

**ANÁLISIS TRANSCRIPTÓMICO DE NEUROPEPTIDOS
EN ESTADIOS EMBRIONARIOS DE INSECTOS**

Trabajo Final de Grado
del alumno



**Escuela de Ciencias Agrarias, Naturales y Ambientales.
Universidad Nacional del Noroeste de la Provincia de Buenos Aires.**

Pergamino, 5 de abril de 2024

**ANÁLISIS TRANSCRIPTÓMICO DE NEUROPEPTIDOS
EN ESTADIOS EMBRIONARIOS DE INSECTOS**

Trabajo Final de Grado

del alumno

NICOLÁS LAVATTI

Aprobada por el Tribunal Evaluador

(Nombre y Apellido)
Evaluador/a

(Nombre y Apellido)
Evaluador/a

(Nombre y Apellido)
Evaluador/a

Rolando Víctor Rivera-Pomar
Co-Director

Agustín Ariel Baricalla
Director

**Escuela de Ciencias Agrarias, Naturales y Ambientales.
Universidad Nacional del Noroeste de la Provincia de Buenos Aires**

Pergamino, 5 de abril de 2024

Tabla de contenido

AGRADECIMIENTOS	5
RESUMEN.....	7
1 INTRODUCCIÓN	8
1.1 INSECTOS Y DESARROLLO.....	8
1.2 NEUROPEPTIDOS Y PROTEÍNAS HORMONALES	10
1.3 SECUENCIACIÓN DE ARN Y EXPRESIÓN GÉNICA DIFERENCIAL.....	12
1.4 ACLARACIONES SOBRE LA CORRECTA CARACTERIZACIÓN DE NEUROPEPTIDOS	13
1.5 ANTECEDENTES DE NEUROPEPTIDOS EN EL DESARROLLO.....	13
2 HIPÓTESIS Y OBJETIVOS	15
2.1 HIPÓTESIS	15
2.2 OBJETIVO GENERAL	15
2.3 OBJETIVOS ESPECÍFICOS	15
2.4 PALABRAS CLAVE.....	15
3 METODOLOGÍA.....	16
3.1 BÚSQUEDA Y RECOLECCIÓN DE DATOS.....	16
3.2 PREDICCIÓN E IDENTIFICACIÓN DE NEUROPEPTIDOS EN CADA ESPECIE	16
3.3 ORDENAMIENTO DE SERIES TRANSCRIPTÓMICAS TEMPORALES DE DESARROLLO ENTRE ESPECIES.....	17
3.4 PROCESAMIENTO DE DATOS	19
3.4.1 <i>Determinación de calidad y filtrado de datos</i>	19
3.4.2 <i>Alineamiento de reads contra los genomas anotados</i>	19
3.4.3 <i>Cuantificación de reads, normalización y expresión génica diferencial</i>	20
4 RESULTADOS.....	21
4.1 COMPARATIVA DEL DESARROLLO EMBRIONARIO ENTRE ESPECIES	21
4.2 PROCESAMIENTO Y EVALUACIÓN DE CALIDAD DE DATOS.....	22
4.3 ALINEAMIENTOS CONTRA LOS RESPECTIVOS GENOMAS ANOTADOS	23
4.4 CUANTIFICACIÓN, NORMALIZACIÓN Y EXPRESIÓN GÉNICA DIFERENCIAL.....	24
4.5 EXPRESIÓN DE NEUROPEPTIDOS A LO LARGO DEL DESARROLLO EMBRIONARIO.....	26
4.5.1 <i>Los neuropéptidos se identificaron a nivel embrionario</i>	26
4.5.2 <i>Los neuropéptidos se expresan en distintas etapas del desarrollo</i>	26
5 DISCUSIÓN	35
6 CONCLUSIÓN	37
7 PERSPECTIVAS Y DIRECCIONES FUTURAS	37
8 BIBLIOGRAFÍA.....	38
9 ANEXO	45
10 PRODUCCIÓN CIENTÍFICA	49

Agradecimientos

Con esta tesis finaliza un ciclo de gran crecimiento, tanto académico como personal. Este viaje no fue realizado individualmente, sino que existen muchas personas que merecen ser mencionadas, y a las cuales debo mi más sincero agradecimiento:

A la UNNOBA, por haberme otorgado formación, hogar y amigos a lo largo de este periodo. Lo que comenzó siendo una apuesta a la universidad local de mi ciudad culminó en una etapa de descubrimiento que disfruté muchísimo. Gracias a los docentes que nos dan esta formación de calidad, en especial a esos que nos transmiten pasión por lo que hacen y contagian el interés por sus materias. En especial aquellos con los cuales uno tiene un acercamiento más personal y nos brindan su apoyo y consejos tanto para la carrera como para el día de mañana.

A mis directores, Agustín Baricalla y Rolando Rivera-Pomar, por permitirme realizar este trabajo con ellos.

A Agus, por confiar en mí desde el primer día. Por formarme y darme un abanico de herramientas enorme. Por impulsarme a superar los desafíos que se presentaban, sabiendo que era capaz de resolverlos. Por estar siempre presente para aclarar esas dudas rápidas que uno tiene. Por las charlas y consejos en lo académico, laboral y personal, gracias por todo. Tanto Gengar como Haunter son grandes Pokémon.

A Rolando, por su ayuda, predisposición y consejo ante las dudas que surgieron en la marcha, y por tener siempre el *paper* indicado bajo la manga.

A los mejores amigos y amigas que la vida puede otorgar. A Marti, Cande y Cache, quienes fueron las caras con las que me crucé esos primeros días de química general y análisis matemático. Si bien hacer carreras distintas dificultaron los mates durante la cursada, nunca faltaron las juntadas y el apoyo en estos años. A Mati, Anita, Ayito y Caro, que me acompañaron a lo largo de toda esta travesía llena de laboratorios, trabajos prácticos, resúmenes grupales y juntadas de estudio virtuales cuando ocurrió la pandemia. A Agus y Juli, con quienes pude compartir algunas materias, tecitos y otros lindos momentos.

A los chicos y chicas de las tutorías de Introducción a la Biología. Lograron que mi interés por la docencia se incrementara exponencialmente. Espero que, junto con el resto de los tutores, hayamos logrado transmitirles por qué vale la pena recorrer este camino. Destaco a Mati y Anita, con quienes compartí mi rol de tutor. Enseñar y finalizar la cursada juntos fue un momento muy emotivo, y uno de los más lindos de toda la carrera.

A toda mi familia, en especial a mis padres y mi hermano, por el apoyo que recibí y por el grandísimo esfuerzo que realizaron para que pueda estar donde me encuentro ahora.

A Ana, por su compañía, optimismo y amor. Por escuchar atentamente la infinidad de explicaciones y quejas que salían de mi boca, y por contrarrestarlas con otra infinidad de preguntas que nos dejaron charlas hermosas sobre esta carrera que tanto nos apasiona. Gracias por todo.
I love you 3000.

"[...] I have only done the easy part, that is, using the excellent computer programs written by others to look at the very large amount of data collected and analyzed by many colleagues. To all of those who made this possible I express my most sincere gratitude."

Jan Veenstra

Resumen

Los neuropéptidos conforman un grupo de moléculas de señalización que actúan como neurotransmisores, neuromoduladores y factores de crecimiento, además de asociarse a la alimentación, cortejo y ciclo circadiano, entre otras funciones.

Investigaciones previas demuestran una alta similitud entre las secuencias aminoacídicas de la hormona protoracicotrópica (PTTH) en *B. mori* con péptidos de la familia de la insulina y factores de crecimiento de tipo insulina. Además, se conoce la participación de PTTH en la vía de señalización de proteínas tipo insulina a través de su unión a un receptor de tirosina quinasa conocido como Torso en *D. melanogaster*, que actúa a nivel de desarrollo embrionario. Por otra parte, Baricalla (2020) ha estudiado la expresión de neuropéptidos en *O. fasciatus* a lo largo de una serie temporal en estadios embrionarios y en adultos. No se detectaron neuropéptidos maduros en embriones, por lo que en su lugar estos podrían existir en su forma inmadura como reservorios para etapas de desarrollo subsiguientes. Estos precursores podrían poseer funciones específicas en el estadio embrionario, diferentes a las descritas en adultos hasta la fecha.

El objetivo del presente trabajo es comprobar la expresión transcriptómica de neuropéptidos en estadios tempranos del desarrollo, en especies representativas de diferentes órdenes de insectos, mediante ensayos de expresión génica diferencial.

Para ello, se recolectaron datos de series temporales transcriptómicas a nivel embrionario de diversas especies de insectos de estudio popular (NCBI-SRA). Se les realizó un chequeo de calidad y remoción de adaptadores, para ser alineados a los respectivos genomas de referencia de cada especie y cuantificados con el fin de realizar los análisis de expresión diferencial. En paralelo a este flujo de trabajo, se desarrollaron dos tareas.

Por un lado, se identificaron los neuropéptidos presentes en cada especie a través de herramientas de búsqueda de ortólogos, a partir de una base de datos curada manualmente con las secuencias aminoacídicas de interés, contra las respectivas anotaciones proteicas de cada especie. Se homogeneizó la nomenclatura de neuropéptidos para todas las especies de trabajo. A partir de los péptidos identificados se obtuvieron los genes de estudio para buscar aquellos diferencialmente expresados.

Por otro lado, se realizó una revisión bibliográfica que permitiera ajustar los datos recolectados a eventos clave del desarrollo embrionario, haciendo posible una comparación interespecífica de los resultados obtenidos.

Los resultados hallados permitieron corroborar la existencia de expresión transcriptómica de neuropéptidos en todas las especies de estudio. Dicha expresión es variable a lo largo del desarrollo embrionario y en general no presenta patrones apreciables y comparables entre las distintas especies. Existen transcritos de neuropéptidos que presentan picos de expresión en un mismo momento del desarrollo embrionario (AKH, BURS, FMRF y ORCK en himenópteros; PBAN en hemípteros). Adicionalmente se identificaron varios neuropéptidos expresados previo a la formación del ovocito en *A. gambiae*, *R. prolixus* y *B. germanica*.

Esta tesis es el primer trabajo en estudiar un conjunto de neuropéptidos expresados durante el desarrollo embrionario en una selección de especies representativas de insectos. Futuras investigaciones permitirán complementar los resultados obtenidos y esclarecer el motivo de la presencia de los neuropéptidos y su participación en el desarrollo embrionario.

1 Introducción

La bioinformática puede ser definida como “la aplicación de herramientas computacionales para organizar, analizar, entender, visualizar y almacenar información asociada con macromoléculas biológicas” (Diniz & Canduri, 2017). Esta disciplina reúne conceptos biológicos, estadísticos y computacionales con la finalidad de otorgar sentido biológico a los datos generados. La bioinformática posee una diversidad de subtemas, entre los cuales se encuentran la genómica, transcriptómica y proteómica—agrupadas como ciencias ómicas—, siendo la transcriptómica de particular interés en el presente manuscrito (figura 1).

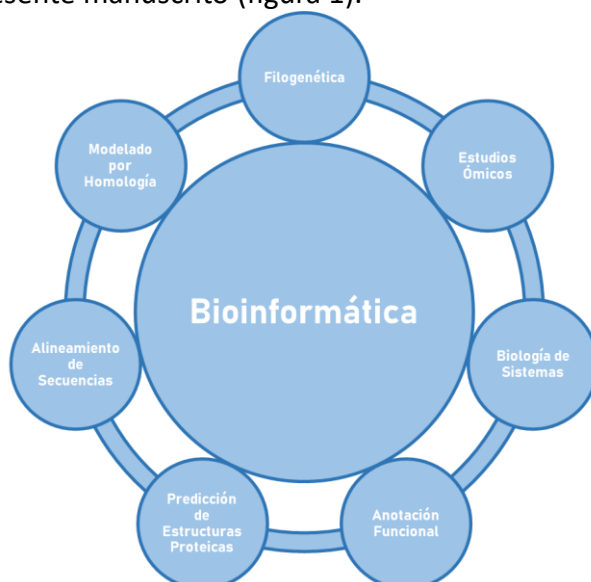


Figura 1. La bioinformática y sus principales enfoques de estudio. Tomado y adaptado de (Diniz & Canduri, 2017).

Los metaanálisis basados en la reutilización de datos biológicos se han convertido rápidamente en un pilar fundamental de la investigación en las ciencias de la vida. En el área de la bioinformática y las ciencias ómicas, se moviliza una importante magnitud de datos mayormente utilizados una única vez para luego ser depositados en repositorios públicos, por lo que su revalorización podría incrementar notoriamente su utilidad (Patra *et al.*, 2020). Existe cierto rebatimiento relacionado a la reutilización de datos, siendo los principales motivos la preocupación por su calidad y fiabilidad, falta de conciencia de su potencial, o bajo conocimiento bioinformático en minería de datos (Sielemann *et al.*, 2020; Tenopir *et al.*, 2020). Por ello, nos aseguramos de implementar el uso de los principios FAIR: encontrable, accesible, interoperable y reutilizable (del inglés, *Findability, Accessibility, Interoperability, and Reusability*). Estas reglas fueron definidas por la comunidad científica y destinadas a productores y publicadores de datos para sobrellevar los potenciales obstáculos en el campo de la bioinformática (Wilkinson *et al.*, 2016).

1.1 Insectos y desarrollo

Los insectos conforman una clase taxonómica de artrópodos de gran diversidad que comprenden más de la mitad de las especies de animales descritas (Tihelka *et al.*, 2021). La mayor parte de ellos puede clasificarse en dos grupos principales: holometábolos y hemimetábolos (figura 2). Los insectos holometábolos (desarrollo completo) pasan por un estado de larva luego de la embriogénesis, cuya morfología no cuenta con patas, alas, ni órganos reproductores. La larva pasa por mudas, sufre metamorfosis dentro de una pupa y sale en estadio adulto, con extremidades, alas y órganos reproductivos. Los hemimetábolos (desarrollo incompleto) emergen del huevo en ninfas en lugar de larvas, pasan por etapas de ninfa similares a un ejemplar adulto, pero sin alas ni órganos reproductores.

La ninfa se somete a varias mudas (sin formar una pupa), aumentando de tamaño progresivamente, y en la última muda desarrolla alas y órganos reproductivos, llegando a su estado adulto (Wanninger, 2015).

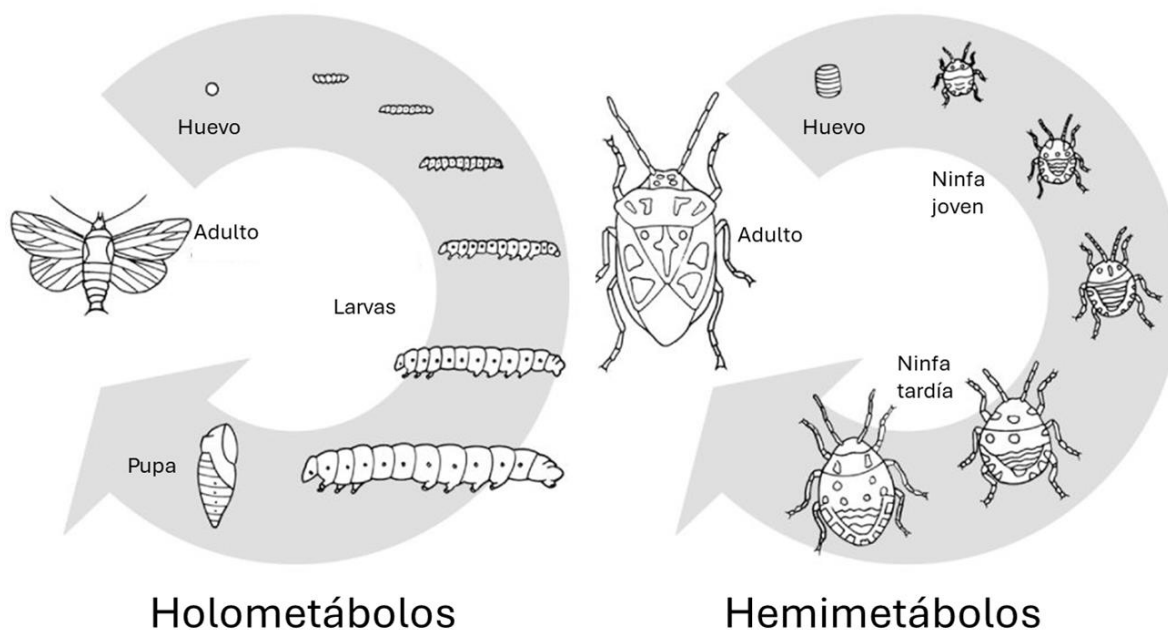


Figura 2. Esquema representativo de la metamorfosis completa (holometábolos) e incompleta (hemimetábolos) de los insectos. Tomado y traducido de <https://www.britannica.com/animal/insect/Natural-history>.

Los insectos comparten una disposición anatómica comúnmente denominada plan corporal, que consiste en una cabeza formada por la fusión de 6-7 segmentos, un tórax con 3 segmentos, y un abdomen compuesto por 8-11 segmentos (Rosenberg *et al.*, 2009). Si bien el proceso de embriogénesis presenta sus diferencias a lo largo de estos organismos, el plan corporal es definido gracias a la acción de varios grupos de genes encargados de dirigir la arquitectura corporal. Por un lado, se encuentran los genes efectores maternos, expresados en el ovario de la madre y productores de ARN mensajero depositados asimétricamente en el ovocito. Por otra parte, se encuentran los genes de segmentación cigóticos que incluyen los grupos de genes *gap*, *pair-rule* y *segment polarity*, activándose en ese mismo orden de forma jerárquica, lo que culmina con la otorgación de identidad a los segmentos corporales del insecto y la aparición de sus respectivos apéndices anatómicos (Gilbert, 2010; Rivera-Pomar & Jäckle, 1996).

Durante el desarrollo temprano, luego de la fertilización del ovocito, el cigoto atraviesa eventos críticos como la segmentación y la gastrulación. La segmentación en insectos incluye una serie de divisiones nucleares dentro de una gran y única membrana que recubre a los núcleos, y la migración de éstos hacia la periferia del embrión para formar el estadio de blastodermo sincitial. Posteriormente, esta membrana se invagina rodeando cada uno de los núcleos, definiendo individualmente a las células, y dando lugar al estadio denominado como blastodermo celular. Por otra parte, la gastrulación involucra una serie de movimientos celulares que relocalizan a las células para dar lugar a una nueva estructura llamada gástrula, a partir de la cual emergen las capas embrionarias presuntivas (mesodermo, endodermo y ectodermo) que darán origen a los futuros tejidos, órganos y sistemas. Los eventos posteriores a la gástrula difieren entre especies, pero a nivel general se mantienen los procesos de elongación/retracción de la banda germinal, la segmentación procedural de las distintas regiones corporales (cabeza, tórax, abdomen), la formación de apéndices y el cierre dorsal (Gilbert, 2010).

1.2 Neuropéptidos y hormonas proteicas

Los neuropéptidos son un grupo muy diverso de pequeñas moléculas de señalización. En insectos 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 (figura 3). Los neuropéptidos 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).

