

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

PRESENTA

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

PRESENTA

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.

A t e n t a m e n t e "POR MI RAZA HABLARA EL ESPIRITU" Morelia, Michoacán a, 22 de abril del 2019.

DR. VICTOR HUGO ANAYA MUÑOZ SECRETARIO GENERAL

#### CAMPUS MORELIA

Apartado Postal 27-3 (Santa Ma. De Guido), 58090, Morelia, Michoacán Antigua Carretera a Pátzcuaro Nº 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 2015/253420: CB "Origen y evolución de los mecanismos endógenos de digestión celulosa de 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 lleva a cabo esta tesis lo dedico con todo el amor a mis padres Mario Diaz 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 revelo 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 revelo las diferencias presentes en la cavidad de unión al substrato, dado que las enzimas del grupo 2 presentan una cavidad más larga y una hendidura con propiedades hidrofóbicas. Mientras que el análisis de electrostática revelo 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	
	6.3 Selección de secuencias	27
	6.4 Análisis filogenéticos	
	6.4.1 Modelo Evolutivo	
	6.4.2 Método Bayesiano	
	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
7	Resultados	
1.	7.1 Púsquada da acquancias	
	7.1 Dusqueda de secuencias	
	7.2 Antileia filoson	
	7.3.1 Arbol Filogenetico 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	
Script 1 Leer y eliminar secuencias repetidas	
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 metá	licos 117
Script 12 Secuencias con mutaciones especificas	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					
Tabla 2 Secuencias de usadas para realizar la búsqueda de secuencias25					
Tabla 3 Secuencias usadas para generar los modelos de la estructura	ı				
terciaria	)				
Tabla 4 Secuencias de celulasas de animales protostomados obtenidas de la	1				
búsqueda realizada en el GenBank	;				
Tabla 5 Secuencias de celulasas con mutaciones o deleciones en los sitios	;				
catalíticos					
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 fi	bras
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 metazoa	arios
	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  $\beta$  -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, Weiset 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- $\beta$ -glucanasas, celobiohidrolasas o exoglucanasas y  $\beta$ -glucosidasas (Bayer et al. 1998, Santoyo et al. 1999, Vries et al.

al. 2001, Wei et al. 2009). Las endoglucanasas rompen los enlaces  $\beta$ -1,4glucosídicos 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  $\beta$ 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, <u>http://www.cazy.org/</u>, 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 (<u>http://www.cazy.org/</u>, 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 liquida, minimizar la perdida 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 arabinoxylana, manana, galactomanana, glucomanana, xiloglucanos. arabinogalactana II y β1\_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, Dgalacturonato, 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  $\beta$ -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 glucooligosacáridos, 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, <u>http://www.cazy.org/</u>, 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) (<u>http://www.cazy.org/</u>, 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, <u>http://www.cazy.org/</u>, 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, <u>http://www.cazy.org</u>, 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, <u>http://www.cazy.org</u>, Lombard et al. 2014). Los clanes GH que incluyen celulasas son (Fig. 5):

• Barril ( $\beta/\alpha$ )8.- Las enzimas pertenecientes a este clan se caracterizan por contar con ocho  $\alpha$ -hélices y ocho  $\beta$ -laminas, las cuales forman un barril central (Gromiha et al. 2004).

• Toroide ( $\alpha/\alpha$ )6.- Las enzimas pertenecientes a este clan se caracterizan por contar con doce  $\alpha$ -hélices, seis de las cuales son internas y se conectan por medio de seis hélices externas (Alzari et al. 1996, Khademi et al. 2002).

•  $\beta$ -"jellyroll".- Las enzimas pertenecientes a este clan se caracterizan por contar con ocho  $\beta$ -laminas, cinco de las cuales son antiparalelas y se encuentran en la posición superior de las tres  $\beta$ -laminas restantes. (Richardson 1981, Gaskell et al. 1995).

•  $\alpha$ -8 superhelíptico.- Las enzimas pertenecientes a este clan se caracterizan por contar con ocho  $\alpha$ -hélices y una pequeña hoja  $\beta$ , compuesta por 3  $\beta$ -laminas (Brás et al. 2012).

• 7-fold  $\beta$ -propeller.- La enzimas pertenecientes a este clan se caracterizan por tener una estructura conformada por dos dominios de siete  $\alpha$ -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 (Articulo 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 (Articulo 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).

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

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

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 órden 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 ( $\alpha/\alpha$ )<sub>6</sub>, 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, Shahriarinour 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 substrato. 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) (<u>https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi,</u> 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 *Olivooides multisulcatus* (Cnidaria; 636.1 a 529 Ma) y *Rusophycus avalonensis* (Ecdyzosoa; 636.1 a 528.82 Ma) identificados (Benton et al. 2015), ubican el origen de los metozarios 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 monofília 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, Deutorostomi, Ecdysozoa y Lophotrochozoa resltados 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 (<u>https://blast.ncbi.nlm.nih.gov/Blast.cgi</u>, Altschul et al.1990, Gish & States 1993) en su versión Protein BLAST utilizando como blanco las secuencias mencionadas.

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

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

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 *Cellumonas 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

(<u>Stamatakis</u> 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).

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

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

Collembolo	Orchesella cincta	ODM94862.1
Orthoptera	Teleogryllus emma	ABV32557.1
Embioptera	Antipaluria urichi	AOV94253.1, AOV94256.1, AOV94250.1
Phasmatodea	Medauroidea extradentata, Peruphasma schultei, Ramulus artemis, Timema cristinae, Timema cristinae	AMH40372.1, AMH40373.1, AMH40379.1, AMH40392.1, AMH40395.1
Blaearoidea	Panesthia cribrata, Blattella germanica	AAF80584.1, PSN31180.1
Termitoidae	Rhynchotermes bulbinasus, Microtermes pakistanicus, Reticulitermes speratus, Mastotermes darwiniensis, Cryptotermes secundus, Zootermopsis nevadensis, Cryptotermes secundus	AGP76427.1, AKV16371.1, BAA31326.1, CAD54726.1, XP_023704929.1, KDR16731.1, XP_023716596
Hemiptera	Acyrthosiphon pisum, Myzus persicae, Nilaparvata lugens	XP_001944774, XP_022167267.1, XP_022200004.1
Phthiraptera	Pediculus humanus corporis	XP_002426465.1
Coleoptera	Tribolium castaneum, Nicrophorus vespilloides, Anoplophora glabripennis	XP_008194455, XP_017785110.1, XP_018568194.1
Hymenoptera	Camponotus floridanus, Dufourea novaeangliae, Fopius arisanus, Apis mellifera	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 ( $\Phi$ ) psi ( $\Psi$ ) 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 ( $\Phi$ ) y psi ( $\Psi$ ) 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 (<u>http://mordred.bioc.cam.ac.uk/~rapper/rampage.php</u>, 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 (<u>http://raptorx.uchicago.edu/</u>, 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 variables la herramienta Evolutionary conservados V con 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 estuvó 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* (IKSC; 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 arrojo 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 arrojo 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.

ΤΑΧΑ	ESPECIES PERTENECIENTES AL GRUPO	NO. DE ACCESO DEL NCBI	REFERENCIAS
CNIDARIA	Exaiptasia pallida, Stylophora pistillata, Orbicella faveolata	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	Ampullaria crossean, Haliotis discus discus, Haliotis tuberculata, Haliotis gigantea, Haliotis discus hannai, Haliotis discus, Bellamya sp. UM,2014, Aplysia kurodai, Pomacea canaliculata, Aplysia californica, Lottia gigantea, Biomphalaria glabrata, Octopus bimaculoides, Crassostrea gigas, Crassostrea virginica, Mizuhopecten yessoensis, Achatina fulica, Corbicula japonica	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, PVD35551.1, PVD35385.1, PVD35551.1, PVD36624.1, PVD36625.1, PVD36624.1, PVD36625.1, PVD36896.1, XP_005101787.1, XP_005101789.1, XP_005111868.2, XP_005111869.2, XP_005111870.1, XP_012935745.1, XP_012938497.1, XP_012941225.1, XP_012945520.1, XP_009049791.1, XP_009049836.1, XP_009049731.1,	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

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

		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_013066281.1, XP_013066370.1, XP_013066371.1, XP_0130691521.1, XP_013069170.1, XP_013069521.1, XP_013069523.1, XP_013084279.1, XP_013084279.1, XP_013084285.1, XP_013084283.1, XP_013084279.1, XP_013084285.1, XP_013084287.1, XP_013084285.1, XP_013084287.1, XP_0130846694.1, EKC17289.1, EKC39023.1, XP_011421782.1, XP_014786694.1, FC17289.1, EKC39023.1, XP_011430172.1, XP_02336001.1, XP_02336153.1, XP_022336001.1, XP_02336153.1, XP_022341787.1, OWF45593.1, OWF56607.1, XP_021363491.1, XP_021363490.1, XP_021363495.1, XP_021363490.1, XP_021363498.1, ATS91249.1 BAF38757.1	
ANNELIDA	Eisenia fetida, Eisenia andrei, Metaphire hilgendorfi, Perinereis aibuhitensis, Perinereis brevicirris	3WC3_A, XP_001810693.2, BAM14716.1, BAT21145.1, AAX92641.1, ACE75510.1, ACE75511.1, BAH22180.1, ANR02619.1, BAK20401.1	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	Adineta ricciae	AKC91349.1	Szydlowski et al. 2015
CRUSTACEA	Hyalella azteca, Eurytemora affinis, Cherax quadricarinatus, Parasesarma erythrodactyla, Gecarcoidea natalis, Mictyris platycheles, Coenobita brevimanus, Coenobita rugosus, Daphnia pulex, Daphnia magna, Limnoria quadripunctata, Neomysis intermedia	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, EFX86622.1, KZS03923.1, KZS05722.1, KZS07093.1, KZS16514.1, KZS16516.1, ADB85440.1, ADB85442.1, BAL60587.1	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	Orchesella cincta, Folsomia candida	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	Teleogryllus emma	ABV32557.1	Kim et al. 2008
EMBIOPTERA	Antipaluria urichi	AOV94250.1,AOV94256.1	Shelomi No publicado
PHASMATODEA	Timema cristinae, Aretaon asperrimus, Extatosoma tiaratum, Medauroidea extradentata, Peruphasma schultei, Ramulus artemis, Sipyloidea sipylus	AMH40392.1, AMH40395.1, AMH40356.1,AMH40391.1	Shelomi et al. 2016
BLABEROIDEA	Blattella germanica, Panesthia cribrata	PSN31180.1, AAF80584.1, AAF80585 1	Lo et al. 2000, Harrison et al. 2018
TERMITOIDAE	Cryptotermes secundus, Coptotermes acinaciformis, Reticulitermes flavipes, Coptotermes formosanus,	XP_023704929.1, XP_023704930.1, XP_023716596.1, XP_023718476.1, AAK12339.1, AAU20853.1, ADB12483.1, BAB40693.1,	Watanabe et al. 1998, Tokuda et al. 1999, Cairo et al. 2013, Nakashima et al. 2002, Li et al. 2003,

	Odontotermes formosanus, Macrotermes barneyi, Amitermes dentatus, Amitermes foreli, Anoplotermes schwarzi, Constrictotermes cavifrons, Constrictotermes guantanamensis, Grigiotermes hageni, Globitermes sulphureus, Hospitalitermes bicolor, Microcerotermes crassus, Microtermes pallidus, Macrotermes carbonarius, Macrotermes gilvus, Macrotermes gilvus, Macrotermes subhyalinus, Nasutitermes corniger, Nasutitermes sp. NpEG, Odontotermes hainanensis, Pericapritermes sp. PpEG1, Pericapritermes sp. PpEG2, Pericapritermes sp. PpEG2, Pericapritermes ballbinasus, Subulitermes baileyi, Syntermes grandis, Sphaerotermes sphaerothorax, Coptotermes sphaerothorax, Coptotermes gestroi, Microtermes pakistanicus, Reticulitermes speratus, Nasutitermes takasagoensis, Nasutitermes walkeri, Mastotermes darwiniensis, Zootermopsis nevadensis	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	Acyrthosiphon pisum, Myzus	XP_001944774.2, XP_008181842.1, XP_	_022167267.1,
PHTHIRAPTERA	Pediculus humanus corporis	XP_002426465.1 XP_002426465.1	Kirkness et al. No publicado
COLEOPTERA	Tribolium castaneum, Nicrophorus vespilloides, Agrilus planipennis, Anoplophora alabripennis	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	Apis mellifera, Camponotus floridanus, Acromyrmex echinatior, Lasius niger, Habropoda laboriosa, Melipona quadrifasciata, Atta colombica, Cyphomyrmex costatus, Trachymyrmex cornetzi, 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	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_018396635.1, KYN25700.1, XP_018347393.1, KYQ52831.1, XP_018307051.1, KZC10616.1, XP_015432538.1, OAD61293.1, XP_001606454.1, XP_003402778.1, XP_0011606454.1, XP_00140661535.1, XP_011164235.1, XP_011296914.1, XP_011296916.1, XP_011632703.1, XP_011296915.1, XP_011867251.1, XP_012273697.1, XP_012526654.1, XP_015126097.1, XP_015592927.1, XP_016907626.1, XP_0289301.1, XP_020289297.1, XP_02289301.1, XP_0202892936.1	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 órden 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.

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 Aminoácid o donador de protones; Aminoácid o original Ac. Glutámico (E)
PFX25363.1	Stylophora pistillata	Cnidaria	E		
BAU45388.1	Bellamya sp.	Molllusca			D
XP_013066277.1	Biomphalaria glabrata	Molllusca			R
XP_013084279.1	Biomphalaria glabrata	Molllusca	V	G	
XP_013084281.1	Biomphalaria glabrata	Molllusca	V	G	
XP_013084282.1	Biomphalaria	Molllusca	L	A	D

Tabla 5. Secuen	ncias de celulasas co	n mutaciones o	deleciones en	los sitios catalíticos.

	glabrata				
XP_013084283.1	Biomphalaria glabrata	Molllusca	L	А	D
XP_013084285.1	Biomphalaria glabrata	Molllusca			D
XP_013084287.1	Biomphalaria glabrata	Molllusca			D
XP_011430172.1	Crassostrea gigas	Molllusca	Ν		Н
XP_022336001.1	Crassostrea virginica	Molllusca	Ν		Н
XP_022338491.1	Crassostrea virginica	Molllusca	Ν		Н
PVD35384.1	Pomacea canaliculata	Molllusca		G	G
PVD35385.1	Pomacea canaliculata	Molllusca		G	D
PVD36625.1	Pomacea canaliculata	Molllusca	Deleción de 9 aminoácidos		
XP_023718476.1	Cryptotermes secundus	Termitodae			I
ADB82658.1	Odontotermes formosanus	Termitodae			V
XP_021935277.1	Zootermopsis nevadensis	Termitodae			Μ
PSN31180.1	Blattella germanica	Blaberoidea	Deleción de 10 aminoácidos		
EFN70196.1	Camponotus floridanus	Hymenoptera		N	G
EFN70197.1	Camponotus floridanus	Hymenoptera		Ν	G
XP_011253653.1	Camponotus floridanus	Hymenoptera		N	G
XP_011253654.1	Camponotus floridanus	Hymenoptera		N	S
XP_011253655.1	Camponotus floridanus	Hymenoptera		N	G
XP_019884179.1	Camponotus floridanus	Hymenoptera		Ν	N
XP_011346866.1	Ooceraea biroi	Hymenoptera	N		D
XP_020289295.1	Pseudomyrmex gracilis	Hymenoptera	N	N 	N
XP_020289297.1	Pseudomyrmex gracilis	Hymenoptera	N	Г 	ĸ
XP_020289301.1	Pseudomyrmex gracilis	Hymenoptera	Y	S	K
XP_020289336.1	Pseudomyrmex gracilis	Hymenoptera	N	Ν	

# 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 (Artropoda: Hexápoda) y el crustáceo *Daphnia magna* (100/51), y el segundo clado (61/17) está conformado por secuencias por el grupo Pancrustacea (Artropoda) que incluye las secuencias de celulasas 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 celulasas 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

43

#### 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  $\leq 0.5$  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 grafico de Ramachandran aplicado a los modelos sin
refinar, refinados con método ligero y refinados con método fuerte.

	Modelos sin refinar		Modelos refinados con método ligero			Modelos refinados con método fuerte				
	Porcentaje de aminoácidos en zonas			Porcer	Porcentaje de aminoácidos en zonas			Porcentaje de aminoácidos en zonas		
No de acceso del NCBI	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  $\beta$ -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 catalitico), 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 RSMD obtenidos indican que las diferencias estructurales entre los 49

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 Dendograma 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 evolucionary 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).



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 melífera (B), Antipaluria urichi (C, D), Timema cristinae (E, F), Zootermopsis nevadensis (G, H). El color rojo representa hidrofobicidad, mientas que el cómo 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 (<u>https://www.ncbi.nlm.nih.gov/</u>)

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 Lophochozotra y Pancrustacea + Rotífera).

Dentro del grupo Lophotrochozoa se forman dos clados, conformados por taxa Mollusca y Anellida, 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* (Rotífera) 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 Rotífera (una secuencia). Esta relación posiblemente errona, puede deberse a una baja representación del filo Lophotrochozoa (Sanderson & Shaffer 2002).

El clado Pancrustacea (que incluye la secuencia de *Adineta riccinae*) 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.

60

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, Jermiin 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 eumetazorios 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 celulasas de organismos evolutivo ancestral para las 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 (Rotifera, Annelida, Mollusca, Crustacea y Hexápoda) utilizan las celulasas de la familia GH9 para degradar la celulosa de zooxante de secomposición (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 filogéneticos 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 subfuncionalizacion 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 c y βlaminas 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, Shahriarinour 2011, Chen et al. 2018). La estabilidad es una propiedad fisioquímica de las enzimas que determina si una enzima puede llevar a cabo una reacción enzimática en diferentes condiciones, ya que la incoporación de iones metálicos puede jugar un papel fundamental en la estructura terciaria de la enzima y/o interactuar con el substrato (Mandels et al. 1957, Juy et al. 1992, Shahriarinour 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/hidrofilicas (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. <u>https://doi.org/10.1016/j.softx.2015.06.001</u>

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, Hélène 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

Abrusán, G., & Marsh, J. A. (2016). Alpha Helices Are More Robust to Mutations than Beta Strands. PLoS Computational Biology, 12(12), 1–16. <u>https://doi.org/10.1371/journal.pcbi.1005242</u>

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. <u>https://doi.org/10.1080/09168451.2015.1075860</u>

Barletta, G. P., & Fernandez-Alberti, S. (2018). Protein Fluctuations and Cavity Changes Relationship. Journal of Chemical Theory and Computation, 14(2), 998–1008. <u>https://doi.org/10.1021/acs.jctc.7b00744</u>

Bayer, E. A., Chanzy, H., Lamed, R., & Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. Current Opinion in Structural Biology, 8(5), 548–557. <u>https://doi.org/10.1016/S0959-440X(98)80143-7</u>

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. <u>https://doi.org/10.1073/pnas.181342398</u>

Bucher. (2008). The genome of the model beetle and pest Tribolium castaneum. Nature, 452(7190), 949–955. <u>https://doi.org/10.1038/nature06784</u>

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

Beguin, P. and Aubert, J. (1994) The biological degration 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., Künzli, 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. <u>https://doi.org/10.1073/pnas.77.11.6678</u>

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

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. <u>https://doi.org/(...)0010-4655(95)00042-e</u>

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

Boudet, a M., Lapierre, C., & Grimapettenati, J. (1995). Tansley Review No-80 - Biochemistry and Molecular-Biology of Lignification. New Phytologist, 129(2), 203–236. <u>https://doi.org/DOI 10.1111/j.1469-8137.1995.tb04292</u>.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  $\beta$ -1,3-glucanase active in the alkaline midgut of Spodoptera frugiperda larvae and its relation to  $\beta$ -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/PDB4D0D/PDB

Bui, T. H. H., & Lee, S. Y. (2014). Endogenous cellulase production in the leaf litter foraging mangrove crab Parasesarma erythodactyla. 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

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. <u>https://doi.org/10.1111/j.1365-2583.2010.00991.x</u>

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. <u>https://doi.org/10.1146/annurev-ecolsys-110411-160312</u>

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. <u>https://doi.org/10.1038/jid.2013.1.Research</u>

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. <u>https://doi.org/10.1111/j.1749-7345.1989.tb00560.x</u>

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. <u>https://doi.org/10.1016/j.ibmb.2012.05.002</u>

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 CeID.

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. <u>https://doi.org/10.1186/s13068-018-1055-6</u>

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. <a href="https://doi.org/10.1126/science.1197761">https://doi.org/10.1126/science.1197761</a>.

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. <u>https://doi.org/10.2210/PDB4CJ0/PDB</u>

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. <u>https://doi.org/10.1038/nrm1746</u>

Crawford, A. C., Kricker, 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. <u>https://doi.org/10.1016/j.gene.2004.06.060</u>

Crippen, V. N. M. & G. M. (1994). Siginificance of Root-Mean-Square Deviation in Comparing Three-dimnsional Structures of Globular Propteins. 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. <u>https://doi.org/10.1073/pnas.1008486107</u>

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

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. <u>https://doi.org/10.1002/anie.200352776</u>

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. <u>https://doi.org/10.1093/gbe/evw134</u>

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. <u>https://doi.org/10.1534/genetics.105.053249</u>

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. <u>https://doi.org/10.1016/j.ibmb.2013.07.007</u>

Gaskell, A., Crennell, S., & Taylor, G. (1995). The three domains of a bacterial sialidase: a  $\beta$ -propeller, an immunoglobulin module and a galactose-binding jelly-roll. Structure, 3(11), 1197–1205. <u>https://doi.org/10.1016/S0969-2126(01)00255-6</u>

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. <u>https://doi.org/10.1111/j.1432-1033.1971.tb01589.x</u>

Gilbert, H. J. (2010). The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiology, 153(2), 444–455. <u>https://doi.org/10.1104/pp.110.156646</u>

Gora, A., Brezovsky, J., & Damborsky, J. (2013). Gates of Enzymes. Chemical Reviews, 113(8), 5871–5923. <u>https://doi.org/10.1021/cr300384w</u>

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. <u>https://doi.org/10.1093/bioinformatics/btl461</u>

Gray, M., Linton, S. M., & Allardyce, B. J. (2018). cDNA sequences of GHF9 endo-β-1,4glucanases in terrestrial Crustacea. Gene, 642(November 2017), 408–422. <u>https://doi.org/10.1016/j.gene.2017.11.030</u>

Gromiha, M. M., Pujadas, G., Magyar, C., Selvaraj, S., & Simon, I. (2004). Locating the Stabilizing Residues in ( $\alpha/\beta$ )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. <u>https://doi.org/10.1105/tpc.13.1.73</u>

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., & Guillénn, R. (1995). Composition of plant cell walls. Zeitschrift Für Lebensmittel-Untersuchung Und -Forschung, 200(1), 24–31. <u>https://doi.org/10.1007/BF01192903</u>

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. <u>https://doi.org/10.1021/ct700301q</u>

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. <u>https://doi.org/10.2210/PDB5DGQ/PDB</u>

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. <u>https://doi.org/10.1186/1471-2105-11-24</u>

Hughes, A. L. (2012). Evolution of the βgRP/GNBP/β-1,3-glucanase family of insects. Immunogenetics, 64(7), 549–558. <u>https://doi.org/10.1007/s00251-012-0610-8</u>

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. <u>https://doi.org/10.1093/protein/13.12.839</u>

Jermiin, L. S., Poladian, L., & Charleston, M. A. (2005). Is the "big bang" in animal evolution real? Science, 310(5756), 1910–1911. <u>https://doi.org/10.1126/science.1122440</u>

Jondelius, U., Ruiz-Trillo, I., Baguñà, 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., Claeyssens, 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. <u>https://doi.org/10.1073/pnas.0914228107</u>

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 Celt. 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. <u>https://doi.org/10.1107/S0907444902002366</u>

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., Thei??en, 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. <u>https://doi.org/10.1016/j.ibmb.2014.06.008</u>

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. <u>https://doi.org/10.1186/1471-2164-13-587</u>

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. <u>https://doi.org/10.1039/c6sc05381e</u>

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. <u>https://doi.org/10.1016/j.cbpc.2004.06.015</u>

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. <u>https://doi.org/10.1016/j.cbpc.2004.12.003</u>

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. <u>https://doi.org/10.1038/ncomms3485</u>

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. <u>https://doi.org/10.1186/1471-2164-13-271</u>

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

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. <u>https://doi.org/10.1128/ec.2.5.1091-</u>

#### 1098.2003

Li, Y., Yin, Q., Ding, M., & Zhao, F. (2009). Purification, characterization and molecular cloning of a novel endo- $\beta$ -1,4-glucanase AC-EG65 from the mollusc Ampullaria crossean. 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. <u>https://doi.org/10.1126/science.290.5494.1151</u>

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. <u>https://doi.org/10.1016/S0960-9822(00)00561-3</u>

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. <u>https://doi.org/10.1098/rsbl.2003.0016</u>

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:  $\phi$ , $\psi$  and C $\beta$  deviation. Proteins: Structure, Function and Genetics, 50(3), 437–450. <u>https://doi.org/10.1002/prot.10286</u>

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. <u>https://doi.org/10.1098/rstb.1991.0078</u>

Maiorov V. N. & Crippen G. M. (1994). Siginificance of Root-Mean-Square Deviation in Comparing Three-dimnsional Structures of Globular Propteins. 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 cellulosedigesting system of the wood-feeding termite, Coptotermes formosanus Shiraki. Insect Biochemistry and Molecular Biology, 32(7), 777–784. <u>https://doi.org/10.1016/S0965-1748(01)00160-6</u>

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. <u>https://doi.org/10.1002/arch.21158</u>

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. <u>https://doi.org/10.2210/PDB3X17/PDB</u>

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. <u>https://doi.org/10.1007/978-3-319-15976-8\_1</u>

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. <u>https://doi.org/10.1098/rspb.2008.1574</u>

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. <u>https://doi.org/10.2210/PDB1IA6/PDB</u>

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. <u>https://doi.org/10.1016/j.ibmb.2009.04.001</u>

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. <u>https://doi.org/10.2210/PDB3EZ8/PDB</u>

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. <u>https://doi.org/10.2210/PDB2XFG/PDB</u>

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. <u>https://doi.org/10.1016/j.cub.2015.10.015</u>

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. <u>https://doi.org/10.1146/annurev-arplant-042110-103809</u>

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. <u>https://doi.org/10.1016/0377-8401(91)90013-I</u>

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: <a href="http://tree.bio.ed.ac.uk/software/figtree/">http://tree.bio.ed.ac.uk/software/figtree/</a>

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. <u>https://doi.org/10.1098/rspb.2004.2917</u>

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

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

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

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: <u>http://www.redalyc.org/articulo.oa?id=47543418</u>

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. <u>https://doi.org/10.1093/jxb/erp245</u>

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. <u>https://doi.org/10.1111/j.1365-2583.2004.00527.x</u>

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. <u>https://doi.org/10.2210/PDB1RQ5/PDB</u>

Secundo, F. (2013). Conformational changes of enzymes upon immobilisation. Chemical Society Reviews, 42(15), 6250–6261. <u>https://doi.org/10.1039/c3cs35495d</u>

Shahriarinour, 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. <u>https://doi.org/10.5897/AJB11.199</u>

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. <u>https://doi.org/10.1016/j.ibmb.2016.02.003</u>

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. <u>https://doi.org/10.1186/1471-2164-15-917</u>

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, Otillar 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

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

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

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 , Psacothea hilaris, 3460(May), 3455–3460. <u>https://doi.org/10.1046/j.1432-1033.2003.03735.x</u>

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. <u>https://doi.org/10.1016/j.gene.2015.04.007</u>

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

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

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

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., Vilcinskas 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

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. <u>https://doi.org/10.1146/annurev-ento-120710-100538</u>

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 Bellamya chinensis laeta. Molecular Biotechnology, 58(4), 241–250. <u>https://doi.org/10.1007/s12033-016-9922-5</u>

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. <u>https://doi.org/10.1038/srep01448</u>

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. <u>https://doi.org/10.1146/annurev-ento-112408-085319</u>

Watanabe, H., Noda, H., Tokuda, G., & Lo, N. (1998). A cellulase gene of termite origin. Nature, 394(6691), 330–331. <u>https://doi.org/10.1038/28527</u>

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. <u>https://doi.org/10.1002/arch.21092</u>

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 GHF9 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. <u>https://doi.org/10.1111/j.1744-7917.2010.01322.x</u>

Wilson, D., Irwin, D., Tsao, G., Brainard, A., Bungay, H., Cao, N., ... Xia, L. (1999). Genetics and Properties of Cellulases\rRecent Progress in Bioconversion of Lignocellulosics, 65, 1–21. <u>https://doi.org/10.1007/3-540-49</u>

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. <u>https://doi.org/10.1021/ar030250v194-5\_1</u>

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. <u>https://doi.org/10.1093/bioinformatics/btq066</u>

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. <u>https://doi.org/10.1074/jbc.M208443200</u>

Yokoe, Y., & Yasumasu, I. (1964). the Distribution of Cellulase in Invertebrates. Comparative Biochemistry and Physiology, 13, 323–338. <u>https://doi.org/10.1016/0010-406X(64)90027-1</u>

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. <u>https://doi.org/10.7717/peerj.4087</u>

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

Zhao, F., Ding, M., Li, Y., Yin, Q., & Teng, Y. (2011). Expression and Characterization of Full-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

Length Ampullaria crossean Endoglucanase EG65s and Their Two Functional Modules . Bioscience, Biotechnology, and Biochemistry, 75(2), 240–246. <u>https://doi.org/10.1271/bbb.100529</u>

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. <u>https://doi.org/10.1016/j.gene.2007.01.004</u>

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. <u>https://doi.org/10.1016/j.gene.2007.01.004</u>

## Anexo 1

Tabla 1.- Secuencias utilizadas en el alineamiento 1.

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

	subhyalinus, Nasutitermes corniger, Nasutitermes sp. NpEG, Odontotermes hainanensis, Pericapritermes nitobei, Pericapritermes sp. PpEG1, Pericapritermes sp. PpEG2, Pericapritermes sp. PpEG3, Rhynchotermes bulbinasus.	AGP76430.1, AGP76431.1, AGP76434.1, AGS32241.1, AKV16371.1, BAA31326.1, BAA33708.1, BAA33709.1, CAD54726.1, KDR16731.1, XP_023704929.1
	Subulitermes baileyi, Syntermes grandis, Sphaerotermes sphaerothorax Coptotermes gestroi, Microtermes pakistanicus, Reticulitermes speratus, Nasutitermes takasagoensis, Nasutitermes walkeri, Mastotermes darwiniensis, Zootermopsis nevadensis Cryptotermes secundus	ç, Ş,
Hemiptera	Acyrthosiphon pisum, Myzus persicae, Nilaparvata lugens	XP_001944774.2, XP_008181842.1, XP_022167267.1, XP_022167268.1, XP_022200004.1
Phthiraptera	Pediculus humanus corporis	XP_002426465.1
Coleoptera	Eurytemora affinis, Tribolium castaneum, Nicrophorus vespilloides, Agrilus planipennis, Anoplophora glabripennis	XP_023337121.1, KYB27092.1, XP_017785110.1, XP_018323591.1, XP_018568194.1
Hymenoptera	Camponotus floridanus, Acromyrmex echinatior, 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, Apis mellifera	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_011296914.1, XP_011632703.1, XP_011498432.1, XP_011632703.1, XP_011700378.1, XP_011632703.1, 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)

Crune de		
organimos	especies pertenecientes al grupo	NO. DE ACCESO dEI NCBI
Collembola	Orchesella cincta, Folsomia candida	ODM94862.1- ODM95900.1- XP_021956548.1- XP_021956549.1- XP_021959948.1- XP_021966378.1- XP_021967587.1
Orthoptera	Teleogryllus emma	ABV32557.1
Embioptera	Antipaluria urichi	AOV94250.1-AOV94256.1
Phasmatodea	Timema cristinae, Aretaon asperrimus, Extatosoma tiaratum, Medauroidea extradentata, Peruphasma schultei, Ramulus artemis, Sipyloidea sipylus	AMH40392.1- AMH40395.1- AMH40356.1-AMH40391.1
Blaberoidea	Blattella germanica, Panesthia cribrata	PSN31180.1- AAF80584.1- AAF80585.1
Termitoidae	Cryptotermes secundus, Coptotermes acinaciformis, Reticulitermes flavipes, Coptotermes formosanus, Odontotermes formosanus, Macrotermes barneyi, Amitermes dentatus, Amitermes foreli, Anoplotermes barneyi, Amitermes dentatus, Amitermes foreli, Anoplotermes schwarzi, Constrictotermes cavifrons, Constrictotermes guantanamensis, Grigiotermes hageni, Globitermes sulphureus, Hospitalitermes bicolor, Microcerotermes crassus, Microtermes pallidus, Macrotermes carbonarius, Macrotermes gilvus, Macrotermes subhyalinus, Nasutitermes corniger, Nasutitermes sp. NpEG, Odontotermes hainanensis, Pericapritermes sp. NpEG, Odontotermes sp. PpEG1, Pericapritermes sp. PpEG2, Pericapritermes sp. PpEG3, Rhynchotermes bulbinasus, Subulitermes baileyi, Syntermes grandis, Sphaerotermes sphaerothorax, Coptotermes gestroi, Microtermes pakistanicus, Reticulitermes takasagoensis, Nasutitermes valkeri, Mastotermes darwiniensis, Zootermopsis nevadensis	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- AGP76420.1-AGP76418.1- AGP76420.1-AGP76437.1- 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
ויפווויטנפומ	persicae, Nilaparvata lugens	XP_022167267.1, XP_022167268.1, XP_02200004 1
Phthiraptera	Pediculus humanus corporis	XP_002426465.1
Coleoptera	Tribolium castaneum. Nicrophorus	
- sicopiaia		

Tabla 2.- Secuencias utilizadas en el alineamiento 2.

	vespilloides, Agrilus planipennis,	XP_008194455.1- XP_015836219.1-
	Anoplophora glabripennis	XP 017785110.1- XP 018323591.1-
		XP_018568194.1- XP_018568196.1
Hymenoptera	Apis mellifera, Camponotus	XP_396791.3- EFN70196.1-
	floridanus, Acromyrmex echinatior,	EFN70197.1- XP_011253653.1-
	Lasius niger, Habropoda laboriosa,	XP_011253654.2- XP_011253655.2-
	Melipona quadrifasciata, Atta	XP_019884179.2- EGI63652.1-
	colombica, Cyphomyrmex	XP_011057333.1- KMQ95413.1-
	costatus, Trachymyrmex cornetzi,	KOC62382.1- XP_017793200.1-
	Trachymyrmex septentrionalis,	KOX76353.1- KYM87195.1-
	Trachymyrmex zeteki, Dufourea	XP_018044452.1- KYN01696.1-
	novaeangliae, Eufriesea mexicana,	XP_018396635.1- KYN22851.1-
	Apis cerana cerana, Nasonia	XP_018359268.1- KYN35700.1-
	vitripennis, Bombus terrestris, Apis	XP_018347393.1- KYQ52831.1-
	florea, Apis dorsata, Solenopsis	XP_018307051.1- KZC10616.1-
	invicta, Fopius arisanus, Ooceraea	XP_015432538.1- OAD61293.1-
	biroi, Ceratosolen solmsi marchali,	XP_017766299.1- PBC31793.1-
	Pogonomyrmex barbatus,	XP_001606454.1- XP_003402778.1-
	Wasmannia auropunctata,	XP_003690676.1- XP_006610535.1-
	Vollenhovia emeryi, Atta	XP_011164235.1- XP_011296914.1-
	cephalotes, Megachile rotundata,	XP_011296916.1- XP_011346866.1-
	Linepithema humile, Bombus	XP_011498432.1- XP_011632703.1-
	impatiens, Orussus abietinus,	XP_011700378.1- XP_011867251.1-
	Monomorium pharaonis,	XP_012061951.1- XP_012146513.1-
	Copidosoma floridanum, Polistes	XP_012224814.1- XP_012248155.1-
	canadensis, Diachasma alloeum,	XP_012273697.1- XP_012526654.1-
	Polistes dominula, Neodiprion	XP_014214612.1- XP_014601854.1-
	lecontei, Cephus cinctus, Apis	XP_015126097.1- XP_015184471.1-
	cerana, Pseudomyrmex gracilis	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.

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

Tabla 3.- Matriz de valores de RMSD

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.
Tipo de sitios So de unión	ecuencias
Sin sitios de K	ZS03923.1, XP_021967587.1, XP_008194455.1, XP_015836219.1, XP_001810693.2,
unión X	P_017785110.1, XP_001944774.2, XP_022167268.1, XP_001606454.1,
X	P 011498432.1, XP 014214612.1, XP 003690676.1, XP 016907626.1, OAD61293.1,
X	P 017766299.1, XP 003402778.1, XP 012248155.1, KOC62382.1, KOX76353.1,
X	P 015432538.1, KZC10616.1, XP 011296916.1, XP 012273697.1, XP 011253655.1,
X	P 020289295.1, XP 020289297.1, KYM87195.1, KYN22851.1, KYQ52831.1,
X	P 018307051.1, XP 011164235.1, XP 011700378.1, XP 012526654.1,
X	P 011867251.1, XP 012224814.1, XP 014601854.1, AMH40395.1, XP 002426465.1,
A	OV94250.1, AMH40370.1, AMH40376.1, AMH40383.1, AMH40377.1, AMH40369.1,
A	MH40393.1, AMH40364.1, AMH40368.1, AMH40371.1, AMH40366.1, AMH40373.1,
A	MH40374.1, AMH40363.1, AOV94254.1, AOV94252.1, ABV32557.1, AAF80585.1,
P	SN31180.1, CAD54726.1, XP_023704929.1, XP_021935277.1, XP_023718476.1,
A	GP76420.1, AGP76421.1, AGP76419.1, AGP76412.1, AGP76414.1, AAU20853.2,
B	AB40695.1, AKV16371.1, AAK12339.1, BAA33708.1, BAA33709.1, AGP76406.1,
A	GP76427.1, AGP76404.1, AGP76408.1, AGP76409.1, AGP76410.1, AGP76411.1,
A	.GP76418.1, AGP76424.1, AGP76425.1, AGP76423.1, XP_018016813.1,
X	.P_018016810.1, XP_018027134.1, ADB85442.1, ADB85441.1, XP_018026496.1,
A	.PF29560.1, AAO61672.2, XP_009049840.1, KOF68403.1, XP_014786694.1,
X	.P_009049791.1, XP_021363495.1, XP_021363497.1, ABD24281.1, ABD24280.1,
X	P_013069522.1, XP_009049941.1, XP_009058133.1, BAD01504.1, BAD44734.1,
A	FQ98380.1, AFQ98379.1, XP_013066371.1, XP_013094438.1, XP_005099315.1,
B	AX09271.1, XP_012939931.1, XP_012938497.1, XP_005101789.1, XP_013066278.1,
Х	P_013066280.1, XP_013066281.1, PVD36986.1, BAU45387.1, PVD36896.1,
B	AK20401.1, ACE75511.1, AAX92641.1, ACE75510.1, BAT21145.1, PFX25363.1
Con sitios de El	FX86822.1, XP_021959948.1, XP_021966378.1, ODM94862.1, ODM95900.1,
unión de X	.P_021956548.1, XP_021956549.1, AKC91349.1, KYB27092.1, XP_018323591.1,
calcio y ZINC XI	P_018568194.1, XP_018568196.1, XP_008181842.1, XP_02216/267.1,
X	P_022200004.1, AOV94253.1, XP_015513382.1, XP_396791.3, XP_006610535.1,
PI	BC31793.1, XP_017793200.1, XP_012146513.1, XP_011296914.1, XP_015126097.1,
X	P_015592927.1, EFN70196.1, XP_011253653.1, EFN70197.1, XP_019884179.1,
	P_UT253654.1, XP_U2U289301.1, XP_U2U289336.1, KMQ95413.1, XP_UT346866.1,
E	GI03032.1, AF_UT1037333.1, AF_UT2001931.1, AF_UT0044432.1, ATN33700.1,
	P_016347393.1, K1N01696.1, XP_016396635.1, XP_016359266.1, XP_011632703.1,
	MU40256 1 AMU40299 1 AMU40257 1 AMU40290 1 AMU40267 1 AMU40262 1
AI A	NIT40300.1, ANTA40300.1, ANTA40307.1, ANTA40309.1, ANTA40307.1, ANTA40302.1,
AI A'	MU40302.1, AMU40301.1, AMU40301.1, AMU40300.1, AMU40300.1, AMU40350.1, AMU40379.1,
	MH40380.1, AMH40385.1, AMH40307.1, AMH40360.1, AMH40387.1, AMH40372.1
	MH 40301.1, AMI 140303.1, AMI 140394.1, AMI 140300.1, AMI 140304.1, AMI 140372.1, MH 40370.1 ΔΩ\/04256.1 ΔΩ\/04251.1 ΔΩ\/04255.1 ΔΔΕ80584.1 ΔΔΕ54727.1
	MINIA0373.1, AUV94230.1, AUV94231.1, AUV94233.1, ARI 00304.1, CAD34727.1,
	CP76/3/1 ACP76/35 1 ACP76/36 1 ACP76/37 1 ACP76/13 1 ACP76/16 1
	GP76415.1, ACT 70433.1, ACT 70430.1, ACT 70437.1, ACT 70413.1, ACT 70410.1, GP76415.1 AED33365.1 BAA31326.1 BAA34050.1 BAB40606.1 BAB40607.1
	DB12483 1 AGS32241 1 AGP76407 1 AGP76308 1 AGP76309 1 AGP76307 1
A A	GP76428 1 AGP76429 1 AGP76431 1 AGP76432 1 AGP76433 1 AGP76402 1
	GP76403 1 AGP76417 1 AGP76405 1 AGP76430 1 AGP76401 1 AGP76426 1
Δι	$\Delta \mathbf{G}_{i}$ is a second relation of the relation of the relation $\mathbf{G}_{i}$ and
A'	GP76422.1 XP 023337121.1 BAL60587.1 XP 018016820.1 XP 018016821.1
	GP76422.1, XP_023337121.1, BAL60587.1, XP_018016820.1, XP_018016821.1, P_018016806.1, XP_018016809.1, ADB85440.1, APE29556.1, APE29557.1
A A X A	GP76422.1, XP_023337121.1, BAL60587.1, XP_018016820.1, XP_018016821.1, P_018016806.1, XP_018016809.1, ADB85440.1, APF29556.1, APF29557.1, PF29558.1, APF29559.1, AIT11911.1, APF29561.1, AAD38027.1, EFX69372.1
A A Xi Al K	GP76422.1, XP_023337121.1, BAL60587.1, XP_018016820.1, XP_018016821.1, P_018016806.1, XP_018016809.1, ADB85440.1, APF29556.1, APF29557.1, PF29558.1, APF29559.1, AIT11911.1, APF29561.1, AAD38027.1, EFX69372.1, ZS07093.1, KZS16514.1, KZS16516.1, EFX80605.1, KZS05722.1, XP_022336153.1
A A X Al K O	GP76422.1, XP_023337121.1, BAL60587.1, XP_018016820.1, XP_018016821.1, P_018016806.1, XP_018016809.1, ADB85440.1, APF29556.1, APF29557.1, PF29558.1, APF29559.1, AIT11911.1, APF29561.1, AAD38027.1, EFX69372.1, ZS07093.1, KZS16514.1, KZS16516.1, EFX80605.1, KZS05722.1, XP_022336153.1, WF45593.1, XP_021363496.1, XP_021363481.1, XP_022341787.1, XP_009049836.1.

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.

	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	EKC17289.1, XP 005101787.1
unión a zinc	
Con sitio de	CAD54728.1, BAB40693.1, AGP76400.1, EFX80604.1, XP_011421782.1,
unión a	XP 021363498.1, XP 011423159.1, XP 013084279.1, XP 013069170.1, PVD36624.1,
calcio	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	Nasutitermes
		takasagoensis
1	AAF80584.1	Panesthia cribrata
2	ABV32557.1	Teleogryllus emma
3	AGP76427.1	Rhynchotermes
		bulbinasus
4	AKV16371.1	Microtermes
		pakistanicus
5	AMH40372.1	Medauroidea
		extradentata
6	AMH40373.1	Peruphasma
		schultei
7	AMH40379.1	Ramulus artemis
8	AMH40392.1	Timema cristinae
9	AMH40395.1	Timema cristinae
10	AOV94250.1	Antipaluria urichi
11	AOV94253.1	Antipaluria urichi
12	AOV94256.1	Antipaluria urichi
13	BAA31326.1	Reticulitermes
		speratus
14	CAD54726.1	Mastotermes
		darwiniensis
15	EFN70196.1	Camponotus
		floridanus
16	KDR16731.1	Zootermopsis
		nevadensis
17	KZC10616.1	Dufourea

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

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 articulo 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 obtienne 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 informaion 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 proseso de solvatacion 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 Minimizacion 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 proseso de solvatacion 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 dirije 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)

###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\_2018-1013.pdb','/home/mario/Downloads/v3todos/v3fuerte/35XP\_021956548\_2018-1013.pdb')</pre>

###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)</pre>

```
####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 dicujar 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 dicujar 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 dicujar 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 dicujar 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 nocos for l in t.iter\_leaves():

```
###Se obtiene los sitios de union con los que cuneta cada nodo
     \cos a = \lim(\sin(\sin(1)[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 especificas

####Se abre el archivo a leer

```
archivo=open("alineadotodo.fas","r")
```

###Se crea el nuevo archivo en el que se depositaran las secuencias con

diferecnias

```
nuevo=open("cons22.csv","w")
```

l=[]

t=""

###Se lee cada uno de los renglones

for i in archivo.readlines():

###Si es un nombre de secuencas:

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 perteneece
```

```
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()

## Anexo 3

64038037 1	YEAH CHEVYEVE A OPECKI BEDORY TWREDEAL NDEEDVENDL TECYYOU CON	ECERMANTA THI AWOOLDEANCURKACOTRYCHAALKWATRYFI	100
AAD36027.1	YSGAL CMST V F TEAGRSGKL PG DGKVT WRKDSAL NDGSDVGHDL TGGYT DAGDHV	FOFFMAFTATMLAWGUI DFANGHSKAGUI STGHAALKWATDYFL	. 100
AAF80584.1	YAOVLOYSEL FYEADRSGKEP ADOKYTWRKDSAL NDKGUNGEDLTGGYTDAGDY	FGFPMATTATVLAWGVISHEUGTUKANALEDARKAVKWATDYFT	100
AAF60305.1	YSTEL FIEL FIEL ORGERED ADOKYTWEEDSAL DEGONGEDE TOGYTOADDT	FGFPMAFTATELAWSLIDTEUGTAKANSVEDARKAVKWATDTFL	100
AAO81672.2	YSOAL CMSYVEYEAOBSCKL BODOPYTWEGDSAL NDGSDYGUDI TGGYYDAGDUY	EGERMAETATMI AWGO I DEANGUSKAGOTSVGUAAI KWATDYEL	100
AAU20853.2	YKTYLSNSLLFYEADRSGKLPSDOKYTWRKDSALNDKGOKGEDLTGGYYDAGDEV	EGEPMAYTYTYL AWGYLDYESAYSAAGALDSGRKALKYGTDYFL	100
AAX92641.1	YDEVLEKS I LEYEAERS GDLPSNNR LPYRGDSAL GDOGNOGODI TGGWYDAGDHY	EGEPMA FATTLIAWGI LEERDGY FAAGOYNIALDS I RWTI NY FL	100
ABD24274.1	YDDLLYKSILFYEAGRSGKLPANNRIPWRGDSALNDHGNAGEDLTGGWYDAGDFV	FNFPMAWSTAVL TWGLLQFKDAYQAAGQLEWMYESI KWPLDYLL	100
ABD24275.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND - GDGGVDLTGGWYDAGDFV	F N F P M A W S T A I L N W G L L Q F K D A Y E A A G Q L E W M Y E S V K W P L D Y L L	. 99
ABD24276.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND-GDGGVDLTGGWYDAGDFV	F N F P M A W S T A I L N W G L L Q F K D A Y E A A G Q L E W M Y E S V K W P L D Y L L	. 99
ABD24277.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND - GDGGVDLTGGWYDAGDFV	FNFPMAWSTAILNWGLLQFKDAYEAAGQLEWMYESVKWPLDYLL	. 99
ABD24278.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND-GDGGVDLTGGWYDAGDFV	FNFPMAWSTAVLTWGLLQFKDAYQAAGQLEWMYESVKWPLDYLL	. 99
ABD24279.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND - GDGGVDLTGGWYDAGDFV	FNFPMAWSTAVL TWGLLQFKDAYQAAGQLEWMYESVKWPLDYLL	. 99
ABD24280.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND - GDGGVDLTGGWYDAGDFV	F N F P M A W S T A V L T W G L L Q F K D A Y Q A A G Q L E W M Y E S V K W P L D Y L L	. 99
ABD24281.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND - GDGGVDLTGGWYDAGDFV	FNFPMAWSTAVLTWGLLQFKDAYQAAGQLEWMYESVKWPLDYLL	. 99
ABO26609.1	YKEVLGLSILFYDAQRSGKLPANNHVSWRADSSLTDKGDNGEDLTGGWYD	FNLPMASATTILSWGFLKWSDAYKTAGQEDHMYDMIKWPLDYFL	. 100
ABV32557.1	YADVIKKSLLFYQAQRSGRLSGDPLVSWRKDSALNDRGNNGEDLTGGYYDAGDFV	F G F P M A Y T I T L L S W G V I D Y E N T Y S S I G A L S A A R A A I K W G T D Y F I	100
ACE75510.1	YDEVLEKSILFYEAERSGDLPANNRIPYRGDSALGDQGNQGQDLTGGWYDAGDHV	F G F P M A F A T T T L A WG I L E F R D G Y E A A G Q Y N L A L D S I R W T L N Y F L	. 100
ACE75511.1	Y DEVLEKSILFYEAERSGYLPADNRIPYRGDSALGDQGNQGQDLIGGWYDAGDHV	FGFPMAISTTTLAWGILEFREAYEAAGQYNWALDSIRWPLEYFI	100
ADB12483.1	YKTVLKNSLLFYEAQRSGKLPADQKVTWRKDSALNDKGDNGEDLTGGYYDAGDFV	F G F P M A Y T A T V L A WG L V D Y E A G Y S S A G V L D D G R K A V K WA T D Y F L	. 100
ADB82658.1	YKTVLRNSLLFYEAQRSGKLPADQKVTWRKDSALNDKGQNGEDLTGGYYDAGDYV	F G F P M A S T A T V L A W G L V D Y A A G Y T S A G A L D D G R K A V K W A T D Y F L	. 100
ADB85440.1	YSQALCMSPLFYEAGRSGKLPADQRVTWRADSGLQDGADVGLDLTGGYYD	FNFPMAASVTALAFGLHAFPSGMSTAGQTNYVKAAVKWGSDYFL	. 100
ADB85441.1	Y SQAL CMAML FY EAORS GY L PADNR V DWRGDS AT GDGS D V GLDLEGGY Y DAGD F V	F G F P L A Y S V A V L A WG L D E Y S G G Y R T A G Q T A Y G R A T I K R A T D Y L L	. 100
ADB85442.1	YAQALCMSPLFYEAQRSGKLPADQRVTWRKDSALDDGSDCGIDLEGGYYDAGDYV	F G Y P F A S AM T A L A Y G L L A Y A D G Y E A A G Q T E Y G K Q P L K W G T D Y M L	. 100
AFD33365.1	YKTVLSNSLLFYEAQRSGKLPADQKVTWRKDSALNDKGQNGEDLTGGYYD	F G F P M A S T A T I L A W G L V D Y E A G Y S S A G A L E D G R K A L K W A T D Y F L	. 100
AFQ98379.1	YGESLAKSILFYDAORSGKLPANNPIPWRGDSALGDKGDNGEDLTGGWYDAGDHV	F S L P M S S T S T V L L W G Y L Q W K D A Y A T M K Q T D M F F D M I K W P L D Y Y L	. 100
AFQ98380.1	Y G E A L G K S I L F Y D A Q R S G K L P A N N P I K W R G D S A L G D K G D N G E D L T G G W Y D A G D H V	FSLPMS STSTVLLWGYLQWKDA YATTKQTDMFFDMI KWPLDYFL	100
AGI61069.1	Y GEALGKSILFY DAGRSGKLPANNPI KWRGDSALGDKGDNGEDLTGGWYDOGDHV	FSLPMSSTSTVLLWGYLQWKDAYATTKQTDMFFDMIKWPLDYFL	. 100
AGP76397.1	YKOGLRDSLLFYEAORSGKLPADSKVTWRKDSALNDEGEQGODLTGGYFDAGDYV	F G F P M A A T T T V L A W G L I D V I A G Y S S A G A L D D G R K A V K W A T D Y F I	100
AGP76398.1	YKOVLRHSLLFYEAQRSGRLPODQKVTWRKDSALNDQGEQGODLTGGYFDAGDFV	CEGEP MAATATVLAWGLIDEEAGYSSAGALDDGRKAVKWATDYFT	100
AGP76399.1	YKVVLRDSLLFYEARSGRLF0DDQKVTWRKDSALNDEGE0G0DLTGGYFDAGDFV	FGFPTAATATVLAWGLRDFEAGTSSAGALDDGRRAVRWATDYFT	100
AGP76400.1	TKKGGRDSILFTEAGRSPRLPSDGKVTWRKDSALNDKGENGEDLTGGTFDAGDFV	FGFPMAYTTTVLAWGLVDFEGAYSSAGALDDGRKAVKWATDYLL	. 100
AGP10401.1	TAGUEDSIL EVELOBE OBLODVY WERDSAL NDROOCODI TOCYVDADDEV	CERMANTOTIL AWOULD VAACYSSAGAL DECRKAVKKATDIEL	100
AGP 16402.1		FOFFMATIGTVLAWGEIDTAAGYERAGALDBORKAVKWATDTFT	100
AGP10403.1	YTOVI BDSLLFY EAOBSCRI BDOKYTWEKDSALNDBGOOGODI TGGYYDAGDEY	COPPMANTS TVLAWGLIDYAAGYSSAGALDDGRKAVKWATDYFI	100
ACD76405.1	VENUE BESCH FERNOR CHER DOK VTWERDS AL NDROEGOEDI TOOVYDER	CEDMANTTAIVEAUGEIDIARGISSAGAEDDORKAVKWATDII	100
AGP76408.1	YKOOL BEST LEVEDORS BELBODOK VTWRKDSALNDOGEOGODI TGOYED GOOD	EGERMAATATVI AWGI MDEEAGYSSAGAL DDGRKAVKWATDYEL	100
AGP76407.1	YKUVI BESLI EVA AOBSODI BADEKVTWEKDSAL NDOGOGODI TOGYEDIGOV	EGERMAATTTVLAWGI MDEEAGYSRAGAL DDGRKAVKWATDYES	100
ACP76409.1	TTOVI BOST I EVELOPS CRI BODOK VTWRKDS AL NDOGOGODI TOCKED CON	EGERMANTATVI AWGI NDEAAGYSSAGAI DDGRKAVKWATDYFI	100
AGP76400.1	YTEVL RDSLL EVOADRSCRL PODOKYTWRKDSAL NDOGOOGODI TGGYEDIGDYY	EGERMAYTATVI AWGI NDYAAGYSSAGAL DDGRKAVKWATDYFI	100
AGP76410.1	YTTVL RDSLI EYFAORSORI PPDOKYTWRKDSAL NDOGOOGODI TOGYFDIODEV	EGEPMASTATYLAWGI NDEAAGYSSAGATDDGRKAVKWATDYFA	100
AGP76411.1	YKKVICDSLI FYSSSSSGRI PTDYKVTWRKDSAL NDBGOHGODI TGGYI DAGDDV	EGEPGAYTATVI AWGI VDI EAGYSSAGARDDGRKAVKWATDYFI	100
AGP76412.1	YKNGL SNTLL FY FADRS GKL PY DOKY TWRK DSAL NDOGON GODI TGGY YDAGDEY	EGEPMAYTATVI AWGLVDYEAGYTSAGALDDGRKAVKWATDYEL	100
AGP76413.1	YKNGLSNTLLFYAAORSGKLPDDHKYTWRKDSALNDOGONGODLTGGYYDAGDFV	F G F P M A Y T A T V L A W G L V D Y E A G Y T S A G A L D D G R K A V K W A T D Y F T	100
AGP76414.1	YTTVLSNSLLFYEAGRSGKLPADQKVTWRKDSALNDKGQNGRDLTGGYYDAGDFV	F G F P M A S T A T V L A W G L V D Q E A G Y T S A G A L D D G R K A V K W A T D Y F L	100
AGP76415.1	YTTVLSNSLLFYEAORSGKLPADOKYTWRKDSALNDKGONGRDLTGGYYDDGDLV	F G F P M A S T A T V L A W G L V D H E A G Y D S A G A L D D G R K A V K W A T D Y F L	100
AGP76416.1	YTTVLSNSLLFYEAQRSGKLPADQKVTWRKDSALNDKGQNGVDLTGGYYDAGDFV	F G F P M A S T A T V L A W G L V D Y E A G Y T S A G A L D D G R K A V K W A T D Y F I	100
AGP76417.1	YKKGQRDSLLFYEAQRSGRLPADQKVTWRKDSALNDRGEQGODLTGGYFDAGDFV	F G F P M A Y T A T V L A W G M I D L E A G Y T S A G A Q D D G R K A V K W A T D Y C L	. 100
AGP76418.1	YKKVHRDSLLFYEAQRSGRLPADQKVTWRKDSALNDRGEQGQDLTGGYFD	F G F P M A Y T A T V L A W G M I D L E A G Y S S A G A L D D G R K A V K W A T D Y F L	. 100
AGP76419.1	YKNGLDLTLLFYAAHRSGKLPSDQKVTWRKDSALNDRGQNGVDLTGGYYDAGDYV	F G F P M A S T A T V L A WG L V D Y E A G Y T S A G A L D D G R K A V K WA T D Y F I	100
AGP76420.1	YKNGLDLTLLFYAAQRSGKLPDDQKVTWRKDSALNDKGQNGRDLTGGYYDAGDYV	F G F P M A S T A T V L A W G L V D H E A G Y T S A G A L D D G R K A V K W A T D Y F L	. 100
AGP76421.1	YKNGLDLTLLFYVAQRSGKLPDDQKVTWRKDSALNDKGQNGRDLTGGYYD <mark>A</mark> GDYV	F G F P M A S T A T V L A W G L V D H E A G Y R S A G A L D D G R K A V K W A T D Y F L	. 100
AGP76422.1	YNQVLRESILFYEAQRSPRLPADLKVTWRKDSALNDRGHKGEDLTGGYYDAGDFV	F G F P M A Y T T T V L A W G L M D Y G V A F S S A G A L D D G R K A L K W A T D Y F L	. 100
AGP76423.1	YN QV L R E S I L F Y E A Q R S G R L P A D L K V T W R K D S A L N D R G H K G E D L T G G Y Y D A G D F V	<b>F G F P M A Y T T T V L A WG L M D Y R V A F S S A G A L D D G R K A H K WA T D Y F L</b>	. 100
AGP76424.1	YN QV L R E S I L F Y E A Q R S G R L P A D L K V T W R K D S A L N D R G Q N G E D L T G G Y Y D G G D L V	F G F P M A Y T T T V L A W G L V D Y E G A F S S A G A L D D G R K A V K C T T D Y F L	. 100
AGP76425.1	YN OV LRESIL FY EAGRS GRL PADL KVT WRKDS AL ND RGONGEDL TGGY YD AGD FY	KFGFPMAYTTTVLAWGLVDYEGAFSSAGALDDGRKAVKWATDYFL	. 100
AGP76428.1	YNQVLRESILFYEDHRSGRLPADLKVTWRKDSALNDRGQNGEDLTGGYYDAGDFV	F G F P M A Y T T T V L A W G L V D S E G A F S S A G A L D D G R K A V K W A T D Y F L	. 100
AGP76427.1	YNQVLRESLLFYEAGRSGRLPVDQKVTWRKDSALNDQGHQGQDLTGGYFDAGDFV	FGFPMAFTTTVLAWGL I DFEGGYSSAGAL DDGRKAVKWATDYFI	100
AGP76428.1	YNQVLRESLLFYEAQRSGRLPVDQKVTWRKDSALNDQGHQGQDLTGGYFDAGDYV	F G F P M A F T T T V L A W G L I D F E G G Y S P A G A L D D G R K A V K W A T D Y F I	100
AGP76429.1	YN QV L RESLL FY EAORS G R L P V DOK V T WRKDS AL ND OGHOGOD L TGGY FDAGDF V	F G F P M A F T T T V L A W G L I D F E G G Y S Y A G A L D D G R K A V K W A T D Y F I	100
AGP76430.1	YNQVLRESIL FYEAQRSGRLPSDQKVTWRKDSALNDKGEQGEDLTGGYYDAGDFV	E G F P M A Y T T T V L A W G M V D Y E V G Y S S A G A L D D G R K A L K W A T D Y L L	. 100
AGP76431.1	YNQVLRESLLFYEAGRSGRLPADQKVTWRKDSALNDUGHOGUDLTGGYFDAGDFV	F G F P M A F T T T V L A W G L I D F E G G Y S S A G A L D D G R K A V K W A T D Y F I	100
AGP76432.1	TNOVI BESLI FYEADRSGREPADORVTWRKDSAL NDUGHQGUDI TGGTFDAGDFV	FGFPMAFTTTVLAWGLIDFEGGYSSAGALDDGRKAVKWATDYFT	100
AGP76433.1	YTTVOKNELLI VAURAGELI POKYTWEKDEAL NDRGNGGDLI TOCYVELOPT	FOFPLARTATVIAWGLIDFEGGTSSAGALDDGRKAVKWATDFF	100
AGP76434.1	KTTVQKTSLLLVDAHRSGSLPADOKVTWRKDSALNDRGONGODITGGYDAGDEV	EGERI ARTATVI AWGI VODRAGEASAGAVODGRKAVKSATOVEL	100
AGP76435.1	VEAVE BASEL LEVADBSGS PADKKYTWRKDSAL ND GGONGI DI TGGYYD GDYY	EGERLAVIATVLAWGI VODTAGETSAGAI DOGRKAVKJATDVL	100
AGP76437.1	RKAVI PNSLLL YVAORSGSL PL DRKYTWRKDSALNDRGONGL DL TGGYMDAGDEV	I GEPCANTATVI AWGG PDATAGENSAGAL DVGPKAVKYATDVI I	100
AG\$32241.1	YKTVLKNSLLEYFADRSGOLPADOKVTWRKDSALNDKGNNGEDLTGGYYDAGDEV	EGEPMAYTATVI AWGI VDYEAGYSSAGAI DDGRKAVKWATDYEI	100
AIT11911.1	YAGAL CMSYVEYEADS GAL PS DORYTWEADS ALADGS DVGVDL TGGYYDAGDH Y	EGEPMAETATMI GWGL V DEEAGHSSAGOL DYGRAAL KWATDYEL	100
AKC91349.1	YASVLGASLLFYEAQRAGKLPANNRLSWRGDSMLMDKGENGEDLTGGYFDAGDYV	F G F P M A A F T T L L A W G A I D Y A D A Y T A S G E L E N M R Q A I R W S T D Y F I	100
AKV16371.1	YKTVLKNSLLFYEAGRSGKLPADQKVTWRKDSALNDKGQKGGDLTGGYYDAGDFV	F G F P M A Y T T T V L A W G V V D Y E S A Y S S A G A L D D G R K A L K W A T D Y F L	100
AMH40356.1	YKLLLQKSLLFFEAQRSGKLPSDQKVTWRHDSALNDKGQNGEDLTGGYYDAGDFV	F G F P M A F S T T V L A W S V L D H P E A Y Q K A N S L E D G R K A V K W A T D Y F I	100
AMH40357.1	YKSVLQKSLLFYEAQRSGKLPSNQKVTWRHNSALNDKGNNGEDLTGGYYDAGDYV	F G F P M A Y S T T V L A W S V L D H A S A Y Q K A G V L A E A R E A I K W A T D Y F I	100
AMH40358.1	YKQVLQQSLLFYEAQRSGKLPADQRVTWRKDSALNDKGDNGEDLTGGYFDAGDFV	FGFPMAYTTTVLSWGLLDHQSGYQSASSVEECLKAIKWATDYFI	100
AMH40359.1	YKEVIQKSLLFYEAQRSGKLPSDQRVTWRHDSAMNDKGVHGEDLTGGYYDAGDFI	FGFQMSFTVTVLAWGLLDHYKGYEAADSVENTRKAIKWGTDYFL	. 100
AMH40360.1	YK TALTDSLLFYEAQRSGKLPSDQRVTWRKDSALNDKGNNGEDLTGGYYD <mark>A</mark> GDFV	FGFPMAYTVTVLSWGVIEYQDVYSSTGNLDYARKAIKWGTDYFL	. 100
AMH40361.1	Y Q E V L R K S L L F Y E A Q R S G K L P A D Q K V K W R A D S A L N D K G R N G E D L T G G Y Y D A G D F V	FGFPMAFTGTML SWSVL DHEAAYTKAGAL DDGRKAIKWVTDYLL	. 100
AMH40362.1	YKTVLRKSLLFYEAQRSGKLPADQKVKWRADSALKDKGWKGEDLTGGYYDAGDYV	VFGFPMAFTTTVL SWSVL DHEKAYTKAGVLGDSHKAIKWATDYLL	. 100
AMH40363.1	TKIALIDSLLFYEAGRSGKLPSDQRVKWRKDSALNDKGNNGEDLTGGYYDAGDFV	F G F P M A Y T I T V L S WG I I D Y E D V Y K S T G N L E Y A R K A V K WG T D Y F L	100
AMH40364.1	TRUVEQRSELFTEAURSGREPADQKVKWRKDSALHDKGNNGEDLTGGYYDAGDFV	F G F P M A F I I I V L S W G L L D H H K G Y E S A G S V E E C M K A I K W A T D Y F I	100
AMU40200	TREVERALEFTEAURSGREPADURVTWRKDSALNDKGVNGEDETGGYYDAGDCV	FOR WORLVIVLAWGLLDHEAGYKAAGVVEDIRKAIKWGTDYFL	100
AMU/0305.1	YKAVI OKELLEVEADEEKLEENOKVKWEBBEEAL BEKOKOVOEDI TOOVOEDI	CECEDMANTTTVI AWTVI DHARAYOVAOVI AEADEAUVWA TOVE	100
- aver ++0.007.1	IN THE REAL PROPERTY AND	ST. WALLING AND THE AREAVEWALDER AND	100
AMH40368.1	YKDALQKSLLFYEAQRSGKLPADQKVTWRHDSALNDKGENGEDLTGGYFDAGDFV	F G F P M A Y T T T V L S W G L L D H H K G Y E T A G S V E E C L K A I K W A T D Y F I	100
AMH40369.1	YKTVLQKSLLFYEAQRSGKLPADQKVKWRADSALNDKGIHGEDLTGGYYDAGDFV	FGFPMAFTATLL SWNILDHEKAYVKAGALEDGRKAIKWATDYLL	. 100
AMH40370.1	YKVLLQKSLLFFEAQRSGKLPADQKVKWRHDSALNDKGQNGEDLTGGYYDAGDYV	F G F P M A F S T T V L A W S V L D H P T A Y Q K A S A L E D G R K A V K W A T D Y F L	. 100
AMH40371.1	YKEVLGKSLLFYEAGRSGKLPADORVKWRHDSALNDKGVHGEDLTGGYYDAGDWV	NFGFQMSFTITVLAWGLLDHEAGYKAAGQVDYTRTAIKWGTDYFL	. 100
AMH40372.1	YKKAFSDSLLFYEAQRSGKLPSDQRVKWRKDSALNDKGNNGEDLTGGYYDAGDFV	F G F P M A Y T I T V L S WG I I D Y E D V Y K S T G N L E Y A Q K A I K WG T D Y F L	100
AMH40373.1	TREVEQUELETEAURS GREPS DORV TWRKDS ALNDKGVNGEDLTGGY YDAGDFV	F G F G M S F F A F I L A W G L I D H Q K G Y E D A G S L E E G R K A L K W A L D Y I L	100
AMH40374.1	TATAFONOLLETEAURSCHLPSDURVKWRKDSALNDKUNNGEDLIGGYYDMGDEV	FOFFWATTTILSWOLLDHEKOVESTUNDELKATKWGTDYFL	. 100
AMH40375.1	TK DV LQK SLLFYEAQKSGKLPADQKVKWRKDSALKDKGQNGEDLTGGYFDAGDFV	FOF PMATIFILS WOLLDHHKGYESANSVSDCLKAIKWATDYFI	100
AMU40077 1	TREELEMALLETEAURSGREPSDURVRWRHDSALNDRGTHGEDLIGGYYDAGDFV	TREFWATATILLAWSVLURKNGYLKANALEUGRKAVKWATDYFL	100
AMH40377.1	YKEVVOKST LEVEADESGKI BADORVTWEKDSAL NDKONNGEDT TOOVODOOL	FOFFWATTATYLAWSYLDHEFAYQQAGYLADAKEATKWATDYFT	100
AMH40270 +	YKTALSDSLLFTEAGROOKEPADORVTWEKDEALNDKOVHGEDLTGGTYDAGDEV	EGEDMAYTITTIAWGI I DEEDVYRETONI EHAOVA I VWOTOVE	100
AMH40380 1	YKDALOKSLLEYOAORSGKLPADOKYTWRHDSALNDKGEKGEDI TGOVEDIGOLV	FGFPMAYTTTVLSWGLLDHHKGYOAAGSVDFCIKAIKW4TDYFL	100
AMH40381 1	YKEVLOKSLLEYEADRSGKLPADORYKWRHDSALNDKGNHGEDLTGGVVDAGDWU	EGEOMSETATILAWGI I DHKKGYEAAGSI EEGRKAI KWATDYEI	100
AMH40382 1	YKTVLOKSLLEFEADRSGKLPADOKVKWRADSALHDKGRKGEDI TGGVVDAGDEV	EGEPMAETATVLSWSVLDHEKAVVKAGVLEDOHKALKWATDVII	100
AMH40383 1	YKAVLQKSLLFYEAQRSGKLPANQKAKWRHDSALODKGNNGEDLTGGY YDAGDY V	FGFPMAYTTTVLAWTVVEHAAAYOKAGVLADAREAVKWATDYFI	100
AMH40384 1	YKTALTDSILEYFADRSGKIPSDORVTWRKDSALNDKGNSGEDLTGGYYDIGDEV	FGFPMAYTITVL SWGVI DYADI YTSTGNLEYARKA I KWGTDYFI	100
AMH40385.1	THE PARTY AND A REPORT OF A		100
654LH03900 4	YK DV VQK SLL FY EAQR SGKL PSDQR V TWRKD SAL ND KGANGEDL TGGY YDAGDFV	F G F Q M S F T A T T L A W G L L D H E K G Y E A A D S L E E G R K A L K W A L D Y F L	
AMP40300.1	YK D V V Q K S L L F Y E A Q R S G K L P S D Q R V T W R K D S A L N D K G A N G E D L T G G Y Y D A G D F V YK D V L Q K S L L F Y E A Q R S G K L P S D Q K V T W R K D S A L N D K G E N G E D L T G G Y Y D A G D F V	F G F Q M S F T A T T L A WG L L D H E K G Y E A A D S L E E G R K A L K WA L D Y F L F G F P M A F T T T V L S WG L L D H Y NG Y K S A G S V D E G L K A T K WA T D Y F I	100
AMH40387.1	YKDVVQKSLLFYEAORSGKLPSDORVTWRKDSALNDKGANGEDLTGGYYDAGDFV YKDVLQKSLLFYEAORSGKLPSDCKVTWRKDSALNDKGENGEDLTGGYYDAGDFV YKEVLEMSTLFYEAORSGKLPSDCRVKWRGDSALNDKGANGEDLTGGYYDAGDFV	F GF GM S F TA TT LAWGLLDHER GY EAAD SLEEGRKALKWALDY FL F GF PM A F T T T V LSWGLLDH YN GY KSAG SV DE GLKATKWALDY FL F GF GM S F T V T V LAWGLLEHKAG Y TAAGAV DD T LKATKWGTDY FL	100 100
AMH40387.1 AMH40388.1	YK DV VQK SLLFYEA OR SGKLP SDOR V TWRK DSALNDKGANGEDLTGGY YDAGDFV YK DV LQK SLLFYEA OR SGKLP SDOK V TWRK DSALNDKGEN GEDLTGGY YDAGDFV YK EV LEM STLFYEA OR SGKLP A DOR V KWR GD SALNDKGAN GEDLTGGY YDAGDFV YKLLLQK SLLFYEA OR SGKLP SDOK V TWRHDSALNDKGAN GEDLTGGY YDAGDFV	FGFOMSFTAITLAWGLLDHERGYEAADSLEEGKKALKWALDYFL FGFPMAFTTTVLSWGLLDHYNGYKSAGSVDEGLKAIKWATDYFI FGFDMSFTVTVLAWGLLEHKAGYTAAGAVDDTLKAIKWGTDYFL FGFPMAFSTTLLAWSVLDHSAAYESANALEDGRKAVKWATDYFI	100 100 100
AMH40387.1 AMH40388.1 AMH40388.1 AMH40389.1	YKDVVQKSLLFYEAQRSGKLPSDQRYTWRKDSALNDKGANGEDLTGGYYDAGDFY YKDVLQKSLLFYEAQRSGKLPSDQKYTWRKDSALNDKGENGEDLTGGYYDAQDFY YKEVLEMSTLFYEAQRSGKLPADQRYKWRGDSALNDKGANGEDLTGGYYDAQDFY YKLLQKSLFYEAQRSGKLPSDQKYTWRHDSALNDKGQNGEDLTGGYYDAQDYY YKSVLQKSLLFYEAQRSGKLPSDQKYTWRHDSALNDKGONGEDLTGGYYDAQDYY	A GFOMSFIATILAWGLLDHKGYEAADSLEEGKAALKWALDYFL GEOPMAFTTVLSWGLLDHYNGYKSAGSVDEGLKAIKWATDYFI GEOFMSFTYTVLAWGLLEHKAGYTAAGAVDDTLKAIKWGTDYFL GEOFMAFSTTLLAWSVLDHSAAYESANALEDGRKAVKWATDYFL GEOFMAYTATVLAWTVIDHASAYQTAGVLSDAREAVKWATDYFL	100 100 100 100
AMH40386.1 AMH40387.1 AMH40388.1 AMH40389.1 AMH40390.1	YKDVVQKSLLFYEAORSGKLPSDQRVTWRKDSALNDKGANGEDLTGGYYDAGDFV YKDVLQKSLLFYEAORSGKLPSDQKVTWRKDSALNDKGANGEDLTGGYYDAGDFV YKEVLEMSTLFYEAORSGKLPSDQRVKWRGDSALNDKGANGEDLTGGYYDAGDFV YKLLQKSLLFYEAORSGKLPSDQKVTWRHDSALNDKGQNGEDLTGGYYDAGDYV YKSVLQKSLLFYEAORSGKLPSNQKVTWRHNSALNDKGGNGEDLTGGYYDAGDYV	F GF OM SF TATTIL AWGLLDH HAG YEAADSLEEG KAALKWALDYFL F GF PMAFTTTVL SWGLLDH YN GYK SACS VDEGLKAI KWATDYFL F GF PMAFSTTLLAWSVLDH SAAYESANALED GRKAVKWATDYFL F GF PMAYTATVLAWTVIDHASAYESANALED GRKAVKWATDYFL F GF PMAYTATVLAWTVIDHASAYETAD SVEECMKAI KWATDYFL	100 100 100 100 100
AMH40386.1 AMH40387.1 AMH40388.1 AMH40389.1 AMH40390.1 AMH40391.1	YKDVVQKSLLFYEAORSGKLPSDORVTWRKDSALNDKGANGEDLTGGYYDA GDFV YKDVLQKSLLFYEAORSGKLPSDCKVTWRKDSALNDKGEN GEDLTGGYYDA GDFV YKEVLEMSTLFYEAORSGKLPSDCKVTWRKDSALNDKGANGEDLTGGYYDA GDYV YKLLLQKSLLFYEAORSGKLPSDCKVTWRHDSALNDKGNNGEDLTGGYYDA GDYV YKSVLQKSLLFYEAORSGKLPSNCKVTWRHDSALNDKGNNGEDLTGGYYDA GDYV YKDVLQKSLLFYEAORSGKLPSNCKVTWRHDSALNDKGNNGEDLTGGYYDA GDYV YKDVLQKSLLFYEAORSGKLPSNCKVTWRHDSALNDKGNNGEDLTGGYYDA GDYV YKSVLQKSLLFYEAORSGKLPSNCKVTWRHDSALNDKGNNGEDLTGGYYDA GDYV	F GF DMS F IAT I LAWG L DH FKG YE FADSLEEG KKA LKWALDYF L F GF PMA F TT T V LSWG L DH YNG YK SAGS VO BC LKA I KWAT DYF I F GF PMA F ST T L LAWG U L BH KAG YT A AGA V DD T LKA I KWGT DYF L F GF PMA F ST T L LAWS V L DH SA AY E SA NA LE DG RKA X KWAT DYF I F GF PMA Y T AT V LAWT V I DH AS AY QT AGVL SD AR EA V KWAT DYF L F GF PMA Y T T T V LSWG L DH KKG YE T A DSY EE GMKA I KWAT DY L F GF PMA F T AT T V LSWN I LD H EKKAYT KAG V LE DG RKA I KWAT DY L	100 100 100 100 100 100
AMH40386.1 AMH40387.1 AMH40388.1 AMH40389.1 AMH40390.1 AMH40391.1 AMH40392.1	YKDVVQKSLLFYEADRSGKLPSDQRYTWRKDSALNDKGANGEDLTGGYYDA GDFY YKDVLQKSLLFYEADRSGKLPSDQKYTWRKDSALNDKGENGEDLTGGYYDA GDFY YKEVLEMSLTFYEADRSGKLPSDQRYKWRKDSALNDKGANGEDLTGGYYDA GDFY YKLLQKSLLFYEADRSGKLPSDQKYTWRHDSALNDKGNGEDLTGGYYDA GDYY YKSVLQKSLLFYEADRSGKLPSDQKYTWRHDSALNDKGENGEDLTGGYYDA GDFY YKSVLQKSLLFYEADRSGKLPSDQKYTWRKDSALNDKGENGEDLTGGYYDA GDFY YKTVLQKSLLFYEADRSGKLPSDQKYTWRKDSALNDKGENGEDLTGGYYDA GDFY YKTVLQKSLLFYEADRSGKLPSDQKYTWRKDSALNDKGENGEDLTGGYYDA GDFY YKTVLQKSLLFYEADRSGKLPSDQKYTWRKDSALNDKGENGEDLTGGYYDA GDFY	F GF GMS F TATTUL SWGLLDH KA GY FAA DSLEEG KRALK WALD YF L F GF GMS F TY TY LSWGLL DH YN GY KSAGS VD GE LKA I KWATD YF I F GF GMS F TY TY LAWGLLE H KA GY TA AGA V DD T LKA I KWATD YF L F GF F MA F STTLLAWS LD H SA AY E SANALE DG RKA V KWATD YF L F GF F MA Y TATY LAW TV I DH AS AY QTA GY LSD AR E A V KWATD YF L F GF F MA Y TATY LSWGLLDH KKGY E TA DSVEE CMKA I KWATD YF L F GF F MA F TATY LSWILD H KKGY T KA GY LE DG RKA I KWATD YF L F GF F MA F TATY LSW VLDH KASY TSASALDDA HKA I KWATD YF I	100 100 100 100 100 100 100

			*					
	AMH40394.1	YATVLQNSLLFYEAQRSGKLPSDQKVTWRKDSALNDKGDNGEDLTGGYY	DAG	DF V	FGFQMAFTATVLSWGL	LDHEAGYSNAGALEDGRKALKWATD	YEL 10	or
	AMH40395.1	YERVLELSLLFYEAQRSGKLPSDNRVTWRGDSALEDRGQKGEDLTGGYY	DAG	DFV	FGFTMASTTTLLAWGF	LSYKEAYISAGQFNHGLNALKWSAD	YFI 10	ar
	ANR02619.1	YRNALQKSIL FYAAQRSGRLPSNNPI DWRGDSALGDQGNNGQDL TGGWY	DAG	DHV	FGLPMAWSATTLIWGM	IDFANGYGA DRNNAMQSVRWALD	YEM 98	3
	AOV94250.1	Y K Q V L Q K S L L F Y E A Q R S G K L P S D Q K V T W R K D S A L N D R G N N G E D L T G G Y Y	DAG	DFV	FGFPMAFTATVLAWTI	I D H E S A Y N S A G S M D D A R K A V K W A T D	YFI 10	00
	AOV94251.1	Y K E V L K Y S M R F Y E A Q R S G K L P A D N H I E W R G D S A L N D R G T R G E D L T G G Y Y	DAG	DFV	FGFPMAYTITILSWGA	IEYEEGYQNAGLLNDIQDAIKWGTD	YFL 10	90
	AOV94252.1	YKQVLNYSLLFYEAQRSGKLPADQHVTWRKDAALDDKGNNGEDLTGGYF	DAG	DFV	FGLPLAYTVTTLSWGV	ILYEDAYVSTNLLDEVHEAIKWGTD	YIL 10	X
	AOV94253.1	YN HVLDLSLR FYEAQRSGKLPKDNR I PWRGDSALNDKGQRGEDLTGGYY	DAG	DFV	FGFTMASTITLLAWGF	LSYRDAYIAAGQEKYGLSAIKWASD	YFI 10	ж
	AOV94254.1	YROVENYSMEFYEAURSGKEPADQHVTWRKDSALGDKGNDGEDETGGYY	DAG	DEVE	FGFPMAYTITMLSWGV	I A Y Q K G Y E D A G L A D N I HE A I K W G I D	IYFL 10	JI.
	AOV94255.1	YKOVI NYSNI EVENORSCHI PADOHVIWKHDSAL NDKGUNGEDLIGGTY	DAG	DIV	FORPMANTITUL SWGA	I E YAUGY DSAGLINDI UEATRWGTD	VEL 1	JL or
	A0V94258.1	TROVENTSMETTEAGREGAL BODORY TWRADSALGDAGNDGEDETGGTT	DAG	BUU	FORDMARTATTIANCI	LEEEACY DEACLADNINEAT KWGTD		
	APP29556.1	YSOAL CMSYVEYEADRSGAL RODORY TWRODSAL DDONDY GHDL TOGYY	DAG	DUV	EGERMAETATNI AWGI	I DEESGYNEAGOL DYGRAGLRWATD		nr nr
	APE200001.1	VAGAL CMSYVEYEAGRSGAL PODGRVTWRGDSAL DDGSDVGHDL TGGYY	DAG	DHV	EGERMAETATMLAWGL	I DEESCYNSAGOL DYGRAGL RWATD		nr
	APE29559 1	EVI I LOLEI GERVVVAVGALEN NGGGGDGSDVGI DL TGGYY	DAG	DHV	EGEPMAETATMLGWGL	I DE EAGHSSAGOL EYGRAAL RWATD	YEL 90	<i>"</i>
	APE29560 1	YAGAL CMSYVEYEAERSGDI PGDORVTWRADSAL DDGSDVGHDI SGGYY	DAG	DHV	EGEPMAETATMLAWGI	I DEEDGHNTAGOL DYGRAAL RWATD	YEL 10	òr
	APF29561.1	YAQVLCMSYLFYEAQRSGPLPADQRYTWRYDSALDDGSDVGHDLTGGYY	DAG	DHV	FGFPMAFSTSLLGWGV	V D F K E G H E A A G Q L D Y G L A A V K WA T D	YEL 10	or
	ATS91249.1	YNDVLYKSIL FYEAORSGKLPANKRI PWRGDSALNDKGDKGEDL TGGWY	DAG	DHV	ENEPMAYTTTVLAWSY	LLYPDAYEAAGOKDYILDCIKWPLD	YLL 10	a
	BAA31326.1	YKTVLSNSLLFYEAORSGKLPSDOKVTWRKDSALNDKGOKGEDLTGGYY	DAG	DEV	EGEPMAYTVTVLAWGV	I DYESAYSAAGAL DSGRKAL KYGTD	YEL 10	ar
	BAA33708.1	YKQVLRDSLLFYEAQRSGRLPADQKVTWRKDSALNDQGDQGQDLTGGYF	DAG	DEV	FGFPMAYTATVLAWGL	I D F E A G Y S S A G A L D D G R K A V K W A T D	YFI 10	ar
	BAA33709.1	YKQVLRDSLLFYEAQRSGRLPADQKVTWRKDSALNDQGEQGQDLTGGYF	DAG	DFV	FGFPMAYTATVLAWGL	I D F E A G Y S S A G A L D D G R K A V K W A T D	YEL 10	)r
	BAA34050.1	YKTVLSNSLLFYEAQRSGKLPSDQKVTWRKDSALNDKGQKGEDLTGGYY	DAG	DFV	FGFPMAYTVTVLAWGV	I DYESAYSAAGALDSGRKALKYGTD	YFL 10	)í
	BAB40693.1	Y K T V L K N S L L F Y E A Q R S G K L P A D K K V T W R K D S A L N D K G Q K G E D L T G G Y Y	DAG	DFV	FGFPMAYTVTVLAWGL	V D Y E S A Y S T A G A L D D G R K A L K W G T D	YFL 10	00
	BAB40695.1	Y K T V L K N S L L F Y E A Q R S G K L P A D K K V T W R K D S A L N D K G Q K G E D L T G G Y Y	DAG	DFV	FGFPMAYTVTVLAWGL	V D Y E S A Y S T A G A L D D G R K A L K W G T D	YFL 10	00
	BAB40696.1	YK T V L K N S L L F Y E A Q R S G K L P A D Q K V T W R K D S A L N D K G Q K G E D L T G G Y Y	DAG	DFV	FGFPMAYTVTVLAWGL	V D Y E S A Y S T A G A L D D G R K A L K W G T D	YFL 1	)(
	BAB40697.1	YKTVLKNSLLFYEAQRSGKLPADQKVTWRKDSALNDKGQKGEDLTGGYY	DAG	DFV	FGFPMAYTVTVLAWGL	V D Y E S A Y S T A G A L D D G R K A L K W G T D	YFL 10	00
	BAC67186.1	YGEALGKSIL FYDAQRSGKLPANNPI KWRGDSALGDKGDNGEDLTGGWY	DAG	DHV	FSLPMSSTSTVLLWGY	LQWKDAYATTKQTDMFFDMIKWPLD	YFL 10	30
	BAD01504.1	Y G E A L G K S I L F Y D A Q R S G K L P A N N P I K W R G D S A L G D K G D N G E D L T G G W Y	DAG	DHV	FSLPMSSTSTVLLWGY	LQWKDAYATTKQTDMFFDMIKWPLD	IYFL 1	30
	BAD44734.1	Y G E A L G K S I L F Y D A Q R S G K L P A N N P I K W R G D S A L G D K G D N G E D L I G G W Y	DAG	DHV	FSLPMSSISIVLLWGY	LQWRDAYATIKQIDMFFDMIKWPLD	IYFL 10	JI.
	BAF38757.1	YGEALGMSTLFFDAURSGRLPANNPTPWRGDSAVND-GDAGHDLSGGWY	DAG	DHV	FNLPMAFSTWVLEYGM	L K F K DG Y Q A A G Q L DMA C DM I K W P L E	YFL 9	3
	BAK20401.1	VREVLOKSILEVAAORSOOLPONNRIDIRGDSALGDRGNGGGDLIGGWI	DAG	DHV	EGI PMARSTITLAWGI	LEFRAATEAAGQTSTALDSTRWPLD	VEM OF	л. 8
	BAI 80587 1	YKOAL CMSEL EVGADESGKLES - SELEWEGDSAL NDGSDVGKDLTGGVY	D & G	DHV	EGERMAGTVTVLGWGV	VEYRDAYTDSGOLEYGLDALKWGTD	YEL 90	á
	BAM147161	YDEVI EKSI I EVEAERSGDI PANNEL PYRGDSAL GDOGNOGODI TOGWY	DAG	DHV	EGEPMAEATTTLAWGI	LEERDGYEAAGOYNLALDSIRWTIN	YEL 10	ór
	BAS18736.1	YG - ALGKSIL FYDAOGSGRLPANNPI PWRGDSALGDC VVGGWY	DAG	DHV	FGFPMASSTTLLLWGL	I R F K D G Y V R A N O L D M M Y D M I K W P L D	YEL 9	3
	BAT21145.1	YDEVLEKSIL FYEAERSGDLPANNR I PYRGDSAL GDQGNQGQDL TGGWY	DAG	DHV	FGFPMAFATTTLAWGI	LEFRDGYEAAGQYNLALDSIRWTLN	YFL 1	àr
	BAU45387.1	YG - ALGKSILFYDAQRSGRLPANNPIPWRGDSALGDC VVGGWY	DAG	DHV	FGFPMASSTTLLLWGL	I R F K D G Y V R A N Q L D M M Y D M I K W P L D	YFL 9	3
	BAU45388.1	YG - ALGKSIL FYDAQRSGRLPANNPIPWRGDSALGDC VVGGWY	DAG	DHV	FGFPMASSTTLLLWGL	IRFKDGYVRANQLDMMYDMIKWPLD	YFL 93	з
	BAX09271.1	YG - ALGLSIK FYEAQRSGPIHD - NIVPWRGNSGMHDC TVGGWY	DAG	DHV	FGLPMAASAHILLWSL	YRFKDGYVKGNQLNQMYHMIKTPLN	YFE 92	2
	CAD54726.1	YN DVL TKSLL FYEAQRSGKL PSDQKV TWRKDSAL NDKGQN GEDL TGGY Y	DAG	DYV	FGFPMAYTATVLAWGL	V D H P A G Y S S A G L L D D G R K A L K W V T D	YLI 10	00
	CAD54727.1	YKDVLTKSLLFYEAQRSGKLPSDQKVTWRKDSALNDKGQNGEDLTGGYY	DAG	DYV	FGFPMAYTATVLAWGL	V D H P A G Y S S A G L L D D G R K A L K W V T D	YLI 10	00
	CAD54728.1	YK DVL T K S L L F Y E A Q R S G K L P S D Q K V T W R K D S A L N D K G Q N G E D L T G G Y Y	DAG	DYV	FGFPMAYTATVLAWGL	V D H P A G Y S S A G L L D D G R K A L K W V T D	YLI 10	X
	CAD54729.1	YK DV L T K S L L F Y E A Q R S G K L P S D Q K V T W R R D S A L N D K G Q N G E D L T G G Y Y	DAG	DYV	FGFPMAYTATVLAWGL	V D H P A G Y S S A G V L D D G R K A V K W V T D	YLI 1	)(
	CAD54730.1	YN DVL TKSLL FYEAQRSGKL PSDQKV TWRKDSAL NDKGQ NGEDL TGGY Y	DAG	DYV	FGFPMAYTATVLAWGL	V D H P A G Y S S A G V L D D G R K A V K W V T D	YLI 10	)(
	EFN70196.1	YAKVLNLSLMFYEAQRSGKLPRENRIPWRNDSALHDAGWEDEDLTGGYY	DSH	NYV	FGFTMAFTTTILAWGA	V S W P E A F N V A N Q L D E M H K T I K W A T D	YFI 10	х
	EFN70197.1	YGEVLNRSLQFYEAQRSGKLPRENRIPWRGDSALHDVGLEGEDLTGGYY	DGH	NYV	FGFTMAFTTTILAWGA	V S W P E A F N V A N Q L D E I R K A I K W A T D	1YFI 10	Л
	EFX69372.1	YAEVITKSLLFYYAQRSGRLPTDNPISYRSDSALGDIGQNGEDLTGGYY	DAG	DFV	FGFPMAYTTTVSAWSV	I D Y P S G Y S S A G Q L T N A R N M V K W A S D	1YF1 10	30
	EFX80604.1	YREVIIKSLLFYYAURSGRLPVDNPIAYRSDSALKDVGUNAEDLIGGYY	DAG	DHL	LGFPLASAMITAAWST	I DYPIGYSSAGQLIIAUNMVKWASD	IYLI 10	30
	EFX80605.1	YTEVTTKSLLFYYAURSGRLPUDNPTPYRSDSALNDRGUDGEDLTGGYY	DAG	DHL	LGFPLASSIFIVAWST	I DYPEGYSSAGKETTARNMVKWATU	ITEI 10	1
	EGI83652.1	YAOVI ELSI LEVEADRSGKI PKNNRI SWRGDSAL VDRGSNGEDI TGGVY	DAG	DEV	EGETMAETTTIIAWGT	VNWPEAYDAAGOLDELERALKWSTD	IVEL 10	, nr
	EKC17289.1	YDEVLMKSILFYEAORSGKLPDDNRIPWRGDSSLLDAGDNGEDLTGGWY	DAG	DNI	FGFPMAASTTLLTWGL	LRYKDAYOHSGOLEHMYSCIRWPLE	WML 10	or
	EKC39023.1	YGOVLEKSILFYEADOSGDLPPTHRINWRGDSALGDSGIYGEDLTGGWY	DAG	DGV	FNLPMSASVINLAVGL	YKWRDAYSASNOLOHMYDSIKWPLD	YFL 10	ő
	KDR16731.1	YSRVLELSLLFYEAGRSGQLPPDNRISWRGDSALGDRGQNGEDLTGGYY	DAG	DEV	FGFTMASTTTLLAWGY	FSYRDAYESAGQAKYALDAIKWAAD	YEL 10	or
	KMQ95413.1	YAEVIKLSLLYFEAGRSGKLPKDNRIPWRGDSAVNDRGPNCEDLSGGYY	DAA	DFV	FTFTTAFTTTILAWGV	ISWPKAYIAANQLDEIYKTIKWATD	YEL 10	ar
	KOC62382.1	YARVLEFSLLFYEAQRSGKLPENNRIPWRGDSALVDRGLNGEDLTGGYY	DAG	DEV	FGFTMASTTTLLAWGG	V S W P E A Y N A A G Q L D E L R K A I K W A T D	YFI 10	ar
	KOF68403.1	EILLKSMLFYEAQRSGYLLPNRRIRWRGHSALNDRSEFGEDLTGGWY	DAG	DHV	FGFTMASATTILLWGL	V E F R E A Y E K T G L I K D M L D C V R W P L D	YFL 9	3
	KOX76353.1	YARVLELSLLFYEAGRSGKLPGNNRVPWRGDSALDDRGLNGEDLSGGYY	DAG	DFV	FGFTMASATTLLAWGA	I SWPEAY I AAGQL DELRKA I KWATD	YFI 10	00
	KXJ18436.1	Y D Q V L H D S I L F Y E A Q R S G K L P S T N R I P W R G D S A L N D K G Q N G Q D L T G G W Y	DAG	DFV	FGFPMASSVTVLAWGV	V E Y R D A Y K A A G E L D N V L D S I K W A T D	YFI 10	00
	KYB27092.1	YRHALKLSLLFYEAQRSGPLPASNRITWRSDSALDDHGLNNEDLTGGYY	DAS	DYV	FSFTMAFTTTILTWGM	LSFRHAYNEAGQYQHGLDAIKWATD	14 FI 10	)I
	KYM87195.1	YAQVLELSLLFYEAQRSGKLPKNNRISWRGDSALYDRGSNDEDLTGGYY	DAG	DFV	FGFTMAFTTTLLAWGT	V N W S E A Y D A A D Q L D E I R R A I K W S T D	IYFI 10	)(
	KYN01696.1	YAKVLELSLLFYEAQRSGRLPKNNRISWRGDSALYDRGSNGEDLTGGYY	DAG	DFV	FGFTMAFTTSLLAWGT	V NWPEAYDAAGQLDEIRRTIKWSTD	IYFI 10	)(
	KYN22851.1	YAOVI ELSI LEVEADRSGKI PKNNR I SWRGDSAL YDRGSNGEDI TGGYY	DAG	DEV	EGETMAETTTILAWGT	VNWPEAYDAAGOLDELERALKWSTD	YEL 10	or
	KYN35700.1	YAOVLELSLL FYEAORSGKLPKNNR I LWRGDSAL YDRGSNGEDL TGGY Y	DAG	DEV	EGETMAETTTLLAWGT	VNWPEAYDAAGOLDEIRRAIKWSTD	YEL 10	ar
	KYQ52831.1	YGQALKFSLLFYEAORSGKLPKNNRI SWRGDSALYDRGSNGEDLTGGYY	DAG	DEV	FGFTMAFTTTILAWGT	VNWPEAYDAADQLDEIRRAIKWSTD	YEL 10	àr
	KZC10616.1	YAQVLELSLLFYEAQRSGKLPKNNRIQWRGDSALDDHGLNGEDLTGGYY	DAG	DFV	FGFTMASTTTLLAWGA	V S W P E A Y D A A G Q L E K L R D A V K W A T D	YEL 10	br
	KZS03923.1	YNLALKKSIL FYEAGRSGKLIN - NRIDWRGDTFLNDGSDVGVDLSGGYF	DAG	DHV	FGFPFAHSMIFLAWGM	I DYRSTYQAAGELDNGLAALRWGAD	WIL 99	3
	KZS05722.1	YRDA I TKSLL FYYAQRSGRL PADNPI PYRSDSALKDRGRNGEDL TGGY Y	DAG	DHV	FGFPMAGTTTIIAWSI	I D Y Q L G Y S S V G Q L K N A L N M V K W A T D	YFI 10	X
	KZS07093.1	YAEVITKSLLFYYVQRSGRLPTDNPIPYRSDSALGDVGQNGEDLTGGYY	DAG	DFV	FGFPMAFMTTVVASSV	LDYPLGYSSAGQLTNARNMVKWASD	YFI 10	00
	KZS16514.1	YHEAITKSLLFYYAQRSGRLPVDNPIPYRFDSAVNDRGQSGEDLTGGYY	DAG	DHL	LGFPLASAITIIAWSI	V D F P L G Y S S A G Q L T K A Q N M V K W A T D	YFI 10	)(
	KZS16516.1	YPDAITKSLLFYYAQRSGRLPIDNPIPYRSDSALNDRGQSGEDLSGGYY	DAG	DHL	LGFPLASTITIIAWSI	I D Y P L G Y S S A G Q L T K A Q N M V K W A T D	YFI 10	X
	OAD61293.1	YARVLELSLLFYEAQRSGKLPENNRIPWRGDSALDDRGLNGEDLTGGYY	DAG	DFV	F G F T M A S T T T L L A W G A '	V S W P E A Y N A A G Q L D E L R E A I K W A T D	14 F I 16	X
	ODM94862.1	YGEVIEKSLLFYEAGRLGRLPSSNRIPWRGDAFVSDRGQNGEDLSGGYC	DAG	DFV	FGFPFTSAMTMLAWGM	A D F K A G Y V K A G Q W D Y A K D A L K W G L D	YMI 10	30
	ODM96900.1	YGEVIEKSILFYEAURLGRLPSSNRIPWRGDSFVNDRGENNUDLIGGYF	DAG	DHV	FGFPFASAMIMLAWGM.	A D F KUGYVKAGQWEQAKDALRWGLD	YMI 10	JI.
	UWF45593.1	YDELLERSIL FYEAURSGREPPINRIPWRGDSALNDHGUSGEDLIGGWY	DAG	DHIN	FGLPMSSTVTLLAWSM	LRFRDAYEHAGQLEYAYDCIRWPLE	WWL 10	л. 
	DWF30007.1	YARVLEDSILFTEAURSGELPTUNKTHTKKDSATTDGSDVGLDLIGGWT	DAG	DEV	FREEMASTTTLLAWGE	SEWRGAYMASGQLGSMLDSTRWPLD VSWREAYNAAGOLDELPKAIKWATO	IVEL 1	JL nr
	ndbl3MC3(A	VDEVLEKS LLEVEAERS GDL PANNR LPVRGDS AL GDOGNOGODI TGGWY	DAG	DHV	EGEPMAEATTTLAWGI	LEEPDGYEAAGOVNLALDS I RWTI N		â
	PEX25363.1	YSEVLTLS ILFYEAORSGKLPSNNRVKWRKNSALGDKGONGEDLTGGWY	EAG	DYV	FGFPMASSVTVLAWGI	VEYKKAYVAAGOYSNVLNSIKWATD	YEL 10	år
	PSN31180.1	YGDV I RKSLL FYEAQRSGKLPSDQQVTWRKDSALNDKGEN	- AG	DYV	FGFPMAYTITTLAWGA	I D Y A N G Y N K A N A L D G V R K A I K W G T D	YFI 90	J
	PVD35337.1	YAEVLQKSIL FYEAQRSGKLPASNRI PWRGDSAL QDRGDKGEDL TGGWY	DAG	DNV	FNFPMAFSTTVLCWSL	LEFKDAYDKAGQLDYMYDSIRWPLE	YFI 10	)(
	PVD35340.1	Y A E V L Q K S I L F Y E A Q R S G K L P A S N R I P W R G D S A L Q D R G D K G E D L T G G W Y	DAG	DNV	FNFPMAFSTTVLCWSL	LEFKDAYEKAGQLDYMYDSIRWPLE	YFI 10	n
	PVD35381.1	YDDVLYKSIL FYEAQRSGKLPANNRIPWRGDSALND - GDGGVDLTGGWY	DAG	DFV	FNFPMAWSTAILNWGL	LQFKDAYEAAGQLEWMYESVKWPLD	YLL 9	3
	PVD35383.1	YDDVLYKSILFYEAQRSGKLPANNRIPWRGDSALND-GDGGVDLTGGWY	DAG	DFV	FNFPMAWSTAVLTWGL	LQFKDAYQAAGQLEWMYESVKWPLD	YLL 9	9
	PVD35384.1	Y D E V L M K S I M F F E A Q R S G K L P A T N R I P W R G D S G L R D R G T G G E D L T G G W Y	DGA	GFV	FTFPTAWSTVILTWGL	LEFRDAYEAAGQVEWMYDCIKWPLD	IYLL 1	n
	PVD35385.1	T DELLMKSIMFYEAURSGKLPAINRIPWRGDSGLQDRGTGGEDLTGGWY	UIGG	GYV	FILPIAWSTVILTWGL	LEFRDAYEAAGQVEWMYDCIKWPLD	17 LL 10	1
	PVD35551.1	YG-ALAKSILFYDAURSGKLPANNPIHWRGDSALGDGVVGGWY	DAG	DHV	FGLPFGAATHILLWGL	Q L F K D G Y Q K I N Q L D N M Y D M I K W G L D	IYML 93	3
	PVD36624.1	ADTIL SPACE - PIPOCTTTOPSPIC - PTIPOTVIKO	GAG	DHV	ENERMAESTIMLAWGL	LETRUATQAAGQLDHMYDSIKWPLD	VIL 1	л л
	PVD38825.1	ADTELSRAGY RTRGGTTTGRSREC PTERCTTERG		DUV	FREPRAFATIVLIWGU		VMU 04	ĩ
	PVD36986.1	YA-ALGKSILEYDADRSGKLPPDNPLHWRGDSALNDC	DAG	DHV	EGEPLASTTHLLLWGI	RIEKDGYIKAGQIDMMYDMIKWPID	YMI OF	3
P	001606454 1	YARVLELSLLFYEAORSGKLTSAGRVAWRGDSALGDRGANGEDI TGGVY	DAG	DEV	FGFTMASTTTLLAWGA	VSWPDAYAAVGOIDVIRETVKWATD	YEV H	or
(P	001810693.2	YRHALKLSLLFYEAQRSGPLPASNRITWRSDSALDDHGLNNEDLTGGYY	DAS	DYV	FSFTMAFTTTILTWGM	LSFRHAYNEAGQYQHGLDAIKWATD	YEL 1	'n
(P	001944774.2	Y D E V L R L S L M F Y E A Q R S G H L P V N N R I P W R A D S A T T D K G Q N G E D L S G G Y Y	DAG	DFV	FGFTMASTTTILAWGY	LSYKDAYEEVNEWENALGSIRWAAE	YFI 10	20
(P	002426465.1	Y D K A L E L S L L F Y E A Q R S G Y L P P D N R I S W R Q N S G L N D S S P N G D D L T G G Y Y	DAG	DFV	FGFTMASTTTLLAWGL	LSYKDAYLASGQYKHCLNAIKWASD	YFI 10	n
(P	003402778.1	Y A R V L E L S L L F Y E A Q R S G K L P E N N R I P W R E D S A L D D R G L N G E D L T G G Y Y	DAG	DFV	FGFTMASTTTLLAWGA	V S W P E A Y N A A G Q L D E L R K A I K W A T D	YFI 10	n
۲P.	003690676.1	Y A R V L E L S L L F Y E A Q R S G K L P E N N R I P W R G D S A L D D R G L N G E D L T G G Y Y	DAG	DFV	FGFTMASTTTLLAWGA	V S W S E A Y N A A G Q L D E L R K A I K W A T D	YFI 10	)0
P.	005099312.1	YA - ALGKSILFYDAQRSGKLPANNPIKWRGDSGLNDC VVGGWY	DAG	DHV	FGLPMASSTTLLLWSL	Y S F K D G Y V K A G Q L N N M Y D M I K W P L D	YFL 93	3
CP.	005099315.1	YG - ALGLSIKFYEAQRSGPI - NDNIVPWRGNSAMSDC TVGGWY	DAG	DHV	FGLPMAATTHILLWSL	YRFKDGYVRANQLNQMYHMIKTPLN	YEM 92	2
KP.	005101787.1	TK-ALKKSILFYEAURSGPI-NDHIIPWRHSSAMHDCTVGGWY	UNG	DHV	FGLPMAASSHLLMWGL	TEFROGYTQAGELNQMYHMLKTPLN	YFK 92	1
NP.	005101789.1	TR-ALRESTLETEAURS OF INU-ULIPWRHSSAMHDGTVGGWY	G	DEV	ENEDMASSAHLLMWGL	TERNUGTVEAGQENQMYHMEKTPEN	ITEK 92	1
(P)	005111868.2	TNULLINGTLETEAURGUKLPSNNKTPWRGDSALGDKGAKGEDLTGGWY	on G	DEV	ENERMAYSTIVLTWGY	LETPUATRAAGQEDKILDCIKWPLQ	VII H	л ог
(P	005111970 4	YNDVLYKSIL FYFTORSGKLPASNPLPYPGDSAMUDOGENCEDLTGGWY	D IC	DHV	ENERMSYSTTVI TWOY	LI FPEAYKAAGOEDKELDCIRWPLE	YIL O	14 01
(P	006610535 1	YARVLELSLLFYEAQRSGKLPENNRIPWRGDSALFDRGLNGEDLTGGVY	DAG	DEV	FGFTMASTTTLLAWGA	V SWPEAYNAAGOLDELRKAIKWATO	YEL 1	01
(P	008181842.1	Y DEVLRLSLMFYEAQRSGHLPVNNR I PWRADSATTDKGQNGEDLSGGYY	DAG	DEV	FGFTMASTTTILAWGY	LSYKDAYEEVNEWENALGSIRWAAE	YEL 1	'n
P	008194455.1	YRHALKLSLLFYEAQRSGPLPASNRITWRSDSALDDHGLNNEDLTGGYY	DAS	DYV	FSFTMAFTTTILTWGM	LSFRHAYNEAGQYQHGLDAIKWATD	YEL 1	or
(P	009049791.1	Y DQ I LQLS I L F Y E SQR S G K L P A N N R I P W R G D S A L N D K G S N G E D L T G G W Y	DAG	DHV	FGLPMAFSTTMLTWSL	LEFQDAYQASGQLQHMYDSIKWPLD	YFL 10	00
(P	009049836.1	YADVLMKSIL FYEAQRSGKLPANNRIPRRGDSALNDKGSNGEDLTGGWY	DAG	DNV	FNFPMAWSTTVLTWGL	IQWKDAYENAGELDKI TLWLI -	93	3
(P	009049839.1	Y A E V L E K S I L F Y E A Q R S G K L P A N N R I P W R G D S A L S D K G S N G E D L T G G W Y	DAG	DNV	FNFPMAWSTTVLTWGM	I Q W K D A Y E K T G Q L N H M Y E S I R W P L D	YLM 1	n
(P	009049840.1	Y D E V L Q K S I L F Y E A Q R S G R L P A I R N I A W R G D S A L Q D Q G I F G E D L S G G W Y	DAG	DCV	FGFPMAASTTLLLWGF	IEFYDGYEYAGLVNEMYDCVQWPLT	YEM 10	)(
P.	009049941.1	Y D D V L R K S I L F Y E A Q R S G K L P A N N R I P W R G D S G L K D G S D V G V D L T G G W Y	DAG	DHV	FGFPMAYSTAILALGL	LQWGDAYKASKQEDHMYDCIKWPLD	YFL 10	)(
(P	009058133.1	Y D D V L Y K S I L F Y E T Q R S G K L P A N N R I P W R G D S A L K D G S D V G V D L T G G W Y	DAG	DHV	FGFPMAYSSAILTWGL	LEWKDAYVKSGQLKWLYESIKWPLD	YFL 1	R
۹P.	009058134.1	YSKPLANSILFYNAGRSGKLPANNPIKWRGDSALKDGSDVGVDLTGGWY	DAG	DHA	SFGLPMAATAHLLLWSY	IUWPDAYSSTSLTDKFFDMIKWPLD	YFL 10	JC

