

Maestría en Bioinformática y Biología de Sistemas

**Genómica, transcriptómica y proteómica
de neuropéptidos de *Oncopeltus fasciatus*
(Hemiptera: Lygaeidae)**

Trabajo final del estudiante:

Lic. Agustin Ariel Baricalla

Director:

Dr. Rolando Rivera Pomar

Este trabajo ha sido presentado como requisito para la obtención del título
de:

Magister en Bioinformática y Biología de Sistemas

Instituto de Posgrado - U.N.N.O.B.A.

Maestría interinstitucional

U.N.N.O.B.A. – U.N.Q.

Pergamino, diciembre 2020

“Nada en Biología tiene sentido si no es a la luz de la Evolución”

Theodosius Dobzhansky

“Basic research may or may not lead to new technologies. Our safest bet in this gamble is to give our brightest minds the funding and freedom to dream big.”

“Individuals need space, freedom and trust to be able to develop to their full potential. Only then will they have the courage to take the risks that make discoveries possible.”

Helmut Schwartz

“The genome sequence of an organism is an information resource unlike any that biologists have previously had access to. But the value of the genome is only as good as its annotation. It is the annotation that bridges the gap from the sequence to the biology of the organism.”

Lincoln Stein

"¿Por dónde hay que empujar este país para llevarlo adelante?"

"¿Y por qué habiendo mundos más evolucionados yo tenía que nacer en este?"

Mafalda

“No dejes para mañana lo que puedes hacer hoy. ¡Desde mañana mismo empiezo!”

Felipe

Agradecimientos

A Rolando Rivera-Pomar, por acompañarme y potenciarme en este desafío que fue la maestría y por hacerme participe de todos los proyectos estamos llevando adelante.

A Centro de Bioinvestigaciones, por potenciar mi desarrollo profesional diariamente.

A Gabriela Fernández, por su enorme dedicación en la dirección de la Maestría.

A Danya Tavela, por ser la impulsora de mi formación, por recordarme lo que es la pasión por la ciencia, a nunca bajar los brazos, a entender la política como una herramienta de cambio social y a poder poner a disposición los conocimientos adquiridos a las necesidades y demandas de la gente (especialmente este año de pandemia).

A Ariel Chipman y Eduardo Callegari, por la entrega y la disposición a ayudarme a la distancia. Eternamente agradecido.

Al Instituto Nacional de Enfermedades Humanas “Dr. Julio Maiztegui”, a todo el grupo de Arbovirus, especialmente a Alejandra Morales, Cintia Fabbri y Victoria Luppo, por demostrarme permanentemente lo que la ciencia argentina es capaz de hacer y recordarme que todo lo que se hace en ciencia es, directa o indirectamente, para el pueblo argentino.

Al equipo de investigación en Maíz del que me han acogido en este año 2020, a Juliana Iglesias, Maria Belén Kitsner, Lucia Montenegro, Sofia Perdomo y Ana Ugarte, por los desafíos superados, los logros y alegrías que hemos compartido y la muestra constante de lo que la multidisciplinariedad y las relaciones interinstitucionales son capaces de llevar a cabo.

A mi familia, en especial a Iris y a Juan, mis padres, porque son los impulsores constantes de mi crecimiento profesional.

A Nicolas Andrada, por ser el mayor sostén en todos estos años. Gracias, amor.

A Eloy Colell, por brindarme todo su conocimiento siempre.

A Maria José Delgado, consultora permanente de cuestiones del desarrollo de insectos y una gran amiga.

A Cecilia Carnovale, editora principal, correctora estricta, consultora bioinformática y una gran amiga.

A Juliana Iglesias, por ser mi consultora en todos los aspectos profesionales y personales de la vida, por compartir debates apasionados de lo que es y lo que debería ser la ciencia, la formación de recursos humanos y el feminismo, por transmitirme su camino recorrido y aprender a estar atento, a disentir como estado de alerta y a reconocer el potencial de las personas. Y por ser una gran amiga.

Resumen

Oncopeltus fasciatus (Dallas, 1852, Hemiptera: Lygaeidae) es un insecto hemimetábolo que se destaca por ser uno de los nuevos modelos de evolución, desarrollo y fisiología. El genoma de *O. fasciatus* ha sido secuenciado, ensamblado y se encuentra en proceso de anotación.

En esta tesis se caracterizó el neuropeptidoma en estadios embrionarios tempranos y en el cerebro adulto a nivel genómico, transcriptómico y proteómico. Un neuropéptido se define como un péptido o proteína (hormonas proteicas) que es liberado desde el cerebro, diversas estructuras nerviosas (ganglios) u otros órganos a la hemolinfa o en un tejido objetivo para regular diversas actividades.

En primera instancia se caracterizó de manera multiómica el neuropeptidoma de cerebro adulto de *O. fasciatus*. A partir de datos de secuenciación de ARN se pudo ensamblar un transcriptoma *de novo* que dió soporte a la corrección de marcos de lectura de neuropéptidos en el genoma. Posteriormente, se estableció la expresión génica relativa de cada neuropéptido. Luego, se logró reconocer y validar por espectrometría de masas a los precursores proteicos y a los neuropéptidos activos. Finalmente se determinó que existe una correlación entre los neuropéptidos con mayores niveles de expresión a nivel transcriptómico y la cantidad de espectros identificados a nivel proteómico por espectrometría de masas.

En una segunda instancia se estudiaron embriones de *O. fasciatus* en una serie temporal de seis momentos que abarcó la primera mitad de este desarrollo. Se pudo concluir que la expresión a nivel de ARN mensajero para la mayoría de los neuropéptidos estudiados es baja y que esta expresión para cada neuropéptido resulta variable en el tiempo durante el desarrollo temprano. Dicha variabilidad en la expresión fue utilizada para clasificar a los neuropéptidos en grupos. La clasificación obtenida sugiere una separación de los neuropéptidos según su expresión materna-temprana, coincidentes con la expresión de genes *GAP* o con la segmentación del embrión, la cual es sincrónica con la formación del sistema nervioso. Finalmente se observó que en embriones tempranos las proteínas precursoras de neuropéptidos se hallan presentes en las primeras 24 horas de desarrollo embrionario aun cuando los neuropéptidos activos no fueron hallados.

Esta tesis permitió caracterizar la secuencia nucleotídica y aminoacídica de neuropéptidos de *O. fasciatus* y sus respectivos niveles de ARN mensajero en cerebro adulto, así como también establecer una línea temporal de expresión transcriptómica de estos en el desarrollo embrional temprano. Este análisis bioinformático provee datos y conclusiones que permitirán elaborar nuevas hipótesis de trabajo experimental que permitan dilucidar la participación de los neuropéptidos en el desarrollo embrionario temprano y en la regulación de procesos fisiológicos.

Índice

Tabla de contenido

Agradecimientos	2
Resumen.....	3
Índice.....	4
1.Introducción y marco teórico.....	6
1.A. Insectos como organismos de estudio.....	6
1.B. <i>Oncopeltus fasciatus</i>	7
1.C. Secuenciación de insectos y <i>Oncopeltus fasciatus</i>	9
1.D. Neuropéptidos	10
1.E. Transcriptómica, proteómica y bioinformática.....	11
2. Hipótesis y Objetivos.....	13
2.A. Hipótesis.....	13
2.B. Objetivos	13
3. Materiales y métodos	14
3.A. Colonias de insectos.....	14
3.B. Procesado y calidad de muestras secuenciadas	14
3.C. Ensamble y caracterización del transcriptoma de cerebro.....	15
3.D. Alineamiento y expresión génica diferencial del transcriptoma de embriones tempranos	19
3.E. Análisis de neuropéptidos por espectrometría de masas.....	21
Resultados	24
4. Capítulo I	25
4.A. Caracterización multiómica del neuropeptidoma de cerebro adulto de <i>Oncopeltus fasciatus</i>	25
4.B. Procesado y calidad de las muestras	27
4.C. Ensamble <i>de novo</i> y eliminación de ARN ribosómico	29
4.D. Revisión y correcciones de marcos de lectura.....	29
4.E. Alineamiento de <i>reads</i> limpias a genoma correctamente anotado.....	30
3.F. Expresión de neuropéptidos	33
4.G. Expresión peptídica en cerebro	36
4.H. Conclusiones capítulo I	38
5. Capítulo II	39
Expresión génica y proteica de neuropéptidos en embriones tempranos	39
5.A Marco histórico	39
5.B Panorama general del desarrollo de <i>Oncopeltus fasciatus</i>	40

5.C Identificación de secuencias en transcriptomas embrionarios	43
5.D Muestras de estadios embrionarios tempranos.....	44
5.E Determinación de calidad de secuenciación, filtrado y alineamiento contra el genoma anotado	45
5.F Cuantificación y expresión génica diferencial de genes codificantes de neuropéptidos.....	47
5.G Expresión peptídica en embriones	56
5.G Conclusiones capítulo II	57
6. Discusión	58
6.A Genómica	58
6.B. Transcriptómica	59
6.C Proteómica	62
6.D. Consideraciones finales	66
7. Bibliografía	67
8. Anexo.....	76
8.A Depleción de ARN ribosómico	76
8.B Ensamble <i>de novo</i> y caracterización del transcriptoma de cerebro	80
8.C Medidas de calidad del ensamble.....	81
8.D Anotación automática de transcritos	83

1.Introducción y marco teórico

La evolución, el desarrollo, la herencia y la variabilidad son temas centrales de la biología moderna y el entendimiento de los procesos inherentes a estos mueven los más diversos campos de la investigación científica actual. Desde el desarrollo de la postulación de las leyes mendelianas de la herencia, la determinación del material genético y su estructura hasta los métodos de secuenciación de nueva generación, se ha logrado redefinir y ampliar las fronteras del conocimiento (Gayon, 2016; Mendel, 1941).

Asimismo, este desarrollo exponencial de bases de datos biológicas ha generado al menos dos inconvenientes, el primero es contar con las capacidades técnicas (hardware y software) para procesar el afluente de datos originados y el segundo es el de contar con las capacidades humanas para producir, manejar, ensamblar, validar y procesar dichos datos sin perder de vista el problema a resolver o las preguntas biológicas a responder.

La Bioinformática es hoy la disciplina científica emergente que no se limita a la solución de problemas complejos usando herramientas computacionales, sino que de manera multidisciplinaria integra áreas como Biología, Bioquímica, Biotecnología, Matemática, Estadística, Computación y Tecnologías de la Información, intentando implementar una serie de procedimientos que permitan la recuperación adecuada de la información biológica que se encuentra almacenada en las bases de datos disponibles (Manzoni et al., 2018).

1.A. Insectos como organismos de estudio

Desde hace décadas se utilizan diferentes organismos para diversos estudios en una amplia variedad de campos. Los organismos modelo son aquellos que nos dan una visión particular de ciertos fenómenos biológicos, con la expectativa de que los descubrimientos realizados en él permitan comprender el funcionamiento de otros organismos (Fields y Johnston, 2005).

De entre la gran variedad de organismos disponibles para el estudio biológico los insectos han tenido un papel preponderante para entender la genética y, actualmente, la genómica siendo utilizados como modelos de estudio.

Esto se debe a la facilidad que surge de la cría controlada en laboratorios, la alta tasa de reproducción, los cortos ciclos entre las etapas embrionarias, larvales/nifales, adultas y reproductivas y su fácil manipulación en estudios a niveles fisiológicos y bioquímicos.

1.B. *Oncopeltus fasciatus*

Una primera clasificación comúnmente consensuada en el campo de la biología consiste en clasificar a los insectos en holometábolos (aquellos que sufren una metamorfosis drástica en la cual las formas juveniles difieren notablemente de la adulta) y hemimetábolos, en los que su metamorfosis no es tan pronunciada y las formas juveniles son similares a las adultas salvo por el aumento de tamaño, la aparición de las alas y el desarrollo de las gónadas (Resh y Cardé, 2009).

En particular, dentro de los insectos hemimetábolos, se encuentra el orden de los hemípteros, caracterizados por la particular división de las alas y la similitud que posee su aparato suctopírador. Este orden de insectos posee una gran variedad de alimentación y muchos son considerados como plagas agrícolas o vectores de enfermedades humanas (Panfilio y Angelini, 2018) de allí el interés de estudio.

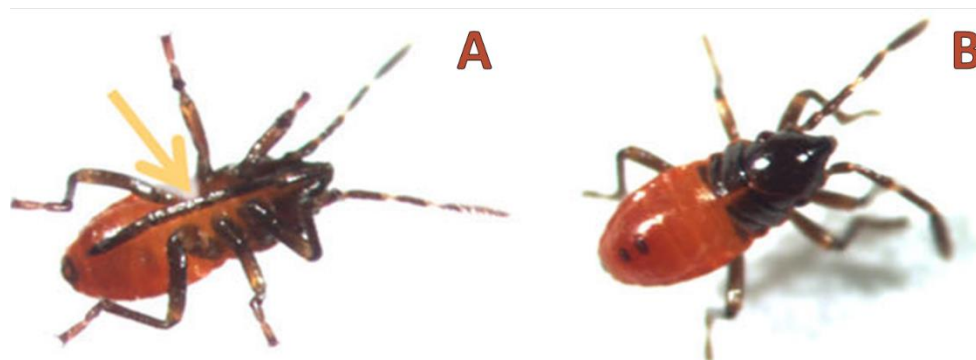


Figura 1. Ninfas en primer estadio (A: ventral – B: dorsal). La flecha marca el labio (estilete), parte de la anatomía bucal de los hemípteros adaptada para la alimentación por perforación y succión. Tomado y modificado de Panfilio et al. (2019)

Oncopeltus fasciatus (Dallas, 1852)(Hemiptera: Lygaeidae) ha sido propuesto como animal modelo (Feir, 1974; Panfilio, 2018) para el estudio científico desde 1960. En las últimas 2 décadas se lo destaca por ser uno de los nuevos modelos de evolución, desarrollo, embriología y fisiología (Chipman, 2017; Auman y Chipman, 2018).



Figura 2. A) Ejemplar macho de *O. fasciatus* sobre un pulgar. B) Ejemplar de *O. fasciatus* sobre un tubo con una colonia de *D. melanogaster* (comparación de tamaños relativos). Fotografías gentileza Rolando Rivera-Pomar.

Oncopeltus fasciatus es un hemíptero hemimetábolo que presenta en sus estadios iniciales ninfas que se asemejan mucho a la estructura del adulto. Esta característica es especialmente destacable en este trabajo ya que el cerebro y sistema nervioso crecen sin atravesar una metamorfosis completa. Este hemíptero se alimenta succionando fluidos de las semillas de plantas, siendo el algodóncillo (*Asclepias syriaca*) la especie más descrita y en donde son fácilmente localizables individuos de la especie. A su vez, *O. fasciatus* ingiere y neutraliza las sustancias tóxicas presentes en los fluidos de esta especie vegetal (Burdfield-Steel & Shuker, 2014).

Una de las principales características de *O. fasciatus* es la distintiva coloración rojo-anaranjada y negra aposemática (de advertencia) dentro de la cutícula y la epidermis que disuade a los depredadores. La misma se encuentra en todas las etapas de desarrollo (ninfales y adultas) de este insecto (Panfilio et al., 2019).

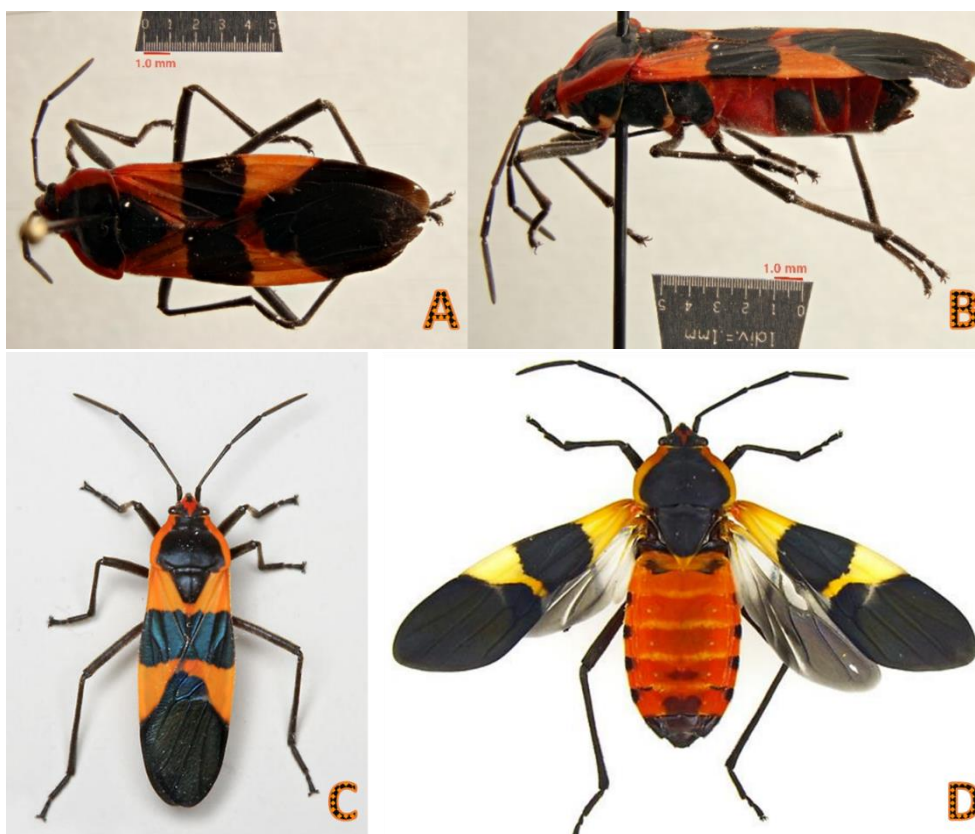


Figura 3. Fotografías de *O. fasciatus*. A y C- Dorsal. B- Lateral. D- Dorsal con alas abiertas. (Imágenes tomadas de <https://eol.org/pages/74463/media> University of California San Diego - 2017 CC-BY-SA 3.0)

La cría de este insecto en laboratorios es relativamente sencilla, siendo la humedad un factor clave en su mantenimiento. Una alimentación sencilla, a base de semillas de girasol (*Helianthus annuus*), y un mínimo de acondicionamiento con algodón humedecido facilitan su cría en cautiverio notablemente (Feir, 1974).

A nivel experimental se destaca que *O. fasciatus* presenta una alta efectividad en la interferencia de genes por la metodología ARN de interferencia (J. Liu et al., 2014) la cual permite evaluar el efecto de los genes por su pérdida de función.

Por las razones previamente enumeradas el estudio de *O. fasciatus*, contribuye ampliamente al conocimiento de su biología básica y experimental, disciplinas que posibilitan asiduamente nuevos desarrollos de herramientas para el estudio exhaustivo y comparativo con especies evolutivamente relacionadas.

1.C. Secuenciación de insectos y *Oncopeltus fasciatus*

Actualmente se desarrolla de manera colaborativa alrededor del mundo la iniciativa i5K (<https://i5k.nal.usda.gov/>), que provee lineamientos para proyectos genómicos de artrópodos con la finalidad de secuenciar, anotar y manejar de manera estándar los datos de 5000 genomas de artrópodos de importancia en la salud humana, la agricultura y el ambiente. (Evans et al., 2013)

Adicionalmente existen estudios transcriptómicos, complementarios a los genómicos, que son de interés para la salud humana como la vectores de la enfermedad de chagas *Triatoma infestans*, *Triatoma brasiliensis* (Marchant et al., 2015; Schwarz et al., 2014) y *Rhodnius prolixus* (Ons et al., 2016; Ribeiro et al., 2014).

Por otro lado, se han realizado estudios genómicos y transcriptómicos en insectos de interés agroeconómico por ser vectores de transmisión de patógenos a cultivos de producción principal (como el arroz (*Oryza sativa*) o el maíz (*Zea mays*)). Entre dichos vectores se destacan los auquenorrincos *Sogatella furcifera* (Huang et al., 2018; Liang et al., 2018), *Nilaparvata lugens* (Bao et al., 2013) y *Laodelphax striatellus* (J. H. Lee et al., 2013).

A los modelos de estudio previamente mencionados se les suman modelos clásicos en estudios biológicos y genéticos como *Drosophila melanogaster* (Brown et al., 2014), *Tribolium castaneum* (Binzer et al., 2014) y *Bombyx mori* (Yokoi et al., 2019). Piezas claves para la comprensión de innumerables procesos biológicos (Yamaguchi & Yoshida, 2018).

El laboratorio Cebio (Centro de Bioinvestigaciones, CIC-UNNOBA, Pergamino, Argentina) se encuentra involucrado en el estudio del genoma y transcriptoma de *O. fasciatus* (con especial interés en el transcriptoma embrionario), así como también en la ovogénesis del modelo de estudio *Rhodnius prolixus* (Mesquita et al., 2015; Pascual, 2019), el transcriptoma y neuropeptidoma de *Nezara viridula* (Lavore et al., 2018) y el transcriptoma de las diferentes etapas del desarrollo embrionario de *Dalbulus maidis* (Palacio, 2018).

En los últimos años se ha llevado a cabo la iniciativa i5K (Evans et al., 2013; Poelchau et al., 2015) con laboratorios distribuidos en diferentes países que han unido esfuerzos para llevar a cabo un estudio con foco en la secuenciación y anotación de diversos genomas entre los que se encuentra el genoma de *O. fasciatus* (Panfilio et al., 2018). Cada uno de ellos capacidades técnicas y humanas para secuenciar, ensamblar, anotar y,

posteriormente, analizar los datos obtenidos de la secuenciación del genoma y el transcriptoma dentro del cual se encuadra esta tesis.

1.D. Neuropeptidos

Los neuropeptidos son uno de los grupos más diverso de moléculas de señalización en los organismos multicelulares. Actúan como neurotransmisores, neuromoduladores, hormonas o factores de crecimiento, son extremadamente variados y participan en el control del crecimiento, el desarrollo, ecdisis, digestión y diuresis, entre otros. Los neuropeptidos también son cruciales para regular una gran variedad de acciones de comportamiento asociadas con la alimentación, el cortejo, el ciclo circadiano, el aprendizaje y la memoria, respuestas al estrés y las interacciones sociales (Schoofs et al., 2017).

En principio, cualquier péptido del sistema nervioso podría llamarse neuropeptido. Según la definición propuesta por Jan Veenstra (Veenstra, 2019):

“El neuropeptido se define como un péptido o proteína, que es liberado en la hemolinfa, directamente en un tejido objetivo, o dentro del sistema nervioso para regular la actividad celular por la interacción con un receptor específico de la superficie celular, por lo general a un receptor acoplado a una proteína G.”

En el trabajo de Veenstra (2009) también se destaca que un gran número de neuropeptidos han sido identificados por su actividad biológica, por la unión a receptores o por la homología con neuropeptidos previamente descriptos.

Los neuropeptidos y las hormonas peptídicas se producen a partir de grandes pre-peptidos dentro de la vía secretoria regulada (Pauls et al., 2014). Al igual que otras proteínas secretadas o unidas a la membrana, las pre-proteínas llevan un péptido de señal N-terminal que las dirige al lumen del retículo endoplásmico rugoso (Pauls et al., 2014).

Tras la eliminación del péptido de señalización, la proteína resultante puede modificarse y luego es exportada a través del aparato de Golgi para finalmente ser empaquetada y transportada a sectores de excreción (Pauls et al., 2014). Dentro de la red trans-Golgi los pro-peptidos se procesan mediante un conjunto de enzimas específicas, incluidas las furinas, convertasas y carboxipeptidasas que llevan a cabo el proceso de clivaje del propéptido a un péptido activo en sitios dibásicos, finalmente diversas enzimas proceden a otra serie de modificaciones postraduccionales entre las que se encuentran la amidación del extremo C-terminal (Pauls et al., 2014), oxidaciones y acetilaciones (Sterkel, 2011).

Han sido hallados motivos de secuencia comunes que corresponden a regiones conservadas en varias familias de péptidos, generalmente asociadas al sitio biológico activo que interactúa con su respectivo receptor (Baggerman et al., 2005). Por el contrario, las regiones peptídicas linderas a la secuencia funcional son muy poco conservadas, dificultando su reconocimiento y caracterización (Veenstra, 2019).

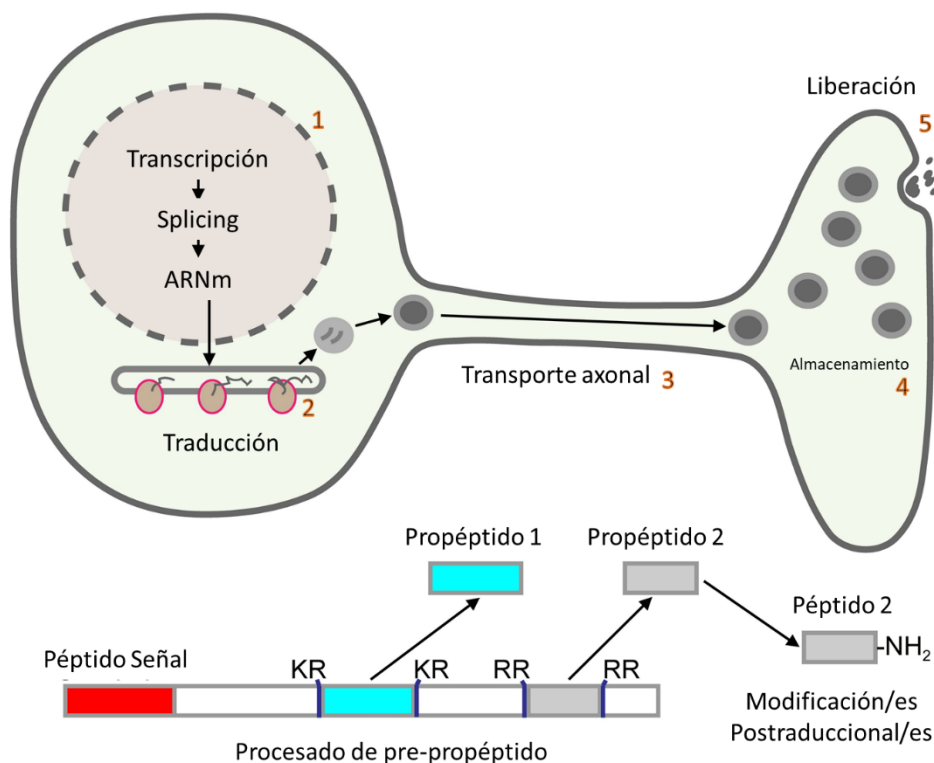


Figura 4. Imagen sintética y representativa del proceso de expresión y maduración de neuropéptidos. Se visualizan los procesos de transcripción, maduración y exportación del ARN mensajero maduro en el núcleo (1), la traducción en ribosomas del retículo endoplasmático (RE) rugoso guiada por el péptido señal (2), el procesamiento, transporte axonal y almacenamiento en vesículas densas del aparato de Golgi (3-4) para su posterior liberación extracelular (5). En la parte inferior se visualiza un pre-propéptido en proceso de maduración. Este se transloca perdiendo su péptido señal (rojo) de localización en el RE, posteriormente es clivado en sitios dibásicos de escisión (KR y RR, K=Lisina, R=Arginina), produciendo dos pro-péptidos (gris y celeste) y finalmente es modificado a péptido 2 -NH₂(péptido amidado). Tomado y modificado de Yeoh et al. (2017).

1.E. Transcriptómica, proteómica y bioinformática

En los primeros años de investigaciones genéticas, la identificación y caracterización de un único gen requería de un gran esfuerzo.

Actualmente, la secuenciación masiva del ADN abre nuevas posibilidades, acelerando los tiempos de investigación y volviendo a estos campos más robustos en la identificación, caracterización y cuantificación de la expresión génica (Amarasinghe et al., 2020; Bentley, 2006). Los métodos de secuenciación masiva, entre los que se encuentra la técnica de *RNA-seq* (secuenciación de ARN), permiten conocer la secuencia de nucleótidos a un bajo costo relativo, facilitando la obtención de grandes volúmenes de datos transcriptómicos pertenecientes a diferentes organismos y/o tejidos (Trapnell et al., 2010).

Con el devenir de la ciencia, no solo se ha hecho foco en la identificación de nucleótidos, sino que también se ha avanzado en la identificación de proteínas. Con este objetivo se ha perfeccionado la utilización de la técnica de espectrometría de masas en la identificación de péptidos y proteínas a partir de una muestra celular, de tejido o inclusive de organismos completos (Clynen et al., 2003; Schulz-Knappe et al., 2012).

La tecnología ha permitido la obtención de cantidades masivas de datos y la bioinformática, consecuentemente, contribuyó aportando las herramientas de control e interpretación de los mismos. Las numerosas técnicas desarrolladas en las últimas décadas, sumadas a la globalización consecuente del desarrollo de Internet han provisto a la comunidad científica de metodologías y estándares que permiten el procesamiento de datos a gran escala. Haciendo posible, de esta manera, un análisis integral donde se combinan diferentes tipos de datos ómicos o multiómicos, incluidas las mutaciones definidas mediante la genómica, los niveles de ARNm inferidos a través de la transcriptómica, la abundancia y el tipo de proteínas comprendidas mediante la proteómica. La combinación de estos datos permite obtener respuestas precisas a preguntas biológicas de gran envergadura (Manzoni et al., 2018).

En este trabajo, el enfoque se encuentra dirigido hacia los neuropéptidos característicos de *O. fasciatus*, determinando su composición a partir de datos peptídicos y su expresión a nivel del ARN mensajero, así como también asistir a la estructura y localización dentro del genoma de los genes que codifican dichos neuropéptidos en el proceso de anotación genómica de esta especie.

Adicionalmente se pretende obtener un flujo de trabajo mediante el cual sea posible extender esta metodología a futuros estudios en otras especies con el objetivo de comparar el sistema de neuropéptidos de insectos genéticamente emparentados a nivel de sus posibles marcos de lectura (ORF) y ARN mensajero maduro.

2. Hipótesis y Objetivos

2.A. Hipótesis

1. Es posible caracterizar e identificar los neuropéptidos de *Oncopeltus fasciatus* mediante técnicas ómicas.
2. La expresión de los neuropéptidos de *Oncopeltus fasciatus* comienza luego de la formación del sistema nervioso.

2.B. Objetivos

2.B.1 Objetivos generales

- 1 – Caracterizar el neuropeptidoma en estadios embrionarios tempranos y el cerebro adulto de *Oncopeltus fasciatus* a nivel genómico, transcriptómico y proteómico.
- 2 – Estudiar la expresión de neuropéptidos de *Oncopeltus fasciatus* durante el desarrollo embrionario temprano.

2.B.2 Objetivos específicos

1. Identificar y validar genes que codifican neuropéptidos a partir del genoma y el transcriptoma del cerebro en estadio adulto de *Oncopeltus fasciatus*.
2. Realizar las determinaciones de calidad y limpieza de datos transcriptómicos para ensamblar un transcriptoma *de novo*.
3. Ensamblar y caracterizar el transcriptoma *de novo* de cerebro adulto.
4. Establecer la presencia de neuropéptidos, su secuencia y modificaciones postraduccionales en péptidos de embriones de hasta 64 horas después de la oviposición y en cerebro de adultos.
5. Determinar la expresión génica diferencial de los genes que codifican neuropéptidos durante las primeras etapas del desarrollo embrionario.
6. Comparar la expresión de neuropéptidos de insectos genéticamente y/o evolutivamente emparentados a nivel de ARN mensajero maduro a través de datos disponibles en repositorios públicos.

3. Materiales y métodos

3.A. Colonias de insectos

Las colonias de insectos fueron mantenidas en estufa en los laboratorios del Cebio bajo un esquema 12 horas de luz: oscuridad a 25°C y humedad relativa de 70%. Los mismos fueron alimentados con semillas de girasol y con agua *ad libitum*.

3.B. Procesado y calidad de muestras secuenciadas

Los análisis correspondientes al procesado de bioinformático de muestras de los puntos 3.C y 3.D fueron llevados a cabo en el servidor “**Mendieta**” y utilizaron recursos computacionales del Centro de Cómputo de alto Desempeño (CCAD) de la Universidad Nacional de Córdoba (<https://ccad.unc.edu.ar/>), que forman parte del Sistema Nacional de Computo de Alto Rendimiento del MinCyT de la República Argentina.

3.B.1 Procesado de muestras, extracción y secuenciación por NGS

Las muestras para la secuenciación se obtuvieron a partir de un pool de 20 cerebros diseccionados de insectos adultos, compuesto de 10 cerebros de cada sexo. A las muestras se les extrajo el ARN total por método de Trizol (Thermo Fisher Scientific), se les realizó una corrida en gel de agarosa para determinar la integridad y, paralelamente, se le realizó una cuantificación mediante el fluorómetro Qubit 2.0 (Thermo Fisher Scientific) (Lavore et al., 2018). En ambos casos se siguieron las especificaciones del fabricante.

Los ARN totales extraídos fueron secuenciados en el servicio de secuenciación del “Max Planck Institute for Molecular Genetics” (Berlín, Alemania) en la plataforma Illumina HiSeq 4000. La secuenciación proveyó 210 millones de pares de *reads* crudas de 151 bases de longitud.

3.B.2 Determinación de calidad de secuenciación, filtrado y depleción de contaminantes

A partir de los datos crudos obtenidos en formato ‘fastq’ se procedió a la evaluación de calidad bioinformática de la secuenciación.

Cada archivo fastq de *reads* crudas o luego de cada paso de filtrado o recorte por criterios de calidad, fueron analizadas mediante el software FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y en cada paso se determinó:

- Cantidad de secuencias (millones de secuencias)
- Largo de las secuencias
- Calidad promedio (*Phred Score*)
- Contenido porcentual de cada base por posición
- Porcentaje de secuencias duplicadas
- Porcentaje de contenido de GC (Guanina + Citosina) en las *reads*
- Presencia de adaptadores de PCR o secuenciación
- Secuencias de *reads* sobrerrepresentadas

Posteriormente se prosiguió con la remoción de adaptadores de PCR o secuenciación mediante el uso de Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) con los valores por defecto que se resumen en la tabla presentada a continuación:

Tabla 1. Parámetros utilizados para la remoción de adaptadores de amplificación y/o secuenciación

Parámetro	Valor
Adaptadores de secuencia a ser escindidos	Detección automática
Recorta nucleótidos del extremos con baja calidad además de la eliminación de adaptadores si no superan un valore de calidad mínimo	Si Valor de calidad mínimo: 20
Eliminación de lecturas con menos de N nucleótidos	20 nucleótidos
Retención de lecturas que no posean el par (Forward + Reverse)	No, solo pares completos

Se optó por depletar *in silico* los ARN ribosómicos, debido a los altos valores de contenido de GC. Para ello se utilizó fastQScreen (Wingett & Andrews, 2018) con las opciones por defecto.

FastQScreen utilizó Bowtie2 (Langmead & Salzberg, 2012) como programa auxiliar para alinear las *reads* a diversas bases de datos de contaminantes. Se incluyeron las bases de datos con contaminantes típicos de la secuenciación NGS. Las mismas fueron la base de ARN ribosómico del “SILVA rRNA database Project” para Eucariotas, Archeas y Bacterias (Glöckner et al., 2017), los contaminantes de laboratorios de la base de UniVec (Schäffer et al., 2018) de NCBI, el genoma de *E. coli* y las secuencias PhiX correspondientes al control interno de corridas en secuenciadores Illumina (San Diego, Estados Unidos).

3.C. Ensamble y caracterización del transcriptoma de cerebro

3.C.1 Alineamiento contra el genoma anotado

Para llevar a cabo el alineamiento de las *reads* limpias al genoma se utilizó la última versión disponible del genoma y sus respectivas anotaciones. Los mismos fueron descargados de https://i5k.nal.usda.gov/Oncopeltus_fasciatus ya que el genoma de *O. fasciatus* es parte de la iniciativa i5K y presenta el mayor grado de curación disponible. Actualmente el genoma se encuentra en proceso de ensamble y posee un genoma compuesto por 17.224 scaffolds y 19.760 transcriptos/genes anotados (Panfilio et al., 2018).

Las *reads* fueron mapeadas con STAR v2.6 (Dobin et al., 2013; Dobin & Gingeras, 2015). Este software realiza un índice a partir del genoma y sus anotaciones y, a partir de ello, mapea las *reads* en los sitios declarados como marcos de lectura anotados.

El programa corre dos instancias para mejorar el mapeo. En un primer paso se realiza un alineamiento contra los marcos de lectura declarados y se anotan todos los sitios de corte y empalme alternativos que difieran de los previamente anotados (*novel splice junctions* o sitios de empalme nuevos). En la segunda instancia las *reads* no asignadas a una posición se re-mapean a todo el genoma sin considerar los marcos de lectura anotados y tomando

en consideración los sitios de corte y empalme noveles obtenidos de las lecturas mapeadas en el primer paso.

Este procedimiento de 2 pasadas con anotación de sitios de *splicing* adicionales permite identificar exones o intrones no asignados o incorrectamente asignados, dar soporte para la localización de nuevos marcos de lectura durante el proceso de anotación y realizar un curado bioinformático de los mismos.

STAR provee estadísticas mínimas del alineamiento que incluyen cantidad de *reads* correctamente mapeadas una única vez, múltiples veces o no mapeadas, así como también los números y tipos de *splicing*.

Previo al alineamiento se corroboraron de manera manual las anotaciones de cada uno de los genes que codifican neuropéptidos. Se realizó una búsqueda exhaustiva mediante tBLASTn (Altschul et al., 1990) de los neuropéptidos de *O. fasciatus* que se encuentran en Uniprot contra todo el genoma. Actualmente se disponen de 46 neuropéptidos, incluyendo secuencias de propéptidos inmaduros completos como de péptidos funcionales maduros. Los mismos se encuentran detallados en la Tabla 2.

Adicionalmente se exploró, también mediante tBLASTn (Altschul et al., 1990), la posibilidad de que existieran marcos de lectura para otros neuropéptidos no anotados previamente en el genoma de *O. fasciatus*. Para la búsqueda se utilizó la base de datos DINER (<http://neurostresspep.eu/diner/insectneuropeptides>) que contiene, al momento de la escritura de este manuscrito, 53 neuropéptidos. Respecto a la versión original de este manuscrito fueron incorporados varios genes codificantes de neuropéptidos, 2 de ellos no poseían un marco de lectura proveniente de las anotaciones del proyecto y se exponen aquí por primera vez (GPB5 y Proctolin), mientras que otros 12 fueron reasignados correctamente (Allatostatina A, B, CC y CCC, Elvenin, GPA2, Neurologina, Neuroparsina 1 a 4 y RY-amida).

Se evaluó la calidad del mapeo de los archivos de ensamble BAM de reads al genoma mediante QualiMap (Okonechnikov et al., 2016). Este programa permitió visualizar un resumen del alineamiento que incluye la región genómica a la que pertenece fuera asignado (intrónica, exónica o intergénica), un análisis de los sitios de empalme y la cobertura de las *reads* a lo largo de las posiciones del transcripto y la cobertura promedio que presenta cada transcripto del transcriptoma.

