

Contribución al modelo de pérdida de ADN para el estudio de la relación evolutiva de las familias de neuropéptidos: LWamida, APGWamida, Hormona Concentradora de Pigmentos Rojos, Hormona Adipocinética, Corazonina y Hormona Liberadora de Gonadotropina

Cristian Enrique Cadena Caballero

Trabajo de grado para optar el título de Biólogo

Co-Investigador a cargo del estudiante (Tutor)

Francisco José Martínez Pérez

Doctor en Ciencias

Director del centro o laboratorio

Carlos Jaime Barrios Hernández

Doctor en Informática

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Bucaramanga

2020

Dedicatoria

A mi país Colombia.

A mi madre Victoria Caballero Barajas que es el amor más grande que tengo, que sin su esfuerzo y su berraquera no sería la persona que soy hoy en día.

A mi padre Milton Marino Cadena López que, sin sus consejos y su cariño jamás podría haber superado esta etapa.

A mi hermano Sender David Cadena Caballero que me enseñó desde pequeño la importancia de la dedicación y del trabajo duro.

A mi tía María Teresa Cadena López que, sin ella, su alegría, sus regaños y su ternura no habría podido realizar todas mis metas.

“Si te tienes que marchar, vete hijo mío y no regreses hasta que cumplas tus sueños”.

MMCL.

Agradecimientos

A los proyectos: Ciencia Básica 2015-255664 del Consejo Nacional de Ciencia y Tecnología (CONACYT) y 5713 de la Vicerrectoría de Investigación de la Universidad Industrial de Santander.

A la Facultad de Ciencias del Mar de la Universidad Autónoma de Sinaloa en Mazatlán, Sinaloa, México y a la Cátedra-CONACYT del Consejo Nacional de Ciencia y Tecnología del CDMX, México.

Al Dr. Francisco José Martínez Pérez, a la Dra. Laura Rebeca Jiménez Gutiérrez y al Dr. Carlos Jaime Barrios Hernández por permitirme hacer parte de su grupo de trabajo.

A la Universidad Industrial de Santander.

A la Facultad de Ciencias de la Universidad Industrial de Santander.

A la Escuela de Biología de la Universidad Industrial de Santander.

Al Laboratorio de Genómica de Celomados (GCL-UIS).

Al grupo de Computo Avanzado y a Gran Escala (CAGE).

Al grupo *Biomedical Imaging, Vision and Learning Laboratory* (BIVL2ab).

Al Grupo de Investigación en Microbiología y Genética (GIMG).

A mis familiares y amigos por todo su apoyo.

A todos ustedes lectores.

TABLA DE CONTENIDO

	Pág.
Introducción.....	13
1. Objetivos	19
1.1 Objetivo de la pasantía	19
1.2 Objetivo general	19
1.3 Objetivos específicos	19
2. Competencias adquiridas	20
2.1 Competencias cognitivas	20
2.2 Competencias actitudinales.....	20
3. Metodología	21
3.1 Identificación de LW, APGW, RPCH, AKH, CRZ y GnRH en las bases de datos	21
3.2 Depuración de las secuencias precursoras neuropeptídicas	21
3.3 Determinación de estructura primaria de los precursores neuropeptídicos	21
3.4 Desarrollo del código de identificación para los precursores neuropeptídicos.....	21
3.5 Construcción de la base de datos neuropeptídica.....	22
3.6 Obtención secuencias precursoras neuropeptídicas por transcriptómica	22
3.7 Alineamiento de secuencias neuropeptídicas.....	22
3.8 Generación de secuencias precursoras consenso	23
3.9 Validación <i>in silico</i> de péptidos virtuales del DNA-LM en el GenBank.....	23
3.10 Relaciones filogenéticas entre las familias neuropeptídicas	23

4. Resultados	24
4.1 Identificación de los precursores neuropeptídicos	24
4.2 Depuración de las secuencias precursoras neuropeptídicas	24
4.3 Determinación de la estructura primaria de los precursores neuropeptídicos.....	26
4.4 Construcción de la base de datos y código de identificación.....	27
4.5 Secuencias precursoras obtenidas por transcriptómica	28
4.6 BLAST péptidos virtuales APGW/RPCH	29
4.7 Alineamiento de secuencias y generación de secuencias precursoras consenso.....	30
4.8 Relaciones filogenéticas.....	32
5. Discusión	36
6. Conclusiones	43
7. Recomendaciones	44
Referencias bibliográficas	45

Lista de tablas

Tabla 1. Características de las familias neuropeptídicas.....	17
Tabla 2. Secuencias precursoras híbridas AKH/CRZ-peptido relacionado (ACP)...	25

Lista de figuras

Figura 1. Biosíntesis de los precursores neuropeptídicos.....	14
Figura 2. Valor porcentual de las familias neuropeptídicas en el GenBank.....	25
Figura 3. Secuencia híbrida ACP de <i>C. toxotes</i> y CRZ de <i>C. arcuatus</i>	28
Figura 4. Representación gráfica mediante el péptido activo del DNA-LM	31
Figura 5. Árbol filogenético realizado para los péptidos virtuales APGW/RPCH.....	34
Figura 6. Árboles filogenéticos concatenados de las familias neuropeptídicas.....	35

Lista de apéndices

Los apéndices se adjuntan en el CD y puede visualizarlos en base de datos de la biblioteca UIS.

Apéndice A. Búsqueda del GenBank mediante el uso de las palabras claves.

Apéndice B. Resultados obtenidos mediante la implementación de BioDataToolKit.

Apéndice C. Resultados obtenidos y descripción de manejo de Proteios.

Apéndice D. Depuración de las secuencias precursoras neuropeptídicas.

Apéndice E. Formatos tipo FASTA de AKH/CRZ-péptido relacionado (ACP).

Apéndice F. Resultados obtenidos mediante la implementación de ProP Server.

Apéndice G. Resultados obtenidos de la identificación manual de los sitios de corte.

Apéndice H. Sitios de escisión (R/K), revisión manual y con software ProP Server.

Apéndice I. Sitios de corte verificados manualmente para las secuencias.

Apéndice J. Estructura promedio del precursor y el péptido activo.

Apéndice K. Resultados de la depuración manual de las familias neuropeptídicas.

Apéndice L. Depuración manual y creación de formatos FASTA.

Apéndice M. Base de datos de precursores neuropeptídicos.

Apéndice N. Descripción de la base de datos de precursores neuropeptídicos.

Apéndice Ñ. Códigos bioinformáticos de identificación neuropeptídica.

Apéndice O. Alineamiento tipo BLAST para CRZ de *C. arcuatus* y ACP de *C. toxotes*.

Apéndice P. Alineamiento tipo BLAST para los péptidos virtuales APGW/RPCH.

Apéndice Q. Grupos taxonómicos asociados a los péptidos virtuales.

Apéndice R. Alineamiento de secuencias con los parámetros del DNA-LM y Kaling.

Apéndice S. Secuencias consenso obtenidas con DNAMAN.

Apéndice T. Relaciones filogenéticas de los precursores neuropeptídicos.

Apéndice U. Descripción de las familias LW, APGW, RPCH, AKHs CRZ y GnRH.

Apéndice V. Imágenes en alta resolución.

RESUMEN

TÍTULO: CONTRIBUCIÓN AL MODELO DE PÉRDIDA DE ADN PARA EL ESTUDIO DE LA RELACIÓN EVOLUTIVA DE LAS FAMILIAS DE NEUROPEPTIDOS: LWAMIDA, APGWAMIDA, HORMONA CONCENTRADORA DE PIGMENTOS ROJOS, HORMONA ADIPOCINÉTICA, CORAZONINA Y HORMONA LIBERADORA DE GONADOTROPINA*

AUTOR: CRISTIAN ENRIQUE CADENA CABALLERO**

PALABRAS CLAVES: LW, APGW, RPCH, AKH, ACP, CRZ, GnRH, Precursor neuropeptídico virtual, Modelo evolutivo, Duplicación de ADN, Pérdida de ADN.

DESCRIPCIÓN:

Los neuropeptidos son una familia de neuromoduladores involucrados en diversos procesos fisiológicos. Uno de los modelos propuestos para explicar su evolución, es el Modelo de Pérdida de ADN (DNA-LM), el cual postula que la pérdida de codones de la LWamida de hidras y la APGW de moluscos produjeron precursores virtuales (APGW/RPCH), que originaron a la RPCH de crustáceos y las AKH 1-4 de insectos. Diversos grupos de investigación han demostrado la existencia del péptido híbrido AKH/CRZ-péptido relacionado (ACP) en insectos y crustáceos, así como su relación con la CRZ de insectos y la GnRH de vertebrados. Para validar, el DNA-LM entre estas familias y el precursor APGW/RPCH, se obtuvieron 9489 secuencias del GenBank. La implementación de tres softwares y un código de selección demostró que 511 secuencias corresponden a precursores neuropeptídicos. Con estos, se construyó una base de datos que permitió determinar RPCH, ACP y AHK 1 no caracterizadas. También permitió establecer la CRZ y la ACP en los transcriptomas de *Callinectes arcuatus* y *C. toxotes*, respectivamente. Además, se confirmó que los precursores virtuales APGW/RPCH presentan homología con sus respectivas familias y con otros *phyla* originalmente no asociados al modelo. La concatenación de las filogenias de cada familia con los parámetros del software Kaling y los del DNA-LM mostraron una división entre las GnRH de protostomados y deuterostomados. La primera mostró un clado de cefalópodos entre la RPCH y una secuencia AKH 4, mientras que las otras familias se agruparon en sus respectivos clados. En tanto, la filogenia con DNA-LM mostró un clado de GnRH, principalmente de vertebrados entre los grupos AKH 3 y CRZ, otro de anfibios entre AKH 1 y CRZ, mientras que los demás clados presentaron combinaciones entre las familias. El DNA-LM permite correlacionar evolutivamente familias neuropeptídicas por pérdida de ADN.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Biología.

Tutor: Francisco José Martínez Pérez. Director del centro: Carlos Jaime Barrios Hernández.

ABSTRACT

TITLE: CONTRIBUTION TO THE DNA LOSS MODEL FOR THE STUDY OF THE EVOLUTIONARY RELATIONSHIP OF NEUROPEPTIDE FAMILIES: LWAMIDE, APGWAMIDE, RED PIGMENT CONCENTRATING HORMONE, ADIPOKINETIC HORMONE, CORAZONIN AND GONADOTROPIN-RELEASING HORMONE*

AUTHOR: CRISTIAN ENRIQUE CADENA CABALLERO**

KEY WORDS: LW, APGW, RPCH, AKH, ACP, CRZ, GnRH, Virtual neuropeptide precursor, Evolutionary model, DNA Duplication, DNA Loss.

DESCRIPTION:

Neuropeptides are a family of neuromodulators involved in various physiological processes. One of the models proposed to explain its evolution is the DNA Loss Model (DNA-LM), which postulates that the loss of codons of the LWamide of hydras and the APGW of mollusks produced virtual precursors (APGW/RPCH), which originated the RPCH of crustaceans and the AKH 1-4 of insects. Several research groups have demonstrated the existence of the hybrid peptide AKH/CRZ-related peptide (ACP) in insects and crustaceans, as well as its relationship with insect CRZ and vertebrate GnRH. To validate, the DNA-LM between these families and the APGW/RPCH precursor, 9489 GenBank sequences were obtained. The implementation of three softwares and a selection code showed that 511 sequences corresponds to neuropeptide precursors. With these, a database was built that allowed to determine RPCH, ACP and AHK 1 not characterized. It also allowed the establishment of CRZ and ACP in the transcriptomes of *Callinectes arcuatus* and *C. toxotes*, respectively. In addition, it was confirmed that APGW/RPCH virtual precursors have homology with their respective families and with other *phyla* originally not associated with the model. The concatenation of the phylogeny of each family with the parameters of the Kaling software and the DNA-LM showed a division between the GnRH of protostomes and deuterostomes. The first showed a clade of cephalopods between the RPCH and an AKH 4 sequence, while the other families were grouped in their respective clades. Meanwhile, the phylogeny with DNA-LM showed a clade of GnRH, mainly of vertebrates between the AKH 3 and CRZ groups, another of amphibians between AKH 1 and CRZ, while the other clades presented combinations between the families. The DNA-LM allows the evolutionary correlation of neuropeptide families by loss of DNA.

* Bachelor thesis

** Science Faculty. School of Biology.

Tutor: Francisco José Martínez Pérez. Center director: Carlos Jaime Barrios Hernández.

Introducción

Los neuropéptidos son cadenas peptídicas de 3 a 100 aminoácidos (Merighi, 2009). Son producidos y liberados por neuronas, mediante la vía secretora regulada e interaccionan sobre diversos sustratos neurales, convirtiéndose en la clase más diversa de elementos de señalización en el cerebro y de comunicación en el sistema nervioso central de vertebrados e invertebrados (Burbach, 2011). Estos se desempeñan como reguladores autocrinos o paracrinos, moduladores de la neurotransmisión y al controlar el crecimiento, desarrollo celular, ecdisis, digestión gástrica y diuresis (Burbach, 2011). De igual forma, son cruciales en el comportamiento alimenticio, el cortejo, el sueño, el aprendizaje, la memoria, el estrés y desordenes psiquiátricos (Belzung, Yalcin, Griebel, Surget, & Leman, 2006; Schoofs, De Loof, & Van Hiel, 2017).

