit-Def cDNA, isolated from *Calomera littoralis*, was 500 bp including poly-A tail, and contained an ORF of 246bp. The complete Clit-Def cDNA sequence and the deduced amino acid sequence are shown in Fig 1. The cDNA encoding protein precursor is composed of 81 amino acid residues, including a predicted 19 amino acid signal peptide, a 11 amino acid propeptide region and a 46 amino acid mature peptide Clit-Def. The predicted molecular weights of the complete and mature peptide defensin were 8.70 and 4.75 KDa, respectively, and the isoelectric points were 9.01 and 8.94, respectively. Blast searches showed diverse grade of identity (42-73%) with defensins from other insect species. Clit-Def shares the highest identity (73%) with the defensin of the Hemiptera *Graminella nigrifrons* (AIY24634), 54% identity with the coleoptera *Copris tripartitus* (ABP97087), and 53% with the Anoplura *Pediculus humanus* (XP002432619). Mature peptide contains six conserved cysteine residues that forms the following motif: C-X_{16}-C-X_3-C-X_{11}-C-X_4-C-X-C.

```
  1 tggggatagtaaactgtgacaagtcgaaaaagaaaaatgtacaagactactttagctgtg
  1                                       M  Y  K  T  T  L  A  V
 61 ctgctccttgctgcaatttgtgcagctcttccattagatgttactgaggatggactagca
  9  L  L  L  A  A  I  C  A  A  L  P  L  D  V  T  E  D  G  L  A
121 gaaagagcgaatcgtcacagaagagttacttgcgatttgctaagcttttccgctaaagga
 29  E  R  A  N  R  H  R  R  V  T  C  D  L  L  S  F  S  A  K  G
181 gtatcagtaaatcatgctgcttgcgctgcccattgtctggcaatgttaaagggctaccga
 49  V  S  V  N  H  A  A  C  A  A  H  C  L  A  M  L  K  G  Y  R
241 ggtggaagatgcattgatggagtttgtcactgcaggagataaacacaattatattgctca
 69  G  G  R  C  I  D  G  V  C  H  C  R  R  *
301 atttatatatattttatgtttaaaaaactttacgctttcaaatatatattcaattagtc
361 taataatatgataatattcagcattaattttaagaacaaattttgaagtagatgagattt
421 tgaattgcattatgttaaataattcttaaataaaataaaaggcaattaaaaataattaaa
481 aaaaaaaaaaaaaaaaaaa
```

Figure 1. **Nucleotide sequence of the Clit-Def cDNA gene and deduced amino acid sequence.** The putative mature peptide is shaded in grey. The asterisk indicates the stop codon.

Three-dimensional structure prediction

For molecular modeling, three templates were chosen, PBD ID: ICA, 2LN4 and 2E3G. These templates share a 67%, 63% and 56% of identity to Clit-Def, respectively. The best model obtained in Modeller has a Z-score on ProSA of -3.72, indicating a valid molecular level. The model was composed of an α-helix (residues Ala^{19}- Ala^{26}) and two anti-parallel β-sheet (residues Gly^{33}-Ile^{37} and Val^{40}-$Arg^{44)}$, that are stabilized by three disulfide bonds with the pattern Cys^{I}-Cys^{IV}, Cys^{II}-Cys^{V} and Cys^{III}-Cys^{VI} (Fig 2) which conforms a typical CSαβ structural fold.

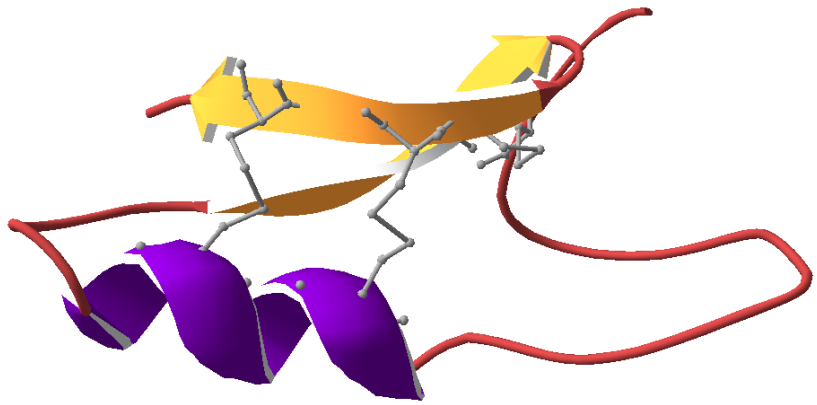


Figure 2. **Predicted structure model of the Clit-Def protein.** α-helix is represented in purple, antiparallel β-sheets are showed in orange, and coils in red. The disulfide bonds are displayed as grey balls and sticks.

Antimicrobial activity prediction

Antimicrobial predicted tools were used to complement the results obtained from structural features. Clit-Def was classified as AMP via AntiBP2 server, and putative antibacterial and antifungal activity was predicted via AVPpred server. No activity against virus was predicted using iAMP-2l server.

Phylogenetic analysis

A phylogenetic tree was generated from 26 defensin amino acid sequences retrieved from GenBank data base from different orders of insect (6 Coleoptera, 5 Lepidoptera, 5 Hymenoptera, 5 Diptera, 4 Hemiptera and 1 Phthiraptera). As showed in Fig 3, *C. littoralis* defensin was grouped together with the Hemiptera

Graminella nifrifroms defensin (bootstrap 53%). The phylogenetic tree shows evolutionary diversity among coleopteran defensin that forms diverse clades.

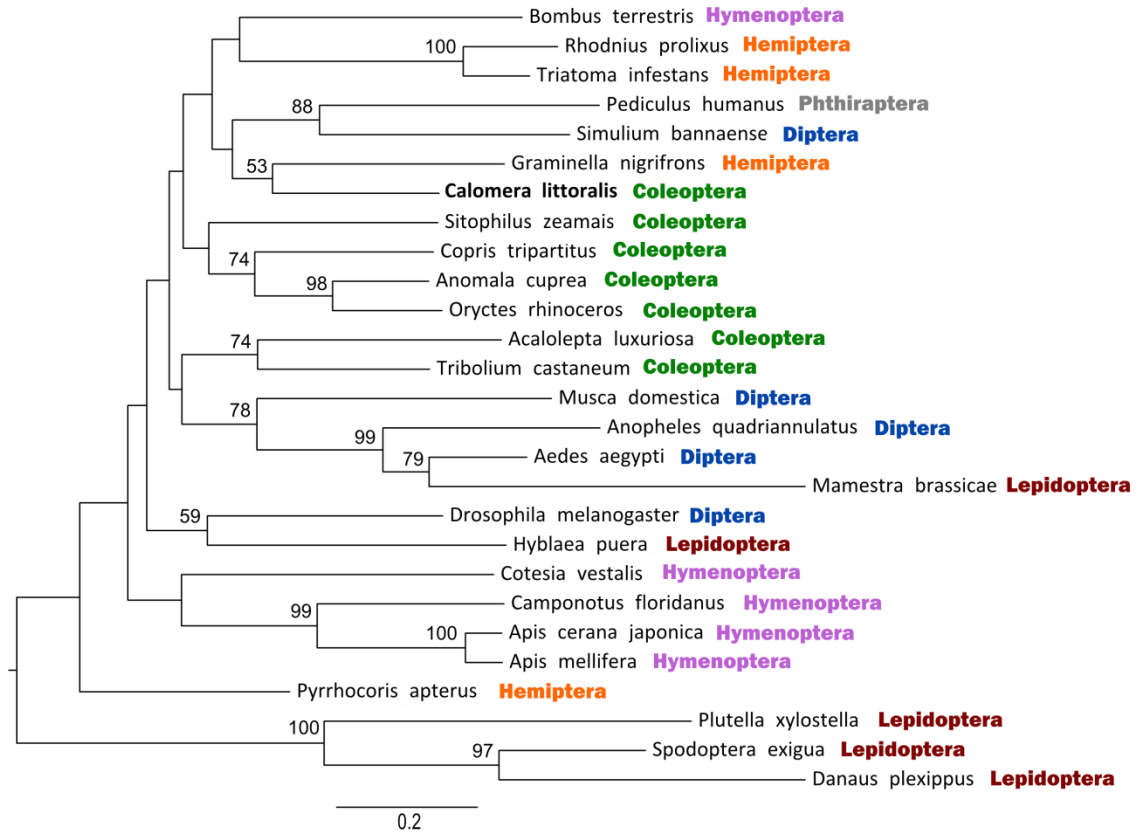


Figure 3. Neighbour-joining phylogenetic tree of defensins from different insect orders. The numbers in the nodes indicate the percentage of bootstrap support. The bar indicates the genetic distance. GenBank ID are as follows: *Sithophilus zeamais* (ABZ80665), *Acalolepta luxuriosa* (AAK35160, *Anomala cuprea* (BAD77966), *Oryctes rhinoceros* (BAA36401), *Tribolium castaneum* (XP968237), *Copris tripartitus* (ABP97087), *Graminella nigrifrons* (AIY24634), *Pediculus humanus* (XP002432619), *Cotesia vestalis* (AGI44428), *Apis cerana* (BAI81896), *Camponotus floridanus* (XP011259830), *Bombus terrestris*(ADB29129), *Rhodius prolixus* (AAO74625), *Phyrrhocoris apterus* (AGI17576) *Apis mellifera* (AGM19140), *Triatoma infestans* (ABD61004), *Simulium bannaense* (AJP36711), *Drosophila melanogaster* (NP523672), *Anopheles quadriannulatus* (ABB00999), *Aedes aegypti* (AAB46807), *Musca domestica* (AIL2468), *Plutella xylosesta* (AIW49877), *Spodoptera exygua* (AEW24427), *Mamestra brassicae* (AAL69980) *Danaus plexippus* (EHJ63539), *Hyblaea purea* (ABQ08056).

Multi-sequence alignment of coleopteran defensins (Fig 4) shows that are divergent, mainly overall in signal peptide and propeptide regions. However, fifteen amino acid residues within the mature peptide are conserved, including the six conserved cysteines and nine additional residues. In addition there are two residues (-RR or -KR) at the C-terminus of propeptide, except for *Sitophilus zeamais*.