Tabla 2. Nombre de los neuropéptidos y su abreviatura correspondiente. Nota: la forma abreviada es coincidente con la forma acortada en el idioma inglés. Ciertos neuropéptidos han sido nombrados por patrones peptídicos típicamente encontradas en su secuencia (CCH, CNM, FMRF, ITG, NVP, SIF) y no son acrónimos de su nombre completo.

Neuropéptido	Abreviatura	Neuropéptido	Abreviatura
Allatostatina A	AstA	Miosupresina	MS
Allatostatina B - péptido myoinhibitorio	AstB	Neuroigin	Nlg
Allatostatina CC	AstCC	Neuroparsina (Net Platelet Angiogenic Activity)	NPAA
Allatostatina CCC	AstCCC	Neuroparsina 1	NP1
Allatotropina	AT	Neuroparsina 2	NP2
Bursicona	Burs	Neuroparsina 3	NP3
Bursicona-beta	pBurs	Neuroparsina 4	NP4
CCH-amida	CCH	Neuropéptido activador de la biosíntesis de feromonas	PBAN
CNM-amida	CNM	Neuropéptido F largo	NPF
Corazonina	Crz	NVP	NVP
Elevenin	Elv	Orcokinina	OK
Factor dispersante de pigmentos	PDF	Péptido cardioactivo de crustáceos	CCAP
Factor liberador de corticotropina - DH44	DH44	Péptido C-terminal (7B2)	c-Ter
FMRF-amida	FMRFa	Péptido relacionado con la insulina 1	ILP1
Glicoproteína alfa 2	GPA2	Péptido relacionado con la insulina 2	ILP2
Glicoproteína beta 5	GPB5	Péptido transportador de iones	ITP
Hormona protoracicotrópica	PTTH	Proctolin	Proc
Hormona Adipoquinética	AKH	Proteína tipo Neuropéptido	NPLP
Hormona de eclosión	EH	RYamida	RYa
Hormona disparadora de la ecdisis	ETH	SIF-amida	SIFa
Hormona diurética DH31	DH31	SNF - Neuropéptido F corto	sNPF
ITG	ITG	Sulfaquinina	SK
Kinina	Kin	Taquicinina	TK

3.C.2 Ensamble “de novo”

En virtud de poseer un genoma en proceso de anotación se realizó de un ensamble *de novo* que no necesita de un genoma de referencia para ensamblar un transcriptoma para indagar la posibilidad de encontrar nuevos marcos de lectura o ayudar a la corrección de aquellos previamente anotados.

Todos los datos de secuenciación fueron unidos en un único archivo que contenía todas las *reads* limpias procedentes de los pasos anteriores (150 millones de pares de *reads*). Una vez unidas y comprimidas fueron cargadas al servidor principal de Galaxy <https://usegalaxy.org/> (Hillman-Jackson et al., 2012). Este servidor fue elegido porque presenta una versión optimizada de Trinity v2.2.0 (Grabherr et al., 2011; Haas et al., 2013) para correr en el cluster “*Bridges*” del “Pittsburgh Supercomputing Center” permitiendo corridas de datos computacionalmente intensas y con una alta demanda de memoria RAM.

Las opciones utilizadas con Trinity fueron las seteadas por default en la página. Esto incluye una configuración para los datos tipo “*pair-end*” (extremos pareados) y “*unstranded*” (cuando se desconoce la dirección de la lectura respecto del transcripto que le dio origen) con una normalización de los datos previos a la ejecución del programa.

3.C.3 Caracterización del ensamble “de novo”

Con las secuencias ensambladas por Trinity (Haas et al., 2013) se procedió a caracterizarlas mediante BUSCO (Seppey et al., 2019), Trinotate (Bryant et al., 2017), TrinotateR (<https://github.com/cstubben/trinotateR>) y RNAQuast (Bushmanova et al., 2016).

Primero se realizó la determinación de secuencias ortólogas mediante BUSCO. La métrica de BUSCO es complementaria a otras métricas técnicas, como N50, y está basado en una búsqueda mediante modelos ocultos de Márkov de los ortólogos evolutiva y funcionalmente anotados previamente provistos en diversas bases de datos. Para este análisis se seleccionaron las bases de datos de ortólogos para eucariotas, artrópodos, insectos y hemípteros provenientes de OrthoDB v10 (Kriventseva et al., 2018).

Trinotate anota computacionalmente los transcriptos obtenidos de Trinity. El funcionamiento está basado en varios programas de soporte para realizar un informe detallado en forma de una base de datos. Este incluye la identificación de regiones codificantes de proteínas predichas por TransDecoder, la determinación de transcriptos de ARNs ribosómicos mediante RNAMMER (Lagesen et al., 2007), las secciones transmembrana de las proteínas por TMHMM (Krogh et al., 2001), los péptidos señal por SignalP (Almagro Armenteros et al., 2019) y los dominios proteicos de Pfam mediante HMMscan (Eddy, 2011).

Con los transcriptos nucleotídicos de Trinity (Haas et al., 2013) y las secuencias proteicas inferidas por TransDecoder (<https://github.com/TransDecoder/>) se realizó una búsqueda con BLAST contra 2 bases de datos, la base de datos Swiss-prot y la base de Uniprot para *O. fasciatus* que consta (al momento de la realización de este manuscrito) de 19760 proteínas totales, 19752 anotadas computacionalmente y 8 secuencias manualmente curadas. La búsqueda en Swiss-prot (<https://www.uniprot.org/uniprot/>) permitió extraer los términos GO (egglog) y las vías metabólicas (Kegg) en las que está involucrada la proteína.

Con las demás herramientas nombradas anteriormente se anotan automáticamente dominios proteicos o transmembrana y péptidos señal. La búsqueda contra la base específica del organismo genera nueva evidencia de la correspondencia entre un transcripto y la proteína extraída computacionalmente obtenidas en el proceso de

anotación del genoma. Con el informe de Trinotate (Bryant et al., 2017) se obtuvieron medidas de resumen en R mediante el paquete trinotateR (<https://github.com/cstubben/trinotateR>).

Por último, se alinearon las secuencias de Trinity (Haas et al., 2013) fueron alineadas al genoma anotado mediante STAR-long (Dobin & Gingeras, 2015). STAR-long es una versión optimizada de STAR enfocada en alinear segmentos de varios cientos o miles de pares de bases, a diferencia de STAR que se encuentra enfocado en lecturas de entre 50 y 250 bases. Este alineamiento permitió identificar modificaciones en la estructura de los marcos de lectura, incluyendo inserciones, deleciones, longitud del marco de lectura o modificaciones en la secuencia nucleotídica. Al igual que con el mapeo de *reads* se evaluaron las métricas de calidad por QualiMap (Okonechnikov et al., 2016) y, adicionalmente, métricas por RNAQuast (Bushmanova et al., 2016).

3.D. Alineamiento y expresión génica diferencial del transcriptoma de embriones tempranos

El primer acercamiento a los análisis transcriptómicos se basó en el ensamble del transcriptoma materno y embrionario temprano que fuera amablemente cedido para su estudio por el laboratorio de Cassandra Extavour (Universidad de Harvard).

Adicionalmente, fue llevado a cabo un intercambio de datos con el grupo del laboratorio de 'Ecología, evolución y comportamiento' a cargo de Ariel Chipman (Universidad Hebrea de Jerusalén) para realizar los análisis de expresión génica diferencial en embriones tempranos de *O. fasciatus*. El mismo consistió en el remapeo de lecturas crudas cedidas por el laboratorio del doctor Chipman a la versión más reciente del genoma de *Oncopeltus fasciatus* disponible (v1.2), la cuantificación y la expresión génica diferencial por parte de nuestro laboratorio.

3.D.1 Identificación de secuencias en transcriptomas embrionarios

Se identificaron mediante alineamiento de secuencias con BLASTx (Altschul et al., 1990) los transcriptos codificantes de neuropéptidos. Los transcriptos utilizados para la búsqueda (*query*) corresponden con los ensamblados en el trabajo de Ewen-Campen y colaboradores (2011). Fue utilizada la base de datos proteica de neuropéptidos de *O. fasciatus* procedente de Uniprot como referencia de búsqueda como se describió en la sección 3.C.3.

Las muestras para el estudio de Ewen-Campen y colaboradores (2011) constaron de ARN total obtenido de ovarios y embriones de diversos estadios de madurez. Los estadios abarcaron los 3 primeros días post - oviposición y se corresponden con, aproximadamente, el 60% de la embriogénesis a 28°C y 12 horas de luz: oscuridad. Se sintetizó cDNA con el ARN total, fue secuenciado en la plataforma de pirosecuenciación 454 Titanium GS-FLX (Roche) y ensamblados con Newbler v2.5 (Roche) con valores por defecto.

Los datos crudos del estudio nombrado en el párrafo anterior se encuentran disponibles en la plataforma SRA de NCBI (<https://www.ncbi.nlm.nih.gov/sra/SRX022014>). Estos fueron descargados, analizados y alineados al genoma anotado como se explica en el punto

3.C.1 con el fin de contrastar los hallazgos del trabajo original (ensamble *de novo*) con el avance de la anotación del genoma actualmente disponible y, adicionalmente, se cuantificó mediante Stringtie (Pertea et al., 2015) para evaluar la distribución relativa de las *reads* en el subconjunto de genes correspondientes a neuropéptidos.

3.D.2 Muestras de estadios embrionarios tempranos

Los datos cedidos para su análisis transcriptómicos fueron obtenidos como describe Ginzburg y colaboradores (2017). Brevemente, las colonias se mantuvieron a 25°C con ciclos de 14/10 horas de luz/oscuridad. Los embriones fueron colectados en ventanas de 2 horas y mantenidos en incubadora hasta cumplir las horas de desarrollo deseadas. Una vez alcanzada la etapa deseada los mismos fueron congelados, conservados a -80°C y criodiseccionado en los segmentos anterior y posterior.

Se colectaron huevos a partir de las 10, 19, 26 y 30 horas posteriores a la oviposición, diseccionados en mitades anterior y posterior y agrupados en 8 unidades muestrales de entre 10 y 15 medios embriones con 3 réplicas para cada momento.

De manera complementaria se obtuvieron 2 set de datos de la plataforma “*Sequence Read Archive*” del NCBI. El primer set de datos es de tejido ovárico (Bewick et al., 2019) (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA430099>) y el segundo se corresponde con embriones de entre 48 a 64 horas después de la oviposición (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100429>). Ambos con 3 réplicas por unidad muestral.

Se obtiene de todos los datos en conjunto una serie temporal constituida de ovarios (pre-oviposición o tiempo 0), embriones tempranos a las 10, 19, 26 y 30 (± 1 hs) horas después de la oviposición y embriones de entre 48 a 64 horas post-oviposición, estos últimos serán denominados posteriormente como grupo “50 horas” para facilitar la lectura e interpretación de resultados (esto se debe a que pasadas las 48 horas post-oviposición el embrión adopta formas consecutivas de diferenciación que contrastan notablemente con los estadios anteriores).

3.D.3 Determinación de calidad de secuenciación, filtrado y alineamiento contra el genoma anotado

Los datos obtenidos por el grupo del Dr. Chipman, incluidos en el estudio de Ginzburg et al. (2017), fueron re-mapeados a la versión actual del genoma con las anotaciones correspondientes.

Se estableció la calidad de secuencia, se eliminaron adaptadores y se filtraron de secuencias de igual manera a la descrita en la sección 3.B.2. La única modificación al protocolo anterior en el caso de las muestras de embriones de entre 10 a 30 fue el cambio de “Trim Galore!” por trimmomatic (Bolger et al., 2014) con las opciones TRAILING:3 SLIDINGWINDOW: 4:25 MINLEN:15, con resultados similares en ambos casos. Al contarse con 30 muestras que fueron analizadas por múltiples herramientas las medidas de calidad fueron resumidas y graficadas con multiQC (Ewels et al., 2016).

El alineamiento al genoma se realizó como fue descrito en la sección 3.C.1. Este proceso se realizó para las muestras de ovario y los embriones de todos los tiempos considerados. Para el caso particular de los embriones de entre 10 y 30 horas los mismos se encontraban partidos en 2 segmentos, por lo que se ensamblaron los 24 sets de datos con 2 enfoques. El primer alineamiento constó de alinear cada muestra por separado con la finalidad de comparar dentro de cada tiempo la expresión en el extremo anterior respecto al posterior. Luego se unieron bioinformáticamente las 2 mitades de cada replica en cada tiempo para obtener 12 archivos BAM de alineamiento para establecer la expresión diferencial a través de los 4 tiempos del desarrollo.

Los ensamblados generados en este trabajo mancomunado pueden hallarse de manera libre en <https://i5k.nal.usda.gov/node/933164>.

3.D.4 Cuantificación del alineamiento, normalización y expresión génica diferencial

Se llevo adelante el protocolo de expresión génica diferencial de manera local en R (R. Team, 2020) a través de RStudio (Rstudio Team, 2020). Se utilizaron las librerías edgeR (Robinson et al., 2010), Rsubread (Yang Liao et al., 2019) y ggplot2 (Wickham, 2009).

Las *reads* mapeadas dentro de los marcos de lectura en el genoma fueron cuantificadas con el módulo featureCounts de Rsubread. La normalización de los datos y la expresión génica diferencial se realizó con edgeR. Fueron removidos, en primera instancia, todos los genes que no poseían al menos una cuenta en al menos tres muestras. Luego se normalizaron los datos por muestra mediante el método TMM ("*trimmed mean of M-values*") como factor de normalización para obtener tamaños de librerías efectivos comparables entre muestras.

Se diagnosticó la variabilidad de los datos con un gráfico de "Biological coefficient of variation" (coeficiente de variación biológica) y la correcta asignación de las muestras a su respectivo grupo mediante un gráfico de Escalado multidimensional (MDS).

Se ajustó un modelo lineal generalizado de tipo binomial negativo con metodología de tipo quasi-verosimilitud ("*quasi-likelihood negative binomial model*"). Este modelo ajusta el valor de cada gen en cada muestra, controla la variación indeseada, estima la variación biológica y disminuye los falsos positivos.

Una vez modelados los ajustes para los datos se procedió con los contrastes de interés. Se determinaron los genes diferencialmente expresados mediante una prueba F *quasi-likelihood*. Se utilizaron como valores de corte al menos 2 veces la expresión de una condición versus otra condición a comparar y un p-valor corregido mediante el ajuste de Benjamini-Hochberg (FDR) de 10^{-3} .

3.E. Análisis de neuropéptidos por espectrometría de masas

Para los análisis de identificación de neuropéptidos por espectrometría de masas se partieron de dos grupos de muestras. El primer grupo se compone de 10 cabezas de machos y 10 cabezas de hembras (con 2 réplicas en cada caso) y un segundo grupo de aproximadamente 100 embriones completos de entre 0 y 24 horas.

Todos los grupos de muestras fueron analizados de igual manera siguiendo los procedimientos establecidos Ons y colaboradores (2009, 2011) y Lavore y colaboradores (2018) como se detalla en los siguientes apartados a continuación.

3.E.1 Procesado de muestras

Para procesar los cerebros el procedimiento comprende la disección manual de las cabezas y su puesta en solución salina sobre hielo inmediatamente después de la disección. El La solución salina utilizado se compone de NaCl 129 mM, KCl, 8.6 mM, CaCl₂ 2.0 mM, MgCl₂ 8.5 mM, NaHCO₃ 10.2 mM, NaH₂PO₄ 4.3 mM, HEPES 8.6 mM a pH 7.

Los grupos de muestras fueron situados en 100 µl de metanol/agua/ácido acético (90:9:1, v/v/v), sonicados por 5 minutos y centrifugados por 10 minutos a 7500g. El sobrenadante fue colectado y el pellet fue re-extraído 2 veces. Se eliminaron los solventes orgánicos al sobrenadante colectado mediante extracción en centrifuga de vacío, fue re-diluido en 20 µl de ácido trifluoroacético (TFA) al 0.1% y desalado utilizando discos de extracción C18 activados con una solución al 80% de Acetonitrilo en TFA al 0.1%.

3.E.2 Cromatografía líquida y espectrometría de masas

Los péptidos resultantes del paso 3.E.1 fueron separados por cromatografía líquida en un cromatógrafo Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) para luego ser analizada por espectrometría de masas. (Lavore et al., 2018)

La metodología elegida para determinar la composición aminoacídica fue espectrometría de masas en tándem en un espectrómetro de tipo Orbitrap. El mismo fue realizado en las instalaciones de proteómica del Instituto Max Planck de Química Biofísica (Göttigen, Alemania) mediante un equipo Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific).

3.E.3 Identificación de péptidos

Los espectros obtenidos por espectrometría de masas en formato RAW fueron procesados para obtener una lista de picos y, luego, se utilizaron para una búsqueda en bases de datos. La identificación se llevó a cabo en PEAKS X-Pro (Bioinformatics Solutions Inc., Ontario, Canadá) con los parámetros detallados en la siguiente tabla:

Tabla 3. Parámetros utilizados para la identificación de espectros de masa correspondientes a neuropéptidos.

Parámetro	Valor
Modificaciones variables	Acetilación N-terminal Amidación C-terminal Glutamina -> piroglutamato (N-term Q) Oxidación (M)
Enzima	Ninguna
Valores de masas	Monoisotópico
Tolerancia de la masa del péptido	50 ppm
Tolerancia de la masa del fragmento	0,3 Da
Máximo número de clivajes perdidos	2
Tipo de instrumento	ESI-FTICR

Se realizó un análisis manual de cada espectro para confirmar las modificaciones e identificar de manera precisa las secuencias, teniendo en cuenta las series de iones “b” e “y”.

Se confeccionaron 2 bases de datos para la identificación de neuropéptidos. La primera fue confeccionada con todas las proteínas de *O. fasciatus* disponibles en Uniprot (<https://www.uniprot.org/>) con 19760 secuencias proteicas al momento del análisis. La segunda base de datos resultó del filtrado manual de las 34 proteínas y péptidos de interés para este trabajo (neuropéptidos y hormonas proteicas).

Resultados

El presente manuscrito de tesis presenta dos secciones, una primera enfocada en la caracterización multiómica del neuropeptidoma de cerebro adulto de *O. fasciatus* y una segunda que procura describir el patrón temporal de expresión génica y la presencia de neuropéptidos y proteínas hormonales en embriones en estadios tempranos del desarrollo basadas en los hallazgos de la sección inicial, razón por la cual se ofrece una breve introducción para facilitar su lectura.

4. Capítulo I

4.A. Caracterización multiómica del neuropeptidoma de cerebro adulto de *Oncopeltus fasciatus*

Al momento de caracterizar el transcriptoma y proteoma de *O. fasciatus*, deben ser tomadas en consideración una serie de cuidados específicos relacionados a la característica más destacable de los neuropéptidos: la longitud de los péptidos activos ronda la decena de aminoácidos. Esto hace que se dificulte la detección, pero ese no es el único factor por considerar.

Jan Veenstra recogió en su publicación de 2019 una serie de recaudos y aclaraciones para la correcta caracterización de neuropéptidos, los cuales se resumen a continuación:

- Algunos neuropéptidos han sido identificados únicamente sobre la base de que se producen después de un procesamiento lítico de proteínas de función desconocida, o incluso sólo sobre la probabilidad de que sus precursores putativos pudieran ser procesados por las convertasas neuroendócrinas en neuropéptidos.
- La calidad de los genomas es variable, por lo que poseer un genoma anotado no es garantía de exhaustiva curación manual.
- En algunos casos las secuencias codificantes previstas para ciertos neuropéptidos se encuentran incompletas o mal descriptas.
- Del procesamiento postranscripcional del ARNm se desprende que no todas los transcriptos producidos por Trinity (ensamblador *de novo*) son copia fiel del ARNm de los genes. Las secuencias parecerían indicar que han sufrido el mismo empalme que el que se produce en el ARNm del neuropéptido en cuestión, pero eso no es necesariamente lo que ocurre.
- La metodología descrita permite demostrar la presencia de determinados neuropéptidos. Sin embargo, cuando un gen particular no se identifica de esta manera, no significa necesariamente su ausencia en la especie en cuestión.
- En el caso de utilizar genes ortólogos para la identificación de genes codificantes de neuropéptidos en el organismo de estudio, a menudo el péptido precursor inmaduro difiere notablemente, aunque la secuencia del neuropéptido activo se encuentre conservada, esto dificulta que el mismo pueda ser reconocido de manera fiable.
- En algunos casos la secuencia que codifica el péptido, o sus partes más conservadas, puede ser interrumpida por un intrón.
- Algunos neuropéptidos no sólo son de corta longitud, sino que también poseen secuencias de aminoácidos cuyos codones se encuentran altamente degenerados.

De la bibliografía actual y bases de datos especializadas se recogen tres publicaciones que han caracterizado a los neuropéptidos presentes en *O. fasciatus*.

En primer lugar, la publicación del genoma de *O. fasciatus* (Panfilio et al., 2019), de los genes allí caracterizados destaca como ejemplares a aquellos que codifican para péptidos que son secretados.

En segundo lugar, el grupo del laboratorio Extavour (Ewen-Campen et al., 2011), analizando el transcriptoma embrionario temprano de *O. fasciatus* con foco en los transcritos relacionados usualmente con el proceso de ecdisis (muda), logró observar que “[...]Como era de esperar en una situación en la que la ecdisona regula el desarrollo embrionario pero no la muda, no se detectaron transcritos que codificaran las hormonas peptídicas de los insectos implicadas en el comportamiento de la muda, como la hormona disparadora de la ecdisis (ETH), la hormona de la eclosión (EH) y el péptido cardioactivo de los crustáceos (CCAP)”. Esto sugeriría la hipótesis de que ciertos neuropéptidos (o posiblemente todos) estarían exclusivamente expresados en estadios embrionales tardíos, estadios ninfales o adultos, pero no en estadios embrionales tempranos.

En tercer lugar, la búsqueda en bases de datos específicas de proteínas como Uniprot, pone en evidencia que de los más de cuarenta neuropéptidos hallados bioinformáticamente en el genoma solo dos proteínas precursoras presentan evidencia a nivel de péptidos observados por espectrometría de masas (seis neuropéptidos activos derivados del precursor de taquicinina y una de alatotropina) y algunas inconsistencias a nivel de secuencia por resolver. (Neupert et al., 2009)

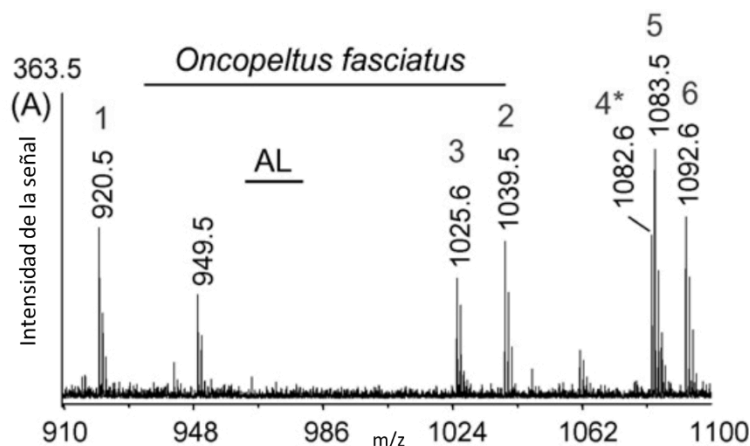


Figura 5. Espectro MS1 de una mezcla de las 6 variantes de Taquicinina en *O. fasciatus*. Cada número marca una variante. El asterisco en la variante 4 indica que las señales de Taquicinina-4 y Taquicinina-6 son indistinguibles entre sí por esta metodología (según los autores) y este neuropéptido (Taquicinina 4) presenta irregularidades en su secuencia. (tomado de Neupert, 2009)

Bajo las consideraciones enumeradas previamente y teniendo en cuenta los conocimientos que existen hasta el momento, surge la necesidad de revisar, curar y determinar de manera fehaciente los neuropéptidos y proteínas hormonales de *O. fasciatus*.

Como primer paso para la determinación de la presencia de neuropéptidos en *O. fasciatus* se procedió a la búsqueda e identificación bioinformática de todos los marcos de lectura correspondiente a estos genes.

Dado que el genoma de este insecto se encuentra en proceso de anotación, revisión y curado, se diseñó una estrategia que incluyera, en un primer paso, la generación de un transcriptoma *de novo* con el fin de facilitar a la anotación de dichos genes.

4.B. Procesado y calidad de las muestras

A partir de ARN extraído de cerebros adultos se realizó la cuantificación y la secuenciación correspondiente para el ensamblado de un transcriptoma. Una vez obtenidas las lecturas (*reads*) se procedió con la evaluación de su calidad. Para ello se utilizó el programa fastQC (Wingett & Andrews, 2018), que permite determinar el número y largo de secuencias, la calidad por base y promedio de la secuencia, la proporción de bases por posición, el contenido de guanina + citosina (%GC), la presencia de secuencias sobrerrepresentadas y los adaptadores de secuenciación.

Tabla 4. Medidas de resumen de calidad. Para cada muestra del par de *reads* se muestra el porcentaje de *reads* duplicadas, el contenido porcentual de Guanina + Citosina, la longitud y la cantidad de secuencias.

Nombre de la muestra	% Duplicados	% GC	Longitud	Millones de secuencias
Ofas-RNA_1.fastq.gz	81.70%	45%	151 pb	211.5
Ofas-RNA_2.fastq.gz	76.70%	45%	151 pb	211.5

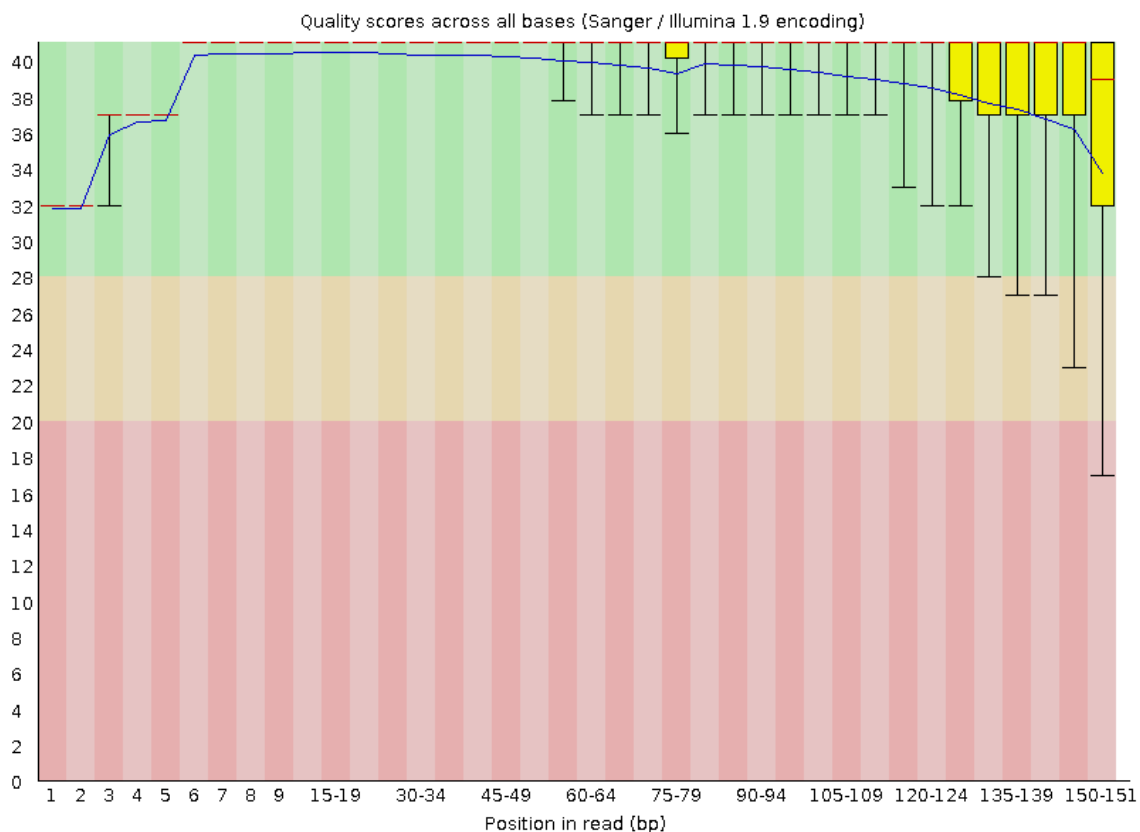


Figura 6. Distribución de Phred Score en función de la posición a lo largo de cada read.

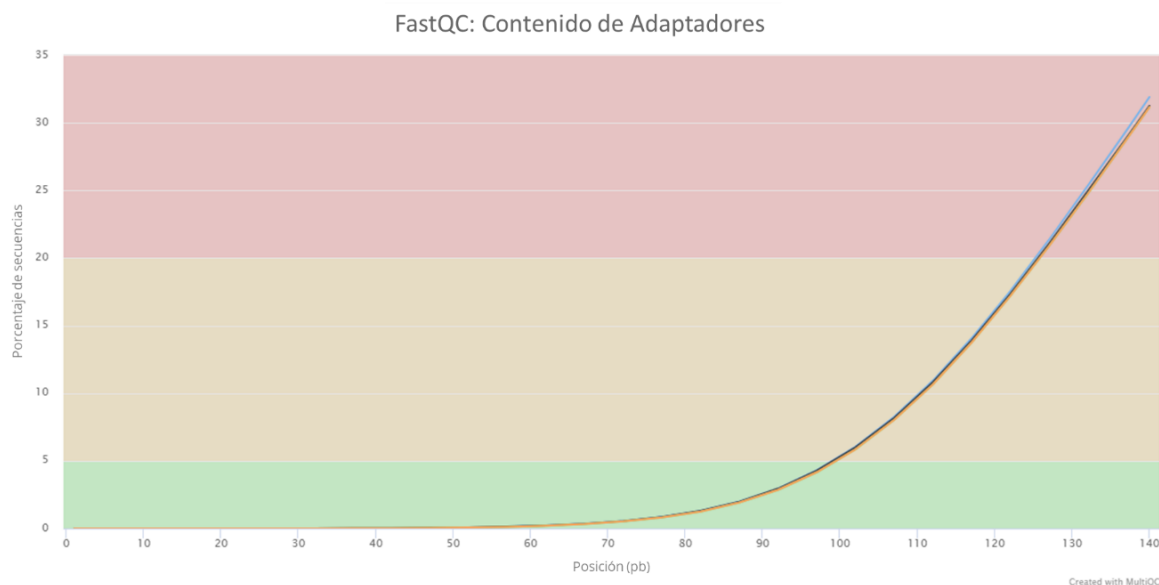


Figura 7. Distribución porcentual de secuencias que presentaron adaptadores de secuenciación en función de la posición dentro de las *reads*.

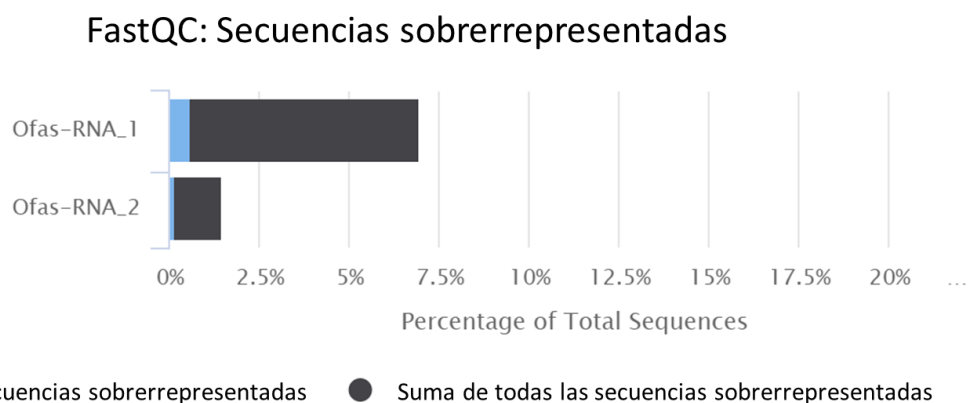


Figura 8. Porcentaje de secuencias sobrerrepresentadas para las lecturas en crudo. En negro se muestran las secuencias sobrerrepresentadas y en azul el porcentaje (sobre el total de secuencias) que presenta mayor grado de representación.

Como puede observarse en las figuras 6, 7 y 8 las métricas fueron satisfactorias. Cabe destacar que se hallaron tres principales problemas, el primero se debió a la presencia de adaptadores de secuenciación, el segundo fue el exceso de secuencias sobrerrepresentadas (mucho mayor al 1% recomendado) y, finalmente, se observó un alto contenido de GC, mayor al 40% del contenido típico promedio esperado para este organismo (ver Figura 9). En este último también se aprecian picos en la cola de la distribución (55-75%) normalmente asociada a la presencia excesiva de moléculas de ARN ribosómico.

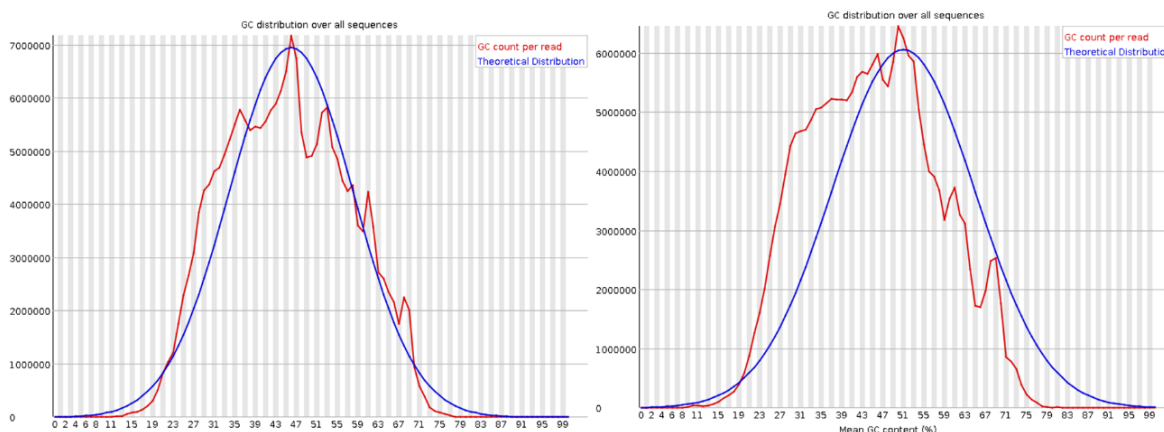


Figura 9. Distribución de la cantidad de *reads* en función del contenido porcentual de GC. Se muestran separados los gráficos para las lecturas que componen el par.

4.C. Ensamble *de novo* y eliminación de ARN ribosómico

Los detalles de la depleción *in silico* de ARN ribosómico, el ensamble *de novo* mediante Trinity (Haas et al., 2013), la caracterización de este transcriptoma y las correcciones de los marcos de lectura de los neuropéptidos en el genoma pueden encontrarse en la sección 8 Anexo.

Brevemente, fueron eliminadas en una primera instancia las secuencias de ARN ribosómico contaminantes mediante la herramienta fastQScreen (Wingett & Andrews, 2018). Luego, fueron eliminadas las subsecuencias de adaptadores de PCR y secuenciación mediante Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), posteriormente fue corroborada la efectividad de los pasos por medio de FastQC (Wingett & Andrews, 2018). Las lecturas limpias de contaminantes fueron ensambladas *de novo* con Trinity (Haas et al., 2013) y caracterizadas por Trinotate (Bryant et al., 2017) para proceder a la revisión y curado de marcos de lectura.

4.D. Revisión y correcciones de marcos de lectura

Con el transcriptoma ensamblado y caracterizado se procedió a alinear los transcriptos ensamblados *de novo* al genoma con su anotación mediante STAR (Dobin et al., 2013) en su versión *STAR-long*, y se revisó manualmente cada marco de lectura de cada neuropéptido con el soporte de las anotaciones obtenidas con Trinotate (Bryant et al., 2017).

El mapeo mostró buenas métricas de alineamiento con un millón de transcriptos correctamente alineados al genoma y menos del 1% de transcriptos mapeados a múltiples regiones y, en promedio, menos del 1% de polimorfismos de nucleótido simple (SNPs) respecto del genoma. Ver tabla 5.

Tabla 5. Resumen de las principales métricas de alineamiento. Estadísticas de mapeo de transcritos ensamblados *de novo* contra el genoma anotado de *O. fasciatus*.

STAR-long	
Numero de Transcritos	1.179.126
Longitud promedio del transcripto	557
Longitud de transcritos N50	665
TRANSCRIPTOS UNÍVOCOS	
Número de transcritos correctamente alineados	1.010.011
Porcentaje de transcritos alineados unívocamente	85,66%
Proporción promedio de bases discordantes (mismatch)	0,93%
TRANSCRIPTOS MULTI-MAPEADOS	
Transcritos mapeados a múltiples sitios	0,52%
TRANSCRIPTOS NO MAPEADOS	
Transcritos no mapeados	13,40%

Al momento de la revisión se corrigió el marco de lectura de gen que codifica para la Taquicinia. El primer exón del gen correspondiente a este neuropéptido no se encontraba correctamente anotado. Se procedió a su extensión en 52 nucleótidos de longitud (14 aminoácidos) al final de dicho exón para contribuir con su correcta anotación según lo observado durante la recolección e integración de los datos de este trabajo (ver figura 10).

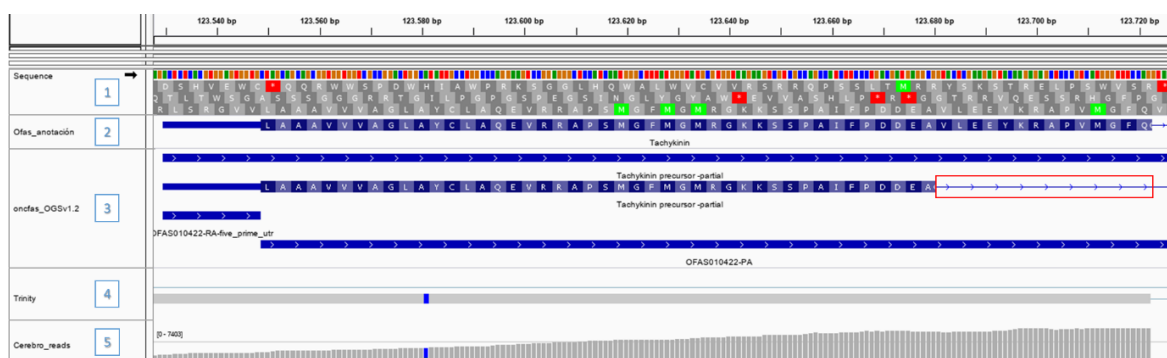


Figura 10. Representación del marco de lectura (ORF) del gen de la Taquicinia en el Scaffold 800. 1) Secuencia del primer exón. 2) Anotación corregida. 3) Anotación original del genoma (oncas_OGSv1.2). Se destaca en rojo la diferencia entre el ORF original y el corregido. 4) Transcripto originado por Trinity (Haas et al., 2013). 5) Reads mapeadas al genoma.