El modelo propuesto de biosíntesis de los neuropéptidos inicia en el núcleo celular a partir de la transcripción del gen que codifica para la formación del precursor neuropeptídico (Fig. 1). Este puede llegar a producir uno o más ARN mensajeros (ARNm), por el procesamiento alternativo y su traducción genera una o más proteínas denominadas precursores neuropeptídicos también conocidos como prepropéptidos, compuestos por: el péptido señal, el péptido activo y el péptido relacionado (Fig. 1) (Rouillé et al., 1995).

El péptido señal de los neuropéptidos tiene una secuencia de 16 a 30 aminoácidos que se subdividen en tres regiones: La región-N de característica preferentemente hidrofílica con carga positiva, seguida de la región-H hidrófoba ubicada en el segmento central y la región-C terminal de tamaño corto con la señal del sitio de escisión para la peptidasa en el Retículo Endoplasmático Rugoso [RER] (Steiner, 1998). Estas regiones varían dependiendo de la especie y estado fisiológico, encontrándose inclusive regiones-N extendidas o dos regiones-H (Von Heijne, 1990).

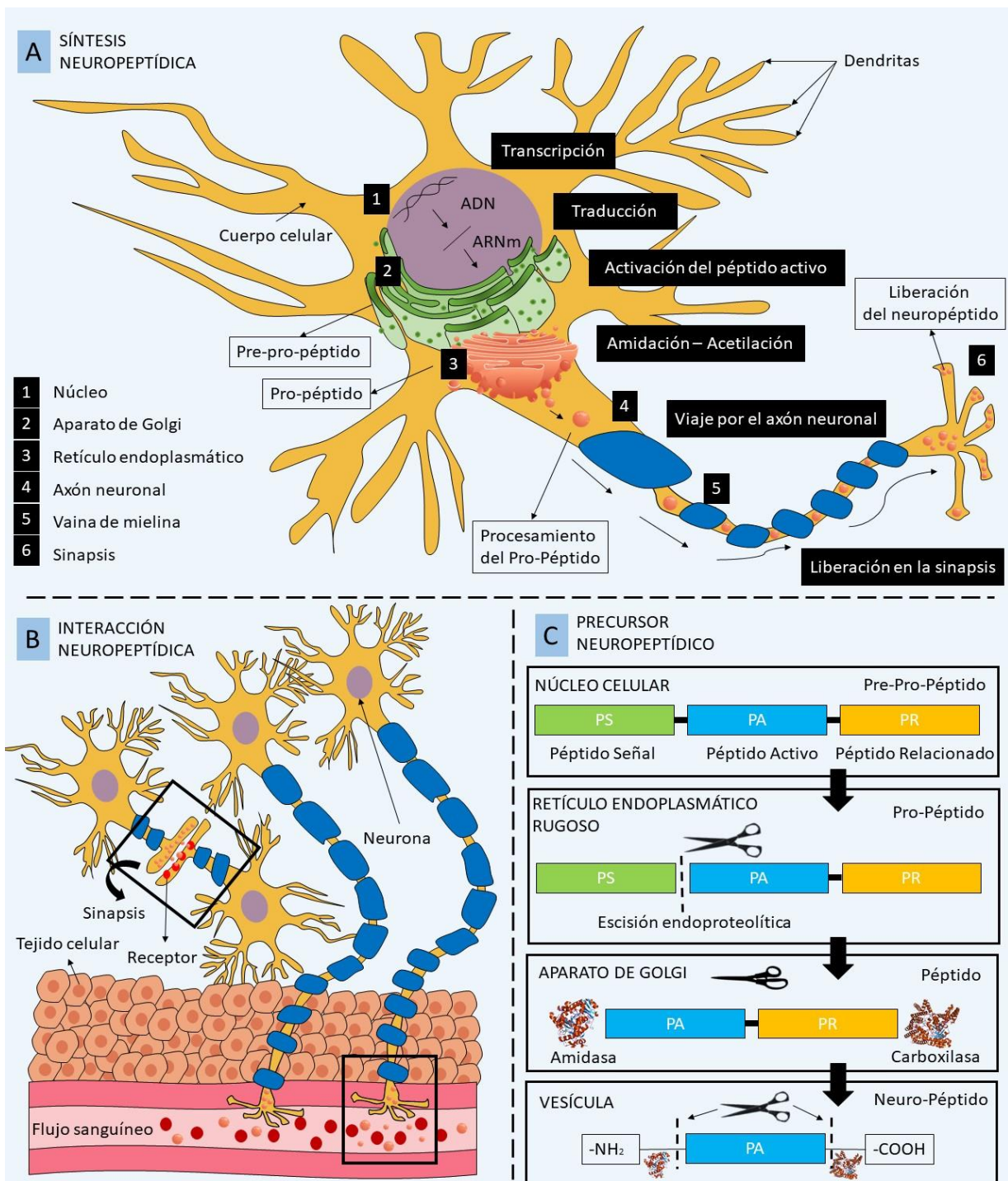


Figura 1. Biosíntesis de los precursores neuropeptídicos en neuronas neuroendocrinas. A) Síntesis neuronal de los precursores neuropeptídicos desde su transcripción en el núcleo celular hasta su liberación en la sinapsis. B) Interacción celular del neuropéptido en su liberación dentro de los organismos. Los recuadros negros indican los sitios de correlación del neuropéptido en la sinapsis neuronal y su liberación por el torrente sanguíneo. C) Procesamiento neuronal en la biosíntesis de precursor neuropeptídico, iniciando en el núcleo celular, el corte del péptido señal en el retículo endoplasmático, la adición de las enzimas, el ensamble postraduccional en el Aparato de Golgi y finalmente la amidación y la carboxilación en la vesícula.

Además, el análisis molecular y bioinformático del precursor neuropeptídico ha facilitado el desarrollo de herramientas que permiten predecir con exactitud el tamaño del péptido señal y los sitios de corte endoproteolíticos del prepropéptido (Nielsen, Brunak, & Von Heijne, 1999).

El péptido activo se define como un polipéptido de 2 a 15 aminoácidos, el cual se considera como la región más conservada evolutivamente dado que al transmitir el mensaje por la sinapsis neuronal o al ser liberado en el torrente circulatorio genera la actividad biológica, que actúa en la célula u órgano receptor (Fig. 1) (De Loof & Schoofs, 1990; Steiner et al., 1980). Posterior a la escisión del péptido señal en el RER, el péptido se transporta al Aparato de Golgi y se encapsula en vesículas, las cuales viajan por el axón neuronal hasta la sinapsis, proceso en donde ocurren las modificaciones postraduccionales, las cuales dependiendo del neuropeptido pueden ser: la formación de enlaces disulfuro, glicosilación, acilación, adición de carbohidratos, fosforilación, ubiquitinación y acetilación (Eipper, Stoffers, & Mains, 1992).

Otra estructura que se genera en el procesamiento del precursor neuropeptídico ocurre por la acción de endoproteasas que reconocen secuencias de aminoácidos básicos y generan una molécula denominada péptido relacionado, la cual no tiene función extracelular y corresponde a la secuencia final del precursor neuropeptídico. Esta, controla y regula la bioquímica de los cortes proteolíticos que activan al neuropeptido (Steiner et al., 1980) y es degradada con proteasas (Rouillé et al., 1995).

Los neuropeptidos maduros en el espacio extracelular son reconocidos por los receptores acoplados a proteínas G (GPCRs), que pertenecen a las familias de receptores rodopsina- β , rodopsina- γ y secretina (Ritschard, Fitak, Simakov, & Johnsen, 2019).

Debido a la importancia de los neuropeptidos tanto en su estructura, señalización y función biológica de las especies, se requiere la generación de conocimiento relacionado a describir su

regulación molecular, sus mecanismos de control, sus procesos fisiológicos y su origen evolutivo (Mains & Eipper, 2006). Actualmente, el uso de tecnologías de Secuenciación de Nueva Generación (NGS, por sus siglas en inglés) producen datos genómicos y transcriptómicos que se reportan en bases de datos públicas como *Nucleotide* o *Protein* en el GenBank del Centro Nacional para la Información Biotecnológica (NCBI, por sus siglas en inglés; Sayers et al., 2019). Estas cuando tienen una correcta caracterización se pueden comparar y clasificar, por su similitud estructural, bioquímica, tamaño, posición o cantidad de aminoácidos agrupándose en familias neuropeptídicas para establecer sus relación evolutiva (Zandawala, Tian, & Elphick, 2018).

Ejemplo de lo anterior, son los neuropéptidos que se emplearon en este estudio, los cuales se clasifican por su secuencia aminoacídica en las familias LWamida y APGWamida. Mientras que las demás familias se nombran por su actividad fisiológica: Hormona Concentradora de Pigmentos, Hormonas Adipocinética, Corazonina y Hormona Liberadora de la Gonadotropina (Tabla 1). Además, ambas clasificaciones se usan para identificar y nombrar nuevos precursores neuropeptídicos homólogos (De Loof & Schoofs, 1990; Nowicki, Bzhalava, & Bała, 2018).

Así, los neuropéptidos LW y APGW comparten simultáneamente un alto número de copias en su prepropéptido, pero se diferencian porque la APGW en el extremo carboxilo, la Leucina fue reemplazada por Glicina y se adicionaron Alanina y Prolina (Martínez-Pérez et al., 2007).

La RPCH y algunas AKH tienen un péptido activo de 8 aminoácidos, mientras que otras AKH contienen de 9 a 12, en donde la cuarta posición presenta preferentemente una Fenilalanina, el extremo carboxilo tiene aminoácidos similares a la LW y a la APGW y difieren en el tamaño del precursor (Gäde, 2009; Martínez-Pérez et al., 2007); la CRZ y GnRH por su parte comparten más del 70% de similitud en el péptido activo y se han planteado como parte de la superfamilia gonadotrópica (Roch, Busby, & Sherwood, 2011; Sakai et al., 2017; Zandawala et al., 2018).

Tabla 1.

Características neuropeptídicas de las familias LW, APGW, RPCH, AKH, CRZ y GnRH.

Neuropéptido	Abreviatura	Número de copias	Clasificación	Descripción	Referencia
LWamida	LW*	> 11	Secuencia aminoacídica	Regeneración tisular y la inducción de la metamorfosis	(Gajewski et al., 1996)
APGWamida	APGW*	10 < 19	Secuencia aminoacídica	Comportamiento sexual y espermatogénesis	(Ahn et al., 2017)
Hormona Concentradora de Pigmentos Rojos	RPCH	1	Función biológica	Ritmo circadiano de concentración de pigmentos en retina y cromatóforos	(Martínez-Pérez, Zinker, Aguilar, Valdés, & Aréchiga, 2005; Mordue & Stone, 1977)
Hormona Adipocinética	AKH	1	Función biológica	β-oxidación y la glucólisis durante el vuelo	(Robinson & Goldsworthy, 1976; Stone, Mordue, Batley, & Morris, 1976)
Corazonina	CRZ	1	Función biológica	Frecuencia cardíaca, jerarquía social, inicio a la ecdisis y desencadena la pigmentación	(Gospocic et al., 2017; Veenstra, 1989)
Hormona Liberadora de la Gonadotropina	GnRH	1	Función biológica	Estimulación para segregar LH y FSH	(Stamatiades & Kaiser, 2018)

Nota: *Precusores neuropeptídicos descritos por su secuencia particular de aminoácidos.

Por lo anterior, se observa que existe relación y/o homología entre ellos, que generalmente ocurre en el péptido activo y aun cuando los estudios fisiológicos y evolutivos del precursor neuropeptídico completo son escasos, el Modelo de Pérdida de ADN (DNA-LM, por sus siglas en inglés), plantea que la evolución de genes neuroendocrinos ocurrió mediante la duplicación de un gen para un precursor neuropeptídico con distinto número de copias y debido al movimiento de intrones y pérdida de codones se generó un nuevo gen en especies filogenéticamente cercanas (Martínez-Pérez et al., 2007). Inicialmente se consideró a la AKH y la RPCH como neuropéptidos independientes (Christensen, Carlsen, & Josefsson, 1978, 1979), sin embargo, por su similitud con la APGW de moluscos (Kuroki et al., 1990) y LW de hidras (Gajewski et al., 1996) y mediante el DNA-LM, se demostró que los primeros cuatro aminoácidos y los restantes de la AKH y RPCH se generaron por la pérdida y fusión de codones de una las copias del prepropéptido LW y APGW, generando precursores híbridos LW/APGW, APGW/RPCH y RPCH/AKH en especies que aún no han sido descritas o ya extintas (Martínez-Pérez et al., 2007).

