



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

LICENCIATURA EN TECNOLOGÍAS PARA LA INFORMACIÓN EN CIENCIAS

Escuela Nacional de Estudios Superiores,
Unidad Morelia

Análisis Filogenético y Estructural de Celulasas
de la Familia GH9 en Insectos

TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN TECNOLOGÍAS PARA LA INFORMACIÓN EN
CIENCIAS

P R E S E N T A

MARIO ADRIÁN DÍAZ CARRILLO

DIRECTORA DE TESIS: DRA. DRA. NANCY CALDERÓN CORTÉS
CO-DIRECTORA DE TESIS: DRA. ALICIA LARA MÁRQUEZ

MORELIA, MICHOACÁN

MAYO, 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

LICENCIATURA EN TECNOLOGÍAS PARA LA INFORMACIÓN EN CIENCIAS

Escuela Nacional de Estudios Superiores,
Unidad Morelia

Análisis Filogenético y Estructural de Celulasas
de la Familia GH9 en Insectos

TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN TECNOLOGÍAS PARA LA INFORMACIÓN EN
CIENCIAS

P R E S E N T A

MARIO ADRIÁN DÍAZ CARRILLO

DIRECTORA DE TESIS: DRA. DRA. NANCY CALDERÓN CORTÉS
CO-DIRECTORA DE TESIS: DRA. ALICIA LARA MÁRQUEZ

MORELIA, MICHOACÁN

MAYO, 2019



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
ESCUELA NACIONAL DE ESTUDIOS SUPERIORES, UNIDAD MORELIA
SECRETARÍA GENERAL
SERVICIOS ESCOLARES

MTRA. IVONNE RAMÍREZ WENCE

DIRECTORA

DIRECCIÓN GENERAL DE ADMINISTRACIÓN ESCOLAR

PRESENTE

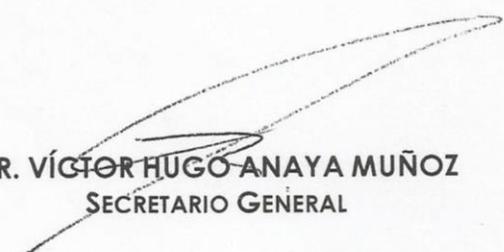
Por medio de la presente me permito informar a usted que en la **sesión ordinaria 12** del **H. Consejo Técnico** de la Escuela Nacional de Estudios Superiores (ENES) Unidad Morelia celebrada el día **05 de diciembre del 2018**, acordó poner a su consideración el siguiente jurado para la presentación del Trabajo Profesional del alumno (a) **Mario Adrián Díaz Carrillo** de la Licenciatura en **Tecnologías para la Información en Ciencias**, con número de cuenta **415071002**, con la tesis titulada: "Análisis Filogenético y Estructural de Celulasas de la Familia GH9 en Insectos." bajo la dirección como **tutora** de la Dra. Nancy Calderón Cortés y como **co-tutora** la Dra. Alicia Lara Márquez.

El jurado queda integrado de la siguiente manera:

Presidente:	Dr. Ernesto Vicente Vega Peña
Vocal:	Dr. César Andrés Torres Miranda
Secretario:	Dra. Nancy Calderón Cortés
Suplente 1:	Dr. Diushi Keri Corona Santiago
Suplente 2:	Dr. Hernando Alonso Rodríguez Correa

Sin otro particular, quedo de usted.

Atentamente
"POR MI RAZA HABLARA EL ESPIRITU"
Morelia, Michoacán a, 22 de abril del 2019.


DR. VÍCTOR HUGO ANAYA MUÑOZ
SECRETARIO GENERAL

CAMPUS MORELIA

Apartado Postal 27-3 (Santa Ma. De Guido), 58090, Morelia, Michoacán
Antigua Carretera a Pátzcuaro N° 8701, Col. Ex Hacienda de San José de la Huerta
58190, Morelia, Michoacán, México. Tel: (443)689.3500 y (55)56.23.73.00, Extensión Red UNAM: 80614
www.enesmorelia.unam.mx

El presente trabajo se realizó en el Laboratorio de Biología Molecular de la Escuela Nacional de Estudios Superiores Unidad Morelia de la UNAM.

Investigación realizada con financiamiento del Fondo Sectorial SEP-CONACYT. Proyecto CB 2015/253420: “Origen y evolución de los mecanismos endógenos de digestión de celulosa en insectos”.

Agradecimientos Institucionales

Agradezco a la Universidad Nacional Autónoma de México y a la Escuela Nacional de Estudios Superiores Unidad Morelia, así como a la carrera de licenciatura de Tecnologías para la Información en Ciencias por todas las enseñadas, conocimientos y apoyos aportados a mi persona.

A la UNAM por la beca alimenticia y la beca de Titulación por proyecto de Investigación recibidas durante la realización de mis estudios.

A la Dra. Nancy Calderón Cortés, por dejarme participar en este proyecto de investigación el cual me permitió poner a prueba los conocimientos adquiridos en la carrera, además de obtener nuevos conocimientos y habilidades. Del mismo modo agradezco por la paciencia que junto con la Dra. Alicia Lara Márquez tuvieron para enseñarme a lo largo de este proceso de investigación.

Al departamento de Servicio Escolares, que siempre atendieron mis dudas y me brindaron su apoyo para realizar distintos trámites.

Al Jurado conformado por:

Dr. Ernesto Vicente Vega Peña
Dr. César Andrés Torres Miranda
Dra. Nancy Calderón Cortés
Dr. Diushi Keri Corona Santiago
Dr. Hernando Alonso Rodríguez Correa

Agradecimientos Personales

A mis primas Andrea e Itzel, tía Miriam, mi hermano Miguel y mi Abuela por darme ánimos de seguir adelante a pesar de que en varias ocasiones pensaba que la carrera no era para mí.

Al Dr. Víctor Anaya por motivarme con su consejo de seguir en la carrera.

A los profesores de la Licenciatura Cesar Torres, Marisol Flores, Sergio Tinoco, Adriana Menchaca, Luis Miguel y Danielle Colossi por todo el conocimiento adquirido y apoyo recibido.

Al Dr. Diushi Corona por su ayuda en el manejo de algunas herramientas bioinformáticas usadas para este estudio.

A mi amigo Damián García por apoyarme en todo momento.

A mis amigos Melisa Patiño, Ángel Uriostegui y Daniel Delgado por apoyarme en momentos de desesperación, ayudándome a seguir adelante.

A mis amigos en la carrera Jorge Quirós, Gerardo Rodríguez, Karla Macedo y Paola Guzmán, Vianey López y Sonia Miranda por estar a mi lado.

A mis amigos Lesslie Pardavé y Edgar de la Cruz por ayudarme con cuestiones de redacción.

Dedicatoria

El esfuerzo realizado para llevar a cabo esta tesis lo dedico con todo el amor a mis padres Mario Díaz Cárdenas y Adriana Esmeralda Carrillo Martínez, dado que sin su apoyo, sacrificio y educación, esta tesis no se podría haber realizado.

Resumen

La celulosa es el biopolímero más abundante del planeta y el principal componente de la pared celular de plantas y algas, por lo que representa una fuente de alimento para una gran cantidad de organismos, entre los cuales se encuentran los insectos. Para llevar a cabo la degradación de celulosa se requiere de enzimas especializadas denominadas celulasas, estas se clasifican en diferentes familias de enzimas Glycosil Hidrolasas (GH), de las cuales solo las familias GH5, GH9, GH45 y GH48 están presentes en insectos. Las celulasas de la familia GH9 tienen una mayor representación y se encuentran en un gran número de taxa de insectos tales como: Embioptera, Phasmatodea, Coleoptera, Hemiptera, Termitoidae, Hymenoptera, Blaberoidea, Orthoptera y Phthiraptera. Además, estas enzimas se encuentran presentes en otros grupos de animales invertebrados, así como en plantas, bacterias y hongos lo cual sugiere un origen ancestral para estas enzimas. Estudios previos proponen un origen monofilético para las celulasas de la familia GH9 en insectos, sin embargo, no existen estudios actuales que incorporen el número creciente de secuencias de estas enzimas disponibles en bases de datos, ni análisis estructurales. Por lo anterior, en el presente estudio se analizó el origen, evolución y estructura de celulasas de la familia GH9 en insectos, para lo cual se realizó una búsqueda de secuencias de proteínas provenientes de metazoarios protostomados en bases de datos. A partir de las secuencias obtenidas se realizaron análisis filogenéticos con los métodos de Máxima Verosimilitud y Bayesiano, usando diferentes grupos externos. Para llevar a cabo los análisis estructurales se generaron distintos modelos de estructuras terciarias generados a partir de secuencias representativas de los clados obtenidos en las filogenias, mediante modelado por homología, con los cuales se realizó un análisis de sobreposición estructural. Posteriormente se realizó un análisis de probables sitios de unión a iones metálicos y un análisis de Evolutionary Trace para identificar diferencias en la estructura primaria y terciaria entre los grupos de enzimas de insectos. Finalmente, se realizaron análisis de electrostática e hidrofobicidad de algunas enzimas representativas de insectos. El

proceso de búsqueda y depuración de secuencias en dos fases permitió identificar 349 secuencias con un tamaño mayor a 440 aminoácidos, provenientes de insectos y otros protostomados. La clase Insecta incluye secuencias de diversos órdenes detritívoros y herbívoros como: Orthoptera, Phasmatodea, Blaberoidea, Termitoidae, Hemiptera, Coleoptera, Embioptera e Hymenoptera, aunque también incluye ordenes de insectos hematófagos tal como el orden Phthiraptera. Los análisis filogenéticos (Bayesiano y Máxima Verosimilitud) resultaron en árboles con topologías similares en los 6 casos y en todos ellos se reconocen dos grupos de insectos: uno conformado por secuencias del super orden Polyneoptera (grupo 1) y otro conformado por los super ordenes Polyneoptera, Holometabola y Paraneoptera (grupo 2). En el análisis de sobreposición no se formaron clados correspondientes a las filogenias, ni muestran patrones relacionados con el hábitat o dieta de los organismos. El análisis de los probables sitios de unión a iones metálicos reveló que estos se presentan de manera irregular en las secuencias de celulasas de la familia GH9. El análisis de evolutionary trace indica la presencia de una mutación en la posición 393 de la estructura cristalizada de *Nasutitermes takasagoensis* presente en las secuencias provenientes del grupo 2. Por su parte el análisis de hidrofobicidad reveló las diferencias presentes en la cavidad de unión al sustrato, dado que las enzimas del grupo 2 presentan una cavidad más larga y una hendidura con propiedades hidrofóbicas. Mientras que el análisis de electrostática reveló que la zona con mayor carga electronegativa está presente en la zona de la cavidad. Estos resultados confirman el origen monofilético de las enzimas de la familia GH9 en eumetazoarios protostomados y muestran la existencia de dos clados de estas enzimas en insectos, sugiriendo un evento de duplicación de genes con una posible neofuncionalización de las enzimas del grupo 2 de insectos.

Abstract

The cellulose is the most abundant biopolymer in the planet and the principal component of the cellular wall in plants and seaweed. It represents a food source to several organisms, including the insects. To carry out the cellulose degradation some specialized enzymes are required, these enzymes are denominated cellulases, and they are classified in different Glycosyl Hydrolases (GH) families, of which only the GH5, GH9, GH45 and GH48 families are present in insects. The cellulases of the GH9 family have a wide representation in insects, including taxa such as Hymenoptera, Blaberoidea, Orthoptera y Phthiraptera. Furthermore, these enzymes are present in other groups of invertebrate animals and other organisms such as plants, bacteria and fungus. This suggest an ancestral origin for these enzymes. Accordingly, previous studies suggest a monophyletic origin of the GH9 cellulases in insects, however there are no recent studies which incorporate the growing number of cellulases sequences available in databases or complement the studies with structural analysis. Based on this, the present study analyzed the origin, evolution and structure of the GH9 family cellulases in insects, for which a search of protein sequences in metazoan protostomes in databases was made. With the obtained sequences the phylogenetics analysis using Maximum Likelihood and Bayesian methods were performed, using different outgroups. To carry out the structural analyzes, different models of tertiary structures were generated from representative sequences of the clades obtained in the phylogenies by homology modelling. Thereafter, an overlap of the structures was made to compare the structural model. Subsequently, an analysis of probable metal ion binding sites and an Evolutionary Trace analysis were performed to identify the differences in the primary and tertiary structure between insect enzyme groups. Finally, electrostatic and hydrophobicity analysis of some enzymes representative of insects were carried out. The process of searching and debugging sequences in two phases allowed the identification of 349 sequences from insects and other protostomes with a length bigger than 440 amino acids. The

Insecta class includes sequences of various orders of detritivores and herbivores such as Orthoptera, Phasmatodea, Blaberoidea, Termitoidae, Hemiptera, Coleoptera, Embioptera and Hymenoptera, although it also includes orders of hematophagous insects such as the order Phthiraptera. The six phylogenetic analyzes (Bayesian and Maximum Likelihood) performed, resulted in trees with similar topologies that recognized two groups of insect cellulases: one conformed with sequences of the superorder Polyneoptera (group 1) and the other one conformed by sequences of the super orders Polyneoptera, Holometabola and Paraneoptera (group 2). In the analysis of structure overlap, the generated clades do not correspond to the ones generated in the phylogenetic analysis, neither to their habitat nor to the diet of the organisms. The analysis of the putative binding sites of metal ions revealed that predicted sites are irregularly distributed across the cellulase sequences of the GH9 family. The evolutionary trace analysis indicates the presence of a mutation at position 393 of the crystallized structure of *Nasutitermes takasagoensis* present in the sequences from the group 2 of insects. On the other hand, the hydrophobicity analysis revealed the differences present in the substrate-binding cavity in the enzymes of the group 2, have a longer cavity and a cleavage with hydrophobic properties than enzymes from group 1. The electrostatic analysis revealed that the zone with the highest electronegative charge is present in the cavity area. These results confirm the monophyletic origin of the enzymes of the GH9 family in eumetazoan protostomes and show the existence of two clades of GH9 enzymes in insects, suggesting an event of gene duplication with a possible neofunctionalization event in the insect group 2 enzymes.

Índice

1. Introducción	1
2. Justificación.....	5
3. Hipótesis	7
4. Objetivos	7
4.1 Objetivo general	7
4.2 Objetivos específicos.....	7
5. Antecedentes	8
5.1 Estructura y composición química de la Pared Celular Vegetal	8
5.1.1 Celulosa.....	10
5.2 Degradación de la celulosa y clasificación de las celulasas	11
5.3 Celulasas en Animales.....	14
5.4 Estructura de las celulasas de la familia GH9	18
5.5 Evolución de animales e insectos.....	20
6. Metodología	25
6.1 Búsqueda de secuencias.....	25
6.2 Alineamiento de secuencias.....	26
6.3 Selección de secuencias	27
6.4 Análisis filogenéticos.....	28
6.4.1 Modelo Evolutivo	28
6.4.2 Método Bayesiano	28
6.4.3 Método de Máxima Verosimilitud.....	29
6.5 Análisis estructural	29
6.5.1 Modelado de proteínas	29
6.5.2 Predicción de sitios de unión a iones	32
6.5.3 Análisis evolutionary trace de celulasas de insectos	33
6.5.4 Análisis de hidrofobicidad, potencial electrostático y cavidad de celulasas de insectos.....	34
7. Resultados	35
7.1 Búsqueda de secuencias.....	35
7.2 Alineamiento.....	38
7.3 Análisis filogenéticos.....	39
7.3.1 Árbol Filogenético 1.....	40

7.3.2	Árbol Filogenético 2.....	43
7.3.3	Árbol Filogenético 3.....	45
7.4	Análisis estructural	47
7.5	Análisis de sitios de unión a iones	51
7.6	Análisis estructural de celulasas de insectos.....	53
7.6.2	Análisis de hidrofobicidad y cavidad	55
8.	Discusión	59
8.1	Origen y evolución de celulasas GH9 en animales.....	59
8.2	Origen y evolución de celulasas de insectos.....	63
8.3	Análisis estructural	66
9.	Conclusiones	71
	Referencias	72
	Anexo 1.....	87
	Anexo 2.....	96
	Script 1 Leer y eliminar secuencias repetidas	96
	Script 2 Información taxonomica	101
	Script 3 Generar modelos de estructura terciaria.....	104
	Script 4 Minimizacion de energía método ligero.....	107
	Script 5 Minimizacion de energía método fuerte.....	108
	Script 6 Eliminacion de moléculas.....	109
	Script 7 Grafico de Ramachandran	110
	Script 8 Analisis estructural RMSD.....	113
	Script 9 RaptorX ID´s	114
	Script 10 RaptorX Informacion de grupos prostéticos	116
	Script 11 Filogenia con información de los sitios de unión a iones metálicos....	117
	Script 12 Secuencias con mutaciones específicas.....	120
	Anexo 3.....	121
	Anexo 4.....	137
	Grupos Taxonómicos Comprimidos del Árbol 1.....	137
	Grupos Taxonómicos Comprimidos del Árbol 2.....	141
	Grupos Taxonómicos Comprimidos del Árbol 3.....	148

Índice de tablas

Tabla 1 Celulasas de la familia GH9 de animales protostomados	16
Tabla 2 Secuencias de usadas para realizar la búsqueda de secuencias.....	25
Tabla 3 Secuencias usadas para generar los modelos de la estructura terciaria.....	29
Tabla 4 Secuencias de celulasas de animales protostomados obtenidas de la búsqueda realizada en el GenBank.....	35
Tabla 5 Secuencias de celulasas con mutaciones o deleciones en los sitios catalíticos.....	38
Tabla 6 Resultados de los análisis del grafico de Ramachandran	47

Índice de figuras

Figura 1 Estructura a distintas escalas de la pared celular.....	9
Figura 2 Pared celular vegetal	9
Figura 3 Estructura de la Celulosa.....	11
Figura 4 Esquema de la degradación de las celulasas de la familia GH9 sobre fibras de celulosa.....	12
Figura 5 Estructura de los distintos clanes de celulasas	14
Figura 6 Filogenia de los principales grupos metazoarios	21
Figura 7 Vista consenso de las relaciones filogenéticas de los animales metazoarios	22
Figura 8 Relaciones filogenéticas de Hexápoda	24
Figura 9 Filogenia de celulasas de la familia GH9 utilizando el alineamiento 1.....	42
Figura 10 Filogenia de celulasas de la familia GH9 utilizando el alineamiento 2.....	44
Figura 11 Filogenia de celulasas de la familia GH9 utilizando el alineamiento 3	46
Figura 12 Sobreposición de modelos generados por homología	49
Figura 13 Dendograma de los valores de RMSD de los modelos	50
Figura 14 Resultados del análisis de los sitios de unión a calcio y zinc	52
Figura 15 Filogenia resultado del análisis realizado en Evolutionary Trace	54
Figura 16 Resultados de Evolutionary Trace	55
Figura 17 Análisis de hidrofobicidad superficial de celulasas de la familia GH9.....	56
Figura 18 Potencial electrostático de celulasas de la familia GH9.....	58

1. Introducción

La mayor fuente de energía del planeta se encuentra almacenada en los tejidos de plantas y algas que representan organismos fotosintéticos. Por lo que estos tejidos son usados por un gran número de organismos como fuente de alimento, los cuales transforman los polisacáridos presentes en la pared celular de los tejidos vegetales a azúcares simples (Bayer et al. 1998, Sarkar et al. 2009, Wei et al. 2009, Gilbert 2010, Popper et al. 2011).

Entre los compuestos presentes en la pared celular, la celulosa representa la molécula más abundante del planeta. Este polisacárido estructural está compuesto por unidades de glucosa unidas por enlaces β -1,4 glucosídicos que forman cadenas lineales que se unen entre sí y con los demás componentes de la pared celular, para formar microfibras de celulosa a través de puentes de hidrógeno y enlaces de Van der Waals (Benziman et al. 1980, Heredia et al. 1995, Bayer et al. 1998, Santoyo et al. 1999, Wilson et al. 1999, Reiter 2002, Kim et al. 2008, Caffall & Mohnen 2009, Sarkar et al. 2009, Wei et al. 2009, Gilbert 2010). No obstante, aunque su estructura pueda parecer simple, las microfibras de celulosa presentan una estructura altamente ordenada y compacta que comprende regiones tanto cristalinas como amorfas, lo cual hace que la celulosa sea insoluble y difícil de degradar (Benziman et al. 1980, Santoyo et al. 1999, Wilson et al. 1999, Reiter 2002, Caffall & Mohnen 2009, Wei et al. 2009, Gilbert 2010).

En este sentido, aun cuando la celulosa es difícil de degradar, existen organismos que son capaces de secretar enzimas especializadas llamadas celulasas, las cuales degradan la celulosa a azúcares simples (Prins & Kreulen 1991, Bayer et al. 1998, Santoyo et al. 1999, Vries et al. 2001, Sugimura et al. 2003, Danchin et al. 2010, Gilbert 2010, Calderón-Cortés et al. 2010, Shelomi et al. 2014). Este conjunto de enzimas incluye: endo- β -glucanasas, celobiohidrolasas o exoglucanasas y β -glucosidasas (Bayer et al. 1998, Santoyo et al. 1999, Vries et

al. 2001, Wei et al. 2009). Las endoglucanasas rompen los enlaces β -1,4-glucosídicos de las regiones amorfas de la celulosa; mientras que las celobiohidrolasas actúan sobre los enlaces de las regiones cristalinas de la celulosa, produciendo oligosacáridos, que finalmente son hidrolizados por las β -glucosidasas para producir unidades de glucosa (Bayer et al. 1998, Vries et al. 2001, Gilbert 2010, Calderón-Cortés et al. 2012).

Las celulasas y otras enzimas involucradas en la degradación de la pared celular vegetal, se clasifican en diferentes familias dependiendo de la forma en la que estas actúan sobre el sustrato, así como en su similitud estructural. En este sistema de clasificación, cada familia presenta una estructura terciaria característica con un origen evolutivo independiente. Las celulasas son agrupadas dentro de las enzimas Glicosil Hidrolasas (GH) (Henrissat & Bairoch 1993, Gilbert 2010, <http://www.cazy.org/>, Lombard et al. 2014), estas se encuentran clasificadas en 19 familias GH: GH1, GH3, GH5, GH6, GH7, GH8, GH9, GH10, GH12, GH16, GH23, GH26, GH39, GH44, GH45, GH47, GH48, GH51, GH74 (Bayer et al. 1998, Sugimura et al. 2003, Gilbert 2010, Shi et al. 2011, <http://www.cazy.org/>, Lombard et al. 2014).

Durante mucho tiempo la única hipótesis aceptada para entender la degradación de celulosa en animales (incluidos los insectos) fue el establecimiento de interacciones simbióticas con microorganismos, aunque en la actualidad es reconocido que algunos animales, especialmente los insectos, pueden secretar sus propias enzimas para llevar a cabo este proceso (Watanabe & Tokuda 2010, Calderón-Cortés et al. 2012). Entre los grupos de animales que pueden sintetizar celulasas, se encuentran: Collembola, Insecta, Cnidaria, Anellida, Mollusca, Crustacea, Chordata, Rotifera y Nematoda (Davison & Blaxter 2005, Danchin et al. 2010, Watanabe & Tokuda 2010, Calderón-Cortés et al. 2012, Szydlowski et al. 2015, Gray et al. 2018).

Las celulasas reportadas para los insectos se clasifican en las familias GH5, GH9,

GH45 y GH48. Las familias GH5, GH45 y GH48 solo se encuentran presentes en coleópteros (Sugimura et al. 2003, Lee et al. 2004, Chang et al. 2012, Eyun et al. 2014); mientras que las celulasas de la familia GH9 están presentes en 9 órdenes de insectos y en más de 145 especies (<http://www.cazy.org/>, Lombard et al. 2014).

Las celulasas de la familia GH9, denominadas endo-1, 4- β -D-glucanasas, son enzimas que actúan sobre los enlaces internos de la celulosa. La característica común para esta familia es la presencia de una estructura terciaria formada por 12 alfa hélices (alfa/alfa), un dominio catalítico y varios dominios multi-catalíticos en el caso de bacterias (Willson et al. 1999, Arimori et al. 2013).

La comparación de las celulasas de la familia GH9 de insectos con sus homólogos en bacterias indica una similitud entre el 42 al 52%, mientras que la comparación de las celulasas entre insectos indica una similitud entre el 75 al 85% (Watanabe et al. 1998, Calderón-Cortés et al. 2010), lo cual sugiere una conservación estructural en la familia. Este grado de conservación fundamenta la hipótesis de que estas enzimas presentan un origen monofilético en el grupo Metazoa (Davison & Blaxter 2005). Sin embargo, el análisis filogenético que plantea esta hipótesis incluyó un número escaso de secuencias de ADN de genes provenientes de insectos (11 secuencias en total). Aunado a lo anterior, la transferencia horizontal de genes ha sido planteada como una hipótesis alternativa que explica la presencia de genes de celulasas en insectos para otras familias de enzimas, tales como la GH5 y la GH45 (Lee et al. 2004, Danchin et al. 2010).

En los últimos años, el número de secuencias de celulasas reportadas para diferentes órdenes de insectos ha incrementado notablemente, lo cual permite realizar nuevos análisis filogenéticos y plantear nuevas hipótesis evolutivas que expliquen mejor el origen de las celulasas de la familia GH9 en insectos (Davison & Blaxter 2005). Es importante resaltar que las secuencias de las celulasas de la familia GH9 tienen un grado de conservación, sin embargo los cambios en algunos aminoácidos podrían reflejar modificaciones en la estructura terciaria de las

enzimas bajo diferentes condiciones, ya que es conocido que las enzimas actúan mejor bajo cierto pH y temperatura (Khademi et al. 2002, Arimori et al. 2013). Por lo tanto, es posible que las celulasas en organismos que viven en diferentes condiciones ambientales presenten cambios a nivel de su estructura. Es por ello, que resulta importante complementar nuevos análisis filogenéticos con análisis estructurales de las enzimas que permitan determinar si existen cambios conformacionales a nivel estructural entre los grupos taxonómicos (Bragatto et al. 2010).

El presente estudio plantea realizar análisis filogenéticos de las secuencias de celulasas de la familia GH9 reportadas en animales protostomados y cnidarios, principalmente en insectos, usando diferentes métodos y estrategias, para determinar el origen evolutivo de las celulasas de la familia GH9 en insectos. El análisis filogenético será complementado con análisis estructurales de las enzimas para determinar si la estructura terciaria de estas refleja las relaciones filogenéticas observadas o si están relacionadas a cambios en estilos de vida y dieta de los organismos.

2. Justificación

En la década pasada, la biología molecular ha tenido un desarrollo sin precedentes, debido al constante desarrollo de nuevas tecnologías para la secuenciación de genomas y transcriptomas, que aunado al desarrollo de la bioinformática (Tan et al. 2014, Carriço et al. 2013), ha permitido que la información presente en bases de datos públicas tales como GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), Cazy (<http://www.cazy.org/>, Lombard et al. 2014) y Uniprot (<https://www.uniprot.org/>, Pichler et al. 2018), aumente de manera considerable.

La acumulación de información en estas bases de datos, así como el surgimiento de nuevas herramientas de análisis permiten someter a evaluación hipótesis previas y proponer nuevas hipótesis para explicar diversos procesos evolutivos importantes. Tal es el caso de la evolución de la capacidad de degradación de celulosa en animales, en especial en los insectos, que constituyen más de la mitad de las especies animales, cuyo estudio permite entender los mecanismos evolutivos involucrados en el origen y preservación de una gran parte de la diversidad del planeta (Strauss & Zangerl 2002).

La degradación de celulosa es un proceso complejo que involucra la acción coordinada de varias enzimas (Beguin & Aubert 1994), entre las cuales destaca la participación de las celulasas. Estudios previos sugieren que poseer celulasas pudo ser una adaptación clave para la evolución y diversificación de los insectos, particularmente en la evolución temprana de estos organismos durante el Silúrico (470 a 423 Ma) (Lo et al. 2003, Calderón-Cortés et al. 2012), período en el cual se asume se presentó la divergencia de los hexápodos a partir de un ancestro crustáceo, y durante el Carbonífero (360-286 Ma) donde se originan y diversifican la mayoría de los linajes de insectos (McGavin 2005, Misof et al. 2014, Calderón-Cortés et al. 2012).

Los insectos y otros grupos de animales invertebrados como cnidarios, moluscos, anélidos y crustáceos poseen genes que codifican celulasas de diferentes familias (<http://www.cazy.org/>, Lombard et al. 2014). Las celulasas de la familia GH9 tienen una mayor presencia en insectos (<http://www.cazy.org/>, Lombard et al. 2014), lo cual plantea preguntas fundamentales sobre el origen y evolución de las celulasas de la familia GH9 en este grupo. Sin embargo, las evidencias disponibles a la fecha sobre la evolución de las celulasas de la familia GH9 en insectos (Lo et al. 2003, Davison & Blaxter 2005) son insuficientes para confirmar la hipótesis evolutiva que postula un origen monofilético de estas enzimas en insectos, ya que los análisis filogenéticos disponibles incluyen un número limitado de secuencias. Por ejemplo, en 2005, en GenBank se reportaban alrededor de 300 secuencias de celulasas de la familia GH9 (Davison & Blaxter, 2005), mientras que en marzo de 2018, se encontraban reportadas más de 2900 secuencias de las cuales alrededor de 450 representan secuencias de animales. Con base en lo anterior, las celulasas de la familia GH9 representan un excelente modelo de estudio, ya que un mayor número de secuencias disponibles en bases de datos permite comprobar o refutar hipótesis sobre el origen y la evolución de estas enzimas en insectos y animales (Davison & Blaxter 2005), además de determinar diferencias a nivel estructural, gracias a la disponibilidad de información sobre proteínas cristalizadas también disponibles en bases de datos. Analizar la evolución de celulasas de la familia GH9, es fundamental para determinar el origen de estas enzimas, así como los cambios a nivel estructural que podrían representar adaptaciones particulares de los organismos que las poseen.

3. Hipótesis

Si las celulasas de la familia GH9 presentes en la clase Insecta provienen de un ancestro común y presentan un origen monofilético, a nivel global, las relaciones filogenéticas de las secuencias de aminoácidos de las enzimas corresponderán a las relaciones filogenéticas de los distintos taxa, mientras que cambios estructurales puntuales podrían reflejar adaptaciones particulares a diferentes presiones de selección.

4. Objetivos

4.1 Objetivo general

- Analizar el origen, la evolución y la estructura de celulasas de la familia GH9 en insectos.

4.2 Objetivos específicos

- Analizar las relaciones filogenéticas de secuencias de proteínas reportadas de celulasas pertenecientes a la familia GH9 en insectos y grupos filogenéticamente relacionados.
- Desarrollar modelos de estructura tridimensional de celulasas de la familia GH9 de insectos y grupos filogenéticamente relacionados.
- Identificar los cambios estructurales de las celulasas de la familia GH9 de insectos entre diferentes grupos y con diferentes hábitos alimenticios.

5. Antecedentes

5.1 Estructura y composición química de la Pared Celular Vegetal

La pared celular vegetal es una estructura macromolecular semirrígida que rodea a la célula y cuyas funciones principales son brindar protección contra patógenos, transportar materia líquida, minimizar la pérdida de agua, proteger de estímulos ambientales, transmitir señales y dar soporte a la planta (Heredia et al. 1995, Cosgrove 2005, Sarkar et al. 2009). La pared celular está compuesta por varias macromoléculas como celulosa, hemicelulosa, lignina, pectina y en algunos casos sustancias cerosas (Heredia et al. 1995, Cosgrove 2005, Lerouxel et al. 2006, Caffall & Mohnen 2009, Sarkar et al. 2009, Wei et al. 2009) que interactúan mediante enlaces no covalentes para formar una estructura altamente organizada (Fig. 1, Fig. 2-B). Las fibras de celulosa, formadas por unidades de glucosa, se unen a fibras de hemicelulosa formadas por cadenas lineales de xilana que incluyen ramificaciones laterales con una gran diversidad de carbohidratos como arabinoxilana, manana, galactomanana, glucomanana, xiloglucanos, arabinogalactana II y β 1_3-Glucano (Heredia et al. 1995, Cosgrove 2005, Caffall & Mohnen 2009). Tanto las fibras de celulosa como de hemicelulosa se encuentran embebidas en una matriz de pectina formada principalmente por largas cadenas de homogalacturonano, rhamnogalacturonano I y rhamnogalacturonano II que incluyen diferentes azúcares como arabinano, galactana, arabinogalactana I, D-galacturonato, apiogalacturonano, con diferente grado de esterificación con grupos metilo o iones metálicos como Ca y Mg (Heredia et al. 1995, Caffall & Mohnen 2009). Esta matriz de pectina que actúa como un agente gelificante, puede ser reemplazada por lignina formada por compuestos aromáticos heterogéneos tales como hidroxifenilo, guaiacilo, syringyl, hidroxilado y metoxilado presentes en algunas zonas de la pared celular vegetal, principalmente en la pared secundaria (Bayer et al. 1998, Santoyo et al. 1999, Cosgrove 2005, Lerouxel et al. 2006, Wei et al. 2009) (Fig. 1, Fig. 2-B).

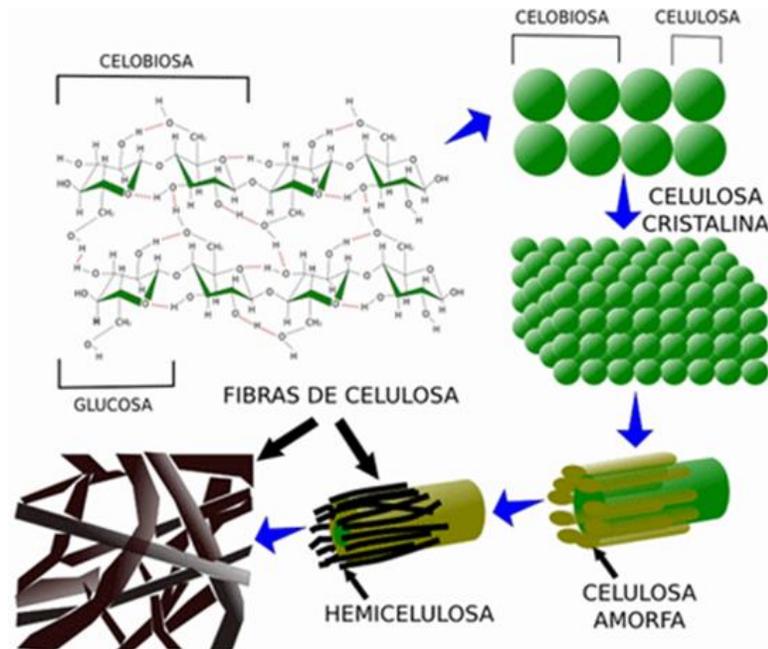


Figura 1 Estructura a distintas escalas de la pared celular. Figura tomada de <https://biologia.laguia2000.com/bioquimica/polisacridos>

La pared celular vegetal se divide en tres zonas principales: lamela media, pared celular primaria y pared celular secundaria que difieren en su composición química y función (Heredia et al. 1995, Cosgrove 2005, Lerouxel et al. 2006) (Fig. 2).

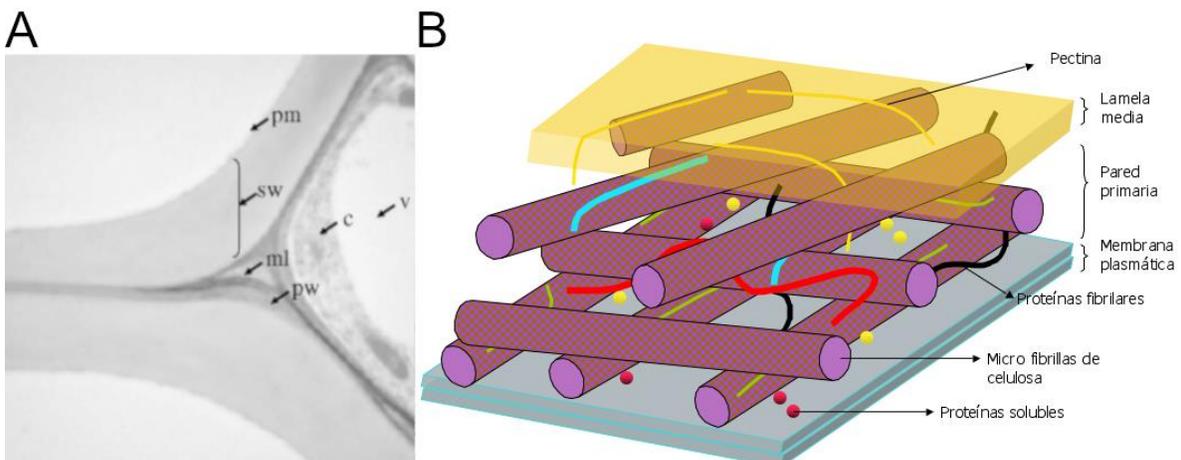


Figura 2 Pared celular vegetal. A.-Pared celular de *Arabidopsis thaliana*. Lamela (ml), pared celular primaria (pw), pared celular secundaria (sw), membrana plasmática (pm), citosol (c) y vacuola (v). Figura tomada de Caffall & Mohnen 2009. B.- Pared celular. Figura realizada por LadyofHats, publicada en Wikipedia

La lamela media, compuesta principalmente por pectina, se forma durante la división celular y es la parte más superficial de la pared celular que permite la deposición de la pared celular primaria (Heredia et al. 1995, Caffall & Mohnen 2009).

La pared celular primaria compuesta de celulosa, hemicelulosa, pectina y una porción variable de proteínas, se encarga de controlar el crecimiento de la planta, dar estructura y adhesión a las células, transmitir señales, entre otras funciones (Heredia et al. 1995, Caffall & Mohnen 2009). Esta zona solo se encuentra en plantas o tejidos jóvenes (Cosgrove 2005, Sarkar et al. 2009).

A diferencia de la pared celular primaria, la pared celular secundaria se caracteriza por la inclusión de altos porcentajes de lignina y polímeros aromáticos heterogéneos que se depositan sobre la pared celular primaria y que brindan a las células mayor rigidez (Heredia et al. 1995, Gilbert 2010). La pared celular secundaria también se caracteriza por tener más celulosa y menos pectina que la pared celular primaria, lo cual la hace más difícil de degradar (Cosgrove 2005, Sarkar et al. 2009, Gilbert 2010, Caffall & Mohnen 2009).

5.1.1 Celulosa

La celulosa es el componente principal de las paredes celulares vegetales y es el biopolímero más abundante del planeta (Bayer et al. 1998, Watanabe & Tokuda 2001, Lerouxel et al. 2006, Caffall & Mohnen 2009) está compuesta por unidades de glucosa unidas por enlaces β -1,4-glucosídicos (Fig. 3), los cuales a su vez se agrupan entre sí por medio de puentes de hidrógeno y enlaces de Van der Waals. Estos enlaces permiten formar una estructura cristalina muy resistente y difícil de degradar, la cual puede llegar a contener hasta 15 mil moléculas de celulosa y tener un diámetro de 2 a 5 nm (Heredia et al. 1995, Bayer et al. 1998, Santoyo et al. 1999, Cosgrove 2005, Caffall & Mohnen 2009, Sarkar et al. 2009, Wei et al. 2009, Gilbert 2010) (Figura 3).

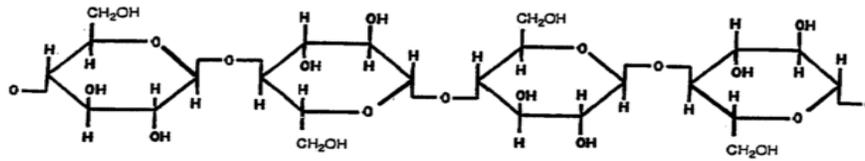


Figura 3 Estructura de la Celulosa. Figura tomada de Heredia 1995

5.2 Degradación de la celulosa y clasificación de las celulasas

Para llevar a cabo la degradación de celulosa se requieren varias enzimas denominadas colectivamente celulasas (Prins & Kreulen 1991, Santoyo et al. 1999, Gilbert 2010). La acción coordinada de estas enzimas explica la degradación de celulosa (Fig. 4). Las celulasas se nombran de acuerdo a la función particular que llevan a cabo sobre el substrato (Sarkar 2009, Gilbert 2010, Willis et al. 2010) de la siguiente forma:

- Endoglucanasas (endo-1,4- β -glucanasas).- Estas enzimas hidrolizan los enlaces glucosídicos de la celulosa para convertirla en gluco-oligosacáridos, haciendo así que la cadena tenga menor longitud y por ende sea más fácil de degradar (Bayer et al. 1998, Santoyo et al. 1999, Vries et al. 2001).
- Exoglucanasas (exo-1,4- β -glucanasas) o celobiohidrolasas.- Estas enzimas degradan la celulosa y los gluco-oligosacáridos a glucosa, esto gracias a que cortan la celobiosa del extremo no reductor de la cadena y la estructura cristalina de la celulosa (Santoyo et al. 1999, Vries et al. 2001, Wei et al. 2009).
- Glucosidasas (β -glucosidasas). - Estas enzimas degradan las cadenas de oligosacáridos, xilooligosacáridos y celobiosa para convertirlas a glucosa (Santoyo et al. 1999, Vries et al. 2001).

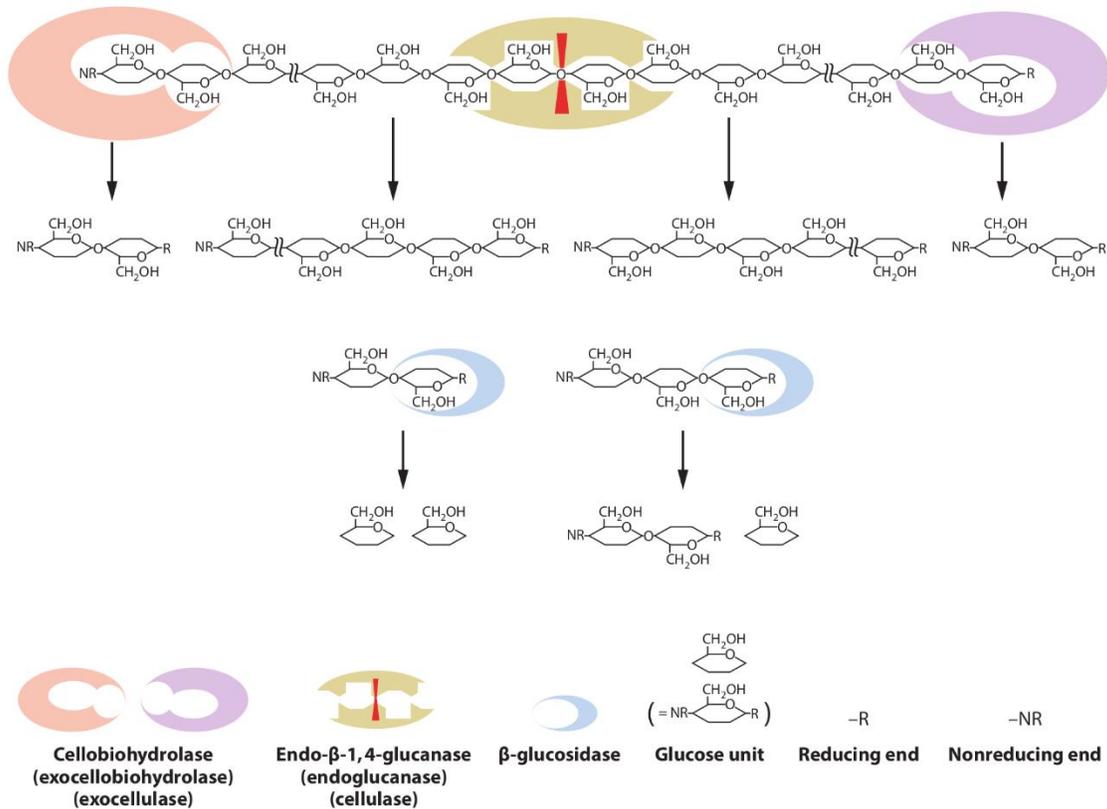


Figura 4 Esquema de la degradación de las celulasas de la familia GH9 sobre fibras de celulosa. Figura tomada de Lo & Watanabe 2010.

Las enzimas se clasifican en base a un sistema que considera la similitud de los aminoácidos y la forma en la que estas actúan sobre el sustrato (Henrissat & Bairoch 1993, Gilbert 2010, <http://www.cazy.org/>, Lombard et al. 2014). Cada familia de enzimas presenta una estructura terciaria característica, así como un origen evolutivo independiente al resto de familias. La clasificación de celulasas las divide en tres grupos de familias principales: los grupos Glicosil Hidrolasas (GH), Modulo de unión a Carbohidratos (CBM) y Actividades auxiliares (AA) (<http://www.cazy.org/>, Lombard et al. 2014).

Dentro del grupo GH, las celulasas se encuentran reportadas en 19 familias distintas las cuales son capaces de hidrolizar oligosacáridos y polisacáridos (Bayer et al. 1998, <http://www.cazy.org/>, Lombard et al. 2014). Las familias GH donde se han reportado celulasas son: GH1, GH3, GH5, GH6, GH7, GH8, GH9, GH10, GH12, GH16, GH23, GH26, GH39, GH44, GH45, GH47, GH48, GH51, GH74

(Bayer et al. 1998, Sugimura et al. 2003, Gilbert 2010, Sukharnikov et al. 2011, <http://www.cazy.org>, Lombard et al. 2014).

Las familias de enzimas se pueden agrupar a su vez en clanes con base en su similitud estructural (Bayer et al. 1998, Santoyo et al. 1999, Sukharnikov et al. 2011, <http://www.cazy.org>, Lombard et al. 2014). Los clanes GH que incluyen celulasas son (Fig. 5):

- Barril (β/α)₈.- Las enzimas pertenecientes a este clan se caracterizan por contar con ocho α -hélices y ocho β -laminas, las cuales forman un barril central (Gromiha et al. 2004).
- Toroide (α/α)₆.- Las enzimas pertenecientes a este clan se caracterizan por contar con doce α -hélices, seis de las cuales son internas y se conectan por medio de seis hélices externas (Alzari et al. 1996, Khademi et al. 2002).
- β -“jellyroll”.- Las enzimas pertenecientes a este clan se caracterizan por contar con ocho β -laminas, cinco de las cuales son antiparalelas y se encuentran en la posición superior de las tres β -laminas restantes. (Richardson 1981, Gaskell et al. 1995).
- α -8 superhelíptico.- Las enzimas pertenecientes a este clan se caracterizan por contar con ocho α -hélices y una pequeña hoja β , compuesta por 3 β -laminas (Brás et al. 2012).
- 7-fold β -propeller.- Las enzimas pertenecientes a este clan se caracterizan por tener una estructura conformada por dos dominios de siete α -hélices cada uno (Yaoi & Mitsuishi 2002).

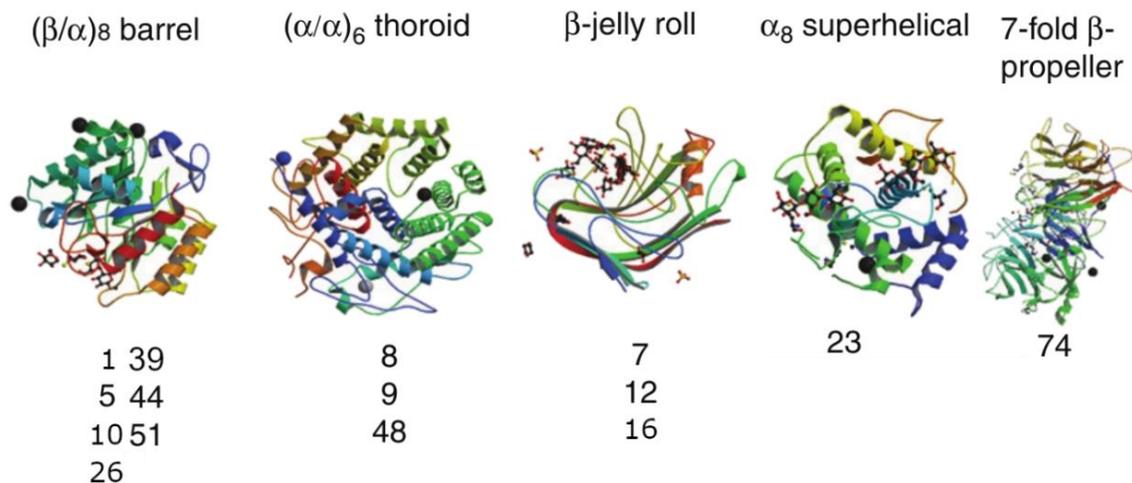


Figura 5 Estructura de los distintos clanes de celulasas. Figura editada de Sukharnikov et al. 2011

5.3 Celulasas en Animales

Los genes que codifican celulasas se encuentran presentes en bacterias, plantas, hongos y animales. A pesar de esto, durante mucho tiempo se pensó que solo las bacterias y los hongos eran los únicos organismos capaces de producir estas enzimas especializadas (Slaytor 1992) y que los animales con una dieta basada en celulosa necesitaban de una relación endosimbiótica para poder subsistir (Cleveland 1924, Yokoe & Yasumasi 1964). En 1952 se postuló que ciertos animales podían generar sus propias celulasas sin depender de bacterias u hongos (Ray & Julian 1952, Yokoe & Yasumasi 1964, Slaytor 1992), aunque esta teoría fue rechazada. Fue en 1998 que se demostró con evidencia molecular que los insectos presentan genes que codifican celulasas de la familia GH9 (Prins & Kreulen 1991, Slaytor 1992, Watanabe et al. 1998, Watanabe et al. 1998, Watanabe & Tokuda 2001).

En la actualidad se sabe que existen varias familias de celulasas presentes en animales (Slaytor 1992, Bayer et al. 1998, Davison & Blaxter 2005, Wei et al. 2005, Danchin et al. 2010). En el grupo de los insectos solo se han reportado celulasas en las familias GH5, GH9, GH45 y GH48 (Calderón-Cortés et al. 2012, Chang et al. 2012, Kirsch et al. 2014, Shelomi et al. 2014, Mayer et al. 2015). Con

la excepción de la familia GH9, el resto de las familias de celulasas en insectos están poco representadas ya que se encuentran en un número limitado de organismos, ejemplo de ello son las familias de celulasas GH5, GH45 y GH48, las cuales únicamente se han reportado en coleópteros (Calderón-Cortés et al. 2012, Chang et al. 2012, Shelomi et al. 2014). Dada la poca representación, el origen evolutivo de estas enzimas en insectos es poco preciso, diversos autores proponen que las familias GH5 y GH45 en animales se adquirieron por medio de transferencia horizontal por parte de bacterias (Calderón-Cortés et al. 2012, Scully et al. 2013, Pauchet et al. 2013, 2014).