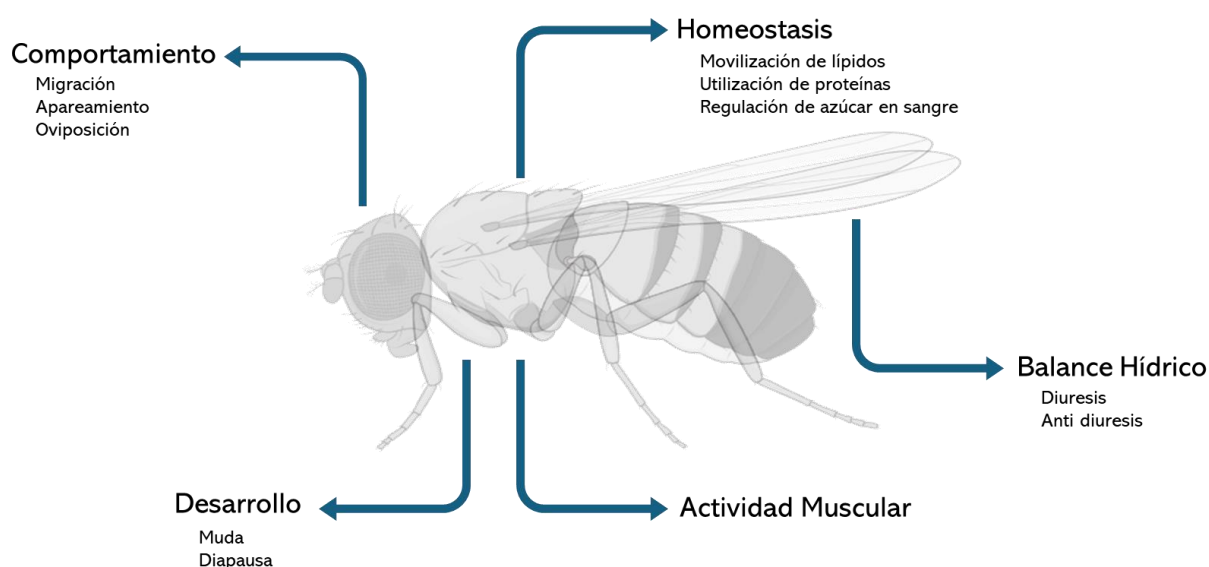


Figura 3. Participación y funciones descriptas de los neuropéptidos en insectos adultos. Tomado y modificado de (Scherkenbeck & Zdobinsky, 2009)

Los neuropéptidos son codificados por un gen precursor que genera un pre-propéptido precursor incluyendo un péptido señal en posición N-terminal de unos 16 a 30 aminoácidos, el cual indica el ingreso al retículo endoplasmático (RE), donde los ribosomas asociados traducen el ARN mensajero y el péptido señal es removido por una enzima peptidasa, liberando al precursor en el lumen del RE. Una vez que son enviados al aparato de Golgi sufren ciertas modificaciones postraduccionales previamente a ser transferidos a gránulos secretorios. Estos péptidos son alterados en el extremo N-terminal, donde un residuo glutámico es oxidado a piroglutamato, al igual que otro tipo de cambios como acetilaciones, fosforilaciones, glicosilaciones, entre otras. Además, pueden ocurrir modificaciones en el extremo C-terminal, que al encontrarse el aminoácido glicina en dicha posición es transformado en una amida mediante la acción de una enzima monooxigenasa (Ons, 2017). Seguidamente, el precursor es procesado por enzimas convertasas en uno o más neuropéptidos maduros luego del clivaje en sitios básicos o con pares dibásicos de aminoácidos Lisina (K) y Arginina (R): KR, KK, RK y RR (Baggerman *et al.*, 2005; Zeng *et al.*, 2021). Aun así, las regiones peptídicas linderas a la secuencia funcional son muy poco conservadas, dificultando su reconocimiento y caracterización (Veenstra, 2019). La figura 4 resume el proceso de generación de un neuropéptido maduro, desde su síntesis hasta su liberación extracelular.

En algunas ocasiones los precursores pueden dar lugar a varias copias de péptidos maduros, sea que se trate de un único tipo de neuropéptido o de distintas formas, sin importar que estén relacionadas estructuralmente o no. Luego de ser secretados, los neuropéptidos y proteínas hormonales actúan generalmente sobre receptores acoplados a proteína G, lo que provoca una cascada de señalización que involucra segundos mensajeros junto a la vía de la proteína quinasa u otras, para desencadenar la respuesta celular correspondiente (Nässel & Larhammar, 2013).

Diferentes células neuronales pueden transcribir distintos subtipos de neuropéptidos. Por ejemplo, se ha demostrado la expresión de la hormona protoracicotrópica (PTTH) en cuatro pares de células neurosecretoras en la región dorso-medial del cerebro de *Bombyx mori*, a través de ensayos de inmunohistoquímica (Ishizaki & Suzuki, 1988). Las variaciones en la expresión espacial y temporal de los neuropéptidos y sus receptores explican la diversidad fenotípica en el comportamiento resultante del organismo en cuestión.

Basado en información sobre genomas anotados de varios organismos de estudio, como ciertas especies de *Drosophila sp.*, *Anopheles gambiae*, *Aedes aegypti*, *Bombyx mori* o *Apis mellifera*, se conoce que en cada especie hay entre 30 y 40 genes que codifican precursores de neuropéptidos (Altstein & Nässel, 2010). Dado que varios de los pre-propéptidos precursores codifican más de un neuropéptido predicho, es posible que haya más de 40 neuropéptidos u hormonas peptídicas funcionales en una especie (Altstein & Nässel, 2010). Se encuentran descritas unas 50 familias de neuropéptidos a lo largo de 400 especies incluidas en la base de datos para la investigación de neuropéptidos en insectos (DINer) (Yeoh *et al.*, 2017).

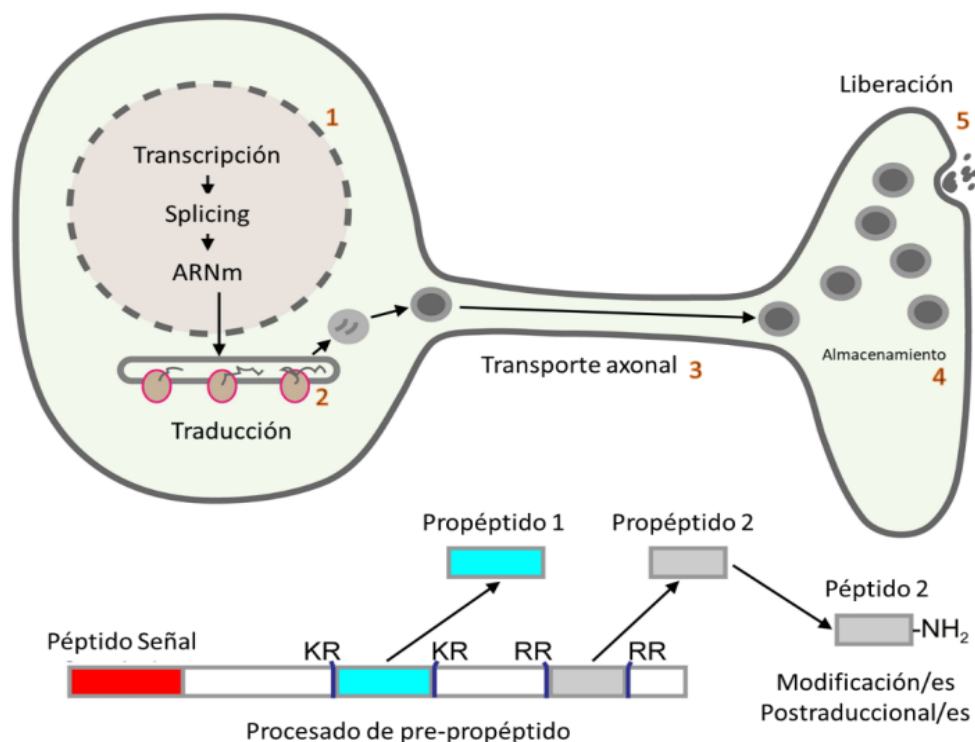


Figura 4. Expresión, procesamiento y secreción de los neuropéptidos. En la imagen superior se observa una célula neurosecretora donde se detalla: (1) transcripción del gen precursor y exportación de ARN mensajero maduro. (2) Traducción del precursor en retículo endoplasmático por la presencia del péptido señal que lo guía. (3-4) Procesamiento, transporte a través del axón y almacenamiento del precursor envuelto en vesículas. (5) Secreción del neuropéptido hacia el medio extracelular. En posición inferior se muestra un esquema del pre-propéptido con su péptido señal (rojo), sitios de clivaje (azul) y sus correspondientes díptidos (K=Lisina, R=Arginina), generando dos propéptidos (celeste y gris) que sufrirán modificaciones postraduccionales. Tomado y modificado de Yeoh *et al.* (2017).

1.3 Secuenciación de ARN y expresión génica diferencial

La secuenciación de ARN (RNA-Seq) es un abordaje que utiliza la tecnología de los métodos de secuenciación de nueva generación para estudiar el transcriptoma, además de proveer información sobre los niveles de expresión de los transcritos analizados (Singh *et al.*, 2012). Algunos de los usos más comunes de esta técnica son el estudio de *splicing* alternativo, la identificación de nuevos genes y variantes, detección de microorganismos, y la expresión génica diferencial (Han *et al.*, 2015).

Chatterjee y colaboradores (2018) han determinado que algunos de los parámetros más importantes a tener en cuenta a la hora de comenzar un proyecto de RNA-Seq son:

- Secuenciación *single-end* o *paired-end*: secuenciar únicamente la hebra en dirección 5'→3' o secuenciar ambas, respectivamente. Esto último permite una identificación más robusta de rearrreglos estructurales, variantes de *splicing*, entre otras.
- Bibliotecas *stranded* o *unstranded*: la información genética de la hebra utilizada para secuenciar el ARN se mantiene en bibliotecas *stranded*, lo que mejora la identificación de transcritos antisentido y, el desempeño al trabajar con genes superpuestos en hebras opuestas.
- Mayor cobertura, y/o mayor cantidad de réplicas: la cobertura hace referencia al número de secuencias que cubren una región genómica. Debido a la magnitud de expresión de los genes, la complejidad del transcriptoma o la muestra de partida existe variación entre distintas muestras, por lo que el uso de réplicas es común.
- Número y tipo de réplicas: trabajar con réplicas biológicas disminuyen la variación entre individuos, mientras que réplicas técnicas permiten estimar los errores de medición.

El análisis de expresión génica diferencial (DGE, del inglés *differential gene expression*) permite estudiar de manera cuantitativa aquellos genes que tengan niveles de ARN mensajero significativamente distintos entre dos o más condiciones (McDermaid *et al.*, 2019). Esta comparación puede estar asignada, por ejemplo, entre tratamientos y controles o estudios longitudinales entre distintos estadios del desarrollo. Si bien no existe un único flujo de trabajo óptimo para los diversos casos que puedan presentarse, hay una serie de pasos fundamentales a ser ejecutados (figura 5). Teniendo a disposición los datos crudos de secuenciación de ARN, inicialmente se realiza un control de calidad sobre ellos para definir si es necesario efectuar una limpieza o remoción de *reads* o adaptadores. Con los datos purificados se lleva a cabo un alineamiento al genoma de referencia, a partir del cual se cuantifican los transcritos alineados a regiones exónicas, y se realiza un contraste estadístico enfrentando dos condiciones para identificar aquellos genes que hayan aumentado o disminuido su expresión (Conesa *et al.*, 2016).

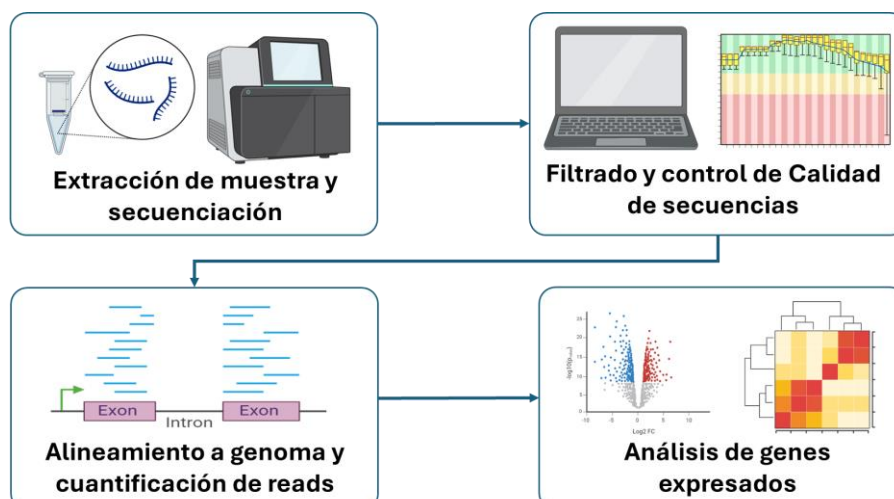


Figura 5. Esquema de trabajo convencional en el análisis de expresión génica diferencial. Ver sección 3.4.3.

1.4 Aclaraciones sobre la correcta caracterización de neuropéptidos

Es necesario mencionar que la detección y caracterización de neuropéptidos puede traer consigo una serie de complicaciones debido a las particulares propiedades que los mismos presentan. Veenstra (2019) detalla ciertas consideraciones a tener en cuenta para lograr una caracterización adecuada. Baricalla (2020) resume estos puntos, los cuales se mencionan 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 (Grabherr et al., 2011) (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.

1.5 Antecedentes de Neuropéptidos en el desarrollo

La hormona protoracicotrópica (PTTH) es un neuropéptido liberado por células neurosecretoras del cerebro que activa las glándulas protorácicas para secretar ecdisona, la cual dirige procesos como la muda y la metamorfosis (Ishizaki, 2004). Estudios previos han reportado una alta homología entre las secuencias aminoacídicas de PTTH en el lepidóptero holometábolo *Bombyx mori* con los péptidos de la familia de la insulina y factores de crecimiento de tipo insulina en vertebrados (Nagasawa et al., 1984). Además, en dicho organismo, existe evidencia de la participación de PTTH en la vía de señalización de proteínas tipo insulina a través de su unión a un receptor de tirosin quinasa conocido como "Torso" en *Drosophila melanogaster*, que actúa a nivel de desarrollo embrionario (Smith et al., 2014).

Investigaciones previas de nuestro laboratorio (Baricalla, 2020) han estudiado la expresión de neuropéptidos en *Oncopeltus fasciatus* a lo largo de una serie temporal en estadios embrionarios y en adultos. Comparando la expresión de los mismos entre las diferentes etapas mencionadas no se detectaron neuropéptidos maduros en embriones, por lo que en su lugar estos podrían existir en su estado de pre-propéptidos como reservorios para etapas de desarrollo subsiguientes.

Asimismo, muchos de estos neuropéptidos carecen de funciones descritas para embriones, de manera que otra posibilidad sugiere que estos precursores podrían poseer funciones específicas en el estadio embrionario, diferentes a las descritas en ninfas o adultos hasta la fecha (Baricalla, 2020).

Estos antecedentes sugieren que podría existir una mayor cantidad de neuropéptidos con patrones similares a los encontrados en PTTH en diferentes organismos. En el presente trabajo, se tiene como objetivo determinar la presencia genómica de neuropéptidos y sus transcritos codificantes en estadios tempranos del desarrollo mediante expresión génica diferencial en un conjunto representativo de especies que contempla los principales órdenes taxonómicos de la clase Insecta.

2 Hipótesis y Objetivos

2.1 Hipótesis

Existe expresión transcriptómica de neuropéptidos a lo largo de los distintos órdenes de insectos durante estadios tempranos del desarrollo.

2.2 Objetivo General

Comprobar la expresión transcriptómica de neuropéptidos en estadios tempranos del desarrollo en especies representativas de diferentes órdenes de insectos, mediante ensayos de expresión génica diferencial.

2.3 Objetivos específicos

- Recopilar datos de RNA-Seq en estadios embrionarios de diferentes especies de insectos a partir de repositorios públicos.
- Determinar la calidad y realizar la limpieza de los datos seleccionados para realizar un alineamiento de dichas secuencias a su genoma anotado.
- Efectuar un análisis de expresión génica diferencial con foco en los genes que codifican neuropéptidos y proteínas hormonales.
- Comparar los resultados obtenidos en estadios embrionarios equiparables entre las especies seleccionadas.

2.4 Palabras clave

Insectos, embriones, neuropéptidos, RNA-Seq, expresión génica diferencial

3 Metodología

3.1 Búsqueda y recolección de datos

La búsqueda inicial de datos de RNA-Seq se realizó mediante la plataforma *Sequence Read Archive* (SRA) de NCBI (<https://www.ncbi.nlm.nih.gov/sra/>) (Katz *et al.*, 2021), filtrando por especies de insectos comúnmente estudiadas, acompañadas de las palabras clave “egg” (huevo) o “embryo” (embrión).

Los datos de interés corresponden principalmente a proyectos que contienen transcriptomas de estadios embrionarios tempranos, de los cuales fueron priorizados aquellos en torno a una serie temporal que permita comparar la expresión de neuropéptidos en distintos momentos del desarrollo temprano (ver sección 4.1).

Además, se seleccionaron solo aquellos proyectos que contengan secuencias realizadas con la tecnología de Illumina, para mantener los estándares de secuenciación lo menos variables posible, ya que son los más comúnmente utilizados en la última década para expresión génica diferencial.

La selección de especies incluidas en este trabajo abarca los principales órdenes taxonómicos de la clase Insecta, teniendo en cuenta la disponibilidad de genomas secuenciados y anotados para cada especie incluida. La tabla 1 indica la lista de especies seleccionadas para este trabajo, junto con los proyectos que incluyen los datos de secuenciación en SRA utilizados de cada una de ellas.

Tabla 1. Especies de insectos seleccionadas. Se indica el tipo de desarrollo que posee (holometábolo o hemimetábolo), orden, nombre específico y proyectos de secuenciación asociados.

Tipo de desarrollo	Orden	Especie	Proyecto
Hemimetabolía	Orthoptera	<i>Schistocerca gregaria</i>	PRJNA524786
	Blattodea	<i>Blattella germanica</i>	PRJNA389591
		<i>Nilaparvata lugens</i>	PRJNA579876
	Hemiptera	<i>Rhodnius prolixus</i>	PRJNA694974, PRJNA924824
		<i>Oncopeltus fasciatus</i>	PRJNA391717
Holometabolía	Coleoptera	<i>Tribolium castaneum</i>	PRJNA645519, PRJNA490167
		<i>Nasonia vitripennis</i>	PRJNA701367
	Hymenoptera	<i>Monomorium pharaonis</i>	PRJNA767561
		<i>Apis mellifera</i>	PRJNA668751, PRJNA257666
	Lepidoptera	<i>Bombyx mori</i>	PRJNA675719
		<i>Anopheles gambiae</i>	PRJNA866506
	Diptera	<i>Aedes aegypti</i>	PRJNA209388
		<i>Drosophila melanogaster</i>	PRJNA495910

3.2 Predicción e identificación de neuropéptidos en cada especie

Teniendo en cuenta la masividad de secuencias totales de neuropéptidos a revisar, se automatizó la búsqueda, predicción e identificación de estos. Se utilizó la herramienta Proteinortho (Klemm *et al.*, 2023), la cual permite detectar genes ortólogos y parálogos partiendo de comparaciones entre secuencias de ADN o proteínas. Realiza una comparación de todas las proteínas anotadas en el genoma (proteoma) para determinar similitudes, un agrupamiento de proteínas similares en grupos ortólogos y culmina con la generación de un archivo de salida en formato tabla asignando los resultados propios a cada especie. En este caso, se ejecutó el programa con las anotaciones proteómicas de cada especie en formato FASTA, descargadas desde NCBI.

Las anotaciones fueron comparadas contra una base de datos curada manualmente, reuniendo las secuencias de neuropéptidos y proteínas hormonales anotadas hasta la actualidad, de las diferentes especies.

Los resultados crudos obtenidos fueron llevados a una planilla de Excel (Microsoft Corporation, 2019) para ser filtrados, descartando así las proteínas que no fueron objeto de estudio.

Se tomó como punto de partida la lista de péptidos de la base de datos (manualmente curada por expertos) para la investigación de neuropéptidos en insectos (DINeR) (Yeoh *et al.*, 2017) para crear una tabla indicando la presencia o ausencia de cada uno, en cada especie estudiada. La tabla 2 muestra los neuropéptidos que fueron objeto de estudio en este trabajo.

Para las proteínas cuya identidad no se logró designar inicialmente, ya sea catalogadas como “proteínas no caracterizadas” o aquellas con resultados dispares, se realizaron digestiones con tripsina a través de la herramienta PeptideCutter (Gasteiger *et al.*, 2005) (https://web.expasy.org/peptide_cutter/), tomando las secuencias proteicas de NCBI. Los fragmentos obtenidos se identificaron utilizando la herramienta Peptide search de UniProt (<https://www.uniprot.org/peptide-search>, (Zaru & Orchard, 2023).

Se obtuvieron los identificadores génicos de neuropéptidos de cada especie a través de herramientas de NCBI como Batch Entrez y tBLASTn (Altschul *et al.*, 1990), utilizando como *query* los identificadores proteicos extraídos de Proteinortho (ver sección 9 “Anexo”).

Finalmente, se realizó una curación manual de los datos recolectados. Esto permitió dilucidar ciertos grupos de neuropéptidos similares en secuencia peptídica (por ejemplo, el grupo de alatostatinas), identificando con mayor precisión los hallados en cada especie. Además, dicha curación facilitó la incorporación de varios neuropéptidos pasados por alto en la búsqueda automatizada, para los cuales se verificó su presencia de manera bibliográfica (Yeoh *et al.*, 2017).

3.3 Ordenamiento de series transcriptómicas temporales de desarrollo entre especies

Debido a las diferencias genéticas, fisiológicas y temporales a lo largo del desarrollo embrionario en las distintas especies seleccionadas, las series temporales fueron asignadas a los principales eventos del desarrollo que representan, con el fin de hacer posible la comparación interespecífica de datos. Para ello, se efectuó una revisión bibliográfica de las etapas del desarrollo embrionario que permitiera localizar y agrupar los datos recolectados, lo que permitió complementar la descripción encontrada en la plataforma SRA de NCBI (ver sección 9 “Anexo”).

Se tuvieron en cuenta las descripciones indicadas en las publicaciones de origen de los datos de secuenciación, así como también aquellos artículos que describieran completamente el desarrollo del insecto, de no ser suficiente con la primera opción. En este caso, se tuvo la precaución de respetar lo mayormente posible las condiciones de cría reportadas de los ejemplares procesados para la obtención de datos, principalmente de su temperatura debido a que dicho factor podría alterar la velocidad de desarrollo (Farnesi *et al.*, 2009).