	Signal peptide	Propeptide	Mature peptide
Calomera littoralis	MYKTTLAVLLLAAICAALP-	LDVTEDGLAERANRHRR---	VTCDLLSFSAKG----VSVNHAACAAHCLAMLKGYRGGRCIDG-VCHCRR
Sitophilus zeamais	MVKVVLFVFLIVLAVGAYC-	APVDEEFQDDLIEGPVRVR-	ATCDLLSFEIKG----FKLNDSACAAHCIQLGK--RGGHCNNSKVCVCRR
Acalolepta luxuriosa	MKFFITFTFVLSLVVLTVYS	APREFAEPEEQDEGHFRVKR	FTCDVLSVEAKG----VKLNHAACGIHCLFRRR--TGGYCNKKRVCICR-
Anomala cuprea	MSKSFLITLVVAMCIVHTLA	APTPEEFEGSIVRQKR----	VTCDLLSFEAKG----FAANHSICAAHCLAIGR--KGGSCQNG-VCVCRN
Oryctes rhinoceros	MSRFIVFAFIVAMCIAHSLA	APAPEALEASVIRQKR----	LTCDLLSFEAKG----FAANHSLCAAHCLAIGR--KGGACQNG-VCVCRR
Tribolium castaneum	MKLLIVALVALFCIFETTA-	FPTDGEHIRVKR--------	FTCDVLSAEGSFRGVSVKLNHSACATHCLFLKK--RGGYCNNKAICVCRN
Copris tripartitus	MAKLIAFALVASLCLSMVLC	NPLPEEVQEEGLVRQKR---	VTCDVLSFEAKG----IAVNHSACALHCIALRK--KGGSCQNG-VCVCRN
			*** ** * * ** ** * * * **

Figure 4. Amino acid sequence of coleopteran defensins. The conserved cysteine residues are shaded. The asterisk indicates equal positions.

Expression patterns of Clit-Def after infection with lipopolysaccharides of *Escherichia coli*

The Clit-Def transcript profiles from control and bacterial-challenged *C. littoralis* were compared at 1, 3, 6, 12, 24 and 36 h p.i. using RT-qPCR (Fig 5) The Clit-Def gene was up-regulated at all times of exposure in treated samples, increasing it response after 3 hours and reaching the highest expression level at 12 hours of treatment. Control samples showed a small increase in expression whose highest values correspond with those obtained for treated samples.

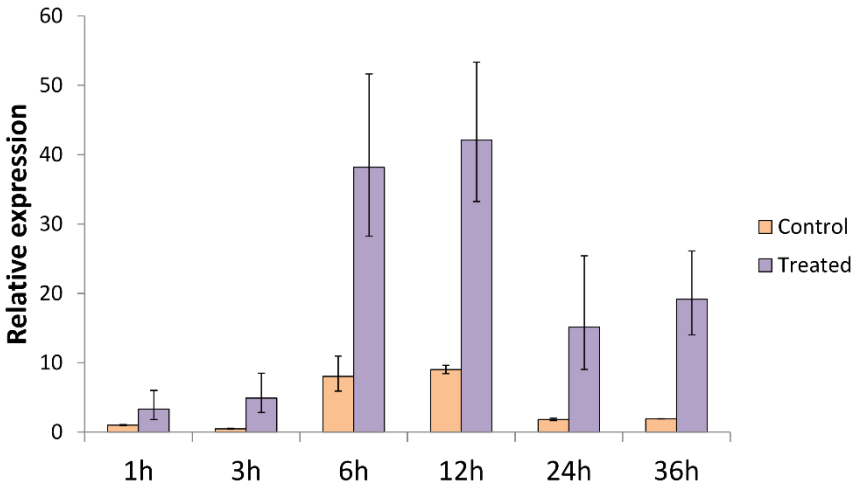


Figure 5. RT-qPCR of the Clit-Def from RNA isolated from the abdomen of control and treated (bacteria-challenged) *Calomera littoralis* at 1, 3, 6, 12, 24, and 36 h p.i.

Additionally, RT-qPCR was performed to examine the expression patterns of the Clit-Def gene in head, thorax and abdomen in control and treated *C. littoralis* beetles (Fig 6). After 12 h of exposure, the highest expression was found in treated

abdomen while Clit-Def gene expression is lower all over the body in control samples.

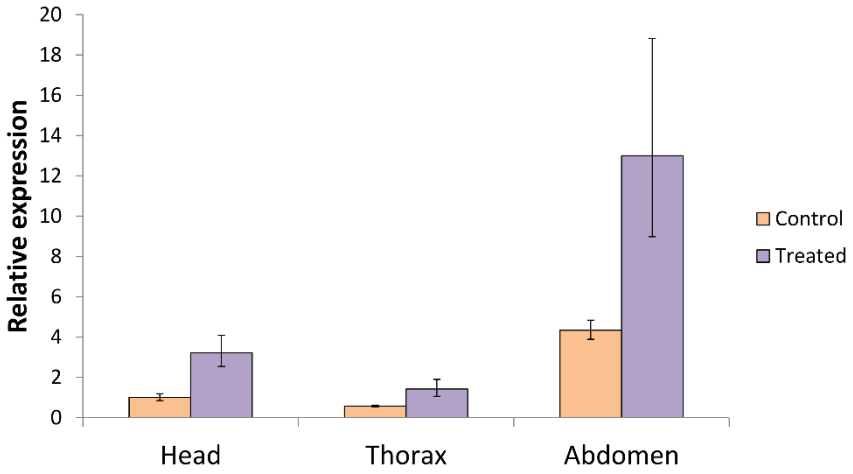


Figure 6. Relative expression of Clit-Def normalized to Arginine kinase endogenous in the head, thorax and abdomen of *Calomera littoralis*.

Discussion

The Clit-Def mature peptide of the tiger beetle *Calomera littoralis* contains the following motif: C-X_{16}-C-X_3-C-X_{11}-C-X_4-C-X-C, which is consistent with the consensus motif of insect defensins C-$X_{5\text{-}16}$-C-X_3-C-$X_{9\text{-}11}$-C-$X_{4\text{-}7}$-C-X_1-C (Čeřovský & Bém, 2014). The cDNA sequence presents the conserved enzymatic processing site (-RR) to release the mature peptide. –RK or –RR dibasic motif have been identified in different insect orders (Bulet & Stöcklin, 2005; Yi *et al.,* 2014). There are several features that are important in the function of defensin; i) the disulfide bonds that conforms the conserved six cysteines (Ganz & Lehrer, 1994; Wanniarachchi *et al.,* 2011); ii) the basic molecular charge, due to the positively charged defensin interacting with the negatively charged microbial membrane components (Sagaram *et al.,* 2011); and iii) the three dimensional structure (Bulet *et al.,* 2004). The results obtained in *C. littoralis* are in agreement with these features, as the Clit-Def motif have six cysteine residues that forms three disulphide bonds, which plays an important role in the folding and stability of proteins (Sevier & Kaiser, 2002). The predicted isoelectric point showed that the Clit-Def is a cationic protein. Homology

modeling of Clit-Def shows one α-helix and two antiparallel β-sheets. The α-helix and β-sheet regions are considered to be the active sites of defensins in general (Tsuji *et al.,* 2007) and contribute to the antimicrobial action. In addition the bioinformatic approaches predicted putative antibacterial and antifungal activity, but no antiviral activity.

Phylogenetic analysis showed that Clit-Def is more closely related to an Hemipteran defensin than to that of other coleopteran taxa. Furthermore, the coleopteran defensins used in the phylogenetic analysis do not form a unique clade, they are interspersed among sequences of representatives of other insect orders. Recent studies suggest that the cysteine-stabilized α-helix β-sheet (CSαβ) motif of insect defensins evolved by gene duplications, followed by divergence due to selective evolutionary pressure, to produce a diverse set of paralogues (Dassanayake *et al.,* 2007). Additionally, other studies focused on the invertebrate defensin evolution did not achieve a conclusion in the evolutionary history of AMPs due to the incomplete data set analyzed (Rodriguez de la Vega & Possani, 2005). The presence of divergent coleopteran defensins suggests a possible diversity in the structure and function of their immune system. The characterization of defensins in other cicindelid species, could help to elucidate the evolutionary dynamics of these genes in Coleoptera.

RT-qPCR was performed to examine the expression patterns of the Clit-Def gene in different tissues (head, thorax and abdomen), in control and treated with LPS of *E coli* in *C. littoralis* specimens. After 12 h of exposure, Clit-Def gene expression was significantly lower throughout the body in control samples, and the highest expression was observed in abdomen after infection. These results are consistent with earlier findings that showed that AMPs are primarily distributed in the fat body, midgut or haemolymph (Wang *et al.,* 2006), tissues that are found in the abdomen. Previous studies have also suggested that other tissues as the salivary glands are involved in the immune response (Wei *et al.,* 2015), what could explain the increase of expression levels in head after treatment. In addition, the expression profile through time was analyzed. The transcript level of the Clit-Def gene was up-regulated after infection with the lipopolysaccharides of *E. coli*, peaked after 12 h p.i. This pattern of increase and decline expression over time was reported in other

insect defensins, with variation in the time of maximum expression: the expression of BmDefensinB in fat body tissue of *B. mori* peaks after 8 h (Kaneko *et al.,* 2008), *Dermacentor variabilis* peaks at 48 h p.i in midgut and fat body (Ceraul *et al.,* 2007), and in *Spodoptera littoralis* the highest expression of defensins in haemolymph was at 48 h p.i. (Seufi *et al.,* 2011). The small increase of expression over time in control samples could be due to the activation of the immune system after the injure resulting from the ininjection of the saline solution.

In conclusion, bioinformatic analysis suggests that the Clit-Def is a defensin gene and meets all the structural criteria to present antimicrobial activity. Phylogenetic analysis showed a high diversity in coleopteran defensins. The Clit-Def was up-regulated in *E-coli* infected individuals which also suggest that Clit-Def is involved in humoral response. However, further works are required to determine how this defensin acts after Gram (+) and Gram (-) bacterial and fungal infections.

Acknowledgements

This work has been funded by project CGL2008-03628 of the Ministerio de Ciencia e Innovación, Spain.

References

Ali, H. (1978) Faunistic study of the Cicindelidae (Coleoptera) of Iraq and southwest Asia. *Coleop Bull* **32**: 1-20.

Anselme, C., Pérez-Brocal, V., Vallier, A., Vincent-Monegat, C., Charif, D., Latorre, A. *et al.* (2008) Identification of the weevil immune genes and their expression in the bacteriome tissue. *BMC Biol* **6**: 43.

Bulet, P. and Stöcklin, R. (2005) Insect antimicrobial peptides: structures, properties and gene regulation. *Protein Expr Purif* **12**: 3-11.

Bulet, P., Hetru, C., Dimarcq, J.L. and Hoffmann, D. (1999) Antimicrobial peptides in insects; structure and function. *Dev Comp Immunol* **23**: 329-344.

Bulet, P., Stöcklin, R. and Menin, L. (2004) Anti-microbial peptides: from invertebrates to vertebrates. *Immunol Rev* **198**: 169-184.

Bulmer, M.S. and Crozier, R.H. (2004) Duplication and diversifying selection among termite antifungal peptides. *Mol Biol Evol* **21**: 2256-2264.