Las celulasas de la familia GH9 tienen una mayor representación en animales, incluyendo los insectos (<http://www.cazy.org>, Lombard et al. 2014). En los animales las celulasas de la familia GH9 se han reportado en los taxa: Cnidaria, Mollusca, Annelida, Rotífera, Crustacea, Collembola e Insecta, los cuales representan diversos taxa de los metazoarios Protostomados (Davison & Blaxter 2005, Calderón-Cortés et al. 2012, Arimori et al. 2013, Szydlowski et al. 2015, Aranda et al. 2016). Dos secuencias similares a las celulasas de la familia GH9 se han identificado en los genomas de una especie de Urochordata y una especie de Echinodermata (Davison and Blaxter 2005), los cuales representan taxa de organismos metazoarios Deuterostomados basales. En el grupo de los insectos (Tabla 1) se han reportado en Embioptera (Artículo no publicado, no. de acceso de las secuencias en NCBI: AOV94250-AOV94256), Phasmatodea (Shelomi et al. 2014, Shelomi et al. 2016), Coleoptera (Kim et al. 2008, Willis et al. 2011), Hemiptera (Artículo no publicado, no. de acceso de la secuencia en NCBI: XP_015378087), Termitoidae (Tokuda et al. 1999, Zhou et al. 2007, King et al. 2010), Hymenoptera (Bonasio et al. 2010), Blaberoidea (Watanabe & Tokuda 2001), Orthoptera (Kim et al. 2008) y Phthiraptera (Kirknes et al. 2010). La detección de celulasas de la familia GH9 en diversos organismos se ha generado por distintos métodos: determinación de actividad enzimática, purificación de enzimas, cristalización de proteínas, clonación de genes, análisis de expresión de genes, hibridación de ADN y ARN y recientemente por medio de anotación de

genomas y transcriptomas (Tabla 1).

Tabla 1.- Celulasas de la familia GH9 de animales protostomados.

Grupo de organismos	Número de Especies	Dieta	Datos disponibles
Cnidaria	3	Filtrador/Herbívoro (Algas y microorganismos)	Anotación genómica, expresión RNA, actividad enzimática en placa
Mollusca	19	Filtrador/Herbívoro	Clonación de genes, actividad enzimática en hepatopáncreas, purificación de enzimas
Annelida	5	Detritívoro	Anotación genómica, purificación de enzimas, cristalización de enzima
Rotifera	1	Herbívoro (Fitoplancton)	Anotación genómica
Crustacea	12	Herbívoro/Detritívoro	Clonación de genes, actividad enzimática en hepatopáncreas, purificación de enzimas
Collembola	2	Detritívoro	Anotación genómica
Insecta			
Orthoptera	1	Herbívoro/Detritívoro	Clonación de genes, expresión de RNA, Actividad Enzimática en intestino, Glándulas salivales
Embioptera	1	Detritívoro	Clonación de genes
Phasmatodea	7	Herbívoro	Anotación genómica, clonación de genes, actividad enzimática en intestino, glándulas salivales y purificación parcial de enzimas
Blattodea			
Blaberidae	2	Detritívoro/Omnívoro	Clonación de genes, anotación genómica
Termitoidae	37	Detritívoro/Xilófago	Expresión RNA, anotación genómica, clonación de genes, Glándulas salivales, actividad enzimática en placa, expresión RNA y cristalización de enzima
Hemiptera	3	Herbívoro	Anotación genómica
Phthiraptera	1	Hematófago	Anotación genómica
Coleoptera	4	Herbívoro/Xilófago	Anotación genómica
Hymenoptera	39	Herbívoro/Polen	Anotación genómica

Las citas de los métodos de identificación de las secuencias se pueden encontrar en a tabla 4.

La dieta o hábitos alimenticios de los animales en los cuales se ha reportado la presencia de celulasas de la familia GH9 son muy diversos (Tabla 1). Por ejemplo, los animales marinos como Cnidaria y Mollusca son herbívoros y filtradores (Koueta 1999, Malaquias et al. 2004, Yoshioka et al. 2017), mientras que los

organismos terrestres de los taxa Anellida y Collembola, son predominantemente detritívoros (Jennings & Gelder 1969, Giangrande et al. 2000, Endlweber et al. 2009). La clase Crustacea presenta tanto organismos detritívoros como herbívoros (Castell et al. 1989). Dentro de los insectos se encuentran detritívoros como Embioptera, Ortóptera y Blattodea, organismos xilófagos tanto del orden Coleoptera (barrenadores) como de la familia Termitoidae (termitas), herbívoros de los órdenes Hymenoptera, Coleoptera, Hemiptera y Phasmatodea, e incluso hematófagos del orden Phthiraptera (McGavin 2005).

Dada la alta representación con la que cuentan las celulasas de la familia GH9 se han propuesto varias hipótesis para explicar el origen de estas enzimas, de las cuales diversos autores apoyan la hipótesis de que tienen un origen común ancestral en la evolución de la vida (Watanabe et al. 1998, Watanabe & Tokuda 2001, Davison & Blaxter 2005, Calderón-Cortés et al. 2012, Grey et al. 2018).

La primera hipótesis que plantea el origen de las celulasas de la familia GH9 en animales es la propuesta por Lo, Watanabe y Sugimura (2003), para ello realizan un alineamiento con secuencias provenientes de los taxa Termitoidae, Mollusca y Urochordata, reportando que las secuencias de ADN provenientes de genes de la familia GH9 tienen tres intrones en la misma posición, lo que sugiere que el último ancestro común de los organismos metazoarios contaba con genes codificantes de celulasas de la familia GH9. Posteriormente Davison y Blaxter (2005), realizaron un análisis evolutivo del origen de las celulasas de la familia GH9 en distintos organismos tales como plantas, animales, bacterias y hongos. En dicho estudio los autores sugieren la hipótesis de eventos de pérdida y duplicación de genes de celulasas en distintos grupos de organismos, lo que explicaría la ausencia de estas enzimas en la mayoría de los organismos Deutorostomados. El resultado del análisis filogenético provee evidencia de que las celulasas de la familia GH9 en animales se dividen en dos clados que representan a los grupos Protostomia y Deutorostomia, el posible origen monofilético de las celulasas de la familia GH9 en animales y otros grupos de organismos provee evidencia que

sustenta la hipótesis de un ancestro común muy antiguo de los eucariontes que contaba con celulasas de la familia GH9.

5.4 Estructura de las celulasas de la familia GH9

La estructura de las celulasas de la familia GH9 se conoce gracias a que existen varias estructuras cristalizadas de estas enzimas, en su mayoría provenientes de bacterias (Sakon et al. 1997, Mandelman et al. 2003, Khademi et al. 2002, Parsiegla et al. 2002, Kesavulu et al. 2004, Schubot et al. 2004, Arimori et al. 2013, 4ZH5/4ZG8).

Las celulasas de la familia GH9 se caracterizan por contar con una estructura de barril (α/α)₆, formada por 12 hélices alfa. Las hélices internas están compuestas mayoritariamente por aminoácidos alifáticos y aromáticos que forman el barril central. La cavidad del sitio de unión al sustrato se forma por medio de conexiones entre las hélices internas en el extremo N-terminal de la enzima. Adicional al cuerpo central, la estructura está formada por hélices alfa cortas y hojas beta variables (Arimori et al. 2002, Khademi et al. 2002).

El dominio catalítico de las celulasas de la familia GH9 está formado por tres residuos indispensables para llevar a cabo la actividad enzimática, estos residuos son dos ácidos aspárticos (D) y un ácido glutámico (E). Los ácidos aspárticos funcionan como aminoácidos catalíticos (primario y secundario), mientras que el ácido glutámico funciona como sitio donador de protones. Dichos aminoácidos se encuentran en las posiciones D74, D77 y E431 de la secuencia de la enzima cristalizada proveniente de la lombriz *Eisenia fetida*, mientras que en la secuencia cristalizada proveniente de la termita arbórea *Nasutitermes takasagoensis* se encuentra en las posiciones D54, D57 y E412 (Arimori et al. 2002, Khademi et al. 2002). El dominio catalítico de las celulasas en insectos tiene una similitud a nivel de secuencia de aminoácidos del 52 al 42% con otros grupos de organismos (Watanabe 1998, Tokuda et al. 1999). La estructura cristalizada de *Nasutitermes*

takasagoensis, comparada con las estructuras de celulasas de bacterias por medio de RMSD (Root-Mean-Square Deviation) presenta una diferencia de 1.6 Å y 2.3 Å (Khademi et al. 2002), estos bajos valores se deben a la gran similitud entre las estructuras (Reva et al. 1998).

Se ha reportado que las celulasas de la familia GH9 presentan sitios de unión a iones metálicos, por ejemplo la estructura de celulasas de bacterias presenta tres sitios de unión a iones de calcio (Chauvaux et al. 1995), de los cuales solo uno brinda una mayor estabilidad a la enzima (Khademi et al. 2002). Además, algunos estudios han reportado que el aumento de la concentración de calcio u otros iones metálicos hierro (Fe), magnesio (Mn) o zinc (Zn) en el medio de cultivo de bacterias, incrementa la producción de celulasas extracelulares en bacterias (Mandels et al. 1957, Shahriarinnour 2011, Chen et al. 2018). Por su parte, las celulasas de la familia GH9 en insectos solo presentan un sitio de unión a calcio y la función que este brinda es desconocida (Arimori et al. 2002, Khademi et al. 2002), aunque existen estudios que han reportado que el aumento de calcio en el medio de incubación de los ensayos enzimáticos tiene un efecto bajo o nulo sobre la actividad de celulasas en la especie de coleóptero *Gryllus bimaculatus* (Weidlich et al. 2013), mientras que en equinodermos, se ha reportado que el calcio podría brindar mayor estabilidad a la enzima al igual que en las celulasas de bacterias (Juy et al. 1992).

Otra característica estructural importante es la carga electrostática de la proteína, la cual puede variar en la superficie y en la cavidad donde ocurre la unión al sustrato. En el caso de las celulasas cristalizadas de la familia GH9 en animales se ha reportado que la cavidad en la que ocurre la unión del sustrato se caracteriza por presentar aminoácidos cargados que le confieren carga electrostática negativa (Arimori et al. 2002). Esta cavidad tiene aproximadamente 30 Å de largo y 9 Å de profundidad (Khademi et al. 2002). Por su parte, la superficie electrostática de estas enzimas presenta pocas diferencias de carga (Arimori et al. 2013), aunque la superficie de la celulasa cristalizada de *Eisenia*

fetida presenta una carga electrostática más negativa, posiblemente para soportar temperaturas bajas (Arimori et al. 2013).

5.5 Evolución de animales e insectos

Para poder entender la distribución y evolución de las celulasas de la familia GH9 es necesario conocer la evolución de los organismos que las poseen. Metazoa es uno de los grupos de organismos más diverso. Este reino se divide en los grupos Mesozoa, Placozoa, Porifera y Eumetazoa. El grupo Eumetazoa se divide en los grupos Bilateria que incluye a los Protostomados, Deuterostomados y tres grupos de difícil asignación taxonómica (Gnathostomulida, Platyhelminthes y Xenacoelomorpha) (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>, Jener 2004, Paps et al. 2009, Benton et al. 2015, Cannon et al. 2016).

De acuerdo al registro fósil, *Kimberella quadrata*, el organismo metazoario bilaterio más antiguo (Protostomia; 880 a 550.25 Ma), comparte algunas características con los moluscos actuales tales como alimentación y morfología (Benton et al. 2015). Otros fósiles de organismos metazoarios como *Olivoides multisulcatus* (Cnidaria; 636.1 a 529 Ma) y *Rusophycus avalonensis* (Ecdyzozoa; 636.1 a 528.82 Ma) identificados (Benton et al. 2015), ubican el origen de los metozoarios a inicios del Cámbrico. La hipótesis de la explosión del Cámbrico propone la existencia de un fuerte proceso de diversificación y cladogénesis en metazoarios en un periodo de pocos millones de años. Los ancestros de los grupos de organismos presentes actualmente surgieron en esta época hace aproximadamente 529 a 543 Ma (Fig. 6) (Carroll & Knoll 1999, Jermiin et al. 2005, Benton et al. 2015, Pisani & Liu 2015).

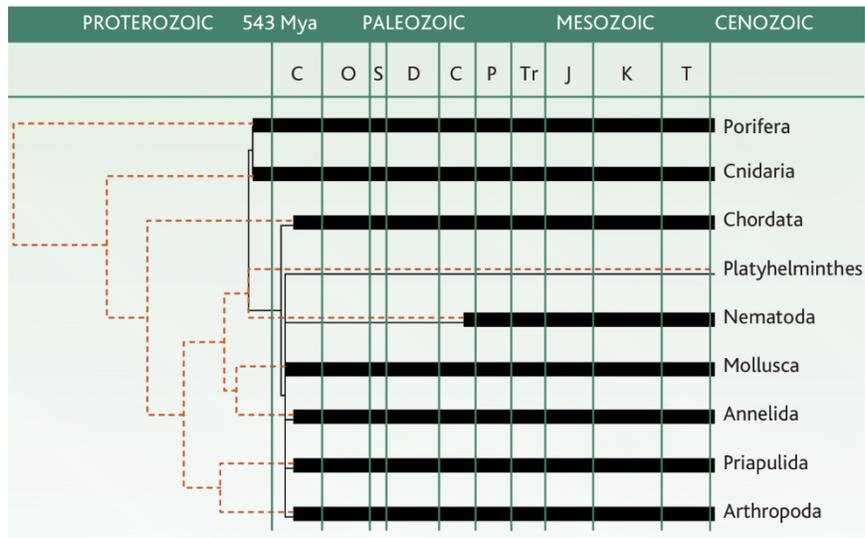


Figura 6 Filogenia de los principales grupos metazoarios, representado la escala de tiempo en la que estos grupos se originaron. Figura extraída de Benton et al. 2015.

Desde que Darwin propuso la teoría de la evolución por medio de selección natural, se han realizado múltiples estudios para intentar esclarecer las relaciones filogenéticas entre los organismos. En la actualidad además de usar datos morfológicos y fósiles, se integran datos genéticos. Sin embargo, a pesar de ello, algunas de las relaciones filogenéticas de los metazoarios, particularmente de taxa basales no están resueltas (Carroll & Knoll 1999, Jermiin et al. 2005, Telford & Copley 2011, Pisani & Liu 2015). Esto se podría deber al origen ancestral de los metazoarios y el posible evento de radiación adaptativa que implicó un proceso de divergencia y diversificación de estos organismos en un período corto de tiempo (Carroll et al. 1999, Regier et al. 2005, Benton et al. 2015).

Los problemas de resolución se pueden ver reflejados en la filogenia de distintos grupos de Eumetazoarios (Protostomia, Deuterostomia, Ecdysozoa, Lophotrochozoa y grupos basales) propuesta por Telford y Copley (2011), en la cual se hace uso de múltiples datos genéticos, tales como ARN mitocondrial y posición de intrones (Fig. 7). En la que los taxa basales (Porifera, Ctenophora, Cnidaria y Placozoa) no presentan relaciones claras, pues se forma una politomía al inicio del árbol. A pesar de ello, se reconoce la monofilia del grupo Bilateria y de sus subgrupos (Deuterostomia y Protostomia) (Telford & Copley 2011).

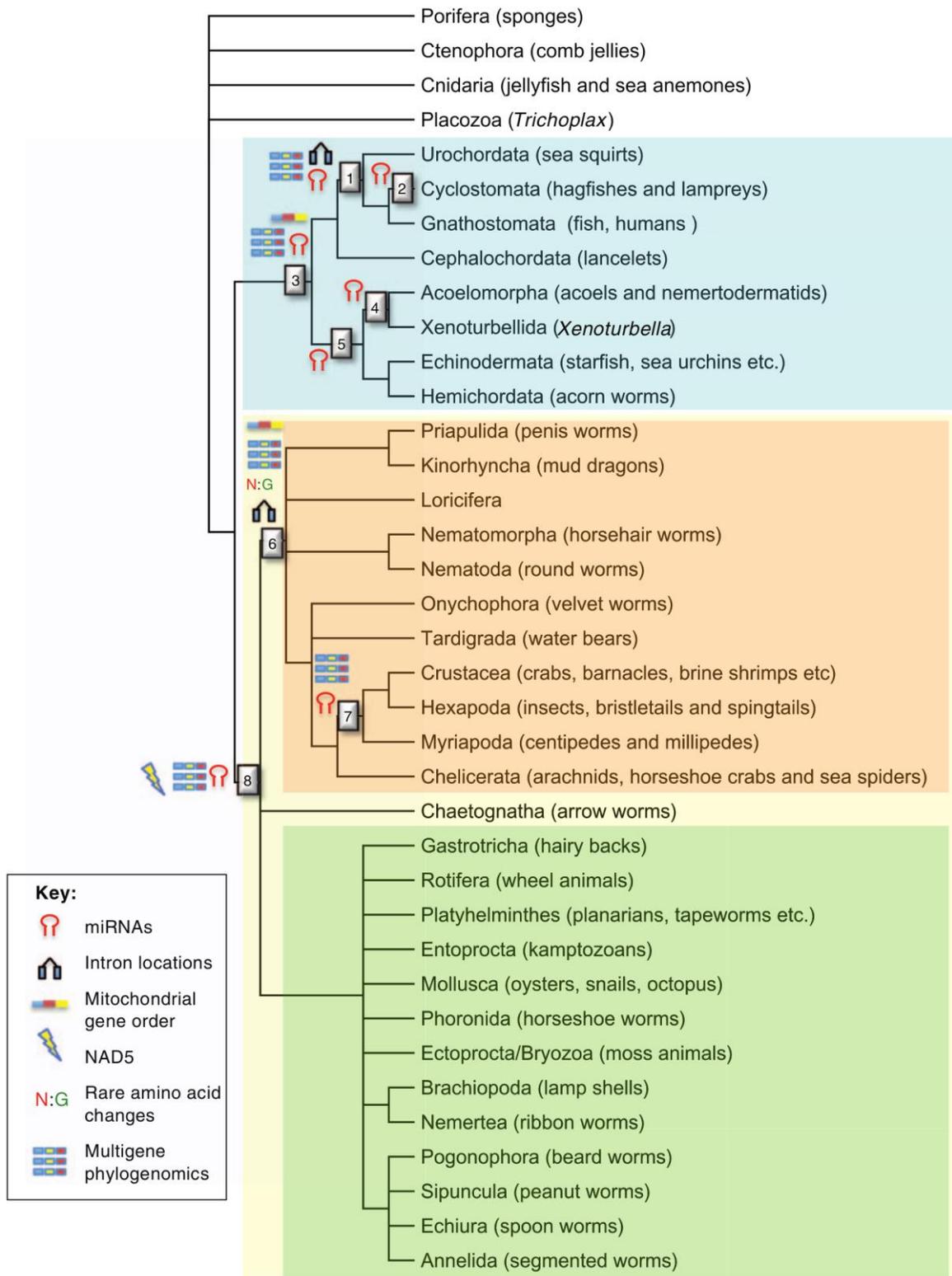


Figura 7 Vista consenso de las relaciones filogenéticas de los animales metazoarios, grupos Protostomia, Deuterostomi, Ecdysozoa y Lophotrochozoa resultados en cuadros color amarillo, azul, naranja y verde, respectivamente. Figura tomada de Telford & Copley (2011).

Dentro de los organismos Protostomados se encuentran el taxon Artropoda, el cual está compuesto por una gran diversidad de organismos, los cuales se pueden encontrar en distintos ecosistemas (McGavin 2005). Dentro del grupo Artropoda se encuentran los insectos, organismos que representan una gran parte de las especies de animales (McGavin 2005). La evolución de estos organismos ocurrió durante el Silúrico hace aproximadamente 470 a 423 Ma (McGavin 2005, Calderón-Cortés et al. 2012, Misof et al. 2014) y fue durante el Carbonífero hace aproximadamente 360 a 286 Ma que se presentó la divergencia y origen de los linajes basales (Calderón-Cortés et al. 2012).

En la filogenia consenso propuesta por Trautwein y colaboradores (2012) (Fig.8), el grupo Crustacea se identifica como grupo externo. Como grupo basal del grupo Hexapoda se encuentran los hexápodos no insectos (Collembola, Diplura y Protura). El clado Insecta se subdivide en 4 clados: un clado basal perteneciente al superorden Palaeoptera (Odontonata y Ephemeroptera), mientras que los otros 3 clados pertenecen a la infraclase Neoptera (Paraneoptera, Polyneoptera y Holometabola). El superorden Polyneoptera que incluye un gran número de insectos con origen Paleozoico está conformado por los órdenes Orthoptera, Phasmatodea, Embioptera, Mantophasmatodea, Grylloblattodea, Dermaptera, Zoraptera, Mantodea, Blattodea y Termitoidae. El superorden Paraneoptera está conformado por los órdenes Psocoptera, Phthiraptera, Thysanoptera y Hemiptera, mientras el superorden Holometabola incluye varios grupos de insectos con procesos de diversificación más recientes conformado por los órdenes Hymenoptera, Strepsiptera, Coleoptera, Neuroptera, Megaloptera, Raphidioptera, Trichoptera, Lepidoptera, Diptera, Siphonaptera y Mecoptera (Trautwein et al. 2012).

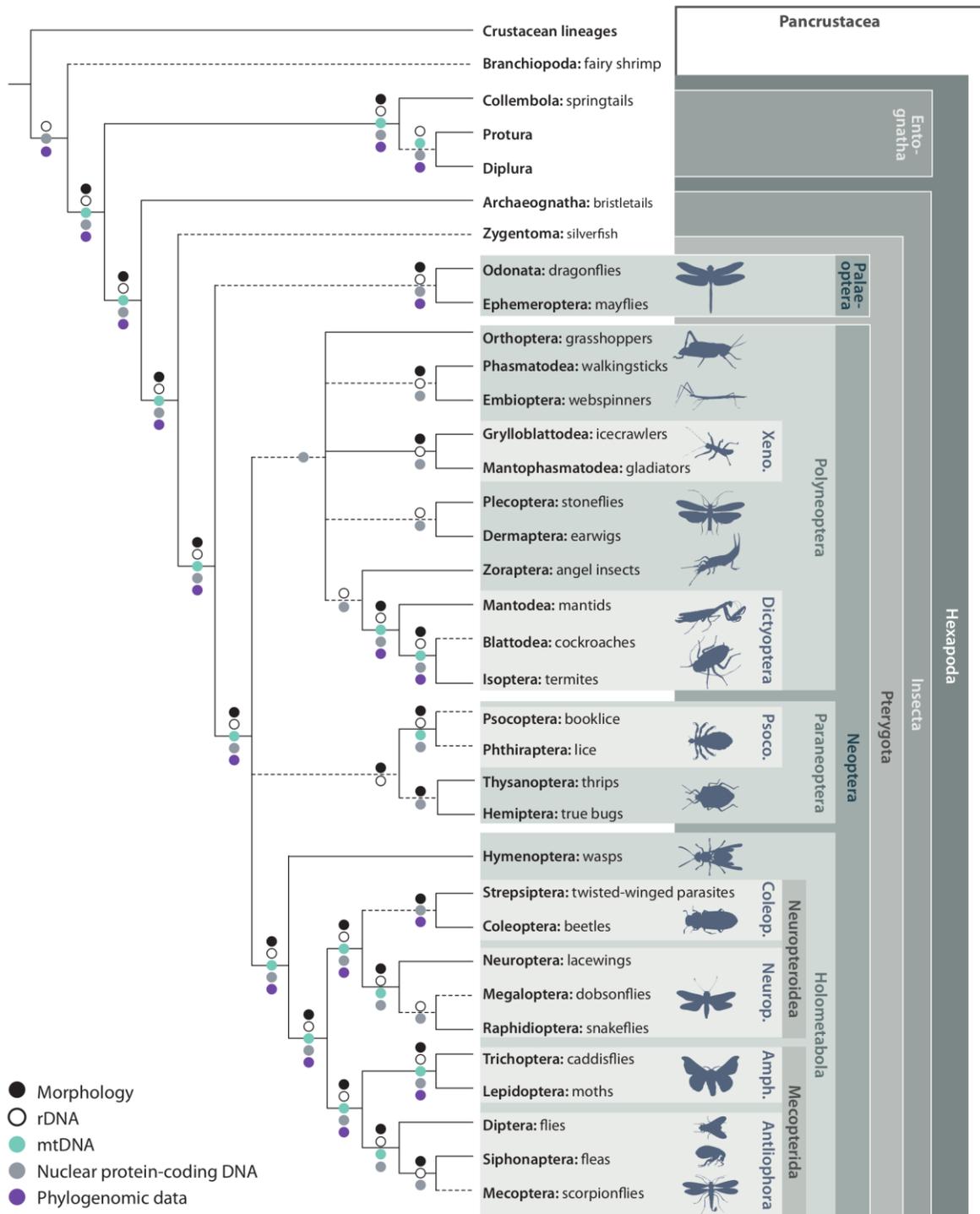


Figura 8 Relaciones filogenéticas de Hexápoda, figura tomada de Trautwein et al. (2012)

6. Metodología

6.1 Búsqueda de secuencias

Para identificar las secuencias de celulasas pertenecientes a la familia GH9 en insectos, animales protostomados y cnidarios se realizó una búsqueda utilizando secuencias reportadas en la literatura (Tabla 2). Para corroborar que dichas secuencias pertenecieran a la familia GH9, se analizó la presencia del dominio Glyco_hydro_9 (pfam00759). Posteriormente, se realizaron múltiples búsquedas en la herramienta BLAST del GenBank (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, Altschul et al.1990, Gish & States 1993) en su versión Protein BLAST utilizando como blanco las secuencias mencionadas.

Tabla 2.- Secuencias de proteínas reportadas en la literatura usadas para realizar la búsqueda de secuencias en la herramienta BLAST.

Numero de acceso en NCBI	Especies	Referencia
1KS8_A	<i>Nasutitermes takasagoensis</i>	Khademi et al. 2002
3WC3_A	<i>Eisenia fetida</i>	Arimori et al. 2013
AAU20853	<i>Reticulitermes flavipes</i>	Scarf et al. 2005
ADB85440, ADB85441	<i>Limnoria quadripunctata</i>	King et al. 2010
ADI24131	<i>Oncideres albomarginata chamela</i>	Calderón-Cortés et al. 2010
AIT11911	<i>Parasesarma erythroductyla</i>	Bui & Lee 2014
AMH40356- AMH40396	<i>Aretaon asperrimus, Extatosoma tiaratum, Medauroidea extradentata, Peruphasma schultei, Ramulus artemis, Sipyloidea sipyulus, Timema cristinae.</i>	Shelomi et al. 2016
AOV94250- AOV94256	<i>Antipaluria urichi</i>	Shelomi et al. 2016
BAA28815	<i>Reticulitermes speratus</i>	Watanabe et al. 1998
BAN82458- BAN82460, BAN82515	<i>Eurycantha calcarata, Entoria okinawaensis</i>	Shelomi et al. 2014
BAS18736	<i>Bellamyia sp. UM-2014,</i>	Ueda et al. 2016
ODM94862	<i>Folsomia candida</i>	Faddeeva-Vakhrusheva et al. 2016

Se descargaron los archivos en formato de valores separados por comas (CSV) de todas las secuencias que arrojaron las búsquedas, sin tomar en cuenta la similitud ni la cobertura con la secuencia blanco, la búsqueda arrojó muchas secuencias duplicadas. Usando un script en Python (Anexo 2; Script 1) (G. van Rossum et al. 1995), se eliminaron los registros de secuencias duplicadas. A continuación se realizó una búsqueda de las secuencias obtenidas en la herramienta del GenBank, Batch Entrez (<https://www.ncbi.nlm.nih.gov/sites/batchentrez>) y se descargó un archivo en formato GenPept con la descripción general de los números de acceso, usando un script en Python (Anexo 2; Script 2). Utilizando minería de texto y el módulo ETE2 (Huerta-Cepas et al. 2010), se obtuvo información de cada número de acceso (p. e. número de aminoácidos que componen la secuencia, clasificación taxonómica del organismo del que proviene, descripción breve, artículo del cual proviene, autores y si la secuencia pertenece al dominio pfam00759. Con esta información se descartaron las secuencias que no eran celulasas y que pertenecían a otras familias GH, así como las secuencias que no correspondían a los grupos Protostomia y Cnidaria.

Para ampliar el número de secuencias, se realizó una segunda búsqueda, usando como blanco las secuencias de celulasas obtenidas en la búsqueda anterior (ver resultados), repitiendo los mismos pasos que en la primera búsqueda y eliminando aquellas secuencias con una longitud menor a 440 aminoácidos.

6.2 Alineamiento de secuencias

Utilizando la herramienta Batch Entrez, se hizo una búsqueda de los archivos fasta de las secuencias obtenidas en la búsqueda, para hacer un alineamiento con Muscle (Edgar R.C. 2004) en Molecular Evolutionary Genetics Analysis versión 6 (Mega6) (Tamura et al. 2013). El alineamiento obtenido se editó manualmente para confirmar que los dominios conservados de las celulasas de la familia GH9 se encontrarán en la misma posición en todas las secuencias incluidas en el

análisis. Durante esta edición manual se descartaron aquellas zonas del alineamiento con más del 70% de gaps, los extremos amino, carboxilo y los sitios de unión a celulosa. Las zonas variables con gran número de gaps pueden generar relaciones erróneas en los análisis filogenéticos (Dwivedi & Gadagkar 2009), por este motivo se descartaron. Resultando así un alineamiento con una longitud de 421 aminoácidos.

6.3 Selección de secuencias

La adición o exclusión de ciertos grupos de secuencias en los análisis filogenéticos puede afectar la reconstrucción de las relaciones del resto de los clados (Sanderson & Shaffer 2002). Por lo cual, se realizaron distintos análisis filogenéticos usando distintos grupos externos (Bacteria, Cnidaria y Collembola), incluyendo y excluyendo diferentes grupos y secuencias de animales para evaluar la robustez de las relaciones filogenéticas recuperadas, generando así 3 alineamientos:

Alineamiento 1.- Se seleccionó una sola secuencia por especie, dando prioridad a las secuencias que no provinieran de anotación de genomas. En este alineamiento se incluyó a una secuencia de bacteria proveniente de la especie *Cellomonas fimi* (AAA23086) como grupo externo, dado que se ha reportado en la literatura que esta secuencia está más emparentada con las primeras secuencias de celulasas reportadas de insectos con respecto al resto de secuencias de bacterias (Tokuda 2005). Este alineamiento incluyó en total 137 secuencias (Anexo 1; Tabla 1).

Alineamiento 2.- Debido a que el primer análisis sólo incluyó una secuencia por especie, se realizó un análisis posterior para incluir todas las secuencias recuperadas en la fase de búsqueda (Tabla 4), con el fin de evaluar cambios en las relaciones entre los clados. Para este alineamiento se incluyeron las 349 secuencias de celulasas depuradas de la fase de búsqueda. Como grupo externo del análisis filogenético con método de máxima verosimilitud en RaxML

(Stamatakis 2014) se usaron todas las secuencias pertenecientes al grupo Cnidaria, mientras que para el análisis filogenético por método bayesiano en Mr. Bayes v3.2 (Huelsenbeck & Ronquist 2001, 2003) se usó como grupo externo a una secuencia proveniente de la especie *Exaiptasia pallida* (Cnidaria; KXJ18436.1).

Alineamiento 3.- Debido a las relaciones filogenéticas de insectos recuperadas y la falta de resolución entre secuencias provenientes de algunos taxa se realizó un tercer análisis para tratar de esclarecer las relaciones filogenéticas de los insectos. Para este alineamiento se seleccionaron las 198 secuencias de celulasas del grupo Hexapoda (Anexo 1; Tabla 2). Como grupo externo del análisis filogenético con método de máxima verosimilitud en RaxML se usaron todas las secuencias pertenecientes al grupo Collembola, mientras que para el análisis filogenético por método bayesiano en Mr. Bayes se seleccionó la secuencia proveniente de la especie *Orchesella cincta* (Collembola; ODM94862 .1).

6.4 Análisis filogenéticos

6.4.1 Modelo Evolutivo

A partir de cada alineamiento obtenido se determinó el modelo de evolución molecular que mejor se ajustará a los datos. Para ello se utilizó la herramienta “Find Best DNA/Protein Models (ML)” la cual está integrada en el programa Mega6.

6.4.2 Método Bayesiano

Para realizar los análisis filogenéticos por método bayesiano se usó el programa Mr. Bayes v3.2, ajustando los parámetros al modelo evolutivo. Se configuró un muestreo de 1 millón de iteraciones (ngen= 1,000,000) con un muestreo cada 1,000 iteraciones (samplefreq = 1,000), lo cual dio como resultado 1,000 árboles muestra. Antes de la estimación de los valores de soporte, se comprobó la

convergencia de las cuatro cadenas del análisis, revisando los valores de “log likelihood”, mediante una gráfica generada en Libre Office. Posteriormente se descartó el 5% de los primeros árboles (“burn in”; sumt burnin = 50), y con los 950 árboles restantes se estimaron los valores de soporte y se obtuvo el árbol consenso de acuerdo al criterio de mayoría (>50%). Los arboles obtenidos, fueron editados manualmente con el programa TreeGraph2 (Stöver et al. 2010) para asignar el grupo externo.

6.4.3 Método de Máxima Verosimilitud

El análisis de máxima verosimilitud se realizó en el programa RAxML-HPC BlackBox en el servidor Cipres (Miller et al. 2010), usando proteínas como tipo de archivo, una matriz de sustitución WAG y la estimación de la proporción de sitios invariables (GTR GAMMA + I).

6.5 Análisis estructural

6.5.1 Modelado de proteínas

Para realizar el modelado de proteínas se usaron algunas de las secuencias obtenidas en el alineamiento, las cuales son representativas de los clados formados en las filogenias (Tabla 3).

Tabla 3.- Secuencias usadas para generar los modelos de la estructura terciaria.

Grupo	Especies	Número de Acceso
Cnidaria	<i>Exaiptasia pallida</i> , <i>Orbicella faveolata</i>	KXJ18436.1, XP_020614867.1
Mollusca	<i>Lottia gigantea</i> , <i>Biomphalaria glabrata</i>	XP_009049791.1, XP_013066277.1
Anelido	<i>Perinereis brevicirris</i>	BAK20401.1
Rotifero	<i>Adineta ricciae</i>	AKC91349.1
Crustasea	<i>Limnoria quadripunctata</i> , <i>Coenobita brevimanus</i> , <i>Daphnia pulex</i>	ADB85440.1, APF29560.1, EFX69372.1, EFX86822.1
Collembola	<i>Folsomia candida</i>	XP_021956548.1

Collembolo	<i>Orchesella cincta</i>	ODM94862.1
Orthoptera	<i>Teleogryllus emma</i>	ABV32557.1
Embioptera	<i>Antipaluria urichi</i>	AOV94253.1, AOV94256.1, AOV94250.1
Phasmatodea	<i>Medauroidea extradentata, Peruphasma schultei, Ramulus artemis, Timema cristinae, Timema cristinae</i>	AMH40372.1, AMH40373.1, AMH40379.1, AMH40392.1, AMH40395.1
Blaearoidea	<i>Panesthia cribrata, Blattella germanica</i>	AAF80584.1, PSN31180.1
Termitoidea	<i>Rhynchotermes bulbinasus, Microtermes pakistanicus, Reticulitermes speratus, Mastotermes darwiniensis, Cryptotermes secundus, Zootermopsis nevadensis, Cryptotermes secundus</i>	AGP76427.1, AKV16371.1, BAA31326.1, CAD54726.1, XP_023704929.1, KDR16731.1, XP_023716596
Hemiptera	<i>Acyrtosiphon pisum, Myzus persicae, Nilaparvata lugens</i>	XP_001944774, XP_022167267.1, XP_022200004.1
Phthiraptera	<i>Pediculus humanus corporis</i>	XP_002426465.1
Coleoptera	<i>Tribolium castaneum, Nicrophorus vespilloides, Anoplophora glabripennis</i>	XP_008194455, XP_017785110.1, XP_018568194.1
Hymenoptera	<i>Camponotus floridanus, Dufourea novaeangliae, Fopius arisanus, Apis mellifera</i>	EFN70196.1, KZC10616.1, XP_011296914.1, XP_396791.3

Se descargó un archivo fasta con las 41 secuencias de proteínas seleccionadas para realizar los modelos con la herramienta Batch Entrez. Se procedió a realizar un alineamiento, usando el método Muscle en Mega6. El alineamiento se editó manualmente, removiendo los extremos amino y carboxilo tomando como punto de corte las zonas de inicio y término del alineamiento anterior.

Para generar los modelos de las estructuras terciarias de las proteínas se usó el método de modelado por homología por medio del servidor SWISS MODEL (<https://swissmodel.expasy.org/>, Bertoni et al. 2017). Mediante el uso de un script en Python (Anexo 2; Script 3), se cargaron cada una de las secuencias alineadas en el servidor, para así generar los modelos con base al templado con mayor similitud de secuencia de aminoácidos. Se seleccionaron y descargaron los modelos con los mayores valores GMQE (Global Model Quality Estimation) (Benkert et al. 2009), mismos que representan la calidad que combina propiedades del alineamiento y el método de búsqueda del templado.

Al usar modelado por homología, algunas de las conformaciones estructurales de

las cadenas laterales de los aminoácidos se pueden encontrar en posiciones energéticamente desfavorables, por lo que es necesario realizar un refinamiento para mejorar la calidad de los modelos (Levitt & Lifson 1969). Para ello se usó el software GROMACS 5.0 (Berendsen et al. 1995, Lindahl et al. 2001, van der Spoel et al. 2005, Hess et al. 2008, Pronk et al. 2013, Páll et al. 2015, Abraham et al. 2015) el cual permite corregir problemas estructurales tales como choques estéricos entre las cadenas laterales de los aminoácidos y corrección de ángulos de torsión, disminuyendo la energía general del sistema, ya que se considera que valores negativos de los ángulos de torsión en los modelos favorecen la termodinámica del sistema, brindando mayor estabilidad. Para el refinamiento de los modelos se usaron dos scripts en Python (Anexo 2; Script 4 y Script 5), el refinamiento fuerte y el refinamiento ligero. El refinamiento ligero simula un proceso de solvatación de los modelos con moléculas de agua, mientras que el refinamiento fuerte además de simular el proceso de solvatación con agua, agrega moléculas de sodio y cloro. El refinamiento fuerte minimiza aún más la energía del modelo en comparación con el refinamiento ligero. Posteriormente con el uso de otro script (Anexo 2; Script 6) se eliminaron las moléculas de agua, sodio y cloro de los modelos refinados.

El análisis de los enlaces phi (Φ) psi (Ψ) es un parámetro fundamental para determinar la calidad de los modelos tridimensionales generados. En este análisis se busca el menor porcentaje de aminoácidos con ángulos de torsión en zonas no permitidas. En particular, los gráficos de Ramachandran permiten conocer y evaluar la proporción de aminoácidos con ángulos phi (Φ) y psi (Ψ) en regiones permitidas, en regiones aceptables (máximo 2% de los aminoácidos) y en regiones no permitidas (Hooft et al. 1997, Lovell et al. 2002, Ho et al. 2009). Entre menor sea el porcentaje de aminoácidos en zonas aceptables y no permitidas, menor será la probabilidad de que el modelo tenga problemas estructurales. Por lo tanto, para evaluar la calidad de los modelos se hizo un análisis del gráfico de Ramachandran para cada modelo generado y refinado, desarrollando un séptimo Script en Python (Anexo 2; Script 7), el cual permite cargar los modelos generados

a la página de Rampage (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>, Lovell et al. 2002). Con el uso de minería de texto se obtuvo información de la conformación espacial de los aminoácidos que componen cada uno de los modelos.

El análisis comparativo de las estructuras de los modelos se realizó por medio de la superposición estructural de los carbonos α con el uso del lenguaje de programación R (R Development Core Team 2008) y la librería Bio3D (Grant et al. 2006), para ello se realizó un script (Anexo 2; Script 8) que permite leer y alinear los modelos. La superposición de las estructuras se realiza con base en un alineamiento de las secuencias de los modelos generado en Muscle. La superposición se realizó en todos los modelos, generando así una matriz con los valores de RMSD (desviación de la media cuadrática); entre menor sea el valor de RMSD la diferencia estructural de los modelos será menor (Maiorov & Crippen 1994). A partir de esta matriz se generó un dendrograma de distancias mínimas con el fin de tener una mejor visualización de los resultados.

6.5.2 Predicción de sitios de unión a iones

Con el fin de analizar la presencia de sitios probables de unión a iones se usó la herramienta BindingSite incorporada en el servidor Raptor X (<http://raptorx.uchicago.edu/>, Källberg et al. 2012). Esta herramienta predice los sitios de unión por medio de homología usando como templado varias estructuras cristalizadas. Para hacer el análisis de todas las secuencias obtenidas en la fase de depuración, se cargaron archivos fasta con 20 secuencias cada uno, esto dado que el servidor RaptorX solo permite trabajar con dicho número de secuencias a la vez.

Para interpretar los resultados de los sitios de unión se usaron dos scripts (Anexo 2; Script 9 y 10). Para obtener una mejor visualización de los resultados, estos se mapearon en el árbol filogenético obtenido en el análisis del alineamiento 2 en Mr.

Bayes, para ello se realizó un script (Anexo 2; Script 11).

6.5.3 Análisis evolutionary trace de celulasas de insectos

Con el fin de identificar mutaciones en las secuencias de celulasas de la familia GH9 en insectos y verificar si estas son consistentes con la formación de los clados dentro de los arboles filogenéticos, se realizó un análisis de residuos conservados y variables con la herramienta Evolutionary Trace (<http://mordred.bioc.cam.ac.uk/~jiye/evoltrace/evoltrace.html>, Innis & Blundell 2000). Esta herramienta genera secuencias consenso de clados basándose en el agrupamiento por similitud de las mismas, posteriormente compara las similitudes y diferencias de las secuencias consenso, para finalmente mapear los cambios en una estructura proteica. Los clados se definen a partir del cladograma generado por el análisis con un método de distancia, definiendo un número determinado de particiones. Cada partición (representada como una línea vertical, donde los clados posteriores son usados para generar las secuencias consenso) contendrá diferentes clados, por lo cual en este análisis es necesario definir cual partición incluye los clados que se desean comparar. Para llevar a cabo el análisis se definieron 20 particiones, ya que con este número de particiones en el análisis, el cladograma obtenido incluía los dos clados de celulasas de insectos obtenidos en los análisis filogenéticos que se requerían analizar.

Para realizar este análisis se cargó el alineamiento usado para generar los modelos, descartando las secuencias ajenas al grupo hexápoda. La estructura cristalizada de la termita *Nasutitermes takasagoensis* (Termitidae) fue utilizada como referencia estructural. Se establecieron 20 particiones para el análisis y se comprobó cuáles zonas del alineamiento presentan mutaciones en ciertos aminoácidos por medio de un Script (Anexo 2; Script 12).

6.5.4 Análisis de hidrofobicidad, potencial electrostático y cavidad de celulasas de insectos

Para llevar a cabo la visualización de la hidrofobicidad y potencial electrostático se seleccionaron distintos pares de secuencias de celulasas de la familia GH9; cada par estuvo conformado por secuencias provenientes de una misma especie y que en el análisis filogenético se encontrasen en los dos clados formados en el grupo de los insectos: *Zootermopsis nevadensis* (KDR16731.1 y XP_021941324.1; Termopsidae), *Timema cristinae* (AMH40392.1 y AMH40395.1; Timematidae) y *Antipaluria urichi* (AOV94250.1 y AOV94253.1; Clothodidae). Como control se usó la estructura cristalizada de *Nasutitermes takasagoensis* (1KSC; Termopsidae) y la secuencia proveniente de *Apis mellifera* (XP_396791.3; Apidae). Para visualizar la hidrofobicidad se usó el programa UCSF Chimera 1.13.1 (Pettersen et al. 2004) en el cual se tomó captura a la parte de la cavidad donde ocurre la unión al sustrato. El análisis de hidrofobicidad permitió observar diferencias en cuanto a la distribución y cantidad de aminoácidos con propiedades hidrofóbicas en la hendidura de las estructuras.

Para visualizar el potencial electrostático de la superficie de las enzimas se utilizó el programa Chimera 1.13.1. Para este análisis fue necesario agregar carga e hidrógenos a los modelos para posteriormente calcular el potencial electrostático por Adaptive Poisson-Boltzmann Solver (APBS) (Baker et al. 2002). Para llevar a cabo la visualización de electronegatividad se tomaron capturas a la parte de la cavidad donde ocurre la unión al sustrato y la zona posterior a esta.

7. Resultados

7.1 Búsqueda de secuencias

El primer proceso de búsqueda arrojó aproximadamente 7000 secuencias de proteínas de celulasas de la familia GH9, de las cuales al eliminar duplicados quedaron 534 secuencias. Al eliminar secuencias ajenas a la búsqueda quedaron 404 secuencias de celulasas de la familia GH9 en animales protostomados y cnidarios.

El segundo proceso de búsqueda de secuencias arrojó aproximadamente 42780 secuencias, al eliminar duplicados quedaron 782 secuencias. Al eliminar secuencias ajenas a la búsqueda y aquellas con una longitud menor a 440 aminoácidos, se obtuvieron 349 secuencias provenientes de 136 especies (Tabla 4). Estas secuencias se usaron para realizar los análisis filogenéticos y estructurales.

Tabla 4.- Secuencias de celulasas de animales protostomados obtenidas de la búsqueda realizada en el GenBank.

