



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO
Doctorado en Ciencias Biomédicas

“Caracterización del transcriptoma de la glándula venenosa y el proteoma del veneno del alacrán *Thorellius atrox*”

TESIS
QUE PARA OPTAR POR EL GRADO DE:
Doctor en Ciencias Biomédicas

PRESENTA:
M. en C. María Teresa Romero Gutiérrez

TUTOR PRINCIPAL
Dr. Lourival Domingos Possani Postay
Instituto de Biotecnología-UNAM

MIEMBROS DEL COMITÉ TUTOR

Dr. Miguel A. Cevallos Gaos
Centro de Ciencias Genómicas

Dr. Gerardo Corzo Burguete
Instituto de Biotecnología

Cuernavaca, Morelos, mayo del 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Esta tesis se realizó en el laboratorio del Dr. Lourival D. Possani Postay, que pertenece al Departamento de Medicina Molecular y Bioprocesos del Instituto de Biotecnología-Universidad Nacional Autónoma de México en Cuernavaca, Morelos.

El trabajo fue financiado por los donativos de la Dirección General de los Asuntos del Personal Académico (DGAPA)-UNAM IN203416 y SEP-CONACyT 237864, ambos concedidos al Dr. Lourival D. Possani. La alumna María Teresa Romero Gutiérrez recibió la beca de doctorado de CONACyT con el número de becario 257751.

Agradecimientos:

Al doctor Fernando Zamudio Zúñiga por su apoyo en la secuenciación de péptidos y espectrometría de masas. Al M. en C. Timoteo Olamendi Portugal y a Leopoldo Güereca Gurrola por su apoyo en la secuenciación de DNA.

Agradezco también a Shirley Elizabeth Ainsworth Gore, encargada de la Unidad de Biblioteca y a Jesús Omar Arriaga Pérez por su apoyo. A la Unidad de Cómputo, especialmente a Juan Manuel Hurtado Ramírez y a Roberto Pablo Rodríguez Bahena por todo el apoyo bioinformático-técnico brindado.

Gracias a la Unidad de Secuenciación Masiva, especialmente a Verónica Jiménez Jacinto, a Jérôme Jean Verleyen y al doctor Alejandro Sánchez por su apoyo bioinformático, al Dr. Ricardo Alfredo Grande Cano y a Gloria Tanahiry Vazquez Castro por su apoyo en la secuenciación.

A Erika Patricia Meneses Romero, a Lorena Hernández Orihuela y al Dr. Cesar Ferreira Batista de la Unidad Universitaria de Proteómica por el apoyo brindado durante la realización de la tesis.

Dr. Possani, muchas gracias por creer en mí y darme la oportunidad de ser su alumna. Las enseñanzas que me llevo no sólo me servirán para mi vida profesional sino personal. Siempre actuar con ética y ser constante y disciplinada, por enseñarme a disfrutar mi trabajo.

A mi comité tutorial, al doctor Miguel A. Cevallos Gaos por no darse por vencido conmigo, por siempre darme ánimos para seguir adelante. Gracias por siempre salvarme el pellejo y por toda la ayuda que me brindó durante este tiempo. Al doctor Gerardo Corzo por la paciencia y los regaños, siempre con un punto de vista acertado y oportuno.

Erne, muchas gracias por siempre regresarme a la tierra, por enseñarme a pensar como científica y ser impecable con mi trabajo; Rita, muchas gracias por ser como mi hermana mayor, por regañarme y consolarme cuando no me quedaban fuerzas para seguir adelante. Vero, gracias por forjar mis raíces y ayudarme a ser independiente. Dra. Gina, gracias por inyectarle alegría a mi vida con esas pláticas tan largas y nutritivas, por siempre apoyarme en mis locuras. Muchas gracias a ustedes por todo el cariño, por su amistad.

A mi familia por no dejar que me rindiera, por ser mi motor y mi motivación. A mis suegros, a Lalo y Nayeli, muchas gracias por adoptarme y quererme como una de los suyos.

A mis queridos amigos, mi segunda familia; Jimena, Juancho, Mari, Rosby, Miriam, Leo, Lili, Karla, Jonathan, Víctor, Perla, Guille, Laura porque a pesar de la distancia o la cercanía, a pesar de todo sigo estando en su corazón.

A todos mis colegas y amigos del laboratorio, muchas gracias por todos los aportes que le dieron a mi trabajo y a mi vida.

Guillermo, mi compañero de vida. Gracias por compartir conmigo este jardín.

Índice

1. Resumen.....	8
2. Abstract.....	9
3. Marco teórico.....	10
4. Antecedentes.....	14
5. Objetivos.....	17
5.1 Objetivo principal.....	17
5.2 Objetivos específicos.....	17
6. Material y métodos.....	18
6.1 Esquema metodológico.....	18
6.2 Metodología.....	18
6.2.1. Colecta de material biológico.....	18
6.2.2 Análisis transcriptómico de la glándula venenosa del alacrán <i>T. atrox</i>	18
6.2.2.1 Extracción de RNA de la glándula venenosa.....	18
6.2.2.2 RNA-Seq, ensamble y anotación del transcriptoma.....	19
6.2.2.3 Cuantificación de RNA-Seq.....	20
6.2.3 Análisis proteómico del veneno del alacrán <i>T. atrox</i>	21
6.2.3.1 Determinación de masas moleculares del veneno del alacrán <i>T. atrox</i>	21
6.2.3.2 Secuenciación de proteínas por LC-MS/MS.....	21
6.2.3.3 Análisis de datos obtenidos por LC-MS/MS.....	22
6.2.4 Síntesis y caracterización biológica de secuencias identificadas en el transcriptoma y el proteoma.....	22
6.2.4.1 Medición de la actividad antimicrobiana.....	23
7. Resultados y discusión.....	24
7.1 RNA-Seq de la glándula venenosa del alacrán <i>T. atrox</i>	24
7.1.1 Extracción de RNA de la glándula venenosa y transcriptoma (RNA-Seq).....	24
7.1.2 Análisis del transcriptoma.....	25
7.1.3 Nomenclatura de los transcritos.....	26
7.1.4 Componentes de veneno identificados en el transcriptoma de las glándulas venenosas del alacrán <i>T. atrox</i>	27
7.1.4.1 Péptidos con enlaces disulfuro.....	27
Toxinas que actúan en los canales de sodio dependientes de voltaje (NaTxS).	28
Toxinas que actúan sobre canales de potasio (Ktx).....	31
Toxinas que actúan en canales de calcio.....	34
7.1.4.2 Péptidos de defensa del huésped (HDPs).....	38

7.1.4.3 Enzimas	42
7.1.4.4 Inhibidores de proteasas	43
7.1.4.5 Otros componentes de veneno	44
Péptidos La1-like	44
Otros componentes de veneno con función desconocida	45
7.1.5 Cuantificación del transcriptoma de <i>T. atrox</i>	45
7.2 Proteómica del veneno del alacrán <i>T. atrox</i>	48
7.2.1 Fingerprint del veneno	48
7.2.2 Secuenciación <i>de novo</i> por LC MS/MS	50
7.3 Síntesis y caracterización biológica del péptido ViCT2 secuenciado en el transcriptoma y el proteoma del alacrán <i>T. atrox</i>	51
8. Conclusiones	55
9. Perspectivas	56
10. Bibliografía	57
11. Anexos	63
Anexo 1. Purificación de RNA a partir de la glándula venenosa usando el kit SV Total RNA Isolation System® Promega	63
Anexo 2. Dominios PFAM Utilizados para hacer búsquedas en los archivos de anotación del transcriptoma de <i>T. atrox</i>	64
Anexo 3. Comandos utilizados para caracterizar el transcriptoma y medir los niveles de expresión de los transcritos relacionados con veneno	65
Anexo 4. Protocolo de microdilución	69
Anexo 5. Cuantificación y medición de integridad de RNA por Bioanalyzer	71
Anexo 6. Gráficos de calidad de la secuenciación	72
Anexo 7. Cuantificación de RNA-Seq	74
Anexo 8. Secuencias LC-MS/MS	78
Anexo 9. Artículo de titulación	81

Índice de figuras

Figura 1: Distribución de la sub-familia Syntropinae. Tomado de (Soleglad & Fet, 2008).....	15
Figura 2: Distribución de términos GO en los transcritos anotados de la glándula venenosa del alacrán <i>T. atrox</i>	25
Figura 3. Diversidad relativa de transcritos anotados con identidad de proteínas de veneno.	26
Figura 4. Transcritos identificados en el transcriptoma de <i>T. atrox</i> con actividad putativa de toxinas de sodio.....	30
Figura 5. Secuencias identificadas en el transcriptoma de <i>T. atrox</i> con posible actividad de toxinas de potasio.....	33
Figura 6. Toxinas putativas de calcio identificadas en el transcriptoma de <i>T. atrox</i>	36
Figura 7 Posibles HDPs identificados en el análisis transcriptómico.....	39
Figura 8. Alineamiento de secuencias que tuvieron identidad con péptidos tipo La1.....	44
Figura 9. Distribución de las abundancias de los transcritos identificados como componentes de veneno contra el logaritmo de la longitud efectiva	47
Figura 10. Distribución relativa de las masas moleculares identificadas en el veneno del alacrán <i>T. atrox</i>	49
Figura 11. Características fisicoquímicas del péptido ViCT2 y su proyección helicoidal.	52
Figura 12. Perfil cromatográfico del péptido sintético ViCT2.	53
Figura 13. Actividad anti-bacteriana del péptido ViCT2 evaluada contra la cepa <i>E. coli</i> ATCC 25922.	54
Figura 14. Actividad anti-bacteriana del péptido ViCT2 contra la cepa <i>P. aeruginosa</i> ATCC 25923. ..	55
Figura 15. Actividad anti-bacteriana del péptido ViCT2 contra la cepa <i>S. aureus</i> ATCC 25923 utilizando como tratamiento el péptido sintético ViCT2.....	55

Índice de tablas

Tabla 1. Revisión de los transcriptomas de glándulas venenosas reportadas en la literatura.	12
Tabla 2 Descripción del archivo de salida de Salmon.	20
Tabla 3. Nomenclatura utilizada para los transcritos identificados en el transcriptoma de <i>T. atrox</i>	27
Tabla 4. Parámetros fisicoquímicos predichos para los péptidos maduros de los transcritos anotados como NDBPs por el software HeliQuest	42
Tabla 5. Resumen de los 10 transcritos con mayor nivel de expresión	46
Tabla 6. <i>Fingerprint</i> de las fracciones de veneno soluble de <i>T. atrox</i>	48
Tabla 7. Masas moleculares identificadas en el transcriptoma de <i>T. atrox</i>	50
Tabla 8. Secuencias aminoacídicas identificadas por LC MS/MS utilizando el transcriptoma de <i>T. atrox</i> como base de datos para la identificación de proteínas por el algoritmo de Sequest.	50

1. RESUMEN

El trabajo realizado en esta tesis está dirigido a caracterizar el transcriptoma de la glándula venenosa y el proteoma del veneno del alacrán *Thorellius atrox*. Los resultados fueron obtenidos a partir de las glándulas venenosas y del veneno soluble del alacrán *Thorellius atrox*, colectados en el estado de Colima. El RNA purificado de las glándulas de veneno permitió obtener un cDNA que se secuenció por medio de la plataforma de Illumina. Como resultado de la secuenciación de este cDNA se obtuvieron más de 40 millones de *reads* (lecturas) ensamblados en más de 20,000 transcritos. El veneno soluble de este alacrán permitió separar el veneno por cromatografía líquida e identificar la masa molecular de sus componentes por medio de espectrometría de masas (LC-MS/MS). Utilizando una estrategia de búsqueda y anotación por medio de librerías especializadas, se identificaron 160 transcritos relacionados con proteínas de veneno.

Como resultado del perfil de masas moleculares del veneno, se identificaron 135 componentes, 5 de ellos coincidieron con la masa teórica predicha para los péptidos maduros identificados en el transcriptoma. La secuenciación *de novo* por LC-MS-MS permitió identificar 42 proteínas codificadas por los transcritos ensamblados, validando así el análisis transcriptómico.

Entre los hallazgos más significativos está la identificación de componentes, cuya secuencia de aminoácidos deducida tiene identidad con una toxina de calcio (ω -agatoxina) encontrada anteriormente en el veneno de arañas y las waprinas anteriormente encontradas en el veneno de serpientes. Así mismo se encontraron secuencias peptídicas semejantes a las secuencias de toxinas que reconocen canales de potasio, sodio y calcio. Los péptidos sin enlaces de disulfuro (NDBPs) identificados son abundantes y corresponden a diferentes familias de los mismos (NDBP-2, NDBP-3 y NDBP-4). Algunas de las secuencias identificadas corresponden a péptidos antimicrobianos descritos anteriormente para alacranes. Péptidos similares a defensinas y al péptido de función desconocida La-1, anteriormente encontrado en otros alacranes también están presentes, así como inhibidores de proteasas. Varios tipos de enzimas fueron encontradas, entre las cuales están: hialuronidasa, fosfolipasa A2, metaloproteinasas, serin-proteasas y enzima convertidora de angiotensina. Además, otros componentes cuya función no es conocida también fueron identificados.

2. ABSTRACT

The work performed in this thesis was directed to characterize the transcriptome of the venomous gland, and the proteome of the venom of the scorpion *Thorellius atrox*. The data reported in this thesis were obtained from venomous glands and soluble venom of the scorpion *T. atrox*, collected in the state of Colima (México). The RNA purified from the venomous glands allowed the preparation of the cDNA, which was sequenced using the platform of Illumina. The cDNA sequencing produced more than 40 million *reads*, which were assembled into over 20,000 transcripts. The whole soluble venom of this scorpion was separated by liquid chromatography and permitted the identification of the molecular masses of its components, by means of mass spectrometry analysis (LC-MS/MS). Using a strategy of search and annotation based on specialized libraries, 160 transcripts were identified, which are related to proteins of the scorpion's venom.

The proteomic analysis of the molecular masses obtained from the soluble venom permitted the identification of 135 components, from which 5 had identical masses, as expected from the mature peptides found in the transcriptomic analysis. *De novo* sequencing by LC-MS/MS method allowed the identification of 42 proteins, encoded by the assembled transcripts, validating the analysis of the transcriptome.

Among the most relevant findings are the identification of components whose deduced amino acid sequences show identity to calcium toxin (ω -agatoxin); previously found in the venom of spiders, and the waprins previously found in snake venoms. In addition, other peptides were found, showing amino acid sequences similar to scorpion toxins known to act on K^+ -channels, Na^+ -channels and Ca^{2+} -channels. Peptides with no disulfide bridges (NDBPs) occurred in high abundance and were shown to correspond to different families (NDBP-2, NDBP-3 y NDBP4). Several sequences identified correspond to anti-microbial peptides previously described to be present in scorpion venoms. But also peptides similar to defensins and a peptide of unknown function called La-1-like, earlier found in scorpion venom were found. Similarly, peptides that are thought to be protease inhibitors were identified. Several types of enzymes were found, among which are: hyaluronidases, phospholipases A2, metalloproteinases, serin-proteinase and angiotensin converting enzymes. In addition, other compounds of known function were also identified in this venom.

3. MARCO TEÓRICO

Desde hace muchos años se considera que los alacranes son los artrópodos más antiguos y conservados que existen, ya que desde su aparición hasta ahora, sólo han sufrido algunos pocos cambios. Los primeros especímenes aparecieron durante la era silúrica (aproximadamente hace 450 millones de años) y eran acuáticos (Lourenço, 2018).

Los alacranes están clasificados en 20 familias que contienen 208 géneros y a su vez 2231 especies descritas hasta la fecha. El 12% de esa diversidad está distribuida en México con 281 especies que pertenecen a 38 géneros y 8 familias que incluyen la Buthidae (de importancia médica debido al problema de salud pública que genera su picadura) y no Buthidae como Caraboctonidae, Chactidae, Diplocentridae, Euscorpidae, Superstitioniidae, Typlochactidae and Vaejovidae. A esta última se le considera la familia más diversa (Santibáñez-López, Francke, Ureta, & Possani, 2015).

El veneno de alacrán es una mezcla de cientos de sustancias biológicamente activas, la mayoría de tipo proteico o peptídico (Utkin, 2015). Dentro de los componentes de tipo proteico se encuentran enzimas como las fosfolipasas, las metaloproteinasas, las serin-proteasas y las hialuronidasas, entre otras. Como componentes de tipo peptídico se han identificado las toxinas que modulan o bloquean la actividad de canales iónicos de sodio, potasio, cloro y calcio, además de los péptidos de defensa del huésped como las defensinas y los péptidos sin enlaces disulfuro. Estos componentes forman parte del arsenal de defensa de estos animales.

Por muchos años, el estudio de veneno de alacranes se ha centrado en el aislamiento y posterior caracterización bioquímica de ciertos componentes, principalmente toxinas y otras moléculas de interés biotecnológico (Rendón-Anaya, Delaye, Possani, & Herrera-estrella, 2012). Se estima que existen más de 200,000 componentes de veneno distribuidos en las más de 2,000 especies de alacranes que se tienen reportadas hasta el momento, sin embargo, menos del 2% de estos componentes se han identificado o caracterizado (Zhang et al., 2015).

Una de las razones de la baja cantidad de componentes caracterizados de veneno es debido a la necesidad de coleccionar un gran número de especímenes para poder tener la cantidad suficiente de veneno y así lograr aislar los componentes menos representados. Este procedimiento implica un impacto negativo en las poblaciones de alacranes silvestres. Los alacranes, además de controlar otras poblaciones de artrópodos, son el soporte de otros depredadores superiores, por lo tanto, juegan un papel importante en sus ecosistemas. Otra razón importante se debe a que los permisos para recolección de especies son cada vez más restrictivos con respecto a especies particulares y número de muestra autorizada para

recolección; por lo tanto, se deben de buscar alternativas para lidiar con la falta de especímenes para la caracterización bioquímica clásica.

En un principio se desarrollaron algunas técnicas para resolver las limitaciones antes expuestas, una de ellas consistía en hacer búsqueda de componentes de veneno en el cDNA obtenido a partir del RNAm (RNA mensajero) de glándulas venenosas (Valdez-Velázquez, Quintero-Hernández, Romero-Gutiérrez, Coronas, & Possani, 2013) y la expresión heteróloga de péptidos para caracterización (Dai et al., 2012; Li & Xia, 2018).

En los últimos años, los estudios proteómicos y transcriptómicos de los venenos y de las glándulas venenosas han permitido la identificación de un gran número de componentes con respecto a las metodologías tradicionales de bioquímica y de biología molecular (Rodríguez de la Vega, Schwartz, & Possani, 2010); sin embargo, la utilización de las nuevas tecnologías de secuenciación masiva de alto rendimiento para los análisis de los venenos y de glándulas venenosas hizo posible comprender la enorme diversidad de componentes de veneno de alacrán de naturaleza peptídica (Santibáñez-López, Cid-Uribe, Batista, Ortiz, & Possani, 2016). Una de las ventajas más resaltables es que sólo se necesitan pocos animales para ambos estudios, la cantidad de muestra dependerá de los tamaños del alacrán y de su telson (que contiene la glándula venenosa), y para especies de tamaño grande el análisis completo se puede realizar con la muestra de un solo espécimen.

La caracterización del transcriptoma a partir del RNAm de la glándula venenosa del alacrán es una de las metodologías más utilizadas hoy en día, ya que nos permite tener una visión global de la maquinaria molecular de este órgano (Utkin, 2015). Una de las ventajas de estudiar los genes que codifican para los diferentes componentes de veneno es que las toxinas principales se logran identificar de manera eficaz y de esa manera se puede contribuir al desarrollo de antivenenos específicos y eficaces.

En los últimos años se han reportado diversos transcriptomas de glándulas venenosas de alacranes (tabla 1) pertenecientes a la familia Buthidae, Caraboctonidae, Euscorpiidae, Hemiscurpiidae, Scorpionidae, Superstitioniidae y Vaejoividae. Sin embargo, la familia Buthidae es la más estudiada por ser de importancia médica. La mayoría de los transcritos se han secuenciado como moléculas de cDNAs con el método de Sanger pero hay otros que se han secuenciado con la tecnología de pirosecuenciación 454 y los más recientes con la plataforma Illumina.

Tabla 1. Revisión de los transcriptomas de glándulas venenosas reportadas en la literatura.

Familia	Especie	Tipo de secuenciación	Referencia
Buthidae	<i>Lychas mucronatus</i>	Sanger	(Ruiming et al., 2010a)
	<i>Hottentotta judaicus</i>	Sanger	(Morgenstern et al., 2011)
	<i>Lychas mucronatus</i>	Sanger	(Ma et al., 2012)
	<i>Isometrus maculatus</i>	Sanger	
	<i>Centruroides noxius</i>	454	(Rendón-Anaya et al., 2012)
	<i>Centruroides tecomanus</i>	Sanger	(Valdez-Velázquez et al., 2013)
	<i>Parabuthus stridulus</i>	Sanger	(Mille, Peigneur, Diego-García, Predel, & Tytgat, 2014)
	<i>Mesobuthus gibbosus</i>	Sanger	(Diego-García, Caliskan, & Tytgat, 2014)
	<i>Mesobuthus eupeus</i>	Sanger	(Kuzmenkov et al., 2015)
	<i>Androctonus bicolor</i>	Sanger	(Zhang et al., 2015)
	<i>Tityus bahiensis</i>	454	(De Oliveira, Candido, Coronado Dorce, & Junqueira-De-Azevedo, 2015)
	<i>Hottentotta conspersus</i>	Sanger	(Mille, Peigneur, Predel, & Tytgat, 2015)
	<i>Odonthubuthus doriae</i>	Sanger	(NaderiSoorki, Galehdari, Baradaran, & Jalali, 2016)
<i>Centruroides hentzi</i>	Illumina	(Ward, Ellsworth, & Rokyta, 2018)	
Caraboctonidae	<i>Hadrurus gertschi</i>	Sanger	(E. F. Schwartz, Diego-Garcia, Rodríguez de la Vega, & Possani, 2007)
	<i>Hadrurus spadix</i>	Illumina	(Rokyta & Ward, 2017)
Euscorpidae	<i>Scorpiops jendeki</i>	Sanger	(Ma et al., 2009)
	<i>Scorpiops margerisonae</i>	Sanger	(Ma et al., 2012)
	<i>Megacormus gertschi</i>	Illumina	(Santibáñez-López et al., 2017)
Hemiscorpiidae	<i>Hadogenes troglodytes</i>	Sanger	(Zhong et al., 2017)
Scorpionidae	<i>Pandinus imperator</i>	454	(Roeding et al., 2009)
	<i>Heterometrus petersii</i>	Sanger	(Yibao et al., 2010)
	<i>Pandinus cavimanus</i>	Sanger	(Diego-García et al., 2012)
	<i>Urodacus yaschenkoi</i>	Illumina	(Luna-Ramírez, Quintero-Hernández, Juárez-González, & Possani, 2015)
	<i>Heterometrus spinifer</i>	Illumina	(Deng et al., 2018)
Superstitioniidae	<i>Superstitionia donensis</i>	Illumina	(Santibáñez-López et al., 2016)
Vaejovidae	<i>Vaejovis mexicanus</i>	Sanger	(Quintero-Hernandez et al., 2015)
	<i>Vaejovis intrepidus</i>	Sanger	
	<i>Vaejovis subcristatus</i>	Sanger	
	<i>Vaejovis punctatus</i>	Sanger	

El primer transcriptoma de glándulas venenosas de alacrán reportado en la literatura pertenece a *Hadrurus gertschi* (E. F. Schwartz et al., 2007). Para la realización de este trabajo, se construyó una librería de cDNA de las glándulas venenosas de este alacrán y se secuenció por el método de Sanger. En este estudio se identificaron 147 ESTs (Expressed Sequence Tags) de buena calidad que fueron agrupadas en 68 transcritos únicos (con 20 contigs y 48 singlets). La diversidad de componentes del veneno que se encontró con esta metodología fue la siguiente: el 3% de las secuencias tuvieron identidad con α y β toxinas de potasio; el 5% tuvo identidad con péptidos sin enlaces disulfuro (NDBPs), 29% de los transcritos tuvieron identidad con proteínas implicadas en procesos celulares, el 34% de las secuencias

no tuvo un marco de lectura abierto (ORF), el 2.9% tuvo identidad con componentes de veneno, el 3% tuvo identidad con proteínas con función desconocida y el 19.1 no tuvo identidad con las bases de datos (E. Schwartz, Diego-garcia, Rodríguez de la Vega, & Possani, 2007).

Además de la secuenciación por Sanger, se han caracterizado transcriptomas mediante otras plataformas, una de ellas la 454, siendo uno de los primeros esfuerzos por obtener una visión más global de la composición molecular de la glándula venenosa. La pirosecuenciación o 454 es una tecnología de segunda generación que por quimioluminiscencia detecta el pirofosfato liberado durante la elongación de la cadena complementaria de ADN, esto permite la detección de las bases en tiempo real (Ambarar, Gupta, Trakroo, Lal, & Vakhlu, 2016).

El transcriptoma del cDNA de *Centruroides noxius*, un alacrán perteneciente a la familia Buthidae reportado por Rendón-Anaya y colaboradores en 2012, fue el primero en ser secuenciado por la tecnología 454. En este reporte se construyeron 3 librerías de cDNA; una del cuerpo completo del animal sin el telson, los otros dos experimentos consistieron en caracterizar el cDNA de las glándulas venenosas antes y después de ser ordeñadas. Como resultados se obtuvieron más de 3 millones de *reads* (lecturas de secuencia) con una longitud media de isotigs de 950 pb. La anotación de las secuencias dio como resultado la formación de 72 isogrupos (que representan el 0.4% del número total de transcritos ensamblados) relacionados con componentes de veneno donde se incluyeron anotaciones de toxinas, péptidos antimicrobianos, enzimas como metaloproteinasas, hialuronidasas y fosfolipasas; además de otros componentes de veneno cuya función aún no se ha descrito (Rendón-Anaya et al., 2012).

Con el paso del tiempo, se han desarrollado plataformas de secuenciación cada vez más eficientes y menos costosas. La secuenciación por la plataforma Illumina es una de ellas y hoy en día es la tecnología que más se utiliza. El primer transcriptoma de glándulas venenosas de alacrán secuenciado por este método fue el de *Urodacus yaschenkoi* que pertenece a la familia Scorpionidae (Luna-Ramírez et al., 2015). En la publicación se reportan más de 80 millones de *reads* (lecturas) secuenciados y ensamblados en más de 200 mil secuencias que, después de ser filtradas por anotación, resultaron en 111 secuencias únicas relacionadas con proteínas de veneno. La diversidad de los componentes identificados se distribuyen de la siguiente manera: Los DBPs que incluyen toxinas de sodio, potasio y calcio representaron el 30% de las secuencias anotadas; los NDBPs el 21%; las enzimas representaron un 31% de los transcritos e incluyeron fosfolipasas, metaloproteinasas y serin proteasas; la abundancia de los péptidos tipo La1 fue de 4%; los inhibidores de proteasas representaron un 3% seguido de alérgenos con

un 2% y por último, aquellas proteínas de veneno sin función definida con un 9% (Luna-Ramírez et al., 2015).