La revisión manual del resto de los neuropéptidos no mostró cambios en los marcos de lectura codificantes, aunque se pudo detectar la presencia de regiones 3' y 5' no traducidas (UTRs) que no fueron modificadas en los subsiguientes análisis de este trabajo, pero serán puestas a consideración de futuros trabajos de revisión y curado del genoma de *O. fasciatus*.

4.E. Alineamiento de *reads* limpias a genoma correctamente anotado

Las *reads* se alinearon dos veces consecutivas con STAR (Dobin et al., 2013) sin presencia de secuencias contaminantes al genoma anotado y corregido, según se detalla en las secciones 4.C y 4.D.

En los controles de calidad se observó un buen desempeño del alineamiento. Aproximadamente el 83% de las *reads* mapearon correctamente a una única posición en el genoma, el 15% lo hizo correctamente en más de una posición y un 1.1% de las *reads* no mapearon en el genoma (Ver tabla 6).

Tabla 6. Resumen de las principales métricas de alineamiento. Estadísticas de mapeo de las *reads* contra el genoma.

STAR	
Número de pares de reads	149.929.134
Longitud promedio del par de reads	275
TRANSCRIPTOS UNÍVOCOS	
Número de reads correctamente alineadas	124.302.270
Reads alineadas unívocamente	82,91%
Proporción promedio de bases discordantes (mismatch)	0,47%
TRANSCRIPTOS MULTI-MAPEADOS	
Reads mapeadas a múltiples sitios	15,75%
TRANSCRIPTOS NO MAPEADOS	
Reads no mapeadas	1,16%

El 42% de las *reads* coincidieron con posiciones exónicas y se detectaron un 27% de sitios alternativos de *splicing* para la unión entre exones ya declarados. La cobertura mostró que más de 11.000 marcos de lectura (~55%) tuvieron al menos 50X de cobertura mientras que unos 3000 (~15%) no mostraron expresión en tejido cerebral (ver figuras 11 y 12).

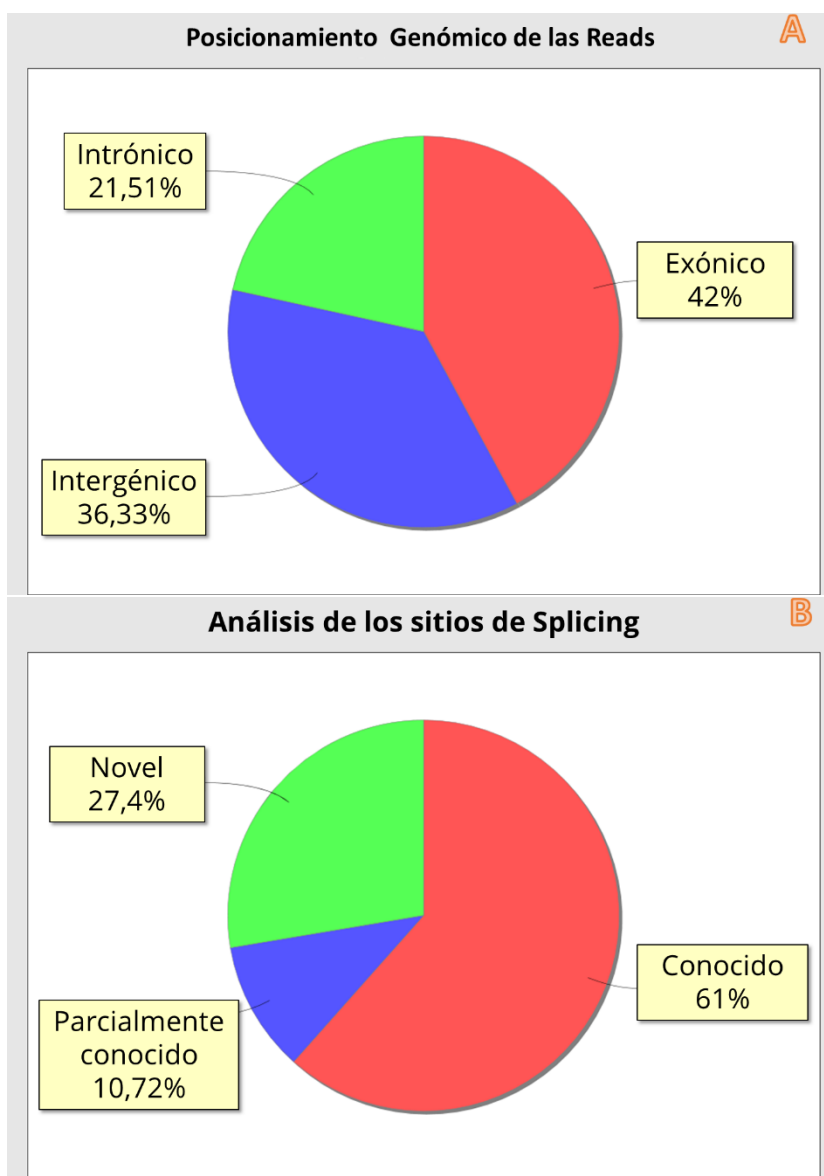


Figura 11. A) Posicionamiento de las *reads* sobre diversos sitios genómicos anotados. B) Resumen de los sitios de corte y empalme (*splicing*) encontrados durante el alineamiento.

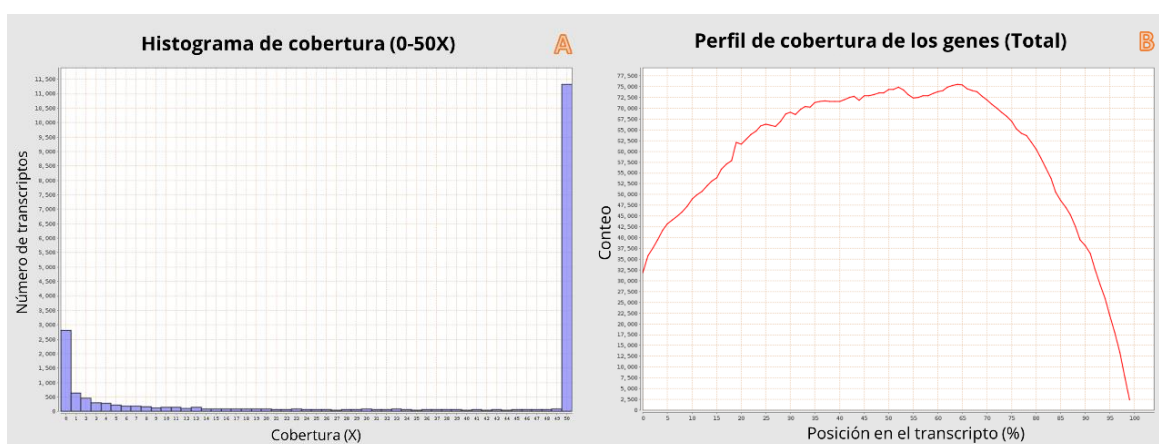


Figura 12. A) Histograma con la distribución de número de transcritos (ORFs) en función de la cobertura.

B) distribución de la cobertura función de la posición dentro del transcrito.

4.F. Expresión de neuropéptidos

Se cuantificó la expresión de los mensajeros de todos los genes anotados. Con las *reads* alineadas al genoma anotado y corregido, se realizó la cuantificación para evaluar la presencia de los neuropéptidos y realizar una comparación cualitativa (más o menos expresados) dentro del mismo tejido en el mismo momento.

Una vez realizada la cuantificación con Stringtie (Pertea et al., 2015) se seleccionaron los genes de interés a este trabajo. Se observó la expresión de todos los neuropéptidos previamente señalados (ver tabla 2). Se desprende de esta cuantificación que:

- Los transcritos relacionados a ecdisis (hormona de eclosión, hormona disparadora de la ecdisis, bursicona-alfa, bursicona-*beta* y hormona protoracicotrópica) están en los niveles más bajos de todos los neuropéptidos, coincidente con el estadio adulto desarrollado del insecto. Aunque resulta destacable que se halla en valores detectables, sugiriendo funciones distintas a las del comportamiento de la muda como podría llegar a ser su implicancia en la reproducción (Meiselman et al., 2017).
- Los neuropéptidos Precursor de neuropéptidos (NPLP), NVP, Hormona adipoquinética (AKH), ITG, Taquicinina, Hormona diurética DH31, SIF-amida y neuropéptido F largo (NPF) se encuentran dentro del 5% de los genes con mayor transcripción y, se destacan como funciones principales de estos la regulación del metabolismo energético, la excreción, la actividad muscular y los ciclos circadianos, incluyendo el control del sueño. Cabe destacarse que algunos de estos neuropéptidos no tienen una función conocida en insectos a la fecha.

En la tabla 7 se muestra la localización genómica de los transcritos y su expresión medida en transcritos por millón (TPM). Complementariamente, en la figura 13 se muestra un gráfico de barras con la expresión de cada neuropéptido medido en TPMs (ordenadas de menor a mayor).

Tabla 7: Localización genómica de los transcritos seleccionados y su expresión medida en TPM. Se observa el Identificador del gen, el nombre, el *scaffold*, la cadena y el inicio y fin del marco de lectura para la versión actual del genoma. La última columna indica el valor medido en TPM de la expresión de cada gen. Se coloreó en un gradiente de tres colores para indicar los valores bajos (rojo), intermedios (amarillo) y altos (verde).

Nombre	NP	Cromosoma	Cadena	Inicio	Fin	TPM
Allatostatina A	AstA	Scaffold6312	-	295	3234	37.92
Allatostatina B - péptido myoinhibitorio	AstB	Scaffold902	+	74116	80992	33.46
Allatostatina CC	AstCC	Scaffold182	+	114184	139649	160.89
Allatostatina CCC	AstCCC	Scaffold452	+	346343	400628	5.60
Allatotropina	AT	Scaffold172	-	530489	532928	44.15
Bursicona	Burs	Scaffold392	-	169176	185392	0.25
Bursicona-beta	pBurs	Scaffold392	-	121566	123552	0.69
CCH-amida	CCHa	Scaffold1073	-	288198	293535	15.54
CNM-amida	CNMa	Scaffold182	+	336428	347978	65.11
Corazonina	Crz	Scaffold297	-	521546	523725	36.39
Elevenin	Elv	Scaffold1822	+	66332	76788	16.43
Factor dispersante de pigmentos	PDF	Scaffold1889	+	87788	91125	33.79
Factor liberador de corticotropina - DH44	DH44	Scaffold128	+	710558	720410	49.84
FMRf-amida	FMRFa	Scaffold13	-	531473	534638	2.55
GPA2	GPA2	Scaffold233	+	364425	395101	30.71
GPB5	GPB5	Scaffold233	.	613905	615432	2.48
Hormona Adipoquinetica	AKH	Scaffold913	-	202365	208050	372.85
Hormona de eclosión	EH	Scaffold1053	-	12342	16520	0.02
Hormona disparadora de la ecdisis	ETH	Scaffold2178	+	8405	10810	1.62
Hormona diuretica DH31	DH31	Scaffold173	-	386135	604737	137.21
ITG	ITGa	Scaffold765	-	65765	140211	238.13
Kinina	Kin	Scaffold3256	+	24919	28904	29.47
Miosupresina	MS	Scaffold1697	+	9936	98962	51.34
Neuroigin	Nlg	Scaffold1758	+	64201	64342	6.46
Neuroparsina1	NP1	Scaffold1515	+	68783	69291	46.89
Neuroparsina2	NP2	Scaffold1515	-	73454	74028	59.79
Neuroparsina3	NP3	Scaffold1515	-	87610	88154	108.00
Neuroparsina4	NP4	Scaffold1515	-	99334	135870	42.05
Neuropéptido F largo	NPF	Scaffold101	+	665947	688414	84.99
NPAA - Neuroparsina	NPAA	Scaffold338	-	313131	314327	59.95
NPLP - Proteina tipo Neuropeptido	NPLP	Scaffold898	+	36355	48626	1713.76
NVP	NVP	Scaffold991	+	265305	279578	431.51
Orcokinina	OK	Scaffold1870	+	9081	26293	52.64
PBAN - Neuropeptido activador de la biosíntesis de feromonas	PBAN	Scaffold819	-	77501	83919	55.30
Peptido cardioactivo de crustaceos	CCAP	Scaffold526	-	438759	443151	21.63
Peptido C-terminal	c-Ter	Scaffold1051	-	320353	332346	394.25
Péptido relacionado con la insulina 1	ILP1	Scaffold4433	-	12071	15567	48.84
Peptido relacionado con la insulina 2	ILP2	Scaffold4433	-	19991	23323	23.81
Peptido transportador de iones	ITP	Scaffold96	-	769020	792120	15.32
Proctolin	Proc	Scaffold3202	-	32333	33162	31.49
PTTH- hormona protoracicotrópica	PTTH	Scaffold57	-	927818	938771	6.21
Ryamida	Rya	Scaffold3769	+	5172	21316	7.26
SIF-amida	SIFa	Scaffold330	+	8712	9875	99.64
SNF - Neuropeptido F corto	sNPF	Scaffold217	+	458808	479265	35.55
Sulfaquinina	SK	Scaffold93	+	498184	502726	22.79
Taquicinina	TK	Scaffold800	+	123530	134353	159.95

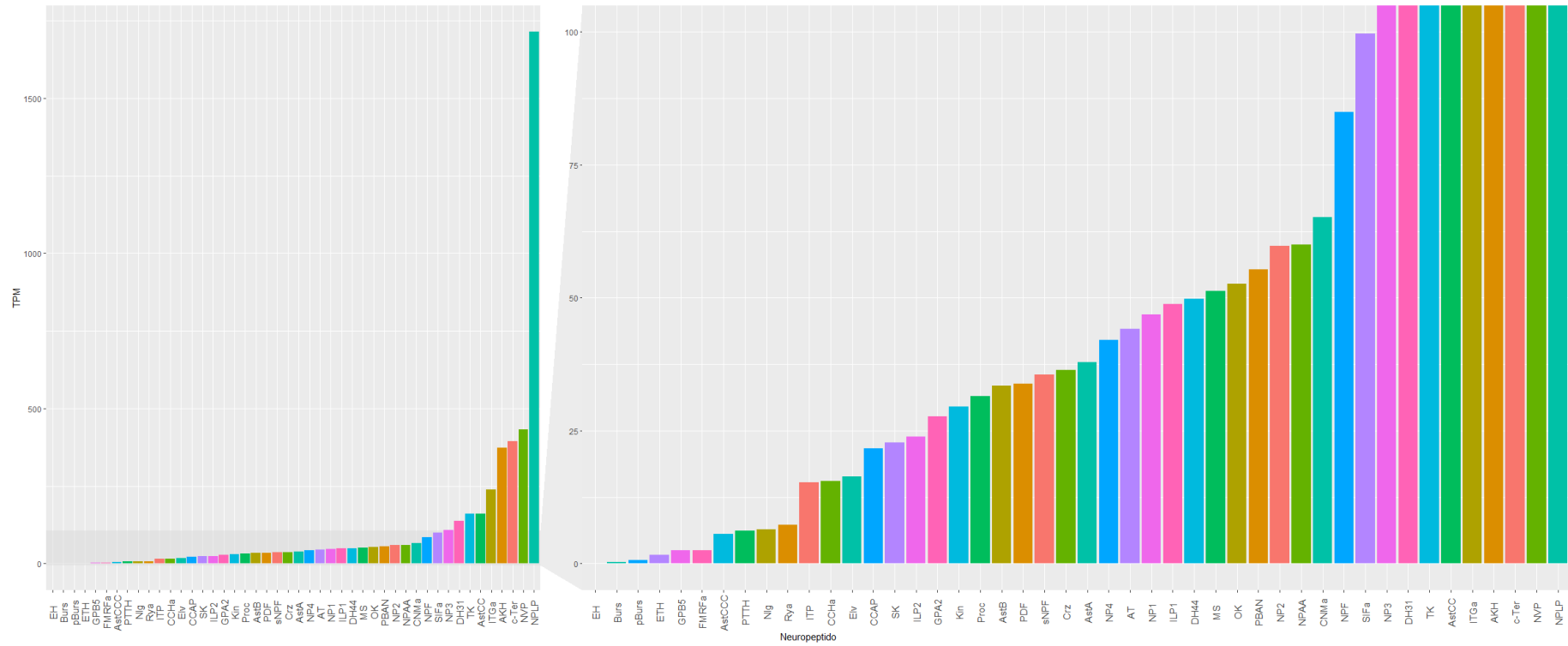


Figura 13. Expresión génica de neuropéptidos. Expresión medida en TPM para cada gen. Nota: Se proporciona en la tabla 2 las siglas de cada neuropéptido

Izquierda: Expresión de genes de neuropéptidos. Derecha: Expresión de genes de neuropéptidos con un recorte del valor de expresión de entre 0 y 100 TPM.

4.G. Expresión peptídica en cerebro

Paralelamente a la determinación de la expresión génica, mediante la técnica de *RNA-seq* se determinó la presencia y las modificaciones postraduccionales de los neuropéptidos a nivel proteico o peptídico por medio de espectrometría de masas.

Se identificaron y caracterizaron mediante Peaks (Tran et al., 2019) los espectros de masa obtenidos a partir de muestras agrupadas de cerebros de individuos adultos. En el anexo 8.D se detallan *in extenso* los neuropéptidos hallados, los péptidos identificados por metodología *bottom-up*, las modificaciones postraduccionales y las masas de estos. Adicionalmente en la tabla 8 se presenta un resumen que incluye el identificador, el logaritmo negativo del valor p para la proteína precursora, las modificaciones halladas y la cobertura porcentual (este último conjunto de datos se grafica en la figura 14). Como puede observarse, una cobertura promedio del 28% que incluye varios neuropéptidos donde se encontraron secuencias parciales o el neuropéptido activo hasta casos que incluyen una cobertura superior a dos tercios de la proteína precursora, también se hallaron las modificaciones comúnmente descritas en la literatura (Ons et al., 2011).

Tabla 8. Tabla descriptiva de las proteínas precursoras de neuropéptidos o neuropéptidos maduros identificados en este trabajo. Se indica el nombre, el identificador de referencia de Uniprot, el valor p asociado, la cobertura sobre la longitud total (resaltado en escala de menor a mayor cobertura, de rojo a verde), modificaciones determinadas mediante el espectro (A: acetilación del extremo N-terminal, M: oxidación de metionina, Q: glutamina a piro-glutámico, N: amidación del extremo C-terminal) y la identificación del neuropéptido activo (columna PA).

Descripción	UNIPROT	-10lgP	Cobertura (%)	A	N	M	Q	PA
AKH - Hormona adipoquinética	A0A2R8WJF3	-	12		N		Q	X
AstA - Allatostatina A	A0A2R7X8E8	44.5	20		N			X
AstB - Allatostatina B	A0A2R7WJ16	195.88	20		N	M		X
AstCC - Allatostatina CC	A0A2R7VZ22	58.16	21					
AstCCC - Allatostatina CCC	A0A2R7W8B9	48.44	12					
AT - Allatotropin	A0A2R7W317	157.82	34		N			X
CCAP - Peptido cardioactivo de crustaceos	A0A2R7WA90	111.75	11			M		
CCH - CCHamida	A0A2R7WQA8	53.03	18					
CNM - CNMamida	A0A2R7VZ25	43.96	8					
C-ter - Peptido C-terminal	A0A2R7WLR0	56.44	10					
DH31 - Hormona diuretica DH31	A0A2R7VZH8	167.21	37		N	M		X
DH44 - Factor liberador de corticotropina (CRF DH44)	A0A2R7WOT8	46.65	37			M		X
Ele - Elevelin	A0A2R7WWS0	111.81	51			M		
FMRF - FMRF-amida	A0A2R7VPA5	71.64	21		N			X
ILP2 - Peptido relacionado con la insulina 2	A0A2R7X8H8	41.37	7	A				
ITG - ITG amida	A0A2R7WFR3	90.87	16			M		X
Kin - Kinina	A0A2R7X7W8	-	4		N			X
MS - Miosupresina	A0A2R7WVK6	189	56	A	N		Q	X
NPAA - Neuropepsina (Net Platelet Angiogenic Activity)	A0A2R7W523	-	15					
NPF - Neuropeptido F largo	A0A2R7VV07	163.97	77		N	M		X
NPLP - Precursor tipo neuropeptido	A0A2R7W1W3	332.38	74	A	N	M		X
NVP - NVP	A0A2R7ZZZ6	393.94	64		N	M		X
Orc - Orcokinin	A0A2R7WXD1	296.31	18		N	M		X
PBAN - Neuropeptido activador de la biosíntesis de feromonas	A0A2R7WHB4	192.42	44		N			X
PDF - Factor dispersante de pigmentos	A0A2R7WXS9	51.76	18		N			X
Proc - Proctolin	A0A2R7ZZZ2	47.28	14					
RYa - Ryamida	A0A2R7XR4	90.97	21		N	M		X
SIF - SIF-amida	A0A2R7W6Z4	114.97	17		N			X
SK - Sulfaquinina	A0A2R7VUK4	71.89	23		N	M		X
SNF - Neuropeptido F corto	A0A2R7W245	125.55	35		N	M		X
TK - Taquicinina (precursor)	A0A2R7ZZ3	199.9	41		N	M	Q	X
TK1 - Taquicinina1	P86582	48.34	100		N	M		X
TK2 - Taquicinina2	P86583	105.47	100		N	M		X
TK3 - Taquicinina3	P86584	92.79	100		N	M		X
TK4 - Taquicinina4	P86585	66.8	100		N	M		X
TK5 - Taquicinina5	P86586	72.55	100		N	M		X
TK6 - Taquicinina6	P86581	49.34	100		N	M		X

4.G.1 Neuropéptidos con baja cobertura del precursor.

El péptido C-terminal obtuvo una cobertura del 10% con 2 péptidos únicos correctamente alineados. Para el caso del neuropéptido F corto (SNF) se observó una cobertura cercana al 35%, pero se identificó el péptido activo, coincidente (SPNNRSPQLRLRF) con el de la especie *Rhodnius prolixus* (Ons et al., 2009). De igual manera se identificó Kinina que presentó su péptido activo (KPVFSSWG-amida) con alta similitud a la Kinina descrita en *Periplaneta americana* (Predel et al., 1997).

La hormona adipoquinética (AKH) también presentó muy baja cobertura, 12%, porque el único péptido que fue identificado es el péptido activo (pyro-QLNFSPNW-amida).

4.G.2 Neuropéptidos no hallados

La identificación por la técnica elegida para este análisis (espectrometría de masas *bottom-up*) utiliza metodologías estadísticas (propios de cada software) para establecer que una proteína está presente en una muestra biológica fue correctamente asignada e identificada.

En el conjunto de datos analizados no fueron hallados los siguientes neuropéptidos, o los mismos no superaron los requerimientos estadísticos que permitan afirmar *bona fide* que los péptidos obtenidos por espectrometría de masas son suficientes:

- Bursicona alfa y beta
- Corazonina
- Hormona de la eclosión (EH)
- Hormona disparadora de la ecdisis (ETH)
- Glicoproteína alfa 2 y beta 5
- Proteína tipo Insulina 1 (ILP1)
- Neuroigin
- Neuroparsina 1, 2, 3 y 4
- Hormona protoracicotrópica (PTTH)

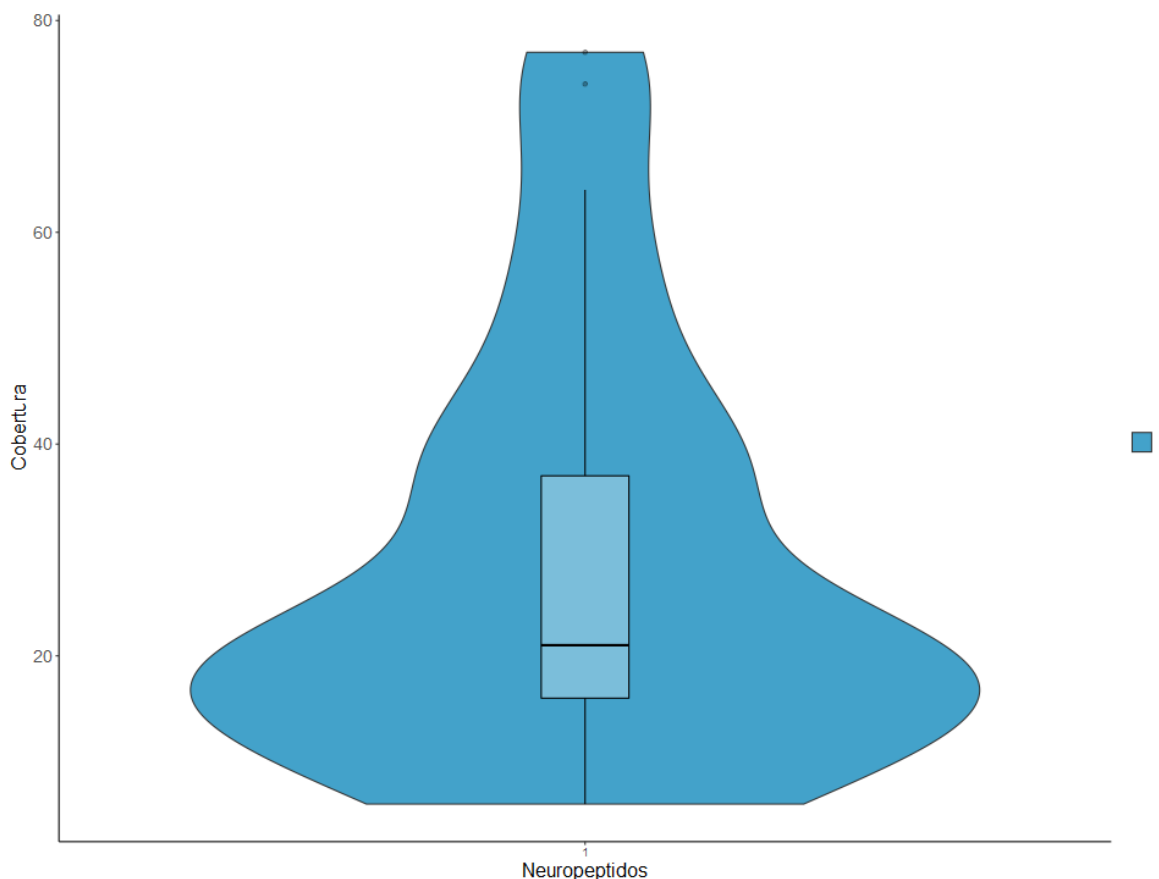


Figura 14. Gráfico de violín para la distribución de la cobertura de las proteínas precursoras de neuropéptidos por los péptidos identificados por espectrometría de masas.

4.H. Conclusiones capítulo I

Se puede establecer que los objetivos planteados para esta sección fueron llevados a cabo exitosamente ya que la estrategia propuesta en este trabajo permitió:

- Ensamblar un transcriptoma de cerebro adulto *de novo* y otro alineado al genoma mediante datos de secuenciación masiva sin presencia de secuencias contaminantes.
- Obtener marcos de lectura de neuropéptidos corregidos y manualmente curados gracias a la alta profundidad de cobertura mayor a 50X en más de la mitad de los genes totales del organismo.
- Establecer la expresión relativa de cada neuropéptido en cerebro adulto.
- Conocer y validar por espectrometría de masas más de la mitad de los precursores proteicos y los neuropéptidos activos de *O. fasciatus*. Incluyendo la confirmación de la secuencia de Taquicinina que previamente no había sido correctamente caracterizada.
- Determinar que existe una correlación entre los neuropéptidos más expresados a nivel transcriptómico coincidente con los neuropéptidos identificados a nivel proteómico por espectrometría de masas.
- Precisar que los transcriptos y péptidos de hormonas neuronales relacionados a los procesos de muda (como péptido cardioactivo de crustáceos, hormona protoracicotrópica, hormona de la eclosión y la hormona disparadora de la ecdisis) fueron menos expresados. Mientras que por el contrario las hormonas relacionadas a los procesos de diuresis, contracción muscular, apetito, crecimiento y/o metabolismo energético fueron los más representados en ambos casos.

5. Capítulo II

Expresión génica y proteica de neuropéptidos en embriones tempranos

5.A Marco histórico

Feir propone a *O. fasciatus* como animal modelo en la década de 1970 (Feir, 1974) tras recopilar más de 150 trabajos de las tres décadas anteriores de trabajo en el que se describen ventajas de su uso como animal de laboratorio en una amplia gama de cuestiones dentro de la entomología y la fisiología, incluyendo aspectos de la digestión, la alimentación, la endocrinología, el desarrollo e incluso la toxicología. Feir sostenía que “Hay muy pocos problemas biológicos o entomológicos básicos para los que *O. fasciatus* no sería un buen animal de experimentación”.

El establecimiento en la década de 1980 de *Drosophila melanogaster* como modelo de la genética y el desarrollo (Nüsslein-volhard & Wieschaus, 1980; Yamaguchi & Yoshida, 2018) desplazó la utilización de varios modelos del desarrollo, incluyendo *O. fasciatus* (Chipman, 2017). Esto puede ser corroborado evaluando las publicaciones indexadas (Pubmed) para este insecto a lo largo de los años, se observa desde mediados de 1980 hasta principios de milenio una tendencia a la baja de publicaciones (Figura 15).



Figura 15. Recuento de publicaciones indexadas en Pubmed. Cantidad de publicaciones agrupadas en periodos de 5 años.

A mediados de la década de 1990 se comenzaron a extender los trabajos de genética del desarrollo realizados previamente en *Drosophila melanogaster* (Nüsslein-volhard y Wieschaus, 1980; Rivera-Pomar y Jäckle, 1996) a otras especies con el fin de ampliar los conocimientos de la genética del desarrollo y la evolución del plan corporal en artrópodos. *O. fasciatus* fue y sigue siendo estudiado en este sentido dado que es un insecto hemíptero hemimetábolo de desarrollo embrional por banda germinal corta como se describe en la siguiente sección. (P. Z. Liu & Kaufman, 2005).

5.B Panorama general del desarrollo de *Oncopeltus fasciatus*

Las hembras de *O. fasciatus* ponen huevos en camadas de 10 a 25 huevos y la duración del proceso de embriogénesis dura aproximadamente 6 días (120 a 140 horas) a 25°C y normalmente la oviposición se da en bolas de algodón cuando se encuentran en condiciones de laboratorio lo que facilita su colección. El desarrollo larval dura un mes y la madurez sexual es alcanzada entre los 2 y 3 días luego de su última muda (Panfilio et al., 2006).

Los huevos son fertilizados en la oviposición y el núcleo cigótico se divide varias veces después de lo cual las enérgidas (núcleos sin membrana) migran a la superficie del huevo formando un blastodermo sincitial (ver figura 16 A) (Chipman, 2017). Entre la hora 18 y 34 (a 25°C y coincidente con la expresión de genes gap) el embrión ingresa a la etapa de blastodermo celular, momento en el cual se determinan los ejes principales (anteroposterior y dorso-ventral) (ver figura 16 B) (Ben-David & Chipman, 2010).

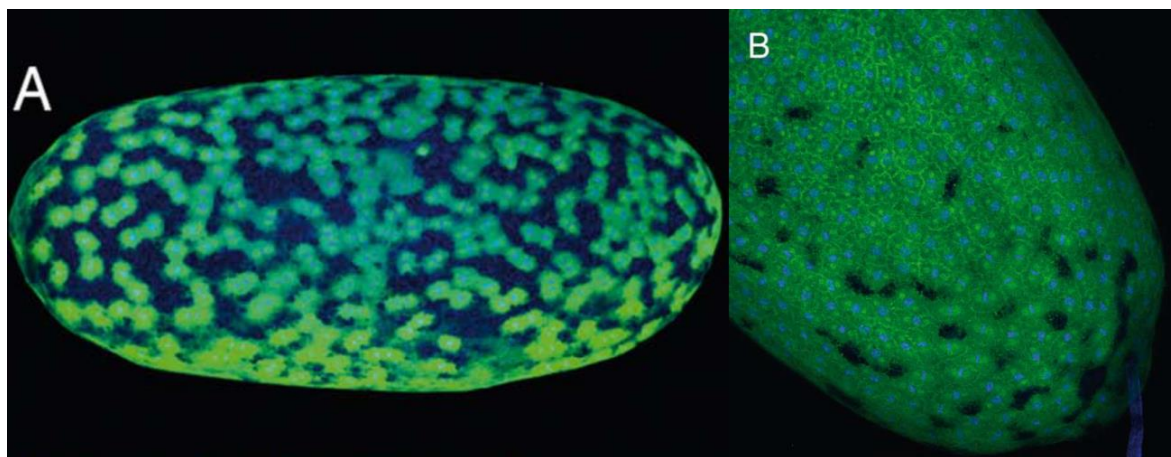


Figura 16. A) Blastodermo sincitial. Núcleos (azul - DAPI) rodeados de citoplasma sin membrana (verde – citoesqueleto – anticuerpos anti-tubulina). B) Blastodermo celular. Membranas rodeando la mayoría de las células. Tomado de Chipman, 2017.

Luego de las 34 horas el blastodermo comienza a experimentar una transición donde la mayoría de las células blastodérmicas se invaginan formando un tubo que se extiende hacia adelante desde la parte posterior del huevo (Figura 17). El tubo está compuesto por el embrión propiamente dicho (embrión en etapa de banda germinal) y una capa de células que forman el amnios (ver Figura 18).

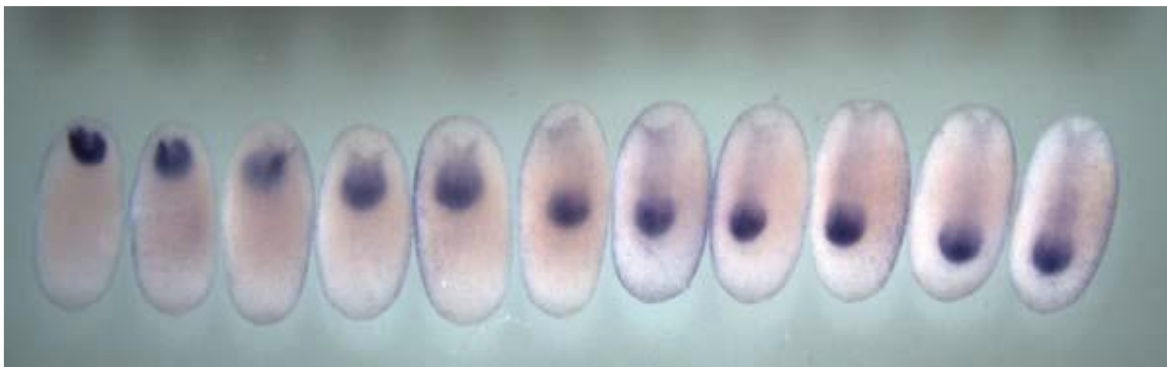


Figura 17. Progreso de la invaginación, resaltado mediante la expresión del gen caudal posterior. Los embriones progresivamente más viejos se ordenan de izquierda a derecha. Los embriones son vistos ventralmente con el posterior a la parte superior. A medida que los embriones se desarrollan, la parte posterior de la banda germinal (marcada por la expresión del gen Caudal) se mueve anteriormente, creando el tubo de la cavidad amniótica a medida que progresa. Tomado de Chipman, 2017.

La punta posterior de la banda germinal forma la zona de crecimiento (o zona de adición de segmentos) de la que se forman los segmentos abdominales. Estos nuevos segmentos aparecen de la zona de crecimiento de manera secuencial (Auman et al., 2017).

Tras la formación de todos los segmentos aparecen en la cabeza y el tórax, y estos se extienden y diferencian. Hacia el final de la embriogénesis, el embrión sufre otra serie de complejos movimientos morfogenéticos conocidos como catatrepsis (Panfilio, 2008). En este proceso el embrión emerge del huevo y se invierte su polaridad con respecto a su posición anterior (Figura 18). El final de la catatrepsis está marcado por el cierre dorsal.

Se ofrece, en la sección “Determinación de calidad de secuenciación, filtrado y alineamiento contra el genoma anotado”, la tabla 10 que resume los estadios descritos en los párrafos anteriores en función de las horas post-oviposición.

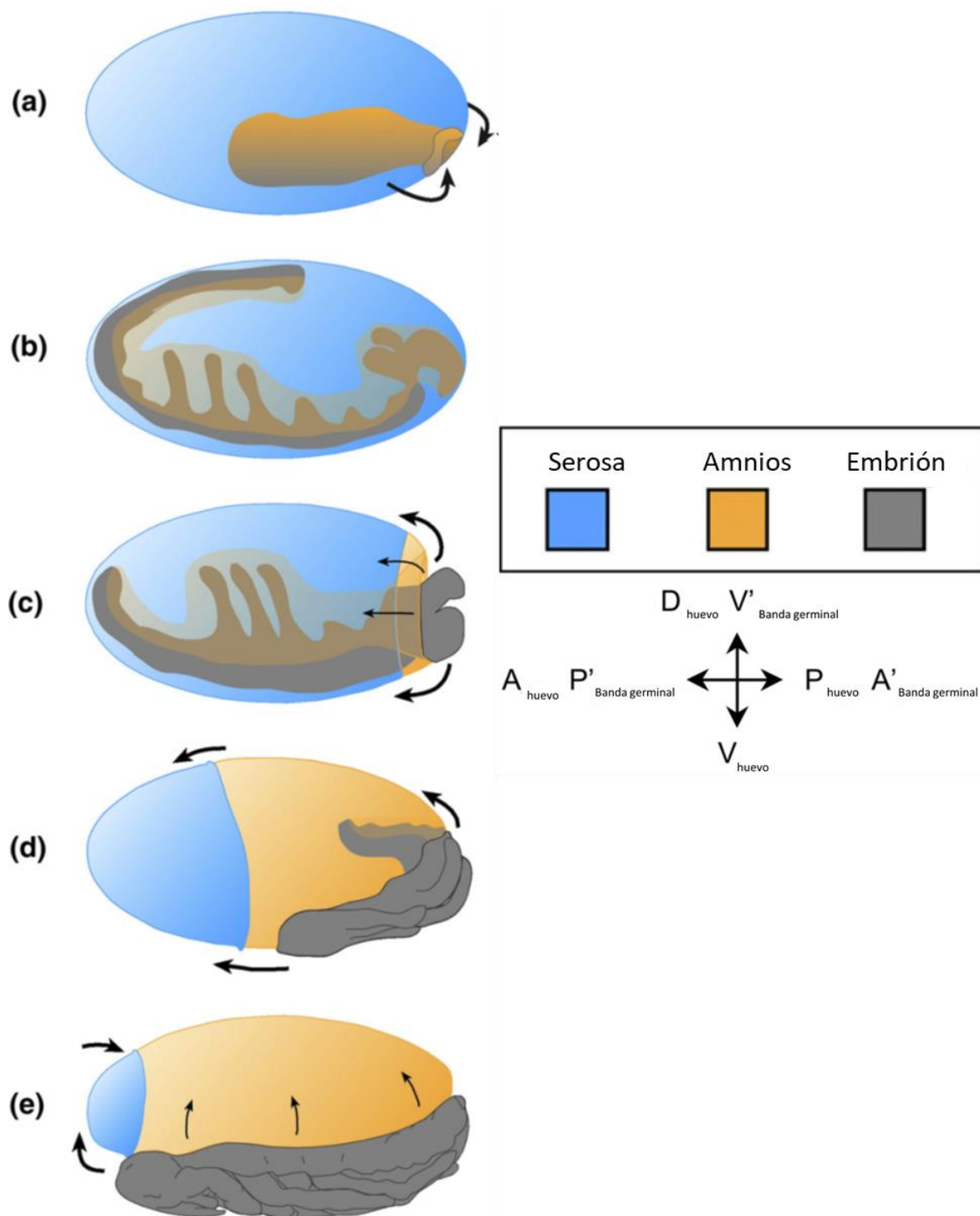


Figura 18. a) anatrepsis de inmersión, b) etapa de banda germinal, c) catatrepsis temprana, d) catatrepsis media, e) catatrepsis tardía. Las flechas negras en los dibujos indican la dirección del movimiento. Abreviaturas: A, P, D y V, ejes anterior, posterior, dorsal y ventral del huevo, respectivamente; A', P' y V', lados del embrión en estadio de banda germinal. Tomado de Panfilio, 2008.