Los neuropéptidos híbridos han sido buscados en invertebrados cuyo plan corporal fuese similar a especies de la explosión del Cámbrico, pero debido las técnicas de extracción genómica y

características intrínsecas de su secuencia nucleotídica, únicamente con la Reacción en Cadena de la Polimerasa (PCR, por sus siglas en inglés), se originaron amplicones que pudieron ser clonados mas no secuenciados lo que sugirió que si podían existir los neuropéptidos híbridos. Sin embargo, con la NGS aplicada a invertebrados marinos y con base a los datos moleculares ya existentes (Hauser & Grimmelikhuijzen, 2014; Tian et al., 2016; Tsai, 2018; Zandawala et al., 2018), el grupo de investigación donde se realizó esta pasantía, confirmó que la posición de los intrones propuestos para la APGW de *Aplysia californica* corresponden al modelo DNA-LM. . En el año 2014, se publicó para esta misma especie un precursor similar a la AKH (Johnson, Kavanaugh, Nguyen, & Tsai, 2014), que pudiese correlacionar con alguno de los péptidos híbridos de APGW/RPCH, como se propuso en el DNA-LM (Martínez-Pérez et al., 2007). Además en este periodo de tiempo también se caracterizó la GnRH en moluscos (Osada & Treen, 2013), por lo que Roch et al., (2011) y Zandawala et al., 2018, postularon que dada la similitud y posición de los aminoácidos entre la RPCH, AKH, CRZ y GnRH, que todas ellas pertenecen a una misma familia evolutiva y cuyo posible origen fue desde la GnRH, lo cual es opuesto al DNA-LM.

No obstante, con los datos de estos neuropéptidos se confirmó la evolución del receptor/ligando, donde estos se duplicaron en dos o más ocasiones, dando diferencias en la señalización dentro de las familias (Patel, Orchard, Veenstra, & Lange, 2014). Puesto que existe una discrepancia entre el DNA-LM y los modelos anteriormente mencionados con relación a la evolución de los neuropéptidos. En esta pasantía, se contribuyó con la generación de conocimiento para esclarecer y determinar las relaciones evolutivas entre ellas al crear una base de datos con datos publicados genómicos y su asociación con las secuencias híbridas para validar o refutar el DNA-LM.

1. Objetivos

1.1 Objetivo de la pasantía

Contribuir al estudio evolutivo con base en el modelo de pérdida de ADN de los neuropéptidos LWamida, APGWamida, Hormona Concentradora de Pigmentos Rojos, Hormona Adipocinética, Corazonina y Hormona Liberadora de la Gonadotropina mediante el análisis bioinformático de sus precursores reportados en el GenBank.

1.2 Objetivo general

Determinar las características de los precursores para la LWamida, APGWamida, Hormona Concentradora de Pigmentos Rojos, Hormona Adipocinética, Corazonina y Hormona Liberadora de la Gonadotropina reportados en el GenBank, para ser considerados en la validación del modelo pérdida de ADN.

1.3 Objetivos específicos

Establecer las características de las secuencias de los precursores para LWamida, APGWamida, Hormona Concentradora de Pigmentos Rojos, Hormona Adipocinética, Corazonina y Hormona Liberadora de la Gonadotropina reportados en el GenBank con base en los elementos necesarios para su procesamiento postraducciona.

Diseñar una base de datos de la LWamida, APGWamida, Hormona Concentradora de Pigmentos Rojos, Hormona Adipocinética, Corazonina y Hormona Liberadora de la Gonadotropina reportados en las bases de datos del GenBank que correspondan a precursores de neuropéptidos.

Asociar los precursores de la LWamida, APGWamida, Hormona Concentradora de Pigmentos Rojos, Hormona Adipocinética, Corazonina y Hormona Liberadora de la Gonadotropina reportados en el GenBank con base al modelo de pérdida de ADN.

2. Competencias adquiridas

2.1 Competencias cognitivas

Genera la exploración y búsqueda digital de secuencias precursoras neuropeptídicas en la base de datos del GenBank para identificar los elementos requeridos en el procesamiento de precursores de neuropéptidos.

Desarrolla la capacidad analítica durante los procesos de selección de las secuencias precursoras neuropeptídicas para construir bases de datos bioinformáticas.

Crea códigos de lenguaje bioinformático para obtener formatos tipo FASTA de precursores neuropeptídicos necesarios en la aplicación de programas computacionales.

Realiza la construcción de bases de datos con precursores neuropeptídicos para identificar e interpretar resultados evolutivos.

Construye secuencias consenso de precursores neuropeptídicos presentes en especies filogenéticamente relacionadas para establecer sus regiones constantes y posiciones variables.

Analiza el modelo evolutivo de pérdida de ADN para establecer su capacidad de generar relaciones filogenéticas entre precursores de neuropéptidos y sus secuencias consenso.

2.2 Competencias actitudinales

Participa activamente en un grupo de investigación para resolver preguntas dirigidas a validar o cuestionar modelos evolutivos.

Integra conocimientos biológicos para tomar decisiones lógicas e imparciales.

Establece líneas de acción concretas a partir de su formación biológica integral para solucionar problemas científicos.

Muestra valores éticos y bioéticos para la ejecución y convivencia en la realización de proyectos de investigación científica.

3. Metodología

3.1 Identificación de LW, APGW, RPCH, AKH, CRZ y GnRH en las bases de datos

Se realizó la búsqueda en la base de datos GenBank del NCBI (Sayers et al., 2019) hasta el 4 de enero del 2019, con 10 palabras claves para cada familia neuropeptídica, como criterios de búsqueda se utilizaron: nombre del neuropéptido, abreviatura, nombre completo del precursor, nombre completo del precursor con la palabra prepropéptido y las combinaciones entre estas. Las secuencias precursoras neuropeptídicas candidatas se descargaron de la plataforma en formato tipo FASTA mediante la ruta establecida por el GenBank

3.2 Depuración de las secuencias precursoras neuropeptídicas

Para la depuración empleó el software BioDataToolKit versión 5.0 (Acuña, 2019), de acuerdo, a los parámetros establecidos por el diseñador. Los resultados se corroboraron y depuraron con el software Proteios versión 1.0. Los resultados finales se clasificaron en carpetas office versión 18.1910.1283.0 y almacenaron en archivos tipo FASTA para cada palabra clave y por familias.

3.3 Determinación de estructura primaria de los precursores neuropeptídicos

La estructura primaria se obtuvo con el software ProP Server versión 1.0 (Duckert, Brunak, & Blom, 2004), con la selección de generación de gráficos y la inclusión de la predicción del péptido señal (Nakayama, 1997). Los resultados se colocaron en archivos de texto enriquecido.

3.4 Desarrollo del código de identificación para los precursores neuropeptídicos

Debido a que las bases de datos contienen códigos numéricos en formato FASTA para usos bioinformáticos para la identificación de cada precursor en la base de datos se desarrolló con un código de identificación que mantenía el formato FASTA y cambió el orden numérico por la abreviatura del neuropéptido, el género en dos letras, la especie en 4, patrones de obtención y su fenotipo, todo separado con guiones bajos.

3.5 Construcción de la base de datos neuropeptídica

Se creó una hoja de cálculo en formato tipo Excel versión 1902. En la cual, se agregaron en forma de tabla cronológica y por familias neuropeptídicas: número total de secuencias, abreviatura del de precursor, el código generado por bioinformática y de identificación del GenBank, la especie de la cual fue obtenido, grupo taxonómico, el nombre común y características reportadas por los autores: presencia o ausencia de isoformas, predicción por computación, identificación de las secuencias similares (*like*), hipotéticas y/o provisionales. El título del artículo y de no encontrarse se describió la información presente en la sección *Model Refseq* del GenBank.

3.6 Obtención secuencias precursoras neuropeptídicas por transcriptómica

Se realizó la secuenciación del transcriptoma de *C. arcuatus* y *C. toxotes* con la tecnología de *Illumina MiSeq*. Las especies se capturaron en el Océano Pacífico (23° 20'N 106° 30'W) y de estas se utilizó un stock de tallo ocular, tejido de ovario y hepatopáncreas, que representó las etapas de madurez, la temporada de captura y el ritmo circadiano. El ARN total se aisló y purificó de todo el organismo de acuerdo a Jimenez-Gutierrez et al., (2019). Los stocks de ARN de cada especie se emplearon para la secuenciación de cada transcriptoma con la tecnología *Illumina MiSeq* de acuerdo con el procedimiento y ensamblaje propuesto por el fabricante. Ambos procesos se efectuaron mediante la contratación de los servicios de la empresa Genoma Mayor, Universidad Mayor de Chile (Jimenez-Gutierrez et al., 2019). Las secuencias de los precursores neuropeptídicos putativos se corroboraron mediante BLAST (Nowicki et al., 2018).

3.7 Alineamiento de secuencias neuropeptídicas

Se realizó el alineamiento con el software Kaling versión 2.0 (Lassmann, Frings, & Sonnhammer, 2009), en la página del Instituto Europeo de Bioinformática (EMBL-EBI) (Madeira et al., 2019), con los parámetros del DNA-LM (Martínez-Pérez et al., 2007).

3.8 Generación de secuencias precursoras consenso

Las secuencias precursoras consenso, se obtuvieron con el software DNAMAN versión 10.0, mediante un alineamiento rápido por pares (<https://www.lynnon.com/index.html>).

3.9 Validación *in silico* de péptidos virtuales del DNA-LM en el GenBank

Se realizaron alineamientos BLAST con los 6 precursores híbridos APGW/RPCH y APGW/RPCH/AKH propuestos en el DNA-LM (Martínez-Pérez et al., 2007).

3.10 Relaciones filogenéticas entre las familias neuropeptídicas

Los árboles se generaron con el software IQtree versión 1.6.12. (L. T. Nguyen, Schmidt, Von Haeseler, & Minh, 2014). Se estableció automáticamente el modelo de sustitución con ModelFinder (Kalyaanamoorthy, Minh, Wong, Von Haeseler, & Jermin, 2017), el tipo de partición *Edge-linked* con heterogeneidad *FreeRate* (+R), con 4 tipos de categorías y estado de frecuencia empírica. El análisis de soporte de rama se obtuvo con *Bootstrap Ultrafast* (Minh, Nguyen, & Von Haeseler, 2013), con número fijo de 1000 alineaciones y 0.99 de coeficiente mínimo de correlación con la prueba de rama única de 1000 réplicas (SH-aLRT) y el test aproximado de Bayes (aBayes) (Anisimova, Gil, Dufayard, Dessimoz, & Gascuel, 2011; Guindon et al., 2010). De igual forma, se utilizó *perturbation strength* de 0.5 y *IQ-Tree stopping rule* de 100. Para la concatenación de los árboles, se evaluaron los modelos de sustitución aminoacídica con ModelFinder. Los alineamientos de cada familia realizados con Kaling se modelaron con Mesquite versión 3.61 (Maddison & Madisson, 2018) y se utilizó *-bsam GENESITE* para remuestrear las particiones y luego los sitios dentro de particiones, con el fin de verificar y reducir los falsos positivos (Gadagkar, Rosenberg, & Kumar, 2005). Para la visualización de los resultados se utilizó el software ItoI versión 5.0 (Letunic & Bork, 2019).

4. Resultados

4.1 Identificación de los precursores neuropeptídicos

La búsqueda en el GenBank con los criterios de selección identificó un total de 9489 secuencias, consideradas por esta base de datos como posibles secuencias precursoras neuropeptídicas. De estas, la menor cantidad correspondió a la APGW con menos de 100, en tanto la LW, RPCH y CRZ presentaron 420, 644 y 763 secuencias, respectivamente. Las secuencias mayoritarias correspondieron a la AKH con 2167 y más del doble para GnRH con 5399 (Apéndice A).

4.2 Depuración de las secuencias precursoras neuropeptídicas

El software BioDataToolKit mostró un listado de 7208 secuencias candidatas de las 9489 iniciales. Muchas de las cuales correspondieron a secuencias: utilizadas en patentes con especies no determinados (*unknown*), información faltante (*empty*), marcadoras tipo *flag*, receptores, especies no correlacionadas con el DNA-LM y microorganismos. Por lo anterior, dichas secuencias se descartaron (Apéndice B).

Al verificar la depuración con el software Proteios, se obtuvieron 722 secuencias similares a precursores neuropeptídicos (Apéndice C). Al comparar los resultados con ambos softwares respecto a la cantidad inicial (Apéndice D), se observó una disminución en el número de secuencias candidatas para cada familia. BioDataToolKit mostró que las familias APGW, CRZ y LW obtuvieron la menor cantidad, las familias RPCH, AKH tuvieron el doble y la GnRH fue 18 a 20 veces mayor. Por su parte, Proteios indicó que la APGW es la familia con menor cantidad, la RPCH, la LW y la CRZ fueron intermedias con menos de 50 secuencias y la AKH y la GnRH fueron mayoritarias. Además, con los criterios de selección se identificó la familia híbrida ACP en artrópodos propuesta por (Hansen et al., 2010). Es de resaltar que los miembros de esta familia

presentaron las características de organización de los precursores neuropéptidos como se propuso por el DNA-LM para los precursores híbridos APGW/RPCH (Tabla 2 y Apéndice E).

