



Facultad de Ciencias Naturales y Museo

Universidad Nacional de La Plata

Análisis genómico-funcional de la  
embriología de *Rhodnius prolixus* (Ståhl,  
1859) (Hemíptera, Reduviidae)

Lic. Andrés E. Lavore

2012

Director: Rolando Rivera-Pomar,

Co-director: Luís Diambra.

Centro Regional de Estudios Genómicos (CREG- UNLP).

Parte de los resultados de esta tesis fueron publicados en:

Lavore, A., Pagola, L., Esponda-Behrens, N., Rivera-Pomar, R, The gap gene giant of *Rhodnius prolixus* is maternally expressed and required for proper head and abdomen formation.. *Developmental Biology*, 2012. Vol. 361, Issue 1, 147-155. doi:10.1016/j.ydbio.2011.06.038.

# Índice

Agradecimientos .....	IV
Resumen .....	V
Abstract .....	VII
Abreviaciones .....	VIII
1 Introducción general .....	1
1.1 Los triatomíneos .....	1
1.2 Genómica de insectos .....	2
1.3 Biología del desarrollo .....	3
1.3.1 Componentes comunes en el desarrollo embrionario del reino animal .....	3
1.3.2 Desarrollo embrionario de <i>Rhodnius prolixus</i> .....	4
1.3.3 Mecanismos de segmentación .....	6
1.4 Objetivos .....	9
2 Materiales y Métodos .....	10
2.1 Colonia de <i>R. prolixus</i> .....	10
2.2 Genoteca de ADNc normalizada de <i>R. prolixus</i> .....	10
2.2.1 Extracción de ARN total .....	10
2.2.2 Producción de ADNc .....	11
2.2.3 Normalización de la genoteca de ADNc .....	13
2.2.4 Fraccionamiento y clonado .....	15
2.3 Búsqueda de genes de segmentación .....	16
2.3.1 Búsqueda bioinformática .....	16
2.3.2 Análisis de secuencias .....	17
2.3.3 Validación de las predicciones por PCR y clonado .....	18
2.4 Técnicas embriológicas .....	19
2.4.1 Colecta de embriones, disección y extracción de ovariolas .....	19
2.4.2 Hidridación <i>in situ</i> .....	19
2.4.2.1 Síntesis de las sondas de ARN .....	20
2.4.2.2 Tinción de embriones y ovariolas con sondas marcadas con Dig. ....	20
2.4.3 ARNi parental .....	22
3 Genoteca de ESTs normalizada de <i>R. prolixus</i> .....	23
3.1 Introducción .....	23
3.1.1 Estrategias y métodos de secuenciación .....	23
3.1.2 Búsqueda de similitud y homología de secuencia .....	25

3.1.3 Métodos para la búsqueda de similitud: BLAST .....	26
3.2 Resultados .....	28
3.2.1 Obtención de ARN total.....	28
3.2.2 Construcción de una genoteca de ADNc.....	28
3.2.3 Análisis de secuencias.....	31
3.3 Discusión.....	33
3.3.1 Representación de transcritos en la genoteca.....	33
3.3.2 Importancia de la construcción de genotecas y la secuenciación de ESTs .....	33
4 Genes de segmentación en <i>R. prolixus</i> .....	35
4.1 Introducción.....	35
4.2 Identificación de genes de segmentación.....	38
4.2.1 Genes Gap .....	38
4.2.1.1 <i>giant</i> .....	38
4.2.1.2 <i>Krüppel</i> .....	41
4.2.1.3 <i>hunchback</i> .....	43
4.2.1.4 <i>knirps</i> .....	45
4.2.1.5 <i>tailless</i> .....	48
4.2.2 Genes <i>gap</i> cefálicos.....	50
4.2.2.1 <i>orthodenticle</i> .....	50
4.2.2.2 <i>forkhead</i> .....	52
4.2.2.3 <i>empty spiracles</i> .....	54
4.2.2.4 <i>buttonhead</i> .....	56
4.2.3 Genes <i>pair-rule</i> .....	57
4.2.3.1 <i>hairy</i> .....	57
4.2.3.2 <i>runt</i> .....	59
4.2.3.3 <i>even-skipped</i> .....	61
4.2.4 Genes de polaridad de segmentos .....	63
4.2.4.1 <i>engrailed</i> .....	63
4.3 Análisis funcional de los genes <i>gap giant</i> y <i>Krüppel</i> .....	66
4.3.1 <i>giant</i> .....	66
4.3.1.1 Agrupamiento de reguladores en el promotor del gen <i>Rp-gt</i> .....	66
4.3.1.2 Expresión de <i>Rp-gt</i> .....	67
4.3.1.3 Fenotipo <i>giant</i> en <i>R. prolixus</i> .....	70
4.3.2 <i>Krüppel</i> .....	74
4.3.2.1 Análisis del promotor de <i>Rp-Kr</i> .....	74
4.3.2.2 Patrón de expresión de <i>Rp-Kr</i> .....	75

4.3.2.3 Fenotipo <i>Kr</i> en <i>R. prolixus</i> .....	77
4.4 Discusión.....	80
4.4.1 <i>Giant</i> .....	80
4.4.1.1 Características conservadas del gen <i>gt</i> .....	80
4.4.1.2 Expresión materna de <i>Rp-giant</i> .....	81
4.4.1.3 Expresión cigótica de <i>Rp-gt</i> y el fenotipo <i>gap giant</i> . .....	82
4.4.2 <i>Krüppel</i> .....	84
4.4.2.1 La conservación del gen <i>Krüppel</i> en insectos. ....	84
4.4.2.2 Expresión cigótica y fenotipo de <i>Rp-Kr</i> .....	85
5 Conclusión.....	89
6 Bibliografía .....	92
Anexo 1 .....	99
Anexo 2 .....	101

## Agradecimientos

Principalmente agradezco y dedico esta tesis a Verónica y Bruno, las dos personas más importantes en mi vida, que me aguantaron y acompañaron durante este proceso.

A mis padres y familia, que siempre confiaron en mí al elegir esta carrera y acompañarme en todo momento.

A mis amigos, los de Quilmes y los de La Plata, que siempre me acompañaron y me ayudaban a cambiar de aire durante el transcurso de la tesis.

A Rolando, un director que siempre estuvo para lo que se necesitó (a veces a último momento.....pero estuvo, je) y me dio todo lo necesario para que esta tesis se realice.

A Luisito, con quien siempre pude contar para discutir y analizar resultados.

A Gustavo Saldaña que fue quien me enseñó todo lo que sé de trabajo de laboratorio.

A Lucia y Natalia, cuya discusión y críticas de los resultados ayudaron a la realización de esta tesis.

A Diego, que me ayudó con algunas filogenias y además de ser un gran amigo, es un excelente compañero de trabajo.

Al resto de la gente del laboratorio con los cuales pude consultar y discutir resultados y que día a día hicieron del trabajo un lugar cómodo y agradable donde además de trabajar me divertía.

## Resumen

La genómica y biología del desarrollo de *Rhodnius prolixus* (vector de la enfermedad del Chagas) no fue estudiada sino hasta recientemente. El genoma de *R. prolixus* fue secuenciado y actualmente se encuentra en proceso de anotación. En este contexto, y dentro del consorcio de secuenciación del genoma de *R. prolixus*, se llevó a cabo la construcción de una genoteca normalizada de ADNc incluyendo todos los estadios de desarrollo, desde huevos hasta insectos adultos, tanto hembras como machos. Tanto esta genoteca como la última versión de anotación del genoma de *R. prolixus* fueron utilizados para la búsqueda de genes de segmentación. Los genes identificados y caracterizados fueron los siguientes: *giant*, *krüppel*, *hunchback*, *knirps*, *tailless*, *orthodenticle*, *empty-spiracles*, *forkhead*, *hairy*, *even skipped*, *runt* y *engrailed*.

A partir de la genoteca normalizada de ADNc se clonaron los genes *Rp-gt* y *Rp-Kr*, los cuales fueron caracterizados, mostrando su patrón de expresión y función durante el desarrollo embrionario. Mediante el análisis bioinformático de la región genómica *upstream* de las unidades transcripcionales de *Rp-gt* y *Rp-Kr*, se pudo identificar (para cada uno de estos genes) un potencial elemento regulatorio. Estos resultaron ser similares en posición y composición a los sitios de unión a factores de transcripción analizados en *Drosophila melanogaster*. Como referencia también se analizó la región equivalente en *Tribolium castaneum*. *Rp-gt* muestra expresión materna, de forma tal que el transcrito se encuentra en ovarios y oocitos sin fertilizar. En embriones en estadio de pre-blastodermo la distribución del transcrito es en parches formando un gradiente posterior. La expresión cigótica de *Rp-gt* se da en dos dominios, uno cefálico y otro abdominal. Mediante ARNi parental, se pudo ver que *Rp-gt* es requerido para la correcta formación de la cabeza y el abdomen. La cabeza pierde los apéndices mandibulares y maxilares, se reduce la longitud del clípeo-labro y el abdomen pierde segmentos anteriores. La expresión de *Rp-Kr* es cigótica. Durante el proceso de invaginación del embrión, *Rp-Kr* se expresa en la mitad posterior del huevo, mientras que durante el estadio de banda germinal se expresa en la parte central del embrión, desde el segmento torácico T2 hasta los primeros segmentos abdominales. Los embriones interferidos para *Rp-Kr* presentan una alteración del patrón de segmentación. Estas alteraciones corresponden a la zona media del embrión, donde se pierden el 2<sup>do</sup> y 3<sup>er</sup> segmento torácico y el abdomen se reduce de diez a seis segmentos. Además, los embriones con fenotipo interferido, muestran dificultades en la formación del segmento labial y T1, y cambios homeóticos, en los cuales aparece un peine tibial ectópico en la tibia de la pata T2.

En la presente tesis doctoral, se muestra un análisis bioinformático de una genoteca normalizada de ESTs y la caracterización de la mayoría de los ortólogos de genes de

segmentación para *R. prolixus*. Mostramos las potenciales regiones regulatorias para los genes *Rp-gt* y *Rp-Kr*, cuya posición se encuentra evolutivamente conservada tanto en *D. melanogaster* como en *T. castaneum*. Por último, se demuestra tanto estructural como funcionalmente que *Rp-gt* y *Rp-Kr* son verdaderos genes gap, el primero para la región cefálica y abdominal, y el segundo en la región central del embrión de *R. prolixus*.

## Abstract

The genomic and developmental biology of the Chagas disease vector *Rhodnius prolixus* has not been studied until recently. The genome of *R. prolixus* has been sequenced and it is actually in annotation process. In this context we performed a normalized cDNA library of all developmental stages from recently laid eggs to adults, both males and females. This library and the last annotation version of the *Rp*-genome are used to search the segmentation genes involved in the development of *R. prolixus*. Thus, we identified and characterized the following segmentation genes orthologs in *R. prolixus*: *giant*, *Krüppel*, *hunchback*, *knirps*, *tailless*, *orthodenticle*, *empty spiracles*, *forkhead*, *hairy*, *even-skipped*, *runt* and *engrailed*.

We cloned the *Rp-gt* and *Rp-Kr* genes from a normalized cDNA library and their expression pattern and function were characterized. The bioinformatic analysis of genomic sequence upstream of the *Rp-gt* and *Rp-Kr* transcriptional units showed a cluster of *bona fide* regulatory binding sites, which are similar in location and structure to the regulatory domains of the *Drosophila melanogaster* orthologs. *Rp-gt* is expressed in ovaries and maternally supplied in the early embryo. The maternal contribution forms a gradient of scattered patches of mRNA in the preblastoderm embryo. Zygotic *Rp-gt* is expressed in two domains that after germ band extension are restricted to the head and the posterior growth zone. Parental RNAi shows that *Rp-gt* is required for proper head and abdomen formation. The head lacks mandibular and maxillary appendages and shows reduced *clypeus-labrum*, while the abdomen lacks anterior segments. *Rp-Kr* expression is only zygotic. That it is expressed in the posterior half of the egg during the invagination process and in a central domain during the germ band stage – from T2 to the first abdominal segments. The *kr*-RNAi embryos have altered their corporal segmentation pattern. These alterations occur in the central part of the embryo, where *Rp-Kr* is expressed. These embryos lack the 2<sup>nd</sup> and 3<sup>rd</sup> thoracic segments, and the abdominal segments are reduced from ten to six. In addition to this, the embryos with *Rp-Kr* strong phenotype show problems in the labial and T1 thoracic segment formation, and they show homeotic changes – tibial combs in the T2 thoracic leg.

In the present doctoral thesis, we show a long bioinformatic analysis of a normalized library, and the characterization of the majority of the segmentation genes orthologs for *R. prolixus*. We showed putative regulation sequences for *Rp-gt* and *Rp-Kr*, which are conserved in position both in *D. melanogaster* and in *Tribolium castaneum*. Finally, we demonstrated that *Rp-gt* and *Rp-Kr* are truly gap genes on the head and abdomen, and in the central part of the embryo, respectively.

## Abreviaciones

ADNc: molécula de ADN complementaria a un mensajero de ARN.  
ADN: Acido desoxirribonucleico.  
ARN: Acido ribonucleico.  
ARNi: refiriéndose a la técnica de ARN de interferencia.  
ARNpol: ARN polimerasa.  
ARNm: Mensajero de ARN.  
ARNt: ARN de transferencia.  
BrEt: Bromuro de Etidio.  
ESTs: del ingles Expressed sequence tags.  
dATP: trifosfato de deoxiadenosina  
dNTP: trifosfato deoxiribonucleotido.  
DMSO: Dimetilsulfoxido.  
PO: Post ovipuesta.  
DTT: Ditiotreititol.  
DO: Densidad óptica  
dH<sub>2</sub>O: Agua destilada.  
Fw: Del ingles *Forward* aplicado a los *primers*.  
Grad: Gradiente.  
Kpb: Kilopares de bases.  
kV: Kilovoltio.  
mF: microfaradio.  
M: Molar.  
mM: milimolar.  
μM: micromolar.  
μl: microlitros.  
nm: nanometros .  
ON: del ingles *Over Nigh* – durante la noche, en relación a incubaciones de 8 a 12 hs de duración.  
ORF: del ingles *Open Reading Frame* – Marco de Lectura Abierto.  
Pb: par de bases nucleotídicas  
PCR: del ingles *Polimerase Chain Reaction*.  
PBS: del ingles *Phosphate Buffer Saline*.  
PFA: paraformaldeido.  
PWMs: del ingles *Position Weight Matrices*.  
RACE: del ingles *Rapid amplification of cDNA ends*.  
rpm: revoluciones por minuto.  
Rv: del ingles *Reverse*, aplicado a los *primers*.  
T°a : Temperatura de *anealling* (Temperatura Hibridacion de los *primers*).  
U: Unidades enzimaticas.  
UTR: del ingles *untranslated region* (región no traducida) de un mensajero.  
Zn-F: del ingles *Zinc Finger* (haciendo referencia al dominio proteico tipo dedo de Zinc)

## Lista de abreviaturas de aminoácidos

A	Alanina	C	Cisteína	D	Aspartato
E	Glutamato	F	Fenilalanina	G	Glicina
H	Histidina	I	Isoleucina	K	Lisina
L	Leucina	M	Metionina	N	Asparagina
P	Prolina	Q	Glutamina	R	Arginina
S	Serina	T	Treonina	V	Valina
W	Triptofano	Y	Tirosina	*	Stop

# 1 Introducción general

## 1.1 Los triatominos

El suborden Heteróptera (Hemíptera) es el grupo de insectos hemimetábolos más grande y morfológicamente más diverso. Es un grupo cosmopolita, encontrándose en todos los continentes (a excepción de Antártida). Esta gran adaptabilidad a diversos ambientes resulto en una amplia diversidad estructural y morfológica (Schuh y Slater 1995). Dentro de los Heterópteros, la familia Reduviidae es la que aporta la mayor diversidad de especies. Los miembros de esta familia se caracterizan por sus hábitos predadores, a excepción de la subfamilia Harpactorinae que tiene hábitos fitófagos y la subfamilia Triatominae que tiene hábitos hematófagos (Schuh y Slater, 1995; Melo, 2006). De esta forma los Redubidos tienen una gran importancia económica y sanitaria.

Los Triatominos habitan los mismos nichos que habitan sus hospederos, alimentándose de la sangre de aves, reptiles y mamíferos, incluido el hombre. Su distribución abarca desde México hasta el sur de América del Sur (su límite más austral es el sur de Argentina). La subfamilia Triatominae contiene 5-6 tribus, con 17 géneros y 130 especies. Dentro de esta familia, los géneros *Rhodnius*, *Panstrongylus* y *Triatoma* son los vectores más importantes de la enfermedad del Chagas. *Rhodnius prolixus* (Ståhl, 1859) tiene un distribución que abarca desde el sur de México hasta el norte y centro de América del sur. En América latina entre 10-15 millones de personas padecen la enfermedad. En Argentina la enfermedad es transmitida por *Triatoma infestans*, y se estima que existen aproximadamente 2.300.000 enfermos de Chagas, lo que representa el 7,2% de la población. *R. prolixus* tiene un ciclo de vida más corto que *T. infestans* (3 meses frente a los 6 meses que dura en *T. infestans*). Esto lo convierte en un modelo de estudio más flexible y práctico para el trabajo de laboratorio.

A lo largo del ciclo de vida de *R. prolixus* un individuo atraviesa cinco estadios larvales (Figura 1) y esto, junto a su modo de alimentación hematófaga, hace que sufran periódicamente cambios abruptos en la activación de los procesos fisiológicos en las horas posteriores a la alimentación. Dentro de estos se incluyen la osmoregulación, ecdisis y reproducción. Esto permitió que desde mediados de siglo XX hasta la fecha, *R. prolixus* haya sido considerado como un excelente modelo para estudios en fisiología, biología celular y reproducción.



**Figura 1:** Estadios de desarrollo post-embionario de *R. prolixus*. Se muestran los cinco estadios larvales desde L1 a L5 y un ejemplar adulto. Foto disponible en internet tomada por Thierry Heger.

## 1.2 Genómica de insectos

Uno de los mayores avances de la biología en los últimos tiempos fue la secuenciación y decodificación del genoma de diversos organismos. Esto fue posible gracias a la aparición de la genómica como una disciplina independiente dentro de la genética. La genómica involucra el estudio del total de la información contenida en el ADN. La genómica, junto al conocimiento de la genética y el genoma de un organismo modelo, como *D. melanogaster*, permitió aplicar sus técnicas y herramientas al estudio de otros grupos de insectos (Evans y Gundersen-Rindal, 2003; Heckel, 2003). Dentro de los insectos que han sido secuenciados y anotados casi en su totalidad, se encuentran: *D. melanogaster*, *Culex quinquefasciatus*, *Aedes aegypti*, *Apis mellifera*, *Tribolium castaneum*, *Pediculus humanus* y *Nasonia vitripennis* (Adams, *et al.*, 2000; Nene *et al.*, 2007; Stephen, *et al.*, 2008; Werren, *et al.*, 2010; Kirkness, *et al.*, 2010; Arensburg, *et al.*, 2010). Actualmente *Rhodnius prolixus* y *Macrosiphum pisi* se encuentran en proceso de secuenciamiento y anotación.

Los genomas de los insectos antes mencionados, fueron secuenciados a partir de genotecas de plásmidos o fagos, esto requiere largos tiempos de producción y secuenciación, además de altos costos. En los últimos años se puso en práctica la secuenciación de genomas utilizando secuenciadores de nueva generación (“*next generation sequencing*”). Esta tecnología está representada por la estrategia de pirosecuenciación 454 (Roche) y *Bridge-PCR* (Illumina). Ambos secuenciadores generan un gran número de secuencias cortas que posteriormente son ensambladas para cubrir la extensión total del genoma del organismo. Esta tecnología es muy eficiente, pudiendo secuenciarse el genoma de un único individuo en menor tiempo y con menores costos que las estrategias tradicionales. Recientemente, Lott y col. (2011), consiguieron analizar el transcriptoma completo de un único embrión de *D. melanogaster* mediante el pirosecuenciador 454. Este es uno de los ejemplos que demuestra que la secuenciación de genomas dejó de ser solo la producción de información global de una especie,

para pasar a ser una herramienta rápida para la generación de información de un determinado individuo en determinado momento de su ciclo de vida y bajo diferentes condiciones de estudio.

El desarrollo de la genómica tiene un gran impacto en la entomología básica contribuyendo a comprender mejor la evolución y taxonomía de un grupo, así como en la entomología aplicada, en relación con el control de plagas o el mejoramiento de insectos beneficiosos (Kaufman *et al.* 2002). Dentro de este marco, conocer el genoma de *R. prolixus* no solo es importante desde el punto de vista médico, sino también desde el punto de vista biológico, dado que es el primer insecto hemimetábolo al cual se le secuenció el genoma.

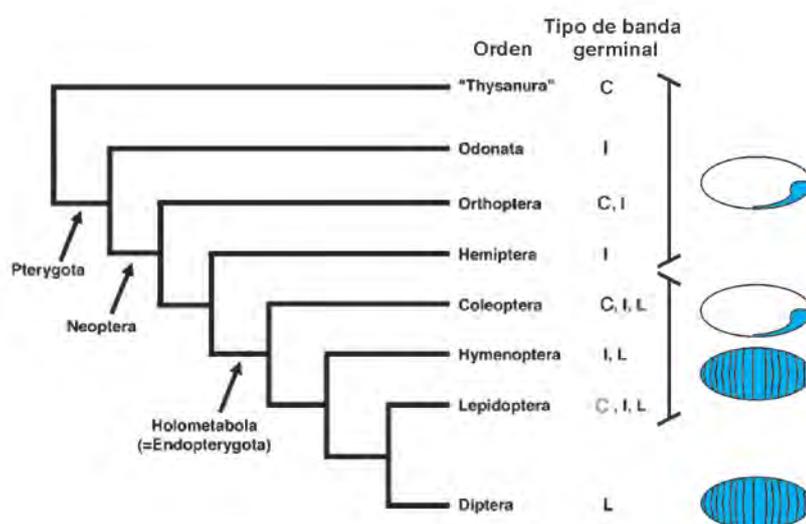
### 1.3 Biología del desarrollo

#### 1.3.1 Componentes comunes en el desarrollo embrionario del reino animal.

Durante el desarrollo embrionario de insectos ocurren una gran variedad de procesos que culminan en un mismo patrón morfológico, la división del cuerpo en los tagmatas característicos de insectos (cabeza, tórax y abdomen), y a la vez en una gran diversidad. En todos los insectos la embriogénesis comienza con repetidas divisiones del núcleo cigótico pero sin divisiones celulares. Estos núcleos luego de un número determinado de ciclos de división forman lo que se denomina blastodermo, el cual rodea al vitelo. Posteriormente, el blastodermo se celulariza, formando el blastodermo celular, a partir del cual se formará el embrión. En algunos insectos, como *G. bimaculatus* y *T. castaneum*, el estadio de blastodermo no se prolonga mucho en el tiempo, sino que a partir de éste se van a diferenciar dos poblaciones celulares, una de estas va migrar hacia la zona posterior del huevo para formar un rudimento embrionario (el “esbozo germinal”); y la otra va a formar los tejidos extraembrionarios.

En insectos se pueden distinguir tres tipos diferentes de desarrollo embrionario en base a la relación entre el esbozo embrionario segmentario y la fracción extraembrionaria (Figura 2). En función de esto la embriogénesis se puede dividir en: embriogénesis de banda germinal corta, intermedia y larga (Krause, 1939; Davis y Patel, 2002). La embriogénesis de banda germinal larga se da en insectos más derivados en la evolución, como son Dípteros e Himenópteros, en los cuales todos los segmentos son determinados durante el estadio de blastodermo, que ocupa la totalidad del huevo (ej. *D. melanogaster* y *N. vitripennis*). Por otro lado, la embriogénesis de banda germinal corta e intermedia es una forma de embriogénesis más primitiva (ej. *G. bimaculatus*, *T. castaneum* y *R. prolixus*); donde durante el estadio de blastodermo sólo se determinan los segmentos cefálicos y torácicos, quedando una menor fracción del blastodermo como región embrional. El resto de los segmentos aparecen más tarde en el desarrollo, de forma secuencial a partir de la “zona de crecimiento” en la parte posterior del embrión. El proceso de segmentación ha sido ampliamente estudiado en *D. melanogaster*

donde los segmentos del embrión se especifican al mismo tiempo en un entorno sincitial. El hecho de que el proceso ocurra en un entrono semejante hace posible la difusión a través del embrión en segmentación de diferentes factores localizados en los polos anterior y posterior, que por interacciones mutuas van a determinar la identidad de cada región del embrión (Rivera-Pomar y Jackle, 1996). En el caso de los insectos de banda germinal corta e intermedia, en los cuales no todos los segmentos se especifican al mismo tiempo, el mecanismo de regulación y determinación de segmentos cambia, no pudiendo realizarse completamente por acción de gradientes en un entorno sincitial. En estos insectos la identidad de segmentos va a estar dada por la interacción de los diferentes factores de transcripción en los segmentos que surgen a partir de la zona de crecimiento en un entorno celularizado que requiere la señalización por proteínas de membrana.

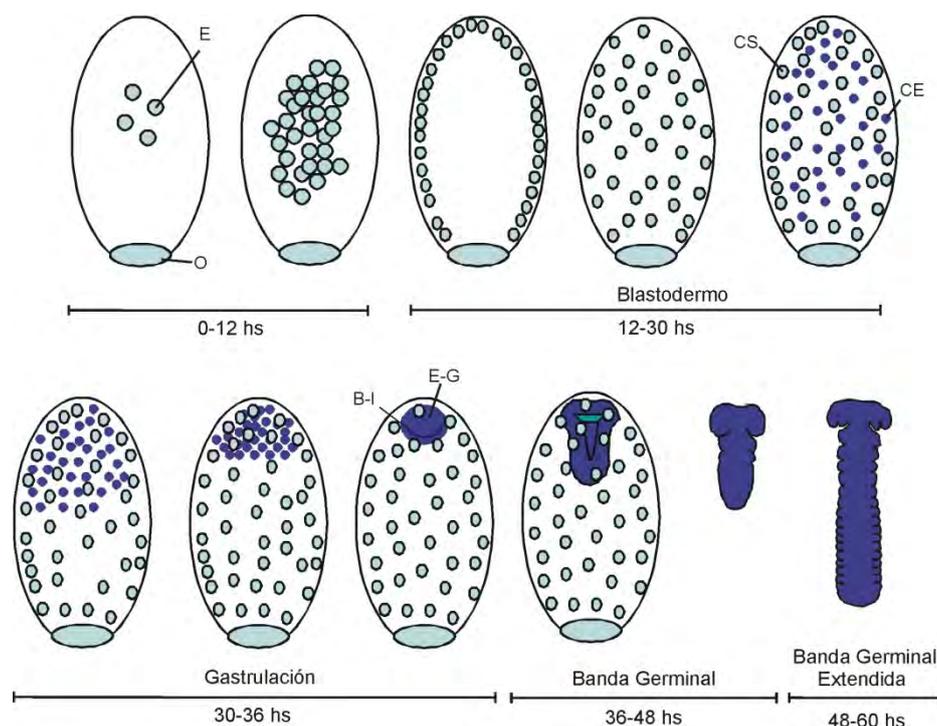


**Figura 2:** Filogenia de insectos y tipos de banda germinal. Se muestran los diferentes tipos de banda germinal en función de los taxones en el árbol filogenético. Como se puede ver, el tipo de banda germinal corta e intermedia se encuentra en taxones más basales de la evolución del grupo. Banda germinal corta en Lepidópteros es tentativa y se marca en gris. Abreviaciones: C: banda germinal corta, I: intermedia y L: larga. Figura basada en la mostrada en Davis y Patel (2002).

### 1.3.2 Desarrollo embrionario de *Rhodnius prolixus*

El desarrollo embrionario de *R. prolixus* ha sido estudiado a nivel morfológico y citológico como parte de los trabajos histo-embriológicos de Mellanby (1934, 1935) y de oogénesis de Kelly y Huebner (1989). Sin embargo resta aún un estudio embriológico detallado que refleje la dinámica de expresión génica y que permita el análisis de patrones de expresión y fenotipos. Ese estudio está realizándose actualmente en nuestro laboratorio y forma parte de la tesis de doctorado de la Lic. Natalia Esponda-Berens y Lic. Lucia Pagola. El huevo de *R. prolixus* mide entre 0,70-0,75 mm de ancho y 1,2-1,5 mm de largo, es dorsalmente convexo y

ventralmente cóncavo y en la parte anterior se encuentra el opérculo, a partir del cual eclosiona la larva I (Mellanby, 1934). Luego de la fecundación y ovoposición (0-12 hs post ovipuesta), tras una serie de divisiones nucleares del cigoto, aparecen un grupo de núcleos rodeados de citoplasma: las enérgidas. Pasadas las 12-24 hs PO, las enérgidas migran al cortex del huevo, donde comienzan una serie de divisiones nucleares asimétricas, para finalizar rodeando al vitelo formando el blastodermo sincitial. En torno a las 30 hs PO la monocapa celular se desdobra y aparecen dos poblaciones celulares en la superficie del huevo. Una de ellas son células cuboides y con núcleos pequeños que posteriormente migraran a la parte posterior del huevo para formar el rudimento embrionario. Para las 30-36 hs PO comienza el proceso de anatrepsis en el polo posterior del huevo, donde las células del esbozo germinal se invaginan. Para las 36-48 hs PO el embrión forma lo que se denomina banda germinal. En este estadio el embrión consta de la región cefálica, los segmentos torácicos y la zona de crecimiento, a partir de la cual los segmentos aparecen secuencialmente. Entre las 48 y 60 hs PO se completa la segmentación del embrión (Figura 3). Aproximadamente a las 120 hs PO se produce la katatrepsis, proceso en el cual el embrión gira  $180^\circ$ , momento en el cual la parte anterior del embrión coincide con la parte anterior del huevo. Durante esta fase del desarrollo los apéndices completan su formación y se produce el cierre dorsal del embrión al mismo tiempo que se produce la organogénesis. A los 14 días a partir de la puesta de los huevos, el embrión está listo a eclosionar. Según datos producidos por Kelly y Huebner (1989), a  $28^\circ\text{C}$  y con una humedad relativa de 70% la embriogénesis se completa en 15 días PO. Los tiempos de cada fase de desarrollo del embrión son relativos y dependen en gran parte de la temperatura a la cual se mantienen los insectos en el bioterio, observándose diferencias entre los datos presentados por Mellanby (1934) y Kelly y Huebner (1989) y los obtenidos a partir de la colonia establecida en el CREG (Esponda-Behrens y Pagola, comunicación personal).



**Figura 3:** Desarrollo embrionario de *R. prolixus*. Se muestra las diferentes etapas del desarrollo embrionario en los tiempos datadas en nuestro laboratorio. 0-12 hs PO, huevo sin núcleos en superficie. 12-24 hs PO, estadio de blastodermo, pasando de blastodermo acelular a celularizado al llegar a las 24 hs. 30-36 hs, migración de células embrionarias al polo posterior culminando en la gastrulación del embrión cercano a las 36 hs post-puesta. 36-48 hs PO, estadio de banda germinal hasta llegar a extensión completa a entre las 48-60 hs PO. Abreviaciones: B-I: Borde de invaginación, CE: células embrionarias, CS: células de la serosa, E: enérgidas y E-G: esbozo germinal.

### 1.3.3 Mecanismos de segmentación

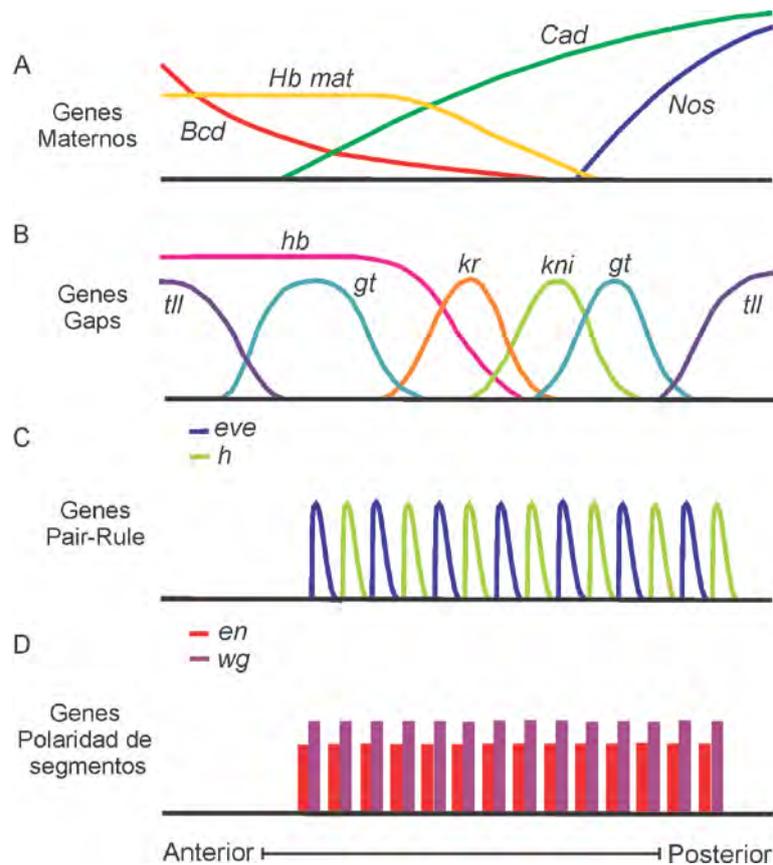
A partir de la década del 80' surgió la genética del desarrollo como una rama de la embriología, para estudiar las bases moleculares que determinan los procesos embriológicos de los organismos. Luego del *screening* de genes embrionarios en *D. melanogaster* realizado por Nüsslein-Volhard y col. (Jurgens, Nüsslein-Volhard y Wieschaus, 1984), se extendieron los estudios genéticos y moleculares a insectos pertenecientes a diferentes órdenes, lo que permitirá entender los procesos moleculares que gobiernan la embriología de los insectos dentro de un marco evolutivo (Peel, 2005).

La maquinaria molecular que gobierna la segmentación de insectos se basa en una estructura jerárquica de los genes identificados por Nüsslein-Volhard y col., los cuales conforman un circuito de activación-represión. Los patrones espacio-temporales de expresión génica que se ejecutan durante la segmentación fueron bien establecidos en *D. melanogaster* (Lawrence, 1992, Rivera Pomar y Jäckle, 1996). En trabajos posteriores se demostró que ésta

red jerárquica también se encuentra en el resto de los artrópodos (Heming, 2003) (Figura 4). Este mecanismo de control de la segmentación comienza con factores que son transferidos al huevo por la madre durante la oogénesis. Estos son los llamados genes maternos, los cuales forman gradientes anteriores y posteriores que determinan la polaridad del embrión. Dentro de esta categoría se encuentran: *bicoid* (*bcd*), *caudal* (*cad*), *nanos* (*nos*) y *hunchback* (*hb*). Un nivel por debajo de los genes maternos, y regulados por estos, se encuentran los genes *gap*: *hb*, *giant* (*gt*), *Krüppel* (*kr*), *knirps* (*kni*), *tailless* (*tll*), *orthodenticle* (*otd*), *buttonhead* (*btd*) y *empty spiracles* (*ems*). Estos genes poseen dominios de expresión solapados y subdividen al embrión en territorios de expresión génica. La actividad conjunta de los genes *gap* regula la expresión de los genes *pair-rule*, los cuales se expresan en siete bandas alternadas a lo largo del eje antero-posterior del embrión, ocupando un parasegmento embrionario (unidad morfológica embrional que ocupa la mitad posterior del segmento anterior y la mitad anterior del segmento posterior del adulto). Al final de esta cascada de segmentación se encuentran los genes de polaridad de segmentos, los cuales se expresan en catorce bandas a lo largo del eje antero-posterior, ocupando una de las mitades de cada uno de los segmentos, determinando así su polaridad (Lawrence, 1992).

Durante la evolución, el proceso de segmentación ha divergido, y la regulación de este proceso como se explica arriba, corresponde a modo derivado de segmentación. En hemimetábolos y algunos holometábolos, con desarrollo de banda germinal corta e intermedia, la segmentación es secuencial. Este proceso es regulado principalmente por Notch y Delta, los cuales determinan un “reloj biológico” en función del cual se adicionan los segmentos corporales (Carroll *et al.*, 1988; Damen *et al.*, 2000; Pueyo *et al.*, 2008; Sommer y Tautz, 1993). Alejados filogenéticamente, en vertebrados el proceso de segmentación es semejante al de hemimetábolos y holometábolos inferiores, a través de la adición sucesiva de segmentos. Esto es un ejemplo de la importancia de la genética del desarrollo en la interpretación de los procesos evolutivos y de cómo los diferentes factores que intervienen en la determinación de un patrón corporal son “reciclados” a lo largo de la evolución en reiterados momentos con esos diferentes.

Hasta el momento no se realizaron trabajos sobre la genética del desarrollo en *R. prolixus*. Siendo este insecto de interés médico y evolutivo, resulta importante conocer qué factores intervienen durante el proceso de segmentación y cómo estos interactúan durante el desarrollo embrionario para generar el patrón corporal. Esto permitirá interpretar el marco evolutivo donde se encuentra *R. prolixus*.



**Figura 4:** Cascada de genes de segmentación en *D. melanogaster*. **A.** Genes maternos. Estos corresponden a transcritos que son incorporados al huevo durante la oogénesis. Forman gradientes anteriores y posteriores de represión complementaria. **B.** Genes *gap*. Estos son activados por los genes maternos y se expresan en dominios solapados, donde ocurren eventos de activación y represión mutua. Son los que determinan la primera regionalización del cuerpo. **C.** Genes *pair-rule*. Estos se expresan en siete bandas a lo largo del eje antero-posterior del embrión, ocupando los futuros segmentos pares o impares, de ahí su nombre. **D.** Genes de polaridad de segmentos. Se expresan en 14 bandas, una para cada segmento corporal, ocupando la mitad anterior o posterior de los mismos.

## 1.4 Objetivos

Actualmente, es posible realizar un análisis completo y la caracterización de los procesos fisiológicos y moleculares que se llevan a cabo en un gran número de organismos gracias a las tecnologías desarrolladas en genómica. Estas tecnologías nos ofrecen la posibilidad de registrar en forma simultánea los niveles de expresión de una gran variedad de genes, bajo ciertas condiciones, estímulos y en el transcurso del desarrollo embrionario de los individuos. Esta información hace posible analizar el funcionamiento de las redes de regulación génica que intervienen en un determinado proceso. Por otro lado, un sinnúmero de fenómenos biológicos presentan una compleja relación temporal y espacial, generando patrones espacio-temporales. Para conocer esto es necesario en primer lugar conocer los genes responsables de la determinación del patrón corporal de un organismo.

**Objetivo general:** generar información genómica de *Rhodnius prolixus* y describir y caracterizar genes que intervienen durante el proceso de segmentación del embrión.

**Objetivos específicos:**

- 1- Generación de genotecas normalizadas de ESTs de *R. prolixus*,
- 2- Identificación de genes involucrados en los procesos de segmentación antero-posterior,
- 3- Determinación de la estructura del gen *in-silico* a partir de la secuencia genómica,
- 4- Análisis funcional de genes de segmentación.

## 2 Materiales y Métodos

### 2.1 Colonia de *R. prolixus*

Los ejemplares de *R. prolixus* utilizados durante esta tesis fueron obtenidos a partir de una colonia preestablecida en el insectario del CREG. Los insectos de la colonia se mantuvieron a una temperatura de 30°C con una humedad relativa del 50%. Los insectos se colocaron cada 15 días con gallinas para la ingesta de sangre. En estas condiciones, la duración de la embriogénesis es de 14±1 días y el ciclo de vida completo es de 3 meses.

### 2.2 Genoteca de ADNc normalizada de *R. prolixus*.

#### 2.2.1 Extracción de ARN total.

Para la extracción de ARN total, se utilizaron insectos adultos (tanto machos como hembras), larvas en los estadios I-V y huevos embrionados. En todos los casos se utilizaron individuos alimentados y no alimentados con el fin de cubrir la mayor cantidad de estados de expresión génica posibles. Los diferentes estadios a considerar se separaron en tres tratamientos diferentes para la extracción de ARN: 1) huevos embrionados a diferentes tiempos de desarrollo; 2) ninfas en estadios de I a V y 3) adultos de ambos sexos. Las distintas muestras de ARNm se sumaron en una sola muestra que representa la mayoría de los genes expresados por estos insectos. Los huevos fueron recolectados cada 5 días para cubrir de manera homogénea los distintos estadios embrionarios. La extracción de ARN total se realizó mediante la técnica de tiocianato de guanidinio-fenol ácido (Chomzynski y Sacchi, 1987) modificada, en la cual la extracción con fenol ácido es reemplazada por fenol neutro y precipitación con cloruro de litio.