5.C Identificación de secuencias en transcriptomas embrionarios

Como primer paso para determinar la presencia de ARN mensajeros que codificaran neuropéptidos durante etapas tempranas del desarrollo embrionario, se buscaron e identificaron por similitud de secuencia los ensambles de nucleótidos a partir de *reads* (*isotigs*) obtenidos en el trabajo de Ewen-Campen y colaboradores (Ewen-Campen et al., 2011). De esta manera se pudieron identificar los ARN mensajeros codificantes obtenidos mediante ensamble *de novo* (tabla 9). Esta aproximación se debe a que dicha publicación es previa a la anotación del genoma de *O. fasciatus* en su versión actualizada, disponible a partir del año 2014.

Tabla 9. Neuropéptidos identificados por identidad de secuencia mediante el algoritmo BLASTX (Altschul et al., 1990). Para cada neuropéptido se provee el *e-value*, el nombre del transcripto provisto en el trabajo original y el porcentaje de identidad de secuencia correspondiente.

Neuropéptido	Evalue	Isotig - identificado	Identidad (%)
Allatostatina CCC	1,36 E-33	isotig14659	100,0
CCH-amida	3,85 E-18	cap3_Contig12791	100,0
CNM-amida	5,65 E-28	isotig18606	100,0
FMRF-amida	2 E-38	cap3_singlet_GEQE5QV02GZ7R7	95,6
Kinina	9,03 E-62	cap3_Contig18371	93,2
Miosupresina	6,22 E-67	isotig19121	99,0
Neuroparsina	8,85 E-20	cap3_singlet_FQTBZRY01EIXXU	84,4
NPAA - Neuroparsina	1,74 E-29	cap3_Contig25747	97,9
Peptido C-terminal	5,47 E-35	cap3_Contig25688	75,6
Peptido relacionado con la insulina 1	1,05 E-29	cap3_Contig3393	89,7
Peptido relacionado con la insulina 2	2,26 E-54	isotig18925	100,0
Peptido transportador de iones	5,72 E-21	cap3_singlet_GESJTKM02HPYUO	94,6
PTTH- hormona protoracicotropica	4,86 E-82	isotig13648	100,0
Sulfaquinina	2,06 E-11	cap3_singlet_FQTBZRY01E2WRC	100,0
Taquicinina	1,14 E-33	cap3_Contig19713	100,0

Se obtuvieron 16 mensajeros que codifican para neuropéptidos. De estos datos se desprende que menos de la mitad de los neuropéptidos esperables se encontraron representados, razón por la cual se procedió al alineamiento de las *reads* limpias no-normalizadas con el genoma anotado y manualmente curado, como se explica en la sección 4.E. Sin embargo, los alineamientos de las lecturas con el genoma no presentaron diferencias con las expuestas por los ensambles y solo se hallaron los mismos genes enumerados en la tabla 9.

5.D Muestras de estadios embrionarios tempranos

Una vez verificada la presencia de un subgrupo de neuropéptidos en las primeras etapas del desarrollo, se estableció la dinámica de expresión de los mismos a lo largo de las etapas tempranas del desarrollo. Cabe destacar que las principales diferencias con la identificación previa (en la tabla 9) incluyen el cambio de tecnología de secuenciación (Roche 454 a Illumina) y su correspondiente aumento en la profundidad de la cobertura, la presencia de réplicas biológicas y técnicas, y la posibilidad de alinear las lecturas al genoma anotado.

Se constituyó una serie temporal de seis etapas: tejido ovárico adulto (o tiempo de referencia inicial, 0), 10, 19, 26 y 30 horas después de la oviposición (± 1 hora) y, finalmente, un pool de embriones de entre 48 a 64 horas (denominados conjuntamente como 50 horas, para simplificar la redacción y lectura tal como se detalla en materiales y métodos).

Como puede apreciarse en la tabla 10, todos los tiempos elegidos para el análisis se corresponden con etapas tempranas de desarrollo (aproximadamente 50%) y, simultáneamente, son en momentos del desarrollo en los cuales los cambios morfológicos y génicos difieren significativamente.

Tabla 10. Muestras que se incluyen en los subsiguientes análisis según el tiempo desde la postura del huevo y el estadio del desarrollo en el que se encuentra el embrión (y su correspondiente estadio blastodérmico).

Horas post-oviposición	Estadio		Muestras
0	Los óvulos son fertilizados en la oviposición	Blastodermo sincitial	Tiempo 0 horas
16	Formación del blastodermo celular		Tiempo 10 horas Genes maternos
		Blastodermo celular	Tiempo 19 horas Inicio de expresión de genes GAP
22	Inicio de la Blastoquinesis (Anatrepsis)		Tiempo 26 horas Pico de expresión genes GAP
28	Primeros signos de segmentación		Tiempo 30 horas Genes segmentalmente expresados
40	Formación del mesodermo		
46	Diferenciación de neuroblastos y células sanguíneas		
56	Diferenciación de la <i>corpora cardiaca</i> : depósito de la primera cutícula embrionaria (primeros signos visibles en la región de la cabeza)		Tiempo 48-64 horas Diferenciación de estructuras
61	Diferenciación de las glándulas protorácicas y los enocitos		
62	Diferenciación de la <i>corpora allata</i>		
63	Fin de la Blastoquinesis (Katatrepsis)		
66-78	Contracción de Serosa: formación e involución del órgano dorsal secundario		
74	Deposición de la segunda cutícula embrionaria (primeros signos visibles en la región de la cabeza)		
78	Cierre dorsal primario		
92	Cierre dorsal secundario		
94	Deposición de la tercera cutícula embrionaria		
123	Eclosión		

5.E Determinación de calidad de secuenciación, filtrado y alineamiento contra el genoma anotado

A los datos crudos (*raw reads*) se les evaluó la calidad de secuencia mediante fastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y posteriormente se prosiguió a la eliminación de los adaptadores de secuenciación por Trimmomatic (Bolger et al., 2014) como se describe en la sección 3.B.2.

Las lecturas crudas presentaron una buena calidad general (ver figura 19), la presencia de adaptadores se encontró en menos del 10% de las *reads* y la distribución de Guanina y Citosina fue uniforme en todos los casos con un valor promedio del 40% (ver figura 20).

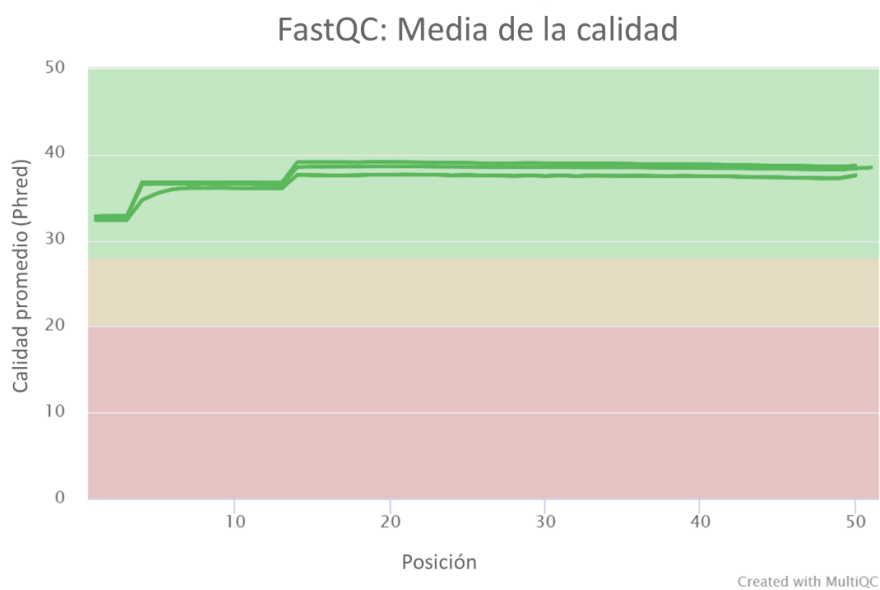


Figura 19. Calidad de secuenciación (phred score). Valor de calidad promedio por posición en cada read.

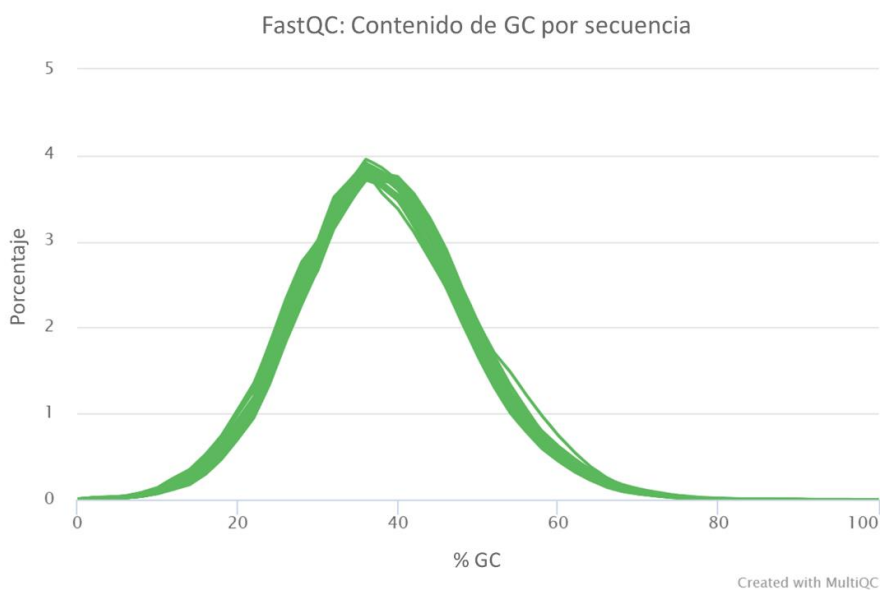


Figura 20. Distribución del contenido porcentual de Guanina-Citosina (%GC) a través de todas las lecturas analizadas en cada muestra.

Una vez descartados los adaptadores y las *reads* cuya longitud fuese menor a 20 bases, las *reads* restantes fueron alineadas al genoma anotado de *O. fasciatus*. Mediante STAR (Dobin et al., 2013) se alinearon todas las muestras y se determinó que, en promedio, el 84.8% de las *reads* mapearon correctamente (del 78.5% al 88%), tal como se puede apreciar en detalle en la figura 21. Este resultado es concordante con los reportados en bibliografía (Dobin & Gingeras, 2015).

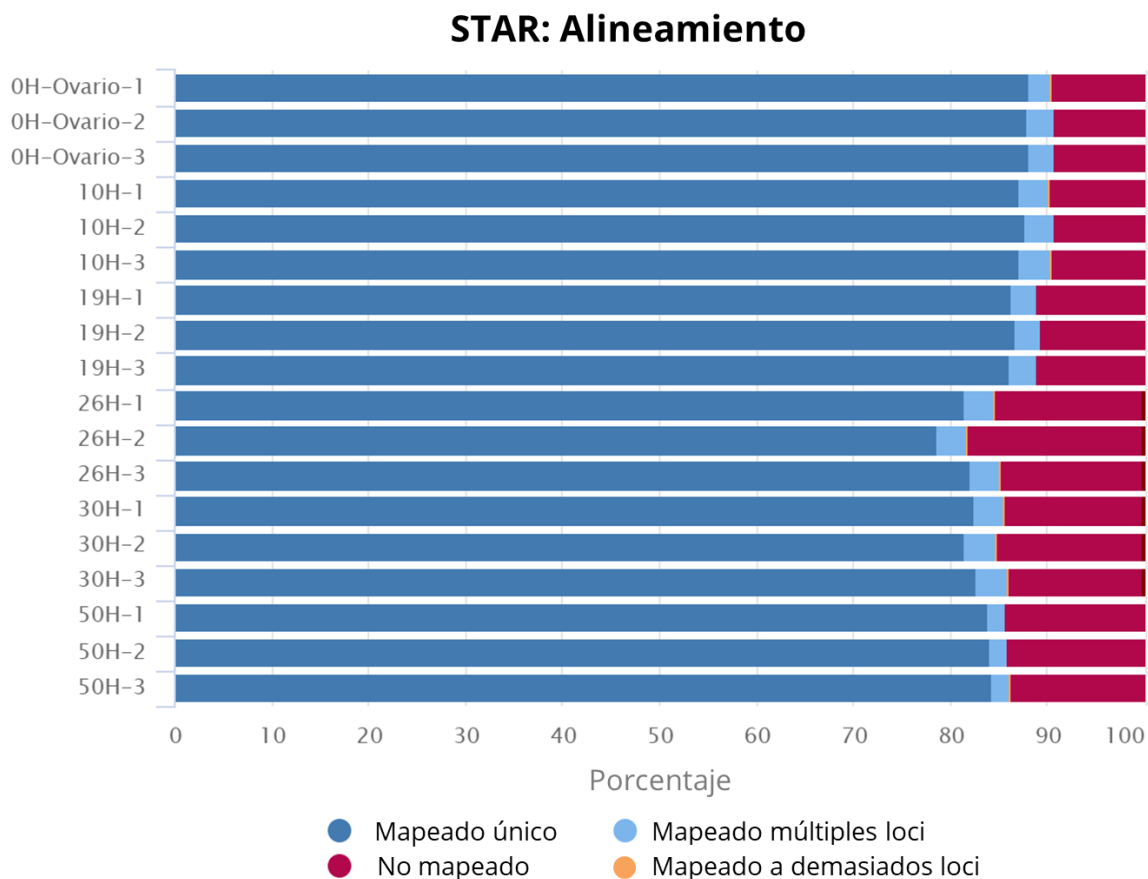


Figura 21. Porcentaje de lecturas por muestra en función del mapeo. Para cada muestra analizada se indican diferentes colores que señalan si las lecturas se alinearon a un único lugar en el genoma (azul), a más de un lugar (celeste y naranja) o no fueron mapeadas (bordo).

Mediante Qualimap (Okonechnikov et al., 2016) se observó que todas las muestras tuvieron buena cobertura y un patrón similar de distribución a lo largo de la extensión de los transcritos. Adicionalmente se desprende que, en promedio, el 65.2% (62.2-68.6%) de las lecturas tienen a los exones como origen genómico, como se muestra en la figura 22.

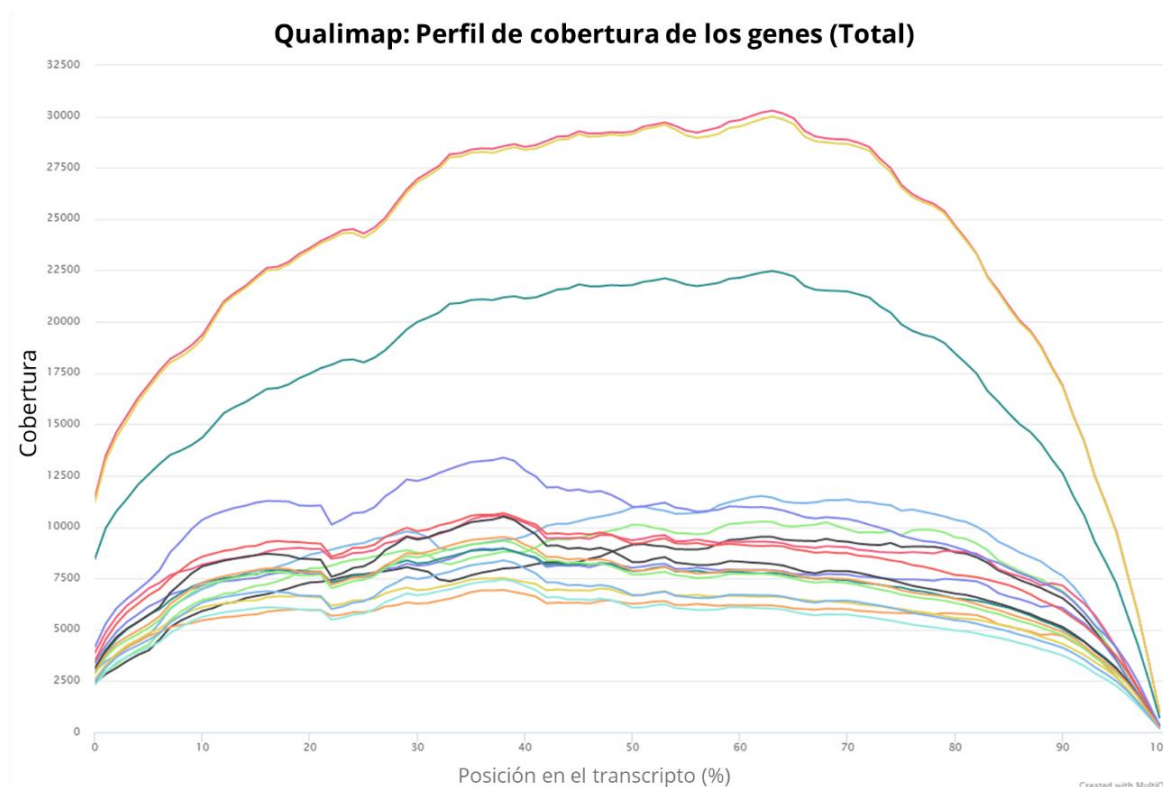


Figura 22. Distribución de la cobertura de las lecturas a lo largo de la posición porcentual ponderada. Cada línea representa una muestra diferente.

5.F Cuantificación y expresión génica diferencial de genes codificantes de neuropéptidos

Los archivos BAM, de lecturas alineadas y ordenadas por coordenadas genómicas, fueron procesadas mediante los paquetes Rsubread (Yang Liao et al., 2019) y edgeR (Robinson et al., 2010) del lenguaje R en RStudio (Rstudio Team, 2020).

Mediante la herramienta featureCounts (Y. Liao et al., 2014) fueron contabilizadas *reads* correctamente alineadas en las secciones exónicas de todos los genes. A partir de ello se filtraron los genes que no presentaron al menos una cuenta (cuentas por millón - CPM) en al menos 3 réplicas (genes no expresados). De los 19760 genes totales, 5716 genes fueron filtrados con este criterio.

Los genes expresados fueron normalizados mediante el método TMM (del mismo paquete (Robinson et al., 2010)). Este normaliza los tamaños de la biblioteca, encontrando un conjunto de factores de escalado para los tamaños de la biblioteca que reduce al mínimo los cambios de “log-fold change” (\log_2 de la magnitud de cambio) entre las muestras para la mayoría de los genes.

Con los valores filtrados y normalizados se procedió a la determinación de la dispersión de los datos y la variación biológica. Como parte de las corroboraciones previas a los análisis de expresión génica diferencial, se estableció el Coeficiente de variación biológica. El coeficiente de variación biológica (BCV, por su sigla en inglés) es el coeficiente de variación con el que la verdadera abundancia (desconocida) del gen varía entre las muestras de ARN replicado (en contraste de aquella variación causada por cuestiones técnicas), siendo esta

variación menor a 0,2 y hallándose en valores esperables y deseables en la literatura (Chen et al., 2016). (Ver figura 23 izquierda).

También se evaluó la dispersión de los datos crudos para ajustarlos estadísticamente. Este proceso reduce la posibilidad de obtener falsos positivos de genes con dispersiones crudas extremadamente altas o bajas, al mismo tiempo que aumenta el poder estadístico para detectar la expresión diferencial del cuerpo principal de los genes expresados (Ver figura 23 derecha).

Por último, se visualizó el perfil de expresión de cada muestra. Partiendo de los 500 genes más expresados en todas las muestras, se realizó una reducción dimensional a 2 coordenadas principales y ésta fue graficada. Mediante este gráfico de escalado multidimensional (figura 24) se puede observar que las muestras pertenecientes al mismo momento del desarrollo han sido correctamente asignadas y agrupadas. Adicionalmente se puede ver un patrón de recorrido, donde, en el eje Y (segunda coordenada), se ordenan secuencialmente los eventos en el tiempo de desarrollo (desde 10 hasta 64 horas post-oposición) desde valores negativos a positivos.

El ajuste estadístico y las comparaciones se realizaron mediante modelos lineales generalizados en edgeR (Robinson et al., 2010). A este se le proveyó una matriz que tiene como factores determinantes los grupos a los que pertenecen cada una de las muestras, esto es cada uno de los estadios evaluados, y las comparaciones fueron realizadas para cada estadio con respecto al anterior inmediato.

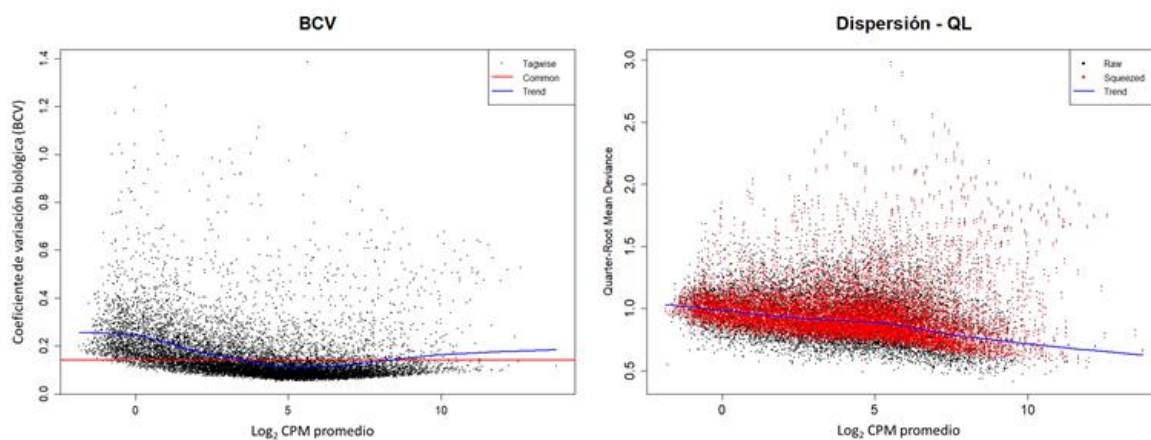


Figura 23. Izquierda: Gráfico del Coeficiente de variación biológica a en función de la expresión. Se muestra la variación de la expresión génica entre las muestras para cada gen (punto negro), la variación estimada a lo largo de los diversos genes con mismo valor de expresión (línea azul) y la variación común a todo el estudio (línea roja). Derecha: Estimación de la dispersión de los datos según su abundancia. Los datos originales (puntos negros) son ajustados (puntos rojos) para establecer la dispersión a través de todos los genes (línea azul) y moderar la presencia de valores atípicos.

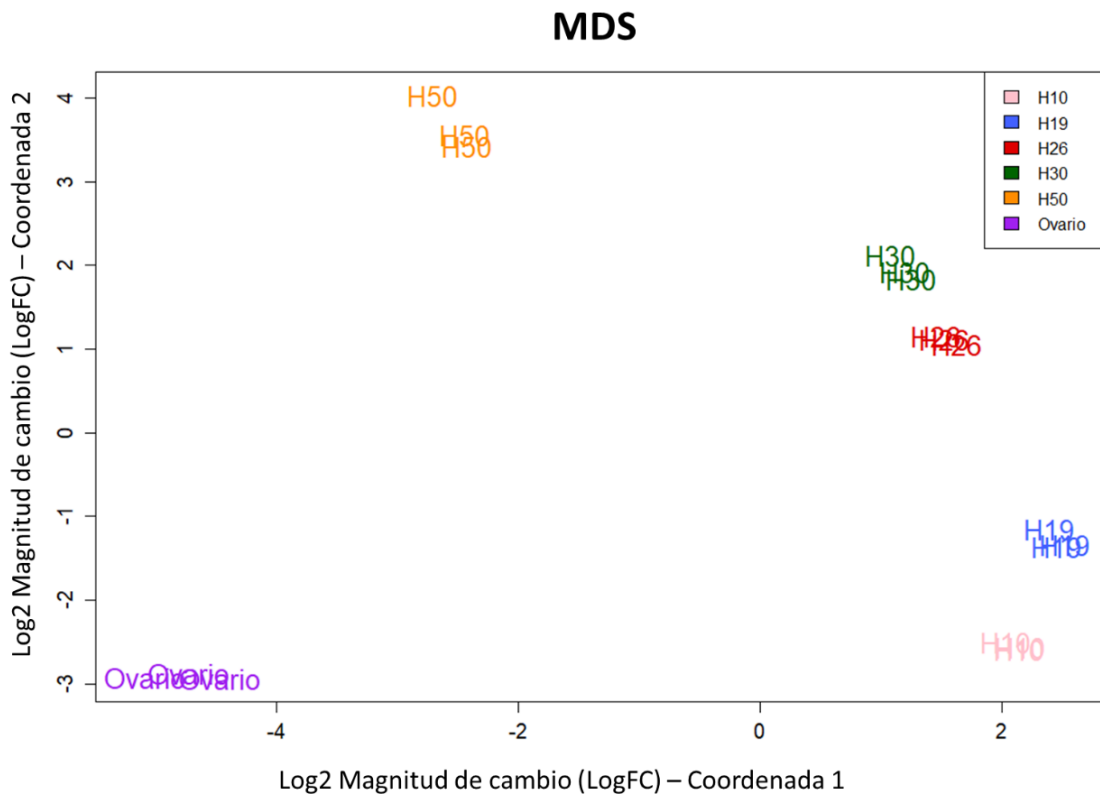


Figura 24. Análisis de coordenadas principales o de escalado multidimensional (MDS). Visualización de las diferencias entre los perfiles de expresión de diferentes muestras en dos dimensiones. Cada color representa un grupo de muestras y cada nombre, una muestra.

Los ajustes realizados permiten obtener, para todos los genes expresados del organismo, un valor promedio normalizado que permite comparar su expresión. Siguiendo los objetivos planteados en este trabajo se seleccionaron todos los genes que codifican neuropéptidos y se evaluó su perfil de expresión temporal en estadios tempranos del desarrollo.

La tabla 11 muestra los valores obtenidos para crear el perfil individual de las tres muestras de cada estadio analizado. En primer lugar, se observa que la mayoría de los ARN mensajeros para estos genes presentan muy baja o nula expresión. Es esperable que dichos genes se expresen en los momentos más tardíos evaluados (48 a 64 horas post-oviposición) ya que es el momento en que se diferencian las células nerviosas en este organismo, sin embargo, se observan algunos péptidos con expresión alta y patrones en algunos de estos genes codificantes de neuropéptidos.

Los patrones emergentes de estos datos fueron clasificados mediante algoritmos de agrupamiento (o clusterización). Los datos de Cuentas por millón (CPM) promedio normalizados y obtenidos para cada tiempo fueron escalados y agrupados mediante el algoritmo de clustering jerárquico “hclust” (R. Team, 2020). El escalado pretende retener la variación entre los diversos tiempos, independientemente del valor absoluto de las CPMs de cada gen, para luego ser agrupados. El método de clustering inicia con cada gen como un cúmulo separado del resto y luego el algoritmo procede de manera iterativa, en cada etapa uniendo los dos cúmulos más similares, continuando hasta que haya un solo cúmulo.

Finalmente, los grupos obtenidos son graficados mediante “pheatmap” (Kolde, 2015) como puede observarse en la figura 25.

5.F.1 Los neuropéptidos se agruparon según su perfil temporal

Se observaron varios patrones distinguibles y similares a los patrones de expresión de genes establecidos previamente como determinantes del desarrollo embrionario. En el siguiente apartado se detallan los neuropéptidos con un pico o patrón de expresión para cada momento y una descripción de los hallazgos realizados:

1. Expresión ovárica: péptido transportador de iones, hormona diurética DH31, Taquicinina, CNMamida, péptido relacionado con insulina 1, neuropéptido F largo y la hormona adipoquinética presentan una expresión moderada y, en todos los casos, resulta significativa la baja desde ovarios a 10 horas. Ver figura 26.
2. Expresión materna: Hormona protoracicotrópica, el péptido C-terminal y las Allatostatinas A, CC y CCC. Están en ovarios y se mantienen en los primeros estadios del desarrollo con expresión temprana sostenida. La expresión cae significativamente en estadios posteriores del desarrollo. Ver figura 27.
3. Expresión embrionaria temprana: péptido cardioactivo de crustáceos, FMRF-amida, Corazonina, Kinina y Neuroparsina. Péptidos con baja expresión en ovarios, pero con pico de expresión diferencialmente incrementada entre las 10 y las 30 horas post-oviposición. Este grupo presenta la expresión a nivel de transcritos más grande del set de datos estudiado. Ver figura 28.
4. Expresión en estadio de segmentación: Alatotropina, Allatostatina B (péptido mioinhibitorio), NVP, hormona de la eclosión, proteína tipo neuropéptido y miosupresina. Todos estos presentan una expresión básicamente nula hasta superar la barrera de las 48 horas de desarrollo, coincidente con el comienzo del desarrollo del sistema nervioso. Ver figura 29.
5. Expresión ubicua. Bursicona alfa y beta, elevenin, CCH-amida, Factor liberador de corticotropina (CRF-DH44), Glicoproteína alfa 2 y beta 5, Hormona disparadora de la ecdisis (ETH), ITG, Neuroparsinas 1 a 4, neuropéptido f corto (SNF), neurologina, péptido relacionado con insulina 2, Orcokinina, neuropéptido activador de la biosíntesis de feromonas (PBAN), Proctolin, PDF (factor de dispersión de pigmentos), RY-amida, SIF-amida y sulfaquinina. No presentan un patrón diferencial asociable a ningún estadio y su expresión es baja en todo momento.

Tabla 11. Expresión medida en Cuentas Por Millón para los tiempos evaluados de cada neuropéptido.

Neuropéptido	Expresión temporal					
	Ovario	10H	19H	26H	30H	50H
Allatostatina A	0.711	0.883	0.779	0.328	0.247	0.165
Allatostatina B - péptido myoinhibitorio	0.065	0.088	0.034	0.077	0.047	0.591
Allatostatina CC	1.428	1.418	0.744	0.667	0.676	0.478
Allatostatina CCC	28.43	26.01	31.14	28.32	15.33	1.36
Allatotropina	0.262	0.044	0.074	0.107	0.099	0.358
Bursicona	0.022	0.009	0.035	0.078	0.169	0.034
Bursicona-beta	0.000	0.019	0.018	0.021	0.016	0.017
CCH-amida	0.626	0.173	0.199	0.316	0.416	0.118
CNM-amida	4.727	0.053	0.018	0.105	0.030	0.034
Corazonina	0.538	2.763	2.408	2.591	2.172	0.153
Elevenin	0.022	0.035	0.069	0.118	0.111	0.000
Factor dispersante de pigmentos	0.000	0.003	0.018	0.490	0.763	0.132
Factor liberador de corticotropina	0.513	0.028	0.104	0.581	0.801	0.099
FMRF-amida	0.130	0.283	0.664	0.196	0.079	0.034
GPA2 - Glicoproteína alfa 2	0.000	0.035	0.121	0.020	0.032	0.000
GPB5 - Glicoproteína beta 5	0.065	0.071	0.069	0.098	0.047	0.000
Hormona Adipoquinética	0.557	0.000	0.018	0.026	0.008	0.034
Hormona de eclosión	0.000	0.000	0.018	0.005	0.079	0.623
Hormona disparadora de la ecdisis	0.000	0.006	0.181	0.221	0.171	0.067
Hormona diurética DH31	0.338	0.141	0.175	0.005	0.025	0.275
ITG	0.000	0.035	0.000	0.043	0.040	0.000
Kinina	3.290	23.082	26.916	6.850	2.073	0.370
Miosupresina	0.000	0.009	0.019	0.009	0.000	20.051
Neuroparsina 1	0.000	0.000	0.017	0.000	0.016	0.000
Neuroparsina 2	0.000	0.000	0.000	0.000	0.032	0.000
Neuroparsina 3	0.000	0.018	0.017	0.000	0.016	0.000
Neuroparsina 4	0.000	0.000	0.017	0.000	0.000	0.000
Neuropéptido F corto	0.000	0.000	0.000	0.000	0.000	0.000
Neuropéptido F largo	0.998	0.303	0.315	0.193	0.073	0.149
NPAA - Neuroparsina	624.48	2790.08	2349.27	3097.58	853.77	118.69
NPLP - Proteína tipo Neuropéptido	0.302	0.013	0.057	0.190	0.169	2.217
NVP	0.129	0.000	0.000	0.021	0.048	0.502
Orcokinina	0.000	0.024	0.026	0.439	0.170	0.085
PBAN - Neuropeptido activador de la biosíntesis de feromonas	0.000	0.000	0.000	0.042	0.000	0.017
Péptido cardioactivo de crustáceos	0.022	0.349	0.421	0.071	0.050	0.034
Péptido C-terminal	4.803	5.536	6.133	5.030	2.655	1.842
Péptido relacionado con la insulina 1	2.851	0.410	0.328	0.025	0.023	0.000
Péptido relacionado con la insulina 2	0.065	0.028	0.035	0.083	0.035	0.000
Péptido transportador de iones	1.775	0.009	0.000	0.041	0.097	0.771
Proctolin	0.000	0.018	0.017	0.020	0.016	0.108
PTTH - Hormona protoracicotrópica	21.353	27.940	21.322	5.511	2.128	0.470
Ryamida	0.022	0.069	0.054	0.056	0.015	0.036
SIF-amida	0.043	0.069	0.035	0.065	0.008	0.050
Sulfaquinina	0.000	0.010	0.101	0.090	0.111	0.034
Taquicinina	5.283	1.056	0.389	0.345	0.510	0.948

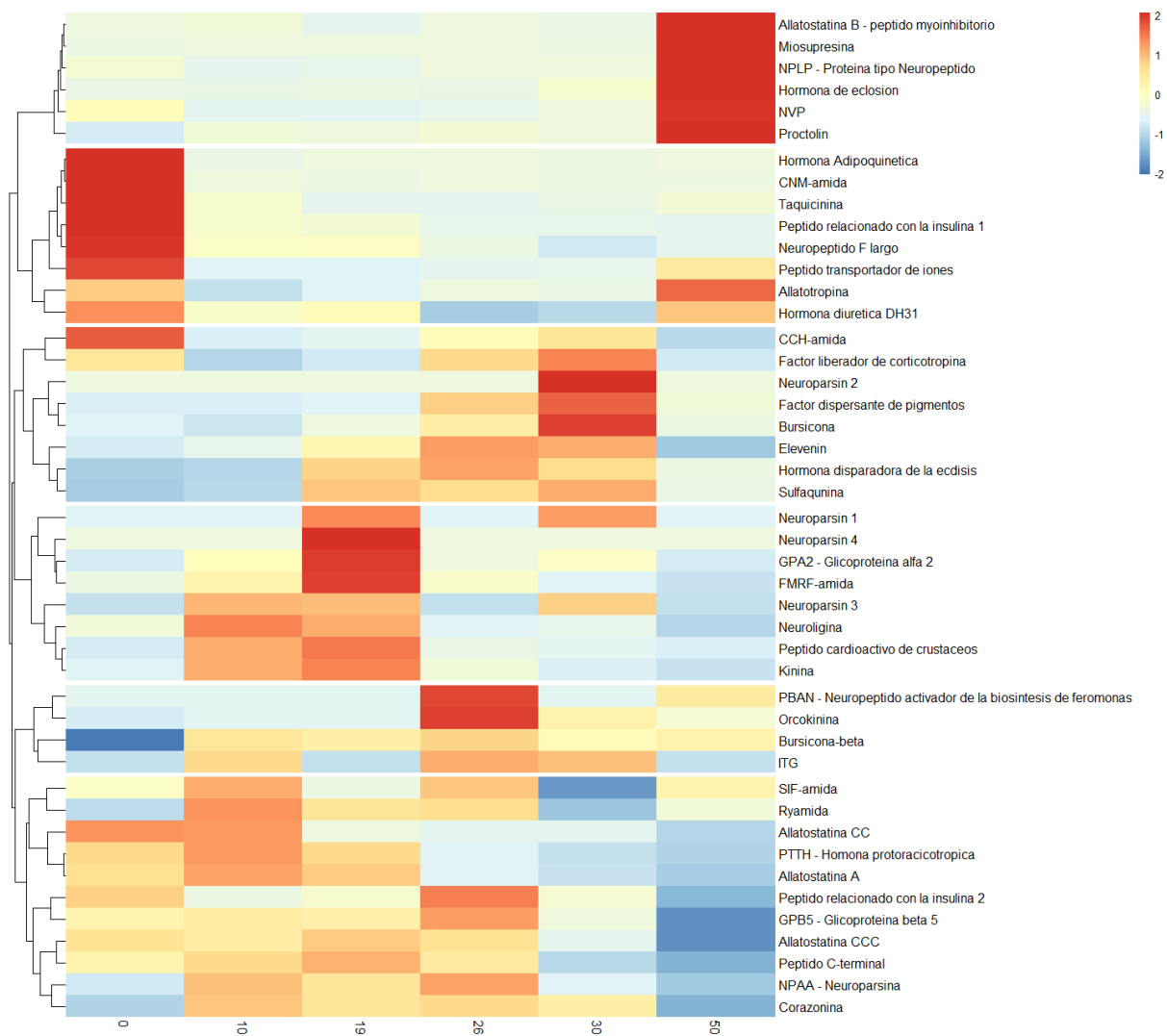


Figura 25. Mapa de calor (*heatmap*) de la expresión promedio escalada por gen codificante de neuropeptidos. Cada gen es agrupado de acuerdo con la clusterización jerárquica en uno de los 5 grupos posibles separados por líneas horizontales blancas. En colores azules se indica expresión baja, en colores amarillo claro, la expresión intermedia y en colores de la gama del rojo, expresiones altas para cada gen en cada momento. (Nota: se excluyó “Neuropeptido F corto” por no presentar expresión en ninguno de los momentos analizados).

