

**DESCUBRIMIENTO DE PEQUEÑOS MARCOS DE
LECTURA EN EL CROMOSOMA 2 DE DROSOPHILA
MELANOGASTER**

Trabajo final de la alumna

Pamela Magjor

Este trabajo ha sido presentado como requisito para la obtención del título
de
Licenciada en Genética

Carrera: Licenciatura en Genética

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

Pergamino, 20 de diciembre de 2011

**DESCUBRIMIENTO DE PEQUEÑOS MARCOS DE
LECTURA EN EL CROMOSOMA 2 DE DROSOPHILA
MELANOGASTER**

Trabajo final de la alumna

Pamela Magjor

Este trabajo ha sido presentado como requisito para la obtención del título
de
Licenciada en Genética

Director: Rolando Rivera Pomar

Carrera: Licenciatura en Genética

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

Pergamino, 20 de diciembre de 2011

AGRADECIMIENTOS

A todas las personas que participaron haciendo posible este proyecto, muchas gracias por su apoyo y enseñanza:

A Rolando Rivera Pomar, director, quien siempre me guió y no dudó en ayudarme incondicionalmente cada vez que lo necesité;

A la Escuela de Ciencias Agrarias Naturales y Ambientales (ECANA) de la UNNOBA, por haber permitido realizar mis estudios y lograr mi formación;

A la Comisión de Investigaciones Científicas (CIC), por darme la posibilidad y el sustento necesarios para realizar este trabajo;

A todos mis profesores y compañeros que fueron fundamentales en este camino;

A Elin, por ser mi compañera y cómplice en las dificultades que se me presentaban;

A mi mamá, mi hermana y mi sobrino, por su apoyo en esta tesis y en toda mi vida;

A Walter, por su infinita paciencia y su fiel compañía;

A mis amigos, que me soportaron y acompañaron;

A mis familiares, por estar siempre que los necesito.

Nota

Los resultados presentados en este trabajo final fueron parte del siguiente trabajo científico: Diambra, L., Ons, S., Gutiérrez, P., Gennero, E., Magjor, P., Rivera-Pomar, R., Genomic analysis of polycistronic transcripts in insects, Sixth International Symposium on Molecular Insect Science, Amsterdam, Holanda, Octubre 2011.

INDICE

Introducción.....	Página 5
Genómica.....	Página 5
ORFs y sORFs.....	Página 9
Razones para analizar a <i>Drosophila melanogaster</i>	Página 17
Hipótesis.....	Página 20
Objetivo.....	Página 20
Materiales y métodos.....	Página 21
Herramientas bioinformáticas.....	Página 21
Bases de datos.....	Página 21
BLAST.....	Página 21
Augustus.....	Página 22
Resultados.....	Página 23
Análisis de nuevos genes correspondientes a regiones genómicas no anotadas y con estructura predicha.....	Página 27
Regiones genómicas con validación tanto por ESTs como por espectrometría de masas.....	Página 27
Regiones genómicas con validación sólo por espectrometría de Masas.....	Página 79
Discusión y conclusiones.....	Página 136
Bibliografía citada.....	Página 139
Resumen.....	Página 140
Anexos.....	Página 141

INTRODUCCIÓN

GENÓMICA

El genoma es el complemento genético entero y haploide de un individuo. Es decir, toda la secuencia de ADN (ácido desoxirribonucleico) de un organismo. Define las características de cada ser vivo, ya que cada individuo se distingue por su material genético. Por lo general, al hablar de genoma en los seres eucarióticos nos referimos sólo al ADN contenido en el núcleo, organizado en cromosomas, aunque también las mitocondrias y los cloroplastos contienen genes.

Un genoma incluye genes que codifican productos proteicos y ARNs (ácido ribonucleico), pseudogenes, los cuales representan secuencias nucleotídicas similares a un gen normal que derivaron por duplicación y posterior divergencia de un gen funcional pero sin capacidad de producir una molécula de ARN o una proteína funcional, es decir, sin expresión (1). Se creía que los pseudogenes eran “fósiles genéticos” que quedaron de un momento en la evolución del genoma en la que pueden haber tenido una función importante, hoy perdida. Sin embargo, nuevas evidencias indican que pueden funcionar como elementos reguladores. Se ha determinado en levaduras que el genoma se transcribe de manera masiva tanto en una como en otra hebra de ADN. Esto se denomina transcripción difusa (pervasive transcription) e implica que se transcriben muchos más ARNs que los esperados como codificantes. Esto llevó al concepto de ARN no codificante (ARNnc), algunos de los cuales son pseudogenes con función regulatoria. Para mas detalles, ver (www.pervasivetranscription.com).

Otros elementos genéticos que podemos encontrar en un genoma son los orfones, los cuales constituyen secuencias de ADN solitarias y dispersas por el genoma de un organismo que derivan de familias multigénicas en tándem; por tanto el orfón es una

secuencia de ADN que ha perdido su familia (2). También encontramos en un genoma regiones no codificantes, las cuales pueden ser regulatorias o estructurales.

Para estudiar un genoma es necesario realizar un mapeo genético, un mapeo físico y una secuenciación exhaustiva. El mapeo genético se basa en el uso de técnicas genéticas, como experimentos de recombinación y estudios de pedigrees, para construir mapas que muestren la posición de diferentes marcadores genéticos. El mapeo físico utiliza técnicas de Biología Molecular para examinar directamente las moléculas de ADN para la construcción de mapas que muestran la posición de fragmentos de ADN (en general de secuencia desconocida) en el genoma.

Como no es posible aún secuenciar un genoma en una sola reacción de secuenciación se lo divide en fragmentos, los cuales se almacenan en forma de clones en bibliotecas genómicas. Una biblioteca genómica es un conjunto de clones, cada uno de los cuales contiene un fragmento del genoma de un organismo dado. Las bibliotecas genómicas se obtienen clonando los fragmentos del genoma de un organismo en vectores plasmídicos.

La genómica es una disciplina moderna que permite estudiar el ADN en diferentes niveles (desde un gen individual hasta el complemento genético completo de un organismo), es decir, se ocupa de caracterizar los genomas. Se divide en Genómica Estructural y Genómica Funcional. La Genómica Estructural tiene como objetivo conocer los genomas de diferentes especies y determinar los genes contenidos en él, así como las secuencias regulatorias; la Genómica Funcional determina las funciones e interrelaciones de los genes con respecto a procesos celulares específicos. Cuando la secuencia de un genoma es determinada completamente y cada gen es anotado (es decir, es asignada una identificación tentativa), se pueden agrupar genes de acuerdo a procesos en los que participa la presunta proteína codificada por el gen (por ejemplo, síntesis de

ADN, síntesis proteica) con lo cual se establece que cada fracción del genoma está asociada a actividades celulares determinadas.

El objetivo de la anotación genómica es predecir genes, secuencias promotoras y reguladoras de la expresión, y definir su función en el contexto del genoma entero. Por otro lado, la genómica implica la comparación de genomas de distintos individuos de la misma especie para detectar polimorfismos, y comparar genomas de individuos de diferentes especies para proveer datos evolutivos. A mayor similitud entre genomas, mayor cercanía evolutiva entre dichos organismos comparados. Este procedimiento se denomina filogenia molecular derivada de secuencias génicas o filogenómica. Las comparaciones genómicas apoyan conclusiones basadas en dos décadas de investigación sobre genes que se han conservado durante cientos de millones de años. Debido a la conservación de genes y proteínas, los resultados de experimentos en un organismo son guías útiles para investigar en otros organismos.

Los proyectos de secuenciación genómica han expandido el conocimiento sobre la estructura y organización del genoma y han ofrecido un desafío para identificar genes funcionales en dichos genomas. La combinación de la secuenciación genómica y de algoritmos computacionales de hallazgo de genes produjo el inventario de una gran parte de los genes codificantes de proteínas para diferentes organismos. Las funciones de la mitad de las proteínas codificadas son conocidas o se produjeron por medio de comparación de secuencias pero muchos genes y proteínas aún esperan ser identificados. Asimismo aún no se ha establecido un método certero que permita determinar con precisión el número de genes. De hecho, los métodos de predicción informáticos varían en función del algoritmo usado y también varían en función del método de análisis. Como ejemplo podemos mencionar la primera aproximación basada en el borrador del genoma humano: algunos predijeron 35.000 a 120.000 genes (3).

La determinación de la secuencia completa del genoma de varios organismos reveló variaciones en el número de genes codificantes de proteínas y en su organización en procariotas y eucariotas. La disposición más frecuente de genes codificantes de proteínas en los procariotas tiene una lógica notable: los genes dedicados a un único objetivo metabólico se suelen encontrar en una disposición contigua en el ADN. Tal disposición de genes en un grupo funcional se denomina operón porque opera como una unidad a partir de un único promotor. La transcripción de un operón produce una hebra continua de ARN mensajero que transporta el mensaje para una serie relacionada de proteínas. Cada secuencia del ARN mensajero representa la unidad o gen que codifica una de las proteínas en la serie. En el ADN procarionte, los genes están estrechamente empaquetados, con muy pocos intervalos no codificadores, y el ADN se transcribe en ARN mensajero colineal (policistrónico), que luego es traducido a proteína.

En cambio, los genes eucariontes dedicados a una única vía están habitualmente separados físicamente en la hebra de ADN y suelen localizarse en distintos cromosomas. Cada gen se transcribe a partir de su propio promotor, produciendo un ARN mensajero, que generalmente es traducido para producir un único polipéptido. Se cree que los organismos eucariontes también podrían incluir ARN mensajeros policistrónicos codificantes de pequeños péptidos (genes no canónicos), ya que los ejemplos conocidos son muy pocos y sin confirmación funcional, tal como se explica en párrafos siguientes. El análisis de este tipo de genes es uno de los objetivos de este trabajo.

Las secuencias genómicas deben ser descriptas y organizadas de una manera confiable y actualizada, para poder obtener información útil a partir de ellas. A este proceso se lo denomina anotación. En el proceso de anotación se utilizan diversas fuentes, entre las cuales están: publicaciones que reportan nuevas secuencias, revisiones que actualizan

periódicamente la anotación de familias o grupos de proteínas, expertos externos y análisis de secuencias. Estos se complementan con la predicción bioinformática de marcos de lectura codificantes y en la predicción de las uniones de exones e intrones. Sin embargo esto presenta limitaciones dadas por la escasa información disponible, en cuanto al número de organismos secuenciados. Un genoma anotado incluye, además de la secuencia, una descripción de la probable función de cada producto génico, la cual se deduce de comparaciones con otras secuencias genómicas y con funciones proteicas establecidas.

ORFs y sORFs

Se denomina marco abierto de lectura (ORF, del inglés *Open Reading Frame*) a cada una de las secuencias de ADN comprendida entre un codón de inicio (ATG) de la traducción y un codón de terminación, descontando las secuencias que corresponden a los intrones en caso de haberlas. Se encuentra acotado por los UTRs (secuencias no traducidas). Un ORF se define por métodos bioinformáticos en secuencias que suelen contener al menos 100 codones. Esto excluye a proteínas conocidas y con menos de 100 aminoácidos, como podría ser la insulina. Esto revela las limitaciones de los métodos empleados al día de hoy y es motivo también del estudio presentado en este trabajo final.

En una secuencia de ADN cualquiera hay, *a priori*, 6 posibles sentidos en los que pueden aparecer marcos abiertos de lectura; dado que cada codón esta formado por 3 nucleótidos, existen 3 posibles lugares de inicio para tomar los nucleótidos de 3 en 3. A lo que hay que sumar los otros 3 posibles marcos abiertos de lectura si el ADN es traducido tomando como molde la hebra complementaria, dando el sentido de lectura

opuesto. Estos marcos abiertos de lectura se denominan +1, +2, +3, -1, -2 y -3, respectivamente.

El número de genes que codifican proteínas en eucariotas superiores identificados por proyectos de genoma es más bajo que lo esperado, mientras que ha sido detectada una considerable cantidad de transcritos putativamente no codificantes. Esto refleja, nuevamente, las limitaciones de los sistemas predictivos. Pero en caso de ser correctas, ¿cómo es posible que la enorme variabilidad fenotípica de los organismos tenga como origen un pequeño número de genes mayormente conservados? Varias respuestas han sido propuestas: interacciones génicas complejas, splicing alternativo, y genes no codificantes. Otra posibilidad menos discutida podría ser muchos genes no canónicos y por lo tanto no identificados, escapen a la detección de los programas de anotación génica.

Algunos métodos de detección génica y anotación permiten encontrar genes con sORFs (pequeños marcos abiertos de lectura, del inglés *small open reading frames*) que codificarían pequeños péptidos, los cuales esperan caracterización. Alguno de estos, sin embargo, son anotados en base a similitudes encontradas entre organismos a partir de péptidos que fueron caracterizados antes de la secuenciación del genoma, como por ejemplo, las histonas, la insulina, la RNAsa, todos ellos con menos de 100 aminoácidos.

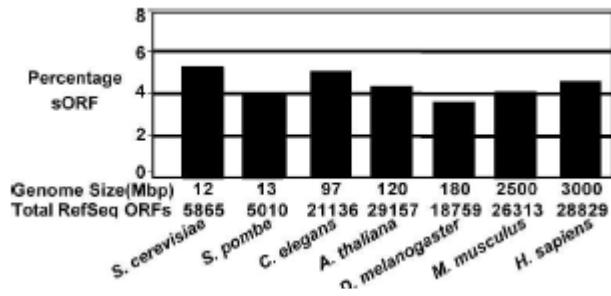
Aquellos sORFs menores a 100 aminoácidos correspondientes a genes funcionales pueden escapar a la detección debido a que se consideran sORFs sin sentido formados por azar.

Es sabido que péptidos pequeños pueden tener funciones biológicas importantes, mayoritariamente en la comunicación neural y endócrina, pero, en todos estos casos descritos, los péptidos son productos maduros clivados de un ORF más largo.

El genoma secuenciado y anotado inicialmente de *Saccharomyces cerevisiae*, tuvo en cuenta todos los ORFs de al menos 100 codones contiguos no contenidos dentro de un ORF más largo. El primer estudio funcional de sORFs se realizó utilizando como modelo a *Saccharomyces cerevisiae* (4). En esta investigación, se realizaron diferentes experimentos y análisis computacionales, hallándose evidencia de 299 sORFs en el genoma de *Saccharomyces cerevisiae*, lo cual representa el 5 por ciento de los ORFs anotados. Estos experimentos representaron la primera evidencia de la transcripción de cientos de marcos de lectura no anotados en *Saccharomyces cerevisiae*, incluyendo muchos sORFs. Uno de los experimentos realizados consistió en crear colecciones de cepas de delección génica haploides, en las cuales se eliminaban diferentes sORFs. Las delecciones dieron como resultado un fenotipo letal o de crecimiento lento, lo cual indica que la función dada por estos sORFs es esencial. Uno de esos sORFs fue caracterizado, y se determinó que es un blanco downstream de la vía mediada por MEC-1, la cual regula la detención del daño en el ADN y de la replicación. Estos resultados validaron la significancia funcional de sORFs, y sugieren que otros sORFs pueden tener funciones importantes. Además, se determinó un porcentaje similar de sORFs anotados en otros eucariotas (Figura 1), y 184 de los sORFs de *Saccharomyces cerevisiae* exhiben similitud con ORFs en otros organismos, de lo cual se desprende que sORFs están conservados en otros organismos y tienen funciones biológicas importantes.

En este trabajo final aplicamos una serie de métodos bioinformáticos que nos ha permitido descubrir cientos de potenciales nuevos genes que codifican sORFs, con la posibilidad de que puedan existir varios miles. En *Populus deltoides* se identificaron sORFs que codifican péptidos, los cuales se encuentran conservados entre *P. deltoides* y otras 14 especies de plantas, sugiriendo la conservación interespecífica de dichos sORFs (5). Estos números están en línea con estimaciones realizadas en otros eucariotas,

Figura 1: sORFs constituyen un porcentaje similar de ORFs anotados en eucariotas representativos

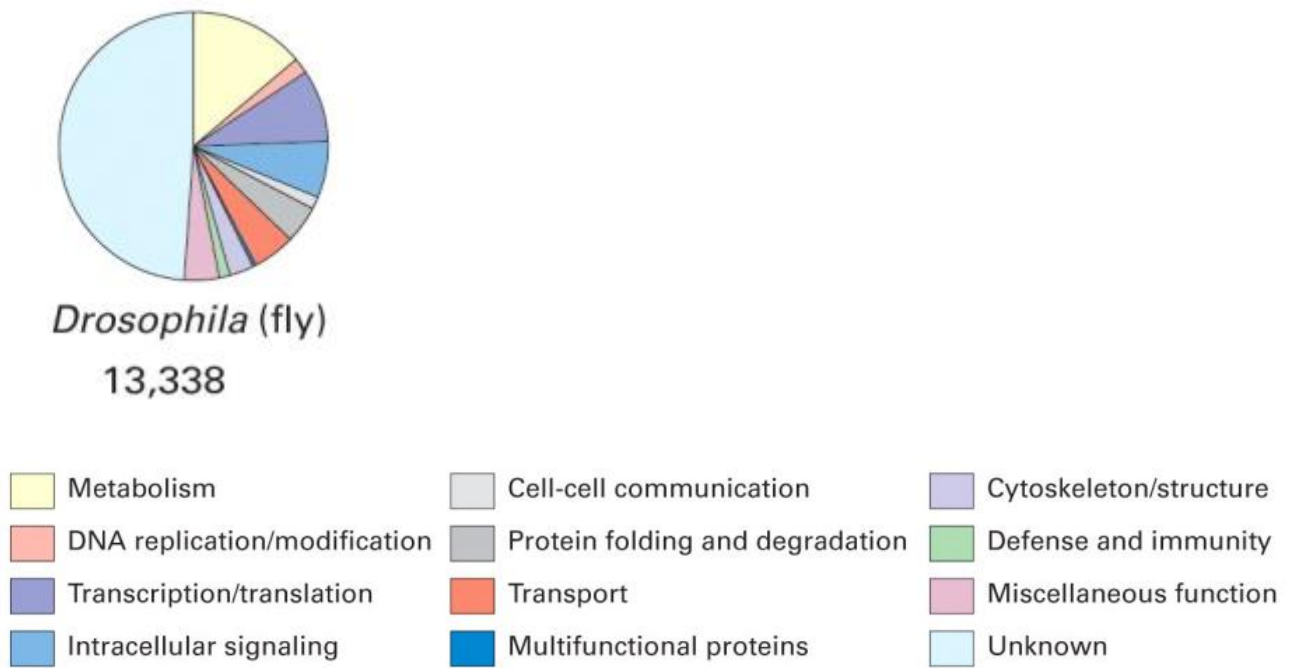


Tomado de Kastenmayer James P. et al., Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*, 2006, Genome Research, 16:365–373 ISSN 1088-9051/06.

sugiriendo que sORFs pueden representar una fracción significativa del genoma eucariótico traducido (Figura 2).

En *Drosophila melanogaster*, se caracterizó el gen *tarsal-less* (*tal*), un nuevo tipo de gen no canónico que ha sido previamente clasificado como ARN no codificante (6). En este estudio se demostró que *tal* controla la expresión génica y la morfogénesis en *Drosophila melanogaster*. La función de *tal* es mediada por varios ORFs de 33 nucleótidos contenidos en un solo transcripto, siendo este un locus policistrónico. *tal* define dos nuevos paradigmas en los genes codificantes eucariotas: la existencia de pequeños péptidos con función biológica, y su arreglo en mensajeros policistrónicos. *tal* se identificó a través de mutantes espontáneos (*tal^l*) cuyo fenotipo era la presencia de patas defectuosas, en las cuales los segmentos del tarso no se desarrollan. La expresión de *tal* en las patas es transitoria, y coincide con la especificación de la región del tarso por activación de genes específicos en un patrón en anillos similares a *tal*. Uno de los genes activados transitoriamente en el mismo momento que *tal*, y requerido para la formación del tarso es el factor de transcripción *rotund* (*rn*), cuya expresión se pierde en mutantes *tal*, pero la pérdida de *rn* no afecta la expresión de *tal*. Estos resultados demuestran que el gen *rn* es un blanco downstream de *tal*. Expresión ectópica de *tal* induce la morfogénesis en las patas, lo cual no es producido por cambios en la expresión de *rn*, ni el fenotipo es rescatado por inducir la expresión de *rn* en mutantes *tal*. Los mutantes nulos de *tal* provocan letalidad embrionaria, demostrando que es un gen esencial.

Figura 2: número y tipos de proteínas codificadas en el genoma de *Drosophila*.



Tomado de Lodish, Biología Celular y Molecular, 5ª edición, Editorial Médica Panamericana.

Existen cinco sORFs en el transcripto *tal*. De 5' a 3' son: 1A, 2A, 3A, AA, y B. Los sORFs tipo A incluyen un motivo conservado de siete aminoácidos (LDPTGXY), el cual se encuentra fuertemente conservado en genes homólogos a *tal* en otras especies, definiendo una antigua y no canónica familia génica eucariótica. sORFs 1A y 2A codifican un péptido idéntico de once aminoácidos. sORF 3A codifica otro péptido de once aminoácidos muy similar a 1A. sORF AA codifica un péptido de treinta y dos aminoácidos. sORF B codifica un péptido de cuarenta y nueve aminoácidos sin dominios conservados en otros insectos.

Para interpretar el rol de cada sORF, se realizó una construcción quimérica que sólo llevaba un sORF tipo A más un sORF B, lo cual fue completamente funcional. Otros dos constructos en los cuales se impidió la traducción del sORF B, ya sea por su delección o por mutación de su codón de comienzo de la traducción, dio como resultado una completa funcionalidad de *tal*. Asimismo, una construcción con un solo tipo de sORF A es completamente funcional. Con esto se desprende que un solo tipo de sORF A es suficiente para proveer la función de *tal*, y que sORF B no es necesario para dicha función.

Los péptidos de 11 aminoácidos codificados por *tal* provocan una señal celular que participa en el desarrollo de la pata de *Drosophila* (7). Esta señal alcanza 2 a 3 células es atribuida a la acción directa de los péptidos *tal*. Durante el desarrollo de la pata, *tal* señala una actividad que determina el límite del tarso, el cual regula la transcripción de varios genes, lo que indicaría que *tal* actúa de manera no autónoma. Esta señal producida por *tal* regula los genes *rotund* (*rn*) y *spineless* (*ss*), y crea un límite requerido para la intercalación, crecimiento y supervivencia de los segmentos dos, tres y cuatro del tarso. Estudios de pérdida de función y de expresión ectópica demuestran que *tal* es requerido de forma no autónoma para el desarrollo del tarso. A nivel molecular,

los péptidos codificados por el gen *tal* cambian la actividad de regulador transcripcional del gen *shavenbaby (svb)*. El locus *svb* codifica tres isoformas proteicas solapantes: *svb* y las proteínas específicas de la línea germinal *OvoA* y *OvoB*. Todas estas proteínas comparten el mismo sitio de reconocimiento del ADN y de activación transcripcional, pero difieren en su amino-terminal. La isoforma más corta, *OvoB*, es un activador transcripcional que induce la formación de tricomas. *OvoA* contiene una región amino-terminal extendida, lo cual cambia la función hacia un represor transcripcional, inhibiendo la formación del tricoma. *Svb* contiene una mayor extensión, comparado con *OvoA*, y promueve la formación de tricomas. *tal* no tiene efecto sobre la isoforma más corta (*OvoB*), sino que controla específicamente la capacidad de *Svb* de inducir tricomas, sin influir en su producción ni en su transporte al núcleo sino que modifica la actividad del producto génico, que pasa de ser un represor a un activador, controlando así la diferenciación epidérmica y de tricomas en *Drosophila* (8).

Homólogos a la familia *tal* se identificaron en artrópodos, conteniendo sORFs homólogos de entre 10 y 13 codones, los cuales son traducidos en péptidos que funcionan durante el desarrollo. Esos datos muestran que genes funcionales que codifican muy pequeños ORFs pueden existir en otros eucariotas superiores. Los sORFs no encajan con la definición estándar de gen, no son considerados en la anotación de genomas y caen, finalmente, en regiones consideradas como no informativas. Sin embargo, los sORFs podrían tener un rol en muchos procesos biológicos, modulando el flujo de información de una manera aún impredecible. Pero para esto necesitan ser identificados como una parte funcional del genoma, tal como hemos hecho en este trabajo.

Debido a su cuidadoso análisis genómico y la cantidad de información disponible, el de *Drosophila* es uno de los mejores genomas anotados y, aún así, una fracción importante

de sORFs podría haber escapado a la detección por parte de métodos de anotación que se han aplicado.

RAZONES PARA ANALIZAR A *DROSOPHILA MELANOGASTER*

Thomas Hunt Morgan introdujo el empleo de la mosca de la fruta, *Drosophila melanogaster*, en la investigación genética. Desde entonces se ha transformado en uno de los organismos modelo más usados en diversas disciplinas científicas. Entre las ventajas que ofrece su uso se destacan las siguientes:

- ✓ Fácil de criar, de manipular, requiere poco espacio y su costo es reducido.
- ✓ La embriogénesis se completa en 24 horas a 25 grados centígrados.
- ✓ El ciclo de vida es corto (9-10 días a 25° C).
- ✓ El genoma, completamente secuenciado, se resume en cuatro cromosomas.
- ✓ Se dispone de amplias colecciones de líneas mutantes accesibles a la comunidad científica.
- ✓ Es considerado el mejor modelo de eucariota superior para realizar estudios genómicos, transcriptómicos y proteómicos.

El complemento cromosómico de *Drosophila melanogaster* consta de 4 parejas de cromosomas: los cromosomas sexuales X e Y (cromosoma I) y 3 parejas de autosomas (II, III, IV). Los cromosomas Y y IV son pequeños y telocéntricos, siendo el cromosoma Y heterocromático y portador de algunos genes relacionados con la espermatogénesis. La mayoría de los genes se localizan en los cromosomas X, II y III, que son metacéntricos y grandes. Es en uno de estos cromosomas en los cuales se desarrolla este trabajo.

Su desarrollo (Tabla 1) incluye un período de embriogénesis dentro del huevo y una sucesión de estadios larvales que culminan con una metamorfosis completa (holometábola), de la que finalmente surge un adulto. El desarrollo se completa en 10 días a 25° C y 60 % de humedad relativa. Del huevo una pequeña larva (primer estadio), la cual muda al cabo de un día y llega al último estadio larval en el tercer día. Durante la metamorfosis, la hormona ecdisona desencadena una serie de cambios en el organismo entre los cuales se cuentan la destrucción de ciertos tejidos y órganos larvarios (histólisis) y la organización de las estructuras del adulto a partir de los discos imaginales y de células larvarias que se diferencian. El estado pupal lleva 3 a 5 días del ciclo hasta la emergencia del adulto.

Tabla 1: Cronología del desarrollo de *Drosophila melanogaster* a 25°C (tomada de la guía de trabajos prácticos de genética, FCNyM, UNLP)

Por día	Por hora	Fase
0	0	Huevo depositado
0-1	0-22	Embrión
1	22	Eclosión del huevo (primer estadio larval)
2	47	Primera muda (segundo estadio larval)
3	70	Segunda muda (tercer estadio larval)
5	118	Formación del pupario
5	122	Muda “prepupal” (cuarto estadio)
5.5	130	Pupa: eversión de cabeza, alas y patas
5.7	167	Pigmentación de ojos pupales
9	214	Adulto emerge del pupario con alas plegadas
	215	Las alas se expanden al tamaño adulto

HIPÓTESIS

1. El genoma de *Drosophila melanogaster* tiene genes no identificados aún en regiones consideradas no codificantes.
2. Los organismos eucariontes expresan transcritos policistrónicos de una zona más amplia que lo esperado.

OBJETIVO

Analizar el genoma del cromosoma 2 de *Drosophila melanogaster* en busca de sORFs codificados en ARNs policistrónicos.

MATERIALES Y MÉTODOS

HERRAMIENTAS BIOINFORMÁTICAS

BASES DE DATOS

La secuencia correspondiente al cromosoma 2 de *Drosophila melanogaster* se obtuvo por FTP a partir de las bases de datos públicas (National Center for Biotechnological Information, NCBI, Flybase y Ensembl). Los mismos fueron obtenidos en formato Fasta y trabajados con formato de texto usando un algoritmo (no publicado) desarrollado en la base Wolfram Mathematica por el Dr. Luis Diambra del Laboratorio de Biología de Sistemas, Centro Regional de Estudios Genómicos, Universidad Nacional de La Plata y cedido gentilmente para nuestros análisis. Se detalla en la sección de resultados el diagrama de flujo de análisis y los datos obtenidos.

BLAST

El programa de computación empleado con el propósito de comparar secuencias es BLAST (basic local alignment search tool). El algoritmo BLAST divide la secuencia de la nueva proteína o gen (secuencia consultada o *query*) en segmentos más cortos y luego busca en la base de datos coincidencias significativas con cualquiera de las secuencias almacenadas. El programa de coincidencias asigna un puntaje alto a los aminoácidos que coinciden idénticamente y un puntaje más bajo a coincidencias entre aminoácidos que son relacionados (por ejemplo, hidrófobos, polares). Cuando se encuentra una coincidencia significativa para un segmento, el algoritmo BLAST buscará localmente para extender la región de similitud. Una vez completada la búsqueda, el programa clasifica las coincidencias entre la secuencia de averiguación y diversas proteínas

conocidas de acuerdo con sus valores p . Este parámetro es una medida de la probabilidad de hallar ese grado de similitud entre dos secuencias de proteínas por casualidad. Mientras más bajo es el valor p , mayor es la similitud de secuencias entre dos secuencias. Un valor menor de 10^{-3} es una evidencia significativa de que dos proteínas comparten un ancestro común.

En este trabajo utilizamos específicamente el programa tBLASTn, el cual busca coincidencias entre la secuencia peptídica ingresada y una base de datos de nucleótidos, cuyas secuencias son traducidas en los 3 marcos de lectura posibles sobre ambas cadenas del ADN dúplex. Esta búsqueda se realizó utilizando los parámetros que ofrece el programa por default, con un valor $E=10^{-3}$.

AUGUSTUS

Augustus Gene Prediction es una herramienta de predicción génica eucariótica que emplea un método adecuado para modelar la distribución en longitud de exones e intrones. Provee información sobre las secuencias nucleotídicas de los exones e intrones, y sobre las secuencias peptídicas. Puede ser utilizado a través de una interface web (<http://augustus.gobics.de/>) o descargado y utilizado localmente.

RESULTADOS

Por medio de una aproximación bioinformática se identificaron sORFs en el cromosoma 2 de *Drosophila melanogaster* utilizando un algoritmo que detecta sORFs policistrónicos que codifican péptidos menores a 5 kilodaltons (ver métodos). El número total de genes con potencial ORF corresponde a varios miles. Para acotar la búsqueda a un número menor, se limitó el análisis a aquellos sORFs que se encontraran en grupos de secuencia similar y menos de 2 kilobases de distancia entre cada uno. El cromosoma 2e de *Drosophila* se dividió entonces en segmentos que contenían 242 clusters de sORFs. Este análisis se acopló a un algoritmo que realiza una vinculación de la masa teórica de dichos péptidos a bancos de datos públicos de espectros de masas (www.flybase.org). Aquellos péptidos para los cuales se obtuvieron evidencias de codificación de los sORFs con espectros de masa en dichos bancos se utilizaron para realizar una búsqueda contra bases de datos de nucleótidos traducidos (tBLASTn). Los péptidos para los que no se encontraron resultados en las bases de datos (158 sORFs) fueron tomados para analizar su secuencia génica con el software Augustus. Este software determina la posible estructura del gen (su organización en exones e intrones) y predice las secuencias de aminoácidos. Dichas secuencias predichas son analizadas utilizando FlyBase para poder determinar si dichos genes estaban previamente anotados. Esto permitió descartar una serie de genes que poseían sORFs, pero dentro de ORFs más grandes. Esto no significa que estos sORFs no fueran funcionales (podrían funcionar como genes-dentro-de-genes), sino que representa un criterio de simplificación de la búsqueda. Los genes que evidenciaron no haber sido anotados previamente y que poseen una estructura génica coherente predicha por el software Augustus (33 sORFs), fueron evaluados teniendo en cuenta los pequeños péptidos

codificados por los diferentes sORFs, su evidencia en espectros de masas, y su evidencia de expresión a partir de bases de datos de ESTs (etiquetas de secuencia expresada, del inglés *Expressed Sequence Tags*) obtenida también de bases de datos públicas (NCBI: www.ncbi.nlm.nih.gov). Estos datos permitieron evaluar la transcripción y traducción de los sORFs, es decir su expresión. Se toman en cuenta los genes no anotados porque el objetivo de este trabajo es descubrir nuevos sORFs. Los nuevos genes finalmente anotados fueron 48. Los genes no anotados o aquellos sin estructura predecible fueron 77. La ausencia de predicción de estos últimos no implica que no sean expresados, pero en ese caso podrían considerarse como ncRNAs. En los párrafos subsiguientes se describen los nuevos genes encontrados en el cromosoma 2 clasificados en aquellos que muestran evidencia de transcripción y traducción y aquellos que evidencian una u otra.

Figura 3: Diagrama de flujo del análisis de anotación genómica del cromosoma 2 de *D. melanogaster*.

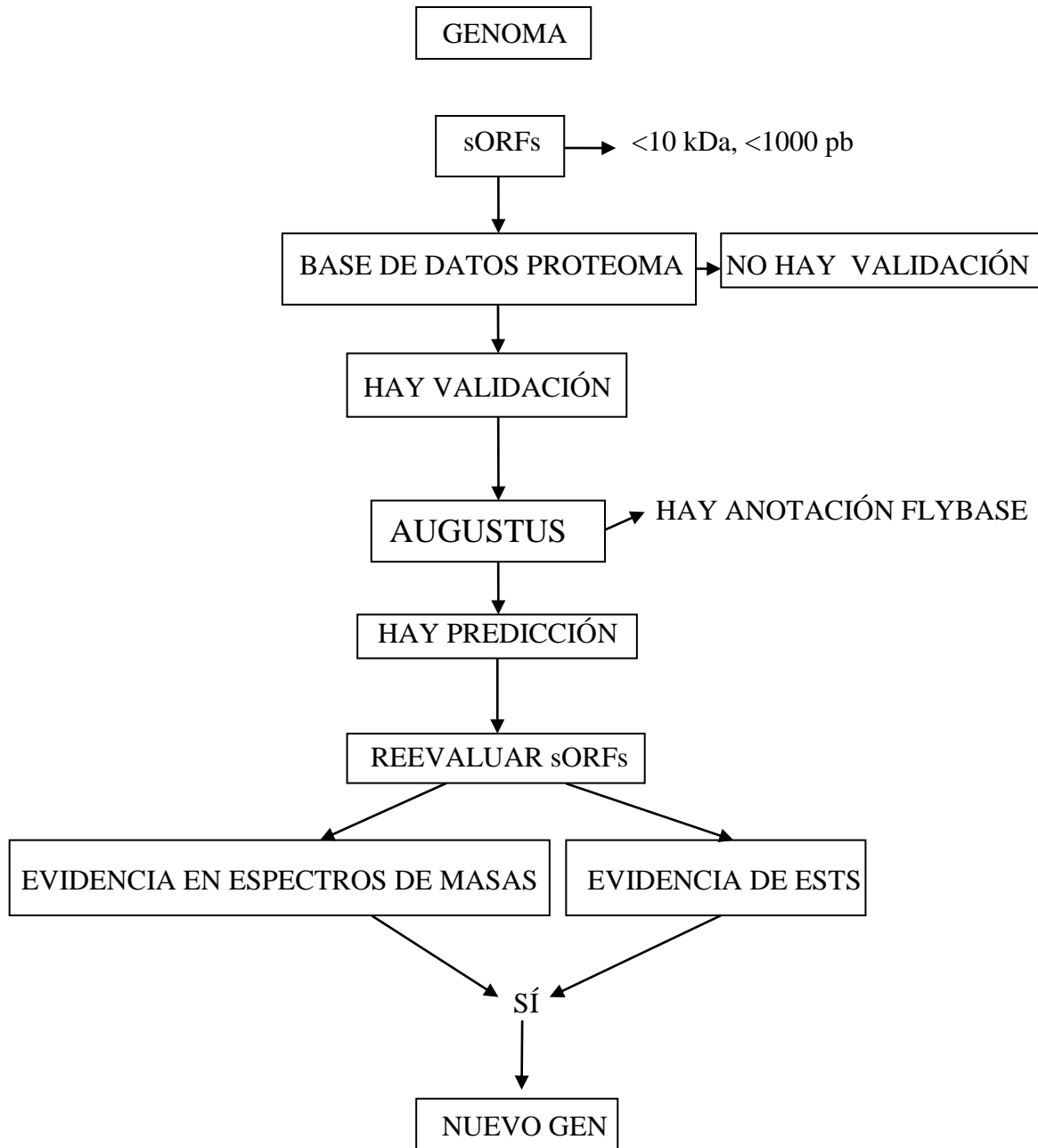


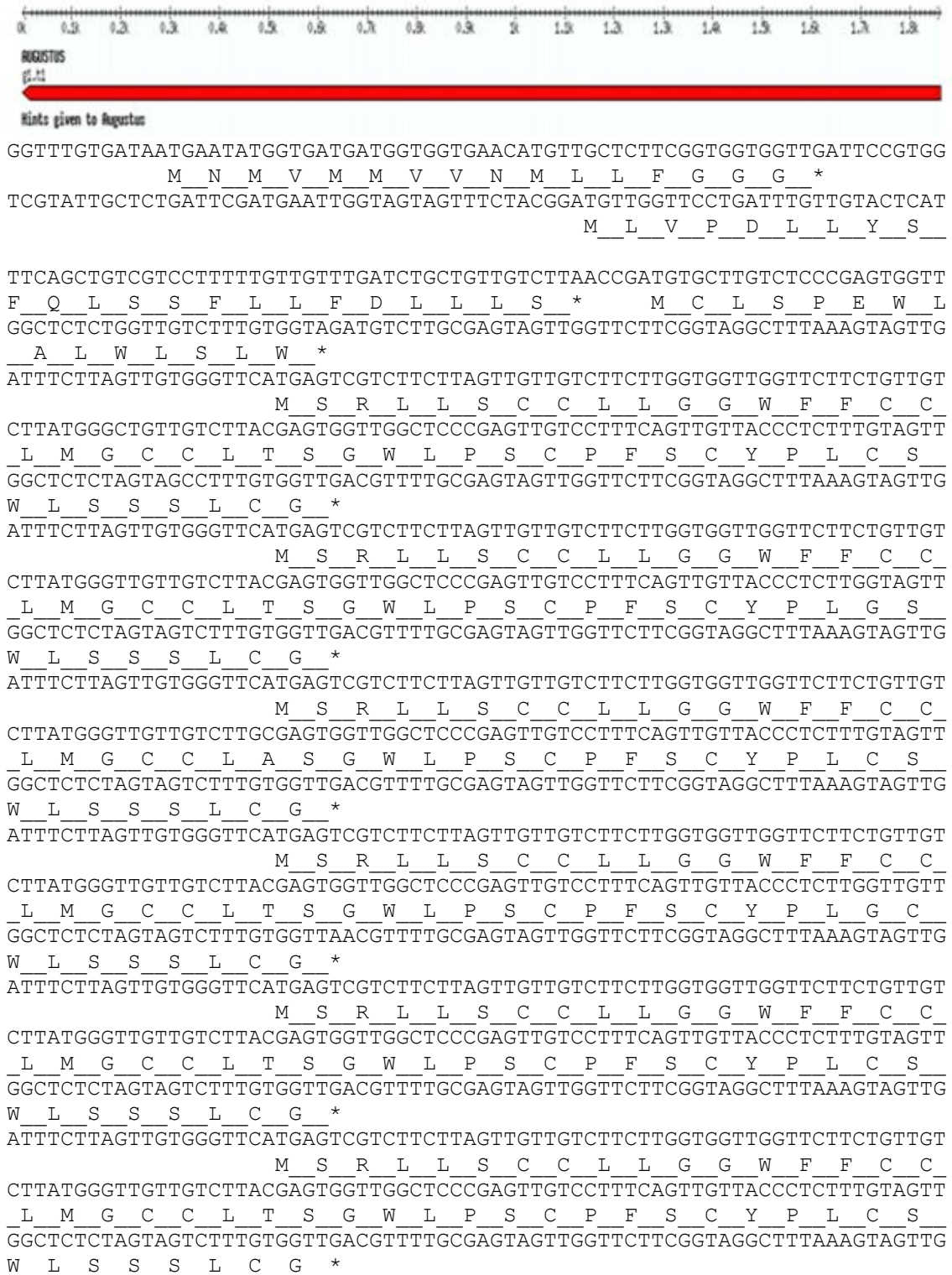
Tabla 2. Presenta los resultados obtenidos a partir de la búsqueda bioinformática descrita anteriormente:

BRAZO CROMOSÓMICO	CLUSTERS DE sORFs	CLUSTERS SELECCIONADOS	GENES ANOTADOS	GENES NO ANOTADOS	
				CON POTENCIAL ESTRUCTURAL (AUGUSTUS)	SIN ESTRUCTURA PREDECIBLE (AUGUSTUS)
2L	132	83	23	17	43
2R	110	75	25	16	34

ANÁLISIS DE NUEVOS GENES CORRESPONDIENTES A REGIONES GENÓMICAS NO ANOTADAS Y CON ESTRUCTURA PREDICHA

1. REGIONES GENÓMICAS CON VALIDACIÓN TANTO POR ESTS COMO POR ESPECTROMETRÍA DE MASAS

Gen 2L_3



```

ATTCTTAGTTGTGGGTTTCATGAGTCGTCTTCTTAGTTGTTGTCTTCTTGGTGGTTGGTTCTTCTGTTGT
      M S R L L S C C L L G G W F F C C
CTTATGGGTTGTTGCTTACGAGTGGTTGGCTCCCGAGTTGTCCTTTCAGTTGTTACCCCTCTTTGTAGTT
  L M G C C L T S G W L P S C P F S C Y P L C S
GGCTCTCTAGTAGTCTTTGTGGTTGACGTTTTCGCGAGTAGTTGGTTCTTCGGTAGGCTTTAAAGTAGTTG
W L S S S L C G *
ATTCTTAGTTGTGGGTTATGAGTCGTCTTCTTAGTTGTTGTCTTATTGGTGGTTGGTTCTTCTGTTGTG
      M S R L L S C C L I G G W F F C C
TTATGGGTTGTTGCTTACGAGTGGTTGGCTCCTTTGTTG
L M G C C L T S G W L L C T G *
    
```

Para la región genómica 2L_3, el programa Augustus Gene Prediction predijo una estructura compuesta por un exón de 1861 pares de bases.

Para este región , se encontró acuerdo con espectros de masas para el péptido NMVMMVVNMLLFGGG con una probabilidad de $\rho=0.005$ y un valor $E=0.000900377925286973$.

Evidencia de EST:

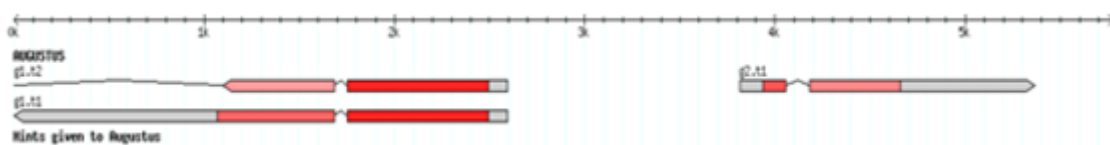
gi|3946654|gb|AI297247.1|:

```

CTCTAGTAGTCTTTGTGGTTGACGTTTTGCGAGTAGTTGGTTCTTCGGTAGGCTTTAAAGTAGTTGATTT
CTTAGTTGTGGGTTTCATGAGTCGTCTTCTTAGTTGTTGTCTTCTTGGTGGTTGGTTCTTCTGTTGTCTTA
TGGGTTGTTGCTTACGAGTGGTTGGCTCCCGAGTTGTCCTTTCAGTTGTTACCCCTCTTGGTTGTTGGCT
CTCTAGTAGTCTTTGTGGTTGACGTTTTGCGAGTAGTTGGTTCTTCGGTAGGCTTTAAAGTAGTTGATTT
CTTAGTTGTGGGTTTCATGAGTCGTCTTCTTAGTTGTTGTCTTATTGGTGGTTGGTTCTTCTGTTGTCTTA
TGGGTTGTTGCTTACGAGTGGTTGGCTCCTTTGTTGTTGTTCTTCTGTTGTTTTCCCTTTTGTAGTTGGCT
CTCTAGTAGTCTTTGTGGTTGACGTTTTGCGAGTAGTTGGTTCTTCGGTAGGCTTTAAAGTAGTTGATTT
TTTAGTTGTGGGTTTCATGCGTTGTCTTTTTGTTGGTTGGTTCTTCTGTTGTCTTATGGGTTGTTGTCTTA
CGAGTGGTTGGCTCCT
    
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 4 sORFs seguidos de una repetición en tándem de 7 sORFs idénticos, lo que podría deberse a un fenómeno de duplicación múltiple. La transcripción se basa en la presencia de un EST en la base de datos pública y la probable traducción en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORFs codificados por este transcripto es traducido. La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_4



AGCTCAGCAGCCGTTCCCGTCTCCGGCGAAAGGACCTACCAACCCAGCGAATAAATGATTAGGGCCGTTT
AATCATAGAACAGTTGGAAAAATAATACAACCTGTTCAAAAAATACTTGTATAGTTAAGATTTTTAAACT
TATTGTTAGTTCTTATACAAGAAGATTCAATAAAATAAAGCAAAGTAAAAAAAAAAAAAAAAAAAAAA
AATTCGGTTTCAGCGGGAAGATGAGCACTAAATAAATGCCGAAAATGCAAGCGGAGTCAATGCACGTAACG
ATATATGTAGCAATGTATGTATGTATATATGTAATGCTAGCTCGCTTAAACACAGCCACTATCACTTGTC
M_Y_I_C_N_A_S_S_L_K_H_S_H_Y_H_L_S
CAACCGATTGCGATTTAATGTATTTATATTTATATAAATGCATGCAAATTAACACTATCACGTCGGGG
N_R_L_R_F_N_V_F_I_F_I_Y_K_C_M_Q_I_N_T_I_T_S_G
TCACCAATATTTAGCTATTGAGCTTTTTCAATAGCTGTTGCGAATAATAAATAGCCGAAGAAAGTTTGT
S_P_I_F_S_Y_*
TATTTTACGATACTGTTTATTTTCGCGAGTTGTTTACAGCAGCTAGGGCCGACAAAGAGCTATATCGAATG
M
CACAAAGTTTAGTCTCTTCCGACAGACTAAGAAGTGGCAATTGGCGGGACAGACAGCCGAAAAACGCAGTG
H_K_F_S_L_F_R_Q_T_K_K_W_Q_L_A_G_Q_T_A_E_K_R_S
CCCAAGCTTAACGTAGGCTTGGCTTCCAACAGAGTCTCAAGGGGAAGCAGACACAGCTTGTTCAGTCC
A_Q_A_*
CGCCTTATTACTCCAGACGCGTCTCAATTCTGCAACTCGAGCGTTGCCGTCTCTGCCAGGAATCAACT
GCATGACTCTCGCAAAGGCCATCTCATTGGGGGTAGATTCTCGTCTTAACAGGAACGACGTTGTCCAC
GGCTACGCCAGGCTTAGGGGTGCGCCAATTGGAGCGCTGCTGGAGTACTCTTCCCTCCATCGCGGCCAAA
ATATATGCTGAAGAAAGGAGATGCGCTTGACTCTTGGCAAGCCGATTATAGTTTGGCCCGTTATATCTG
M_L_K_K_G_D_A_L_D_S_W_Q_A_D_Y_S_L_G_P_L_Y_L
M_R_L_T_L_G_K_P_I_I_V_*
GCTCGTCAAACGACGAAGGCGGACCACCATTGAGAAAAATGCGCCGGAGTGAGGACGTCCAGATCGGCAGG
A_R_Q_T_T_K_A_D_H_H_* M_R_R_S_E_D_V_Q_I_G_R
GTTCTCTGAAATGGGAATAAAGGTGATGCTCGCTGCTGAGAAATAACCTTCGAAGTTCAGAAGTTCAT
V_L_* M_G_T_K_G_D_A_R_C_*
TTTTGGCGCCGACAAAGTTGGTGGCGTTGTCTGACCGAAGTTGCTTCGGCTTCCGCCGGGTGCATATGAA
M_N
CCTCTTGAGTCCGCACAGAAACACAACCTGTGAGAGATCCTTGATCAGCTCCAGGTGCAGTGCCTTGGTT
L_L_S_P_H_R_N_T_T_V_E_R_S_L_I_S_S_R_C_T_A_L_V
GCAAAACATACAAATACGCAGACGTAGCATTTAACC GCGGGCTTGTGCGAGTATCCGACTTGTGAAAGA
A_K_H_T_N_T_Q_T_*
AGGGTCCACAGAAGTCTATACCAGCTACCTCAAAAGCGTGAGATCCTTCCAAGCGCTCATTGGGAAGGTC
CGCCATTATGTGCTCTATCAGCCGCGGCTTAATCCGAAAACATCTGATGCACCTGTTCCAGGCTTGGTA
M_C_S_I_S_R_G_L_I_R_K_H_L_M_H_L_F_T_A_L_V
ACCGTCTTCTCCCCCAAAAGGCCAATATTGGGATCGAATTGCTCCCAGAAGAGCTCGAGGTCCGGCAT
T_V_F_L_P_P_K_G_Q_Y_W_D_R_I_A_P_R_R_A_R_G_P_A
M
GAAGATTGCTTTCATGGTAATGCGAAATAATTGCCAGAGTCACCAATGACGAATTACGGAGGCGACCGCC
*
K_I_A_F_M_V_M_R_N_N_C_Q_S_H_Q_*
M_T_N_Y_G_G_D_R
TACTCTAAGAAGTCCAAATTGATCCAGGAATGGCGAGAGCGAGGATATGGGAGTAGACGAGGAAACTCTG
L_L_* M_G_V_D_E_E_T_L_H_Y_G_K_A_C
CACTATGGCAAAGCATGCAACCAAGAGAATTTTGGGTGGTAGCAACTAGCGTCAACGATTTTTGCCAC
N_Q_E_N_F_G_W_*
M_A_K_H_A_T_K_R_I_L_G_G_S_N_*
M_Q_P_R_E_F_W_V_V_A_T_S_V_E_R_F_L_P_T
CTGAACGCCTGGCGCATAGGTTGCCATGGCGCAATCCACGGCTCCAAAGTCTACATCGCTGCTCCAGG
*
AATGCAGCCATCGATTCCCAGACGGAATAAGGTTGGCAAAGACCGAATCCTCTAAGCGCTCCTCCCCT
M_Q_P_S_I_P_T_T_E_*
TAGCTTGGCTCGCCGATCCAGCTTTTGCAGCAGCACTTGCAGTATGATGCAGCCAGCGATTTGCTCAGT
GGTGCCCAACCCCTTCATCGCACGAATGTGAGCGTTAAACTTGTCCGACAATTCCCGAAGCGATGCCACT
M_P_L
GAACCATTCGGTACTACCTTTAAACCCAGAATCTCGGTGATGTGCGCCTGAAAAACGAAACGCCGATTAT
N_H_S_V_L_P_L_N_P_E_S_R_*
CAAACCTTTTTGAAGCAACTCTAAAGCCGCTTCGTAATTGCTGTTTGAATCTCCAATGATCGGATAGT
TTCCAGCGCTGAATCCCTCAAACATGACCGCAGATGTTGAAATTTCTCAATGTTGGAGAGGTCCGGATGG
M_T_A_D_V_E_I_S_Q_C_W_R_G_P_D_G_C_R_L_C_S_*

CTGTGCGATTATGCTCGTGAACACGGAGTAAAAATCCGTCCAGTTTGCCTAGCCTCCGCCAAACGTTGGCA
M L E R S G W
L S I M L V N T E *
GCTGCACAGGCGGCAACGGCCTCGGGCTCGGCCGCGGCTGCGTTTGGAGACCCTACACTTGTGGGGCAC
AACACCTTCCGCAAGCGTTAAGCGCGGCGGAAAATGCACATTGCGTTTGGCTATTTTCGAGCGCACACTA
GCCTTGAATTCAACGATGAATTCATCGAAAACTCGCTCATTTCACCTCTCGTCGAAATCCAATTCCTCCA
GTTCTGTATGCAGAACATTGAAGGCATTTGCAGCTCATCAATCTGCTCCAGCCTCACCGTAAGAGTGGG
TTCATCGTATCCGCTTAGCGTCTCCGCTTAGCGTCTCATGCGCTGAACAGCGCATTGGCTTTGCGCTTTA
AAAAGGTCACCTGTTTTTGTACCTTCAGCGGGCATAGCAACAAATCGCCTTTAATTCGGTTTCAGCGGG
AAGATGAGCACAAAATAAATGCCGAAAATGCAAGCGGAGTCAATGCACGTAACGATATATGTAGCAATGT
ATGTATGTATATATGTAATGCTAGCTCGCTTAAACACAGCCACTATCACTTGTCCAACCGATTGCGATTT
M Y I C N A S S L K H S H Y H L S N R L R F
AATGTATTTTATATTTTATTTATAAATGCGATGCAAATTAACACTATCACGTCGGGGTACCATAATGTTTAGCT
N V F I F I Y K C M Q I N T I T S G S P M F S
ATTGAGCTTTTTCAATAGCTGTTGCGAATAATAAATAGCCGAAGAAAGTTTGTGTTGTTGTTGCCAA
Y *
AAGGAAAACAGCTTTATAAGACTATAATAAAAAGCCACTAATAAACTTGACCATGTTTCCCCGGGAGTCC
M F P P G V
TACATACTGGCCATCAGACCTCAATAAGCTGCCAGACCTGATCGACTTCGCAGTTACGAAAAATATTTCC
L H T G H Q T S I S C Q T *
CGCAGTTTGGTTAAAGCTGAATGTCTGCCGGATCTGTCTATCTGATCACTCGCCTGTACTAATTCACCTCC
GCCGGTACGCAGAAAACGTGAAACCACCAATCAGATTGACCTCTAGCAAAAACAACTGGCTCAGGTATAA
AAAATATATAAGTTACATATTTAGCTAAGCCAAAACCTCAATACTGAATCTGATATAGAGAGCTGCACG
TGTGCATTGCAATCCATCCTTACTGCAGCAGCTCTTACTGCAACACCCAAAATAACAAATAATACAATTA
ATTCAAAAAGACCAACGTACAAAATCGAGCAACTCGTCCACGTAAAACGTGCTTACGCAGAGAATGGCA
ATCTTCCAGATCCCCAACTGCAAAAACAAAAGCTAAAAGCAGCCACACGGAAAACGGCCAAACGCTTGAAA
CAAGAAGAGGACGATCAGCGCCGATACATAGCAACTCACACCAACAGGCACAAAACAAAAGTACAC
TGTGGCGAGCCACTCAACTCTTCGCCACCCGACTGAAACCGTTTTTGGCGATAAGGAATTCCTCAGGTGG
CTGGGCCCGTAGTGATGAAGACAGAGCCACCACATTTGCCGCTCACCTACAAAATGTGTTACGCCAAAC
M K T E P P H L P L T Y K M C S R Q T
CAGGCTACTAGCACATTCGCGCTACCGTCTTATCCCGTAAACCGCCATCAGCAACACACCCCAATTGTGT
R L L A H S R Y R P I P *
TTCGTCCTAAAGAAAATAACTAAAATAATCAAAGACAATCTCAGCCCGAAAAATCGCCCAGCTGCGACCT
TATAACACCGAAAATGATCATCCAGCTGCCACATTTCTGCATTTTCGCTACATAACCAAGCTCTTTAATACC
ATCACTAAACTTGTTACTTTCCACAACGATGGAAGAGGTGATTATCATAATGATTCCAAAGCCTGGTA
M E E V D Y H N D S K A W *
AGAACCACACAGTCGCTTCATCTTACAGACCAATAAGTCTACTCTCATGCAATTTGAAACTATTCGAAAA
M H F E T I R K
ATGCCTGCTGATCCGACTTAATCAACATCTGATATATCACAATATAATCCCAGCCACCAATTTGGATTT
M P A D P T *
CGCGAAAGCCATGGAACCATTGAACAGGTGAATCGTATTACAACGGAAATAAGAAGTGCATTTGAATATC
GCGAATACTGTACAGCAGTATTTTTAGACGTATCCCAAGCATTCGGCAGAGTCTGGCTCGACGGCCTAAT
GTTTAAAATTAACAATCCCTACCCGAAAAGCACACAAAATCTTAAAGTCTGACTCTATGACAGAAAAG
TTTGCAAGTCCGCTGCAACACTGCCACTTCCACTGATCATAACAATGAGGCTGGAGTCCCCAAGGCAGCG
TTCTTGGGCGCACTTATACCTCATCTATACAGCCGACATCCCTACAAATAGTCGCACATATCCAGCAATTT
GCCGACGATACAGCTATCCTTAGCCGTTCAAGGTCCCCATCCAAGCTACAGCACAGTTGGCACTGTACC
TCATCGACATTGAGAAGTGGTCTCTGACTGGCGAATAAAAAGTAAACGAGCAAAAATGCAAGCACGTGAC
M Q A R D
GGTTACGCTAAACAGACAAGACTGTCCCTCCGCTCTTATTGAACAACATAACCACTCCCAGAAAGCAGATGAG
G Y A K Q T R L S S A L I E Q H T T P E S R *
GTAGCGTACCTAGGAGTACACCTTGACAGAAGACTCACATGGCGCAGGCACATTGAAGCCAAAAAACCC
AACTTAAACTCAAAGCCAACAACCTTACAGTGGCTCATCAACTCTGGTCTCCGCTCAGCCTAGATCACAA
GGTCTTGCTCTACAATTTATATTTGAAACCAATCTGGACCTATGGCTCACAGTTATGGGGGCAATGCCAG
M G A M P
CAACAGCAATATTGACATCATTGAGCGAGCACAATCAAAGATTCTGAGAACCATCACTGGGGCACCGTGG
A T A I L T S F S E H N Q R F *
TACGTTTCGGAGTGAAAAACATCCAAAAGAGACTTAAATATCCCATCAGTTACCAACGCAATCACGGAACCTA
AGGAAAAATACCATAGCAAGCTTCACACGCACCCCAACCACCTAGCGCGAGGTCTAATCCAGCTCAGCAG
CCGTTCCCGTCTCCGGCGAAAGGACCTACCAACCCAGCGAATAAATGATTAGGGCCGTTTAAATCATAGAA
CAGTTGAAAAATAATACAACCTGTTCAAAAAATACTTGTATAGTTAAGATTTTTAACTTATTGTTAGT
TCTTATACAAGAAGATTCAATAAATAAAAAGCAAAGTAAAAAAAAAAAAAAAAAAAAAAAAAATTCGGTTC

AGCGGGAAGATGAGCACTAAATAAATGCCGAAAATGCAAGCGGAGTCAATGCACGTAACGATATATGTAG
 CAATGTATGTATGTATATATATGTAATGCTAGCTCGCTTAAACACAGCCACTATCACTTGTCCAACCGATTG
M Y I C N A S S L K H S H Y H L S N R L
 CGATTTAATGTATTTATATTTTATTTATAAATGCATGCAAATTAACACTATCACGTGCGGGGTCACCAATAT
R F N V F I F I Y K C M Q I N T I T S G S P I
 TTAGCTATTGAGCTTTTCAATAGCTGTTGCGAATAATAAATAGCCGAAGAAAGTTTGTTTATTTTACGA
 F S Y *

TACTGTTTATTTTCGCGAGTTGTTTACAGCAGCTAGGGCCGACAAAAGAGCTATATCGAATGCACAAGTTTA
M H K F
 GTCTCTTCCGACAGACTAAGAAAGTGGCAATTGG
 S L F R Q T K K W Q L G *

Para el cluster de sORFs 2L_4, el programa Augustus Gene Prediction predijo para el gen 1, transcripto 1.1, una estructura de 2 exones, de los cuales uno contiene 844 pares de bases y el otro 1682 pares de bases, y un intrón de 69 pares de bases. Para el gen 1, transcripto 1.2, predijo una estructura de 2 exones, uno de 844 pares de bases, y el otro de 583 pares de bases, y un intrón de 69 pares de bases. Para el gen 2.1 predijo una estructura de 2 exones, uno de 1508 pares de bases y, el otro de 241 pares de bases, y un intrón de 127 pares de bases. La estructura indica una secuencia repetitiva de

Se encontró acuerdo con espectro de masas para el péptido KWQLAGQTAEK con una probabilidad de $\rho=0.068$ y un valor $E=0.000285737204312655$, y para el péptido HLMHLFTALVTVFLPPKGQYWDR con una probabilidad de $\rho=0.06$ y un valor $E=0.000305832558931141$.

Evidencia de EST:

gi|133838108|gb|EL871737.1|:

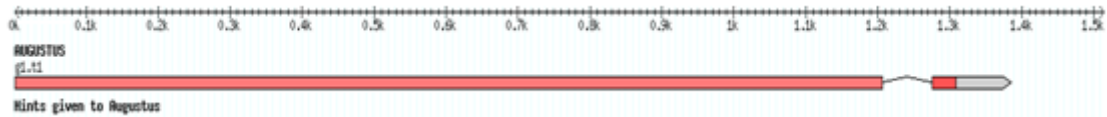
TGCATTGCCTTAGCTGATGACCTGCCTTTAGGCAATTTAGGCAAAGTGCTAGTCTCTTGAGTGCATGCAA
 CCAAGAGAATTTTGGGTGGTAGCAGCTAGCGTCAACGATTTCTGCCCACCTGAACGCCTGGCGCATAGG
 TTGCCATGGCGCAATCCCCGGCCTCCAAAGTCTACATCGCTGCTCCAGGAATGCAGCCAACGATTCCCA
 CGACGGAATAAGGTTGGCAAAGACCGGATCCTCCAAGCGCTCTCCCACTTAGCTTGGCTCGACGCATCC
 AGCTTTTGAGCAGCACTTGCACTATGATGCAGCCAGCGATTTGCTCAGTGGTGCCCAAACCCCTCAATG
 CACGAATGTGATCGTTAACTTGTCGACAAATCCCCGAAGCGATGCCACTGAACCATTCGGTACTACCTT
 TAAACCCAGAATCTCGGTTATGTGCGCCTGAAAAACGAGACGCCGATTATCAAACCTTTTTGAAGCAACT
 CTAAAGCCGCTTCGTAATTGCTGTTTGAATCTCAAATGATCGGATAGTTTCCAGCGCTGAATTCCTCAA
 ACATGACCTCAGATGTTGAAATGTCTCAATGTTGGAGAGGTCGGGATGGCTGTCGATTATGCTCGTGAAC
 ACGGATTAATAATCCGCCAGTTTGCCTAGCCTCCGCCAAACGTTGGCAGCTGCACAGGCGGCAACGGCA
 TCGGCCTCGGCCCGGCTGCTTTGAGGTCCACTACACTGGTGGGGCACAACACCTCCGCAAGCGTTGA
 GTGCGACGCGAAATGCACATTGCGTTTGGCTATTTCCGAGCGCACACTAGCCTTGAATTC AACGATGAAT
 TCATCGAAACTATCGCTCATCTCACTGCCAATTTTCGTCGAAATCCAATTC

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 26 sORFs. El sORF 18 es una duplicación desplazada del sORF 1. La transcripción se basa en la presencia de un EST en la base de datos pública y la probable traducción en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORFs codificados por este transcripto es traducido. En el caso del primer péptido, corresponde al marco de lectura

AAGTGGCAATTGGCGGGACAGACAGCCGAAAAA, donde dicho péptido se encuentra entre dos sitios de dos aminoácidos básicos, que es el sitio de corte de la tripsina usada para la digestión de muestras proteicas por espectrometría de masas.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_67



AAGGAACTGCTTCCAAAGGACCAGTTCAAGTTGGAGAAGTTATTCGAACGAAGGCACCGTATCTGGCGG
 ACTTGATATGATTCAAATGAAATCCAAAAATTTAATCAAGGTGGAATATACATTCCACATGTAGGCCAA
M_I_Q_M_K_S_K_N_L_I_K_V_E_Y_T_F_H_M_*
 GGAAATAGTGACAATAGATATTTATCCAAATCCGAAATCGGAGCAGTTCCAGTCGGAAATCCTATTCCGA
 ATGAAGGCAATGCATTTGGCGGTCGTGGATATGATTTAATGAAATCCAAAAACATAATCAAGGTGGGAG
M_K_A_M_H_L_A_V_V_D_M_I_L_M_K_S_K_N_I_I_K_V_G
 ATACATTCCACATGTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAACCTCTTAAAATCGGAGGAAAT
D_T_F_H_M_* M_E_M_V_P_M_H_L_K_N_L_L_K_S_E_E_I
 ATTCCAAATGTAGGCCAAGGAAACAGGGTAATAGATATGATCTAAATCCTGAAATCGGAGCAGTTCCAG
F_Q_M_*
 TCGGAAATCCTATTCCGAATGAAGGCAATGCATTTGGCGGTCGTGGATATGATTTAAATGAAATCCAAAA
M_K_A_M_H_L_A_V_V_D_M_I_* M_K_S_K
 ACATAATCAAGGTGGAAGATACATTCCACATGTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAACTT
N_I_I_K_V_E_D_T_F_H_M_* M_E_M_V_P_M_H_L_K_N_F
 CTTAAAATCGGAGGAAATATTCCAAATGTAGGCCAAGGAAACAGAGGGAATATATATGATCCAAATCCGG
L_K_S_E_E_I_F_Q_M_* M_I_Q_I_R
 AAATCGGAGCAGTTCCAGTCGGAAATCCTATTCCGAATGAAGGCAATGCATTTGGCGGTCGTGGATATGA
K_S_E_Q_F_Q_S_E_I_L_F_R_M_K_A_M_H_L_A_V_V_D_M
 TTATTTAAATGAAATCCAAAAACATAATCAAGGTGGAAGATACATTCCACATGTAGGTCATGGAAATGGT
I_* M_K_S_K_N_I_I_K_V_E_D_T_F_H_M_* M_E_M_V
 CCAATGCACCTAAAGAACCTCTTAAAATCGGAGGAAATATTCCAAATGTAGGCCAAGGAAACAGAGGGA
P_M_H_L_K_N_L_L_K_S_E_E_I_F_Q_M_*
 ATATATATGATCCAAATCCGAAATCGGAGCAGTTCCAGTCGGAAATCCTATTCCGAATGAAGGCAATGC
M_I_Q_I_R_K_S_E_Q_F_Q_S_E_I_L_F_R_M_K_A_M
 ATTTGGCGGTCGTGGATATGATTTAAATGAAATCCAAAAACATAATCAAGGTGGAAGATACATTCCACAT
H_L_A_V_V_D_M_I_* M_K_S_K_N_I_I_K_V_E_D_T_F_H_M
 GTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAACCTCTTAAAATCGGAGGAAATATTCCAAATGTAG
_* M_E_M_V_P_M_H_L_K_N_L_L_K_S_E_E_I_F_Q_M_*
 GCCAAGGAAACAGAGGGAATAGATATGATCCAAATCCGAAATCGGAGCAGTTCCAGTCGGAAATCCTAT
M_I_Q_I_R_K_S_E_Q_F_Q_S_E_I_L
 TCCGAATGAAGGCAATGCATTTGGCGGTCGTGGATATGATTTAAATGAAATCCAAAAACATAATCAAGGT
F_R_M_K_A_M_H_L_A_V_V_D_M_I_* M_K_S_K_N_I_I_K_V
 GGAAGATACATTCCACATGTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAACCTCTTAAAATCGGAG
E_D_T_F_H_M_* M_E_M_V_P_M_H_L_K_N_L_L_K_S_E
 GAAATATTCCAAATGTAGGCCAAGGAAACAGAGGTAATAGATATGATCTAAATCCTGAAATCGGACATGC
E_I_F_Q_M_* M
 CACATTTGGTGCACCCGGTACCGGAATGCAAGGAAATAATATAGATAATCCATATAAAAAACAATCTTGGT
P_H_L_V_Q_P_V_P_E_C_K_E_I_I_*
 GACAGAGCAGAAAATGAATTGGAAGGTTTCCTCTTGAAAATCCCAATAGGGATCATGTTATCAGGCTTA
M_N_W_K_G_F_L_L_K_I_P_I_G_I_M_L_S_G_L
 GTGATTTTTTAATCACAACGCCACTCCGAGTGCACCTATAAAATT
V_I_F_*

Para la región genómica 2L_67, el programa Augustus Gene Prediction predijo una estructura compuesta por 2 exones, uno de 1205 pares de bases, y el otro de 108 pares de bases, y un intrón de 60 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido PHLVQPVPECKEII con una probabilidad de $p=0.037$ y un valor $E=4.1569837010652e-05$.

Evidencia de ESTs:

gi|103649946|gb|EC051174.1|;

```
ATGCACCAAATGTGGCATGTCCGATTTTCAGGATTTAGATCATATCTATTACCTCTGTTTCCTTGGCCTAC
ATTTGGAATATTTCCCTCCGATTTTAAGAGGTTCTTTAGGTGCATTGGGACCATTTCCATGACCTACATGT
GGAATGTATCTTCCACCTTGATTATGTTTTGGATTTTCATTTAAATCATATCCACGACCGCCAAATGCAT
TGCCATTCATTCGGAATAGGATTTCCGACTGGAAGTCTCCGATTTCCGGATTTGGATCATATCTATTCCC
TCTGTTTCCTTGGCCTACATTTGGAATATTTCCCTCCGATTTTAAGAGGTTCTTTAGGTGCATTGGGACCA
TTTCCATGACCTACATGTGGAATGTATCTTCCACCTTGATTATGTTTTGGATTTTCATTTAAATCATATC
CACGACCGCCAAATGCATTGCCATTCGGAATAGGATTTCCGACTGGAAGTCTCCGATTTCCGGATT
TGGATCATATATATTTCCCTCTGTTTCCTTGGCCTACATTTGGAAT
```

gi|103660570|gb|EC054701.1|;

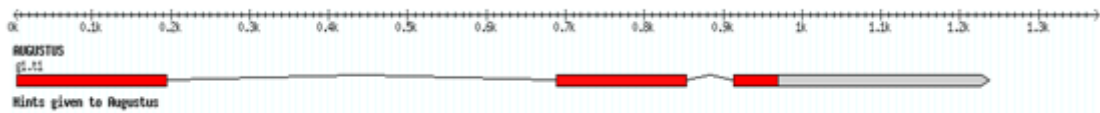
```
AGCCGGCGGAGAACCTGGCAAGGATACGTGCAAAGGCGATGGAGGTGCCCCACTTGCCTGCCCACTGCAA
AGCGATCCTAACCGGTACGAACTGTTGGGAATCGTAAACTTTGGATTCCGGATGTGGAGGGCCACTTCCTG
CTGCCTACACTGATGTTTCCCAAATTCGAAGCTGGATTGATAATTGTATCCAAGCGGAAGCAGTCCACTA
TTCTCCTCAGTTGGGAAATGTAGGCCAATCACCTGCTCCACTCGATCGCTATATACCAAATATAGGCCTA
GAAACTCAAAATGAAGTCCACAGCCCGGTTCAAGGATCTCTCCATAATGTTCGTATATCAAACCTGGTGGTG
TTGGATATATTCGAGGTATGGGCCAAGGAACTGCTTTCCAAGGACCAGTTCAAGTTGGAGAACTTATTCC
GAACGAAGGCACCGTATCTGGCGGACTTGTATATGATTCAAATGAAATCCAAAAATTTAATCAAGGTGGA
ATATACATTCCACATGTAGG
```

gi|103649961|gb|EC051179.1

```
AGCCGGCGGAGAACCTGGCAAGGATACGTGCAAAGGCGATGGAGGTGCCCCACTTGCCTGCCCACTGCAA
AGCGATCCTAACCGGTACGAACTGTTGGGAATCGTAAACTTTGGATTCCGGATGTGGAGGGCCACTTCCTG
CTGCCTACACTGATGTTTCCCAAATTCGAAGCTGGATTGATAATTGTATCCAAGCGGAAGCAGTCCACTA
TTCTCCTCAGTTGGGAAATGTAGGCCAATCACCTGCTCCACTCGATCGCTATATACCAAATATAGGCCTA
GAAACTCAAAATGAAGTCCACAGCCCGGTTCAAGGATCTCTCCATAATGTTCGTATATCAAACCTGGTGGTG
TTGGATATATTCGAGGTATGGGCCAAGGAACTGCTTTCCAAGGACCAGTTCAAGTTGGAGAACTTATTCC
GAACGAAGGCACCGTATCTGGCGGACTTGTATATGATTCAAATGAAATCCAAAAATTTAATCAAGGTGGA
ATATACATTCCACATGTAGG
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 17 sORFs. El sexto y el noveno sORF son repeticiones desplazadas del tercero; el séptimo, octavo y noveno sORF se repiten en tándem dos veces más. La transcripción se basa en la presencia de tres ESTs en la base de datos pública, y la probable traducción en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORFs codificados por este transcripto es traducido. La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_69



CTCCATGGAGATCAAGCACCTTGATGGACACCTGGTGCCCGTCACGAGAGAGAAGGTAACGTGGCCCCGGC
M D T W C P S R E R R *

GCTAGAATCCGCAAGAAGGGCGAGGGCATGCCAACCTTTGAGAACAACAACCTCACCGGCAACCTGTACA
TCACCTTCGATGTGGAGTTCCCAAGAAGGATCTTACGGAGGAGGACAAGGAAGGTTAGCATATACATTA
TCACAAGAGTAATTGTTTCATTAATAAACGTTTCTAATTCCTAGCGCTCAAGAAAATACTCGACCAATCCT
CCATCAATCGCATATACAATGGACTGTGAATACCCATGTTTCATCCCCAATAAAAAGTGCTCGCCGGCTGT
M D C E Y P C F I P N K S A R R L
ACTCTAGTTTTTTAAATGCCTAACTAACAGGGCGACTTTGATTTGGGAACGCCCTAATTC AAGAGATGAC
Y S S F *

TTTGTTCAAAAAAGTTGTGCGAAAACAATTTAATTTTAGTTTGAATTTTGTAAAATAGCAAAAAGTGTA
L F K K V V R K Q F N F S L E F C K I A K V *

TGAGAGGGAAC TATTAAGGCGGTTCC TGCCGACAAAAAGGAATTGCAACAAAAACCAACCCCAAAAA
TAAAGCGGATTTAAATTATATACAAC TTTGCGGAAGCTACTATTTTTGATTTACTGTTTTAAAAGGGAT
TTTTTGATTGTCTCAGAAATAATTTAATTTGAAAAC TATTATATTAAGACAATCGATAGGCTGCATTATT
GCAACCTTGCTCAGCTGTGCATAACTGGCGCAGGAATCCGCAAGAAGGGCGAGGGCATGCCAACCTTG
M P N F
AGAACAACAACCTCACCGGCAACCTGTACATCACCTTCGATGTGGAGTTCCCAAGAAGGATCTTACGGA
E N N N L T G N L Y I T F D V E F P K K D L T E
GGAGACAAGGAAGGTTAGCATATACATTATCACAAGAGTAATTGTTTCATTAATAAACGTTTCTAATTC
E D K E G *

TAGCGCTCAAGAAAATACTCGACCAATCCTCCATCAATCGCATATACAATGGACTGTGAATACCCATGTT
M D C E Y P C
TCATCCCCAATAAAAAGTGCTCGCCGGCTG TACTCTAGTTTTTTAAATGCCTAACTAACAGGGCGACTTTGA
F I P N K S A R R L Y S S F *

TTTTGGGAACGCCCTAATTC AAGAGATGACTTTGTTCAAAAAAGTTGTGCGAAAACAATTTAATTTTAGT
M T L F K K V V R K Q F N F S
TTAGAATTTTGTAAAATAGCAAAAAGTGTAATGAGAGGGAAC TATTAAGGCGGTTCC TGCCGACAAAAAGG
L E F C K I A K V *

AATTGCAACAAAAACCAACCCCAAAAAATAAAGCGGATTTAAATTATATACAAC TTTGCGGAAGCTACT
ATTTTTGATTTACTGTTTTTAAAAGGGATTTTTTGATTGTCTCAGAAATAATTTAATTTGAAAAC TATTA
TATTAAGACAATCGATAGGCTGCATTATTGCAACCTTGCTCAGCT

Para la región genómica 2L_69, el programa Augustus Gene Prediction predijo una estructura compuesta por 3 exones, uno de 189 pares de bases, el segundo de 164 pares de bases, y el tercero de 323 pares de bases, y 2 intrones, uno de 495 pares de bases y el segundo de 60 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido DLTEEDKEG con una probabilidad de $\rho=0.068$ y un valor $E=0.00078340696005875$.