TAXA	ESPECIES PERTENECIENTES AL GRUPO	NO. DE ACCESO DEL NCBI	REFERENCIAS
CNIDARIA	<i>Exaiptasia pallida</i> , <i>Stylophora pistillata</i> , <i>Orbicella faveolata</i>	KXJ18436.1, XP_020893295.1, XP_020893299.1, PFX25363.1, XP_020614867.1	Lehnert et al. 2012, Baumgarten et al. 2015, Voolstra et al. No publicado
MOLLUSCA	<i>Ampullaria crossean</i> , <i>Haliotis discus discus</i> , <i>Haliotis tuberculata</i> , <i>Haliotis gigantea</i> , <i>Haliotis discus hannai</i> , <i>Haliotis discus</i> , <i>Bellamyia sp. UM, 2014</i> , <i>Aplysia kurodai</i> , <i>Pomacea canaliculata</i> , <i>Aplysia californica</i> , <i>Lottia gigantea</i> , <i>Biomphalaria glabrata</i> , <i>Octopus bimaculoides</i> , <i>Crassostrea gigas</i> , <i>Crassostrea virginica</i> , <i>Mizuhopecten yessoensis</i> , <i>Achatina fulica</i> , <i>Corbicula japonica</i>	ABD24274.1, ABD24281.1, ABO26609.1, BAD44734.1, AFQ98379.1, AFQ98380.1, AGI61069.1, BAD01504.1, BAC67186.1, BAS18736.1, BAU45387.1, BAU45388.1, BAX09271.1, PVD35337.1, PVD35339.1, PVD35340.1, PVD35381.1, PVD35383.1, PVD35385.1, PVD35551.1, PVD36624.1, PVD36625.1, PVD36896.1, XP_005099312.1, XP_005099315.1, XP_005101787.1, XP_005101789.1, XP_005111868.2, XP_005111869.2, XP_005111870.1, XP_012935745.1, XP_012938497.1, XP_012939931.1, XP_012939933.1, XP_012941225.1, XP_012945520.1, XP_009049791.1, XP_009049836.1, XP_009049839.1,	Albertin et al. No publicado, Liu & Sui No publicado, Nikapitiya et al. No publicado, Niwa & Aono No publicado, Sade et al. No publicado, Sakamoto et al. No publicado, Simakov et al. 2013, Suzuki et al. 2003, Suzuki et al. No publicado, Tsuji et al. No publicado, Ueda et al. 2016, Wang et al. 2017, Zhan et al. No publicado, Zhang et al. 2012

		<p>XP_009049840.1, XP_009049941.1, XP_009058133.1, XP_009058134.1, XP_009060290.1, XP_009064976.1, XP_009064980.1, XP_009065702.1, XP_013066277.1, XP_013066278.1, XP_013066280.1, XP_013066281.1, XP_013066283.1, XP_013066286.1, XP_013066370.1, XP_013066371.1, XP_013069169.1, XP_013069170.1, XP_013069521.1, XP_013069523.1, XP_013084279.1, XP_013084281.1, XP_013084283.1, XP_013084285.1, XP_013084287.1, XP_013088631.1, XP_013094438.1, XP_013066268.1, KOF68403.1, XP_014786694.1, EKC17289.1, EKC39023.1, XP_011421782.1, XP_011423159.1, XP_011430172.1, XP_011447501.1, XP_011447502.1, EKC24431.1, EKC29105.1, XP_022328778.1, XP_022329668.1, XP_022336001.1, XP_022336153.1, XP_022338490.1, XP_022338491.1, XP_022341787.1, OWF45593.1, OWF56607.1, XP_021345983.1, XP_021361084.1, XP_021363481.1, XP_021363490.1, XP_021363495.1, XP_021363496.1, XP_021363498.1, ATS91249.1, BAF38757.1, 3WC3_A, XP_001810693.2, BAM14716.1, BAT21145.1, AAX92641.1, ACE75510.1, ACE75511.1, BAH22180.1, ANR02619.1, BAK20401.1</p>	
ANNELIDA	<i>Eisenia fetida</i> , <i>Eisenia andrei</i> , <i>Metaphire hilgendorfi</i> , <i>Perinereis aibuhitensis</i> , <i>Perinereis brevicirris</i>		Arimori et al. 2013, Baik et al. No publicado, Ito et al. No publicado, Lee et al. No publicado, Nozaki et al. No publicado, Ueda et al. No publicado
ROTIFERA	<i>Adineta ricciae</i>	AKC91349.1	Szydowski et al. 2015
CRUSTACEA	<i>Hyalella azteca</i> , <i>Eurytemora affinis</i> , <i>Cherax quadricarinatus</i> , <i>Parasesarma erythroductyla</i> , <i>Gecarcoidea natalis</i> , <i>Mictyris platycheles</i> , <i>Coenobita brevimanus</i> , <i>Coenobita rugosus</i> , <i>Daphnia pulex</i> , <i>Daphnia magna</i> , <i>Limnoria quadripunctata</i> , <i>Neomysis intermedia</i>	<p>XP_018016806.1, XP_018016809.1, XP_018016810.1, XP_018016813.1, XP_018016820.1, XP_018016821.1, XP_018026496.1, XP_018027134.1, XP_023337121.1, AAD38027.1, AAO61672.1, AIT11911.1, APF29556.1, APF29561.1, EFX69372.1, EFX80604.1, EFX80605.1, EFX86822.1, KZS03923.1, KZS05722.1, KZS07093.1, KZS16514.1, KZS16516.1, ADB85440.1, ADB85442.1, BAL60587.1</p>	Byrne et al. 1999, Crawford et al. 2004, King et al. 2010, Colbourne et al. 2011, Bui & Lee 2014, Gilbert et al. No publicado, Gray & Linton No publicado, Sato et al. No publicado
COLLEMBOLA	<i>Orchesella cincta</i> , <i>Folsomia candida</i>	ODM94862.1, ODM95900.1, XP_021956548.1, XP_021956549.1, XP_021959948.1, XP_021966378.1, XP_021967587.1	Faddeeva-Vakhrusheva et al. 2016
ORTHOPTERA	<i>Teleogryllus emma</i>	ABV32557.1	Kim et al. 2008
EMBIOPTERA	<i>Antipaluria urichi</i>	AOV94250.1, AOV94256.1	Shelomi No publicado
PHASMATODEA	<i>Timema cristinae</i> , <i>Aretaon asperrimus</i> , <i>Extatosoma tiaratum</i> , <i>Medauroidea extradentata</i> , <i>Peruphasma schultei</i> , <i>Ramulus artemis</i> , <i>Sipyloidea sipyulus</i>	AMH40392.1, AMH40395.1, AMH40356.1, AMH40391.1	Shelomi et al. 2016
BLABEROIDEA	<i>Blattella germanica</i> , <i>Panesthia cribrata</i>	PSN31180.1, AAF80584.1, AAF80585.1	Lo et al. 2000, Harrison et al. 2018
TERMITOIDAE	<i>Cryptotermes secundus</i> , <i>Coptotermes acinaciformis</i> , <i>Reticulitermes flavipes</i> , <i>Coptotermes formosanus</i> ,	<p>XP_023704929.1, XP_023704930.1, XP_023716596.1, XP_023718476.1, AAK12339.1, AAU20853.1, ADB12483.1, BAB40693.1,</p>	Watanabe et al. 1998, Tokuda et al. 1999, Cairo et al. 2013, Nakashima et al. 2002, Li et al. 2003,

	<i>Odontotermes formosanus</i> , <i>Macrotermes barneyi</i> , <i>Amitermes dentatus</i> , <i>Amitermes foreli</i> , <i>Anoplotermes schwarzi</i> , <i>Constrictotermes cavifrons</i> , <i>Constrictotermes guantanamensis</i> , <i>Grigiotermes hageni</i> , <i>Globitermes sulphureus</i> , <i>Hospitalitermes bicolor</i> , <i>Microcerotermes crassus</i> , <i>Microtermes pallidus</i> , <i>Macrotermes carbonarius</i> , <i>Macrotermes gilvus</i> , <i>Macrotermes subhyalinus</i> , <i>Nasutitermes corniger</i> , <i>Nasutitermes sp. NpEG</i> , <i>Odontotermes hainanensis</i> , <i>Pericapritermes nitobei</i> , <i>Pericapritermes sp. PpEG1</i> , <i>Pericapritermes sp. PpEG2</i> , <i>Pericapritermes sp. PpEG3</i> , <i>Rhynchotermes bulbinasus</i> , <i>Subulitermes baileyi</i> , <i>Syntermes grandis</i> , <i>Sphaerotermes sphaerotherax</i> , <i>Coptotermes gestroi</i> , <i>Microtermes pakistanicus</i> , <i>Reticulitermes speratus</i> , <i>Nasutitermes takasagoensis</i> , <i>Nasutitermes walkeri</i> , <i>Mastotermes darwiniensis</i> , <i>Zootermopsis nevadensis</i>	BAB40695.1, BAB40696.1, BAB40697.1, ADB82658.1, AGP76419.1, AFD33365.1, AGP76397.1, AGP76418.1, AGP76420.1, AGP76437.1, AGS32241.1, AKV16371.1, BAA31326.1, BAA34050.1, BAA33708.1, BAA33709.1, CAD54726.1, CAD54727.1, CAD54728.1, CAD54729.1, CAD54730.1, KDR16731.1, XP_021924915.1, XP_021935277.1, XP_021941324.1	Scharf et al. 2005, Ni et al. 2011, Terrapon et al. 2014, Yoon et al. 2015, Bujang et al. No publicado, Chen & Mo No publicado, Wu No publicado, Zhang et al. No publicado
HEMIPTERA	<i>Acyrtosiphon pisum</i> , <i>Myzus persicae</i> , <i>Nilaparvata lugens</i>	XP_001944774.2, XP_008181842.1, XP_022167267.1, XP_022167268.1, XP_022200004.1	
PHTHIRAPTERA	<i>Pediculus humanus corporis</i>	XP_002426465.1	Kirkness et al. No publicado
COLEOPTERA	<i>Tribolium castaneum</i> , <i>Nicrophorus vespilloides</i> , <i>Agrilus planipennis</i> , <i>Anoplophora glabripennis</i>	KYB27092.1, XP_001810693.2, XP_008194455.1, XP_015836219.1, XP_017785110.1, XP_018323591.1, XP_018568194.1, XP_018568196.1	Richards et al. 2008
HYMENOPTERA	<i>Apis mellifera</i> , <i>Camponotus floridanus</i> , <i>Acromyrmex echinatior</i> , <i>Lasius niger</i> , <i>Habropoda laboriosa</i> , <i>Melipona quadrfasciata</i> , <i>Atta colombica</i> , <i>Cyphomyrmex costatus</i> , <i>Trachymyrmex cornetzi</i> , <i>Trachymyrmex septentrionalis</i> , <i>Trachymyrmex zeteki</i> , <i>Dufourea novaeangliae</i> , <i>Eufriesea mexicana</i> , <i>Apis cerana cerana</i> , <i>Nasonia vitripennis</i> , <i>Bombus terrestris</i> , <i>Apis florea</i> , <i>Apis dorsata</i> , <i>Solenopsis invicta</i> , <i>Fopius arisanus</i> , <i>Ooceraea biroi</i> , <i>Ceratosolen solmsi marchali</i> , <i>Pogonomyrmex barbatus</i> , <i>Wasmannia auropunctata</i> , <i>Vollenhovia emeryi</i> , <i>Atta cephalotes</i> , <i>Megachile rotundata</i> , <i>Linepithema humile</i> , <i>Bombus impatiens</i> , <i>Orussus abietinus</i> , <i>Monomorium pharaonis</i> , <i>Copidosoma floridanum</i> , <i>Polistes canadensis</i> , <i>Diachasma alloeum</i> , <i>Polistes dominula</i> , <i>Neodiprion lecontei</i> , <i>Cephus cinctus</i> , <i>Apis cerana</i> , <i>Pseudomyrmex gracilis</i>	XP_396791.3, EFN70196.1, EFN70197.1, XP_011253653.1, XP_011253654.2, XP_011253655.2, XP_019884179.2, EGI63652.1, XP_011057333.1, KMQ95413.1, KOC62382.1, XP_017793200.1, KOX76353.1, KYM87195.1, XP_018044452.1, KYN01696.1, XP_018396635.1, KYN22851.1, XP_018359268.1, KYN35700.1, XP_018347393.1, KYQ52831.1, XP_018307051.1, KZC10616.1, XP_015432538.1, OAD61293.1, XP_017766299.1, PBC31793.1, XP_001606454.1, XP_003402778.1, XP_003690676.1, XP_006610535.1, XP_011164235.1, XP_011296914.1, XP_011296916.1, XP_011346866.1, XP_011498432.1, XP_011632703.1, XP_011700378.1, XP_011867251.1, XP_012061951.1, XP_012146513.1, XP_012224814.1, XP_012248155.1, XP_012273697.1, XP_012526654.1, XP_014214612.1, XP_014601854.1, XP_015126097.1, XP_015184471.1, XP_015513382.1, XP_015592927.1, XP_016907626.1, XP_020289295.1, XP_020289297.1, XP_020289301.1, XP_020289336.1	Bonasio et al. 2010, Diao et al. No publicado, Konorov et al. No publicado, Nygaard & Zhang No publicado, Nygaard et al. No publicado, Pan & Kapheim No publicado

Las secuencias de celulasas encontradas en el GenBank pertenecen a distintos

grupos de animales tales como: Cnidaria, Lophotrochozoa, Crustacea, Collembola e Insecta. Particularmente este último grupo es el que tiene una mayor representación e incluye secuencias de diversos órdenes de insectos herbívoros, detritívoros y xilófagos tales como: Orthoptera, Phasmatodea, Blaberoidea, Termitoidae, Hemiptera, Coleoptera, Embioptera e Hymenoptera, aunque también incluye ordenes de insectos hematófagos tal como el orden Phthiraptera. Es importante mencionar que algunos órdenes sólo están representados por 1 a 3 especies, por ejemplo: Orthoptera, Blaberoidea, Hemiptera, Coleoptera, Phthiraptera y Embioptera.

7.2 Alineamiento

El alineamiento obtenido (Anexo 3; Alineamiento 1) muestra varias zonas conservadas de las secuencias de celulasas de la familia GH9. La mayoría de las secuencias obtenidas presentan los sitios catalíticos y el sitio donador de protones, sin embargo 30 secuencias carecen de alguno de estos sitios o presentan una mutación en estos aminoácidos (Tabla 5). Es importante mencionar que estas secuencias provienen de anotación de genomas.

Tabla 5. Secuencias de celulasas con mutaciones o deleciones en los sitios catalíticos.

Numero de Acceso en NCBI	Especie	Grupo	Mutación en el Aminoácido. Catalítico primario; Aminoácido original Ac. Aspártico (D)	Mutación en el Aminoácido. Catalítico secundario; Aminoácido original Ac. Aspártico (D)	Mutación en el Aminoácido o donador de protones; Aminoácido original Ac. Glutámico (E)
PFX25363.1	<i>Stylophora pistillata</i>	Cnidaria	E		
BAU45388.1	<i>Bellamyia sp.</i>	Mollusca			D
XP_013066277.1	<i>Biomphalaria glabrata</i>	Mollusca			R
XP_013084279.1	<i>Biomphalaria glabrata</i>	Mollusca	V	G	
XP_013084281.1	<i>Biomphalaria glabrata</i>	Mollusca	V	G	
XP_013084282.1	<i>Biomphalaria</i>	Mollusca	L	A	D

<i>glabrata</i>						
XP_013084283.1	<i>Biomphalaria glabrata</i>	Mollusca	L		A	D
XP_013084285.1	<i>Biomphalaria glabrata</i>	Mollusca				D
XP_013084287.1	<i>Biomphalaria glabrata</i>	Mollusca				D
XP_011430172.1	<i>Crassostrea gigas</i>	Mollusca	N			H
XP_022336001.1	<i>Crassostrea virginica</i>	Mollusca	N			H
XP_022338491.1	<i>Crassostrea virginica</i>	Mollusca	N			H
PVD35384.1	<i>Pomacea canaliculata</i>	Mollusca			G	G
PVD35385.1	<i>Pomacea canaliculata</i>	Mollusca			G	D
PVD36625.1	<i>Pomacea canaliculata</i>	Mollusca		Delección de 9 aminoácidos		
XP_023718476.1	<i>Cryptotermes secundus</i>	Termitodae				I
ADB82658.1	<i>Odontotermes formosanus</i>	Termitodae				V
XP_021935277.1	<i>Zootermopsis nevadensis</i>	Termitodae				M
PSN31180.1	<i>Blattella germanica</i>	Blaberoidea		Delección de 10 aminoácidos		
EFN70196.1	<i>Camponotus floridanus</i>	Hymenoptera			N	G
EFN70197.1	<i>Camponotus floridanus</i>	Hymenoptera			N	G
XP_011253653.1	<i>Camponotus floridanus</i>	Hymenoptera			N	G
XP_011253654.1	<i>Camponotus floridanus</i>	Hymenoptera			N	S
XP_011253655.1	<i>Camponotus floridanus</i>	Hymenoptera			N	G
XP_019884179.1	<i>Camponotus floridanus</i>	Hymenoptera			N	N
XP_011346866.1	<i>Ooceraea biroii</i>	Hymenoptera				D
XP_020289295.1	<i>Pseudomyrmex gracilis</i>	Hymenoptera	N		N	N
XP_020289297.1	<i>Pseudomyrmex gracilis</i>	Hymenoptera	N		T	K
XP_020289301.1	<i>Pseudomyrmex gracilis</i>	Hymenoptera	Y		S	K
XP_020289336.1	<i>Pseudomyrmex gracilis</i>	Hymenoptera	N		N	

7.3 Análisis filogenéticos

El modelo de evolución molecular que mostró un mejor ajuste para los tres alineamientos fue el modelo WAG + G + I (Goldman And Whelan) (Goldman & Whelan 1995). Los análisis filogenéticos de las secuencias de celulasas de la familia GH9 usando los tres alineamientos y dos métodos estadísticos distintos por cada alineamiento (Bayesiano y Máxima Verosimilitud) resultaron en seis árboles

con topologías similares en todos los casos. Los tres árboles obtenidos por método bayesiano tienen un mayor número de politomías, así como valores de soporte más altos en comparación con los árboles generados por método de máxima verosimilitud.

7.3.1 Árbol Filogenético 1

El árbol filogenético 1 (Fig. 9) es el resultado de unir los valores de soporte obtenidos por método bayesiano y máxima verosimilitud sobre la topología obtenida en Mr. Bayes, usando el alineamiento que contiene una secuencia por especie (Anexo 1; Tabla 1) y la secuencia de celulasa de *Cellulomonas fimi* (Bacteria) como grupo externo. Este análisis reconoce un origen monofilético para los Eumetazoa, con dos clados principales que presentan un origen monofilético: Cnidaria/Lophotrochozoa (Mollusca y Annelida) (99/71, valores obtenidos por método Bayesiano y Maxima Verosimilitud respectivamente) y animales Protostomados (100/73).

El clado Cnidaria/Lophotrochozoa se divide en dos clados: como clado basal se reconoce a las secuencias provenientes del filo Cnidaria (100/98), mientras el otro clado (99/60) se divide en las secuencias provenientes de los filos Annelida (100/100) y Mollusca (100/81).

El clado de los Protostomados se subdivide a su vez en dos clados: el primero formado por secuencias provenientes de los grupos Rotífera, Collembola (Artrópoda: Hexápoda) y el crustáceo *Daphnia magna* (100/51), y el segundo clado (61/17) está conformado por secuencias por el grupo Pancrustacea (Artrópoda) que incluye las secuencias de celulastas de insectos. La mayoría de las secuencias de crustáceos que se agrupan en un solo clado (89/37), con excepción de una secuencia de *D. pulex*, la cual se agrupa con las secuencias de celulastas de insectos del super orden Polynoptera (76/17).

Las celulasas de insectos se agrupan en dos clados (100/63 y 100/100), un clado basal conformado por las secuencias provenientes del superorden Polyneoptera (Grupo 1) y otro clado conformado por secuencias provenientes de los superórdenes Polyneoptera, Holometabola y Paraneoptera (Grupo 2). Las celulasas de termitas (Termitoidea: Polyneoptera) se encuentran en ambos clados, mientras que las celulasas de los insectos del superórden Paraneoptera (Phthiraptera y Hemiptera) no tienen un origen monofilético.



Figura 9.- Filogenia de celulasas de la familia GH9 utilizando el alineamiento 1 (Anexo 1; Tabla 1). La filogenia muestra la topología generada por método Bayesiano, los valores de soporte (Bayesiano/Máxima verosimilitud) se muestran sobre las ramas; "*" indica que el clado no es soportado por el análisis. La barra de escala representa el número de sustituciones por sitio. Los grupos taxonómicos que forman clados fueron colapsados. Árbol editado en TreeGraph 2 (Stöver et al. 2010), visualizado en FigTree v1.4 (Rambaut 2012) y figura editada en GIMP 2.10 (Gimp Team).

7.3.2 Árbol Filogenético 2

En términos generales, las relaciones filogenéticas de los grupos en ambos análisis (una secuencia y todas las secuencias por especie; Figs. 9 y 10) son similares. Los resultados del segundo análisis filogenético (Fig. 10) indican que además de Termitoidea, algunas secuencias de Embioptera y Phasmatodea se agrupan en ambos clados de celulasas de insectos.

El segundo análisis que usa como grupo externo las secuencias de celulasas del taxón Cnidaria reconoce la monofilia del grupo Protostomia (100/99). En este análisis se identifican dos subclados dentro de Pancrustacea + Rotifera: Rotífera/Collembola/Crustacea (100/61) y Pancrustacea (Crustacea/Insecta; 51/27). En este clado, las celulasas de la mayoría de los crustáceos forman un grupo (100/100), aunque varias secuencias provenientes del género *Daphnia* se agrupan en el clado de las celulasas de insectos Grupo 1 (74/19).

En este análisis, también se identifican dos clados para las celulasas de insectos. El primer clado (Grupo 1) conformado por secuencias del superorden Polyneoptera (100/99) (Termitoidea, Blaberoidea, Orthoptera, Embioptera y Phasmatodea), para el cual Phasmatodea presenta la mayor diversidad de secuencias (6 linajes), seguido de Embioptera (2 linajes). En este clado las secuencias de Termitoidea y Blaberoidea que pertenecen al mismo orden (Blattaria), forman linajes independientes con valores de soporte 100/90 y 97/61 respectivamente.

Por su parte, el segundo clado (Grupo 2) está conformado por los superordenes Paraneoptera (Hemiptera y Phthiraptera), Holometabola (Coleoptera e Hymenoptera) y Polyneoptera (Termitoidea, Embioptera y Phasmatodea). En este grupo las secuencias que presentan mayor divergencia (mayor cantidad de mutaciones y ramas largas) pertenecen al orden Hymenoptera. Al igual que en el primer análisis, este análisis tampoco reconoce el origen monofilético de las

celulasas de insectos.

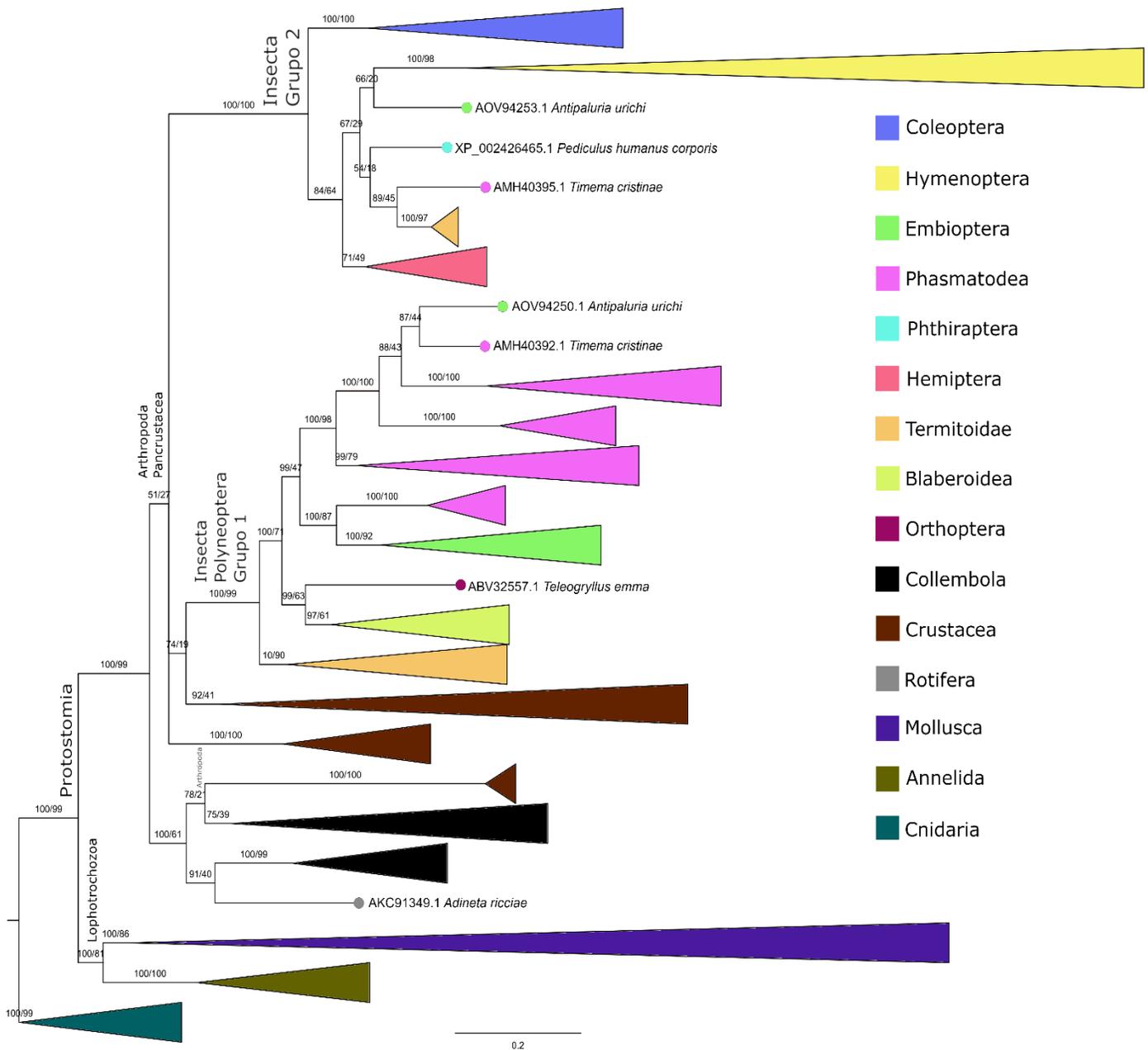


Figura 10.- Filogenia de celulasas de la familia GH9 utilizando el alineamiento 2 (Tabla 4). La filogenia muestra la topología generada por método Bayesiano, los valores de soporte (Bayesiano/Maxima verosimilitud) se muestran sobre las ramas. La barra de escala representa el número de sustituciones por sitio. Los grupos taxonómicos que forman clados fueron colapsados. Árbol editado en TreeGraph 2 (Stöver et al. 2010), visualizado en FigTree v1.4 (Rambaut 2012) y figura editada en GIMP 2.10 (Gimp Team).

7.3.3 Árbol Filogenético 3

En este análisis, nuevamente se recupera la formación de dos clados de celulasas de insectos (Fig. 11), con las mismas secuencias que en los análisis anteriores y con valor de soporte de 100 en ambos clados. Sin embargo, algunos valores de soporte para relaciones filogenéticas interiores son más bajos que los valores obtenidos en los análisis que incluyen secuencias de crustáceos.

Dentro del clado Polyneoptera (Grupo 1) hay dos grupos reconocibles, uno conformado por secuencias de la familia Termitoidae, el cual funge como clado basal (69/82), mientras que el resto de secuencias (90/28) forma dos clados compuestos por las secuencias provenientes de los taxa Blaberoidea/Ortoptera (91/27) y Embioptera/Phasmatodea (90/28).

Dentro del clado conformado por secuencias de los superórdenes Polyneoptera, Holometabola y Paraneoptera (Grupo 2), se identifican dos subclados: un clado basal compuesto por secuencias provenientes del orden Coleoptera (100/100), y un segundo clado formado a su vez por dos grupos. Un grupo basal con origen monofilético compuesto por secuencias provenientes del orden Hemiptera (85/50), y un grupo interno que incluye secuencias de los órdenes Hymenoptera, Embioptera, Phthiraptera, Termitoidae y Phasmatodea (60/26). A diferencia de los análisis anteriores, la secuencia del Phthiraptero *Pediculus humanus* no se relaciona con ninguna secuencia en este grupo.

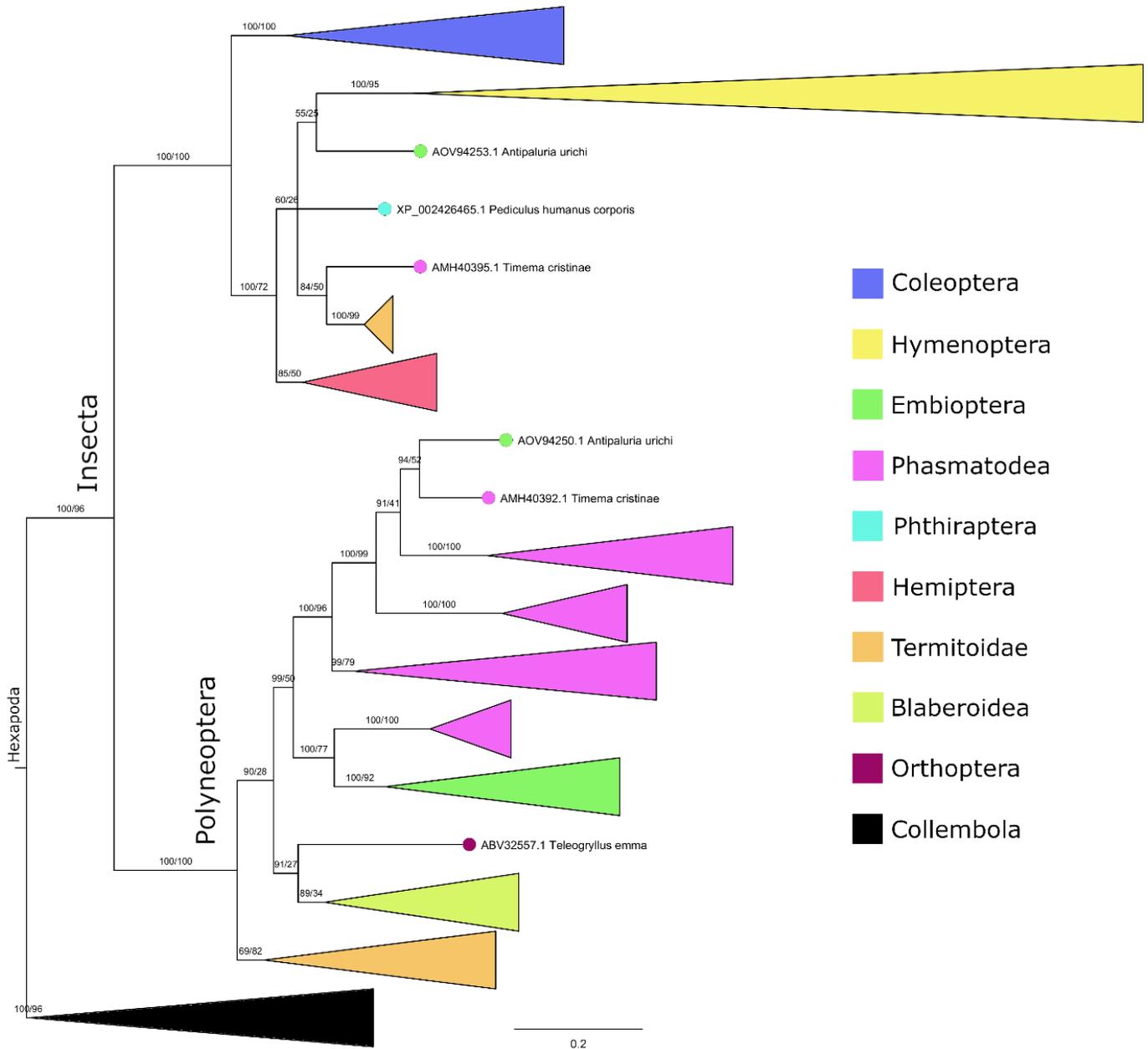


Figura 11.- Filogenia de celulasas de la familia GH9 utilizando el alineamiento 3 (Anexo 1; Tabla 2). La filogenia muestra la topología generada por método Bayesiano, los valores de soporte (Bayesiano/Maxima verosimilitud) se muestran sobre las ramas. La barra de escala representa el número de sustituciones por sitio. Los grupos taxonómicos que forman clados fueron colapsados. Árbol editado en TreeGraph 2 (Stöver et al. 2010), visualizado en FigTree v1.4 (Rambaut 2012) y figura editada en GIMP 2.10 (Gimp Team).

7.4 Análisis estructural

De acuerdo con los resultados del gráfico de Ramachandran (Tabla 6), los modelos refinados por ambos métodos presentan ligeras diferencias entre ellos y entre su versión sin refinar. Los modelos sin refinar presentan un mayor número de modelos (33 modelos) con porcentajes de aminoácidos ≤ 0.5 en zonas no permitidas y menor promedio (0.37%) respecto a los modelos refinados con métodos ligero y fuerte (19 y 14 modelos; 0.6% y 0.68%, respectivamente). Con base en estos resultados, se decidió utilizar los modelos sin refinar para los análisis posteriores.

Tabla 6.- Resultados de los análisis del gráfico de Ramachandran aplicado a los modelos sin refinar, refinados con método ligero y refinados con método fuerte.

No de acceso del NCBI	Modelos sin refinar			Modelos refinados con método ligero			Modelos refinados con método fuerte		
	Porcentaje de aminoácidos en zonas			Porcentaje de aminoácidos en zonas			Porcentaje de aminoácidos en zonas		
	Optima	Aceptable	No permitida	Optima	Aceptable	No permitida	Optima	Aceptable	No permitida
AAF80584.1	96.9	2.6	0.2	94.0	5.5	0.5	93.6	5.7	0.7
ABV32557.1	95.0	3.3	0.9	92.9	6.0	1.2	92.4	6.4	1.2
ADB85440.1	95.3	3.1	0.5	93.3	5.8	1.0	92.8	6.3	1.0
AGP76427.1	94.7	2.3	0.2	95.9	4.1	0.0	95.5	4.5	0.0
AKC91349.1	92.5	3.4	0.2	94.5	5.2	0.2	94.5	5.0	0.5
AKV16371.1	92.5	2.7	0.5	95.0	5.0	0.0	93.6	5.7	0.7
AMH40372.1	92.7	2.7	0.0	93.3	6.0	0.7	94.5	5.0	0.5
AMH40373.1	91.3	3.0	0.9	95.0	4.3	0.7	93.3	5.5	1.2
AMH40379.1	92.1	2.9	0.0	93.8	6.2	0.0	93.6	5.7	0.7
AMH40392.1	91.6	3.0	0.2	94.7	4.8	0.5	93.6	5.7	0.7
AMH40395.1	92.1	2.3	0.2	94.3	5.5	0.2	93.8	6.2	0.0
AOV94250.1	90.1	3.4	0.0	93.0	6.3	0.7	93.8	5.3	1.0
AOV94253.1	90.7	2.4	0.2	94.5	5.0	0.5	92.8	6.4	0.7
AOV94256.1	89.2	4.0	0.0	92.9	6.9	0.2	92.2	7.3	0.5
APF29560.1	88.6	3.1	1.1	93.6	5.2	1.2	93.6	5.7	0.7
BAA31326.1	96.9	2.9	0.2	95.5	4.5	0.0	93.6	5.7	0.7
BAK20401.1	95.1	2.3	0.0	93.4	5.6	1.0	92.7	6.8	0.5
CAD54726.1	95.7	3.1	0.7	94.3	4.1	1.7	93.8	4.5	1.7
EFN70196.1	94.7	3.0	0.9	91.0	7.8	1.2	92.0	6.9	1.2

EFX69372.1	94.6	3.3	0.5	94.0	5.3	0.7	94.3	5.0	0.7
EFX86822.1	94.4	3.5	0.2	93.6	6.1	0.2	93.6	6.1	0.2
KDR16731.1	93.8	2.5	0.5	93.1	6.4	0.5	92.4	6.9	0.7
KXJ18436.1	90.6	3.6	0.2	91.2	8.1	0.7	92.8	6.0	1.2
KZC10616.1	89.7	3.6	0.9	92.9	5.9	1.2	93.1	5.9	1.0
ODM94862.1	87.8	4.9	1.1	91.7	7.3	0.9	91.0	8.0	0.9
PSN31180.1	86.5	3.9	0.5	91.6	6.8	1.5	93.6	5.7	0.7
XP_00194477 4.1	93.9	3.5	0.5	94.0	5.7	0.2	92.6	7.2	0.2
XP_00242646 5.1	96.5	2.6	0.0	93.3	6.4	0.2	91.9	7.9	0.2
XP_00819445 5	93.6	3.2	0.2	92.9	6.4	0.7	92.7	6.9	0.5
XP_00904979 1.1	92.1	3.7	0.7	94.5	4.8	0.7	93.6	5.7	0.7
XP_01129691 4.1	94.1	1.8	0.5	94.8	4.5	0.7	93.6	5.7	0.7
XP_01306627 7.1	93.7	1.9	0.2	93.4	5.6	1.0	92.7	6.8	0.5
XP_01778511 0.1	91.3	2.2	0.4	94.3	4.8	1.0	93.8	5.7	0.5
XP_01856819 4.1	89.1	3.5	0.4	93.9	5.2	0.9	92.5	6.8	0.7
XP_02061486 7.1	89.9	2.6	0.0	94.8	5.2	0.0	93.6	5.7	0.7
XP_02195654 8.1	88.8	3.1	0.4	94.3	5.0	0.7	94.0	5.2	0.7
XP_02216726 7.1	88.4	3.3	0.4	93.1	6.4	0.5	93.6	5.7	0.7
XP_02220000 4.1	87.5	4.2	0.2	92.6	6.4	1.0	92.1	6.9	1.0
XP_02370492 9.1	88.9	2.6	0.0	97.6	2.4	0.0	96.2	3.8	0.0
XP_396791.3	94.7	2.8	0.2	92.6	6.9	0.5	91.9	7.4	0.7
XP_02371659 6.1	89.1	2.4	0.2	94.3	5.3	0.5	93.6	6.0	0.5

La visualización de la superposición indica similitud estructural, aunque existen algunas zonas que presentan diferencias notorias entre los modelos (Fig. 12). Estas zonas están representadas principalmente por giros-c y β -hélices se encuentran localizadas en zonas particulares localizadas con las que cuentan algunas secuencias, lo que se ve reflejado en la estructura terciaria, en la cual el dominio catalítico se mantiene conservado en todos los modelos, así como la posición de los aminoácidos catalíticos.

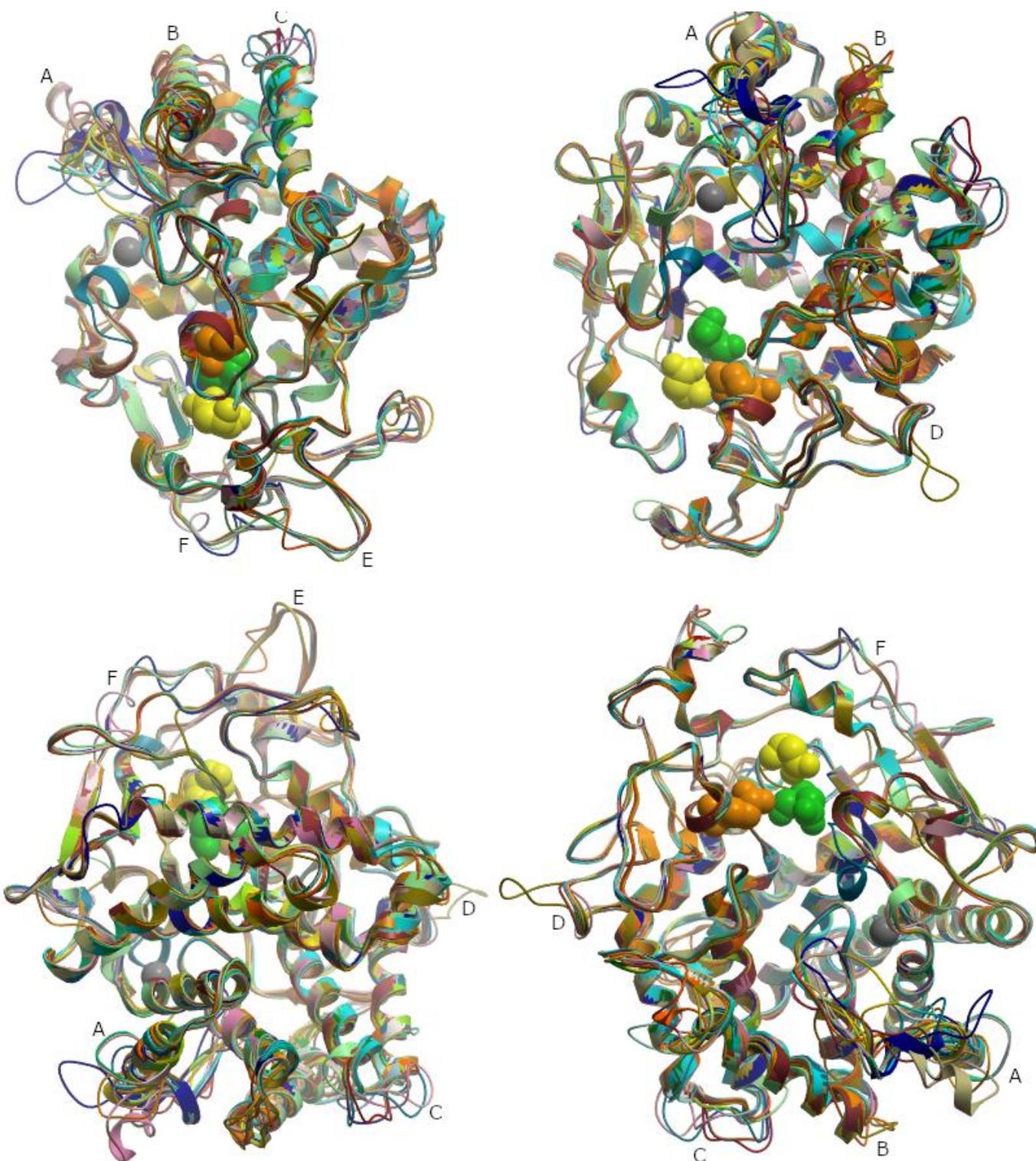


Figura 12 Sobreposición de modelos generados por homología (Tabla 3), los aminoácidos que participan en la actividad enzimática están resaltados de color verde (aminoácido catalítico), amarillo (probable aminoácido catalítico secundario) y naranja (aminoácido). Las zonas que presentan variabilidad están señaladas con letras (A-F) y el calcio de la enzima cristalizada está representado como una esfera color gris. Superposición, visualización y captura de imágenes de los modelos se realizó en el software Molsoft ICM-Browser (Molsoft LLC), la edición de las imágenes se realizó en GIMP 2.10 (Gimp Team).

Los valores de RMSD obtenidos indican que las diferencias estructurales entre los

modelos varían de 0.3 a 2.0 Å (Anexo 1, Tabla 3). La representación de estos valores en un dendrograma (Fig. 13), muestra que estas diferencias no corresponden a las relaciones filogenéticas obtenidas, ni a los hábitos alimenticios o dietas de los organismos. El modelo que presenta la mayor diferencia estructural es el generado a partir de la secuencia proveniente de *Blattella germanica*, la cual cuenta con una inserción de 669 aminoácidos en el extremo amino.

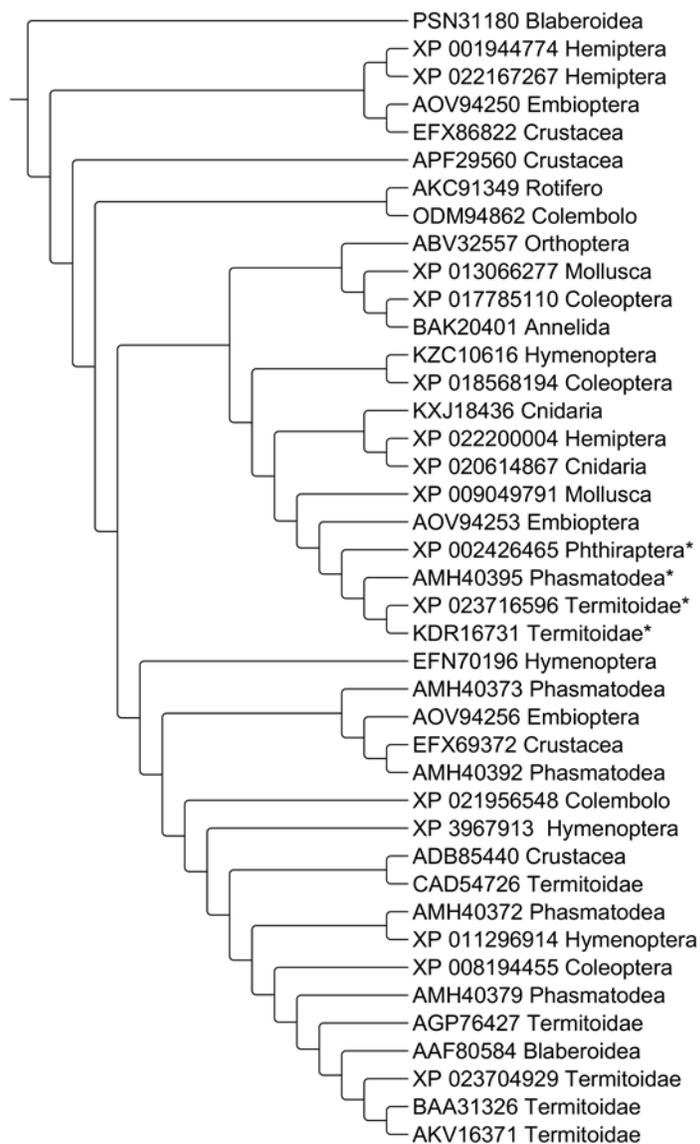


Figura 13 Dendrograma de los valores de RMSD de los modelos. Las hojas resaltadas con un “*” representan a los modelos del grupo Polyneoptera pertenecientes al grupo 2 visualizado en el software TreeGraph2(Rambaut 2012) y figura editada en GIMP 2.10 (Gimp Team).

7.5 Análisis de sitios de unión a iones

El resultado del análisis de los sitios de unión a iones, indica que algunas de las secuencias de celulasas de la familia GH9 presentan sitios probables de unión a iones metálicos (calcio y zinc) (Anexo 1; Tabla 4). De acuerdo con los resultados, 208 secuencias presentan sitios de unión a ambos iones, 14 secuencias presentan solo sitio de unión al calcio, 2 secuencias presentan solo sitio de unión al zinc y 125 secuencias no presentan sitios de unión a iones metálicos (Anexo 1; Tabla 4). El mapeo de los resultados muestra que la presencia de sitios probables de unión a iones metálicos no es consistente dentro de los clados, debido a que secuencias provenientes de un mismo taxón difieren en los sitios de unión.

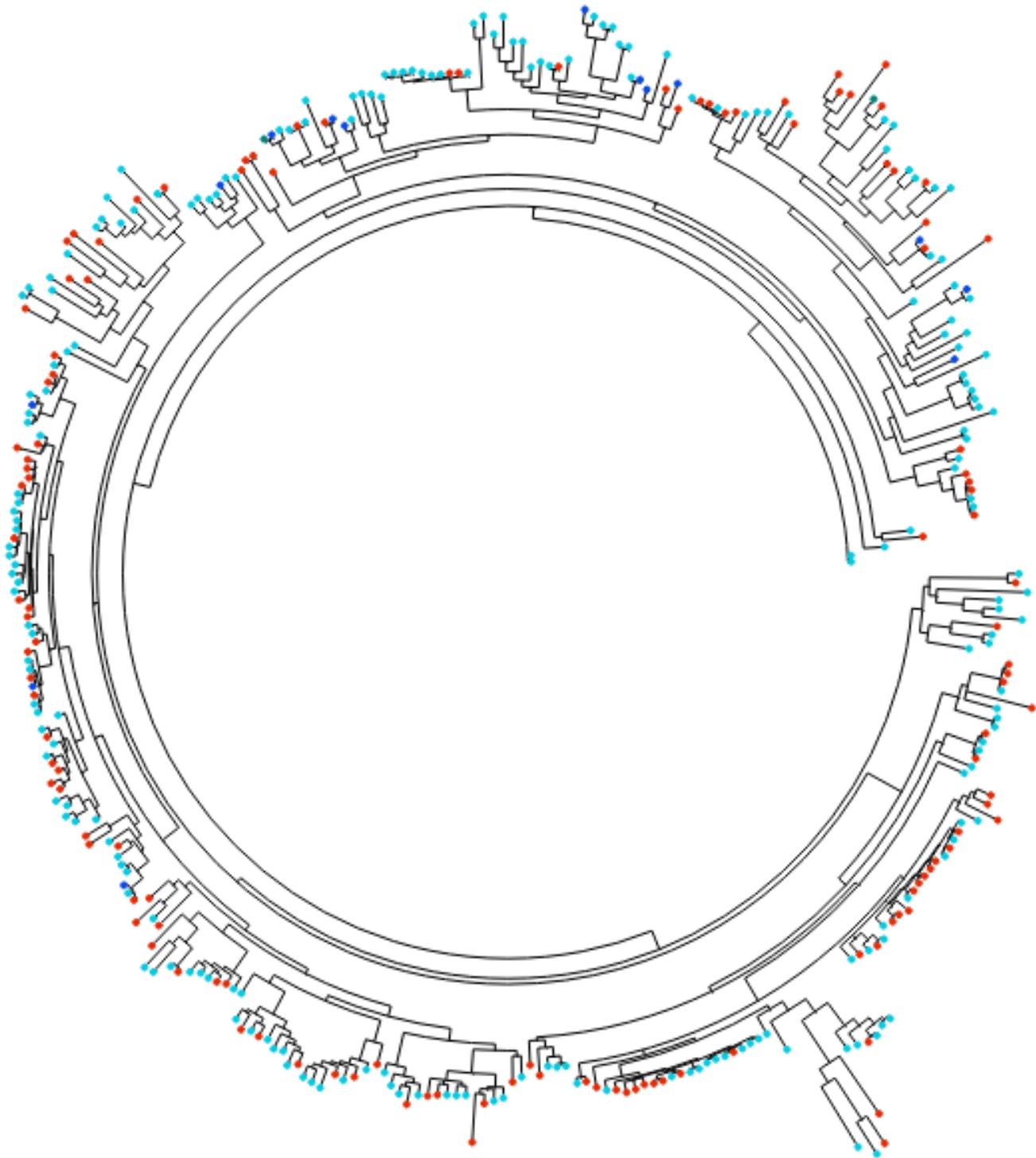


Figura 14 Resultados del análisis de los sitios de unión a calcio y zinc, representados sobre la filogenia obtenida por método estadístico Bayesiano realizado en Mr. Bayes usando el alineamiento 2, generada y visualizada en el lenguaje de programación Python y la librería Ete2. El color azul marino indica que la secuencia presenta sitio de unión a calcio, el color cian indica que la secuencia presenta sitio de unión a zinc, el color azul claro indica que la secuencia presenta sitio de unión a calcio y zinc y el color rojo indica que la secuencia no presenta sitios de unión a iones metálicos. Figura editada en GIMP 2.10 (Gimp Team).

7.6 Análisis estructural de celulasas de insectos

7.6.1 Evolutionary Trace

El análisis realizado con la herramienta evolutionary trace sobre secuencias representativas de celulasas de la familia GH9 de hexapodos, mostró como grupo basal la secuencia del colémbolo *Orchesella cincta* y la formación de dos clados (Fig. 15) al igual que en los análisis filogenéticos (Figs. 9 a 11). La comparación de las secuencias consenso obtenidas en la segunda partición del análisis, corresponde a la formación de los dos grupos principales de las celulasas de insectos. Esta comparación indica la presencia de una mutación en la posición 393 de la estructura cristalizada de *Nasutitermes takasagoensis* (Fig. 16), dado que en el Grupo 1 de insectos este aminoácido es una glicina, mientras que en el Grupo 2 cambia a una serina.

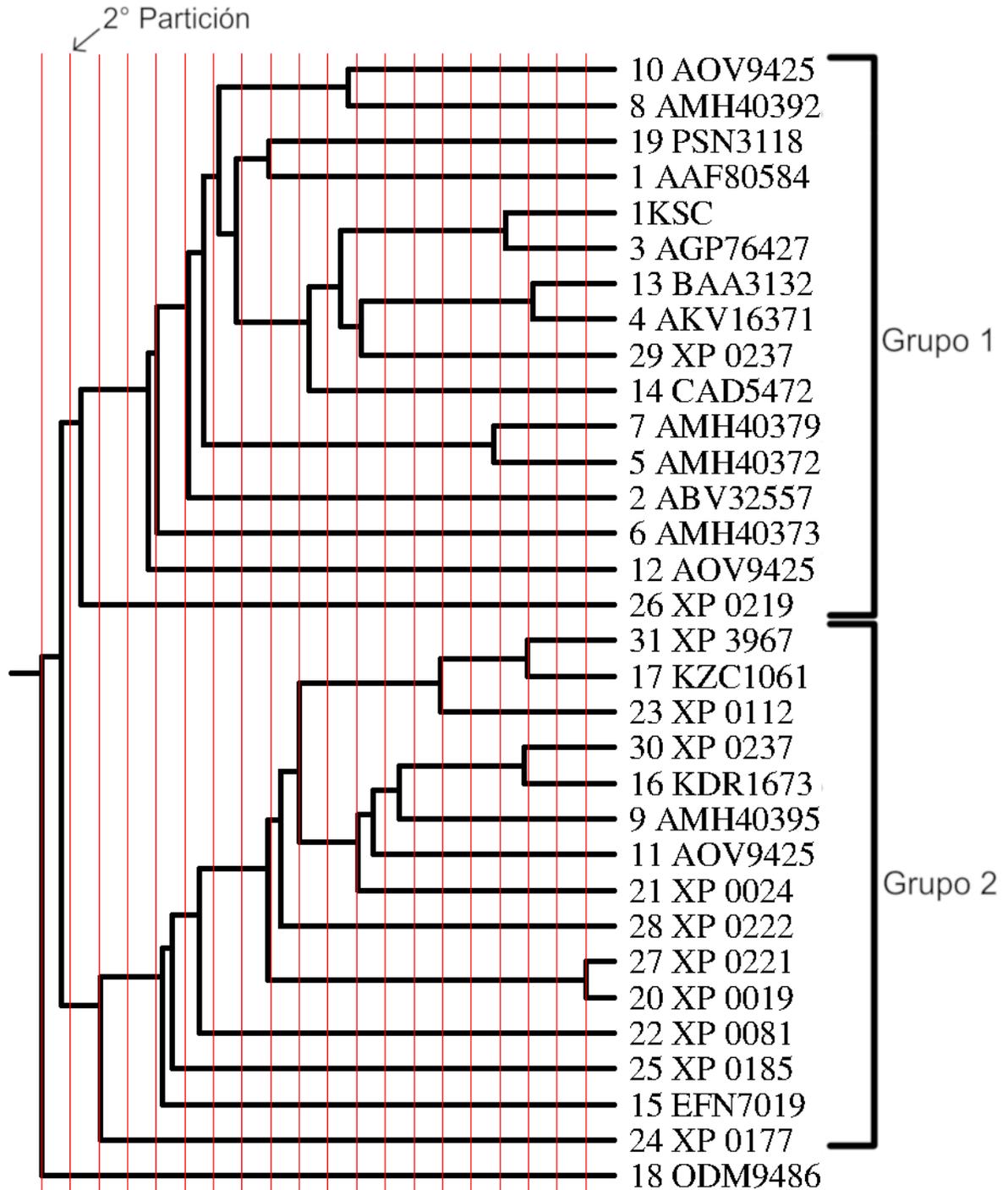


Figura 15.- Filogenia resultado del análisis realizado en Evolutionary Trace, las líneas rojas representan las distintas particiones realizadas en el programa, la segunda partición de la cual se forman los dos clados está marcada con una flecha. Las secuencias usadas en este análisis, la especie de la que provienen y el número asignado por el análisis se encuentran presentes en la tabla 5 del anexo 1 (Anexo 1; Tabla 5). Figura editada en GIMP 2.10 (Gimp Team).