Los datos fueron agrupados en cuatro secciones principales: huevo no fertilizado, embrión temprano, medio y tardío. El primero de estos parte de la disponibilidad de datos de este tipo en algunas de las especies del análisis, sin especificar su tiempo o etapa, interpretándose como etapa pre-embrionaria general por la corta descripción de los datos en su origen. Los otros tres grupos fueron establecidos teniendo en cuenta la formación del blastodermo (sin diferenciar entre blastodermo sincitial y celular), el proceso de gastrulación, y aquellos eventos posteriores al desarrollo (retracción de la banda germinal, cierre dorsal, especificación de segmentos corporales), respectivamente. Se registraron los tiempos totales de desarrollo embrionario para cada especie, de manera tal de representar porcentualmente las secciones y grupos de datos estudiados.

Tabla 2. Neuropeptidos analizados en este trabajo, junto a las abreviaturas con las cuales serán mencionados a lo largo del manuscrito.

Neuropeptido	Abreviatura	Neuropeptido	Abreviatura
Péptido relacionado a corazonina	ACP	Hormona paratiroidea de insecto	IPTH
Factor antidiurético	ADF	Péptido relacionado con la insulina	IRP
Agatoxina	AGA	ITG	ITG
Hormona adipoquinética	AKH	Péptido transportador de iones	ITP
Apidaecina 1	APD	Quinina	KIN
Alatostatina A	ASTA	Leucoquinina	LK
Alatostatina B	ASTB	Limostatina	LST
Alatostatina C	ASTC	Miosupresina	MS
Alatostatina doble C	ASTCC	Neuroligina	NLGN
Alatostatina triple C	ASTCCC	Proteína neuroendócrina 7B2	NP7B2
Alatotropina	AT	Neuroparsina	NPA
Vasopresina	AVLP	Neuroparsina (<i>Net Platelet Angiogenic Activity</i>)	NPAA
Péptido cerebral	BP	Neuropeptido F largo	NPF
Bursicon alfa	BURS	Proteína precursora de neuropeptidos	NPLP
Calcitonina A	CALa	Natalisina	NTL
Calcitonina B	CALb	NVP	NVP
Capability	CAPA	Orcoquinina	ORCK
Precursor de neuropeptido similar a Carausius	CNLP	Neuropeptido activador de la biosíntesis de feromonas	PBAN
Péptido cardioactivo crustáceo	CCAP	Asociado a bursicon	PBURS
CCH-amida	CCH	Factor de dispersión de pigmentos	PDF
CNM-amida	CNM	Precursor de neuropeptido similar a Periplaneta	PNLP
Factor liberador de corticotropina - relacionado con hormona diurética	CRF-DH	Piroquinina	PK
Corazonina	CRZ	Prohormona 2	PRO2
Hormona diurética 31	DH31	Prohormona 4	PRO4
Hormona diurética 44	DH44	Proctolina	PROC
Hormona de eclosión	EH	Hormona protoracicotrópica	PTTH
Elevenina	ELE	RFL-amida	RFL
Hormona disparadora de la ecdisis	ETH	RY-amida	RY
FMRF-amida	FMRF	Péptido estimulante de salivación de la glándula salival	SGSSP
Gonadulina	GON	SIF-amida	SIF
Glicoproteína alfa 2	GPA2	Sulfaquinina	SK
Glicoproteína beta 5	GPB5	Neuropeptido F corto	SNPF
HAN	HAN	Péptido sexual	SP
Péptido cefálico	HP	Taquicinina	TK
Hugin	HUG	Trisina	TRIS
IDLSRF-like peptide	IDLSRF	Trunk	TRUNK
IMF-amida	IMF	Similar a Torso (<i>torso-like</i>)	TSL

3.4 Procesamiento de datos

El análisis bioinformático realizado con los sets de datos mencionados en el curso de este trabajo fue llevado a cabo en el servidor “Mulatona” del Centro de Cómputo de alto Desempeño (CCAD) de la Universidad Nacional de Córdoba (<https://ccad.unc.edu.ar/>). Dicho espacio es accesible de forma gratuita para cualquier miembro de la comunidad universitaria y/o científica nacional pública.

Los datos de RNA-Seq de los proyectos seleccionados (ver tabla 1) se importaron en formato FASTQ a través de SRA Explorer (<https://sra-explorer.info/>). Esta herramienta permite generar un *script* con los experimentos elegidos de cada proyecto, logrando una carga ágil y directa de datos al servidor. La figura 6 describe el flujo de trabajo implementado para el procesamiento de datos, detallado a lo largo de la sección 3.4.

3.4.1 Determinación de calidad y filtrado de datos

Si bien los datos que provienen de repositorios superan procesos de revisión al momento de ser publicados, se realizó un control de calidad de secuenciación para cada *set* de datos con la finalidad de revisar fehacientemente su condición. Esto fue realizado mediante la herramienta FastQC (Andrews, 2010), que permite analizar parámetros como cantidad y longitud de secuencias, calidad promedio, porcentaje de GC (Guanina + Citosina), secuencias sobre representadas y presencia de adaptadores.

Los resultados pertenecientes a una misma especie fueron revisados de manera conjunta a través del *software* MultiQC (Ewels *et al.*, 2016), que tiene la capacidad de tomar los resultados del programa anterior y generar un reporte completo con el total de archivos por proyecto.

En los casos en que se identificó baja calidad o presencia de adaptadores, estos fueron eliminados mediante Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), utilizando sus parámetros por defecto.

3.4.2 Alineamiento de *reads* contra los genomas anotados

Una vez revisada la calidad de las secuencias y realizada la limpieza correspondiente, se procedió a efectuar los alineamientos contra los genomas anotados en cada especie. Para ello, se importaron al servidor sus genomas de referencia y anotaciones genómicas de RefSeq desde NCBI. La tabla 3 indica las versiones utilizadas de cada genoma, junto con las referencias correspondientes.

Tabla 3. Versiones de los genomas de cada especie que fueron utilizados para realizar los alineamientos de los transcriptomas procesados y la identificación de neuropéptidos presentes.

Especie	Genoma	Referencia
<i>Schistocerca gregaria</i>	GCF_023897955.1_iqSchGreg1.2	Verlinden <i>et al.</i> , 2020
<i>Blattella germanica</i>	GCA_000762945.2_Bger_2.0	Harrison <i>et al.</i> , 2018
<i>Nilaparvata lugens</i>	GCF_000757685.1_NilLug1.0	Ma <i>et al.</i> , 2021
<i>Rhodnius prolixus</i>	GCA_000181055.3_Rhodnius_prolixus-3.0.3	Mesquita <i>et al.</i> , 2015
<i>Oncopeltus fasciatus</i>	GCA_000696205.2_Ofas_2.0	Panfilio <i>et al.</i> , 2019
<i>Tribolium castaneum</i>	GCF_000002335.3_Tcas5.2	Richards <i>et al.</i> , 2008
<i>Nasonia vitripennis</i>	GCF_009193385.2_Nvit_psr_1.1	Dalla Benetta <i>et al.</i> , 2020
<i>Monomorium pharaonis</i>	GCF_013373865.1_ASM1337386v2	Gao <i>et al.</i> , 2020
<i>Apis mellifera</i>	GCF_003254395.2_Amel_HAv3.1	Wallberg <i>et al.</i> , 2019
<i>Bombyx mori</i>	GCF_014905235.1_Bmori_2016v1.0	Lu <i>et al.</i> , 2020
<i>Anopheles gambiae</i>	GCF_000005575.2_AgamP3	Holt <i>et al.</i> , 2002
<i>Aedes aegypti</i>	GCF_002204515.2_AaegL5.0	Matthews <i>et al.</i> , 2018
<i>Drosophila melanogaster</i>	GCF_000001215.4_Release_6_plus_ISO1_MT	Adams <i>et al.</i> , 2000

Las *reads* fueron mapeadas utilizando la herramienta HISAT2 (Kim *et al.*, 2015). Esta hace uso de un índice genómico global construido previamente para buscar rápidamente las posiciones dentro del genoma. Los archivos de salida en formato SAM fueron comprimidos mediante SAMtools (Danecek *et al.*, 2021) para obtener archivos BAM. La calidad del mapeo de los archivos obtenidos en formato BAM fue evaluada mediante QualiMap (Okonechnikov *et al.*, 2016), y los resultados fueron visualizados de manera global utilizando MultiQC (Ewels *et al.*, 2016).

3.4.3 Cuantificación de *reads*, normalización y expresión génica diferencial

La cuantificación del número de *reads* mapeadas se realizó con el programa featureCounts (Liao *et al.*, 2014). Los resultados de la cuantificación fueron descargados y procesados de manera local haciendo uso del lenguaje de programación R (R Core Team, 2022) a través de RStudio (Posit team, 2022). La normalización de los datos y los ensayos de expresión génica diferencial se efectuaron utilizando las librerías edgeR (Robinson *et al.*, 2010) y ggplot2 (Wickham, 2009) para la visualización de resultados. Para aquellos proyectos que no contaban con réplicas técnicas o biológicas se reemplazó la librería edgeR por NOIseq (Tarazona *et al.*, 2011). Este abordaje permite simular las réplicas tomando como punto de partida la variabilidad intrínseca de los datos, buscando genes que muestren suficiente evidencia de cambios en su expresión.

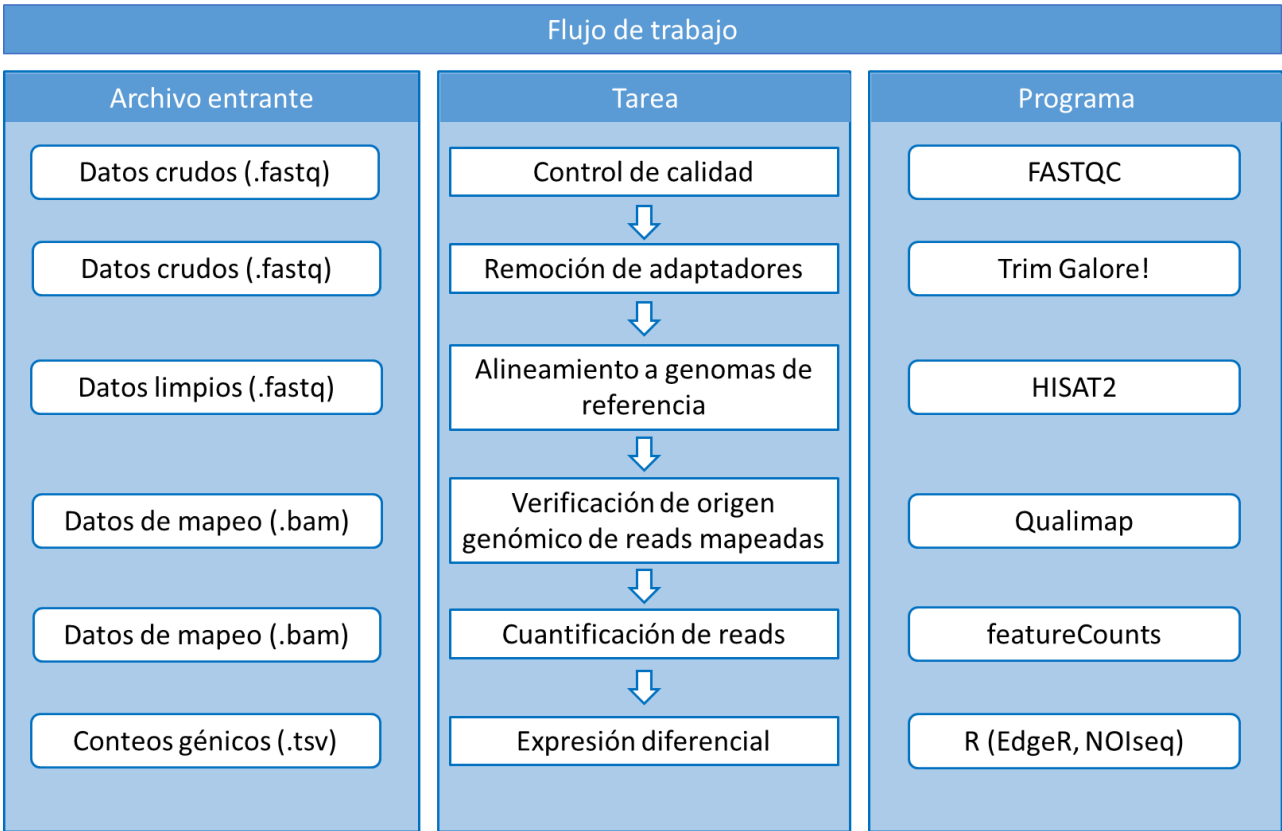


Figura 6. Flujo de trabajo a partir de los datos de RNA-Seq recolectados, para el análisis de expresión génica diferencial. El procesamiento desde la carga de datos hasta la cuantificación de *reads* fue realizado haciendo uso del *cluster* “Mulatona” del CCAD. Posteriormente los datos fueron trabajados de manera local en R.

4 Resultados

4.1 Comparativa del desarrollo embrionario entre especies

La búsqueda bibliográfica permitió establecer un marco comparativo para el desarrollo embrionario de las especies en estudio. Por un lado, se logró localizar con mayor precisión ciertos grupos de datos para los cuales no se tenía una referencia que acompañara a los datos transcriptómicos recolectados. Por otra parte, los valores de tiempos de desarrollo embrionario total se relativizaron porcentualmente para poder estimar la fracción que representa cada etapa graficada en el tiempo total. Para facilitar la comparativa posterior de expresión de neuropéptidos, se agruparon los eventos embriológicos en tres secciones: embrión temprano, medio y tardío (tal como se describe en la sección 3.3).

Se puede observar una distribución variada de los datos en las diferentes especies, encontrando ausencia de datos tardíos o la totalidad de datos en embrión temprano. Esto depende de la finalidad original con la que fueron generados en los trabajos de procedencia (figura 7).

Dentro de la etapa de embrión temprano se hallan casos como *S. gregaria*, *A. aegypti*, o *D. melanogaster* con un único set de datos perteneciente a este estadio, y otros en el extremo opuesto como *R. prolixus* o *N. vitripennis* con un número mayor de transcriptomas. Las especies *B. germanica*, *R. prolixus* y *A. gambiae* contienen adicionalmente datos de huevos no fertilizados.

Los datos se remarcaron tal como figuran descritos en la base de datos de la cual fueron tomados (SRA-NCBI), designando aquellos pertenecientes a un tiempo prolongado con líneas horizontales, y con una línea vertical a los que fueron generados en un horario puntual durante el desarrollo (figura 7). Todos los transcriptomas que se encuentren en una misma etapa de desarrollo embrionario (huevo no fertilizado, temprano, medio o tardío) son considerados como comparables entre sí a los efectos prácticos de este trabajo.

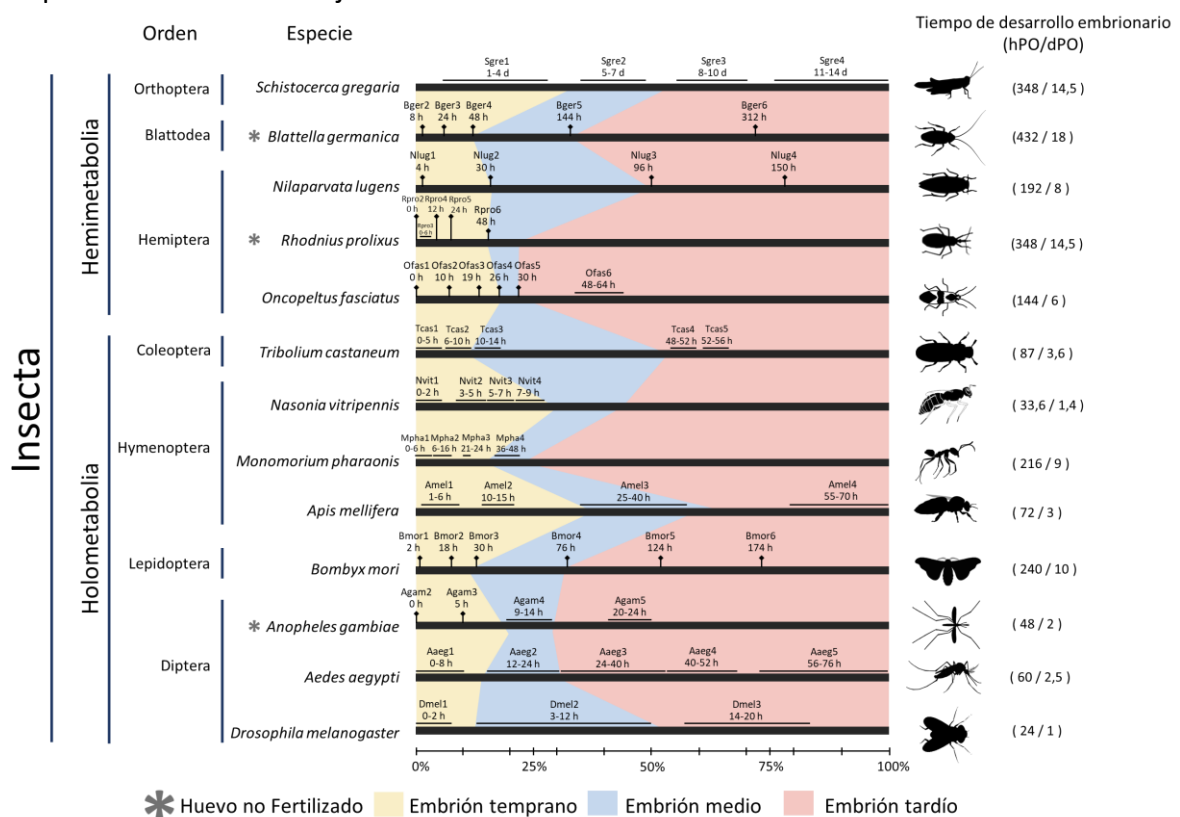


Figura 7. Alineamiento de las etapas embrionarias del desarrollo. Se muestra la posición relativa (0-100%) de los sets de datos de RNA-Seq de cada especie. Las áreas coloreadas marcan las tres regiones comparables a través de las especies; embrión temprano (amarillo), medio (azul) y tardío (rojo). Las etapas fueron establecidas como se describe en la sección 3.3. Adicionalmente se indica la disponibilidad de datos de huevos no fertilizados en ciertas especies (*). A la derecha, el tiempo total de desarrollo es indicado en horas/días después de la oviposición (hPO/dPO).

4.2 Procesamiento y evaluación de calidad de datos

Los datos de RNA-Seq recolectados presentaron un buen estado de calidad. Del total de parámetros evaluados por la herramienta FastQC (Andrews, 2010), se puso principal interés en aquellos tales como la calidad por base y promedio de la secuencia, porcentaje de citosina-guanina (%GC), secuencias sobrerepresentadas y presencia de adaptadores de secuenciación. De todos estos, el que mayormente presentó resultados regulares fue el %GC, observándose picos dispares en los gráficos respectivos, y un mayor porcentaje del esperado para algunas de las especies. Los adaptadores presentes en las muestras lograron ser removidos mediante Trim Galore!.

Los porcentajes de secuencias duplicadas presentes en los datos se ubicaron desde el 40% al 90%, encontrando valores atípicos que sobrepasan estos extremos siendo el mínimo un 22% y el máximo un 99%. Estos resultados son variables de manera intra- e interespecífica, aun teniendo en cuenta aquellos datos correspondientes a un mismo proyecto de RNA-Seq. La longitud media de las *reads* abarcó desde los 36 pares de base (pb) para algunas muestras de *A. mellifera*, hasta los 301 pb en *B. germanica*. La profundidad de secuenciación mostró un promedio de 26 millones de *reads* (M), encontrando muestras desde 3 M hasta los 88 M (Figura 8).

La figura 9 muestra algunos parámetros relacionados a la calidad de secuenciación, tomando de ejemplo a la especie *S. gregaria* para representar los resultados obtenidos en este análisis. La tendencia de dichos resultados se mantuvo para el resto de las especies.