A diferencia de los estudios transcriptómicos, la proteómica de los venenos tiene un poco más de historia: los primeros esfuerzos que se hicieron para caracterizar el veneno de alacranes consistieron en aislar directamente toxinas del veneno y posteriormente evaluarlas para conocer su actividad biológica (Batista et al., 2018; Rochat, Rochat, Miranda, & Lissitzky, 1967). En tiempos recientes, se han creado técnicas de secuenciación masiva de proteínas que funcionan por medio de cromatografía en línea. Esta metodología consta de un cromatógrafo de líquidos de alto rendimiento acoplado a un espectrómetro de masas (LC-MS/MS). Una de las ventajas de esta tecnología es la poca cantidad de veneno que se necesita para el análisis, permitiendo la identificación de una mayor cantidad de componentes del veneno.

Se tienen pocos reportes de proteómica masiva de venenos de alacrán, entre ellos se encuentran el análisis del veneno de *Tytius serrulatus* (Rates et al., 2008), *Superstitioniae donensis* (Santibáñez-López et al., 2016), *Megacormus gertschi* (Santibáñez-López et al., 2017), *Tytius obscurus* (De Oliveira et al., 2018), *Centruroides hentzi* (Ward et al., 2018). La identificación de los componentes de veneno en los proteomas de cada uno de ellos fue por medio de LC-MS/MS, identificándose 28, 26, 23, 41 y 59 proteínas respectivamente.

4. ANTECEDENTES

La familia Vaejovidae es la más diversa de México y está ampliamente distribuida desde Canadá hasta Guatemala. México tiene una gran diversidad ya que cuenta con 149 especies de alacranes que pertenecen a 21 géneros. Esta familia está constituida por 3 sub-familias: Smeringurinae, Vaejovinae y Syntropinae (Soleglad & Fet, 2008) y es esta última la que tiene mayor distribución (figura 1). Dentro de esta familia se encuentra el género *Thorellius* que es uno de los Vaejovidos más grandes. Este género se encuentra distribuido en los estados de Aguascalientes, Colima, Guanajuato, Guerrero, Jalisco, Hidalgo, Puebla, Oaxaca, Michoacán, Nayarit, Sinaloa, Estado de México, Querétaro, Veracruz y Zacatecas (Soleglad & Fet, 2008).

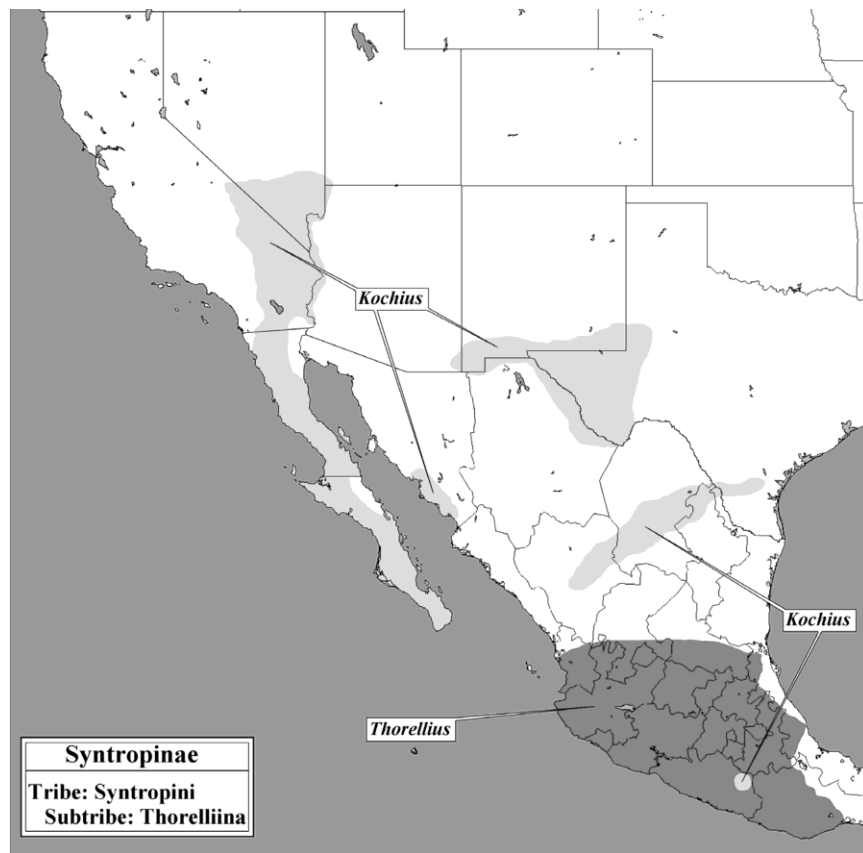


Figura 1: Distribución de la sub-familia Syntropinae. Tomado de (Soleglad & Fet, 2008)

Este trabajo se enfocará en la especie *T. atrox* la cual se distribuye en Jalisco y en Colima. Existe poca información relacionada con la ecología de *Thorellius*, los animales de este género sólo son activos durante la época de lluvias y se ocultan durante la época seca (González-Santillán & Prendini, 2013), lo que dificulta su colecta.

Quintero-Hernández y colaboradores en 2015 reportaron el transcriptoma de la glándula venenosa de 4 miembros de la familia Vaejovidae: *Vaejovis intrepidus* (ahora nombrado *T. atrox*), *V. mexicanus*, *V. subcristatus* y *V. punctatus*. La tecnología utilizada para la secuenciación del transcriptoma fue por el método de Sanger. En ese trabajo para *Vaejovis intrepidus* (*T. atrox*) se secuenciaron 103 clonas de las cuales se encontraron 55 ESTs agrupadas en 17 contigs y 48 singlets. Durante el proceso de anotación, 10.7% de las secuencias tuvieron identidad con toxinas, 22.3% tuvo identidad con péptidos de defensa del huésped, 19.4% tuvo identidad con proteínas sin función definida, 3.9% tuvo identidad con péptidos La1, 1.9% tuvo identidad con otros componentes de veneno y, por último, 28.1% no presenta similitud con las secuencias reportadas en las bases de datos.

Con los antecedentes descritos anteriormente, se concluyó que esta especie tiene una gran cantidad de transcritos con potencial biotecnológico. El análisis transcriptómico mencionado permitió,

tener un panorama global de la composición del veneno y la glándula venenosa de *T. atrox*. Para ampliar aún más este panorama y poder tener una visión completa de los componentes venenosos de *T. atrox*, en este trabajo se planteó hacer el análisis transcriptómico utilizando la plataforma de Illumina, además de un análisis proteómico de los venenos.

5. OBJETIVOS

5.1 Objetivo principal

Caracterizar el transcriptoma de la glándula venenosa y el proteoma del veneno del alacrán *T. atrox*, así como escoger una secuencia candidata para comprobar su actividad biológica.

5.2 Objetivos específicos

-Colectar especímenes de *T. atrox*.

-Caracterizar el transcriptoma de la glándula venenosa del alacrán *T. atrox*:

- Obtener una librería de cDNA a partir del RNAm de la glándula venenosa.
- Realizar un ensamble *de novo* con las lecturas obtenidas de la secuenciación de la librería de cDNA por la plataforma Illumina.
- Identificar aquellos transcritos que tengan identidad con proteínas de veneno a partir de librerías previamente reportadas en la literatura.
- Analizar los niveles de expresión por medio de la cuantificación de los transcritos obtenidos del RNA-Seq para identificar las secuencias más expresadas.

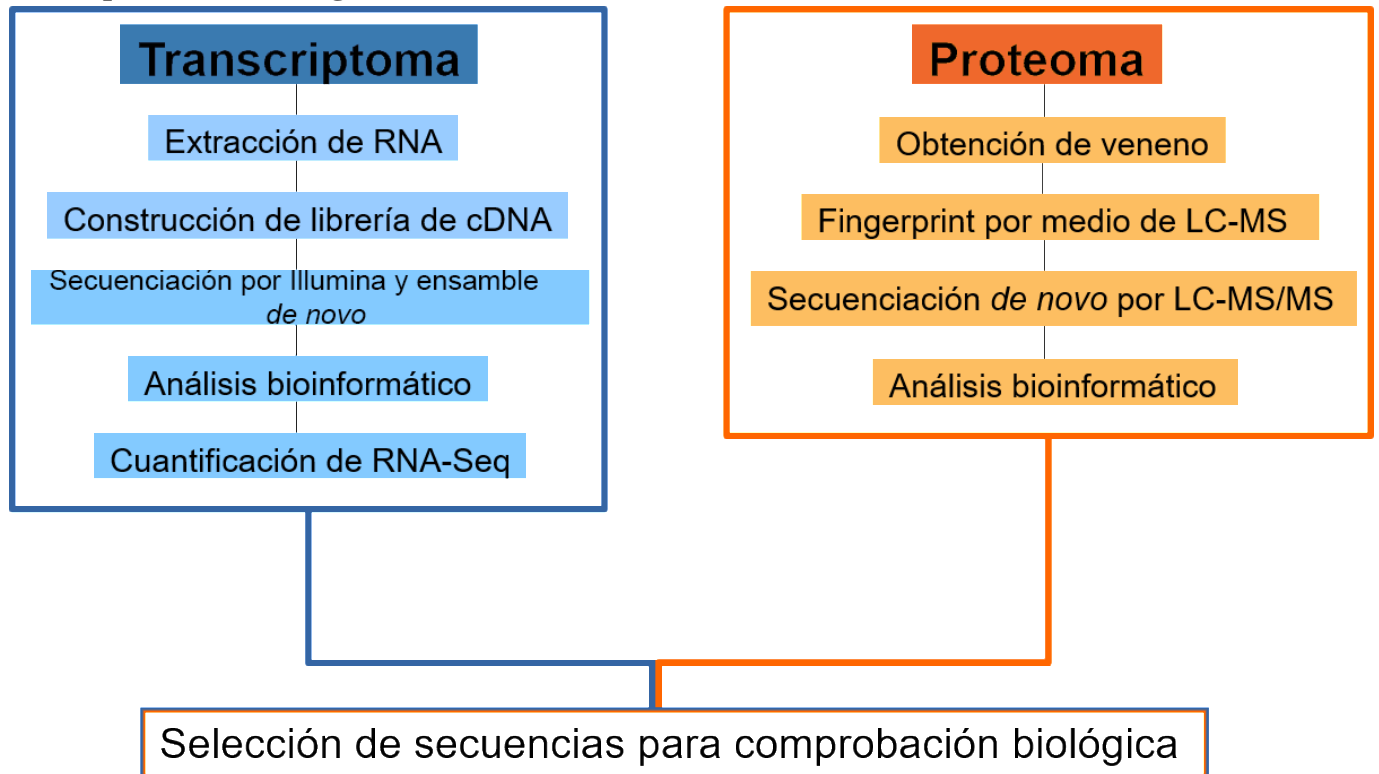
-Caracterizar el proteoma del alacrán *T. atrox*:

- Obtener el veneno para realizar un fingerprint por medio de LC-MS.
- Digerir veneno y secuenciar *de novo* los fragmentos de proteínas por LC-MS/MS.
- Identificar las proteínas secretadas en el veneno utilizando como template la colección de transcritos obtenida en el análisis transcriptómico.

-Seleccionar una proteína presente en el transcriptoma y en el proteoma para caracterizar su actividad biológica.

6. MATERIAL Y MÉTODOS

6.1 Esquema metodológico



6.2 Metodología

6.2.1. Colecta de material biológico

Se colectaron cuatro especímenes que fueron identificados como *Thorellius atrox* por el Doctor Oscar Francke del Instituto de Biología-UNAM. Los alacranes fueron recolectados en el municipio de Coquimatlán, Colima (19°12'39''N 103°48'24''W). Se mantuvieron en cautiverio a temperatura ambiente, con ciclos naturales de luz/oscuridad, alimentados mensualmente con grillos y con acceso libre a agua. Los alacranes se colectaron con el permiso otorgado por la SEMARNAT (SGPA/DGVS/12063/15) y concedido a la Doctora Laura Valdez Velazquez, profesora investigadora de la Universidad de Colima.

6.2.2 Análisis transcriptómico de la glándula venenosa del alacrán *T. atrox*

6.2.2.1 Extracción de RNA de la glándula venenosa

Los alacranes se ordeñaron por electroestimulación cinco días previos a la extracción del RNA, para poner en marcha la maquinaria molecular para la producción de veneno. El veneno obtenido con este procedimiento se utilizó para el análisis proteómico que se describirá en la sección de proteómica. Los animales no se alimentaron durante este procedimiento.

Para la extracción del RNA se cortó el telson de 4 especímenes previamente identificados como *T. atrox*, el tejido se homogenizó con un pistilo para microtubo (Daigger Vernon Hills, IL, USA). Se utilizó el kit extracción de RNA “SV Total RNA Isolation System Kit (Promega, Madison, WI, USA)” (Anexo 1). La pureza del material obtenido se evaluó en el equipo Agilent RNA 2100 Bioanalyzer system.

6.2.2.2 RNA-Seq, ensamble y anotación del transcriptoma

La construcción de la librería de cDNA (cDNA de doble cadena) se realizó con el kit True seq strandet mRNA simple (Illumina). El proceso de secuenciación pareada se hizo con el equipo Illumina Genome Analyzer xII. Posteriormente, el ensamble *de novo* y la anotación de los *reads* se realizó con el software Trinity (disponible en <https://github.com/trinityrnaseq/trinityrnaseq/wiki>) y Trinotate (disponible en <https://trinotate.github.io/>).

Trinotate es parte del esquema de trabajo de Trinity y se utiliza para hacer la anotación funcional de los transcritos resultantes del ensamble *de novo* hechos con Trinity. Este software utiliza diferentes herramientas como BLAST y Swiss Prot para buscar homólogos en bases de datos conocidas; PFAM y HMMER para identificar dominios funcionales, signalP y tmHMM para identificar péptido señal y/o dominios transmembranales y también hace búsquedas en diferentes bases de datos como eggNOG (búsqueda de ortólogos), GO (búsqueda de ontología) y Kegg (rutas metabólicas).

Se utilizaron herramientas como blastp, blastn, blastx (disponibles en <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) para identificar coincidencias en las secuencias nucleotídicas y aminoacídicas con las secuencias reportadas en las bases de datos. Para determinar la masa molecular del precursor se utilizó ProtParam (disponible en <http://web.expasy.org/protparam/>), adicionalmente se utilizaron las herramientas SpiderP (<http://www.arachnoserver.org/spiderP.html>) y ProP (<http://www.cbs.dtu.dk/services/ProP/>) para identificar modificaciones post-traduccionales (propéptidos).

Para la selección de las secuencias relacionadas con veneno se tomaron en cuenta los siguientes criterios: a) El transcrito tiene similitud con alguna de las secuencias reportadas previamente para *T. atrox*. b) La secuencia contiene alguno de los 22 dominios funcionales relacionados con animales venenosos como alacranes, serpientes, arañas, conos y abejas (Anexo 2). C) El transcrito tuvo identidad de secuencia con alguna de las 6,000 secuencias reportadas en el Proyecto de anotación de toxinas de animales venenosos (Animal Toxin Annotation Project (http://www.uniprot.org/biocuration_project/Toxins/statistics)). D) El *E-value* de la secuencia identificada en el transcriptoma contra las bases de datos reportadas debe de ser menor a 1×10^{-4} . Este

valor de e representa la probabilidad de que la anotación asignada por similitud de secuencia se haya dado por azar o no.

Todos los alineamientos se realizaron con el software Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) con los parámetros estándar.

6.2.2.3 Cuantificación de RNA-Seq

Una vez realizada la identificación de los transcritos relacionados con veneno, se cuantificaron los niveles de expresión *in silico* del transcriptoma por medio del algoritmo Salmon (disponible en <https://combine-lab.github.io/salmon/>). Se utilizó como referencia el ensamble obtenido previamente por Trinity para el mapeo de lecturas y se emplearon los parámetros estándar para las cuantificaciones por transcrito.

Para llevar a cabo el proceso sólo se requiere el ensamble obtenido por Trinity y el archivo que contiene los *reads* provenientes de la secuenciación. Como resultado de la ejecución del programa se obtuvieron 2 archivos: “archivo.index” que contiene el índice o las coordenadas de cada uno de los transcritos ensamblados (código binario) y una carpeta “salmon.quant” que contiene los resultados de la cuantificación. Dentro de la carpeta “salmon.quant” existe un archivo de texto llamado “quant.sf” que contiene información como el número de *reads*, la longitud efectiva, los transcritos por millón y el número de *reads* mapeados (tabla 2).

Tabla 2 Descripción del archivo de salida de Salmon.

Nombre de la columna	Significado
Name (nombre)	Identificador de Trinity
Lenght (longitud)	Longitud del transcrito cuantificado
Effective length (longitud efectiva de transcrito)	Longitud efectiva del transcrito cuantificado
TPM (transcritos por millón)	Abundancia relativa del transcrito en unidades de transcritos por millón. Es una cuantificación normalizada.
NumReads (número de <i>reads</i>)	Número de <i>reads</i> que mapearon en un transcrito en particular según el algoritmo de salmon. Esta cuantificación no está normalizada.

Para realizar la cuantificación de transcritos del RNA-Seq se utilizaron los parámetros estándar (anexo 3) y adicionalmente se le indicó al programa que los *reads* utilizados para el análisis son pareados, por lo tanto, el programa se alimentó de dos archivos fastq. Una vez obtenida la tabla de resultados, se hizo un ordenamiento de mayor a menor número de TPMs.

6.2.3 Análisis proteómico del veneno del alacrán *T. atrox*

6.2.3.1 Determinación de masas moleculares del veneno del alacrán *T. atrox*

Para la determinación de las masas moleculares se utilizaron 4 µg de la fracción soluble del veneno disuelta en ácido fórmico al 0.1% (solvente A). Esta mezcla se desaló con una columna Zip Tip C18 (Millipore; Billerica, MA, USA) y se aplicó en un sistema LC-MS (Liquid Chromatography-Mass Spectrometry por sus siglas en inglés) compuesto de una bomba de nanoflujo EASY-nLC II (Thermo-Fisher Co.; San Jose, CA, USA) acoplado a un espectrómetro de masas LTQ-Orbitrap Velos (Thermo-Fisher Co., San Jose, CA, USA) con una fuente de ionización tipo nano-electrospray (ESI).

La cromatografía de fase reversa se realizó en un gradiente de 10-80% de solvente B (agua/acetonitrilo con 0.1% de ácido fórmico) en 290 minutos utilizando una columna capilar C18 hecha en casa (diámetro interno de 0.75 µm y 10 cm largo). El flujo del sistema LC fue de 300 nanolitros/minuto.

Las masas se obtuvieron como [M] por deconvolución automática utilizando el software Xcalibur (versión 2.2, Waltham, MA, USA, 2011). Los datos crudos se filtraron con una calculadora de Excel hecha en casa (basada en la calculadora de aductos de espectrometría de masas del laboratorio de Fiehn de la UC Davis) para eliminar abductos comunes, oxidación de aminoácidos, deshidratación, desaminación y pérdida de monóxido de carbono. Para péptidos con una masa molecular menor de 3,000 Da se tomó en cuenta la masa monoisotópica y para aquellas mayores de 3,000 Da, se eligió la masa molecular promedio.

6.2.3.2 Secuenciación de proteínas por LC-MS/MS

La identificación de las proteínas presentes en el veneno se realizó mediante digestión proteolítica seguida de LC-MS/MS. La reducción de cisteínas se hizo utilizando 10 mM de ditioneitol (DTT) a 56°C por 30 minutos y después alquilada con 50 mM de yodocetamida por otros 30 minutos; esta reacción se hace resguardada de la luz; posteriormente, la muestra es desalada.

Para la reacción enzimática con tripsina (Promega, Madison, WI) se usaron 50 µg de proteínas solubilizadas en 40 mM de bicarbonato de amonio con la enzima en una proporción 1:25. Esta mezcla se incubó toda la noche a 37 °C y a un pH de 8.1.

Se utilizaron 5 µg de proteína digerida para la identificación de fragmentos por LC-MS/MS. La muestra se fraccionó por cromatografía de fase reversa en una columna capilar C18 casera (30 cm de longitud por 0.5 mm de diámetro interno). La resina que se usó como fase estacionaria es una sílica

Júpiter C-18 con un tamaño de partícula de 4 μM (Torrance, CA, USA). El gradiente que se utilizó para la corrida fue de 5 a 80% de solvente B durante 120 minutos con un flujo de 400 nL/min.

6.2.3.3 Análisis de datos obtenidos por LC-MS/MS

Todos los espectros de los fragmentos obtenidos por LC-MS/MS se analizaron con el software Proteome Discoverer utilizando el algoritmo de Sequest (Thermo Fisher Scientific, San Jose, CA, USA) para identificar aquellos fragmentos de proteínas que tienen similitud con la base de datos de interés. En este caso, la base de datos consiste en las secuencias ensambladas en el transcriptoma traducidas en los 6 marcos de lectura.

Los parámetros utilizados para la búsqueda de coincidencias fueron: modificación constante: carbamidometilación de cisteínas, como modificaciones variables la oxidación de metioninas y desamidación, la enzima proteolítica fue tripsina, número máximo de sitios donde no cortó la enzima: 2, la tolerancia de la masa molecular del precursor de 20 ppm y una tolerancia de la masa de los fragmentos de 0.6 Da. Adicionalmente se tomó un FDR (False Discovery Rate o tasa de falsos positivos) de 0.01 y 0.05 como estricto y relajado, respectivamente.

6.2.4 Síntesis y caracterización biológica de secuencias identificadas en el transcriptoma y el proteoma

El péptido seleccionado para caracterizar su actividad biológica fue aquel que cumplió con siguientes criterios: a) El transcrito de interés tuvo identidad con algún componente de veneno en las bases de datos con un E-value menor a 1×10^{-5} ; b) la secuencia madura del transcrito codificado se encontró completo, c) por medio de cuantificación de RNA-Seq se determinó que la secuencia tiene altos niveles de expresión y d) la secuencia madura del péptido candidato también se encontró en el proteoma.

La estrategia utilizada para producir el péptido de interés fue la síntesis química de péptidos en fase sólida (Merrifield, 1963), empleando el grupo Fmoc (N-(9-fluorenil) metoxycarbonil) como protector del amino terminal de los aminoácidos.

Las condiciones para la síntesis fueron las siguientes: se utilizó la resina Rink amida MBHA (con una densidad de 0.54 mol/g de resina) como soporte sólido. La activación del carboxilo de los aminoácidos Fmoc se realizó con HATU (Hexafluorofosfato de 2-(1H-7-azabenzotriazol-1-il)-1,1,3,3-tetrametiluronio), HOBT (N-hidroxibenzotriazol) y se usó DIPEA (Diisopropiletilamina) para la formación del enlace peptídico, la proporción de los compuestos antes mencionados fue de 1:0.9:0.9:2 (aminoácido Fmoc:HBTU:HOBT:DIPEA), esto para la reacción de acoplamiento. Para desproteger el

amino terminal del último aminoácido Fmoc acoplado a la resina se utiliza piperidina al 20% en DMF (dimetilformamida). Como prueba de monitoreo de desprotección/acoplamiento de cada uno de los aminoácidos se utiliza la prueba de kaiser (reacción con ninhidrina).

6.2.4.1 Medición de la actividad antimicrobiana

Para medir la actividad antibacteriana del péptido se utilizó el protocolo de microdilución (anexo 4); las cepas a evaluar fueron *Escherichia coli* (ATCC 25922), *Staphylococcus aureus* (ATCC 25923) y *Pseudomonas aeruginosa* (ATCC 9027) el medio de cultivo usado fue Mueller Hinton. Se partió de un título de 1×10^4 unidades formadoras de colonia (UFC) por mililitro y una concentración de péptido de 40 μM , 20 μM , 10 μM , 5 μM y 2.5 μM ; se hicieron las cinéticas de crecimiento bacteriano en un tiempo de 24 horas a una temperatura de 37 °C haciendo una medición de densidad celular en un lector de ELISA a 650 nm cada dos horas.

7. RESULTADOS Y DISCUSIÓN

7.1 RNA-Seq de la glándula venenosa del alacrán *T. atrox*

7.1.1 Extracción de RNA de la glándula venenosa y transcriptoma (RNA-Seq)

Se obtuvieron un total de 212 µg de RNA total; la calidad y la integridad se midieron en el bioanalyzer (anexo 5). Un parámetro importante al momento de medir la calidad del RNA es el RIN, que es la proporción del gen ribosomal 28S contra el 18S. En los artrópodos ocurre un fenómeno de procesamiento del 28S que hace que el fragmento quede del mismo tamaño que el 18S (Winnebeck, Millar, & Warman, 2010), así que no se puede hacer la estimación de este valor; sin embargo, se observó que no había degradación del RNA, por lo tanto, se procedió a la secuenciación de la muestra.

Como resultado de la secuenciación, se obtuvieron un total de 44,049,844 *reads* de buena calidad (anexo 6) y éstas a su vez se ensamblaron en 129,950 transcritos con una N50 (mediana de la longitud de los transcritos) de 1,849 bases. De los ensamblados obtenidos se anotaron 20,851 utilizando Trinotate, esto quiere decir que sólo los anotados tuvieron identidad en las bases de datos.

Debido a la gran diversidad de anotaciones de las secuencias del transcriptoma, se decidió utilizar como primer parámetro de distribución de funciones la clasificación de GO (Gen Ontology), esta clasificación describe cada uno de los genes y las características generales del mismo; basados en esto, las secuencias traducidas fueron clasificadas dentro de las siguientes ontologías: 33.3 % correspondiente a componentes celulares, 40.4 % fueron incluidos dentro de procesos biológicos y 26 % dentro de funciones moleculares (figura 2).

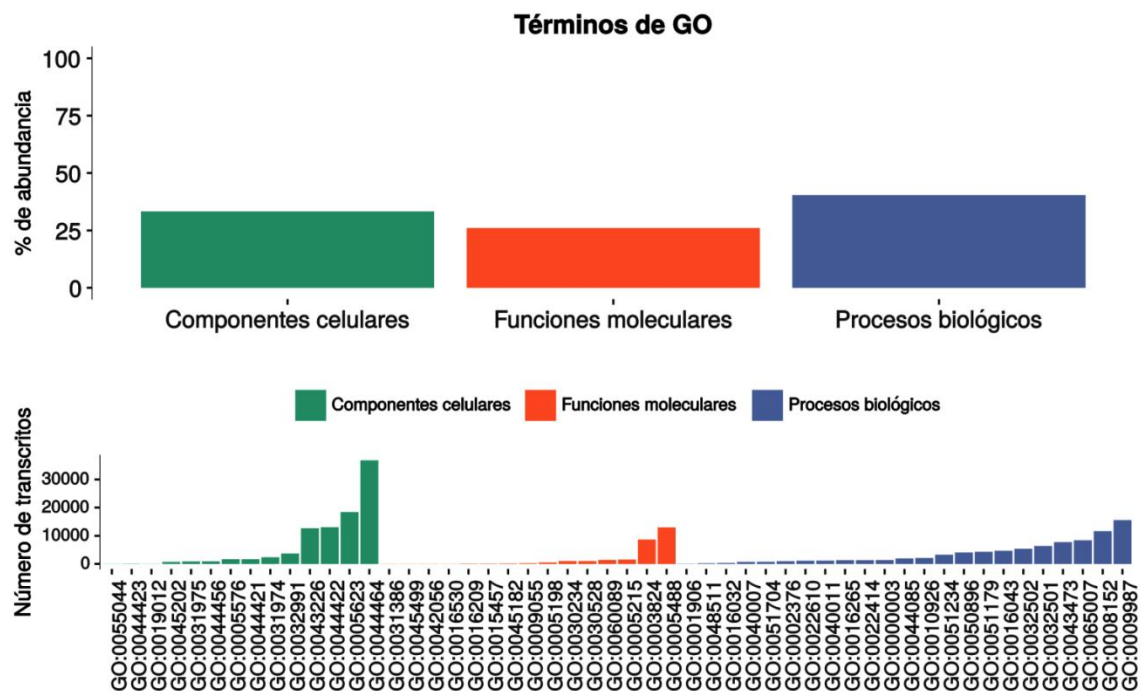


Figura 2: Distribución de términos GO en los transcritos anotados de la glándula venenosa del alacrán *T. atrox*. **A)** Distribución de los tres dominios principales de GO: Procesos biológicos (41 %), Componentes celulares (33 %) y Funciones moleculares (26 %). **B)** Distribución de las categorías de las ontologías del gen con su número de acceso, las barras de color verde corresponden a aquellas ontologías que están relacionadas con procesos biológicos, las barras de color naranja corresponden a la abundancia de las ontologías relacionadas con componentes celulares y las barras azules corresponden a aquellas ontologías relacionadas con funciones moleculares.