Tabla 2.

Secuencias precursoras híbridas identificadas como AKH/CRZ-péptido relacionado (ACP).

N°	Neuropéptido	Código de identificación	Código del GenBank	Especies	Taxonomía
1	ACP	Int_AKH_CRZ_Ma_rose*	ANT96502	<i>Macrobrachium rosenbergii</i>	Crustacea
2	ACP	Int_AKH_CRZ_Ca_maen*	AVA26881	<i>Carcinus maenas</i>	Crustacea
3	ACP	Int_AKH_CRZ_Na_vitr	NP_001161199	<i>Nasonia vitripennis</i>	Hymenoptera
4	ACP	Int_AKH_CRZ_Rh_prol	AKO62855	<i>Rhodnius prolixus</i>	Hemiptera
5	ACP	Int_AKH_CRZ_He_vire	ADW77572	<i>Heliothis virescens</i>	Lepidoptera
6	ACP	Int_AKH_CRZ_Tr_cast	NP_001159497	<i>Tribolium castaneum</i>	Coleoptera
7	ACP	Int_AKH_CRZ_Pe_vann_1*	XP_027229238	<i>Penaeus vannamei</i>	Crustacea
8	ACP	Int_AKH_CRZ_Pe_vann_2*	ROT69296	<i>Penaeus vannamei</i>	Crustacea
9	ACP	Int_AKH_CRZ_Di_citr	XP_008480698	<i>Diaphorina citri</i>	Hemiptera
10	ACP	Int_AKH_CRZ_Ne_norv*	QBX89024	<i>Nephrops norvegicus</i>	Crustacea

Nota: *Las secuencias precursoras AKH/CRZ-péptido relacionado (ACP) obtenidas de especies de Crustáceos, también pueden nombrarse RPCH/CRZ por cuestiones de nomenclatura (Zandawala et al., 2018). Para efectos de evitar confusiones en este estudio serán nombradas AKH/CRZ.

Se observó que las secuencias precursoras neuropéptidas obtenidas de las familias LW, APGW, RPCH y ACP representaron menos del 10% de los neuropéptidos evaluados en este estudio, mientras que las familias AKH, CRZ y GnRH presentan más del 90% de especies aquí evaluadas (Fig. 2).

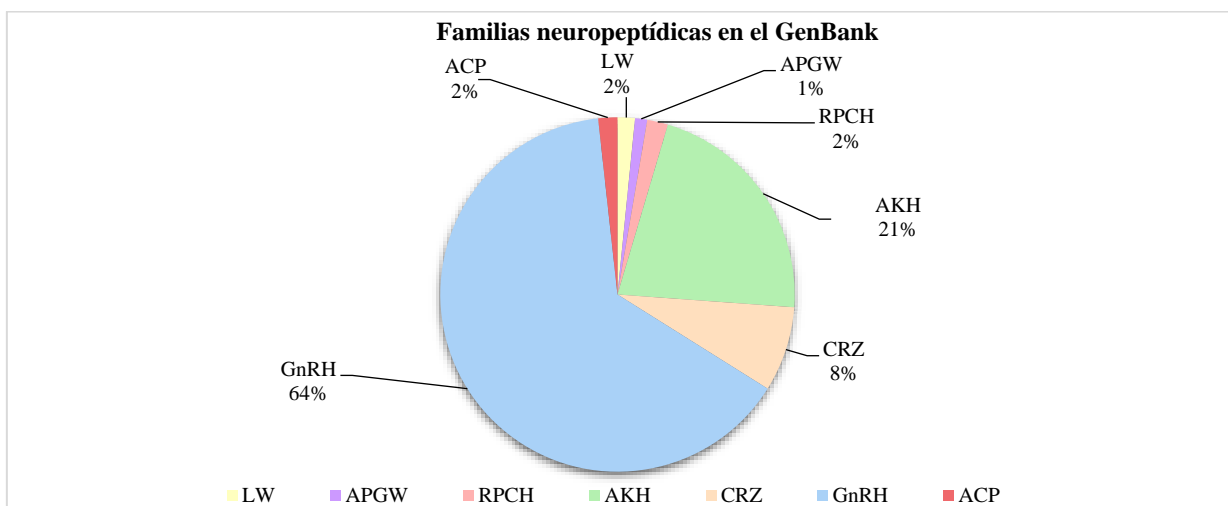


Figura 2. Porcentaje de las secuencias presentes en la base de datos del GenBank para las familias neuropéptidas LW, APGW, RPCH, AKH, ACP, CRZ y GnRH.

4.3 Determinación de la estructura primaria de los precursores neuropeptídicos

La determinación de la estructura primaria con la predicción del software ProP Server versión 1.0 generó resultados incongruentes, respecto a los otros sitios de corte endoproteolíticos. Aun cuando el software está diseñado para el reconocimiento de los sitios Arginina/Lisina (R/K) en ocasiones únicamente reconoció un sitio y no identificó a los otros.

Además, por su diseño no permitió el reconocimiento de los sitios de corte K/R – R/K – K/K – R/R, de los precursores neuropeptídicos evaluados y considero Prolina/Arginina (P/R) y Glicina/Serina (G/S) como sitios de corte (Apéndice F).

Debido a estos resultados se realizó la identificación y comparación manual con los sitios de corte R/K, producidos por el software (Apéndice G). Los precursores de la GnRH y la CRZ tuvieron una cantidad superior respecto a los identificados por el software. Además, las familias LW, APGW, AKH 1 y ACP fueron predichas con mayor número de cortes por el software. La familia AKH 3 con el software y en la depuración manual, mostró la misma cantidad y las familias RPCH, AKH 2 y AKH 4 generaron resultados falsos positivos, ya que el software mostró sitios de corte, pero la revisión manual mostró la ausencia de los sitios (Apéndice H).

La determinación manual de los sitios de corte endoproteolíticos mostró un total de 1465. La mayoría correspondieron a los sitios K/R, sitios R/K y R/R con valores similares y el sitio K/K fue minoritario. Las familias GnRH, APGW y AKH 1 tuvieron la mayor cantidad; la CRZ, la LW y la AKH 3 se consideraron intermedios respecto a la ACP y la RPCH; la AKH 2 y la AKH 4 que fueron minoritarias. Así mismo, la GnRH presentó los 4 cortes, seguida por la APGW, AKH 1, CRZ, LW y AKH 3; mientras que la familia RPCH presentó solo los sitios K/R, K/K y R/R; la familia AKH 4 presentó K/R y K/K y, la AKH 2 mostró solo K/R (Apéndice I).

Pese a lo anterior el software ProP Server, si permitió calcular el tamaño promedio de los precursores neuropeptídicos y el péptido señal. La longitud promedio de los precursores fue de 65 a 282 aminoácidos, observándose que LW y APGW correspondieron a los más extensos; mientras que ACP, CRZ y RPCH fueron intermedias, y las familias con los cortos son GnRH y AKH. Además, el péptido señal no presentó variaciones significativas entre las familias, ya que la longitud mínima y máxima fue 17 - 26 aminoácidos en promedio (Apéndice J).

Al corroborar de forma manual las secuencias precursoras resultantes de los tres softwares se confirmó que se obtuvieron 511 precursores donde: LW, APGW, RPCH y ACP, fueron minoritarias, CRZ fue intermedia con más del triple respecto a las anteriores y las 4 AKHs aumentaron 10 veces más y la GnRH cerca de 30 veces más (Apéndice K). En este sentido la mayor cantidad de cortes proteolítico correspondió a la LW y a la APGW, en tanto que el alto número de sitios para las AKHs y GnRH, fue debido a que el número de precursores fue mayor (Apéndice L).

4.4 Construcción de la base de datos y código de identificación

La base de datos de las 7 familias neuropeptídicas, tuvo las 511 secuencias neuropeptídicas distribuidas en 260 especies dentro de 70 grupos taxonómicos. Estas contuvieron los criterios indicados en la metodología. Además, se incluyó las 60 palabras claves utilizadas en el proceso de selección. También se mostró: a las 239 secuencias depuradas que se establecieron como idénticas con diferente código de acceso, parte del proceso de depuración y eliminación con la ejemplificación de 2268 códigos generados en la corroboración manual, que mostró que no son precursores neuropeptídicos y que 10 secuencias publicadas en artículos (Apéndice M, N y Ñ).

4.5 Secuencias precursoras obtenidas por transcriptómica

El análisis del transcriptoma de *C. toxotes* y *C. arcuatus* mostró las secuencias precursoras neuropeptídicas: la ACP y la CRZ, respectivamente. La secuencia del primer precursor neuropeptídico presentó en total 690 pb, donde la región 5' no traducida (UTR) tuvo 18 pb, el ORF fue de 306 pb compuesto por los codones para la secuencia de direccionamiento al RER, el péptido activo y el relacionado; la región 3' UTR fue de 366 pb (Fig. 3A). Por su parte, la secuencia de ADNc de *C. arcuatus* mostró un total de 234 pb. Cuyo extremo 5' UTR fue de 81 pb y el ORF mostro únicamente a la señal de direccionamiento al RER el péptido activo y parte del relacionado. No se determinó el codón de paro y la región 3' UTR (Fig. 3B).

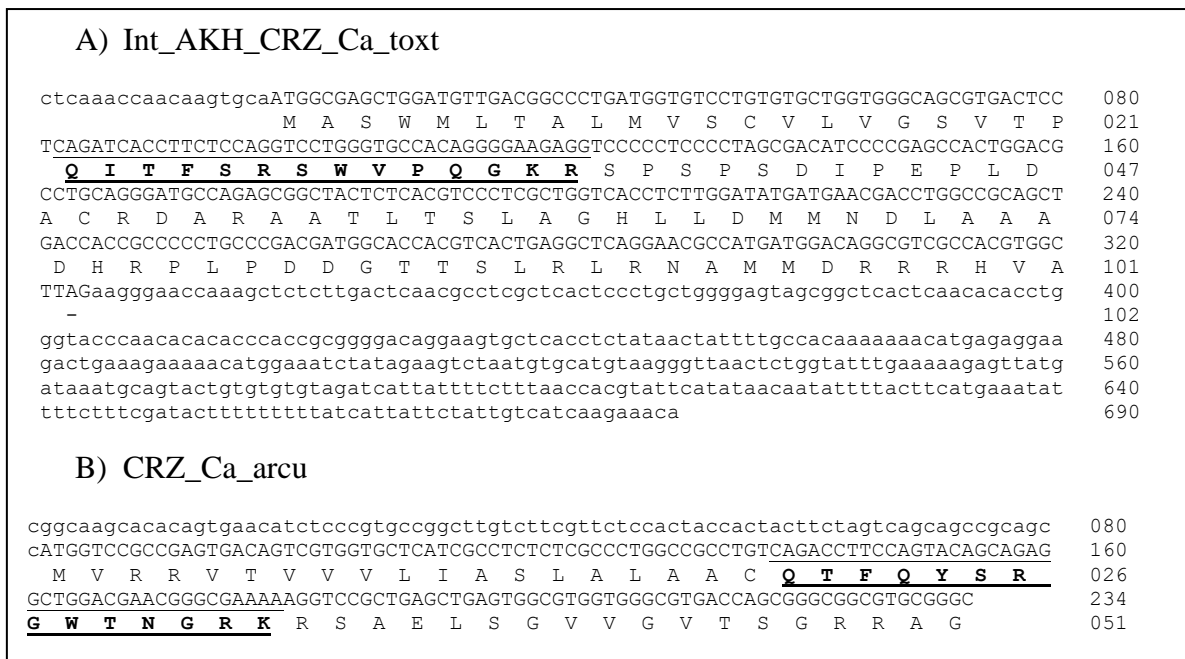


Figura 3. Secuencia híbrida AKH/CRZ (ACP) de *C. toxotes* y Corazonina (CRZ) de *C. arcuatus*. Se observa el código generado para los análisis bioinformáticos y en negrita el péptido activo subrayado inserto dentro del precursor neuropeptídico, en minúsculas las regiones no traducidas.

El alineamiento BLAST confirmó que la ACP de *C. toxotes* mostró una identidad del 76% respecto a la ACP de *Carcinus maenas* y del 37,0% al 44,6% con otros crustáceos. Mientras que el precursor de la CRZ de *C. arcuatus* tuvo homología con lepidópteros, dípteros, himenópteros y

quelicerados. Además, con secuencias reportadas sin caracterizar o hipotéticos de CRZs de los tardígrados *Ramazzottius varieornatus*, *Hypsibius dujardini*, y de los gasterópodos *Pomacea canaliculata* y *Biomphalaria glabrata* (Apéndice O). Ambas secuencias serán públicas GenBank en el año 2020.