Previamente a la lisis, los insectos fueron congelados en nitrógeno líquido y triturados mecánicamente para disgregar los tejidos blandos. A las muestras se les agregó buffer de lisis (tiocianato de guanidinio 4 M, citrato de sodio 30 mM, β-mercaptoetanol 30 mM, pH 7,0-7,5) y fueron lisadas en homogeneizador de vidrio. Posteriormente, el lisado se centrifugó por 5 min a 20.000 rpm en tubos corex en centrifuga refrigerada Centrikon T-124 (Kontron Instrument, rotor 8.24). El sobrenadante se pasó a un nuevo tubo, se adicionó un volumen igual de fenol, (pH 7,0-8,0), y se mezcló hasta homogeneizar. Luego, se le agregó 1/5 del volumen de cloroformo-alcohol isoamílico (24:1) y se mezcló (la muestra se agitó en 4 ciclos de 1 min, incubando el tubo en hielo entre paso y paso). Se centrifugó el tubo a 20.000 rpm en centrifuga por 30 min a 4°C. El sobrenadante se pasó a un nuevo tubo y este paso se repitió dos veces más para eliminar la mayor cantidad posible de proteínas de la muestra. Finalmente, se agregó 1

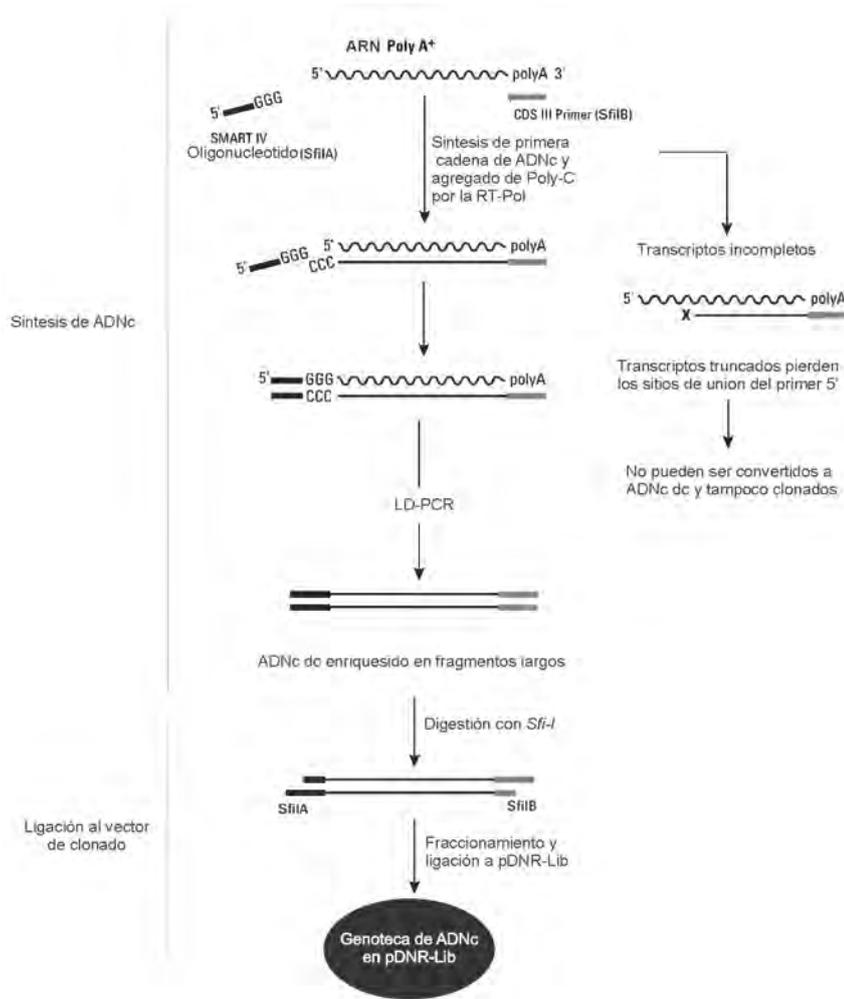
volumen de etanol 96% y se mezcló por inversión. Inmediatamente, se centrifugó a 20.000 rpm por 10 min. El precipitado se lavó con 500  $\mu$ l de etanol 80%, y posteriormente secado al aire para eliminar restos de alcohol. El precipitado se disolvió en 100  $\mu$ l de H<sub>2</sub>O<sub>bd</sub> estéril. La solución de ARN fue precipitada por agregado de un volumen igual de LiCl 12 M y se incubó a -20°C por 30 min. Luego, se centrifugó por 15 min a máxima velocidad en microcentrifuga Eppendorf (modelo *Centrifuge 5417C*). El precipitado se lavó con 500  $\mu$ l de etanol 80% para eliminar el exceso de sales y se seco como se indicó previamente. El ARN total se resuspendió en 40  $\mu$ l de agua H<sub>2</sub>O<sub>bd</sub> estéril. Para determinar la calidad del ARN purificado, se tomaron 2  $\mu$ l de la solución de ARN y se analizó por electroforesis en gel de agarosa 1%, en presencia de BrEt para marcar el ARN. Así mismo, el ARN total se cuantificó usando el espectrofotómetro GeneQuant II (Pharmacia Biotech) por medición de absorbancia a 260 nm.

### 2.2.2 Producción de ADNc.

Para la síntesis del ADNc se utilizó el sistema *SMART cDNA Library Construction Kit* (Clontech). Esta estrategia proporciona un método para generar genotecas de ADNc de alta calidad y enriquecida en fragmentos largos, a partir de nanogramos de ARN total o poli A+ (Figura 5). Esta estrategia, al igual que otras metodologías, utiliza transcriptasa reversa para transcribir la totalidad de los mensajeros de ARN a ADNc de simple cadena (ADNc ss) y LD PCR (*Long Distance PCR*) para obtener ADNc doble cadena enriquecida en fragmentos largos.

Para la construcción de la genoteca se utilizó 1  $\mu$ g de ARN total, el cual sirvió de molde para la síntesis de ADNc ss con un cebador de tipo oligo dT (CDS III/3' PCR *Primer*). Éste tiene una secuencia que aumenta la especificidad de unión a ARNm (indicado abajo como N<sub>1</sub>N-), una serie de 30 dT que va a hibridar con la cola de poli-A de los mensajeros y un sitio *Sfi*I-A para direccionar el clonado. Así el primer resultante es:





**Figura 5:** Estrategia SMART para la síntesis de una genoteca de ESTs. El lado derecho del diagrama muestra el destino de los transcritos incompletos a causa de degradación de ARN o terminación temprana de la retro-transcripción por la RT. Esto hace que la genoteca este enriquecida en fragmentos largos.

Para la síntesis de la primera cadena de ADNc se procedió de la siguiente manera: 1  $\mu$ g de ARN total en 1  $\mu$ l de solución, 1  $\mu$ l de CDS III/3' primer (10  $\mu$ M), 1  $\mu$ l de Smart IV primer (10  $\mu$ M), se llevó a un volumen final de 5  $\mu$ l con agua H<sub>2</sub>O<sub>bd</sub>. Se incubó por 2 min a 72°C, y luego en hielo por otros 2 min. Luego, a la mezcla se le agregó 2  $\mu$ l de 5X First-Strand Buffer (250 mM Tris Ph 8.3, 30 mM MgCl<sub>2</sub>, 375 mM KCl), 1  $\mu$ l de DTT (20 mM), 1  $\mu$ l de dNTP (10 mM) y 1  $\mu$ l de Power Script Reverse Transcriptase. La mezcla se incubó a 42°C por 1,5 hs. De la reacción se tomó una fracción de 5  $\mu$ l y se analizó su calidad por electroforesis en gel de agarosa 1,5% en presencia de BrEt para la tinción del ADN.

La segunda cadena se generó por el método LD-PCR usando ADN polimerasa de *Escherichia coli*, el primer CDS III/3'PCR y un primer adaptador en el extremo 5' (5'PCR primer), el cual cuenta con un sitio de restricción para la enzima Sfi-B. La mezcla de reacción fue la siguiente: 2  $\mu$ l de ADNc sc, 10  $\mu$ l de 10X Advantage 2 PCR Buffer (400 mM Tricine-

KOH, pH 8,7; 150 mM AcKO; 35 mM Mg(OAc)<sub>2</sub>; 37.5 µg/ml BSA; 0,05 % Tween 20 y 0,05 % Nonidet-P40), 2 µl de dNTP (10 mM), 2 µl 5'PCR *primer* (10 µM), 2 µl *primer* CDS III/3'PCR (10 µM), 2 µl 50X *Advantage 2 Polymerase mix*, y H<sub>2</sub>O<sub>bd</sub> estéril hasta un volumen final de 100 µl. El programa de PCR utilizado para la síntesis de ADNc fue: 1 ciclo de 20 seg 95°C → 20 ciclos (5 seg. 95°C → 6 min. 68°C). El producto de ADNc obtenido por esta estrategia se purificó mediante el sistema *QuiaQuick PCR Purification Kit* (QUIAGEN) y el ADNc se eluyó en 50 µl de agua H<sub>2</sub>O<sub>bd</sub> estéril. 5 µl del eluido se usaron para analizar la integridad del ADNc por electroforesis en gel de agarosa 1,5% con BrEt y para cuantificar por espectrometría. Debido a que la concentración del eluido es menor que la necesaria para la posterior normalización de la genoteca, la muestra se concentró por precipitación con acetato de sodio 3 M pH 5.2 y etanol 100%. El precipitado resultante se resuspendió en fracciones 10 µl c/u con una concentración aproximada de 130 ng/µl.

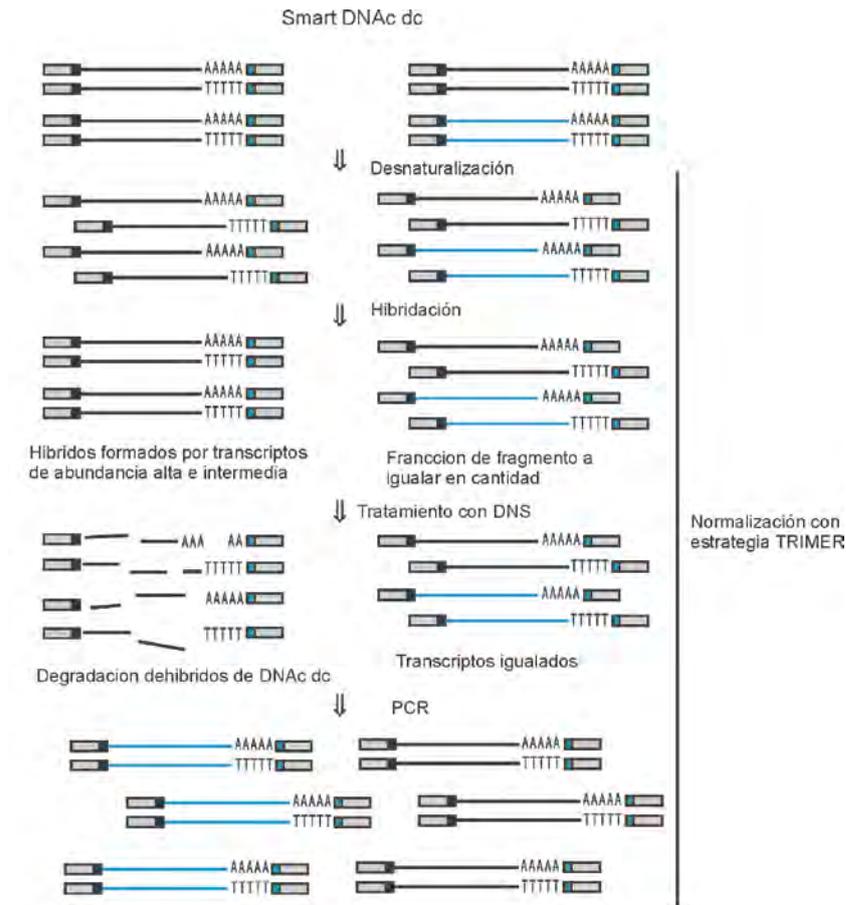
### 2.2.3 Normalización de la genoteca de ADNc.

Durante el proceso de normalización se equilibran los niveles relativos de los mensajeros de ARN dentro de la muestra. De esta forma, los mensajeros que menos se expresan aumentan en proporción y los que se expresan abundantemente disminuyen. Esto hace posible la identificación de la mayoría de los transcritos en una muestra por secuenciación al azar.

Existen varias metodologías para normalizar genotecas primarias, las cuales en su gran mayoría incluyen la tediosa utilización de columnas de hidroxipatita (Soares *et al.*, 1994; Bonaldo *et al.*, 1996). En esta tesis, se utilizó una nucleasa termoestable (nucleasa específica de ADN de doble cadena del cangrejo de Kamchatka o *Kamchatka Crab Duplex-Specific Nuclease*, KCDSN, Zhulidov *et al.*, 2004), y LD-PCR (*Long Distance-PCR*) de acuerdo al protocolo descrito por Anisimova y col (2006). Esta nucleasa termoestable es provista por el sistema de normalización *TRIMMER-DIRECT cDNA Normalization* (Evrogen).

Los ADNc completos generados a partir de los ARNm por medio del sistema *SMART*, proveen secuencias específicas en los extremos 5' y 3' del ADNc, permitiendo la amplificación del ADNc por PCR. La normalización se realiza, a diferencia de otros métodos, antes de la ligación al plásmido. En la figura 6 se muestra un esquema de la estrategia de normalización.

Siguiendo este esquema de trabajo, a 10 µl de ADNc (1,3 µg totales) se le agregó 4 µl del buffer de hibridación 4X (200 mM Hepes, pH 7,5, y 2 M NaCl), y 2 µl de agua H<sub>2</sub>O<sub>bd</sub> estéril. La mezcla se incubó por 2 min a 98°C para desnaturalizar el ADN y luego por 5 hs a 68°C para su renaturalización. Posteriormente, la mezcla se separó en 4 tubos conteniendo cada uno de estos 4 µl de mezcla, esto se hace porque cada tubo será tratado con diferentes concentraciones de nucleasa (1 U, 0,5 U, 0,25 U de la nucleasa, además de un control negativo sin enzima). El



**Figura 6:** Esquema de la metodología de normalización por DNS. Las líneas negras representan los transcritos abundantes dentro de la muestra, mientras que las azules los transcritos poco frecuentes. Los rectángulos grises corresponden a los *primers* provenientes de la síntesis de ADNc con la estrategia *SMART*.

tratamiento con la nucleasa específica de doble cadena se realizó así: a cada uno de los tratamientos arriba descritos se le agregó 5  $\mu$ l de *DNS Master Buffer* (100 mM Tris-HCl, pH8,0, 10 mM MgCl<sub>2</sub>, 2 mM DTT) y 1  $\mu$ l de la nucleasa (DNS) en la cantidad correspondiente. Una vez hecha la mezcla, se incubó a 68°C por 25 min. Todos aquellos ADNc que hayan formado dobles cadenas son cortados, mientras que los que no han hibridado, no son atacados por la nucleasa. El proceso de reasociación de las hebras de ADN tiene una cinética de segundo orden en la que los transcritos más abundantes se asocian más rápidamente que los de menor concentración. Por lo tanto, al cabo de un cierto tiempo la nucleasa cortará mayormente transcritos abundantes. Pasado el tiempo de hibridación, la nucleasa es inactivada por la adición de 10  $\mu$ l de *DNS Stop Solution* (5 mM EDTA) y puesto en hielo. Posteriormente, se le agregó 20  $\mu$ l de H<sub>2</sub>O<sub>bd</sub> estéril. De esta forma, cada tratamiento queda con una concentración aproximada de ADNc de 8 ng/ $\mu$ l. El próximo paso en el proceso de normalización, es la amplificación por PCR de aquellos ADNc no cortado por la nucleasa, usando *primers* específicos para los extremos 5' y 3' (el ADNc cortado no amplifica). La mezcla para la amplificación por PCR consiste en: 1  $\mu$ l de ADNc (8 ng/ $\mu$ l), 5  $\mu$ l de 10X *Advantage 2 PCR*

*Buffer*, 1 µl dNTP (10 mM), 1 µl 5' PCR *primer* (10 µM), 1 µl CDS III/3' PCR *primer* (10 µM) y 1 µl 50X *Advantage 2 Polimerase Mix*. La mezcla se llevó a un volumen final de 50 µl con agua H<sub>2</sub>O<sub>bd</sub> estéril. El programa de PCR utilizado fue: 1 ciclo de 20 seg 95°C → 20 ciclos de (5 seg. 95°C → 6 min. 68°C). Una vez terminada la amplificación, se realizó una electroforesis en gel de agarosa 1,5% con BrEt para determinar cuál fue la concentración óptima de nucleasa utilizada y, por ende, determinar en cuál de los tratamientos se produjo una mejor normalización.

#### **2.2.4 Fraccionamiento y clonado.**

El producto final de la reacción de amplificación y normalización de la genoteca fue separado por tamaño en cromatografía de exclusión usando columnas de *Chroma Spin-400* (Clontech). Cada fracción fue analizada por electroforesis en gel de agarosa en presencia de *Sybr Green* (MobiTec), colorante más sensible para la detección de ADN. Las fracciones conteniendo insertos entre los 0,5 y 5,0 Kpb fueron reunidas, precipitadas con etanol y glucógeno como agregante y el producto ligado en el vector pDNR-lib (Clontech). Las bacterias electrocompetentes utilizadas para la transformación fueron *E. coli* cepa DH5α (ElectroMax DH5α- Invitrogen). La transformación se realizó en cubas de electroporación de 0,2 cm de espacio entre placas (*Gene Pulser Cuvette*, BIO-RAD) utilizando el electroporador *Gene Pulser* II (BIO-RAD) a 1,8 kV y 200 mF. La mezcla de ligación transformada fue plaqueada en medio LB selectivo con cloranfenicol y los clones fueron almacenados a -80°C en 20% DMSO hasta su secuenciación.

#### **2.2.5 Análisis bioinformático de la calidad de la genoteca normalizada.**

La genoteca normalizada así preparada fue enviada a *Washington University Genome Center* para su secuenciamiento junto a otras 3 librerías como parte del proyecto de ESTs de *R. prolixus*. Todas las secuencias resultantes de este proyecto fueron depositadas en la base de datos de EST de NCBI. La librería normalizada de interés aquí se encuentra bajo el nombre de *Rhodnius prolixus*\_EST\_NADL\_RpcDNAorm ([http://www.ncbi.nlm.nih.gov/nucest/?term=Rhodnius\\_prolixus\\_EST\\_NADL\\_RpcDNAorm](http://www.ncbi.nlm.nih.gov/nucest/?term=Rhodnius_prolixus_EST_NADL_RpcDNAorm)).

Como se verá más adelante, el análisis de la NADL se basa en 1645 secuencias (860 secuencias ADNc 5' y 785 secuencias ADNc 3'). En un primer paso, un subconjunto pequeño de 243 secuencias fueron anotadas por homología con ortólogos por del consorcio de secuenciación del genoma. Varios pueden ser los motivos por los cuales muchas de las secuencias no fueron anotadas, pero las dos causas más frecuentes son la presencia de

contaminación con secuencias de origen vectorial o que la longitud de las secuencias era muy cortas para su identificación por homología. Estas secuencias requerirían un procesamiento previo para su posterior análisis. Con el fin de identificar la mayor cantidad posible de secuencias de la librería y constatar la correcta normalización, se desarrolló un proceso de análisis que permita eliminar las posibles contaminaciones y extender, sobre la secuencia genómica, el tamaño de las que eran muy pequeñas para encontrar homología. Como primer paso se hizo un BLAST local de las secuencias de la librería NADL no anotadas contra la última versión del genoma de *R. prolixus* disponible en NCBI (eso requiere armar previamente una base de datos BLAST a partir de los *contigs* del genoma en formato FASTA). La base de datos BLAST se arma usando el programa *formatdb* de la suite BLAST de NCBI. El comando de BLAST usado para buscar las secuencias de la librería NADL fue:

```
blastall -p blastn -d RhodniusGEN3.0 -i Rhodnius_NADL.fa -o Rhodnius_NADL.txt -v1 -b1 -e 2e-15
```

donde *RhodniusGEN3.0* identifica la base de datos BLAST del genoma de *R. prolixus*; *Rhodnius\_NADL.fa* es el archivo en formato FASTA de las secuencias de la librería NADL y *Rhodnius\_NADL.txt* es el archivo donde se guarda la salida generada por el programa *blastall*. Allí se guarda no más de un alineamiento por cada secuencia siempre que el mismo tenga un valor de expectación inferior al límite impuesto, en este caso fijamos el límite en  $2 \times 10^{-15}$ .

El segundo paso consistió en analizar los alineamientos escritos en *Rhodnius\_NADL.txt*. Para ello se desarrolló un *script* con el *software* Mathematica (Wolfram Research Inc.) que identifica las secuencias de la genoteca que tenían homología con el genoma y luego extrae del genoma los fragmentos de secuencias homólogo extendidos 250 pb a 5' y 250 pb a 3' siempre que ésto sea posible. De esta manera, usando el genoma como referencia, se eliminan las posibles contaminaciones y se extienden las secuencias cortas.

Posteriormente, se realizó BLASTX con cada una de las secuencias ampliadas con el fin de identificar nuevos genes dentro de la genoteca normalizada. Como línea de corte de valores de similitud se tomó un *e-value* igual o mayor a 0,5.

## 2.3 Búsqueda de genes de segmentación

### 2.3.1 Búsqueda bioinformática.

Para la búsqueda de genes de segmentación en *R. prolixus* se utilizó la base de datos de proteínas del *Genbank*. Una vez obtenidas las secuencias proteicas para cada gen de interés, se procedió a su comparación y alineamiento, para así determinar los dominios de mayor conservación de secuencia. Para ésto se utilizó el programa MegAlign del *software* Lasergene 7

(DNASTar). Los dominios conservados entre las diferentes especies de insectos se utilizaron para la búsqueda de dichos genes en el genoma de *R. prolixus*.

Para realizar la búsqueda bioinformática por similitud de secuencia de los diferentes genes de segmentación se utilizaron las bases de datos del genoma de *R. prolixus*. La búsqueda se realizó mediante el algoritmo TBLASTX y la matriz BLOSUM62. Dichas bases de datos han sido elaboradas en “*Washington University Genome Center*”, y pueden ser consultadas en [http://genome.wustl.edu/genomes/view/R.prolixus\\_prolixus](http://genome.wustl.edu/genomes/view/R.prolixus_prolixus). Una vez identificada la región codificante del gen de interés se procedió a la ampliación de la secuencia tanto para el extremo 5' como 3'. Para las predicciones bioinformáticas de estructura de genes y marcos de lectura abiertos (ORF) se utilizaron los programas Lasergene 7 (DNASTar) y la utilidad online del FGENESH (<http://www.softberry.com/berry.phtml?topic=fgenesh>). Los resultados obtenidos fueron corregidos manualmente.

### 2.3.2 Análisis de secuencias.

Las secuencias de los diferentes genes de segmentación putativos de *R. prolixus* identificados en el genoma se alinearon con ortólogos conocidos mediante el *software* Clustal W. Estos alineamientos fueron utilizados para realizar un análisis filogenético de cada uno de los genes a partir de los alineamientos de proteínas. El análisis filogenético se realizó mediante el algoritmo Bayesiano usando Bayes Online <http://www.phylogeny.fr> (Dereeper *et al.*, 2008; Huelsenbeck *et al.*, 2001). Los parámetros usados para el análisis fueron: *Number of substitution types*: 6 (GTR), Modelo de sustitución: Blosum62, Número de generaciones: 100.000 y *Sample a tree every*: 10 generaciones.

A partir de las secuencias genómicas de cada gen, obtenidas del genoma de *R. prolixus*, se identificaron las secuencias *upstream* respecto al origen de transcripción, con el fin de analizar la secuencias regulatorias. Habiendo identificado esta región, se analizó la presencia de sitios de unión para diferentes factores de transcripción mediante el uso de matrices de conservación de secuencia (*Position Weight Matrices* - PWMs). Se usaron las matrices descritas por Berman *et al.* (2002) para los sitios de unión de Hb, Cad, Bcd, Kr, y Kni, a las cuales se le sumaron las matrices generadas en esta tesis para Otd y Gt (Ver anexo 1). Para la búsqueda de las PWMs que se encuentren en la secuencia genómica de los diferentes genes se usó el *software* PATSER ([http://rsat.ulb.ac.be/rsat/patser\\_form.cgi](http://rsat.ulb.ac.be/rsat/patser_form.cgi)). Los parámetros de búsqueda usados fueron los siguientes: *Lower threshold estimation- weight store*: 5; *alphabet*: a:t 0.297 c:g 0.203 (Berman *et al.*, 2002). Como método de validación y búsqueda de sitios de unión que formen agrupamientos se usó el *software* STUBB (<http://stubb.rockefeller.edu/>).

### 2.3.3 Validación de las predicciones por PCR y clonado

Contando con la secuencia genómica de los diferentes genes de segmentación, se procedió a diseñar *primers* específicos para poder extender, por medio de PCR, la región codificante de cada uno de los genes. Se diseñaron dos poblaciones de *primers* diferentes: unos con parte del promotor viral T7 en el extremo 5' del *primer* y otros sin ese fragmento de secuencia. Los que tienen parte del promotor T7 se utilizarán posteriormente para realizar experimentos de hibridación *in situ* y de ARNi parental mediante transcripción *in vitro*. Del conjunto de *primers* diseñados los que se muestran a continuación corresponden a los genes con los que se realizaron análisis más exhaustivos:

GTFwT7	<u>CGACTCA CTATAGGGA</u> ACCACCGTGAAGAAGAGAC
GTRvT7	<u>CGACTCACTATAGGGA</u> AAGAAAAGCCGCTCGTATAGC
GTFw	GACCATTTAAAGCGTATCCAAAAG
GTRv	CTTTCCC AATAGGCGGCATC
KRFwT7	<u>CGACTCACTATAGGGT</u> GTGCAGGGTGTGTAACCATT
KRRvT7	<u>CGACTCACTATAGGG</u> AAGAAGGGTTTTTCACCACGATG
KRFw	AAAATCACGAACGAACCCACAC
KRRv	CGTTCTCCGGTGTGTACCCTTAG

Para todas las amplificaciones por PCR la mezcla de reacción utilizada fue: *Buffer* (1X), Cl<sub>2</sub>Mg (3 mM), DNTP (25 μM), *primer* Fw (250 μM), *primer* Rv (250 μM), Taq pol 1U, 20 ng del ADNc normalizado como molde y agua estéril y se realizó un gradiente de temperaturas de hibridación (T°a) para así determinar la T°a óptima. El programa de amplificación usado fue: un ciclo de 1 min 94°C → 40 ciclos (30 sec. 94°C → 30 sec. 50°C [Grad. 10°] → 1 min 74°C o 68°C) → 5 min 74°C o 68°C. Para cada par de *primers* fue necesaria una temperatura de hibridación (T°a) y enzima determinada, para así lograr los mejores resultados. De esta forma para *Rp-gt* la T°a utilizada fue 57°C y la enzima PFU (PBL), para *Rp-kr* 60°C y *Expand Long* (Roche), y para *Rp-h* 59°C y Taq pol (Haiway). Los programas de PCR para la posterior reamplificación de cada uno de los fragmentos fue una modificación del indicado antes, donde sólo se cambió la T°a y la temperatura de extensión de la enzima. El ciclador térmico utilizado para la PCR fue MBS 0.2G Hybaid (Thermo Scientific).

Los amplicones fueron posteriormente clonados en el vector pGEM-T Easy (Promega). Este vector tiene como característica que contiene una T colgante agregada en los extremos 3' del vector, esto genera una baja religación del vector disminuyendo el ruido de transformación. Para la ligación se necesita que el inserto sea un producto de PCR amplificado con Taq polimerasa, la cual como “error” de amplificación agrega en los extremos 3' del amplicón una A. Esto permite una ligación eficiente al vector. Para los fragmentos correspondientes a *Rp-gt* y

*Rp-Kr*, los cuales fueron amplificados con polimerasas que no agregan nucleótidos extra (*proofreading polymerases*), fue necesario previamente realizar la poliadenilación de los fragmentos, para así ligarlos al vector. La poliadenilación se realizó usando 1 U de Taq polimerasa en 1X Buffer, 3 mM Cl<sub>2</sub>Mg, 25 μM dATP, 20 ng de molde y agua estéril hasta un volumen de 10 μl, se incubó a 74°C por 2 hs y luego se procedió a la ligación. La relación molar entre inserto y vector fue de 3:1, para favorecer la incorporación de inserto. La mezcla de ligación se introdujo en bacterias competentes químicas *E. coli* de la cepa XL1-*Blue*. Los clones positivos se analizaron por digestión con enzimas de restricción y aquellas que dieron el patrón esperado fueron posteriormente secuenciadas (Macrogen Inc.).

## 2.4 Técnicas embriológicas

### 2.4.1 Colecta de embriones, disección y extracción de ovariolas.

Se colectaron embriones de diferentes estadios de desarrollo (24, 36, 48 y 72 hs PO). Para la disección de los embriones se utilizaron dos estrategias diferentes de decorionación y fijación. Algunos huevos fueron decorionados a mano luego de tres ciclos de incubación en calor y frío. Los huevos fueron colocados en criotubos e incubados durante 10 min en nitrógeno líquido, luego otros 10 min en baño de agua a 60°C. Luego de los ciclos de calor-frío, los huevos fueron fijados en 4% paraformaldehído en PBS (PFA 4%) por 1 h, decorionados a mano y posteriormente guardados en metanol 100%. Con este método, los embriones y el vitelo se separan del corion, facilitando la extracción manual del corion sin causar daños en el embrión. Por otro lado, parte de los embriones fueron decorionados y devitelinizados a mano en PFA 4%, incubados por 1 h y posteriormente deshidratados y guardados en metanol al 100%. Para embriones de pocas horas PO, en particular aquellos en estadio de blastodermo, la metodología utilizada fue diferente. Éstos fueron desoperculados manualmente y fijados en PFA 4% por 1 h, luego de lo cual fueron utilizados directamente en la hibridación *in situ* sin remover el corion. Esto se debe a que es imposible remover el corion en estos estadios sin destruir el blastodermo.

Las ovariolas fueron colectadas a partir de hembras alimentadas y en estado reproductivo de oogénesis activa. La fijación de las ovariolas se realizó como se detalla en Osborne y Dearden (2005).

### 2.4.2 Hidridación *in situ*

Para la determinación del dominio de expresión de cada uno de los genes, se diseñaron sondas específicas de ARN antisentido marcadas con digoxigenina para determinar la localización de los mensajeros tanto en embriones como en ovarios.

#### 2.4.2.1 Síntesis de las sondas de ARN.

Los amplicones correspondientes a cada gen producido por PCR, tiene en los extremos 5' y 3' parte de la secuencia del promotor del fago T7. Esto hace posible la transcripción *in vitro* de dichos fragmentos en un sentido, en otro o en ambos. Para el caso de producción de sondas de ARN se usaron los amplicones que tienen la secuencia del promotor de fago T7 sólo en el *primer* Rv, haciendo de esta forma una 5'RACE seguida de una extracción fenólica y precipitación alcohólica. La reacción de transcripción *in vitro* es: 1 µl de 10x buffer de transcripción, 1 µl de NTP *labeling mix* para ARN (Roche), 1 µl de inhibidor de ARNasa (20U), 0,5-1,0 µg de molde de ADN linearizado, 1µl de ARN polimerasa de T7 (20 U). La mezcla se incubó 2 hs a 37°C, se le agrega 15 µl H<sub>2</sub>O<sub>bd</sub> y 25 µl de 2x buffer carbonato (120 mM Na<sub>2</sub>CO<sub>3</sub>, 80 mM NaHCO<sub>3</sub>, pH 10,2) y se incubó a 70°C por 30 min para inactivar la enzima. Posteriormente se agrega 50 µl de solución de *Stop* (0,2 M NaAc, 1% HAc, pH 6,0), 8 µl de LiCl 5 M, 10 µl de ARNt (20 mg/ml) y 300 µl de EtOH 100% y se incubó a -20 °C ON. Luego, se centrifuga en microcentrifuga a 14.000 rpm por 20 min a 4°C y el precipitado se lava con EtOH 70% frío. Se centrifuga 5 min a 14.000 rpm, se elimina cualquier traza de alcohol de la muestra por secado y posteriormente se disuelve el precipitado en 150 µl de buffer de hibridación Hybe (50% formamida, SSC 5X, 0,1% Tween-20, 0,1 mg/ml ADN de testículo de salmón sonicado, 0,1 mg/ml ARNt de *Torula sp.* y 0,05 mg/ml Heparina).

#### 2.4.2.2 Tinción de embriones y ovariolas con sondas marcadas con Dig.

El protocolo de hibridación *in situ* utilizado aquí es una modificación del desarrollado para *D. melanogaster* por Martin Klingler (comunicación personal), al cual se le realizaron ciertas modificaciones para poder utilizarlo en *R. prolixus*. La optimización del protocolo fue llevada a cabo en colaboración con Lucia Pagola y Natalia Esponda-Beherens. El protocolo utilizado fue:

- 1) enjuagar los embriones en MetOH 100%, 50:50 MetOH:PBT;
- 2) post fijarlos en 50:50 PBT:Solución fijadora (50 mM EGTA-Na<sub>2</sub> y 10% PFA en 1x PBS) por un tiempo de 20 min en agitación;
- 3) enjuagar 2 veces con PBT, lavar 5 min con PBT en agitación y enjuagar nuevamente con PBT;
- 4) incubar por 1 h en una mezcla de detergente RIPA (150 mM NaCl, 1% NP-40, 0,5% desoxicolato de sodio, 0,1% SDS, 50 mM EDTA, 50 mM Tris-HCl, pH8,0);
- 5) idem paso 3;
- 6) luego incubar en Buffer de Hibridación (Hybe) sin sonda a 70°C por 30 min (esto se hace a fin de inactivar las fosfatasas endógenas);

- 7) pre-hibridar a 60°C x 30 min antes de la hibridación con sonda;
- 8) eliminar tanto Hybe como sea posible, y agregar la sonda (~1µl) diluida en 30 µl Hybe.
- 9) la hibridación se realiza por 24 hs a 60°C en baño húmedo;
- 10) una vez transcurrido el periodo de hibridación agregar 500 µl de la solución Hybe-B (50% formamida; 5x SSC y 0,1% Tween-20) a temperatura ambiente. Mantener los embriones a 60°C hasta que los embriones decantan al fondo del tubo; reemplazar con 500µl de Hybe-B, incubar a 60°C por 15 min. Descartar el Hybe-B y repetir el paso anterior;
- 11) eliminar todo el Hybe-B que sea posible y reemplazar por buffer Hybe, incubar ON a 60°C en baño húmedo;
- 12) enjuagar los embriones con Hybe-B dos veces y luego lavar durante 5 min en agitación con Hybe-B;
- 13) eliminar todo el Hybe-B como sea posible y hacer dos enjuagues con Hybe-C y luego un lavado de 5 min. Los lavados cortos en soluciones de baja salinidad ayudan a reducir el ruido de señal;
- 14) reemplazar el Hybe-C por PBT, realizar dos enjuagues y posteriormente dos lavados de 5 min en agitación, con PBT;
- 15) incubar en solución de Hibridación-Anticuerpo (2mg/ml albúmina de suero bovino (BSA) y 5% suero normal de cabra en PBTw) por 2 hs;
- 16) incubar ON a 4°C con 1 ml αDig-AP (1:1000), en agitación;
- 17) luego enjuagar en PBT dos veces y lavar 4 veces por 10 min en PBT;
- 18) revelar la fosfatasa alcalina. Eliminar el PBT y lavar los embriones con Buffer Alcalino;
- 19) dejar suficiente buffer alcalino como para cubrir los embriones y agregar 2 µl de NBT-BCIP (Roche). El tubo con los embriones en tinción se incuba en oscuridad para no generar fondo producto de la descomposición del reactivo por la luz. La tinción se realiza a temperatura ambiente y puede llevar de 30 min a varias horas;
- 20) detener la reacción con 2 enjuagues con PBT y luego un lavado de 20 min con PBT;
- 21) los embriones se deshidrataron en escalas seriadas de etanol (30%, 50%, 70% y 100%).

Luego del revelado de la sonda, los embriones se tiñeron con DAPI para la determinación de estructuras embrionarias. Los embriones en estadio de banda germinal disecados manualmente fueron montados en bálsamo de Canadá. Para la toma de imágenes de los diferentes embriones se utilizó un microscopio de fluorescencia o binocular estereoscópico (Leica DM 1000) y una cámara “CCD camera, Cool SNAP-Procf color” (Media Cybernetics).

### 2.4.3 ARNi parental.

La técnica de ARN de interferencia se basa en el mecanismo de defensa de las células de insectos contra virus de ARN. Cuando un ARN dc ingresa a la célula, se activa la maquinaria de reconocimiento y degradación del ARN iniciada por la proteína Dicer, la cual junto con otras proteínas involucradas en el procesamiento del mensajeros forman el complejo RISC. Ésto produce la fragmentación del ARN en fragmentos de aproximadamente 21 nt. En el caso que el ARN introducido sea endógeno, estos fragmentos de 21 nt (microARNs y siARNs) van a reclutar los mensajeros producidos por el organismo, los cuales seran enviados a degradación para así suprimir su expresión génica (Tomari y Zamore, 2005). Este principio biológico se utiliza para la generación de fenotipos por genética reversa (Bucher, 2002).

El ARN doble cadena (ARNdc) se sintetizo siguiendo la estrategia de transcripción *in vitro* de amplicones de PCR con parte de la secuencia del promotor de fagosT7 en los extremos 5' y 3' desarrollada por Dekanty (2008). El ARNdc fue cuantificado e inyectado en hembras de *R. prolixus* vírgenes (2,0 µg en un volumen no superior a 5 µl). La inyección se realizo entre el tercer y cuarto segmento abdominal con una jeringa Hamilton de 5 µl. Luego de la inyección se dejó que las hembras se recuperen del estrés por 3 días. Posteriormente, fueron alimentadas para estimular es proceso de oogénesis y colocadas junto con machos para el apareamiento. Los huevos fueron colectados a diferentes horas posteriores a la puesta para abarcar diferentes estadios de desarrollo. Los embriones fueron inmediatamente fijados como se dijo en la sección 2.4.1, o bien se los dejó desarrollar por completo para evaluar letalidad y para realizar preparados cuticulares.

Los preparados cuticulares se realizaron clarificando los embriones con 10% hidróxido de potasio por 16 hs a temperatura ambiente. Una vez clarificados, se deshidrataron sucesivamente en etanol y xilol y se montaron en bálsamo de Canadá. La adquisición de imágenes se realizó de dos formas diferentes: por un lado se tomaron imágenes en microscopio binocular de luz blanca trasmitida (Leica DM 1000) y por otro lado se aprovechó la autofluorescencia de la cutícula para tomar secciones ópticas por microscopia confocal. La autofluorescencia de la cutícula está en el rango de absorción de 520 a 660 nm, y para su detección se utilizó el microscopio confocal Zeiss 510-META y un laser de Argón para excitar la muestra a 488 nm. La emisión se registró usando un filtro de paso de banda de 505-535 nm. Se ensamblaron imágenes de proyecciones generadas a partir de 15 secciones de los embriones. Para cuantificar el grado de modificación de las estructuras, éstas se midieron usando una escala para después compararlos con los datos obtenidos de los insectos salvajes.

## 3 Genoteca de ESTs normalizada de *R. prolixus*.

### 3.1 Introducción

#### 3.1.1 Estrategias y métodos de secuenciación.

El genoma es el conjunto de información genética de un organismo, y la genómica estudia su estructura, función y evolución. Conceptos relacionados incluyen el transcriptoma (conjunto de transcritos presentes en un determinado estado celular), y el proteoma (el conjunto de las proteínas presentes en un determinado estado celular). Este énfasis en el estudio simultáneo del “todo” marca una nueva tendencia en biología. En este tipo de estudios existen dos perspectivas:

- Considerar el genoma como la suma de las partes, realizando avances desde la descripción parcial del genoma hasta llegar a un conocimiento total, ensamblando todos los fragmentos que lo constituyen.
- Considerar al genoma como una unidad. De esta manera la genómica procura revelar y comprender nuevas propiedades que emergen a este nivel superior de análisis.

La secuenciación de un genoma completo es un objetivo perseguido por la genómica. Pero la secuenciación completa no tiene sentido sin interpretar esa información, lo cual se denomina anotación del genoma. Durante la anotación se establece la presencia de los genes, sus intrones, exones, promotores, secuencias reguladoras, secuencias aislantes, transposones, telómeros, centrómeros y regiones sin información aparente. Ésta es una tarea que sólo ha sido completada en pocas especies e implicó el esfuerzo cooperativo entre laboratorios.

La generación de genotecas puede darse de dos formas diferentes: una es secuenciando directamente el ADN genómico del organismo (genotecas y mapas físicos), previo clonado de los fragmentos, otra es a partir de la secuenciación parcial del ADNc que representa los genes que se expresan en el organismo (EST's – *expressed sequence tags*), lo cual es muy útil para el mapeo de cada gen en el genoma.

- **Genotecas y mapas físicos del genoma**

Un mapa físico es el conjunto de elementos genéticos contiguos en el cromosoma (solapamiento de fragmentos de ADN, o clones) que cubren parcial o totalmente el genoma. La clave para la construcción de los mapas físicos ha sido la capacidad de clonar grandes fragmentos de ADN en cósmidos, YAC (*Yeast Artificial Chromosome*), BAC (*Bacterial*

*Artificial Chromosome*) y PAC (*Artificial Chromosome P1*), así como también una eficiente determinación del solapamiento de muchos clones. Los mapas físicos pueden ser usados como el sustrato para el secuenciamiento de la totalidad del genoma o para aislar rápidamente alguna secuencia deseada para futuras manipulaciones.

- **Proyectos de ESTs, secuenciación del genoma expresado**

Una base de datos de ESTs es un conjunto de secuencias correspondientes a los extremos 5' o 3' de clones de ADNc elegidos al azar de una genoteca. Los ESTs representan, por lo tanto, ARNm. Estas genotecas abarcan un estado particular del transcriptoma en un tejido u organismo en un momento fisiológico o estadio del ciclo de vida determinado. Las genotecas pueden ser directas, en las cuales la repetición de clones refleja la abundancia natural de los mensajeros de ARN correspondientes, o normalizadas, que son aquellas en las que se reduce la cantidad de mensajeros abundantes, incrementando la probabilidad de seleccionar al azar un clon raro (Bonaldo *et al.*, 1996). Las secuencias de ESTs son habitualmente depositadas en una base de datos (dbEST) con la correspondiente información sobre la especie y tejido de origen pero sin información adicional, es decir, sin indicar cuál es la posible estructura del gen, ya que en la mayoría de los casos los EST representan sólo fracciones de los UTRs u ORFs de los ARNm. Esto hace que también sea complejo determinar su identidad por homología.

- **Secuenciación masiva**

Hace pocos años comenzó la secuenciación de genomas utilizando secuenciadores de nueva generación (“*next generation sequencing*”). Los dos sistemas de secuenciación masiva más difundidos son: la pirosecuenciación (454-Roche) y la secuenciación por síntesis (Illumina-Solid AP) (Revisado en Metzker, 2010). El 454 se basa en la estrategia de secuenciación desarrollada por Ronaghi y col (1998), donde los nucleótidos incorporados durante un ciclo de amplificación del ADN se detectan por emisión de luz. El ADN genómico se rompe en fragmentos de 300 a 500 pb. A estos se ligan los adaptadores A y B en ambos extremos. Estos adaptadores proporcionan las secuencias de hibridación para la posterior amplificación y secuenciación de los fragmentos de la genoteca. El adaptador B está biotinilado en su extremo 5' lo cual permite la inmovilización de los fragmentos a esferas recubiertas de estreptavidina. Cada esfera contiene una única molécula de ADN. Las esferas se emulsionan con reactivos de amplificación por PCR en una suspensión de agua y aceite (emPCR) para así amplificar cada fragmento clonalmente. La secuenciación consta de una amplificación del ADN donde, además de polimerasa, la mezcla de reacción contiene nucleótidos acoplados a un fluoroforo, luciferasa y sulfurilasa. Así, durante la incorporación de nucleótidos al fragmento, el pirofosfato

inorgánico (PPi) liberado genera la degradación de la luciferasa y la emisión de luz. La luz es detectada por una cámara CCD de forma tal que se genera un espectro de picos en función de los nucleótidos incorporados.