Evidencia de ESTs:

gi |114070047|dbj|BP553224.1|;
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGATATAAATTTAAATCCGCTTTATTTTTGTGGGGGTTGGTTTTT
GTTGCAATTCCTTTTTGTGCGCAGGAACCGCCTTAATAGTTCCTCTCATTACACTTTTGCTATTTTACA
AAATTTTAACTAAAATTAATTTGTTTTCGCACAAC TTTTTTGAACAAAGTCATCTCTTGAATTAGGGGC
GTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCATTTAAAAC TAGAGTACAGCCGGCGAGCACTTTTAT
TGGGGATGAAACATGGGTATTCACAGTCCATTGTATATGCGATTGATGGAGGGTTGGTTCGAGTATTTTCT
TGAGCGCTTCCTTGTCCTCCTCCGTAAGATCCTTCTTGGGGAAC TCCACATCGAAGGTGATGTACAGGTT
GCCGGTGAGGTTGTTGTTCTCAAAGTTGGGCATGCCCTCGCCCTTCTTGCGGATTC TAGCGCCGGGCCAC
GTTACCTTCTCTCTCGTGACGGGCACCAGGTGCCATCAAGGTGCTTGATCTCCATGGAGAAGCCGACAA
GGGCATCTGCAGACTGATGGTTCAGTTCGTGTACAGATCATATTCTTGCGCAAGAATCGCGGATGCGG

CATTTGTTGAACCCGCACAATGAAGTCGCCGGGCCCCCATCGATATGTGGCTCCCC

gi | 8120729 | gb | AW943021.1 | ;

ATAATTTAAATCCGCTTTATTTTGTGGGGTTGGTTTTTGTGCAATTCCTTTTTGTTCGGCAGGAACCCG
CTTAATAGTTCCCTCTCATTACACTTTTGTCTATTTTACAAAATTCATAACTAAAATTAATTTGTTTTCGC
ACAACCTTTTTGAACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAG
GCATTTAAAACTAGAGTACAGCCGGCGAGCACTTTTATTGGGGATGAAACATGGGTATTCACAGTCCAT
TGTATATGCGATTGATGGAGGATTGGTCGAGTATTTCTTGAGCGCTTCCTTGTCCTCCTCCGTAAGATC
CTTCTTGGGGAACCTCCACATCGAAGGTGATGTACAGGTTGCCGGTGAGGTTGTTGTTCTCAAAGTTGGGC
ATGCCCTCGCCCTTCTTGCGGATTCTAGCGCCGGGCCACGTTACCTTCTCTCTCGTGACGGGCACCAGGT
GTCCATCAAGGTGCTTGATCTCCATGGAGAAGCCGACAAGGGCATCCTGCAGACTGATGGTCACGTTCTGT
GT

gi | 103772525 | gb | EC090092.1 | ;

ATGACCGCTTTATTTTGTGGGGTTGGTTTTTGTGCAATTCCTTTTTGTTCGGCAGGAACCCGCTTAATA
GTTCCCTCTCATTACACTTTTGTCTATTTTACAAAATTCATAACTAAAATTAATTTGTTTTCGCACAACCT
TTTTGAACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCATTTA
AAAAGTACAGTACAGCCGGCGAGCACTTTTATTGGGGATGAAACATGGGTATTCACAGTCCATTGTATAT
GCGATTGATGGAGGATTGGTCGAGTATTTCTTGAGCGCTTCTTGACCTCCTCCGTAAGATCCTTCTTG
GGAACTCCACATCGAAGGTGATGTACAGGTTGCCGGTGAGGTTGCTGTACTCAAAGTTGGGCATGC

gi | 49379138 | gb | CO320704.1 | ;

ATGACCGCTTTATTTTGTGGGGTTGGTTTTTGTGCAATTCCTTTTTGTTCGGCAGGAACCCGCTTAATA
GTTCCCTCTCATTACACTTTTGTCTATTTTACAAAATTCATAACTAAAATTAATTTGTTTTCGCACAACCT
TTTTGAACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCATTTA
AAAAGTACAGTACAGCCGGCGAGCACTTTTATTGGGGATGAAACATGGGTATTCACAGTCCATTGTATAT
GCGATTGATGGAGGATTGGTCGAGTATTTCTTGAGCGCTTCTTGACCTCCTCCGTAAGATCCTTCTTG
GGAACTCCACATCGAAGGTGATGTACAGGTTGCCGGTGAGGTTGCTGTACTCAAAGTTGGGCATGC

gi | 107222932 | gb | EC203778.1 | ;

GCTTTATTTTGTGGGGTTGGTTTTTGTGCAATTCCTTTTTGTTCGGCAGGAACCCGCTTAATAGTTCCC
TCTCATTACACTTTTGTCTATTTTACAAAATTCATAACTAAAATTAATTTGTTTTCGCACAACCTTTTTGA
ACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCATTTAAAACT
AGAGTACAGCCGGCGAGCACTTTTATTGGGGATGAAACATGGGTATTCACAGTCCATTGTATATGCGATT
GATGGAGGATTGGTCGAGTATTTCTTGAGCGCTTCTTGACCTCCTCCGTAAGATCCTTCTTGGGGAAC
TCCACATCGAAGGTGATGTACAGGTTGCCGGTGAGGTTGTTGTTCTCAAAGTTGGGCATGC

gi | 49230633 | gb | CO308755.1 | ;

GCTTTATTTTGTGGGGTTGGTTTTTGTGCAATTCCTTTTTGTTCGGCAGGAACCCGCTTAATAGTTCCC
TCTCATTACACTTTTGTCTATTTTACAAAATTCATAACTAAAATTAATTTGTTTTCGCACAACCTTTTTGA
ACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCATTTAAAACT
AGAGTACAGCCGGCGAGCACTTTTATTGGGGATGAAACATGGGTATTCACAGTCCATTGTATATGCGATT
GATGGAGGATTGGTCGAGTATTTCTTGAGCGCTTCTTGACCTCCTCCGTAAGATCCTTCTTGGGGAAC
TCCACATCGAAGGTGATGTACAGGTTGCCGGTGAGGTTGTTGTTCTCAAAGTTGGGCATGCTCTCGCCCT
TCTTGCGGATTCTAGCGCCGGGCCACGTACCTTTTCTCTCGTGACGGGCACC

gi | 103773345 | gb | EC090523.1 | ;

ATGACCGCTTTATTTTGTGGGGTTGGTTTTTGTGCAATTCCTTTTTGTTCGGCAGGAACCCGCTTAATA
GTTCCCTCTCATTACACTTTTGTCTATTTTACAAAATTCATAACTAAAATTAATTTGTTTTCGCACAACCT
TTTTTGAACAAAAGTCATCTTTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCATT
TAAAAACTAGAGTACAGCCGGCGAGCACTTTTATTGGGGATGAAACATGGGTATTCACAGTCCATTGTAT
ATGCGATTGATGGAGGATTGGTCGAGTATTTCTTGAGCGCTTC

gi | 148890392 | gb | ES690303.1 | ;

AGTCGTCGACCGCTTTATTTAGTGGGGTAGGTTTTTGTGCAATTCCTTTTTGTTCGGCAGGAACCCGCTT
AATAGTTCCCTCTCATTACACTTTTGTCTATTTTACAAAATTCATAACTAAAATTAATTTGTTTTCGCACA
ACTTTTTTGAACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCA
TTTTAAAACTAGAGTACAGCCGGCGAGCACTTTTATTGGGGATGAAACATGGGTATTCACAGTCCATTGT
ATATGCGATTGATGGAGGATTGGTCGAGTATTTCTTGAGCGCTTCTTGACCTCCTCCGTAAGATCCTT

CTTGGGAACTCCACATCGAAGGTGATGTACAGGTTGCCGGTGAGGTTGTTGTTCTCAAAGTTGGGCATG
 CCCTCGCCCTTCTTGGGATTC TAGCGCCGGGCCACGTTACCTTCTCTCTCGTGACGGGCACCAGGTGTC
 CATCAAGGTGCTTGATCTCCATGGAGAAGCCGACAAGGGCATCCTGCAGACTGATGGTCACGTTTCGTGTA
 CAGATCATCATTCTTGCAGGAATCGCGGATGCGGCATTTGTTGAACCCGCACAATGAGGTGCGCCGGGC
 TCCCCATCGATATGTGGCTCCCCCTC

gi | 148890367 | gb | ES690278.1 | ;

CGTCGTCGACGCTTTATTTTGTGGGGTGGTATAGTTGCAATTCCTTTTTGTGCGGCAGGAACCGCCTTAA
 TAGTTCCTCTCATTACACTTTTGTATTTTACAAAATTC TAAACTAAAATTAATTTGTTTTCGCACAAAC
 TTTTTTGAACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCGCCCTGTTAGTTAGGCATT
 TAAAAACTAGAGTACAGCCGGCGAGCACTTATATTGGGGATGAAACATGGGTATTCACAGTCCATTGTAT
 ATGCGATTGATGGAGGATTGGTCGAGTATTTTCTTGAGCGCTTCCTTGTCTCCTCCGTAAGATCCTTCT
 TGGGAACTCCACATCGAAGGTGATGTACAGGTTGCCGGTGAGGTTGTTGTTCTCAAAGTTGGGCATGCC
 CTCGCCCTTCTTGGGATACTAGCGCCGGGCCACGTTACCTTCTCTCTCGTGACGGGCACCAGGTGTCCA
 TCACGGTGCTTGATCTCCATGGAGAAGCCGACATGGGCATCCTGCAGACTGATGGTCACGTTTCGTGTACC
 GATCACCATTCTTGCAGGAATCGCAGATTGCGCATTTGTTGAACCCGCCAAATGAGGTGCGCA

gi | 1704395 | gb | AA140799

ATAATNTAAATCCGCTTTATTTTGTGGGGTGGTTTTTGTGCAATTCCTTTTTGTTGGCAGGAACCGC
 CTTAATANTTCCCTCTCATTACACTTTTGTATTTTACAAAATTC TAAACTAAAATTAATTTGTTTTCGC
 ACAACTTTTTTGAACAAAAGTCATCTCTTGAATTAGGGGCGTTCCCAAATCAAAGTCCCCCTGTTAGTTAG
 GCNTTTAAAACTATAGTACAGCCGGCGAACACTTTTATTGGGGATGAAACATGGGTATTCNCAGTCCAT
 TGTCTATGCGATTGATGGAGGATTGGTCNAATNTTTTCTTGAACGCNTCCNTGTCTCCCTCCGTAATAAT
 CCNTCTTGGGAACTCCNCATCCAA

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 6 sORFs. El quinto y sexto sORF son repeticiones del segundo y tercer sORF. La transcripción se basa en la presencia de diez ESTs en la base de datos pública, lo que indicaría un alto nivel de expresión y la probable traducción en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORFs codificados por este transcrito es traducido.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_70



ACTCTCGTGGCTGGAATTTCCGATGGAATAGTTGTATCGGGTGAACTGCAGGATGAGTAGATTTTACTA
 M_S_R_F_Y_Y
 TAGTTTCAGTTGTTCTGACGCTCGGAGGTGCAGTAGTCTGTTCTGTCACTTGCCCACCTGTTGAACCAAT
 _S_F_S_C_S_D_A_R_R_C_S_S_L_F_C_H_L_P_T_C_*
 TGGAACATTTGTAGAGGTTTCTGAAGTGACTGATGGAACGGTAGTTTCCAGAGATACTGGACTGGGAGTA
 GATGTCTCAAATGGACGGGTGGTAAGAACGTCCTCTAGTAGATTTCAGTTGTTTGTCTAAGTGAGGTCTAG
 TAGTGAAGTCTGAAGGTAATGTTGTGGGTTTCTCAGGACTTTCGGATGGAAGGTTCTATCCGTTGACCT
 M_L_W_V_S_Q_D_F_R_M_E_K_F_Y_P_L_T
 TGAAGGAAGTGTAGATTCTTCAACACGTATTGTTGTTCTCACTTCGCTCGGAGAAGATGTACTCTGTTCT

L_K_E_V_* M_Y_S_V_L
 GTC AATTGCC CACCTGTTGAACCAATTGGAACATTTGTAGTGGTTTCAGAAGTGACTGATGGAACAGTTG
 S_I_A_H_L_L_N_Q_L_E_H_L_* M_E_Q_L_
 TTTCCAGAGATGCTGGACTGGGAGTAGATGCCTCAAATGGCCGGGTGGTTGGAACGTCTCTAGTAGATTC
 F_P_E_M_L_D_W_E_* M_P_Q_M_A_G_W_L_E_R_L_*
 AGTCGTTTGTCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGTTGGTGTCTCAGGACTTTTCG
 M_L_L_V_S_Q_D_F_R
 GATGGAGTAGTTCTATCCGCTGACCTTGAAGGAAGTGTAGATTTCAACGCGTATTGTAGTTCTGACTT
 M_E_*
 CGCTCGGCGGTGCAGTTGTCTGTCCTGCCACTTGCCACCTGTTGAACCAATTGGAACATTTGTGGTGGT
 TTCTGAAGTTACTGGTGAACAGCAGTATCTAGAGTTACTGGTCTGGGAGTAGATGCCTCAAATGGCTGG
 M_P_Q_M_A_G
 GTGGTTGGAACGTCTCTAGTAGATTCAGTCGTTTGATCTGAGTGAGGTCTTGTGTAAGTCTGAAGGTA
 W_L_E_R_L_*
 ATGTTGTGGGTGTCTCAGGACTTTCCGATGGAGAAGTTCTATCCGCTGACCTAGGAGGAAGTGTAGATTC
 M_L_W_V_S_Q_D_F_R_M_E_K_F_Y_P_L_T_*
 TTCAACACGTATTGTAGTTCTGACTTCGCTCGGCGGTGCAGTTGTCTGTCCTGTCACTTGCCCACCACAC
 CTGTTGAACCAATTGGAACATTTGTGGTGGTTTCTGAAGTTACTGATGGAACAGCAGTTTCTAGAGTTAC
 M_E_Q_Q_F_L_E_L_
 TGGTCTGGGAGTAGATGCCTCAAATGGCCGGGTGGTTGGAACGTCTCTAGTAGATTCAGTCGTTTGATCT
 L_V_W_E_* M_P_Q_M_A_G_W_L_E_R_L_*
 GAGTGAGGTCTAGTAGTGAAGTCTGAAGGTAATGTTGTGGGTGTCTCAGGACTTTCCGATGGAGTAGTTC
 M_L_W_V_S_Q_D_F_R_M_E_*
 TATCCGTTGACCTTGAAGGAAGTGTAGATTTCTCAACACCTATTGTAGTTCTGACTTCGCTCGGCGTTGC
 AGTAGTCTGTCTGTCACTTGCCACCTGTTGAACCGATTGGAACATTTGTAGTGGTTTCTAAAGTAAC
 GATGGAACAGTTGTTTCCAGAGATGCTGGACTGGGAGTAGATGCCTCAAATGGCCGGGTGGTTGGAACGT
 M_E_Q_L_F_P_E_M_L_D_W_E_* M_P_Q_M_A_G_W_L_E_R_L_*
 CTCTAGTAGATTGATCGTTTGATCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGTGGGTGT
 L_* M_L_W_V_
 CTCAGGACTTTCCGATGGAGTAGTTCTATCCGTTGACCTTGAAGGAAGTGTAGATTTCTCAACACCTATT
 S_Q_D_F_R_M_E_*
 GTAGTTCTGACTTCGCTCGGCGTTGCAGTAGTCTGTCTGTCACTTGCCACCTGTTGAACCGATTGGAA
 CATTGTAGTGGTTTCTGAAGTTACTGGGGGAACAGCAATTTCTAGAGTTACTGGTCTGGGAGTAGATGA
 M_
 CTCAAATGGCTGGGTGGTTGGAACGTCTCTAGAAGATTCAGTCGTTTGATCTGAGTGAGGTCTAGTTGTG
 T_Q_M_A_G_W_L_E_R_L_*
 AAGTCTGAAGGTAATGTTGTGGGTGTCTCAGGACTTTCCGATGGAGGAGTTCTATCCGTTGACCTCGAAG
 M_L_W_V_S_Q_D_F_R_M_E_*
 GAAGTGTAGATTTCTCAACGCGTATTGTAGTTCTGACTTCGCTCGGCGGTGCAGTAGTCTGTCTGTCACT
 TTGCCACCTGTTGAACCCATTGGAACACTTGTAGTGGTTTCTAAAGTAACGTGGAACAGTTGTTTCC
 M_E_Q_L_F_P
 AGAGATGCTGGACTGGGAGTAGATGCCTCAAATGGCCGGGTGGTTGGAACGTCTCTAGTAGATTGATTCAGTCG
 E_M_L_D_W_E_* M_P_Q_M_A_G_W_L_E_R_L_*
 TTTGATCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGTTGGTGTCTCAGGACTTTCCGATGG
 M_L_L_V_F_Q_D_F_R_M_
 AGTAGTTCTATCCGTTGACCTTGAAGGAAGTGTAGATTTCTCAACACGTATTGTAGTTCTGACTTCGCTC
 E_*
 GGAGAAGATGTAGTTTGTCTGTCACTTGCCACCTGTTGAACCGATTGGAACATTTGTAGTGGCTTCTG
 AAGTGAAGTGTAGTTGTTTTCAGAGATGCTGGACTGGGAGTAGATGCTTCAAATGGACGGGTGGT
 M_E_Q_L_F_P_E_M_L_D_W_E_* M_S_R_M_V_G_W_
 TGAACGTCTCTAGTAGATTGATTTGTTTTTCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTT
 M_L_L_E_R_L_* M_L
 GTTGGTGTCTCAGGACTTTCCGATGGAGTAGTTCTATCCGTTGACCTTGAAGGAAGTGTAGATTTCTCA
 L_V_F_Q_D_F_R_M_E_*
 ACACGTATTGTAGTTCTGACTTCGCTCGGAGAAGATGTAGTTTGTCTGTCACTTGCCACCTGTTGAAC
 CGATTGGAACATTTGTAGTGGCTTCTGAAGTGAAGTGTAGTTGTTTTCAGAGATGCTGGACTGGG
 M_E_Q_L_F_S_E_M_L_D_W_
 AGTAGATGCCTCAAATGGACGGGTGGTTGGAACGTCTCTAGTAGATTGATTCAGTCGTTTGTCTGAGTGAGGT
 E_* M_P_Q_M_D_G_W_L_E_R_L_*
 CTAGTTATGAAGTCTGAAGGTAATGTTGTGGGTGTCTCAGGACTTTCCGATGGAGAAGTTCTATCCGTTG
 M_L_W_V_S_Q_D_F_R_M_E_K_F_Y_P_L_

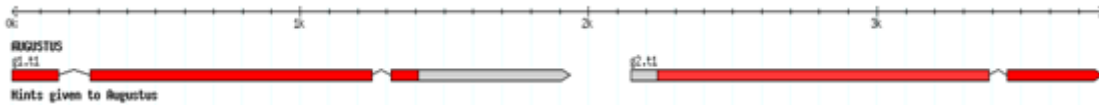
ACCTTGAAGGAAGTGTAGATTCTTCAACACGTATTGTAGTTCTGACTTCGCTCGGCGGTGCAGTAGTCTG
T L K E V *
TCCTGTCACTTGCCACCTGTTGAACCCATTGGAACACTTGTAGTGGTTTCTAAAGTAACTGATGGAACA
M E Q
GTTGTTTCCAGAGATGCTGGACTGGGAGTAGATGTCTCGAATGGTCGGGTGGTTGGAACGTCTCTAGTAG
L F P E M L D W E * M S R M V G W L E R L *
ATTGAGTTGTTTTTCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGTTGGTGTTCAGGACT
M L L V F Q D
TTCGGATGGAGTAGTTCTATCCGTTGACCTTGAAGGAAGTGTAGATTCTTCAACACGTATTGTAGTTCTG
F R M E *
ACTTCGCTCGGAGAAGATGTAGTTTGTCTGTCACTTGCCACCTGTTGAACCGATTGGAACATTTGTAG
TGGCTTCTGAAGTACTGATGGAACAGTTGTTTTTCAGAGATGCTGGACTGGGAGTAGATGCTTCAAATGG
M E Q L F S E M L D W E * M L Q M
ACGGGTGGTTGGAACGTCTCTAGTAGATTTCAGTCGTTTGTCTGAGTGAGGTCTAGTTATGAAGTCTGAA
D G W L E R L *
GGTAATGTTGTGGGTGTCTCAGGACTTTCGGATGGAGAAGTTCTATCCGTTGACCTTGAAGGAAGTGTAG
M L W V S Q D F R M E K F Y P L T L K E V *
ATTCTTCAACACCTATTGTAGTTCTGACTTCGCTCGGCGGTGCAGTAGTCTGTCCCTGTCACTTGCCACCC
TGTTGAACCAATTGGAACATTTGTAGTAGTTTCTGAAGTACTGATGGAACAGTAGTTTCCAGAGATGCT
GGACTGGGAGTAGATGCCCAAATGGCCTGGTGGTTGGAACGTCTCTAGTAGATTTCAGTCGTTTGTCTG
M P Q M A W W L E R L *
AGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGGTGTCTCAGGACTTTCGGATGGAGTAGTTCTATC
M L V S Q D F R M E *
CGCTGACCTTGAAGGAAGTGTAGATTCTTCAACGCGTATTGTAGTTCTCACTTCGCTCGGAGAAGATGTA
GTTTGTCTGTCACTTGCCACCTGTTGAACCGATTGGAACATTTGTAGTGGTTTCTAAAGTAACTGATG
M
GAACAGTTGTTTCCAGAGATGCTGGACTGGGAGTAGATGTCTCGAATGGTCGGGTGGTTGGAACGTCTCT
E Q L F P E M L D W E * M S R M V G W L E R L
AGTAGATTCAGTTGTTTTTCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGTTGGTGTTC
* M L L V F Q
GGACTTTCGGATGGAGTAGTTCTATCCGTTGACCTTGAAGGAAGTGTAGATTCTTCAACACGTATTGTAG
D F R M E *
TTCTGACTTCGCTCGGAGAAGATGTAGTTTGTCTGTCACTTGCCACCTGTTGAACCGATTGGAACATT
TGTAGTGGCTTCTGAAGTACTGATGGAACAGTTGTTTTTCAGAGATGCTGGACTGGGAGTAGATGCTTCA
M E Q L F S E M L D W E * M L Q
AATGGACGGGTGGTTGGAACGTCTCTAGTAGATTTCAGTCGTTTGTCTGAGTGAGGTCTAGTTATGAAGT
M D G W L E R L *
CTGAAGGTAATGTTGTGGGTGTCTCAGGACTTTCGGATGGAGAAGTTCTATCCGTTGACCTTGAAGGAAG
M L W V S Q D F R M E K F Y P L T L K E
TGTAGATTCTTCAACACCTATTGTAGTTCTGACTTCGCTCGGCGGTGCAGTAGTCTGTCCCTGTCACTTGC
V *
CCACCTGTTGAACCAATTGGAACATTTGTAGTAGTTTCTGAAGTACTGATGGAACAGTAGTTTCCAGAG
ATGCTGGACTGGGAGTAGATGCCCAAATGGCCTGGTGGTTGGAACGTCTCTAGTAGATTTCAGTCGTTT
M P Q M A W W L E R L *
TTCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGGTGTCTCAGGACTTTCGGATGGAGTAGTT
M L V S Q D F R M E *
CTATCCGCTGACCTTGAAGGAAGTGTAGATTCTTCAACGCGTATTGTAGTTCTCACTTCGCTCGGAGAAG
ATGTAGTTTGTCTGTCACTTGCCACCTGTTGAACCGATTGGAACATTTGTAGTGGTTTCTAAAGTAAC
TGATGGAACAGTTGTTTCCAGAGATGCTGGACTGGGAGTAGATGTCTCGAATGGTCGGGTGGTTGGAACG
M E Q L F P E M L D W E * M S R M V G W L E R
TCTCTAGTAGATTTCAGTTGTTTTTCTGAGTGAGGTCTAGTTGTGAAGTCTGAAGGTAATGTTGTTGGT
L * M L L V
TTTCAGGACTTTCGGATGGAGTAGTTCCATCCGTTGACCTTGAAGGAAGTGTAGATTCTTCAACACGTA
F Q D F R M E *
TTGTAGTTCTGACTTCGCTCGGAGAAGATGTAGTTTGTCTGTCACTTGCCACCTGTTGAACCGATTGG
AACATTTGTAGTGGCTTCTGAAGTACTGATGGAACAGTTGTTTTTCAGAGATGCTGGACTGGGAGTAGAT
M E Q L F S E M L D W E * M
GCCTCAAATGGACGGGTGGTTGGAACGTCTCTAGTAGATTTCAGTCGTTTGTCTGAGTGAGGTCTAGTTA
P Q M D G W L E R L *
TGAAGTCTGAAGGTAATGTTGTGGGTGTCTCAGGACTTTCGGATGGAGAAGTTCTATCCGTTGACCTTGA
M L W V S Q D F R M E K F Y P L T L

gi|15479873|gb|BI588451.1|

GTCTCAGGACTTTCGGATGGAGTAGTTCTATCCGCTGACCTTGAAGGAAGTGTAGATTCTTCAACGCGTA
 TTGTAGTTCTCACTTCGCTCGGAGAAGATGTAGTTTGTCTGTCACCTGCCACCTGTTGAACCGATTGG
 AACATTTGTAGTGGTTTCTAAAGTAACTGATGGAACAGTTGTTTCCAGAGATGCTGAACTGGGAGTAGAT
 GCCTCAAATGGGCGGGTGGTTGGAACGTCTCTAGTAGATTTCAGTCGTTTGTATCTGAGTGAGGTCTAGTTA
 TGAAGTCTGAAGGTAATGTTGTGGGTGTCTCAGGACTTTCGGATGGAGAAGTTCTATCCGTTGACCTTGA
 AGGAAGTGTAGATTCTTCAAGACCTATTGTAGTTCTGACTTCGCTCGGAGAAGATGTAGTTTGTCTGT

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 60 sORFs. Se pueden observar varias repeticiones en tándem y varias desplazadas. La transcripción se basa en la presencia de un EST en la base de datos pública y la probable traducción en la presencia en bases de datos de 4 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos 4 de los sORFs codificados por este transcrito son traducidos. La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada

Gen 2L_80



GTGTGGCCAGTGCCTCAGCCACGTCCCCTGACCGAGAAGCTGCCCGCCAACCACCCCTGCTCACC GGAC
 AGCGTGTGCTCGACTCGCTCTTCCCTGTGTCCAGGGCGGTACCACC GCCATTCCCGGAGCTTTCGGTTG
 CCGCAAGACTGTGATCTCGCAGGTGAGAGTCCCACAAATTGAGAATTTAAGGAGCGATGCCTCGTGTAGC
 CTCCATACTCAAGTTTTCATAAAAAATACAATCCCTAATAAATCATTTACTTGCTTG CAGGCTCTGTCCA
 AGTACTCCA ACTCCGATGTCATCATCTACGTCGGTTGCGGTGAGCGTGGTAACGAGATGTCTGAGGTACT
 GCGTGACTTCCCGAGCTGTCCGTGGAGATCGACGGTGT CACCGAGTCCATCATGAAGCGTACC GCCCTT
 GTGGCCAACACCTCCAACATGCCTGTGGCTGCTCGTGAGGCC TCCATCTACACTGGTATCACCTTGTCCG
 AATACTCCGTGATATGGGTTACAACGTGTCCATGATGGCTGATTCCACCTCCC GTTGGGCTGAGGCTCT
 TCGTGAAAATTTCTGGTCTGCTCGCTGAGATGCC TGCCGATTCCGGCTACCCAGCCTACTTGGGAGCCCGT
 CTGGCCTCCTTCTACGAGCGTGCCGGTCGCGTTAAGTGCTTGGGTAACCCCGAGCGCGAGGGATCCGTGT
 CCATTGTGCGGAGCTGTGTCTCCTCCTGGTGGTGACTTCTCCGATCCC GTGACCTCCGCCACTCTGGGTAT
 CGTGCAGGTGTTCTGGGGTCTCGACAAGAAGTTGGCCAGCGCAAGCACTTCCCTTCGATCAACTGGCTC
 ATCTCTACTCGAAGTACATGCGTGTCTGGATGACTTCTATGACAAGA ACTTCCCGAATTCGTGCCGC
 TGCGTACCAAGGTCAAGGAGATCCTGCAGGAGGAGGAGGATCTGTCTGAGATCGTGA ACTGGTCCGCAA
 GGCTCTCTGGCCGAAACCGACAAGATCACGCTGGAGGTGGCCAAGCTGCTGAAGGACGATTTCTGCAG
 CAGAACTCCTACTCCTCGTACGATCGCTTCTGCCCTTCTACAAGACCGTGGGCATGTTGAGGAACATCA
 TCGACTTCTACGACATGGCCCGTCACTCCGTGGAGTCTACGGCTCAGTCTGAGAACAAGATCACCTGGAA
 CGTGATTTCGTGAGGCAATGGGCAACATTATGTACCAGCTGT CATCCATGAAGTTCAAGGTGGGTTAACAC
 GCAAATTTATCCATTGCC TAGACACTGGGTGACCACATTTTTCAATCCATTT CAGGACCCCGTTAAGGAT

M
 GGTGAGGCCAAGATCAAGGCTGACTTCGAGCAGCTGCACGAGGACCTGCAGCAGGCCTTCAGAAATCTGG
V R P R S R L T S S S C T R T C S R P S E I W
 AGGACTAGAGACCGCGCTGGCCCTACTTTTACTCTAATCTTATATTTGTTATATAGTTAACGTTTAA
R T R D R A G P T F T L *
 AATGAAAGCAGTCAAAAACCATCCGAAAAAGCCTAATCAAACACCAACAATCCGTGCTGCATTCCGATGA
M K A V K N H P K K P N Q T P T I P C C I R *
 M
 AAAACAAAAGTCCAACAAATACCACA ACTTCTTGGTGCCTGCGAGAGATGTAAACATTCGGCCTGCGGT
K N K S P T N T T T S W C L R E M *

TAATACTTTCCCCTAACCCAGCCCCCTCCGCCCTTGAAGGGCAACTCTAGGCAACAGCAACTACAACGT
 CCTGCTATGTACTTCCATTTACAACAACAACCAACATACACTTGAATAAAAGTACACGGACTGGCG
M_Y_F_H_L_Q_Q_Q_H_Q_H_T_L_E_*
 CACACACAACACATACATAAAAAGACACAAATACAAATGCATGCATAAAATAGTATTATTGTTTAAATGAATG
M_N
 GAAATTCTTGTATTATTGTGAAAAAAGTCATGTTTTCTCCCTGTTTGTGTTTGTAAATTTATGTAATATT
G_N_S_C_L_F_V_K_K_V_M_F_S_P_C_L_F_V_K_F_M_*
 TAAAGTATGAAATATTTAAATGTACGAATAAAGTGCAACAACAATAACATTTAATGTAATTGAAAGTGAAT
 TTCACTGGCAGCAGAATGGATATTAATAATGTGTCAACTCGATAAAAAGATAATAAGTTAAAATATTTTT
M_D_I_K_N_V_S_T_R_*
M_C_Q_L_D_K_K_I_I_S_*
 TTGAATTTTAAACCTTCATTATATAAACATACTTGACTATATGAAAGCTAAGAAAATGGGAATATATTGA
M_K_A_K_K_M_G_I_Y_*
 AAAGCATAATCTAAAAATCCTATATCATTCTCAGTCATCACCTTAGTCTTTCCATAAACTTCTGCCCAT
 TATTTTATTTATTTGACAATATTAGTAAAAATTTGTACCAAGTGAAAAAATTTGTATATATTTAAACCAAT
 GGAAAAGAAGTCTGGGTTGCACCGCTTACCTCGTTCACCAACAGTAACGCAAGTAACGAAGACAGAGCC
 CGCAACTCGGGATCGTTAAACCAAGATGGGGAGCCAAGGTCCGTGCCTCACCAACGGTTCCGGCTAAGA
M_G_S_Q_G_P_C_L_T_Q_R_F_R_L_R
 AGTGCACATGTCCAACCTGCGCTTGCCAACGTAATGACCATAAAATCCAATCCGGTACCCCTCCCCCTCA
S_A_H_V_Q_P_A_L_A_N_V_M_T_I_N_P_I_R_Y_P_L_P_L
 ATCACGCAAGCCTGTAGCCGAGCCGGAAAGTCCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGAC
N_H_A_S_L_*
M_I_Q_S_M_T_R_T_V_*
 TTGAGACGTTTCGACGGACCAATCCCACAAGAGCGCTCACATCGCCTTGGAGAAGAATGAGGACTCGGGTT
 TTGTGATCGAGCAGGTGGTTGATACGCACAAATATTCGTCCGATGAAGAAGAGGAGGAGGCGCAGATGGG
 TCGCATTTTCGGAGTCTCCGGTCCGGTGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTG
 GTTCGCGTTGGACACTCCAGCTTGTGGTGGATCATTCGACTGGAGGGTGATATGGCCACCATTCCAGG
 TTTACGAGGATACTTCGGGTGTGAGCGTGGGTGATCCCGTCTACCAGACGGGAAAGCCGCTCTCCGTGGA
 ATTGGGACCCGGCATCATGGGCAGCATTTTCGATGGTATCCAGCGACCATTGAGGTCCATCAGTGAACTA
 ACCAACTCCATATACGTGCCAAGGGCATCGATACGCCCTCCCTGCCAGGAACATTGCGTACGAATTCA
 CACCCGAAAAATTGAAGATCGATGCTCTGATCACCGCGGAGACATCTACGGATCTGTTTTCGAAAACAG
 CATGATGCACGATCACCGCTGATACTACCGCCCCGCACCAAGGGGCGCATCCGGTGGTTGGCCCCGCC
 GGGAACTACTGCGTGGACGAGGTGATCGTGGAGACGGAGTTCAACGACGAGATCACCAGCACACCATGC
 TCCAGGTGTGGCCGTACGGAGGTGTCGTCCGGTGGAGGATAAGCTCCCCAGCAATTCACCACTCTTGAC
 TGGCCAGCGCTCTGGACGCCCTTCTTTCCATGTGTCCAGGGCGGAACCACTGCCATTCCAGGAGCGTTT
M_C_P_G_R_N_H_C_H_S_R_S_V
 GGATGTGAAAGACCGTCATCTCACAGGTGAGAGGGTTTCTAAGAGTTTAGTTGACAAATGATTACATTC
W_M_W_K_D_R_H_L_T_G_E_R_V_S_K_S_L_V_D_K_*
 CAATCAACTTATACCCCTAGGCCCTGTCCAAATACTCCAACCTCAGATGTCATCATCTACGTGGGCTGCGG
 TGAGCGCGGGAACGAAATGTCCGAGGTTCTTCGGGACTTTCCGCAGCTTGAGGTTGAGGTCAATGGCACC
 ATGGAATCGATTATGAAGCGGACTGCCCTGGTGGCCAATACCTCCAATATGCCGGTGGCTGCCCGAGAAG
 CCTCGATCTACACGGGCATCACCTCTCGGAATACTTCCGCGACATGGGCTACCATGTGTCCATGATGGC
 GGACTCCACATCTCGATGGGCGGAGGCGCTGCGTGAGATTTCCGGACGTCTGGCCGAGATGCCTGCCGAT
 G

Para el cluster de sORFs 2L_80, el programa Augustus Gene Prediction predijo, para el gen1.1, una estructura compuesta por 3 exones de 162 pares de bases el primero, el segundo de 977 pares de bases y el tercero de 619 pares de bases, y 2 intrones, el primero de 109 pares de bases y el segundo de 68 pares de bases; para el gen 2.1 predijo una estructura de 2 exones, el primero de 1241 pares de bases y el segundo de 530 pares de bases, y un intrón de 64 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido SPTNNTTTSWCLR con una probabilidad de $\rho=0.017$ y un valor $E=0.000844543161337096$.

Se halló evidencia de ESTs:
 gi|13692404|gb|BF500563.2|;

ATTTTATTTATTTGACAATATTAGTAAAAATTTGTACCAAGTGAAAAAATTTGTATATATTTAAACCAATG

GAAAAGAAGTCTCTGGGTTGCACCGCTTACCTCGTTCACCAACAGTAACGCAAGTAACGAAGACAGAGCCC
GCAACTCGGGATCGTTAAACCAAGATGGGGAGCCAAGTCCGTGCCTCACCCAACGGTTCGGGCTAAGAA
GTGCACATGTCCAACCTGCGCTTGCCAACGTAATGACCATAAAATCCAATCCGGTACCCCCCTCCCCCTCAA
TCACGCAAGCCTGTAGCCGAGCCGGAAAAGTCCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGACT
TGAGACGTTTCGACGGACCAATCCCACAAGAGCGCTCACATCGCCTTGGAGAAGAATGAGGACTCGGGTTT
TGTGATCGAGCAGGTGGTTGATACGCACAAATATTCGTCCGATGAAGAAGAGGAGGAGGCGACGATGGGT
CGCATTTTCGGAGTCTCCGGTCCGGTGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTGG
TTCGCGTTGGACACTCCCAGCTTGTGGTGAGATCATTCGACTGGAGGGTGATATGGCCACCATTTCAGGT
TTACGAGGATACTTCGGGTGTGAGCGTGGGTGATCCCGTCTACCAGACGGGAAAAGCCACTCTCCGTTGAA
TTGGGACCCGGCATCATGGGCAGCATCTTTGATGGTATCCAGCGACCATTGAGGTCATCAGTGAACATA
CCAACCTCCATATACGTGCCCAAGGGCATCGATACGCCCTCCCTGCCAGGAACAT

gi|13693150|gb|BF506709.2|;

CATTCTCAGTCATCACCTTAGTCTTTCCATAAACTTCTGCCCATTTATTTTATTTATTTGACAATATTA
GTAAAAATTTGACCAAGTGAATAAATTTGATATATTTAAACCAAATGGAAAAGAAGTCTGGGTTGCACC
GCTTACCTCGTTCACCAACAGTAACGCAAGTAACGAAGACAGAGCCCGCAACTCGGGATCGTTAAACCAA
GATGGGGAGCCAAGGTCCGTGCCTCACCCAACGGTTCGGCTAAGAAGTGCACATGTCCAACCTGCGCTT
GCCAACGTAATGACCATAAAATCCAATCCGGTACCCCCCTCCCCCTCAATCACGCAAGCCTGTAGCCGAGCC
GGAAAGCCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGACTTGAGACGTTTCGACGGACCAATCC
CACAAAGCCCTCACATAGCCTTGGAGAAGAATGAGGACTCGGGTTTTGTGATCGAGCAGGTGGTTGATA
CGCACAAATATTCGTCCGATGAAGAAGAGGAGGAGGCGACGATGGGTTCGATTTTCGGAGTCTCCGGTCC
GGTGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTGGTTCGCGTTCGGACACTCCCAGCTA
GTTGGTGAGATCATTCGACTGGAGGGTGATATGGCCACCATTTCAGGTTTACGAGGATACTTCGGGTGTGA
GCGTGGGTGATCCCGTCTACCAGACGGGAAAAGCCGCTCTCCGTTGAATTGGGACCCGGCATCATGGGCAG
CATCTTCGATGGTATCCAGCGACCATTGAGGGTCATCAGTGAAC

gi|13698237|gb|BF486355.2|;

AAAAATTTGACCAAGTGAATAAATTTGATATATTTAAACCAAATGGAAAAGAAGTCTGGGTTGCACCGC
TTACCTCGTTCACCAACAGTAACGCAAGTAACGAAGACAGAGCCCGCAACTCGGGATCGTTAAACCAAGA
TGGGGAGCCAAGGTCCGTGCCTCACCCAACGGTTCGGCTAAGAAGTGCACATGTCCAACCTGCGCTTGC
CAACGTAATGACCATAAAATCCAATCCGGTACCCCCCTCCCCCTCAATCACGCAAGCCTGTAGCCGAGCCGG
AAAGTCCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGACTTGAGACGTTTCGACGGACCAATCCCA
CAAGAGCGCTCACATCGCCTTGGAGAAGAATGAGGACTCGGGTTTTGTGATCGAGCAGGTGGTTGATACG
CACAAATATTCGTCCGATGAAGAAGAGGAGGAGGCGACGATGGGTTCGATTTTCGGAGTCTCCGGTCCGG
TGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTGGTTCGCGTTCGGACACTCCCAGCTTGT
TGGTGAGATCATTCGACTGGAGGGTGATATGGCCACCATTTCAGGTTTACGAGGATACTTCGGGTGTGAGC
GTGGGTGATCCCGTCTACCAGACGGGAAAAGCCGCTCTCCGTTGAATTGGGACCCGGCATCATGGGCAGCA
TTTTTCGATGGTATCCAGCGACCATTGAGGGTCATCAGTGAACATAACCAACTCCATATACGT

gi|13755044|gb|BF489384.2|;

AAAAATTTGACCAAGTGAATAAATTTGATATATTTAAACCAAATGGAAAAGAAGTCTGGGTTGCACCGC
TTACCTCGTTCACCAACAGTAACGCAAGTAACGAAGACAGAGCCCGCAACTCGGGATCGTTAAACCAAGA
TGGGGAGCCAAGGTCCGTGCCTCACCCAACGGTTCGGCTAAGAAGTGCACATGTCCAACCTGCGCTTGC
CAACGTAATGACCATAAAATCCAATCCGGTACCCCCCTCCCCCTCAATCACGCAAGCCTGTAGCCGAGCCGG
AAAGTCCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGACTTGAGACGTTTCGACGGACCAATCCCA
CAAGAGCGCTCACATCGCCTTGGAGAAGAATGAGGACTCGGGTTTTGTGATCGAGCAGGTGGTTGATACG
CACAAATATTCGTCCGATGAAGAAGAGGAGGAGGCGACGATGGGTTCGATTTTCGGAGTCTCCGGTCCGG
TGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTGGTTCGCGTTCGGACACTCCCAGCTTGT
TGGTGAGATCATTCGACTGGAGGGTGATATGGCCACCATTTCAGGTTTACGAGGATACTTCGGGTGTGAGC
GTGGGTGATCCCGTCTACCAGACGGGAAAAGCCGCTCTCCGTTGAATTGGGACCCGGCATCATGGGCAGCA
TTTTTCGATGGTATCCAGCGACCATTGAGGGTCATCAGTGAACATAACCAACTCCATATACGT

gi|13691260|gb|BF499396.2|;

ATGGGTTACAACGTGTCCATGATGGCTGATTCACCTCCCGTGGGCTGAGGCTCTTCGTGAAATTTCTG
GTCGTCTCGTGAGATGCCTGCCGATTCGGCTACCCAGCCTACTTGGGAGCTCGTCTGGCCTCCTTCTA

CGAGCGTGCCGGTCGCGTTAAGTGCTTGGGTAACCCCGAGCGGAGGGATCCGTGTCCATTGTCGGAGCT
GTGTCTCCTCCTGGTGGTACTTCTCCGATCCCGTGACCTCCGCCACTCTGGGTATCGTGCAGGTGTTCT
GGGGTCTCGACAAGAAGTTGGCCCAGCGCAAGCATTTCGCCCTCGATCAACTGGCTCATCTCCTACTCGAA
GTACATGCGTGCTCTGGATGACTTCTATGACAAGAACTTCCCCGAATTTCGTGCCGCTGCGTACCAAGGTC
AAGGAGATCCTGCAGGAGGAGGAGGATCTGTCTGAGATCGTGCAACTGGTCCGGCAAGGCCTCTCTGGCCG
AAACCGACAAGATCACGCTGGAGGTGGCCAAGCTGCTGAAGGACGATTTCTGCAGCAGAACTCCTACTC
CTCGTACGATCGCTTCTGCCCTTCTACAAGACCGTGGGCATGTTGAGGAACATCATCGACTTCTACGAC
ATGGCCCGTCACTCCGTGGAGTCTACGGCTCAGTCTGAGAACAAGATCACCTGGAACGTGATTCGTGAGG
CAATGGGCAACATTATGTACCAGCTGTCATCCATGAAGTTCAAGGACCCCGTTAAGGATGGTGAGGCCAA
GATCAAGGCTGA

gi|13688793|gb|BF506216.2|;

AAATTGTACCAAGTGAAAAAATTGTATATATTTAAACCAAATGGAAAAGAAGTCTGGGTTGCACCGCTT
ACCTCGTTACCAACAGTAACGCAAGTAACGAAGACAGAGCCCGCAACTCGGGATCGTTAAACCAAGATG
GGGAGCCAAGGTCCGTGCCTCACCCAACGGTTCCGGCTAAGAAGTGCACATGTCCAACCTGCGCTTGCCA
ACGTAATGACCATAAATCCAATCCGGTACCCCTCCCCCTCAATCACGCAAGCCTGTAGCCGAGCCGGAA
AGTCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGACTTGAGACGTTGACGGACCAATCCCACA
AGAGCGCTCACATCGCCTTGAGAGAAGATGAGGACTCGGGTTTTGTGATCGAGCAGGTGGTTGATACGCA
CAAATATTCGTGGATGAAGAAGAGGAGGAGGCGACGATGGGTGCGATTTTCGGAGTCTCCGGTCCGGTG
GTCAATCCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTGGTTCGCGTTGGACACTCCCAGCTTGTG
GTGAGATCATTTCGACTGGAGGGTGATATGGCCACCATTACAGTTTTACGAGGATACTTCGGGTGTGAGCGT
GGGTGATCCCGTCTACCAGACGGGAAAGCCGCTCTCCGTGGAATTGGGACCCGGCATCATGGGCAGCATT
TTCGATGGTATCCAGCGACCATTGAGGTCCA

gi|13693563|gb|BF496063.2|;

TTCTGCCCCATTTATTTTATTTTATTTGACAATATTAGTAAAAAATTGTACCAAGTGAAAAAATTGTATATA
TTTAAACCAAATGGAAAAGAAGTCTGGGTTGCACCGCTTACCTCGTTACCAACAGTAACGCAAGTAAC
GAAGCAGAGCCCGCAACTCGGGATCGTTAAACCAAGATGGGGAGCCAAGGTCCGTGCCTCACCCAACGG
TTCCGGCTAAGAAGTGCACATGTCCAACCTGCGCTTGCCAACGTAATGACCATAAATCCAATCCGGTACC
CCCTCCCCCTCAATCACGCAAGCCTGTAGCCGAGCCGAAAGCCCCCATGATACAGTCAATGACGAGGAC
AGTTTGAAGGACTTGAGACGTTGACGGACCAATCCCACAAGAGCGCTCACATAGCCTTGGAGAAGAATG
AGGACTCGGGTTTTGTGATCGAGCAGGTGGTTGATACGCACAAATATTTCGTGGATGAAGAAGAGGAGGA
GGCGACGATGGGTGCGATTTTCGGAGTCTCCGGTCCGGTGGTCAATGCCGAGGAGATGGCCGGCGCAGCC
ATGTACGAGCTGGTTCGCGTCCGACACTCCCAGCTAGTTGGTGAGATCATTCGACTGGAGGGTGATATGG
CCACCATTACAGTTTACGAGGATACTTCGGGTGTGAGCGTGGGTGATCCCGTCTACCAGACGGGAAAGCC
GCTCTCCGTTGAATTGGGACCCGGCATCATGGGCAGCATCTTCGATGG

gi|13694048|gb|BF496495.2|;

AAAAATTGTACCAAGTGAAAAAATTGTATATATTTAAACCAAATGGAAAAGAAGTCTGGGTTGCACCGC
TTACCTCGTTACCAACAGTAACGCAAGTAACGAAGACAGAGCCCGCAACTCGGGATCGTTAAACCAAGA
TGGGGAGCCAAGGTCCGTGCCTCACCCAACGGTTCCGGCTAAGAAGTGCACATGTCCAACCTGCGCTTGC
CAACGTAATGACCATAAATCCAATCCGGTACCCCTCCCCCTCAATCACGCAAGCCTGTAGCCGAGCCGG
AAAGCCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGACTTGAGACGTTGACGGACCAATCCCA
CAAGAGCGCTCACATAGCCTTGAGAGAAGATGAGGACTCGGGTTTTGTGATCGAGCAGGTGGTTGATACG
CACAAATATTTCGTGGATGAAGAAGAGGAGGAGGCGACGATGGGTGCGATTTTCGGAGTCTCCGGTCCGG
TGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTGGTTCGCGTCCGACACTCCCAGCTAGT
TGGTGAGATCATTTCGACTGGAGGGTGATATGGCCACCATTACAGTTTTACGAGGATACTTCGGGTGTGAGC
GTGGGTGATCCCGTCTACCAGACGGGAAAGCCGCTCTCCGTTGAATTGGGACCCGGCATCATGGGCAGCA
TCTTCGATGGTATCCAGCGACCATTGAGGTCCA

gi|13762330|gb|BG634793.1|;

AAAAATTGTACCAAGTGAAAAAATTGTATATATTTAAACCAAATGGAAAAGAAGTCTGGGTTGCACCGC
TTACCTCGTTACCAACAGTAACGCAAGTAACGAAGACAGAGCCCGCAACTCGGGATCGTTAAACCAAGA
TGGGGAGCCAAGGTCCGTGCCTCACCCAACGGTTCCGGCTAAGAAGTGCACATGTCCAACCTGCGCTTGC

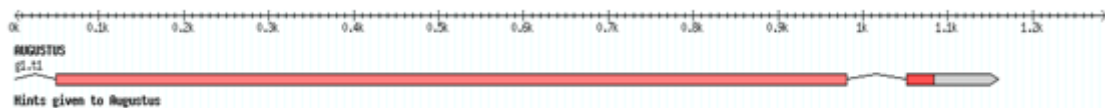
CAACGTAATGACCATAAAATCCAATCCGGTACCCCTCCCCCTCAATCAGCAAGCCTGTAGCCGAGCCGG
 AAAGTCCCCATGATACAGTCAATGACGAGGACAGTTTGAAGGACTTGAGACGTTTCGACGGACCAATCCCA
 CAAGAGCGCTCACATCGCCTTGGAGAAGAATGAGGACTCGGGTTTTGTGATCGAGCAGGTGGTTGATACG
 CACAAATATTCGTCGGATGAAGAAGAGGAGGAGGCGCAGATGGGTTCGCATTTTCGGAGTCTCCGGTCCGG
 TGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTACGAGCTGGTTCGCGTTGGACACTCCCAGCTTGT
 TGGTGAGATCATTCGACTGGAGGGTGATATGGCCACCATTTCAGGTTTACGAGGATACTTCGGGTGTGAGC
 GTGGGTGATCCCGTCTACCAGACGGGAAAGCCGCTCTCCGTGGAATTGGGACCCGGCATCATGGGCAGCA

gi|13693804|gb|BF496314.2|

AACTTCTGCCCCATTTATTTTATTTATTTGACAATATTAGTAAAAATTGTACCAAGTAAAAAATTGTAT
 ATATTTAATCCAAATGAAAAAGAAGTCTGGGTTGCACCGCTTACCTCGTTCACCAACAGTAACGAAGAC
 AGAGCCCGCAACTCGGGATCGTTAAACCAAGATGGGGAGCCAAGGTCCGTGCCTCACCAACGGTTCCGG
 CTAAGAAGTGCACATGTCCAACCTGCGCTTGCCAACGTAATGACCATAAAATCCAATCCGGTACCCCTCC
 CCCTCAATCACGCAAGCCTGTAGCCGAGCCGAAAGCCCCATGATACAGTCAATGACGAGGACAGTTTG
 AAGGACTTGAGACGTTTCGACGGACCAATCCCACAAGAGCGCTCACATCGCCTTGGAGAAGAATGAGGACT
 CGGGTTTTGTGATCGAGCAGGTGGTTGATACGCACAAATATTCGTCGGATGAAGAAGAAGAGGAGGCGAC
 GATGGGTTCGCATTTTCGGAGTCTCCGGTCCGGTGGTCAATGCCGAGGAGATGGCCGGCGCAGCCATGTAC
 GAGCTGGTTCGCGTCGGACACTCCCAGCTAGTTGGTGAGATCATTCGACTGGAGGGTGATATGGCCACCA
 TTCAGGTTTACGAGGATACTTCGGGTGTGAGCGTGGGTGATCCCGTCTACCAGACGGGAAAGCCGCTCTC
 CGTTGAATTGGGACCCGGCATCATGGGCAGCATCTTCGATGGT

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 11 sORFs. Se pueden observar varias repeticiones en tándem y varias desplazadas. La transcripción se basa en la presencia de 10 ESTs en la base de datos pública y la probable traducción en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORFs codificados por este transcrito es traducido. La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_92



TTGGCGGTCTGGATATGATTTTAAATGAAATCCAAAAACATAATCAAGGTGGGAGATAACATCCACATGT

M I L M K S K N I I K V G D T F H M

AGGTCATGGAAATGGTCCAATGCACCTAAAGAACCTCTTAAAATCGGAGGAAATATTCCAAATGTAGGC

* M E M V P M H L K N L L K S E E I F Q M *

CAAGGAAACAGGGTAATAGATATGATCTAAATCCTGAAATCGGAGCAGTTCCAGTCGGAAATCCTATTC
 CGAATGAAGGCAATGCATTTGGCGGTCTGGATATGATTTAAATGAAATCCAAAAACATAATCAAGGTGG

M_K_A_M_H_L_A_V_V_D_M_I_* M_K_S_K_N_I_I_K_V_
 AAGATACATTCCACATGTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAACTTCTTAAAATCGGAGGA
 E_D_T_F_H_M_* M_E_M_V_P_M_H_L_K_N_F_L_K_S_E_E
 AATATTCCAAATGTAGGCCAAGGAAACAGAGGGAATATATATGATCCAAATCCGGAAATCGGAGCAGTTC
 I_F_Q_M_* M_I_Q_I_R_K_S_E_Q_F_
 CAGTCGGAAATCCTATTCCGAATGAAGGCAATGCATTTGGCGGTCTGGATATGATTTAAATGAAATCCA
 Q_S_E_I_L_F_R_M_K_A_M_H_L_A_V_V_D_M_I_* M_K_S_
 AAAACATAATCAAGGTGGAAGATACATTCCACATGTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAA
 K_N_I_I_K_V_E_D_T_F_H_M_* M_E_M_V_P_M_H_L_K_N
 CCTCTTAAAATCGGAGGAAATATTCCAAATGTAGGCCAAGGAAACAGAGGGAATATATATGATCCAAATC
 L_L_K_S_E_E_I_F_Q_M_* M_I_Q_I_
 CGGAAATCGGAGCAGTTCAGTTCGGAAATCCTATTCCGAATGAAGGCAATGCATTTGGCGGTCTGGATA
 R_K_S_E_Q_F_Q_S_E_I_L_F_R_M_K_A_M_H_L_A_V_V_D_
 TGATTTAAATGAAATCCAAAAACATAATCAAGGTGGAAGATACATTCCACATGTAGGTCATGGAAATGGT
 M_I_* M_K_S_K_N_I_I_K_V_E_D_T_F_H_M_* M_E_M_V
 CCAATGCACCTAAAGAACCTCTTAAAATCGGAGGAAATATTCCAAATGTAGGCCAAGGAAACAGAGGGA
 P_M_H_L_K_N_L_L_K_S_E_E_I_F_Q_M_*
 ATATATATGATCCAAATCCGGAAATCGGAGCAGTTCAGTTCGGAAATCCTATTCCGAATGAAGGCAATGC
 M_I_Q_I_R_K_S_E_Q_F_Q_S_E_I_L_F_R_M_K_A_M_
 ATTTGGCGGTCTGGATATGATTTAAATGAAATCCAAAAACATAATCAAGGTGGAAGATACATTCCACAT
 H_L_A_V_V_D_M_I_* M_K_S_K_N_I_I_K_V_E_D_T_F_H_M
 GTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAACCTCTTAAAATCGGAGGAAATATTCCAAATGTAG
 * M_E_M_V_P_M_H_L_K_N_L_L_K_S_E_E_I_F_Q_M_*
 GCCAAGGAAACAGAGGGAATAGATATGATCCAAATCCGGAAATCGGAGCAGTTCAGTTCGGAAATCCTAT
 M_I_Q_I_R_K_S_E_Q_F_Q_S_E_I_L_
 TCCGAATGAAGGCAATGCATTTGGCGGTCTGGATATGATTTAAATGAAATCCAAAAACATAATCAAGGT
 F_R_M_K_A_M_H_L_A_V_V_D_M_I_* M_K_S_K_N_I_I_K_V
 GGAAGATACATTCCACATGTAGGTCATGGAAATGGTCCCAATGCACCTAAAGAACCTCTTAAAATCGGAG
 E_D_T_F_H_M_* M_E_M_V_P_M_H_L_K_N_L_L_K_S_E_
 GAAATATTCCAAATGTAGGCCAAGGAAACAGAGGTAATAGATATGATCTAAATCCTGAAATCGGACATGC
 E_I_F_Q_M_* M_
 CACATTTGGTGCACCCGGTACCCGGAATGCAAGGAAATAATATAGATAATCCATATAAAAAACAATCTTGGT
 P_H_L_V_Q_P_V_P_E_C_K_E_I_I_*
 GACAGAGCAGAAAATGAATTGGAAAGGTTTCTCTTGAATAATCCCAATAGGGATCATGTTATCAGGCTTA
 M_N_W_K_G_F_L_L_K_I_P_I_G_I_M_L_S_G_L_
 GTGATTTTTTAATCACAACGCCACTCCGAGTGCACCTATAAA
 V_I_F_*

Para el gen 2L_92, el programa Augustus Gene Prediction predijo una estructura compuesta por 2 exones de 931 pares de bases el primero, y el segundo de 108 pares de bases, y 1 intrón de 71 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido PHLVQPVPECKEII con una probabilidad de $p=0.028$ y un valor $E=5.29465057429542e-06$.

Se halló evidencia de EST:

gi|103649946|gb|EC051174.1|

ATGCACCAAAATGTGGCATGTCCGATTTTCAGGATTTAGATCATATCTATTACCTCTGTTTCCCTTGGCCTAC
 ATTTGGAATATTTTCTCCGATTTTAAAGAGGTTCTTTAGGTGCATTTGGGACCATTTCCATGACCTACATGT
 GGAATGTATCTTCCACCTTGATTATGTTTTTGGATTTTCATTTAAATCATATCCACGACCGCCAAATGCAT
 TGCCCTCATTCGGAATAGGATTTCCGACTGGAACCTGCTCCGATTTCCGGATTTGGATCATATCTATTCCC
 TCTGTTTCCCTTGGCCTACATTTGGAATATTTCTCCGATTTTAAAGAGGTTCTTTAGGTGCATTTGGGACCA
 TTTCCATGACCTACATGTGGAATGTATCTTCCACCTTGATTATGTTTTTGGATTTTCATTTAAATCATATC
 CACGACCGCCAAATGCATTTGCCCTCATTCGGAATAGGATTTCCGACTGGAACCTGCTCCGATTTCCGGATT
 TGGATCATATATATTTCCCTCTGTTTCCCTTGGCCTACATTTGGAAT

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 19 sORFs. Se pueden observar varias repeticiones en tándem y varias desplazadas. Los sORFs octavo, decimoprimeros, decimocuarto y decimoseptimo son repeticiones del segundo sORF. Los sORFs noveno, decimosegundo y decimoquinto son repeticiones del sexto sORF. Los sORFs décimo, decimotercero y decimosexto son repeticiones del séptimo sORF. La transcripción se basa en la presencia de un EST en la base de datos pública y la probable traducción en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORFs codificados por este transcripto es traducido.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_106



CCGAATCCTTGTGGCCATTTCCACCTTGTCCGAATCCTTGCTGACCACGTCCACCCTGGCCGTGTCCAC
 CCTGTTGTCCATAGCCAGCGCAAAGACCACAGCGATCAAGGCGAAGAAAACAATCTAAAAGAAGAGAAA
 TAAAACAATGTTAATTCGAGCCCAAAAATCCTTGAAAGGAATTTGCAGAGCTTACGAAGAATTTCA
M_F_N_S_E_P_T_K_I_L_E_R_N_L_Q_S_L_R_R_I_S
 TTTTGGTAGATTGATTGTTTACACTGTCGAAAAGCGTTCTAACTAATTCATGTAATGGCAGGGCTTTTA
F_W_*
 TAGCAGGTGGAAGCCGAATGGAGCGATTTGCACCAAACAAAGTAATAAAAACAAAATTTTCTTATTTTT
M_E_R_F_A_P_K_Q_S_N_K_N_K_I_F_L_F_F
 TGCTTATTATTTCTATCGAAATCTATATACCCCAAGCTGTGATTATTAGATTAAGCGTCTGCGATGAGG
A_Y_Y_F_L_S_K_S_I_Y_P_K_L_*
 TAGCAAAATTTCTAGCACTGCATTAACAGAGCTGATTTCTGATCAGGGGTTGAGGACCTTTAGCCTGATCT
 TTCTTCTGTTAATTTGTTTTGCTTATTTTGGAGCTGATTAGCTATATACCGATTAAGCTTTTCAATT
 GCCACTAAAAAAGTATAAGAAACGGCGAAGACTTTTTAGCGCTGCATTTAAATTCACAAGCTTAACA
 GTATTCTTTTAGGCTGAATTTGCTCTCAAAAAAAGTACAATGGTTTGGAAAATTTTAAATGTTTTTGAA
M_F_L_N
 TGTAACCAAACACAGCGAAAAAATACAAATAATTTCTTTAAAAAAGGATGTAAAGAAATAAACT
V_T_K_H_S_E_K_K_Y_K_*
 AGTCTTATTAGTTTAAATAAATAACAATTTTTTAAAGACACATAATTTCTATGTTTTTATAGACGGTACTTGT
 GGAGTAATTGAGTATACCAAATTTCTTTAAAAAGTCTGTGGGAGTTTGAAGAAGGCGTTGCCGAACCTTTA
 CCTATACAGAATATATATATTTCTTACCAGCATCACCAGCGGAGTCTATCTAGCCGATTTCTTTGGCCA
 TCTGTCTGTAAGCCTGTCTGTATGAACTGAATGAGATTGAGACCTTAAATAGTTCCAATTTCTTTCAGAA
M_R_F_R_P_L_N_S_S_N_S_F_R
 CCTGTGCCAACAACTCTAGATTCTTCATTGATTTTCATCTAGTCTTGTCTTGTAAATATCTTAGTTCTT
T_C_A_N_N_T_L_D_S_S_L_I_S_S_S_L_V_L_*
 CTATTAGCTAAGAAATGGGTATCTGATAATCGAGGAATGCGACTATAGTGGTCTAACCTGTTAAGCGCTA
 AAATTCAAAGACATAACATAATTATACGTATGTATAAAAAATTAATCCATGATTACAAGGATTTTCACT
M_Y_K_K_L_I_H_D_Y_K_D_F_I_L
M_I_T_R_I_S_Y
 TAATTATCCCTCCTCATTTTCGAAAGTCTAATGATAACCCAGAAAAGCAAACAAGAACATTTGTTGCCGTA
N_Y_P_S_S_F_R_K_S_N_D_N_P_E_K_Q_T_R_T_F_V_A_V
L_I_I_P_P_H_F_E_S_L_M_I_T_Q_K_S_K_Q_E_H_L_L_P_*
 ACACGTTTGTGATGCGTGAAAACGGAACATAAATGATCTAAAGTAATCTCAAACAACGTCTTCTTCTAAG

T R L *
 TCGCCCTTAGTTTTTCTTAGCAACTGTTGCGGACTTATTAGATTAGGCTCGACTTCCCAGATAAGATTGA
 TGTTGAAGGGTTCCCCCGTTGAGCATAGGCATTGATACAGTGCCTCAAATATTAATTTCCCAGACGTCTG
 GAAATTTTTTTTCGTATTTTTGTTGGGTTAATTGTACTTGTATTTGTGTCAGAACATTGATAATACTTCGGCG
 CGTGACGACATGATGAGATTGCAGGAGAGTATAAAAAGATCGGAGGCGTCCAAAAATTTTATTCACATTAG
M M R L Q E S I K D R R R P K I L F T L
 CAAGACTAACCAGTATCAGCACCATGAAGTTTTTGGTACGTTTTTACATATACGCTTCAAAGAAAAACC
A R L T E V S A P *
M K F L V R F T Y T L Q R K T
 TTTAAGAACTAAGTTGATTTTAATATTGTTACAGATCCTCTTCATTGCCCTATTTCGCCGCGGCCGCTGCC
F K N *
 CTACCTCAGTTTTGGATTTCGGCGGTGGATTTCGGTGATCAGCAGCAGCAGCAAGGAGGATTTGGAGGTG
 GATTTGGTGATCAGCAGCAGCAGCAAGGAGGATTTGGCGGTGGATTTCGGAGGTGGATTTCGGTGATCAGCA
 GCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTCGGAGGTGGATTTCGGTGATCAGCAGCAGCAACAGCAA
 CAAGGCTGGTAAAAACCAATTCATAGCACCCTTAACGCAGATAACCTCAATAAAAAATATGACACCTTTA
 TGATAAAAAACAATTGGAATTTTTTACTTTATCGTTCTCGAAAATTTATGAGTTTTATGAAAATAAATT
M I K K Q L E F F T L S F S K I Y E F Y E N K F
M S F M K I N
 TATATTTATTATTCATTTATATTTATTAATAAGTGCAGAAATAAATAACACTTAACATAAGGGACAACG
I F I I H L Y L L N S A Q K *
L Y L L F I Y I Y *
 GGAAATATCTTCTCTTTCAGATGTTCAAATGGTTTTACTTAAAACATTGTAATATAAATAAAAGCCAAA
M F K W F Y L K H C K Y K *
 ATATAAAAGCCATATTATAAATATAAGCCAAATTTTCACAAAAATTTTGTTCCTTTGAAATATGTAAC
 GCAACTAAAAAGCGGATTCACATGATTGAAAAACAGATAACAATTTTCGAGTATCCTAATCAATGGCCTC
M I E K T D T I S S I L I N G L
 CTCAGCAGCCACCCTTTTGTGCTCCTCGATCTGCACCTTGTCCAGATCGACTCCATCTTTTTTGGCTGCC
L S S H P F V L L D L H L V Q I D S I F F G C
 CGTATCTCGCCTTCCCCCTCCGATCCGTCTAGCTGGTTGGTCTGCCAACTCCTGTCCACGGCAGTGACA
P Y L A F P L R S V *
 TCACTGACATGACCTTGACCAGACTGCTATGGTTTTCTGGTCTAGATCACTCTCCCCATTTCTTCCCG
 TATTCGTCTGTCTCCATCCGGCAACTGCAACTCCTCGCAGCACAGCATGGAAATCAGCAAGTGCTCAACG
 AAGAGGGTTCCGCCGACGCTTTGAGATTCGTCCGTATCGTGGCACTGCACTTTTTGTGTGCAATCCACT
 GAAGTAGAACATATCTGTCTGTCTGTGTTTGAATGGCCCCCGGAACCTCAAGCCTAACGCCAGGGACTC
 TGAATTGGCTGCCTCCTCTGGAGGATCATTGTTGCTTAGGAGCTGGGGATCGGAGCTCACACGTCCAGTT
 GACAGCCATACGGATATATCGAATATCCTTGAATTCGGAGTTGTTTGAACCTGCGATTGGCGGTGGTGGC
 GCATGTGATTACGGGTAAATATTCGGCGTAGTGGCGGTGTGTGGGCTCTAGGTAATGGCGTGATCGAT
 M
 GTGGTTGTGACAGGAGATGTTTGGAAAGGTGGCGTATGGCCACCAGTTCGGCGGCCGACAGTCCGGCGCAC
W L C R R C L E R W R M A T S S A A D S P A H
M F G K V A Y G H Q F G G R Q S G A
 ATGGGGCAGACCACATATTCATTGGACACCCGATGTTGAACGTAAACATGTTTCGATGAACCGCTTCGCTG
M G Q T T Y S L D T R C *
H G A D H I F I G H P M L N V N M F D E P L R C
M N R F A
 TGAGCGCAAGTAGGCCCAATGTGGACAGGTGAAGCAATTGGACAGGTGGACCAGCTCTGGGATTTCCGCC
E R K *
V S A S R P Q C G Q V K Q L D R W T S S G I S P
 GTAAAGCAGCGGAACCAGATCGTCCGAGTCAGGGAACAATTGCATGGGATGATTGGCGCGATGCTCTTCC
 *
 GTTCTATCTGATGATCGTAGCATATGTCACATAGGTTATAGTTACGCAGCGCAAACACCGGAAGCGGC
 TACCCTGGAAGATCGTCATTTGACAGCCGTTGCAGCAAATGTTGCGGTGACCCATCCTCGATCCTTGTGC
 TAACTTCGTTCTTAGCGAATATACCTAAACTCTGGTCGAAAAAACGGGCAAAGGGTTCTAAATTTTAAC
 AAATCGAATCTATACAATGACATGATTTTCATAGATGTAATAATTTTCGGGGCGTTTTGGTAATAACATAT
M I F I D V K Y F G A F G N N I
 AATTCTACTTTTTAAAACATTTTTTCAACCTAAATGTATCAAACCTTACTCAAGCTTTCTACATATTTCTT
 *
M Y Q T Y S S F L H I S F
 TGCTCTTTTTAAACAAATGTTTATTTTTGGAACCTTAAGTACAGCGAAAAAAGTTAAATAAACATGCA
A S F K Q M F I F G T L S T A K K K L N K H A
 M H
 CTGTACTATTTTGTCCGGTGGACAGCTATCTACGTGACAAATAACGATAATGCAATTTGTTTTCCAAA

L Y Y F V R *
C T I L S G E H S Y L R D K * M Q F V F Q
CCTATATGAAACGAGAAATAAGCCCCATTATTGATAAGGAACATAAATCTTTTATCTAAATCTTTTGACA
T Y M K R E I S P H Y *
GTAGAAGATAACTCAAAGGATATTTTATATTTGCTCAAATCAGAATGCAGTTATCGAAAATTAACCTCC
M Q L S K I K L P
CTTGAGCTGAAAATTTAGCTTGAATTTATAATTATACATATGTCCAAATGTAATACAGCTTCAGTGTGTG
L S *
TACACATAAAATTCAAAACCCAAAAAATCCACAAAAATGCTATTTTAGTAGGTATTGCGTTGAACACAT
CGGTTCTAAATGTGGCTGTTGTGCCATTAGTTACTCTTCTGTCTCTGTCTCGGTTTGAATCTCCATCTCG
CTCTCTCTTTTTAATTTTCCCGACTTGGTGC GCGGCAGTTGTAATAGAAGTTTTCCGGCTTCATGCACC
TCCGCCCTTGCAACCACCCCTTTTCGGCGACCATTGTGGTGGATTAGTACACTCTTAAAAATAAGATGTT
TAATAAACATATAGCAGGTGGAAGCCGAATGGAGCGATTTGCACCAAAACAAAGTAATAAAAAACAAAATT
M E R F A P K Q S N K N K I
TTCTTATTTTTGCTTATTATTTCCCTATCGAAATCTATATACCCCAAGCTGTGATTATTAGATTAAGCGT
F L F F A Y Y F L S K S I Y P K L *
CTGCGATGAGGTAGCAAAATCTAGCACTGCATTAACAGAGCTGATTTCTGATCAGGGGTTGAGGACCTT
TAGCCTGATCTTTCTTCTTGTAAATTTGTTTTGCTTATTTTGGAGCTGATTAGCTATATACCGATTA
GCTTTTCAATTGCCACTAAAAAAAAGTATAAGAAACGGCGAAGACTTTTTAGCGCTGCATTTAAATTCA
CAAGCTTAACAGTATTCTTTTAGGCTGAAATTTGCTCTCAAAAAAAGTACAATTGGTTTTGAAAATTTTAA
ATGTTTTTGAATGTAACCAAAACACAGCGAAAAAATACAATAATATTTCTTTAAAAAAAAGGATGTAA
M F L N V T K H S E K K Y K *
AGAAATAAACTAGTCTTATAGTTTAAATAAATACAATTTTTTAAAGACACATAATTTCTATGTTTTTATAG
ACGGTACTTGTGGAGTAATTGAGTATAACCAATTTTAAAAAGTCTGTGGGAGTTCAAGAAGGCGTTG
CCGAACCTTTACCTATACAGAATATATATATTTCTTACCAGCATCACCAGCGGAGTCTATCTAGCCGTAT
TCGTTTTGGCCATCTGTCTGTAAGCCTGTCTGTATGAACTGAATGAGATTAGACCTTAAATAGTTCCAA
M R F R P L N S S N
TTCTTTCAGAACCTGTGCCAACCAACTCTAGATTCTTTCATTGATTTTCATCTAGTCTTGTCTTGTAAATA
S F R T C A N N T L D S S L I S S S L V L *
TCTTAGTTCTTCTATTAGCTAAGAAATGGGTATCTGATAATCGAGGAATGCGACTATAGTGGTCTAACCT
GTTAAGCGCTAAAAATCAAAGACATAACATAATTATACGTATGTATAAAAAATTAATCCATGATTACAAG
M Y K K L I H D Y K
M I T R
GATTCATACTTAATTATCCCTCCTCATTTTCGAAAGTCTAATGATAACCCAGAAAAGCAAACAAGAACAT
D F I L N Y P S S F R K S N D N P E K Q T R T
I S Y L I I P P H F E S L M I T Q K S K Q E H
TTGTTGCCGTAACACGTTTGTGATGCGTGAACCGAACATAAATGATCTAAAGTAATCTCAAACAACGT
F V A V T R L *
L L P *
CTTCTTCTAAGTCGCCCTTAGTTTTTCTTAGCAACTGTTGCGGACTTATTAGATTAGGCTCGACTTCCCA
GATAAGATTGATGTTGAAGGGTTCCCCCGTTGAGCATAGGCATTGATACAGTGCTCAAATATTAATTTT
CCAGACGTCTGGAAATTTTTTTCGTATTTTTGTTGGGTTAATTGTACTTGATTTTGTGAGAACATTGATA
ATACTTCGGCGCGTGACGACATGATGAGATTGCGAGGAGTATAAAAAGATCGGAGCGTCCAAAAATTTT
M M R L Q E S I K D R R R P K I L
ATTACATTAGCAAGACTAACCGAAGTATCAGCACCATGAAGTTTTTGGTACGTTTTTACATATACGCTTC
F T L A R L T E V S A P *
M K F L V R F T Y T L
AAAGAAAAACCTTTAAGAACTAAGTTGATTTTAAATATTGTTACAGATCCTCTTCATTGCCCTATTCGCCG
Q R K T F K N *
CGGCCGCTGCCCTACCTCAGTTTGGATTTCGGCGGTGGATTTCGGTGATCAGCAGCAGCAGCAAGGAGG
ATTTGGAGGTGGATTTGGTGATCAGCAGCAGCAGCAAGGAGGATTTGGCGGTGGATTTCGGAGGTGGATTC
GGTGATCAGCAGCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTCGGAGGTGGATTTCGGTGATCAGCAGC
AGCAACAGCAACAAGGCTGTTAAAAACCAATTCATAGCACCGTCTTAACGCAGATACCTCAATAAAAAATA
TGACACCTTTATGATAAAAAACCAATTTGGAATTTTTTACTTTTATCGTTCTCGAAAATTTATGAGTTTTAT
M I K K Q L E F F T L S F S K I Y E F Y
M S F M
GAAAATAAATTTATATTTATTTATTTATTTATTTAATAGTGCCCAGAAATAAATACACTTAACAT
E N K F I F I I H L Y L L N S A Q K *
K I N L Y L L F I Y I Y *
AAGGGACAACGGGAAATATCTTCTTTTCAGATGTTCAAATGGTTTTTACTTAAACATTGTAAATATAAA
M F K W F Y L K H C K Y K