A

BAA31326.1	G	A	L	V	G	G	P	D	S	N	D	S	Y	T	D
AAF80584.1	G	A	L	V	G	G	P	G	Q	N	D	D	Y	E	D
CAD54726.1	G	A	L	V	G	G	P	D	S	N	D	N	Y	E	D
ABV32557.1	G	A	L	V	G	G	P	D	A	N	D	Y	Y	N	D
XP_002426465.1	G	A	L	V	S	G	P	D	E	N	D	Y	Y	E	D
EFN70196.1	G	A	L	V	S	G	P	N	E	N	D	F	F	Y	D
XP_001944774.2	G	A	L	V	S	G	P	D	E	N	D	N	Y	K	D
XP_396791.3	G	A	L	V	S	G	P	D	E	A	D	K	F	H	D
AGP76427.1	G	A	L	V	G	G	P	D	Q	N	D	N	Y	E	D
XP_008194455.1	G	A	L	V	S	G	P	D	Q	N	D	H	Y	E	D
KDR16731.1	G	A	L	V	S	G	P	D	E	N	D	H	Y	M	D
XP_011296914.1	G	A	L	V	S	G	P	D	E	A	D	K	F	R	D
AKV16371.1	G	A	L	V	G	G	P	D	S	N	D	S	Y	T	D
AMH40372.1	G	A	M	V	G	G	P	D	S	N	D	G	Y	T	D
AMH40373.1	G	A	L	V	G	G	P	G	A	D	D	Y	T	D	
AMH40379.1	G	A	M	V	G	G	P	D	S	N	D	G	Y	K	D
AMH40392.1	G	A	L	V	G	G	P	G	S	N	D	D	Y	T	D
AMH40395.1	G	A	L	V	S	G	P	D	E	N	D	N	Y	K	D
KZC10616.1	G	A	L	V	S	G	P	D	E	A	D	K	F	H	D
XP_017785110.1	G	A	L	V	S	G	P	S	Q	I	D	H	Y	E	D
ODM94862.1	G	A	L	V	G	G	P	N	R	Q	D	Q	Y	N	D
AOV94250.1	G	A	L	V	G	G	P	G	Q	N	D	D	W	T	D
XP_018568194.1	G	A	L	V	S	G	P	D	Q	N	D	H	Y	E	D
XP_021956548.1	G	A	L	V	G	G	P	G	P	N	D	D	Y	V	D
XP_022167267.1	G	A	L	V	S	G	P	D	E	N	D	N	Y	K	D
XP_022200004.1	G	A	L	V	S	G	P	D	E	N	D	L	Y	E	D
XP_023716596.1	G	A	L	V	S	G	P	D	E	N	D	Y	Y	K	D
XP_023704929.1	G	A	L	V	G	G	P	D	S	S	D	N	Y	K	D
PSN31180.1	G	A	L	V	G	G	P	G	A	N	D	D	Y	Q	D
1KSC	G	A	L	V	G	G	P	D	Q	N	D	N	Y	V	D

B

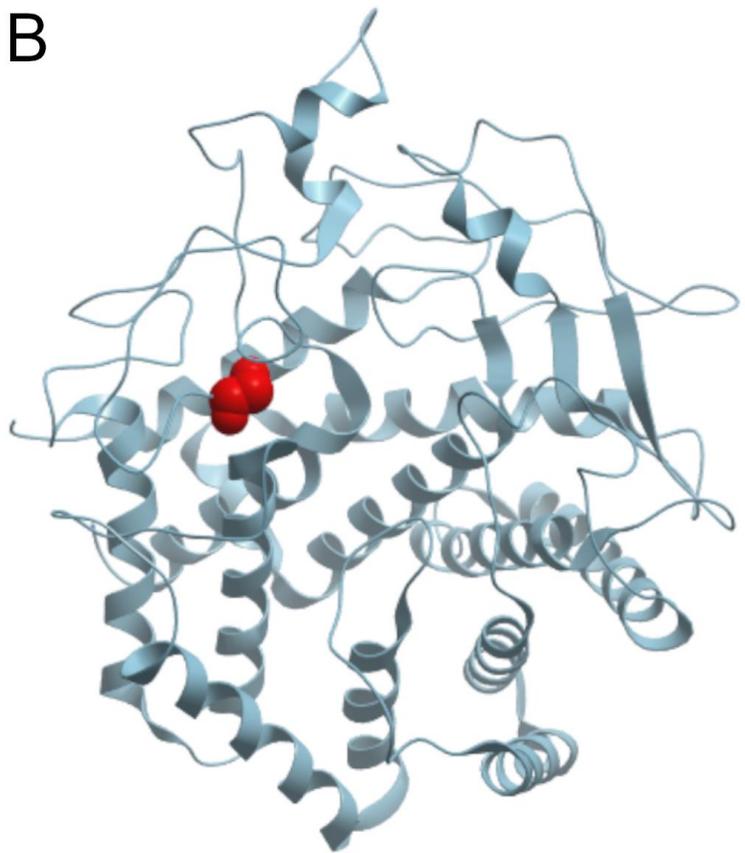


Figura 16. Resultados de Evolutionary Trace. A.- Parte del alineamiento usado para realizar el análisis en Evolutionary Trace, el aminoácido que presenta mutación en el grupo 2 esta resaltado con un "*" sobre el alineamiento. B.- Estructura de la celulasa cristalizada de *Nasutitermes takasagoensis*, en la que el aminoácido que presenta mutación en el grupo 2 esta resaltado en color rojo. Figura editada en GIMP 2.10 (Gimp Team).

7.6.2 Análisis de hidrofobicidad y cavidad

La visualización de hidrofobicidad también permitió analizar las diferencias presentes en la cavidad de los modelos cuyas secuencias pertenecen al grupo 2 de insectos. Los resultados indican que la cavidad de los modelos del grupo 2 es más larga y presenta una hendidura con propiedades hidrofóbicas en la parte superior de esta (Fig. 16: B, D, E y H) en comparación con sus contrapartes del grupo 1 (Fig. 17: A, C, E y G).

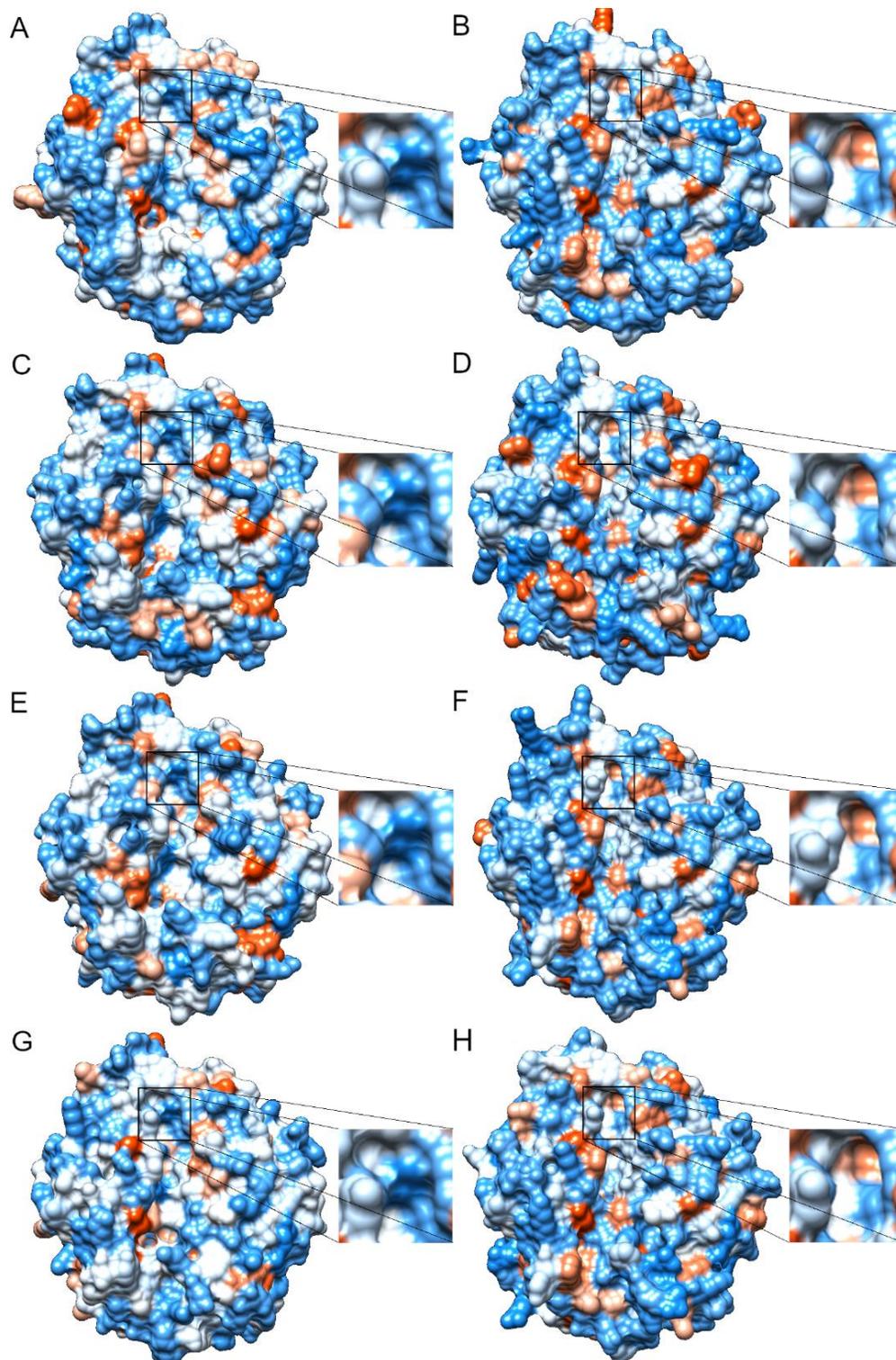


Figura 17.- Análisis de hidrofobicidad superficial de celulasas de la familia GH9. Se muestran la vista de la cavidad. Las imágenes A, C, E y G pertenecen a modelos de celulasas de organismo pertenecientes al grupo Polyneoptera mientras que las imágenes B, D, F y H representan al grupo 2. Estructura cristalizada de *Nasutitermes takasagoensis* (A), *Apis mellífera* (B), *Antipaluria urichi* (C, D), *Timema cristinae* (E, F), *Zootermopsis nevadensis* (G, H). El color rojo representa hidrofobicidad, mientras que el color azul representa zonas hidrofílicas. Figura editada en GIMP 2.10 (Gimp Team)

7.6.3 Análisis de potencial electrostático

En la visualización del potencial electrostático de la superficie de las enzimas realizada sobre los modelos, se encontró una carga altamente negativa en la zona de la cavidad, siendo esta la zona más negativa de todos los modelos analizados. En la zona opuesta a la cavidad se presentan los valores menos electronegativos (Fig. 18). En general, los resultados indican que no existen patrones que diferencien los dos grupos de modelos de celulasas, dado que los valores de carga electrostática son variables en cada uno.

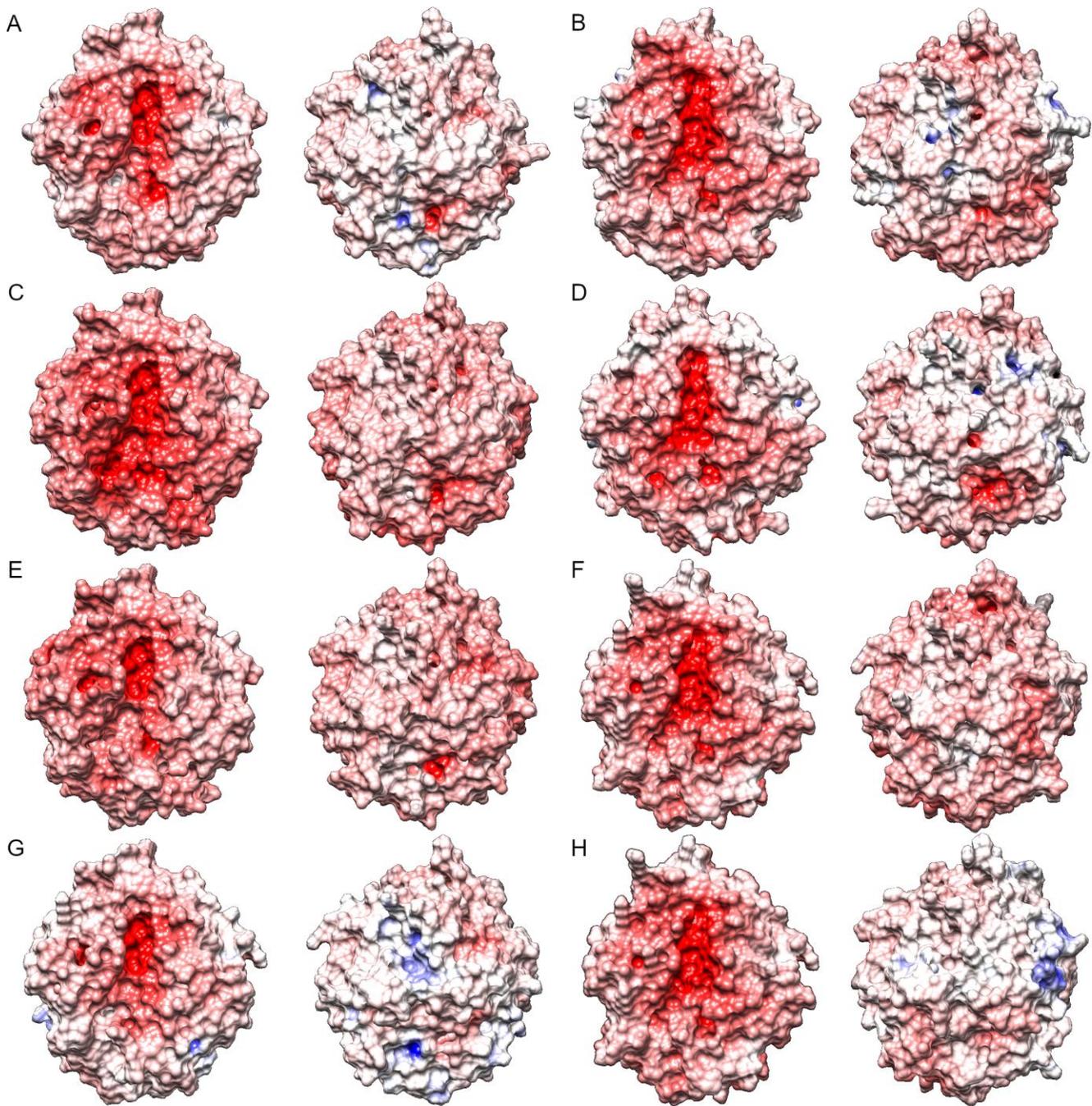


Figura 18.- Potencial electrostático de celulasas de la familia GH9. Se muestran el potencial electrostático de cada enzima en las vistas frontales incluyendo la cavidad (izquierda) y vista trasera (derecha), donde el potencial electrostático va de una escala de -15 (rojo) a 4 (azul) julios. Las imágenes A, C, E y G pertenecen a modelos de celulasas de organismo pertenecientes al grupo Polyneoptera mientras que las imágenes B, D, F y H representan al grupo 2. Estructura cristalizada de *Nasutitermes takasagoensis* (A), *Apis mellifera* (B), *Antipaluria urichi* (C, D), *Tinema cristinae* (E, F), *Zootermopsis nevadensis* (G, H).. Figura editada en GIMP 2.10 (Gimp Team).

8. Discusión

La capacidad de digerir celulosa se ha propuesto como una adaptación clave para la evolución y diversificación de los insectos (Martin et al. 1991, Pauchet et al. 2010 Tokuda 2010, Calderón-Cortés et al. 2012, Pauchet et al. 2014, Shelomi et al. 2014). Sin embargo, debido a la diversidad de celulasas que pertenecen a diferentes familias GH con orígenes evolutivos independientes, a la distribución irregular de estas enzimas en varios grupos de animales, así como a eventos de transferencia horizontal de genes reportados para algunas familias (Calderón-Cortés et al. 2012, Pauchet et al. 2014), el origen y evolución de la capacidad de degradación de celulosa en insectos no están claramente establecidos.

En el presente trabajo se analizó el origen y evolución de celulasas de la familia GH9. De manera general y de acuerdo con la hipótesis planteada, los resultados de este estudio indican que los insectos poseen un mecanismo ancestral de degradación de celulosa que depende de enzimas de la familia GH9 heredado a partir de un ancestro eumetazoario de acuerdo a lo reportado en estudios previos (Lo et al. 2003, Davison & Blaxter 2005). Sin embargo, la historia evolutiva de estas enzimas es compleja y parece involucrar la divergencia temprana de dos clados de enzimas en los insectos.

8.1 Origen y evolución de celulasas GH9 en animales

Los análisis filogenéticos de 421 aminoácidos, los cuales corresponden al 100% de los aminoácidos reportados para el dominio catalítico de enzimas de la familia GH9 (<https://www.ncbi.nlm.nih.gov/>)

En el primer análisis filogenético dentro del grupo Eumetazoa se observan dos grandes clados, el primero está compuesto por secuencias provenientes de los taxa Lophotrochozoa y Cnidaria, mientras que el segundo clado se compone de secuencias provenientes de los taxa Pancrustacea y Rotifera, ambos clados con valores de soporte de 99/71 y 100/73 respectivamente. Cuando las secuencias del filo Cnidaria son usadas como grupo externo también se recupera la formación del clado Protostomia (conformado por los clados Lophotrochozoa y Pancrustacea + Rotifera).

Dentro del grupo Lophotrochozoa se forman dos clados, conformados por taxa Mollusca y Annelida, el clado Mollusca presenta valores de 100/81 y 100/86 en el primer y segundo análisis respectivamente, mientras el clado Annelida presenta valores de soporte de 100/100 en ambos análisis. Sin embargo, la secuencia proveniente de *Adineta ricciae* (Rotifera) no se agrupa con el resto de las secuencias pertenecientes al filo Lophotrochozoa, en cambio se agrupa con las secuencias provenientes de los taxa Collembola y un par de secuencias del género *Daphnia*. Esta relación se podría deber a la baja representación con la que cuenta Rotifera (una secuencia). Esta relación posiblemente errónea, puede deberse a una baja representación del filo Lophotrochozoa (Sanderson & Shaffer 2002).

El clado Pancrustacea (que incluye la secuencia de *Adineta ricciae*) presenta un origen monofilético y valores de soporte de 100/73 y 100/99 en el primer y segundo análisis respectivamente, además es el clado con el mayor número de secuencias, en el que se encuentran representados los grupos Hexapoda y Crustacea. Sin embargo, en ninguno de los análisis se recupera la monofilia reportada previamente para hexápodos (Collembola e Insecta) (Kjer et al. 2006, Xie et al. 2009, Von Reumont et al. 2012, Legg et al. 2013, Regier et al. 2017), debido a que las secuencias de colémbolos e insectos se agrupan en distintos clados.

De acuerdo con la literatura, las relaciones evolutivas dentro del subfilo Crustacea son difíciles de esclarecer, por ejemplo en algunos estudios la clase Remipedia se recupera como grupo hermano de Hexapoda (Von Reumont et al. 2012, Legg et al. 2013), mientras que en otros estudios el grupo hermano de Hexapoda es Branchiopoda (Kjer et al. 2006, Xie et al. 2009, Regier et al. 2017). Por lo anterior, es posible que la inclusión o exclusión de algunas secuencias de crustáceos en los análisis expliquen las relaciones obtenidas en este estudio dentro del clado Pancrustacea. Ejemplo de ello son las secuencias provenientes de *Daphnia*, las cuales no se relacionan con el resto de las secuencias de crustáceos. Estudios previos a partir del análisis del genoma de *Daphnia pulex*, indican que estos organismos presentan una alta tasa de mutaciones y divergencia en la mayoría de los genes (Colbourne et al. 2012), lo cual podría generar conflicto en las relaciones de los análisis filogenéticos en los que se incluyen (Sanderson & Shaffer 2002). De igual manera, la falta de secuencias de crustáceos basales reportados como posibles grupos hermanos del subfilo Hexapoda (Remipedia y otras especies de Braquiópodos) pueden generar conflicto en las relaciones obtenidas (Sanderson & Shaffer 2002).

De manera general en el presente estudio reconocen un origen monofilético para las celulasas de Protostomia (Mollusca, Anellida, Rotífera, Crustacea, Collembola e Insecta), además de reflejar las relaciones filogenéticas de los taxa Protostomia reportadas en estudios previos (Carroll & Knoll 1999, Jermin et al. 2005, Telford & Copley 2011, Benton et al. 2015). Lo anterior implica que por lo menos un gen que codificaba una proteína de la familia GH9 estuvo presente en el último ancestro común de los Metazoarios protostomados. De acuerdo al registro fósil, el organismo metazoario más antiguo del que se tiene registro (*Kimberella quadrata*) que representa un organismo Protostomado, data de hace 636.1 a 550.25 Ma (Benton et al. 2015), sugiriendo un origen evolutivo ancestral de las enzimas de la familia GH9. Sin embargo, no existen reportes de estos genes en clados basales de Metazoa como Porífera, Mesozoa y Placozoa (<http://www.cazy.org/>, Lombard et

al. 2014), por lo que la identificación de genes de la familia GH9 en estos taxa es indispensable para confirmar esta hipótesis.

Estudios previos que incluyen dos secuencias de eumetazoarios deuterostomados reportan un origen monofilético para las celulasas en animales (Lo et al. 2003, Davison & Blaxter, 2005). En el presente estudio no fue posible incluir las secuencias de estas dos especies de deuterostomados porque la divergencia de las secuencias generaba alineamientos inestables que podían afectar la reconstrucción de las relaciones filogenéticas de los eumetazoarios protostomados, y porque no existe evidencia que demuestre que estas secuencias pertenecen al genoma de los organismos analizados y no representan secuencias contaminantes durante el procesamiento de las muestras y secuenciación. Por lo tanto, con los resultados del presente estudio sólo se puede inferir un origen evolutivo ancestral para las celulasas de organismos eumetazoarios protostomados.

La función que realizaban las enzimas de la familia GH9 en el último ancestro común de los organismos bilaterios protostomados es un misterio, aunque se hipotetiza que este organismo ancestral, probablemente era parecido a un filtrador marino que, se alimentaba de la celulosa presente en algas (Lo et al. 2003). Estudios recientes sugieren que organismos filtradores actuales como Cnidarios usan las celulasas de la familia GH9 para degradar detritos provenientes de algas presentes en el medio acuático, así como para degradar la celulosa de zooxantelas con las cuales tienen una relación simbiótica (Yoshioka et al. 2017). El resto de los taxa de eumetazoarios protostomados usados en los análisis (Rotífera, Annelida, Mollusca, Crustacea y Hexápoda) utilizan las celulasas de la familia GH9 para degradar la celulosa proveniente de tejidos de plantas y algas, tanto vivas como en estado de descomposición (Bayer et al. 1998, Watanabe & Tokuda 2001, Suzuki et al. 2003, Zhao et al. 2007, Sarkar et al. 2009, Wei et al. 2009, Gilbert 2010, Calderón-Cortés et al. 2012), lo que sugiere que estas enzimas adquirieron una función digestiva en el último ancestro común de los

animales protostomados.

8.2 Origen y evolución de celulasas de insectos

Las secuencias de enzimas de la familia GH9 reportadas para insectos incluidas en este estudio están presentes en diversos órdenes, en particular de los grupos Polyneoptera, Paraneoptera y algunos órdenes de Holometabola. De acuerdo con el registro fósil, los insectos de estos grupos se originaron en el Paleozoico hace 360 a 286 Ma (Calderón-Cortés et al. 2012, Trautwein et al. 2012), período que corresponde al origen reportado para los insectos, lo cual sugiere que estos genes que permiten la degradación de la celulosa presente en los tejidos vegetales consumidos por los insectos, pudieron jugar un papel clave en la evolución del grupo. De esta manera, es posible que las relaciones filogenéticas de las enzimas de la familia GH9 reflejen procesos de divergencia y adaptación ancestrales que ocurrieron en este grupo de organismos.

No obstante, es importante mencionar que no existen secuencias de celulasas de la familia GH9 en los genomas de Lepidoptera y Diptera reportados (Shirley 2012, Calderon-Cortes et al. 2012, Kirsch et al. 2014). Esto sugiere que insectos con procesos de divergencia y diversificación más recientes (Misof et al. 2014) que pueden explotar otros recursos presentes en las plantas (néctar, polen, frutos, etc) perdieron la capacidad de degradación de celulosa (Calderón-Cortés et al. 2012).

Aunque los análisis filogenéticos reconocen la monofilia de las secuencias de algunos taxa de insectos (Blaberoidea, Hymenoptera, Coleoptera y Hemiptera), sin importar el grupo externo que se use, ni la inclusión o exclusión de grupos de secuencias de celulasas de la familia GH9, la monofilia de las celulasas de insectos no se recupera en ninguno de los análisis filogenéticos realizados. Las relaciones de los clados de insectos son consistentes en los 6 árboles obtenidos y muestran que las secuencias se agrupan en dos clados, el clado Polyneoptera o Grupo 1 uno presenta valores de soporte de 100/99,100/98 y 100/100, para el

primer, segundo y tercer análisis, mientras el clado Polyneoptera + Holometabola + Paraneoptera o Grupo 2 presenta valores de soporte de 100/100 en los tres análisis. La formación de estos clados no corresponde a las relaciones filogenéticas reportadas, ya que las secuencias de los taxa: Termitoidae (*Zootermopsis nevadensis* y *Cryptotermes secundus*), Embioptera (*Tinema cristinae*), y Phasmatodea (*Antipaluria urichi*) están presentes en ambos grupos, lo cual indica un posible evento de duplicación que ocurrió en un ancestro del superorden Polyneoptera, antes de la diversificación de los linajes basales de insectos que se ha propuesto para el carbonífero tardío hace 360 a 286 Ma (Calderón-Cortés et al. 2012). Los resultados también indican que existieron otros eventos de duplicación de genes al interior de órdenes específicos, como Embióptera y Phasmatodea, los cuales forman tres y seis clados respectivamente, coincidiendo con un estudio previo que demuestra que estos eventos de duplicación en Phasmatodea generaron neofuncionalización y subfuncionalización de enzimas por lo que estas tienen una mayor o menor actividad enzimática sobre distintos sustratos (Celulosa y Hemicelulosa) (Shelomi et al. 2016).

La duplicación de genes es un proceso que permite aumentar el material genético y obtener nuevos genes (Ohno 1970, Lynch & Conery 2000, Bielawski & Yang 2003, He & Zhang 2005, Soskine & Tawfik 2010); se genera por retroposición, duplicación y entrecruzamiento cromosómicos (Zhang 2003). En la mayoría de los casos la duplicación de genes permite liberar a una de las copias del gen de la presión de selección natural provocando que este acumule mutaciones (Bielawski & Yang 2003, Zhang 2003). Existen distintas variantes de lo que puede ocurrir a una de las copias del gen: esta se convierte en pseudogen, adquiere una subfuncionalización o neofuncionalización (Hughes 1994, Zhang 2003). En la neofuncionalización, las mutaciones obtenidas en una de las copias del gen alteran la estructura terciaria de este, al grado de adquirir una función diferente a la original; mientras que, en la subfuncionalización las dos copias del gen conservan una parte de la función del gen original aunque realicen funciones distintas (Lynch & Conery 2000, Zhang 2003, He & Zhang 2005, Soskine & Tawfik

2010). Por su parte la generación de pseudogenes ocurre cuando al ser redundante la copia acumula mutaciones al grado de perder funcionalidad (Lynch & Conery 2000, Zhang 2003, He & Zhang 2005, Soskine & Tawfik 2010).

En este sentido, la mayoría de las secuencias de genes de celulasas de la familia GH9 presentes en el grupo 1 de insectos han sido clonados y caracterizados mediante la expresión de proteínas (Tabla 1). Algunos estudios han demostrado que los genes del grupo 1 codifican enzimas funcionales que actúan sobre los enlaces internos de la celulosa generando oligosacáridos como celobiosa, celotriosa y celotetraosa, los cuales son usados por los insectos como fuente de energía (Vries et al. 2001, Shelomi et al. 2014, Shelomi et al. 2016). Esto sugiere que los genes de celulasas de insectos del grupo 1, representan enzimas funcionales que participan en la digestión de celulosa, al igual que sus ancestros protostomados.

Por el contrario, la mayoría de las secuencias de insectos presentes en el grupo 2 provienen de anotación genómica, por lo que hay pocos estudios que comprueben la actividad de estas enzimas (Shelomi et al. 2016). Además, los resultados de los análisis del dominio de la familia de enzimas GH9 revelan que la mayoría de las secuencias provenientes de insectos que carecen de alguno de los aminoácidos implicados en la actividad enzimática (Hymenoptera) se encuentran en este grupo de insectos, lo cual podría indicar que algunos genes de este grupo codifican proteínas sin actividad celulolítica o representan pseudogenes.

Estudios recientes han reportado que enzimas de la familia GH9 provenientes de *Tinema cristinae* y *Anoplophora glabripennis*, presentes en el grupo 2, no presentan actividad enzimática sobre celulosa ni xilano (McKenna et al. 2016, Shelomi et al. 2016). Como se mencionó previamente algunas secuencias de la familia GH9 de *T. cristinae* han sufrido procesos de diversificación de genes (Shelomi et al. 2016), por lo que la pérdida de actividad puede indicar un proceso de neofuncionalización sin fines digestivos.

Las enzimas β -1,3-glucanasas de la familia GH16 son ejemplos de enzimas que inicialmente tenían una función digestiva que presentaron un evento de neofuncionalización no digestivo. Estas enzimas están presentes en animales invertebrados, plantas, bacterias y hongos, pero dentro del grupo de los insectos se ha reportado un proceso de neofuncionalización en el que algunas copias de este gen ahora intervienen en la respuesta inmune de organismos de algunos taxa (Termitoidea, Lepidoptera, Diptera y Coleoptera) (Bragatto et al. 2010, Hughes 2012, Kirsch et al. 2012). No obstante, se desconoce cuál podría ser la función de las proteínas de la familia GH9 del grupo 2 de insectos.

La formación de pseudogenes y la posterior pérdida de genes, también es una consecuencia común de la duplicación génica y por tanto de la evolución de familias de genes (Zhang 2003, Demuth & Hahn 2009). La ausencia de la copia de celulasas funcionales (grupo 1) de la familia GH9 en Paraneoptera y Holometabola, sugiere que la copia de este gen se perdió antes del origen y diversificación de estos clados, hace 345 Ma de acuerdo con registro fósil (Misof et al. 2014). Una posible explicación a esta pérdida, son los cambios ocurridos durante la evolución de estos taxa como la diferencia drástica en la estructura y función del sistema digestivo (Terra 1990), la adaptación a diferentes dietas (Calderón-Cortés et al. 2012), así como la metamorfosis completa en el caso de Holometabola (Nicholson et al. 2009).

8.3 Análisis estructural

Como se mencionó previamente, después de un proceso de duplicación genética comúnmente se mantiene la estructura y función de las proteínas por un proceso de selección purificadora o negativa. Como consecuencia, pocas mutaciones que podrían dar ventajas funcionales a la enzima se fijan por selección natural positiva (Bielawski & Yang 2003, Fares and Travers 2006). El estudio funcional de una enzima, requiere por lo tanto de análisis que permitan identificar estas mutaciones

que alteran la estructura de una proteína.

De acuerdo con los análisis realizados en este estudio, las enzimas de la familia GH9 presentan una estructura que consiste en un barril formado por distintas hélices ($\alpha / \alpha 6$) (Fig. 16). De igual manera, los análisis revelaron una similitud estructural, lo que sugiere que los genes codificantes están sujetos a una fuerte presión de selección (Hughes 1994, Zhang 2003), debido a que estas enzimas llevan a cabo funciones importantes para los organismos al estar involucradas en distintos procesos, de los cuales destaca el proceso digestivo (Slaytor 1992, Watanabe et al. 1998, Watanabe & Tokuda 2001, Sami & Sakori 2008, Wei et al. 2009). Además, las hélices confieren estabilidad a las enzimas, dado que estas a diferencia de otras estructuras secundarias, son menos vulnerables a generar cambios por mutaciones (Abrusán & Marsh 2016). Es de esperarse entonces, que la selección purificadora y su estabilidad natural mantengan la conservación estructural observada. Sin embargo, la sobreposición de estructuras indica que existen algunas regiones variables en su mayoría compuestas por giros α y β -láminas que podrían corresponder a inserciones, deleciones y mutaciones acumuladas por selección positiva, las cuales son específicas de cada grupo de organismos.

Los análisis de diferencias estructurales por sobreposición y valores RMSD presentados en el dendograma, no reflejan las relaciones obtenidas en los análisis filogenéticos y tampoco presentan patrones evolutivos relacionados a condiciones en las que actúan las enzimas tales como: temperatura, hábitos alimenticios, tipo de hábitat, etc. Este resultado se puede deber a que las enzimas de la familia GH9 son altamente variables en algunas zonas, lo cual podría afectar los resultados de agrupamiento. Aun cuando la comparación estructural por medio de RMSD es usada comúnmente, algunos autores han reportado que este método es muy sensible a variaciones locales en las estructuras (Xu & Zhang 2010). Considerando lo anterior es necesario que estudios futuros incluyan análisis alternativos que hagan uso de otros métodos de comparación estructural de

enzimas menos sensibles a la variación local (Xu & Zhang 2010, Wang et al. 2013). Sin embargo, no se puede descartar que la diversificación asociada a diferentes eventos de duplicación que observamos en los análisis filogenéticos (asociada a diferentes eventos de duplicación) no se refleje en comparaciones estructurales globales. Por tal motivo, se decidió analizar características estructurales más puntuales, las cuales podrían estar implicadas en la evolución de estas enzimas.

De acuerdo a la literatura, la unión de las enzimas de familia GH9 con iones metálicos es importante para la estabilidad de algunas de estas enzimas (Mandels et al. 1957, Khademi et al. 2002, Shahriarinnour 2011, Chen et al. 2018). La estabilidad es una propiedad fisicoquímica de las enzimas que determina si una enzima puede llevar a cabo una reacción enzimática en diferentes condiciones, ya que la incorporación de iones metálicos puede jugar un papel fundamental en la estructura terciaria de la enzima y/o interactuar con el sustrato (Mandels et al. 1957, Juy et al. 1992, Shahriarinnour 2011, Chen et al. 2018). Bajo este supuesto, es posible esperar que las enzimas de diferentes taxa de organismos adaptados a condiciones y hábitos alimenticios distintos, presenten sitios de unión conservados dentro del grupo que pueden diferir a los sitios de unión de otros grupos de enzimas. No obstante, los resultados del análisis de predicción de sitios de unión a iones metálicos realizado en el presente estudio no muestra ningún patrón, la presencia de los sitios de unión a iones es variable dentro de los clados formados en los análisis filogenéticos (Fig. 16). Por lo tanto, es probable que los iones metálicos no cumplan un papel fundamental en la actividad enzimática de las celulasas de la familia GH9. Sin embargo, es necesario mencionar que los resultados obtenidos se podrían deber a que el modelado por homología se basa en las estructuras reportadas (Evers & Klege 2003) y en el caso de las celulasas de la familia GH9 en eumetazoarios solo se han cristalizado 3 estructuras (Khademi et al. 2002, Arimori et al. 2013, 4ZH5/4ZG8).

En el análisis de evolutionary trace se obtuvieron dos clados (en la segunda

partición), los cuales corresponden a los grupos obtenidos en los análisis filogenéticos. Además, el análisis reveló que la diferencia de las secuencias consenso de estos grupos corresponde a una mutación en la posición 393 de la estructura cristalizada de *Nasutitermes takasagoensis* (cerca del sitio de unión al sustrato), en la cual se reemplaza una glicina por una serina en el grupo dos. Esta mutación podría representar un cambio significativo en la función de estas enzimas, dado que si bien los aminoácidos involucrados en esta mutación (serina y glicina) pertenecen al grupo de aminoácidos pequeños, el cambio de la glicina por otro aminoácido puede alterar la función de las enzimas (Betts & Russell 2016). También se ha reportado que el grupo hidroxilo de la serina puede formar puentes de hidrogeno con una gran cantidad de sustratos polares en algunas hidrolasas (proteasas y lipasas) (Betts & Russell 2016). Por lo tanto no se puede descartar que el cambio de una glicina por una serina esté relacionado con diferencias en las propiedades de unión y degradación del sustrato en ambos grupos de insectos.

El análisis de hidrofobicidad indica que existe una diferencia estructural en las celulasas de la familia GH9 del grupo 2 de insectos, dado que los modelos pertenecientes a este grupo presentan diferencias en el sitio de unión al sustrato, tales como una cavidad más larga y una hendidura con propiedades hidrofóbicas en la parte superior de esta (Fig. 21). Las cavidades están involucradas en una gran variedad de procesos como el reconocimiento a ligandos, unión, transporte y/o transferencia de moléculas pequeñas y el tránsito de sustratos y productos (Barletta and Fernandez-Alberti 2018, Gora et al. 2013). Estas crean microambientes específicos para generar interacciones que participan en la catálisis enzimática (Gora et al. 2013), los cuales se pueden ver afectados por la rotación de un residuo, los cambios conformacionales de interdominios (Barletta & Fernandez-Alberti 2018) y los cambios en las propiedades hidrofóbicas/hidrofílicas (Gora et al. 2013). Por lo anterior, el cambio en las propiedades hidrofóbicas en la parte superior de la cavidad de las enzimas del grupo 2 podría estar relacionado con un cambio en la función (Nakai 1983).

Por su parte, el análisis de potencial electrostático de la superficie indicó que la zona más negativa de las enzimas de la familia GH9 analizadas se encuentra presente en la cavidad, al igual que como se ha reportado en estudios previos de estas enzimas (Arimori et al. 2013). En el caso de la celulasa de la familia GH9 en *Eisenia fetida* reportan cargas altamente negativas como adaptación de la enzima a bajas temperaturas (Arimori et al. 2013). Cambios electrostáticos en la superficie de las enzimas, se han relacionado con adaptaciones particulares a diferentes condiciones de temperatura y pH (Gertler 1971, Montagu et al. 1993, Secundo 2013), lo cual explicaría las variaciones electrostáticas particulares encontradas para cada enzima, sin importar si provienen de un mismo organismo o grupo.

En general, los resultados indican que la cavidad presenta diferencias estructurales y bioquímicas (hidrofobicidad) en los aminoácidos que podrían estar relacionadas con la divergencia de las secuencias del grupo 1 y 2 de insectos, así como con posibles cambios en la función de las enzimas del grupo 2. Estos resultados resaltan la importancia de combinar los análisis filogenéticos con análisis estructurales puntuales que permitan predecir cambios entre los grupos recuperados en las filogenias e identificar líneas de investigación futuras.

9. Conclusiones

- Los análisis filogenéticos confirmaron el origen monofilético de las enzimas de la familia GH9 en Protostomia y revelaron la existencia de dos clados de estas enzimas en insectos.
- La formación de los dos clados de celulasas de insectos indica un evento de duplicación de genes antes de la divergencia de los grupos Paraneóptera y Holometabola.
- Los análisis estructurales revelaron una mutación en un aminoácido cercano a los aminoácidos catalíticos entre el grupo 1 y 2 de las secuencias de insectos y diferencias de hidrofobicidad en la cavidad de las enzimas de estos grupos.
- Es probable que las diferencias estructurales puntuales estén relacionadas con un evento de neofuncionalización de las proteínas de la familia de GH9 en el grupo 2 de insectos, que debe confirmarse en estudios posteriores.

Referencias

- Abrahama M. J., Murtolad T., Schulzb R., P´alla S., Smithb J. C., Hessa B., Lindahla E., D., & A. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Aik, W. S., McDonough, M. A., Thalhammer, A., Chowdhury, R., & Schofield, C. J. (2012). Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Current Opinion in Structural Biology*, 22(6), 691–700. <https://doi.org/10.1016/j.sbi.2012.10.001>
- Alzari P. M., H el ene S. & Dominguez R. (1996). New species of spiriferids (Brachiopoda) from the Devonian and Carboniferous of Eastern Europe. *Paleontological Journal*, 40(5), 507–517. <https://doi.org/10.1134/S0031030106050054>
- Arimori, T., Ito, A., Nakazawa, M., Ueda, M., & Tamada, T. (2013). Crystal structure of endo-1,4-??-glucanase from *Eisenia fetida*. *Journal of Synchrotron Radiation*, 20(6), 884–889. <https://doi.org/10.1107/S0909049513021110>
- Abrus an, G., & Marsh, J. A. (2016). Alpha Helices Are More Robust to Mutations than Beta Strands. *PLoS Computational Biology*, 12(12), 1–16. <https://doi.org/10.1371/journal.pcbi.1005242>
- Akazawa, S. I., Ikarashi, Y., Yarimizu, J., Yokoyama, K., Kobayashi, T., Nakazawa, H., ... Morikawa, Y. (2016). Characterization of two endoglucanases for the classification of the earthworm, *Eisenia fetida* Waki. *Bioscience, Biotechnology and Biochemistry*, 80(1), 55–66. <https://doi.org/10.1080/09168451.2015.1075860>
- Barletta, G. P., & Fernandez-Alberti, S. (2018). Protein Fluctuations and Cavity Changes Relationship. *Journal of Chemical Theory and Computation*, 14(2), 998–1008. <https://doi.org/10.1021/acs.jctc.7b00744>
- Bayer, E. A., Chanzy, H., Lamed, R., & Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology*, 8(5), 548–557. [https://doi.org/10.1016/S0959-440X\(98\)80143-7](https://doi.org/10.1016/S0959-440X(98)80143-7)
- Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2002). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18), 10037–10041. <https://doi.org/10.1073/pnas.181342398>
- Bucher. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 452(7190), 949–955. <https://doi.org/10.1038/nature06784>
- Baumgarten, S., Simakov, O., Esherick, L. Y., Liew, Y. J., Lehnert, E. M., Michell, C. T., ... Voolstra, C. R. (2015). The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proceedings of the National Academy of Sciences*, 112(38), 11893–11898. <https://doi.org/10.1073/pnas.1513318112>
- Beguın, P. and Aubert, J. (1994) The biological degradation of cellulose. *FEMS Microbiol. Rev*, 13, 25-28.
- Benton, M. J., Donoghue, P. C. J., Asher, R. J., Friedman, M., Near, T. J., & Vinther, J. (2015). Constraints on the timescale of animal evolutionary history, 1–107.
- Benkert, P., K unzli, M., & Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Research*, 37(SUPPL. 2), 510–514. <https://doi.org/10.1093/nar/gkp322>

Benziman, M., Haigler, C. H., Brown, R. M., White, a R., & Cooper, K. M. (1980). Cellulose biogenesis: Polymerization and crystallization are coupled processes in *Acetobacter xylinum*. *Proceedings of the National Academy of Sciences of the United States of America*, 77(11), 6678–6682. <https://doi.org/10.1073/pnas.77.11.6678>

Betts, M J. & Russell., R B. (2016). Amino Acid Properties and Consequences of Substitutions. *Physical Review A*, 93(5). <https://doi.org/10.1103/PhysRevA.93.053607>

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS - A MESSAGE-PASSING PARALLEL MOLECULAR-DYNAMICS IMPLEMENTATION. *Computer Physics Communications*, 91(1-3), 43-56. [https://doi.org/10.1016/0010-4655\(95\)00042-e](https://doi.org/10.1016/0010-4655(95)00042-e)

Berendsen, H. J. C., Hess, B., Lindahl, E., Van Der Spoel, D., Mark, A. E., & Groenhof, G. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701–1718. <https://doi.org/10.1002/jcc.20291>

Bielawski, J. P., & Yang, Z. (2003). Maximum likelihood methods for detecting adaptive evolution after gene duplication. *Journal of Structural and Functional Genomics*, 3: 201–212, 201–212.

Bonasio, R., Zhang, G., Ye C., Mutti, N. S., Fang X., Qin N., Donahue, G., Zhang P., Yang P., Li Q., Li C., Zhang P., Huang Z., Berger S. L., Reinberg D., Wang J. & Liebig J. (2013). Genomic comparison of the ants, 329(5995), 1–11. Retrieved from [papers3://publication/uuid/1B6CF02F-C93D-4188-AD62-ABC797148D81](https://pubmed.ncbi.nlm.nih.gov/24111111/)

Boudet, a M., Lapierre, C., & Grimapettenati, J. (1995). Tansley Review No-80 - Biochemistry and Molecular-Biology of Lignification. *New Phytologist*, 129(2), 203–236. <https://doi.org/10.1111/j.1469-8137.1995.tb04292.x>

Brás JL., Cartmell A., Carvalho AL., Verzé G., Bayer E. A., Vazana Y., Correia M. A., Prates J. A., Ratnaparkhe S., Boraston A. B., Romão M. J., Fontes C. M & Gilbert H. J. (2011). Structural insights into a unique cellulase fold and mechanism of cellulose hydrolysis. *Proceedings of the National Academy of Sciences*, 108(13), 5237–5242. <https://doi.org/10.1073/pnas.1015006108>

Bragatto, I., Genta, F. A., Ribeiro, A. F., Terra, W. R., & Ferreira, C. (2010). Characterization of a β -1,3-glucanase active in the alkaline midgut of *Spodoptera frugiperda* larvae and its relation to β -glucan-binding proteins. *Insect Biochemistry and Molecular Biology*, 40(12), 861–872. <https://doi.org/10.1016/j.ibmb.2010.08.006>

Brunecky R., Alahuhta M., Xu Q., Donohoe B. S., Crowley M. F., Kataeva I. A., Yang S. J., Resch M. G., Adams M. W., Lunin V. V., Himmel M. E. & Bomble Y. J. (2013). Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science*, 342, 1513–1516. <https://doi.org/10.2210/PDB4DOD/PDB>

Bui, T. H. H., & Lee, S. Y. (2014). Endogenous cellulase production in the leaf litter foraging mangrove crab *Parasesarma erythroactyla*. *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology*, 179, 27–36. <https://doi.org/10.1016/j.cbpb.2014.09.004>

Byrne, K. A., Lehnert, S. A., Johnson, S. E., & Moore, S. S. (1999). Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. *Gene*, 239(2), 317–324. [https://doi.org/10.1016/S0378-1119\(99\)00396-0](https://doi.org/10.1016/S0378-1119(99)00396-0)

Caffall, K. H., & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research*, 344(14), 1879–1900. <https://doi.org/10.1016/j.carres.2009.05.021>

Calderón-Cortés N., Oyama K. & Watanabe H.. (2014). Evolution of cellulose digestion in insects:

insights from midgut and salivary glands transcriptome analyses of the apterygotan insect *Thermobia domestica*, 20855.

Calderón-Cortés, N., Watanabe, H., Cano-Camacho, H., Zavala-Páramo, G., & Quesada, M. (2010). CDNA cloning, homology modelling and evolutionary insights into novel endogenous cellulases of the borer beetle *Oncideres albomarginata* chamela (Cerambycidae). *Insect Molecular Biology*, 19(3), 323–336. <https://doi.org/10.1111/j.1365-2583.2010.00991.x>

Calderón-Cortés, N., Quesada, M., Watanabe, H., Cano-Camacho, H., & Oyama, K. (2012). Endogenous Plant Cell Wall Digestion: A Key Mechanism in Insect Evolution. *Annual Review of Ecology, Evolution, and Systematics*, 43(1), 45–71. <https://doi.org/10.1146/annurev-ecolsys-110411-160312>

Carriço, J. A., Sabat, A. J., Friedrich, A. W., & Ramirez, M. (2013). Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution, on behalf of the ESCMID Study Group for Epidemiological Markers (ESGEM). *Eurosurveillance*, 18(4), 1–9. <https://doi.org/10.1038/jid.2013.1.Research>

Carroll, S. B., & Knoll, A. H. (1999). Early Animal Evolution: Emerging Views from Comparative Biology and Geology. *Science*, 284(June), 2129–2137.

Castell, J. D., Kean, J. C., D'Abramo, L. R., & Conklin, D. E. (1989). A Standard Reference Diet for Crustacean Nutrition Research. I. Evaluation of Two Formulations. *Journal of the World Aquaculture Society*, 20(3), 93–99. <https://doi.org/10.1111/j.1749-7345.1989.tb00560.x>

Chang, C. J., Wu, C. P., Lu, S. C., Chao, A. L., Ho, T. H. D., Yu, S. M., & Chao, Y. C. (2012). A novel exo-cellulase from white spotted longhorn beetle (*Anoplophora malasiaca*). *Insect Biochemistry and Molecular Biology*, 42(9), 629–636. <https://doi.org/10.1016/j.ibmb.2012.05.002>

Chauvaux S., Souchon H., Alzari P. M., Chariot P. & Beguin P. (1995). Structural and Functional Analysis of the Metal-binding Sites of *Clostridium thermocellum* Endoglucanase CelD.

Chen, Y., Shen, Y., Wang, W., & Wei, D. (2018). Mn²⁺modulates the expression of cellulase genes in *Trichoderma reesei* Rut-C30 via calcium signaling. *Biotechnology for Biofuels*, 11(1), 1–19. <https://doi.org/10.1186/s13068-018-1055-6>

Cleveland L. R. (1924) The physiological and symbiotic relationships between the intestinal protozoa of termites and their hosts, with special reference to *Reticulitermes flavipes* KoUar. *Biol. Bull.* 6, 117-227.

Colbourne J. K., Pfrender M. E., Gilbert D, Thomas W. K., Tucker A, Oakley T. H., Tokishita S, Aerts A, Arnold G. J., Basu M. K., Bauer D. J., Cáceres C. E., Carmel L, Casola C, Choi J. H., Detter J. C., Dong Q, Dusheyko S, Eads B. D., Fröhlich T,, Geiler-Samerotte K. A., Gerlach D, Hatcher P, Jogdeo S, Krijgsveld J, Kriventseva E. V., Kültz D, Laforsch C, Lindquist E, Lopez J, Manak J. R., Muller J, Pangilinan J, Patwardhan R. P., Pitluck S, Pritham E. J., Rechtsteiner A, Rho M, Rogozin I. B., Sakarya O, Salamov A, Schaack S, Shapiro H, Shiga Y, Skalitzky C, Smith Z, Souvorov A, Sung W, Tang Z, Tsuchiya D, Tu H, Vos H, Wang M, Wolf Y. I., Yamagata H, Yamada T, Ye Y, Shaw J. R., Andrews J, Crease T. J., Tang H, Lucas S. M., Robertson H. M., Bork P, Koonin E. V., Zdobnov E. M., Grigoriev I. V., Lynch M, Boore J. L. (2012). *Daphnia pulex* genome, 331(6017), 555–561. <https://doi.org/10.1126/science.1197761>.