			6		
XP 009060290.1	YSKALGNSILEYDAQRSGKLPANNPI	KWRGDSALKDGSDVGLDLTGGWY	NGDH V	GEGLEMAASTTNLLWGYLOWSDAYKAAGLTDKEEDMIKWELDYEL 1	00
XP_009064976.1	YDDILRKSILFYEAQRSGKLPGNNRI	PWRGDSGLEDGSDVGLDLTGGWY	DAGDHV	KFGFPMAYSTAILTMGLLQWKDAYQTSRQLDHMYSCIKWPLDYFI 1	00
XP_009064980.1	YGDVVQKSILFYYAQRSGKLPAYNPV	/ P W R G D S A V N D Q G D N G E D L S G G W Y	DAGDLV	K FNMPASSAVRVLLSGIVRWKDAYQALGQLDLALDSVKWELDYFL 1	00
XP_009065702.1	YGDALGKSILFFDAQRSGKLPSNNPI	KWRGDSALHDKGSHGEDLTGGWY	DAGDNV	KFNLPMASSTTLLLWGFIQWKDAYQSSHQINQMYDSVKWPLDYFL 1	00
XP_011057333.1	YAQVLELSLLFYEAQRSGKLPKNNR	SWRGDSALYDRGSNGEDLIGGYY		REGETMAETTTLLAWGTVNWPEAYDAAGQLDEIRRAIKWSTDYFI 1	00
XP_011253653.1	YAKVLNLSLMEYEAORSGKLPRENRI	PWRNDSALHDAGWEDEDLTGGYY	DSHNYV	AFGETMAETTTILAWGAVSWPEAENVANOLDEMHKTIKWATDYFI 1	00
XP_011253654.1	YAKVLNLSLMFYEAQRSGKLPRENRI	PWRGDSALYDSGWENEDLTGGYY	D G H N Y V	K FGFTMAFTTTILAWGAVSWPEAFNVANQLDEIRKTIKWSTDYLI 1	00
XP_011253655.1	YGEVLNRSLQFYEAQRSGKLPRENRI	PWRGDSALHDVGLEGEDLTGGYY	DGHNYV	KFGFTMAFTTTILAWGAVSWPEAFNVANQLDEIRKAIKWATDYFI 1	00
XP_011296914.1	YAKVLELSLMFYEAQRSGELPPDNRI	PWRGNSALDDRGLHGEDLTGGYY	DAGDEV	K FGFTMASTTTLLAWGVVSWPEAYEAAGQLDQVRKSIKWATDYFI 1	00
XP_011296916.1	YAKVLELSLMFYEAGRSGELPPDNRI	PWRGNSALDDRGLHGEDLTGGYY	DAGDEV	KFGFTMASTTTLLAWGVVSWPEAYEAAGQLDQVRKSIKWATDYFI 1	00
XP_011421782.1	YDEVLMKS LLEYEAORSGKLPDDNRI	PWRGDSSLIDAGDNGEDLTGGWY	DAGDNI	EGEPMAASTTILLTWGLLRYKDAYOHSGOLEHMYSCIRWPLEWML 1	00
XP_011423159.1	YNEVMMKSILFYEAQRSGKLPPTNRI	PWRGDSALTDHGDNGEDLTGGWY	DAGDHI	KFGFPMASSTTLLAWGLLEYKDAYEHSGQLENMYDCIRWPLEWML 1	00
XP_011430172.1	YSEALSISILFYDAQRSGVLPADNPI	PWRSDSAVDDHGDYGEDLSGGWY	NAGDHV	KFNLPMASSAHLLSWGLEIWKDAYQSAGQLDMMYDMLKWPLDYFL 1	00
XP_011447501.1	YGQVLEKSILFYEAQQSGDLPPTHR	NWRGDSALGDSGIYGEDLTGGWY	DAGDGV	KFNLPMSASVINLAVGLYKWRDAYSASNQLQHMYDSIKWPLDYFL 1	00
XP_011447502.1	YGLVLEKSILFYEAQQSGDLPPTHRI	NWRGDSALGDSGIYGEDLTGGWY	DAGDTV	KENLPMSASVINLAVGLYKWKDAYSASNQLQHMYDSIKWPLDYFL 1	00
XP_011498432.1 XP_011832703.1	YARVLELSLLFYEAURSGKLAPDNRI YAOVIGISIIEVEAOPSGKIPKNNPI	AWRGDSGLDDRGDDGEDLIGGYY	DAGDEV	NEGETMASTITLLAWGAVSWPEAYARLGQLDVIREALKWAIDYFT 1 Negetmaetttilawgavkwpesykaagoideipptikwstdyel 1	00
XP 011700378.1	YAQVLGLSLLFYEAQRSGKLPRNNRI	SWRGDSALYDRGSNGEDLTGGYY	DAGDEV	SFGFTMAFTTTLLAWGTVSWPEGYDAASQLDEVRRAIKWSTDYFI 1	00
XP_011867251.1	YAQVLELSLLFYEAQRSGNLPKNNRI	PWRGDSALYDRGSNGEDLTGGYY	DAGDEV	FGFTMAFTTTLLAWGAVNWPEAYDAAGQLDEVRRAIKWSTDYFI 1	00
XP_012061951.1	YAQVLELSLLFYEAQRSGKLPKNNR	SWRGDSALYDHGSNGEDLTGGYY	DAGDEV	K FGFTMAFTTTLLAWGTVNWSEAYDAADQLDEIRRAIKWSTDYFI 1	00
XP_012146513.1	YAEVLELSLLFYEAGRSGKLPRDNRI	PWRGDSALNDRGLNGEDLTGGYY	DAGDEV	K F G F T M A S T T T L L A W G A V S W P E A Y N A A G Q L D E L R K A I K W A T D Y F I 1	00
XP_012224014.1 XP_012248155.1	YARVIELSLIFTEAGRSGKLPKDNRI	PWREDSAL DDRGLNGEDL TGGTT	DAGDEV	AFGETMASANIVLAWGVDSWPQATRAAGQLDEVRRAIKWASDIFI I Regetmastitiiawgavswpeaynaagoideirkaikwatdyfi 1	00
XP 012273697.1	YARVLELSLLFYEAGRSGELPASNRV	PWRGNSAMNDRGLDGEDLTGGYY	DAGDEV	SFGFTMASTTTLLAWGAVSWPEAYELAGQTEMLRNAIKWATDYFI 1	00
XP_012526654.1	YAQVLELSLLFYEAQRSGKLPKNNR	SWRGDSALCDRGSNGEDLTGGYY	DAGDEV	KFGFTMAFTTTLLAWGAVNWPEAYDAAGQLDEVRSAIKWSTDYFI 1	00
XP_012935745.1	YNDLLYKSILFYEAQRSGKLPNNNH I	PWRGDSALGDKGAKGEDLTGGWY	DAGDHV	KFNFPMSYSTTILTWGYLLYPDAYRAAGQESKILDCVRWPLEYLL 1	00
XP_012938497.1	YG-ALGLSIKFYMAQRSGPI-HDSIV	PWRGNSGMNDC TVGGWY	DAGDNV	K F T F P M A T S A H L L M W S Y Y K Y K D G Y V K A N Q A T Q F L N M I K T P L N F F L 9	12
XP_012939931.1 XP_012939933.1	Y-OALSKSILEYEADRSGPL-NDHIL	PWRHSSAMHDC TVGGWY	DAGDHV	REGIPMA DSA HILLWSLTREK DGYTOAGOI NOMYHMYK TPINYEK 9	12
XP 012941225.1	YRDSLAKSMLFYFVQRSGPIDD - DI	PWRKDSATNDGSDVGKDLVGGWY	DAGDHV	AFGFPMAASATTLLWGFLQWRTSYDAMDKTEMMLELVKHPLDYFL 9	99
XP_012945520.1	YNDILFKSFLFYEAQRSGKLPASNRI	PWRGDSALNDKGENGEDLTGGWY	DAGDHV	KFNFPMSYSTTILAWGYLLYPEAYKAAGQEDALLDGIKWPLEYLL 1	00
XP_013066277.1	YS-ALGKSILFYNAQRSGQLPGNNPI	SWRSHSALDDC VLGGWY	DGSDHL	KFGLPMAAATHLLLWSVYKYKDAYISAGQLTQMYDMVKWPLDYFT 9	13
XP_013066278.1	YG-ALEKSILFYNAORSGKLPANNPI	PWRGDSALNDC VVGGWY		FGLPASFATTALLWSLVRFKDGYQRAGQLDQMYNMIKWPLDYFL 9	13
XP_013066280.1	YA-ALSKSILFFDTORSGKLPADNPI	SWRGDSALNDC VTGGWY	DAGDHV	KETLEMSASTTVLLWSLNREKDAYSSSGTLDOMYDMVKWPLDYFL 9	33
XP_013066283.1	Y - DALKKSILFFDAQRSGKLPADNPI	SWRGDSALNDC VMGGWY	DAGDHI	K FTLPMSASTTILLWSLNRFKDAYSSSGTLDQMYDMVKWPLDYFL 9	13
XP_013066264.1	YA-ALSKSILFLNTQRSGKLPINNPI	PWRGDSGLNDC VVGGWY	DAGDHV	SFGQPMASAATILLWSLYRFKDAYSSTATLDSMYDMIKWPLDYFL 9	3
XP_013066285.1	YA - ALAKSIL FYDAQRSGKLPGSNPI	HWRGDSALNDC VVGGWY	DAGDHV	K FGLPASAATTVLLWSLNRFRDGYQKAGQLNQMYDSVKWGLEYFL 9	13
XP_013066286.1	Y - DALGKSILFYDAQRSGKLPSNNPI	SWRGDSALGDCVVGGWY		K F G L P A S A A A T V L L W S L N R F K D G Y Q T A N Q L D Q M Y D M I K W G L D Y F L 9	13
XP_013086370.1	YT-ALGKSILEYEAORSGKLPANNEN	PWRGDSALNDC VPGGWY	DAGDHV	AFGLEMASSTTELLEWSLIKEKDGYVOSGOINNAMDOVKWTIDYFI 9	13
XP_013069169.1	YNDVLYKSILFYETQRSGKLPANKRI	PYRGDSALGDKGDAGEDLTGGWY	DAGDHV	KENFPMAYSTAILAWGYLMYADAYKAAGQDTYLLDSIKWPLDYLL 1	00
XP_013069170.1	YNDVLYKSILFYETQRSGKLPANKR	PYRGDSALGDKGDAGEDLTGGWY	DAGDHV	KFNFPMAYSTAILAWGYLMYADAYKAAGQDTYLLDSIKWPLDYLL 1	00
XP_013069521.1	YNDVLQKSILFYEAQRSGKLPANKRI	PWRFDSALGDKGEAGEDLTGGWY	DAGDHV	KFNFPMAYSTTILTWSYLLYADAYESAGQTDYMLDCIKWPLDYLL 1	00
XP_013069522.1	YNDVLQKSILFYEAQRSGKLPANKRI	PWRFDSALGDKGEAGEDLTGGWY	DAGDHV	KENEPMAYSTTILTWSYLLYADAYESAGQTDYMLDCIKWPLDYLL 1	00
XP_013069523.1 XP_013064279.1	YNDVLEKSILEYEAORTOKI PATNRI	PYRGDSLLODKGDGGEDLSGGYE	VEDENV	KENEPIAYSATII SWGYLLYEDAYKAAGOEDHIIESIKWALDYLL 1	00
XP 013084281.1	YNDVLFKSILFYEAQRTGKLPATNRI	PYRNNSLVGDKGDGGEDLSGGYF	GHGHV	KENFPIAYSATILSWGYLLYEESYKAAGQEDHVLESIKWALDYLI 1	00
XP_013084282.1	YNDVLFKSILFYEAQRTGKLPATKRI	PYRDDSCLTDKGDNGEDLSGGYF	LAHAHV	KFNFPMAYSATILSWGYLLYEDAYKSAAEDDHVLESIKWALDYLI 1	00
XP_013084283.1	YNDVLFKSILFYEAQRTGKLPATKRI	PYRDDSCLTDKGDNGEDLSGGYF	LAHAHV	KFNFPMAYSATILSWGYLLYEDAYKSAAEDDHVLESIKWALDYLI 1	00
XP_013084285.1	YNDLLYKSILFYEAGRSGKLPANNRI	PYRGDSALNDTGDNHEDLTGGWY		KENERMAYSVTVLAWGYIMYPEAYKSAGQEDYLLDSIKWPLDYLL 1	00
XP_013084281.1	YA - ALGKSILFYDAORSGPLPANNPI	HWRGDSAVHDC VVGGWY	DAGDHV	FGLPMAATANILLWSLYKWKDAYORANOLNOMYDMIKWPLDYEL 9	33
XP 013094438.1	Y - EALAKSILFYDAQRSGKLPANNPI	KWRGDSAVNDC VTGGWY	DAGDHV	K FGLPLASSTTLLLWGFVSFKDGYVKAGQLDAMYDMIKWPLDYYL 9	33
XP_014214612.1	YATVLGMSLMFYEAQRSGKLTPDNR	PWRGDSGLNDRGDNGEDLTGGYY	DAGDEV	KFGFTMASTTTLLAWGAISWPGAYERAHQLEAIRAAVKWATDYFV 1	00
XP_014601854.1	YARVLELSLLFYEAQRSGKLPEDNRI	SWRGDSALTDHGINGEDLTGGYY	DAGDEV	FGFTMASTTTLLAWGAASWPEAYVEAGQLDEIRKAIKWATDYFI 1	00
XP_014786694.1	YQEILLKSMLFYEAQRSGYLLPNRRI	RWRGHSALNDRSEFGEDLTGGWY		REGETMA SATTILLWGLVEFREAYEKTGLIKDMLDCVRWPLDYFL 1	00
XP_015184471.1	YARVLELSLIFTEAGRSGELFGDNRI YARVLELSLIFTEAGRSGELFGDNRI	SWRGDSALNDHGINGEDLTGGYY	DAGDEV	AFGETMASTTTLLAWGAASWPEAYVEAGOLDEURKSIKWATDYFL 1	00
XP_015432538.1	YAQVLELSLLFYEAQRSGKLPKNNRI	QWRGDSALDDHGLNGEDLTGGYY	DAGDEV	K FGFTMASTTTLLAWGAVSWPEAYDAAGQLEKLRDAVKWATDYFI 1	00
XP_015513382.1	YFEVLERSLLFYEAGRSGRLPKNNRI	PWRGDSGLEDRGQNGEDLTGGYY	DAGDEV	K FGFTMASMTTLLAWGAVSWPQAYEAAGQLDELREAVKWATDYFI 1	00
XP_015592927.1	YANVLKLSLLFYEAQRSGKLTDDNRI	SWRGDSGMNDRGLNGEDLTGGYY	DAGDEV	K FGFTMASTTTLLAWGAVSWPEAYAAAGQLEELRRAIKWATDYFM 1	00
XP_015836219.1	YRHALKLSLLFYEAQRSGPLPASNRI VADVIELSLIEVEAODSCKIDENNDI	TWRSDSALDDHGLNNEDLTGGYY	DASDYV	KESETMAETTTILTWGMLSERHAYNEAGQYQHGLDAIKWATDYFI 1 KECETMASTTTILAWCAYSWBEAYNAACOLDELBKAIKWATDYFI 1	00
XP 017766299.1	YARVLELSLLFYEAGRSGKLPENNRI	PWRGDSALDDRGLNGEDLTGGYY	DAGDEV	FGFTMASTTTLLAWGAVSWPEAYNAAGQLDELREAIKWATDYFI 1	00
XP_017785110.1	YKEALKLSLLFYEAQRSGILPESNRI	SWRSDSALNDQGIKGEDLTGGYY	DAADEV	FGFTIAYSMTILAWGLVSYEETYRRIGQHNATAEAIKWGTDYII 1	00
XP_017793200.1	YARVLEFSLLFYEAQRSGKLPENNR	PWRGDSALVDRGLNGEDLTGGYY	DAGDEV	K FGFTMASTTTLLAWGGVSWPEAYNAAGQLDELRKAIKWATDYFI 1	00
XP_018016806.1	YKEALCMSFLFYEAQRSGALPPDQR	TWRHDSALSDGLDNGIDLTGGYY	DAGDHV	KFNFPMAFVMTALSWGLLSYREGYEAAGQVAHAEAAIKWTVDYFL 1	00
XP_018016809.1 XP_018016810.1	VKOAL CMSTLEVEANRSGDLPSSQR	DWRGDSAL DDGSDVGVDL TGGVY	DAGDHV	NEGYDMAETTTTLGWALLSEKAGEDKAGQIGYAMDAVKWATDYEL 1 NEGYDMAYTVTVLSWGLLSYRAGYETAGOVEYVEAATRWGTDYEL 1	00.
XP 018016813.1	YRELLCHALVFYEANRSGYLPSDOR	/ EWRGDSALNDGADNGVNLEGGYY	DAGDLV	SFGFPLAYTVTFLSWGYITYGEGYVQAGQTQYLEKTIKWATDFIL 1	00
XP_018016820.1	YRELLCOSLVFYEANRSGKLPSDOR	/ EWRYDS AVDDGAADGVDLEGGYY	DAGDEV	KFAFPMGFTMTFLAWGYISYEDGFVQAGQVEYFEKTLRWATDYFL 1	00
XP_018016821.1	YRELLCQSLVFYEANRSGKLPSDQR	/ EWRYDS AVDDGA ADGVDLEGGYY	DAGDEV	K FAFPMGFTMTFLAWGYISYEDGFVQAGQVEYFEKTLRWATDYFL 1	00
XP_018026496.1	YPQVLCMSILFYEAQRSGRLPADQRI	TWRWDSALGDGGDVGVDLTGGYY	DAGDHV	K F G L P L A Y T G T V L A W G A I A Y E A G M R S A G E Y D N A R A A V R W V A E Y L L 1	00
XP_018027134.1 XP_018044452.1	YAOVI ELSI I EYEADRSGKI PKNNRI	SWEGDSAL YDEGSDOGEDL TGGYY	DAGDEV	NFGFFMATSVIVLSWGLLSTRAGTEAAGQVIAAENATRFGIDTFL 1 KEGETMAETTTIIAWGTVNWSEAYDAADOIDEIRRAIKWSTDYEL 1	00
XP_018307051.1	YGQALKFSLLFYEAQRSGKLPKNNRI	SWRGDSALYDRGSNGEDLTGGYY	DAGDEV	FGFTMAFTTTILAWGTVNWPEAYDAADQLDEIRRAIKWSTDYFI 1	00
XP_018323591.1	YKEALSLSLLFYEAQRSGPLPPDNR	PWRGDSALHDTGQNGEDLSGGYY	DASDEV	KFGFTMTFSTTLLAWGLLSYKDAYKYAGQYKYGLEAVKWATDYFV 1	00
XP_018347393.1	YAQVLELSLLFYEAQRSGKLPKNNR	LWRGDSALYDRGSNGEDLTGGYY	GDFV	AFGFTMAFTTTLLAWGTVNWPEAYDAAGQLDEIRRAIKWSTDYFI 1	00
XP_018359268.1 VD_018306635.1	YAQVLELSLLFYEAQRSGRLPKNNRI VAKVLELSLLEVEAORSGRLPKNNRI	SWRGDSALYDRGSNGEDLIGGYY SWRGDSALYDRGSNGEDLIGGYY		REGETMAETTSIIAWGTVNWPEAYDAAGQLDEIRRAIKWSIDYEL 1	00
XP_018568194.1	YKEALKLSLLFYEAQRSGKLPSNNRI	KWRGDSALNDRGVNGEDLTGGYY	DGSDFV	GETMA FATTIL SWGVIKYKDA YQAADQYTEVLDA I KWSTDYFI 1	00
XP_018568196.1	YKEALKLSLLFYEAQRSGKLPSNNRI	KWRGDSALNDRGVNGEDLTGGYY	DGSDFV	K FGFTMAFATTILSWGVIKYKDAYQAADQYTEVLDAIKWSTDYFI 1	00
XP_019884179.1	YYKVLKLSMLFYEAQRSGEL - RNNRI	PWRDDSALQDAGWEDEDLTGGYY	DGHNYV	K FGFTMAFTTTILAWGAVSWPEAFNVANQLDEMHKTIKWATDYFI 9	19
XP_020289295.1	YARALELSFLFFEAQRSGKLPKDNR	PWRGDSAMNDRGHDGEDLTGGYY	NGRNYI	KYTLPTAFTTTMLAWSVLTWPEGYKEAGQLDEAYKAIKWATDYFL 1	00
XP_020269297.1 XP_020289301.1	Y. ALMISIWEEDAORSGKMPLNNRI	GERHDSMLHDHGYEGEDLSGGYY	VICUSIVV	NYVEPTAETTTILAWGULTWIEATRNAGGLEIVRENIRWASDIFI I	00
XP_020289336.1	YAKALSLSLLFFEAQRSGKLPTTNRI	PWRYDSLLKPTSH LNLTGGYY	NGINYV	KSTEPTAETTTMLAWGLLTWTEGEYLAGQLNEIYGTIKWATDYEI 9	8
XP_020614867.1	YDSVLHDSILFYEAQRSGKLPISNR	/ PWRKDSALNDRGQNGEDLSGGWY	DAGDYV	K FGFPMAFSVTMLSWGLVEYRDAYVDAGELNYMLDCIKWPLDYFM 1	00
XP_020893295.1	YDQVLHDSILFYEAQRSGKLPSTNR	PWRGDSALNDKGQNGQDLTGGWY	DAGDEV	K F G F P M A S S V T V L A W G V V E Y R D A Y K A A G E L D N V L D S I K W A T D Y F I 1	00
XP_020893299.1 XP_021345083 4	YORVIEDSILEVEAORSGKLPSTNRI	HYRKDSALNDKGQNGQDLTGGWY	DAGDEV	REFERENCESSVIVLAWGVVEYRDAYKAAGELDNVLDSIKWATDYFI 1	00
XP_021361084.1	YAEALRUSILFYAAQRSGKUPANNPI	PWRGDSALQD - GHDHTDLTGGWY	DAGDLI	KENFPMASTTTLLVWGLIRWPDAYQAAGQLDNMYDSVKWPLDYFL 9	99
XP_021363481.1	YAEVLEKSILFYEAQRSGKLPPTNRI	PWRGDSAMNDRGQNGEDLTGGWF	DAGDHI	KFALPLASSVSMLAWGLVEFKDAYVGAGQLDNMYDSIRWPLEWMM 1	00
XP_021363490.1	YGNALGLSILFYDAQRSGRLPGNNP	PWRSNSATGD - GSDGTDLSGGWY	DAGDLV	KFNFPMASSATILAWGLSRWKDGYEAAGQLEMMYDCLKWPLDYFL 9	19
XP_021363495.1	YDELLEKSIL FYEAGRSGKLPPTNRI	PWRGDSALNDHGQSGEDLTGGWY	DAGDHI	AFGLPMSSTVTLLAWSMLRFKDAYEHAGQLEYAYDCIRWPLEWML 1	00
XP_021363496.1 XP_021363497.1	YDEVLEKSILFYEAORSGKIPPINRI	SWRGDSAMNDOGANGEDLIGGWY	DAGDHV	FGLPMSASVTMLAWGFLOYADAYENAGOTFYMYDCIRWPLEWML 1	00
XP_021363498.1	YDEVLEKSILFYEAQRSGKLPPNNRI	SWRGDSAMNDQGANGEDLTGGWY	DAGDHV	KFGLPMSASVTMLAWGFLQYADAYENAGQTEYMYDCIRWPLEWLI 1	00
XP_021924915.1	YSRVLELSLLFYEAGRSGQLPPDNR	SWRGDSALGDRGQNGEDLTGGYY	DAGDEV	KFGFTMASTTTLLAWGYFSYRDAYESAGQAKYALDAIKWAADYFI 1	00
XP_021935277.1	YAEVLRDSLLFYEAQRSGKLPPDQK	TWRKDSALNDKGENGEDLTGGYY	DSGDYV	FGFAMAYTTTVLAWGLVDHEGGYTSSGALDDGLKAIKWATDYFL 1	00
XP_021941324.1 XP_02105654P 4	YEEVLGISIEEVEADSGDIDESND	MYRGDSELDDK DGDIDIVCOVE	DAGDYV	NEGEPEAYETTVINWGIVDYKOGYESAGELONALDAVKWSLOVELO	UÜ 20
XP_021956549.1	YEEVLGLSLEFYEAQRSGRLPESNRI	MYRGDSFLDDK-DGDIDLVGGYF	DAGDHV	FGFPFAYFTTVLNWGIVDYKQGYESAGELQNALDAVKWSLDYFI 9	19
XP_021959948.1	YRQVLEKSLLFYEGORSGRLPANNR	PWRGDSMLED - GDGDVDLTGGYF	DAGDHV	K F N F P Q A WALST I A M G G L D F K E G F I A S G Q L N N L E A Y K W G A D Y Y L 9	19
XP_021966378.1	YASVISKSLLFYEAGRTGRLPADNR	PWRGDSFLDDKGLSGEDLSGGYF	DAGDEV	K FGFPLASAMSMLAWGMVDSVEGYKKAGEWENALASLKWGMDYMI 1	00
XP_021967587.1	YAEVLRUSNUFYEANRSGRUPOSNRI	SWRSNAFVDDGQREGIDLDGGYF	DAGDHV	KEGEPMAGETTILNWGLLDWRNGYDKANDYANSLSTVKWSLDEFI 1	00
XP_02210/20/.1 XP_022167268.1	YDEVLRLSLMFYEAORSGHLPANNRI	PWRGDSATTDKGONGEDLSGGYY	DAGDEV	A FGFTMASTTTILAWGYLSYKDAYEEVNEWENALGSIRWAAEYFI 1	00
XP_022200004.1	YSKVVHLSLLFYEAGRSGQLPANNR	PWRGDSSLNDRGLKGEDLTGGYY	AGDTV	REGETMASTTTLLAWGCLSYKDAYKAAGEWKEVLNALKWATDYFI 1	00
XP_022328778.1	YPLVLEKSILFYEAQQSGALPSNHRI	DWRGDSALGDSGVNGEDLTGGWY	DAGDGV	KFNLPMATSVINLALGLYQWKDSYAAANQLENMYNSIKWPLDYFL 1	00
XP 022329668.1	YPLVLEKSILFYEAQQSGALPSNHRI	. DWRGDSALGDSGVNGEDLTGGWY	DAGDGV	FNLPMATSVINLALGLYQWKDSYAAANQLDNMYNSIKWPLDYFL 1	00

VP 022226001 1		100
XP_022336153.1	YDEVLMKSILFYEAQRSGKLPDSNRIPWRGDSSLDAGNNGEDLTGGWYDAGDNIF FGFPMAASTTMLTWGLLRYKDAYGNSGELENMYACVRWFLEWML	100
XP_022338490.1	YGNALGLSILFFDAQRSGKMPASNPIPWRHDSAMNDHGDGGVDLTGGWYD <mark>DAGDHVK</mark> FNLPMAFSAHVLAYGLNRWKDGYSSSNQLNNMYEMLRTPLDYFM	100
XP_022338491.1	YSEALSLSILFFDAQRSGVLPANNPIPWRNDSAGDDHGDNGEDLSGGWYNAGDHMKFNLPMASSAHLLSWGLERWKDAYQNAGQLEMWYDMLKWPLDYFL	100
XP_022341767.1 XP_023337121.1	T DEVEM NSTEFTEAURS GREPANNREFTER BOS ALL DUGDINGED EGWIPD GOH VAR GEFTMAAS TILLAWGALDET KDATEGS GUETMT DET KWAT YN EVILLSNEFY OA ORS GNEDAF GMIPY GNS AS LODGS DNGVNES GY YV DIG GDEVY KEFTPMAAATTEL AWGALD FKG YEAS GEYNATEN TI KWATDYFI	100
XP_023704929.1	YKEVLRDSLLFYEAQRSGKLPADQKVTWRKDSALNDKGONGEDLTGGYYDAGHYVKFGFPMAYTATVLAWGLVDNQAGYSAAGALSDGLKAVKWATDYFI	100
XP_023704930.1	YKEVLRDSLLFYEAQRSGKLPADQKVTWRKDSALNDKGQNGEDLTGGYYDGGDFYGFGFPMAYTITYLSWGVVDHEIGYKDANAVDDARKAIKWGTDYFL	100
XP_023716596.1	YSHVLELSLLFYEAORSGKLPTDNRVSWRODSALDDRGONGEDLTGGYYDAGDFVRFGFTMASTTTLLAWGYLSYROAYESAGQARYGLSTIKWAADYFI VAOALDSLLFYEAORSGKLDDDNGYWWBLNSSLDNDRGONGEDLTGGYYDAGDFVWFGFDMASTTTLLAWGYLSYROAYESAGQARYGLSTIKWADYSE	100
XP_396791.3	YARVLELSLLFYEAQRSGKLPENNR I PWRGDSALDDRGLDDRGEDLTGGY I DAGDFVAGDFVAASTTTLLAWGAVSWPEAYNAAGQLDELRKA I KWATDYFI	100
AAA23086.1	YAEALQKSMFFYQAQRSGDLPADFPVSWRGDSGLTDGADVGKDLTGGWYDAGDHVKFGFPMAFSATMLAWGAIESPTGYSKAGSLDELKDNLRFVSDYFV	100
AAD38027.1 AAE80584.1	KAH TA INE FYGQVGEGN I DHAYW - KPEDMI MSKPSTKI DQSKPGSULAGELAAALAAASTVFKSSUSSTSSUVLSVAKELTEFADQHKU TYIAATUDAAS KAHVSENE FYGDVGRGD I DHAWWGBPE DMTMERBAYKI DA OHDGSDLAGELAAALAAASTVFKSSUSTSSUVLSVAKELTEFADQHKU TYIAAD	200
AAF80585.1	KAHVSEHEFY GOVGEGNLDHN SWGRPEDMTMERPAYK I DE GNPGTELAAETAAALAAAS I VFKSVDPSY SNTLLTHAKQLYDFGDNFRGKYSES I NDAQQ	200
AAK12339.1	KAHTAATELYGQVGDGDADHAYWGRPEDMTMARPAYKIDASRPGSDLAGETAAALAAASIVFKGVDSSYSDNLLAHAKQLFDFADNYRGKYSDSITQASN	200
AAO61672.2	KAHTATNEFY GQVGEGNTDHAYW-RPEDMTMSRPSYK I DOSRPGSDLAGETAAALAAASI V FKSSDSSYSSQVL SVAKELYE FADOHRDI Y TNA I TDAAS	199
AAX92641.1	KAHVSDNEFY GOVGDANTDHAWWGRPEDMTMERPAWSISPSAPGSDLAAETAAALAAGYLVFRDLDAAFANNLL HHARTLYDFALNNEGI YSOSISNAAG	200
ABD24274.1	KCHVSDNVLYVQVGDGGVDHGSWGRPEDM - MARPAFKIDASKPGSEVAMETAAAFAAGHLAFKEKDPSYSAKLLQHAKSLWQFAVTHKGKYSDSVSAAAG	199
ABD24275.1	KCHSSDDVLYVQVGDGGADHGSWGRPEDM - MARPAFKVDASRPGSEVAMETAVAFVTGHLAFKEKDPPYSAKLLQHAKSLWEFAVTHKGKYSESVSAAAS	198
ABD24276.1	KCHSSDDVLYVQVGDGGADHGSWGRPEDM. MARPAFKVDASRPGSEVAMETAVAFVTGHLAFKEKDPPYSAKLLQHAKSLWEFAVTHKGKYSESVSAAAS VCHASDDVLYVQVGDGADHGSWGRPEDM. MARPAFKVDASRVGSEVAMETAVAFVTGHLAFKEKDPPYSAKLLQHAKSLWEFAVTHKGKYSESVSAAAS	198
ABD24278.1	KCHVSDNVLYVQVGDGGVDHGSWGRPEDM - MARPAFKIDATKPGSEVAMETAAAFAAGHLAFKEKDPSYSAKLLQHAKSLWQFAVTHKGKYSDSVSAAAG	198
ABD24279.1	KCHVSDNVLYVQVGDGGVDHGSWGRPEDM - MARPAFKIDATKPGSEVAMETAAAFAAGHLAFKEKDPSYSAKLLQHAKSLWQFAVTHKGKYSDSVSAAAG	198
ABD24280.1	KCHASDNVLYVQVGDGGADHGSWGRPEDM - MARPAFKIDATKPGSEVAMETAAAFAAGYLAFKEKDPSYSAKLLQHAKSLWQFAVTHKGKYSDSYSAAAG	198
ABU24281.1 ABO26609.1	KCHVSDNVLVQVGDGGVDHSSWGRPEDM - MARVARKIDA IKPGSVAMBI AMAFAAGHLAFKEKUPSTSAKLLQHAKSLWQFAVIHKGKISUSVSAAAG KCNPVKKEVVNOVGESSI DSTEMGEPEDII KVEKVTASKPGSDVABI TAAAFAAGHLAFKEKUPSTSAKLLQHAKSLWQFAVIHKGKISUSVSAAAG	198
ABV32557.1	KAHVSANELY GQVGNGGADHSWWGRPEDMNMDRPAYK I DISRPGSDLAAETAAAMAAASI VIKNADSNYANT LLRHAKELYN FADNYRGKYSDSI SDAAA	200
ACE75510.1	KAHVSDNEFYGQVGDANTDHAWWGRPEDMTMERPAWS I SPSAPGSDLAAETAAALAAGYLVFRDLDAAFANNLLAHSRTLYDFALNNRGI Y SQS I SNAAG	200
ACE75511.1	KAHVSDNEFYGOVGDGNIDHSYWGRPEEMTMERPAWSITPSNPGSDLAAETAAALAAGYLVFRDLDAAYASTLLDHSRTLYDFAYNNRGIYSDSIPNAGE	200
ADB12463.1 ADB82658.1	KANTAATETI YOOVGEGEADHATWIGKPEDMI MAKPATKI DAAKPGSDLAGE LAAALAAASI VEKAVDESI SANLL DKAKQLEDFADKHKGKI SDSI TDANS KANTAETEI YOOVGEGE DHATWIGKPEDMI MAKPATKI DASNPGSDLAGE TAAALAAASI VEKAVDESI SANLL DKAKQLEDFANNYKKI SDSI TDANS	200
ADB85440.1	KCVQGPTTLWGQVGDPDVDHAYW-RPEDMTETRVSYKIDASAPGTDLAAETAAALAGAAVVHSGSYASECVSAAEQMYTFADEHRGKYSNSIDKAAN	196
ADB85441.1	KCHAGSKVLYGQCGDGDIDHEYW-RPEEMTMSRPSYKIDEQNPGADLAAETAAGLAAASIVFQGTDASYANQLLDFAKEMLELADEHQTSYVNSIP-AGD	198
ADB85442.1 AED33385.1	KAKTGINEFY GQVGAGNIDHSFW - RPEDMIED RPCWKIDI TAPGIDLAAETAAALAAASIVFQD - DRAYSSECLAAARELYDFADQVREKYQKSINAARG VAHTASNEFY GOVGODIDHSFW - RPEDMII SPDAEKIDIGVDGSDIAAETAAALAAASIVFRWINDESVSNNII I HAVOIFDEANNYDGVSDNIIDINN	198
AFQ98379.1	KGIPATETLYAQVGGGNDDHHFWGRAEDM RPAYKLTPSKPGSDVAGEIAASLAAGYLAFKERDAKYAATLL ITAKELPDFMINIKGKT906 I YSOSI SDAGQ	197
AFQ98380.1	KCIPKSQTLYAQVGEGNDDHHFWGRAEDMRPAYKLTPSKPGSDVAGEIAASMAAGYLAFKQRDAKYAATLLSTSKEIYEFGKKYPGTYSSSIQDAGQ	197
AGI61069.1	KCIPKSQTLYAQVGEGNDDHHFWGRAEDM RPAYKLTPSKPGSDVAGEIAASLAAGYLALKQRDAKYAATLLSTSKEIYEFGKKYPGIYSSSIQDAGQ	197
AGP76397.1	KAHTSQNEFYGQVGQGDLDHAYWGRPEDMTMARPAYKIDMSRPGSDLAGETAAALAASSIVFKNVDGTYSNNLLTHARQLFDFAKNYRGKYSDSITDARN	200
AGP76399.1	KANTSPLEFYGOVGOGDLDHATWORFEDMINARFAINIDMSRFØSDLAGETAAALAASSIGIKNUDGTISNULLTHARQLEPFAKNIRGKISDSIIDARN	200
AGP76400.1	KAHTAPKEFY GQV GQGD V DHAY WGRPEDMT MARPAYKIDT SPPGSDLAIETAAALAAPSIV FKYV DPNY SKNLLTHAKQLFD FANNY RGKY SDWITGARL	200
AGP76401.1	KAHTAPKEFYGQYGQGDVDHAFWGRPEDMTMPRPAYKIDTSPPGSDLAIETAAALAAPSINFKYVDPNYSQNLLTHAKQLFDFANNYRGKYSDWITGARL	200
AGP76402.1	KAHTSQKEFYGQVGQGVVHAYWGRPEDMTMARPAYKIDTSRPGSDLAGETAAALAAASIVFKNADCTYSNALLTHAKQLFDFANNYRGKYSDSITGARN	200
AGP76404.1	KANTSQKEFYGQVGQGDVDHATWGRPEDMTMARPATKIDTSRPGSDLAGETAAALAAASIVFKNADCTYSNALLTHAKQLPDFANNTKGKYSDSITGARN	200
AGP76405.1	KAHTDLTEFY GQV GRGDV DHARWGRPEDMT MERPAYK I DP SPPGSDLA I ETAAALAAPSI V FKYV DP NY SNNLLTHAKQLFDFANNY RGKY SDWI TGARN	200
AGP76406.1	KAHTAQNEFYGQVGQGDVDHAYWGRPEDMTMARPAYK I DTSRPGSDLAGETAAALAATS I V FKNVDGTYSNNLL THARQLFDFANNYRGKYSDS I TDARN	200
AGP76407.1	KAHTAQSEFYGQVGQGDQDHAYWGRPEDMIMPRPAYKIDMSRPGSDLAGEIAAALAAASIVFKNVDVIYSNNLLIHARQLFDFANNYRGKYSDSIITAAN YALTSOSEEYGQVGQGDVHAYWGBPEDMIMPRPAYKIDTSPBGSDLAGEIAAALAAASIVFKNVDVIYSNNLLIHARQLFDFANNYRGKYSDSIITAAN	200
AGP76409.1	KAHTSQLEFYGQVGQGVDHAYWGRPEDMTMARPAYKIDTSRPGSDLAGETAALAAASIVFNFVDGTYSNALLTHARQLFPFANNYRGKYSDSITGARN	200
AGP76410.1	KAHTSQKEFY GQV GQGD V DHA Y WGR P E DMT MAR PAYK I DT S R PGS DLA R E TAAALAASF I V F K NDDR T Y S NALL THARQL F D F AN NY RGKY S D S I T G A R N	200
AGP76411.1	KAHEAPNEFYGQVGEGDEDHAYWGRPEDMTMPRPAYKIDSSRPGSDLAYETAAALAATSIIYKYVDSTYSNILLTHARQLFDFANNYRGKYSDSITGARP	200
AGP76412.1 AGP76413.1	KAHTAPTEFY GQVGRGDFDHSY WGRPEOMT MSRPAFK I DASNPGSDLATE TAAALAATSI VYKTVDPSYSNNLI THAKOLFDFANNYRGKYSDSI TDARS VAHTTDTEFY GOVGRDFDHSY WGRPEOMT MSRPAFK I DASNPGSDLATE TAAALAATSI VYKTVDPSYSNNI I THAKOLFDFANNYRGKYSDSI TDARS	200
AGP76414.1	KAHTHATELYGQVGGGDVDHSYWGRPEOMTMSRPAFKIDASNPGSDLAAETAAALAATSIVFKTVDPSYSNDLITHAKQLFDFANNYRGKYSDSITDAKS	200
AGP76415.1	KAHTTSNELY GQV GEGDV DHSY WGRPEDMTMSRPAFKIDSSFPGSDLAAETAAALAATSIVY KTV DPSY SNNLITHAKQLFDFANNY RGKY SDSITDAKS	200
AGP76416.1	KAHTAQTELYGQVGEGDFDHSYWGRPEDMTMSRPAVKIDASIPGSDLATETAAALAAASIVFKTVDPSYSNHLLTHAKQLFDFANNYRGKYSDSITDANS	200
AGP76417.1 AGP76418.1	KAHTAPNETYGQVGQGDEDHAYWGRPEDMI MARPATKI DI SKPQSULAGEI AAALAASGI VERYVUSI I SNNLLI HARQLEDFANNYKGKYSUSI I GARP VAHTAPNEEYGDVGOGDEDHAYWGBPEDMI MARPAYKI DI SKPQSULAGEI AAALAASGI VERYVUSI I SNNLLI HARQLEDFANNYKGKYSUSI I GARP	200
AGP76419.1	KAHTAPTELY GOV GEGDFDH SY WGRPEDMT MSRPAFKI DA SNPGSDLATE TAAALAAASI V FKTVDPSY SNDLLT HAKQLFDFANNY RGKY SDSI TDANS	200
AGP76420.1	KAHTTATEYYGQVGEGDLDHSYWGRPEDMTMSRPAFKIDASNPGSDLACETAAALAAASIVLKTVDPSYSNNLITHAKQLFDFANNYRGKYSDSITDAAS	200
AGP76421.1	KAHTAATELYGQVGEGDIDHCYWGRPEDMTMSRPALKIDASKPGSDLADETAAALAATSIVFKTVDGSYSNNLITHAKQLFDFANNYRGKYSDSITDAKS	200
AGP76422.1 AGP76423.1	KAHTAPIEFYGQVGRGDVDHSTWGRPEDMIMERPATKIDISIPGSDLAGEIAAALAAASIVFREVDPNTSDNLLIHAKOLPDFANNTRGKTSDSIIDARL KAHTAPIEFYGQVGRGDVDHSTWGRPEDMIMERPATKIDISIPGSDLAGEIAAALAAASIVFREVDPNTSDNLLIHAKOLPDFANNTRGKYSDWITDARL	200
AGP76424.1	KAH TAQNELY GQV GRGDV DH SY WGRPEDMT MERPAYK I DT SNPGSDLAGE TAAALAAAS I V FKHV DPNY SDNLL THAKQL FDFANNY RGKY SDSI TDARL	200
AGP76425.1	KAHTAPNELY GQY GRGDY DHSY WGRPEDMTMERPAYK I DTSNPGSDLAGETAAALAAASI Y FKHY DPNY SDNLL THAK QLFDFANNY RGKY SDSI TDARL	200
AGP76426.1	KAHTAPNEFYGQVGRGDVDH IYWGRPEDMTMERPAYK ID TSNPGSDLAGETAAALAAASIV FKNVDPNYSDNLL THAKQLFDFANNYRGKYSDWI TDARL VAHTSONEEYGQVGODI HONWRDBEDMTMABDAYK ID TSPDGSDLAGETAAALAAASIV FKNVDPNYSDNLL THAKQLEDGANNYRGKYSDWI TDARL	200
AGP76428.1	KAHTSQHEFY GQV GQG DWDHDY WGR PEDMT MAR PAYKIDTSR PGSDLAGETAALAAASIVY KNVDGTYSNILLTHARQLLDFANNY RGKYSDSITDARN	200
AGP76429.1	KAHTSQNEFY GQVGQGDLDHDY WGRPEDMT MARPAYKIDT SRPGSDLAGETAAALAAASIVY KNVDGTY SNALLTHARQLFDFANNY RGKY SDSITDARN	200
AGP76430.1	KAHTAPTEFY GQVGRGDVDHAYWGRPEDMTMERPAYK I DTSPPGSDLA I ETAAALAAPSI V FKYVDPNYSNNLL THATQLFDFANNYRGKYSDWI TGARN	200
AGP76431.1 AGP76432.1	KAHISQNEFYGQVGUGD DHDYWGRPEDMIMARPAYKIDISRPGSDLAGEIAAALAAASIVSKNVDGIYSNNLLIHARQLPDFANNYRGKYSDSIIDARN KAHISQNEFYGQVGUGD DHDYWGRPEDMIMARPAYKIDISRPGSDLAGEIAAALAAASIVSKNVDGIYSNNLLIHARQLPDFANNYRGKYSDSIIDARN	200
AGP76433.1	KAHTSQNELYGQVGQGDLDHYYWGRPEDMTMARPAYKIDTSRPGSDLAGETAAALAAASIVSKNVDGTYSNNLLTHARQLFDFADNYRGKYSDSITDARN	200
AGP76434.1	KAHTAATELYGQVGEGDLDHAFWGRPEDMTMWRPAYKIDASNPGSDLAGETAAALAAASIVFKALDPSYSNDLLTHAKQLFDFANNYRGKYSDSITDASS	200
AGP76435.1	KAHTA TELYGQVGEGDWDHAFWGRPEDMI MRRPAYKI DASNPGSDLAGEI AAALAAASI VFKALDPSYSNDLLI HAKQLFDFANNYRGKYSDSI I DASS YALTAATELYGQVGEGDI DHYVMGBDEDMI MWRPAYKI DYSNBGSDLAGEI AAALAAASI VFKALDPSYSNDLLI HAKQLE DEANNYRGKYSDSI TDASS	200
AGP76437.1	KAHTAATELY GOVGEGDLDH YY WGRPEDMT MWRPAFKI DV SY PGSDLAGE TAAALAAASI V FKALDP SY SNDLLTHAK QLFD FANNY RGKYSDSI TDAKS	200
AG\$32241.1	KAHTAASEFY GQV GQGD V DHAY WGR PEDMT MDR PAYKIDAAHPGSDLAGETAAALAAASIV FRGVDSSY SEKLL DRAKQLFD FADKN RGKY SDSITDARN	200
AIT11911.1	KAHTSATELYGQVGDGDTDHAFW-RPEDMTMWRPSYKIDAGASQLAGETAAALAAASIVFQNSDASYSATLLSHARELFDFADQVRDVYTNSITAAAN	199
AKV16371.1	KAHTAATEFY GOVGOGN VDHASWIRFEDM - ARPAWKVIRI UPOSELVOEIAAALAAASMVFKSADVITSSKELSTARGUTDFANERKOST- DSIANAAN KAHTAATEFY GOVGOGN VARAYWGRPEDM TINSRPAYKID ISKPGSDLAATAIAYKN VOSSYANNLI THAKOLFDFANNYRDKYSDSITDARN	200
AMH40356.1	KCHVSEFEFYGQVGQGQLDHDFW-QPEDMTEERTPYKITAQAPGSDLAAETAAAMAAAYLVFKDVDATYAATLLTHAEQLMNFAVTYLGKYSDSITDAKN	199
AMH40357.1	KCHVSEFVFYGQVGQGQVDHNSW-QPEDMTEARPAYKITTSAPGSDLAAETAAAFAASYLVFKSADANYAKKLLVHAEQLMKFALNYKGKYSDSISDAKN	199
AMH40358.1	KCHVSEYEYYGQVGSGNVDHNYW - RPEDMAEARPAYK I TTSKPGSDLAAETAAAFAAASLVFKSSNPTYAKTLLTHAEQMYDFGYKYRGLFSDSI TDSKC VALVSENVEVAOVESCHI DHNYW BDEDMTEBDAYM NUTNDOSDL ACETAAAFAAASLVFKSSNPTYAKTLLTHAEQMYDFGYKYRGLFSDSI TDSKC	199
AMH40360.1	KAHVSANELYGOVGQGSPDHSYW-RPEDMTMARPAYKITASSPGSDLAGETAAALAAASIVLKSTDASYAANLLNHAKQLFDFANNHRGIYSDAISDASG	199
AMH40361.1	K CH V G D N K L Y G Q V G L G Q L D H D S W - R P E D M T M E R P A F M I N A S A P G T D L A A E T A A AMA A A S L V F K T V N S A Y S T Q L L Q H A R Q L F D F A N N Y R G K Y S D S I P D A N N	199
AMH40362.1	KCHVGENQLYGQVGLGQLDHVYW-RPEDMTMERPAFMINASAPGTDLAGETAAAMAAASLVFKTVNATYSATLLKHARQLFDFAYKYRGKYSDVIADANT	199
AMH40363.1 AMH40364 1	nanyseneli guvgugaadha tmi nfeumi marfatni i asarosulade i Aaalaaasi vikti anssyssi leunhakolin han thegi ysosi SDASS Kohvseyygo yogoson vihaywi reedmitearpayki i taskrosollaaftaafaaasi vikksin vyhysi van vii kiakomvyfovykymi i ensi tinsko	199
AMH40365.1	KAHVSENVFYAQVGSGY I DHQYW-RPEDLTMERPAYMVNVTNPGSDLAAETAAALAASYLVFKDVNATYANI LLTHAKQLYNFAYKNRGKYTDAFADVRC	199
AMH40366.1	KAHVSENVLYÄQVGSCTTDHNNW- RPEDMTEA RPAFMINTTHPGSDMÄÄETÄÄÄLÄÄGYMVFKDVNÄTYANTLLTHAKQLYSFÄYKYRGKYSDAFTDVEG	199
AMH40367.1	KCHVSEFEFYGQVGQQVDHNEW-QPEDMTEARPAYK I TASAPGSDLAAETAAAFAASYLVFKSVDANYAKKLLVHAEQLMKFATTYKGKYSDSI ADAKS	199
AMH40368.1	KCHVSEYEYYGQVGSGTVDHAYW-RPEDMTMARPAYKITASKPGSDLAAETAAALAAASLVFRSSNATYANILLQHAKQLYDFGYKYRGIFSNSITDSQC	199
AMH40359.1	KCHVBEFFYQQVGQQLDHDFWL PEDMINEERFAFMINAAAFOSDLAAEIAAMMAAASLVFKIAMAITAALLEHAAUQLFDANDYKQKYSDSIADAEIAAMMAAYLVFKSVDAITAALLIHAAQLMAAVAAVYKGKYSDSIADAEIAAMMAAYLVFKSVDAITAALLIHAAQLMAFAVYQGKYSDSIADAEIAAMMAAYLVFKSVDAITAAL	199
AMH40371.1	KAHVSDNVFYAQVGSGY I DHQTW - RPEDMTE I RPAYT I NVTNPGSDLAAETAAALAASYLVFKDVNATYAN I LLEHAKTLYNFAYKYRGKYTDAFDDVRC	199
AMH40372.1	KAHVSENELYGOVGOGSDHSYW-RPEDMKMARPAYKITASAPOSDLAGETAAALAAASIVFKTANSSYSATLLNHAKQLFNFADTHRGIVSESISDASG	199
AMH40373.1 AMH40374 1	nanyagnyr rayausuyaunn twinreum i carritmin i thrusul aaciaaacasti yy KULUSKYAQTLL THAKQLYNFAYTYRGKY SDAFDDVRC KAHVSENEL Y GOVGGGP DHSYWI RPE DMTMARPAYK I TTSAPGSDLAGETAALAAASTI YFKSADSRYSATLL THAKOLFNFANTHRGI YSASTI SNAAC	199
AMH40375.1	KCHVSEYEYYGQVGSGTVDHAHW-RPEDMTEERPAYKITASOPAADLAGETAAAFAAASLVFRSTNSSYANTLLNHAKOMYDFAYKHRGLTDAIPOSKC	199
AMH40376.1	KCHVSEFELYGQVGEGTVDHIYW-LPEDMPYKRNAFMITAKAPGSDLAAETAAAMAASYLVFKDVDSKYASTLLTHAKQLMNFAVKYLGKYSDSIADANK	199
AMH40377.1	KCHVSEYEFYGQQGGGQUDHNSW-QPEDMTEARPAYKITASAPGSDLAAETAAFFAASYLVFKNVDVNYARTLLTHAEQLMKFAVTYKGKYSDSISDAKN	199
AMH40379.1	KAHVSENELYGOVGGSSDHAYW RPEDMINARPANIN I TASAPOSDLAGETAALAAASMYTKI VUPATSIELLIHAKQLTIPATTIKKKI SDAFAUKA	199
AMH40380.1	KCHVSEYEYYGQVGSGNADHAYW-RPEDMTMARPAYKITPSKPGTDLAAETAAALAAASMVFRSSNASYADLLLKHAKQMYDFGYKYRGLVSDSITDSKC	199
AMH40381.1	KAHVSENVMYAQVGSCSLDHDYW-RPEDMTEARPAFMINTTHPGSDMAAETAAALAASYMVFKEVDATYANTLLTHAKQLYNFAYKYRGKYSDAFSDYRG	199
AMH40382.1 AMH40383.1	KCHVSERUT GAVGLGAUDNEGN-NFEUMINENFAFNINASAFOSULASEIAAANAAASLVIKKVNAINSALLLKHANGLPDANNHRGKYSDSIADANN KCHVSERVFYGGVGGOVDHDEW-OPEDMAEARPAVKITAOAPGSDLAAETAAFAASYVYKKVNAINSANLIHAEOLMKFAVTYKKGKYDSIIDAKN	199

AMH40384.1	I KAHVSENEFYGQVGEGSSDHSYW-RPEDMTMARPAYKITTSAPGSDLAGETAAALAAASIVFKTANSSYSATLLTHAKQLFNFANTYRGIYSQSISD/	4 S S 199
AMH40385.1	I KAHVSENEFYAQAGSGVLDHNYW-RPEDMIEARPINMINIINPGSDLAAEIAAAMAASYLVFKEVNAIYANILLIHAKQLYSFAYKYRGKYIDAFDD I KCHVKENEVYCOVGSGNUDHAYW-BEENEFEEDDAVYITTEVEGSDLAAEIAAAIAAAIAAXIVEBEVNNTYANIILIHAKQLYSFAYKYRGYIESNSITO	RC 195
AMIT40300.1	KUTYKENET TOUVOSON V DHATW-RPEDMIENKPATKII ISKPOSOLAKETAAALAASIY FKSYNNITANVLLI HAKULTYPOTKING FSKSIII S	100 195
AMH40388 1	KCHVSEFFFYQVQQQQIDHDFW-RFEDMTERTPYKITATSQCDLAAETAAAMACIASTMYEKSVATTATII THAFOIMAFAVTYQQKYSDSITD	KN 196
AMH40389.1	KCHISEYEFYGQVGGGQVDHNYW-QPEDMTEARPAYKITTSAPGSDLAAETAAAFAASYLVFKSVDATYAKKLLTHAEQLMKFALNYKGKYSDSISD/	4 K N 190
AMH40390.1	KCHVSEYEYYGQVGSGNVDHAHW-RPEDMTEARPAYKITTSKPGSDLAAETAAAFAAASLVFRSSNSTYADILLTHAKQMYDFGYKYRALVSNSITD	3 Q C 19
AMH40391.1	I KCHVGDNQLYGQVGLGQLDHDSW - RPEDMTMERPAFMINASAPGTDLAGETAAAMAAASLVFKSVNATYSALLLKHARQLFDFANNYRGKYSDAITD/	ANT 195
AMH40392.1	I KAHVSEFEFYGQVGQGQVDHDYW-RPEDMTMARPAYMITTSAPGSDLAAETAAALAAASLVFKTVDSSYSATLLQHAIQLFNFANTYRGIYSNSITD/	NS 195
AMH40393.1	I KAHVSANEYYGQVGSGTVDHASW-RPEDMTMERPAFKITTSAPGSDLAAETAAALAAASLVFQSVDSTYAATLLEHAKQLYDFGYNYRGLFSDSITDS	QC 199
AMH40394.1	I KAHTSEFEFYGQVGDG I ADHNYW - RPEDMTMDRPAFK I TESAPGSDLAAETAAA LAAASL V FKTVDSSYSATL LQHA I QLFNFANTYRGKFSDS I VN	/ R C 196
AMH40395.1	KCHVSPNELYGQVGDFNLDHEFWGRPEELNMSRPAYKIDAQHPGSDLAGETAAALAAVSLVFRTENPKYADLCLEHARQLYTFASDFRGLY-EPIKG/	AQ 199
ANR02619.1	I K C H V S D N E F Y G Q V G D G H A D H A Y W G R P E E M N M N R P A W S I R P G A P G S D L A G E T A A A L A A G S I L F S D S D A N Y A N Q L R N H A R T L Y D F A Y N N R G V Y H D S I P N A	AD 198
AOV94250.1	IKCHVDETTLYGQVGNGQLDHNSW-RPEDMTMERPAYKIDASAPGSDLAAETAAAFAAASLVFKSVDQSYSNTLLTHAKQLFDFANNHRAIYSDSIAD/	KL 199
AOV94251.1	I KAHVSENEIYGQVGNGEVDHAWW-RPEDMTMERPAYIVSANNGGSELAGETAAALAAASMVFRDVDPGYSSELIKHAEQLFNLANNHRNAYTNAIPD'	/ V N 199
AOV94252.1	I KCHVDDYVFYGQVGNGELDHTYW-RPEDMTMERPAYQINA SAPGSDLAGETAAALAAASIVFKTVDPDYSATLLEHARTMFDFAVTYRGKYSDSITD/	NG 199
AOV94253.1	IKCHVNDTEFYGQVGDFTLDHKFWGRPEELNMSRPAYKIDTEHPGSDLAGETAAALAAVSIIYKFDDPDYSAKCLMHARQLYQFATQFRGLY-EAIIG/	AQ 199
AOV94254.1	K CHVSENELY GQVGNG DI DHAWW - RPEDMTMERPAYKLTE SAPGSELAGETAAA LAAASIV FQNVNVDYSDELLEHARQLFDFANNNRKSYVDSITD/	A K 195
AOV94255.1	I KCHIQDLILYGQVGNGEA DHAWW - KPEEMPMDKPAYATI QDNAGSELAGEIAAALAAASIV FQNVDSGYANILLQHAQILFNFANDHKKIYVQSIID.	VS 195
AD 094256.1	ECHYSENELT GUVONG DI DHAWW - KPEDMI MEKPATKLI ESAPOSELAGEI AAALAAASI VEKNYNYD TSDELLEHAKULF DFANNNKKSTYD DSI DD	A N 195
APF20557.1	KCHTAERDEVGOVGEGETENLAVW, RDEEMTMVRPAVKIDTGARGSDIAGETAAALAASIVEGSSDSSSSSTILTHALOLEDEADONRGVVSDSIAD	5 A N 100
APF295581	KCHTAEREIYGOVODONTDHAYW-RPEDMINYRPAYKIDTSAPOSDIAGETAAALAAASIYFOSSDSTSSTILTHAROLFDFADONRGIYSDSISA	AD 196
APF29559.1	KCHSSANELWGQVGDGYADHAFW-RPEDMTMYRPAYKIDTGAPGSDLAGETAAALYKIDTGAPGSDLAGETAAALHAQELYSFADQYRA-YGDSIWAA	AD 190
APF29560.1	I KAHTGYFEFYGQVGDGYVDHAYW- RPEDMTMERPSYKIDTNAPGSELAGEAAAA LAAASIVFQNVDSSYASTVLDAAKELYQFANERRENYHVSIPAJ	AD 19
APF29561.1	KAHTGYREFYGQVGQGDVDHAYW-RPENMTMERPSYKIDEAAPGSELAGETAAALAAASIAFADVDSAYSTELLAVAEELYLFGDELRENYHISIED/	AN 199
ATS91249.1	I KCHTGPEELYVQVGDAGPDHSYWGSPETM - MARPAYKITASKPGTDVTMETAAAMAAGSLAFKDKDPAYSATLLEHAKQLWDFGIKYQVKYSDSVQAA	AG 199
BAA31326.1	U KAHTAANEFYGQVGQGDVDHAYWGRPEDMTMSRPAYKIDTSKPGSDLAAETAAALAATAIAYKSADATYSNNLITHAKQLFDFANNYRGKYSDSITD/	Q N 200
BAA33708.1	I KAHTSQNEFYGQVGQGDADHAFWGRPEDMTMARPAYKIDTSRPGSDLAGETAAALAAASIVFRNVDGTYSNNLLTHARQLFDFANNYRGKYSDSITD/	RN 200
BAA33709.1	I KAHTSQNEFYGQVGQGDVDHAYWGRPEDMTMARPAYKIDTSRPGSDLAGETAAALAAASIVFKNVDGTYSNNLLTHARQLFDFANNYRGKYSDSITD/	KRN 200
BAA34050.1	I KAHTAANEFY GQVGQG DV DHAY WG RP EDMTMS RPAY KIDTSKPGSDLAAETAAALAATAIAY KSADATY SNNLITHAKQLFDFANNY RGKYSDSITD/	4 K N 200
BAB40693.1	I KAHTAANEFY GQ VG GG DV DHAY WG RP E DM TMS RP AY KID TSK PG S DLAAETAAA LAATAI AY K SADS TY SNNL I THAK QL F DF ANNY RGK YS DS I TDJ	4 K N 200
BAB40695.1	I KAHTAANEFY GUVGGUV DHAY WGRDE DMINSRPAYKIDISKPGSDLAAETAAALAATA TAY KSADSITSNNLI THAKQLE PEANNYKGKTSDSITD/	KN 200
DA D40090.1	KANTAANEFTQUVQQQDVDNATWORDEDNINGRDATVEDISKPOODLAAETAAALAATATATATASADSITSNNLITINAKULEDEANNYROKSODSITD	KN 200
BAC67186 1	KATI MARET I GUVGGGD DINAT WORFEDMI MORFATAT DI SKIPOSILA KETAMA LAMA TATATA KOADOTTONKET I NAKKET VERAKKVPGI VSSSI ODI	GO 10
BAD01504.1	I K C I P K SQ T L Y AQ V G E G N D D H H F W G R A E D M R P A Y K L T P S K P G S D V A G E I A A S L A A G Y L A F K Q R D A K Y A A T L L S T S K E I Y E F G K K Y P G I Y S S S I Q D /	4 G Q 19
BAD44734.1	I KCIPKSQTLYAQVGEGNDDHHFWGRAEDMRPAYKLTPSKPGSDVAGEIAASLAAGYLAFKQRDAKYAATLLSTSKEIYEFGKKYPGIYSSSIQD/	4 GQ 19
BAF38757.1	KCVPESNTLYVQVGDGGQDHSFWGRPENM RPAFKVTTSCKGSDVAGDTVSALAAGYLVFKDSDTTFANNLLTAAKSLYTFTKNNRGIYSQCVNA	AA 19
BAH22180.1	KA HVSD NEFY GQ VGDG NA DHSY WG RPEDMI MA RPAWSITPSA PGA DLA A ETAA A LAAGYLVFRDSD A GY A A N LLDHA R RLYTFA YN NRGI Y SQSISN	AQ 200
BAK20401.1	I KCHVSDNELYGQVGDGHADHAYWGRPEEMTMDRPAWSLTPSAPGSDLAGETAAALAAGSIL FSDSDASYANQLLDHARTIYDFAYNNRGIYSESIPN	AD 198
BAL60587.1	I KAHTASSELYGQVGNGGADHSFW - RPEDMTMSRPSAK I DN NAPGSDLAGETAAALAAAS I AFKDTDSSYSSKLL SAAKELFEFANNRRATYDSS I TD/	GN 198
BAM14716.1	I KAHVSDNEFYGQVGDANTDHAYWGRPEDMTMERPAWSISPSAPGSDLAAETAAALAAGYLVFRDSDAAFANNLLAHSRTLYDFALNNRGIYSQSISN/	AG 200
BAS18736.1	I NANPQTRQLVVQVGDGNADHAFWGRPEDM - MERPCFRLEPGKP - SDAAAETAASLAAGSIAFKD KTAAEVC TRS PRTNRGIYSAA	- N 177
BAT21145.1	I KAHVSDNEFYGQVGDANTDHAYWGRPEDMTMERPAWSISPSAPGSDLAAETAAALAAGYLVFRDSDAAFANNLLAHSRTLYDFALNNRGIYSQSISN/	A G 200
BAU45387.1	NANPQTRQLVVQVGDGNADHAFWGRPEDM - MERPCFRLEPGKP - SDAAAETAASLAAGSIAFKD	N 177
BAU45388.1 BAV00271.1	NANPQIRQLVVUVGUGNA DHAFWGRPEDM - MERPCKPCSDAAEIAASLAAGSLAFKD	- N 1//
CAD54726 1	A A DEAMLE LA VOIDOUT D'HA Y WORD DE MITTARD A Y KIDTSNDOSTI A GETAAA LA LOLLEFRAN YANTI I TAKOLEDEAN YANGKYSDSI TO	SN 200
CAD54727.1	KAHVSENELYGQ VGDGDADHAY WGRPEDMTMARPAY KIDTSRPGSDLAGETAAALAAASIVFTSTDSN YANTLLTHAKOLFDFANNY RGKYSDSITO/	ASN 200
CAD54728.1	KAHVSENELYGQ VGDG DA DHAY WG RPEDMTMARPAY KIDTSNPG SDLAGETAAA LAAASIV FTSTDSNYANTLLTHAKQLFDFANNY RGKYSDSITQ/	ASN 200
CAD54729.1	I KAHVSKNELYGQVGDGDADHAYWGRPEDMTMARPAYKIDTSRPGSDLAGETAAALAAASIVFKSTDSNYANTLLTHAKQLFDFANNYRGKYSDSITQ/	SN 200
CAD54730.1	I KAHVSKNELYGQVGDGDADHAYWGRPEDMTMARPAYKIDTSRPGSDLAGETAAALAAASIVFKSTDSNYANTLLTHAKQLFDFANNYRGKYSDSITQJ	SN 200
EFN70196.1	I KCHVKENELYGQVGEFYLTETF-GRPEELETYRPAYKIDPEHPGSDVAGETAAALAASSIVFKKVDPEYSAKCLKHAKEIYNLAKRYRGFFLNEIQN	/ Q K 199
EFN70197.1	I KCHVKENELYAQVGEFYL IQDFYGRPEE INTKRRAYK IDPEHPGSDVAGETAAALAASS IV FKKVDPEYSAK CLKHAKE IYNLAKRYRGFFLNE IQN	/QK 200
EFX69372.1	I KAHVAPNKLYGQVGDGYADHSYWGRPEEM - MARPASFISTSNPGSDLAGETAAALAATSIVFSQVNATYSAICLAHAKDLYAFAKQYQGLYS - SIPQ/	GS 198
EFX80604.1	I KAHVQTNKFYGQVGDGTLDHNYWGRPEDM - MSRPAYFISTTNPGSDLAGETAAALAATSMVFASVNSTYSALCLTHAKQLYDFAKQYQAKYQ - SMPO/	AA 198
EFX80605.1	I KAHVQTNKFYGQVGDGDLDHDYWGRPEEM - MSRPAFFISTSKPGSDLAGETAAALAATSIVFKTVDSTYSALCLTHAKQLYAFAKTYRGIY - ESIPQU	/ A D 198
EF A86822.1	I KAKEDUGKYT GUAGNGGADHSTW- KAEDWPASKPSTATI I GOPGI DEANTASALAAI SYAI GSSDAATSANELSYAKULTDPAKTI KGI YSUSI ADJ	GA 198
EC17280 1	KCHTADNELVOVOSCO-SHSIWDRDESITKOOA VKVDDTHDOSDVAGEVAAAMAAGIVA EKDKDDA FATKI I EHAKO I NDFAVAVKKVVDDVTA	AF 100
EKC39023.1	K CONSTOLIY AQVGNGDTDHA I WTRPEDI RPPYYLDSSKRGSDVAGGTAAALAAGALAFLSEDSSYSSKLLESAESLYAFAKAHQGTYNTDI PEC	3 A K 19
KDR16731.1	KCHVSPNELYGOVGDEN LDHAFWGRPED LNMSRPAYK I DAOHPGSDLAGETAAA LAAVSL V FROVNPGYSSK CLDHAKOLYRFASOFRGLY - FA I KG	AQ 19
KMQ95413.1	KCHVSKNVLYGQVGDFMSDQMYWGRPEEMTICRPAFKIDAEHPGSDLAGETAAALAASSIIFRSVDRDYSLWCLQHAIELYEFADTYRGFY-EAITN/	AKE 19
KOC62382.1	I KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETASALAASSIIFRNHDPEYSALCLKHAKELYKFANKYRGLY-EAIRG/	AQ 195
KOF68403.1	I KAHTMKYEFY VQ VGN PY I DHSY WT RAED M T MER PAFKVT KEN PGSDVV AETAAA LAAGAM V FN RTD R LYA T Q L Q HA I Q L FEFA DEY RN KYSD SV P V	/ R E 198
KOX76353.1	I KCHVSEYVFYGQVGEFFLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETAAALAATSIVFRNHDPEYSDKCLKHAKELYKFANRYRGLY-EAIPG/	AL 199
KXJ18436.1	I KAHTAKFEFYGQVGDGHVDHAYWGRPEEMTMKRPAFKIXTXKPGSELAAETAAALAATSMAFQQTNPSYANKLIKHAEELYEFADKYRGKYSDSITN/	AS 200
KYB27092.1	I KCHASKYEFYGQIGDFSLDHAFWGRPEEMNMTRPAYKIDIEHPGSDLAGEAVAALAAASLLFENFNQSYSDELLRHATELYDFATMYRGLY-DSIPG/	4 K V 199
KYM87195.1	KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFSSVDPEYSAKCLEHAKDLYKFADVYRGLY-EAIRG/	AER 190
KYN01696.1	I KCHVSENALYGQVGDFPNDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFKNVDPEYSDKCLEHAKDLYKFADVYRGLY-EAIRG/	AER 199
KYN22851.1	I KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLASETAAALAASSIVFSSVDSEYSAKCLEHAKNLYKFADVYRGLY-EAIRG/	AER 199
KYN35700.1	I KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFSSVDPEYSAKCLEHAKDLYKFADVYRGLY-EAIRG/	AER 199
KYQ52831.1	I K C H V S E N V L Y G Q V G D F P I D H M F W G R P E E L N T T R P T Y K I D A E H P G S D L A S E T A A A L A A S S I V F S S V D P K Y S A K C L E H A K D L Y K F A D V Y R G L Y - E A I R G /	AER 199
KZC10616.1	I KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETAAALAASSIVFRDHEPEYSARCLKHAKELYRFATRYRGLY-EAIRG/	AQ 199
KZS03923.1	I KAKL DQGKVY GQAGNGGA DHAYW - RP EDWPASRP SYA I I I SQPGI DLA ANYASA LAAASVA I GSSDA I YSANL LNVAKQL YDFAKNYRG I YSNS I SD/	4 G A 198
KZ803022.1	KANPAIN LIGUVADONI DHKFWORPEDMUNNKPAFSISPNAPOSDLAGEIAAALAANSIVFISVDKOTSSKOLI HAKULTDPAKHRUGKI	CD 10
KZS16514.1	I KAHVAANKI YOVOOTA DHOI WORPEEM - MARPASTISI SAROSDLAGETAALAY SAITISAY VALI ISAI OLSHAKKI YIFAKOYORIYO. SI POY	/ A D 190
KZS16516.1	I KAHVAANKLYGQ VODGT LDHNYWGRPEEM - MNRPAFFI SASNPOSDLAGETAAALAATSI VFKSVDSTYSAL CLTRAKQLYAFAKQYQGKY - QSI PQ'	AD 190
OAD61293.1	KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETAAALAASSIVFRNHDPEYSTKCLKHAKELYKFASRYRGLY-EAIRG/	A Q 19
ODM94862.1	I KSHPEPNVLYVQVGDANVDHSYWGRPEDWDE I RPTL RATTRL PASEVA AEQAAAMAAASG I Y RAGDNA YASTL L QHARDL YTFATTNRGKY - DSFPE	AE 195
ODM95900.1	I KTHPEPNVLYVQVGDGNRDHAYWGRPEDWDDVRPVLRATTDRPASEVA AEQAAAMAAALKFTAMMGIPLSHQLFSNMLGIFTLLPLSTGPSATHSQ	- N 197
OWF45593.1	K C H T A P E E L Y VQ V G D G G P D H A Y WG R P E D M - F A R P S Y K V D A T K PG S D V A G E Y A A A MAA S Y L V F K D K D P D F A A K L L T H A K Q L N D F A N K Y H G R Y S D S V T A A	AA 199
OWF56607.1	I KCRASSQEYYAQVGNGGLDHSFWGRPEEMTMQRPSYKVTTSNPGSDVAAETAAALAIGAYVFKRGNTAYANELLSAAETLYTFANTYKGKYSDAIKD/	ARN 200
PBC31793.1	I KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPDHPGSDLAGETAAALAASSIVFRNYNPEYSANCLKHAKELYKFASKYRGLY-EAIRG/	AQ 199
pdb[3WC3]A	K KAHVSDNEFY GQVGDAN I DHAY WGRPEDM I MERPAWSI SPSAPGSDLAAEI AAA LAAGYL VFRDSDAAFANNL LAHSRI LYDFALNNRGI YSGSI SNA	1 A G 200
PFX25363.1	I KAHIKKEEFY GQ VGDGDLDHALWGRPEDMINRRPAWKII POEPOSDLAAEIAAALAAASIAFQAQNKKYARKLLIHAKQLYVFADKYRGKYSDSIADIPAN	4 G K 200
PSN31160.1	I KANYSUSI LI GUVGEGUA DINAWOKPEDMI MAKPATKI DI SKROSDLA AEI AAA LAAASI VYKNYDOSI SNI LLI NAKULYSPADSI KOKI SDI SDI S I KANYGDEI YVOVGTGES DIGSWIYSPEDMI ONDKAYKVI AASDGSDVAGETAAA MACGA IAFOKVNY TEEKI I ONSEOI EDEAOAHOGEVSSSVADI	1 N N 194
PVD35340.1	KCHTGPNELY VVGROGT DHGTWTSPELM. UNRPAFK I TASSPOSDVAGET AAAI TGGVI AF GSKOKT FAEKEL LURSEQLEDFAQAHQGE I SSSVAA	A A 191
PVD35381.1	. KCHASDDVI V VO	AS 18
PVD35383 1	I KCHVSDNVLYVQVGDGGADHGSWGRPEDM - MARPAFKIDATKPGSEVAMETAAAFAAGYLAFKEKDPSYSAKLLOHAKSLWGFAVTHKGKYSDSVSA	AG 19
PVD35384.1	KCHSADNVLWVQVGDGMIDSSSWGRPEDM-MSRPAFKIDSTKGGADVAMETAAAFASGYMAFKERDTEYAATLLGHAVTLYDFGLKHPGTYSSYILE	3 Q G 19
PVD35385.1	KCHSSDNVVWVQVGDGMIDQAHWGRPEDM-MSRPSFKVDSIGGGADVAMEMAAAFASGYLAFRERDAEYAETLLHHAQTVYDYGIKHPGVSSQWIIE	3 Q G 196
PVD35551.1	KANPSTKELV VQIGDGNADHSFWGRPEDM - MNRPCLKVRN PQDGADIAGEWAAALAAGSIV FKDKDSS VASQLLTAAESL VAFAKTL NGVFGGALPI/	GP 192
PVD36624.1	I KCHTGPNELY VQ VGNADADH SY WGPPEKM - MARPAFK I DA SHPGSDVAMETAAAFAAGSLAFADKNSGY SSTLL SHAKQL WEFAHSHQGKY SDSVSAA	AG 199
PVD36625.1	I KCHTGPNELY VQ VGNAGA DHSYWGPPELM - MARPA FKIDA SHPGSDVA METAAA MAAGSLA FA EKDPFRNITNLGTLS RVPVFNSEPTTKFVRVLAL	SS 183
PVD36896.1	INSDPFREELVIQ VGDGDIDHQ FWGRPEDM - MPRPCLKIG V NRDGSDVA A DYAAA LAAGSIV FREGDVQ YARELLLTAESIYDFAVNNRGLYSNTLSA/	4 Q Q 193
PVD36986.1	I KAN MSIKEFV VU IGEGYPDHHFWGRPEDM - MDRPCFK IGPGKP - TOVAAETAASFALAYM IFKEKGAAAESIFDFANSEKG IYSSSQ/	GA 181
KP_001606454.1	I KUHYSEFELTGU YGDFA I DHI FWGRPEELN ISRPAFKI DPEHPGSDLAGETAAALAAASLV FRTSEPSYSEKL LRHARELYN FANKYRGLY - EA IRG/ KOUNAVEEVED	AQ 199
CP_001044774_0	: non a nieriou iouro Lunarworreem mirratni Diehrosola Geavaalaaas Lien FNGS YSDell RHATELYDFATMYRGLY-DSIPG/ ) KCHVSDYFFYGOVGDFS: DOGFWG RDEED NINS RPSYVID RUNDBOG NI AGETSAAL SU SUN VEVINS DE SOOI GUIDAD I WEATOVGOV (S	AO 400
(P 002426465 1	KCHVSPDEFYGOVGDFALDHAFWGRPEDINMSRPAYKIDKEHPGSDLAGETAALAAVSIAFKVVDPEVSASICIKHRKUTMFATQYRGLY-BAIKG	ASO 100
(P 003402778 1	KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETAAALAASSMVFRNHDSFYSAKCIKHAKFIVKFANFVRA	AQ 10
(P 003690676 1	I KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPDHPGSDLAGETAAALAASSIVFRNYDPEYSANCLKHAKELVKFASKYRGIY-FAIRGY	AAQ 196
(P_005099312.1	KADPHKKELVVQ I GDG L A DH H F W G R P E D M - MA R P C T K V N Q Q K K - S D I A A E T A A A L A L G S I T F K R K G T A A E S L Y S F A K N N R G Y F R D /	F - 17
(P_005099315.1	S A D E Q H H R L T A Q V G D G G T D H N Y W G R P E D M - M S R P C K Y V D S N H V - S D V A G E T A A S L A L G Y L T F K D K D A K A K T L Y T F A K A H P G V F T G S /	G - 178
<pre>KP_005101787.1</pre>	I A A N A H A H K L T A Q V G D G D A D H A Y W G R P E D M - M S R P C M Y I D S H N P - S D I A G E T A A A M A L G Y L V F K Q K D S S A K S L Y A F A K A H P G V F T K S - E	- K 178
KP_005101789.1	I AANAHAHKLTAQVGDAHADHAYWGRPEDM - MSRPCVY I DNNTP - SNIAGETAASLALGYLVFKQKD SSAKSLYAFAKAHPGSFTGS - G	- G 178
CP_005111868.2	2 K CH T G P N E LY VQ VG NG N L DH AS WG P P E QM - MA R P SY K I T A SK PG SD VA M E T AA AMAAGYL A F KQKD VH F A N T L L Q HA K QL WE F A I NH RGKY SD SV P E /	HD 19
ur_005111869.2	<pre>c num i a peel t vu voda ga dh sy wg ppelm - ma kp a ski i a a kp go su va Me TAAA MSA GY LA FK DK DP TF AAD L T HA EQ LW DF AMK Y PGKY S DS VQA Kali ta peel t vu voda ga dh sy wg oppek w ma pa pa vi ta a kp go go wa Me TAAA MSA GY LA FK DK DP TF AAD L T HA EQ LW DF AMK Y PGKY S DS VQA Kali ta peel t vu voda ga dh sy wg oppek wa na pa pa vi ta a kp go go wa Me TAAA MSA GY LA FK DK DP TF AAD L T HA EQ LW DF AMK Y PGKY S DS VQA Kali ta peel t vu voda ga dh sy wg oppek wa ma pa pa pa vi ta a kp go go wa me ta a pa pa pa pa pa pa pa pa pa Kali ta peel t vu voda ga dh sy wg oppek wa ma pa pa</pre>	AG 199
1.D/8111 CUV 1	I NUMIOFALLI YUYUUUNKUMNI WUFFENMI MFRFOINI I AONFOODYAMEI AAAMSAUYLAFKNKUAHFANKLLASAKSLWI FAWNHOGKYSDSVPA/	100 195