TAAAAGCCAAAATATAAAAAGCCATATTATAAATATAAGCCAAATTTTCACAAAATTTTGTTTTTCTTTG
*
AAATATGTAACGCAACTAAAAAGCGGATTCAACATGATTGAAAAACAGATACAATTTTCGAGTATCCTAA
M I E K T D T I S S I L
TCAATGGCCTCCTCAGCAGCCACCCTTTTGTGCTCCTCGATCTGCACCTTGTCCAGATCGACTCCATCTT
I N G L L S S H P F V L L D L H L V Q I D S I F
TTTTGGCTGCCCGTATCTCGCTTCCCCCTCCGATCCGTCTAGCTGGTTGGTCTGCCAAACTCCTGTCCA
F G C P Y L A F P L R S V *
CGGCAGTGACATCACTGACATGACCTTGACCAGACTGCTATGGTTTTTCCCTGGTCTAGATCACTCTCCCA
TTTCTTCCCGTATTCTGCTCTCCATCCCGCAACTGCAACTCCTCGCAGCACAGCATGGAAATCAGCA
AGTGCTCAACGAAGAGGGTTCCGCCGAGCCTTTGAGATTTCGTCCGTATCGTGGCACTGCACTTTTTGCTG
TGCAATCCACTGAAAGTAGAACATATCTGTCTGTCTGTCTGTTTGAATGGCCCCCGGAACTCAAGCCTAAC
GCCAGGGACTCTGAATTGGCTGCCCTCCTCTGGAGGATCATTGTTGCTTAGGAGCTGGGGATCGGAGCTCA
CACGTCCAGTTGACAGCCATACGGATATATCGAATATCCTTGAATTCGGAGTTGTTTGAACACTGCGATTG
GCGGTGGTGGCGCATGTGATTACGGGTAAATATTCGGCGTAGTGGCGGTGTGTCCGGCTCTAGGTAATTG
GCGTGATCGATGTGGTTGTGCAGGAGATGTTTGGAAAGGTGGCGTATGGCCACCAGTTCGGCGGCCGACA
M W L C R R C L E R W R M A T S S A A D
M F G K V A Y G H Q F G G R Q
GTCCGGCGCACATGGGGCAGACCACATATTCATTGGACACCCGATGTTGAACGTAAACATGTTTCGATGAA
S P A H M G Q T T Y S L D T R C *
S G A H G A D H I F I G H P M L N V N M F D E
M N
CCGCTTCGCTGTGAGCGCAAGTAGGCCGCAATGTGGACAGGTGAAGCAATTGGACAGGTGGACCAGCTCT
P L R C E R K *
R F A V S A S R P Q C G Q V K Q L D R W T S S
GGGATTTCCGCTAAAGCAGCGGAACCAGATCGTCCGAGTCAGGGAACAATTGCATGGGATGATTGGCGC
G I S P *
GATGCTCTCCGTTTCTATCTGATGATCGTAGCATATGTCACATAGGTTATAGTTCACGCAGCGCAAACA
CCGGAAGCGGCTACCCTGGAAGATCGTCATTTGACAGCCGTTGCAGCAAATGTTGCGGTGACCCATCCTC
GATCCTTGTGCTAACTTCGTTCTTAGCGAATATACCTAAACTCTGGTTCGAAAAACGGGCAAAAGGGTTC
TAAATTTTAAACAAATCGAATCTATACAATGACATGATTTTCATAGATGTAAAATATTTCCGGGCGTTTGG
M I F I D V K Y F G A F G
TAATAACATATAAATTTACTTTTTAAACTATTTTTCAACCTAAATGTATCAAACCTTACTCAAGCTTTCTA
N N I * M Y Q T Y S S F L
CATATTTCCCTTTGCCTCTTTTAAACAAATGTTTATTTTTGGAACCTTAAGTACAGCGAAAAAAGTTAA
H I S F A S F K Q M F I F G T L S T A K K K L
ATAAACATGCACTGTAATTTTTGTCCGGTGGAGCACAGCTATCTACGTGACAAATAACGATAATGCAATT
N K H A L Y Y F V R *
M H C T I L S G E H S Y L R D K * M Q F
TGTTTTCCAAACCTATATGAAACGAGAAATAAGCCCCATTATTGATAAGGAACTAAATCTTTTATCTAA
V F Q T Y M K R E I S P H Y *
ATCTTTTGACAGTAGAAGATAACTCAAAGGATATTTTATATTTGCTCAAATCAGAATGCAGTTATCGAAA
M Q L S K
ATTAACCTTCCCTTGAGCTGAAAATTTAGCTTGAATTTATAATTATACATATGTCCAAATGTAATACAGC
I K L P L S *
TTCAGTGTGTGTACACATAAAAATTCAAAACCCAAAAATCCACAAAAATGCTATTTTAGTAGGTATTGC
GTTGAACACATCGGTTCTAAATGTGGCTGTTGTGCCATTAGTTACTCTTCTGTCTCTGTCTCGGTTTGAA
TCTCCATCTCGCTCTCTTTTAAATTTTCCCGACTTGGTGCAGCGGCAGTTGTAATAGAAGTTTTCCGG
CTTCATGCACCTCCGCCCTTGACCACCCCTTTCCGGCGACCATTGTGGTGGATTTCAGTACACTCTTAAA
AATAAGATGTTTAAATAACATATAGCAGGTGGAAGCCGAATGGAGCGATTTGCACCAAACAAAGTAATA
M E R F A P K Q S N
AAAACAAAATTTTCTTATTTTTTGGCTTATTATTTTCTATCGAAATCTATATACCCCAAGCTGTGATTATT
K N K I F L F F A Y Y F L S K S I Y P K L *
AGATTAAGCGTCTGCGATGAGGTAGCAAAATTCAGCACTGCATTAACAGAGCTGATTTCTGATCAGGGG
TTGAGGACCTTTAGCCTGATCTTTCTTCTTGTAAATTTGTTTTGCTTATTTTGGAGCTGATTAGCTATAT
ACCGATTAAGGCTTTTCAATTGCCACTAAAAAAGTATAAGAAACGGCGAAGACTTTTTAGCGCTGC
ATTTAAATTCACAAGCTTAACAGTATTTCTTTTAGGCTGAATTTGCTCTCAAAAAAAGTACAATTGGTTTG
GAAAATTTTAAATGTTTTTGAATGTAACCAAACACAGCGAAAAAATAACAATAATATTTCTTTAAAAA
M F L N V T K H S E K K Y K *
AAAGGATGTAAGAAAATAACTAGTCTTATTAGTTTAAATAAATAACAATTTTTTAAAGACACATAATTTCTA
TGTTTTTATAGACGGTACTTGTGGAGTAATTGAGTATACCAAATTTTAAAAAGTCTGTGGGAGTTCGA

AGAAGGCGTTGCCGAACCTTTACCTATACAGAATATATATATATTCCTTACCAGCATCACCAGCGGAGTCTA
TCTAGCCGTATTTCGTTTGGCCATCTGTCTGTAAGCCTGTCTGTATGAAGTGAATGAGATTCAGACCTCTA
M_R_F_R_P_L
AATAGTTCCAATTCTTTTCAGAACCTGTGCCAACAACACTCTAGATTCTTCATTGATTTTCATCTAGTCTTG
N_S_S_N_S_F_R_T_C_A_N_N_T_L_D_S_S_L_I_S_S_S_L
TCTTGTAATAATCTTAGTTCTTCTATTAGCTAAGAAATGGGTATCTGATAATCGAGGAATGCGACTATAG
V_L_*
TGGTCTAACCTGTTAAGCGCTAAAAATCAAAGACATAACATAATTATACGTATGTATAAAAAATTAATCC
M_Y_K_K_L_I
ATGATTACAAGATTTCATACTTAATTATCCCTCCTCATTTCGAAAGTCTAATGATAACCCAGAAAAGCA
H_D_Y_K_D_F_I_L_N_Y_P_S_S_F_R_K_S_N_D_N_P_E_K_Q
M_I_T_R_I_S_Y_L_I_I_P_P_H_F_E_S_L_M_I_T_Q_K_S
AACAAAGAACATTTGTTGCCGTAAACACGTTTGTGATGCGTGAAAACGGAACATAAATGATCTAAAGTAATC
T_R_T_F_V_A_V_T_R_L_*
K_Q_E_H_L_L_P_*
TCAAACAACGCTTCTTCTAAGTCGCCCTTAGTTTTTCTTAGCAACTGTTGCGGACTTATTAGATTAGGC
TCGACTTCCAGATAAGATTGATGTTGAAGGGTTCCCCCGTTGAGCATAGGCATTGATACAGTGCTCAA
ATATTAATTTCCAGACGCTCGAAAATTTTTTTCGTATTTTTTGTGGGTTAATTGTACTTGATTTTTGTCA
GAACATTGATAATACTTCGGCGCGTGACGACATGATGAGATTGCAGGAGAGTATAAAAAGATCGGAGGCGT
M_M_R_L_Q_E_S_I_K_D_R_R_R
CCAAAAATTTTATTTCACATTAGCAAGACTAACCGAAGTATCAGCACCATGAAGTTTTTGGTACGTTTTTAC
P_K_I_L_F_T_L_A_R_L_T_E_V_S_A_P_*
M_K_F_L_V_R_F_T
ATATACGCTTCAAAGAAAAACCTTTAAGAATAAGTTGATTTTAATATTGTTACAGATCCTCTTCATTGC
Y_T_L_Q_R_K_T_F_K_N_*
CCTATTCGCGCGCCGCTGCCCTACCTCAGTTTGGATTTCGGCGGTGGATTTCGGTGATCAGCAGCAGCAG
CAGCAAGGAGGATTTGGAGGTGGATTTGGTGATCAGCAGCAGCAGCAAGGAGGATTTGGCGGTGGATTTCG
GAGGTGGATTTCGGTGATCAGCAGCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTCGGAGGTGGATTTCGG
TGATCAGCAGCAGCAACAGCAACAAGGCTGGTAAAAACCAATTCATAGCACCGTCTTAACGCAGATACCT
CAATAAAAAATATGACACCTTTATGATAAAAAACAATTGGAATTTTTTACTTTATCGTTCTCGAAAATTT
M_I_K_K_Q_L_E_F_F_T_L_S_F_S_K_I
ATGAGTTTTATGAAAAATAAATTTATATTTATTTATTTATTTATTTATTTAAATAGTGCCCAAGAAATAAAT
Y_E_F_Y_E_N_K_F_I_F_I_I_H_L_Y_L_L_N_S_A_Q_K_*
M_S_F_M_K_I_N_L_Y_L_L_F_I_Y_I_Y_*
ACACTTAACATAAGGGACAACGGGAAATATCTTCTCTTTCAGATGTTCAAATGGTTTTACTTAAAACATT
M_F_K_W_F_Y_L_K_H
GTAAATATAAATAAAAAGCCAAAATATAAAAAGCCATATTATAAATATAAGCCAAATTTTACAAAAATTTT
C_K_Y_K_*
GTTTTTCTTTGAAAATATGTAACGCAACTAAAAAGCGGATTCAACATGATTGAAAAACAGATACAATTTTC
M_I_E_K_T_D_T_I_S
GAGTATCCTAATCAATGGCCTCCTCAGCAGCCACCCTTTTTGTGCTCCTCGATCTGCACCTTGTCAGATC
S_I_L_I_N_G_L_S_S_H_P_F_V_L_L_L_D_L_H_L_V_Q_I
GACTCCATCTTTTGGCTGCCGTATCTCGCTTCCCCCTCCGATCCGCTAGCTGGTTGGTCTGCCAA
D_S_I_F_T_T_G_C_P_Y_L_A_F_P_L_R_S_V_*
ACTCCTGTCCACGGCAGTGACATCAGTACATGACCTTGACCAGACTGCTATGGTTTTTCTGGTCTAGAT
CACTCTCCCCATTTCTTCCCGTATTCTGTCTGTCTCCATCCGGCAACTGCAACTCCTCGCAGCACAGCAT
GGAAATCAGCAAGTGCTCAACGAAGAGGGTTTCGCCGCAGCCTTTGAGATTTCGTCCGTATCGTGGCACTGC
ACTTTTTGCTGTGCAATCCACTGAAGTAGAACATATCTGTCTGTCTGTGTTTGAATGGCCCCCGGAAC
TCAAGCCTAACGCCAGGGACTCTGAATTGGCTGCCCTCCTCTGGAGGATCATTGTTGCTTAGGAGCTGGGG
ATCGGAGCTCACACGTCCAGTTGACAGCCATACGGATATATCGAATATCCTTGAATTCGGAGTTGTTTGC
AACTGCGATTGGCGGTGGTGGCGCATGTGATTACGGGTAAATATTCGGCGTAGTGGCGGTGTGTCTGGGCT
CTAGGTAATTGGCGTGATCGATGTGGTTGTGCAGGAGATGTTTGGAAAGGTGGCGTATGGCCACCAGTTC
M_W_L_C_R_R_C_L_E_R_W_R_M_A_T_S_S
M_F_G_K_V_A_Y_G_H_Q_F
GGCGGCCGACAGTCCGGCGCACATGGGGCAGACCACATATTCATTGGACACCCGATGTTGAACGTAACA
A_A_D_S_P_A_H_M_G_Q_T_T_Y_S_L_D_T_R_C_*
G_G_R_Q_S_G_A_H_G_A_D_H_I_F_I_G_H_P_M_L_N_V_N
TGTTTCGATGAACCGCTTCGCTGTGAGCGCAAGTAGGCCGCAATGTGGACAGGTGAAGCAATTGGACAGGT
M_F_D_E_P_L_R_C_E_R_K_*
M_N_R_F_A_V_S_A_S_R_P_Q_C_G_Q_V_K_Q_L_D_R
GGACCAGCTCTGGGATTTCCGCGTAAAGCAGCGGAACCAGATCGTCCGAGTCAGGGAACAATTGCATGGG

WTSSGISP*
ATGATTGGCGGATGCTCTTCCGTTTCTATCTGATGATCGTAGCATATGTCACATAGGTTATAGTTCACG
CAGCGCAAACACCGGAAGCGGCTACCCCTGGAAGATCGTCATTTGACAGCCGTTGCAGCAAATGTTGCGGT
GACCCATCCTCGATCCTTGTGCTAACTTCGTTCTTAGCGAATATACCTAAACTCTGGTCGAAAAACGGG
CAAAAGGGTCTAAATTTTAAACAAATCGAATCTATACAATGACATGATTTTCATAGATGTAAAATATTT
GGGCGTTTGGTAATAACATATAATTCTACTTTTAAAACATTTTTTCAACCTAAATGTATCAAACCTTACT
GAFGNNI* MIFIDVKYF
CAAGCTTCTACATATTTCCCTTTCCTTTTAAACAAATGTTTATTTTTTGGAACCTTAAGTACAGCGAA
SSFLHISFASFKQMFIFGTLSTAK
AAAAAGTTAAATAAACATGCACTGTACTATTTTGTCCGGTGAGCACAGCTATCTACGTGACAAATAACG
KKLNKHALYYFVR*
MHCTILSGEHSYLRDK*
ATAATGCAATTTGTTTTCCAAAACCTATATGAAACGAGAAATAAGCCCCATTATTGATAAGGAACTAAAT
MQFVFQTYMKREISPHY*
CTTTTATCTAAATCTTTTGACAGTAGAAGATAACTCAAAGGATATTTTATATTTGCTCAAATCAGAATGC
M
AGTTATCGAAAATTAAACTTCCCTTGAGCTGAAAATTTAGCTTGAATTTATAATTATACATATGTCCAAA
QLSKIKLPLS*
TGTAATACAGCTTCAGTGTGTGTACACATAAAATTCAAAACCCAAAAATCCACAAAAATTGCTATTTTA
GTAGGTATTGCGTTGAACACATCGGTTCTAAATGTGGCTGTTGTGCCATTAGTTACTCTTCTGTCTCTGT
CTCGGTTTGAATCTCCATCTCGCTCTCTCTTTTTAATTTTCCCGACTTGGTGCGCGGCAGTTGTAATAG
AAGTTTTCCGGCTTCATGCACCTCCGCCCTTGACCACCCCTTTCGGCGACCATTGTGGTGGATTGAGT
ACACTCTTAAAAATAAGATGTTTAAATAAACATATAGCAGGTGGAAGCCGAATGGAGCGATTGCACAAA
MERFAPK
ACAAAGTAATAAAAAACAAAATTTTCTTATTTTTTGCTTATTATTTCCATCGAAATCTATATACCCCAAG
QSNKNKIFLFAYYFLSKSIYPK
CTGTGATTATTAGATTAAGCGTCTGCGATGAGGTAGCAAAATTTCTAGCACTGCATTAACAGAGCTGATTT
L*
CTGATCAGGGGTTGAGGACCTTTAGCCTGATCTTCTCTTGTAAATTTGTTTTGCTTATTTTGGAGCTG
ATTAGCTATATACCGATTAAAAGCTTTTCAATTGCCACTAAAAAAAAGTATAAGAAACGGCGAAGACTT
TTTAGCGCTGCATTTAAATTCACAAGCTTAACAGTATTTCTTTTAGGCTGAATTTGCTCTCAAAAAAGTA
CAATTGGTTTGGAAAATTTTAAATGTTTTTGAATGTAACCAAACACAGCGAAAAAAAATACAAATAATAT
MFLNVTKHSEKKYK*
TTCTTTAAAAAAGGATGTAAGAAATAAACTAGTCTTATTAGTTTAATAAATACAATTTTTTAAAGACA
CATAATTTCTATGTTTTTATAGACGGTACTTGTGGAGTAATTGAGTATACCAAATTTCTTTAAAAGTCTG
TGGGAGTTCGAAGAAGCGTTGCCGAACCTTTACCTATACAGAATATATATATTTCTTACCAGCATCACC
AGCGGAGTCTATCTAGCCGTATTCGTTTGGCCATCTGTCTGTAAGCCTGTCTGTATGAACTGAATGAGAT
MR
TCAGACCTCTAAATAGTTCCAATTTCTTTCAGAACCTGTGCCAACAACTCTAGATTCTTCATTGATTTCT
FRPLNSSNSFRTCANNTLDSSLIS
ATCTAGTCTTGTCTGTAAATATCTTAGTTCTTCTATTAGCTAAGAAATGGGTATCTGATAATCGAGGAA
SSLVL*
TGCGACTATAGTGGTCAACCTGTTAAGCGCTAAAATTCAAAGACATAACATAATTATACGTATGTATAA
MYK
AAAATTAATCCATGATTACAAGGATTTTCACTTAATTATCCCTCCTCATTTCGAAAGTCTAATGATAAC
KLIHDYKDFILNYPSSFRKSNDN
MITRISYLIIPPHFESLMIT
CCAGAAAAGCAAACAAGAACATTTGTTGCCGTAACACGTTTGTGATGCGTGAAAACGGAACATAAATGAT
PEKQTRTFVAVTRL*
QKSKQEHLLP*
CTAAAGTAATCTCAAACAACGTCTTCTTCTAAGTCGCCCTTAGTTTTTCTTAGCAACTGTTGCGGACTTA
TTAGATTAGGCTCGACTTCCCAGATAAGATTGATGTTGAAGGGTCCCCCGTTGAGCATAGGCATTGAT
ACAGTGCTCAAATATTAATTTCCCAGACGTCTGGAAATTTTTTTTTCGTATTTTTTGTGGGTTAATTGACT
TGATTTTTGTCAGAACATTGATAATACTTTCGGCGCTGACGACATGATGAGATTGCAGGAGAGTATAAAG
MMRLQESIK
ATCGGAGGCGTCCAAAAATTTTATTCACATTAGCAAGACTAACCAGGATATCAGCACCATGAAGTTTTTG
DRRRPKILFTLARLTEVSAP*
MKFL
GTACGTTTTACATATACGCTTCAAAGAAAAACCTTTAAGAACTAAGTTGATTTTAAATATTGTTACAGATC
VRFTYTLQRKTFKN*

CTCTTCATTGCCCTATTTCGCCGCGGCCGCTGCCCTACCTCAGTTTGGATTTCGGCGGTGGATTTCGGTGATC
AGCAGCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTGGTGATCAGCAGCAGCAGCAAGGAGGATTTGG
CGGTGGATTTCGGAGGTGGATTTCGGTGATCAGCAGCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTCGGA
GGTGGATTTCGGTGATCAGCAGCAGCAACAGCAACAAGGCTGGTAAAAACCAATTCATAGCACCGTCTTAA
CGCAGATACCTCAATAAAAAATATGACACCTTTATGATAAAAAACAATTGGAATTTTTACTTTTATCGTT
M I K K Q L E F F T L S F
CTCGAAAAATTTATGAGTTTTATGAAAAATAATTTATATTTATTTATTTATTTATTTATTTAAATAGTGCC
S K I Y E F Y E N K F I F I I H L Y L L N S A
M S F M K I N L Y L L F I Y I Y *
CAGAAATAAATACTAACATAAGGGACAACGGGAAATATCTTCTCTTTTCAGATGTTCAAATGGTTTTTA
Q K * M F K W F Y
CTTAAAAACATTGTAAATATAAAATAAAAGCCAAAATATAAAAGCCATATTATAAATATAAGCCAAATTTTC
L K H C K Y K *
ACAAAAATTTTGTTTTTCTTTGAAATATGTAACGCAACTAAAAAGCGGATTCAACATGATTGAAAAACA
M I E K T
GATACAATTTTCGAGTATCCTAATCAATGGCCTCCTCAGCAGCCACCCTTTTGTGCTCCTCGATCTGCACC
D T I S S I L I N G L L S S H P F V L L D L H
TTGTCCAGATCGACTCCATCTTTTTTGGCTGCCCGTATCTCGCCTTCCCCCTCCGATCCGTCTAGCTGGT
L V Q I D S I F F G C P Y L A F P L R S V *
TGGTCTGCCAAACTCCTGTCCACGGCAGTGACATCACTGACATGACCTTGACCAGACTGCTATGGTTTTTC
CTGGTCTAGATCACTCTCCCCATTTCCCTTCCCGTATTCGTCTGTCTCCATCCGGCAACTGCAACTCCTCG
CAGCACAGCATGGAAATCAGCAAGTGCTCAACGAAGAGGGTTTCGCCGAGCCTTTGAGATTCGTCCGTAT
CGTGGCACTGCACTTTTTGTGTGCAATCCACTGAAGTAGAACATATCTGTCTGTCTGTCTCGTTCGAATGG
CCCCCGGAACTCAAGCCTAACGCCAGGACTCTGAATTGGCTGCCTCCTCTGGAGGATCATTGTTGCTT
AGGAGCTGGGGATCGGAGCTCACACGTCCAGTTGACAGCCATACGGATATATCGAATATCCTTGAATTCG
GAGTTGTTTGAACCTGCGATTGGCGGTGGTGGCGCATGTGATTACGGGTAAATATTCGGCGTAGTGCCGG
TGTGTCCGGCTCTAGGTAATTGGCGTGATCGATGTGGTTGTGCAGGAGATGTTTGGAAAGGCGCGTATG
M W L C R R C L E R W R M
M F G K V A Y
GCCACCAGTTTCGGCGGCCGACAGTCCGGCGCACATGGGGCAGACCACATATTCATTGGACACCCGATGTT
A T S S A A D S P A H M G Q T T Y S L D T R C
G H Q F G G R Q S G A H G A D H I F I G H P M L
GAACGTAAACATGTTTCGATGAACCGCTTCGCTGTGAGCGCAAGTAGGCCGCAATGTGGACAGGTGAAGCA
*
N V N M F D E P L R C E R K *
M N R F A V S A S R P Q C G Q V K Q
ATTGGACAGGTGGACCAGCTCTGGGATTTTCGCCGTAAGCAGCGGAACCAGATCGTCCGAGTCAGGGAAC
L D R W T S S G I S P *
AATTGCATGGGATGATTGGCGCGATGCTCTTCCGTTTCTATCTGATGATCGTAGCATATGTCACATAGGT
TATAGTTTACGCAGCGCAAAACCCGGAAGCGGCTACCCGGAAGATCGTCAATTTGACAGCCGTTGCAGCA
AATGTTGCGGTGACCCATCCTCGATCCTTGTGCTAACTTCGTTCTTAGCGAATATACCTAAACTCTGGTC
GAAAAACGGGCAAAAGGGTTCTAAATTTTAAACAAATCGAATCTATACAATGACATGATTTTCATAGATG
M I F I D
TAAATATTTTCGGGCGTTTGGTAATAACATATAATTTCTACTTTTTAAACTATTTTTCAACCTAA
V K Y F G A F G N N I *
ATGTATCAAACCTTACTCAAGCTTTCTACATATTTTCTTTTGCCTTTTTTAAACAAATGTTTATTTTTGGAA
M Y Q T Y S S F L H I S F A S F K Q M F I F G
CCTTAAAGTACAGCGAAAAAAGTTAAATAAACATGCACTGTACTATTTTGTCCGGTGAGCACAGCTATC
T L S T A K K K L N K H A L Y Y F V R *
TACGTGACAAATAACGATAATGCAATTTGTTTTTCCAAACCTATATGAAACGAGAAATAAGCCCCCATTAT
M Q F V F Q T Y M K R E I S P H Y
TGATAAGGAACTAAATCTTTTATCTAAATCTTTTGGACAGTAGAAGATAACTCAAAGGATATTTTATATTT
*
GCTCAAATCAGAATGCAGTTATCGAAAATTAACCTCCCTTGAGCTGAAAATTTAGCTTGAATTTATAAT
M Q L S K I K L P L S *
TATACATATGTCCAAATGTAATACAGCTTCGTGTACACATAAAATTCAAAACCCAAAAATCCACAAAA
TTGCTATTTTAGTAGGTATTGCG

Para el cluster de sORFs 2L_106, el programa Augustus Gene Prediction predijo, para el gen 1.1, una estructura compuesta por 2 exones de 125 pares de bases el primero y el

segundo de 34 pares de bases, y 1 intrón de 86 pares de bases; para el gen 2.1 predijo una estructura de 2 exones, el primero de 41 pares de bases y el segundo de 314 pares de bases, y un intrón de 68 pares de bases; para el gen 3.1 predijo una estructura de un exón de 1394 pares de bases; para el gen 4.1 predijo una estructura de 2 exones, el primero de 42 pares de bases y el segundo de 314 pares de bases, y un intrón de 68 pares de bases; para el gen 5.1 predijo una estructura de un exón de 1393 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido FRPLNSSNSFRTCANNTLDSSLISSSLVL con una probabilidad de $\rho=0.042$ y un valor $E=4.2191514764815e-05$; para el péptido FNSEPTKILER con una probabilidad de $\rho=0.042$ y un valor $E=3.30202273667463e-07$.

Evidencia de ESTs:

gi|133847614|gb|EL881080.1|;

```
CTCTTTCAGATGTTCAAATGGTTTTACTTAAAACATTTGTAATATAAATAAAAAGCCAAAATATAAAAGCC
ATATTATAAATAAAGCCAAATTTTACAAAAATTTTGTTTTTCTTTGAAATATGTAACGCAACTAAAAA
GCGGATTCAACATGATTGAAAAACAGATACAATTTGAGTATCCTAATCAATGGCCTCCTCAGCAGCCA
CCCTTTTGTGCTCCTCGATCTGCACCTTGTCCAGATCGACTCCATCTTTTTTGGCTGCCCGTATCTCGCC
TTCCCCCTCCGATCCGTCTAGCTGGTTGGTCTGCCAACTCCTGTCCACGGCAGTGACATCACTGACATG
ACCTTGACCAGACTGCTATGGTTTTCCGGTCTAGATCACTCTCCCCATTTCTTCCCGTATTCGTCTGT
CTCCATCCGGCAACTGCAACTCCTCGCAGCACAGCATGGAATCAGCAAGTGCTCAACGAAGAGGGTTTCG
CCGCAGCCTTTGAGATTTCGTCCGTATCGTGGCACTGCACTTTTTTGTGTGCAATCCACTGAAGTAGAACA
TATCTGTGCTCGTGTTCGAATGGCCCCCGGAACCAAGCCTAACGCCAGGGACTCTGAATTGGCTG
CCTCCTCTGGAGGATCATTTGTTGCTTAGGAGCTGGGGATCGGAGCTCACACGTCCAGTTGACAGCCATAC
GGATATATCGAATATCCTTGAATTCGGAGTTGTTTGAACCTGCGATTG
```

gi|133847726|gb|EL881190.1|;

```
GAAATATGTAACGCAACTAAAAAGCGGATTCAACATGATTGAAAAACAGATACAATTTGAGTATCCTA
ATCAATGGCCTCCTCAGCAGCCACCCTTTTGTGCTCCTCGATCTGCACCTTGTCCAGATCGACTCCATCT
TTTTTGGCTGCCCGTATCTCGCCTTCCCCCTCCGATCCGTCTAGCTGGTTGGTCTGCCAACTCCTGTCC
ACGGCAGTGACATCACTGACATGACCTTGACCAGACTGCTATGGTTTTCTGGTCTAGATCACTCTCCCC
ATTTCTTCCCGTATTCGTCTGTCTCCATCCGGCAACTGCAACTCCTCGCAGCACAGCATGGAATCAGC
AAGTGCTCAACGAAGAGGGTTTCGCCGAGCCTTTGAGATTCGTCCGTATCGTGGCACTGCACCTTTTTGT
GTGCAATCCACTGAAGTAGAACATATCTGTGCTGCTGTTTGAATGGCCCCCGGAACCAAGCCTAAC
CGCCAGGACTCTGAATTGGCTGCCTCCTCTGGAGGATCATTTGTTGCTTAGGAGCTGGGGATCGGAGCTC
ACACGTCCAGTTGACAGCCATACGGATATATCGAATATCCTTGAATTCGGAGTTGTTTGAACCTGCGATT
GGCGGTGGTGCGCATGTGATTACGGGTAAATATTCGGCGTAGTGCGGTGTGTGCGGCTCTAGGTAATT
GGCGTGATCGATGTGGTTGTGCAGGAGATGTTTTGGAAAGTGGNGTATG
```

gi|133847951|gb|EL881415.1|;

```
ATGTCATTGTATAGATTCGATTTGTTAAAATTTAGAACCTTTTTGCCGTTTTTTTCGACCAGAGTTTAGG
TATATTCGCTAAGAACGAAGTTAGCACAAGGATCGAGGATGGGTACCAGCAACATTTGCTGCAACGGCTG
TCAAATGACGATCTTCCAGGGTAGCCGCTTCCGGTGTTCGCGTGCCTGAACATAACCTATGTGACATA
TGCTACGATCATCAGATAGAAAACGGAAGAGCATCGCGCAATCATCCCATGCAATTGTTCCCTGACTCGG
ACGATCTGGTTCCGCTGCTTTACGGCGAAAATCCAGAGCTGGTCCACCTGTCCAATTGCTTACCTGTCC
ACATTGCGGCCTACTTGCGCTCACAGCGAAGCGGTTTCATCGAACATGTTTACGTTCAACATCGGGTGTCC
AATGAATATGTGGTCTGCCCATGTGCGCCGACTGTCCGGCCGCGCAACTGGTGGCCATACGCCACCTTT
CCAAACATCTCCTGCACAACCACATCGATCACGCCAATTACCTAGAGCCCGACACACCGCCACTACGCCG
AATATTTACCCGTAATCACATGCGCCACCACCGCAATCGCAGTTGCAAACAACCTCCGAATTCAAGGATA
TTCGATATATCCGATATGGCTGTCAACTGGACGTGTGAGCTCCGATCCCCAGCTCCTAAGCAACAATGATC
CTCCAGAGGAGGCAGCCAATTCAGAGTCCCTGGCGTTAG
```

gi|133847727|gb|EL881191.1|;

```
CAGAGTTTAGGTATATTCGCTAAGAACGAAGTTAGCACAAGGATCGAGGATGGGTACCAGCAACATTTGC
TGCAACGGCTGTCAAATGACGATCTTCCAGGGTAGCCGCTTCCGGTGTTCGCGTGCCTGAACATAAC
TATGTGACATATGCTACGATCATCAGATAGAAAACGGAAGAGCATCGCGCAATCATCCCATGCAATTGTT
```

CCCTGACTCGGACGATCTGGTTCCGCTGCTTTACGGCGAAATCCCAGAGCTGGTCCACCTGTCCAATTGC
TTCACCTGTCCACATTTGCGGCTACTTGGCTCACAGCGAAGCGGTTTCATCGAACATGTTTACGTTCAAC
ATCGGGTGTCCAATGAATATGTGGTCTGCCCCATGTGCGCCGGACTGTGCGCCGCCAACTGGTGGCCAT
ACGCCACCTTTCCAAAACATCTCCTGCACAACCACATCGATCACGCCAATTACCTAGAGCCCCGACACACCG
CCACTACGCCGAATATTTACCCGTAATCACATGCGCCACCACCGCCAATCGCAGTTGCAAACAACCTCCGA
ATTTCAAGGATATTCGATATATCCGTATGGCTGTCAACTGGACGTGTGAGCTCCGATCCCCAGCTCCTAAG
CAACAATGATCCTCCAGAGGAGGCAGCCAATTCAGAGTCCCTTGGGTTAGGCTTGAGTTCCGGGGGCCA
TTCGAACACCACGACGACAGATATGT

gi|133847950|gb|EL881414.1|;

GAAAATATGTAACGCAACTAAAAAGCGGATTCAACATGATTGAAAAACAGATAACAATTTGAGTATCCTA
ATCAATGGCTCCTCAGCAGCCACCTTTTTGTGCTCCTCGATCTGCACCTTGTCCAGATCGACTCCATCT
TTTTTGGCTGCCGATCTCGCCTCCCCCTCGATCCGTCTAGCTGGTGGTCTGCCAACTCCTGTCC
ACGGCAGTGACATCACTGACATGACCTTGACCAGACTGCTATGGTTTTCTGGTCTAGATCACTCTCCCC
ATTTCTTCCGATTCGTCTGTCTCCATCCGGCAACTGCAACTCCTCGCAGCACAGCATGGAAATCAGC
AAGTGCTCAACGAAGAGGGTTCCGCCGACGCTTTGAGATTCGTCCTGTGCGTGTTCGAATGGCCCCCGGAACTCAAGCCTAA
CGCCAGGACTCTGAATTGGCTGCCCTCTGGAGGATCATTGTTGCTTAGGAGCTGGGGATCGGAGTTT
ACACGTCCAGTTGACAGCCATACCGATATATCGAATATCCTTGAATTCGGAGTTGTTTGCAACTGCCATT
GGCGGTGGTGGCGCATGTGATTA

gi|14692315|gb|BI214591.1|;

ATTCACATTAGCAAGACTAACCAGAGTATCAGCACCATGAAGTTTTTGATCCTCTTCATTGCCCTATTCCG
CCGCGCCCGCTGCCCTACCTCAGTTTTGGATTTCGGCGGTGGATTTCGGTGATCAGCAGCAGCAGCAAGG
AGGATTTGGAGGTGGATTTGGTGATCAGCAGCAGCAGCAAGGAGGATTTGGCGGTGGATTTCGGAGGTGGA
TTCGGTGATCAGCAGCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTCGGAGGTGGATTTCGGTGATCAGC
AGCAGCAACAGCAACAAGGCTGGTAAAAACCAATTCATAGCACCGTCTTAACGCAGATACCTCAATAAAAA
ATATGACACCTTTATGATAAAAAACAATTTGGAATTTTTTACTTTTATCGTTCTCGAAAATTTATGAGTTT
TATGAAAATAAATTTATATTTATTTATTCATTTTATTTTAAATAGTGCCAGAAATAAATACACTTAA
CATAAGGGACAACGGGAAATATCTTCTCTTTCAGATGTTCAAATGGTTNTACTTAAAACATTGTAAATAT
AAATAAAAGCCAAAAATATAAAAGCCATATTTATAAATATAAGCCAAAAAAAAAAAAAAAA

gi|14639167|gb|BI173360.1|;

ATTCATTAGCAAGACTAACCAGAGTATCAGCACCATGAAGTTTTTGATCCTCTTCATTGCCCTATTCCG
CGCGGCCGCTGCCCTACCTCAGTTTTGGATTTCGGCGGTGGATTTCGGTGATCAGCAGCAGCAGCAAGGA
GGATTTGGAGGTGGATTTGGTGATCAGCAGCAGCAGCAAGGAGGATTTGGCGGTGGATTTCGGAGGTGGAT
TCGGTGATCAGCAGCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTCGGAGGTGGATTTCGGTGATCAGCA
GCAGCAACAGCAACAAGGCTGGTAAAAACCAATTCATAGCACCGTCTTAACGCAGATACCTCAATAAAAA
TATGACACCTTTATGATAAAAAACAATTTGGAATTTTTTACTTTTATCGTTCTCGAAAATTTATGAGTTTT
ATGAAAATAAATTTATATTTATTTATTTATTCATTTTATTTTAAATAGTGCCAGAAATAAATACACTTAA
ATAAGGGACAACGGGAAATATCTTCTCTTTCAGATGTTCAAATGGTTTACTTAAAACATTGCAAATATA
AATAAAAGCCAAAAATATAAAAGCCATATTTATAAATATAAGCCAAAAAAAAAAAAAAAA

gi|133847615|gb|EL881081.1|;

GTAAAAATTTAGAACCTTTTGGCCGTTTTTTCGACCAGAGTTTAGGTATATTCGCTAAGAACGAAGTTA
GCACAAGGATCGAGGATGGGTCACCGCAACATTTGCTGCAACGGCTGTCAAATGACGATCTTCCAGGGTA
GCCGCTTCCGGTGTGTCGCTGCGTGAACCTATAACCTATGTGACATATGCTACGATCATCAGATAGAAAC
GGAAGAGCATCGCGCCAATCATCCATGCAATTTGTTCCCTGACTCGGACGATCTGGTTCGCTGCTTTAC
GGCGAAATCCCAGAGCTGGTCCACCTGTCCAATTGCTTTCACCTGTCCACATTGCGGCCACTTTCGCTCA
CAGCGAAGCGGTTTCATCGAACATGTTTACGTTTCAACATCGGGTGTCCAATGAATATGTGGTCTGCCCCAT
GTGCGCCGGACTGTGCGCCGCCGAACTGGTGGCCATACGCCACCTTTCCAAACATCTCCTGCACAACCAC
ATCGATCACAGCGAAGCGGTTTCATCGAACATGTTTACGTTTCAACATCGGGTGTCCAATGAATATGTGGTC
TGCCCCATGTGCGCCGGACTGTGCGCCGCCGAACTGGTGGCCATACGCCACCTTTCCAAACATCTCCTGC
ACAACCACATCGATCACGCCAATTACCTAGAGCCCACACACCGCCACTACGCCGAATATTTACCCGTAA
TCACATGCGCCACCACCGCCAATCGCAGTTGCAAACAACCTCCGAATTCAGGATATTCGATATATCCGTA
TGGCTGTCAACTGGACGTGGGAGCTCCGATCCCCAGCTCCTAAGCAACAATGA

gi|114061072|dbj|BP556658.1|;

TTTACAAAAATTTGGCTTATATTTATAATATGGCT

TTTATATTTTGGCTTTTATTTATATTTACAATGTTTAAAGTAAAACCATTTGAACATTTGAAAGAAAAA
TATTTCCCGTGTGCCCTTATGTTAAGTGTATTTATTTCTGGGCCCTATTTAATAAATATAAATGAATAAT
AAATATAAATTTATTTTCATAAAACTCATAAATTTTCGAGAACGATAAAGTAAAAAATTTCCAATTGTTTT
TTTATCATAAAGGGGTCATATTTTTTATTGAGGTATCTGCGTTAAAACGGGGCTATGAATTGGTTTTTACC
AGCCTTGTGCTGTTGCTGCTGCTGATCACCGAATCCACCTCCGAATCCACCTCCAAATCCTCCTTGCTG
CTGCTGCTGCTGATCACCGAATCCACCTCCGAATCCACCGCAAATCCTCCTTGCTGCTGCTGCTGATCA
CCAAATCCACCTCCAAATCCTCCTTGCTGCTGCTGCTGCTGATCACCGAATCCACCTCCGAATCCACTTC
CAAATCCTCCTTGCTGCTGCTGCTGCTGCTGATCACCGAATCCACC

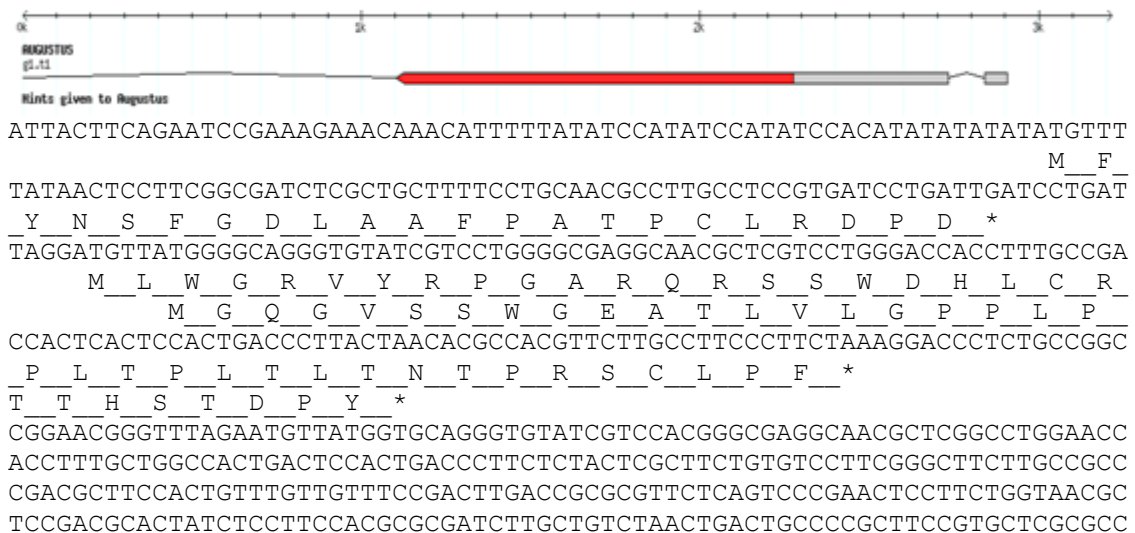
gi|114061073|dbj|BP556659.1|

GATTACACATTAGCAAGACTAACCGAAGTATCAGCACCATGAAGTTTTTGGATCCTCTTCATTGCCCTATTC
GCCGCGGCCGCTGCCCTACCTCAGTTTGGATTTCGGCGGTGGATTTCGGTGATCAGCAGCAGCAGCAGCAGC
AAGGAGGATTTGGAAGTGGATTTCGGAGGTGGATTTCGGTGATCAGCAGCAGCAGCAGCAAGGAGGATTTGG
AGGTGGATTTGGTATCAGCAGCAGCAGCAAGGAGGATTTGGCGGTGGATTTCGGAGGTGGATTTCGGTGAT
CAGCAGCAGCAGCAGCAAGGAGGATTTGGAGGTGGATTTCGGAGGTGGATTTCGGTGATCAGCAGCAGCAAC
AGCAACAAGGTGGTAAAAACCAATTCATAGCACCGTCTTAACGCAGATACCTCAATAAAAAATATGACAC
CTTTATGATAAAAAACAATTTGGAATTTTTTACTTTTACTCGTTCTCGAAAATTTATGAGTTTTATGAAAAT
AAATTTATATTTATTATTTATTTATTTATTTAATAAGTGCCAGAAATAAATACACTTAACATAAGGGA
CAACGGGAAATATCTTCTCTTTCAGATGTTCAAATGGTTTTTACTTAAAACATTTGTTAAATATAATAAAG
CCCAATATAAAGCCATATTATAAATA

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 77 sORFs. Se pueden observar varias repeticiones. La transcripción se basa en la presencia de 10 ESTs en la base de datos pública y la probable traducción en la presencia en bases de datos de 2 péptidos definidos por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos 2 de los sORFs codificados por este transcrito son traducidos.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_1



TTTTATAGCCGCAGGGATCCTGCATCCTTCTCAGCCTTCACGTCTCAGCCTGCCGTTAGGCATCTCGAC
GGGGCATCTTGACCCCTTCACATATCCTTCTCCGTTCTCCTCTGTCCCTCGCGGTCTTCTTTACCTTCTAT
CTGCTGCTGCTGTCTGCTCTGCTTACCAGTTCGATATTTGAATTTTCCCGCCTTCCATTACCCGCTCAT
CAATTTTTCCGCCTGGCAACATTGCCGAACCTCGTGTATCGATATCGGTAGGCGGGCGTTGCCACTTCGCT
TCTCCCGGGTCTCCTCCTTCGGGTTCGCCCTCATTCGGGGTCTCCTTCTCCGGCCTCTTATCTTCTGCTC
GGGCGGTAGCCGCCGATCGCTTCCGCCTCCTCCAGACACGGGCTTCCGGCGCGTCTCCGGCCTCCAT
CGACGCGGGCCACGTCCACGTACCGACTTTTTGTGCCACAGACGCCCTCCCATCTTGACCGTGGCGACC
ACCTGTCCGACGTGTGCCTGTAGAATCTCCACGCCCTGCTCTCTGTACATACCGTGGGCTCTCCCCCTCCT
GGGGTGGTGGCGTCAGCAGCGGCGCTGCGCAGGCGTTATCCGTGGTGGTGTCCGTGGCGGTGACTGCGT
TGGCGCGGGCGGTGTCTGTGCCAACCGCTCCTGCCATGCCGCTTGGCGTCCGGGACCTTTCCGCGCATC
M_R_P_W_R_R_R_P_F_A_H
CACCATCTGGAAACGCCGAGGGTGGCGACTCGCGCATCCATGAATCTTTCATGTCCACCGGATCCTCG
P_P_S_G_N_A_A_G_W_R_L_A_H_P_*
M_N_L_F_M_S_T_G_S_S
GTCGGCGAGCGACTCCGAGGCTTATCCGTCCCTGGCCCTGCGTAGCCCGCCGGAGGCTTCGGTAAGCCA
V_G_G_A_T_P_R_L_I_R_P_W_P_C_V_A_R_R_R_L_R_*
TCCGCGGGTCTCATCCTCGTCCGATGACACCTCCGGGGTACCCTTGCCTCCGTGCTGCTTCCCTC
ACTCTCCGAAGAGAGGGTGTGATGACCTCAGCCGCGCCCTCTGTATCCGCCATCGGTTGGTCTTCCCTAAGT
M_T_S_A_A_P_S_V_S_A_I_G_W_S_S_L_S
TCCACGTCCGGAATTTCTCCACACTCCCCGGGGTGGAGATCGCGACACCAGGGGTGACGCCAGGAGCT
S_T_S_G_I_S_S_T_L_P_G_G_D_R_D_T_R_G_D_A_R_S
GGAGATAGGAATCGGAATCCCTATTCATATCGCCCCGGGCTGTTCGCCGTATCGCTGCTTCCCTGCT
W_R_*
GACCTGGTCCGCCAGATATGTCATGAAACCTCTAGGCAGGAAAAATCCTGGAAGACCCATGTCTCCAG
M_S_P_R
AATTTGTTGTTGGCAATAAACCACTTCAAGGCCCTGCCCGGTAGTAGTTTCGGCATCGCTCTTGGGATCT
I_C_C_W_Q_*
GGTTGATGTGATTCCATATGTCATTGCTGACCATTCCAGCTGCTCTAGGAACCTCAGCGGCTTCTCGTT
M_S_L_L_T_I_P_A_A_L_G_T_S_A_A_S_R
CCCGTCTAACCAGAACTCCACTCACGCACCTGTTTCGCCACCTTTGCGTAGTCTTCCGGTGTGGCCTT
S_R_R_N_G_N_S_T_H_A_P_V_S_P_P_L_R_S_P_A_V_L_A_L
GCGGTCTCCCGCTATCCCGGTGACTTCCGCGCTTCCCTTCTTTCGCGGCTGCCGTCTCGGCGTTCCT
R_S_P_A_Y_P_G_*
GTCCGATTTCTTGCCTTTCAGACCTGCTTCCGCTGGCCTCTGCCATTCTAGCGAAGTCGGGGCTCCTCAC
CAGGCCCTCTACGCGGTGACTTTCAACTCTACGCGCGGTCTCCTTCCATACTCCTTTTCCAGGGCTTCT
ATAATTGCGACAGATTCCGGCTCATCTGTGGTCTGGTTCGATGAACCTCGGACAACGTCTTTCGCATTTCT
M_N_S_D_N_V_F_R_I_S
CCACCAAACCTTCCAGCTTTATCCCGAGTGTATCTTTCTTCAAGCTGTATAGCCACGACTTGCCCATCT
S_T_K_P_S_S_F_I_P_S_V_I_F_L_Q_A_V_*
TGCTCTGCTTTTCTTTTGGCCGGACAATCACGCTTGGGCGCCAGATGTAACGAACCTGATTTTCTTATTTT
GCTCGCTACGGATTGCAGTGGCTTGTCTCAGAGAGTTTTTGGCTTGTCCACCAAATATGATTTGTGT
M_I_C_V
GATTGCCCGATCACGAAAAGCAATGGGTTCTTAGTATAGTATTTTATTTCTACAGCCTTATGGATAACT
I_A_R_S_R_K_S_N_G_F_L_V_* M_D_N_F
CTTCCGCAATCTCGGTGCTACTGCGGCGTTTCTCCTGCAACGCCTTGCCCTCCGTGGTCCCTGCTTGTCTCT
F_G_N_L_G_A_T_A_A_F_L_L_Q_R_L_A_S_V_V_L_L_A_P
TGACACGGACGACGACGACGGGTAGGGTGTATCGTCTCGGGCGAGGCAAACGCTCGTCTCGGGACCAC
*
CTTTGCCGACCACTGACTCCACTGTCCCTTTTCGAACACGCTACGTCGTTATCGTCTCCTCTGGAAGACCA
CCTGCCGGTCCGATGGGTTTAGGATTATTATGGGGCAGGGTATATCGTCTCGGGCGAGGCAACGCTCG
TCCTGGGGCCACCTGTACCGACCACTGACTCCTCTCACCTTACGAACACGCCAAGTTCTTGCCTTCCCT
TCGGAACGTCCCTCTGCCGCGCGAACGGGTTTAGGATGTTATGGTGCAGGGTGTATCGTCCACGGGCGA
M_V_Q_G_V_S_S_T_G_E
GGCAACGCTCGGCCTGGAACCACCTTTGCTGGCCACTGACTCCACTGACCCTTTCTGACACGCTAGATCC
A_T_L_G_L_E_P_P_L_L_A_T_D_S_T_D_P_F_*
TTCTCCTCCACTCCAGGATCTCTCCTCGCTTGTGGTGTATCGTCTTCCATGTCCTTCGGGCTTCTTGGC
GCCAGACGTTTCCACTGTTGGTTATTTCCCACTTGACCGCGGTTCTCACTCCCGAACTCCTTCAGTTAA
CGTCTGATGCACGTTCTCCTTCCACTCGCGATCTGGCTGGCTGACTGAATGACCCTCATCCGT
M_H_V_L_L_P_L_A_I_W_L_A_D_*

Para la región genómica 2R_1, el programa Augustus Gene Prediction predijouna estructura compuesta por 2 exones de 1629 pares de bases el primero y 66 pares de bases el segundo, y un intrón de 108 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido MSLLTIPAALGTSAASR con una probabilidad de $p=0.057$ y un valor $E=0.000525837993509083$.

Evidencia de ESTs:

gi|10609532|gb|BE978250.1|;

```
GCACGAGACGGTACAGATGGTCCCAGGCCGAGCGTTGCCCGCCCGTGGACGATACACCATGCTCCATAA
CACCTAAACCCATACCGACTGGCAGGTAGTCTTCCAGCGCAGACGACAGCGACATAGCGGGTTTCGAGAG
GGACAGTGGAGTCAGTGGTTCGGCAAAGGTGGTCTCCTGACGAGCATTGTCTCGCCCCAGGACGATACACC
CTGCCCCATAACATCCTAGTCCCAGGCCGCGGCTCGTCTGTCCTGGTCAAGAAGCAATCAGGACCACGG
AGGCAAGGCGTTGCAGGAAAAGCGACGCGAGGAGCAGCGAGATCGCCGAATTAATTATATATAACAACAGTA
GAATAAACGGCTATCATANGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

gi|13690304|gb|BF498414.2|;

```
CTAGAACCAACCAATTCTAAAATTAACATTTCAACACTGTGAGGAGTCTTAAAACCTTACAAATCTCG
CGTAGGGGAGTGGCTAGCAACAGCCGCGGTGCAGGGCAAAGCGGAAAGACCGTGAATTAACGGTACCGC
TGTGGACCTTGGACTCCAGCGGGCCAAGGGAGGGCAGGTGCAACGGTGCAGGTGCGTGCAGCAAGGAGGAC
GCACCGTGAATTAACGGCACAATATGTGGACCTTGTACCCGAGCGGGCCGTGCGCAAAAAGAGAAATAGC
GGGATCCCTGCGGCCTATAAAAGGGTAGGCGCGAGCACGGATGAGAGGCAGTTAGTGAGACAGAAAGGTTCG
CGCACGAAAAGACAGAGCGTTCGCAAGTGTAAACGGAAGGAGTTTCGGAAGTGAGAACGCGCGGTCAAGTC
AGAAGGAAAAGGCACAGTGGAAAGCATCGGGCGGCAAGAAGCCCGAAGGACACGGAAGGACGGTACCACAA
GCAGGTAGAGAACCCTTGGAGTGGAGGAGAAGGCGTGTGTAAGGATCAGTGGAGTCAGTGGCCAGCAAAG
GTGGTAAACGGCTATAATAAGAACCATTGCTTTTCTTGGTTCAGGCAATCACAAAATCTAAATTTGGT
GGGGCGAACGCACAAACTCTCTGAGCTAGCCGCTGCAATCCGTAGCGAGCGAAATAAGAAAATCAGTTTG
TTACATCTGGCGCTAAGCGTGATTGTCCGGGCAAAGGAAAACAGAGCAAGAT
```

gi|133845428|gb|EL878894.1|

```
TCTCATCAGAAAATTAATGAAATTAATTTTATCCATATTGTATTCTGTGCAACTAGTGTAAAAGTTTTCA
ATTCAAAAATCAATTCAGAGTTCAGTGTGAGCCTTCCGACCAGAGAGCGCGCTCTTGCTCTCAGGGATTG
CCAAACACAGTTCACACAAAAGACCCTAGAATAACCTAGAAGACCCTAGAATAGCTTGAAGCAATACGAAA
AAGAGTGTCTCAAATTTGCAAGATACAGACAAATACGGCAGAACAAATATGGTAAATTTACCAACGCA
AAGGATTACATTAATACTGGAGTAACTACGCGGGGCCATATCATATAAGGTGCTCAACAAATCGTAAT
CAGAAAATCGTATAAGGGTTATACTCCTGGGAGTGGAGGAGAAGGATCTGGCGTTTTCAAAAAGGGTCAGTG
GAGTCAGTGGCCAGCAAAGGTGGACCCAGG
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 13 sORFs. La transcripción se basa en la presencia de 3 ESTs en la base de datos pública y la probable traducción en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORFs codificados por este transcrito es traducido.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_2



TAACAATTGCGTGTATTTTACAAACAGTACGATTACTTAAGCTAATAGTTCGAAAGGTCTTCGGGGGAACA
 GTCCGTAGTGCATCCGCTTTCATAGAGTTCTAGTGATGACGGGGGTGCTCGGGTGCCTGGTGGGTGGGA
 CTCCAGCCAGGCGACGGCGCGGGCGATGGCCTCTGGGATTGGAGGAGGAGTGGGGATCCAGGCTCCCAG
 GGCTGGTATCCGTTCTCATTGGCGACGTACTTAACCTCGACGTGCTCGCCCTCGGGTGGATCCAGCCGA
 AGTTGCCGTGGATGTTGCCATGGGCATCACCGCTGGCGGCCCTGCTCGATTCCGTTGGAGGTGTGCAGGCT
 M G I T A G G L L D S V G G V Q A
 GGAATCGAATCCATCGGCACGAAACATCATCGGATTGGGAAAGGACATCGGCGTGGACATCCTCCGAACGG
 G I E S I G T N I I G L G K D I G V D I L R T
 CCTAGGGAATGGGGCACCGGGGGTGGCCACCGCCAGGCCAAAACCTGCGCAGATCATGACCTGAAGGA
 A * M G H R G V G H R Q A Q N C A D H D L K E
 GAAATAGGCATTGTTAGGGATCGAGTATGTTGGTTCGTATCCTGCGGACACTTACAACTGAACATAT
 K * M L V R I L R T L T N L N I
 TGGCTGACTTCTGTCTGGTTCGAGAACGAACGTTGACTGATAACTGATTGTGGTTCGGGCAGAGCCTTTTTA
 L A D F C L V E N E R *
 TATGATTGCCACCCAAAGTGAAACCATATCAAAAACGCTACATCTTGCATTCAATTGGTTCGCTGAGTTAT
 GGGTTGACTTCGCGGATGCGCCACACCCATTAAAAAGCATCGTCACATGCGACAGCTTAGCAAATGCCA
 M R Q L S K C Q
 M P
 AACAAACGGGATAAGTACTATATATAATCGCTTAATAGATGCTAAAATGAAACAATCGCATCAATAAGGGG
 T T G *
 N N G I S T I Y N R L I D A K M K Q S H Q *
 CAAACGGGTTCGGGTGGCATCCCAGAAATCGAAACGAGTTTGACCCGCCCTTCATTAATCCCCTCCTTTG
 AGCTTGACTTGGCATTGCCAGGCCGTGAACCACTTTTGTAGTCCCATCAAGACTCCATGGACTACAAAGC
 M D Y K A
 ATCGCCGACTTCTTAGACTCGATGTTTTGTAATATATAAACGTGTGTCTCTACCTTGGCACCATTTCTTC
 S P T S *
 TTTTTTTATCAATTCCTCTCGACGAACAGGTTGCAGTTTGCTTTTAAACAAATTCAGTGTGTTTTTACA
 TTATATTTTCATATAGCATAGCGCCTTATATTTTTTTTTATATTTTCGTAAAATTAATAAGATAATAAGAC
 M H S A L Y Y F F I F R K I K *
 CTGTTCCATATAGGAGGCTACCGAAATACATAGTTTTTTTAGTGACAATTAATAATGTTAAGAAAACCTAA
 TGTTAAACCTAAAAGAGGATATTTACTTGTCTGTATAGGCATATAAACTAAAAGTTTACTTAAAATTTATAA
 TTCTATAGATACATTTTCATGGTATTTAAATGTGCAAGCCGAACATCTAATATAAGAGCTTGAATAGTTT
 M V F K C A S R T S N I R A *
 GTTTTTCAGATCAGTTAGACCGCACAGTATCCTCACAAATTTGGTTTTGAAAGTTAAGATTAATTTGTAAT
 CAAACAGAATAGCCTCTTCTGAGATTTATGCTAGTGACTATCGCACCTTTAACAATCCAAGCTTCTAATA
 M L V T I A P L T I Q A S N
 TTAAGTGTGACGAAAATAGAGTTTTTATTCACATAAGACAATACAATATCTAAAAGTTACTCGTTTCG
 I N C C D E N R V F I Q L R Q Y N I *
 GGACGCTTTATGACATTATTTATTTGAGGTGCATTTGCAGACAGACAAACAACCAACAAATCCCAGTC
 TTATTTTAAAAAGAGAGAAATGCTATACTCAGCTGGAAGTGTGACCGATATATTTCAATACTTTTTTGAAT
 M L Y S A G S V T D I F Q Y F F E
 ATCGGTAAAAATTTGAAAACAAATAAATAATGAAAAAATTTATGAATTGTTCAAACGTGTGGCGCTG
 Y R * M K K I I E L F K T C G R
 GCAGTTTTAGCCGTTTGTGGGCTCTGGAATCTTCATATCATCTAATCTTAACTCTCTAGTTTTTTGAAG
 G S F R P F V G L W N L H I *
 TCGAGGCGTTCATACGGACATACTCTGCTATTGATCGCTTTTTAAAAATATATATACTTTATATGGTGTGA
 AACGCTCCTTCTACCTGTTACATAATTTTCCCTAGTATTCATTTTACTTTTCGAGTAACGGGTATTATCA
 TTTTCAGTTTATCTTCGTATGTTTTTATCTCATTATCTTCTTAGATTATTTTACGGTTCATCTCTTTACT
 TTTTATGCTGCCTTTCTATTGGTTATCACACAATATAATAAGACAACCTACAGCAGATAGAAGTTGTCTTT
 GTACTCTCTGGTGCCTTCGTCACCTACGTGGAATGCCGTTTCATGGTGATTCTTCTCTCTTTTTGAAACCA
 M P F M V I L L S F *
 TTTACTTATGTCGGGTAAAAAATTTATTTGAATAATATATAACATGTCGATGGAAGTCCGGTGGATGGGT
 M S M E S R W M G
 CTGAGATCGGGGTGAAGAGTCTTGGTGTGGTGGGATGCTCGGGTACCGGTTCAGTGAACGCCTTCGGG
 L R S G * M V R M L G Y R S V E R L R
 ATTGGTGCAGGAGTCGGAATCAAGGCTGATCGAAGCCTGGTATCCGTTCTCACCAGCGTCGTACGTGACC

D_W_C_R_S_R_N_Q_G_*
 TCGACGTGATCGCCCTCGGGGGTGATCCATTTTATAGTGCCTGAAAAATACCATCGATATCTGCTTTGG
 CGGCCCCGCTCGATTCCGTTGAATATGTGCAGACTGCTGGAATCGAATCCATCGGTATAAACATCGTATGA
 M_C_R_L_L_E_S_N_P_S_V_* M
 TTGGTAGTGGACATCGGGCGTTGTCACCTTTCGAAACGTCTACGGATGGGGGATGGGGGATGAGCCACTCA
 I_G_S_G_H_R_R_C_H_F_R_N_V_L_R_M_G_D_G_G_*
 M_G_D_E_P_L
 M_S_H_S
 ATCCAAAACTGCACAGATCACCACGTGAACGAGAGCATAGACATTGAATACGTTTCGTATCCTGCCGTTA
 N_P_K_T_A_Q_I_T_T_*
 I_Q_K_L_H_R_S_P_R_E_R_E_H_R_H_*
 CTCACGAGCATGAACATTTTAGATTCCAATTTCTGTTTGCTCGAGAACGAACGTTGACTGATAACCACCA
 M_N_I_L_D_S_N_F_C_L_L_E_N_E_R_*
 AAACTGAACCCAAAGTCAATACGCTACATCTTTCATACGATTGGTGTCTGAAATCTGGGTTGACTATGAA
 GATACCCACACCCATTA AAAAGCATCGCATGGCAATCCACGGGATTAGTACTACGAGTATATACTAGGG
 M_A_I_H_G_I_S_T_T_S_I_Y_*
 TAATCGGCTCGTGTGGCATTGACCATTGCTTTTCTATTTCAATAAGTTTGTCCATTGGTATTGGCTTCTC
 TAATTGGCTTTTAAATACTACTTAAAAATAAAAAAAAAAATGCATTGAGACTTTTGTAGTGAATAAAAATTGT
 ATGCAAACATTTGCTTTTTTGTTTTTTCAGATCTTTTAGACTGCCAGATCCTTCTCAAATTTTAAAAGAC
 M_Q_T_F_A_F_L_F_F_R_S_F_R_L_P_R_S_F_S_N_F_K_R
 AAAATTAATTTTTAATTCAGCACATCTTTTCTTTATGGACACATATACTCATTTTTTGATAAATGGATTTA
 Q_N_*
 GTGAATCGAATATAAAAAGGTAGAAGGAAGCATTCCCAGCGAAGGTAGAGGTATAAAAAGCTCGTATTATTA
 CAGGAACTACCTAAAAAGTGAATGAAGAACACACACATTTTTACTGCTATAAGCAAGTGCATGTATTTT
 M_K_N_T_H_I_F_T_A_I_S_K_C_M_Y_F
 TTTGAAAGTCAAGAAGTACAGCTAAACAACCTGGGCGATTTGTTTTCGGGAGAACAGTCCGTTGGAGTGGTC
 F_E_S_Q_E_V_Q_L_N_N_L_G_D_L_F_R_E_N_S_P_W_S_G
 CGGGATCCGGGTGACGAGTCTAGTGATGACGGGGGTGCTCGGGTGTGGGGGTGAGACTCCAGCCAGG
 P_G_S_G_*
 CAACGGCGCGGGCGATGGCCCTCGGGATTGGAGGAGGAGTGGGGATCCAGGCTCCCGAGGGCTGGTATCC
 GTTTTTCATTTCGCGACGTACTTTACCTCAACGTGCTCGCCCTCGGGTGAGATCCAGCCGAAGTTGCCGTGG
 ATGTTGCCATGGGCATCACCGCTGGCGGCCGTCTCGATTCCGTTTGAGGTGTGCAGGCTGGAGTCAATC
 M_G_I_T_A_G_G_L_L_D_S_V_*
 CGTCGGCACGAACGTCTCCGATCGGGAAAGGACATCAGCGTGTACATCATCGGAGCGGGAAACTGGGGC
 TAGGGCGGTAGCCACTCCACAACGGCGAGAATCATCACCTGGATAGAGGAAATGATACCAATTAGATGG
 CGAGTTCATAAATGTCCTGTGAGCCACTCACAACTTGAACATGTTGGCTGATTTCTGTCTGGTTCGAGAA
 M_S_C_E_P_L_T_N_L_N_M_L_A_D_F_C_L_V_E_N
 CGAACGTTGACTGATAACTGACTTTGGTC
 E_R_*

Para la región genómica 2R_2, el programa Augustus Gene Prediction predijouna estructura compuesta por 5 exones de 482 pares de bases el primero, 55 pares de bases el segundo, 235 pares de bases el tercero, 229 pares de bases el cuarto, y 27 pares de bases el quinto, y 4 intrones de 65 pares de bases el primero, 1829 pares de bases el segundo, 921 pares de bases el tercero, y 63 pares de bases el cuarto.

Se encontró acuerdo con espectros de masas para el péptido LIDAKMK con una probabilidad de $\rho=0.035$ y un valor $E=0.000668890167392407$; y para el péptido MLGYRSVER con una probabilidad de $\rho=0.059$ y un valor $E=5.70180330807161e-06$.

Se halló evidencia de ESTs:

gi|48954582|gb|CO181701.1|;

GAAATCAGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTGGCTACCGCCCTAGCCCCA
 GTTTCCCGCTCCGATGATGTACACGCTGATGTCTTTCCCGATCGGACGACGTTTCGTGCCGACGGATTCCG
 ACTCCAGCCTGCACACCTCAAACCGAATCGAGCAGGCCCGCCAGCGGTGATGCCCATGGCAACATCCACGG
 CAACTTCGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTCCGAATGAAAACGGATAC
 CAGCCCTCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCAGAGGCCATCGCCCGCCGCTTGCCTGGC

TGGAGTCTCACCCCCAGCACCCGAGCACCCCGTTCATCACTAGGACTCGTCACCCGGATCCCGGACCAC
TCCACGGACTGTTCTCCCGAAACAAATCGCCCAAGTTGTTTAGCTGTACTTCTTGACTTTCAAAAAAATA
CATGCACTTGCTTATAGCAGTAAAAAAAAAAAAAAAAAATATTAGAGCGCCGCTCTAGAGGATCCAA
GCTTACGTACGCGTGCATGCGACGTCATAGC

gi|41402352|gb|CK658827.1|;

CTCGACCAGACAGAAATCAGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTGGCTACC
GCCCTAGCCCCAGTTTCCCGCTCCGATGATGTACACGCTGATGTCCTTTCCCGATCGGACGACGTTTCGTG
CCGACGGATTTCGACTCCAGCCTGCACACCTCAAACGGAATCGAGCAGGCCGCCAGCGGTGATGCCCATGG
CAACATCCACGGCAACTTCGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTCGCGAAT
GAAAAACGGATAACCAGCCCTCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCGCCCCG
CCGTTGCTTGGCTGGAGTCTCACCCCCAGCACCCGAGCACCCCGTTCATCACTAGGACTCGTCACCCGG
ATCCCGGACCACTCCACGGACTGTTCTCCCGAAACAAATCGCCCAAGTTGTTTAGCTGTACTTCTTGACT
TTCAAAAAAATACATGCACTTGCTTATAGCAGTAAAAAAAAAAAAAAAAA

gi|38623711|gb|CK131775.1|;

GTTTCGTTCTCGACCAGACAGAAATCAGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGT
GGCTACCGCCCTAGCCCCAGTTTCCCGCTCCGATGATGTACACGCTGATGTCCTTTCCCGATCGGACGAC
GTTTCGTGCCGACGGATTTCGACTCCAGCCTGCACACCTCAAACGGAATCGAGCAGGCCGCCAGCGGTGATG
CCCATGGCAACATCCACGGCAACTTCGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGT
CGCGAATGAAAAACGGATAACCAGCCCTCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATC
GCCCCGCGCGTTGCCTGGCTGGAGTCTCACCCCCAGCACCCGAGCACCCCGTTCATCACTAGGACTCGT
CACCCGGATCCCGGACCACTCCACGGACTGTTCTCCCGAAACAAATCGCCCAAGTTGTTTAGCTGTACTT
CTTGACTTTCAAAAAAATACATGCACTTGCTTATAGCAGTAAAAAAAAAAAAAAAAA

gi|3946833|gb|AI297426.1|;

TTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTGGCTACCGCCCTAGCCCCAGTTTCCCGCTCCGATG
ATGTACACGCTGATGTCCTTTCCCGATCGGACGACGTTTCGTGCCGACGGATTTCGACTCCAGCCTGCACAC
CTCAAACGGAATCGAGCAGGCCGCCAGCGGTGATGCCCATGGCAACATCCACGGCAACTTCGGCTGGATC
TCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTCGCGAATGAAAAACGGATAACCAGCCCTCGGGAGC
GGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCGCCCCGCGCGTTGCCCTGGCTGGAGTCTCACCCCC
AGCACCCGAGCACCCCGTTCATCACTAGGACTCGTCCCGGATCCCGGACCACTCCACGGACTGTTCTC
CCGAAACAAATCGCCCAAGTTGTTTAGCTGTACTTCTTGACTTTCAAAAAAATACATGCACTTGCTTATA
GCAGTAAAAAAAAAAAAAAAAA

gi|3946548|gb|AI297141.1|;

TTCGTTCTCGACCAGACAGAAATCAGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTG
GCTACCGCCCTAGCCCCAGTTTCCCGCTCCGATGATGTACACGCTGATGTCCTTTCCCGATCGGACGACG
TTCGTGCCGACGGATTTCGACTCCAGCCTGCACACCTCAAACGGAATCGAGCAGGCCGCCAGCGGTGATGC
CCATGGCAACATCCACGGCAACTTCGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTC
GCGAATGAAAAACGGATAACCAGCCCTCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCG
CCCCGCGCGTTGCCCTGGCTGGAGTCTCACCCCCAGCACCCGAGCACCCCGTTCATCACTAGGACTCGTC
ACCCGGATCCCGGACCACTCCACGGACTGTTCTCCCGAAACAAATCGCCCAAGTTGTTTAGCTGTACTTC
TTGACTTTCAAAAAAATACATGCACTTGCTTATAGCAGTAAAAAAAAA

gi|3945292|gb|AI295885.1|;

TTCGTTCTCGACCAGACAGAAATCAGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTG
GCTACCGCCCTAGCCCCAGTTTCCCGCTCCGATGATGTACACGCTGATGTCCTTTCCCGATCGGACGACG
TTCGTGCCGACGGATTTCGACTCCAGCCTGCACACCTCAAACGGAATCGAGCAGGCCGCCAGCGGTGATGC
CCATGGCAACATCCACGGCAACTTCGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTC
GCGAATGAAAAACGGATAACCAGCCCTCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCG
CCCCGCGCGTTGCCCTGGCTGGAGTCTCACCCCCAGCACCCGAGCACCCCGTTCATCACTAGGACTCGTC
ACCCGGATCCCGGACCACTCCACGGACTGTTCTCCCGAAACAAATCGCCCAAGTTGTTTAGCTGTACTTC
TTGACTTTCAAAAAAATACATGCACTTGCTTATAGCAGTAAAAAAAAA

gi|3945124|gb|AI295717.1|;

TTCGTTCTCGACCAGACAGAAATCAGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTG
GCTACCGCCCTAGCCCCAGTTTCCCGCTCCGATGATGTACACGCTGATGTCCTTTCCCGATCGGACGACG
TTCGTGCCGACGGATTTCGACTCCAGCCTGCACACCTCAAACGGAATCGAGCAGGCCGCCAGCGGTGATGC

```
CCATGGCAACATCCACGGCAACTTCGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTC
GCGAATGAAAACGGATACCAGCCCTCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCG
CCCGCGCCGTTGCTTGGCTGGAGTCTCACCCCCAGCACCCGAGCACCCCGTCATCACTAGGACTCGTC
ACCCGGATCCCGGACCACTCCACGGACTGTTCTCCCGAAAACAAATCGCCCAAGTTGTTTTAGCTGTACTTC
TTGACTTTCAAAAAAAAAATACATGCCTTGCTTATAGCAGTAAAAAAAAAAAAAAAAAAAA
```

gi|3866732|gb|AI259207.1|;

```
CCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTGGCTACCGCCCTAGCCCCAGTTTCCCG
CTCCGATGATGTACACGCTGATGTCTTTCCCGATCGGACGACGTTTCGTGCCGACGGATTTCGACTCCAGC
CTGCACACCTCAAACCGAATCGAGCAGGCCGCGCAGCGGTGATGCCCATGGCAACATCCACGGCAACTTCG
GCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTTCGCGAATGAAAACGGATACCAGCCCTC
GGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCGCCCGCGCCGTTGCCTGGCTGGAGTCT
CACCCCCAGCACCCGAGCACCCCGTCACTACTAGGACTCGTCACCCGGATCCCGGACCACTCCACGGA
CTGTTCTCCCGAAAACAAATCGCCCAAGTTGTTTTAGCTGTACTTCTTGACTTTCAAAAAAATACATGCCT
TGCTTATAGCAGTAAAAAAAAAAAAAAAAAAAA
```

gi|48956556|gb|CO183675.1|;

```
AGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTGGCTACCGCCCTAGCCCCAGTTTCC
CGCTCCGATGATGTACACGCTGATGTCTTTCCCGATCGGACGACGTTTCGTGCCGACGGATTTCGACTCCA
GCCTGCACACCTCAAACCGAATCGAGCAGGCCGCGCAGCGGTGATGCCCATGGCAACATCCACGGCAACTT
CGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTTCGCGAATGAAAACGGATACCAGCCC
TCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCGCCCGCGCCGTTGCCTGGCTGGAGT
CTCACCCCCAGCACCCGAGCACCCCGTCACTACTAGGACTCGTCACCCGGATCCCGGACCACTCCACG
GACTGTTCTCCCGAAAACAAATCGCCCAAGTTGTTTTAGCTGTACTTCTTGACTTTCAAAAAAATACATGCA
CTTGCTTATAGCAGTNNAAAAAAAAAAAAAAAAAAAAAATATTAGGGCGGCCCTCTAGAGGATCCA
AGCTTACGTACCGCTGCATGCGACGTCATAGC
```