Correa A, Pacheco S, Mechaly A. E., Obal G, Béhar G., Mouratou B., Oppezzo P., Alzari P. M, Pecorari F. (2014). Potent and Specific Inhibition of Glycosidases by Small Artificial Binding Proteins (Affitins). *Plos One*, 9, 97438. <https://doi.org/10.2210/PDB4CJ0/PDB>

Cosgrove, D. J. (1997). Assembly and enlargement of the primary cell wall in plants. *Annual Review*

of Cell and Developmental Biology, 13, 171–201. <https://doi.org/10.1146/annurev.cellbio.13.1.171>

Cosgrove, D. J. (2005). Growth of the plant cell wall. Nature Reviews. Molecular Cell Biology, 6(November), 850–861. <https://doi.org/10.1038/nrm1746>

Crawford, A. C., Krickler, J. A., Anderson, A. J., Richardson, N. R., & Mather, P. B. (2004). Structure and function of a cellulase gene in redclaw crayfish, *Cherax quadricarinatus*. Gene, 340(2), 267–274. <https://doi.org/10.1016/j.gene.2004.06.060>

Crippen, V. N. M. & G. M. (1994). Significance of Root-Mean-Square Deviation in Comparing Three-dimensional Structures of Globular Proteins. J. Mol Biol.

Danchin, E. G. J., Rosso, M.-N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., & Abad, P. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proceedings of the National Academy of Sciences of the United States of America, 107(41), 17651–17656. <https://doi.org/10.1073/pnas.1008486107>

Davison, A., & Blaxter, M. (2005). Ancient origin of glycosyl hydrolase family 9 cellulase genes. Molecular Biology and Evolution, 22(5), 1273–1284. <https://doi.org/10.1093/molbev/msi107>

de Vries, R. P., Visser, J., Ronald, P., de Vries, R., & P. (2001). Aspergillus Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides. Microbiology and Molecular Biology Reviews, 65(4), 497–522. <https://doi.org/10.1128/MMBR.65.4.497>

Demuth, J. P., & Hahn, M. W. (2009). The life and death of gene families. BioEssays, 31(1), 29–39. <https://doi.org/10.1002/bies.080085>

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput Nucleic Acids Res. 32(5):1792-1797

Ehrlich PR, Raven PH. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586–608

Fratti F, Galletti R, De Lorenzo G, Salerno G, Conti E. 2006. Activity of endo-polygalacturonases in mirid bugs (Heteroptera: Miridae) and their inhibition by plant cell wall proteins (PGIPs). Eur. J. Entomol. 103:515–22. <https://doi.org/10.14411/eje.2006.067>

Ellinghaus T. L., Pereira J. H., McAndrew R. P., Welner D. H., DeGiovanni A. M., Guenther J. M., Tran H. M., Feldman T. 1., Simmons B. A., Sale K. L. & Adams P. D. (2018). Engineering glycoside hydrolase stability by the introduction of zinc binding. Acta Crystallogr D Struct Biol, 74, 702–710. <https://doi.org/10.2210/PDB5U0H/PDB>

Evers, A., & Klebe, G. (2003). Ligand-Supported Homology Modeling of G-Protein-Coupled Receptor Sites: Models Sufficient for Successful Virtual Screening. Angewandte Chemie - International Edition, 43(2), 248–251. <https://doi.org/10.1002/anie.200352776>

Eyun S. I., Wang H., Pauchet Y., Ffrench-Constant R. H., Benson A. K., Valencia-Jiménez A., Moriyama E. N., Siegfried B. D., (2014). Molecular evolution of glycoside hydrolase genes in the western corn rootworm (*Diabrotica virgifera virgifera*). PLoS ONE, 9(4). <https://doi.org/10.1371/journal.pone.0094052>

Faddeeva-Vakhrusheva A., Derks M. F., Anvar S. Y., Agamennone V., Suring W., Smit S., van Straalen N. M. & Roelofs D. (2016). Gene family evolution reflects adaptation to soil environmental stressors in the genome of the collembolan *orchesella cincta*. Genome Biology and Evolution, 8(7), 2106–2117. <https://doi.org/10.1093/gbe/evw134>

- Fares, M. A., & Travers, S. A. A. (2006). A novel method for detecting intramolecular coevolution: Adding a further dimension to selective constraints analyses. *Genetics*, 173(1), 9–23. <https://doi.org/10.1534/genetics.105.053249>
- Farrell BD (1998) “Inordinate Fondness” explained: why are there So many beetles? *Science* 281: 555–559.
- Cairo J. P., Oliveira L. C., Uchima C. A., Alvarez T. M., Citadini A. P., Cota J, Leonardo F. C., Costa-Leonardo A. M., Carazzolle M. F., Costa F. F., Pereira G. A. & Squina F. M. (2013). Deciphering the synergism of endogenous glycoside hydrolase families 1 and 9 from *Coptotermes gestroi*. *Insect Biochemistry and Molecular Biology*, 43(10), 970–981. <https://doi.org/10.1016/j.ibmb.2013.07.007>
- Gaskell, A., Crennell, S., & Taylor, G. (1995). The three domains of a bacterial sialidase: a β -propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure*, 3(11), 1197–1205. [https://doi.org/10.1016/S0969-2126\(01\)00255-6](https://doi.org/10.1016/S0969-2126(01)00255-6)
- Gertler, A. (1971). Selective, Reversible Loss of Elastolytic Activity of Elastase and Subtilisin Resulting from Electrostatic Changes Due to Maleylation. *European Journal of Biochemistry*, 23(1), 36–40. <https://doi.org/10.1111/j.1432-1033.1971.tb01589.x>
- Gilbert, H. J. (2010). The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiology*, 153(2), 444–455. <https://doi.org/10.1104/pp.110.156646>
- Gora, A., Brezovsky, J., & Damborsky, J. (2013). Gates of Enzymes. *Chemical Reviews*, 113(8), 5871–5923. <https://doi.org/10.1021/cr300384w>
- Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., & Caves, L. S. D. (2006). Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*, 22(21), 2695–2696. <https://doi.org/10.1093/bioinformatics/btl461>
- Gray, M., Linton, S. M., & Allardyce, B. J. (2018). cDNA sequences of GHF9 endo- β -1,4-glucanases in terrestrial Crustacea. *Gene*, 642(November 2017), 408–422. <https://doi.org/10.1016/j.gene.2017.11.030>
- Gromiha, M. M., Pujadas, G., Magyar, C., Selvaraj, S., & Simon, I. (2004). Locating the Stabilizing Residues in (α/β)₈ Barrel Proteins Based on Hydrophobicity, Long-Range Interactions, and Sequence Conservation. *Proteins: Structure, Function and Genetics*, 55(2), 316–329. <https://doi.org/10.1002/prot.20052>
- Guo, D., Chen, F., Inoue, K., Blount, J. W., & Dixon, R. a. (2001). Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa. impacts on lignin structure and implications for the biosynthesis of G and S lignin. *The Plant Cell*, 13(1), 73–88. <https://doi.org/10.1105/tpc.13.1.73>
- He, X., & Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics*, 169(2), 1157–1164. <https://doi.org/10.1534/genetics.104.037051>
- Hehruch P. R. & Raven P. (1964). This Week’ s Citation Classic. *Psychometric Theory*, 18, 586–608.
- Henrissat, B., & Bairoch, a. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical Journal*, 293 (Pt 3(January), 781–788. <https://doi.org/066/0>

- Heredia, A., Jiménez, A., Guillén, R., & Guillén, R. (1995). Composition of plant cell walls. *Zeitschrift Für Lebensmittel-Untersuchung Und -Forschung*, 200(1), 24–31. <https://doi.org/10.1007/BF01192903>
- Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3), 435–447. <https://doi.org/10.1021/ct700301q>
- Honda, Y., Arai, S., Suzuki, K., Kitaoka, M., & Fushinobu, S. (2016). The crystal structure of an inverting glycoside hydrolase family 9 exo-beta-D-glucosaminidase and the design of glycosynthase. *Biochem.J.*, 473, 463–472. <https://doi.org/10.2210/PDB5DQG/PDB>
- Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17:754-755.
- Huerta-Cepas, J., Dopazo, J., & Gabaldón, T. (2010). ETE: A python Environment for Tree Exploration. *BMC Bioinformatics*, 11. <https://doi.org/10.1186/1471-2105-11-24>
- Hughes, A. L. (2012). Evolution of the β gRP/GNBP/ β -1,3-glucanase family of insects. *Immunogenetics*, 64(7), 549–558. <https://doi.org/10.1007/s00251-012-0610-8>
- Hughes, A. L., & Hughes, A. L. (1994). The Evolution of Functionally Novel Proteins after Gene Duplication Stable URL : <http://www.jstor.org/stable/50343> Linked references are available on JSTOR for this article : The evolution of functionally novel proteins after gene duplication, 256(1346), 119–124.
- Innis, C. A., Shi, J., & Blundell, T. L. (2000). Evolutionary trace analysis of TGF- and related growth factors: implications for site-directed mutagenesis. *Protein Engineering Design and Selection*, 13(12), 839–847. <https://doi.org/10.1093/protein/13.12.839>
- Jermiin, L. S., Poladian, L., & Charleston, M. A. (2005). Is the “big bang” in animal evolution real? *Science*, 310(5756), 1910–1911. <https://doi.org/10.1126/science.1122440>
- Jondelius, U., Ruiz-Trillo, I., Bagnà, J., & Riutort, M. (2002). The Nemertodermatida are basal bilaterians and not members of the Platyhelminthes. *Zoologica Scripta*, 31(2), 201–215. <https://doi.org/10.1046/j.1463-6409.2002.00090.x>
- Juy, M., Amrt, A. G., Alzari, P. M., Poljak, R. J., Claeysens, M., Béguin, P., & Aubert, J. P. (1992). Three-dimensional structure of a thermostable bacterial cellulase. *Nature*, 357(6373), 89–91. <https://doi.org/10.1038/357089a0>
- King A. J., Cragg S. M., Li Y, Dymond J, Guille M. J., Bowles D. J., Bruce N. C., Graham I. A. & McQueen-Mason S. J. (2010). Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proceedings of the National Academy of Sciences*, 107(12), 5345–5350. <https://doi.org/10.1073/pnas.0914228107>
- Kesavulu, M. M., Tsai, J.-Y., Lee, H.-L., Liang, P.-H., & Hsiao, C.-D. (2012). Structure of the Catalytic Domain of the *Clostridium Thermocellum* Cellulase C_{elt}. *Acta Crystallogr., Sect. D*, 68, 310. <https://doi.org/10.2210/PDB2YIK/PDB>
- Khademi, S., Guarino, L. A., Watanabe, H., Tokuda, G., & Meyer, E. F. (2002). Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. *Acta Crystallographica Section D: Biological Crystallography*, 58(4), 653–659. <https://doi.org/10.1107/S09074444902002366>
- King, A.J., Cragg, S.M., Li, Y., et al., 2010. Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *PNAS* 107 (12), 5345–5350.

<https://doi.org/10.1073/pnas.0914228107King>

Kim N., Choo Y. M., Lee K. S., Hong S. J., Seol K. Y., Je Y. H., Sohn H. D. & Jin B. R.. (2008). Molecular cloning and characterization of a glycosyl hydrolase family 9 cellulase distributed throughout the digestive tract of the cricket *Teleogryllus emma*. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 150(4), 368–376. <https://doi.org/10.1016/j.cbpb.2008.04.005>

Kirkness E. F., Haas B. J., Sun W, Braig H. R., Perotti M. A., Clark J. M., Lee S. H., Robertson H. M., Kennedy H. M., Elhaik E., Gerlach D., Kriventseva E. V., Elsik C. G., Graur D., Hill C. A., & PNAS July 6, 2010 107 (27) 1216. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences*, 107(27), 12168–12173. <https://doi.org/10.1073/pnas.1003379107>

Kirsch, R., Gramzow, L., Theißen, G., Siegfried, B. D., Ffrench-Constant, R. H., Heckel, D. G., & Pauchet, Y. (2014). Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles. *Insect Biochemistry and Molecular Biology*, 52(1), 33–50. <https://doi.org/10.1016/j.ibmb.2014.06.008>

Kirsch, R., Wielsch, N., Vogel, H., Svatoš, A., Heckel, D. G., & Pauchet, Y. (2012). Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. *BMC Genomics*, 13, 587. <https://doi.org/10.1186/1471-2164-13-587>

Kjer K. M., Carle F. L., Litman J. & Ware J. (2017). Simultaneous engineering of an enzyme's entrance tunnel and active site: The case of monoamine oxidase MAO-N. *Chemical Science*, 8(5), 4093–4099. <https://doi.org/10.1039/c6sc05381e>

Lee S. J., Kim S. R., Yoon H. J., Kim I, Lee K. S., Je Y. H., Lee S. M., Seo S. J., Dae S. o. & Jin B. R. (2004). cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 139(1), 107–116. <https://doi.org/10.1016/j.cbpc.2004.06.015>

Lee S. J., Lee K. S., Kim S. R., Gui Z. Z., Kim Y. S., Yoon H. J., Kim I, Kang P. D., Sohn H. D. & Jin B. R. (2005). A novel cellulase gene from the mulberry longicorn beetle, *Apriona germari*: Gene structure, expression, and enzymatic activity. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 140(4), 551–560. <https://doi.org/10.1016/j.cbpc.2004.12.003>

Legg, D. A., Sutton, M. D., & Edgecombe, G. D. (2013). Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nature Communications*, 4, 1–7. <https://doi.org/10.1038/ncomms3485>

Lehnert, E. M., Burriesci, M. S., & Pringle, J. R. (2012). Developing the anemone *Aiptasia* as a tractable model for cnidarian-dinoflagellate symbiosis: the transcriptome of aposymbiotic *A. pallida*. *BMC Genomics*, 13(1), 1. <https://doi.org/10.1186/1471-2164-13-271>

Levitt, M., & Lifson, S. (1969). Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of Molecular Biology*, 46(2), 269–279. [https://doi.org/10.1016/0022-2836\(69\)90421-5](https://doi.org/10.1016/0022-2836(69)90421-5)

Lerouxel, O., Cavalier, D. M., Liepman, A. H., & Keegstra, K. (2006). Biosynthesis of plant cell wall polysaccharides - a complex process. *Current Opinion in Plant Biology*, 9(6), 621–630. <https://doi.org/10.1016/j.pbi.2006.09.009>

Li, L., Fröhlich, J., Pfeiffer, P., & König, H. (2003). Termite Gut Symbiotic Archaezoa Are Becoming Living Metabolic Fossils. *Eukaryotic Cell*, 2(5), 1091–1098. <https://doi.org/10.1128/ec.2.5.1091->

[1098.2003](#)

Li, Y., Yin, Q., Ding, M., & Zhao, F. (2009). Purification, characterization and molecular cloning of a novel endo- β -1,4-glucanase AC-EG65 from the mollusc *Ampullaria crosseana*. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 153(2), 149–156. <https://doi.org/10.1016/j.cbpb.2009.02.011>

Lynch, M. & Conery J. S. (2002). The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, 290(5494), 1151–1155. <https://doi.org/10.1126/science.290.5494.1151>

Lo N, Tokuda G, Watanabe H, Rose H, Slaytor M, Maekawa K, Bandi C & Noda H. (2000). Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Current Biology*, 10(13), 801–804. [https://doi.org/10.1016/S0960-9822\(00\)00561-3](https://doi.org/10.1016/S0960-9822(00)00561-3)

Lo, N., Watanabe, H., & Sugimura, M. (2003). Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. *Proceedings of the Royal Society B: Biological Sciences*, 270(Suppl_1), S69–S72. <https://doi.org/10.1098/rsbl.2003.0016>

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42(D1), 490–495. <https://doi.org/10.1093/nar/gkt1178>

Lovell, S. C., Davis, I. W., Arendall, W. B., De Bakker, P. I. W., Word, J. M., Prisant, M. G. & Richardson, D. C. (2003). Structure validation by $C\alpha$ geometry: ϕ, ψ and $C\beta$ deviation. *Proteins: Structure, Function and Genetics*, 50(3), 437–450. <https://doi.org/10.1002/prot.10286>

Mandelman, D., Belaich, A., Belaich, J. P., Aghajari, N., Driguez, H., & Haser, R. (2003). X-Ray Crystal Structure of the Multidomain Endoglucanase Cel9G from *Clostridium cellulolyticum* Complexed with Natural and Synthetic Cello-Oligosaccharides. *J.BACTERIOL.*, 185, 4127–4135. <https://doi.org/10.2210/PDB1GA2/PDB>

Mandels, M., & Reese, E. T. (1957). Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals. *Journal of Bacteriology*, 73(2), 269–278. <https://doi.org/10.1002/path.1700730133>

Martin, M. M., Jones, C. G., & Bernays, E. A. (1991). The evolution of cellulose digestion in insects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 333(1267), 281–288. <https://doi.org/10.1098/rstb.1991.0078>

Maierov V. N. & Crippen G. M. (1994). Significance of Root-Mean-Square Deviation in Comparing Three-dimensional Structures of Globular Proteins. *J. Mol Biol.*

McGavin George C. (2005). *MANUAL DE IDENTIFICACION Insectos arañas y otros artrópodos terrestres. Manuales de identificación.*: Omega.

McKenna D. D., Scully E. D., Pauchet Y., Hoover K, Kirsch R., Geib S. M., Mitchell R. F., Waterhouse R. M., Ahn S. J., Arsala D., Benoit J. B., Blackmon H., Bledsoe T., Bowsher J. H., Busch A., Calla B., Chao H., Childers A. K., Childers C., Clarke D. J., Cohen L., Demuth J. P., Dinh H., Doddapaneni H., Dolan A., Duan J. J., Dugan S., Friedrich M., Glastad K. M., Goodisman M. A., Haddad S., Han Y., Hughes D. S., Ioannidis P., Johnston J. S., Jones J. W., Kuhn L. A., Lance D. R., Lee C. Y., Lee S. L., Lin H., Lynch J. A., Moczek A. P., Murali S. C., Muzny D. M., Nelson D. R., Palli S. R., Panfilio K. A., Pers D., Poelchau M. F., Quan H., Qu J., Ray A. M., Rinehart J. P., Robertson H. M., Roehrdanz R., Rosendale A. J., Shin S., Silva C., Torson A. S., Jentzsch I. M., Werren J. H., Worley K. C., Yocum G., Zdobnov E. M., Gibbs R. A. & Richards S. (2016). Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biology*,

17(1), 1–18. <https://doi.org/10.1186/s13059-016-1088-8>

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8.

Misof B., Liu S., Meusemann K., Peters RS., Donath A., Mayer C., Frandsen PB., Ware J., Flouri T., Beutel RG., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust J., Aberer AJ., Aspöck U., Aspöck H., Bartel D., Blanke A., & Insects. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210), 763–767. <https://doi.org/10.1126/science.1257570>

Montagu, J., Configurations, M., Deflection, B., & Technologies, S. (1993). United States Patent (19), (19).

Nakai, S. (1983). Structure-Function Relationships of Food Proteins with an Emphasis on the Importance of Protein Hydrophobicity. *Journal of Agricultural and Food Chemistry*, 31(4), 676–683. <https://doi.org/10.1021/jf00118a001>

Nakashima, K., Watanabe, H., Saitoh, H., Tokuda, G., & Azuma, J. I. (2002). Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. *Insect Biochemistry and Molecular Biology*, 32(7), 777–784. [https://doi.org/10.1016/S0965-1748\(01\)00160-6](https://doi.org/10.1016/S0965-1748(01)00160-6)

Ni, J., Wu, Y., Yun, C., Yu, M., & Shen, Y. (2014). cDNA cloning and heterologous expression of an endo- β -1,4-glucanase from the fungus-growing termite *Macrotermes barneyi*. *Archives of Insect Biochemistry and Physiology*, 86(3), 151–164. <https://doi.org/10.1002/arch.21158>

Nicholson, D. B., Ross, A. J., & Mayhew, P. J. (2009). Fossil evidence for key innovations in the evolution of insect diversity : Electronic Supplementary Material, 1–12.

Onho, O. S. Susumi. (1970). *Evolution by Gene Duplication* (Ed. rev.). New York, USA: Springer Science+Business.

Okano, H., Kanaya, E., Ozaki, M., Angkawidjaja, C., & Kanaya, S. (2014). Structure, activity and stability of metagenome-derived glycoside hydrolase family 9 endoglucanase with an N-terminal Ig-like domain. *Protein Sci.* <https://doi.org/10.2210/PDB3X17/PDB>

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). Tackling exascale software challenges in molecular dynamics simulations with GROMACS. *Lecture Notes in Computer Science* (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8759, 3–27. https://doi.org/10.1007/978-3-319-15976-8_1

Paps, J., Baguna, J., & Riutort, M. (2009). Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes. *Proceedings of the Royal Society B: Biological Sciences*, 276(1660), 1245–1254. <https://doi.org/10.1098/rspb.2008.1574>

Parsiegla, G., Belaich, A., Belaich, J. P., & Haser, R. (2002). Crystal structure of the cellulase Cel9M enlightens structure/function relationships of the variable catalytic modules in glycoside hydrolases. *Biochemistry*, 41, 11134–11142. <https://doi.org/10.2210/PDB1IA6/PDB>

Pauchet, Y., Wilkinson, P., Chauhan, R., & Ffrench-Constant, R. H. (2010). Diversity of beetle genes encoding novel plant cell wall degrading enzymes. *PLoS ONE*, 5(12). <https://doi.org/10.1371/journal.pone.0015635>

Pauchet, Y., Wilkinson, P., van Munster, M., Augustin, S., Pauron, D., & Ffrench-Constant, R. H. (2009). Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae*

reveals new gene families in Coleoptera. *Insect Biochemistry and Molecular Biology*, 39(5–6), 403–413. <https://doi.org/10.1016/j.ibmb.2009.04.001>

Pauchet Y, Heckel DG (2013) The genome of the mustard leaf beetle encodes two active xylanases originally acquired from bacteria through horizontal gene transfer. *Proc R Soc B* 280: 201310

Pauchet Y, Saski CA, Feltus FA, Luyten I, Quesneville H, Heckel DG (2014) Studying the organization of genes encoding plant cell wall degrading enzymes in *Chrysomela tremula* provides insights into a leaf beetle genome. *Insect Mol Biol* 23(3):286–300. <https://doi.org/10.1111/imb.12081>

Pereira, J. H., Sapra, R., Volponi, J. V., Kozina, C. L., Simmons, B., & Adams, P. D. (2009). Structure of endoglucanase Cel9A from the thermoacidophilic Alicyclobacillus acidocaldarius. *Acta Crystallogr., Sect. D*, 65, 744–750. <https://doi.org/10.2210/PDB3EZ8/PDB>

Petkun S., Rozman Grinberg I., Lamed R., Jindou S., Burstein T., Yaniv O., Shoham Y., Shimon L. J., Bayer E. A. & Frolow F.. (2015). Reassembly and Co-Crystallization of a Family 9 Processive Endoglucanase from its Component Parts: Structural and Functional Significance of Intermodular Linker. *PeerJ*, 3, E1126. <https://doi.org/10.2210/PDB2XFG/PDB>

Pichler K, Warner K, Magrane M, UniProt Consortium
SPIN: Submitting Sequences Determined at Protein Level to UniProt
Curr. Protoc. Bioinformatics 62(1):e52 (2018)

Pisani, D., & Liu, A. G. (2015). Animal Evolution: Only Rocks Can Set the Clock. *Current Biology*, 25(22), R1079–R1081. <https://doi.org/10.1016/j.cub.2015.10.015>

Popper Z. A., Michel G, Hervé C, Domozych D. S., Willats W. G., Tuohy M. G., Kloareg B, Stengel D. B. (2011). Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. *Annual Review of Plant Biology*, 62(1), 567–590. <https://doi.org/10.1146/annurev-arplant-042110-103809>

Prins, R. a., & Kreulen, D. a. (1991). Comparative aspects of plant cell wall digestion in insects. *Animal Feed Science and Technology*, 32(1–3), 101–118. [https://doi.org/10.1016/0377-8401\(91\)90013-I](https://doi.org/10.1016/0377-8401(91)90013-I)

Pronk, S., et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29, 845–854. <http://dx.doi.org/10.1093/bioinformatics/btt055>

Rambaut Andrew 2009 Computer program distributed by the author, website:
<http://tree.bio.ed.ac.uk/software/figtree/>

Ray D. L. & Julian J. R. (1952) Occurrence of cellulase in Limnoria. *Nature, Lond.* 169, 32–33.

Regier, J. C., Shultz, J. W., & Kambic, R. E. (2005). Pancrustacean phylogeny: Hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proceedings of the Royal Society B: Biological Sciences*, 272(1561), 395–401. <https://doi.org/10.1098/rspb.2004.2917>

Reiter, W. D. (2002). Biosynthesis and properties of the plant cell wall. *Current Opinion in Plant Biology*, 5(6), 536–542. [https://doi.org/10.1016/S1369-5266\(02\)00306-0](https://doi.org/10.1016/S1369-5266(02)00306-0)

Reva, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Folding and Design*, 3(2), 141–147. [https://doi.org/10.1016/S1359-0278\(98\)00019-4](https://doi.org/10.1016/S1359-0278(98)00019-4)

Python Software Foundation. Python Language Reference, version 2.7. Available at <http://www.python.org>

- Sakon, J., Irwin, D., Wilson, D. B., & Karplus, P. A. (1997). Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat.Struct.Mol.Biol.*, 4, 810–818. <https://doi.org/10.2210/PDB1JS4/PDB>
- Sami, A. J., & Shakoori, A. R. (2008). Biochemical characterization of endo-1, 4- β -D-glucanase activity of a green insect pest *Aulacophora foveicollis* (Lucas). *Life Science Journal*, 5(2), 30–36.
- Sanderson, M. J., & Shaffer, H. B. (2002). Troubleshooting Molecular Phylogenetic Analyses. *Annual Review of Ecology and Systematics*, 33(1), 49–72. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150509>
- Santoyo, H., Hernández, G., Romero, R., Hernández-santoyo, A., & Rodríguez-romero, E. G. A. (1999). Disponible en: <http://www.redalyc.org/articulo.oa?id=47543418>
- Sarkar, P., Bosneaga, E., & Auer, M. (2009). Plant cell walls throughout evolution: Towards a molecular understanding of their design principles. *Journal of Experimental Botany*, 60(13), 3615–3635. <https://doi.org/10.1093/jxb/erp245>
- Scharf, M. E., Wu-Scharf, D., Zhou, X., Pittendrigh, B. R., & Bennett, G. W. (2005). Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. *Insect Molecular Biology*, 14(1), 31–44. <https://doi.org/10.1111/j.1365-2583.2004.00527.x>
- Schubot, F. D., Kataeva, I. A., Chang, J., Shah, A. K., Ljungdahl, L. G., Rose, J. P., & Wang, B. C. (2004). Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*. *Biochemistry*, 43, 1163–1170. <https://doi.org/10.2210/PDB1RQ5/PDB>
- Secundo, F. (2013). Conformational changes of enzymes upon immobilisation. *Chemical Society Reviews*, 42(15), 6250–6261. <https://doi.org/10.1039/c3cs35495d>
- Shahriarinnour, M., Noor, M., Wahab, A., Mohamad, R., Mustafa, S., & Ariff, A. B. (2011). Effect of medium composition and cultural condition on cellulase production by *Aspergillus terreus*. *African Journal of Biotechnology*, 10(38), 7459–7467. <https://doi.org/10.5897/AJB11.199>
- Shelomi, M., Heckel, D. G., & Pauchet, Y. (2016). Ancestral gene duplication enabled the evolution of multifunctional cellulases in stick insects (Phasmatodea). *Insect Biochemistry and Molecular Biology*, 71, 1–11. <https://doi.org/10.1016/j.ibmb.2016.02.003>
- Shelomi, M., Jasper, W. C., Atallah, J., Kimsey, L. S., & Johnson, B. R. (2014). Differential expression of endogenous plant cell wall degrading enzyme genes in the stick insect (Phasmatodea) midgut. *BMC Genomics*, 15(1), 917. <https://doi.org/10.1186/1471-2164-15-917>
- Shelomi, M., Watanabe, H., & Arakawa, G. (2014). Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut. *Journal of Insect Physiology*, 60(1), 25–30. <https://doi.org/10.1016/j.jinsphys.2013.10.007>
- Shi, W., Ding, S. Y., & Yuan, J. S. (2011). Comparison of insect gut cellulase and xylanase activity across different insect species with distinct food sources. *Bioenergy Research*, 4(1), 1–10. <https://doi.org/10.1007/s12155-010-9096-0>
- Simakov O. 1., Marletaz F, Cho S. J., Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo D. H., Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, de J. o., Grimwood J, Chapman J. A., Shapiro H, Aerts A, Otilar R. P., Terry A. Y., Boore J. L., Grigoriev I. V., Lindberg D. R., Seaver E. C., Weisblat D. A., Putnam N. H. & Rokhsar D. S. (2014). Insights into bilaterian evolution from three

spiralian genomes. *Nature*, 493(7433), 526–531. <https://doi.org/10.1038/nature11696>. Insights

Shirley, D. M. (2011). Functional characterization of an endoglucanase from *Tribolium castaneum* (TcEG1) in *Saccharomyces cerevisiae* and characterization of cellulolytic activity in the digestive fluids of *Thermobia domestica* (*Zygentoma* : *Lepismatidae*)

Slaytor, M. (1992). Cellulose digestion in termites and cockroaches: What role do symbionts play? *Comparative Biochemistry and Physiology -- Part B: Biochemistry And*, 103(4), 775–784. [https://doi.org/10.1016/0305-0491\(92\)90194-V](https://doi.org/10.1016/0305-0491(92)90194-V)

Soskine, M., & Tawfik, D. S. (2010). Mutational effects and the evolution of new protein functions. *Nature Reviews Genetics*, 11(8), 572–582. <https://doi.org/10.1038/nrg2808>

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>

Strauss, S. & Zangerl. 2002 Plant-insect interactions in terrestrial ecosystems. Pages 77-106
Herrera, C. & Pellmyr O. Plant-Animal Interactions, An evolutionary Approach. Blackwell Publishing, UK

Sugimura, M., Watanabe, H., Lo, N., & Saito, H. (2003). Purification , characterization , cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle , *Psacotha hilaris*, 3460(May), 3455–3460. <https://doi.org/10.1046/j.1432-1033.2003.03735.x>

Sukharnikov, L. O., Cantwell, B. J., Podar, M., & Zhulin, I. B. (2011). Cellulases: Ambiguous nonhomologous enzymes in a genomic perspective. *Trends in Biotechnology*, 29(10), 473–479. <https://doi.org/10.1016/j.tibtech.2011.04.008>

Suzuki, K. I., Ojima, T., & Nishita, K. (2003). Purification and cDNA cloning of a cellulase from abalone *Haliotis discus hannai*. *European Journal of Biochemistry*, 270(4), 771–778. <https://doi.org/10.1046/j.1432-1033.2003.03443.x>

Szydlowski, L., Boschetti, C., Crisp, A., Barbosa, E. G. G., & Tunnacliffe, A. (2015). Multiple horizontally acquired genes from fungal and prokaryotic donors encode cellulolytic enzymes in the bdelloid rotifer *adineta ricciae*. *Gene*, 566(2), 125–137. <https://doi.org/10.1016/j.gene.2015.04.007>

Tamura, K., Stecher, G., Peterson, D., Filipowski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–2729. <https://doi.org/10.1093/molbev/mst197>

Telford, M. J., & Copley, R. R. (2011). Improving animal phylogenies with genomic data. *Trends in Genetics*, 27(5), 186–195. <https://doi.org/10.1016/j.tig.2011.02.003>

Terra, W. (1990). Evolution Of Digestive Systems Of Insects. *Annual Review of Entomology*, 35(1), 181–200. <https://doi.org/10.1146/annurev.ento.35.1.181>

Terrapon N., Li C., Robertson H. M., Ji L., Meng X., Booth W., Chen Z., Childers C. P., Glastad K. M., Gokhale K., Gowin J., Gronenberg W., Hermansen R. A., Hu H., Hunt B. G., Huylmans A. K., Khalil S. M., Mitchell R. D., Munoz-Torres M. C., Mustard J. A., Pan H., Reese J. T., Scharf M. E., Sun F., Vogel H., Xiao J., Yang W., Yang Z., Yang Z., Zhou J., Zhu J., Brent C. S., Elsik C. G., Goodisman M. A., Liberles D. A., Roe R. M., Vargo E. L., Vilcinskis A., Wang J., Bornberg-Bauer E., Korb J., Zhang G., Liebig J. (2014). Molecular traces of alternative social organization in a termite genome. *Nature Communications*, 5(May). <https://doi.org/10.1038/ncomms4636>

- Tokuda, G., Lo, N., Watanabe, H., Slaytor, M., Matsumoto, T., & Noda, H. (1999). Metazoan cellulase genes from termites: Intron/exon structures and sites of expression. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1447(2–3), 146–159. [https://doi.org/10.1016/S0167-4781\(99\)00169-4](https://doi.org/10.1016/S0167-4781(99)00169-4)
- Trautwein, M. D., Wiegmann, B. M., Beutel, R., Kjer, K. M., & Yeates, D. K. (2012). Advances in Insect Phylogeny at the Dawn of the Postgenomic Era. *Annual Review of Entomology*, 57(1), 449–468. <https://doi.org/10.1146/annurev-ento-120710-100538>
- UCSF Chimera a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem*. 2004 Oct;25(13):1605-12.
- Ueda, M., Maruyama, T., Kawasaki, K., Nakazawa, M., & Sakaguchi, M. (2016). Purification, Characterization, and Gene Cloning of a Cold-Adapted Endo-1,4-β-glucanase from *Bellamyia chinensis laeta*. *Molecular Biotechnology*, 58(4), 241–250. <https://doi.org/10.1007/s12033-016-9922-5>
- Von Reumont, B. M., Jenner, R. A., Wills, M. A., Dell’Ampio, E., Pass, G., Ebersberger, I., ... Misof, B. (2012). Pancrustacean phylogeny in the light of new phylogenomic data: Support for remipedia as the possible sister group of hexapoda. *Molecular Biology and Evolution*, 29(3), 1031–1045. <https://doi.org/10.1093/molbev/msr27>
- Wang, S., Ma, J., Peng, J., & Xu, J. (2013). Protein structure alignment beyond spatial proximity. *Scientific Reports*, 3. <https://doi.org/10.1038/srep01448>
- Wang S., Zhang J., Jiao W., Li J., Xun X., Sun Y., Guo X., Huan P., Dong B., Zhang L., Hu X., Sun X., Wang J., Zhao C., Wang Y., Wang D., Huang X., Wang R., Lv J., Li Y., Zhang Z., Liu B., Lu W., Hui Y., Liang J., Zhou Z., Hou R Li X Liu Y., Li H., Ning X., Lin Y., Zhao L., Xing Q., Dou J., Li Y., Mao J., Guo H., Dou H., Li T., Mu C., Jiang W., Fu Q., Fu X., Miao Y Liu J., Yu Q., Li R., Liao H., Li X., Kong Y., Jiang Z., Chourrout D., Li R & Bao Z. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology and Evolution*, 1(5), 1–12. <https://doi.org/10.1038/s41559-017-0120>
- Watanabe, H., & Tokuda, G. (2001). Animal cellulases. *Cell Mol Life Sci*, 58, 1167–1178. <https://doi.org/10.1007/PL00000931>
- Watanabe, H., & Tokuda, G. (2010). Cellulolytic Systems in Insects. *Annual Review of Entomology*, 55(1), 609–632. <https://doi.org/10.1146/annurev-ento-112408-085319>
- Watanabe, H., Noda, H., Tokuda, G., & Lo, N. (1998). A cellulase gene of termite origin. *Nature*, 394(6691), 330–331. <https://doi.org/10.1038/28527>
- Wei, H., Xu, Q., Taylor, L. E., Baker, J. O., Tucker, M. P., & Ding, S. Y. (2009). Natural paradigms of plant cell wall degradation. *Current Opinion in Biotechnology*, 20(3), 330–338. <https://doi.org/10.1016/j.copbio.2009.05.008>
- Wei Y. D., Lee K. S., Gui Z. Z., Yoon H. J., Kim I, Je Y. H., Lee S. M., Zhang G. Z., Guo X, Sohn H. D., Jin B. R. (2006). N-linked glycosylation of a beetle (*Apriona germari*) cellulase Ag-EGase II is necessary for enzymatic activity. *Insect Biochemistry and Molecular Biology*, 36(6), 435–441. <https://doi.org/10.1016/j.ibmb.2006.03.007>
- Weidlich, S., Müller, S., Hoffmann, K. H., & Woodring, J. (2013). Regulation of amylase, cellulase and chitinase secretion in the digestive tract of the two-spotted field cricket, *gryllus bimaculatus*. *Archives of Insect Biochemistry and Physiology*, 83(2), 69–85. <https://doi.org/10.1002/arch.21092>

- Whelan, S., & Goldman, N. (1995). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach, 691–699.
- Willis, J. D., Oppert, B., Oppert, C., Klingeman, W. E., & Jurat-Fuentes, J. L. (2011). Identification, cloning, and expression of a GH9 cellulase from *Tribolium castaneum* (Coleoptera: Tenebrionidae). *Journal of Insect Physiology*, 57(2), 300–306. <https://doi.org/10.1016/j.jinsphys.2010.11.019>
- Willis, J. D., Oppert, C., & Jurat-Fuentes, J. L. (2010). Methods for discovery and characterization of cellulolytic enzymes from insects. *Insect Science*, 17(3), 184–198. <https://doi.org/10.1111/j.1744-7917.2010.01322.x>
- Wilson, D., Irwin, D., Tsao, G., Brainard, A., Bungay, H., Cao, N., ... Xia, L. (1999). Genetics and Properties of Cellulases Recent Progress in Bioconversion of Lignocellulosics, 65, 1–21. <https://doi.org/10.1007/3-540-49>
- Wise, E. L., & Rayment, I. (2004). Understanding the Importance of Protein Structure to Nature's Routes for Divergent Evolution in TIM Barrel Enzymes. *Accounts of Chemical Research*, 37(3), 149–158. https://doi.org/10.1021/ar030250v194-5_1
- Wu, L., & Davies, G. J. (2018). Structure of the GH9 glucosidase/glucosaminidase from *Vibrio cholerae*. *Acta Crystallogr F Struct Biol Commun*, 74, 512–523. <https://doi.org/10.2210/PDB6GDT/PDB>
- Xie, Q., Tian, X., Qin, Y., & Bu, W. (2009). Phylogenetic comparison of local length plasticity of the small subunit of nuclear rDNAs among all Hexapoda orders and the impact of hyper-length-variation on alignment. *Molecular Phylogenetics and Evolution*, 50(2), 310–316. <https://doi.org/10.1016/j.ympev.2008.10.025>
- Xu, J., & Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, 26(7), 889–895. <https://doi.org/10.1093/bioinformatics/btq066>
- Yaoi, K., & Mitsuishi, Y. (2002). Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase. *Journal of Biological Chemistry*, 277(50), 48276–48281. <https://doi.org/10.1074/jbc.M208443200>
- Yokoe, Y., & Yasumasu, I. (1964). the Distribution of Cellulase in Invertebrates. *Comparative Biochemistry and Physiology*, 13, 323–338. [https://doi.org/10.1016/0010-406X\(64\)90027-1](https://doi.org/10.1016/0010-406X(64)90027-1)
- Yoshioka, Y., Tanabe, T., & Iguchi, A. (2017). The presence of genes encoding enzymes that digest carbohydrates in coral genomes and analysis of their activities. *PeerJ*, 5, e4087. <https://doi.org/10.7717/peerj.4087>
- Zhang G., Fang X., Guo X., Li L., Luo R., Xu F., Yang P., Zhang L., Wang X., Qi H., Xiong Z., Que H., Xie Y., Holland PW., Paps J., Zhu Y., Wu F., Chen Y., Wang J., Peng C., Meng J., Yang L., Liu J., Wen B., Zhang N., Huang Z., Zhu Q., Feng Y., Mount A., Hedgecock D., Xu Z., Liu Y., Domazet-Lošo T., Du Y., Sun X., Zhang S., Liu B., Cheng P., Jiang X., Li J., Fan D., Wang W., Fu W., Wang T., Wang B., Zhang J., Peng Z., Li Y., Li N., Wang J., Chen M., He Y., Tan F., Song X., Zheng Q., Huang R., Yang H., Du X., Chen L., Yang M., Gaffney PM., Wang S., Luo L., She Z., Ming Y., Huang W., Zhang S., Huang B., Zhang Y., Qu T., Ni P., Miao G., Wang J., Wang Q., Steinberg CE., Wang H., Li N., Qian L., Zhang G., Li Y., Yang H., Liu X., Wang J., Yin Y. & Wang J. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, 490(7418), 49–54. <https://doi.org/10.1038/nature11413>
- Zhang, J. (2003). Evolution by gene duplication: An update. *Trends in Ecology and Evolution*, 18(6), 292–298. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)

Zhao, F., Ding, M., Li, Y., Yin, Q., & Teng, Y. (2011). Expression and Characterization of Full-Length Ampullaria crossean Endoglucanase EG65s and Their Two Functional Modules .
Zhang, Z., Lin, D., Xin, G., Yan, G., & Jingfa Xiao. (2012). Bioinformatics clouds for big data manipulation. *Biology Direct*, 7, 43; discussion 43. Retrieved from <http://www.biologydirect.com/content/7/1/43>

Bioscience, Biotechnology, and Biochemistry, 75(2), 240–246. <https://doi.org/10.1271/bbb.100529>

Zhou, X., Smith, J. A., Oi, F. M., Koehler, P. G., Bennett, G. W., & Scharf, M. E. (2007). Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*. *Gene*, 395(1–2), 29–39. <https://doi.org/10.1016/j.gene.2007.01.004>

Zhou, X., Smith, J. A., Oi, F. M., Koehler, P. G., Bennett, G. W., & Scharf, M. E. (2007). Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*. *Gene*, 395(1–2), 29–39. <https://doi.org/10.1016/j.gene.2007.01.004>

Anexo 1

Tabla 1.- Secuencias utilizadas en el alineamiento 1.

Grupo de organismos	Especies pertenecientes al grupo	No. De Acceso del NCBI
Bacteria	<i>Cellulomonas Fimi</i>	AAA23086.1
Cnidaria	<i>Exaiptasia pallida</i> , <i>Stylophora pistillata</i> , <i>Orbicella faveolata</i>	KXJ18436.1, PFX25363.1, XP_020614867.1
Mollusca	<i>Ampullaria crossean</i> , <i>Haliotis discus discus</i> , <i>Haliotis tuberculata</i> , <i>Haliotis gigantea</i> , <i>Haliotis discus hannai</i> , <i>Haliotis discus</i> , <i>Bellamya sp. UM,2014</i> , <i>Aplysia kurodai</i> , <i>Pomacea canaliculata</i> , <i>Aplysia californica</i> , <i>Lottia gigantea</i> , <i>Biomphalaria glabrata</i> , <i>Octopus bimaculoides</i> , <i>Crassostrea gigas</i> , <i>Crassostrea virginica</i> , <i>Mizuhopecten yessoensis</i> , <i>Achatina fulica</i> , <i>Corbicula japonica</i>	ABD24274.1, ABO26609.1, AFQ98379.1, AFQ98380.1, AGI61069.1, BAC67186.1, BAS18736.1, BAX09271.1, PVD35337.1, XP_005099312.1, XP_009049791.1, XP_013066277.1, KOF68403.1, EKC17289.1, XP_022328778.1, OWF45593.1, ATS91249.1, BAF38757.1
Annelida	<i>Eisenia fetida</i> , <i>Eisenia andrei</i> , <i>Metaphire hilgendorfi</i> , <i>Perinereis aibuhitensis</i> , <i>Perinereis brevicirris</i>	pdb 3WC3 A, AAX92641.1, BAH22180.1, ANR02619.1, BAK20401.1
Rotifera	<i>Adineta ricciae</i>	AKC91349.1
Crustacea	<i>Hyalella azteca</i> , <i>Daphnia pulex</i> , <i>Daphnia magna</i> , <i>Limnoria quadripunctata</i> , <i>Neomysis intermedia</i> , <i>Cherax quadricarinatus</i> , <i>Parasesarma erythrodactyla</i> , <i>Gecarcoidea natalis</i> , <i>Mictyris platycheles</i> , <i>Coenobita brevimanus</i> , <i>Coenobita rugosus</i>	XP_018016806.1, EFX69372.1, KZS03923.1, ADB85440.1, BAL60587.1, AAD38027.1, AIT11911.1, APF29556.1, APF29559.1, APF29560.1, APF29561.1
Collembola	<i>Orchesella cincta</i> , <i>Folsomia candida</i>	ODM94862.1, XP_021956548.1
Orthoptera	<i>Teleogryllus emma</i>	ABV32557.1
Embioptera	<i>Antipaluria urichi</i>	AOV94250.1
Phasmatodea	<i>Aretaon asperrimus</i> , <i>Extatosoma tiaratum</i> , <i>Medauroidea extradentata</i> , <i>Peruphasma schultei</i> , <i>Ramulus artemis</i> , <i>Sipyloidea sipyulus</i> , <i>Timema cristinae</i>	AMH40356.1, AMH40362.1, AMH40366.1, AMH40373.1, AMH40379.1, AMH40384.1, AMH40392.1
Blaberoidea	<i>Blattella germanica</i> , <i>Panesthia cribrata</i>	PSN31180.1, AAF80584.1
Termitoidae	<i>Coptotermes acinaciformis</i> , <i>Reticulitermes flavipes</i> , <i>Coptotermes formosanus</i> , <i>Odontotermes formosanus</i> , <i>Macrotermes barneyi</i> , <i>Amitermes dentatus</i> , <i>Amitermes foreli</i> , <i>Anoplotermes schwarzi</i> , <i>Constrictotermes cavifrons</i> , <i>Constrictotermes guantanamoensis</i> , <i>Grigiotermes hageni</i> , <i>Globitermes sulphureus</i> , <i>Hospitalitermes bicolor</i> , <i>Microcerotermes crassus</i> , <i>Microtermes pallidus</i> , <i>Macrotermes carbonarius</i> , <i>Macrotermes gilvus</i> , <i>Macrotermes</i>	AAK12339.1, AAU20853.2, ADB12483.1, ADB82658.1, AFD33365.1, AGP76397.1, AGP76398.1, AGP76400.1, AGP76402.1, AGP76404.1, AGP76405.1, AGP76406.1, AGP76408.1, AGP76411.1, AGP76412.1, AGP76414.1, AGP76415.1, AGP76416.1, AGP76417.1, AGP76418.1, AGP76420.1, AGP76422.1, AGP76424.1, AGP76425.1, AGP76426.1, AGP76427.1,

	<i>subhyalinus</i> , <i>Nasutitermes corniger</i> , <i>Nasutitermes</i> sp. <i>NpEG</i> , <i>Odontotermes</i> <i>hainanensis</i> , <i>Pericapritermes nitobei</i> , <i>Pericapritermes</i> sp. <i>PpEG1</i> , <i>Pericapritermes</i> sp. <i>PpEG2</i> , <i>Pericapritermes</i> sp. <i>PpEG3</i> , <i>Rhynchotermes bulbinasus</i> , <i>Subulitermes baileyi</i> , <i>Syntermes</i> <i>grandis</i> , <i>Sphaerotermes sphaerotherax</i> , <i>Coptotermes gestroi</i> , <i>Microtermes</i> <i>pakistanicus</i> , <i>Reticulitermes speratus</i> , <i>Nasutitermes takasagoensis</i> , <i>Nasutitermes walkeri</i> , <i>Mastotermes</i> <i>darwiniensis</i> , <i>Zootermopsis nevadensis</i> , <i>Cryptotermes secundus</i>	AGP76430.1, AGP76431.1, AGP76434.1, AGS32241.1, AKV16371.1, BAA31326.1, BAA33708.1, BAA33709.1, CAD54726.1, KDR16731.1, XP_023704929.1
Hemiptera	<i>Acyrtosiphon pisum</i> , <i>Myzus persicae</i> , <i>Nilaparvata lugens</i>	XP_001944774.2, XP_008181842.1, XP_022167267.1, XP_022167268.1, XP_022200004.1
Phthiraptera	<i>Pediculus humanus corporis</i>	XP_002426465.1
Coleoptera	<i>Eurytemora affinis</i> , <i>Tribolium</i> <i>castaneum</i> , <i>Nicrophorus vespilloides</i> , <i>Agrius planipennis</i> , <i>Anoplophora</i> <i>glabripennis</i>	XP_023337121.1, KYB27092.1, XP_017785110.1, XP_018323591.1, XP_018568194.1
Hymenoptera	<i>Camponotus floridanus</i> , <i>Acromyrmex</i> <i>echinator</i> , <i>Lasius niger</i> , <i>Habropoda</i> <i>laboriosa</i> , <i>Melipona quadrifasciata</i> , <i>Atta colombica</i> , <i>Cyphomyrmex</i> <i>costatus</i> , <i>Trachymyrmex cornetzi</i> , <i>Trachymyrmex septentrionalis</i> , <i>Trachymyrmex zeteki</i> , <i>Dufourea</i> <i>novaeangliae</i> , <i>Eufriesea mexicana</i> , <i>Apis cerana cerana</i> , <i>Nasonia</i> <i>vitripennis</i> , <i>Bombus terrestris</i> , <i>Apis</i> <i>florea</i> , <i>Apis dorsata</i> , <i>Solenopsis</i> <i>invicta</i> , <i>Fopius arisanus</i> , <i>Ooceraea</i> <i>biroi</i> , <i>Ceratosolen solmsi marchali</i> , <i>Pogonomyrmex barbatus</i> , <i>Wasmannia auropunctata</i> , <i>Vollenhovia emeryi</i> , <i>Atta cephalotes</i> , <i>Megachile rotundata</i> , <i>Linepithema</i> <i>humile</i> , <i>Bombus impatiens</i> , <i>Orussus</i> <i>abietinus</i> , <i>Monomorium pharaonis</i> , <i>Copidosoma floridanum</i> , <i>Polistes</i> <i>canadensis</i> , <i>Diachasma alloeum</i> , <i>Polistes dominula</i> , <i>Neodiprion</i> <i>lecontei</i> , <i>Cephus cinctus</i> , <i>Apis</i> <i>cerana</i> , <i>Pseudomyrmex gracilis</i> , <i>Apis</i> <i>mellifera</i>	EFN70196.1, EGI63652.1, KMQ95413.1, KOC62382.1, KOX76353.1, KYM87195.1, KYN01696.1, KYN22851.1, KYN35700.1, KYQ52831.1, KZC10616.1, OAD61293.1, PBC31793.1, XP_001606454.1, XP_003402778.1, XP_003690676.1, XP_006610535.1, XP_011164235.1, XP_011296914.1, XP_011346866.1, XP_011498432.1, XP_011632703.1, XP_011700378.1, XP_011867251.1, XP_012061951.1, XP_012146513.1, XP_012224814.1, XP_012248155.1, XP_012273697.1, XP_012526654.1, XP_014214612.1, XP_014601854.1, XP_015126097.1, XP_015184471.1, XP_015513382.1, XP_015592927.1, XP_016907626.1, XP_020289295.1, XP_396791.3 PR

Para este alineamiento se utilizó una secuencia por especie, de las secuencias obtenidos y depuradas de la búsqueda en el GenBank, agregando la secuencia de Bacteria (*Cellulomonas Fimi*)

Tabla 2.- Secuencias utilizadas en el alineamiento 2.