XP 006610535.1	1 KCHVSEYVEYGOVGDESIDHTEWGRREEINTTRRAYKIDRDHRGSDIAGETAAAIAASSIVERNYNREYSANCIKHAKEIYKEANKYRGIY.EAIRU	GAAO 190
XP_0081818421	I KCHVSDYFFYGOVGDESI DOG FWGRPEDMMMSRPSYKIDKNRPGSDI AGETSAAI SAVSMVETKIDSNESASCI SHAKOI YMFATOYRGI Y.DAIK	GAAQ 196
XP 008 194 455.1	1 KCHASKYEFYGQIGDFSLDHAFWGRPEEMNMTRPAYKIDIEHPGSDLAGEAVAALAAASLLFENFNGSYSDELLRHATELYDFATMYRGLY-DSIP	GAKV 195
XP 009049791.1	1 KAHTKPNEFY VQVGDGGKDHSYWGRPEDM - MARPSYKITPSCAGSDVA AETAAAMAAGYMVFKNKDSA YANKLLSHSKOLYDFALAHOGIYTHCINI	EAAA 196
XP_009049836.1	1LHWRVGDGG I DHGSWTSPERM - GNRPSFK I DASSPGSDVAMETAAAMAAGS I A FQTKDSA YSA T L L THAKQLYTFAKAHPSYYSNSVN	AAAA 184
XP_009049839.1	1 KCHTKPNELYVQVGDGGKDHGSWTSPERM-GSRPAFKIDTTRGGADVGMETAAALAAGSIAFKTKDSAYSHKLLTQAKSIYQFARKTNNLYSNSVN	AAAA 199
XP_009049840.1	1 KAHPTKFEFYAQVGDPFIDHDYWGRPEDMTMYRPAYRLSPERPGSDVIGETAAALAAGSIVFRTRDPALAGEMLRHSEELFEFAQNYQGLYSESIP	SVKE 200
XP_009049941.1	.1 KAHVSKYELYVQVGDGGADHGFWGRPEEM - MRRPSYKIDASHPGSDVAGETCAAFAAGYLAFKERDPKYAAQLKQHAMDLYDFGMKHQGRYTDTIS/	AASG 199
XP_009058133.1	.1 KSHTGKNELY VQ VGDASA DHSFWGRPESL - MARPAYKITA SKPGTDVA GATSA A FAAGHLA FKEKDPKYAAKLLQ HAKELY DFAMTHKGVYSSSVP/	A A A G 199
XP_009058134.1	1 KCIPSKNEFYMQVGNGDDDHHFWGRAEDMRPSLKNTASKPGSDVAGETAAAMAAGSIAFKKRDHAYSTKLLTAAKSLYVFAKTYTGKYSQSVQ	GAGQ 197
XP_009060290.1	.1 KCIPAKNEFYVQIGDGTPDHHFWGRAEDMRPAYKITASKPGSDISGDTAASLAAGSIAFRSRDPAYSKKLLDASKSLYAFGKKYTGIYSQSVS/	A A A A 197
XP_009064976.1	.1 KAHNNTNELYVQVGDGKADHNFWGRPEDM-MARPSYKIDASHPGSDVAGETCAAFAAGYLAFKDKDPKYAATLKQHATELYEFGMKYQGVYTDTIT/	A A V G 199
XP_009064980.1	1 KCIADKEIYYYQVGDADIDHKFWGRPEDM RPAFYLTKDKPGSDVAGGTAATFAAGAIAFNITDAVYSQKLLKAAESLYKFAVTYKGKYSQTHSI	DTGT 197
XP_009065702.1	.1 KCNPHSQTYYAQVGNGNADHSFWGRPEDM RPAYKVTASKPGSDVAGGTAAALAAGSIAFKSKDHGYSSKLLNSAKSLYQFAKNHKGIYSSSIPI	EASQ 197
XP_011057333.1	1 KCHVSENVLYGQVGDFPVDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFSSVDSEYSAKCLEHAKDLYKFADVYRGLY-EAIR(	GAER 199
XP_011164235.1	1 KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFSIVDPEYSAKCLEHAKGLYKFADVYRGLY-EAIRO	GAER 199
XP_011253653.1	1 KCHVKENELYGQVGEFYLTETF-GRPEELETYRPAYKIDPEHPGSDVAGETAAALAASSIVFKKVDPEYSAKCLKHAKEIYNLAKRYRGFFLNEIQI	NVQK 199
XP_011253654.1	.1 KCH I SENELY GQ V GEF DL NE I FFG R PEDMNT FR PVYK I DAKH PG SDLAGETAAA LAASS I V FKKVD PEYSAK CLKHAKELY I LANH YRG FY - H TVQ	SAKQ 199
XP_011253655.1	1 KCHVKENELYAQVGEFYLIQDFYGRPEEINTKRRAYKIDPEHPGSDVAGETAAALAASSIVFKKVDPEYSAKCLKHAKEIYNLAKRYRGFFLNEIQI	NVQK 200
XP_011296914.1	1. KCHVSEFVFYGQVGDFSLDHTFWGRPEELNTIRPAFKIDKDHPGSDLAGETAAALAASSTVFKMIDPSYSSECLRHARELYKFANIHRGLY-ESIKI	NAAQ 199
XP_011296916.1	. KCHVSEFVFYGQVGDFSLDHTFWGRPEELNTRPAFKIDKDHPGSDLAGETAAALAASSTVFKMTDPSYSSECLKHARELYKFANTHRGLY-ESTRI	NAAQ 195
XP_011346866.1	1. KCHVSENVFYGQVGDFSIDHAFWGRPELNISRPAYKIDAEHPGSDLAGEIAAALAASSIIFRNVDPEYSVKCLQHAKDLYKFAUHTRGFY-IVIP	GGQE 195
XP_011421782.1	. I KOHTAPNELTVQVGSGQ-SHSLWDKPESTI KURA-TAVIDI I HPGSDVAGETAAAMAAGTLAFKDKDPAFATKLLEHAKQTNDFAVATKGATSDSVIJ	AAAE 190
XP_011423139.1	. Konigrneli vyvodog podi vokredni kokredni kokred v zakravani v rekuperakli i nakoli pravni kokredni kokredni k	DANN 195
XP_011430172.1	I KONFULUETTAU OKOKEDTRI WERDEDMERPEMERFATIOLIKUU - SUVAGETAALASGATVERDKNA ISNELLIAAKSLAFAFAKSUUGITOLIIMI	EGAK 10
XP_011447502.1	I KORNSAD IVI AUVONGDI DHA I WI KPEDI BEDVYI NSSKTOSDVA GOTAA LAAGALA EMEDSSYSSKELESAESI VA GAKA HOGI VNI VI	EGAK 197
XP_011408432.1		GAAO 10
XP_011632703.1		GAER 196
XP 011700378.1	1 KCHVSDNVLYGQVGDFPIDHMFWGRPEELNITRPTYKIDAEHPGSDLAGETAAALAASSIIFSSVDPEYSAKCLEHAKGLYKFADVYRGLY-EAIR	GAER 195
XP 011867251.1	1 KCHVGENVLYGQVDDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFNSVDPEYSAVCLEHAKDLYKFADVYRGLY-EAIR	GAER 196
XP 012061951.1	1 KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFSSVDPEYSAKCLEHAKDLYKFADVYRGLY-EAIR	GAER 199
XP_012146513.1	1 KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDAEHPGSDLAGETAAALAASSIVFRNQDPEYSAKCLKHAKELYKFADKYRGLY-EAIRU	GAAQ 19
XP_012224814.1	1 KCHVSENVLYGQVGDFSIDHTFWGRPEDLNTSRPAYKIDPEHPGSDLAGETAAALAASSIVFRSIDPAYSAECLKHAMELYKFADAYRGLY-EAIRI	GTEL 199
XP_012248155.1	1 KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETAAALAASSMVFRNHDPEYSAKCLKHAKELYKFANEYRGLY-EAIRI	GAAQ 199
XP_012273697.1	1 KCHVSELVFYGQVGDFALDHTFWGRPEELNMTRPAFKIDPEHPGSDLAGETAAALAASSIVFKNTEPAYSLECLRHARELYKFANTYRGLY-EAIRO	GAAQ 199
XP_012526654.1	1 KCHVSENVLYGQVGDFPVDHMFWGRPEKLNTTRPTYKIDAEHPGSDLAGETAAALAATSIVFSDVDPEYSAKCLEHAKGLYRFADVYRGLY-EAIRO	GAER 199
XP_012935745.1	1 KCHTAPEELYVQVGNGGIDHGYWGSPEAM-MSRPSYKITASNPGSDVAMETAAAMAAGYLAFKNKDAHFASNLLSHAKQLWEFGNKYRGKYSNSVT/	A A A G 199
XP_012938497.1	1 AAWNGHDKLTGQVGDANADHAQWQPPEKM-TSRTCQYITKGTK-SDVAGATAASLALGYLIFKSSNNKAKSLYSFGKSNRGFYGTS	P 177
XP_012939931.1	1 A A D P A L R R L T A Q V G D G T A D H A F W G R P E D M - M S R P C M Y V D S S H V - S D I A G E T A A S L A L G Y L T F K Q	Q D 158
XP_012939933.1	1 AANANAHKLTAQVGDGAADHAYWGRPEDM - MSRPCKYIDNNTP - SDIAGETAASLALGYLVFKQKD SSAKSLYAFAKAKPGVFTGS (	G D 178
XP_012941225.1	1 KCDNDKQEFYGQVGDGHIDHSTW-RPEDMQMSRPSFKLDRNNGGSDLAGETAAALAAGSLAFKSTDPAYASELLTAATQLYKFADKYRREYHLTITI	NAVD 198
XP_012945520.1	1 KCHTAPEELYVQVGNPGPDHGYWGRPEDM - MPRPSYKITASRPGSDVAMETAAAMSAGYLVFKDKNSTFANTLLTHAKQLWDFGMKYKGKYSDSVS/	A A A G 199
VD 012066277 1	A RANGEVOEL VOUADED I DRAVWERDEDNI NVRRCVAANACUR CRVACATVAALAACVI VEHRVE CVAVTI VVEATRVRCAV. OT	NI 477
XP_013066278.1	. RANSSKUEFT TUVADED TUKAT WORFEDM "MIRP COVINTSIR. SDVA GATVA LAGRIVEK KKG	A F 175
XP 013066280.1	1 RANPSTRILTVQVGDGYADHEWWGRPEDM - MSRPCQTVTTSNR - SDVAAGTAAALAAGAVAFMNNG VAAESLYDFAKTYRGFYGAA	S 178
XP 013066281.1	1 KADPTAKTLVAQVGDGNADHGWWGRPEDM - MARPCTY I SATEK - SDISGGVAAALAAGA I AFKDKG SAAESL YAFAKAN RG I FNAS	H G 175
XP 013066283.1	1 KADPTAKTLVAQVGDGNADHGWWGRPEDM - MARPCTYISATEK - SDISGGVAAALAAGAIVFKDKGSAAESLYAFAKANRGISI	NASG 178
XP_013066284.1	1 NANSSAQQFVAQIGDGNTDHAYWGRPEDM - MVRPCIYINTTFK - SDVLGETVAALAAGSITFQAKG SAAKSLYAFAMDNRGLYNGS I	R G 179
XP_013066285.1	.1 NSDPVHLEFVAQ IGDGNLDHGYWGRPEDM - MDRPCLK INTTSP - SDIA AGTAAA LAAGA IAYKTKG TAAE SLYAFAKNYRGIFNGS I	N C 179
XP_013066286.1	1 NSDPVKKEFVVQVGDGTLDHGFWGRPEDM - MARPCLKVNTSYP - SDIAGGTAAALAAGAIAFKAKG TAAESLYSFAKTYRGLFTGS I	H G 179
XP_013066370.1	1 KAKPQSRELVVQVGDGNADHSFWGRPEDM - MARPCQTVSSSNK - SDIAGETAAALAVGSIVFKEKG TAAESLFAFAKSNRGIFRGSAP	179
XP_013066371.1	.1 KGNPQTKQLVAQVGDGRADHSYWGRPEDM - MPRPCQFINETV - NSDIAAETAAALAAGSIAFKENG TAAESLYSFAKTHRGVFKGS /	A T 179
XP_013069169.1	1 KCHTKPDELYVQVGDGGADHSYWGPPELM - MARPAYKITASKPGTDVAMETAAAFAIGSLVFKNSDPTYSATLLTHAKQLWELGEKHRGIYSDSVQ/	AAAA 199
XP_013069170.1	1 KCHTKPDELYVQVGDGGADHSYWGPPELM - MARPAYKITASKPGTDVAMETAAAFAIGSLVFKNSDPTYSATLLTHAKQLWELGEKHRGIYSDSVQ/	AAAA 199
XP_013069521.1	1 KCHVKPNELYVQVGDGGQDHGYWGPPELM - MPRPAFKITTSKPGSEVAMETAAAFAAGSLAFKTKDPTYSSTLLDHAKQLWDFAMNYRGKYSDSVN/	AAAG 199
XP_013069522.1	1 KCHVKPNELYVQVGDGGQDHGYWGPPELM - MPRPAFKITTSKPGSEVAMETAAAFAAGSLAFKTKDPTYSSTLLDHAKQLWDFAMNYRGKYSDSVN/	AAAG 199
XP_013069523.1	1 KCHVKPNELYVQVGDGGADHSYWGPPELM-MPRPSFKITASKPGSDIAMDTAAAFVTGYFAFKEKDPTYANTLLEHAKQLWDFGMNHKGKYSESVG/	A A A G 199
XP_013084279.1	1 KCHTKPDEFY VQVSDYLA DYWSWNHPE - FYGDRPPYRINA THPGTDIA MEAAAA FAAGSLVFKTKDPSYSTRLLTHAKQLWDFADKNRGIYSDSIE'	VAAH 199
XP_013084281.1	1 KCHTKPDELYVQVSDYIEDYWAWNQPE - FYGERPPYKINDTHPGTDIAMETAAAFAAGSLAFKTKDPSYSATLLTHAKQLWDFGNKYHGLYSDSIVI	MAAY 199
XP_013084282.1	1 KCHTKPEELYVQVSDYIEDYWNWLHPEFM - GQRPPYKINATHPATDIAMETAAALAAGSLAFKTKNPSYSTTLLTHAKQLWEFGNKYHGLYSDSIS.	A A A H 199
XP_013084283.1	1 KCHTKPEELY VQVSDYTEDYWNWLHPEFM - GORPPYKINA THPATDIA METAAA LAAGSLA FKTKNPSYSTILLTHAKOLWEFGNKYHGLYSDS IS.	AAAH 195
XP_013084285.1	1. KCHI KDEELY VQVQNAI KDHAFWGPPEKM - QKPPSFKINSI KPGI DVAMEI AAAFAAGSI AFKAKDPSYSI I LLNHAKKLWDFGNKYNELYSNSL <sup>1</sup> V AUTOELYV VQVQNATKDHAFWGPBEKM - QKPBSFEDINSTKPGI DVAMEI AAAFAAGSI AFKAKDPSYSTI I LLNHAKKLWDFGNKYNELYSNSL <sup>1</sup>	GVSP 199
XP_013004287.1	. RUNTINDEELI VUVGNATINDAF WOPPERM-UKRPSTRINSTRYGT DVAMETAAAFAAGSTAFRANDYSTSTTLELMAKKEWPGNATINELISASEY	G V 3 F 198
XP_013066631.1	I RANFOLTEFIFU GNESLUHHT WORPEUM - NORPEN VASVARF-SEVAAETAALAIGSIVENDRGFARESLINFANEHDU 1933	S 170
XP 014214612.1	1 KCHVSDDVLYGQVGDFDVDHLYWGRPEELNGTRPAYKIDAEHPGSELAGETAAALAASSILFRATDGPYSEKLLGHARRLYSFAVTHRGLY-EAIK	GAAQ 195
XP 014601854.1	1 KCHVSEYVEYGQVGDESVDHAYWGRPEELNTTRPVYKIDPDHPGSDLAGETAASLAASSIVEKNVYPDYSERCLKHAKELYNEANKYRGLY-SAIR	GAAO 196
XP 014786694.1	1 KAHTMKYEFY VQVGNPY I DHSYWTRAEDMTMERPAFKVTKENPGSDVVAETAAALAAGAMVFNRTDRLYATQLQQHA I QLFEFADEYRNKYSDSVP	VVRE 200
XP_015126097.1	1 KCHVSEFVFYGQVGDFSLDHTFWGRPEELNTTRPAFKIDSDHPGSDLAGETAAALAATSIVFKTIDPLYSSKCLRHALELYKFANTYRGLY-ESIRI	NAAQ 198
XP_015184471.1	1 KCHVSEFVFYGQVGDFSVDHAYWGRPEELNTTRPVYKIDPDHPGSDLAGETAASLAASSIVFKNVNPDYSERCLKHAKELYNFANKYRGLY-SAIRI	GAAQ 199
XP_015432538.1	1 KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETAAALAASSIVFRDHEPEYSARCLKHAKELYRFATRYRGLY-EAIRI	GAAQ 199
XP_015513382.1	1 KCHVSEDVLYGQVGDFALDHRFWGRPEELNTSRPAFKIDSDHPGSDLAGETAAALAVTSILFRDSNASYSAECLRHAKELYHFANTRRGLY-EVIKO	GAAQ 199
XP_015592927.1	1 KCHVSEFVFYGQVGDFSLDHTFWGRPEELNTTRPSYKIDPEHPGSDLAGETAAALAASSIVFRDSDLEYSRLCTRHAKELYKFANTYRGLY-EAIRI	GAAQ 199
XP_015836219.1	1 KCHASKYEFYGQIGDFSLDHAFWGRPEEMNMTRPAYKIDIEHPGSDLAGEAVAALAAASLLFENFNQSYSDELLRHATELYDFATMYRGLY-DSIP(	GAKV 199
XP_016907626.1	.1 KCHVSEYVFYGQVGDFSLDQTFWGRPEELNTTRPAYKIDPDHPGSDLAGETAAALAASSIVFRNYNPEYSANCLKHAKELYKFASKYRGLY-EAIRO	GAAQ 199
XP_017766299.1	1. KCHVSEYVFYGQVGDFSLDHTFWGRPEELNTTRPAYKIDPEHPGSDLAGETAAALAASSIVFRNHDPEYSTKCLKHAKELYKFASRYRGLY-EAIRU	GAAQ 199
XP_017785110.1	. KCHTAPPETTAQVGDFPSDHKFWGRPEELNMTRPAHKIDKDQPGSDVAAEAAAALASASTVFQKIDPDTSHELTKHSTELFDFATETKGLP.DTFT	G 5 K H 199
XP_018016806.1	I KCHVSENTI NGVODOVI DHAWW. PREELA INFAI KUPENTAPOSELAGETASA CAASSI VERESDA LAND VENTAPELENEANOVAANVHESID	NAVD 195
XP 018016809 1	1 KAHSSNMLLYGQVGQGDIDHAYW-RPEDMTMQRPAWKIDTSAPGSDLAADTAAALATASMIFRDSDSGYADRCLAASKDLFFFADQYRALYHISIPI	DAAY 196
XP 018016810.1	1 RAHSGDMLLWGQVGEGQVDHSYW-RPEEMTMSRPSMKIDVNAPGSDLAGETAAALTAASLVFQDGDPAYAAVCLEAAKNLFDFADQYRLMYHESIT	DAAN 195
XP 018016813.1	1 KAHTTPTTIWAQVGDGDLDHSLW-PPEDMTMARPSYKLDQNAPGSEVAAESAAGLAAASIVFKNSNPSYSATLLQAAKELFDFANNYREMYTNSLP	EAAD 199
XP_018016820.1	1 KSHTAPTTIWAQVGDSDLDHTFW-RPENMTMARPSFVLNETNPGTEVGAELAAGLAASSMVFQRSDPAYAATLLTAAKELFDFANTYRAVYSDNIPI	EGDA 199
XP_018016821.1	1 KSHTAPTTIWAQVGDSDLDHTFW-RPENMTMARPSFVLNETNPGTEVGAELAAGLAASSMVFQRSDPAYAATLLTAAKELFDFANTYRAVYSDNIP	EGDA 199
XP_018026496.1	1 KAHSAPGELYGQVGDGGLDHAYW - RPEEMTMPRPAAKIT TQAPGSDLAGETAAALAAASIL FSD - DAQFSTACL DAAR DLYDLADLHRQIY TVSIP (	GASG 198
XP_018027134.1	.1 KAHSASMTLWGQVGDGNADHAYW-RPEDMTMARPSWKIDTNAPGSDLAGETAAALAAASIVFYNSDNAYSVRLLDAAKDLFAFADQHRGKYSDSIP	A A A N 199
XP_018044452.1	1 KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFSSVDPEYSAKCLEHAKDLYKFADVYRGLY-EAIRO	GAER 199
XP_018307051.1	1 KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLASETAAALAASSIVFSSVDPKYSAKCLEHAKDLYKFADVYRGLY-EAIR	GAER 199
XP_018323591.1	1 KCHPSAYEFYGQVGDFSTDHRFWGRPEDLNMSRPAYKIDKDHPGSDLAGEAAAALAASSLVFRETNKTYADIILKHAEELFDFATTYRGLY-DVIP	GAKK 199
XP_018347393.1	1 KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLAGETAAALAASSIVFSSVDPEYSAKCLEHAKDLYKFADVYRGLY-EAIRO	GAER 199
XP_018359268.1	1. KCHVSENVLYGQVGDFPIDHMFWGRPEELNTTRPTYKIDAEHPGSDLASETAAALAASSIVFSSVDSEVSAKCLEHAKNLYKFADVYRGLY-EAIR	GAER 199
AP_018396635.1	1. RUTYSENALT GUVGEP FNUHMFWGRPELINT I RPTYK I DA EHPGSDLAGET AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUMB AFV I DA FHORSDLAGET AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUMB AFV I DA FHORSDLAGET AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUMB AFV I DA FHORSDLAGET AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUMB AFV I DA FHORSDLAGET AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUMB AFV I DA FHORSDLAGET AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUM I DA FHORSDLAGET AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUM I DA FWO BEENNUM I DA FHORSDLAGT AAA LAASS I V FKNUDPE VSDKCL EHAKDLYKFADVYRGLY - EA IR V CHUSYNELVCOVARED I DA EWO BEENNUM I DA FWO BEENNUM I DA	GAER 199
XP_018568194.1	. ROMYGRNEL I GUYGNEFFLUHAFWURFEEMNMNRFAR I DE I HFU DLAGEASALFSSASI VERNVNESYSKELLKHAT ELYDFANNYRG I USUU 199 A VENVENET VANVANED DLAGWADDEEMNMNDDAEN DE TUDAT DI AGEASALFSSASI VERNVNESYSKELLKHAT ELYDFANNYRG I SAVIA	GAKO 199
XF_018088196.1	. ROUNDRALLIGUVORFFLUDAFWORFEEMAMARFAFIJEINFOIDLAGEASALFSSASIVYKNVNESYSKELLRHAIELYDFANNYRG-LSDVIP	NVOV
XP_019884179.1	, ROUTRERELIGUIGETELEETWORFEELDMIRFALINFERTOODVAGETAALAASSIVEKKUDETSAKGLMARETYLLANLINKYKYKY Y KCHKKYEEYINOSETIENNYKKODERMINTROAVKANSOLDASDVIAETVALIASSIVEKEETVOVEVENTUUSTINUS VERTOOPY KGTPI	D 404
XP 020209295.1	. ROHMANTELLER NUMBER EINER NUMBER EINER EINER VAL AUS VERLANDEL VAL LANDE VERLANDE IN TANKEN KER ANN ANN ANN A	GYEN 20
XP 020280301 1	KCHPOFELEY LOVGDEMVERELEGREENVEVHRPAYAANSSHPATDI I SETTALAAASI VEKINDU HI SOVOI THAKAI VELANVODVODVODVODV	SYEN 100
XP 020289336 1	I KCHRDKN IFV I QVGN I TY EANYWGRPEEL NYDRP I LETNSTYPASDV I AETAAALAAASI VEKEVDYENSRI CI I HAI GI VEFAMI I NEEHDI RODU	NLSD 10
XP 020614867 1	1 KAHIKQEEFYGQVGDGHTDHAYWGRPEDMTMARPAWKITPQKPGSDLAAETAAALAAASIAFRFSDVSVSMKLLQHAADLVFFADKYRGVSDSID	NAAS 200
XP_020893295.1	1 KAHTAKFEFYGQVGNGDI DHAYWGRPEDMKMKRPAYKI TESKPGSDLAAETAAALAATSMAFKSTDROYANKL LQHAKELYDFADKYRGKYSDSI LI	DAAX 200
XP_020893299.1	1 KAHTAKFEFY GQ V GDG H V DHAY WGRP E E M TMK RP A F K I X T X K P G S E LA A E T AAA LAA T SMA F QQ T N P S Y AN K L I K HA E E L Y E F A D K Y R G K Y S D S I T I	NAAS 200
XP_021345983.1	1 KCRASSQEYY AQVGNGG L DHS FWGRP EEM TMQRP SY KVT TSNPGSDVA AETAAA LA IGAYV FK RGN TA YANE LLSAAE TLY TFAN TYKGK YSDA I KI	DARN 200
XP_021361084.1	1 KCKPSRHTLY IQCGNGHAGHDYWGRPEDM RPTYKVTAQKPGSDVAGETAAAMAAGS IAFQSKDTVYANRLLTASRSLYVFAKSHRGKYSDSVHI	DVHE 196
XP_021363481.1	1 KAHTGPNELYVQVGDVDLDHHFWGRPEDM - MARPAYKITAEKPGSDVAGEFAAAMAASYIAFKDKDPQFAADLLRHAREIYAFALQYNGTYSKSLP	QAGA 199
XP_021363490.1	1 KCKPSQNVYYAQVGNGGTDHAVWGRPEDM RPSYKVDAGKPGSDVAGETAAALAAGSIVFKERDAGYSTKLLTAAKSLYEFAKNHKGIYSQSVPI	DAQS 196
XP_021363495.1	.1 KCHTAPEELYVQVGDGGPDHAYWGRPEDM - FARPSYKVDATKPGSDVAGEYAAAMAASYLVFKDKDPDFAAKLLTHAKQLNDFANKYHGRYSDSVT/	A A A A 199
XP_021363496.1	A KONTAREELVNOVORCORRUAVWORRERN, EARROYVNATKROSRVACEVAAANAARVI VEKRKRRRRAKI I TUAKOLNREANKVUORVORVORVA	AAAA 196
the second se	. KONTAPEELIVUVBBOOPDHATWORPEDM-PARPSTRUDATRPOSDVAGETAAAMAASTLUPRDRDPDPAARLLIHARULNDPANRTHORTSDSVI	
XP_021363497.1	1 KAHTAPNELY IQVGDGGDHATMGRPEDM - MARPAFKVDATNPGSDVAGETAAMMASTLVFRDKDFDFAALLTHARQLNDFANK HGKTDDSVI 1 KAHTAPNELY IQVGDGGQDHGFWGRPEDM - MARPAFKVDATNPGSDVAGEYTAAMAAGYLVFKDKDFDFAALLTHARQLNDFANK HGKTDDSVI	QAAA 199
XP_021363497.1 XP_021363498.1	I KAHTAPNELY IQVGDGGODHG FWGRPEDM - MARPAFKVDATNPGSDVAGEYTAAMAAGYLVFKDKDPTFAAQLLTNAGEIYKFAVTYKGIYSNSVN I KAHTAPNELY IQVGDGGODHG FWGRPEDM - MARPAFKVDATNPGSDVAGEYTAAMAAGYLVFKDKDPTFAAQLLTNAEEIYKFAVTYKGIYSNSVN I KAHTAPNELY IQVGDGGODHG FWGRPEDM - MARPAFKVDATNPGSDVAGEYTAAMAAGYLVFKDKDPTFAAQLLTNAEEIYKFAVTYKGIYSNSVN	QAAA 199 QAAA 199
XP_021363497.1 XP_021363498.1 XP_021924915.1	I KAHTAPNELY IQVGDGGQDHGFWGRPEDM - MARPAFKVDATNPGSDVAGEYTAAMAAGYLVFKDKDPTFAAQLLTNACEIYKFAVTYKGIYSNSVN 1 KAHTAPNELY IQVGDGQDHGFWGRPEDM - MARPAFKVDATNPGSDVAGEYTAAMAAGYLVFKDKDPTFAAQLLTNAEEIYKFAVTYKGIYSNSVN 1 KAHTAPNELY IQVGDGQDHGFWGRPEDM - MARPAFKVDATNPGSDVAGEYTAAMAAGYLVFKDKDPTFAAQLLTNAEEIYKFAVTYKGIYSNSVN 1 KCHVSPNELYGQVGDFNLDHAFWGRPEDLNMSRPAYKIDACHPGSDLAGETAAALAAVSLVFRDVNPGYSSKCLDHAKQLYRFASQFRGLY - EAIK	QAAA 196 QAAA 199 GAAQ 199

0010110011	KAUMOONELA				000
_021941324.1	KAHVSQNEL	GUVU	SOG DV DHA Y WG RP	EDM IMSKPATKIDISHPGSDLAGETAAALAAASIVPGSADSGTASTLLIHAKQLPDPANNTRGKTSDSIIDAQS	200
021956549.1	KAHPTDEEE	GOTO	SDG NODHGY WGRP		106
021959948.1	KCHTGETELI	FVQVO	SNATADHNSWDIP	ERW PRPAYKITAERPGSDIAGETSALFSAGTLLFOTVNDTYASILLDHAKSLYTFANOYRGKY - DSVPEATP	196
021966378.1	KNHPSPNVL	vavo	SNGEIDHNYWGRP	E D WN G P R P V L Q A N S S K P A S E V A G E Q S A A MAA A S I A F R Q G N T S Y A D T L L Q H A R E L Y T F A T T Y R G H Y - E S F P E V K E	199
021967587.1	KAHVAPNEL	GQTO	G D G Y Q D H S F W G R P	EDW - GPRQSWK I TQQQPGSELAGETAAA LASGYL V FREEDP TYAN TL LQHAREL YN FAK TYRGDY - NA I PARD -	197
_022167267.1	KCHVSDYEFY	GQVO	GDFSLDQGFWGRP	E D M N M S R P S Y K I D K N R P G S D L A G E T S A A L S A V S I V F T K I D S N F S A S C L S H A K Q L Y L F A T Q Y R G L Y - D A I K G A A Q	199
_022 167 268.1	KCHVSDYEF	GQVO	3 D F S L D Q G F W G R P	E D M N M S R P S Y K I D K N R P G S D L A G E T S A A L S A V S I V F T K I D S N F S A S C L S H A K Q L Y L F A T Q Y R G L Y - D A I K G A A Q	199
_022200004.1	KCHVSPNEF	GQVO	3 D F T L D H E Y WG R P	E D M N M S R K A Y K I D A E H P G S D L A G E T A A MAA V S M V F K E E D P E Y S D L C L E H A K Q L F K F A T T Y R G L Y - E A I K G A A Q	199
_022328778.1	KCRNDSQKY	r AQ VO	SNGDIDHAYWGRP	EDM RP AT YLDS I KIGSDVIGATAAALIAGSMAFL SKDPSYSSRLLESAESLYVFAKARQGIYNADISDGAK	197
022336001.1	KCHOOTOEV	AQVO	SNGKEDHRYWGRP	EDM RPATELES INTOS DVIGATAAA LAAGAMA FLSK DPS TS SKLLESAESLTVFAKARAGITINAD ISDGAK	197
022336153.1	KCHTGTNEL	vovo	SNGK - SHSVWDRP	ESI - TKQP AY KVDD T KPGSDVAG EYAAA MAAGYLAFKERDP TFAAKLLEHAKQINDFAV TYKGKYSDSVPEASE	198
022338490.1	KCRPQSQEY	AQVO	G N G G A D H A F W G R P	EDM RPAYKCTASNGGSDVEG I TAAALAAGSMAFKASDAAYSQRLLSSAKSLYEFANAHKG I YNKG I SDATS	197
_022338491.1	KCHPQTQEY	AQVO	3 K G K E D H R Y WG R P	E D M F R L R P A Y K C T R N D T - S D V A G E T AA A L A S G A I V F K D K D P R Y S S D L L E A A E S L Y D F A K R N Q G Y Y V Q D V N G	199
_022341787.1	KCHTGHNEL	rvava	3 N G G Q D H G F W G R P	EDM - MSRPAYK I TANRPGSDI AAEYAAA FAVSHLVFKERDAE FADKLLTHSKQLYDFAVHHKARYTDSVSEAAG	199
_023337121.1	KCHTNTNEL	VAQVO	3DG NA DH S Q W - R V	E QM T I HRP S FK I DSQ NPG S DLA GET SAA LAAAS I A FK DV DP S YSS T L L EAK SL FN FAD T Y RGA YSD S VSAA GS	199
023704929.1	KAHVSANVL	GOV	3066 P D H E Y WG R P	EDMINISTRATETISTINGSDLAAETAAALAAASTVEKSSDSSTANTELTHAKULEDAANTKOKTSDSTTDAUS	200
023716596.1	KCHVSPNEL	GQVO	OF N L D H T F WG R P	E D L S MS R P A Y K I D T Q H P G S D L A G E T A A A L A A V A L V F R D V N P G Y S S K C L E H A K Q L Y R F A S Q Y R G L Y - E A I K G A A Q	199
023718476.1	KAHVSPNEF	GQVO	GE V D H S I WC R P	E D I P MH R P A Y K I D T S H P G S D L A S E T A A A L A A A S I V F Q R N D S S Y A D I L L T H A K Q L F D F A N K Y R G K Y T E S I T G A K G	198
XP_396791.3	KCHVSEYVF	GQVO	3 D F S L D H T F WG R P	E E L N T T R P A Y K I D P D H P G S D L A G E T A A A L A A SS I V F R NY N P E Y S A N C L K H A K E L Y K F A S K Y R G L Y - E A I R G A A Q	199
AAA23086.1	KAHTAPNEL	r v q v q	G D G E A D H K W W G P A	E V M T MA R P S H K I S A S C P G S D V A A E T AA A L AS SA I V L KGD D P A Y A A T L V S HA K Q L Y T F A D T Y RGA Y S D C V T A A S A	200
AAD20027 4	EVNEWEVOD	li a sell		I DRAVCHUSEENICTRS OFOWDSVAQUOALI VII DOSSEVITALNOEIN EVDNOADTREOLVELDAWOSND	207
AAD30027.1	VYSSYGYED	LVW	ALWLARATODNS	LITKA FOYY TO FGI Y DG Y FSWDOKK A GVEVI I A FITK DOGH I TO YCD MV RGOT RTPKGI VY I SDWGSK	208
AAF80585.1	FYRSNEFED	LVW	ALWLYKATMDES	LITKAQQYYDDFGINPWFSWDQKCTSSQLLAQITQEQQYIKITAYCDHMISGQQRTPKGLVYIDTWGSLR	297
AAK12339.1	FYASGDYKD	LVW	ATWLYRATNDNT	V LT KAESLYN EFGLGNWNGAFN WDN KVSGVQVLLAKLTSKQAYKTVQGYVDYL I NNQQKT PKGLLY I DQWGTLR	300
AAO61672.2	FYNSWSYGD	LAW	ALWLARATGDNS	DRAKGH VSEFN LGT PS-QFGWDDKKAGVQALL VLLDGSSE Y TAL NQFLN - FV RNQAPT PEGL V FLDAWGSN R	297
AAU20853.2	FYASGDYKD	LVW	AAWLYRATNDNT	T T KAESLYN EFGLGNWNGAFN WDN KISGVQVLLAKLTSKQAYKKVQGYVDYL I SSQKKT P KGLVY I DQWGTLR	300
AAX92641.1	FYASSAYED	LAW	AAWLYRATEEQE	L DRAYEFG T T T AW AY DWNEK I V GYQL LLT T SAGQ T D F LRV EN F LRNWF PGGS V T PLGLAWLAQWGPNR	295
ABD24274.1	YYNSANVTD	LCW	SLWLYKATKEPK	LEEALKHYD AS PDW GMS WDDVF I GNQVLYEL TKEAK YKAVEGTFKEWFPGGTVTPKGLAYRLQWGALR	295
ABD24275.1 ABD24276.1	YYNSVNVTD	LCW	SLWLYKATKEPK	LEEALKHYD - AS PDW - GMSWDDVFIGNQVLLYELIKEEKYKAVEGIFKEWFPGGIVIPKGLAVRIQWGALR	294
ABD24277.1	YYNSVNVTD	LCW	SLWLYKATKEPK	LEEALKHYD AS PDW GMSWDDVF I GNQVLLYELTKEEKYKAV EGT FKE WF PGGT VT PKGLAYRLOWGALR	294
ABD24278.1	YYNSANVTD	LCW	SLWLYEATKEPK	LEEALKHYD AS PDW GMS WDDV F I GNQVLL YE LTKEAKYKAV EGT FRE WF PGGT VT PKGL AY RLQWGSLR	294
ABD24279.1	YYNSANVTD	LCW	SLWLYKATKEPK	LEEALKHYD ASPDW GMSWDDVFIGNQVLLYELTKEAKYKAVEGTFREWFPGGTVTPKGLAYRLQWGSLR	294
ABD24280.1	YYNSANVTD	LCW	SLWLYKATKEPK	V LEEALKHYD AS PDW GMS WDDVF I GNQVLLYELTKEAKYKAVEGTFREWFPGGTVTPKGLAYRLQWGSLR	294
ABD24281.1	YYNSANVTD	LCW	SLWLYKATKEPK	VLEEALKHYD AS PDW GMSWDDVFIGNQVLLYELTKEAKYKAVEGTFREWFPGGTVTPKGLAYRLQWGSLR	294
AB026609.1	EVNEVEVED		A VWLYKA I RDKK	LEDAKIYHS - SASPW - AYSWDDIIVGCQLMLYDMIRDVIYREVQRFLISWKPGGS - IPCGMAFRSKWGSLK	292
ACE75510 1	FYASSAVED		A AWLYRATGEOF	I DRAYEFG TTTAW. AYDWNEKI VGYOLLI TTSAGODELRV ENELRWEPGGSVTPLGI AWLAOWGPNR	205
ACE75511.1	FYPSNGYED	LAW	AAWLYRATGEQY	LDRAYEFG STVAWAYDWDEKTVGYQLLLTTFAGYTDFQPVEDFLRNWFPGGSVTPLGLAWLIEWGANR	295
ADB 12483.1	FYASGDYKD	LVW	ATWLYRATNDNT	T L T KAESLYN EFGLGNWNGALNWDNKVSGVQVLLAKLTSKQAYK TV KGYVDYL I NDQQKT P KGLLY I DQWGTLR	300
ADB82658.1	YYTSYDYRD	LVW	AAWLYRATNDIT	IN TAESLHNQFGLQDWNGGFSWDAKVSGVQLLLAKLTNKQQYKAIKGYVDYLINAQQKTPKGLLFIDVWGSLR	300
ADB85440.1	FYSSYSFHD	LPW	ALWLYLATGENS	LSDAEGYIQEFGLDVL NVDWDDKTAGVYALMVE IDGGSQYTALQGYVDNLL SYQTPGGMVY ISDWGSLR	292
ADB85441.1	FYNSWSFQD	LGW	A CWLYKATGEQS	VY NOAE GY YN SFNMNDQE VNFY WDDK TA GVYALMT E FOQDSK YNR FAQFLD - SMRTSQRT NAGL I FINDWOSI R	297
ADB65442.1	EVESYDYED	LYW	ALWLYRATGEEE	INTAARGETEEFGEMDUKVDWDVKIGAITALIAELDUGGETEKLEETAEMLIITTIPEGLIFIDVGGER	294
AFQ98379.1	FYSSSGYKD	MCE	AMWLYRATGDKS	LADAKG-YNAWALGWDDKKIACQLLLYEATRDTAYKEYEGFFKGWLPGGS-TPCGQAWRDKWGSNR	290
AFQ98380.1	FYSSSGYKD	MCE	AMWLYKATGDKS	ADAKG - YN AW - AW ALGWDDKK I ACQLLLYEATKDTAYKEV EGFFKGWLPGGS - TPCGQAWRDKWGSNR	290
AG (61069.1	FYSSSGYKD	MCE	AMWLYKATGDKS	X L A D A K G - Y N A W - A W A L G W D D K K I A C Q L L L Y E A T K D T A Y K E V E G F F K G W L P G G S - T P C G Q A W R D K W G S N R	290
AGP76397.1	FYTSADYSD	LVW	AAWLYRATNDSS	INDAESLYNEFGLQSWGGGLNWDSKVSGVQVLLAKLTNKQEYKTIQSYVNYLINKQQKTPKGLLYIDKWGTLR	300
AGP76398.1	FYASADYRD	LVW	AAWLYSATNDSS	V LN TAESLYNEFGLLSWGGGLNWDSKVSGVQVLLAKLTNKQEYKT I QSYVNYL I NNQQKTPKGLLY I DKWGTLR	300
AGP76399.1	TYASADYADI	LVW	AAWLYSATNDSS	VLNTAESLYN EFGV XSWGGGHNWDSKVSGVQVLLAKLTNKQEYKTI QSYVNYLINNQQKTPKGLLYIDKWGTLR	300
AGP76400.1	FYGSWDYRD		AAWLYRATNDM I AAWLYRATNDMS	I TAESLYNEFGI VYWGVLCWDHKVIGVQVLLAKI RAKUEYKSI RGYMDYLLNDUGKI PKGL FLDEWGSLR	300
AGP76402.1	FYASADYRD	LVW	AAWLYRATNDST	LT TAESLYN EFG LQNWFGGLNWDSKASGVQVLLAK FTNK QAYKN I QTYVDYL I YNQQKTPKGLL FIDMWGSLR	300
AGP76403.1	FYASADYRD	LVW	AAWLYRATNDST	T T A E S L Y N E F G L Q N W F G G L N W D S K A S G V Q V L L A K F T N K Q A Y K N I Q T Y V D Y L I Y N Q Q K T P K G L F I D M W G S L R	300
AGP76404.1	FYASADYRD	LVW	AAWLYSATNDST	TTAESLYN EFG LQNWFGGLNWDSKASGVQVLLAK FTNKQA YKN I Q TYVDYL I YNQQKTPKGLL FI DMWGSLR	300
AGP76405.1	FYGSWDYRD	LVW	AAWLYRATNDMT	Y L T T A E S L Y N E F G L V Y W F G V I S W D N K V T G V Q V L L A K L T A K Q E Y K S I K G Y M D Y L L N D Q Q K T P K G L F L D E W G S L R	300
AGP/6406.1	FYASADYRD	LVW	AAWLYSATNDSS	LENNAESLYN DFGLQNWGGSLNWDSKASGVQVLLAKLINKQEYKTLQSYVNYLINNQQKTPKGLLYIDMWGSLR	300
AGP76407.1 AGP76408.1	FYASADYQD		AAWLYRAINDSS AAWIYRAINDSS	I KYAESLYN EFGLONWGGSLNWDSKASGVQVLLAKLINKQWYKII QSYVNYLII NQQKIPKGLLYIDMWGSLK	300
AGP76409.1	LYASADYRD	LVW	AAWLYRATNDSS	L K TA E SLYN E FG LQ SWF GG LN WD S KV S G VQ V LL A K LT N KQ E Y K N LQ S Y N Y L I Y NQ K T P KG LL D I D MWG T L R	300
AGP76410.1	FYASADYRD	LVW	AAWLYSATNDSI	V L K TA E S L Y N V F G L Q N W F G G L N W D S K V S G V Q L L L A K L T N K Q E Y K N I Q S Y V N Y L I Y N Q K T P K G L D I D M W G T L R	300
AGP76411.1	FYSSEDYRD	L V W	ATWLYKATNDSI	IN IVESLTDEFGLQNWGGSINWDSKATGVQVILAKLATKQAYKTVQAYVDYLIKTQHKTPMGLTHIDMWGSLR	300
AGP76412.1	YYSSSDYRD	LVW	AAWLYRATNDST	Y L NNAESLYN OFGLODWNGGFSWDAKVSGVOLLLAKLANKQQYKEI QGYVDYLVYTQOKTPKGLVFI DEWGSLR	300
AGP/6413.1	YYSSYDYRD		AAWLYRAINDST	INNAESLYNGFGLQDWRGAFSWDARVSGVQULLARLANRQQYREIQGYDDLUYTQQRIPRGLVFIDEWGSLR	300
AGP76415.1	YYSSYDYRD	LVW	AAWLYRATNDST	L NKAESLYN EFGLODWNG ANRWDSKYSGYOVLLAKLISKOH YKEI TGYVDYLVYTOOKTPKGLVFI DVWGTLR	300
AGP76416.1	YYSSYDYRD	LVW	AAWLYRATNDST	V LN TAESLYN EFGLQDWNGDLSWDAKVSGVQLLLAKLANKQQYKAILGYVDYL I DTQQKTPKGLLH I DVWGSLR	300
AGP76417.1	FYASADYRD	LVW	AAWLYRATNDSN	V L N N A E S L H N E F G L Q K WG G S L N WD S K A T G V Q V L L A K L T N K Q E Y K T V Q A Y V D Y L I N N Q Q K T P MG L L Y I D M WG S L R	300
AGP76418.1	FYASADYRD	LVW	A A W L Y K A T N D S N	Y L N N V E S L Y N E F G L Q N WG G S L N WD S KA T G V Q V L L A K L T N K Q E Y K T V Q A Y V D Y L I N N Q Q K T P MGL L Y I D M WG S L R	300
AGP76419.1	FYSSYDYRD	LVW	AAWLYRATNDST	VLNTAESLYNQFGLQDWNGRLSWDAKVSGVQLLLAKLANKQQYKAILGYVDYLIDTQQKTPKGLLFIDVWGSLR	300
AGP76420.1	YYSSYDYRD	IVW	AAWLYRATNDST	I NSAESI INOFGI ODWNGGERWDAKYSGVOLLEAKT TIKOYKAT RGYVDYL INTOKT KGI FIDI WGT R	300
AGP76422.1	LYASGDYMD	LVW	AAWLYRATNDNN	V LNIAESLYDEFGIQYWCGAFNWDNKATGVQVLLAKLSNKQEYKTIKGYVDYLINNQQKTPMGLFYIDMWGPLR	300
AGP76423.1	PYASGDYMD	LVW	A A W L Y R A T N D G N	IN IVESLYDEFG I QYWCGAFN WDN KA TGVQVLLAK LSN KQR YK T I KGYVA YLMN NQQ KT PMGLFY I DMWGPLR	300
AGP76424.1	FYASGDYRD	LVW	AAWLYRATNDNN	I L N I A E S L YN E F G I Q YWPG A F NWD N KA S G V Q V L L A K F S T K Q E Y K T I R G Y VD Y L I N NQ Q K T P K G L V Y I D HWGP L R	300
AGP76425.1	FYASGDYRD	LVW	AAWLYSATNDNN AAWLYBATNDNN	L NIAESLYN EFGI QYWGGAFNWDNKASGVQVLLAK FSTKQEYKTI RGYVDYL INNQQKTPMGLLYIDHWGPLR	300
AGP76420.1	FYASADYRD	IVW	ATWLYRATNDSS	I NTAESLYN FEGI RAWUGAEN WDRAASGYQYLLAKT TN GEYK NESSYMNYL I NDOKT FKGLLYD DWGFLR	300
AGP76428.1	FYASSDYRD	LVW	ATWLYRATNDSS	VIN TAESLYN EFGLQSWFGGLNWDSKVSGVQVLLAKLT I KQEYKNSSSYMNYL I NDQQKT PKGLLYLDKWGTLR	300
AGP76429.1	FYASADYRD	LVW	ATWLYSATNDSS	N TAESLYN EFGLQSWGGGLNWDSKLSGVQVLLAKLTNKQEYKNSSSYMNYLLNDQQKTPKGLTLLGKWGTLR	300
AGP76430.1	FYGSWDYSD	LVW	AAWLYRATNDMT	LT TAESLYNEFGLVYWFGILSWDNKVTGVQVLLAKLTAKQQYKSIKGYMDYLLNDQQKTPKGLLFLDEWGSLR	300
AGP76431.1	FYASADYRD	LVW	ATWLYRATNDSS	V L N TA E S L Y N E F G L Q S W G G G L N W D S K V S G V Q V L L A K L T N K Q E Y K N S S S Y M N Y L I N D Q Q K T P K G L L Y L D K W G T L R	300
AGP/6432.1 AGP76432.1	FYASADYRD		ATWLYRAINDSS ATWLYSATNDSS	INTAESLYNEFGLQSWGGGLNWDSKVSGVQVLLAKLINKQEYKNSSSYMNYLINDQQKIPKGLLYLDKWGILK	300
AGP76434 1	FYTSCDYRD	I VW	AAWI YRATNDST	I NTAESI HNOFGI OHWNGGET WDAKYSGYOLLI AKI TNKOFYKA I KSYYDYI I NAOKKTPKGI I YIDYWGTI R	300
AGP76435.1	FYTSCDYRD	LVW	AAWLYRATNDST	IN TAESLHNQFGLQHWNGGFTWDAKVSGVQVLLAKLTNKGEYKSLKSYVDYL INAGKKTPKGLSY I DVWGTLK	300
AGP76436.1	YYTSADYRD	LVW	AAWLYRATNDST	N TAESLYKEFGLQHWNGGFSWDAKVSGVQLLLAKLTNKQKYKSIKSYVNYLIKAQQKTPKGLLFIDEWGSLR	300
AGP76437.1	YYTSADYRD	LVW	AAWLYRATNDSS	VLNTAESLFKEFGLQHWNGGISWDAKVSGVQLLLAKLTNKQEYKSIKSYVNYLIKTQQKTPKGLLFIDDWGSLR	300
AGS32241.1	FYASGDYKD	LVW	ATWLYRATNDDA	L TRAESLYNEFGLGNWNGALNWDNKVSGVQVLLARLTKKQEYRTVQGYVDYLINNQQKTPKGLLYIDQWGTLR	300
AC 11911.1 AKC91349 4	FYKSWSYGD	LAW	AAWLLRATNEHD	Q I EVEK HFMEFG LKRPT - EFGWDNKLAGSQALLSL I DGGSI TIALSSFLNSVK - SHAIPGGMVY I SEWGSAR	295
AKV16371 1	FYASGDYKD	LVW	AAWLYRATNDNT	LTKAESLYNEFGLGNWNGAFNWDNKISGVQVLLAKLTSKQAYKKVQGYVDYLISSOKKTPKGLVYIDOWGTIR	300
AMH40356.1	FYSSSGYVD	LVW	AAWLYRATQKQT	LDIVDKYWSTVMYND-GFGWDSKKSGADAIMYKLTKDDKYKRLVAYCDYCFNSVQKTPKGLIFLGEWGSLR	296
AMH40357.1	FYASSGYVD	LVW	AAWLYRATNSKT	LNIVDQYYS SVMYND - GFGWDNKKSGADAVLYKVTKESKYKRLVAYCDYSFNSAKKTPKGL I FLGEWGSLR	296
AMH40358.1	CYTSTHYED	LVW	AAWLYRATNDKK	VLGYVDELWD SFNMY GFDWDQKFGGAVVLLAKLTGDTKYKA I TNYCNYVYNQGTRTSKGL I FIYDWGPLR	295
AMH40359.1	GYNSSGYED	LVW	A A W L Y R A T N D S T	L UT TA DV TT P T Q Y T - GF NWDQ K T GGADVLLAK L TGESKYLKMEN YCDNMLANQQ RTP KGL F TA EWGSLM	296
AMH40361 4	FYPSSNVPD	LAW	AAWLYRATGEKK	MTVVDEIVP - TIQYWD-GFSWDDKKSGLDLLLATLTKESKYLKLKSSCOVAENTOKKTPKGLLYIDQWGTLR	299
AMH40362.1	FYPSSNYPD	LAW	AAWMYRATREOK	MNVVNEIYP TIQYWD - GFSWDDKKGGLDLLLATLTGESKYLKLKTFCDYAFNKGTRTPKGLIFYSFSASLR	296
AMH40363.1	FYSSSSYAD	LVW	AVWLYKATKDSS	TLTKAESLYQEGNLQYSNGGFDWDTKTSGVEILMSQISSNSIHKKAKSYLDYMVNNQQKTPKGLLYIDQWGTLR	299
AMH40364.1	CYTSTHYED	LVW	AAWLYRATGDKT	LKYVDALWD KVSMY GFDWDQKFGGAMVLLAKLTGEAKYKAITNYCNYVFEKGTRTSKGLIFIYDWGPLR	295
AMH40365.1	CYNSSGYED	LIW	AAWLYRATKDKT	LTYVEEVYP TIQYYI - GFNWDQKIGGADVLLAKLTGDSKYLKTKNYCDNMIENQQRTPKGLIFIAEWGSLM	296
AMH40366.1	CYNSSGYED	LVW	AAWLYRATEDKT	LTYVDKEYT PIQYYI - GFSWDQKAGGADVLLAK ITGDSKYLKMKNYCDQIMTTQQRTPKGLIFISQWGSLS	296
AMH40367.1	FYASSGYVD	LVW	AAWLYRATNNKE	I T I A E QN YN SY MYND - GF GWDNKK SGADAVL YK I TKEAK YKR LEA YCDYQ I KSAKKTPKGL I FLGEWGSLR	296
AMH40369.1	FYPSSNYPD	IVW	AAWMYRATREOK	MKVVDELYP - TIQYWD GFSWDDKKGGIVLLANLIGDSKYLLKILKTFCFYAFNKOSRTPKALIFYOFSASIR	295
AMH40370.1	FYASSGYVD	LVW	AAWLYRATQKOT	LDIVDKYWSTIMYND-GFGWDSKKSGADAIMYKVTKEEKYKRLKIYCDYGFNNVGKTPKGLIFLGEWGSLR	296
AMH40371.1	CYNSSGYED	LIW	AAWLYRATKDKT	LTYVEEVYP VIQYYIG - FNWDQKIGGADVLLAKLTGDSKYLKTKNYCDNMLATTPRTPKGLIFITQWGSLM	296
AMH40372.1	FYRSSSYAD	LVW	AVWLYKATNDKT	T L T KAESLYQTGNLQYSNGGFDWDTKTSGVEILMSQVSSNTIHKKAKSYLDYMVTNQQKTPKGLLYIDQWGTLR	299