4.6 BLAST péptidos virtuales APGW/RPCH

El alineamiento tipo BLAST mostró similitud para los seis péptidos virtuales APGW/RPCH descritos por el DNA-LM (Martínez-Pérez et al. 2007). Se obtuvieron 40 neuropéptidos en diferentes especies. La búsqueda con los precursores virtuales APGW/RPCH de *A. californica* se determinó que las regiones presentaron homología con el péptido activo y péptido relacionado del precursor de RPCH de *M. rosenbergii*. En tanto, que los precursores virtuales APGW/RPCH de *Lymnaea stagnalis* reportado de Molusco-Crustáceo fueron homólogos con la AKH 1 de los hemípteros *Plautia stali*, *Nezara viridula*, *Laodelphax striatellus*, *Nilaparvata lugens*, y en el tardígrado *H. dujardini*. Mientras que, el precursor Molusco-Crustáceo-Insecto mostró homología con las copias APGW. En tanto que los péptidos virtuales de *Mytilus edulis* y las copias de APGW de los precursores neuropeptídicos virtuales presentaron identidad con moluscos (Apéndice P).

El porcentaje de identidad de los 6 precursores virtuales fue mayor para las clases Gastropoda y Bivalvia junto con los órdenes Hymenoptera y Hemiptera; el *phylum* Platyhelminthes y el orden Lepidoptera mostro la mitad de la cantidad, mientras que los órdenes Blattodea y Passeriformes, el *subphylum* Crustacea y los *phyla* Rotifera y Tardigrada fueron minoritarios (Apéndice Q). Los péptidos virtuales APGW/RPCH completos no se determinados lo que sugiere que aún no han sido validados experimentalmente o no se ha reportado la especie que realiza su biosíntesis.

4.7 Alineamiento de secuencias y generación de secuencias precursoras consenso

Se realizaron 42 alineamientos de las secuencias precursoras neuropeptídicas que incluyeron las 6 secuencias de péptidos virtuales. La mitad correspondió al DNA-LM y la otra a los establecidos por el software Kaling (Lassmann et al., 2009) (Apéndice R). Los resultados corroboraron las secuencias consenso y mostraron que la LW y APGW comparten homología en sus péptidos activos, como postula el DNA-LM donde la Leucina que caracteriza a la LW no estuvo en APGW.

Dicho patrón continuó al incluir en el alineamiento a los precursores de la RPCH, AKH 1 y la AKH 2. Es de resaltar, que la Leucina no se determinó en la APGW y en las tres familias antes mencionadas. Además, se observó que los 3 primeros aminoácidos y la Fenilalanina se localizó en otra región río abajo, mientras que los últimos 4 y el Triptófano de la posición 8, que genera la actividad con su respectivo receptor RPCH o AKH, tuvieron posiciones comunes con una de las copias de los precursores LW y APGW. Por lo tanto, los resultados confirman el DNA-LM.

La inclusión de la AKH 3 y la AKH 4 en el alineamiento, mostró dos tipos de extremos amino; el primer grupo constituido por ambos precursores neuropéptidos y se caracterizó por mostrar 4 aminoácidos antes de la primera Glicina del péptido activo y fue mayoritario respecto al segundo que contuvo 6 en esta misma región. Las posiciones dos, tres y cuatro del péptido activo se compartieron con la RPCH y AKH 1, estando ausentes en la LW y APGW, lo cual ratifica el DNA-LM. Además, los últimos 4 aminoácidos presentaron un patrón similar al antes expuesto en algunas AKH 3 que mostraron un primero grupo que conservó sus posiciones respecto a la RPCH, AKH 1 y AKH 2, el segundo presentó tres aminoácidos posteriores al triptófano y anteriores a los sitios de corte. Todo lo anterior confirmó nuevamente el DNA-LM.

La distribución de aminoácidos para el péptido activo se conservó con la inclusión de la CRZ, ACP y GnRH. Todos los precursores junto con los péptidos virtuales APGW/RPCH, mostraron que estos últimos comparten una de las copias de APGW las posiciones en el extremo carboxilo con todas las familias más no con la región de RPCH virtual. En tanto, esta última no presentó identidad con respecto a la región de todas las familias. La región posterior a la señal de internalización al RER de los precursores virtuales fue diferente respecto a todas las familias y uno de los precursores virtuales de *M. edulis* que no presentó homología en esta región (Fig. 4).

A partir de los alineamientos de cada familia neuropeptídica se generaron 20 secuencias consenso, 10 con el software DNAMAN, con los parámetros del software y las otras empleadas como control con el formato ClustalW. El resultado mostró que la región del péptido activo y sus sitios de corte son más conservados respecto los a las demás partes del precursor neuropeptídico. La región más variable correspondió a la señal de direccionamiento de internalización al lumen del RER y que el extremo carboxilo fue menos conservado (Apéndice S).

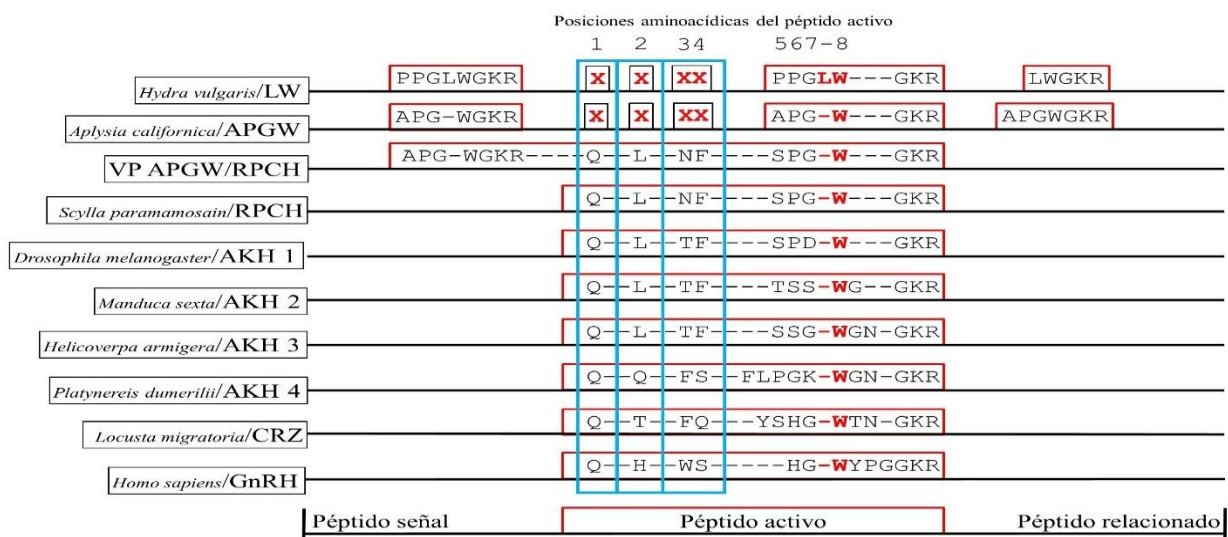


Figura 4. Representación gráfica de los alineamientos de la región del péptido activo generados por el DNA-LM. El alineamiento muestra en recuadros azules la variación de las posiciones en el extremo amino entre las familias. La X roja representan la ausencia de las posiciones 1 a 4 de la LW y APGW que corresponden al péptido relacionado y que generaron el extremo N terminal de las familias. La Leucina que se perdió para dar origen a la APGW y el triptófano que da la actividad se indican en rojo y negritas. Los péptidos activos se indican en cajas rojas.

4.8 Relaciones filogenéticas

Se realizaron 13 árboles filogenéticos que correlacionan las 7 familias neuropeptídicas. De estas, 11 corresponden a el agrupamiento por familia neuropeptídica y 2 con árboles concatenados, uno con los parámetros del software Kaling y el otro con los parámetros DNA-LM (Apéndice T). Los detalles de las filogenias de las familias LW, APGW, RPCH, AKHs CRZ y GnRH se detallan en el Apéndice U. Las relaciones de las familias ACP, péptidos virtuales APGW/RPCH y las dos filogenias concatenadas se describen a continuación.

La relación filogenética entre los péptidos virtuales APGW/RPCH confirmó la homología que presentaron respecto a otras especies (Fig. 5). A partir del tercer internodo se agrupó un clado APGW/RPCH virtual de *M. edulis* con las APGW de bivalvos y gasterópodos, es de resaltar, que estos precursores virtuales no presentan a la RPCH. En el cuarto internodo se generó un clado con APGW/RPCH virtual para gasterópodos, en los cuales si se presentaban la RPCH dentro de sus respectivos precursores. A partir de este se generó un quinto internodo en donde se agruparon las AKHs y RPCHs de platelmintos, rotíferos y tardígrados y en el sexto se agruparon los clados para las AKHs y RPCHs de himenópteros, lepidópteros, hemípteros, un crustáceo y un blatodeo.

Al concatenar las filogenias, con los parámetros del software Kaling y los del DNA-LM. Se determinó que ambas filogenias presentaron un clado para la mayoría de las GnRHs de deuterostomados y protostomados. La diferencia entre ellas es que la filogenia obtenida con los parámetros del software Kaling, se obtuvo un clado para 3 cefalópodos entre las AKH 1 y la RPCH; mientras que con el modelo DNA-LM se generó un clado compuesto por deuterostomados para las GnRH de Anfibios y otro entre las AKH 3 y las CRZ. Por otra parte, se observó una diferencia entre el patrón de distribución de ambas filogenias, ya que los nodos y clados para los protostomados, con los parámetros por defecto mostraron que en la raíz del árbol la familia LW

estuvo en un solo clado, la CRZ en otro el cual incluyó la AKH 2 y una AKH 4. El tercer clado correspondió a la AKH 1 que presentó a los péptidos virtuales APGW/RPCH. Posterior a este se encontró el clado de las RPCH que contuvo una AKH 3 y dos APGW; mientras que el último clado mostró a las ACP una AKH 1, AKH 4 y AKH 3 (Fig. 6A). En tanto, la filogenia con los parámetros del DNA-LM, cuatro LW se encontraron al inicio de la filogenia y posteriormente se generó un segundo clado compuesto por: cuatro LW, RPCHs, AKH 2-3, ACP y CRZ. El tercer clado estuvo formado por dos LW y AKH 2-3-4 y se insertó las GnRH de mamíferos. El cuarto clado fue compuesto por péptidos virtuales APGW/RPCH, AHK 2, y CRZ. El quinto contuvo AKH 1-3, una AKH 4 y una CRZ. El sexto clado mostró a los péptidos virtuales APGW/RPCH, AKH 1, CRZ, una ACP. Posteriormente, el séptimo nodo mostró APGW, AKH 1, una AKH 4, RPCH y CRZ, mientras que el octavo y último nodo fue exclusivo de RPCH, AKH 1, CRZ y las GnRH de anfibios (Fig. 6B).

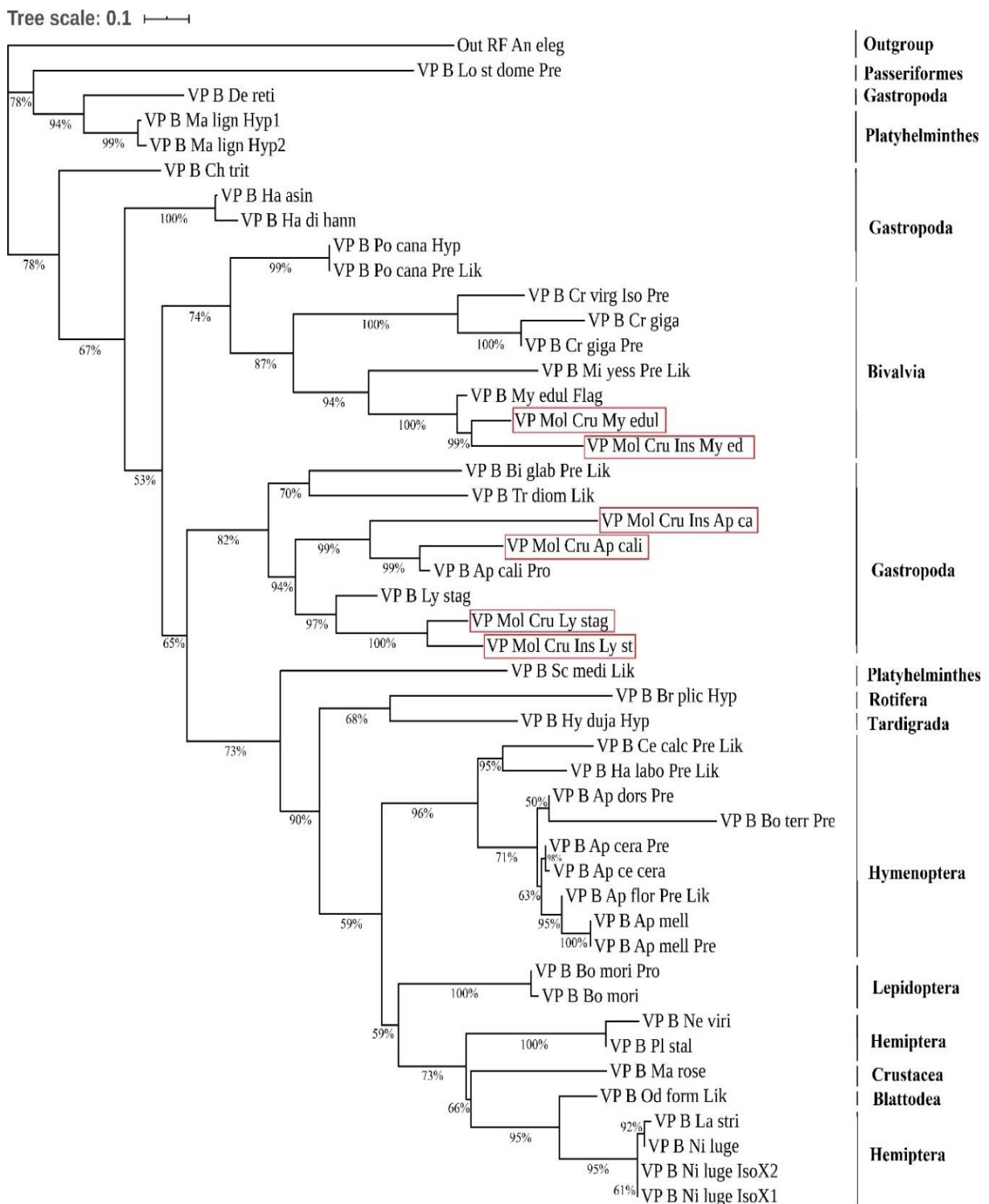


Figura 5. Árbol filogenético realizado con el software IQtree para las secuencias precursoras neuropeptídicas de los péptidos virtuales APGW/RPCH descritos por Martínez-Pérez et al. (2007) y las secuencias encontrados mediante BLAST. Se utilizó como *outgroup* el precursor neuropeptídico RFamida de *Anthopleura elegantissima*, los resultados se visualizaron mediante el software ItoI.