Grupo de organismos	Especies pertenecientes al grupo	No. De Acceso del NCBI
Collembola	<i>Orchesella cincta</i> , <i>Folsomia candida</i>	ODM94862.1- ODM95900.1- XP_021956548.1- XP_021956549.1- XP_021959948.1- XP_021966378.1- XP_021967587.1
Orthoptera	<i>Teleogryllus emma</i>	ABV32557.1
Embioptera	<i>Antipaluria urichi</i>	AOV94250.1-AOV94256.1
Phasmatodea	<i>Timema cristinae</i> , <i>Aretaon asperrimus</i> , <i>Extatosoma tiaratum</i> , <i>Medauroidea extradentata</i> , <i>Peruphasma schultei</i> , <i>Ramulus artemis</i> , <i>Sipyloidea sipyulus</i>	AMH40392.1- AMH40395.1- AMH40356.1-AMH40391.1
Blaberoidea	<i>Blattella germanica</i> , <i>Panesthia cribrata</i>	PSN31180.1- AAF80584.1- AAF80585.1
Termitoidae	<i>Cryptotermes secundus</i> , <i>Coptotermes acinaciformis</i> , <i>Reticulitermes flavipes</i> , <i>Coptotermes formosanus</i> , <i>Odontotermes formosanus</i> , <i>Macrotermes barneyi</i> , <i>Amitermes dentatus</i> , <i>Amitermes foreli</i> , <i>Anoplotermes schwarzi</i> , <i>Constrictotermes cavifrons</i> , <i>Constrictotermes guantanamoensis</i> , <i>Grigiotermes hageni</i> , <i>Globitermes sulphureus</i> , <i>Hospitalitermes bicolor</i> , <i>Microcerotermes crassus</i> , <i>Microtermes pallidus</i> , <i>Macrotermes carbonarius</i> , <i>Macrotermes gilvus</i> , <i>Macrotermes subhyalinus</i> , <i>Nasutitermes corniger</i> , <i>Nasutitermes sp. NpEG</i> , <i>Odontotermes hainanensis</i> , <i>Pericapritermes nitobei</i> , <i>Pericapritermes sp. PpEG1</i> , <i>Pericapritermes sp. PpEG2</i> , <i>Pericapritermes sp. PpEG3</i> , <i>Rhynchotermes bulbinasus</i> , <i>Subulitermes baileyi</i> , <i>Syntermes grandis</i> , <i>Sphaerotermes sphaerothorax</i> , <i>Coptotermes gestroi</i> , <i>Microtermes pakistanicus</i> , <i>Reticulitermes speratus</i> , <i>Nasutitermes takasagoensis</i> , <i>Nasutitermes walkeri</i> , <i>Mastotermes darwiniensis</i> , <i>Zootermopsis nevadensis</i>	XP_023704929.1- XP_023704930.1- XP_023716596.1- XP_023718476.1- AAK12339.1- AAU20853.1- ADB12483.1- BAB40693.1- BAB40695.1- BAB40696.1- BAB40697.1- ADB82658.1- AGP76419.1- AFD33365.1- AGP76397.1-AGP76418.1- AGP76420.1-AGP76437.1- AGS32241.1- AKV16371.1- BAA31326.1- BAA34050.1- BAA33708.1- BAA33709.1- CAD54726.1- CAD54727.1- CAD54728.1- CAD54729.1- CAD54730.1- KDR16731.1- XP_021924915.1- XP_021935277.1- XP_021941324.1
Hemiptera	<i>Acyrtosiphon pisum</i> , <i>Myzus persicae</i> , <i>Nilaparvata lugens</i>	XP_001944774.2, XP_008181842.1, XP_022167267.1, XP_022167268.1, XP_022200004.1
Phthiraptera	<i>Pediculus humanus corporis</i>	XP_002426465.1
Coleoptera	<i>Tribolium castaneum</i> , <i>Nicrophorus</i>	KYB27092.1- XP_001810693.2-

	<i>vespilloides, Agrilus planipennis, Anoplophora glabripennis</i>	XP_008194455.1- XP_015836219.1- XP_017785110.1- XP_018323591.1- XP_018568194.1- XP_018568196.1
Hymenoptera	<i>Apis mellifera, Camponotus floridanus, Acromyrmex echinator, Lasius niger, Habropoda laboriosa, Melipona quadrifasciata, Atta colombica, Cyphomyrmex costatus, Trachymyrmex cornetzi, Trachymyrmex septentrionalis, Trachymyrmex zeteki, Dufourea novaeangliae, Eufriesea mexicana, Apis cerana cerana, Nasonia vitripennis, Bombus terrestris, Apis florea, Apis dorsata, Solenopsis invicta, Fopius arisanus, Ooceraea biroi, Ceratosolen solmsi marchali, Pogonomyrmex barbatus, Wasmannia auropunctata, Vollenhovia emeryi, Atta cephalotes, Megachile rotundata, Linepithema humile, Bombus impatiens, Orussus abietinus, Monomorium pharaonis, Copidosoma floridanum, Polistes canadensis, Diachasma alloeum, Polistes dominula, Neodiprion lecontei, Cephus cinctus, Apis cerana, Pseudomyrmex gracilis</i>	XP_396791.3- EFN70196.1- EFN70197.1- XP_011253653.1- XP_011253654.2- XP_011253655.2- XP_019884179.2- EGI63652.1- XP_011057333.1- KMQ95413.1- KOC62382.1- XP_017793200.1- KOX76353.1- KYM87195.1- XP_018044452.1- KYN01696.1- XP_018396635.1- KYN22851.1- XP_018359268.1- KYN35700.1- XP_018347393.1- KYQ52831.1- XP_018307051.1- KZC10616.1- XP_015432538.1- OAD61293.1- XP_017766299.1- PBC31793.1- XP_001606454.1- XP_003402778.1- XP_003690676.1- XP_006610535.1- XP_011164235.1- XP_011296914.1- XP_011296916.1- XP_011346866.1- XP_011498432.1- XP_011632703.1- XP_011700378.1- XP_011867251.1- XP_012061951.1- XP_012146513.1- XP_012224814.1- XP_012248155.1- XP_012273697.1- XP_012526654.1- XP_014214612.1- XP_014601854.1- XP_015126097.1- XP_015184471.1- XP_015513382.1- XP_015592927.1- XP_016907626.1- XP_020289295.1- XP_020289297.1- XP_020289301.1- XP_020289336.1

Para este alineamiento se utilizaron las secuencias obtenidas y depuradas de la búsqueda en el GenBank pertenecientes al subfilo hexapoda.

Tabla 3.- Matriz de valores de RMSD

Modelo / Modelo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1/AOV94253.1	0	1.1	1.1	0.4	1	1	0.4	0.5	1	1	0.5	1	1	1	1	1.2	1.5	1.8	1	0.6
2/XP_021956548.1	1.1	0	0.7	1.1	0.6	0.6	1.1	1.2	0.6	0.9	1.3	0.6	1.1	0.5	0.7	0.9	1.7	1.6	1.3	1.2
3/XP_396791.3	1.1	0.7	0	1	0.4	0.5	1.1	1.1	0.4	0.8	1.2	0.5	1.1	0.4	0.6	0.7	1.7	1.6	1.3	1.1
4/AMH40395.1	0.4	1.1	1	0	1	0.9	0.1	0.4	1	1	0.7	1	0.9	0.9	1	1.2	1.4	1.8	1	0.7
5/AMH40372.1	1	0.6	0.4	1	0	0.4	1	1	0.2	0.7	1.2	0.3	1	0.2	0.5	0.8	1.6	1.5	1.2	1.1
6/EFX69372.1	1	0.6	0.5	0.9	0.4	0	1	1	0.4	0.6	1.1	0.4	1	0.3	0.5	0.8	1.7	1.6	1.2	1
7/XP_023716596.1	0.4	1.1	1.1	0.1	1	1	0	0.4	1	1	0.7	1	0.9	0.9	1	1.2	1.4	1.8	1	0.7
8/XP_002426465.1	0.5	1.2	1.1	0.4	1	1	0.4	0	1.1	1.1	0.8	1	1	1	1	1.2	1.4	1.8	1.1	0.8
9/XP_011296914.1	1	0.6	0.4	1	0.2	0.4	1	1.1	0	0.7	1.2	0.4	1	0.3	0.5	0.8	1.6	1.6	1.2	1.1
10/AMH40373.1	1	0.9	0.8	1	0.7	0.6	1	1.1	0.7	0	1	0.7	1	0.7	0.6	1	1.8	1.7	1.5	1
11/XP_022200004.1	0.5	1.3	1.2	0.7	1.2	1.1	0.7	0.8	1.2	1	0	1.2	1	1.1	1.2	1.4	1.6	1.9	1.2	0.7
12/XP_008194455.1	1	0.6	0.5	1	0.3	0.4	1	1	0.4	0.7	1.2	0	1	0.2	0.5	0.8	1.6	1.5	1.2	1

13/XP_017785110.1	1	1.1	1.1	0.9	1	1	0.9	1	1	1	1	1	0	1	1	1.3	1.8	1.8	1.3	0.9
14/AAF80584.1	1	0.5	0.4	0.9	0.2	0.3	0.9	1	0.3	0.7	1.1	0.2	1	0	0.5	0.7	1.6	1.5	1.2	1
15/AOV94256.1	1	0.7	0.6	1	0.5	0.5	1	1	0.5	0.6	1.2	0.5	1	0.5	0	0.8	1.7	1.6	1.3	1.1
16/EFN70196.1	1.2	0.9	0.7	1.2	0.8	0.8	1.2	1.2	0.8	1	1.4	0.8	1.3	0.7	0.8	0	1.8	1.7	1.4	1.3
17/XP_001944774.1	1.5	1.7	1.7	1.4	1.6	1.7	1.4	1.4	1.6	1.8	1.6	1.6	1.8	1.6	1.7	1.8	0	1.4	1.6	1.5
18/AOV94250.1	1.8	1.6	1.6	1.8	1.5	1.6	1.8	1.8	1.6	1.7	1.9	1.5	1.8	1.5	1.6	1.7	1.4	0	1.9	1.8
19/AKC91349.1	1	1.3	1.3	1	1.2	1.2	1	1.1	1.2	1.5	1.2	1.2	1.3	1.2	1.3	1.4	1.6	1.9	0	1.2
20/KXJ18436.1	0.6	1.2	1.1	0.7	1.1	1	0.7	0.8	1.1	1	0.7	1	0.9	1	1.1	1.3	1.5	1.8	1.2	0
21/ADB85440.1	1.1	0.7	0.6	1	0.4	0.5	1	1.1	0.5	0.8	1.2	0.4	1.1	0.4	0.6	0.8	1.7	1.6	1.3	1.1
22/AGP76427.1	1	0.5	0.4	0.9	0.2	0.3	0.9	1	0.3	0.7	1.1	0.2	1	0.1	0.5	0.7	1.6	1.5	1.2	1
23/XP_023704929.1	1	0.5	0.4	0.9	0.2	0.3	0.9	1	0.3	0.7	1.1	0.3	1	0.1	0.5	0.7	1.6	1.5	1.2	1
24/ODM94862.1	1.2	1.3	1.4	1.2	1.3	1.3	1.2	1.2	1.3	1.5	1.3	1.3	1.4	1.3	1.4	1.4	1.7	1.9	0.6	1.2
25/KZC10616.1	0.9	1.4	1.3	1	1.3	1.3	1	0.9	1.3	1.3	1	1.3	1.3	1.3	1.3	1.4	1.7	2	1.3	1
26/ABV32557.1	1.1	1.3	1.2	1.1	1.2	1.1	1.1	1.1	1.2	1.2	1.2	1.2	1.1	1.1	1.2	1.4	1.6	1.9	1.3	1
27/KDR16731.1	0.4	1.1	1	0.1	1	1	0.1	0.4	1	1	0.7	1	0.9	0.9	1	1.2	1.4	1.8	1	0.7
28/XP_013066277.1	1.2	1.3	1.3	1.2	1.2	1.2	1.2	1.2	1.3	1.2	1.2	1.2	0.8	1.2	1.2	1.4	1.9	1.9	1.5	1.2
29/BAK20401.1	0.9	1.1	1.1	0.9	1	1	0.9	1	1	1	1	1	0.5	1	1	1.3	1.8	1.8	1.3	0.9
30/BAA31326.1	1	0.5	0.4	0.9	0.2	0.3	0.9	1	0.3	0.7	1.1	0.3	1	0.1	0.5	0.7	1.6	1.5	1.2	1
31/XP_018568194.1	0.9	1.4	1.3	0.9	1.3	1.3	0.9	0.9	1.3	1.2	0.9	1.3	1.2	1.3	1.3	1.4	1.6	2	1.3	1
32/AKV16371.1	1	0.5	0.4	0.9	0.2	0.3	0.9	1	0.3	0.7	1.1	0.3	1	0.1	0.5	0.7	1.6	1.5	1.2	1
33/XP_020614867.1	0.6	1.3	1.2	0.8	1.2	1.1	0.8	0.8	1.2	1	0.5	1.2	0.9	1.1	1.2	1.4	1.7	1.9	1.3	0.6
34/PSN31180.1	2	1.9	1.8	2	1.8	1.8	2	2	1.8	1.9	2.1	1.8	2.1	1.8	1.8	1.9	2.1	1.9	2.1	2
35/XP_022167267.1	1.5	1.7	1.7	1.4	1.6	1.7	1.4	1.5	1.7	1.8	1.6	1.6	1.8	1.6	1.7	1.8	0.1	1.4	1.6	1.5
36/AMH40392.1	0.9	0.7	0.6	0.9	0.4	0.3	0.9	1	0.5	0.5	1.1	0.5	1	0.4	0.4	0.8	1.7	1.6	1.3	1
37/AMH40379.1	1	0.5	0.4	0.9	0.2	0.3	0.9	1	0.3	0.7	1.1	0.3	1	0.1	0.5	0.7	1.6	1.5	1.2	1
38/APF29560.1	1.6	1.4	1.3	1.6	1.3	1.3	1.6	1.6	1.3	1.4	1.7	1.3	1.6	1.3	1.3	1.5	2	2	1.7	1.6
39/CAD54726.1	1	0.6	0.5	1	0.3	0.4	1	1.1	0.4	0.7	1.2	0.4	1	0.3	0.5	0.8	1.6	1.6	1.2	1.1
40/XP_009049791.1	0.6	1.2	1.1	0.5	1	1	0.5	0.6	1.1	1	0.7	1	0.8	1	1.1	1.3	1.6	1.8	1.2	0.7
41/EFX86822.1	1.7	1.5	1.4	1.6	1.3	1.4	1.6	1.7	1.3	1.5	1.8	1.4	1.7	1.3	1.4	1.5	1.3	1	1.7	1.7

Modelo / Modelo	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
1/AOV94253.1	1.1	1	1	1.2	0.9	1.1	0.4	1.2	0.9	1	0.9	1	0.6	2	1.5	0.9	1	1.6	1	0.6	1.7
2/XP_021956548.1	0.7	0.5	0.5	1.3	1.4	1.3	1.1	1.3	1.1	0.5	1.4	0.5	1.3	1.9	1.7	0.7	0.5	1.4	0.6	1.2	1.5
3/XP_396791.3	0.6	0.4	0.4	1.4	1.3	1.2	1	1.3	1.1	0.4	1.3	0.4	1.2	1.8	1.7	0.6	0.4	1.3	0.5	1.1	1.4
4/AMH40395.1	1	0.9	0.9	1.2	1	1.1	0.1	1.2	0.9	0.9	0.9	0.9	0.8	2	1.4	0.9	0.9	1.6	1	0.5	1.6
5/AMH40372.1	0.4	0.2	0.2	1.3	1.3	1.2	1	1.2	1	0.2	1.3	0.2	1.2	1.8	1.6	0.4	0.2	1.3	0.3	1	1.3
6/EFX69372.1	0.5	0.3	0.3	1.3	1.3	1.1	1	1.2	1	0.3	1.3	0.3	1.1	1.8	1.7	0.3	0.3	1.3	0.4	1	1.4
7/XP_023716596.1	1	0.9	0.9	1.2	1	1.1	0.1	1.2	0.9	0.9	0.9	0.9	0.8	2	1.4	0.9	0.9	1.6	1	0.5	1.6
8/XP_002426465.1	1.1	1	1	1.2	0.9	1.1	0.4	1.2	1	1	0.9	1	0.8	2	1.5	1	1	1.6	1.1	0.6	1.7
9/XP_011296914.1	0.5	0.3	0.3	1.3	1.3	1.2	1	1.3	1	0.3	1.3	0.3	1.2	1.8	1.7	0.5	0.3	1.3	0.4	1.1	1.3
10/AMH40373.1	0.8	0.7	0.7	1.5	1.3	1.2	1	1.2	1	0.7	1.2	0.7	1	1.9	1.8	0.5	0.7	1.4	0.7	1	1.5

11/XP_022200004.1	1.2	1.1	1.1	1.3	1	1.2	0.7	1.2	1	1.1	0.9	1.1	0.5	2.1	1.6	1.1	1.1	1.7	1.2	0.7	1.8
12/XP_008194455.1	0.4	0.2	0.3	1.3	1.3	1.2	1	1.2	1	0.3	1.3	0.3	1.2	1.8	1.6	0.5	0.3	1.3	0.4	1	1.4
13/XP_017785110.1	1.1	1	1	1.4	1.3	1.1	0.9	0.8	0.5	1	1.2	1	0.9	2.1	1.8	1	1	1.6	1	0.8	1.7
14/AAF80584.1	0.4	0.1	0.1	1.3	1.3	1.1	0.9	1.2	1	0.1	1.3	0.1	1.1	1.8	1.6	0.4	0.1	1.3	0.3	1	1.3
15/AOV94256.1	0.6	0.5	0.5	1.4	1.3	1.2	1	1.2	1	0.5	1.3	0.5	1.2	1.8	1.7	0.4	0.5	1.3	0.5	1.1	1.4
16/EFN70196.1	0.8	0.7	0.7	1.4	1.4	1.4	1.2	1.4	1.3	0.7	1.4	0.7	1.4	1.9	1.8	0.8	0.7	1.5	0.8	1.3	1.5
17/XP_001944774.1	1.7	1.6	1.6	1.7	1.7	1.6	1.4	1.9	1.8	1.6	1.6	1.6	1.7	2.1	0.1	1.7	1.6	2	1.6	1.6	1.3
18/AOV94250.1	1.6	1.5	1.5	1.9	2	1.9	1.8	1.9	1.8	1.5	2	1.5	1.9	1.9	1.4	1.6	1.5	2	1.6	1.8	1
19/AKC91349.1	1.3	1.2	1.2	0.6	1.3	1.3	1	1.5	1.3	1.2	1.3	1.2	1.3	2.1	1.6	1.3	1.2	1.7	1.2	1.2	1.7
20/KXJ18436.1	1.1	1	1	1.2	1	1	0.7	1.2	0.9	1	1	1	0.6	2	1.5	1	1	1.6	1.1	0.7	1.7
21/ADB85440.1	0	0.4	0.4	1.3	1.4	1.2	1	1.3	1.1	0.4	1.4	0.4	1.2	1.8	1.7	0.6	0.4	1.3	0.3	1.1	1.4
22/AGP76427.1	0.4	0	0.1	1.3	1.3	1.1	0.9	1.2	1	0.1	1.3	0.1	1.1	1.8	1.6	0.4	0.1	1.3	0.3	1	1.3
23/XP_023704929.1	0.4	0.1	0	1.3	1.3	1.1	0.9	1.2	1	0	1.3	0	1.1	1.8	1.6	0.4	0.1	1.3	0.3	1	1.3
24/ODM94862.1	1.3	1.3	1.3	0	1.4	1.4	1.2	1.5	1.4	1.3	1.4	1.3	1.4	2.2	1.7	1.4	1.3	1.7	1.3	1.3	1.8
25/KZC10616.1	1.4	1.3	1.3	1.4	0	1.3	1	1.4	1.3	1.3	1	1.3	1.1	2.2	1.7	1.3	1.3	1.8	1.3	1.1	1.9
26/ABV32557.1	1.2	1.1	1.1	1.4	1.3	0	1.1	1.2	1.1	1.1	1.3	1.1	1.3	2.1	1.6	1.1	1.1	1.6	1.2	1.3	1.8
27/KDR16731.1	1	0.9	0.9	1.2	1	1.1	0	1.2	0.9	0.9	0.9	0.9	0.8	2	1.4	0.9	0.9	1.6	1	0.5	1.6
28/XP_013066277.1	1.3	1.2	1.2	1.5	1.4	1.2	1.2	0	0.7	1.2	1.4	1.2	1.2	2.2	1.9	1.2	1.2	1.7	1.3	1	1.8
29/BAK20401.1	1.1	1	1	1.4	1.3	1.1	0.9	0.7	0	1	1.2	1	0.9	2	1.8	1	1	1.6	1	0.8	1.7
30/BAA31326.1	0.4	0.1	0	1.3	1.3	1.1	0.9	1.2	1	0	1.3	0	1.1	1.8	1.6	0.4	0.1	1.3	0.3	1	1.3
31/XP_018568194.1	1.4	1.3	1.3	1.4	1	1.3	0.9	1.4	1.2	1.3	0	1.3	1	2.2	1.7	1.3	1.3	1.8	1.3	1	1.9
32/AKV16371.1	0.4	0.1	0	1.3	1.3	1.1	0.9	1.2	1	0	1.3	0	1.1	1.8	1.6	0.4	0.1	1.3	0.3	1	1.3
33/XP_020614867.1	1.2	1.1	1.1	1.4	1.1	1.3	0.8	1.2	0.9	1.1	1	1.1	0	2.1	1.7	1.1	1.1	1.7	1.2	0.6	1.8
34/PSN31180.1	1.8	1.8	1.8	2.2	2.2	2.1	2	2.2	2	1.8	2.2	1.8	2.1	0	2.1	1.8	1.8	2.2	1.8	2	1.9
35/XP_022167267.1	1.7	1.6	1.6	1.7	1.7	1.6	1.4	1.9	1.8	1.6	1.7	1.6	1.7	2.1	0	1.7	1.6	2	1.7	1.6	1.3
36/AMH40392.1	0.6	0.4	0.4	1.4	1.3	1.1	0.9	1.2	1	0.4	1.3	0.4	1.1	1.8	1.7	0	0.4	1.3	0.5	1	1.5
37/AMH40379.1	0.4	0.1	0.1	1.3	1.3	1.1	0.9	1.2	1	0.1	1.3	0.1	1.1	1.8	1.6	0.4	0	1.3	0.3	1	1.3
38/APF29560.1	1.3	1.3	1.3	1.7	1.8	1.6	1.6	1.7	1.6	1.3	1.8	1.3	1.7	2.2	2	1.3	1.3	0	1.3	1.6	1.8
39/CAD54726.1	0.3	0.3	0.3	1.3	1.3	1.2	1	1.3	1	0.3	1.3	0.3	1.2	1.8	1.7	0.5	0.3	1.3	0	1.1	1.4
40/XP_009049791.1	1.1	1	1	1.3	1.1	1.3	0.5	1	0.8	1	1	1	0.6	2	1.6	1	1	1.6	1.1	0	1.7
41/EFX86822.1	1.4	1.3	1.3	1.8	1.9	1.8	1.6	1.8	1.7	1.3	1.9	1.3	1.8	1.9	1.3	1.5	1.3	1.8	1.4	1.7	0

Resultados obtenidos de comparar las estructuras de los modelos generados por SWISS Model.

Tabla 4.- En esta tabla se agrupan las secuencias de celulasas de la familia GH9 en base a los resultados del análisis de sitios de unión.

Tipo de sitios de unión	Secuencias
Sin sitios de unión	KZS03923.1, XP_021967587.1, XP_008194455.1, XP_015836219.1, XP_001810693.2, XP_017785110.1, XP_001944774.2, XP_022167268.1, XP_001606454.1, XP_011498432.1, XP_014214612.1, XP_003690676.1, XP_016907626.1, OAD61293.1, XP_017766299.1, XP_003402778.1, XP_012248155.1, KOC62382.1, KOX76353.1, XP_015432538.1, KZC10616.1, XP_011296916.1, XP_012273697.1, XP_011253655.1, XP_020289295.1, XP_020289297.1, KYM87195.1, KYN22851.1, KYQ52831.1, XP_018307051.1, XP_011164235.1, XP_011700378.1, XP_012526654.1, XP_011867251.1, XP_012224814.1, XP_014601854.1, AMH40395.1, XP_002426465.1, AOV94250.1, AMH40370.1, AMH40376.1, AMH40383.1, AMH40377.1, AMH40369.1, AMH40393.1, AMH40364.1, AMH40368.1, AMH40371.1, AMH40366.1, AMH40373.1, AMH40374.1, AMH40363.1, AOV94254.1, AOV94252.1, ABV32557.1, AAF80585.1, PSN31180.1, CAD54726.1, XP_023704929.1, XP_021935277.1, XP_023718476.1, AGP76420.1, AGP76421.1, AGP76419.1, AGP76412.1, AGP76414.1, AAU20853.2, BAB40695.1, AKV16371.1, AAK12339.1, BAA33708.1, BAA33709.1, AGP76406.1, AGP76427.1, AGP76404.1, AGP76408.1, AGP76409.1, AGP76410.1, AGP76411.1, AGP76418.1, AGP76424.1, AGP76425.1, AGP76423.1, XP_018016813.1, XP_018016810.1, XP_018027134.1, ADB85442.1, ADB85441.1, XP_018026496.1, APF29560.1, AAO61672.2, XP_009049840.1, KOF68403.1, XP_014786694.1, XP_009049791.1, XP_021363495.1, XP_021363497.1, ABD24281.1, ABD24280.1, XP_013069522.1, XP_009049941.1, XP_009058133.1, BAD01504.1, BAD44734.1, AFQ98380.1, AFQ98379.1, XP_013066371.1, XP_013094438.1, XP_005099315.1, BAX09271.1, XP_012939931.1, XP_012938497.1, XP_005101789.1, XP_013066278.1, XP_013066280.1, XP_013066281.1, PVD36986.1, BAU45387.1, PVD36896.1, BAK20401.1, ACE75511.1, AAX92641.1, ACE75510.1, BAT21145.1, PFX25363.1
Con sitios de unión de calcio y ZINC	EFX86822.1, XP_021959948.1, XP_021966378.1, ODM94862.1, ODM95900.1, XP_021956548.1, XP_021956549.1, AKC91349.1, KYB27092.1, XP_018323591.1, XP_018568194.1, XP_018568196.1, XP_008181842.1, XP_022167267.1, XP_022200004.1, AOV94253.1, XP_015513382.1, XP_396791.3, XP_006610535.1, PBC31793.1, XP_017793200.1, XP_012146513.1, XP_011296914.1, XP_015126097.1, XP_015592927.1, EFN70196.1, XP_011253653.1, EFN70197.1, XP_019884179.1, XP_011253654.1, XP_020289301.1, XP_020289336.1, KMQ95413.1, XP_011346866.1, EGI63652.1, XP_011057333.1, XP_012061951.1, XP_018044452.1, KYN35700.1, XP_018347393.1, KYN01696.1, XP_018396635.1, XP_018359268.1, XP_011632703.1, XP_015184471.1, KDR16731.1, XP_021924915.1, XP_023716596.1, AMH40392.1, AMH40356.1, AMH40388.1, AMH40357.1, AMH40389.1, AMH40367.1, AMH40362.1, AMH40382.1, AMH40391.1, AMH40361.1, AMH40386.1, AMH40358.1, AMH40375.1, AMH40380.1, AMH40390.1, AMH40387.1, AMH40365.1, AMH40359.1, AMH40378.1, AMH40381.1, AMH40385.1, AMH40394.1, AMH40360.1, AMH40384.1, AMH40372.1, AMH40379.1, AOV94256.1, AOV94251.1, AOV94255.1, AAF80584.1, CAD54727.1, CAD54729.1, CAD54730.1, XP_021941324.1, XP_023704930.1, ADB82658.1, AGP76434.1, AGP76435.1, AGP76436.1, AGP76437.1, AGP76413.1, AGP76416.1, AGP76415.1, AFD33365.1, BAA31326.1, BAA34050.1, BAB40696.1, BAB40697.1, ADB12483.1, AGS32241.1, AGP76407.1, AGP76398.1, AGP76399.1, AGP76397.1, AGP76428.1, AGP76429.1, AGP76431.1, AGP76432.1, AGP76433.1, AGP76402.1, AGP76403.1, AGP76417.1, AGP76405.1, AGP76430.1, AGP76401.1, AGP76426.1, AGP76422.1, XP_023337121.1, BAL60587.1, XP_018016820.1, XP_018016821.1, XP_018016806.1, XP_018016809.1, ADB85440.1, APF29556.1, APF29557.1, APF29558.1, APF29559.1, AIT11911.1, APF29561.1, AAD38027.1, EFX69372.1, KZS07093.1, KZS16514.1, KZS16516.1, EFX80605.1, KZS05722.1, XP_022336153.1, OWF45593.1, XP_021363496.1, XP_021363481.1, XP_022341787.1, XP_009049836.1, XP_009049839.1, PVD35337.1, PVD35340.1, ABD24275.1, ABD24276.1, ABD24277.1,

	PVD35381.1, ABD24274.1, ABD24278.1, ABD24279.1, PVD35383.1, PVD35384.1, PVD35385.1, XP_012935745.1, XP_005111868.2, XP_012945520.1, XP_005111870.1, XP_005111869.2, ATS91249.1, XP_013069521.1, XP_013069523.1, XP_013084281.1, XP_013084282.1, XP_013084283.1, XP_013084285.1, XP_013084287.1, XP_013069169.1, PVD36625.1, BAC67186.1, AGI61069.1, XP_009058134.1, XP_009060290.1, XP_013066370.1, XP_005099312.1, XP_012939933.1, XP_013066277.1, XP_013088631.1, XP_013066285.1, XP_013066286.1, XP_013066283.1, XP_013066284.1, BAU45388.1, PVD35551.1, XP_009065702.1, XP_011430172.1, XP_022338491.1, XP_022338490.1, XP_021363490.1, BAF38757.1, XP_021361084.1, XP_009064980.1, XP_022328778.1, XP_022329668.1, XP_011447501.1, EKC39023.1, XP_011447502.1, XP_012941225.1, XP_021345983.1, OWF56607.1, ANR02619.1, BAH22180.1, 3WC3, BAM14716.1, XP_020614867.1, XP_020893295.1, XP_020893299.1, KXJ18436.1
Con sitio de unión a zinc	EKC17289.1, XP_005101787.1
Con sitio de unión a calcio	CAD54728.1, BAB40693.1, AGP76400.1, EFX80604.1, XP_011421782.1, XP_021363498.1, XP_011423159.1, XP_013084279.1, XP_013069170.1, PVD36624.1, XP_009064976.1, BAS18736.1, XP_022336001.1, ABO26609.1

Tabla 5.- Secuencias usadas para realizar el análisis de evolutionary trace.

Numero asignado	Numero de acceso	Especie
	1KSC	<i>Nasutitermes takasagoensis</i>
1	AAF80584.1	<i>Panesthia cribrata</i>
2	ABV32557.1	<i>Teleogryllus emma</i>
3	AGP76427.1	<i>Rhynchotermes bulbinasus</i>
4	AKV16371.1	<i>Microtermes pakistanicus</i>
5	AMH40372.1	<i>Medauroidea extradentata</i>
6	AMH40373.1	<i>Peruphasma schultei</i>
7	AMH40379.1	<i>Ramulus artemis</i>
8	AMH40392.1	<i>Timema cristinae</i>
9	AMH40395.1	<i>Timema cristinae</i>
10	AOV94250.1	<i>Antipaluria urichi</i>
11	AOV94253.1	<i>Antipaluria urichi</i>
12	AOV94256.1	<i>Antipaluria urichi</i>
13	BAA31326.1	<i>Reticulitermes speratus</i>
14	CAD54726.1	<i>Mastotermes darwiniensis</i>
15	EFN70196.1	<i>Camponotus floridanus</i>
16	KDR16731.1	<i>Zootermopsis nevadensis</i>
17	KZC10616.1	<i>Dufourea</i>

		<i>novaeangliae</i>
18	ODM94862.1	<i>Orchesella cincta</i>
19	PSN31180.1	<i>Blattella germanica</i>
20	XP_001944774	<i>Acyrtosiphon pisum</i>
21	XP_002426465.1	<i>Pediculus humanus corporis</i>
22	XP_008194455	<i>Tribolium castaneum</i>
23	XP_011296914.1	<i>Fopius arisanus</i>
24	XP_017785110.1	<i>Nicrophorus vespilloides</i>
25	XP_018568194.1	<i>Anoplophora glabripennis</i>
26	XP_021956548.1	<i>Folsomia candida</i>
27	XP_022167267.1	<i>Myzus persicae</i>
28	XP_022200004.1	<i>Nilaparvata lugens</i>
29	XP_023704929.1	<i>Cryptotermes secundus</i>
30	XP_023716596	<i>Cryptotermes secundus</i>
31	XP_396791.3	<i>Apis mellifera</i>

En la tabla se muestra la especie y numero de acceso del NCBI de la que proviene la secuencias, así como el numero asignado en el análisis.

Anexo 2

Script 1 Leer y eliminar secuencias repetidas

```
####Este script permite leer un archivo de texto con
####los números de acceso y especie de las secuencias
### y permite eliminar las secuencias repetidas

###Esta función convierte una lista a un string con el nombre de la especie
def ponernom(nombre):
n=""
#nombre=nombre.split()
for i in nombre:
n+=(i.lower()+ " ")
return n[:-1]

def leer(nombre):
###Se lee el archivo a analizar
archivo = open(nombre, "r")
###Se crean los archivos donde se guardaran los nombres y números de acceso
de las secuencias
nom= open("nombress.txt","w")
num= open("numeross.txt","w")
##Se crea un lista vacía en donde se guardaran los números de acceso
no=[]
###Se lee cada una de las líneas del texto
for linea in archivo.readlines():
###Se parte la línea del texto
linea= linea.split()
#print linea.split()[0]
###Se asigna el nombre y numero
```

```

numero= linea[0]
nombre =ponernom(linea[1:])
###Si el número de acceso no ha sido leído, se escribe el número de acceso
###y nombre de la especie
###Por último se agrega el número de acceso a la lista , para evitar volver a
agregarlo
if numero not in no:
num.write(numero+"\n")
nom.write(nombre+"\n")
no.append(numero)
###Se crean los archivos en donde se guardaran la información de las secuencias
###el título y autores tienen sus propios archivos ya que alguno de estos tienen
comas
###Lo cual interfiere con el formato csv
au= open("autores.txt","w")
titu=open("titulo.txt","w")
nuevo= open("escribir2.csv","w")
####Importar el modulo ETE2
from ete2 import NCBITaxa
ncbi = NCBITaxa()
###Se crean lista auxiliares
todo=[]
t=[]
s=""
u="n/a"
###Se crean contadores para tener un mejor manejo de la información
au0=0
tit0=0
auu=0
###Se leer el archivo gp
archivo = open("sequence.gp","r")

```

```

###Se lee cada una de las líneas del archivo
for i in archivo.readlines():
    ###Si la línea tiene autor
    ###El contador autor aumenta en uno
    ###Y se asigna la variable autor
    if "AUTHORS" in i and auu ==0:
        autor=(i[12:-1]+\n")
        au0=1
        auu=1
        print autor
    ###Si se cuenta con el título del artículo al que la secuencias está relacionada
    ###El contador titulo aumenta en uno
    ###Y se asigna la variable titulo
    if "TITLE" in i and "Direct Submission" not in i:
        titulo=(i[12:-1]+\n")
        tit0=1
    ###Como el formato gp empieza con locus, esta línea lo que permite
    ###es reiniciar los contadores, y la información
    ###además de agregar la información ya obtenida a las listas creadas
    anteriormente
    if "LOCUS" in i:
        if tit0==1:
            titu.write(titulo)
        else :
            titu.write("n/a"+\n")

        if au0==1:
            au.write(autor)
        else :
            au.write("n/a"+\n")
        au0=0

```

```
tit0=0
t.append(u)
t.append(s)
todo.append(t)
auu=0
u="n/a"
t=[(i.split())[2]]
```

```
###Nos da una breve descripción de las enzimas
```

```
if "DEFINITION" in i:
```

```
    s=(i[12:-1].lower())
```

```
####Nos permite saber si la enzima pertenece al dominio Glyco hydro 9
```

```
if '/region_name="Glyco_hydro_9"' in i:
```

```
    u=(i[22:-1])
```

```
###Nos permite saber a qué organismo pertenece la secuencia, además de
agregar
```

```
###La información taxonómica con el uso de la librería ete2
```

```
if "ORGANISM" in i:
```

```
    t.append(i[12:-1])
```

```
    nomb= i[12:-1]
```

```
    name2taxid = ncbi.get_name_translator([nomb])
```

```
    valores = name2taxid.values()
```

```
    newlist=[]
```

```
    for k in valores:
```

```
        newlist.append(k[0])
```

```
    for k in newlist:
```

```
        lineage = ncbi.get_lineage(k)
```

```
#
```

```

rango=(ncbi.get_rank(lineage))
rango= (list(rango.values()))

names = ncbi.get_taxid_translator(lineage)
names=(list(names.values()))

if "order" in rango:
    t.append(str(names[rango.index("order")]))
else:
    t.append("n/a")

if "phylum" in rango:
    t.append(str(names[rango.index("phylum")]))
else:
    t.append("n/a")

if "family" in rango:
    t.append(str(names[rango.index("family")]))
else:
    t.append("n/a")

if "kingdom" in rango:
    t.append(str(names[rango.index("kingdom")]))
else:
    t.append("n/a")

```

###Nos permite saber el número de acceso de la secuencias

```

if "ACCESSION" in i :
    t.append(i[12:-1])

```

```

###Se escribe en los archivos toda la información que se guardó en las listas
for i in todo:
    if len(i)>1:
        texto=""
        for j in range(0,len(i)-1):
            texto+= i[j]+", "
        texto+=i[j+1]+\n"
        nuevo.write(texto)

###Se cierran los archivos
nuevo.close()
au.close()
titu.close()

```

Script 2 Información taxonomica

```

####Este scrip permite obtener la informacion taxnomocia
###Se importa la librería a usar
from ete3 import NCBITaxa
ncbi = NCBITaxa()
newlist=[]
###Se abre el archive con nombre p
archivo = open("p", "r")
###Se crea un archivo donde se escribira la informacion
escribir = open("infotaxa.csv", "w")

###Para cada linea del archivo, donde los datos están delimitados por comas
for linea in archivo.readlines():
    ###Se obtiene el nombre de la especie
    nomb= [(linea[:-1])]

```

```

pala=nombre
###Se obtiene el id de la especie en el NCBI
name2taxid = ncbi.get_name_translator(nomb)
###Se obtienen los datos taxonomicos
valores = name2taxid.values()

newlist=[]
###Se obtiene el ID de los caracteres taxonomicos y se agregan a una
lista
for i in valores:
    newlist.append(i[0])
for i in newlist:
    ##### Se obtienen los taxa asignados
    lineage = ncbi.get_lineage(i)
    rango=(ncbi.get_rank(lineage))
    rango= (list(rango.values()))
    names = ncbi.get_taxid_translator(lineage)
    names=(list(names.values()))
    if "genus" in rango:
        print (names[rango.index("genus")])
    else:
        print ("n/a")
    if "phylum" in rango:
        pala+=" " + (names[rango.index("phylum")])
    else:
        pala+=" " + ("n/a")
    if "family" in rango:
        pala+=" " + (names[rango.index("family")])
    else:
        pala+=" " + ("n/a")
###Se escriben la informacion obtenida en el archivo

```

```
escribir.write(">" + pala + " \n" )
```

```
else:
```

```
###Si no encontró información taxonómica solo guarda la especie
```

```
escribir.write(linea)
```

Script 3 Generar modelos de estructura terciaria

```
###Se importa la librería Splinter la cual nos permite abrir navegadores
###De internet y navegar en ellos desde Python
import splinter as sp
###Se importa la librería time, ya que en algunas partes
###del proceso se tiene que esperar un tiempo
import time

###Se selecciona el archivo .fas del cual se van a
###generar los modelos
archivo = open("paracalcio.fas", "r")

###Se crean listas las cuales van a contener
##Los nombres de las secuencias y las secuencias
texto=""
seq=[]
ide=[]

###Se lee el archivo fasta, y se agregan
###los nombres de las secuencias y las secuencias
for i in archivo:
    if ">" in i:
        print (((i[:-1]).split())[0])[1:])
        ide.append(((i[:-1]).split())[0])[1:])
        seq.append(texto)
        texto=""
    else :
        texto+=(i[:-1])
seq.append(texto)
seq.pop(0)
```

```
archivo.close()
```

```
###Para cada una de las secuencias se repetirá este proceso
```

```
for i in range(len(seq)):
```

```
    ###Se abre el explorador
```

```
    browser = sp.Browser('chrome')
```

```
    ###Se visita la pgaina de SWISS model
```

```
    browser.visit('https://swissmodel.expasy.org/interactive')
```

```
    ###Se espera 5 segundos, esto con el fin de que la pagina
```

```
    ###Cargue bien
```

```
    time.sleep(5)
```

```
    ###Se selecciona el cuadro donde se inserta la secuencia
```

```
    ###Se escribe la secuencia
```

```
    browser.fill('target', seq[i])
```

```
    ###Se selecciona el cuadro donde se inserta el título del proyecto
```

```
    ###Se escribe el título del proyecto
```

```
    browser.fill('project_title',str(i)+".-"+ide[i][:12])
```

```
    ###Se selecciona el botón validar
```

```
    button= browser.find_by_id('validateInputButton')
```

```
    button.click()
```

```
    time.sleep(2)
```

```
    ###Se presiona el botón para empezar el modelado
```

```
    button2= browser.find_by_id('buildButton')
```

```
    button2.click()
```

```
#####
```

```
###
```

```
    ##El tiempo de espera del modelo varía dependiendo del tamaño de la  
    secuencia
```

```
    time.sleep(360)
```

###Una vez pasado el tiempo se descarga el archivo con la información del modelo

```
nuevo=((str(browser.windows.current).split())[2][: -1])+"report.zip"
```

###Se cierra el navegador

```
browser.quit()
```

#####

#####

Script 4 Minimizacion de energía método ligero

```
###Se importa la librería os, para poder trabajar con archivos del
###Sistema operativo
####Se tiene que tener instalado GROMACS 5 para poder ejecutar este código
###El archivo em.mdp se tiene que encontrar en la misma ubicación de este script
import os

###Se asigna la dirección en la que se encuentran los archivos a refinar
directorio = "/home/mario/Downloads/v3/"
###Se obtienen los archivos en el directorio asignado
lista= (os.listdir(directorio))
####Este proceso se pone en un "for" ya que se va
## a repetir con todos los archivos(modelos)
for i in lista[:]:
    ### Se obtiene la topología del modelo
    os.system("gmx pdb2gmx -f "+directorio+i+" -water spc -ff gromos53a5")
    ### Se realiza el proceso de solvatación del modelo
    os.system("gmx editconf -f conf.gro -bt cubic -d 0.7 -o box.gro")
    os.system("gmx solvate -cp box.gro -cs spc216.gro -p topol.top -o solvated.gro")
    os.system("gmx grompp -f em.mdp -p topol.top -c solvated.gro -o em.tpr")
    ###Se obtiene el modelo refinado
    os.system("gmx mdrun -v -s em.tpr -deffnm em -c "+ i)
```

Script 5 Minimización de energía método fuerte

```
###Se importa la librería os, para poder trabajar con archivos del
###Sistema operativo
####Se tiene que tener instalado GROMACS 5 para poder ejecutar este código
###El archivo em.mdp se tiene que encontrar en la misma ubicación de este script
import os

###Se asigna la dirección en la que se encuentran los archivos a refinar
directorio = "/home/mario/Downloads/v3/"

###Se obtienen los archivos en el directorio asignado
lista= (os.listdir(directorio))

####Este proceso se pone en un "for" ya que se va
## a repetir con todos los archivos(modelos)
for i in lista[:]:
    ### Se obtiene la topología del modelo
    os.system("gmx pdb2gmx -f "+directorio+i+" -water spc -ff gromos53a5")
    ### Se realiza el proceso de solvatación del modelo
    os.system("gmx editconf -f conf.gro -bt cubic -d 0.7 -o box.gro")
    os.system("gmx solvate -cp box.gro -cs spc216.gro -p topol.top -o solvated.gro")
    os.system("gmx grompp -f em.mdp -p topol.top -c solvated.gro -o em.tpr")
    ####Se simula el proceso en el que se agregan moléculas de sodio y cloro
    os.system('printf "SOL" |gmx genion -s em.tpr -p topol.top -pname NA -nname
CL -neutral -conc 0.15 -o ionized.pdb')
    os.system("gmx grompp -f em.mdp -c ionized.pdb -p topol.top -o em2.tpr")
    ###Se obtiene el modelo refinado
    os.system("gmx mdrun -v -s em2.tpr -deffnm em -c "+ i)
```

Script 6 Eliminacion de moléculas

```
##Se importa la librería os, la cual nos permitirá manejar archivos
###dentro del sistema operativo
import os

###Se selecciona la carpeta que contiene los archivos
###.pdb que tienen moléculas de agua, cloro y sodio
directorio = "/home/mario/Descargas/v3_refinadose/"
lista= (os.listdir(directorio))

####Esto se repetirá en cada archivo dentro de la carpeta
for i in lista:
    ##Se abre el archivo a leer
    archivo=open(directorio+i,"r")
    ##Se crea el nuevo documento pdb que no tenga las moléculas no
deseadas
    nuevo=open("new"+i,"w")
    ###Se leen cada uno de los renglones del archivo
    for i in archivo.readlines():
        ####Si la línea no contiene alguna de las moléculas no deseadas
        ###Se escribe el renglón en el nuevo documento
        if "CL" not in i and "NA" not in i and "OW" not in i and "HW1" not in i and
"HW2" not in i:

            nuevo.write(i)
        ###Se cierra el documento
        nuevo.close()
```

Script 7 Grafico de Ramachandran

```
###Se importa la librería Selenium

###Esta librería nos permite usar navegadores

###Web desde Python además de cargar archivos

###a los mismos

from selenium import webdriver

from selenium.webdriver.common.keys import Keys

###Se importa la librería os, para poder trabajar con archivos del

###Sistema operativo

import os

###Se importa la librería time, ya que en algunas partes

###del proceso se tiene que esperar un tiempo

import time

###Se especifica el directorio del cual se van a cargar los archivos pdb

###a analizar

directorio = "/home/mario/Descargas/v3levebien/"

###Se obtiene una lista de todos los archivos

lista= (os.listdir(directorio))

###Se crea un archivo donde se almacenaran los resultados

ta=open("tabla.csv","w")

###esto se repite para todos los archivos

for i in lista:
```

```

s=(i[:2])

###Se abre el navegador

driver = webdriver.Firefox()

###Se dirige a la pagina de rampage, esta permite

###hacer el análisis del grafico de Ramachandran

driver.get('http://mordred.bioc.cam.ac.uk/~rapper/rampage.php')

##Esperamos a que la pagina cargue

time.sleep(2)

###Seleccionamos y cargamos el archivo para el análisis

element=driver.find_element_by_name("pdbfile")

element.send_keys(directorio+i)

driver.find_element_by_name("SUBMIT").click()

###Se carga la página de los resultados

page = driver.page_source

p2 = (page.split("\n"))

#print (p2)

p=""

n=""

###Con análisis de texto se obtiene la información del número y

###Porcentaje de aminoácidos en las distintas zonas del análisis

for i in p2:

    if "%" in i and ":" in i:

        #print (i)

```

```
i=[i.index(":")+1:i.index("(")]  
  
#print (i)  
  
nu=(int(i[i.index(":")+1:i.index("(")]))  
  
po=(float(i[i.index("(")+1:i.index(")")] - 1))  
  
n+= str(nu)+", "  
  
p+= str(po)+", "  
  
###Se escriben los resultados en el archivo  
  
ta.write(s+", "+n+p[:-2]+"\\n")  
  
####Se cierra el navegador  
  
driver.close()  
  
###Se cierra el archivo  
  
ta.close()
```

Script 8 Analisis estructural RMSD

```
###Se tiene que tener instalado Muscle
###Se importa la librería Bio3D
library(bio3d)
###Se importa la librería Ape
library(ape)
###Se asignan las direcciones de los archivos .pdb a analizar
files<-c('/home/mario/Downloads/v3todos/v3fuerte/30AOV942531_c_2018-10-
13.pdb','/home/mario/Downloads/v3todos/v3fuerte/35XP_021956548_2018-10-
13.pdb')

###Se alinean los modelos
pdbs <- pdbaln(files)
###Se asignan los ids
pdbs$id <- substr(basename(pdbs$id),1,6)
### Calcula la identidad de los modelos
seqidentity(pdbs)
## Calcula el valor de RMSD
rd<-rmsd(pdbs, fit=TRUE)

####SE genera el dendograma
hc <- hclust(as.dist(rd))
hc.rd <- hclust(as.dist(rd))
hclustplot(hc.rd, k=3, labels=pdbs$id, cex=0.5, fillbox=FALSE)
###Se convierte el dendograma a un archivo .tre
my_tree <- as.phylo(hc)
write.tree(phy=my_tree, file="fuerte.nex")
```

Script 9 RaptorX ID's

```
###Se importa la librería os la cual nos permite
###Administrar archivos del sistema operativo
import os
###Se especifica la ruta en donde están los archivos con
##Los nombres y códigos de las secuencias
ruta ="C:\\Users\\Mario\\Documents\\Codigo5"
###Se crea un archivo .CSV donde se guardaran los números de acceso y
códigos de RaptorX
doc=open("sec.csv","w")
###Se obtienen los archivos de la ruta especificada
dirs = os.listdir(ruta)
###Para cada archivo se va a hacer :
for i in dirs :
    ###Se abre y lee el archivo
    archivo = open (ruta+ "\\"+i,"r")
    for j in archivo.readlines():
        if ">" in j:
            j= j[1:-1]
            print j
            ###Se lee el número de acceso y el código de RaptorX relacionado el
análisis
            ###de dicha secuencia
            if "XP" not in j and "P" != j[0]:
                numero=j.split("_")[0]
                idd=j.split("seqID: ")[1]
            elif "P" == j[0]:
                numero=j
                idd=j.split("seqID: ")[1]
            else :
```

```
numero=j.split("_")[0]+"_"+j.split("_")[1]
idd=j.split("seqID: ")[1]
###Se escribe la información en el archivo .csv
doc.write(numero+","+idd+"\n")
###Cerramos el archivo leído
archivo.close()
###Cerramos el archivo .csv
doc.close()
```