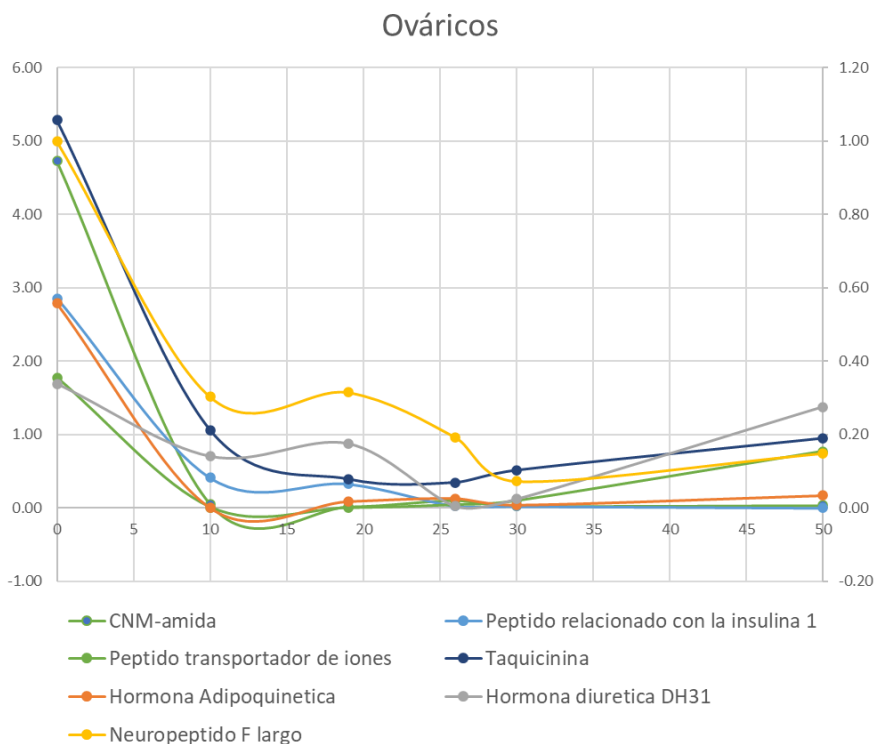


Figura 26. Expresión temporal de transcritos de neuropéptidos en función del tiempo. Valores de expresión temporal agrupados por expresión principal en ovario de *O. fasciatus*. Eje derecho: de Hormona adipoquinética, Hormona diurética DH31, Neuropéptido F largo. Eje izquierdo: Proteína tipo Insulina 1, péptido transportador de Iones, CNMamida, y Taquicinina (eje izquierdo).

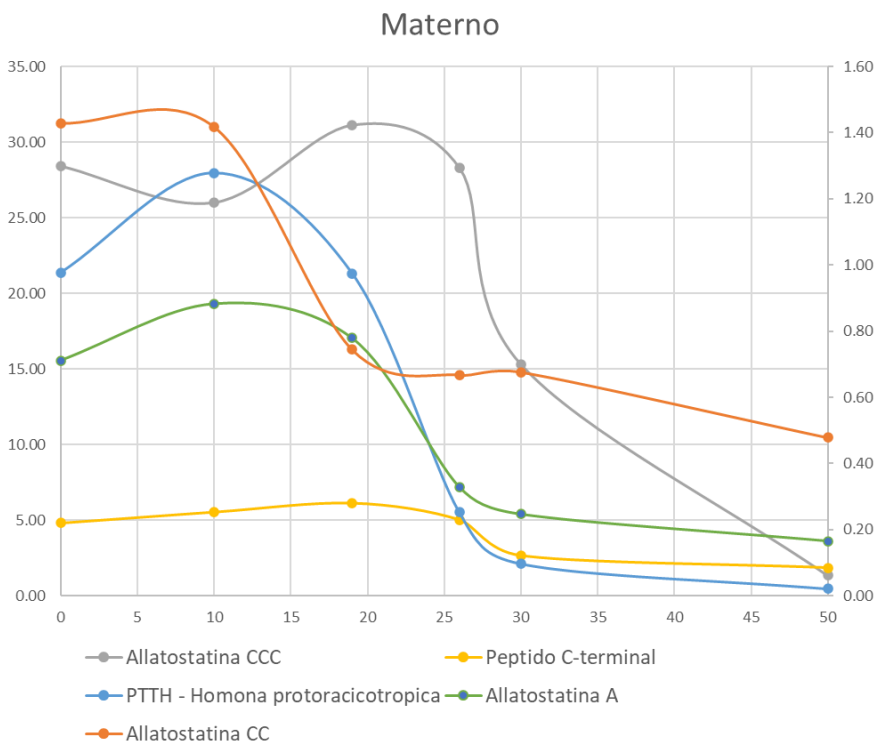


Figura 27. Expresión temporal de transcritos de neuropéptidos en función del tiempo. Neuropéptidos con expresión materna y temprana. Eje derecho: valores de expresión Allatostatina A y Allatostatina CC. Eje izquierdo: valores de expresión de PTTH, péptido C-terminal y Allatostatina CCC.

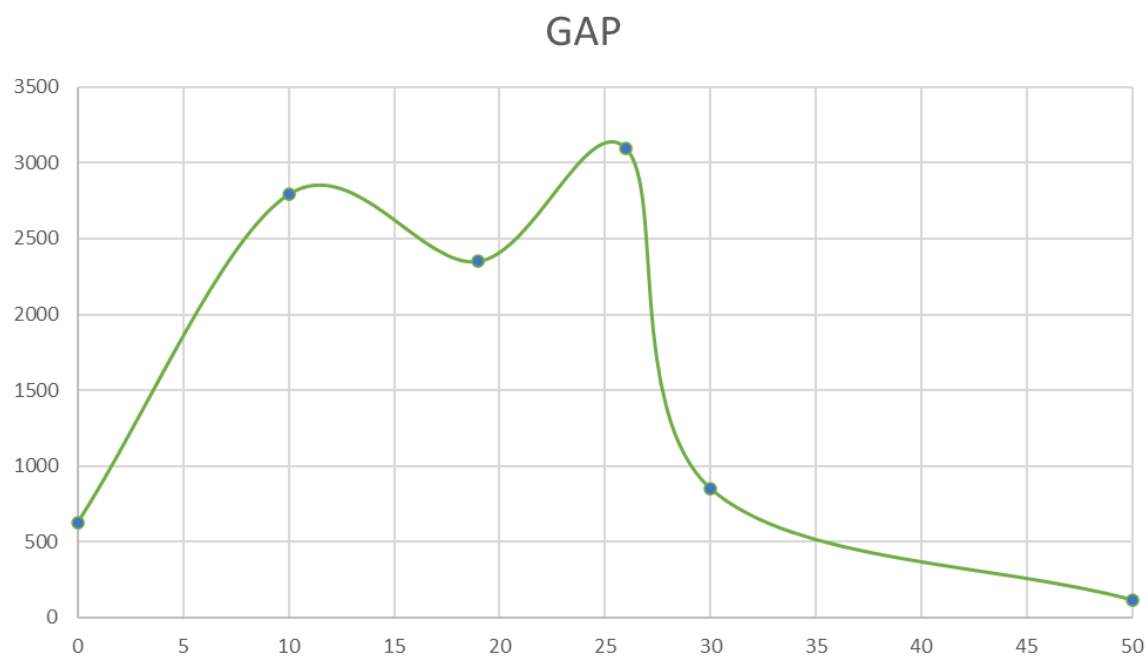
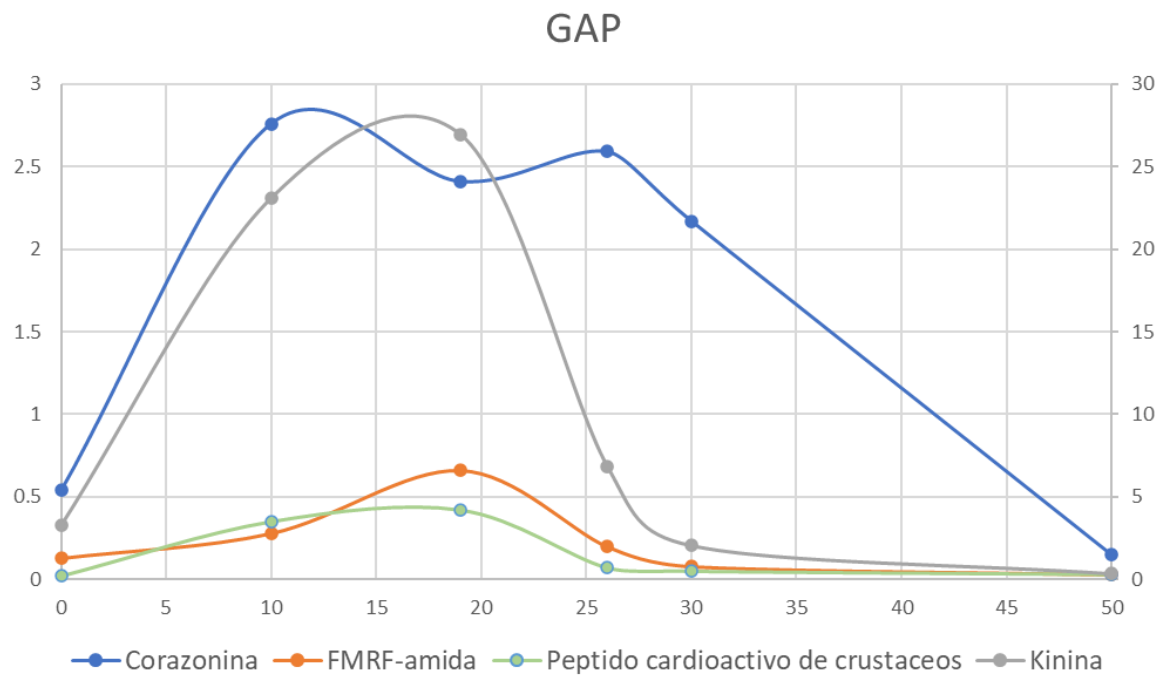


Figura 28. Expresión temporal de transcritos de neuropeptidos en función del tiempo. Neuropeptidos con expresión temprana coincidentes con la expresión de genes del desarrollo tipo GAP. Imagen superior: Eje izquierdo: valores de péptido Cardioactivo de crustáceos, Corazonina, FMRF-amida Eje derecho: Kinina. Imagen Inferior: valores de expresión Neuropeptidase. Nótese en esta ultima la diferencia en la escala respecto a la expresión.

Segmentación

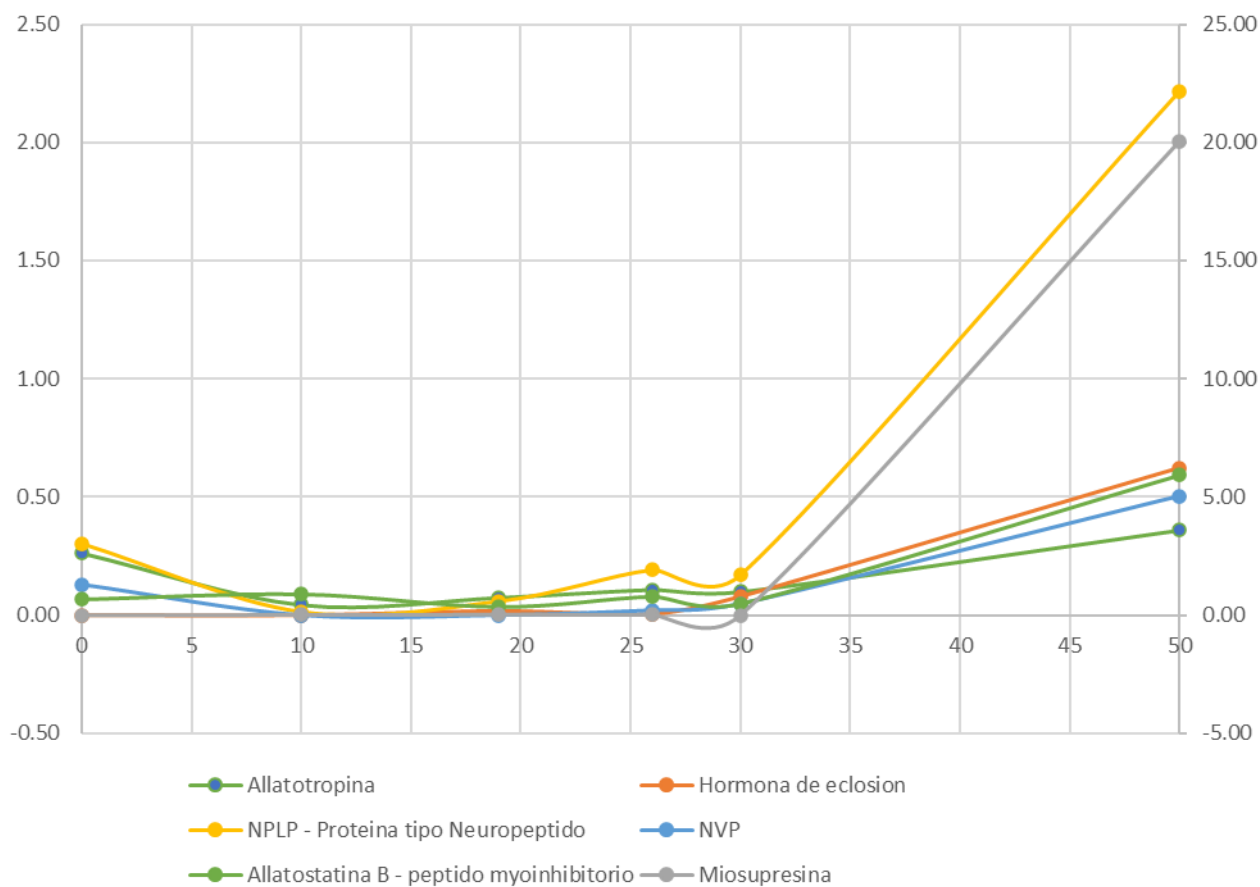


Figura 29. Expresión temporal de transcritos de neuropéptidos en función del tiempo. Valores de expresión temporal agrupadas por expresión principal luego de las 48 horas de oviposición de *O. fasciatus*, coincidentes con la segmentación y la aparición del sistema nervioso. Eje derecho: valores en la expresión de Allatotropina, Allatostatina B, Hormona de la eclosión, Proteína tipo Neuropeptido y NVP, Eje izquierdo: valores de la expresión de Miosupresina.

5.F.2 Los neuropéptidos se expresan de manera uniforme a lo largo de todo el embrión en etapas tempranas

Se evaluó la expresión génica diferencial de todos los genes a los tiempos 10, 19, 26 y 30 horas post-oviposición entre los segmentos anterior y posterior. No fueron halladas diferencias significativas de expresión para ninguno de los neuropéptidos estudiados en ninguno de los tiempos evaluados entre el segmento anterior y posterior respectivamente.

5.G Expresión peptídica en embriones

En pos de corroborar la presencia de neuropéptidos a nivel de proteína los mismos se identificaron por espectrometría de masas en una muestra de huevos de entre 0 y 24 horas post-oviposición.

Se observa en la tabla 12 el listado de proteínas precursoras de neuropéptidos halladas y la cobertura de la proteína por parte de los péptidos (ver figura 30).

Tabla 12. Tabla descriptiva de las proteínas precursoras de neuropéptidos. Se indica el nombre, el identificador de referencia de Uniprot y la cobertura sobre la longitud total (resaltado en escala de menor a mayor cobertura, de rojo a verde).

Neuropéptido	UNIPROTID	Cobertura (%)
PBAN Neuropéptido activador de la biosíntesis de feromonas	A0A2R7WHB4	65
SIF-amida	A0A2R7W6Z4	61
Taquicinina	A0A2R7WHG2	45
Hormona diurética DH31	A0A2R7VZH8	44
PTTH - Hormona protoracicotropica	A0A2R7VSI2	39
Corazonina	A0A2R7W3M4	38
Miosupresina	A0A2R7WVK6	36
NPLP- Precursor tipo neuropeptido	A0A2R7WIW3	35
NVP	A0A2R7ZZZ6	35
Allatostatina A	A0A2R7X8E8	31
CNMamida	A0A2R7VZ25	30
FMRF-amida	A0A2R7VPA5	29
Orcokinin	A0A2R7WXD1	26
Péptido relacionado con la insulina 2	A0A2R7X8H8	25
Pbur Bursicona-beta	A0A2R7W6Q3	23
GPB5 Glicoproteina beta 5	S/D	21
Kinina	A0A2R7X7W8	17
Allatostatina B - péptido myoinhibitorio	A0A2R7WJ16	16
NPAA - Neuroparsina	A0A2R7W523	14
CCAP - Péptido cardioactivo de crustáceos	A0A2R7WA90	14
ILP1 - Péptido relacionado con la insulina 1	A0A2R7X7H3	13
Bursicona alfa	A0A2R7W8C2	12
Péptido C-terminal	A0A2R7WLR0	10

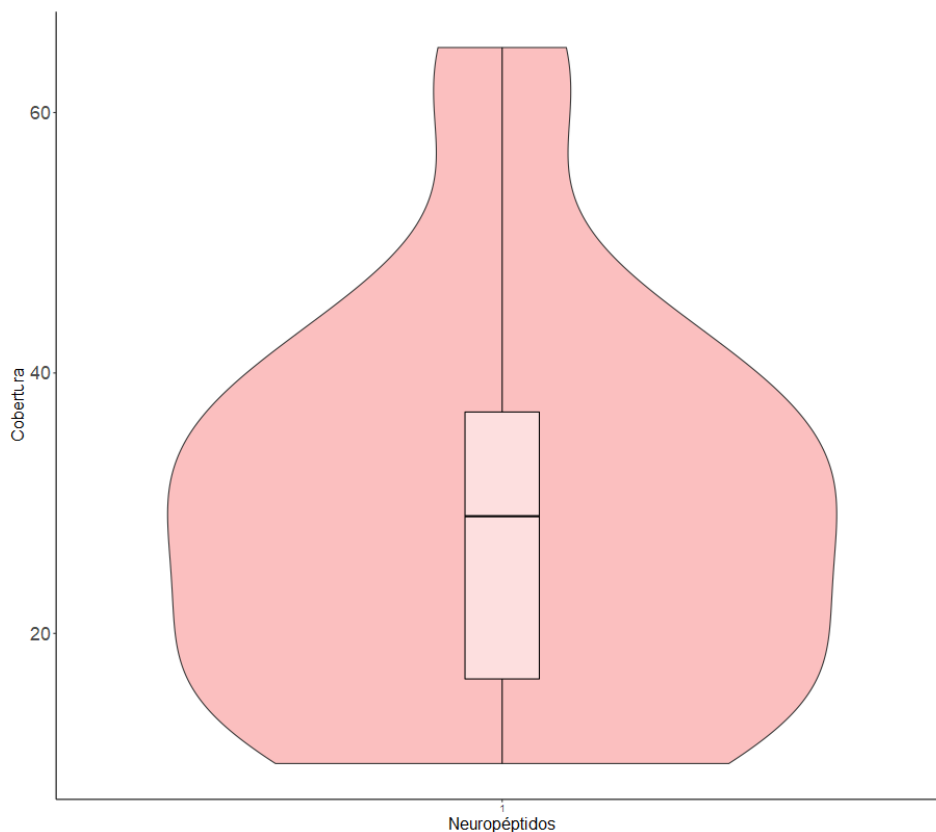


Figura 30. Gráfico de violín para la distribución de la cobertura de las proteínas precursoras de neuropéptidos por los péptidos identificados por espectrometría de masas.

5.G Conclusiones capítulo II

Mediante el análisis integrado de datos de expresión transcriptómica en una serie temporal de desarrollo embrional temprano y su correspondiente expresión proteica de neuropéptidos de hasta un día después de la oviposición se pudo establecer:

- La expresión a nivel de ARN mensajero para la mayoría de los neuropéptidos estudiados es baja.
- La expresión transcriptómica resulta variable en el tiempo durante el desarrollo temprano.
- La variabilidad de la expresión puede ser utilizada para clasificar a los neuropéptidos en grupos según el momento en el que esta muestra un pico de expresión.
- Los genes con mayor expresión transcriptómica pueden ser asociados a etapas tempranas del desarrollo cuyos picos se observan durante las primeras 20 horas post-oviposición (como Neuroparsina, hormona protoracicotrópica o allatostatina CCC).
- Algunas proteínas precursoras de neuropéptidos se hallan presentes en lo que, posiblemente, sea una forma no procesada en los embriones tempranos de hasta 24 horas.

Los péptidos activos maduros en embriones de hasta 24 horas no pudieron ser detectados en este trabajo.

6. Discusión

6.A Genómica

El primer registro de genes a escala ómica de *O. fasciatus* fue el transcriptoma de las glándulas salivales (sialotranscriptoma), donde se describió el ensamble de poco más de mil etiquetas de secuencia expresada (EST) y una baja complejidad de transcritos (Francischetti et al., 2007).

Posteriormente Ewen-Campen y colaboradores (2011) publicaron un transcriptoma más completo, basado en ovarios y diversas etapas embrionarias tempranas. Esto proporcionó una base útil para la clonación de genes para muchos de los estudios detallados.

El genoma completo ha sido secuenciado como parte del proyecto i5K (Poelchau et al., 2015) que pretende secuenciar 5000 genomas de artrópodos. La publicación del trabajo conjunto del ensamble y la anotación del genoma fue finalmente publicada en el año 2019, siendo el genoma de *O. fasciatus* (Panfilio et al., 2019) en su segunda ronda de anotación la cual actualmente se encuentra disponible (https://i5k.nal.usda.gov/Oncopeltus_fasciatus). Esta anotación actualmente dista de estar completa, ya que se encuentra un ensamble con *scaffolds* (aproximadamente 17.000 segmentos genómicos de longitud variable) y no cromosomas completos ($2n=6+XY$).

Al momento de la escritura de este manuscrito se está llevando a cabo una nueva tanda de re-secuenciación del genoma con tecnología PacBio (de lecturas largas) y soporte de datos por secuenciación Hi-C (secuenciación utilizada para analizar la organización de la cromatina en todo el genoma) (Kristen Panfilio, comunicación personal).

La anotación da significado al genoma proporcionando la ubicación y función de los genes (codificantes de proteínas o de otro tipo) y las regiones reguladoras, que subyacen a la biología de organismos vivos. Determinar cuáles son las regiones funcionales del genoma puede facilitar análisis posteriores, como por ejemplo, la evaluación del efecto de las mutaciones causadas por variantes de un solo nucleótido en individuos o poblaciones (Abril & Castellano, 2018).

Las anotaciones de un genoma normalmente se inician con la predicción *ab initio* o *de novo*, con la identificación de los genes que codifican las proteínas utilizando las señales que definen y caracterizan sus estructuras genéticas a lo largo del genoma. Esto se hace sin la ayuda de sus ARNs y proteínas codificadas (Abril & Castellano, 2018). Luego son incorporadas las predicciones por similitud de secuencia, normalmente mediante BLAST (Altschul et al., 1990) o modelos ocultos de Márkov (Eddy, 2011).

La baja en los costos y la mejora continua en la secuenciación de ARN ha provisto en la última década un incremento exponencial en el uso de transcriptomas para anotar genes, regiones reguladoras o ARN interferentes (Abril & Castellano, 2018).

El análisis de los datos de transcriptómica de *O. fasciatus* nos permitió abordar una primera caracterización genómica donde se vio que si se analiza en su conjunto el total de las *reads* desde los archivos crudos producidos mediante secuenciación, hasta obtener aquellas que

son efectivamente utilizadas en los análisis de expresión génica diferencial, las lecturas posicionadas en exones caen al valor (promedio) de 54.6 lecturas cada 100 lecturas originales. Ver Figura 31.

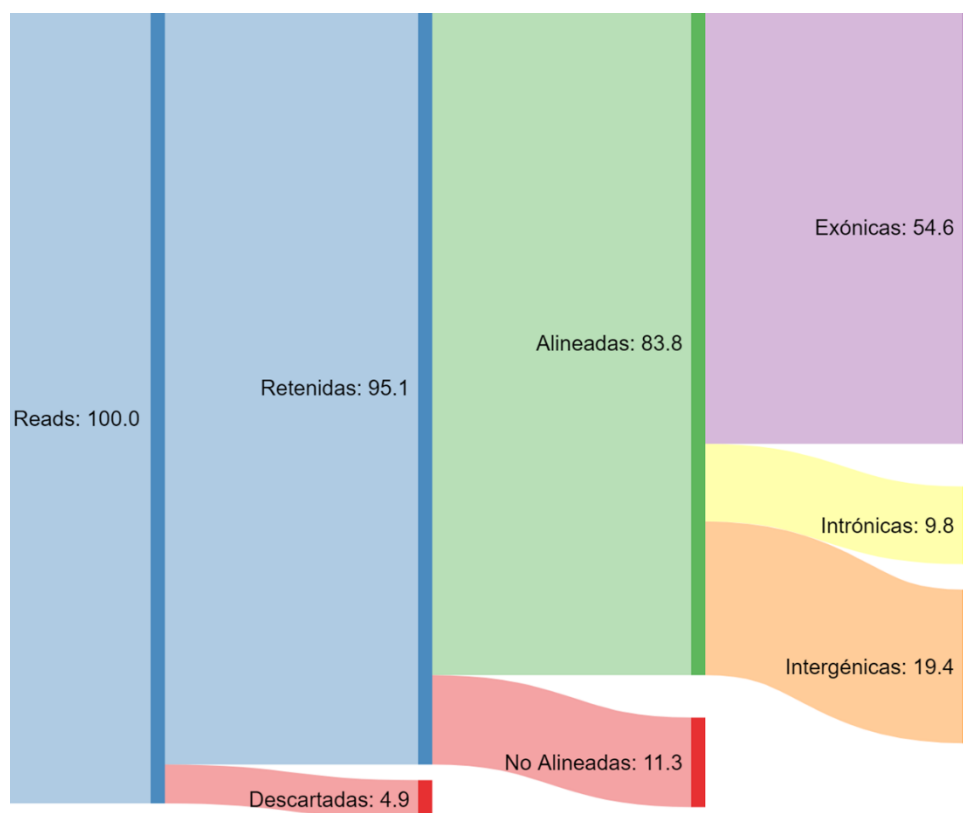


Figura 31. Valores promedio de la selección de lecturas a través de todas las muestras durante el proceso de filtrado y alineamiento, con un valor inicial de 100. Se descartaron 4.9% por no cumplir los criterios de calidad mínima. Del valor original retenido, solo un 83.8% de las *reads* fueron correctamente alineadas al genoma anotado. El 54.6% del total se corresponde a lecturas alineadas a exones de genes anotados utilizados en análisis ulteriores.

En promedio el 30% de las lecturas originales se corresponde a secuencias no alineadas al genoma o con una ubicación en sitios intergénico, además de que se observó un 39% de los sitios de corte y empalme noveles o parcialmente conocidos. Estas métricas serán mejoradas en un futuro con la asistencia de la secuenciación transcriptomas o la re-secuenciación del genoma en el futuro.

Los procedimientos de filtrado por calidad o por provenir de sitios distintos a los exónicos no resultaron relevantes en este trabajo dado que se tenía por objetivo la evaluación de genes conocidos, pero sí podrían serlo en futuras investigaciones destinadas a entender uso diferencial de exones o intrones, ayudar al ensamblado del genoma de *O. fasciatus* o anotar genes no anotados previamente con evidencia de expresión aquí presente disponible a la comunidad científica (<https://i5k.nal.usda.gov/node/933164>)

6.B. Transcriptómica

Del análisis de expresión génica en cerebro de *O. fasciatus* se pudieron obtener dos aportes principales. En primer lugar este análisis dio lugar al ensamble *de novo* de un transcriptoma

útil para la anotación de los marcos de lectura de los neuropéptidos, comparable con los realizados previamente por Ewen-campen y colaboradores (2011) con su transcriptoma materno.

En segundo lugar se observó una marcada diferencia de expresión entre los diferentes neuropéptidos estudiados como se observa en la tabla 13 (tomado de resultados del capítulo I, tabla 7).

Tabla 13. expresión de neuropéptidos en cerebro adulto, medidos en Transcriptos por millón (TPM) y ordenados por valor decreciente de TPMs. Expresión baja marcada en fondo rojo, media en fondo amarillo y alta en fondo verde. Datos tomados de tabla 7.

Neuropéptido	NP	TPM	Neuropéptido	NP	TPM
NPLP - Proteína tipo Neuropéptido	NPLP	1713.76	Corazonina	Crz	36.39
NVP	NVP	431.51	Neuropéptido F corto	sNPF	35.55
Péptido C-terminal	C-ter	394.25	Factor dispersante de pigmentos	PDF	33.79
Hormona Adipoquinética	AKH	372.85	Allatostatina B - péptido myoinhibitorio	AstB	33.46
ITG	ITG	238.13	Potolina	Proc	31.49
Alatostatina CC	AstCC	160.89	Glicoproteína alfa 2	GPA2	30.71
Taquicinina	TK	159.95	Kinina	Kin	29.47
Hormona diurética DH31	DH31	137.21	Péptido relacionado con la insulina 2	ILP2	23.81
Neuroparsina3	NP3	108.00	Sulfaquinina	SK	22.79
SIF-amida	SIF	99.64	Peptido cardioactivo de crustaceos	CCAP	21.63
Neuropéptido F largo	NPF	84.99	Elevenin	Elv	16.43
CNM-amida	CNM	65.11	CCH-amida	CCHa	15.54
NPAA - Neuroparsina	NPAA	59.95	Peptido transportador de iones	ITP	15.32
Neuroparsina2	NP2	59.79	Ryamida	Rya	7.26
PBAN - Neuropéptido activador de la biosíntesis de feromonas	PBAN	55.30	Neuroigin	Nlg	6.46
Orcokinina	OK	52.64	PTTH- homona proracicotropica	PTTH	6.21
Miosupresina	MS	51.34	Allatostatina CCC	AstCCC	5.60
Factor liberador de corticotropina - DH44	DH44	49.84	FMRf-amida	FMRFa	2.55
Péptido relacionado con la insulina 1	ILP1	48.84	Glicoproteína beta 5	GPB5	2.48
Neuroparsina1	NP1	46.89	Hormona disparadora de la ecdisis	ETH	1.62
Allatotropina	AT	44.15	Bursicona-beta	pBurs	0.69
Neuroparsina 4	NP4	42.05	Bursicona	Burs	0.25
Allatostatina A	AstA	37.92	Hormona de eclosión	EH	0.02

Al comparar esta expresión con datos provenientes de búsquedas bibliográficas que refieren a datos de medición de la expresión de ARN mensajero se hallaron numerosas coincidencias que se detallan en los siguientes apartados.

6.B.1.a Expresión neuropeptidérgica en cerebro

Los patrones de expresión media (mayor a 15 TPMs) y alta (mayores a 50 TPMs) de mensajeros codificantes medidos por *RNA-seq* para neuropéptidos, son coincidentes con los declarados para *Locusta migratoria* (Hou et al., 2015), *Pterolamus puparum* (G. Xu et al., 2020), *Diaphorina citri* (Wang et al., 2018) y *Drosophila melanogaster* (datos evaluados en Flybase (Drysdale et al., 2005)).

6.B.1.b Neuropéptidos relacionados con la ecdisis en cerebro

Los principales componentes de las vías de señalización peptídicas de la muda son la hormona de eclosión (EH), la hormona disparadores de la ecdisis (ETH), el péptido cardioactivo de los crustáceos (CCAP) y la bursicona (heterodímero de Bursicona y compañero de Bursicona o bursicona-beta) (Arakane et al., 2008). Estos neuropéptidos sumados a la hormona protoracicotrópica (PTTH) que regula la sincronización del desarrollo mostraron baja expresión en el tejido cerebral, efecto esperable para el estadio adulto.

Este fenómeno se repite en varios insectos, ya que la expresión de ARN mensajero medido por *RNA-seq* en *Diaphorina citri* (Wang et al., 2018) y *Drosophila melanogaster* (datos evaluados en Flybase (Drysdale et al., 2005)) muestra el mismo comportamiento para estos neuropéptidos. Para el caso de *Locusta migratoria* y *Pterolamus puparum* el patrón se mantiene para los neuropéptidos con la excepción de CCAP que está expresado en niveles intermedios (Hou et al., 2015; G. Xu et al., 2020).

No puede descartarse que existan funciones alternativas de estos neuropéptidos en este estadio adulto. Como ejemplo de esto puede marcarse la implicancia en procesos reproductivos por parte de ETH (S. S. Lee et al., 2017; Meiselman et al., 2017).

6.B.2 Expresión neuropeptidérgica en embriones tempranos

Para comparar la expresión de neuropéptidos se recabaron datos de expresión en embriones. En la mayoría de los casos no se encontraron series temporales en diversos momentos del desarrollo embrional (como las presentadas en este trabajo), pero si se hallaron comparaciones con etapas tardías del desarrollo (larval, pupal o adulta). Por esta razón se detallan comparaciones generales del estadio embrionario (a menos que se indique específicamente un momento) y se divide la información recabada en los neuropéptidos en la bibliografía disponible.

Hormona protoracicotrópica (PTTH)

En comparación con *O. fasciatus* donde éste neuropéptido mostró expresión embrional temprana, *P. puparum* mostró igual tendencia con valores mayores en los embriones respecto a los determinados en ovarios (Xu et al., 2020), mientras que en *D. citri* se comprobó que el pico de expresión medido por RT-qPCR estuvo presente en los embriones y ésta disminuye hasta llegar a la madurez sexual (Wang et al., 2018).

El mayor contraste de esta hormona se da con *D. melanogaster* que no presenta expresión transcriptómica en el embrión (datos evaluados en Flybase (Drysdale et al., 2005)).

Allatostatina (AST)

El perfil de expresión de allatostatina mostró que tanto el transcripto de allatostatina A, como allatostatina CC y CCC son detectables en ovarios y este se mantiene durante las primeras 30 horas (~25% del desarrollo embrional) sugiriendo un comportamiento semejante al encontrado en genes maternos (como bicoid, nanos o caudal) de *Drosophila melanogaster* (Rivera-Pomar y Jäckle, 1996).

En contraste a *O. fasciatus*, *D. melanogaster* expresa los genes de allatostatina A, B (péptido myoinhibitorio) y C luego de superar el 50% del desarrollo embrionario, en cambio Allatostatina CC no muestra expresión de ARN mensajero (medido por RNAseq) durante el desarrollo embrional, más sin embargo la interferencia del gen Allatostatina CC (perteneciente a la familia de Allatostatina) muestra un fenotipo letal embrional sugiriendo que este es un gen esencial para el correcto desarrollo (Veenstra, 2009).

Blatella germánica (Maestro et al., 2010) y *Diptera punctata* (Woodhead et al., 2003) muestran patrones similares a los descritos para *O. fasciatus* dado que se comprobó por ELISA (ensayo inmunoabsorbente ligado a enzimas) y Southern-blot la presencia detectable en bajas dosis en los primeros momentos del desarrollo, luego esta expresión decae y vuelve a expresarse tardíamente previo a la eclosión.

Neuropéptido activador de la biosíntesis de feromonas (PBAN)

El neuropéptido activador de la biosíntesis de feromonas no mostró señales de expresión temprana en *O. fasciatus*, coincidente con lo descrito para *Bombyx mori*, organismo cuya expresión se vuelve apreciable luego de superar el 50% del desarrollo (Xu et al., 1995).

Neuroparsina.

En *O. fasciatus* la neuroparsina presentó la mayor expresión de todos los neuropéptidos analizados en embriones tempranos. *D. citri* también mostró que la expresión disminuyó gradualmente desde el huevo hasta la madurez sexual (Wang et al., 2018) mientras que por el contrario *P. puparum* presentó una expresión baja (G. Xu et al., 2020).

Neuropéptidos relacionados con la ecdisis en embriones tempranos

Ewen-campen (2011) señalaba en la caracterización del transcriptoma embrional temprano que “[...]no se detectaron transcritos que codificaran las hormonas peptídicas de los insectos implicadas en el comportamiento de la ecdisis, como la hormona disparadora de la ecdisis, la hormona de la eclosión y el péptido cardioactivo de los crustáceos.”

El análisis realizado en este trabajo mostró baja expresión de todos los transcritos mencionados y mediante la identificación por espectrometría de masas tampoco fueron hallados.

6.C Proteómica

Cuando se compara la identificación de neuropéptidos a nivel peptídico en embriones y en cerebro adulto las principales diferencias encontradas incluyen los neuropéptidos encontrados, la cobertura de los mismos y la extensión en los péptidos identificados.

La diferencia de cuáles fueron identificados en cada caso es atribuible al momento del desarrollo y no son contrastables, igualmente se adjuntan los listados de neuropéptidos comunes a ambos momentos, aquellos que no fueron hallados y la tabla 14 que especifica cuales fueron encontrados en cada tejido estudiado (ver figura 32).

Neuropéptidos comunes a tejido embrional o cerebro adulto:

- AstA - Allatostatina A
- AstB - Allatostatina B
- CCAP - Péptido cardioactivo de crustáceos
- CNM - CNMamida
- C-ter - péptido C-terminal
- DH31 - Hormona diurética DH31
- FMRF - FMRF-amida
- ILP2 - péptido relacionado con la insulina 2
- MS - Miosupresina
- NPAA - Neuroparsina
- NPLP - Precursor tipo neuropéptido
- NVP – NVP
- Orc - Orcokinina
- PBAN - neuropéptido activador de la biosíntesis de feromonas
- SIF - SIF-amida
- TK - Taquicinina

Neuropéptidos no identificados ni en tejido embrional ni en cerebro adulto:

- GPA2 - Glicoproteína alfa 2
- Hormona de eclosión
- Hormona disparadora de la ecdisis
- Neuroligina
- Péptido transportador de iones
- NP1 - Neuroparsina 1
- NP2 - Neuroparsina 2
- NP3 - Neuroparsina 3
- NP4 - Neuroparsina 4

Tabla 14. Neuropéptidos hallados exclusivamente en tejido embrional o cerebro adulto.