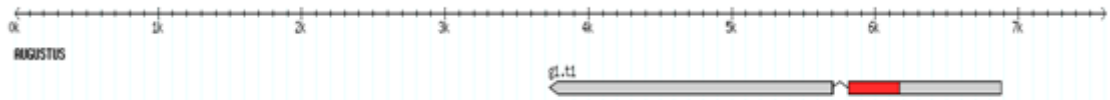
gi|49006763|gb|CO195588.1|

```
GAAATCAGCCAACATGTTCAAGTTTGTGATGATTCTCGCCGTTGTGGGAGTGGCTACCGCCCTAGCCCCA
GTTTCCCGCTCCGATGATGTACACGCTGATGTCTTTCCCGATCGGACGACGTTTCGTGCCGACGGATTTCG
ACTCCAGCCTGCACACCTCAAACCGAATCGAGCAGGCCGCGCAGCGGTGATGCCCATGGCAACATCCACGG
CAACTTCGGCTGGATCTCACCCGAGGGCGAGCACGTTGAGGTAAAGTACGTTCGCGAATGAAAACGGATAC
CAGCCCTCGGGAGCCTGGATCCCCACTCCTCCTCCAATCCCAGAGGCCATCGCCCGCGCCGTTGCCTGGC
TGGAGTCTCACCCCCAGCACCCGAGCACCCCGTCACTACTAGGACTCGTCACCCGGATCCCGGACCAC
TCCACGGACTGTTCTCCCGAAAACAAATCGCCCAAGTTGTTTTAGCTGTACTTCTTGACTTTCAAAAAAATA
CATGCCTTGCTTATAGCAG
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 23 sORFs. Se pueden observar varias repeticiones. La transcripción se basa en la presencia de 10 ESTs en la base de datos pública y la probable traducción en la presencia en bases de datos de 2 péptidos definidos por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos 2 de los sORFs codificados por este transcrito son traducidos.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_5



CCAGAAAGAAATCGTTTCGACCATACAAAGTATATATACTTACGTATTTTACTTTTTCGAGTCGAACAAAT
M
GATCCGTCGTTTACTAAGATCAACGCCTTTAAAGAAGTTTCGGATTTGAAAGAATCAACACGAGGTATT
I R R L T K I N A F K E V S D L K E S T R G I
TTTTACTTTAGATCTTTATTTCTTCGTTGTCGGCTTAAGACTAAACGTACAGAGGTTTTTGATCGGCTCTT
F T L D L Y F F V V G L R L N V Q R F L I G S
ATTGAGTCGCGAGAAACGGAAATGTTTTGTCAGCATTTTTTATTTTATTTGTTTTTAAGTGCATACAAA
Y_*
M F L S A F F I L F V F K C I Q
AAGAATGAAGAAGTTATACAATGATCATGTGTTTTACTTTGTTGCAGGTTTCAAGTTTCGTTACGTTATG
K E_*
M K K L Y N D H V F Y F V A G F K F R Y V M
M I M C F T L L Q V S S F V T L C
TTCTGTGCATGCAAGATAGTATACTTAGCGACTAATGCAATGTAAGCATAATATGAGAAGCATATTATGT
F C A C K I V Y L A T N A M_*
S V H A R_* M Q C K H N M R S I L C
M
TATGTAAGTGCAGACTGTATGTTTGGCAATTAGTGTAAGTATGTAAACGTAATATGAGAAGCATATTAT
Y V S A R L Y V W Q L V_*
L C K C E T V C L A I S V S M_* M R S I L
GTTACTTCCCTCCCTTTAAAAGGGGAAAGCTTATGCTTCCACTCCTTCTATTGGTGTGATGCCCGCAATCT
C Y F P P F K R E A Y A S T P S I G V D A R N L
M L P L L L L V L M P A I
TGGTGTATGCAGTTGGATTTGCATCTTGTGATTCAATGATTATTTTTGTTTTTTAGTTTTAAGTAAATACAA
G D A V G F A S C D S M I I L F F S F K_*
L V M Q L D L H L V I Q_*
GCAATACAAGATGTGAGGCTTATGATTGATACTATTATACAAACTGTGATATTATTATTTGATATTAATT
GTACAATGTCCTTATGTTCTATTGGTTTTATTTTTATTAGGCTTAAGGGTGTGTACAATGGGGTAAGGTT
TATCTCGGGGTTAGCTTGTTCGTTAAGTATTATGTTTCCTATTTTGTATTTTACATTGTTTTATTGTT
ATTATTTTGTGTCGTTAACCGTTACATTTTTATTTATTTCTTGGCAATTTTGTTCCTATGATAGTGTGTT
TTAGATTCCAAGTATGATACTATTTGGTTCCTACATACTTAATTTTGTTCGACTGTTATAGGGACAAA
GTTACATATTGGATTTGTTCTTTTAATAATATTTTTGATACATTCGTTGTTAATTTCTTGATAGTTCTGG
TTATAAATTTTATCTTGTCTTCGAAATAAGTGTCTTTGTCTATATGATCAAGTTGATAGCCATTATTAT
M I K L I A I I I
CTGGATAGGGTATTATGTTAATAGTATGAATGGTTTTTTGTTTTAGTAGGGATATGAGCTATGATAAGTAG
W I G Y Y V N S M N G F L F S R D M S Y D K_*
M I S S
TTGGTTAGTGTGATCCAGGTTGAGGTTTTAGTCTGTAGTATGTTGCTACTGTCAACGTTTTCTAATTTG
W L V L I Q V E V L V C S M L L L S T F S N L
TCATAGTTAAGTAATTTTGGGTTAAACAGTCCTAGTCGTGAGAGTTGCATTCCCATTGCGATGTCTTCAA
S_* M S S
TGTATTCGGTAAATTTGATATTAGTTCGTATAGCATTTGTTCTCTGTTATTATGTTTCGTCTTCAAATCTTT
M Y S V N C I S S Y S I C S L L L C S S S N S L
M F V F K F F
GATCTTTTGCATGTTACTATTGATAAGTTGAATTATGTCGCTCATTTTCGTGTATGTTAATTGAGTTTTTT
I F C M L L L I S_*
D L L H V T I D K L N Y V A H F V Y V N_*
M S L I S C M L I E F F
GTCGCGGACTCTAATTGGTTATGAATTTGTATCCGGTCGTTATTGTCTAGTGTCCCGAAAAGGTATTTAA
V A D S N W L_*
M N L Y P V V I V_*
ATGCAGATCCCATTTATGTTGATTAGTCCGCGTTTCTGTCTATTTGCTGGTGTAAACGTTATAGCGTTAAG
M L I S P R F C L F V W C N V I A L S
TTCCCTTAACATTTTATTGTGTAGAAGTTTAAATTTGGGGGATATGTGTATAGTTGGTTACTTTAGTGTTA
S L N I L L C R S L I W G I C V_*
M C I V G Y F S V

AAATAGTGAGTTAGGTTGTCTATATCCAAGAGGTTTATGCGGAGACAGTGGTGTTCGTATTCTGAAGGGA
K_I_V_S_* M_R_R_Q_W_C_S_Y_S_E_G
TTTGAATAGGTTTGTCTCAGAGAATAGGAGATATCCGTGGTGGTGTTCGATGTTGTTTATTTGGAGTTGCTG
I_* M_L_F_I_W_S_C_W
GGCTTGGCATAGTTTTGTAAGAGTAAGAAAAAGTATTATGGTGAACAAGGGGCTGTAAAATGGGAAGTT
A_W_H_S_F_V_R_V_R_K_S_I_M_V_N_K_G_P_V_K_W_E_V
GATTAGTTTTCTCTGTTTTTTAAATTGGGTCTTATAGTAATGTGTCTCTCTGTTTCTGTAGTAATTT
D_* M_C_L_S_V_S_V_S_N_F
GTAATGGTCTTCGTTCGTTTGTTCGACGTTCTTTGTTAATTTGAAAGGGGCTTCTAATTTTCCTTTTTGT
V_M_V_F_V_G_L_F_D_V_L_C_*
AATGGTCTTTTTCTGTATCTGGAGTCGACATTATATTAGTTCTATTCTTATTTATTTTCGTTAATTTTCT
M_V_L_F_C_I_W_S_R_H_Y_I_Q_F_Y_S_Y_L_F_R_*
TTTCTTTATTGTTTTGTGTATCTAAGTTAGGTTGTCCGGCGTAAATGAATATGTGAGCAGGGGTTTGTCC
GGTAGTATCGTGTGGTCTGTGGTTGTAATGTAAGGATTGTTTCTATTTTACTTAATTTGGTCTCG
TTGTCGTCAGAGTTTTTATGATGCGGATTTTTTCGTTTATAGTTTTGTGGAGTCTCTCTATGTCAGCGA
CACCTGTTTTGTAGTATTAAGTTGTAACCTACTTCTTCGGTTTTCGAGCCAACGTTTGAGGGCTAGGCT
GGAGAATGCACCATCTCGATCTGCTTTGAGAAGTTTAGGTTCCCTAATTGATTAATATGCACATTAAG
M_H_I_K
GCATTTTTGCATTCTACCCAGTCTTTTGTCTTAATCTGTTCAAGGGTGGCGAATTTAGAGTAAATATCGA
A_F_L_H_S_T_Q_S_F_V_L_I_C_S_R_V_A_N_L_E_*
TACAGCTCAGATAATGGTTTCTTCGGAAGAATAGATGTCTATTACGAACCTATCAGACAATGTTCCGG
M_V_S_F_G_R_I_D_V_Y_Y_E_L_I_T_T_M_F_G
M_S_I_T_N_L_S_R_Q_C_S_G
TTTAGGGGTAATTTTCATTGGCATTTCGGTGTACGGTGTCTGTTTTGGCTAGGTTGCAGACTTGACAT
F_R_G_N_F_H_W_H_F_R_V_T_V_F_C_F_G_*
L_G_V_I_F_I_G_I_S_V_L_R_C_S_V_L_A_R_L_Q_T_*
TCGTTGATTATATTCTGGATAAGGAGTTGGCTATTGGGGAAGTAATAATTCCTACTAAAGAGTTTTACAG
TTTTTTGAATTCGGGGTGTTCCTTAAACTTGCATAAGTAGGGATGTTTTTAATAAACGAGGCTTCG
M_F_F_N_K_T_R_L_R
GAAAATTTTTGTACATTTCCGATTTCATTGTTGTTGTATATGCATTTTGGATCATTTCGAAGTCTGCATCG
K_I_F_V_H_F_G_F_I_V_V_V_Y_A_F_W_I_I_S_K_S_A_S
M_H_F_G_S_F_R_S_L_H_R
CTTTCGATGTAAAGAGCGCTATTCTTTGAGCAAAAGTGATTGAGTAGATATTGTTTGGCTGCTTCTAGGG
L_S_M_*
F_R_C_K_E_R_Y_S_L_S_K_S_D_*
TCATTTCTTTGTAGAAAATGTGTATGATTTGTTTTTTGAAAATAAGTTTTCGTGTTCTATACTACTATTGGG
ACCTTTTGGAGAAAATAATTTGCTTTTATGGTAATTAATGGCTTTTTTCGGACAACCTATTAAGTCGCTA
M_V_I_N_G_F_F_G_Q_P_Y_*
TTATCTTCTTCTGCGCTATGTTGCGTAGATTCTCCGAAAAATGCTTGTCTATTTTGATTCTAGACAGTG
M_L_R_R_F_S_E_K_C_L_F_Y_F_D_S_R_Q_C
CGTCTGCGACATGGTTTTCTTTCTTTGATATATTTTATATCGAAGTCGTAAGTCTCGGATAATCTAATCTT
V_C_D_M_V_F_L_S_F_D_I_F_Y_I_E_V_V_L_G_*
CCACCTAGTCAGCTTTGAGTTAGGTTCTTTAATTTGTGTAGCCAACACAATGGTTGATGGTCAGAAGCT
M_V_D_G_Q_K_L
M_V_R_S
ATTTGAAAATGTCGTCCTAACAGGTAGTGTCTAAAGGCTTTTGTGGCCCAAACCTATTGCTAGTAGTCTT
F_R_N_V_V_L_T_G_S_V_*
Y_F_E_M_S_S_*
TTTCTATAGCACTGTAATTAACCTCATGTTTCGTTTAGGGTCTGCTAATGTAGCTAATGGGATGTCCATC
TTGAGATAGAACAGCACCTAATGCTACATTGCTTGCCTCAGTTGTTAATACAAATTTCTTTCAAATTA
M_L_H_C_L_R_Q_L_L_I_Q_I_F_F_Q_N_*
GGTATTTTTAATATGGGGTCTTCGAAAATAAGTAATTTTTAATTTTTGAAATGCTTCTATATATTCTTTAT
M_G_S_S_E_I_S_N_F_N_F_* M_L_L_Y_I_L_Y
GTTTTGTATCTATTTTTACTCCCTTTTTGAGACATTTAGTCAAGTGGTTTTCGCTATGTCAGAGAAGTTGGG
V_L_Y_L_F_L_L_P_F_* M_S_E_K_L_G
TATGAATTTCTGTAGTATCCTGTGATTCCAAGGAATGCTTTAATTTCTTTAGGTTTAGATGGAAGTGGG
M_N_F_L_*
TAATCTTGTATAGCTTTGATTTTTTGGGATTTGGTTTAAATTCGGTCTTTAGTGATAACGTGTCCTAGAAA
GGTAGTTTCTTGTCTTAGGAATTCACATTTATCTAGCTGTAGTTTAAAGGTTGGCTTGAGATAATTTCTCG
AAGACTGCTTCAAGGGATTGAAGGTGTTCTCTAGAGAGGTAGAAAAGACAATGATATCGTCTAGATATA

```

                                                                M_I_S_S_R_Y
CTAAACAATTTTTATTTAAAAGTGGACGAAGTAGGTTATTCATGCAGCGTTGGAAGGTAGCTGGGGCGTT
T_K_Q_F_L_F_K_S_G_R_S_R_L_F_M_Q_R_W_K_V_A_G_A_F
TTTAAGCCCAATGGCATTCTAGTATATTCGTAATGTCCGATTTTAGTAGAGAAAGCCGTTTTGGCTATT
_L_S_P_N_G_I_L_V_Y_S_*M_S_V_F_S_R_E_S_R_F_G_Y
GATTCTGAATCCATCTCTATTTGATGGAAGCCTTTTGCCAGGTCTATGGTGGTAAAGTAATTAGACCTCC
*
                                                                M_E_A_F_C_Q_V_Y_G_G_K_V_I_R_P_P
CTAACTTTCTAATATTTTCGTCTATATTTGGCATAGGGTATCGGTCTATTTATTTGTATTTTCATTCACTT
*
CCTGTAGTCTATGACTATTCGAAGTTTCTGCTGTCCGGAAGCATCTAATTTTTTTTGGTACTACCCAGATG
                                                                M_T_I_R_S_F_C_C_P_E_A_S_N_F_F_G_T_T_Q_M
GGGCTGCAGTACGGTGAGTTGCTTTCTCGAATGATGCCTTGTTCTAACATTTCTTGAATTTGTCTTTCTA
_G_L_Q_Y_G_E_L_L_S_R_M_M_P_C_S_N_I_S_*
CTTCTTGTTTCGTATGATTGTTGGAAGCTATACATTTTAGAGTAAACTGGAATATTATGTGTGGTATTAAT
                                                                M_C_G_I_N
TGTATGTTTTTCTTGGTTTTGTGAAGCTAAGTTTGTCCGCTCGTGAAGTGAATGTCTTTGAATCTTGTCT
_C_M_F_F_L_V_C_E_A_K_F_V_A_L_V_K_L_N_V_F_E_S_C_
                                                                M_S_L_N_L_V
AGTAAGTTTGTAAATTTGTATTTTCTTCTGCATTTAAATGGTTTAGTTCGGAACAGATTCTCCTGAAGTG
Q_*
_S_K_F_V_N_L_Y_F_S_S_A_F_K_W_F_S_R_N_R_F_S_*
TGGAGGTTTCTAAAAAGGTAGGGTCTTGTGGGAACGAAGAGTCTACTTGAGCATGATTTTGGTCTGTAAG
                                                                M_I_L_V_C_K
TTTTTCAGTGTCTTTTATTTTATAGACTTTGTTGAAGAGGGTGGCAGTTTTCTTTTGGATAATCTATTATG
_F_F_S_V_F_Y_F_I_D_F_V_E_E_G_G_S_F_L_L_I_I_Y_Y_
                                                                M
GCTTTAGCTTTAGACAACAATTTGCGTCCAATCAGCATATCATAATGATCTGAGAATTTGTGAATTAGAA
G_F_S_F_R_Q_Q_F_A_S_N_Q_H_I_I_M_I_*
_A_L_A_L_D_N_N_L_R_P_I_S_I_S_*
ACTCCTGAGCAATAGGAAAAGTAGTATTGTTGGTGGTATAGTGGTGGTCTCGTTAATAGTTAAACAAGCCATT
ACTTGTATGAATTAGGTGGTGGTTGTTGTATAGGTGTTTTCGAAAATGTTCCAAGTAAGCATGTTTTATG
GTGGATCCAGTGTCAATTAGGCACTTTAAGTCTATGTCATTATGTTTAATAATTACATATTGGTTTGATC
CGGGCAACTAGCCGAAAATTTCTGCTTCTTTGGTATTGTCTGATTCGTCATCGTTTGGTGTCTGTCTGT
TCTTGTGTGTTTTGTCCGAAGTGTGCATGATCTTGTTCGAGAGGGATTTAAGTTGGGCTGGGATAAAC
                                                                M_I_L_F_E_R_D_L_S_W_V_W_D_I_T
CCATTTGATAAGGATATGTTTGTGGTTGGAAGTTATCTGGGTAGTATGGATAGTAATAATGGTAAGGGTA
_H_L_I_R_I_W_L_W_L_E_V_I_W_V_V_W_I_V_I_M_V_R_V_
TTCGGGTTGGTTTGGATAACCATTTCTTGGTATCTTAGTTCTTCTATGCTCATTTTTAGTCTGTCCGCTGTCT
_F_G_L_V_*
                                                                M_L_I_L_V_C_P_L_S
GATGGTTCGTGTCTTTTTACCGGTTGTTGGTTTGTGTGTTCCAGTTGGAGGGAAGTTATTCCGGTGTGT
_D_G_R_C_L_F_T_G_L_L_V_C_V_F_P_V_G_G_K_L_F_G_V_
_M_V_V_V_F_L_P_G_C_W_F_V_C_S_Q_L_E_G_S_Y_S_V_C_
AAGTTTGATTAGGTCCGAAACGATTTTGGTTCGAGATATCTGTCTGATTAAAGACAGGATTCGGTCTGGG
*
_K_F_D_*
TATGTTAGGTTGTGGGTAGGTTGGTCTTGGTTGATATTGGGGAAAAGGGTTTATTGGACCTGGGTGTA
TGGATTGGTTCGCTATGGGTATGGTGGATAAAGGGTTGAAAGACGTTGTTATGGTACGGTTGAAAAGGTG
M_D_W_C_V_W_V_W_W_I_I_G_L_K_D_V_V_M_V_R_L_K_R_C
                                                                M_G_M_V_D_N_R_V_E_R_R_C_Y_G_T_V_E_K_V_
TTATAGGTTTATTTTGGCCATTAAGTGTCTATTTGGTAAAGTTACTTGGCGTCTTATTAGAATTAGGAGC
_Y_R_F_I_L_A_I_N_L_S_I_W_*
_L_*
TGGCTTAAAGTTATCTTTGTTTAAATGTTGAGGTTTCATATAGACCCTTTTCTTGTGAAGATATTAATAAGG
                                                                M_F_E_V_H_I_D_P_F_L_V_R_Y_*
GATCTTAAATCGGGAATTTGGCAGTTAACTATTCTTAGATGAATATGGTTAGGGAGTTTTAGGATCAAGT
TTTCTATGCTATCTTGAATAAGTCTGGTAAATAAGCGGTTTTAGCTGAGTTATTTTCTAGTTCTAACTT
ACTCATTAAACTTTGTCTGCTTTTCTCGGCTTCCCTCAGAAAAGTGCCTGATGTTGCCTTTGTATTGAATG
TTCCGAAAATCCTCCAAAAGTTGATGACTCGGTGCCCTGGGGTTTGTAAATATAGGATCAATCTGGTCTTAA
                                                                M_T_R_C_L_G_F_V_I_*
GTTCAATCCAGCTGGTGAAGTTGTTGGTCCCGAGAGTACGGATGACGTCGCCGCTGATATTGCGTTCAAT
                                                                M_T_S_P_L_I_L_R_S_M
```


GTGTCCATAAAATTATCAACTTCTGCCTTTCGTCAGTAGTTTGGTATAAGGCCAGAATGAAGTCGATGCGA
C_P_* M_K_S_M_R
 GTTATAAAGGTATGTAGAGTATTCGGGTCTCCATTGAAGATTTTCGACAGCATTATCTGTCTCTCGACCT
V_I_K_V_C_R_V_F_G_S_P_L_K_I_S_T_A_F_I_C_L_S_T
 GACTCATATTTGAGTCGGATAGTATTATAGGGCTTGCCTGTTGCCATTGTTGTTGTTGTTGTTATTTATTGTTAT
 TTAATGTTTTAGGTAAGGTTACAGTTCGGGTAAGGCTTGCAGTTTCGGATAGGTTTCGTTTTTCGATAA
 GGTTCAAGTTTTACTGTAGGTTTTGTCTCTGGTCCGAACTTTAGGTTTTATGTTTTGGATAGGTTTTGG
 * M_F_W_I_G_F_W
 TTTTCACTATGGGTTCTTGATTTCCACTATTGGTTTGTATTCCAGGTCTGAGCTTTGTGTTTGTGTTTGT
F_S_L_W_V_L_D_F_H_Y_W_F_V_F_Q_V_*
 TGTTTGTTCACCTATTAATTTAATTCACCTATAATGTTTGTATTTCACCTTAGTTATAGTTCTTTAGCTTA
M_F_V_I_H_L_V_I_V_L_*
 GGTTTTGTTGAATTTTCGCACCTTATTTGTTTCGTAATATTTCTGTTATTCACCTATTTGTAAGTTTTTAAT
 TTAAGTTTGTGTTGTTGCGCTCTTAGCTGTTTCGCTTTGGTTTGGAGTAACCTTTTGTGGATTTACTACTC
 AGCGAATTCAAAGGCGTCTTCTTAAAATTTTTTTGTAATTTTAAAGTTATTTCACTTCGTGGTCACTTTAT
 GGGATTGAGTACCATTGGTTAACACTTTTCCGCGGTGAACAGGTTAACTGACACGATCCGCGTTTGTAT
 AAGGTTAGTTATACCGTTGGTTAGCACTTTCCACTTAGCACGACGCGACTGCGTTAGTTCTTTTGTGGAT
 AATAATTGAAGAATTGAATAGGTTACTATCGCATCAACAATGTATTATCCTACCGACTGCGCCACTTAC
M_Y_Y_P_T_D_C_A_T_Y
 GTATTTTACTTTTCGAGTCGAACAAATGATCTGTCGTTTACTAAGATCAACGCCTTTAAAGAAGTTTCG
V_F_Y_F_S_S_R_T_N_D_L_S_F_D_*
M_I_C_R_L_T_K_I_N_A_F_K_E_V_S
 GATTTGAAAGAATCAACACGAGGTATTTTACTTTAGATCTTTATTTCTTCGTTGTGCGGCTTAAGACTAA
D_L_K_E_S_T_R_G_I_F_T_L_D_L_Y_F_F_V_V_G_L_R_L
 ACGTACAGAGGTTTTGATCGGCTCTTATTGAGTCGCGAGAAACGAAATGTTTTTGTGAGCATTTTTTTA
N_V_Q_R_F_L_I_G_S_Y_* M_F_L_S_A_F_F
 TTTTATTTGTTTTAAGTGCATACAAAAGAATGAAGAAGTTATACAATGATCATGTGTTTTACTTTGTT
I_L_F_V_F_K_C_I_Q_K_E_*
M_K_K_L_Y_N_D_H_V_F_Y_F_V
M_I_M_C_F_T_L_L
 GCAGGTTTCAAGTTTCGTTACGTTATGTTCTGTGCATGCAAGATAGTATACTTAGCGACTAATGCAATGT
A_G_F_K_F_R_Y_V_M_F_C_A_C_K_I_V_Y_L_A_T_N_A_M
Q_V_S_S_F_V_T_L_C_S_V_H_A_R_* M_Q_C
 AAGCATAATATGAGAAGCATATTATGTTATGTAAGTGCAGACTGTATGTTTGGCAATTAGTGTAAAGTAT
 *
K_H_N_M_R_S_I_L_C_Y_V_S_A_R_L_Y_V_W_Q_L_V_*
M_L_C_K_C_E_T_V_C_L_A_I_S_V_S_M
 GTAAACGTAATATGAGAAGCATATTATGTTACTATATATTCTTGATCAGAGAACAGAGAAATATAAGGTA
 * M_R_S_I_L_C_Y_Y_I_F_L_I_R_E_Q_R_N_I_R_*
M_L_L_Y_I_L_D_Q_R_T_E_K_Y_K_V
 ACAAATAAATCCGAGGGCGAACAAACACTTTGTCACACGACTGTGACCAGATCTATAGCCTGACACAAGG
T_N_K_S_E_G_E_Q_T_L_C_H_T_T_V_T_R_S_I_A_*
 AACCGATCGCCCTAGAGAACATGAACGACCAACCGACCACATCCAGATGAAACAGTTAAACTCAGTAGAA
M_N_D_Q_P_T_T_S_R_* M_K_Q_L_N_S_V_E
 AACAAAGAGAGAATGCTATAGTCAAGATCCCCACTATCAGATACCCG
N_K_R_E_C_Y_S_Q_D_P_H_Y_Q_I_P_G_*

Para la región genómica 2R_5, el programa Augustus Gene Prediction predijouna estructura compuesta por 2 exones de 1972 pares de bases el primero, y 1062 pares de bases el segundo, y 1 intrón de 110 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido SLNLVSK con una probabilidad de $\rho=0.002$ y un valor $E=0.000113327640664902$; para el péptido GIFTLDLYFFVVLRLNVQR con una probabilidad de $\rho=0.078$ y un valor $E=3.79387154351475e-05$; para el péptido SITNLSR con una probabilidad de $\rho=0.004$ y un valor $E=0.000894416822897423$; para el péptido MIMCFTLLQVSSFVTLCSVHAR con una probabilidad de $\rho=0.035$ y un valor $E=0.000351924336191395$; para el péptido IMCFTLLQVSSFVTLCSVHAR con una

probabilidad de $\rho=0.029$ y un valor $E=0.000732423818564053$; para el péptido LISPRFCLFVWCNVIALSSLNILLCR con una probabilidad de $\rho=0.033$ y un valor $E=0.000363678919483202$; para el péptido ERYSLSK con una probabilidad de $\rho=0.064$ y un valor $E=0.0001266509436431$; para el péptido VDGQkLFR con una probabilidad de $\rho=0.016$ y un valor $E=0.00016400775031939$; para el péptido MEAFCQVYGGK con una probabilidad de $\rho=0.038$ y un valor $E=0.000943166356804488$.

Evidencia de ESTs:

gi|49216375|gb|CO294829.1|;

```
GAGATAAACCTTACCCCATTTGTACACACCCCTTAAGCCTAATAAAAAATAAAACCAATAGAACATAAGGACA
TTGTACAATTAATATCAAATAATAATATCACAGTTTGTATAATAGTATCAATCATAAGCCTCACATCTTG
TATTGCTTGTATTTACTTAAAACTAAAAACAAAAATAATCATTGAATCACAAGATGCAAATCCAACCTGC
ATCACCAAGATTGCGGGCATCAACACCAATAGAAGGAGTGGAAAGCATAAGCTTCCCTTTTAAAGGGAGGG
AAGTAACATAATATGCTTCTCATATTACGTTTACATACTTACACTAATTGTACATAACAATCTTGCACATG
CATAAACACATGAAACCAGTTTACATTTTTTACTTACACTTAAGCGCATAATTTGTTGTGCATCCATACCG
TTATTTTTCCGTTCTTTTTTGTACACATATACTGATTAGACATTTCCCGTTTCTCGCGACTCACTTCGAGC
CGATCAAAAACCTCTGTACAGTCAGTCTTAAGCCGA
```

gi|49401009|gb|CO340714.1|;

```
TCGGCTTAAGACTGACTGTACAGAGTTTTTGTATCGGCTCGAAGTGAGTCGCGAGAAACGGGAATGTCTAA
TCAGTATATGTGTACAAAAAAGAACGGAAAAATAACGGTATGGATGCACAACAAATTATGCGCTTAAGTG
TAAGTAAAAATGTAACCTGGTTTCATGTGTTTATGCATGTGCAAGATTGTATGTACAATTAGTGTAAGTA
TGTAACGTAATATGAGAAGCATATTATGTTACTTCCCTCCCTTTAAAAGGGAAGCTTATGCTTCCACTC
CTTCTATTGGTGTGATGCCCGCAATCTTGGTGTATGCAGTTGGATTTGCATCTTGTGATTCAATGATTAT
TTTGTTTTTTAGTTTTAAGTAAATACAAGCAATACAAGATGTGAGGCTTATGATTGATACTATTATACAA
ACTGTGATATTATTATTTGATATTAATTGTACAATGTCTTATGTTCTATTGGTTTTATTTTTATTAGGC
T
```

gi|103678882|gb|EC060256.1|;

```
AAGTTTTTAATTTAAGTTTGTGTTGTTTCACTTATTAATTTAATTCACCTATAATGTTTGTATTTCAC
TTAGTTATAGTTCTTTAGCTTAGGTTTTGTTGAATTTTCGCACTTATTTGTTTCGTAATATTTCTGTTATT
CACTTATTTGTAAGTTTTTAATTTAAGTTTGGTGGGGTTGCGCTCTTAGCTGTTTCGCTTTGGTTTTGAGT
AACCTTTTGTGATTTACTACTCAGCGAATTCAAAGGCGTCTTCTTAAAATTTTTGTAATTTTAAGTTAT
TTCACTTCGTGGTCACTTTATGGGATTGAGTACCATTGGTTAACACTTTCCACTT
```

gi|107316563|gb|EC259523.1|;

```
ATCCATACCGTTATTTTTCTGTTCTTTTTTGTACACATATACTGATTAGACATTTCCCGTTTCTCGCGACT
CACTTCGAGCCGATCAAAAACCTCTGTACAGTCAGTCTTAAGCCGACAACGAAGAAATAAAGATCTAAACT
AAAAAATACCTCGTGTGATTCTGAAACTTCTTTAAAGGCGTTGATCTTAGTCAAACGACGGATCATTG
TTCGACTCGAATAGTAAAATACGTAAGTGGCGCAGTCGGTAGGATAATACATTGTTGATGCGATAGTAAC
CCTATTCAATTCTTCAATTATTATCCACAAAAGAACTAACGCAGTCGCGTCGTGCTAAGTGGAAAGTGT
AACCAATGGTACTCAATCCCATAAAGTGACCACGAAGTGAATAACTTAAAATTACAAAAAT
```

gi|49403391|gb|CO343084.1|

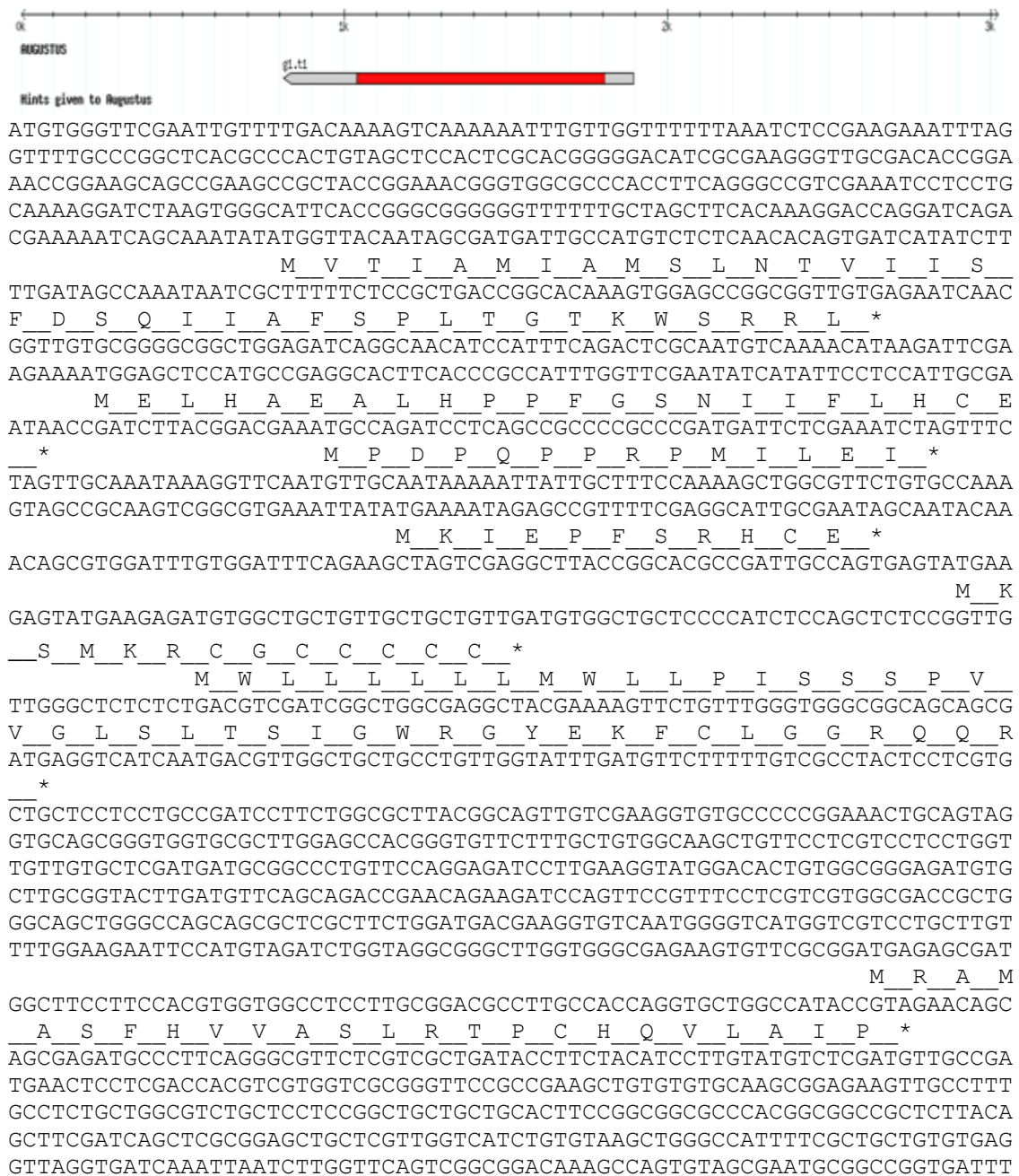
```
GCATCCATACCGTTATTTTTCTGTTCTTTTTTGTACACATATACTGATTAGACATTTCCCGTTTCTCGCGA
CTCACTTCGAGCCGATCAAAAACCTCTGTACAGTCAGTCTTAAGCCGACAACGAAGAAATAAAGATCTAAA
CTAAAAAATACCTCGTGTGATTCTGAAACTTCTTTAAAGGCGTTGATCTTAGTCAAACGACGGATCATT
TGTTTCGACTCGAATAGTAAAATACGTAAGTGGCGCAGTCGGTAGGATAATACATTGTTGATGCGATAGTA
ACCCTATTCAATTCTTCAATTATTATCCACAAAAGAACTAACGCAGTCGCGTCGTGCTAAGTGGAAAGTGT
TTAACCAATGGTACTCAATCCCATAAAGTGACCACGAAGTGAATAACTTAAAATTACAAAAAT
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 63 sORFs. La transcripción se basa en la presencia de 5 ESTs en la base de datos pública y la probable traducción en la presencia

en bases de datos de 9 péptidos definidos por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos 9 de los sORFs codificados por este transcripto son traducidos.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_35



GAACTGAACAGCTCTCCACATTGCAGCCCTTTTATATCCGCCTGTTGTGAAATCGAACTTCGAAAAAGC
TTGAGGCATTGGGGAGAGCCAATAAGAGAGGGAAGGTCTATGTTTTTGGTATTATTGCTCAGAGTGAATC
M F L V L L L R V N
ATCACGAAACGGGTGGCGCCACCTTCAGGGCCGTCGAAATCCTCCTGCAAAAGGATCTAAGTGGGCATT
H H E T G G A H L Q G R R N P P A K G S K W A F
CACCGGGCGGGGGTTTTTGCTAGCTTCACAAAGGACCAGGATCAGACGAAAAATCAGCAAATATATGG
T G R G V F C * M
TTACAATAGCGATGATTGCCATGTCTCTCAACACAGTGATCATATCTTTTGATAGCCAAATAATCGCTTT
V T I A M I A M S L N T V I I S F D S Q I I A F
TTCTCCGCTGACCGCACAAAGTGGAGCCGGCGGTTGTGAGAATCAACGGTTGTGCGGGCGGCTGGAGA
S P L T G T K W S R R L *
TCAGGCAACATCCATTTTCAGACTCGCAATGTCAAAACATAAGATTTCGAAGAAAATGGAGCTCCATGCCGA
M E L H A E
GGCACTTCACCCGCCATTTGGTTCGAATATCATATTCCTCCATTGCGAATAACCGATCTTACGGACGAAA
A L H P P F G S N I I F L H C E *
TGCCAGATCCTCAGCCGCCCGCCGATGATTCTCGAAATCTAGTTTCTAGTTGCAAATAAAGGTTCAAT
M P D P Q P P R P M I L E I *
GTTGCAATAAAAAATTATTGCTTTCCAAAAGCTGGCGTTCTGTGCCAAAGTAGCCGCAAGTCGGCGTGAAA
TTATATGAAAATAGAGCCGTTTTTCGAGGCATTGCGAATAGCAATACAAACAGCGTGGATTTGTGGATTC
M K I E P F S R H C E *
AGAAGCTAGTCGAGGCTTACCGGCACGCCGATTGCCAGTGAGTATGAAGAGTATGAAGAGATGTGGCTGC
M K S M K R C G C
M W L
TGTTGCTGCTGTTGATGTGGCTGCTCCCCATCTCCAGCTCTCCGGTTGTTGGGCTCTCTCTGACGTCGAT
C C C C *
L L L L L M W L L P I S S S P V V G L S L T S I
CGGCTGGCGAGGCTACGAAAAGTTCTGTTTGGGTGGGCGGCAGCAGCGATGAGGTCATCAATGACGTTGG
G W R G Y E K F C L G G R Q Q R *
CTGCTGCTGTTGGTATTTGATGTTCTTTTTGTCGCCCTACTCCTCGTGTGCTCCTCCTGCCGATCCTTC
TGGCGCTTACGGCAGTTGTGCAAGGTGTGCCCCGAAACTGCAGTAGGTGCAGCGGGTGGTGGCCTGG
AGCCACGGGTGT

Para la región genómica 2R_35, el programa Augustus Gene Prediction predijo una estructura compuesta por 1 exón de 1082 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido VNHHETGGAHLQRR con una probabilidad de $\rho=0.097$ y un valor $E=0.000424343980856626$.

Evidencia de ESTs:

gi|13696898|gb|BF502961.2|;

CTCCGCTTGCACACACAGCTTCGGCGGAACCCGGGACCACGACGTGGTCGAGGAGTTCATCGGCAACATC
GAGACATACAAGGATGTAGAGGGTATCAGCGACGAGAACGCCCTGAAGGGCATCTCGCTGCTGTTCTACG
GTATGGCCAGCACCTGGTGGCAAGGCGTTTCGCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCAT
CCGCGAACACTTCTCGCCACCAAGCTCGCCTACCAGATCTACATGGAATTTCTCCAAAACAAGCAGGAC
GACCATGACCCCATTGACACCTTCGTCATCCAGAAGCGAGCGCTGCTGGCCAGCTGCCAGCGGTTCGCC
ACGACGAGGAAACGGAACCTGGATCTTCTGTTTCGGTCTGCTGAACATCAAGTACCGCAAGCACATCTCCCG
CCACAGTGTCCATACCTTCAAGGATCTCCTGGAACAGGCGCCATCATCGAGCACAAACAGGAGGAC
GAGGAACAGCTTGCCACAGCAAAGAACACCCGTTGGCTCCAAGCGCACCACCCGCTGCACCTACTGCAGTT
TCCGGNGGCACACTTCGACAACCTGCGGTAAGCGCCAGAAGGATCGGCAGGAGGAGCAGCACGAGGAGTA
GGCGACAAAAAGAACATCAAATAACCAACAGGCAGCAGCCAACGTCATTGATGACCTCATCGCTGCTGCCG
NCCACCCAAAACAGAACCTTTTCGGAGCCTCGCCAGCCGATCGACGTCAGAGAGAGCCCAACACCCGGAGAG
CTGA

gi|13756611|gb|BF490911.2|;

CTGAACCAAGATTAATTTGATCACCTAACCTCACACAGCAGCGAAAATGGCCAGCTTACACAGATGACC
AACGAGCAGCTCCGCGAGCTGATCGAAGCTGTAAGAGCGCCGCCGTTGGGCGCCGCCGGAAGTGCAGCAG
CAGCCGGAGGAGCAGACGCCAGCAGAGGCAAAGGCAACTTCTCCGCTTGACACACAGCTTCGGCGGAAC

CCGGGACCACGACGTGGTTCGAGGAGTTCATCGGCAACATCGAGACATAACAAGGATGTAGAGGGTATCAGC
GACGAGAACGCCCTGAAGGGCATCTCGTGTCTTCTACGGTATGGCCAGCACCTGGTGGCAAGGCGTTC
GCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCATCCGCGAACACTTCTCGCCCACCAAGCTCGC
CTACCAGATCTACATGGAATTTCTTCCAAAACAAGCAGGACGACCATGACCCCATTTGACACCTTCGTTCATC
CAGAAGCGAGCGCTGCTGGCCCAGCTGCCAGCGGTTCGCCACGACGAGGAAACGGAACCTGGATCTTCTGT
TCGGTCTGCTGAACATCAAGTACCGCAAGCACATCTCCCGCCACAGTGTCCATACCTTCAAGGATCTCCT
GGAACAGGGCCGCATCATCGAGCACATCAACCAGGAGGACGAGGAACAGCTTGCCACAGCAAAGAACC
CGTGGCTCCAAGCGCACACCACCCGCTGCACCTACT

gi|26255784|gb|CA806835.1|;

GCGTGGGCGCCGCCGGAAGTGCAGCAGCAGCCGCAGGAGCAGACGCCAGCAGAGGCAAAGGCAACTTCTC
CGCTTGACACACAGCTTCGGCGGAACCCGGGACCACGACGTGGTTCGAGGAGTTCATCGGCAACATCGAG
ACATACATGATGTAGAGGGTATCAGCGACGAGAACGCCCTGAAGGGCATCTCGTGTCTTCTACGGTAT
GGCCAGCACCTGGTGGCAAGGCGTTCGCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCATCCGC
GAACACTTCTCGCCCACCAAGCCCGCTACCAGATCTACATGGAATTTCTTCCAAAACAAGCAGGACGACC
ATGACCCCATTTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGGCCCAGCTGCCAGCGGTTCGCCAC
CGAGGAAACGGAACCTGGATCTTCTGTTCGGTCTGTTGAACATCAAGTACCGCAAGCACATCTCCCGCCAC
AGTGTCCATACCTTCAAGGATCTCCTGGAACAGGGCCGCATCATCGAGCACAACAACCAGGAGGACGAGG
AACAGCTTGCCACAGCAAAGAACCACCCGTGGCTCCAAGCGCACACCACCCGCTGCACCTACTGTAGTTTCCG
GGGGCACACCTTCGACAACCTGCCGTAAGCGCCAGAAGGATCGGCAGGAGGAGCAGCACGAGGAGTAGGCG
ACAAAAAGAACATCAAATACCAACAGGCAG

gi|13696731|gb|BF502793.2|;

CTGGCTTTGTCCCGGACTGAACCAAGATTAATTTGATCACCTAACCTCACACAGCAGCGAAAATGGCCC
AGCTTACACAGATGACCAACGAGCAGCTCCGCGAGCTGATCGAAGCTGTAAGAGCGGCCCGCCGTGGGCGC
CGCCGGAAGTGCAGCAGCAGCCGCAGGAGCAGACGCCAGCAGAGGCAAAGGCAACTTCTCCGCTTGACACA
CACAGCTTCGGCGGAACCCGGGACCACGACGTGGTTCGAGGAGTTCATCGGCAACATCGAGACATAACAAGG
ATGTAGAGGGTATCAGCGACGAGAACGCCCTGAAGGGCATCTCGTGTCTTCTACGGTATGGCCAGCAC
CTGGTGGCAAGGCGTTCGCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCATCCGCGAACACTTC
TCGCCACCAAGCCCGCTACCAGATCTACATGGAATTTCTTCCAAAACAAGCAGGACGACCATGACCCCA
TTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGGCCCAGCTGCCAGCGGTTCGCCACGACGAGGAAAC
GGAACCTGGATCTTCTGTTCGGTCTGTTGAACATCAAGTACCGCAAGCACATCTNCCGCCACAGTGTCCAT
ACCTTCAAGGATCTCCTGGAACAGGGCCGCATCATCGAGCACAACAACCAGGAGGACGAGGAACAGCTTG
CCACAGCAAAGAACCACCCGTG

gi|26255761|gb|CA806812.1|;

GCGTGGGCGCCGCCGGAAGTGCAGCAGCAGCCGCAGGAGCAGACGCCAGCAGAGGCAAAGGCAACTTCTCC
GCTTGACACACAGCTTCGGCGGAACCCGGGACCACGACGTGGTTCGAGGAGTTCATCGGCAACATCGAGA
CATAAAGGATGTAGAGGGTATCAGCGACGAGAACGCCCTGAAGGGCATCTCGTGTCTTCTACGGTAT
GGCCAGCACCTGGTGGCAAGGCGTTCGCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCATCCGC
GAACACTTCTCGCCCACCAAGCCCGCTACCAGATCTACATGGAATTTCTTCCAAAACAAGCAGGACGACC
ATGACCCCATTTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGGCCCAGCTGCCAGCGGTTCGCCACGA
CGAGGAAACGGAACCTGGATCTTCTGTTCGGTCTGTTGAACATCAAGTACCGCAAGCACATCTCCCGCCAC
AGTGTCCATACCTTCAAGGATCTCCTGGAACAGGGCCGCATCATCGAGCACAACAACCAGGAGGACGAGG
AACAGCTTGCCACAGCAAAGAACCACCCGTGGCTCCAAGCGCACACCACCCGCTGCACCTACTGTAGTTTCCG
GGGGCACACCTTCGACAACCTGCCGTAAGCGCCAGAAGGATCGGCAGGAGGAGCAGCACGAGGAGTAGGCG
ACAAAAAGAACAT

gi|13687890|gb|BF505287.2|;

CTCCGCTTGACACACAGCTTCGGCGGAACCCGGGACCACGACGTGGTTCGAGGAGTTCATCGGCAACATC
GAGACATAAAGGATGTAGAGGGTATCAGCGACGAGAACGCCCTGAAGGGCATCTCGTGTCTTCTACG
GTATGGCCAGCACCTGGTGGCAAGGCGTTCGCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCAT
CCGCGAACACTTCTCGCCCACCAAGCTCGCCTACCAGATCTACATGGAATTTCTTCCAAAACAAGCAGGAC
GACCATGACCCCATTTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGGCCCAGCTGCCAGCGGTTCGCC
ACGACGAGGAAACGGAACCTGGATCTTCTGTTCGGTCTGTTGAACATCAAGTACCGCAAGCACATCTCCCG
CCACAGTGTCCATACCTTCAAGGATCTCCTGGAACAGGGCCGCATCATCGAGCACAACAACCAGGAGGAC
GAGGAACAGCTTGCCACAGCAAAGAACCACCCGTGGCTCCAAGCGCACACCACCCGCTGCACCTACTGCAGTT
TCCGGGGGGCACACCTTCGACAACCTGCCGTAAGCGCCAGAAGGATCGGCAGGAGGAGCAGCACGAGGAGTA
GGCGACAAAAAGAACATCAAATACCAACAGGCAGCAGCCAACGTTCATTTGATGACCTCATCGTGTCTG

gi|15454856|gb|BI565664.1|;

```
GTCAGTTCAAATCACCCGGCCGCATTCGCTACACTGGCTTTGTCCGCCGACTGAACCAAGATTAATTTGAT
CACCTAACCTCACACAGCAGCGAAAAATGGCCCAGCTTACACAGATGACCAACGAGCAGCTCCGCGAGCTG
ATCGAAGCTGTAAGAGCGGCCGCCGTGGGCGCCGCCGGAAGTGCAGCAGCAGCCGGAGGAGCAGACGCCA
GCAGAGGCAAAGGCAACTTCTCCGCTTGACACACAGCTTCGGCGGAACCCGCGACCACGACGTGGTTCGA
GGAGTTCATCGGCAACATCGAGACATAACAAGGATGTAGAAGGTATCAGCGACGAGAACGCCCTGAAGGGC
ATCTCGCTGCTGTTCTACGGTATGGCCAGCACCTGGTGGCAAGGCGTCCGCAAGGAGGCCACCACGTGGA
AGGAAGCCATCGCTCTCATCCGCGAACACTTCTCGCCACCAAGCCCGCCTACCAGATCTACATGGAATT
CTTCCAAAACAAGCAGGACGACCATGACCCCATTTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGGCC
CAGCTGCCCAGCGGTGCGCCACGACGAGGAAACGGAAGTGGATCTTCTGTTTCGGTCTGCTGAACATCAAGT
ACCGCAAGCACATCTCCCGCCACAGTGTCCATACCTTCAGGATCTC
```

gi|3946290|gb|AI296873.1|;

```
CTACACTGGCTTTGTCCGCCGACTGAACCAAGATTAATTTGATCACCTAACCTCACACAGCAGCGAAAAAT
GGCCCAGCTTACACAGATGACCAACGAGCAGCTCCGCGAGCTGATCGAAGCTGTAAGAGCGGCCGCCGTG
GGCGCCGCCGGAAGTGCAGCAGCAGCCGGAGGAGCAGACGCCAGCAGAGGCAAAGGCAACTTCTCCGCTT
GCACACACAGCTTCGGCGGAACCCGCGACCACGACGTGGTTCGAGGAGTTCATCGGCAACATCGAGACATA
CAAGGATGTAGAAGGTATCAGCGACGAGAACGCCCTGAAGGGCATCTCGCTGCTGTTCTACGGTATGGCC
AGCACCTGGTGGCAAGGCGTCCGCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCATCCGCGAAC
ACTTCTCGCCACCAAGCCCGCCTACCAGATCTACATGGAATTCTTCCAAAACAAGCAGGACGACCATGA
CCCCATTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGGCCAGCTGCCCAGCGGTGCGCCACGACGAG
GAAACGGAAGTGGATCTTCTGTTTCGGTCTGCTGAACATCAAGTACCGCAAGCACATCTCCCGCCACAGTG
TCCATACCTTCAAGGATCTTCTGGAAC
```

gi|15509291|gb|BI613766.1|;

```
ACCATGGTTCAAATCACCCGGCCGCATTCGCTACACTGGCTTTGTCCGCCGACTGAACCAAGATTAATTTG
ATCACCTAACCTCACACAGCAGCGAAAAATGGCCCAGCTTACACAGATGACCAACGAGCAGCTCCGCGAGC
TGATCGAAGCTGTAAGAGCGGCCGCCGTGGGCGCCGCCGGAAGTGCAGCAGCAGCCGGAGGAGCAGACGC
CAGCAGAGGCAAAGGCAACTTCTCCGCTTGACACACAGCTTCGGCGGAACCCGCGACCACGACGTGGTTC
GAGGAGTTCATCGGCAACATCGAGACATAACAAGGATGTAGAAGGTATCAGCGACGAGAACGCCCTGAAGG
GCATCTCGCTGCTGTTCTACGGTATGGCCAGCACCTGGTGGCAAGGCGTCCGCAAGGAGGCCACCACGTG
GAAGGAAGCCATCGCTCTCATCCGCGAACACTTCTCGCCACCAAGCCCGCCTACCAGATCTACATGGAA
TTCTTCCAAAACAAGCAGGACGACCATGACCCCATTTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGG
CCCAGCTGCCCAGCGGTGCGCCACGACGAGGAAACGGAAGTGGATCTTCTGTTTCGGTCTGCTGAACATCAA
GTACCGCAAGCACACA
```

gi|4419244|gb|AI516144.1|

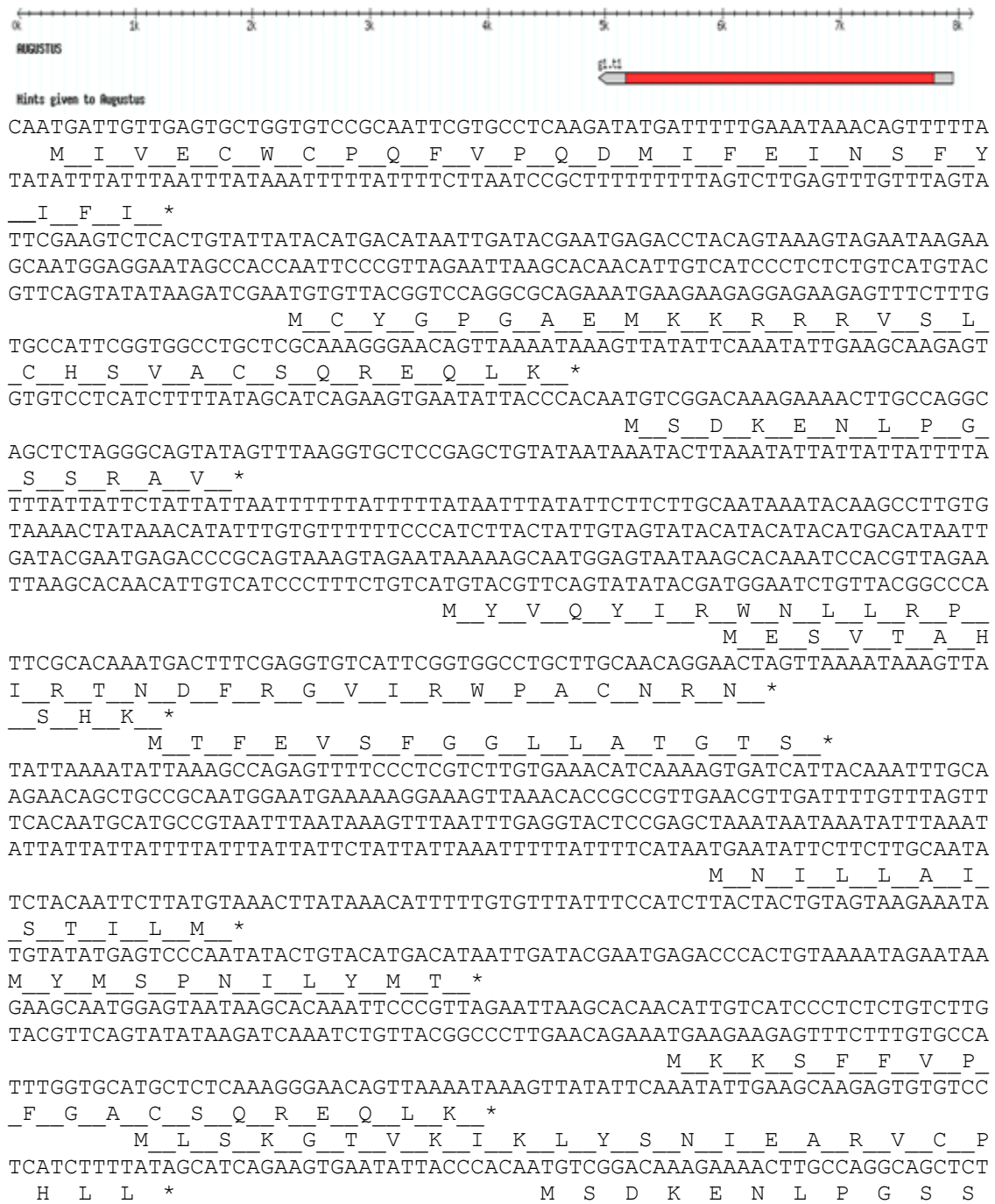
```
CTACACTGGCTTTGTCCGCCGACTGAACCAAGATTAATTTGATCACCTAACCTCACACAGCAGCGAAAAAT
GGCCCAGCTTACACAGATGACCAACGAGCAGCTCCGCGAGCTGATCGAAGCTGTAAGAGCGGCCGCCGTG
GGCGCCGCCGGAAGTGCAGCAGCAGCCGGAGGAGCAGACGCCAGCAGAGGCAAAGGCAACTTCTCCGCTT
GCACACACAGCTTCGGCGGAACCCGCGACCACGACGTGGTTCGAGGAGTTCATCGGCAACATCGAGACATA
CAAGGATGTAGAAGGTATCAGCGACGAGAACGCCCTGAAGGGCATCTCGCTGCTGTTCTACGGTATGGCC
AGCACCTGGTGGCAAGGCGTCCGCAAGGAGGCCACCACGTGGAAGGAAGCCATCGCTCTCATCCGCGAAC
ACTTCTCGCCACCAAGCCCGCCTACCAGATCTACATGGAATTCTTCCAAAACAAGCAGGACGACCATGA
CCCCATTGACACCTTCGTTCATCCAGAAGCGAGCGCTGCTGGCCAGCTGCCCAGCGGTGCGCCACGACGAG
GAAACGGAAGTGGATCTTCTGTTTCGGTCTGCTGAACATTAAGTACCGGAAGCACATTTCCCGCCACAGTGT
CCATACCTTCAAAGATCTCCT
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 14 sORFs. Se pueden observar varias repeticiones en tándem y varias desplazadas. Los sORFs noveno, décimo, decimoprimer, decimosegundo, decimotercero y decimocuarto son duplicaciones desplazadas de los sORFs primero, segundo, tercero, cuarto, quinto y sexto, respectivamente. La transcripción se basa en la presencia de 10 ESTs en la base de datos pública y la probable traducción en la presencia en bases de datos de un péptido

definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos uno de los sORF codificado por este transcripto es traducido.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_36



CGGACAGTATAGTTTAAGGTACTCCGAGCTGTATAATAAATACTTAAATATTATTATTATTTTATTTATT
R T V *
 ATTCTATTATTAATTTTTTATTTTTATAATTTATATTCTTCTTGAATAAATACACGCCTTGAGTAAAC
 TATAAACATATTTGCGTTTTTTTCCCATTTTTACTATTGTAGTATACATACATACATGACATAATTGATACG
 AATGAGACCCGCAGTAAAGTAGAATAAAAAAGCAATGGAGTAATAAGCAAAAATCCACGTTAGAATTCAGC
 ACAACATTGTCATACCTTTCTGTCATGTACGCTCAGTATATACGATGGAATCTGTTACGGCCCATTCGCA
M Y A Q Y I R W N L L R P I R
M E S V T A H S H
 CAAATGACTTTTCGAGGTGTCATTCGGTGGCCTGCTTGCAAAGGGAACAGTTAAAATAAAGTTATATTA
T N D F R G V I R W P A C K G N Q L K *
K *
M T F E V S F G G L L A K G T S *
 AATATTAAGCCAGAGTTTCTCCTCGTCTTGTAACATCAGAAGTGATCATTACAAATTTGCAAGAACA
 GCTGCCGCAATGGAATGAAAAAGGAAATTTAAACACCGCCGTTGAACGTTGGTTTTTTTAGTTTCACAAT
 GCATGCCGTAATTTAATAAAGTTTAATTTGAGGTACTCCGAGCTAAATAATAAATATTTAAATATTATTA
 TTATTTTATTATTATTCTATTATTAAATTTTTATTTTCATAATGAATATCTTCTTGAATATCTACAA
M N I L L A I S T
 TTCTTATGTAACCTTATAACATTTTTGTGTTTATTTCCATCTTACTGTTGTAGTAAGAAATATGTATAT
I L M * M Y M
 GAGTCCACAATACTGTACATGACATAATTGATACGAATGAGACCTACAGTAAAGTAGAATAAAAAGCAA
S P T I L Y M T *
 TGGAAAGATAGCCACAAAATTTCCGTTAGAATTAAGCACAACATTGTCATTCCTCTCTGTCATGTACGTT
M Y V
 AGTATATAAGATCGAATGTGTTACGGTCCAGGCCAAGAAATAAAGAAGAGGAGAAGAGTTTCTTTGTGCC
Q Y I R S N V L R S R R R N K E E E K S F F V P
M C Y G P G E E I K K R R R V S L C
 ATTCGGTGGCCTGCTCGCAAAGGGAACAGTTAAAATAAAGTTATATTAATAAATAAAGCAGGAGTGTGT
F G G L L A K G T G *
H S V A C S Q R E Q V K *
 CCTCGTCTTGTGCAACATCAGAAGTGATCATTACAAATTTGCAAGAACAGCTGCAGCACTGGAATGAAAA
 AGGAAAGTTAAACACCGCCGTTGAACGTTGATTTTTTTTAGTTTCACAATGCATGCCTTATTTAATAAA
 GTTTAATTTGAGGTACTCCGAGCTAAATAATAAATATTTAAATATTATTATTATTTTATTATTATTCTA
 TTATTAATTTTTGCAATCTCTACAATCTTATGTAACTTATAAACATTTTTGTGTTTATTTCCATCTT
 ACTATTGTAGTAAGAAATATGTATATGAGTCCCAATATACTGTACATGACATAATTGATACGAATGAGAC
M Y M S P N I L Y M T *
 CCACTGTAAAATAGAATAAGAAGCAATGGAGTAATAAGCACAAATTTCCCGTTAGAATTTTTAAGGTACTC
 CGAGCTGTATAATAAATACTTAAATATTATTATTTTATTTATTCTATTATTAATTTTTTATTTT
 TATAATTTATATTCTTCTTGAATAAATACACGCCTTGAGTAAACTATAAACATATTTGCGTTTTTTTCC
 CATTTTACTATTGTAGTATACATACATACATGACATAATTGATACGAATAAGACCCGCAGTAAAGTAGAA
 TAAAAAGCAATGGAGTAATAAGCAAAAATCCACGTTAGAATTCAGCACAACATTGTCATACCTTTCTGTC
 ATGTACGCTCAGTATATACGATGGAATCTGTTACGGCCCATTCGCACAAATGACTTTCGAGGTGCCATTC
M Y A Q Y I R W N L L R P I R T N D F R G A I
M E S V T A H S H K *
M T F E V P F
 GGTGGCCTGCTTGC AAAGGGAACAGTTAAAATAAAGTTATATTAATAAATAAAGCCAGAGTTTCTCCT
R W P A C K G N Q L K *
G G L L A K G T S *
 CGTCTTGTGCAACATCAGAAGTGATCATTACAAATTTGCAAGAACAGCTGCCGCAATGGAATGATAAAGG
 AAAGTTAAACACCGCCGTTGAACGTTGATTTTTTTTAGTTTCACAATGCATGCCGTAATTTAATAAAGTT
 TAATTTGAGGTACTCCGAGCTAAATAATAAATATTTAAATATTATTATTATTTTATTGTTATTCTATTA
 TTAATTTTTTATTTTCATAATGAAATTTCTTCTTGAATATCTACAATCTTATGTAACTTATAAACAT
M K F L L A I S T I L M *
 TTTTGTGTTTATTTCCATCTTACTATTGTAGTAAGAAATATGTATATGAGTCCCAATATACTGTACATGA
M Y M S P N I L Y M
 CATAATTGATACGAATGAGACCTACAGTAAAGTAGAATAAGAAGCAATGGAGGAATAGCCACCAATCCC
T *
 GTTAGAATTAAGCACAACATTGTCATCCCTCTCTGTCATGTACGTTTCAGTATATAAGATCGAATGTGTTA
M C Y
 CGGTCCAGGCGCAGAAATGAAGAAGAGGAGAAGAGTTTCTTTGTGCCATTCGGTGGCCTGCTCGCAAAGG
G P G A E M K K R R R V S L C H S V A C S Q R
 GAACAGTTAAAATAAAGTTATATTCAAATATTGAAGCAAGAGTGTGTCCTCATCTTTTATAGCATCAGAA

E_Q_L_K_*
 GTGAATATTACCCACAATGTCCGACAAAGAAAACCTGCCAGGCAGCTCTAGGGCAGTATAGTTTAAGGTG
M_S_D_K_E_N_L_P_G_S_S_R_A_V_*
 CTCCGAGCTGTATAATAAACTTAAATATTATTATTATTTATTATTATTCTATTATTAATTTTTTAT
 TTTTATAATTTTATATTCTTCTTGTCAATAAAATACAAGCCTTGTGTAAAACATAAAACATATTTGTGTTTT
 TCCCATCTTACTATTGTAGTATACATACATACATGACATAATTGATACGAATGAGACCCGCAGTAAAGTA
 GAATAAAAAGCAATGGAGTAATAAGCACAAATCCACGTTAGAATTAAGCACAAACATTTGTCATCCCTTTCT
 GTCATGTACGTTTCAGTATATACGATGGAATCTGTTACGGCCCATTCGCACAAATGACTTTTCGAGGTGTCA
M_Y_V_Q_Y_I_R_W_N_L_L_R_P_I_R_T_N_D_F_R_G_V_*
M_E_S_V_T_A_H_S_H_K_*
M_T_F_E_V_S_*
 TTCGGTGGCCTGCTTGCAACAGGAAGTAGTTAAAATAAAGTTATATTAATAATTAAGCCAGAGTTTTTC
I_R_W_P_A_C_N_R_N_*
F_G_G_L_L_A_T_G_T_S_*
 CCTCGTCTTGTGAAACATCAAAAGTGATCATTACAAATTTGCAAGAACAGCTGCCGCAATGGAATGAAAA
 AGGAAAGTTAAACACCCGCGTTGAACGTTGATTTTGTGTTAGTTTCACAATGCATGCCGTAATTTAATAAA
 GTTTAATTTGAGGTACTCCGAGCTAAATAAATAAATTTAAATTTGTTGTTTGTGAAAAGGTGCTCCAAG
 TGTGGAACACTGACAACCATATTGGGAAGATCCTAATCTTCACCCAAACTCAGCGCATTGAGAGCAAAGG
 CCGAATGAGTAAACCGGAAATAAAAAGTCTTGAGCCGAGCGTCCGAACACGAAGAAGTTTTATTTGAAAC
M_S_K_R_E_I_K_S_L_E_P_S_V_R_T_R_R_S_F_I_*
 TGAAACTGATACACATAAAAAATGCCAGCTTCATGCGGCTGCTTATTTACTAATATATATACATACATA
 GTTCATGTGTATGTATGTATGTTCTCATAACAGAAATACATATGCACATGTGCAATGGCCTATATTCAAC
M_Y_V_L_I_Q_N_Y_I_C_T_C_A_N_G_L_Y_S_T_*
M_F_S_Y_R_I_T_Y_A_H_V_Q_M_A_Y_I_Q_*
 ACACATGTGGGCATTGCCTATATTCAACAGCCTTCTCAATGCCAACATGTTTACAAAAATTAATAAAA
H_M_W_A_L_P_I_F_N_S_L_P_Q_C_Q_H_V_Y_K_N_*
 H_T_C_G_H_C_L_Y_S_T_A_F_L_N_A_N_M_F_T_K_I_K_T_N
 TTTAACATTTTCATAAACCCATTATGCTTTCCCTTTGAAAGATTCTTTGTCATAATGTCTGCTATCATCT
L_T_F_S_* M_S_A_I_I_*
 CATTCTGTAAGTGTACTCTAACACAACGTGACCCTCTTTCATGACTTCTCTAATGAAATGATATCGAAT
S_F_V_E_V_Y_S_N_T_T_* M_I_S_N_*
 GTCGATGTGCTTCGTCTTAGAGTGATGAACCGGGTCTTCGCTAAGTGTGCGCACTCAGGTTGTCGCCA
V_D_V_L_R_P_R_V_M_N_R_V_L_R_*
 TGCATAACCGTCCGGGTCTTCAGATCACCGCATCCGATCTCCACTATTAGCCTTCGTAAAGCTATAGCTT
M_H_N_R_R_G_L_Q_I_T_A_S_D_L_H_Y_*
 CCTTGCAAGCCGTGCTCAGAGCCATATACTCGGCTTCAGTACTGCTCAACGCCAGCTCTGCTGCTTCTC
 GGACCTCCATGATACTGGTCCGCCAGACAGGAAAAACATACCCTGTATATGACTTTCCGGTCCAAACGG
M_I_L_V_R_Q_T_G_K_T_H_T_L_Y_M_T_F_G_P_N_G_*
 TCGCCTCCCCAATCTGCATCCACAAAGCCGGTAAATGCCTGACCGCACTTTTGATAATGCAGCTTGACGT
R_L_P_N_L_H_P_Q_S_R_* M_P_D_R_T_F_D_N_A_A_*
 CCACAGTTGACGCCAAGTACCGGAGGATGTGCTTCACAGCCACCATGTGCTCAGAATGCGGGTCTTATT
M_C_F_T_A_T_M_C_S_E_C_G_S_L_F_* M_R_V_L_I_*
 CCTCTGAGCCAACTTCGCCACCGAATGTAGTATGTCTGGTCTGGTAGTAAGCCCAAGCCACATTAGCTCA
L_*
P_L_S_Q_L_R_H_R_M_* M_S_G_L_V_V_S_P_S_H_I_S_S_*
 CCAATTGTAGACTGATACTGCCCTGCGTGCACCTTCTGGCACTGCTCACCCGCGCACAAAACCTTGATGCC
P_I_V_D_* M_P_*
 CTGCATCCAAAGGTGTGCTGCTGGTCTACAGTTCTCGCTGCCATATGCCCGCAATAGTTTCTTGATATA
C_I_Q_R_C_R_R_W_S_T_V_L_A_A_I_C_P_Q_*
 TTGCGAATGGCCAAAGTGATTTCTCAAGGTGCCATCTCGTTGCACCTCCATGCCTAAGAACAAATGC
M_A_Q_S_D_F_S_K_V_A_I_S_L_H_L_H_A_* M_P_K_N_K_C_*
M_*
 AGTGGACCCTTGTCCGTGCACTCGAAAGACTCTGAAATCTTGGCTTTTCAGATCCTCCATATCTTCTCTTG
S_G_P_L_S_V_H_S_K_D_S_E_I_L_A_F_R_S_S_I_S_S_L_*
Q_W_T_L_V_R_A_L_E_R_L_*
 ACTGGCAGCTAGAATTAATCATCAACATATACTAAGATGAGCATCAGATTACCTTGACCACTTTGCTG
D_W_H_A_R_I_K_S_S_T_Y_T_K_M_S_I_R_L_P_*
 ATAAAGACATGGTTCATGATTACAGGCCTTAAATCCCAAGTCTTTTGAACACCCGTCGAGCTTGGAGTTC
M_I_T_G_L_K_S_Q_V_F_*

CACTCTCTGCCTGACTGCTTCAAGCCGTATATTGCCTTCCTCAGCAATAACACCTGGTTCGGGATTAGCAG
CATCTGTGAACCCCTGGGGCTGCTTCATGTACACAGTATCCTTTAGCTCGCTATTTAAGTACGCCGTGCA
M_Y_T_V_S_F_S_S_L_F_K_Y_A_V_H
TACGTCCATGTGATGCAAGTACAATTGCATCTCTGCTGCCAATGCCAAAATGAGCCTCACACTCTCGAGC
T_S_M_*_M_Q_V_Q_L_H_L_C_C_Q_C_Q_N_E_P_H_T_L_E_
M_S_L_T_L_S_S
CTGCACACGGGTGAAAAAGTCTCGAAGTAGTCCACTCCGAACCTTCTGCGAACACCCCTTTGCTACTAGTC
P_A_H_G_*
L_H_T_G_E_K_V_S_K_*
GTGCTTTGAAGCGCTCAATTCTACCAGAGACGTCTCGTTTTCAGGGAATACACCCACTTGCAAGCCACACA
CCGGCGATTTCTTGTAAGTCAGCCAGCTTCCATGTCTCATTTCGCAAGTAGCGCCTTGTACTCCAGGCC
M_S_H_L_Q_V_A_P_C_T_P_G_P
ATTGCCCTTTCCCACTTTGCAGAATACTGCCGAATTGATGGCCTCCTCATAGGACTTGGGAATTTTCGACGT
L_P_L_P_T_L_Q_N_T_A_N_*
CGCTAGCCATCAACACGCCAAGTACATTGTATTGCTTCTTCGGGCGCCCTGGTCTGCCCGTCCGAACAAT
CTTTGGCCGTCCAGGCCCCACATGTACTTCAGCTTCTTCTGTCTCAGCTGCACTCTCATAGTCGTTCGCTG
CTGCCACGGGTCTGTATCGTCACCTTCCCCTGCAATCGAATCACTCTGCGGCTCCGGGTTCATCAGTTG
CGGGAACTGGAATCAATGGTATTTCCATGATTTACCAAATACCAGACTCATCAAAAAGGACATCTCG
CTTCTCGATCACACACCGCTTCTCTTTGTCAAACAGACGGTACCCCTTAGCCGCTATTGAATATCCAATC
ATACGATATTCCTTTTCTTTGGATTTCGAATTTGCCTTTATGGACTCCTTTGTCCAATGCCACTGCTATGG
M_D_S_F_V_Q_C_H_C_Y_G
M_P_L_L_W
CACCAAAAACCCCTCAAGTGGCTCACTGCTGGTATTTTTCCGGTCCACTCTTCCATAGGGGTTTTGCTTTG
T_K_N_P_Q_V_A_H_C_W_Y_F_S_G_P_L_F_H_R_G_F_A_L_
H_Q_K_P_S_S_G_S_L_L_V_F_F_R_S_T_L_P_*
TAATGCTCTGCTCGTTGATCGGTTCCCTCAGATACACCCGAGTGTATTCGCTCAGCCCATAGAGCCTCA
*M_L_C_S_L_I_G_S_S_D_T_P_Q_C_L_S_P_Q_P_I_E_P_H
CCCAACTCCGATTGCAGCAACATGCACCTAGCCATTTCTACTAGCGTGCGGTTGGCTCGTTCTGCAACTC
P_T_P_I_A_A_T_C_T_*
CATTTTGTGGGGAGTGTGTGGAATAGTACAGCTGTCTAGCGATCCCATGTGCCTTTAAATAGTCATCAAA
AACATTATTGACGAACTCACCACCATTATCGCTCCGGATACATTTTATCTTTCTACCTGTTTGTGCTCG
ACCAGTTTCTTAAACTCGACAAACTTAGTGAAGACTTCGTCTTCTTCCGCAAGAAATATACAAAAATCC
GCCTGGACTTGTGCTATGAAAGTGAGAAAGTACTTTGATCCAGCAAGTGACGGTGTGCTAAATGGCC
M_A
GCACAGGTCTGAATGGATCATATCCAACAGCTCTTCAGCTCTGCTCCTCGTTATCTTCGGAAATGGTTGC
R_T_G_L_N_G_S_Y_P_T_A_L_Q_L_C_S_S_L_S_S_E_M_V_A
ACATGGATTTTTCAGCATGCACGTCTTGCACTAGCGTCTGGTTTAAAACGACCTTTTCAACACCGT
H_G_F_L_P_A_C_T_S_C_I_L_R_L_V_*
ACACCATCTTCTTCCATCACCATCTCCTGTAGGCTGCTTGTATTCAAATGGCCATACCTTTTATGCCATAG
TGAACCATCAGCATCAACGGCCGCAAAACAACGTTATGTTTTCCCTTGAACATATACAAATACCAGCT
M_F_P_L_K_H_I_Q_I_T_S
CTCATTACACGCAGTATTCGCTCGCCTTCCCTGAATGACGTCAGCGTAATGTGGTCCAAAATGACAAAAC
S_H_Y_T_Q_Y_S_L_A_F_L_N_D_V_S_V_M_W_S_K_I_D_K_T
ATTTATACTGAGCTGCACGGCTGACTGACATAAAGTTGCCGTTCAAATCTGGGACGAAGAGTACGTTATT
F_I_L_S_C_T_A_D_*
CAATACCAGAGTACATAAAATCAGTCTTCAGCTTCACTGTTTCT

Para la región genómica 2R_36, el programa Augustus Gene Prediction predijo una estructura compuesta por 1 exón de 3010 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido VAISLHLHA con una probabilidad de $\rho=0.026$ y un valor $E=0.000229369724183258$; para el péptido HIQITSSHYTQYSLAFLNDVSVMSKIDK con una probabilidad de $\rho=0.081$ y un valor $E=3.03744367302705e-05$; para el péptido LYSNIEARVCPHLL con una probabilidad de $\rho=0.092$ y un valor $E=4.09554946328805e-06$; para el péptido VSQCQSVACSQREQLK con una probabilidad de $\rho=0.061$ y un valor $E=0.000449345514508928$; para el péptido TNTIGVEFGLLAK con una probabilidad de $\rho=0.088$ y un valor $E=0.000508807322509452$; para el péptido SDKEILPGSSR con

una probabilidad de $\rho=0.075$ y un valor $E=0.000636722643072906$; para el péptido ESVTELSHK con una probabilidad de $\rho=0.054$ y un valor $E=0.000450049714182085$; para el péptido TYHGAPVMVLR con una probabilidad de $\rho=0.003$ y un valor $E=1.48439998597412e-06$; para el péptido SDKEILPGSVEITL con una probabilidad de $\rho=0.003$ y un valor $E=0.000794405317951904$; para el péptido GIIDYSSERAD con una probabilidad de $\rho=0.091$ y un valor $E=0.000241072869175419$; para el péptido MIVECWCPQFVPQDMIFEINSFYIFI con una probabilidad de $\rho=0.063$ y un valor $E=0.000698357520840089$; para el péptido SLTLSSLHTGEEK con una probabilidad de $\rho=0.021$ y un valor $E=0.000685695690553343$; para el péptido STLTDNIFILIPK con una probabilidad de $\rho=0.07$ y un valor $E=0.000401250037091144$; para el péptido GYLSLFGCFHSERTTVPQ con una probabilidad de $\rho=0.003$ y un valor $E=0.000241335605173436$; para el péptido RFQSVGEELSGLATVTFVLCLLV con una probabilidad de $\rho=0.052$ y un valor $E=2.57365994909838e-05$; para el péptido RACTQSSSSNNTATLQYMSTI con una probabilidad de $\rho=0.1$ y un valor $E=0.000840819062297212$; para el péptido ILDNPLNRLIGR con una probabilidad de $\rho=0.012$ y un valor $E=0.000104857096110441$; para el péptido MQLELTGVCV con una probabilidad de $\rho=0.075$ y un valor $E=0.000190342314803903$; para el péptido VMFISGFYRPIKLFVFWVHK con una probabilidad de $\rho=0.018$ y un valor $E=0.000135603127855901$; para el péptido LVFYQKLYPNIECCQPIVLESLVFATVYQEDNEVSFFL con una probabilidad de $\rho=0.066$ y un valor $E=0.000345276025242189$; para el péptido MDLTVSIGSLSSPDK con una probabilidad de $\rho=0.002$ y un valor $E=1.52517535190515e-05$.

Evidencia de ESTs:

gi|4447938|gb|AI533803.1|;

```
AATGCATGCCGTAATTTAATAAAAGTTTAAATTTGAGGTACTCCGTGCTAAATAATAAATATTTAAATATTA
TTATTATTTTATTTATTTATTTCTAATATTTAAATTTTTATTTTCATAATGAATATTTCTTCTTGCAATATCTA
CAATTCCTTATGTAACTTATAAACATTTTTGTGTTTTATTTCCATCTTACTGTTGTAGTAAGAAATATGTA
TATGAGTCCCACAATACTGTACATGACATAATTGATACGAAATGAGACCTACAGTAAAGTAGAATAAAAAAG
CAATGGAGGAATAGCCACAAATTCCTGTTAGAAATTAAGCACAACATTTGGCATCCCTCTCTGTCATGTAAG
TTCAGTATATAAGATCGAATGTGTTACGGTCCAGGCGAAGAAATGAAGAAGAGGAGAAGAGTTTCTTTGT
GCCATTCGGTGGCCTGCTCGCAAAGGGAACAGGTTAAATAAAAGTTATATTTAAATATAAAAGCAGGAGTG
TGTCCTCGTCTTGTGCAACATCAGAAAGTGATCATTACAAATTTGCAAGAACAGCTGCCGCAATGGAATGA
AAAAAGAAA
```

gi|13771353|gb|BG639775.1|

```
ACTCAAGGCGTGTATTTATTGCAAGAAGAATATAAATTATAAAAAATAAAAAATTAATAATAGAATAATAA
ATAAAAAATAAGGATAAGGGTTAAGTATTTATTATACAGCTCGGAGTACCTTAAACTATACTGTCCGAGAGC
TGCCCTGGCAAGTTTTCTTTGTCCGACATTTGTGGGTAATATTCACCTTCTGATGCTATAAAAGATGAGGACA
CACTCTTGCTTCAATATTTGAATATAACTTTATTTTAACTGTTCCCTTTGAGAGCATGCACCAAATGGCA
CAAAGAAACTCTTCTTCAATTTCTGTTCAAGGGCCGTAACAGATTTGATCTTATATACTGAACGTACAAGA
CAGAGAGGGATGACAAATGTTGTGCTTAATTTCTAACGGGAATTTGTGCTTATTACTCCATTG
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 54 sORFs. La transcripción se basa en la presencia de 2 ESTs en la base de datos pública y la probable traducción en la presencia en bases de datos de 21 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos dos de los sORFs codificados por este transcrito son traducidos.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por

espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_44



S Y S P A L L *
A I V P H Y C K I K S A R G R C R L C I *
TATTGCGCAAAAAAATTTAAGTTCGAGTTTTTTAGACGCATCACCATTAATTTTTATAAATAAAAGCC
AATGCACAAGTCAAAAAGTATTTCACTTTTTAATAAGACATCAAGAACAACACTACAGTTGAATCTTTTGTAT
M H K S K V F H F L I R H Q E Q L Q L N L L Y
CTAATTTTTCACGCCATTTAATTTACCCCATGGAGGGTGGCCCCAGGTCTCTTGTCGATTATT
L I F T P F N L P P W R V A P G L L S I I *
M E G G P R S L V D Y T *

Para el gen 2R_44, el programa Augustus Gene Prediction predijo una estructura compuesta por 3 exones de 103 pares de bases el primero, 30 pares de bases el segundo, y 97 pares de bases el tercero, y 2 intrones de 56 pares de bases el primero y 939 pares de bases el segundo .