Script 10 RaptorX Informacion de grupos prostéticos

```
###Se importa la librería os la cual nos permite
###Administrar archivos del sistema operativo
import os
###Se especifica la ruta en donde están los archivos con
##Los nombres y códigos de las secuencias
ruta ="C:\\Users\\Mario\\Documents\\calcio4"
###Se crea un archivo .CSV donde se guardaran los códigos de RaptorX y los
resultados de
###La predicción de los sitios de unión
doc=open("sec.csv","w")
###Se obtienen los archivos de la ruta especificada
dirs = os.listdir(ruta)
###Para cada archivo
for i in dirs :
    ###Se abre y lee dicho archivo
    archivo = open (ruta+ "\\"+i,"r")
    texto = archivo.read()

    ###estas variables sirven para control
    ##y verificar que secuencias tienen sitios de unión
    ###a calcio o zinc
    ca=""
    zn=""

    ###Si el archivo contiene las palabras CA o ZN
    ###se registra que tienen dichos sitios de unión
    if "CA" in texto:
```

```

        ca="ca"
    else :
        ca="sin"
    if "ZN" in texto:
        zn="zn"
    else :
        zn="sin"

    ##Se escribe en el archivo .cvs que secuencia con el número de acero de
raptorX tiene los sitios de unión
    doc.write(i[:6]+","+ ca + ","+ zn +"\n")
    ###Cerramos el archivo leido
    archivo.close()
    ###Cerramos el archivo ,CSV
doc.close()

```

Script 11 Filogenia con información de los sitios de unión a iones metálicos

```

#####Se importa la libreria ete2, la cual permitira
###a
from ete2 import Tree, TreeStyle, NodeStyle
###Se abre el archivo de los sitios de union a calcio
###tiene que estar ordenado de mismo modo que los nodos en el analisis
filogenetico
archivo=open("CAL.txt","r")
li= archivo.readlines()

def get_example_tree():
    ###Se lee la topologia del arbol la cual esta en un archivo .txt

```

```
t = Tree((open("arbol.txt")).read())
##Es el estilo de nodo que se va a dibujar cuando hay sitios de union a
###Calcio y zinc
style1 = NodeStyle()
style1["fgcolor"] = "#08C9DF"
style1["shape"] = "circle"
style1["size"] = 60
```

```
##Es el estilo de nodo que se va a dibujar cuando hay sitios de union a
###zinc
style2 = NodeStyle()
style2["fgcolor"] = "#04847A"
style2["shape"] = "circle"
style2["size"] = 60
```

```
##Es el estilo de nodo que se va a dibujar cuando hay sitios de union a
###Calcio
style4 = NodeStyle()
style4["fgcolor"] = "#0849DF"
style4["shape"] = "circle"
style4["size"] = 60
```

```
##Es el estilo de nodo que se va a dibujar cuando no hay sitios de union
style3 = NodeStyle()
style3["fgcolor"] = "#DF2C08"
style3["shape"] = "circle"
style3["size"] = 60
```

```
###Esto se hara para cada uno de los nodos
for l in t.iter_leaves():
```

```

####Se obtiene los sitios de union con los que cuneta cada nodo
cosa = li[int(str(l)[3:])-1][:-1]
####Se colores el nodo segun el tipo de sitios de union
if cosa=="D" :
    l.img_style = style1
if cosa=="C" :
    l.img_style = style4
if cosa=="Z":
    l.img_style = style2
if cosa=="S":
    l.img_style = style3
####Se crea el arbol a dibujar
ts = TreeStyle()
ts.mode = "c"
####Se dibuja el arbol el forma circular
ts.root_opening_factor = 1
####Se dibujan los nodos de colores
ts.show_leaf_name = False
return t, ts
if __name__ == "__main__":
    t, ts = get_example_tree()
####Se dibuja y muestra el arbol
t.show(tree_style=ts)
t.render("node_style.png", w=400, tree_style=ts)

```

Script 12 Secuencias con mutaciones específicas

```
####Se abre el archivo a leer
archivo=open("alineadotodo.fas","r")
###Se crea el nuevo archivo en el que se depositaran las secuencias con
diferencias
nuevo=open("cons22.csv","w")
l=[]
t=""
###Se lee cada uno de los renglones
for i in archivo.readlines():
    ###Si es un nombre de secuencias:
    if ">" in i :
        ####Si es la secuencias y tienen una "S" serina en el lugar
        ### 2895 del alineamiento se guardara el numero de acceso y especie a la
que pertenece
        if len(t)>1 and t[2895]=="S":
            s=l[len(l)-1].split("_")
            if "XP" in s[0]:
                s1= s[0]+"_" +s[1]
            else :
                s1= s[0]
            nuevo.write(s1+" "+s[-2]+" "+s[-1]+"\n")
        l.append(i[1:-1])
        t=""
    else :
        t+=i[:-1]
###Se cierra el archivo donde se escribe
nuevo.close()
```


021941324.1 KAHVSNQELYGQVGGQDGDHAYWGRPEDMTMSRPAYKIDTSHPGSDLAGETAAALAAASIVFQSDADSGYASTL LTHAKQLDFDANNYRKYSDSITDAQS 200
021956548.1 KAHPSDFVFFYGGTGGDGGDGHYWGPEW-GARNSYKIQQQPQSGELAGEAASAFSGYLAFKDNPVTYASTL LNHAHTLLYDFANQFRGTY-DAIP-AGG 196
021956549.1 KAHPTDFVFFYGGTGGDGGDGHYWGPEW-GARNSYKIQQQPQSGELAGEAASAFSGYLAFKDNPVTYASTL LNHAHTLLYDFANQFRGTY-DAIP-AGG 196
021959948.1 KCHTGETELFVQVGNATA DHNSVDI PERW--PRPAYKITAERPSSDIAGETSALFSGATLLFOTVNDTYASIL LDHAKSLYTFANQYRKY-DSVPEATP 196
02196378.1 KHNPSPNVLYVQVNGEIDHNYWGRPEWNGRPRVLAQNSKPSBEVAGEAASAAAMAAASIAFRQNTS YADTL LQHARELYTFATYRGRHY-ESFPVEKE 199
021967587.1 KAHVAPNELYGGTGGDGGDGHYWGPEW-GPRQSWKITQQQPSGELAGEAASAFSGYLAFKDNPVTYASTL LNHAHTLLYDFANQFRGTY-NAIPARD- 197
022167267.1 KCHVSDYEFYGGQVDFSLDGGFWGRPEDMNMSRPSYKIDKNRPPGSDLAGETAAALAAASIVFVKIDSNFASCLSHAKQLYLFATQYRGLY-DAIKGAAQ 199
022167288.1 KCHVSDYEFYGGQVDFSLDGGFWGRPEDMNMSRPSYKIDKNRPPGSDLAGETAAALAAASIVFVKIDSNFASCLSHAKQLYLFATQYRGLY-DAIKGAAQ 199
022200041.1 KCHVSPNEFYGGQVDFSLDGGFWGRPEDMNMSRPSYKIDAEHPGSDLAGETAAALAAASIVFVKIDSNFASCLSHAKQLYLFATQYRGLY-DAIKGAAQ 199
022328781.1 KCRNDSQYVQVNGEIDHAYWGRPEW--RPAYYLDSTKTQSDVGTATAALTAGSMALFKSDPSYSRLLSESAESLYVFAKARQGTYNADI SDGAK 197
022329688.1 KCRNDSQYVQVNGEIDHAYWGRPEW--RPAYYLDSTKTQSDVGTATAALTAGSMALFKSDPSYSRLLSESAESLYVFAKARQGTYNADI SDGAK 197
022336001.1 KCHQTOQYVQVNGEKDHRVWGRPEWFKLRPACKTRNDT-SDVAGEAALASGAVFVKDKDPPYSDDLLEAAKSLYDFAKRQGTYGVTVQVNG 196
022336153.1 KCHQTHNELYVQVNGK-SHVSVDPEI-TKOPAYKVDNTPKPSDVAAGEYAAAGYLAFLHAKHLEHAKIDNDFAYTKYKGSVDSVPEASE 198
022338401.1 KCRPQSOEYVQVNGGDHAFWGRPEW--RPAYKCTASNGGSDVEGITAAALTAGSMALFKASDAAYSQRLSSAKSLYEFANAHGVIYNGKISDATS 197
022338401.1 KCRPQSOEYVQVNGGDHAFWGRPEW--RPAYKCTASNGGSDVEGITAAALTAGSMALFKASDAAYSQRLSSAKSLYEFANAHGVIYNGKISDATS 197
022338401.1 KCRPQTOEYVQVNGEKDHRVWGRPEWDMFRLRPACKTRNDT-SDVAGEAALASGAVFVKDKDPPYSDDLLEAAKSLYDFAKRQGTYGVTVQVNG 199
022341787.1 KCHQTHNELYVQVNGGDHAFWGRPEW-MSRPAYKIDANRPPGSDIAEYAAAFVSHLVFERDEAFADKLLTHSKQLDYFAVHHKARYTDSVSEAG 199
022337121.1 KCHTNTNELWQVGGDNDHDSQW-RVEQMTIHRPSFKIDSNQPPGSDLAGETAAALAAASIAFKDVPDPSYSTLLLEAKSLFNADYRGRAYDSVSAAGS 199
023704929.1 KAHVSNVLYGQVGGQDGHAYWGRPEDMTMSRPAYKIDTSHPGSDLAETAAALAAASIVFKSSDSSYANTL LTHAKQLDFDANNYRKYSDSITDAQS 200
023704930.1 KAHVSNVLYGQVGGQDGHAYWGRPEDMTMSRPAYKIDTSHPGSDLAETAAALAAASIVFKSSDSSYANTL LTHAKQLDFDANNYRKYSDSITDAQS 200
023716596.1 KAHVSNVLYGQVGGQDGHAYWGRPEDMTMSRPAYKIDTSHPGSDLAETAAALAAASIVFKSSDSSYANTL LTHAKQLDFDANNYRKYSDSITDAQS 200
023718476.1 KCHVSPNEFYGGQVDFSLDGGFWGRPEW--VDSHICWRPEWIKIDTSHPGSDLAGETAAALAAASIVFVRNDSYADIL LTHAKQLDFDANNYRKYSDSITDAQS 198
XP_3967913.1 KCHVSEYVYGGQVDFSLDGGFWGRPEW--VDSHICWRPEWIKIDTSHPGSDLAGETAAALAAASIVFVRNDSYADIL LTHAKQLDFDANNYRKYSDSITDAQS 199
AA22086.1 KAHYAPNELYGGTGGDGGDGHYWGPEWPAEVMTPARPSHKIDSPCGSDVAGEAALAAASIVLKGDDPAAAL LTHAKQLYTFADYRGRAYDSVCAVTA 200

AAD38027.1 FYNSSVYGD ELVWAAWLYRATNDNT LDRAGHVSFNLGTPS-QFGWDDKAGVQALLVLLDGSSEYALNQFLN-FVRNQAPTPEGLVFLDAWGSNR 297
AAF80581.1 YSSVYGED ELVWAAWLYRATNDNT LTKAQYVDFGLYDQ--YFSWDDKISGVEVLLAEITKDDGHIRITIQYCDNMVRRGQTRPKGLVYIDWGS LR 298
AAF80581.1 YSSVYGED ELVWAAWLYRATNDNT LTKAQYVDFGLYDQ--YFSWDDKISGVEVLLAEITKDDGHIRITIQYCDNMVRRGQTRPKGLVYIDWGS LR 298
AAK12391.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 297
AAO61672.1 FYNSSVYGD ELVWAAWLYRATNDNT LDRAGHVSFNLGTPS-QFGWDDKAGVQALLVLLDGSSEYALNQFLN-FVRNQAPTPEGLVFLDAWGSNR 300
AAU20852.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKISGVOVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 297
AA92641.1 FYNSSVYGD ELVWAAWLYRATNDNT LDRAYEFG--TTTAW--AYDWNEKIVGYQLLTSAGOTDFLRVENFLRNWFPGGSVTLPLGLAWLQWGNR 295
ABD24274.1 YNSANVYD ELVWAAWLYRATNDNT LEEALKHVD--ASPDW--GMSWDDVFIIGNOVLLYELTKEAKYKAVEGTFREWFPGGTVPKGLAYRLQWGLR 294
ABD24280.1 YNSANVYD ELVWAAWLYRATNDNT LEEALKHVD--ASPDW--GMSWDDVFIIGNOVLLYELTKEAKYKAVEGTFREWFPGGTVPKGLAYRLQWGLR 294
ABD24281.1 YNSANVYD ELVWAAWLYRATNDNT LEEALKHVD--ASPDW--GMSWDDVFIIGNOVLLYELTKEAKYKAVEGTFREWFPGGTVPKGLAYRLQWGLR 294
ABO26609.1 AYASTGYND ELVWAAWLYRATNDNT LEDAKTYHS--ASAPW--AYSWDDTIVGQVLLDMTRDVTYRQVRFLLTSWPKGS-TPCGAARWPKWGS LR 292
ABV2557.1 FYNSSVYGD ELVWAAWLYRATNDNT LNKAATQYVNGQYSIQYKNSPLSDWDDKTSAGASALLKTDGQDQYKAVQSCDGFYVNNQKTPKGLIWSWGS LR 300
ACE75510.1 FYNSSVYGD ELVWAAWLYRATNDNT LDRAYEFG--TTTAW--AYDWNEKIVGYQLLTSAGOTDFLRVENFLRNWFPGGSVTLPLGLAWLQWGNR 295
ACE75511.1 FYNSSVYGD ELVWAAWLYRATNDNT LDRAYEFG--STVAV--AYDWNEKIVGYQLLTSAGOTDFLRVENFLRNWFPGGSVTLPLGLAWLQWGNR 295
ADB12483.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 297
ADB82658.1 YTSYDYRD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 297
ADB85410.1 FYNSSVYGD ELVWAAWLYRATNDNT LSADEAGHYDFGLDVL--NVDDDKTAGVYLLMVEIDGGSQYALOGYVNDLL--SYOTPGMVIYDWSGS LR 292
ADB85411.1 FYNSSVYGD ELVWAAWLYRATNDNT YNQAEGYVSNFMNDQYVYVDDKTAGVYLLMVEIDGGSQYALOGYVNDLL--SMRTSQNTAGLFIYDWSGS LR 297
ADB85421.1 FYNSSVYGD ELVWAAWLYRATNDNT MTAAKGFIIEF--GFMDKXVYVDDKTAGVYLLMVEIDGGSQYALOGYVNDLL--TYTPEGLIFDVGWGLR 294
AFD3385.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AFQ98379.1 FYNSSVYGD ELVWAAWLYRATNDNT LADAKG-YN--AW-AW--ALGWDDKIIACOLLLYEATKDTAYKEVEGFFKQWLPGGS-TPCGAARWPKWGS LR 290
AG61069.1 FYNSSVYGD ELVWAAWLYRATNDNT LADAKG-YN--AW-AW--ALGWDDKIIACOLLLYEATKDTAYKEVEGFFKQWLPGGS-TPCGAARWPKWGS LR 290
AGP76397.1 FYNSSVYGD ELVWAAWLYRATNDNT LNDAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76398.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76399.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76400.1 FYNSSVYGD ELVWAAWLYRATNDNT LTTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76401.1 FYNSSVYGD ELVWAAWLYRATNDNT LTTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76402.1 FYNSSVYGD ELVWAAWLYRATNDNT LTTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76403.1 FYNSSVYGD ELVWAAWLYRATNDNT LTTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76404.1 FYNSSVYGD ELVWAAWLYRATNDNT LTTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76405.1 FYNSSVYGD ELVWAAWLYRATNDNT LTTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76406.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76407.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76408.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76409.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76410.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76411.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76412.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76413.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76414.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76415.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76416.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76417.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76418.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76419.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76420.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76421.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76422.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76423.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76424.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76425.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76426.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76427.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76428.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76429.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76430.1 FYNSSVYGD ELVWAAWLYRATNDNT LTTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76431.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76432.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76433.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76434.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76435.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76436.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGP76437.1 FYNSSVYGD ELVWAAWLYRATNDNT LNTAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AGS32411.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AKC01349.1 FYNSSVYGD ELVWAAWLYRATNDNT LDRAGHVSFNLGTPS-QFGWDDKAGVQALLVLLDGSSEYALNQFLN-FVRNQAPTPEGLVFLDAWGSNR 297
AKV16371.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40356.1 FYNSSVYGD ELVWAAWLYRATNDNT LDIYDQYWS--TVMYND--GFWDKQFGGAVYLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40357.1 FYNSSVYGD ELVWAAWLYRATNDNT LNIYDQYWS--SVMYND--GFWDKQFGGAVYLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40358.1 FYNSSVYGD ELVWAAWLYRATNDNT LGYVDELDW--SVMYND--GFWDKQFGGAVYLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40359.1 FYNSSVYGD ELVWAAWLYRATNDNT LQYIADVYV--PIQYI--GFNWDKIGGADVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40360.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40361.1 FYNSSVYGD ELVWAAWLYRATNDNT MNTVDEIYV--TIQYD--GFWDKQFGGAVYLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40362.1 FYNSSVYGD ELVWAAWLYRATNDNT MNTVDEIYV--TIQYD--GFWDKQFGGAVYLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40363.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40364.1 FYNSSVYGD ELVWAAWLYRATNDNT LKYVDAWLD--KVMYND--GFWDKQFGGAVYLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40365.1 FYNSSVYGD ELVWAAWLYRATNDNT LTYVEEYVYV--PIQYI--GFNWDKIGGADVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40366.1 FYNSSVYGD ELVWAAWLYRATNDNT LTYVDEYVYV--PIQYI--GFNWDKIGGADVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40367.1 FYNSSVYGD ELVWAAWLYRATNDNT LTYVDEYVYV--PIQYI--GFNWDKIGGADVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40368.1 FYNSSVYGD ELVWAAWLYRATNDNT LTYVDEYVYV--PIQYI--GFNWDKIGGADVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40369.1 FYNSSVYGD ELVWAAWLYRATNDNT LTYVDEYVYV--PIQYI--GFNWDKIGGADVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40370.1 FYNSSVYGD ELVWAAWLYRATNDNT LDIYDQYWS--TVMYND--GFWDKQFGGAVYLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40371.1 FYNSSVYGD ELVWAAWLYRATNDNT LTYVEEYVYV--PIQYI--GFNWDKIGGADVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 300
AMH40372.1 FYNSSVYGD ELVWAAWLYRATNDNT LTKAESLYNEFGLQWNGGAFNNDKSVSGVQVLLAKLTKQAYKTVQGVYDYLINQQKTPKGLLIDWGT LR 299

XP_002426465.1 YVES TDYAD E L T W AAWLYKASNDL T L DDAEHHYMKFHLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFIV - DQKTPKGLLYIDKFGTLC 296
XP_003002778.1 YVES TDYDG E L A W AVWLFKATNDT I L L EDAEHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_003006076.1 YVES TDYDG E L A W AVWLFKATNDT M L L EEEAEHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_005099312.1 FYSSSGDRD E L C E SMWLYRATGNRO L QDANSTW -- AW - AW -- ALGWDDKVAQOELLYEETHNADYQOY - PGSGT PCGLAWRDMWGANR 273
XP_005099315.1 FYSSSGYKD E L C E AVWLKRAATGNS L LNDAKN - YT - NW - AW -- AEGWDDKMLACQLLMYEETKSEYKAVVGF - PGSGT PCGLAWRDMWGANR 271
XP_005101787.1 FYKDGHYE -- GV - PW -- ALSWDDKVAQOELLYEETHNADYQOY - NNGYKDVTFNLQOFLPG -- LPGLAARDDWGSNR 289
XP_005101789.1 FYRSDGKY E M C L AVWLKRAATKDSN L LNDAKSFY - EAGV - PW -- ALSWDDKVAQOELLYEETHNADYQOY - NNGYKDVTFNLQOFLPG -- LPGLAARDDWGSNR 270
XP_005111868.2 FYRSDNID E L A W SLWLKATRNDSK L MVEAKHFD -- PAPW -- GMSWDEKTIQOVLNQLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 295
XP_005111869.2 FYRSDNID E L A W SLWLKATRNDSK L LDEAKLFD -- PAPW -- GMSWDEKTIQOVLNQLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 295
XP_005111870.1 FYRSDNID E L A W SLWLHEATNDPM L LNKAKTLFD -- DP - AW -- GMSWDEKTIQOVLNQLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 294
XP_006101535.1 YVES TDYDG E L A W AVWLFKATNDT M L L EEEAEHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_008181842.1 YVES TDYDG E L T W AAWLYKATKESQ L LDDAEHYMKFHLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_008194455.1 YVES TDYDG E L T W AAWLYKATKESQ L LDDAEHYMKFHLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_009049791.1 FYSSMGFKD E L A W GAWLYKATKEAK L L T SAEKYHQ -- GK - SW -- GLSWDDKVAQOELLYEETHNADYQOY - IYKDI EATFDQWMPGGIIPKGLAARLWGSNR 294
XP_009049836.1 YKSNITD E L N S W AAWLYKATSDKX L LADAEAAHI -- GE - SW -- GFSWDEKTIQOVLNQLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 294
XP_009049839.1 YKSNITD E L G W AAWLYKATSDKX L IAEAKHHE -- GA - AW -- GFSWDEKTIQOVLNQLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 279
XP_009049940.1 FYSSQFQD E L C W AAWLYRATGKFT L LDTAEQYK -- HS - HS -- AFNWDKTDATSLMLFMLTKKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 294
XP_009049941.1 FYSSQNMFD E L C W SLWMYAITNETK L L T EAEKWFQ -- GP - SW -- GMSWDDTQAQOELLYEETHNADYQOY - IYKDI EATFDQWMPGGIIPKGLAARLWGSNR 294
XP_009058133.1 YKSNITD E L C W SLWMYRVTKEAK L L T EAEKWFQ -- GP - AW -- GMSWDDTQAQOELLYEETHNADYQOY - IYKDI EATFDQWMPGGIIPKGLAARLWGSNR 294
XP_009058134.1 FYSSGYKD E L C E AAWLYRATKDKQ L LNDAKG - FT - AW - AW -- ALSWDDKVAQOELLYEETHNADYQOY - IYKDI EATFDQWMPGGIIPKGLAARLWGSNR 290
XP_009060290.1 FYSSSYKD E L C J AAWLYRATNDKQ L L NDAISFNH -- AW - AW -- ALSWDDKVAQOELLYEETHNADYQOY - IYKDI EATFDQWMPGGIIPKGLAARLWGSNR 292
XP_009064976.1 FYKSNYTD E L C W SLWMYAITNETK L L T EAEKWFQ -- GP - SW -- GMSWDDTQAQOELLYEETHNADYQOY - IYKDI EATFDQWMPGGIIPKGLAARLWGSNR 294
XP_009064980.1 YVESATGNT E L C V AAWLYKATGNT F L L T EAKLYHS -- SATPN -- RLWDNDSLPLNOVLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 292
XP_009065702.1 FYRSGYKD E L C E AAWLYKATKDHQ L LNDAKS - FN - AW - AW -- ALSWDDKVAQOELLYEETHNADYQOY - IYKDI EATFDQWMPGGIIPKGLAARLWGSNR 290
XP_011057333.1 YVES TDYDG E L A W AAWLYKATGNT F L L T EAKLYHS -- SATPN -- RLWDNDSLPLNOVLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 292
XP_011164235.1 YVES TDYDG E L A W AAWLYKATGNT F L L T EAKLYHS -- SATPN -- RLWDNDSLPLNOVLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 292
XP_011253653.1 YNNTDYDG E L A W AAWLYKATGNT F L L T EAKLYHS -- SATPN -- RLWDNDSLPLNOVLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 292
XP_011253654.1 YNNTDYDG E L A W AAWLYKATGNT F L L T EAKLYHS -- SATPN -- RLWDNDSLPLNOVLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 292
XP_011253655.1 YNNTDYDG E L A W AAWLYKATGNT F L L T EAKLYHS -- SATPN -- RLWDNDSLPLNOVLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 292
XP_011296914.1 FYESTDYDG E L A W AAWLYKATNDAK L L EDAEHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSSK - ROQKTPKGLLYIDKFGTLC 297
XP_011296916.1 FYESTDYDG E L A W AAWLYKATNDAK L L EDAEHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSSK - ROQKTPKGLLYIDKFGTLC 297
XP_011346866.1 FYESTDYDG E L A W AAWLYKATNDAK L L EDAEHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSSK - ROQKTPKGLLYIDKFGTLC 297
XP_011421721.1 FYRSDYND E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011423159.1 YRSDYND E L T W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011430172.1 YRSDYND E L T W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011447501.1 FYSSSEYRD E M C V AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011447502.1 FYSSSEYRD E M C V AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011498432.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011632703.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011700378.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_011867251.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012061951.1 YVES TDYDG E L T W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012146513.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012224814.1 YVES TDYDG E L V W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012248155.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012273697.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012526854.1 YVES TDYDG E L A W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012536745.1 FYTNSYD E L C W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012536847.1 FYTNSYD E L C W AAWLYKATNETK L L YDAIHHYQHFLRPN -- EFFYNNKVVAGVQLLQAQGTORPEYIAAQAFCDFSVY - ROQKTPKGLLYIDKFGTLC 297
XP_012539031.1 FYSSGSD E L C L AAWLYKATGDSH L LNDAKS FHN -- GP - PW -- AQS WDDKTIQOVLNQLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 295
XP_012539031.1 FYSSGSD E L C L AAWLYKATGDSH L LNDAKS - YL - NK - PW -- AQS WDDKTIQOVLNQLLYKLTGKDYKAI - IGTFSWFGGDIIPKGLAARLWGSNR 295
XP_012539933.1 YH - DDYK E M C L AAWLYRATKDSK L LNDAKSFYE -- GV - PW -- ALSWAEKKAQO -- ENV -- LF -- 294

XP_018568194.1 HAANI AFLCL IASDP ---SEKYTTFAKEQIDYI LGATGRSYVVG YGFENFTQOPYHSASSCPDKPLACGWDEYHSKQPNPQILYGLVSGPQQNDHYEDRR 394
XP_018568196.1 HAANI AFLCL IASDP ---SEKYTTFAKEQIDYI LGATGRSYVVG YGFENFTQOPYHSASSCPDKPLACGWDEYHSKQPNPQILYGLVSGPQQNDHYEDRR 394
XP_019884179.1 HAANVAF ICLVVAAD-GLHQ--YYKFAKEQIDYI LKAGRSYVVG YGINAPQQRHLIASSCPHLKWCWGKQYRKNEPNQILYGLVSGPNENDEFLDIR 393
XP_020289295.1 LAANVAF LCLVVAAD--DLKRYIYFAKQIINYML GSGRSYVVG YDATSPKQPRHLIASSCPKVPVCEQEMIQQNVI NPQTLYGLVSGPNVEDDLTDYR 388
XP_020289297.1 LAANAAF LCL IALSM--NLHKYYDFAKEQIDYI LKTERDYVVG LTKSPTQIPHKASSCPKWPANCCHTKMMSNEPNHEILYGLVSGPNVGHDFVDIR 392
XP_020289301.1 LAANAAF LCL IAA--SVEHDEPYFAKEQIDYI LKTRSYVVG YFSSSSPTIHSKASSCPRMRNCI REI TSNESNYQVLYGLVSGPNFKNDVYDDR 391
XP_020289306.1 FAANAAF VCLAAASY--QSQFYNFAEQIDYI LKSGRSYVVG YGFENFTQOPYHSASSCPDKPLACGWDEYHSKQPNPQILYGLVSGPQNDHYEDRR 394
XP_020614867.1 YSANTA FLALVAADGGVXSA-FREFAKKQIHYML GDSGRSFVVG YFGTNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 397
XP_020893295.1 YAANTA FLALVAADNGMKND-YRTFAKKQIHYML GDSGRSFVVG YFGKSPQPRPHRRSSCPKXPCWWDNRNPGPNPQILYGLVSGPQNDHYEDRR 397
XP_020893299.1 YSANTA FLALVAADNGLKDII-YRNFARKQIHYML GDSGRSFVVG YFGANPPKQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 397
XP_021345983.1 YSANSFA IALMAAKTG IADTSYRDYARQGLHYML GNEIRSYVVG YFGNFPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 396
XP_021361084.1 FAANAAF VCLAAEDGD---YKTWALSQIDYAL GNNHFSYVVG YFGSNYPKQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 396
XP_021363481.1 HAANAAF VCLAAEHGINSI-YRAWAKSQIHYAL GDTGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_021363490.1 YAANAAF VCLAAEDGND---YKTFALSQIDYI LKGRMSFVVG YFGSNYPKQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 394
XP_021363495.1 YASNMA FMAL IAAEDG IHSI-YRAWAKSQIHYAL GDTGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 394
XP_021363496.1 YASNMA FMAL IAAEDG IHSI-YRAWAKSQIHYAL GDTGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 394
XP_021363497.1 YASNMA FMAL IAAEDG IHSI-YRAWAKSQIHYAL GDTGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_021363498.1 YASNMA FMAL IAAEDG IHSI-YRAWAKSQIHYAL GDTGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_021928495.1 HAANVAF VCLQAAD---PRRYREFARQIDYML GDSGRSFVVG YFGYVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_021928577.1 H IANA AAF ICFQLEDLQASGAFYRQFQTKQIDYI LKDSGRSFVVG YFGKSNPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 394
XP_021941324.1 MAANAAF ICAQADLGIINAASYRAFAKKQIDYI LKDTGRSFVVG YFGTNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 400
XP_021965548.1 HSSNVAF ICLQAADVGIKDV-YRAFAKQIINYAL GDSGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 392
XP_021965549.1 HSSNVAF ICLQAADVGIKDV-YRAFAKQIINYAL GDSGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 394
XP_021965948.1 AASNSL FVCGAQLGLNEK-YRNFQTKQIDYML GDMWGRSFVVG YFGTNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_021966378.1 HANSVYFLCQAANLQV--YDFGKQOIDYML GNSGRSFVVG YFGSNYPKQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 397
XP_021967587.1 HAANVAFACITFGEVGA NEK-YRTFAKKQIHYML GNTGRSFVVG YFGKSNPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_022167267.1 HAANI VFLCQAAD---TOYREFAKQIINYI LKGGQSYVVG YFGKSNPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_022167268.1 HAANI VFLCQAAD---TOYREFAKQIINYI LKGGQSYVVG YFGKSNPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_022200004.1 HAANI AFVCMQAAD---EKYRDFAKKQIDYI LKDSGRSFVVG YFGYVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 392
XP_022328778.1 YSANA AFIALLAADVGA---YKEWAMSQIHYML GNAFNSYVVG YGSSYPLKPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 391
XP_022329668.1 YSANA AFIALLAADVGA---YKEWAMSQIHYML GNAFNSYVVG YGSSYPLKPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 391
XP_022336001.1 YAANAAF VCLMAVDE---KEDI IAWLKOIINYML GNKEMSYVVG YGKNYPMQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 398
XP_022336153.1 YASNMA FMAL IAAEDG LHSI-YRRAWAKSQIHYAL GDSGRSFVVG YFGVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 392

XP_022338490.1 YAANAAF VCLAAEDGD---QFKSWALSQIINYI LKNNHMSYVVG YFGNNYRQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 388
XP_022338491.1 YAANAAF VCLMAVDE---KEDI IAWLKOIINYML GNKEMSYVVG YGKNYPMQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 385
XP_022341787.1 YTSNMA FMAL IAAEDG VGHQ-YROWARSQIHYAL GDTGRSFVVG YFGTNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_022337121.1 HASNA AAFACLEAAK IISIDTK-FQGFAMDQIHYAL GDTGRSFVVG YFGTNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 391
XP_023704929.1 MAANAAF ICA SAADLGIINAASYRAFAKKQIDYI LKDSGRSFVVG YFGKSNPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 400
XP_023704930.1 MYANA AYLFCQAADLGIISVDTRYQYFAKQIHYML CDAGRYSYVVG YFGNNYRQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 399
XP_023716596.1 HAANVAF VCLQAAD---PTKYREFARQIDYML GDSGRSFVVG YFGSNYPKQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 393
XP_023718476.1 K IAAAAY ICFQLEDLQASGAFYRQFQTKQIDYI LKDSGRSFVVG YFGYVNPPTQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 398
XP_396791.3 HAANVAF VCLAAEDG---DQK-YREFAEQIINYML GGGRSYVVG YGWRNPPKQPHRRSSCPDRPAPCTHDALNAPGANPQVLYGLVSGPQNDHYEDRR 394
AAA23086.1 YAANTS FVALVYSDATRKARYHDFGVRQIINYAL GDNPRSYVVG YFGANPPPTAPHRT---AHGSWDSITTPAQSRHVLVGLVSGPQNDHYEDRR 393

AAD38027.1 QDYQHNE IACDYNAA YTGALA 416
AAF80584.1 SDYVHNE IADDYNA AFGALA 419
AAF80585.1 YDYVHNE IADDYNA GYGALA 419
AAK12339.1 SDYVHNE IADYNA GFQSAVA 421
AAO61672.1 QDYQHNE IACDYNAA YTGALA 416
AAU20853.2 SDYVHNE IADYNA GFQSAVA 421
AAX92641.1 SDYVHNE IADYNA GFQSAVA 416
ABD24274.1 ENYVHNE IACDYNAA GFQSAVA 415
ABD24275.1 ENYVHNE IADYNA GFQSAVA 414
ABD24276.1 ENYVHNE IADYNA GFQSAVA 414
ABD24277.1 ENYVHNE IADYNA GFQSAVA 414
ABD24278.1 ENYVHNE IADYNA GFQSAVA 414
ABD24279.1 ENYVHNE IADYNA GFQSAVA 414
ABD24280.1 ENYVHNE IADYNA GFQSAVA 414
ABD24281.1 ENYVHNE IADYNA GFQSAVA 414
ABO26609.1 SDYVHNE IADYNA GFQSAVA 410
ABV32557.1 SDYVHNE IADYNA GFQSAVA 421
ACE75510.1 SDYVHNE IADYNA GFQSAVA 416
ACE75511.1 SDYVHNE IADYNA GFQSAVA 416
ADB12483.1 SDYVHNE IADYNA GFQSAVA 421
ADB82658.1 NDYVQNE IADYNA GFQSAVA 421
ADB85440.1 DDYIHNE IACDYNAA LFTGLLA 412
ADB85441.1 SNWQME IGLDYNA YFSSTLA 417
ADB85442.1 SDYVHNE IADYNA YFSSTLA 414
AFD33365.1 NDYVHNE IADYNA GFQSAVA 421
AFQ98379.1 EDYVHNE IADYNA GFQSAVA 408
AFQ98380.1 EDYVHNE IADYNA GFQSAVA 408
AGI81089.1 EDYVHNE IADYNA GFQSAVA 408
AGP76397.1 SDYVHNE IADYNA GFQSAVA 421
AGP76398.1 SDYVHNE IADYNA GFQSAVA 421
AGP76399.1 SDYVHNE IADYNA GFQSAVA 421
AGP76400.1 SDYVHNE IADYNA GFQSAVA 421
AGP76401.1 SDYVHNE IADYNA GFQSAVA 421
AGP76402.1 SDYVHNE IADYNA GFQSAVA 421
AGP76403.1 SDYVHNE IADYNA GFQSAVA 421
AGP76404.1 SDYVHNE IADYNA GFQSAVA 421
AGP76405.1 SDYVHNE IADYNA GFQSAVA 421
AGP76406.1 SDYVHNE IADYNA GFQSAVA 421
AGP76407.1 SDYVHNE IADYNA GFQSAVA 421
AGP76408.1 SDYVHNE IADYNA GFQSAVA 421
AGP76409.1 SDYVHNE IADYNA GFQSAVA 421
AGP76410.1 SDYVHNE IADYNA GFQSAVA 421
AGP76411.1 SDYVHNE IADYNA GFQSAVA 421
AGP76412.1 SDYVHNE IADYNA GFQSAVA 421
AGP76413.1 SDYVHNE IADYNA GFQSAVA 421
AGP76414.1 NDYVQNE IADYNA GFQSAVA 421
AGP76415.1 NDYVQNE IADYNA GFQSAVA 421
AGP76416.1 NDYVQNE IADYNA GFQSAVA 421
AGP76417.1 SDYVHNE IADYNA GFQSAVA 421
AGP76418.1 SDYVHNE IADYNA GFQSAVA 421
AGP76419.1 SDYVHNE IADYNA GFQSAVA 421
AGP76420.1 SDYVHNE IADYNA GFQSAVA 421
AGP76421.1 SDYVHNE IADYNA GFQSAVA 421
AGP76422.1 SDYVHNE IADYNA GFQSAVA 421
AGP76423.1 SDYVHNE IADYNA GFQSAVA 421
AGP76424.1 SDYVHNE IADYNA GFQSAVA 421
AGP76425.1 SDYVHNE IADYNA GFQSAVA 421
AGP76426.1 SDYVHNE IADYNA GFQSAVA 421
AGP76427.1 SDYVHNE IADYNA GFQSAVA 421
AGP76428.1 SDYVHNE IADYNA GFQSAVA 421
AGP76429.1 SDYVHNE IADYNA GFQSAVA 421
AGP76430.1 SDYVHNE IADYNA GFQSAVA 421
AGP76431.1 SDYVHNE IADYNA GFQSAVA 421
AGP76432.1 SDYVHNE IADYNA GFQSAVA 421
AGP76433.1 SDYVHNE IADYNA GFQSAVA 421
AGP76434.1 SDYVHNE IADYNA GFQSAVA 421
AGP76435.1 SDYVHNE IADYNA GFQSAVA 421
AGP76436.1 SDYVHNE IADYNA GFQSAVA 421

AGP76437.1	SDYVQNE	ACDYN	NAGFQSAVP	421
AGS32241.1	SDYISNE	ATDYN	NAGFQSAVA	421
AIT11911.1	NDYVKNE	ACDYN	NAFTGALA	414
AKC91349.1	NDYVKNE	ATDYN	NAGFQSAVA	415
AKV16371.1	SDYISNE	ATDYN	NAGFQSAIA	421
AMH40356.1	KDAQHNE	SCDYN	NAGFQGVIA	417
AMH40357.1	KDAQHNE	ACDYN	NAGFQGVIA	417
AMH40358.1	SDWRHTE	GCDYN	NACFQSILA	416
AMH40359.1	SDYVHNE	DTVYN	NAGFQGVLA	417
AMH40360.1	SDYVHNE	ACDYN	NAAYQGALA	419
AMH40361.1	KDAPHNE	ACDEN	NAFQSCIA	417
AMH40362.1	KDAPHNE	ACDEN	NASFOACIA	417
AMH40363.1	SDYVHNE	ACDYN	NAAFQGVIA	419
AMH40364.1	TDWRHSE	GCDYN	NACFQSMLA	416
AMH40365.1	SDYVHNE	ATYYN	NSGFOTVVA	417
AMH40366.1	NDYVHNE	ATYYN	NSGFOTVVA	417
AMH40367.1	KDAQHNE	ACDYN	NAAFQGVIA	417
AMH40368.1	TDWRHSE	GCDYN	NACFQSMLA	416
AMH40369.1	KDAPHNE	ACDEN	NASFOACIA	417
AMH40370.1	KDAQKNE	ACDYN	NAAFQSVIA	417
AMH40371.1	SDYVHNE	DTVYN	NAGFQGVLA	417
AMH40372.1	SDYVHNE	ACDYN	NAAYQGALA	420
AMH40373.1	SDYVHNE	ATYYN	NSGFOTIIA	417
AMH40374.1	SDYVHNE	ACDYN	NAAYQGALA	419
AMH40375.1	TDYQHTE	GCDYN	NACFQSMLA	416
AMH40376.1	KDATHNE	TCDNN	NAGYHGVLV	417
AMH40377.1	NDAQHNE	ACDYN	NAGFQAVIA	417
AMH40378.1	SDYVHNE	DTVYN	NAGFQSVVA	417
AMH40379.1	SDYVHNE	ACDYN	NAAYQGALA	420
AMH40380.1	TDYQHTE	GCDYN	NACFQSMLA	416
AMH40381.1	SDYVHNE	DTVYN	NAGFQGVLA	417
AMH40382.1	KDAPHNE	ACDEN	NAFQACIA	417
AMH40383.1	KDAQHNE	ACDYN	NAAFQGVIA	417
AMH40384.1	SDYVHNE	ACDYN	NAAYQGALA	419
AMH40385.1	SDYVHNE	ATYYN	NAGFQSVLA	417
AMH40386.1	TDWRHTE	GCDYN	NACFQSMLA	416
AMH40387.1	SNYVVNE	DTVYN	NAGFQGVLA	417
AMH40388.1	KDAQHNE	ACDYN	NAAFQGVIA	417
AMH40389.1	KDAQHNE	ACDYN	NAAFQGVIA	417
AMH40390.1	TDYRHSE	GCDYN	NACFQSMLA	416
AMH40391.1	KDAPHNE	ACDEN	NAFQSCIA	417
AMH40392.1	QDAQHNE	ACDYN	NAAYQGALS	417
AMH40393.1	SDWRHTE	GCDYN	NAAFQSTLA	416
AMH40394.1	SDYQHNE	ACDYN	NAAFQGVIA	417
AMH40395.1	EYVYVNE	TLDYN	NAGFQSAVA	413
ANR02619.1	QDYVSN	ACDYN	NAGFQGALA	412
AOV94250.1	QDAQHNE	ADDYN	NAAFQGSVLV	417
AOV94251.1	TDYQHNE	ACDYN	NAAFQGVLA	417
AOV94252.1	YDAVHNE	ACDYN	NAAFQGVIA	420
AOV94253.1	RDYVYNE	TLDYN	NAGFQSAVA	412
AOV94254.1	TDYEHNE	ACDYN	NAAFQGVLA	420
AOV94255.1	TDYQHNE	ACDYN	NAAFQGVLA	420
AOV94256.1	TDYEHNE	ACDYN	NAAFQGVLA	420
APF29556.1	NDYVKNE	ACDYN	NAGFTGALA	414
APF29557.1	NDYVKNE	ACDYN	NAGFTGALA	413
APF29558.1	NDYVKNE	ACDYN	NAAYTGALA	413
APF29559.1	NDYVKNE	ACDYN	NAFTGALA	403
APF29560.1	NDYVKNE	ACDYN	NAAYTGALA	415
APF29561.1	DDYVHNE	ACDYN	NAGITAGAA	416
ATS91249.1	TNYVHNE	ACDYN	NAGFTGAVA	414
BAA31326.1	SDYISNE	ATDYN	NAGFQSAVA	421
BAA33708.1	SDYVHNE	ATDYN	NAGFQSALA	421
BAA33709.1	SDYVHNE	ATDYN	NAGFQSALA	421
BAA34050.1	SDYISNE	ATDYN	NAGFQSAVA	421
BAB40693.1	SDYISNE	ATDYN	NAGFQSAVA	421
BAB40695.1	SDYISNE	ATDYN	NAGFQSAVA	421
BAB40696.1	SDYISNE	ATDYN	NAGFQSAVA	421
BAB40697.1	SDYISNE	ATDYN	NAGFQSAVA	421
BAC67166.1	EDYVHNE	ACDYN	NSGFQSALA	408
BAD01504.1	EDYVHNE	ACDYN	NSGFQSALA	408
BAD44734.1	EDYVHNE	ACDYN	NSGFQSALA	408
BAF38757.1	DDYVKNE	ACDYN	NAGFQSALA	404
BAH21460.1	SDYISNE	ACDYN	NAGFQGAVA	416
BAK20401.1	QDYISNE	ACDYN	NAGFQGALA	412
BAL69587.1	NDYVHNE	ACDYN	NAGYTSLLA	413
BAM14716.1	SDYIRNE	ACDYN	NAGFQGAVA	415
BAS16736.1	KDYVLLNE	ATDYN	NSGFHSALA	390
BAT21145.1	SDYIRNE	ACDYN	NAGFQGAVA	415
BAL45387.1	KDYVLLNE	ATDYN	NSGFHSALA	390
BAL45388.1	KDYVLLNE	ATDYN	NSGFHSALA	390
BAX06271.1	SDYTHNE	ATDYN	NSGFQGALA	387
CAD54726.1	SDYVANE	ATDYN	NAAFQSLLS	420
CAD54727.1	SDYVANE	ATDYN	NAAFQSLLS	420
CAD54728.1	SDYVANE	ATDYN	NAAFQSLLA	420
CAD54729.1	SDYVANE	ATDYN	NAAFQSLLA	421
CAD54730.1	SDYVANE	ATDYN	NAAFQSLLA	421
EFN70196.1	HDGNYTG	GIDYN	NAGFTCVLA	415
EFN70197.1	HDGNYTG	GIDYN	NAGFTCVLA	416
EFX69372.1	SNYVSN	ATDYN	NAGFQGAVA	410
EFX60604.1	SNYVTNE	ACDYN	NAGFQGSVA	411
EFX80605.1	HNFFTNE	ACDYN	NAGFQASVA	411
EFX80622.1	DDFRANE	ALDYN	NAGFQSALA	415
EGI63652.1	EYAYTNE	TLDYN	NAGFTSALA	415
EKC17289.1	LDFQSN	ACDYN	NAGFQSAVA	413
EKC39023.1	DNFRTNE	ACDYN	NAGFQSAVA	412
KDR16731.1	EYVYVNE	TLDYN	NAGFQSAVA	414
KMQ95413.1	EEDIYTE	VLINYN	NAGFTSALA	416
KOC62382.1	EDYVYTE	TLDYN	NAGFTSALA	415
KOF69403.1	SDFIHNE	TCDYN	NAGFHSAVS	414
KOX76353.1	KDYVYTE	TVEYN	NGGFTSALA	415
KXJ16436.1	KDYIENE	ATDYN	NAGFQSAVA	416
KYB27032.1	EFLVNE	TLDYN	NAGFQSTLA	415
KYB87195.1	EYAYTNE	TLDYN	NAGFTSALA	415
KYN01696.1	EYAYTNE	TLDYN	NAGFTSALA	415
KYN22851.1	EYAYTNE	TLDYN	NAGFTSALA	415
KYN35700.1	EYAYTNE	TLDYN	NAGFTSALA	415
KYQ52831.1	EYAYTNE	TLDYN	NAGFTSALA	415
KZC10616.1	EDYVYTE	TLDYN	NGGFTGALA	415
KZS03923.1	DDYRSNE	ALDYN	NAGFQSALA	418
KZS05722.1	ENYQTN	ALDYN	NAGFQASLA	409
KZS07093.1	QNYVTNE	ATDYN	NAGFQAAVA	411
KZS16514.1	DNVYVNE	ACDYN	NAGFQASLA	414
KZS16516.1	ENYVTNE	ACDYN	NAGFQASLA	410
OAD61293.1	EDYVYTE	TLDYN	NAGFTSALA	415

ODM94862.1	EDYVAN	ETIDY	AGFQGALA	418
ODM95900.1	GNYVSN	EIAIDY	AGFQGVAA	414
OWF45593.1	ENYVNN	WACDY	SGFQGVSS	415
OWF56607.1	SDYIQN	WACDY	AGFQSVVA	417
PBC31733.1	EDYVYT	ETLDY	AGFTTOLA	415
PH59403JA	SDYIRN	EANDY	AGFQGAVA	416
PK25363.1	KDYVMN	ETTDY	AGFQSSVA	418
PSN31180.1	NDAQHN	WACDY	AGFQSALA	411
PVD35327.1	KDYVHN	WACDY	AGFQSAVA	414
PVD35340.1	GNYINN	WACDY	AGFQGSVA	414
PVD35381.1	ENYVNN	WACDY	AGFQTAVA	399
PVD35383.1	ENYVNN	WACDY	AGFQTAVA	414
PVD35384.1	DQYLYN	WAIITY	AGFQTAVA	415
PVD35385.1	EYYVND	WAIIVY	AGFQTAVA	414
PVD35551.1	KDYVMN	EIAIDY	SGFQSALA	407
PVD36624.1	SNYVNN	WACDY	AGFQTAVA	412
PVD36625.1	SNYVNN	WACDY	AGFQTAVA	399
PVD36896.1	PDYVHN	WALDY	SGFQSALA	389
PVD36896.1	EDYVKN	EIAIDY	SGFQSALA	394
XP_001806454.1	EDYVYT	ETLDY	AGFTGALA	415
XP_001810693.2	EEFLYN	ETLDY	AGFQSTLA	415
XP_001944774.2	EEIYNN	ETLDY	AGFQSAVA	414
XP_002426465.1	EEYVYN	ETLDY	AGFQSAVA	414
XP_003402778.1	EDYVYT	ETLDY	AGFTSALA	415
XP_003690076.1	EDYVYT	ETLDY	AGFTSALA	415
XP_005099312.1	LNYYTL	WATDY	SGFQGGALA	393
XP_005099315.1	GDYVHN	WATDY	SGFQGGALA	388
XP_005101787.1	TDYQHN	WATDY	SGFQGGALA	388
XP_005101789.1	TDYKHN	WATDY	SGFQGGALA	389
XP_005111688.2	GNFQNN	WATDY	AGFTTAVA	415
XP_005111689.2	SNYVNN	WATDY	AGFTTAVA	415
XP_005111870.1	LNYYNN	WATDY	AGFTTALA	414
XP_006810535.1	EDYVYT	ETLDY	AGFTSALA	415
XP_008181842.1	EEIYNN	ETLDY	AGFQSAVA	414
XP_008194455.1	EEFLYN	ETLDY	AGFQSTLA	414
XP_009049791.1	DDYIHN	WATDY	AGFOASIA	414
XP_009049836.1	NDYIHN	WACDY	AGFQSAVA	399
XP_009049839.1	KDYTKN	WACDY	AGFQSTVA	414
XP_009049840.1	EDFIQN	WACDY	AGFQSALA	415
XP_009049941.1	GNYVNN	WATDY	AGFQGAVA	414
XP_009058133.1	QNYINN	WATDY	AGFQSAVA	414
XP_009058134.1	EDYVHN	WACDY	SGFHSALA	408
XP_009060290.1	EDYVQN	WACDY	AGFQSGLA	410
XP_009064976.1	DNYVSN	WATDY	AGFQSAVA	414
XP_009064980.1	GDYVRN	WALDY	AGLQSCACA	410
XP_009065702.1	DDYVHN	WACDY	AGFQSALA	405
XP_011057333.1	EEYAYT	ETLDY	AGFTSALA	415
XP_011164235.1	EEYAYT	ETLDY	AGFTSALA	415
XP_011253653.1	HGDNYS	WGLDY	AGFTCVLA	415
XP_011253654.1	HGDNYS	WGLDY	AGFTCVLA	414
XP_011253655.1	HGDNYS	WGLDY	AGFTCVLA	416
XP_011296914.1	EDYVYT	ETLDY	AGFTSALA	415
XP_011296916.1	EDYVYT	ETLDY	AGFTSALA	415
XP_011346896.1	DDYIYT	ETLDY	AGFTSALA	415
XP_011421792.1	LDFQSN	WACDY	AGFQSAVA	413
XP_011423159.1	KDYVSN	WACDY	AGFQSAVA	414
XP_011430172.1	---	NAGCF	AGFQSALA	404
XP_011447501.1	DNFRIN	WACDY	AGFQSAVA	412
XP_011447502.1	DNFRIN	WACDY	AGFQSAVA	410
XP_011488432.1	EDYVYT	ETLDY	AGFTGALA	415
XP_011632703.1	AEYIYT	ETLDY	AGFTSALA	415
XP_011700378.1	EEYAYT	ETLDY	AGFTSALA	415
XP_011867251.1	EEYAYT	ETLDY	AGFTSALA	415
XP_012061951.1	EEYAYT	ETLDY	AGFTSALA	415
XP_012146613.1	EDYVYT	ETLDY	AGFTSALA	415
XP_012224814.1	EEYIYT	ETLDY	AGFTGALA	416
XP_012248155.1	EDYVYT	ETLDY	AGFTSALA	415
XP_012273697.1	EDYVYT	ETLDY	AGFTSALA	415
XP_012526654.1	EEYAYT	ETLDY	AGFTSALA	415
XP_012935745.1	DNYQNN	WATDY	AGFTTALA	415
XP_012938497.1	DNVGNN	WATDY	AGFQGGALA	389
XP_012939931.1	SDYIHN	ETTDY	SGFQGGALA	386
XP_012939933.1	NDHKHN	WATDY	SGFQGGALA	351
XP_012941225.1	GDIKNN	ETTDY	AGFQSAVA	415
XP_012945520.1	ENYQNN	WATDY	AGFTTAVA	415
XP_013066277.1	PDWIYN	WEIDY	SGFQGGALA	390
XP_013066278.1	SDYVQN	WISISY	SLHGLALA	393
XP_013066280.1	SDYVHN	WATDY	SGFQGGALA	393
XP_013066281.1	DDYVTN	WATDY	AGFQGGALA	394
XP_013066283.1	DDYVTN	WATEY	AGFQGGALA	393
XP_013066284.1	DNYVLN	WSTDY	AGYQGGALS	394
XP_013066285.1	DDYVLN	WALDY	AGFQGGALA	393
XP_013066286.1	EDYVLN	WALDY	AGFQGGALA	393
XP_013066370.1	TDYILN	WATDY	SGFQGGALA	393
XP_013066371.1	SDYILN	WATDY	SGFQGGALA	392
XP_013069169.1	SNYVNN	WACDY	AGFQTAVA	415
XP_013069170.1	SNYVNN	WACDY	AGFQTAVA	415
XP_013069521.1	SNYVNN	WACDY	AGFOAAVA	415
XP_013069522.1	SNYVNN	WACDY	AGFOAAVA	415
XP_013069523.1	SNYVNN	WACDY	AGFOAAVA	415
XP_013084279.1	DSVLLN	WTEFN	AGFQTAVA	415
XP_013084281.1	DSVLLN	WTEFN	AGFQTAVA	415
XP_013084282.1	TNEVNN	WSEYN	AGFQTAVA	415
XP_013084283.1	TNEVNN	WSEYN	AGFQTAVA	415
XP_013084285.1	TNYLYN	WACDY	AGFQSAVA	414
XP_013084287.1	TNYLYN	WACDY	AGFQSAVA	414
XP_01308631.1	SDYVLN	WACDY	SGFHGLALA	391
XP_013084438.1	TDYVHN	WATDY	SGFQGGALA	391
XP_014214612.1	EDFIYT	ETLDY	AGFTGALA	414
XP_014601854.1	EEYIYT	ETLDY	AGFTSALA	415
XP_014786694.1	SDFIHN	WTCDY	AGFHSVVS	416
XP_015126097.1	EDYVYT	ETLDY	AGFTSALA	415
XP_015184471.1	EEYIYT	ETLDY	AGFTSALA	415
XP_015432538.1	EDYVYT	ETLDY	GGFTGALA	415
XP_015513382.1	EDYIYT	ETLDY	AGFTSALA	415
XP_015592927.1	EDYVYT	ETLDY	AGFTSALA	415
XP_015836219.1	EEFLYN	ETLDY	AGFQSTLA	413
XP_016907626.1	EDYVYT	ETLDY	AGFTSALA	415
XP_017786289.1	EDYVYT	ETLDY	AGFTSALA	415
XP_017785110.1	EEYLYN	ETIDY	AGLHSLLA	412
XP_017793200.1	EDYVYT	ETLDY	AGFTSALA	415
XP_018016806.1	SDYVHN	WAVGY	TGLTASLA	414
XP_018016809.1	NDYVQN	WACDY	AGLTGALA	415
XP_018016810.1	MDYMHN	WACDY	AGLTGALA	414