El otro de los secuenciadores de nueva generación ampliamente difundido es síntesis directa por PCR en puente (Bridge-PCR). En esta metodología, los fragmentos son ligados a un soporte que permite la amplificación de los mismos por PCR-puente, en el cual cada fragmento generado se liga nuevamente al soporte hasta alcanzar una amplificación clonal. Luego durante la secuenciación, al igual que la pirosecuenciación, la detección se produce por emisión lumínica, donde hay un fluoróforo para cada nucleótido, siendo estos detectados por una cámara CCD. Esta metodología genera secuencias más pequeñas que la pirosecuenciación, estos rondan entre 75 y 100 pb. De esta forma, las técnicas de secuenciación masiva tienen como desventaja la producción de secuencia cortas, lo cual requiere un gran trabajo bioinformático posterior para el ensamblado de un genoma; pero tiene como ventaja los tiempos cortos de secuenciación, mayor cobertura del genoma y, sobre todo, una disminución importante de los costos frente a las técnicas de secuenciación tradicional.

### **3.1.2 Búsqueda de similitud y homología de secuencia.**

Cuando se quiere determinar la identidad de una secuencia de ADN, una de las primeras preguntas es si hay en las bases de datos existentes alguna secuencia parecida. Existen grandes bases de datos de ADN y proteínas (Gene Bank, <http://www.ncbi.nlm.nih.gov/genbank/>; UniProtKB/Swiss-Prot, <http://www.uniprot.org> ) que ofrecen un vasto número de secuencias para comparar. Además, existen programas que permiten comparar rápidamente una secuencia dada con las existentes en bases de datos. La comparación de secuencias es el método más poderoso y confiable para establecer las primeras relaciones funcionales o evolutivas entre genes.

Es un error común hablar de homología cuando se analizan comparaciones de secuencias en una base de datos. Es común afirmar que dos secuencias tienen una homología del 50%, cuando la homología es todo o nada. En estos casos se debe hablar de similitud. El término homología implica una relación, y que dos secuencias sean similares (en nucleótidos o aminoácidos) no significa que ellas sean descendientes de un ancestro común. Por otra parte, dos proteínas homólogas pueden divergir al punto de no mostrar ninguna similitud estadísticamente significativa conservando, sin embargo, una estructura común. Las proteínas homólogas deben tener la misma estructura tridimensional, mientras que proteínas con similitud de secuencia, no son necesariamente homólogas. Es recomendable llamar a la búsqueda en bases de datos búsqueda de similitud. Por supuesto, un alto nivel de similitud es una indicación significativa de homología, pero no son llamadas secuencias homólogas, sino ortólogas,

solucionando de esta forma la controversia de si están o no ligadas evolutivamente. Los ortólogos pueden ser similares en toda su longitud o solo en la totalidad de un dominio funcional.

### 3.1.3 Métodos para la búsqueda de similitud: BLAST

Todos los métodos de búsqueda de similitud se basan en conceptos de alineación y distancia entre secuencias. Una secuencia similar es calculada desde una distancia medible (por ejemplo, el número de bases de ADN o aminoácidos que son diferentes entre dos secuencias), y esta distancia sólo puede ser medida entre secuencias alineadas. Así, la búsqueda de similitud se reduce a un proceso de alineamiento de una secuencia problema con alguna otra secuencia en la base de datos.

Las primeras herramientas de similitud, desarrolladas por Needleman y Wunch (1970), calculaban la similitud global en toda la longitud de las secuencias que estaban siendo comparadas. Pero este tipo de algoritmo es más aplicable a la construcción de árboles evolutivos que a la búsqueda de similitud en los enormes bancos de datos actuales, ya que no es sensible para secuencias altamente divergentes y es muy lento por implicar comparación de secuencias largas. Los algoritmos más utilizados hoy en día permiten focalizar en regiones más cortas de similitud para búsqueda de similitud local, como el basado en el método Smith-Waterman (1981). Estos son el *Basic Local Alignment Search Tool* (BLAST) (Altschul *et al.*, 1990) y el FASTA (Pearson, 1997). En esta tesis se ha utilizado exclusivamente el BLAST, que es el más difundido.

El BLAST es un algoritmo que usa una aproximación heurística (basada en un proceso de aproximación sucesiva) para realizar búsquedas rápidas de similitud. Esto puede resultar en alguna pérdida de rigor en la comparación y, en consecuencia, alguna similitud baja pero relevante puede ser omitida. Además puede ocurrir que el BLAST identifique secuencias similares a nuestra secuencia problema, aún cuando la similitud no sea biológicamente significativa. Es necesario, por lo tanto, prestar especial atención al significado estadístico de los resultados de similitud. El algoritmo BLAST busca automáticamente similitud entre secuencias de ADN usando tanto la secuencia problema que se ingresa en la base de datos como su secuencia complementaria.

Las bases de datos públicas, como las contenidas en el Centro Nacional de Información Biológica (NCBI) de los Institutos Nacionales de Salud de los Estados Unidos, permiten una búsqueda con BLAST en internet ([http:// www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). Este medio es muy rápido y consta de subprogramas que combinan la búsqueda de secuencias de proteínas y ADN

En el marco del proyecto de secuenciación de *R. prolixus*, se generó una genoteca normalizada ADNc a partir de EST's de *R. prolixus*, las cuales se han identificado usando

BLAST. Esta genoteca está destinada a rellenar los espacios sin lectura de la secuenciación genómica realizada por el *Washington University Genome Center*.

## 3.2 Resultados

### 3.2.1 Obtención de ARN total

El aspecto más importante en la preparación de una genoteca es la obtención del ARN. La calidad del ARN usado en la síntesis de ADNc afecta la calidad de la genoteca resultante. Se debe asegurar la representación de todos los transcritos que se hallan en un tejido en un determinado estado fisiológico de la célula, por lo que el ARN no debe estar degradado y debe ser biológicamente activo. De acuerdo a lo descrito en Materiales y Métodos, se obtuvieron aproximadamente 300 µg de ARN total. La calidad del ARN total se analizó por electroforesis en gel de agarosa al 1,5% (Figura 7 A). Se ve que el ARN no tiene signos de degradación, ya que la mayoría de los transcritos se distribuyen uniformemente desde los 500 pb hasta de 10 Kpb. Las bandas definidas más intensas que se observan, corresponden a los ARN ribosómicos 23S y 18S que migran en forma equivalente a fragmentos de ADN de 1 Kpb y 700 pb respectivamente. Otras bandas intensas que se observan corresponden a mensajeros muy abundantes.

### 3.2.2 Construcción de una genoteca de ADNc

El ADNc se sintetizó a partir de 1 µg de ARN total usando el método *SMART* (ver Materiales y Métodos). El análisis en gels de agarosa mostró que el rango de tamaño del ADNc sintetizado coincidió con el observado para el ARN, con lo que se concluye que se obtuvo una buena síntesis (Figura 7 B y C columna 1). Una vez sintetizado el ADNc, se procedió a la normalización.

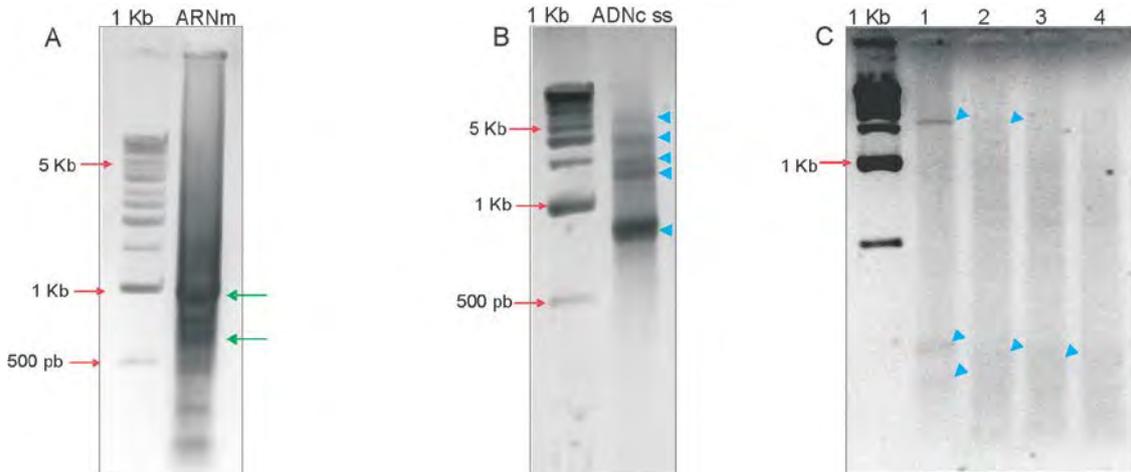
El ADNc fue dividido en cuatro alícuotas de 350 ng con el fin de realizar la titulación de la DSN y encontrar las condiciones óptimas de normalización. Luego de la digestión y posterior inactivación de la DSN, cada uno de los tratamientos fue sometido a 20 ciclos de amplificación por PCR utilizando los *primers* 5' y 3' ya descritos. Los resultados de la digestión con la DSN se muestran en la figura 7 C. Aquí, se ve claramente que en los tratamientos con la DSN los transcritos abundantes que se aprecian en el control empiezan a disminuir en cantidad a medida que se incrementa la concentración de enzima (Figura 3 C, marcados con flechas). Se puede ver que en el tratamiento con 1 U de DSN la proporción de mensajeros es homogénea, no observándose ningún mensajero que se encuentre representado más que otro. Ésto indica de modo grosero que hay un equilibrio entre especies, aunque no es un indicio de una buena normalización.

Una vez terminada la presunta normalización, se procedió a purificar el ADNc normalizado mediante cromatografía, para luego digerirlo con la endonucleasa de restricción *Sfi-I*.

Luego de la digestión, el ADNc fue fraccionado para disminuir la proporción de las moléculas de tamaño menor a 400 pb. Éstas representan una alta fracción molar y compiten eficientemente durante la ligación al vector de clonado frente a los fragmentos de tamaño mayor. Las 16 fracciones obtenidas fueron analizadas por electroforesis en gel de agarosa 1,5%, donde se sembró 1/10 del volumen total de cada una (Figura 8). De esta manera, se determinó que el ADNc de las fracciones 6, 7, 8, 9 y 10 están por encima de los 400 pb, por lo que estas fracciones fueron reunidas (aprox. 180 µl) y precipitadas para ser ligadas al vector pDNR-LIB.

La ligación del vector pDNR-LIB y los fragmentos de ADNc fueron introducidos en bacterias competentes *E. coli* DH5α, resultando en una genoteca primaria de 300.000 clones. Para analizar el porcentaje de inserción en el plásmido y el tamaño promedio de los insertos se prepararon plásmidos a partir de colonias elegidas al azar y se digirió con la enzima *Sfi-I*. Se comprobó que la mayoría de los insertos estaban en un rango de 100 a 500 pb. Este rango no es lo que se esperaba en la construcción, ya que el paso de fraccionamiento, debería haber disminuido los fragmentos inferiores a 400 pb en tamaño. Al repetir el chequeo, se obtuvo el mismo resultado, con un 80% de los fragmentos en el rango de tamaño antes mencionado. Una explicación posible es que restos de fragmentos pequeños que aún quedan en las fracciones seleccionadas, compiten con los fragmentos más grandes en el proceso de clonado en el vector pDNR-LIB, incorporándose con mayor facilidad. Esto, de acuerdo con los datos obtenidos por otros investigadores (G. Saldaña, comunicación personal), parece ser un problema implícito del vector pDNR-LIB. Sin embargo, debido al buen número de clones obtenidos, se decidió continuar con el proceso para determinar la calidad de la normalización.

La genoteca fue almacenada en bacterias en 20% DMSO a -80°C. Una parte de la genoteca fue enviada a secuenciar al *Washington Genome Center* para realizar una secuenciación de prueba.



**Figura 7:** Diferentes etapas en la producción de la genoteca normalizada. **A.** Se muestra el resultado de la extracción de ARN total, en donde las flechas verdes marcan los ARNr. **B.** Síntesis de la primera cadena de ADNc. **C.** Aquí se muestra la producción de la segunda hebra de ADN y el resultado del tratamiento con la DSN a diferentes concentraciones. 1: segunda cadena de ADNc, 2, 3, y 4 normalización de la genoteca a diferentes concentraciones de DSN (dilución de la enzima 0,25, 0,5 y 1 U respectivamente). Las flechas celestes marcan transcritos frecuentes, los cuales luego de la normalización disminuyen en cantidad.



**Figura 8.** Aquí se muestran, en un gel de agarosa 1,5% las fracciones de ADNc eluidas mediante las columnas Croma-Spin 400. Las calles 6-10, marcadas con el rectángulo verde fueron las utilizadas para la producción de la genoteca. La línea de puntos roja marca la altura en el gel correspondiente a las 500pb.

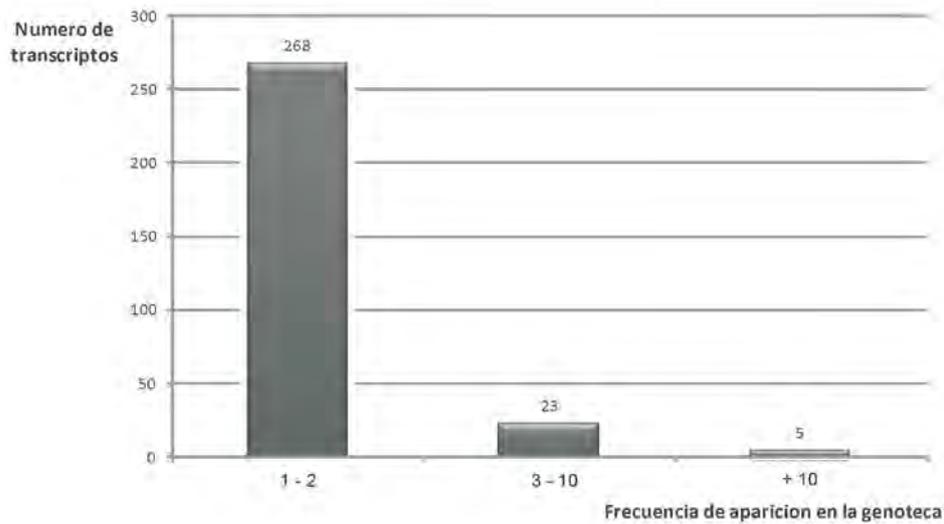
### 3.2.3 Análisis de secuencias

De la secuenciación de prueba realizada por el consorcio de secuenciación del genoma de *R. prolixus*, se obtuvieron 1645 secuencias de EST's con el nombre de "Rhodnius\_prolixus\_EST\_NADL RpcDNAnorm" (aquí las llamaremos NADL). De estas secuencias 860 corresponden a secuencias ADNc 5' y 785 a secuencias ADNc 3'. En un primer paso de búsqueda por similitud de secuencia mediante el algoritmo BLAST, un subconjunto pequeño de 243 secuencias fueron anotadas por homología con ortólogos (estas secuencias fueron anotadas en el GenBank por el consorcio). La anotación de las 1402 secuencias restantes no se realizó en esta instancia debido a que no se encontraba similitud significativa con ninguna de las secuencias de la base de datos. Esto puede ser debido a dos causas, la presencia de contaminación con secuencias de origen vectorial o a que las secuencias son cortas para su identificación por similitud.

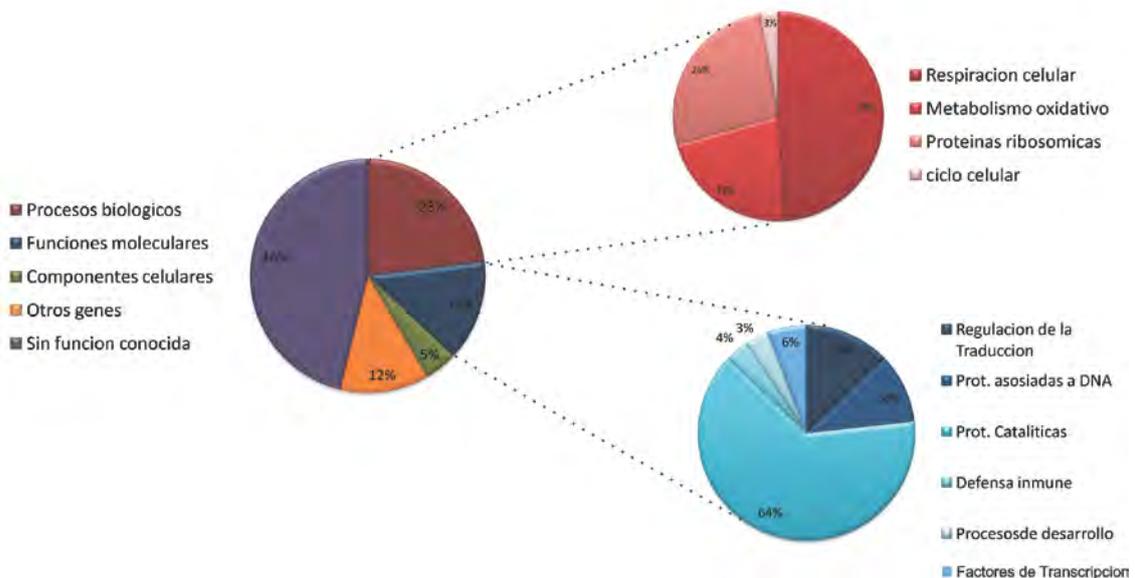
Pensando en esto último, se procedió a identificar la mayor cantidad posible de secuencias de la genoteca (en particular las secuencias que mostraron no tener similitud significativa) y chequear la correcta normalización. Para esto se desarrolló un *pipeline* para el procesamiento de las secuencias tal que elimine las posibles contaminaciones y extienda el tamaño de las mismas. Como primer paso se realizó un BLAST de las secuencias de la genoteca NADL no anotadas contra la última versión del genoma de *R. prolixus* disponible en NCBI. Como resultado de este análisis 1113 secuencias fueron encontradas en el genoma, mientras que 289 no dieron similitud (quizás contaminación vectorial). De las 1113 secuencias que dieron similitud con el genoma, 877 corresponden a regiones codificantes del genoma libre de intrones, mientras que 236 corresponden regiones con intrones y exones. Como se explicó en la sección de Materiales y Métodos, estas secuencias fueron ampliadas 250 pb hacia cada uno de los extremos 5' y 3' de la secuencia original sobre el genoma. Como resultado de la búsqueda por similitud de estas secuencias, se encontraron 373 nuevos genes con una alta probabilidad de homología con las proteínas de la base de datos.

En el Anexo 2 se muestran los EST's secuenciados de la genoteca normalizada de *R. prolixus*, acompañadas del número de identificación del clon donde fue identificado y la frecuencia de aparición de ese EST's en la genoteca. Como era de esperar, los transcritos más abundantes corresponden a genes del metabolismo celular, como ser NADH y citocromo oxidasa C. En el figura 9 se muestra una representación de la relación entre el número de transcritos en función de la frecuencia de aparición en la genoteca. Esto nos permite ver el grado de normalización que posee la genoteca. La mayoría de los transcritos aparecen representados sólo una o dos veces, mientras que muy pocos aparecen más de cinco veces (en su mayoría proteínas del metabolismo energético). Por otra parte, se observa que los ARNm correspondientes a proteínas ribosomales, que normalmente son muy abundantes, se hallan en

baja proporción. Esto nos indica que la genoteca fue eficientemente normalizada. Analizando los datos se puede ver que la proporción de EST's que fueron identificados con alguna proteína (37%) es menor frente al número de secuencias sin similitud alguna (63%). Como se puede ver, la proporción de estos últimos es muy elevada lo que se discute en la próxima sección. De los EST's con identidad, un 55% corresponden a proteínas con una función conocida en otros organismos, mientras que el 45% restante corresponde a proteínas que en otros organismos se desconoce su función (Figura 10).



**Figura 9:** Aquí se muestra la relación entre transcritos y frecuencia de aparición en la genoteca. Como se puede ver la gran mayoría de los transcritos aparecen sólo 1 o 2 veces en la genoteca. Esto nos da un indicio de que la genoteca está bien normalizada.



**Figura 10:** ESTs de *R. prolixus* agrupados por su función. Se las divide principalmente en genes involucrados en procesos biológicos, funciones moleculares y componentes celulares.

### 3.3 Discusión

#### 3.3.1 Representación de transcritos en la genoteca

El análisis transcriptómico presentado aquí, como dijimos anteriormente, es un análisis preliminar de la genoteca normalizada de *R. prolixus*. Estos datos serán utilizados para refinar la secuencia genómica total de *R. prolixus*, rellorando los agujeros en la lectura de los secuenciadores y así obteniendo un genoma más íntegro y sin errores. La genoteca normalizada se construyó por otro lado para la búsqueda de transcritos (y variantes de *splicing* de estos) de genes poco expresados. Un caso de estos pueden ser los genes de desarrollo, los cuales se encuentran de forma abundante en periodos restringidos del desarrollo o en pequeños grupos celulares, esto hace que se encuentren sub-representados en las genotecas no normalizadas y son poco frecuentes en proyectos de ESTs.

Una de las características de las genotecas no normalizadas es la presencia de un gran número de transcritos pertenecientes a metabolismo celular básico, como ser proteínas ribosomales y de metabolismo energético. Recientemente, Madeiros y col (2011) publicaron el transcriptoma de ovarios en *R. prolixus*. Esta genoteca consta de una gran cantidad de transcritos metabólicos y muy pocos transcritos embrionarios o involucrados en la vitelogénesis. Esto demuestra que en las genotecas no normalizada la presencia de transcritos abundantes interfiere en la identificación de genes raros. A diferencia de esto, en la genoteca mostrada en esta tesis, aparecen ESTs con una alta similitud con proteínas metabólicas, receptores de membrana, proteínas ribosomales, proteínas involucradas en la regulación de la traducción, factores de transcripción, proteínas involucradas en el desarrollo embrionario y vitelogénesis. Estos datos fueron agrupados en grupos funcionales de proteínas. Examinando esto, el hecho de que aparezcan proteínas involucradas en variados procesos fisiológicos, como por ejemplo vitelogénesis, desarrollo embrionario y diferentes proteínas del metabolismo básico, nos dice que todos los diferentes estadios del ciclo vital de *R. prolixus* considerados para la construcción de la genoteca están representados. Si a esto le sumamos el hecho de que la gran mayoría de los ESTs son poco frecuentes, apareciendo sólo una o dos veces en la genoteca (Figura 9), nos encontramos con una genoteca bien normalizada y representativa de las diferentes condiciones fisiológicas del insecto a lo largo de su ciclo vital.

#### 3.3.2 Importancia de la construcción de genotecas y la secuenciación de ESTs

La genómica de insectos cobra particular importancia cuando consideramos plagas y vectores de enfermedades. Esto es considerado clave para el descubrimiento de genes, la identificación de nuevos blancos para pesticidas o para comprender el proceso de transmisión de

patógenos. Actualmente el genoma de *R. prolixus* está en las fases finales de su anotación (D. Lawson, comunicación personal). Esto genera una gran cantidad de información que puede ser utilizada en diversos campos. Una de las metas del proyecto genoma es la búsqueda de genes que permitan el control de la enfermedad del Chagas. Estos genes pueden ser aquellos involucrados en la respuesta inmune del insecto frente a la infección del parásito (para así poder impedir su parasitación), así como también la búsqueda de genes a partir de los cuales se puedan diseñar pesticidas específicos menos dañinos con el medio ambiente que permitan controlar al vector de la enfermedad.

Otra ventaja de conocer el genoma de un organismo es la búsqueda de genes involucrados en diferentes procesos biológicos, permitiendo conocer la estructura génica, ya sea desde la secuencia codificante hasta las secuencias reguladoras; ubicación del gen en el cromosoma, alelos, etc. En el capítulo siguiente realizamos la búsqueda de diferentes genes involucrados en el proceso de segmentación durante el desarrollo embrionario de *R. prolixus* utilizando para ello esta genoteca.

Dado la calidad de la normalización, esta genoteca se encuentra en proceso de secuenciación completa por pirosecuenciación usando el ADNc generado y no clonado. Esto servirá para mejorar el borrador del genoma y completar su anotación. Los resultados de dicha secuenciación y el análisis de los datos no forman parte de esta tesis.

## 4 Genes de segmentación en *R. prolixus*

### 4.1 Introducción.

La embriogénesis en insectos muestra una gran variedad de mecanismos que culminan por formar siempre el patrón corporal típico de los insectos. Existen tres formas de desarrollo embrionario en función de cómo se forme la banda germinal: embriogénesis de banda germinal larga, intermedia y corta. La embriogénesis de banda germinal corta es la más difundida entre los insectos y, a la vez, la más primitiva (Davis y Patel, 2002; Liu y Kaufman, 2005). Aunque este modo de desarrollo embrionario sea el más difundido entre los insectos, la mayoría del conocimiento en genética y mecanismos moleculares en biología del desarrollo deriva de estudios realizados en *D. melanogaster*, un insecto derivado.

A partir de los trabajos fundacionales de Nüsslein-volhard y Wieschaus (1980) se expandió los límites del conocimiento en genética del desarrollo. En estos trabajos se realizó una mutagénesis en *D. melanogaster*, analizando el patrón segmental de los embriones afectados. Los genes que producían fenotipos mutantes fueron clasificados en función del efecto producido en embrión. Estos se clasificaron en: genes *gap*, en cuyos mutantes se afectan regiones completas del embrión; genes *pair-rule* (“de la regla par”), en donde los mutantes perdían los segmentos pares o impares, y los de polaridad de segmentos, en los cuales la mutación producía un cambio de polaridad de los segmentos. Del análisis en profundidad de cada mutante se estableció que un número pequeño de genes son necesarios para el establecimiento del patrón del embrión. (Jurgens *et al*, 1984; Wieschaus *et al*, 1984 y Nüsslein-volhard *et al*. 1984). Como consecuencia de esto, *D. melanogaster* se convirtió en el modelo por excelencia en genética del desarrollo.

Durante la determinación del eje antero-posterior de los insectos, en particular en *D. melanogaster*, actúa una amplia cascada génica que se organiza en forma jerárquica (Rivera-Pomar y Jackle, 1996). En esta red, los genes maternos forman gradientes en los polos anterior y posterior (recopilado en, Hulskamp y Tautz, 1991; St. Johnston y Nüsslein-Volhard, 1992 y Pankratz y Jackle, 1993; Rivera-Pomar *et al.*, 1995). Estos van a activar la expresión de los genes *gap*, que van a determinar la identidad de cada región embrionaria. Dentro de esta jerarquía los genes cefálicos *otd*, *btd*, *esm* y *fkf*, son los más anteriores y están únicamente involucrados en la formación de la cabeza; el resto de los genes *gap*, *hb*, *kni*, *gt*, *Kr* y *tll* están involucrados en la formación del resto del cuerpo. En *D. melanogaster* estos genes se expresan en dominios contiguos y solapados, en un entorno sincitial. Sin embargo el modo de desarrollo de *D. melanogaster* no es el más difundido en la naturaleza.

La aparición de nuevos modelos de estudio en biología del desarrollo como son *T. castaneum*, *O. fasciatus* y más recientemente *N. vitripennis* (otro insecto de banda germinal larga) mostró diferencias con todo lo conocido previamente en *D. melanogaster*. Una de las diferencias más importantes es que *bicoid*, como determinante anterior, no existe en otro grupo de insectos salvo dípteros superiores. En otras especies, esta función la realizan otros genes: *orthodenticle* y *hunchback* en *T. castaneum* (Schroder, 2003), *orthodenticle* en *N. vitripennis* (Brent *et al.*, 2007; Lynch *et al.*, 2006). Existe un gran contraste entre el proceso de segmentación que se ve en insectos con embriogénesis de banda germinal corta e intermedia y los de banda germinal larga. En insectos de banda germinal corta e intermedia, los segmentos anteriores (segmentos cefálicos y torácicos) se determinan durante el estadio de blastodermo en la parte posterior del huevo. Los segmentos posteriores devienen más tarde en el desarrollo. Este “crecimiento secundario” se da a partir de una población celular en el polo posterior del esbozo embrionario, denominada “zona de crecimiento”. Aquí, los segmentos son secuencialmente agregados y a la vez determinados. De esta forma, sólo los segmentos más anteriores se definen en un entorno sincitial, mientras que los segmentos que surgen a partir de la “zona de crecimiento” se definen en un entorno celularizado. Esto hace que la difusión de moléculas esté restringida y por lo tanto la interacción entre los diferentes genes de segmentación ocurra de una forma diferente a lo conocido en *D. melanogaster* (Davis y Patel, 2002). En esta tesis se abordan en profundidad dos genes *gap* involucrados en la formación de la parte media del embrión, *giant* y *Krüppel*.

El gen *gap giant* codifica un factor de transcripción con motivo “*leucine-zipper*”, el cual se encuentra muy conservado en insectos (Brent *et al.*, 2007; Bucher y Klingler, 2004; Capovilla *et al.*, 1992; Liu y Patel, 2010; Wilson *et al.*, 2010). Éste es regulado por otros genes *gap* y *pair-rule* por represión (Arnosti *et al.*, 1996; Eldon y Pirrotta, 1991; Kraut y Levine, 1991; Small *et al.*, 1992). Mutantes para *gt* en *D. melanogaster* se ven afectados en la formación de la cabeza y en la segmentación del abdomen. (Mohler *et al.*, 1989; Petschek *et al.*, 1987). En otros insectos de banda germinal larga, como *N. vitripennis* y *A. mellifera*, la expresión de *giant* es materna, y la falta de éste resulta en la pérdida de los segmentos cefálicos y torácicos, además de la fusión de segmentos abdominales (Brent *et al.*, 2007; Wilson *et al.*, 2010). En insectos de banda germinal corta, la expresión de *gt* es siempre cigótica, pero en *T. castaneum* los autores suponen que la expresión podría potencialmente ser materna, aunque no lo comprobaron experimentalmente (Bucher y Klingler, 2004). Aquí, la ausencia de expresión de *Tc-gt* resulta en cambios homeóticos del segmento maxilar y labial, los cuales toman identidad torácica. Al mismo tiempo, se ven cambios en la zona abdominal, donde se fusionan y pierden segmentos (Bucher y Klingler, 2004). De esta forma, *Tc-gt* no es netamente un gen *gap*, dado que no produce la pérdida de segmentos en la región correspondiente al dominio de expresión anterior. Recientemente, fue reportada la expresión de *gt* en *O. fasciatus*, en donde el

fenotipo *gt* es más parecido al visto en *D. melanogaster* que para otros insectos de banda germinal corta. De esta forma, se puede ver claramente que, con la diversidad de insectos y los diferentes caminos evolutivos que llevan a cada especie, esto podría representar un caso particular más que una regla.

Otro de los genes *gap* en el cual se focaliza en esta tesis es *Krüppel*. Este gen codifica para un factor de transcripción del tipo “*zinc-finger*” (Rosenberg, *et al.*, 1986). *Kr* fue descubierto durante los experimentos de búsqueda de genes de desarrollo en *D. melanogaster* por Nüsslein-volhard y col (1984). Los mutantes producto de la ausencia de *Kr* se caracterizaron por la ausencia de los segmentos torácicos y la ausencia y/o fusión de los primeros segmentos abdominales (Wieschaus *et al.*, 1984; Ingham *et al.*, 1986; Jackle *et al.*, 1985). *Kr* se expresa en la parte media del embrión, en la región correspondiente a los segmentos torácicos y en los primeros abdominales (Jackle *et al.*, 1985; Gaul *et al.*, 1987).

Además de *D. melanogaster*, *Krüppel* ha sido estudiado en pocos insectos como son: *G. bimaculatus*, *O. fasciatus*, *T. castaneum* y *A. mellifera* (Liu y Kaufman, 2004; Cerny *et al.*, 2005, Mito *et al.*, 2006 y Wilson *et al.*, 2010). En todos, el dominio de expresión se extiende a lo largo de los segmentos torácicos y en los primeros segmentos abdominales. Luego, cuando el embrión completó la formación, la expresión se hace más generalizada. En *O. fasciatus*, la ausencia de expresión de *Kr*, produce pérdida del 2<sup>do</sup> y 3<sup>er</sup> segmento torácico y los primeros segmentos abdominales (Liu y Kaufman, 2004). En *G. bimaculatus* tiene como consecuencia una amplia gama de fenotipos. Estas van desde la fusión de segmentos abdominales hasta la pérdida del 2<sup>do</sup> y 3<sup>er</sup> segmento torácico y varios de los segmentos abdominales, quedando solo un segmento torácico y unos pocos del abdomen (Mito *et al.*, 2006). Tanto para *Of-Kr* como para *Gm-Kr* el fenotipo mutante es muy similar a lo visto para *D. melanogaster*. En *T. castaneum* la ausencia de *Kr* no resulta en un clásico fenotipo *gap*, con pérdida de segmentos, sino que a la pérdida de algunos segmentos se le suman cambios homeóticos, donde los segmentos abdominales toman identidad torácica y el 2<sup>do</sup> segmento torácico toma identidad de segmento labial. De esta forma se puede ver claramente que la función de *Kr* varía en diferentes grupos de insectos.

La aparición de los insectos de banda germinal corta como modelos de estudio en genética del desarrollo, combinado con la genómica y las técnicas de genética reversa ayuda a comprender mejor los procesos evolutivos que determinan los diferentes modos de segmentación y a la biodiversidad de los insectos. A la fecha, no se han realizado estudios de genética del desarrollo en *R. prolixus*. En este contexto, reflatamos este modelo para identificar algunos de los factores que intervienen en el proceso de segmentación y caracterizar funcionalmente a los genes *gap* *Kruppel* y *giant*.

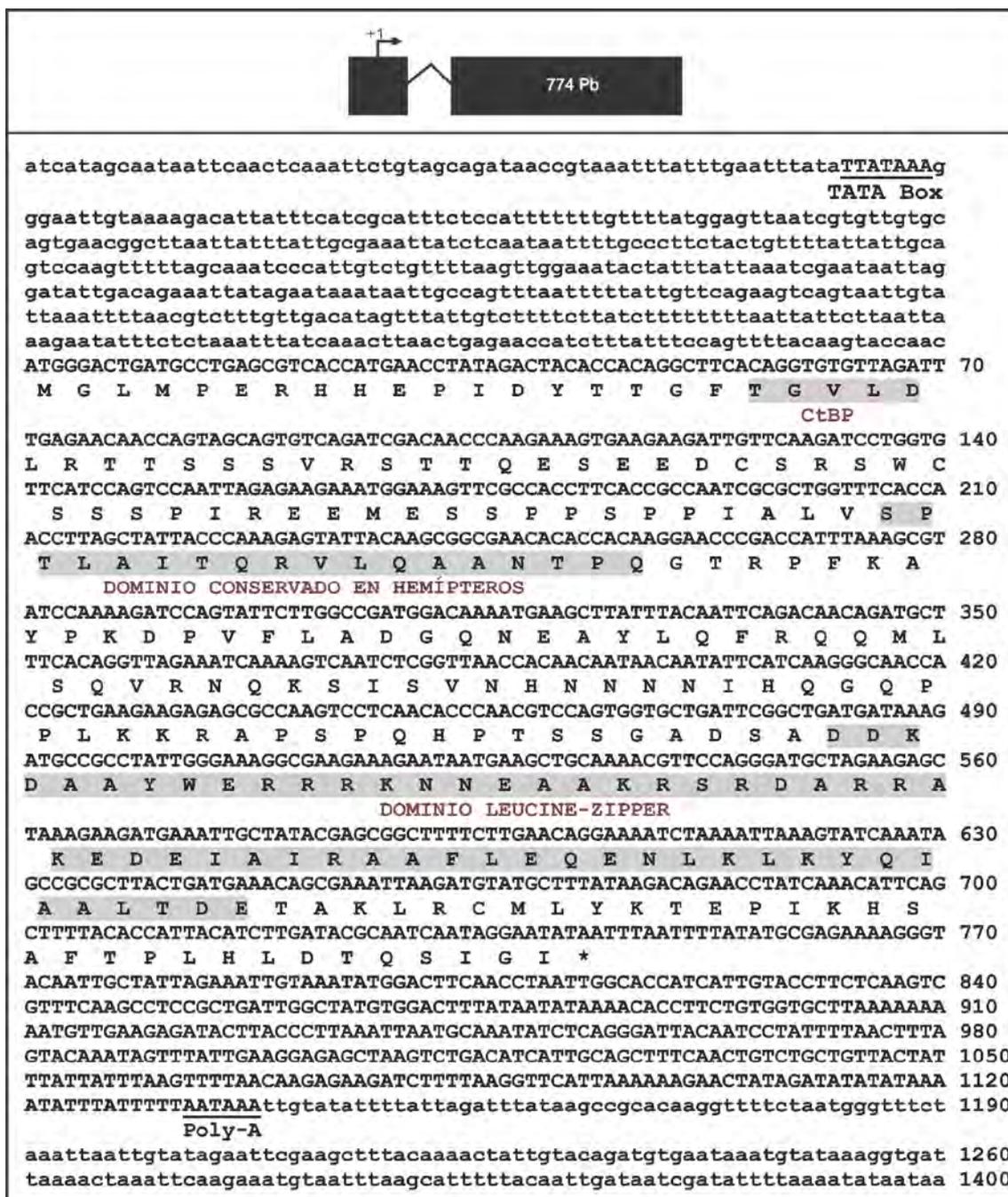
## 4.2 Identificación de genes de segmentación

Con el fin de identificar ortólogos para diversos genes que intervienen en la formación del eje antero-posterior de *R. prolixus* se realizó una búsqueda por homología sobre las secuencias disponibles de su genoma. Esta estrategia resultó en la identificación de algunos ortólogos de genes de segmentación, los cuales se citan a continuación: *Krüppel*, *giant*, *hunchback*, *knirps*, *tailless*, *orthodenticle*, *empty spiracles*, *forkhead*, *hairy*, *even-skipped*, *runt* y *engrailed*. Estos fragmentos se ensamblaron y fueron analizados ampliamente mediante diferentes herramientas bioinformática con el fin de caracterizar su organización génica, región codificante y análisis evolutivo comparándolos con ortólogos de otras especies de insectos. A continuación se describirán detalladamente cada uno de los genes analizados.

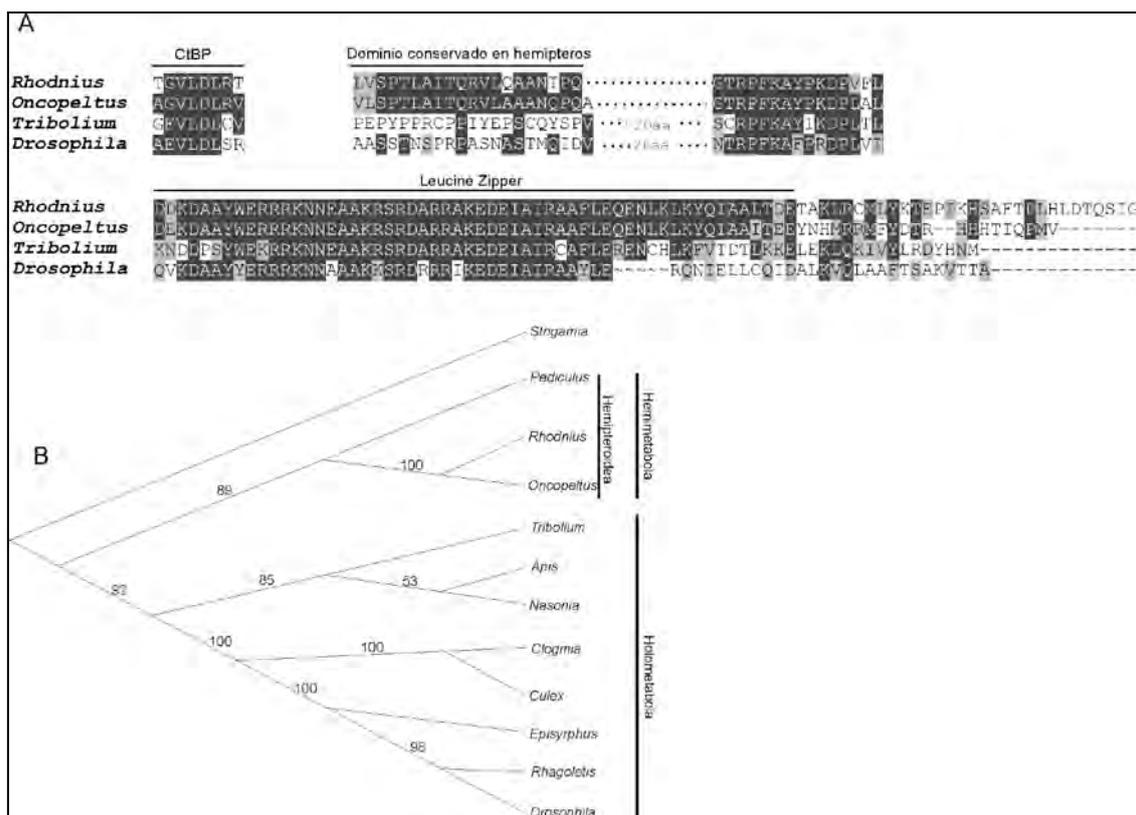
### 4.2.1 Genes *Gap*

#### 4.2.1.1 *giant*

Como resultado del ensamblado de los fragmentos del genoma se logró ensamblar una región genómica de 12,5 Kpb. Esta región contiene al gen *Rp-giant*, el cual posee un marco de lectura abierto (ORF) de 248 aminoácidos libre de intrones (Figura 11). En las secuencias genómicas disponibles no se detectaron evidencia de secuencias duplicadas ni parálogas, soportando el hecho de que éste es el único ortólogo de *gt* en *R. prolixus*. A unas 420 pb *upstream* al origen de traducción, se encuentra el sitio de inicio de la transcripción, el cual posee un motivo de secuencia tipo TATA-Box. *Upstream* al origen de este, se detectó una potencial zona reguladora, de la cual hablaremos más adelante. Considerando la región reguladora, el gen *giant* en *R. prolixus* se extiende unas 4.138 pb, desde la región reguladora aguas arriba, hasta la señal de poliadenilación en el 3' UTR. El ORF de *Rp-gt* posee una alta similitud de secuencia con ortólogos de *giant* conocidos para otras especies de insectos (GeneBank HQ853222). Este alto grado de similitud de secuencia, incluye el dominio básico *Leucine-Zipper* (BLZ), el dominio CtBP, y un dominio adicional, compartido con *O. fasciatus* que es un potencial sitio de fosforilación (Figura 12 A). El análisis evolutivo realizado mediante algoritmo Bayesiano, reveló que *Rp-gt* se ajusta muy bien dentro del árbol filogenético de los hexápodos, dividiendo a los insectos en dos grupos: los hemimetábolos y los holometábolos (Figura 12 B). Esto indica que la evolución de la secuencia codificante del gen acompañó la evolución del grupo.