Cerebro	Embrión
SNF - Neuropéptido F corto	Kin - Kinina
SK - Sulfaquinina	PTTH - Hormona protoracicotrópica
RYa - Ryamida	Crz - Corazonina
CCH - CCHamida	Burs - Bursicona alfa
NPF - Neuropeptido F largo	GPB5 - Glicoproteína beta 5
Ele - Elevenina	ILP2 - Péptido relacionado con la insulina 2
AstCCC - Allatostatina CCC	pBur - Bursicona-beta
AT - Allatotropina	
PDF - Factor dispersante de pigmentos	
ITG - ITG amida	
AKH - Hormona adipoquinética	
Proc - Proctolin	
DH44 - Factor liberador de corticotropina (CRF DH44)	
AstCC - Allatostatina CC	

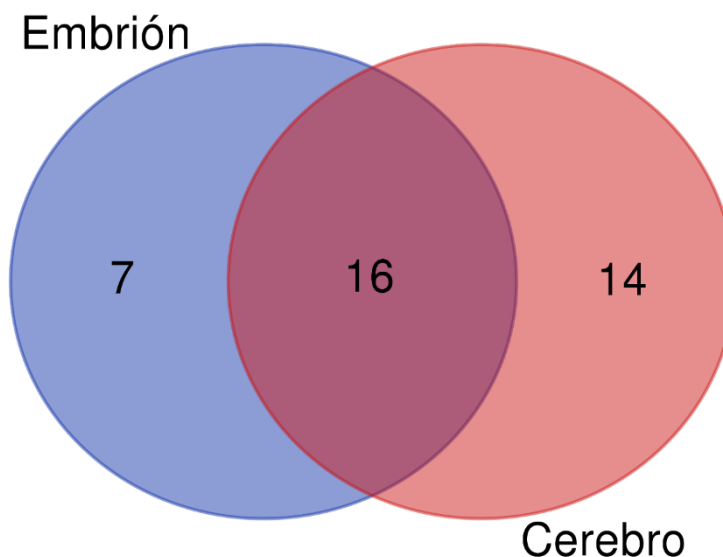


Figura 32. Diagrama de Venn del número de neuropéptidos identificados en cada tejido.

Respecto a los resultados de la cobertura y la extensión de los péptidos identificados en embriones temprano de *O. fasciatus* nuestra hipótesis de trabajo es que existen proteínas precursoras completas (pre-propéptidos) y no neuropéptidos maduros.

Los neuropéptidos maduros no fueron hallados como tal ya que los analitos identificados por espectrometría de masas en secuencias polipeptídicas presentaron de manera intacta los sitios dibásicos (previos y/o posteriores al péptido activo) y las glicinas (que ceden su grupo amino para amidar el extremo C-terminal) (ver figura 33). De esto se puede inferir que la fragmentación proteica es producto de la ionización y clivaje efectuado técnicamente y no por escisión proteolítica por dentro de la célula. Futura evidencia experimental complementaria convalidara o refutara esta hipótesis aquí expresada.

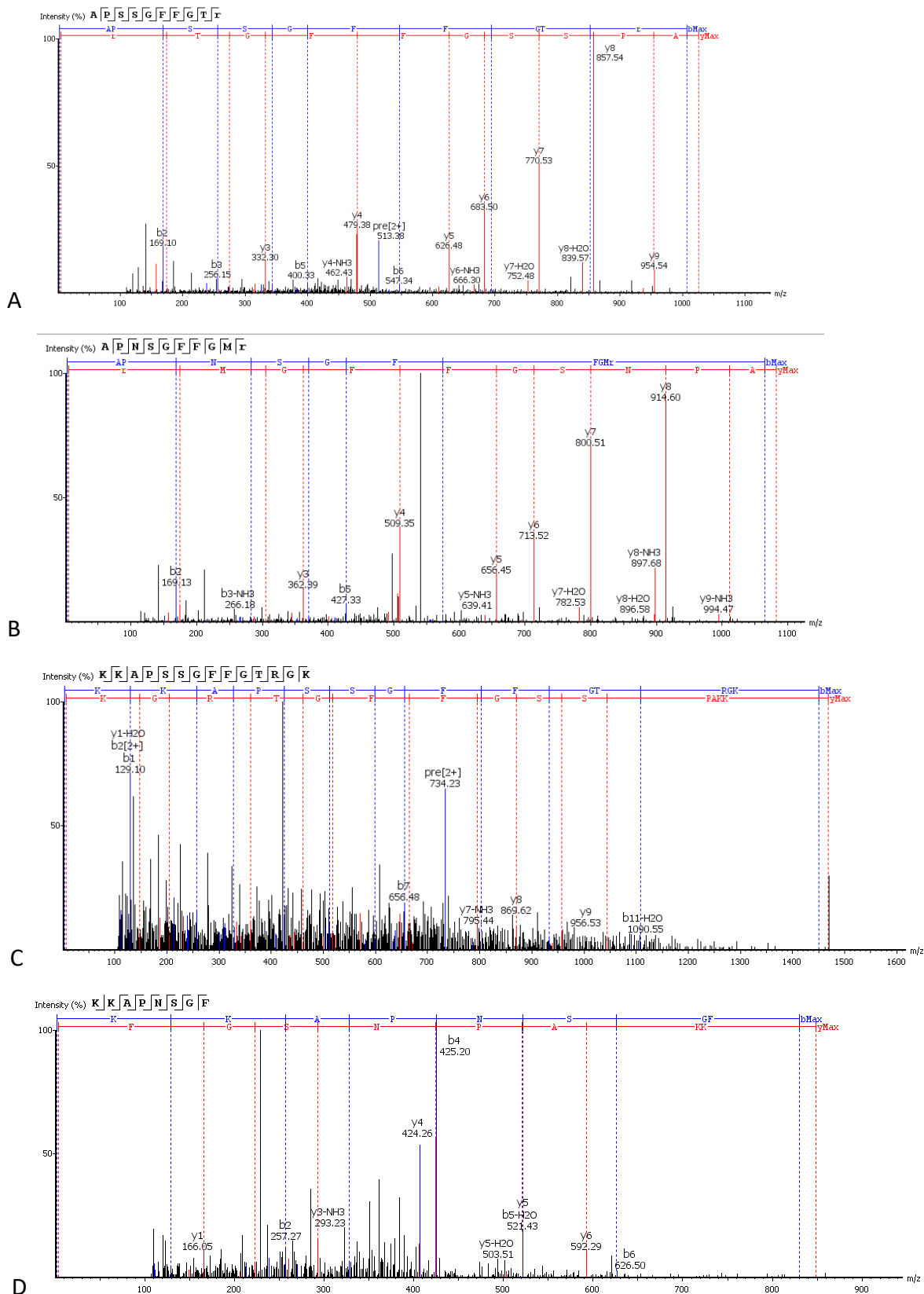


Figura 33. Espectros de masa identificados para A-B) péptido activo maduro de Taquicinina 3 y 4 obtenido de tejido cerebral adulto, en la parte superior de la imagen se observa la secuencia de aminoácidos. (r denota Arginina modificada por acetilación).

C-D) Pre-propéptido precursor de Taquicinina 3 y 4 obtenidos de muestras de embriones de hasta 24 horas, en la parte superior de la imagen se observa la secuencia de aminoácidos que consta de una secuencia dibásica de Lisinas (KK), la secuencia del péptido activo y una secuencia Glicina-Lisina (GK).

Estos datos abren paso al planteo de nuevas hipótesis de trabajo sugiriendo que las proteínas podrían cumplir un rol diferencial al conocido en estadios larvales o adultos, considerando que los datos analizados (embriones de hasta 24 horas post-oviposición) son coincidentes a la formación del blastodermo sincicial o celular y anteriores a la formación de la banda germinal y el sistema nervioso.

6.D. Consideraciones finales

Este es el primer trabajo de integración de datos de origen ómico que se realiza en *O. fasciatus* y que sienta las bases a nuevos estudios en la fisiología de los neuropéptidos.

Adicionalmente se abren las puertas a la exploración de las funciones putativas de las proteínas precursoras en embriones que, como se ha demostrado en esta tesis están presentes en niveles considerables de transcritos y proteínas, podrían ser pensadas como reservorios para etapas de desarrollo subsiguientes o, alternativamente, con funciones específicas en los embriones diferentes a las descritas hasta la fecha.

Futuros estudios experimentales que incluyan la localización de mensajeros por hibridación *in situ* o de proteínas por inmunohistoquímica en el embrión o la interferencia de mensajeros por ARN interferente permitirán esclarecer el rol de los neuropéptidos y proteínas precursoras aquí propuestos.

7. Bibliografía

- Abril, J. F., & Castellano, S. (2018). Genome annotation. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vols. 1–3, pp. 195–209). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20226-4>
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. In *Genome Biology* (Vol. 21, Issue 1, pp. 1–16). BioMed Central Ltd. <https://doi.org/10.1186/s13059-020-1935-5>
- Arakane, Y., Li, B., Muthukrishnan, S., Beeman, R. W., Kramer, K. J., & Park, Y. (2008). Functional analysis of four neuropeptides, EH, ETH, CCAP and bursicon, and their receptors in adult ecdysis behavior of the red flour beetle, *Tribolium castaneum*. *Mechanisms of Development*, *125*(11–12), 984–995. <https://doi.org/10.1016/j.mod.2008.09.002>
- Auman, T., & Chipman, A. D. (2018). Growth zone segmentation in the milkweed bug *Oncopeltus fasciatus* sheds light on the evolution of insect segmentation. *BioRxiv*, 327106. <https://doi.org/10.1101/327106>
- Auman, T., Vreede, B. M. I., Weiss, A., Hester, S. D., Williams, T. A., Nagy, L. M., & Chipman, A. D. (2017). Dynamics of growth zone patterning in the milkweed bug *Oncopeltus fasciatus*. *Development (Cambridge)*, *144*(10), 1896–1905. <https://doi.org/10.1242/dev.142091>
- Baggerman, G., Liu, F., Wets, G., & Schoofs, L. (2005). Bioinformatic analysis of peptide precursor proteins. *Annals of the New York Academy of Sciences*, *1040*, 59–65. <https://doi.org/10.1196/annals.1327.006>
- Bao, Y.-Y., Qu, L.-Y., Zhao, D., Chen, L.-B., Jin, H.-Y., Xu, L.-M., Cheng, J.-A., & Zhang, C.-X. (2013). The genome- and transcriptome-wide analysis of innate immunity in the brown planthopper, *Nilaparvata lugens*. *BMC Genomics*, *14*, 160. <https://doi.org/10.1186/1471-2164-14-160>
- Ben-David, J., & Chipman, A. D. (2010). Mutual regulatory interactions of the trunk gap genes during blastoderm patterning in the hemipteran *Oncopeltus fasciatus*. *Developmental Biology*, *346*(1), 140–149. <https://doi.org/10.1016/j.ydbio.2010.07.010>
- Bentley, D. R. (2006). Whole-genome re-sequencing. In *Current Opinion in Genetics and Development* (Vol. 16, Issue 6, pp. 545–552). Elsevier Current Trends. <https://doi.org/10.1016/j.gde.2006.10.009>

- Bewick, A. J., Sanchez, Z., McKinney, E. C., Moore, A. J., Moore, P. J., & Schmitz, R. J. (2019). Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, *Oncopeltus fasciatus* 06 Biological Sciences 0604 Genetics. *Epigenetics and Chromatin*, 12(1). <https://doi.org/10.1186/s13072-018-0246-5>
- Binzer, M., Heuer, C. M., Kollmann, M., Kahnt, J., Hauser, F., Grimmelikhuijzen, C. J. P., & Schachtner, J. (2014). Neuropeptidome of *Tribolium castaneum* antennal lobes and mushroom bodies. *Journal of Comparative Neurology*, 522(2), 337–357. <https://doi.org/10.1002/cne.23399>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., Wan, K. H., Yu, C., Zhang, D., Carlson, J. W., Cherbas, L., Eads, B. D., Miller, D., Mockaitis, K., Roberts, J., ... Celniker, S. E. (2014). Diversity and dynamics of the *Drosophila* transcriptome. *Nature*, 512(7515), 393–399. <https://doi.org/10.1038/nature12962>
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T. H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Peshkin, L., Tabin, C. J., ... Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Reports*, 18(3), 762–776. <https://doi.org/10.1016/j.celrep.2016.12.063>
- Burdfield-Steel, E. R., & Shuker, D. M. (2014). The evolutionary ecology of the Lygaeidae. *Ecology and Evolution*, 4(11), 2278–2301. <https://doi.org/10.1002/ece3.1093>
- Bushmanova, E., Antipov, D., Lapidus, A., Suvorov, V., & Prjibelski, A. D. (2016). RnaQUAST: A quality assessment tool for de novo transcriptome assemblies. *Bioinformatics*, 32(14), 2210–2212. <https://doi.org/10.1093/bioinformatics/btw218>
- Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline [version 2; referees: 5 approved]. *F1000Research*, 5, 1438. <https://doi.org/10.12688/F1000RESEARCH.8987.2>
- Chipman, A. D. (2017). *Oncopeltus fasciatus* as an evo-devo research organism. *Genesis*, 55(5), e23020. <https://doi.org/10.1002/dvg.23020>
- Clynen, E., De Loof, A., & Schoofs, L. (2003). The use of peptidomics in endocrine research. *General and Comparative Endocrinology*, 132(1), 1–9. [https://doi.org/10.1016/S0016-6480\(03\)00158-8](https://doi.org/10.1016/S0016-6480(03)00158-8)
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-seq Reads with STAR. *Current Protocols in Bioinformatics*, 51, 11.14.1-19. <https://doi.org/10.1002/0471250953.bi1114s51>

- Drysdale, R. A., Crosby, M. A., Gelbart, W., Campbell, K., Crosby, M., Emmert, D., Matthews, B., Russo, S., Schroeder, A., Smutniak, F., Zhang, P., Zhou, P., Zytkevich, M., Ashburner, M., Drysdale, R., de Grey, A., Foulger, R., Millburn, G., Sutherland, D., ... Shu, S. Q. (2005). FlyBase: Genes and gene models. *Nucleic Acids Research*, 33(DATABASE ISS.), D390–D395. <https://doi.org/10.1093/nar/gki046>
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, 7(10), 1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- Evans, J. D., Brown, S. J., Hackett, K. J. J., Robinson, G., Richards, S., Lawson, D., Elisk, C., Coddington, J., Edwards, O., Emrich, S., Gabaldon, T., Goldsmith, M., Hanes, G., Misof, B., Muñoz-Torres, M., Niehuis, O., Papanicolaou, A., Pfrender, M., Poelchau, M., ... Zhou, X. (2013). The i5K initiative: Advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, 104(5), 595–600. <https://doi.org/10.1093/jhered/est050>
- Ewels, P., Magnusson, M., Lundin, S., & Källner, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Ewen-Campen, B., Shaner, N., Panfilio, K. A., Suzuki, Y., Roth, S., & Extavour, C. G. (2011). The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genomics*, 12(January). <https://doi.org/10.1186/1471-2164-12-61>
- Feir, D. (1974). *Oncopeltus Fasciatus* : A Research Animal. *Annual Review of Entomology*, 19(1), 81–96. <https://doi.org/10.1146/annurev.en.19.010174.000501>
- Fields, S., & Johnston, M. (2005). Whither model organism research? In *Science* (Vol. 307, Issue 5717, pp. 1885–1886). American Association for the Advancement of Science. <https://doi.org/10.1126/science.1108872>
- Francischetti, I. M. B., Lopes, A. H., Dias, F. A., Pham, V. M., & Ribeiro, J. M. C. (2007). An insight into the sialotranscriptome of the seed-feeding bug, *Oncopeltus fasciatus*. *Insect Biochemistry and Molecular Biology*, 37(9), 903–910. <https://doi.org/10.1016/j.ibmb.2007.04.007>
- Gayon, J. (2016). From Mendel to epigenetics: History of genetics. *Comptes Rendus Biologies*, 339(7–8), 225–230. <https://doi.org/10.1016/J.CRVI.2016.05.009>
- Ginzburg, N., Cohen, M., & Chipman, A. D. (2017). Factors involved in early polarization of the anterior-posterior axis in the milkweed bug *Oncopeltus fasciatus*. *Genesis*, 55(5), 1–10. <https://doi.org/10.1002/dvg.23027>
- Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., & Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*, 261(June), 169–176. <https://doi.org/10.1016/j.jbiotec.2017.06.1198>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference

- genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hillman-Jackson, J., Clements, D., Blankenberg, D., Taylor, J., Nekrutenko, A., & Team, G. (2012). Using Galaxy to Perform Large-Scale Interactive Data Analyses. In *Current Protocols in Bioinformatics* (Vol. 38, Issue 1, pp. 10.5.1-10.5.47). John Wiley & Sons, Inc. <https://doi.org/10.1002/0471250953.bi1005s38>
- Hou, L., Jiang, F., Yang, P., Wang, X., & Kang, L. (2015). Molecular characterization and expression profiles of neuropeptide precursors in the migratory locust. *Insect Biochemistry and Molecular Biology*, 63, 63–71. <https://doi.org/10.1016/j.ibmb.2015.05.014>
- Huang, H. J., Lu, J. B., Li, Q., Bao, Y. Y., & Zhang, C. X. (2018). Combined transcriptomic/proteomic analysis of salivary gland and secreted saliva in three planthopper species. *Journal of Proteomics*, 172(September), 25–35. <https://doi.org/10.1016/j.jprot.2017.11.003>
- Kolde, R. (2015). *pheatmap: Pretty heatmaps [Software]*.
- Kriventseva, E. V, Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., Simão, S., & Zdobnov, E. M. (2018). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 47, 807–811. <https://doi.org/10.1093/nar/gky1053>
- Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3), 567–580. <https://doi.org/10.1006/jmbi.2000.4315>
- Lagesen, K., Hallin, P., Rødland, A., Staerfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357. <https://doi.org/10.1038/nmeth.1923>
- Lavore, A., Perez-Gianmarco, L., Esponda-Behrens, N., Palacio, V., Catalano, M. I., Rivera-Pomar, R., & Ons, S. (2018). *Nezara viridula* (Hemiptera: Pentatomidae) transcriptomic analysis and neuropeptidomics. *Scientific Reports*, 8(1), 17244. <https://doi.org/10.1038/s41598-018-35386-4>
- Lee, J. H., Choi, J. Y., Tao, X. Y., Kim, J. S., Kim, W., & Je, Y. H. (2013). Transcriptome Analysis of the Small Brown Planthopper, *Laodelphax striatellus* Carrying Rice stripe virus. *The Plant Pathology Journal*, 29(3), 330–337.

<https://doi.org/10.5423/PPJ.NT.01.2013.0001>

- Lee, S. S., Ding, Y., Karapetians, N., Rivera-Perez, C., Noriega, F. G., & Adams, M. E. (2017). Hormonal Signaling Cascade during an Early-Adult Critical Period Required for Courtship Memory Retention in *Drosophila*. *Current Biology*, *27*(18), 2798–2809.e3. <https://doi.org/10.1016/j.cub.2017.08.017>
- Liang, A. W., Zhang, H., Lin, J., & Wang, F. H. (2018). De Novo Assembly and Analysis of the White-Backed Planthopper (*Sogatella furcifera*) Transcriptome. *Journal of Insect Science (Online)*, *18*(4), 1–7. <https://doi.org/10.1093/jisesa/iey074>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Liao, Yang, Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, *47*(8), e47–e47.
- Liu, J., Lemonds, T. R., & Popadić, A. (2014). The genetic control of aposematic black pigmentation in hemimetabolous insects: Insights from *Oncopeltus fasciatus*. *Evolution and Development*, *16*(5), 270–277. <https://doi.org/10.1111/ede.12090>
- Liu, P. Z., & Kaufman, T. C. (2005). Short and long germ segmentation: Unanswered questions in the evolution of a developmental mode. *Evolution and Development*, *7*(6), 629–646. <https://doi.org/10.1111/j.1525-142X.2005.05066.x>
- Maestro, J. L., Pascual, N., Treiblmayr, K., Lozano, J., & Bellés, X. (2010). Juvenile hormone and allatostatins in the German cockroach embryo. *Insect Biochemistry and Molecular Biology*, *40*(9), 660–665. <https://doi.org/10.1016/j.ibmb.2010.06.006>
- Manzoni, C., Kia, D. A., Vandrovcova, J., Hardy, J., Wood, N. W., Lewis, P. A., & Ferrari, R. (2018). Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, *19*(2), 286–302. <https://doi.org/10.1093/BIB/BBW114>
- Marchant, A., Mougél, F., Almeida, C., Jacquín-Joly, E., Costa, J., & Harry, M. (2015). De novo transcriptome assembly for a non-model species, the blood-sucking bug *Triatoma brasiliensis*, a vector of Chagas disease. *Genetica*, *143*(2), 225–239. <https://doi.org/10.1007/s10709-014-9790-5>
- Meiselman, M., Lee, S. S., Tran, R. T., Dai, H., Ding, Y., Rivera-Perez, C., Wijesekera, T. P., Dauwalder, B., Noriega, F. G., & Adams, M. E. (2017). Endocrine network essential for reproductive success in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(19), E3849–E3858. <https://doi.org/10.1073/pnas.1620760114>
- Mendel, G. (1941). Versuche über Pflanzen-Hybriden. *Der Züchter*, *13*(10–11), 221–268. <https://doi.org/10.1007/BF01804628>
- Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., Spieth, J., Carvalho, A. B., Panzera, F., Lawson, D., Torres, A. Q., Ribeiro, J. M. C., Sorgine, M. H. F., Waterhouse, R. M., Montague, M. J., Abad-Franch, F.,

- Alves-Bezerra, M., Amaral, L. R., Araujo, H. M., ... Oliveira, P. L. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(48), 14936–14941. <https://doi.org/10.1073/pnas.1506226112>
- Neupert, S., Russell, W. K., Russell, D. H., López, J. D., Predel, R., & Nachman, R. J. (2009). Neuropeptides in Heteroptera: Identification of allatotropin-related peptide and tachykinin-related peptides using MALDI-TOF mass spectrometry. *Peptides*, *30*(3), 483–488. <https://doi.org/10.1016/j.peptides.2008.11.009>
- Nüsslein-volhard, C., & Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, *287*(5785), 795–801. <https://doi.org/10.1038/287795a0>
- Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292–294. <https://doi.org/10.1093/bioinformatics/btv566>
- Ons, S., Lavore, A., Sterkel, M., Wulff, J. P., Sierra, I., Martínez-Barnette, J., Rodríguez, M. H., & Rivera-Pomar, R. (2016). Identification of G protein coupled receptors for opsins and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis. *Insect Biochemistry and Molecular Biology*, *69*(May), 34–50. <https://doi.org/10.1016/j.ibmb.2015.05.003>
- Ons, S., Richter, F., Urlaub, H., & Pomar, R. R. (2009). The neuropeptidome of *Rhodnius prolixus* brain. *Proteomics*, *9*(3), 788–792. <https://doi.org/10.1002/pmic.200800499>
- Ons, S., Sterkel, M., Diambra, L., Urlaub, H., & Rivera-Pomar, R. (2011). Neuropeptide precursor gene discovery in the Chagas disease vector *Rhodnius prolixus*. *Insect Molecular Biology*, *20*(1), 29–44. <https://doi.org/10.1111/j.1365-2583.2010.01050.x>
- Palacio, V. (2018). *Secuenciación del transcriptoma de Dalbulus maidis*. Universidad Nacional del Noroeste de Buenos Aires.
- Panfilio, K. A. (2008). Extraembryonic development in insects and the acrobatics of blastokinesis. In *Developmental Biology*. <https://doi.org/10.1016/j.ydbio.2007.11.004>
- Panfilio, K. A., & Angelini, D. R. (2018). By land, air, and sea: hemipteran diversity through the genomic lens. In *Current Opinion in Insect Science* (Vol. 25, pp. 106–115). Elsevier. <https://doi.org/10.1016/j.cois.2017.12.005>
- Panfilio, K. A., Liu, P. Z., Akam, M., & Kaufman, T. C. (2006). *Oncopeltus fasciatus* zen is essential for serosal tissue function in katatrepsis. *Developmental Biology*, *292*(1), 226–243. <https://doi.org/10.1016/j.ydbio.2005.12.028>
- Panfilio, K. A., Vargas Jentzsch, I. M., Benoit, J. B., Erezylmaz, D., Suzuki, Y., Colella, S., Robertson, H. M., Poelchau, M. F., Waterhouse, R. M., Ioannidis, P., Weirauch, M. T., Hughes, D. S. T., Murali, S. C., Werren, J. H., Jacobs, C. G. C., Duncan, E. J., Armisen, D., Vreede, B. M. I., Baa-Puyoulet, P., ... Richards, S. (2018). Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the

milkweed bug genome. *BioRxiv*, 201731. <https://doi.org/10.1101/201731>

- Panfilio, K. A., Vargas Jentzsch, I. M., Benoit, J. B., Ereyilmaz, D., Suzuki, Y., Colella, S., Robertson, H. M., Poelchau, M. F., Waterhouse, R. M., Ioannidis, P., Weirauch, M. T., Hughes, D. S. T., Murali, S. C., Werren, J. H., Jacobs, C. G. C., Duncan, E. J., Armisén, D., Vreede, B. M. I., Baa-Puyoulet, P., ... Richards, S. (2019). Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biology*, 20(1), 64. <https://doi.org/10.1186/s13059-019-1660-0>
- Pascual, A. (2019). *Genómica del desarrollo embrionario de Rhodnius prolixus*. Universidad de La Plata.
- Pauls, D., Chen, J., Reiher, W., Vanselow, J. T., Schlosser, A., Kahnt, J., & Wegener, C. (2014). Peptidomics and processing of regulatory peptides in the fruit fly *Drosophila*. *EuPA Open Proteomics*, 3, 114–127. <https://doi.org/10.1016/j.euprot.2014.02.007>
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290–295. <https://doi.org/10.1038/nbt.3122>
- Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C. Y., Lin, H., Lin, J. W., & Hackett, K. (2015). The i5k Workspace@NAL-enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, 43(D1), D714–D719. <https://doi.org/10.1093/nar/gku983>
- Predel, R., Kellner, R., Rapus, J., Penzlin, H., & Gáde, G. (1997). Isolation and structural elucidation of eight kinins from the retrocerebral complex of the American cockroach, *Periplaneta americana*. *Regulatory Peptides*, 71(3), 199–205. [https://doi.org/10.1016/S0167-0115\(97\)01029-X](https://doi.org/10.1016/S0167-0115(97)01029-X)
- Resh, V. H., & Cardé, R. (2009). Encyclopedia of Insects. In *Encyclopedia of Insects* (pp. 56–57). <https://doi.org/10.1016/B978-0-12-374144-8.X0001-X>
- Ribeiro, J. M. C., Genta, F. A., Sorgine, M. H. F., Logullo, R., Mesquita, R. D., Paiva-Silva, G. O., Majerowicz, D., Medeiros, M., Koerich, L., Terra, W. R., Ferreira, C., Pimentel, A. C., Bisch, P. M., Leite, D. C., Diniz, M. M. P., da S G V Junior, J. L., Da Silva, M. L., Araujo, R. N., Gandara, A. C. P., ... Oliveira, P. L. (2014). An insight into the transcriptome of the digestive tract of the bloodsucking bug, *Rhodnius prolixus*. *PLoS Neglected Tropical Diseases*, 8(1), e2594. <https://doi.org/10.1371/journal.pntd.0002594>
- Rivera-Pomar, R., & Jäckle, H. (1996). From gradients to stripes in *Drosophila* embryogenesis: Filling in the gaps. In *Trends in Genetics* (Vol. 12, Issue 11, pp. 478–483). Elsevier Ltd. [https://doi.org/10.1016/0168-9525\(96\)10044-5](https://doi.org/10.1016/0168-9525(96)10044-5)
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140.
- Schäffer, A. A., Nawrocki, E. P., Choi, Y., Kitts, P. A., Karsch-Mizrachi, I., & McVeigh, R. (2018). VecScreen-plus-taxonomy: Imposing a tax(onomy) increase on vector

- contamination screening. *Bioinformatics*, 34(5), 755–759.
<https://doi.org/10.1093/bioinformatics/btx669>
- Schoofs, L., De Loof, A., & Van Hiel, M. B. (2017). Neuropeptides as Regulators of Behavior in Insects. In *Annual Review of Entomology* (Vol. 62, Issue 1, pp. 35–52). Annual Reviews Inc. <https://doi.org/10.1146/annurev-ento-031616-035500>
- Schulz-Knappe, P., Hans-Dieter, Z., Heine, G., Jurgens, M., & Schrader, M. (2012). Peptidomics The Comprehensive Analysis of Peptides in Complex Biological Mixtures. *Combinatorial Chemistry & High Throughput Screening*, 4(2), 207–217.
<https://doi.org/10.2174/1386207013331246>
- Schwarz, A., Medrano-Mercado, N., Schaub, G. A., Struchiner, C. J., Bargues, M. D., Levy, M. Z., & Ribeiro, J. M. C. (2014). An Updated Insight into the Sialotranscriptome of *Triatoma infestans*: Developmental Stage and Geographic Variations. *PLoS Neglected Tropical Diseases*, 8(12), e3372. <https://doi.org/10.1371/journal.pntd.0003372>
- Sepey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. In *Methods in Molecular Biology* (Vol. 1962, pp. 227–245). Humana Press Inc. https://doi.org/10.1007/978-1-4939-9173-0_14
- Sterkel, M. (2011). *Genómica y proteómica de neuropéptidos en Rhodnius prolixus* [Universidad Nacional de Quilmes].
<https://ridaa.unq.edu.ar/handle/20.500.11807/123>
- Team, R. (2020). *R: a language and environment for statistical computing. Version 3.6. 3.*
- Team, Rs. (2020). RStudio: Integrated development for R. RStudio Inc.(2015).
Acknowledgements This Work Was Supported by Fundação Para a Ciência e Tecnologia (FCT), Azores.
- Tran, N. H., Qiao, R., Xin, L., Chen, X., Liu, C., Zhang, X., Shan, B., Ghodsi, A., & Li, M. (2019). Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nature Methods*, 16(1), 63–66.
<https://doi.org/10.1038/s41592-018-0260-3>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), 511–515.
<https://doi.org/10.1038/nbt.1621>
- Veenstra, J. A. (2009). Allatostatin C and its paralog allatostatin double C: The arthropod somatostatins. *Insect Biochemistry and Molecular Biology*, 39(3), 161–170.
<https://doi.org/10.1016/j.ibmb.2008.10.014>
- Veenstra, J. A. (2016). Allatostatins C, double C and triple C, the result of a local gene triplication in an ancestral arthropod. *General and Comparative Endocrinology*, 230–231, 153–157. <https://doi.org/10.1016/j.ygcen.2016.04.013>
- Veenstra, J. A. (2019). Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. *PeerJ*, 2019(6), e7144.
<https://doi.org/10.7717/peerj.7144>

- Wang, Z., Zhou, W., Hameed, M. S., Liu, J., & Zeng, X. (2018). Characterization and expression profiling of neuropeptides and g-protein-coupled receptors (GPCRs) for neuropeptides in the asian citrus psyllid, *diaphorina citri* (hemiptera: Psyllidae). *International Journal of Molecular Sciences*, *19*(12).
<https://doi.org/10.3390/ijms19123912>
- Wickham, H. (2009). Ggplot2. In *Ggplot2* (Vol. 4, Issue 1). <https://doi.org/10.1007/978-0-387-98141-3>
- Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, *7*, 1338.
<https://doi.org/10.12688/f1000research.15931.2>
- Woodhead, A. P., Thompson, M. E., Chan, K. K., & Stay, B. (2003). Allatostatin in ovaries, oviducts, and young embryos in the cockroach *Diploptera punctata*. *Journal of Insect Physiology*, *49*(12), 1103–1114. <https://doi.org/10.1016/j.jinsphys.2003.08.006>
- Xu, G., Teng, Z. W., Gu, G. X., Qi, Y. X., Guo, L., Xiao, S., Wang, F., Fang, Q., Wang, F., Song, Q. S., Stanley, D., & Ye, G. Y. (2020). Genome-wide characterization and transcriptomic analyses of neuropeptides and their receptors in an endoparasitoid wasp, *Pteromalus puparum*. *Archives of Insect Biochemistry and Physiology*, *103*(2).
<https://doi.org/10.1002/arch.21625>
- Xu, W. H., Sato, Y., Ikeda, M., & Yamashita, O. (1995). Stage-dependent and temperature-controlled expression of the gene encoding the precursor protein of diapause hormone and pheromone biosynthesis activating neuropeptide in the silkworm, *Bombyx mori*. In *Journal of Biological Chemistry* (Vol. 270, Issue 8, pp. 3804–3808).
<https://doi.org/10.1074/jbc.270.8.3804>
- Yamaguchi, M., & Yoshida, H. (2018). Drosophila as a model organism. *Advances in Experimental Medicine and Biology*, *1076*(5330), 1–10. https://doi.org/10.1007/978-981-13-0529-0_1
- Yeoh, J. G. C., Pandit, A. A., Zandawala, M., Nässel, D. R., Davies, S. A., & Dow, J. A. T. (2017). DINeR: Database for Insect Neuropeptide Research. *Insect Biochemistry and Molecular Biology*, *86*, 9–19. <https://doi.org/10.1016/j.ibmb.2017.05.001>
- Yokoi, K., Tsubota, T., Sun, J., Jouraku, A., Sezutsu, H., & Bono, H. (2019). Reference transcriptome data in silkworm *Bombyx mori*. *BioRxiv*, *14*, 1–22.
<https://doi.org/10.1101/805978>

8. Anexo.

8.A Depleción de ARN ribosómico

Al momento de evaluarse los controles de calidad, se evidenció la presencia de un alto contenido de guanina y citosina, indicio de una posible contaminación por ARN ribosómico. Por esta razón se procedió con la depleción bioinformática de estos contaminantes.

Para realizar este procedimiento se constituyó una serie de bases de datos que incluyeron datos de:

- ARN ribosómico proveniente de laboratorio Silva para Eucariotas y Procariotas, al que se le incluyeron las secuencias de ARN ribosómico de *O. fasciatus*
- El genoma de *Escherichia coli*.
- Adaptadores de secuenciación comúnmente encontrados en la tecnología Illumina.
- Controles internos de secuenciación, como PhiX.
- Secuencias comúnmente encontradas en contaminaciones de laboratorio procedentes de la base de datos de NCBI Univec.

Una vez obtenidas las bases de datos se realizó un índice para cada una de ellas. Posteriormente las *reads* fueron alineadas a todas las bases y, de esta manera, se evaluó la presencia de contaminantes evaluados.

Como puede verse en la figura 33, se halló una gran contaminación de los datos por ARN ribosómico del propio organismo. Una vez identificadas el mismo programa separó esas lecturas y devolvió un set enriquecido con lecturas correspondientes a ARN mensajeros codificantes.

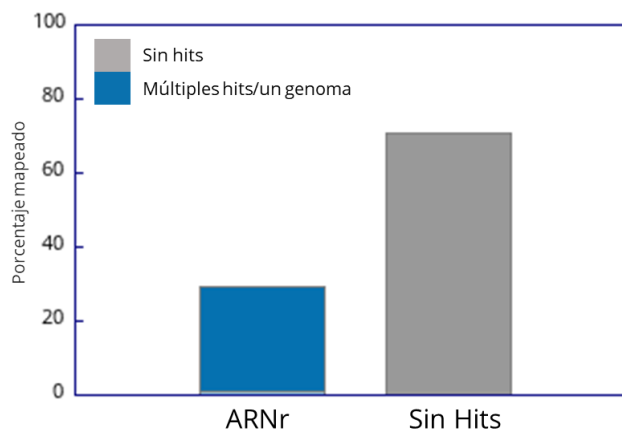


Figura 33. Porcentaje de reads mapeadas a la base de datos de ARNr.

En este caso se pudo determinar que alrededor del 28% de las *reads* se correspondían con ARN ribosómico.

Paralelamente se evaluó la presencia detallada en el resto de las bases de datos analizadas y se determinó que menos del uno por ciento de los datos podrían proceder de secuencias de otras fuentes de contaminación por lo cual no fueron eliminadas.

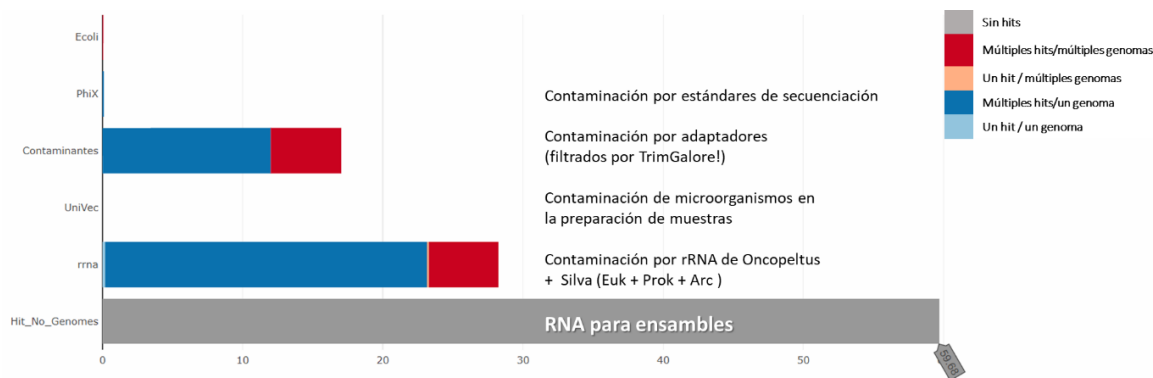


Figura 34. Distribución porcentual de *reads* en función de diversas bases de datos.

En el último paso FastQScreen devuelve los pares completos de lecturas que no sean coincidentes con los ARN ribosómicos. Los adaptadores de secuenciación representan subsecuencias de las *reads*, por lo cual se utilizó “Trim Galore!” especializado en la remoción parcial y específica de estas subsecuencias y de los vectores de amplificación. Estas *reads* serán utilizadas en los análisis subsecuentes.

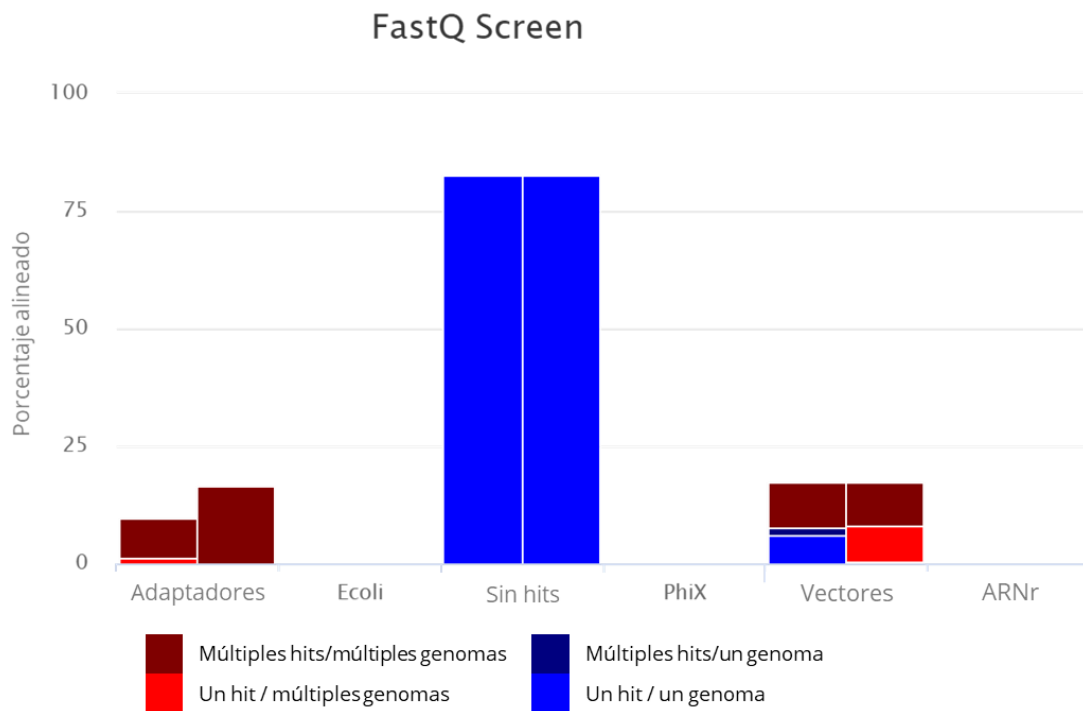


Figura 35 distribución porcentual de *reads* en función de diversas bases de datos. Se muestra la depleción *in silico* de ARNr.