Cassola, F. and Pearson, D.L. (2000) Global patterns of tiger beetle species richness (Coleoptera: Cicindelidae): their use in conservation planning. *Biol Conserv* **95**: 197-208.

Ceraul, S., Dreher-Lesnick, S., Gillespie, J., Rahman, M. and Azad, A. (2007) New tick defensin isoform and antimicrobial gene expression in response to *Rickettsia montanensis* challenge. *Infec Immun* **75**: 1973-1983.

Čeřovský, V. and Bém, R. (2014) Lucifensins, the insect defensins of biomedical importance: the story behind maggot therapy. *Pharmaceuticals* **7**: 251-264.

Dassanayake, R.S., Gunawardene, Y.S. and Tobe, S.S. (2007) Evolutionary selective trends of insect/mosquito antimicrobial defensin peptides containing cysteine-stabilized α/β motifs. *Peptides* **28**: 62-75.

Ganz, T. (2003) Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol* **3**: 710-720.

Ganz, T. and Lehrer, R.I. (1994) Defensins. *Curr Opin Immunol* **6**: 584-589.

Gao, B. and Zhu, S. (2012) Alteration of the mode of antibacterial action of a defensin by the amino-terminal loop substitution. *Biochem Biophys Res Commun* **426**: 630-635.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18**: 2714-2723.

Hazlett, L. and Wu, M. (2011) Defensins in innate immunity. *Cell Tissue Res* **343**: 175-188.

Hoffmann, J.A. (1997) Immune responsiveness in vector insects. *P* **94**: 11152-11153.

Hwang, J. S., Lee, J., Kim, Y. J., Bang, H. S., Yun, E. Y., Kim, S. R. *et al.* (2009) Isolation and characterization of a defensin-like peptide (coprisin) from the dung beetle, *Copris tripartitus*. *Int J Pep* **2009**.

Ishibashi, J., Saido-Sakanaka, H., Yang, J., Sagisaka, A. and Yamakawa, M. (1999) Purification, cDNA cloning and modification of a defensin from the coconut rhinoceros beetle, *Oryctes rhinoceros*. *Eur J Biochem* **266**: 616-623.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Kaneko, Y., Tanaka, H., Ishbashi, J., Iwasaki, T. and Yamakawa, M. (2008) Gene expression of a novel defensin antimicrobial peptide in the silkworm, *Bombyx mori*. *Biosci Biotechnol Biochem* **72**: 2353-2361.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.

Lambert, J., Keppi, E., Dimarcq, J. L., Wicker, C., Reichhart, J. M., Dunbar, B. *et al.* (1989) Insect immunity: isolation from immune blood of the dipteran *Phormia terranovae* of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides. *P Natl Acad USA* **86**: 262-266.

Landon, C., Barbault, F., Legrain, M., Guenneugues, M. and Vovelle, F. (2008) Rational design of peptides active against the Gram positive bacteria *Staphylococcus aureus*. *Proteins* **72**: 229-239.

Lazzaro, B.P. and Little, T.J. (2009) Immunity in a variable world. *Philos Trans R Soc Lond B Biol Sci* **364**: 15-26.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2<- ΔΔCT< Method. *Methods* **25**: 402-408.

Matsuyama, K. and Natori, S. (1988) Molecular cloning of cDNA for sapecin and unique expression of the sapecin gene during the development of *Sarcophaga peregrina*. *J Biol Chem* **263**: 17117-17121.

Nordahl Petersen, T., Brunak, S., vonHeijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785-786.

Pasupuleti, M., Schmidtchen, A. and Malmsten, M. (2012) Antimicrobial peptides: key components of the innate immune system. *Crit Rev Biotechnol* **32**: 143-171.

Pearson, D. L. (1984) The tiger beetles (Coleoptera: Cicindelidae) of the Tambopata Reserved Zone, Madre de Dios, Peru. *Rev Peru Entomol* **27**: 15-24.

Pearson, D. L. (1988) Biology of tiger beetles. *Ann Rev Entomol* **33**: 123-147.

Pearson, D.L. (1992) Tiger beetles as indicators for biodiversity patterns in Amazonia. *Res Explor* **8**: 116-117.

Pearson, D. L. and Cassola, F. (2005) A quantitative analysis of species descriptions of tiger beetles (Coleoptera: Cicindelidae), from 1758 to 2004, and notes about related developments in biodiversity studies. *Coleops Bull* **59**: 184-193.

Pearson, D.L. and Ghorpade, K. (1989) Geographical and ecological history of tiger beetles (Coleoptera: Cicindelidae) of the Indian subcontinent. *J Biogeogr* **16**: 333-344.

Pearson, D.L., Blum, M.S., Jones, T.H., Fales, H.M., Gonda, E. and White, B.R. (1988) Historical perspective and the interpretation of ecological patterns: defensive compounds of tiger beetles (Coleoptera: Cicindelidae). *Am Nat* **132**: 404-416.

Rajamuthiah, R., Jayamani, E., Conery, A. L., Fuchs, B. B., Kim, W., Johnston, T. *et al.* (2015) A Defensin from the Model Beetle *Tribolium castaneum* Acts Synergistically with Telavancin and Daptomycin against Multidrug Resistant *Staphylococcus aureus*. *PloS One* **10**: e0128576.

Rodriguez de la Vega, R.C. and Possani, L.D. (2005) On the evolution of invertrebate defensins. *Trends Genet* **21**: 330-332.

Rodríguez-García, M.J., Machado, V. and Galián, J. (2015) Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles. *BMC Genomics* **16:** 391.

Sagaram, U.S., Pandurangi, R., Kaur, J., Smith, T.J. and Shah, D.M. (2011) Structure–activity determinants in antifungal plant defensins MsDef1 and MtDef4 with different modes of action against *Fusarium graminearum*. *PLoS One* **6**: e18550.

Sagisaka, A., Ishibashi, J., Miyanoshita, A., Saido-Sakanaka, H. and Yamakawa, M. (1999) cDNA cloning and gene expression of defensin, an antibacterial peptide from a beetle, *Allomyrina dichotoma* (Coleoptera: Scarabaeidae). *Appl Entomol Zool* **34**: 421-427.

Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. and Karplus, M. (1995) Evaluation of comparative protein modelling by MODELLER. *Proteins* **23**: 318-326.

Schneider, T., Kruse, T., Wimmer, R., Wiedemann, I., Sass, V., Pag, U. *et al.* (2010) Plectasin, a fungal defensin, targets the bacterial cell wall precursor Lipid II. *Science* **328**: 1168-1172.

Seufi, A. M., Hafez, E. E. and Galal, F. H. (2011) Identification, phylogenetic analysis and expression profile of an anionic insect defensin gene, with antibacterial activity, from bacterial-challenged cotton leafworm, Spodoptera littoralis. *BMC Mol Biol* **12**: 47.

Sevier, C.S. and Kaiser, C.A. (2002) Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Biol* **3**: 836-847.

Söding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**: W244-W248.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729.

Theodorides, K., De Riva, A., Gómez-Zurita, J., Foster, P.G. and Vogler, A.P. (2002) Comparison of EST libraries from seven beetle species: towards a framework for phylogenomics of the Coleoptera. *Insect Mol Biol* **11**: 467-47.

Tsuji, N., Battsetseg, B., Boldbaatar, D., Miyoshi, T., Xuan, X., Oliver Jr, J.H. *et al.* (2007) Babesial vector tick defensin against Babesia sp. parasites. *Infect Immun* **75**: 3633-3640.

Ueda, K., Imamura, M., Saito, A. and Sato, R. (2005) Purification and cDNA cloning of an insect defensin from larvae of the longicorn beetle, *Acalolepta luxuriosa*. *Appl Entomol Zool* **40**: 335-345.

Vilcinskas, A. (2013) Evolutionary plasticity of insect immunity. *J Insect Physiol* **59**: 123-129.

Viljakainen, L. (2015) Evolutionary genetics of insect innate immunity. *Brief Funct Genomics* elv002.

Viljakainen, L. and Pamilo, P. (2008) Selection on an antimicrobial peptide defensin in ants. J Mol Evol **67**: 643-652.

Wang, J. X., Zhao, X. F., Liang, Y. L., Li, L., Zhang, W., Ren, Q. *et al.* (2006) Molecular characterization and expression of the antimicrobial peptide defensin from the housefly (*Musca domestica*). *Cell Mol Life Sci* **63**: 3072-3082.

Wanniarachchi, Y.A., Kaczmarek, P., Wan, A. and Nolan, E.M. (2011) Human defensin 5 disul-fide array mutants: disulfide bond detection attenuates antibacterial activity against *Staphylococcus aureus*. *Biochemistry* **50**: 8005-8017.

Wei, L., Mu, L., Wang, Y., Bian, H., Li, J., Lu, Y. *et al.* (2015) Purification and characterization of a novel defensin from the salivary glands of the black fly, *Simulium bannaense*. Parasites & vectors **8**: 1-11.

Wiederstein, M. and Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35:** W407-W410.

Willis, H. (1972) Species density of North American Cicindela. *Cicindela* **4**: 29-34.

Yi, H.Y., Chowdhury, M., Huang, Y.D. and Yu, X.Q. (2014) Insect antimicrobial peptides and their applications. *Appl Microbiol Biotechnol* **98:** 5807-5822.

Zhu, S., Gao, B., Harvey, P.J. and Craik, D.J. (2012) Dermatophytic defensin with antiinfective potential. *P Natl Acad USA* **109**: 8495- 8500.

General comments and conclusions

In this thesis reproduction and the immune related genes have been identified and characterized for the first time in cicindelids, which are one of the most studied groups of beetles, although the information on their transcriptome is limited. From the analysis of EST libraries of different tiger beetles (*Calomera littoralis, Cephalota litorea and Cicindela campestris*) four seminal fluid proteins (Acp01-Acp04), the SUMO conjugating enzyme E2 (Cc-UBC9) and one ubiquitin-ribosomal fusion protein (Cc-UBS27), have been identified and characterized, all of them involved in the reproductive process. A defensin (Clit-Def) involved in the immune system has also been characterized. The results obtained from the characterization of its sequence and expression pattern have been analysed from an evolutionary perspective.

The description of these genes provides important information on tiger beetles biology, and also provides a foundation to better understand their diversification process, leading to a better comprehension of their phylogenetic relationships and providing clues on speciation processes.

Conclusions

Chapter1. Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles

1. Of the 14 proteins selected as candidates, four of them met the criteria of seminal fluid proteins, they are secreted and show male-biased expression.

2. Homology searches with other species to infer function in the four seminal fluid proteins characterised, did not produce results likely because there are not homologous proteins characterised or homology is obtained with proteins whose function has not been annotated.

3. Bioinformatic predictions of function from sequence and structure of proteins, confirm that two proteins AcpC01 (immune system) and AcpC03 (hormone) are seminal proteins as they have functions already described for those proteins.