7.1.2 Análisis del transcriptoma

Durante el proceso de anotación se hizo una búsqueda de transcritos con identidad de componentes de veneno. Se identificaron 160 transcritos de esta categoría de los cuales el 25.5 % pertenecen a secuencias que tienen identidad con DBPs (incluyendo en este grupo toxinas que afectan a canales iónicos de sodio, potasio y calcio), 10.7 % pertenecen a HDPs (en este grupo se incluyen transcritos que tienen identidad con péptidos de la familia NDBP-2, NDBP-3, NNBP-4, péptidos aniónicos, waprín-like y defensinas), 34.4 % tuvo identidad con enzimas (incluyendo metaloproteinasas, fosfolipasas, hialuronidasas y serín proteasas), 4.4 % de los transcritos tuvieron identidad con péptidos tipo La1, 15 % con inhibidores de proteasas, 5 % de las secuencias se anotaron como CRISP (miembros de la superfamilia CAP) y por último, el 5 % se anotó como otros componentes de veneno (figura 3).

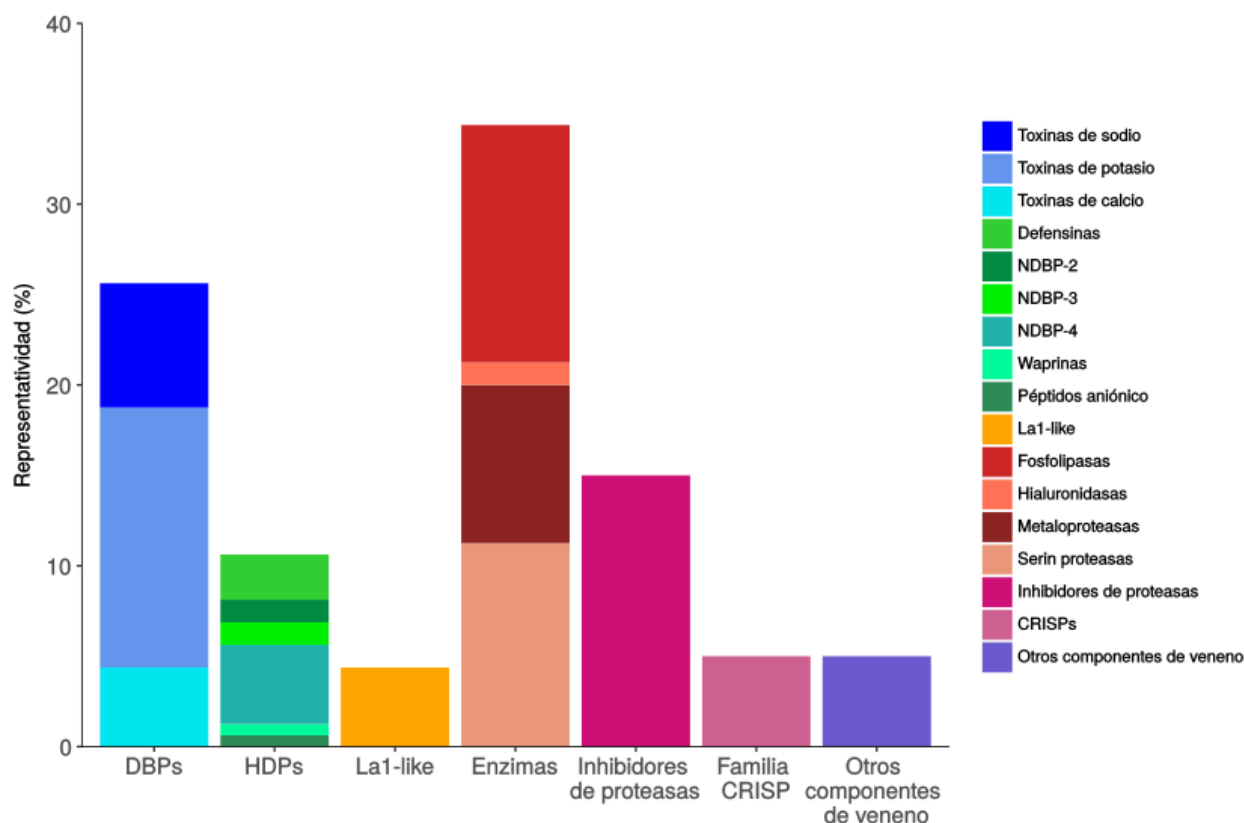


Figura 3. Diversidad relativa de transcritos anotados con identidad de proteínas de veneno. El grupo más representado es el que tiene identidad con enzimas, este grupo incluye fosfolipasas, metaloproteinasas, hialuronidasas y serin proteasas. La categoría DBPs incluye toxinas que afectan canales iónicos de sodio, potasio y calcio y dentro de la categoría HDPs se encuentran transcritos que tienen identidad con defensinas, waprín-like, NDMP-2, NDBP-3 y NDBP-4.

7.1.3 Nomenclatura de los transcritos.

Actualmente no existe una nomenclatura estándar para nombrar los transcritos identificados por RNAseq que son anotados como proteínas de veneno; con frecuencia se conserva el identificador que genera el ensamblador, esto dificulta la clasificación de las secuencias y puede generar confusión.

En este proyecto de tesis se propone una forma estándar para identificar los transcritos que resulta intuitiva para el consultante, esta nomenclatura se compone de los siguientes elementos: Los tres primeros dígitos corresponden a la especie descrita (Tat en el presente trabajo, que corresponde al alacrán *T. atrox*). Los siguientes tres caracteres definen la familia que codifica para el péptido/proteína con respecto a su función putativa, seguido de otros tres caracteres relacionados con la sub-familia (cuando se presenta el caso). Los últimos dos caracteres asignan el número de transcrito (Tabla 3). Existen casos donde se encuentra una secuencia reportada con anterioridad; en estas circunstancias, se conserva el nombre original para evitar proteínas duplicadas con diferente nombre.

Tabla 3. Nomenclatura utilizada para los transcritos identificados en el transcriptoma de *T. atrox*.

Código de la especie	Significado	Código de la familia	Significado	Código de la sub-familia	Significado	Ejemplo
Tat	<i>T. atrox</i>	NaT	Toxinas de sodio (NaTx)	Alp	Alfa-Na toxinas	TatNaTAlp01
				Bet	Beta-Na toxinas	TatNaTBet01
		KTx	Toxinas de potasio (KTx)	Alp	Alfa-K toxinas	TatKTxAlp01
				Bet	Beta-K toxinas	TatKTxBet01
				Kap	Kappa-K Toxinas	TatKTxKap01
				Del	Delta-K Toxinas	TatKTxDel01
				Scr	Tipo escorpina	TatKTxScr01
		CaT	Toxinas de calcio (CaTx)	Clc	Calcina	TatCaTClc01
				Lio	Tipo liotoxina	TatCaTLio01
				Ome	Omegascorpinas	TatCaTOme01
		HDP	Péptidos de defensa del huésped	Def	Defensinas	TatHDPDef01
				ND1-5	NDBPs familias 1-5	TatHDPND201
				Ani	Péptidos aniónicos	TatHDPAni01
				Wap	Tipo waprinas	TatHDPWap01
				PA2	Fosfolipasas A2	TatEnzPA201
		Enz	Enzimas	PLB	Fosfolipasas B	TatEnzPLB01
				PLD	Fosfolipasas D	TatEnzPLD01
				SeP	Serín proteasas	TatEnzSeP01
				MtP	Metaloproteasas	TatEnzMtP01
				Hya	Hialuronidasas	TatEnzHya01
				Srp	Tipo serpinas	TatPlnSrp01
		Pin	Inhibidores de proteasas	Kun	Tipo kunitz	TatPlnKun01
				La1	Tipo La1	TatOthLa101
Oth	Otros componentes de veneno	CRI	CRISP	TatOthCRI01		
		Und	Indefinido	TatOthUnd01		

7.1.4 Componentes de veneno identificados en el transcriptoma de las glándulas venenosas del alacrán *T. atrox*

7.1.4.1 Péptidos con enlaces disulfuro

En los venenos de alacrán, las toxinas que actúan sobre los canales iónicos están principalmente representados por los péptidos con enlaces disulfuro (DBPs), Estos péptidos tienen una longitud entre 28-120 aminoácidos y su estructura está conformada por 3 a 4 enlaces disulfuro. Estas toxinas actúan específicamente en canales de sodio, potasio o calcio, modificando la fisiología de las células, tejidos y órganos y pueden causar intoxicaciones severas que muchas veces terminan con la muerte del animal que sufrió la picadura de alacrán (Cao, Di, Wu, & Li, 2014; V. Quintero-Hernández, Jiménez-Vargas, Gurrola, Valdivia, & Possani L. D., 2013). En la literatura se reporta que los DBPs caracterizados hasta el momento son tóxicos para mamíferos, insectos y crustáceos, estas toxinas constituyen el arsenal de defensa y predación de estos arácnidos (Zhong et al., 2017).

Las toxinas que se describen a continuación se clasificaron de acuerdo con su familia estructural y el canal iónico que afectan.

Toxinas que actúan en los canales de sodio dependientes de voltaje (NaTxS).

Las toxinas de sodio (NaTxS) son las responsables de la mayoría de los síntomas de intoxicación. Se ha reportado que los venenos de los alacranes de la familia Buthidae (que incluye la mayor parte de las especies peligrosas para humanos) son más diversos en toxinas de sodio que aquellos que pertenecen a familias no Buthidae (De Oliveira et al., 2015; Ruiming et al., 2010b; Santibanez-Lopez & Possani, 2015)

Estos péptidos tienen una longitud de 58 a 76 aminoácidos, son estabilizados por 3 o 4 enlaces disulfuro (Possani, Becerril, Delepierre, & Tytgat, 1999), modifican la cinética de apertura-cierre del canal y para su estudio se dividen en dos familias que han sido clasificadas de acuerdo a su efecto fisiológico en los canales: las α -NaTx y las β -NaTx (Couraud, Jover, Dubois, & Rochat, 1982).

Las α -Na toxinas se unen en el sitio 3 del canal iónico de sodio e inhiben el proceso normal de inactivación de los canales. Las toxinas β se unen en el sitio 4 del canal cambiando el umbral de la activación, abriéndolo a potenciales más negativos (de la Vega & Possani, 2007; Mille et al., 2015; V. Quintero-Hernández et al., 2013).

En el transcriptoma de *T. atrox* se identificaron 13 secuencias que tienen identidad con toxinas de sodio, tres de ellas mostraron identidad de secuencia con α -Na toxinas previamente reportadas en la literatura y 10 β -Na toxinas (Figura 4).

De los tres transcritos identificados como posibles α -toxinas de sodio, se recuperaron dos precursores y una secuencia parcial, los CDS (secuencias codificantes) tuvieron similitud con el precursor de una toxina de *Anuroctonus phaiodactylus* (ahora conocido como *Anuroctonus pococky bajae* (Soleglad & Fet, 2004)) (acceso UniProt Q5MJP5), con un 47% y un 43% respectivamente (Figura 4B). Esta toxina de referencia, originalmente llamada phaiodotoxina, es una insectotoxina específica para canales de sodio y tiene actividad en el canal de sodio de insecto para/tipo E, mostrando efecto alfa (predominantemente) y beta (Valdez-Cruz et al., 2004).

Se identificaron 10 transcritos anotados como posibles β -toxinas de sodio, de los cuales se recuperaron 6 CDS. En el alineamiento de la figura 4B se muestran dos transcritos, TatNaTBet01 y TatNaTBet02 que tienen identidad con el precursor CsEI (acceso UniProt P01491) identificado en el alacrán *Centruroides sculpturatus* y la LVP1-alpha (acceso UniProt P0CI48) identificada en el alacrán *Lychas mucronatus*. La toxina CsEI es letal para aves y es medianamente toxica para ratones y grillos (Carcamo-Noriega et al., 2018), mientras que un fragmento del transcrito LVP-alpha (Lipolysis-

- 1 activating peptide 1-alpha chain por sus siglas en inglés) podría codificar para un péptido con similitud
- 2 a la neurotoxina BmKBTx identificada en el alacrán *Mesobuthus martensii* (Ruiming et al., 2010b).



Figura 4. Transcritos identificados en el transcriptoma de *T. atrox* con actividad putativa de toxinas de sodio. A) Distribución de los transcritos identificados en las sub-familias α y β NaTx. B) Alineamiento de los transcritos completos traducidos que potencialmente codifican para α -NaTx con las secuencias que tienen mayor identidad de acuerdo al algoritmo de BLAST. C) Alineamiento de dos precursores que codifican potencialmente para β -NaTxs con sus coincidencias más cercanas. En todos los alineamientos mostrados en las figuras, los puntos indican identidad de secuencia y los guiones medios indican *gaps* (indica la falta de un aminoácido en una posición determinada de la secuencia). Cuando está presente, cada uno de los elementos de las secuencias se mostrarán de la siguiente manera: El péptido señal está subrayado, el **péptido maduro** se muestra en letras negritas y las cisteínas están marcadas con color azul. El porcentaje de identidad fue calculado de acuerdo con la secuencia completa (incluyendo el péptido señal y el péptido maduro).

Toxinas que actúan sobre canales de potasio (Ktx).

Las toxinas que actúan sobre los canales de potasio tienen un arreglo estructural estabilizado por 3 o 4 enlaces disulfuro y tienen un efecto bloqueador del canal (Tytgat et al., 1999). Se clasifican en 5 familias: α -, β -, y γ -KTxs con un motivo estructural estabilizado por cisteínas α/β (CS α/β), la familia κ con un motivo CS α/α y la δ -KTx con un arreglo tipo Kunitz. En este grupo de toxinas también se encuentran las escorpinas, proteínas con dos dominios, uno de ellos con una similitud de secuencia a las β -Ktxs (Conde, Zamudio, Rodriguez, & Possani, 2000; Flores-Solis et al., 2016; Ramirez & Jimenez Vargas, 2016), esta es la razón por la que estas proteínas se consideran como una sub-familia de las toxinas β -Txs de potasio. En el transcriptoma de la glándula venenosa del alacrán *T. atrox* se identificaron 21 transcritos que codifican putativamente para toxinas de potasio.

La sub-familia α -KTx se considera la más diversa hasta el momento (Mille et al., 2015), con más de 170 péptidos reportados a la fecha de acuerdo a la base de datos Kalium (<http://kaliumdb.org/>), estas toxinas tienen una longitud media de 20 a 40 residuos y presentan un arreglo estructural CS α/β y son bloqueadoras de los canales de potasio por la interacción de la diada catalítica de las toxinas de potasio, esta involucra una lisina que interacciona con residuos del poro del canal y un residuo aromático (generalmente triptófano) que interactúa con el *loop* (asa) extracelular del canal (Soleglad & Fet, 2004).

En el transcriptoma de *T. atrox* se identificaron 15 secuencias que codifican potencialmente para α -KTxs. La Figura 5B muestra dos transcritos alineados con otras toxinas o precursores previamente reportados. Una de las referencias es la toxina Vm23 (acceso UniProt P0DJ32) identificada en el veneno del alacrán *Vaejovis mexicanus* y bloqueadora del canal de potasio dependiente de voltaje Kv1.3 (US 8,394,770 B2, 2008). La secuencia identificada como TatKTxAlp15 y la toxina Vm23 comparten el 83% de identidad con 6 cambios en el péptido maduro. Las otras dos referencias derivan de secuencias de cDNAs; la primera codifica para la BmTX1 (acceso UniProt A0RZD1) identificada en la glándula venenosa del alacrán *Mesobuthus martensii* y la segunda referencia se identificó previamente en la librería de cDNA del alacrán *T. atrox* (GenBank JZ8183). El transcrito TatKTxAlp15 y las últimas dos referencias codifican para α -KTxs estabilizadas por 3 enlaces disulfuro.

Las β -KTxs son péptidos largos de 50 a 75 aminoácidos con 3 enlaces disulfuro. En el transcriptoma de *T. atrox* no se identificaron toxinas β de potasio clásicas, pero se identificó una sub familia de estas: las escorpinas. Como anteriormente se indicó, las escorpinas están compuestas de dos dominios, el N-terminal que presenta similitud de secuencia con la cecropina y se ha reportado que este dominio presenta actividad antimicrobiana (Conde et al., 2000), por otro lado se encuentra el dominio C-terminal con

identidad y estructura parecida a una toxina β de potasio. Dado que el N-terminal de las escorpinas tienen actividad antimicrobiana, también se les considera como miembros de las defensinas, péptidos de defensa del huésped (Zhu & Tytgat, 2004).

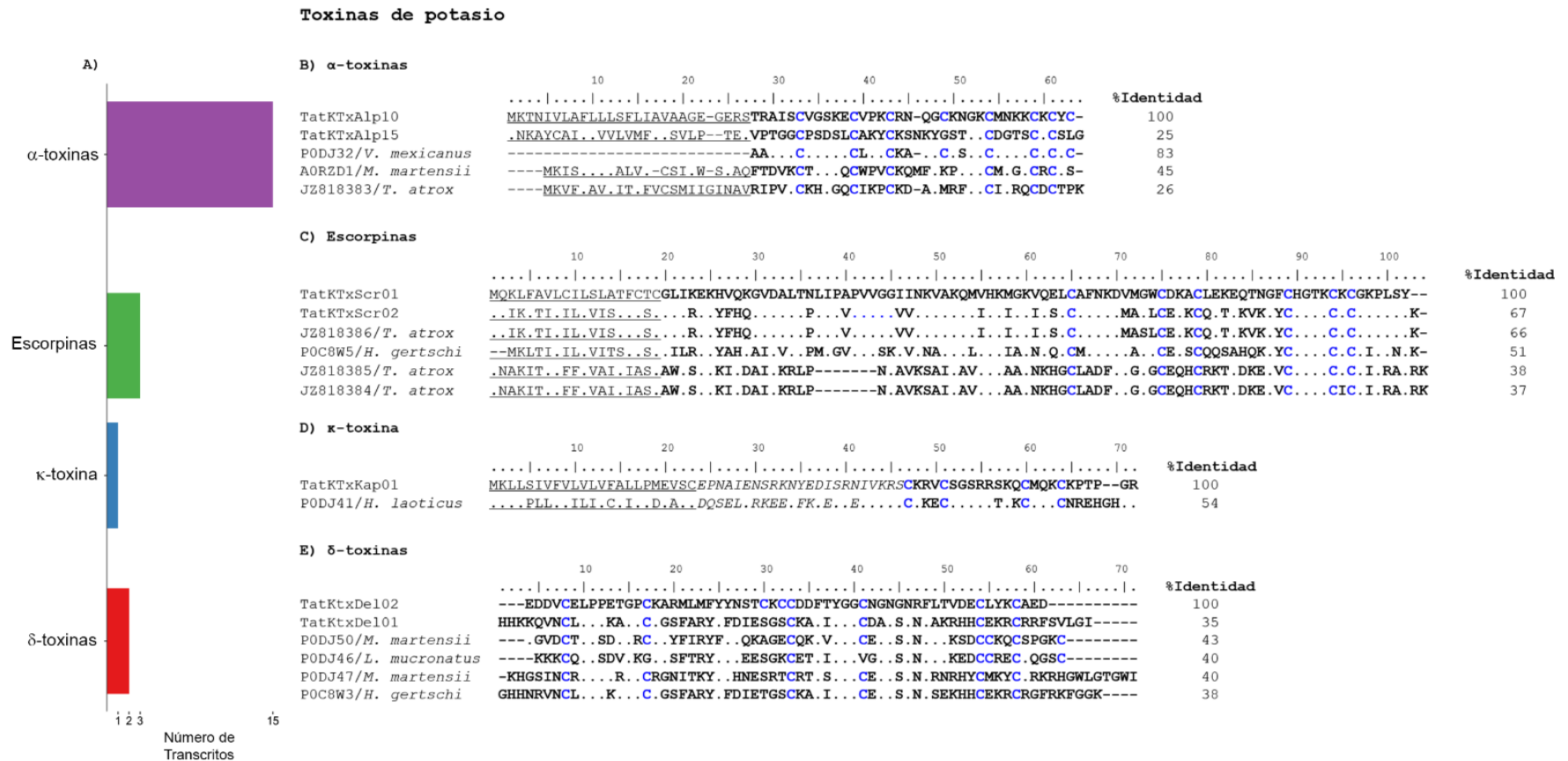


Figura 5. Secuencias identificadas en el transcriptoma de *T. atrox* con posible actividad de toxinas de potasio. A) Distribución de los transcritos identificados con respecto a la familia de toxinas de potasio. B) Alineamiento de dos de los precursores que tienen identidad de las α -KTxs y sus coincidencias más cercanas. C) Alineamiento de precursores con anotación de escorpinas contra una secuencia previamente identificada en la librería de cDNA de las glándulas venenosas del alacrán *T. atrox* y la HgeScpl2 como proteína de referencia. D) Alineamiento del precursor anotado como κ -KTx y su coincidencia más cercana identificada por BLAST. E) Alineamiento de las secuencias maduras anotadas como δ -KTx contra otras toxinas tipo Kunitz previamente reportadas. El péptido señal está subrayado, el **péptido maduro** se muestra en letras negritas, las cisteínas están marcadas con color azul y el *propéptido* en cursivas. El porcentaje de identidad fue calculado de acuerdo con la secuencia completa (incluyendo el péptido señal, el maduro y el propéptido).

En el transcriptoma de *T. atrox* se identificaron 3 transcritos que tienen anotación como escorpinas (Figura 5C). La comparación de las secuencias que se identificaron en este trabajo, como las reportadas previamente en su librería de cDNA (Verónica Quintero-Hernández et al., 2015) revela que uno de los transcritos nuevos codifica exactamente para la secuencia identificada como ViScplp2 y es 99% idéntica a la ViScplp1 con un cambio en un aminoácido en el péptido maduro. También se puede observar que los péptidos ViScplp1 y ViScplp2 (acceso GenBank JZ818384 y JZ818385 respectivamente) son más cortos que el péptido de referencia HgeScplp2 (acceso UniProt P0C8W5), identificado en la librería de cDNA del alacrán *Hoffmanihadrurus gertschi* (E. Schwartz et al., 2007); JZ818384 y JZ818385 tienen una delección de 7 aminoácidos en el dominio de cecropina.

La sub-familia κ es muy interesante, porque a diferencia de las otras familias de toxinas que afectan canales de potasio, esta tiene un motivo estructural CS α / α (Chagot et al., 2005). Se identificó un transcrito que codifica putativamente para una κ -KTx (Figura 5D). Para el alineamiento de los transcritos identificados con otras κ -KTx se utilizaron como referencia las secuencias más cercanas obtenidas por el algoritmo de BLAST. La toxina HelaTx1 (acceso UniProt P0DJ41) identificada en el veneno del alacrán *Heterometrus laoticus* bloquea el canal de potasio dependiente de voltaje Kv1.1 y Kv1.6 (Vandendriessche et al., 2013). El precursor identificado como TatKTxKap01 contiene un péptido señal seguido de un propéptido y la secuencia madura que dentro de su secuencia contiene un arreglo típico de 4 cisteínas.

Finalmente, la sub-familia δ -KTx que está integrada por toxinas tipo Kunitz, con actividad tanto de inhibidor de serin proteasas como de bloqueadora de canales de potasio, principalmente el Kv1.3 (Chen et al., 2013). Se identificaron dos transcritos con anotación de δ -KTxs; las secuencias maduras se muestran en la figura 5E se alinearon contra las toxinas BmKTT (UniProt P0DJ50), LmKTT-1a (UniProt P0DJ46), BmKTT-3 (UniProt P0DJ47) identificadas en el alacrán *Mesobuthis martensii* y la Hg1 (UniProt P0C8W3) identificada en el alacrán *Hadrurus gertschi*.

Toxinas que actúan en canales de calcio

A partir del veneno de algunos alacranes se han caracterizado péptidos que afectan canales de calcio. Se ha descrito que pueden actuar en dos subfamilias de canales de calcio: canales de calcio activados por ligando y en canales de calcio dependientes de voltaje. Hasta el momento la mayoría de las toxinas que actúan en canales de calcio identificadas en venenos de alacrán corresponden a dos tipos de toxinas. Estas se conocen como calcinas (E. F. Schwartz et al., 2009) y liotoxinas (Smith et al., 2011); ambas tienen interacción con los receptores de rianodina, que pertenecen a la sub-familia de canales de

calcio activadas por ligando. La interacción de las toxinas con el receptor que induce la aparición de un estado de sub-conductancia que incrementa la probabilidad de apertura del canal y la liberación de calcio (Xiao et al., 2016).

En el transcriptoma del alacrán *T. atrox* se identificaron 7 secuencias que tienen anotación de toxinas de calcio (figura 6), donde se pueden observar secuencias con identidad de calcinas (B), liotoxinas (C) y ω -calcitoxinas (D).