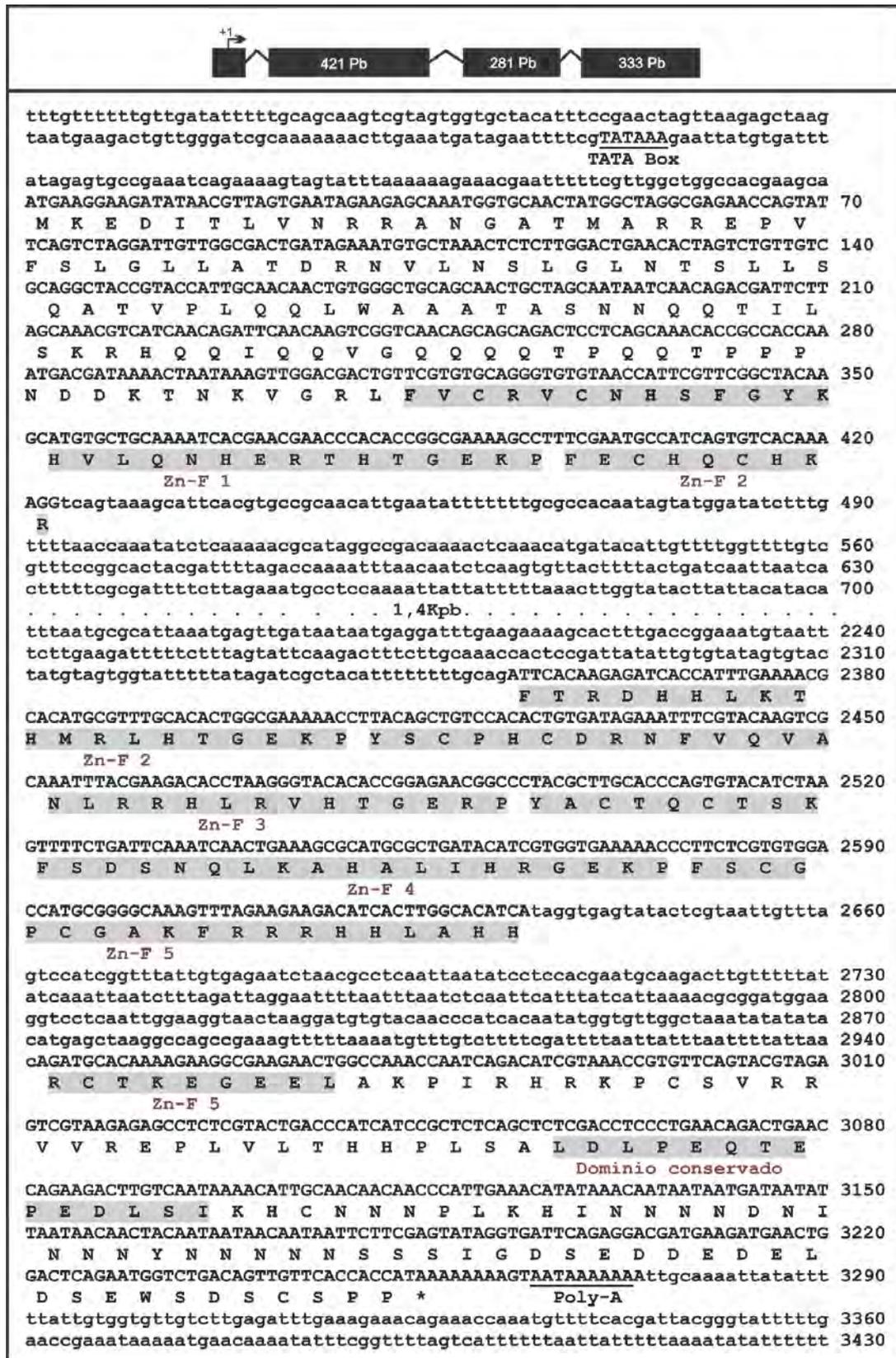
**Figura 11:** Gen *Rp-gt*. En la parte superior se muestra un esquema del gen. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-gt*. Los dominios proteicos conservados, CtBP, dominio en común de hemipteros y el dominio básico *Leucine-Zipper*, se marcan en gris. El supuesto TATA-Box y la señal de poliadenilación se muestran subrayados. En la secuencia nucleotídica, se marca en mayúscula la región exónica, y en minúscula la región no codificante.



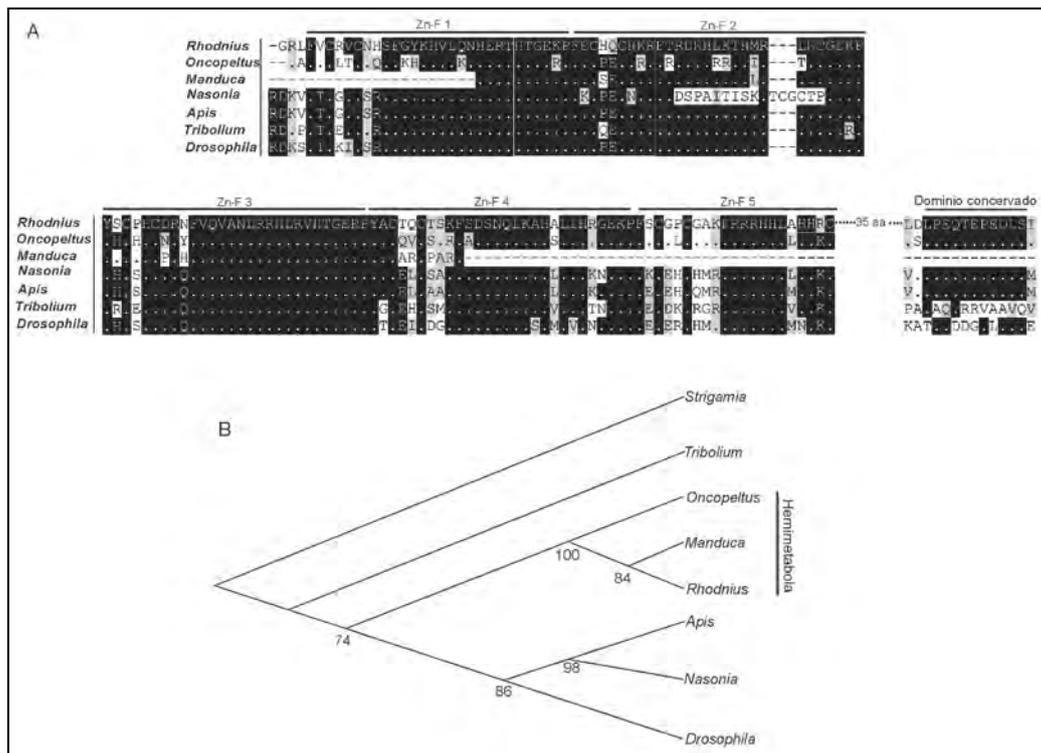
**Figura 12.** Análisis del gen *giant* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-gt* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total; son evidentes tres bloques; el CtBP, el dominio *Leucine-Zipper* y un dominio que sólo está conservado en hemipteros. **B.** Análisis filogenético de *Rp-gt*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

#### 4.2.1.2 *Krüppel*

A partir de las trazas del genoma de *R. prolixus*, y utilizando la misma estrategia de búsqueda que se usó para la identificación de *Rp-gt*, se logró identificar el ortólogo de *Krüppel*. Esto resultó en el ensamblado de una región genómica de 53,6 Kpb, conteniendo un ORF de 339 aminoácidos con alta similitud de secuencia con otros ortólogos del gen *Kr* (Gene Bank JN092576). La secuencia genómica conteniendo a *Rp-Kr* fue analizada bioinformáticamente con el fin de identificar su estructura génica. Como resultado de esto se pudo observar que el ORF de *Rp-Kr* se encuentra formado por 3 exones, uno de 421 pb, otro de 281 pb y el último de 281 pb (Figura 13). El intrón que separa el primer y segundo exón tiene una longitud de 1859 pb, mientras que el segundo 307 pb. De esta forma, el gen *Kr* en *R. prolixus* se extiende unas 3484 pb desde el origen de transcripción (TATA-Box) hasta la señal de Poly-A (Figura 13). La secuencia proteica fue comparada con la de ortólogos en otras especies de insectos y se reveló un alto grado de similitud de secuencia en la zona correspondiente a los ZnF1-5, los cuales son del tipo C2H2. Además, se observa la aparición de un dominio conservado en la mayoría de los insectos analizados, el cual no mostró tener ninguna función biológica conocida (Figura 14 A). El análisis evolutivo aplicando matemática Bayesiana, reveló que hay ciertas diferencias en la secuencia proteica que hacen que la secuencia codificante de *Kr* no se ajuste muy bien dentro del árbol filogenético de los hexápodos. Dentro del árbol, el grupo de los hemimetábolos está bien soportado, aunque *R. prolixus* y *O. fasciatus* no formen un grupo monofilético siendo ambos hemípteros (Figura 14 B).



**Figura 13:** El gen *Rp-Kr*. En la parte superior se muestra un esquema del gen, indicando la longitud de cada exón. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-Kr*. En gris se marcan los dominios proteicos conservados: Zn-F 1-5 y el dominio conservado hemípteros. El supuesto TATA-Box y la señal de poliadenilación se muestran subrayados. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.



**Figura 14.** Análisis del gen *Krüppel* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-Kr* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total; como se puede ver los cinco dominios Zn-F están altamente conservados. Además se muestra un dominio muy conservado en la mayoría de los taxones. **B.** Análisis filogenético de *Rp-Kr*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

#### 4.2.1.3 *hunchback*

A partir de las secuencias parciales producto del secuenciamiento del genoma de *R. prolixus* se ensambló un *contig* de alrededor de 16 Kpb que contiene la región codificante para el gen *hunchback*. El gen consta de 3 exones, uno de 66 pb, otro de 525 pb y el último y más grande de 2062 pb. En la figura 15 se muestra detalladamente la estructura de *Rp-hb*, acompañada de la secuencia proteica en el marco de lectura correspondiente. El ORF de *Rp-hb*, codifica para una proteína de 651 aa, la cual fue comparada con ortólogos de *hb* en otros insectos. *Rp-hb* conserva los seis motivos *zinc-finger*, típicos para este factor de transcripción, además del dominios *Box-C* y *Basic Box* encontrados en otros insectos (Figura 16 A). Como dato remarcable, surgió la aparición de dos dominios Zn-F adicionales a los conocidos para otros insectos, los cuales se nombran como Zn-F 1 y 2. Estos están altamente conservados en hemípteros, mostrando una alta similitud de secuencia en *R. prolixus* y *O. fasciatus*. Todos los *zinc-finger* son de tipo C2H2. En la figura 16 B se muestra el resultado del análisis filogenético, en el cual se ve que la secuencia proteica del gen *hb* no se acopla a la evolución del grupo, pero los dípteros están bien agrupados dentro del árbol. *O. fasciatus* y *R. prolixus* forman un grupo monofilético, quizás como resultado de que ambos comparten dos dominios Zn-F además de los seis que comparten con el resto de las especies.

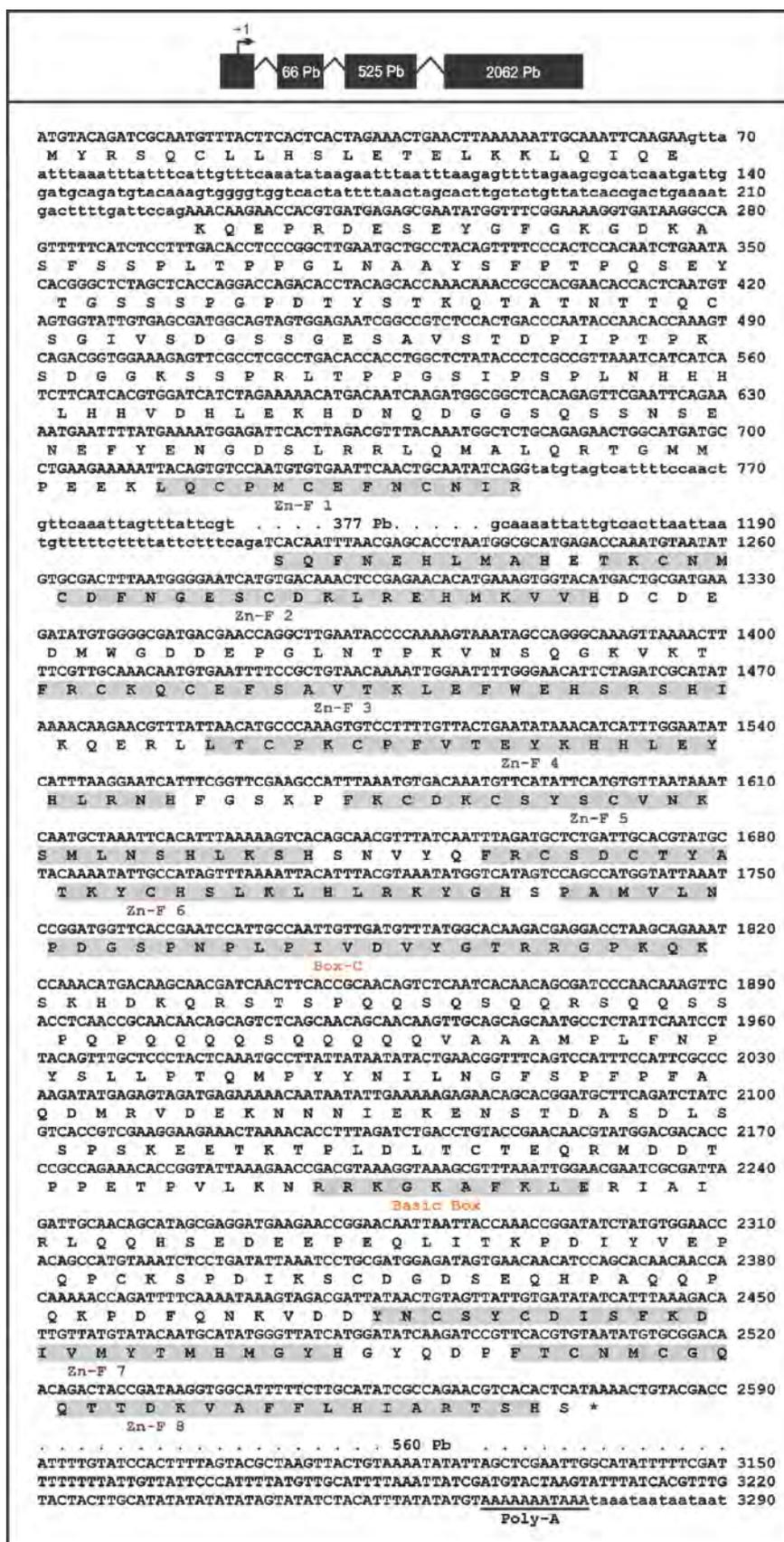
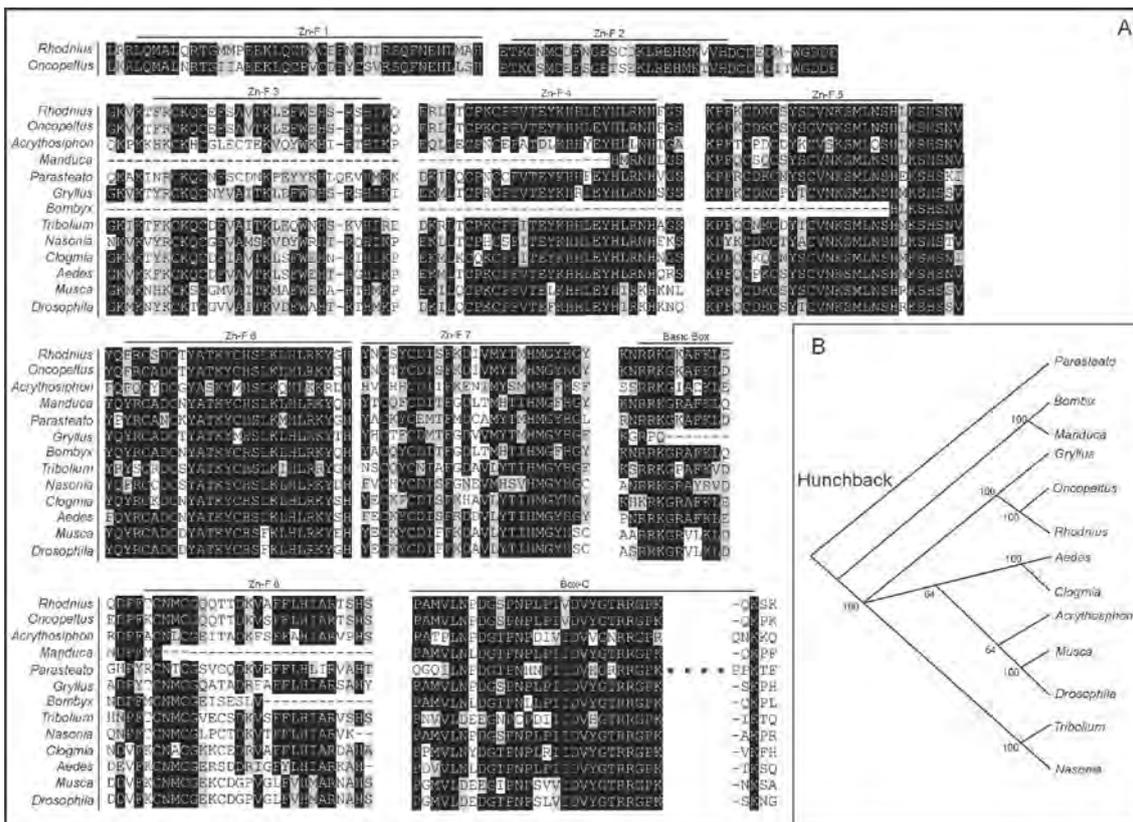


Figura 15: El gen *Rp-hb*. En la parte superior se muestra un esquema del gen. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica. En gris se marcan los dominios conservados: Zn-F 1-8, el dominio Box-C y el Basic Box. La señal de poliadenilación se muestra subrayada. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.

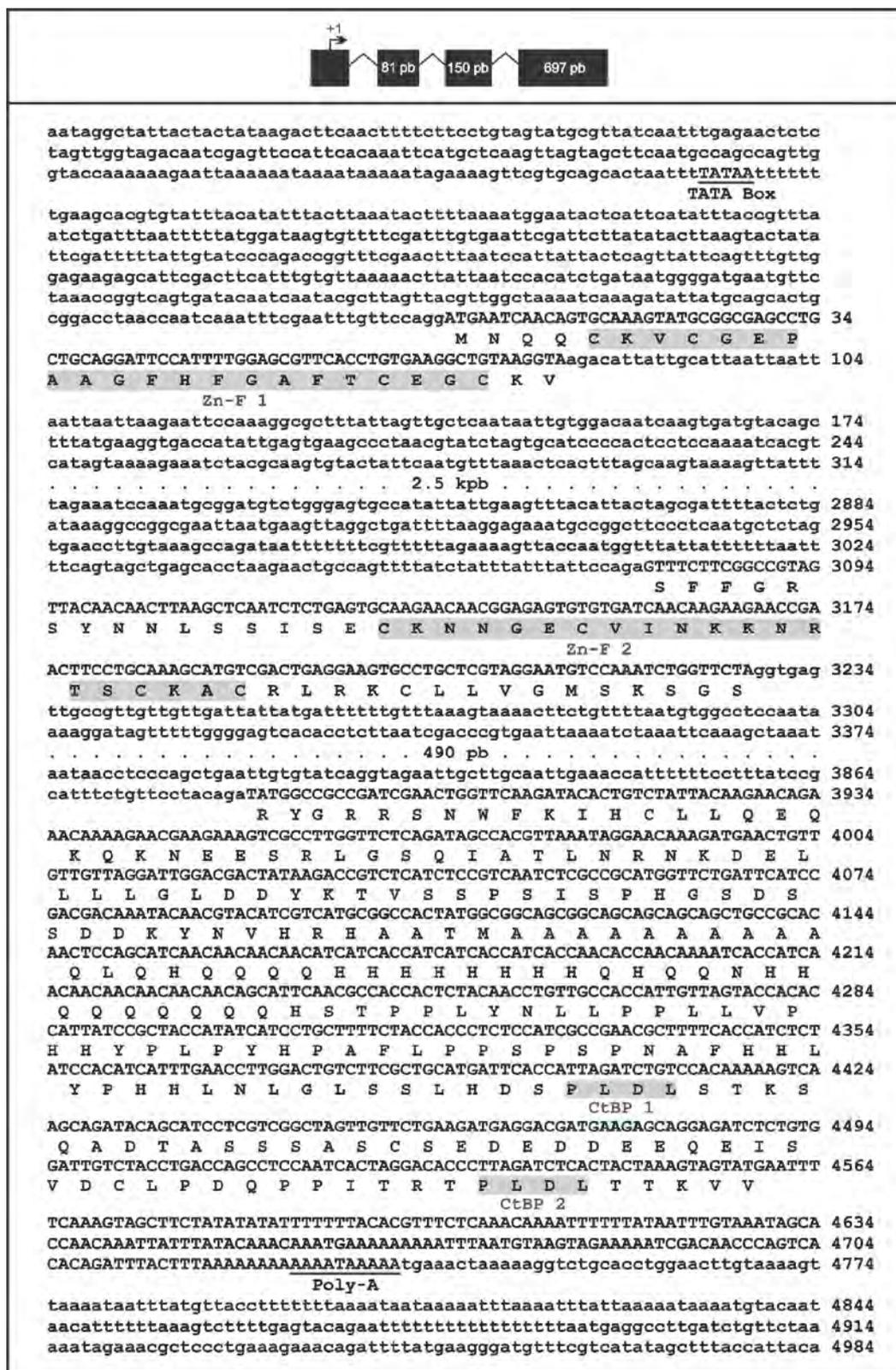


**Figura 16.** Análisis del gen *hunchback* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-hb* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total. Se muestran los ocho dominios Zn-F (de los cuales los primeros dos se encuentran solo en hemipteros), el *Basic Box* y el dominio *Box C*. **B.** Análisis filogenético de *Rp-hb*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

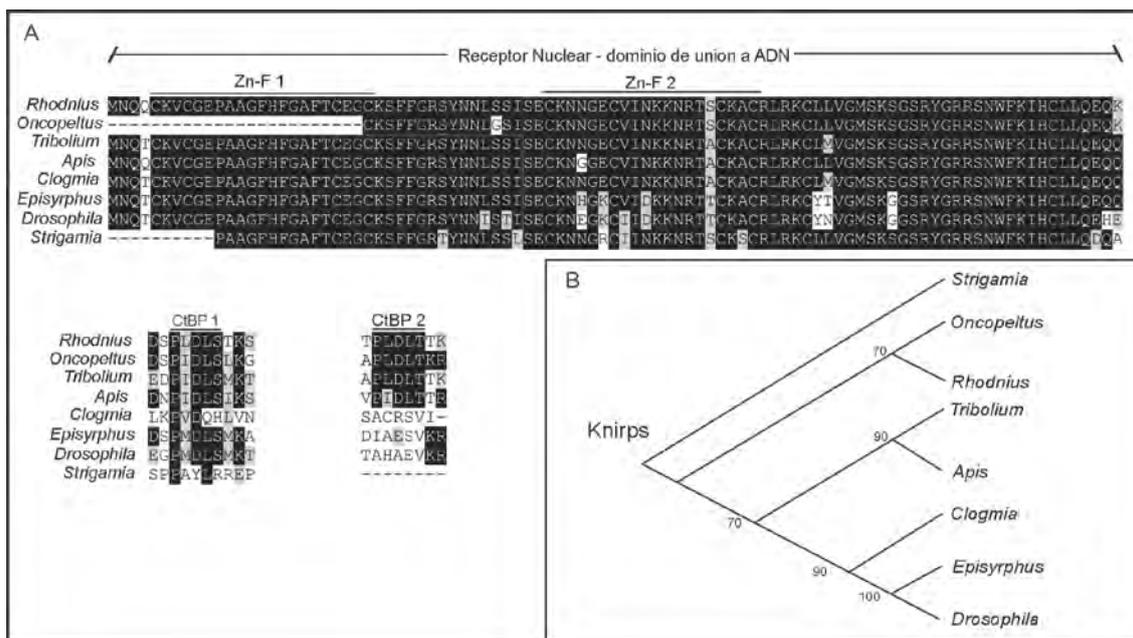
#### 4.2.1.4 *knirps*

Para la búsqueda de los restantes genes de segmentación se utilizó la última versión del genoma ensamblado de *R. prolixus*. Como resultado, se encontró que el *contig* 17922.18 contenía un marco de lectura abierto que codificaba para un ortólogo del factor de transcripción *knirps*. Este *contig* tiene una longitud de 57,5 Kpb, el cual contiene a *Rp-kni*. El gen está formado por 3 exones, el primero tiene una longitud de 81 pb, el segundo 150 pb y el tercero 697 pb. El gen *Rp-kni*, desde el supuesto origen de transcripción a la señal de Poly-A, tiene una longitud de 5,2 Kpb (Figura 17). El ORF de *Rp-kni* mostró un alto grado de similitud de secuencia con otros ortólogos del mismo gen en insectos. Estas similitudes se dan en el dominio de unión a ADN, la cual está formada por dos dominios *zinc-finger* de tipo C4, y 2 dominios CtBP en el extremo C-terminal de la proteína (Figura 18 A). El segundo dominio CtBP no se encuentra en *S. marítima* y tampoco en dípteros, lo cual podría indicar que este dominio apareció temprano en la evolución de los insectos y se perdió en el transcurso de la evolución de los dípteros. El análisis filogenético del gen *knirps* divide a los hemimetábolos de los holometábolos como grupos monofiléticos independientes, separando dentro de este último

grupo a Coleópteros e Himenópteros de los Dípteros (Figura 18 B). Para este análisis se excluyó del alineamiento a las especies *Pediculus humanus* y *Acrythosiphon pisum* debido a que dichas secuencias no tenían gran similitud de secuencia con el resto de los ortólogos, salvo en los dominios Zn-F. Al sacar ambas secuencias del análisis, el árbol filogenético de la evolución del gen *knirps* se ajusta mucho mejor a la evolución del clado.



**Figura 17:** El gen *Rp-kni*. En la parte superior se muestra un esquema del gen, indicando la longitud de cada exón. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-kni*. En gris se marcan los dominios proteicos conservados: Zn-F 1, Zn-F 2 y los dos dominios CtBP. El supuesto TATA-Box y la señal de poliadenilación se muestran subrayados. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.



**Figura 18.** Análisis del gen *knirps* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-kni* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total; son evidentes tres bloques, los dos Zn-F típicos de *knirps* y dos dominios CtBP, de los cuales el segundo no se encuentra en dípteros. **B.** Análisis filogenético de *Rp-kni*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

#### 4.2.1.5 *tailless*

Otro de los genes *gap* que se encontraron dentro de las secuencias ensambladas del genoma de *R. prolixus* fue un ortólogo del gen *tailless*. Este gen se localizó dentro del *contig* 17871.12, y se extiende unas 9,8 Kpb. *Rp-tll* posee un marco de lectura abierto compuesto por 8 exones que codifican para dos dominios Zn-F de unión a ADN en el extremo N-terminal y un dominio de unión a ligandos (normalmente clasificado como de tipo hormona esteroide) en el extremo C-terminal (Figura 19). *Rp-tll* tiene un alto grado de similitud de secuencia con ortólogos conocidos para otros insectos principalmente en la región de unión a ADN, la cual cuenta con dos Zn-F de tipo C4 altamente conservados (Figura 20 A). A partir del alineamiento de las secuencias de los ortólogos de *tll* en insectos se realizó un análisis filogenético mediante matemática Bayesiana para establecer como la evolución del gen acompañó a la evolución de los insectos. Como resultado, se vio que *R. prolixus* y *A. pisum* conforman un grupo monofilético, representando el orden Hemiptera. Al mismo tiempo todos los dípteros forman un grupo monofilético, así como también en el caso de *A. mellífera* y *Polyrhachis vicina*, los cuales conforman a los Himenóptera (Figura 20 B). Más allá de la importancia evolutiva de incluir en el análisis al ortólogo de *tll* en *O. fasciatus*, éste no fue incluido debido a que no se cuenta con el gen completo. Esto introduce ruido durante el procesamiento de los datos y distorsiona los resultados del análisis filogenético.

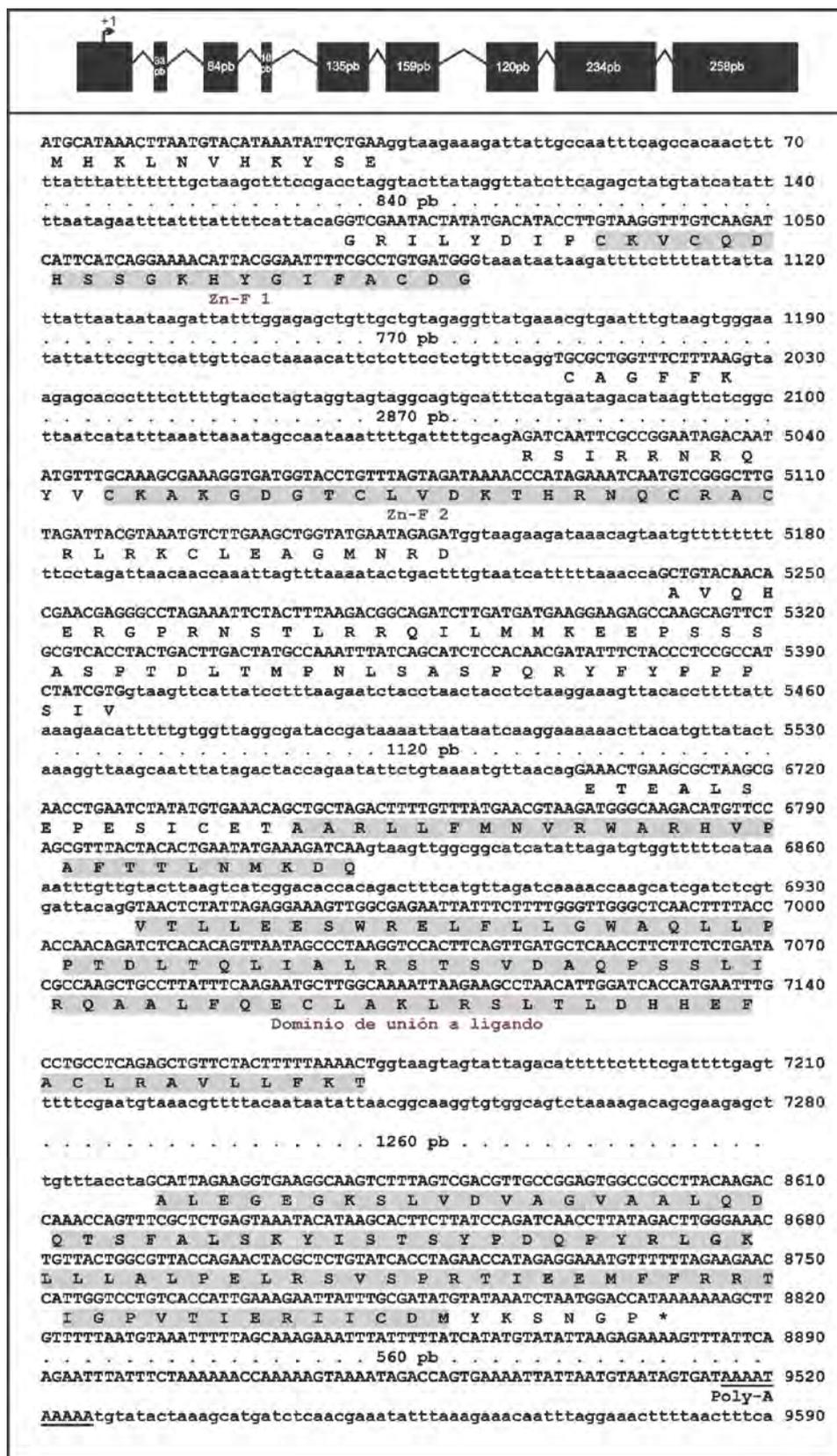
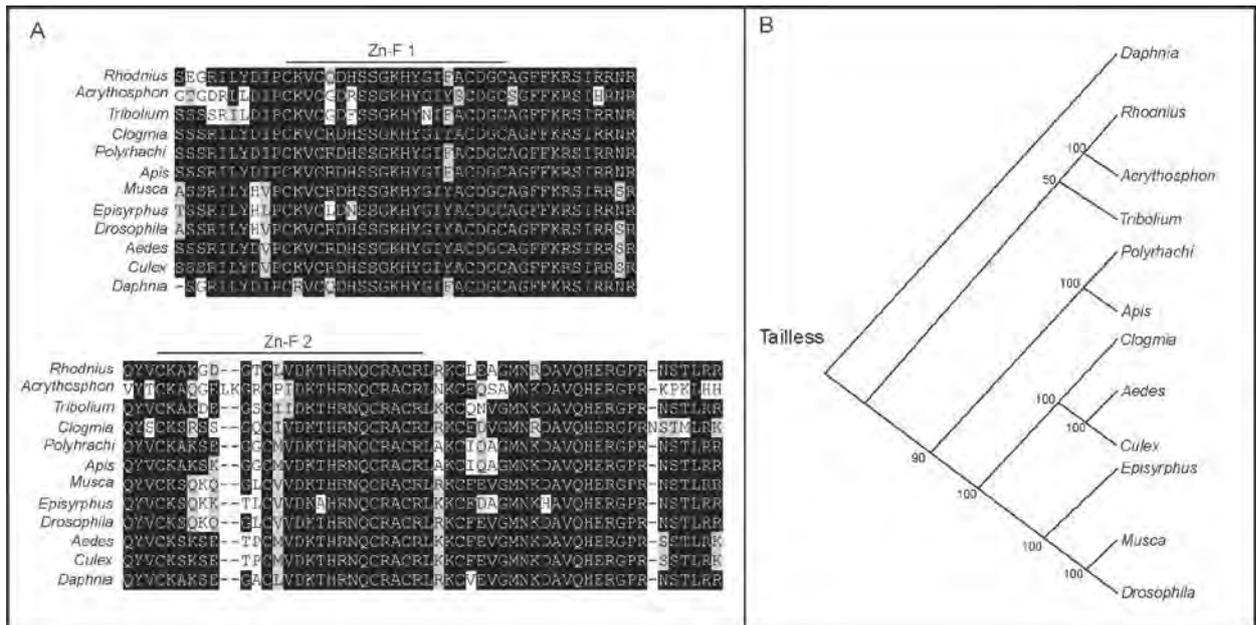


Figura 19: El gen *Rp-III*. En la parte superior se muestra un esquema del gen. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-III*. En gris se marcan los dominios proteicos conservados: Zn-F 1, Zn-F 2 y el dominio tipo esteroide de unión a ligandos. La señal de poliadenilación se muestra subrayada. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencia no codificantes.



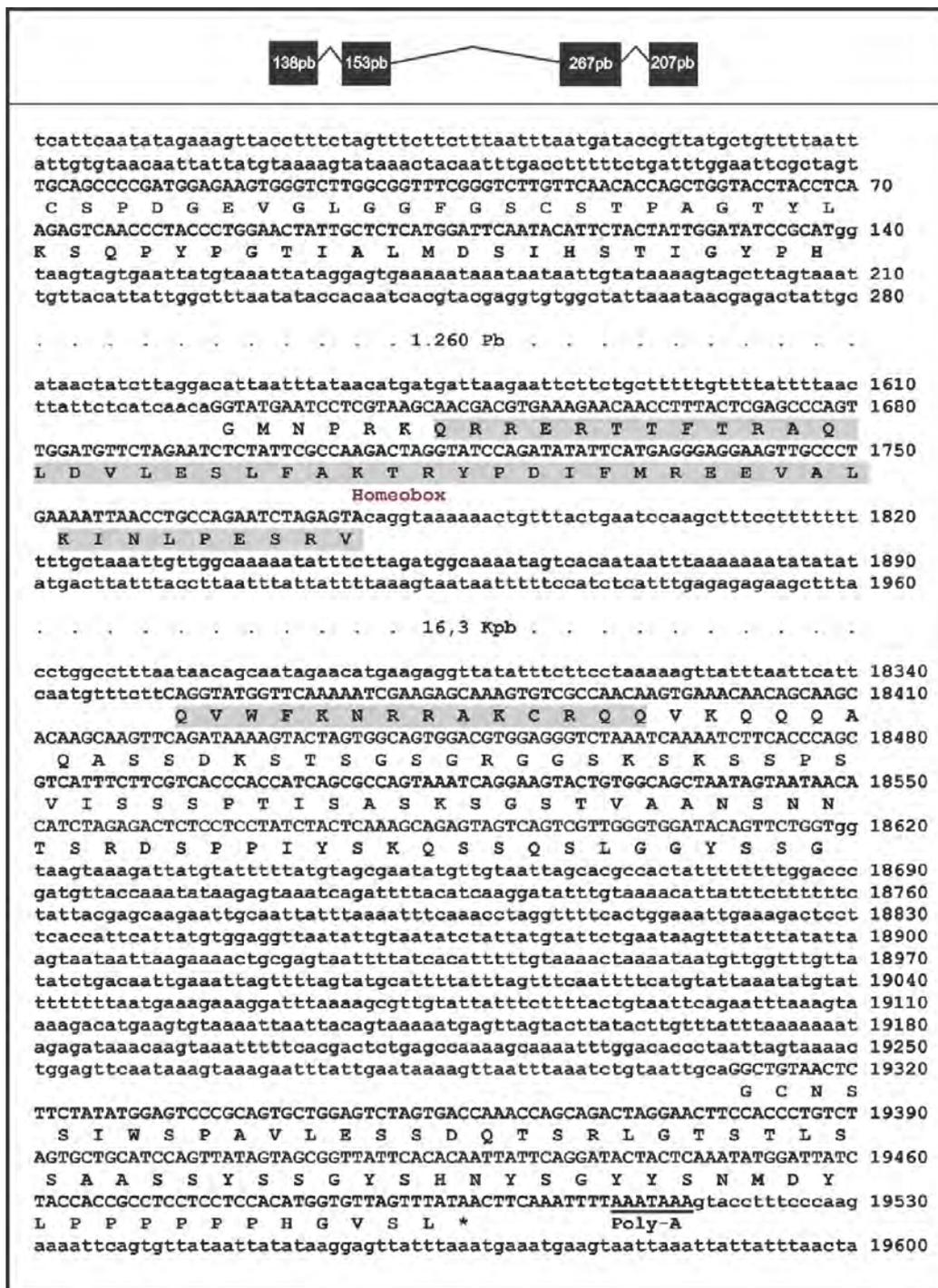
**Figura 20.** Análisis del gen *tailless* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-tll* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total. Se muestran los dos dominios Zn-F altamente conservados. **B.** Análisis filogenético de *Rp-tll*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

## 4.2.2 Genes gap cefálicos

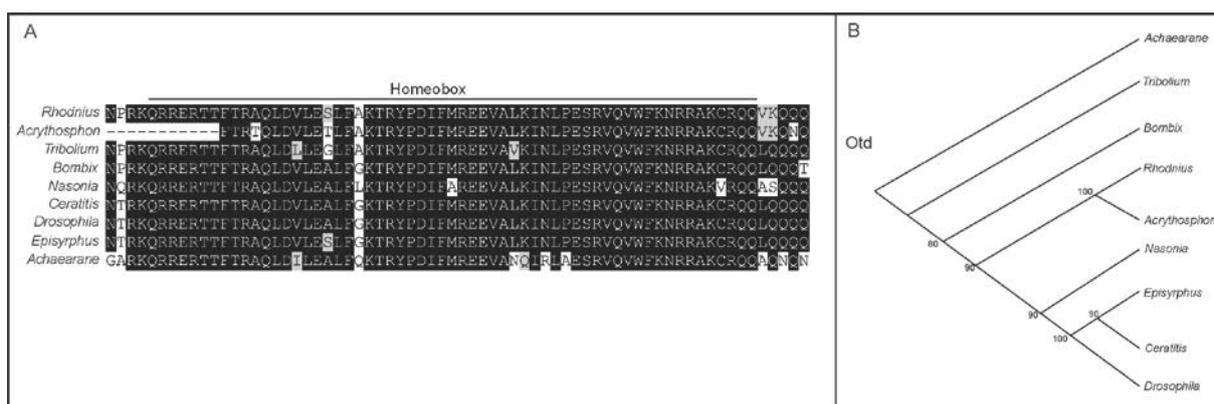
### 4.2.2.1 *orthodenticle*

Uno de los genes *gap* cefálicos encontrados en el genoma de *R. prolixus* fue *orthodenticle*. Este se localiza en el *contig* 17699.21, que tiene una longitud cercana a las 25 Kpb. El análisis de similitud de secuencia resultó en la identificación de parte del ORF de *otd*, en particular la región correspondiente al homeobox. De esta forma, se procedió a la predicción de la estructura génica de *Rp-otd*, la cual resultó en la identificación de un marco de lectura abierto de 19.516 pb. No fue posible identificar la región 5' UTR, pero sí la totalidad del dominio homeobox y la región 3' del gen hasta la señal de Poly-A (Figura 21). El ORF identificado está compuesto por cuatro exones que en conjunto tiene una longitud de 765 pb. El ORF de *Rp-otd* está interrumpido por un gran intrón, que separa los exones 2 y 3, que tiene una longitud de 16.573 pb. La secuencia codificante de *Rp-otd* se comparó con otros ortólogos de insectos, lo cual mostró un alto grado de similitud de secuencia en la región correspondiente al homeodominio (Figura 22 A). En este alineamiento de secuencias no se consideraron los ortólogos *Of-otd* y *Gb-otd* debido a que sólo se cuenta con unos pocos aminoácidos correspondientes a parte del homeobox de *otd*, y no la secuencia proteica total. Estos tampoco fueron considerados para el análisis filogenético. El análisis filogenético resultó en un buen agrupamiento de *R. prolixus* y *A. pisum* formando el grupo hemiptera, y *D. melanogaster*, *Episyrphus balteatus* y *Ceratitis capitata* formando el clado Díptera (Figura 22 B). Como se ve

en la figura 22 B, *T. castaneum*, *N. vitripennis* y *B. mori* no se ajustan bien dentro del árbol a la evolución del grupo, esto quizás se debe a errores en el alineamiento o secuencias con erróneas o incompletas.