4. AcpC03 only amplified in *C. littoralis* and *Lophyra flexuosa,* species that are phylogeneticaly closed-related and thus, meets the requiriments to be considered a rapidly evolving protein.

5. AcpC04 amplified in all tiger beetle species analysed, so it most likely is a conserved protein at least in tiger beetles.

6. AcpC01 amplified in all cicindelid species analysed, except for *Cephalota maura*, which is the most phylogenetically distant species analysed, and therefore it most likely is a conserved protein in at least one group of tiger beetles.

Chapter2. Characterisation and expression analysis of UBC9 and UBS27 genes in developing gonads of cicindelids (Coleoptera: Cicindelidae)

7. After EST annotation, and further characterization of the complete sequence it is concluded that the gene described in *C. campestris* as Cc-UBC9 encoded a conjugating enzyme E2 SUMO and the gene described as Cc-UBS27 encoded a ubiquitin-ribosomal fusion protein.

8. The amino acid sequence and the phylogenetic analysis show that both genes are highly conserved in cicindelids and other beetles.

9. The selection analysis revealed that both genes are under purifying selection what highlights the importance of these proteins.

10. Both genes show higher levels of expression in gonads compared to the other tissues tested, indicating that they play an active role in gametogenesis in cicindelids.

11. The higher expression of these genes in immature testis relative to the mature stage, indicates that these genes are mostly involved in the early stages of gametogenesis.

Chapter 3. Identification, structural characterisation and expression analysis of a defensin gene from the tiger beetle *Calomera littoralis* (Coleoptera: Cicindelidae)

12. The gene characterised as Clit-Def in *Calomera littoralis* meets the structural features of a defensin.

13. The structural and bioinformatic analyses suggest that Clit-Def has antimicrobial activity and therefore is involved in the immune system.

14. Clit-Def is overexpressed in *C. littoralis* after infection with lipopolysaccharides *of Escherichia coli* in different tissues and mainly in the abdomen, suggesting that this gene plays a role in the humoral response.

15. The phylogenetic analysis using the defensin Clit-Def together with other coleopteran sequences obtained from GenBank, do not show clades congruent with the taxonomy indicating that they are highly diverse proteins.

Anexes

Rodríguez-García et al. BMC Genomics (2015) 16:391
DOI 10.1186/s12864-015-1619-9

BMC Genomics

RESEARCH ARTICLE Open Access

Identification and characterisation of putative seminal fluid proteins from male reproductive tissue EST libraries in tiger beetles

María Juliana Rodríguez-García*, Vilmar Machado and José Galián

Abstract

Background: The study of proteins transferred through semen can provide important information for biological questions such as adaptive evolution, the origin of new species and species richness. The objective of this study was to identify seminal fluid proteins (SFPs) that may contribute to the study of the reproductive system of tiger beetles (cicindelids), a group of more than 2,500 species distributed worldwide that occupy a great diversity of habitats.

Results: Two cDNA libraries were constructed from the male gonads of *Calomera littoralis* and *Cephalota litorea*. Expressed sequence tags (ESTs) were analysed by bioinformatics approaches and 14 unigenes were selected as candidate SFPs, which were submitted to Reverse Transcription Polymerase Chain Reaction (RT-PCR) to identify patterns of tissue-specific expression. We have identified four novel putative SFPs of cicindelids, of which similarity searches did not show homologues with known function. However, two of the protein classes (immune response and hormone) predicted by Protfun are similar to SFPs reported in other insects. Searches for homology in other cicindelids showed one lineage specific SFPs (rapidly evolving proteins), only present in the closely related species *C. littoralis* and *Lophyra flexuosa* and two conserved SFP present in other tiger beetles species tested.

Conclusions: This work represents the first characterisation of putative SFPs in Adephagan species of the order Coleoptera. The results will serve as a foundation for further studies aimed to understand gene (and protein) functions and their evolutionary implications in this group of ecologically relevant beetles.

Keywords: Seminal fluid proteins, Adephagan, Tiger beetles, Expressed sequences tags, Gene expression

Background

Tiger beetles or cicindelids belong to the Adephaga, a suborder of Coleoptera that includes conspicuous, brightly coloured, non-pest species that are significant components of ecosystems, being important links in food chains. Cicindelids are organisms that are commonly used as bioindicators, as their presence or absence can provide information on the quality, alterations and successional stage of habitats, and are considered to be good bioindicators of general biodiversity [1–5]. There have been more than 2,500 species of cicindelid beetles described [6], which are distributed worldwide (except for Tasmania, Antarctica, and some Oceanic islands) and occupy a great diversity of habitats (alpine meadows, dessert grassland, among others) [7, 8]. Nevertheless, they do not show a homogenous pattern of distribution with relation to latitude or biogeography [6] and species tend to specialise in particular habitats [1]. They have several mechanisms or adaptations to reduce competition and contribute to maintain species richness; seasonal temporal segregation (life cycles), spatial segregation (differences in habitat or microhabitats preferences) [9] and temporal partitioning on the diel scale (diel activity patterns) have been previously studied as the most common strategies.

Tiger beetles have been deeply studied at different levels: (i) morphology [10], (ii) taxonomy [11–13], (iii) biology [7, 14], (iv) physiology [15, 16], (v) thermoregulation [17], (vi) evolution, ecology and diversity [18], (vii) chromosome evolution [19] and (viii) conservation strategies [2]. Nevertheless, the reproductive biology has not

* Correspondence: mjulia.rodriguez@um.es
Department of Zoology and Physical Anthropology, Faculty of Veterinary, University of Murcia, Campus Mare Nostrum, E-30100 Murcia, Spain

been studied, on physiological or molecular grounds, and little information on the transcriptome and gene expression related to physiological processes is available. The transcriptome analysis is an important tool to help identifying putative function of genes, translating the sequence of nucleotides in a sequence of amino acids, which is more likely to be conserved [20, 21]. In addition, the availability of a variety of bioinformatics tools allows the characterisation of these genes [22].

Some studies show that proteins with high expression levels in male reproductive tissues and with characteristics that meet the criteria of extracellular secretion, are good candidates to be considered as seminal fluid proteins (SFPs) [23–27]. The seminal fluid of insects contains sperm and a complex mixture of proteins, inorganic solutes, carbohydrates and lipids that are transferred to females during mating via the spermatophore. These proteins, which are produced in male gonads (testes, vas deferens and accessory glands), are important in the reproduction process by inducing physiological and behavioural changes in females, reducing responsiveness to other males, increasing the ovulation and egg laying rates, and altering feeding activity and also immune response [28–30].

The study of proteins transferred through the semen provides information for important biological questions such as the origin of new species and the origin of new molecules involved in sperm competition and coevolution between males and females [29, 31–36]. Seminal fluid proteins have two characteristics that according to several theoretical models might lead to speciation. i) These proteins are related with sexual selection and sexual conflict and ii) the rapid evolutionary rate of these proteins may also contribute to the evolution of reproductive barriers between populations. Furthermore, there is experimental evidences indicating a correlation between features that undergo the action of sexual selection and the speciation process [37–39].

In addition, several studies have demonstrated that many of these SFPs have similar characteristics to those found in the taxonomically restricted genes (TRGs), such as high evolutionary rate and low similarity between closely related species [25, 29, 40, 41]. According to Avila *et al*., the analysis of SFPs provides insight into the evolutionary patterns of reproductive traits [30]. Therefore, a better understanding of cicindelid reproductive molecules and their actions provides opportunities to reveal functionally conserved mechanism in cicindelids reproduction (highly conserved SFPs), as well as mechanisms involved in the reproductive isolation between species (lineage-specific SFPs) as a subset of seminal proteins is among the most rapidly evolving proteins [42–45].

To date, several SFPs have been described in insects, such as flies and mosquitoes (Diptera), field crickets (Orthoptera), honeybee (Hymenoptera), moths and butterflies (Lepidoptera) and in the beetle genus *Tribolium* [25, 26, 45–50]. Apart from the Polyphagan genus *Tribolium*, which is considered a model organism, no other species of beetles have been analysed for these proteins.

The aim of this study was to identify and characterise genes encoding proteins that are transferred to females during mating through seminal fluids in Cicindelids to contribute to the knowledge of the nature and function of insect SFPs and particularly in this ecologically important group of Adephagan beetles. To identify these proteins, EST libraries from gonads and accessory glands of male *Calomera littoralis* and *Cephalota litorea* were made and bioinformatically analysed. Proteins selected as candidate SFPs were submitted to gene expression analysis in female and male tissues. The complete sequence of genes showing differential expression patterns was obtained and the prediction of the function was inferred either by comparing with other insects or based on the sequence properties. Searches for homology of these putative SFP were performed in other cicidelid species. This work represents the first characterisation of putative SFP in Adephagan beetles.

Results and discussion

Library construction and EST assembly

Two separate cDNA libraries were constructed from RNA extracted from reproductive tissues (testes and accessory glands) of two males *C. littoralis* and one male of *C. litorea*. A total of 1,144 EST sequences were generated; 568 clones were sequenced from the *C. littoralis* library and 576 clones were sequenced from *C. litorea*. These sequences were trimmed for the removal of vector sequences and sequences <100 bp were excluded by Seqman (DNAstar, Inc. Madison, WI). EST sequences with high quality were previously deposited in GenBank under the following accession numbers: *C. litorea* (CV156657: CV157115) and *C. littoralis* (CV157116: CV157483). The high quality sequences of *C. littoralis* were assembled to 101 contigs (two or more sequences) and 84 singletons (single sequence), and *C. litorea* ESTs were assembled to 154 contigs and 58 singletons. Each cDNA library had a minimum average inset size ranging from 114 to 1,245 bp in *C. littoralis*, and from 139 to 1,246 in *C. litorea* library. The maximum number of ESTs that formed each contig was 15 ESTs in *C. littoralis* and 18 ESTs in *C. litorea* (Table 1).

Annotation – gene ontology

Blast2GO software showed that 82 contigs of *C. litorea* and 72 contigs of *C. littoralis* had no blast hits against the non-redundant protein database at National Center for Biotechnology Information (NCBI). The annotation of the 64 ESTs of *C. litorea* and 75 ESTs from *C. littoralis* were designated by database search algorithms BlastX

Table 1 Summary of EST analyses from *C. littoralis* and *C. litorea* male gonad cDNA libraries

	C. littoralis	*C. litorea*
Number of ESTs generated	568	576
Number of High quality ESTs	368	459
Number of contigs	101	154
Number of singletons	84	58
Average length of contigs	523 bp	628 bp
Number of EST range in the contig	2-15	2-18

for proteins in the NCBI web server (Table 2). Additionally, gene ontology annotations of all contigs were performed using Blast2GO.