Toxinas de calcio

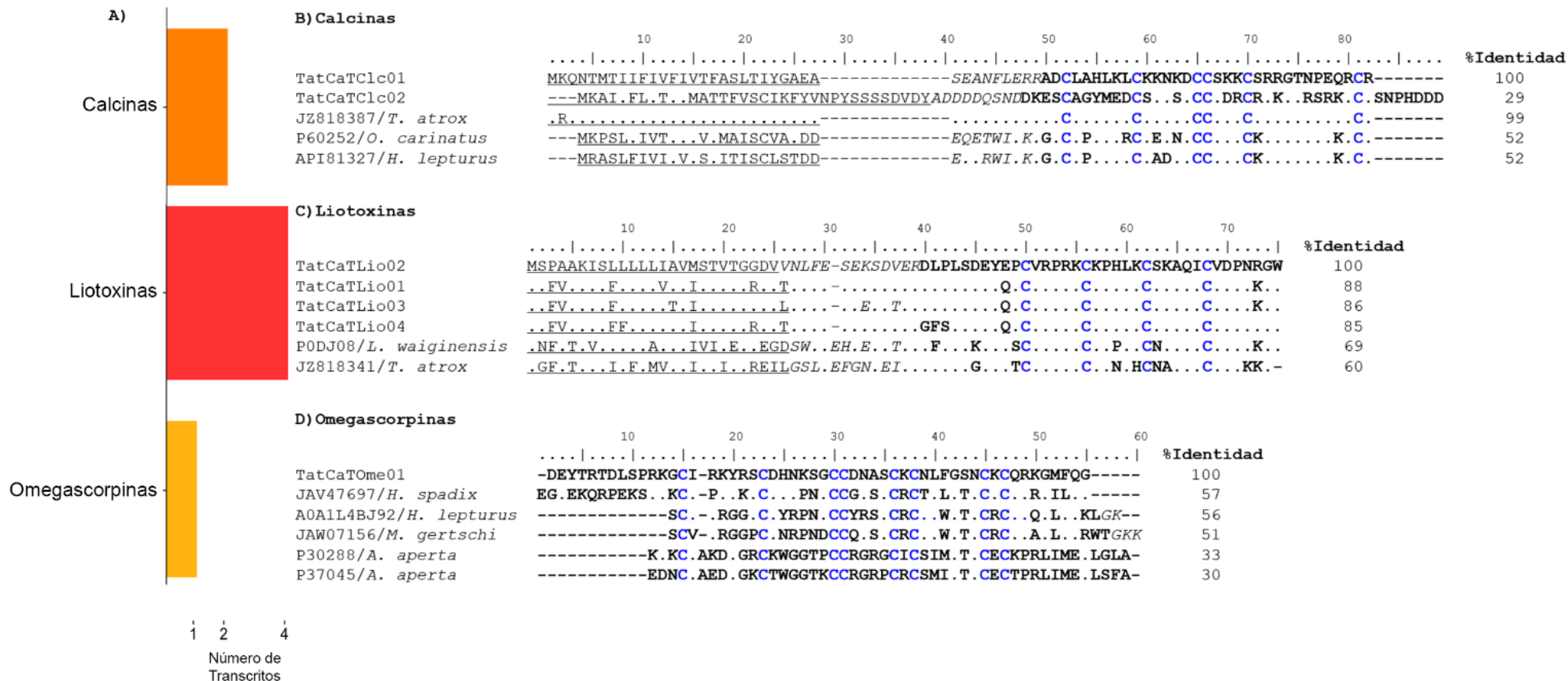


Figura 6. Toxinas putativas de calcio identificadas en el transcriptoma de *T. atrox*. A) Distribución de los transcritos de acuerdo con su sub-familia. B) Alineamiento de secuencias que tienen identidad con calcinas, incluyendo las referencias más cercanas de acuerdo con BLAST. C) Alineamiento de precursores que tienen identidad con las liotoxinas. D) Alineamiento de péptidos maduros de secuencias que tienen identidad con ω -agatoxinas de *A. aperta* y su referencia. El péptido señal está subrayado, el **péptido maduro** se muestra en letras negras, las cisteínas están marcadas con color azul y el *propéptido* en cursivas. El porcentaje de identidad fue calculado de acuerdo con la secuencia completa (incluyendo el péptido señal, el maduro y el propéptido).

Las calcinas son toxinas cortas de 33 a 35 aminoácidos. Estos péptidos básicos, capaces de traspasar la membrana celular, estructuralmente se caracterizan por presentar un dominio ICK. En el transcriptoma de *T. atrox* se identificaron dos secuencias que codifican putativamente para calcinas. Los precursores traducidos a aminoácidos se muestran en la figura 6. Para el alineamiento se utilizaron como secuencias de referencia aquellas que tuvieron el porcentaje de similitud más alto de acuerdo con el algoritmo de BLAST; una de ellas, la intrepicalcina (antes llamada ViCaTx1) (acceso GenBank JZ818387), identificada previamente en la librería de cDNA de *T. atrox*. Otra secuencia es la opicalcina-1 (acceso UniProt P60252) identificada en el alacrán *Opisthophthalmus carinatus* y por último la hemicalcina-1 (acceso UniProt API81327) identificada en el alacrán *Hemiscorpius lepturus*.

El segundo grupo de toxinas que actúan sobre canales de calcio son las liotoxinas. Estos péptidos tienen una longitud similar a las calcinas y están estabilizados por dos enlaces disulfuro. La primera liotoxina reportada se aisló del veneno del alacrán *Liocheles australasiae* (Matsushita, Miyashita, Sakai, Nakagawa, & Miyagawa, 2007). Más tarde se demostró que un péptido similar, Phi-LITX-Lw1a, identificado en el veneno de *Liocheles waigiensis*, actuaba sobre los canales de RyR con un mecanismo de acción similar al de las calcinas, pero con una potencia significativamente mayor. Se identificaron 4 transcritos con anotación putativa de liotoxinas. Los precursores traducidos de esos transcritos se muestran en la figura 6C. Como secuencias de referencia se utilizaron los precursores VmCaTx1 (acceso GenBank JZ818341) identificado en el cDNA del alacrán *Vaejovis mexicanus* y la Phi-LITX-Lw1a (acceso UniProt P0DJ08).

Existen toxinas que actúan sobre canales de calcio dependientes de voltaje; sin embargo, no se reportan con frecuencia. La kurtoxina, identificada en el alacrán *Parabuthus transvaalicus* y las toxinas kurtoxin-like I y II de *Parabuthus granulatus* son algunos representantes de esta familia. Estos péptidos están más relacionados con α -NaTxS en términos de secuencia que con otras toxinas de calcio; sin embargo se tienen reportes de que la kurtoxina modifica la apertura de los canales de calcio tipo T-, L-, N- y P/Q presentes en neuronas (Chuang, Jaffe, Cribbs, Perez-Reyes, & Swartz, 1998; Sidach & Mintz, 2002). Los péptidos kurtoxin I y II disminuyen la actividad en los canales de calcio tipo T en células espermatogénicas de ratón (López-González et al., 2003; Olamendi-Portugal T, García BI, López-González I, Van Der Walt J, Dyason K, Ulens C, Tytgat J, Felix R, Darszon A, 2002). En contraste, otros arácnidos como la araña *Agelenopsis aperta* tiene una batería de toxinas que actúan en canales de calcio dependientes de voltaje; estas toxinas, llamadas ω -agatoxinas, son antagonistas presinápticos de los canales de calcio dependientes de voltaje y tienen diferente especificidad hacia varios subtipos de estos canales en insectos y en mamíferos (Adams, 2004).

En el transcriptoma de *T. atrox* se identificó una secuencia con similitud a ω -agatoxinas. Este tipo de transcritos se han reportado en los transcriptomas de las glándulas venenosas de los alacranes *U. yaschenkoi* (Luna-Ramírez et al., 2015), *Megacormus gertschi* (acceso UniProt JAW07156) dos secuencias no publicadas de *Hadrurus spadix* (UniProt A0A1L4BJ92) y *Hemiscorpius lepturus* (GenBank JAV47697). La figura 6 muestra el alineamiento de los péptidos maduros predichos para dos secuencias identificadas en *A. aperta* anotadas como ω -agatoxinas (acceso UniProt P30288 y P37045 respectivamente); estas secuencias son las más cercanas en términos de secuencia y arreglo de cisteínas.

7.1.4.2 Péptidos de defensa del huésped (HDPs)

Los péptidos antimicrobianos están presentes en todos los seres vivos y son parte del sistema de defensa innato contra agentes externos (Hancock, Haney, & Gill, 2016). Algunos de esos péptidos, además de tener actividad antimicrobiana, pueden modular el sistema inmune (Cuperus, Coorens, van Dijk, & Haagsman, 2013; Steinstraesser, Kraneburg, Jacobsen, & Al-Benna, 2011). Los HDPs pueden ser divididos en aquellos que tienen cisteínas y forman enlaces disulfuro como las defensinas (Oppenheim, Biragyn, Kwak, & Yang, 2003) y los péptidos sin enlaces disulfuro (NDBPs).

En el transcriptoma de *T. atrox* se identificaron 17 transcritos que tuvieron identidad con HDPs, esto representa más del 10% de todas las secuencias anotadas relacionadas con péptidos de veneno; estos porcentajes coinciden con hallazgos previos en otras especies que no pertenecen a la familia Buthidae (figura 7).

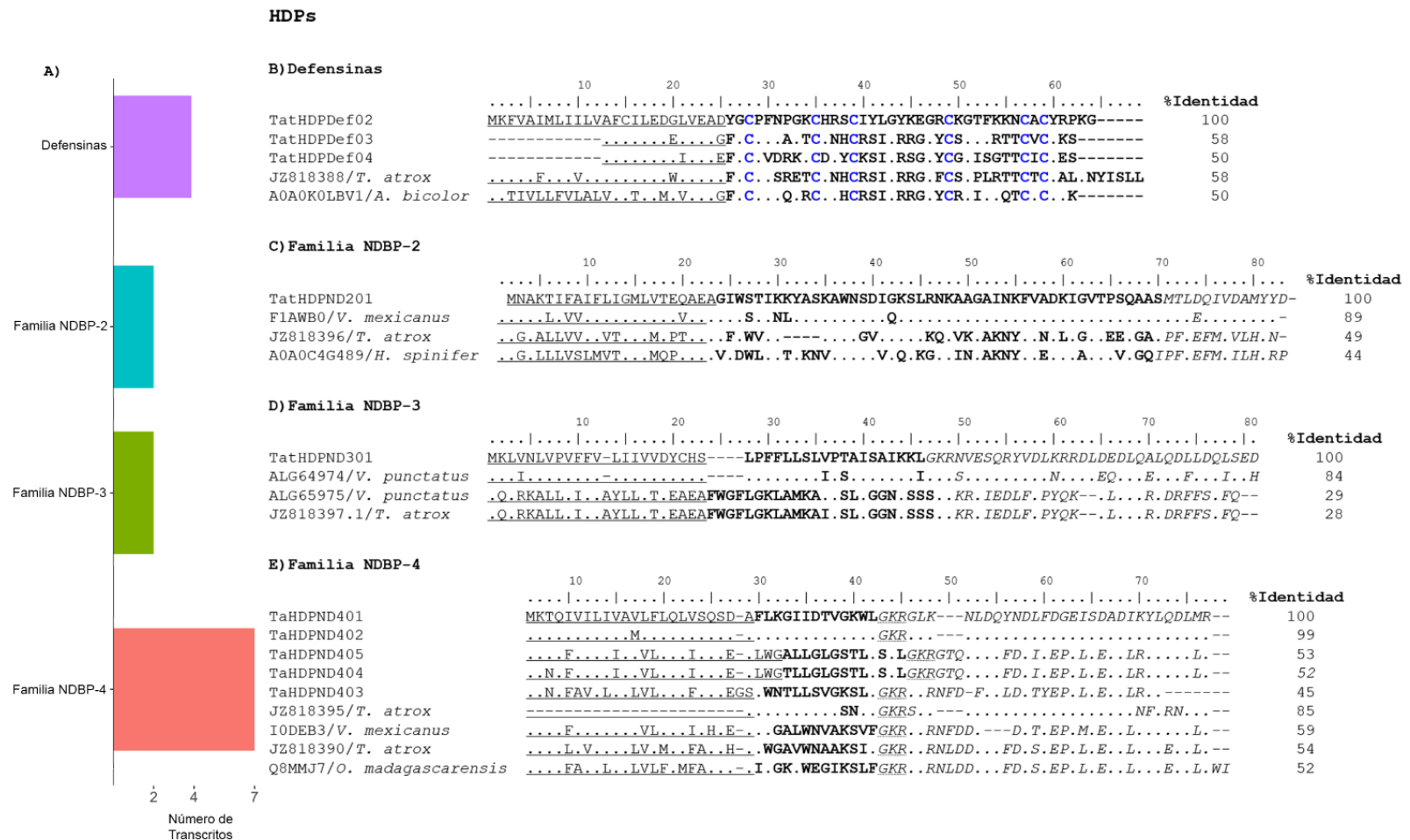


Figura 7 Posibles HDPs identificados en el análisis transcriptómico. A) Distribución de los transcritos encontrados clasificados por familias. B) Alineamiento de precursores con identidad de β -defensinas contra secuencias de referencia. C), D), E) Alineamiento de precursores con identidad de NDBPs de las familias 2, 3 y 4 con sus respectivas secuencias de referencia. El péptido señal está subrayado, el **péptido maduro** se muestra en letras negras, las cisteínas están marcadas con color azul y el *propéptido* en cursivas. El porcentaje de identidad fue calculado de acuerdo con la secuencia completa (incluyendo el péptido señal, el maduro y el propéptido).

Las defensinas se han identificado en tres de los cuatro reinos de las eucariotas: Animalia, Plantae y Fungi (Wilmes & Sahl, 2014). Se han identificado dos sub-familias: α y β -defensinas y esta última es la más distribuida. Estos péptidos catiónicos son pequeños (2-6 kDa), tienen una estructura CS $\alpha\beta$ estabilizada por 3 enlaces disulfuro y son activos contra bacterias, hongos y virus.

En los venenos de alacrán las defensinas se encuentran con regularidad, incluso se ha propuesto la vinculación de estos péptidos con las neurotoxinas en el sentido funcional y evolutivo (Meng et al., 2016). Se identificaron 4 transcritos que tienen anotación de β -defensinas, de los cuales se identificaron 3 precursores y una secuencia parcial. Los primeros 3 son los que se utilizaron para el alineamiento (figura 7A). Como secuencias de referencia se usaron aquellos precursores más cercanos identificados por el algoritmo de BLAST; la primera secuencia, ViDef, una β -defensina identificada en el cDNA de la especie de *T. atrox* (GenBank JZ818388), previamente reportada por Quintero y cols en 2014 y la AbDef-1 (acceso UniProt A0A0K0LBV1) de *Androctonus bicolor*.

Otros miembros de la familia de HDP son los péptidos sin enlaces disulfuro (NDBPs). Los NDBPs normalmente son de tamaño pequeño, catiónicos y ricos en residuos hidrofóbicos. Estas características dan como resultado péptidos con una estructura tridimensional desordenada en soluciones acuosas y de α -hélice en el entorno de la membrana celular (Brogden, 2005). Con respecto a la composición del precursor, los NDBPs tienen un péptido señal, péptido maduro que posteriormente termina con un propéptido carboxi-terminal que consta de una glicina y dos residuos básicos (normalmente lisina (K) y arginina (R), eso permite que el péptido maduro que se expresa se encuentre amidado en el carboxilo terminal.

En los últimos años, los NDBPs han captado la atención debido a la gran cantidad de actividades que presentan estos péptidos. Se ha descrito que los NDBPs tienen actividad antimicrobiana, citolítica, inmunomoduladora, potenciadora de bradiquinina y algunos autores demuestran actividad antitumoral, por lo cual se han propuesto como buenos candidatos para desarrollo de fármacos (Ortiz, Gurrola, Schwartz, & Possani, 2015). La clasificación de los NDBPs basados sólo en la secuencia de aminoácidos no ha dado resultado ya que son muy variables; sin embargo, Almaaytah y Albalas en 2014 propusieron la clasificación de los péptidos sin enlaces disulfuro en 5 familias (NDBP-1, NDBP-2, NDBP-3, NDBP-4 y NDBP-5), basados en la identidad de secuencia, longitud y actividad farmacológica (Almaaytah & Albalas, 2014).

En el transcriptoma de la glándula venenosa del alacrán *T. atrox* se identificaron 11 transcritos que tuvieron identidad de NDBPs, de los cuales 2 secuencias tuvieron identidad con miembros de la

familia NDBP2, 2 con la familia NDBP-3 y 7 con la familia NDBP-4.

La familia NDBP-2 está compuesta por péptidos multifuncionales largos que oscilan entre 40 y 60 residuos, además de ser ricos en residuos básicos como lisina y arginina. Se consideran multifuncionales porque se ha reportado que tienen actividad antimicrobiana, potenciadora de bradicinina, insecticida y antitumoral (Almaaytah & Albalas, 2014).

Se identificaron 2 transcritos que codifican putativamente para esta familia, uno de ellos se identificó previamente en la librería de cDNA; la secuencia ViVlp1 (acceso GenBank JZ818396) se utilizó como referencia para el alineamiento mostrado en la figura 7C. Las otras dos referencias son la vejovina (acceso UniProt F1AWB0) identificada en el alacrán *Vaejovis mexicanus*, efectiva contra bacterias multirresistentes Gram (-) (Hernández-Aponte et al., 2011) y la heterin-1 (acceso UniProt A0A0C4G489) identificada en el alacrán *Heterometrus spinifer*, efectiva contra bacterias Gram (+) y Gram (-) (Wu et al., 2014).

Los miembros de la familia NDBP-3 tienen longitud media entre 20 y 30 aminoácidos. Se identificaron dos transcritos con identidad de NDBP-3; uno de ellos descrito previamente en la librería de cDNA previamente reportada para *T. atrox* ViAMP1 (acceso GenBank JZ818397.1). El otro miembro de esta clasificación se alineó contra ViAMP1, VpAMP1.0 (acceso UniProt ALG64974) y VpAMP2.0 (UniProt ALG64975), estos dos péptidos se identificaron en la librería de cDNA de *Vaejovis punctatus* (Verónica Quintero-Hernández et al., 2015) (Figura 7D).

Los péptidos que pertenecen a la familia NDBP-4 tienen una longitud corta, son catiónicos y anfipáticos. Se identificaron 7 secuencias que tienen identidad con esta familia, dos de ellas previamente identificadas, ViCT2 (GenBank JZ818390) y la segunda ViCT7 (GenBank JZ818395). Los precursores de las 5 secuencias restantes se muestran en la figura 7. Para el alineamiento se utilizaron las secuencias VmCT1 (acceso UniProt I0DEB3) e IsCT (UniProt Q8MMJ7) de *V. mexicanus* y *Opisthantus madagascarensis*, respectivamente, las dos secuencias más cercanas según el algoritmo de BLAST. Cabe destacar que las secuencias TatHDPND403 y TatHDPND404 comparten el mismo péptido maduro y difieren en un aminoácido en el péptido señal.

En la tabla 4 se muestran las características fisicoquímicas más relevantes de las secuencias maduras identificadas en el transcriptoma de *T. atrox* como hidrofobicidad, momento hidrofóbico y carga. Los péptidos maduros de los miembros de la familia NDBP-2 son los más cargados y los menos hidrofóbicos y los miembros de la familia NDBP-4 son los más hidrofóbicos y los menos cargados

positivamente.

Tabla 4. Parámetros fisicoquímicos predichos para los péptidos maduros de los transcritos anotados como NDBPs por el software HeliQuest (<http://heliquest.ipmc.cnrs.fr/cgi-bin/ComputParams.py>).

ID	Familia NDBP	Longitud del péptido maduro	Hidrofobicidad	Momento hidrofóbico	Carga
ViVlp1	NDBP-2	56	0.31	0.176	+3
TatHDPND201	NDBP-2	60	0.326	0.1	+5
TatHDPND301	NDBP-3	19	0.903	0.377	+2
ViAMP1	NDBP-3	24	0.557	0.299	+3
ViCT2	NDBP-4	13	0.793	0.595	+1
ViCT7	NDBP-4	13	0.778	0.769	0
TatHDPND401	NDBP-4	13	0.752	0.792	+1
TatHDPND403	NDBP-4	13	0.819	0.606	+1
TatHDPND404	NDBP-4	13	0.742	0.458	+1
TatHDPND405	NDBP-4	13	0.746	0.456	+1

Además de las defensinas y los NDBPs, se identificaron otros péptidos de defensa del huésped. Se identificó un transcrito TatHDPAni01, que codifica para un péptido aniónico; su coincidencia más cercana en términos de secuencia fue la Hta1 (con 55 % de identidad), identificada en el transcriptoma del alacrán *Hadogenes troglodytes*. Los péptidos aniónicos de este tipo se han identificado previamente en otros alacranes (De Oliveira et al., 2015; Melo et al., 2017) y otros phyla. Son parte integral del sistema de defensa de vertebrados, invertebrados y plantas (Harris, Dennison, & Phoenix, 2009).

El segundo transcrito identificado, TatHDPWap01, cuya secuencia madura predicha comparte un 64% de identidad con una waprin-Enh1-like identificada en el análisis transcriptómico de la araña *Parasteatoda tepidariorum* (GenBank XP_015928629.1). Las waprinas tienen una longitud de 50 aminoácidos y se encuentran frecuentemente en venenos de serpiente (Liu et al., 2013). Son homólogos estructurales de la familia WAP (whey acid protein por sus siglas en inglés), con un arreglo conservado de 4 enlaces disulfuro. Se ha visto que estas proteínas tienen actividad inhibidora de proteasas, antimicrobianas y forman parte del sistema inmune innato (Hagiwara et al., 2003). No se tienen reportes previos de esta familia de proteínas en alacranes.

7.1.4.3 Enzimas

Se sabe que en el veneno de alacranes existen diferentes enzimas que juegan un papel importante en la toxicidad y sirven también para que el veneno se difunda en el sistema de la presa (Fletcher, Fletcher, Weninger, Anderson, & Martin, 2010). Entre estas enzimas encontramos tres tipos principalmente:

fosfolipasas, hialuronidasas, y proteasas. En el transcriptoma de *T. atrox* se identificaron 55 transcritos que tuvieron identidad con enzimas; esto representa alrededor de un tercio de todos los transcritos anotados. La proporción observada ya se había reportado en otros transcriptomas de alacranes (Almeida et al., 2012; Santibáñez-López et al., 2017).

Existen diferentes tipos de fosfolipasas (A2, B, D) que se encuentran presentes en los venenos de alacranes, arañas y serpientes. Se sabe que estas enzimas pueden degradar la membrana celular provocando daño en los tejidos circundantes al sitio de la inyección del veneno (Vines & Bill, 2015). En este trabajo los transcritos más diversos fueron los que tuvieron identidad con fosfolipasas (21 transcritos). Sólo se identificaron 4 precursores, 3 pertenecientes a secuencias anotadas como fosfolipasas A2 y una secuencia identificada como fosfolipasa B. En los alacranes, las fosfolipasas más representadas son del tipo A2, sin embargo, en este trabajo también se identificaron transcritos que codifican para los tipos B y D.

Las hialuronidasas hidrolizan el ácido hialurónico, este lo podemos encontrar en la matriz extracelular de los tejidos. Esta enzima facilita la difusión del veneno del sitio de inyección hacia los capilares. Existen evidencias de la presencia de hialuronidasas en el veneno de los alacranes y de otros animales venenosos (Bordon, Wiesel, Amorim, & Arantes, 2015). En este trabajo identificamos solo dos transcritos que codifican para hialuronidasas.

Las proteasas son enzimas presentes en los venenos de diferentes animales. Las proteasas más representadas en los venenos de alacranes son las metaloproteasas (Fletcher et al., 2010) y las serin proteasas (Cera, 2009), estas favorecen la difusión del veneno en el tejido al ser capaces de hidrolizar proteínas de la matriz extracelular. En este trabajo se identificaron 18 secuencias que codifican para serin proteasas y 14 transcritos con identidad a metaloproteasas.

Cabe mencionar que la mayoría de los transcritos identificados para enzimas son parciales, las únicas que se identificaron precursores completos son algunas fosfolipasas.

7.1.4.4 Inhibidores de proteasas

Las proteasas que son secretadas pueden causar daño celular si no están reguladas (Armstrong, 2001), por lo tanto, también se espera que los venenos de alacrán que son ricos en proteasas contengan inhibidores de proteasas. Esto se refleja en el análisis transcriptómico ya que se identificaron 24 secuencias codifican putativamente para inhibidores de proteasas. A pesar de que sólo se identificaron secuencias parciales, se pudo recuperar el dominio PF0014 que corresponde a serpinas o inhibidores de

tripsina tipo Kunitz, ambos inhibidores de proteasas. La mayoría de los transcritos tuvieron identidad a inhibidores de serin proteasa tipo serpina (19 secuencias) y las 5 secuencias restantes tuvieron identidad con inhibidores tipo Kunitz.

7.1.4.5 Otros componentes de veneno

Este grupo incluye transcritos anotados como péptidos de veneno, pero la función o el blanco molecular no se ha demostrado experimentalmente.

Péptidos La1-like

Los péptidos La1 definen una familia caracterizada de manera estructural por el dominio llamado factor Von Willebrand tipo C (SVWC) con 4 enlaces disulfuro. Uno de los primeros reportes fue el hallazgo de esta proteína en el veneno del alacrán *Liocheles australasiae* el cual se reportó como el péptido más abundante (Miyashita, Otsuki, Hanai, Nakagawa, & Miyagawa, 2007). Estos péptidos se encuentran prácticamente en todos los géneros de alacranes, sin embargo, su función o su blanco molecular aún se desconoce.

Existen algunos reportes que mencionan la posible función de los péptidos La1. Por ejemplo la spermaurina, un péptido tipo La1 identificado en el veneno del alacrán *Scorpio maurus palmatus* que incrementa la motilidad de los espermatozoides en mamíferos (Martinez et al., 2017). Se identificaron 7 transcritos con identidad de La1, de los cuales se recuperaron 6 precursores y una secuencia parcial.

Para el alineamiento de las secuencias identificadas como La1 se utilizaron como referencias la secuencia previamente identificada en el cDNA del alacrán *T. atrox* ViLa1lp1 (GenBank JZ818417), el péptido La1 original (UniProt P0C5F3) identificado en el alacrán *Liocheles australasiae* y por último HtLa1 (UniProt A0F40202) identificado en el veneno del alacrán *Hadogenes troglodytes* (Figura 8).

Péptidos La1-like

	10	20	30	40	50	60	70	80	%Identidad			
TaLa1lp004	YIRTC	PIG-DKDISNGE	EWADQKRCV	KYKCV	QVRGPD	AALLITR	CPSVGV	YPTDKCRE	VPGK--GDFPTCCPKLQCD-----	100		
TaLa1lp005	...C...	...N...	...C...	...C...	...S...	...C...	...C...	...SCC...	...CD-----	95		
TaLa1lp001	VGEICQV	-SMS.NV.KKM	Q.P.SC.I.EC	VEQNYRIL	.SKM	SCSPQV--	.KRG	C.N...PVD	AP.DCC.TSL	CRGKQWDE	31	
TaLa1lp002	FGE.CQA	.GKYNVRV	.QPIQ.PNS	C.L..CLN	YNRRYV	.QTLSC	ATQT--L	KSGC.F	AAAAPGTP	.NCC.MVI	CKGSG---	28
A0F40202/H. troglodytes	LGE.CKV	-QFT.PV.QQ	AN.LTAC.Q	.CLN	YNRRYA	.ETLSC	AKQT--L	KSGCHNTL	.AAGTP	.NCC.TVI	CRGG----	26
P0C5F3/L. australasiae	FGESCIA	.RFIVPL	.QQVT.RD	CAL..CV	YNKKF	.ETK.C	ATVN--L	KSGCKT	...GAGAA	.SCC.MV	TCK-----	25
JZ818417/T. atrox	VGEICQV	-SMS.NV.KKM	Q.P.SC.I.EC	VEQNYRIL	.SKM	SCSPQV--	.KRG	CRN...PVD	AP.DCC.TSL	CRGKQWDE	25	

Figura 8. Alineamiento de secuencias que tuvieron identidad con péptidos tipo La1. Sólo los péptidos maduros fueron utilizados para el alineamiento. El **péptido maduro** se muestra en letras negras, las cisteínas están marcadas con color azul.