8.A.2 Comparación de calidad en el pre-procesado.

Para evaluar la eficiencia de las correcciones en los pasos de depleción de ARNr y adaptadores se realizó una comparación de calidad. Se corroboró la disminución en la cantidad de *reads*, la mejora en la calidad de las muestras (*Phred Score*), la eliminación de adaptadores y la corrección de los valores de contenido porcentual de GC. Éstos últimos, de alrededor del 40%, resultaron coincidentes con lo establecido en el proyecto de secuenciación del genoma (Panfilio et al., 2018). Se muestran los resultados en las tablas y gráficos presentados a continuación.

Tabla 15. Medidas de resumen de calidad. Ofas-RNA: muestras originales, Ofas-filtrado: luego de la depleción de ARNr, Trim-Galore-Ofas: *reads* filtradas y sin adaptadores.

Nombre de la muestra	% Duplicación	% GC	Longitud	Millones de Secuencias
Ofas-RNA_1.fastq.gz	81.70%	45%	151 bp	211.5
Ofas-RNA_2.fastq.gz	76.70%	45%	151 bp	211.5
Ofas_1_filtrado.fastq.gz	64.50%	41%	151 bp	150
Ofas_2_filtrado.fastq.gz	64.50%	41%	151 bp	150
Trim Galore_Ofas_1_filtrado.fastq.gz	64.10%	40%	138 bp (20-151)	149.9
Trim GaloreOfas_2_filtrado.fastq.gz	64.20%	40%	138 bp (20-151)	149.9

FastQC: Cantidad de Secuencias

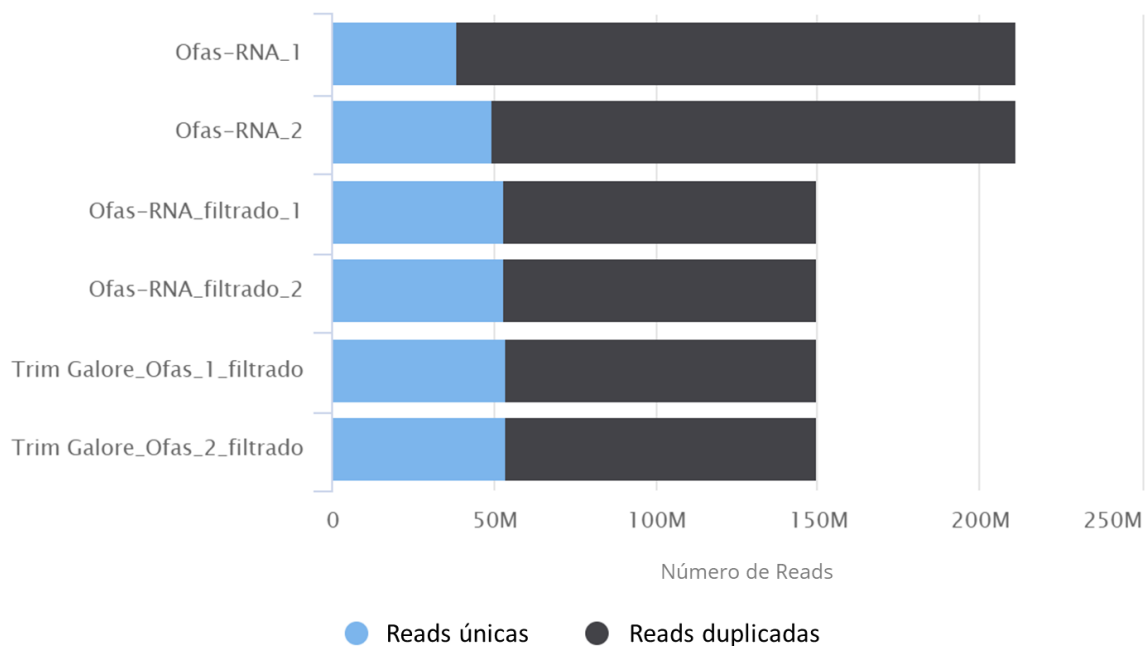


Figura 36. Cantidad de secuencias presentes en cada etapa. Ofas-RNA: muestras originales, Ofas-filtrado: luego de la depleción de ARNr, Trim-Galore-Ofas: *reads* filtradas y sin adaptadores.

FastQC: Puntaje de calidad por Secuencia

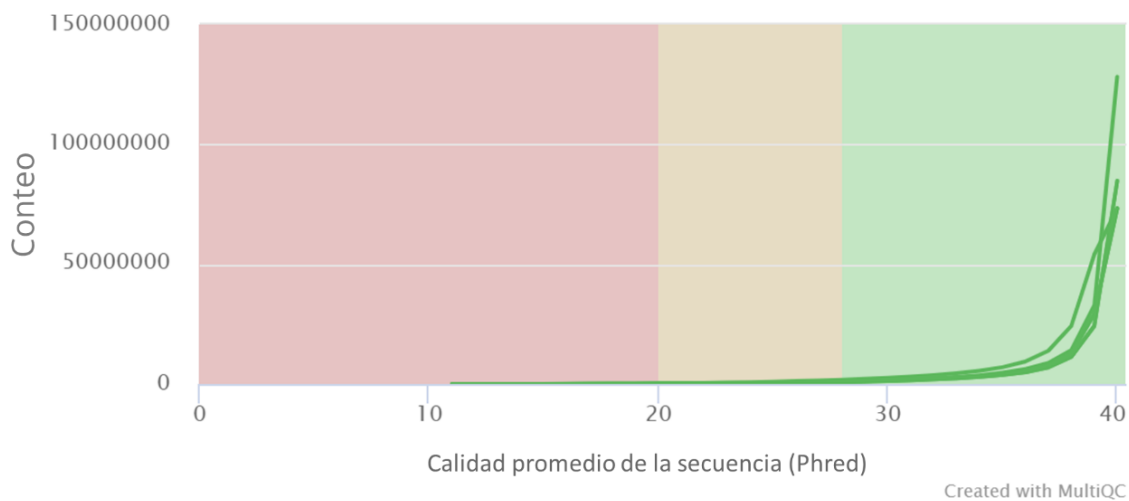


Figura 37. Calidad de secuenciación (phred score). Millones de secuencia por valor de calidad promedio.

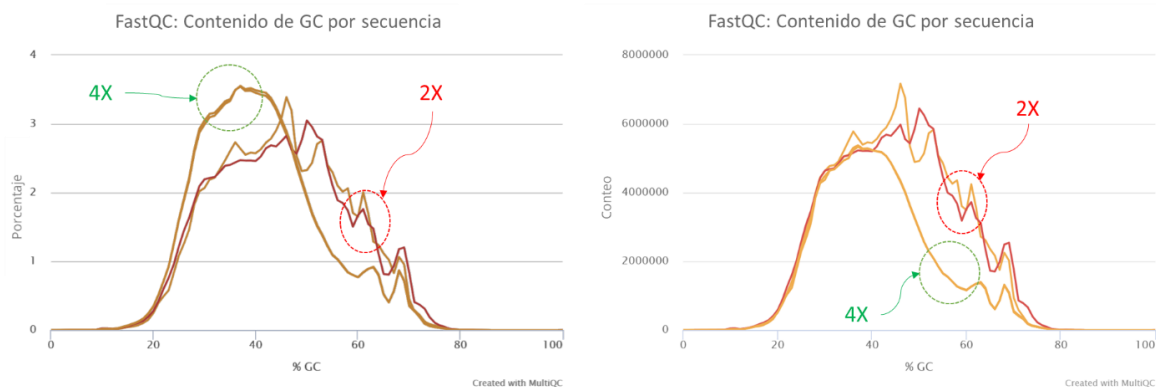


Figura 38. Contenido de GC. Izquierda: Distribución de los valores porcentuales de GC. Derecha: Distribución de la cantidad de GC en el total de reads. Se marcan en verde las reads luego del filtrado de calidad y en rojo sin correcciones.

FastQC: Contenido de Adaptadores

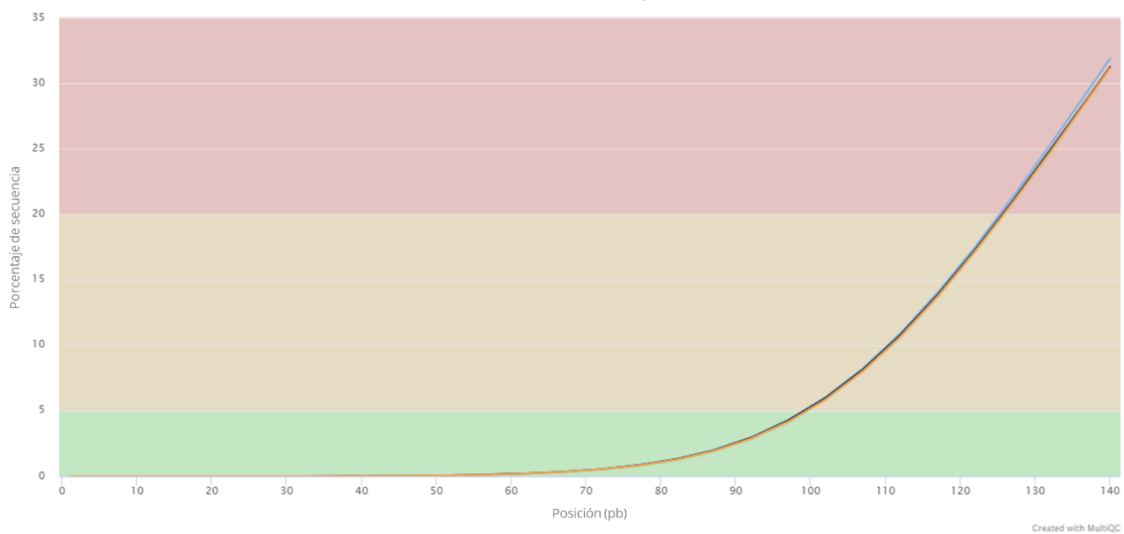


Figura 39. Presencia adaptadores. En celeste y naranja reads originales, en negro las reads luego de la depleción de ARNr. Las reads limpias no presentan contaminación.

Se puede observar en la figura 40 que este procedimiento redujo la cantidad de secuencias sobrerrepresentadas de más del 5% antes del filtrado, a menos del 1% luego del mismo llegando a niveles recomendables.

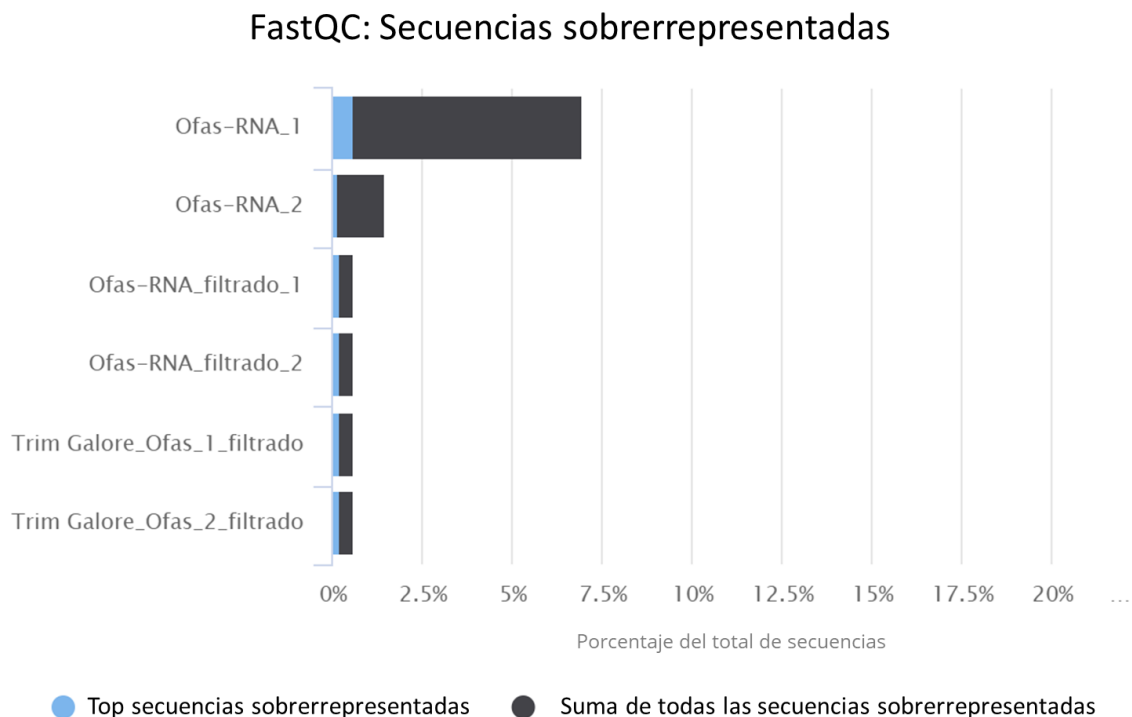


Figura 40. Porcentaje de secuencias sobrerrepresentadas. Ofas-RNA: muestras originales, Ofas-filtrado: luego de la depleción de ARNr, Trim-Galore-Ofas: *reads* filtradas y sin adaptadores

8.B Ensamble *de novo* y caracterización del transcriptoma de cerebro

Con las lecturas de secuenciación filtradas se procedió al ensamble *de novo* del transcriptoma de *Oncopeltus fasciatus*. El mismo fue llevado a cabo en Galaxy dados los requerimientos técnicos de este proceso.

Este software consta de un primer paso de normalización de datos dada la profundidad de estos. Este paso tiene como propósito mantener representado la proporción de bases al tiempo que reduce el tamaño del archivo a procesar.

Se partió de 150 millones de pares de lecturas de entre 20 y 151 pares de bases al inicio del experimento (aproximadamente 40.000 millones de bases). Fueron ensamblados 1.1 millones de transcriptos que contaron compuestos de 650 millones de bases (~1,46%). De los transcriptos se desprenden aproximadamente un millón de 'genes' con una longitud de entre 200 y 45.000 bases y una longitud promedio de 665 bases ($N50 \pm 586$ d.e.). Ver tabla 16 y gráfico 41.

Cabe destacar que se realizó un ensamble sin depleción de lecturas correspondientes ARN ribosómico. Se puede apreciar como diferencia que el ensamble sin filtrado brindó 7000 (~0.6%) transcriptos de diferencia manteniéndose el resto de los valores estadísticos asociados y la diferencia numéricas entre ambos ensambles resultó despreciable.

Tabla 16. Comparación de estadísticas y descriptores para muestras con o sin depleción bioinformática de ARN ribosómico.

	ARNr Depletado		"Contaminado" con ARNr	
Total Trinity 'genes':	1.088.927		1.094.040	
Total trinity transcritos:	1.179.126		1.186.275	
Porcentaje GC:	33.59 %		33.6 %	
Estadísticas basadas en: →	Contigs de todos los transcritos	Solo la isoforma más larga por "Gen"	Contigs de todos los transcritos	Solo la isoforma más larga por "Gen"
Contig N10	2414	1762	2408	1756
Contig N20	1491	1180	1490	1178
Contig N30	1082	900	1078	897
Contig N40	836	719	834	717
Contig N50	665	584	663	583
Largo de contig (mediana)	385	373	385	373
Contig Promedio	557.53	509.59	557.16	509.28
Total bases ensambladas	657.393.540 (1,46%)	554.905.366	660.947.260	557.178.124

Histograma de distribución Longitud del Isotig

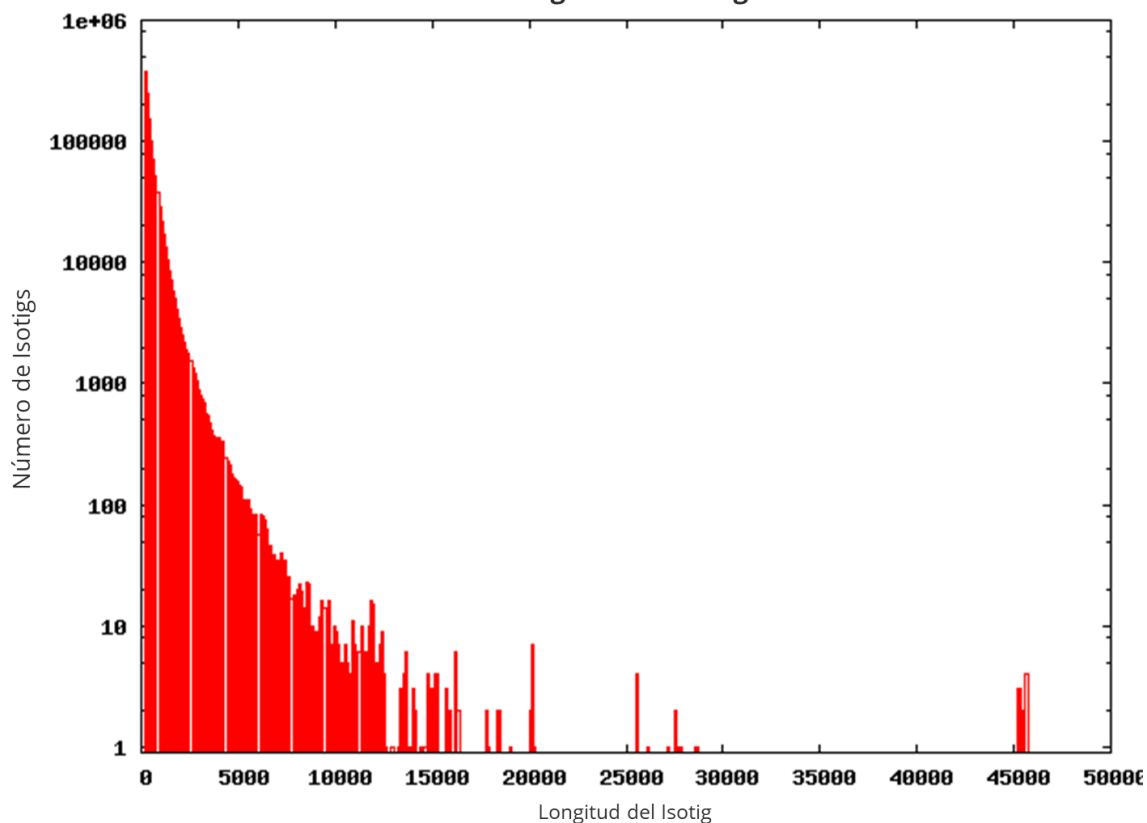


Figura 41. Histograma de distribución de longitud de los 'Isotigs'.

8.C Medidas de calidad del ensamble

Una vez finalizado el ensamble se procedió con la determinación de la calidad de este. Este proceso se ejecutó con BUSCO y RNAQuast.

BUSCO (Seppey et al., 2019) intenta proporcionar una evaluación cuantitativa de la completitud en términos del contenido genético esperado de un conjunto de genoma, transcriptoma o conjunto de genes anotado. Compara los transcritos obtenidos mediante el ensamble *de novo* con una base de datos compuesta por perfiles de Márkov de un consenso de genes caracterizados como “núcleos” para el grupo en el cual se realiza la comparación. A su vez, determina si dichos ortólogos se encuentran completos, fragmentados o perdidos.

En este caso se procedió a comparar el transcriptoma con cuatro niveles de pertenencia en la clasificación taxonómica de *O. fasciatus* disponibles en OrthoDB v10.

Los niveles evaluados en este análisis corresponden a eucariota, artrópoda, insecta y hemíptera. Siendo el 96,4%, el 94,9%, el 96,2% y el 97% de cobertura para cada nivel respectivamente para la versión 10 de las bases de datos (Kriventseva et al., 2018). Estos resultados muestran que los genes núcleos de cada estrato se encuentran correctamente representados y el ensamble puede considerarse exitoso.

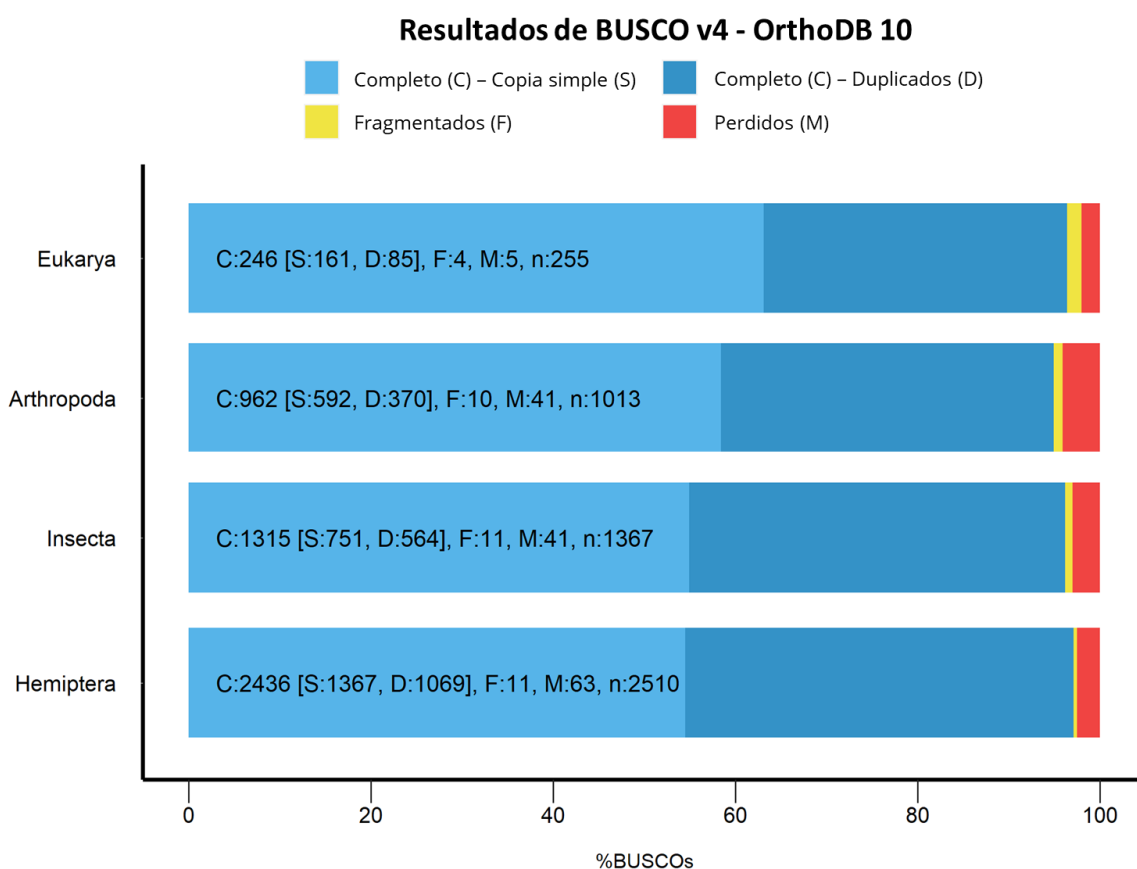


Figura 42. Porcentaje de Ortólogos cubiertos por cada nivel. C = completos = Simples + duplicados.

RNAQuast (Bushmanova et al., 2016), a diferencia de BUSCO, utiliza las anotaciones del genoma y, mediante alineamientos y cuantificación de los mismos permiten determinar entre otras cosas, la cobertura de la base de datos los genes ensamblados o cubiertos y la presencia de isoformas en el transcriptomas respecto al genoma anotado. Cómo se resume en la tabla 17 aproximadamente dos tercios de la base de datos fue cubierta, se destaca que en promedio existen 6 transcritos por isoforma, aproximadamente 30,000 fueron emparejados (*match*) en por lo menos la mitad de su longitud mientras que 10.000 genes presentaron un emparejamiento mínimo del 95%.

Al momento de evaluarse RNAQuast (Bushmanova et al., 2016) se obtuvieron estadísticas similares a las descritas en el punto anterior 8.B y este análisis predice la presencia de aproximadamente 30000 genes codificantes. Ver tabla 17.

Tabla 17. Métricas de Sensibilidad y sensibilidad del ensamble.

METRICAS / TRANSCRIPTO	Trinity_Ofas
== INTEGRIDAD DEL ENSAMBLE (SENSIBILIDAD) ==	
Cobertura de la base de datos	0.67
Proporción de duplicación	2.078
Promedio de isoformas por transcripto	6.326
Genes ensamblados (50%)	8393
Genes ensamblados (95%)	2987
Genes cubiertos (50%)	11879
Genes cubiertos (95%)	5681
== ESPECIFICIDAD DEL ENSAMBLE ==	
Emparejamiento (match) (50%)	29336
Emparejamiento (match) (95%)	10037

8.D Anotación automática de transcritos

Con la certeza de contar con un transcriptoma apropiadamente ensamblado se llevó adelante la identificación de transcritos codificantes. En este paso se utilizaron trinotate (Bryant et al., 2017) y trinotateR.

En el procedimiento realizado por trinotate (Bryant et al., 2017) se incluyó la identificación por homología de secuencias mediante BLAST (Altschul et al., 1990), la identificación de dominios proteicos a través de modelos ocultos de Márkov por HMMER (Eddy, 2011), la presencia de péptido señal por signalP (Almagro Armenteros et al., 2019), la identificación de pertenencia a vías metabólicas, los procesos y/o la localización celular mediante términos GO (*gene ontology*, ontología génica o 'atributos del producto génico').

Trinotate (Bryant et al., 2017) genera un archivo símil base de datos para facilitar su recorrido, filtrado e identificación de genes de interés.

La clasificación de estas proteínas permitió establecer que los principales procesos en los que se encuentran involucrados incluyen:

- Transporte intracelular
- Modificación postraduccional y plegado proteico
- Mecanismos de transducción de señales
- Transcripción

Todos estos coinciden con las funciones principales de neuronas y otras células del sistema nervioso central.

Se pueden destacar de la figura 45 que los componentes celulares incluyen una alta proporción de transcritos involucrados en cuerpos neuronales, dendritas, axones, sinapsis celular o membranas postsinápticas. También se desprende de esta tabla que los procesos biológicos en los que participan incluyen el guiado axonal, la transmisión sináptica, el desarrollo del sistema nervioso, la neurogénesis o la morfogénesis de dendritas. Esto permite establecer que el ensamble fue exitosamente completado.

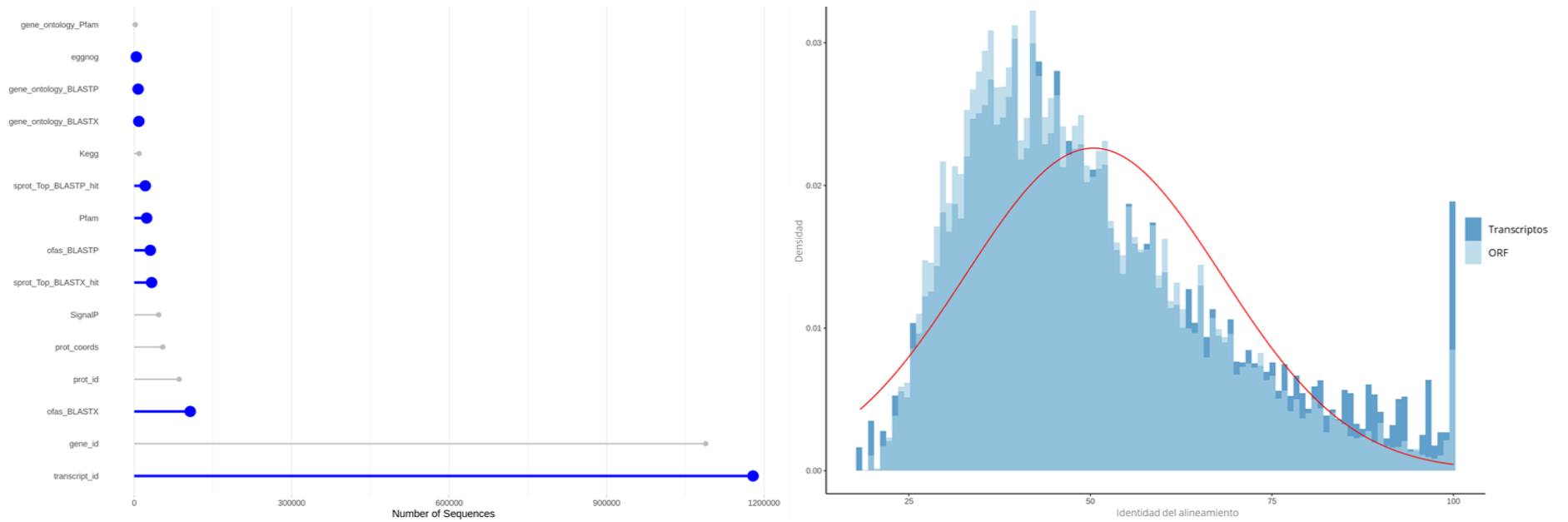


Figura 43. Izquierda: Distribución del número de secuencias identificadas por Trinotate en sus diferentes instancias de caracterización. Derecha: Distribución de la identidad de los alineamientos de BLAST identificados por Trinotate.

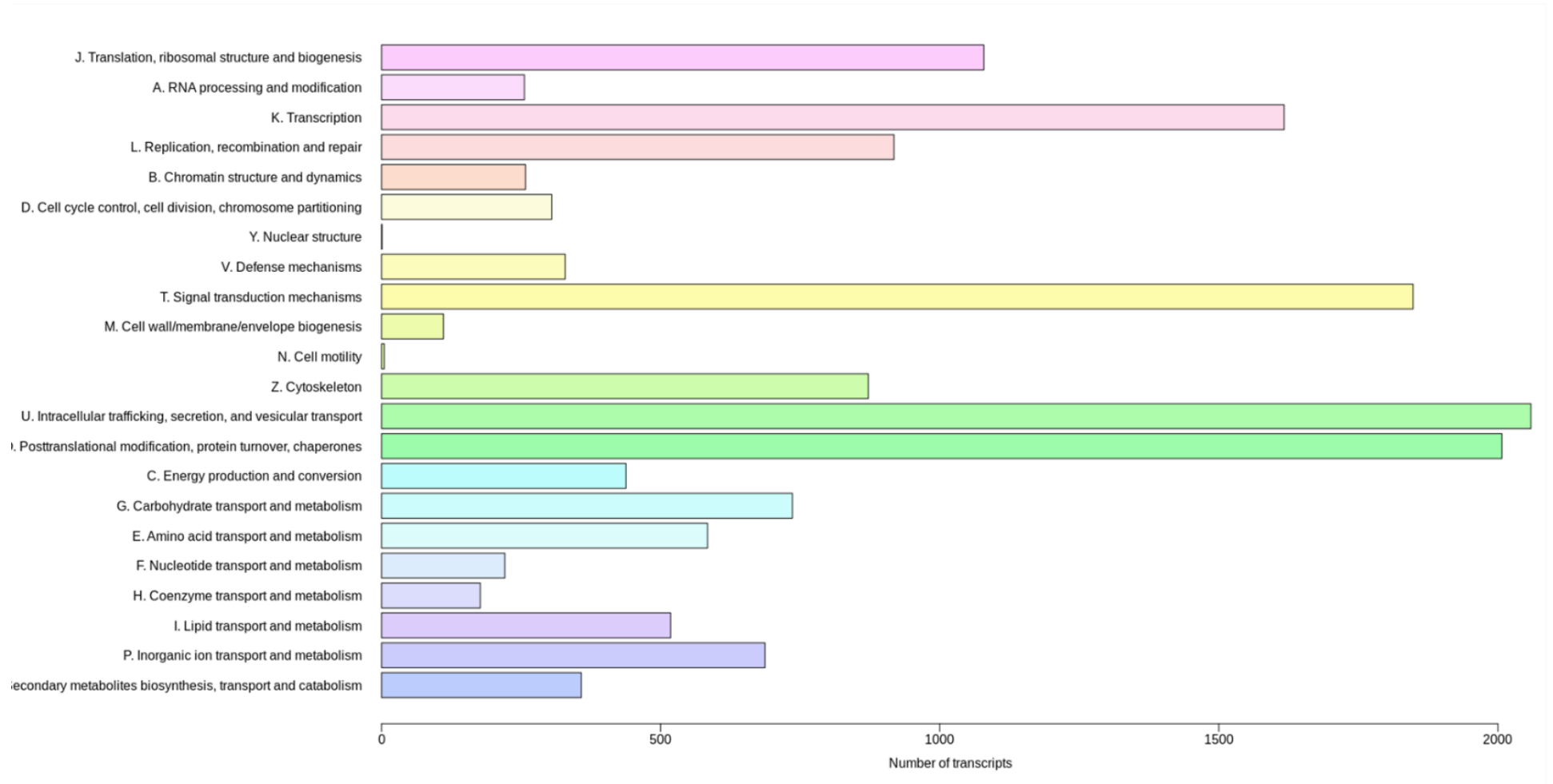


Figura 44. Número de transcritos por categoría ontológica. Cuantificación del número de transcritos cuya homología permite clasificarlo en las diferentes categorías ontológicas.

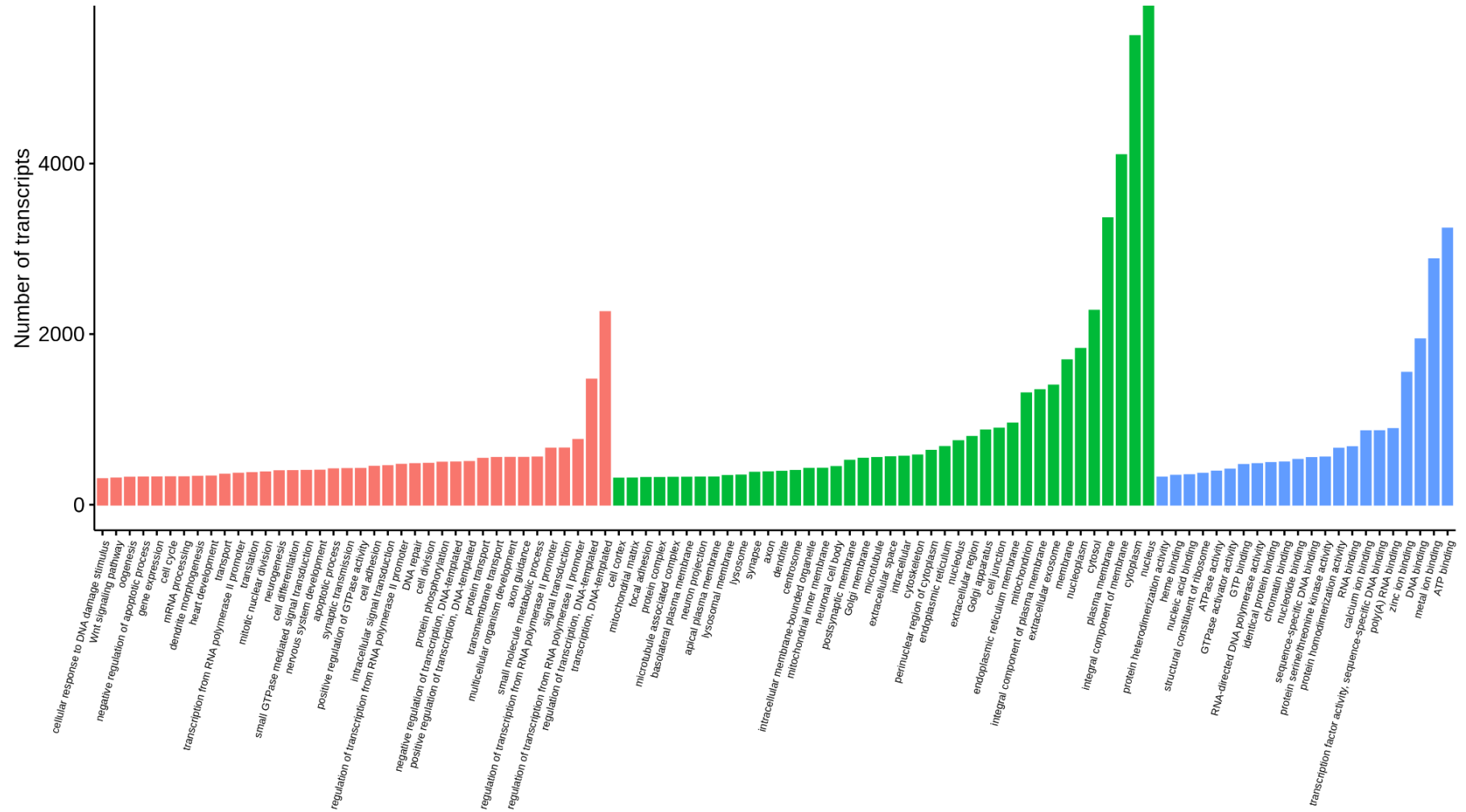


Figura 45. Número de transcritos en función del proceso biológico (barras rojas), componente celular (barras verdes) o función molecular (barras azules) a la que los transcritos pertenecen.

8.D Péptidos identificados por espectrometría de masas en cerebros adultos

Se detallan *in extenso* los péptidos hallados por espectrometría de masas para las muestras de cerebro adulto de *O. fasciatus*.

En esta lista se especifican aquellos péptidos que superaron los criterios establecidos en materiales y métodos para su correcta identificación y poseen curación manual.

En las referencias a continuación se puntualizan aspectos destacables considerados para el análisis.

- Referencias:

KK – KR – RR: sitios dibásicos de corte proteolítico

GK – GR – GKK – GKR – GRK – GRR: Sitios de corte y amidación en el extremo C-terminal

Modificaciones postraduccionales señaladas: oxidación de metionina (**M** y **(M-OH)**), amidación del extremo c-terminal (**X** o **X-NH₂**), Glutamina en piroglutamico (**Q** y **pyroQ**), acetilación del N-terminal (**X** y **acetil-X**).

Secuencia itálica: porción de la proteína predicho por Signal-P como péptido señal.

Números entre paréntesis: Se indica la masa molecular de cada péptido.

Subrayado doble: péptido identificado por espectrometría de masas.

Subrayado doble color verde: péptido identificado por espectrometría de masas homologo a uno conocido como neuropéptido activo/maduro en especies filogenéticamente relacionadas.

C : Cisteínas potencialmente implicadas en un puente disulfuro. (Se especifica en particular el organismo que presenta alta similitud de secuencia y del que se desprende la implicancia de la cisteína en la participación del puente)

Subrayado simple color rojo: péptido con similitud de secuencia a otro péptido conocido como neuropéptido activo/maduro en especies filogenéticamente relacionadas pero que no fue hallado mediante MS/MS en estos experimentos.