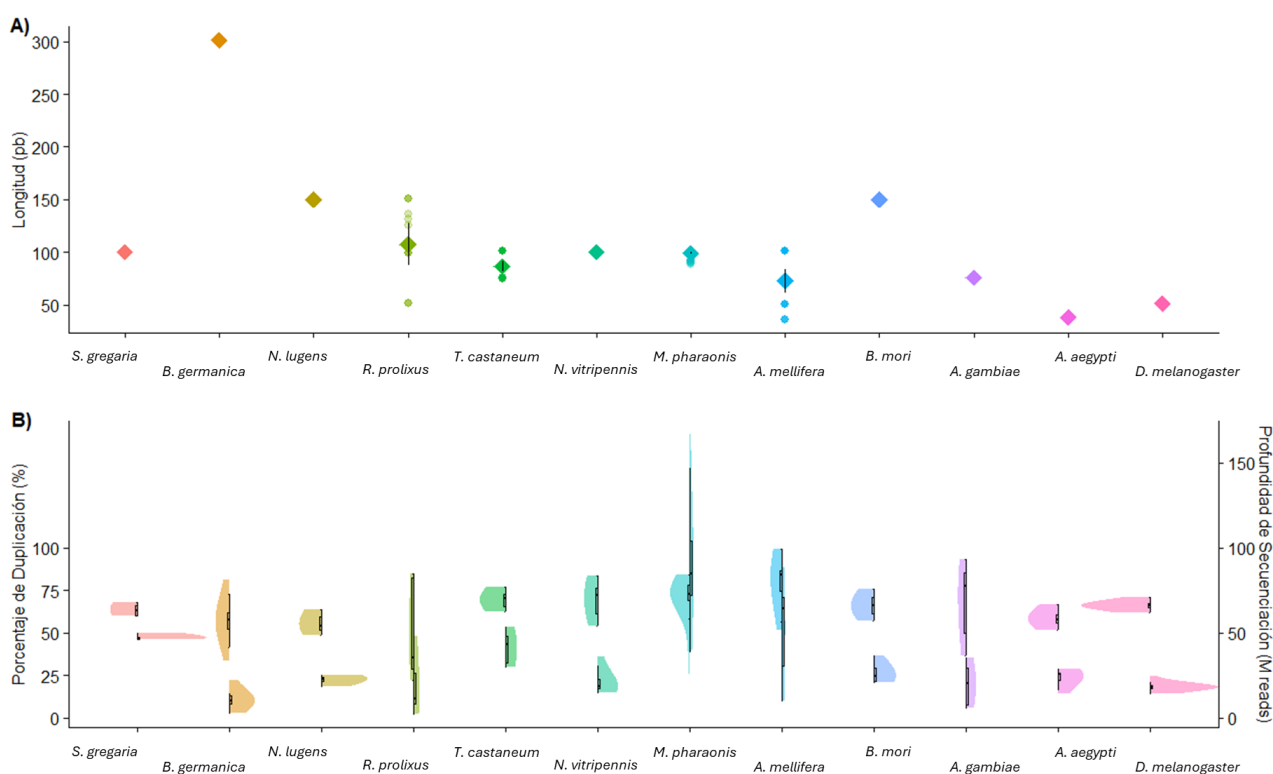


Figura 8. Estadísticas generales de la calidad de los grupos de datos, para cada una de las especies. **A)** Longitud media de *reads*. Es importante destacar la utilización de dos proyectos de secuenciación en las especies *R. prolixus*, *T. castaneum* y *A. mellifera*, razón por la cual las mismas presentan variabilidad en sus métricas. **B)** Porcentaje de duplicación de *reads* (eje izquierdo) y profundidad de secuenciación en millones de *reads* (eje derecho). La posición de los ejes se corresponde con la orientación de las áreas de densidad.

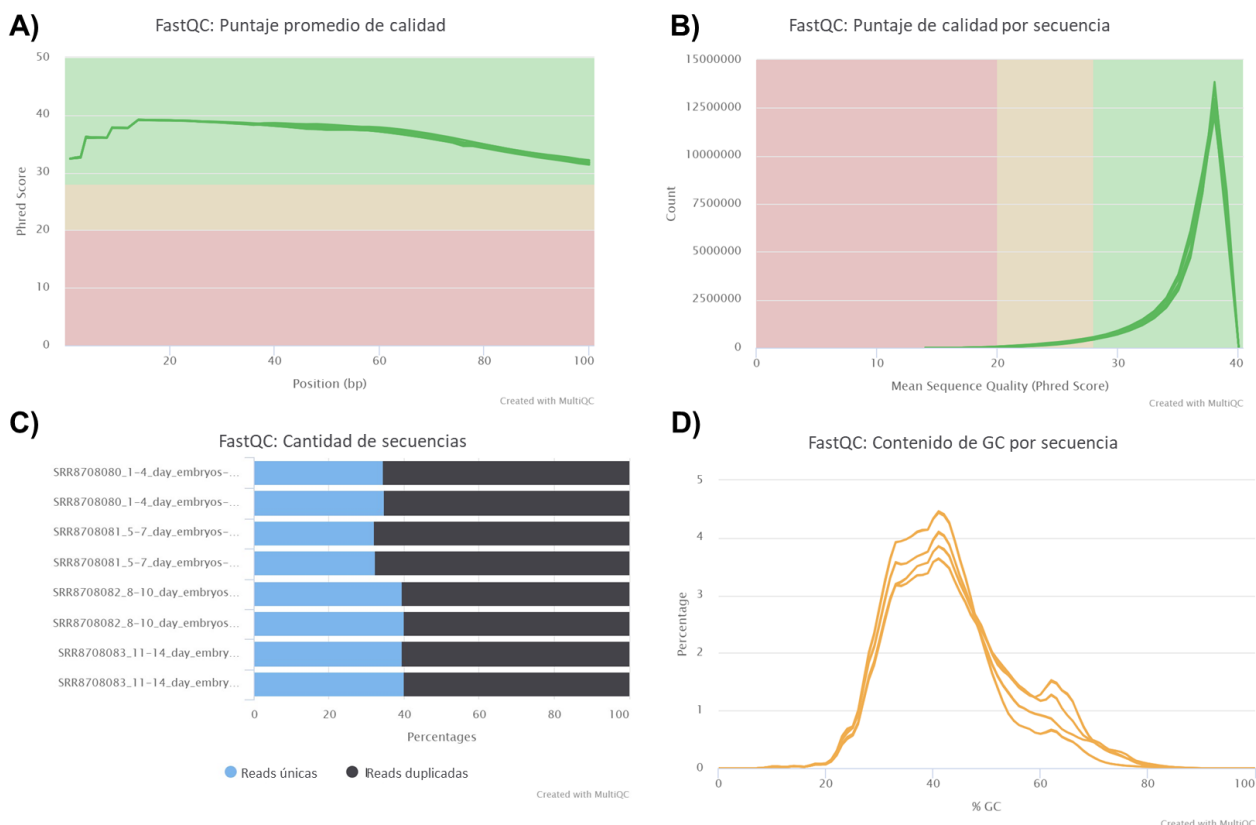


Figura 9. Ejemplo representativo del control de calidad realizado por FastQC, para la especie *Schistocerca gregaria*. Se muestran algunos de los parámetros de mayor interés al analizar datos de RNA-Seq. **A)** Puntaje promedio de calidad por posición en cada *read*. **B)** Puntaje promedio de calidad por conteo total de secuencias. **C)** Porcentaje de secuencias únicas y sobrerrepresentadas para cada muestra que integra el proyecto. **D)** Distribución del contenido porcentual de guanina-citosina (%GC) en las *reads* analizadas.

4.3 Alineamientos contra los respectivos genomas anotados

Una vez removidos los adaptadores de secuenciación y PCR existentes, se alinearon las *reads* a los genomas de referencia correspondientes a cada especie. Los resultados del mapeo realizado con HISAT2 se muestran en la figura 10. En general se obtuvieron resultados satisfactorios, logrando porcentajes de alineamiento total superiores al 80%. En el caso de *B. germanica*, se observó un rango de mapeo al genoma de referencia del 34-65%. Cabe destacar que la base de datos de la cual se extrajo dicho genoma (NCBI) indicaba la presencia de contaminantes en el ensamble.

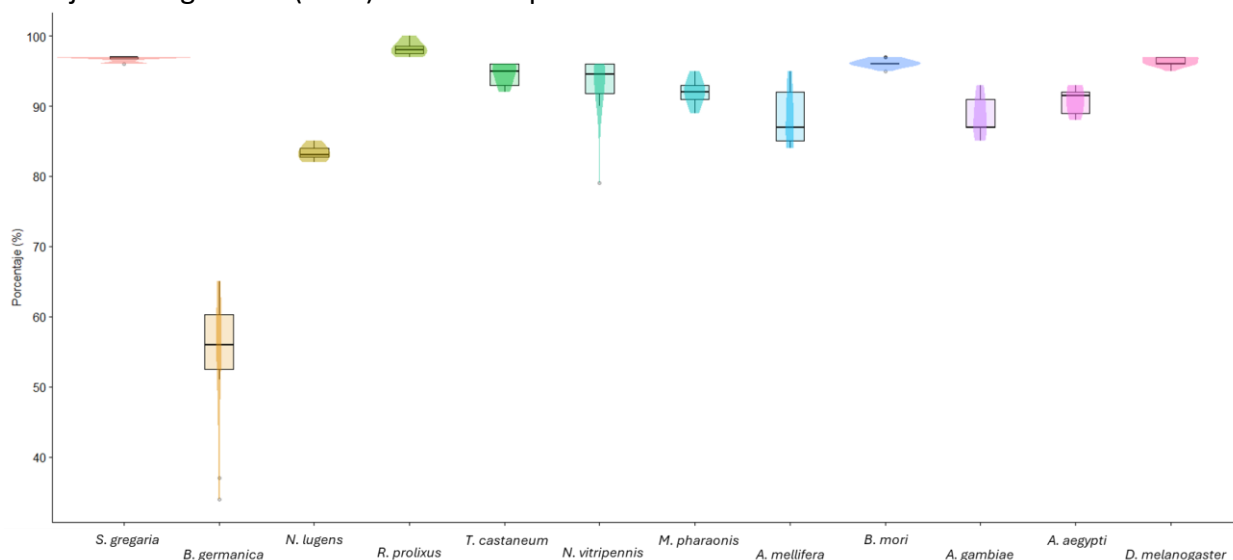


Figura 10. Porcentajes generales de mapeo de *reads* a los respectivos genomas de referencia de cada especie. Las versiones utilizadas de los genomas se encuentran para su consulta en la Tabla 3.

El análisis arrojado por Qualimap mostró un patrón similar en el origen de las *reads* a lo largo de la mayoría de las especies. A excepción de *A. aegypti* y *B. germanica*, cuyo porcentaje de *reads* de origen exónico fue del 48% y 36% respectivamente, el promedio para el resto de los organismos fue del 82,6% (62,7-98,4%). Es de interés resaltar el porcentaje de *reads* posicionadas en sitios intergénicos en *A. aegypti*, *A. gambiae* y *B. germanica*, y de aquellas de origen intrónico en *A. mellifera* y *B. germanica*. Los valores exactos resultantes del análisis en cada especie se muestran en la figura 11.

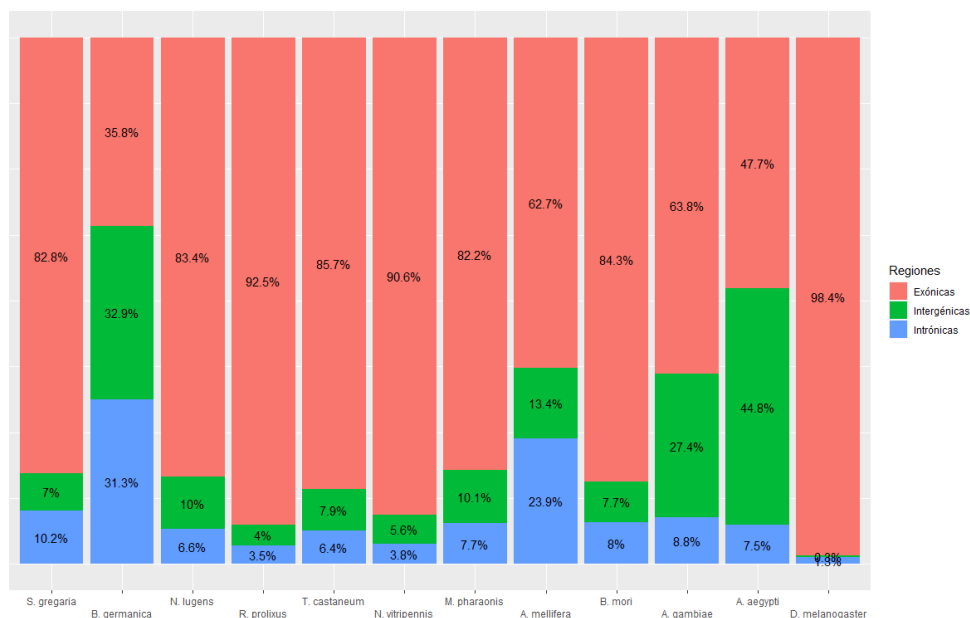


Figura 11. Origen genómico de las *reads* alineadas a los genomas de referencia.

4.4 Cuantificación, normalización y expresión génica diferencial

Las *reads* de los archivos BAM resultantes del alineamiento fueron cuantificadas mediante featureCounts. Las salidas de dicho programa fueron descargadas desde el servidor de trabajo del CCAD (Mulatona) y a partir de este punto fueron procesadas de manera local en R (R Core Team, 2022) a través de RStudio (Posit team, 2022). Se filtraron los genes que no superaron un valor de 0,5 conteos por millón (CPM – *counts per million*) en un número de réplicas respectivo a cada proyecto y especie.

Los perfiles de expresión de las muestras de cada especie fueron visualizados mediante gráficos de escalado multidimensional (MDS, del inglés *multidimensional scaling*), para lo cual se realizó una reducción dimensional a dos coordenadas principales. Se puede observar que la disposición espacial en grupos definidos es la esperada según los tiempos de desarrollo correspondientes a los datos (figura 12).

En un grupo de especies, *D. melanogaster*, *A. aegypti* y *A. gambiae*, el análisis mostró una superposición de muestras de tiempos cercanos, formando conjuntos mayores. Debido a esto, se optó por reorganizar los grupos manualmente teniendo en cuenta los eventos del desarrollo correspondientes a cada insecto, junto con lo comentado anteriormente para proceder con los siguientes pasos (figura 13). De esta forma, las “pseudorréplicas” otorgan mayor robustez estadística y una comparación más sencilla en torno a la cantidad de transcriptomas a comparar.

La normalización de los genes filtrados se realizó mediante el algoritmo TMM (del inglés *trimmed mean of M values*). Seguidamente se determinaron la dispersión de los datos mediante el algoritmo *quasi likelihood*, y el Coeficiente de Variación Biológica (CVB), con el fin de distinguir la variación génica específica de la proveniente por fuentes biológicas y/o técnicas.

El ajuste estadístico de los datos fue realizado mediante modelos lineales generalizados a través de edgeR (Robinson *et al.*, 2010). Cada archivo de datos fue acompañado de un archivo adicional indicando las agrupaciones de las muestras según correspondía. De esta forma, resultan definidos los grupos contrastados estadísticamente durante el ensayo de expresión diferencial, entre un estadio temporal respecto del anterior, dentro de cada especie.

Para aquellos datos sin réplicas, los ensayos de expresión diferencial se realizaron mediante NOIseq (Tarazona *et al.*, 2011) a través de la aplicación web IDEA (del inglés *Interactive Differential Expression Analyzer*) (Zhao *et al.*, 2018). En esta se cargaron los archivos salientes de featureCounts con un archivo indicador de muestras agrupadas, tal como se indicó en el párrafo anterior para edgeR.

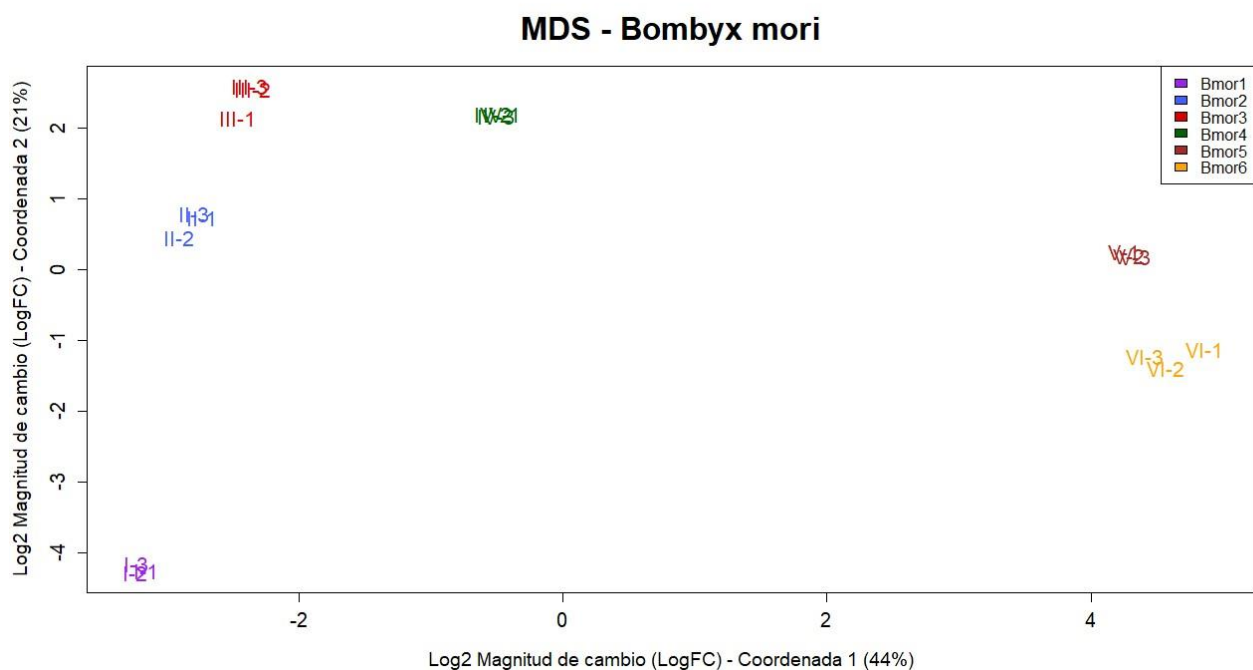


Figura 12. Análisis representativo de escalado multidimensional (MDS) para la especie *Bombyx mori*. Se puede observar que las repeticiones de cada tratamiento comparten una posición muy similar en el gráfico. Por otra parte, se aprecia una distancia marcada entre las muestras correspondientes a los distintos tiempos de desarrollo, indicando perfiles de expresión diferentes. Se observan tres repeticiones por cada muestra, agrupadas bajo un mismo color y nombre, tal como se muestra en la leyenda de la figura.

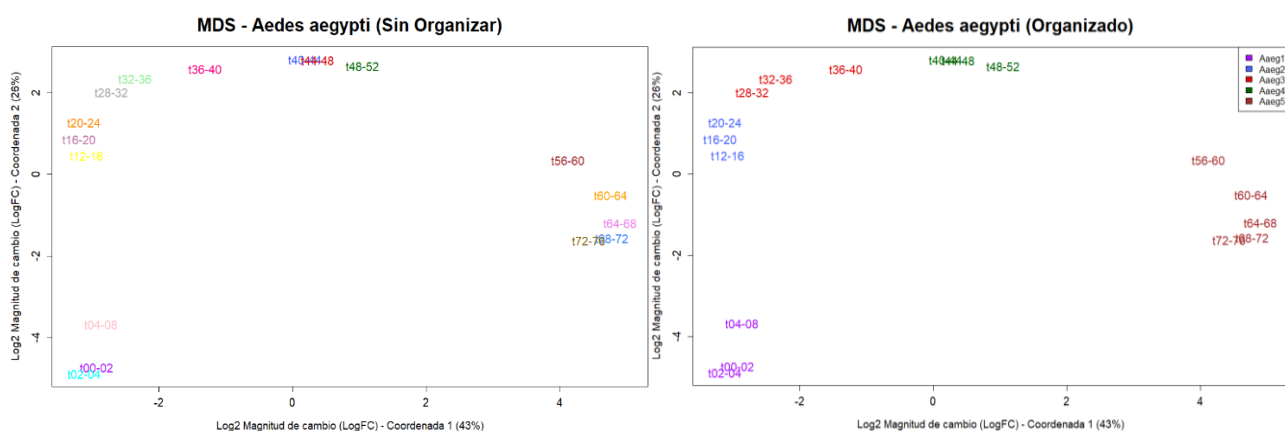


Figura 13. Análisis representativo de escalado multidimensional (MDS) para la especie *Aedes aegypti*. **Izquierda:** posición de los transcriptomas de trabajo. Se observa cierta superposición en algunos rangos temporales, dado que no existe gran variación en la expresión génica entre dichos datos. **Derecha:** transcriptomas organizados en 5 grupos, designados como Aaeg1-5 y diferenciados por colores. Este agrupamiento fue realizado teniendo en cuenta la proximidad de los datos observada en el gráfico, además de la descripción de los eventos embriológicos detallados en la bibliografía consultada.

4.5 Expresión de neuropéptidos a lo largo del desarrollo embrionario

4.5.1 Los neuropéptidos se identificaron a nivel embrionario

Una vez curada la lista de neuropéptidos totales de estudio junto con sus identificadores génicos en cada especie, se estableció una matriz binaria de presencia/ausencia con dichos datos. En la figura 14.A se distinguen aquellos neuropéptidos presentes en la mayoría de las especies (7 o más, hasta estar presentes en todas), mientras que en la figura 14.B se observan los neuropéptidos de menor presencia, llegando a aquellos hallados en una única especie.