Familia CRISP

Las proteínas CRISP (Cysteine-Rich Secretary Proteins por sus siglas en inglés) forman parte de la superfamilia CAP que incluye familias como la de Antígeno-5 y otras proteínas relacionadas con patogenicidad tipo 1 (CAP por sus siglas en inglés). Se han identificado miembros de esta superfamilia en muchos reinos animales, incluidos aquellos donde se encuentran los animales venenosos. Estas proteínas se asocian con numerosas funciones endócrinas y parácrinas (Gibbs, Roelants, & O'Bryan, 2008).

Existen reportes que señalan la presencia de bajas proporciones de proteínas CRISP en venenos de alacrán (Santibáñez-López et al., 2017) y se han aislado con éxito algunas proteínas CRISP en el veneno de serpientes (Adade et al., 2014) y en el veneno de abejas como el alérgeno-5. En el transcriptoma del alacrán *T. atrox* sólo se identificaron transcritos con posible identidad de proteínas tipo CRISP, siendo 8 los transcritos parciales anotados.

A pesar de que se han identificado estos componentes en el veneno de diferentes animales, no se le ha descrito una función asociada.

Otros componentes de veneno con función desconocida

Es común identificar transcritos con anotación de componentes de veneno sin función conocida. Se anotan como tal porque no tienen un dominio asociado que corresponda a alguna familia de proteínas u otra información de actividad o blanco molecular; sin embargo, estos péptidos se han identificado en el veneno o los transcritos en las glándulas venenosas y, en el caso de los transcritos, ha identificado el CDS (Coding DNA Sequence por sus siglas en inglés). En el caso del transcriptoma de *T. atrox*, se identificaron 8 secuencias con estas características que se agruparon en la categoría “desconocida” (Und).

7.1.5 Cuantificación del transcriptoma de *T. atrox*

Conocer la diversidad de componentes de interés en un transcriptoma es un paso esencial en la caracterización del RNA-Seq. La cuantificación del transcriptoma nos permite estimar la abundancia relativa de los transcritos para comprender mejor la biología del animal estudiado. Ya que no se tiene un genoma de referencia para mapear y cuantificar los *reads*, sólo se pueden saber los niveles de expresión de los transcritos, no de los genes.

En la literatura existen diferentes herramientas de cuantificación de RNA-Seq, una de ellas es Salmon (documentación disponible en <https://combine-lab.github.io/salmon/>), que sirve para cuantificar la expresión de los transcritos ensamblados utilizando los datos de RNA-Seq. Una de las ventajas de

utilizar este algoritmo es el bajo poder de cómputo que se requiere gracias a un *quasi*-mapeo de dos fases que consiste en indizar y cuantificar (Patro, Duggal, Love, Irizarry, & Kingsford, 2017); otra característica importante es que evita el sesgo de estimación de niveles de expresión por tamaño de transcrito, es decir, Salmon considera los transcritos por millón por longitud efectiva de transcrito (TPM) en vez de sólo hacer conteo de *reads* por transcrito. Utilizar esta herramienta nos ayuda a evitar sesgos por la probabilidad de mapear una lectura en un cierto gen cuando este sea más largo o más corto.

No existen reportes de cuantificación de transcritos por el algoritmo salmon a partir de RNA-Seq de glándulas venenosas, sin embargo, existe un trabajo de cuantificación de RNA-Seq de la glándula venenosa del alacrán *Centruroides hentzi*, ellos utilizaron Bowtie2 para inferir la cantidad de *reads* existentes por transcrito, pero no consideran la longitud de los transcritos en esa cuantificación. Por lo tanto, el usar este algoritmo nos da un acercamiento más real de los niveles de expresión.

En la tabla 5 se muestran los 10 transcritos más representados en función del valor de TPM. Los transcritos relacionados con veneno que tienen los niveles de expresión más altos pertenecen a la familia de HDPs con representantes de las sub-familias NDBP-4, NDBP-3, NDBP-2 y péptidos aniónicos; también se identificaron 2 escorpinas, clasificadas como toxinas de potasio y miembros de la familia de DBPs.

Tabla 5. Resumen de los 10 transcritos con mayor nivel de expresión. La posición en el transcriptoma se designó de acuerdo con el lugar del transcrito definido por el valor de TPM, el ID es el identificador asignado de acuerdo a la anotación, el TPM es el valor de transcritos por millón determinado por el algoritmo de Salmon y, por último, la identidad es la anotación definida por el algoritmo de BLAST.

ID	Posición en el transcriptoma	Longitud (nt)	Longitud efectiva	TPM	Número de <i>reads</i>
ViCT7	8	275	132.36	14581.9	149773
TatHDPND401	11	470	327.195	7356.36	186781
TatHDPND402	12	598	455.163	6955.35	245669
TatHDPND403	13	495	352.189	6367.39	174021
TatHDPND201	14	487	344.191	6136.84	163911
ViScplp2	15	728	585.14	6122.72	278015
TatKTxAlp07	16	546	403.18	5818.71	182050
ViVlp1	18	607	464.161	5438.79	195900
ViCT2	21	1791	1648.07	3815.24	487936
TatHDPAni01	22	1209	1066.07	3731.18	308672

Los transcritos ViCT7, TaHDPND401 y 402 tienen identidad con el péptido maduro de la secuencia VsCT1 (ViCT7 tiene 85% mientras que TaHDPND401 y 402 tienen un 100% de identidad).

Esta secuencia se identificó en una librería de cDNA de la glándula venenosa del alacrán *Vaejovis subcristatus*.

El transcrito ViCT2 se identificó previamente en la librería de cDNA del alacrán *T. atrox*. El péptido maduro de esta secuencia tiene un 62% de identidad con VmCT1, previamente identificado en una librería de cDNA de la glándula venenosa del alacrán *Vaejovis mexicanus*. La actividad hemolítica de este péptido es de 10% de hemólisis a una concentración máxima de 50 μ M. A diferencia del péptido VsCT1, VmCT1 tuvo actividad antimicrobiana contra Gram (+) y Gram (-) a concentraciones desde 5 hasta 25 μ M.

En resumen, la figura 9 muestra la distribución de los niveles de expresión de los transcritos relacionados con proteínas de veneno contra su longitud efectiva. De manera general se puede observar que los HDPs son el grupo de proteínas más representado seguido de los DBPs, específicamente las toxinas de potasio. En relación con las enzimas, se observa que las fosfolipasas y algunas hialuronidasas son las secuencias que tienen niveles de expresión relativamente altos comparados con las serin proteasas y las metaloproteasas.

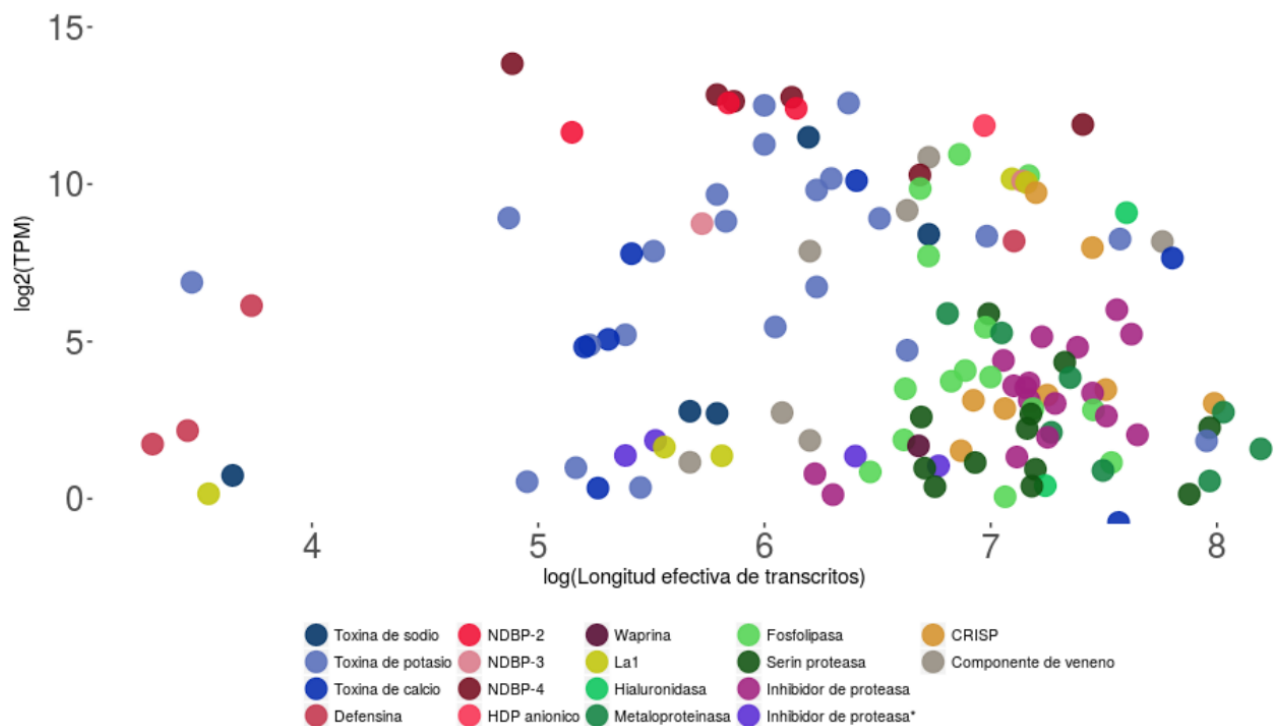


Figura 9. Distribución de las abundancias de los transcritos identificados como componentes de veneno contra el logaritmo de la longitud efectiva. Los miembros de la familia de DBPs están resaltados en colores de la gama azul, los HDPs están resaltados con colores de la gama rosa, los La1 están resaltados en amarillo, las enzimas están resaltadas

en colores de la gama del verde, los inhibidores de proteasas están resaltados en colores de la gama del morado, los miembros de la familia CRISP están resaltados en naranja y por último, los transcritos identificados como componentes de veneno están resaltados en gris.

7.2 Proteómica del veneno del alacrán *T. atrox*.

7.2.1 Fingerprint del veneno

Se utilizaron 5 µg de veneno para identificar las masas moleculares de los componentes del veneno de *T. atrox*. Para esto se usó la técnica de cromatografía líquida acoplada a un espectrómetro de masas (LC-MS) por medio de la cual se lograron identificar 135 masas moleculares (tabla 6) que oscilaron entre los 1,077 y 16,920 Da con una mediana de 4,506 Da.

Tabla 6. Fingerprint de las fracciones de veneno soluble de *T. atrox*. La distribución de los componentes de veneno por LC-MS fue reportada en intervalos de 20 minutos.

TR ¹ (min)	PM ² (Da)	TR (min)	PM (Da)
1-20	1462.7, 2057.24, 2117.68, 2265.06, 2796.27, 3111.96, 9115.86, 10663.93, 11123.28,	140-160	1944.15, 2645.50, 2815.60, 6330.03, 6473.90, 6714.40, 7438.62, 7639.27, 7843.00, 8049.12, 8213.16, 8829.81, 8950.11, 9535.2
20-40	1076.62, 1205.68, 1212.80, 1673.85, 1817.88, 3427.38, 3499.92, 3586.92, 3878.10, 4197.53, 12306.36	160-180	1337.72, 1497.81, 2193.06, 2248.28, 2347.32, 3338.30, 7040.46, 7956.10, 8201.97, 8727.14
40-60	1331.64, 1799.04, 1886.82, 2333.32, 2411.36, 2447.40, 2592.26, 3777.63, 3945.62, 5813.52	180-200	1296.10, 2151.20, 4171.38, 4302.42, 4389.42, 4697.56, 4762.08, 6195.66
60-80	2377.16, 2850.1, 2944.70, 3606.60, 4485.10, 4595.04, 5279.52, 5654.40	200-220	10039.5, 13729.41, 14079.03
80-100	3332.90, 3535.47, 3718.60, 3787.85, 4113.96, 4125.80, 4204.00, 4279.05, 4290.36, 5196.42, 5756.56, 7011.33, 7123.96, 7236.99, 8126.40, 8328.51	220-240	1828.00, 6554.31, 6750.45, 6946.57, 7269.84, 8272.50, 10545.20, 12430.9, 13591.92, 13815.51, 14614.72
100-120	3223.80, 3243.80, 3569.92, 3767.15, 4250.67, 8468.54, 8581.60, 8716.70, 9056.88, 9490.25	240-260	3821.44, 5409.48, 10882.9, 16915.41

120-140	1198.64, 1648.86, 3267.39, 4036.16, 4348.40, 4561.84, 4815. 2	260-290	2038.11, 3347.5, 4505.55, 4791.65, 4949.7, 8355.48, 11174.46, 11833.92, 11847.44, 11899.27, 13891.59, 14257.71, 14705.56, 14741.70
----------------	---	----------------	---

¹ TR (tiempo de retención en minutos)

² PM (masa molecular en Daltones).

En la figura 10 se observa que la mayoría de los componentes identificados rondan entre 1,000 Da y 5,000 Da; en este rango se pueden encontrar péptidos con posible actividad antimicrobiana y toxinas que afectan los canales de potasio y calcio, entre otros. Cabe mencionar que no se tienen reportes previos de análisis proteómico del veneno de algún miembro de la familia Vaejovidae.

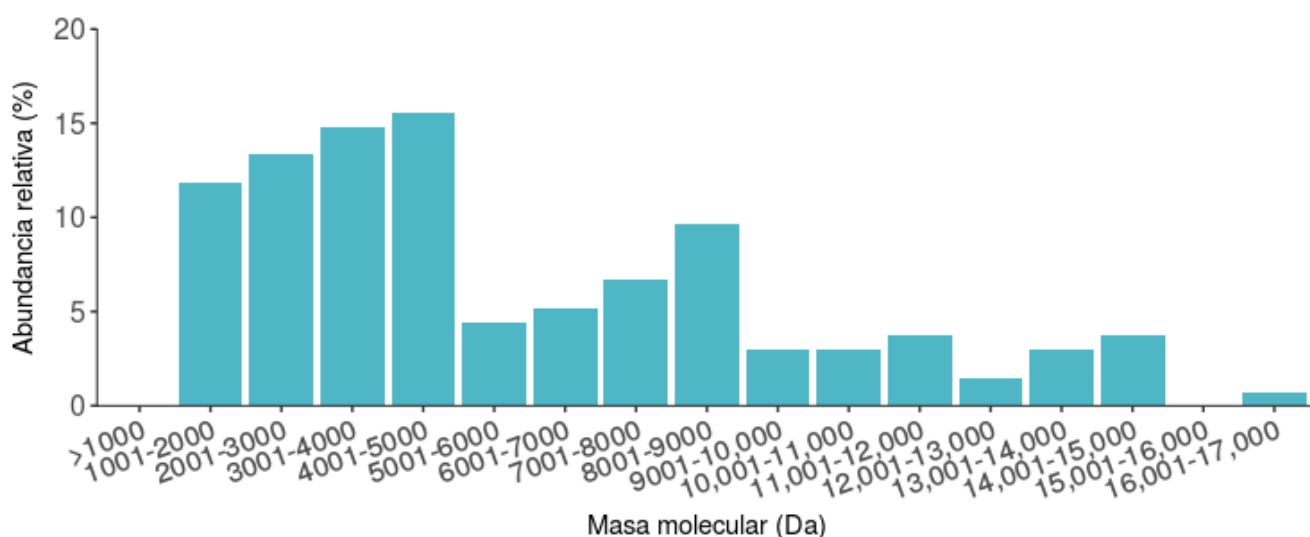


Figura 10. Distribución relativa de las masas moleculares identificadas en el veneno del alacrán *T. atrox*.

Los péptidos que se encuentran en el rango de 1,000 a 5,000 Da son los más abundantes, cubriendo más del 50% de los componentes identificados en el *fingerprint*.

Se utilizó como base de datos todos las masas moleculares identificados en el *fingerprint* para buscar transcritos del transcriptoma de *T. atrox* que tuvieran la misma masa molecular, para esto se utilizó la secuencia madura de cada uno de los transcritos relacionados con péptidos de veneno y se hicieron cálculos para determinar la masa molecular dependiendo de la modificación (oxidación de cisteínas para el caso de los péptidos con enlaces disulfuro y para el caso de los NDBPs las modificaciones post-traduccionales como amidación del carboxilo terminal), esto permitió encontrar 8 coincidencias posibles para: β -toxinas de sodio (TatNaTBet03 y TatNaTBet08), α -toxinas de potasio (TatKTxAlp10 y TatKTxAlp12) y calcinas (TatCaTClc01) (Tabla 7).

Tabla 7. Masas moleculares identificadas en el transcriptoma de *T. atrox*.

<i>Toxinas de sodio</i>			
Identificador transcriptoma	Masa teórica	Masa experimental	Rango de TR
TatNaTBet03	5196.8	5196.4	100-120
TatNaTBet08	6195.9	6195.7	180-200
<i>Toxinas de potasio</i>			
TatKTxAlp10	3607.4	3606.6	80-100
TatKTxAlp12	4114.9	4113.9	100-120
<i>Toxinas de calcio</i>			
TatCaTClc01	3788.5	3787.9	100-120

7.2.2 Secuenciación *de novo* por LC MS/MS

Producto de la digestión del veneno de *T. atrox* y posterior paso por LC MS/MS se obtuvieron más de 10,000 espectros cuyas secuencias de aminoácidos se ensamblaron para identificar secuencias de proteínas presentes en el veneno utilizando como base de datos todos los péptidos maduros predichos como componentes de veneno en el transcriptoma (anexo 7).

La secuenciación *de novo* por LC-MS/MS permitió identificar fragmentos de 42 proteínas (Tabla 8) codificadas por transcritos ensamblados distribuidos de la siguiente manera: proteínas con identidad a componentes celulares (15 proteínas); enzimas (7) que incluyen fosfolipasas, hialuronidasas y metaloproteinasas; HDPs (7) como miembros de la familia NDBP-2, -3 y -4; DBPs (3) que incluye miembros de la familia de toxinas de potasio (como escorpinas y k-KTx) y toxinas de calcio tipo calcinas; péptidos tipo La1 (2); miembros de la familia CRISP (2); y proteínas sin anotación en las bases de datos.

Tabla 8. Secuencias aminoacídicas identificadas por LC MS/MS utilizando el transcriptoma de *T. atrox* como base de datos para la identificación de proteínas por el algoritmo de Sequest.

Identificador del transcriptoma	Score	Cobertura	Tipo de proteína	Número de acceso de la proteína de referencia
comp8310_c0_seq1	46.06	19.1 %	Alostatina	XP_013775495
comp32030_c1_seq1	28.07	34.8 %	Enzima convertidora de angiotensina	XP_013773749
comp32030_c2_seq1	32.73	7.7 %	Enzima convertidora de angiotensina	XP_013773749
comp33161_c0_seq1	535.88	24.5 %	Enzima convertidora de angiotensina	XP_013773749
comp33725_c0_seq1	65.33	16.8 %	Enzima convertidora de angiotensina	XP_013773749
comp33936_c0_seq1	64.74	13.1 %	Enzima convertidora de angiotensina	XP_013773749

TatCaTClc01	88.83	24.2 %	Toxina de calcio. Calcin	A0A1L4BJ42
comp32319_c0_seq1	18.83	7.56 %	Ectonucleótido trifosfato difosfohidrolasa tipo 2	XP_013778001
comp881_c0_seq1	452.55	18.7 %	Proteína tipo elastasa	CAX51421
TatHDPND201	513.12	46.7 %	HDP. Familia NDBP-2	F1AWB0
TatHDPND301	22.46	94.7 %	HDP. Familia NDBP-3	ALG64974
ViVlp1	762.84	28.6 %	HDP. Familia NDBP-2	AGK88593
ViAMP1	188.70	70.8 %	HDP. Familia NDBP-3	ALG64975
TaHDPND401	254.77	61.5 %	HDP. Familia NDBP-4	I0DEB5
TatHDPND403	37.65	100 %	HDP. Familia NDBP-4	I0DEB5
ViCT2	882.53	76.9 %	HDP. Familia NDBP-4	I0DEB3
TatEnzHya01	161.40	39.1 %	Hialuronidasa	API81375
comp15335_c0_seq1	92.74	10.3 %	Proteína hipotética	CAX51393
comp30560_c0_seq1	67.45	9.1 %	Proteína hipotética	AEX09195
comp31101_c0_seq1	103.49	29.1 %	Proteína hipotética (tipo alérgeno)	CAX51409
comp30730_c0_seq1	16.64	6.5 %	Proteína hipotética RvY_03950	GAU91754
ViLa1lp1	47.91	63.3 %	Tipo La1	AOF40216
TatOthLa101	469.44	45.5 %	Tipo La1	AOF40202
comp34524_c0_seq1	74.23	11.5 %	Metaloproteinasa	XP_009865190
TatEnzMtp04	31.25	18.8 %	Metaloproteinasa	AMO02513
comp32637_c0_seq1	828.12	42 %	Nucleotidasa	XP_013774694
comp26928_c1_seq1	214.51	23.2 %	Otros componentes de veneno	N/A
comp27809_c1_seq1	1255.98	24.4 %	Otros componentes de veneno	N/A
comp30392_c0_seq1	34.49	32.9 %	Otros componentes de veneno	CAX51433
comp32982_c0_seq3	20.83	13.7 %	Otros componentes de veneno	N/A
comp43100_c0_seq1	70.29	15.3 %	Otros componentes de veneno	N/A
comp31198_c0_seq1	20.19	3.13%	Otros componentes de veneno	N/A
TatEnzPA201	1616.89	45.5 %	Fosfolipasa A2	API81339
TatEnzPA213	253.38	27 %	Fosfolipasa A2	API81335
TatEnzPA215	877.33	50.2 %	Fosfolipasa A2	API81335
TatEnzPA202	94.96	31.3 %	Phospholipase A2	API81335
comp20627_c0_seq1	14.64	0.9 %	Protein cinasa C	XP_022243213
TatOthCRI06	24.43	15.9 %	CRISP	JAT91149
TatOthCRI07	10.69	23.1 %	CRISP	API81352
comp30427_c0_seq1	16.87	2.27 %	Esteril sulfatasa	XP_0193859

7.3 Síntesis y caracterización biológica del péptido ViCT2 secuenciado en el transcriptoma y el proteoma del alacrán *T. atrox*

Se optó por sintetizar químicamente y evaluar la actividad biológica del péptido ViCT2 por las siguientes razones: este transcrito se identificó en la librería de cDNA, en el transcriptoma con un 100 %

de identidad, también se identificó la secuencia madura en la secuenciación *de novo* con altos valores de confiabilidad, en lo que se refiere a las propiedades fisicoquímicas, este péptido tiene carga +1 y una hidrofobicidad de 0.79, además de tener alta identidad con la secuencia previamente reportada VmCT1 que es activa contra bacterias Gram (+) y Gram (-). En la figura 11 se observa la proyección helicoidal del péptido ViCT2 y su secuencia de aminoácidos.

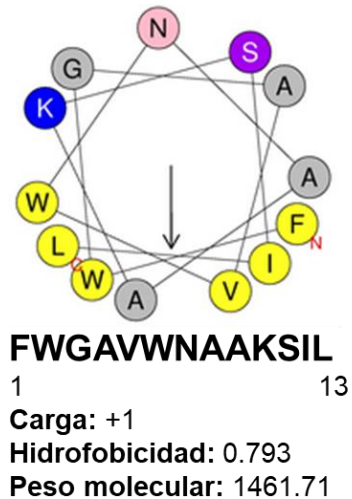


Figura 11. Características fisicoquímicas del péptido ViCT2 y su proyección helicoidal. Se muestra la secuencia de aminoácidos, la longitud del péptido, la carga, la hidrofobicidad y el peso molecular teórico. La N indica el N-terminal del péptido y la C el C-terminal del péptido. La flecha indica el ángulo polar.

A partir de la síntesis de ViCT2 se obtuvieron 100 mg de péptido crudo que se purificó por medio de cromatografía en fase reversa utilizando una columna C18 semipreparativa, con un gradiente de 0 a 100% B (acetonitrilo + 0.1% de TFA) en 60 minutos. En la figura 12 se observa el perfil cromatográfico del péptido crudo; el pico mayoritario a los 41.6 minutos (68.4 % de B). El péptido tiene una masa molecular experimental de 1461 Da y coincidió con el peso molecular teórico.

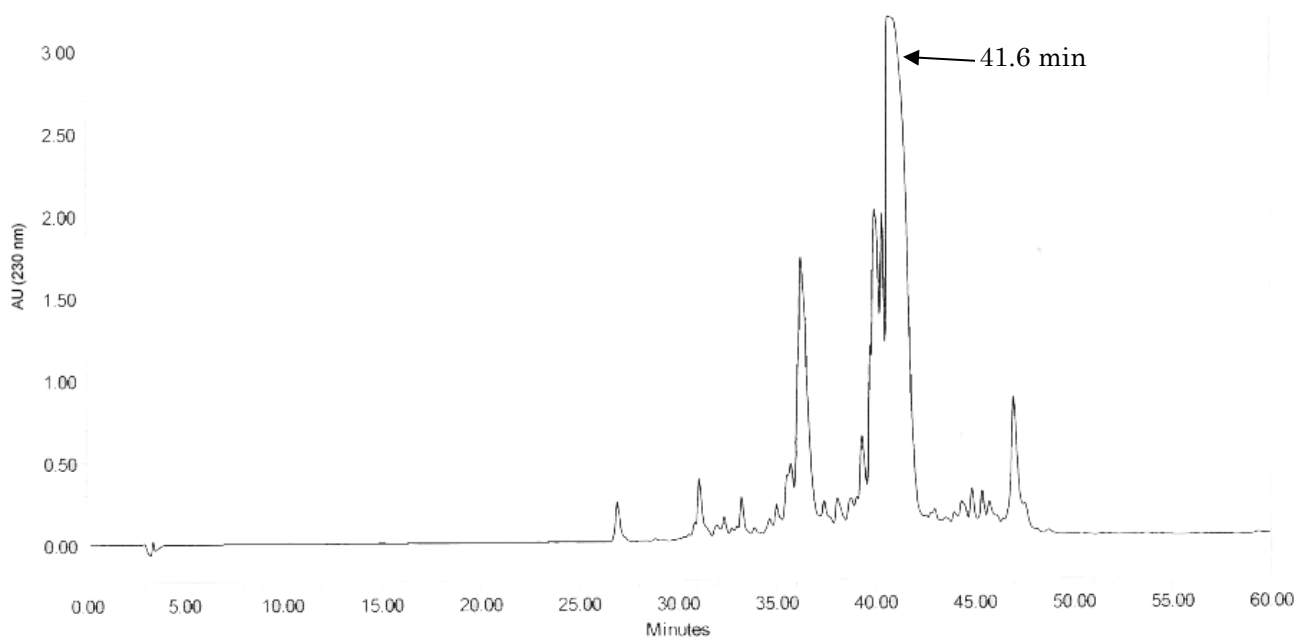


Figura 12. Perfil cromatográfico del péptido sintético ViCT2. Separación por fase reversa con una columna C18 con un gradiente de 0 a 100% de B en 60 minutos.

Una vez purificado el péptido, se cuantificó y se prepararon diluciones para hacer las pruebas de susceptibilidad anti-bacteriana. En la figura 13 se observan las curvas de crecimiento de la bacteria Gram (-) *E. coli* ATCC 25922 utilizando el péptido sintético ViCT2 como tratamiento; como control positivo se utilizó ampicilina a 500 μM y como control negativo se utilizó agua destilada. Al realizar el análisis de las curvas se observó que el péptido sintético ViCT2 tiene acción anti-bacteriana a partir de una concentración de 5 μM .