AMU/0272 1	CANSSOVED	I VIN	AAWMYPATNDST	I KYVDNEVR. TIOVVI, GESWDOKAGGADVI I AK ITGOSKVI KITNYCEO IVNSOORTBKGI I EISOWGSI S 2	200
AMH40373.1	EVERGEEADE	LVW	AAWMYRAINUSI AVWI YKATEDKEN	LTO A SET YO NE TO STATE A	290
AMH40375.1	CYVSTHYEDE	IVW	AAWI YRATNDKT	TYVDALWD SENMY GEDWDOKYAGAVVLLAKI TGDSKYKALTNNCNYVEQQAARSNKGLLYLANWGPLR	295
AMH40376.1	FYMSSSYODE	LVW	AAWLYRATKDKS	LDIINKYWP VIMYWD - GFDWDNKKSGAEVILYMTTRDDKYRRVKSYCDNTFFKTPKTPKGLLLTNAGGNLR 2	296
AMH40377.1	FYASSGYVDE	LVW	AAWLYRATNSKS	L D I V N Q N Y N S V M Y N D - G F G W D N K K S G A D V I L Y K Q T K D D K Y K R L I S F C D Y N I N S A Q K T P K G L I F L G E W G S L R 2	296
AMH40378.1	CYNSSGYEDE	LVW	AAWLYRATQDKT	LQYAEAVYP PIQYYI - GFNWDQKIGGADVLLAKLTGDSKYLKMQKYCDDTLANQQRTPKGLIFIAEWGSLM 2	296
AMH40379.1	FYRSSSYADE	LVW	AVWLYKATNDKT	LTQAESLYQEGNLQNSNGGFDWDTKTSGIEILMSQISSNAIHKKAKSYLDYMVNNQQKTPKGMLYIDQWGTLR 2	299
AMH40380.1	CYVSTHYEDE	LVW	AAWLYRATNDST	LKYVDGLWD TFGMY GFDWDQKFGGAVVLLANLTGDSKYTAITNFCNYVFQQAPKTSKGLIFIVNWGPLR 2	295
AMH40381.1	CYNSSGYEDE	LVW	AAWLYRATGDKT	LTYAEKEYT PIQYYI - GFSWDQKAGGADVLLAKITKDSKYLKMKNYCDQIMNSQQRTPKGLIFISQWGSLS 2	296
AMH40382.1	FYPSSNYPDE	IVW	BAAWLYRATREQK	MNVVNEIYP TIQYWD - GFSWDDKKGGVDLLLADLTKDSKYLKLKTFCDYAFNKQAKTPKGLIFYSESASLR 2	296
AMH40383.1	FYDSSGYVDE	IVW	AAWLYRATNSKN	LNIVNQNYN SVANND - GFGWDNKKSGADAILYKVTKEAKYKKLESYCDYQIKSAKKTPKGLIFLGEWGSLR 2	296
AMH40384.1	FYSSSSYADE	LVW	AVWLYKATQDSS	LTKAESLYQEGNLQYSNGGFDWDTKTSGVEILMSDISSNTIHKKAKSYLDYMVNNQQKTPKGLLYIDQWGTLR 2	299
AMH40385.1	CYNSSGYEDE	LVW	AAWLYRATGDKT	LTYVDTEYT NIGYYI - GFSWDQKAGGADVLLAKLTGESKYLKMKNYCDQIINSGQRTPKGLIFISQWGSLS 2	296
AMH40386.1	CYPSIHYEDE	LVW	AAWLYRAINDKI	LNYVDILWD - SVSMY - GFDWDQKFGGAVVLLAKLIGDIKYKAIINYCNYVFQQGSRISKGLIFIYDWGPLR 2	295
AMH40387.1	EVAREGUIDE	LIVIO	AAWLYRAINUSK AAWLYRAINUSK	L STAEEVTP I UTT - GENWOORI GGADVLLAKLIGDSKTLKIKNTCDNMLANQQRIPKGLIFI I UWGSLM 2	290
AMH40388.1	FYASSGYVDE	LVW	AAWLYRATORUT	LDIVDRIWS WWIND - GFOWDSKKSGADAINI VVIKEDRINKERI I GDI GFNNVGRIFRGI IFLGEWGSLK 2	206
AMH40309.1	CYVSTHYEDD	LVW	AAWLYRATNOST	I TV DALWD - TESMYN - GERWDARY GADVLLIAOL TROSKYTALTNE CDYVEDOARKTSKOLLE LVNWGOLP 2	205
AMH40391.1	EYPSSNYPDE	LAW	AAWMYRATGEKK	MSVVDELYP TIOYWD-GFSWDDKKGGLDLLATLTGOSTYLKLKSYCDYOFTKTTKTPKGLIYYSDSANLB 2	296
AMH40392.1	FYKSSSYEDE	LVW	ATWLYRATNDNO	LSVVNELYS GIQYNDG - FGWDSKKSGADAVLYRLTKDSTYSRLTAYCDYTFNSAPKTPKGLIFLGDWGSLR 2	296
AMH40393.1	CYTSTHYEDE	LVW	AAWLYRATNDQK	LGYVDELYS TV SMY GFDWDQKFAGAMTMLYKVTKDEKYADLKNYCDYVINSSSRTPKGLIFIYDWGPAR 2	295
AMH40394.1	CYNSSNYNDE	LVW	AAWLYRATEDNO	LNVIEQEYP TVQYWI - GFDWDQKTSGADVLLAQLTGDSSYSKTKNYCDNIINNQQRTPKGLIFISEWGSLR 2	296
AMH40395.1	YYESTDYGDE	LTW	AAWLYKATSESQ	L D E A E H N Y M K F R L R P N E F F Y N K K V A G V Q M L L A Q L T L Q P E Y M A V Q N F C D F T V - N Y Q K T P K G L V Y I D K F G T L G 2	296
ANR02619.1	FYRSGGYQDE	LCW	ALWLYRATGEQD	MNKANEFLGRPWAYSWDAKEAGSLVLLTSF-GNNNARQLEDFLQSWFPGGGITPQGLAWRDTWGSLR 2	291
AOV94250.1	YYASTSYEDE	LIW	AAWLYRATNDQS	MSVVDELWE GIMYDDG - FGWDNKKSGADVILYKLTQDSKYTRLKAMCDYVFNSTPRTPKGLIFLGDWGSLR 2	296
AOV94251.1	FYNSYNYQDE	IVW	AAWLAKATYDQK	LEQVYRLYQEFGFEYRYTNLSWDDKTAGIEALMAELQIDGNHYKLQSYVNDIVNSRQKTPKGLVFIDEWGTLR 2	299
AOV94252.1	FYRTITYEDE	LVW	A M W L Y K A T E E A D	L S Q A E D L Y Q E L R P E A Q R T G F D WE Q K I S G A E I L L A E Y S S N A T F K N T Q N Y V D Y L V N N Q T R T P K G L A F I A T W G P L R 2	299
AOV94253.1	YYESTDYGDE	LTW	AAWLYKATGEVR	L D D A E H H Y Y R L E R P N - E F F Y N K K V A G V Q I L L A Q L T H Q Q E Y V A A Q A F C D F T V - D L Q K T P K G L V Y I D K F G T L C 2	295
AOV94254.1	FYDSFNYGDE	LVW	AAWLAKATNDQS	L D KAND L YE EFS I AS L KTNLS WG D KTAG I EALMAQ I TDSDSHKKLKS YVED VMNNRPRTPKGL VYLNDWGSLR 2	299
AOV94255.1	FYNSYNFNDE	IVW	ALWIAKATNDOS	T Q R A Q S L Y D E F Q M S T I D L N F S W D D K S A A I E A L F T E L T N D D N H K K L Q S Y A E D V L H S R P R T P M G L V F I D A W G S L R 2	299
A0 V94 256.1	FYDSFNYGDE	LVW	AAWLAKATNDQS	LDKANDLYEEFSTASLKINLSWGDKIAGTEALMSQIIDSDSHKKLKSYVEDVMNNRPIPKGLYYLNDWGSLR 2	299
APF29000.1	EVGSWSYGDE	LCW	ALWLYRAINDAS	LIKAUGEWDETGIATGI-DTSWDQRSAGAMALFSLIDGGSTTTALSSFLINIRSQPIPGGMVFVSEWGSAR 2	290
APF29558 1	FYKSWSYGDE	LCW	ALWLYRATNDES	I TKAKGYWDEFGETPS - FESWDNKNAGSOALI SI I NGGSTYTAL SSELNS IR - SEPTPGGNVVVSEWGSAR 2	205
APF295591	FYRSWSYGDE	LCW	SLWLYRATGDVN	LNKAKQFWGEFGVPSEFSWDNKNAGSQALFSMLDGGSFYTALSSFINSMRSKPTPGGMVVVSFWGSAR 2	285
APF295601	FYRSWSYGDE	LCW	ALWLYRATWDTG	LDAAKQHWNDFGINPPG-EFSWDNKIGGIYALFTLLQDDATYSALNOFLTNLRYLPTPGGMVFIFFWGSAR	296
APF29561.1	YYMSSGYIDE	LLW	SMWLYRATGDTS	R D R A R A H H G E L D P Y H N M - Q F G W D D K R A G A Y A L G M V L D E F T D Y E A I R N Y F D - WL H N E A T T P D G L V F L D A W G A N R 2	297
ATS91249.1	FYRSNNITDE	VCW	SLWLHKATNDDK	LAEAEKQFD DP - AW GQSWDEKT I GNQLLLYLATKKDNYKS I LNSFEGWFPGGKVTPKGLPFRLEWGSLR 2	294
BAA31326.1	FYASGDYKDE	LVW	AAWLYRATNDNT	LTKAESLYNEFGLGNWNGAFNWDNKISGVQVLLAKLTSKQAYKKVQGYVDYLISSQKKTPKGLVYIDQWGTLR 3	300
BAA33708.1	FYASADYRDE	LVW	AAWLYRATNDNT	LN TAESLYDEFGLQNWGGGLNWDSKVSGVQVLLAKLTNKQAYKTVQSYVNYL INNQQKTPKGLLYIDMWGTLR 3	300
BAA33709.1	FYASADYRDE	LVW	AAWLYRATNDNS	LN TAESLYN EFGLQNWGGGLNWDSKVSGVQVLLAKLTNKQEYKTIQSYVNYL INNQQKTPKGLLYIDMWGTLR 3	300
BAA34050.1	FYASGDYKDE	LVW	AAWLYRATNDNT	LTKAESLYNEFGLGNFNGAFNWDNKVSGVQVLLAKLTSKQVYKKVQSYVDYLISSQKKTPKGLVYIDQWGTLR 3	300
BAB40693.1	FYASGDYKDE	LVW	AAWLYRATNDNT	LTKAESLYNEFGLGSWNGAFNWDNKISGVQVLLAKLTSKQAYKKVQGYVDYLVSSQKKTPKGLVYIDQWGTLR 3	300
BAB40695.1	FYASGDYKDE	LVW	AAWLYRATNDNT	L T KAESLYN EFGLGSWNGAFNWDNKISGVQVLLAKLTSKQAYKKVQGYVDYLVSSQKKTPKGLVYIDQWGTLR 3	300
BAB40696.1	FYASGDYKDE	LVW	AAWLYRAINDNI	LIKAESLYNEFGLGSWNGAFNWDNRISGVQVLLARLISROAYKRVQGYVDYLVSSQKRIPRGLVYIDQWGILR 3	300
BAB40697.1	FYASGDYKDE	LVW	AAWLYRAINDNI AMWLYKATODKS	LIKAESLYNEFGLGSWNGAFNWDNKISGVQVLLAKLISKQAYKKVQGYVDYLVSSQKKIPKGLVYIDQWGILR	300
BAC0/100.1	EVERECYKDE	MCE	AMWLYKATODKS	LADAKG IN - AW - ALGWDDKK LACQULLI VEATKDTAINEVEGFFKGWLFGGS - FPCGQAWRDKWGSNR 2	200
BAD44734 1	FYSSSGYKDE	MCE	AMWLYKATODKS	LADARG YN - AW - ALGWDDKK LACOLLLYEATKDTAYKEVEGEFKGWLPGGS - TPCGOAWDDKWGSNE 2	200
BAE38757.1	FYGSSGERDE	VAT	AAWMYKATNDNG	I T DA OS I Y G T PW G FA WN DA OS GAAL I I Y F A T OD NA YK DI VA F IK S YM PGGG - T P CGLAWR DOWGPNR 2	289
BAH22180.1	FYSSSSYVDE	LAW	AAWLYRATNEQTY	LNYALEFAS A I - SW AYDWNEKIVGYQLLLFSSAGQTVFQPVEGYIRSWMPGGSVTPQGLAWRQQWGPNR 2	295
BAK20401.1	FYRSSAYEDE	LCW	ALWLYRATGEOD	MDKANEFLG RPW AFSWDSKEAGSLVLLTSFG - NSNARQLEDFLQSWFPGGD I TPLGLAWRDTWGSLR 2	291
BAL60587.1	YYKSWSFNDE	LVW	AAWLAKA - GESS	LSKAEQLFDEFNLGAY - AQFGWDDKRPGVLALLTDLTGSNKYKLLTKWTNDMM NKPSPGGMVYIDQWGTLR 2	294
BAM14716.1	FYASSAYEDE	LAW	AAWLYRATEEQE	LDRAYE - F GTTAW AYDWNEKIVGYQLLLTTSAGQTDFLRVENFLRNWFPGGSVTPLGLAWLAQWGPNR 2	294
BAS18736.1	FYGSSGDKDE	MCV	AMWLYKATRNNO	LEDAKAHEN AW - GW ALAWDDKKVACQLLLFEETNDAAYREVEGFFQGWLPGG TPCGLAWRDKWGANR 2	270
BAT21145.1	FYASSAYEDE	LAW	AAWLYRATEEQE	LDRAYE - F GTTAW AYDWNEKIVGYQLLLTTSAGQTDFLRVENFLRNWFPGGSVTPLGLAWLAQWGPNR 2	294
BAU45387.1	FYGSSGDKDE	MCV	AMLLYKATRNNQ	LEDAKAHEN AW - GW ALAWDDKKVACQLLLFEETNDAAYREVEGFFQGWLPGG TPCGLAWRDKWGANR 2	270
BAU45388.1	FYGSSGDKDE	MCV	AMWLYKATRNNQ	LEDAKAHEN AW - GW ALAWDDKK VACQLLLFEETNDAAYREVEGFFQGWLPGG TPCGLAWRDKWGANR 2	270
BAX09271.1	FYQSGNWKDE	MCL	AVWLHRATGTHS	LNDAKNYVDLKPWAEGWDDKMLACQLLMYEETKEAKYKAVVSFFNDWLPGATPCGLAWRDLWGANR 2	271
CAD54726.1	FYSSSDYRDE	LVW	AAWLYRAINDOI	LN TAESENDDYGLOYWNYHFTWDSKALGATALLAK TIGKOTYKGLTSYCDSLV - SROOTPKGLTFLDOWGSLR 2	299
CAD54727.1	FYSSSDYRDE	LVW	AAWLYRAINDUT	INTAESINDDYGLQYWNYHFIWDSKALGATALLAK ITGKQTYKGLTSYCDSLY-SKQQPKGLTFLDQWGSLK 2	299
CAD54728.1	FYSSSDYKDE	LVW	AXWLYRATNDOT	I T TA ESEN DE LA LANNA GET MOTAL CATALLA KITAKATI KULI SI COLUS ANA I PAGI PAGI PAGI VI DEWASIL PA	200
CAD54730 1	FYSSSDYKDE	IVW	AVWLYRATNDOT	I T TAEK I YS DI GLOSWNGGET WDT KISGVEVI I AK I TGKOAYKKY KGYCDY I SGSOOKT PKGI YY I DKWGSI R 3	300
EFN70196.1	YYNNTDYGDE	LAW	AIWLHKAEPHED	LNDAENHYQDFHLKRPN - EFFYNNKI AGVOILLAQLTGKLEYKAY I DFCNFSY - HNOTTPKGLLY I DKLGTLS 2	297
EFN70197.1	YYNNTDYGDE	LAW	AIWLHKAEPHFD	LNDAENHYQDFHLKRPN-EFFYNNKIAGVQILLAQLTGKLEYKAVIDFCNFSV-HNQTTPKGLLYIDKLGTLS 2	298
EFX69372.1	FYSSSGFGDE	LAW	AIWLYRATQDVS	LNDAKTFFNQYGLGAG - AFSWDNKAAGVQVLLAKLTNDPTYISIRTFCDSYV NQPTPKGLLFISQWGSLR 2	294
EFX80604.1	FYPSSDFGDE	LAW	AAWLYRATNNVT	LNDAKAFYTQFGLNNPS-EFSWDSKTAGVQVLMAKLTNDTLYKATKTFCDSKVNQPTPKGLLFISTWGSLR 2	295
EFX80605.1	FYSSSGFGDE	LAW	ASWLYRATNDNA	LDDAKAFYTEFNLDNPS-EFSWDTKTAGVQVLLAKLTNENLYRAAQTFCDSKVNQPTPKGLLFVTKWGSLG 2	295
EFX86822.1	YYSAYAYEDE	LTF	AGMLALATNEQA	K T D A A N F W D Q F G T Q T F F D W D N K M A G I S V L L S R I L G D Q K Y K S A Q A H C D Y W I - N T E R T P K G L V F I N E W G S L R 2	294
EG/63652.1	YYESTDYGDE	LAW	AIWLFKATGDTK	F DDA I HH YQ H F H L E R P N - A F F Y N K V A G I Q V L A E L T G Q S E Y Q A T R A F C D F S V - R Q Q K T P K G L F I D K S G T L S 2	297
EKC1/289.1	FYRSVDYNDE	LAW	GAWLYKATNETK	LI QAE I Y Y GA - SW GQ SWDDK I A GCQVLL Y E EI GKDK YK DI EA I FQD WMPGG SV SPKGL A FRS QWGS LR 2	293
EKG39023.1	FYSSSEYRDE	M C V	AALLYLAINKIG ASWLYPATDDDDO	R SDATTFYN SYSPGRW - AYDWDDRIMLCDVLL YCH QETRYRAVENFVI SYMPGGS - IPCGLAWRLEWGSLR 2	294
KMO95413.1	FESSTDYGDE	LAW	A SWL FRATDDFG	I FHAI HHYOKEK KERN - FEFYNKK AGVOVILLAU FROPE I MAAGA RAFODESVO, KOKTPKOL VI I NKTOLUS 2	297
KOC62382.1	YYESTDYGDE	LTW	AVWLFKATNDTM	LEDAEHHYQHFHLERPN - EFFYNKKVAGVQVLLAQMTGQPEYQAARAFCDFSV - RQQKTPKGLLYIDKFGTLC 2	297
KOF68403.1	FYPSSGYQDE	LTW	AIWLYRATGMKK	LHYAERNYI GGWVWSFSWDEKE I GSMVLLYN LTRKAKYKDL TQ I LQDWL PGGTL TPKGL VYRSSWGVLR 2	294
KOX76353.1	YYESTDYGDE	LAW	ALWLFKATNNTM	FEEAEHHYQHFHLERPN - EFFYNKKVAGVQMLLAQ I TGQLEYQAA RAFCDFSVR - QQKTPKGLLY I DKFGTLC 2	297
KXJ18436 1	FYKSSGYNDE	LVW	AAWLYRATNKAF	LSKAESYYQQFNMKAWTFSWDDKKAGAOLLLAE ITGKSNYKDIKGYIDSWI PGRDVTPKGI AWPDOWGPNP	298
KYB27092.1	YYESTGYGDE	LTW	AIWLYKKTKDVK	IEQAETFYTKFRIDRPN-EFFYNKKVAGIQLLLAELTLRPEYITIKNFCDYTI-KEQITPKGLIFIDKSGTLS 2	297
KYM87195.1	YYESTDYGDE	LAW	AIWLFKATGDTK	FDDAIHHYQHFHLERPN - AFFYNKKVAGIQVLLAELTGQSEYQATRAFCDFSV - RQQKTPKGLLFIDKSGTLS 2	297
KYN01696.1	YYESTDYGDE	LAW	AIWLFKATGDTK	FDDAVHHYQHFHLERPN - AFFYNKKVAGIQILLAELTGQSEYQAS RAFCDFSV - RQQKTSKGLLFIDKSGTLS 2	297
KYN22851.1	YYESTDYGDE	LAW	AIWLFKATGDTK	FDDAVHHYQHFHLERPN - AFFYNKKVAGIQVLLAELTGQSEYQAARAFCDFSV - RQQKTPKGLLFIDKSGTLS 2	297
KYN35700.1	YYESTDYGDE	LAW	AIWLFKATGDTK	F DDA I HH Y QHFHLERPN - AFFYNKKVAG I QVLLAELTGQSEY QATRAFCDFSV - RQQKTPKGLLFI DKSGTLS 2	297
KYQ52831.1	YYESTDYGDE	LAW	A IWLEKATRDTK	F DDAVHHYQHFHLERPN-AFFYNKKVAGIQVLLAELTGOSEYQATRAFCDFSV-RQQKTPKGLLFIDKSGTLS 2	297
KZS03022 4	YYRAVAVED	LAW	AAMLEKAINDIM	KTDAAN FWN OF OF FEDWIN KMAG I VULAULI GUPET QAARAF CDE SV-RQQKIPKGLY I DKEGTLC 2	:5/
KZS057224	YYDSSSEEDE	LAW	A IWI YRATKOPE	INFAKAFYAKGIRNW-AFSWDDKSPOVOMIAKITGEKYKDVKOVAVD-NOEKIPKOLVFINEWOSLKZ	202
KZS070931	FYRSSGYGDE	LAW	AVWLYRATODVA	LNDAKAYYNQYGLGGA - GEFSWDSKAAGVOVMLAKLTNDPTYIAIKSFCDSKV - NOFTFKGI FISOWGALP 2	295
KZS16514.1	FYPSSGFGDE	LAW	ASWLYRATNDVT	LNDAKTFYTQFGLNPW EFSWDSKNPGVQVMLAKLTGETQYKAVKVFCDSKV NQPTPKGLLF I NKWGSLR 2	294
KZS16516.1	FYPSSGYGDE	LAW	ASWLYRASNDVA	LNDAEVFYSQFGLNPWEFSWDSKEPGVQIMLARLTGETQYKAVKAFCDSKVNQPTPKGLLFVNKWGSLR 2	294
OAD61293.1	YYESTDYGDE	LAW	AVWLFKATNDTM	LEDAEHHYQHFHLERPN - EFFYNKKVAGVQVLLAQMTGQPEYQAARAFCDFSV - RQQKTPKGLLY I DKFGTLC 2	297
ODM94862.1	YYESSGFGDE	FLW	AAWLYRVTNESK	QQDYERFYTEFDLNGRPVEASWDDKAAQAQVLLAKVDGSNKYVAARTFCD-WVVDQAPTPKGLVFISQWGSLR 2	298
ODM95900.1	SFSG FGDE	LLW	A A W L Y R V T N E D R F	K Q D Y E T F F T E F N L N R Q P V E A S W D D K T A Q A Q V L L A K V D R S A K Y V A A R T F C D - W V V D Q A P T P K G L V F I S Q W G S L R 2	294
OWF45593.1	FYRSDAYEDE	LSW	GAWLYKVTNETK	LTQAEQFYV GGAAW GQSWDEKIASATILLYELTGKDVYLDIETTFTEWMPGGTVTPKGLAFRLQWGALR 2	295
OWF56607.1	YYNSSGYDDE	LCY	AAMLYKVTGKLE	LE GAEIIFI QSKPW SFSWDDKTVGCQVLLFE ATGKAVYRLVASFLQSYMAGGDVTPKGLAWRDQWGALR 2	296
PBC31/93.1	EVACEAUCE	LAW	A VWLFRAINDIM	LEGAENTINGTELEKEN-EFFTNKKVAGVUVLLAUMIGQPEYQAANAFCDESVY-QQKIPKGLLYIDKFGTLC	10:
PEX25262	FYKSESEED	LAW	AAWL FRAIEEQE	LIKATETO TITAW - ATOWNERTYOTULLI ISAGUIDTEKYENTEKNIKAGYIPLGEAWLAQWGPNE 2	CC:
PSN31180 4	FYSSSGYFDE	LVW	AVWLYKATGDET	LNKAEQLYSOFGMESWIGNEGWEDKISGEVLLSQITKNSKYTKVKGCDWMPROOVIPROLAWKIKWGSNK Z	.00
PVD35337.1	FYKSNNYSDE	VSW	AGWLFKATNONK	LEIAKKGYLGP-AWGFSWDEKLAGNMLLMYSFTNGEAYRDIEDSMKAWSKAGGMTPKCLAFRLOWGALR 2	294
PVD35340.1	YYKSGNYTDE	MTW	AAWLYKATNDQQ	LTDAEQTYIGA-AWGFSWDEKLAGNMVLLYDATHKDLYKDIVATMTDWMPGGSISPKCLAFRLQWGSLR 2	294
PVD35381.1	YYNSVNVTDE	LCW	SLWLYKATKEPK	LEEALKHYD ASPDW GMSWDDVFIGNQVLLYELTKEEKYKAVEGTFKEWFPGGTVTPKGLAYRLQWGALR 2	279
PVD35383.1	YYNSANVTDE	LCW	SLWLYKATKEPK	LEEALKHYD AS PDW GMSWDDVF I GNQVLLYELTKEAKYKAVEGTFREWFPGGTVTPKGLAYRLQWGSLR 2	294
PVD35384.1	YYNSMDYNDD	VCW	SLWMYKATNKTK	MTEALKYCHITPWY-EMNWDHVFLGSQILLYQVTKQETYKSVEVFFQHWFPNGTVSPKGLDYYALPGSLG 2	295
PVD35385.1	YYNSWDYNDD	ICW	SLWMYKATNKNK	MTEALKYC HIYAW EMDWNHVFLGSQILLYQVTKQETYKSVEAYFQHWFPNGTITPKGLGYKEIPGSLR 2	294
PVD35551.1	TYGSSGFKDE	LCE	AMWLYRVTNNVQ	LEDARSNEU - DLWAW - ALSWDDKKIACQLLLYESTGEEVYKTVLGFMQGWY - PDPETPCGLAWRDTWGANR 2	287
PVD36624.1	CHREENICO	MCW	SLWLTRAIKDAK	LICEARAWED AP - AW GMSWDER LAGNQULM VQLI NEAAAVKGIYKYWEPGGEVSPKGLPFRLQWGSLR 2	:92
PVD36906 4	YYRSTNDUDG	MCV	GVWL FKATODDY	I YOAT - A - AV RGH REAHVR GG	19
PVD36986 1	FYSSSAVIDE	LCF	GMWLYRATGNAK	LEIAKAVET - AW - AW - A LGWDDKK I VCOFLLWEETKDASYKAVVGYFDAWT - GGOTPGGI AWPDOWGSNR 2	274
001606454 1	YYESTDYGDE	LAW	AAWLYKASNESR	LEEAEHHYQHFHLERPN-EFFYNKKVAGVVLLAQLTGTOEYQAAHSFCNFSVR-ROKTRKGI VIDKFGTIG	297
_001810693.2	YYESTGYGDE	LTW	AIWLYKKTKDVK	I EQAETFYTKFRIDRPN-EFFYNKKVAGIQLLLAELTLRPEYITIKNFCDYTI-KEQITPKGLIFIDKSGTLS 2	297
_001944774.2	YYESTDYGDE	LTW	AAWLYKATKESQ	L D D A E Y M Y M K Y R L E R P N - E F Y Y N K K V A G V QALL A E L T N Q S E Y T A I Q N F C D Y N L - Y V Q K T P K G L L Y I E K S G T L C 2	297
				109	R
				120	٠
XP (002426466 1		M			
--	---	---	---	--	
VD 002402779 1	YYESIDYAD	EL I WE	AAWLYKASNDLI AVWLEKATNDTI	L DDAEHHYMKFRLRPN EFFYNKKVAGVQTLLAQLIGRPEYIAAQAFCDFIV - DYQKIPKGLIYIDKLGILC 200	
XP_003690676.1	YYESTDYGD	EL AW	AVWLFKATNDTH	LEFAEHHYOHFHIERPN - FEFYNKKVAGVOLLAOMTGOPEYOAARAFODFSVY - GOKTPKGLI Y DKFGTLO 20	
XP_005099312.1	FYSSSGDRD	ELCE	SMWLYRATGNRO	LODANSTWT AW - AW ALGWDDKK VACOELLYEETHNADYRDYEGYESSWLPGGT - TPCGLAWRSLWGANR 27	
XP_005099315.1	FYQSGNWKD	MCL	AVWLKRATGSNS	LNDAKN - YT NW - AW AEGWDDKMLACQLLMYEETKESKYKAV VGFFRGWL - PGSGTPCGLAWRDMWGANR 27	
XP_005101787.1	FYKDGGYKD	EMCL	AVWLYRATKDSK	LNDAKHFYE GV - PW ALSWGDKKVACQFLLYQEG - NNGYKDVTNFLNQFLPG LPCGLAARDDWGSNR 26	
XP_005101789.1	NYRDGGYKD	EMCL	AVWLYRATKDSN	LNDAKSFY - EAGV - PW ALSWGDKKPACQFLLFQE - GNTGYKDVVGFLNQFLPG LPCGLACRDDWGSNR 27	
XP_005111868.2	FYASNDITD	ELAW	SLWLFKATRDNK	MVEAKKHFD PAPGW GQSWDDKTVGNQILFYKLSEKPKYKAIIGTFKSWFPGGDITPKGLAFRLVWGSLR 29	
XP_005111869.2 XP_005111870.1	EVERENITO	ELCW	SLWLYQAINDSK SIWIHEATNDDM	LDEAEKLED PAPAW GMSWDEKTIGNQVLLYKLIGKDKYKATEGSEKSWEPGGATSPKGLAYNQWGSLR 28	
XP_006610535.1	VYESTOYOD		AVWLEKATNDTM		
XP_008181842.1	YYESTDYGD	LTW	AAWLYKATKESO	L DDA EY MY M KYR L ERPN - EFYYN K KVA GVQALL AEL TN QS EY TA I ON F CDYNL - YVQ K TP KGLLY I EK SGTLC 28	
XP_008194455.1	YYESTGYGD	ELTW	AIWLYKKTKDVK	I E Q A E T F Y T K F R I R P N E F F Y N K K V A G I Q L L L A E L T L R P E Y I T I K N F C D Y T I - K E Q I T P K G L I F I D K S G T L S 29	
XP_009049791.1	YYRSMGFKD	ELAW	GAWLYYATKEAK	LTSAEKYHQ GK - SW GLSWDDKQAAEQVLLYKLTKKDIYKDIEATFQDWMPGGGITPKGLAFRLKWGALR 29	
XP_009049836.1	YYKSENITD	ENSW	AIWLYKATSDKK	LADAEAAHI GE - SW GFSWDEKIFRIILLEYNLKKKSNYLDIKATMKAWMPGGSVTPKCLAFRLQWGALR 27	
XP_009049839.1	YYKSNNITD	ELGW	ALWMYKATGDKS	IAEAEKHHE GA - AW GFS WDEKTAGVNMLLYE ITKNNKYKDL IDTFKSWMPGGDLTPKCLAFRLEWGALR 29	
XP_009049840.1	FYPSSQFQD	ELCW	AAWLYRAIGKFI	LDTAEQYYKHSWAFNWEDKIDAISLMLFMLIKKEKYRDIERSLRDWLPGGIVIPKGLVFRAKWGSLR 29	
XP_009049941.1 XP_009058133.1	YYKSTNITD	MCW	SLWMYRVTKEAK	LTEAEKWEQ GP - SW GMS WDDTQA INQVLITKTINKDITKKVIDIENTWMPGGTITEKGIAWRLEWGSLR 29	
XP 009058134.1	FYSSSGYTD	ELCE	AIWLYRATKDQK	LNDAKG-FTAW-AWALSWDDKKVACQLMLYQETQDDMAGAVIGFMKSWLPGGG-TPCGLAWRDQWGSTR 29	
XP_009060290.1	FYSSSTYKD	LCI	AIWLYRATNDSK	LNDAISFHN AW - AW ALSWDDKKIACQMLVYQYTLKASIKEVEGFFNNWLPDSGTTPCGFAWRDKWGSTR 293	
XP_009064976.1	FYKSGNTTD	ELCW	ALWMYMITNETK	LTEAEKLFQ GP - AW GLSWDDTTPLNQVLLYKLTKKDIYKAVIDTMADWMPGGSVTPKGLAWRSQWGSLR 29	
XP_009064980.1	YYESSAYED	ELCV	AAWLYKATGNTF	LTEAKLYHS SATPN RLDWDNSLPSCQLMLYELTKNDLYKEIQKFLTVWKPGGN - TPCGLGIFDRWGTAR 28	
XP_009065702.1	FYRSSGYKD	LCE	AMWLYKATKDHO	LNDAKS - FN AW - AW ALSWDDKK I ACQMLLYE ATKDNQYKE I EGYFKGWKPGGG - TPCGLAWRDOWGSTR 29	
XP_011057333.1 XP_011164225.1	YYESTDYGD	L A W	AIWLEKATODIR	EDDATHHYQHFHLERPN - AFFYNKKVAGIQVLLAELIGQSEYQATKAFCDFSV - RQKIPKGLLFIDKSGILS 20	
XP_011253653.1	YYNNTDYGD	LAW	AIWLHKAEPHED	INDAENHYODFHLKRPN-EFFYNNKIAGVOILLAOLTGKLEYKAVIDFCNFSV-HNOTTPKGLLYIDKLGTLS 28	
XP_011253654.1	FYNNTDYGD	ELAW	AIWLHKAEPHFE	LNDAENHYQNFHLRPN EFSYNNKIAGVQVLLAQLTRKLKYKAIIDFCNFSM - HYQTTPKGLLYIDKLGTLS 29	
XP_011253655.1	YYNNTDYGD	ELAW	AIWLHKAEPHFD	LNDAENHYQDFHLKRPN - EFFYNNKIAGVQILLAQLTGKLEYKAV IDFCNFSV - HNQTTPKGLLYIDKLGTLS 29	
XP_011296914.1	FYESTDYGD	ELAW	AAWLFKATNDAK	LEDAEHHYQHFHLERPN - EFFYNKKVAGVQVLLAQLTGQSEYEAARAFCDFSSK - QQKTPKGLVYIDKFGTLC 28	
XP_011296916.1	FYESTDYGD	ELAW	AAWLFKATNDAK	LEDAEHHYQHFHLERPN - EFFYNKKVAGVQVLLAQLTGQSEYEAARAFCDFSSK - QQKTPKGLVY I DKFGTLC 29	
XP_011346866.1	FYESTGYGD	ELAW	AAWL FNATAEPK	LYDAIHHYQDFHLERPT - RFFYNNKVAGVQVLLAOLTGQSEYQAVRAFCDFSV - ROOTTPKGLLYIDKSGTLS 29	
XP_011421762.1 XP_011423159.1	YYRSDKYED	ELTW	AAWI FRVTNDTK	INVAFOYYI - GP - DW - GOSWDDKIAGCOVLITEEIGKDKIKDIEA IFODWNPGGSVSPKGLAFRSOWGSLR 20	
XP 011430172.1	YYRSNSYKD	LCT	AMELYKLTNDAS	LODA ISFYN - SK - SW - AFSWDDKN VOCALMLYRAOGSRHYAOVDDFIRSYMPGGE - TPCGLAWRLGVGPLR 28	
XP_011447501.1	FYSSSEYRD	MCV	AALLYLATNKTG	R S D A I T F Y N S Y S P G R W A Y D W D D K T M L C D V L L Y E I T Q E T K Y K A V E N F V T S Y M P G G S - T P C G L A W R L E W G S L R 29	
XP_011447502.1	FYSSSEYRD	MCV	AALLYLATNKTG	R S D A I T F Y N S Y G R W A Y D W D D K T M L C D I L L Y E I T Q E T K Y K A V E N F V T S Y M P G G S - T P C G L A W R L E W G S L R 29	
XP_011498432.1	YYESTDYGD	ELAW	AAWLFKASNETK	L D E A E H H Y Q H F H L E R P N - E F F Y N K K V A G V Q V L L A Q L T G Q Q E Y R A K A F C N F S V - R Q Q K T P K G L L Y I D K F G T L C 20	
XP_011632703.1	YYESTDYGD	ELAW	AIWLFKATEDTK	FDDAVHHYQHFHLERPN - AFFYNKKVAGIQLLLAELTELSEYQATRAFCDFSML - QQKTPKGLLFIDKSGTLS 29	
XP_011700378.1	YYESTDYGD	ELAW	A I WL F K A T G D M K	FDDAVHHYQHFHLERPN - AFFYNKKVAGIQVLLAELTGQPEYLATRAFCDFSVR - QQKTPKGLLFIDKSGTLS 29	
XP_011667251.1	TTESTDIGD	EL TW	AIWLEKATODIK	EDDAVIHIYOHFHLERPN AFFTNKKVAGIOVLLAELIGUSEYOATBAEODESVR-UUKIPKGLLFIDKSGILS 20	
XP_012061951.1 XP_012146513.1	YYESTDYGD	LAW	AAWLEKATNETV	I FDAKHHYQOFHI FRYN - FFFYNKKVAGTQVLLADI TGOPFYQAARAFCDFSV - HEOKTSKEL Y DKSGT C 20	
XP 012224814.1	YYESTDYGD	ELVW	AAWLFKATEDPK	L DAAERHYQOFHLKRPN - EFFYNKKLAGIQVLLAOLTGQSEYQATRTFCNFSV - REOKTPKGLLY I DKFGTLC 29	
XP_012248155.1	YYESTDYGD	ELAW	AVWLFKATNDTI	LEDAEHHYQHFHLERPN - EFFYNKKVAGVQVLLAQMTGQPEYQAARAFCDFSV - RQQKTPKGLLY I DKFGTLC 29	
XP_012273697.1	YYESTDYGD	ELAW	AAWLYKATNDSR	LEDAEHHYQHFHLERPN - EFFYNKKVAGVQVLLAQLTGQPEYQAAQAFCNFSVR - HQKTPKGLLYIDKFGTLC 29	
XP_012526654.1	YYESTDYGD	ELAW	AIWLFKATGDVR	FEDAVHHYQHFHLERPN - AFFYNKKVAGIQVLLAELTGLSEYQATRAFCDFSVR - QQKTLKGLLFIDKTGTLG 20	
XP_012935745.1	FYTSNNYTD	ELAW	SLWLYQATNDHK	LAEAERQFD PEPAW GMS WDEKT I GNQLLLYKLTHKEKYKAV I DTYKSWFPGGD I TPKGLAFRLEWASLR 29	
XP_012938497.1	FYSSGSDED	ENCL	ALWLYKAIGDSH AVWLYRATGTOSK	LNDAKSPHNGMPSWAQSWDDNKLACRILLLEAIHDNKLKEIQNDLKPWYEKIPCGLAWKLEWAPIR 20	
XP_012939933.1	YYH-DDYKD	MCL	AVWLYRATKDSK	INDAKSFYE - GV-PW- ALSWERKAACO	
XP_012941225.1 XP_012045520.1	FYKSWSYQD	EL CW	AAWLLKAIGDEA SIWIYOATNESKI	NVKARAAMA SMQNP - TEYSWDSKSTGCYVLMHELSEKSEYRKTENFMSNWLGFKNGTPCGLAWLREWGALR 28	
XP_012945520.1 XP_012066277.1	SYPEVEVID	ELCV	AVWI VRATGTAON	LIDAAEKLWD PEPAW - OMSWDEKIISNOVLLIKLIGKUSIKAVEGITISWP960DIIPNGLAFKLWASILK 28	
XP_013066278.1	FYNPSSDKD	LCE	GMWLYKATNTAO	LTDARTFAE - RYAAW - AYSWEDKQLACHQLLYEATRENQYRLYYTYFNSWLPGGS - TPCGLAWRHLWGANG 27	
XP_013066280.1	FYSSSGDRD	LCE	SIWLYKATRLSK	LNDARSFVD T G T AW A LS WDD KRAACQLLLYE E THEN QYRAV VN F LQS WQ PGGG - T P C GLAWRD KWGAN R 27.	
XP 013066281 1	LYDSNSEKD	MCE	SIWIHKATOKSN	I DDAKTYLD NYPPW SEDWENKOAACROLLYEETOSSYYKAYVEYLTTWI PGGT. TPCGLAWRSKWGSLR. 27	
			or we have a down	ebbaktieb - with w- bibwelkkaadkaeeteetabott karterettiinetoot - ti boekakokaboek b	
XP_013066283.1	LYDSYSEKD	ЕМСЕ	SLFLHKATQKSN	L DDAKTYLE NY PPW SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27	
XP_013066283.1 XP_013066284.1	L Y D S Y S E K D F Y D S T S E S D	EMCE	S L F L H K A T Q K S N A MWL Y R A T N D S T	LDDAKTYLE NYPPWSEDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT-TPCGLAWRIKWGSLR 27. LDYARTLA-DFNV-PWSSVDWDNKQAACCLLLYEETHEDKYQALVNYFDSWLPGGT-TPCGLAWRFKWGSLR 27.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1	L Y D S Y S E K D F Y D S T S E S D C Y T S D D D R D	E M C E E M C N E M C E	SLFLHKATQKSN AMWLYRATNDST SIWLYRATQNAQ	LDDAKTYLE NY PPW SEDWEN KQAACRQLLYE ETQSSVYKAVVEYLTTWLPGGT - TPCGLAWR I KWGSLR 27. LDYARTLA - DFNV - PW SYDWDN KQASCLLLYE ETAESVYKAVVEYLTTWLPGGT - TPCGLAWR FKWGSLR 27. LADARSFV DLLAW DLSWSSKKPSCQLYLYE ETQSDQYKAIQNFFNWLPGGT - TPCGLAWRSKWGSNR 27.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1 XP_013066286.1	L Y D S Y S E K D F Y D S T S E S D C Y T S D D D R D F Y D S D D D R D E Y S S S A D D D	E M C E E M C N E M C E E L C E	SLFLHKATQKSN AMWLYRATNDST SIWLYRATQNAQ SIWLYRATQNGQ AAWIYRATQNGQ	LDDAKTYLE - NY PPW - SFDWENKQAAGRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPGGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SFDWENKQAAGRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPGGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQASCLLLLYEETQSDQYKAIQNFFNNWLPGGT - TPGGLAWRSKWGSNR 27. LDAARSFV DLLAW - DLSWSKKPSCOLYLFEETQSDQYKAIQNFFNNWLPGGT - TPGGLAWRSKWGSNR 27. LEDAKTFV DLLAW - DLSWLSKKPSCOLYLFEETQSNQYKAIQNYFNSWLPGGT - TPGGLAWRSKWGSNR 27.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066286.1 XP_013066370.1 XP_013066371.1	L Y D S Y S E K D F Y D S T S E S D C Y T S D D D R D F Y D S D D D R D F Y S S S A D Q D Y Y S S R D E K D	E M C E E M C N E M C E E L C E E M C V E L C M	SLFLHKATQKSN AMWLYRATNDST SIWLYRATQNAQ SIWLYRATQNGQ AAWLYRATRNNQ TLWLYRATHNTQ	LDDAKTYLE NYPPW SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW SYDWDNKQASCLLLYEETHEDKYQALVNYFDSWLPGGT - TPCGLAWRFKWGSLR 27. LADARSFV DLLAW DLSWSKKPSCQLYLYEETQSDQYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCQLYLFEETQSNQYKAIQNYFNSWLPGGT - TPCGLAWRSKWGSNR 27. LNDANSLVD AW - AALSWDDKKIACQLILYEETQKSTLKAVVGFMQAWLPGGG - TPCGLAWRDAWQANR 27. LNDANSLVD W - AW ALGWDDKKIACQLILYEETDQKSTLKAVVGFMQAWLPGGG - TPCGLAWRDAWQANR 27.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066371.1 XP_013069169.1	L Y D S Y S E KD F Y D S T S E S D C Y T S D D D D D F Y D S D D D D D F Y S S S A D Q D Y Y S S R D E KD Y Y R S Y N I T D	E M C E / E M C N E M C E / E L C E / E M C V / E L C M	SLFLHKATQKSN AMWLYRATNDST SIWLYRATQNQQ AAWLYRATQNQQ AAWLYRATRNNQ TLWLYRATHNTO SIWLYQATNENK	LDDAKTYLE NYPPW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQASCLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRFKWGSLR 27. LADARSFV DLLAW - DLSWSSKKPSCQLYLYEETQSDQYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - DLSWLSKKPSCQLYLFEETQSNQYKAIQNYFNSWLPGGT - TPCGLAWRSKWGSNR 27. LNDANSLVD - AW - AW - ALGWDDKKIACQLLYEETQSNQYKAIQNYFNSWLPGGT - TPCGLAWRSKWGSNR 27. LHDAYD - YD LW - AW - ALGWDDKKIACQLLYEETQSNYKAVGFMQAWL - DGINTPCGLAWRDAWGANR 27. LHDAYD - YD LW - AW - ALSWDDKYVACQLILYEETHDDRFSAVYGFMQAWL - DGINTPCGLAWRDAWRGNR 27. LALAEKYFD PAPAW - GMSWDDKTIVNQYMLYKITKDEKYKAVEDSFKFWFPGGGITPKGLAYRLOWGSLR 28.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066286.1 XP_013066370.1 XP_013066370.1 XP_013066371.1 XP_013069169.1	L Y D S Y S E K D F Y D S T S E S D C Y T S D D D R D F Y D S D D D R D F Y S S A D Q D Y Y S S R D E K D Y Y R S Y N I T D Y Y R S Y N I T D	E M C E / E M C N ( E M C E / E L C E / E L C M ( E L A W ( E L A W (	SLFLHKATQKSN AMWLYRATNDST SIWLYRATQNAQ AWLYRATQNAQ AAWLYRATRNNQ TLWLYRATHNTQ SIWLYQATNENK SIWLYQATNENK	LDDAKTYLE NY PPW SFDWENKQAAGRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWR IKWGSLR 27. LDYARTLA - DFNV - PW SY DWDNKQASCLLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWR IKWGSLR 27. LDARSFY DLLAW DLSWSKKPSCOLYLYEETQSDQYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCOLYLFEETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAARSFV DLLAW ALSWDDKKIKOLILYEETQSNQYKAIQNYFNSWLPGGT - TPCGLAWRSKWGSNR 27. LDAAKTFV DLLAW ALSWDDKKIKOLILYEETDSNYKAIQNYFNSWLPGGT - TPCGLAWRDMGANR 27. LHDAYD - YD LW - AW ALSWDDKKIKOLILYEETHDDRFSAVYGFMQAWL - DG INTPCGLAWRDMWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGG ITPKGLAYRLOWGSLR 28.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066286.1 XP_013066371.1 XP_013069371.1 XP_013069170.1 XP_013069170.1	L Y D S Y S E K D F Y D S T S E S D C Y T S D D D R D F Y D S D D D R D F Y S S A D Q D Y Y S S R D E K D Y Y S S N I T D Y Y R S Y N I T D Y Y N S N N I T D	E M C E E M C N E M C E E L C E E L C C E L C M E L A W E L A W E L C W	SLFLHKATQKSN AMWLYRATQNAO SIWLYRATQNAO SIWLYRATQNAO AAWLYRATRNNO TLWLYRATRNNO SIWLYQATNENK SIWLYQATNENK SIWLYQATNENK	LDDAKTYLE NY PFW SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - FW SY DWDNKQASCLLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDARSFV DLLAW DLSWSKKPSCOLYLYEETQSDQYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCOLYLFEETQSNQYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LNDANSLVD AW - AW ALGWDDKKIACQLLYEETDKSTLKAVVGFMQSWLPGGG - TPCGLAWROKMRGANR 27. LNDANSLVD AW - AW ALGWDDKKIACQLLYEETDRSTVVGFMQAWL - DGINTPCGLAWROKMGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNOVMLYK ITKDEKYKAVEDSFKFWFFQGGITPKGLAYRLOWGSLR 29. LALAEKYFD PAPAW GMSWDDKTIVNOVMLYK ITKDEKYKAVEDSSFKFWFFQGGITPKGLAYRLOWGSLR 29. LALAEKYLD DE PAW GMSWDDKTIANQVLLYKLTKKDRYKAVEDSSFKFWFFQGSITFKGLAYRLOWGSLR 29.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066286.1 XP_013066371.1 XP_013069371.1 XP_013069170.1 XP_013069152.1 XP_013069521.1 XP_0130695252.1	L Y D S Y S E KD FY D S T S E S D C Y T S D D D RD FY D S D D D RD FY S S S A D Q D Y Y S S R D E KD Y Y R S Y N I T D Y Y R S Y N I T D Y Y N S N N I T D Y Y N S N N I T D	E M C E E M C E E L C E E L C E E L C W E L C W E L A W E L A W E L C W E L C W	SLFLHKATQKSN AMWLYRATQNGO SIWLYRATQNGO AAWLYRATQNGO TLWLYRATRNO SIWLYQATNENK SIWLYQATNENK SLWLYQATNDPK	LDDAKTYLE NY PFW SFDWEN KQAACRQLLYE ETQSSVYKAV VEYLTTWLPGGT - TPCGLAWR I KWGSLR 27. LDYARTLA - DFNV - PW SY DWDN KQAACLLLLYE ETHEDK YQAL VN YFDSWLPGGT - TPCGLAWR I KWGSLR 27. LADARSFV DLLAW DLSWSKKPSCQLYLYE ETQSDQYKAI QN FFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCQLYLFE ETQSNQYKAI QN YFNSWLPGGT - TPCGLAWRSKWGSNR 27. LNDANSLVD AW - AW - ALGWDDKKI ACQL ILYE ETQKSTLKAV VG FMGAWLPGGG - TPCGLAWRDAWGANR 27. LNDANSLVD WM - ALSWDDKXI ACQL ILYE ETDRFSAV VG FMGAWLPGGG - TPCGLAWRDAWGANR 27. LALAEKYFD PAPAW GMSWDDKTI KOLILYE THDRFSAV VG FMGAWL - DG INTPCGLAWRDAWGANR 27. LALAEKYFD PAPAW GMSWDDKTI VNQVMLYK I TKDEKYKAVED SFKFWFPGGG I TPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW GMSWDDKTI NOVLLYK LTKKDRYKAVE GSYTMWFPGGS I TPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS I TPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS I TPKGLAYRLQWGSLR 29.	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066371.1 XP_013069169.1 XP_013069170.1 XP_013069522.1 XP_013069522.1 XP_013069522.1	L Y D S Y S E KD FY D S T S E S D FY D S D D D R D FY D S D D D R D Y Y S S R D E KD Y Y S S N I T D Y Y S S N I T D Y Y N S N N I T D Y Y N S N N I T D Y Y N S N N I T D FY P S N N F T D	E M C E / E M C N C E L C E / E M C V E L C W E L A W E L A W E L C W E L C W E L C W	SLFLHKATQKSN AMWLYRATQNGO SIWLYRATQNGO AWLYRATRNNGO TLWLYRATRNNG SIWLYQATNENK SIWLYQATNENK SLWLYQATNERK SLWLYQATNERK SIWLYQATNERK SIWLYQATNERK	LDDAKTYLE - NYPFW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - FW - SYDWDNKQASCLLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - DLSWLSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAAKTFV DLLAW - ALSWDDKKIACQLIYEETDSDQYKAIQNYFNSUPGG - TPCGLAWRSKWGSNR 27. LDAAYD - YD LW - AW ALSWDDKKIACQLIYEETHDDRFSAVVGFMQSWLPGG - TPCGLAWRDAWGANR 27. LALAEKYFD PAPAW - GMSWDDKTIVNQVMLYKITKDEKYKAVEDSFKFWFPGGGITPKGLAYRLQWGSLR 28. LALAEKYFD PAPAW - GMSWDDKTIVNQVMLYKITKDEKYKAVEDSFKFWFPGGSITPKGLAYRLQWGSLR 28. LADAEKYFD DEPAW - GMSWDDKTINNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW - GMSWDDKTINNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LADAEKYLD - DEPAW - GMSWDDKTINNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAEKYLD - DLPAW - GMSWDDKTINNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAEKYLD - DLPAW - GMSWDDKTINNQVLLYNITKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAEKYLD - DLPAW - GMSWDDKTINNQVLLYNITKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28.	
XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013069170.1 XP_013069521.1 XP_013069521.1 XP_013069521.1 XP_0130694281.1	L Y D S Y S E KD FY D S T S E S D C Y T S D D D R D FY D S D D D R D FY D S D D D R D Y Y S S R D E KD YY R S Y N I T D YY R S Y N I T D YY N S N N I T D FY P S N N F T D N Y S S H G F N D	E M C E / E M C N C E M C E / E L C E / E L C W E L A W E L A W E L A W E L C W	SLFLHKATQKSN AMWLYRATQNGO SIWLYRATQNGO AWLYRATRNNO TLWLYRATRNNO SIWLYQATNENK SIWLYQATNENK SLWLYQATNDPK SLWLYQATNDPK SIWLYQATNENK SIWLYQATNENK SLWLYQATNENK SLWLYQATNENT SLWLYQATNENT	LDDAKTYLE NY PFW SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - FW SYDWDNKQAACQLLYE ETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDARSFV DLLAW DLSWSKKPSCOLYLYEETQSDQYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDARSTV DLLAW DLSWLSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LNDANSLVD AW - AW ALSWDDKKIACQLILYEETHDRFSAVVGFMQSWLPGGG - TPCGLAWRSKWGSNR 27. LNDANSLVD AW ALSWDDKKIACQLILYEETHDRFSAVVGFMQAWL - DGINTPCGLAWRDMWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQVMLYKITKDEKYKAVEDSFKFWFPGGSITPKGLAYRLQWGSLR 28. LALAEKYFD PAPAW GMSWDDKTIVNQVMLYKITKDEKYKAVEDSFKFWFPGGSITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW GMSWDDKTI ANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAEKYLD DEPAW GMSWDDKTI NNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAEKYLD DEPAW GMSWDDKTI NNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHYD DEPAW GMSWDDKTI NNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NNQVLLYKLTKKDRYKAVE MSYTWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NNQVLLYKLTKKDRYKAVE MSYTWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NNQVLLYKLTKKDRYKAVE MSTWWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NNQVLLYKLTKDKYKAVE MSTWWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NNQVLLYKLTKDCHYKAVE NTFQUWFFGGSITPKGLAHRSDRAALS 28.	
XP_013066283.1 XP_013066284.1 XP_013066286.1 XP_013066286.1 XP_013066370.1 XP_013066370.1 XP_013066160.1 XP_01306651.1 XP_01306652.1 XP_013066523.1 XP_013066223.1 XP_013064282.1	L Y D SY SE K D F Y D ST SE S D C Y T S D D R D F Y D S D D R D F Y S S R D E K D Y Y S S R D E K D Y Y R SY N I T D Y Y R SY N I T D Y Y R S N N I T D Y Y N S N N I T D Y Y S S H G F N D N Y S S H G F N D	E M C E / E M C N C E / E L C E / E L C W C V / E L A W C E L A W C E L A W C E L C W C W C W C W C W C W C W C W C W C	SLFLHKATOKSN AMWLYRATONST SIWLYRATONAO SIWLYRATONAO AWLYRATRNNO TLWLYRATRNNO SIWLYQATNENK SLWLYQATNENK SLWLYQATNDPK SIWLYQATKEAK SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT	L DDAKTYLE NY PPW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. L DYARTLA - DFNV - PW - SY DWDNKQAACQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. L ADARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDYKAIQNYFNNWPGGT - TPCGLAWRSKWGSNR 27. L EDAKTFV DLLAW - DLSWLSKKPSCOLYLYEETQSDYKAIQNYFNNWPGGT - TPCGLAWRSKWGSNR 27. L DANSVD - AW - ALGWDDKKIACQLIYEETQSDYKAIQNYFNSUPGGT - TPCGLAWRSKWGSNR 27. L HDAYD - YD LW - AW ALGWDDKKIACQLIYEETHDDRFSAVVGFMQSWLPGGT - TPCGLAWRDAWGANR 27. L HDAYD - YD LW - AW ALGWDDKKIACQLIYEETHDDRFSAVVGFMQSWLPGGT - TPCGLAWRDAWGANR 27. L HDAYD - YD LW - AW ALGWDDKKIACQLIYEETHDDRFSAVVGFMQSWLPGGT - TPCGLAWRDAWGANR 27. L ALAEKYFD PAPAW GMSWDDKTI VNOVMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLQWGSLR 29. L ALAEKYFD PAPAW GMSWDDKTI NOVLLYK ITKDEKYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L ADAEKYLD DEPAW GMSWDDKTI ANQVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L TAAEKYLD DEPAW GMSWDDKTI ANQVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L TAAECHFD PLPAA DMSWDNKTAANQVLLYKLTKKDKYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L TAAECHFD PLPAA DMSWDNKTAANQVLLYKLTKKDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. L TAAECHFD PLPAA DMSWDNKTAANQVLLYKLTKKDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. L TAAECHFD PLPAA DMSWDNKTAANQVLLYKLTKKDKYKAVE DTFKYWFPGGS ITPKGLAHRSDRAALS 20. L TAAECHFD PLPAP DMSWDNKTAANQVLLYKLTKEN KWA PLTFNYWFPGGS ITPKGLAHRSDRAALS 20. L TAAECHFD PLPAP DMSWDNKTAANQVLLYKLTKEN KVAVEDTFKYWFPGGSVTPKGLAHRSDRAALS 20. L TAAECHFD PLPAP DMSWDNKTAANQVLLYKLTKEN VKAVEDTFKYWFPGGSVTPKGLAHRSDRAALS 20.	
XP_013066283.1 XP_013066284.1 XP_013066284.1 XP_013066286.1 XP_013066370.1 XP_013066371.1 XP_013066371.1 XP_013066521.1 XP_013066521.1 XP_013066522.1 XP_01306422.1 XP_01304282.1 XP_01308428.1	L Y D S Y S E KD FY D S T S E S D C Y T S D D D R D FY D S D D D R D FY D S D D D R D Y S S A D Q D Y Y S S N F T D Y Y S S N I T D Y Y N S N I T D Y Y N S N I T D FY P S N F T D FY P S N F T D N Y S S H G F N D N Y S S H G F N D N Y S S H G F N D	E M C E / E M C N C E / E L C E / E L C E / E L C W C E L C W E I C W E I C W E I C W	SLFLHKATQKSN AMWLYRATQNGO SIWLYRATQNGO AWLYRATRNNGO TLWLYRATRNNG SIWLYQATNENK SIWLYQATNENK SLWLYQATNDFK SLWLYQATNDFK SLWLYQATNDFK SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT	LDDAKTYLE NY PFW SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT . TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW SY DWDNKQAACLLLLYEETQSSVYKAVVEYLTTWLPGGT . TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW SY DWDNKQASCLLLLYEETQSDQYKAIQNYFNNWLPGGT . TPCGLAWRSKWGSNR 27. LDAARSFV DLLAW DLSWLSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT . TPCGLAWRSKWGSNR 27. LDAAKTFV DLLAW DLSWDDKKIACOLILYEETQSDQYKAIQNYFNNWLPGGT . TPCGLAWRSKWGSNR 27. LDAAYD - YD LW - AW ALSWDDKKIACOLILYEETASTKAVVGFMQSWLPGG - TPCGLAWRDAWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGGITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGSITPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTIVNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTIVNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTIVNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTIANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTAANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDNKTAANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDNKTAANQVLLYKLTKKDRYKAVE GSYTMWFPGGSITPKGLAHRSDRAALS 29. LTAAECHFD PLPAP DMSWDNKTAANQVLLYKLTKKDRYKAVE OTFKYWFPGGSITPKGLAHRSDRAALS 29. LTAAECHFD PLPAP DMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGSITPKGLAHFFNRGALS 29. LTAAECHFD PLPAP DMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGSVTPKGLAHPFNRGALS 29. LTAAECHFD PLPAP DMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGSVTPKGLAHPFNRGALS 29.	
XP_013066283.1 XP_013066284.1 XP_013066286.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066160.1 XP_013066152.1 XP_013066521.1 XP_013066521.1 XP_013066521.1 XP_01306622.1 XP_01306625.1 XP_01306625.1 XP_01306625.1 XP_01306625.1 XP_01306625.1 XP_01306625.1 XP_01306625.1 XP_01306625.1 XP_01306625.1 XP_01306657.1 XP_0130657.1 XP_013057.1 XP_013057.1 XP_013057.1 XP_013057.1 XP_013057.1 XP_013057.1XP_0157.1 XP_0157.1 XP_0157.1XP_0157.1 XP_0157.1 XP_0157.1XP_0157.1 XP_0157.1 XP_0157.1XP_0157.1 XP_0157.1 XP_0157.1XP_0157.1 XP_0157.1 XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.11XP_0157.	L Y D S Y S E K D F Y D S T S E S D F Y D S T D S D D R D F Y D S D D R D Y S S A D C D Y Y S S R D E K D Y Y S S N I T D Y Y S S N I T D Y Y S S N F T D N Y S S H G F N D N Y S S H G F N D N Y S S H G F N D F Y R S N N Y D F Y R S N N Y D	E M C E A E M C N C E E L C E E L C E E L C M E L A W E L A W E L A W E L A W E L C W E L C W E L C W E I C W	SLFLHKATQKSN AMWLYRATQNGO SIWLYRATQNGO AAWLYRATRNNO SIWLYRATRNNO SIWLYQATNENK SIWLYQATNENK SLWLYQATNENK SLWLYQATNDPK SIWLYQATNDPK SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT	LDDAKTYLE NYPRW SEDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW SYDWDNKQASCLLLLYEETQSSVYKAVVEYLTTWLPGGT TPCGLAWRIKWGSLR 27. LDAARSYV DLLAW DLSWSKKPSCQLYLYEETQSDQYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCQLYLFEETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LNDANSLVD AW ALSWDDKKIACQLILYEETHDDRFSAVVGFMQSWLPGGG - TPCGLAWRDWWGANR 27. LNDANSLVD AW ALSWDDKKIACQLILYEETHDDRFSAVVGFMQSWLPGGG - TPCGLAWRDWWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGG ITPKGLAYRLQWGSLR 28. LADAEKYFD PAPAW GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PAPAM GMSWDDKTI NOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA DMSWDDKTI NOVLLYK LTKKDRYKAVEG SYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NOVLLYK LTKKDRYKAVEG SYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NOVLLYK LTKKDRYKAVEG SYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAA GMSWDDKTI NOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKKDRYKAVE NTFKYWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS STPKGLAHFSDRAALS 28. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS STPKGLAHPFNRGALS 28. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS STPKGLAHPFNRGALS 28. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS STPKGLAHPFNRGALS 28. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066371.1 XP_013066371.1 XP_013066170.1 XP_013069170.1 XP_013069523.1 XP_013069523.1 XP_013069523.1 XP_013069282.1 XP_013084282.1 XP_013084282.1 XP_013084282.1	L Y D S Y S E KD FY D S T S E SD FY D S T S E SD FY D S D D RD FY D S D D RD Y S S R D E KD Y S S R O E KD Y S S R N I T D Y S S N I T D Y S S N I T D Y S S N G F ND N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND Y S S H G F ND		S L F L H K A T Q K S N AMWL Y R A T Q N D T S I WL Y R A T Q N G Q AWL Y R A T Q N G Q AWL Y R A T R N N G S I WL Y Q A T N D K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D T K S I WL Y Q A T N D T T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D T S L WL Y Q A T N D T	LDDAKTYLE - NY PPW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQAACQLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRFKWGSLR 27. LDAARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - DLSWLSKKPSCOLYLYEETQSDYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAANSTV - DLLAW - ALGWDDKKIACQLLYEETQSDYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAANSTV - NU - AW - ALGWDDKKIACQLLYEETDSNYKAIQNYFNSWLPGGT - TPCGLAWRSKWGSNR 27. LHDAYD - YD LW - AW ALGWDDKKIACQLLYEETHDDRFSAVVGFMQSWLPGGT - TPCGLAWRDWGANR 27. LHAAKYFD PAPAW - GMSWDDKTI VNOVMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTI NOVMLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW - GMSWDDKTI NOVLLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLQWGSLR 29. LTAAEKYLD DLPAW - GMSWDDKTI NOVLLYK ITKDEKYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAA - DMSWDDKTI ANOVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAA - DMSWDDKTI ANOVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAA - DMSWDNKTAANOVLLYK LTKKDKYKAVE TTGYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAA - DMSWDNKTAANOVLLYK LTKKDKYKAVE DTFKYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS VTPKGLAHFSDRAALS 29. LTAAECHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS VTPKGLAHFSDRAALS 29. LTAAECHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS VTPKGLAHPFNRGALS 29. LTAAECHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LAEAEKYFD HP - AS EFSWDDKTYANOVLLYK LKKLDTYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LAEAEKYFD HP	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066522.1 XP_013066522.1 XP_013066522.1 XP_013066229.1 XP_013064283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1	L YD SYSEKD FYD SYSEKD FYD SYSERD FYD SYDD RD FYD SYDD RD YYSSR O EKD YYR SYN I TD YYR SYN I TD YYR SYN I TD YYR SYN I TD YYN SN I TD FYP SNN FTD NYSSH GF ND NYSSH GF ND NYSSH GF ND FYR SNN YTD FYR SNN YTD FYR SNN YTD FYR SNN YTD		SLFLHKATQKSN AMWLYRATQNGO ASULYRATQNGO SIWLYRATQNGO SIWLYRATRNNO TLWLYRATRNNO SIWLYQATNENK SIWLYQATNENK SLWLYQATNDFK SLWLYQATNDFK SLWLYQATNDFT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDT SLWLYQATNDT SLWLYQATNDT SLWLYQATNDT SLWLYQATNDT	LDDAKTYLE - NYPFW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - FW - SYDWDNKQASCLLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFY DLLAW - DLSWLSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAAKTYU ALSWDDKKIACOLILYEETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAAYD - YD LW - AW ALSWDDKKIACOLILYEETDSSDQYKAIQNYFNWLPGGT - TPCGLAWRSKWGSNR 27. LDAAKTYU NU - AW ALSWDDKKIACOLILYEETHDDRYSAV GFMGSWLPGG - TPCGLAWRDAWGANR 27. LALAEKYFD PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGGITPKGLAYRLQWGSLR 28. LALAEKYFD PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGSITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW - GMSWDDKTIVNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM - GMSWDDKTIVNQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM - GMSWDDKTIANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM - GMSWDDKTIANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM - GMSWDNKTAANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM - GMSWDNKTAANQVLLYKLTKKDRYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM - GMSWDNKTAANQVLLYKLTKDDKYKAVEDTFKYWFPGGSITPKGLAHRSDRAALS 28. LTAAECHFD PLPAM - DMSWDNKTAANQVLLYKLTKEDKYKAVEDTFKYWFPGGSITPKGLAHRSDRAALS 28. LTAAECHFD PLPAM - DMSWDNKTAANQVLLYKLTKEDKYKAVEDTFKYWFPGGSITPKGLAHPFNRGALS 28. LTAAECHFD PLPAM - DMSWDNKTAANQVLLYKLTKEDKYKAVEDTFKYWFPGGSYTPKGLAHPFNRGALS 28. LTAAECHFD PLPAM - DMSWDNKTAANQVLLYKLTKEDKYKAVEDTFKYWFPGGSYTPKGLAHPFNRGALS 28. LAAEAEKYFD HP - AS EFSWDDKTVANQVLLYKLRKLDTYKAVENTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LAAEAEKYFD HP - AS EFSWDDKTVANQVLLYKLRKLDTYKAVENTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD NLAS FSWDDKTVANQVLLYKLRKLDTYKAVENTFKYWFGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD NLAS - EFSWDDKTVANQVLLYKLRKLDTYKAVENTFKYWFGGSSTPKGLAYSMEWGSLR 29. LTAAEAEKYFD HP - AS EFSWDDKTVANQVLLYKLRKLDTYKAVENTFKY	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066522.1 XP_013066522.1 XP_013066222.1 XP_01306428.1 XP_01306428.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1	L YD SYSEKD FYDST SESD FYDST SESD FYDSDDRD FYDSDDRD YYSSADCD YYRSYN ITD YYRSYN ITD YYRSYN ITD YYRSYN ITD YYRSYN ITD YYSSHGFND NYSSHGFND NYSSHGFND FYSSHGFND FYSSHFFND FYSSTYN		SLFLHKATQKSN AMWLYRATQNAO SIWLYRATQNAO SIWLYRATQNAO TLWLYRATRNNO TLWLYRATHNTO SIWLYQATNENK SLWLYQATNENK SLWLYQATNDPK SIWLYQATNDPT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT	LDDAKTYLE NY PW SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW SY DWDNKQAACQLLYE ETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW SY DWDNKQASCLLLLYE ETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCOLYLYE ETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW ALSWDDKKIACQLILYE ETDSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LHDAYD - YD LW - AW ALSWDDKKIACQLILYE ETDSNYKAIQNYFNSUPGGT - TPCGLAWRDMWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGG ITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVEDSFKFWFPGGS ITPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI NOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTAANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKKDKYKAVE NTFGYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE NTFKYWFPGGS STPKGLAHRSDRAALS 29. LAAEAEKYFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS STPKGLAHPFNRGALS 29. LAEAEKYFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS STPKGLAHPFNRGALS 29. LAEAEKYFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LAEAEKYFD HP - AS EFSWDDKTVANOVLLYK LTKDDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LEAACHYFD HP - AS EFSWDDKTVANOVLLYK LTKDTWKAVENTFKYWFPGGS STPKGLAYSMEWGSLR 29. LEAACHYFD HP AL TLENKQLSCNQLLYE TTKNA KNENTFKYWFPGGS TPCGLAWMMWGANR 27. LEEAXTHYN FMFSLRPM EFYVNNKXGVULLYK LTKLOK	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_013066522.1 XP_013066422.1 XP_013064279.1 XP_013064225.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064438.1	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D F Y D S D D R D Y Y S S N I T D Y Y S S N I F D Y S S H G F ND N Y S S H G F ND N Y S S H G F ND Y S S H G F ND F Y S S A R D I D F Y S S A R D I D F Y E S T Y G D		S L F L H K A T Q K S N AMWL Y R A T Q N D S T S I WL Y R A T Q N G Q A WL Y R A T Q N G Q A WL Y R A T Q N G Q A WL Y R A T H N N G S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y R A T N N F O A L WL Y K A T N D T N A AWL F K A T N D T N	LDDAKTYLE - NYPPW - SFDWENKQAAGRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQAAGCLLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRFKWGSLR 27. LDAARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFY DLLAW - DLSWLSKKPSCOLYLYEETQSNYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAANSFYD NAPAW - ALGWDDKKIACOLILYEETHDDRFSAVVGFMGSWLPGG - TPCGLAWRSKWGSNR 27. LHDAYD - YD LW - AW ALGWDDKKIACOLILYEETHDDRFSAVVGFMGSWLPGG - TPCGLAWRDAWGANR 27. LHDAYD - YD LW - AW ALGWDDKKIACOLILYEETHDDRFSAVVGFMGSWLPGG - TPCGLAWRDAWGANR 27. LHDAYD - YD LW - AW ALGWDDKKIACOLILYEETHDDRFSAVVGFMGSWLPGG - TPCGLAWRDAWGANR 27. LALAEKYFD PAPAW GMSWDDKTI VNOYMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLOWGSLR 28. LALAEKYFD PAPAW GMSWDDKTI NOYMLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 28. LADAEKYLD DEPAW - GMSWDDKTI ANQVLLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 28. LTAAEGHFD PLPAA DMSWDDKTI ANQVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLOWGSLR 28. LTAAEGHFD PLPAA DMSWDDKTI ANQVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 28. LTAAEGHFD PLPAA DMSWDNKTAANQVLLYK LTKKDKYKAVE DTFKYWFPGGS ITPKGLAHRSDRAALS 28. LTAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKKDKYKAVE DTFKYWFPGGS VTPKGLAHRSDRAALS 28. LTAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS VTPKGLAHPFNRGALS 28. LTAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAHPFNRGALS 28. LTAAEGKYFD HP - AS EFSWDDKTVANQVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAHPFNRGALS 28. LTAAEGHFD PLPAP MSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAYSMEWGSLR 28. LTNAKPMAS - GY - IW - ALTENKQUSCQLLLYE DTGNS YKAV GFHGWVF	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_0130669170.1 XP_013066922.1 XP_013066922.1 XP_013066922.1 XP_013066922.1 XP_013064287.1 XP_013064285.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_01408485.1 XP_014786694.1	L Y D S Y S E KD FY D S T S E SD FY D S T S E SD FY D S D D RD FY D SS D D RD Y Y SS RD E KD Y Y SS RD E KD Y Y SS RD E KD Y Y SS NN T D Y Y SS NN T D Y Y SS NG F ND NY SS HG F ND NY SS HG F ND NY SS HG F ND Y SS HG F ND Y SS HG F ND FY RS NN Y TD - YG ST S Y KD FY SS A RD I D FY SS A RD I		S L F L H K A T Q K S N AMWL Y R A T Q N D T S I WL Y R A T Q N Q O S I WL Y R A T Q N Q O T L WL Y R A T H N T Q S I WL Y Q A T N E N K S I WL Y Q A T N E N K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D T T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D T S L WL Y Q A T N D T S L WL Y Q A T N D T S L WL Y Q A T N D T S L WL Y Q A T N D T S L WL Y A T N N K Q S W L Y R A T N N K Q A WL Y K A T N D T N A WL Y K A T D T N	LDDAKTYLE - NYPPW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQASCLLLLYETQSDYKAI QNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDYKAI QNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - OLSWDSKKPSCOLYLYEETQSDYKAI QNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAANSTVO - AW - ALGWDDKKI ACOLILYEETQSDYKAI QNYFNSUPGGT - TPCGLAWRSKWGSNR 27. LDAANSTVO - AW - ALGWDDKKI ACOLILYEETDSNYKAI QNYFNSUPGG - TPCGLAWRSKWGSNR 27. LDAASTVO - AW - ALSWDDKKI VOLILYEETHDDRFSAVVGFMQSWLPGGG - TPCGLAWRDWGANR 27. LALAEKYFD - PAPAW - GMSWDDKTI VNQVMLYK ITKDEKYKAVEDSFKFWFPGGG ITPKGLAYRLQWGSLR 28. LALAEKYFD - PAPAW - GMSWDDKTI VNQVMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLQWGSLR 28. LADAEKYLD - DEPAW - GMSWDDKTI NNQVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAACHFD - PLPAPAW - GMSWDDKTI NNQVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD - PLPAPA - GMSWDDKTI NNQVLLYKLTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD - PLPAPA - GMSWDDKTI ANOVLLYKLTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD - PLPAPA - DMSWDNKTAANOVLLYKLTKKDRYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD - PLPAP - DMSWDNKTAANOVLLYKLTKKDRYKAVE TTQYWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD - PLPAP - DMSWDNKTAANOVLLYKLTKEDKYKAVE DTFKYWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD - PLPAP - DMSWDNKTAANOVLLYKLTKEDKYKAVE DTFKYWFPGGS TPKGLAHPFNRGALS 28. LAEAEKYFD - HP - AS - EFSWDDKTVANOVLLYKLTKEDKYKAVE DTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LAEAEKYFD - HP - AS - EFSWDDKTVANOVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD - PLPAP - DMSWDNKTAANOVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LAEAEKYFD - HP - AS - EFSWDDKTVANOVLLYKLTKLDTYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD - PLPAP - DMSWDNKTAANOVLLYKLTKLDTYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD - HP - AS - EFSWDDKTVANOVLLYKLRKLDTYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD - HP - AS - EFSWDDKTVANOVLLYKLRKLDTYKAVE NTFKYWFPGGSTFNGLAYSMEWGSLR 29. LEAAKYFD - HP - AS - EFSWDDKTANOVLLYKLRKLDTYKAVENTFKYWFQGSSTPKGLAYSMEWGSLR 29.	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_013066522.1 XP_013066522.1 XP_013066252.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_01506631.1 XP_015126607.1	L YD SYSEKD FYD SYSEKD FYD SYSERD FYD SYDD RD FYD SYDD RD YYSS RO EKD YYR SYN I TD YYR SYN I TD YYSSHGF ND NYSSHGF ND NYSSHGF ND YYSSHGF ND FYR SN YTD - YG ST SYKD FYR SN YTD - YG ST SYKD FYE ST EYGD YYE ST DYGD FYE ST CYGD FYE ST CYGD		SLFLHKATQKSN AMWLYRATQNGO SIWLYRATQNGO AWLYRATQNGO TLWLYRATRNNO TLWLYRATRNNO SIWLYQATNENK SIWLYQATNENK SLWLYQATNENK SLWLYQATNDRT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYQATNDNT SLWLYATNDN SAWLYRATNDN ALWLYKATNDTN	LDDAKTYLE NY PFW SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - FW SYDWDNKQAACQLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - FW SYDWDNKQASCLLLLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCOLYLFEETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW ALSWDDKKIACQLILYEETGSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LHDAYD - YD LW - AW ALSWDDKKIACQLILYEETGSNQYKAIQNYFNWLPGGT - TPCGLAWRDMWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGG TTPKGLAWRLDWGSLR 29. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGG TTPKGLAYRLDWGSLR 29. LADAEKYLD DEPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLDWGSLR 29. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLDWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI NOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLDWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLDWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLDWGSLR 29. LTAAECHFD PLPAM GMSWDDKTAANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 20. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 20. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS YTPKGLAHRSDRAALS 20. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS YTPKGLAHPSNRGALS 20. LTAAECHFD PLPAM DFYNNNX 400VLLYK LTKDDKYKAVE TTKYWFPGGS YTPKGLAHPSNCALS 20. LTAAECHFD PLPAM DFYNNX 400VLLYK LTKDDKYKAV	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066571.1 XP_01306652.1 XP_01306652.1 XP_01306622.1 XP_01306622.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_014746664.1 XP_014746664.1 XP_01476664.1 XP_01516447.1 XP_01516447.1 XP_01516447.1 XP_01516447.1	L Y D S Y S E KD FY D S T S E SD FY D S T S E SD FY D S D D RD FY D S D D RD Y S S N D FD Y S S N F TD Y S SN I TD Y S SN I TD Y S SN I F D Y S SN I F D Y S SN G F DD Y S SN G F DD Y S SN G F DD Y S SN G F ST FY S SN Y TD F Y S SN Y TD F Y S SN Y TD F Y S ST S Y SD F Y E ST D Y GD F Y E ST D Y GD Y Y E ST D Y GD		S L F L H K A T Q K S N AMWL Y R A T Q N G O S I WL Y R A T Q N G O AWL Y R A T Q N G O S I WL Y Q A T Q N G O T L WL Y Q A T N D K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D F K S I WL Y Q A T N D T K S I WL Y Q A T N D T T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y Q A T N D N T S L WL Y A T N N K O A L WL Y K A T N D T N A WL Y K Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	LDDAKTYLE - NYPPW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQAACQLLYLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRRKWGSLR 27. LDAARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - DLSWLSKKPSCOLYLYEETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAANSFYD NAPAW - ALGWDDKKIACQLIYEETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LHDAYD - YD LW - AW ALGWDDKKIACQLIYEETDSSNYKAIQNYFNSWLPGGT - TPCGLAWRSKWGSNR 27. LHDAYD - YD LW - AW ALGWDDKKIACQLIYEETHDDRFSAVVGFMQSWLPGGT - TPCGLAWRDWGANR 27. LALAEKYFD PAPAW - GMSWDDKTI VNOYMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLOWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTI NOYMLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 29. LADAEKYLD DEPAW - GMSWDDKTI NOYLLYK LTKNDEKYKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAA - OMSWDDKTI ANQVLLYK LTKNDEKYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAA - OMSWDDKTI ANQVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAA - OMSWDDKTI ANQVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAA - OMSWDDKTAANQVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAA - OMSWDDKTAANQVLLYK LTKKDKYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKKDKYKAVE TTFWYFPGGS ITPKGLAHFSNRALS 20. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKKDKYKAVE DTFKYWFPGGS TPKGLAHFSNRALS 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS TPKGLAHFSNRALS 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - OMSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS TPCGLAWANNGNSLR 27. LEAAKYFD H	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066952.1 XP_013066952.1 XP_013066952.1 XP_013064279.1 XP_013064283.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_015126007.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01542553.1 XP_01552007.1 XP_01542553.1 XP_01542553.1 XP_01542553.1 XP_01542553.1 XP_01552007.1 XP_01542553.1 XP_01552007.1 XP_01542553.1 XP_01552007.1	L Y D S Y S E KD FY D S T S E SD FY D S T S E SD FY D S D D RD FY D S D D RD Y S S R D E KD Y Y S S N I TD Y S S I G F ND N Y S S I G F ND N Y S S I G F ND N Y S S I G F ND Y S S I G S N Y TD FY S S TO Y GD FY S S TO Y GD Y Y S S TO Y GD		S L F L H K A T Q K S N A M W L Y R A T Q K S N S I W L Y R A T Q N G Q S I W L Y R A T Q N G Q A W L Y R A T R N N Q T L W L Y R A T H N T Q S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y A T N N K Q A L W L Y K A T N D T N A A W L F K A T N D T R A A W L F K A T N D T R A A W L F K A T N D T M	LDDAKTYLE - NYPPW - SFDWENKQAACRQLLYETTOSSVYKAVVEYLTTWLPGGT - TPCGLAWR IKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQASCLLLLYETTOSD QYKAIQNFFNNWLPGGT - TPCGLAWR KWGSLR 27. LDAARSFY DLLAW - DLSWSKKPSCOLYLYETOSD QYKAIQNFFNNWLPGGT - TPCGLAWR SKWGSNR 27. LEDAKTFV DLLAW - OLSWDSKKPSCOLYLYETOSD QYKAIQNYFNNUPGGT - TPCGLAWR SKWGSNR 27. LDAANSFV - DLLAW - ALSWDDKXIACOLIYYETOSD QYKAIQNYFNNUPGG - TPCGLAWR SKWGSNR 27. LDAANSFV DLLAW - ALSWDDKXIACOLIYYE ETOSNYKAIQNYFNSUPGG - TPCGLAWR SKWGSNR 27. LDAANSFV NAW - ALSWDDKXIACOLIYYE ETOSNYKAIQNYFNSUPGG - TPCGLAWR SKWGSNR 27. LALAEKYFD PAPAW - GMSWDDKTI VNQVMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLGWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTI VNQVMLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLGWGSLR 29. LADAEKYJD DEPAW - GMSWDDKTI NOVNLYK LTKDRYKAVEG SYTMWFPGGS ITPKGLAYRLGWGSLR 29. LTAACHFD PLPAP - GMSWDDKTI NOVNLYK LTKDDKYKAVE GSYTMWFPGGS ITPKGLAYRLGWGSLR 29. LTAAECHFD PLPAM - GMSWDDKTI ANOVLLYK LTKDDKYKAVE GSYTMWFPGGS ITPKGLAYRLGWGSLR 29. LTAAECHFD PLPAM - GMSWDDKTI ANOVLLYK LTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAM - GMSWDDKTI ANOVLLYK LTKDDKYKAVE TTOYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYK LTKDDKYKAVE NTFKYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS STPKGLAHRSDRAALS 29. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS STPKGLAHRSDRAALS 29. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS STPKGLAHSDRAALS 29. LAEAEKYFD HP - AS - EFSWDDKTVANOVLLYK LTKEDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LEAAEKYFD HP - AS - EFSWDDKTVANOVLLYK LTKEDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LAEAEKYFD HP - AS - EFSWDDKTVANOVLLYK LTKEDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LEAAEKYFD HP - AS - EFSWDDKTVANOVLLYK LTKEDKYKAVE NTFKYWFPGGS STPKGLAYSMEWGSLR 29. LEAAELYFD GWWWSFSWDKE KE JGSMVLLYN LTKAKYKN FUN TYN NYFSWFPGGS TPCGLAWRMRWGSLR 29. LEAAE	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_013066522.1 XP_013066522.1 XP_013066522.1 XP_013066223.1 XP_013064287.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_01512607.1 XP_01512807.1 XP_01513822.1	L Y D S Y S E KO FY D S T S E SO C Y T S D D D R D FY D S T D D R D Y S S A D O D Y Y S S N E K D Y Y S S N I T D Y S S H G F N D N Y S S H G F N D N Y S S H G F N D N Y S S H G F N D Y S S H G F N D F Y R S N Y T D F Y R S N Y T D F Y S S N F T D F Y S S N F T D F Y S S N T D F Y S S T D Y G D Y Y E S T D Y G D		S L F L H K A T Q K S N A M W L Y R A T Q K S N S I W L Y R A T Q N G Q A W L Y R A T Q N G Q T L W L Y R A T N N G Q T L W L Y R A T N N T Q S I W L Y Q A T N E N K S L W L Y Q A T N E N K S L W L Y Q A T N D F K S L W L Y Q A T N D T K S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y A T T N T T S L W L Y A T T N T T S L W L Y A T T N T T A A W L F K A T N D T R A A W L F K A T N D T R A A W L F K A T N D T M A A W L F K A T N D T M	LDDAKTYLE NYPRW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW SYDWDNKQAACQLLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDAARSYV DLLAW DLSWSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCOLYLYEETQSDQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAARSYD AW ALGWDDKKIACQLILYEETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAAKTFV DLLAW ALSWDDKKIACQLILYEETGSNQYKAIQNYFNNWLPGGT - TPCGLAWRDAWGANR 27. LDAAKTYD AW ALSWDDKKIACQLILYETHDDR FSAVVGFMQSWLPGG - TPCGLAWRDAWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGG ITPKGLAYRLQWGSLR 28. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW GMSWDDKTI NOVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LADAEKYLD DEPAW GMSWDDKTI NOVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYKLTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 28. LTAAECHFD PLPAM GMSWDDKTAANQVLLYKLTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD PLPAM DMSWDNKTAANQLLYKLTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 28. LTAAECHFD PLPAM DMSWDNKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS YTPKGLAHPFNRGALS 28. LTAAECHFD PLPAM DMSWDNKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS YTPKGLAHPFNRGALS 28. LAAAEKYFD HP - AS EFSWDDKTVANOVLLYKLKKLDTYKAVE NTFKYWFPGGS YTPKGLAHPFNRGALS 28. LTAAECHFD PLPAM DMSWDNKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS TPKGLAHPFNRGALS 28. LTAAECHFD PLPAM DMSWDNKTAANQVLLYKLKKLDTYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 28. LTAAECHYFD HP AS EFSWDDKTVANOVLLYKLKKLDTYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 28. LTAAECHYFD HP AS EFSWDDKTVANOVLLYKLKKLDTYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 28. LTAAECHYFD HP AS EFSWDDKTVANOVLLYKLKLTOTOKSYKAV GFLGGSTPKGLAYSMEWGSLR 28. LEDAEHHYQHFHLERPN - EFFYNKXVAGVOVLLAQTGQSEYQAARAFCDFSVR QCKTPKGLLYID	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306652.1 XP_01306622.1 XP_01306427.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_015164427.6 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_01556207.1	L Y D S Y S E KD FY D S T S E S D FY D S T S E S D D R D FY D S D D R D FY D S D D R D Y Y S S N F D Y S S H G F D F Y S S A R D F Y E S T Y S D F Y E S T Y S D Y E S T D Y G Y E S T D Y G Y Y E S T D Y G Y E S T D Y G Y Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y G Y E S T D Y C T D Y E S T D Y C T D Y		S L F L H K A T Q K S N A M W L Y R A T Q K S N S I W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O S I W L Y Q A T N C N S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S I W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y A T N N K O A L W L Y K A T N D T N A A W L F K A T N D S R A W L F K A T N S S K A W L F K K T N D S R	LDDAKTYLE - NYPPW - SFDWENKQAAGRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQAAGCLLLYETQSDYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSLR 27. LDAARSFY DLLAW - DLSWSKKPSCOLYLYEETQSDYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDDAKTFV DLLAW - DLSWLSKKPSCOLYLYEETQSDYKAIQNYFNNUPGGT - TPCGLAWRSKWGSNR 27. LDDANSLYD - AW - ALGWDDKKIACOLILYETASTLKAVVGFMQSWLPGGT - TPCGLAWRSKWGSNR 27. LDAASTF7 PAPAW - GMSWDDKTIAQVMLYK ITKDEKYKAVEDSKFWFPGGGITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTI NOVMLYK ITKDEKYKAVEDSKFWFPGGSITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTI NOVMLYK ITKDEKYKAVEDSKFWFPGGSITPKGLAYRLQWGSLR 29. LALAEKYFD DEPAW - GMSWDDKTI NOVMLYK ITKDEKYKAVEDSKFWFPGGSITPKGLAYRLQWGSLR 29. LADAEKYLD - DEPAW - GMSWDDKTI NOVLLYK LTKNDKYKAVEDSKTWFPGGSITPKGLAYRLQWGSLR 29. LTAAEGHFD PLPAA - DMSWDDKTI ANOVLLYK LTKNDKYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAEGHFD PLPAA - DMSWDDKTI ANOVLLYK LTKKDKYKAVEGSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAEGHFD PLPAA - DMSWDDKTAANOVLLYK LTKKDKYKAVE GSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAEGHFD PLPAA - DMSWDDKTAANOVLLYK LTKKDKYKAVE GSYTMWFPGGSITPKGLAYRLQWGSLR 29. LTAAEGHFD PLPAA - DMSWDNKTAANOVLLYK LTKKDKYKAVE DTFKYWFPGGSITPKGLAHFSNRALS 20. LTAAEGHFD PLPAP - DMSWDNKTAANOVLLYK LTKKDKYKAVE DTFKYWFPGGSITPKGLAHFSNRALS 20. LTAAEGHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGSSTPKGLAHFSNRALS 20. LAEAEKYFD H - AS - EFSWDDKTVANOVLLYK LTKEDKYKAVE NTFGYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAEGHFD PLPAP - DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGSTPCGLAWAMMWGSLR 27. LEAAKYFD H - AS - EFSWDDKTVANQVLLYK LTKEDKYKAVE NTFKYWFPGGSTPCGLAWAMMWGSLR 27. LEAAKYFD H - AS - EFSWDDKTVANQVLLYK LTKEDKYKAVE NTFKYWFPGGSTPCGLAWAMMWGSLR 27. LEAAKYFD H - AS - EFSWDDKTVANQVLLYK LTKEDKYKAVE NTFKYWFPGGSTP	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066522.1 XP_013066522.1 XP_01306627.1 XP_013064279.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_015126007.1 XP_015126007.1 XP_015522007.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1	L Y D S Y S E KD FY D S T S E SD FY D S T S E SD FY D S D D RD FY D S D D RD YY S S R D E KD YY S S N I TD YY R S N I TD YY R S N I TD YY R S N I TD YY S S HG F ND NY S S HG F ND NY S S HG F ND NY S S HG F ND YS S HG F ND YS S HG F ND FY R S N Y TD FY R S N Y TD FY R S N Y TD FY S S R G ND FY S S TO Y CD FY S S TO Y CD YY E S TO Y CD		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N O T L W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y R A T N S T A W L F K A T N D T R A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M	LDDAKTYLE - NYPPW - SFDWENKQAACRQLLYEETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQASCLLLLYETQSD QYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAARSFY DLLAW - DLSWSKKPSCOLYLFETQSD QYKAIQNFFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - OLSWLSKKPSCOLYLFETQSNQYKAIQNYFNNUPGGT - TPCGLAWRSKWGSNR 27. LDAARSFV - DLLAW - ALSWDDKKIACOLILYETGSD QYKAIQNYFNSUPGGT - TPCGLAWRSKWGSNR 27. LDAANSTV - AW - ALSWDDKKIACOLILYETDSTAVGFMQSWLPGGG - TPCGLAWRSKWGSNR 27. LALAEKYFD PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAV EDSFKFWFPGGG ITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LADAEKYLD - DEPAW - GMSWDDKTIVNQVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAP - OMSWDDKTI NOVVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM - GMSWDDKTI NOVVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM - GMSWDDKTI ANOVLLYKLTKDDKYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYKLTKDDKYKAVE TTQYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYKLTKDDKYKAVE TTQYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYKLTKDDKYKAVE DTFKYWFPGGS STPKGLAHPFNRGALS 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYKLTKDDKYKAVE DTFKYWFPGGS STPKGLAHPSNRGALS 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYKLTKDDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYKLTKEDKYKAVE DTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LAEAEKYFD HP - AS - EFSWDDKTVANOVLLYKLTKLDTYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LTAAECHFD PLPAM - OMSWDNKTAANOVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. LEAAEKYFD HP - AS - EFSWDDKTVANOVLLYKLTKLDTYKAVE	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_01306652.2 XP_01306652.2 XP_01306652.2 XP_01306622.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_01512607.1 XP_01512607.1 XP_01512807.1 XP_01513822.1 XP_01551382.2	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D F Y D S D D R D F Y D S D D R D Y Y S S N T D Y Y S S N F T D F Y S S H G F N D N Y S S H G F N D N Y S S H G F N D N Y S S H G F N D Y S S H G F N D		S L F L HK A T Q K S N AMWL Y R A T Q N D T S I WL Y R A T Q N G Q AWL Y R A T Q N G Q AWL Y R A T R N N G Q T L WL Y R A T H N T Q S I WL Y Q A T N E N K S I WL Y Q A T N E N K S L WL Y Q A T N D N T S L WL Y Q A T N D T S L WL Y A T N T N T A A WL Y R A T N D T R A A WL F K A T N D T R A A WL F K A T N D T R A WWL F K A T N D T R A I WL Y K K T K D V K A V WL F K A T N D T M A V WL F K A T N D T M	LDDAKTYLE NYPRW - SFDWENKQAACRQLLYE ETQSSVYKAVVEYLTTWLPGGT - TPCGLAWR IKWGSLR 27. LDYARTLA - DFNV - PW SYDWDNKQASCLLLLYE ETQSSVYKAVVEYLTTWLPGGT - TPCGLAWR IKWGSLR 27. LDAARSYV DLLAW DLSWLSKKPSCOLYLF ETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW DLSWLSKKPSCOLYLF ETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAARSYD AW ALGWDDKKIACQLILYE ETDSNYKAIQNYFNNWLPGGT - TPCGLAWRDMWGANR 27. LHDAYD - YD LW - AW ALSWDDKKVACQLILYE ETDSNYKAIQNYFNSUPGGT - TPCGLAWRDMWGANR 27. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVED SFKFWPPGGG ITPKGLAYRLQWGSLR 29. LALAEKYFD PAPAW GMSWDDKTIVNQYMLYK ITKDEKYKAVED SFKFWPPGGS ITPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LADAEKYLD DEPAW GMSWDDKTI NOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI NOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. LTAAECHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAM DMSWDNKTAANOVLLYK LTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE OTFKYWFPGGS YTPKGLAHRSDRAALS 29. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS YTPKGLAHPSNRGALS 29. LTAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE DTFKYWFPGGS YTPKGLAHPSNCH 2000000000000000000000000000000000000	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066427.1 XP_013066427.1 XP_01306427.1 XP_013064282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_013044282.1 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_015164427.4 XP_01516427.4 XP_01556292.7 XP_01556292.7 XP_01556292.7 XP_01765629.1 XP_017785110.0 XP_017785110.4 XP_0177785110.4 XP_0177785110.4 XP_01777777777777777777777777777777777777	L Y D S Y S E KD F Y D S T S E S D F Y D S T S E S D F Y D S D D R D F Y D S D D R D F Y D S D D R D Y Y S S N I T D Y Y R S Y N I T D Y Y R S Y N I T D Y Y N S N I T D Y Y N S N I T D Y Y S S H G F ND N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND Y		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O A W L Y R A T R N N G O S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S I W L Y Q A T N D T T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T A L W L Y K A T N D T T A A W L F K A T N D S R A W L F K A T N D S R A W L F K A T N D T M A W L F K A T N D T M A W W F K A T N D T M A W W F K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M	L DDAKTYLE NY PPW SFDWENKQAAGRQLLYE ETGSSVYKAVVEYLTTWLPGGT . TPGGLAWRIKWGSLR 27. L DYARTLA - DFNV - PW SYDWDNKQAAGCLLLYE ETGSSVYKAVVEYLTTWLPGGT . TPGGLAWRRKWGSLR 27. L DAARSFY DLLAW DLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT . TPGGLAWRSKWGSNR 27. L EDAKTFV DLLAW DLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT . TPGGLAWRSKWGSNR 27. L DAARSFYLD APAW ALGWDDKKIACOLILYE ETGSNYKAIQNYFNNWLPGGT . TPGGLAWRSKWGSNR 27. L HDAYD - YD LW - AW ALGWDDKKIACOLILYE ETGSNYKAIQNYFNSUPGGG . TPGGLAWRDKWGANR 27. L HDAYD - YD LW - AW ALGWDDKKIACOLILYE ETHDDRFSAVVGFMGSWLPGG . TPGGLAWRDKWGANR 27. L ALAEKYFD PAPAW GMSWDDKTI VNOYMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLOWGSLR 29. L ALAEKYFD PAPAW GMSWDDKTI ANQVLLYK LTKNDEKYKAVED SFKFWFPGGS . TPKGLAYRLOWGSLR 29. L ADAEKYLD DEPAW GMSWDDKTI ANQVLLYK LTKNDEKYKAVE DSYTMWFPGGS . TPKGLAYRLOWGSLR 29. L TAAEGHFD PLPAA DMSWDDKTI ANQVLLYK LTKNDKYKAVE GSYTMWFPGGS . TPKGLAYRLOWGSLR 29. L TAAEGHFD PLPAA DMSWDDKTI ANQVLLYK LTKNDKYKAVE GSYTMWFPGGS . TPKGLAYRLOWGSLR 29. L TAAEGHFD PLPAA DMSWDDKTI ANQVLLYK LTKNDKYKAVE GSYTMWFPGGS . TPKGLAHRSDRAALS 20. L TAAEGHFD PLPAA DMSWDDKTAANQVLLYK LTKNDKYKAVE GSYTMWFPGGS . TPKGLAHRSDRAALS 20. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKNDKYKAVE GTFKYWFPGGS . TPKGLAHRSDRAALS 20. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS . TPKGLAHPSNRGALS 20. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS . TPKGLAHPSNRGALS 20. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS . TPKGLAYSMEWGSLR 29. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS . TPGGLAWSMMRGSLR 29. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS . TPGGLAWSMMRGSLR 29. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS . TPGGLAWSMMRGSLR 29. L TAAEGHFD PLPAP DMSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS . TPGGLAWSMMRGSLR 29. L TAAEGHFD FFYNKXXAGVVLLAQLTGGSEYQAARAFCDFSV - RQAKTPKGL	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066522.1 XP_013066522.1 XP_013066522.1 XP_01306622.1 XP_01306423.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552207.1 XP_01552217.1 XP_0155217.1 XP_01552217.	L Y D S Y S E KD F Y D S T S E SD F Y D S T S E SD F Y D S D D RD Y Y S S D D RD Y Y S S N D KD Y Y S S N I TD Y S S H G F ND N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND Y S S H G ND Y		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N O T L W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y Q A T N D M S W L Y R A T M D M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W W F K A T N D T M	L DDAKTYLE NY PFW SFDWENKQAACRQLLYE ETQSSVYKAVVEYLTTWLPGGT . TPCGLAWR IKWGSLR 27. L DYARTLA - DFNV - FW SY DWDNKQASCLLLLYE ETQSSVYKAVVEYLTTWLPGGT . TPCGLAWR KWGSLR 27. L DAARSY DLLAW DLSWLSKKPSCOLYLYE ETQSNYKAIQNYFNNWLPGGT . TPCGLAWR SKWGSNR 27. L EDAKTFV DLLAW DLSWLSKKPSCOLYLYE ETQSNYKAIQNYFNNWLPGGT . TPCGLAWRSKWGSNR 27. L EDAKTFV DLLAW ALSWDDKKI ACOLILYE ETQSNYKAIQNYFNNWLPGGT . TPCGLAWRSKWGANR 27. L HDAYD - YD LW - AW ALSWDDKKI VCOLILYE ETDSNYKAIQNYFNSUPGG . TPCGLAWRDAWGANR 27. L ALAEKYFD PAPAW GMSWDDKT I VNQVMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLGWGSLR 29. L ALAEKYFD PAPAW GMSWDDKT I VNQVMLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLGWGSLR 29. L ALAEKYFD DEPAW GMSWDDKT I VNQVMLYK ITKDEKYKAVEG SYTMWFPGGS ITPKGLAYRLGWGSLR 29. L ADAEKYLD DEPAW GMSWDDKT I VNQVLLYK LTKKDRYKAVEG SYTMWFPGGS ITPKGLAYRLGWGSLR 29. L TAAECHFD PLPAM GMSWDDKT I NOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLGWGSLR 29. L TAAECHFD PLPAM GMSWDDKTI ANOVLLYK LTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. L TAAECHFD PLPAM OMSWDNKTAANOVLLYK LTKDDKYKAVE OT FYWFPGGS ITPKGLAHRSDRAALS 29. L TAAECHFD PLPAM OMSWDNKTAANOVLLYK LTKDDKYKAVE OT TFWYFPGGS ITPKGLAHRSDRAALS 29. L TAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE DT FKYWFPGGS STPKGLAHPFNRGALS 29. L TAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKDDKYKAVE DT FKYWFPGGS STPKGLAHPFNRGALS 29. L TAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE DT FKYWFPGGS STPKGLAYSMEWGSLR 29. L TAAECHFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE DT FKYW	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_01306652.2 XP_01306652.2 XP_01306652.2 XP_01306622.1 XP_01306622.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_013064283.1 XP_01512607.1 XP_01512807.1 XP_01551382.2 XP_01551382.2 XP_01551382.2 XP_01551382.2 XP_01551382.2 XP_01650267.1 XP_01765382.9 XP_01650267.1 XP_01765382.9 XP_017765382.1 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_0151382.2 XP_016382.2 XP_017765382.1 XP_017765382.1	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D F Y D S D D R D F Y D S D D R D F Y D S N T D Y Y S S N I F D F Y S S H G F N D N Y S S H G F N D N Y S S H G F N D N Y S S H G F N D Y S S H G F		S L F L H K A T Q K S N AMWL Y R A T Q N D S T S I W L Y R A T Q N G Q A W L Y R A T Q N G Q A W L Y R A T R N N G Q A W L Y R A T R N N G Q S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y A T N D T S W L Y A T N D T S W L Y K A T N D T A W L F K A T N D S A W L F K A T N D S A W L F K A T N D T A W L F K A T N D	L DDAKTYLENY PWSFDWENKQAAGRQLLYE ETGSSVYKAVVEYLTTWLPGGT .TPCGLAWR IKWGSLR 27. L DDAKTYLENY PWSY DWDNKQAAGCLLLYE ETGSSVYKAVVEYLTTWLPGGT .TPCGLAWR IKWGSLR 27. L DAKTFV DLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT .TPCGLAWRSKWGSNR 27. L EDAKTFVDLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT .TPCGLAWRSKWGSNR 27. L EDAKTFVDLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT .TPCGLAWRSKWGSNR 27. L EDAKTFVDLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT .TPCGLAWRSKWGSNR 27. L HDAYDYD LW - AWALSWDDKKI XOLLIYE ETHDDR FSNV VGFMQSWL -DG ITPCGLAWRDMWGANR 27. L HDAYDYD LW - AWALSWDDKT I NOVMLYK ITKDEKYKAVE DSFK FWFPGGG ITPKGLAYRLQWGSLR 29. L ALAEKYFDPAPAWGMSWDDKT I NOVMLYK ITKDEKYKAVE DSFK FWFPGGS ITPKGLAYRLQWGSLR 29. L ADAEKYLDDE PAWGMSWDDKT I NOVLLYK ITKDEKYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L ADAEKYLDDE PAWGMSWDDKT I ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L TAAECHFDPLPAADMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L TAAECHFDPLPAADMSWDDKT ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLQWGSLR 29. L TAAECHFDPLPAADMSWDNKTAANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 20. L TAAECHFDPLPAPDMSWDNKTAANOVLLYK LTKKDKYKAVE GTFKYWFPGGS TPKGLAHRSDRAALS 20. L TAAECHFDPLPAPDMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS STPKGLAHPFNRGALS 20. L TAAECHFDPLPAPDMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGSSTPKGLAYSMEWGSLR 29. L EAAKYFDHP - ASEFSWDDKTVANOVLLYK LTKEDNYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. L EAAKYFDHP - ASEFSWDDKTVANOVLLYK LTKEDNYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. L EAAKYFDHP - ASEFSWDDKTVANOVLLYK LTKEDNYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. L EAAKTYPUHPASEFSWDDKTVANOVLLYK LTKEDNYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. L EAAKTYPUHPASEFSWDDKTVANOVLLYK LTKEDNYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 29. L EAAKTYPUHPFYYNKXAGVVLLAQLTGQSEYQAARAFCDFSVQQKTPKGLLYIDKFGTLC 29. L EAAKTYPUHPFYYNKXAGVV	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066952.1 XP_013066952.1 XP_013066952.1 XP_013064279.1 XP_013064279.1 XP_013064282.1 XP_013064282.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_0151542532.1 XP_015542535.1 XP_015542535.1 XP_015542535.1 XP_015542535.1 XP_015542535.1 XP_015542535.1 XP_015542535.1 XP_017785100.1 XP_017785100.1 XP_017785100.1 XP_017785100.1 XP_017785100.1 XP_017810680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1	L Y D S Y S E KD FY D S T S E S D FY D S T S E S D FY D S T D S D D R D FY D S D D R D Y Y S S N D T D Y Y S S N I T D Y S S H G F N D Y S H G F N D Y S S H G F N D Y S H G N D Y S S H G F N D Y S H G N D Y S S H G F N D Y S H G N D Y S S H G F N D Y S H G N		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O A W L Y R A T R N N G O S I W L Y Q A T N D K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S L W L Y Q A T N D T K S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y A T N N K O A L W L Y K A T N D T N A A W L F K A T N D S R A A W L F K A T N D S R A A W L F K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A I W L Y K A T N D T M A V W L F K A T N C M M A V W L F K A T N C M M M T M W T K A T N K T M M M W T K K A T N K T M M M W T K K A T N K T M M M W T K K A T N K T M M M W T K K A T N K T M M M W T K K A T N K T M M M W T K K A T N K T M M M W T K K A T N K T M M M W T K K A T N K T M M	LDDAKTYLENYPPW - SFDWENKQAAGRQLLYEETGSSVYKAVVEYLTTWLPGGT -TPGGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQAAGCLLLYETGSDGYKAIQNFFNNWLPGGT -TPGGLAWRSKWGSNR 27. LDAARSFY DLLAW - DLSWLSKKPSCOLYLFEETGSNGYKAIQNFFNNWLPGGT -TPGGLAWRSKWGSNR 27. LDAAKTFV DLLAW - ALSWDDKKIACOLIYEETGSNGYKAIQNYFNNUPGGGT -TPGGLAWRSKWGSNR 27. LDAAKTFV DLLAW - ALSWDDKKIACOLIYEETHDDRFSAVVGFMGSWLPGGT -TPGGLAWRDKWGANR 27. LHDAYD - YD LW - AW ALSWDDKKIACOLIYEETHDDRFSAVVGFMGSWLPGGT -TPGGLAWRDKWGANR 27. LHDAYD - YD LW - AW ALSWDDKKIACOLIYEETHDDRFSAVVGFMGSWLPGGT -TPGGLAWRDKWGANR 27. LALAEKYFD PAPAW - GMSWDDKTI NQVMLYK ITKDEKYKAVED SFKFWFPGGG ITPKGLAYRLOWGSLR 29. LALAEKYFD PAPAW - GMSWDDKTI NQVMLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 29. LALAEKYFD DEPAW - GMSWDDKTI NQVMLYK ITKDEKYKAVEGSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LADAEKYLD - DEPAW - GMSWDDKTI NQVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAM - GMSWDDKTI NQVLLYKLTKDDKYKAVEGSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAM - GMSWDDKTI NQVLLYKLTKDDKYKAVEGSYTMWFPGGS ITPKGLAYRLOWGSLR 29. LTAAEGHFD PLPAM - OMSWDNKTAANQVLLYKLTKDDKYKAVEGSYTMWFPGGS ITPKGLAHRSDRAALS 20. LTAAEGHFD PLPAM - OMSWDNKTAANQVLLYKLTKDDKYKAVEGSYTMWFPGGS ITPKGLAHRSDRAALS 20. LTAAEGHFD PLPAM - OMSWDNKTAANQVLLYKLTKEDKYKAVE TTFWYFPGGSSTPKGLAHRSDRAALS 20. LTAAEGHFD PLPAM - OMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGSSTPKGLAHRSDRAALS 20. LTAAEGHFD PLPAM - OMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGSSTPKGLAHRSDRAALS 20. LTAAEGHFD PLPAM - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAHRSDRAALS 20. LTAAEGHFD PLPAM - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAHRSDRAALS 20. LEAEKYFD - HP - AS - EFSWDDKTVANQVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAHPSNGALS 27. LEAACHYYG - HP - AS - EFSWDDKTVANQVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 27. LEAACHYYG - HP - AS - EFSWDDKTVANQVLLYKLTKEDKYKAVE NTFKYWFPGGSSTPKGLAYSMEWGSLR 27. LEAACHYYG HFHLERN - EFYYNKXAGVVLLAQLTGGSEYGAARAFCDFSV- RQKTPKGLLYIDKFGTLC 28. LEDAEHHYGHFHLERN - EFYYNKXAGVVLLAQLTGGSEYGAARAFCDFSV- RQ	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306652.1 XP_013064273.1 XP_01304283.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01304282.1 XP_01512607.1 XP_01512607.1 XP_01552207.1	L Y D S Y S E KD F Y D S Y S E KD F Y D S Y S E KD F Y D S D D RD Y Y S S D D RD Y Y S S N D E KD Y Y S S N I TD Y S S H G F ND N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND Y S S H G F NG Y S S H G F ND Y S S H G F ND		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N O T L W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S L W L Y Q A T N D T K S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T A W L Y K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L Y K A T G D S S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I S M W L Y K A T G D S I	LDDAKTYLENYPEW - SFDWENKQAACRQLLYE ETQSSVYKAVVEYLTTWLPGGT TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQAACQLLLYE ETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRRKWGSLR 27. LDAARSYV DLLAW - DLSWLSKKPSCOLYLYE ETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - ALSWDDKKIACQLIYE ETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAANSYD - AW - ALSWDDKKIACQLIYE ETQSNQYKAIQNYFNNWLPGGT - TPCGLAWRSKWGSNR 27. LDAANSYD - AW - ALSWDDKKIVACQLIYE ETDSNQYKAIQNYFNSWLPGG - TPCGLAWRSKWGSNR 27. LALAEKYFD - PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGG ITPKGLAYRLGWGSLR 29. LALAEKYFD - PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLGWGSLR 29. LALAEKYFD - PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEGSYTMWFPGGS ITPKGLAYRLGWGSLR 29. LADAEKYLD - DEPAW - GMSWDDKTIVNQVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLGWGSLR 29. LTAAECHFD - PLPAM - GMSWDDKTI NOVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLGWGSLR 29. LTAAECHFD - PLPAM - GMSWDDKTI NOVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLGWGSLR 29. LTAAECHFD - PLPAM - GMSWDDKTI ANOVLLYKLTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKDDKYKAVE TTQYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKDDKYKAVE TTGYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKDDKYKAVE DTFKYWFPGGS TPKGLAHPFNRGALS 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKEDKYKAVE DTFKYWFPGGS TPKGLAHPKNGALS 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKEDKYKAVE DTFKYWFPGGS TPKGLAHPKNGALS 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAHPKNGALS 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LAAAEKYFD - HP - AS - EFSWDDKYVANOVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LAAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LAAAECHFD - PLPAM - DMSWDNKTAANOVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LAAAECHFD - NEFFYNKXXAGVVLLXQUTALACTGOS VAAAFCDFSV - NQKTYKGLAYSMEWGSL	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_013066170.1 XP_013066172.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_0151384271.1 XP_0151384271.1 XP_015513822.1 XP_015513822.1 XP_015513822.1 XP_015513822.1 XP_015513822.1 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_01561382.21 XP_016607665.1 XP_01616680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1 XP_018016680.1	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D Y Y S S N F T D Y S S H G F N D F Y S S H G F N D Y T S S H F N D Y T S S H F N D Y T S S H F N S H F N D Y T S S H F N S		S L F L H K A T Q K S N AMWL Y R A T Q N D S T S I W L Y R A T Q N G Q A W L Y R A T Q N G Q A W L Y R A T R N N G Q T L W L Y R A T H N T Q S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D T N S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D T S W L Y Q A T N D T S W L Y A T N D T A W L Y K A T N D T A W L Y K A T N D S A W L Y K A T N D S A W L F K A T N D S A W W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L F K A T N D T A W L Y K A T R D T A W L Y K A T R D T A W L Y K A T R D T A W L Y K A T N D T A W	L DDAKTYLENY PWSFDWENKQAAGRQLLYE ETGSSVYKAV VEYLTTWLPGGT .TPGGLAWR IKWGSLR 27. L DDAKTYLENY PWSY DWDNKQAAGCLLLYE ETGSSVYKAV VEYLTTWLPGGT .TPGGLAWR KWGSLR 27. L DAKTFV DLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT .TPGGLAWRSKWGSNR 27. L EDAKTFVDLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT .TPGGLAWRSKWGSNR 27. L EDAKTFVDLLAWALSWDDKKIACOLIYE ETGSNYKAIQNYFNNWLPGGT .TPGGLAWRSKWGSNR 27. L HDAYDYD LW - AWALSWDDKKIACOLIYE ETGSNYKAIQNYFNSWLPGGT .TPGGLAWRDKWGANR 27. L HDAYD - YD LW - AWALSWDDKKI VOLIYE ETHDDR FSAV VG FMGSWLPGG .TPGGLAWRDMWGANR 27. L HLAKYFD PAPAWGMSWDDKTI NOVMLYK ITKDEKYKAV ED SFK FWFPGG ITPKGLAYRLOWGSLR 29. L ALAEKYFD PAPAWGMSWDDKTI NOVMLYK ITKDEKYKAV ED SFK FWFPGGS ITPKGLAYRLOWGSLR 29. L ADAEKYLDDE PAWGMSWDDKTI ANQVLLYK LTKKDRYKAV EG SYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAECHFDPLPAADMSWDDKTI ANQVLLYK LTKKDRYKAV EG SYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAECHFDPLPAADMSWDDKTI ANQVLLYK LTKKDRYKAV EG SYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAECHFDPLPAADMSWDDKTAANQVLLYK LTKKDKYKAV EG SYTMWFPGGS ITPKGLAHRSDRAALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYK LTKKDKYKAV EG SYTMWFPGGS ITPKGLAHRSDRAALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYK LTKEDKYKAV ED TFKYWFPGGS YTPKGLAHRSDRAALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYK LTKEDKYKAV ED TFKYWFPGGS YTPKGLAHPFNRGALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYK LTKEDKYKAVE DTFKYWFPGGS YTPKGLAHPFNRGALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYK LTKEDKYKAVE NTFKYWFPGGS YTPKGLAHPFNRGALS 20. L TAAECHFDPLPAPDMSWDNKAACOULLYK LTKEDKYKAVE NTFKYWFPGGS YTPKGLAHPFNRGALS 20. L EAKT HYWALSWEDKKVSCOLLLYE LTGDAS YAARAFCDFSVQQKTPKGLUYINKGYKGNR 27. L ELAKT HYWALSWEDKKVSCO	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066952.1 XP_013066952.1 XP_013066952.1 XP_013064279.1 XP_013064279.1 XP_013064282.1 XP_013064282.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_013084285.1 XP_0151084255.1 XP_0151642558.1 XP_01559297.1 XP_01559297.1 XP_01559297.1 XP_01559297.1 XP_01559297.1 XP_01559297.1 XP_01801680.1 XP_01801680.1 XP_018016860.1 XP_01801680.1	L Y D S Y S E KD F Y D S T S E S D F Y D S T S E S D F Y D S T D S D D R D F Y D S D D R D Y Y S S N I T D Y S S H G F N N Y S S H G F N N Y S S H G F N N Y S S H G F N Y S S H G F N Y S S H G F N F Y S S N Y T D F Y S S N Y T D F Y S S T Y G Y S S S S S S S S S S S S S S S S S S S		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O T L W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y A T N N C A L W L Y K A T N D T A A W L F K A T N D S R A A W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A L W L F K A T N D T A W W T F K A T N T A W T T T T T T T A W T T T T T T T T T T T T T T T T T T	L DDAKTYLE NY PPW - SFDWENKQAAGRQLLYE ETGSSVYKAVVEYLTTWLPGGT TPGGLAWRIKWGSLR 27. L DYARTLA - DFNV - PW SY DWDNKQASCLULYE ETGSSVYKAVVEYLTTWLPGGT TPGGLAWRFKWGSLR 27. L DAARSFY DLLAW DLSWLSKKPSCOLYLFE ETGSNYKAIQNYFNNWLPGGT TPGGLAWRSKWGSNR 27. L EDAKTFV DLLAW DLSWLSKKPSCOLYLFE ETGSNYKAIQNYFNNWLPGGT TPGGLAWRSKWGSNR 27. L HDAYD - YD LW - AW ALSWDDKKIACOLILYE ETHDDR FSAVVGFMGSWLPGG TTPGGLAWRDKWGSNR 27. L HDAYD - YD LW - AW ALSWDDKKI VQUNLYK ITKDE YKAV GFMGSWLPGG TTPGGLAWRDKWGSNR 27. L HDAYD - YD DPAW GMSWDDKTI VNOVMLYK ITKDE YKAVED SFKFWFPGGG ITPKGLAYRLOWGSLR 29. L ALAEKYFD PAPAW GMSWDDKTI NOVMLYK ITKDE YKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 29. L ALAEKYFD PAPAW GMSWDDKTI NOVMLYK ITKDE YKAVEG SYTMWFPGGS ITPKGLAYRLOWGSLR 29. L ADAEKYLD DEPAW GMSWDDKTI ANOVLLYK LTKKDRYKAVEG SYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAEGHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVEG SYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAEGHFD PLPAM GMSWDDKTI ANOVLLYK LTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAEGHFD PLPAM GMSWDDKTI ANOVLLYK LTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 20. L TAAEGHFD PLPAM GMSWDDKTI ANOVLLYK LTKDDKYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 20. L TAAEGHFD PLPAM OMSWDNKTAANOVLLYK LTKDDKYKAVE TTGYWFPGGS VTPKGLAHRSDRAALS 20. L TAAEGHFD PLPAP OMSWDNKTAANOVLLYK LTKDDKYKAVE TTFKYWFPGGS VTPKGLAHRSDRAALS 20. L TAAEGHFD PLPAP OMSWDNKTAANOVLLYK LTKEDKYKAVE DTFKYWFPGGS VTPKGLAHRSDRAALS 20. L TAAEGHFD PLPAP OMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAHRSDRAALS 20. L TAAEGHFD PLPAP OMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAHPSNRGALS 29. L TAAEGHFD PLPAP OMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAHPSNRGALS 29. L TAAEGHFD PLPAP OMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAYSMEWGSLR 29. L TAAEGHFD PLPAP OMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAYSMEWGSLR 29. L EAAEKYFD PLPAP DMSWDNKTAANOVLLYK LTKEDKYKAVE NTFKYWFPGGS VTPKGLAYSMEWGSLR 2	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306652.1 XP_01306627.1 XP_013064283.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_013084281.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01552207.1 XP	L Y D S Y S E KD F Y D S Y S E KD F Y D S Y S E KD F Y D S D D RD Y Y S S D D RD Y Y S S N D E KD Y Y S S N I TD Y S S I G F ND N Y S S I G F ND N Y S S I G F ND N Y S S I G F ND Y S S I G F ND Y S S I G S TO Y S S S S S TO		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N O T L W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S W L Y Q A T N D T K S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y Q A T N D T T S W L Y A T T N K O A W L Y K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T M A W L F K A T M T M T M A W L F K A T M T M T M A W L F K A T M T M T M A W L F K A T M T M T M A W L F K A T M T M T M A W L F K A T M T M T M A W L F K A T M T M T M A W L F K A T M T M T M A W L F K A T M T M T M T M A W L Y K A T M T M T M T M T M A W L Y K A T M M T M T M T M T M T M T M T M T M	LDDAKTYLENYPPW - SFDWENKQAACRQLLYE ETQSSVYKAVVEYLTTWLPGGT -TPCGLAWRIKWGSLR 27. LDYARTLA - DFNV - PW - SYDWDNKQASCLLLYE ETQSSVYKAVVEYLTTWLPGGT - TPCGLAWRSKWGSLR 27. LDAARSYV DLLAW - DLSWLSKKPSCOLYLYE ETQSNQYKAIQNYFNNWLPGGT -TPCGLAWRSKWGSNR 27. LEDAKTFV DLLAW - ALSWDDKKIACQLIYE ETQSNYKAIQNYFNNUPGGT -TPCGLAWRSKWGNR 27. LHDAYD - YD - LW - AW - ALSWDDKKIVQCULIYE ETQSNYKAVGFMQSWLPGGG -TPCGLAWRSKWGSNR 27. LHDAYD - YD - LW - AW - ALSWDDKKIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGG ITPKGLAWRLWGSLR 29. LALAEKYFD - PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEDSFKFWFPGGS ITPKGLAYRLWGSLR 29. LALAEKYFD - PAPAW - GMSWDDKTIVNQVMLYK ITKDEKYKAVEGSYTMWFPGGS ITPKGLAYRLWGSLR 29. LADAEKYLD - DEPAW - GMSWDDKTIVNQVMLYK LTKDDKYKAVEGSYTMWFPGGS ITPKGLAYRLWGSLR 29. LADAEKYLD - DEPAW - GMSWDDKTINNQVLLYKLTKKDRYKAVEGSYTMWFPGGS ITPKGLAYRLWGSLR 29. LTAAECHFD - PLPAP - OMSWDDKTIANQVLLYKLTKDDKYKAVEGSYTMWFPGGS ITPKGLAYRLWGSLR 29. LTAAECHFD - PLPAA - OMSWDDKTIANQVLLYKLTKDDKYKAVEGSYTMWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKDDKYKAVE TTQYWFPGGS ITPKGLAHRSDRAALS 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKDDKYKAVE TTQYWFPGGS ITPKGLAHPFNRGALS 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS TPKGLAHPFNRGALS 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS TPKGLAHPFNRGALS 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS TPKGLAHPKNRGALS 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LAAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LTAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LAAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAYSMEWGSLR 29. LAAAECHFD - PLPAP - OMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGS TPKGLAY	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_013066170.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_01306622.1 XP_0151684471.1 XP_01551382.1 XP_01551382.1 XP_01551382.1 XP_01551382.1 XP_01551382.1 XP_01551382.1 XP_01551382.1 XP_01561863.1 XP_01660762.5 XP_01778510.0 XP_01660762.5 XP_016076.5 XP_016076.5 XP_016076.5 X	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D Y Y S S N F T D Y Y S S N I T D Y Y S S N I T D Y Y S S N I F D Y Y S S N F T D N Y S S H G F N D N Y S S H G F N D N Y S S H G F N D Y S S H G F N D Y S S H G F N D F Y S S H G F N D Y H S H F H H H H H H H H H H H H H H H H		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O S I W L Y Q A T N D N K S I W L Y Q A T N D N K S I W L Y Q A T N D N K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y A T N K O A L W L Y K A T N D T N A W L Y K A T N D T N A AWL F K A T N D S R A W L F K A T N D S R A W L F K A T N D T M A W L Y K A T G D S A I W L Y K A T G D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A W L Y K A T G D D S A I	L DDAKTYLENY PWSFDWENKQAAGRQLLYE ETGSSVYKAVVEYLTTWLPGGT -TPGGLAWRIKWGSLR 27. L DYARTLA - DFNV - PWSY DWDNKQAAGCLLLLYE ETGSSVYKAVVEYLTTWLPGGT -TPGGLAWRFKWGSLR 27. L DAARSFY DLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT -TPGGLAWRSKWGSNR 27. L EDAKTFYDLLAWDLSWLSKKPSCOLYLYE ETGSNYKAIQNYFNNWLPGGT -TPGGLAWRSKWGSNR 27. L HDAYD - YD LW - AWALSWDDKKI ACOLILYE ETHDDR FSAVVGFMQSWLPGG -TPGGLAWRDKWGANR 27. L HDAYD - YD LW - AWALSWDDKKI XCOLILYE ETHDDR FSAVVGFMQSWLPGG -TPGGLAWRDKWGANR 27. L HLAKYFD PAPAWGMSWDDKTI VNOVMLYK ITKDEKYKAVED SFKFWFPGG ITPKGLAYRLOWGSLR 29. L ALAEKYFD PAPAWGMSWDDKTI NOVMLYK ITKDEKYKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 29. L ALAEKYFD PAPAWGMSWDDKTI NOVLLYK LTKKDRYKAVED SFKFWFPGGS ITPKGLAYRLOWGSLR 29. L ADAEKYLDDEPAWGMSWDDKTI ANQVLLYKLTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAECHFDPLPAADMSWDDKTI ANQVLLYKLTKKDRYKAVE GSYTMWFPGGS ITPKGLAYRLOWGSLR 29. L TAAECHFDPLPAADMSWDDKTAANQVLLYKLTKKDRYKAVE GSYTMWFPGGS ITPKGLAHRSDRAALS 29. L TAAECHFDPLPAADMSWDDKTAANQVLLYKLTKDDKYKAVE DTFKYWFPGGS ITPKGLAHRSDRAALS 29. L TAAECHFDPLPAPDMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGS VTPKGLAHRSDRAALS 29. L TAAECHFDPLPAPDMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGS VTPKGLAHPFNRGALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGS VTPKGLAHPFNCALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGS VTPKGLAHPFNCALS 20. L TAAECHFDPLPAPDMSWDNKTAANQVLLYKLTKEDKYKAVE DTFKYWFPGGS VTPKGLAYSMEWGSLR 20. L EAAEKYFDPLPAPDMSWDNKTAANQVLLYKLTKEDKYKAVE NTFKYWFPGGS VTPKGLAYSMEWGSLR 20. L EAAELYFDPLPAP	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066161.1 XP_013066952.1 XP_013066952.1 XP_013066952.1 XP_013064279.1 XP_013064279.1 XP_013064283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_013084287.1 XP_015126207.1 XP_015184471.11 XP_0151842476.12 XP_015184471.11 XP_0151842476.11 XP_015592927.11 XP_01680686.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801682.1	L Y D S Y S E KD FY D S T S E S D FY D S T S E S D FY D S T S E S D FY D S T D R D FY D S D D R D Y Y S S N F T D Y Y S S H G F N D Y S S H F N D Y S S H G F N D Y S S H F N D Y S S H F N D Y S H H H H H H H H H H H H H H H H H H		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G Q S I W L Y R A T Q N G Q A W L Y R A T R N N G Q A W L Y R A T R N N G Q S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D K S I W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y Q A T N D T S L W L Y A T N N K O A L W L Y K A T N D T A AW L F K A T N D S A AW L F K A T N D S A AW L F K A T N D T A AW L F K A T N D T A L W L Y K A T R D T A AW L F K A T N D S A W W L F K A T N D T A L W L Y K A T R D T A W W L F K A T N D T A L W L Y K A T R D T A W W L F K A T N D T A W W F K A T N D T A W W F K A T N T	L DDAK TY LE NY P PW SFDWEN KGA ACROLLYE E TGSSVYKA VVEYLTTWL PGGT - TP CGLAWR FKWGSLR 27. L DYARTLA - DF NY - PW SYDWD NKGA ACROLLYE E TGSSVYKA VVEYLTTWL PGGT - TP CGLAWR FKWGSLR 27. L DAARSY DLLAW DLSWL SKKP SCOLYL E E TGSN YKA I QN YF SWL PGG - TP CGLAWR SKWGSNR 27. L EDAKTFY DLLAW DLSWL SKKP SCOLYL E E TGSN YKA I QN YF SWL PGG - TP CGLAWR SKWGSNR 27. L HDAYD - YD UW - AW ALSWD DKKI A COLI Y E E TGSN YKA I QN YF SWL PGG - TP CGLAWR DWGANR 27. L HDAYD - YD WW ALSWD DKKI A COLI Y E E TGSN YKA I QN YF SWL PGG - TP CGLAWR DWGANR 27. L HDAYD - YD DL PAW GMSWD DKTI V NOVML YK I TKDE KYKA VE D SFK FWF PGG I TP KGLAYR LQWGSLR 29. L ALAEKYFD PA PAW GMSWD DKTI V NOVML YK I TKDE KYKA VE D SFK FWF PGG I TP KGLAYR LQWGSLR 29. L ADAEKYLD DE PAW GMSWD DKTI ANOVLLYK LTKD RYKA VE GSYT MWF PGGS I TP KGLAYR LQWGSLR 29. L ADAEKYLD DE PAW GMSWD DKTI ANOVLLYK LTKD RYKA VE GSYT MWF PGGS I TP KGLAYR LQWGSLR 29. L TAAE QHF D PL PAM GMSWD DKTI ANOVLLYK LTKD RYKA VE GSYT MWF PGGS I TP KGLAYR LQWGSLR 29. L TAAE QHF D PL PAM OMSWD NK TAANOVLLYK LTKD DK YKA VE D' FFY WF PGGS I TP KGLAHFS DRAALS 20. L TAAE QHF D PL PAM OMSWD NK TAANOVLLYK LTKD DK YKA VE D' FFY WF PGG S' TP KGLAHF SNRAALS 20. L TAAE QHF D PL PAM OMSWD NK TAANOVLLYK LTKD DK YKA VE D' TFKY WF PGG S' TP KGLAHF SNRAALS 20. L TAAE QHF D PL PAM OMSWD NK TAANOVLLYK LTKD DK YKA VE D' TFKY WF PGG S' TP KGLAHF SNRAALS 20. L TAAE QHF D PL PAM DMSWD NK TAANOVLLYK LTKD DK YKA VE D' TFKY WF PGG S' TP KGLAHF SNRAALS 20. L TAAE QHF D PL PAM DMSWD NK TAANOVLLYK LTKD DK YKA VE D' TFKY WF PGG S' TP KGLAY MEWGSLR 29. L TAAE QHF D PL PAM DMSWD NK TAANOVLLYK LTKD DX YKA VE WT FKY WF PGG S' TP KGLAY MEWGSLR 29. L TAAE QHF D PL PAM DMSWD NK TAANOVLLYK LTKD DX YKA VE WT FKY WF PGG S' TP KGLAY MEWGSLR 20. L EAAE YF D HP - AS EFS WD KTV ANOVLLYK LTKD DX YKA VE WT FKY WF PGG S' TP KGLAY MEWGSLR 29. L EAAE YF D HP S	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306652.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_013084283.1 XP_01512607.1 XP_01512607.1 XP_01552207.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801682.1 XP_01801682.1 XP_01801682.1 XP_01802466.1 XP_01802466.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802462.1 XP_01802463.1 XP_0180245.1 XP_0180245.1 XP_0180245.1 XP_0180245.1 XP_01	L Y D S Y S E KD F Y D S Y S E KD F Y D S Y S E KD F Y D S D D RD Y Y S S N D E ND Y Y S S N D E ND Y Y S S N I TD Y S S H G F ND N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND Y S S H G F NC Y S S H F NC Y S S H F NC Y		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N O T L W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y A T N N K O A I W L Y A T N N K O A I W L Y K A T N D T N A W L F K A T N D T N A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L Y K A T G D E A I W L Y K A T G D T A I W L Y K A T G D T A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W L Y K A T R K L A I W	L DDAK TY LE NY P PW SFDWEN KGA AGROLL YE E TGS SV YKA V VE YL TTWL PGGT - T PGGLAWR FKWGSLR 27. L DDAK TFV DLLAW - DLSWL SKAP SGOLYL YE E TGS SV YKA V VE YL TTWL PGGT - T PGGLAWR FKWGSLR 27. LEDAK TFV DLLAW DLSWL SKAP SGOLYL YE E TGS NG YKA I GN YF NWL PGGT - T PGGLAWR SKWGSLR 27. LEDAK TFV DLLAW DLSWL SKAP SGOLYL YE E TGS NG YKA I GN YF NWL PGG - T PGGLAWR SKWGSNR 27. LEDAK TFV DLLAW DLSWL SKAP SGOLYL YE E TGS NG YKA I GN YF NWL PGG - T PGGLAWR DGWGANR 27. LHDAN SLUD AW ALSWD DKKI VA GOLI YE E THD DR FS AV VG FMG NUL PGG I T PFGLAWR DGWGANR 27. LHDAN SLUD AW ALSWD DKKI VA GOLI YE E THD DR FS AV VG FMG NUL PGG I T PFGLAWR DGWGANR 27. LA LA KYF D PA PAW GMS WD DKT I VNOVUL YK ITKDE KYKAV ED SFK FWP GGG I T PFGLA YR LGWGSLR 29. LA AL KYF D DE PAW GMS WD DKT I ANOVL YK ITKDE KYKAV ED SFK FWP GGG I T PFGLA YR LGWGSLR 29. LA DA EKYL D DE PAW GMS WD DKT I ANOVL YK LTKD RYKAV EG SYT MWF PGG SI T PFGLAYR LGWGSLR 29. L TAAE GHF D PL PAM GMS WD DK TI ANOVL YK LTKD DK YKAV E G SYT MWF PGG SI T PFGLAYR LGWGSLR 29. L TAAE GHF D PL PAM DMS WD NK TA ANOVL YK LTKD DK YKAV EN T FG YWF PGG SI T PFGLAHRS DRAALS 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL YK LTKD DK YKAV EN T FK YWF PGG SI T PFGLAHRS DRAALS 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL LYK LTKD DK YKAV EN T FK YWF PGG SI T PFGLAHRS DRAALS 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL LYK LTKD DK YKAV EN T FK YWF PGG SI T PFGLAHRS DRAALS 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL LYK LTKD DK YKAV EN T FK YWF PGG SI T PFGLAWR MWGSLR 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL LYK LTKD DK YKAV EN T FK YWF PGG SI T PFGLAWAR WOSLR 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL LYK LTKD DK YKAV EN T FK YWF PGG SI T PFGLAWAR WOSLR 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL LYK LTKD DK YKAV EN T FK YWF PGG SI T PFGLAWAR WOSLR 20. L TAAE GHF D PL PAM DMS WD NK TA ANOVL LYK LTKD DK YKAV EN T FK YWF PGG SI T PFGLA	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066422.1 XP_013066422.1 XP_013064270.1 XP_013046285.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_0154214612.1 XP_0154214612.1 XP_0154262.1 XP_0154262.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01563282.1 XP_01801685.1 XP_01801685.1 XP_01801685.1 XP_01801685.1 XP_01801685.1 XP_01801685.1 XP_01801685.1 XP_01801685.1 XP_01801682.1 XP_01801683.1 XP_01801	L Y D S Y S E KO F Y D S T S E S D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D F Y D S D D R D F Y S S D S A D D Y Y S S N I T D Y Y S S N I T D Y Y S S N I F D Y Y S S N I F D Y Y S S N F F D N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND Y S S H G F ND F Y S S N F T D F Y S S N F T D F Y S S I D G D Y E S T D G D Y Y S S G G D F		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O S I W L Y Q A T N D K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y A T N N K O A L W L Y K A T N D T N A W L Y K A T N D T N A W L Y K A T N D T N A W L F K A T N D S R A W L F K A T N D S R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L F K A T N D T R A W L Y K A T G D S A I W L Y K A T G D S A I W L Y K A T G D T S A W L F K A T G D T K A W L F K A T G D T K A W L F K A T R T T K K A W L F K A T R T T K K A W L F K A T R T T T T T T T T T T T T T T T T T	L DDAK TYLE NY PPW - SFDWEN KGA AGRGLLYE ETGSSVYKAVVEYLTTWLPGGT - TPGGLAWR FKWGSLR 27. L DYARTLA - DFNY - PW - SFDWEN KGA AGRGLLYE ETGSSVYKA VEYLTTWLPGGT - TPGGLAWR FKWGSLR 27. L DAARSY DLLAW - DLSWLSKHPSGDLYIE ETGSNQYKA I QN YFNSWLPGGT - TPGGLAWR FKWGSLR 27. L EDAKTFY DLLAW DLSWLSKHPSGDLYIE ETGSNQYKA I QN YFNSWLPGG - TPGGLAWR SKWGSNR 27. L EDAKTFY DLLAW ALGWDDKKI AGGLLYE ETGSNQYKA I QN YFNSWLPGG - TPGGLAWR DWGANR 27. L HDAYD AW - ALGWDDKKI AGGLLYE ETGSNQYKA I QN YFNSWLPGG - TPGGLAWR DWGANR 27. L HDAYD AW - ALGWDDKKI AGGLLYE ETGSNGYKA I QN YFNSWLPGG - TPGGLAWR DWGANR 27. L HDAYD DL AW ALSWDDKKI AGGLLYE ETGSNGYKA I QN YFNSWLPGG - TPGGLAWR DWGANR 27. L HDAKYD DPAPAW GMSWDDKTI VNOVMLYK I TKDEKYKA VEDSFK FWP GGG I TPKGLAYR LQWGSLR 28. L ALAEKYFD PAPAW GMSWDDKTI ANQULLYK I TKDEKYKA VEDSFK FWP GGG I TPKGLAYR LQWGSLR 28. L ADAEKYLD DL FAW GMSWDDKTI ANQULLYK I TKDR YKA VEGSYT MWF PGGS I TPKGLA YR LQWGSLR 28. L TAAE QHFD PL FAA DMSWDNKTA ANQULLYK I TKDDK YKA VEGSYT MWF PGGS I TPKGLAHRSDRAALS 28. L TAAE QHFD PL FAA DMSWDNKTA ANQULLYK I TKDDK YKA VED TFK YWF PGGS T FKGLAHF SDRAALS 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDK YKA VED TFK YWF PGGS T FKGLAHP FNRGALS 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDK YKA VED TFK YWF PGGS T FKGLAHP FNRGALS 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDYKA VENT FK YWF PGGS T FKGLAYSMEWGSLR 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDYKA VENT FK YWF PGGS T FKGLAYSMEWGSLR 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDYKA VENT FK YWF PGGS T FKGLAYSMEWGSLR 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDY YA VENT FK YWF PGGS T FKGLAYSMEWGSLR 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDYYA VENT FK YWF PGGS T FKGLAYSMEWGSLR 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKDDYYA VENT FK YWF PGGS T FKGLAYSMEWGSLR 28. L TAAE QHFD PL FAP DMSWDNKTA ANQULLYK L TKAE DYYA VENT FK YWF PGGS T FKGLAYSMEWGS	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066522.1 XP_013066522.1 XP_013066522.1 XP_01306622.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_014786504.1 XP_015184258.1 XP_015164258.1 XP_015562927.1 XP_015562927.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801682.1 XP_01801683.	L Y D S Y S E KO F Y D S T S E S D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D Y Y S S N D E D Y Y S S N F T D Y S S H G F N D N Y S S H G F N D N Y S S H G F N D N Y S S H G F N D Y S S H G F N Y S S H G S N Y S S H S N S N S N Y S S H S N S N S N S N S N S N Y S S H S N S N S N S N S N S N S N S N S		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O T L W L Y Q A T N D N K S I W L Y Q A T N D N K S I W L Y Q A T N D N K S I W L Y Q A T N D N K S I W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D T N A W L Y K A T N D T N A W L Y K A T N D T N A W L F K A T N D T N A L W L F K A T N D T N A L W L F K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L F K A T N D T M A L W L Y K A T N D T M A W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L W L Y K A T N D T M A L	L DDAK TY LE NY PPW - SFDWENKQAACRQLLYE ETQSSVYKAVYEYLTTWLPGGT - TPGGLAWRIKWGSLR 27. L DTAK TY LE NY PPW - SFDWENKQASCLLLYE ETGSSVYKAVYEYLTTWLPGGT - TPGGLAWRFKWGSLR 27. L EDAK TFV DLLAW - DLSWS KKPSCOLYLFE ETQSNQYKAIONYFNSWLPGGT - TPGGLAWRSKWGSNR 27. L EDAK TFV DLLAW ALGWDDKKIACQLILYE ETGSNQYKAIONYFNSWLPGGT - TPGGLAWRGKWGSNR 27. L HDAYD - NW ALGWDDKKIACQLILYE ETGSNQYKAIONYFNSWLPGGT - TPGGLAWRDWGANR 27. L HDAYD - NW ALGWDDKKIACQLILYE ETTQSNQYKAIONYFNSWLPGGT - TPGGLAWRDWGANR 27. L HDAYD - NU AW ALGWDDKKIACQLILYE ETTOSNQYKAIONYFNSWLPGGT - TPGLAWRDWGANR 27. L HDAKT D PAPAW GMSWDDKTIVNQYMLYKITKDEKYAX EDSFKFWFPGGGITPKGLAYRLGWGSLR 20. L ALAEKYFD PAPAW GMSWDDKTIVNQYMLYKITKDEKYAX EDSFKFWFPGGSITPKGLAYRLGWGSLR 20. L ADAEKYLD DEPAW GMSWDDKTINNQYLLYKLTKNDYKAYE GSYTWFPGGSITPKGLAYRLGWGSLR 20. L ADAEKYLD DEPAW GMSWDDKTINNQYLLYKLTKNDYKAYE GSYTWFPGGSITPKGLAYRLGWGSLR 20. L TAAEQHFD PLPAM GMSWDDKTINNQYLLYKLTKNDYKAYE GSYTWFPGGSITPKGLAYRLGWGSLR 20. L TAAEQHFD PLPAM GMSWDNKTAANQYLLYKLTKNDYKAYE GSYTWFPGGSITPKGLAHRSDRAALS 20. L TAAEQHFD PLPAM DMSWDNKTAANQYLLYKLTKNDYKAYE GYYMWFPGSITPKGLAHRSDRAALS 20. L TAAEQHFD PLPAM DMSWDNKTAANQYLLYKLTKDTYKAYE TYKYPFGSYTPKGLAHRSDRAALS 20. L TAAEKYK D HP AS EFSWDDKTYANQYLLYKLTKDTYKAYE TYKYPFGSYTPKGLAHRSDRAALS 20. L TAAEKYK D HP AS EFSWDDKTYANQYLLYKLTKDTYKAYE TYKYPFGSYTPKGLAHRSDRAALS 20. L TAAEKYK D HP AS EFSWDDKTYANQYLLYKLTKDTYKAYE TYKYPFGSYTPKGLAHRSDRAALS 20. L ELAVDHYHFKRTYN . EFYYNKKYAGYYLLAQLTGDYE YGAARAFCDFSY . RQGKTPKGLLY DKFGTLC 20. L ELAVDHYHFKRTYN . EFYYNKKYAGYYLLAQLTGDYE YGAARAFCDFSY . RQGKTPKGLLY DKFGTLC 20. L ELAVHYHFKRTYN . EFYYNKKYAGYY	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306652.1 XP_01306623.1 XP_01306423.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01308428.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_015522607.1 XP_015522607.1 XP_015522607.1 XP_015522607.1 XP_01563621.1 XP_01563621.1 XP_01616806.1 XP_01801680.1 XP_01801682.1 XP_01801682.1 XP_0180375.1 XP_0183255.1 XP_0	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D F Y D S D D R D F Y D S D D R D Y Y S S N F T D Y Y S S N I T D Y Y S S N I T D Y Y S S N I F D F Y S S H G F ND N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND F Y S S H G F ND Y Y S S H G F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H S F N T D F ND Y S S H		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N O T L W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N E N K S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S I W L Y Q A T N D M S W L Y R A T N N K O A I W L Y K A T N D T N A W L F K A T N D T N A W L F K A T N D T N A W L F K A T N D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T M A W L F K A T M D T K A W L F K A T M D T K A W L F K A T M D T K A W L F K A T M T M K K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A W L F K A T M T M K A	L DDAKTYLE HY PPW - SFDWENKGAACRGLLYE ETGSSVYKAVYEYLTTWLPGGT - TPGGLAWRIKWGSLR 27. L DDAKTYLE - NUY - PW - SFDWENKGASCLLLYE ETGSSVYKAVYEYLTTWLPGGT - TPGGLAWRIKWGSLR 27. L DDAKTFY - DLLAW - DLSWSKKPSGLYLYE ETGSNGYKAIONYFNSWLPGGT - TPGGLAWRSKWGSNR 27. L EDAKTFY - DLLAW - OLSWSKKPSGLYLYE ETGSNGYKAIONYFNSWLPGGT - TPGGLAWRGWGGANR 27. L HDAYD - D - AW - AW - ALGWDKKIACGLILYE ETHDR SAVYGFMGSWLPGG I TPGGLAWRGWGGANR 27. L HDAYD - D - AW - AW - ALGWDKKIACGLILYE ETHDR SAVYGFMGSWLPGG I TPGGLAWRGWGGANR 27. L HDAYD - D - PAPAW - ONSWDDKTIVNQYMLYKITNDEKYKAYEDSFFKWFPGGG I TPGLAWRGWGGANR 27. L HDAKTYD - PAPAW - ONSWDDKTIVNQYMLYKITNDEKYKAYEDSFFKWFPGGG I TPGLAWRGWGSLR 29. L ALAEKYFD - PAPAW - ONSWDDKTIVNQYMLYKITNDEKYKAYEGSYTWFPGGS I TPGLAYRLQWGSLR 29. L ADAEKYLD - DEPAW - ONSWDDKTI NAQVLLYKLTKDRYKAYEGSYTWFPGGS I TPGGLAYRLQWGSLR 29. L ADAEKYLD - DEPAW - ONSWDDKTI NAQVLLYKLTKDKYKAYEGSYTWFPGGS I TPGGLAYRLQWGSLR 29. L TAAEQHFD - PLPAA - DWSWDNKTAANQYLLYKLTKDKYKAYEGSYTWFPGGS I TPGGLAYRLQWGSLR 29. L TAAEQHFD - PLPAA - DWSWDNKTAANQYLLYKLTKDKYKAYEGYKWWFPGGS I TPGGLAHFNGALS 20. L TAAEQHFD - PLPAA - DWSWDNKTAANQYLLYKLTKDKYKAYEGYKWWFPGGS I TPGGLAHFNRGALS 20. L TAAEQHFD - PLPAA - DWSWDNKTAANQYLLYKLTKDKYKAYEGYKWWFPGGS I TPGGLAHFNRGALS 20. L TAAEQHFD - PLPAA - DWSWDNKTAANQYLLYKLTKDKYKAYEGYKWWFPGGS TPKGLAYSMEWGSLR 29. L TAAEQHFD - PLPAA - SFSWDDKTVANQYLLYKLTKDKYKAYEGYKAYEGGS TPGGLAWRMWGSLR 27. L TAAEQHFD - PLPAA - SFSWDDKTVANQYLLYKLTKDTYKAYENTFKYWFPGGSTPKGLAYSMEWGSLR 29. L AEAEKYFD - HP - AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSTPKGLAYSMEWGSLR 29. L EAAEKYFD - HP - AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSTPKGLAYSMEWGSLR 29. L EAAEAKYFD - HP - AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSTPKGLAYSMEWGSLR 29. L ELAADHYHRFRLERN - EFFYNKKYAGUSKULAGUTGGEYAARAFCOFSY. RQKTPKGLYI DKGTLC 20. L ELAADHYHRFRLERN - EFFYNKKYAGUSKULAGUTGGEYAARAFCOFSY. RQKTPKGLYI DKGTLC 20. L ELAADHYHRFRLERN - EFFYNKKYAGUSKULAGUTGGEYAARAFCOFSY. RQKTPKGLYI DKGTLC 20. L EDAEHHYHRFRLERN - EFFYNKKYAGUSKULAGUTGGEYAARAFCOFSY. RQKTPKGLYI DKGTLC 20. L EDAEHHYHRFRLERN - FFYNKK	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306622.1 XP_01306622.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01306428.1 XP_01542567.1 XP_0156184471.1 XP_01542567.1 XP_0156184471.1 XP_0156184471.1 XP_0156184471.1 XP_0156184471.1 XP_01562827.1 XP_01562827.1 XP_016607668.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_018016805.1 XP_0180371.51.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_01832751.1 XP_018327551.1 XP_01	L Y D S Y S E KO F Y D S T S E S D F Y D S T S E S D F Y D S T D S D D R D F Y D S D D R D F Y D S D D R D F Y S S A D D D Y Y S S N I T D Y Y S S N F N Y S S N F N Y S S N F N F Y S S I S S N F Y S S I S S N Y E S T D Y G Y Y E S T D Y G F Y S S N F D Y Y S S N F N Y S S N F N Y S S S S S Y S S S S S Y S S S S S Y S S S S S Y S S S S S S S S S S S S S S S S S S S		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D T K S I W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y A T N N K O A L W L Y K A T N D T N A W L Y K A T N D T N A W L Y K A T N D T A A W L F K A T N D T A A W L F K A T N D T A W L Y K A T G D E A I W L Y K A T G D E A I W L Y K A T G D E A I W L Y K A T G D E A I W L Y K A T G D E A I W L Y K A T G D E A I W L Y K A T G D E A W L Y K A T G D E A W L Y K A T G D E A W L Y K A T G D E A W L Y K A T G D E A W L Y K A T G D T A W L Y K A T G D T A W L Y K A T G D T A W L Y K A T G D T A W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W T K K T G D T A W W F K A T K F K F K K K A W W F K A T K K F	L DDAK TYLE NYPPW - S DWENKOAACROLLYE ET DSS VYKAVVEY TT WLPGOT - TPCGLAWRFKWOS LR 27. L DYARTLA - DFNV - PW - S DWENKOAACROLLYE ET DSS VYKAVVEY TT WLPGOT - TPCGLAWRFKWOS NR 27. L DAARSY DLLAW - DLSWSSKKPSCOLYLFE ET OSNOYKA I ON FFN WLPGOT - TPCGLAWRSKWOSNR 27. L EDAKTFV - DLLAW - DLSWSSKKPSCOLYLFE ET OSNOYKA I ON FFN WLPGOT - TPCGLAWRSKWOSNR 27. L EDAKTFV - DLLAW - ALGWODKKI ACOLILYE ET OSNOYKA I ON FFN WLPGOT - TPCGLAWRSKWOSNR 27. L HDANSLVD - AW - AW - ALGWODKKI ACOLILYE ET OSNOYKA I ON FFN WLPGOT - TPCGLAWRDKWOSNR 27. L HDAYD - VD - LW - AW - ALGWODKKI ACOLILYE ET OSNOYKA I ON FFN WLPGOS I TPCGLAWRDMWGANR 27. L HDAKE V D - PAPAW - ONSWODKT I VNOVMLYK I TNDEKY KAY ED SFNF WFPGOG I TPKGLAYRLGWOSLR 28. L ALAEKYFD - PAPAW - ONSWODKT I NNOVMLYK I TNDEKY KAY ED SFNF WFPGOS I TPKGLAYRLGWOSLR 28. L ADAEKYL D - DE PAW - ONSWODKT I NNOVMLYK I TNDEKY KAY EG SYT WWFPGOS I TPKGLAYRLGWOSLR 28. L ADAEKYL D - DE PAW - ONSWODKT I NNOVLIYK LTKKDKYK AY EG SYT WWFPGOS I TPKGLAYRLGWOSLR 28. L TAAEQHFD - PLPAA - DDSWODNKT AANOVLIYK LTKNDKYK AY EG SYT WWFPGOS I TPKGLAYRLGWOSLR 28. L TAAEQHFD - PLPAA - DDSWODNKT AANOVLIYK LTKNDKYK AY EG TFHY WFPGOS YT PKGLAHF FNROALS 28. L TAAEQHFD - PLPAA - DDSWODNKT AANOVLIYK LTKNDKYK AY EG TFHY WFPGOS YT PKGLAHF FNROALS 28. L TAAEQHFD - PLPAA - DDSWODNKT AANOVLIYK LTKNDKYK AY EG TFHY WFPGOS YT PKGLAHF FNROALS 28. L TAAEQHFD - PLPAA - DDSWODNKT AANOVLIYK LTKNDYK YA VE DT FHY WFPGOS YT PKGLAHF FNROALS 28. L TAAECHYFD - HP - AS - EF S WDDKT VANOVLIYK LTKNDYK YA VE DT FHY WFPGOS YT PKGLAHF FNROALS 28. L TAAECHYFD - HP - AS - EF SWDDKT VANOVLIYK LTKNDYK YA VE DT FHY WFPGOS YT PKGLAHF FNROALS 28. L TAAECHYFD - HP - AS - EF SWDKT VANOVLIYK LTKNDYK YA VE DT FHY WFPGOS YT PKGLAHF FNROALS 28. L TAAECHYFD - HP - AS - EF SWDKT VANOVLIYK LTKNDYK YA VE DT FHY WFPGOS YT PKGLAHF FNROALS 28. L TAAECHYFD - HP - AS - EF SWDKT VANOVLIYK LTKNDYK YA VE DT FHY WFPGOS YT PKGLAHF FNROALS 28. L EAAEKYFD - HP - AS - EF SWDKT VANOVLIYK LTKNDYK YA VE DT FHY WFPGOS YT PKGLAHF FNROALS 28. L EAAEK	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306622.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_01478650.1 XP_015562927.1 XP_015562927.1 XP_015562927.1 XP_0156007625.1 XP_016007625.1 XP_01801680.1 XP_01802846.1 XP_01832581.1 XP_018326851.1 XP_0183686184.1	L Y D S Y S E KO F Y D S T S E S D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D Y Y S S D C R D Y Y S S N I T D Y Y R S Y N I T D Y Y S N I T D Y S S I G R D N Y S S I G R D N Y S S I G R D N Y S S I G R D Y S S		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O T L W L Y Q A T N D N K S I W L Y Q A T N D N K S I W L Y Q A T N D N K S I W L Y Q A T N D K S I W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y Q A T N D N T S L W L Y A T N N K O A L W L Y K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A W W L F K A T N T A W W L F K A T N D T A W W L F K A T N T A W W L F K A T N D T A W W L F K A T N T A W W L F K A T N D T A W W L F K A T N D T A W W L F K A T N T A W W L F K A T N T A W W L F K A T N D T A W W L F K A T N D T A W W L F K A T N D T A W W L F K A T N T A W W L F K A T N D T A W W L F K A T N D T A W W L F K A T N D T A W W L K A T K N P Y A W W L K A T K N P Y A W W L K A T K N P Y A W W L K A T K N P Y A W W L K A T K N P Y A W W L K A T K N P Y A W W L K A T K N P Y A W W L K	LDAKTYLE         LV         ETGLAKTYLE         LV         ETGLAKTYLE         LV         ETGLAKTYLE         LV         ETGLAKTYLE         LV         ETGLAKTYLE         LV         LV         ETGLAKTYLE         LV         LV         ETGLAKTYLE         LV         LV         LV         LV         LV         LV         ETGLAKTYLE         LV	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306652.1 XP_01306423.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01304283.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01512607.1 XP_01552207.1 XP_01552207.1 XP_01563261.1 XP_01563261.1 XP_01563261.1 XP_01616806.1 XP_01616806.1 XP_01801680.1 XP_01803753.1 XP_01830753.1	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D F Y D S D D R D F Y D S D D R D F Y S S D E K S Y S S N F T D Y R S Y N T D Y R S Y N T D Y Y S S H G F ND N Y S S H G F ND N Y S S H G F ND N Y S S H G F ND F Y S S H G F ND Y Y S S H G ND Y Y S H S H H H H H H H H H H H H H H H		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G Q A W L Y R A T Q N G Q A W L Y R A T Q N G Q A W L Y R A T R N N G Q A W L Y R A T R N N G Q A W L Y Q A T N D Y K S L W L Y Q A T N D Y K S L W L Y Q A T N D T K S L W L Y Q A T N D T K S L W L Y Q A T N D T K S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y K A T N D T N A W L Y K A T N D S N A W L Y K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D T M A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W L F K A T N D S N A W W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L F K A T N D T M A W L Y K A T G D S A I W L Y K A T G D D T A W L Y K A T G D T A W L Y K A T G D T A W L Y K A T K L Y K T A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W L Y K A T K L Y K A W	L DAKTYLE NY PPMSFDWENKGAAGROLLY ETGUSSY VYAAV EYLTTW. PGGI-TPGGLAWR KWGSLR 72 L DYARTIA. DPN -WSYDWDNKGASGLLLY ETGUSSY VYAAV EYLTTW. PGGI-TPGGLAWR KWGSLR 72 L DAARSY VDLLAWDLSWSKKPSGQLYLF ETGSSNGYKAIGN YFNSWLPGGT-TPGGLAWRSKWGSNR 72 LEDAKTYVDLLAWDLSWSKKPSGQLYLF ETGSSNGYKAIGN YFNSWLPGGT-TPGGLAWRGMANG 82 LEDAKTYVDLLAWDLSWSKKPSGQLYLF ETGSSNGYKAIGN YFNSWLPGGT-TPGGLAWRGMANG 82 LEDAKTYVQLLAWDLSWSKKPSGQLYLF ETGSSNGYKAIGN YFNSWLPGGT-TPGGLAWRGMANG 82 LEDAKTYVQLLAWGLWAWDKIACQLLYE ETGSSNGYKAIGN YFNSWLPGGT-TPGGLAWRGMANG 82 LEDAKTYD - PAFAWGLWAWDKYICQLIF ETGSSNGYKAIGN YFNSWLPGGITFKGLAYRLGWSSNR 72 LEDAKTYD - PAFAWGLWAWDKYICQLIF ETGSSNYKAY DSFKFWFPGGGITFKGLAYRLGWSSLR 28 LALAEKYFD - PAFAWGLWAWDKYICNUYKITK ITKDEKYKAYEDSFKFWFPGGSITFKGLAYRLGWSSLR 28 LADAEKYLD - DEFAWGMSWDDKTIANGYLLYKLTKLTKKDKYKAYEDSFKFWFPGGSITFKGLAYRLGWSSLR 28 LADAEKYLD - DEFAWGMSWDDKTIANGYLLYKLTKLTKKDKYKAYEDSTKWFPGGSITFKGLAYRLGWSSLR 28 LTAAEGHFD - PLFAADMSWDNKTAANGYLLYKLTKDKYKAYEDSTKWFPGGSITFKGLAYRLGWSSLR 28 LTAAEGHFD - PLFAADMSWDNKTAANGYLLYKLTKDKYKAYEDTFKYWFPGGSVTFKGLAHPFNRGALS 28 LTAAEGHFD - PLFAADMSWDNKTAANGYLLYKLTKDKYKAYEDTFKYWFPGGSVTFKGLAHPFNRGALS 28 LEAAEKYFD - HP-AS - EFSWDDKTVANQYLLYKLTKDKYKAYEDTFKYWFPGGSSTFKGLAYSMWKGSLR 28 LTAAEGHFD - PLFAADMSWDNKTAANGYLLYKLTKDTKYKAYEDTFKYWFPGGSSTFKGLAYSMWKGSLR 28 LTAAEGHFD - HP-AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSSTFKGLAYSMWKGSLR 28 LTAAEGHYFD - HP-AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSSTFKGLAYSMWKGSLR 28 LTAAEGHYFD - HP-AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSSTFKGLAYSMWKGSLR 28 LTAAEGHFD - HP-AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSSTFKGLAYSMWKGSLR 28 LEAAEKYYFD - HP-AS - EFSWDDKTVANQYLLYKLTKLDTYKAYENTFKYWFPGGSSTFKGLAYSMWKGSLR 28 LEAAEKYYFD - HP-AS - EFSWDXKTVANGVLAGUTGGS VAAVAFCDTSY AND CGSTFKGLAYSMWKGSLR 28 LEAAEKYYFD - HP-AS - EFSWDXKTVANGVLAGUTGGS VAAFGCFSY - NOKTFKGLAYSMWKGSLR 28 LEAAEKYYFD - HP-AS - EFSWDXKTVANGVLAGUTGGS VAAFGCFSY - NOKTKGLAYSMKGLAYSKUGSLR 28 LEAAEKYYFD - HP-AS - EFSWDXKANGVGAGAGA LECACHYYYFYYCHYLGGSTFKGLAYSWKGLAYKGLAYSKUGSKG LEC	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306622.1 XP_01306427.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_015164427.6 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164471.1 XP_015164670.1 XP_016166065.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007662.1 XP_016007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007663.1 XP_007653.1 XP_0076	L Y D SY S EK S D F Y D ST D S S D D R D F Y D S T D S D D R D F Y S S A D D D Y Y S S N I T D Y S S N F N D F Y S S N S N F Y S S N S N S N F Y S S N S N S N F Y S S N S N S N S N F Y S S N S N S N S N S N S N S N S N S N		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T T S I W L Y Q A T N D T T S I W L Y Q A T N D T T S I W L Y Q A T N D T T S I W L Y Q A T N D T T S I W L Y Q A T N D T T S I W L Y Q A T N D T T S I W L Y Q A T N D T T S I W L Y K A T N D T T A W L F K A T N T T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T N D T T A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F	L DAAK TYLE NY PEW SFDUENKGAAGRULYE ETHGS SVYKAVVEYLETT TW. PGG1 - TPGLAWR KWGSLR 22         L DAAR SFV	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_01306652.1 XP_01306652.1 XP_01306652.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064279.1 XP_013064281.1 XP_013064281.1 XP_013064287.1 XP_013064285.1 XP_013064285.1 XP_013064285.1 XP_014786561.1 XP_014786561.1 XP_01478651.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01550207.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801682.1 XP_01801682.1 XP_01802465.1 XP_01802681.1 XP_01802681.1 XP_01803681.1 XP_01803681.1 XP_018368184.1 XP_018368184.1 XP_0183686184.1 XP_00282020.7 XP_0022820.7 XP_00	L Y D S Y S E KO F Y D S Y S E KO F Y D S Y S E KO F Y D S D D RD F Y D S D D RD Y Y S S N F D Y Y S S H G F DD N Y S S H G F DD Y S S H G F DD Y Y S S H G F D Y Y S S H G F D Y Y S S H G F D Y Y S H F G F D Y H S H F F F G F D Y H S H F F F G F D Y H S H F F F G F D Y H S H F F F F F F F F G F D Y H S H F F F F F F F F F F F F		S L F L H K A T Q K S N A M W L Y R A T Q K S N A M W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O T L W L Y R A T R N N G O T L W L Y Q A T N D H K S I W L Y Q A T N D H K S I W L Y Q A T N D H K S I W L Y Q A T N D H K S I W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y Q A T N D H T S L W L Y A T N N K O A L W L Y K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A A W L F K A T N D T A I W L Y K A T R L S A I W L Y K A T R L S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A I W L F K A T G D T A W L Y K A T K N P Y A A W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W Y K A T K N P Y A W W Y K A T K N P Y A W W Y K A T K N P Y A W W Y K A T K N P Y A W W Y K A	DDAK TY LE         - NY PW SFDWENKGAA GROLL YE ETGS VY KAVYEYIT TW. PGGT - TP GCLAWR FKWGSLR 27           LADAK TY LA         - OFN - PW SV DWENKGAS CILL YE ETHE BY KOAL VN YF DW. PGGT - TP GCLAWR FKWGSLR 27           LADAK SF V OLL AW OLS WS SKP SCOULY FE ETOS DG YKA I GN FFNNU PGGT - TP GCLAWR FKWGSLR 27           LNDAN SL VD AW AU GWDEKK I ACOLI LY E ETOS NG YKA I GN FFNNU PGGT - TP GCLAWR SKWGSNR 27           LNDAN SL VD AW AW ALG WDDKK I ACOLI LY E ETOS NG YKA I GN FFNNU PGGT - TP GCLAWR SKWGSNR 27           LALAK YFD PAPAW ALG WDDKKI ACOLI LY E ETOS NG YKA VE MW PGG SI TP GCLAWR DWGANR 27           LALAK YFD PAPAW GMS WDKTI VNOVULY XI TKDE KYKAV E D SFK FW PGG TP FGCLAWR DWGANR 27           LALAK YFD PAPAW GMS WDKTI VNOVULY XI TKDE KYKAV E D SFK FW PGG TP FGCLAWR DWGSLR 28           LADAK YL D - DE PAW GMS WDKTI ANOVLLYK LTK KDR YKAV E G SY MW PGG SI TP GCLAWR DWGSLR 28           LADA EXYL D - D PAW - GMS WDKTI ANOVLLYK LTK KDR YKAV E G SY MW PGG SI TP GCLAHRS DRAALS 28           LTAAE CHFD - P. PAP - DMS WDNK TAANOVLLYK LTK KDR YKAV E O TF KYW PGG SY TP GCLAHRS DRAALS 28           LTAAE CHFD - P. PAP - DMS WDNK TAANOVLLYK LTK KD XYKAV E D TF KYW PGG SY TP GCLAHP FNRGALS 28           LTAAE CHFD - P. PAP - DMS WDNK TAANOVLLYK LTK KD XYKAV E D TF KYW PGG SY TP GCLAHP FNRGALS 28           LTAAE CHFD - P. PAP - DMS WDNK TAANOVLLYK LTK KD XYKAV E D TF KYW PGG SY TP GCLAHP FNRGALS 28           LTAAE CHFD - P. PAP - DMS WDNK TAANOVLLYK LTK LT KD XYKAV E D TF KYW PGG SY TP GCLAHP FNRGALS 28           LTAAE CHFD - P. PAP - DMS WDNK TAANOV	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_013066172.1 XP_013066172.1 XP_01306427.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064282.1 XP_013064285.1 XP_013064285.1 XP_01513424.1 XP_01513424.1 XP_01513424.1 XP_015513382.1 XP_015513382.1 XP_015513382.1 XP_015513382.1 XP_015513382.1 XP_0156184471.1 XP_0156184471.1 XP_015618482.1 XP_015618482.1 XP_015618482.1 XP_015618482.1 XP_015618482.1 XP_015618482.1 XP_015618482.1 XP_016807628.1 XP_01801680.1 XP_01802681.1 XP_01802681.1 XP_00280380.1 X	L Y D S Y S E K D F Y D S T S E S D F Y D S T S E S D F Y D S T S E S D F Y D S D D R D Y Y S S N F T D F Y S S N S T D F Y S S N S S S S S F Y S S S S S S F Y S S S S S S S S F Y S S S S S S S S F Y S S S S S S S S F Y S S S S S S S S F Y S S S S S S S S F Y S S S S S S S S S F Y S S S S S S S S S S F Y S S S S S S S S S S S S S S S S S S		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O S I W L Y Q A T N D N G S I W L Y Q A T N D N K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D P K S I W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y Q A T N D N T S W L Y A T N N K O A W L Y K A T N D T N A W L Y K A T N D T A W L Y K A T N D S A I W L Y K A T N D S A I W L Y K A T N D S A W L F K A T N D S A W L F K A T N D S A I W L Y K A T G D S A I W L Y K A T G D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D D S A I W L Y K A T G D T A W L F K A T G D T A W L F K A T G D T A W L Y K A T G D D S A I W L Y K A T G D T A W L Y K A T G D D S A I W L Y K A T G D T A W L Y K A T G D D S A I W L Y K A T G D T A W L Y K A T G D T A W W L F K A T G D T A W W L F K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T G D T A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N P Y A W W L Y K A T K N F Y A W W T Y K A T K N F Y A W W T Y K A T K N Y A W W T Y K A T K N Y A W W T Y K A T K N Y A W W T Y K A T K N Y A W W T Y K A T K N Y A W W Y K A T K N Y A W W Y K A T K N Y A W W Y K A T K N Y A W W Y K A T K N Y A W W Y Y	L DOAK TY LE NY PFW - SFDWENKGAACGUL YE ETGS YY KAYYEYI TTWL PGG T TP GGLAWR FKWGER Z L ADARS FY DLLAW - DLSWGSKAPSGOL LY E ETGS DOYKA I ON FFNNU PGG T TP GGLAWR FKWGEN Z L ADARS FY DLLAW DLSWGSKAPSGOL LY E ETGS DOYKA I ON FFNNU PGG T TP GGLAWR FKWGEN Z L DAKT SY DLLAW DLSWGSKAPSGOL LY E ETGS DOYKA I ON FFNNU PGG T TP GGLAWR FKWGEN Z L DAKT SY	
XP_013066283.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066427.1 XP_013066427.1 XP_01306427.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_01304428.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184471.1 XP_015184618.1 XP_01518461.1 XP_01607628.1 XP_01607628.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01801680.1 XP_01803255.1 XP_0028235.1 XP_00208125.1	L Y D SY S E KO F Y D SY D S KO F Y D SY D D RD F Y D SY D D RD F Y D SY D D RD F Y D SY D D RD Y Y S SY D F KO Y Y S SY N I TD Y Y R SY N I TD Y Y R SY N I TD Y Y S S N F F D Y Y S S N F F D Y Y S S N F F D Y Y S S N F F D F Y S S N F F D F Y S S N F TD F Y S S N F TD Y S S N F TD F Y S S N F N F Y S S N F TD F Y S S N F TD F Y S S N F TD Y S S N F TD		S L F L H K A T Q K S N A M W L Y R A T Q K S N A W W L Y R A T Q N G O S I W L Y R A T Q N G O A W L Y R A T R N N G O S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D F K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T K S I W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T S L W L Y Q A T N D T T A W L Y K A T N D T T A W L Y K A T N D T T A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T R D T K A W L F K A T T D S N A W L Y K A T T D S N A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K A T K D K K A W L Y K	D DAK TY LE NY PW SF DW NK GA A GROLL YE ET DS VY A VY LY TTW. PGGT - TP GCL AWR FKWGSLR 27 LADARS FY SY DW NK GAS CILL YE ET HE DK YG AL UN YF DW PGGT - TP GCL AWR FKWGSLR 27 LADARS FY	