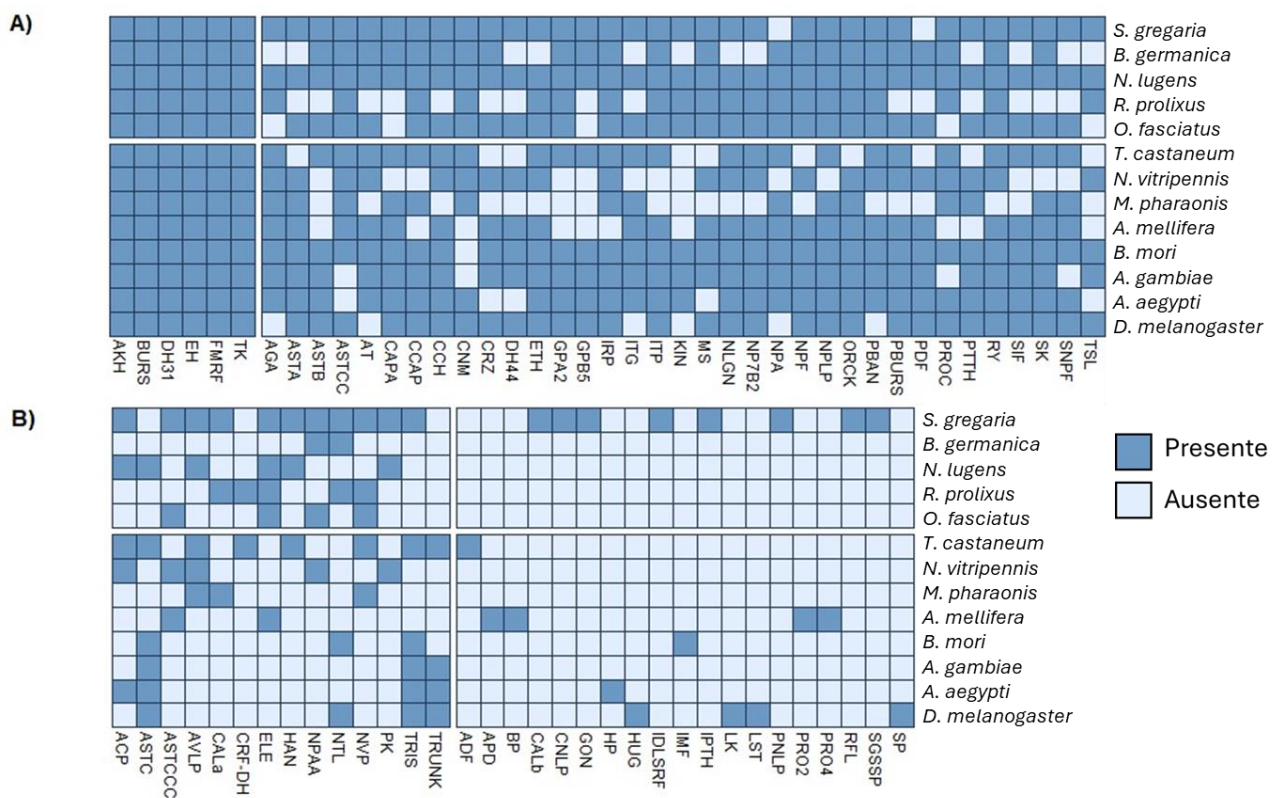


Figura 14. Neuropéptidos presentes por especie a nivel embrionario. **A)** Presentes en 7 especies o más, hallados en todas las especies de estudio (división vertical hacia la izquierda). **B)** Neuropéptidos presentes en 6 especies o menos, contando aquellos presentes en una única especie (división vertical hacia la derecha). La separación blanca horizontal divide los insectos holometábolos (arriba) de los hemimetábolos (abajo), en ambos gráficos.

4.5.2 Los neuropéptidos se expresan en distintas etapas del desarrollo

Paralelamente al desarrollo del análisis de expresión diferencial, se tomaron las matrices de cuantificación iniciales y se transformaron los valores a transcritos por millón (TPM). Estos valores se promediaron dentro de cada réplica y se escalaron para permitir una comparación entre las distintas especies.

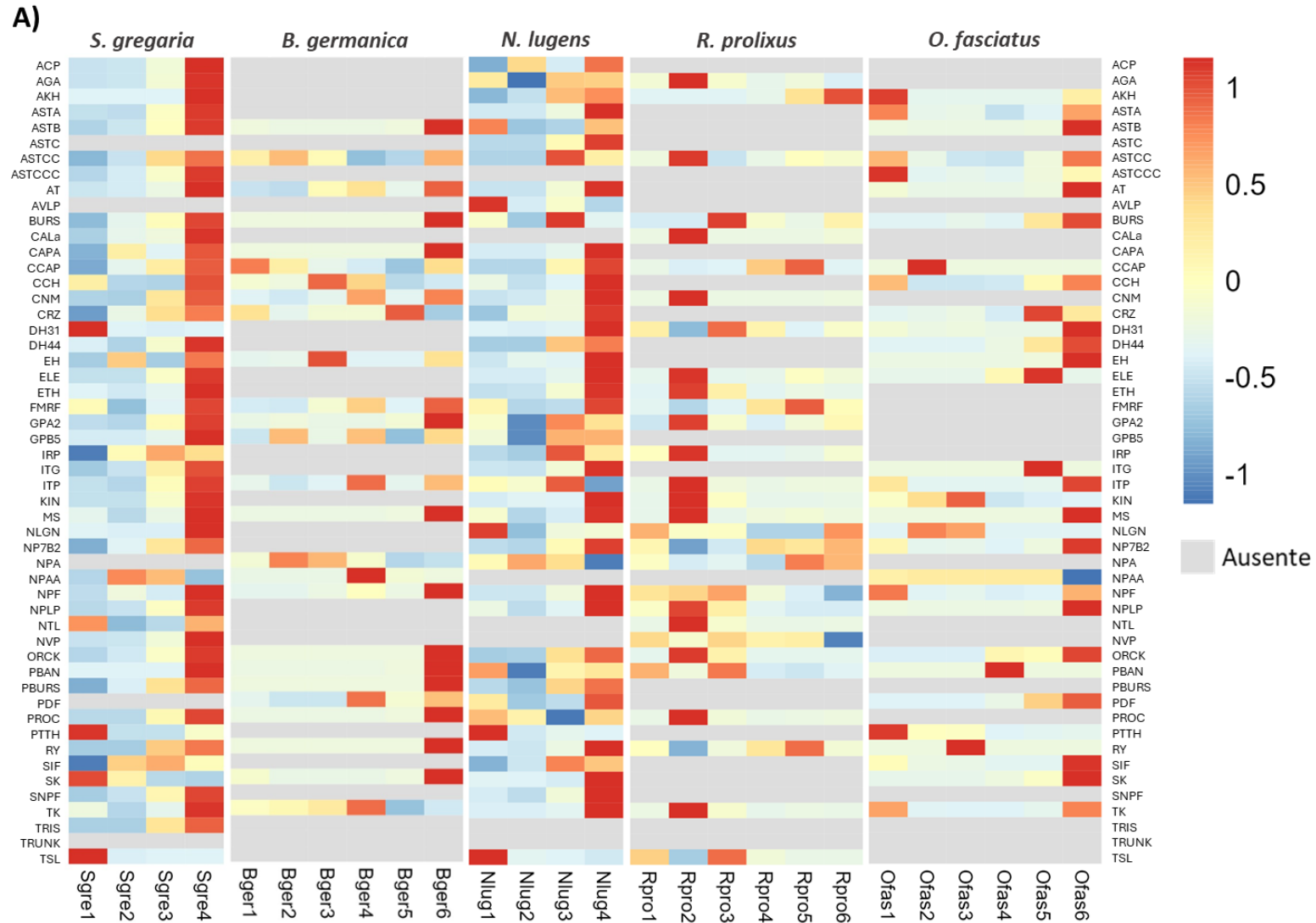


Figura 15. Mapa de calor de genes precursores de neuropéptidos. Los valores de expresión fueron medidos en transcritos por millón (TPM), normalizados por especie (columnas), y luego escalados por neuropéptidos (filas). Los colores indican un bajo (azul), medio (amarillo) y alto (rojo) nivel de expresión. Los neuropéptidos ausentes o no expresados a nivel embrionario se muestran en gris. Se observan en **A)** insectos hemimetábolos y **B)** insectos holometábolos.

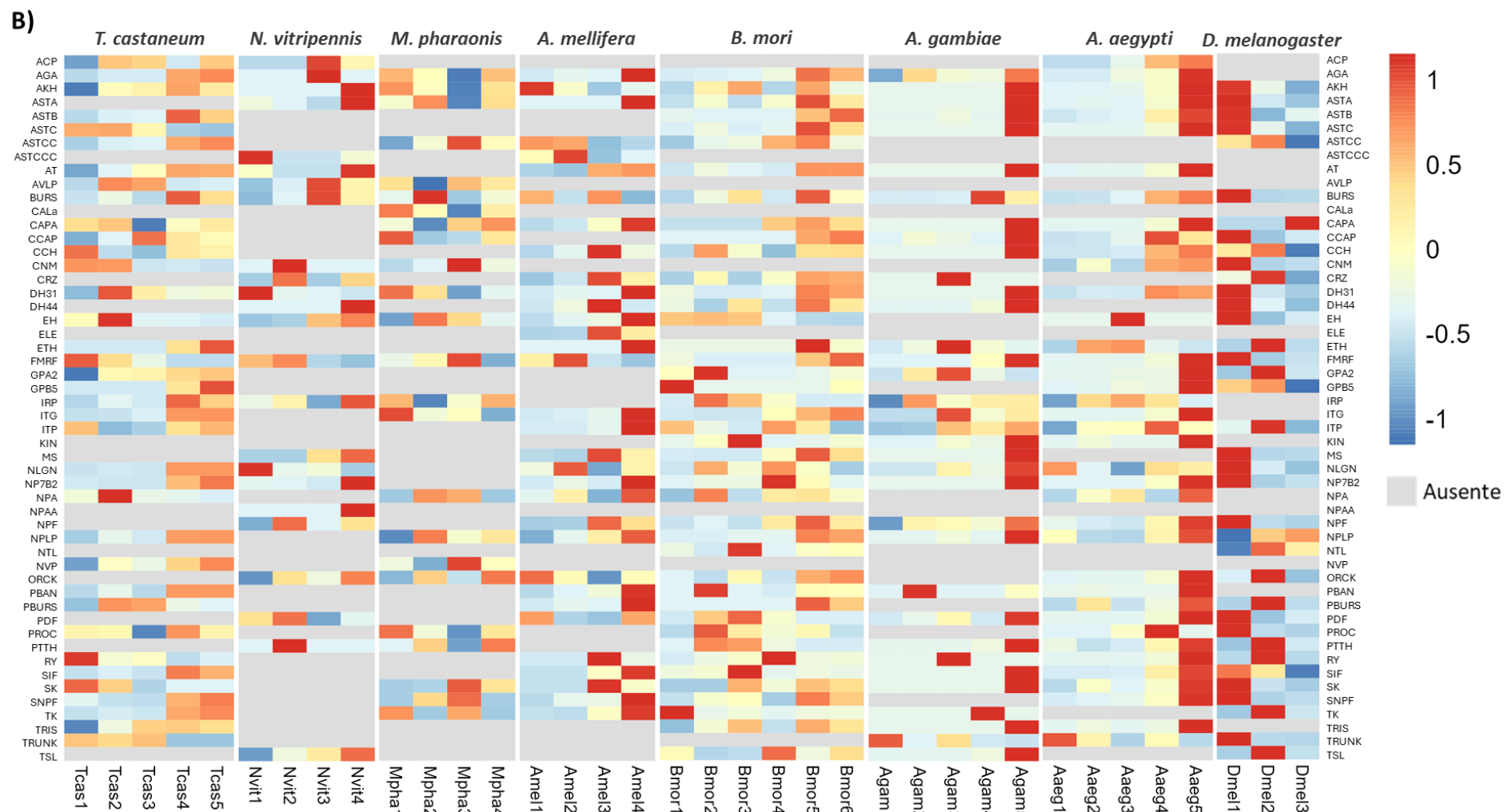


Figura 15. Continuación.

Se continuó el análisis de expresión con aquellos neuropéptidos mayormente presentes en el grupo de especies de estudio, tal como se indicó en la figura 14.A. De todos los genes expresados, es de interés resaltar ciertos patrones de expresión a nivel de embrión temprano/medio, lo que reúne aproximadamente el primer 50% del desarrollo embrionario (Figura 16.A):

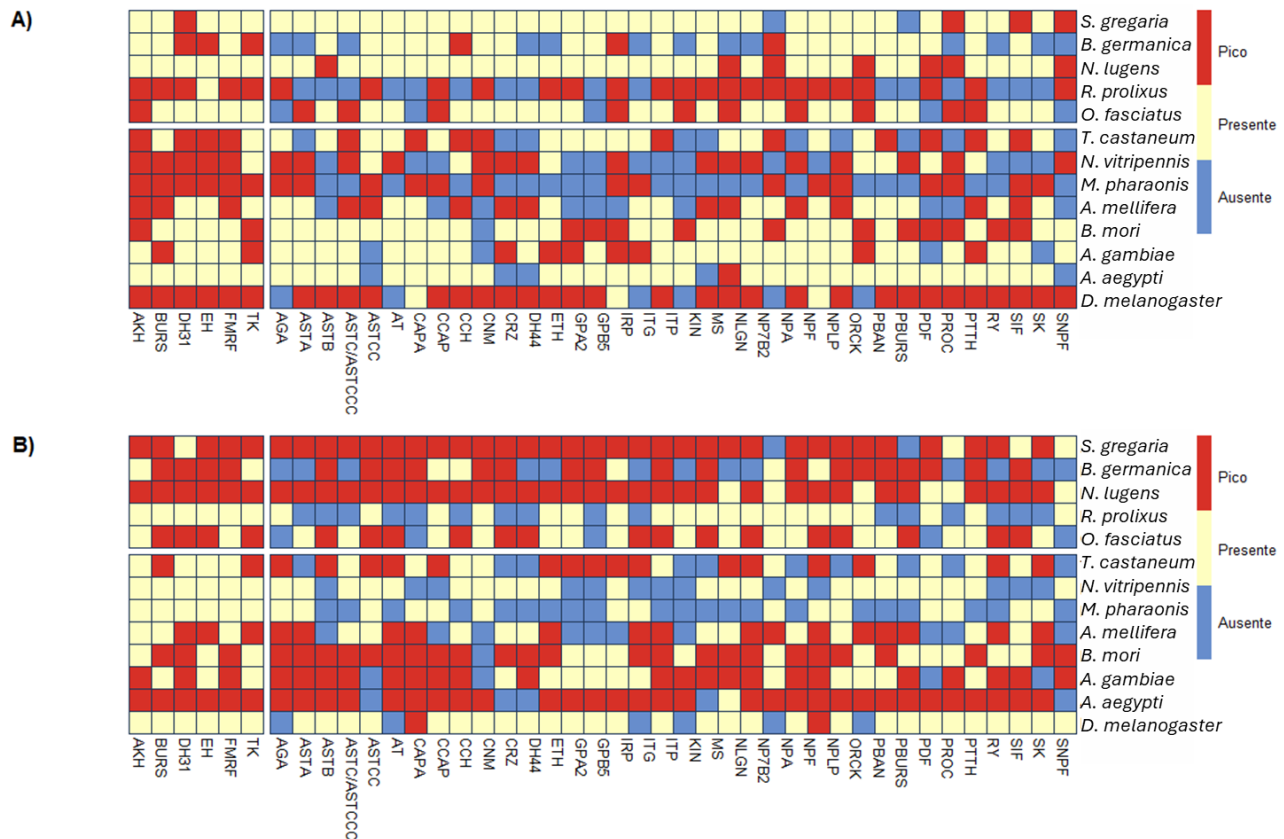


Figura 16. Picos máximos de expresión de genes precursores de neuropéptidos. **A)** Expresión en embrión temprano y medio. **B)** Expresión en embrión tardío. Se muestran aquellos neuropéptidos ausentes en la especie en color azul, aquellos presentes en amarillo, y los que presentan un pico de expresión en su respectiva etapa en rojo (un cuadrado rojo en ambos gráficos indica que el nivel se mantiene en ambas etapas). La separación blanca horizontal divide los insectos holometábolos (arriba) de los hemimetábolos (abajo), mientras que la separación vertical divide aquellos neuropéptidos presentes en todas las especies de aquellos que no lo están.

- AKH, BURS, FMRF y ORCK en todos los himenópteros.
 - o AKH: expresión nula en *N. vitripennis*, hasta el final de la etapa de embrión temprano donde se observa una mínima expresión (Nvit4). Picos de expresión al inicio del desarrollo en *M. pharaonis* y *A. mellifera*, con variaciones a lo largo de las etapas agrupadas mencionadas (figura 17).

HORMONA ADIPOQUINÉTICA (AKH)

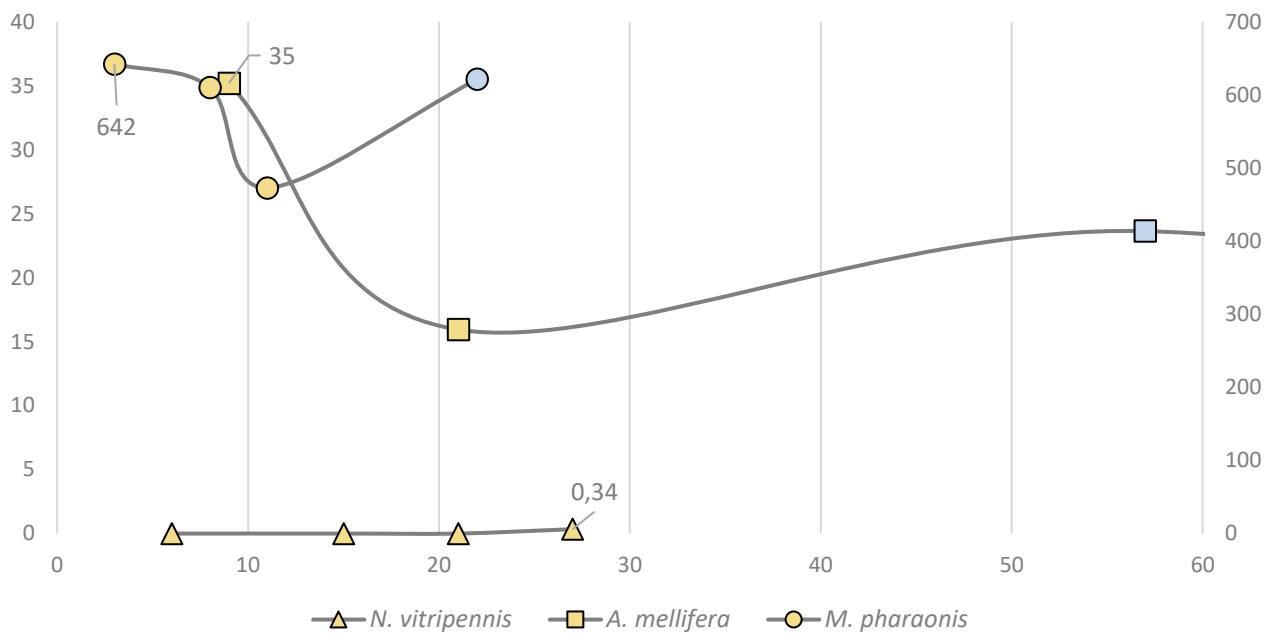


Figura 17. Niveles de expresión temporal de AKH (en TPM), en función del porcentaje de desarrollo. El eje izquierdo corresponde a las especies *N. vitripennis* y *A. mellifera*, mientras que el eje derecho corresponde a los valores en *M. pharaonis*. Los colores de los símbolos que diferencian cada especie marcan a su vez el estadio de desarrollo embrionario, tal como fue clasificado en la figura 7 (gris: huevo no fertilizado; amarillo: embrión temprano; azul: embrión medio; rojo: embrión tardío). Adicionalmente se indican los picos máximos de expresión en números grises.

- BURS: mínimo nivel de expresión en *N. vitripennis*. Expresión a lo largo del desarrollo en todas las especies. Pico de expresión en embrión temprano en *M. pharaonis*, y en embrión medio en *A. mellifera*, con fluctuaciones a lo largo de las etapas (figura 18).