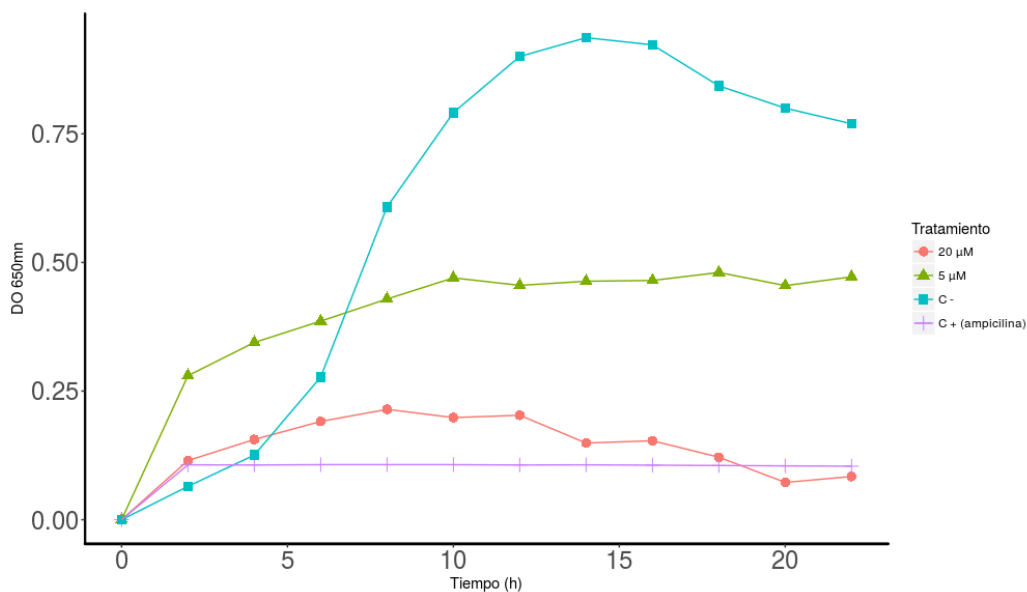


Figura 13. Actividad anti-bacteriana del péptido ViCT2 evaluada contra la cepa *E. coli* ATCC 25922. Incubación por 22 horas en medio Mueller-Hinton con concentraciones de 20 y 5 µM de ViCT2. Como control positivo se utilizó ampicilina a 500 µM y como control negativo agua destilada.

Además de probar ViCT2 en *E. coli* se decidió probar en otra bacteria Gram (-) *P. aeruginosa* ATCC 9027 y adicionalmente contra la bacteria Gram (+) *S. aureus* ATCC 25923. La actividad antibacteriana del péptido ViCT2 contra estas cepas se muestra en la figura 14 y figura 15, respectivamente. La actividad antimicrobiana de ViCT2 inhibe el crecimiento de *P. aeruginosa*, a partir de 2.5 µM.

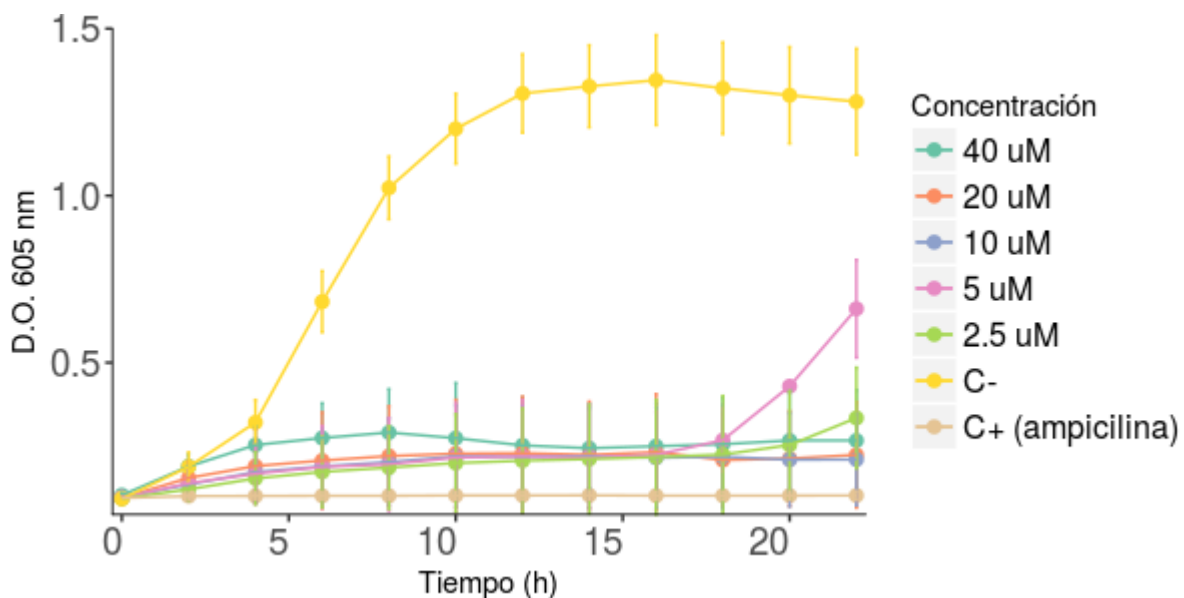


Figura 14. Actividad anti-bacteriana del péptido ViCT2 contra la cepa *P. aeruginosa* ATCC 25923. Incubación por 22 horas en medio Mueller-Hinton con concentraciones de 40 μM , 20 μM , 10 μM , 5 μM y 2.5 μM . Como control positivo se utilizó ampicilina a 500 μM y como control negativo agua destilada.

Con respecto a la actividad de ViCT2 contra la cepa *S. aureus* ATCC 25923, se observó que este péptido tiene actividad antibacteriana desde 2.5 μM .

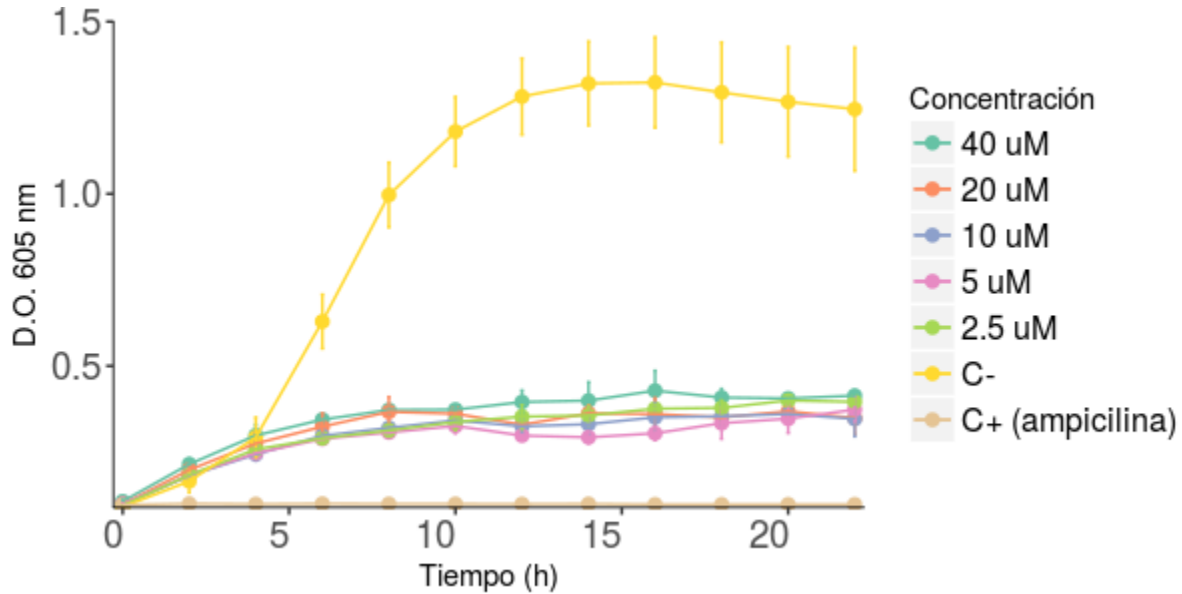


Figura 15. Actividad anti-bacteriana del péptido ViCT2 contra la cepa *S. aureus* ATCC 25923 utilizando como tratamiento el péptido sintético ViCT2. Incubación por 22 horas en medio Mueller-Hinton con concentraciones de 40 μM , 20 μM , 10 μM , 5 μM y 2.5 μM . Como control positivo se utilizó ampicilina a 500 μM y como control negativo agua destilada.

8. CONCLUSIONES

Transcriptoma:

- Se identificaron 160 secuencias con anotación de componentes de veneno dentro de las cuales se encontraron toxinas de sodio, potasio y calcio, siendo estas últimas las más conservadas presentándose pocos cambios de aminoácidos a lo largo de su secuencia. Se identificaron también péptidos de defensa del huésped como defensinas, péptidos sin enlaces disulfuro de las familias 2, 3 y 4 y otros péptidos como las waprinas identificadas en serpientes; también se identificaron enzimas como metaloproteinasas, fosfolipasas y serinproteasas; inhibidores de proteasas, péptidos tipo La1 y algunos miembros de la familia CRISP.

-La caracterización del transcriptoma de *T. atrox* mostró la presencia de 1 subfamilia no descrita previamente en alacranes; perteneciente a la sub-familia de péptidos de defensa del huésped, las waprinas.

-La cuantificación del transcriptoma demostró que los transcritos con identidad de HDPs fueron los que tuvieron mayores niveles de expresión.

Proteoma

-*Fingerprint*. En el veneno de *T. atrox* se identificaron 135 masas moleculares, 5 de ellas se lograron correlacionar con el transcriptoma.

-*Secuenciación de novo*. Se reconstruyeron 42 proteínas codificadas para transcritos ensamblados, validando así el análisis transcriptómico.

- El péptido ViCT2 mostró actividad antibacteriana contra bacterias Gram (+) y Gram (-).

9. PERSPECTIVAS

-Hacer una caracterización completa del péptido ViCT2, incluyendo pruebas como dicroísmo circular para conocer la estructura secundaria del péptido, medir la actividad antimicrobiana contra cepas resistentes a diferentes antibióticos (MDR), pruebas de actividad inmunomoduladora, antiparasitaria, antifúngica, entre otras.

-Expresar la ω -calcitoxina y caracterizar su actividad biológica.

-Expresar o sintetizar químicamente un péptido representante de cada familia de componentes de veneno identificado en el transcriptoma y caracterizar su actividad biológica.

- Medir los niveles de expresión de los componentes de veneno por PCR en tiempo real para corroborar los resultados obtenidos por el análisis de cuantificación de RNA-Seq.

-Hacer una cuantificación de los componentes obtenidos por el proteoma para conocer los péptidos más abundantes en el veneno.

10. BIBLIOGRAFÍA

- Adade, C. M., Carvalho, A. L. O., Tomaz, M. A., Costa, T. F. R., Godinho, J. L., Melo, P. A., ... Souto-Padr??n, T. (2014). Crovirin, a Snake Venom Cysteine-Rich Secretory Protein (CRISP) with Promising Activity against *Trypanosomes* and *Leishmania*. *PLoS Neglected Tropical Diseases*, 8(10). <https://doi.org/10.1371/journal.pntd.0003252>
- Adams, M. E. (2004). Agatoxins: Ion channel specific toxins from the american funnel web spider, *Agelenopsis aperta*. *Toxicon*, 43(5), 509–525. <https://doi.org/10.1016/j.toxicon.2004.02.004>
- Almaaytah, A., & Albalas, Q. (2014). Scorpion venom peptides with no disulfide bridges: A review. *Peptides*, 51, 35–45. <https://doi.org/10.1016/j.peptides.2013.10.021>
- Almeida, D. D., Scortecchi, K. C., Kobashi, L. S., A., L. F., Medeiros, S. R., Silva-Junior, A. A., ... de Fernandes-Pedrosa, M. (2012). Profiling the resting venom gland of the scorpion *Tityus stigmurus* through a transcriptomic survey. *BioMed Central*, 13(362). <https://doi.org/10.1186/1471-2164-13-362>
- Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology*, 56(4), 394–404. <https://doi.org/10.1007/s12088-016-0606-4>
- Armstrong, P. B. (2001). The contribution of proteinase inhibitors to immune defense. *Trends in Immunology*, 22(1), 47–52. [https://doi.org/10.1016/S1471-4906\(00\)01803-2](https://doi.org/10.1016/S1471-4906(00)01803-2)
- Batista, C. V. F., Martins, J. G., Restano-Cassulini, R., Coronas, F. I. V., Zamudio, F. Z., Procópio, R., & Possani, L. D. (2018). Venom characterization of the Amazonian scorpion *Tityus metuendus*. *Toxicon*, 143, 51–58. <https://doi.org/10.1016/j.toxicon.2018.01.006>
- Bordon, K. C. F., Wiesel, G. A., Amorim, F. G., & Arantes, E. C. (2015). Arthropod venom Hyaluronidases: Biochemical properties and potential applications in medicine and biotechnology. *Journal of Venomous Animals and Toxins Including Tropical Diseases*, 21(1), 1–12. <https://doi.org/10.1186/s40409-015-0042-7>
- Brogden, K. A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews. Microbiology*, 3(3), 238–50. <https://doi.org/10.1038/nrmicro1098>
- Cao, Z., Di, Z., Wu, Y., & Li, W. (2014). Overview of scorpion species from China and their toxins. *Toxins*, 6(3), 796–815. <https://doi.org/10.3390/toxins6030796>
- Carcamo-Noriega, E. N., Olamendi-Portugal, T., Restano-Cassulini, R., Rowe, A., Uribe-Romero, S. J., Becerril, B., & Possani, L. D. (2018). Intraspecific variation of *Centruroides sculpturatus* scorpion venom from two regions of Arizona. *Archives of Biochemistry and Biophysics*, 638(November 2017), 52–57. <https://doi.org/10.1016/j.abb.2017.12.012>
- Cera, E. Di. (2009). NIH Public Access. *International Union of Biochemistry and Molecular Biology Life*, 61(5), 510–515. <https://doi.org/10.1002/iub.186>.Serine
- Chagot, B., Pimentel, C., Dai, L., Pil, J., Tytgat, J., Nakajima, T., ... Ferrat, G. (2005). An unusual fold for potassium channel blockers: NMR structure of three toxins from the scorpion *Opisthacanthus madagascariensis*. *The Biochemical Journal*, 388(Pt 1), 263–271. <https://doi.org/10.1042/BJ20041705>
- Chen, Z., Luo, F., Feng, J., Yang, W., Zeng, D., Zhao, R., ... Wu, Y. (2013). Genomic and Structural Characterization of Kunitz-Type Peptide LmKTT-1a Highlights Diversity and Evolution of Scorpion Potassium Channel Toxins. *PLoS ONE*, 8(4), 1–10. <https://doi.org/10.1371/journal.pone.0060201>
- Chuang, R. S., Jaffe, H., Cribbs, L., Perez-Reyes, E., & Swartz, K. J. (1998). Inhibition of T-type voltage-gated calcium channels by a new scorpion toxin. *Nature Neuroscience*, 1(8), 668–674. <https://doi.org/10.1038/3669>
- Conde, R., Zamudio, F. Z., Rodriguez, M. H., & Possani, L. D. (2000). Scorpine, an anti-malaria and anti-bacterial agent purified from scorpion venom. *FEBS Letters*, 471(2–3), 165–168.
- Couraud, F., Jover, E., Dubois, J. M., & Rochat, H. (1982). Two types of scorpion toxin receptor sites, one related to the activation, the other to the inactivation of the action potential sodium channel. *Toxicon*, 20(1), 9–16. [https://doi.org/10.1016/0041-0101\(82\)90138-6](https://doi.org/10.1016/0041-0101(82)90138-6)

- Cuperus, T., Coorens, M., van Dijk, A., & Haagsman, H. P. (2013). Avian host defense peptides. *Developmental and Comparative Immunology*, *41*(3), 352–369. <https://doi.org/10.1016/j.dci.2013.04.019>
- Dai, H., Yin, S., Li, T., Cao, Z., Ji, Y., Wu, Y., & Li, W. (2012). Recombinant expression, purification, and characterization of scorpion toxin BmTX14. *Protein Expression and Purification*, *82*(2), 325–331. <https://doi.org/10.1016/j.pep.2012.02.001>
- de la Vega, R. C. R., & Possani, L. D. (2007). Novel paradigms on scorpion toxins that affects the activating mechanism of sodium channels. *Toxicon*, *49*(2), 171–180. <https://doi.org/10.1016/j.toxicon.2006.09.016>
- De Oliveira, U. C., Candido, D. M., Coronado Dorce, V. A., & Junqueira-De-Azevedo, I. D. L. M. (2015). The transcriptome recipe for the venom cocktail of *Tityus bahiensis* scorpion. *Toxicon*, *95*, 52–61. <https://doi.org/10.1016/j.toxicon.2014.12.013>
- De Oliveira, U. C., Nishiyama, M. Y., Dos Santos, M. B. V., De Paula Santos-Da-Silva, A., De Menezes Chalkidis, H., Souza-Imberg, A., ... Junqueira-de-Azevedo, I. de L. M. (2018). Proteomic endorsed transcriptomic profiles of venom glands from *Tityus obscurus* and *T. serrulatus* scorpions. *PLoS ONE*, *13*(3), 1–23. <https://doi.org/10.1371/journal.pone.0193739>
- Deng, Y., Gu, J., Yan, Z., Wang, M., Ma, C., Zhang, J., ... Xiao, L. (2018). De novo transcriptomic analysis of the venomous glands from the scorpion *Heterometrus spinifer* revealed unique and extremely high diversity of the venom peptides. *Toxicon*, *143*, 1–19. <https://doi.org/10.1016/j.toxicon.2017.12.051>
- Diego-García, E., Caliskan, F., & Tytgat, J. (2014). The Mediterranean scorpion *Mesobuthus gibbosus* (Scorpiones, Buthidae): Transcriptome analysis and organization of the genome encoding chlorotoxin-like peptides. *BMC Genomics*, *15*(1), 1–16. <https://doi.org/10.1186/1471-2164-15-295>
- Diego-García, E., Peigneur, S., Clynen, E., Marien, T., Czech, L., Schoofs, L., & Tytgat, J. (2012). Molecular diversity of the telson and venom components from *Pandinus cavimanus* (Scorpionidae Latreille 1802): Transcriptome, venomics and function. *Proteomics*, *12*(2), 313–328. <https://doi.org/10.1002/pmic.201100409>
- Fletcher, P. L., Fletcher, M. D., Weninger, K., Anderson, T. E., & Martin, B. M. (2010). Vesicle-associated membrane protein (VAMP) cleavage by a new metalloprotease from the Brazilian scorpion *Tityus serrulatus*. *Journal of Biological Chemistry*, *285*(10), 7405–7416. <https://doi.org/10.1074/jbc.M109.028365>
- Flores-Solis, D., Toledano, Y., Rodríguez-Lima, O., Cano-Sánchez, P., Ramírez-Cordero, B. E., Landa, A., ... del Rio-Portilla, F. (2016). Solution structure and antiparasitic activity of scorpine-like peptides from *Hoffmanniadrurus gertschi*. *FEBS Letters*, *590*, 2286–2296. <https://doi.org/10.1002/1873-3468.12255>
- Gibbs, G. M., Roelants, K., & O'Bryan, M. K. (2008). The CAP superfamily: Cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins - Roles in reproduction, cancer, and immune defense. *Endocrine Reviews*, *29*(7), 865–897. <https://doi.org/10.1210/er.2008-0032>
- González-Santillán, E., & Prendini, L. (2013). Redefinition and generic revision of the north american Vaejovid scorpion subfamily Syntropinae Kraepelin 1905, with descriptions of six new genera. *Bulletin of the American Museum of Natural History*, *382*, 341–405. <https://doi.org/10.1111/cla.12091>
- Hagiwara, K., Kikuchi, T., Endo, Y., Usui, K., Takahashi, M., Shibata, N., ... Nukiwa, T. (2003). Mouse SWAM1 and SWAM2 Are Antibacterial Proteins Composed of a Single Whey Acidic Protein Motif. *The Journal of Immunology*, *170*(4), 1973–1979. <https://doi.org/10.4049/jimmunol.170.4.1973>
- Hancock, R. E. W., Haney, E. F., & Gill, E. E. (2016). The immunology of host defence peptides: beyond antimicrobial activity. *Nature Reviews Immunology*, *16*(5), 321–334. <https://doi.org/10.1038/nri.2016.29>
- Harris, F., Dennison, S., & Phoenix, D. (2009). Anionic Antimicrobial Peptides from Eukaryotic Organisms. *Current Protein & Peptide Science*, *10*(6), 585–606. <https://doi.org/10.2174/138920309789630589>
- Hernández-Aponte, C. A., Silva-Sanchez, J., Quintero-Hernández, V., Rodríguez-Romero, A., Balderas, C., Possani, L. D., & Gurrola, G. B. (2011). Vaejovine, a new antibiotic from the scorpion venom of *Vaejovis mexicanus*. *Toxicon*, *57*(1), 84–92. <https://doi.org/10.1016/j.toxicon.2010.10.008>

- Kuzmenkov, A. I., Vassilevski, A. A., Kudryashova, K. S., Nekrasova, O. V., Peigneur, S., Tytgat, J., ... Grishin, E. V. (2015). Variability of potassium channel blockers in *Mesobuthus eupeus* scorpion venom with focus on Kv1.1: An integrated transcriptomic and proteomic study. *Journal of Biological Chemistry*, 290(19), 12195–12209. <https://doi.org/10.1074/jbc.M115.637611>
- Li, H., & Xia, Y. (2018). Recombinant production of the insecticidal scorpion toxin BjaIT in *Escherichia coli*. *Protein Expression and Purification*, 142, 62–67. <https://doi.org/10.1016/j.pep.2017.10.003>
- Liu, D., Wang, Y., Wei, L., Ye, H., Liu, H., Wang, L., ... Lai, R. (2013). Snake venom-like waprin from the frog of *Ceratophrys calcarata* contains antimicrobial function. *Gene*, 514(2), 99–104. <https://doi.org/10.1016/j.gene.2012.11.007>
- López-González, I., Olamendi-Portugal, T., De La Vega-Beltrán, J. L., Van Der Walt, J., Dyason, K., Possani, L. D., ... Darszon, A. (2003). Scorpion toxins that block T-type Ca²⁺ channels in spermatogenic cells inhibit the sperm acrosome reaction. *Biochemical and Biophysical Research Communications*, 300(2), 408–414. [https://doi.org/10.1016/S0006-291X\(02\)02859-0](https://doi.org/10.1016/S0006-291X(02)02859-0)
- Lourenço, W. R. (2018). The evolution and distribution of noxious species of scorpions (Arachnida: Scorpiones). *Journal of Venomous Animals and Toxins Including Tropical Diseases*, 24(1), 1–12. <https://doi.org/10.1186/s40409-017-0138-3>
- Luna-Ramírez, K., Quintero-Hernández, V., Juárez-González, V. R., & Possani, L. D. (2015). Whole transcriptome of the venom gland from *Urodacus yaschenkoi* scorpion. *PLoS ONE*, 10(5), 1–33. <https://doi.org/10.1371/journal.pone.0127883>
- Ma, Y., He, Y., Zhao, R., Wu, Y., Li, W., & Cao, Z. (2012). Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: Implication for proteome evolution of scorpion venom arsenal. *Journal of Proteomics*, 75(5), 1563–1576. <https://doi.org/10.1016/j.jprot.2011.11.029>
- Ma, Y., Zhao, R., He, Y., Li, S., Liu, J., Wu, Y., ... Li, W. (2009). Transcriptome analysis of the venom gland of the scorpion *Scorpiops jendeki*: Implication for the evolution of the scorpion venom arsenal. *BMC Genomics*, 10, 1–15. <https://doi.org/10.1186/1471-2164-10-290>
- Martinez, G., Hograindleur, J. P., Voisin, S., Nahed, R. A., Aziz, T. M. A. El, Escoffier, J., ... Arnoult, C. (2017). Spermaurin, an La1-like peptide from the venom of the scorpion *Scorpio maurus palmatus*, improves sperm motility and fertilization in different mammalian species. *Molecular Human Reproduction*, 23(2), 116–131. <https://doi.org/10.1093/molehr/gaw075>
- Matsushita, N., Miyashita, M., Sakai, A., Nakagawa, Y., & Miyagawa, H. (2007). Purification and characterization of a novel short-chain insecticidal toxin with two disulfide bridges from the venom of the scorpion *Liocheles australasiae*. *Toxicon*, 50(6), 861–867. <https://doi.org/10.1016/j.toxicon.2007.06.014>
- Melo, M. M. A., Daniele-Silva, A., Teixeira, D. G., Estrela, A. B., Melo, K. R. T., Oliveira, V. S., ... Fernandes-Pedrosa, M. F. (2017). Structure and in vitro activities of a Copper II-chelating anionic peptide from the venom of the scorpion *Tityus stigmurus*. *Peptides*, 94(September 2016), 91–98. <https://doi.org/10.1016/j.peptides.2017.05.009>
- Meng, L., Xie, Z., Zhang, Q., Li, Y., Yang, F., Chen, Z., ... Wu, Y. (2016). Scorpion potassium channel-blocking defensin highlights a functional link with neurotoxin. *Journal of Biological Chemistry*, 291(13), 7097–7106. <https://doi.org/10.1074/jbc.M115.680611>
- Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society*, 85(14), 2149–2154. <https://doi.org/10.1021/ja00897a025>
- Mille, B. G., Peigneur, S., Diego-García, E., Predel, R., & Tytgat, J. (2014). Partial transcriptomic profiling of toxins from the venom gland of the scorpion *Parabuthus stridulus*. *Toxicon*, 83, 75–83. <https://doi.org/10.1016/j.toxicon.2014.03.001>
- Mille, B. G., Peigneur, S., Predel, R., & Tytgat, J. (2015). Transcriptomic approach reveals the molecular diversity of *Hottentotta conspersus* (Buthidae) venom. *Toxicon*, 99, 73–79. <https://doi.org/10.1016/j.toxicon.2015.03.015>
- Miyashita, M., Otsuki, J., Hanai, Y., Nakagawa, Y., & Miyagawa, H. (2007). Characterization of peptide components in the