Se encontró acuerdo con espectros de masas para el péptido CIISRANIYLFNVVQHLI con una probabilidad de $\rho=0.083$ y un valor $E=0.000666519259242421$; para el péptido SKVFHFLIR con una probabilidad de $\rho=0.033$ y un valor $E=7.09239329981059e-05$; para el péptido VAPGLLSII con una probabilidad de $\rho=0.026$ y un valor $E=0.000142575835728852$.

Se halló evidencia de ESTs:

gi|38626652|gb|CK134716.1|;

ATTATTTCCAGGGCAAATATTTATTTAATTTAACGTGGTTCAACATTTAATTTAGTCTAAAAATTACGCAT
TTCATGATTCCATGGGTGAAAGCGATGCATCCAATCCTGGTTGTCTATGTGCTTGCCGTACGTGGAATTG
TAAAAATAGTTAGAACTTACTTGTGTTGTTGCGCCTTCAACCTGGCCTACAAGCGTTAGATAGAATAGCCTT
GTCTGAGGAGAATTTCTCGACATTGGGGCGACTGGGGCAAAGGGGACAGCTCTCCAAATCAAAGTTCGCGAT
GCCAACGAGTCAGGCCGTATCTCCTCTTGATCCTGAACCTCAAGAATTTGGATGTGGGCTTAAGGGAATA
AATACGCCACATCAGTAGACTTGGATGACACATTCCTGATCAGGGCCTCTCCACGGCATCCAAAAGTTTC
CACTCTCAGCCGGTTCCCGAATTTCCAGATCTTTTGAATATGGTACTTGGCAGCTGGGACAATGGGATTA
GCTATAGTCTCGCATTATTGTAAAATGAAAAGCGCTAGAGGGCGCTGTGCGCTTTGTATATAAGTATTAT
ATTGCGCAAAAAAATATCAAGTTCGAGTTTTTTGCACACGTCAAAGTATTTCACTTTTT

gi|15470698|gb|BI579276.1|;

AGTTCCAAAATGCAGCTCATCGTGAAATTACTGATTGTCCTTTTGGCATTATTTGCTGTTCTTTATGGAG
AAGTAAAACATCCACAAAAGATCCTGGAGCCACCTTCCATGGCGGTAAATTAAGTGGCGTGAAAATTAGAT
ACAAAAGATTCAACTGTAGTTGTTCTTGATGTCTTATTA AAAAGTGAATACTTTTACTTGTGCAAAAA
ACTCGAACTTGATAGTTTTTTTTGCGCAATATAAATACTTATATACAAAGGCGACAGCGCCCTCTAGCGCTT
TTCATTTTACAATAATGCGGGACTATAGCTAATCCCATTTGTCCAGCTGCCAAGTACCATATTTCAAAGA
TCTGGAAAATTCGGGAACCGGCTGAGAGTGGAAACTTTTGGATGCCGTGGAGAGGCCCTGATACGGAATGT
GTCATCCAAGTCTACTGATGTGGCGTATTTATTTCCCTTAAGCCACATCCAAATTTCTGGAGTTTCAGGAT
CAAGAGGAGATACGGCCTGACTCGTTGGCATCCGCAGTTTGTATTTGGAGAGCTGTCCCTTTGCCCCAGT
CGCCCAATGTGAGATATTTCTCCTCAGACAAGGCTATTTCTATCTAACGCTTGTAGGCCAGGTTG

gi|15470762|gb|BI579340.1|;

AGTTCCAAAATGCAGCTCATCGTGAAATTACTGATTGTCCTTTTGGCATTATTTGCTGTTCTTTATGGAG
AAGTAAAACATCCACAAAAGATCCTGGAGCCACCTTCCATGGCGGTAAATTAAGTGGCGTGAAAATTAGAT
ACAAAAGATTCAACTGTAGTTGTTCTTGATGTCTTATTA AAAAGTGAATACTTTTACTTGTGCAAAAA
ACTCGAACTTGATAGTTTTTTTTGCGCAATATAAATACTTATATACAAAGGCGACAGCGCCCTCTAGCGCTT
TTCATTTTACAATAATGCGGGACTATAGCTAATCCCATTTGTCCAGCTGCCAAGTACCATATTTCAAAGA
TCTGGAAAATTCGGGAACCGGCTGAGAGTGGAAACTTTTGGATGCCGTGGAGAGGCCCTGATACGGAATGT
GTCATCCAAGTCTACTGATGTGGCGTATTTATTTCCCTTAAGCCACATCCAAATTTCTGGAGTTTCAGGAT
CAAGAGGAGATACGGCCTGACTCGTTGGCATCCGCAGTTTGTATTTGGAGAGCTGTCCCTTTGCCCCAGT
CGCCCAATGTGAGATATTTCTCCTCAGACAAGGCTATTTCTATCTAACGCTTGTAGGCCAGGTTG

gi|14631551|gb|BI165744.1|;

GCAGTTCCAAAATGCAGCTCATCGTGAAATTACTGATTGTCCTTTTGGCATTATTTGCTGTTCTTTATGG
AGAAGTAAAACATCCACAAAAGATCCTGGAGCCACCTTCCATGGCGGTAAATTAAGTGGCGTGAAAATTAG
ATACAAAAGATTCAACTGTAGTTGTTCTTGATGTCTTATTA AAAAGTGAATACTTTTACTTGTGCAAAA

```
AAACTCGAACTTGATAGTTTTTTTTGCGCAATATAAATACTTATATACAAAGGCGACAGCGCCCTCTAGCGC  
TTTTTCATTTTACAATAATGCGGGACTATAGCTAATCCCATTTGTCCCAGCTGCCAAGTACCATATTCAAAA  
GATCTGGAAATTCGGGAACCGGCTGAGAGTGGAAACTTTTGGATGCCGTGGAGAGGCCCTGATACGGAAT  
GTGTCAATCCAAGTCTACTGATGTGGCGTATTTATTCCTTAACCCACATCCAAATTCCTGGAGTTCANGA  
TCAAGAGGAGATACGGCCTGACTCGTTGGCATCCGCAGTTTGATTTGGAGAGCTGTCCCT
```

gi|15325207|gb|BI485029.1|;

```
AGTTCCAAAATGCAGTCATCGTGAAATTACTGATTGTCCTTTTGGCATTATTTGCTGTTCTTTATGGAG  
AAGTAAAACATCCACAAAGATCCTGGAGCCACCTTCCATGGCGGTAAATTAAGTGGCGTGAAAATTAGAT  
ACAAAAGATTCAACTGTAGTTGTTCTTGATGTCTTATTAATAAGTGAATACTTTTGACTTGTGCAAAAA  
ACTCGAACTTGATAGTTTTTTTTGCGCAATATAAATACTTATATACAAAGGCGACAGCGCCCTCTAGCGCTT  
TTCATTTTACAATAATGCGGGACTATAGCTAATCCCATTTGTCCCAGCTGCCAAGTACCATATTCAAAAGA  
TCTGGAAATTCGGGAACCGGCTGAGAGTGGAAACTTTTGGATGCCGTGGAGAGGCCCTGATACGGAATGT  
GTCATCCAAGTCTACTGATGTGGCGTATTTATTCCTTAAGCCACATCCAAATTCCTGGAGTTCAGGAT  
CAAGAGGAGATACGGCCTGACTCGT
```

gi|15469717|gb|BI578295.1|;

```
AGGGCCAAAAGGGGAGCTCATCGTGAAATTACTGATTGTCCTTTTGGCATTATTTGCTGTTCTTTATGGAG  
AAGTAAAACATCCACAAAGATCCTGGAGCCACCTTCCATGGGGGAAATTAAGTGGCGTGAAAATTAGAT  
ACAAAAGATTCAACTGTGGGGGGGTTGATGTCTTATTAATAAGTGAATACTTTTGACTTGTGCAAAAA  
ACTCGAACTTGATAGTTTTTTTTGCGCAATATAAATACTTATATACAAAGGCGACAGCGCCCTCTAGCGCTT  
TTCATTTTACAATAATGCGGGACTATAGCTAATCCCATTTGTCCAGCTGCCAAGTACCATATTCAAAAGA  
TCTGGAAATTCGGGAACCGGCTGAGAGTGGAAACTTTTGGATGCCGTGGAGAGGCCCTGATACGGAATGT  
GTCATCCAAGTCTACTG
```

gi|48957813|gb|CO184932.1|

```
TGAAATTAAGTATTGTCCTTTTGGCATTATTTGCTGTTCTTTATGAAGAAGTAAAACATCCACAAAGATC  
CTGGAGCCACCTTCCATGGCGGTAAATTAATGGCGTGAAAATTAGATACAAAAGATTCAACTGTAGTTG  
TTCTTGATGTCTTATTAATAAGTGAATACTTTTGACTTGTGCAAAAACTCGAAAAAAAAAAAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAGGAAA
```

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 19 sORFs. Se pueden observar varias repeticiones en tándem y varias desplazadas. Los sORFs decimoquinto, decimosexto y decimoséptimo son duplicaciones desplazadas del octavo, noveno y décimo sORF, respectivamente. La transcripción se basa en la presencia de 7 ESTs en la base de datos pública y la probable traducción en la presencia en bases de datos de 3 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación. Esto significa que al menos 3 de los sORFs codificados por este transcrito son traducidos.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

2. REGIONES GENÓMICAS CON VALIDACIÓN SÓLO POR ESPECTROMETRÍA DE MASAS

Gen 2L_16



GTGATCTTAAATTTTCAGGATGGTTCGAATTGTAGTAGGCTAGAGGTACATGTGCGCAGCGTTGACATTTTCT
TGTGCCTGATCTCCCATGTTGGCAATTAACATTTTGTAGTGAAGCGAATTTCCAACCGCGTGCGCGATCG
TCTGTATGCAACTGATCGTATTTTCTACGCATCGCATTCTACACATGCTAGCTGGACTACACTAGGCGAC
M_Q_L_I_V_F_S_T_H_R_I_L_H_M_L_A_G_L_H_*
AGCGCGACAACGCGTCGCTCGACACTCGCCTCACTCGCCTCACTCGTAACGGTACTTTGGCGCCAGTC
ATATTCTTCGTGGTAGTCAAATTTCTAAGAATGACCTCTCAATATACCCAACCTCATTGCGCAGCTGCGTCT
M_T_S_Q_Y_T_Q_L_I_A_Q_L_R_L
TGCAAGCACAGTGGTCAAACCATGTTAATAATATGACAGACACGTACAGTGGTCAGAATACATCCGTAAT
A_S_T_V_V_K_P_C_* M_T_D_T_Y_S_G_Q_N_T_S_V_M
GCGACACGGCCTTCCTGACATAATACACGTCCCTCGTGTGTTGTCTTAAATCTACCACATAAGTGATAATA
R_H_G_L_P_D_I_I_H_V_L_V_C_C_L_K_S_T_T_*
CTTCAGGATTCCAAGGCTCTCTGGATATTTCTAAGAATGTTTTCCATTTATTCTATTTCTATTTCTTTTT
M_F_S_I_Y_S_I_S_I_S_F_
ATTCTCAAATTTTCTTTGTCTACCTCATCATCGTTATCTTCGTTTAGCGGCAACAAGCCAAAAGGTCGAG
*
CTTCTTTACCAATCAATAATTCAAAAGGGCTAAACTTAGTCACACGATTAATCGTGCAATTCATTCCTAG
CTGCACTTCAGCCAACGCATCCTGCCATGATCTCTGACTAGACTCGACAGCCGTTAGCATACTTTTCAGG
GTGCTCATGACACGTCTTACCTGACCGTTTGTCTACTTTTACCTGTAGCAATTAATGTAGATTAATTT
TTTGAGAGGAGCAAAAATCGCGAAAGCTTTTACTTGGCAAACCTACGTCCCTGATCTGCTATAATTCGAGT
GGGTAAACCAACAAGGACACGGCTGATTTTACAGCCTTAATGCAGCTATCAGAATATACATTTAGTGTG
M_Q_L_S_E_Y_T_F_S_V_
TGATGAAAATAAACAACTTGGTGAAGGCATCTAAAGGACGATGAAATATTCCTTCAAATCATTTTTGCC
*
M_K_I_N_K_L_G_E_G_I_*
ACTCAATTTGCCTGTGATGTCAATATGTATCGTATGCCACGGTATTTTCTATTTTAGGAATAGGATTCAGC
M_Y_R_M_P_R_Y_F_Y_F_R_N_R_I_Q_
TCAGCCTGCACTTTTCCCGTAGTGGCTTAGAAAATTTACATGTAATGCAATTTTCTACGAATCTACACA
L_S_L_H_F_S_R_*
CATATTTAGGCATATTTTTCGAACCAGTAATAACAATAGACTTTTGTCCTTTATGTTCTTCAAGCACTGGAT
TTCGAGAGAGAAAATCCACATGCGCCAATTCGCTCGCCTTTTATATACTCTATTTTCGAAATCAAATGCCT
GTAAAAACGCCACCATCTATGCACCTCGGTGTTAAATCTAGTTTTGTACGGCTGGCCTTTAAAGAATT
ACAGTCAGTATGTAATAAAATTTACGTCCCTGCAAATAATGCCGGAAGTGTTTTATAGCATTTACCACT
M_Y_Y_K_F_T_S_L_Q_I_M_P_E_V_F_Y_S_I_Y_H_
GCTAACGTCTCCAGTTCGTATGAATGGTATTTAGATTCGGCGGGAGACGTACACTTACTGTAATACTCTA
C_* M_V_F_R_F_G_G_R_R_T_L_T_V_I_L_Y
CAACATGAGGTTTATTGTCAATTTTATGTAACAGGATCGCTCCATACCCAATACAACCTTGCCTCAGTATG
N_M_R_F_I_V_N_F_M_* M_
CAGCTCTATTTGGTATTGCGGATCGAATATGATTAATAACTGGCGTCTGGGTCAATATAGTTATTATTTT
Q_L_Y_L_V_L_R_I_E_Y_D_*
M_I_K_T_G_V_W_V_N_I_V_I_I_F_
ATTGCGACCTGTTTCATGCACTGGTAAACCAAACAAAGTCAGTGTTTTTCTTAGATGTGAGAAGATACAGAG
I_R_T_C_S_C_T_G_N_Q_T_K_S_V_F_F_L_D_V_R_R_Y_R_
GTTTAAAGTGTGGGAGAAAATTTAACACAAATTTGACGAAAATATGTGGCTAAACCAATGAATTTGTCTTAG
G_L_S_V_W_E_K_F_N_T_N_* M_N_C_L_S
TTGTGAGACATTTTGGGGTATGGTAATGCTTTTAAAGCTTGAATTTTTCGTGTATTTGGCCTGACTTCT
C_E_T_F_W_G_D_G_N_A_F_K_A_*
M_V_M_L_L_K_L_E_F_F_V_Y_L_A_*
CCCTTGCTTATTTCAAATCCGAGATACTCTATTTTTGTGGTGAAGGAACACTTAGTGACATTAATG

M
 AAAATCCAGCCTTACTTAATGTCTCCAACACTCTATGCAATCTAACTAGCGCATCTTCTTTACTTTCTGC
 K I Q P Y L M S P T L Y A I *
 TATTATGAGAACATCATCAATATACACAATTGCATAAGTGTAGGCTAAATCTCCTAGGGCCTTCATTATA
 M R T S S I Y T I A *
 GCCCCTGGAAAAACAGATGGTGGCTTTTTCAAACCAAACGGCATTGTCAAGAAGCTCGTATTGACCCTCTG
 M V R F S N Q T A L S R T R I D P L
 GCGTTACAAACGCTGTTTCGTTCAATTGAGTCCATGGACTGGAATTTGATAAAAAGCCGCTTGCCATATC
 A L Q T L F V Q L S P H G L E F D K S R L P Y
 M D W N L I K A A C H I
 TATGCACGAAAAATATTTAGCACCCCCGAAGCCTAGCAATTTGGTCTGAGATGAGGGGCAATGGATACTTG
 L C T K N I *
 Y A R K I F S T P K P S N L V *
 M H E K Y L A P R S L A I W S E M R G N G Y L
 M D T C
 TCAGCCACTGTATTTTCATTTAGCATGCGATAGTCCACACCCAGTCGATCGGTGCCATTCTTTTTTTTTTGC
 S A T V F S F S M R *
 Q P L Y F H L A C D S P H P V D R C H S F F L
 ACTAGCATCATGGGACTAGCGAAAGTTGAACAGCTTGGCCGAATAAGCTTTTAGTAATTAATCAACTTTA
 T S I M G L A K V E Q L G R I S F *
 CCCCCACTTTTGCTTTTCTCTTCTTCACTAAGAGTCTTTTAAATTTGTGGCCATGTGATGCCAGCAAAGCA
 TATTTGGGACAGCCATTGTGATGCACTTCCCTCTAGCGCCTTACTTAAAGTTATCATCAGGCTGCTACCT
 M H F P L A P Y L K L S S G C Y L
 TCGGGCATTTTATCAGCAAAAATGAGATCCACTGTTGTACACCAAGCTGAGGCATCTGCACCAGCTCCCT
 R A F Y Q Q K *
 CAGGATGAAACTTTGGGAGAATTATATATTTCCCTTTTCATCAGTTCCGATGCCTTTGGTGTCTGCAT
 M K L W E N Y I F S L F I S F R C L W C L H
 ATTTTGTAGATCTCCATTAGGTTGTTGTTTTGCACCTGTAAAGCTGCCAGGTATGCATCGGTGATATTT
 I F D D L F *
 M I S I R L L F C T C K A A R Y A S V I F
 TTAGGCAAAGCAGATCCCCTTCTGATGTCGCGAGCGTTGACATTTTCTTGTGCCTGATCTCCCATGTTGG
 L G K A D P T S D V A A L T F S C A *
 CAATTAAACTATTTGAGTGAAGCGAATTCACACCGCGTGCAGCATCGTCTGTATGCAACTGATCGTATT
 M Q L I V F
 TTCTACGCATCGATCTATCCTTTACACAGCTAGCTGGACTACACTAGGCGACAGCGGACAACGCCGTC
 S T H R S I L Y T R *
 GCCTCACTCGTAACGGTACTTTGGCGCCAGTCATATTCTTCGTGGTAGTCAAATTTCTAAGAATGACCTCT
 M T S
 CAATATAACCCAACTCATTGCGCAGCTGCGTCTTGCAAGCACAGTGGTCAAACCATGTTAATAATATGACA
 Q Y T Q L I A Q L R L A S T V V K P C * M T
 GACACGTACAGTACACATAAGTTCAGTTCATAATAGCAACATAGTATATAAACATTAAGTCATTGT
 D T Y S T H T *
 TCGTTTTCGGGTGCATTCTCTCTGTTTCGCAACATCGGCTCTGTAGCAACAAGCATAGAGGCAGATCGCA
 AACGGAGAGGGAGTCAGTGAATGCTAAAGCTGAACACGGGAAGTAACAAATAAACCAAAGAGTCCGACT
 M L K L N T G S N K *
 CCTGCTTTTGTG

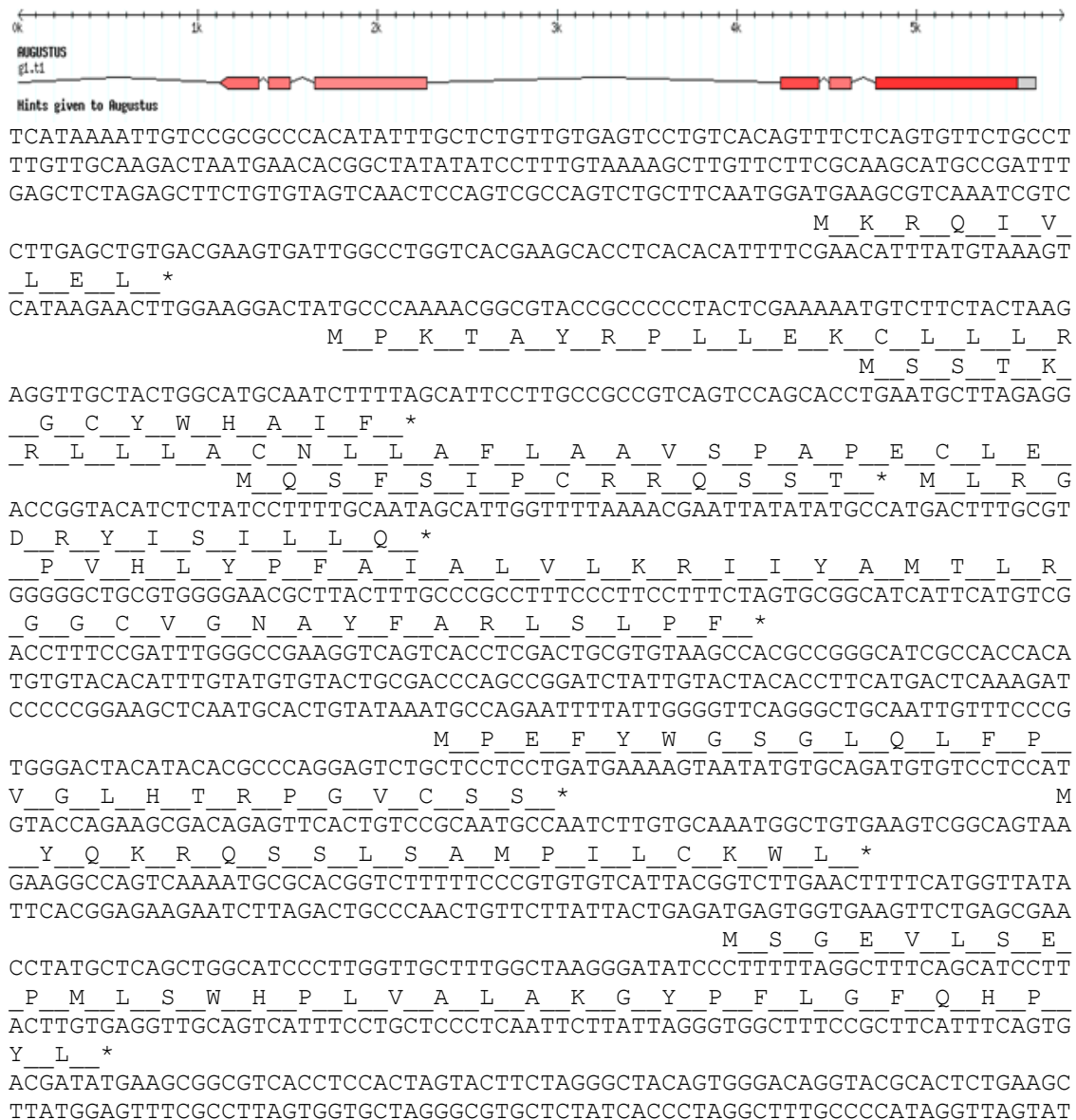
Para el cluster de sORFs 2L_16, el programa Augustus Gene Prediction predijo una estructura de 2 exones, uno de 428 pares de bases y el otro de 50 pares de bases, y un intrón de 1734 pares de bases.

Se encontró acuerdo con espectro de masas para el péptido MQLIVFSTHR con una probabilidad de $\rho=0.031$ y un valor $E=1.77918840258577e-05$; para el péptido MTDYSGQNTSVMR con una probabilidad de $\rho=0.042$ y un valor $E=0.000643597686829289$; para el péptido ISIRLLFCTCK una probabilidad de $\rho=0.02$ y un valor $E=0.00065656819598243$; y para el péptido QLYLVLRIEYD una probabilidad de $\rho=0.019$ y un valor $E=0.000174133872588637$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 26 sORFs. El sORf vigésimocuarto es una duplicación desplazada del segundo sORF. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 4 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_39



M E F R L S G A R A C S I T L G F A P * M
GGGTCTGACTACCATGAGGTACAGCCATCTTATTATGCGTGGTGGTTCACCCACGACTTACCAGCAATG
G L T T M R Y S H L I M R G E V P H D L P A M
TTTCTACACGTGAAAAGTGTCTGGTGCACCTTATCAATTTTGTATCGACATGAGTTTTGAAGCTGAGTT
F L H V K S A L V H L S I L L S T *
M S F E A E F
TGGAGTAAAGGGCTATCCCAAGATACTTAACTTCATTGCTGGCCTCTATGCCACATTTTGACAGGGCTAT
G V K G Y P K I L N F I A G L Y A T F *
M P H F D R A I
TGCCCTCATTCTCTGCTAGGATTTAGGCTCAGGCCTACTTCCTTGAACCATTCGCTGGTCATATTGATGC
A F I L C *
CCAGTTGAATGATATCGCAGAGTGTTTCCCTCATATTTACCTCTGGCTATAATGACAATATCGTCAGCATA
M I S Q S V S S Y L P L A I M T I S S A *
GCCTTGGCAAAGTATACCTCTGCGGGTCAGTCTGTCCAGCAACTCGCGTACCAGCAAGCTCCACAGGAGA
GGCGAGAGGACACCCCTTGGGAGCAACCTCTTGTGGTAGGTACTGTGGACGACTGCCCTCCTATAGTGG
CCATGGCTCGCTTAGATGCCAGTAGTACTTGTATCCATCTGCTGGTTGTTGCATCTAGTCCCTCTTCTTGC
CAGGGATGCGTTGACAGCTTCATGGGAGGTATTGTGCAATGCACCCTGTATGTCTAAGAAGGCACATAAC
M G G I V E C T L Y V *
M H P V C L R R H I T
GCCACTTCCTTATGAGTCAGTGAATCCTCAATAAGGCTTTTCAGGTGATATAGGGCAGTATCAGTTGATC
P L P Y E S V N P Q *
TTCCCGCCATGTAGGCATGCTGCGTTCGTTGGCGACGCCTGGTGCCCATATTACGTGATGTGCATTTGC
M L R S L A T P G A P Y Y V M C I C
ATCACATCCAATGATGATGGTTGAGTGCATCCTCTTGCAGTGTTCAGCCAGTGCAGTGATTTGGGCGGT
I T S N D D G *
GGGCACGCCTCGTCACCGTCAAAGTATGCCGATGCCACCACAGTCTCCTTGGCTGTTCCACATCTTCTA
M P M P P Q S P W L F P H L L
CCACGATCGGTACACAGTCCCGCGTGAGAAATCTGTAAGTATATAATAATAACTACAACATGCCAG
P R S V H S P A *
GATACAGGCTCTAGGTCTAGAAGCCCATTTGATCCCAAATTACCCAGGCGTTCCTTACGTACAGTCCACGT
ATGCCCTGGTTATACCATGGTTTCTGTACCAGGGCTAGCCCAAGATGTTGCTTGGTGAACATCCTTGCCA
M V S V P G L A Q D V A W *
M L L G E H P C Q
ATCGACGTGAGCCGTGATTTGCAGTACTCCCTTGAATGGATGTTTTACCTACTTAACATTACCCTCCA
S T *
ACAGCGGGCCACAGCTTGGGCGACTGAGGAGAATTCGAATTAAGCTTGGATCTATGAGGTTCTCGTAATC
CCCCTGAAGCTTTCTGAAGAACGCCTTCATAAAATTGTCCGCGCCACATATTTGCTCTGTTGTGAGTCC
TGTCACAGTTTTCTCAGTGTCTGCCTTTTGTGCAAGACTAATGAACACGGCTATATATCCTTTGTA
CTTGTCTTTCGCAAGCATGCCGATTTGAGCTCTAGAGCTTCTGTGTAGTCAACTCCAGTCGCCAGTCTGC
TTCAATGGATGAAGCGTCAAATCGTCCCTTGGAGCTGTGACGAAGTATTGGCCTGGTCACGAAGCACCTCA
M K R Q I V L E L *
CACATTTTCGAACATTTATGTAAGTCAATAAGAACTTGGAAAGACTATGCCAAAACGGCGTACC
M P K T A Y R P
CTACTCGAAAAATGCTTCTACTAAGAGGTTGCTACTGGCATGCAATCTTTTAGCATTCCTTGCCGCGT
L L E K C L L L R G C Y W H A I F *
M S S T K R L L L A C N L L A F L A A V
M Q S F S I P C R R
CAGTCCAGCACCTGAATGCTTAGAGGACCGGTACATCTCTATCCTTTTGCAATAGCATTTGGTTTTAA
S P A P E C L E D R Y I S I L L Q *
Q S S T * M L R G P V H L Y P F A I A L V L K
GAATTATATATGCCATGACTTTGCGTGGGGGCTGCGTGGGGAACGCTTACTTTGCCCGCCTTTCCCTTCC
R I I Y A M T L R G G C V G N A Y F A R L S L P
TTTCTAGTGCGGCATCATTATGTCGACCTTCCGATTTGGGCCGAAGGTCAGTCACCTCGACTGCGTGT
F *
AAGCCACGCCGGGCATCGCCACCACATGTGTACACATTTGTATGTGTACTGCGACCCAGCCGGATCTATT
GTACTACACCTTCATGACTCAAAGATCCCCCGGAAGCTCAATGCACGTGTATAAATGCCAGAATTTTATTG
M P E F Y W
GGGTTACAGGGCTGCAATTTGTTCCCGTGGGACTACATACACGCCCCAGGAGTCTGCTCCTCCTGATGAAA
G S G L Q L F P V G L H T R P G V C S S *
GTAATATGTGCAGATGTGTCTCCATGTACCAGAAGCGACAGAGTTCACTGTCCGCAATGCCAATCTTGT
M Y Q K R Q S S L S A M P I L

GCAAATGGCTGTGAAGTCGGCAGTAAGAAGGCCAGTCAAAATGCGCACGGTCTTTTTCCCGTGTGTCATT
C_K_W_L_*
ACGGTCTTGAACCTTTTCATGGTTATATTCACGGAGAAGAATCTTAGACTGCCTAAGCCCTGGTAGGTTTT
CCCGGAAGAACAGACGTTTTTCTCTATGGTGACCAACAGGACAGAAGGGCTCGGGCCCAATTAGAGGGTT
M_V_T_N_R_T_E_G_L_G_P_N_*
TCTTTGCCTCTTTCCTGGCTAGGTTGTCGGCTAGCTCATTTCCCGGGATGTTGTTGTGTCCCGGGACCCA
M_L_L_C_P_G_T_H
CCTTATGTGGGTTTGTGACAGCTCCTAGTTTGTCTAGGGCTGTCCTCACTTTACTTACTACCTTAGAT
L_I_V_G_L_L_T_A_P_S_L_S_R_A_V_L_T_L_L_T_T_L_D
GTTATCTTAGCTTTGCTGAGCGCCTTTATGGCTGCTTGACTATCAGACAGTATAGCAATGTCCCTGCCAC
V_I_L_A_L_L_S_A_F_M_A_A_*
GATAGTTTCTTTGCAGATTAACCTCCGCACAGCGTACAATAGCACAGACTTCAGCTTGAAAAATGCTGGT
M_L_V
GTATCTGCCCATTTGATTCGTAGTATTTCAACGTGGGGCTACAACGCCAGCCCACTTCCCTCGTCGGTG
Y_L_P_I_D_S_*
AGGGATCCGTCTGTATAACACTTTTATTGCTTGTGGTTTTCATAGGGTGTACTTTCCCCAGGGATGTCCAAC
TGTTCTTATTACTGAGATGAGTGGTGAAGTCTGAGCGAACCTATGCTCAGCTGGCATCCCTTGGTTGCT
M_S_G_E_V_L_S_E_P_M_L_S_W_H_P_L_V_A
TTGGCTAAGGATATCCCTTTTTAGGCTTTCAGCATCCTTACTTGTGAGGTTGCAGTCATTTCCCTGCTCC
L_A_K_G_Y_P_F_L_G_F_Q_H_P_Y_L_*
CTCAATTCTTATTAGGGTGGCTTTCCCTCTTCAATTCATGACGATATGAAGTGGCGTCACCTCCACTAGT
ACTTCTAGGGTTGCAGTGGGACAGGTACGCACTCTGAAGCTTATGGAGTTTCGCCTTAGTGGTGTAGGG
M_E_F_R_L_S_G_A_R
CGTGCTCTATCACCCCTAGGCTTTGCCCCATAGGTTAGTATGGGTCTGACTACCATGAGGTACAGCCACT
A_C_S_I_T_L_G_F_A_P_* M_G_L_T_T_M_R_Y_S_H_L
TATTATGCGTGGTGGTTCCTCCACGACTTACCAGCAATGTTTCTACACGTGAAAAGTGTCTGGTGCAC
I_M_R_G_E_V_P_H_D_L_P_A_M_F_L_H_V_K_S_A_L_V_H
TTATCAATTTTGTATCGACATGAGTTTTGAAGCTGAGTTTGGAGTAAAGGGCTATCCCAAGATACTTAA
L_S_I_L_L_S_T_*
M_S_F_E_A_E_F_G_V_K_G_Y_P_K_I_L_N
CTTCATTGCTGGCCTCTATGCCACATTTTGACAGGGCTATTGCCTTCATTCTCTGCTAGGATTTAGGCTC
F_I_A_G_L_Y_A_T_F_*
M_P_H_F_D_R_A_I_A_F_I_L_C_*
AGGCCTACTTCTTGAACCATTTCGCTGGTCATATTGATGCCAGTTGAATGATATCGCAGAGTGTTTCTCT
M_I_S_Q_S_V_S
CATATTTACCTCTGGCTATAATGACAATATCGTCAGCATAGCCTTGGCAAAGTATACCTCTGCGGGTCAG
S_Y_L_P_L_A_I_M_T_I_S_S_A_*
TCTGTCCAGCAACTCGCGTACCAGCAAGCTCCACAGGAGAGGCGAGAGGACACCCCTTGGGAGCAACCT
CTTGTGGTAGGTAAGTGTGGACGACTGCCCTCCTATAGTGGCCATGGCTCGCTTAGATGCCAGTAGTACT
TGATCCATCTGCTGGTTGTTGCATCTAGTCCCTTCTTGGCAGGGATGCGTTGACAGCTTCATGGGAGGT
M_G_G
ATTGTCGAATGCACCCCTGTATGTCTAAGAAGGCACATAACGCCACTTCCCTTATGAGTCAGTGAATCCTCA
I_V_E_C_T_L_Y_V_*
M_H_P_V_C_L_R_R_H_I_T_P_L_P_Y_E_S_V_N_P_Q
ATAAGGCTTTTTCAGGTGATATAGGGCAGTATCAGTTGATCTTCCCGCCATGTAGGCATGCTGCGTTTCGTT
* M_L_R_S_L
GGCGACGCCTGGTGCCCCATATTACGTGATGTGCATTTGCATCACATCCAATGATGATGGTTGAGTGCAT
A_T_P_G_A_P_Y_Y_V_M_C_I_C_I_T_S_N_D_D_G_*
CCTCTTGCAGTGTTCAGCCAGTGCAGTGATTTCCGGCGGTGGGCACGCCTCGTCACCGTCAAAGTATGCC
M_P
GATGCCACCACAGTCTCCTTGGCTGTTCCACATCTTCTACCACGATCGGTACACAGTCCCAGCTGAGAA
M_P_P_Q_S_P_W_L_F_P_H_L_L_P_R_S_V_H_S_P_A_*
ATTCTGTAAGTATATAATAATAATACTACAACATGCCAGGATACAGGCTCTAGGTCTAGAAGCCATTG
ATCCCAAATTACCCAGGCGTTCTTACAGTACAGTCCACGTATGCCCTGGTTATACCATGGTTTCTGTACC
M_V_S_V_P
AGGGCTAGCCCAAGATGTTGCTTGGTGAACATCCTTGCCAATCGACGTGAGCCGTCGATTTGCAGTGACT
G_L_A_Q_D_V_A_W_*
M_L_L_G_E_H_P_C_Q_S_T_*
CCCTTGAATGGATGTTTACCTACTTAACTTACCCCTCAACAGCGGGCCACAGCTTGGGCGACTGAGGA
GAATTGCAATTAAGCTTGGATCTATGAGGTTCTCGTAATCCCCCTGAAGCTTCTGAAGAACGCCTTCAT
AAAATTGTCCGCGCCACATATTTGCTCTGTTGTGAGTCTGTACAGTTTCTCAGTGTCTGCCTTTGT

TGCAAGACTAATGAACACGGCTATATATCCTTTGTAAAAGCTTGTTCTTCGCAAGCATGCCGATTTGAGC
TCTAGAGCTTCTGTGTAGTCAACTCCAGTCGCCAGTCTGCTTCAATGGATGAAGCGTCAAATCGTCCCTTG
M_K_R_Q_I_V_L
AGCTGTGACGAAGTGATTGGCCTGGTCCACGAAGCACCTCACACATTTTCGAACATTTATGTAAAGTCATA
E_L_*
AGAACTTGGAAAGGACTATGCCCAAAACGGCGTACCGCCCCCTACTCGAAAAATGTCTTCTACTAAGAGGT
M_P_K_T_A_Y_R_P_L_L_E_K_C_L_L_L_R_G
M_S_S_T_K_R
TGCTACTGGCATGCAATCTTTTAGCATTCCTTGCCGCCGTCAGTCCAGCACCTGAATGCTTAGAGGACCG
C_Y_W_H_A_I_F_*
L_L_L_A_C_N_L_L_A_F_L_A_A_V_S_P_A_P_E_C_L_E_D_R
M_Q_S_F_S_I_P_C_R_R_Q_S_S_T_*_M_L_R_G_P
GTACATCTCTATCCTTTTGAATAGCATTTGGTTTTAAACGAATTATATATGCCATGACTTTGCGTGGGG
Y_I_S_I_L_L_Q_*
V_H_L_Y_P_F_A_I_A_L_V_L_K_R_I_I_Y_A_M_T_L_R_G
GCTGCGTGGGGAACGCTTACTTTGCCCGCCTTCCCTTCTTTCTAGTGCGGCATCATTCATGTGACCT
G_C_V_G_N_A_Y_F_A_R_L_S_L_P_F_*
TTCCGATTTGGGCCGAAGGTCAGTCACCTCGACTGCGTGTAAAGCCAGCCGGGCATCGCCACCACATGTG
TACACATTTGTATGTGACTGCGACCCAGCCGGATCTATTGTACTACACCTTCATGACTCAAAGATCCCC
CGGAAGCTCAATGCACTGTATAAATGCCAGAATTTTATTGGGGTTCAGGGCTGCAATGTTTCCCCTGGG
M_P_E_F_Y_W_G_S_G_L_Q_L_F_P_V_G
ACTACATACACGCCCAGGAGTCTGCTCCTCCTGATGAAAAGTAATATGTGCAGATGTGTCCTCCATGTAC
L_H_T_R_P_G_V_C_S_S_*_M_Y_Q_K_R_Q_S_S_L_S_A_M
CAGAAGCGACAGAGTTCACTGTCCGCAATGCCAATCTTGTGCAAATGGCTGTGAAGTCGGCAGTAAGAAG
P_I_L_C_K_W_L_*
GCCAGTCAAAATGCGCACGGTCTTTTTCCCCTGTGTGTCATTACGGTCTTGAACCTTTTCATGGTTATATTCA
CGGAGAAGAATCTTAGACTGCCTAAGCCCTGGTAGGTTTTCCCGAAGAACAGACGTTTTTCTCTATGGT
M_V
GACCAACAGGACAGAAGGGCTCGGGCCCAATTAGAGGGTTTTCTTTGCCTCTTTCCTGGCTAGGTTGTCCG
T_N_R_T_E_G_L_G_P_N_*
CTAGCTCATTTCCCAGGATGTTGTTGTGTCGCCGGGACCCACCTTATTGTGGGTTTTGTTGACAGCTCCTAG
M_L_L_C_P_G_T_H_L_I_V_G_L_L_T_A_P_S
TTTGTCTAGGGCTGTCTCACTTTACTTACTACCTTAGATGTTATCTTAGCTTTGCTGAGCGCCTTTATG
L_S_R_A_V_L_T_L_L_T_T_L_D_V_I_L_A_L_L_S_A_F_M
GCTGCTTGACTATCAGACAGTATAGCAATGCCCTGCCACGATAGTTTTCTTTGCAGATTAAACTCCGCAC
A_A_*
AGCGTACAATAGCACAGACTTCAGCTTGAAAAATGCTGGTGTATCTGCCATTGATTCGTTAGTATTTCAA
M_L_V_Y_L_P_I_D_S_*
CGTGGGGCCTACAACGCCCAGCCCACTTCCCTCGTCCGGTGGGGATCCGCTGTATACCCTTTATTGCT
TGTGGTTTTCATAGGGTGTACTTTCCCCAGGGATGTCCAACGTCTTATTACTGAGATGAGTGGTGAAGT
M_S_G_E_V
TCTGAGCGAACCTATGCTCAGCTGGCATCCCTTGGTTGCTTTGGCTAAGGGATATCCCTTTTAGGCTTT
L_S_E_P_M_L_S_W_H_P_L_V_A_L_A_K_G_Y_P_F_L_G_F
CAGCATCCTTACTTGTGAGGTTGCAGTCAATTCCTGCTCCCTCAATTCCTATTAGGGTGGCTTTCCCTCT
Q_H_P_Y_L_*
CATTTCAATGACGATATGAAGTGGCGTCACCTCCACTAGTACTTCTAGGGTTGCAGTGGGACAGGTACGC
ACTCTGAAGCTTATGGAGTTTCGCCTTAGTGGTGTAGGGCGTGTCTATACCCTAGGCTTTGCCCAT
M_E_F_R_L_S_G_A_R_A_C_S_I_T_L_G_F_A_P
AGGTTAGTATGGGTCTGACTACCATGAGGTACAGCCATCTTATTATGCGTGGTGGGTTCCCCACGACTT
*_M_G_L_T_T_M_R_Y_S_H_L_I_M_R_G_E_V_P_H_D_L
ACCAGCAATGTTTCTACACGTGAAAAGTGTCTGGTGCACCTTATCAATTTTGTATCGACATGAGTTTTG
P_A_M_F_L_H_V_K_S_A_L_V_H_L_S_I_L_L_S_T_*
M_S_F
AAGCTGAGTTTGGAGTAAAGGGCTATCCCAAGATACTTAACTTCATTGCTGGCCTCTATGCCACATTTTG
E_A_E_F_G_V_K_G_Y_P_K_I_L_N_F_I_A_G_L_Y_A_T_F_*
M_P_H_F
ACAGGGCTATTGCCTTCATTCTCTGCTAGGATTTAGGCTCAGGCCTACTTCTTGAACATTTCGCTGGT
D_R_A_I_A_F_I_L_C_*
ATATTGATGCCAGTTGAATGATATCGCAGAGTGTTCCTCATATTTACCTCTGGCTATAATGACAATAT
M_I_S_Q_S_V_S_S_Y_L_P_L_A_I_M_T_I
CGTCAGCATAGCCTTGGCAAAGTATACCTCTGCGGGTTCAGTCTGTCCAGCAACTCGCGTACCAGCAAGCT

S _ S _ A _ *
 CCACAGGAGAGGCGAGAGGACACCCCCCTTGGGAGCAACCTCTTGTGGTAGGTACTGTGGACGACTGCCCT
 CCTATAGTGGCCATGGCTCGCTTAGATGCCAGTAGTGACTTGATCCATCTGCTGGTTGTTGCATCTAGTC
 CTCTTCTTGCCAGGGATGCGTTGACAGCTTCATGGGAGGTATTGTGCGAATGCACCCTGTATGTCTAAGAA
 M _ G _ G _ I _ V _ E _ C _ T _ L _ Y _ V _ *
 M _ H _ P _ V _ C _ L _ R _
 GGCACATAACGCCACTTCTTATGAGTCAGTGAATCCCTCAATAAGGCTTTTCAGGTGATATAGGGCAGTA
 R _ H _ I _ T _ P _ L _ P _ Y _ E _ S _ V _ N _ P _ Q _ *
 TCAGTTGATCTTCCCGCCATGTAGGCATGCTGCGTTGCTGGCGACGCCTGGTGCCCATATTACGTGAT
 M _ L _ R _ S _ L _ A _ T _ P _ G _ A _ P _ Y _ Y _ V _ M _
 GTGCATTTGCATCACATCCAATGATGATGGTTGAGTGCATCCTCTTGCAGTGTTCAGCCAGTGCAGT
 C _ I _ C _ I _ T _ S _ N _ D _ D _ G _ *
 TTCCGGGCGGTGGGCACGCCTCGTCACCGTCAAAGTATGCCGATGCCACCACAGTCTCCTTGGCTGTTCC
 M _ P _ M _ P _ P _ Q _ S _ P _ W _ L _ F _ P _
 ACATCTTCTACCAGATCGGTACACAGTCCCCTGAGAAATCTGTAAAGTATATAATAATTAATACTAC
 H _ L _ L _ P _ R _ S _ V _ H _ S _ P _ A _ *
 AACATGCCAGGATACAGGCTCTAGGTCTAGAAGCCATTGATCCCAAATTACCCAGGCGTTCTTCACGTA
 CAGTCCACGTATGCCCTGGTTATACCATGGTTTCTGTACCAGGGCTAGCCCAAGATGTTGCTTGGTGAAC
 M _ P _ W _ L _ Y _ H _ G _ F _ C _ T _ R _ A _ S _ P _ R _ C _ C _ L _ V _ N _
 M _ V _ S _ V _ P _ G _ L _ A _ Q _ D _ V _ A _ W _ *
 M _ L _ L _ G _ E _
 ATCCTTGCCAGCACAGCTCAGGCCGCTTGGCGCGATGTATGTTTACCTGAGCTATCCCTGTACTATCCC
 I _ L _ A _ S _ T _ A _ Q _ A _ A _ L _ A _ R _ C _ M _ F _ T _ *
 H _ P _ C _ Q _ H _ S _ S _ G _ R _ L _ G _ A _ M _ Y _ V _ Y _ L _ S _ Y _ P _ C _ T _ I _ P _
 TGTATGATCGGCTCCATCGTTCGGCGTTACTAATCTGGAGATGCCCCGATCACTATCCGCCGGTACCGCTT
 V _ *
 M _ I _ G _ S _ I _ V _ G _ V _ T _ N _ L _ E _ M _ P _ R _ S _ L _ S _ A _ G _ T _ A _
 CTTCTTACTTACTAGCAGGTCTAGTTCCATGACTAGGTCTGAGAGGATATCGGGGTCTGATCCTCCATCTC
 S _ S _ L _ L _ S _ R _ S _ S _ S _ M _ T _ R _ S _ *
 GATGACGGTTCGCATCGAGATCCATATCATCCGCAAGAACCCTCGCTGGA
 M _ T _ V _ A _ S _ R _ S _ I _ S _ S _ A _ R _ T _ S _ L _ G _ A _ *

Para el cluster de sORFs 2L_39, el programa Augustus Gene Prediction predijo una estructura de 6 exones, uno de 215 pares de bases, el segundo de 122 pares de bases, el tercero de 623 pares de bases, el cuarto de 215 pares de bases, el quinto de 122 pares de bases y el sexto de 895 pares de bases, y 5 intrones, uno de 55 pares de bases, el segundo de 135pares de bases, el tercero de 1970 pares de bases, el cuarto de 55 pares de bases y el quinto de 135 pares de bases.

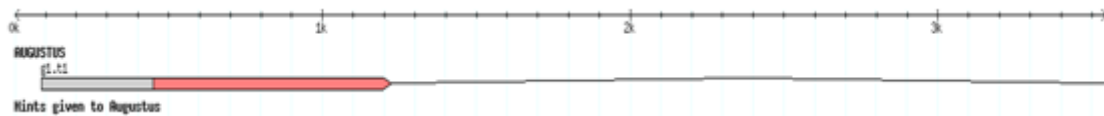
Se encontró acuerdo con espectro de masas para el péptido QSFSIPCRP con una probabilidad de $\rho=0.028$ y un valor $E=0.000572181997750584$; para el péptido LLCPGTHLIVGLLTAPSLSR con una probabilidad de $\rho=0.058$ y un valor $E=1.70994459689005e-06$; para el péptido LLLACNLLAFLAAVSPAPECLEDR una probabilidad de $\rho=0.02$ y un valor $E=0.000683845203177432$; para el péptido GEVPHDLPAMFLHVK una probabilidad de $\rho=0.086$ y un valor $E=0.000419547838544714$; para el péptido RQIVLEL una probabilidad de $\rho=0.006$ y un valor $E=0.000419717048509459$; para el péptido TAYRPLLEKCLLLR una probabilidad de $\rho=0.044$ y un valor $E=0.000307576662455124$; para el péptido CLLLRGCVWHAIF una probabilidad de $\rho=0.056$ y un valor $E=0.000424692340298027$; y para el péptido MLLGEHPCQHSSGR una probabilidad de $\rho=0.098$ y un valor $E=0.000876828754197293$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 65 sORFs. Se pueden observar varias repeticiones. No se observa presencia de ESTs, con lo cual no se halla evidencia de

transcripción. La probable traducción se basa en la presencia en bases de datos de 8 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_63



TTCGGAATATGTATTTACATTC AATATTATTTATCAGCGTCAAATTAGCGTCGGTAAGTTAGCGTCGTTT
M_Y_L_H_S_I_L_F_I_S_V_K_L_A_S_V_S_*
TTTTACATCCGTCGTGTTCTAAACTGTGCTAGACAACAGCGAAGATACATACATGTATATGTCGGAGCGC
M_Y_M_S_E_R_
GACTCTTCGCATTCAGAAAATTCCTTCTGTGATTTTCATATTTCTGATGAAATTGTAATATTGCGGTAAAAT
D_S_S_H_S_R_N_S_F_C_D_F_I_F_*
TCTGTGAGCGCTGCGGCAGAGGCACGAACAGCCTCTGCAGCTGAGTGAGGTCCGATCGAGTGGTCATAA
ATAGTGTGTTAGAGAGAGATGACAATGTGGCACACGCCAGTGTGTATAGCCATTAGAGAATATGATGAAG
M_T_M_W_H_T_P_V_C_I_A_I_R_E_Y_D_E_
M_M_K_
AAGGGACATGTAAGAAGATCCCTTCAGTGAAGTTTGACTGCTGACGTCGATCGGAACTTGCTGCGCTGAC
E_G_T_C_K_K_I_P_S_V_K_F_D_C_*
K_G_H_V_R_R_S_L_Q_*
GTACAAAATCGCGAAGTGAATAAATAATATGGATGAGACTCCTGTTTTCGCCGACATCAGAAGTGGGATCG
M_R_L_L_F_R_R_H_Q_K_W_D_R_
CCTTCGCCACAAGAGACCAGACCTGATCAGTATATGGCTGCGTTGGAGGCACAAAACCGTAACCTCATGG
L_R_H_K_R_P_D_L_I_S_I_W_L_R_W_R_H_K_T_V_T_S_W_
AGATAATAAGAACCATGCACGCACCGAGAGCATCGGCGTCAAATGAAAAGTCCCTTGACGTCGCACATGCC
R_* M_K_S_P_C_T_S_H_C_
AAAATTCTGCGCTGACAGCGCTGGAGCAGATCCATCTGCATGGTGCACCACCGTGGATTTAATCTTTGCA
Q_N_S_A_L_T_A_L_E_Q_I_H_L_H_G_A_P_P_W_I_*
M_V_H_H_R_G_F_N_L_C_
GATAATGCGCTTGTAGGCAGTGCACCTCGTAATAGCGTTAAGCAAAGCATTAGAAGGCAGTGCATCGCAAT
R_* M
GGCTGTGCGAGATATGCTTCGCTGGCATCACGTGGCCGAGTTTAAAGAACTGTTTCATACAGCGATTTCGT
A_V_A_D_M_L_R_W_H_H_V_A_A_V_*
AGGAATGGAGACGTCGGCTGCTATTTTGATGAACGTTTTGAACGGACGTCCAACACCTGGAGAGAGCTTT
GCCCAGTATGGAAGTCGCATTGTCACCTTGTGCTGTCTAAGTGAAGGCCAAGGATTTGGAGGAGATTG
M_E_V_A_L_S_P_C_C_C_L_S_G_R_P_R_I_W_R_R_L_
CGGTTTTCGGTAGCGTTGGCTCACATGGCACAAATCGATAATAAGTTGTTGCTGTGGGTGTTTACGACTAA
R_F_R_*
TGTGGCTACGCGCAATGAGCTTCAGCAACAACCTGCAAGCTTACGCCTTCAAGAAACGTAACAACGAAGAT
M_W_L_R_A_M_S_F_S_N_N_C_K_L_T_P_S_R_N_V_T_T_K_M
GATTCTGGCCAGAAAAAAGTTAAGAATGCAGTTGCAGACCATGTGCCATTTTTGTGGGAAAGCTGGCC
I_L_A_Q_K_K_S_*
ACAAATTCGCCGAATGCCGTGCTCGAAAGGAAGGTACATCAAACACAAAAGGAAGAACTACAGCGAAAG
M_P_C_S_K_G_R_Y_I_K_H_K_R_K_K_L_Q_R_K_
CAACACGCCTGGGTTAAAAGATCGTTGGAACATAAAATGTTTTAAATGCGACGAAATGGGACATGTGGCG
Q_H_A_W_V_K_R_S_F_E_H_K_M_F_* M_R_R_N_G_T_C_G_
M_W_R_
TCTGTATGTCCCAAAGGCCACAACAAATACATCGAAAAGCGAGTTGACGTGTGCGAAACTAAGTGAGCCA
V_C_M_S_Q_R_P_Q_Q_I_H_R_K_A_S_*
L_Y_V_P_K_A_T_T_N_T_S_K_S_E_L_T_C_A_K_L_S_E_P_

AGTGGACTGTGTTTTTCAATTGGGTGAGCCCTTCCATTTTTTTTTTCGATTCGGACGCCGAATGTTCTTT
S_G_L_C_F_S_I_G_*
AATAAAGGATAAATTAAGTAACAAATTTGGCTGGCAAGAGAATTAATAACCTAGTTGTTTTAAGAGGAATT
GGTGACAATGTAGTTAAGAGCCACTTGCAAATACTATAAACAGTCCAAATGTCTCAATTTGAAATTTAGAAA
TTTTGTTTTCATGTTGTAGTGACAACTATTTAAAGTACGACATAATGATTGGCCAAGACATATTAGAGCT
TGGTTTCGTTTTATTGATTAATCTAATAAATTTAGCATGTACAAATCCAAAGTTATCAATGCCGTTCAA
M_P_F_K
AATAATGCACATGAGAATAGTTTGGAAAACTTTTTGACATCAATACAGAAGTAAGCGATAATGATAAAT
I_M_H_M_R_I_V_W_K_N_F_L_T_S_I_Q_K_*
GTAAATTAAGAAATATATTGAAAAGTATGCCGATAGTTTCGTAACGGGATACCAAATAAGGCGCATTG
GAAATTCGTTTTAAGTAATAAAAAATAAAAGTGTTCAGGGCGCCCATATCGACTCAGTTTCCATTGCTACT
GATTTTGGATCAAAATAGATCGTTTGCAGGTGTGGTAAATATTTTTTTCATTTGTTTTAGACATGGCCAGCG
M_A_S
GATTCTATAGAAGAGTCAAGTATGTCTTTCCTTAGAGCTCCCAGTCGCATCATAGTGGATCAAGGCAGAA
G_F_Y_R_R_V_K_Y_V_F_P_*
GCTTTGTCAGCAACAAGTTCCGTGAGCTTTGTTCAACAAACAGGATTGAGTTGTTTTCAATAGCTACAGA
TGCCAGCAGAGCAAATGGGCAGGTTGAAAACAAATGAGCGTACAAGCAAGGAAAACATGCGAAAGGATGT
M_P_A_E_Q_M_G_R_L_K_T_N_E_R_T_S_K_E_N_M_R_K_D_V
M_S_V_Q_A_R_K_T_C_E_R_M
AGAGTAGTTCGGTATGAGGTCATGTGTTAAGTGAGAAGTATCATCGAGTGATTAATGTACTAGATGGTCA
E_*
*
TCGACACAATTTGAAGTCGTTAGTTATCAAGCGAACATATAAATATTCGCATGAATGTAATGGTGACTCC
M_N_V_M_V_T_P
CAACAGAGAAAAGAGTTGAAGGATGGCACTAGTGAACGTGCATCTGTGTCCCAAGAGAGAAAAGAGTTTA
N_R_E_K_S_* M_A_L_V_N_V_H_L_C_P_K_R_E_K_S_L
AAGATGGCACTAGTGCATCGTGCATCTGTGTCCCAAGAAAGGAAACATTTGGCGAACGTTATGAATAGAAT
K_M_A_L_V_I_V_H_L_C_P_K_K_G_N_I_W_R_T_L_*
GGCGTGTGACCGCACTATAACAGCTAACTCGCAGGGACGAGAAGTGTATGGCGGGGGTCAAGTGATGAT
GACTGCACTATAACAGCTAACTCGCAGGGACGAGAAGTGTATGGCGGGGGTCAAGTGATGATGACTGCA
CTATAACAGCTAACTCGCAGGGACGAGAAGTGTATGGCGGGGGTCAAGTGATGATGACTGCACTATAAC
AGCTAACTCGCAGGGACGAGAAGTGTATGGCGGGGGTCAAGAGATTGGGTGTGAGAGAAGACACGCCAA
M_A_G_V_K_R_L_G_V_R_E_D_T_P
TGTTATTATGAGAAATTAAGTCATGAAAAATGTAAATAGTTTGAAGTTTTGATATGTAAATTTGGAGATG
M_L_L_* M_E_N_V_N_S_L_K_F_*
TCTTTGTAAAAGAAAAATCAGTCATGAGATGAATTGTCAATTAATAATTACTGATTATTACTTGTGTGTC
ATTAATTGTTCTTAAGTTGACGAAAGTGTGTGACTTGGACTTGATTGGTGGATTAGGCACACGAGGACGT
GTGAAAAGGTCAGGAAGGCCGTGTCGGAGCGCGACTCTTCGCATTCCAGAAATTCCTTCTGTGATTTTCATA
TTCTGATGAAATTTGTAATATTGCGGTAATAATCTGTGTCGAGCGCTGCGGCAGAGGCACGAACAGCCTCTGC
AGCTGAGTGAGGTCCGATCGAGTGGTTCATAAATAGTGTGTTAGAGAGAGATGACAATGTGGCACACGTCA
M_T_M_W_H_T_S
GTGTGTATAGCCATTAGAGAATATGATGAAGAAGGGACATGTAAGAAGATCCCTTCAGTGAAGTTTGACT
V_C_I_A_I_R_E_Y_D_E_E_G_T_C_K_K_I_P_S_V_K_F_D
M_M_K_K_G_H_V_R_R_S_L_Q_*
GCTGACGTCGATCGGAACTTGTGCGCTGACGTACAAAATCGCGAAGTGAATAAATAATATGGATGAGAC
C_* M_D_E_T
M_R
TCCTGTTTTCGCCGACATATATAAGAAAAATTTGTCTTCTCAAAAATTTGTGAAAAACAAATTTGTAATCCAAC
P_V_S_P_T_Y_I_R_K_I_V_F_S_K_I_V_K_N_K_L_*
L_L_F_R_R_H_I_*
TTTTTACATAAATAAATAAATAAAGTAAAAATCTTTTAAGGTACCGTACAACCTGGAACAGGTTCTGTGA
GCGAAGTAATTAATATTTGTAGTTATCGTCGGAATCTTGTATAT

Para la región genómica 2L_63, el programa Augustus Gene Prediction predijo una estructura compuesta por un exón de 1137 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido EYDEEGTCKK con una probabilidad de $\rho=0.001$ y un valor $E=5.80894615997063e-05$; para el péptido MYLHSILFISVK con una probabilidad de $\rho=0.069$ y un valor

$E=0.000404532766085626$; para el péptido MALVIVHLCPK con una probabilidad de $\rho=0.031$ y un valor $E=0.000313099870333286$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 25 sORFs. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 3 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_75



ATTTTCATATTCTGATGAAATTGTAATATTGCGGTAAAAATTCGTGCGAGCGCTGCGGCAGAGGCACGAACA
 GCCTCTGCAGCTGAGTGAGGTCCGATCGAGTGGTCATAAAATAGTGTGTAGAGAGAGATGACAATGTGGC
 ACACGCCAGTGTGTATAGCCATTAGAGAATATGATGAAGAAGGGACATGTAAGAAGATCCCTTCAGTGAA
 H_T_P_V_C_I_A_I_R_E_Y_D_E_E_G_T_C_K_K_I_P_S_V_K
 M_M_K_K_G_H_V_R_R_S_L_Q
 GTTTGACTGCTGACGTCGATCGGAACTTGCTGCGCTGACGTACAAAATCGCGAAGTGAATAAATAATATG
 F_D_C_*
 *
 GATGAGACTCCTGTTTCGCCGACATCAGAAGTGGGATCGCCTTCGCCACAAGAGACCAGACCTGATCAGT
 M_R_L_L_F_R_R_H_Q_K_W_D_R_L_R_H_K_R_P_D_L_I_S
 ATATGGCTGCGTTGGAGGCACAAAACCGTAACCTCATGGAGATAATAAGAACCATGCACGCACCGAGAGC
 I_W_L_R_W_R_H_K_T_V_T_S_W_R_*
 ATCGGCGTCAAATGAAAAGTCCCTGACGTCGCACTGCCAAAATTCGTGCGCTGACAGCGCTGGAGCAGAT
 M_K_S_P_C_T_S_H_C_Q_N_S_A_L_T_A_L_E_Q_I
 CCATCTGCATGGTGACCACCGTGGATTTAATCTTTGCAGATAATGCGCTTGTAGGCAGTGCACCTCGTAA
 H_L_H_G_A_P_P_W_I_*
 M_V_H_H_R_G_F_N_L_C_R_*
 TAGCGTTAAGCAAAGCATTAGAAGGCAGTGCATCGCAATGGCTGTGCGAGATATGCTTCGCTGGCATCAC
 M_A_V_A_D_M_L_R_W_H_H
 GTGGCCGAGTTTAAAGAACTGTTTCATACAGCGATTTCGTAGGAATGGAGACGTGCGCTGCTATTTTGTATG
 V_A_A_V_*
 AACGTTTTGAACGGACGTCCAACACCTGGAGAGAGCTTTGCCAGTATGGAAGTCGCATTGTCACCTTGT
 M_E_V_A_L_S_P_C
 TGCTGTCTAAGTGAAGGCCAAGGATTTGGAGGAGATTGCGGTTTCGGTAGCGTTGGCTCACATGGCACA
 C_C_L_S_G_R_P_R_I_W_R_R_L_R_F_R_*
 AATCGATAATAAGTTGTTGCTGTGGGTGTTTACGACTAATGTGGCTACGCGCAATGAGCTTCAGCAACAA
 M_W_L_R_A_M_S_F_S_N_N
 CTGCAAGCTTACGCCTTCAAGAAACGTAACAACGAAGATGATTCTGGCCCAGAAAAAAGTTAAGAATGC
 C_K_L_T_P_S_R_N_V_T_T_K_M_I_L_A_Q_K_K_S_*
 AGTTGCAGACCATGTGCCATTTTTGTGGGAAAGCTGGCCACAAATTCGCCGAATGCCGTGCTCGAAAGGA
 M_P_C_S_K_G
 AGGTACATCAAACACAAAAGGAAGAACTACAGCGAAAGCAACACGCCTGGGTAAAAGATCGTTCGAAC

R Y I K H K R K K L Q R K Q H A W V K R S F E
 ATAAAATGTTTTAAATGCGACGAAATGGGACATGTGGCGTCTGTATGTCCCAAAGGCCACAACAAATACA
H K M F * M R R N G T C G V C M S Q R P Q Q I H
M W R L Y V P K A T T N T
 TCGAAAAGCGAGTTGACGTGTGCGAAACTAAGTGAGCCAAGTGGACTGTGTTTTCAATTGGGTGAGCCC
R K A S *
S K S E L T C A K L S E P S G L C F S I G *
 TTTCCATTTTTTTTCGATTCGGACGCCGAATGTTCTTTAATAAAGGATAAAATTAAGTAACAAATTTGGCT
 GGCAAGAGAATTAATAACCTAGTTGTTTTAAGAGGAATTTGGTGACAATGTAGTTAAGAGCCACTTGCAAA
 TACTATAAACAGTCCAAATGTCTCAATTGAATTTAGAAAATTTTGTTCATGTTGTAGTGGACAACATTTT
 AAAGTACGACATAATGATTTGGCCAAGACATATTAGAGCTTGGTTTTCGTTTTATTGATTAATCTAATAAA
 TTTAGCATGTACAAATCCAAAGTTATCAATGCCGTTCAAATAATGCACATGAGAATAGTTTGGAAAAAC
M P F K I M H M R I V W K N
 TTTTTGACATCAATACAGAAAGTAAGCGATAATGATAAATGTAAATTAAGAAATATATTGGAAAAGTATGC
F L T S I Q K *
 CGATAGTTTCGTAACCTGGGATACCAAATAAGGCGCATTGGAAAATTCGTTTTAAGTAATAAAAAATAAAGTG
 TTCAAGGGCGCCATATCGACTCAGTTTCCATTGCTACTGATTTTGGATCAAATAGATCGTTTGGCGTGT
 GGTAATATTTTTTTCATTTGTTTTAGACATGGCCAGCGGATTCATAGAAAGAGTCAAGTATGTCTTTCC
M A S G F Y R R V K Y V F P
 TTAGAGCTCCCAGTCGCATCATAGTGGATCAAGGCAGAAGCTTTGTGAGCAACAAGTTCCGTGAGCTTTG
*
 TTCAACAAACAGGATTGAGTTGTTTTCAATAGCTACAGATGCCAGCAGAGCAAATGGGCAGGTTGAAAAC
M P A E Q M G R L K T
 AAATGAGCGTACAAGCAAGGAAAACATGCGAAAGGATGTAGAGTAGTTCCGGTATGAGGTCATGTGTTAAG
N E R T S K E N M R K D V E *
M S V Q A R K T C E R M *
 TGAGAAGTATCATCGAGTGATTAATGTACTAGATGGTCAATCGACACAATTTGAAGTCGTTAGTTATCAAG
 CGAACATATAAATATTCGCATGAATGTAATGGTGACTCCCAACAGAGAAAAGAGTTGAAGGATGGCACTA
M A L
M N V M V T P N R E K S *
 GTGAACGTGCATCTGTGTCCCAAGAGAGAAAAGAGTTTAAAGATGGCACTAGTGATCGTGCATCTGTGTC
V N V H L C P K R E K S L K M A L V I V H L C
 CCAAGAAAAGAAACATTTGGCGAACGTTATGAATAGAATGGCGTGTGACCGCACTATAACAGCTAACTCG
P K K G N I W R T L *
 CAGGGACGAGAAGTGTTATGGCGGGGGTCAAGTGATGATGACTGCACATAACAGCTAACTCGCAGGGAC
 GAGAAGTGTTATGGCGGGGGTCAAGTGATGATGACTGCACATAACAGCTAACTCGCAGGGACGAGAAGT
 GTTATGGCGGGGGTCAAGTGATGATGACTGCACATAACAGCTAACTCGCAGGGACGAGAAGTGTTATGG
M
 CGGGGGTCAAGAGATTGGGTGTGAGAGAAGACACGCCAATGTTATTATGAGAAATTAAGTCATGGAAAA
A G V K R L G V R E D T P M L L * M E N
 TGAAAATAGTTTGAAGTTTGTATATGTAATTTGGAGATGTCTTTGTAAAAGAAAAATCAGTCATGAGATG
V N S L K F *
 AATTGTCAATTAATAATTAAGTATTACTTGTGTGTCATTAATTTGTTCTTAAGTTGACGAAGTTGTGT
 GACTTGACTTGATTTGGTGGATTAGGCACACGAGGACGTGTGAAAGGTCAGGAAGGCCGTGTCGGAGCGC
 GACTCTTCGCATTCAGAAATTCCTTCTGTGATTTTCATATTCTGATGAAATTGTAATATTGCGGTAATAAT
 TCTGTGCGAGCGCTGCGGCAGAGGCACGAACAGCCTCTGCAGCTGAGTGAGGTCCGATCGAGTGGTCATAA
 ATAGTGTGTTAGAGAGAGATGACAATGTGGCACACGTCAGTGTGTATAGCCATTAGAGAATATGATGAAG
M T M W H T S V C I A I R E Y D E
 AAGGGACATGTAAGAAGATCCCTTCAGTGAAGTTTACTGCTGACGTCGATCGGAACCTTGTGCGCTGAC
E G T C K K I P S V K F D C *
 GTACAAAATCGCGAAGTGAATAAATAATATGGATGAGACTCCTGTTTTCGCCGACATATATAAGAAAAATT
M D E T P V S P T Y I R K I
M R L L F R R H I *
 GTCTTCTCAAAAATTTGTGAAAAACAAATTTGTAATCCAATTTTTACATAAATAAATAAATAAAGTAAAAA
V F S K I V K N K L *
 TCTTTTAAAGGTACCGTACAACCTTGGAACAGGTTCTGTGAGCGAAGTAATTAATATTTGTAGTTATCGTCG
 GAATCTTGTATATGTGCAAGAGTCATCAGGACACTTTTTTCTTCTTCACTATTCTCGTTGTTGCATAAC
M S K S H Q D T L F P L H Y S R C C I T
 GTTTTCTGGCGCGTGAAGGAGACATGTCTGTTCTGCTCAACGATCGATCACTAACTGACCCAATCTCTCT
F S G A * M S V L L N D R S L T D P I S L
 TC

__C__*

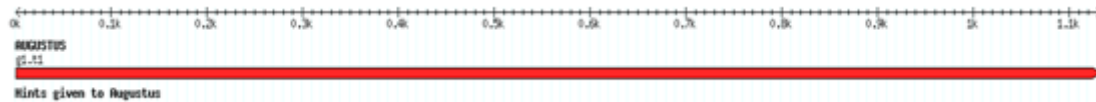
Para el gen 2L_75, el programa Augustus Gene Prediction predijo una estructura compuesta por 1 exón de 976 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido EYDEEGTCKK con una probabilidad de $\rho=0.001$ y un valor $E=5.80894615997063e-05$; para el péptido MALVIVHLCPK con una probabilidad de $\rho=0.031$ y un valor $E=0.000313099870333286$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 24 sORFs. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 2 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_96



```

TAATTTTGATTGCCAGCGGGCTCTATTCCGCCTCTTGGATCACCAAATCAAATTGGACAACCACCAGGA
TCACTGTTGCCTCCTAATGTGGGAATATATCCCCATCCACGGGCTGGATTCCGCCATCTGGGCCACTAA
      M_W_E_Y_I_P_H_P_R_A_G_F_R_H_L_G_H_*
CTCAAGGTGGACATCCAACGGGATCACTGTTGCCTCCTAATACTGGATTGCCACCGGGTTCCATTCCGCC
TCTTCGACCACCAAATCAAGGTGGACATCCACCAGGATCACAGAAGCCTCCTAATGTGGGAATATACCCC
      M_W_E_Y_T_P
CCATCCACGGGCTGGATTCCGCCATTTGGGCCACTAACTCAAGGTGGACATCCACCAGGATCACTGTTGC
_H_P_R_A_G_F_R_H_L_G_H_*
CTCCTAATACTGGATTGCCACCGGGCTCTATTCCGCCGCTTGGATCACCGAATCAAATTGGACATCCACC
AGGATCACAGAAGCCTCCTAATGTGGGAATATACCCCCATCCACGGGCTGGACTCCGCCATCTGGGCCA
      M_W_E_Y_T_P_H_P_R_A_G_L_R_H_L_G_H
CTAACTCAAGGTGGACATCCACCAGGATCACTGTTGCCTCCTAATACTGGATTGCCACCGGGTTCCATTC
_*
CGCCGCTTGGATCACCGAATCAAATTGGACATCCACCAGGATCACAGAAGCCTCCTAATACTGGATTGCC
ACCGGGTTCCATTCCGCCGCTTGGATCACCGAATCAAATTGGACATCCACCAGGATCACAGAAGCCTCCT
AATGTGGGAATATACCCCCATCCACGGGCTGGATTCCGCCATCTGGGCCACTAACTCAAAGTGGACATC
      M_W_E_Y_T_P_H_P_R_A_G_F_R_H_L_G_H_*
CACCAGGATCACTGTTGCCTCCTAATACTGGATTGCCACCGGGCTCTATTCCGCCTCTTCGACCACCAA
TCAAGGTGGACATCCACCAGGATCACAGAAGCCTCCTAATGTGGGAATATACCCCCATCCACGGGCTGG
      M_W_E_Y_T_P_H_P_R_A_G
ATTCCGCCATCTGGGCCACTAACTCAAGGTGGACATCCACCAGGATCACTGTTGCCTCCTAATACTGGAT
_F_R_H_L_G_H_*
TGCCATCGGGCTCTATTCCGCCGCTTGGATCACCGAATCAAATTGGACATCCACCAGGATCACAGAAGCC
TCCTAATCTGGGAATATATCCCCATCCACGGGCTCGATTCCGCCATCTGGGCCACTAACTCAAGGTGGA
    
```

CATCCACCAGGATCACTGTTGCCTCCTAATACTGGATTGCCATCGGGCTCTATTCCGCCGCTTGGATCAC
CGAATCAAATTGGACATCCACCAGGA

Para la región genómica 2L_96, el programa Augustus Gene Prediction predijo una estructura compuesta por 1 exón de 1132 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido WEYTPHPRAGFR con una probabilidad de $p=0.053$ y un valor $E=0.0008127839200649$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 5 sORFs. Se pueden observar repeticiones en tándem y desplazadas. Los sORFs cuarto y quinto son duplicaciones idénticas del segundo sORF. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_105



ACGATGGAGCCGATCATA CAGGGATAGTACAGGGATAGCTCAGGTAAACATACATCGCGCCAAGGCGGCC
TGAGCTGTGCTGGCAAGGATGTT CACCAAGCAACATCTTGGGCTAGCCCTGGTACAGAAACCATGGTATA
ACCAGGGCATACTGGACTGTACGTGAAGAACGCCTGGGTAATTTGGGATCAATGGGCTTCTAGACCTAG
AGCCTGTATCCTGGCATGTTGTAGTATTAATTATTATATACTTACAGAATTTCTCACGCGGGACTGTGTA
CCGATCGTGGTAGAAGATGTGGGAACAGCCAAGGAGACTGTGGTGGCATCGGCATACTTTGACGGTGACG
AGGCGTGCCACCGCCGAAATCACTGCCTGGCTGAACACTGCAAGAGGATGCACCTCAACCATCATCAT
TGGATGTGATGCAAATGCACATCACGTAATATGGGGCACCAGGCGTCGCCAACGAACGCAGCATGCCTAC