XP_021345983.1	YYNSSGYDDELCYSAAMLYKVTGKLEV LE	QAETIFI QSKPW	S F S W D D K T V G C Q V L L F E A T G K A V Y R L V A S F L Q S Y M A G G	DVTPK	GLAWRDQWGALR 296
XP_021361084.1	FYASSGYEDELCTAAAWLYKATNETHYLT	DAQQFYD TAY	A L S WD D K Q I S C Q V L L WE L T G K A K Y K S V Q H F L H N Y M P G G	N - T P C	GLVFRAKWGSLA 289
XP_021363481.1	AYGSSDYDDEMCWGGAWLYLATGEEKFLT	DSKKYHH GT - SW	I GQSWGNKDIGCKVLLYNVTGEDKYKDITGTFTDWMPGG	SLFKS	GLAFRDQWGSLR 294
XP_021363490.1	YYGSTGYNDELCEAAAWLHKATQEAKYLQ	D A K G F Y D T S W	ALSWDDKK ISCQLLLFEATKEAKYKNVESFVNSYKPGG	3 - T P C C	GLAWRDQWGSLR 289
XP_021363495.1	FYRSDAYEDELSWAGAWLYKVTNETKYLT	QAEQFYV GGAAW	GQSWDEKIASATILLYELTGKDVYLDIETTFTEWMPGG	TVTPK	GLAFRLQWGALR 295
XP_021363496.1	FYRSDAYEDELSWAGAWLYKVTNETKYLT	QAEQFYV GGAAW	GQSWDEKIASATILLYELTGKDVYLDIETTFTEWMPGG	TVTPK	GLAFRLQWGALR 295
XP_021363497.1	YYRSSAYEDEL CWEALWLYMAINKISYLI	DAKIYYV GP - DW	GQSWDDKGAGCQILLYN IIGEQIYKNIEAIFIDWLPGG	SISPK	SLAFRSQWGSLR 294
XP_021363498.1	YYRSSAYEDIL CWEALWLYMATNRISNLT	DAKIYYIAVGP - DW 5 A FUUYMK FDI FDD	GQSWDDRGAGCQILLYN IIGEQIYRNIEAIFIDWLPGG	SISPK	GLAFRSQWGSLR 296
XP_021924915.1	SYNSYSYDDEL YWGAAWLYDA TODYOLLD		N-EFFINKKVAGVQILLAQLIKQFEIMAAQAFCDFIV-NI		SLVTIDKFGILC 29
XP_021935277.1	EXASONY DEL VINGAVILLY RATION TO LT	CAEDLYN DEGLONW	INCOLTWOOK I SCI OVI LAK I KOTOS YKKVOCYCDY I TTSO	NKIPN	SLVYMNNVGPLR 299
XP_021941324.1	FYASGDYKDEL VWOAVWLYRAINDNINLI EXNEWSYGDEL GWOAAWLYRAINUT RULD	KAEQLYNDFGLGNW	EESWDDKKAGLOVIL AB ITNNOMYKSA EAECDY II I SU		SLVYIDQWGSLR 300
XP_021950548.1 XP_021956549.1	EXNSWSYGDELGWEAAWLYRATNYTRULD	DVATHWOEFGERPL	PMEESWDDKKAGLOVII AR ITNNOI YKAAEAFCDY I - ND	PTPK	SLVWLSEWGSLR 293
XP 021959948 1	EXTSWAFEDE LAW AAWLYRASGEERVIL	EAEALYAEENLOYV	- NDESWDGKASGVDILL VOTTGRAFYGKLISVCD- WVVDQ	APTPK	BLIELSEWGSLR 294
XP_021966378.1	EXDSSGYGDELAWSASWLYRATGEAGYKA		PSEASWDWKOAHVOVLLASVDGAARYRAAESECN, WVVNT	APTPK	31 VELSMWGSLR 294
XP_021967587.1	EYNSWSYGDELGWAAWLYRATNESAELT	DLDGEWTEENLGRP	G - EFGWDDKKAGLOILMAKLTNDAKYVPATOFCDWVV - DO	APSPLO	3MVFLSEWGALR 295
XP 022167267.1	YYEST DYGDEL TWAAAWLYKATKESQYLD	DAEYMYMKYRLERP	N - EFYYNKKVAGVQALLAELTNQSEYTAIQNFCNYNL - YV	KTPK	3LLYIEKSGTLC 297
XP 022167268.1	YYEST DYGDELT WAAAWLYKATKESQYLD	DAEYMYMKYRLERP	N - EFYYNKKVAGVQALLAEL TNQSEYTAI QNFCNYNL - YV	KTPK	GLLYIEKSGTLC 297
XP_022200004.1	YYESTEYGDELTWAAVWLYKATGENNYID	EAEYFYMRFRLRPN	EFYFNKKVAGVQVLFAQLTKRDDYVAARAFCDFSV - DV	KTPK	GLIVIEKIGTLC 296
XP_022328778.1	FYSSSEYRDELCVSAALLYKATQKPGYLS	DAISVYNSYSPSRW	AYDWDDKSVLCDVLLYD I TKETKYKATEGFVSSYMPGG	S-TPC	GLAWRLEWGSLR 294
XP_022329668.1	FYSSSEYRDELCVSAALLYKATQKPGYLS	DAISVYNSYSPSRW	AYDWDDKSVLCDVLLYD ITKETKYKATEGFVSSYMPGG	S-TPC	GLAWRLEWGSLR 294
XP_022336001.1	YYHSNSYKDELCTAAMELYKATNNFTYMQ	D A M S F F E P R S W	A F S W D D K N V Q C A L M L H R V Q G S P H Y R L V E S F L R S Y M P G G	E - T P C	GLAWRTAIGSLR 293
XP_022336153.1	FYRSVDYNDELCWAGAWLYKATNESKYLT	QAETYYV GA - SW	GQSWDDKTAGSQVLLYEETGKDKYKDIEATFQDWMPGG	SVSPK	GLAFRSQWGSLR 293
XP_022338490.1	YYGSTGYKDELCVAAMELYKATKDAKYLA	D A K A N F N G A - A W	I - ALSWDDNHVMCELLLYEETKDNRYKLVESFVRSYMPGG	S-TPC	GLAWRDQWGSLR 291
XP_022338491.1	YYHSNSYKDELCTAAMELYKATNNFTYMQ	DAMSFFP NSRSW	A F S WD D K N V Q C A L M L Y R V Q G S P H Y R L V E S F L R S Y M P G G	E - TPC	GLAWRTAIGSLR 294
XP 022341787 1	YYRSDKYEDEL TWOAAWLYRYTNDTKYLT	MAEOXYL. GP. DW		SVSPK	SLAFRSOWGSLR 204
XP_0233371211	EXPSWSYMDEL VWGGAWLYKATKETSYLD	RATOLYK	G - FESWDDKT I GGOVI I AOI TGD I KY I KY SDECNRMI - NSI	DVTPK	SMVELOOWGPLR 202
XP_023307121.1	YYASGDYKDEL VWGAVWLYRATNDNSYLT	TAFOLYNEEGLONW	INGGETWOSKI SGVEVI LAKI TNKOSYKKV SGYCTY I STSO	OKTPK	SIVYIDOWGSIR 300
XP 023704930.1	FYASGDYKDELVWGAVWLYRATNDSSYLT	TAEQQYNEFGLQYW	NVPFTWDAKSSAVEVVLAKVTGKONYKRVKSYCDSLI-SK	RTPR	GLTFLDTWGSLR 299
XP 023716596.1	YYEST DYGDELTWAAAWLYKAT DDPLELD	EAEHHYMKERLERP	N - EFFYNKKVAGVOILLAOLTROPEYMAAOAFCDFTV - DY	KTPK	SLVYIDKEGTLC 297
XP_023718476.1	SYETY SYYDELVWGATWLYRATGNESYLK	AEQLYKEFDLLYW	TTGFTWDKKISGIEVLLARMLGKPLYKRV ISYCGTLVDSQ	KRTPK	GLVYIEDYGSLS 298
XP_396791.3	YYESTDYGDELAWAAVWLFKATNDTMYLE	EAEHHYQHFHLERP	N - EFFYNKKVAGVQVLLAQMTGQPEYQAARAFCDFSV - YQ	KTPK	<b>SLLYIDKFGTLC 297</b>
AAA23086.1	YYKSWGYQDELVWGAYWLYKATGDATYLA	KAEAEYDKLGTENY	K W T I A W D N K Q F G T Y A L L A M E T G K Q K Y V D A N R W L D Y W V N G Q	KVSPG	3QAVLDSWGALR 300
AAD38027.1	HAANVAFIALYAAKLGIDGT - NQQWARGQ	G Q L <mark>L G</mark> N S R Q S F V V	G F G V N P P T R P H H R S S S C P D R P A D C S N - G L T N S G P N P Q T L W	SALVG	SPAQDGSYNDDR 395
AAF80584.1	MAANAVYLCLEAAKDGHNPTQYREFATQQ	IGYILGDTGKSYVV	G F G Q N Y P T H E S H R S S S C P D A P A A C D W N T Y N G N Q P N A H V L Y	GALVG	GPGQNDDYEDLR 398
AAF80585.1	MAANAAYLCLEAASAGLKPTEYRAFATEQ	IGYALGDTGKSFVV	F G V N P P S HE S H R S S S C P D A P A P C D W V T Y G S V D P N P H V L Y	GAIVG	SPGPNDEYDDQR 397
AAK12339.1	HAANAALIILQAADLGISADSYRQFAKKQ	IDYALGDGGRSYVV	C F G D N P P T H P H H R S S S C P D A P A V C D W N T F N S P D P N F H V L T	SALVG	SPDQNDNYVDDR 400
AAO61672.2	HAANVAFIALYAAKLGIDGT - NQQWARGQ	IGQLLGNSRQSFVV	G F G V N P P T R P H H R S S S C P D R P A D C S N - G L T N S G P N P Q T L W	GALVG	SPAQDGSYNDDR 395
AAU20853.2	HAANSALIALQAADLGINAATYRAYAKKQ	I D Y A L G D G G R S Y V V	C F G T N P P V R P H H R S S S C P D A P A V C D W N T Y N S A G P N A H V L T	SALVG	BDSNDSYTDAR 400
AAX92641.1	YAANAAFIALVSAKYNILASESEQFARSQ	I H Y M L G D A G R S Y V V	© F G N N P P Q Q P H H R S S S C P D Q P A E C D W D E F N Q P G P N Y Q I L Y	GALVG	SPDQNDQFEDLR 395
ABD24274.1	YASNMAMAALMAAEAGIHPE - YRHWAMCQ	THYALGDIGRSFVV	GFGKNPPVSPHHRSSSCPNLPVRCNMNYLHLDIPNTHMLC	SALVG	SPDSSDGYKDSR 394
ABD24275.1	Y S S NMAMAA L MAAEAG I H PE - Y R HWAMCQ	THYATGDIGRSFVV	GFGKNPPVSPHHRSSSCPNLPVRCNMNYLHLDIPNIHMLC	SAL VG	SPDSSDGYKDSR 393
ABD24276.1	Y S S NMAMAA L MAAEAG I H PE - Y RHWAMCQ		GEGKN PPVSPHHRSSSCPNL PVRCNMNYLHLDIPNIHMLC	SALVG	PDSSDGYKDSR 393
ABD24277.1	YASNMAMAALMAAEAGIHPE-TRHWAMCQ	LHYAL ODTORSEVV	GEGKNPPVSPHHRSSSCPNI PVRCNMNTLHLDTPNTHMLC	SALVG	SPDSSDGYKDSR 393
ABD24279 1	YASNMAMAAI MAAFAGIHPE-YRHWAMCO	LHYA GDTGRSEVV	GEGKNPPVSPHHRSSSCPNI PVRCNMNYLHI DTPNTHMLC	ALVG	PDSSDGYKDSR 393
ABD24280.1	YASNMAMAALMAAEVGIHPE - YRHWAMCO	HYANGDIGRSEVV	FGKNPPVSPHHRSSSCPNLPVRCNMNYLHLDTPNTHMLC	ALVG	PDSSDGYKDSR 393
ABD24281.1	YASNMAMAALMAAEAGIHPE - YRHWAMCQ	HYAL GDTGRSFVV	G F G K N P P V S P H H R S S S C P N L P V R C N M N Y L H L D T P N T H M L C	ALVG	PDSSDGYKDSR 393
AB 026609.1	Y DANVAFIALMAAEDGVE N RQWALSQ	LNYILGNKQLSFVV	G F G S K Y P L Q P R H G A S S C P D Q P A T C D W S N F H S P G P N P H V L K	SALVG	SPDGTDTYSDKR 389
ABV32557.1	Q S M N A V W V C L Q A A D A G V K T G E Y R S L A K K Q	LDYALGDAGRSFVV	G F G N N P P S HE Q H R A A S C P D A P A A C D W N T Y N G G Q S N Y H V L Y	GALVG	SPDANDYYNDVR 400
ACE75510.1	YAANAAFIALVSAKYNILASESEQFARSQ	I H Y M L G D A G R S Y V V	G F G N N P P Q Q P H H R S S S C P D Q P A E C D W D E F N Q P G P N Y Q I L Y	ALVG	PDQNDQFEDLR 395
ACE75511.1	YAANAAFIALVAAKYNILAAEAEQFARSQ	I H Y M L G D A G R S Y V V	G F G N N P P Q Q P H H S S S S C P D L P A E C D W D E F E N P G P N Y Q I L Y	ALVG	GPDQNDQYQDLR 395
ADB 12483.1	HAANAALIILQAADLGISADSYRQFAKKQ	IDYALGDGGRSYVV	G F G D N P P V R P H H R S S S C P D A P A V C D W N T F N S P D P N F H V L T	ALVG	SPDQNDSYEDDR 400
ADB82658.1	HASNAAFVILQAADLGISAVSYRQFAKKQ	I D Y A <mark>L G</mark> D G G R S L V V	<b>G</b> F G N N P P T H P H H A S S S C P D A P A V C D W S T Y S S P D P N F H V L T	GALVG	GPDVNDNYVDDR 400
ADB85440.1	EAVNAAFIAARAASLGIKSS-YNNWAQSQ	A D Y A L <mark>G</mark> S I G H S Y V C	G F G N N P P T H P H H G S S S C P V A P Q P C G W N E Y N S A D P N P Q T L F	GALVG	GPDAQDNYVDDR 391
ADB85441.1	HALGLAFIAI RGADLGINAT - N R N F A Q T Q	D Y A <mark>V G</mark> S Q G H S F V V	G F G T N P P T H A H H R A A S C E D P P A P C S Y D D M N K V D P N P H V I Y	GLVG	SPTDNDNWQDNR 396
ADB85442.1	EAMNVAFVAARAASLGLQDK - FNDWAQSQ	IDYALGSVGHSFVC	G F G E N P P T H P H H A S S S C P D A P A T C D W D T Y D S S E P N P H T L F	GALVG	GPDENDAYTDDR 393
AFD33365.1	HASNAAFVILQAADLGISADSYRQFAKTQ	IDYALGDGGRSFVV	C F G N N P P T H P H H G S S S C P P A P A V C D W S T Y S S P D P N Y N V L T	SALVG	SPDENDNYVDDR 400
AFQ98379.1	Y A A D A A F V A L L A A D A G A T Y R K W A V E Q	INYILGNKYISYQI	G F G S K Y P R S P H H R S A S C P D R P A P C T E A N L H A S G P S P Q V L V	SAIVG	SPSNDDSYKDNR 387
AFQ98380.1	Y A A N S A F V A L M A A D A G V T Y R K W A V E Q	INYILGNKYISYQI	G F G T K Y P R N P H H R S A S C P D I P A P C S E T N L H T A G P S P H I L V	GAIVG	DNDDSYKDNR 387
AG161069.1	YAANSAFAALVAADAGVIYRKWAVEQ	MNY TEGN KY ISYQ T	G F G T K Y P R NP H H RS AS C P D T P A P C S E T N L H T A G P S P H T L V	SAIVG	DNDDSYKDNR 387
AGP76398.1	HAANAAF IMLEAADLGL TASSY ROFACTO	DYALGDGGRSEVC	GEGEN PPTHPHHRSSSCPPAPATCDWNTEN I SDPN YN YPS	ALVG	BOONDNYEDDR 400
AGP76399.1	HAANAAF IMI FAADIGI TASSY ROFAL TO	DYALGDGGRSEVS	E E G S N P P T H P H H R S S S C P P A P A T C D W N T E N I S D P P Y N V P S	ALVG	PDONDNYEDDR 400
AGP76400.1	LAANAALIMLQAAGLGLNPTSYCQFAKTQ	DYILGDAGRSEVR	G F G N N P P T H P H H R S S S C P P L D G T Q D W R V L G L S C R F V C L S V	ALVG	PDLNDNYVDDR 400
AGP76401.1	LAANAALIMLQAAGLGLNPTSYCQFANAQ	IDYILGDAGRSFVR	G F G N N P P T H P H H R S S S C P P L D G T Q D W R V L G L S C R F V C V S V	ALVG	GPDLNDNYVDDR 400
AGP76402.1	HAANAAFIMLEAADLGLSASTYRQFAQTQ	IDYILGDAGRSFVC	G F G S N P P T R P H H R S S S C P P A P A A C D W N T F N S P D P T Y N V L S	ALVG	SPDQNDNYVDVR 400
AGP76403.1	HAANAAF IMLEAADLGLSASTY RQYAQTQ	I D Y I <mark>L G</mark> D A G R S F V V	<b>G</b> F G S N P P T R P H H R S S S C P P A P A D C D W N T F N S P D P T Y N V L S	GALVG	GPDQNDNYVDVR 400
AGP76404.1	HAANAAF IMLEAADLGLSASTYRQFAQTQ	I D Y I L <mark>G</mark> D A G R S F V C	G F G S N P P T R P H H R S S S C P P A P A A C D W N T F N S P D P T Y N V L S	GALVG	GPDQNDNYVDVR 400
AGP76405.1	LAANAALIMLQAAGLGLNPTSYCQFAKTQ	D Y   <mark>L G</mark> D A G R S F V R	G F G N N P P T H P H H R S S S C P P L D G T Q D W R V L G L S C R F V C L S V	S A L V G	SPDLNDNYVDDR 400
AGP76406.1	HAANAAFIMLEAADLGLSATSYRQFAQTQ	D Y A L <mark>G</mark> D G G R S L V C	C F G S N P P T R P H H R S S S C P P A P A T C D W N T F N I S D P Y H N V L S	GALVG	GPDQNDNYVDDR 400
AGP76407.1	HAANAAYIMLEAADLGLSATSYRQFAQTQ	I D Y A L G D G G R S F V C	G F G S N P P I R P H H R S S S C P P A P A T C D W N T F N I S H P N H N V L S	5 A L V G	SPDQNDNYVDDR 400
AGP76408.1	HAANAAFMULEAADLGLSASSYRUFAUTU	DYTEGDAGRSFVC	CFGSNPPTRPHHRSSSCPPAPATCDWNTLNSPDPNYNVLS	AL G	SPDQNDNYVDDR 400
AGP76409.1	HAANAAFIMLEAADLGLSASSIRGFAGIG	DYLLODAGRSFVC	CEGSNEET DELLESSSCEPAPATCOWNTLESEDEN YN VLS	ALVO	DONDNYVDDR 400
AGP76411.1	HASNAAFIMI QAADIGI SATSY ROIII TO	DYALGDAGESYVE	GEGSN PPT CPHHRSSS CPPAPAS CDWN TENS PYPSI NV IN	SALVG	SPDONDNYVDDR 400
AGP76412.1	HASNAALVILQAADLGLSADTYRQFAWKQ	IDYALGDGGRSLVV	G F G N N P P T H P H H A S S S C P P A P A V C D W S T S S S P G P N V H V L T	ALVG	SPDVNDSYADDR 400
AGP76413.1	HASNAALVILQAADLGLSADTYRQFAAKQ	IDYALGDGGRSKVV	G F G N N P P T H P H H A S S S C P P A P A V C D W S T S S S P G P N V H V L S	SALVG	SPDENDSYADDR 400
AGP76414.1	HAANAAFIILQAADLGISADSYRQFAYKQ	I D Y A L G D G G H S Y V V	G F G N N P P T H P H H G S S S C P P A P A V C D W S T Y S S P G P N V H V L T	GALVG	SPDENDNYADDR 400
AGP76415.1	HAANAAFIILQAADLGISADSYRQFAVKQ	I D Y A <mark>L G</mark> D G G H S L V V	G F G N N P P T H P H H G S S S C P P A P A V C D W S T Y S S P G P N V H V L T	GALVG	DENDNYADDR 400
AGP76416.1	HASNAALVILQAADLGISADTYRQFAYQQ	IDYALGDGGRSLVV	G F G N N P P T H P H H G S S S C P P A P A V C D W S T Y S S P G P N V N V L T	SALVG	SPDENDNYADDR 400
AGP76417.1	HAANAAFIMLQAADLGLSATSYRQFAQTQ	DYALGDGGRSFVR	OF GSN PPT RPHHRS SSCPP A PAPC DWN TFNS PYPN RNV I T	ALVG	DQNDYYVDDR 400
AGP76418.1	HAANAAF IMLAAAGLGLSATCYRQFAQTQ	DYALGUGGRSFVR	OF GSN PFI KPHHKSSSCPPAPARCDWNTFNSPYPNPNVI T	JAL VG	DUNDYYVDDR 400
AGP76419.1	HASNAAFVILQAADLGISAVTYRQFARKQ	I D Y ALLODGGRSLVV	OF GNN PPT HPH HGS SSCPD A PAVCDWS TYSS PGPN VHVL T	ALVG	DVNDNYADDR 400
AGP76420.1	HASNAAFTI LQAADLGI SAVSY RQFARKQ	I DYAL ODGGRSTVV	GEGNN PPTHPHHAS RSCPD A PAVCDWSTY SSPDPN VHVLT		SPDVNDNYADDR 400
AGP76421.1	HAANAAFIMLQAAGLGLNPTOYCOYAKTO	DYILGDAGRSEVO	GEGNN PPTHPHHRS SSCPPL DGTODWRVIGI SCREVCISV	SALVG	SPDENDNYVDDR 400
AGP76423 1	HAANAAVIMLQAAGIGINPTOYCHYANTO	DYILGDAGRSEVO	FGNNPPTHPHHRSSSCRPLDGTODWRVIGISCRFVCISV	ALVG	DENDNYVDDR 400
AGP76424.1	HAANAAFIMLQAAGLGLNPTQYCQFAKTQ	IDYILGDAGRSFVC	C F G N N P P T H P H H R S S S C P P L D G T Q D W R V L G L S C R F V C I S V	ALVG	SPDENDNYVDDR 400
AGP76425.1	HAANAAFIMLQAAGLGLNPTQYCQFAKTQ	IDYILGDAGRSFVC	G F G N N P P T H P H H R S S C P P L D G T Q D W R V L G L S C R F V C I S V	ALVG	SPDENDNYVDDR 400
AGP76426.1		DYILGDAGRSFVC	G F G N N P P T H P H H R S S C P P L D G T O D W R V L G L S C R F V C I S V	SALVG	SPDENDNYVDDR 400
AGP76427.1	HAANAAFIMLQAAGLGLNPIQYCQFAKIQ				The second se
AGP76428.1	LAANAAF IMLQAAGLGLNPIQYCQFAKIQ	I D Y A <mark>L G</mark> D G G R S F V C	C FGSN P P T H P H H R S N S C P P A P A T C D W N T F N I S D P N Y N V P S	SAL VG	SPDQNDNYEDDR 400
the second second second	HAANAAFIMLQAAGLGLNPIQYCQFAKIQ LAANAAFIMLQAADLGLTPSSYRQFAQTQ LAANAAFIMLQAADLGLTPSSYRQFAQTQ		G F G S N P P T H P H H R S N S C P P A P A T C D W N T F N I S D P N Y N V P S G F G S N P P T H P H H R S N S C P P A P A T C D W N T F N I S D P S Y N V P S	GALVG	SPDQNDNYEDDR 400 SPDQNDNYEDDR 400
AGP76429.1	HAANAAF IMLQAADLGL N PIQY CQFAK IQ LAANAAF IMLQAADLGL TPSSY RQFAQ TQ LAANAAF IMLQAADLGL TPSSY RQFAQ TQ LAANAAF IMLQAADLGL TPSSY RQFLQ TQ	I D Y A L <mark>G</mark> D G G R S F V C I D Y A L G D G G R S F V C I D Y A L <mark>G</mark> D G G R S F V C	G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P S Y N V P S G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P P Y N V P S	5 A L V G 5 A L V G	SPDQNDNYEDDR 400 SPDQNDNYEDDR 400 SPDQNDNYEDDR 400
AGP76429.1 AGP76430.1	HAANAAF IMLQAAGUGL NPIQYQQFAK IQ LAANAAF IMLQAADLGL TPSSYRQFAQTQ LAANAAF IMLQAADLGL TPSSYRQFAQTQ LAANAAF IMLQAADLGL TPSSYRQFLQTQ LAANAAL IMLEAAGLGL NPTCYQQFAKTQ	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y I L G D A G R S F V R	GFGSN PPTHPHHRSNSCPP APATCDWNTFN I SDPN YNVPS GFGSN PPTHPHRSNSCPP APATCDWNTFN I SDPS YNVPS GFGSN PPTHPHRSNSCPP APATCDWNTFN I SDPPYNVPS GFGNNPPTHPHRSSSCPPHLDGTQDWRVLGLSCRFVCLSV	5 A L V G 5 A L V G 5 A L V G 5 A L V G	SPDQNDNYEDDR 400 SPDQNDNYEDDR 400 SPDQNDNYEDDR 400 SPDQNDNYEDDR 400 SPDHNDNYVDGR 400
AGP76429.1 AGP76430.1 AGP76431.1	HAANAAF I MLQAAGLGL NP IQY CQFAK IQ LAANAAF I MLQAADLGL TPSSY RQFAQ TQ LAANAAF I MLQAADLGL TPSSY RQFAQ TQ LAANAAF I MLQAADLGL TPSSY RQFLQ TQ LAANAAF I MLQAADLGL TPSSY RQFQ TQ HAANAAF I MLQAADLGL TPSSY RQFAQ TQ	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y I L G D A G R S F V R I D Y A L G D G G R S F V C	GFGSN PPTHPHHRSNSCPPAPATCDWNTFN I SDPN YNVPS GFGSN PPTHPHHRSNSCPPAPATCDWNTFN I SDPSYNVPS GFGSN PPTHPHHRSNSCPPAPATCDWNTFN I SDPYNVPS GFGNN PPTHPHHRSSSCPPLDGTQDWRVLGLSCRFVCLSV GFGSN PPTHPHHRSSSCPPAPATCDWNTFN I SDPN YNVPS	5 A L V G 5 A L V G 5 A L V G 6 A L V G 5 A L V G	SPDQNDNYEDDR         400
AGP76429.1 AGP76430.1 AGP76431.1 AGP76432.1	HAANAAFIMLQAAGLGLNPIQYCGFAKIQ LAANAAFIMLQAADLGLTPSSYRQFAQTQ LAANAAFIMLQAADLGLTPSSYRQFAQTQ LAANAAFIMLQAADLGLTPSSYRQFAQTQ LAANAAFIMLQAADLGLTPSSYRQFAQTQ HAANAAFIMLQAADLGLTPSSYRQFAQTQ	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C	G F G SN P P T H P H H R SN S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T H P H H R SN S C P P A P A T C D WN T F N I S D P S Y N V P S G F G SN P P T H P H H R SN S C P P A P A T C D WN T F N I S D P Y Y N V P S G F G SN P P T H P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T H P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S	5 A L V G 5 A L V G	S P D Q N D N Y E D D R 400 S P D Q N D N Y E D D R 400 S P D Q N D N Y E D D R 400 S P D H N D N Y V D G R 400 S P D Q N D N Y E D D R 400 S P D Q N D N Y E D D R 400
AGP76429.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76433.1	HAANAAF I MLQAABGGGL NP I QY CQFAK I Q LAANAAF I MLQAADLGL T PSSY RQFAQ TQ LAANAAF I MLQAADLGL T PSSY RQFAQ TQ LAANAAF I MLQAADLGL T PSSY RQFAQ TQ HAANAAF I MLQAADLGL T PSSY RQFAQ TQ HAANAAF I MLQAADLGL T PSSY RQFAQ TQ HAANAAF I MLQAADLGV PSSY RQFAQ TQ	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y I L G D A G R S F V C I D Y I L G D A G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C	GFGSN PPTHPHHRSNSCPPAPATCDWNTFN I SDPN YNVPS GFGSN PPTHPHHRSNSCPPAPATCDWNTFN I SDPS YNVPS GFGSN PPTHPHHRSNSCPPAPATCDWNTFN I SDPPYNVPS GFGNN PPTHPHHRSSSCPPLDGTQDWRVLGLSCRFVCLSV GFGSN PPTHPHHRSSSCPPAPATCDWNTFN I SDPN YNVPS GFGSN PPTHPHHRSSSCPPAPALCDWNTFN I SDPN YNVPS GFGSN PPTHPHHRSSCPPAPALCDWNTFN I SDPN YNVPS		SPDQNDNYEDDR 400 SPDQNDNYEDDR 400 SPDQNDNYEDDR 400 SPDHNDNYVDGR 400 SPDHNDNYVDGR 400 SPDQNDNYEDDR 400 SPDQNDNYEDDR 400 SPDQNDNYEDDR 400
AGP76429.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76433.1 AGP76433.1	HAANAAF I MLQAABG US LNP I QY CQFAK TQ LAANAAF I MLQAADLG LT PSSY RQ FAQ TQ LAANAAF I MLQAADLG LT PSSY RQ FAQ TQ LAANAAF I MLQAADLG LT PSSY RQ FAQ TQ HAANAAF I MLQAADLG US PSSY RQ FAQ TQ HAANAAF I MLQAADLG I SAV SY RQ FAX KQ	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y I L G D A G R S F V R I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S L V V	GFGSNPPTHPHHRSNSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSNSCPPAPATCDWNTFNISDPSYNVPS GFGSNPPTHPHHRSNSCPPATATCDWNTFNISDPSYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSCPPAPATCDWNTFNISDPNYNVPS		SP DQNDNYEDDR 400 SP DQNDNYEDCR 400 SP DQNDNYEDCR 400 SP DQNDNYEDCR 400 SP DQNDNYEDCR 400 SP DQNDNYEDCR 400
AGP76429.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76433.1 AGP76433.1 AGP76434.1 AGP76435.1	HAANAAF I MLQAAGLGL NP IQY CQFAK IQ LAANAAF I MLQAADLGL TPSSY KQFAQTQ LAANAAF I MLQAADLGL TPSSY KQFAQTQ LAANAAF I MLQAADLGL TPSSY KQFAQTQ LAANAAF I MLQAADLGL TPSSY KQFAQTQ HAANAAF I MLQAADLGL TPSSY KQFAQTQ HAANAAF I MLQAADLGU TPSSY KQFAQTQ HAANAAF I MLQAADLGV PSSY KQFAQTQ HAANAAF I MLQAADLGU SY KQF KAKQ HAANAAF VV LQAADLG I SAVSY RQFAKKQ HAANAAF VV LQAADLG I SAVSY RQFAKKQ	I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S L V V	G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P Y N V P S G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P S Y N V P S G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P Y Y N V P S G F G S N P T T H P H H R S S C P P A P A T C D WN T F N I S D P Y Y N V P S G F G S N P P T H P H H R S S C P P A P A T C D WN T F N I S D P Y Y N V P S G F G S N P T D P H H R S S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P T D P H H R S S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P T D P H H R S S C P P A P A L C D WN T F N I S D P N Y N V P S G F G S N P P T H P H H R S S C P E A P A A C D WN T F N I S D P N Y N V P S G F G N N P Y V P H H A S S C P E A P A A C D WS T Y S S P V P N F H V L R G F G N N P P V H P H A S S C P E A P A A C D WS T Y S S P V P N F H V L R G F G N N P P V H P H A S S C P E A P A A C D WS T Y S S P V P N F H V L R		SP DQNDNYEDDR 400 SP DMNDTTCDTR 400 SP DMNDTTCDTR 400 SP DMNDTTCDTR 400
AGP76429.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76432.1 AGP76434.1 AGP76435.1 AGP76435.1 AGP76435.1	HAANAAF I MLQAAGLGL NP I QY CQFAK I Q LAANAAF I MLQAADLGL T PSSY RQFA O T Q LAANAAF I MLQAADLGL T PSSY RQFA O T Q LAANAAF I MLQAADLGL T PSSY RQFA O T Q HAANAAF I MLQAADLGL T PSSY RQFA O T Q HAANAAF I MLQAADLGL T PSSY RQFA O T Q HAANAAF I MLQAADLGU T PSSY RQFA C T Q HAANAAF I VLQAADLG I SAUSY RQFA K K Q HAANAAF VVL QAADLG I SAUSY RQFAK K Q HAANAAF VVL QAADLG I SAUSY RQFAK K Q HAANAAF VVL PAADLG I SAUSY RQFAK K Q HAANAAF VVL PAADLG I SAUSY RQFAK K Q	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S L V V I D Y A L G D G G R S L V V I D Y A L G D G G R S L V V I D Y A L G D G G R S L V V	G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P P T H P H H R S N S C P P A P A T C D WN T F N I S D P Y N V P S G F G S N P P T H P H H R S N S C P A A T C D WN T F N I S D P Y N V P S G F G S N P P T H P H H R S S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P P T D P H H R S S S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P P T D P H H R S S S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P P T D P H H R S S C P P A P A T C D WN T F N I S D P N Y N V P S G F G S N P P T D P H H R S S C P P A P A L C D WN T F N I S D P N Y N V P S G F G S N P P T D P H H R S S C P E A P A L C D WN T F N I S D P N Y N V P S G F G N N P V H P H H R S S C P E A P A C D W S T Y S S P V P N F H V L R G F G N N P P V H P H H A S S S C P E A P A A C D W S T F S S P V P N F H V L R G F G Y N P P V H P H H A S S C P D A P A A C D W S T F S S P V P N F H V L R G F G Y N P P V H P H H A S S C P D A P A A C D W S T F S S P V P N F H V L R		G         D         D         N         Y         D         R         4000         S         D         N         N         Y         D         R         0400         S         D         N         Y         D         R         0400         S         D         N         Y         D         R         0000         S         D         N         Y         D         N         Y         D         N         Y         N         N         Y         D         N         S         D         N         Y         Y         Y         N         N         N         Y         N         N         Y         Y         N <t< td=""></t<>
AGP76429.1 AGP76430.1 AGP76430.1 AGP76432.1 AGP76432.1 AGP76433.1 AGP76435.1 AGP76435.1 AGP76436.1 AGP76437.1 AGP76437.1	HAANAAF IMLQAAGLGL NP IQY CQFAKIQ LAANAAF IMLQAADLGL TPSSY RQFAOTQ LAANAAF IMLQAADLGL TPSSY RQFAOTQ LAANAAF IMLQAADLGL TPSSY RQFAOTQ HAANAAF IMLQAADLGL TPSSY RQFAOTQ HAANAAF IMLQAADLGL TPSSY RQFAOTQ HAANAAF IMLQAADLGL TPSSY RQFAOTQ HAANAAF IMLQAADLG IS AVSY RQFAKKQ HAANAAF VVL QAADLG IS AVSY RQFAKKQ HAANAAF VMLQAADLG LS AVSY RQFAKRQ HAANAAF VMLQAADLG LS AVSY RQFAKRQ HAANAAF IMLQAADLG LS AVSY RQFAKRQ HAANAAF IMLQAADLG LS AVSY RQFAKRQ	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S V V I D Y A L G D G G R S L V V I D Y A L G D G G R Y V V I V Y I M Y A L M Y A L M Y A V I M Y A L M Y A V I M Y	GFGSNPPTHPHHSNSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSNSCPPAPATCDWNTFNISDPSYNVPS GFGSNPPTHPHHRSNSCPPAPATCDWNTFNISDPYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTDPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTDPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTDPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSNPPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGNNPPVHPHASSSCPPAPACDWSTYSSPVPNFHVLT GFGYNPPVHPHASSSCPDAPACDWSTFSSPVPNFHVLT GFGYNPPVHPHASSCPDAPAACDWSTFSSPVPNFHVLT GFGNPPVHPHASSCPDAPAACDWSTFSSPVPNFHVLT		G         D         ND         Y         D         R         400         FD         ND         Y         FD         R         400         FD         ND         Y         FD         R         400         FD         ND         Y         FD         R         400         FD         ND         Y         FD         R         400         FD         ND         Y         FD         R         400         FD         FD         ND         Y         MD         FD         FD         FD         Y         MD         Y         MD         FD         FD         Y         MD         Y         Y         FD         FD         Y         Y         Y         Y
AGP76429.1 AGP76430.1 AGP76432.1 AGP76432.1 AGP76432.1 AGP76435.1 AGP76436.1 AGP76436.1 AGP76436.1 AGP76436.1 AGP76436.1 AGP76437.1 AGS32241.1 AIT11911.1	HAANAAF I MLQAAAGLGL NP IQY CQFAK IQ LAANAAF I MLQAADLGL TP SSY RQFAQ TQ LAANAAF I MLQAADLGL TP SSY RQFAQ TQ LAANAAF I MLQAADLGL TP SSY RQFAQ TQ LAANAAF I MLQAADLGL TP SSY RQFAQ TQ HAANAAF I MLQAADLGL TP SSY RQFAQ TQ HAANAAF I MLQAADLGL TP SSY RQFAQ TQ HAANAAF I MLQAADLG I SAYSY RQFAK TQ HAANAAF VVL QAADLG I SAYSY RQFAK KQ HAANAAF VVL QAADLG I SAYSY RQFAK KQ HAANAAF VMLQAADLG I SAYSY RQFAK KQ HAANAAF VMLQAADLG I SAYSY RQFAK KQ HAANAAF VMLQAADLG I SAYSY RQFAK KQ HAANAAF I MLQAADLG I SAYSY RQFAK KQ HAANAAF VMLQAADLG I SAYSY RQFAK KQ HAANAAF I GLYANK AQ V DS TMY - NWAK GQ	I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S F V C I D Y A L G D G G R S L V V I D Y A L G D G G R S L V V I D Y A L G D G G R S L V V I D Y A L G D G G R S L V V I D Y A L G D G G R S V V I D Y A L G D G G R S V V I D Y A L G D G G R S V V	G FGSN PPTHPHHRSNSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTHPHHRSNSCPPAPATCDWNTFNISDPYNVPS GFGSN PPTHPHRSSSCPPAPATCDWNTFNISDPYNVPS GFGSN PPTHPHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTDPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTDPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTUPHHRSSSCPPAPALCDWNTFNISDPNYNVPS GFGNN PPVHPHASSSCPAPALCDWSTFSSPVPNFHVLR GFGYN PPVHPHASSSCPAPAACDWSTFSSPVPNFHVLR GFGYN PPVHPHRSSSCPDAPAACDWSTFSSPVPNFHVLR GFGYN PPVHPHHASSSCPDAPAACDWSTFSSPVPNFHVLR GFGYN PPVHPHHRSSSCPDAPAACDWSTFSSPVPNFHVLR GFGYN PPVHPHHRSSSCPDAPAACDWSTFSSPVPNFHVLR GFGYN PPVHPHHRSSSCPDAPAACDWSTFSSPVPNFHVLT GFGYN PPVHPHRSSSCPDAPAACDWSTFSSPVPNFHVLR GFGYN PPVHPHRSSSCPDAPAACDWSTFSSPVPNFHVLR GFGSN PPVHPHRSSSCPDAPAACDWSTFSSPVPNFHVLT GFGSN PPVRPHRSSSCPDAPACDWSTFSSPVPNFHVLT		G         D         N         Y         D         R         4005           G         D         O         N         Y         D         R         4005           G         D         O         N         Y         D         R         4005           G         D         O         N         Y         D         R         4005           G         D         N         N         T         C         R         4005           G         D         N         N         M         D         R         4005           G         D         N         N         M         D         R         4005           G         D         N
AGP76429.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76432.1 AGP76434.1 AGP76435.1 AGP76436.1 AGP76436.1 AGP76436.1 AGP76437.1 AGS32241.1 AT11911.1	HAANAAF IMLQAAGLGL NP IQY CGFAKIQ LAANAAF IMLQAADLGL TPSSY ROFAOTQ LAANAAF IMLQAADLGL TPSSY ROFAOTQ LAANAAF IMLQAADLGL TPSSY ROFAOTQ LAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF VULQAADLGI SAVSY ROFAKKQ HAANAAF VULQAADLGI SAVSY ROFAKKQ HAANAAFVULQAADLGI SAVSY ROFAKKQ	I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S F V V I DYA L G D G G R S F V V I DYA L G D G G R S F V V I DYA L G D G G R S F V V I DYA L G D G G R S F V V I DYA L G D G G R S F V V I DYA L G S G R S F V V	E F G SN P P T H P H H S N S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T H P H H R S N S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T H P H H R S N S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T H P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G S N P T D P H H R S S C P A P A T C D WN T F N I S D P Y N V P S G F G N N P V H P H A S S C P A A A C D W S T Y S S P V P N F H V L R G F G N N P V H P H A S S C P D A P A C D WS T F S S P V P N F H V L T G F G N N P V H P H H A S S C P D A P A C D WS T F S S P V P N F H V L T G F G N N P V H P H H A S S C P D A P A A C D WS T F S S P V P N F H V L T G F G N N P V H P H H A S S C P D A P A A C D WS T F S S P V P N F H V L T G F G N N P V H P H H A S S C P D A P A A C D WS T F S S P V P N F H V L T G F G N N P V H P H H A S S C P D A P A A C D WS T F S S P V P N F H V L T G F G N N P V H P H H A S S C P D A P A A C D WS T F S P P N P H H V L T G F G N N P O N P H H H R S S C P D A P A C D WS T F S P P N P N P N L T G Y D S S S P E P H H R S S C C P D A P A C D WS T F S S P V P N F H V L T G Y D S S P E P H H R S S C C P S F A V C D W A T M S G P N P O T L Y		G         D         N         Y         D         R         400         S         S         D         N         N         Y         D         R         400         S         D         N         N         Y         D         R         400         S         D         N         N         Y         D         R         400         S         D         N
AGP76429.1 AGP76430.1 AGP76431.1 AGP76431.1 AGP76435.1 AGP76436.1 AGP76436.1 AGP76436.1 AGP76437.1 AG756436.1 AGP76437.1 AKG97349.1 AKC97349.1	HAANAAF I MLQAAGLGL NP IQY CQFAK IQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ HAANAAF I MLQAADLGL TPSSY RQFAQTQ HAANAAF I MLQAADLGL TPSSY RQFAQTQ HAANAAF I MLQAADLGU TPSSY RQFAXTQ HAANAAF I MLQAADLGU TPSSY RQFAXTQ HAANAAF I MLQAADLGU SY RQFAKKQ HAANAAF VU LQAADLGI SAXSY RQFAKKQ HAANAAF VU LQAADLGL SAYSY RQFAKRQ HAANAAF VI LQAADLGL SAYSY RQFAKRQ HAANAAF I LQAADLGU SY RQFAKRQ HAANAAF I MLQAADLGU SY RQFAKRQ HAANAAF I LQAADLGU SY RQFARRQ HAANAAF I LQAADLGU SY RQFARRQ HAANAAF I LQAADLGU SY RQFARRQ HAANAAL I LQAADLGU SY RQFARRQ HAANAAFI LU QAADLGU SY RQFARRQ HAANAAFI LU QAADLGU SY RQFARRQ HAANYAFI GU YANKAGV DST NY - NWAKGG HAANYAFI GU YANKAGV DST NY - NWAKGY HAANYAFI LALQAADLGU SY RQFARQ	I DYA L GD GG R F V C I DYA L GD GG R F V C I DYA L GD AG R F V C I DYA L GD AG R F V C I DYA L GD GG R F V V I DYA L GD GG R F V V I I DYA L GD GG R F V V I I DYA L DYA L GD GG R F V V I I DYA L DYA L GD F GG R F V V I I DYA L DYA L DYA L GD F G F V V I I DYA L DYA L DYA L GD F G F V V I I DYA L D	GFGSN PPTHPHHRSNSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTHPHHRSNSCPPAPATCDWNTFNISDPSYNVPS GFGSN PPTHPHHRSNSCPPAPATCDWNTFNISDPSYNVPS GFGSN PPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTHPHRRSSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPTHPHHRSSSCPPAPALCDWNTFNISDPNYNVPS GFGSN PPTHPHHRSSSCPPAPATCDWNTFNISDPNYNVPS GFGSN PPYHPHHASSSCPEAPACDWSTYSSPVPNFHVLR GFGNN PPVHPHASSSCPEAPACDWSTYSSPVPNFHVL GFGNN PPVHPHASSSCPEAPACDWSTSSPVPNFHVL GFGNN PPVHPHASSSCPEAPACDWSTFSSPVPNFHVL GFGNN PPVHPHRSSCPDAPACDWSTFSSPVPNFHVL GFGNN PPVHPHRSSCPDAPACDWSTFSSPVPNFHVLT GFGNN PPVHPHRSSCPDAPACDWSTFSSPVPNFHVLT GFGNN PPVHPHRSSCPDAPACDWSTFSSPVPNFHVLT GFGNN PPVHPHRSSCPDAPACDWSTFSSPVPNFHVLT GFGNN PPVHPHRSSCPDAPACDWSTFSSPVPNFHVLT GFGNN PPVHPHRSSCPDAPACDWSTFSSPVPNFHVLT GFGNN PPVHPHRSSCPDAPACDWSTSSSPVPNFHVLT GFGNN PPVHPHRSSCPDAPACDWSTSSSPVPNFHVLT		C D C N D N Y E D D R         400           S D C N D N Y E D D R         400           S D C N D N Y E D D R         400           S D C N D N Y E D D R         400           S D C N D N Y E D D R         400           S D C N D N Y E D R         400           S D C N D N Y E D R         400           S D C N D N Y E D R         400           S D C N D N Y E D R         400           S D D N N Y E D C R         400           S D D N N D T C D T R         400           S D D N N D T C D T R         400           S D D N N D T C D T R         400           S D D N N D T C D T R         400           S D D N N D T C D T R         400           S D D N N D T C D T R         400           S D D N N D T C D T R         400           S D D N N D T C D T R         400           S D D N N D Y E D R         400           S D C N N S Y E D R         303           S D S N D S Y T D A R         304           S D S N D S Y T D A R         304
AGP76429.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76434.1 AGP76434.1 AGP76435.1 AGP76436.1 AGP76436.1 AGP76437.1 AGS32241.1 AIC191349.1 AKC91349.1 AKV6371.1 AKV6371.1 AKV6371.1	HAANAAF I MLQAAGLGL NP IQY CQFAK IQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ LAANAAF I MLQAADLGL TPSSY RQFAQTQ HAANAAF I MLQAADLGL TPSSY RQFAQTQ HAANAAF I MLQAADLGL TPSSY RQFAQTQ HAANAAF I MLQAADLGU SAYSY RQFAK HAANAAF VU LQAADLG I SAYSY RQFAKKQ HAANAAF VU LQAADLG I SAYSY RQFAKKQ HAANAAF VU LQAADLG I SAYSY RQFAKKQ HAANAAF VU LQAADLG I SAYSY RQFAKRQ HASNAAF VU LQAADLG I SAYSY RQFAKRQ HASNAAF IGL VANKQY DSTNY - NWAKGQ HAANAAF IGL VANKGY DSTNY - NWAKGQ HAANAAF I LG LQAADLG I SAYSY RQFAKRQ HAANAFI LGU VANKGY DSTNY - NWAKGQ HAANAFI LGU AANKGY DSTNY - NWAKGQ HAANAFI LG SAAGGY DSTNY - NWAKGQ	I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G R S I V V I DYA L G D G R S I V V I DYA L G D G R S V V I DYA L G D G G R S V V I DYA L G D G R S V V I C S V G R S V V I C S V C S V V I C S V C S V S V S V S V S V S V S V S V	E F GSN P P T H P H H R S NS C P P A P A T C D WN T F N I S D P Y N V P S F G SN P P T H P H H R S NS C P P A P A T C D WN T F N I S D P Y N V P S C F G SN P P T H P H H R S SS C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T H P H H R S SS C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R S SS C P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R SS SC P P A P A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R SS C P A P A T C D WN T F N I S D P Y N V N P S G F G SN P P T D P H H R S SS C P A A A T C D WN T F N I S D P Y N V P S G F G SN P P T D P H H R SS C P A A A C D W T F N S P V P N F H V L R G F G N N P V H P H A S SS C P A A A C D W T F N S P V P N F H V L R G F G N N P V H P H A SS C P A A A C D W T F N S P V P N F H V L T G F G N N P V H P H A SS C P A A A A C D W T F N S P V P N F H V L T G F G N N P V H P H A S SS C P A A A A C D W T F N S P O P N F H V L T G F G N N P V H P H H R S S C P D A P A C D W T F N S P O P N F H V L T G F G N N P P V R P H H R S SS C P D A P A V C D W N T F N S P O P N F H V L T G F G N N P V R P H H R S SS C P D A P A V C D W T F N S P O P N F H V L T G F G N N P V R P H H R S SS C P D A P A V C D W T F N S A O P O L L Y G F G N N P V R P H H R S SS C P D R A P A V C D W N T Y N S A S A N P O L L Y G F G N N P V R P H H R S SS C P D R A P A V C D W N T Y N S A S A N P O L L Y G F G N N P V R P H H R S SS C P A P A A V C D W N T Y N S A S A N P O L L Y G F G N N P V R P H H R S SS C P A P A A V C D W N T Y N S A S A N P O L L Y G F G N N P V R H H R S SS C P A P A A V C D W N T Y N S A S A N P O L L Y G F G N N P V R H H R S SS C P D A A A A V C D W N T Y N S A S A N P O L L Y G F G N N P V R H H R S SS C P D H A A T C G W D F N M G N T N N H H H L E		G         D         ND         N         FD         R         400         FD         ND         N         ND         N         ND         N         ND         <
AGP76429.1 AGP76430.1 AGP76431.1 AGP76431.1 AGP76434.1 AGP76434.1 AGP76434.1 AGP76436.1 AGP76436.1 AGP76437.1 AGS2241.1 AIC19191.1 AKC91349.1 AKC91349.1 AKC91349.1 AKC91349.1 AKC91349.1 AKC91349.1 AKC91349.1	HAANAAF IMLQAAGLGL NP IQY CGFAK IQ LAANAAF IMLQAADLGL TPSSY ROFAOTQ LAANAAF IMLQAADLGL TPSSY ROFAOTQ LAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF VVLQAADLG IS AVSY ROFAKKQ HAANAAF VVLQAADLG IS AVSY ROFAKRQ HAANAAF VVLQAADLG IS AVSY ROFAKRQ HAANAAL I LQAADLG IS ATSY ROFAKRQ HAANAAL I LQAADLG IS ATSY ROFAKRQ HAANSAL I ALQAADLG IS ATSY ROYAKKQ NNANLL I CS EAAGMG IN TATY ROYAT VKRQ	I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S L V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I G Y V G D T G F S V I I	E F GSN P P T H P H H R S NS C P P A P A T C D WN T F N I S D P Y NV P S E F GSN P P T H P H H R S NS C P P A P A T C D WN T F N I S D P Y NV P S E F GSN P P T H P H H R S S C P P A P A T C D WN T F N I S D P Y NV P S G F G S N P P T H P H H R S S C P P A P A T C D WN T F N I S D P Y NV P S E F G S N P T D P H H R S S C P P A P A T C D WN T F N I S D P Y NV P S E F G S N P T D P H H R S S C P P A P A T C D WN T F N I S D P Y NV P S E F G S N P T D P H H R S S C P P A P A T C D WN T F N I S D P Y NV P S E F G S N P T D P H H R S S C P P A P A C D WN T F N I S D P Y NV P S E F G S N P T D P H H R S S C P P A P A C D W T T N S P Y P N F H V L R E F G N N P V H P H H A S S C P E A P A C D W T T S S P V P N F H V L R E F G N N P V H P H H A S S C P D A P A A C D W T F S S P V P N F H V L R E F G N N P V H P H H A S S C P D A P A A C D W T F S S P V P N F H V L R E F G N N P V H P H H A S S C P D A P A A C D W T F S S P V P N F H V L R E F G N N P V H P H H A S S C P D A P A A C D W T F S S P V P N F H V L R E F G N N P V H P H H A S S C P D A P A A C D W T F S S P V P N F H V L T E Y D S S S P E P H H R S S C P S C P A V C D W N T N S A S A N P O L V E F G N N P V A P H H R S S C P A A P A V C D W N T Y N S A S A N P O L V E F G N N P V A P H H R S S C P A A P A V C D W N T Y N S A S A N P O L V E F G N N P V A P H H R S S C P A A P A V C D W N T Y N S A S N P O L V V E F G N N P V A P H H R S S C P A A P A V C D W T Y N S A S N P N O L V V E Y G T N P V Y A H H R S S C P D H A A C G W D F F N G N T P N H V L E E Y S K Y P X A H H R S S C P D H P A T C G W D F N G N T P N H V L E		G         D         N         Y         D         R         400         50         N         N         Y         D         R         400         50         N         N         Y         D         R         400         50         N         N         Y         D         R         400         50         D         N         <
AGP76420.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76434.1 AGP76434.1 AGP76435.1 AGP76435.1 AGP76437.1 AG75437.1 AKC91349.1 AKV16371.1 AMH40355.1 AMH40355.1	HAANAAF I MLQAAGLGL NP IQY CQFAAT Q LAANAAF I MLQAADLGL TP SSY RQFAQT Q HAANAAF I MLQAADLGL TP SSY RQFAQT Q HAANAAF I MLQAADLGL TP SSY RQFAQT Q HAANAAF I MLQAADLG I P SSY RQFAQT Q HAANAAF I MLQAADLG I SAVSY RQFAK KQ HAANAAF I MLQAADLG I SAVSY RQFAK KQ HAANAAF I MLQAADLG I SAVSY RQFAK KQ HAANAAF VULQAADLG I SAVSY RQFAK KQ HAANAAF VULQAADLG I SAVSY RQFAK KQ HAANAAF I GL VANKG V STNY - NWAK GG HAANAAF I LGU ANKG V DSTNY - NWAK G HAANAAF I LU QAADLG I SAVSY RQFAK KQ HAANAAF I LU QAADLG I SAVSY RQFAK KQ HAANAAF I LI LOAADLG I SAT SY RQFAK KQ HAANAAF I LI LOAADLG I SAT SY RQFAK KQ HAANAX I LI LOAADLG I SAT SY RQFAK KQ HAANAXI LI LOAADLG I SAT SY RQFAK KQ NAANLU KS EAANLG I N TASY RAT V KXQ YAI N I AY VCLQAADLG I N ASY RAFAK KXQ	I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R S L VV I DYA L G D G G R S L VV I DYA L G D G G R S L VV I DYA L G D G G R S L VV I DYA L G D G G R S VV I DYA L G D G G R S VV I DYA L G D G G R S VV I DYA L G D G G R S VV I DYA L G D G G R S VV I DYA L G D G G R S VV I DYA L G D G G R S VV I G V V G D T G F S VV I I G Y L G G G R S VV I I G Y L G D G R S VV I I G Y L G D G G R S VV I I G Y L G D G G R S VV I I G Y L G D G G R S VV I I G Y L G D G G R S VV I I G Y L G D G R S VV I I G Y L G D G R S VV I I G Y L G D Z R S V I I G Y L G D Z R S V I I G Y L G D Z R S V I I G Y L G D Z R S V I I G Y L G D Z R S V I I G Y L G D Z R S V I I G Y L G D Z R S V I I G Y L G D Z R	G F GSN P P T H P H H RS NS C P P A P A T C D WN T F N I S D P Y NV P S G F G SN P P T H P H H RS NS C P P A P A T C D WN T F N I S D P Y NV P S G F G SN P P T H P H H RS NS C P P A P A T C D WN T F N I S D P Y NV P S G F G SN P P T H P H H RS SS C P P A P A T C D WN T F N I S D P Y NV P S G F G SN P P T H P H H RS SS C P P A P A T C D WN T F N I S D P N Y NV P S G F G SN P P T D P H H RS SS C P P A P A T C D WN T F N I S D P N Y NV P S G F G SN P P T D P H H RS SS C P P A P A L C D WN T F N I S D P N Y NV P S G F G SN P P T D P H H RS SS C P E A P A A C D WS T Y SS P V P N F H V L R G F G NN P P V H P H H A SS SC P E A P A A C D WS T Y SS P V P N F H V L T G F G N N P P V H P H A SS SC P D A P A A C D WS T Y SS P V P N F H V L T G F G N N P D V H P H A SS SC P D A P A A C D WS T F SS P V P N F H V L T G F G N N P D V H P H A SS SC P D A P A A C D WS T F SS P V P N F H V L T G F G N N P D V H P H A SS SC P D A P A A C D WS T F SS P V P N F H V L T G F G N N P D V H P H A SS SC P D A P A A C D WS T F SS P V P N F H V L T G F G N N P D V H P H H A SS SC P D A P A C D W N T F N S D P D N F H V L T G F G N N P D Y H P H H A SS SC P D A P A C D W N T F N S A G N A O V L T G F G N N P D Y H P H H R SS SC P D A P A C D W T F N S A G N A O V L T G F G N N P D Y H P H H R SS SC P A A P A V C D W N Y N S A G N A O V L T G F G N N P V H P H H R SS C P A A P A V C D W T Y N S A G N A O V L T G F G N N P V H P H H R SS C P A H P A C G W D F F N G N T P N H V L E G F G N N P Y Y A H H R SS C P D H P A A C G W D F F N G N T P N H V L E G F G N N P Y Y A H H R S SC P D H P A A C G W D F F N G N T P N H V L E G F G N N P T H A H R A A C P D K P A V C N W T Y S G G P N A H V L		G         D         N         Y         D         R         400         F         D         N         N         Y         D         R         400         F         D         N         N         Y         D         R         400         F         D         N         N         T         D         N         D         N         N         T         T         T         N         D         D         N         N         T         T         T         N         D         N         N         N         T         T         T         T         N         N         T         T         N         N         T         T         N         N         N         N         N         N         T         N         N         N
AGP76420.1 AGP76430.1 AGP76431.1 AGP76432.1 AGP76434.1 AGP76435.1 AGP76435.1 AGP76437.1 AGP76437.1 AG776437.1 AG7349.1 AKU16371.1 AMH40356.1 AMH40356.1 AMH40356.1	HAANAAF IMLQAABGUGL NP IQY CGFAKIQ LAANAAF IMLQAADGUGL TPSSY ROFAOTQ LAANAAF IMLQAADGUGL TPSSY ROFAOTQ LAANAAF IMLQAADGUGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL TPSSY ROFAOTQ HAANAAF IMLQAADLGL SAVSY ROFAKKQ HAANAAF VVLQAADLG ISALSY ROFAKKQ HAANAAF VVLQAADLG ISALSY ROFAKKQ HAANAAF VVLQAADLG ISAVSY ROFAKKQ HAANAAF VLQAADLG ISAVSY ROFAKKQ HAANAAFI ILQAADLG ISATY ROFAKKQ HAANAAL ILQAADLG ISATY ROFAKKQ HAANAAL ILQSAAGMO INTATY ROYAKKQ NAANLI ICSEAAMGO INTASY RAFAKKQ TTAY VUCQAADLG INAASY RAFAKKQ TTAY LUE IC LEFAALG IN AASY QAEAK KQ	I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R F V C I DYA L G D G G R S F V C I DYA L G D G G R S F V C I DYA L G D G G R S I V V I DYA L G D G G R S I V V I DYA L G D G G R S I V V I DYA L G D G G R S I V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I DYA L G D G G R S V V I G Y V G D T G F S V I I G Y L G D T G F S Y V I N V L G D T G F S Y V I N V L G D T G F S Y V	E F GSN P P T H P H H S NS C P P A P A T C D WN T F N I S D P Y N V P S F G S N P P T H P H H R NS C P P A P A T C D WN T F N I S D P Y N V P S G F G S N P P T H P H H R S NS C P P A P A T C D WN T F N I S D P Y N V P S G F G S N P P T H P H H R S S C P P A P A T C D WN T F N I S D P Y N V P S G F G S N P T D P H H R S S C P A P A T C D WN T F N I S D P Y N V P S G F G S N P T D P H H R S S C P A P A T C D WN T F N I S D P Y N V P S G F G S N P T D P H H R S S C P A P A T C D WN T F N I S D P Y N V P S G F G S N P T D P H H R S S C P A P A T C D WN T F N I S D P Y N V P S G F G S N P P T D P H H R S S C P A P A T C D WN T F N I S D P Y N V P S G F G N N P V H P H A S S C P A P A T C D WN T F N S P Y P N F H V L R G F G N N P V H P H A S S C P D A P A C D WS T F S S P Y P N F H V L T G F G N N P V H P H H A S S C P D A P A A C D WS T F S S P V P N F H V L T G F G N N P V H P H H A S S C P D A P A A C D WS T F S S P V P N F H V L T G F G N N P O Y H P H H S S S C P D A P A A C D WS T F S S P V P N F H V L T G F G N N P O Y A P H H R S S C P D A P A A C D W M T F N S A D P N O T L Y G F G N N P V R P H H R S S C P D A P A C D WN T Y N S A S A N P O L L Y G F G N N P V R P H H R S S C P D A P A C D WN T Y N S A G N P O L L Y G F G N N P V R P H H R S S C P D A P A C D WN T Y N S A G N N P O L L Y G F G N N Y Y A H H R S S C P D A P A C D W A T Y N S A G N N A D L Y G F G N N Y Y A H H R S S C P D H P A A C G W D F F N G N T P N H V L E G Y G S K Y P K Y A H H R S S C P D H P A A C G W D F F N G N T P N H V L E G F G N N P Y L A Y H R A S C P D L P P A C M W T Y S S C P D P N H V L E G F G N N P Y L A Y A H R S S C P D P A A C D W Y D Y S S P D P N H V L E G F G N N P Y L A Y A H R S S C P D P A A C D W Y S S P D P N P U V L G F G N N P Y Y A H H R S S C P D P A A C D W Y S S P D P N P U V L G F G N N P Y S H N A B N S C P C F O T N P Y Y S T N P N H Y L E G F G N N P Y S H N A B S C D C C P P A C D W N T Y S S C P D P N H Y L E G F G N N P Y S H N A B S C D C P A A C D W S S P D P N P U V L T Y S N O S D D N Y S S		B         D         N         Y         E         D         A         400           S         D         O         N         Y         E         D         A         000         S         D         N         N         Y         D         A         000         S         D         O         N         Y         D         A         000         S         D         N         N         Y         D         A         000         S         D         D         N         N         Y         D         A         000         S         D         N         N         Y         D         A         000         S         D         D         N         Y         D         A         000         S         D         N         N         Y         D         A         000         S         D         N         N         Y         D         N