- venom of the scorpion *Liocheles australasiae* (Hemiscorpiidae). *Toxicon*, 50(3), 428–437. <https://doi.org/10.1016/j.toxicon.2007.04.012>
- Morgenstern, D., Rohde, B. H., King, G. F., Tzachy, T., Sher, D., & Zlotkin, E. (2011). The tale of a resting gland: Transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*. *Toxicon*, 57(5), 695–703. <https://doi.org/10.1016/j.toxicon.2011.02.001>
- NaderiSoorki, M., Galehdari, H., Baradaran, M., & Jalali, A. (2016). First venom gland transcriptomic analysis of Iranian yellow scorpion “*Odonthubuthus doriae*” with some new findings. *Toxicon*, 120, 69–77. <https://doi.org/10.1016/j.toxicon.2016.07.010>
- Olamendi-Portugal T, García BI, López-González I, Van Der Walt J, Dyason K, Ulens C, Tytgat J, Felix R, Darszon A, P. L. (2002). Two new scorpion toxins that target voltage-gated Ca²⁺ and Na⁺ channels. *Biochem Biophys Res Commun*, 299(4), 562–568.
- Oppenheim, J. J., Biragyn, A., Kwak, L. W., & Yang, D. (2003). Roles of antimicrobial peptides such as defensins in innate and adaptive immunity. *Ann. Rheum. Dis*, 62, ii17–ii21. https://doi.org/10.1136/ard.62.suppl_2.ii17
- Ortiz, E., Gurrola, G. B., Schwartz, E. F., & Possani, L. D. (2015). Scorpion venom components as potential candidates for drug development. *Toxicon: Official Journal of the International Society on Toxinology*, 93, 125–135. <https://doi.org/10.1016/j.toxicon.2014.11.233>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417–419. <https://doi.org/10.1038/nmeth.4197>
- Possani, L. D., Becerril, B., Delepierre, M., & Tytgat, J. (1999). Scorpion toxins specific for Na⁺-channels. *European Journal of Biochemistry*, 264(2), 287–300.
- Postay, Possani, Lourival, D. (MX), Gurrola-Briones, G. (MX), Salas-Castillo, S. P. (MX), Ferreira Batista, C. V. (MX), Varga, Zoltán, S. . (HU), Panyi, G. (HU), & Gáspár, R. (HU). (2008). *US 8,394,770 B2*.
- Quintero-Hernández, V., Jiménez-Vargas, J. M., Gurrola, G. B., Valdivia, H. H. F., & Possani L. D. (2013). Scorpion venom components that affect ion-channels function. *NIH Public Access*, 86(12), 3279–3288. <https://doi.org/10.1007/s11103-011-9767-z>.Plastid
- Quintero-Hernandez, V., Ramirez-Carreto, S., Romero-Gutierrez, M. T., Valdez-Velazquez, L. L., Becerril, B., Possani, L. D., & Ortiz, E. (2015). Transcriptome analysis of scorpion species belonging to the *Vaejovis* genus. *PLoS One*, 10(2), e0117188. <https://doi.org/10.1371/journal.pone.0117188>
- Ramirez, K. L., & Jimenez Vargas, J. M. (2016). Scorpine-Like Peptides. *Single Cell Biology*, 5(2), 2–4. <https://doi.org/10.4172/2168-9431.1000138>
- Rates, B., Ferraz, K. K. F., Borges, M. H., Richardson, M., De Lima, M. E., & Pimenta, A. M. C. (2008). Tityus serrulatus venom peptidomics: Assessing venom peptide diversity. *Toxicon*, 52(5), 611–618. <https://doi.org/10.1016/j.toxicon.2008.07.010>
- Rendón-Anaya, M., Delaye, L., Possani, L. D., & Herrera-estrella, A. (2012). Global Transcriptome Analysis of the Scorpion *Centruroides noxius* : New Toxin Families and Evolutionary Insights from an Ancestral Scorpion Species. *PLoS One*, 7(8), 1–12. <https://doi.org/10.1371/journal.pone.0043331>
- Rochat, C., Rochat, H., Miranda, F., & Lissitzky, S. (1967). Purification and Some Properties of the Neurotoxins of *Androctonus australis* Hector. *Biochemistry*, 6(2), 578–585. <https://doi.org/10.1021/bi00854a028>
- Rodríguez de la Vega, R. C., Schwartz, E. F., & Possani, L. D. (2010). Mining on scorpion venom biodiversity. *Toxicon: Official Journal of the International Society on Toxinology*, 56(7), 1155–61. <https://doi.org/10.1016/j.toxicon.2009.11.010>
- Roeding, F., Borner, J., Kube, M., Klages, S., Reinhardt, R., & Burmester, T. (2009). A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (*Pandinus imperator*). *Molecular Phylogenetics and Evolution*, 53(3), 826–834. <https://doi.org/10.1016/j.ympev.2009.08.014>

- Rokyta, D. R., & Ward, M. J. (2017). Venom-gland transcriptomics and venom proteomics of the black-back scorpion (*Hadrurus spadix*) reveal detectability challenges and an unexplored realm of animal toxin diversity. *Toxicon*, *128*, 23–37. <https://doi.org/10.1016/j.toxicon.2017.01.014>
- Ruiming, Z., Yibao, M., Yawen, H., Zhiyong, D., Yingliang, W., Zhijian, C., & Wenxin, L. (2010a). Comparative venom gland transcriptome analysis of the scorpion *Lychas mucronatus* reveals intraspecific toxic gene diversity and new venomous components. *BMC Genomics*, *11*, 452. <https://doi.org/10.1186/1471-2164-11-452>
- Santibanez-Lopez, C. E., & Possani, L. D. (2015). Overview of the Knottin scorpion toxin-like peptides in scorpion venoms: Insights on their classification and evolution. *Toxicon : Official Journal of the International Society on Toxinology*, *107*(Pt B), 317–326. <https://doi.org/10.1016/j.toxicon.2015.06.029>
- Santibáñez-López, C. E., Cid-Urbe, J. I., Batista, C. V. F., Ortiz, E., & Possani, L. D. (2016). Venom gland transcriptomic and proteomic analyses of the enigmatic scorpion *Superstitionia donensis* (Scorpiones: Superstitioniidae), with insights on the evolution of its venom components. *Toxins*, *8*(12), 3–6. <https://doi.org/10.3390/toxins8120367>
- Santibáñez-López, C. E., Cid-Urbe, J. I., Zamudio, F. Z., Batista, C. V. F., Ortiz, E., & Possani, L. D. (2017). Venom gland transcriptomic and venom proteomic analyses of the scorpion *Megacormus gertschi* Díaz-Najera, 1966 (Scorpiones: Euscorpiidae: Megacorminae). *Toxicon*, *133*, 95–109. <https://doi.org/10.1016/j.toxicon.2017.05.002>
- Schwartz, E., Diego-garcia, E., Rodríguez de la Vega, R., & Possani, L. D. (2007). Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones). *BMC Genomics*, *8*(119), 1–12. <https://doi.org/10.1186/1471-2164-8-119>
- Schwartz, E. F., Capes, E. M., Diego-Garcia, E., Zamudio, F. Z., Fuentes, O., Possani, L. D., & Valdivia, H. H. (2009). Characterization of hadrucalcin, a peptide from *Hadrurus gertschi* scorpion venom with pharmacological activity on ryanodine receptors. *British Journal of Pharmacology*, *157*(3), 392–403. <https://doi.org/10.1111/j.1476-5381.2009.00147.x>
- Sidach, S. S., & Mintz, I. M. (2002). Kurtoxin, a gating modifier of neuronal high- and low-threshold Ca channels. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *22*(6), 2023–2034. <https://doi.org/10.1523/JNEUROSCI.2262-02.2002> [pii]
- Smith, J. J., Hill, J. M., Little, M. J., Nicholson, G. M., King, G. F., & Alewood, P. F. (2011). Unique scorpion toxin with a putative ancestral fold provides insight into evolution of the inhibitor cysteine knot motif. *Proceedings of the National Academy of Sciences*, *108*(26), 10478–10483. <https://doi.org/10.1073/pnas.1103501108>
- Soleglad, M. E., & Fet, V. (2004). The Systematics of the Scorpion Subfamily *Uroctoninae* (Scorpiones : Chactidae). *Revista Ibérica de Aracnología*, *10*, 81–128.
- Soleglad, M. E., & Fet, V. (2008). Contributions to Scorpion Systematics. III. Subfamilies Smeringurinae and Syntropinae (Scorpiones: Vaejovidae). *Occasional Publications in Scorpology*, (71).
- Steinraesser, L., Kraneburg, U., Jacobsen, F., & Al-Benna, S. (2011). Host defense peptides and their antimicrobial-immunomodulatory duality. *Immunobiology*, *216*(3), 322–333. <https://doi.org/10.1016/j.imbio.2010.07.003>
- Tytgat, J., Chandy, G., Garcia, M. L., Gutman, G. A., Martin-Eauclaire, M., van der Walt, J., & Possani, L. D. (1999). A unified nomenclature for short-chain peptides isolated from scorpion venoms : α -KTx molecular subfamilies. *TiPS*, *20*, 444–447.
- Utkin, Y. N. (2015). Animal venom studies: Current benefits and future developments. *World Journal of Biological Chemistry*, *6*(2), 28. <https://doi.org/10.4331/wjbc.v6.i2.28>
- Valdez-Cruz, N. A., Batista, C. V. F., Zamudio, F. Z., Bosmans, F., Tytgat, J., & Possani, L. D. (2004). Phaiodotoxin, a novel structural class of insect-toxin isolated from the venom of the Mexican scorpion *Anuroctonus phaiodactylus*. *European Journal of Biochemistry*, *271*(23–24), 4753–4761. <https://doi.org/10.1111/j.1432-1033.2004.04439.x>
- Valdez-Velázquez, L. L., Quintero-Hernández, V., Romero-Gutiérrez, M. T., Coronas, F. I. V., & Possani, L. D. (2013). Mass Fingerprinting of the Venom and Transcriptome of Venom Gland of Scorpion *Centruroides tecomanus*. *PLoS ONE*, *8*(6). <https://doi.org/10.1371/journal.pone.0066486>
- Vandendriessche, T., Kopljar, I., Jenkins, D. P., Diego-Garcia, E., Abdel-Mottaleb, Y., Vermassena, E., ... Tytgat, J. (2013).

Purification, molecular cloning and functional characterization of HelaTx1 (*Heterometrus laoticus*): The first member of a new k-KTX subfamily. *Biochemical Pharmacology*, 83(9), 1307–1317. <https://doi.org/10.1016/j.bcp.2012.01.021>.PURIFICATION

- Vines, C. M., & Bill, C. A. (2015). Phospholipases. *eLS*, (May), 1–9. <https://doi.org/10.1002/9780470015902.a0001394.pub3>
- Ward, M. J., Ellsworth, S. A., & Rokyta, D. R. (2018). Venom-gland transcriptomics and venom proteomics of the Hentz striped scorpion (*Centruroides hentzi*; Buthidae) reveal high toxin diversity in a harmless member of a lethal family. *Toxicon*, 142, 14–29. <https://doi.org/10.1016/j.toxicon.2017.12.042>
- Wilmes, M., & Sahl, H. G. (2014). Defensin-based anti-infective strategies. *International Journal of Medical Microbiology*, 304(1), 93–99. <https://doi.org/10.1016/j.ijmm.2013.08.007>
- Winnebeck, E. C., Millar, C. D., & Warman, G. R. (2010). Why Does Insect RNA Look Degraded? *Journal of Insect Science*, 10(159), 1–7. <https://doi.org/10.1673/031.010.14119>
- Wu, S., Nie, Y., Zeng, X. C., Cao, H., Zhang, L., Zhou, L., ... Liu, Y. (2014). Genomic and functional characterization of three new venom peptides from the scorpion *Heterometrus spinifer*. *Peptides*, 53, 30–41. <https://doi.org/10.1016/j.peptides.2013.12.012>
- Xiao, L., Gurrola, G. B., Zhang, J., Valdivia, C. R., SanMartin, M., Zamudio, F. Z., ... Valdivia, H. H. (2016). Structure–function relationships of peptides forming the calcin family of ryanodine receptor ligands. *The Journal of General Physiology*, 147(5), 375–394. <https://doi.org/10.1085/jgp.201511499>
- Yibao, M., Yong, Z., Ruimming, Z., Weiping, Z., Yawen, H., Yingliang, W., ... L., W. (2010). Molecular diversity of toxic components from the scorpion *Heterometrus petersii* venom revealed by proteomic and transcriptome analysis. *Proteomics*, 10, 2471–2485. <https://doi.org/10.1002/pmic.200900763>
- Zhang, L., Shi, W., Zeng, X. C., Ge, F., Yang, M., Nie, Y., ... E, G. (2015). Unique diversity of the venom peptides from the scorpion *Androctonus bicolor* revealed by transcriptomic and proteomic analysis. *Journal of Proteomics*, 128, 231–250. <https://doi.org/10.1016/j.jprot.2015.07.030>
- Zhong, J., Zeng, X. C., Zeng, X., Nie, Y., Zhang, L., Wu, S., & Bao, A. (2017). Transcriptomic analysis of the venom glands from the scorpion *Hadogenes troglodytes* revealed unique and extremely high diversity of the venom peptides. *Journal of Proteomics*, 150, 40–62. <https://doi.org/10.1016/j.jprot.2016.08.004>
- Zhu, S., & Tytgat, J. (2004). The scorpine family of defensins: gene structure, alternative polyadenylation and fold recognition. *Cellular and Molecular Life Sciences*, 61(14), 1751–1763. <https://doi.org/10.1007/s00018-004-4149-1>

11. ANEXOS

Anexo 1. Purificación de RNA a partir de la glándula venenosa usando el kit SV Total RNA Isolation System® Promega

1. Limpiar el exterior del alacrán con etanol y agua para retirar impurezas, cortar el telson y colocarlo en un tubo de 1.5 ml con 175 μ l de buffer de lisis (Tiocianato de guanidinio 4 M, Tris pH 7.5 0.01 M, β - mercaptoetanol 0.97 %) homogenizar con un pistilo eppendorf.
2. Adicionar 300 μ l de solución de dilución de RNA mezclar por inversión 3-4 veces, centrifugar por 10 min a 14 000 x g.
3. Transferir el sobrenadante a un tubo limpio y adicionar 200 μ l de etanol 95 % mezclar con la pipeta 3-4 veces, transferir la mezcla a una columna de purificación ensamblado con un colector (columna-colector) y centrifugar 1 min a 14 000 x g. Desechar el sobrenadante.
4. Adicionar 600 μ l de RNA Wash Solution (acetato de potasio 60 mM, Tris-ácido clorhídrico pH 7.5 y etanol 60 %) en la columna y centrifugar 1 min a 14 000 x g. Desechar el sobrenadante y ensamblar columna-colector.
5. Preparar la mezcla 40 μ l de Yellow Core Buffer (Tris pH 7.5 0.0225 M, Cloruro de sodio 1.125 M, colorante amarillo 0.0025%), 5 μ l de Cloruro de manganeso 0.09 M, y 5 μ l de DNAsa I (siempre usar la mezcla recién preparada) y aplicarla sobre la membrana de la columna de purificación. Incubar 15 min a 20- 25 °C.
6. Adicionar 200 μ l de DNase Stop Solution (Isotiocianato de guanidinio 2 M, Tris-ácido clorhídrico pH 7.5 4 mM, etanol 57%) en la columna-colector y centrifugar 1 min a 14 000 x g. Desechar el sobrenadante.
7. Adicionar 600 μ l de RNA Wash Solution en la columna-colector y centrifugar 1 min a 14 000 x g. Desechar el sobrenadante.
8. Adicionar 250 μ l de RNA Wash Solution en la columna-colector y centrifugar 2 min a 14 000 x g. Desechar el sobrenadante.
9. Adicionar 100 μ l de agua libre de nucleasas sobre la membrana de la columna ensamblada a un tubo de 1.5 ml limpio y centrifugar 1 min a 14 000 x g. En el sobrenadante tenemos el RNA purificado.
10. Cuantificar a 260 nm en un Nanodrop 1000 spectrophotometer ® Thermo scientific.
11. Almacenarlo a -70 °C.

Anexo 2. Dominios PFAM Utilizados para hacer búsquedas en los archivos de anotación del transcriptoma de *T. atrox*.

Dominio PFAM	Identificador	Descripción
Toxin_2	PF00451	Scorpion short toxin, BmKK2
Toxin_3	PF00537	Scorpion toxin-like domain
Toxin_5	PF05294	Scorpion short toxin
Toxin_6	PF05453	BmTXKS1/BmP02 toxin family
Toxin_17	PF08086	Ergtoxin family
Toxin_25	PF08095	Hefutoxin family
BmKX	PF09132	BmKX
Toxin_27	PF08099	Scorpion calcine family
Antimicrobial_7	PF08102	Scorpion antimicrobial peptide
Toxin_31	PF08119	Scorpion acidic alpha-KTx toxin family
Toxin_32	PF08120	Tamulustoxin family
Toxin_38	PF14866	Potassium channel toxin
Defensin_2	PF01097	Arthropod defensin
Kunitz_BPTI	PF00014	Kunitz/Bovine pancreatic trypsin inhibitor domain
Toxin_34	PF08396	Spider toxin omega agatoxin/Tx1 family
Serpin	PF00079	Serpin (serine protease inhibitor)
HSP90	PF00183	Hsp90 protein
Peptidase_M24	PF00557	Metallopeptidase family M24
Glyco_hydro_56	PF01630	Hyaluronidase
WAP	PF00095	Whey Acidic Protein
Phospholip_A2_2	PF05826	Phospholipase A2
Antimicrobial_4	PF08024	Ant antimicrobial peptide

Anexo 3. Comandos utilizados para caracterizar el transcriptoma y medir los niveles de expresión de los transcritos relacionados con veneno.

FastQC.

FastQC es un programa que, basado en el archivo fastq computa indicadores de calidad de la secuenciación. Existen dos modalidades de esta herramienta: online y en línea de comandos.

Uso:

```
>fastqc <archivo.fastq>
```

El resultado es un archivo en html que contiene todos los gráficos de calidad de la secuenciación.

Más información en: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Trinity

Trinity es la herramienta que ensambla *de novo* los *reads* provenientes de secuenciación por Illumina. El comando toma en cuenta que la librería hecha es paired-end.

Uso:

```
>Trinity --seqType fq --left file_R1.fastq -right file_R2.fastq --  
max_memory 184G --SS_lib_type RF -min_contig_length 100 --CPU 48
```

Donde:

--seqtype: tipo de archivos de secuenciación. En este caso fueron fastq.

--left: Archivo fastq directo.

--right: Archivo fastq reverso.

--max_memory:

--SS_lib_type: Orientación de los archivos fastq. F, reverso; R, directo.

-min_contig_length: Tamaño mínimo de contig ensamblado reportado.

--CPU: Número de CPUs que serán utilizados en el proceso.

Más información:

<https://github.com/trinityrnaseq/trinityrnaseq/wiki/Running-Trinity>

Uso de Blast en línea de comandos:

1.Instalación del programa.

Blast en línea de comandos puede ser bajado de los repositorios. Este programa está disponible para Mac OSX y Lynux/Unix (ver el manual <https://www.ncbi.nlm.nih.gov/books/NBK279690/>).

```
>sudo apt-get install ncbi-blast+
```

2. Construcción de la base de datos.

Para ejecutar blast se necesita tener la librería de transcritos donde se van a hacer las búsquedas que pueden ser locales o las bases de datos ya publicadas en NCBI.

```
>makeblastdb -in <libreria_de_secuencias> -dbtype prot -out database.out
```

Donde:

-in: Librería de secuencias que se va a convertir en la base de datos. Tiene que estar en formato FASTA.

-dbtype: Es el tipo de molécula que tiene la base de datos, prot es para proteína y nucl es para DNA.

-out: Es el nombre que se le da a la base de datos resultante.

Como resultado de este comando se obtendrán 3 diferentes archivos con extensión *.nhr, *.nir y *.nsq.

Para consultar la ayuda, teclear makeblastdb -help.

3. Búsqueda de secuencias en bases de datos:

El algoritmo de blast tiene 4 modalidades de búsqueda:

Comando	Descripción
Blastn	Busca en bases de datos de nucleótidos las secuencias de interés en nucleótidos.
Blastp	Busca en bases de datos de proteínas las secuencias de interés en aminoácidos.
Blastx	Busca en bases de datos de proteínas las secuencias de interés traducidas a nucleótido en los 6 marcos de lectura.
tblastx	Traduce a nucleótidos tanto la base de datos como las secuencias de interés.
tblastn	Busca en bases de datos de nucleótidos las secuencias de interés traducidas de aminoácidos a nucleótidos.

El comando es el siguiente:

```
>blastn -db <base_de_datos> -evaluate 0.0001 -query <secuencia de interés>  
-outfmt 6 -out <output.out>
```

Donde:

-db: Base de datos construida previamente. Sólo escribir el nombre, no la extensión.

-evaluate: valor de corte. Lo ideal es 0.0001.

-query: Secuencias de interés.

-outfmt: Formato separado por tabuladores.

-out: Formato de salida.

Más información en: <https://www.ncbi.nlm.nih.gov/books/NBK279690/>

Cuantificación de RNA-Seq con Salmon

El pipeline para hacer la cuantificación del RNA-Seq con Salmon se da en dos etapas: Primero se indiza el transcriptoma y luego se mapean los *reads* al índice.

1. Index del transcriptoma:

```
>salmon index -t <ensamble.fasta> -i <output>
```

Donde:

-t: Archivo .fasta que contiene el ensamble de transcritos.

-i: nombre del índice que se va a crear.

2. Mapeo de reads al index:

```
>salmon quant -i index -l IU -1 lib_1.fq -2 lib_1_2.fq -o out
```

Donde:

-i: Índice del ensamble.

-l: Tipo de librería.

-1: *Reads* en formato fastq directo.

-2: *Reads* en formato fastq reverso.

-o: Archivo de salida.

Más información en: <https://salmon.readthedocs.io/en/latest/>

Anexo 4. Protocolo de microdilución

1.- Preparar medio Muller-Hinton (MH) caldo, la cantidad suficiente para la cantidad de muestras a ensayar, esterilizar.

2.- La preparación de los antibióticos se realiza a una concentración 10 veces mayor a la de trabajo a ensayar: 512, 256, 128, 64, 32, 16, 8, 4, 2, 1, 0.5, 0.25, 0.125, 0.0625, etc. microgramos / mililitro en el caso de **antibióticos**,

Por ejemplo: Para una concentración de inicio 512 $\mu\text{g/ml}$, se prepara un stock de 5120 $\mu\text{g/ml}$, de acuerdo con las especificaciones del CLSI.

1 ml de MH + 9 ml de stock (5120) = 512 $\mu\text{g/ml}$

2 ml de MH + 2 ml de dilución 512 $\mu\text{g/ml}$ = 256 $\mu\text{g/ml}$

3 ml de MH + 1 ml de dilución 512 $\mu\text{g/ml}$ = 128 $\mu\text{g/ml}$

7 ml de MH + 1 ml de dilución 512 $\mu\text{g/ml}$ = 64 $\mu\text{g/ml}$

Y se continua hasta la dilución a ensayar.

3.- En cajas de Elisa de fondo redondo con tapa de 96 pozos, se coloca en cada columna una dilución a ensayar de las anteriores diluciones de 50 μl , dejando la primera o última columna sin MH con antibiótico, en una de estas dos hileras solo se coloca MH sin antibiótico 50 μl , como control de las bacterias a probar.

4.- Se prepara el inóculo de las bacterias a una concentración de 0.5 de McFarland (en espectrofotómetro a 625 nm, a una observancia de 0.08 -0.135) en MH.

5.- Del inóculo anterior realizar una dilución 1:100 en caldo MH y colocar en la caja de Elisa que contiene las diluciones del antibiótico y el blanco 50 μl de la dilución 1:100. Todo el procedimiento se realiza por duplicado.

6.- Incubar a 35°C durante 16-18 horas.

7.- Leer las cajas.

En el caso de PEPTIDOS

1.- Prepara medio MH caldo, esterilizar

2.- Preparar el péptido a una concentración milimolar (mM) que será el stock.

3.- Los ensayos se realizan a concentración micromolar (μM), hacer los cálculos respectivos,

Si iniciamos con una concentración 64 μM

En una caja de Elisa de 96 pozos, colocar en los pozos de la primera columna 190 μ l de la concentración de 64 μ M, dependiendo de los péptidos a probar

De la segunda columna hasta la doceava se colocan 95 μ l de MH.

Ya con todos los pozos llenos, del primer pozo con 190 μ l se toman 95 μ l y mezclarlos con los 95 de MH del pozo siguiente con lo que se obtiene una concentración de 32 μ M, realizar el procedimiento anterior hasta la dilución final a ensayar descartando 95 μ l de la última dilución, todos los pozos contendrán solamente 95 μ l, tomando en cuenta que el doceavo pozo no colocarle dilución de péptido, por lo que será el control bacteriano.

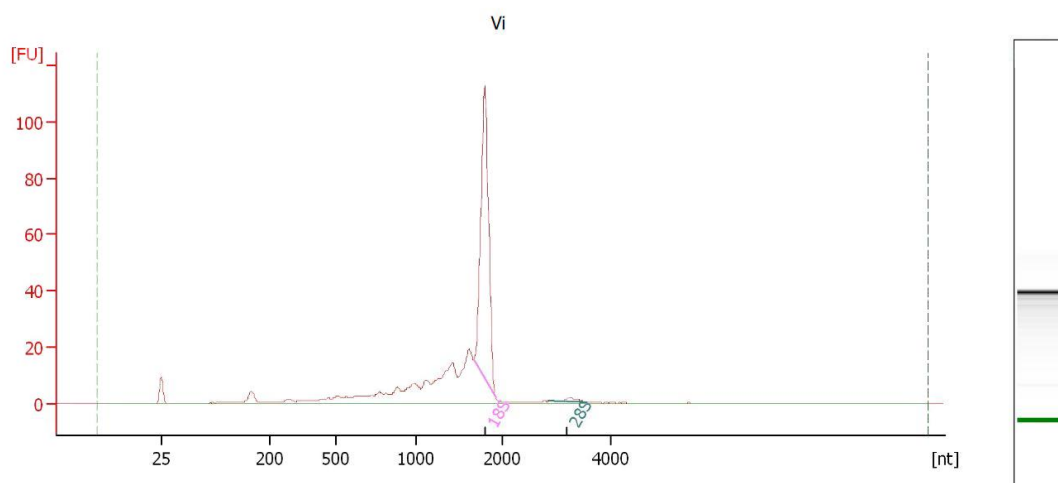
4.- Se prepara el inóculo de las bacterias a una concentración de 0.5 de McFarland (en espectrofotómetro a 625 nm, a una observancia de 0.08 -0.135) en MH.

5.- Del inóculo anterior realizar una dilución 1:100 en caldo MH y colocar en la caja de Elisa que contiene las diluciones del péptido y el blanco, 5 μ l de esta dilución bacteriana. Todo el procedimiento se realiza por duplicado.

6.- Incubar a 35°C durante 16-18 horas.

7.- Leer las cajas.

Anexo 5. Cuantificación y medición de integridad de RNA por Bioanalyzer.



Overall Results for sample 3 : Vi

RNA Area: 335.0
RNA Concentration: 212 ng/ μ l
rRNA Ratio [28s / 18s]: 0.0

RNA Integrity Number (RIN): N/A (B.02.08)
Result Flagging Color:
Result Flagging Label: RIN N/A

Anexo 6. Gráficos de calidad de la secuenciación.

La calidad phred refleja directamente la confiabilidad de la secuenciación a través de cada una de las bases de los *reads*. Este indicador está basado el código ASCII y su valor indica la probabilidad de que una base específica sea asignada a una posición en las lecturas. Una calida phred de 0 a 20 tiene una probabilidad de 99% de que una base sea asignada correctamente a una posición en la lectura (en la figura 16 es el área coloreada de rojo); si el valor de calidad está entre 21 y 30, la probabilidad sube a 99.99% y si está en el área verde, arriba de 28, se tiene una certeza de la secuenciación de un 99.999%.