BURSICONA ALFA (BURS)

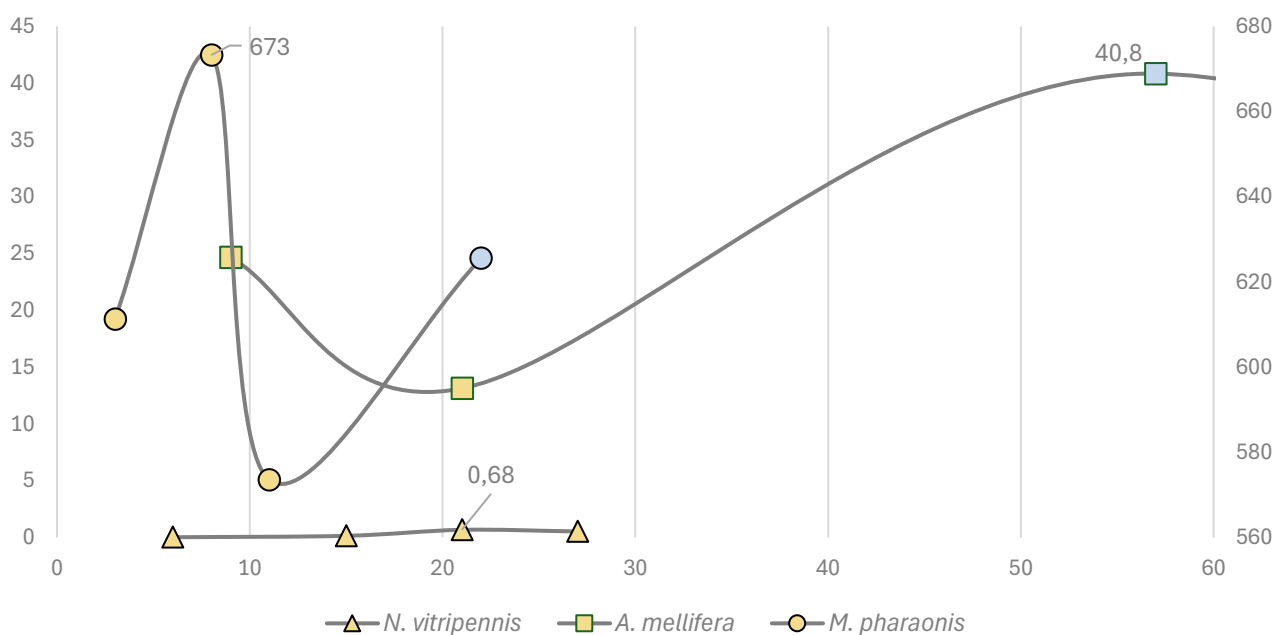


Figura 18. Niveles de expresión temporal de BURS (en TPM), en función del porcentaje de desarrollo. El eje izquierdo corresponde a las especies *N. vitripennis* y *A. mellifera*, mientras que el eje derecho corresponde a los valores en *M. pharaonis*. Los colores de los símbolos que diferencian cada especie marcan a su vez el estadio de desarrollo embrionario, tal como fue clasificado en la figura 7 (gris: huevo no fertilizado; amarillo: embrión temprano; azul: embrión medio; rojo: embrión tardío). Adicionalmente se indican los picos máximos de expresión en números grises.

- FMRF: expresión a lo largo de toda la formación del embrión en cada especie, las tres con picos de expresión en embrión temprano (figura 19).

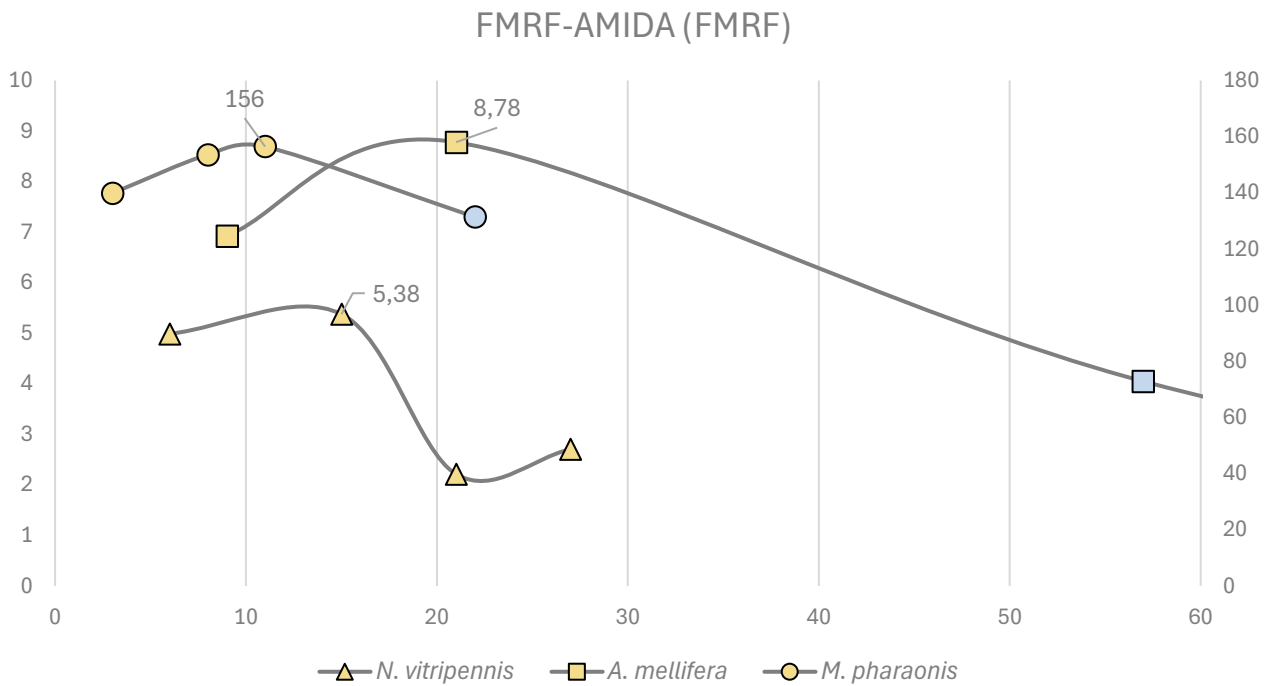


Figura 19. Niveles de expresión temporal de FMRF (en TPM), en función del porcentaje de desarrollo. El eje izquierdo corresponde a las especies *N. vitripennis* y *A. mellifera*, mientras que el eje derecho corresponde a los valores en *M. pharaonis*. Los colores de los símbolos que diferencian cada especie marcan a su vez el estadio de desarrollo embrionario, tal como fue clasificado en la figura 7 (gris: huevo no fertilizado; amarillo: embrión temprano; azul: embrión medio; rojo: embrión tardío). Adicionalmente se indican los picos máximos de expresión en números grises.

- ORCK: expresión mínima en *N. vitripennis* pero constante en todas las especies. Pico de expresión en embrión temprano en *A. mellifera*, y dos picos de expresión en embrión temprano y medio en *M. pharaonis*, siendo este último ligeramente mayor (figura 20).

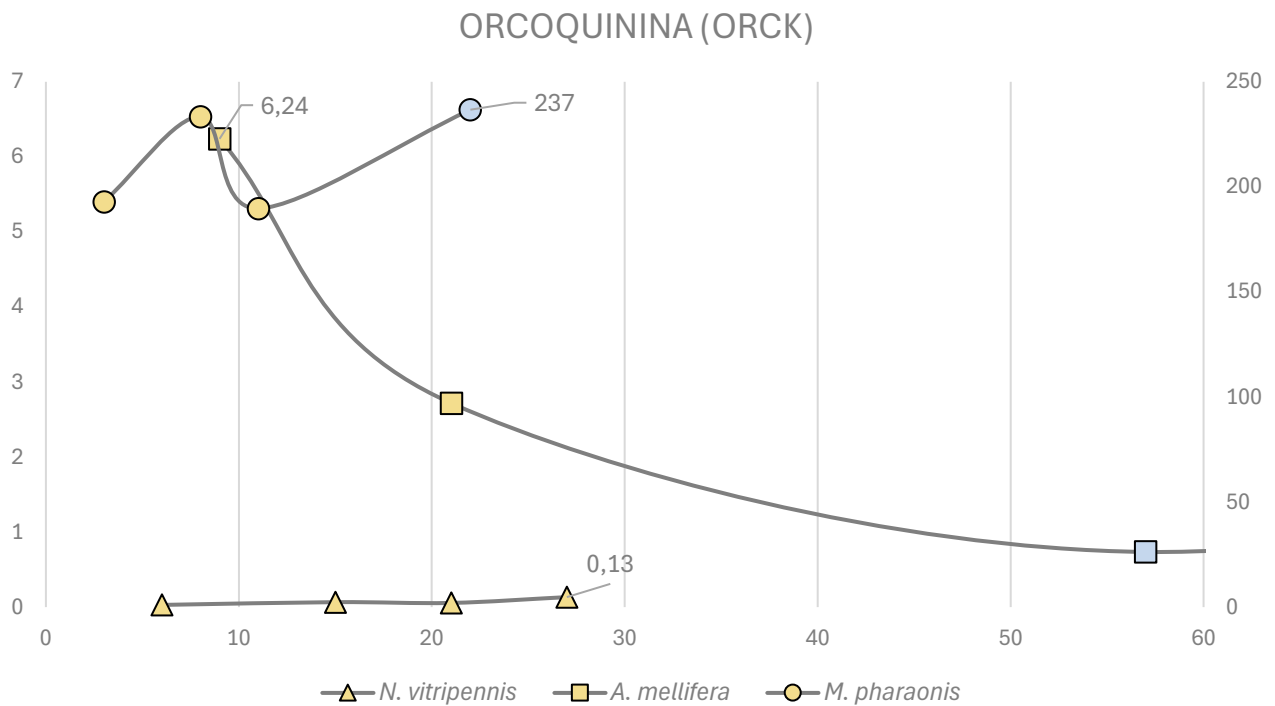


Figura 20. Niveles de expresión temporal de ORCK (en TPM), en función del porcentaje de desarrollo. El eje izquierdo corresponde a las especies *N. vitripennis* y *A. mellifera*, mientras que el eje derecho corresponde a los valores en *M. pharaonis*. Los colores de los símbolos que diferencian cada especie marcan a su vez el estadio de desarrollo embrionario, tal como fue clasificado en la figura 7 (gris: huevo no fertilizado; amarillo: embrión temprano; azul: embrión medio; rojo: embrión tardío). Adicionalmente se indican los picos máximos de expresión en números grises.

- PBAN en todos los hemípteros
 - Expresión nula en *O. fasciatus*, a excepción del inicio del desarrollo medio (Ofas4) con un mínimo valor de expresión. Expresión continua en *R. prolixus* y *N. lugens*, la primera con un patrón fluctuante y la segunda aumentando su expresión hasta llegar a un marcado pico en embrión tardío (figura 21).

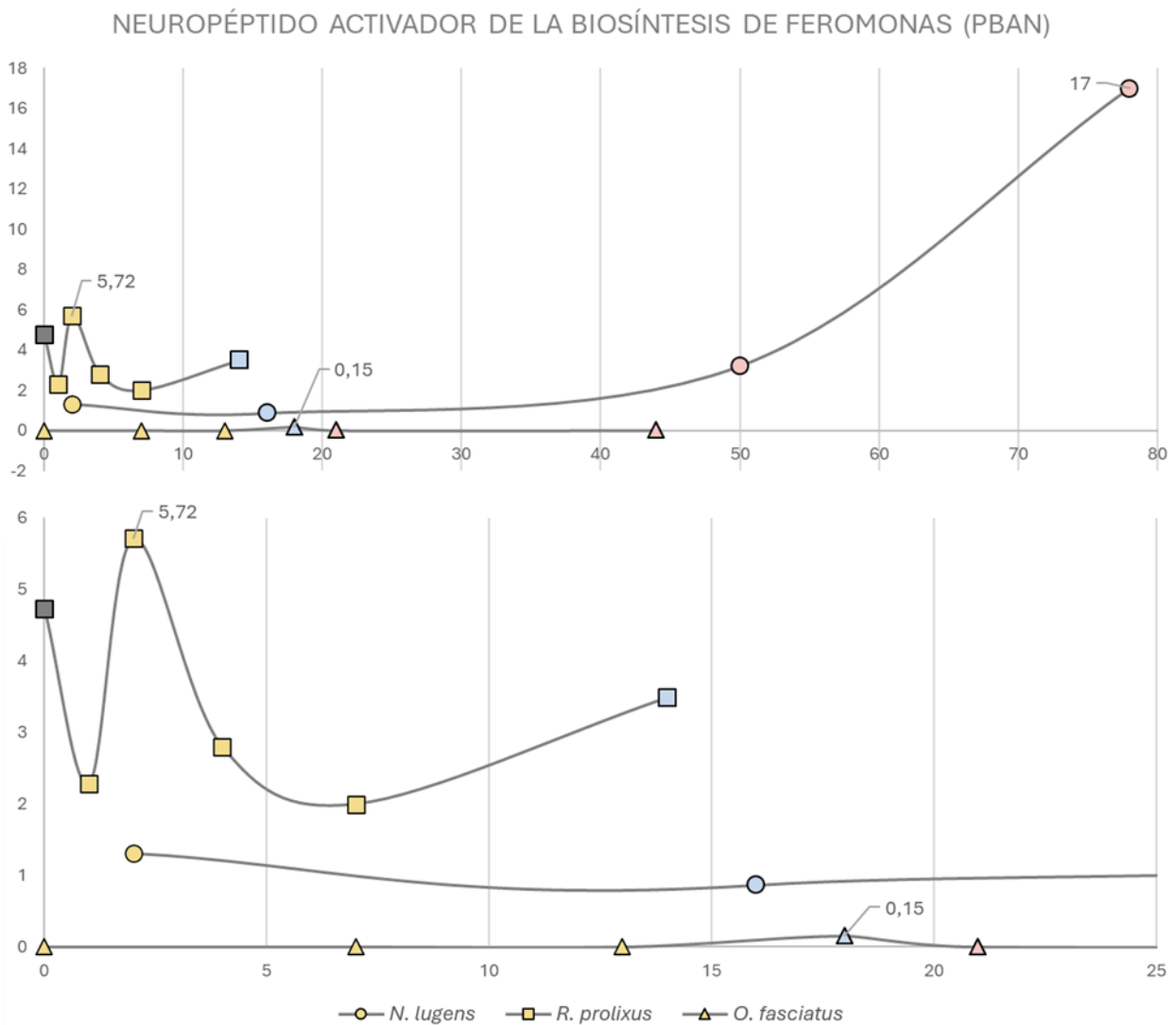


Figura 21. Niveles de expresión temporal de PBAN (en TPM), en función del porcentaje de desarrollo. Se muestran las especies *N. lugens*, *R. prolixus* y *O. fasciatus*. Los colores de los símbolos que diferencian cada especie marcan a su vez el estadio de desarrollo embrionario, tal como fue clasificado en la figura 7 (gris: huevo no fertilizado; amarillo: embrión temprano; azul: embrión medio; rojo: embrión tardío). Adicionalmente se indican los picos máximos de expresión en números grises. **Arriba:** gráfico con los grupos de datos completos. **Abajo:** aumento del gráfico anterior, donde se aprecia mejor la variación de la expresión.

Por otra parte, se ha detectado expresión de transcritos de neuropéptidos en huevos no fertilizados, teniendo picos de expresión de AKH y CCAP en *B. germanica*, y de TRUNK en *A. gambiae* (figura 22). Otros neuropéptidos han mostrado expresión en el mismo estadio, aunque en niveles menores (figura 23).

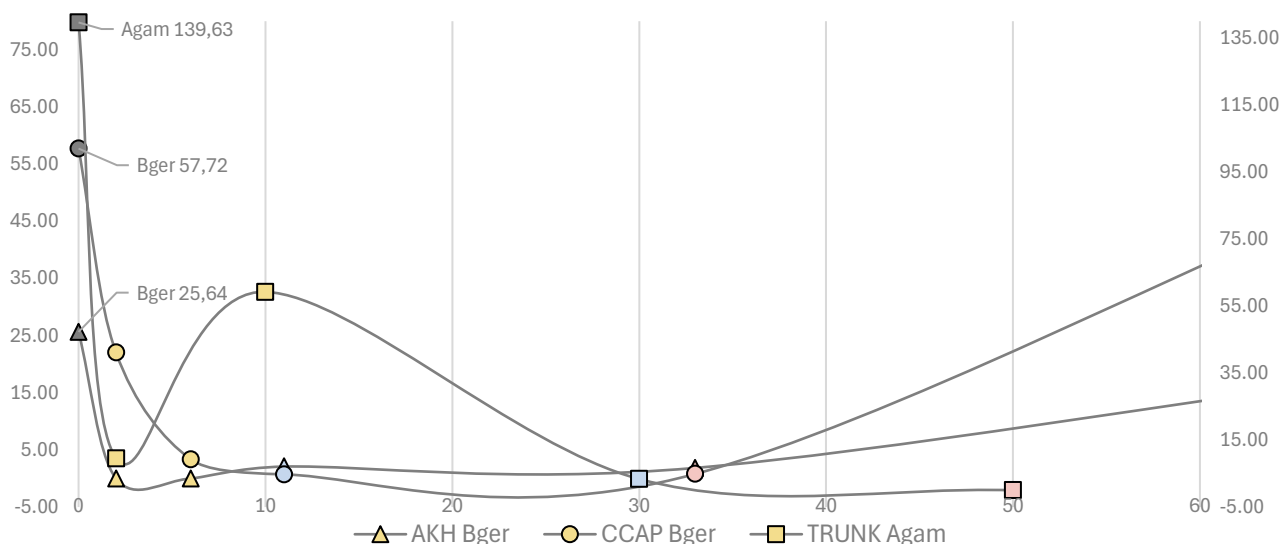


Figura 22. Expresión temporal de transcritos de neuropéptidos (TPM) que presentan picos de expresión en la etapa de huevo no fertilizado. El eje izquierdo corresponde a la especie *B. germanica* (Bger), mientras que el eje derecho corresponde a *A. gambiae* (Agam). Los colores de los marcadores referencian el estadio de desarrollo embrionario, tal como fue clasificado en la figura 7 (gris: huevo no fertilizado; amarillo: embrión temprano; azul: embrión medio; rojo: embrión tardío).

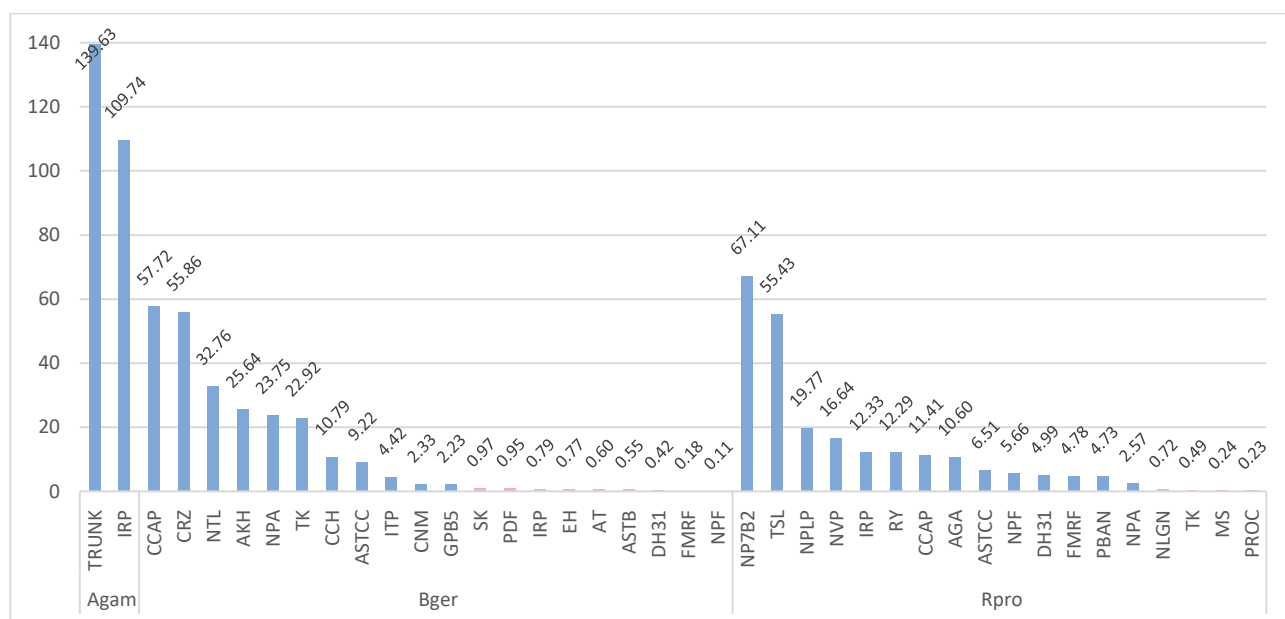


Figura 23. Niveles de expresión de transcritos de neuropéptidos (TPM) en la etapa de huevo no fertilizado. Se muestran las especies *A. gambiae* (Agam), *B. germanica* (Bger) y *R. prolixus* (Rpro) junto con los neuropéptidos hallados en cada una, en orden de expresión decreciente. Los neuropéptidos cuyos valores de expresión resultaron menores a 1 TPM se marcan en color rosa.