Annotated sequences were classified according to their gene ontology (GO) into three categories: biological process, molecular function and cellular component (Fig. 1, 2 and 3). The proportions of genes associated with the different categories were highly similar among the two libraries; it is important to note that a sequence could be included in different categories and be associated to multiple GO. Within the category "Cellular Component", the subcategories "cellular" and "organelle" were the most abundant in both libraries. However, the "extracellular" subcategory, where putative SFPs should be included, was not present in the analysis performed, and no sequences were annotated as SFPs. This could be due to both the low number of ESTs obtained in the cicindelid libraries and to the small number of Coleopteran libraries available for comparisons.

Identification of putative SFPs

Nine genes putatively encoding SFPs were identified in *C. littoralis* and five in *C. litorea* by indirect approaches (Table 3). Candidate genes were selected after detection of one or more of the following characteristics i) the presence of a computationally predicted signal peptide inferred via Signal P 4.1 software [51, 52], ii) its localisation as extracellular and/or with plasma membrane destination inferred via ProtComp , iii) the recognition of a membrane helix inferred via TMHMM (InterproScan).

Of the 14 candidates, it was only possible to design useful RT-PCR primers for 12 (*C. littoralis*: eight and *C. litorea*: four). Tissue-specific expression patterns were obtained in the 12 available candidates. RT-PCR revealed strong amplification from the male abdomen but weak or not expression in female abdomens and male thoraxes in four of the genes (*C. littoralis*: AcpC01, AcpC02 and AcpC03 and *C. litorea*: AcpC04). The positive control gene (arginine kinase) amplified in all tissues (Fig. 4). The identification of an extracellular component and the tissue-specific patterns of expression found in these four genes suggest that they encode seminal fluid proteins. The other eight candidate genes with an identified extracellular component did not show any differential expression; some studies have demonstrated that not all SFPs have robust expression in male reproductive glands [29, 40], which could explain the low number of putative SFPs characterised in the present survey.

Table 2 Summary of unigenes from *C. littoralis* and *C. litorea* male gonad analyses and annotated by Blast2go software

	C. littoralis	*C. litorea*
Annotated contigs	75	64
No annotated contigs	19	38
No mapping contigs	19	28
No blast hits contigs	72	82
No blast contigs	0	0
Total	185	212

SFP characterisation

Full length sequences and complete Open Reading Frames (ORFs) were obtained via Rapid amplification of cDNA ends PCR (RACE-PCR) for the four putative SFPs. Sizes ranged from 66 to 218 amino acids, which is in agreement with ACPs characterised in *Drosophila* [53] and references therein. Sequences are available in GenBank (KP164546–KP164549). Two of these proteins (AcpC01 and AcpC02) did not show any significant BlastX similarity (E-value > 10^{-4}) against GenBank proteins, and appear to be novel proteins (Table 4). However, the protein AcpC01 yielded a similarity close to 30 % with two seminal fluid proteins (HACP050: *Heliconius hereto* and CSSFP014: *Chilo suppressalis*). This similarity, although low, may be a sign of the high evolutionary rates documented for some of these SFPs [43, 45, 54–56]. The proteins AcpC03 and AcpC04 showed similarity to sequences from *Drosophila yakuba* and *D. mojavensis*, respectively, which have not yet been characterised. Searches against several other insect genomes, run in Flybase, yielded the same results.

Additionally, we tried to determine protein structure (3D fold) using PHYRE protein fold recognition metaserver and protein domains using InterProScan software. PHYRE did not yield consistent results for any of the genes tested. This approach, which proves the annotation in the tertiary structure of the proteins, was useful to annotate *Drosophila* SFPs [44], suggesting that candidate cicindelid SFPs do not meet the criteria found in *Drosophila*, due to differences in the structure and/or function.

In the AcpC03 gene, a Single domain von Willebrand factor type C (SVWC) was detected via InterProScan. SVWC family proteins, which are largely present in arthropods, normally contain ten cysteines, and are thought to respond to environmental challenges, such as bacterial infection and nutritional status [57, 58]. Several studies have pointed out that SFPs may be involved in the immune

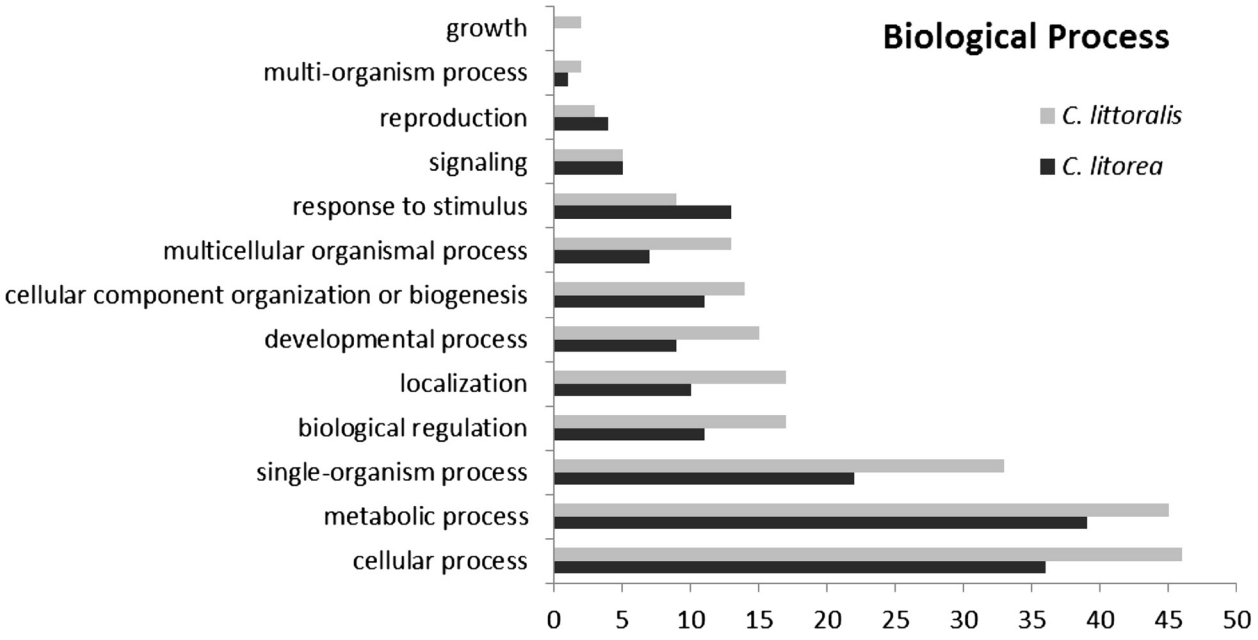


Fig. 1 GO term distribution in biological processes for *C. littoralis* and *C. litorea* gonad unigenes. Percentages are in proportion to the total biological process GO annotations

response, as mating processes can transfer numerous pathogens into the female tract, jeopardising the reproductive success. Several SFPs analysed in *D. melanogaster* seem to have direct antimicrobial activity, protecting the male and subsequently the female reproductive tracts and even eggs against bacterial infection [59, 60] and/or stimulation of antimicrobial gene expression levels [61]. Other putative SFPs have been identified in other Diptera, such as *A. aegypti* and *A. gambiae* [48, 62], which are related to immune response. In Coleoptera, South *et al*. identified a putative SFP in *Tribolium* which is a predicted prophenoloxidase, an important component of the innate immune response in Arthropoda [50]. However, Protfun identifies AcpC03 as a hormone. This result could be in accordance with protein classes that are found in seminal proteins in different animals. Wolfner stated in her

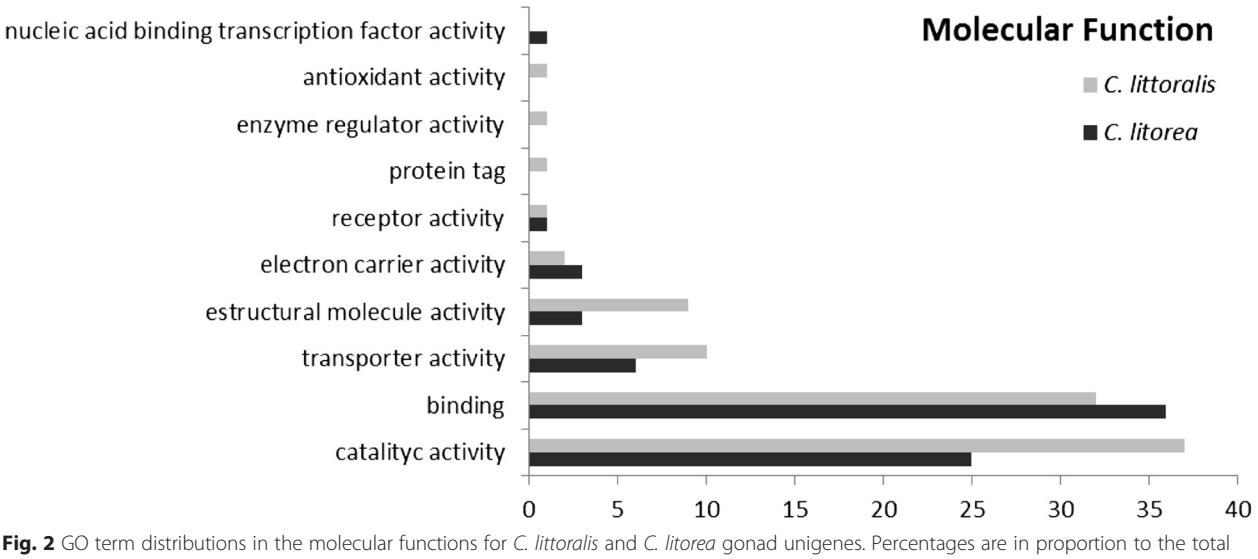


Fig. 2 GO term distributions in the molecular functions for *C. littoralis* and *C. litorea* gonad unigenes. Percentages are in proportion to the total molecular function GO annotations

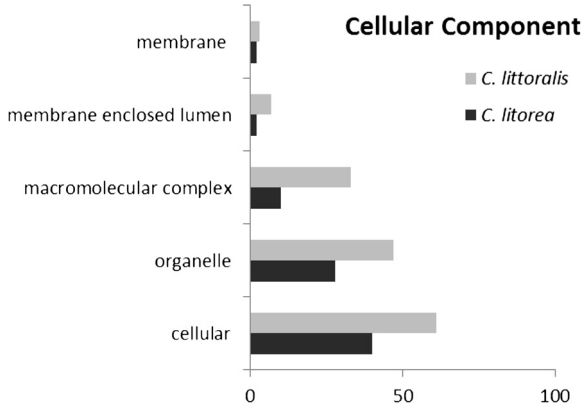


Fig. 3 GO term distributions in the cellular components for *C. littoralis* and *C. litorea* gonad unigenes. Percentages are in proportion to the total cellular component GO annotations

work that 40 % of accessory gland proteins appear to be peptide hormones or prohormones [63], and in *Drosophila melanogaster* ACP26Aa SFP was found to have similarity with califin C, a hormone from *Aplysia californica* [64] which is involved in the egg-laying process [65]. This could be an example of how a function assignment based on the sequence and structure similarity (InterproScan) could actually be different from a function assignment based not only on the structure but also on the physical/chemical and functional biological properties (Protfun). In other words, a conserved structure of a protein does not ensure a conserved function [66].