M _ G _ H _ Q _ A _ S _ P _ T _ N _ A _ A _ C _ L _
M _ P _ T

ATGGCGGGAAGATCAACTGATACTGCCCTATATCACCTGAAAAGCCTTATTGAGGATTCACCTGACTCATA

H _ G _ G _ K _ I _ N _ *

W _ R _ E _ D _ Q _ L _ I _ L _ P _ Y _ I _ T _ *

AGGAAGTGGCGTTATGTGCCTTCTTAGACATACAGGGTGCATTCGACAATACCTCCCATGAAGCTGTCAA

M _ C _ L _ L _ R _ H _ T _ G _ C _ I _ R _ Q _ Y _ L _ P _ *

M _ K _ L _ S _

CGCATCCCTGGCAAGAAGAGGACTAGATGCAACAACCAGCAGATGGATCAAGTCACTACTGGCATCTAAG

T _ H _ P _ W _ Q _ E _ E _ D _ * M _ Q _ Q _ P _ A _ D _ G _ S _ S _ H _ Y _ W _ H _ L _ S

CGAGCCATGGCCACTATAGGAGGGCAGTCCGTCACAGTACCTACCACAAGAGGTTGCTCCCAAGGGGGTG

E _ P _ W _ P _ L _ *

TCCTCTCGCTCTCCTGTGGAGCTTGTGGTACGCGAGTTGCTGGACAGACTGACCCGCAGAGGTATACT

TTGCCAAGGCTATGCTGACGATATTGTCATTATAGCCAGAGGTAATATGAGGAAACACTCTGCGATATC

M _ R _ K _ H _ S _ A _ I _ S

ATTCAACTGGGCATCAATATGACCAGCGAATGGTTCAAGGAAGTAGGCCTGAGCCTAAATCCTAGCAGAG

F _ N _ W _ A _ S _ I _ *

M _ V _ Q _ G _ S _ R _ P _ E _ P _ K _ S _ *

AATGAAGGCAATAGCCCTGTCAAATGTGGCATAGAGGCCAGCAATGAAGTTAAGTATCTTGGGATAGCC
CTTTACTCCAAACTCAGCTTCAAACCTCATGTGATAACAAAATTGATAAGTGCACCAGAGCACTTTTCA
M S I T K L I S A P E H F S
CGTGTAGAAAACATTGCTGGTAAGTCGTGGGGAACCTCACCACGCATAATAAGATGGCTGTACCTCATGGT
R V E T L L V S R G E P H H A * M A V P H G
AGTCAGACCCATACTAACCTATGGGGCAAAGCCTAGGGTGATAGAGCACGCCCTAGCACCCTAAGGCGA
S Q T H T N L W G K A *
AACTCCATAAGCTTCAAGTGCCTACCTGTCCCACTGCAACCCTAGAAGTACTAGTGGAGGTGACGCCAC
TTCATATCGTCATTGAAATGAAGAGGAAAGCCACCCTAATAAGAATTGAGGGAGCAGGAAATGACTGCAA
M T A
CCTCACAAGTAAGGATGCTGAAAGCCTAAAAAGGGATATCCCTTAGCCAAAGCAACCAAGGGATGCCAGC
T S Q V R M L K A *
TGAGCATAGGTTTCGCTCAGAACTTCACCACTCATCTCAGTAATAAGAACAGTTGGACATCCCTGGGGAAA
GTACACCCTATGAAACCACAAGCAATAAAGTGGTATACAGACGGATCCCTCACCAGCAGGGAAAGTGGGC
TGGGCGTTGTAGGCCCCACGTTGAAATACTACGAATCAATGGGCAGATACACCAGCATTTTTCAAGCTGA
AGTCTGTGCTATTGTACGCTGTGCGGAGTTAATCTGCAAAGAACTATCGTGGCAGGGACATTGCTATA
CTGTCTGATAGTCAAGCAGCCATAAAGCGCTCAGCAAAGCTAAGATAACATCTAAGGTAGTAAGTAAAG
TGAGGACAGCCCTAGACAAACTAGGAGCTGTCAACAAACCACAATAAGGTGGTCCCGGGACACAACAA
CATCCCGGGAAATGAGCTAGCGGACAACCTAGCCAGGAAAGAGGCAAAGAAACCCTCTAATTGGGCCCCGA
GCCCTTCTGTCTGTGGTACCATAGAGAAAAACGTCTGTCTTCCGGGAAAACCTACCAGGGCTTAGG
CAGTCTAAGATTCTTCTCCGTGAATATAACCATGAAAAGTTCAAGACCGTAATGACACACGGGAAAAAGA
CCGTGCGCATTTTACTGCGCTTCTTACTGCGGACTTCACAGCCATTTGCACAAGATTGGCATTGCGGAC
AGTGAACCTCTGTGCTTCTGGTACATGGAGGACACATCTGCACATATTACTTTTCATCAGGAGGAGCAGA
M E D T S A H I T F H Q E E Q
CTCCTGGGCGTGTATGTAGTCCCACGGGAAACAATTGCAGCCCTGAACCCCAATAAAAATTCTGGCATTTA
T P G R V C S P T G N N C S P E P Q *
TACAGTGCATTGAGCTTCCGGGGATCTTTGAGTCATGAAGGTGTAGTACAATAGATCCGGCTGGGTGCG
AGTACACATAACAAATGTGTACACATGTGGTGGCGATGCCCGGCTGGCTTACACGCAGTCGAGGTGACTG
M C T H V V A M P G V A Y T Q S R *
M W W R C P A W L T R S R G D *
ACCTTCGGCCCAAAATCGGAAAGGTGACATGAATGATGCCGCACTAGAAAGGAAGGAAAGGCGGGCAA
M N D A A L E R K G K A G K
GTAAGCGTTCACACGAGCCCCACGCAAAGTCATGGCATATATAATTCGTTTTAAACCAATGCTATT
V S V P H A A P T Q S H G I Y N S F *
M A Y I I R F K T N A I
M L L
GCAAAAGGATAGAGATGTACCGGTCTCTAAGCATTGAGGTGCTGGACTGACGGCGGCAAGGAATGCTAA
A K G *
Q K D R D V P V L *
M Y R S S K H S G A G L T A A R N A K
M L
AAGATTGCATGCCAGTAGCAACCTCTTAGTAGAAGACATTTTTTCGAGTAGGGGGCGGTACGCCGTTTTGG
R L H A S S N L L V E D I F R V G G G T P F W
K D C M P V A T S *
GCATAGTCCCTCCAAGTTCTTATGACTTTACATAAATGTTGCAAAATGTGTGAGGTGCTTCGTGACCAGG
A * M T L H K C S K M C E V L R D Q
CCAATCACTTCGTACAGCTCAAGGACGATTTGACGCTTCATCCATTGAAGCAGACTGGCGACTGGAGTT
A N H F V T A Q G R F D A S S I E A D W R L E L
GACTACACAGAAGCTCTAGAGCTCAAATCGGCATGCTTGCGAAGAACAAGCTTTTACAAAGGATATATAG
T T Q K L * M L A K N K L L Q R I Y S
CCGTGTTTATTAGTCTTGCAACAAAGGCAGAACACTGAGAACTGTGACAGGACTCACAACAGAGCAAAT
R V H *
ATGTGGGCGCGGACAATTTTATGAAGGCGTCTTCAGAAAGCTTCAGGGGGATTACGAGAACCCTCATAGA
M K A F F R K L Q G D Y E N L I D
TCCAAGCTTAATTCGAATTTCTCCTCAGTCGCCCCAAGCTGTGGCCCGCTGTTGGAGGGTAATGTTAAGTAG
P S L I R I L L S R P S C G P L L E G N V K *
GTAAAACATCCATTCAGGGAGTCACTGCAAAATCGACGGCTCACGTGATTTGGCAAGGATGTTCCCAAG
CAACATCTTGGGCTAGCCCTGGTACAGAAACCATGGTATAACCAGGGCATAACGTGGACTGTACGTGAAGA
ACGCCTGGGTAATTTGGGATCAATGGGCTTCTAGACCTAGAGCCTGTATCCTGGCATGTTGTAGTATTAA
TTATTATATACTTACAGAATTTCTCACGCGGACTGTGTACCGATCGTGGTAGAAGATGTGGGAACAGCC
AAGGAGACTGTGGTGGCATCGGCATACTTTGACGGTGACGAGGCGTCCCACCAGCCGAAATCACTGCAC

TGGCTGAACACTGCAAGAGGATGCACTCAACCATCATCATTGGATGTGATGCAAATGCACATCACGTAAT
ATGGGGCACCAGGCGTCGCCAACGAACGCAGCATGCCTACATGGCGGGAAGATCAACTGATACTGCCCTA
M_G_H_Q_A_S_P_T_N_A_A_C_L_H_G_G_K_I_N_*
M_P_T_W_R_E_D_Q_L_I_L_P_Y
TATCACCTGAAAAGCCTTATTGAGGATTCCTGACTCATAAGGAAGTGGCGTTATGTGCCTTCTTAGACA
I_T_* M_C_L_L_R_H
TACAGGGTGCATTCGACAATACCTCCCATGAAGCTGTCAACGCATCCCTGGCAAGAAGAGGACTAGATGC
T_G_C_I_R_Q_Y_L_P_*
M_K_L_S_T_H_P_W_Q_E_E_D_* M_
AACAAACCAGCAGATGGATCAAGTCACTACTGGCATCTAAGCGAGCCATGGCCACTATAGGAGGGCAGTCG
Q_Q_P_A_D_G_S_S_H_Y_W_H_L_S_E_P_W_P_L_*
TCCACAGTACCTACCACAAGAGGTTGCTCCCAAGGGGGTGTCTCTCGCCTCTCCTGTGGAGCTTGTCTGG
TACGCGAGTTGTCTGGACAGACTGACCCGCAGAGGTATACTTTGCCAAGGCTATGCTGACGATATTGTCAT
TATAGCCAGAGGTAATAATGAGGAAACACTCTGCGATATCATTTCAACTGGGCATCAATATGACCAGCGAA
M_R_K_H_S_A_I_S_F_N_W_A_S_I_*
TGGTTCAAGGAAGTAGGCCTGAGCCTAAATCCTAGCAGAGAATGAAGGCAATAGCCCTGTCAAATGTGG
M_V_Q_G_S_R_P_E_P_K_S_*
CATAGAGGCCAGCAATGAAGTTAAGTATCTTGGGATAGCCCTTTACTCCAAACTCAGCTTCAAACACTCAT
M
GTCGATAACAAAATTGATAAGTGCACCAGAGCACTTTTTCACGTGTAGAAACATTGCTGGTAAGTCGTGGG
S_I_T_K_L_I_S_A_P_E_H_F_S_R_V_E_T_L_L_V_S_R_G
GAACCTCACCACGCATAATAAGATGGCTGTACCTCATGGTAGTCAGACCCATACTAACCTATGGGGCAA
E_P_H_H_A_* M_A_V_P_H_G_S_Q_T_H_T_N_L_W_G_K
GCCTAGGGTGTAGAGCACGCCCTAGCACCCTAAGGCCGAACTCCATAAGCTTCAGAGTGCCTACCTGT
A_*
CCCCTGTAGCCCTAGAAGTACTAGTGGAGGTGACGCCGCTTCATATCGTCACTGAAATGAAGCGGAAAG
CCACCCTAATAAGAATTGAGGGAGCAGGAAATGACTGCAACCTCACAAGTAAGGATGCTGAAAGCCTAAA
M_T_A_T_S_Q_V_R_M_L_K_A_*
AAGGATATCCCTTAGCCAAAGCAACCAAGGGATGCCAGCTGAGCATAGGTTTCGCTCAGAACTTACCAC
M_P_A_E_H_R_F_A_Q_N_F_T_T
TCATCTCAGTAATAAGAACAGTTGGGCGTCTAAGATTCTTCTCCGTGAATATAACCATGAAAAGTTCAA
H_L_S_N_K_N_S_W_A_V_*
GACCGTAATGACACACGGGAAAAAGACCGTGCGCATTTTGTACTGGCCTTCTTACTGCCGACTTCACAGCC
ATTTGCACAAGATTGGCATTGCGGACAGTGAACCTCTGTGCTTCTGGTACATGGAGGACACATCTGCACA
M_E_D_T_S_A_H
TATTACTTTTTCATCAGGAGGAGCAGACTCCTGGGCGTGTATGTAGTCCCACGGGAAACAATTGCAGCCCT
I_T_F_H_Q_E_E_Q_T_P_G_R_V_C_S_P_T_G_N_N_C_S_P
GAACCCCAATAAAAATTCTGGCATTATACAGTGCATTGAGCTTCCGGGGGATCTTTGAGTCATGAAGGTG
E_P_Q_*
TAGTACAATAGATCCGGCTGGGTCGCAGTACACATACAAATGTGTACACATGTGGTGGCGATGCCCGGCG
M_C_T_H_V_V_A_M_P_G
M_W_W_R_C_P_A
TGGCTTACACGCAGTCGAGGTGACTGACCTTCGGCCCAAATCGGAAAGGTCGACATGAATGATGCCGCAC
V_A_Y_T_Q_S_R_*
W_L_T_R_S_R_G_D_* M_N_D_A_A
TAGAAAAGGAAGGGAAAAGCGGGCAAAGTAAGCGTTCCCCACGCAGCCCCACGCAAAGTCATGGCATATA
L_E_R_K_G_K_A_G_K_V_S_V_P_H_A_A_P_T_Q_S_H_G_I_Y
M_A_Y
TAATTCGTTTTAAAACCAATGCTATTGCAAAGGATAGAGATGTACCGGTCCTCTAAGCATTACAGGTGCT
N_S_F_*
I_I_R_F_K_T_N_A_I_A_K_G_*
M_L_L_Q_K_D_R_D_V_P_V_L_*
M_Y_R_S_S_K_H_S_G_A
GGACTGACGGCGCAAGGAATGCTAAAAGATTGCATGCCAGTAGCAACCTCTTAGTAGAAGACATTTTTTC
G_L_T_A_A_R_N_A_K_R_L_H_A_S_S_N_L_L_V_E_D_I_F
M_L_K_D_C_M_P_V_A_T_S_*
GAGTAGGGGGCGGTACGCCGTTTTGGGCATAGTCCCTCCAAGTTCTTATGACTTTACATAAATGTTTCGAA
R_V_G_G_G_T_P_F_W_A_* M_T_L_H_K_C_S_K
AATGTGTGAGGTGCTTCGTGACCAGGCCAATCACTTCGTACAGCTCAAGGACGATTTGACGCTTCATCC
M_C_E_V_L_R_D_Q_A_N_H_F_V_T_A_Q_G_R_F_D_A_S_S
ATTGAAGCAGACTGGCCTGAGTTGACTACACAGAAGCTCTAGAGCTCAAATCGGCATGCTTGCGAAG

```

_I_E_A_D_W_R_L_E_L_T_T_Q_K_L_*           M_L_A_K
AACAAAGCTTTTACAAAGGATATATAGCCGTGTTTCATTAGTCTTGCAACAAAGGCAGAACACTGAGAAACT
_N_K_L_L_Q_R_I_Y_S_R_V_H_*
GTGACAGGACTCACAAACAGAGCAAATATGTGGGCGCGGACAATTTTATGAAGGCGTTCTTCAGAAAGCTT
                                     M_K_A_F_F_R_K_L
CAGGGGGATTACGAGAACCTCATAGATCCAAGCTTACTTCTCGAATTCTCCTCAGTCGCCCAATATCGCCCA
_Q_G_D_Y_E_N_L_I_D_P_S_L_L_R_I_L_L_S_R_P_I_S_P
AGCTGTGGCCCGCTGTTGGAGGGTAATGTTAAGTAGGTAAAACATCCATTCAAGGGAGTCACTGCAAATC
K_L_W_P_A_V_G_G_*
GACGGCTCACTTAT
    
```

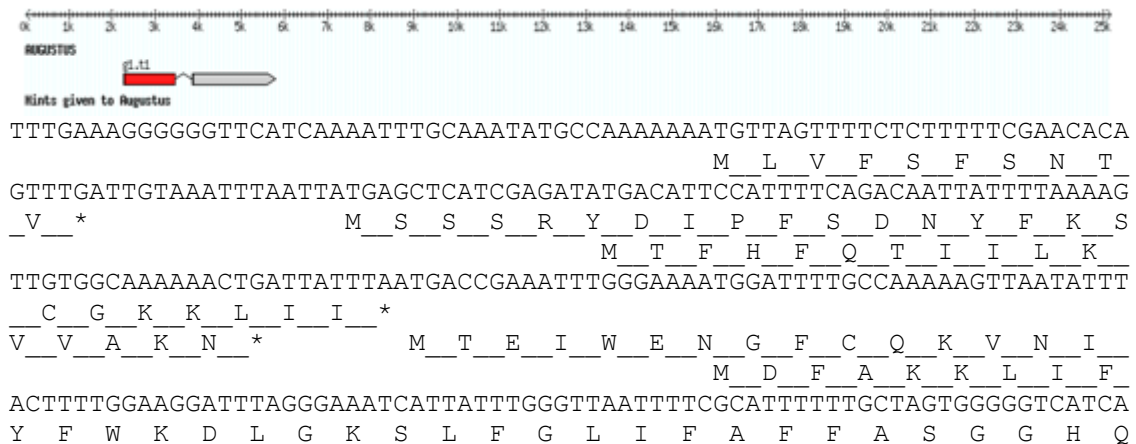
Para el gen 2L_105, el programa Augustus Gene Prediction predijouna estructura compuesta por 6 exones de 839 pares de bases el primero, 122 pares de bases el segundo, 215 pares de bases el tercero, 623 pares de bases el cuarto, 122 pares de bases el quinto, y 375 pares de bases el sexto, y 5 intrones de 135 pares de bases el primero, 55 pares de bases el segundo, 1970 pares de bases el tercero, 135 pares de bases el cuarto y 55 pares de bases el sexto.

Se encontró acuerdo con espectros de masas para el péptido FAQNFTTHLSNKN~~S~~WAV con una probabilidad de $\rho=0.062$ y un valor $E=0.000251395072294928$, y para el péptido DQANHFVTAQGRFDASSIEADWR, con una probabilidad de $\rho=0.027$ y un valor $E=0.000168306701192261$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 43 sORFs. Se pueden observar repeticiones. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 2 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_108



T F G R I *
 AAATTTGAAAAGATATGCCAAAAAATGTAATTTTCGTTGTTGAACACAGTTTCATTGGAAATTTAATTA
N L K D M P K K M *
 TGAGCTCATCGAGATATGACATTCCATTTTCAGACAATTATTTTAAAAGTTGTGGCAAAAACTGATTAT
M S S S R Y D I P F S D N Y F K S C G K K L I I
M T F H F Q T I I L K V V A K N *
 TTAATGACCGAAATTTGGAAAAACGGATTTTGCCAAAAAGTTAATATTTCCTTTTTGAAGGATTTAGGGA
* M T E I W K N G F C Q *
 AATCATTATTTGTGTACATTTTCGCATTTTTTGAAAAGGGGGTTCATCAAATTTGCAAATATGCCAAAA
M P K
 AAATGTTATTTTTTCTCTTTTTTCGAACACAGTTTGATTGTAAATTTAATTATGAGCTCATCGAGATATGA
K C Y F F S F S N T V *
M L F F L F F E H S L I V N L I M S S S R Y D
M
 CATTCCATTTTCAGACAATTATTTTAAAAGTTGTGGCAAAAACTGATTATTTAATGACCGAAATTTGGA
I P F S D N Y F K S C G K K L I I *
T F H F Q T I I L K V V A K N * M T E I W
 AAAACGGATTTTGCCAAAAAGTTAATATTTACTTTTTGGAAGGATTTAGGGAAATCATTATATGGGTTAAT
K N G F C Q K V N I Y F W K D L G K S L Y G L I
 TTTTCGCATTTTTTGTAGTGGGGTTCATCAAATTTGAAAAGATATGCCAAAAAATGTAATTTTCATTG
F A F F A S G G S S K F E R Y A K K M *
M P K K C N F H C
 TTGAACACAGTTTCATTGGAAATTTAATTACGAGCTCAGCGAGGTATGACATTTTCATATTCAGACAATTA
* M T F H I Q T I
 ATTTTAAAAGTTGTGGCAAAAAACTGATTATTTAATGACCGAAATTTAGGGAATCAGATTTTGCCAAAAAG
N F K V V A K K L I I * M T E I R E S D F A K K
 TTGTTATTTACTTTTTGAAGGATTTAGGGAAATCAATTTTTGCGTTAATTTTCGCATTTTTTGAAGGGG
L L F T F *
 GGTTCATTAATAATTTGCAAATATGCCGAAAAAATGTAATTTTCATTGTTGAACACAGTTTCATTGGAAAT
M P K K C N F H C *
 TTAATTACGAGCTCAGCGAGGTATGACATTTTCATATTCAGACAATTATTTTAAAAGTTGTGGCAAAAA
M T F H I Q T I I L K V V A K K
 CTGATTATTTAATGACCGAAATTTAGGGAATCGGATTTTGCCAAAAAGTTTTTATTTACTTTTTGAAGGAT
L I I * M T E I R E S D F A K K F L F T F *
 TTAGGGAAATCAATTTTTTGGGTTAATTTTCGCATTTTTTGAAGGGGGTTCATCAAATTTGCAAATAT
M
 GCCAAAAAATGTTATTTTTCTCTTTTTTGCACACAGTTTGATTGTAAATTTAATTATGAGCTCATCGAG
P K K C Y F S L F C T Q F D C K F N Y E L I E
M L F F S F L H T V * M S S S R
 ATATGACATTTCCATTTTCAGACAATTATTTTAAAAGTTGTGGCAAAAACTGATTATTTAATGACCGAAA
I *
Y D I P F S D N Y F K S C G K K L I I *
M T F H F Q T I I L K V V A K N * M T E
 TTTGAAAAATGATTTTGCCAAAAAGTTAATATTTACTTTTTGGAAGGATTTAGGGAAATCATTATTTGG
I W K N G F C Q *
M D F A K K L I F T F G R I *
 GTTAATTTTCGCATTTTTTGTAGTGGGGTTCATCAAATTTGAAAAGATATGCCAAAAAATGTAATTTT
M P K K C N F
 CATTGTTGAACACAGTTTCATTGGAAATTTAATTACGAGCTCAGCGAGGTATGACATTTCCATTTTCAGAC
H C * M T F H F Q T
 AATTATTTTTAAAAGTTGTGGCAAAAAACTGATTATATAATGACCGAAATTTAGGGATACGGATTTTGCCA
I I F K V V A K K L I I * M T E I R D T D F A
 AAAAGTTGTTATTTACTTTTTGAAGGATTTAGGAAAATCAATTTTTTGGGTTAATTTTCGCATTTTTTGA
K K L L F T F *
 AGGGGGTTCATCAAATTTGCAAATATGCTGTTGGAATATACTATTCAACCTACAAAAATAACGTTAAA
 CAACACTACCTTTTATATTTGATATGAATGGCCACACCTTTTATGCCATAAAACATATTTGTGAGAGAATAC
M N G H T F Y A I K H I V R E Y
 CACTCTTTTTATTCCTTCTTTTCTTCTTGTACGTTTTTGTGAGTAGGTGCTGGTGGTGGTGGTGGTGGCA
H S F Y S F F S S L Y V F A V S R S W C W C C
 GTTGAATAACTTAAAAATATAAATCATAAACTCAAACATAAACTTGACTATTTATTTATTTATTTAAGAA
S *

AGGAAATATAAAATTATAAAATTACAACAGGTTATGGGCCAGTCCATGCCTAATAAACAATTAATTTGTGA
M_P_N_K_Q_L_N_C_E
ATTAAGATTGTGAAAATAAAATTTGTGAAATAGCATTTCCTTGTGAAATAGCTTTTTTTTTTCA
L_K_I_V_K_I_N_C_E_I_A_F_F_H_I_L_V_K_*
CATTCTTGTGAAATTAATTTCTTCTCAGAAATTTGAGTGAAAAATGGACAAGGCATAACCGTAATATTAAGC
CGTTTGTATGGCGAGAAGTACGCGATTTGGAAATTTAGAATTAGGGCTCTTTTAGCCGAGCAAGATGTGCT
M_A_R_S_T_R_F_G_N_L_E_L_G_L_F_*
TAAAGTAGTTGATTTAATGCCTAACGAGGTAGATGACTCCTGGAAAAAGGCAGAGCCTTTGTGCAAAA
M_T_P_G_K_R_Q_S_V_V_Q_K
AGTACAATAATAGAGTACCTAAGCGACTCGTTTTTAAATTTGCAACAAGCGACATTACGGCGCGTCAGA
V_Q_*
TTCTTTGAGAATTTGGACGCCGTTTATGAACGAAAAAGTTTGGCGTCGCAACTGGCGCTGCGAAAACGTTT
M_N_E_K_V_W_R_R_N_W_R_C_E_N_V
GCTTTCTCTGAAGCTATCGAGTGAGATGTCACATTAAGCCATTTTCATATTTTGTGACGAACTTATAAGT
C_F_L_*
GAATTGTTGGCAGCTGGTGCAAAAATAGAAGAGATGGATAAAATTTCTCATCTACTGATCAGATTGCCTT
CGTGTTACGATGGAATTATTACAGCGATAGAGACATTATCTGAAGAAAATTTGACATTGGCGTTTGTGAA
AAATAGATTGCTGGATCAAGAAATTAATAAAATGACCACAACGATAACAAGCAAGAAAGTTATGAAC
M_T_T_T_I_Q_A_R_K_L_*
GCGATCGTGACAACAATAAACAACCTTATAAAAAATAATTTGTTTTAAAAATCGGGTAACATAACCAAAGA
AAATATTCAAGGGAAAATTTCAAAGTATAAAGTCAAGTGTCAACCTGTGGCAGAGAAGGCCACATTAATAAA
AGATTGTTTCCATTATAAAAAGAAATATTAATAAATAAAAAATAAGAAAATGAAAAACAAGTTCAAAGTCA
M_K_N_K_F_K_L_Q
ACATCACACGGCATTGCGTTTTATGGTAAAAAGAAGTGAATAATACTTCAGTGTGACAACCTGCGGGTTTG
H_H_T_A_L_R_L_W_*
TCCTTGTATCTGGTGTAGTACCATCTTATAAATGATGAGTGCCTGTATACCGACAGTGTGGAGTTGT
GCCTCCACTTAAGATTGCAGTGGCCAAGCAAGGCGAATTTATTTATGCCACTAAGCGTGGTATTGTCCGA
CTACGGAATGACCATGAGATTACACTGGAGGATGTACTCTTTTGTAAAGGAAGCTGCTGGTAATTTGATGT
CCGTAAAGCGTCTCCAAGAGGCAGGAATGTGATCGAATTTGACAAAAGCGGTGTAACCATTTTCGAAAA
AAATGGGTTAATGGTTGTCAAAAATTCAGGTATGTTAAACAATGTACCTGTGATCAATTTTCAAGCATAT
TCTATAAATGCTAAGCATAAAAAATAATTTTCGTTTTATGGCATGAGAGTTTGGCCATATAAGCAATGGCA
M_L_S_I_K_I_I_F_V_Y_G_M_R_G_L_A_I_*
AATTATTAGAAAATAAACGAAAAGAATATGTTTAGTGATCAAAGTCTTCTAAACAACCTTAGAGTTATCATG
TGAAATTTGTGAACCTGTTTAAATGGTAAACAGGCAAGACTTCCTTTTAAACAATTTGAAAGATAAGACC
M_V_N_R_Q_D_F_L_L_N_N_*
CATATTAAGAAGACCACTTTTTGTAGTACACTCAGATGTCTGTGGGCCTATTAATCCAGTTACTTTAGATG
M_S_V_G_L_L_I_Q_L_L_*
ATAAAAAATTATTTTGTGATCTTTGTTGATCAGTTTACACATTATTGTGTAACCTATTTAATTAATATAAA
ATCTGATGTGTTTAGCATGTTTTCAAGATTTTGTAGCCAAGAGTGAAGCTCATTTTAATTTAAAGGTTGTG
M_C_L_A_C_F_K_I_L_*
TACTTATACATTGACAATGGTAGAGAATACTTGTCAAATGAGATGAGACAATTTTGTGTTAAGAAAGGAA
M_V_E_N_T_C_Q_M_R_*
TTTCTTATCACTTAACAATGCCACATACACCTCAGTTAAATGGTGTTCCTGAGAGAATGATAAGAACCAT
TACGGAAAAAGCTCGAACCATGGTTAGTGGTGCAAAGCTAGATAAAAGCTTTTGGGGCGAAGCAGTATTA
ACTGCTACTTATTTAATCAACAGAATTCCTAGCAGAGCACTTGTGATAGTTCAAAGACCCCATATGAGA
M_R_
TGTGGCACAATAAGAAGCCATACTTAAAACATTTGAGAGTGTTCCTGGTCAACTGTTTTATGTGCATATTAA
C_G_T_I_R_S_H_T_* M_C_I_L
AAACAAACAAGGAAAAGTTTGTGATAAATCATTAAAGTATTTTTGTGGGCTATGAACCCAATGGTTTT
K_T_N_K_E_S_L_M_I_N_H_L_K_V_F_L_W_A_M_N_P_M_V_L
AAGTTGTGGGATGCTGTAATGAAAAATTTATTGTCGCAAGAGATGTTGTTGTCGATGAAACCAATATGG
S_C_G_M_L_* M_K_N_L_L_S_Q_E_M_L_L_S_M_K_P_I_W
TTAATTCTAGAGCTGTTAAATTTGAAACAGTGTTCCTGAAAGATAGTAAGGAAAGTGAATAAAAAATTT
L_I_L_E_L_L_N_L_K_Q_C_S_*
TCCGAATGACAGTAGGAAAATAATACAAACAGAGTTCCCGAATGAGAGTAAGGAATGCGACAACATACAA
M_R_V_R_N_A_T_T_Y_N
M_R_Q_H_T
TTCTGAAAGATAGTAAGGAAAAGTGAATAAAAAATTTTCCGAATGACAGTAGGAAAATAATACAAACAG
S_*
I_P_E_R_*
AATTCGGAATGAGAGTAAGGAATGCGACAACATACAATTCCTGAAAGATAGTAAGGAAAGTAATAAATA

M R V R N A T T Y N S *
M R Q H T I P E R *
 TTTTCTGAATGAGAGTAAGAAAAGAAAGCGAGATGATCACCTGAATGAAAGTAAGGGATCAGGCAACCCG
M R V R K E S E M I T * M K V R D Q A T R
 AATGAGAGTAGGGAAAAGTGAACAGCAGAGCACTTAAAAGAAATTGGAATTGATAATCCAATAAAAATG
M R V G K V K Q Q S T *
 ATGGCATAGAAATTATTAATAGAAAGTGAGAGATTAAGACTAAGCCTCAGATATCCTATAATGAAGA
 GGATAATAGTCTAAATAAAGTTGTTCTAAATGCTCACACTATATTTAACGATGTCCCAAATTCATTTGAT
M L T L Y L T M S Q I H L M
 GAAATTC AATATAGGGATGATAAATCTTCTGGGAAGAAGCCATCAATACAGAGTTAAATGCTCATAAAA
K F N I G M I N L L G K K P S I Q S * M L I K
 TTAATAATACTTTGGACAATTACAAAAAGGCCGTGAAAACAAAAATATTGTAGATAGCAGATGGGTATTTTC
L I I L G Q L Q K G L K T K I L *
 TGTTAAATATAATGAACTTGGAAATCCAATTAGATACAAAGCTAGATTGGTTGCACGAGGATTCCTCAA
M N L E I Q L D T K L D W L H E D S L K
 AAATACCAAATAGACTATGAAGAGACATTTGCTCCTATAGCTAGAATTTCAAGTTTCCGATTTATATTGT
N T K *
 CATTAGCAATACAGTATAACTTAAAAGTCCATCAAATGGATGGATGTAAAAACAGCTTTCTTAAATGGCA
 CGTTAAAAGAGGAAATTTATATGAGACTTCCCTCAAGGTATATCGTGTAAATAGTGACAATGTGTGTAAATT
 GAATAAGGCAATTTACGGACTCAAGCAAGCGGCTAGATGCTGGTTTGAAGTATTTGAGCTAGCATTGAAA
 GAGTGTGAGTTTGTAAACTCTTCAGTTGATCGCTGTATATATATTTTAGACAAAGGTAACATCAATGAAA
M K
 ACATATATGTATTATATATGTAGATGATGTGGTTATAGCTACAGGAGATATGACAAGAATGAATAACTT
T Y M Y Y Y M *
 CAAAAGGTATTTAATGGAAAAGTTTAGGATGACTGACCTAAATGAAATAAAACATTTTATTGGAATTAGG
 ATAGAGATGCATGAAGATAAAATCTATTTAAGCCAATCTGCATATGTTAAAAAATTTTAAGTAAATTTA
 ACATGGAAAATTGTAATGCAGTTAGTACTCTTTACCTAGTAAAATAAATTATGAATTACTTAATTCAGA
M Q L V L L Y L V K * M N Y L I Q
 TGAAGACTGCAATACCCCATGCCGTAGCCTCATAGGATGTTTAAATGTACATAATGCTTTGTACACGCCCA
M K T A I P H A V A S * M F N V H N A L Y T P
 GATTTAACTACTGCAGTAAATATCTTGAGCAGATATAGTAGCAAAAATAACTCCGAATTATGGCAGA
R F N Y C S K Y L E Q I *
 TAAAAAGAGTTCTTAGATATTTGAAGGGCACTATCGATATGAAATTGATTTTTTAAAAGAAGCTTGGCATT
 TGAAAATAAAATTTATGGTTATGTGGATTCTGATTGGGCTGGTAGTGAATTTGATAGAAAAGTACAACA
M W I L I G L V V K L I E K V Q Q
 GGGTATTTATTCAAATGTTTGAATTTAATCTCATTGTTGGAATACAAAGAGACAGAAGCTCAGTAGCAG
G I Y S K C L I L I S F V G I Q R D R T Q *
 CCTCATCAACTGAAGCTGAGTATATGGCCCTATTTGAAGCCGTGAGAGAAGCTCTATGGCTTAAATTTTT
 ATTAAGTATTAACATTAACCTAGAAAACCCATTAATTTTACGAAGACAATCAAGGCTGTATTAGC
 ATAGCAAAACCCCTCATGTCATAAACGAGCTAAACATATTGATATTAATATCATTGTCAGAGAGC
 AAGTTCAGAATAATGTGATTTGCTTGTAGTATATTCCTACAGAGAATCAACTGGCTGACATTTTACAAA
 ACCGTTGCCTGCTGCGAGATTTGTGGAGTTACGAGACAAATTTGGGTTTGTGCAAGACGACCAATCGAAT
 M
 GCTGAATGAAATTTTTATATATATTTTTCAAATTTAAATTCCTGTAAACATATTTTGTTACAATGATCTG
L N E I F I Y I F Q I *
M K F L Y I F F K F K F L *
 ATCGGGTTTTTCTGGGTTTTCCCGTATCCTCGCAGCAAATGCTGGATCAGTTAACACTTCCCAGAATGC
M L D Q L T L P R M
 ACACCACCACATTTGATAGTTACTAATGAATATTATTGTTATGTTTTTAAATTATAGACGTTATTTTTTGA
H T T H I * M N I I V M F L I I D V I F E
 GGGGGCGTGTGGAATATACTATTCAACCTACAAAAATAACGTTAAACAACACTACTTTATATTTGATAT
G A C W N I L F N L Q K * M
 GAATGGCCACACCTTTTATGCCATAAAACATATTGTGAGAGAATACCCTCTTTTTTATTCCTTCTTTTCT
N G H T F Y A I K H I V R E Y H S F Y S F F S
 TCCTTGTACGTTTTTGTGAGTAGGTCGTGGTGTGGTGTGTCAGTTGAAATAACTTAAAATATAAAT
S L Y V F A V S R S W C W C C S *
 CATAAACTCAAACATAAACTTACTATTTATTTATTTATTAAGAAAGGAAATATAAATTATAAATTACA
 ACATATGCCGAAAAAATGTAATTTTCATTTGTTGAACACAGTTTCATTTGAAATTTAATTACGAGCTCAGC
M P K K C N F H C *
 GAGGTATGACATTCATATTCAGACAATTAATTTTAAAGTTGTGGCAAAAAAAGCTGATTATTTAATGACC
M T F H I Q T I N F K V V A K K L I I *

CAAATTTGGAAAAACGGATTTTGCACAAAAGTTAATATTTCCTTTTGGAAGGATTTAGGAAAATCATTAT
 TTGGGTTAATTTTCGCATTTTTTGTCTAGTGGGGTTCATCAAAATTTGAAAGATATGCCAAAAAATGTGAT
 TTTTCATTGTTGAACACAGTTTCATTGGAAAATTTAATTACGAGCTCAGCGAGGTATGACATTCCATATTC
 F I V E H S F I G N L I T S S A R Y D I P Y S
 M P K N V I
 M T F H I P
 AACAAATATTTTTAAAGTTGTGGCAAAAAAAGTATTATTTAATGACCGAAATTTAGGGAATCAGATTTTG
 N N Y F *
 T I I F K V V A K K L I I * M T E I R E S D F
 CAAAAAGTTGTTATTTACTTTTTGAAGGATTTAGGGAATCAATTTTTTGCCTTAATTTTCGCATTTTTT
 A K K L L F T F *
 GAAAGGGGGTTCATTTAAATTTGCAAATATGCCGAAAAAATGTAATTTTCATTGTTGAACACAGTTTCA
 M P K K C N F H C *
 TTGGAAAATTTAATTACGAGCTCAGCGAGGTATGACATTTTCATATTCAGACAATTAATTTTAAAGTTGTGG
 M T F H I Q T I N F K V V
 CAAAAAAGTATTATTTAATGACCGAAATTTAGGATACGGATTTTGCACAAAAGTATTGTTATTTACTTTT
 A K K L I I * M T E I R D T D F A K N L L F T F
 TGAAGGATTTAGGAAAATCAATTTTTTGGGTTAATTTTCGCATTTTTTGAAGGGGGTTCATCAAAATTT
 *
 GAAAGATATGCCGAAAAAATGTGATTTTCATTGTTGAACACAGTTTCATTGGAAAATTTAATTACGAGCTC
 M P K K C D F H C *
 AGCGAGGTATGACATTTCCATATTTCCAACAATTTATTTTTAAAGTTGTGGCAAAAAAAGTATTATTTAATG
 M T F H I P T I I F K V V A K K L I I * M
 ACCGAAATTTAGGATACGGATTTTGCACAAAGTTGTTATTTACTTTTTGAAGGATTTAGGGAATCAAT
 T E I R D T D F A N K L L F T F *
 TTTTGGGTTAATTTTCGCATTTTTTGAAGGGGGTTCATCAAAATTTGCAAATATGCCAAAAAATTTA
 M P K K F
 ATTTTCATTGTTGAACACAGTTTCATTGGAAAATTTAATTACGAGCTCAGCGAGGTATGACATTTCCATATTC
 N F I V E H S F I G N L I T S S A R Y D I P Y S
 M T F H I
 AAACAATATTTTTAAAGTTGTGGCAAAAAAAGTATTATTT
 N N Y F *
 Q T I I F K V V A K K L I I *

Para el gen 2L_108, el programa Augustus Gene Prediction predijouna estructura compuesta por 2 exones de 1190 pares de bases el primero y el segundo de 1933 pares de bases, y 1 intrón de 407 pares de bases.

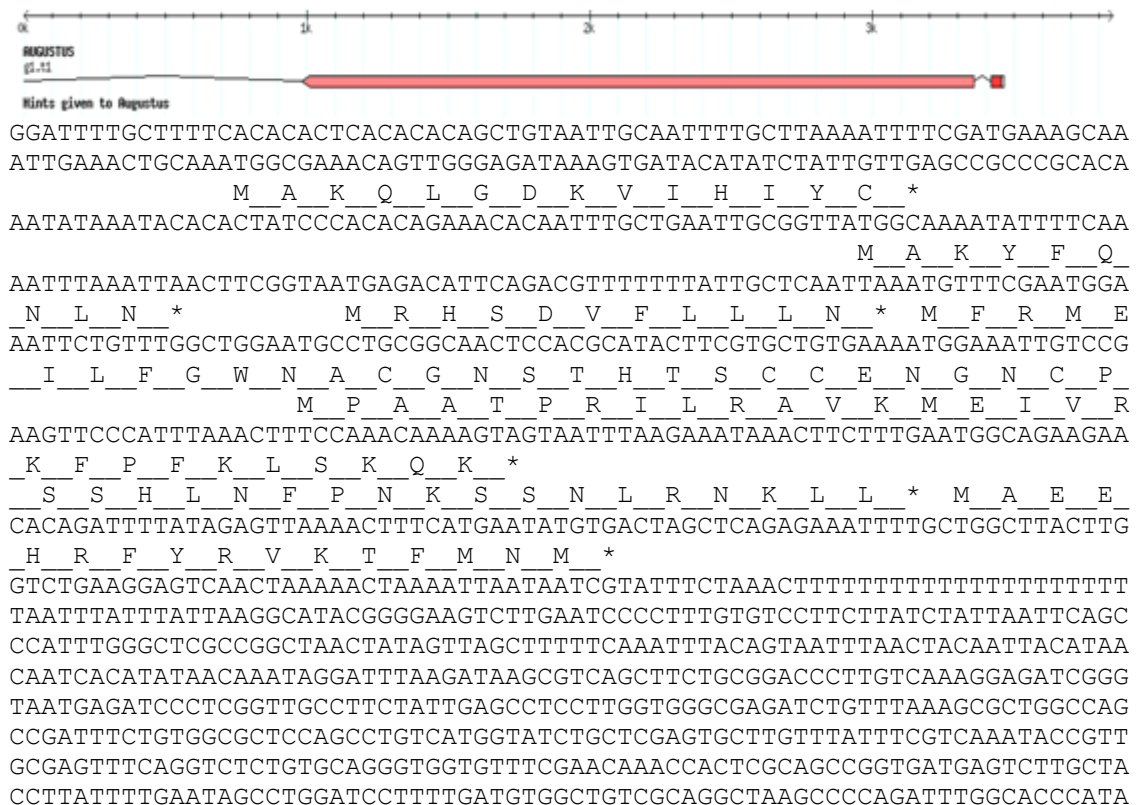
Se encontró acuerdo con espectros de masas para el péptido LFFSFLHTV con una probabilidad de $\rho=0.02$ y un valor $E=2.18235564894947e-05$; para el péptido MTFHFQTIIFK con una probabilidad de $\rho=0.048$ y un valor $E=0.000541665926594377$; para el péptido TFHIQTIIFKVVAK con una probabilidad de $\rho=0.077$ y un valor $E=1.74494226863423e-05$; para el péptido LFFSFLNTV con una probabilidad de $\rho=0.012$ y un valor $E=1.95068835446376e-05$; para el péptido IVKINCEIAFFHILVK con una probabilidad de $\rho=0.033$ y un valor $E=0.000103692423625058$; para el péptido KFNFIVEHSFIGNLITSSAR con una probabilidad de $\rho=0.007$ y un valor $E=0.000672127891608788$; para el péptido GVHQLNLI con una probabilidad de $\rho=0.083$ y un valor $E=0.000381138123940895$; para el péptido TLHFQTIIFKVVAK con una probabilidad de $\rho=0.077$ y un valor $E=0.000121002071622241$; para el péptido MTFHIQTINFK con una probabilidad de $\rho=0.088$ y un valor $E=4.77278654852789e-05$; para el péptido ESLMINHLKVFLWAMNPMVLSCGML con una probabilidad de $\rho=0.065$ y un valor $E=0.000261882533564455$; para el péptido VFLWAMNPMVLSCGML con una probabilidad de $\rho=0.091$ y un valor $E=0.000896475383408014$; para el péptido FNIGMINLLGK con una probabilidad de $\rho=0.044$ y un valor $E=0.00021280221052333$; para el péptido LIKLILGQLQK con una probabilidad de

$\rho=0.051$ y un valor $E=9.11628258769806e-05$; para el péptido VQGGIYSKCLILISFVGIQR con una probabilidad de $\rho=0.012$ y un valor $E=3.55732257113095e-05$; para el péptido MWILIGLVVK con una probabilidad de $\rho=0.01$ y un valor $E=1.93761506518799e-05$; para el péptido NIIVMFLIIDVIFEGACWNILFNLQK con una probabilidad de $\rho=0.065$ y un valor $E=0.000232678886555133$; para el péptido DLGKSIFALIFAFFER con una probabilidad de $\rho=0.099$ y un valor $E=0.000145335264384092$; para el péptido VVIYFLKDLGK con una probabilidad de $\rho=0.064$ y un valor $E=0.000143336854414044$; para el péptido DLGKSLFGLIFAFFVSGGHQNLK con una probabilidad de $\rho=0.047$ y un valor $E=0.000584690172889705$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 73 sORFs. Se pueden observar repeticiones. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 19 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_120



CTTCCAGTTTGGTGCTAGGATCTGTTTGTAAATTGTCAGCTTGTGGATAGCGACAATTTACTGCGCGAA
M_W_L_S_Q_A_K_P_Q_I_W_H_P_Y
_F_Q_F_G_A_R_I_C_L_*
GAAAGTAGCCAGTAGTGTCTCGCCACCTTAGCACGTAGGCGCGTTCTGATGTCGGTCACATGCTTGGAAA
M_S_V_T_C_L_E
M_L_G_K
ATGTGAGTCTGCGATCCAGAAGCACTCCAAGGTACTTTGCTGCGTTCCGGCTGTGGTACGGGGACTTCCTC
N_V_S_L_R_S_R_S_T_P_R_Y_F_A_A_F_G_C_G_T_G_T_S_S
_C_E_S_A_I_Q_K_H_S_K_V_L_C_C_V_R_L_W_Y_G_D_F_L
GATGTAGACGGGCGGTGGCGTTCTCCGCTTTAAAGTGAAGCAGACGTTGTGGATTTACTGCTATTGATG
M_*
_D_V_D_G_R_W_R_S_P_L_* M
CCAATATTCCATCGTCTTGGCCATTCCGAGAACCAGGTATGCAAAGTCCTGGATACCATCGGCTGCGTCGT
P_I_F_H_R_L_A_H_S_E_N_R_Y_A_K_S_W_I_P_S_A_A_S
M_Q_S_P_G_Y_H_R_L_R_R
GCTCGCATCGGGACCTGTAGGTGACGCACACGTCATCGGCAAATGTGGCCAACATAGATTTCCCGTAAAG
C_S_H_R_D_L_*
_A_R_I_G_T_C_R_*
GCTTACATCTGGCTGCGGCATGTCGTGGCTATACAGGCAGTAGAGCAGGGGGCCGAGGACACTACCTTGT
GGAACACCAGCTGCCACGTTGTGCTCGGTGAAAATTGCTGAAATGAAAGCGCACAGCGAACCTCCTTCCTT
CCAGGTACGATTTTCAGCAACCCAAAATATGGAGCAGGCAGCGTCTGCTTGATTTTCAGTTGGAGTCCAAT
M
GTGCCACACTCGATCAAACGCTTCTCGAATGTCCAAGAAGAGGCTGTTACAATATTCCTTACTGTCGTAG
_C_H_T_R_S_N_A_S_R_M_S_K_K_R_L_L_Q_Y_S_L_L_S_*
GCAGTCAGAATTTGCTCGACGACTCGATGGAGCTGCTCGACAGTACAGTGGCCAGCACGGAAACCGAACT
GGTGTCAAGGGTAATCCCTGGGCTTCCATAATCCTTACAATCCGGACAGCAATCAGTCTCTCAAACAC
CTTCGAAATTGAAGGGAGGAGACTGATCGCGCGTAGGAGGGGCTCCCTTTCCGGTTGCCAGGTTTG
TGGATCATGGAGATTATCCCGAGCTTCCACTGATAAGGGAAGTATTGCAACCTACAATAGCGTTGAAAA
CCAACGTTATGTAGAGGATCGCTTGTCTGGGCAGGGCCTTCAATGTGGCGTTGCAGATGACGTCATGTCC
M_T_S_C_P
AGGTGCTTTGTTGTTACTCTGGCGCACAAATGACTTCGGCTACCTCGCAGGGTTCAAACGGCGTGATTGGC
_G_A_L_L_L_L_W_R_T_M_T_S_A_T_S_Q_G_S_N_G_V_I_G
ATATCCATTTGAAGCGCTTGGTTCAGCTGGTCTGAGTTTCTTCAACCTGTTGCAGGCTGGCAAGCTTGA
_I_S_I_*
ATGGTTGAAATCGCTCGCGAGGTGCGCAGCGAATACCTCCGCTTGTCCCAAGTCCGTTCCGACACCATGT
M_V_E_I_A_R_R_G_A_Q_R_I_P_P_L_V_P_S_R_S_D_T_M
CCCGTTGCTATCGACTAGAGGCGCTTCCCTCGTGCAACGCTTTTGATTGCGCGCGTGGCCTTCCACAGC
S_R_C_Y_R_L_E_A_P_S_S_C_N_A_F_*
GAATGCGTTCGCTTTCGACTCCGGTATTGGCAAGCCTTTCGTCGAACCATGCGGCTTTGTGCAGCA
M_R_L_C_A_A
CCAACGCGCATCGGAGCCTACTCAGAGCTCGGTTCCATTCCGTTTTGTCCGACGGGTGCCGAGACCTCAT
_P_T_R_I_G_A_Y_S_E_L_G_S_I_P_F_C_P_T_G_A_E_T_S
CCACCTAGTGGTAAGCGCTTCTCTCAGCTATAAGCAGCAGGATCTCCCTGGGTATTGGGATTTCGATCT
S_T_*
GCCGCGGGCTGCCGAGATATGCTGCGAGTTGCCAATCTAGCTGCTGACTTGATATTGTTGGTGAGCAGTT
CGATAGCATCATCAACGTCCTGTCCAGAGTTTTATGGCAGTGTGAGGGGCAGTGTGGACTCCAGTTGCCA
TTGGAATACCTCCAGGTTAGTACGCCTAGTGATAAGCCGCTCCATCTTAGGGTAGGATATTGCCCCCGCA
TCCAGCGTAATTACCAGAGGCAGGTGGTCCGAGCTAAGCTCAACAACCGCACTTATGTTAGCGTGGATGC
CCAGCACACCCTTTGTCAGTGCAAAAATCGATGTACCCTGGTGGAGCCTCTACAGCCGTACGGGTATCTTGT
TGGTCTCCTGTTGCTAGCGAGTCCACCTCCGAATTTAGGACGAGATTGCTAATGCAATGCCTCTTTGG
M_Q_C_L_F_G
TTGCTCCTTCCCGCACCCAGAGCCAGTGGGACGCGTTCCAATCGCCTGCAACAAGAACTTCGTCTGGAA
_C_S_F_P_H_P_R_A_S_G_T_R_S_N_R_L_Q_Q_*
CTCTTCCAAAATGTCTTAAACTCGTCCGTAGTCCATGCAAATCTCGGTGGGCAGTAGACCGCTCCAAG
M_Q_I_S_V_G_S_R_P_L_Q_S
CTTATAGGACCAAGTGCCGTTTCAATAACCGCCGGCGCAAGTTGCACCTTGGCAGTGGCTATTGGTGTTA
L_*
GCGGAAAGTGGGCAAGGCTATTTTTGATTAAGATAGCTGCTCCGCCTTTGGCGTTGCCACCACTCGGATC
ATTGGCAGTGTAGGCTACAAATCCGAATAGCTTAGGCGTCTCTGCCCCCTTGCAGTGTGTCTCGCTGAGC
AGTAATACATCGATTTCTGTGACGTCGCACGAAGCACTTACTTCAGGCACTTACTCTACGCAGGCTGGC
ATGCCTGCGGCAACTCCACGCATACTTCGTGCTGTGAAAATGAAAATTGTCCGAAGTTCCCATTTAAACT

M_P_A_A_T_P_R_I_L_R_A_V_K_M_E_I_V_R_S_S_H_L_N
 TTCCAAACAAAAGTAGTAATTTAAGAAATAAACTTCTTTGAATGGCAGAAGAACACAGATTTTATAGAGT
 F_P_N_K_S_S_N_L_R_N_K_L_L_*_M_A_E_E_H_R_F_Y_R_V
 TAAAACTTTCATGAATATGTGACCAGAGAAATTTTGCTGGCTTACTTGGTCTGAAGAAGTCAACTAAAA
 K_T_F_M_N_M_*
 CTAATAATTAATAATCGTATTTCTAAACTAAACTATAATATTTTTTACTTGTGATAAACGATATAACTG
 AAAAGCCAATTGATTTCCGTGGCACCCATTTTCAATTTACAAACGAACCAGATATTGCCATTTAATCCGT
 GCCACAGGGCGAACTAATTTCTGTTTGCTTTCAATTTCCAAACGTTAATTGCAACAAATTGGCTTG

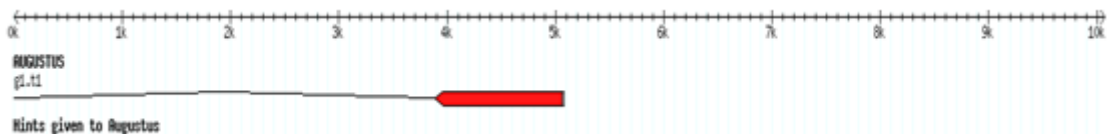
Para el gen 2L_120, el programa Augustus Gene Prediction predijouna estructura compuesta por 2 exones de 2373 pares de bases el primero y el segundo de 45 pares de bases, y 1 intrón de 62 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido SVTCLENVSLR con una probabilidad de $\rho=0.009$ y un valor $E=2.74974465259169e-05$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 19 sORFs. Se pueden observar repeticiones en tándem y desplazadas. Los sORFs decimooctavo y decimonoveno son repeticiones desplazadas de los sORFs quinto y sexto. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2L_122



GCAATAAGTCTCTCACATAATTTAGAAAAGGGAGGATCATCTGCGACTCAGGATCATCTTCAGGTTTTCCAGGCTTGTGGATCAACAAAATAGCAGCCATTTCCACTGTCTTGAAACACTTGGATCCTTAAAATAGCAT TAAAGATTAGTGCTGTATATATATTATCGCTTTTATAGGCAAACCGTTAGCGTGGCATGACAAATCCTGTC
 M_T_N_P_V
 AATTCTGGAGCTTTATGTTTCGGCAGTTGTTGTAATTTATAATTTATATTTCTTTCTTAATAAATAAA
 N_S_W_S_F_M_F_R_Q_L_L_*
 TAAATAGTCAAGTTTATGTTGAGTTTATGATTTTATAATTTAAGTTATTTCAACTGCAACACCAGCACC
 M_F_E_F_Y_D_L_Y_F_K_L_F_Q_L_Q_H_Q_H
 ACGACCTACTCACAGCAAAAACGTACAAGAAGGAAAGAAGGAATAAAAGAGTGGTATTCTCTCACAATA
 H_D_L_L_T_A_K_N_V_Q_E_G_K_K_E_*
 TGTTTTATGGCATAAAAAGGTGTGGCCATTCATATCAAATATAAAGTAGTGTGTTAACGTTATTTTTGT
 AGGTTGAATAGTATATTTCCAACACGCCCTCAAAAATAACGTCTATAATTTAAAACATAACAATAATAT
 TCATTAGTAACTATCAAATGTGGGTGGTGTGCATTTCTGGGAAGTGTTAACTGATCCAGCATTGCTGCGA
 M_W_V_V_C_I_L_G_S_V_N_*
 GGATACGGGGAAAACCCAGAAAAACCCGATCAGATCATTTGTAACAAAATATGTTTACAGGAATTTAAATT
 TGAAAAATATATATAAAAAATTTCAATTCAGCATTCGATTGGTCTGCTTGCAGCAAACCCAATTTGTCTCGT

AACTCCACAAATCTCGCAGCAGGCAACGGTTTTGTAAATATGTCAGCCAGTTGATTCTCTGTAGGAATAT
ACTCAAGACAAATCACATTATTCTGAACTTGCTCTCTGGCAAATGATATTTAATATCAATATGTTTAGC
M I F N I N M F S
TCGTTTATGACATGAGGGGTTGTTTGCTATGCTAATACAGCCTTGATTGTCTTCGTAAATTTTAATGGGG
S F M T *
TTTTCTAGTTTAATGTTAATACTAGTTAATAAAAAATTTAAGCCATAGAGCTTCTCTCACGGCTTCAAATA
GGCCATATACTCAGCTTCAGTTGATGAGGCTGCTACTGAGTTCTGTCTCTTTGTATTCCAACAAATGAG
M R L L L S S V S L Y S N K *
ATAAAAATCAAACATTTTGAATAAAATACCCTGTTGTACTTTTTCTATCAATTTCACTACCAGCCCAATCA
GAATCCACATAACCAATAATTTTATTTTCAAATGCCAAGTTCTTTTTAAAAATCAATTTTCATATCGATAG
TGCCCTTCAAATATCTAAGAACTCTTTTTAAGTTCTGCCATAATTCGGAGTTATTTTTGCTACTATATCT
GCTCAAGATATTTACTGCAGTAGTTAAATCTGGGCGTGTACAAAGCATTATGTACATTAACATCCTATG
M Y I K H P M
AGGCTACGGCATGGGGTATTACAGTCTTCATCTGAATTAAGTAATTCATAATTTATTTTACTAGGTAAAG
R L R H G V L Q S S S E L S N S *
M G Y Y S L H L N *
GAGTACTAACTGCATTACAATTTCCATGTTAAATTTACTTAAAATTTTTTTTTAACATATGCAGATTGGC
M L N L L K I F F N I C R L A
M Q I G
TTAAATAGATTTTATCTTCATGCATCTCTATCCTAATTCCAATAAAATGTTTTATTTTCATTTAGGTCAGT
*
L N R F Y L H A S L S *
M H L Y P N S N K M F Y F I *
CATCCTAACTTTTCCATTAATAACCTTTTGAAGTTATTCATTCCTGTGCATATCTCCTGTAGCTATAACC
ACATCATCTACATATAATAATACATATATGTTTTCAATGATGTTACCTTTGTCTAAAATATATATACAGC
M F S L M L P L S K I Y I Q
GATCAACTGAAGAGTTTACAACTCACACTCTTTTCAATGCTTGCTCAAATACTTCAAACCAGCATCTAGC
R S T E E F T N S H S F N A C S N T S N Q H L A
M L A Q I L Q T S I *
CGCTTGCTTGAGTCCGTAATTTGCCTTATTCAATTTACACACATTTGTCACTATTACACGATATACCTTGA
A C L S P *
GGAAGTCTCATATAAATTTCCCTCTTTTAAACGTGCCATTTAAGAAAGCTGTTTTTACATCCATTTGATGGA
M D
CTTTCAAGTTATACTGTATTGCTAATGACAATATAAATCGGAAACTTGAAATTCCTAGCTACAGGAGCAAA
F Q V I L Y C *
TGTCTCTTCATAGTCTATTTGGTATTTTTGAGTGAATCCTCGTGCAACCAATCTAGCTTTGTATCTAATT
M S L H S L F G I F E *
GGATTTCCAAGTTTCAATATATTTAACAGAAAATACCCATCTGCTATCTACAATATTTTTGTTTTCAGGCC
TTTTTTGTAATTTGTTCAAGTATTATTAATTTTATGAGCATTAACTCTGTATTGATGGCTTCTTCCAAGA
M A S S Q E
AGATTTATCATCCCTATATTGAATTTTCATCAAATGAATTTGGGACATCATTAAATATAGTGTGAGCATT
D L S S L Y *
AGAACAACCTTTATTTAGACTATTATCCTCTTCATTATAGGATATCTGAGGCTTAGTCTTTAATCTCTCAC
TTCTTCTATTAATAATTTCTATGCCATCATTTTTAGTTGGATTATCAATTCCAATTTCTTTTTAAGTGCTC
M P S F L V G L S I P I S F K C S
TGCTGTTTTCACTTTCCCTACTCTCATTTCGGGTTGCCTGATCCCTTACTTTTCAATTCAGGTGATCATCTCGC
A V S L S L L S F G L P D P L L S F R *
TTTTTTTTCTTACTCTCATTTCAGAAAATATTTATTACTTTTCCCTTACTATCTTTTCAGGAATTGTATGTTGT
CGCATTCCCTACTCTCATTTCGGGAATTTCTGTTTGTATTATTTTCCCTACTGTCATTTCGGAAAATTTTTATT
TTCATTTCCCTACTATCTTTTCAGGAATTGTATGTTGTGCGCATTCCCTTACTCTCATTTCGGGAATCTGTT
TGTATTATTTTCCCTACTGTCATTTCGGAAAATTTTTATTTTCACTTTCCCTTACTATCTTTTCAGGAATTGTA
TGTTGTCGATTCCCTACTCTCATTTCGGGAATCTGTTTGTATTATTTTCCCTACTGTCATTTCGGAAAATTT
TTATTTTCACTTTCCCTTACTATCTTTTCAGGAACACTGTTTCAAATTTAACAGCTCTAGAATTAACCATAT
TGGTTTCATCGACAACAACATCTCTTGCAGACAATAAATTTTTTCAATTTACAGCATCCCACAACCTAAAACC
ATTGGGTTTCATAGCCCACAAAAATACTTTTAAATGATTTATCATCAAACCTTTCCCTGTTTGTTTTTAATA
M I Y H Q T F L V C F *
TGCACATAAACAGTTGCACCAAACACTCTCAAATGTTTTAAGTATGGCTTCTTATTGTGCCACATCTCAT
M H I N S C T K H S Q M F * M A S Y C A T S H
ATGGGGTCTTTGAACTATCAACAAGTGTCTGCTAGGAATTTCTGTTGATTAAATAAGTAGCAGTTAATAC
M G S L N Y Q Q V L C *
TGCTTCGCCCCAAAAGCTTTTATCTAGCTTTGCACCCTAACCATGGTTCGAGCTTTTTTCCGTAATGGTT

M_V_R_A_F_S_V_M_V
 CTTATCATTCTCTCAGAAACACCATTAACTGAGGTGTATGTGGCACTGTTAAGTGATAAGAAATTCCTT
 L_I_I_L_S_E_T_P_F_N_*
 TCTTAACACAAAATGTCTCATCTCATTTGACAAGTATTCTCTACCATTGTCAATGTATAAGTACACAAC
 M_Y_K_Y_T_T
 CTTTAAATTTAAATGAGCTTCACTCTTGGCTACAAAATCTTGAAACATGCTAAACACATCAGATTTTATAT
 F_K_L_K_*
 M_S_F_T_L_G_Y_K_I_L_K_H_A_K_H_I_R_F_I
 M_L_N_T_S_D_L_Y
 TTAATTTAAATAAGTTACACAATAATGTGTAACTGATCAACAAAGATCACAAAATAATTTTTTATCATCTA
 F_N_*
 L_I_K_* M_C_K_L_I_N_K_D_H_K_I_I_F_I_I_*
 AAGTAACTGGAGTAATAGGCCACAGACATCTGAGTGTACTACAAAAGTGGTCTTTTAAATATGGGTCTT
 M_G_L
 ATCTTTCAATTGTTTAAAAGGAAGTCTTGCCTGTTTACCATTTAAACAGGGTTCACAAATTTACATGAT
 I_F_Q_L_F_K_R_K_S_C_L_F_T_I_* M_I
 AACTCTAAGTTGTTTGAAGACTTTGATCACTAAACATATTCTTTTCGTTTTATTTCTAATAATTTGCCAT
 T_L_S_C_L_E_D_F_D_H_*
 CGCTTATATGGCCAAACCTCTCATGCCATAAACGAAAATTATTTTTATGCTTAGCATTTATAGAATATGC
 M_L_S_I_Y_R_I_C
 TTGAAAATTTGATCACAGGTACATTGTTTAAACATACCTGAATTTTTGACAACCATTAACCCATTTTTTCGAA
 L_K_I_D_H_R_Y_I_V_*
 ATGGTTACACCGCTTTTGTCAAATTCGATCGACATTCCTGCCTCTTGGAGACGCTTTACGGACATCAAAT
 TACCAGCAGCTTCTTACAAAAGAGTACATCCTCCAGTGAATCTCATGGTCATTCGGTAGTCGGACAAT
 ACCACGCTTAGTGGCATAAAATAAATTCGCCCTGCTTGGCCACTGCAATCTTAAGTGGAGGCACAACCTCC
 AACTGTCCGTATACAGCGACTCATCTTTATAAGATGGTCCACTAGCACCAGAATCAAGGACAAAACCCGC
 M_V_T_S_T_R_I_K_D_K_P_A
 AGTTGTCCATCACTGAAGTATTATTCACCTTCTTTTACCATAAACGCAATGCCGTGTGATGTTGCAGTTTG
 V_V_H_H_* M_L_Q_F
 AACTTGTTTTTTCATTTTTCTTTATTTTTTATTATTTAATATTCTTTTTATAATGGAAACAATCTTTTTTAATG
 E_L_V_F_H_F_L_Y_F_Y_Y_L_I_F_F_Y_N_G_N_N_L_F_*
 M_E_T_I_F_F_N
 TGGCCTTCTCTGCCACAGTGGTGACACTTGACTTTTATACTTTGAATTTCCCTTGAATATTTTCTTTGGTT
 V_A_F_S_A_T_V_V_T_L_D_F_I_L_*
 TAGTTACCCGATTTTTTAAACAAATTTATTTTTATAAGTGTATTATTGTTGTGCACGATCGCGTTCATAAC
 TTTCTTGCTTGTATCGTTGTGGTCATTTTTAATTTTTAATTTCTTGATCCAGCAATCTATTTTTTCACAAAC
 GCCAATGTCAAATTTTCTTACAGATAATGTCTCTATCGCTGTAATAATTCATCGTAACACGAAGGCAATG
 M_S_N_F_L_Q_I_M_S_L_S_L_*
 TGATCAGTAGATGAGAAAATTTTATCCATCTCTCTATTTTTTGACCAGCTGCCAACAAATTCACCTATAAG
 M_R_N_F_I_H_L_F_Y_F_C_T_S_C_Q_Q_F_T_Y_K
 TTCGTCAAAAATATGAAAATGGCTTAATAGTGACATCTCACTCGATAGCTTCAGAGAAAGCAAACGTTTT
 F_V_K_N_M_K_M_A_*
 CGCAAGCCAGTTGCGACGCCAAACTTTTTCGTTTATAAACGGCGTCCAAATTTCTCAAGAATCTGACGCG
 CCGTAAATGTCTGTTGCGAAATTTAAAACGAGTTCGCTTAGGTACTCTATTATTGTACTTTTTTGACA
 ACGCTCTGCCTTTTTCCAGGAGTCACTTACCTCGTTAGGCATTTAAACCATCAACTACTTTAAGCACATCT
 TGCTCGGCTAAAAGAGCCCTAATTTCTAAATTTCCAAATCGCGTACTTCTCGCCATCAAACGGCTTAATAT
 TACGTTTAGCCTTGTCCATTTTTCACTCAAATTTCTGAGAAGGAAATAATTTCAAGAATGTGAAAAAAA
 GCTATTTTCAAGAATGTGAAAAAAAATGCTATTTTCAAAATTTATTTTCAAAATCTTTAATTTCAAAAT
 M_L_F_H_N_L_F_S_Q_S_L_I_H_N
 TAATTGTTTATTAGGCATGGACTGGGCCATAACCTGTTGTAATTTATAATTTATATTTCTTTCTTAAT
 L_I_V_Y_*
 AAATAAATAAATAGTCAAGTTTATGTTTGGAGTTTTATGATTTATATTTTAAAGTTATTTCAACTGCAACAC
 M_F_E_F_Y_D_L_Y_F_K_L_F_Q_L_Q_H
 CAGCACCACGACCTACTCACAGCAAAAACGTACAAGAAGGAAAGAAGGAATAAAAGAGTGGTATTCTCT
 Q_H_H_D_L_L_T_A_K_N_V_Q_E_G_K_K_E_*
 CACAATATGTTTTATGGCATAAAAAGGTGGCCATTCATATCAAATATAAAGTAGTGTGTTTAAACGTTA
 TTTTTGTAGGTTGAATAGTATATTTCCAACAGCAGTAATTTAATTTTCAAAGATTTCTTCTATCCTTTT
 TGGCTCAATAGCCCACGACAGAACAAATAAGAGAGACTCATTTGTGCACTTAATGCACTGCTCAACAACA
 M_H_C_S_T_T
 ATTTGTTTTATACACATTTGTATTTAATTTTTTAAATATTTTCAAAAACTTTAAAGGTGCATTAGGCCCT
 I_C_F_I_H_I_C_I_*

AATCTGTTCTCTTCTCGGTAACGACATTTCTGTGCAAGGGAAAATGCGCTCGGCTGCTGCACTACTTGC
GGTATAGAAAAGCACCTGCAGCGGAACTTAGAAAAGTTAGGAATAAAAGGTCTCATTGTTTTTCCACA
G_I_E_S_T_C_S_A_N_L_E_S_*
TCCAAGTCTTCGCTAATGAAGACTTTTTTCATTGCTGTATTTTTTTTAAATTTTCATCATACGCGGTAAGGC
M_K_T_F_S_L_L_Y_F_F_*
TGAGCAAGTTTGAAGAGGAATTTTTTCGGGGAAAAAGAATTTTTCTTTTAAAGTATAGGAGAATCTGTTTCG
GAATTTTTTGATGAGGATATAGCCGGGCTATGTCGCTACTTGTAGGCGTTTGTGTGAAAAGTAGCTAAA
M_S_L_L_V_G_V_C_C_E_K_*
ACTTAGTTCTTCCATTTTGGTTAAAATAAAATTTCTTAATTTTATATACAGTATAAATTCGCTTAGCTGCAT
CGATAGTTAGCTGCATCGGCAAGATATCTGCATTATTTTTCCATTTTTTTGTGTGAATAGAAAATTTGTA
CGAAAATTCATACGTTTGTGCTGCATCGCAGATAACAGCCTTTTTAACTTAAGTGCATCATATCAGCTGTTT
TTTTTGCCAATTTCAATGAATATCATCAAAGTTAGCTGCGCCATCTATGAATCATTTTTTGCATATCTAAA
M_N_I_I_K_V_S_C_A_I_Y_E_S_F_L_H_I_*
AGATGGCAAGAATGCCAACTCGTTTCAGTATCTGCGCATGTCCGTTTTTGTTTTTGCTTTGATCGTGATT
M_A_R_M_P_T_R_F_S_I_C_A_C_P_F_L_F_L_L_*
TTTTGTTTTGTTTTCTTATGACACAAAGTTATTTAAAATGGGTA AAAACAAAGCGTGTCTGTTGGACTAACA
F_V_F_L_F_L_M_T_Q_S_Y_*
CTAAAGGAAAAGCTTCAAATAATCGAGTTAGTGACCAACAAAGTGGACAAAAGGAAATTTGTGCCAAGT
TCAAATGCGACAGATCCACAGTCAACCGCATTTTACAAAAACAAATGAAATTCATGAAGCTGTGGTCGC
M_R_Q_I_H_S_Q_P_H_F_T_K_N_K_*
GTCAGGTTTAAAAAGAAAGCGTCAAAGAAAAGGAGCGCACGACTTAGTAGAAGAAGCCTTATACATTTGG
R_Q_V_*
TTCGGACAGCAGGAATCAAAGAACGTAATTTCTTGACCGGCACGTCATATTAGCAAAAGCGAAAGAATTTT
GCCAAAAATTTAACGACGCCTTTGAACCTGACGCCAGCTGGCTTTGGCGCTTTAATTCACAATTTAATTG
TTTTATTAGGCATGGACTGGGCCATAACCTGTTGTAATTTATAATTTATATTTCCTTTTCTTAATAATAA
ATAAATAGTCAAGTTTATGTTTGAGTTTTATGATTTTATATTTTTAAGTTATTTCAACTGCAACACCAGCAC
M_F_E_F_Y_D_L_Y_F_K_L_F_Q_L_Q_H_Q_H_
CACGACCTACTCACAGCAAAAAACGTACAAGAAGGAAAGAAGGAATAAAAGAGTGGTATTCTCTCACAAT
H_D_L_L_T_A_K_N_V_Q_E_G_K_K_E_*
ATGTTTTATGGCATAAAAAGGTGTGGCCATTCATATCAAATATAAAGTAGTGTGTTTAAACGTTATTTTTG
TAGGTTGAATAGTATATTCCAACAGCAGTAATTTAATTATTTCAAAGATTTCTTCTATCCTTTTTGGCTC
AATAGCCCACGACAGAACAAATAAGAGAGACTCATTGTGCACTTAATGCACCTGCTCAACAACAATTTGT
M_H_C_S_T_T_I_C_
TTTATACACATTTGTATTTAATTTTTTTAAAATATTTTCAAAAACTTTAAGGTGCATTAGGCCCTAATCTG
F_I_H_I_C_I_*
TTCTCTTCTCGGTAACGACATTTCTGTGCAAGGGAAAATGCGCTCGGCTGCTGCACTACTTGCAGGTATA
M_R_S_A_A_A_L_L_A_G_I_
GAAAGCACCTGCAGCGGAACTTAGAAAAGTTAGGAATAAAAGGTCTCATTGTTTTTCCACAACCTCCAAG
E_S_T_C_S_A_N_L_E_S_*
TCTTCGCTAATGAAGACTTTTTCATTGCTGTATTTTTTTTTTAAATTTTCATCATACGCGGTAAGGCTGAGCA
M_K_T_F_S_L_L_Y_F_F_*
AGTTTGAAGAGGAATTTTTTCGGGGAAAAAGAATTTTTCTTTTAAAGTATAGGAGAATCTGTTTCGGAATTT
TTTTGATGAGGATATAGCCGGGCTATGTCGCTACTTGTAGGCGTTTGTGTGAAAAGTAGCTAAAACCTTAG
M_S_L_L_V_G_V_C_C_E_K_*
TTCTTCCATTTTGGTTAAAATAAAATTTCTTAATTTTATATACAGTATAAATTCGCTTAGCTGCATCGATAG
TTAGCTGCATCGGCAAGATATCTGCATTATTTTTTCCATTTTTTTGTGTGAATAGAAAATTTGTACGAAAA
TTCATACGTTTGTGCTGCATCGCAGATAACAGCCTTTTTAACTTAAGTGCATCATATCAGCTGTTTTTTTTG
CCAATTTCAATGAATATCATCAAAGTTAGCTGCGCCATCTATGAATCATTTTTTGCATATCTAAAAGATGG
M_N_I_I_K_V_S_C_A_I_Y_E_S_F_L_H_I_* M_
CAAGAATGCCAACTCGTTTCAGTATCTGCGCATGTCCGTTTTTGTTTTTGCTTTGATCGTGATTTTTTGTG
A_R_M_P_T_R_F_S_I_C_A_C_P_F_L_F_L_L_*
TTTTTGTCTTATGACACAAAGTTATTTAAAATGGGTA AAAACAAAGCGTGTCTGTTGGACTAACACTAAAG
F_L_F_L_M_T_Q_S_Y_*
GAAAAGCTTCAAATAATCGAGTTAGTGACCAACAAAGTGGACAAAAGGAAATTTGTGCCAAGTTCAAAT
M
GCGACAGATCCACAGTCAACCGCATTTTACAAAAACAAATGAAATTCATGAAGCTGTGGTCGCGTCAGG
R_Q_I_H_S_Q_P_H_F_T_K_N_K_*

TTTAAAAAGAAAGCGTCAAAGAAAAGGAGCGCAGACTTAGTAGAAGAAGCCTTATACATTTGGTTCCGGA
 V__*
 CAGCAGGAATCAAAGAACGTAATTCTTGACCGGCACGTTCATATTAGCAAAGCGAAAGAATTTTGCCAAA
 AATTTAACGACGCTTTGAACCTGACGCCAGCTGGCTTTGGCGCTTTAATTCACAATTTAATTTGTTTATT
 AGGCATGGACTGGGCCATAACCTGTTGTAATTTATAATTTATATTTCCCTTTCTTAATAAATAAATAAAT
 AGTCAAGTTTATGTTGAGTTTATGATTTATATTTTAAGTTATTTCAACTGCAACACCAGCACCACGAC
 M_F_E_F_Y_D_L_Y_F_K_L_F_Q_L_Q_H_Q_H_H_D
 CTACTCACAGCAAAAAACGTACAAGAAGGAAAAGGAATAAAAAGAGTGGTATTCTCTCACAAATATGTTT
 _L_L_T_A_K_N_V_Q_E_G_K_K_E_*
 TATGGCATAAAAAGGTGTGGCCATTCATATCAAATATAAAGTAGTGTGTTTAAACGTTATTTTTGTAGGTT
 GAATAGTATATTCCAACAGCAGTAATTTAATTATTTCAAAGATTTCTTCTATCCTTTTTGGCTCAATAGC
 CCACGACAGAACAAATAAGAGAGACTCATTTGTGCACTTAATGCCTGCTCAACAACAATTTGTTTTATA
 M_H_C_S_T_T_I_C_F_I
 CACATTTGTATTTAATTTTTTAAAAATATTTTCAAAAACTTTAAGGTGCATTAGGCCCTAATCTGTTCCCTC
 _H_I_C_I_*
 TTCTCGGTAACGACATTTCTGTGCAAGGAAAATGCGCTCGGCTGCTGCACTACTTGCAGGTATAGAAAGC
 M_R_S_A_A_A_L_L_A_G_I_E_S
 ACCTGCAGCGCAACTTAGAAAAGTTAGGAATAAAAAGGTCTCATTGTTTTTCCACAACCTCCAAGTCTTCG
 _T_C_S_A_N_L_E_S_*
 CTAATGAAGACTTTTTTCATTGCTGTATTTTTTTTTAAATTTTCATCATAACGCGGTAAGGCTGAGCAAGTTTG
 M_K_T_F_S_L_L_Y_F_F_*
 AAGAGGAATTTTTTCGGGGAAAAAGAAATTTTTCTTTTAAAGTATAGGAGAATCTGTTTCGGAATTTTTTGAT
 GAGGATATAGCCGGGCTATGTCGCTACTTGTAGGCGTTTGTGTGAAAAGTAGCTAAAACCTAGTTCTTC
 M_S_L_L_V_G_V_C_C_E_K_*
 CATTTTGGTTAAAAATAAATTTCTTAATTTTATATACAGTATAAATTCGCTTAGCTGCATCGATAGTTAGCT
 GCATCGCAAGATATCTGCATTATTTTTCCATTTTTTTGTGTGAATAGAAAATTTGTACGAAAATTCATA
 CGTTTGTGTCATCGCAGATAACAGCCTTTTTAACTTAAAGTGCATCATATCAGCTGTTTTTTTTGCCAATT
 TCAATGAATATCATCAAAGTTAGCTGCGCCATCTATGAATCATTTTTTGCATATCTAAAAGATGGCAAGAA
 M_N_I_I_K_V_S_C_A_I_Y_E_S_F_L_H_I_* M_A_R
 TGCCAACTCGTTTTCAGTATCTGCGCATGTCCGTTTTTGTTTTTGCTTTGATCGTGATTTTTTGTGTTTTTG
 M_P_T_R_F_S_I_C_A_C_P_F_L_F_L_L_*
 M_S_V_F_V_F_A_L_I_V_I_F_V_F_L
 TTTCTTATGACACAAAGTTATTTAAAAATGGGTAAAACAAAGCGTGTCTGTTGGACTAACACTAAAGGAAAAG
 _F_L_M_T_Q_S_Y_*
 CTTCAAATAATCGAGTTAGTGACCAACAAAGTGACAAAAGGAAATTTGTGCCAAGTTCAAATGCGACA
 M_R_Q
 GATCCACAGTCAACCGCATTTTACAAAAACAAA
 _I_H_S_Q_P_H_F_T_K_N_K_*