AKH - Hormona Adipoquinética - OFAS011374

MYKFLSVYIFLGISVFCCFA.QLNFSPNWGKRMVSQEECKISGEALLYTYKTIQNEIRKLLLECEKMM
S

Péptidos identificados:

QLNFSPNW-NH₂ (1003.48)

(pyro-Q) LNFSPNW-NH₂ (986.47)

AstA - Allatostatina A - OFAS019254

MMWLPLLASLILQICLA.EQGKDAQPKRLYDFGLGKRAAYTYVSEYKRLPIYNFGLGKRYDDSMLSGY
PLGRQNDIDQEDYDEDGLKGSEYKRMKQYSFGLGKRLPTKQYSFGLGKRVVPYDEPELNPDSFLDMADE
EEKRSNPSSNGHR

Péptidos identificados:

EQGKDAQPKRLYDFGL-NH₂ (1862.96)

AAITYVSEYKRLPIYNFGL-NH₂ (2266.18) (similar a ALL1_RHOPR)

LPTKQYSFGL-NH₂ (1151.63)

AstB- Allatostatina B (péptido myoinhibitorio) - OFAS011244

MNLMYLFIVGSI AALAYG.ETPSAENPLSPMQQLEEA VLLDNIENEEKRAWKDLTKTWGKRAWSDLQS
AGWGKRGIEDLHQMALDQREDKRAWNLHSTGWGKRAWSDLPSAGWGKRAWSDLQSAGWGKRAWSDLQ
SAGWGKRAWSDLPSAGWGKRAWSDLQSAGWGKRAWSDLQSAGWGKRAWSDLQSAGWGKRAWSDLQSAGWGK
GWNDLQSAGWGKRAEQGYAAEEKDYDEKTLESPHTGFGKRAADWASFRGKI

Péptidos identificados:

AWKDLTKTW-NH₂ (1146.62)

GIEDLHQ (M-OH) ALDQ (1384.63)

GIEDLHQMALDQRE (1653.78)

GIEDLHQ (M-OH) ALDQRE (1669.77)

GIEDLHQMALDQRED (1768.8)

GIEDLHQ (M-OH) ALDQRED (1784.8)

EDLHQ (M-OH) ALDQRED (1614.69)

AEQGYAAEEKDYDEKTLESPHTGF-NH₂ (2713.21)

AstCC - Allatostatina CC -TRINITY_DN282951_c0_g1_i1 (969-610)

(Modifica OFAS008438)

MCSPRQLLLAAMRWSIRWALLASLCLAVPRLVT.APSQKYTDYKEAGVQYDEYPVVVPKRAAMLLDRI
MVALQKAVDDDNVRSVKSRI PLEDTMDLQRRGQKGGRIYWRCYFNAVTCF

Puentes disulfuro (Veenstra, 2016): C 112->119

Péptidos identificados:

APSQKYTDYKEAGVQYDEYPVVVPK (2873.41)

AstCCC - Allatostatina CCC - TRINITY_DN211533_c0_g1_i1 (954-670)
(Modifica OFAS025029)

MGRGMWLAQSLAVLLLCLLMARSFA.AQPAEKERLINELDKLVDDGDSVETALINYLAFARQMVNKLRS
QVDVTDLQRKRSYWKQCAFNAVSCFGK

-Puentes disulfuro (Veenstra, 2016): C 85->92

Péptido identificado:

LRSQVDVTDLQ (1272.67)

AT - Allatotropin - OFAS004904

MRWLLGVWLVVCLCVAVLG.RPPQARARATRGFKNVALSTARGFGKRAPLDQEVPOPRLKADWLAEM
 SASPELARLFVNRFLDSDGYLSAEELLTRPQHH

Péptidos identificados:

ATRGFKNVALSTARGF-NH₂ (1693.94)

GFKNVALSTARGF-NH₂ (1366.75)

FKNVALSTARGF-NH₂ (1308.73)

APLDQEVPOPRL (1248.65)

Bur - Bursicona alfa - OFAS008719

MFILKISSFHAFQGNISINHASRSITYISGKRNAVEDCISMASSPDKGKNVLEETVLVTK
 NWQHSVVFIVHQRGTRLLNDLNLAEEKHEQAKENRILKISFDMKIKLLWTGSRGDDKRAS
 DECQVTPVIHVLQYPGCVPKPIPSFACTGRCSSYIQVSIQYAFWSVTDWKQRRQCACFR
 RVSSDSDTRPPVSRRLRAQADTFLRLHGKVFLHSGDNESPTGVHVQALHFCGRECCDTA
 GDCRSSGRRPPQSALPQAPVARTTIDLQLQSKELLAIL

CCAP - Péptido cardioactivo de crustáceos -
TRINITY_DN247112_c0_g1_i1 (modifica OFAS007216)

MQLYATLLLCMVCIVLG.DDVLLQKRMFYPAEGVEPYEQKVKRPFCNAFTGCGKRSAKRSGAGDNLAS
LVELNSEPAVEELSRQILSEAKLWEAIQEARMELSNRKQQQEDQVPIHPMPLPLAGHKRQAQYYT

-Puentes disulfuro cisteínas (inferidos por similitud de secuencia con *Tenebrio molitor*): C 46->52

Péptidos identificados:

MFYP AEGVEPYEQKV (1786.84)

(M-OH) FYP AEGVEPYEQKV (1801.82)

YP AEGVEPYEQKV (1507.72)

PAEGVEPYEQKV (1344.66)

CCHamida - OFAS011242

ERDGDGSRKPKPLRRGGCSSFGHSCFGGHGKRSEDYLAQAQRLQRISPPTDIRQWLQAYHSSSSLLE

Péptidos identificados:

SEDYLAQAQRLQ (1420.69)

CNMamida - TRINITY_DN218363_c1_g1_i1 (modificac OFAS002244)

MGGLRLLLLIAVAMALLYGVSS.SPIDDSALQDIAYSLQSQEMSGSDSSLQQLANLVEQARRTGVDLE
 DPDPNPFQPAIMEILRDHELQARAEKRASYLSLCHF KICNMGRKRRNPGNLWQRT

Péptidos identificados:

RNPGNLWQRT (1241.65)

NPGNLWQRT (1085.55)

Crz - Corazonina - OFAS005241

MIWVRQLLFLLAIAILAGSALS.QTFQYSRGWTNGKRAGQTACQLQRLRAVLEGRAPSADPSYLLCEL
 YRSPQEELRLDHLGKTQPREPIPVDIAEK

C-ter (7B2 - SCG) - Péptido C-terminal - OFAS011270

SGENNENNNQVNGQQQQQQNSNRQDSLPA YCDPPNPCPLGYTAEEGCLEDFENTASFSR
 EYQAQQDCMCDSEHMFDCSNNPNNNDISDTDFDRLVEDGHKNLVAKKYHMKNSNPYLD
 GEKLPVAAKKGIDVIY

Péptidos identificados:

NSNPYLDGEKLPVA (1515.76)

NSNPYLDGEKLPV (1444.72)

DH31 - Hormona diurética DH31 - OFAS003583

MVTNVLVTFSLTLGTLILLSAAA.ESVPYIGHRSYLGIDIDNEPDPEVVLEILAKLGQTIMRANDLEN
 SKRGLDLGLSRGFSGSQAAKHLMGLAAANYAGGPGRRRRQA

Péptidos identificados:

AESVPYIGH (971.47)

GLDLGLSRGFSGSQ (1392.7)

GLDLGLSRGFSGSQA (1463.74)

GLDLGLSRGFSGSQAAKHLMGLAAANYAGGP-NH₂ (2985.52)

GLDLGLSRGFSGSQAAKHL (M-OH) GLAAANYAGGP-NH₂ (3001.52)

GLDLGLSRGFSGSQAAKHL (M-OH) GL (2230.15)

DH44 - Factor liberador de corticotropina (CRF - DH44) - OFAS003428

MRLNGAVHRTASFAFSSAEGHCPLLKAVSGCRWTMWWKVLTTWVLAALALVAGSREWTSORLYPPSAQ
 LLOAEDDDWRMKRTGPSLSVSNPIEVLRNRLLEIARRKMOKQGOQNREFLKNIGKRHTLNSDRLLDS
 LDQDREQSIDYLLSNSGDWDSS

Péptidos identificados:

SREWTSORLYPPSAQQLLOAEDDDWR (M-OH) (3193.49)

SREWTSORLYPPSAQQLLOAEDDDWRM (3177.49)

TGPSLSVSNPIEVLRNRLLEIARRKMOKQGOQNREFLKNI-NH₂ (4772.68)

HTLNSDRLLDSLQDREQSIDYLLSNSGDWDSS (3793.73)

EH - Hormona de la eclosión - OFAS009237

MFSTCSDVISYVIVRRAIINEIVILAKTGTDIVEMQLKLLLLLA AVLVVVTLAGPSQVGV
 IRNCAQCKKMFQYFEGQLCAETCLKFKGKLI PD CEDVTSIGPFLNRID

Ele - Elevenin - OFAS015577

MWFTARRGLSAICLCLLILSILVAFA.SSQRGREFVNCKGVAAKRAGNSHPFKAQPOVPYSMRGTPQYS
PSDKQSPYWFYQARLAERAFF

Péptidos identificados:

AGNSHPFKAQPOVPYSMRGTPQYSPDKQSPYWFYQARLA (4585.20)
 AGNSHPFKAQPOVPYS (M-OH) RGTPQYSPDKQSPYWFYQARL (4530.16)
 AGNSHPFKAQPOVPYS (M-OH) RGTPQYSPDKQSPYWFYQARLA (4601.2)
 AGNSHPFKAQPOVPYS (M-OH) RGTPQYSPDKQSPYWFYQARLAE (4714.25)
 AGNSHPFKAQPOVPYS (M-OH) RGTPQYSPDKQSPYWFYQARLAERAFF (5201.5)

ETH - Hormona disparadora de la ecdisis - OFAS025148 -

TRINITY_DN279982_c0_g1_i1

MSTVTGKREAEGFTVLCLTASCQSEPLEDVSEDFVPLLMRRNDFFLKAAKSIPRI GRRND
 FFFTKAVKSVPRI GRRNDFFMKAAKSIPRI GRR EDLTTVSILI

FMRF - FMRF-amida - TRINITY_DN27373_c0_g1_i1 (modifica OFAS025208)

MIQHCCAVNQNEKSRNSFLIDQLSRRSSLEKNFMRFGRSLPQIEERKEALARLSRGADFMRFGRPSK
 AAVIGKRKIQERSGSNDFMRFGRSAYPGDKRLARNKVGFIREFGRFAENPVNKEKVIANTPLKELEM
 EDQRETAGQAGLFRFGRSGFLQSGRFASTGEELEGRKQESGYIRFERSGALEDGGHGIPFTSEMKRGR
 KDTGFLRFGRSSHLDPVEKRLKQGRSAFLRFAPSGFRLPGNLTGKEYEDFFSFGRKADSIRPDFTSHPY
 SFQLPVLGPPGLTAGLPTCCGQSVPPSDCRAPFVTGRMVVSLANQGRICALWRLVLTKAILASWRFP
 LPSRSLSVYGHSLGIYSTSGGR

Péptidos identificados:

SSLEKNFMRF-NH₂ (1256.63)
 KIQERSGSNDFMRF-NH₂ (1712.84)
 FAENPVNKEKVIANTPLKELEMGEDQRETAGQAGLFRF-NH₂ (4275.17)
 LARNKVGFIREF-NH₂ (1405.83)
 GKDTGFLRF-NH₂ (1039.58)

GPA2 - Glicoproteína alfa 2 - TRINITY_DN208695_c0_g1_i1 (modifica OFAS004661)

MLRGTAHSVPSRGELVRTRYFLLFLLHFDIFILQFFPAQSPALRIRKSQELESHSRDNQDFPQRYRGRD
 PRMANRIVCCVVLICLSLVAFSKQPD SWGKPGCHKVGH SRKVRI PD C VEF E V T T N A C R G Y C E S W A V P S
 SNEILLSNPRQLITSVGQCCNIMETEDVQVRVLCLEGMKDLTFKSARSCSCYHCKKD

**GPB5 - Glicoproteína beta 5 - TRINITY_DN305720_c1_g2_i1
(Scaffold233:617.111-616395 + 614103-613905)**

MSTLLSSFYRLLAAMAVAVLVN SSA.EPGSTLECHKRLYQYTARQTDSEGRVCWDNVTVLSCWGRCD
 SNEISDWRFPYKRSHHPVCLHGIREPRTVVL RHCEEGAEPGTELFTALHAVSCKCHVCKSSEASCQGL
 RYRPNH

ILP1 - Péptido relacionado con la insulina 1 - OFAS018844

MIKEMLRVFLLSFGLAVLA.EWDVIPSAREKRSHILCGENLVQTLNIICKGKFKAPFSDTMELQPL
SSMKKSYDNSDDSTSNNNAIDSKEVTLNFPFKSRNKALSILEHRRIKRQVYDRCKEKGCM LSDLQSYCG

ILP2 - Péptido relacionado con la insulina 2 - OFAS018845

MFICFRTIMNYWNLGIILLFLMELVYG.RQFCGSDLTHAMQLVCNCRGYNVAFRSDHSPSEKRFKRG
IVDECKRGCSWSTLETYCSPTPSNIQIRKRSDFASKWQPESSEVQTKIEKFKYDINEMKVEDSDKSK
VKISFDNNRTPNRFNALRDEKNRGGQAIRGHSIVNKVKRNDIAEELPLHPKRSSLSFGINWQSGKL
EPGKIGDAVDGIIVPRIPIRHRHRKADTLRKNTKVSVFIYIHSLSLTKYLASIFKIQSLLTNWyGNCCLLP

Péptidos identificados:

[pyroQ](#)-GGQAIRGHSIVNKVKRNDI (2231.22)

ITG - OFAS025288

MLPAVLLTAVLQPALG.WGGLFNRFSPPEMLSNLGYGGHGTSLRVQPFLQNGQPMETLHDLQSEIEG
PCYGRCTANEHCCSGYVCVDVDGGGSCMFPYGLGQELCRRHSDCETGLVCSDTGEGSKTCMPFFSA
PKQYSEECTMSSECDIHRGLCCQFQRRHRQVARKMCSYFTDPMVCIGPVASDHAKVEIERTAGEKRIT
[GKTPVFL](#)

Péptidos identificados:

WGGLFNRFSPPE (M-OH)LSNLGYGGHGTSL (2712.28)

[ITGKTPVFL](#) (974.58)

ITP - Péptido transportador de iones - OFAS003278

MSSLLMHQERRALASLVVASTLLSCLVAVPTSR.SIMGHPLNKRFFDLQCKGVYDKSIFARLDRIE
DCYNLRFREPQLHSLCRSNCFGSKYFKGCLEALLLGEEEKKYL RMVEFLG

-Puentes disulfuro cisteínas (putativos, inferidos por homología con *Schistocerca gregaria*): 50->86, 76->82, 69->95

VPTSRSIMGHPLN (1407.73)

Kin - Kinina - OFAS017943

MGVLWLLSVAVICQTSHISTA.SSPKNFETSTNSTTPAQRINITVTLHEKDNCTEEIKLREKRGTILQ
DNNLEKRGGQKLFSSWAGKRADEIVEDKRAKFSWGGKRFDEYLNNIYEDNNDLKRKAFSSWGGK
DFMEDQSDKRAKFSWAGKRDIGITEEDKRSKFSWAGKRAPDGDKRAKFSWAGKRGKFSWAGKRS
DLNIHSGEEQGIYPWDEGPENEQTGEYLSEEKRSKFSWAGKRSYVSSWNKVNPE SRLKYYSKM
EEESGSENIMKLWQDLMSRKRARFSSWAGKRSDESISDMESLERLLEPQDEDKRAKFSWAGKRAIE
GMEKRKFSSWGGKRDGAVVKRSWSSYKRKPVFSSWGGKRSNEHSIQARSVGRSLVFKGRPWGSLMRP
VRKQPEFYAWGKRKRSV

Péptidos identificados:

GTILQDNNLE (1115.55)

KPVFSSWG-NH₂ (905.48)

MS - Miosupresina - OFAS013557

MKVSTVCLCMVVLMAACSNA.APGPDCSPAVISELPARIRNMCAALYQFSSALQQYIDENPSYQPARE
VVAREASPVYENGVKRQDLDHVFLRFGRRR

Péptidos identificados:

acetil-SNAAPGPD CSPAVISELP (1765.82)

IDENPSYQ PAREVVA (1686.82)

IDENPSYQ PAREVVAREASPVYENGV (2888.39)

VVAREASPVYENGV (1488.76)

(pyro-Q) DLDHVFLR**F**-NH₂ (1270.65)

QDLDHVFLR**F**-NH₂ (1287.67)

(pyro-Q) DLDHVFLRF (1271.63)

(pyro-Q) DLDHVFLR (1124.56)

DLDHVFLR**F**-NH₂ (1159.61)

LDHVFLR**F**-NH₂ (1044.58)

DHVFLR**F**-NH₂ (931.50)

Nlg - Neuroligin - OFAS015910

MVCRLSFTSMESLTSGIPETPMTAVS

NP1 - Neuroparsina 1 - OFAS014019

MKYSFTGERCEVDLKSCYGTVRLCGQVLCAGPGMRCGGYMNIR**GK**CAEGLTCNCCGYC
TGCFKDSLDCYKSDLVMCYAD**RKK**

NP2 - Neuroparsina 2 - OFAS014020

MIRIPGGERCEVDLKSCYGTVRLCGQVLCAGPGMRCGGYMNIR**GK**CAEGLTCNCCGY
CTGCFKDSLDCYKSDLVMCYAD**RKK**

NP3 - Neuroparsina 3 - OFAS014021

MVSKGRDSNQLEVQMDRPRHTRGERCEVDLKSCYGTVRLCGQVLCAGPGMRCGGYMNIR**GK**CAEGLTCNCCGYCTGCFKDSLDCYKSDLVMCYAD**RKK**

NP4 - Neuroparsina 4 - OFAS014022

MVCLGSTSAFLASARAQVPSSRRRGGEAGWNLRLWVAFQRELSLVVREDRIRDLCYRGE
RCEVDLKSCYGTVRLCGQVLCAGPGMRCGGYMNIRGNRYSE

Péptido identificado (común a las 4 isoformas de Neuroparsina):

RLCGQVLCAGPG (**M-OH**) **RC**-NH₂ (1705.83)

NPAA - Neuroparsina - OFAS007227

MTLLLVNIFFLFGIALIET.SEDFLCYPCRGDDCDQOPDGCLYGTVRDPC**GK**LVCAAGPGDRCG**GR**DN
HLGKCGEGMTCKC**GK**CRGCSIVRLMSGVIECEYTNPICT

Péptido identificado:

QPDGCLYGTVRDPC (1579.67)

NPF - Neuropéptido F largo - OFAS025089

MKSWLACWCISLVLMVCOA.EPAPADAMSRPARPKSFASPDLLRLYLDQLGQYYAVAGRPRF**FGKRTSS**
HPQRFHLPLEASSYRYNGPDPSPDLYDLLYQPLAAE

Péptidos identificados:

EPAPADAMSRPA (1211.56)

EPAPADA (**M-OH**) SRPA (1227.55)

APADAMSRPA (985.47)

PKSFASPDLLRLYLDQLGQYYAVAGRPR**F**-NH₂ (3341.73)

LRLYLDQLGQYYAVAGRPR**F**-NH₂ (2397.30)

QYYAVAGRPR**F**-NH₂ (1382.72)

TSSHPQRFHLPLEASSYRYNGPDPSPDLYDLLYQPLAAE (4347.09)

TSSHPQRFHLPLEASSYRYNGPDPSPDLYDLLYQPLA (4147.01)

TSSHPQRFHLPLEASSYRY (1955.95)

NPLP - Precursor tipo neuropéptido - OFAS010770

MAYFNKVTGEVD**KR**SISTLARTGQL**P****GKR**YIAALAKNGEYPRPSWN**KK**YSSLEELAEEDPYA**KR**FLGS
LARSGDLRTREE**KR**TEVDGLVRDMEAAERI**RR**LRLDALKEELIKDGVDEDPEDDKTALPD**KR**SVEALA
RIGLLKPITTTTHDFTEDINDYE**KK**ISSDEFEED**KR**GGISSLARNGWNN**KR**TVDEELEELMNEVYGIGE
KRNVASLARGFNLP**NGKR**SEEEDE**KR**NIASVMRDM**SGKR**DSYTNNAI PFVNDD**KR**NVGALARNWDRP
YIY**KY****GKR**EVSEEEKEEV**KKR**YVATLLRDGPLVGPDANTSNQDESQAKETES**M**TEGREDQSLSEEKG
ANQR**SK**KEASEEDTHS**RKKR**ETFISPVKDELAAPSSDIEFGGLIGAGNKWADDT**SPLNRR**YFAVRGGG
KMTG**GR**LPKVGA**PVRGGG**KMPGGRLPKVGT**RR**SRSRHENSG**RRRH**

Péptidos identificados:

SISTLARTGQL**P**-NH₂ (1241.71)

ISTLARTGQL**P**-NH₂ (1154.67)

YIAALAKNGEYPR (1464.77)

YIAALAKNGEYPRPSWN (1948.97)

YIAALAKNGEYPRPSWN**KK**YSSLEELAEEDPYA (3801.85)

AALAKNGEYPRPSWN (1672.83)

TEVDGLVRDMEAAERI (1802.88)

LRLDALKEELIKDGVDEDPEDDKTALPD (3151.58)

LDALKEELIKDGVDEDPEDDKTALPD (2882.39)

LIKDGVDEDPEDDKTALPD (2083.97)

IKDGVDEDPEDDKTALPD (1970.90)

PEDDKTALPD (1099.50)

SVEALARIGLLKPITTTTHDFTEDINDYE (3160.59)

SVEALARIGLLKPITTTTHDFTEDINDYE**KK**ISSDEFEED (4468.18)

SVEALARIGLLKPITTTTHDFTEDINDYE**KK**ISSDEFEE (4353.15)

SVEALARIGLLKPITTTTHDFTEDIND (2868.48)

EALARIGLLKPITTTTHDFT (2096.16)

GGISSLARNGWNN (1344.65)

GGISSLARNGWN (1230.61)

GISSLARNGWNN (1287.63)
 NVASLARGFNLPGN-NH₂ (1429.78)
 VASLARGFNLPGN-NH₂ (1313.72)
 NIASVMDMS-NH₂ (1123.55)
 NIASV (M-OH) RDMS-NH₂ (1137.53)
 NIASVMD (M-OH) S-NH₂ (1137.53)
 NVGALARNWDRPYIYKY-NH₂ (2097.09)
 GALARNWDRPYIYKY-NH₂ (1883.97)
 EVSEEEKEEV (1205.53)
 YVATLLRDGPLVGPDANTSNQDESQAKETESMTEGREDQSLSEEKGANQRS (5634.62)
 YVATLLRDGPLVGPDANTSNQDESQAKETES (M-OH) TEGREDQSLSEEKGANQRS (5650.62)
 YVATLLRDGPLVGPDANTSNQDESQAKETESMTEGREDQSLSEEKGANQ (5391.49)
 TLLRDGPLVGPDANTSNQDESQAKETES (M-OH) TEGREDQSLSEEKGANQRS (5317.45)
 LRDGPLVGPDANTSNQDESQAKETESMTEGREDQSLSEEKGANQRS (5087.32)
 LRDGPLVGPDANTSNQDESQAKETES (M-OH) TEGREDQSLSEEKGANQRS (5103.32)
 LRDGPLVGPDANTSNQDESQAKETESMTEGREDQS (3887.74)
 RDGPLVGPDANTSNQDESQAKETES (M-OH) TEGREDQSLSEEKGANQRS (4990.24)
 RDGPLVGPDANTSNQDESQAKETES (M-OH) TE (3118.36)
 RDGPLVGPDANTSNQDESQAKETES (M-OH) (2888.28)
 RDGPLVGPDANTSNQDESQAKETESMTEGREDQSLSEEKGANQRSKKEASEEDTHS (6215.79)
 PVGPDANTSN (1012.45)
 ETFISPVKDELAAPSSDIEFGGLIGAGNKWADDTSPLN (3960.93)
 VRGGKMPGGRLPKVGT (1665.95)
 VRGGK (M-OH) PGGRLPKVGT (1681.95)

NVP - NVP - OFAS010896

MGLLQELSIYLLCFGAVATA. IPVSLIEDLKEAQLQTPLRTNKV KRAQEVLMFGNQONRAINNYGSAR
PEKR DIGDNTLAEDVATLPNILPQVEDPGYDTNQLGSGAQYEKAYPYKGRVPTYGMLLR NAALKESLH
SDPVLGYGDLPLYNMMDA RRKR DAHKSAPALQANKNR

NVP - NVP - TRINITY_DN265883_c0_g1_i1 (isoforma más larga a partir de N171 "I")

MGLLQELSIYLLCFGAVATA IPVSLIEDLKEAQLQTPLRTNKV KRAQEVLMFGNQONRAINNY
GSARPE KRDIGDNTLAEDVATLPNILPQVEDPGYDTNQLGSGAQYEKAYPYKGRVPTYGMLLR NAALK
ESLHSDPVLGYGDLPLYNMMDA RRKR DAHKSAPALQANKN I SFRS KREYDLTPEELVSLMRLLDNQRR
QNSRLHEQSRGWPMYGSDESMDDFGSDNPDDNEIDTESDSYSRNGVWMDG PVAFGQPQ R DGDLWPR RKY
KRFMVS RR SDMEGLGQVNSPYDFPEGVPLQ RR FIL

Péptidos identificados:

IPVSLIEDLKEAQLQTPLRTN (2377.32)
IPVSLIEDLKEAQLQTPLRTNKV (2604.48)
VSLIEDLKEAQLQTPLRTN (2167.18)
VSLIEDLKEAQLQTPLRTNKV (2394.35)
LIEDLKEAQLQTPLRTN (1981.08)
LIEDLKEAQLQTPLRTNKV (2208.25)
LKEAQLQTPLRTN (1510.85)
LKEAQLQTPLRTNKV (1738.01)
AQEVLM (M-OH) FGNQONRAINNYGSARPE (2722.29)

AQEVLMEFGNQONRAINNYGSARPE (2706.29)
 MFGNQONRAINNYGSARPE (2166.00)
 FGNQONRAINN (1274.61)
 FGNQONRAINNYGSARPE (2034.96)
 AINNYGSARPE (1190.57)
 INNYGSARPE (1119.53)
 DIGDNTLAEDVATLPNLPQVEDPGYDTNQLGSGAQYKAYPYKGRVPNY-NH₂ (5467.65)
 PQVEDPGYDTNQLGSGAQYKAYPYKGRVPNYG (M-OH) L (3920.83)
 NAALKESLHSDPVLGYGDLPLYN (2485.25)
 NAALKESLHSDPVLGYGDLPLYN (M-OH) (2632.28)
 NAALKESLHSDPVLGYGDLPLYN (M-OH) (M-OH) D (2894.35)
 NAALKESLHSDPVLGYGDLPLYN (M-OH) (M-OH) DA (2965.38)
 ESLHSDPVLGYGDLPLYN (1987.95)
 DAHKSAPALQANKN (1463.75)
 DAHKSAPALQANKNS (1550.78)
 DAHKSAPALQANKNSF (1697.85)
 DAHKSAPALQANKNSFRS (1940.98)
 HKSAPALQANKNSF (1511.78)
 HKSAPALQANKNSFRS (1754.92)
 APALQANKNSFRS (1402.73)
 LQANKNSFRS (1163.6)
 QANKNSFRS (1050.52)
 DGPVAFGQPQ (1014.48)
 DGDLWPR (857.4)
 DGDLWPRKY (1148.56)
 RSDMEGLGQVNSP (1388.64)
 RSDMEGLGQVNSPYDFPEGVPLQ (2534.17)
 RSD (M-OH) EGLGQVNSPYDFPEGVP (2309.03)
 RSD (M-OH) EGLGQVNSPYDFPEGVPL (2422.11)
 RSD (M-OH) EGLGQVNSPYDFPEGVPLQ (2550.17)
 RSD (M-OH) EGLGQVNSPYDFPEGVPLQ (2550.17)
 RSD (M-OH) EGLGQVNSPYDFPEGVPLQR (2706.27)
 RSDMEGLGQVNSPYDFPEGVPLQRRF (2993.45)
 RSD (M-OH) EGLGQVNSPYDFPEGVPLQRR (2862.37)
 RSD (M-OH) EGLGQVNSPYDFPEGVPLQRRF (3009.44)
 RSD (M-OH) EGLGQVNSPYDFPEGVPLQRRFI (3122.52)
 RSDMEGLGQVNSPYDFPEGVPLQRRFIL (3219.61)
 RSD (M-OH) EGLGQVNSPYDFPEGVPLQRRFIL (3235.61)
 (M-OH) EGLGQVNSPYDFPEGVPLQ (2192.01)
 QVNSPYDFPEGVPLQ (1688.80)

Orc - Orcokinina - OFAS015031

MEKKKRIWNVSEKEQRFQTLKELGAWSYSMKDTGGTLDTIGGGHLLRNVEDEPPVSSEGLEEDELLO
 NTEESQIDNRNLDTIGGGHLIRNLDTIGGGHLIRNLDTIGGGHLIRNLDTIGGGHLIRNLDTIGEISS
 RLYLRGIKMNALPLISREREELVAVCPYQALPSTFHGDKSTKNAQFSADIVKAIKLMKLMGKR
 NGMDEIDRSGFNSFIKKKNFDEIDRSGFDGFVKNFDEIDRSGFNGFVKKRATHN

Péptidos identificados:

*NGMDEIDRSGF (1239.52)
 NG (M-OH) DEIDRSGF (1255.51)
 NGMDEIDRSGFN (1353.56)
 NGMDEIDRSGFNS (1440.59)
 NG (M-OH) DEIDRSGFNS (1456.59)
 NGMDEIDRSGFNSF (1587.66)
 NG (M-OH) DEIDRSGFNSF (1603.66)
 NGMDEIDRSGFNSFI (1700.75)
 NG (M-OH) DEIDRSGFNSFI (1716.74)
 (M-OH) DEIDRSGFNSFI (1545.68)
 DEIDRSGFNSFI (1398.64)
 EIDRSGFNSFI (1283.61)
 *KNFDEIDRSGFDGFV (1744.81)
 NFDEIDRSGFDGF (1517.64)
 NFDEIDRSGFDGFV (1616.71)
 FDEIDRSGFDGF (1403.60)
 DEIDRSGFDGFV (1355.60)
 *NFDEIDRSGFN (1312.57)
 NFDEIDRSGFN-NH₂ (1311.58)
 NFDEIDRSGFN (1516.66)
 NFDEIDRSGFN (1615.73)
 DEIDRSGFN (1255.55)
 DEIDRSGFN (1354.62)
 EIDRSGFN (1170.53)

PBAN - Neuropéptido activador de la biosíntesis de feromonas - OFAS010321

MINEDRCLWTGIKGSCWNSESHTMASYAVLALSLLLLAASATGLMEEGRRRWEDEDERPEETLAEVLRE
 TPWVIIPLPSGKRSAFTPRLGREEEAILAESSRSPPFAPRLGRVMKFSPRLGRHIP

Péptidos identificados:

SLLLLAASATGLME (1275.67)
 WEDEDERPEETLAEV (1845.79)
 SAFTPRL-NH₂ (789.45)
 EEEAILAESSRSPPFAPRL-NH₂ (2097.09)
 EEEAILAESSRSPPFAP (1828.88)
 EAILAESSRSPPFAPRL-NH₂ (1839.00)
 ILAESSRSPPFAPRL-NH₂ (1638.92)
 AESSRSPPFAPRL-NH₂ (1412.75)
 ESSRSPPFAPRL-NH₂ (1341.72)
 SRSPPFAPRL-NH₂ (1212.67)
 SRPPFAPRL-NH₂ (1125.64)
 SPPFAPRL-NH₂ (882.51)
 MKFSPRL-NH₂ (876.50)

Pbur - Bursicon-beta - OFAS008715

MTDQQDQITNGQYEEYDELGRLQRTCSGDIAVNKCEGACNSQVQPSVITPTGFLKECYCCRESFLRER
AVRLEHCYDPDGVRLERDRGTMEVRLREPADCKCYKCSYTR

PDF - Factor dispersante de pigmentos - OFAS015779

MKWGEVSGPHIQQLEVRRTPNAPQVSMGRHTLTASGNPQFPVEITKHSFRSVLEALNQKEGQAWVPEL
LRSEEHGR **KRN**SEIINSLLGIPKVIND**AGRK**

Péptidos identificados:

NSEIINSLLGIPKVIND**A**-NH₂ (1908.07)

SLLGIPKVIND**A**-NH₂ (1237.74)

Proc - Proctolin - TRINITY_DN295884_c0_g1_i1 - Scaffold3202:32.333-33.162

MARHTIVGVLALMLIAAVLTEG. RYLPT**R**SQDDKLLRLRQLLKDLLENDVEPMEHPIGGYDPRLY**KRE**
APAYDRPMHYMH

Péptido identificado:

EAPAYDRPMHYM (1348.59)

PTTH- Hormona protoracicotrópica - OFAS025263

MLFGNIAILVAMISSINKSGLG. CLLGQAECTGHNHLLDLGRGYYPRI ISAVECGGHNCVPYSCTSRI
YKVKVLKKGIEDVIPDGSLPESLQDSWRFVYVPSVACVCSLYKQHNK

RYa - Ryamida - OFAS018696

MTGMKY**P**RKSENFF**M**GSRY**GKR**SGNADDGEASFKSRFDIDLCKINKYIYPCANSR

Péptidos identificados:

KSENFFMGSRY-NH₂ (1420.66)

KSENFF (M-OH) GSR**Y**-NH₂ (1436.65)

SIF - SIF-amida - OFAS005914

MSRTLCLTFLVLAFALLVLDVASA. SYKKPPFNGSI**F****GKR**SGPQVDYETAG**K**ALSTMCEIASEACVAW
FPIQDNSAN

Péptidos identificados:

SYKKPPFNGSI**F**-NH₂ (1455.79)

YKKPPFNGSI**F**-NH₂ (1384.75)

KKPPFNGSIF (1133.62)

SNF (sNPF) - Neuropéptido F corto - OFAS003796

MKSKEENE**KK**SKSIRDLYELLMTRDAQQL**MD**GHKLV**RK**SPNNRSPQLRL**R****GRR**SDPSFLQIQKNVYK
YLSNVYYDENMKL

Péptidos identificados:

MTRDAQQLMDGHKLV (1741.86)
 RDAQQLMDGHKLV (1509.77)
 RDAQQL (M-OH) DGHKL (1525.77)
 SPNNRSPQLRLRF-NH₂ (1582.88)

Sk - Sulfaquinina - OFAS003841

MPTGLTSGIILISLGIQRALLASLPETSINMRKADPSAQSSDFSTRTTMIGLLMTVGCWLLLSAE
 ARNYREAEDPNEV **GKR** QFNEYGHMR **FGKR** GGVTEEFDDYGY **MR** **GR** GQP

Péptidos identificados:

YREAEDPNEV-NH₂ (1348.59)
 QFNEYGHMR-NH₂ (1326.59)
 Pyro-QFNEYGHMR-NH₂ (1309.56)
 GGVTEEFDDYGY (M-OH) RF-NH₂ (1927.84)

Tk - Taquicinina - OFAS010422

LAAAVVAGLAYCLA.QEVR **RR**APSMGFMGM **RGKK**SSPAIFPDDEAGMR **GKK**APSMGFMGM **GKK**EDMA
 DFWDED **KR**APASGFFGM **RGKK**APNSGFFGM **RGKK**APSSGFFGT **RGKK**GPSGFIGMR **GKK**DDFEDIDSL
 LDYLRAADPRIGGEDDFMDSRA **KR**YAPEESLQENSSEQEF

Tk - Taquicinina - TRINITY_DN266630_c0_g2_i1

MRLSRGVVLA AAVVAGLAYCLA.QEVR **RR**APSMGFMGM **RGKK**SSPAIFPDDEAVLEEY **KR**APVMGFQ
MR**GKK**APSMGFMGM **RGKK**EDMA DFWDED **KR**APASGFFGM **RGKK**APNSGFFGM **RGKK**APSSGFFGT **RGK**
KGPSGFIGMR **GKK**DDFEDIDSL LDYLRAADPRIGGEDDFMDSRA **KR**YAPEESLQENSSEQEF

Péptidos identificados (por aparición secuencial en el precursor):

Pyro-QEVRRAPSMGFMGM-NH₂ (1733.83)

>sp|P86581|TRP6_ONCFA Tachykinin-related peptide 6

APSMGFMGM-NH₂ (1082.48)
 APS (M-OH) GFMGM-NH₂ (1098.48)
 APSMGFMG (M-OH) R-NH₂ (1098.48)
 APS (M-OH) GF (M-OH) GMR-NH₂ (1114.47)
 APS (M-OH) GFMG (M-OH) R-NH₂ (1114.47)

>sp|P86586|TRP5_ONCFA Tachykinin-related peptide 5

APVMGFQGM-NH₂ (1091.54)
 APV (M-OH) GFQGM-NH₂ (1107.53)
 APVMGFQG (M-OH) R-NH₂ (1107.53)
 APV (M-OH) GFQG (M-OH) R-NH₂ (1123.53)

>sp|P86581|TRP6_ONCFA Tachykinin-related peptide 6 (repetido)
 APSMGFMGM

>sp|P86583|TRP2_ONCFA Tachykinin-related peptide 2

APASGFFGMR-NH₂ (1038.51)

APASGFFG (M-OH) R-NH₂ (1054.50)

PASGFFG (M-OH) R-NH₂ (983.46)

>sp|P86585|TRP4_ONCFA Tachykinin-related peptide 4

NPASGFFGMR (según Neupert et al., 2009)

APNSGFFGMR (Secuencia modificada a partir de datos transcriptómicos)

APNSGFFGMRG (1139.52)

APNSGFFGMR-NH₂ (1081.51)

APNSGFFG (M-OH) R-NH₂ (1097.51)

>sp|P86584|TRP3_ONCFA Tachykinin-related peptide 3

APSSGFFGTR-NH₂ (1024.51)

PSSGFFGTR-NH₂ (953.47)

>sp|P86582|TRP1_ONCFA Tachykinin-related peptide 1 (**Isoleucina**
determinada por transcriptómica)

GPSGFIGMR-NH₂ (919.47)

PRIGGEDDFMDSRA (1564.69)

PRIGGEDDF (M-OH) DSRA (1580.69)

Taquicinina 3 unida a Taquicinina 1 sin procesado proteolítico:

APSSGFFGTRGKKGPSGFIGMR-NH₂ (2240.16)