Finally, analysis with Protfun identified the gene AcpC01 to be an immune response protein. The other analyses did not assign a function to this protein based on similarity searches. However, Protfun analysis based on amino acid-derived input features did identify a function for this protein. This could be explained when considering the AcpC01gene as a novel putative SFP class that is either present only in tiger beetles (taxonomically restricted gene) or has not yet been characterised in other insects. The protein AcpC04 was identified as a transcription regulation factor by Protfun and also has similarities in Flybase and GenBank with a non-characterised protein in *D. mojavensis*. Although transcription regulation factors are not included within the described protein classes of SFP across animals [63], AcpC04 meets the requirements to be considered an SFP; therefore, further genetic studies may corroborate the biological function of this protein in tiger beetles. A similar consideration can be made in relation to the protein AcpC02, which according to the ontogenetic categories, is considered to be a growth factor by Protfun, although in this case, no homology has been found in the databases.

Homology in cicindelid species

RT-PCR and RACE-PCR primers were used in an attempt to amplify homologous sequences in the available cicindelid species (*C. litorea, C. littoralis, Lophyra flexuosa*, *Cephalota maura*, *Cephalota deserticoloides* and *Cylindera trisignata*). The AcpC04 gene, found originally in *C. litorea,* yielded homologous sequences in male abdomens of all of the analysed species (Fig. 5). This result is not surprising considering that although SFP are considered to have high evolutionary rate and low similarity between closely related species [25, 29, 40, 41], not all SFPs evolve rapidly and some loci are conserved between divergent taxa [25]. In this line of evidence, AcpC01 showed clear amplification in *C. litorea, C. littoralis, L. flexuosa* and *C. deserticoloides,* although weak

Table 3 Summary of the characteristics present in the candidate genes

	Contig	Cellular location	Signal peptide	Membrane helix
C. littoralis	13_59 (AcpC01)	Extracellular	No	Yes
	31_59	Plasma membrane	No	Yes
	50_59 (AcpC02)	Plasma membrane	No	Yes
	95_59	Extracellular	No	No
	129_59 (AcpC03)	Extracellular	Yes	Yes
	139_59	Plasma membrane	No	Yes
	161_59	Plasma Membrane	No	No
	171_59	Extracellular	No	No
	173_59	Plasma membrane	No	No
C. litorea	46_58	Plasma membrane	No	Yes
	70_58 (AcpC04)	Extracellular	No	No
	126_58	Plasma membrane	No	Yes
	101_58	Extracellular	Yes	No
	204_58	Extracellular	No	No

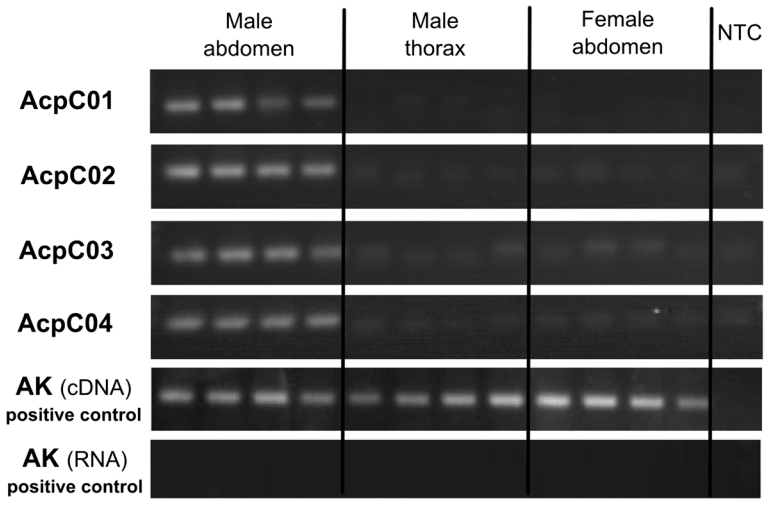


Fig. 4 Results from RT-PCR, showing the expected pattern of expression in the different tissues for the four putative SFPs. Arginine kinase gene was used as positive control and was amplified from cDNA and RNA (free of DNAse) in all samples. NTC indicates negative template control

signal in *C. trisignata* and no signal at all in *C. maura*. This result is coherent with the phylogenetic relationships obtained using cytochrome oxidase I gene (COI) (unpublished data) with the cicindelid species under study (Fig. 6). In fact *C. maura* is the most distantly related species.

However, AcpC03 showed amplification only in male abdomens of *L. flexuosa* (Fig. 5). The phylogenetic tree (Fig. 6) showed that *C. littoralis* and *L. flexuosa* are closely related species. This close relationship could explain why these putative SFPs obtained from *C. littoralis* were also found in *L. flexuosa.* The lack of amplification in the other cicindelid species tested might be interpreted as a consequence of the rapid evolution that is generally considered for SFPs [42, 43, 45, 54–56, 67]. AcpC02 showed amplification in *C. littoralis* and *C. deserticoloides* with the RACE-PCR primers, but using the RT-PCR primers only *C. littoralis* gave positive results. However the detection of this gene in these two species suggest that it might be present in the most recent common ancestor of these two species, although with the primer pairs used was not detected.

Conclusions

Two cDNA libraries were constructed from gonads of *C. littoralis* and *C. litorea* as a foundation to understanding the male reproductive system. A total of 568 and 576 ESTs were sequenced and analysed, assembled in 185 and 212 unigenes, respectively. Also, 75 and 64 contigs were annotated via Blast2go and no SFPs were found. We have identified 14 putative SFPs by bioinformatics analysis and found that four of them met the criteria of tissue-specific expression patterns, which led to them being considered as putative SFP. Functional annotation was difficult due to the fact that the four SFPs either do not show homology via similarity searches or reassemble with sequences whose function is unknown; only Interpro assigns a function in the immune response for AcpC03, as occurs with others insect SFPs. The predicted assignment of function via Protfun for AcpC01 and AcpC03 was immune response and hormone, respectively; these protein functions are included within the conserved protein classes of SFPs that have already been reported in several insect species. Searches for homology in other cicindelids showed that AcpC03 is

Table 4 Summary of the characterisation of cicindelids seminal fluid proteins

GEN	Amino acids size	CDS	GenBank hit	Flybase	PHYRE	Interpro	Protfun (gen ontology category)
AcpC01	66	Complete	No	No	NA	NA	Immune response
AcpC02	218	Complete	No	No	NA	NA	Growth factor
AcpC03	105	Complete	*Drosophila yakuba*/XP_002100905.1 (e-value 7e–4, 30 % similarity)	Yes/Dyak/GE17317-PA	NA	Svwc domine	Hormone
AcpC04	206	Complete	*Drosophila mojavensis*/XP_002007890.1 (e-value 8e–4, 30 % similarity)	Yes/Dmoj/GI12127-PA	NA	NA	Transcription regulation

The homologies and different functions assigned by the different software packages used are indicated

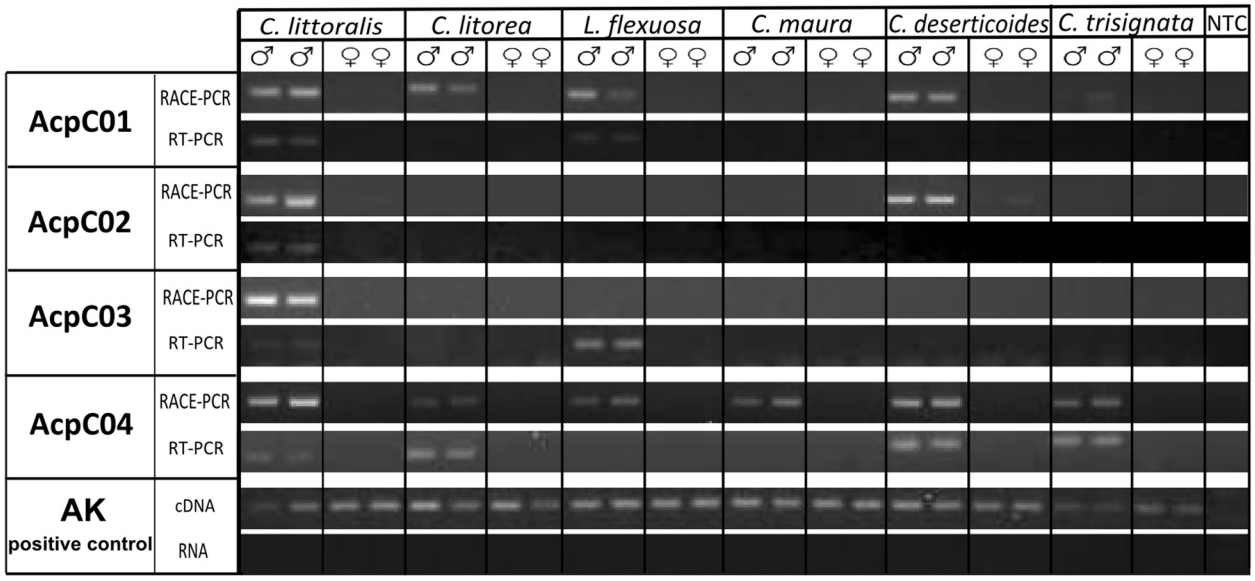


Fig. 5 Results of the SFPs homology searches. RT-PCR and RACE-PCR primers were used to amplify the four putative SFPs in six cicindelid species. Arginine kinase gene was used as positive control and was amplified from cDNA and RNA (free of DNAse) in all samples. NTC indicates negative template control

only present in *C. littoralis* and *Lophyra flexuosa*, species which have a close phylogenetic relationship. SFPs are among the most rapidly evolving proteins and therefore this new putative SPF might represent a linage-specific SFP involved in reproductive isolation between species. In contrast, AcpC04 is present in all cicindelid species analysed here, and could be an example of a highly conserved SFP, at least in tiger beetles. The same can be said about AcpC01, with is present in all species tested except for the most distantly related species *C. maura*. This work represents the first identification of putative SFPs in tiger beetles that are one of the best studied worldwide distributed non-pest insects (more than 2,500 species described), are important in ecosystems as predators and are commonly used as bioindicators. The identification of cicindelid SFPs (both the rapidly evolving and the highly conserved) could represent a significant approach for understanding the male reproductive system in cicindelids and furthermore the species richness and adaptive evolution in Adephagan beetles.