Para el gen 2L_122, el programa Augustus Gene Prediction predijouna estructura compuesta por 1 exón de 1187 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido LFQLQHQHHDLLTAKNVQEGK con una probabilidad de $\rho=0.032$ y un valor $E=0.000340108579655668$; para el péptido LLQMIPYQQMSTKMI con una probabilidad de $\rho=0.055$ y un valor $E=0.000666173665229102$; para el péptido LNLLKIFFNICR con una probabilidad de $\rho=0.054$ y un valor $E=0.000310923654121335$; para el péptido ASSQEDLSSLY con una probabilidad de $\rho=0.003$ y un valor $E=3.23564926321861e-05$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 62 sORFs. Se pueden observar repeticiones. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 4 péptidos definidos

por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_39



K T Q S F A V L R S V V Y S I L G R S A S V H
TATAGATTTGTGTAACCTTCAGCACCTTCTTTGGATATCGTTCATCGATGAATCCCTCTGGCTGCCTCATA
I * M N P S G C L I
TATACCGTTTTCGTGAAGATCACCATTCAAGTATGCAGACGAGACGTCCATTTGGTGCATGTAGAGACCGT
Y T V S * M Q T R R P F G A C R D R
TTTCCACCGCCAAAAGCAATTACTAATTTAATTGATGAGTAGCGTGCCACTGGTGAACGTATCCGTGAA
F P P P K Q L L I * M S S V P L V R T Y P *
GTTAACGCCGTATCGCTGAGCGCACCTTTTGGGACAAGTCGAGACTTAAACCTCTCGACTTTGCCATC
TTGTTACGCTTCACGCTTTCGACCCAATGGCCTTCTCACCTGCCGGCAAATCCACCAATGACCGTTTT
M T R F
GTTTGCACGTAGGCTCTCCAACCTCTTCTGCATTGAGGTTCTCCATTCTTTCAGAGTTCCTTATGGTTTCC
C L H V G S P T P S A L R F S I L Q S S L W F P
GCCTCCTTGACGCTTCTTGGAGTTGGAACCTCATCCGTACTCTTGCCTTACCCTCAGCAACACAGACTGG
P P *
AAGGTGCTGCATCTCCACAAGTTGAACATCGGGAGAAGGTACTGCTTGCCCTGCCGTACATACCATCATCC
TGGTCGTTCTCAAATACAATTACGTACAGCAACACAGATATTTCTTCCGTAGATTGAACAGGCGAT
ATCCCTTAGAAGCATCAGAGTATCCAACCAAAATATTTTCTTCGCCCTTCGCAGAAAACCTCCTTTTTGT
GGACTTGTTAAGCACGATAGCCTTCGAACCAAAATATCTTCAGGTGGGACACATATGGCTTCCTTCCAGTC
M A S F Q S
CAAGACTCGAACGGTGTTACATCAGATAATGCTGACGCTCAGCGGATTGCGTAGATAAGCAGCAGTAT
K T R T V L H Q I M L T S Q R D C V D K Q Q Y
TGATAGCTTTCAGCCATAGTGAGTGGCCGACGCCCTGCGTGAATCATCATAGTCTTGCCTTCCATCTCGACCAG
*
GGTTCGGTTAGCGCGCTCCGCTACACCGTTTTGTTGTGGGGTGTGCGGTACCATCAATTGTCTCTTTATG
CCGTTTTCTGTCAAGATACATTGAAACGCTTTACTTAAATACTCACGCCCGTTGTCGCTCCTTATTGCTT
TCAACTTCTTTGGTTTTTCGGCAAATGCTTTAAACTCCTTGAATTTGGTCAGTATTTTCATCTCTCGTCTTT
ATGAAATACACAAACATGTATCTCGATTTGTCATCTACGAAAGTGACAAAATAGCGTGCACCTCCAACAG
M K Y T N M Y L D L S S T K V T K *
ACGTCGTTTTGCATTGGACCACAGATATCACTATGCACTAGCTCCAGTACATCAATTGTGCGATTTCCAGA
GGCTTTTTGGGAAACTCTTCACACATATCTTGCCTTTTTGCGCAGCTCAAACAGTCATTTTTATTTGCGCCA
CTTAAATATTTTTTCTTTAGTCCGAGTACCATTTCTTGTGTTGATCATTTTTGTTCAAGCTTGCAAAGTTCA
AATGACCGAATCTCTGGTGCCTTTCTCATTAATTTGTCTGTGCAACGAAACATGCCTTTCAACTTT
M T E S L V P L S H * M H F Q L
TCTTTCATTTTGAACATAAAACAAATCACTTTGCAGTTCAGCAATCAAACAACCTTTATCTTATCATT
F F H F G T *
TTTTAAAATTCGCGCTGTTTTTCGAACGACACACGAAATCCGTTTTCAATTGCCTTCGACACCGACATAA
AGTTGCATTGTAAATTCCTTACAAAGAGTACGTGATCAATTCAAATGATGAGTTGCGCCAGGCTATTTT
M M S C A R L F
CACTGTACCACGACCGTTCAGCCATAATATACTTATTGCCAACCAAAAGAATGCGCTCTTTATGCTCTCTG
S L Y H D R Q P * M R S L C S L
M L S
AAGCTTTTAAACATCGATCTCTCACCGCACAAATGAGCCGTTGCTCCTGAATCAAGGCACCACATATTTT
K L L N I D L S P H K *
E A F K H R S L T A Q M S R C S *
TTGGCAATTCGCCAAGCTGTACAGTTGAGCAGAGCACAGCTTTGTGGCCGCTGCGACCACAGTTCCAACA
ATTAACGTCTTTTTCCCGTTAATCAGTTGCTGATTATTTACTTGCACACTCTTCTTAATGTCACTTTTC
M S L F
GCTTTTTGTTTCGACAGCCGAATTCCAAATGCACTTTGTTTCGCTTTTTGCGAGAATGTTCTTAATTTGCCA
A F C F A D R I P N A L C S L F A E C S *
M F L I C H
TACGCCCTTGCCTTCTCTGTAATTTTTATTTTTAACATTTTCAATGTGGGCAACTCATCGCGCGTTTC
T P L P F L L *
TATTGCAACTACGAACTTTCAAACGATTCGCGCAAACCTGGACAACAGGAGAATACTCAGGACTTCTTCT
TGGATCACCAATTGAATTTCTTTTAGTTTTTCAACAGCCGCACAAAATTTGTTACATGACTCGAAACAA
M T R N N
CTTCTGTTTCAGACATTCTCATGTGCAACAACCTGCTTTAGCAAACATATACGTGCGCGGTCCTGAGGG
F C F R H S H V Q Q L L *
AGTGTGGATTTCACTAAGTCTTTGCCATGCCCTTACAGCAGTCTCACAGTTTTTTATATGGTTAATTTGG
M P L Q Q S H S F L Y G *
M V N L
GATGCTTAATGCATAGCATTATACTTGCTAATGCTTTTTTTCGCTTTCGCGTGAACAAAGCCTTTAGCT

G_C_L_N_A_* M_L_F_R_P_S_R_R_T_K_P_L_A
 CAGCACTATCACTTTTCGTTTTTAACCCACACGTCCACACACCACCCCCATAATTCTGCATGAATTAAGAC
Q_H_Y_H_F_R_F_*
 ACTCTTCATTTGTACAGCCCACGCACTGTAGTTTTTCATCATCAAGTTTTTCTATTTGGTTTTATTGATGAA
 CTCATTTCCACTTCGATAAGTTTTTCGTATTTAAATCCCTGAATCACAATATTTAAAGATGCCGAATATTC
M_P_N_I
 TAATTCGTTTAGAACTATTCACTTTTATGTCCGCGTGGGCTGGGCCATAACCCCATTCGTTTCAAGTA
L_I_R_L_E_L_F_T_F_M_S_A_W_A_G_P_I_T_P_I_R_F_K_Y
 TTTATTGTGACTGATCTAGAGTACAATGTTACTGCACGCAGGCAGCCAAATTACACGCTAAAGAGGATAA
L_L_* M_L_L_H_A_G_S_Q_I_T_R_*
 CACAATAAGAGAAAAAAGCTTTTTGGTTACCGCGACCAAAGAGGGAGACACAGTGCATTGACTTTATTAC
 ATATTATTATATATACATACATATTACATACATATTTTCAACACTCCCTCTCAATGCAATATGTCGACCC
M_S_T
 TTACGGATAATATTTCTATCTTAATTCCAAATCAAATACATTTTTGTAGTTTAAACCCATCCCTTGG
L_T_D_N_I_S_I_L_I_P_K_S_K_Y_I_F_V_S_L_N_P_S_L_G
 ATACACTTTTCATGTTTAAATCTATTCAAATCTTGGTTAGAACATCTGCCATCATGTTTTCTGTGGTA
Y_T_F_H_V_*
M_F_K_S_I_Q_I_L_G_*
 CATAATTAACCTCTATTATATTGTTCTTATACAGTCTCTTATATAATGATATTTTATGCTTACTACGTG
M_I_F_Y_A_Y_Y_V
M_L_T_T_C
 CATGGAAAGTTTCATTTTTCACTAAGCACTGGGAECTCTGGTTGTCGCTATAAACCTTAGTTGCCTGTGT
H_G_K_F_H_F_S_L_S_T_G_N_S_G_C_R_Y_K_P_*
M_E_S_F_I_F_H_*
 TTCACAAAACCCATTTTCGTTGACCAATTTCCGAACATACGCCACATCCTTCGCTGCCATGGAAAGTGCC
 ACATACTCTGACTCTGTGCTGCTTAGGGAAACCACTGATTGTTTCTTTGAGTCCCACGTAAGAGCAGCTC
 CTGCAGCAATGACTGCCATCCAGTAAAGAATTTTCGGTCCGTGGTGTACCAGCCCTATCCGCGTCAAC
 ATAGCACTATATAGCACACCAGTACGCTCGTAGTGTAACTGCAAATCAGCTGTGCCCTCAAGTATCGC
 AGAACTCGCTTTGCACCGCTTCGTACTCTTTATGTGGGTGAGAGTTCGGTTGTGATAGTTAATGACCG
M_T
 AATGCATAATATCAGGACGAGTTGTGATTGCCAGATACATTAATGAGCCAATCAGCGATTGAAAGCTGAC
E_C_I_I_S_G_R_V_V_I_A_R_Y_I_N_E_P_I_S_D_*
M_H_N_I_R_T_S_C_D_C_Q_I_H_*
 TTGATTCACCTTCTGACAATCGGCCTTGTACATTTTACTTCAAAGCCTGCTTCCAGTGGGATGAGATTC
M_R_F
 GGCTTACAATCCTGCATCGAATAATCATTGATAAGGCTCTCAATATACTGCTTGTGGCCAATCCTGATTC
G_L_Q_S_C_I_E_*
 CACCAGTTTCTCCGTCACGTTCAATTTTCGATACCCAGGAAATGTCTTAGCTCGCCCCGCTACGACATC
 GAATTCCTTTTGAATTTGATGCTTTAATATCGCACAGTTCCTCCTTGCCTGAACTGGCTATGATAAGGTCG
M_I_R_S
 TCTACGTACATTACAACAAGGTTCTTACTATTGCCAGAATTACGGGTGTAACACATGGCTCGCTAGGGC
S_T_Y_I_T_T_R_F_L_L_L_P_E_L_R_V_*
 ATGAAGAAAACCAATCTTTGCACTACTTCGTTTTCAGTAGTTTATTCATTCTCGGCCGCTCTGCTTCAG
M_K_K_T_Q_S_F_A_V_L_R_S_V_V_Y_S_I_L_G_R_S_A_S
 TCCATATATAGATTTGTGTAACCTTCAGCACCTTCTTTGGATATCGTTCATCGATGAATCCCTCTGGCTGC
V_H_I_* M_N_P_S_G_C
 CTCATATATACCGTTTTTCGTGAAGATCACCATTCAAGTATGCAGACGAGACGTCCATTTGGTGCATGTAGA
L_I_Y_T_V_S_* M_Q_T_R_R_P_F_G_A_C_R
 GACCGTTTTCCACCGCCAAAGCAATTACTAATTTAATTGATGAGTAGCGTGCCACTGGTGAAGACGTATC
D_R_F_P_P_P_K_Q_L_L_I_* M_S_S_V_P_L_V_R_T_Y
 CGTGAAGTTAACCGGTATCGCTGAGCGCACCTTTTTGGGACAAGTCGAGACTTAAACCTCTCGACTTTG
P_*
 CCCATCTTGTACGCTTCACGCTTTCGACCAATGGCCTTCTCACCTGCCGGCAAATCCACCAATGACCA
M_T
 GGTTTTGTTTGCACGTAGGCTCTCCAATCCTTCTGCATTGAGGTTCTCCATTCTCAGAGTTCCCTTATG
R_F_C_L_H_V_G_S_P_T_P_S_A_L_R_F_S_I_L_Q_S_S_L_W
 GTTTCCGCTCCTTGCAGCTTCTTGGAGTTGGAACCTCATCCGTACTCTTGCCTTACCCTCAGCAACACA
F_P_P_P_*
 GACTGGAAGGTGCTGCATCTCCACAAGTTGAACATCGGGAGAAGGTACTGCTTGCCTGCCGTCATACCA
 TCATCCTGGTCGTTCTCAAATACAATTACGTCACGAGCAACACAGATATTCTTCTGATAGGATTGAACA

GGCGATATCCCTTAGAAGCATCAGAGTATCCAACCAAAATATTTTCTTCGCCCTTCGCAGAAAACCTCCT
TTTTGTGGACTTGTTAAGCACGATAGCCTTCGAACCAATATCTTCAGGTGGGACACATATGGCTTCCTT
M_A_S_F
CCAGTCCAAGACTCGAACGGTGTACATCAGATAATGCTGACGTCTCAGCGCGATTGCGTAGATAAGCAG
Q_S_K_T_R_T_V_L_H_Q_I_M_L_T_S_Q_R_D_C_V_D_K_Q
CAGTATTGATAGCTTCAGCCCATAGTGAGTGGCCGACGCCTGCGTGAATCATCATAGTCCTTGCCATCTC
Q_Y_*
GACCAGGGTTCGGTTAGCGCGCTCCGCTACACCGTTTTTGTGTGGGGTGTGCGGTACCATCAATTGTCTC
TTTATGCCGTTTTCTGTCAAGATACATTGAAACGCTTTACTTAAATACTCACGCCCGTTGTCGCTCCTTA
TTGCTTTCAACTTCTTTGGTTTTTCGGCAAATGCTTTAAACTCCTTGAATTTGGTCAGTATTTTCATCTCTC
GTCTTTATGAAATACACAAACATGTATCTCGATTTGTCTATCTACGAAAGTGACAAAATAGCGTGCACCTC
M_K_Y_T_N_M_Y_L_D_L_S_S_T_K_V_T_K_*
CAACAGACGTCGTTTTGCATTGGACCACAGATACACTATGCCTAGCTCCAGTACATCAATTGTGCGATT
TCCAGAGGCTTTTGGGAAACTCTTCACACATATCTTGCTTTTTTGGCGAGCTCAAACAGTCATTTTATTTT
GCGCCACTTAAATATTTTTTCTTTAGTCCGAGTACCATTTCTTGTGTTGATCATTTTGTTCAGCTTGCAA
AGTTCAAATGACCGAATCTCTGGTGCCACTTTCTCATTAATTTGTCTGTGCAACGAAACATGCACCTTC
M_T_E_S_L_V_P_L_S_H_* M_H_F_
AACTTTTCTTTTCAATTTTGAACATAAACAATCACTTTGCAGTTCAGCAATCAAACAACCTTTATTCTTA
Q_L_F_F_H_F_G_T_*
TCATTTTTTAAAATTCGCGCCTGTTTTTCGAACGACACACGAAATCCGTTTTCAATTGCCTTCGACACCG
ACATAAAGTTGCATTGTAATTTCTTTACAAAGAGTACGTGATCAATTCAAATGATGAGTTGCGCCAGGC
M_M_S_C_A_R_
TATTTTCACTGTACCACGACCGTCAGCCATAATATACTTATTGCCAACCAAAAGAATGCGCTCTTTATGC
L_F_S_L_Y_H_D_R_Q_P_* M_R_S_L_C_
M_
TCTCTGAAGCTTTTTAAACATCGATCTCTCACCGCACAAATGAGCCGTTGCTCCTGAATCAAGGCACCACA
S_L_K_L_L_N_I_D_L_S_P_H_K_*
L_S_E_A_F_K_H_R_S_L_T_A_Q_M_S_R_C_S_*
TATTTTTTGGCAATTCGCCAAGCTGTACAGTTGAGCAGAGCACAGCTTTGTGGCCGCTGCGACCACAGTT
CCAACAATTAAGTGTCTTTTCCCGTTAATCAGTTGCTGATTATTTACTTGCACACTCTTCTTAATGTCA
M_S_
CTTTTCGCTTTTTGTTTTCGCAGACCGAATTCCAAATGCACCTTTGTTTCGCTTTTTGCAGAATGTTCTTAAT
L_F_A_F_C_F_A_D_R_I_P_N_A_L_C_S_L_F_A_E_C_S_*
M_F_L_I
TTGCCATACGCCCTTTGCCCTCCTCTTGTAAATTTTATTTTTAACAATTTTCAATGTGGGCAACTCATCGCG
C_H_T_P_L_P_F_L_L_*
CGTTTCTATTGCAACTACGAACTTTCAAACGATTCGGGCAAACCTGGACAACAGGAGAATACTCAGGACT
TCTTCTTGATCACCAATTTGAATTTCTTTTAGTTTTTCAACAGCCGCACAAAATTTGTTCCATGACTCG
M_T_R_
AAACAACCTTCTGTTTCAGACATTTCTCATGTGCAACAACCTGCTTTAGCAAACATATACGTGCGCGGTTCC
N_N_F_C_F_R_H_S_H_V_Q_Q_L_L_*
TGAGGGAGTGTGGATTTCACTAAGTCTTTGCCATGCCCTCTACAGCAGTCTCACAGTTTTTTATATGGTTA
M_P_L_Q_Q_S_H_S_F_L_Y_G_*
M_V_
ATTTGGGATGTCTTAATGCATAGCATTATACTTGCTAATGCTTTTTTTCGCTCCTTCGCGTGAACAAAGCCT
N_L_G_C_L_N_A_* M_L_F_R_P_S_R_R_T_K_P_
TTAGCTCAGCACTATCACTTTTCGTTTTTAAACCACACGTCCACACACCACCCCATTAATTTCTGCATGAAT
L_A_Q_H_Y_H_F_R_F_* M_H_
TAAGACACTCTTCATTTGTACAGCCACGCACTGTAGTTTTTCATCATCAAGTTTTTCTATTTGGTTTTATT
F_Q_L_F_F_H_F_G_T_*
GATGAACCTCATTCCACTTCGATAAGTTTTTTCGTATTAATCCCTGAATCACAATATTTAAAGATGCCGA
M_P_
ATATTCTAATTCGTTTAGAACTATTCACTTTTATGTCCGCGTGGGCTGGGCCATAACCCCATTCGTTT
N_I_L_I_R_L_E_L_F_T_F_M_S_A_W_A_G_P_I_T_P_I_R_F
CAAGTATTTATTGTGACTGATCTAGAGTACAATGTTACTGCACGCAGGCAGCCAAATTACACGCTAAAGA
K_Y_L_L_* M_L_L_H_A_G_S_Q_I_T_R_*
GGATAACACAATAAGAGAAAAAGCTTTTGGTTACCAGGACCAAAAGAGGGAGACACAGTGCATTGACTT
TATTACATATATATATATATATGTCGGAGCGGACTCTTCGCATTCAGAAATTCCTTCTGTGATTTTCAT
M_S_E_R_D_S_S_H_S_R_N_S_F_C_D_F_I
ATTCTGATGAAATTTGAATATTTGCGGTAATAATTTCTGTGAGCGCTGCGGCAGAGGCACGAACAGCCTCTG
F_*

CAGCTGAGTGAGGTCCGATCGAGTGGTCATAAATAGTGTGTTAGAGAGAGATGA

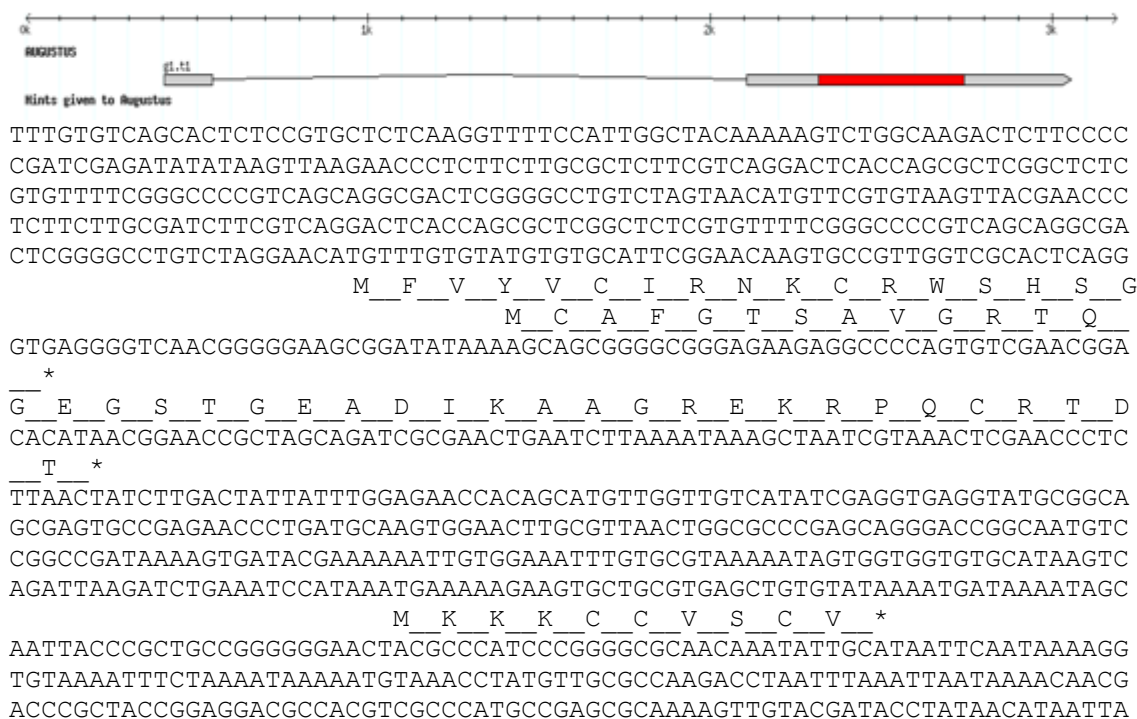
Para la región genómica 2R_39, el programa Augustus Gene Prediction predijo una estructura compuesta por 4 exones de 957 pares de bases el primero, 136 pares de bases el segundo, 957 pares de bases el tercero y 429 pares de bases el cuarto, y 3 intrones, el primero de 284 pares de bases, el segundo de 2689 pares de bases y el tercero de 284 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido STLTDNISILIPKSK con una probabilidad de $\rho=0.003$ y un valor $E=0.000847262945717522$; para el péptido MPNILIR con una probabilidad de $\rho=0.002$ y un valor $E=0.000568375656334029$; para el péptido SERDSSHSR con una probabilidad de $\rho=0.096$ y un valor $E=0.000505724310511363$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 60 sORFs. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 3 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_43



M_P_S_A_K_V_V_R_Y_L_*

AAACACGATCAACCCACTGCGGCGGTACGGCTTGTGGGAAAATTTTTTTTTTCTCTCCTTGCCAATTCGC
GAGTGCAAAAAGATTGTGTATAATAAACCAATAATTAACCATTCGAGCAGTTTACCTGCGGCAGTACGAGT
AATATGAGCGCCAGAGTGATAAGGTGGTGTGTGGCAGCTTGTGGATACGTTAAGTGGTGTGGAATGCA
M_H

CCCCAAAAAACC GCCCAACAAGTTGTGTGGCGGCCGTACCTTAGTAGGCAACCAGCCAAAAGGGATAC
P_K_K_N_R_P_T_S_C_V_A_A_V_P_*

TACGGAACCACCGTGCCAGTGCCGAAAATAAATTAGAGGTCATCAATAAAAACTGTAACAGCAGCAGC
CAAGGAAAAAATATTGCAAAATGGAATAGCGCACAAAAATTGTATAAACACATGCACAACACCACAATTC
M_H_N_T_T_I

AAAGGAAAAACAAGTATAGTACAATTTCCCAATTTCCCGAAACGAAATGAATAAAAAATTTCTCTACAATAC
Q_R_K_T_S_I_V_Q_F_P_Q_F_P_E_T_K_*

M_N_K_N_S_L_Q_Y

ACTAATTCACCTTTTTTTACTTATCCCAATAGAATATTCATGCTGTAGGGGTACAACCTAAACGACGAAA
T_N_S_P_F_F_T_Y_P_N_R_I_F_M_L_*

ACTAATAAAGAGCATACAAGGGTGAGTGAAATATTTTCATTAAACTTTATTGCCATATTTGCTAAATTTAG
AGAAATAAAGAAAAAGCAAAGAAGAACAGATATCTTTTTTATCGGGTTAAAACCGTTGTCTCACATTTTC
CGTAAAGTAATAACAGTTAAAACAGATCAGGAAAAATTTTCCACAAATCTGATCTAAGAAATTGAGACG
TGGCTTTCACATTTAAAAAGAAACGCGAAAAAGAAATAACGAAAGTAATAAAAGTACGTTGTGGCAGCTA
ATGAAATATTTCCCCCATGCATACCCATATATAAAAAAACATTAGTTAAAAAATAAAAAAAAAAAAAAAAAA
M_K_Y_S_P_H_A_Y_P_I_*

M_H_T_L_Y_K_K_N_I_S_*

AAAAAAAAAAAAAAAAAGAAATGAGTTAAGAAATACAAAAAGAAATACAAAAAACTATAAAAAAATAATA
TAAAAAATACAGATTATAAGAAATAAGAAATAAGAAATATAAAAAAATAAAAAATAAGTACACAAAAT
GTACCGTACCCCCACACACTACGTAGTCTTAGAACAACTTAGACGACCAGATATTTACGAATTGTCTTTT
TGTAAGCGGATTTCTGCATGCGGCGCAAATCCCGCTCACTGGACTGGCTGGGGTCGGCTTGGAAATGGG
M_G

TAGTGGATCTCCAGATGCTGCTGATTGGAACGCCGCTTGGCCGCGCAAGCGACGGCTTCGAGGAACTG
S_W_I_S_R_C_C_*

M_L_L_I_G_T_P_S_W_P_R_K_R_R_L_R_G_T

CAAAAACTGGAGGAGGCTAGCTGTATCCCTCGGCTACTGAAGTAACCAACGAGTGGTTAAGCAAGTCGAC
A_K_T_G_G_G_*

GATGGAATGCTCCTCTGACCAACTTCAACGGAACCTAAGAACGGCTGCAGAGAACTACGACCTGATCG
GCTCCTTATCATCCAATTCGACAATGAGACGATAATGGTCAACGGTCAAAAATTTCCAGTTACTCGGT
CAGTCATCTAATGGCGATGCCGCGGTGTTGAGCCACATAACGGCCAGCAACTTTCAACTTTCTCTGGAA
TACGTCCACGACGTGAGCATGAAGAATTTGGAAAAGATGTCCAACATGGCGAGTGAGCTACTAGCCTCTC
TTCTCACCGAGGCGGCACTCGCAATCTGCATATTCCTAGGCTTTTTATTTCCATGGAAGAAGCTGATGTC
CACCAAAGGCATGCCCCGATGTCGCGGAGATTGCCGCAAACCTTAGAAGCATTGGGCCAAACCGAGCTGAAC
AAGGCTCACTAATCTGCGGGACGCAGATCTTGAGGGGGGAGGAGTTAAGAACCCTCTTCTTGCGCTCTTC
GTCAGGACTCACCAGCGCTCGGCTCTCGTGTTTTCGGGCCCCGTCAGCAGGCGACTCGGGCCCTGTCTAG
GAACATGTTTGTGTATGTGTGCATTCGGAACAAGTGCCGTTGGTTCGCACTCAGGGTGAGGGTCAACGGG
M_F_V_Y_V_C_I_R_N_K_C_R_W_S_H_S_G_*

M_C_A_F_G_T_S_A_V_G_R_T_Q_G_E_G_S_T_G

GGAAGCGGATATAAAAAGCAGCGGGGAGAGAGGCCCCAGTGTGCAACGGACACATAACGGAACCGC
E_A_D_I_K_A_A_G_R_E_K_R_P_Q_C_R_T_D_T_*

TAGCAGATCGCGAACTGAATCTTAAAAATAAGCTAATCGTAAACTCGAACCCCTTAACTATCTTGACTA
TTATTTGGAGAACCACAGCATGTTGGTTGTTCATATCGAGGTGAGGTATGCGGCAGCGAGTGCCGAGAACC
M_R_Q_R_V_P_R_T

CTGATGCAAGTGGAACCTGCGTTAACTTATATGTATA
L_M_Q_V_E_L_A_L_T_Y_M_Y_A_*

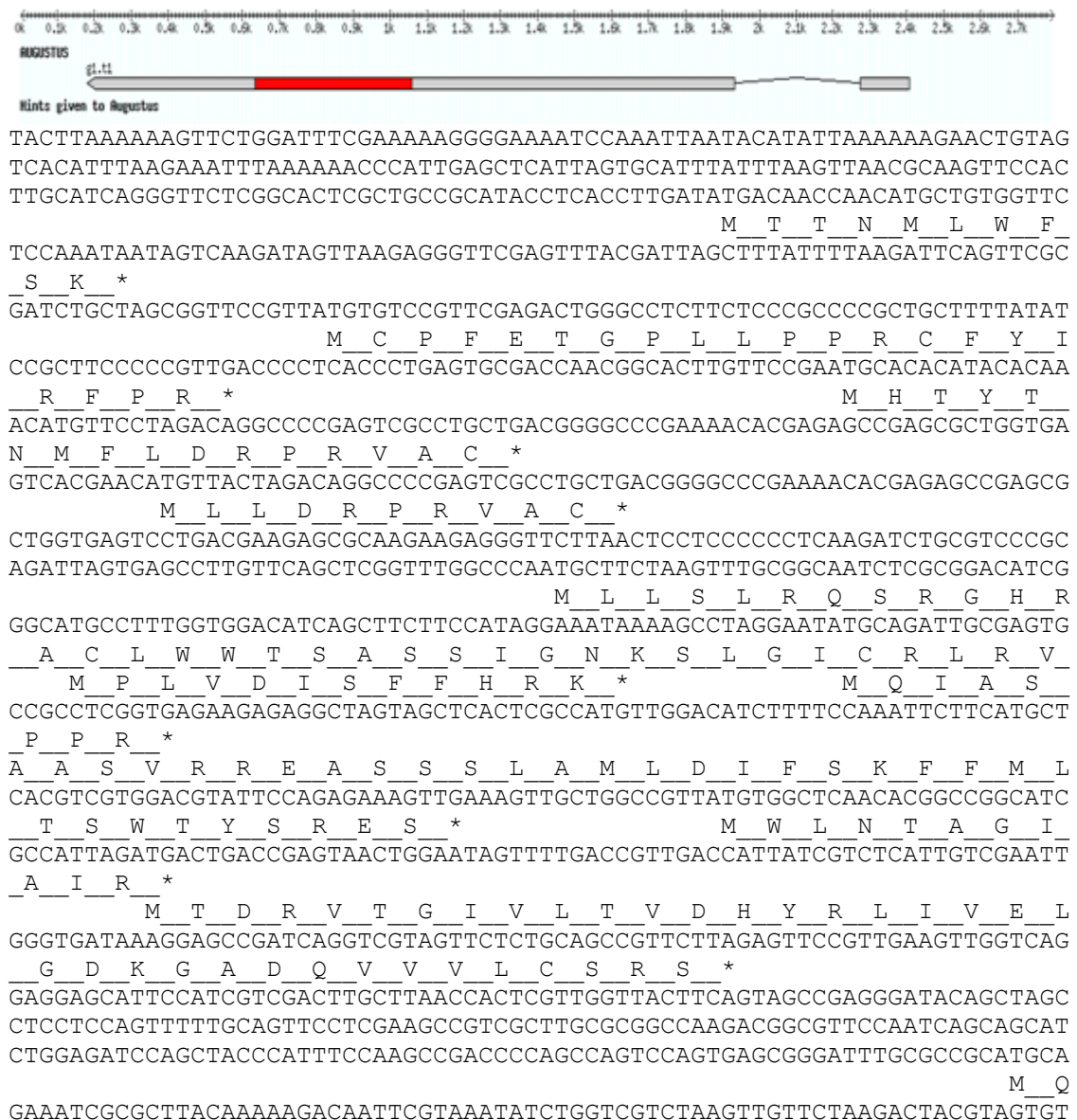
Para el gen 2R_43, el programa Augustus Gene Prediction predijo una estructura compuesta por 2 exones de 139 pares de bases el primero y 948 pares de bases el segundo, y 1 intrón de 1562 pareos de bases.

Se encontr acuerdo con espectros de masas para el péptido HNTTIQR con una probabilidad de $\rho=0.019$ y un valor $E=0.000167662151123091$; para el péptido NRPTSCVAAVP con una probabilidad de $\rho=0.081$ y un valor $E=0.000320259711550587$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 14 sORFs. Se puede observar una repetición desplazada. El sORF decimosegundo es una repetición desplazada del segundo sORF. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 2 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_54



K S R L Q K D N S *
 GTGGGGGTACGGTACATTTTGTGTACTTATATTTTTATTTATTTTCGTTTTCGGGAAATTGGGGAAATTGTA
 CTATACTTGTTTTTTCTTTTGAATTGTGGTGTGTGCATGTGTTTATACAATTTTTTGTGCGCTATTCCATTT
M C L Y N F C A L F H F
 TGCAATATTTTTTCTTTGCGTGCCTGCTGTTACAGTTTTTTTATGATGACCTCTAATTTATTTTCGGCACT
A I F F P C V R A V T V F Y *
 GGGCACGGTGGTTCCTAGTATCCCTTTTGGCTGGTTGCCTACTAAGGTACGGCCGCCACACAACCTTGT
 GGGCGGTTTTTTTTGGGTGCATTCCACACCACCTAACGTATCCAACAAGCTGCCACACACCACCTTATCA
 CTCTGGGCGCTCATATTACTCGTACTGCCGAGGTAACCTGCTGCAATGGTTAATTATTTGGTTTATTATA
M V N Y W F I I
 CACAATCTTTTGCCTCGCAATTGGCAAGGAGAGAAAAAAAAAATTTTCCACAAGCCGTACCGCCGCA
H N L L H S R I G K E R K K K F S H K P Y R R
 GTGGGTTGATCGTGTTTTAAATTATGTTATAGGTATCGTACAACTTTTGCGCTCGGCATGGGCGACGTGGC
S G L I V F * M G D V A
 GTCTCCGGTAGCGGGTCTGTTGTTTTATTAATTTAAATTAGGTCTTGGCGCAACATAGGTTTACATTTTT
S S G S G S L F Y *
 ATTTTAGAAATTTTACACCTTTTATTGAATTATGCAATATTTGTTGCGCCCCGGGATGGGCGTAGTTCCC
M Q Y L L R P G M G V V P
 CCCGGCAGCGGTAATTGCTATTTTATCATTTTATACACAGCTCACGCAGCACTTCTTTTTTCATTTATGG
P G S G *
 ATTTTCAGATCTTAATCTGACTTATGCACACCACCACCTATTTTTACGCACAATTTTCCACAATTTTTTCGT
 ATCACTTTTATCGGCCGGACATTGCCGGTCCCTGTTTCGGGCGCCAGTTAACGCAAGTTCCACTTGCATCA
 GGGTTCTCGGCACTCGCTGCCGCATACCTCACCTTGATATGACAACCAACATGCTGTGGTTCTCCAAATA
 ATAGTCAAGATAGTTAAGAGGGTTCGAGTTTACGATTAGCTTTATTTTAAAGATTAGTTTCGCGATCTGCT
 AGCGGTTCCGTTATGTGTCCGTTTCGAGACTGGGCCTCTTCTCCCGCCCCGCTGCTTTTATATCCGCTTCC
M C P F E T G P L L P P R C F Y I R F
 CCCGTTGACCCCTCACCTGAGTGCACCAACGGCACTTGTTCGAATGCACACATACACAACATGTTCC
P R * M H T Y T N M F
 CTAGACAGCCCCGAGTCGCTGCTGACGGGGCCCCGAAAACACGAGAGCCGAGCGCTGGTGAGTCACGAA
L D R P R V A C *
 CATGTTACTAGACAGGCCCGAGTCGCTGCTGACGGGGCCCCGAAAACACGAGAGCCGAGCGCTGGTGAG
M L L D R P R V A C *
 TCCTGACGAAGAGCGCAAGAAGAGGGTTCTTAACTTTTATGTTACGTTTCCATTTTTTTTTTATGTGTACA
M L R F H F F L C V Q
M C T
 GAATGGCATTGGGCCGAAACCACAGTTCGTGTTCCGCTGTTGGCCAGTGTATTTGCGTCCAGATGCA
N G I G P K P Q F V F R C W P V Y L R P D A *
E W H W A E T T V R V P L L A S V F A S R C A *
M A L G R N H S S C S A V G Q C I C V Q M C A
 *

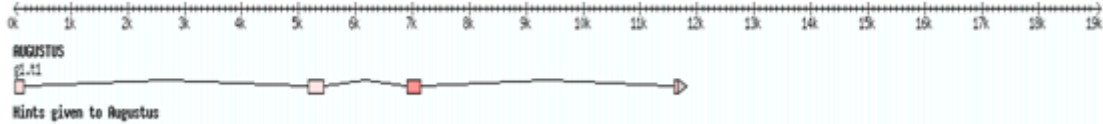
Para el gen 2R_54, el programa Augustus Gene Prediction predijo una estructura compuesta por 2 exones de 1752 pares de bases el primero y 132 pares de bases el segundo, y 1 intrón de 340 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido CPFETGPLLPPR con una probabilidad de $\rho=0.084$ y un valor $E=0.000609518336030885$; para el péptido SLGICRLR con una probabilidad de $\rho=0.008$ y un valor $E=0.000524605278366567$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 20 sORFs. Se puede observar una repetición desplazada. El sORF decimoséptimo es una duplicación desplazada del cuarto sORF. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 2 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_58



ATGCTCTATGAAACGTTCTAGAATGAAGGTCGAAACCGTTAGAGAGGTAAAACAAGATCCTCGGTCAGAA
M_K_R_S_R_M_K_V_E_T_V_R_E_V_K_Q_D_P_R_S_E
ATAAATTCGTTTTGATCTATGTGAATCGTATGAAAAGGTTATCACGGTTTTGGATTACAATGGAGTGTGCG
I_I_R_F_D_L_C_E_S *M_K_R_L_S_R_F_W_I_T_M_E_C_A
ATCTTGAATGGTTTTCGGTAAGTGTATATACACGGTAAAGAATTTCTACTTTTATTTTCATAGCT
I_L_N_G_F_R_Y *
M_V_F_G_T_E_C_Y_I_H_G_K_E_F_P_T_F_I_F_I_A
GGAAACCAATAGAATCTTTTCCAGATACTGTATTCAGTACAGATGAGCTGTGTTCAATTACTTTTCTAT
G_N_Q * M_S_C_V_Q_L_L_F_Y
CCATATTCTTTAAAATTTGTTTCATAACAGCATTAAATTTTAATTAAGAATTAATAATATGTTGATCAATA
P_Y_S_L_K_F_V_S * M_L_I_N
ATTTTCAATTCATATAACCTGTTTTTCTTAGTCGTCTACCGGTACAGGGCAAATCCGGTGTATA
N_F_Q_F_N_F_I *
ATCCGCTTGTAGTCTCCAGTTTTGTTTTGATTGTGACGACGCGAACTGCGCCATCAGAATAGGGGTGGA
ATTCCCAGGACCCCGGTGCTGTGCTGTCCCTCCGCTATGTTGCAGGCGGATGCCACTTGCTGCGCTGCTG
M_L_Q_A_D_A_H_L_L_R_C_C
TATTAAGTTAAGTATTCGGAGATCCATAGATCTCAGAAAAGATTATGACGACCCACTTAAAAAAAAC
I_K_V_K_Y_S_E_I_H_R_S_Q_K_R_L *
M_T_T_H_L_K_K_K
CTCAGTGTGCTCTTTAGTCCCTTGCTCACCAGGACTGCCAAGGTATGACTTCCGGTGAGCAGGATGTG
P_Q_C_A_L *
AGGCCAAGTCGCACTCCCTTCCCTTGAGTGTAGTATATGAGGGCTCCGCCTTCCACCCTTTAGTCGCCATT
TAAGATAAGTTAAGAACGATGAGGTAGTATTTCTCGCCCGCTAACTATAAAGGAGGTAAAGACGTGTTC
GCACTGGATCTTCCCTGGACTTTGCAGTTAATAGTTTAAACCGTTATAACTTCTGAGTTATCTACCACTCT
TGCGTAGATGGCGGCCCAAAGGCACGACAAGAAGCATCGGAGAAACCACACAGTTGCAAGTCATGAATT
M_A_A_P_K_A_R_Q_E_A_S_E_K_P_H_S_C_K_S *
M_N
TCACTATCTATCGACACTACGGCAGGGTAAATTGAAGTTTTTCACGGTATAGTTCCCACTGTTGGGCATAG
F_T_I_Y_R_H_Y_G_R_V_N *
CTGTGCATTTGCATTTGCCAGCCGATTCCTGTGATAACTTAGCCTTTCATGTTAACGACGTAAGTTCTGG
TTTATTTTCGTCAAATACCGTGGGCAGTTTCAAATCCTTGTGCAGGGTGGTGTTCACACATACCACTTGC
AATTTTTTCCGAGTCACAAGCCAAGCCCCAGATTTGGCACCCAGAGCAGAGGGCCGAGGACACTTCACTG
AGGGACTCCTACAGCGACTTCATGCTCAGTAAAAGAGCCGGTTATACCTAACTGCGAATCGCCTTCCTT
CAAGATATAATCCAAGCAATCCAAAATATGGAGTCTGTTTGATAATGACCAGGATTCAGTGTGCCACAT
TCGATCAAAAAGCTTCTCAAATTTCCATAAAGAGGGCGTTTCAATATTTTCATATTGTCGTATGAAGTTGCT
M_K_L_L
CAAGGTACTGTGGTCATCACGGAATCCGAACTGGTGTGCTCAGGAGTAATCCCTTGGGATTCCATGATCCTG
K_V_L_W_S_S_R_N_P_N_W_C_S_G_V_I_P_W_D_S_M_I_L
ACTATCTGACTACCTGACAGCAATTAGTCTCTCAAACACCTTTGAGATCGAAGGTAGAAGGAGGATCGGC
T_I *
CTGTAGAAAAGCGGGTTTTCGGCTTTCCAGGCTTGTGGATCATGGAAATTATCCCGAGCTTCTACTAAGAAG
GGAAGTATTACAGCCTGACAATACCAACGTTATGAAGATGATCGCACGACCTGGTAGTGCCTTTAAAGTG
M_K_M_I_A_R_P_G_S_A_F_K_V
GCATTGCAAATGACGTGCTCCCTTATGCTCTGTTGCAGTTCTAGAGTGCCATGACATCAGCTCACTCTC
A_L_Q_M_T_S_L_P_Y_A_L_L_Q_F *
M_L_C_C_S_S_R_V_P *
M_T_S_A_H_S

TGGGCTCAAACGGCGTGGTTGGAAGTCCATCTGAATTGCTTGTGTTCAGATGCTTATGAGTCTCTTCTAC
L_G_S_N_G_V_V_G_R_S_I_*
 TTGCTGCTGACTGGCAAGCTTACATGGCTGAAGTTCGCTCAGCGAGGGGAGCAGCGAATACATCCGCTTTC
M_A_E_V_A_Q_R_G_E_Q_R_I_H_P_L_S
 CTCAAGTCAGTCCGGCACCAAGCAATGTTGCTATCCATCAGTTGCGCCTTCCTCGTGCAACGCCTTTTGA
S_S_Q_S_G_T_K_Q_C_C_Y_P_S_V_A_P_S_S_C_N_A_F_*
M_L_L_S_I_S_C_A_F_L_V_Q_R_L_L
 TTTCCGAGTTATCGGTGACGGCACGAGCACGCCAGATATAGAACTCAATCCGGACAGAAAGCTACAAA
I_S_E_L_S_V_T_A_R_A_R_Q_I_*
 CTAGCAAGGTCAACAACCAGCAGAAAATCAGGATTCAGAAGAACTGAACGCAGGGGTATCGCAGCAGCAA
 CGAATTGTACTAGACTAAGTGAAGTGAAGAAAGACTGAAATAAAATTCACCTCTGATTCTGACCACTGA
 TACATTCAGTCCAACCAGAAATTCCTCCTCTACTCAGATACCTTTTATGCGAAATTAATCACTTTTCTGG
 GATATCATTAGCTGTTGGGAAAATAAACACAACTTTTTTACAAATGTGGTAGCAGTTTTGTGCGGTTA
 GTGTGCCGTTAGAATGGGTGTGCCAAAAAGATTTATCAATTTTACAAGATTTTACAATTTACAAGACTAA
M_G_V_P_K_R_F_I_N_F_T_R_F_Y_N_L_Q_D_*
 TAATATTATGAAAAAAGTGGGCGTGGTGGTTTTGGGCCACGAGTAACGGATATACAAACCGGGATAAA
M_K_K_K_W_A_W_W_F_W_A_T_S_N_G_Y_T_N_R_D_K
 TTTAAAAAAGTAAAATGAAAAACAAGAGACAACATTGCGGTTGAGCCTCAGACAACCACGACTTT
F_K_K_K_C_K_M_K_N_K_R_Q_H_C_G_*
 ACTAGTCTAGTCGGGTTGTTATGAAGAAATTTCCAATTTTCGTTTTTCATATCAATAACACTAATGGACGTT
M_K_K_F_P_I_S_F_S_Y_Q_* M_D_V
 AATACAGACAGACAGGGGAAATGCCTTCTGAACATCCGAGAAAAGCACCATGTTTCGATAACAAATAAAGT
N_T_D_R_Q_G_K_C_L_L_N_I_R_E_K_H_H_V_R_*
M_P_S_E_H_P_R_K_A_P_C_S_I_T_N_K_V
 TGGTGGAGCTGCGTTGGTGGAGTTGTGTGTGCCCGAATTACATAAACATTGAATTCATGAAAAATATCT
G_G_A_A_L_V_E_L_C_V_P_E_L_H_K_H_*
 CTAACTTCTAGCGCTCAAATAAAGATTTGACTAATCATCGAAAGCACTTGACCATTTACGACATCCC
 CACCGACGCTCGACGAGCTACTCTCTCCGGTGAACACCTCTTCTACGCCTTGTTCCTCTCTTTTGG
 GTATGGCCGATCACATATGAACACCAGTTCTCCTCGACGATTGGCTAGACGCAATTAAGATCCATTCC
M_A_G_S_H_M_N_T_S_S_P_R_R_L_A_R_R_N_*
 ACTTCTTGACATGCACTGTCTCGTGAAGTATGGTGACAACCCATACAGTCGGCTGCCCTTGACGGCC
M_H_C_P_R_R_S_M_V_T_T_H_T_V_G_C_P_*
 TTTCTTGCTGTCTGAGCAAACGCTCCTCTCTGCGATTGCCCTTATAAGTGGGCGCTCTGCCCTCAGTGC
 ACTCCCTCGTTCAGTCCGTGGTCCCTTTATCGGAACGTAGTCCACGACATCGTGAATAAATCTGTG
 CACCCATGCGGTGCTCCTCACCAACCGGTTATAGCTCGAAAATCTATGCAACGATATAAAGTTCGCCGGG
M_R_C_S_S_P_T_G_Y_S_S_K_I_Y_A_T_I_*
M_Q_R_Y_K_V_R_R
 ACCAGGACAAACTCACATTTTCATCTCCTCTTAATCTGACGTCGGGGGATTGCAGTCTTTGCGCGATAAAT
D_Q_D_K_L_T_F_H_L_L_L_I_*
 TTGTGGGGACCGTTTCATCTGTCTGTTGCTGGCTGCCACTACATTCATAGTACCTTCAAGTGGGTTTC
 AACTATCTAAGCACACTTGCCGAACAGGAAGTAAATCGGCAGAAATCAAGCCCAGCGTCTGCATGTATAT
M_Y_I
 ATCTCATCCGTTAAGAAACGGCGATAGCTTCTTCTTTACTGGAGAACTGTGACGCTAGTCTTCCATGTA
S_H_P_L_R_N_G_D_S_F_F_F_T_G_E_T_V_D_V_V_F_H_V
 GTGATACTCGTTGATCCGTGCTGCTTAACCTACAAAGCACTCTCCGTCTTCACATAATCGTCGAGCAAGG
V_I_L_V_D_P_C_C_L_T_T_K_H_S_P_S_S_H_N_R_R_A_R
 TGATCAGGCCGCTCCGAAAGTCATCATCTGCATCTCGTATATGTGGGCTCCAGACGATCGTCTCCTGTG
 * M_S_G_S_R_R_S_S_P_V
 CTCCATAAGAACCCTGGGAACCTCCCCTCTGGCTGCATCAATACTTGTAAAACAGATTATGTGATCTA
L_H_K_N_P_W_E_P_P_L_W_L_H_Q_Y_L_L_K_Q_I_M_*
 GTCGGGGTCTGACTTGGTTGGGCGCCAGATATGTTTTAATGATGAACGCGGCAAGGTAGTTGCCCGCG
 AACTCCTCCTCCAGCACCGTTGTTGCAAGGGGATGTCAACGTGGCACGTTGCTGCTCACGTAGTTGCTGA
 AGTGTGTTGCGGCGAGTTGCTATCCACAACACAGTCACCTCGTTGTCGCCTCTGTAGCTGTGGTGTGGTA
 CAGAGTGGGATCAGGACCAGGTGGTGGCCACGATCGGTGACTATAGCTGAAGTGACCAAGTGAATAAG
 CCGGTAGCAGGGTCTAGCTCCGACTTGATATTCACGGCAGTTGTTTTACGATAACGTCACCTTTATTTTCG
 GCTGTGAGTTCAGAACTAAATAATACAAAGAGAGAGCTTAAGATTACCTATCTCGCACAGATCAAGG
 CCGCCGACCAGCATCCGAAGTATAGCTTCGTGTGCTGTTGCTGTCACACGCAGCTTTTCTGATGTTAC
M_L
 TTGTTAACTTATATAATAACTTATACTTATTGGCTCCGATAATTTACCTAACATAGATCTTACGATTCTA
L_V_N_L_Y_N_N_L_Y_L_L_A_P_I_I_Y_L_T_*
 TTCACTCTCTACTTGCCTTTAGTTTAAAGGTACGCTGTAGCCACCTTTATATACTCTATGAAACGTT

CTAGAATGAAGGTCGAAACCGTTAGAGAGGTAACAAGATCCTCGGTCAGAAATAATTCGTTTTGATCT
 S_R_M_K_V_E_T_V_R_E_V_K_Q_D_P_R_S_E_I_I_R_F_D_L
 ATGTGAATCGTAGGAAAAGGTTATCACGGTTTTGGATTACAATGGAGTGTGCGATCTTGAATGGTTTTTCG
 C_E_S_* M_E_C_A_I_L_N_G_F_R
 GTACACGGTAAAGAATTTCTACTTTAATTTTCATAGCTGGAAACCAATAGAATCTTTTTCCAGGTACAG
 Y_T_V_K_N_F_L_L_*
 TATTTCAGTACAGATGAGCTCTGTTAAATTACTTTTTCTATCCATATTCTTTAAAATTTGTTTCATAACAGC
 M_S_S_V_K_L_L_F_Y_P_Y_S_L_K_F_V_S_*
 ATTACATTTTAATTAAGAATTACAATATGTTGATTCCACAGTCTTCGGGTGGAGACGTGTTTCTTTCAAG
 M_L_I_P_Q_S_S_G_G_D_V_F_L_S_S
 CTACGAATAGCAAGTTCTAAAACTACAACAGTATAGTGAAAGTTAAACACAAAGTGTAAGTGCAGTTT
 Y_E_*
 GCACAACCTAACAATTTATGACTATAGTAATTTACTAAAATAAATAATTTATCCATATTGTTCTGGTA
 ATTGTTATATGTGGACTTAGAACAATGAATCAAACGACATACGTTCTCAGCGACAATGTGAACAAGACG
 AGCGCCGGCTCTCTTTACAACGCAACAATGCATACTTTTTCTTTCTCGTCTCACCAGCAATCGGTGATCGAGC
 ACCCTCACCTTCACTAAGTCTGAACTTTTGCCTCAGCGAACGACAGACCGGCTTCTGCTCTCCCTCT
 CTGCCTGCTTCGGCTCACAAGTCTGGAGCGAAGAGACCGCCTCTCCTACCCCGCTCCTCTCGCAGCGCC
 AAACGACCGTCCCGGTAACGTAACTGCAATAACGAGTGCAGTGACCTCAGTGGCAACTGCCACAAC
 ATCAACTTCGTCAGCGGCCAATAATATCGCTGTGCCAGCTGTAAATAATTCAGCAGCACTGACCGTT
 TGCAACAACAATAATGCACGTAAGAAGAAATCAAACAAGCAGAAGTCGATTTGACTGTGCAGACTG
 GCATGGATCGCTACATCCAAATCAAGAGAAAGCTCAGCCCTCAAACAATAAGGCAGGTAATCAACCCAA
 AATCAATCGAACCAACAACGGCAATGAAAACCTCTGCAGTAAATAATTCAAACCGATATGCTATCTGGCT
 M_L_S_W_L
 GATTCTGCGACCGAACCAACCAACGAAAAACGGTAGGGGAACCAAAAAGACCAGGCCTCCACCAATTT
 I_L_R_P_N_N_P_T_K_K_R_*
 TCATACGAGAACAAGTACAATGCACTTGTAATAAATAAAGTCTGTTGCTTTGATTGGTGACAGCAAGTTCCA
 CATTATCCCACTTAAAAAAGGAAATATTCATGAAATAAATAAAGTCTCAGATCCAAACAGAAGCAGACCACCGT
 ATAGTGACTAAATACCTAAATGATGCTGGTAAAACTACTACACATACCAATTAATAAAGTTGCAAAGGGC
 M_M_L_V_K_T_T_T_H_T_N_*
 TACAGGTAGTACTTAAGGGCATTGAAGCAACAGTGACACCAGCTGAGATAATTGAGGCTCTGAAGGCCAA
 AAACTTTTCTGCAAAGACAGCTATTAATATTTTAAACAAGACAAAGTTCGCGAGCCACTATTCAAAATA
 GAACTCGAACCAGAGCTCCAGGCACTAAAGAAAAACGAAGTGCACCCAATATAACAATTTACAGTACTTGC
 TACATCGGAGGATCACCGTGGAGGAGCCGCACAAACGTATCAATCCAGTTCATGTACTAATTTGCCAAGA
 ATACGGCCACACCAAGGCATACTGCACCCTTAAGTCCGTATGTGTTGTCTGTAGCGAACCTCATACTACC
 GCAAACGCCCCAAAAACAAGGACGATAAGTCTGTGAAGAAATGCAGTAACTGCGGGGAAAAACATACTG
 CAAACTACAGAGGCTGTGTGGTGACAAAGAAATGAAGAGCCGCTAAACAACAGTATTGCCACAGCACA
 TACATACAACAAGTCAATTTCTACTCTCCGCAACCGATTTTTCAACCACCCCTAACTGTCCCAAGCACT
 ACTCCAACAATTTCTTTTCGCTAGCGCCCTAAAAATCCGGACTAGAAGTGCCCGCCCCACCGACAAGAAGT
 CTCATTCCGAACATACACCGACAAACATCCAACAAACACAACAAGTGGCATCGAAGCTATGATGCTATC
 CCTACAGCAAAGCATGAAAGACTTTATGACGTTTCATGCAAAATACTTTGCAAGAGCTCATGAAAAACCAA
 AATATCCTGATTCAACTTCTTGTATCTTCAAAATCCCCATAATGGCTTCCCTACGGATATCTCTGTGGAA
 CGAAATGGCGTTTTCAGGCATACACAAGAGCTCACACAGTTCATTTACGAAAAAACAATCGACGTAATG
 M_A_F_H_G_I_H_K_S_S_H_S_S_F_T_K_K_T_S_T_*
 CTACTATCAGAAACGCACCTCACAAATAAAAAACAATTTTCATATACCAGGATACTTGTTCATGGTACAA
 M_V_Q
 ATCATCCAGATGGTAAAGCTCATGGAGGCACTGGAATACTCATCAGAAATCGCATAAAAACACCACCACTT
 I_I_Q_M_V_K_L_M_E_A_L_E_Y_S_S_E_I_A_*
 AAACAATTTTGACAAAAACTACTTACAATCTACGTCCATAGCCTTACAACCTACAACATGGTTCAACGACT
 CTAGCCGAGTCTACTGCCACCGCGCTTTCCAATCTCTGAGGATCAATTCATGGAATTTCTTTAACACAC
 TAGGTGACAGGTTTCATCGCAGCGGGTACTATAACGCCAAGCACACCCATTTGGGGATCTCGACTTGTGTC
 GCCAAAGGGTAAGCAATTTGTAATGCGCTTACGAAGCCAGAAAACAAGCTAGACTATGTATCCCCGGGT
 M_R_L_R_S_Q_K_T_S_* M_Y_P_R_V
 AAGCCTACATACTGGCCAGCAGACCCAAAGAAAAATCCAGACCTGATCGATTTTGAATTTACTAAACATG
 S_L_H_T_G_Q_Q_T_Q_E_K_S_Q_T_* M
 TCCCCCGCAACATGGTCAACCGCGAAGCACTAGCAGATTTATCATCAGATCACTCACCTGTTTTTCTAAA
 S_P_A_T_W_S_P_P_K_H_*
 TATGCTAACTCGCCCCACATCGTTCGACCCACCGTATAGACTCACAAATTTTAGAACAACCTGGCCAAGG
 TATCAAAAAGTATGTCTGTTTACACATAGAATAACGACGGCATTATCTACAAAGGAGGATATAGACAAGT
 CAACGGAACTCTTGAACAATTTTAGTTTTCGGCTGCAAAGGCTTCAACCCCGCAGTGACGTATGCAAA
 M_Q

ACCAAACACTACATCAAAACTAATCGCGAAATCGAGCGGCTGGTATTAGATAAACGACGCCTACGAAGGGAT
N_Q_T_T_S_K_L_I_A_K_S_S_G_W_Y_*
TGGCAGTCTAATAGATCACCAATTACTAAGCACATGCTTAAGATAGCCACACGCAGGCTTACCAATGCTC
M_L
TCAAACAAGAGGAAAAAACAGCCAACGTTTCATATATCGAGCAACTCTCTCCCACCAGCACTAAGTACCC
S_N_K_R_K_K_T_A_N_V_H_I_S_S_N_S_L_P_P_A_L_S_T
TCTTTGGAGAGCTCACAGAAACCTAAAGACTCCAATAGCGCCAATTATGCCACTACGAAGTCCCTCTGGC
L_F_G_E_L_T_E_T_*

ACCTGGTTTCGAAGTGATGAAGAAAGAGCCAATGCTTTTCGCTGACCATTACAAAATGTATTCCGACCAA
M_K_K_E_P_M_L_S_L_T_I_Y_K_M_Y_S_D_Q
ACCCCTCTACCAACACATTTATTTCTCCCTCCTTTAATAGCAGCCAATCTAGATCCTCAAGAACCCTTTGA
T_P_L_P_T_H_L_F_S_L_L_*

ATTCCGACCATGTGAACTAGCAAAGGTTATCAAAGAGCAACTGAACCCAAGAAAATCGCCTGGCTACGAC
CTAATAACTCCAAGAAATGCTCATTGAACTCCCAAAGTGTGCTATTCTTCACATCTGCCTGTTGTTCAACG
CAATCGCCAAGCTTGGATACTTCCCTCAAAAATGGAAAAAGTCGACCATAGTAATGATTCCAAAGCCAGG
M_E_K_V_D_H_S_N_D_S_K_A_R
AAAAGATA
K_R_*

Para el gen 2R_58, el programa Augustus Gene Prediction predijo una estructura compuesta por 4 exones de 155 pares de bases el primero, 270 pares de bases el segundo, 236 pares de bases el tercero, y 226 pares de bases el cuarto, y 3 intrones, de 4998 pares de bases el primero, 1455 pares de bases el segundo, y 4461 pares de bases el tercero.

Se encontró acuerdo con espectros de masas para el péptido NPNWCSGVIPWDSMILTIcon con una probabilidad de $\rho=0.088$ y un valor $E=0.000421188502997275$; para el péptido VALQMTSLPYALLQFcon con una probabilidad de $\rho=0.015$ y un valor $E=8.00954401243523e-05$; para el péptido QEASEKPHSCKcon con una probabilidad de $\rho=0.082$ y un valor $E=0.000341167814746965$; para el péptido CLLNIREKcon con una probabilidad de $\rho=0.085$ y un valor $E=0.000236707608832088$; para el péptido ECAILNGFRcon con una probabilidad de $\rho=0.009$ y un valor $E=0.000872799071745708$; para el péptido FWITMECAILNGFRYTVKcon con una probabilidad de $\rho=0.022$ y un valor $E=9.13024638785673e-05$; para el péptido MVQIIQMVKcon con una probabilidad de $\rho=0.063$ y un valor $E=0.000842331872273208$; para el péptido FVLLLNIENTAQPFcon con una probabilidad de $\rho=0.046$ y un valor $E=6.78997988356888e-05$; para el péptido MLALHLVWSSKcon con una probabilidad de $\rho=0.001$ y un valor $E=0.000686304066554346$; para el péptido SFSNFQKEGVAIFHNVCSSYCGHHGTRcon con una probabilidad de $\rho=0.019$ y un valor $E=0.00048312585066518$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 42 sORFs. Se pueden observar repeticiones. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 10 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura,

pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_76



Hints given to Augustus

TCCGTGCTCTCAAGGTTTTCCATTGGCTACAAAAAGTCTGGCAAGACTCTTCCCCGATCGAGATATATA
AGTTAAGAACCCTCTTCTTTCGCTCTTCGTCAGGACTCACCAGCGCTCGGCTCTCGTGTTCGGGCCCC
GTCAGCAGGCGACTCGGGCCTGTCTAGTAACATGTTTCGTGTAAGTTACGAACCCTCTTCTTTCGATCTT
CGTCAGGACTCACCAGCGCTCGGCTCTCGTGTTCGGGCCCCGTCAGCAGGCGACTCGGGCCTGTCTA
GGAACATGTTTGTGTATGTGTGCATTTCGGAACAAGTGCCGTTGGTTCGCACTCAGGGTGAGGGTCAACGG
M_F_V_Y_V_C_I_R_N_K_C_R_W_S_H_S_G_*
M_C_A_F_G_T_S_A_V_G_R_T_Q_G_E_G_S_T
GGGAAGCGGATATAAAAAGCAGCGGGGCGGGAGAAGAGGCCCCAGTGTGCAACGGACACATAACGGAACCG
G_E_A_D_I_K_A_A_G_R_E_K_R_P_Q_C_R_T_D_T_*
CTAGCAGATCGCGAACTGAATCTTAAAAATAAGCTAATCGTAAACTCGAACCCTCTTAACTATCTTGACT
ATTATTTGGAGAACCACAGCATGTTGGTTGTCATATCGAGGTGAGGTATGCGGCAGCGAGTGCCGAGAAC
CCTGATGCAAGTGGAACTTTCGTTAACTGGCGCCCAGCAGGGACCGCAATGTCCGGCCGATAAAAAGTG
ATACGAAAAAATTTGTGAAAATTTGTGCGTAAAAATAGTGGTGGTGTGCATAAGTCAGATTAAGATCTGAA
ATCCATAAATGAAAAAGAAGTGTGCGTGAGCTGTGTATAAAATGATAAAATAGCAATTACCCGCTGCCG
M_K_K_K_C_C_V_S_C_V_*
GGGGAACTACGCCCATCCCCGGGGCGCAACAAATATTGCATAATTCAATAAAAGGTGTAATAATTTCTAAA
ATAAAAAATGTAAACCTATGTTGCGCCAAGACCTAATTTAAATTAATAAAAACAACGCCCGCTACCCGAGG
ACGCCACGTGCCCCATGCCGAGCGCAAAAGTTGTACGATACCTATAACATAATTAACACGATCAACCC
M_P_S_A_K_V_V_R_Y_L_*
ACTGCGGCGGTACGGCTTGTGGGAAAAATTTTTTTTTCTCTCCTTGCCAATTCGCGAGTGCAAAAGATTG
TGTATAATAAACCAATAATTAACCATTCGAGCAGTTTACCTGCGGCAGTACGAGTAATATGAGCGCCAG
AGTGATAAGGTGGTGTGTGGCAGCTTGTGGATACGTTAAGTGGTGTGGAATGCACCCAAAAAAAACCG
M_H_P_K_K_N_R
CCCAACAAGTTGTGTGGCGCCGTACCTTAGTAGGCAACCAGCCAAAAGGGATACTACGGAACCACCGTG
P_T_S_C_V_A_A_V_P_*
CCCAGTGGCGAAATAAATTAGAGGTCATCAATAAAAAACTGTAACAGCAGCAGCAAGGAAAAAATATT
GCAAAATGGAATAGCGCACAAAAATTTGTATAAACACATGCACAACACCACAATTCAAAGGAAAAACAAGTA
M_H_N_T_T_I_Q_R_K_T_S
TAGTACAATTTCCCAATTTCCCGAAACGAAATGAATAAAAAATTTCTTACAATACCTAATTCACCTTTT
I_V_Q_F_P_Q_F_P_E_T_K_*
M_N_K_N_S_L_Q_Y_T_N_S_P_F
TTTACTTATCCCAATAGAATATTCATGCTGTAGGGGTACAACCTAAACGACGAAAACCTAATAAAGAGCAT
F_T_Y_P_N_R_I_F_M_L_*
ACAAGGGTGAGTGAAATATTTTCATTAACCTTTATTGCCATATTTGCTAAATTTAGAGAAATAAAGAAAA
GCAAAAGAAGAACAGATATTTCTTTTATCGGGTTAAAACCGTTGTCTCACATTTCCGTAAAGTAATAACA
GTTAAAACAGATCAGGAAAAATTTCCACAAATCTGATCTAAGAAATTTGAGACGTGGCTTTCACATTTA
AAAAAGAAACGCGAAAAAGAATAACGAAAGTAATAAAAAGTACGTTGTGGCAGCTAATGAAATATTTCCCC
M_K_Y_S_P
CATGCATACCCTATATAAAAAAACATTAGTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
H_A_Y_P_I_*
M_H_T_L_Y_K_K_N_I_S_*
GAATGAGTTAAGAAATACAAAAAGAAATACAAAAAACTATAAAAAAATAATATAAAAAAATACAGAT
TATAAGAAATAAGAAATAAGAAATATAAAAAAATAAAAAATAAAGTACACAAAATGTACCGTACCCCCAC
ACACTACGTAGTCTTAGAACAACTTAGACGACCAGATATTTACGAATTTGTCTTTTTTGTAAAGCGGATTTT
TGCATGCGGCGCAAATCCCGCTCACTGGACTGGCTGGGGTTCGGCTTGGAAATGGGTAGCTGGATCTCCAG
M_G_S_W_I_S_R
ATGCTGCTGATTGGAACGCCGCTTTGGCCGCGCAAGCGACGGCTTCGAGGAACTGCAAAAACCTGGAGGAG
C_C_*
M_L_L_I_G_T_P_S_W_P_R_K_R_R_L_R_G_T_A_K_T_G_G
GCTAGCTGTATCCCTCGGCTACTGAAGTAACCAACGAGTGGTTAAGCAAGTCGACGATGGAATGCTCCTC

G__*
 CTGACCAACTTCAACGGAACTCTAAGAACGGCTGCAGAGAACTACGACCTGATCGGCTCCTTTATCATCC
 AATTCGACAATGAGACGATAATGGTCAACGGTCAAAACTATTCCAGTTACTCGGTCAGTCATCTAATGGC
 GATGCCGGCCGTGTTGAGCCACATAACGGCCAGCAACTTTCAACTTTCTCTGGAATACGTCCACGACGTG
 AGCATGAAGAATTTGGAAAAGATGTCCAACATGGCGAGTGAGCTACTAGCCTCTCTTCTCACCGAGGCGG
 CACTCGCAATCTGCATATTCCTAGGCTTTTATTTCCATATGGAAGAAGCTGATGTCCACCAAAGGCATGCC
 CGATGTCCGCGAGATTGCCGCAAACCTTAGAAGCATTTGGGCCAAACCGAGCTGAACAAGGCTCACTAATCT
 GCGGGACGCAGATCTTGAGGGGGGAGGAGTTAAGAACCCTCTTCTTGCCTCTTCGTGAGGACTCACCAG
 CGCTCGGCTCTCGTGTTCGGGCCCGTCAGCAGGCGACTCGGGCCCTGTCTAGGAACATGTTTGTGTA
 M__F__V__Y
 TGTTGTCATTTCGGAACAAGTGCCGTTGGTTCGCACTCAGGGTGAGGGGTCAACGGGGGAAGCGGATATAAA
 V__C__I__R__N__K__C__R__W__S__H__S__G__*
 M__C__A__F__G__T__S__A__V__G__R__T__Q__G__E__G__S__T__G__E__A__D__I__K
 AGCAGCGGGGCGGGAGAAAGAGGCCCCAGTGTCGAACGGACACATAACGGAACCGCTAGCAGATCGCGAAC
 A__A__G__R__E__K__R__P__Q__C__R__T__D__T__*
 TGAATCTTAAAATAAAGCTAATCGTAAACTCGAACCCCTCTTAACTATCTTGACTATTATTTGGAGAACCA
 CAGCATGTTGGTTGTCATATCGAGGTGAGGTATGCGGCAGCGAGTGCCGAGAACCCTGATGCAAGTGGAA
 M__R__Q__R__V__P__R__T__L__M__Q__V__E__
 CTTGCGTTAACTTATATGTATATAC
 L__A__L__T__Y__M__Y__I__C__*

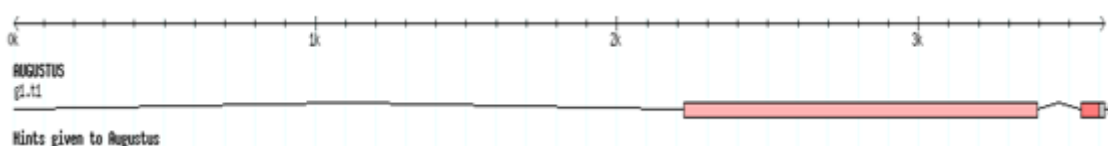
Para el gen 2R_76, el programa Augustus Gene Prediction predijo una estructura compuesta por 2 exones de 58 pares de bases el primero, y 948 pares de bases el segundo, y 1 intrón de 546 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido HNTTIQR con una probabilidad de $\rho=0.019$ y un valor $E=0.000167662151123091$; para el péptido NRPTSCVAAVP con una probabilidad de $\rho=0.081$ y un valor $E=0.000320259711550587$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 14 sORFs. Se puede observar una repetición desplazada. Los sORFs decimosegundo y decimotercero son repeticiones desplazadas de los sORFs primero y segundo. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 2 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_86



GCATTCCCTTTTAAACATATACCAAAGTTTAAATTGTTTCCCTTCGGCAAATCCAAACCACAAATAAGGCT
TGAAAGCGAGATACGATTCGGTGATCCAGGAATATCTCGACTTAAAGCACATGCGACAAGTTCTTCCCTAC
CCATGATTGCAACGCTATAATATGCCACATCACGCCGTCTTAAAACCGGAAAGCGTAACACTACTACACGG
M I A T P I I C H I T P S *
AGTTTAGTATTCAATGCCCTCCAGCCCTTCATCGAATGGTACCAGCTTAAATGATATCCTTCATGCTGGGC
M P P A L H R M V P A * M I S F M L G
CTGTCTTACAGTCCGACTTGACCATTCAAATTTCTGAAGCAACAACGTGTGATGCCGGCCACGGCAAGTAA
L S Y S P T *
AACAAATTGTGCGTGCTTTTGCACCTCACGAAGCTGATGTCCCTTTGCAAAGCAGTTTAAAGCATAACTGCTT
CCGTTAAATGTAGGCTGACCGGTCTCAACCGACATTTGGAGAAAACGCGGACATACACGGACAGGATGG
M
TTTTCTTGTGTCACAAGTCGCAACTTTTTCGATTTTGGAGCCACTTTCGTCTCATAGGAATTTAGTTTTCT
V F L V A Q V A T F R F W S H F R L I G I *
TAGAGGGCCCACTTGAGTTCACCGCTTGGAGTGCAGCTGACTTGGTACGGACTGTCTCACATCATCGAT
GGCCTCTAGGGTTCGATGACGTTCTGTGAGGAAGGTGTTCAAGTCTCCCCATGTCGGGATGTCGGCTTTC
M T F C E E G V Q V S P C R D V G F
M S G C R L S
M S A F
TTATGTAGCGACTGCGAGTTATCTTCGGGAGCTTGGATGAACACAGATATACCAGCAGGCAGTCCCAGTT
L M *
Y V A T A S Y L R E L G *
L C S D C E L S S G A W M N T D I P A G S P S
CTCAGTGTGATGCCTGACATTTCTAAGGCAGTCAAGCTACCTTGAACAGTACTTTGCAGTACCTTCAAG
S Q C * M P D I S K A V K L P *
GCCGCCCCAGATTCTGTGGTATCGACTGCACATTAACAGTATTTTCAATTGACTGTTTACCAACAATC
GTTTTATTTTGCAAAACGCTCTATTAGGTTTTTACACGCAGAGCGGAAACCCCTCGTTGGTGAGAGGCGAAAT
CGAAACTATGGCATGCACGTCGCCACTTGTTTTGGCATTAAATGGAATAACTATTCAACTGGAGTTAGC
M A C T S P L V L A F K W N N Y S T G V S
M H V A T C F G I *
CGTGGATTATTGATATAAATGGCTGTGAAAAGATCCCGGAAAGTCCGCAAGATAGTCGCCTGCGA
R G L L I * M A V K R S R K V G Q R R *
AAACTTCTGTATCGCATGGAGGCAACCGACAGCCAGAGGAAATGTAGGCTTGGGGCGCAGCGTTTCGCGGC
M E A T D S Q R K C R L G A Q R S R
TGGGATGGACTGAGACGTGCCCTGCTCGATTTTATCAACGAGCTGGGCAACACACCTTTCATAGACTGAA
L G W T E T C P A R F Y Q R A G Q H T F H R L N
TAGCAATAACTGTATTTAGCCCTGAGAATAGGCATGTTGCCTGCTGCCGCTTCGCCTGCTGACACAAGGC
S N N C I *
ACTCTGAGCAGAGGTGCAATTCCTTTTTCAACCTTGTCCCATAGGGTTCGGACTTGGTCGCGACGGACGCT
AATCATCGTGACGGTTGGAGCTGCGGATTCGGAGTGTGATCTGAGCCTCAAATTCGCTTAAAGCGGGCA
GAAACCGAAATGAATTTGGCTAGCGCTAGATCGGAAGCAGCCATATTTTTCGCTGTGCGCTGTACTGTGG
M N L A S A R S E A A I F F A V R C T V
CTTTGGATGGAGTAGCACAGTCTCGGAAATCTGCGCAGGCGTTTTTACCAGCATGCTTGGATTCTTGG
A L D G V A Q S S E I C A G V F T D M L E I L G
GCCCTTGTGGTGAAAAAATAGACCTGCGTTTGTCTGCGGACAATTTCTTCTTATCGTCCCCGATGGGCAT
P L W * M G I
AATCGAAGAAATACGAAATGAAAGGGGAGCGTTTTTGGAGCCCGGTCACACGTTTTGAAAAGTAGACAGA
I E E I R N E R E A V L E P G H T F E K V D R
M K G K R F W S P V T R L K K *
TTCTCAGAATGCACTTAATAAATTTAAGCTTGCCTGCGGCGTGAACGGTATAAAGCAGCGGAAAAAGACCGT
F L R M H L I N L S L P P *
ATAGGATGTAATGGGTGAGAAGGTGGTGTAATAATTTACGAAGTGGAAACGCCGTAGTTTACTACTAATTGG
ACAGGTGACTCGGAGTGAACAGCGAATTTATATAAATAATATAAATAAATAAATAAATAAATAAATAAATAA
AAAAAATAAATTTAATAAATAATGAAAAAGTTATTTGTTAAAAAAGAGAAAAAACACTTTTAATTT
ATTGCTCGGAAATTAATTTGGGAAATTTACGGAAAAATTTGATGGAAAAACAAACGCCTTTTCCGCGTCCG
M E N K T P F P R R
CACTTGAATAAATTTTATTTGTGGTTTGGATTTCCCGCCACACGTTGTTGCACCGAAACAACCTTTTT
H L E *
CCAGCAATTCAAGCCCTTCAAATCCTGCAAGGCCAATTTCCGCTAATCAGGCCAATTTTCGTTCCAAATGA
M
GCAAGCCGGTGTTCAAAATTTATTTAACCCACGGGCTCAAGAGCTATCCTTCTTGGCACTGCCATAATCAAT
S K P V F K I I *

ATTTCCCATCTTGGCACTAACTTTAAGGCACGCGCCCTGATCGACTCCGGATCAGAGGCGACATTCATAA
 CCGAGCGACTGTTCAATCTAATTAGATTGCCATTCCAGGTGGTTCAAGTCCAGGTCTCGGGCTTAAACCA
 AACAGTAGCTGCTCAGTCCAAGAAGCTCTGCAGTTTCACCATCCGATCTCCGACTAGGCCCGCGTTGCAG
 TTGGAGACGACGGCCTATGTCTCCCTCAACTAGCCGGAAATCTGCCTTCCCTACCCAATTCGCAAAATT
 TCCTTCGGGATCTTCCCGATTTTCCACTGGCGGATCCAAAATCTATGAGAGCGCACAAATAGATGTACT
M_Y
 TATCGGAGCCGACATCCTGCCTTCGGTGTCTTCTGAGTGGAGCAAAAACCAACATCTGTGGCTCTCTTTG
 L_S_E_P_T_S_C_L_R_C_F_*
 GGGCAAGAGACCATTTTCGGCTGGGTACTAACTGGGCCAGTGTGAGCCTCAGCCCAAAGCAAAATTTCT
 CTTTTTCGACACAGATCTCCACGCGTACGATAAATTCAGTGGACAACTCCTCACAAAATTTTGGGAGGT
 GGAGGATATAACCAACAAAGTTGGTAAAAGAATCCGATTCCATGTGCGAGAAGAATTTCTTCAAACGACC
 ACGAGAAAACGAGTGC GGCAAATATGTCGTTACTCTGCCTTTTCGCGACCCCGAACATATCGGTTCCGGGC
M_S_L_L_C_L_F_A_T_P_N_I_S_V_P_G
 TAGGGCATTCTAGGTCTTTCGCGTTGGCTCAGTTCCTTAAGAAATGAGCAGCGTCTAAAAAGAGATGAGGC
 *
 CTTGAAAGCGAGATACGATTCGGTGTATCCAGGAATATCTCGACTTAAAGCACATGCGACAAGTTCTTCT
 ACGCATGATTGCAACGCTATTATATGCCACATCAGCCGTCTTAAAACCGGAGAGCGTAACTACTAAAC
M_I_A_T_P_I_I_C_H_I_T_P_S_*
 TCCGTGTAGTATTCAATGCCCTCAGCCATCATCATCGAATGGTACCAGTTTAAATGATATCCTTCATGC
M_P_P_A_H_H_H_R_M_V_P_V_* M_I_S_F_M
 TGGCCCTGTCTTACAGTCCGACTTGACCAATCAAATCTGAAGTGGCGCTATTTCCGATACGTGTTCAAC
 L_A_L_S_Y_S_P_T_*
 GCCGATATCGAGAAAATGTATCGGCAGATCTGGGTAGGTCCGAAACACACTCCATTCCAGCGAATACTTT
 TCCGTAACAATAGAGGGGAAATCAGAGATTTGAAATGAAAACAGTAACCTTTGGAGTCAAAGCTAGCA
 ATGTCATTCGAAATTTGATGATGTGGATGATGTTTTAGCCGGAGCGGACTCCACGGAAGAAGCGAAGCT
M_S_F_E_I_S_C_M_W_M_M_F_*
M_Y_V_D_D_V_L_A_G_A_D_S_T_E_E_A_K_L
 CATGGTGCAAGAGCTCCGAGACGCTCTGAATTCGCGGATTTCCA
M_V_Q_E_L_R_D_A_L_N_S_A_G_F_P_*

Para la región genómica 2R_86, el programa Augustus Gene Prediction predijo una estructura compuesta por 2 exones de 1172 pares de bases el primero, y 76 pares de bases el segundo, y 1 intrón de 147pares de bases.