Finalmente, se han podido reconocer ciertos genes precursores que son expresados una vez finalizado el desarrollo embrionario. Esto es debido a que no muestran expresión en las series temporales de transcriptomas analizados, pero fueron identificados en los respectivos genomas de cada especie:

- *O. fasciatus*: SNPF, NPA, GPA2, ETH, CNM, FMRF, PBURS
- *N. vitripennis*: RY, PROC, PBURS, PBAN, ETH, CCH, ASTCC, TK
- *A. gambiae*: PBURS, ORCK, NPA, GPB5, EH
- *R. prolixus*: EH
- *D. melanogaster*: IRP
- *B. germanica*: NPLP

5 Discusión

El análisis de datos transcriptómicos es utilizado en un amplio rango de áreas para observar el efecto génico de la actividad metabólica. En este contexto, los datos originados mediante RNA-Seq son una opción popular debido a una baja proporción de falsos positivos y una alta reproducibilidad, frente a otras técnicas como el uso de *microarrays* (Lefol *et al.*, 2023). A lo largo de los años, la reducción de costos y optimización de técnicas han provocado un surgimiento de diseños experimentales que integraron la dimensión del tiempo como variable, obteniendo series temporales continuas de transcriptomas como metodología de comparación de expresión (Lefol *et al.*, 2023).

Asimismo, la disponibilidad de dichos datos en repositorios bioinformáticos y su acceso facilitado otorgan integridad, transparencia, y permiten su reutilización ya sea en relación a su propósito original o para responder una pregunta biológica distinta (Curty *et al.*, 2017; Custers & Uršič, 2016). Los resultados obtenidos permiten asistir al estudio de la expresión de neuropéptidos durante el desarrollo embrionario de insectos.

En primera instancia, se identificaron aquellos neuropéptidos presentes en cada especie mediante una búsqueda automatizada por ortología con Proteinortho. Esto permitió homologar los neuropéptidos en una nomenclatura fácilmente comparable de manera interespecífica. A su vez, se logró establecer un listado de neuropéptidos más completo y robusto que los actualmente presentes en bases de datos especializadas.

En segundo lugar, se realizó una revisión bibliográfica de las etapas de desarrollo embrionario de cada especie, para asociar los datos de RNA-Seq recolectados a su momento de desarrollo correspondiente. El proceso de desarrollo del embrión se dividió en tres secciones (embrión temprano, medio, tardío) sumado a una etapa previa materno-ovárica (huevo no fertilizado) para un análisis más claro (figura 7). La clasificación se efectuó tomando los eventos canónicos del desarrollo: formación de blástula, gastrulación y desarrollo secundario, que a pesar de las diferencias encontradas entre los diferentes insectos, dan lugar a un plan corporal básico compartido (Tautz *et al.*, 1994).

La proporción del desarrollo total que ocupa cada sección en general varía ligeramente a lo largo de las especies. El porcentaje que ocupa la etapa de desarrollo temprano culmina aproximadamente entre el 15% y el 25%, a excepción de las especies *S. gregaria* y *A. mellifera* que exceden este valor. La etapa de embrión medio logra extenderse hasta pasando el 50% en algunos casos. Si bien en este trabajo se toman todos los datos bajo una misma sección de desarrollo como comparables, analizando a un nivel más resolutivo esto no es exactamente así y podría haber diferencias a destacar. Aun así, en aquellas especies donde contamos con varios *sets* de datos tempranos los gráficos de escalado multidimensional realizados (ver sección 4.4) no mostraron separación apreciable como sí ocurrió con el resto de los grupos. Por este motivo, los datos se agruparon interpretándolos como réplicas, lo que a su vez permitió una comparación más directa con otras especies cuyo número de *sets* de datos de expresión temprana era más reducido.

Como se mencionó al inicio, los neuropéptidos son pequeñas moléculas de señalización que actúan como neurotransmisores, neuromoduladores o neurohormonas. La liberación de los mismos junto con estas funciones están evolutivamente ligadas a las propiedades y funcionamiento ancestral de las células neuronales (Elphick *et al.*, 2018). Para ello, versiones inmaduras de los neuropéptidos son procesadas y liberadas a través de vesículas secretoras para ejercer su función.

Por otra parte, en la etapa temprana del desarrollo embrionario el sistema nervioso aún no se encuentra formado. Por ello, los resultados de expresión temprana de transcriptos de neuropéptidos presentados permiten pensar que dichos neuropéptidos podrían estar expresándose

para cumplir una función diferente a las definidas en etapas posteriores al desarrollo embrionario (larva, pupa, adulto)(Baricalla, 2020).

Los transcritos de neuropéptidos encontrados en ovocitos (previo a la formación de la célula huevo) en las especies *A. gambiae*, *B. germanica* y *R. prolixus* apoyan esta hipótesis presentada.

Observando de forma general la expresión de neuropéptidos en todas las especies (figura 15), se puede ver una tendencia de valores máximos de expresión llegando a las etapas más tardías, donde el sistema nervioso está formado y el embrión está próximo a completar su desarrollo. A su vez, se puede notar que los patrones de expresión transcriptómica en su mayoría no son específicos de los grupos taxonómicos analizados. Se han encontrado pocos neuropéptidos compartidos por todas las especies de himenópteros y hemípteros analizadas, con picos de expresión en etapas similares del desarrollo embrionario.

Existe evidencia previa de expresión transcriptómica embrionaria de neuropéptidos de manera experimental. Del Valle (2022) ha reportado la expresión de ASTA (Alatostatina A), ASTCCC (Alatostatina CCC), CCAP (Péptido cardioactivo crustáceo), NPA (Neuroparsina) y PTTH (Hormona protoracicotrópica) en *O. fasciatus*, de manera continua a lo largo del desarrollo embrionario. Estos resultados se condicen con lo hallado mediante el abordaje bioinformático.

Los resultados muestran la necesidad de realizar estudios adicionales, tanto *in-silico* como experimentales, que permitan esclarecer el motivo por el cual se encuentra expresión transcriptómica embrionaria de neuropéptidos.

6 Conclusión

A través del análisis de datos transcriptómicos durante el desarrollo embrionario de las diferentes especies de insectos estudiadas en el presente trabajo, se puede concluir que:

- Los niveles de expresión de ARN para la mayoría de los neuropéptidos estudiados son bajos.
- La expresión transcriptómica resulta variable en el tiempo durante el desarrollo embrionario.
- Tanto a nivel de especies como de genes de neuropéptidos, los incrementos en niveles de expresión generalmente tienden a darse hacia la etapa tardía del desarrollo embrionario. Esto podría relacionarse con la especialización de células neuronales durante la aparición del sistema nervioso.
- La expresión de un mismo neuropéptido presenta variabilidad entre las distintas especies de estudio, sin poder describirse patrones válidos para grupos enteros (holometábolos/hemimetábolos; órdenes taxonómicos estudiados).
- Existen transcriptos de neuropéptidos que presentan picos de expresión en un mismo momento del desarrollo embrionario. Esto ocurre en todas las especies de himenópteros (AKH, BURS, FMRF, ORCK) y hemípteros (PBAN).
- Los datos previos a la formación del ovocito conseguidos de *A. gambiae*, *R. prolixus* y *B. germanica* permitieron identificar un grupo de neuropéptidos expresados en dicha etapa del desarrollo.

7 Perspectivas y direcciones futuras

Esta tesis es el primer trabajo en estudiar un conjunto de neuropéptidos expresados durante el desarrollo embrionario de una selección de especies representativas de insectos.

A partir de este punto se originan nuevas direcciones que permitirán complementar los resultados aquí presentados, tomando en cuenta no solo un análisis *in-silico* sino también abordajes experimentales.

Avances bioinformáticos:

- Identificar la presencia de los distintos receptores asociados a los neuropéptidos considerados en presente trabajo.
- Explorar las funciones putativas de los neuropéptidos, pensando que podrían actuar como reservorios no procesados o, alternativamente, cumplir alguna función no reportada hasta la fecha (Baricalla, 2020).
- Extender el análisis bioinformático a nuevas especies revisando constantemente la publicación de nuevos datos en repositorios públicos. Integrar especies de órdenes no incluidos en este trabajo, además de complementar aquellos órdenes con un solo representante por falta de datos disponibles.
- Analizar datos transcriptómicos de polaridad anteroposterior para estudiar la posible existencia de expresión diferencial en porciones específicas del embrión.
- El análisis de expresión de enzimas convertasas a lo largo del desarrollo embrionario no ha presentado patrones coincidentes con lo previamente mostrado en este trabajo para la expresión de los grupos de proteínas, con lo cual no existen resultados concluyentes y se necesitan más investigaciones al respecto.

Avances experimentales:

- Aplicar técnicas de ARN de interferencia, inmunohistoquímica o fluorescencia *in-situ* para detectar experimentalmente los transcriptos mensajeros de neuropéptidos en el embrión, incluyendo diferencias anteroposteriores y/o dorsoventrales.

8 Bibliografía

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., ... Venter, J. C. (2000). The genome sequence of *Drosophila melanogaster*. *Science (New York, N.Y.)*, 287(5461), 2185–2195. <https://doi.org/10.1126/science.287.5461.2185>
- Akbari, O. S., Antoshechkin, I., Amrhein, H., Williams, B., Diloreto, R., Sandler, J., & Hay, B. A. (2013). The developmental transcriptome of the mosquito *Aedes aegypti*, an invasive species and major arbovirus vector. *G3 (Bethesda, Md.)*, 3(9), 1493–1509. <https://doi.org/10.1534/g3.113.006742>
- 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)
- Altstein, M., & Nässel, D. R. (2010). Neuropeptide signaling in insects. *Advances in Experimental Medicine and Biology*, 692, 155–165. https://doi.org/10.1007/978-1-4419-6902-6_8
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. 2010 (p. <http://www.bioinformatics.babraham.ac.uk/projects/>).
- 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>
- Baricalla, A. (2020). *Genómica, transcriptómica y proteómica de neuropéptidos de *Oncopeltus fasciatus* (Hemiptera : Lygaeidae)*. Universidad Nacional del Noroeste de la Provincia de Buenos Aires.
- Bull, A. L. (1982). Stages of Living Embryos in the Jewel Wasp (Hymenoptera : Pteromalidae). *International Journal of Insect Morphology and Embryology*, 11(1), 1–23. <http://www.sciencedirect.com/science/article/pii/0020732282900344>
- Chatterjee, A., Ahn, A., Rodger, E. J., Stockwell, P. A., & Eccles, M. R. (2018). A guide for designing and analyzing RNA-seq data. En *Methods in Molecular Biology* (Vol. 1783). https://doi.org/10.1007/978-1-4939-7834-2_3
- Chipman, A. D. (2017). *Oncopeltus fasciatus* as an evo-devo research organism. *genesis*, 55(5), e23020. <https://doi.org/10.1002/dvg.23020>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 1–19. <https://doi.org/10.1186/s13059-016-0881-8>
- Curry, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists' data reuse. *PLoS ONE*, 12(12), 1–22. <https://doi.org/10.1371/journal.pone.0189288>
- Custers, B., & Uršič, H. (2016). Big data and data reuse: A taxonomy of data reuse for balancing big data benefits and personal data protection. *International Data Privacy Law*, 6(1), 4–15. <https://doi.org/10.1093/idpl/ipv028>
- Dalla Benetta, E., Antoshechkin, I., Yang, T., Nguyen, H. Q. M., Ferree, P. M., & Akbari, O. S. (2020). Genome elimination mediated by gene expression from a selfish chromosome. *Science Advances*, 6(14), eaaz9808. <https://doi.org/10.1126/sciadv.aaz9808>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Del Valle, D. (2022). *Estudio de la expresión de genes codificantes de neuropéptidos y sus*

receptores durante el desarrollo embrionario de Oncopeltus fasciatus. Universidad Nacional del Noroeste de la Provincia de Buenos Aires.

- Diniz, W. J. S., & Canduri, F. (2017). Bioinformatics: An overview and its applications. *Genetics and Molecular Research*, 16(1). <https://doi.org/10.4238/gmr16019645>
- Elphick, M. R., Mirabeau, O., & Larhammar, D. (2018). Evolution of neuropeptide signalling systems. *Journal of Experimental Biology*, 221(3). <https://doi.org/10.1242/jeb.151092>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, 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>
- Fan, X. Bin, Pang, R., Li, W. X., Ojha, A., Li, D., & Zhang, W. Q. (2020). An Overview of Embryogenesis: External Morphology and Transcriptome Profiling in the Hemipteran Insect *Nilaparvata lugens*. *Frontiers in Physiology*, 11(February), 1–14. <https://doi.org/10.3389/fphys.2020.00106>
- Farnesi, L. C., Martins, A. J., Valle, D., & Rezende, G. L. (2009). Embryonic development of *Aedes aegypti* (Diptera: Culicidae): Influence of different constant temperatures. *Memorias do Instituto Oswaldo Cruz*, 104(1), 124–126. <https://doi.org/10.1590/S0074-02762009000100020>
- Fleig, R., & Sander, K. (1986). Embryogenesis of the honeybee *Apis mellifera* L. (hymenoptera : apidae): An sem study. *International Journal of Insect Morphology and Embryology*, 15(5–6), 449–462. [https://doi.org/10.1016/0020-7322\(86\)90037-1](https://doi.org/10.1016/0020-7322(86)90037-1)
- Gao, Q., Xiong, Z., Larsen, R. S., Zhou, L., Zhao, J., Ding, G., Zhao, R., Liu, C., Ran, H., & Zhang, G. (2020). High-quality chromosome-level genome assembly and full-length transcriptome analysis of the pharaoh ant *Monomorium pharaonis*. *GigaScience*, 9(12). <https://doi.org/10.1093/gigascience/giaa143>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. En J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press. <https://doi.org/10.1385/1-59259-890-0:571>
- Gilbert, S. F. (2010). Developmental Biology. En *Nature* (Vol. 218, Número 5137). <https://doi.org/10.1038/218207a0>
- Goltsev, Y., Rezende, G. L., Vranizan, K., Lanzaro, G., Valle, D., & Levine, M. (2009). Developmental and evolutionary basis for drought tolerance of the *Anopheles gambiae* embryo. *Developmental Biology*, 330(2), 462–470. <https://doi.org/10.1016/j.ydbio.2009.02.038>
- 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>
- Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., Artieri, C. G., Van Baren, M. J., Boley, N., Booth, B. W., Brown, J. B., Cherbas, L., Davis, C. A., Dobin, A., Li, R., Lin, W., Malone, J. H., Mattiuzzo, N. R., Miller, D., ... Celniker, S. E. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature*, 471(7339), 473–479. <https://doi.org/10.1038/nature09715>
- Gurska, D., Vargas Jentzsch, I. M., & Panfilio, K. A. (2020). Unexpected mutual regulation underlies paralogue functional diversification and promotes epithelial tissue maturation in *Tribolium*. *Communications Biology*, 3(1). <https://doi.org/10.1038/s42003-020-01250-3>
- Han, Y., Gao, S., Muegge, K., Zhang, W., & Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinformatics and Biology Insights*, 9, 29–46.

<https://doi.org/10.4137/BBI.S28991>

- Harrison, M. C., Jongepier, E., Robertson, H. M., Arning, N., Bitard-Feildel, T., Chao, H., Childers, C. P., Dinh, H., Doddapaneni, H., Dugan, S., Gowin, J., Greiner, C., Han, Y., Hu, H., Hughes, D. S. T., Huylmans, A.-K., Kemena, C., Kremer, L. P. M., Lee, S. L., ... Bornberg-Bauer, E. (2018). Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution*, 2(3), 557–566. <https://doi.org/10.1038/s41559-017-0459-1>
- Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M. C., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., ... Hoffman, S. L. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science (New York, N.Y.)*, 298(5591), 129–149. <https://doi.org/10.1126/science.1076181>
- Ishizaki, H. (2004). Molecular characterization of the brain secretory peptides, prothoracicotropic hormone (PTTH) and bombyxin, of the silkworm *Bombyx mori*. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, 80(5), 195–203. <https://doi.org/10.2183/pjab.80.195>
- Ishizaki, H., & Suzuki, A. (1988). An insect brain peptide as a member of insulin family. *Hormone and Metabolic Research*, 20(7), 426–429. <https://doi.org/10.1055/s-2007-1010852>
- Kaiser, M. L., Duncan, F. D., & Brooke, B. D. (2014). Embryonic development and rates of metabolic activity in early and late hatching eggs of the major malaria vector *Anopheles gambiae*. *PLoS ONE*, 9(12), 1–15. <https://doi.org/10.1371/journal.pone.0114381>
- Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J. R., & O'Sullivan, C. (2021). The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Research*, 50(D1), D387–D390. <https://doi.org/10.1093/nar/gkab1053>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360.
- Klemm, P., Stadler, P. F., & Lechner, M. (2023). Proteinortho6: pseudo-reciprocal best alignment heuristic for graph-based detection of (co-)orthologs. *Frontiers in Bioinformatics*, 3(December). <https://doi.org/10.3389/fbinf.2023.1322477>
- Klingler, M., & Bucher, G. (2022). The red flour beetle *T. castaneum*: elaborate genetic toolkit and unbiased large scale RNAi screening to study insect biology and evolution. *EvoDevo*, 13(1), 1–12. <https://doi.org/10.1186/s13227-022-00201-9>
- Konopová, B., Buchberger, E., & Crisp, A. (2020). Transcriptome of pleuropodia from locust embryos supports that these organs produce enzymes enabling the larva to hatch. *Frontiers in Zoology*, 17(1), 1–22. <https://doi.org/10.1186/s12983-019-0349-2>
- Kuntz, S. G., & Eisen, M. B. (2014). *Drosophila* Embryogenesis Scales Uniformly across Temperature in Developmentally Diverse Species. *PLoS Genetics*, 10(4). <https://doi.org/10.1371/journal.pgen.1004293>
- Lefol, Y., Korfage, T., Mjelle, R., Prebensen, C., Lüders, T., Müller, B., Krokan, H., Sarno, A., Alsøe, L., Berdal, J. E., Sætrom, P., Nilsen, H., & Domanska, D. (2023). TiSA: TimeSeriesAnalysis - A pipeline for the analysis of longitudinal transcriptomics data. *NAR Genomics and Bioinformatics*, 5(1), 1–11. <https://doi.org/10.1093/nargab/lqad020>
- 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>
- Lu, F., Wei, Z., Luo, Y., Guo, H., Zhang, G., Xia, Q., & Wang, Y. (2020). SilkDB 3.0: visualizing and exploring multiple levels of data for silkworm. *Nucleic Acids Research*, 48(D1), D749–D755. <https://doi.org/10.1093/nar/gkz919>
- Ma, W., Xu, L., Hua, H., Chen, M., Guo, M., He, K., Zhao, J., & Li, F. (2021). Chromosomal-level

- genomes of three rice planthoppers provide new insights into sex chromosome evolution. *Molecular Ecology Resources*, 21(1), 226–237. <https://doi.org/10.1111/1755-0998.13242>
- Matthews, B. J., Dudchenko, O., Kingan, S. B., Koren, S., Antoshechkin, I., Crawford, J. E., Glassford, W. J., Herre, M., Redmond, S. N., Rose, N. H., Weedall, G. D., Wu, Y., Batra, S. S., Brito-Sierra, C. A., Buckingham, S. D., Campbell, C. L., Chan, S., Cox, E., Evans, B. R., ... Voss hall, L. B. (2018). Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature*, 563(7732), 501–507. <https://doi.org/10.1038/s41586-018-0692-z>
- McDermaid, A., Monier, B., Zhao, J., Liu, B., & Ma, Q. (2019). Interpretation of differential gene expression results of RNA-seq data: Review and integration. *Briefings in Bioinformatics*, 20(6), 2044–2054. <https://doi.org/10.1093/bib/bby067>
- Mellanby, H. (1934). The Early Embryonic Development of *Rhodnius prolixus* (Hemiptera , Heteroptera). *Department of Zoology, University College London*.
- 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>
- Microsoft Corporation. (2019). *Microsoft Excel*. <https://office.microsoft.com/excel>
- Morita, A., Niimi, T., & Yamashita, O. (2003). Physiological differentiation of DH-PBAN-producing neurosecretory cells in the silkworm embryo. *Journal of Insect Physiology*, 49(12), 1093–1102. <https://doi.org/10.1016/j.jinsphys.2003.08.009>
- Mundim-Pombo, A. P. M., Carvalho, H. J. C. de, Rodrigues Ribeiro, R., León, M., Maria, D. A., & Miglino, M. A. (2021). *Aedes aegypti*: egg morphology and embryonic development. *Parasites and Vectors*, 14(1), 1–12. <https://doi.org/10.1186/s13071-021-05024-6>
- Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Ishizaki, H., Mizoguchi, A., Fujiwara, Y., & Suzuki, A. (1984). Amino-terminal amino Acid sequence of the silkworm prothoracicotropic hormone: homology with insulin. *Science (New York, N.Y.)*, 226(4680), 1344–1345. <https://doi.org/10.1126/science.226.4680.1344>
- Nässel, D. R., & Larhammar, D. (2013). Neurosciences - From Molecule to Behavior: a university textbook. *Neurosciences - From Molecule to Behavior: a university textbook*, 213–237. <https://doi.org/10.1007/978-3-642-10769-6>
- 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. (2017). Neuropeptides in the regulation of *Rhodnius prolixus* physiology. *Journal of Insect Physiology*, 97, 77–92. <https://doi.org/10.1016/j.jinsphys.2016.05.003>
- Panfilio, K. A., Vargas Jentsch, I. M., Benoit, J. B., Erezyilmaz, 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. (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., & Rivera-Pomar, R. (2022). Dynamics of maternal gene expression in *Rhodnius prolixus*. *Scientific Reports*, 12(1), 1–17. <https://doi.org/10.1038/s41598-022-09874-7>
- Patra, B. G., Maroufy, V., Soltanalizadeh, B., Deng, N., Zheng, W. J., Roberts, K., & Wu, H. (2020). A content-based literature recommendation system for datasets to improve data reusability – A