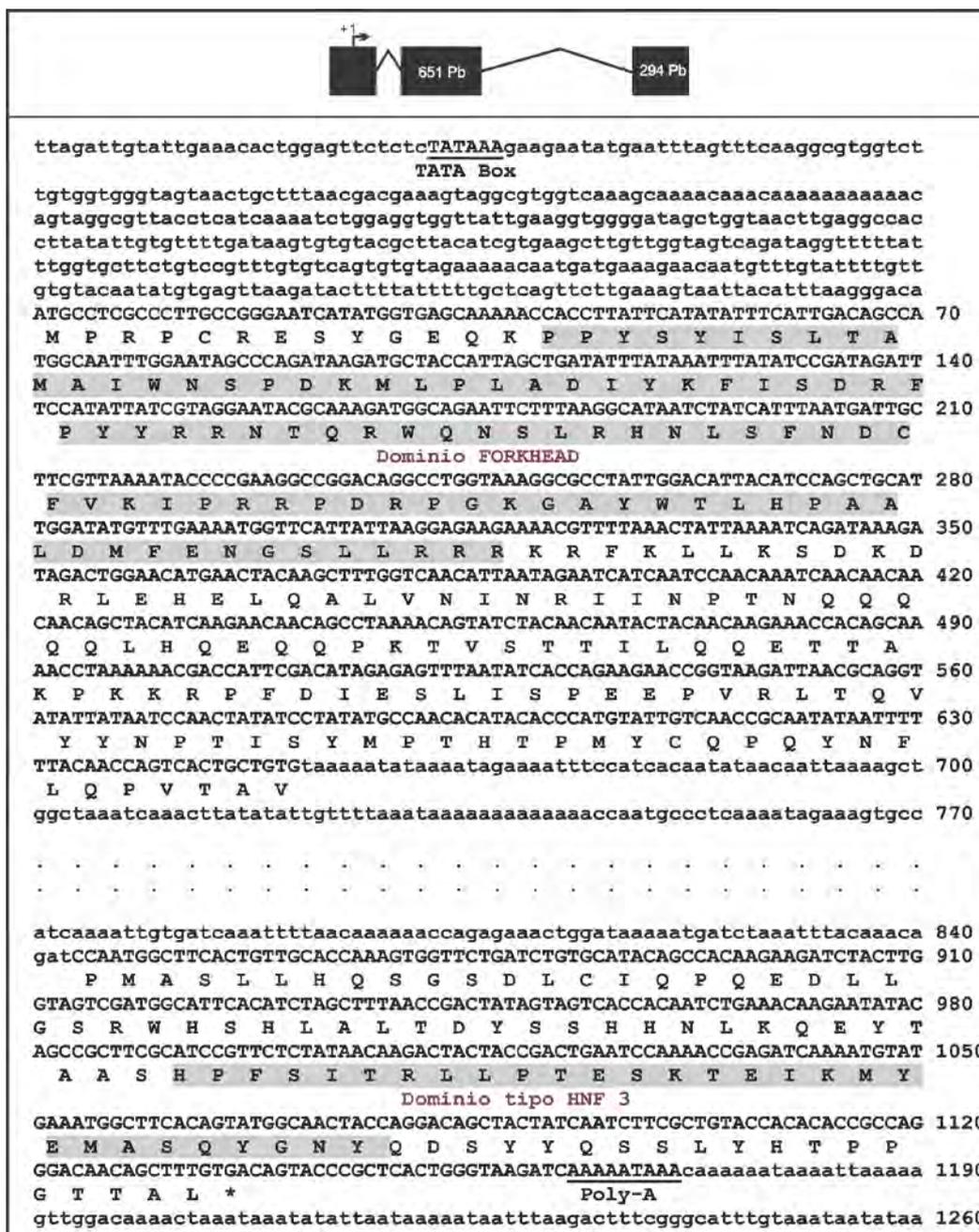
**Figura 21:** El gen *Rp-otd*. En la parte superior se muestra un esquema del gen. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-otd*. En gris se marcan los dominios homeobox altamente conservado en artrópodos. La región 5' no pudo ser identificada. La señal de poliadenilación se muestra subrayada. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.



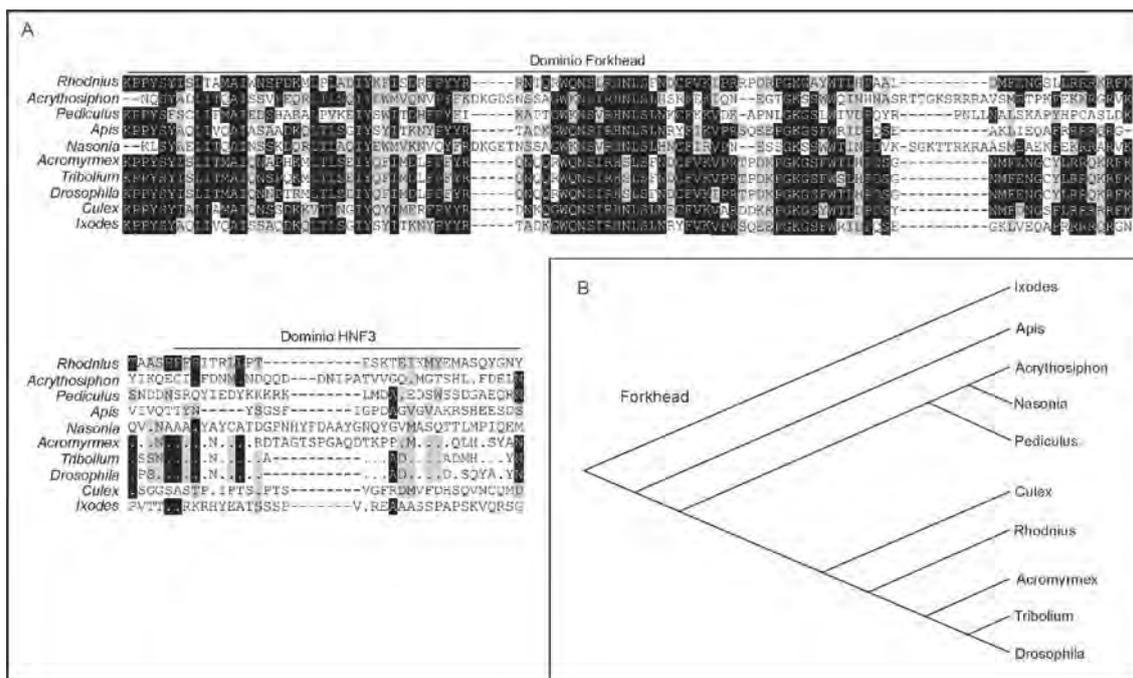
**Figura 22.** Análisis del gen *orthodenticle* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-otd* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total y como se puede ver el dominio homeobox de *Rp-otd* se encuentra muy conservado. **B.** Análisis filogenético de *Rp-otd*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

#### 4.2.2.2 *forkhead*

*Rp-fkd* fue encontrado en dos *contigs* diferentes dentro de la base de datos del genoma de *R. prolixus*. La región codificante de este gen está formada por dos exones y cada uno de estos se encuentra en un *contig* diferente. Uno de éstos es el *contig* 17794.92, el cual tiene una longitud de 3.556 pb y contiene la región 5' de *Rp-fkd*. El segundo *contig*, el cual contiene la región 3' de *Rp-fkd*, es el *contig* 17924.3 y tiene una longitud de 5.130 pb. Como dijimos, este gen está formado por dos exones, el primero de 651 pb que codifica para una región de la proteína *Rp-fkd* de 217 aa que contiene al dominio Forkhead, y el segundo exón de 294 pb que contiene la región C-terminal de la proteína y contiene el dominio tipo HNF 3 (Figura 23). La longitud del intrón que separa ambos exones no se pudo estimar debido a que éstos se encuentran en *contigs* separados y los extremos de los mismos no son superponibles. La secuencia aminoacídica de *Rp-fkd* fue comparada con ortólogos de *fkd* en otros insectos para ver el grado de conservación de secuencia. Como se puede ver en la figura 24 A, el dominio Forkhead está altamente conservado en todos los insectos. Observando el alineamiento se identificó una región conservada solo en parte de los insectos comparados, el dominio tipo HNF 3. Este dominio se encuentra conservado solo en *R. prolixus*, *T. castaneum*, *A. achinator* y *D. melanogaster*. El análisis filogenético mediante inferencia Bayesiana no es bueno ya que la agrupación de los taxones en el árbol no es coherente. Esto puede deberse a la presencia del dominio HNF 3 en los taxones mencionados anteriormente, los cuales dentro del árbol forman un grupo monofilético. Esto quizás distorsiona los datos de forma tal que la evolución del gen *fkd* no se corresponda con la evolución los insectos (Figura 24 B).



**Figura 23:** El gen *Rp-fkd*. En la parte superior se muestra un esquema del gen, indicando la longitud de cada exón. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-fkd*. En gris se marcan el dominio Forkhead altamente conservado en artrópodos, como si también el dominio tipo HNF3. El TATA-Box y la señal de poliadenilación se muestran subrayadas. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.



**Figura 24.** Análisis del gen *forkhead* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-fkd* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total y como se puede ver el dominio homeobox de *Rp-fkd* se encuentra muy conservado. **B.** Análisis filogenético de *Rp-fkd*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

#### 4.2.2.3 empty spiracles

El último de los genes *gap* cefálicos encontrados en el genoma de *R. prolixus* fue *empty spiracles*. Este fue localizado en los *contigs* 17146.7 (de 43.669 pb), 17493.5 (de 26.336 pb) y 17493.4 (de 5.580 pb). El gen está compuesto por 3 exones que codifican para una proteína de 285 aa con alta similitud de secuencia con ortólogos de *ems* de otros insectos. El primer exón tiene una longitud de 462 pb y codifica para la región N-terminal de *Rp-ems*. El segundo exón 167 pb codifica para la región central de la proteína, la cual posee casi la totalidad del dominio homeobox. El tercer exón tiene una longitud de 225 pb y codifica para parte del homeobox y para un dominio rico en glicinas en el extremo N-terminal de la proteína (Figura 25). El dominio homeobox de *ems* se encuentra conservado en insectos. De la comparación de *Rp-ems* con los ortólogos de *ems* de otros insectos, se destacan otros dos dominios conservados además del dominio homeobox (Figura 26 A). Uno de éstos es un dominio en la región N-terminal que aquí llamamos Dominio N, el otro en un dominio cercano a la zona C-terminal de la proteína, que es rico en glicinas, al cual denominamos Dominio C. El análisis filogenético del gen *ems* resultó en un árbol filogenético en el cual dípteros forman un grupo monofilético, al igual que coleópteros e himenópteros. *P. humanus* y *R. prolixus* también forman un grupo monofilético pero es excluido *A. pisum*, el cual enraza el grupo como *outgroup* (Figura 26 B).

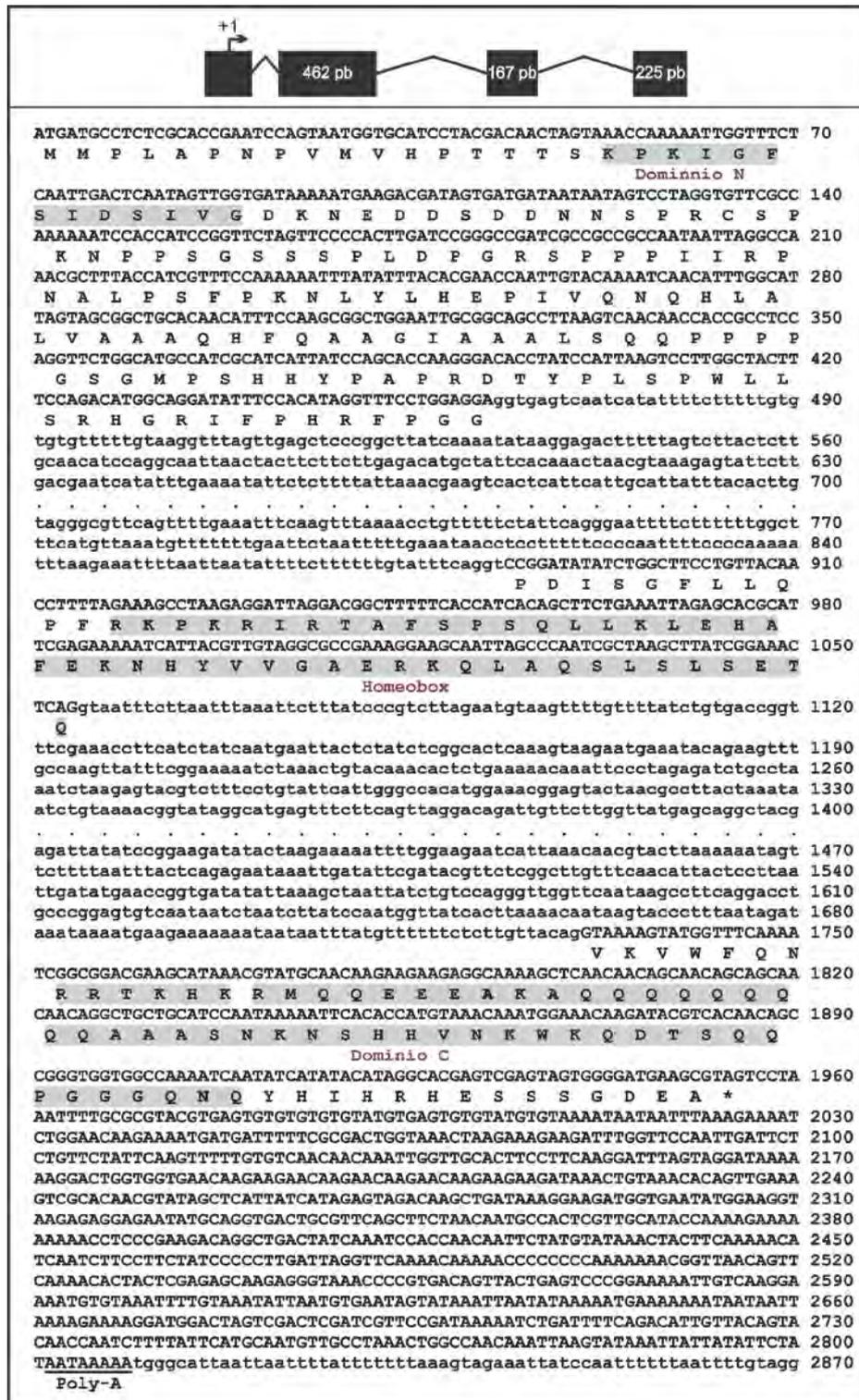


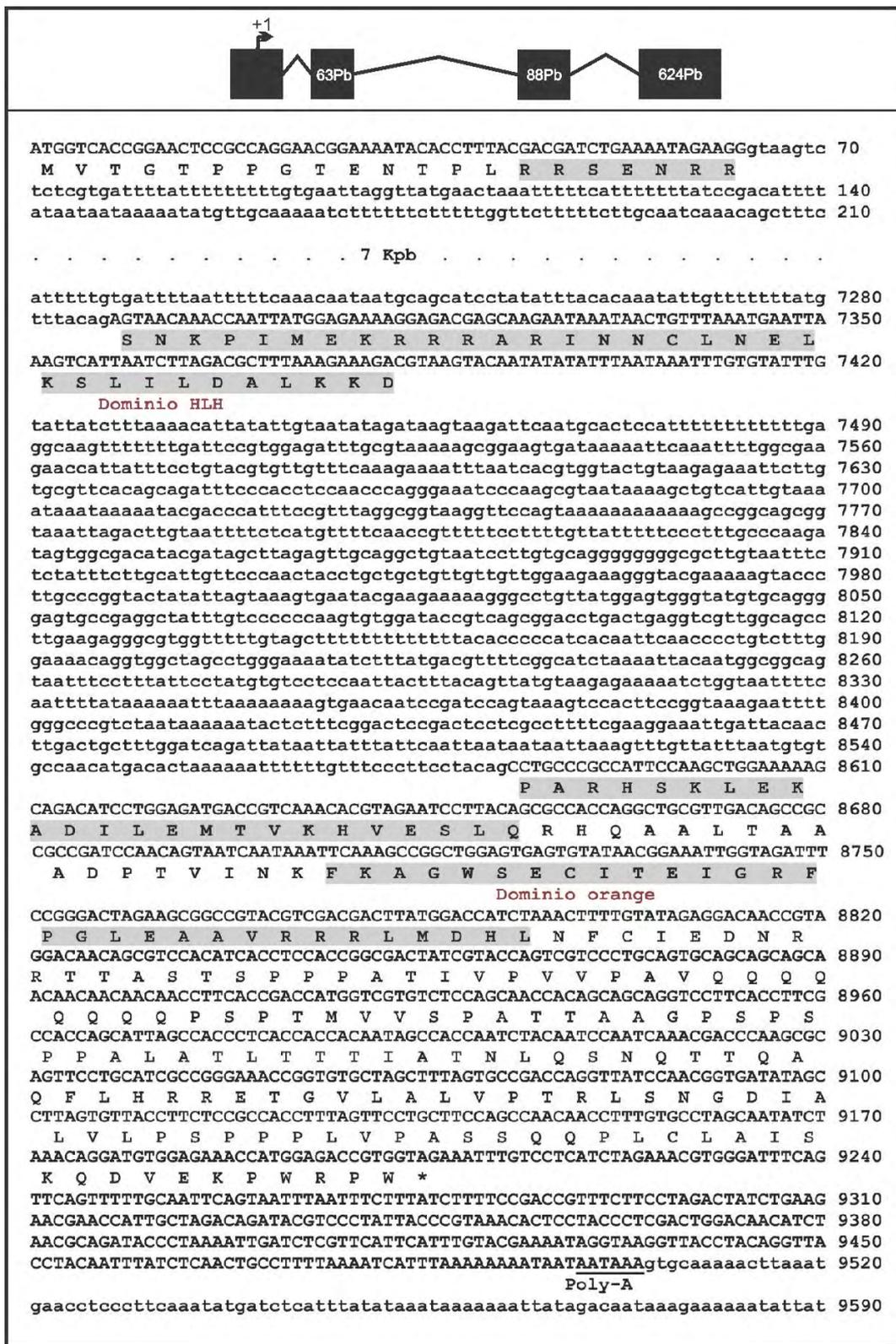
Figura 25: El gen *Rp-ems*. En la parte superior se muestra un esquema del gen, indicando la longitud de cada exón. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-ems*. En gris se marcan los dominios tres dominios conservados, el dominio homeobox, y los dominios N y C. La señal de poliadenilación se muestra subrayada. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.



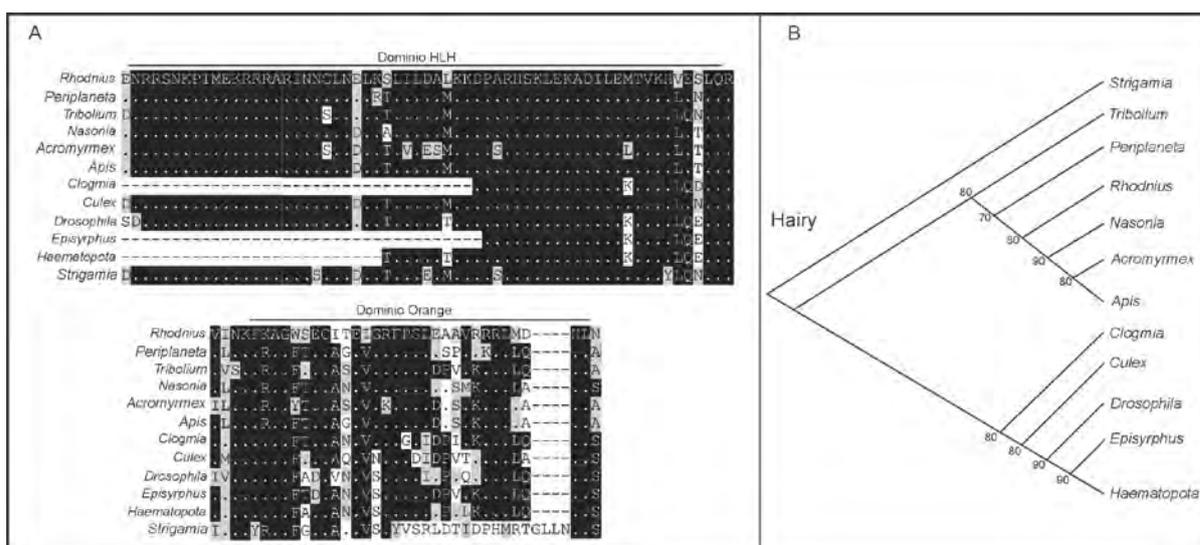
### 4.2.3 Genes *pair-rule*

#### 4.2.3.1 *hairy*

A partir de la búsqueda por similitud de secuencia sobre los *contigs* del genoma de *R. prolixus*, se logró identificar el ortólogo del gen *pair-rule hairy*. Éste se encuentra en el *contig* 6472.5, el cual tiene una longitud de 47,3 Kpb. *Rp-h* se extiende unas 16.187 pb, desde el supuesto TATA-Box hasta la señal de Poly-A, y contiene un ORF de 260 aa. La secuencia genómica conteniendo el gen *hairy* fue analizada bioinformáticamente con el fin de identificar su estructura génica. Como resultado de esto se pudo observar que *Rp-h* se encuentra formado por 3 exones, uno de 66 pb, otro de 88 pb y el último de 624 pb (Figura 27). El gen *Rp-h* posee un gran intrón que separa el primer y segundo exón tiene una longitud de 7,2 Kpb, mientras que el segundo tiene una longitud de 1209 pb. La secuencia proteica fue comparada con la de ortólogos en otras especies de insectos y reveló un alto grado de similitud de secuencia, en la región correspondiente a los dominios hélice-vuelta-hélice (*Helix-Loop-Helix* - HLH) y *Orange* (Figura 28 A). El análisis evolutivo aplicando matemática Bayesiana, reveló que la evolución del gen *hairy* acompaña en parte a la evolución del grupo, principalmente en dípteros, los cuales forman un grupo monofilético (Figura 28 B).



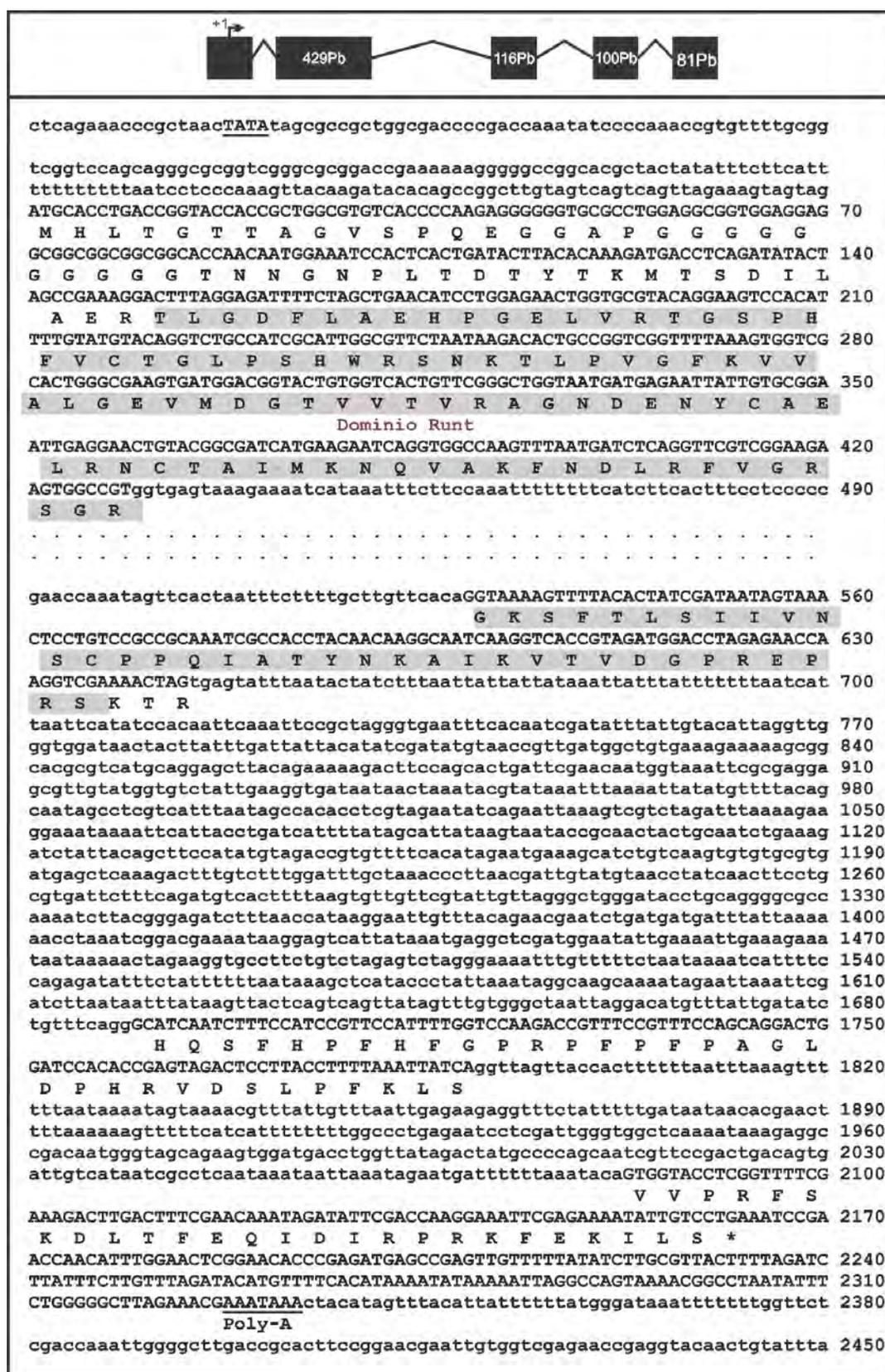
**Figura 27:** El gen *Rp-h*. En la parte superior se muestra un esquema del gen, indicando la longitud de cada exón. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-h*. En gris se marcan los dominios proteicos conservados: El dominio *Helix-Loop-Helix* (HLH) y el dominio *Orange* típico de *hairy*. La señal de poliadenilación se muestra subrayada. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.



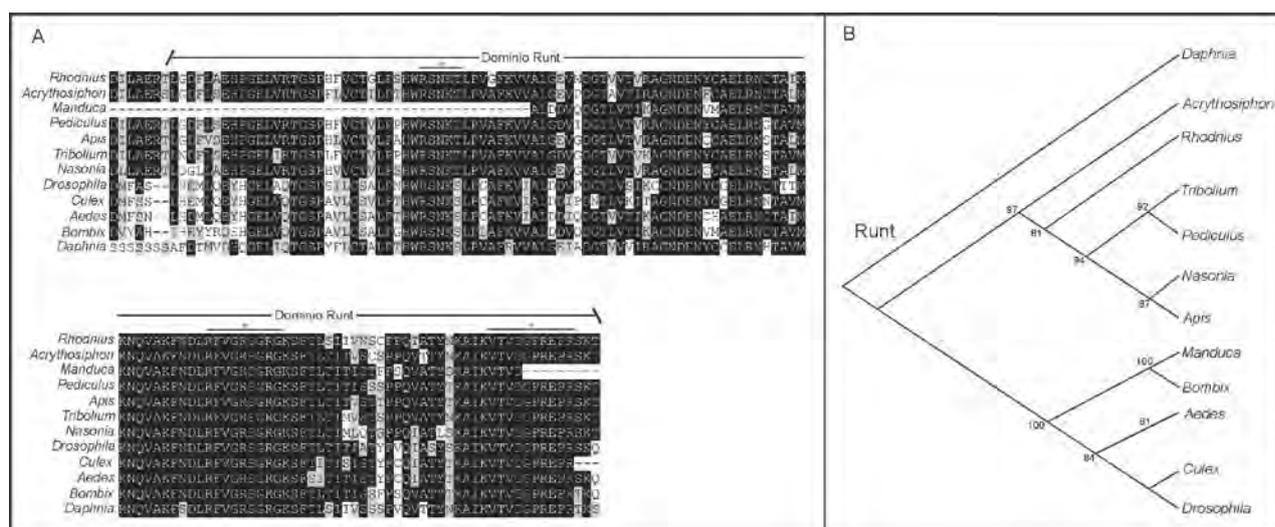
**Figura 28.** Análisis del gen *hairy* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-h* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total y como se puede ver el dominio *Helix-Loop-Helix* (HLH) y el Dominio *Orange* se encuentran muy conservados. **B.** Análisis filogenético de *Rp-h*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

#### 4.2.3.2 runt

Otro de los genes *Pair-Rule* encontrados dentro del genoma de *R. prolixus* fue el gen *runt*. Este gen está altamente conservado tanto en insectos como en vertebrados. *Rp-run* se encontró en dos *contigs* diferentes dentro de la base de datos de *R. prolixus*: el *contig* 16559.1 (de 19,4 Kpb) y el *contig* 17847.29 (de 19,2 Kpb). El primero contiene un marco de lectura abierto que codifica para la región N-terminal del dominio Runt de *Rp-run*, mientras que el segundo *contig* contiene un ORF que codifica para la región C-terminal. Ambos *contig* no son superponibles, por lo tanto, cabe la posibilidad de que contengan un intrón de gran extensión que los separa. En la figura 29 se muestra la estructura del gen *Rp-run*, donde se puede ver que está formado por cuatro exones, en donde el primero y el segundo codifican para el dominio Runt, separados ambos por un intrón de longitud desconocida. La proteína para la cual codifica *Rp-run* fue comparada con ortólogos de *runt* en otros insectos. Esto demostró que esta proteína, en particular el dominio Runt y sus tres sitios de unión a ADN, están altamente conservada dentro del grupo (Figura 30 A). El análisis filogenético realizado, resultó en un árbol que no se corresponde con la evolución del grupo (Figura 30 B). Esto se debe a que la secuencia proteica de los diferentes ortólogos de *run* es altamente variable entre taxones, salvo en la región correspondiente al dominio Runt. Esto hace que durante el análisis se distorsionen los resultados.



**Figura 29:** El gen *Rp-run*. En la parte superior se muestra un esquema del gen. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp-run*. En gris se marca el dominio Runt que se encuentra muy conservado en artrópodos. El TATA-Box y la señal de poliadenilación se muestran subrayadas. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencias no codificantes.



**Figura 30.** Análisis del gen *runt* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-run* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad tota. Como se puede ver el dominio Runt está bien conservado (los asteriscos marcan los dominios de unión a ADN). **B.** Análisis filogenético de *Rp-run*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

#### 4.2.3.3 *even-skipped*

Otro de los genes *Pair-Rule* encontrado dentro de la base de datos del genoma de *R. prolixus* fue *even-skipped*. Este se localizó en el *contig* 16644.1, el cual tiene una longitud cercana a 34,3 Kpb. Contiene un ORF compuesto por tres exones que codifican para el gen *Rp-eve*, teniendo el gen *eve* de *R. prolixus* una longitud de 7.165 pb (Figura 31). Como resultado de la comparación de *Rp-eve* con ortólogos de *eve* en otros insectos, se determinó que el dominio homeobox está altamente conservado dentro del grupo. Además de este dominio, se observa un dominio en el extremo C-terminal (que nombramos como Box C), el cual está conservado en la gran mayoría de los insectos a excepción de *B. mori* y *C. albipunctata* (Figura 32 A). El análisis filogenético muestra que la evolución del gen se acopla en gran medida a la evolución del grupo, en donde dípteros y lepidópteros forman un grupo monofilético, al igual que himenópteros y coleópteros. Los insectos hemimetábolos, como *R. prolixus*, *O. fasciatus*, *S. gregarina* y *A. pisum* forman un grupo monofilético a excepción de *G. bimaculatus* el cual, según el análisis realizado, se agrupa junto con coleópteros e himenópteros (Figura 32 B).

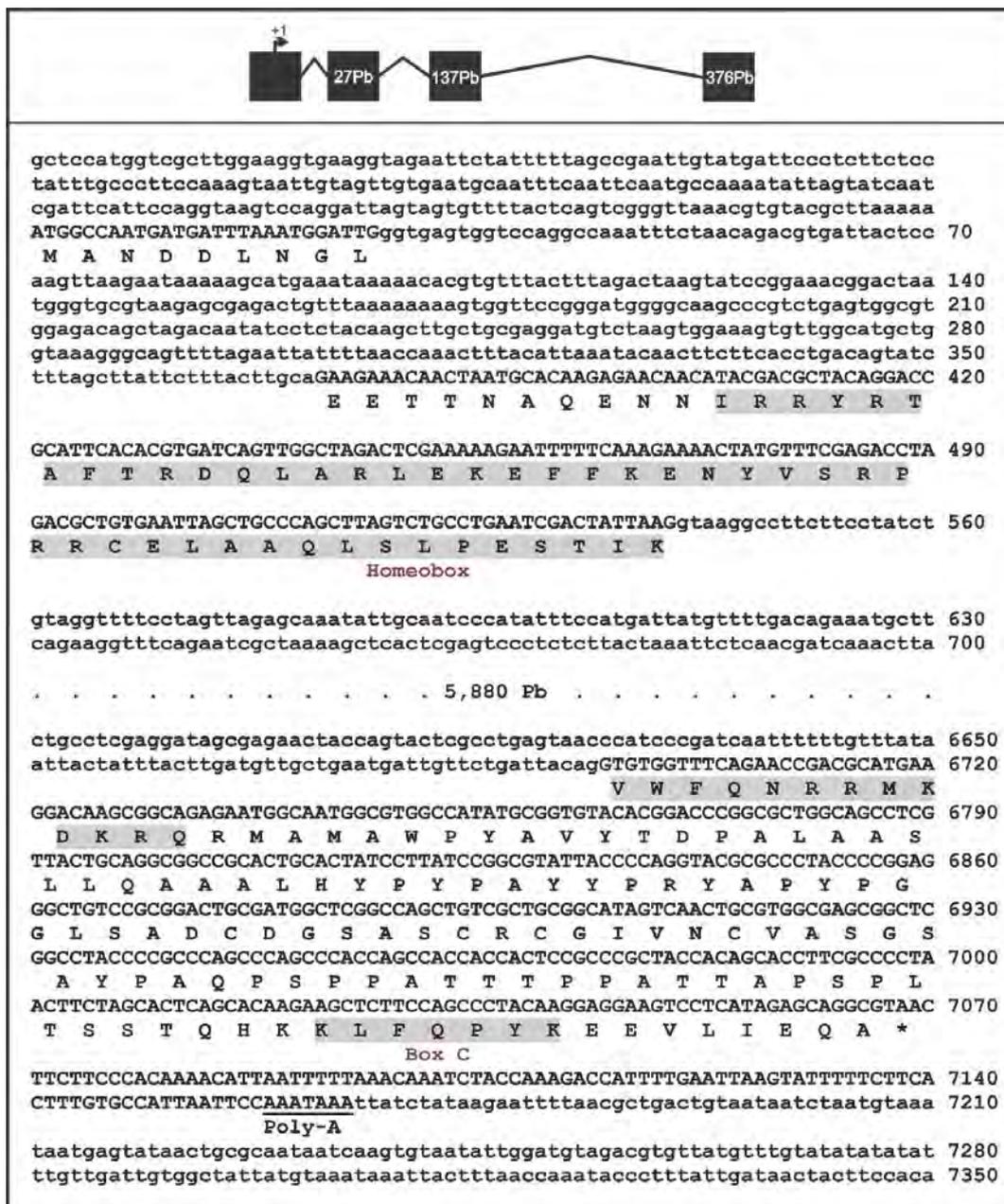
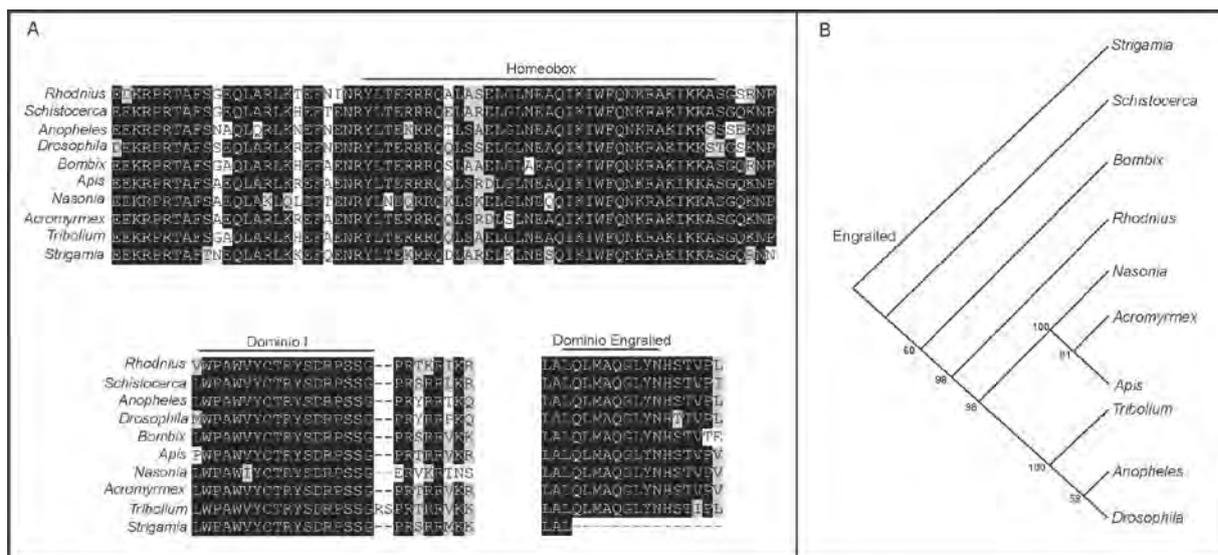


Figura 31: El gen *Rp- eve*. En la parte superior se muestra un esquema del gen, indicando la longitud de cada exón. En la parte inferior se muestra la secuencia nucleotídica y aminoacídica de *Rp- eve*. En gris se marcan los dominios proteicos conservados: dominio homeobox y el Box C. La señal de poliadenilación se muestra subrayada. En la secuencia nucleotídica, se marca en mayúscula la región exónica y en minúscula los intrones y secuencia no codificantes.







**Figura 34.** Análisis del gen *engrailed* en *R. prolixus*. **A.** Alineamiento de la secuencia proteica de *Rp-en* con ortólogos de otros insectos. Las zonas marcadas con negro indican identidad total. Se muestran el dominio homeobox y el motivo homeobox típico en, y el Dominio I que se encuentra altamente conservado cerca del extremo N-terminal se encuentra muy conservado. **B.** Análisis filogenético de *Rp-en*. El árbol fue generado por medio de inferencia Bayesiana; los valores de los nodos representan los valores de credibilidad de cada clado.

### 4.3 Análisis funcional de los genes *gap giant* y *Krüppel*.

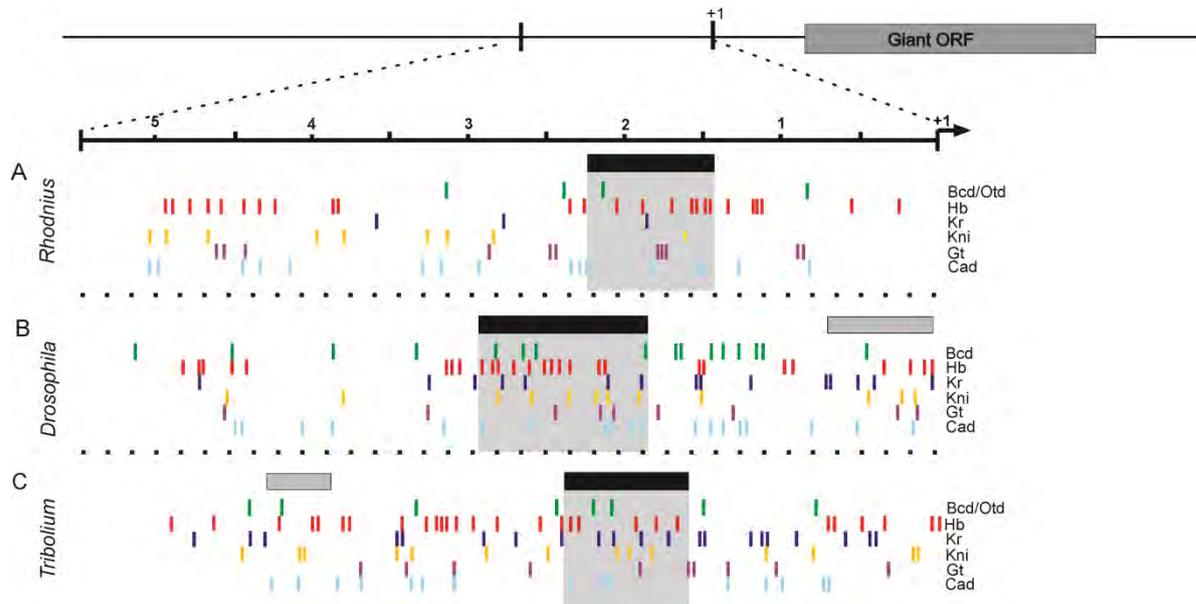
#### 4.3.1 *giant*

Como se dijo en el capítulo anterior, como resultado del ensamblado de los fragmentos del genoma se obtuvo una región genómica de 12,5 Kpb que contiene al gen *Rp-giant*. Este tiene una longitud de 4.138 pb y posee un marco de lectura abierto de 248 aminoácidos libre de intrones (remítase a la Figura 11). Para validar la actividad transcripcional del supuesto gen *Rp-gt*, se diseñaron *primers* específicos a partir de este ORF para realizar su amplificación por PCR utilizando como molde ADNc normalizado. El producto de PCR resultó en un amplicón de 332 pb, éste fue clonado en el vector pGEM-T *Easy* y secuenciado. La secuencia contiene el dominio *Leucine-zipper* (BLZ) completo y se encuentra subida a la base de datos del *GeneBank* bajo en número de acceso GU724146.1.