Accession	Sequence	Position
XP_018016813.1	GNWEGNEGGDDYNAALAAANLA	412
XP_018016820.1	TNFDVQNEATDYNAGFVGTLA	416
XP_018016821.1	TNFDVQNEATDYNAGFVGTLA	416
XP_018026486.1	MDFVHNEACDYNAAFVGVLA	414
XP_018027134.1	NDYVKNEACDYNAAFTGDLA	413
XP_018044452.1	E EYAYTE TLDYNAAGFTSALA	415
XP_018307051.1	E EYAYTE TLDYNAAGFTSALA	415
XP_018323591.1	E EFLYNE TLDYNAAGFQGVVA	414
XP_018347393.1	E EYAYTE TLDYNAAGFTSALA	415
XP_018359268.1	E EYAYTE TLDYNAAGFTSALA	415
XP_018396635.1	E EYAYTE TLDYNAAGFTSALA	415
XP_018568194.1	E EFLYNE TVDYNAAGFQSALA	415
XP_018568196.1	E EFLYNE TVDYNAAGFQSALA	415
XP_019884179.1	Y E P D Y T N W R I D Y N A G F T C V L A	414
XP_020289295.1	N I R C H T N W K I E N N A G F T G V L A	409
XP_020289297.1	F N - K Y T K R I D Y N A G F T G A L A	412
XP_020289301.1	F R - E H T K W R I D Y N A G F T G A L A	411
XP_020289336.1	Y G N - Y T E N I E Y N A G F T G A L A	414
XP_020614867.1	K D F I K N E W A T D Y N A G F Q S A V A	418
XP_020893295.1	S D Y V Q N E W T T D Y N A G F Q S A V A	418
XP_020893299.1	K D Y I E N E W A T D Y N A G F Q S A V A	418
XP_021345903.1	S D Y I Q N E W A C D Y N A G F Q S V V A	417
XP_021361004.1	K D Y V K N E W A C D Y N S G F Q S A L A	407
XP_021363481.1	O D Y K Q N E W A C D Y N A G F Q S A V A	414
XP_021363490.1	S D Y T K N E W T C D Y N A G F Q S A L A	405
XP_021363495.1	E N Y V N N E W A C D Y N S G F Q G A S S	415
XP_021363496.1	E N Y V N N E W A C D Y N S G F Q G A S S	415
XP_021363497.1	G N Y I N N E W A C D Y N A G F Q S A V A	414
XP_021363498.1	G N Y I N N E W A C D Y N A G F Q S A V A	418
XP_021924915.1	E E Y V Y N E W T L D Y N A G F Q S A V A	414
XP_021935277.1	K E T I H N M W S C D C N A A F Q S A V A	415
XP_021941324.1	S N Y I S N E W A T D Y N A A F Q G L L A	421
XP_021950548.1	G D Y I K N E W A T D Y N A A F Q G A L A	413
XP_021966549.1	G D Y I K N E W A T D Y N A A F Q G A L A	415
XP_021969048.1	N D A V R N E W A V D Y N A G F Q G V V A	414
XP_021969378.1	E D Y V A N E W T L D Y N A G F Q G A V S	418
XP_021967587.1	N D Y V K N E W T T D Y N A G F Q S A L A	414
XP_021967287.1	E E Y I Y N E W T L D Y N A G F Q S A V A	414
XP_022167288.1	E E Y I Y N E W T L D Y N A G F Q S A V A	414
XP_022200004.1	E D Y I Y N E W T I D Y N A G F Q S A V A	413
XP_022328778.1	D N Y R T N E W A C D Y N A G F Q S A V A	412
XP_022329688.1	D N Y R T N E W A C D Y N A G F Q S A V A	412
XP_022336001.1	- - - - N H G C D F N A G F Q S A L A	404
XP_022336153.1	L D F Q S N E W A C D Y N A G F Q S A V A	413
XP_022338490.1	K D Y T K N E W T C D Y N A G F Q S A L A	409
XP_022338491.1	- N Y A H N H W G C D F N A G F Q S A L A	405
XP_022341787.1	K D Y V S N E W A C D Y N A G F Q S A V A	414
XP_023337121.1	S D Y I K N E W A C D Y N A G F Q S A L A	412
XP_023704929.1	S N Y I S N E W A T D Y N A A F T S L L A	421
XP_023704930.1	S N Y I S N E W A T D Y N A A F T S T L A	420
XP_023716596.1	E E Y V Y N E W T L D Y N A G F Q S A V A	414
XP_023718476.1	L D A V H N I W A C D Y N A A F Q S A L A	419
XP_386191.3	E D Y V Y T E W T L D Y N A G F T S A L A	415
AAA23086.1	Q D Y V A N E W A T D Y N A G F T S A L A	414

Figura 1 Alineamiento de las 349 secuencias de celulasas de la familia GH9 obtenidas en la fase de búsqueda y la secuencias de *Cellulomonas fimi*. Los sitios conservados en todos los aminoácidos se encuentran resaltados en color negro y gris. Los sitios important

Anexo 4

Grupos Taxonómicos Comprimidos del Árbol 1

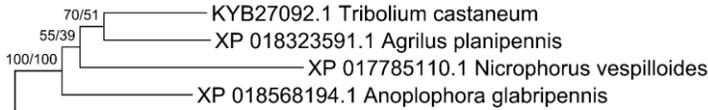


Figura 1 Relaciones filogenéticas formadas por secuencias provenientes del orden Coleoptera

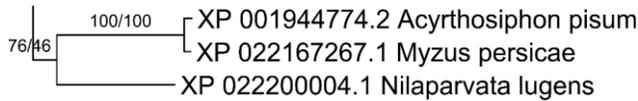


Figura 2 Relaciones filogenéticas formadas por secuencias provenientes del orden Hemiptera

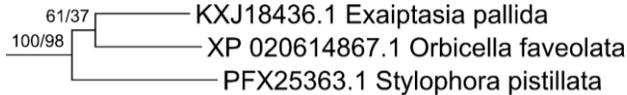


Figura 3 Relaciones filogenéticas formadas por secuencias provenientes del orden Embioptera

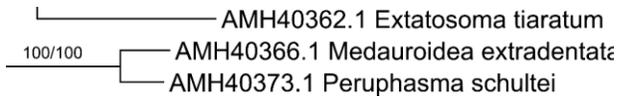


Figura 4 Relaciones filogenéticas formadas por secuencias provenientes del orden Phasmatodea, Clado I

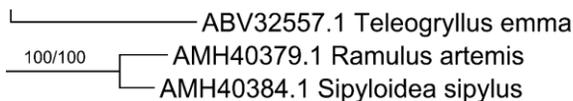


Figura 5 Relaciones filogenéticas formadas por secuencias provenientes del orden Phasmatodea, clado II

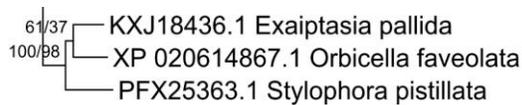


Figura 6 Relaciones filogenéticas formadas por secuencias provenientes del filo Cnidaria



Figura 7 Relaciones filogenéticas formadas por secuencias provenientes del filo Annelida

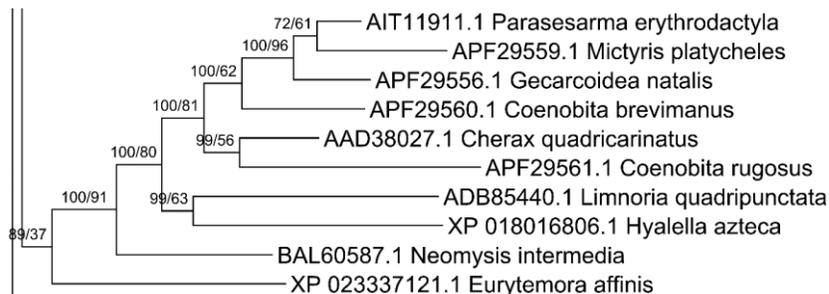


Figura 8 Relaciones filogenéticas formadas por secuencias provenientes del subfilo Crustacea

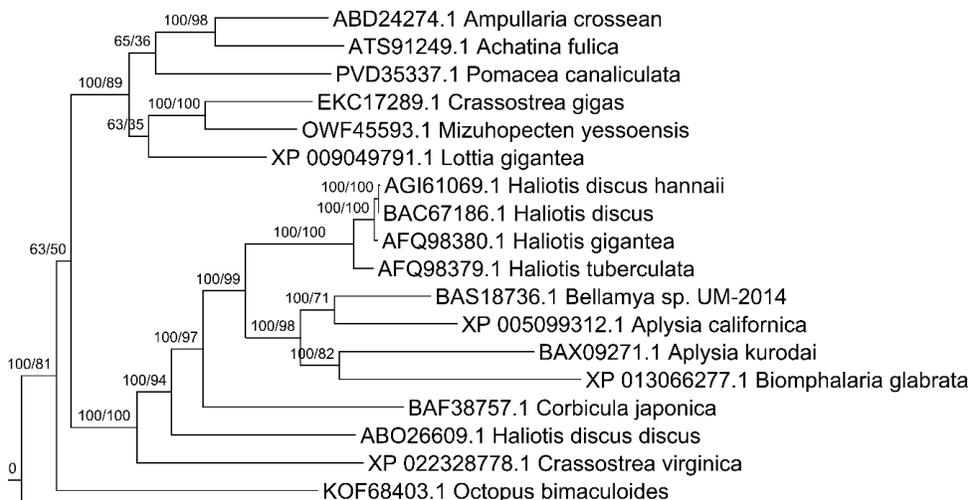


Figura 9 Relaciones filogenéticas formadas por secuencias provenientes del filo Mollusca

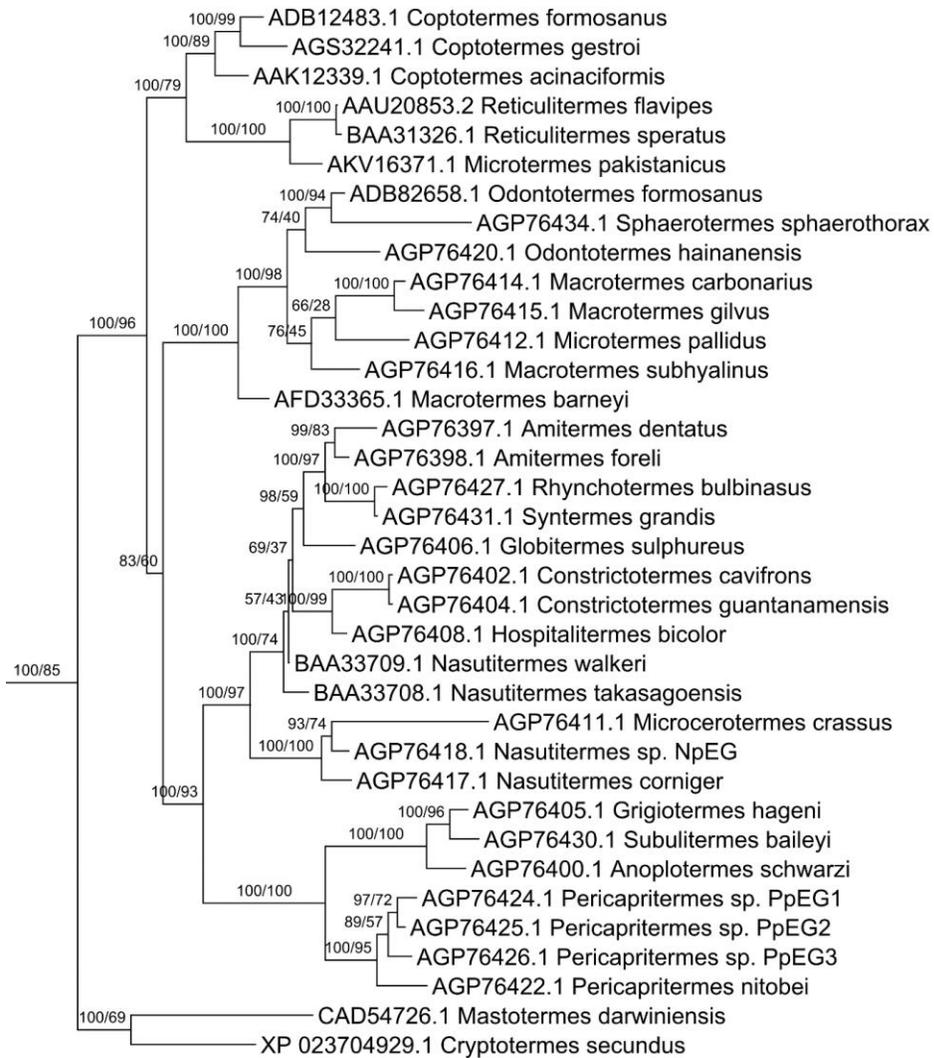


Figura 10 Relaciones filogenéticas formadas por secuencias provenientes de la familia Termitoidae

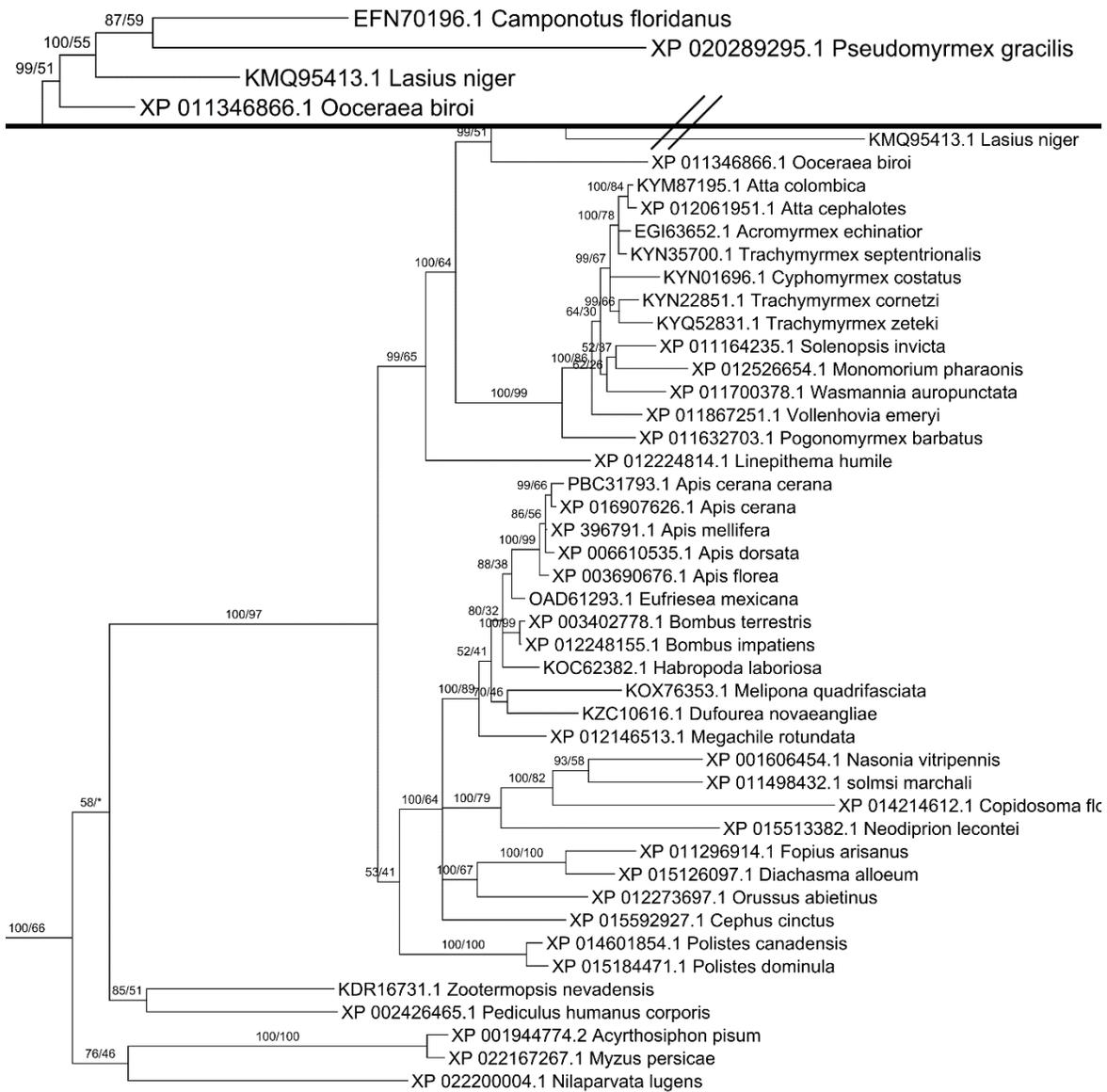


Figura 11 Relaciones filogenéticas formadas por secuencias provenientes del orden Hymenoptera

Grupos Taxonómicos Comprimidos del Árbol 2

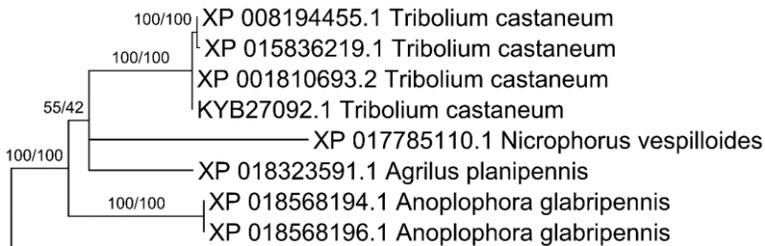


Figura 12 Relaciones filogenéticas formadas por secuencias provenientes del orden Coleoptera

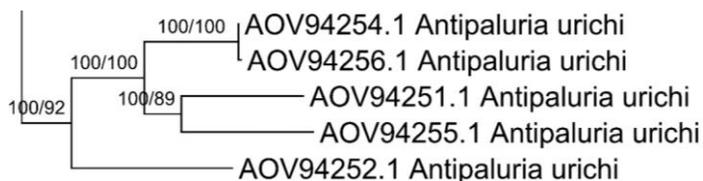


Figura 13 Relaciones filogenéticas formadas por secuencias provenientes del orden Embioptera



Figura 14 Relaciones filogenéticas formadas por secuencias provenientes de la familia Termitidae, Clado I, presentes en el clado Polyneoptera/ Holometabola/Paraneoptera



Figura 15 Relaciones filogenéticas formadas por secuencias provenientes de la superfamilia Blaberoidea



Figura 16 Relaciones filogenéticas formadas por secuencias provenientes del orden Collembola

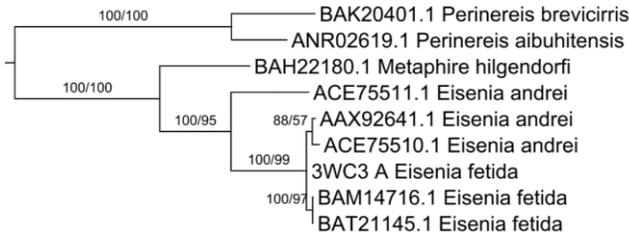


Figura 17 Relaciones filogenéticas formadas por secuencias provenientes del filo Annelida

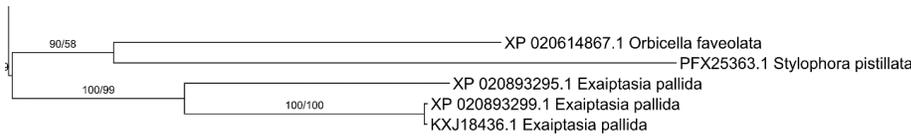


Figura 18 Relaciones filogenéticas formadas por secuencias provenientes del filo Cnidaria

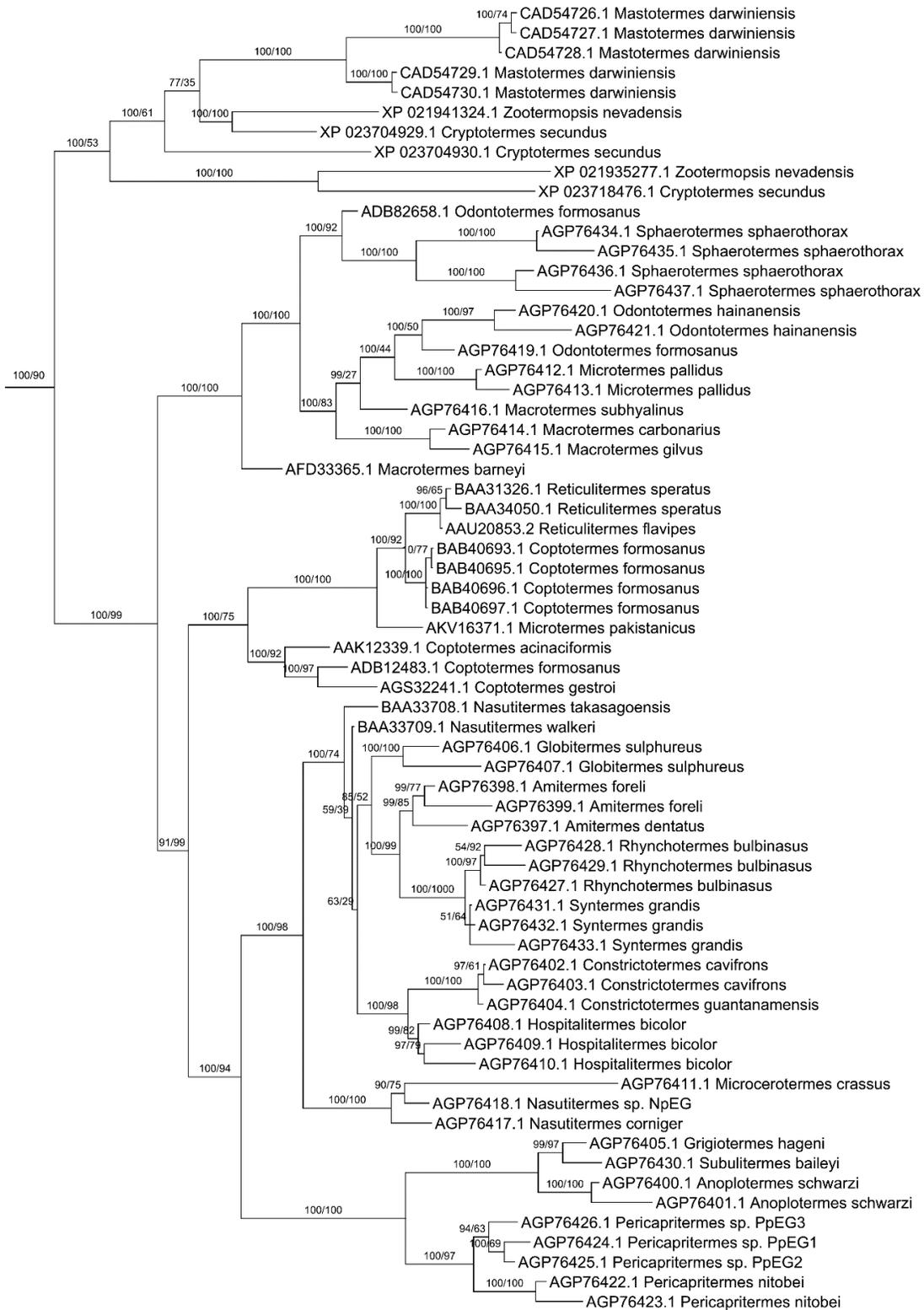


Figura 19 Relaciones filogenéticas formadas por secuencias provenientes de la familia Termitidae, Clado II, presentes en el clado Polyneoptera

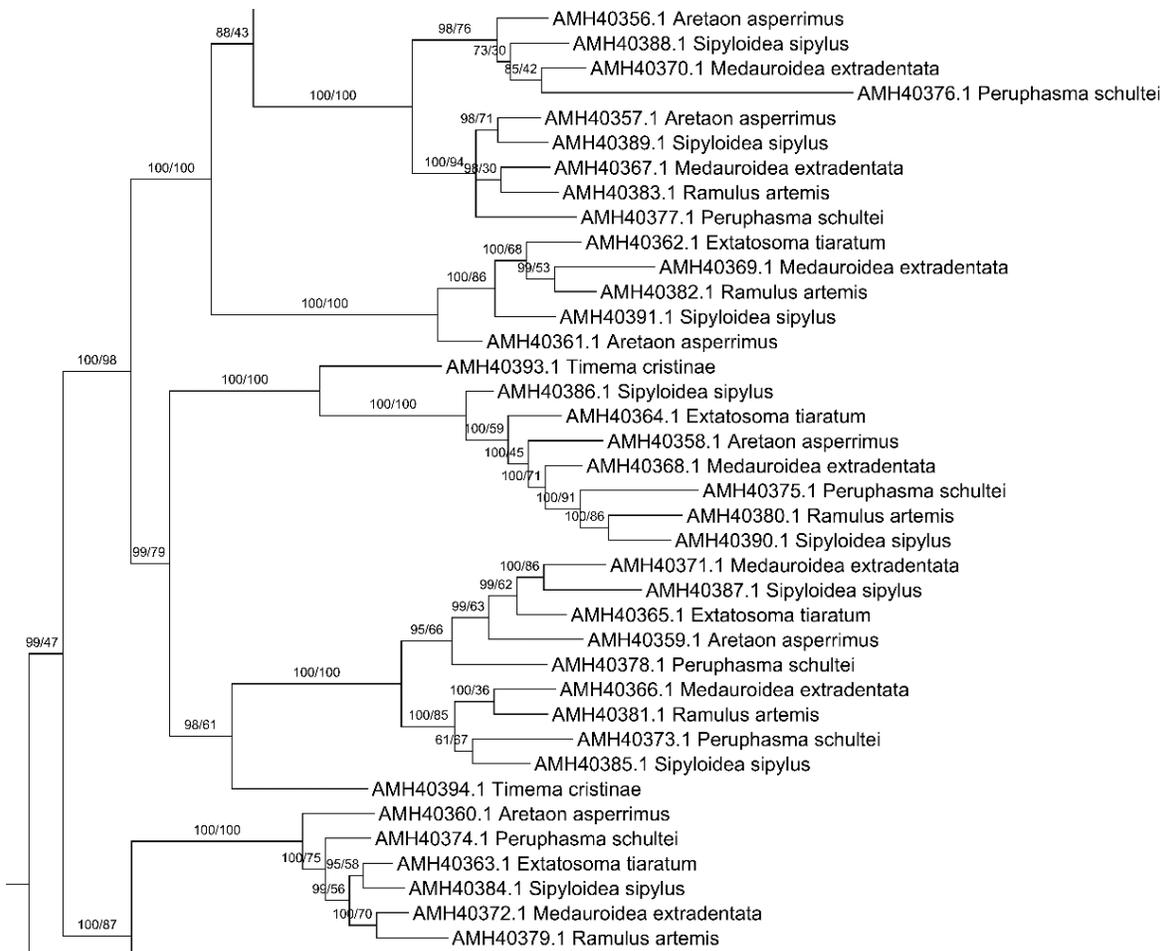


Figura 20 Relaciones filogenéticas formadas por secuencias provenientes del orden Phasmatodea

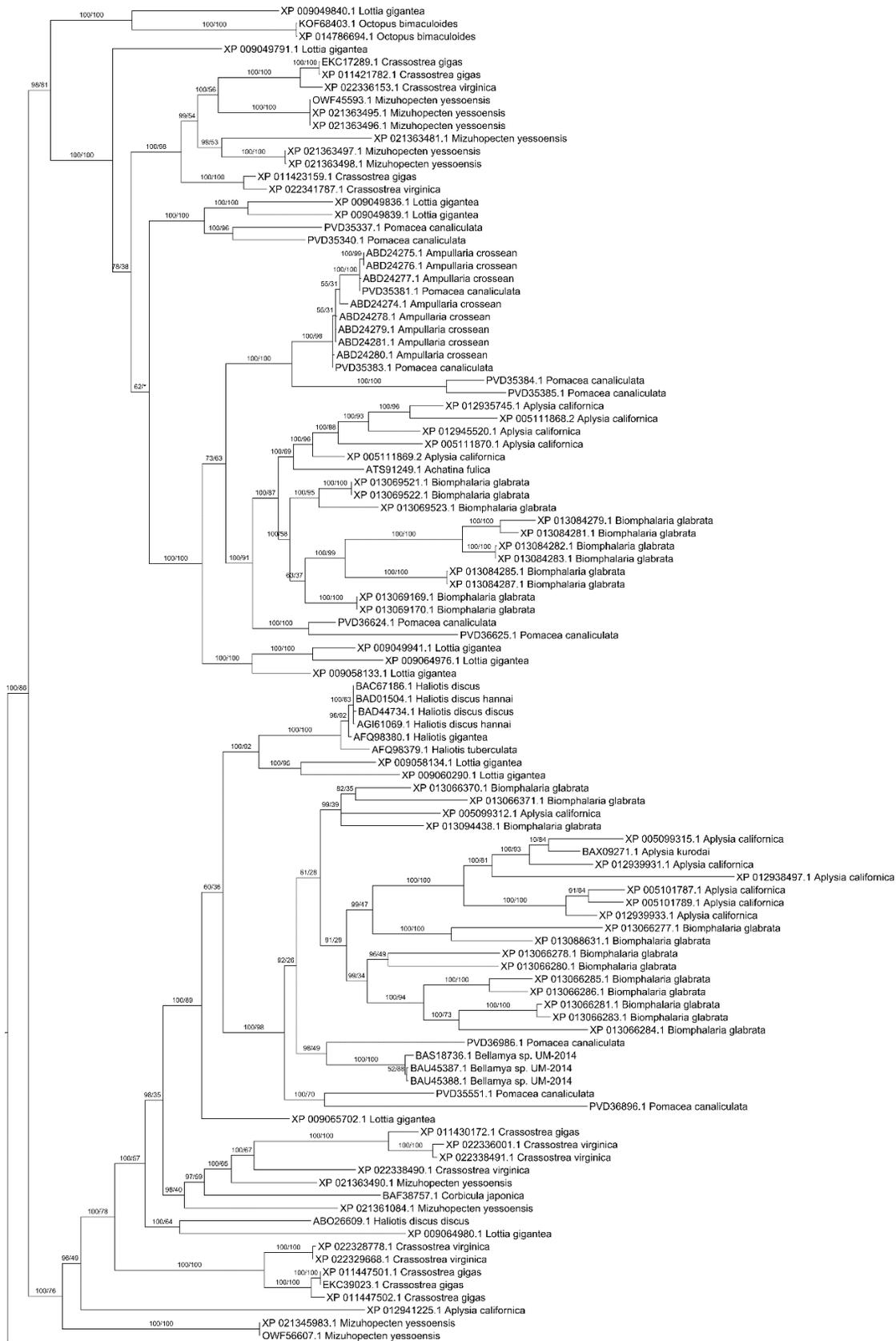


Figura 21 Relaciones filogenéticas formadas por secuencias provenientes del filo Mollusca

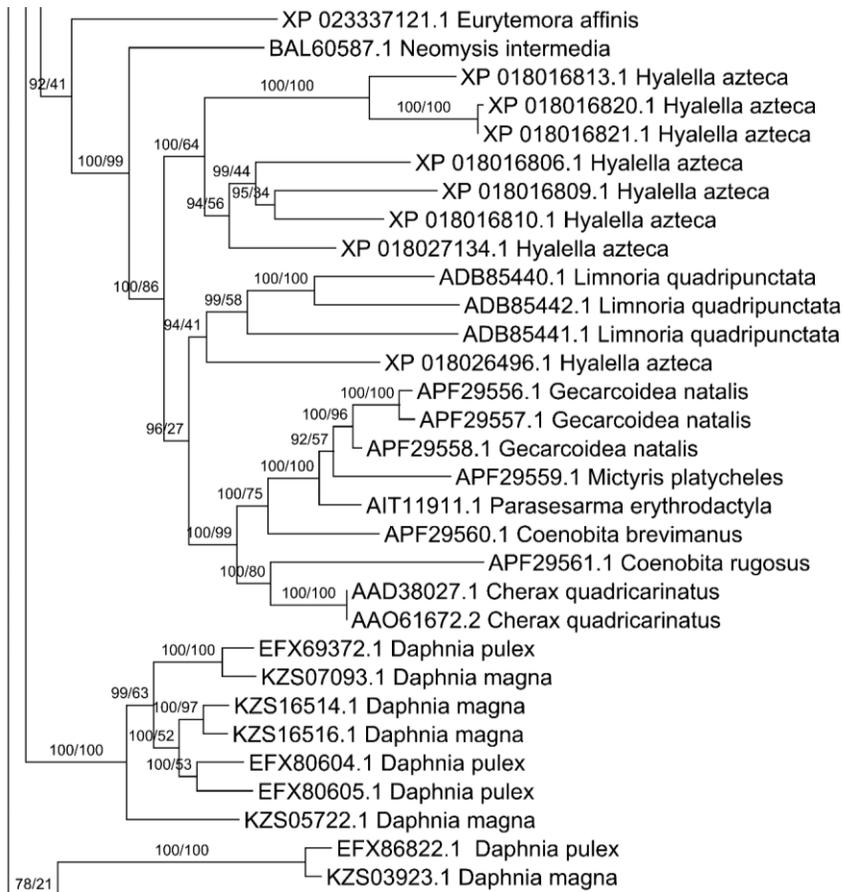


Figura 22 Relaciones filogenéticas formadas por secuencias provenientes del subfilo Crustacea

Grupos Taxonómicos Comprimidos del Árbol 3

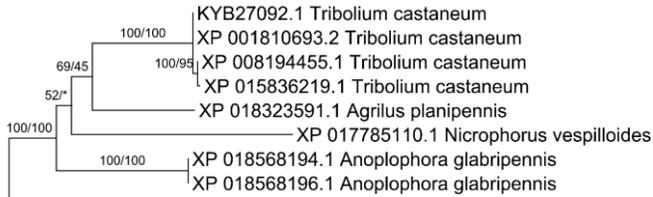


Figura 24 Relaciones filogenéticas formadas por secuencias provenientes del orden Coleoptera

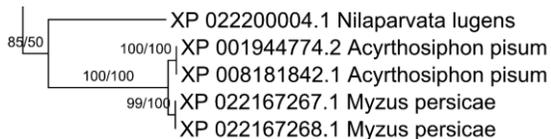


Figura 25 Relaciones filogenéticas formadas por secuencias provenientes del orden Hemiptera

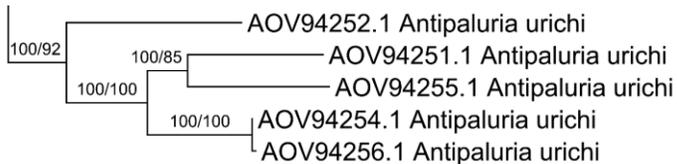


Figura 26 Relaciones filogenéticas formadas por secuencias provenientes del orden Embioptera

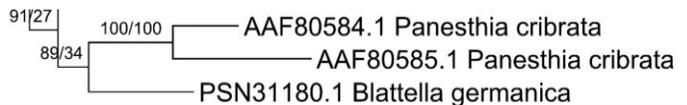


Figura 27 Relaciones filogenéticas formadas por secuencias provenientes de la familia Blaberoidea

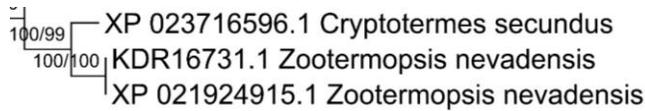


Figura 28 Relaciones filogenéticas formadas por secuencias provenientes de la familia Termitidae, Clado I, presentes en el clado Polyneoptera/ Holometabola/Paraneoptera

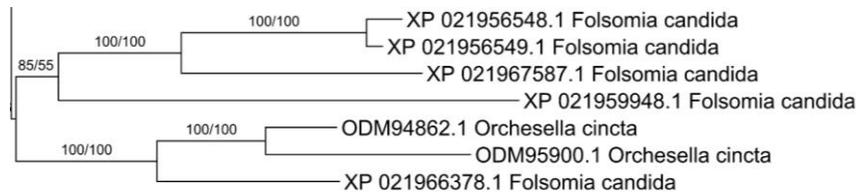


Figura 30 Relaciones filogenéticas formadas por secuencias provenientes del orden Collembola

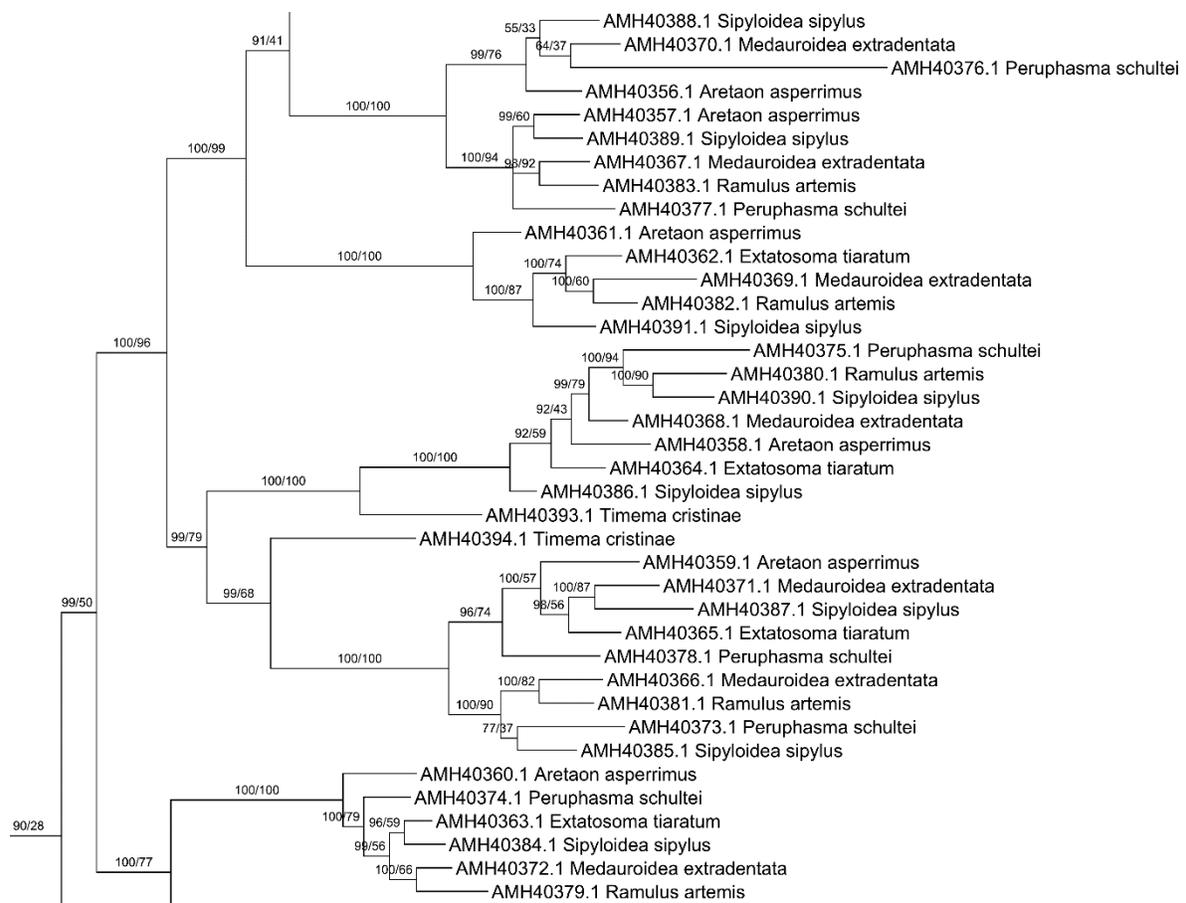


Figura 29 Relaciones filogenéticas formadas por secuencias provenientes del orden Phasmatodea

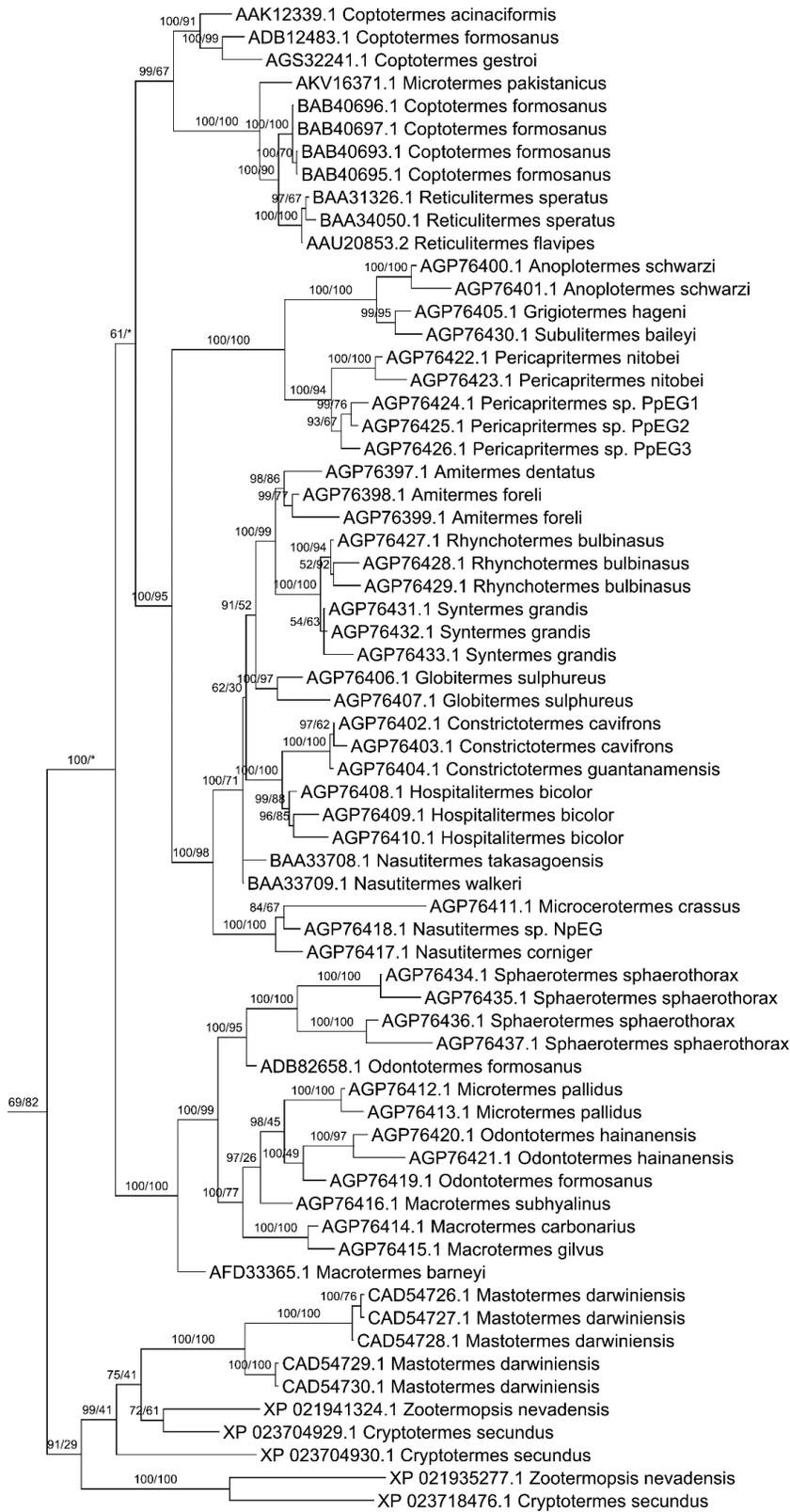


Figura 31 Relaciones filogenéticas formadas por secuencias provenientes de la familia Termitodae, Clado II, presentes en el clado Polyneoptera

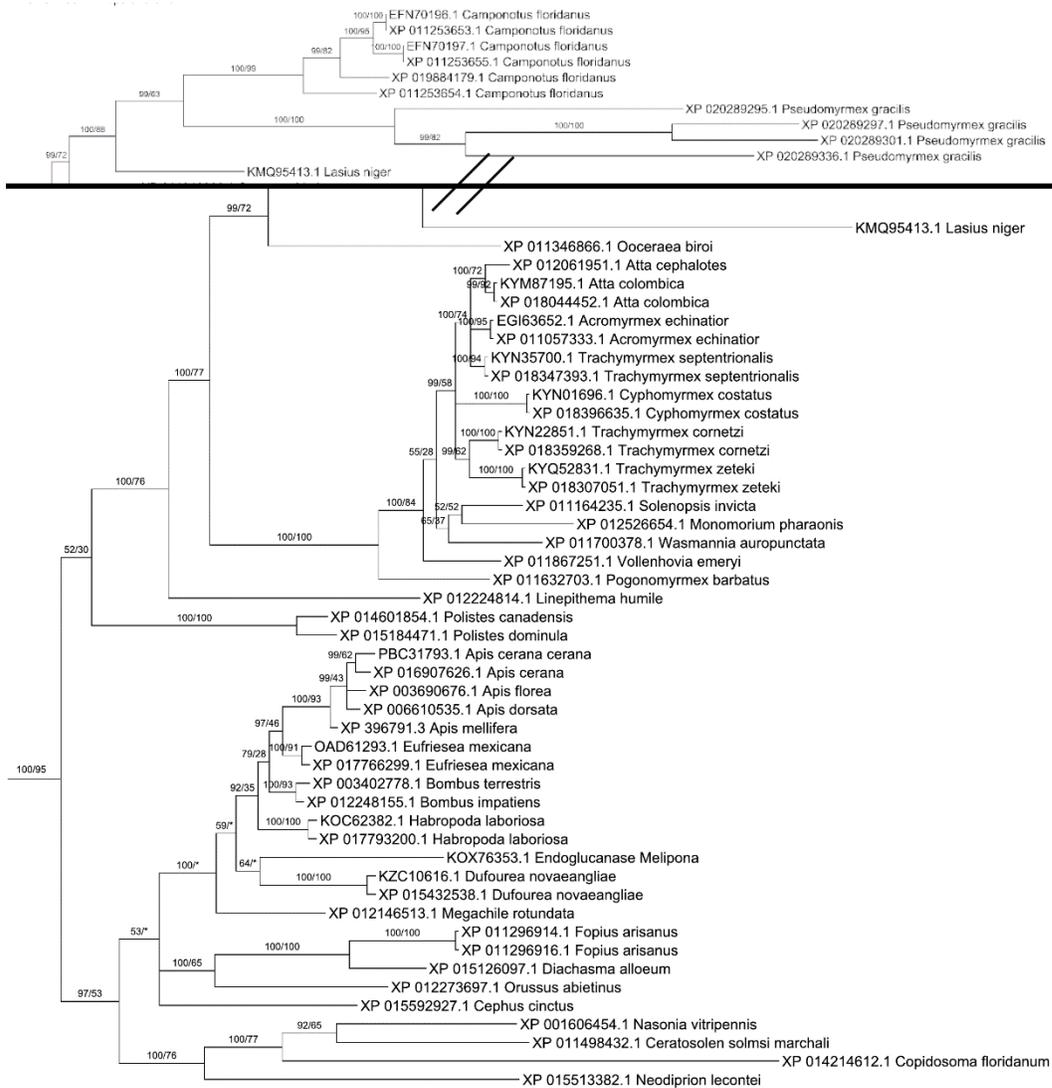


Figura 32 Relaciones filogenéticas formadas por secuencias provenientes del orden Hymenoptera