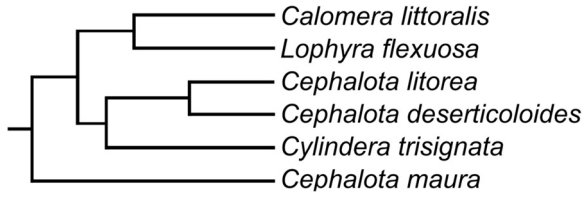


Fig. 6 Phylogenetic tree based on COI sequences (unpublished data) from the cicindelids species analysed

Methods

Library construction and EST assembly

Tiger beetles

Two cDNA libraries were constructed from male reproductive tissue (testes and accessory glands) of two *Calomera littoralis* males and a *Cephalota litorea* male. The developmental stage of the testes was that of sexually mature males with the final half of the testes full of spermatozoa (pearl white colour) and the anterior part with active meiosis (transparent white colour). Reproductive tissues were extracted and preserved in RNAlater (Qiagen, Crawley, UK) and stored at −20 °C prior to extraction.

cDNA libraries

According to the manufacturer's instructions, total RNA was precipitated using the RNeasy Protect Mini Kit (Qiagen, Crawley, UK). The RNA sample concentrations were quantified by spectrometry. mRNA was purified using the Oligotex ARNm mini kit (Qiagen, Crawley, UK). The BD SMART PCR cDNA Synthesis kit (BD Biosciences) was used for cDNA libraries construction. The first strand was synthesised using BD PowerScript Reverse Transcriptase, the SMART IIA Oligonucleotide and the CDS IIA primer provided in the kit. The double-stranded cDNA was synthesised by PCR and purified using Micropure-EZ (Millipore). The cDNA products were analysed using agarose gel electrophoresis to determine their quality before cloning. cDNA was ligated into a TOPO vector and was transformed by TOPO TA cloning (Invitrogen) using TOP10 chemically

competent *E. coli* cells. Libraries were plated on LB medium and grown at 37 °C overnight. Colonies were manually picked up for PCR amplification with M13 and T7 universal vector primers and subsequent sequencing with poly-T primer on an ABI Prism 3,700 sequencer (Applied Biosystems).

Annotation – gene ontology

The Seqman module of DNAStar (ver. Madison, WI) was used to remove the vector sequence, trim ends using program defaults and assemble sequences. Assembling parameters were: 80 % for minimal match percentage, 100 for minimal sequence length, 0 for gap penalty, and 0. 7 for gap length penalty.

Assembled sequences (contigs and singletons) were subjected to a similarity search for assigned putative protein functions using BlastX of the Blast2GO v2.5.0 software [68] with 10^{-3} for the cut-off E-value. BlastX reference was used from the non-redundant protein database of GenBank in the NCBI.

Gene ontology enrichment analysis was performed with Blast2GO mapping to determine protein functions in biological processes.

Identification of putative SFP

Indirect strategies based on bioinformatics tools were previously employed in insects to identify putative SFP unigenes [25, 27, 40, 69]. First, ORF of each unigene generated in Orfinder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) were selected and SignalP 4.1 [70] software was used to identify a predicted signal peptide. Additionally, integral prediction of protein location was analysed by ProtComp v9 software (http://www.softberry.com) and TMHMM tool in InterProScan [71] was used to determine the presence of a membrane helix. Those candidate genes with similar predicted functions to known SFPs (extracellular and/or with membrane destination, as the signal peptide can sometimes be recognised as a membrane helix) were also selected.

Second, patterns of tissue-specific expression were examined for each candidate SFP via RT-PCR. PCR primers were designed with PrimerExpress 3.0 (Applied Biosystems) (Table 5). Total RNA was isolated from the male abdomen, male thorax (pronotum) and female abdomen, in three males and three females of *C. littoralis* and *C. litorea*. RNA was extracted using the RNeasy Protect Mini Kit (Qiagen) following the manufacturer's protocol. Around 1,2 μg of each RNA extraction were treated with TURBO DNA free (Ambion, Life Technologies) and reverse-transcribed using the QuantiTect Reverse Transcription kit (Qiagen). One μl of a ten-fold dilution of cDNA was used as template for a 12,5 μl RT-PCR experiments. PCR was performed using the following cycling parameters: one cycle of 2 min at 96 °C, 35 cycles of 30 s at 96 °C, 30 s at 60 °C and 1 min at 72 °C, and a final extension of 10 min at 72 °C. The arginine kinase gene (AK) was used as a positive control using in this case cDNA and also RNA (DNAse treated) to discard genomic DNA amplification. PCR amplicons were electrophoresed on agarose gel with RedSafe™ (INTRON Biotechnologies, Korea).

Table 5 RACE-PCR and RT-PRC primers for the four putative seminal fluid proteins and the amplicon sizes

	Primers sequence	Amplicon size
AcpC01	**RACE_F1** 5′-GTATTCCATTGTGTCCACCACCTCCGG-3′	**128**
	RACE_R1 5′-TGGTGGACAAGGTGGACAACATGGAAC -3′	
	RT_F1 5′-TTGCCCTCCATGTGCAGTAC-3′	**139**
	RT_R1 5′-TGGCTTCTGTGGCTCAAATTT-3′	
AcpC02	**RACE_F2** 5′- TGAGGAACCAGCCGCACAAGTAAAGAC-3′	**191**
	RACE_R2 5′-AGACCGACTCTGCAGTTTTTGTCTCGG-3′	
	RT_F2 5′-AGGAACCAGCCGCACAAGTA-3′	**110**
	RT_R2 5′-CTCCTTGTTGGGTGGTGCAT-3′	
AcpC03	**RACE_F3** 5′-TCATAACGATGATTCTGCCGCTCGTGG-3′	**176**
	RACE_R3 5′-GACACTCGAGATGCCTACAGTCCGGTA-3′	
	RT_F3 5′-ATGCTGTGCTGCTTGTGCAT-3′	**100**
	RT_R3 5′-GGACAACAGGCCGGAAATG-3′	
AcpC04	**RACE_F4** 5′-ACCAGTTTGTGATTGTCCGCCGTTACG-3′	**300**
	RACE_R4 5′-GTGTAACTGAACGCACGGGAAATAGCC-3′	
	RT_F4 5′-GCTATTTCCCGTGCGTTCAG-3′	**100**
	RT_R4 5′-CGGAGATCTCGTCTGCGTTT-3′	

SFP characterisation

The SMARTer™ RACE cDNA Amplification Kit (Clontech Laboratories, Inc. Kyoto, Japan) was used to obtain the full length cDNA sequences. Total RNA from one *C. littoralis* male and one *C. litorea* male was used to obtain the first strand of 3′ and 5′ RACE Ready CDNA. Gene-specific primers for 5′ and 3′ RACE were designed (Table 5). The synthesis of first strand cDNA was performed following PCR conditions, as indicated the manufacturer's instructions. The amplification products were sequenced in triplicate by SAI at the University of Murcia (Spain) using an ABI Prism 3,130 Sequencer (Applied Biosystems) and assembled by GENEIOUS v5 [72] to obtain the complete cDNA sequence.

The complete ORFs derivate from the full length sequences were generated in Orfinder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and were used as queries in BlastX to search for homologues in other species with an e-value cut-off of 10^{-4} and identities >30 %. We also searched for similarity between our putative SFPs and other known insect proteins using Flybase [73]. Additionally, protein domains were searched again using InterProScan [74] and sequences were submitted to the PHYRE protein fold recognition server [75] to generate the protein structure.

Sequences were submitted to ProtFun 2.2 Server (http://www.cbs.dtu.dk/services/ProtFun/) for analysis of the GO ontology to predict function based on amino acid sequence-derived input features (physical/chemical and functional biological properties) such as predicted protein secondary structure, transmembrane helices, subcellular localisation and post-transcriptional modifications [66].

Homology in cicindelid species

To determine the presence/absence of these putative SFPs in other species, total RNA was extracted from the abdomen of two males and the abdomen of two females of the following species: *Lophyra flexuosa*, *Cephalota maura*, *Cephalota deserticoloides* and *Cylindera trisignata.* RNA was extracted using the RNeasy Protect Mini Kit (Qiagen) following the manufacturer's instructions. For each tissue type around 1, 2 μg of Extracted RNA was treated with TURBO DNA free (Ambion, Life Technologies) to remove DNA contamination, and reverse-transcribed using the QuantiTect Reverse Transcription kit (Qiagen). RACE-PCR and RT-PCR primers that had been previously designed (Table 5) were tested in all of the available samples under the same conditions previously described for each PCR reaction. The arginine kinase gene (AK) was used as a positive control using cDNA and RNA.

Abbreviations

EST: Expressed sequence tag; GO: Gene ontology; ORF: Open reading frame; RACE: Rapid amplification of cDNA ends polymerase chain reaction; RT-PCR: Reverse transcription polymerase chain reaction; SFP: Seminal fluid protein; TRG: Taxonomically restricted genes.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

MJR-G participated in the conception and design of the study, performed the bench work and data analysis and drafted the manuscript. VM participated in the conception, design of the study and data analysis. JG participated in the conception, design, and coordination of the study and helped write the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We greatly thank Alfried P. Vogler, Alessandra de Riva and Kosmas Theodorides for contributing library construction and sequencing. Thanks are due to PRS Proofreading Service for language editing of the paper and to two anonymous reviewers for providing valuable comments on this manuscripts. This work has been funded by project CGL2008–03628 of the Ministerio de Ciencia e Innovación, Spain. The authors thank the support of CNPq for granting a postdoctoral scholarship (PDE) to V.M.