AMH40362.1	STNNML LICLEAAERG INANTYRKESKKQ I DYM	GDSGFSYVV	I GNKYPKHAHHRSSSCPDHPAACGWTFFNSDVPNHHMVE	AIVG	GSNDSYIDTR	396
AMH40363.1	HATNVAFLLLQAVKLGIGG-TYTTMAKSQVDYAL	GSTGFSYVI	CLGSKYPTHAQHRSSSCPDAPASCDWN TYNGQQPNAHILY	AMVG	PDSNDGYTDVR	398
AMH40364.1	YAINIAYVCL RAADLG INTASY RAEAKKQ MNYAL	GDAGRSYVI	<b>G</b> F G H N P P T H A H H R A A S C P D K P A V C N W N T Y H G S Q P N A H V L Y	GLVG	PNNNDQFVDNR	395
AMH40365.1	TTAYLMEICLEAAALGANAEKYRAEAKKQIDYMV	GSTGFSYVV	S F G S K Y P L H V H H R A A S C P N P P A K C D W T Q F S S T E P N P H V L T	ALVGO	GADDDYTDVR	396
AMH40366.1	STAYVMQ ICLETAALGVNAATY RAEAKKQ INYML	GDTGFSYLV	6 F G S K Y P T H C H H R A A S C P D A P A T C G W A Q F S T P D P N P H V L T	GAL VGO	<b>F</b> GANDDYTDVR	396
AMH40367.1	NNANFLFICSEAANTGINTASYRAVVKKQIGYFL	GDTGFSYVI	S Y G S K Y P K Y A H H R S S S C P D H P A K C G W D F F N S N T P N H H V L E	GALVGO	<b>F G A N D D Y T D T R</b>	396
AMH40368.1	YALNIAYVCLQAADLGINTASYRAEAKKQMDYA	GDAGRSYVI	5 F G N N P P T H A H H R A A S C P D K P A V C N W D T Y H G S Q P N A H V L Y	GLVG	PNKSDIFVDDR	395
AMH40369.1	STNNVLLICLEAAQRGVNAESYRKESKKQLDYML	GDSGSSYVV	S I G K N Y P T H A H H R S S S C P D H P K P C G W T F F N S D L P N Y H M V E	GAIVGO	<b>P</b> GANDTHVDTR	396
AMH40370.1	NNANMMVICSEAANMGIETTTYRPLIKNQVGYY	GDTGFSYIV	G Y G T N Y P K Y A H H R S S S C P D H P A T C G W D F F N S N T P N T H A V E	GALVGO	<b>F</b> GANDDYTDTR	396
AMH40371.1	TTAYLMEICLEAAALGAN PEKYRAEAKTQIDYMV	GSTGFSYVV	E F G D K F P V H V H H K A A S C P D P P A K C D W T Q F S S P D P N P Q Q I T	GALVGO	G A N D D Y T D V R	396
AMH40372.1	HATNIAYLLLQAARLGIGDASYRKMAQSQVDYAL	GSTGFSYVI	E G S K Y P T H A Q H R S S S C P D A P A T C D W N T Y S G Q Q P N A H I L Y	SAMVGO	<b>B</b> D S N D G Y T D V R	399
AMH40373.1	STSYVMQVCLETAALGVNTAKYRAETKKQIDYML	GDTGYSYVV	E F G S K Y P T H C H H R A S C P D K P A T C D W N Q F S T T E P N P H V L T	GALVGO	G A D D D Y T D L R	396
AMH40374.1	HATNVAFLLLQAVKLGIGS - SYTAAAKSQVDYAL	GSTGFSYVI	G N G N S Y P T H A Q H R S S S C P D A P A T C D W N T Y N G N Q P N A H I L Y	SAMVGO	D S N D G Y T D V R	398
AMH40375.1	Y A V N Y A F V C L Q A A D L G I N T Q S Y R T E A K K Q L D Y A L	GDAGRSYVI	<b>6</b> F G N N P P T H A H H R A A S C P D A P A V C N W D T Y H G S Q P N A H V L Y	GMVG	AKDDSFKDDR	395
AMH40376.1	SNANAML VCSEAANLG IEASSY RPL I KQQ V GY YV	GDTGFSYVV	6 F G N K Y P R Y A H H R S S S C P D H P A K C G W D F F N S N T P N H H I V D	SALVGO	SP N A S D Y Y F D T R	396
AMH40377.1	N N A N F L L V C S E A A D MG I N T A H Y R D V V K K Q I G Y Y U	GDTGFSYVI	G F G N K Y P K F A H H R S S S C P D H P A Q C G W N F F G S S T P N H H V L E	GLVG	GKDDSYKDVR	396
AMH40378.1	TTQYLMEICLETAALGVNAEKYRAEAKKQVDYIL	GDTGFSYVV	S F G S K Y P T H V S H R A A S C P D L P A T C D W T Q F S T T Q P N P Q V L T	SALVG0	G V N D D Y A D V R	396
AMH40379.1	HACNVAYLLLQASRLGIGDSSYTKMAKSQIDYAL	GSTGFSYVI	S V G S N Y P T H A Q H R S S S C P D A P A K C D W S T Y S G Q Q P N A H I L Y	SAMVGO	DSNDGYKDER	399
AMH40380.1	YAINMAYVCL RAADIG IN TESYRTEAKKQIDYAL	GDAGRSYVV	B F G N N P P T HA H H R A A S C P D A P A V C N M D T Y H G S Q P N V H V L Y	GMVG	PNKSDIFTDDR	395
AMH40301.1	STNNMI LICI EAAEBOVNAATV DKESKKO I DYM	GDIGISILV	F G S K T F T H C H H R A A S C F D A F A K C D W T Q T S T T S F N F Q V L T		GANDDYIDVR	390
AMH40382.1	STANGLI UCSEAABROVNAATTRESKKQIDTMI	GD SGF STVV	TONK TPTHAHHRSSSCPDHPAQCOWSFFNSDAPNHHMVE		GSKDATVDTR	390
AMH40383.1	SNANFLLVGSEAANMGINTASTRAVVKKQIGTP	GUTGESTVV	T G S N T P K T A H H R S S G P D H P A T G G W D F P N S N T P N H H V L E		BOSNDDTTDTR	390
AMI140304.1	NTAYVNOICI ETAALGVNADTV DAEAKKO INVNI	GDTGESVVV	EGTN VETHCHHEAASCENA BATCEWTOVSTTEAN BHVI T		GANDOYTOVR	206
AMH40386 1	VALNIAVVCI RAADIGINTASY RAFAKKOMDYA	CDAGRSYVI	Y GNN P PT HA H H RAAS CPD K PAV CNWN TY HG SOPN A HVL Y	GI VG	DKNDOFTDNR	305
AMH40387.1	TTAYI MELCI FAAALGANAFOY RAFAKKO I DYMU	STOFSYVV	YORKYPI HVHHKAASCPD PPA TCDWTOESS TAPN PHVIK	ALVO	BGANDDYEDVR	306
AMH40388 1	NNANI LI LOS FAAKMG LETDTY RPI LKNO LGYY	OTGESYVV	Y G S N Y P K NA H H R S S S C P D H PO T C G W D E F N S D T P N H H V I F	ALVG	PSSTDAYTDTR	396
AMH40389 1	NNANELEVCSEAANAG INTASY RAVVKKO LGYE	ODTGESYVI	Y GSK Y PKYAHHRSSSCPDHPOTCGWDEENGNTANHHVI E	GL NG	RSSTDAYTDTR	306
AMH40390 1	YA I NMAYVCI RAADIG I NTDTY RTAAKKO I DYA	GDAGRSYVI	EGNN PPT HAHHRSASCPD A PAVCNWD TY HOSOPNAHVI Y	GMIG	PNKSDIYVDDR	395
AMH40391.1	SVNNLLLICLEAAERG INADSYRKEAKKQIDYM	GDSGFSYVV	I G N N Y P T HA H H R S S S C P D H P A E C G W S F F N S D L K N Y H M V E	AIVGO	GANDSYIDTR	396
AMH40392.1	NIANAVFICLEAASLGINADTYRSGAKTQIDYML	GDGGHSYVV	YGDN Y PT HAHHRSSSCPD A PAVCDWN L FNSDTPN YHVL Y	ALVG	PGSNDDYTDTR	396
AMH40393.1	YAANLAFIFMQAADLGINADTYRAEAKKQIDYIL	GDGGRSYVI	SYGDN Y P T HA H H R S S S C P D A P A T C D WN T Y S G T Q P N V H V L Y	GMVG	PDKNDAFEDLR	395
AMH40394.1	E SANAMFICL KAAELG ISTDSYRTEAKTQIDYML	GDSGRSYIV	SYGTN PPT HPH HRSSSCPD APA TCD WTQYSTTDPN PHVL F	ALVG	PGANDDYEDVR	396
AMH40395.1	HAANVAFICLQVADS EKYRDFVKSQIHYML	GESGRSYVV	F G L N Y P T Q P Y H A A S S C P D L P E P C G W K D F H R R G P N S Q I L Y	SALVSO	PDENDNYKDDR	392
ANR02619.1	YAANSAF IALLAADQG VLPTQS RNFARNQ INYML	GSTGRSFVV	<b>G</b> F G N N P P Q R P H H R A S S C P D L P A N C G W D Q A S D P A P N P Q I L N	ALVGO	PDAQDNYNDDR	391
AOV94250.1	NAANSMFICLEAANLGIDADNYRTQSKSQIDYML	GDGGHSYVV	G F G D N Y P K Y A H H R S S S C P D A P A V C D W N F F G S D Q P N A H V L E	ALVGO	PGQNDDWTDTR	396
AOV94251.1	HAANAAFILMRAADLGVGGT DFAKKQIGYM	GDTGRSYVV	GYGENYPKHAHHRSSSCPNLPAPCGWDDYNSDGPNPQILY	ALVGO	P D R N D Q Y T D V R	396
AOV94252.1	Y T A N A V F I A L R A A D L G I G N E S Y R S F A Q S Q V G Y M	GDTGRSFVI	S F G E N Y P T H E H H R S S S C P D A P A T C D W M Q Y G M N D S N Y H I L Y	SALVGO	PDQDDQYTDIR	399
AOV94253.1	HAANVAFICLEAADP LKYQEFARSQIHYM	GDAGRSYVV	S F G N N P P Q Q P H H A A S S C P D V P E P C G W Q E Y Q H D G P N H Q L L K	SALVSO	PDENDYYLDKR	391
AOV94254.1	Y S A N A V F I A M R T A D L G I D G D S Y I S F A K S Q I G Y M L	GDTGRSFVV	S F G E N F P T H A H H R A S S C P D I P I P C D Y H Q Y N A T T P N P H M L Y	ALVGO	PDENDQYTDVR	399
AOV94255.1	HAGNAAFITLRAAELGVGNDDYINFARQQLGYML	GDTGRSYVV	S F G Q N Y P T H P H H R A S S C P D A P A T C N W D Q F N T N D P N P Q L L Y	GALVGO	PDQNDQYTDIR	399
AOV94256.1	Y S A N A V F I A M R T A D L G I D G D S Y I S F A K S Q I G Y M B	GDTGRSFVV	FGENFPTHAHHRASSCPDIPIPCDYHQYNATTPNPHMLY	ALVGO	PDENDQYTDVR	399
APF29556.1	HAANVAFIALVANKAGVDSANY - AFAKQQIDYIL	GSTGRPFVV	GYGNG - PERPHHRSSSCPDLPEACDW - A I NQSGANPQTLY	ALVG0	S P G Q D G S Y V D D R	393
APF29557.1	AAANVAFIALLANKAGVDSANY - AFAKQQIDYIL	GSTGRPFVV	GYGNG - PERPHHRSSSCPDLPEACDW - A I NQSGANPQTLY	GALVGO	<b>P</b> G Q D G S Y V D D R	392
APF29558.1	HAANVAFIALVANKAGVDSTNY - EWGKQQIDYIL	GSTGRSFVV	GYGSG - PERPHHRSSSCPDMPEACNW - AMTQSGAN PQTLY	GALVGO	<b>F</b> GQDGSYTDDR	392
APF29559.1	HAANVAFIALMANKVGVDSTNY - NWGKQQIDYL	GYSGRPFVV	GYGSG - PERPHHRSSSCPDMPEYCNW - AMTQGGPNPQTLY	GALVGO	<b>F</b> GQDGSYTDDR	382
APF29560.1	HAANVAFVALMANKLGVDDT - N RQWAETQINYIL	GSSGRSFVV	EFGSNYPVRPHHRASSCPDIPEYCSWAQYQ - EDANPQTLW	SALVGO	PDQNGNYVDDR	394
APF29561.1	HAANVALLHY FAAKNGLDEA - N RQWATGQ I N L I L	GDTGRSYMV	BYGVDPPQQPHHRSSSCPSLPGSCGW-AVSQEGPNPQPLY	GALVGO	D E N G V Y V D D R	395
ATS91249.1	YASNMAMAALVAADAGLRPE - YRRWAMCIIHYAI	GDTGFSYLI	S F G S K Y P H S P H H R S S S C P N P P A P C G S Y V M S M K E P N V H T L Y	SALVG0	GQQDDYKDER	393
BAA31326.1	HAANSAL IA LQAADLG INAATY RAYAKKQ IDYAL	GDGGRSYVI	S F G T N P P V R P H H R S S S C P D A P A V C D W N T Y N S A G P N A H V L T	SALVG0	DSNDSYTDAR	400
BAA33708.1	HAANAAF IMLEAAELGLSASSY RQFAQ IQIDYA	GDGGRSFVC	G F G S N P P T RP H H R S S C P P A P A T C D W N T F N S P D P N Y H V L S	SAL VGU	DONDNYVDDR	400
BAA33709.1	HAANAAF IMLEAADLGLSASST RUFAUTUTDTA	GDGGRSFVC	SFGSNPPTRPHHRSSSCPPAPATCOWNTFNSPDPNTNVLS	AL GO	DONDNYVDDR	400
BAA34050.1	HAANSAL TALQAADLG INAATY KAYAKKQ IDYA	GDGGRSTVI	6 F G I N P P V RP H H RS S S C P D A P A V C D WN I Y N S A G P N A H V L T		DENDSYTDER	400
BAB40693.1	HAANSAL TALQAADLG INAAST RUTAKKQ IDTAL	COGGRSTVV	5 F G T N P P V RP H H RS S S C P D A PA A C D WN T Y NS A G P N A H V L T		P D S N D S Y I D S R	400
BAB40695.1	HAANSAL TALQAADLG INAASY ROTAKKO IDYAL	GDGGRSTVV	6 F G T N D D V D D H H D C C C D D A D A A C D WN T Y N C A C D N A H V L T		DONDOTIDOR	400
BAB40696.1	HAANSAL TALQAADLG INAASY RQTARRQID TAL	GDGGRSTVV	E C TN D DV D D H D C C C D A DA A C D WN TY NG A C D N A HVI T		DONDOTIDOR	400
BAC67186 1	VAANSAL VAADAGVT V RKWAVEOMNY I	GNKYISYOU	EGTKYPRNPHHRSASCPD I PAPCSETNI HTACPSPHII V		DONDOSYKONR	397
BAD01504.1	YAANSAFAAL VAADAGVT YRKWAVEQMNY I	GNKYISYQI	FGTK Y PRNPHHRSASCPD I PAPCS ETN LHTAGPS PHILV	AIVG	PDNDDSYKDNR	387
BAD44734 1	YAANSAFAAL VAADAGVT YRKWAVEOMNOL	GNKYISYOL	EGTK VPRNPHHRSASCPD I PAPESETNI HTAGPS PHILIV	AING	PDNDDSYKDNR	387
BAE38757 1	YAANAAF JAJAAAFE GDO EKNEAMSO INYI	GNKLISYEL	FGSK YPOHPHHRGSSCPTTTTOC SIGDTGPNPNI I K	GING	PDNSDNYDDRR	383
BAH22180.1	YAANSAFIALVAAKYNILTAEAQNFARSQIHYMI	GDTGKSFVV	F G N N P P Q Q P H H R S S S C P D Q P N P C D W D E Y N N P G P N Y Q I L Y	ALVG	PDQNDNYNDAR	395
BAK20401.1	Y SANSAF JALLAAEEGVLTSQARTFARAQLDYM	GSTGRSEVV	E G T N P P L R P H H R A A S C P D M P A S C G W D Q A S D P A P N P Q V L D	ALVG	PDDDDNYNDDR	391
BAL60587.1	HAANVAFVMTRAAHLGINAA - NKAFALRQLHYI	GNPQRSFVV	F G N N P P Q R P H H R S S S C P D P P A S C D N - A M N N S G P N P H T L F	AFVG	PANNDNYTDDR	392
BAM14716.1	YAANAAFIALVSAKYNILASESEQFARSQIHYM	GDAGRSYVV	G F G N N P P Q Q P H H R S S S C P D Q P A E C D W D E F N Q P G P N Y Q I L Y	ALVGO	PDQNDQFEDLR	394
BAS18736.1	YAGNAAFAALVAAEDGIQTR - YRSWGVEQINYIL	GNHHGSFQI	G F G T K Y P R N P H H R A A S C P D R P A P C S H E Q L D A P G P S P Q L L V	AIVGO	PDAQDNYVDDR	369
BAT21145.1	YAANAAFIALVSAKYNILASESEQFARSQIHYML	GDAGRSYVV	F G N N P P Q Q P H H R S S S C P D Q P A E C D W D E F N Q P G P N Y Q I L Y	ALVG	PDQNDQFEDLR	394
BAU45387.1	YAGNAAFAALVAAEDGIQTR - YRSWGVEQINYI	GNHHGSFQI	F G T K Y P R N P H H R A A S C P D R P A P C S H E Q L D A P G P S P Q L L V	GAIVGO	PDAQDNYVDDR	369
BAU45388.1	YAGNAAFAALVAAEDGIQTR - YRSWGVEQINYIL	GNHHGSFQI	S F G T K Y P R N P H H R A A S C P D R P A P C S H E Q L D A P G P S P Q L L V	SAIVGO	PDAQDNYVDDR	369
BAX09271.1	YAADAAFIALVAAEDGINAN - YRKFAVEQINYI	GNHHGSYAI	GYGSKYPL RPHHRGSSCPSH - GHCNK DAHGPNPHVLL	SAVVGO	PNADDSYTDVR	366
CAD54726.1	MISNAVYICLQAADLSINSDTYRQFAKSQLNYIL	GDAGRSFVV	8 F G N N P P T H P H H R S S S C P D A P A T C D W N N Y N S A D P N P H V L Y	SALVGO	PDSNDNYEDLR	399
CAD54727.1	MISNAVYICLQAADLSINSDTYRQFAKSQLNYI	GDAGRSFVV	<b>G</b> F G N N P P T H P H H R S S S C P D A P A T C D W N N Y N S A D P N P H V L Y	SALVGO	PDSNDNYEDLR	399
CAD54728.1	MISNAVYICLQAADLSINSDTYRQFAKSQLNYI	GDAGRSFVV	8 F G N N P P T H P H H R S S S C P D A P A T C D W N N Y N S A N P N P H V L Y	GALVGO	PDSNDNYQDLR	399
CAD54729.1	MAANAAYICAVAADVGISSTAYRQFAKTQINYI	GDAGRSFVV	SYGNN PPTHPHHRSSSCPD A PATCDWNNYNS ANPN PHVL Y	SALVGO	DSNDNYQDLR	400
CAD54730.1	MAANAAY I CAVAADVG I S STAY RQ FAK TQ I NY I	GDAGRSFVV	SYGNN PPTHPHHRSSSCPD A PATCDWNNYNS ANPN PHVLY	SALVGO	DSNDNYQDLR	400
EFN70196.1	HAANIAFICLEASEG PHQYDKFAKEQIDYIL	GIPGRSYVV	BYGINAPQQPRHIASSCPHLPKSCGWEQYRKNEPNLQILY	SALVSC	PNENDFFYDNR	394
EFN70197.1	HAANIAFICLEASEG PHQYDKFAKEQIDYIL	GIPGRSYVV	STGIN A POUPRHIASS CPHL PKWCGWKQYRKNESN POILY	ALVSO	PNENDFFYDNR	395
EFX093/2.1	HAAN LAEVOLONING	GDTGRSFVC	BEGYN DDY YDUHESSSCON DDAUGDWETY SETTENSET.	ALVGO	UNNUNYSUDR	389
EFX8060F 4	AASNIAFICIONINI	GDTGPSEVV	EGHN PPI KPHHASSSCON PAVODWOTT SOTTPN PQILY		PONNOHYVODO	300
EFX868221	HAANAAFGCLLVADSGIGNAAYKAFAKOOIDVA	GSTGRSYVV	GEGNN PPVKCHHRGASCPDMPAPCGWNEYNSPNPNGOLLE	ALVG	BDVNDNYNDAR	394
EG/63652 1	HVANVAFVCLVAADG EQEYROFAKEQIDYM	GGAGRSYVV	SYGKNSPKQPHHAASSCPNKPTPCGWQEFSKNAPNPQ I I Y	ALVS	PDEVDLEKDCR	394
EKC17289.1	YASNMAFMALLAADDGLHSS - YRTWAKSQINYA	GDAGRSFVC	S F G V N P P E Q P H H R G A S C P T L P A P C S WA D Q T K H A P N P H V L Y	ALVG	PDGHDSYRDSR	392
EKC39023.1	YSANAAFIALLAADAGAN YKQWAMSQIHYM	GNADFSYVI	G Y G S N Y P L K P H H R A S S C D S P P A V C D W N D Y N S P D P N P H V L Q	ALVGO	PDQYDNYEDAR	391
KDR16731.1	HAANVAFVCLQAADPSRYREFARRQIDYM	GDSGRSFVV	F G Y N Y P R Q P H H A A S S C P T R P A P C G W D A F G W P R D N P Q L L Q	SALVSO	PDENDHYMDNR	393
KMQ95413.1	HAANVAFLCLEAAD-GDSLI-YRQFAEEQIAYI	GSTGRSYVV	GYGKN SPKQPLHAASSCPD RPAPCGWADY DKNESN PQILY	SALVSO	PNENDEFNDHR	395
KOC62382.1	HAANVAFVCLEAADG NQEYREFAEQQIYYM	GGGGRSYVV	WGRNPPKQPHHAASSCPDRPAACGWPEFDKDAPNPQILY	SALVSO	PDEADKFQDHR	394
KOE68403 1	YAGNMAFIALLAAFSGINSTVYRYWALOO HYL	GDSGRSYVV	AGHN PPMRPHHAGSSCKV SG - PCSWDD FESKGTN PILLY	AVVG	PDINDSYLDDR	393
KOX76353.1	HAANVAFVCLQAAD-GDQE YRQFAEQQIHYM	GGSGRSYVI	SWGRN PPKQPHHVASSCPD RPAPCGWSELDK NAPN PQ ILY	ALVSO	PDEADKFHDRR	394
KXJ18436.1	Y SANTAFLAL VAADNGL K D I - Y RN FAKKQ I H Y ML	GDSGRSFVV	5 F G A N P P K R P H H R S S S C P D R P A P C T H D A L N A P G A N P Q V L X	ALVGO	PDVNDNFKDDR	397
KY B27092.1	HAANVAFICLHA GLSQAAYVSFAKEQINYM	GSTGQSFVV	GYGQNYPKQPHHSASSCPNLPEPCGWKQFTWKGPNPQILY	ALVSO	PDQNDHYEDVR	394
KYM87195.1	HVANVAFVCLVAADG EQEYRQFAKEQIDYML	GAGRSYVV	YGKN SPKQPHHAASSCPN RPTPCGWQE FSK NAPN PQILY	ALVSO	PDEVDLFKDCR	394
KYN01696.1	HAANVAFVCLVAADG EQEYRQFAKEQIDYM	GGAGRSYVV	Y G K N S P K Q P H H A A S S C P N R P T P C G W Q E F S K N A P N P Q I L Y	SALVSO	PDEVDSFKDCR	394
KYN22851.1	HAANVAFVCLVAADG EQEYRRFAKEQIDYM	GGAGRSYVV	YGKNSPKQPHHAASSCPNRPTSCGWQEFSKNAPNPQILY	SALVSO	PDEVDLFKDCR	394
KYN35700.1	HVANVAFVCLVAADG EQEYRQFAKEQIDYM	GGAGRSYVV	SYGKN SPKQPHHAASSCPS RPTPCGWQE FSK NAPN PQILY	ALVSO	PDEVDLFKDCR	394
KYQ52831.1	HAANVAFVCLVAADG EQEYRQFAKEQIDYM	GGAGRSYVV	GYGKN SPKQPHHAASSCPN RPTSCGWQE FSK NAPN PQILY	SALVSO	PDEVDLFKDYR	394
KZC10616.1	HAANVAFVCLEAADG NQE - YREFAEQQIYYM	GGGGRSYVI	WGRN P P K Q P H H A A S S C P D R P A A C G W S E F D K D A P N P Q I L Y	SALVSO	PDEADKFHDHR	394
KZS03923.1	HASNAAFGCLLVADSGIGNAAYKAFAKQQIDYA	GSTGRSFVV	6 F G N N P P V K C H H R G A S C P D A P A P C G W N E Y N S P N P N G Q V L E	ALVG	PDINDYYNDAR	397
KZS05722.1	YAANSAFICLQNINQ TYRKFAQKQIHYIL	GDTGRSFVV	<b>G</b> F G V N P P V K P H H R S S S C P R R P A P C G N A N L V S K A P N P Q I L Y	SAMVGO	PSDRDSYVDDR	388
KZS07093.1	HASNIAYICLQGINL AYRKLAQQQIHYAL	GDTGRSFVC	G F G T N P P V R S H H R S S S C P N K P A T C D W N T Y N S P A P N A H I L Y	GALVGO	P D S N D N Y D D N R	390
KZS16514.1	HASNIAFICLQVILVNISLT-YRQFAQKQIHYA	GDTGRSFVV	S F G V N P P V K P H H S S S C P N R P A P C D W S T Y S S T G P N P Q I L Y	SAL VG	P D D K D V Y V D D R	393
KZS16516.1	HASNVAFICLQNISL TYRQFAQKQIHYA	GDTGRSFVV	G F G V N P P G K P H H S S S C P N R P A L C D W S A Y S S T A P N P Q I L Y	GAL VG	PDNNDFYVDDR	389
OAD61293.1	HAANVAFVCLEAADG DQEYREFAEQQIYYM	GGGGRSYVV	GWGRN P P K Q P H H A A S S C P D R P A I C G W P E F D K D A P N P Q I L Y	SALVSO	PDEADKFHDHR	394
ODM94862.1	HAANVALTCLQAAEAGINSK - YRAFAKQQIDYM	GSTGRSFVC	SYGNN PPQ RPHHASSSCPN RPAPCD WDD FSK PGPN SQ V L H	SALVGO	NRQDQYNDDR	397
ODM95900.1	HASNVAFICLQAAEAGINEK - YRAFAKTQLDYA	GSTGRSFVC	S F G N N P P Q R P H H R S S S C P N R P A P C D W Q A Y N S P D P N P Q V L T	ALVGO	DQQDQYTDDR	393
OWF45593.1	YASNMAFMALIAAEDGIHSA - YRAWAMSQIHYAL	GDTGRSFVV	B F G V N P P T H P H H R A A S C P S M P A P C S H A D L K N V G P S P H T L Y	ALVGO	GGTDDYKDNR	394
OWF56607.1	T SANSAFIAL MAAKTGIADTSYRDYARGQLHYMU	ONEIRSYVI	BEGNNEPTQPHHKSSSCDDLPAPCNDSNQYRAGPSPQTLY	LVGO	GRNGQFTDRR	396
PBC31793.1	HAANVAFVGLEAAD-GDQKYREFAEQQIYYML	GGGGRSYVV	SWORN PPROPHHAASSCPD RPAICGWSEFDK DAPN PQILY	LVSO	DEADKFHDHR	394
pabi3MC3 A	TAANAAF IAL VAADOOL VAADOOL VAANTAELAL VAADOOL VAA	OD AGRSYVVI	BEGKN PROPHENSSSCPDQ PAECDWDE EN QPGPN YQ I L Y	AL VGC	PONDQFEDLR	395
PEA25363.1	SASNVAFICI FAASAGISSSTVDNEAVVOINVA	GDGGBSVVV	BEGNN PPTHEHHRSSSCPDKPAACDWN TERCCOAN AUUUU	ALVG	PGANDOYKUDR	300
PVD35227 4	YSANMAFFAL VAAKYGI HAD - V DOWAMCO I HYA	GDTGPSEVV	6 FGVN PPT RPHHRASSCPMI PAPCGWEAGON DGDN DUVI V		PGGSDDYVDDP	303
PVD35340.4	YASNTAFIALLAARRGINAK-YNOWGMCOLHYA	GDTGRSEVV	FGVNPPTKPHHRASSCPCKPFPCTWDAORFPGPNPKTI V	ALVG	PGSDDSYTDDR	303
PVD35381.1	YSSNMAMAALMAAEAGIHPE - YRHWAMCQIHYA	GDTGRSFVV	G F G K N P P V S P H H R S S S C P N L P V R C N M N Y L H L D T P N T H M L C	ALVGO	PDSSDGYKDSR	378
					and a second second product of the second seco	