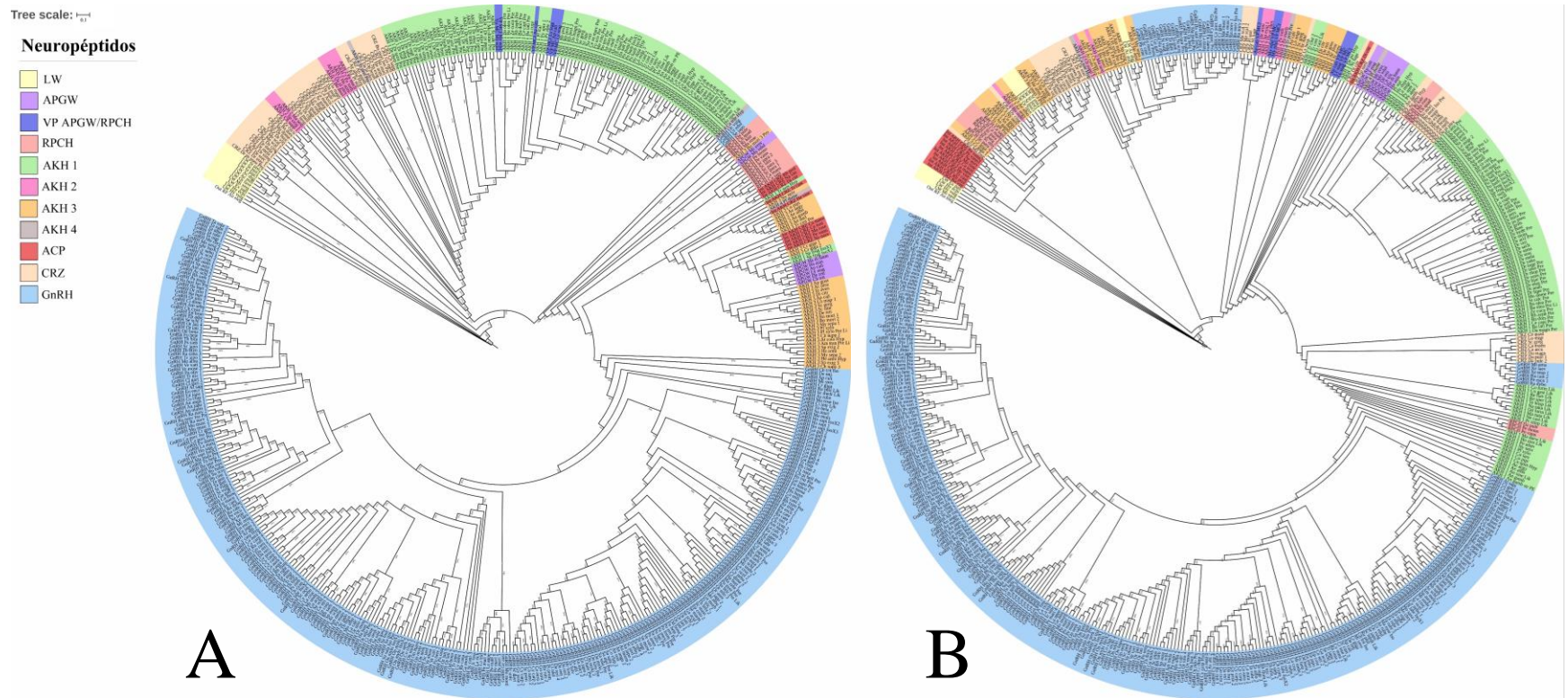


Figura 6. Árboles filogenéticos concatenados realizados mediante el software IQtree para las familias neuropeptídicas LW, APGW, péptidos virtuales APGW/RPCH, RPCH, AKHs, ACP, CRZ y GnRH. A) Muestra las relaciones filogenéticas con los parámetros del DNA-LM. B) Muestra las relaciones filogenéticas con los parámetros de alineamiento del software Kaling. Se utilizó como *outgroup* el precursor neuropeptídico Rfamida de *A. elegantissima*, los resultados se visualizaron mediante el software ItoI. Para ver detalles en alta resolución Apéndice V.

5. Discusión

La congruencia evolutiva entre protostomados y deuterostomados (Sherwood & Parker, 1990; Tsai & Zhang, 2008). ha planteado diversas hipótesis para el origen de los neuropéptidos (Derst et al., 2016; Semmens & Elphick, 2017), no obstante, cualquiera de ellas deberá de considerar la estructura y organización de los precursores neuropeptídicos y con ello poder comprender los eventos evolutivos que pudieron generar a sus familias. Por esta razón, en esta pasantía se contribuyó a la validación del DNA-LM con las familias neuropeptídicas LW, APGW, RPCH, AKHs, ACP, CRZ y GnRH mediante la confirmación de los precursores híbridos en especies de moluscos y crustáceos *in silico* junto a su identificación en el transcriptoma de las especies marinas *C. arcuatus* y *C. toxotes*, tal como se postuló en el año 2007 (Martínez-Pérez et al., 2007). El éxito de este estudio es debido a las tecnologías de NGS que permiten generar datos genómicos y transcriptómicos de especies que habitan distintos ecosistemas y que son reportadas en bases de datos públicas (Levy & Myers, 2016; Sayers et al., 2019). Esto ha permitido solucionar algunas de las incógnitas evolutivas en relación a: las características de los precursores de neuropéptidos inconclusos, la organización estructural de aquellos prepropéptidos que únicamente se conocían por su actividad o composición de aminoácidos y los que muestran una posible correlación evolutiva (Semmens & Elphick, 2017; Semmens et al., 2016; Zandawala et al., 2018).

Sin embargo, hay que destacar que no toda la información descrita en las bases de datos públicas son precursores de neuropéptidos con las características requeridas para su identificación (Rouillé et al., 1995; Steiner et al., 1980). Lo anterior, sumado a un desconocimiento de los neuropéptidos y sus precursores en nuevas especies, ha generado una incorrecta interpretación biológica y por ende análisis evolutivos cuestionables. Una de estas causas es el establecimiento inexacto de los parámetros bioinformáticos para el ensamble de las secuencias obtenidas por NGS, los cuales, no

permiten una correcta identificación de las características de los precursores neuropeptídicos (Foulkes et al., 2017; Kremer, Schneider, & Terstappen, 2005). Por lo tanto, es necesario generar nuevos enfoques para el almacenamiento y el intercambio de datos genómicos de precursores neuropeptídicos, para evitar que las bases de datos pierdan la capacidad de estructurar y almacenar la caracterización de nuevos prepropéptidos (Ekblom & Galindo, 2011). De mantenerse esta tendencia continuará el aumento de secuencias: descritas de baja calidad, incorrecta identificación y déficit en su caracterización (Pible, Hartmann, Imbert, & Armengaud, 2014). Por lo tanto, se incrementa el fracaso de recuperación de secuencias neuropeptídicas ocurrió con las relaciones evolutivas de los genes *ACP*, *AKH*, *CRZ* y *GnRH* (Plachetzki, Tsai, Kavanaugh, & Sower, 2016).

Al implementar los softwares BioDataToolKit y Proteios en la depuración de las secuencias precursoras obtenidas del GenBank con las palabras claves se garantizó un manejo eficiente y rápido de las 9489 identificadas y las generadas en la secuenciación transcriptómica de *C. arcuatus* y *C. toxotes*. Los resultados de la búsqueda en el GenBank con los criterios empleados, concuerdan con Hauser & Grimmelikhuijzen, (2014), sin la inclusión de receptores acoplados a proteínas G.

El software BioDataToolKit que se evaluó por Acuña et al. (2019) y en esta pasantía corroboró su funcionalidad, pero como indicó este autor, requiere optimizar el *script* y la reestructuración de comandos para acceder a una mayor cantidad de secuencias completas.

El software Proteios mostró una mayor eficiencia respecto al anterior para depurar las secuencias, debido a que su interfaz gráfica es más amigable y permitió observar paso a paso el filtrado mediante los sitios de corte proteolíticos. Proteios al igual BioDataToolKit plantean una solución eficaz de búsqueda y depuración de secuencias para usuarios sin conocimientos en programación.

La validación de los resultados de ambas herramientas con el software ProP Sever, el cual identifica a precursores neuropeptídico con los posibles sitios de escisión sitios K/R que reconoce la Furina (Duckert et al., 2004) pero su diseño no incluye a K/K, R/K y/o R/R indico que era necesario la verificación manual. Con ello se demostró que, de las 9489 secuencias obtenidas, únicamente el 5,4% correspondieron precursores neuropeptídicos. Esta diferencia ocurrió porque la identificación de secuencias genómicas y/o transcriptómicas es en función de un genoma de referencia o con las bases de datos disponibles. Por lo tanto, si la nueva secuencia presenta identidad y similitud con alguna de las regiones de un precursor neuropeptídico ya descrito será denominada con un nombre que indique su similitud más no implica que sea un precursor.

Por otra parte, los códigos de tipificación de secuencias aminoacídicas que se desarrollaron e implementaron en esta pasantía permitieron la correcta identificación de las secuencias precursoras neuropeptídicas, por ejemplo, la secuencias del GenBank contienen números de identificación y las versiones de las secuencias previamente implementadas. Ambos son los más usados en la identificación, pero al diferir en formato e implementación no cumplen con el papel fundamental facilitar la caracterización y la rápida identificación de las secuencias. Por el contrario, el código neuropeptídico diseñado en esta pasantía permite observar claramente las características del precursor, el tipo de neuropéptido, la especie y/o las características genotípicas.

Además, con la depuración de las secuencias, el cambio de la nomenclatura del GenBank a los códigos desarrollados en esta pasantía y su implementación en la hoja de cálculo de Excel, se confirmó que la mayor cantidad de precursores correspondió a organismos relacionados de forma directa o indirecta con actividades de investigación y/o comerciales (Sajwan et al., 2015; Steven et al., 2003). Lo anterior se demostró al establecer que la mayor cantidad de GnRH correspondió a mamíferos y peces. Mientras que la AKH y CRZ pertenecieron principalmente a insectos como

se demuestra en la genómica de especies del género *Drosophila* (Audsley, Down, & Isaac, 2014) y en las obtenidas en esta pasantía.

Este patrón de cantidad de precursores no se determinó para las especies marinas ya que la LW, la APGW y la RPCH fueron minoritarias respecto a las anteriores. En este sentido, la investigación y caracterización genómica y transcriptómica se ha centrado preferentemente en algunas especies modelo (Larhammar, 2009; Sajwan et al., 2015; Steven et al., 2003), que han acelerado la generación de conocimiento. Sin embargo, los estudios comparativos de especies no modelo complementarán la investigación de especies modelo e incluso llegarán a producir estudios más rigurosos con especies silvestres aún no descritas, como se corroboró en esta pasantía al demostrar que la expresión de la secuencia precursora híbrida AKH/CRZ-péptido relacionado (ACP) de *C. toxotes*, que habita de forma silvestre en el Océano Pacífico y que fue considerada en su momento exclusiva de insectos (Hansen et al., 2010), es expresada en crustáceos (Alexander et al., 2018; Suwansa-ard et al., 2016). Además, se ha descrito que los sistemas de señalización de AKH y ACP son parálogos, al surgir por una duplicación de genes en el linaje de artrópodos (Hauser & Grimmelikhuijzen, 2014). Sin embargo, después de esta duplicación, el sistema de señalización de ACP se ha perdido de forma independiente en otros artrópodos (Zandawala et al., 2018).