##### 4.3.1.1 Agrupamiento de reguladores en el promotor del gen *Rp-gt*

Contando con 12,5 Kpb identificadas en el genoma que contiene a *Rp-gt*, se analizó la secuencia *upstream* de la unidad transcripcional, en busca la presencia de sitios de unión para potenciales reguladores de este gen. El análisis bioinformático de la secuencia con el *software* PATSER resultó en numerosos sitios de unión a la secuencia promotora para diferentes factores de transcripción que potencialmente regularían la actividad de *Rp-gt*. El número de sitios de unión encontrados para cada uno de los factores de transcripción fueron los siguientes: Hb, 30; Bcd/Otd, 4; Cad, 16; Kr, 3; Kni, 9 y Gt, 11. Como la presencia de sitios de unión puede ocurrir al azar a lo largo del genoma, también se evaluó otro parámetro, el agrupamiento de los mismos, que se denomina “*clustering*”. Con el uso del *software* STUFF se pudo predecir, con un alto grado de confianza, que los sitios de unión ubicados entre las 0,8 Kpb y las 1,5 Kpb *upstream* a partir del origen de transcripción podrían conformar un *cluster* regulatorio. Este fragmento de secuencia contiene los siguientes sitios de unión a factores de transcripción: Bcd/Otd, 1; Hb, 7; Kr, 1; Kni, 1; Gt, 3 y Cad, 3 (Figura 35 A). Este tipo de análisis fue previamente realizado en *D.melanogaster*, permitiendo identificar la secuencia responsable de la regulación del dominio de expresión abdominal de *Dm-gt* (Berman *et al.*, 2002). Para evaluar la fidelidad de nuestros resultados, se analizó el *clustering* de factores de transcripción de la secuencia *upstream* a *Dm-gt* (como control) y de *Tc-gt*. Para esto se usaron los mismos parámetros de análisis que en *R. prolixus*. Esto resultó en la identificación de un *cluster* regulatorio entre los 0,7 y 1,7 Kpb a partir del origen de transcripción en *T. castaneum* y, como era de esperar, en *D. melanogaster* entre los 1 Kpb y 1,8 Kpb a partir del nucleótido +1 (Figura 35 B y C). Como se ve, la posición

de los *cluster* para las tres especies es relativamente similar, y esto podría estar definiendo una potencial secuencia regulatoria conservada para *Rp-gt*, *Tc-gt* y *Dm-gt*.

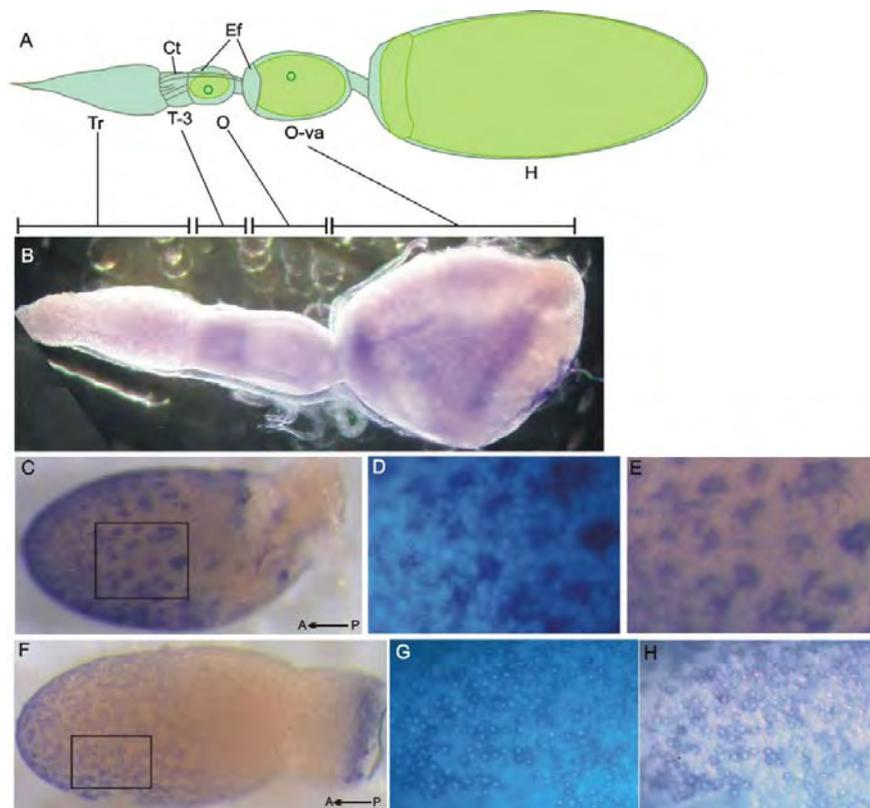


**Figura 35.** Análisis comparativo de la potencial secuencia promotora de *giant* en *R. prolixus*, *D. melanogaster* y *T. castaneum*. Secuencia reguladora de *R. prolixus* (A), *D. melanogaster* (B) y *T. castaneum* (C). A la derecha del gráfico se muestran los diferentes factores de transcripción. Las rayas de color representan los sitios de unión de los reguladores a la secuencia, los cuales fueron predichos por el software PATSER. Los *clusters* regulatorios predichos por el software STUBB son marcados con los rectángulos negros (mayor probabilidad) y grises. La escala marca la distancia en Kpb a partir del origen de transcripción.

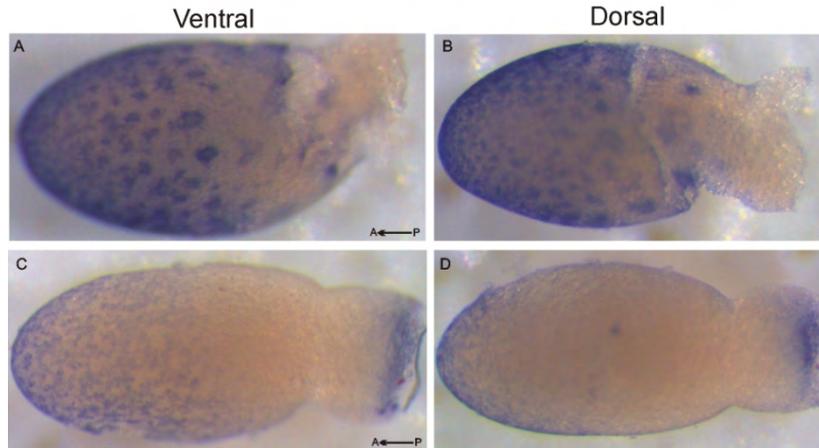
#### 4.3.1.2 Expresión de *Rp-gt*.

Como se comentó en la sección de Materiales y Métodos 2.4.2, se realizó la hibridación *in situ* de sondas de ARN en ovarios y embriones para determinar los dominios de expresión de *Rp-gt* en el embrión. En ovarios, el mensajero de *Rp-gt* se detectó en el trofario y oocitos. En los oocitos la expresión se da en aquellos que se encuentran en estadios tempranos de maduración (los cuales son comparables con los oocitos en estadio T-3 de Bjorsson y Huebner, 2004), y en los que se encuentran en vitelogénesis avanzada (O-va) (Figura 36 A y B). El transcripto de *Rp-gt* es también detectado en embriones que cursan las primeras fases del desarrollo. En embriones con 0-12 hs PO el mensajero de *Rp-gt* se encuentra distribuido en forma de manchas o parches a lo largo de la mitad posterior del huevo (Figura 36 C). Para este estadio, los núcleos, que posteriormente formarán el blastodermo, todavía no han migrado a la periferia del huevo. De esta forma, los parches de mensajero de *Rp-gt* no están asociados a ninguna estructura celular, lo cual se puede ver claramente en las imágenes teñidas con DAPI (Figura 36 D y E). Para las 12-24 hs PO, los núcleos se hacen presentes en la superficie del huevo y la distribución de *Rp-gt* se hace más homogénea, mostrando una distribución perinuclear (Figura 36 F-H). En

la figura 36 G y H se muestra claramente la distribución perinuclear para el estadio de blastodermo. A lo largo de éste estadio la expresión de *Rp-gt* se da en la mitad posterior del huevo, siendo más intensa en la cara ventral que en la cara dorsal (Figura 37). El polo posterior del huevo corresponde a la región anterior del embrión antes de la katatrepsis. Nótese que la parte anterior del huevo (el opérculo) corresponde al polo posterior del embrión durante el estadio de banda germinal, se remarca esto para que no quepa lugar a confusión más adelante en el texto. La expresión de *Rp-gt* para los estadios mencionados arriba se da en una etapa pre-cigótica del desarrollo embrionario, donde no se ha diferenciado ningún tipo de células embrionarias. De esta forma la expresión observada es expresión materna, producto de la transferencia direccional de mensajeros al huevo por la hembra.

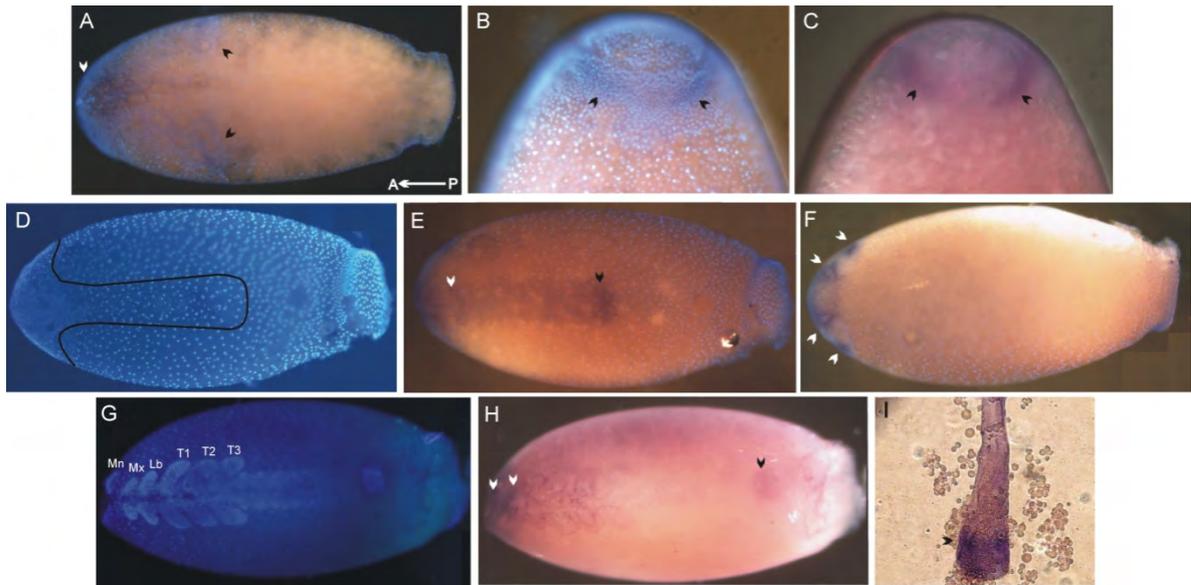


**Figura 36:** Expresión materna de *Rp-gt*. **A.** Esquema donde se muestra la estructura de una ovariola de *R. prolixus*. **B.** Distribución de los mensajeros de *Rp-gt* en ovarios. Hay expresión en el trofario (Tr), oocitos tempranos (T-3) y oocitos en vitelogénesis avanzada (O-va), oocitos fuera de la etapa de vitelogénesis (O) no muestran expresión. **C - H.** Distribución de ARNm de *Rp-gt* en embriones tempranos. **C.** Embrión con pocas horas post ovoposición, en el estadio de pre-blastodermo, donde el transcripto de *Rp-gt* se distribuye en parches a lo largo de la mitad anterior del huevo. El rectángulo indica la ampliación mostrada en D y E. La flecha indica el polo anterior del embrión (cabeza), el cual coincide con el polo posterior del huevo (opuesto al opérculo) antes de la katatrepsis. **D.** Imagen del embrión teñido con DAPI (no hay núcleos en la superficie del huevo). **E.** La misma imagen con epiiluminación, donde se muestra el patrón de distribución del transcripto. **F.** Embrión de 12-24 hs PO. La distribución del transcripto es más homogénea que en C. **G.** Blastodermo teñido con DAPI. **H.** La misma imagen que en G pero con epiiluminación, mostrando la distribución perinuclear de *Rp-gt*. Esta imagen fue adquirida por doble exposición por epiiluminación y epifluorecencia UV. Abreviaciones: Ct, cordón trófico; Ef, epitelio folicular; H, huevo; O, oocito; O-va, oocito en vitelogénesis avanzada; T-3, oocitos tempranos según Bjorsson y Huebner (2004) y Tr, trofario.



**Figura 37:** Localización ventralizada del transcrito de *Rp-gt* transcrito. **A-B.** Embrión con 0-12 hs PO, distribución del mensajero en parches. **C-D.** Embrión con 24 hs PO, *Rp-gt* con distribución perinuclear. A y C, vista ventral del huevo. B y D, vista dorsal. Para los diferentes tiempos de desarrollo mencionados, la distribución del transcrito es más intensa en la cara ventral del huevo.

Durante el proceso de gastrulación (30-36 hs PO), la expresión deviene en expresión cigótica, en donde el dominio de expresión se retrae hacia la parte posterior del huevo, acompañando a las células en migración que formaran luego la banda germinal (Figura 38 A, B y C). En la cara ventral del huevo se detectan dos zonas donde se expresa *Rp-gt*: una anterior (donde se formará la cabeza) (Figura 38A; punta de flecha blanca) y una en la región media del huevo (Figura 38A; punta de flecha negra). El dominio de expresión anterior se subdivide en dos dominios a cada lado del centro de gastrulación (Figura 38 B y C). Los embriones fueron co-teñidos con DAPI para determinar la morfología de los embriones y poder datar los dominios de expresión. Para el estadio de banda germinal (36-48 hs PO), la expresión de *Rp-gt* tiene lugar en tres dominios: uno en la parte más anterior, un dominio de expresión medio, y uno posterior (Figura 38 D, E y F). El dominio de expresión más anterior se encuentra en la región correspondiente a los futuros segmentos pre-gnatales. Este dominio está formado por cuatro sub-dominios, dos a cada lado de la línea media del embrión (Figura 38 F). El dominio de expresión medio coincide con la zona que posteriormente ocuparan los segmentos mandibular y maxilar (ver luego). El dominio posterior corresponde a la “zona de crecimiento” donde los segmentos abdominales son agregados consecutivamente (Figura 38 E). Luego que se completó la extensión de la banda germinal (48-60 hs PO), el dominio de expresión anterior queda restringido a los segmentos mandibular y maxilar (Figura 38 G y H; puntas de flecha blancas), mientras que el dominio posterior a la zona terminal del abdomen (Figura 38 H, puntas de flecha negras y 38 I). Los dominios de expresión de *gt* correspondientes a los segmentos gnatales y abdominales están muy conservados en los todos insectos para los cuales se estudió previamente, pero el dominio pre-gnatal solo se describió para *D. melanogaster*.



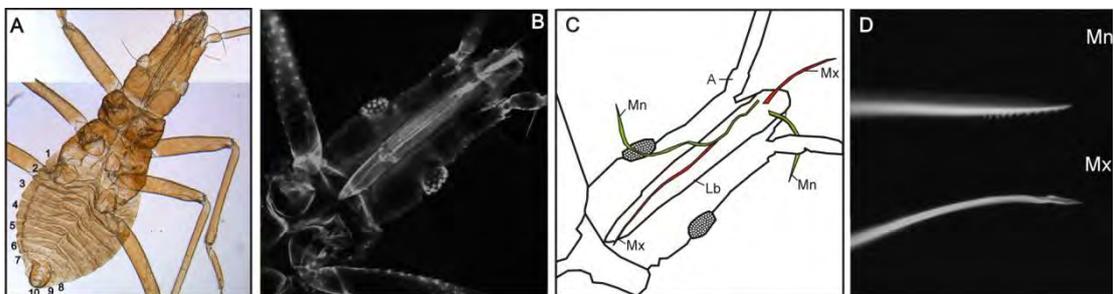
**Figura 38.** Expresión cigótica de *Rp-gt*. **A-C.** Embrión en invaginación (30-36 hs PO). **A.** Vista ventral donde se muestra el dominio de expresión anterior y el dominio posterior temprano. **B-C.** Vista dorsal mostrando la distribución del transcrito *Rp-gt* en las cercanías del borde de invaginación. **B.** Hibridación *in situ* junto con la tinción con DAPI. **C.** el mismo embrión, donde sólo se muestra la marcación de la hibridación *in situ*. **D-E.** Vista ventral de embriones en estadio de banda germinal en extensión. **D.** Tinción con DAPI. La línea negra marca la posición del embrión en el huevo. **E.** Hibridación *in situ* mostrando el mismo embrión en D. La flecha negra marca el dominio de expresión posterior en la zona de crecimiento. **F.** Vista dorsal del mismo embrión en E, donde las flechas blancas muestran el dominio de expresión más anterior (en el futuro cípeo-labro). **G-I.** Embriones con extensión de la banda germinal completa. **G.** Imagen en vista ventral teñida con DAPI donde se indican los apéndices cefálicos y torácicos (Mn, mandíbulas; Mx, maxilas; Lb, labio; T1-T3, segmentos torácicos 1 a 3). **H.** Hibridación *in situ* del mismo embrión en G, donde se muestran los diferentes dominios de expresión: mandibular y maxilar (flechas blancas) y abdominal (flecha negra). **I.** Detalle del dominio de expresión posterior de *Rp-gt* en la zona de crecimiento en un embrión disectado.

#### 4.3.1.3 Fenotipo *giant* en *R. prolixus*.

Para determinar el fenotipo resultante de la pérdida de expresión de *Rp-gt*, se generaron fenocopias por ARNi parental. Las hembras inyectadas produjeron un total de 199 huevos en dos experimentos independientes. Los embriones fueron estudiados a diferentes tiempos post-puesta para determinar el fenotipo resultante (Tabla 1). El 78% de los embriones interferidos mostraron un fenotipo diferente al salvaje (*Wild type – wt*). Dentro de estos se puede dividir dos grupos que llamamos: fenotipo leve (8%) y fuerte (92%). Las fenocopias leves completaron el desarrollo pero no sobrevivieron. Estos embriones corresponden a las hembras interferidas con ARN sc antisentido y a las primeras puestas de las hembras interferidas con ARN dc antisentido, donde la penetrancia de la técnica es menor.

<b>Tabla 1:</b> Resultados de la ARNi parental de <i>Rp-gt</i>				
	Fuerte (%)	Leve (%)	WT (%)	Total
ARNsc antisen	0 (0%)	8 (17%)	39 (83%)	47
ARNdc antisen	143 (94,1%)	5 (3,3%)	4 (2,6%)	152
	143 (71,9%)	13 (6,5%)	43 (21,6%)	199 (100%)

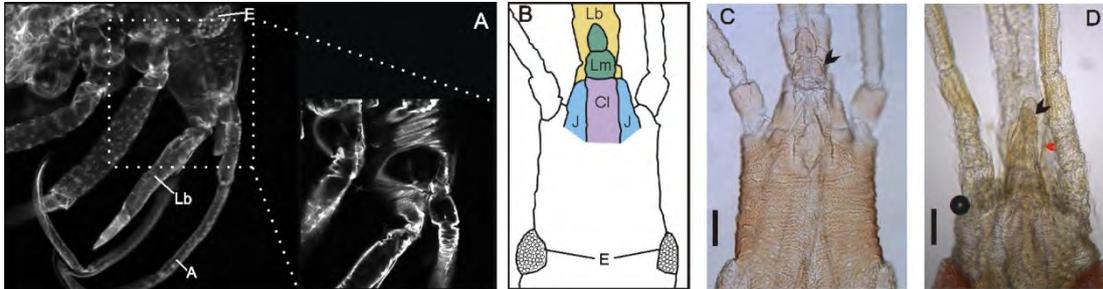
A continuación se describe brevemente el fenotipo *wt* de *R. prolixus* con sus características más importantes. El abdomen del insecto *wt* está formado por un total de 10 segmentos (Figura 39 A). La cabeza consiste de los segmentos pre-gnatales (formados por el clipeo, el labro y el segmento antenal) y los segmentos gnatales (mandíbulas, maxilas y labio) formando en conjunto la proboscis (Figura 39 B y C). *R. prolixus* (y el resto de los hemípteros) tiene una cabeza altamente modificada en función de los hábitos alimenticios, en este caso una dieta hematófaga, en donde los segmentos más anteriores se encuentran plegados ventralmente. El clipeo y el labro son las estructuras más anteriores de la cabeza y en conjunto forman la capsula cefálica. El labro es segmentado y está formado por dos segmentos, mientras que el clipeo no es segmentado y se encuentra por detrás del labro. En el *wt* la longitud de la cabeza es cercana a los 637  $\mu\text{m}$ , y la longitud del “clipeo-labro” (considerando a estos como una estructura única) es de 363  $\mu\text{m}$ , siendo 213  $\mu\text{m}$  para el clipeo y 150  $\mu\text{m}$  para el labro. La proboscis está formada por el labio que rodea a los estiletes, siendo estos las mandíbulas y las maxilas modificadas. Las mandíbulas tienen el extremo distal aserrado y las maxilas los tienen espatulado (Figura 39 D).



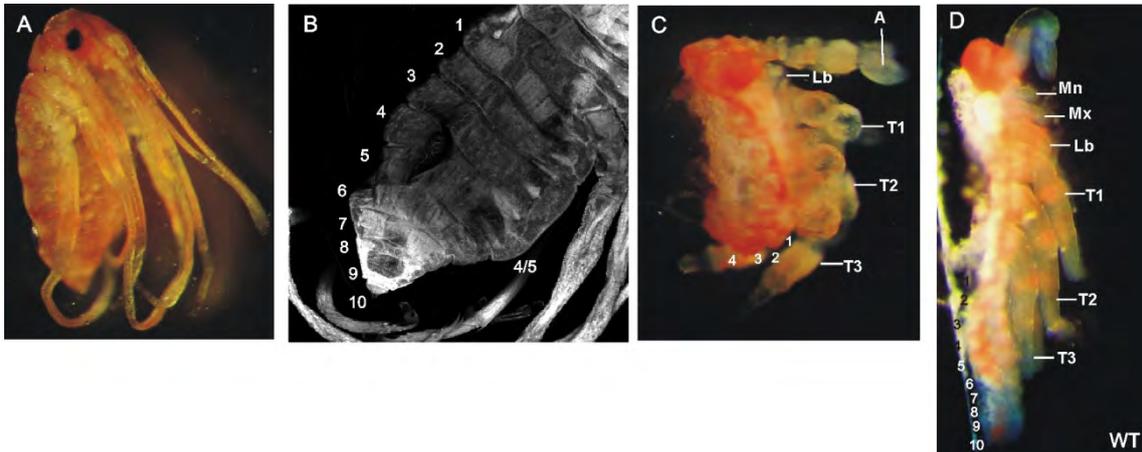
**Figura 39.** Características morfológicas del 1<sup>er</sup> estadio larval de *R. prolixus*. **A.** Preparado cuticular en vista dorsal donde se marcan los segmentos abdominales. **B.** Imagen de confocal de la cabeza. **C.** Esquema de la imagen mostrada en B, indicando los diferentes apéndices cefálicos. (A, antena; Lb, labio; Mn, mandíbula y Mx, maxila). **D.** estructura de los estiletes, mandíbula (punta aserrada) y maxila (punta espatulada).

Los embriones interferidos muestran un inequívoco fenotipo que corresponde a las regiones donde se expresa *Rp-gt*: reducción de la longitud de la cabeza, pérdida de segmentos gnatales (Figura 40) y pérdida y fusión de segmentos abdominales (Figura 41). Todos muestran un desarrollo normal del segmento labial y una proboscis normal (Figura 40 A-D). El análisis de las cutículas larvales clarificadas utilizando secciones de microscopia confocal, demostró un desarrollo normal de la proboscis, pero con ausencia de estiletes en su interior, tanto Mn como Mx (Figura 40 A). Otra de las modificaciones más evidentes en la cabeza es el acortamiento del clipeo y el labro, comparados con el *wt* (Figura 40 C-E). Estas dos estructuras se encuentran fusionadas y es imposible distinguir una de otra, por eso nos referimos a esta estructura como “clipeo-labro”. En estas fenocopias la longitud del “clipeo-labro” es de 248  $\mu\text{m}$  y la longitud de la cabeza por delante del borde anterior de los ojos, es de 387  $\mu\text{m}$ . Si comparamos estos parámetros con las mediciones del *wt*, se ve que el “clipeo-labro” se redujo 31,6% en longitud y la región cefálica por delante de los ojos un 39,2%. Este fenotipo se correlaciona con el dominio de expresión más anterior, en la región pre-gnatale. La pérdida o reducción de estas estructuras ha sido descrita también para los mutantes *gt* de *D. melanogaster*, aportando esto evidencias de que la similitud de función entre *Rp-gt* y *Dm-gt* es mayor que para otros grupos de insectos analizados hasta la fecha. En algunas larvas, donde la penetrancia de la técnica fue menor, se observaron mandíbulas muy cortas (flecha roja en Figura 40 D). Otra de las características que se aprecia en las fenocopias leves, es la pérdida y/o fusión de segmentos abdominales. (Figura 41 A-D). En la figura 41 B se muestra una imagen de confocal de una larva de estadio I, en la cual claramente los segmentos abdominales 4<sup>to</sup> y 5<sup>to</sup> se encuentran fusionados (Figura 41 B).

Los embriones con fenotipo interferido fuerte (N=143) están altamente modificados en comparación con el *wt* (Figura 41 C y D). Estos embriones consistentemente arrestaron su desarrollo a las 72 hs PO. La cabeza presenta el mismo fenotipo que para las fenocopias leves, donde se pierden las mandíbulas y maxilas pero el labio está siempre presente. Por otro lado, en estos embriones el fenotipo correspondiente a la región abdominal es más pronunciado, donde el número de segmentos abdominales se reducen a un total de cuatro segmentos. Se observaron defectos adicionales como, por ejemplo, el cierre dorsal incompleto, aunque esto se lo atribuye al arresto del desarrollo más que a un efecto de la ausencia de *Rp-gt*. En todos los casos se vieron afectados los apéndices torácicos, los cuales están deformados, en particular el fémur y la tibia. Esto podría deberse como un efecto indirecto de la falta de *Rp-gt* que afectaría a otros genes *gap* o *hox*, los cuales serían necesarios para la correcta formación de la patas.



**Figura 40.** Fenotipos *gt-ARNi*. **A-D.** Fenotipo cefálico. **A.** Imagen de confocal de una larva I interferida para *Rp-gt*. El detalle muestra un corte transversal de la proboscis, donde se puede ver la ausencia de estiletes en su interior. **B.** Esquema de la cabeza *wt* en vista dorsal. Aquí se marcan las diferentes estructuras que forman la región pre-gnata de la cabeza. **C-D.** Preparado cuticular de la cabeza *wt* (**C**) y *gt-ARNi* (**D**) donde se muestra la morfología del “clipeo-labro”. La flecha negra marca el “clipeo-labro”, el cual es más pequeño en el ejemplar interferido comparado con el *wt*. La flecha roja en D marca unas pequeñas mandíbulas presentes en ese ejemplar interferido. Abreviaciones: A, antena; E, ojo; Cl, clipeo; Lb, labio; Lm, labro; Mn, mandíbula; Mx, Maxila, J, jugum. Escala = 100 $\mu$ m.



**Figura 41.** Fenotipo abdominal. **A.** Vista lateral de una larva I *gt-ARNi*, la cual presenta fusión de segmentos abdominales. **B.** imagen de confocal de la misma larva en A, donde se muestran los segmentos 4 y 5 fusionados. **C.** Embrión con fenotipo fuerte. Imagen de un embrión con 72 hs PO en vista lateral, donde el labio es el único apéndice bucal y los segmentos abdominales se redujeron a cuatro. **D.** Vista lateral de un embrión *wt* en el mismo estadio de desarrollo que el embrión en C.

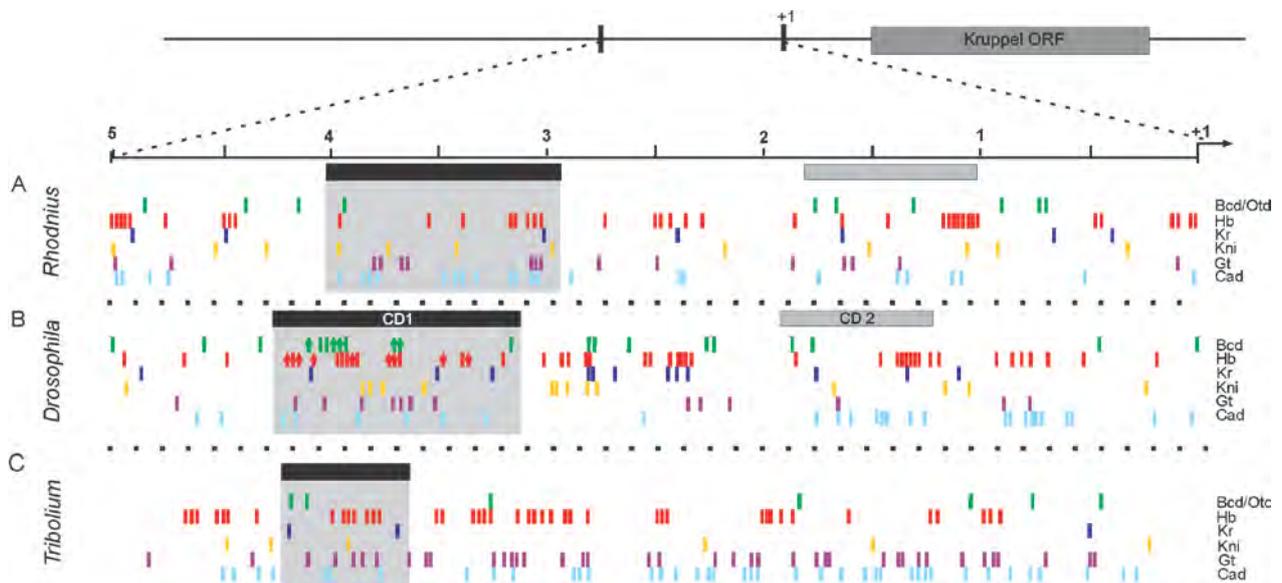
### 4.3.2 *Krüppel*

El gen *Rp-Kr* se identificó en una región genómica de 53,6 Kpb que contiene un ORF que codifica para una proteína de 339 aminoácidos con alta similitud de secuencia con ortólogos de ese mismo gen en otros insectos. Para la validación de la supuesta unidad transcripcional de *Kr* en *R. prolixus* se diseñaron *primers* específicos que hibriden en el primer y segundo exón del gen, para amplificar mediante PCR la región evolutivamente más conservada del gen. El producto de dicha PCR fue un amplicón de unas 336 pb, el cual incluye a los primeros cuatro dominios *zink-finger* (ZN-F) del gen. La secuencia fue subida al Gene Bank bajo el número de acceso GU724144.1.

#### 4.3.2.1 Análisis del promotor de *Rp-Kr*

Habiendo identificado un fragmento de 43 Kpb del genoma de *R. prolixus* que contiene al gen *Rp-Kr*, se realizó un análisis de la secuencia *upstream* a la región codificante en busca de posibles sitios de unión para factores de transcripción. El análisis bioinformático de las 5 Kpb *upstream* del origen de transcripción mediante el *software* PATSER resultó en la identificación de numerosos sitios de unión para algunos de los potenciales reguladores de *Rp-Kr*. La cantidad de sitios encontrados para cada uno de los reguladores fue la siguiente: Hb, 40; Bcd/Otd, 10; Cad, 26; Kr, 7; Kni, 12 y Gt, 16. Se pudo determinar la presencia de dos posibles *clusters* regulatorios: uno ubicado entre 1 y 1,7 Kpb y otro entre las 3 y 4 Kpb a partir del origen de transcripción. Este último fragmento contiene los siguientes sitios de unión: Bcd/Otd, 1; Hb, 8, Kr, 1; Kni, 4; Gt, 7 y Cad, 12 (Figura 42 A). La región reguladora de *Kr* en *D. melanogaster* fue descrita y caracterizada experimentalmente por Hoch y col. (1990 y 1991). En estos trabajos se describen dos dominios regulatorios: CD1 y CD2, de los cuales el primero es el que mayor actividad tiene durante las etapas tempranas de expresión de *Kr*. Para evaluar la fidelidad de nuestros resultados, se analizó el *clustering* de factores de transcripción a la secuencia *upstream* de *Dm-Kr* (como control) y de *Tc-Kr*. Para esto se usaron los mismos parámetros de análisis que en *R. prolixus*. Esto resultó en la identificación de un *cluster* regulatorio entre los 3,7 y 4,3 Kpb a partir del origen de transcripción en *T. castaneum* y en *D. melanogaster*, al igual que en los trabajos de Hoch y col., se observan dos potenciales *clusters*: uno entre los 3 Kpb y 4,3 Kpb (misma localización que el CD1) y otro entre los 1,3 y 2 Kpb (similar localización que CD2) a partir del nucleótido +1 (Figura 42 B y C). La predicción de los *clusters* regulatorios para *Dm-Kr* no sólo tiene una posición similar que los dominios CD1 y CD2, sino que también se predicen con una alta precisión los sitios de unión para Hb y Bcd. En la predicción realizada, los sitios de unión con mayor *score* para Bcd y Hb en el *cluster* CD1 se

encuentran en igual número y localización que los comprobados experimentalmente en los trabajos de Hoch y col. (sitios marcados con flechas en la figura 42 B). Comparando las predicciones sobre *D. melanogaster* y *R. prolixus*, se puede ver que los *clusters* predichos para *R. prolixus* tienen una localización similar a los fragmentos CD1 y CD2 de *D. melanogaster*. Las coincidencias entre la predicción para *D. melanogaster* y su comprobación experimental refuerzan las predicciones realizadas sobre *R. prolixus* y *T. castaneum*, y esto podría estar definiendo una potencial secuencia regulatoria conservada para *Rp-Kr*, *Tc-kr* y *Dm-Kr*.

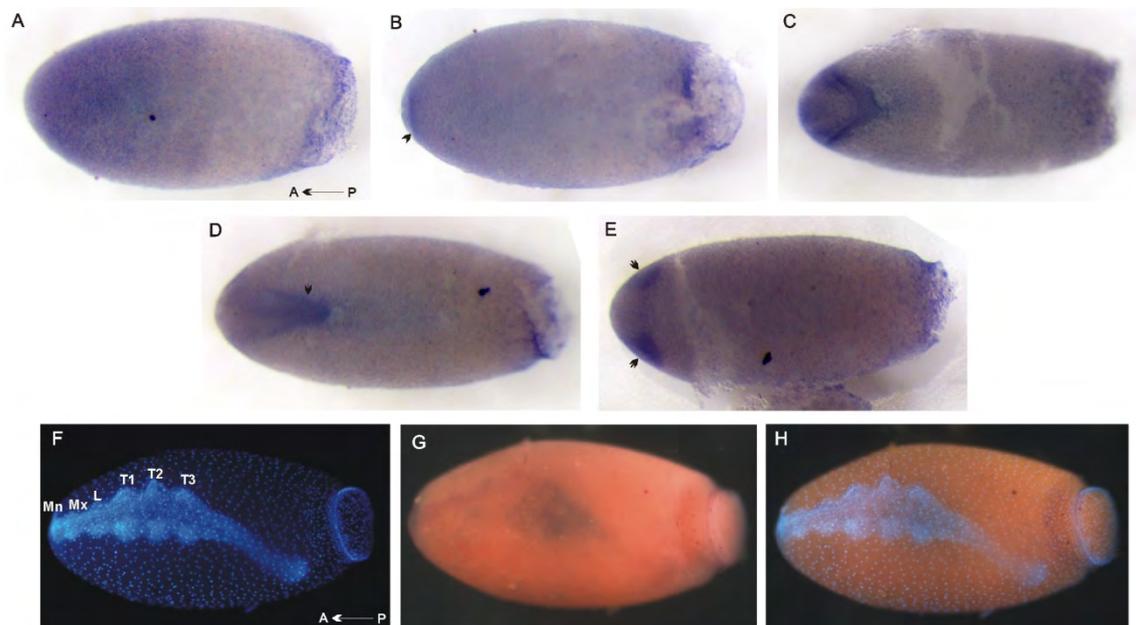


**Figura 42.** Análisis comparativo de la potencial secuencia promotora de *Krüppel* en *R. prolixus*, *D. melanogaster* y *T. castaneum*. Secuencia regulatoria de *R. prolixus* (A), *D. melanogaster* (B) y *T. castaneum* (C). A la derecha del gráfico se muestran los diferentes factores de transcripción. Las rayas de color representan los sitios de unión de los reguladores a la secuencia, los cuales fueron predichos por el *software* PATSER. Los *clusters* regulatorios son marcados con los rectángulos negros y grises. La escala marca la distancia en Kpb a partir del origen de transcripción. Los *clusters* regulatorios mostrados para *D. melanogaster* se marcan como CD1 y CD2 debido a la similitud en posición a los hallados en el trabajo de Hoch (1990). Los sitios marcados con flechas son aquellos que tienen la misma localización que los comprobados experimentalmente por Hoch y col. (1991).

#### 4.3.2.2 Patrón de expresión de *Rp-Kr*

Para la determinación del dominio de expresión de *Kr* en *R. prolixus*, se realizó una hibridación *in situ* con la sonda anti-sentido cuya secuencia corresponde a los cuatro dominios Zn-F más conservados en insectos. Se analizaron todos los estadios embrionarios considerados anteriormente, huevos con pocas horas post puesta, blastodermos, embriones en gastrulación, y bandas germinales, tanto en proceso de extensión como en extensión completa. La expresión de *Rp-Kr* es siempre cigótica, no evidenciándose presencia de transcritos antes de la 30 hs PO (Figura 43). La expresión cigótica se hace evidente en embriones con 30-36 hs PO, en los cuales

se observa la marcación en la región posterior del huevo. Estos transcritos se localizan junto a los núcleos embrionales que migran hacia el polo posterior del huevo durante el proceso de invaginación (Figura 43 A). Para este mismo estadio, pero en la cara dorsal del huevo, se observa una zona de mayor expresión sobre el borde de invaginación (Figura 43 B). Para las 36-40 hs PO, el dominio de expresión de *Rp-Kr* se retrae totalmente hacia la región posterior del huevo y se localiza exclusivamente en la zona de invaginación (Figura 43 C). Durante las primeras etapas del estadio de banda germinal (cerca de las 40 hs PO) el transcripto de *Rp-Kr* posee dos dominios de expresión: un dominio anterior, en la zona de los lóbulos cefálicos y un dominio central, en la zona donde posteriormente se desarrollarán los segmentos torácicos. En embriones entre 48 y 60 hs PO se observa una fuerte expresión de *Rp-Kr* (Figura 43 F-H). Los embriones, fueron co-tenidos con DAPI para datar morfológicamente los dominios de expresión (Figura 43 F). En éstos se ve claramente que la distribución del mensajero se localiza en la región media del embrión, en la zona correspondiente al segundo y tercer segmento torácico y los primeros segmentos abdominales (Figura 43 G y H). Este dominio está bien definido y hay poco fondo en el resto del embrión. En la siguiente sección se detallan los cambios morfológicos en los mutantes para *Rp-Kr*, los cuales coinciden con este patrón de expresión.



**Figura 43.** Expresión cigótica de *Rp-Kr*. **A-B.** Hibridación *in situ* de *Rp-Kr* en un embrión en invaginación (30-36 hs PO). **A.** Embrión en vista ventral, donde se ve la distribución del mensajero en el polo posterior del huevo. **B.** Mismo huevo que en A en cara dorsal. La flecha marca el borde de invaginación, donde la marcación es más intensa. **C.** Embrión de 36-40 hs PO en vista ventral, donde la marcación se localiza solo en la zona de invaginación. **D-E.** Embrión en banda germinal temprana (cerca de las 40 hs PO). Aquí la expresión de *Rp-Kr* se da en dos dominios, uno anterior y otro en la región media del embrión. **F-H.** Embrión con 48-60 hs PO. **F.** Tinción con DAPI, donde se marcan los apéndices torácicos y cefálicos. **G.** Hibridación *in-situ* de *Rp-Kr* sobre el mismo embrión en F. **H.** Imagen con doble exposición, epiiluminación (donde se muestra la distribución del transcripto de *Rp-Kr* en la *in situ*) y epifluorescencia donde se ve al embrión teñido con DAPI. Como se puede ver la marcación se da en los segmentos T 2-3 y en los primeros segmentos abdominales. Abreviaciones: Mn, mandíbulas; Mx, maxilas; Lb, labio; T1-T3, segmentos torácicos 1 a 3.

#### 4.3.2.3 Fenotipo *Kr* en *R. prolixus*

Con el fin de conocer el fenotipo resultante de la ausencia de expresión de *Rp-Kr*, se generaron fenocopias por ARNi parental. Se realizaron dos experimentos independientes, y como resultado las hembras pusieron un total de 279 huevos. Los embriones fueron colectados a diferentes tiempos post-puesta, y sólo una fracción (n=209) fue analizada para la determinación del fenotipo *Kr* en *R. prolixus* (Tabla 2). El 97,6% de los embriones interferidos mostraron un fenotipo diferente al *wt*. Estos fueron separados en dos fenotipos: leve (1,6%) y fuerte (96%). Todos los embriones llegaron a desarrollo completo, pero ninguno llegó a eclosionar.