PVD35384 1	TASNMAMAALMAAEAGIHPE-TRHWAMCQIHTAL	GDTGRSFVV	F G K N P P V S P H H R S S S C P N L P V R C N M N Y L H L D T P N T H M L C	GAL	VGC	DSSDGYKDSR	393
	YAADMAMAALMAAEAGIHPE - YRHWAMCQIHYIL	GDVGHSFVV	G F G K N P P V S P R H R A S S C P N P P A T C N M H T L H E P G P N P H L L C	GAV	GG	PDGRDGYTDER	394
PVD35385.1	YAANMALAALMAAEAGIHPE - YRHWAMCQIHYS	OD AGHS YVV	GEGEN PPVSPHHRSSSCPSPPASCIPSVLHAHGPNPHVLG	GAL	GG	DGRDGYKDDR	393
PVD35551.1 PVD36624.1	VASNMAAMAI JAADSNSAI E. VRHWAMCVIHVA	GDTGESVLI	E FONT Y PL NPH HRAASCP DE PAP COWDAL SA AGPS PHILY	GAL	G	PDWSGNYEDSR	385
PVD36625.1	YASNMAMVALISADSDIMPE-YRHWAMCIIHYAL	GDTGFSYLI	S F G D Q W A H S P H H R S S S C P N Q P A P C G P F V M S S T H P N V H T L Y	GAL	VGC	PDQNDAYTDLR	378
PVD36896.1	HAANVAFAALVAAEDGIAEV - NRKWAAEQINYI	GNPHGSYEI	G Y G N K Y P L R P H H R A A S C P D P P A T C D Q R N L N T D N P S P H I L Y	GGL	VGC	PILNDGYNDTR	368
PVD36986.1	YAGNTAFLALAAAEAGLTPK - LRKWAIEQINFI	GNHQGSFQI	GYGSKYPRSPHHRGASCPD I PQPCSEANLHSSEPSPHVLV	GAI	VGG	P D L A G K Y E D N R	373
XP_001606454.1	HAANVAFVCLQAAD-GDKEYREFARQQIHYIL	GGTGRSYVV	G W G R N P P K Q P H H A A S S C P D R P A P C G W P E F D R N R P N P Q V L Y	GAL	SC	PDEADRFHDKR	394
XP_001810693.2	HAANVAFICLHA GLSQAAYVSFAKEQINYMU	GSTGQSFVV	G Y G Q N Y P K Q P H H S A S S C P N L P E P C G W K Q F T W K G P N P Q T L Y	GAL	S S	DONDHYEDVR	394
XP_001944774.2 XP_002426465.1	HASNVAEVCLOAADS SSKYRKEAKROIDYI	OD SGRSYLV	FGKDYPKOPHHAASSCEKRPAKCGWEAYSOPGPNPOLLY	GAL	S S	DENDYYEDKR	393
XP 003402778.1	HAANVAFVCLEAADG DOEYREFAEQQIYYMI	GGGGRSYVV	W G R N P P K Q P H H A A S S C P D R P A V C G W P E F D K D A P N P Q I L Y	GAL	s	PDEADKFHDHR	394
XP_003690676.1	HAANVAFVCLEAAD-GDQKYREFAEQQIYYM	GGGGRSYVV	WGRNPPKQPHHAASSCPDRPAICGWSEFDKDAPNPQILY	GAL	vsc	PDEADKFHDHR	394
XP_005099312.1	Y A G N S A F L A L I A A E A G I G T R - Y R K WA V E Q I N F I L	GNKHGSYQI	G Y G S K Y P L R P H H R A A S C P D R P K P C T E A Q L N D P K P S P Q L L V	GAL	VGC	PLQSGEYKDDR	372
XP_005099315.1	YAADAAFIALVAAEDGINST - YRKWAAEQINYIL	GNHHSSFEI	G Y G S S F P R K P H H R G S S C V A K P -  -  C -  -  -  S I G D A G N N P H V L L	GAV	VGC	PNADDSYTDTR	365
XP_005101787.1	YAANAAFLALLAAKENINPH - FRKFAVEQINYIL	GNHIGSFEI	G YGS K F P R K P H H R G A S C P T H - G T C G S A D L N S H A N S P Q I L Q	GAV	GG	DLGNVYTDSR	367
XP_005101789.1 XP_005111868.2	YSSNMAMIAL VAADAG INSE - YRHWGMCO I HYAL	GDAGRSEVV	FGHN PPVSPHHRGASCPTH - GPUNNSNK DSPVNSPQTLQ	GAL	G	BOSSDNYKDER	305
XP 005111869.2	YASNMAMSALVAADMGIQPE-YRHWAMCQIHYAL	GDTGFSFLI	G F G D K Y P H S P H H R S S S C P N L P A P C G P F V M S S K E P N V H T L Y	GAL	VGC	DOSDGYTDDR	394
XP_005111870.1	YSSNMAMGALIAAKLGVNPP - YRHWAMCQIHYAL	GDTGFSYLI	F G N K Y P R S P H H R S S S C P N L P A P C G P F I Q S S S E P N V H I L Y	GAL	VGC	F G D S D N Y S D K R	393
XP_006610535.1	HAANVAFVCLEAAD-GDQKYREFAEQQIYYML	GGGRSYVV	G WGR N P P K Q P H H A A S S C P D R P A I C G W S E F D K D A P N P Q I L Y	GAL	VSC	PDEADKFHDHR	394
XP_008181842.1	HAANIVFLCLQAADS AQYRAFAKYQIDYI	GDGGQSYVV	F G K N Y P L Q P H H A A S S C L D R P A P C D W E A Y R S P K P N P Q V L Y	GAL	SC	PDENDNYKDLR	393
XP_008194455.1	HAANVAFICLHA GLSQAAYVSFAKEQINYMU	GSTGQSFVV	GYGQN YPKQPHHSASSCPNL PEPCGWK QFI WRGPN PQIL Y	GAL	1 Se	DONDHYEDVR	393
XP_0090498361	YASNAAFIAI VAADIGVNPI - YRKWAISOINYA	GDTGRSFVI	F G V N P F K N P H H R G S S C P M M P A P C A W P D G I G P G P N P H I L I	GAL	N G	GSSDDYVDOR	378
XP_009049839.1	YSANTALIALMAADQGVNPL - YRKWALSQINYAL	GDAGRSFVV	G F G K N P P T R P H H R A S S C P L L P A P C G W S A Q Q N P G P N P H T L Y	GAL	VGC	PGKNDDYADDR	393
XP_009049840.1	FAGNVAFLALVAANADINPSMYRRWAKQQIHYI	GDTGRSYVV	G F G N N P P T R P H H A S S S C N S P P L P C T W D D F S K S T P N S H I L Y	GAV	VGC	PDKFDHFVDDR	394
XP_009049941.1	YSSNMAMVALIAAENGINPK-YRKWAMSQIHYAL	GDTGLSYVI	G F G D K Y P L R P H H R S A S C P W P P A P C S P N V V H R P Q A S V H T L Y	GAL	VGC	P G S G D Q Y K D E R	393
XP_009058133.1	YSSNMALIALMAADEGLHPE - YRKWAMSQIHYAL	GDTGFSFVI	G Y G K N Y P H S P H H R G A S C P L P P A P C G P F V M S S K Q P N V H T L H	GAL	GGG	PDSDGSYQDSR	393
XP_009058134.1	YAANTAFLALMAADAGIK YRKWAQSQINYIL	GNKQFSYVI	GEGKKEPRNPHHRSASCPMIPAPCSEANLHSSGPSPHILV	GAL	GG	DAKGAYIDKR	387
XP_009064976.1	YSANMAMVAL MAAEVG I NEK - Y RDWAMGO I HYAL	GDTGESYVI	OVGDK YPT HPH HRAASCPWPPAPCS PEEV NR PEPSTHTLY	Gell	G	DSTDATVDNR	393
XP_009064980.1	HAADNSYAALWAAELGLE N R NWALSQINYIL	GNNYMSYVI	G F G N K Y P L K P H H R P S S C P D L N M Y C G W D A F Y S T A P N P H R L E	GAL	VGC	PNKTDGWDDDR	389
XP_009065702.1	YAANTAFMALMAADHGL SLRKWAQGQINYI	GNNKFSYMI	FGKHYPLRPHHRGASC HAHCDSSILHSSAPSPHVLV	GAL	VGQ	PDEHDKYVDKR	384
XP_011057333.1	HVANVAFVCLVAADG EQEYRQFAKEQIDYML	GAGRSYVV	Y G K N S P K Q P H H A A S S C P N K P T P C G W Q E F S K N A P N P Q I L Y	GAL	SC	PDEVDLFKDCR	394
XP_011164235.1 XP_011252652.4	HAANVAFICLVAAD-G-DQE-YRQFAKEQIDYM	GLEGESVVV	OTUKNSPKQPHHAASSCPURPASCGWQEFSKNAPNPQILY	SAL	SC	ENENDEEVDUR	394
XP_011253654 1	HAAN LAFICMEAADG PHOYDIFAKEO IDY	GKAGRSYVV	YGVNSPQOPHHIASSCPDRPKFCRIOFYNOTIANPOVIV	GAL	V S	DENDDEEDNR	393
XP_011253655.1	HAAN I AF I CLEASEG PHQY DKFAKEQ I DY I	GIPGRSYVV	GYGIN APQQPRHIASSCPHL PKWCGWKQY RK NESN PO IL Y	GAL	vs	PNENDFFYDNR	395
XP_011296914.1	HAANVAFICLEAAD-GDHEYREFAKHQIHYM	GGGGRSYVV	WGINPPQQPHHAASSCPNLPTHCGWPEFDRKAPNPQILY	GAL	Vsc	PDEADKFRDHR	394
XP_011296916.1	HAANVAFICLEAAD-GDHEYREFAKHQIHYML	GGGGRSYVV	G WG I N P P Q Q P H H A A S S C P N L P T H C G W P E F D R K A P N P Q I L Y	GAL	vsc	PDEADKFRDHR	394
XP_011346866.1	HAANVALICLAAADG DREYRQFAKEQIDYML	GAAGRSYVV	G Y G K D P P K Q P H H S A A S C S N R P V P C G W P E L D R N A P N P Q I L Y	GAL	SC	PDVADQFQDRR	394
XP_011421782.1 XP_011422150.1	YASNMAFMALLAADUGLHSS-YRTWAKSQINYA	GD AG R S F V C	6 F G V N P P E Q P H H R G A S C P T L P A P C S WA D Q T K H A P N P H V L Y	GAL	GGG	DGHDSYRDSR	392
XP_011423139.1 XP_011430172.1	YAANAAL VALMAVDEGD AYKIWALKOINYMI	GNKQMSYQI	GYGTKYPOKPYHRASSCPDPPSPCGWDNINNPGPSPOIL R	GAL	G	PDVNDVYVD	388
XP_011447501.1	YSANAAFIALLAADAGAN YKQWAMSQIHYML	GNADFSYVI	G Y G S N Y P L K P H H R A S S C D S P P A V C D W N D Y N S P D P N P H V L Q	GAL	GC	PDQYDNYEDAR	391
XP_011447502.1	YSANAAFIALLAADAGAN Y KQWAMSQIHYML	GNANFSYVI	G Y G S N Y P L K P H H R A S S C D Y P P E V C D W N D Y N S P D P N P H V L Q	GAL	VGC	PDQYDNYEDAR	389
XP_011498432.1	HAANVAFICLQAADG DQE - YREFAKQQIHYIL	GGAGRSYVV	G WG H N P P T Q P H H A A S S C P D K P T A C G W S E F D R N A P N P Q V L Y	GAL	VSC	P D E A D R F H D R R	394
XP_011632703.1	HAANVAFLCLVAAD-G-DQK-YRQFAKEQIDYML	GRAGRSYVI	G Y G N N P P K Q P H H A A S S C P N R P A P C G W Q E F I K N T P N P H I L Y	GAL	SC	PDEVDLFKDYR	394
XP_011867251.1	HAANVAFVCLVAAD-G-DRE-YRREAREOIDYM	GAGRSYVV	HGKNSPKOPHHAASSCPN RPAPCGWOEFSKNAPN POLLY	GAL	S	DEVDJEKDCR	394
XP 012061951.1	HVANVAFVCLVAAD-G-EQE-YRQFAKEQIDYML	GGAGRSYVV	G Y G K N S P K Q P H H A A S S C P N R P T P C G W Q E F S K N A P N P Q I L Y	GAL	VSC	PDEVDLFKDCR	394
XP_012146513.1	HAANIAFVCLEAADG DQVYREFAEQQIYYML	GGGGRSYVV	G WG R N P P K Q P H H A A S S C P N K P A V C G W S E F D K D A P N P Q I L Y	GAL	vsc	PDEADKFHDHR	394
XP_012224814.1	HAANVAFICLEAADS D DQEY RQFAKEQ I D Y M	GAGRSYVV	G	GAL	VSC	PDEADKFQDRR	395
XP_012248155.1	HAANVAFVCLEAADG DQE - YREFAEQQIYYML	GGGGRSYVV	G W G R N P P K Q P H H A A S S C P D R P A V C G W P E F D K D A P N P Q I L Y	GAL	SC	PDEADKFHDHR	394
XP_0122/369/.1	HAANVAFVCLQAAD-GDKEYREFAEQQIQYM	GGGGRSYVV	6 W G E N P P Q Q P H H A A S S C P K R P A P C G W P E F D R K A P N P Q I L F	GAL	5	IDEADRFQDHR	394
XP_012935745.1	YASNMAMAALIAADAGINPE-YRHWAMCOIHYA	GDSGRSFVV	FGHN PPKSPHHRGSSCPVL PARCGS FAMRA PGPN VHTLY	GAL	G	PGSSDDYKDER	394
XP_012938497.1	YAANTAFMALLAADASINSN-NRKFAVQQINYIL	GNKHGSYQI	GYGQK Y PQ R P H H R G S S C P - K G S T C G W D F Y H S S K A N V N K L L	GAL	VGC	PDAHDGYQDNR	368
VD 012020021 1	VAANAAE LAL VAAEDGI NASY DKEAVEO INVI	NERARVEL			Ge	KUDDSVTDVP	245
XP 012939933.1	ANAAFLALLAAKENINPP - FRKFAVEQINYIL	GNKIGSFEI	GYGSK YPL KPHHRGASCPTH - GPCNTSNRDSPGNSPQILQ	GGV	GG	PDLQNGYEDAR	330
XP_012941225.1	YAANTALIALMAAEQGIGTS - YRTFAETQINYML	CNGFSYQV	F G S N F P L R P H H Q S A T C P L S P A T C D W S T F N D P A P N A H E L T	GAL	VGC	PGKNDDYVDKR	394
XP_012945520.1	YSSNMAMMALIAADAGIHPE-YRHWAMCQIHYA	GDTGFSYVV	FGDDYPEKPHHRGSSCPV I PAQCGSYVLRSHAPNTHTLY	GAL	VGC	<b>P</b> GRSDDYKDDR	394
XP_013066277.1	LAANSAFLALLAADAGI - EVSYRKWAVEQINYI	GNHHGSFEV	G Y G D R Y P L Q P H H R G A S C P N K P A P C G Q N E A G S A E P N P Q V L T			PDEYDQYVDSR	369
XP_013066278.1 XP_013066280.1	YAASSALVALVAAESGIDSR-YRRWAAEQVNYI			GAL	VGC	DIDDDYVDLR	312
Pdf010000000000.1	YAISNVAELAL AISEMISTIAK-YRKWAVEDVNEL	GNNHGSYVV	G Y G S K Y P Q Q P H H A G A S C P D R P A S C G W A Q Y S S P S P N P Q V L H N Y G S K Y P L O P H H R S S S C P S P P O A C N W D T Y N L H S P N P H V L N	GAL GAL		PRONDNYVDIR	372
XP 013066281.1	EAANNAFIALVAAESGIESK - YRKWAVEQINYI	G N N H G S Y V V G N P H G S Y O V	Ğ Y G S K Y P Q Q P H H A G A S C P D R P A S C G W A Q Y S S P S P N P Q V L H 9 Y G S K Y P L Q P H H R S S S C P S P P Q A C N W D T Y N L H S P N P H V L N 9 Y G T K Y P L H P H H R G S S C P D R P A P C D W N Y F N S A A P N P O V L Y	GAI GAL GAL GAL		P D Q N D N Y V D L R P D E N D I Y Q D K R	372 373
XP_013066281.1 XP_013066283.1	EAANNAFIALVAAESGIESK - YRKWAVEQINYIL EAANNAFIALVAAESGIESK - YRKWAVEQINYIL	G N N H G S Y V V G N P H G S Y Q V G N P H G S Y Q V	Ϫ Υ G S Κ Υ Ρ Q O P H H A G A S C P D R P A S C G WA O Y S S P S P N P O V L H S Y G S K Y P L O P H H R S S S C P S P P Q A C N WD T Y N L H S P N P H V L N G Y G T K Y P L H P H H R G S S C P D R P A P C D W N N F N S A A P N P Q V L V G Y G T K Y P L H P H H R G S S C P D R P A P C D W N N F N S T A P N P Q V L V	GAL GAL GAL GAL		P D Q N D N Y V D L R P D E N D I Y Q D K R P D E N D I Y E D K R	372 373 372
XP_013066281.1 XP_013066283.1 XP_013066284.1	EAANNAFIAL VAAESGIESK - YRKWAVEQINYIL EAANNAFIAL VAAESGIESK - YRKWAVEQINYIL EMANTAFLAL LAAESGIQTR - LRQWAVQQINYIL	G N N H G S Y V V G N P H G S Y Q V G N P H G S Y Q V G N P H G S Y Q V	G YGSK Y POQPHHAGASCPD RPASCGWAQYS B PSPN POULH YGSK YPL QPHHRSSSCPS PPAOLCNWD T YN LHSPN PHVL N G YGTK YPL HPHHRGSSCPD RPAPCD WNN FNS AAPN POULV YGTK YPL HPHHRGSSCPD RPAPCD WNN FNS TAPN POULV G YGTK YPL UPHHRAASCPD KPAPCGWDN FNS PLPN PHVLQ	G A I G A L G A L G A L G A L G A L		P D Q N D N Y V D L R P D E N D I Y Q D K R P D E N D I Y E D K R P L V D D S Y E D K R	372 373 372 373
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066285.1	TAGMYAFLALLAGENGIDAK-YNXMAVEQVNFLL EAANNAFIALVAAESGIESK-YRXWAVEQINYIL EMANTAFLALLAAESGIESK-YRXWAVEQINYIL EMANTAFLALLAAESGIYTK-LRQWAVQOINYIL	G N N HG S Y V V G N P HG S Y Q V G N P HG S Y Q V G N T F G S F Q I G N N HG S Y Q I	G YGSK Y POQPHHAGASCPD RPASCGWAQYS SPSPNPOULH YGSK YPL QPHHRSSCPS PPAOL NWD T YN LHSPN PHVL N G YG TK YPL HPHHRGSSCPD RPAPC DWNN FNS AAPN PQVL V G YG TK YPL HPHHRGSSCPD RPAPC DWNN FNS TAPN PQUL V G YG TK YPL QPHHRAASCPD KPAPC GWD N FHA TTPS PQVL Y G YG TK YPT HPHHRGASCPD KPAPC GWD N FHA TTPS PQVL Y	G A L G A L G A L G A L G A L G A L		P D Q N D N Y V D L R P D E N D I Y Q D K R P D E N D I Y E D K R P L V D D S Y E D K R P E V D D S Y E D K R	372 373 372 373 372
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1 XP_013066286.1	EAANNAFIALLAGENGIDAA - YNWAVEQUNFL EAANNAFIALVAAESGIESK - YNWAVEQUNYIL EMANNAFIALVAAESGIESK - YNKWAVEQUNYIL EMANNAFIALVAAESGIQTR - LRQWAVQQUNYIL EAANNAFLALLAAESGUDSR - LRKWAVEQUNFIL EAANNAFLALAAAESGUDSR - LRKWAVEQUNFIL	GNNHGSYVV GNPHGSYQV GNPHGSYQV GNTFGSFQI GNNHGSYQI GNNHGSYQI	G YGSK Y POQPH HAGASCPD RPASCGWAQY SS PSPN POVL H G YGSK Y PL QPH HRSSSCP SP POACNWD T YN LHSPN PHVL N G YGTK Y PL HPH HRGSSCPD RPAPCDWNN F NS AAPN POVL V G YGTK Y PL HPH HRGSSCPD RPAPCDWNN F NS TAPN POVL V G YGTK Y PL QPH HRAASCPD KPAPCGWDN F NS TAPN POVL V G YGTK Y PT HPH HRGASCPD KPAPCGFEN LHATTPS POVL V G YGTK Y PT HPH HRGASCPD KPAPCGFEN LHATTPS POVL V G YGTK Y PT HPH HRGASCPD D KPAPCGFEN LHATTPS POVL V	G A L G A L G A L G A L G A L G A L G A L		P D Q N D N Y V D L R P D E N D I Y Q D K R P D E N D I Y E D K R P L V D D S Y E D K R P E V D D S Y E D K R P E A D D S Y A D N R	372 373 372 373 372 372 372 372
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1 XP_013066286.1 XP_013066370.1 XP_013066371.1	TAGMYAFLALLAGEMGIDAA - YNWAVEQVNFL EAANNAFIALVAAESGIESK - YNKWAVEQINYIL EAANNAFIALVAAESGIESK - YNKWAVEQINYIL EMANYAFLALLAAESGIOTR - LRQWAVEQINFIL EAANVAFLALAAESGVDSR - LRKWAVEQINFIL YAGNSAFLALLAAESGIFPR - LRKWAVEQINFIL YAANSAFLALLAAEAGINFD - YNKWAVEQNNYI	G N N H G S Y V V G N P H G S Y V V G N P H G S Y Q V G N T F G S F Q I G N N H G S Y Q I G N N H G S Y Q I G N N H G S F Q I G N K H G S F Q I	GYGSK Y PDQPHHAGASCPD RPASCGWAQYS B PSPNPQU H YGSK Y PL APHHRGSSCP S PPAC CNWD T YN L HSPN PHVL N YGTK Y PL HPHHRGSSCP D RPAPC DWN F NS A PN PQU L Y YGTK Y PL HPHHRGSSCP D RPAPC DWN F NS TAPN POVL Y VGNK Y PL QPHHRAASCPD KPAPC GWD N F NA T PS PQU L Y YGK Y PT HPHHRGASCPD K PAPC GWD N F NA T T PS PQU L YGTK Y PT HPHHRGASCP N K PAPC GE N L HAT T PS PQU L Y YGK Y PT HPHHRGASCP D RPAPC GS F A QL NA T PS PQU L Y YGB X F PRQPHHRGASCP D RPAPC S F A QL NA PGPS PHVL	G A L G A L		P D Q N D N Y V D L R P D E N D I Y Q D K R P D E N D I Y E D K R P L V D D S Y E D K R P E V D D S Y E D K R P E A D D S Y A D N R P E A N D N Y R D V R P E Y N D N H H D L R	372 373 372 373 372 372 372 372 372
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066286.1 XP_013066370.1 XP_013066371.1 XP_013069169.1	TAGMYAFLALLAGEMGIDAA - YAWAYEQ NY L EAANNAFIALVAAESGIESK - YAWAYEQ NY I EAANNAFIALVAAESGIESK - YAWAYEQ NY I EMANTAFLALAAESGIOTA - LROWAYQO NY I EAANYAFLALAAAESGYDSR - LRKWAYEQ NY I YAGNSAFLALLAAESGINTD - YRKWAYEQ NY I YAGNSAFLALLAAESGINTD - YRKWAYEQ NY I YSSNMAMAALVAADAGLNAC - YRKWAYEQ NY I	G N N H G S Y V V V G N P H G S Y Q V G N P H G S Y Q V G N T F G S F Q I G N N H G S Y Q I G N N H G S Y Q I G N N H G S F Q I G N K H G S F Q I G D T G F S Y L I	GYGSK Y POQPHHAGAS CPD RPASCGWAQYS SPSPNPQU H YGSK YPL QPHHRGSSC PS PPAC NWD TYN HSPN PHVL N YGTK YPL HPHHRGSSC PD RPAPCD WNN FNS A PN PQU LY GYGTK YPL HPHHRGSSC PD RPAPCD WNN FNS TAPN PQU LY GYGTK YPT HPHHRGASC PD KPAPCGWD N FNA TTPS PQU LY GYGTK YPT HPHHRGASC PD KPAPCGWD N FNA TTPS PQU LY GYGTK YPT HPHHRGASC PD RPAPCG FL ATTPS PQU LY GYGTK YPT HPHHRGASC PD RPAPCG GAQL ND PGPS PQI LY GYGK YPL APHHRASC PD RPAPCS GAQL ND PGPS PUL L GYGDK YPL QPHHRASC CPD RPASCS FAQL NA PGPS PVU LI	6 A L 6 A L		DQNDNYVDLR DENDIYQDKR DENDIYQDKR DEVDDSYEDKR DEVDDSYEDKR DEVDDSYEDKR DEADDSYADNR DEANDNYRDVR DEVNDNHDLR PEQNDNHDLR	372 373 372 373 372 372 372 372 371 394
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066371.1 XP_013069170.1	TAGMYAFLALLAGENGIDAA - Y KWAVEQ (Y NFLE EAANNAFIALVAAESGIESK - Y KWAVEQ (NY IL EAANNAFIALVAAESGIESK - Y KWAVEQ (NY IL EMANTAFLALLAAESGVDSR - L RKWAVEQ (NY IL EAANVAFLALAAAESGVDSR - L RKWAVEQ (NY IL YAGNSAFLALLAAAESGINTD - Y KWAVEQ (NY IL YAGNSAFVALLSAEAGINTD - Y RKWAVEQ (NY IL YSSMAMAALVAADAGLNAQ - Y RHWAMCQ (HY AL	G N N H G S Y V V G N P H G S Y Q V G N P H G S Y Q V G N T F G S F Q I G N N H G S Y Q I G N N H G S Y Q I G N N H G S F Q I G N K H G S F Q I G D T G F S Y L I G D T G F S Y L I	GYGSK Y POQPHHAGASC PD RPASC GWAQY S P S PN PQ V L H YGSK Y PL PP HHRGSS C P S PPAC C NW D T YN L H S PN PH V L N YG TK Y PL HP HHRGSS C PD R PAP C D WNN F NS A AP N PQ V L V GYGTK Y PL HP HHRGSS C PD K PAP C GWD N F NS T AP N PQ V L YG NK Y PL U P HHRAASC PD K PAP C GWD N F HA T T PS PQ V L YG TK Y PT HP HHRGASC PD K PAP C GWD N F HA T T PS PQ V L YG TK Y PT HP HHRGASC PD K PAP C G F E N L HA T T PS PQ V L YG SK F PC PH HRGASC PD R PAP C SG A Q L ND PG PS P0 I L V YG SK Y P C PD HHRAASC PD R PAP C SG A Q L ND PG PS P0 I L V GYG SK Y PL Q P HHRAASC PD R PAP C SG A Q L ND PG PS P0 I L V S FG DK Y PL Q P HHRAASC PN L PAP C G F I M SS K E PN Y H I L Y S FG DK Y PL SP HHRASC PN L PAP C G P F I M SS K E PN Y H I L Y	6 A L 6 A L		DQNDNYVDLR           PDENDIYQDKR           PDENDIYQDKR           PDENDIYQDKR           PDENDIYQDKR           PEVDDSYEDKR           PEVDDSYEDKR           PEADDYADNR           PEADDYADNR           PEADNYRDVR           PEQNNHHDLR           PGQSDDYTDDR           PGQSDDYTDDR	<ul> <li>372</li> <li>373</li> <li>372</li> <li>373</li> <li>372</li> <li>372</li> <li>372</li> <li>372</li> <li>371</li> <li>394</li> <li>394</li> </ul>
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066286.1 XP_013066286.1 XP_013066370.1 XP_013066370.1 XP_013069169.1 XP_013069170.1 XP_0130699770.1	TAGMYAFLALLAGEMGIDAA - YAWAYEQ VIFL EAANNAFIALVAAESGIESK - YRKWAYEQ INYIL EAANNAFIALVAAESGIESK - YRKWAYEQ INYIL EAANNAFLALAAESGIQTR - LRQWAYEQ INYIL EAANNAFLALAAESGYDSR - LRKWAYEQ INFIL YAGNSAFLALLAAESGITDR - LRKWAYEQ INFIL YAGNSAFVALLSAEAGINTO - YRKWAYEQ INYIL YSSMMAMAALVAADAGLNAO - YRHWAMCQ IHYAL YSSMMAMAALVAADAGLNAO - YRHWAMCQ IHYAL YSSMMAMAALVAADAGLNAO - YRHWAMCQ IHYAL	G N N HG S Y V V G N P HG S Y Q V G N P HG S Y Q V G N T FG S FQ I G N N HG S YQ I G N N HG S YQ I G N N HG S FQ I G N K HG S FQ I G D T G F S YL I G D T G F S YL I G D T G F S YV I	Y GSK Y PDQPHHAGASCPD RPASCGWAQYS SPSPNPQU H Y GSK Y PL QPHHRSSCPS PPACC NWD T YN L HSPN PHVL N Y GT K Y PL HPHHRGSSCPD RPAPC DWNNFNS A PN PQU L Y V GT K Y PL HPHHRGSSCPD RPAPC DWNNFNS TAPN PQU L Y V GT K Y PL HPHHRGASCPD K PAPC GWDNFNS FLPN PHVL Q Y GT K Y PT HPHHRGASCPD K PAPC GWDNFHA TT PS PQV L Y GYG K Y PT HPHHRGASCPN K PAPC GF EN LHA TT PS PQV L Y GYG K Y PL UPHHRASCSCPN K PAPC GF EN LHA TT PS PQV L Y GYG K Y PL UPHHRASCSCPN K PAPC GF EN LHA TT PS PQV L Y GYG K Y PL UPHHRASCSCPN K PAPC GF EN LHA TT PS PVV L GYG K Y PL SPHHRGSSCPN K PAPC GF FI MSS K EPN YHI L Y GYG K Y PL SPHHRASCCPN L PAPC GF FI MSS K EPN YHI L Y GYG K Y PL SPHHRASCCPN L PAPC GF FI MSS K EPN YHI L Y GYG K Y PL SPHHRASSCPN L PAPC GF FI MSS K EPN YHI L Y GYG K Y PL RPHRSSCCPN L PAPC GF T I MSS K EPN YHI L Y GYG N Y PL RPHRSSCCPN L PSPC GT QYMPL K EPN YHI L Y GYG N Y PL RPHRSSCCPN L PSPC GT QYMPL K EPN YHI L Y	6 A L 6 A L		D D D N Y V D L R P D E N D I Y Q D K R P D E N D I Y E D K R P L V D D S Y E D K R P E V D D S Y E D K R P E A D D Y A D N R P E A N D N Y R D V R P E A N D N Y R D V R P G Q S D D Y T D D R P G Q S D D Y T D D R P G Q S D D Y T D D R P G Q S D D Y T D D R	372 373 372 373 372 372 372 372 371 394 394 394
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066286.1 XP_013066370.1 XP_013066370.1 XP_013069171.1 XP_013069170.1 XP_013069170.1 XP_013069521.1 XP_013069523.1	TAGMYAFLALLAGENGI DAK - YRKMAVEG INYIL EAANNAFI ALVAAESGI ESK - YRKMAVEG INYIL EAANNAFI ALVAAESGI ESK - YRKMAVEG INYIL EAANNAFLALAAESGI OTR - LRKMAVEG INYIL EAANNAFLALAAAESGI OTR - LRKMAVEG INFIL YAGNSAFLALLAAESGI PR - LRKWAVEG INFIL YAGNSAFLALLAAESGI PR - LRKWAVEG INFIL YSSINAMAALVAADAGLN GO - YRKWAVEG MIYIL YSSINAMAALVAADAGLN GO - YRHWANCGI HYAL YSSINAMAALVAADAGLN GO - YRHWANCGI HYAL YSSINAMAALVAADAGLN SE - YRHWANCGI HYAL YSSINAMAALVAADAGLH SE - YRHWANCGI HYAL	G N N HG S Y V V G N P HG S YQ V G N P HG S YQ V G N N HG S YQ I G N K HG S FQ I G D TG FS YV I G D TG FS YV I G D TG FS YV I	GYGSK YPQOPHHAGASCPD RPASCGWAQYS B PSPN POULH YGSK YPL OPHHRSSC PS PPAOL CNWD T YN LHSPN PHVL N YGTK YPL HPHHRGSSC PD RPAPCD WNN FNS A APN PQU L V YGTK YPL HPHHRGSSC PD KPAPCGWDN FNS FLPN PHVL Q YGTK YPT HPHHRGASC PD KPAPCGWDN FNA TTPS PQU L Y YGK YPT HPHHRGASC PD KPAPCGWDN FNA TTPS PQU L Y YGK YPT HPHHRGASC PD KPAPCG FL NH TTPS PQU L Y YGK YPT UPHHRASC PD KPAPCG FL NH TTPS PQU L Y YGK YPT UPHHRGASC PD KPAPCG FL NH AT TPS PQU L Y YGDK YPL QPHHRASC PD RPAPCSG AQL ND PGPS P0 I L Y YGDK YPL QPHHRASC PD RPAPCSG FL NH AF PSPV L I YGDK YPL QPHHRASSC PN L PAPCG PF I MSSK EPN VH I L Y GFGDK YPL RPHHRSSC PN L PAPCG FG TO YMPL KEPN VH I L Y GFGK YPL RPHHRSSC PN L PSPCG TO YMPL KEPN VH I L Y GFGK YPL RPHHRSSC PN L PSPCG TO YMPL KEPN VH I L Y	6 A L 6 A L 7 A L		F D G ND N Y V D L R           O D E N D I Y Q D K R           O D E N D I Y E D K R           P L V D D S Y E D K R           P L V D D S Y E D K R           P E V D D S Y E D K R           P E A D D S Y A D N R           P E A D D S Y A D N R           P E A N D N Y R D V R           P E S N D N Y D D R           P G S N D D Y T D D R           P G S N D Y T D D R           P G G S N D Y T D D R           P G G N D Y T D D R           P G G N D Y T D D R           P G G N D Y T D D R	372 373 372 373 372 372 372 371 394 394 394 394 394
XP_013066281.1 XP_013066283.1 XP_013066284.1 XP_013066285.1 XP_013066285.1 XP_013066370.1 XP_013066371.1 XP_013069170.1 XP_013069170.1 XP_013069522.1 XP_013069523.1 XP_013069523.1	TAGMYAFLALLAGEMGIDAK - YRWAVEQINYIL EAANNAFIALVAAESGIESK - YRWAVEQINYIL EMANTAFLALAAAESGIESK - YRWAVEQINYIL EMANTAFLALAAAESGIOTR - LRWAVEQINYIL YAGNSAFLALLAAAESGITPR - LRWAVEQINYIL YAGNSAFLALLAAAESGINTD - YRWAVEQINYIL YSSNMAMAALVAADAGLNAD - YRWAVEQINYIL YSSNMAMAALVAADAGLNAD - YRWANCQIHYAL YSSNMAMAALVAADAGLNAD - YRWAMCQIHYAL YSSNMAMAALVAADAGLNSCH YRWAMCQIHYAL YSSNMAMAALVAADAGLNSCH YRWAMCQIHYAL YSSNMAMAALVAADAGLNSCH YRWAMCQIHYAL YSSNMAMAALVAADAGLNSCH YRWAMCQIHYAL	G N NHGS YVV G N PHGS YQV G N PHGS YQV G N TFGS FQ I G N NHGS YQ I G N NHGS YQ I G N NHGS FQ I G N HGS FQ I G D TGFS YL I G D TGFS YV I G D TGFS YV I G D TGFS YV I G D TGFS YV I	GYGSK Y POQPHHAGASCPD RPASCGWAQYS B PSPN PQU H YGSK Y PL PPHHRGSSCP B PPAC NWD T YN L HSPN PHVL N YGTK Y PL HPHHRGSSCP D RPAPCD WNN F NS A PN PQU L Y GYGTK Y PT HPHHRGASCPD K PAPCGWD N F NS FLPN PHVL Q YGTK Y PT HPHHRGASCPD K PAPCGWD N F HA TT PS PQU L Y YGTK Y PT HPHHRGASCPN K PAPCG F L HA TT PS PQU L Y YGTK Y PT HPHHRGASCPN K PAPCG F L HA TT PS PQU L Y YGSK F PRQ PHHRGASCPD K PAPCG F L HA TT PS PQU L Y YGSK F PRQ PHHRGASCPD R PAPCSG AQL ND F PSP VI YGSK F PRQ PHHRGASCPD R PAPCSG AQL ND F PSP VI YGSK F PRQ PHHRGASCPD R PAPCG F L MSSK EPN VH I L Y GFGDK Y PHS PHHRASSCPN L PAPCG F I MSSK EPN VH I L Y GFGDK Y PHS PHHRASSCPN L PAPCG F I MSSK EPN VH I L Y GFGDK Y PL RPHHRSSCPN L PAPCG F I MSSK EPN VH I L Y GFGDK Y PL RPHHRSSCPN L PSPCGT Q VM PL KEPN VH I L Y GFGDK Y PL RPHHRSSCPN L PSPCGT Q VM PL KEPN VH I L Y GFGDK Y PL RPHHRSSCPN L PSPCGT Q VM PL KEPN VH I L Y GFGDK Y PL RPHHRSSCPN P PAKCGS Q I M MKEPS PHV L I Y GFGDK Y PL RPHHRASSCPN P PAKCGS Q I M MKEPS PHV L I Y GFGDK Y PL RPHHRASSCPN P PAKCGS Q I M MKEPS PHV L I Y GFGDK Y PL RPHHRASSCPN P PAKCGS Q I M MKEPS PHV L I Y GFGD K Y PL RPHRASSCPN L PSPCGT Q VM PL KEPN VH I L Y GFGDK Y PL RPHHRASCPN PAKCGS Q I M MKEPS PHV L I Y GFGD K Y PL RPHHRASCPN PAKCGS Q I M MAC A MPN I H I L Y GFGD K Y PL RPH RASCPN PAKCGS Q I M MACAM PN I H I L Y GFGD K Y PL RPH R M NG PN F PAKCGS Q I M MACAM PN I H I L Y GFGD K Y PL R PH NG M NG PN F PAKCGS Q I M MACAM PN I H I L Y GFGD K Y PL R PH NG M NG PN F PAKCGS Q I M MACAM PN I H I L Y GFGD K Y PL R PH NG M NG PN F PAKCGS Q I M MACAM PN I H I L Y GFGD K Y PL R PH NG M NG PN F PAKCGS Q I M MACAM PN I M N F PS F M L Y GFGD K Y PL R P N NG M NG PN F PAKCGS Q I M MACAM PN I H I M NG PN F PN L Y M NG PN F PN L M NG P	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6		F D Q ND N Y U D LR           F D E N D I Y Q D KR           F D E N D I Y Q D KR           F D L D S Y E D KR           F L V D S Y E D KR           F E A D S Y A D NR           F E A D S Y A D NR           F E A N D Y T D VR           F G Q S D D Y T D DR           F G S N D Y T D DR           F G S N D Y T D DR           F G Q N D Y T D DR           F G Q N D Y T D DR           F G Q N D Y T D DR           F G Q N D Y T D DR           F G Q N D Y T D T DR           F G Q N D Y T D T R	372 373 372 373 372 372 372 372 371 394 394 394 394 394 394
XP_013066281.1 XP_013066284.1 XP_013066284.1 XP_013066284.1 XP_013066285.1 XP_013066370.1 XP_013066371.1 XP_013066371.1 XP_013066371.1 XP_01306952.1 XP_01306952.1 XP_01306952.1 XP_01306952.1 XP_01306952.1 XP_01306952.1	TAGMYAFLALLAGENG I DA - Y RKWAVEQ I NY IL EAANNAF I AL VAAESG I E SK - Y RKWAVEQ I NY IL EMANTAFLALLAAESG I E SK - Y RKWAVEQ I NY IL EMANTAFLALLAAESG I OTR - L RCWAVQO I NY IL EAANVAFLALLAAAESG VD SR - L RKWAVEQ I NF IL YAGNSAFLALLAAAESG I NTD - Y RKWAVEQ I NY IL YAGNSAFLALLAAAESG I NTD - Y RKWAVEQ I NY IL YSSNMAMAAL VAADAGL NAQ - Y RHWAMCQ I HY AL YSSNMAMAAL VAADAGLN AQ - Y RHWAMCQ I HY AL YSSNMAMAAL VAADAGLN BE - Y RHWAMCQ I HY AL YSSNMAMAAL VAADAGLH SE - Y RHWAMCQ I HY AL YASNMAFAALLAADTGLH PE - Y RHWAMCQ I HY AL SAAMAMLALI AADAGLH AE - Y RHWAMCQ I HY AL	G N N H G S Y V V G N P H G S Y V V G N T F G S F Q I G N N H G S Y Q I G N N H G S Y Q I G N N H G S Y Q I G N N H G S F Q I G D T G F S Y L I G D T G F S Y V I G D T G F S Y V I G D T G F S Y V V G D T G F S Y V V	GYGSK YPQOPHHAGASCPD RPASCGWAQYS B PSPN POULH YGSK YPL PPHHRGSSCPD RPAPCOWNN FNS AAPN POULU YGTK YPL HPHHRGSSCPD RPAPCOWNN FNS AAPN POULU UGTK YPL HPHHRGASCPD KPAPCGWDN FNS FLPN PHULQ YGTK YPT HPHHRGASCPD KPAPCGWDN FNS FLPN PHULQ YGTK YPT HPHHRGASCPN KPAPCGWDN FNA TTPS POULU YGTK YPT HPHHRGASCPN KPAPCG GAULNA TTPS POULU YGTK YPT HPHHRGASCPN KPAPCG GAULNA TTPS POULU YGTK YPT HPHHRGASCPN KPAPCG FINSK SKEPN VHI YGDK YPL OPHHRAASCPD RPAPCG FINSK SKEPN VHI FGDK YPHSPHHRASSCPN LPAPCG FINSK SKEPN VHI FGOK YPHSPHHRASSCPN LPAPCG FINSK SKEPN VHI FGOK YPL SPHHRASSCPN LPAPCG FINSK SKEPN VHI FGOK YPL RPHHRSSCPN LPSPCG TO VMPL KEPN VHI LY FGOK YPL RPHHRSSCPN LPSPCG TO VMPL KEPN VHI LY FGOK YPL RPHHRASSCPN LPSPCG TO VMPL KEPN VHI LY FGOK YPL RPHHRASSCPN LPSPCG TO VMPL KEPN VHI LY FGOD YPL RPH HRASSCPN PAKCGSOI MPM KEPS PHVL Y FGOD YPL RPY OR ISMCPN HPAHCG SE VLA AMPN I HI LY	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6		FD GNDNYUDLR           SP DENDIYQDKR           SP DENDIYQDKR           SP DENDIYQDKR           PE VDDSYEDKR           PE VDDSYEDKR           PE ADDSYADNR           PE ADDSYADNR           PE ANDNYRDYR           PE GSSDDYDDR           PG GSSDDYDDR           PG GSNDYDDR           PG GSNDEYDDR           PG GSNDEYLDKF           PG GANDEYLDKF           PG GANEYLDKF	372 373 372 373 372 372 371 394 394 394 394 394 394 394 394
XP_013066281.1 XP_013066284.1 XP_013066284.1 XP_013066284.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013068190.1 XP_013068190.1 XP_013068170.1 XP_013068170.1 XP_01306822.1 XP_01306822.1 XP_013084281.1 XP_01308428.1	TAGMYAFLALLAGEMGIDAR - YRWMAVEGINYIL EAANNAFIALVAAESGIESK - YRWMAVEGINYIL EAANNAFIALVAAESGIESK - YRWMAVEGINYIL EAANTAFLALAAESGIOTR - LRWMAVEGINYIL EAANTAFLALAAAESGIOTR - LRWMAVEGINFIL YAGNSAFLALLAAESGIPR - LRWWAVEGINFIL YSSIMAMAALVAADAGLING - YRWMAVEGINYIL YSSIMAMAALVAADAGLNG - YRWMAVEGINYAL YSSIMAMAALVAADAGLNG - YRWMAVEGINYAL YSSIMAMAALVAADAGLNG - YRWMAVEGINYAL YSSIMAMAALVAADAGLNG - YRWMAVEGINYAL YSSIMAMAALVAADAGLNG - YRWMAVEGINYAL YSSIMAMAALVAADAGLNG - YRWMAVEGINYAL YSSIMAMAALVAADAGLNS - YRWMAVEGINYAL SSIMAMAALVAADAGLNS - YRWMAVEGINYAL SANNAMLALIAADAGLNS - YRWMAVEGINYAL SANNAMLALIAADAGLNS - YRWMAVEGINYAL SANNAMLALIAADAGLNS - YRWMAVEGINYAL	G N N H G S Y V V G N P H G S Y Q V C N T F G S F Q I G N H G S Y Q I G D T G F S Y L I G D T G F S Y L I G D T G F S Y V I	Y GSK Y POQPHHAGASCPD RASCGWAQYS SPSPN POVLH Y GSK Y PL PPHHRGSSCP SPPACC NWD T YN LHSPN PHVL N Y GT K Y PL HPHHRGSSCP D RPAPC DWNN FNS A APN POVL V Y GT K Y PL HPHHRGSSCP D KPAPC DWNN FNS A PN POVL V Y GT K Y PT HPHHRGASCPD K PAPC GWD N FHA T TPS POVL V Y GT K Y PT HPHHRGASCP D K PAPC GWD N FHA T TPS POVL V Y GT K Y PT HPHHRGASCP D K PAPC GY A ALT TPS POVL V Y GT K Y PT L PHHRGASCP D K PAPC GY A ALT PS POVL Y Y GT K Y PT L PHHRGASCP D K PAPC GY A ALT PS POVL Y Y GD K Y PL O PHHRA SC C PD R PAPC SG A Q L ND POPS PO I L Y Y GD K Y PL O PHHRA SSC PN L PAPC G PF I MS SK EPN VH I L Y F GD K Y PL SP HHRASSC PN L PAPC GY T MS SK EPN VH I L Y F GD K Y PL R PHHRSSSC PN L PS PC GT Q V M PL K EPN VH I L Y F GS K Y PL R PHHRSSSC PN L PS PC GT Q V M PL K EPN VH I L Y F GS K Y PL R PHHRSSSC PN PAPL GS GS O I M M K EPS PHVL I S F GD Y PL R PHHRSSSC PN PAPL G S D I L Q A D M PN I H I L Y F GD D Y PL K PY Q R I SMC PN H PAHC G S D I L Q A D M PN I H I L Y F GD D Y PL K PY Q R I SMC PN H PAHC G S D I L Q A D M PN I H I L Y F GD D Y PL R PY HR MSMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY HR MSMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY HR MSMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY HR MSMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY V R I SMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY HR MSMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY HR MSMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY HR MSMC PN H PAHC G F D I L Q A D M PN I H I L Y F GD D Y PL R PY HR M SMC PN H PAHC G F D I L Q A S M PN Y H I L Y F GD D Y PL R PY HR M SMC PN H PAHC G F D I L Q A S M PN Y H I L Y F GD D Y PL R PY H R M PN D PN H PAHC G F D I L Q A S M PN Y H I L Y F GD D Y PL R PY H R M PN D PN H PAHC G F D I L Q A S M PN Y H I L Y F GD D Y PL R PY H R M PN D PN H PN H C T P H L A A K L PN Y H I L Y F GD D Y PL R PN H R A S C PN PN H PN H C T P H L A A K L PN Y H I L Y F GD D Y PL R PN H R A S C PN H PN H C T P H L A A K L PN Y H I L Y	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6		F D G ND N Y V D LR           G D E N D I Y G D KR           G D E N D I Y E D KR           F L V D D S Y E D KR           F E V D D S Y E D KR           F E V D D S Y E D KR           F E A D D S Y A D NR           F E A D D S Y A D NR           F E A D S S D Y T D DR           F G G S D D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G S N D Y T D DR           F G G N D E Y V D K F           F G A N D E Y V D K F           F G A N D E Y V D K F           F G O N D Y Y D C F           F G O N D Y Y D C F           F G O N D Y Y D K F           F G O N D Y Y D K F	372 373 372 372 372 371 394 394 394 394 394 394 394 394
XP_013066281.1 XP_013066284.1 XP_013066284.1 XP_013066284.1 XP_013066285.1 XP_013066370.1 XP_013066370.1 XP_013066370.1 XP_013066170.1 XP_013066170.1 XP_013066523.1 XP_013066523.1 XP_013064524.1 XP_013064282.1 XP_013064282.1 XP_013064282.1	TAGMYAFLALLAGEMGIDAK - YRWAVEG NYL EAANNAFIALVAAESGIESK - YRWAVEG NYL EAANNAFIALVAAESGIESK - YRWAVEG NYL EAANNAFLALAAAESGIOTR - LRWAVEG NYL EAANVAFLALAAAESGVDSR - LRWAVEG NYL YAGNSAFLALLAAESGVDSR - LRWAVEG NYL YAGNSAFLALLAAESGINTD - YRWAVEG NYL YSSNMAMAALVAADAGLNAG - YRWAVEG NYL YSSNMAMAALVAADAGLNAG - YRWANCG NYA YSSNMAMAALVAADAGLNAG - YRWAMCG NYA YSSNMAMAALVAADAGLNAG - YRWAMCG NYA YSSNMAMAALVAADAGLNSCH YRWAMCG NYA YSSNMAMAALVAADAGLNAG - YRWAMCG NYA YSSNMAMAALVAADAGLNAG - YRWAMCG NYA YSSNMAMAALVAADAGLNAG - YRWAMCG NYA SSNMAMAALVAADAGLNAG - YRWAMCG NYA SSNMAMAALVAADAGLNAG - YRWAMCG NYA SSNMAMAALVAADAGLNAG - YRWAMCG NYA SAANMAMLALAADAGLNAG - YRWAMCG NYA AASNMAMLALIAADAGLNAG - YRWAMCG NYA AAANMAMLALIAADAGLNAG - YRWAMCG NYA	G N H H G S Y V V G N P H G S Y Q V G N T H G S Y Q V G N T H G S Y Q I G N N H G S Y Q I G D T G F S Y V V G D T G F S Y V V G D T G F S Y V V G D T G F S Y V V	Y Y SK Y P Q P HH AGAS C P D R P A S C G WA Q Y S P S P N P Q V H Y SK Y P L P P HH R G S S C P D R P A C C M W T Y N L H S P N P H V L N Y G T K Y P L P P HH R G S S C P D R P A P C D W N F N S A P N P Q V V Y G T K Y P L Q P HH R A S C P D K P A P C G W D N F N S T A P N P Q V V Y G T K Y P L Q P HH R A S C P D K P A P C G W D N F N S T A P N P Q V V Y G T K Y P T H P HH R G S C P D K P A P C G W D N F H A T T P S P Q V V Y G T K Y P T H P HH R G S C P D K P A P C G W D N F H A T T P S P Q V V Y G T K Y P T H P HH R G S C P D K P A P C G F N L H A T T P S P Q V V Y G D K Y P L Q P HH R A S C P D R P A P C G F N L H A T T P S P Q V V Y G D K Y P L Q P HH R A S C P N L P A P C G F C I M S K E P N V H L Y G F G D K Y P H S P HH R A S C P N L P A P C G F F I M S K E P N V H I L Y G F G D K Y P H S P HH R A S C P N L P A P C G F I M S K E P N V H I L Y G F G D K Y P L R P HH R S S C P N L P A P C G F I M S K E P N V H I L Y G F G D K Y P L R P HH R S S C P N L P A P C G F I M S K E P N V H I L Y G F G D K Y P L R P H H R S S C P N L P A C G G T Q M P L K E P N V H I L Y G F G D Y P L R P H H R S S C P N P A H C G S O I M M K E P S P H V L Y G F G D Y P L R P H R S S C P N P A H C G S O I M M K E P S P H V L Y G F G D Y P L R P Y A R I S M C P N H A H C S P E V L L A E M P N I H I L Y G F G D Y P L R P Y H R M M C P N H P A L C T F E H L A A K L P N V H I L Y G F G D Y P L R P Y H R M M C P N H P A I C T F E H L A A K L P N V H I L Y G F G D Y P L R P Y H R M S C P F L P H C G P M I N I T F P N V H I L Y G F G D Y P L R P Y H R M S C P F L P H C G P M I N I T F P N V H I L Y G F G D Y P L R P Y H R M S C P F L P A H C G P M I N I T F P N V H I L Y G F G D Y P L R P Y H R S S C P F L P H C G C P M I N I T F P N V H I L Y G F G D Y P L R P Y H R S S C P F L P A H C G P M I N I T F P N V H I L Y G F G D Y P L R P H H R S S C P F L P A H C G P M I N I T F P N V H I L Y G F G D Y P L R P H H R S S C P F L P A H C G P M I N I T F P N V H I L Y G F G D Y P L R P H H R S S C S C P F L P A H C G C P M I N I T F P N V H I Y			P G ND N Y UDLR           P D E N D I Y G D K R           P L V D S Y E D K R           P L V D S Y E D K R           P E V D S Y D S Y D D K R           P E A D S Y A D N R           P E A D S Y A D N R           P E A D S Y A D N R           P E A N D N Y D V R           P E A S N D Y T D R           P G S N D Y T D D R           P G S N D Y T D D R           P G S N D Y T D D R           P G S N D Y T D D R           P G G N D Y T D D R           P G S N D Y T D D R           P G S N D Y T D D R           P G S N D Y T D T D R           P G S N D Y T D T R           P G S N D Y T D T R           P G S N D Y T D R           P G S N D Y T D R           P G S N D Y T D R           P G S N D Y T D R           P G S N D Y T D R           P G S N D Y T D R           P G S N D Y T D R           P G S N D Y T D R           P G S N D Y Y D F           P G Y N D Y V D F           P G Y D D Y V D F           P G Y D D Y Y D F           P G Y D D Y Y D F           P G Y D D Y Y D F           P G Y D D Y Y D F	372 373 372 372 372 372 372 371 394 394 394 394 394 394 394 394 394 394
XP_01306628.1 XP_01306628.1 XP_01306628.1 XP_01306628.1 XP_01306628.1 XP_013066370.1 XP_013066370.1 XP_013066190.1 XP_013066190.1 XP_013066192.1 XP_01306652.1 XP_01306652.1 XP_01306628.1 XP_013064279.1 XP_01306428.1 XP_01306428.1 XP_01306428.1	TAGMYAFLALLAGEMG I DA - Y RKWA VEQ I NY IL EAANNAF I AL VAAESG I E SK Y RKWA VEQ I NY IL EMANTAFLALLAAESG I E SK Y RKWA VEQ I NY IL EMANTAFLALLAAESG I OTR - L RCWA VQO I NY IL EAANVAFLALLAAAESG VD SR - L RKWA VEQ I NF I L YAGNSAFLALLAAAESG I NTD - Y RKWA VEQ I NY IL YAGNSAFLALLAAESG I NTD - Y RKWA VEQ I NY IL YSS NIAMAAL VAADAGL NAQ - Y RKWA VEQ I NY IL YSS NIAMAAL VAADAGL NAQ - Y RHWA MCQ I HY AL YSS NIAMAAL VAADAGL NAQ - Y RHWA MCQ I HY AL YSS NIAMAAL VAADAGL NAQ - Y RHWA MCQ I HY AL YSS NIAMAAL VAADAGL H SE - Y RHWA MCQ I HY AL YSS NIAMAAL VAADAGL H SE - Y RHWA MCQ I HY AL YASN MAFAAL LAADTGL H SE - Y RHWA MCQ I HY AL YASN MAFAAL LAADTGL H SE - Y RHWA MCQ I HY AL YASN MAFAAL LAADTGL H SE - Y RHWA MCQ I HY AL AASAN MAMLAL I AADAGL H SE - Y RHWA MCQ I HY AL HAANMAMLAL I AADAGL H TE - Y RHWA MCQ I HY AL HAANMAMLAL I AADAGL H TE - Y RHWA MCQ I HY AL HAANMAMLAL I AADAGL H TE - Y RHWA MCQ I HY AL HAANAMALAL I AADAGL H TE - Y RHWA MCQ I HY AL	G N N HG S Y V V G N P HG S Y Q V G N T HG S Y Q V G N T HG S Y Q I G N N HG S Y Q I G N N HG S Y Q I G N N HG S Y Q I G D T G F S Y L I G D T G F S Y V I G D T G F S Y V I G D T G F S Y V V G D T G F S Y V V	Y GSK Y PDQ PH HAGAS C PD R PASC GWAQY S P S P N PQ V L H Y GSK Y PL PP HHRGSS C P S P PAC C NWD T YN L HSP N PH V L N Y GT K Y PL HP HHRGSS C PD R PAP C D WNN F NS A PN PQ V L V VGT K Y PL HP HHRGSS C PD K PAP C G WD N F NS T A P N PQ V L V Y GT K Y PT HP HHRGAS C PD K PAP C G WD N F NS T T PS PQ V L Y G YG K Y PT HP HHRGAS C PD K PAP C G WD N F NA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD K PAP C G F E N L HA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD K PAP C G F E N L HA T T PS PQ V L Y G YG K Y P L Q P HHRA ASC PD R PAP C G F E N L HA T T PS PQ V L Y G YG K Y P L Q P HHRA S C PD R PAP C G F E N L M T PS PN V I L Y G YG K Y P L Q P HHRA S C PD R PAP C G F I M S K E PN V HI L Y G F G K Y P H S P HHRA S C PN L PAP C G F I M S K E PN V HI L Y G F G K Y P L R P HHRS S C PN L PAP C G T O Y M P L K PN V HI L Y G F G K Y P L R P HHR S S C PN L PAP C G T O Y M P L K PN V HI L Y G F G K Y P L R P H HRA S C PN P PAK C G G O I M PM K E PS P HV L Y G F G D Y P L R P Y G R I S M C PN H PAH C S P E V L A EMP N H I L Y G F G D Y P L R P Y G R I S M C PN H PA H C S P E V L A EMP N H I L Y G F G D Y P L R P Y A R I S M C PN H PA H C S P E V L A EMP N Y H I L Y G F G D Y P L R P Y A R I S M C PN H PA H C S P E V L A EMP N Y H I L Y G F G D Y P L R P Y A R I S M C PN H PA H C S P E V L A EMP N Y H I L Y G F G D Y P L R P Y A R I S M C PN H PA H C S P E V L A EMP N Y H I L Y G F G D Y P L R P Y H R S S C P F L PA H C G P M L N L T E PN Y W I L Y G F G D Y P L R P Y H R S S C P F L PA H C G P M L N L T E PN Y W I L Y G F G D Y P L R P Y H R S S C P F L PA H C G P M L N L T E PN Y W T L Y G F G D Y P L R P Y H R S S C P F L PA H C G P M L N L T E PN Y W T L Y G F G D Y P L R P H H R S S C P F L PA H C G P M L N L T E PN Y W T L Y G F G D Y P L R P H H R S S C P F L PA H C G P M L N L T E PN Y W T L Y G F G D Y P L R P H H R S S C P F L PA H C G P M L N L T E PN Y W T L Y			P         P         D         N         V         D         P         D         D         N         D         K         D         D         K         D         D         N         D         D         D         N         D         D         D         N         D	372 373 372 373 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
AP_013066281           AP_013066281           AP_013066281           AP_013066281           AP_013066281           AP_013066281           AP_013066371           AP_013066371           AP_013066371           AP_013066371           AP_013066371           AP_013066371           AP_013066371           AP_013066371           AP_013066371           AP_013066321           AP_013066321           AP_013066321           AP_013064281           AP_013084281           AP_013084281           AP_013084281           AP_013084281           AP_013084281           AP_013084281           AP_013084281           AP_013084281           AP_013084281           AP_0130842851           AP_0130842851           AP_030842851           AP_030842851           AP_030842851	TAGMYAFLALLAGEMGIDAR - YRWAVEG NYFL EAANNAFIALVAAESGIESK - YRWAVEG NYFL EAANNAFIALVAAESGIESK - YRWAVEG NYFL EAANTAFLALAAAESGIOTR - LRWAVEG NYFL EAANTAFLALAAAESGIOTR - LRWAVEG NFIL YAGNSAFLALLAAESGIPR - LRWAVEG NFIL YSSMAMAALVAADAGIPR - YRWAVEG NYFL YSSMAMAALVAADAGING - YRWAVEG NYFL YSSMAMAALVAADAGLNG - YRWAVEG NYFL YSSMAMAALVAADAGLNS - YRWAVEG NYFL YSSMAMAALVAADAGLNS - YRWAVEG NYFL SSMAMAALVAADAGLNS - YRWAVEG NYFL SSMAMAALVAADAGLNS - YRWAVEG NYFL ASSMASAALVAADAGLNS - YRWAVEG NYFL ASSMASAALVAADAGLNS - YRWAVEG NYFL ASSMASALLAADAGLNS - YRWAVEG NYFL AANNAALA AAAAGGLNS - YRWAVEG NYFL HAANMAMLALIAADAGLNS - YRWAVEG NYFL HAANALALVALVAADAG NTE - YRWAVEG NYFL	G N HIGS Y V V G N PIGS Y Q V G N PIGS Y Q V G N TIGS FQ I G N HIGS Y Q I G N HIGS Y Q I G N HIGS Y Q I G N HIGS FQ I G D TGFS Y L I G D TGFS Y V I G D TGFS Y V VI G D TGFS Y V VI	Y GSK Y POQPHHAGASCPD RASCGWAQYS SPSPN POVLH Y GSK Y PL PPHHRGSSCPD RPAPC DWN F NS A PPN POVL V Y GT K Y PL HPHHRGSSCPD RPAPC DWN F NS A PPN POVL V Y GT K Y PT HPHHRGSSCPD RPAPC DWN F NS FI APN POVL V Y GT K Y PT HPHHRGASCPD K PAPC GWD N F HA TT PS POVL V Y GT K Y PT HPHHRGASCPD K PAPC GWD N F HA TT PS POVL V Y GT K Y PT HPHHRGASCPD K PAPC GG F L HA TT PS POVL V Y GT K Y PT HPHHRGASCPD K PAPC GG F L HA TT PS POVL V Y GT K Y PT L PHHRGASCPD K PAPC GG F L HA PC PS PV L I S FOK Y PHS PHHRGASCPD K PAPC GG F L MS SKEPN VH I L Y G FGD K Y PHS PHHRASSCPN L PAPC GF F I MS SKEPN VH I L Y G FGD K Y PL SP HHRASSCPN L PAPC GF I MS SKEPN VH I L Y G FGD K Y PL RP HHRSSSCPN L PSPC GT QV MP L KEPN VH I L Y G FGD K Y PL RP HHRSSCPN PAPL GS GA I L A BMP N I HI L Y G FGD K Y PL RP HHRSSCPN PAPL G SG I I M MKEPS PHV L I G FGD Y PL RP HHRSSCPN PAPL G SG I U A D MN KEPS PHV L I F GOD Y PL RP Y R MS MC PN H PA H C S P V L L A EMP N I HI L Y G FGD D Y PL R Y HR MS MC PN H PA I C T PE HL A A KL PN V HI L Y G FGD D Y PL R P H HRSSCP FL PA HC G P PM L N L T EPN V H T L Y G FGD K Y PL R PH HRSSCP FL PA HC G P PM L N L T EPN V H T L Y G FGD Y PL R P H HR SSC P FL PA HC G P PM L N L T EPN V H T L Y G FGD Y PL R P H HR SSC P FL PA HC G SG SH GN PN V Y T L Y G FGD Y PL R P H HR SSC P FL PA HC G SG SH GN PN P V H T L Y G FGD Y PL R P H HR SSC P FL PA HC G SG SH GN PN P V H T L Y G FGD Y PL R P H HR SSC P FL PA HC G SG SH GN FN P V Y T L Y G FGD K Y PL R P H HR SSC P FL PA HC G SG SH GN FN P V Y T L Y G FGD K Y PL R P H HR SSC P FL PA HC G SG SH GN SH PN Y T L Y G SG K F P R P H HR SSC P FL PA HC G SG SH GN SH PN Y T V T L Y G SG K Y PL R P H HR SSC P FL PA HC G SG SH GN SH PN Y T V T L Y G SG K Y PL R P H HR SSC P FL PA HC G SG SH GN SH PN Y T V T L Y G SG K Y P R P H HR SSC P FL PA HC G SG SH GN SH PN Y T V T L Y G SG K Y P C P H HR SSC P FL PA HC G SG SH GN SH PN Y T V T L Y G SG K Y P R P H HR SSC P N R PA FL G SG SA G SH GN H PN Y T V T L Y G SG K Y P R P H HR SSC P FL PA HC G SG SH GN SH M PN Y T V T Y SK SK F P R P H H SSC P N T A T C F			F D Q ND N Y V D LR           SP D E N D I Y Q D KR           SP D E N D I Y E D KR           SP L V D D S Y E D KR           SP E V D D S Y E D KR           SP E V D D S Y E D KR           SP E V D D S Y E D KR           SP E A D D S Y A D NR           SP E S ND D Y T D DR           SP G Q S D D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D DR           SP G S ND D Y T D R           SP G S ND D Y T D R           SP G S ND D Y N D F           SP G P D S Y S D VR           SP G D S Y S D VR A LR           SP D D S Q Y A LR	372 373 372 373 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306627.1           XP_01306637.1           XP_01306657.1           XP_01306657.1           XP_01306657.1           XP_01306657.1           XP_01306657.1           XP_01306657.1           XP_01306422.1           XP_01308428.1           XP_01308438.1	TAGMYAFLALLAGEMGIDAR - YRWAYUG NHYL EAANNAFIALVAAESGIESK - YRWAYUG NHYL EAANNAFIALVAAESGIESK - YRWAYUG NYIL EAANYAFLALAAAESGIOTR - LRWAYUG NYIL EAANYAFLALAAAESGIOTR - LRWAYUG NYIL EAANYAFLALAAAESGIOTR - LRWAYUG NYIL YAGNSAFLALLAAAESGINTD - YRWAYUG NYIL YSSNMAMAALVAADAGLNAD - YRWAYUG NYIL YSSNMAMAALVAADAGLNAD - YRWAYUG NYAL YSSNMAMAALVAADAGLNAD - YRWAYUG NYAL ASAMAALAAADAGLNAD - YRWAYUG NYAL ASAMAALAAADAGLNAG - YRWAYUG NYAL ASAMAALLAADAGLNAG - YRWAYUG NYAL AANNAMLALIAADAGLNAG - YRWAYUG NYAL HAANNAMLALIAADAGLNAG - YRWAYUG NYAL HAANNAMLALIAADAGLNTE - YRWAYUG NYAL	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N T H G S Y Q I G N H H G S Y Q I G N H H G S Y Q I G D T G F S Y V V G D T G F S Y V V G D T G F S Y V V G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V I	Y Y SK Y P Q P HH AGAS C P D R P A S C G WA Q Y S P S P N P Q V H Y S K Y P L P P HH R G S S C P D R P A C D WN N F N S A P N P Q V V Y G T K Y P L H P HH R G S S C P D R P A P C D WN F N S T A P N P Q V V Y G T K Y P L Q P HH R A S C P D K P A P C G WD N F N S T A P N P Q V V Y G T K Y P L Q P HH R A S C P D K P A P C G WD N F N S T A P N P Q V V Y G T K Y P T H P HH R G S C P D K P A P C G WD N F H A T T P S P Q V V Y G T K Y P T H P HH R G S C P D K P A P C G W D N F H A T T P S P Q V V Y G T K Y P L Q P HH R A S C P D R P A C G F N L H A T T P S P Q V V Y G T K Y P L Q P HH R A S C P D R P A C G F N L H A T P S P Q V V Y G K Y P L Q P HH R A S C P N L P A P C G F N L H A T P S P N V I L Y G G D K Y P H S P HH R G S C P N L P A P C G F I M S K E P N V H I L Y G G D K Y P H S P HH R S S C P N L P A P C G F I M S K E P N V H I L Y G G D K Y P L R P HH R S S C P N L P A P C G F I M S K E P N V H I L Y G G D K Y P L R P HH R S S C P N L P A P C G T Q Y M P L K E P N V H I L Y G F G D K Y P L R P H H R S S C P N P A H C G S O I M M K E P S P H V L I G F G D Y P L R P Y A R S C P N P A H C G S D I L Q A D M P N H I L Y G F G D Y P L R P Y A R I M M C P N H P A I C T P E V L A A K L P N V H I L Y G F G D Y P L R P Y H R M S C P F L P A H C G P P M L N L T E P N V H I L Y G F G D Y P L R P Y H R M S C P F L P A H C G P M L N L T E P N V H I L Y G F G D Y P L R P Y H R M S C P F L P A H C G P M L N L T E P N V H I L Y G F G D Y P L R P Y H R S S C P F L P A H C G P M L N L T E P N V H L Y G F G D Y P L R P H H R S S C P F L P A H C G P M L N L T E P N V H L Y G F G D Y P L R P H H R S S C P F L P A H C G P M L N L T E P N V H L Y G G G X Y P L R P H H R S S C P F L P A H C G P M L N L T E P N V H L Y Y G S K F P R Q P H H R A S C P N R P A P C G S A D A Q S H G P N V Y T Y Y G S K F P R Q P H H R A S C P D R P A P C S H D Q L N A P G P S P O V L Y	G G G G G G G G G G G G G G G G G G G		P G ND N Y UDLR           P D E N D I Y E D K R           P L V D S Y E D K R           P L V D S Y E D K R           P E V D S Y D S Y D D K R           P E A D S Y A D N R           P E A N D N Y D D R           P E A N D N Y D D R           P E A S D S Y D D N R           P G G S D D Y T D D R           P G G N D D Y T D D R           P G G N D D Y T D D R           P G G N D D Y T D D R           P G G N D D Y T D D R           P G G N D D Y T D D R           P G G N D Y Y D D R           P G G N D Y Y D D R           P G G N D Y Y D D R           P G G N D Y Y D D R           P G Q D S Y S D Y R           P G P D D S Y S D Y R           P G P D D S Y S D Y R           P D D S O Y A D L R           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Y S D Y R	372 373 372 373 372 372 372 374 394 394 394 394 394 394 394 394 394 39
AP_013066281,1           AP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066371,1           XP_013066371,1           XP_013066170,1           XP_013066522,1           XP_013066522,1           XP_01306628,1           XP_01306622,1           XP_01306428,1           XP_01306428,1           XP_01304428,1           XP_01304451,1           XP_01304461,1           XP_01448,1           XP_01448,1	TAGMYAFLALLAGENG IDAK - YRKWAVEG INYIL         EAANNAF IALVAAESG IE SK - YRKWAVEG INYIL         EAANNAF IALVAAESG IE SK - YRKWAVEG INYIL         EMANTAFLALLAAESG IT FR - LRCWAVEG INYIL         EAANNAF IALVAAESG IT FR - LRCWAVEG INYIL         EAANNAF LALLAAAESG IT FR - LRCWAVEG INYIL         YAGNSAFLALLAAAESG INT D - YRKWAVEG INYIL         YASNMAMAALVAADAGLNAG - YRHWAMCG IHYAL         YSSNMAMAALVAADAGLNAG - YRHWAMCG IHYAL         YSSNMAMAALVAADAGLNAG - YRHWAMCG IHYAL         YSSNMAMAALVAADAGLNE - YRHWAMCG IHYAL         YASNMAFAALLAADTGLHE - YRHWAMCG IHYAL         YASNMAFAALLAADAGLH FE - YRHWAMCG IHYAL         YASNMAFAALLAADAGLH TE - YRHWAMCG IHYAL         HAANMALALI AADAGLH TE - YRHWAMCG IHYAL         HAANMAMLALI AADAGLH TE - YRHWAMCG IHYAL         HAANMAMLALI AADAGLH TE - YRHWAMCG IHYAL         HAANAMLALI AADAGLH TE - YRHWAMCG IHYAL         HAANAK ALVALVAADAGI INTE - YRHWAMCG IHYAL         HAANAK ALVALVAADAGI INTE - YRHWAMCG IHYAL         HAANAK ALVALVAADAGI INTE - YRHWAMCG IHYAL         HAANAK ALVALVAADAG IN	G N H H G S Y V V G N P H G S Y Q V G N T H G S Y Q V G N T H G S Y Q V G N T H G S F Q I G N H H G S Y Q I G N H H G S F Q I G D T G F S Y V V G D T G F S Y V I G D T G F S Y V I	Y GSK Y PLOP HH AGASC PD D RASC GWAQYS S PS PN PQ V L H Y GSK Y PL PP HH RGSS C PD R PA PC D WNN F NS A AP N PQ V L Y Y GT K Y PL HP HH RGSS C PD R PA PC D WNN F NS A AP N PQ V L Y VGT K Y PL HP HH RGSS C PD R PA PC G WD N F NS T AP N PQ V L Y Y GT K Y PL HP HH RGASC PD K PA PC G WD N F NS PL PN PV V L Q Y GT K Y PT HP HH RGASC PD K PA PC G WD N F NA TT PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PA PC G F E N L HA TT PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PA PC G F E N L HA TT PS PQ V L Y G YG K Y PL Q PH HR A SC PD R PA PC G F E N L HA TT PS PV V H Y GD K Y PL Q PH HR ASS C PD R PA PC G F I MS S K E PN V H I L Y F GD K Y PH SP HH RGSS C PN L PA PC G F I MS S K E PN V H I L Y F GD K Y PL SP HH RASSC PN L PA PC G F I MS S K E PN V H I L Y F GD K Y PL R PH HR SS C PN L PA PC G T O YM PL K E PN V H I L Y F GD K Y PL R PH HR SS C PN L PA PC G T O YM PL K E PN V H I L Y F GD K Y PL R PH HR SS C PN L PA PC G T O YM PL K E PN V H I L Y F GD K Y PL R PH HR SS C PN L PA PC G T O YM M K E PS PH VL Y G GD D Y PL R PY AR I SMC PN HP A H C S PE V L L A EMPN I H I L Y F GD Y PL R PY AR I SMC PN HP A H C S PE V L L A EMPN I H I L Y F GD Y PL R PY HR MS SC P F L PA H C G PM L N L T E PN V W I L Y F GD Y PL R PY HR MS SC P F L PA H C G PM L N L T E PN V H I L Y F GD X Y PL R PY HR MS SC P F L PA H C G PM L N L T E PN V H I L Y F GD X Y PL R PH HR SS C P F L PA HC G PM L N L T E PN V H I L Y F GD X Y PL R PH HR SS C P F L PA HC G PM L N L T E PN V H I L Y Y GS K F PR Q P HHR A SC P D R PA PC G S AD A Q S H G PN Q V T Y WG R D PO Q P HHR A A SC P D R PA PC G S M A Q S H G PN Q V T Y WG R D PO Q Q PH HR A A SC P D R PA PC G S Y C N Q N A PG PS PO V V WG R D PO Q Q PH HR A A SC P D R PA PC G S Y E N A A PG PS PO V V Y	G G G G G G G G G G G G G G G G G G G		P         P         P         D         N         V         D         P         D         D         N         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         K         D         D         K         D         D         K         D	372 373 372 373 372 372 372 372 372 372
μP         013066281.1           μP         01306281.1           μP         01306281.1           μP         01306281.1           μP         01306281.1           μP         01306281.1           μP         01306371.1           μP         01306837.1           μP         01306817.0           μP         01306817.1           μP         01306817.1           μP         01306828.1           μP         01306428.1           μP         01306428.1           μP         01306428.1           μP         01306428.1           μP         01306428.1           μP         01306428.1           μP         01476612.1           μP         01476612.1	TA GINYA FLA LLAGE BIG I DA - Y RKWA VE G I NY IL EA ANNA FI AL VA AESG I E SK Y RKWA VE G I NY IL EA ANNA FI AL VA AESG I E SK Y RKWA VE G I NY IL EA ANNA FI AL LA AESG I O TR - L RCWA VE G I NY IL EA ANNA FI AL LA AESG VD SR - L RKWA VE G I NY IL Y ASNA ANA AL VA ADAG I NP G - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG I NP G - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG I ND G - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG I ND G - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG I ND G - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN G - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN G - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN SE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LH SE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LH SE - Y RHWA MCG I HY AL A ANNA AL I AADAG LH FE - F RHWA MCG I HY AL HAA NIAMILAL I AADAG LH TE - Y RHWA MCG I HY AL HAA NIAMILAL I AADAG LH TE - Y RHWA MCG I HY AL HAA NIAMILAL I AADAG LH TE - Y RHWA MCG I HY AL HAA NIAMILAL I AADAG LH TE - Y RHWA MCG I HY AL HAA NIAMILAL I AADAG LH TE - Y RHWA MCG I HY AL HAANI AL VAL VAADAG I HTE - Y RHWA MCG I HY AL HAANI AL VAL VAADAG I S SI S Y RHWA MCG I HY AL HAANI AL VAL VAADAG I S SI SA T - Y RKWA VE G I NY IL Y AG NIAL J AL LA ADAG I S SI S Y RKWA VE G I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG G I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN TAL Y ALLTA EAG I S AT - Y RA WANG O I NY IL Y AGN NA FY CLO AAD O G D E Y RA WANG Y D Y WI Y AGN NA FY CLO AAD O G D C NY WI Y W Y W Y Y Y Y Y Y Y Y Y Y Y Y Y Y	G N HIGS Y V V G N PHGS Y Q V G N THGS Y Q V G N THGS Y Q V G N THGS Y Q I G N HIGS Y Q I G N HIGS Y Q I G N HIGS Y Q I G D TGFS Y V I G D TGFS Y V I G D TGFS Y V V G N FIGS Y C S TG R S V V G N G S G S V V V C D TGFS Y V V C D TGFS Y V V V V V V V V V V V V V V V V V V	Y GSK Y PLOP PH HAGAS C PD R PASC G WAQ Y S P S P N P Q V L M Y GSK Y PL PP HHRGSS C P S P PAC C NWD T YN L HSP N P HV L N Y GT K Y P L HP HHRGSS C PD R PAP C D WN N F NS A P N P Q V L V V GT K Y P L HP HHRGSS C PD R PAP C D WN N F NS A P N P HV L Q Y GT K Y P T HP HHRGAS C PD K PAP C G WN P HA T T P S P Q V L V Y GT K Y P T HP HHRGAS C PD K PAP C G WN P HA T T P S P Q V L V Y GT K Y P T HP HHRGAS C PD K PAP C G F L HA T T P S P Q V L Y Y G K Y P L Q P HHRA AS C P D R PAP C S G A Q L N P G P S P Q I L Y Y G K Y P L Q P HHRA S C P D R PAP C S G A Q L N P G P S P VI L I G F G K Y P L S P HHRA S C P D R PAP C S F A Q L N A P G P S P HV L I F G D K Y P L S P HHRA S C C P N F P P C G F I M S S K E P N V H I L Y G F G K Y P L R P HHRS S C P N L P P C G F I M S S K E P N V H I L Y G F G D K Y P L R P HHRS S C P N L P P C G F I M S K E P N V H I L Y G F G D K Y P L R P HHRS S C P N L P P C G F I M S K E P N V H I L Y G F G D K Y P L R P HHRS S C P N L P A C C S Q I M P M K E P S P HV L I G F G D Y P L R P H HR S S C P N P A H C G S D I L Q A D M P N I H I L Y G F G D Y P L R P Y H R S M C P N H P A I C T P E H L A A K L P N V H I L Y G F G D Y P L R P Y H R S S C P F L P A HC G P P M L N I T E P N V H I L Y G G D K Y P L R P H HR S S C P F L P A HC G P P M L N I T E P N V H I L Y G G D K Y P L R P H H R S S C P F L P A HC G P P M L N I T E P N V H I L Y G G D K Y P L R P H H R S S C P F L P A HC G P P M L N I T E P N V H I L Y G G D K Y P L R P H H R S S C P F L P A HC G P P M L N I T E P N V H I L Y G G D K Y P L R P H H R S S C P F L P A HC G P P M L N I T E P N V H I L Y G G D K Y P L R P H H R S S C P F L P A HC G S A D A G H G N P N V T L Y Y G S K F P R O P H H R A S C P D R P P C S S D A O S H G N P N O V L Y Y S N K P R O P H H A A S C P D R P P C S M D Q N A P P O V L Y Y G N P P A Q P H H A A S C P D R P P C S M D Q N A P P O V L Y Y G N N P P A Q P H H A A S C P D K P M C S S P G S D G N A P N P O V L Y Y G N N P P A Q P H H A A S C P V R F P C G S S D G S F D R O P V L Y Y A C N P A Q P H Y A A S C P Y C Y			P D ON NY VD LR           P D END I Y D D KR           P D END I Y D KR           P L VD D S Y E D KR           P L VD D S Y E D KR           P E A D D S Y A D NR           P E A D D S Y D NR           P E A D D S Y D NR           P E A N D NY D VR           P G QS D D Y D D R           P G S N D Y T D D R           P G A N D E Y U D KF           P G A N D E Y U D KF           P G A N D E Y U D KF           P G A N D E Y U D KF           P G A N D E Y U D KF           P G Y N D D Y N D F           P G P D D S Y S D VR           P G P D D S Y S D VR           P G D D S Y S D VR           P G D D S Y S D VR           P G D D S Y S D VR           P G D S S O Y D N R           P G D S S O Y D N R           P G D S S O Y D N R           P G D S O Y D N R           P G D S S O Y D N R           P G D S S O Y D N R           P G D S S O Y D N R           P G D S S O Y N R           P G D S S O Y N R           P D E 1 O C F Y D K R           P D I N S Y S N N R           P D I N S Y S N N R	372 373 372 372 372 372 372 372 372 372
AP         013066281.1           AP         013066281.1           AP         013066281.1           AP         013066281.1           AP         013066281.1           AP         013066281.1           AP         013066371.1           AP         013066371.1           AP         01306697.1           AP         01306697.1           AP         01306672.1           AP         01306472.1           AP         01306422.1           AP         01306422.1           AP         01306422.1           AP         01306422.1           AP         01306422.1           AP         01306424.1           AP         01306424.1           AP         01306424.1           AP         0140424612.1           AP	TAGMYAFLALLAGEMGIDA - YRXWAYEG NYFL EAANNAFIALVAAESG IESK - YRXWAYEG NYFL EAANNAFIALVAAESG IESK - YRXWAYEG NYIL EAANTAFLALAAESG VDSR - LRXWAYEG NYIL EAANTAFLALAAAESG VDSR - LRXWAYEG NYIL YAGNSAFLALLAAESG IPR - LRXWAYEG NYIL YSSNMAMAALVAADAG NO - YRXWAYEG NYIL YSSNMAMAALVAADAG NO - YRWANCG NYA YSSNMAMAALVAADAGLNAO - YRWANCG NYAL YSSNMAMAALVAADAGLNAO - YRWANCG NYAL YSSNMAMAALVAADAGLNAO - YRWANCG NYAL YSSNMAMAALVAADAGLNS - YRWANCG NYAL YSSNMAMAALVAADAGLNS - YRWANCG NYA SSNMAMAALVAADAGLNS - YRWANCG NYAL SSNMAMAALVAADAGLNAO - YRWANCG NYAL SSNMAMAALVAADAGLNAO - YRWANCG NYAL SSNMAMAALVAADAGLNAG - YRWANCG NYAL ASSNMAMAALVAADAGLNA - YRWANCG NYAL ASSNMAMAALVAADAGLNAG - YRWANCG NYAL ASSNMAMAALVAADAGLN E - YRWANCG NYAL ASNMAFAALLAADAGLN E - YRWANCG NYAL AANNAMLALI AADAGLH TE - YRWANCG NYAL HAANNAMLALI AADAGLH TE - YRWANCG NYAL AAANAMLALI AADAGLH TE - YRWANCG NYAL AAANAMLALI AADAGLH TE - YRWANCG NYAL AAANALALVALVAADAG N ST - YRWANCG NYA LAANSAFUC O AADG DOEYRSFAKGG NYM HAANNAFVCL QAADG OEYRSFAKGG NYM	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N T H G S Y Q I G N H H G S Y Q I G N H H G S Y Q I G N H H G S Y Q I G D T G F S Y V V G D T G F S Y V V G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G D T G F S Y V V I G N K H G S F Q I G N K G R S F V V I G D S G R S Y V V I G D S G R S Y V V I	Y GSK Y PDQ PHHAGAS C PD R PASC G WAQ Y S P S P N PQ V H Y GSK Y PL PP HHRGS S C P S P PAC C M WN F N S A P N PQ V L V Y GT K Y P L H P HHRGS S C PD R PAP C D WN F N S A P N PQ V L V V G N K Y P L Q P HH R A S C P D K PAP C D WN F N S T A P N PO V L V Y GT K Y P T H P HHRG AS C P D K PAP C G WD N F HA T T P S PQ V L V Y G N K Y P L Q P HH R A S C P D K PAP C G WD N F HA T T P S PQ V L V Y G N K Y P L Q P HH R A S C P D K PAP C G M D F HA T T P S PQ V L V Y G N K Y P L Q P HH R A S C P D R PAP C S G A Q L N D P O P S P O I L V Y G N K Y P L Q P HH R A S C P D R PAP C S G A Q L N D P O P S P O I L V Y G D K Y P L Q P HH R A S C P D R PAP C S G A Q L N D P O P S P V I I Y G D K Y P L Q P HH R A S C P N L P P C G F I M S S K E P N V H I L Y G G D K Y P L R P HH R S S C P N L P P C G T Q V M P L K E P N V H I L Y G G D K Y P L R P HH R S S C P N L P S P C G T Q V M P L K E P N V H I L Y G G D K Y P L R P HH R S S C P N L P S P C G T Q V M P L K E P N V H I L Y G G D Y P L R P Y H R S S C P N L P S P C G T Q V M P L K E P N V H I L Y G G D Y P L R P Y H R S S C P N L P S P C G T Q V M P L K E P N V H I L Y G G D Y P L R P Y H R S S C P N L P A C G S P U L A E M P N I H I L Y G G D Y P L R P Y H R S S C P N L P A H C G P D I L Q A D M N I H I L Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N V H I L Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N V H I L Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N V H I L Y G G D Y P L R P Y H R S S C P R P P C G S A D A G S H O P N Q V L Y Y G S K F P R P H H R A S C P D R P A P C S S A D G S H O P N Q V L Y Y G S K Y P L Q P H H R A S C P D R P A P C S S A D G S H O P N Q V L Y Y G S N P P A Q P H H A A S S C P D R P P C S M D Q L N A P G S P Q V L Y G G N N P M A P H A A S S C P D R P P C S M D Q L N A P G S P N Q V L Y G S N N P M A Q P H H A A S S C P D R P P C S M D Q L N A P G S P N V L Y G S N N P M A Q P H H A A S S C P D R P P C S M D Q L M A P G S N P N Q L Y G S N N P M A Q P H H A A S S C P D K P A C G S M D D F E	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6		P G ND N Y UD LR           P D END I Y UD KR           P D END I Y ED KR           P L VD D S Y ED KR           P L VD D S Y ED KR           P E AD D S Y AD NR           P E AD D S Y AD NR           P E AD D S Y AD NR           P E AN D N Y D VR           P E AN D N Y D VR           P G Q S DD Y T D DR           P G Q S ND Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G Q ND D Y T D DR           P G D D S Y S D VR           P D D D S Q Y AD LY           P D D D S U Y D L P           P D D D S U Y D L P           P D D S U Y D L P           P D L D C Y D KR           P D E AD U C Y D C F           P D E AD U C Y D C F           P D E AD U C Y D C F           P D E AD U C Y D C F           P D E AD U C Y D C F           P D E AD U C Y D C F           P D E AD U C Y D C F           P D E AD C Y C Y C F </td <td>372 373 372 372 372 372 372 372 372 372</td>	372 373 372 372 372 372 372 372 372 372
AP_013066281,1           AP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066371,1           XP_013066371,1           XP_013066170,1           XP_01306652,1           XP_01306652,1           XP_01306628,1           XP_01306622,1           XP_01306428,1           XP_0140448,1           XP_01460448,1           XP_01460454,1           XP_015126067,1           XP_015126067,1           XP_01526067,1           XP_01512607,	TAGMYAFLALLAGEMG I DA - Y RKWA VEQ I NY L EAANNAF I AL VAAESG I ESK - Y RKWA VEQ I NY I L EAANNAF I AL VAAESG I ESK - Y RKWA VEQ I NY I L EAANNAF I AL LAAESG I DT R - L RKWA VEQ I NY I L EAANNAFLALAAAESG I DT D - Y RKWA VEQ I NY I L YAGNSAFLALLAAAESG I NT D - Y RKWA VEQ I NY I L YAGNSAFLALLAAEG I NT D - Y RKWA VEQ I NY I L YSSNIAMAAL VAADAGLN AO - Y RHWA MCQ I HY AL YSSNIAMAAL VAADAGLN AO - Y RHWA MCQ I HY AL YSSNIAMAAL VAADAGLN AO - Y RHWA MCQ I HY AL YSSNIAMAAL VAADAGLN AO - Y RHWA MCQ I HY AL YSSNIAMAAAL VAADAGLN AO - Y RHWA MCQ I HY AL YSSNIAMAAL VAADAGLN AO - Y RHWA MCQ I HY AL ANNAMAAL VAADAGLN AO - Y RHWA MCQ I HY AL ASSNIAMAAL VAADAGLN AO - Y RHWA MCQ I HY AL ASSNIAMAAL VAADAGLN AC - Y RHWA MCQ I HY AL ASSNIAMAAL VAADAGLN AC - Y RHWA MCQ I HY AL ASSNIAMAAL VAADAGL HSE - Y RHWA MCQ I HY AL AANNAMLALI AADAGLH TE - Y RHWA MCQ I HY AL HAANNAMLALI AADAGLH TE - Y RHWA MCQ I HY AL HAANNAMLALI AADAGLH TE - Y RHWA MCQ I HY AL ASANAMAMLALI AADAGL HTE - Y RHWA MCQ I HY AL HAANNAMLALI AADAG I HTE - Y RHWA MCQ I HY AL HAANNAMLALI AADAG I HTE - Y RHWA MCQ I HY AL HAANNAMLALI VAADAG I HTE - Y RHWA MCQ I HY AL HAANNAMLALI VAADAG I HTE - Y RHWA MCQ I HY AL HAANNA AL VAADAG I SAT - Y RHWA MCQ I HY AL HAANNAFI ALLAADAG I SAT - Y RHWA MCQ I HY AL HAANNAFI ALLAADAG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAADAG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABAG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABAG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABAG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABAG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABAG I SAT - Y RKWA MCQ I HY AL HAANNAFI ALLAABAG I SAT - Y RKWA MCQ I HY I HAANNAFI ALLAABAG I SAT - Y REFARQQ I HY IL HAANNAFI VCLQAADG OG E Y REFARQQ I HY IL HAANNAFI VCLQAADG OG DE Y REFARQQ I HY IL HAANNAFI VCLQAADG OG DE Y REFARQQ I HY IL HAANNAFI VCLQAADG OG VE - P ROFATQOI SY ML	G N HHG S Y V V G N PHG S Y Q V G N THG S Y Q V G N THG S Y Q V G N THG S Y Q I G N HHG S Y Q I G N HHG S Y Q I G D TG F S Y V V G D G G S S Y V V G D G G S S Y V V	Y GSK Y PLOP PH HAGAS C PD R PASC GWAQYS S PS PN PQ V L M Y GSK Y PL PP HHRGSS C PS PP AC C NWD T YN L HSPN PH V L N Y GT K Y PL HP PH HRGSS C PD R PAP C D WNN F NS A PN PQ V L V V GT K Y PL HP PH HRGAS C PD K PAP C G WD F HA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD K PAP C G WD N F HA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD K PAP C G WD N F HA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD K PAP C G F L NL HA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD K PAP C G F L NL HA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD K PAP C G F L NL HA T T PS PQ V L Y Y GT K Y PT HP HHRGAS C PD R PAP C G F L NL HA T T PS PV V I Y GD K Y PL Q PH HRA S C PD R PAP C G F L NS K E PN V HI L Y F GD K Y P HS PH HRASS C PN L PAP C G F I M S K E PN V HI L Y F GD K Y PL R PH HRS S C PN L PAP C G F I M S K E PN V HI L Y F GD K Y PL R PH HR S S C PN L PAP C G F I M S K E PN V HI L Y F GD K Y PL R PH HR S S C PN L PAP C G F I M S K E PN V HI L Y F GD K Y PL R PH HR S S C PN L PAP C G F I M S K E PN V HI L Y F GD K Y PL R PH HR S S C PN L PAP C G T O Y M PL K E PN V HI L Y F GD K Y PL R PY G R I S M C PN HP A H C G P L L A E M PN I H I L Y F GD Y PL R PY G R I S M C PN HP A H C G P M L N L T E PN V H I L Y F GD K Y PL R PY H R S S C P F L PAH C G P M L N L T E PN V H I L Y Y G S K F P R O P H H R A S C P D R PAP C G S M D Q S H G PN O V L Y W G R D P O Q P H HA A S C P D R PAP C G S M D Q S H G PN O V L Y W G R D P O Q P H HA A S S C P D R P P C G S M D Q L A P O P S O V L Y W G R D P O Q P H HA A S S C P D R P P C G S M D Q L A P O P P O V L Y G A N N P A Q P H HA A S S C P D R P P C G S M D G R S G N N P N Q U L Y G A N N P A Q P H HA A S S C P D R P P C G S M D G R S G N N P N Q U L Y G Y N N P A Q P H HA A S S C P D R P P C G S W D F C B N A S N P N I L Y G Y N N P A Q P H HA A S S C P D R P P C G S W D F C B N A S N P I I L Y G Y N N P A Q P H HA A S S C P D R P P C G S W D F C B N A S N P I I L Y G Y N N P A Q P H HA A S S C P D R P P C G S W D F C B N A S N P I I L Y G Y N N P A Q P H HA A S S C P D R P P C G S W D F E D R N A S N P I	G G G G G G G G G G G G G G G G G G G		P G QN DN Y VD LR           P D E N D I Y G D K R           P D E N D I Y G D K R           P D E N D I Y G D K R           P D E N D S Y G D K R           P E V D S Y D K R           P E A D S Y A D N R           P E A D S Y A D N R           P E A D S Y D V R           P E A D S Y D V R           P E A D S Y D V R           P G QS D D Y T D D R           P G S N D Y T D D R           P G S N D Y T D D R           P G A N D Y V D V R           P G A N D Y V D P R           P G A N D Y V D P R           P G A N D Y V D P R           P G O N D Y V D P R           P G O N D Y V D P R           P G O N D Y V D P R           P G P D D S Y S D V R           P G P D D S Y S D V R           P D S D Q Y A D L R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F Q D R           P D E A D K F R Q H R           P D E A D K F R Q H R           P D E A D K F R Q H R	372 373 372 372 372 372 372 372 374 394 394 394 394 394 394 394 393 393 370 370 393 393 393 393 393 394 394 394 394
μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066521,1           μP_013066523,1           μP_013066523,1           μP_013066523,1           μP_013066523,1           μP_013064281,1           μP_01404641,1           μP_015126067,1           μP_0151264471,1           μP_015184471,1           μP_015184471,1           μP_015184471,1           μP_015184471,1           μP_015184471,1           μP_015184471,1           μP_015425381,4 <td>TA GINYA FLA LLA GE BIG I DA - Y RKWA VE G I NY IL EA ANNA FI AL VA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FI AL VA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FI AL LA AE SG V DSR LRKWA VE G I NY IL Y AG NSA FI AL LA AE SG V DSR LRKWA VE G I NF IL Y AG NSA FI AL LA AE SG I PR - L RKWA VE G I NF IL Y SS NIAMAAL VA AD AG LI NG - Y RKWA VE G I NY IL Y SS NIAMAAL VA AD AG LI NG - Y RKWA VE G I NY IL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI HE - Y RHWA MCG I HY AL A ANNA AL LA AD AG LI HE - Y RHWA MCG I HY AL A ANNA AL LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA AL LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA SI VAL VAA DNG I HTE - Y RHWA MCG I HY AL HAA NIA SI VAL VAA DNG I HTE - Y RHWA MCG I HY AL HAA NIA SI VAL VAA DNG I S AS - Y RE Y RKWA VE O I NY IL HAA NIA SI VAL QAA DG DOE Y RE FA AC O I Y M HAA NIA SI VCL QAA DG DOE Y RE FA AC O I Y M HAA NIA FI CL EA AG C SI NS - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - G DR E - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - G DR E - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - D G DR E - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - D G RE - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - D G DR E - Y RE FA AC O I HY ML</td> <td>G N HHG S Y V V G N PHG S Y Q V G N THG S Y Q V G N THG S Y Q V G N THG S Y Q I G N HHG S Y Q I G N HHG S Y Q I G N HHG S Y Q I G D TG F S Y V I G D TG F S Y V</td> <td>Y GSK Y PLOP PH HAGAS C PD R PASC GWAQY S P S P N P Q V L Y Y GSK Y PL PP H HR GSS C P D R PAP C D WN F NS A PN P Q V L Y Y GT K Y P L H P H HR GSS C PD R PAP C D WN F NS A PN P Q V L Y V GN K Y P L Q P H H R AS C P D K PAP C D WN F NS T A P N P Q V L Y Y GT K Y P T H P H HR GAS C P D K PAP C G WD F H A T T P S P Q V L Y Y GT K Y P T H P H HR GAS C P D K PAP C G WD F H A T T P S P Q V L Y Y GT K Y P T H P H HR GAS C P D K PAP C G F L H A T T P S P Q V L Y Y G K Y P L Q P H H R AS C P D K PAP C G F L H A T T P S P Q V L Y Y G D K Y P L Q P H H R AS C P D R PAP C G F L M S K P P V H I L Y G G D K Y P H S P H HR S S C P N L PAP C G F I M S S K E P N Y H I L Y F G D K Y P H S P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D K Y P L R P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D K Y P L R P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D Y P L R P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D Y P L R P H HR S S C P N P P A K C S Q I M P M K E P S H V L I G G D Y P L R P Y H R S S C P N P P A K C S Q I M M K E P S H V L I G G D Y P L R P H H R S S C P N P P A K C G P U L A E M N I H I L Y G G D Y P L R P Y H R S S C P N P P A K C G P M L N L T E P N Y H I Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N Y H I Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N Y H I Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N Y H I Y G G D Y P L R P H H R S S C P N R P A P C G S A D A G H G P N P Q V L I Y G S K Y P L Q P H H R A S C P D R P A P C G S A D A G H G P N V T L Y Y G S K Y P L Q P H H R A S C P D R P A P C G S S D D A R P A P Q V L Y G N N P A Q P H H A A S C P D R P A P C G S S D D A R P A P Q V L Y G N N P A Q P H H A A S C P D K P M P C G W S E F D R A P N Q I L Y Y G N N P A Q P H H A A S C P D L P M C G W S E F D R A S N P O I L Y Y G N N P A Q P H H A A S S C P D L P M C G W S E F D R A S N P O I L Y Y G N N P A Q P H H A A S S C P D L P M C G W S E F D R A S N P O I L Y Y G N N P A Q P H H A A S S C P D L P M C G W S E</td> <td>G G G G G G G G G G G G G G G G G G G</td> <td></td> <td>F D Q ND NY VD LR           SP D E ND I Y Q D KR           SP D E ND I Y E D KR           SP D E ND I Y E D KR           SP L VD S Y E D KR           SP E V D S Y E D KR           SP E V D S Y E D KR           SP E V D S Y E D KR           SP E V D S Y D D KR           SP E A D M Y D V R           SP E A D M Y D V R           SP G SN D D Y D D R           SP G SN D D Y D D R           SP G SN D D Y D D R           SP G A N D E Y U D K F           SP G A N D E Y U D K F           SP G O N D E Y D D R           SP G O N D E Y D D K F           SP G O N D E Y D D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O D E Y D K F O D R           SP D E A D K F O D R           SP D E A D K F O D R           SP D E A D K F O D R           SP D E A D K F O D R           SP D E A D K F R D H R           SP D E A D K F R D H R           SP D E A D K F R D H R           SP D E A D K F R D H R           SP D E A D K F R D H R<!--</td--><td>372 373 372 372 372 372 372 372 374 394 394 394 394 394 394 393 393 393 39</td></td>	TA GINYA FLA LLA GE BIG I DA - Y RKWA VE G I NY IL EA ANNA FI AL VA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FI AL VA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FI AL LA AE SG V DSR LRKWA VE G I NY IL Y AG NSA FI AL LA AE SG V DSR LRKWA VE G I NF IL Y AG NSA FI AL LA AE SG I PR - L RKWA VE G I NF IL Y SS NIAMAAL VA AD AG LI NG - Y RKWA VE G I NY IL Y SS NIAMAAL VA AD AG LI NG - Y RKWA VE G I NY IL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA AD AG LI HE - Y RHWA MCG I HY AL A ANNA AL LA AD AG LI HE - Y RHWA MCG I HY AL A ANNA AL LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA AL LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI I AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI HE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA LI AA DAG LI TE - Y RHWA MCG I HY AL HAA NIA SI VAL VAA DNG I HTE - Y RHWA MCG I HY AL HAA NIA SI VAL VAA DNG I HTE - Y RHWA MCG I HY AL HAA NIA SI VAL VAA DNG I S AS - Y RE Y RKWA VE O I NY IL HAA NIA SI VAL QAA DG DOE Y RE FA AC O I Y M HAA NIA SI VCL QAA DG DOE Y RE FA AC O I Y M HAA NIA FI CL EA AG C SI NS - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - G DR E - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - G DR E - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - D G DR E - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - D G RE - Y RE FA AC O I HY ML HAA NIA FI CL EA AD - D G DR E - Y RE FA AC O I HY ML	G N HHG S Y V V G N PHG S Y Q V G N THG S Y Q V G N THG S Y Q V G N THG S Y Q I G N HHG S Y Q I G N HHG S Y Q I G N HHG S Y Q I G D TG F S Y V	Y GSK Y PLOP PH HAGAS C PD R PASC GWAQY S P S P N P Q V L Y Y GSK Y PL PP H HR GSS C P D R PAP C D WN F NS A PN P Q V L Y Y GT K Y P L H P H HR GSS C PD R PAP C D WN F NS A PN P Q V L Y V GN K Y P L Q P H H R AS C P D K PAP C D WN F NS T A P N P Q V L Y Y GT K Y P T H P H HR GAS C P D K PAP C G WD F H A T T P S P Q V L Y Y GT K Y P T H P H HR GAS C P D K PAP C G WD F H A T T P S P Q V L Y Y GT K Y P T H P H HR GAS C P D K PAP C G F L H A T T P S P Q V L Y Y G K Y P L Q P H H R AS C P D K PAP C G F L H A T T P S P Q V L Y Y G D K Y P L Q P H H R AS C P D R PAP C G F L M S K P P V H I L Y G G D K Y P H S P H HR S S C P N L PAP C G F I M S S K E P N Y H I L Y F G D K Y P H S P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D K Y P L R P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D K Y P L R P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D Y P L R P H HR S S C P N L PAP C G F I M S K E P N Y H I L Y G G D Y P L R P H HR S S C P N P P A K C S Q I M P M K E P S H V L I G G D Y P L R P Y H R S S C P N P P A K C S Q I M M K E P S H V L I G G D Y P L R P H H R S S C P N P P A K C G P U L A E M N I H I L Y G G D Y P L R P Y H R S S C P N P P A K C G P M L N L T E P N Y H I Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N Y H I Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N Y H I Y G G D Y P L R P Y H R S S C P F L P A H C G P P M L N L T E P N Y H I Y G G D Y P L R P H H R S S C P N R P A P C G S A D A G H G P N P Q V L I Y G S K Y P L Q P H H R A S C P D R P A P C G S A D A G H G P N V T L Y Y G S K Y P L Q P H H R A S C P D R P A P C G S S D D A R P A P Q V L Y G N N P A Q P H H A A S C P D R P A P C G S S D D A R P A P Q V L Y G N N P A Q P H H A A S C P D K P M P C G W S E F D R A P N Q I L Y Y G N N P A Q P H H A A S C P D L P M C G W S E F D R A S N P O I L Y Y G N N P A Q P H H A A S S C P D L P M C G W S E F D R A S N P O I L Y Y G N N P A Q P H H A A S S C P D L P M C G W S E F D R A S N P O I L Y Y G N N P A Q P H H A A S S C P D L P M C G W S E	G G G G G G G G G G G G G G G G G G G		F D Q ND NY VD LR           SP D E ND I Y Q D KR           SP D E ND I Y E D KR           SP D E ND I Y E D KR           SP L VD S Y E D KR           SP E V D S Y E D KR           SP E V D S Y E D KR           SP E V D S Y E D KR           SP E V D S Y D D KR           SP E A D M Y D V R           SP E A D M Y D V R           SP G SN D D Y D D R           SP G SN D D Y D D R           SP G SN D D Y D D R           SP G A N D E Y U D K F           SP G A N D E Y U D K F           SP G O N D E Y D D R           SP G O N D E Y D D K F           SP G O N D E Y D D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O N D E Y D K F           SP G O D E Y D K F O D R           SP D E A D K F O D R           SP D E A D K F O D R           SP D E A D K F O D R           SP D E A D K F O D R           SP D E A D K F R D H R           SP D E A D K F R D H R           SP D E A D K F R D H R           SP D E A D K F R D H R           SP D E A D K F R D H R </td <td>372 373 372 372 372 372 372 372 374 394 394 394 394 394 394 393 393 393 39</td>	372 373 372 372 372 372 372 372 374 394 394 394 394 394 394 393 393 393 39
AP_013066281.1           AP_013066281.1           AP_013066281.1           AP_01306628.1           AP_01306628.1           AP_01306628.1           AP_01306637.1           AP_01306837.1           AP_01306817.1           AP_01306817.1           AP_01306817.1           AP_01306817.1           AP_01306817.1           AP_01306828.1           AP_01306828.1           AP_01308428.1           AP_01478661.1           AP_01478661.1           AP_01508425.1           AP_01508425.1           AP_015184471.1           AP_01551847.1           AP_01551382.1           AP_01551382.1	TA GINYA FLA LLAGE NG I DA - Y RKWA VE G I NY IL EA ANNA FI AL VA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FI AL LA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FI AL LA AE SG I O TR - L RKWA VE G I NY IL EA ANNA FI AL LA AAE SG VD SR - L RKWA VE G I NF IL Y AG NS AF LA LLA AE SG I P DR - L RKWA VE G I NF IL Y SS NIAMAAL VA ADAG I ND - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG I ND - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG LN AD - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN AD - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL SA ANMAAL VA ADAG LH BE - Y RHWA MCG I HY AL A SN NAFA AL LA AD TG LH PE - Y RHWA MCG I HY AL A SN ANAFAL LA ADAG LH TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG LH TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG LH TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG LH TE - Y RHWA MCG I HY AL AA ANNA LA LA ADAG LH TE - Y RHWA MCG I HY AL AA ANNAFI LA LA ADAG LH TE - Y RHWA MCG I HY AL AA ANNAFI LA LA ADAG LH TE - Y RHWA MCG I HY AL AANNAFI LA LA ADAG LH TE - Y RHWA MCG I HY AL AANNAFI LA LA ADAG LH TE - Y RHWA MCG I HY AL AANNAFI LA LA ADAG LH TE - Y RHWA MCG I HY AL AANNAFI CLQ AADG I S AS - Y RKWA VE O I NY ML HAANNAFI CLQ AAD G D G - D GE Y RG FA TG I SY ML Y AG NTAL Y AL LT AEAG I S AT - Y RKWA VE O I NY ML HAANNAFI CLQ AAD G - D GE - Y NG FA TD G I SY ML Y AG NIAF I ALL AAE SG LN ST VY RY WA LQ I HY ML HAANNAFI CLC AAD G - D O DE - Y RG FA TD G I SY ML Y AG NIAF I ALL AAE SG LN ST VY RY FF FA RQ I HY ML HAANNAFI CLC AAD G - D O DE - Y RG FA TD G I SY ML Y AG NIAF I ALL AAE SG LN ST VY RY FF FA RQ I HY ML HAANNAFI CLC AAD G - D O DE - Y RG FA TD G I SY ML Y AG NIAF I ALL AAE SG LN ST VY RY FF FA RQ I HY ML HAANNAFI CLC EAAD G - D O DE - Y RG FA TD G I SY ML Y AG NIAF I LL G FA AG G - N O DE Y RF FA RQ I HY ML	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N T H G S Y Q I G N H H G S Y Q I G N H H G S Y Q I G N H H G S Y Q I G D T G F S Y V V G D S G G R S Y V V G D G G G R S Y V V G G G G R S Y V V G G G G R S Y V V	Y GSK Y PLOP PH HAGAS C PD RPAC C UMAQY S P S P N P O V L Y Y GSK Y PL PP H HR GSS C PD RPAP C D WN F NS A AP N P Q V L Y Y GT K Y P L H P H HR GSS C PD RPAP C D WN F NS A AP N P Q V L Y V GT K Y P L H P H HR GSS C PD RPAP C D WN F NS T AP N P O V L Y Y GT K Y P T H P H HR GAS C PD K PAP C G WD N F HA T T P S P Q V L Y Y GT K Y P T H P H HR GAS C PD K PAP C G WD N F HA T T P S P Q V L Y Y GT K Y P T H P H HR GAS C PD K PAP C G WD N F HA T T P S P Q V L Y Y GT K Y P T U P H HR GAS C PD K PAP C G WD N F HA T T P S P Q V L Y Y G K Y P L Q P H HR A S C P D R P A C G F A L HA T T P S P Q V L Y Y G K Y P L Q P H HR A S C P D R P A C G F A L HA P P P S P H V L I F G D K Y P L S P H HR S S C P N L P A P C G P F I M S S K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P A P C G P F I M S S K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P S P C G T Q V M P L K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P S P C G T Q V M P L K E P N V H I L Y F G D Y P L R P H HR S S C P N L P S P C G T Q V M P L K E P N V H I L Y F G D Y P L R P Y H R S S C P N L P A L C S P V L L A E M P N I H I L Y F G D Y P L R P Y H R S S C P N L P A L C S P V L L A E M P N I H I L Y F G D Y P L R P Y H R S S C P N L P A L C G P P M L N L T E P N V H I L Y F G D Y P L R P Y H R S S C P N L P A L C G P P M L N L T E P N V H I L Y F G D Y P L R P Y H R S S C P L P A H C G P P M L N L T E P N V H I L Y F G D Y P L R P H H R S S C P D R P A C G S D A A S H G N P O V L Y Y G S K F P R Q P H HA A S C P D R P P C G S M D A S H G N P O V L Y Y G N N P T Q P H HA A S S C P D K P P C G W S E F D R A S N P O V L Y G G N N P T Q P H HA A S S C P D K P M C G W S E F D R A S N P O U L Y Y G N N P T Q P H HA A S S C P D K P M C G W S E F D R A S N P O I L Y Y G N N P T Q P H HA A S S C P D K P M C G W S E F D R A S N P O I L Y W G N N P T Q P H HA A S S C P D K P M C G W S E F D R A S N P O I L Y W G N N P T Q P H H A A S S C P D K P M C G W S E F D R A S N P O I L Y W G N N P T Q P H H A A S C P D R P A C G W S E F D R A S N P O I L Y W G N N P T	G G G G G G G G G G G G G G G G G G G		P D Q ND N Y U D LR           P D E ND I Y D D KR           P D E ND I Y D D KR           P L V D D S Y D KR           P E V D D S Y D KR           P E V D D S Y D KR           P E A D D Y N D KR           P E A D D Y N D KR           P E A D D Y D D R           P G Q S D D Y D D R           P G Q S D D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G Q N D Y T D D R           P G D S Y D V R F           P G D D S Y S D V R           P D D S Y S D V R           P D D S Y S D V R           P D E A S D I Y E D L R           P D E A S D I Y E D L R           P D E A S D I Y E D L R           P D E A S D I Y E D L R           P D E A D R C D R           P D E A D R C D R           P D E A D R C D R           P D E A D R C R D R           P D E A D R C R D R           P D E A D R C R D R           P D E A D R C R D R           P D E A	372 373 372 372 372 374 394 394 394 394 394 394 394 394 394 39
AP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066371,1           XP_0130661281,1           XP_0130661281,1           XP_0130661281,1           XP_0130661281,1           XP_013066128,1           XP_013066128,1           XP_01306628,1           XP_01306628,1           XP_01306628,1           XP_01306428,1           XP_01404428,1           XP_01404642,1           XP_01512664,1           XP_01512664,1           XP_01512664,1           XP_01542538,21           XP_01554267,1           XP_01554267,1           XP_01554267,1           XP_0	TAGMYAFLALLAGEMGIDA - YRXWAYEG NYFL EAANNAFIALVAAESGIESK - YRXWAYEG NYFL EAANNAFIALVAAESGIESK - YRXWAYEG NYIL EAANYAFLALAAAESGIOTR - LRXWAYEG NYIL EAANYAFLALAAAESGIOTR - LRXWAYEG NYIL YAGNSAFLALLAAESGINTD - YRXWAYEG NYIL YAGNSAFLALLAAESGINTD - YRXWAYEG NYIL YSSNMAMAALVAADAGLNAC - YRWAMCG NYAL YSSNMAMAALVAADAGLNAC - YRWAMCG NYAL ASSNMAMAALVAADAGLNAC - YRWAMCG NYAL YSSNMAMAALVAADAGLNAC - YRWAMCG NYAL ASSNMAMAALVAADAGLNAC - YRWAMCG NYAL YASNMAFALLAADTGLHFE - YRHWAMCG NYAL AANNAMLALI AADAGLN E - YRHWAMCG NYAL HAANNAMLALI AADAGLN TE - YRHWAMCG NYAL HAANNAYFYCLQAADG DOEYRSFAKQO NYML HAANYAFYCLQAADG OUEYRFARQQ NYML HAANYAFYCLQAADG DOEYRSFAKQO NYML HAANYAFYCLQAADG DOEYRSFARQQ NYML HAANYAFYCLQAADG DOEYRFARQQ NYML HAANYAFYCLQAADG DOEYRFARQQ NYML HAANYAFYCLQAADG DOEYRFARQQ NYML HAANYAFYCLQAADG NEE - YRFARQQ NYML HAANYAFYCLQAADG DOEYRFARQQ NYML HAANYAFYCLGAADG DOEYRFARQQ NYML HAANYAFYCLGAADG DOEYRFARQQ NYML HAANYAFYCLGAADG DOEYRFARQQ NYML HAANYAFYCLGAADG DOEYRFARQQ NYML HAANYAFYCLGAADG DOEYRFARQQ NYML	G N HHG S Y V V G N PHG S Y Q V G N PHG S Y Q V G N THG S FQ I G N HHG S YQ I G N HHG S FQ I G N HHG S FQ I G D TG FS Y L I G D TG FS Y V I G G G G S Y V V G G G G S Y V V C G G G S Y V V C G G G S Y V V	Y GSK Y PLOP PH HAGAS C PD R PAC C UMAQ Y S P S P N P Q V L M Y GSK Y PLOP PH HR S S C P S P PAC C WN N F N S A P N P Q V L Y Y GT K Y P L H P H HR GSS C P D R PAP C D WN N F N S A P N P Q V L Y V GT K Y P L H P H HR GAS C P D K PAP C G WD N F N S T A P N P Q V L Y Y GT K Y P L H P H HR GAS C P D K PAP C G WD N F N S P L P N P V L Q Y GT K Y P T H P H HR GAS C P D K PAP C G WD N F N S P L P N P V L Q Y GT K Y P T H P H HR GAS C P D K PAP C G F E N L H T T P S P Q V L Y Y G T K Y P T H P H HR GAS C P D K PAP C G F E N L H T T P S P Q V L Y Y G T K Y P T H P H HR GAS C P D K PAP C G F E N L H A T T P S P V L I Y G D K Y P L Q P H HR A S C P D R PAP C G F E N L M T P S P N V I L Y G T G K Y P H S P H HR AS C P D R PAP C G F I M S K E P N V H I L Y F G D K Y P L P P H HR S S C P N L PAP C G F I M S K E P N V H I L Y F G D K Y P L R P H HR S S C P N L PAP C G F I M S K E P N V H I L Y F G D K Y P L R P H HR S S C P N L PAP C G T O Y M P L K E P N V H I L Y F G D K Y P L R P H HR S S C P N L PAP C G T O Y M P L K E P N V H I L Y F G D Y L R P Y G R I S M C P N H P A H C G P D L L A E M P N I H I L Y F G D Y P L R P Y G R I S M C P N H P A H C G P P M L N L T E P N V H I L Y F G D Y P L R P Y A R I S M C P N H P A H C G P M L N L T E P N V H I L Y Y G S K Y P L R P H H R S S C P F L P A H C G P M L N L T E P N V H I L Y Y G S K Y P L R P H H R S S C P L P A P C G S M D A P O S M O N P A Q L N H H A S C P D R P A C G P M L N L T E P N V H I L Y Y G S K Y P L P H H A S S C P D R P P C S M D D A P O S P O V L Y Y G N N P A Q P H HA A S C P D R P P C G S M D D F A G P N P Q L H Y G N N P A Q P H H A A S S C P D K P M P C S M D D R A P P O Y L Y Y G N N P M Q P H H A A S S C P D R P A C G W D F D R A S M P O I L Y Y G N N P K Q P H H A A S S C P D R P A C G W D F F D R A S M P O I L Y Y G N N P K Q P H H A A S S C P D R P A C G W D F F D R N A P N O I L Y Y G N N P K Q P H H A A S S C P D R P A C G W D F F D R A S M P O I L Y Y G N N P K Q P H H A A S S C P D R P A C G W D F F D R A S M P O I L Y Y G N N P K Q P H H A A S	G G G G G G G G G G G G G G G G G G G		P G ND NY VD LR           P D E ND I Y G D K R           P D E ND I Y G D K R           P L VD D S Y E D K R           P L VD D S Y D K R           P E A D D S Y A D N R           P E A D D S Y A D N R           P E A D D S Y A D N R           P E A D D S Y A D N R           P E A D S Y A D N R           P E A D S Y A D N R           P E A D S Y A D N R           P G S N D D Y T D D R           P G S N D Y T D D R           P G S N D D Y T D D R           P G A N D E Y D V D F           P G A N D E Y D V D F           P G A N D E Y D D R           P G O D S Y S D V R           P G P D D S Y S D V R           P D E A D K F C D H R           P D E A D K F C D H R           P D E A D K F C D H R           P D E A D K F D D H R           P D E A D K F D D H R           P D E A D K F D H R           P D E A D K F D H R           P D E A D K F D H R           P D E A D K F D H R           P D E A D K F D H R           P D E A D K F D H R           P D E A D K F D H R           P D E A D K F D H R           P D E A D K F R D F R           P D E A D K F R D F R           P D E A D K F R D F R </td <td><ul> <li>372</li> <li>373</li> <li>372</li> <li>372</li> <li>372</li> <li>372</li> <li>372</li> <li>374</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>393</li> <li>370</li> <li>370</li> <li>370</li> <li>370</li> <li>393</li> <li>394</li> </ul></td>	<ul> <li>372</li> <li>373</li> <li>372</li> <li>372</li> <li>372</li> <li>372</li> <li>372</li> <li>374</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>393</li> <li>370</li> <li>370</li> <li>370</li> <li>370</li> <li>393</li> <li>394</li> </ul>
μP_013066281,1           μP_013064281,1           μP_013084281,1           μP_015128607,1           μP_015128607,1           μP_015138247,11           μP_015513827,1           μP_01563827,10           μP_01563827,10           μP_01563827,10	TA GIN AF LA LL AGE ING ID AR - Y RK WA VE GINY IL EA ANNAF I AL VA AE SGIE SK Y RK WA VE GINY IL EA ANNAF I AL LA AE SGIE SK Y RK WA VE GINY IL EA ANNAF I AL LA AE SGIT RF LR WA VE GINY IL Y AGNSAF LA LLA AE SGIT RF LR WA VE GINY IL Y SK MAMAAL VA AD AG LING Y RK WA VE GINY IL Y SK MAMAAL VA AD AG LING Y RK WA VE GINY IL Y SK MAMAAL VA AD AG LING Y RK WA VE GINY IL Y SK MAMAAL VA AD AG LING Y RK WA VE GINY IL Y SK MAMAAL VA AD AG LING Y RH WA MC GINY AL Y SK MAMAAL VA AD AG LING Y RH WA MC GINY AL Y SK MAMAAL VA AD AG LING Y RH WA MC GINY AL Y SK MAMAAL VA AD AG LING Y RH WA MC GINY AL Y SK MAMAAL VA AD AG LING Y RH WA MC GINY AL Y SK MAMAAL VA AD AG LING Y RH WA MC GINY AL A SA MAMAL I AA AD AG LING Y RH WA MC GINY AL A SA MAMAL ALIA AD AG LING Y RH WA MC GINY AL HAA MAMAL ALIA AD AG LING Y RH WA MC GINY AL HAA MAMAL ALIA AD AG LING Y RH WA MC GINY AL HAA MA LIA LIAAD GLINTE - Y RH WA MC GINY AL HAA MA LIA LIAAD GLINTE - Y RH WA MC GINY AL HAA MA LIA LIAAD GLINTE - Y RH WA MC GINY AL HAA MA LIA LIAAD GLINTE - Y RH WA MC GINY AL HAA MA LIA LIAAD GLINTE - Y RH WA MC GINY AL HAA MA LIA LIAAD GLINTE - Y RH WA MC GINY AL HAA MA SI VAL VAA ONG INTE - Y RH WA MC GINY AL HAA MA SI LIAAD SI SI SI SI Y RK WA VE GINY ML HAA MAYAF VC LOAAD GIN - D GOE Y RE SFA KOO INY AL HAA MA Y FV CLOAAD GIN - D GOE Y RE SFA KOO INY ML HAA MAYAF Y CLOAAD GIN - D GOE Y RE SFA KOO INY ML HAA MAYAF Y CLEAAD GIN - D GOE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - D GOE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - D GOE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - D GOE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - D GOE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - N WE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - N WE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - N WE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD GIN - N WE Y RE FA RE GO INY ML HAA MAYAF Y CLEAAD CINY ML HAANYAF Y CLEAAD Y SE A Y WE FA RE GO INY ML HAA MAYAF Y CLEAAD CINY ML Y RE FA RE GO INY ML	G N HIGS Y VV G N PIGS YQ V G N FIGS YQ V G N TIGS FQ I G N HIGS YQ I G D TGFS YV VI G G G G S YV VI G G G G S YV VI G G G G S YV VI	Y GSK Y PLOP PH HAGAS C PD R PA C G WA QY S P S P N P Q V L Y Y GSK Y PL PP H HR GSS C P D R PA P C D WN F N S A P N P Q V L Y Y GT K Y P L H P H HR GSS C P D R PA P C D WN F N S A P N P Q V L Y V G N K Y P L Q P H HR AS C P D K PA P C G WD F H A T T P S P Q V L Y Y G N K Y P L Q P H HR AS C P D K PA P C G WD F H A T T P S P Q V L Y Y G N K Y P L Q P H HR AS C P D K PA P C G WD F H A T T P S P Q V L Y Y G N K Y P L Q P H HR AS C P D K PA P C G F L N A T P S P Q V L Y Y G N K Y P L Q P H HR AS C P D K PA P C G F L N A T P S P Q V L Y Y G N K Y P L Q P H HR AS C P D K PA P C G F L N A T P S P Q V L Y Y G D K Y P L Q P H HR AS C P D R P A P C G F L N S K E P N V H I L Y G F G D K Y P S P H HR S S C P N L P A P C G F I M S K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P A P C G F I M S K E P N V H I L Y G F G D K Y P L R P H HR S S C P N L P A P C G F I M S K E P N V H I L Y G F G D K Y P L R P H HR S S C P N L P A K C S Q I M P M K E P S P H V L I G F G D Y P L R P H HR S S C P N L P A K C S Q I M P M K E P S P H V L I G F G D Y P L R P Y HR S S C P N L P A K C G S Q I M P M K E P S P H V L I G F G D Y P L R P Y HR S S C P N P A H C G P D M L N L T E P N V H I L Y G F G D Y P L R P Y HR S S C P N P A H C G P M L N L T E P N V H I L Y G F G D Y P L R P Y H R S S C P N R A P C G S A D A G B H M P M P V I I Y G N N P A Q P H HA A S C P D R P A P C G S A D A G B H M P N P Q L Y Y G N N P A Q P H HA A S C P D R P A P C S M D A N N P Q V L Y Y G N N P A Q P H HA A S C P D R P A P C S M D A N N P Q V L Y Y G N N P A Q P H HA A S S C P D R P A P C S M D A N N P Q V L Y Y G N N P A Q P H HA A S S C P D R P A P C S M D A N N P Q V L Y Y G N N P A Q P H HA A S S C P D R P A P C G W E F D R A S N P Q I L Y Y G N N P A Q P H HA A S S C P D R P A P C G W E F D R A S N P Q I L Y Y G N N P A Q P H HA A S S C P D R P A P C G W E F D R A S N P Q I L Y Y G N N P R Q P H HA A S S C P D R P A P C G W E F D R A S N P Q I L Y Y G N N P R Q P H HA A S S C P D R P A P C G W E F D R A S N P Q I L Y Y G N N P R Q P H HA A S S C P D R	G G G G G G G G G G G G G G G G G G G		P D ON NY VD LR           P D END I Y D D XY           P D END I Y D D XY           P L VD D S Y E D KR           P L VD D S Y E D KR           P E A D D S Y A D NR           P E A D D S Y D D XR           P E A D D S Y D D XR           P E A D D S Y D D XR           P G QS D D Y D D R           P G QS D D Y D D R           P G G S N D Y T D D R           P G G S N D Y T D D R           P G QS D D Y D D S Y D D R           P G QS D D Y V D D F           P G QS D D Y V D D F           P G QS D D Y V D D F           P G QS D D Y V D D F           P G QS D D Y V D D F           P G QS D D Y V D D F           P G QS D D S Y S D V R           P G P D D S Y S D V R           P G P D D S Y S D V R           P D E A D K F D D R           P D E A D K F D D R           P D E A D K F N D H R           P D E A D K F N D H R           P D E A D K F L D F R           P D E A D K F L D F R           P D E A D K F L D F R           P D E A D K F L D H R           P D E A D K F L D H R           P D E A D K F L D F R           P D E A D K F L D F R           P D E A D K F L D F R           P D E A D K F H D H R </td <td><ul> <li>372</li> <li>373</li> <li>372</li> <li>373</li> <li>372</li> <li>372</li> <li>372</li> <li>374</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>393</li> <li>370</li> <li>370</li> <li>370</li> <li>393</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> </ul></td>	<ul> <li>372</li> <li>373</li> <li>372</li> <li>373</li> <li>372</li> <li>372</li> <li>372</li> <li>374</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>394</li> <li>393</li> <li>370</li> <li>370</li> <li>370</li> <li>393</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> <li>395</li> <li>394</li> </ul>
μP         013066281.1           μP         013066371.1           μP         013068170.1           μP         01306828.1           μP         01308428.1           μP         01308428.1           μP         01308428.1           μP         01308428.1           μP         01424612.1           μP         01424612.1           μP         01424612.1           μP         01424612.1           μP         0154250.1           μP         0154250.1           μP         0154250.1           μP         0155134471.1           μP         015502827.1	TA GINYA FLA LLAGE MG I DA - Y RKWA VE G I NY IL EA ANNA FI AL VA AE SG I E SK - Y RKWA VE G I NY IL EA ANNA FI AL LA AE SG I E SK - Y RKWA VE G I NY IL EA ANNA FLA LLA AE SG I O TR - LR WA VE G I NF IL Y AG NS AF LA LLA AAE SG V DSR - LR KWA VE G I NF IL Y AG NS AF LA LLA AAE SG I P DR - LR KWA VE G I NF IL Y SS NIAMAAL VA ADAG LN PG - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG LN AG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN AG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN AG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL A SN NAFA AL LL AAD TGL HP E - Y RHWA MCG I HY AL A SN MA FA LL AA ADG LH HE - F RHWA MCG I HY AL A ANNAM LA LI AADAG LH TE - Y RHWA MCG I HY AL AA ANNAM LA LI AADAG LH TE - Y RHWA MCG I HY AL AA ANNAM LA LI AADAG LH TE - Y RHWA MCG I HY AL AA ANNAM LA LI AADAG LH TE - Y RHWA MCG I HY AL AA ANNAM LA LI AADAG LH TE - Y RHWA MCG I HY AL AANNA FI AL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FI CL QAADG I SA SI Y Y RHWA MCG I HY AL AANNA FI CL QAADG - D GU E Y STA FA GU I NY ML HAAN VA FI VCL QAADG - D GU E - Y RG FA TD G I SY ML Y AG NTAL Y AL LT AEAG I SA TI Y REFA RQ O I NY ML HAAN VA FI VCL QAAD G - D GU E - Y RG FA TD G I SY ML Y AG NIAF I AL LA AESG LN ST Y Y REFA RQ O I NY ML HAAN VA FI VCL QAAD G - D GU E - Y RG FA TD G I SY ML Y AG NIAF I AL LA AESG LN ST Y REFA RQ O I NY ML HAAN VA FI CL EAAD - G D GE - Y RG FA TD G I SY ML HAAN VA FI CL EAAD - G D GE - Y RG FA TD G I SY ML HAAN VA FI CL EAAD - G D GE - Y RG FA TD G I SY ML HAAN VA FI CL EAAD - G D GO A - Y RE FA RQ O I HY ML HAAN VA FI CL EAAD - G D GA - Y RE FA RQ O I HY ML HAAN VA FI CL EAAD - G D GA - Y RE FA RQ O I HY ML HAAN VA FI CL EAAD - G D GA - Y RE FA RQ O I HY ML HAAN VA FI CL EAAD - G D GA - Y RE FA RQ O I HY ML HAAN VA FI CL EAAD - G D GA O - Y RE FA RQ O I HY ML HAAN VA FI CL EAAD - G D GA - Y RE FA RQ O I HY ML	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N H H G S Y Q V G D T G F S Y V V G G G G R S Y V V	Y GSK Y PLOP PH HAGAS C PD R PAC C G WA QY S P S P N P Q V L W Y GSK Y PL PP HH RGSS C P D R PAP C D WN F NS A PN P Q V L V Y GT K Y P L H P HH RGSS C P D R PAP C D WN F NS A PN P Q V L V V G T K Y P L H P HH RGSS C P D R PAP C G WD F NS P L PN P V L Q Y G T K Y P T H P HH RGAS C P D K PAP C G WD F NS P L PN P Y L Q Y G T K Y P T H P HH RGAS C P D K PAP C G WD F NS P L PN P Y L Q Y G T K Y P T H P HH RGAS C P D K PAP C G WD F NS P L PN P Y L Q Y G T K Y P T P HH RGAS C P D R PAP C G A Q L D P G P S P Q L Y Y G T K Y P T P HH RGAS C P D R PAP C G A Q L D P G P S P Y L I Y G D K Y P L P P HH RGAS C P D R PAP C G F I M S K E P N Y I L Y G C D K Y P L P P H H RGAS C P D R PAP C G F I M S K E P N Y I I L Y G C D K Y P L P P H H RGAS C P D R PAP C G F I M S K E P N Y I I L Y F G D K Y P L P P H H R S S C P N L P P C G F I M S K E P N Y I I L Y F G D K Y P L P P H H R S S C P N L P P C G T Q Y M P L K P N Y I I L Y F G D K Y P L R P H H R S S C P N L P P C G T Q Y M P L K P N Y I I L Y F G D Y P L R P Y H R S S C P N L P P C G T Q Y M P L K P N Y I I L Y F G D Y P L R P Y H R S S C P N L P P C G T Q Y M L K P N Y I I L Y F G D Y P L R P Y H R S S C P N L P P C G T Q Y M L K P N Y I I L Y F G D Y P L R P Y H R S S C P F L P A H C G P D L N A L M P N Y I I L Y F G D Y P L R P Y H R S S C P F L P A H C G P M L N L T E P N Y H L Y F G D Y P L R P Y H R S S C P F L P A H C G P M L N L T E P N Y H L Y F G D Y P L R P Y H R S S C P F L P A H C G P M L N L T E P N Y H L Y Y G N N P P A Q P H H A A S C P D R P A P C S M D Q L N A P M P Q V L Y Y G N N P P A Q P H H A A S C P D R P A P C S M D Q L N A P M P Q L Y Y G N N P P A Q P H H A A S C P D R P A P C G W E F D R A S N Q I L Y Y G N N P P C Q P H H A A S C P D R P A C G W E F D R A S N P Q I L Y W G N N P P C Q P H H A A S S C P D R P A C G W E F D R A S N P Q I L Y Y G N N P P C Q P H A A S S C P D R P A C G W E F D R A S N P Q I L Y W G N N P P C Q P H A A S S C P D R P A P C G W E F D R A S N P Q I L Y W G N N P R Q P H A A S S C P D R P A C G W E F D R A S N P Q I L Y	666666666666666666666666666666666666666		P D Q ND N Y U D LR           P D E N D I Y E D KR           P D E N D I Y E D KR           P L U D S Y E D KR           P L U D S Y E D KR           P E V D D S Y D N R           P E A D D Y N D N R           P E A D D Y D N R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q N D D Y V D F F           P G N D D Y V D F F           P G N D D S Q Y A D L R           P D E A D L P D L R           P D E A D L Y D L R           P D E A D L Y D L R           P D E A D L Y D L R           P D E A D L P D L R           P D E A D L P D L R           P D E A D L P D L R           P D E A D L P D L R           P D E A D L P D L R           P D E A D L P D L R           P D E A D L P D L R           P D E A D L P D R           P D Q N D L P U N R	372 373 372 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
AP_013066281,1           AP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066281,1           XP_013066371,1           XP_013066120,1           XP_013066120,1           XP_013066120,1           XP_013066121,1           XP_01306622,1           XP_01306622,1           XP_013064281,1           XP_01562607,1           XP_01562607,1           XP_01552267,1           XP_01552267,1           XP_0156007626,1           XP_01776628,1           XP_01776628,1           XP_01776130,1	TA GINYA FLA LLA GE MG I DA - Y RKWA VE G I NY I L EA ANNA FI AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNA FI AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNA FI AL LA AE SG V DSR - LRKWA VE G I NY I L EA ANNA FI AL LA AE SG V DSR - LRKWA VE G I NY I L Y AGN SA FLA LLA AE SG I P R - LRKWA VE G I NY I L Y AGN SA FLA LLA AE SG I P R - LRKWA VE G I NY I L Y SS NMAMAAL VA AD AG LN AG - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LN AG - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LN AG - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LN AG - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LN BE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LN BE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LH BE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LH AE - Y RHWA MC G I HY A L A AN AS AL LA AD C AG LH AE - Y RHWA MC G I HY A L Y AS NMA FAL LA AD C G LH PE - Y RHWA MC G I HY A L AA ANAMI AL I A AD AG LH AE - Y RHWA MC G I HY A L HA ANNAMI AL I A AD AG LH TE - Y RHWA MC G I HY A L HA ANNAMI AL I A AD G LH TE - Y RHWA MC G I HY A L HA ANNAMI AL I A AD G LH TE - Y RHWA MC G I HY A LA ANNAMI AL I A AD G LH TE - Y RHWA MC G I HY A L HA ANNAFI C L AA D G G I SA S - Y RKWA VE Q I NY I L HA ANNAFI C L AA D G D G D GE Y RF SA KQ I NY I L HA ANNAFI C L GA AD - G D G Y RE Y SA KQ I NY I L HA ANNAFI C L GA AD - G D G Y RE F A RQ I NY I L HA ANNAFI C L GA AD - G D G - Y RE F A RQ I NY I L HA ANNAFI C L GA AD - G D G - Y RE F A RQ I NY I L HA ANNAFI C L GA AD - O G N E - Y RE F A RQ I NY I L HA ANNAFI C L GA AD - O G N E - Y RE F A RQ I NY I L HA ANNAFI C L GA AD - O G N E - Y RE F A RQ I NY I L HA ANNAFI C L GA AD - G N E - Y RE F A RQ I NY I L HA ANNAFI C L GA AD - O G N E - Y RE F A RQ I NY I L HA ANNAFI C L GA AD - O G N E - Y RE F A RQ I NY I L HA ANNAFI C L GA AD D O E Y RE F A E Q I Y M L HA ANNAFI C L E AA D G D G N AY Y RE F A E Q I Y M L HA ANNAFI VC L E AA DG D O E Y RE F A E Q I Y M L HA ANNAFI VC L E AA DG D O E Y RE F A E Q I Y M L HA ANNAFI VC L E AA DG D O E Y N RE F A E Q I Y M L HA ANNAFI VC L E AA	G N HHG S Y V V G N FHG S Y Q V G N FHG S Y Q V G N THG S Y Q V G N THG S Y Q I G N HHG S Y Q I G N HHG S Y Q I G D TG F S Y L I G D TG F S Y V I G D TG F S Y V I G D TG F S Y V V G G G G R S Y V V	Y GSK Y PLOP PH RAGASC PD R PAC G WAQ Y S P S P N P Q V L Y Y GSK Y PL PP H HRAGASC P S P PAC C WN N F N S A P N P Q V L Y Y GT K Y P L H P H HR GSS C P D R PA P C D WN N F N S A P N P Q V L Y Y GT K Y P L H P H HR GASC P D K PA P C G WD N F N S T A P N P Q V L Y Y GT K Y P L P H H HR GASC P D K PA P C G WD N F N S T A P N P Q V L Y Y GT K Y P T H P H HR GASC P D K PA P C G WD N F N S P L P N Y L Q Y GT K Y P T H P H HR GASC P D K PA P C G WD N F N A T P S P Q V L Y Y GT K Y P T H P H HR GASC P D K PA P C G F E N L H A T T P S P Q V L Y Y GT K Y P T H P H HR GASC P D K PA P C G F E N L H A T T P S P Q V L Y Y G D K Y P L Q P H HR A S C P D R PA P C G F E N L A T P S P N V I L Y G T G K Y P H S P H HR GS C P D R PA P C G F I M S K E P N V H I L Y F G D K Y P I S P H HR A S C P D P P A C G F I M S K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P P C G T Q Y M P L K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P A P C G T Q Y M P L K E P N V H I L Y F G D Y P L R P Y G R I S M C P N P A K C G D I M M K E P S P Y V L Y F G D Y P L R P Y G R I S M C P N P A H C G P M L N L T E P N V H I L Y F G D Y P L R P Y G R I S M C P N H P A H C G P M L N L T E P N V H I L Y F G D Y P L R P Y A R I S M C P N H P A H C G P M L N L T E P N V H I L Y F G D Y P L R P Y A R I S M C P N H P A H C G P M L N L T E P N V H I L Y F G D Y P L R P Y H R S S C P F L P A H C G P M L N L T E P N V H I L Y F G D Y P L R P Y H R S S C P C R P A P C G S M D G L A M P N Q V L Y W G R D P Q Q P H HA A S C P D R P P C G S M D G E S M O D P Q V L Y W G R D P Q Q P H HA A S S C P D R P P C G S M D G E S M O N P A I Y G N N P R C P H H A S S C P D R P A C G S W D F C D R A S M P N Q I L Y W G N N P R Q P H H A A S S C P D R P A C G W S E F D R A P N Q I L Y W G N N P R Q P H H A A S S C P D R P A C G W S E F D R A P N Q I L Y W G N N P K Q P H H A A S S C P D R P A C G W S E F D R A P N Q I L Y W G N N P K Q P H H A A S S C P D R P A C G W S E F D K D A P N Q I L Y W G N N P K Q P H H A A S S C P D R P A C G W S E F D K D A P N Q I L Y W G	G G G G G G G G G G G G G G G G G G G	X × × × × × × × × × × × × × × × × × × ×	P D Q ND NY VD LR           P D E ND I Y D D K R           P D E ND I Y D D K R           P D E ND I Y D D K R           P D E ND I Y D D K R           P L V D D S Y D D K R           P E A D D S Y A D N R           P E A D D S Y A D N R           P E A D D S Y A D N R           P E A D D S Y A D N R           P E A D S Y A D N R           P E G S D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G S N D D Y T D D R           P G A D L Y D V E F           P G A D L Y D V E F           P G A D L Y D L R           P D E A D K F D D R           P D E A D K F D D R           P D E A D K F D D R           P D E A D K F D D R           P D E A D K F D D R           P D E A D K F D D R           P D E A D K F D D R           P D E A D K F H D H R           P D E A D K F H D H R           P D E A D K F H D H R <td>372 373 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39</td>	372 373 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
μP_013066281,1           μP_013064281,1           μP_013084281,1           μP_01508427,1           μP_015128607,1           μP_015138247,11           μP_0155138247,11           μP_0155138247,11           μP_0155138247,11           μP_0156382927,11           μP_0177652941,11           μP_0177782094,11           μP_0177782094,11           μP_0177782094,11           μP_0177782094,11           μP_0177782094,1	TA GIN AF LA LL AGE BIG ID AK - Y RKWA VE G INY IL EA ANNAF I AL VA AE SG IE SK - Y RKWA VE G INY IL EA ANNAF I AL LA AE SG VI SK - Y RKWA VE G INY IL EA ANNAF I AL LA AE SG VI SK - L RKWA VE G INY IL Y AGNSAF LA LL AA EGG IN TO - Y RKWA VE G INF IL Y AGNSAF LA LLAAE SG IP PR - L RKWA VE G INF IL Y SK NAMAAL VA AD AG LN GO - Y RKWA VE G INY IL Y SK NAMAAL VA AD AG LN GO - Y RHWA MCG IHY AL Y SK NAMAAL VA AD AG LN GO - Y RHWA MCG IHY AL Y SK NAMAAL VA AD AG LN GO - Y RHWA MCG IHY AL Y SK NAMAAL VA AD AG LN SC - Y RHWA MCG IHY AL Y SK NAMAAL VA AD AG LN SC - Y RHWA MCG IHY AL Y SK NAMAAL VA AD AG LH SE - Y RHWA MCG IHY AL Y SK NAMAAL VA AD AG LH SE - Y RHWA MCG IHY AL Y SK NAMAAL VA AD AG LH SE - Y RHWA MCG IHY AL A SK AN AMAL LA AD AG LH RE - Y RHWA MCG IHY AL A SK AN AMAL LI AA AD AG LH TE - Y RHWA MCG IHY AL HAA NAMAL AL I AAD AG LH TE - Y RHWA MCG IHY AL HAA NAMAL AL I AAD AG LH TE - Y RHWA MCG IHY AL HAA NAMAL AL I AAD AG LH TE - Y RHWA MCG IHY AL HAA NAMAL AL I AAD AG LH TE - Y RHWA MCG IHY AL HAA NA SK Y CL QA ADG IN TE - Y RHWA MCG IHY AL HAA NA SK Y CL QA ADG IN TE - Y RHWA MCG IHY AL HAAN Y SY CL QA ADG DOE Y NG FA TD OI Y NI HAAN Y SY CL QA ADG DOE - Y NG FFA TD OI Y NI HAAN Y SF VCL QA AD G - DOE - Y NG FFA TD OI Y NI HAAN Y SF VCL QA AD G DOE - Y NG FFA RG OI HY M HAAN Y SF VCL GA AD G - O DOE - Y RE FFA RG OI HY M HAAN Y SF VCL FA AD G ON C - Y NG FFA TD OI Y NI HAAN Y SF ICL EA AD - G DR E - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD - G DR E - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD - G DQ E - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD - G DQ E - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD - G DQ E - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD - G DQ E - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD - C DUG - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD - C DUG - Y RE FFA RG OI HY ML HAAN Y SF ICL EA AD O LS Y NE FFA RG OI HY ML HAAN Y SF UCL EA ADG DUG Y NE FFA RG OI HY ML HAAN Y SF UCL EA ADG O LS Y NE FFA RG OI HY ML HAAN Y SF UCL EA ADG O LS Y NE FFA RG OI HY ML HAAN Y SF UCL EA ADG	G N HIGS Y WV G N FIGS Y QV G N FIGS Y QV G N TIGS FQ I G N HIGS Y QI G D TGFS Y VI G G G G S Y VI	Y GSK Y PLOP PH HAGAS C PD R PA C G WA QY S P S P N P Q V L Y Y GSK Y PL PP H HR GSS C P D R PA P C D WN F N S A P N P Q V L Y Y GT K Y P L H PH HR GSS C PD R PA P C D WN F N S A P N P Q V L Y V GN K Y P L Q PH HR AS C P D K PA P C G WD F HA T T P S P Q V L Y GT K Y P T H P H HR GA S C P D K PA P C G WD F HA T T P S P Q V L Y GT K Y P T H P H HR GA S C P D K PA P C G WD F HA T T P S P Q V L Y GT K Y P T H P H HR GA S C P D K PA P C G F L N A T P S P Q V L Y G T K Y P T H P H HR GA S C P D K PA P C G F L N A P C B P P Y V L Y G T K Y P T H P H HR GA S C P D K PA P C G F L N A T P S P Q V L Y G T K Y P T H P H HR GA S C P D K PA P C G F L N S K P P Y H L Y G D K Y P L P P H HR S S C P N L P P C G F I M S K E P N Y H L Y F G D K Y P L P P H HR S S C P N L P P C G F I M S K E P N Y H L Y F G D K Y P L P P H HR S S C P N L P P C G F I M S K E P N Y H L Y F G D K Y P L R P H HR S S C P N L P P C G F I M S K E P N Y H L Y F G D Y P L R P H HR S S C P N L P P C G F I M S K E P N Y H L Y F G D Y P L R P H HR S S C P N L P A K C S Q I M P M K E P S P H V L I F G D Y P L R P Y HR S S C P N L P A K C S Q I M P M K E P S P H V L I F G D Y P L R P Y H R S S C P N L P A H C G P M L N L T E P N Y H I Y F G D Y P L R P Y H R S S C P N P A P C G S A D A G B H M P M P O V L T F G D Y P L R P Y H R S S C P N R P A P C G S A D A G B H M P M P O V L T Y G N N P A Q P H HA A S C P D R P A P C S M D A N N P O V L Y Y G N N P A Q P H HA A S C P D R P A P C S M D A N N P O V L Y Y G N N P A Q P H HA A S S C P D R P A P C S M D A N N P O V L Y Y G N N P A Q P H HA A S S C P D R P A P C G W S E F D R A D P N O I L Y Y G N N P A Q P H HA A S S C P D R P A P C G W S E F D R A D P N O I L Y Y G N N P A Q P H HA A S S C P D R P A P C G W S E F D R A D N N O I L Y Y G N N P A Q P H HA A S S C P D R P A P C G W S E F D R A D N N O I L Y Y G N N P A Q P H HA A S S C P D R P A P C G W S E F D R A D N N O I L Y Y G N N P A Q P H HA A S S C P D R P A P C G W S E F D R A D N N O I L Y Y G N N P K Q P H HA A S S C P D R P A P C G W S E F D R A D N N	G G G G G G G G G G G G G G G G G G G	0         0	P         D         N         Y         D           P         D         N         Y         D         K           P         D         D         Y         D         K         K           P         D         D         S         Y         D         K	372 373 372 372 372 372 372 371 394 394 394 394 394 394 394 394 393 303 370 370 370 370 370 370 370 393 393 394 394 394 394 394 394 394 394
PP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306628.1           XP_01306637.1           XP_01306637.1           XP_01306637.1           XP_0130667.1           XP_0130667.1           XP_0130667.1           XP_0130667.1           XP_0130667.1           XP_0130667.1           XP_0130667.1           XP_0130667.1           XP_0130667.1           XP_01306428.1           XP_01308428.1           XP_01308428.1           XP_01308428.1           XP_01308428.1           XP_01308428.1           XP_01308428.1           XP_01308428.1           XP_01308428.1           XP_01424612.1           XP_015425.1           XP_015425.1           XP_015425.1           XP_01551382.1           XP_01551382.1           XP_01551382.1           XP_01551382.1           XP_01551382.1           XP_01551382.1           XP_01551382.1           XP_01778510.2           XP_01778510.2           XP_01778510.2	TA GINYA FLA LLAGE MG I DA - Y RKWA VE G I NY IL EA ANNA FI AL VA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FI AL LA AE SG I E SK Y RKWA VE G I NY IL EA ANNA FLA LLA AE SG I O TR - L RKWA VE G I NY IL EA ANNA FLA LLA AA ESG VD SR - L RKWA VE G I NF IL Y AG NS AF LA LLA AAE SG I P DR - L RKWA VE G I NF IL Y AG NS AF LA LLA AAE SG I P DR - L RKWA VE G I NF IL Y SS NIAMAAL VA ADAG LN PG - Y RKWA VE G I NY IL Y SS NIAMAAL VA ADAG LN AG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN AG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN AG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG LN BE - Y RHWA MCG I HY AL A SN NAFA AL LL AAD TGL HP E - Y RHWA MCG I HY AL A SN AN AAL LA ADAG LH TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG LH TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG LH TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG LH TE - Y RHWA MCG I HY AL AANNA FLAL LA ADAG L S AS - Y REWA Y E O I NY ML HAANNA FLAL LA AE SG I S AS - Y REWA Y E O I NY ML HAANNA FVCL GA ADG - D O E - Y RE FA RQ O I NY ML HAANNA FVCL GA ADG - D O E - Y RE FA RQ O I NY ML HAANNA FVCL GA ADG - D O E - Y RE FA RQ O I NY ML HAANNA FVCL GA ADG - D O E - Y RE FA RQ O I NY ML HAANNA FVCL E AADG - N O E - Y RE FA RQ O I NY ML HAANNA FVCL E AADG - D O E - Y RE FA RQ O I NY ML HAANNA FVCL E AADG D C Y RE FA RQ O I NY ML HAANNA FVCL E AADG D C Y RE FA RQ O I NY ML HAANNA FVCL E AADG D C Y RE FA RQ O I NY ML HAANNA FVCL E AADG D C Y RE FA RQ O I NY ML HAANNA FVCL E AADG D C Y NE FA RQ O I NY ML HAANNA FVCL E AADG D C Y NE FA RQ O I NY ML	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N H H G S Y Q V G D T G F S Y V V G G G G R S Y V V G S T G R S Y V V C S T G R S Y V V C S S T G S Y V V C S S T G S S Y V V C S S T G S Y V V C S S T G S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V C S S T S S S Y V V S S S S S Y V V C S S T S S S Y V V S S S S S Y V V S S S S S Y V V S S S S	Y GSK Y PLOP PH HAGAS C PD R PASC G WAQ Y S P S P N P Q V H Y GSK Y PL PP HHRASS C P S P PAC C NWD T YN LHSP N PH V L N Y GT K Y P L HP HHRGSS C PD R PAP C D WN N F NS A P N P Q V L V V GT K Y P L HP HHRGSS C PD R PAP C G WN F NS T A P N P Q V L V Y GT K Y P L HP HHRGSS C PD K PAP C G WN F NS P L P N P Y L Q Y GT K Y P T HP HHRGAS C PD K PAP C G WN F NS P L P N P Y L Q Y GT K Y P T HP HHRGAS C PD K PAP C G WN F NS P L P N Y L Q Y GT K Y P T HP HHRGAS C PD K PAP C G WN F NS P L P N Y L Q Y GT K Y P T HP HHRGAS C PD R PAP CS G A Q L N P G PS P Q L Y Y GT K Y P T HP HHRGAS C PD R PAP CS G A Q L N P G PS P N Y L I Y GD K Y P L P P HHRGAS C PD R PAP CS G A Q L N P G PS P N Y L I Y GD K Y P L P P HHRGAS C PD R PAP CS G T A Q L N P G PS P N Y L I Y GD K Y P L P P HHRS SS C P N L P P C G F I M S K E P N Y H I L Y F GD K Y P L R P HHRS SS C P N L P P C G T Q Y M P L K E P N Y H I L Y F GD K Y P L R P HHRS SS C P N L P P C G T Q Y M P L K E P N Y H I L Y F GD Y P L R P Y H R M SK C P N P P A K C G D I M P M K E P S P H Y L I F GD Y P L R P Y H R M SK C P N P A H C S P U L L A E M P N I H L Y F GD Y P L R P Y H R M SK C P N P A H C S P U L L A E M P N I H L Y F GD Y P L R P Y H R M SK C P N P A A I C T P E H L A A K L P N Y H I L Y F GD Y P L R P Y H R M SK C P N F A P C S A D A Q B H O P N Q L I Y Y S N Y P A Q P H HAASS C P D R P A P C S M Q L N A P G S P Q V L Y W G N P P A Q P H HAASS C P D R P A P C S M D Q L N A P G S P Q V L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P T Q P H H A A SS C P D R P A C G W E F D R A S N P Q I L Y Y G N N P K Q P H H A A SS C P D R P A C G W E	666666666666666666666666666666666666666		P D Q ND N Y U D LR           P D E ND I Y U D KR           P D E ND I Y U D KR           P L U D S Y U D KR           P L U D S Y U D KR           P L U D S Y U D KR           P L U D S Y U D KR           P E A D D Y N D KR           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S D D Y T D D R           P G Q S N D Y T D D R           P G Q S N D Y T D D R           P G Q S N D Y T D D R           P G Q S N D Y T D D R           P G Q S N D Y T D D R           P G Q N D Y Y D D R           P G Q N D Y Y D D R           P G Q N D Y Y D D R           P G Q N D Y Y D D R           P G A N D E Y L D KR           P D D S Y S D Y R           P D D S Y S D Y R           P D D S Q Y A D L Y C L L R           P D E A S D L Y E L L R           P D E A S D L Y E L R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R	372 373 372 372 372 372 371 394 394 394 394 394 394 394 394 394 393 370 370 370 370 370 393 394 394 394 394 394 394 394 394 394
AP_013066281,1           AP_013066281,1           AP_013066281,1           AP_013066281,1           AP_013066281,1           AP_013066281,1           AP_013066281,1           AP_013066371,1           AP_013066371,1           AP_013066521,1           AP_01306652,1           AP_01306652,1           AP_01306652,1           AP_01306652,1           AP_01306652,1           AP_01306652,1           AP_013064281,1           AP_013084281,1           AP_015084281,1           AP_015084281,1           AP_015126097,1           AP_015522697,1           AP_015582297,1           AP_015582297,1           AP_015582297,1           AP_015582297,1           AP_015582297,1           AP_015582297,1           AP_0155828219,1           AP_01776289,1	TA GINYA FLA LLAGE MG I DA - Y RKWA VE G I NY L EA ANNA FI AL VA AE SG I E SK - Y RKWA VE G I NY I EA ANNA FI AL VA AE SG I E SK - Y RKWA VE G I NY I EA ANNA FI AL LA AE SG V DSR - LRKWA VE G I NY I EA ANNA FI AL LA AE SG V DSR - LRKWA VE G I NY I Y ANNA ST AL LA AA ESG V DSR - LRKWA VE G I NY I Y SS NMAMAAL VA AD AG LN AG - Y RKWA VE G MY I Y SS NMAMAAL VA AD AG LN AG - Y RKWA VE G MY I Y SS NMAMAAL VA AD AG LN AG - Y RKWA VE G MY I Y SS NMAMAAL VA AD AG LN AG - Y RHWA MC G I HY A Y SS NMAMAAL VA AD AG LN AG - Y RHWA MC G I HY A Y SS NMAMAAL VA AD AG LN BE - Y RHWA MC G I HY A Y SS NMAMAAL VA AD AG LN BE - Y RHWA MC G I HY A Y SS NMAMAAL VA AD AG LN BE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG LH BE - Y RHWA MC G I HY A L AS ANNAMLAL I A AD AG LH BE - Y RHWA MC G I HY A HAAN LA LI AA DAG LH TE - Y RHWA MC G I HY A L AANNAMLAL I AAD AG LH TE - Y RHWA MC G I HY A HAAN LAL VA LVAA DNG I HT E - Y RHWA MC G I HY A L AANNAMLAL I AAD AG LH TE - Y RHWA MC G I HY A HAAN LAL VAL VAAD NG I HT E - Y RHWA MC G I HY A HAAN LAL VAL VAAD SG I S AS - Y NG FA TO G I Y M HAAN Y F VC LQ AAD G G D E - Y NG FS TA G I NY M HAAN Y F VC LQ AAD G D C - D GU FY SF F A KQ I HY I HAAN Y F VC LQ AAD G O D C - Y NG FS TA CQ I HY I HAAN Y F VC LQ AAD G O D C - Y RE FS A KQ I HY I HAAN Y F VC LQ AAD G O D C - Y RE F A RQ I HY I HAAN Y F VC LQ AAD G O D C - Y RE F A RQ I HY I HAAN Y F VC LQ AAD G O D C - Y RE F A RQ I HY I HAAN Y F VC LQ AAD G O D C - Y RE F A RQ I HY I HAAN Y F VC LQ AAD G O D C - Y RE F A RQ I HY I HAAN Y F VC LG AAD G - D O C - Y RE F A RQ I HY I HAAN Y F VC LG AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY I HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY M HAAN Y F VC LE AAD G - D C C Y RE F A RQ I HY M HAA	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N T H G S Y Q V G N T H G S Y Q I G N H H G S Y Q I G D T G F S Y V V G D G G R S Y V V G G G G G S Y V V G G G G G S S Y V V G G G G G S Y V V G G G G S Y V V G G G G S Y V V Y G G G G S Y V Y Y G G G G S Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y	Y GSK Y PLOP PH RAGAS C PD R PAC C G WA QY S P S P N P Q V L Y Y GSK Y PL PP H HR AGAS C P S P PAC C WN N F NS A P N P Q V L Y Y GT K Y P L H PH HR GSS C P D R PAP C G WN N F NS A P N P Q V L Y V GT K Y P L H PH HR GAS C P D K PAP C G WD N F NS T A P N P Q V L Y Y GT K Y P L H PH HR GAS C P D K PAP C G WD N F NS P L P N P V L Q Y GT K Y P T H P H HR GAS C P D K PAP C G WD N F N S P L P N P V L Q Y GT K Y P T H P H HR GAS C P D K PAP C G WD N F N S P L P N Y L Y G T K Y P T H P H HR GAS C P D K PAP C G F E N L H A T T P S P Q V L Y Y G T K Y P T H P H HR GAS C P D K PAP C G F E N L H A T T P S P N V L I Y G D K Y P L Q P H HR A S C P D R PAP C G F E N L M T P S P N V L I Y G D K Y P L P P H HR GS C P D R PAP C G F I M S K E P N V H I L Y F G D K Y P I S P H HR A S C P D R PAP C G F I M S K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P P C G T Q V M P L K E P N V H I L Y F G D K Y P L R P H HR S S C P N L P P C G T Q V M P L K E P N V H I L Y F G D Y P L R P Y G R I S M C P N P A K C G O I M P M K E P S P H V L Y F G D Y P L R P Y G R I S M C P N P A H C S P E V L A EMP N I H I L Y F G D Y P L R P Y G R I S M C P N H P A H C S P E V L A EMP N I H I L Y F G D Y P L R P Y G R I S M C P N H P A H C S P E V L A EMP N I H I L Y F G D Y P L R P Y G R I S M C P N H P A H C G P M L N L T E P N V H L Y F G D Y P L R P Y M R S S C P F L P A H C G P M L N L T E P N V H L Y F G D Y P L R P Y H R S S C P D R P A C G S M D G L A E M P N V H L Y Y G S K Y P L P H H R A S C P D R P A C G S M D G L A E M P N Q V L Y G Y G N N P A Q P H HA A S S C P D R P A C G S M D G L A M P N Q V L Y Y G N N P R Q P H HA A S S C P D R P A C G W S E T D R A P N P Q V L Y G Y G N N P R Q P H H A A S S C P D R P A C G W S E F D R A P N Q I L Y Y G N N P K Q P H H A A S S C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P K Q P H H A A S S C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P K Q P H H A A S S C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P K Q P H H A A S S C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P K Q	G G G G G G G G G G G G G G G G G G G		P D AND NY UDLR           P D ENDI Y ODKR           P D ENDI Y ODKR           P D ENDI Y ODKR           P L UDDS Y EDKR           P E V DDS Y DKR           P E ADDS Y ADNR           P E ANDN H DLR           P G SN DD Y T DDR           P G SN DD SY S DVR           P D S S OY AD LR           P D E AD K F D HR	372 373 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
μP         013066221.1           μP         01306628.1           μP         01306828.1           μP         01306828.1           μP         01306828.1           μP         01306828.1           μP         01306828.1           μP         01306828.1           μP         01306428.1           μP         01308428.1           μP         0150847.1           μP         0151328.1           μP         01551328.2           μP         01551328.2           μP         01563629.1           μP	TA GIN AF LA LLAGE BIG ID A Y RK WA VE G INY IL EA ANNAF I AL VA AE SG IE SK Y RK WA VE G INY IL EA ANNAF I AL VA AE SG IE SK Y RK WA VE G INY IL EA ANNAF I AL LA AE SG VD SR LRK WA VE G INY IL Y AGNSAF LA LLAAE SG I D R - LRK WA VE G INF IL Y AGNSAF LA LLAAE SG I PR - LRK WA VE G INF IL Y AGNSAF LA LLAAE SG I PR - LRK WA VE G INF I Y SS NIAMAAL VA AD AG LN G - Y RK WA VE G INY I Y SS NIAMAAL VA AD AG LN G - Y RHWA MC G I HY A Y SS NIAMAAL VAAD AG LN G - Y RHWA MC G I HY A Y SS NIAMAAL VAAD AG LN G - Y RHWA MC G I HY A Y SS NIAMAAL VAAD AG LN G - Y RHWA MC G I HY A Y SS NIAMAAL VAAD AG LN G - Y RHWA MC G I HY A Y SS NIAMAAL VAAD AG LN G - Y RHWA MC G I HY A Y SS NIAMAAL VAAD AG LN SE - Y RHWA MC G I HY A A N SS AF VAL SA G AG LN G - Y RHWA MC G I HY A A SS NIAMAAL VAAD AG LH SE - Y RHWA MC G I HY A A SN AN AAL LAAD AD LH HE - F RH WA MC G I HY A HAANNAAL LI AAD AG LH AE - Y RHWA MC G I HY A LAAN AG LI AAD AG LH AE - Y RHWA MC G I HY A HAAN AL AL I AAD AG LH TE - Y RHWA MC G I HY A HAAN AL AL I AAD AG LH TE - Y RHWA MC G I HY A HAAN NAFU C LO AAD G H TE - Y RHWA MC G I HY A HAAN NAFU C LO AAD G I TT E - Y RHWA MC G I HY A HAAN AL AL VAL VAAD NG I HT E - Y RHWA MC G I HY A HAAN Y FV C LO AAD G - D OD E - Y NG F A RG O I NY IL Y AG NAFI LL AE SG IS SA - Y NG F S A F O I NY IL HAAN Y FV C LO AAD G - D OD E - Y NG F A RG O I NY IL HAAN Y FV C LO AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LO AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY M HAAN Y FV C LE AAD G - D OD E - Y NG F A RG O I HY	G N N H G S Y V V G N P H G S Y Q V G N P H G S Y Q V G N T H G S Y Q V G N H G S Y Q V G D T G F S Y V V G G G G R S Y V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V	Y GSK Y PLOP PH RAGASC PD R PACC GWAQYS S PS PN PQ V L N Y GSK Y PL PP H PH RAGSS C PD R PAP C D WNN F NS A PN PQ V L V Y GT K Y P L HP H H RGSS C PD R PAP C D WNN F NS A PN PQ V L V VG NK Y P L Q PH H RAASC PD K PAP C G WD F NS F L PN P V V L Q Y GT K Y P T HP H H RGASC PD K PAP C G WD F NS F L PN PY V L Q Y GT K Y P T HP H H RGASC PD K PAP C G F L NA T T PS P Q V L Y Y GT K Y P T HP H H RGASC PD K PAP C G F L NA T T PS P Q V L Y Y GT K Y P T HP H H RGASC PD K PAP C G F L NA T T PS P Q V L Y Y G K Y P L Q PH H RAASC PD K PAP C G F L NA T PS P Y V L I F G DK Y P L P P H H RGASC PD K PAP C G F I M SS K E PN Y H I L Y F G DK Y P L P P H H RGASC PD L PAP C G F I M SS K E PN Y H I L Y F G DK Y P L P P H H RS SC PN L PAP C G F I M SS K E PN Y H I L Y F G DK Y P L R P H H RS SC PN L PAP C G F I M SS K E PN Y H I L Y F G D Y P L R P H H RS SC PN L PAP C G F I M SS K E PN Y H I L Y F G D Y P L R P H H RS SC PN L PAP C G F I M S K E PN Y H I L Y F G D Y P L R P H H R SS C PN L PAP C G F I M SS K E PN Y H I L Y F G D Y P L R P Y H R SS C PN L PAP C G P M L N L T E PN Y H I L Y F G D Y P L R P Y H R SS C PN P PA K C S Q I M P M K E PS P H V L Y F G D Y P L R P Y H R SS C P N P PA C G P M L N L T E PN Y H I L Y F G D Y P L R P Y H R SS C P F L PA H C G P M L N L T E PN Y H T L Y F G D Y P L R P Y H R SS C P N R PA P C G S A D A G H G PN N Q V L T Y G N N P P A Q P H H R A SC P N R PA P C S M D A C M A P N P Q V L T Y G N N P P A Q P H H A A SC P D R PA P C S M D A L A K L P N Y H I L Y Y G N N P P A Q P H H A A SC P D R P A C G W S E F D R A SN P Q I L Y W G N N P P A Q P H H A A SC P D R P A C G W S E F D R A SN P Q I L Y W G N N P F A Q P H H A A SC P D R P A C G W S E F D R A SN P Q I L Y W G N N P K Q P H H A A SC P D R P A C G W S E F D R A SN P Q I L Y W G N N P K Q P H H A A SS C P D R P A C G W S E F D R A SN P Q I L Y W G N N P K Q P H H A A SS C P D R P A C G W S E F D K D A P N Q I L Y W G N N P K Q P H H A A SS C P D R P A C G W S E F D K D A P N Q I L Y W G N N P K Q P H H A A SS C P D R P A C G W S E F D K D	G G G G G G G G G G G G G G G G G G G		P         P           P         P           P         P           P         P           P         P           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         P	372 373 372 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
NP_01306628.1           NP_01306617.1           NP_01306628.1           NP_01306628.1           NP_01306628.1           NP_01306628.1           NP_01306428.1           NP_01306428.1           NP_01306428.1           NP_01306428.1           NP_01306428.1           NP_01306428.1           NP_01306428.1           NP_01306428.1           NP_01306428.1           NP_01308428.1           NP_01308428.1           NP_01308428.1           NP_0140428.1           NP_0142428.1           NP_01508428.1           NP_01518428.1           NP_0154258.1           NP_0154258.2           NP_015582627.1           NP_0156828.2           NP_0157820.1           NP_0157820.1           NP_01580188.1           NP_01580188.1           NP_01580188.1	TA GIN AF LA LL AGE ING ID AR - Y RKWA VE GUNY LL EA ANNAF I AL VA AE SGIE SK Y RKWA VE GUNY IL EA ANNAF I AL LA AE SGUT R LR WA VE GUNY IL EA ANNAF LA LLA AAE SGUT R LR WA VE GUNY IL EA ANNAF LA LLA AAE SGUT R LR WA VE GUNY IL YAG NS AF LA LLA AAE SGUT R LR WA VE GUNY IL YAG NS AF LA LLA AAE SGUT R LR WA VE GUNY IL YAG NS AF LA LLA AAE SGUT R LR WA VE GUNY IL YAG NS AF LA LLA AAE SGUT R LR WA VE GUNY IL YS SNAMAAL VA ADAG LN AG - Y RHWA MC GUNY IL YS SNAMAAL VA ADAG LN AG - Y RHWA MC GUNY IL YS SNAMAAL VA ADAG LN AG - Y RHWA MC GUNY IL YS SNAMAAL VA ADAG LN AG - Y RHWA MC GUNY IL YS SNAMAAL VA ADAG LN BE - Y RHWA MC GUNY IL YS SNAMAAL VA ADAG LN BE - Y RHWA MC GUNY IL YS SNAMAAL VA ADAG LN BE - Y RHWA MC GUNY IL YS SNAMAAL VA ADAG LN BE - Y RHWA MC GUNY IL AA SNAFAL LLA AD TGL HP E - Y RHWA MC GUNY IL AA SNAFAL LA ADAG LH HE - F RHWA MC GUNY IL AA ANNAMLAL I AADAG LH TE - Y RHWA MC GUNY IL AAANAAL AL VA LVA AD MG INT E - Y RHWA MC GUNY IL AAANAAL AL VA LVA AD MG INT E - Y RHWA MC GUNY IL AAANA AL VALVA AD GUNG INT E - Y RHWA MC GUNY IL AAANAF VC LQA AD GONG E - Y ROFAT DGUNY IL YAG NNAF FVC LQA AD GONG E - Y ROFAT DGUNY IL AAANVAF VC LQA AD GONG E - Y ROFAT DGUNY IL AAANVAF VC LQA AD GONG E - Y ROFAT DGUNY IL AAANVAF VC LQA AD GONG E - Y ROFAT DGUNY IL AAANVAF VC LQA AD GONG E - Y ROFAT DGUNY IL AAANVAF VC LQA AD GONG E - Y ROFAT DGUNY IL AAANVAF VC LGA AD GONG E - Y ROFAT DGUNY IL AAANVAF VC LGA AD G - DOR E - Y ROFAT DGUNY IL AAANVAF VC LGA AD G - DOR E - Y ROFAT DGUNY IL AAANVAF VC LGA AD G - DOR E - Y ROFAT DGUNY IL AAANVAF VC LGA AD G - DOR E - Y ROFAT DGUNY IL AAANVAF VC LGA AD G - DOR E - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR E - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR - Y ROFAT DGUNY ML AAANVAF VC LGA AD G - DOR - Y ROFAT DGUNY ML AAANVAF VC LGA AD G	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N H H G S Y Q V G N H H G S Y Q V G D T G F S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G S T G R S F N V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V G S T G R S F N V V	Y GSK Y PLOP PH HAGAS C PD R PAC G WAQ Y S P S P N P Q V H Y GSK Y PL PP H HR HAGAS C P D R PA P C D WN F N S A P N P Q V L Y Y GT K Y P L H P H HR GSS C P D R PA P C D WN F N S A P N P Q V L Y Y GT K Y P L H P H HR GSS C P D R PA P C G WD N F N S P L P N P V L Q Y GT K Y P T H P H HR GA S C P D K PA P C G WD N F N S P L P N P Y L Q Y GT K Y P T H P H HR GA S C P D K PA P C G WD F N S P L P N P Y L Q Y GT K Y P T H P H HR GA S C P D K PA P C G WD F N A T T P S P Q V L Y Y GT K Y P T H P H HR GA S C P D K PA P C G WD F N A T T P S P Q V L Y Y GT K Y P T H P H HR GA S C P D K PA P C G F L N A T T P S P Q V L Y Y GT K Y P T H P H HR GA S C P D R PA P C S G A Q L N P F O P S P V L I Y GD K Y P L Q P H HR A S C P D R PA P C S G A Q L N P F O P N V I L Y G D K Y P L P P H HR S S C P D L P P C G F I M S K E P N V I I L Y G G D K Y P L P P H HR S S C P N L P P C G F I M S K E P N V I I L Y F G D K Y P L R P H HR S S C P N L P P C G T O Y M P L K E P N V I I L Y F G D Y P L R P H HR S S C P N L P P C G T O Y M P L K E P N V I I L Y F G D Y P L R P Y H R M S C P N P A K C G D I M M K E P S P V V I L Y F G D Y P L R P Y H R M S C P N H P A I C T P E H L A A K L P N V H I L Y F G D Y P L R P Y H R M S C P N H P A I C T P E H L A A K L P N V H I L Y F G D Y P L R P Y H R M S C P N H P A I C T P E H L A A K L P N V H I L Y F G D Y P L R P Y H R M S C P N H P A I C T P E H L A A K L P N V H I L Y F G D Y P L R P Y H R M S C P N F P A C G S M D A D A P O P N O I L Y Y G N N P N Q P H HA A S C P D R P A P C S M D Q L N P O P O I L Y Y G N N P P Q P H H A A S C P D K P A P C S M D Q L N P O P O I L Y Y G N N P T Q P H H A A S C P D R P A C G W E F D R A S N O I L Y Y G N N P T Q P H H A A S C P D R P A C G W E F D R A S N O I L Y Y G N N P T Q P H H A A S S C P D L P A C G W E F D R A S N P O I L Y Y G N N P T Q P H H A A S S C P D R P A C G W E F D R A S N P O I L Y Y G N N P T Q P H H A A S S C P D R P A C G W E F D R A S N P O I L Y Y G N N P T Q P H H A A S S C P D R P A C G W E F D R A S N P O I L Y Y G N N P T Q P H H A A	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		P D Q ND N Y U D LR           P D E ND I Y U D KR           P D E ND I Y U D KR           P D E ND I Y U D KR           P L U D S Y U D KR           P L U D S Y U D KR           P E V D D S Y D KR           P E A ND N Y D VR           P E A ND N Y D VR           P E S ND N T D D R           P G Q S D D Y T D D R           P G Q S ND Y T D D R           P G Q S ND Y T D D R           P G Q S ND Y T D D R           P G Q S ND Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y T D D R           P G A ND E Y L D KF           P G A ND E Y L D KF           P G A ND E Y D D R           P D A D N E Y D D R           P D A D N E Y D L R           P D E A S D I Y E D L R           P D E A D K F H D H R           P D E A D K F H D H R           P D E A D K F H D H R           P D E A D K F H D H R           P D E A D K F H D H R           P D E A D K F H D H R           P D E A D K F H D H R           P D E A D N F R D H R           P D A N Y V D R      P	372 373 372 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066281,1           μP_013066371,1           μP_013066371,1           μP_013066321,1           μP_013066321,1           μP_01306632,1           μP_01306632,1           μP_01306622,1           μP_01306622,1           μP_01306622,1           μP_01306422,1           μP_01306428,1           μP_01306428,1           μP_01304428,1           μP_01308428,1           μP_01512667,1           μP_01512667,1           μP_015526297,1           μP_015626361,1           μP_015612637,1           μP_01561	TA GWA FLALL AGE MG I DA - Y RKWA VE G I NY I L EA ANNAF I AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNAF I AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNAF I AL LA AE SG V DSR - L RKWA VE G I NY I L EA ANNAF I AL LA AE SG V DSR - L RKWA VE G I NY I L Y AGN SAF LA LA AA ESG V DSR - L RKWA VE G I NY I L Y AGN SAF LA LA AAE SG I P P A - L RKWA VE G I NY I L Y SS NMAMAAL VA AD AG L NAG - Y RKWA VE G MY I L Y SS NMAMAAL VA AD AG L NAG - Y RKWA VE G MY I L Y SS NMAMAAL VA AD AG L NAG - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L NAG - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L NAG - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N BE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H BE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H BE - Y RHWA MC G I HY A L A AN AS AL LA AD C GL H PE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H AE - Y RHWA MC G I HY A L A AN AMMAL AL I A AD AG L H TE - Y RHWA MC G I HY A L AA AN AS AL LA AD GL H TE - Y RHWA MC G I HY A L AA AN AS L AL LA AD GL H TE - Y RHWA MC G I HY A L AA AN AS L L I AA DAG L H TE - Y RHWA MC G I HY A L AA AN AS FLAL VA AD NG I H TE - Y RHWA MC G I HY A L AA AN AS FLAL VA AD NG I H TE - Y RHWA MC G I HY A L AA AN AS FLAL VA AD G S I S AS - Y NG FA T D G I NY ML HAA NN AF VC L Q AA DG G I S AS - Y NG FA T D G I NY ML HAA NN AF I C L G AAD - G D G E Y RG FA T D G I NY ML HAA NN AF I C L G AAD - G D G A - N U F S FA KQ O I HY I L HAA NN AF I C L G AAD - G D G A - Y RE FA AG G I HY ML HAA NN AF I C L G AAD - G D G A - Y RE FA AG G I HY ML HAAN NA FI C L G AAD - G D G A - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE FA AG G I HY ML HAAN NA FI C L E AAD G - N G E - Y RE	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N H H G S Y Q V G D T G F S Y V V G G G G R S Y V V G G S S G R S Y V V G G S S G R S Y V V G G S S G R S Y V V G S S G R S Y V V	Y GSK Y PLOP HH AGASC PD D RASC GWAQYS S PS PN PQ U H Y GSK Y PLOP HH RASSC PD N PACC NWD T YN LHSPN PH VU N Y GT K Y PL HP HH RGSS C PD R PAP C D WNN F NS A PN PQ U U Y GT K Y PL HP HH RGSS C PD R PAP C G WD N F NS T AP N PQ U U Y GT K Y PT HP HH RGASC PD K PAP C G WD N F NS T AP N PQ U U Y GT K Y PT HP HH RGASC PD K PAP C G WD N F NA T PS PQ U U Y GT K Y PT HP HH RGASC PD K PAP C G WD N F NA T PS PQ U U Y GT K Y PT HP HH RGASC PD K PAP C G F L NL HA T T PS PQ U U Y GT K Y PT HP HH RGASC PD K PAP C G F L NL HA T T PS PQ U U Y GT K Y PT HP HH RGASC PD K PAP C G F L NL HA T T PS PQ U U Y GD K Y PL Q PH HR A SC PD R PAP C G F L NL A T PS PN U L F GD K Y P S PH HR SS C PN L PAP C G F I M SS K E PN V H I L Y G D K Y PL R PH HR SS C PN L PAP C G F I M SS K E PN V H I L F GO K Y PL R PH HR SS C PN L PAP C G F I M SS K E PN V H I L F GO K Y PL R PH HR SS C PN L PAP C G T O Y M PL K E PN V H I L Y G GD Y PL R P Y G R I SM C PN HP AH C S PC I U A A MP N H I L F GO L Y PL R PY HR SS C PN L PAP C G S A U N M K E PS PH V L Y F GO D Y PL R PY G R I SM C PN HP AH C S PE V L A EMP N I H I L Y G SK Y PL R PH HR SS C P F L PAH C G PM L N L T E PN V H I L Y G SK Y PL R PY H R SS C P F L PAH C G PM L N L T E PN V H I L Y G SK Y PL R PY H R SS C P F L PAH C G PM L N L T E PN V H I L Y G SK Y PL Q PH HR A SS C P D R PAP C G SM D G S H O P Q U V Y G NN P PA Q P HH R A SS C P D R PAP C G SM D G S H O P Q U V Y G NN P PA Q P HH A A SS C P D R PAP C G SM D G E SK G N N P 1 I L Y G NN P PA Q P HH A A SS C P D R PAP C G SW D F E D R A SN P Q U L Y G NN P PA Q P HH A A SS C P D R PAP C G WS E F D R A PN Q I L Y G NN P PA Q P HH A A SS C P D R PA C G WS E F D R A PN Q I L Y G NN P PA Q P HH A A SS C P D R PA C G WS E F D R A PN Q I L Y G NN P R Q P HH A A SS C P D R PA C G WS E F D R A PN Q I L Y Y G NN P K Q P HH A A SS C P D R PA C G WS E F D M A PN P I I L Y Y G N N P K Q P HH A A SS C P D R PA A C G WS E F D M A PN P Q I L Y Y G N N P K Q P HH A A SS C P D R PA C G WS E F D M A PN P Q I L Y Y G N N P K Q P H H A A SS C P D R PA C G W S	000000000000000000000000000000000000000		P         P	372 373 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
μP         013066221.1           μP         01306628.1           μP         01306627.1           μP         01306623.1           μP         01306623.1           μP         01306623.1           μP         01306623.1           μP         01306428.1           μP         01308428.1           μP         0150847.1           μP         0151328.1           μP         0151328.2           μP         01551328.2           μP         01551328.2           μP	TA GIN AF LA LLAGE ING IDA - YI KNWA VE GUN FL EA ANNAFIAL LVA AE SGIESK - YI KNWA VE GUN YI EA ANNAFIAL LA AE SGUESK - YI KNWA VE GUN YI EA ANNAFIAL LA AE SGUESK - YI KNWA VE GUN YI EA ANNAFIAL ALAAE SGUESK - YI KNWA VE GUN YI YA ANSAF VALLSAE AG IN PG - YI KNWA VE GUN YI YA ANSAF VALLSAE AG IN PG - YI KNWA VE GUN YI YA ANSAF VALLSAE AG IN PG - YI KNWA VE GUN YI YA ANSAF VALLSAE AG IN PG - YI KNWA VE GUN YI YA ANSAF VALLSAE AG IN PG - YI KNWA VE GUN YI YA ANSAF VALLSAE AG IN PG - YI KNWA VE GUN YI YS SIMAMAAL VAA DAGLN AG - YI KNWA VE GUN YI YS SIMAMAAL VAA DAGLN AG - YI KNWA NC GUN YA YS SIMAMAAL VAA DAGLH SE - YI KNWA NC GUN YA YS SIMAMAAL VAA DAGLH SE - YI KNWA NC GUN YA AL YS SIMAMAAL VAA DAGLH SE - YI KNWA NC GUN YA AL YA SIMAFAAL LAA DAGLH TE - YI KNWA NC GUN YA HAANNAAL VAA DAGLH AE - YI KNWA NC GUN YA HAANNAAL ALAADAGLH TE - YI KNWA NC GUN YA HAANIA ALI AA DAGLH TE - YI KNWA NC GUN YA HAANIA ALI AA DAGLH TE - YI KNWA NC GUN YA HAANIA ALI AA DAGLH TE - YI KNWA NC GUN YA HAANIA ALI AA DAGLH TE - YI KNWA NC GUN YA HAANIA ALI YA LU AA DAGLH TE - YI KNWA NC GUN YA HAANIA ALI YA LU AA DAGLH TE - YI KNWA NC GUN YA HAANIA ALI YA LU AA DAGLH TE - YI KNWA NC GUN YA HAANIA FU CU GAADG SUN SI SI SI YI KI YI KI YA KU HAANIA YI YA LU TA EGGUN SI YI YI KI YA KU HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI GY FA FA GUN YI YI HAANIYA FU CU GAADG - DOG - YI RE FA GUN YI YI HAANIYA FU CU GAADG - D	0         N         H         S         Y         V           0         N         H         S         Y         V         G         N         H         G         S         Y         V         G         N         H         G         S         Y         V         G         N         H         G         S         Y         V         G         N         H         G         N         H         G         N         H         S         G         N         H         G         N         H         G         N         H         G         D         G         F         S         I         G         N         H         G         D         G         F         G         D         G         F         G         D         G         F         G         D         G         F         G         D         G         F         G         D         G         F         G         D         G         F         G         D         G         F         G         G         G         G         G         G         G         G         G         G         G         G         G	Y GSK Y PLOP PH RAGASC PD R PACC GWAQYS S PS PN PQ V L M Y GSK Y PLOP PH RAGSC PD R PAPC D WNN F NS A PN PQ V L V Y GTK Y PL HP PH RGSS C PD R PAPC D WNN F NS A PN PQ V L V VGNK Y PLOP HH RAASC PD K PAPC GWD F HA T T PS PQ V L Y GTK Y PT HP PH RGASC PD K PAPC GWD F HA T T PS PQ V L Y GTK Y PT HP PH RGASC PD K PAPC G F L NA T T PS PQ V L Y GTK Y PT L PH PH RGASC PD K PAPC G F L NA T T PS PQ V L Y GTK Y PT L PH HR RGASC PD K PAPC G F L NA T T PS PQ V L Y GTK Y PT L PH HR RGASC PD K PAPC G F L NA T T PS PQ V L Y GTK Y PT L PH HR RGASC PD K PAPC G F L NS K PP V H L Y G T G X Y PT S PH HR ASSC PN L PAPC G F L NS K E PN V H L Y F G DK Y PL S PH HR ASSC PN L PAPC G F L NS K E PN V H L Y F G DK Y PL R PH HR S SC PN L PAPC G F L NS K E PN V H L Y F G DK Y PL R PH HR S SC PN L PAPC G F L NS K E PN V H L Y F G D Y PL R PH HR S SC PN L PAPC G F L NS K E PN V H L Y F G D Y PL R PH HR S SC PN L PAPC G F L NS K E PN V H L Y F G D Y PL R PH HR S SC PN L PAPC G F L NS K E PN V H L Y F G D Y PL R P HR S SC PN P PA K C S Q I M P M K E PS PH V L Y F G D Y PL R PY Y R NS MC PN H PA H C G PD L Q A D M PN H I L Y F G D Y PL R PY HR S SC P PL PA H C G P M L N L T E PN V H T L Y F G D Y PL R PY HR S SC P F L PA H C G P M L N L T E PN V H T L Y F G D Y PL R PY HR S SC P F L PA H C G P M L N L T E PN V H T L Y F G D Y PL R P Y HR S SC P F L PA H C G P M L N L T E PN V H T L Y F G D Y PL R P H HR A SC P D R PA PC S M D Q N A PN P Q V L T Y GNN P P Q P H HA A SC P D R PA PC S M D Q N A PN P Q V L T Y GNN P P Q P H HA A SC P D R PA PC G W E F D R A SN P Q I L Y WG N N P R Q P H HA A SC P D R PA PC G W E F D R A SN P Q I L Y WG N N P R Q P H HA A SC P D R PA PC G W E F D R A SN P Q I L Y WG N N P R Q P H HA A SC P D R PA C G W E F D R A SN P Q I L Y WG N N P R Q P H HA A SC P D R PA C G W E F D K D A PN P Q I L Y WG N N P R Q P H HA A SC P D R PA C G W E F D K D A PN P Q I L Y WG N N P R Q P H HA A SC P D R PA C G W E F D K D A PN P Q I L Y WG N N P R Q P H HA A SS C P D R PA C G W E F D K D A PN P Q I L Y WG N N P R Q P H HA A SS C P D R PA	000000000000000000000000000000000000000		P         P         P           P         P         N         V         D           P         P         N         D         V         D         K         O           P         P         D         D         S         V         D         K         O         K         N         D         K         D         C         D         C         D	372 373 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
μP         01306628.1           μP         0130628.1           μP         0130628.1           μP         0130628.1           μP         0130628.1           μP         0130628.1           μP         0130628.1           μP         01306628.1           μP         01306628.1           μP         01306628.1           μP         01306628.1           μP         01306652.1           μP         01306852.1           μP         01306828.1           μP         01306828.1           μP         01308428.1           μP         01308428.1           μP         01408428.1           μP         01408428.1           μP         0140428.1           μP         0140428.1           μP         0140428.1           μP         0140428.1           μP         0140428.1           μP         014	TA GINA F LA LLAGE ING I DA - Y RKWA VE G I NY I L EA ANNA F I AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNA F I AL LA AE SG I E SK - Y RKWA VE G I NY I L EA ANNA F I AL LA AE SG I O TR - L RKWA VE G I NY I L EA ANNA F LA LLA AAE SG VD SR - L RKWA VE G I NY I L Y AG NS AF LA LLA AAE SG I P DR - L RKWA VE G I NY I L Y AG NS AF LA LLA AAE SG I P DR - L RKWA VE G I NY I L Y SS NIAMAAL VA ADAG L NG - Y RKWA VE G I NY I L Y SS NIAMAAL VA ADAG L NG - Y RKWA VE G I NY I L Y SS NIAMAAL VA ADAG L NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG L NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG L NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG L NG - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG L N SE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG L H SE - Y RHWA MCG I HY AL Y SS NIAMAAL VA ADAG L H SE - Y RHWA MCG I HY AL A SN NAFA AL L AAD T GL H PE - Y RHWA MCG I HY AL A SN NAFAL L AAD GL H HE - F RHWA MCG I HY AL A ANNAMLAL I AADAG L H TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG L H TE - Y RHWA MCG I HY AL AA ANNAMLAL I AADAG L H TE - Y RHWA MCG I HY AL HAAN I AL VA LVA AD NG I H TE - Y RHWA MCG I HY AL HAAN I AL VA LVA AD NG I H TE - Y RHWA MCG I HY AL HAAN I AL VA LVA AD NG I H TE - Y RHWA MCG I HY AL HAAN I AL VA LVA AD G I S AS - Y RE FA RQ I NY IL HAAN Y F VC L Q AAD G - D GD E - Y RG FA TD G I SY ML HAAN Y F VC L Q AAD G - D GD E - Y RG FA TD G I SY ML HAAN Y F VC L Q AAD G - D GD E - Y RG FA TD G I SY ML HAAN Y F VC L Q AAD G - D GD E - Y RG FA TD G I SY ML HAAN Y F VC L E AADG D GE - Y RG FA TD G I SY ML HAAN Y F VC L E AADG D GE - Y RG FA TD G I SY ML HAAN Y F VC L E AADG D GE - Y RG FA TD G I SY ML HAAN Y F VC L E AADG D GE - Y RG FA TD G I SY ML HAAN Y F VC L E AADG D GE Y RE FF A EQ G I NY ML HAAN Y F VC L E AADG D GE Y NG FF A EQ G I NY ML HAAN Y F VC L E AADG D GE Y NG FF A EQ G I NY ML HAAN Y F VC L E AADG D GE Y NG FF A EQ G I NY ML HAAN Y F VC L E AADG D GE Y NG FF A EQ G I NY ML HAAN Y F VC L E AADG D GE Y NG FF A EQ G I NY ML HAAN Y F VC L E AADG D GE Y NG FF A EQ G I NY	0         N         H         N	Y GSK Y PLOP HH AGASC PD R PACC GWAQYS S PS PN PQ U H Y GSK Y PLOP HH RGSS C PD R PAPC DWN F NS A PN PQ U L Y GT K Y PL HP HH RGSS C PD R PAPC DWN F NS A PN PQ U L Y GT K Y PL HP HH RGSS C PD R PAPC GWN F NS T APN PQ U L Y GT K Y PT HP HH RGASC PD K PAPC GWN F NS PL PN PH U L Y GT K Y PT HP HH RGASC PD K PAPC GWN F NS PL PN PH U L Y GT K Y PT HP HH RGASC PD K PAPC GWN F NS PL PN PH U L Y GT K Y PT HP HH RGASC PD K PAPC GG F L N HA T T PS PQ V L Y GT K Y PT HP HH RGASC PD R PAPC GG F L N HA T T PS PQ V L Y GT K Y PT HP HH RGASC PD R PAPC G F L N SK F PN V H L Y GD K Y PL SP HH RASSC PD R PAPC G F I MS SK F PN V H L F GD K Y PL SP HH RASSC PN L PAPC G F I MS SK F PN V H I L Y GD K Y PL RP HH RS SC PN L PAPC G F I MS SK F PN V H I L Y GD K Y PL RP HH RS SC PN L PAPC G F I M SK K PN V H I L Y GD K Y PL RP HH RS SC PN L PAPC G F I M SK K PN V H I L Y GD K Y PL RP HH RS SC PN L PAPC G F I M SK K PN V H I L Y GD K Y PL RP HH RS SC PN L PAPC G F I M SK K PN V H I L Y GD K Y PL RP HH RS SC PN L PAPC G F I M SK K PN V H I L Y GD Y PL R PY H R M SK C PN H PAHC S P V L L A EM PN I H I L Y GG K Y PL R PH HR SS C PN L PAPC G SA DA Q B M N H I L Y Y GD K Y PL R P H H R SS C P F L PAHC G PD M L N L T PN V H I L Y GD Y PL R PY H R M SK C PN H PA I C T PE H L A AK L PN V H I L Y GG L Y PL R P H H R SS C P F L PAHC G PM L N L T E PN V H I L Y GS K Y PL Q PH HA ASS C PD R PAPC G SA DA Q B H PA PQ I L Y GN N P AQ P H AASS C PD R PAPC G SM D Q L N AP OP S PQ V L Y GN N P AQ P H H AASS C PD R PAPC G SW E F D R AP N PQ I L Y GN N P PL Q P H H AASS C PD L PAPC G W E F D R ASN PQ I L Y Y GN N P F Q P H H AASS C PD R PAPC G W E F D R ASN PQ I L Y Y GN N P F Q P H H AASS C PD L PAPC G W PE F D R ASN PQ I L Y Y GN N P F Q P H H AASS C PD L PAPC G W PE F D R ASN PQ I L Y Y GN N P F Q P H H AASS C PD L PAPC G W PE F D R ASN PQ I L Y Y GN N P F Q P H H AASS C PD L PAPC G W PE F D R ASN PQ I L Y Y GN N P F Q P H H AASS C PD L PAPC G W PE F D R ASN PQ I L Y Y GN N P F Q P H H AASS C PD L PAPC G W PE F D R ASN P Q I L Y Y GN N P F Q P H H AASS C	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		P D Q ND NY VD LR           P D E ND I Y D D KR           P D E ND I Y E D KR           P L VD D S Y E D KR           P L VD D S Y E D KR           P E V D D S Y D D KR           P E A D D S Y A D NR           P E A D D S Y A D NR           P E A ND N Y D D R           P E G Q S D D Y T D D R           P G Q S ND D Y T D D R           P G Q S ND D Y T D D R           P G Q S ND D Y T D D R           P G Q S ND D Y T D D R           P G Q S ND D Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y T D D R           P G Q ND D Y V D F F           P G A ND E Y L D K F           P G A ND E Y L D K F           P G A D N E Y D D R           P G A D A F H D H R           P D E A S D I Y E D L R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D E A D K H D H R           P D Q N N Y A D N R	372 373 372 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
μP_013066281,1           μP_013066282,1           μP_013066282,1           μP_013066282,1           μP_013066282,1           μP_013066282,1           μP_013064281,1           μP_0151642631,1           μP_015126067,1           μP_015502607,1           μP_015502607,1           μP_015502607,1           μP_0150660,1           μP_01801660,1           μP_01801680,1           μP_01801680,1           μP_01801680,1           μP_01801680,1           μP_01801680,1           μP_01801680,1	TA GWAF LA LLAGE MG I DA - Y RKWA VE G I NY I L EA ANNAF I AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNAF I AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNAF I AL LA AE SG V DSR - LRKWA VE G I NY I L EA ANNAF LA LLA AE SG V DSR - LRKWA VE G I NF I L YAGNSAF LA LLAAE SG I PR - L RKWA VE G I NF I Y ASNSAF LA LLAAE SG I PR - L RKWA VE G I NY I L Y SS NMAMAAL VAA DAGL NAG - Y RKWA VE G MY I Y SS NMAMAAL VAA DAGL NAG - Y RKWA VE G MY I Y SS NMAMAAL VAA DAGL NAG - Y RHWA MC G I HY A L Y SS NMAMAAL VAA DAGL NAG - Y RHWA MC G I HY A L Y SS NMAMAAL VAA DAGL NAG - Y RHWA MC G I HY A L Y SS NMAMAAL VAA DAGL H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VAA DAGL H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VAA DAGL H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VAA DAGL H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VAA DAGL H TE - Y RHWA MC G I HY A L A ANNAM LA LI AA DAGL H TE - Y RHWA MC G I HY A L AA ANAM LA LI AA DAGL H TE - Y RHWA MC G I HY A L HAAN I AL VAA VAA DAGL H TE - Y RHWA MC G I HY A L AANNAM LA LI AA DAGL H TE - Y RHWA MC G I HY A L AANNAM LA LI AA DAGL H TE - Y RHWA MC G I HY A L HAAN I AL VAL VAA DNG I H TE - Y RHWA MC G I HY A L HAAN I AL VAL VAA DNG I H TE - Y RHWA MC G I HY A L AANNAM LA LI AA DAGL H TE - Y RHWA MC G I HY A L HAAN I AL I AA DAG C - D GU F Y SF A K Q I NY I L HAAN Y SF VC LQ AAD - G D GE - Y RE F A R Q I NY I L HAAN Y SF VC LQ AAD - G D GE - Y RE F A R Q I HY I L HAAN Y SF VC LQ AAD - G D GE - Y RE F A R Q I HY I L HAAN Y SF VC LQ AAD - G D GE - Y RE F A R Q I HY I L HAAN Y SF VC LG AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LG AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LG AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LE AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LE AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LE AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LE AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LE AAD G - D GE - Y RE F A Q C I HY I L HAAN Y SF VC LE AAD G - D GE - Y RE F A Q C I HY M HAAN Y SF VC LE AAD G - D GE - Y RE F A R Q U HY M HAAN Y SF VC LE AAD	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N H H G S Y Q V G N H H G S Y Q V G D T G F S Y V V G S T G F S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G S S G R S Y V V G N P S R S Y V V G N P S R S Y V V G N P S R S Y V V G G S G G R S Y V V G G S G G R S Y V V G N P S R S Y V V G G S G G R S Y V V G N P S R S Y V V G G S G G R S Y V V G G S G G R S Y V V G N P S R S Y V V G G G G G R S Y V V G N P S R S Y V V G G S G G R S Y V V G G S G G R S Y V V G N P S R S Y V V G G S G G R S Y V V G N P S R S Y V V G G S G G R S Y V V G G S G G R S Y V V G N P S R S Y V V G G S G R S Y V V G G S G R S Y V V	Y GSK Y PLOP HH AGASC PD D RASC GWAQYS S PS PN PQ V L Y Y GSK Y PLOP HH RASSC PD N PACC NWD T YN LHSPN PH V L N Y GT K Y PL HP HH RGSS C PD N PAP C D WNN F NS A PN PQ V L V Y GT K Y PL HP HH RGASC PD K PAP C GWD F HA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAP C GWD F HA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAP C G WD F HA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAP C G F E N L HA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAP C G F E N L HA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAP C G F E N L HA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAP C G F E N L HA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD R PAP C G F E N S K E PN V H I L Y G T G K Y P HS PH HRASSC PN L PAP C G F F I MS S K E PN V H I L Y F G D K Y P L P PH HR S SC PN L PAP C G F I M S K E PN V H I L Y F G D K Y P L R PH HR S SC PN L PAP C G F I M S K E PN V H I L Y F G D K Y P L R PH HR S SC PN L PAP C G T O Y M P L K E PN V H I L Y F G D Y P L R P Y G R I S M C PN HP A H C G P L L A E M PN I H I L Y F G D Y P L R P Y G R I S M C PN HP A H C G P M L N L T E PN V H I L Y F G D Y P L R P Y G R I S M C PN HP A H C G P M L N L T E PN V H I L Y F G D Y P L R P Y G R I S M C PN HP A H C G P M L N L T E PN V H I L Y F G D Y P L R P Y H R S SC P F L P A H C G P M L N L T E PN V H I L Y F G D Y P L R P Y H R S SC P F L P A H C G P M L N L T E PN V H T Y Y G S K Y P L Q P H HA A SC P D R P P C G S M D G E S M O D P Q V L Y W G N P P A Q P HH A A SS C P D R P P C G S M D G E S M O D P Q V L Y Y G N N P A Q P HH A A SS C P D R P P C G S M D G E S M O D P Q V L Y Y G N N P P C Q P H H A SS C P D R P A C G S W D F F D R A S M P Q U L Y W G N P P K Q P HH A A SS C P D R P A C G W S E F D K D A P N Q I L Y W G N N P K Q P HH A A SS C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P K Q P H H A A SS C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P F Q P H H A A SS C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P F Q P H H A A SS C P D R P A C G W S E F D K D A P N Q I L Y Y G N N P F Q P H H A A SS C P D R P A C G W S E F D	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		P         P           P         P	372 373 372 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
NP_013066221.1           NP_013066281.1           NP_013064281.1           NP_015136427.1           NP_01513282.01.1           NP_01513282.01.1           NP_017562827.1           NP_017562827.1           NP_01766281.1           NP_01801680.1           NP_01801680.1           NP_01801680.1           NP_01801680.1           NP_01801680.1           NP_01801680.1 <tr< td=""><td>TA GIVA F LA LLA GE BIG I DA - Y RKWA VE G I NY I L EA ANNA F I AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNA F I AL LA AE SG V DSR - LRKWA VE G I NY I L EA ANNA F I AL LA AE SG V DSR - LRKWA VE G I NY I L Y AGNS AF LA LLA AE SG V DSR - LRKWA VE G I NY I L Y AGNS AF LA LLA AE SG I P R - LRKWA VE G I NY I L Y AS NAMAAL VA AD AG L N G - Y RKWA VE G I NY I L Y SS NMAMAAL VA AD AG L N G - Y RKWA VE G I NY I L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H SE - Y RHWA MC G I HY A L A ANNA AL L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN MAMLA L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN MAMLA L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN MAMLA L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN I AL U AL VA A DNG I HT E - Y RHWA MC G I HY A L HA AN I AL U AL VA A DNG I HT E - Y RHWA MC G I HY A L HA AN Y F V C L G AA D G - P G DE - Y NG F S A C O I NY I L Y AG NA T AL Y AL LT A EAG I S AS - Y RG F A R Q I HY M L HA AN Y F V C L G AA D G - D G D E - Y NG F S A K O I NY I L HA AN Y F V C L G AA D G - D G D E - Y NG F S A K O I NY I L HA AN Y F V C L G AA D G - D G D E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D R E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D R E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A G O I NY M L HA AN Y S F I C L E AA D - O D Q Y NG F S A G O I NY M L HA AN Y S F I C L E AA D - G D Q - Y NG F S A G O I NY M L HA AN Y S F I C L E AA D G - N O E - Y NG F S A G O I NY M L HA AN Y S Y C L E AA D G - N O E - Y NG F S A G O</td><td>0         N         HIGS S         Y         V         O         O         N         HIGS S         Q         V         O         O         N         HIGS S         Q         V         O         <t< td=""><td>Y GSK Y PLOP HH AGASC PD D RASC GWAQYS S PS PN PQ V L M Y GSK Y PLOP HH RASSC PD R PAPC D WNN F NS A APN PQ V L V Y GT K Y PL HP HH RGSS C PD R PAPC D WNN F NS A APN PQ V L V Y GT K Y PL HP HH RGASC PD K PAPC G WON F NS T APN PQ V L V Y GT K Y PT HP HH RGASC PD K PAPC G WON F NS T APN PQ V L V Y GT K Y PT HP HH RGASC PD K PAPC G WON F NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G G A L NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T PS PY V L I Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T PS PY V L I Y GT K Y PT HP HH RGASC PD K PAPC G F L NS K E PN Y H L Y F GD K Y PL SP HH RASSC PN L PAPC G F I MSS K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSS K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSK K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSK K E PN Y H I L Y F GD K Y PL R PH HRS SC PN PA K C SQ I M P K E PS PH V L I F GD Y PL R PY HR SS C PN L PAPC G F G I M SK E PN Y H I L Y F GD K Y PL R PH HRASSC PN PA K C SQ I M P K E PS PH V L Y F GD Y PL R PY Y R SM K PN HP A I C T PE H L A A K L PN Y H I L Y F GD Y PL R PY HR SS C P N PA K C G P M L N L T E PN Y H I L Y F GD Y PL R PY HR SS C P N PA C G SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P N R PAPC G SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P N R PAPC G M SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P D R PAPC G M SE F D R A PN P Q U L Y Y GN N P P Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P P Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D K O PN P Q I L Y Y G N N P F Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y Y G N N P F Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y Y G N N P F C P H H A SSC P D R PAPC G W SE F D R C A D N P P Q I L Y Y G N</td><td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td></td><td>P         P           P         P           P         P           P         P           P         P           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         P</td><td>372 373 372 372 372 372 372 394 394 394 394 394 394 394 394 394 394</td></t<></td></tr<>	TA GIVA F LA LLA GE BIG I DA - Y RKWA VE G I NY I L EA ANNA F I AL VA AE SG I E SK - Y RKWA VE G I NY I L EA ANNA F I AL LA AE SG V DSR - LRKWA VE G I NY I L EA ANNA F I AL LA AE SG V DSR - LRKWA VE G I NY I L Y AGNS AF LA LLA AE SG V DSR - LRKWA VE G I NY I L Y AGNS AF LA LLA AE SG I P R - LRKWA VE G I NY I L Y AS NAMAAL VA AD AG L N G - Y RKWA VE G I NY I L Y SS NMAMAAL VA AD AG L N G - Y RKWA VE G I NY I L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L N G - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H SE - Y RHWA MC G I HY A L Y SS NMAMAAL VA AD AG L H SE - Y RHWA MC G I HY A L A ANNA AL L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN MAMLA L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN MAMLA L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN MAMLA L I AA DA G L H T E - Y RHWA MC G I HY A L HA AN I AL U AL VA A DNG I HT E - Y RHWA MC G I HY A L HA AN I AL U AL VA A DNG I HT E - Y RHWA MC G I HY A L HA AN Y F V C L G AA D G - P G DE - Y NG F S A C O I NY I L Y AG NA T AL Y AL LT A EAG I S AS - Y RG F A R Q I HY M L HA AN Y F V C L G AA D G - D G D E - Y NG F S A K O I NY I L HA AN Y F V C L G AA D G - D G D E - Y NG F S A K O I NY I L HA AN Y F V C L G AA D G - D G D E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D R E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D R E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A K O I NY M L HA AN Y S F I C L E AA D - G D Q E - Y NG F S A G O I NY M L HA AN Y S F I C L E AA D - O D Q Y NG F S A G O I NY M L HA AN Y S F I C L E AA D - G D Q - Y NG F S A G O I NY M L HA AN Y S F I C L E AA D G - N O E - Y NG F S A G O I NY M L HA AN Y S Y C L E AA D G - N O E - Y NG F S A G O	0         N         HIGS S         Y         V         O         O         N         HIGS S         Q         V         O         O         N         HIGS S         Q         V         O <t< td=""><td>Y GSK Y PLOP HH AGASC PD D RASC GWAQYS S PS PN PQ V L M Y GSK Y PLOP HH RASSC PD R PAPC D WNN F NS A APN PQ V L V Y GT K Y PL HP HH RGSS C PD R PAPC D WNN F NS A APN PQ V L V Y GT K Y PL HP HH RGASC PD K PAPC G WON F NS T APN PQ V L V Y GT K Y PT HP HH RGASC PD K PAPC G WON F NS T APN PQ V L V Y GT K Y PT HP HH RGASC PD K PAPC G WON F NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G G A L NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T PS PY V L I Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T PS PY V L I Y GT K Y PT HP HH RGASC PD K PAPC G F L NS K E PN Y H L Y F GD K Y PL SP HH RASSC PN L PAPC G F I MSS K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSS K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSK K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSK K E PN Y H I L Y F GD K Y PL R PH HRS SC PN PA K C SQ I M P K E PS PH V L I F GD Y PL R PY HR SS C PN L PAPC G F G I M SK E PN Y H I L Y F GD K Y PL R PH HRASSC PN PA K C SQ I M P K E PS PH V L Y F GD Y PL R PY Y R SM K PN HP A I C T PE H L A A K L PN Y H I L Y F GD Y PL R PY HR SS C P N PA K C G P M L N L T E PN Y H I L Y F GD Y PL R PY HR SS C P N PA C G SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P N R PAPC G SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P N R PAPC G M SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P D R PAPC G M SE F D R A PN P Q U L Y Y GN N P P Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P P Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D K O PN P Q I L Y Y G N N P F Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y Y G N N P F Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y Y G N N P F C P H H A SSC P D R PAPC G W SE F D R C A D N P P Q I L Y Y G N</td><td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td><td></td><td>P         P           P         P           P         P           P         P           P         P           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         P</td><td>372 373 372 372 372 372 372 394 394 394 394 394 394 394 394 394 394</td></t<>	Y GSK Y PLOP HH AGASC PD D RASC GWAQYS S PS PN PQ V L M Y GSK Y PLOP HH RASSC PD R PAPC D WNN F NS A APN PQ V L V Y GT K Y PL HP HH RGSS C PD R PAPC D WNN F NS A APN PQ V L V Y GT K Y PL HP HH RGASC PD K PAPC G WON F NS T APN PQ V L V Y GT K Y PT HP HH RGASC PD K PAPC G WON F NS T APN PQ V L V Y GT K Y PT HP HH RGASC PD K PAPC G WON F NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G G A L NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T T PS PQ V L Y Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T PS PY V L I Y GT K Y PT HP HH RGASC PD K PAPC G F L NA T PS PY V L I Y GT K Y PT HP HH RGASC PD K PAPC G F L NS K E PN Y H L Y F GD K Y PL SP HH RASSC PN L PAPC G F I MSS K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSS K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSK K E PN Y H I L Y F GD K Y PL R PH HRS SC PN L PAPC G F G I MSK K E PN Y H I L Y F GD K Y PL R PH HRS SC PN PA K C SQ I M P K E PS PH V L I F GD Y PL R PY HR SS C PN L PAPC G F G I M SK E PN Y H I L Y F GD K Y PL R PH HRASSC PN PA K C SQ I M P K E PS PH V L Y F GD Y PL R PY Y R SM K PN HP A I C T PE H L A A K L PN Y H I L Y F GD Y PL R PY HR SS C P N PA K C G P M L N L T E PN Y H I L Y F GD Y PL R PY HR SS C P N PA C G SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P N R PAPC G SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P N R PAPC G M SA DA Q S H G PN Q V L Y G SK Y PL Q P HH RA SC P D R PAPC G M SE F D R A PN P Q U L Y Y GN N P P Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P P Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D K O PN P Q I L Y Y G N N P F Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y WG N N P K Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y Y G N N P F Q P H HA ASSC P D R PAPC G W SE F D R C A PN P Q I L Y Y G N N P F C P H H A SSC P D R PAPC G W SE F D R C A D N P P Q I L Y Y G N	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		P         P           P         P           P         P           P         P           P         P           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         V           P         P	372 373 372 372 372 372 372 394 394 394 394 394 394 394 394 394 394
NP_01306628.1           NP_01306429.1           NP_01306428.1           NP_0130428.1           NP_0130428.1           NP_0130428.1           NP_0130428.1           NP_0130428.1           NP_0130428.1           NP_0150428.1           NP_0150428.1           NP_015128067.1           NP_01562827.1           NP_0150606.1           NP_01801680.1           NP_01801680.1           NP_01801680.1	TAGMYAFLALLAGEMGIDAR - YRXWAVEG NYIL         EAANNAFIALVAAESGIESK - YRXWAVEG NYIL         EAANNAFIALVAAESGIESK - YRXWAVEG NYIL         EAANNAFIALVAAESGITEK - YRXWAVEG NYIL         EAANNAFIALALAAESGVD SR - LRXWAVEG NYIL         EAANNAFIALALAAESGVD SR - LRXWAVEG NYIL         YAGNSAFLALLAAAESGVD SR - LRXWAVEG NYIL         YAGNSAFLALLAAAESGVD SR - LRXWAVEG NYIL         YAGNSAFLALLAAAESGVD SR - LRXWAVEG NYIL         YASNMAMALVAADAGLN GO - YRHWAMCG NYA         YSSNMAMAALVAADAGLN GO - YRHWAMCG NYA         YSSNMAMAALVAADAGLN GO - YRHWAMCG NYA         YSSNMAMAALVAADAGLH SE - YRHWAMCG NYA         YSSNMAMAALVAADAGLH SE - YRHWAMCG NYA         YSSNMAMAALVAADAGLH TE - YRHWAMCG NYA         YAANNAFALLAADAGLH TE - YRHWAMCG NYA         YAANNAFALLAADAGLH TE - YRHWAMCG NYA         HAANNAHLALIAADAGLH TE - YRHWAMCG NYA         HAANNAHLALIAADAGLH TE - YRHWAMCG NYA         HAANNAHLALIAADAGLH TE - YRHWAMCG NYA         HAANNAFUCLQAADG - DOE - YRGYATAG NYA         HAANNAFUCLQAADG - DOE - YRGYATAG NYA         HAANNAFUCLQAADG - DOE - YRGYATAG NYA         HAANNAFUCLQAADG - DOE - YRGYATADG NYA         HAANNAFUCLAADAG	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N H H G S Y Q V G N H H G S Y Q V G D T G F S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G G G G G R S Y V V G S T G G S F S V V G S T G G S F S V V G S T G G S F S V V G S S G G S S Y V V G S S G S S Y V S S S S S Y V V G S S G S S S Y V V S S S S S S S S Y V V S S S S S S S S S S S S S S S S S S	Y GSK Y PLOP HH AGASC PD D RASC GWAQ Y S P S P N P Q V L Y Y GSK Y PL PH HH RGSS C P D R PA P C D WN N F NS A P N P Q V L Y Y GT K Y P L H P HH RGSS C P D R PA P C D WN N F NS A P N P Q V L Y Y GT K Y P L H P HH RGSS C P D R PA P C G WD N F N S P L P N P V L Q Y GT K Y P T H P HH RGASC P D K PA P C G WD N F N S P L P N P Y L Q Y GT K Y P T H P HH RGASC P D K PA P C G WD F N S P L P N P Y L Q Y GT K Y P T H P HH RGASC P D K PA P C G WD F N A T T P S P Q V L Y Y GT K Y P T H P HH RGASC P D K PA P C G G E N L H A T T P S P Q V L Y Y GT K Y P T H P HH RGASC P D R PA P C G F E N L H A T T P S P Q V L Y Y GT K Y P T H P HH RGASC P D R PA P C G F I M S K E P N V H I L Y G D K Y P L P P HH R S SC P N L PA P C G F I M S K E P N V H I L Y F G D K Y P L P P HH R S SC P N L PA P C G F I M S K E P N V H I L Y F G D K Y P L P P HH R S SC P N L PA P C G T O Y M P L K E P N V H I L Y F G D K Y P L P P HH R S SC P N L PA P C G T O Y M P L K E P N V H I L Y F G D Y P L R P H H R S SC P N L PA P C G T O Y M P L K E P N V H I L Y F G D Y P L R P Y H R M S C P N P A K C G D I M M K E P S P V V I L Y F G D Y P L R P Y H R M S C P N P A L C T P E L A A K L P N V H I L Y F G D Y P L R P Y H R M S C P N P A L C T P E L L A K L P N V H I L Y F G D Y P L R P Y H R M S C P N P A L C T P E L L A K L P N V H I L Y F G D Y P L R P Y H R M S C P N F A P C G S M D L N A P O S N O V L Y W G N P P A O P H H A A S C P D R P A P C G S M D A D A P O S N O V L Y W G N P P A O P H H A A S C P D R P A P C G S W D F A A P O N V U L Y G G N N P T O P H H A A S S C P D L P A C G WP E F D R G P P N O I L Y W G I N P P A O P H A A S S C P D L P A C G W P E F D R G P N P O I L Y W G N N P T A P H A A S S C P D L P A C G W P E F D R G P N P O I L Y W G N N P T A P H A A S S C P D L P A C G W P E F D R G P N P O I L Y W G N N P T A P H H A S S C P D L P A C G W P E F D R G P N P O I L Y W G N N P T A P H H A A S S C P D L P A C G W P E F D R G P N P O I L Y W G N N P T A P H H R S S C P D L P P C G W P E F D R G P N P O I L Y Y G N N P T A P H H R S S C P N P	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		P         P           P         P	372 373 372 372 372 372 372 372 374 394 394 394 394 394 394 394 394 394 39
NP_013066281.1           NP_013066281.1           NP_013066281.1           NP_013066281.1           NP_013066281.1           NP_013066281.1           NP_013066371.1           NP_013066371.1           NP_013066371.1           NP_013066371.1           NP_013066371.1           NP_01306632.1           NP_013066282.1           NP_01306628.1           NP_01306628.1           NP_01306628.1           NP_01306428.1           NP_01304428.1           NP_01308428.1           NP_015126067.1           NP_015126067.1           NP_015126067.1           NP_015528267.1           NP_01558267.11           NP_0150680.1           NP_0180680.1           NP_01801680.1           NP_01801680.1           NP_01801680.1           NP_01801680.1           NP_01801680.1           NP_01801	TAGMYAF LALLAGE MG I DA - Y RKWA VE G I NY I L EAANNAF I AL VAAESG I ESK - Y RKWA VE G I NY I L EAANNAF I AL VAAESG I ESK - Y RKWA VE G I NY I L EAANNAF LALLAAESG I OTR - LRKWA VE G I NY I L EAANNAF LALLAAESG VD SR - LRKWA VE G I NY I L YAGNSAF LALLAAESG I PTR - LRKWA VE G I NY I L YAANSAF VALLSAEGG I PTR - LRKWA VE G I NY I L YS SMAMAAL VAADAGL NAG - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL NAG - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL NAG - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL NAG - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL NAG - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL HSE - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL HSE - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL HSE - Y RHWA MC G I HY A L YS SMAMAAL VAADAGL HSE - Y RHWA MC G I HY A L Y S SMAMAAL VAADAGL H SE - Y RHWA MC G I HY A L A SA MAMAL AL I AADAGL H TE - Y RHWA MC G I HY A L AANNAM LA LI AADAGL H TE - Y RHWA MC G I HY A L HAANNAH LA LI AADAGL H TE - Y RHWA MC G I HY A L AANNAM LA LI AADAGL H TE - Y RHWA MC G I HY A L AANNAH LA LI AADAGL H TE - Y RHWA MC G I HY A L HAANNAF VC LQ AAD G I HTE - Y RHWA MC G I HY A L HAANNAF VC LQ AAD G I S AS - Y NG FA TD G I Y M L HAANNAF VC LQ AAD - O D G P - Y NG FA TD G I Y M L HAANNAF VC LQ AAD - O D G P - Y NG FA TD G I Y M L HAANNAF VC LQ AAD - O D G P - Y RF FA QG I HY M L HAANNAF VC LQ AAD - O D G P - Y RF FA QG I HY M L HAANNAF VC LQ AAD - O D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD - O D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD - O D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD - O D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD - O D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C - Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D C C Y RE FA AGG I HY M L HAANNAF VC LQ AAD C - D	G N H H G S Y V V G N F H G S Y Q V G N F H G S Y Q V G N T H G S Y Q V G N H H G S Y Q V G N H H G S Y Q V G D T G F S Y V V G S T G F S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G G G G C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S S C R S Y V V G S S C R S Y V V G S S C R S Y V V G S S S C R S Y V V S S S S S V V S V V S S S S S V S Y V V S S S S	Y GSK Y PLOP HH AGASC PD R PACC GWAQ Y S P S P N PQ V L Y Y GSK Y PLOP HH R S SC PD R PAP C D WN N F NS A PN PQ V L Y Y GT K Y PL HP HH R GSS C PD R PAP C D WN N F NS A PN PQ V L Y Y GT K Y PL HP HH R GASC PD K PAP C G WD N F NS T A PN PQ V L Y Y GT K Y P L PP HH R GASC PD K PAP C G WD N F NS T A PN PQ V L Y Y GT K Y P L PP HH R GASC PD K PAP C G WD N F N S PL PN Y L L Y GT K Y PT HP HH R GASC PD K PAP C G F L N L H T T PS PQ V L Y Y GT K Y P L PP HH R GASC PD K PAP C G F L N L H T T PS PQ V L Y Y GT K Y P T PP HH R GASC PD K PAP C G F L N L H T T PS PQ V L Y Y GT K Y P T PP HH R GASC PD K PAP C G F L N S K E PN Y H L Y G T G K Y P S PH H R A SSC P N L PAP C G F F I M SS K E PN Y H I L Y F G D K Y P L PP HH R SSC C PN L PAP C G F I M SS K E PN Y H I L Y F G D K Y P L PP HH R SSC C PN L PAP C G F I M SS K E PN Y H I L Y F G D K Y P L PP HH R SSC C PN L PAP C G F I M SK K E PN Y H I L Y F G D K Y P L PP HH R SSC P N L PAP C G F I M SK K E PN Y H I L Y F G D Y P L R P H H R SSC P N L PAP C G F I M SK K E PN Y H I L Y F G D Y P L R P Y A R I SM C PN H PA H C S P L L A EMP N I H I L Y F G D Y P L R P Y A R I SM C PN H PA H C S P L L A EMP N I H I L Y F G D Y P L R P Y A R I SM C PN H PA H C S P L L A EMP N I H I L Y F G D Y P L R P Y A R I SM C PN H PA H C S P L D L A EMP N Y H I L Y F G D Y P L R P Y A R I SM C PN H PA H C S P M L N L T E PN Y H T L Y F G D Y P L R P Y A R I SM C PN H PA H C S P L D L A EMP N I I L Y Y G S K Y P L P H H A SS C P D R PA P C G SM D G E S M C N P P Q V L Y W G R D P Q Q P H H A A SS C P D R PA C G SM D G E S K G N N P Q U L Y Y G N N P R C P H H A A SS C P D R PA C G SM D G E S K G N N P Q U L Y W G N N P R Q P H H A A SS C P D R PA C G W S E F D R A P N Q I L Y W G N N P K Q P H H A A SS C P D R PA C G W S E F D K D A P N Q I L Y Y G N N P K Q P H H A A SS C P D R PA C G W S E F D K D A P N Q I L Y Y G N N P K Q P H H A A SS C P D R PA C G W S E F D K D A P N Q I L Y Y G N N P F C P H H R A SS C P D R PA C G W Y E T D R D P N I L Y Y G N N P K Q P H H A A SS C P D R P A C G W Y E T M	0 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	• • • • • • • • • • • • • • • • • • •	P         P           P         P	372 373 372 372 372 372 372 394 394 394 394 394 394 394 394 394 394