También se confirmó la importancia de la verificación *in silico* y la correlación de los grupos taxonómicos de sus familias neuropeptídicas, ya que al igual que lo reportado (Sakai et al., 2017; Tian et al., 2016) ciertos neuropeptidos se consideraban exclusivos de algunas especies están en grupos taxonómicos que no habían sido correlacionados. Por ejemplo, la GnRH del *subphylum* Vertebrata está en el *subphylum* Tunicata y los *phyla*: Echinodermata, Arthropoda, Mollusca y Platyhelminthes (Adams et al., 2003; Collins et al., 2010; Hasunuma & Terakado, 2013; Semmens et al., 2016; Suwansa-ard et al., 2016; Zhang, Tello, Zhang, & Tsai, 2008). Otro ejemplo, es la

CRZ que se consideró exclusiva de insectos, pero la transcriptómica de *C. arcuatus* confirmó su expresión como en otros crustáceos (T. V. Nguyen, Cummins, Elizur, & Ventura, 2016).

Se demostró que los péptidos virtuales reportados en año 2007 obtenidos por el DNA-LM (Martínez-Pérez et al., 2007), presentaron homología con especies que fueron consideradas en su momento y otras inesperadas, pero debido a: la accesibilidad de su hábitat, las características intrínsecas de sus secuencias nucleotídicas y/o aminoacídicas y los sistemas de secuenciación de ese entonces que hacían imposible validar el modelo. Lo anterior, se confirmó en esta pasantía con los precursores virtuales APGW/RPCH de *A. californica* y *L. stagnalis* (Martínez-Pérez et al., 2007), ya que la secuencia del RPCH virtual presentó homología con la proteína hipotética BV898_10396 del tardígrado *H. dujardini*, las RPCHs de crustáceos y algunas AKHs. Mientras que la región correspondiente a las copias de APGW fueron homólogas con el péptido activo e inclusive con el péptido señal y péptido relacionado de moluscos, lo que sugiere fuertemente que el modelo corresponde a un mecanismo evolutivo para la generación de precursores neuropeptídicos aun cuando el péptido virtual APGW/RPCH aún no ha sido descrito.

Las filogenias obtenidas para cada precursor confirmaron las relaciones entre los motivos de cada uno de ellos e inclusive la filogenia obtenida con los péptidos virtuales. Sin embargo, al concatenar todas las familias de neuropéptidos LW, APGW, péptidos virtuales APGW/RPCH, RPCH, AKHs, ACP, CRZ y GnRH en una sola filogenia, sea bien con los parámetros del programa Kaling o con DNA-LM. Se observó una marcada diferencia en el orden de los nodos para cada familia como lo propuso Tsai (2018), quien postuló la hipótesis de la superfamilia GnRH que presenta dos variantes: 1) La AKH de moluscos, artrópodos y nemátodos dio origen a las GnRH de moluscos y vertebrados junto con la CRZ de artrópodos. 2) La GnRH de vertebrados dio origen a la CRZ de artrópodos, está a su vez generó la GnRH de moluscos dividiéndose en la AKH de

moluscos y la AKH de artrópodos. Otra propuesta similar la realizó Plachetzki et al. (2016), al postular una superfamilia neuropeptídica compuesta por ACP, AKH, CRZ y GnRH (Plachetzki et al., 2016).

A diferencia de lo propuesto por estos autores el DNA-LM se fundamenta en la correlación de codones que generan motivos funcionales neuropeptídicos mientras que los alineamientos con los parámetros propuestos por el software Kaling son producto de la similitud aminoacídica. Por lo tanto, al no incluir los parámetros del DNA-LM para generar la matriz de alineamiento aminoacídica no es posible determinar la variación de los precursores a nivel de los motivos funcionales entre precursores que son aparentemente diferentes y han variado desde aproximadamente 500 millones en la explosión de Cámbrico.

Por lo anterior, los resultados de esta pasantía mostraron que los alineamientos y las filogenias obtenidas con el DNA-LM y con los parámetros establecidos por el software Kaling son similares en la división entre protostomados y deuterostomados con respecto a la GnRH, como lo propusieron Plachetzki et al., (2016) y Tsai, (2018). Sin embargo, es de resaltar que las conclusiones de ellos no incluyeron a: La LW, la APGW, la RPCH, los cuatro tipos de AKH y los péptidos virtuales APGW/RPCH que se generan a partir de la pérdida de codones de los precursores LW y APGW. Puesto que los alineamientos y las filogenias realizadas en esta pasantía se generaron únicamente con las secuencias aminoacídicas de los precursores neuropeptídicos, para poder afirmar o postular cualquier hipótesis y/o relación filogenética entre los neuropéptidos este estudio, se hace necesario e indispensable realizar el análisis de los genes de las especies de cada familia, ya que esto permitiría observar la pérdida de codones de los precursores LW y APGW para establecer cuál de las copias del péptido activo pudiese haber generado a la ACP, AKH 2 exclusiva de lepidópteros, AKH 3, AKH 4, CRZ y GnRH o inclusive si estas se originaron de la

misma copia del péptido activo u alguna otra, tal como se corroboró en esta pasantía observar la relación LW/APGW/RPCH/AKH 1, que se propuso originalmente en el DNA-LM.

Por lo tanto, se requiere la aplicación del DNA-LM para obtener los péptidos virtuales entre LW/AKH 3/CRZ/GnRH o diferentes alternativas que permitan observar la pérdida de codones las regiones y copias del péptido activo de la LW u otro precursor neuropeptídico que permita determinar la relación de origen entre la GnRH de protostomados y deuterostomados. Ya que los precursores neuropeptídicos muestran un alto grado de divergencia en sus secuencias aminoacídicas, y solo en pequeñas regiones de los genes como el péptidos activos o motivos específicos dentro de la secuencia de péptidos presentan actividad biológica las cuales son altamente conservadas (Liu, Baggerman, Schoofs, & Wets, 2006).

6. Conclusiones

La caracterización de las familias de neuropéptidos LW, APGW, APGW/RPCH, RPCH, AKH, ACP, CRZ y GnRH en especies filogenéticamente relacionadas y/o con otras sin interés comercial a partir de bases de datos públicas y de secuenciación genómica contribuye a la generación de nuevo conocimiento para establecer los procesos evolutivos en relación con los neuropéptidos que participan en la comunicación celular de protostomados y deuterostomados por medio del DNA-LM.

La aplicación de los elementos requeridos para el procesamiento postraducciona l de los precursores neuropeptídicos que generan el péptido activo de las familias son imprescindibles para su correcta identificación y selección a partir de bases de datos públicas con palabras claves y de proyectos de transcriptómica para la validación del DNA-LM y otros modelos evolutivos.

El cambio de códigos numéricos con formatos FASTA de secuencias de bases de datos por códigos que lo incluyen por otro código diseñado en función de: principios moleculares, celulares y procesamiento de precursores neuropeptídicos de las familias genera bases de datos confiables, accesibles y eficientes para el uso de la información reportada en el GenBank y/o de proyectos de transcriptómica aplicados a la generación de conocimiento evolutivo.

Los péptidos virtuales del precursor APGW/RPCH que dio origen al DNA-LM comparten homología con secuencias de moluscos y artrópodos, lo que indica su presencia en alguna especie no determinada u otra ya extinta.

El DNA-LM permite establecer las relaciones filogenéticas de los dominios funcionales aminoacídicos de los precursores neuropeptídicos entre las familias LW, APGW, RPCH, AKHs, ACP, CRZ y GnRH, pero es necesaria la inclusión de sus genes para establecer la pérdida o ganancia de codones que corroboren la evolución entre estas familias.

7. Recomendaciones

Aun cuando existen otras herramientas bioinformáticas que permiten el acceso y descarga de secuencias reportadas en bases de datos públicas, se recomienda que se empleen programas como BioDataToolKit y Proteios en la depuración de secuencias relacionadas a neuropéptidos para evitar el sesgo y la mala interpretación de los resultados.

El empleo de los códigos de identificación neuropeptídica desarrollados en esta pasantía en proyectos de minería de datos y estudios que involucren precursores neuropeptídicos para un mejor manejo, desarrollo y rendimiento de los metadatos.

La secuenciación masiva del transcriptoma de especies aún no correlacionadas con el DNA-LM generará una mejor comprensión de sus precursores neuropeptídicos y establecerá la presencia de los precursores para péptidos virtuales que se generan por la pérdida de codones.

El análisis filogenético de los genes de los precursores empleados en este estudio para consolidar o refutar el DNA-LM y las hipótesis propuestas por otros investigadores en relación con evolución de AKH, CRZ, ACP y GnRH o las propuestas en esta pasantía.

Referencias bibliográficas

- Acuña, C. I. (2019). *Evaluación de un programa para la generación de bases de datos con secuencias del año 2017 del gen de la Hemaglutinina del virus de influenza A H1N1* (Tesis de pregrado). Universidad Industrial de Santander, Bucaramanga, Colombia.
- Adams, B. A., Tello, J. A., Erchegeyi, J., Warby, C., Hong, D. J., Akinsanya, K. O., ... Sherwood, N. M. (2003). Six novel gonadotropin-releasing hormones are encoded as triplets on each of two genes in the protochordate, *Ciona intestinalis*. *Endocrinology*, *144*(5), 1907–1919.
- Ahn, S. J., Martin, R., Rao, S., & Choi, M. Y. (2017). Neuropeptides predicted from the transcriptome analysis of the gray garden slug *Deroceras reticulatum*. *Peptides*, *93*(2), 51–65.
- Alexander, J. L., Oliphant, A., Wilcockson, D. C., Audsley, N., Down, R. E., Lafont, R., & Webster, S. G. (2018). Functional characterization and signaling systems of corazonin and red pigment concentrating hormone in the green shore crab, *Carcinus maenas*. *Frontiers in Neuroscience*, *11*(752), 1–18.
- Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C., & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, *60*(5), 685–699.
- Audsley, N., Down, R. E., & Isaac, R. E. (2014). Genomic and peptidomic analyses of the neuropeptides from the emerging pest, *Drosophila suzukii*. *Peptides*, *68*(1), 33–42.
- Belzung, C., Yalcin, I., Griebel, G., Surget, A., & Leman, S. (2006). Neuropeptides in psychiatric diseases: An overview with a particular focus on depression and anxiety disorders. *CNS & Neurological Disorders - Drug Targets*, *5*(1), 135–145.
- Burbach, P. (2011). What are neuropeptides? In A. Merighi (Ed.), *Neuropeptides: Methods and protocols* (1st ed., pp. 1–36). Grugliasco, Italy: Humana Press.

- Choo, K. H., & Ranganathan, S. (2008). Flanking signal and mature peptide residues influence signal peptide cleavage. *BMC Bioinformatics*, *9*(12), 1–11.
- Christensen, M., Carlsen, J., & Josefsson, L. (1978). Structure-function studies on red pigment-concentrating hormone. The significance of the terminal residues. *Hoppe-Seyler's Zeitschrift Für Physiologische Chemie*, *813*(7), 813–814.
- Christensen, M., Carlsen, J., & Josefsson, L. (1979). Structure-function studies on red pigment-concentrating hormone, II. The significance of the C-terminal tryptophan amide. *Hoppe-Seyler's Zeitschrift Für Physiologische Chemie*, *360*(8), 1051–1060.
- Collins, J. J., Hou, X., Romanova, E. V., Lambrus, B. G., Miller, C. M., Saberi, A., ... Newmark, P. A. (2010). Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biology*, *8*(10), 1–21.
- De Loof, A., & Schoofs, L. (1990). Homologies between the amino acid sequences of some vertebrate peptide hormones and peptides isolated from invertebrate sources. *Comparative Biochemistry and Physiology*, *95*(3), 459–468.
- Demidyuk, I. V., Shubin, A. V., Gasanov, E. V., & Kostrov, S. V. (2010). Propeptides as modulators of functional activity of proteases. *Biomolecular Concepts*, *1*(3), 305–322.
- Derst, C., Dircksen, H., Meusemann, K., Zhou, X., Liu, S., & Predel, R. (2016). Evolution of neuropeptides in non-pterygote hexapods. *BMC Evolutionary Biology*, *16*(51), 1–10.
- Duckert, P., Brunak, S., & Blom, N. (2004). Prediction of proprotein convertase cleavage sites. *Protein Engineering, Design and Selection*, *17*(1), 107–112.
- Eipper, B. A., Stoffers, D. A., & Mains, R. E. (1992). The biosynthesis of neuropeptides: Peptide α -amidation. *Annual Review of Neuroscience*, *15*(1), 57–85.
- Eklblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology

of non-model organisms. *Heredity*, 107(1), 1–15.

Foulkes, A. C., Watson, D. S., Griffiths, C. E. M., Warren, R. B., Huber, W., & Barnes, M. R. (2017).

Research techniques made simple: Bioinformatics for genome-scale biology. *Journal of Investigative Dermatology*, 137(9), 163–168.