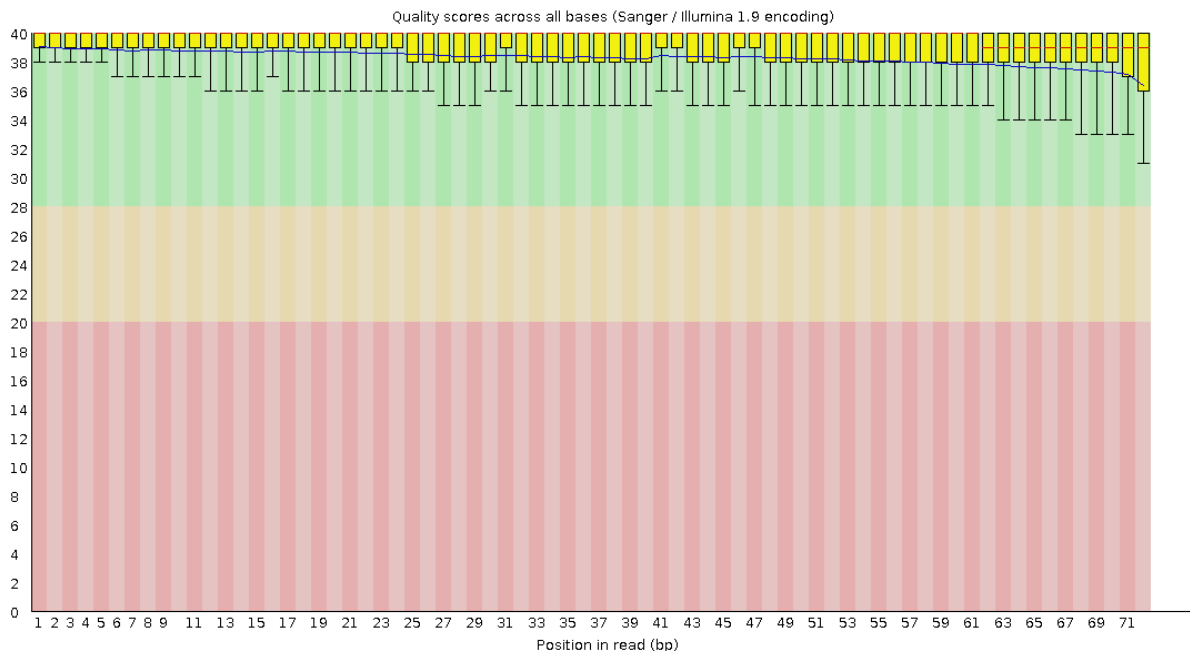


Figura16 Gráfico de caja y bigotes de la calidad phred acumulada resultado de la secuenciación del cDNA en un formato de 2x72 pb. En el eje de las x aparece cada una de las 72 posiciones que conforman una lectura y en el eje de las y aparece la es la calidad phred asociada. La línea azul es la medida total de calidad por base. También se representan la mediana y los cuartiles.

El programa fastQC da un panorama global de la calidad de la secuenciación. En la figura 17 se muestra el contenido de GC (guanina-citosina) de la muestra secuenciada. La línea azul corresponde al contenido de GC teórico calculado a partir de la secuenciación misma y la línea roja muestra el porcentaje experimental. Si se observa que el contenido GC teórico es diferente que el experimental, podría indicar que existe una contaminación por alguna secuencia sobrerrepresentada o que hubo un error sistemático durante la secuenciación.

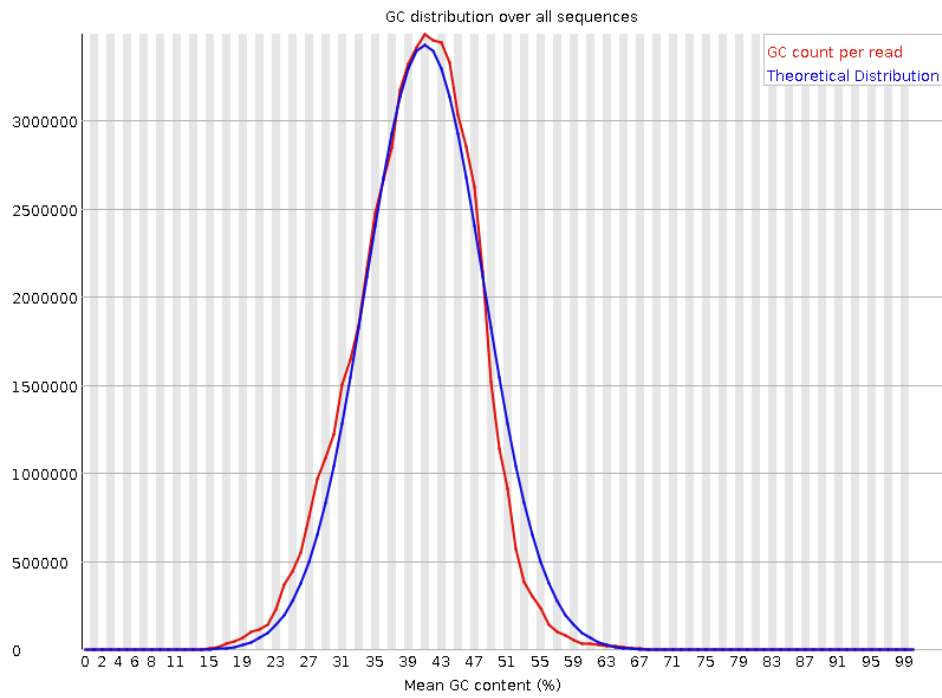


Figura 17 Contenido de GC de la secuenciación. La línea azul indica la distribución teórica del contenido de GC, la línea roja indica el contenido de GC experimental.

Anexo 7. Cuantificación de RNA-Seq

Identificador	Longitud (nt)	Longitud efectiva	TPM	Número de reads	Clasificación	Sub familia
ViCT7	275	132.36	14581.9	149773	HDP	NDBP-4
TatHDPND402	470	327.195	7356.36	186781	HDP	NDBP-4
TatHDPND401	598	455.163	6955.35	245669	HDP	NDBP-4
TatHDPND403	495	352.189	6367.39	174021	HDP	NDBP-4
TatHDPND201	487	344.191	6136.84	163911	HDP	NDBP-2
ViSep12	728	585.14	6122.72	278015	DBP	Escorpina
TatKTxAlp07	546	403.18	5818.71	182050	DBP	Toxina de potasio
ViVlp1	607	464.161	5438.79	195900	HDP	NDBP-2
ViCT2	1791	1648.07	3815.24	487936	HDP	NDBP-4
TatHDPAni01	1209	1066.07	3731.18	308672	HDP	HDP aniónico
TatHDPND405	315	172.303	3210.29	42924.2	HDP	NDBP-2
TatNaTAlp03	633	490.154	2883.55	109679	DBP	Toxina de sodio
TatKTxAlp08	546	403.18	2465.73	77145.2	DBP	Toxina de potasio
TatEnzPA215	1098	955.073	1978.07	146603	enzima	Fosfolipasa
TaVeCOth07	977	834.096	1853.16	119948	Componente de veneno	Componente de veneno
TaVeCOth06	546	403.18	1385.63	43352	Componente de veneno	Componente de veneno
TatHDPND404	945	802.098	1255.12	78122.7	HDP	NDBP-4
TatEnzPA201	1440	1297.07	1238.34	124643	enzima	Fosfolipasa
TatKTxAlp10	685	542.142	1160.96	48842	DBP	Toxina de potasio
ViLa1lp1	1348	1205.07	1153.65	107883	La1	La1
TatCaTOME01	749	606.136	1106.59	52050	DBP	Toxina de calcio
TatHDPND302	1406	1263.07	1100.05	107821	HDP	NDBP-3
TatOthLa101	1428	1285.07	1063.42	106047	La1	La1
TatKTxScr02	1748	1605.07	1023.76	127514	DBP	Escorpina
TatEnzPA213	946	803.098	929.385	57920	enzima	Fosfolipasa
TatKTxKap01	651	508.15	902.12	35573	DBP	Toxina de potasio
TatOthCRI07	1483	1340.07	851.108	88507	CRISP	CRISP
TatKTxAlp03	470	327.195	816.448	20730	DBP	Toxina de potasio
TaVeCOth05	901	758.104	576.04	33888	Componente de veneno	Componente de veneno
TatEnzHya01	2142	1999.07	546.597	84793	enzima	Hialuronidasa
TatKTxAlp09	273	130.364	487.035	4927	DBP	Toxina de potasio
TatKTxAlp04	814	671.123	482.685	25138	DBP	Toxina de potasio
TatKTxAlp12	483	340.191	452.253	11939	DBP	Toxina de potasio
TatHDPND301	449	306.202	429.523	10206.1	HDP	NDBP-3
TatNaTAlp02	977	834.096	339.399	21968	DBP	Toxina de sodio
TatKTxAlp06	1221	1078.07	326.205	27290	DBP	Toxina de potasio
TatKTxScr01	2085	1942.07	306	46116	DBP	Escorpina
TatHDPDef02	1359	1216.07	292.737	27625	HDP	Defensina
TaVeCOth01	2485	2342.07	289.833	52676	Componente de veneno	Componente de veneno
TatOthCRI06	1862	1719.07	253.529	33821	CRISP	CRISP
TaVeCOth08	636	493.153	234.431	8971.42	Componente de veneno	Componente de veneno
TatKTxAlp15	390	247.236	234.394	4497	DBP	Toxina de potasio
TatCaTLio03	367	224.258	221.876	3861.2	DBP	Toxina de calcio
TatEnzPA203	976	833.096	210.541	13611.2	enzima	Fosfolipasa

TatCaTCle01	2591	2448.07	201.573	38293.2	DBP	Toxina de calcio
TatKTxAlp11	156	32.1211	117.949	294	DBP	Toxina de potasio
TatKTxDel01	651	508.15	106.561	4202	DBP	Toxina de potasio
TatNaTBet01	463	320.198	74.3126	1846.48	DBP	Toxina de sodio
TatHDPDef03	171	41.809	70.5832	229	HDP	Defensina
TatEnzSeP08	2058	1915.07	64.5176	9588	Inhibidor de proteasa	Inhibidor de proteasa
TatEnzMtp13	1049	906.073	59.3117	4170.31	enzima	Metaloproteinasa
TatEnzSeP06	1230	1087.07	58.9516	4973	enzima	serin proteasa
TatKTxAlp13	566	423.175	44.1249	1449	DBP	Toxina de potasio
TatEnzPA206	1214	1071.07	43.8184	3642	enzima	Fosfolipasa
TatEnzMtp12	1294	1151.07	38.6683	3454	enzima	Metaloproteinasa
TatPInSrp02	2186	2043.07	37.6424	5967.95	Inhibidor de proteasa	Inhibidor de proteasa
TatKTxAlp05	361	218.262	37.1372	629	DBP	Toxina de potasio
TatPInSrp12	1518	1375.07	35.5448	3792.86	Inhibidor de proteasa	Inhibidor de proteasa
TatCaTLio01	345	202.273	33.5715	526.956	DBP	Toxina de calcio
TatKTxAlp14	329	186.289	29.6761	429	DBP	Toxina de potasio
TatOthLa106	722	579.14	28.9463	1300.89	La1	La1
TatCaTLio02	325	182.293	28.4109	401.901	DBP	Toxina de calcio
TatPInSrp16	1753	1610.07	28.2421	3528.64	Inhibidor de proteasa	Inhibidor de proteasa
TatEnzPA202	901	758.104	26.4125	1553.82	DBP	Toxina de potasio
TatPInSrp13	1303	1160.07	21.1146	1900.79	Inhibidor de proteasa	Inhibidor de proteasa
TatEnzSeP12	1664	1521.07	20.2786	2393.6	enzima	serin proteasa
TatEnzPA211	1123	980.073	16.8804	1283.82	enzima	Fosfolipasa
TatEnzPA210	1239	1096.07	14.6447	1245.62	enzima	Fosfolipasa
TatEnzMtp05	1701	1558.07	14.4574	1748	enzima	Metaloproteinasa
TatEnzPA212	1064	921.073	13.3357	953.176	enzima	Fosfolipasa
TatPInSrp17	1442	1299.07	12.8532	1295.71	Inhibidor de proteasa	Inhibidor de proteasa
TatPInSrp05	1357	1214.07	12.0684	1137	Inhibidor de proteasa	Inhibidor de proteasa
TatPInSrp04	1424	1281.07	11.5982	1153	Inhibidor de proteasa	Inhibidor de proteasa
TatEnzPA209	895	752.105	11.3171	660.51	enzima	Fosfolipasa
TatOthCRI05	1967	1824.07	11.1	1571.2	CRISP	CRISP
TatPInSrp18	1863	1720.07	10.2938	1374	Inhibidor de proteasa	Inhibidor de proteasa
TatOthCRI04	1551	1408.07	9.84109	1075.31	CRISP	CRISP
TatOthCRI02	1159	1016.07	8.76372	691	CRISP	CRISP
TatPInSrp08	1443	1300.07	8.75243	883	Inhibidor de proteasa	Inhibidor de proteasa
TatOthCRI01	3088	2945.07	8.20122	1874.3	CRISP	CRISP
TatPInSrp15	1601	1458.07	8.18403	926	Inhibidor de proteasa	Inhibidor de proteasa
TatOthCRI03	1310	1167.07	7.30963	662	CRISP	CRISP
TatEnzPA214	1465	1322.07	7.28116	747	enzima	Fosfolipasa
TatEnzPhB01	1870	1727.07	7.08926	950.115	enzima	Fosfolipasa
TatNaTBet05	433	290.208	6.88268	155	DBP	Toxina de sodio
TatEnzMtp02	3210	3067.07	6.74909	1606.32	enzima	Metaloproteinasa

TaVeCOth03	579	436.171	6.6771	226	Componente de veneno	Componente de veneno
TatNaTBet02	470	327.195	6.5583	166.518	DBP	Toxina de sodio
TatEnzSeP14	1453	1310.07	6.48225	659	enzima	serin proteasa
TatPInSrp03	1972	1829.07	6.19817	879.75	Inhibidor de proteasa	Inhibidor de proteasa
TatEnzSeP11	950	807.098	6.06727	380	enzima	serin proteasa
TatEnzSeP13	3029	2886.07	4.80888	1077	enzima	serin proteasa
TatEnzSeP01	1431	1288.07	4.71443	471.231	enzima	serin proteasa
TatHDPDef04	155	31.5159	4.49778	11	HDP	Defensina
TatOthLa105	622	479.158	4.49421	167.108	La1	La1
TatEnzMtp14	1577	1434.07	4.32477	481.281	enzima	Metaloproteinasa
TatPInSrp14	2241	2098.07	4.09439	666.614	Inhibidor de proteasa	Inhibidor de proteasa
TatPInSrp07	1553	1410.07	3.8826	424.843	Inhibidor de proteasa	Inhibidor de proteasa
TatEnzPA205	890	747.106	3.66577	212.526	enzima	Fosfolipasa
TaVeCOth04	636	493.153	3.62306	138.651	Componente de veneno	Componente de veneno
TatPInKun03	392	249.235	3.6193	70	Inhibidor de proteasa	Inhibidor de proteasa*
TatPInKun02	2990	2847.07	3.5893	793	DBP	Toxina de potasio
TatHDPDef01	147	26.997	3.34132	7	HDP	Defensina
TatHDPWap01	940	797.098	3.21719	199	HDP	Waprina
TatOthLa104	402	259.227	3.13181	63	La1	La1
TatEnzMtp06	3762	3619.07	3.00525	844	enzima	Metaloproteinasa
TatOthCRI08	1105	962.073	2.88663	215.508	CRISP	CRISP
TatKTxDel02	361	218.262	2.59783	44	Inhibidor de proteasa	Inhibidor de proteasa*
TatOthLa103	477	334.192	2.58354	67	La1	La1
TatPInKun01	746	603.136	2.5639	120	Inhibidor de proteasa	Inhibidor de proteasa*
TatPInSrp10	1374	1231.07	2.51436	240.202	Inhibidor de proteasa	Inhibidor de proteasa
TaVeCOth02	433	290.208	2.23771	50.3938	Componente de veneno	Componente de veneno
TatEnzSeP09	1167	1024.07	2.22482	176.803	enzima	serin proteasa
TatEnzPhD03	2013	1870.07	2.22453	322.821	enzima	Fosfolipasa
TatPInKun04	1014	871.073	2.07028	139.942	Inhibidor de proteasa	Inhibidor de proteasa*
TatKTxAlp02	318	175.3	1.98481	27	DBP	Toxina de potasio
TatEnzSeP18	962	819.097	1.96657	125	enzima	serin proteasa
TatEnzSeP10	1479	1336.07	1.91937	199	enzima	serin proteasa
TatEnzMtp10	1944	1801.07	1.86743	261	enzima	Metaloproteinasa
TatEnzPA204	787	644.131	1.81003	90.4741	enzima	Fosfolipasa
TatPInSrp19	647	504.151	1.73814	68	Inhibidor de proteasa	Inhibidor de proteasa
TatNaTAlp01	166	38.4588	1.67537	5	DBP	Toxina de sodio
TatEnzMtp08	3029	2886.07	1.48095	331.675	enzima	Metaloproteinasa
TatKTxAlp01	284	141.34	1.45878	16	DBP	Toxina de potasio
TatEnzHya02	1539	1396.07	1.33843	145	enzima	Hialuronidasa
TatEnzSeP07	1458	1315.07	1.32288	135	enzima	serin proteasa
TatEnzSeP16	1000	857.073	1.30809	87	enzima	serin proteasa

TatNaTBet09	376	233.249	1.28405	23.2416	DBP	Toxina de potasio
TatCaTLio04	336	193.281	1.26753	19.0112	DBP	Toxina de calcio
TatOthLa102	160	34.5801	1.11797	3	La1	La1
TatEnzSeP05	2781	2638.07	1.10885	227	DBP	serin proteasa
TatPInSrp06	689	546.142	1.09698	46.4908	Inhibidor de proteasa	Inhibidor de proteasa
TatEnzPA216	1312	1169.07	1.04717	95	enzima	Fosfolipasa
TatPInKun05	618	475.159	0.976336	36	DBP	Toxina de potasio
TatNaTBet04	146	26.4599	0.974043	2	DBP	Toxina de sodio
TatEnzPhD02	1395	1252.07	0.891922	86.6605	enzima	Fosfolipasa
TatNaTBet10	263	120.397	0.856268	8	DBP	Toxina de potasio
TatEnzSeP15	1559	1416.07	0.828116	91	enzima	serin proteasa
TatNaTBet06	256	113.462	0.79503	7	DBP	Toxina de sodio
TatEnzSeP02	591	448.165	0.776358	27	enzima	serin proteasa
TatEnzSeP03	409	266.222	0.726076	15	enzima	serin proteasa
TatEnzMtp09	1821	1678.07	0.698821	91	enzima	Metaloproteinasa
TatEnzSeP17	799	656.128	0.687409	35	enzima	serin proteasa
TatEnzPhD04	1543	1400.07	0.666898	72.456	enzima	Fosfolipasa
TatEnzPA207	2496	2353.07	0.624455	114.025	enzima	Fosfolipasa
TatNaTBet03	371	228.255	0.621024	11	DBP	Toxina de sodio
TatEnzSeP04	626	483.156	0.613446	23	enzima	serin proteasa
TatEnzMtp04	473	330.193	0.546381	14	enzima	Metaloproteinasa
TatEnzPA208	2191	2048.07	0.510851	81.1903	enzima	Fosfolipasa
TatEnzMtp11	432	289.209	0.490136	11	enzima	Metaloproteinasa
TatEnzMtp01	604	461.161	0.475042	17	enzima	Metaloproteinasa
TatNaTBet07	601	458.162	0.420627	14.9548	DBP	Toxina de sodio
TatEnzMtp07	2521	2378.07	0.399785	73.7762	enzima	Metaloproteinasa
TatEnzMtp03	2677	2534.07	0.358723	70.5412	enzima	Metaloproteinasa
TatEnzPhD01	1011	868.073	0.338089	22.7747	enzima	Fosfolipasa
TatPInSrp11	1885	1742.07	0.318866	43.1061	Inhibidor de proteasa	Inhibidor de proteasa
TatPInSrp01	642	499.153	0.193864	7.50921	Inhibidor de proteasa	Inhibidor de proteasa
TatPInSrp09	3441	3298.07	8.57E-05	0.0219403	Inhibidor de proteasa	Inhibidor de proteasa
TatCaTCle02	2073	1930.07	0	0	DBP	Toxina de calcio

Anexo 8. Secuencias LC-MS/MS

Tabla 9. Secuencias identificadas con el algoritmo Sequest del software Proteome Discoverer.

Identificador del transcriptoma	Secuencias	Xcorr	
comp8310_c0_seq1	KSYDLPLDDSTAPDDYMIGEPGVK	4.34	
comp32030_c1_seq1	DLDAPSLYHVPADISYAR	4.67	
	AMMSLGISKPWPEALK	4.1	
comp32030_c2_seq1	LQEEYDKFITDIVK	4.91	
	TADLSAWTESYLR	3.87	
	SIGTPELDAAFWR	3.46	
	YNGPIPEHLIGLIPQK	3.16	
	ATSNLHYPFQGAANPGFLVAIGR	5.48	
	IINPFENDFDVSİK	4.26	
	SVGLFDES YGDDRDİSVLLR	3.84	
comp33161_c0_seq1	QİWSFDWQHFSNPELR	3.99	
	SKPWPEVLSLLTSGR	3.13	
	IALENİPIİFSSLSmEK	4.8	
	YnGPIPEHLİGLİPQK	3.9	
	LİGAADİHWLK	3.23	
	TEENFDPASSMHLANGLQLVR	4.17	
	FFTLPSLLQTAK	3.39	
comp33725_c0_seq1	AMMSLGISKPWPEALK	4.1	
	ASQYDWTFSFDPLIR	4.08	
	VDELFrQVLPLYK	4.36	
	KGANEGFHEAİGDİALSVFİPTHWK	5.67	
	GANEGFHEAİGDİALSVFİPTHWK	7.32	
comp33936_c0_seq1	SVYEVDNLSNRVDELFNQVLPLYK	5.37	
	TatCaTCIc01	ADcLAHLK	2.19
	comp32319_c0_seq1	LGNPLLAAGİİLEVK	4.36
		EYGFĐENTWQNİMFVMKVR	1.4
comp881_c0_seq1	TPLTNYDNNİQNIcLPK	5.71	
	RLVVHPGYDADYNADİALİEFK	7.22	
	QGELPwMIQLQVSK	4.69	
	LVVHPGYDADYNADİALİEFK	6.64	
TatHDPND201	AWNSDIGK	3.03	
	GIWSTIK	2.65	
	GIWSTİKK	2.02	
TatHDPND301	LPFFLLSLVPTAISAIKK	5.94	
	KAWNSGVGK	3.27	
ViVlp1	GİFSWVK	2.5	
	AWNSGVGK	2.1	
ViAMP1	AİPSLİGGNK	3.01	
	FWGFLGK	2.79	
TatHDPND401	GİİDTVGK	2.31	
ViCT2	FWGAVWNAAK	2.06	
TatEnzHya01	WcYYLFPDcYNYFGK	4.84	

	ILVNNGESFIGDKIALIYENGIGKYPHIDPK	4.26
	IYPYINYMVHVSR	4.22
	cTWPNDPFTSWK	3.38
	YLTDINSDFDSK	2.29
comp15335_c0_seq1	ELGAGALADFIK	3.27
comp30560_c0_seq1	MPQIEIIQSLcK	4.02
	TIDTIPLPDDPGSTSLK	3.44
comp31101_c0_seq1	TFSGQVFNADTADILFR	6.03
	TLSTAVEGPLKDAIGLALR	5.01
comp30730_c0_seq1	ATDNLLSVVcTLYK	4.28
	YFHQGVDAITPLIPVPPVGGVVNK	7.67
	EKYFHQGVDAITPLIPVPPVGGVVNK	6.73
TatKTxScr02	YFHQGVDAITPLIPVPPVGGVVNKVAK	4.51
	IQSLcAFNK	1.96
	KcQETEKVK	1.48
	GVDALTNLIPAPVVGGIINK	5.76
TatKTxScr01	VQELcAFNKDVMGWcDK	3.94
ViLa1lp1	NVPGPVDAPFPDccPTSLcR	3.87
	FAAAAPGTPFPNccPMVlcK	4.67
TatOthLa101	YVLQTLScATQTLK	5.89
	NKATIDAVLGVVEYNKDLR	5.96
	ATIDAVLGVVEYNKDLR	4.9
comp34524_c0_seq1	NKWPEIVSFSVSSVNAFYIPHQNNFVLPR	6.26
	SILHPPIYDENNPNYLSFGSIGAVIGHEITHGFDSEGR	5.72
TatEnzMtp06	LKGcFTEPQDAIcGNEVVEK	1.61
	LVFTDEIEcLK	3.08
	AEFPIIccNLDVSR	4.33
	KVGIIGYTPDmFLSR	4.61
	VGIIGYTPDmFLSR	4.56
	AEKLVFTDEIEcLKDAVK	4.89
	SLKGSVDIIIALGHSGFPK	5.2
	VGIIGYTPDmFLSR	4.94
	GSGVDIIIALGHSGFPK	4.79
comp32637_c0_seq1	LVFTDEIEcLKDAVK	4.63
	EKYPNSLFLSAGDYQGTfM YTLHK	7.05
	LGHdVmAIGNHELDDGVAGLVPLIEK	7.12
	QYGWTSAAISIWNSSGIR	5.6
	IVVPAYILGGGDGITVFKEK	4.78
	TLVVQDYTFGKYIGFLK	4.29
	IVVPAYILGGGDGITVFK	5.39
	LGHdVMAIGNHELDDGVAGLVPLIEK	7.05
	AVTVFNTGVLDSdVIQTYLSR	5.81
	mEEcNLGNmLADALVDYFTDSPK	4.34
comp26928_c1_seq1	EGYNWGTcVGLLEGLcWHQK	4.98
comp27809_c1_seq1	IVSTcEGcAPIILR	4.05
	QAYEATSDLEILcYSDR	6.08
comp30392_c0_seq1	TLGcWFTYAFSYNPTADIPTKTEGQK	6.77
comp32982_c0_seq3	HcAEDIMPGFLGLYVK	4.95
comp43100_c0_seq1	cVDGcPEVPNGVIETALVGTGK	5.25

comp31198_c0_seq1	NLVVESFLISDVLK	5.3
	ALcANPTAEYTGESGFAK	5.74
TatEnzPA201	TIWGTK	1.68
	VVEGTSWYDYLATLGLIK	6.67
	cLDKVVEGTSWYDYLATLGLIK	5.66
TatEnzPA213	WTYFSAYSPK	2.95
	AFDScLGDAADKEYLWNSMR	5.85
	YGLTNTGSFTLLNcDcENTFDR	7.49
TatEnzPA215	WTYFTAYSPK	3.08
	FTMIYPGTK	3.22
	WKENYKL	2.31
	cLSDAAEKEGWFK	3.18
	SKYGLTNTGSFTLLNcDcENTFDR	7.98
TatEnzPA202	NYYFNILHVQcYR	5.16
comp20627_c0_seq1	KASEILSK	2.82
TatOthCRI06	MHSLQcVEQPDcDLcHQIGYFPVEQNFAVK	4.59
TatOthCRI07	mSAVSFGFSAPGIIFGDPIK	5.75
comp30427_c0_seq1	WLKLITInKAVK	2.85

Anexo 9. Artículo de titulación

El artículo publicado en la revista Toxins con el nombre de “A Deeper Examination of Thorellius atrox Scorpion Venom Components with Omic Technologies” puede ser consultado en la siguiente url: <http://www.mdpi.com/2072-6651/9/12/399/htm>.