Received: 25 December 2014 Accepted: 5 May 2015
Published online: 16 May 2015

References

1. Pearson DL. Tiger beetles as indicators for biodiversity patterns in Amazonia. Res Explor. 1992;8:116–7.
2. Pearson DL, Cassola F. World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): indicator taxon for biodiversity and conservation studies. Conserv Biol. 1992;6:376–91.
3. Carroll SS, Pearson DL. Spatial modeling of butterfly species richness using tiger beetles (Cicindelidae) as a bioindicator taxon. Ecol Appl. 1998;8:531–43.
4. Rodriguez JP, Pearson DL, Barrera R. A test for the adequacy of bioindicator taxa: are tiger beetles (Coleoptera: Cicindelidae) appropriate indicators for monitoring the degradation of tropical forests in Venezuela? Biol Conserv. 1998;83:69–76.
5. Cassola F, Pearson DL. Global patterns of tiger beetle species richness (Coleoptera: Cicindelidae): their use in conservation planning. Biol Conserv. 2000;95:197–208.
6. Pearson DL, Cassola F. A quantitative analysis of species descriptions of tiger beetles (Coleoptera: Cicindelidae), from 1758 to 2004, and notes about related developments in biodiversity studies. Coleops Bull. 2005;59:184–93.
7. Pearson DL. Biology of Tiger Beetles. Ann Rev Entomol. 1988;33:123–47.
8. Pearson DL, Blum MS, Jones TH, Fales HM, Gonda E, White BR. Historical perspective and the interpretation of ecological patterns: defensive compounds of tiger beetles (Coleoptera: Cicindelidae). Am Nat. 1988;132:404–16.
9. Willis HL. Bionomics and zoogeography of tiger beetles of saline habitats in the central United States (Coleoptera: Cicindelidae). Univ Kans Sci Bull. 1967;47:145–313.
10. Singh T, Gupta S. Morphology and histology of the mandibular gland in Cicindela sexpunctata Fabr. (Coleoptera: Cicindeliadae). Uttar pradesh. J Zool. 1982;2:14–8.
11. Horn W. Coleoptera Adephaga (family Carabidae, subfamily Cicindelinae). In: Wystman P, editor. Genera insectorum. Brussels: Desmet-Vereneuil; 1915. p. 1–486.
12. Horn W. Pars 86: Carabidae: Cicindelinae. In Junk and Schenkling Berlin editors. Belin: W.Junk; *Coleopterum catalogus*. 1926.
13. Wiesner J, Verzerichnis der Sabdlaufkäfer der Welt. Checklist of the tiger beetles of the world. Keltern, Germany: Verlag Erna Bauer; 1992.
14. Sota T. Variation of carabid life cycles along climatic gradients: an adaptive perspective for life-history evolution under adverse conditions. In: Danks HV, editor. Insect life-cycle polymorphism: theory, evolution; and ecological consequences for seasonal and diapause control. Dordrecht, Netherlands: Kluwer Academic Publishers; 1994. p. 91–112.
15. Spangler HG. Hearing in tiger beetles (Cicindelidae). Physiol Entomol. 1988;13:447–52.
16. Toh Y, Okamura JY. Behavioural responses of the tiger beetle larva to moving objects: role of binocular and monocular vision. J Exp Biol. 2001;204:615–25.
17. Dreisig H. Daily activity, thermoregulation and water loss in the tiger beetle *Cicindela hybrid*. Oecologia. 1980;44:376–89.

18. Pearson DL, Vogler AP. Tiger beetles: the evolution, ecology and diversity of the cicindelids. Nueva York: Cornell University Press; 2001.
19. Galián J, Proença SJR, Vogler AP. Evolutionary dinamics of autosomal-heterosomal rearrangements in a multiple-X chromosome system of tiger beetles (Cicindelidae). BMC Evol Biol. 2007;7:158.
20. Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, et al. Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol. 2002;3:re-search0029.
21. Nagaraj SH, Gasser RB, Ranganathan S. A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform. 2007;8:6–21.
22. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, et al. Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet. 2009;41:299–307.
23. Davies SJ, Chapman T. Identification of genes expressed in the accessory glands of male Mediterranean fruit flies (*Ceratitis capitata*). Insect Biochem Mol Biol. 2006;36:846–56.
24. Walters JR, Harrison RG. EST analysis of male accessory glands from *Heliconius* butterflies with divergent mating systems. BMC Genomics. 2008;9:592.
25. Walters JR, Harrison RG. Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. Mol Biol Evol. 2010;27:2000–13.
26. Almeida FC, DeSalle R. Orthology, function and evolution of accessory gland proteins in the Drosophila repleta group. Genetics. 2009;181:235–45.
27. Scolari F, Gomulski LM, Ribeiro JMC, Siciliano P, Meraldi A, Falchetto M, et al. Transcriptional profiles of mating-responsive genes from testes and male accessory glands of the Mediterranean fruit fly. Ceratitis capitata PLoS ONE. 2012;7, e46812.
28. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, et al. Evolution in the fast lane: rapidly evolving sex-related genes in drosophila. Genetics. 2007;177:1321–35.
29. Findlay GD, Yi XH, MacCoss MJ, Swanson WJ. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. PLoS Biol. 2008;6:1417–26.
30. Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. Insect seminal fluid proteins: identification and function. Annu Rev Entomol. 2011;56:21–40.
31. McGraw LA, Gibson G, Clark AG, Wolfner MF. Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. Curr Biol. 2004;14:1509–14.
32. Fiumera AC, Dumont BL, Clark AG. Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins. Genetics. 2005;169:243–57.
33. Wigby S, Chapman T. Sex peptide causes mating costs in female *Drosophila melanogaster*. Curr Biol. 2005;15:316–21.
34. Wigby S, Sorot LK, Linkalater JR, Buehner N, Calboli FCF, Bretman A, et al. Seminal fluid protein allocation and male reproductive success. Curr Biol. 2009;19:751–7.
35. Fedorka KM, Winterhalter WE, Ware B. Perceived sperm competition intensity influences seminal fluid protein production prior to courtship and mating. Evolution. 2011;65:584–90.
36. LaFlamme BA, Ram KR, Wolfner MF. The *Drosophila melanogaster* seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. PLoS Genet. 2012;8, e1002435.
37. Ritchie MG. Sexual selection and speciation. Annu Rev Ecol Evol Syst. 2007;38:79–102.
38. Snook RR, Chapman T, Moore PJ, Wedell N, Crudgington HS. Interactions between the sexes: new perspectives on sexual selection and reproductive isolation. Evol Ecol. 2009;23:71–91.
39. Castillo DM, Moyle LC. Intraspecific sperm competition genes enforce post-mating species barriers in Drosophila. P Roy Soc B-Biol Sci. 2014;281:2014–50.
40. Findlay GD, MacCoss MJ, Swanson WJ. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. Genome Res. 2009;19:886–96.
41. Marshall JL, Huestis DL, Garcia C, Hiromasa Y, Wheeler S, Noh S, et al. Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by post-mating pre-zygotic phenotypes. Mol Biol Evol. 2011;28:423–35.
42. Clark NL, Aagaard JE, Swanson WJ. Evolution of reproductive proteins from animals and plants. Reproduction. 2006;131:11–22.
43. Clark NL, Swanson WJ. Pervasive adaptive evolution in primate seminal proteins. PLoS Genet. 2005;1, e35.
44. Mueller JL, Ripoll DR, Aquadro CF, Wolfner MF. Comparative structural modelling and inference of conserved protein classes in *Drosophila* seminal fluid. Proc Natl Acad Sci U S A. 2004;101:13542–7.
45. Andres JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG. Molecular evolution of seminal proteins in field crickets. Mol Biol Evol. 2006;23:1574–84.
46. Andres JA, Maroja LS, Harrison RG. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc R Soc Lond B Biol Sci. 2008;275:1975–198.
47. Collins AM, Caperna TJ, Williams V, Garrett WM, Evans JD. Proteomic analyses of male contributions to honey bee sperm storage and mating. Insect Mol Biol. 2006;15:541–9.
48. Sirot LK, Poulson RL, McKenna MC, Girnary H, Wolfner MF, Harrington LC. Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction. Insect Biochem Mol Biol. 2008;38:176–89.
49. Baer B, Heazlewood JL, Taylor NL, Eubel H, Millar AH. The seminal fluid proteome of the honeybee *Apis mellifera*. Proteomics. 2009;9:2085–97.
50. South A, Sirot LK, Lewis SM. Identification of predicted seminal fluid proteins in *Tribolium castaneum*. Insect Mol Biol. 2011;20:447–56.
51. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 1997;10:1–6.
52. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: signalP 3.0. J Mol Biol. 2004;340:783–95.
53. Chapman T, Davies SJ. Functions and analysis of the seminal fluid proteins of male Drosophila *melanogaster* fruit flies. Peptides. 2004;25:1477–90.
54. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci U S A. 2001;98:7375–9.
55. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. Nat Rev Genet. 2002;3:137–44.
56. Clark AG, Aguade M, Prout T, Harshman LG, Langley CH. Variation in sperm displacement and its association with accessory-gland protein loci in *Drosophila melanogaster*. Genetics. 1995;139:189–201.
57. Sheldon TJ, Miguel-Aliaga I, Gould AP, Taylor WR, Conklin D. A novel family of single VWC-domain proteins in invertebrates. FEBS Lett. 2007;581:5268–74.
58. Zinke I, Schutz CS, Katzenberger JD, Bauer M, Pankratz MJ. Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. EMBO J. 2002;21:6162–73.
59. Samakovlis C, Kylsten P, Kimbrell DA, Engstrom A, Hultmark A. The andropin gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster*. EMBO J. 1991;10:163–9.
60. Lung O, Kuo L, Wofner MF. Drosophila males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. J Insect Physiol. 2001;47:617–22.
61. Mueller JL, Page JL, Wolfner M. An ectopic expression screen reveals the protective and toxic effects of *Drosophila* seminal fluid proteins. Genetics. 2007;175:777–83.
62. Rogers DW, Whitten MMA, Thailayil J, Soichot J, Levashina EA, Catteruccia F. Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. Proc Natl Acad Sci U S A. 2008;105:19390–5.
63. Wolfner MF. Battle and Ballet: molecular interactions between the sexes in *Drosophila*. J Hered. 2009;100:399–410.
64. Heifetz Y, Oliver L, Frongillo EA, Wolfner MF. The Drosophila seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. Curr Biol. 2000;10:99–102.
65. Rothman BS, Hawke DH, Brown RO, Lee TD, Dehghan AA, Shively JE, et al. Isolation and primary structure of the califins, three biologically active egg-laying hormone-like peptides from the atrial gland of *Aplysia californica*. J Biol Chem. 1986;261:1616–23.
66. Jensen LJ, Ussery DW, Brunak S. Functionality of system components: conservation of protein function in protein feature space. Genome Res. 2003;13:2444–9.
67. Panhuis TM, Clark NL, Swanson WJ. Rapid evolution of reproductive proteins in abalone and *Drosophila*. Philos Trans R Soc B. 2006;361:261–8.
68. Conesa A, Gotz S, Garcia-Gomez J, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.
69. Sonenshine DE, Bissinger BW, Egekwu N, Donohue KV, Khalil SM, et al. First transcriptome of the testis-vas deferens-male accessory gland and proteome of the spermatophore from *Dermacentor variabilis* (Acari: Ixodidae). PLoS ONE. 2011;6, e24711.

70. Nordahl Petersen T, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.
71. Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17:847–8.
72. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C et al. Geneious v5.5 [http://www.geneious.com] 2010
73. St Pierre SE, Ponting L, Stefancsik R, McQuilton P, the FlyBaseConsortium. FlyBase 102-advanced approaches to interrogating FlyBase. Nucleic Acids Res. 2014;42:D780–8.
74. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
75. Bennett-Lovsey RM, Herbert AD, Sternberg MJE, Kelley LA. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins. 2008;70:611–25.