XP_018508194.1 XP_0145681961 XP_019884179.1 XP_0202892051 XP_020289297.1 XP_02028939.1 XP_02028939.1 XP_0214867.1 XP_020832851 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_021363495.1 XP_02196557.1 XP_02195657.1 XP_02195657.1 XP_02195657.1 XP_02219567.1 XP_02219567.1 XP_02219567.1 XP_02219567.1 XP_0222677.6 XP_0222677.6 XP_0222677.6 XP_0222677.6 XP_0222367.0 XP_0222367.6 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0222367.0 XP_0235.0 XP_0235.0 XP_0235.0 XP_0235.0 XP_02	HAAN I AF LCL I AS HAAN VAF LCL IAS HAAN VAF LCL VAA LAAN VAF LCL VAA LAAN AAF LCL IAI FAAN AAF LCL IAI YAAN AF LCL IAI YAAN AF LALVAJ YSAN TAF LALVAJ YSAN TAF LALVAJ YSAN TAF LALVAJ YSAN TAF LALVAJ YSAN TAF LALVAJ YAS NMAF MALLAJ YAS NMAF MALLAJ YAS NMAF MALLAJ YAS NMAF MALLAJ YAS NMAF MALLAJ YAS NMAF MALLAJ YAS NMAF LCQAJ HAAN VAF LCQAJ YAS NMAF MALLAJ YAS NMAF HALLAJ YAS NMAF HALLAJ	SDP SEKYTTFAKEQIDYI           SDP SEKYTTFAKEQIDYI           SDP SEKYTTFAKEQIDYI           SDP SEKYTTFAKEQIDYI           SDP DLKRYYFAKEQIDYI           ATD - DLKRYYFAKEQIDYI           SM NLKRYYFAKEQIDYI           A - SVEHDEPYAFAKEQIDYI           A - SVEHADEYAFAKEQIDYI           A - SVEHADEYAFAKEQINYA           A - SVEHADEYAFAKEQINYA           A - SVEHADEYAFAKEQINYA           A - SPAYAFAKASOIHYA           A - SPAYAFAKASOIHYA           A - SPAYAFAKASOIHYA           A - SPAYAFAKASOIHYA           A - SPAYAFAKAGOINYA           A - TAYRAFAKQOINYA	L G A T G G R S S Y Y Y Y G G R S S Y Y Y Y Y G G R S S Y Y Y Y Y G G R S D S S F Y Y G G R S S S S F Y Y G G R S S S S S F Y Y G G S S S S S S S S S S S S S S S S S S	Y GEN F TO OP Y HS ASS C P D K P Y GEN F TO OP Y HS ASS C P D K P Y GEN F TO OP Y HS ASS C P K L P Y GEN P TO P K HASS C P K S P Y GEN P P TO P HK ASS C P K S P Y GEN P P TO P HK K ASS C P K S P F G K S P O R P H H R S S C C P K P F G K S P O R P H H R S S C C P K P F G K S P O R P H H R S S C C P K P F G K Y P K V P H Y G S C C P S P F G K Y P K O P H H R S S C C P S P F G K Y P K O P H H R S S C C P S P F G K Y P K O P H H R S S C C P S P F G K Y P K O P H H R S S C C P S P F G V N P F T P H H R S S C C P S P F G V N P T T P H H R S S C C P S P F G V N P T T P H H R S S C C P S P F G V N P T T P H H R S S C C P S P F G V N P T T P H H R S S C C P S P F G V N P T P T P H H R S S C C P S P F G V N P T P T P H H R S S C P S P F G V N P T P T P H H R S S C P S P F G V N P T P T P H H R S S C P S P F G V N P S O P H A S S C P S P F G V N P S O P H A S S C C P S P F G V N P S C P H A S S C C P S P F G V N P S C P H A S S C C P S P F G V N P S C P H A S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G V N P S C P H H R S S C C P S P F G S V P L K P H H R S S C C P S P F G S V P L K P H H R S S C C P S P F G S V P L K P H H R S S C C D S P F G S V P L K P H H R S S C C P S P F G S V P L K P H H R S S C C D S P F G S V P L K P H H R S S C C D S P F G S V P L K P H H R S S C C D S P F G S V P L K P H H R S S C C D S P F G S V P L K P H H R S S C C T N P Y G S Y P L K P H H R S S C C T N P	L A C G W D E Y H S K O P N P Q I L Y G L A C G W D E Y H S K O P N P Q I L Y G W P C E Q E M I Q O N Y N P O T L Y G W P C E Q E M I Q O N Y N P O T L Y G A C G H T K M N S P P N H E I L Y G C L C E W R Q A Y E D K P N P Q I L Y G Y P C S W D N F N P G N P Q I L T G W C S W D N F N P G N P Q I L T G A C S H A D L N A P G A N P G T L Y G A C S H A D L N A P G A N P G T L Y G A C S H A D L N A P G A N P G T L Y G A C S H A D L N A P G A N P G T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S H A D L K N Y G P S P H T L Y G A C S W A Q O A K G P N P H T L Y G A C S W A Q O A K G P N P H T L Y G A C S W A Q O A K G P N P H T L Y G A C S W A Q O K K S T F N Y H Y L T G S P C D Y NGM N N P G P N Y C L W G S P C D Y NGM N N P G P N Y C T L W G S P C D Y NGM N N P G P N Y C T L W G A C C W A Y S T K F N P Q Y L Y G W K G W E A Y R S T K D N P Q Y L Y G W K G W E A Y R S T K D N P Q Y L Y G V K G W E A Y N S T K D N P Q Y L Y G V C G W S T N S P P N P Q L L Y G V C G W S T N S P P N P Q L L Y G V C G W S T N S P P N P Q L L Y G V C G W M S T N S P P N P Q L L Y G V C G W M T N S Y P D N P Q L L Y G V C N W N F N S P D N P Q L L Y G V C N W N F N S P D N P Q L L Y G V C N W N F N S P D N P Q L L Y G V C N W N T N S P S P N P Q L L Y G V C N W N T N S P S P N P Q L L Y G V C N W N T N S P S P N P Q L L Y G V C N W N T N S P S P N P Q L L Y G V C N W N T N S P S P N P Q L L Y G V C N W N T N S P S P N P Q L L Y G V C N W N T N S P S P N P N L Y G M C S W D N N N P S P S D N P N L Y G M C S W D N N N S S P N L Y S N L Y S Y C N L Y S Y C N Y S N L Y S Y C N L Y S Y C N L Y S Y C N Y S Y C N L Y Y S Y C N L Y S Y C N Y S Y C N Y S Y C N	L         V S G P D OND H Y E D RR         394           L         V S G P D OND H Y E D RR         394           L         V S G P N E D D E L D T R         398           A         V S G P N E D D E L D T R         398           A         V S G P N V E D D L T D T R         398           A         V S G P N V E D D L T D T R         398           A         V S G P N V E D D T F V D T R         394           A         V S G P O N D N F K D E R         397           A         V G G P O N D N F K D E R         397           A         V G G P O N D N F K D E R         397           A         V G G P O N D N F K D E R         394           A         V G G P O M D N F K D E R         394           A         V G G P G G D O Y A D K R         384           A         V G G P G G D D Y K D N R         394           A         V G G P G G S D S Y T D S R         395           A         V G G P G S D S Y T D S R         392           A         V G G P G S D N Y V D R         392           A         V G G P G N D D Y V D D R         392           A         V G G P D C N D Y V D R         392           A         V G G P G D N N Y V D R         392
XP_022338490.1 XP_022338491.1 XP_022341787.1 XP_0233704929.1 XP_023704929.1 XP_023704930.1 XP_023716596.1 XP_023718476.1 XP_396791.3	YAANAAFVALAA) YAANAALVALAA) YTSNMAFMALLA HASNAAFACLEAJ MAANAAFICASA) MVANAAYLCFQAJ HAANVAFVCLQAJ KIAAAAYICFQLJA	NEDGD Q F K SWA L S Q I N Y I / DE KE D I K I WA L K Q I N Y M A D V G V H T Q - Y R QWA R S Q I G Y A K K I S I D T K - F Q Q F A M D Q I H Y A N D L G I N A D S N R Q F A K K Q L D Y I A D L G I S V D T Y R Q Y G K K Q I D Y I A D P T K Y R E F A R R Q I D Y M A D L G A S D F Y R Q F F A E Q Q I Y Y M	L G N N H M S Y E I L G N K E M S Y Q I L G D T G R S F V V L G D T G R S F V V L G D G G R S Y V V L G D S G R S Y V V L G D S G R S Y V V L G D S G R S Y V V L G D G G G R S Y V V	G G N N Y P R O P H H R G S S C P Y P P Y G N K Y P T K P Y H R A S S C P O P P F G T N P P Q R P H H R S S S C P T N P G G N N P T R P H H R S S S C P T A P Y G N N P T H P H H R S S C P T A P Y G N N P K T P H P H R S S C P T A P F G N N P K E P H H A A S S C P T A P F G N N P K C P H A A S S C P T A P	A GCSWNE FNSGGPN PQVLKG SPCGWDNINNPGPSPQILKG SPCSFADQQPGPNPHTLYG AVCDWNSYNNPGPNPSTLWG AVCDWNTNSASPNYHVLG AVCDWNTNSESPNYHVLG APCGWDAFKWPRENPQLLKG AVCNWDTYWSAEPNYHILTG A LCGWSEFDKDAPNPQILKG	A L V G G P G G N D E Y A D S R 386 A L V G G P D V N D 385 A L V G G P N G D Q Y T D N 393 A L V G G P G L D D S Y A D L R 391 A L V G G P D S D N Y K D F R 396 A L V S G P D E N D Y K D K R 398 A L V G G P D E N D Y K D K R 398 A L V G G P D E A D K FH D H R 394
AAA23086.1 AAA38027.1 AAF8058.1 AAF8058.1 AAF8058.1 AAF8058.1 AAC8058.1 AAC8058.1 AAC8058.1 AAC8058.1 AAC8058.1 AAC8058.1 AAC8058.1 AAC8058.1 AAC8058.1 ABC24276.1 AB		SDDATRKARYHDFGVRQINYA         AAYTGALA         AAFGGALA         AGYQGCLA         AGFQSAVA         AGFGSAVA         AGFGCAAVA         AGFGCAAVA         AGFGTAVA         AGFGSAVA         AGFGSAVA         AGFGSAVA         AGFGSAVA         AGFGSAVA         AGFGSAVA         AGFGSAVA         AGFGSALA         AGFGSALV         AGFGSALV         AGFGSALV         AGFGSALA <td><b>G</b>DNPRSYVV</td> <td><b>g</b>FGAN ₽₽ TAPHHRT</td> <td>ahgswdsi tt padsrhvl y</td> <td>L         G         G         S         NDAYTDS         999           418         419         419         419           419         419         421         416           421         416         421         416           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414         414           414         414         414         414         414         414<!--</td--></td>	<b>G</b> DNPRSYVV	<b>g</b> FGAN ₽₽ TAPHHRT	ahgswdsi tt padsrhvl y	L         G         G         S         NDAYTDS         999           418         419         419         419           419         419         421         416           421         416         421         416           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414           414         414         414         414         414           414         414         414         414         414         414 </td

	±
AGP76437.1	SDYVQNEWACDYNAGFQSAVF
AUT 11011 1	NDYVKNEWACDYNAAETCALA
AKC91349.1	NDYVKNEVATDYNAGFOSAVA
AKV16371.1	SDYISNEVATDYNAGFQSAIA
AMH40356.1	KDAQHNEVSCDYNAGFQGVIA
AMH40357.1	KDAQHNEWACDYNAGFQGVIA
AMH40358.1	ADXVHNEWDTVXNAGEOGVLA
AMH40360.1	SDYVHNEMACDYNAAYOGALA
AMH40361.1	KDAPHNEVACDENAALQSCIA
AMH40362.1	KDAPHNEWACDENASFQACIA
AMH40363.1	SDYVHNEVACDYNAAFQGVIA
AMH40364.1	TDWRHSEVGCDYNACFQSMLA
AMH40365.1	NDYVHNEMATYYNSGFOTVVA
AMH40367.1	KDAQHNEWACDYNAAFQGVIA
AMH40368.1	TOWRHSEVGCOYNACFOSMLA
AMH40369.1	KDAPHNEWACDFNASFQACIA
AMH40370.1	K DAQKNEWACDYNAAFQSVIA 8 DYNHNEMDTYNNACEOCYLA
AMH40371.1	SDYVHNEMACDYNAAYOGALA
AMH40373.1	SDYVHNEWATYYNSGFQTIIA
AMH40374.1	SDYVHNEWACDYNAAYQGALA
AMH40375.1	TDYQHTEWGCDYNACFQSMLA
AMH40376.1	KDATHNEWTCDNNAGYHGVLV
AMH40377.1	S D Y Y H NEWD T Y YN A G F O S Y Y A
AMH40379.1	SDYVHNEWACDYNAAYQGALA
AMH40380.1	TDYQHTEVGCDYNACFQSMLA
AMH40381.1	SDYVHNEWDTVYNAGFQGVLA
AMH40382.1	KDAPHNEWACDFNVAFQACIA
AMH40383.1	K DAQH NEWACDYNAAFQGV I A
AMH40385.1	SDYVHNEWATYYNAGFOSVLA
AMH40386.1	TOWRHTENGCOYNACFOSMLA
AMH40387.1	SNYVVNEWDTVYNAGFQGVLA
AMH40388.1	KDAQHNEWACDYNAAFQGVIA
AMH40389.1	KDAQHNEWACDYNAAFQGVIA
AMH40390.1	KDARHNEWACDENAAEOSCIA
AMH40392.1	QDAQHNEWACDYNSAYQGAL S
AMH40393.1	SDWRHTEVGCDYNAAFQSTLA
AMH40394.1	SDYQHNEVACDYNAAFQGILA
AMH40395.1	EEYVYNEWTLDYNAGFQSAVA
ANR02619.1	ODAOHNEWACDYNAGFQGALA
AOV94251.1	TDYOHNEWACDYNAAFSGVLA
AOV94252.1	YDAVHNEVACDYNAAFQGVIA
AOV94253.1	RDYVYNEVTLDYNAGFQSAVA
AOV94254.1	TDYEHNEWACDYNAAFSGVLA
AOV94255.1	TDPQHNEWACDYNAAFSGALA
ADC00556 1	ND Y V K NEWA CD YN A AFSGVLA
APF29557.1	NDYVKNEVACDYNAGFTGALA
APF29558.1	NDYVKNEVACDYNAAYTGALA
APF29559.1	NDYVKNEWACDYNAAFTGALA
APF29560.1	NDYVKNEVACDYNAAYTGALA
APF29561.1	DDYVHNEWACDYNAGITAGAA
A1591249.1 BAA31328.1	SDY I SNEWATDYNAGEOSAVA
BAA33708.1	SDYVHNEWATDYNAGFQSALA
BAA33709.1	SDYVHNEVATDYNAGFQSALA
BAA34050.1	SDYISNEWATDYNAGFQSAVA
BAB40693.1	SDYISNEVATDYNAGFQSAVA
BAB40695.1	SDYISNEWATDYNAGFOSAVA
BAB40697.1	SDYISNEWATDYNAGFOSAVA
BAC67186.1	EDYVHNEVACDYNSGFQSALA
BAD01504.1	EDYVHNEVACDYNSGFQSALA
BAD44734.1	EDYVHNEWACDYNSGFQSALA
BAH38757.1 BAH22180.1	SDY I SNEWACDYNAGEOGAVA
BAK20401.1	Q D Y I S NEVAC D Y NAGFQGALA
BAL60587.1	NDYVHNEVACDYNAGYTSLLA
BAM14716.1	SDYIRNEVANDYNAGFQGAVA
BAS18736.1	K D Y V L N E V A T D Y N S G F H S A L A
BA121140.1 BAU45387 4	KDYVINEWATDYNAGEQGAVA
BAU45388.1	KDYVLNDVATDYNSGFHSALA
BAX09271.1	SDYTHNEWATDYNCGFQGALA
CAD54726.1	SDYVANEVATDYNAAFQSLLS
CAD54727.1	S DY VANEVA T DYNAAFQSLLS
CAD54728.1	SDYVANENATDYNAAFOSILA
CAD54730.1	SDYVANEVATDYNAAFOSLLA
EFN70196.1	HDGNYTGVGIDYNAGFTCVLA
EFN70197.1	HDGNYTGVGIDYNAGFTCVLA
EFX69372.1	SNYVSNEVA TDYNAGFOGAVA
EEX80605.1	HNEETNEWACDYNAGEOASVA
EFX86822.1	DDFRANEWALDYNAGFQSALA
EGI63652.1	EEYAYTEVTLDYNAGFTSALA
EKC17289.1	LDFQSNEVACDYNAGFQSAVA
EKC39023.1	UNFRINE ACDYNAGFOSAVA
KMO954+9-4	EEDLYTENI INVNACCTOALA
KOC62382 1	EDYVYTENTLDYNAGFTSALA
KOF68403.1	SDFIHNEVICDYNAGFHSAVS
KOX76353.1	EDYVYTEVTVEYNGGFTSALA
KXJ18436.1	KDYIENEVATDYNAGFQSAVA
KYB27092.1	EEFLYNEWTLDYNAGFOSTLA
KYN01696 4	EEYAYTENTLDYNAGETSALA
KYN22851.1	EEYAYTENTLDYNAGFTSTLA
KYN35700.1	EEYAYTETLDYNAGFTSALA
KYQ52831.1	EEYAYTEVTLDYNAGFTSALA
KZC10616.1	EDYVYTENTLDYNGGFTGALA
KZS03923.1 KZS05722.4	ENYOTNE ALDYNAGEOASLA
KZS07093.1	QNYVTNEWATDYNAGFQAAVA
KZS16514.1	DNYVTNEWACDYNAGFQASLA
KZS16516.1	ENYVTNEWACDYNAGFQASLA
OAD61293.1	EDYVYTEWTLDYNAGFTSALA

0011010001		Cells.			
ODM94862.1	EDYVAN	빌	OT T D Y	N	AGFQGALA
OW/E45603 (	ENVVNN		ACDY	N	RCEOCARS
OWF56607.1	SDYION	E	ACDY	N	AGEQSVVA
PBC31793.1	EDYVYT	E	TLDY	N	AGFTTCIA
pdb/3WC3/A	SDYIRN	E	ANDY	N	AGFQGAVA
PFX25363.1	KDYVMN	E	VTTDY	N	AGFQSSVA
PSN31180.1	NDAQHN	E	ACDY	N	AGFQSALA
PVD35337.1	KDYVHN	E	VACDY	N	AGFQSAVA
PVD35340.1	GNYINN	E	ACDY	N	AGFQGSVA
PVD35381.1	ENYVNN	E	ACDY	N	AGFQTAVA
PVD35383.1	ENYVNN	E	ACDY	N	AGFQTAVA
PVD35384.1	DQYLYN	G	AITY	N	AGEQTAVA
PVD35385.1	EEYYYN	2	AIVY	N	AGEQTAVA
PVD35551.1	RDYVMN	빝	AIDY	N	SGFQSALA
PVD30024.1	SNTVNN	150	ACDY		AGEQTAVA
PVD38825.1	PRVVUN		ALDY	N	SGEOSALA
PVD36986 1	EDYVKN	E	ALDY	N	SGEOSALA
XP 001606454.1	EDYVYT	E	TLDY	N	AGFTGALA
XP 001810693.2	EEFLYN	E	TLDY	N	AGFQSTLA
XP_001944774.2	EEYIYN	E	TLDY	N	AGFQSAVA
XP_002426465.1	EEYVYN	E	TLDY	N	AGFQSAVA
XP_003402778.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_003690676.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_005099312.1	LNYTLN	E	ATDY	N	SGFQGALA
XP_005099315.1	GDYVHN	E	VATDY	N	CGFQGSLA
XP_005101787.1	TDYQHN	E	ATDY	N	SGFQGALA
XP_005101789.1	TDYKHN	E	ATDY	N	SGFQGAMA
XP_005111868.2	GNFQNN	E	VATDY	N	AGFTTAVA
XP_005111869.2	SNYVNN	E	ATDY	N	AGFTTAVA
XP_005111870.1	LNYVNN	비	AIDY	N	AGETTATA
XP_006610635.1	EDIVII	1	TIDY		AGFISALA
XP 008404466 4	FFFIV	j.	TINY		AGEOSTIA
XP 009049791 1	DDYLHN		ATOV	N	AGEOASIA
XP 009049838 4	NDYIHN	F	ACDY	N	AGEOSAVA
XP 009049839 1	KDYTKN		ACDY	N	AGEQSTVA
XP_009049840.1	EDFION		ACDY	N	AGFQSALA
XP 009049941.1	GNYVNN	E	ATDY	N	AAFQGAVA
XP_009058133.1	QNYINN	E	ATDY	N	AGFQSAVA
XP_009058134.1	EDYVHN	E	ACDY	N	SGFHSALA
XP_009060290.1	EDYVQN	E	ACDY	N	AGFQSGLA
XP_009064976.1	DNYVSN	E	ATDY	N	AGFQSAVA
XP_009064980.1	GDYVRN	E	ALDY	N	AGLQSACA
XP_009065702.1	DDYVHN	E	ACDY	N	AGFQSALA
XP_011057333.1	EEYAYT	E	TLDY	N	AGFTSALA
XP_011164235.1	EEYAYT	E	TLDY	N	AGFTSALA
XP_011253653.1	HDGNYT	G	GIDY	N	AGFTCVLA
XP_011253654.1	HDGDYT	S	RLDY	N	AGFICVLA
XP_011253655.1	HDGNYT	9	GIDY	N.	AGFICVLA
XP_011296914.1	EDIVII	2			AGETSALA
XP_011296910.1	DDVIVT	5	TLDY	N	AGETSALA
XP_011421782.1	LDEOSN	F	ACDY	N	AGEOSAVA
XP 011423159 1	KDYVSN	F	ACDY	N	AGEOSAVA
XP 011430172.1	N	H	GCDF	N	AGFQSALA
XP 011447501.1	DNFRTN	E	ACDY	N	AGFQSAVA
XP_011447502.1	DNFRTN	E	ACDY	N	AGFQSAVA
XP_011498432.1	EDYVYT	E	TLDY	N	AGFTGALA
XP_011632703.1	AEYIYT	E	TLDY	N	AGFTSALA
XP_011700378.1	EEYAYT	E	TLDY	N	AGFTSVLA
XP_011867251.1	EEYAYT	E	TLDY	N	AGFTSALA
XP_012061951.1	EEYAYT	E	TLDY	N	AGFTSALA
XP_012146513.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_012224814.1	EEYIYI	E	TLDY	N	AGFTGALA
XP_012248155.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_012273697.1	EEVAVT	120	UTLDY		AGFISALA
AP_012020004.1	EETATI			TN IN	AGFISALA
XP_012935745.1	DNYQNN	E	ATDY	N	AGETTAIA
XP_012938497.1	DNVGQN	비	ATDY	N	AGFQGALA
XP_012939931.1	SUTIEN	E	ATDY		SCEOCALA
XP 012941225 1	GDYIKN	j.	TTDY	1	AGEOSALA
XP_012945520.1	ENYONN	E	ATDY	N	AGETTAVA
XP 013066277.1	PDWIYN	R	EIDY	N	SGFQGALA
XP_013066278.1	SDYVQN	E	SISY	N	SGLHGALA
XP_013066280.1	SDYVHN	E	ATDY	N	SGFQGALA
XP_013066281.1	DDYVTN	E	ATDY	N	AGFQGALA
XP_013066283.1	DDYVTN	E	ATEY	N	AGFQGALA
XP_013066284.1	DNYVLN	E	STDY	N	AGYQGALS
AP_013066285.1	EDVIL	E.	ALDY	Ň	AGEQGALA
XP_013066286.1	TDVILN		ALDY	N	SGEOGALA
XP_013086370.1	SOVILN		ATDY		SGEOGALA
XP 013069169 1	SNYVNN	F	ACDY	Ň	AGFQTAVA
XP_013069170.1	SNYVNN	E	ACDY	Ň	AGFQTAVA
XP 013069521.1	SNYVNN	E	ACDY	N	AGEQAAVA
XP_013069522.1	SNYVNN	Ē	ACDY	N	AGFQAAVA
XP_013069523.1	SNYVNN	E	ACDY	N	AAFQGAVA
XP_013084279.1	DSVLNN	E	TIEF	N	AGFQTAVA
XP_013084281.1	DSVLNN	E	TIEF	N	AGFQTAVA
XP_013084282.1	TNEVNN	D	SIEY	N	AGFQTAVA
XP_013084283.1	TNEVNN	Р	SIEY	N	AGFQTAVA
XP_013084285.1	TNYLYN	P	ACDY	N	AGFQSAVA
XP_013084287.1	TNYLYN	P	ACDY	N	AGFQSAVA
XP_013088631.1	SDYVLN	E	ACDY	N	SGFHGALA
AP_013094438.1	TUYVHN	j5)	AIDY	N	SGFQGALA
AP_014214612.1	EDFIYT	[]	TIDY	Ň	AGETONI
XP 014786604 4	SDELLA	[]	TCOV	Į,	AGEHSAVO
XP 015126097 1	EDYVYT		TLDY	N	AGETSALA
XP 015184471 1	EEYIYT	F	TLDY	N	AGETSVIA
XP_015432538.1	EDYVYT	E	TLDY	N	GGFTGALA
XP_015513382.1	EDYIYT	E	TLDY	N	AGFTSAVA
XP_015592927.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_015836219.1	EEFLYN	E	TLDY	N	AGFQSTLA
XP_016907626.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_017766299.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_017785110.1	EEYLYN	E	TIDY	N	AGLHSLLA
XP_017793200.1	EDYVYT	E	TLDY	N	AGFTSALA
XP_018016806.1	SDYVHN	E	AVGY	N	AGLTCAL
XP_018016810.1	MDYMHN	E	ACDY	N	AALTGALA

	+
XP_018016813.1	QNWEQNE GCDYNAALAANLA
XP_018016820.1	TNFVQNEWATDYNAGFVGTLA
XP_018016821.1	TNFVQNEWATDYNAGFVGTLA
XP_018026496.1	MDFVHNEWACDYNAAFVGVLA
XP_018027134.1	NDYVKNEWACDYNAAFTGDLA
XP_018044452.1	EEYAYTEWTLDYNAGFTSALA
XP_018307051.1	EEYAYTEVTLDYNAGFTSALA
XP_018323591.1	EEFLYNEWTLDYNAGFQGVVA
XP_018347393.1	EEYAYTEWTLDYNAGFTSALA
XP_018359268.1	EEYAYTEVTLDYNAGFTSTLA
XP_018396635.1	EEYAYTENTLDYNAGFTSALA
XP_018568194.1	EEFLYNEWTVDYNAGFQSALA
XP_018568196.1	EEFLYNEVTVDYNAGFQSALA
XP_019884179.1	YEPDYTNURIDYNAGFTCVLA
XP_020289295.1	NIRCHTNVKIENNAGFTGVLA
XP_020289297.1	FN - KYTKIRIDYNAGFTGALA
XP_020289301.1	FR - EHTKWRIDYNAGFTGALA
XP_020289336.1	YGN - YTENNIEYNAGFTGALA
XP_020614867.1	KDFIKNEWATDYNAGFQSAVA
XP_020893295.1	SDYVQNEVTTDYNAGFQSAVA
XP_020893299.1	KDYIENEWATDYNAGFQSAVA
XP_021345983.1	SDYIQNEWACDYNAGFQSVVA
XP_021361084.1	KDYVKNEWACDYNSGFQSALA
XP_021363481.1	Q D Y K Q N E V A C D Y N A G F Q S A V A
XP_021363490.1	SDYTKNEWTCDYNAGFQSALA
XP_021363495.1	ENYVNNEWACDYNSGFQGASS
XP_021363496.1	ENYVNNEWACDYNSGFQGASS
XP_021363497.1	GNYINNEWACDYNAGFQSAVA
XP_021363498.1	GNYTNNEWACDYNAGFOSAVA
XP_021924915.1	EEYVYNEWTLDYNAGFQSAVA
XP_021935277.1	KETTHNMUSCDCNAAFQSAVA
XP_021941324.1	SNYTSNEWATDYNAAFQGLLA
XP_021950048.1	GDY I KNEWAT DY NAAFOGALA
XP_021996049.1	GDTTKNEGATDTRAAFQGALA
AP_021939940.1	EDVUANENTI DVNAGEOGAVE
VD 001007597.1	NDYVKNEWTTDYMACEOSALA
XP_021907001.1	EEVI VNEWTI DYNAGEOSAVA
XP_022187268.1	EEVI VNEWTL DVNAGEOSAVA
XP 022200004 1	EDV LYNEWT LDYNAGEOSAVA
XP 022328778 1	DNYRTNEWACDYNAGEOSAVA
XP 022329668.1	DNYRTNEMACDYNAGEOSAVA
and an and a start	the proceeding of the processing of the second
XP_022336001.1	NHUGCDENAGEQSALA
XP_022336153.1	LDFQSNEWACDYNAGFQSAVA
XP_022338490.1	KDYTKNEWTCDYNAGFQSALA
XP_022338491.1	- NYAHNHOGCDFNAGFQSALA
XP_022341787.1	KDYVSNEWACDYNAGFQSAVA
XP_023337121.1	SUTIKNEWACDYNAGFQSALA
XP_023704929.1	SNTISNEGAIDYNAAFTSLLA
AP_023704930.1	SNTISNEGALDY MAAPQSILA
AP_023710096.1	L DAVINI ACDYNAGEOSAVA
AF_023110410.1	EDVUVTENTI DVNAGETCALA
AP_395791.3	ODVU ANDUATOVNACETCALA
AAA23086.1	UDIVANE AIDIMAGEISALA

 $\begin{array}{c} 412\\ 4416\\ 4414\\ 4416\\ 4413\\ 4415\\ 4415\\ 4415\\ 4415\\ 4415\\ 4415\\ 4415\\ 4416\\$ 

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

<sup>76/46</sup> XP 001944774.2 Acyrthosiphon pisum XP 022167267.1 Myzus persicae XP 022200004.1 Nilaparvata lugens

Figura 2 Relaciones filogenéticas formadas por secuencias provenientes del orden

Hemiptera

<sup>61/37</sup> KXJ18436.1 Exaiptasia pallida <sup>100/98</sup> XP 020614867.1 Orbicella faveolata PFX25363.1 Stylophora pistillata

Figura 3 Relaciones filogenéticas formadas por secuencias provenientes del orden

#### Embioptera

AMH40362.1 Extatosoma tiaratum <u>100/100</u> AMH40366.1 Medauroidea extradentata AMH40373.1 Peruphasma schultei

Figura 4 Relaciones filogenéticas formadas por secuencias provenientes del orden

Phasmatodea, Clado I

ABV32557.1 Teleogryllus emma <u>100/100</u> AMH40379.1 Ramulus artemis AMH40384.1 Sipyloidea sipylus

Figura 5 Relaciones filogenéticas formadas por secuencias provenientes del orden

Phasmatodea, clado II

61/37 KXJ18436.1 Exaiptasia pallida 100/98 XP 020614867.1 Orbicella faveolata PFX25363.1 Stylophora pistillata

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

KDR16731.1 Zootermopsis nevadensis <sup>100/100</sup> KDR16731.1 Zootermopsis nevadensis <sup>89/45</sup> XP 021924915.1 Zootermopsis nevadensis <sup>89/45</sup> XP 023716596.1 Cryptotermes secundus

Figura 14 Relaciones filogenéticas formadas por secuencias provenientes de la

familia Termitodae, Clado I, presentes en el clado Polyneoptera/

Holometabola/Paraneoptera







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 Termitodae, Clado II, presentes en el clado Polyneoptera







## Figura 21 Relaciones filogenéticas formadas por secuencias provenientes del filo Mollusca







Figura 23 Relaciones filogenéticas formadas por secuencias provenientes del orden Hymenoptera

# **Grupos Taxonómicos Comprimidos del Árbol 3**



Figura 24 Relaciones filogenéticas formadas por secuencias provenientes del





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

XP 023716596.1 Cryptotermes secundus 100/100 KDR16731.1 Zootermopsis nevadensis XP 021924915.1 Zootermopsis nevadensis

Figura 28 Relaciones filogenéticas formadas por secuencias provenientes de la

familia Termitodae, 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