A continuación se describe en detalle el fenotipo de los embriones interferidos, los cuales muestran alteraciones en la región torácica y en los primeros segmentos abdominales, región donde se expresa *Rp-Kr*. Para comparar los diferentes caracteres con el fenotipo *wt*, remitirse a la figura 39.

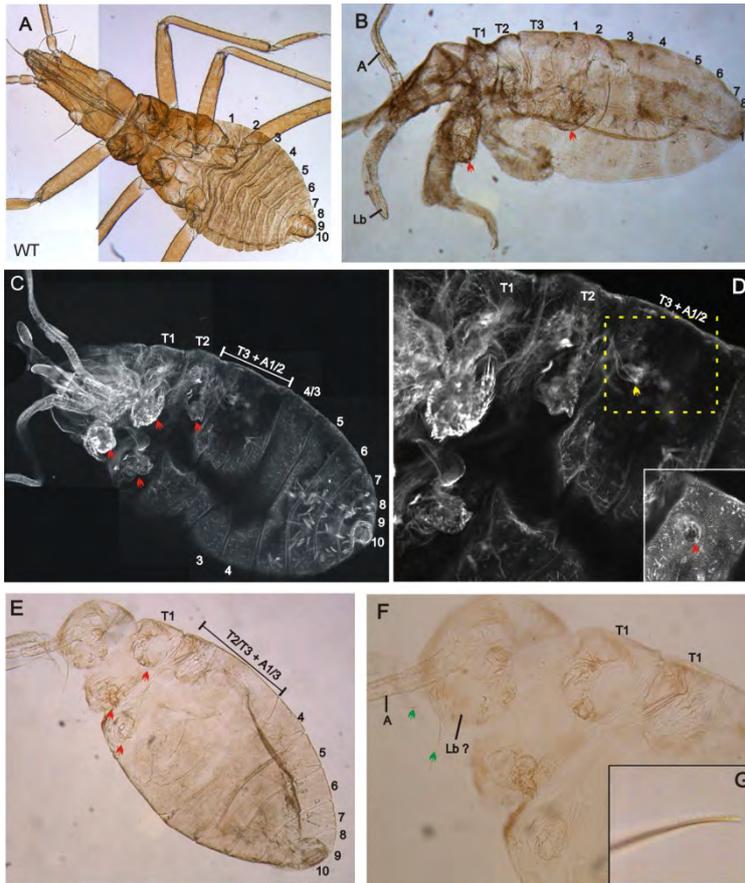
Tabla 3: Resultados de la RNAi parental de <i>Rp-Kr</i>				
	Fuerte (%)	Leve (%)	WT (%)	Total
ARNdc antisense	201 (96%)	3 (1,6%)	5 (2,4%)	209 (100%)

En embriones con fenotipo leve se ve fuertemente afectado el desarrollo de los apéndices torácicos. Estos embriones poseen todos los segmentos torácicos, pero muestran pérdida o modificación de sus apéndices (Figura 44 B). El primer y segundo segmento torácico desarrollan patas, las cuales estas están deformadas y reducidas en longitud. Las patas de T2 están aun más reducidas que las del segmento T1. Por otro lado, el segmento T3 se desarrolla normalmente pero carece de apéndices. Con esto se puede ver que el fenotipo se acentúa hacia la parte posterior del tórax, coincidente con la localización del dominio de expresión de *Rp-Kr*. El abdomen se desarrolla sin alteraciones, contando con el total de los segmentos abdominales. La región cefálica también se desarrolla normalmente. En las larvas con fenotipo leve ocurren ciertos cambios homeóticos que se muestran más adelante.

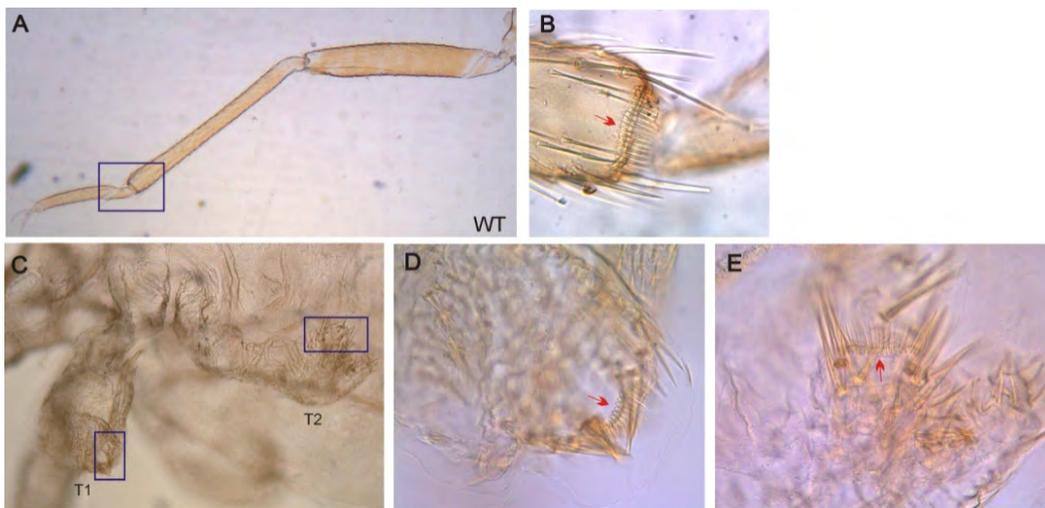
Los mutantes con fenotipo fuerte se encuentran altamente modificados. En estos, el fenotipo correspondiente al tórax se acentúa, y a esto se suman alteraciones en el cabeza y abdomen (Figura 44 C-G). Estas fenocopias sufren pérdida y fusión de segmentos, principalmente la fusión se da entre el segmento T3 y los primeros segmentos abdominales. En la figura 44 C y D se muestra uno de los embriones con fenotipo fuerte en donde el segmento T3 se encuentra fusionado a los primeros dos segmentos abdominales. La presencia del tercer segmento torácico se hace evidente mediante el uso de la microscopia confocal, gracias a la cual

se pueden ver los esbozos de la pata T3 (Figura 44 D). Este esbozo de la pata T3 puede verse claramente en vista ventral como un abultamiento (detalle en figura 44 D). Cuando se analiza internamente a esa estructura, se ve una serie de estructuras cuticulares (flecha amarilla en la figura) que corresponden a las apodemas de inserción muscular. Esto prueba que ese abultamiento es un esbozo apendicular y que el segmento T3 se encuentra formando parte del *gap* junto a los primeros segmentos abdominales. En todos los embriones la segmentación del abdomen se ve ampliamente afectada, reduciéndose a un total de siete (Figura 44 C y D). En la figura 44 D se puede ver que los segmentos 3 y 4 se fusionan parcialmente, mientras que el primero y el segundo se pierden o están fusionados con el último segmento torácico. En el fenotipo más extremo, se observa un gran *gap* desde los segmentos T2 a A3. En este *gap* la única estructura identificable corresponde al esbozo apendicular del segundo segmento torácico (Figura 44 D). En todas las larvas con fenotipo fuerte analizadas, se observa la deformación de la proboscis (Figura 44 F). Los cambios vistos en la proboscis afectan al labio, el cual se encuentra altamente modificado, mientras que los estiletes, tanto mandíbulas como maxilas, tienen un desarrollo normal (Figura 44 G). Los cambios en la región cefálica y en el primer segmento torácico, podrían deberse a cambios en los niveles de expresión de genes regulados por *Rp-Kr* que afectan a zonas corporales por fuera del dominio de expresión de este gen.

Como dijimos anteriormente, en las larvas con fenotipo leve, aparecen algunos cambios homeóticos. Estos cambios están relacionados con la aparición ectópica de peines tibiales. En los insectos salvajes, en el extremo distal de la tibia de la pata T1 aparecen un serie de pelos ordenados de forma paralela, formando una suerte de peine (Figura 45 A y B). En los mutantes de *Rp-Kr* aparecen peines tibiales ectópicos en la pata T2 (Figura 45 C-D). Como se muestra en la figura 45, estos mutantes poseen peines tibiales tanto en la pata T1 como en T2, lo cual podría deberse a una expresión ectópica del gen *Hox Scr*.



**Figura 44.** Fenotipo *Kr*-ARNi. **A.** Larva I *wt*, donde se muestra el número total de segmentos abdominales. **B.** Fenotipo leve. Larva I en vista lateral con la totalidad de los segmentos corporales, pero con apéndices torácicos reducidos o ausentes (pata T3 ausente). **C-G.** Fenotipo fuerte. **C.** Imagen de confocal de una larva I con fusión y pérdida de segmentos. Segmentos torácicos (T3) y abdominales (A1-4). **D.** Ampliación de la región torácica del embrión mostrado en C. El recuadro amarillo marca la posición del esbozo de pata T3 (detalle al pie de la imagen- T3 en vista superficial). La flecha amarilla marca las apodemas. **E.** Ninfa I con T2 y 3 perdidos o fusionados a los segmentos abdominales, los cuales se redujeron a un total de 7. Labio, al igual que en C, altamente modificado. **F.** Ampliación de la cabeza de la ninfa en E. Flechas verdes marcan los estiletes (en G se muestra el detalle de una de las mandíbula). Abreviaciones: A, antena; E, ojo; Cl, clipeo; Lb, labio; Lm, labro; Mn, mandíbula; Mx, Maxila, J, jugum. Escala = 100µm.



**Figura 45:** Cambios homeóticos en la pata del tercer segmento torácico de las larvas interferidas para *Rp-Kr*. **A.** Pata T1 de una larva I *wt*. El recuadro azul marca la localización del peine tibial. **B.** Detalle del peine tibial en el extremo distal de la tibia. **C.** Imagen donde se muestra la región torácica de una larva I interferida para *Rp-Kr*. Los recuadros azules marcan los peines tibiales. **D.** Detalle del peine presente en la pata T1. **E.** Detalle del peine ectópico presente en pata T2. T1: primer segmento torácico; T2: segundo segmento torácico; WT: embrión normal.

## 4.4 Discusión.

### 4.4.1 *Giant*

#### 4.4.1.1 Características conservadas del gen *gt*

El gen *giant* se encuentra altamente conservado en insectos. *Rp-gt* posee los dominios proteicos característicos de *gt* presentes en otras especies de insectos, como son el dominio de represión CtBP (Strunk *et al.*, 2001) y dominio *Leucine-Zipper* (Capovilla *et al.*, 1992). Además, posee un dominio proteico que comparte con *O. fasciatus*, el cual contiene un potencial sitio de fosforilación. La similitud de secuencia entre los diferentes ortólogos de *gt* se ve reflejada en el análisis filogenético realizado, donde *O. fasciatus* y *R. prolixus* forman un grupo monofilético (el orden Hemiptera), y junto con *P. humanus*, forman el grupo Hemimetábola. El agrupamiento de las diferentes especies en el árbol y los valores que soportan cada rama, generan un árbol filogenético robusto que indica que la evolución de la secuencia codificante de *gt* acompañó a la evolución del grupo.

La similitud de secuencia del gen dentro de los insectos se extiende también a la secuencia regulatoria. La regulación de la expresión de *gt* durante el estadio de blastodermo ha sido estudiada ampliamente en *D. melanogaster*, donde se identificaron diferentes factores de transcripción con función activadora y represora (Capovilla *et al.*, 1992; Kraut y Levin, 1991). La región reguladora responsable del control de la expresión de *Dm-gt* no había sido definida sino hasta hace poco tiempo. Berman y col. (2002) predijeron y validaron una región *upstream* al inicio de transcripción del gen, la cual es responsable de la regulación del dominio de expresión posterior en *D. melanogaster*. En esta tesis se definió un potencial *cluster* de regulación en *Rp-gt*, utilizando el mismo acercamiento experimental y los mismos parámetros usados por Berman y col. Este *cluster* es similar al descrito para *D. melanogaster*, tanto en posición como en la representatividad de los sitios de unión para los factores de transcripción. Se realizó el mismo análisis para la secuencia *upstream* a *Tc-gt*, lo cual resultó en un *cluster* con las mismas características que los de *R. prolixus* y *D. melanogaster*. Con todo esto, se puede ver que los ortólogos de *gt* no sólo muestran similitudes a nivel de su secuencia proteica, sino que también a nivel de las regiones no codificantes (a este nivel es más importante la funcionalidad de la misma que la similitud de secuencia), en particular relacionado a la potencial región reguladora del dominio abdominal.

#### 4.4.1.2 Expresión materna de *Rp-giant*

En *D. melanogaster*, los genes *gap* son expresados cigóticamente (a excepción de *hb*) y se activan durante el 13<sup>er</sup> ciclo de clivaje del embrión. En algunos insectos, como son *N. vitripennis* y *T. castaneum*, se detectó que algunos genes *gap* se expresan de forma materna y regulan el inicio del proceso de segmentación (Schroder, 2003; Pultz *et al.*, 2005; Lynch *et al.*, 2006; Brent *et al.*, 2007). Trabajos previos en *A. mellífera* y *N. vitripennis* documentaron una expresión materna semejante en particular del gen *giant* (Olesnick y Desplan, 2007; Wilson *et al.*, 2010). En *N. vitripennis* se demostró que *gt* cumple la función de determinante anterior como resultado de un efecto indirecto sobre la expresión de *hb* (Brent *et al.*, 2007). En insectos de banda germinal corta e intermedia no se observó hasta el momento expresión materna de *gt*, aunque Bucher y Klingler infieren una contribución materna en *T. castaneum* (Bucher y Klingler, 2004). Esta presunción no se muestra claramente, y solo se basa en observaciones de marcación en embriones con pocas horas PO. En esta tesis se mostró que *Rp-gt* se acumula durante el transcurso de la ovogénesis. *Rp-gt* se expresa maternamente en el trofario, en oocitos tempranos y en huevos con pocas horas PO. Trabajos previos estudiaron la estructura, citología y fisiología del ovario telotrófico de *R. prolixus* (Huebner y Anderson, 1972a; Huebner y Anderson, 1972b; Huebner, 1981; Bjorsson y Huebner, 2004). Durante la ovogénesis se da lugar a la síntesis de ácidos nucleicos, ARN, vitelo y proteína por parte de dos tejidos: el trofario y las células foliculares que rodean al oocito (Vanderberg, 1963). Todos estos productos son transportados desde el trofario (a través del “cordón trófico”) y las células foliculares hacia los oocitos. Para cada estadio de oocito la carga de ADN, ARN y proteínas es diferente, y esto es proporcional al diámetro de cordón trófico. En oocitos en vitelogénesis avanzada el calibre del cordón trófico es mayor que en oocitos más inmaduros, de forma tal que la incorporación de ac. nucléicos y proteína va a ser mayor (Vanderberg, 1963). Estos datos son importantes, y comprender esto es necesario para interpretar los resultados mostrados en relación a la expresión materna de *gt* en *R. prolixus*.

La expresión de *Rp-gt* en ovariolas está restringida al trofario, oocitos tempranos, y oocitos en vitelogénesis avanzada. La marcación que se observa en el trofario es probablemente resultado de la síntesis de transcritos en ese tejido, los cuales son posteriormente transportados a los oocitos. Los oocitos tempranos, en los cuales se detectó mensajeros de *Rp-gt*, corresponden a los clasificados como oocitos T-3 (de acuerdo con la nomenclatura usada por Bjorsson y Huebner, 2004). Los altos niveles de expresión en estos oocitos se atribuyen al estrecho agrupamiento de los oocitos en esa región de la ovariola más que a la elevada carga de mensajeros. En oocitos en vitelogénesis avanzada (O-va) se observa la presencia de mensajeros de *Rp-gt*, mientras que no hay marcación en oocitos que se encuentran en la fase final de desarrollo (O). Esto se atribuye a que la transferencia direccional de vitelo y ácidos nucleicos a

través del cordón trófico es mayor en oocitos en vitelogenénesis avanzada. La hibridación *in situ* de *Rp-gt* en embriones en estadio de preblastodermo (con pocas horas PO, antes de la aparición de los núcleos en la superficie del huevo), muestra un patrón de distribución disperso y en forma de parches. Estos “manchones” de mensajeros de *Rp-gt* se acumulan en la corteza del huevo en islas citoplasmáticas entre los gránulos de vitelo. Para las 12 hs PO, momento en el cual los núcleos han migrado a la periferia del huevo (blastodermo sincitial), la localización del transcripto deviene en una distribución más homogénea y perinuclear. Esta distribución perinuclear de *gt* también fue observada en *N. ventripennis* durante la ovogénesis (Olesnický y Desplan, 2007). De esto se puede hipotetizar que la distribución perinuclear de los mensajeros de *Rp-gt* en el cortex del huevo resultaría de una redistribución de los mensajeros durante la formación del blastodermo sincitial. No se puede establecer si la co-localización de los mensajeros de *Rp-gt* con los núcleos es una causa o consecuencia del proceso de migración nuclear. Se puede pensar en esto como un proceso estocástico, en donde algunos núcleos caen dentro de los “islas” de transcripto y como consecuencia de esto se desencadenarán ciertos programas de expresión que influirán en el desarrollo embrionario.

#### 4.4.1.3 Expresión cigótica de *Rp-gt* y el fenotipo *gap giant*.

En *D. melanogaster*, luego del 13<sup>er</sup> ciclo de clivaje del embrión, *gt* se expresa en tres dominios: el dominio más anterior (91 a 97% de la longitud del huevo), el dominio cefálico (75 a 83%) y el dominio abdominal (25 a 33% de la longitud del huevo, según Mohler *et al.*, 1989). Estos dominios de expresión se correlacionan con el fenotipo observado: el clípeo-labro (defectos del dominio más anterior), esqueleto céfalofariengeo y labio (defectos del dominio cefálico) y segmentos abdominales (A5-A7), producto de defectos en el dominio de expresión abdominal. En insectos de banda germinal intermedia como *T. castaneum* y *O. fasciatus*, se describieron dos dominios de expresión: uno anterior (para los segmentos gnatales) y otro posterior (para la zona de crecimiento). En *T. castaneum*, el ARNi parental resultó en un típico fenotipo *gap* abdominal y en transformaciones homeóticas para los segmentos maxilar y labial, no siendo requeridos para la formación de estos segmentos (Bucher y Klingler, 2004). En *O. fasciatus*, los embriones interferidos muestran el fenotipo *gap* clásico, donde se pierden tanto los segmentos gnatales (maxilar y labial) como los abdominales (Liu y Patel, 2010). Sumado a esto se ven algunos cambios homeóticos, como es la aparición de un peine tibial ectópico en la pata T3, producto de la expresión ectópica de *Scr* en el tercer segmento torácico. Nuestros datos muestran un fenotipo *gap* similar en *R. prolixus*, pero no se observan cambios homeóticos de este tipo en los embriones interferidos para *Rp-gt*.

En esta tesis se determinaron dos dominios de expresión cigótica para *giant* en *R. prolixus*: el dominio de expresión anterior (el cual se divide en dos) y el posterior (que corresponde al abdomen y zona de crecimiento). El dominio de expresión anterior es complejo y se puede dividir en dos sub-regiones: el dominio de expresión más anterior y el dominio medio, lo cual hace recordar a los dominios de expresión en *D. melanogaster*. Los resultados del ARNi parental muestran que *Rp-gt* es necesario para la correcta formación de la parte más anterior de la cabeza, maxilas, mandíbulas y el abdomen. Trabajos previos describen defectos en la parte más anterior de embriones de *N. vitripennis* y *A. mellifera* como respuesta a la ausencia de *gt*, pero en éstos *giant* actúan como determinante anterior, reemplazando a *bcd* en su función (Brent *et al.*, 2007; Lynch *et al.*, 2006). En *R. prolixus*, *gt* no tiene esta función, si no que está involucrado en la formación de la cápsula cefálica (clípeo-labro), maxilas y mandíbulas comportándose como un gen *gap*. De esta forma, aquí se muestra claramente que *Rp-gt* es un verdadero gen *gap*, el cual afecta la región cefálica y abdominal, y como en *D. melanogaster*, es requerido para la formación de la parte más distal de la cabeza, el clípeo-labro.

#### 4.4.2 *Krüppel*

Como parte de los resultados que se muestran en esta tesis se presenta el patrón de expresión y función del gen *Rp-Kr*, gen que es imprescindible para el correcto desarrollo de la región torácica y abdominal de *R. prolixus*. El dominio de expresión observado y el fenotipo de falta de función obtenido por ARNi confirman que *Rp-Kr* es un verdadero ortólogo del gen *Krüppel*. Al mismo tiempo se analizó una supuesta región reguladora *upstream* del origen de transcripción, la cual posee similitudes en localización y composición de sitios de unión a factores de transcripción con *D. melanogaster* y *T. castaneum*, pudiendo representar esto un potencial promotor del *Rp-Kr*. Muchas de las características de *Krüppel* observadas en *R. prolixus* también son observadas en otros insectos y esto se discute a continuación.

##### 4.4.2.1 La conservación del gen *Krüppel* en insectos.

El gen *Krüppel* está altamente conservado en insectos. En *R. prolixus*, *Kr* está formado por tres exones, los cuales codifican para una proteína que contiene los cinco dominios Zn-F descritos para *D. melanogaster* (Rosenberg, 1986). Estos dominios están muy conservados en otros insectos para los cuales *Kr* fue analizado (Liu y Kaufman, 2004; Cerny *et al.*, 2005; Mito *et al.*, 2006 y Wilson *et al.*, 2010). A partir del análisis realizado para *Rp-Kr* se pudo definir un nuevo dominio proteico compartido por la gran mayoría de los insectos, para el cual no se pudo establecer función biológica alguna. Esta similitud de secuencia de *Kr* se extiende también a la secuencia regulatoria. La regulación de *Kr* ha sido estudiada ampliamente en *D. melanogaster*, tanto en estadios tempranos del desarrollo como tardíamente (Hoch *et al.*, 1990). En estos trabajos se describe detalladamente los dominios regulatorios encargados de la correcta expresión de *Kr* en la región central del embrión. Se describieron dos *clusters* regulatorios para *Dm-Kr*, CD1 y CD2. El primero de estos, actúa en las primeras etapas de expresión del gen, mientras que el segundo interviene más tardíamente y los niveles de expresión dependientes de él son menores que los de CD1. Este último fue estudiado en mayor detalle y mediante análisis *in vitro* e *in vivo* se determinó la composición de sitios de unión a los factores de transcripción Hb y Bcd en este fragmento de ADN (Hoch *et al.* 1991). Se localizaron 10 sitios de unión para Hb y 6 para Bcd a lo largo del fragmento de 730 pb identificado como CD1. Como dijimos anteriormente, en esta tesis se determinó la presencia de dos potenciales *clusters* regulatorios de *Rp-Kr*. En paralelo se analizó la región *upstream* de *Tc-Kr* y *Dm-Kr*, tomando a este último como control, dado que gracias a los trabajos de Hoch y col. se conoce con exactitud la localización de los mismos. Como resultado, se vio que la localización de los *clusters* para cada una de las tres especies consideradas es similar. La predicción de dominios regulatorios sobre *Dm-Kr* mostró grandes similitudes con los dominios regulatorios comprobados

experimentalmente por Hoch y col. Esta similitud no es solo a nivel de la localización de los dominios CD1 y CD2, sino también en la posición donde se encuentran los sitios de unión para Hb y Bcd. Hasta el momento no se había realizado un análisis de predicción de sitios de unión de reguladores en *Dm-Kr*, y la precisión en la predicción sobre el *cluster* de *Dm-Kr*, hace que la predicción para *Rp-Kr* tenga más peso.

La regulación transcripcional de los genes *gaps* ha sido estudiada con gran detalle en *D. melanogaster*, mientras que la información en otros insectos todavía es escasa. La secuencia involucrada en la regulación transcripcional de *hb* y *cad* fue analizada y comparada entre *D. melanogaster* y *T. castaneum*, como así también entre diferente especies de dípteros (Wolff *et al.*, 1998; Shaw *et al.*, 2002). En esta tesis se presenta, el primer análisis comparativo de la región genómica *upstream* al origen de transcripción de *giant* y *Krüppel* en diferentes especies de insectos. El análisis realizado aquí, sugiere la existencia potenciales dominios regulatorios en gen *gt* y *Kr*. Actualmente se están realizando experimentos *in vitro* para comprobar la presencia de los sitios de unión de los diferentes factores de transcripción en estas supuestas regiones reguladoras.

#### 4.4.2.2 Expresión cigótica y fenotipo de *Rp-Kr*

En *D. melanogaster*, la expresión de *Kr* es siempre cigótica y se da en la región media del embrión cuando este transcurre en el 13<sup>er</sup> ciclo de clivaje. Este dominio central se localiza en los últimos dos segmentos torácicos y el primer segmento abdominal (Jackle *et al.* 1995). La falta de expresión resulta en una serie fenotípica que abarca desde embriones con la pérdida de segmentos torácicos e inversión de la polaridad de segmentos abdominales, hasta el fenotipo más extremo con ausencia de segmentos torácicos y conservando sólo los últimos tres segmentos abdominales (Wieschaus *et al.*, 1984 y Preiss *et al.*, 1985) (Figura 46). En *A. mellifera* (otro insecto de banda germinal larga) la expresión de *Am-Kr* se hace evidente durante el estadio de blastodermo sincitial en la región central del embrión. En los sucesivos estadios este dominio central se desdobra en tres, y surgen dos dominios más: uno en el extremo anterior y otro en el posterior. A partir del 6<sup>to</sup> estadio de desarrollo, la expresión se hace generalizada. Los embriones de *A. mellifera* carentes de *Kr* sufren alteraciones en la segmentación de la zona central del cuerpo, donde se fusionan los segmentos T1-A6 formando un gran *gap* (Wilson *et al.*, 2010). El gen *Kr* también fue estudiado en insectos de banda germinal corta (Liu y Kaufman, 2004; Cerny *et al.*, 2005; y Mito *et al.*, 2006). En *T. castaneum*, *Kr* se expresa durante el estadio de blastodermo en la parte más posterior del huevo y durante el estadio de banda germinal temprana en la región correspondiente a T1-T3. Luego de este estadio, la expresión se hace generalizada (Cerny *et al.*, 2005). La falta de expresión de *Kr* en *T. castaneum*

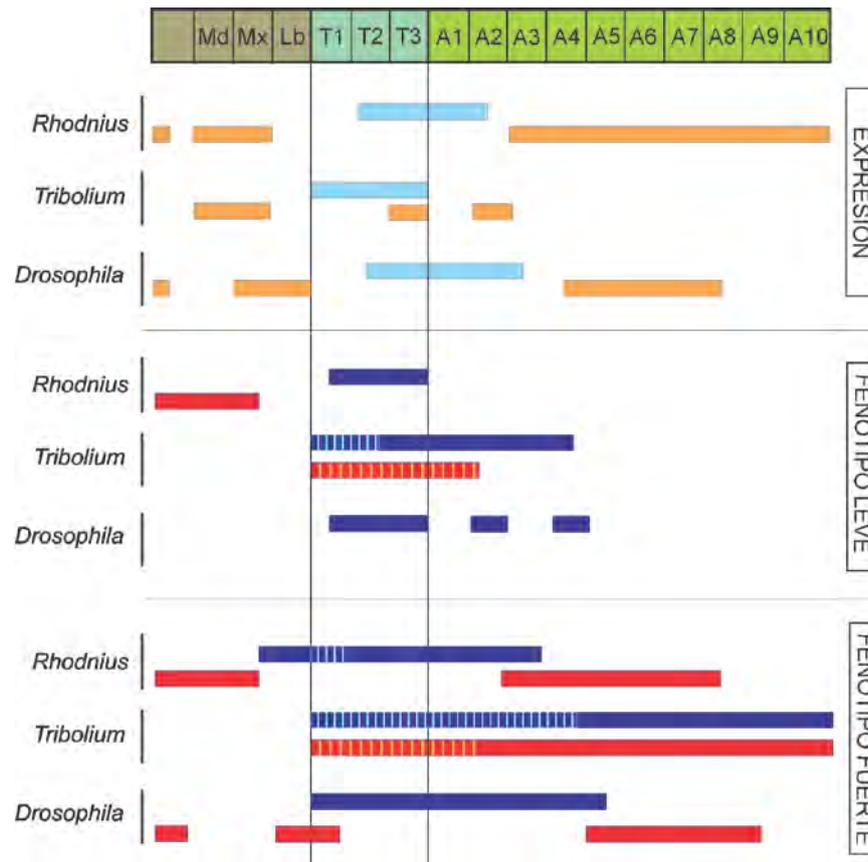
causa pérdida y fusión de segmentos torácicos y abdominales. Sumado a esto, se dan ciertos cambios homeóticos como ser la aparición de apéndices con identidad cefálica (principalmente Mx y Lb) en segmentos torácicos y abdominales (Figura 46). Dentro de hemimetábolos el gen *Kr* fue estudiado en *G. bimaculatus* y *O. fasciatus*. En ambos, la expresión está delimitada solamente a la región torácica, abarcando los segmentos T1-T3 (en *Gryllus*, el dominio de expresión de *Kr* se expande anteriormente hacia el segmento labial). La ausencia de expresión de *Kr* en *G. bimaculatus* altera la segmentación del embrión, desde la pérdida de los segmentos abdominales más anteriores hasta la pérdida de la gran mayoría de los segmentos torácicos y abdominales, quedando en el fenotipo más extremo solo los últimos tres segmentos abdominales. En éstos siempre la región cefálica se desarrolla sin modificaciones, contando con la totalidad de los apéndices cefálicos. En *O. fasciatus* la falta de *Kr* en el embrión produce la pérdida de gran parte de la región media del cuerpo, desde el segmento T2-A6. En estos embriones la región cefálica tampoco se ve afectada (Liu y Kaufman, 2004).

En *R. prolixus*, la expresión de *Rp-Kr* se da desde temprano en el desarrollo, en donde entre las 30 y 36 hs PO se observa una importante expresión en la región posterior del huevo. Este transcripto se localiza junto con los núcleos embrionales en migración hacia el polo posterior del huevo durante el proceso de invaginación. Para este mismo estadio, en la cara dorsal del huevo, se ve una zona de mayor expresión sobre el borde de invaginación. Una expresión semejante se observa en *T. castaneum* y *O. fasciatus*. Luego, el dominio de expresión de *Rp-Kr* se retrae hacia la zona posterior del huevo y se localiza exclusivamente en la zona de invaginación. Durante el estadio de banda germinal temprana, *Rp-Kr* se expresa en dos dominios diferentes: uno más anterior en la zona de los lóbulos cefálicos y otro en la zona donde posteriormente se desarrollarán los segmentos torácicos. El dominio de expresión anterior aparece en otros insectos pero es temporal y la ausencia del mismo no genera cambios apreciables en el patrón corporal del embrión (Liu y Kaufman, 2004; Cerny *et al.*, 2005; y Mito *et al.* 2006). Para el estadio de banda germinal avanzada, el único dominio de expresión que persiste es el dominio central, el cual abarca los últimos dos segmentos torácicos y los primeros abdominales. Como se dijo anteriormente, y al igual que en el resto de los artrópodos para los cuales se estudio el gen *Kr*, la ausencia de expresión de este gen en *R. prolixus* ocasiona la pérdida y fusión de los segmentos torácicos y abdominales causando la formación de un *gap* en la región central del embrión. Esto nos recuerda al típico fenotipo *gap* de los homocigotas de *Krüppel* en *D. melanogaster* (Wieschaus *et al.* 1994). En *R. prolixus* se ve una clara influencia torácica de la acción de *Krüppel* sobre el embrión más que en otros insectos. Esto se postula en base a que el fenotipo *Kr*-ARNi leve sólo afecta la región torácica, no influyendo en la formación del abdomen. El abdomen se ve afectado en los embriones con fenotipo fuerte, donde queda reducido a los últimos siete segmentos. Esto podría estar indicando que la región torácica

del embrión es más sensible a la regulación por parte de *Kr* que la región abdominal, siendo la región torácica más sensible a pequeños cambios en la carga de *Kr* en el embrión.

En *T. castaneum*, la ausencia de *Kr* genera, además de la pérdida de segmentos corporales, cambios homeóticos donde los apéndices torácicos y abdominales toman identidad de apéndices cefálicos, principalmente Mx y Lb (Cerny *et al.*, 2005). En *O. fasciatus*, la ausencia de expresión de otro gen *gap*, en este caso *gt*, genera la aparición de un peine tibial ectópico en la tibia de la pata correspondiente al tercer segmento torácico, como producto de la expresión ectópica del gen *hox scr* en ese segmento (Liu y Patel, 2010). Aquí, en *R. prolixus*, también se dan cambios homeóticos en relación a la ausencia de *Rp-Kr*, donde se desarrolla un peine tibial ectópico en el extremo distal de la tibia del segundo segmento torácico. Quizás en *R. prolixus* se esté produciendo una serie de eventos de desregulación génica ante la ausencia de *Rp-Kr* que generan la expresión ectópica de *scr* en el segundo segmento torácico.

En insectos existen una gran variedad de estrategias que determinan un mismo patrón corporal. Esto se ve reflejado en los diversos efectos que produce la falta de expresión de los genes de segmentación durante el desarrollo embrionario. Es así como, en algunos insectos, la ausencia de *Krüppel* o *giant* causa cambios morfológicos diferentes, aún en organismos evolutivamente emparentados. Esto demuestra que existe una gran plasticidad biológica que, a lo largo de la evolución, adapta la carga genética y la maquinaria regulatoria a la morfología de cada organismo de forma diferente, favoreciendo en todos los casos a la conservación de un mismo patrón morfológico, en el caso de los insectos los clásicos tagmatas, cabeza, tórax y abdomen.



**Figura 46:** Comparación de la expresión y fenotipo de *Kr* y *gt* en *R. prolixus*, *T. castaneum* y *D. melanogaster*: **A.** Dominios de expresión de *Kr* (barras celestes) y *gt* (barras naranjas). **B-C.** Regiones corporales afectadas por la falta de expresión de *Kr* (barras azules) y *gt* (barras rojas). En B se muestra el fenotipo leve para cada una de las especies, mientras que en C se muestra el fenotipo fuerte. Las barras azules y rojas sólidas marcan las regiones afectadas donde se produce pérdida de segmentos o alteraciones de los mismos. Las barras rayadas indican regiones corporales que sufren cambios homeóticos.

## 5 Conclusión

En esta tesis se mostró el análisis de la primer genoteca normalizada de *R. prolixus*. Si bien está conformada por una gran cantidad de fragmentos de corta longitud, la gran mayoría de los transcritos secuenciados e identificados corresponden a mensajeros de copia única o con dos copias dentro de la genoteca. A su vez, dentro de estos se encuentran transcritos relacionados al metabolismo celular básico, los cuales normalmente se encuentran en gran cantidad. Esto indicaría que el proceso de normalización al cual fue sometida esta genoteca fue efectivo. Actualmente, esta genoteca está siendo resecuenciada, utilizando en esta ocasión la tecnología de pirosecuenciación. Con esto se espera alcanzar una mayor cobertura de los genes expresados durante todo el ciclo de vida de *R. prolixus* y así utilizar estos datos para la corrección y finalización de la anotación del genoma de este insecto, la cual ha demostrado ser compleja (Daniel Lawson-EBI, comunicación personal).

Por otro lado, se realizó la búsqueda e identificación de la mayoría de los genes involucrados en el proceso de segmentación del embrión de *R. prolixus*. Se identificaron la totalidad de los genes *gap*, *gt*, *kr*, *kni*, *hb*, *tll*, *otd*, *fkf* y *ems*. Sólo el gen *gap* cefálico *btd* no pudo ser identificado con certeza, debido a su gran similitud de secuencia con los factores de transcripción tipo *Sp1-9*. Además de los genes *gap*, también se identificaron algunos de los genes *pair-rule* (*h*, *run* y *eve*) y el gen de polaridad de segmentos *en*. Este representa el primer análisis de genes de segmentación en insectos. Se espera a corto plazo realizar un *screening* del patrón de expresión de cada uno de estos genes, además de un *screening* fenotípico mediante ARNi.

Haciendo foco en los genes *gap* se caracterizó funcionalmente a dos de estos genes: *Rp-gt* y *Rp-Kr*. Para estos dos genes se identificó una potencial región reguladora *upstream* al origen de transcripción la cual estaría conservada en *D. melanogaster* y *T. castaneum*. En función de esto se espera realizar el análisis *in vitro* e *in vivo* de estas regiones genómicas, de forma tal de corroborar la presencia y funcionalidad de los sitios predichos. Como se mostró en la última parte de esta tesis, estos dos genes están implicados principalmente en la formación la parte central del embrión. *Rp-gt* se expresa en tres dominios corporales: el primero en la parte más anterior (clípeo-labro), el segundo en la región gnatal (mn y mx) y el tercero en el abdomen. La ausencia de expresión de este gen causa drásticos cambios en la segmentación del embrión, produciendo la reducción de los segmentos abdominales a un total de cuatro, la pérdida de los estiletes bucales y el acortamiento del clípeo-labro. *Rp-Kr* se expresa en la región central del embrión, la cual corresponde a los segmentos torácicos T2, T3 y los primeros segmentos abdominales. La ausencia de expresión de *Kr* causa severos cambios en la segmentación, principalmente en la región torácica y abdominal, manteniéndose sólo el primer

segmento torácico y los últimos seis segmentos abdominales. Como se puede ver, ambos genes (*Rp-gt* y *Rp-Kr*) poseen dominios de expresión compartidos, principalmente la región abdominal. Esto hace posible que ambos genes puedan regularse mutuamente. Interacciones de este tipo entre estos dos genes fueron ampliamente documentados en la bibliografía, principalmente en *D. melanogaster*. Con esto, se espera a corto plazo poder analizar el patrón de expresión de estos genes en embriones interferidos tanto para *Rp-gt* como para *Rp-Kr*, de modo tal de poder evaluar el tipo de interacción existente entre ellos (represión o activación) en función de los cambios de expresión para cada gen. De esta forma, contando con los resultados del *screening* del resto de los genes de segmentación identificados en esta tesis y el establecimiento de sus interacciones, se espera poder generar un modelo de regulación del proceso de segmentación en *R. prolixus*.

El descubrimiento de la gran mayoría de los genes de segmentación y la identificación, clonado y caracterización funcional de *Rp-Kr* y *Rp-gt* a partir de una genoteca normalizada de *R. prolixus*, representan los primeros estudios en genética del desarrollo en *Rhodnius prolixus*, impulsando a este insecto vector como un nuevo modelo para el estudio de los procesos genéticos-moleculares que determinan el patrón corporal en insectos.



## 6 Bibliografia

- Adams M. D., Celniker S. E., Holt R. A., Evans C. A., Gocayne J. D., Amanatides P. G., Scherer S. E., Li, P. W., Hoskins, R. A., Galle R. F., *et al.*, 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185-2195.
- Anisimova, V. E., Rebrikov, D. V., Zhulidov, P. A., Staroverov, D. B., Lukyanov, S. A., Shcheglov, A. S., 2006. Renaturation, activation, and practical use of recombinant duplex-specific nuclease from Kamchatka crab. *Biochemistry (Mosc)*. 71:513-519.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215. 403-410.
- Arensburger P, Megy K., Waterhouse, R. M., Abrudan, J., Amedeo P., Antelo, B., Bartholomay, L., *et al.*, 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science*. 330:86–88.
- Arnosti, D. N., Barolo, S., Levine, M., Small, S., 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development*. 122, 205-14.
- Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M., Eisen, M. B., 2002. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl. Acad. Sci. USA*. 99, 757-62.
- Bjornsson, C. S., y Huebner, E., 2004. Extracellular H<sup>+</sup> dynamics during oogenesis in *Rhodnius prolixus* ovarioles. *J. Exp. Boil.* 207, 2835-44.
- Bonaldo, M. F., Lennon, G., Soares, M. B., 1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res*. 6:791-806.
- Brent, A. E., Yucel, G., Small, S., Desplan, C., 2007. Permissive and instructive anterior patterning rely on mRNA localization in the wasp embryo. *Science*. 315, 1841-3.
- Bucher, G., Scholten, J., Klingler, M., 2002. Parental RNAi in *Tribolium* (Coleoptera). *Current Biology*. 12, R85-R86.
- Bucher, G., y Klingler, M., 2004. Divergent segmentation mechanism in the short germ insect *Tribolium* revealed by giant expression and function. *Development*. 131, 1729-40.
- Capovilla, M., Eldon, E. D., Pirrotta, V., 1992. The giant gene of *Drosophila* encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development*. 114, 99-112.
- Carroll, S. B., Laughon, A., and Thalley, B. S., 1988. Expression, function, and regulation of the hairy segmentation protein in the *Drosophila* embryo. *Genes Dev*. 2, 883-90.
- Cerny, A. C., Bucher, G., Schroder, R., Klingler, M., 2005. Breakdown of abdominal patterning in the *Tribolium Krüppel* mutant jaws. *Development*. 132, 5353-5363.

- Chomczynski, P., y Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156-9.
- Damen, W. G., Weller, M., Tautz, D., 2000. Expression patterns of *hairy*, *even-skipped*, and *runt* in the spider *Cupiennius salei* imply that these genes were segmentation genes in a basal arthropod. *Proc. Natl. Acad. Sci. USA.* 97, 4515-9.
- Davis, G. K., Patel, N. H., 2002. Short, long, and beyond: molecular and embryological approaches to insect segmentation. *Annu. Rev. Entomol.* 47, 669-99.
- Dekanty, A., 2008. Identificación y caracterización de nuevos genes necesarios para la respuesta transcripcional a hypoxia. Tesis doctoral. UBA. FCEYN.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465-9.
- Eldon, E. D., y Pirrotta, V., 1991. Interactions of the *Drosophila* gap gene giant with maternal and zygotic pattern-forming genes. *Development.* 111, 367-78.
- Evans, J. D., y Gundersen-Rindal, D., 2003. Beesomes to *Bombyx*: future directions in applied insect genomics. *Genome Biol.* 4:107.
- Gaul, U., Seifert, E., Schuh, R., Jackle, H., 1987. Analysis of *Krüppel* protein distribution during early *Drosophila* development reveals posttranscriptional regulation. *Cell.* 50, 639-647.
- Heckel, D. G., 2003. Genomics in pure and applied entomology. *Annu. Rev. Entomol.* 48:235-260.
- Heming B. S., 2003. Insect development and evolution. Cornell Univ Press.
- Hoch, M., Schroeder, C., Seifert, E., Jackle, H., 1990. Cis-acting control element for *Krüppel* expression in the *Drosophila* embryo. *EMBO J.* 9, 2587-2595.
- Hoch, M., Seifert, E., Jackle, H., 1991. Gene expression mediated by cis-acting sequences of the *Krüppel* gene in response to the *Drosophila* morphogens *bicoid* and *hunchback*. *EMBO J.* 10, 2267-2278.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., Bollback, J. P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science.* 294, 2310-4.
- Huebner, E., 1981. Nurse cell-oocyte interaction in the telotrophic ovarioles of an insect, *Rhodnius prolixus*. *Tissue Cell.* 13, 105-125.
- Huebner, E., y Anderson, E., 1972a. A cytological study of the ovary of *Rhodnius prolixus*. I. The ontogeny of the follicular epithelium. *J. Morph.* 136, 459-494.
- Huebner, E., y Anderson, E., 1972b. A cytological study of the ovary of *Rhodnius prolixus*. II. Oocyte differentiation. *J. Morph.* 137, 385-416.
- Huebner, E., y Anderson, E., 1972c. A cytological study of the ovary of *Rhodnius prolixus*. III. Cytoarchitecture and development of the trophic chamber. *J. Morph.* 138, 1-40.