Gadagkar, S. R., Rosenberg, M. S., & Kumar, S. (2005). Inferring species phylogenies from multiple

genes: Concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 304(1), 64–74.

Gäde, G. (2009). Peptides of the adipokinetic hormone/red pigment-concentrating hormone family:

A new take on biodiversity. *Annals of the New York Academy of Sciences*, 1163(1), 125–136.

Gajewski, M., Leitz, T., Schloßherr, J., & Plickert, G. (1996). LWamides from Cnidaria constitute a

novel family of neuropeptides with morphogenetic activity. *Roux's Archives of Developmental Biology*, 205(5), 232–242.

Gierasch, L. M. (1989). Signal Sequences. *Biochemistry*, 28(3), 923–930.

Gospocic, J., Shields, E. J., Glastad, K. M., Lin, Y., Penick, C. A., Yan, H., ... Bonasio, R. (2017).

The neuropeptide corazonin controls social behavior and caste identity in ants. *Cell*, 170(4), 748–759.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New

algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321.

Hansen, K. K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M., ...

Grimmelikhuijzen, C. J. P. (2010). Discovery of a novel insect neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin hormonal systems. *Journal of Biological Chemistry*, 285(14), 10736–10747.

- Hasunuma, I., & Terakado, K. (2013). Two novel Gonadotropin-Releasing Hormones (GnRHs) from the urochordate ascidian, *Halocynthia roretzi*: Implications for the origin of Vertebrate GnRH isoforms. *Zoological Science*, *30*(4), 311.
- Hauser, F., & Grimmelikhuijzen, C. J. P. (2014). Evolution of the AKH/corazonin/ACP/GnRH receptor superfamily and their ligands in the Protostomia. *General and Comparative Endocrinology*, *209*(1), 35–49.
- Jex, A. R., Nejsum, P., Schwarz, E. M., Hu, L., Young, N. D., Hall, R. S., ... Gasser, R. B. (2014). Genome and transcriptome of the porcine whipworm *Trichuris suis*. *Nature Genetics*, *46*(7), 701–706.
- Jimenez-Gutierrez, S., Cadena-Caballero, C. E., Barrios-Hernandez, C., Perez-Gonzalez, R., Martinez-Perez, F., & Jimenez-Gutierrez, L. R. (2019). The crustaceans vitellogenin: A systematic and experimental analysis of their genes, genomes, mRNAs and proteins, perspective to next generation sequencing. *Crustaceana*, *92*(10), 1169–1205.
- Johnson, J. I., Kavanaugh, S. I., Nguyen, C., & Tsai, P. S. (2014). Localization and functional characterization of a novel adipokinetic hormone in the mollusk, *Aplysia californica*. *PLoS ONE*, *9*(8), 1–14.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.
- Kremer, A., Schneider, R., & Terstappen, G. C. (2005). A bioinformatics perspective on proteomics: Data storage, analysis, and integration. *Bioscience Reports*, *25*(2), 95–106.
- Kuroki, Y., Kanda, T., Kubota, I., Fujisawa, Y., Ikeda, T., Miura, A., ... Muneoka, Y. (1990). A molluscan neuropeptide related to the crustacean hormone, RPCH. *Biochemical and*

Biophysical Research Communications, 167(1), 273–279.

- Larhammar, D. (2009). Neuropeptides phylogeny and evolution. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (1st ed., pp. 873–880). San Diego: Oxford: Academic Press.
- Lassmann, T., Frings, O., & Sonnhammer, E. L. L. (2009). Kalign2: High-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Research*, 37(3), 858–865.
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(1), 256–259.
- Levy, S. E., & Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, 17(1), 95–115.
- Liu, F., Baggerman, G., Schoofs, L., & Wets, G. (2006). Uncovering conserved patterns in bioactive peptides in Metazoa. *Peptides*, 27(12), 3137–3153.
- Maddison, W. P., & Maddison, D. R. (2018). Mesquite: A modular system for evolutionary analysis. Version 3.61. <http://www.mesquiteproject.org>.
- Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., ... Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(1), 636–641.
- Mains, R. E., & Eipper, B. A. (2006). Peptides. In G. J. Siegel, W. R. Albers, S. T. Brady, & D. L. Price (Eds.), *Basic neurochemistry: Molecular, cellular and medical aspects* (7th ed., pp. 317–332). California, USA: Elsevier Academic Press.
- Martínez-Pérez, F., Durán-Gutiérrez, D., Delaye, L., Becerra, A., Aguilar, G., & Zinker, S. (2007). Loss of DNA: A plausible molecular level explanation for crustacean neuropeptide gene evolution. *Peptides*, 28(1), 76–82.

- Martínez-Pérez, F., Zinker, S., Aguilar, G., Valdés, J., & Aréchiga, H. (2005). Circadian oscillations of RPB gene expression in the eyestalk of the crayfish *Cherax quadricarinatus*. *Peptides*, 26(12), 2434–2444.
- Merighi, A. (2009). Neuropeptides and coexistence. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (1st ed., pp. 843–849). Turin, Italy: Academic Press.
- Minh, B. Q., Nguyen, M. A. T., & Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195.
- Mordue, W., & Stone, J. V. (1977). Relative potencies of locust adipokinetic hormone and prawn red pigment-concentrating hormone in insect and crustacean systems. *General and Comparative Endocrinology*, 33(1), 103–108.
- Nakayama, K. (1997). Furin: A mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochemical Journal*, 327(3), 625–635.
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274.
- Nguyen, T. V., Cummins, S. F., Elizur, A., & Ventura, T. (2016). Transcriptomic characterization and curation of candidate neuropeptides regulating reproduction in the eyestalk ganglia of the Australian crayfish, *Cherax quadricarinatus*. *Scientific Reports*, 6(38658), 1–19.
- Nielsen, H., Brunak, S., & Von Heijne, G. (1999). Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Engineering*, 12(1), 3–9.
- Nielsen, H., Engelbrecht, J., Brunak, S., & Von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*, 10(1), 1–6.

- Nowicki, M., Bzhalava, D., & Bała, P. (2018). Massively parallel implementation of sequence alignment with basic local alignment search tool using parallel computing in Java library. *Journal of Computational Biology*, 25(8), 871–881.
- Osada, M., & Treen, N. (2013). Molluscan GnRH associated with reproduction. *General and Comparative Endocrinology*, 181(1), 254–258.
- Patel, H., Orchard, I., Veenstra, J. A., & Lange, A. B. (2014). The distribution and physiological effects of three evolutionarily and sequence-related neuropeptides in *Rhodnius prolixus*: Adipokinetic hormone, corazonin and adipokinetic hormone/corazonin-related peptide. *General and Comparative Endocrinology*, 195(1), 1–8.
- Pible, O., Hartmann, E. M., Imbert, G., & Armengaud, J. (2014). The importance of recognizing and reporting sequence database contamination for proteomics. *EuPA Open Proteomics*, 3(1), 246–249.
- Plachetzki, D. C., Tsai, P. S., Kavanaugh, S. I., & Sower, S. A. (2016). Ancient origins of metazoan gonadotropin-releasing hormone and their receptors revealed by phylogenomic analyses. *General and Comparative Endocrinology*, 234(1), 10–19.
- Ritschard, E. A., Fitak, R. R., Simakov, O., & Johnsen, S. (2019). Genomic signatures of G-protein-coupled receptor expansions reveal functional transitions in the evolution of cephalopod signal transduction. *Proceedings of the Royal Society B: Biological Sciences*, 286(1897), 1–10.
- Robinson, N. L., & Goldsworthy, G. J. (1976). Adipokinetic hormone and the regulation of carbohydrate and lipid metabolism in a working flight muscle preparation. *Journal of Insect Physiology*, 23(1), 9–16.
- Roch, G. J., Busby, E. R., & Sherwood, N. M. (2011). Evolution of GnRH: Diving deeper. *General*

and Comparative Endocrinology, 171(1), 1–16.

- Rouillé, Y., Duguay, S. J., Lund, K., Furuta, M., Gong, Q., Lipkind, G., ... Steiner, D. F. (1995). Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides: The subtilisin-like proprotein convertases. *Frontiers in Endocrinology*, 16(4), 322–361.
- Sajwan, S., Sidorov, R., Stašková, T., Žaloudíková, A., Takasu, Y., Kodrík, D., & Zurovec, M. (2015). Targeted mutagenesis and functional analysis of adipokinetic hormone-encoding gene in *Drosophila*. *Insect Biochemistry and Molecular Biology*, 61(1), 79–86.
- Sakai, T., Shiraishi, A., Kawada, T., Matsubara, S., Aoyama, M., & Satake, H. (2017). Invertebrate gonadotropin-releasing hormone-related peptides and their receptors: An update. *Frontiers in Endocrinology*, 8(217), 1–11.
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 47(1), 94–99.
- Schoofs, L., De Loof, A., & Van Hiel, M. B. (2017). Neuropeptides as regulators of behavior in insects. *Annual Review of Entomology*, 62(1), 35–52.
- Semmens, D. C., & Elphick, M. R. (2017). The evolution of neuropeptide signalling: Insights from echinoderms. *Briefings in Functional Genomics*, 16(5), 288–298.
- Semmens, D. C., Mirabeau, O., Moghul, I., Pancholi, M. R., Wurm, Y., & Elphick, M. R. (2016). Transcriptomic identification of starfish neuropeptide precursors yields new insights into neuropeptide evolution. *Open Biology*, 6(2), 1–31.
- Sherwood, N. M., & Parker, D. B. (1990). Neuropeptide families: An evolutionary perspective. *Journal of Experimental Zoology*, 256(4), 63–71.
- Stamatiades, G. A., & Kaiser, U. B. (2018). Gonadotropin regulation by pulsatile GnRH: Signaling and gene expression. *Molecular and Cellular Endocrinology*, 463(1), 131–141.

- Steiner, D. F. (1998). The proprotein convertases. *Current Opinion in Chemical Biology*, 2(1), 31–39.
- Steiner, D. F., Patzelt, C., Chan, S. J., Quinn, P. S., Tager, H. S., Nielsen, D., ... Rubenstein, A. H. (1980). Formation of biologically active peptides. *Proceedings of the Royal Society of London - Biological Sciences*, 210(1178), 45–59.
- Steven, C., Lehnen, N., Kight, K., Ijiri, S., Klenke, U., Harris, W. A., & Zohar, Y. (2003). Molecular characterization of the GnRH system in zebrafish (*Danio rerio*): Cloning of chicken GnRH-II, adult brain expression patterns and pituitary content of salmon GnRH and chicken GnRH-II. *General and Comparative Endocrinology*, 133(1), 27–37.
- Stone, J. V, Mordue, W., Batley, K. E., & Morris, H. R. (1976). Structure of locust adipokinetic hormone, a neurohormone that regulates lipid utilisation during flight. *Nature*, 263(5574), 207–211.
- Suwansa-ard, S., Zhao, M., Thongbuakaew, T., Chansela, P., Ventura, T., Cummins, S. F., & Sobhon, P. (2016). Gonadotropin-releasing hormone and adipokinetic hormone/corazonin-related peptide in the female prawn. *General and Comparative Endocrinology*, 236(1), 70–82.
- Tian, S., Zandawala, M., Beets, I., Baytemur, E., Slade, S. E., Scrivens, J. H., & Elphick, M. R. (2016). Urbilaterian origin of paralogous GnRH and corazonin neuropeptide signalling pathways. *Scientific Reports*, 6(28788), 1–7.
- Tsai, P. S. (2018). Gonadotropin-releasing hormone by any other name would smell as sweet. *General and Comparative Endocrinology*, 264(1), 58–63.

- Tsai, P. S., & Zhang, L. (2008). The emergence and loss of gonadotropin-releasing hormone in Protostomes: Orthology, phylogeny, structure, and function. *Biology of Reproduction*, *79*(5), 798–805.
- Veenstra, J. A. (1989). Isolation and structure of corazonin, a cardioactive peptide from the American cockroach. *FEBS Letters*, *250*(2), 231–234.
- Von Heijne, G. (1985). Signal sequences: The limits of variation. *Journal of Molecular Biology*, *184*(1), 99–105.
- Von Heijne, G. (1990). The signal peptide. *The Journal of Membrane Biology*, *115*(3), 195–201.
- Yoshida, Y., Koutsovoulos, G., Laetsch, D. R., Stevens, L., Kumar, S., Horikawa, D. D., ... Arakawa, K. (2017). Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. *PLoS Biology*, *15*(7), 1–40.
- Zandawala, M., Tian, S., & Elphick, M. R. (2018). The evolution and nomenclature of GnRH-type and corazonin-type neuropeptide signaling systems. *General and Comparative Endocrinology*, *264*(1), 64–77.
- Zhang, L., Tello, J. A., Zhang, W., & Tsai, P. S. (2008). Molecular cloning, expression pattern, and immunocytochemical localization of a gonadotropin-releasing hormone-like molecule in the gastropod mollusk, *Aplysia californica*. *General and Comparative Endocrinology*, *156*(2), 201–209.