- case study on Gene Expression Omnibus (GEO) datasets. *Journal of Biomedical Informatics*, 104, 103399. <https://doi.org/https://doi.org/10.1016/j.jbi.2020.103399>
- Pontieri, L., Rajakumar, A., Matteen Rafiqi, A., Stenbak Larsen, R., Abouheif, E., & Zhang, G. (2020). From egg to adult: a developmental table of the ant *Monomorium pharaonis*. *bioRxiv*, 2020.12.22.423970. <https://doi.org/10.1101/2020.12.22.423970>
- Posit team. (2022). *RStudio: Integrated Development Environment for R*. <http://www.posit.co/>
- Qiu, B., Dai, X., Li, P., Larsen, R. S., Li, R., Price, A. L., Ding, G., Texada, M. J., Zhang, X., Zuo, D., Gao, Q., Jiang, W., Wen, T., Pontieri, L., Guo, C., Rewitz, K., Li, Q., Liu, W., Boomsma, J. J., & Zhang, G. (2022). Canalized gene expression during development mediates caste differentiation in ants. *Nature Ecology and Evolution*, 6(11), 1753–1765. <https://doi.org/10.1038/s41559-022-01884-y>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. <https://www.r-project.org/>
- Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., Gibbs, R., Beeman, R. W., Brown, S. J., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P., Klingler, M., Lorenzen, M., Richards, S., Roth, S., Schröder, R., Tautz, D., Zdobnov, E. M., ... Bucher, G. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 452(7190), 949–955. <https://doi.org/10.1038/nature06784>
- Rivera-Pomar, R., & Jäckle, H. (1996). From gradients to stripes in *Drosophila* embryogenesis: Filling in the gaps. *Trends in Genetics*, 12(11), 478–483. [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.
- Rosenberg, M. I., Lynch, J. A., & Desplan, C. (2009). Heads and tails: Evolution of antero-posterior patterning in insects. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1789(4), 333–342. <https://doi.org/10.1016/j.bbagrm.2008.09.007>
- Scherkenbeck, J., & Zdobinsky, T. (2009). Insect neuropeptides: Structures, chemical modifications and potential for insect control. *Bioorganic and Medicinal Chemistry*, 17(12), 4071–4084. <https://doi.org/10.1016/j.bmc.2008.12.061>
- Schoofs, L., De Loof, A., & Van Hiel, M. B. (2017). Neuropeptides as Regulators of Behavior in Insects. *Annual Review of Entomology*, 62(October), 35–52. <https://doi.org/10.1146/annurev-ento-031616-035500>
- Schröder, R., Beermann, A., Wittkopp, N., & Lutz, R. (2008). From development to biodiversity - *Tribolium castaneum*, an insect model organism for short germband development. *Development Genes and Evolution*, 218(3–4), 119–126. <https://doi.org/10.1007/s00427-008-0214-3>
- Sielemann, K., Hafner, A., & Pucker, B. (2020). The reuse of public datasets in the life sciences: potential risks and rewards. *PeerJ*, 8, e9954.
- Singh, D., Singh, P. K., Chaudhary, S., Mehla, K., & Kumar, S. (2012). Exome sequencing and advances in crop improvement. *Advances in Genetics*, 79, 87–121. <https://doi.org/10.1016/B978-0-12-394395-8.00003-7>
- Smith, W. A., Lamattina, A., & Collins, M. (2014). Insulin signaling pathways in lepidopteran ecdysone secretion. *Frontiers in Physiology*, 5, 19. <https://doi.org/10.3389/fphys.2014.00019>
- Strobl, F., Stelzer, E. H. K., Strobl, F., & Stelzer, E. H. K. (2014). Erratum to Non-invasive long-term fluorescence live imaging of *Tribolium castaneum* embryos (*Development (Cambridge)* 141, (2331-2338)). *Development (Cambridge)*, 141(11), 2361. <https://doi.org/10.1242/dev.112706>
- Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression

- in RNA-seq: A matter of depth. *Genome Research*, 21(12), 2213–2223.
<https://doi.org/10.1101/gr.124321.111>
- Tautz, D., Friedrich, M., & Schröder, R. (1994). Insect embryogenesis - What is ancestral and what is derived? *Development*, 120(SUPPL.), 193–199.
<https://doi.org/10.1242/dev.1994.supplement.193>
- Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, 15(3), 1–26.
<https://doi.org/10.1371/journal.pone.0229003>
- Tihelka, E., Cai, C., Giacomelli, M., Lozano-Fernandez, J., Rota-Stabelli, O., Huang, D., Engel, M. S., Donoghue, P. C. J., & Pisani, D. (2021). The evolution of insect biodiversity. *Current Biology*, 31(19), R1299–R1311. <https://doi.org/10.1016/j.cub.2021.08.057>
- Veenstra, J. A. (2019). Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. *PeerJ*, 2019(6).
<https://doi.org/10.7717/peerj.7144>
- Verlinden, H., Sterck, L., Li, J., Li, Z., Yssel, A., Gansemans, Y., Verdonck, R., Holtof, M., Song, H., Behmer, S. T., Sword, G. A., Matheson, T., Ott, S. R., Deforce, D., Van Nieuwerburgh, F., Van de Peer, Y., & Vanden Broeck, J. (2020). First draft genome assembly of the desert locust, *Schistocerca gregaria*. *F1000Research*, 9, 775.
<https://doi.org/10.12688/f1000research.25148.2>
- Wallberg, A., Bunikis, I., Pettersson, O. V., Mosbech, M.-B., Childers, A. K., Evans, J. D., Mikheyev, A. S., Robertson, H. M., Robinson, G. E., & Webster, M. T. (2019). A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics*, 20(1), 275. <https://doi.org/10.1186/s12864-019-5642-0>
- Wanninger, A. (2015). Evolutionary Developmental Biology of Invertebrates 5: Ecdysozoa III: Hexapoda. En *Evolutionary Developmental Biology of Invertebrates 5: Ecdysozoa III: Hexapoda* (Vol. 5). <https://doi.org/10.1007/978-3-7091-1856-6>
- Wickham, H. (2009). Ggplot2. En *Ggplot2* (Vol. 4, Número 1). <https://doi.org/10.1007/978-0-387-98141-3>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Xu, G. F., Gong, C. C., Lyu, H., Deng, H. M., & Zheng, S. C. (2022). Dynamic transcriptome analysis of *Bombyx mori* embryonic development. *Insect Science*, 29(2), 344–362.
<https://doi.org/10.1111/1744-7917.12934>
- 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>
- Ylla, G., Piulachs, M. D., & Belles, X. (2018). Comparative Transcriptomics in Two Extreme Neopterans Reveals General Trends in the Evolution of Modern Insects. *iScience*, 4, 164–179.
<https://doi.org/10.1016/j.isci.2018.05.017>
- Zaru, R., & Orchard, S. (2023). UniProt Tools: BLAST, Align, Peptide Search, and ID Mapping. *Current Protocols*, 3(3), 1–26. <https://doi.org/10.1002/cpz1.697>
- Zeng, H., Qin, Y., Du, E., Wei, Q., Li, Y., Huang, D., Wang, G., Veenstra, J. A., Li, S., & Li, N. (2021). Genomics- And Peptidomics-Based Discovery of Conserved and Novel Neuropeptides in the American Cockroach. *Journal of Proteome Research*, 20(2), 1217–1228.

<https://doi.org/10.1021/acs.jproteome.0c00596>

Zhao, Q., Xie, Y., Nie, P., Diao, R., Sun, L., Zuo, Z., & Ren, J. (2018). IDEA: a web server for Interactive Differential Expression Analysis with R Packages. *bioRxiv*, 360461.

9 Anexo

Tabla 4. Referencias consultadas para agrupar los datos transcriptómicos a las etapas correspondientes del desarrollo de cada especie. Cabe aclarar que en algunos casos la descripción encontrada en la publicación de origen de los datos tuvo que ser reforzada por otras con mayor detalle.

Especie	Referencias
<i>S. gregaria</i>	(Konopová <i>et al.</i> , 2020)
<i>B. germanica</i>	(Ylla <i>et al.</i> , 2018)
<i>N. lugens</i>	(Fan <i>et al.</i> , 2020)
<i>R. prolixus</i>	(Mellanby, 1934; Pascual & Rivera-Pomar, 2022)
<i>O. fasciatus</i>	(Baricalla, 2020; Chipman, 2017)
<i>T. castaneum</i>	(Gurska <i>et al.</i> , 2020; Klingler & Bucher, 2022; Schröder <i>et al.</i> , 2008; Strobl <i>et al.</i> , 2014)
<i>N. vitripennis</i>	(Bull, 1982)
<i>M. pharaonis</i>	(Pontieri <i>et al.</i> , 2020; Qiu <i>et al.</i> , 2022)
<i>A. mellifera</i>	(Fleig & Sander, 1986)
<i>B. mori</i>	(Morita <i>et al.</i> , 2003; Xu <i>et al.</i> , 2022)
<i>A. gambiae</i>	(Goltsev <i>et al.</i> , 2009; Kaiser <i>et al.</i> , 2014)
<i>A. aegypti</i>	(Akbari <i>et al.</i> , 2013; Farnesi <i>et al.</i> , 2009; Matthews <i>et al.</i> , 2018; Mundim-Pombo <i>et al.</i> , 2021)
<i>D. melanogaster</i>	(Graveley <i>et al.</i> , 2011; Kuntz & Eisen, 2014)

Tabla 5. Listado de identificadores génicos de los genes de precursores de neuropéptidos hallados en cada especie.

NP	<i>S. gregaria</i>	<i>B. germanica</i>	<i>N. lugens</i>	<i>R. prolixus</i>	<i>O. fasciatus</i>	<i>T. castaneum</i>	<i>N. vitripennis</i>	<i>M. pharaonis</i>	<i>A. mellifera</i>	<i>B. mori</i>	<i>A. gambiae</i>	<i>A. aegypti</i>	<i>D. melanogaster</i>
ACP	126266656, 126365747		111056970			100142386	100313509					5574131	
ADF						100141528, 662321							
AGA	126354006		111056802	RPRC001280-PA		656848	116738550	105835423	100576512	101744163	4578402	110676679	
AKH	126355897	BGER010868, BGER019360, BGER019359	111047315	RPRC000416-PA	OFAS011374	100135764, 663044	100679819	105837209	102656552	100127025, 100174925, 100174839	4578190, 3290616	5575676	38495
APD									406140				
ASTA	126281795		111057660		OFAS019254		100679594	105833574	724880	692588	1274391	5579498	42947
ASTB	126282023	BGER000061	111051590		OFAS011244	100142117				692434	1277431	5575871	39933
ASTC			111058146			100216359		114255333		100174830	1279573	5566967	34537
ASTCC	126354783	BGER027815	111056723	RPRC000300-PA	OFAS008438	100142376	100680495	105837647		101746311			34538
ASTCCC	126354341				OFAS025029		100679486	114255333	725616				
AT	126354277	BGER010050	111052211		OFAS004904	100216358	100680367	105839195	102656116	101740561	1280550	5572123	
AVLP	126281711		111052819			100038343	100122939	105831761					
BP									409634				
BURS	126271888, 126278425	BGER016582	111058210, 111054895	RPRC008554-PA	OFAS008719	654923	100124285, 100115714	105827899	100049551	100049169	5667449, 1272023, 3291626	110677973, 5569473, 110680467, 5572659	42560
CALa	126316172			RPRC000977-PA				105836878					
CALb	126267544			RPRC000977-PB									
CAPA	126281579	BGER000067	111053583			103312856		105829981	100577997	100174831	3289665	5566490	43541
CCAP	126304998	BGER003619	111055511	RPRC000466-PA	OFAS007216	664330		105829336		100174843	1279137	5564839	42683
CCH	126293515, 126293434	BGER000084	111059747		OFAS011242	103315160, 103312182	107981500, 100679516	105833689	725573	100170645, 101735342	4576789	5565599	41711, 41648
CNLP	126299196												
CNM	126356001	BGER013166	111050004	RPRC010893-PA	OFAS002244	103314591	100679460	105838482				5573484	38242
CRF-DH				RPRC000596-PA		659151, 661908		105836878					
CRZ	126355922	BGER016651	111045052		OFAS005241		100116119	105830031	503862	692443	4576842		41742
DH31	126298939	BGER014739	111051109, 111052788	RPRC010407-PA	OFAS003583	103314709	100119132, 100115139	105833436	113218809	100174854	1270451	5569983	34181
DH44	126335335		111046823		OFAS003428	661908	100679179	105836878	100576130	100174841	4577248		41170
EH	126266898, 126266900, 126266901	BGER026455	111062396, 111062397	RPRC014242-PA	OFAS009237	657624, 103312180	100302296, 100313499, 100115151, 107981062	105830409	726377	692740	4577883	5574558, 5579347, 5568324	42101

Tabla 5. Continuación.

NP	S. gregaria	B. germanica	N. lugens	R. prolixus	O. fasciatus	T. castaneum	N. vitripennis	M. pharaonis	A. mellifera	B. mori	A. gambiae	A. aegypti	D. melanogaster
ELE	126354997		111043329	RPRC003084-PA, RPRC003083-PA	OFAS015577			105833586	100576796				
ETH	126297876		111043989	RPRC014486-PA	OFAS025148	100337636	100117747	105839257	100216489	100337635	1270042	5572299	44125
FMRF	126281568	BGER020366	111051572	RPRC014988-PA	OFAS025208	103312524	100120410	105833915	100578542	692549	3289990	5578330	36030
GON	126282128												
GPA2	126278307	BGER011420	111064632	RPRC007092-PA	OFAS004661	659603				100174849	1277683	110676562	5740142
GPB5	126278287	BGER011418	111063980			659543				100174855, 101737993	3291695	5570903	3355097
HAN	126298401					103314778							
HP												110674066	
HUG													41547
IDLSRF	126298687												
IMF										100174833, 119630807			
IPTH	126284832												
IRP	126282268, 126281555	BGER020359, BGER013097	111053537, 111046284, 111053538	RPRC009717-PA	OFAS018844, OFAS018845	100142381	100117514	105832901	726057	*	1275326	5565037, 5567483	39149, 2768992
ITG	126299204		111050122		OFAS025288	662646	100119132	118644053	726472	119631002, 100174834	1279955	5573079	
ITP	126354358	BGER014246	111057547	RPRC000519-PA	OFAS003278	657659	100120410	105836100	724270	100127118	1274736	5567317	37921
KIN	126343458		111054501	RPRC000022-PA	OFAS017943					100174837	11176097	5572957	
LK													47746
LST													36853
MS	126297601	BGER009443	111052056	RPRC000203-PA	OFAS013557		100116855	105836520	726861	100379613	1281700		44324
NLGN	126330110, 126325620, 126336045, 126365371		111059450, 111058443, 111051396, 111044814, 120350722	RPRC002570-PA, RPRC006236-PA	OFAS015910	662358, 663865, 658512, 659717, 659779	100122116, 100123248, 100114473, 100123771	105838638	724137, 411321, 410779, 409003	101742943, 101744567, 101739014, 101739442	1273785, 1281357, 1280747	23687515, 5564665	42402, 33962
NP7B2	126278220		111056952	RPRC015183-PA	OFAS011270	662508	100116354	105838638	408614	101736661	1277869	5573386	40644
NPA		BGER007679	111045335	RPRC010064-PA, RPRC002095-PA	OFAS014019, OFAS014020, OFAS014021, OFAS014022	103313182		105837629	678511	100174836	1272160	5564261	
NPAA	126267355	BGER007679, BGER025495			OFAS007227		100313506						
NPF	126298941, 126293535	BGER000113	111045479	RPRC008107-PA	OFAS025089		100313503		726526	100174829, 100174835	1275883	5575886	42018
NPLP	126331832	BGER012245	111057907, 111058001, 111046511	RPRC011668-PA, RPRC009481-PA, RPRC006960-PA	OFAS010770	658567, 103314142, 657788		105832746	100577694	101747139, 100174827, 105842386	1272677, 1276816	5565029	39798, 50190, 50286, 38003, 36186, 59235, 5740215, 37666, 50191, 317945, 42967
NTL	126343442	BGER011069		RPRC003680-PA						110386111			5740700
*	100169724, 119629139, 100169720, 100169721, 100146109, 100169718, 100169726, 100147698, 100169657, 100169708, 119629137, 101736783, 100169712, 100169719, 100169716, 100169723, 100169658, 119629138, 100169722, 119629140, 100147697, 100169656, 100169725, 100169707, 100169715												

Tabla 5. Continuación.

NP	S. gregaria	B. germanica	N. lugens	R. prolixus	O. fasciatus	T. castaneum	N. vitripennis	M. pharaonis	A. mellifera	B. mori	A. gambiae	A. aegypti	D. melanogaster
NVP	126336393			RPRC003052-PA	OFAS010896	103314006		105837869					
ORCK	126355637	BGER026739, BGER026740	111046352	RPRC014678-PA	OFAS015031		100680225	105836186	726294	100174840	1280471	5569132	37806
PBAN	126324585	BGER003223	120349819	RPRC001168-PA	OFAS010321	103314974	116416908	105830114	100126690	692749	1269266	5575794	
PBURS	126295353	BGER016583	111049233		OFAS008715	100135757	100116744	105827899	413045	692730	1274652	5578519	34845
PDF		BGER005904	111044058		OFAS015779		100302295, 100678187	105829752	100577073	692465	1276445	5572286	43193
PK	126282095, 126324585	BGER003223	120350398				100313507						
PNLP	126273067												
PRO2									409314				
PRO4									409241				
PROC	126298899	BGER009478	111048524, 111048047	RPRC000390-PA		103312231	107980658	105832977		101744130		5565664	250710
PTTH	126304856		111054290		OFAS025263	LOC103315167	103315910	105836183		692767	3289772	110677702	33238
RFL	126272508												
RY	126299556	BGER004936	111044744	RPRC000461-PA	OFAS018696	100750223	107980750		100577155	101742726	1270298	5575249	5740597
SGSSP	126354890												
SIF	126297877	BGER020102	111061372		OFAS005914	100142131	100313499	105838647	100578023	100174832, 119628326, 101741959	1270047	5572527	578
SK	126273165	BGER004998	111060391	RPRC000519-PA	OFAS003841	103313277		105839297	102654729		4578092	110677123	45845
SNPF	126293576	BGER000113	111043799		OFAS003796	103314419	100678187	105838249	100576163	100174974		5580248	35286
SP												5564261	39494, 2768691
TK	126343426	BGER011075	111053004	RPRC000843-PA	OFAS010422	664261	100123761	105833215	406083	100174838, 693012, 100169709	1279443	5572020	41456
TRIS	126267461	BGER020113				103315201				101745081	4397725	5571024	41890
TRUNK						660640					3290689	5579958	34376
TSL	126281616		111049901	RPRC006513-PA			100118858	105831785	100576512	101738782, 101744755	1269277		42564

10 Producción científica

Parte de los resultados de este trabajo final de grado dieron origen a la siguiente presentación en congreso nacional:

Lavatti, Nicolás; Baricalla, Agustín; del Valle, Daiana; Rivera-Pomar, Rolando. “Transcriptomic meta-analysis to elucidate the presence and expression of neuropeptides in insect embryos”. Presented in the lightning talk sessions of the XIII CAB2C - 13th Argentinian Conference in Bioinformatics and Computational Biology. XIII SolBio - 13th International Conference of the Iberoamerican Society of Bioinformatics. III Annual Meeting of the Ibero-American Artificial Intelligence Network for Big BioData. From 1-3 November 2023, Rosario, Prov. de Santa Fe, Argentina