- Hulskamp, M., y Tautz, D., 1991. Gap genes and gradients--the logic behind the gaps. *Bioessays*. 13, 261-268.
- Ingham, P. W., Ish-Horowicz, D., y Howard, K. R., 1986. Correlative changes in homoeotic and segmentation gene expression in *Krüppel* mutant embryos of *Drosophila*. *EMBO J.* 5, 1659-1665.
- Jackle, H., Rosenberg, U. B., Preiss, A., Seifert, E., Knipple, D. C., Kienlin, A., y Lehmann, R., 1985. Molecular analysis of *Krüppel*, a segmentation gene of *Drosophila melanogaster*. *Cold Spring Harb. Symp. Quant. Bio.* 50, 465-473.
- Jurgens, G., Wieschaus, E. F., Nüsslein-volhard, C., 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* II. *Roux's Archives of Dev. Biol.* 193, 283-295.
- Kaufman, T. C., Severson, D. W., Robinson, G. E., 2002. The *Anopheles* genome and comparative insect genomics. *Science*. 298:97-98.
- Kelly, G. M. y Huebner, E., 1989. Embryonic Development of the Hemipteran Insect *Rhodnius prolixus*. *Journal of Morphology*. 199, 175-196.
- Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, A. M., Clark, J. M., Hyeock, S., *et al.*, 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. USA*. 107:12168–12173.
- Krause G., 1939. Die Eitypen der Insekten. *Biol. Zentralbl.* 59:495–536
- Kraut, R. y Levine, M., 1991. Spatial regulation of the gap gene giant during *Drosophila* development. *Development*. 111, 601-9.
- Lawrence, P. A., 1992. *The making of a fly: the genetics of animal design*. Oxford, UK: Blackwell Scientific.
- Liu, P. Z. y Kaufman, T. C., 2004. *Krüppel* is a gap gene in the intermediate germband insect *Oncopeltus fasciatus* and is required for development of both blastoderm and germband-derived segments. *Development*. 131, 4567-4579.
- Liu, P. Z., y Kaufman, T. C., 2005. Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. *Evol. Dev.* 7, 629-46.
- Liu, P. Z. y Patel, N. H., 2010. *giant* is a bona fide gap gene in the intermediate germband insect, *Oncopeltus fasciatus*. *Development*. 137, 835-44.
- Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A., Eisen, M. B., 2011. Noncanonical Compensation of Zygotic X Transcription in Early *Drosophila melanogaster* Development Revealed through Single-Embryo RNA-Seq. *PLoS Biol.* 9(2), e1000590.

- Lynch, J. A., Brent, A. E., Leaf, D. S., Pultz, M. A., Desplan, C., 2006. Localized maternal *orthodenticle* patterns anterior and posterior in the long germ wasp *Nasonia*. *Nature*. 439, 728-32.
- Medeiros, M. N., Logullo, R., Ramos, I. B., Sorgine, M. H. F., Paiva-Silva, G. O. Mesquita, R. D., Alcantara Machado, E., Alice Coutinho, M., Masuda, H., Capurro, M. L., Ribeiro, J. M. C., Cardoso Braz, G. R. y Oliveira, P. L., 2011. Transcriptome and gene expression profile of ovarian follicle tissue of the triatomine bug *Rhodnius prolixus*, *Insect Biochemistry and Molecular Biology*, 41. 10, 823-831.
- Mellanby, H., 1934. The Early Embryonic Development of *Rhodnius prolixus* (Hemiptera, Heteroptera). In *Development of Rhodnius* (London: Department o Zoology, University College), 71-90.
- Mellanby, H., 1935. The Later Embryology of *Rhodnius prolixus*. In *Embryology of Rhodnius* (London: Department o Zoology, University College), 1-40.
- Metzker, M. L., 2010. Sequencing technologies - the next generation. *Nat. Rev. Gen.* 11, 31-46.
- Melo, M. C., 2006. Biodiversidad de Reduviinae (Reduviidae, Heteroptera): análisis cladístico y biogeográfico del genero *Leogorrus* Stal. Tesis de doctorado. FCNYM. UNLP.
- Mito, T., Okamoto, H., Shinahara, W., Shinmyo, Y., Miyawaki, K., Ohuchi, H., Noji, S., 2006. *Krüppel* acts as a gap gene regulating expression of *hunchback* and *even-skipped* in the intermediate germ cricket *Gryllus bimaculatus*. *Dev. Biol.* 294, 471-481.
- Mohler, J., Eldon, E. D., Pirrotta, V., 1989. A novel spatial transcription pattern associated with the segmentation gene, *giant*, of *Drosophila*. *EMBO J.* 8, 1539-48.
- Needleman, S. B. y Wunsch, C. D., 1970. "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins", *J.Mol. Biol.* 48, 443-453.
- Nene V., Wortman J. R., Lawson D., Haas, B., Kodira, C., Tu, Z., Loftus B., *et al.*, 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*. 316:1718–1723.
- Nüsslein-Volhard, C., y Wieschaus, E., 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature*. 287, 795-801.
- Nüsslein-volhard, C., 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* I. *Roux's Archives of Dev. Biol.* 193, 267-282.
- Olesnicky, E. C., y Desplan C., 2007. Distinct mechanisms for mRNA localization during embryonic axis specification in the wasp *Nasonia*. *Dev. Biol.* 306, 134-142.
- Osborne, P. Y Dearden, P. K., 2005. Non-radioactive in-situ hybridisation to honeybee embryos and ovaries. *Apidologie*. 36, 113–118

- Pankratz, M., y Jackle, H., 1993. Blastoderm segmentation. In: The development of *Drosophila melanogaster*. Cold Spring Harbor (New York): Cold Spring Harbor Laboratory Press. 467–516.
- Pearson, W. R., Wood, T., Zhang, Z., Miller, W., 1997. Comparison of DNA sequences with protein sequences. *Genomics*, 46:24–36
- Peel, A. D., Chipman, A. D., Akam, M., 2005. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat Rev Genet.* 6, 905-16.
- Petschek, J. P., Perrimon, N., Mahowald, A. P., 1987. Region-specific defects in *l(1)giant* embryos of *Drosophila melanogaster*. *Dev Biol.* 119, 175-89.
- Preiss, A., Schmidt-ott, U., Kienlin, A., Seifert, E., Jackle, H., 1985. Molecular Genetics of *Kruppel*, a gene required for segmentation of the *Drosophila* embryo. *Nature.* 313.
- Pueyo, J. I., Lanfear, R., Couso, J. P., 2008. Ancestral Notch-mediated segmentation revealed in the cockroach *Periplaneta americana*. *Proc. Natl. Acad. Sci. USA.* 105, 16614-9.
- Pultz, M. A., Westendorf, L., Gale, S., Hawkins, K., Lynch, J., Pitt, J. N., Reeves, N. L., Yao, J. C. Y., Small, S., Desplan, C., Leaf, D., 2005. A major role for zygotic *hunchback* in patterning the *Nasonia* embryo. *Development.* 132, 3705-3715.
- Rivera-Pomar, R. V., Lu, X., Perrimon, N., Taubert, H., Jackle, H., 1995. Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature.* 376, 253-6.
- Rivera-Pomar, R. V., y Jackle, H., 1996. From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet.* 12, 478-83.
- Ronaghi M, Uhlén M, Nyren P, 1998. A sequencing method based on real-time pyrophosphate. *Science.* 281(5375):363, 365.
- Rosenberg, U., Schroeder, C., Preiss, A., Kienlin, A., Kienlin, A., Cote' S., Riede, I. y Jackle, H., 1986. Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. *Nature.* 319, 336-339.
- Schuh, R.T. y Slater, J.A., 1995. True Bugs of the World. Cornell University Press, Ithaca, NY.
- Schaeper N.D., Prpic N. M., Wimmer E. A., 2010. A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, developmental expression patterns and chromosomal location. *BMC Evo. Biol.* 2010, 10:88.
- Schroder, R., 2003. The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium*. *Nature.* 422, 621-5.
- Shaw, P. J., Wratten, N. S., Mc. Gregor, A. P., Dover, G. A., 2002. Coevolution in *bicoid*-dependent promoters and the inception of regulatory incompatibilities among species of higher diptera. *Evo. Dev.* 4, 265–277.

- Small, S., Blair, A., Levine, M., 1992. Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. EMBO J. 11, 4047-57.
- Smith, T. F. y Waterman, M. S., 1981. Identification of common molecular subsequences, J. Mol. Biol., 147, 195-197.
- Soares M. B., Bonaldo M. F., Jelene P., Su L., Lawton L., Efstratiadis A., 1994. Construction and characterization of a normalized cDNA library. Proc. Natl. Acad. Sci. USA. 91:9228-9232.
- Sommer, R. J., y Tautz, D., 1993. Involvement of an orthologue of the *Drosophila* pair-rule gene *hairy* in segment formation of the short germ-band embryo of *Tribolium* (Coleoptera). Nature. 361, 448-50.
- St Johnston, D., y Nüsslein-Volhard, C., 1992. The origin of pattern and polarity in the *Drosophila* embryo. Cell. 68, 201-219.
- Stephen R., 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.
- Strunk, B., Struffi, P., Wright, K., Pabst, B., Thomas, J., Qin, L., and Arnosti, D. N. 2001. Role of CtBP in transcriptional repression by the *Drosophila* *giant* protein. Dev. Biol. 239, 229-40.
- Tomari, Y. y Zamore, P.D., 2005. Perspective: machines for RNAi. Gen. Dev. 19:517–529.
- van Helden, J., 2003. Regulatory sequence analysis tools. Nucleic Acids Res. 31, 3593-6.
- Vanderberg, J.P., 1963. Synthesis and Transfer of DNA, RNA, and Protein during vitellogenesis in *Rhodnius prolixus* (Hemiptera). Biological Bulletin. 125, 3, 556-575.
- Werren, JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK; *et al.*, 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science. 15; 327 (5963):343-8.
- Wieschaus, E. F., Nüsslein-volhard, C., and Jurgens, G., 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* III. Roux's Archives of Dev Biol. 296-307.
- Wilson, M. J., Havler, M., Dearden, P. K., 2010. *Giant*, *Krüppel*, and *caudal* act as gap genes with extensive roles in patterning the honeybee embryo. Dev Biol. 339, 200-11.
- Wimmer, E. A., Jäckle, H., Pfeifle, C., Cohen S. M., 1993. A *Drosophila* homologue of human Sp1 is a head-specific segmentation gene. Nature 366, 690 – 694.
- Wimmer, E. A., 1995 Untersuchungen zur Struktur, Funktion und Regulation von *buttonhead*, einem Kopf-Gap-Gen von *Drosophila melanogaster*, *Dissertation, Universität München*, Cuvillier Verlag, Goettingen, Germany.
- Wolff, C., Schroder, R., Schulz, C., Tautz, D., Klingler, M., 1998. Regulation of the *Tribolium* homologues of *caudal* and *hunchback* in *Drosophila*: evidence for maternal gradient systems in a short germ embryo. Development 125, 3645-3654.

Zhulidov, P. A., Bogdanova, E. A., Shcheglov, A. S., Vagner, L. L., Khaspekov, G. L., Kozhemyako V. B., Matz, M. V., Meleshkevitch, E., Moroz, L. L., Lukyanov, S. A., Shagin, D. A., 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res* 32:e37.

## Anexo 1

### Orthodenticle (Otd)

#### Matriz

A	3	1	0	16	16	0	0	0	7	9
C	4	11	0	0	0	0	16	16	5	2
G	8	3	0	0	0	0	0	0	4	2
T	1	1	16	0	0	16	0	0	0	3

#### Secuencia Logo

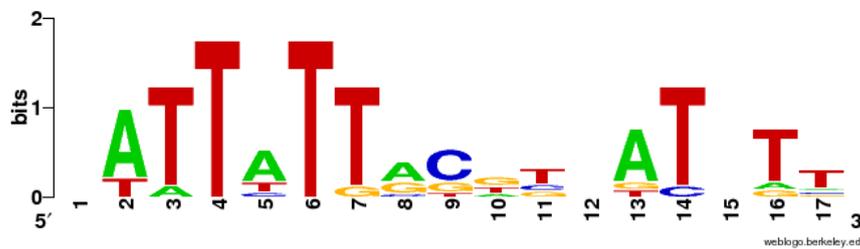


#### Secuencia (referencia 1 y 2)

GCTAATCCAA  
 GCTAATCCAA  
 GCTAATCCAA  
 GCTAATCCAA  
 GCTAATCCAA  
 CCTAATCCAA  
 ACTAATCCGA  
 GATAATCCCA  
 GGTAATCCCA  
 GGTAATCCCG  
 TGTAATCCAC  
 ACTAATCCCC  
 CCTAATCCGT  
 CCTAATCCCT  
 ACTAATCCGG  
 CTTAATCCGT

Giant (Gt)Matriz

A	1	7	1	0	6	0	0	5	0	2	0	3	7	0	3	1	1
C	4	0	0	0	1	0	0	1	6	0	2	3	0	1	2	0	1
G	2	0	0	0	0	0	1	3	2	4	2	1	1	0	1	1	1
T	2	2	8	9	2	9	8	0	1	3	5	2	1	8	3	7	6

Secuencia LogoSecuencias (referencias 3 y 4)

CATTATTACTCTATTTT  
 CATTATTATGTGTTTT  
 GATTATTAGTCAATTGC  
 CTTTATTGCAGCATCTT  
 TTTTATTGCAGCATCTT  
 CATTATGCGTTTGTGTG  
 AATTTTTACGTAATATT  
 GATTCTTGCATATAAA  
 TAATTTTTACGTAACATT

Referencias

- Gao, Q. y Finkelstein, R., 1998. Targeting gene expression to the head: the *Drosophila orthodenticle* gene is a direct target of the Bicoid morphogen. *Development*. 125, 4185-4193.
- Gao Q, Wang Y, y Finkelstein R., 1996. Orthodenticle regulation during embryonic head development in *Drosophila*. *Mech Dev*. 56, 3-15.
- Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H., Jackle, H., 1995. Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature*. 376, 253-256.
- Capovilla, M., Eldon, E. D., Pirrotta, V., 1992. The giant gene of *Drosophila* encodes a b-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. *Development*. 114, 99-112.

## Anexo 2

Tabla1: Resultados de la búsqueda por similitud de los clones de la genoteca normalizada		
Identidad de los clones por similitud	Numero de acceso	Frecuencia en la genoteca
NADH dehydrogenase subunit 1 [Triatoma dimidiata]	gi 159779469	34
NADH dehydrogenase subunit 2 [Triatoma dimidiata]	gi 159693582	2
NADH dehydrogenase subunit 3 [Triatoma dimidiata]	gi 159777191	6
NADH dehydrogenase subunit 4 [Strongylus vulgaris]	gi 159626217	3
NADH dehydrogenase subunit 5 [Triatoma dimidiata]	gi 159763892	6
NADH dehydrogenase subunit 6 [Triatoma dimidiata]	gi 159743165	2
NAPO-cytochrome P450 reductase[Drosophila melanogaster]	gi 159741260	2
NADH:ubiquinone oxidoreductase ESSS subunit, putative [N. vitripennis]	gi 159625016	1
ATP synthase F0 subunit 6 [Oribacterium niger]	gi 159608708	1
ATP synthase, H <sup>+</sup> transporting, mitochondrial F0 complex, subunit g; Mus musculus]	gi 159592171	1
ATP synthase 6 [Homo sapiens]	gi 159763041	1
ATP synthase F0 subunit 6 [Triatoma dimidiata]	gi 159762626	5
ADP-ribosylation factor, putative [Pediculus humanus corporis]	gi 159718578	2
ATP-dependent RNA helicase WM6 [Culex quinquefasciatus]	gi 159741735	2
ADP/ATP translocase [Lucilia cuprina]	gi 159617531	1
GTPase, IMAP family member 4-like [Danio rerio]	gi 159598383	1
cytochrome oxidase 1 [Rhodnius prolixus]	gi 159775626	7
cytochrome oxidase subunit 2 [Zaprinus tuberculatus]	gi 159697598	1
cytochrome b [Triatoma dimidiata]	gi 159776633	2
cytochrome c [validated] - honeybee	gi 159623822	1
cytochrome c oxidase subunit II [Triatoma dimidiata]	gi 159763043	3
cytochrome c oxidase subunit III [Triatoma dimidiata]	gi 159661483	14
cytochrome P450 [Tribolium castaneum]	gi 159723129	1
cytochrome c oxidase subunit VIc [Tribolium castaneum]	gi 159742213	1
Superoxide dismutase [Cu-Zn](Drosophila virilis)	gi 159632943	1

60S ribosomal protein L17 [ <i>Bombyx mori</i> ]	gi 159717732	2
40S ribosomal protein S14, putative [ <i>Pediculus humanus corporis</i> ]	gi 159614990	2
Ribosomal protein S11 [ <i>Heliothis virescens</i> ]	gi 159778317	2
Ribosomal protein S12 [ <i>Spodoptera frugiperda</i> ]	gi 159777195	2
Ribosomal protein L24 [ <i>Spodoptera frugiperda</i> ]	gi 159758066	2
Ribosomal protein S27 [ <i>Spodoptera frugiperda</i> ]	gi 159758481	3
Ribosomal protein L35 [ <i>Spodoptera frugiperda</i> ]	gi 159761773	1
Ribosomal protein S15A CG2033-PA [ <i>Drosophila melanogaster</i> ]	gi 159758068	2
Ribosomal protein L32 [ <i>Argopecten irradians</i> ]	gi 159763051	2
Ribosomal protein L14 [ <i>Rattus norvegicus</i> ]	gi 159724234	4
Ribosomal protein L35A [ <i>Spodoptera frugiperda</i> ]	gi 159768484	1
Ribosomal protein L39 [ <i>Spodoptera frugiperda</i> ]	gi 159778321	2
Ribosomal protein S7 [ <i>Theileria parva</i> strain Muguga]	gi 159716897	2
60S ribosomal protein L27A or L22 [ <i>Platichthys flesus</i> ]	gi 159742682	2
Ribosomal protein S15 [ <i>Spodoptera frugiperda</i> ]	gi 159608704	2
Ribosomal protein S23 [ <i>Spodoptera frugiperda</i> ]	gi 159662747	4
23S rRNA pseudouridine synthase D [ <i>Citrobacter</i> sp. 30_2]	gi 159762609	2
Ribosomal protein L28e [ <i>Tribolium castaneum</i> ]	gi 159630242	1
mitotic control protein dis3 [ <i>Aedes aegypti</i> ]	gi 159761778	2
Cylicin-1 [ <i>Mus musculus</i> ]	gi 159700679	1
Inhibitor of apoptosis protein [ <i>Amsacta moorei</i> ]	gi 159742208	1
Translation initiation factor 5a [ <i>Triatoma infestans</i> ]	gi 159607499	2
Elongation factor 1-delta, putative [ <i>Pediculus humanus corporis</i> ]	gi 159762622	2
eEIF-5A, isoform A [ <i>Drosophila melanogaster</i> ]	gi 159608710	1
Elongation factor-1 gamma [ <i>Acyrtosiphon pisum</i> ]	gi 159758073	1
ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA)eEF-1 alpha chain [ <i>Apis mellifera</i> ]	gi 159757672	1
Translation elongation factor-1 gamma [ <i>Locusta migratoria</i> ]	gi 159723134	3
Elongation factor-1alpha F2 [ <i>Apis mellifera</i> ]	gi 159627416	1
DNA polymerase epsilon catalytic subunit A [ <i>Xenopus tropicalis</i> ]	gi 159763036	2
Histone 2, H3ca1 [ <i>Mus musculus</i> ]	gi 159741268	1

Histone H1-II-1, putative [ <i>Pediculus humanus corporis</i> ]	gi 159741736	2
DNA polymerase III delta' subunit [ <i>Candidatus Sulcia muelleri</i> CARI]	gi 159618817	2
Activator 1 140 kDa subunit (Replication factor C large subunit)[ <i>A. mellifera</i> ]	gi 159742212	2
ribose-5-phosphate isomerase [ <i>Tribolium castaneum</i> ]	gi 159628745	1
Brain chitinase and chia [ <i>Nasonia vitripennis</i> ]	gi 159716901	1
Chitinase CHT-1 (66.9 kD) (cht-1) [ <i>Caenorhabditis elegans</i> ]	gi 159740796	1
Endonuclease and reverse transcriptase-like protein [ <i>Tribolium castaneum</i> ]	gi 159779470	2
Midgut chitinase [ <i>Lutzomyia longipalpis</i> ]	gi 159612495	1
Zinc metalloprotease [ <i>Nasonia vitripennis</i> ]	gi 159662744	2
Transketolase [ <i>Bombyx mori</i> ]	gi 159609935	2
7-cyano-7-deazaguanine reductase ( <i>Homalodisca coagulata</i> )	gi 159613745	1
Exosome complex exonuclease RRP41, putative [ <i>Tribolium castaneum</i> ]	gi 159613744	1
Disulfide-isomerase 1 precursor, putative [ <i>Pediculus humanus corporis</i> ]	gi 159760947	1
Histidine kinase [ <i>Flavobacterium johnsoniae</i> UW101]	gi 159607498	1
Ubiquitin specific protease 7 (herpes virus-associated);[ <i>Homo sapiens</i> ]	gi 159632937	1
Tubulin alpha-6 chain(Alpha-tubulin isotype M-alpha-6) [ <i>Apis mellifera</i> ]	gi 159724231	1
ORF MSV061 putative LINE reverse transcriptase, <i>Caenorhabditis elegans</i>	gi 159762192	1
Cathepsin L-like cysteine protease [ <i>Rhodnius prolixus</i> ]	gi 159741738	7
Aldolase A; fructose-bisphosphate aldolase [ <i>Homo sapiens</i> ]	gi 159723460	6
Adenosine triphosphate synthase subunit 6 [ <i>Cherax destructor albidus</i> ]	gi 159727338	1
Synthase A chain subunit 6 [ <i>Geocica utricularia</i> ]	gi 159696591	1
Protein phosphatase-1 [ <i>Herdmania curvata</i> ]	gi 159741730	2
S-adenosylmethionine synthetase [ <i>Acyrtosiphon pisum</i> ]	gi 159723050	1
Nucleoside diphosphate kinase(Abnormal wing discs protein)( <i>Drosophila melanogaster</i> )	gi 159778315	2
Effete Ubiquitin-conjugating enzyme E2-17 kDa[ <i>Drosophila melanogaster</i> ]	gi 159762615	2

proteasome 26S non-ATPase subunit 8 [Mus musculus]	gi 159761354	1
Vacuolar ATP synthase subunit G (V-ATPase G subunit) [Manduca sexta]	gi 159761353	1
beta-carotene 15,15'-dioxygenase [Gallus gallus]	gi 159760154	2
Ligase III, DNA, ATP-dependent isoform alpha [Homo sapiens]	gi 159685321	2
Glutathione-requiring prostaglandin D synthase	gi 159661488	1
Glutathione S-transferase GST-9 (gst-9) [Caenorhabditis elegans]	gi 159641422	1
Superfamily I DNA/RNA helicase [Cyanotheca sp. PCC 7822]	gi 159645530	1
Fumarylacetoacetate hydrolase domain-containing protein, putative [Pediculus humanus corporis]	gi 159625017	1
Neprilysin (EC 3.4.24.11) - rabbit	gi 159740329	1
Protein tyrosine phosphatase, non-receptor type 11 [Tribolium castaneum]	gi 159620125	1
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase [Methanosarcina acetivorans str. C2A]	gi 159758480	1
Zinc metalloproteinase 5 [Ancylostoma ceylanicum]	gi 159685323	1
Serine threonine-protein kinase [Acyrtosiphon pisum]	gi 159616265	1
Cyclophilin-like CG13892-PA [Apis mellifera]	gi 159640448	1
Defensin C [Rhodnius prolixus]	gi 159724225	2
Nuclear Distribution gene C homolog,nuclear migration gene required for embryonic development[Caenorhabditis elegans]	gi 159741256	1
4Snc-Tudor domain protein, putative [Ixodes scapularis]	gi 159688196	2
Zinc finger protein [Culex quinquefasciatus]	gi 159723712	1
CREB 7 protein [Tribolium castaneum]	gi 159759314	2
Phd finger transcription factor [Tribolium castaneum]	gi 159630244	1
Zinc finger protein [Nasonia vitripennis]	gi 159761348	1
Paramyosin, long form, putative [Pediculus humanus corporis]	gi 159612492	1
Kinesin-73 CG8183-PB [Drosophila melanogaster]	gi 159638405	1
Myosin alkali light chain	gi 159779472	2
Tropomyosin 1 isoform B [Nasonia vitripennis]	gi 159716906	1
Beta actin [Pan troglodytes]	gi 159631723	1
Lamin [Aedes aegypti]	gi 159607496	2
Cuticle protein 7 (LM-7) (LM-ACP 7) [Locusta migratoria]	gi 159606305	2

Titin, putative [Pediculus humanus corporis]	gi 159763456	2
Novex-3 (titin isoform), partial [Acyrtosiphon pisum]	gi 159631715	2
Cuticle protein 21 (LM-ACP 21)	gi 159761763	4
Vinculin CG3299-PA [Drosophila melanogaster]	gi 159768488	2
Tubulin alpha-6 chain (Alpha-tubulin 6) (Alpha-tubulin isotype M-alpha-6) [Apis mellifera]	gi 159668222	2
Muscle myosin heavy chain isoform 2 [Acyrtosiphon pisum]	gi 159687336	2
Collagen alpha-1(II) chain precursor, partial [Apis mellifera]	gi 159758902	2
Myosin alkali light chain 1 isoform b [Acyrtosiphon pisum]	gi 159726290	2
Peci protein [Ciona intestinalis]	gi 159763454	2
U2 small nuclear riboprotein auxiliary factor 50 CG9998-PA [A. mellifera]	gi 159760579	2
Death-associated small cytoplasmic leucine-rich prot SCLP [T. castaneum]	gi 159759737	4
GluClalpha, isoform F [Drosophila melanogaster]	gi 159742203	1
Cyclic nucleotide-binding domain containing protein [Tetrahymena thermophila]	gi 159716900	3
Sugar transporter protein ERD6 [imported] - Arabidopsis thaliana	gi 159768485	1
Sodium/potassium-dependent atpase beta-2 subunit [Acyrtosiphon pisum]	gi 159696589	1
ABC1 transporter [Leishmania braziliensis MHOM/BR/75/M2904]	gi 159763034	1
4Fe-4S ferredoxin iron-sulfur binding domain protein [Clostridium cellulolyticum]	gi 159692003	4
WAP four-disulfide core domain 2 isoform 1 [Apis mellifera]	gi 159645528	2
Nuclear distribution protein NUDC, putative [Ixodes scapularis]	gi 159639409	1
Putative sensor histidine kinase [Clostridium botulinum]	gi 159638409	1
Calmin [Aedes aegypti]	gi 159638408	1
Ganglioside induced differentiation associated protein, Pediculus humanus corporis	gi 159635046	1
EmrB/QacA family drug resistance transporter [Frankia sp. Cc13]	gi 159626219	1
PE repeat family protein [Arthroderma benhamiae CBS 112371]	gi 159622613	1
LPXTG-motif cell wall anchor domain protein [Lactobacillus jensenii JV-V16]	gi 159602191	2
Shaggy CG2621-PA [Acyrtosiphon pisum]	gi 159614994	2

CLOSPI_00859 [Clostridium]	gi 159606295	2
Cullin-1, putative [Pediculus humanus corporis]	gi 159603526	2
CRE-BEN-1 protein [Caenorhabditis remanei]	gi 159603521	1
OTU domain containing 6B [Apis mellifera]	gi 159726790	1
Salivary platelet aggregation inhibitor 2 [Rhodnius prolixus]	gi 159779471	2
Ejaculatory bulb protein III CG11390-PA [Drosophila melanogaster]	gi 159778887	2
90-kDa heat shock protein HSP83 [Spodoptera frugiperda]	gi 159759320	2
hsp70 family member [Locusta migratoria]	gi 159694645	3
70 kDa heat shock protein [Artemia franciscana]	gi 159763467	1
Calmodulin CG8472-PB [Drosophila melanogaster]	gi 159763044	2
transferrin [Riptortus clavatus]	gi 159759732	2
TPA: hypothetical protein [Homo sapiens]	gi 159743172	2
5'-phosphoribosylaminoimidazole carboxylase-5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole synthetase [Drosophila melanogaster]	gi 159742701	2
Ras-related protein Rab-27A(GTP-binding protein Ram)	gi 159741266	2
Vitellogenin - Riptortus clavatus gb AAB72001.1	gi 159740792	2
Jaguar CG5695-PA [Drosophila melanogaster]	gi 159694641	1
Surfeit 1 CG9943-PA [Drosophila melanogaster]	gi 159602197	2
mp20 [Drosophila melanogaster]	gi 159644415	1
JH-inducible protein [Galleria mellonella]	gi 159648719	1
CSP precursor [Locusta migratoria]	gi 159668223	1
HEAT repeat-containing protein, putative [Pediculus humanus corporis]	gi 159763465	2
Odorant-binding protein 2 [Acyrtosiphon pisum]	gi 159644420	1
Odorant-binding protein 56a, putative [Aedes aegypti]	gi 159762199	1
Ornithine decarboxylase antizyme (ODC-Az), partial [Nasonia vitripennis]	gi 159696590	1
E2F-associated phosphoprotein [Equus caballus]	gi 159695608	1
Protein takeout precursor, putative [Pediculus humanus corporis]	gi 159664011	2
Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase family protein [Anaerococcus vaginalis ]	gi 159646567	1
Binding protein (AGAP002905-PA) [Anopheles gambiae]	gi 159636286	1

Apolipoprotein D-like [Macaca mulatta]	gi 159602199	1
LOC100428052 [Macaca mulatta]	gi 159689091	1
hypothetical protein TSTA_111970 [Talaromyces stipitatus]	gi 159717733	2
VVM_01209 [Vibrio vulnificus]	gi 159741725	1
Protein LOAG_16331 [Loa loa]	gi 159761772	2
Protein XP_860064 [Canis familiaris]	gi 159761759	7
CG33521-PA, isoform A [Apis mellifera]	gi 159775627	1
CG8709-PA [Apis mellifera]	gi 159718571	2
CG12367-PA [Apis mellifera]	gi 159617536	2
CG3862-PA [Apis mellifera]	gi 159723459	2
CG30421-PA isoform 1 [Apis mellifera]	gi 159763037	2
CG33521-PA, isoform A [Apis mellifera]	gi 159598379	1
CG3702-PA [Apis mellifera]	gi 159758897	1
Phum_PHUM055770 [Pediculus humanus corporis]	gi 159627410	1
Phum_PHUM055770 [Pediculus humanus corporis]	gi 159743653	1
protein Phum_PHUM269490 [Pediculus humanus corporis]	gi 159763032	2
ref XP_002426862.1  conserved hypothetical protein [Pediculus humanus corporis]	gi 159736499	1
ref XP_002425052.1  conserved hypothetical protein [Pediculus humanus corporis]	gi 159758482	2
ref XP_002569521.1  hypothetical protein [Schistosoma mansoni]	gi 159742221	2
AGAP009743-PA [Tribolium castaneum]	gi 159612493	2
AGAP004877-PA [Tribolium castaneum]	gi 159721332	1
CG18472 CG18472-PA [Tribolium castaneum]	gi 159717736	2
CG14650 CG14650-PA [Tribolium castaneum]	gi 159688198	2
CG11563 CG11563-PA [Tribolium castaneum]	gi 159763468	2
CG7368-PB [Tribolium castaneum]	gi 159716899	2
ref XP_966482.1  conserved protein [Tribolium castaneum]	gi 159760157	2
ref XP_002723895.1  PREDICTED: hypothetical protein [Oryctolagus cuniculus]	gi 159664014	2
Hypothetical protein GTHECHR1003 [Guillardia theta]	gi 159665390	1
Protein PCNPT3_00010 [Psychromonas sp. CNPT3]	gi 159762627	2

KLTH0A06006p [Lachanea thermotolerans]	gi 159690995	4
KLTH0F01056p [Lachanea thermotolerans]	gi 159779465	2
Ac1147 [Gallus gallus]	gi 159779463	11
AGAP007483-PA [Acyrtosiphon pisum]	gi 159719520	1
ref XP_001944236.1  PREDICTED: hypothetical protein isoform 2 [Acyrtosiphon pisum]	gi 159633993	1
ref XP_001947025.1  PREDICTED: similar to conserved hypothetical protein [Acyrtosiphon pisum]	gi 159741731	1
CG11138-PC [Acyrtosiphon pisum]	gi 159777750	2
ref XP_001864833.1  conserved hypothetical protein [Culex quinquefasciatus]	gi 159600879	1
ref XP_001867863.1  conserved hypothetical protein [Culex quinquefasciatus]	gi 159763039	1
Hypothetical protein LOC100406103 [Callithrix jacchus]	gi 159644417	2
SORBIDRAFT_2432s002010 [Sorghum bicolor]	gi 159776114	2
BRAFLDRAFT_76460 [Branchiostoma floridae]	gi 159776113	37
amPTPN6 [Branchiostoma belcheri]	gi 159666822	2
BRAFLDRAFT_72305 [Branchiostoma floridae]	gi 159691002	1
Protein SCHCODRAFT_58249 [Schizophyllum commune H4-8]	gi 159761362	1
SORBIDRAFT_0057s002150 [Sorghum bicolor]	gi 159760946	1
hypothetical protein Bm1_04940 [Brugia malayi]	gi 159762208	7
Bm1_11025 [Brugia malayi]	gi 159614991	2
ref XP_001601525.1  conserved hypothetical protein, partial [Nasonia vitripennis]	gi 159762630	1
CG33715-PB [Nasonia vitripennis]	gi 159762186	2
ref XP_001604071.1  hypothetical protein [Nasonia vitripennis]	gi 159760945	5
ENSANGP00000015219 [Nasonia vitripennis]	gi 159726293	1
GA15837-PA [Nasonia vitripennis]	gi 159716896	2
ENSANGP000000031916 [Nasonia vitripennis]	gi 159719516	1
GA21542-PA [Nasonia vitripennis]	gi 159665387	2
GG16908 [Drosophila erecta]	gi 159633991	1
GI16853 [Drosophila mojavensis]	gi 159779474	1
GE14003 [Drosophila yakuba]	gi 159763473	1

GK20166 [ <i>Drosophila willistoni</i> ]	gi 159768486	25
GI11098 [ <i>Drosophila mojavensis</i> ]	gi 159761765	1
GA24531 [ <i>Drosophila pseudoobscura pseudoobscura</i> ]	gi 159742209	2
CG18812-PC [ <i>Drosophila melanogaster</i> ]	gi 159760944	1
GI13339 [ <i>Drosophila mojavensis</i> ]	gi 159613746	1
GI24566 [ <i>Drosophila mojavensis</i> ]	gi 159625012	1
CG2926-PA [ <i>Drosophila melanogaster</i> ]	gi 159668221	1
CG10327-PC [ <i>Drosophila melanogaster</i> ]	gi 159664013	2
CG6842-PA [ <i>Drosophila melanogaster</i> ]	gi 159661482	1
CG5181-PA [ <i>Drosophila melanogaster</i> ]	gi 159637296	1
CG4111-PB [ <i>Drosophila melanogaster</i> ]	gi 159633998	1
CG10050-PA [ <i>Drosophila melanogaster</i> ]	gi 159613739	2
CG6459-PA [ <i>Drosophila melanogaster</i> ]	gi 159697603	1
CG6105-PA [ <i>Drosophila melanogaster</i> ]	gi 159759321	2
CG1490-PB [ <i>Drosophila melanogaster</i> ]	gi 159761359	1
CG6843-PA [ <i>Drosophila melanogaster</i> ]	gi 159761760	2
CG8332-PA [ <i>Drosophila melanogaster</i> ]	gi 159725282	1
CG10997-PA [ <i>Drosophila melanogaster</i> ]	gi 159763047	2
CG5808-PA [ <i>Drosophila melanogaster</i> ]	gi 159775636	1
GH11894 [ <i>Drosophila grimshawi</i> ]	gi 159763046	2
GH18857 [ <i>Drosophila grimshawi</i> ]	gi 159777196	1
GJ20513 [ <i>Drosophila virilis</i> ]	gi 159761349	1
GK24869 [ <i>Drosophila willistoni</i> ]	gi 159632936	1
AaeL_AAEL011959 [ <i>Aedes aegypti</i> ]	gi 159762206	2
AaeL_AAEL014559 [ <i>Aedes aegypti</i> ]	gi 159616264	1
AaeL_AAEL007344 [ <i>Aedes aegypti</i> ]	gi 159639417	1
ENSANGP00000015232 [ <i>Anopheles gambiae</i> ]	gi 159778318	2
AaeL_AAEL000901 [ <i>Aedes aegypti</i> ]	gi 159614997	2
ENSANGP00000020070 [ <i>Anopheles gambiae</i> ]	gi 159777751	1
ENSANGP00000020393 [ <i>Anopheles gambiae</i> ]	gi 159777189	2
ENSANGP00000012983 [ <i>Anopheles gambiae</i> ]	gi 159776112	2

ENSANGP0000001645 [Anopheles gambiae]	gi 159763895	1
ENSANGP00000012072 [Anopheles gambiae]	gi 159763891	2
ENSANGP00000009361 [Anopheles gambiae]	gi 159763464	1
ENSANGP00000021620 [Anopheles gambiae str. PEST]	gi 159762628	1
ENSANGP00000022432 [Anopheles gambiae str. PEST]	gi 159762610	3
ENSANGP00000020306 [Anopheles gambiae str. PEST]	gi 159762188	2
ENSANGP00000019453 [Anopheles gambiae str. PEST]	gi 159760948	1
ENSANGP00000019175 [Anopheles gambiae str. PEST]	gi 159760942	1
ENSANGP00000019192 [Anopheles gambiae str. PEST]	gi 159760156	1
ENSANGP00000021738 [Anopheles gambiae str. PEST]	gi 159760153	2
ENSANGP00000013046 [Anopheles gambiae str. PEST]	gi 159759733	1
ENSANGP00000019896 [Anopheles gambiae str. PEST]	gi 159758896	2
ENSANGP00000010751 [Anopheles gambiae str. PEST]	gi 159758070	2
ENSANGP00000011934 [Anopheles gambiae str. PEST]	gi 159758067	2
ENSANGP00000019081 [Anopheles gambiae str. PEST]	gi 159757674	4
ENSANGP00000010800 [Anopheles gambiae str. PEST]	gi 159744126	2
ENSANGP00000016671 [Anopheles gambiae str. PEST]	gi 159743168	2
ENSANGP00000025304 [Anopheles gambiae str. PEST]	gi 159742206	2
ENSANGP00000022885 [Anopheles gambiae str. PEST]	gi 159742202	1
ENSANGP00000022635 [Anopheles gambiae str. PEST]	gi 159741262	1
ENSANGP00000020045 [Anopheles gambiae str. PEST]	gi 159740801	2
ENSANGP00000018910 [Anopheles gambiae str. PEST]	gi 159740786	2
ENSANGP00000013163 [Anopheles gambiae str. PEST]	gi 159693581	2
ENSANGP00000011747 [Anopheles gambiae str. PEST]	gi 159726785	2
ENSANGP00000017104 [Anopheles gambiae str. PEST]	gi 159726295	2
ENSANGP00000002653 [Anopheles gambiae str. PEST]	gi 159725280	2
ENSANGP00000007563 [Anopheles gambiae str. PEST]	gi 159722191	2
ENSANGP00000015316 [Anopheles gambiae str. PEST]	gi 159687337	2
ENSANGP00000016830 [Anopheles gambiae str. PEST]	gi 159647674	1
ENSANGP00000020734 [Anopheles gambiae str. PEST]	gi 159665383	2
ENSANGP00000011823 [Anopheles gambiae str. PEST]	gi 159662751	1

ENSANGP0000006861 [Anopheles gambiae str. PEST]	gi 159642418	1
ENSANGP00000022280 [Anopheles gambiae str. PEST]	gi 159641431	2
ENSANGP00000015585 [Anopheles gambiae str. PEST]	gi 159632940	1
ENSANGP00000010158 [Anopheles gambiae str. PEST]	gi 159613742	1
ENSANGP00000013212 [Anopheles gambiae str. PEST]	gi 159609927	1
AGAP012073-PA [Anopheles gambiae str. PEST]	gi 159760587	2
AGAP003713-PA [Anopheles gambiae str. PEST]	gi 159602193	1
AGAP010920-PA [Anopheles gambiae str. PEST]	gi 159647671	1
ENSANGP00000009396 [Anopheles gambiae str. PEST]	gi 159608706	1
KIAA1911 protein [Homo sapiens] [Rattus norvegicus]	gi 159625021	1
NEMVEDRAFT_v1g6787 [Nematostella vectensis]	gi 159625011	1
ref XP_001628638.1  predicted protein [Nematostella vectensis]	gi 159763452	2
ref XP_001638814.1  predicted protein [Nematostella vectensis]	gi 159725772	2
RIKEN cDNA 6720456J22 gene [Mus musculus]	gi 159611164	1