Se encontró acuerdo con espectros de masas para el péptido MGIIEEIR con una probabilidad de $\rho=0.084$ y un valor $E=2.97470105064206e-07$; para el péptido YVDDVLAGADSTEEAKLMVQELR con una probabilidad de $\rho=0.093$ y un valor $E=5.81298966985286e-05$; para el péptido MYVDDVLAGADSTEEAK con una probabilidad de $\rho=0.023$ y un valor $E=7.71602349819706e-05$; para el péptido TFCEEGVQVSPCR con una probabilidad de $\rho=0.068$ y un valor $E=2.88417332453133e-05$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 23 sORFs. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 4 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_88



ACTCAGCTGCAGAGGCTGTTTCGTGCCCTCTGCCGACGCTCGACAGAATTTTACCGCAATATTACAATTT
CATCAGAATATGAAATCACAGAAGGAATTTCTGGAATGCGAAGAGTTCGCGCTCCGACATATATATAATAA
M_K_S_Q_K_E_F_L_E_C_E_E_S_R_S_D_I_Y_I_I
M_R_R_V_A_L_R_H_I_Y_N_N
TATGTAATAAAGTCAATGCACTGTGTCTCCCTCTTTTGGTTCGCGGTAACCAAAGCTTTTTTCTCTTATT
I_C_N_K_V_N_A_L_C_L_P_L_L_V_A_V_T_K_S_F_F_L_L_L
M_* M_H_C_V_S_L_F_W_S_R_*
GTGTTATCCTCTTTAGCGTGAATTTGGCTGCCTGCGTGCAGTAACATTGTACTCTAGATCAGTCACAAT
C_Y_P_L_*
AAATACTTGAAACGAATGGGGTTATGGGCCAGCCACGCGGACATAAAAGTGAATAGTTCATAACGAA
M_G_V_M_G_P_A_H_A_D_I_K_V_N_S_S_K_R
TTAGAATATTCGGCATCTTAAATATTTGTGATTCAGGGATTTAATACGAAAACTTATCGAAGTGAAAT
I_R_I_F_G_I_F_K_Y_C_D_S_G_I_*
GAGTTCATCAATAAACCAATAGAAAACTTGATGATGAAAACTACAGTGCCTGGGCTGTACAAATGAAG
M_M_K_T_T_V_R_G_L_Y_K_*
AGTGTCTTAATTCATGCAGAATTATGGGGGGTGGTGTGTGGACGTGTGGTTAAAAACGAAAGTGATAGTG
M_Q_N_Y_G_G_W_C_V_D_V_W_L_K_T_K_V_I_V
M_G_G_G_V_W_T_C_G_*
CTGAGCTAAAGGCTTTGTTCGACGCGAAGGACGAAAAAGCATTAGCAAGTATAATGCTATGCATTAAGAC
L_S_*
ATCCCAATTAACCATATAAAAACTGTGAGACTGCTGTAGAGGCATGGCAAAGACTTAGTGAAATCCAC
ACTCCCTCAGGACCGGCACGACGTATATGTTTGCATAAGCAGTTGTTGCACATGAGAATGTCTGAAACAG
M_F_A_K_A_V_V_A_H_E_N_V_*
AAGTTGTTTCGAGTCATGTGAACAATTTTGTGCGGCTGTTGAAAACTAAAAGAAATTC AATTGGTGAT
CCAAGAAGAAGTCTTGAGTATTCTCCGTGTGTCAGTTTGC CGGAATCGTTTGAAAGTTTCGTAGTTGCA
ATAGAAACGCGCGATGAGTTGCCACATTTGAAAATGTTAAAAATAAAATTACAAGAGGAAGGGCAAAGGC
GTATGGCAAAATTAAGAACATTTCTGCAAAAAGCGAACAAAGTGCATTTGGAATTCGGTCTGCGAAACAAAA
AGCGAAAAGTGACATTAAGAAGAGTGTGCAAGTAAATAATCAGCAACTGATTAACGGGAAAAGAACAGTT
AATTGTTGGAAGTGTGGTTCGACGCGCCACAAAGCTGTGCTCTGCTCAACTGTACAGCTTGCGCAATTGC
CAAAAAATATGTGGTGCCTTGATTCAGGAGCAACGGCTCATTTGTGCGGTGAGAGATCGATGTTTAAAG
CTTCAGAGAGCATAAAGAGCGCATTTCTTTGGTTGGCAATAAGTATATTTATGGCTGACGGTCGTGGTACA
GTGAAAATAGCCTGGCGCAACTCATCATTTGAATTGATCGACGTACTCTTTGTAAAGAATTTACAATGCA
M_Q
ACTTTATGTCGGTGTGCGAAGGCAATTGAAAACGATTTTCGTGTGTCGTTTCGAAAACAGGCGCGGAATTTT
L_Y_V_G_V_E_G_N_*
AAAAAATGATAAGAATAAAGTTGTTTTGATTGCTGAACTGCAAAGTGATTTGTTTATGTTCCAAAATGAA
M_K
AGAAAAGTTGAAAGTGCATGTTTCGTTGCACAGACAAATTTAATGAGAAAGTGGCACCAGAGATTCGGTC
E_K_L_K_V_H_V_S_L_H_R_Q_I_*
M_F_R_C_T_D_K_F_N_E_K_V_A_P_E_I_R_S
ATTTGAACTTTGCAAGCTTGAACAAAATGATCAAACAAGAAATGGTACTCGGACTAAAGAAAAAATATTT
F_E_L_C_K_L_E_Q_N_D_Q_T_R_N_G_T_R_T_K_E_K_I_F
AAGTGGCGCAAAAATAAATGACTGTTTGGAGCTGCGCAAAAAGCAAGATATGTGTGAAGAGTTTCCAAAA
K_W_R_K_I_K_* M_C_E_E_F_P_K
GCCTCTGGAATTCGCACAATTGATGTACTGGAGCTAGTGCATAGTGATATCTGTGGTCCAATGCAAACGA
S_L_W_K_S_H_N_*
CGTCTGTTGGAGGTGCACGCTATTTTGTCACTTTTCGTAGATGACAAATCGAGATACATGTTTGTGATTT
M_T_N_R_D_T_C_L_C_I
CATAAAGACGAGAGATGAAATACTGACCAATTTCAAGGAGTTTAAAGCATTTGCCGAAAACCAAAGAAGT
S_*
TGAAAGCAATAAGGAGCGACAACGGGCGTGAGTATTTAAGTAAAGCGTTTCAATGTATCTTGACAGAAAA
M_Y_L_D_R_K
CGGCATAAAGAGACAATTGATGGTACCGCACACCCCAACAAAACGGTGTAGCGGAGCGCGCTAACCGA

R H K E T I D G T A H P T T K R C S G A R *
ACCTGGTTCGAGATGGCAAGGACTATGATGATTCACGCAGGCGTCGGCCACTCACTATGGGCTGAAGCTA
TCAATACTGCTGCTTATCTACGCAATCGCGCTGAGACGTCAGCATTATCTGATGTAACACCGTTTCGAGTC
TTGGACTGGAAGGAAGCCATATGTGTCCCACCTGAAGATATTTGGTTCGAAGGCTATCGTGCTTAACAAG
TCCACAAAAAGGAAGTTTTCTGCGAAGGGCGAAGAAAATATTTGGTTGGATACTCTGATGCTTCTAAGG
M L L R
GATATCGCCTGTTCAATCCTATCAGGAAGAATATCTGTGTTGCTCGTGACGTAATTGTATTTGAGAACGA
D I A C S I L S G R I S V L L V T *
CCAGGATGATGGTATGACGGCAGGGCAAGCAGTACCTTCTCCCGATGTTCAACTTGTGGAGATGCAGCAC
M F N L W R C S T
CTTCCAGTCTGTGTTGCTGAGGGTAAGGCAAGAGTACGGATGAGGTTCCAACCTCAAGAAGCGTCAAGGA
F Q S V L L R V R Q E Y G *
GGCGAAACCATAAGGAACCTCTGAAGAATGGAGAACCCTCAATGCAGAAGGAGTTGGAGAGCCTACGTGCA
M E N L N A E G V G E P T C
M Q K E L E S L R A
AACAAAACCTGGTCATTGGTGGATTTGCCGGCAGGTGAGAAGGCCATTGGGTCGAAAGCGTGAAGCGTAA
K Q N L V I G G F A G R *
N K T W S L V D L P A G E K A I G S K A *
CAAGATGGGCAAAGTCGAGAGGTTTAAGTCTCGACTTGTCCCAAAGGGTGCCTCAGCGATACGGCGTT
AACTTACGGATACGTTCTCACCAGTGGCAGGCTACTCATCAATTAATTAAGTAATTGCTTTGGCGGTGG
AAAACGGTCTCTACATGCACCAATGGACGCTCTCGTCTGCATACTTGAATGGTGTCTTACGAAACGGT
M V I F T K R
ATATATGAGGCAGCCAGAGGGATTTCATCGATGAACGATATCCAAAGAAGGTGCTGAAGTTACACAAATCT
Y I *
ATATATGGACTGAAGCAGAGCGGCCGAGAATGGAATAAACTACTGAACGAAGTACTGCAAAGATTGGGT
TTTCTTTCATGCCCTAGCGAGCCATGTGTTTACACCCGTAATTCTGGCAATAGTAAGAACCCTTGTGTGAAT
GTACGTAGACGACCTTATCATAGCCAGTTCAAGCAAGGAGGAACCTGTGCGATATTAAGCATCAATTTCA
AAAGAATTCGATGTCGTAGACGGGGCGAGCTAAGACATTTCCCTGGGTATCGAAATTAACGTCGACGGAG
AAACTGGTGGAAATCAGGATTTGGCCACAAGCAGTATATTGAGAGCCTTATCAATGATTATTCGATGCAGGA
M I I R C R
TTGTAAGCCGAATCTCATCCCCTGGAAGCAGGCTTTGAAGTAAAATGTGACAAGGCCGATTGTGAGAAG
I V S R I S S H W K Q A L K *
GTGAATCAAGTCAGCTTTCAATCGCTGATTGGCTCATTAATGTATCTGGCAATCACAACCTCGTCTGATA
TTATGCATTCGGTCATTAACCTATCACAACGGAACCTCTGACCCACATAAAGAGTACGAAGCCGGTGCAAA
GCGAGTTCTGCGATACTTGAGGGGCGACAGTGTATTTGCAGTTACACTACGAGCGTACTGGTGTGCCTATA
CAGTGTATGTTGACGCGGATAGGGCTGGTGACACCACGGACCGAAAATTTCTTTACTGGATGGGCATGTA
M L T R I G L V T P R T E N S L L D G H V
TTGCTGCAGGAGCTGCTTTTACGTGGGACTCAAAGAAACAATCAGTGGTTTTCCCTAAGCAGCACAGAGTC
L L Q E L L L R G T Q R N N Q W F P *
AGAGTATGTGGCACTTTCCATGGCAGCGAAGGATGTGGCGTATGTTCCGAAATTTGGTCAACGAAATGGGT
M W H F P W Q R R M W R M F G N W S T K W V
TTTGGTGAAAACACAGGCAACTAAGGTTTATAGCGACAACCAGAGTTCCCAGTGTCTAGTAAAAATGAAA
L V K H R Q L R F I A T T R V P S A *
CTTTCCATGCACGTAGTAAGCATAAAAATATCATTTATATAAGAGAACTGTATAAGAACAATATAATAGAAG
TTAATTATGTACCCACAGAAAACATGATGGCAGATGTTCTAACCAAGAATTTGAATAGATTTAAACATGA
M
AAAGTGTATCCAAGGGATGGGTTTAAACTAACAAAAATGTATTTTGATTTTGGAAATTAAGATAGAAAATAT
K S V S K G W V * M Y F D F G I K I E I
TATCCGTAAGGGTCGACATATGTCATTGAGAGGGAGTGTGAAAATATGTATGTAATATGTATGTATATA
L S V R V D I L H * M Y V Y
TAATAATATGTAATAAAGTCAATGCACTGTGTCTCCCTCTTTTGGTTCGCGGTAACCAAAGCTTTTTTCT
I I I C N K V N A L C L P L L V A V T K S F F L
M H C V S L F W S R *
CTTATTGTGTTATCCTCTTTAGCGTGTAATTTGGCTGCCTGCGTGCAGTAACATTGTACTCTAGATCAGT
L L C Y P L *
CACAATAAACTACTTGAACGAATGGGGTTTATGGGCCAGCCCACGGGACATAAAAGTGAATAGTTCTA
M G V M G P A H A D I K V N S S
AACGAATTAGAATATTCGGCATCTTAAATATTGTGATTCAGGGATTTAATACGAAAACCTTATCGAAGT
K R I R I F G I F K Y C D S G I *
GGAAATGAGTTTCATCAATAAACCAAAATAGAAAACTTGATGATGAAAACCTACAGTGCCTGGGCTGTACAA
M M K T T V R G L Y K

ATGAAGAGTGTCTTAATTCATGCAGAATTATGGGGGGTGGTGTGTGGACGTGTGGTTAAAAACGAAAGTG
* M Q N Y G G W C V D V W L K T K V
M G G G V W T C G *
ATAGTGCTGAGCTAAAGGCTTTGTTCGACGCGAAGGACGAAAAAGCATTAGCAAGTATAATGCTATGCAT
I V L S *
TAAGACATCCCAAATTAACCATATAAAAACTGTGAGACTGCTGTAGAGGCATGGCAAAGACTTAGTGAA
ATCCACACTCCCTCAGGACCGGCACGACGTATATGTTTGCATAAGCAGTTGTTGCACATGAGAATGTCTG
M F A K A V V A H E N V *
AAACAGAAGTTGTTTCGAGTCATGTGAACAATTTTTGTGCGGCTGTGAAAAACTAAAAGAAATTC AATT
GGTGATCCAAGAAGAAGTCTGAGTATTCTCCTGTTGCCAGTTTGC CGGAATCGTTTGAAAGTTTCGTA
GTTGCAATAGAAACGCGCGATGAGTTGCCACATTGAAAATGTTAAAAATAAAATTACAAGAGGAAGGGC
AAAGGCGTATGGCAAATTAAGAACATTTCTGCAAAAAGCGAACAAAGTGCATTTGGAATTCGGTCTGCGAA
ACAAAAAGCGAAAAAGTGACATTAAGAAGAGTGTGCAAGTAAATAATCAGCAACTGATTAACGGGAAAAGA
ACAGTTAATTGTTGGAACGTGGTTCGCGAGCGCCACAAAGCTGTGCTCTGCTCAACTGTACAGCTTGGCG
AATTGCCAAAAAATATGTGGTGCCTTGATTCAGGAGCAACGGCTCATTGTGCGGTGAGAGATCGATGTT
TAAAAGCTTCAGAGAGCATAAAGAGCGCATTTCTTTGGTTGGCAATAAGTATATTTATGGCTGACGGTCGT
GGTACAGTGAAAAATAGCTTGGCGCAACTCATCTTTGAATTGATCGACGTACTCTTTGTAAAGAATTTAC
AATGCAACTTTATGTGCGTGTGCAAGGCAATTGAAAACGGATTTCTGTGTCTGTTGAAAACAGGCGCGG
M Q L Y V G V E G N *
AATTTTAAAAAATGATAAGAATAAAGTTGTTTTGATTGCTGAACTGCAAAGTGATTTGTTTATGTTCCAA
AATGAAAAGAAAAGTTGAAAGTGCATGTTTCGTTGCACAGACAAATTTAATGAGAAAAGTGGCACCAGAGAT
M K E K L K V H V S L H R Q I *
M F R C T D K F N E K V A P E I
TCGGTCATTTGAACTTTGCAAGCTTGAACAAAATGATCAAACAAGAAATGGTACTCGGACTAAAGAAAA
R S F E L C K L E Q N D O T R N G T R T K E K
ATATTTAAGTGGCGCAAATAAATGACTGTTTGGAGCTGCGCAAAAAGCAAGATATGTGTGAAGAGTTTC
I F K W R K I K * M C E E F
CCAAAAGCCTCTGAAAATCGCACAAATGATGTACTGGAGCTAGTGCATAGTGATATCTGTGGTCCAATGC
P K S L W K S H N *
AAACGACGTCTGTTGGAGGTGCACGCTATTTTGTCACTTTTCGTAGATGACAAATCGAGATACATGTTTGT
M T N R D T C L
GTATTTCATAAAGACGAGAGATGAAATACTGACCAATTTCAAGGAGTTTAAAGCATTTGCCGAAAACCAA
C I S *
AGAAGTTGAAAGCAATAAGGAGCGACAACGGGCGTGAGTATTTAAGTAAAGCGTTTCAATGTATCTTGAC
M Y L D
AGAAAACGGCATAAAGAGACAATTGATGGTACCGCACACCCCAACAACAAACGGTGTAGCGGAGCGCGCT
R K R H K E T I D G T A H P T T K R C S G A R
AACCGAACCTTGGTCGAGATGGCAAGGACTATGATGATTCACGCAGGCGTTCGGCCACTCACTATGGGCTG
*
AAGCTATCAATACTGCTGCTTATCTACGCAATCGCGCTGAGACGTGAGCATTATCTGATGTAACACCGTT
CGAGTCTTGGACTGGAAGGAAGCCATATGTGTCCCACCTGAAGATATTTGGTTCGAAGGCTATCGTGCTT
AACAAGTCCACAAAAAGGAAGTTTTCTGCGAAGGGCGAAGAAAATATTTTGGTTGGTACTCTGATGCTT
M L
CTAAGGGATATCGCCTGTTCAATCCTATCAGGAAGAATATCTGTGTTGCTCGTGACGTAATTGTATTTGA
L R D I A C S I L S G R I S V L L V T *
GAACGACCAGGATGATGGTATGACGGCAGGGCAAGCAGTACCTTCTCCCGATGTTCAACTTGTGGAGATG
M F N L W R C
CAGCACCTTCCAGTCTGTGTTGCTGAGGGTAAGGCAAGAGTACGGATGAGGTTCCA ACTCCAAGAAGCGT
S T F Q S V L L R V R Q E Y G *
CAAGGAGGCGGAAAACCATAGGAACTCTGAAGAATGGAGAACCCTCAATGCAGAAGGAGTTGGAGAGCCTA
M E N L N A E G V G E P
M Q K E L E S L
CGTGCAAAACAAAACCTGGTCATTTGGTGGATTTGCCGGCAGGTGAGAAGGCCATTTGGGTCGAAAGCGTGAA
T C K Q N L V I G G F A G R *
R A N K T W S L V D L P A G E K A I G S K A *
GCGTAACAAGATGGGCAAAGTGCAGAGGTTTAAAGTCTCGACTTGTCCAAAAGGGTGC GCTCAGCGATAC
GGCGTTAACTTCACGGATACGTTCTCACCAGTGGCAGCTACTCATCAATTAATTTAGTAATGCTTTGG
CGGTGGA AAAACGGTCTCTACATGCACCAAATGGACGTCTCGTCTGCATACTTGAATGGTGTCTTCACGA
M V I F T
AACGGTATATATGAGGCAGCCAGAGGGATTCATCGATGAACGATATCCAAGAAGGTGCTGAAGTTACAC
K R Y I *

AAATCTATATATGGACTGAAGCAGAGCGGCCGAGAATGGAATAAACTACTGAACGAAGTACTGCAAAAGA
 TTGGGTTTTCTTCATGCCCTAGCGAGCCATGTGTTTACACCCGTAATTCTGGCAATAGTAAGAACCCTTGT
 TGTAATGTACGTAGACGACCTTATCATAGCCAGTTC AAGCAAGGAGGAAC TGTGCGATATTAAGCATCA
 ATTTCAAAAAGAAATTCGATGTCGTAGACGGGGGCGAGCTAAGACATTTCCCTGGGTATCGAAATTGAACGTG
 ACGGAGAAAAC TGGTGGAA TCAGGAT TGGCCACAAGCAGTATATTGAGAGCCTTATCAATGATTATTCGAT
 GCAGGATTGTAAGCCGAATCTCATCCCCTGGAAGCAGGCTTTGAAGTAAAATGTGACAAGGCCGATTGT
 C_R_I_V_S_R_I_S_S_H_W_K_Q_A_L_K_*
 CAGAAGGTGAATCAAGTCAGCTTTCAATCGCTGATTGGCTCATTAATGTATCTGGCAATCACAACCTCGTC
 CTGATATTATGCATTCGGTCATTAACCTATCACACGGAACCTGACCCACATAAAGAGTACGAAGCCGG
 TGCAAAGCGAGTTCTGCGATACTTGAGGGGCACAGCTGATTTGCAGTTACACTACGAGCGTACTGGTGTG
 CCTATACAGTGCTATGTTGACGCGGATAGGGCTGGTGACACCACGGACC GAAAATTTCTTTACTGGATGGG
 M_L_T_R_I_G_L_V_T_P_R_T_E_N_S_L_L_D_G
 CATGTATTGCTGCAGGAGCTGCTTTTACGTGGGACTCAAAGAAACAATCAGTGGTTTCCCTAAGCAGCAC
 H_V_L_L_Q_E_L_L_L_R_G_T_Q_R_N_N_Q_W_F_P_*
 AGAGTCAGAGTATGTGGCACTTTCCATGGCAGCGAAGGATGTGGCGTATGTTCCGAAATTTGGTCAACGAA
 M_W_H_F_P_W_Q_R_R_M_W_R_M_F_G_N_W_S_T_K
 ATGGGTTTTGGTGAAACACAGGCAACTAAGGTTTATAGCGACAACCAGAGTTCCCAGTGCTTAGTGAAAA
 W_V_L_V_K_H_R_Q_L_R_F_I_A_T_T_R_V_P_S_A_*
 ATGAAACTTTCCATGCACGTAGTAAGCATAAAAATATCATTATATAAGAGAAC TGTATAAGAACAATATAA
 TAGAAGTTAATTATGTACCCACAGAAAACATGATGGCAGATGTTCTAACCAAGAATTTGAATAGATTTAA
 AC

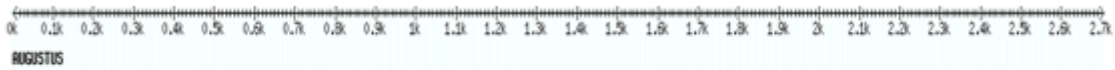
Para el gen 2R_88, el programa Augustus Gene Prediction predijo una estructura compuesta por 3 exones de 1377 pares de bases el primero, 136pares de bases el segundo, y 604 pares de bases el tercero, y 2 intrones de 2689pares de bases el primero, y 284 pares de bases el segundo.

Se encontró acuerdo con espectros de masas para el péptido ENLNAEGVGEPTCK con una probabilidad de $\rho=0.092$ y un valor $E=7.73884335297232e-05$; para el péptido FNEKVAPEIR con una probabilidad de $\rho=0.052$ y un valor $E=5.56497689739542e-05$; para el péptido ELES LRANK con una probabilidad de $\rho=0.044$ y un valor $E=0.000188347883869834$; para el péptido YVYIIICNK con una probabilidad de $\rho=0.005$ y un valor $E=9.64541922302759e-05$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 45 sORFs. Se pueden observar repeticiones. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 4 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_89



Hints given to Augustus

CAGAGCCAGATCAGATGCATCTGGACGCAAATACACTGGCCAACAGCGGAACACGAACTGTGGTTTCGGC
M H L D A N T L A N S G T R T V V S A
 CCAATGCCATTCTGTACACATAAAAAAAAAATGGAAACGTAACATAAAAAGTTAAGAACCCTCTTCTTGCGC
Q C H S V H I K K N G N V T *
 TCTTCGTCAGGACTCACCAGCGCTCGGCTCTCGTGTTCGCGGCCCGTCAGCAGGCGACTCGGGGCCTG
 TCTAGTAACATGTTCTGTACTCACCAGCGCTCGGCTCTCGTGTTCGCGGCCCGTCAGCAGGCGACTCG
 GGGCCTGTCTAGGAACATGTTTGTGTATGTGTGCATTGGAAACAAGTGCCGTTGGTTCGACTCAGGGTGA
M F V Y V C I R N K C R W S H S G *
 GGGGTCAACGGGGGAAGCGGATATAAAAAGCAGCGGGGCGGAGAAGAGGCCAGTCTCGAACGGACACAT
 AACGGAACCGCTAGCAGATCGCGAACTGAATCTTAAAATAAAGCTAATCGTAAACTCGAACCCCTCTTAAC
 TATCTTGACTATTATTTGGAGAACCACAGCATGTTGGTTGTTCATATCAAGGTGAGGTATGCGGCAGCGAG
 TGCCGAGAACCCTGATGCAAGTGGAACCTTGCCTTAACTGGCGCCCGAACAGGGACCGGCAATGTCCGGCC
 GATAAAAAGTGATACGAAAAAATTTGTGAAAAATTTGTGCGTAAAAATAGTGGTGGTGTGCATAAGTCAGATT
 AAGATCTGAAATCCATAAATGAAAAAGAAGTGCTGCGTGAGCTGTGTATAAAATGATAAAATAGCAATTA
M K K K C C V S C V *
 CCCGCTGCCGGGGGAACCTACGCCATCCCGGGGCGCAACAAATATTGCATAATTCATAAAAAGGTGTAA
 AATTTCTAAAATAAAAATGTAAACCTATGTTGCGCCAAGACCTAATTTAAATTAATAAAAACAACGACCCG
 CTACCGGAGGACGCCACGTGCCCCATGCCGAGCGCAAAAGTTGTACGATACCTATAACATAATTAACA
M P S A K V V R Y L *
 CGATCAACCCACTGCGGGGTACGGCTTGTGGGAAAATTTTTTTTTTCTCTCCTTGCCAATTCGCGAGTG
 CAAAAGATTGTGTATAATAAACCAATAATTAACCATTCAGCAGTTTACCTGCGGCAGTACGAGTAATAT
 GAGCGCCAGAGTGATAAGGTGGTGTGTGGCAGCTTGTGGATACGTTAAGTGGTGTGGAATGCACCCAA
M H P
 AAAAAACGCCCAACAAGTTGTGTGGCGCCGTACCTTAGTAGGCAACCAGCCAAAAGGGGACTACGGA
K K T A Q Q V V W R P Y L S R Q P A K R D T T E
 ACCACCGTGCCAGTGCCGAAATAAATTAGAGGTCATCAATAAAAAACTGTAACAGCACGCACGCAAGGA
P P C P V P K *
 AAAAAATATTGCAAAATGGAATAGCGCACAAAAATTTGTATAAACACATGCACAACACCACAATTCAAAGGA
M H N T T I Q R
 AAACAAGTATAGTACAATTTCCCAATTTCCCGAAACGAAATAAATAAAAATATAAGTACACAAAATGTA
K T S I V Q F P Q F P E T K *
 CCGTACCCACACACTACGTAGTCTTAGAACAACTTAGACGACCAGATATTTACGAATTGTCTTTTTGT
 AAGCGCATTTCTGCATGCGGCGCAAATCCCGCTCACTGGACTGGCTGGGGTTCGGCTTGGAAATGGGTAG
M G S
 CTGGATCTCCAGATGCTGCTGATTGGAACGCCGTCTTGGCCGCGCAAGCGACGGCTTCGAGGAACTGCAA
W I S R C C *
M L L I G T P S W P R K R R L R G T A
 AAAGTGGAGGAGGCTAGCTGTATCCCTCGGCTACTGAAGTAACCAACGAGTGGTTAAGCAAGTCGACGAT
K T G G G *
 GGAATGCTCCTCCTGACCAACTTCAACGGAACCTAAGAACGGCTGCAGAGAACTACGACCTGATCGGCT
 CCTTATCACCAATTCGACAATGAGACGATAATGGTCAACGGTCAAACTATTCCAGTTACTCGGTCAG
 TCATCTAATGGCGATGCCGCGCGTGTGGAAACATAACGGCCAGCAACTTTCAACTTTCTCTGGAATAC
 GTCCACGACGTGAGCATGAAGAATTTGGAAAAGATGTCCAACATGGCGAGTGAGCTACTAGCCTCTCTTC
 TCACCGAGGCGGCACTCGCAATCTGCATATTCCTAGGCTTTTATTTTCTATGGAAGAAGCTGATGTCCAC
 CAAAAGCATGCCGATGTCCGCGAGATTGCCGCAAACCTTAGAAGCATTTGGGCCAAACCGAGCTGAACAAG
 GCTCACTAATCTGCGGGACGCAGATCTTGGAGGGGGAGGAGTTAAGAACCCTCTTCTTGCCTCTTCGTC
 AGGACTCACCAGCGCTCGGCTCTCGTGTTCGCGGCCCGTCAGCAGGCGACTCGGGGCCTGTCTAGTAA
 CATGTTCTGTACTCACCAGCGCTCGGCTCTCGTGTTCGCGGCCCGTCAGCAGGCGACTCGGGGCCTGT
 CTAGGAACATGTTTGTGTATGTGTGCATTTCGGAACAAGTGCCGTTGGTTCGACTCAGGGTGAGGGTCAA
M F V Y V C I R N K C R W S H S G *
 CGGGGAAGCGGATATAAAAAGCAGCGGGGCGGAGAAGAGGCCAGTCTCGAACGGACACATAACGGAAC
 CGCTAGCAGATCGCGAACTGAATCTTAAAATAAAGCTAATCGTAAACTCGAACCCCTCTTAACCTATCTTGA
 CTATTATTTGGAGAACCACAGCATGTTGGTTGTTCATATCAAGGTGAGGTATGCGGCAGCGAGTGCCGAGA
M R Q R V P R
 ACCCTGATGCAAGTGGAACTTGCCTTAACTTAAATAAATGCACT

_T_L_M_Q_V_E_L_A_L_T_*

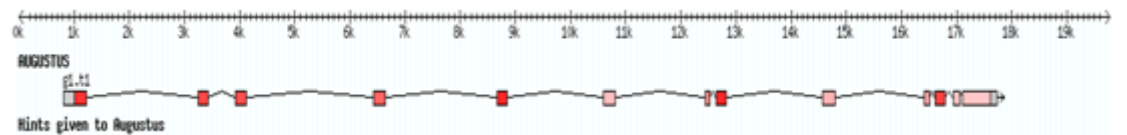
Para el gen 2R_88, el programa Augustus Gene Prediction predijo una estructura compuesta por 2 exones de 134 pares de bases el primero, y 1688 pares de bases el segundo, y 1 intrón de 340 pares de bases.

Se encontró acuerdo con espectros de masas para el péptido HNTTIQR con una probabilidad de $\rho=0.019$ y un valor $E=0.000167662151123091$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 10 sORFs. Se puede observar una repetición. El noveno sORF es una duplicación desplazada del segundo sORF. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de un péptido definido por espectrometría de masas que cumpliría con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

Gen 2R_96



TAGGCAAATGTACATTCAGCCCTACAAGTCTCCTGGTCCAGATGACATTCAGCCAATCCTGCTTCAAC
 GATCAGGAAAATACTGATCAGCAGCCTAGCTTTAGGACTCATACTACTGCCTGAAGTATAGCAAAGTG
 GTCTTCATTCGAAAGCTGGAAAGAAAGACATCACTGTTCCGAAGTCGTTTCAGACCCATCAGCTTGACAT
 CGTTTTGCTAAAAACGTTTAAAAAGCTAGTGGATGTCAGCATTAGAAGCACTCTGCTTGTGAAGCATCC
M_S_A_L_E_A_L_C_L_*
 GCTCCAAAGGACGCATCATGCCTACAGGGCGGGCAGATCAACTGATACTACCTTATATCACCTGAAAAGC
M_P_T_G_R_A_D_Q_L_I_L_P_Y_I_T_*
 CTTATTGAGGATTCAGTACTGACTCATACTGGAAGTGGCGTTATGTGGCTTCTTAGACATACAGGGTGCCTTCA
M_W_L_L_R_H_T_G_C_F_Q
 ACAATACCTCCCATGAGGCTGTCAACGCCTCCCTGGCAAGAAGAGGACTAGATGCGACAACCTAGCAGATG
Q_Y_L_P_*
M_R_L_S_T_P_P_W_Q_E_E_D_* M_R_Q_L_A_D_
M
 GATCAAATCACTACTGGCATCACTACAGTATCCACCACAAGGGGTTAACCAAGGGGGTGTCTCTCGCCT
G_S_N_H_Y_W_H_H_Y_S_I_H_H_K_G_L_T_K_G_V_S_S_R_L
D_Q_I_T_T_G_I_T_T_V_S_T_T_R_G_*
 CTCTGTGCAGCTTGTCTGGACAGACTGACCCGCATGGGTATACTTTGCCAAGGCTATGCTGACCATATTG
S_C_A_A_C_W_T_D_* M_G_I_L_C_Q_G_Y_A_D_H_I
 TCATTATAGTACTGCTTTAGGCCTGAGCCTAAATCCTGGCAAACCTGTTATTGTTCTTTACAAATAGGT
V_I_I_A_S_F_R_P_E_P_K_S_W_Q_N_C_Y_C_S_F_H_K_*
 ACAAGCTACAGAGAATAAAGGCAATAACCTGTGCAAGTGTGCGCATAGAGGTCAACAAAGAAGTTAAGTA
M_S_H_R_G_Q_Q_R_S_*
 TCTTGGGATAACCCCTTACTCCAACTTAGCTTCAAACTCATGTCATAACACAATTGATAAGTGCACCA
 GAGCAATTTACGTGTAGAGCATTATCTCAGTAATAAAAACAGTTGGACATCCCTGTGGAAAGTAACCA
 CAAAACCACAAAATAAAGTGGTATACAGATGGATCCCTCACCGACGAGGGAAGTGGGCTGGGGTTGTAGG

CCCCATGTTGAAATACCACGAATCAATGGGCAGATACACCAGCATTTCCTCAAGCTGAAGCCTATGCTATT
M D P S P T R E V G W G L *
M L L
GGACGCTGTGCAGAGTTTAACTCTGCAAAGGAAGTATCGTGACAAGGACATTGATATACTGTCTGATAGTC
D A V Q S L I C K G T I V T R T L I Y C L I V
AAGCAGCCATAAAGGGCGCTCAGCAAAGCTAAGATAACATCTAAGCTAGTAAATGAAGTGAGAACAGCCCC
K Q P *
AGACAAACTAGGAGCTGTCAACAAACTCACAATAAGGTGGGTCCCGGGACACAACAACATCCCGGGAAT
GAGCTAGCGGACAACCTAGCCAGGAAGGGCCAGAGAACCCTCTAATTGGGCCCGAACATTCTGTGGTGT
TGGTCACCACAGAGTACTGGGCTTACTTAAGTCAATAGCGGAAGAAAAACGTCTGTCTTCTGGGAACAC
CTACCAGGACTTAGGCAGTCTAAGATTCTCCTCCGTGAATATAACCATAAAAAGGTTCAAGACCTTAATGA
CATAACGGGAAAAACACCGTGCGCATTTTGGACTGGCCTTCTTACGGGGCATTGCCGACTTCACAGCCATTT
GCACAAGATTGGCATTGAAGACTTCTGCTGCACGGAGGAGAGCTCAGCACATATTATATGTGACTGC
ATGGCGCTTCAATCAAGAGGAACAGATTCCCTCGGCATGTATGTAGTCCCACGGGAAACAATTACCCCAAT
M Y V V P R E T I T P I
AAAATTCTGGCATTATACAGTGCATTGGACTTCAGGGGGATCTTTGAGTCATGAAGGGGTAGTACAATA
K F W H L Y S A L D F R G I F E S *
GATCTGACTGGGTGCGAGTACACAGACAAAATCACTTAATAATAATACTTCATAAGTCAAGCTAAGAGGC
CTGCTTCACAAAAGTTTCAGCTTAGGAACAGTAATGGTCTTTAATGGACCAACACGAGATTCGAGCAATA
M V F N G P T R D F E Q Y
M D Q H E I S S N
TAACTAGCCGGAACGCTTTCCTCCTTTGGACACAATGTAAACACAAGCACCATAAGCCTTAAGGCTAGCA
N *
I T S R N A F L L W T Q C K H K H H K P *
TCGCTAAACCCATGAATTACACGATCTGAATCTGAGTCCAGCACAAAACGTGGGAATTCATCGCTAAAGC
TAATACAAAAGCTAATATCGAGCCATTTCTGATTTTTGAGACTCGGGTATGCTTTAATGCCAGGATAACTT
CTCCTTACAGAGCTGTTGCAGAAAGATTTTACACTTTGCGATGACGTGTCCAAAGTCCAATCGTAAACC
M T C P K S N R K T
TCGCAAGTGAGGATGCCTACCGGAATAAACTGACCATGTATGAGTTTAGATCAGGAGATCAGGAAATCCA
S Q V R M P T G I N * M Y E F R S G D Q E I Q
M S L D Q E I R K S
AACGAAGGAATTATATAACCTTCTGGGAGGCAATAGTAAACACATGTGAAGGCGCTAGACTTCACTGAGC
T K E L Y N L L G G N S K H M *
K R R N Y I T F W E A I V N T C E G A R L H *
AATCTCCAAAGGAGTGACAGAATCAAATCAGGCACTATGAAAGGAGGATAACTCTTATACTATAGGAAGT
AAGGAGAACTTGAGTACTACTGGCAACTCACTTTCCGGGAAGCACGCTTCAAACGGATGGGAATACCC
CGGTAAGTGCACAAAGTAGACCGCGTGCTACTGATTGGGCCAAAGCAAATCACAAGTAACAACAGAGAG
ATTAGGATAGGCAAATTTGATACATTCAGCCCTACAAGTCTCCTGGTCCAGATGACATTCAGCCAATCCTG
M T F S Q S C
CTTCAACGATCAGGAAAATACTGATCAGCAGCCTAGCTTTAGGACTCATACTACTGCCTGGAGTATAGC
F N D Q E N T D Q Q P S F R T H T Y C L E Y S
AAAAGTGGTCTTCATTCGAAAAGCTGGAAAAGAAAGACATGACTGATCCGAAGTCGTTTAGACCCATCAGC
K S G L H S E S W K E R H D *
M T D P K S F R P I S
TTGATATCGTTTTTGTCTAAAACCTTGGAAAAGCTAGTGGATGTTAGCATTAGAAGCACTCTGCTTGTGGA
L I S F L L K P W K S * M L A L E A L C L W
GCATCCGCTCCAAAAGGACGCAGCATGCCCTACAGGGCGGGCAGATCAACTGATACTACCCTATATCACCTG
S I R S K G R S M P T G R A D Q L I L P Y I T *
AAAAGCCTTATTGAGGATTCATGACTCATAAGGAAGTGACGTTATGTGCCTTCTTAGACATACAGGGTG
M T H K E V T L C A F L D I Q G
M C L L R H T G C
CTTTCGACAATACCTCCCATGAGGCTGTCAACGCATCCCTGGCAAGAAGAGGACTAGATGCGACAACCAG
A F D N T S H E A V N A S L A R R G L D A T T S
F R Q Y L P *
M R L S T H P W Q E E D * M R Q P
CAGATGGATTAAATCACTACTGGCATGACCACTGTAGGAGGACAGTTCGTCACGGTACTTGCTGGTCTAC
R W I K S L L A *
A D G L N H Y W H D H C R R T V V H G T C W S T
TTGCTGGTAGACGAGTTGCTGGACAGACTGACAGGCAGGGGTCATTACAGCTGGAGGTAAATATGAGGAA
C W *

ACACTCTGCGATATCATTCAACTGGGCCTCGAGTGGTGAAGGAAGTAGGCCTGAGAGTAAATCCTAGCA
AATCTGTTATTGTTCCCTTTCACGAATAGGTACAAGCTACAGAGAATGAAGGCAATAACCCTGTGCAATG
TCGCATAGAGGCCAGCAAAGAAATTAAGTATCTTGGGATAACCCTTAACTCCAAACTTAGCTTCAA
S H R G Q Q R N *
CATGTCGATAACACAATTGATAAATGCACCAGAGCACTTTTTCACGTGCAGAAAAATGCTGGTGGTAAGT
M S I T Q L I N A P E H F S R A E K L L V V S
M H Q S T F H V Q K N C W W *
CGTGGGGAACCTCACTACTAAGATGGCTGACCTCATGGTAGTCAGACCCATACTAACTTATCTGGCAA
R G E P H Y * M A V P H G S Q T H T N L S G K
TAGCCTGGGGTGATAGAGCTGCAAAGAATGGCCTGCGTATGCATGCCAGGAGAAATGCGTACATGTTTCA
* M H A R R N A Y M F H
CTGCAGCCCTTGAAGTACTAATGGAGGGCAGCCGCTTCATATCGTCATTCAAATGAAACGAAAGCCAT
C S P *
CCTAATAAGAATTGAGGGAGCAGGAAATGACTGTAACCTCACAAGTAAGGATGCTGAAAGCCTAAAAAGG
M T V T S Q V R M L K A *
GATATCCCTTTGCGAATGCAACCAAGGGATGAGATGCCAGCTGAGCATAGGTTTGGCTCAGGACATCCCTG
GTGAAAGTACACCCTATGAAAACACAAACAATAAAGTGGTATACAGACGGATCCCTCACCAGCAGGGGAA
GTGGGCTGGGGTGTAGGCCGAGGTTAAAATACCACGAAAAAGTAGGCAGATAAATTTTCAGCATTTTT
CATGCTGAAGTCTGTGCTGTGGACGCTGTGCGGAGTTTAACTGCAAAGGAACATCGTGGCAAGGACA
M L K S V L L D A V R S L I C K G T I V A R T
TTTCTATACTGTCTGATAGTCAAGCAGCATTAAAGGCGCTCAGCAAAGCTAAGATAACATCTAAGCTAGT
F L Y C L I V K Q H *
AAATGAAGTGAGGACAGCCCTAGACAAACTAGGAGCTGTCAACAACTCACAATAAGGTTGGTCCCAGGGA
CACAACAACATCCCAGGAAATGAGCTAGCAGACAACCTAGCCAGGAAAGGAGCAGAGAACCCTCTAATTG
GGCCGAAACCCTTCTGTGGTGTGGTACCACGGAGTACGGGGCTTGCTAAAGTCACTAGAGGAAGAAAA
ACGTCTGTTCTTCTGGGAACACCTACCAGGACTTAGGCACCTAAGATTCTTCTACGTGAATAAAAGGTT
TAAGACCTTAATGACACACGGGAAAAACACCGTGCATTTTTGACTGGCCTTTTTACGGGGCATTGACGA
M T H G K N T V R I L T G L F T G H *
CTTCACAGCCATTTGCACAAGATTGGCATTGAAGACTTCTGCTGCACGGAGGAGGAGAGCTCTGCACATA
TTATATGTGACTGCATGGCGCTTCAATCAAGAGGAACAGATTCCCTCGGCATGTATGTAGTCCCACGGAA
M A L Q S R G T D S S A C M *
M Y V V P R E
ACAATTGCAGCCCTGAACCCCAATAAAAATTTCTGGCATTATACAGTGCATTGGACTTCAGGGGGATCTTT
T I A A L N P N K I L A F I Q C I G L Q G D L
GAGTCATGAAGGGTGTAGTACAATAGATCTGACTGGGTGCGAGTACACAGACAAAATCACTTAATAATAAT
ACTTCATAAGTCAAGCTAAGAGGCCTGCTTACAAAAGTTTTCAGCTTAGGAACAGTAATGGTCTTTAATGG
* M V F N G
M
ACCAACACGAGATTTTCGAGCAATATAACTAGCCGGAACGCTTTCCTCCTTTGGACACAATGTAAACACAA
P T R D F E Q Y N *
D Q H E I S S N I T S R N A F L L W T Q C K H K
GCACATAAGCCTAAGGCTAGCATCGCTAAACCCATGAATTACACGATCTGAATCTGAGTCCAGCACAA
H H K P *
AACGTGGGAATTCATCGCTAAAGCTAATACAAAAGCTAATATCGAGCCATTTTCGTATTTTGGACTCGGG
TATGCTTTAATGCCAGGATAACTTCTCCTTACAGAGCTGTTGCAGAAAGATTTTACACTTTGCGATGACG
M T
TGTCAAAAGTCCAATCGTAAAACTCGCAAGTGAAGGATGCCTACCAGGAAATAACTGACCATGTATGAGTT
C P K S N R K T S Q V R M P T G I N * M Y E F
M S
TAGATCAGGAGATCAGGAAATCCAAACGAAGGAATTATATAACCTTCTGGGAGGCAATAGTAAACACATG
R S G D Q E I Q T K E L Y N L L G G N S K H M
L D Q E I R K S K R R N Y I T F W E A I V N T C
TGAAGGCGCTAGACTTCACTGAGCAATCTCAAAGGAGTGACAGAATCAAATCAGGCACTATGAAAGGAG
*
E G A R L H *
GATAACTCTTATACTATAGGAAGTAAGGAGAACTTGAGCTACTACTGGCAACTCACTTTCCGGGAAGCA
CGCTTCAAACGGATGGGAATACCCCGGTAACCTGCACAAAGTAGACCGCGTACTGATTGGGCCAAAGC
AAAATCACAAGTAACAACAGAGAGATTAGGATAGGCAAATTTGTACATTCAGCCCTACAAGTCTCCTGGT
CCAGATGACATTCAGCCAATCTGCTTCAACGATCAGGAAAATACTGATCAGCAGCCTAGCTTTAGGACT
M T F S Q S C F N D Q E N T D Q Q P S F R T

CATACTACTGCCTGGAGTATAGCAAAAGTGGTCTTCATTCCGAAAGCTGGAAAGAAAGACATGACTGAT
H T Y C L E Y S K S G L H S E S W K E R H D *
M T D
 CCGAAGTCGTTTGTAGACCCATCAGCTTGATATCGTTTTTGGCTAAAACCTTGGAAAAGCTAGTGGATGTTAG
P K S F R P I S L I S F L L K P W K S *
 CATTAGAAGCACTCTGCTTGTGGAGCATCCGCTCCAAAGGACGCAGCATGCCTACAGGGCGGGCAGATCA
A L E A L C L W S I R S K G R S M P T G R A D Q
 ACTGATACTACCCTATATCACCTGAAAAGCCTTATTGAGGATTCAATGACTCATAAGGAAGTGACGTTAT
L I L P Y I T *
M T H K E V T L
 M
 GTGCCTTCTTAGACATACAGGGTGTTCGACAATACCTCCCATGAGGCTGTCAACGCATCCCTGGCAAG
C A F L D I Q G A F D N T S H E A V N A S L A R
C L L R H T G C F R Q Y L P *
M R L S T H P W Q
 AAGAGGACTAGATGCGACAACCAGCAGATGGATTAAATCACTACTGGCATGACCACTGTAGGAGGACAGT
R G L D A T T S R W I K S L L A *
E E D *
 M R Q P A D G L N H Y W H D H C R R T V
 CGTCCACGGTACTTGTGGTCTACTTGTGGTAGACGAGTTGTGGACAGACTGACAGGCAGGGGTCATT
V H G T C W S T C W *
 ACAGCTGGAGGTAATAATGAGGAAACACTCTGCGATATCATTCAACTGGGCCTCGAGTGGTGAAGGAAG
 TAGGCCTGAGAGTAAATCCTAGCAAACTGTTATTGTTTCCTTTCACGAATAGGTACAAGCTACAGAGAAT
 GAAGGCAATAACCTGTGCAATGTGCGATAGAGGCCAGCAAAGAAATTAAGTATCTTGGGATAACCTT
M S H R G Q Q R N *
 AACTCCAACTTAGCTTCAAACTCATGTGATAACACAATTGATAAATGCACCAGAGCACTTTTTCAGT
M S I T Q L I N A P E H F S R
M H Q S T F H V
 GCAGAAAAATTGCTGGTGGTAAGTCGTGGGGAACCTCACTACTAAGATGGCTGTACCTCATGGTAGTCAG
A E K L L V V S R G E P H Y *
Q K N C W W *
M A V P H G S Q
 ACCCATACTAACTTATCTGGCAAATAGCCTGGGGTGATAGAGCTGCAAAGAATGGCCTGCGTATGCATGC
T H T N L S G K *
M H A
 CAGGAGAAATGCGTACATGTTTCACTGCAGCCCTTGAAGTACTAATGGAGGCGACGCCGCTTCATATCGT
R R N A Y M F H C S P *
 CATTCAAATGAAACGAAAGCCATCCTAATAAGAATTGAGGGAGCAGGAAATGACTGTAACCTCACAAGT
M T V T S Q V
 AAGGATGCTGAAAGCCTAAAAAGGATATCCCTTTCGGAATGCAACCAAGGGATGAGATGCCAGCTGAGC
R M L K A *
 ATAGGTTTTGCTCAGGACATCCCTGGTAAAAGTACACCCTATGAAAACACAAACAATAAAGTGGTATACAG
 ACGGATCCCTCACCGACGAGGGAAAGTGGGCTGGGGTGTAGGCCGCAGGTTAAAATACCACGAAAAAGT
 AGGCAGATAAAATTTTCAGCATTTTTCATGCTGAAGTCTGTGCTGTTGGACGCTGTGCGGAGTTAATCTGC
M L K S V L L D A V R S L I C
 AAAGGAACATCGTGGCAAGGACATTTCTATACTGTCTGATAGTCAAGCAGCATTAAGGGCGCTCAGCAA
K G T I V A R T F L Y C L I V K Q H *
 AGCTAAGATAACATCTAAGCTAGTAAATGAAGTGAGGACAGCCCTAGACAAACTAGGAGCTGTCAACAAA
 CTCACAATAAAGTTGGTCCCGGACACAACAACATCCCGGAAATGAGCTAGCAGACAACCTAGCCAGGA
 AAGGAGCAGAGAACCCTCTAATTGGGCCCGAACCCTTCTGTGGTGTGGTACCACGGAGTACGGGGCTT
 GCTAAAAGTCACTAGAGGAAGAAAAACGCTGTCTTCTTCTGGGAACACCTACCAGGACTTAGGCCTCTAAG
 ATTCTTCTACGTGAATAAAAAGGTTAAGACCTTAATGACACACGGGAAAAACACCGTGCGCATTTTACT
M T H G K N T V R I L T
 GGCTTTTTACGGGGCATTGACGACTTACAGCCATTTGCACAAGATTGGCATTGAAGACTTCTGCTGCA
G L F T G H *
 CGGAGGAGGAGACTCTGCACATATTATATGTGACTGCATGGCGCTTCAATCAAGAGGAACAGATTCCCT
M A L Q S R G T D S S
 GGCATGTATGTAGTCCCACGGGAAACAATTGCAGCCCTGAACCCCAATAAAATTTCTGGCATTATACAGT
A C M *
M Y V V P R E T I A A L N P N K I L A F I Q
 GCATTGGACTTACGGGGATCTTTGAGTCAATGAAGGGTAGTACAATAGATCTGACTGGGTGCGAGTACA
C I G L Q G D L *
 CAGACAAAATCACTTAATAATAACTTTCATAAGTCAAGCTAAGAGGCCCTGCTTACAAAAGTTTTCAGCTT
 AGGAACAGTAATGGTCTTTAATGGACCAACACGAGATTTTCGAGCAATATAACTAGCCGGAACGCTTTCCT
M V F N G P T R D F E Q Y N *

M D Q H E I S S N I T S R N A F L
 CCTTTGGACACAATGTAAACACAAGCACCATAAGCCTTAAGGCTAGCATCGCTAAACCCATGAATTACAC
L W T Q C K H K H H K P *
 GATCTGAATCTGAGTCCAGCACAAAACGTGGGAATTCATCGCTAAAGCTAATACAAAAGCTAATATCGAG
 CCATTTTCGTATTTTGGAGACTCGGGTATGCTTTAATGCCAGGATAAATTCTCCTTACAGAGCTGTTGCAGA
 AAGATTTTACACTTTGCGATGACGTGTCCAAAAGTCCAATCGTAAAACCTCGCAAGTGAGGATGCCTACCG
M T C P K S N R K T S Q V R M P T
 GAATAAACTGACCATGTATGAGTTTAGATCAGGAGATCAGGAAAATCCAAACGAAGGAATTATATAACCTT
G I N * M Y E F R S G D Q E I Q T K E L Y N L
M S L D Q E I R K S K R R N Y I T F
 CTGGGAGGCAATAGTAAACACATGTGAAGGCGCTAGACTTCACTGAGCAATCTCCAAAGGAGTGACAGAA
L G G N S K H M *
W E A I V N T C E G A R L H *
 TCAAATCAGGCCTATGAAAAGGAGGATAAATCTTATACTATAGGAAGTAAGGAGAAAATGAGCTACTAC
 TGGCAACTCACTTCCGGGAAGCACGCTTCAAACGGATGGGAATACCCCGGTAAGTGCACAAAGTAGACC
 GCGTGTACTGATTGGGCCAAAAGCAAAATCACAAGTAACAACAGAGAGATTAGGATAGGCAAATTTGTACA
 TTCCAGCCCTACAAGTCTCTGGTCCAGATGACATTCAGCCAATCCTGCTTCAACGATCAGGAAAATACT
M T F S Q S C F N D Q E N T
 GATCAGCAGCCTAGCTTTAGGACTCATACCTACTGCCTGGAGTATAGCAAAGTGGTCTTTCATTCCGAAA
D Q Q P S F R T H T Y C L E Y S K S G L H S E
 GCTGGAAAAGAAAGACATGACTGATCCGAAGTCGTTTAGACCCATCAGCTTGATATCGTTTTTGTAAAAC
S W K E R H D *
M T D P K S F R P I S L I S F L L K
 CTTGAAAAGCTAGTGGATGTTAGCATTAGAAGCACTCTGCTTGTGGAGCATCCGCTCCAAAGGACGCAG
P W K S * M L A L E A L C L W S I R S K G R S
 CATGCCCTACAGGGCGGGCAGATCAACTGATACTACCCTATATCACCCTGAAAAGCCTTATTGAGGATTCAA
M P T G R A D Q L I L P Y I T *
 TGACTCATAAGGAAGTGACGTTATGTGCCTTCTTAGACATACAGGGTGCCTTTCGACAATACCTCCCATGA
M T H K E V T L C A F L D I Q G A F D N T S H E
M C L L R H T G C F R Q Y L P *
M
 GGCTGTCAACGCATCCCTGGCAAGAAGAGGACTAGATGCGACAACCAGCAGATGGATTAATCACTACTG
A V N A S L A R R G L D A T T S R W I K S L L
R L S T H P W Q E E D * M R Q P A D G L N H Y W
 GCATGACCACTGTAGGAGGACAGTCGTCCACGGTACTTGCTGGTCTACTTGCTGGTAGACGAGTTGCTGG
A *
H D H C R R T V V H G T C W S T C W *
 ACAGACTGACAGGCAGGGGTCATTACAGCTGGAGGTAAATATGAGGAAACACTCTGCGATATCATTCAAC
 TGGGCTCGAGTGGTGCAAGGAAGTAGGCCGTGAGAGTAAATCCTAGCAAATCTGTTATTGTTCTTTTAC
 GAATAGGTACAAGCTACAGAGAATGAAGGCAATAACCCGTGCAGAATGTCGCATAGAGGCCAGCAAAGAA
M S H R G Q Q R
 ATTAAGTATCTTGGGATAACCCCTAACTCCAAACTTAGCTTCAAACCTCATGTCGATAACACAATTGATA
N * M S I T Q L I
 AATGCACCAGAGCACTTTTACGTCGAGAAAAATTGCTGGTGGTAAGTCGTGGGAACCTCACTACTAAG
N A P E H F S R A E K L L V V S R G E P H Y *
M H Q S T F H V Q K N C W W *
 ATGGCTGTACCTCATGGTAGTCAGACCCATACTAATCTTATCTGGCAAATAGCCTGGGGTGATAGAGCTGC
M A V P H G S Q T H T N L S G K *
 AAAGAATGGCCTGCGTATGCATGCCAGGAGAAATGCGTACATGTTTCACTGCAGCCCTTGAAGTACTAAT
M H A R R N A Y M F H C S P *
 GGAGGCGACGCCGCTTCATATCGTCATTCAAATGAAACGGAAAGCCATCCTAATAAGAATTGAGGGAGCA
 GGAAATGACTGTAACCTCACAAGTAAGGATGCTGAAAGCCTAAAAAGGGATATCCCTTTGCGAATGCAAC
M T V T S Q V R M L K A *
 CAAGGGATGAGATGCCAGCTGAGCATAGGTTTGTCTCAGGACATCCCTGGTCAAAGTACACCCATGAAAA
 CACAAACAATAAAGTGGTATACAGACGGATCCCTCACCGACGAGGGAAGTGGGCTGGGGGTTGTAGGCCG
 CAGGTTAAAAATACCACGAAAAAGTAGGCAGATAAATTTTTCAGCATTTTTTCATGCTGAAGTCTGTGCTGTTG
M L K S V L L
 GACGCTGTGCGGAGTTTAAATCTGCAAAGGAACTATCGTGGCAAGGACATTTCTATACTGTCTGATAGTCA
D A V R S L I C K G T I V A R T F L Y C L I V
 AGCAGCATTAAGGCGCTCAGCAAAGCTAAGATAACATCTAAGCTAGTAAATGAAGTGAGGACAGCCCTA
K Q H *

GACAACTAGGAGCTGTCAACAACTCACAATAAGGTTGGTCCCGGGACACAACAACATCCCGGGAAATG
AGCTAGCAGACAACCTAGCCAGGAAAGGAGCAGAGAACCCTCTAATTGGGCCCGAACCCCTTCTGTGGTGT
TGGTCACCACGGAGTACGGGGCTTGC TAAAGTCACTAGAGGAAGAAAAACGTCTGTTCTTCTGGGAACAC
CTACCAGGACTTAGGCACTCTAAGATTCTTCTACGTGAATAAAAGGTTTAAAGACCTTAATGACACACGGG
M T H G
AAAAACACCGTGCGCATTTTACTGGCCTTTTTACGGGGCATTGACGACTTCACAGCCATTTGCACAAGA
K N T V R I L T G L F T G H *
TTGGCATTGAAGACTTCTGCTGCACGGAGGAGGAGACTCTGCACATATTATATGTGACTGCATGGCGCT
M A L
TCAATCAAGAGGAACAGATTCTCTCGGCATGTATGTAGTCCCACGGGAAACAATTGCAGCCCTGAACCCCA
Q S R G T D S S A C M *
M Y V V P R E T I A A L N P
ATAAAAATTCTGGCATTTATACAGTGCATTGGACTTCAGGGGGATCTTTGAGTCATGAAGGGGTAGTACAA
N K I L A F I Q C I G L Q G D L *
TAGATCTGACTGGGTGCGAGTACACAGACAAAATCACTTAATAATAATACTTCATAAGTCAAGCTAAGAG
GCCTGCTTCACAAAAGTTTACAGTTAGGAACAGTAATGGTCTTTAATGGACCAACACGAGATTTGAGCAA
M V F N G P T R D F E Q
M D Q H E I S S N
TATAACTAGCCGGAACGCTTTCCTCCTTTGGACACAATGTAAACACAAGCACCATTAAGCCTTAAGGCTAG
Y N *
I T S R N A F L L W T Q C K H K H H K P *
CATCGCTAAACCCATGAATTACACGATCTGAATCTGAGTCCAGCACAAAACGTGGGAATTCATCGCTAAA
GCTAATACAAAAGCTAATATCGAGCCATTTGCTATTTTGGAGACTCGGGTATGCTTTAATGCCAGGATAAC
TTCTCCTTACAGAGCTGTTGCAGAAAAGATTTTACACTTTGCGATGACGTGTCCAAAGTCCAATCGTAAAA
M T C P K S N R K
CCTCGCAAGTGAGGATGCCTACCGGAATAAACTGACCATGTATGAGTTTAGATCAGGAGATCAGGAAATC
T S Q V R M P T G I N * M Y E F R S G D Q E I
M S L D Q E I R K S
CAAACGAAGGAATTATATAACCTTCTGGGAGGCAATAGTAAACACATGTGAAGGCGCTAGACTTCACTGA
Q T K E L Y N L L G G N S K H M *
K R R N Y I T F W E A I V N T C E G A R L H *
GCAATCTCCAAAAGGAGTGACAGAATCAAATCAGGCACTATGAAAGGAGGATAACTCTTATACTATAGGAA
GTAAGGAGAAACTTGAGTACTACTGGCAACTCACTTTCCGGGAAGCACGCTTCAAACGGATGGGAATAC
CCCGGTAAGTGCACAAAAGTAGACCGCGTGCTACTGATTGGGCCAAAGCAAATCACAAGTAACAACAGAG
AGATTAGGATAGGCAAATGTACATTCCAGCCCTACAAGTCTCCTGGTCCAGATGACATTCAGCCAATCC
M T F S Q S
TGCTTCAACGATCAGGAAAATACTGATCAGCAGCCTAGCTTTAGGACTCATACCTACTGCCTGGAGTATA
C F N D Q E N T D Q Q P S F R T H T Y C L E Y
GCAAAAAGTGGTCTTCAATTCGAAAAGCTGGAAAAGAAAGACATGACTGATCCGAAGTCGTTTAGACCCATCA
S K S G L H S E S W K E R H D *
M T D P K S F R P I
GCTTGATATCGTTTTTGTCTAAAACCTTGGAAAAGCTAGTGGATGTTAGCATTAGAAGCACTCTGCTTGTG
S L I S F L L K P W K S * M L A L E A L C L W
GAGCATCCGCTCCAAAAGGACGCGCATGCCTACAGGGCGGGCAGATCAACTGATACTACCCTATATCACC
S I R S K G R S M P T G R A D Q L I L P Y I T
TGAAAAGCCTTATTGAGGATTCAATGACTCATAAGGAAGTGACGTTATGTGCCTTCTTAGACATACAGGG
* M T H K E V T L C A F L D I Q G
M C L L R H T G
TGCTTTGACAATACCTCCCATGAGGCTGTCAACGCATCCCTGGCAAGAAGAGGACTAGATGCGACAACC
A F D N T S H E A V N A S L A R R G L D A T T
C F R Q Y L P *
M R L S T H P W Q E E D * M R Q P
AGCAGATGGATTAAATCACTACTGGCATGACCCTGTAGGAGGACAGTCGTCCACGGTACTTGCTGGTCT
S R W I K S L L A *
A D G L N H Y W H D H C R R T V V H G T C W S
ACTTGCTGGTAGACGAGTTGCTGGACAGACTGACAGGCAGGGGTCATTACAGCTGGAGGTAAATATGAGG
T C W *
AAACACTCTGCGATATCATTTCAACTGGGCCTCGAGTGGTGAAGGAAGTAGGCCTGAGAGTAAATCCTAG
CAAATCTGTTATTGTTCTTTCACGAATAGGTACAAGCTACAGAGAATGAAGGCAATAACCCTGTGAGAA
TGTCGCATAGAGGCCAGCAAAGAAATTAAGTATCTTGGGATAACCCCTTAACCTCAAACCTTAGCTTCAAAA
M S H R G Q Q R N *

CTCATGTCGATAACACAATTGATAAATGCACCAGAGCACTTTTCACGTGCAGAAAAATTGCTGGTGGTAA
M S I T Q L I N A P E H F S R A E K L L V V
M H Q S T F H V Q K N C W W *
GTCGTGGGGAACCTCACTACTAAGATGGCTGTACCTCATGGTAGTCAGACCCATACTAACTTATCTGGCA
S R G E P H Y * M A V P H G S Q T H T N L S G
AATAGCCTGGGGTGATAGAGCTGCAAAGAAATGGCCTGCGTATGCATGCCAGGAGAAATGCGTACATGTTT
K * M H A R R N A Y M F
CACTGCAGCCCTTGAAGTACTAATGGAGGCGACGCCGCTTCATATCGTCATTCAAATGAAACGGAAAGCC
H C S P *
ATCCTAATAAGAATTGAGGGAGCAGGAAATGACTGTAACCTCACAAGTAAGGATGCTGAAAGCCTAAAA
M T V T S Q V R M L K A *
GGGATATCCCTTTGCGAATGCAACCAAGGGATGAGATGCCAGCTGAGCATAGGTTTGCTCAGGACATCCC
TGGTGAAGTACACCCCTATGAAAAACAAAATAAAGTGGTATACAGACGGATCCCTCACCGACGAGGG
AAGTGGGCTGGGGTGTAGGCCGCGAGTTAAAATACCACGAAAAAGTAGGCAGATAAAATTTTACGATTT
TTCATGCTGAAGTCTGTGCTGTTGGACGCTGTGCGGAGTTTAACTCTGCAAAGGAACATCGTGGCAAGGA
M L K S V L L D A V R S L I C K G T I V A R
CATTTCTATACTGTCTGATAGTCAAGCAGCATTAAGGCGCTCAGCAAAGCTAAGATAACATCTAAGCTA
T F L Y C L I V K Q H *
GTAAATGAAGTGAAGGACAGCCCTAGACAACTAGGAGCTGTCAACAACTCACAATAAGGTTGGTCCCGG
GACACAACAACATCCCGGAAATGAGCTAGCAGACAACCTAGCCAGGAAAGGAGCAGAGAACCCTCTAAT
TGGGCCCGAACCCCTTCTGTGGTGTGGTCCACCGGAGTACGGGGCTTGCTAAAGTCACTAGAGGAAGAA
AAACGTCTGTTCTTCTGGGAACACCTACCAGGACTTAGGCACCTAAGATTTCTTACGTGAATAAAAGG
TTTAAAGACCTTAATGACACACGGGAAAAACACCGTGCGCATTTTACTGGCCTTTTTACGGGGCATTGAC
M T H G K N T V R I L T G L F T G H *
GACTTACAGCCATTTGCACAAGATTGGCATTGAAGACTTCTGCTGCACGGAGGAGGAGAGCTCTGCACA
TATTATATGTGACTGCATGGCGCTTCAATCAAGAGGAACAGATTCCCTCGGCATGTATGTAGTCCCACGGG
M A L Q S R G T D S S A C M *
M Y V V P R
AAACAATTGCAGCCCTGAACCCCAATAAAAATTCTGGCATTATACAGTGCATTGGACTTCAGGGGGATCT
E T I A A L N P N K I L A F I Q C I G L Q G D L
TTGAGTCATGAAGGGGTAGTACAATAGATCTGACTGGGTGCGAGTACACAGACAAAATCACTTAATAATA
*
ATACTTCATAAGTCAAGCTAAGAGGCCTGCTTACAAAAGTTTACAGCTTAGGAACAGTAATGGTCTTTAAT
M V F N
M
GGACCAACACGAGATTTTCGAGCAATATAACTAGCCGGAACGCTTTCCTCCTTTGGACACAATGTAAACAC
G P T R D F E Q Y N *
D Q H E I S S N I T S R N A F L L W T Q C K H
AAGCACCATAAGCCTTAAGGCTAGCATCGCTAAACCCATGAATTACACGATCTGAATCTGAGTCCAGCAC
K H H K P *
AAAACGTGGGAATTTCATCGCTAAAAGCTAATACAAAAGCTAATATCGAGCCATTTTCGTATTTTGGACTCG
GGTATGCTTTAATGCCAGGATAACTTCTCCTTACAGAGCTGTTGCAGAAAAGATTTTACACTTTGCGATGA
M
CGTGTCCAAAGTCCAATCGTAAAACCTCGCAAGTGAGGATGCCACCAGGAATAAAGTACCATGTATGAG
T C P K S N R K T S Q V R M P T G I N * M Y E
M S
TTTAGATCAGGAGATCAGGAAATCCAAACGAAGGAATTATATAACCTTCTGGGAGGCAATAGTAAACACA
F R S G D Q E I Q T K E L Y N L L G G N S K H
L D Q E I R K S K R R N Y I T F W E A I V N T
TGTGAAGGCGCTAGACTTCACTGAGCAATCTCAAAGGAGTGACAGAATCAAATCAGGCACTATGAAAGG
M *
C E G A R L H *
AGGATAACTCTTATACTATAGGAAGTAAGGAGAACTTGAGCTACTACTGGCAACTCACTTTCCGGGAAG
CACGCTTCAAACGGATGGGAATACCCCGGTAAGTGCACAAAGTAGACCGCGTACTGATTGGGCCAAA
GCAAAATCACAAGTAACAACAGAGAGATTAGGATAGGCAAAATTTGTACATTCCAGCCCTACAAGTCTCCTG
GTCCAGATGACATTCAGCCAATCCTGCTTCAACGATCAGGAAAATACTGATCAGCAGCCTAGCTTTAGGA
M T F S Q S C F N D Q E N T D Q Q P S F R
CTCATACTACTGCCTGGAGTATAGCAAAAGTGGTCTTTCATTCGAAAGCTGGAAGAAAGACATGACTG
T H T Y C L E Y S K S G L H S E S W K E R H D *
M T
ATCCGAAGTCGTTTAGACCCATCAGCTTGATATCGTTTTTGGCTAAAACCTTGGAAAAGCTAGTGGATGTT

D_P_K_S_F_R_P_I_S_L_I_S_F_L_L_K_P_W_K_S_* M_L
 AGCATTAGAAGCACTCTGCTTGTGGAGCATCCGCTCCAAAGGACGCAGCATGCCTACAGGGCGGGCAGAT
A_L_E_A_L_C_L_W_S_I_R_S_K_G_R_S_M_P_T_G_R_A_D
 CAACTGATACTACCTTATATCACCTGAAAAGCCTTATTGAGGATTC AATGACTCATAAGGAAGTGACGTT
Q_L_I_L_P_Y_I_T_* M_T_H_K_E_V_T_L
 ATGTGCCTTCTTAGACATAACAGGGTGCCTTCGACAATACCTCCCATGAGGCTGTCAACGCATCCCTGGCA
C_A_F_L_D_I_Q_G_A_F_D_N_T_S_H_E_A_V_N_A_S_L_A
M_C_L_L_R_H_T_G_C_F_R_Q_Y_L_P_* M_R_L_S_T_H_P_W_Q
 AGAAGAGGACTAGATGCGACAACCAGCAGATGGATCAAATCACTACTGGCATCACTACAGTATCCACCAC
R_R_G_L_D_A_T_T_S_R_W_I_K_S_L_L_A_*
E_E_D_* M_R_Q_P_A_D_G_S_N_H_Y_W_H_H_Y_S_I_H_H
M_D_Q_I_T_T_G_I_T_T_V_S_T_T
 AAGGGGTTAACCAAGGGGGTGTCTCTCGCCTCTCCTGTGCAGCTTGCTGGACAGACTGACCCGCATGGG
K_G_L_T_K_G_V_S_S_R_L_S_C_A_A_C_W_T_D_*
R_G_*
 TATACTTTGCCAAGGCTATGCTGACCATATTGTCATTATAGCTAGCTTTAGGCCTGAGCCTAAATCCTGG
 CAAAAGTGTATTGTTCTTTTACAAAATAGGTACAAGCTACAGAGAATAAAGGCAATAACCCTGTCCAGAA
 TGTCGCATAGAGGTCAACAAAGAAGTAAAGTATCTTGGGATAACCCCTGACTCCAAACTTAGCTTCAAAA
M_S_H_R_G_Q_Q_R_S_*
 CTCATGTCATAACACAATTGATAAGTGCACCAGAGCAATTTACGTGTAGAAGCATTATCTCAGTAATAA
 AAACAGTTGGACATCCCTGTGGAAAGTAACCACAAAACCACAAATAAAGTGGTATACAGATGGATCCCTC
M_D_P_S
 ACCGACGAGGGAAGTGGGCTGGGGTGTAGGCCCATGTTGAAATACCACGAATCAATGGGCAGATACA
P_T_R_E_V_G_W_G_L_*
 CCAGCATTTTTCAAGCTGAAGCCTATGCTATTGGACGCTGTGCAGAGTTAATCTGCAAAGGAACCTATCG
M_L_L_D_A_V_Q_S_L_I_C_K_G_T_I
 TGACAAGGACATTGATATACTGTCTGATAGTCAAGCAGCCATAAAGGCGCTCAGCAAAGCTAAGATAACA
V_T_R_T_L_I_Y_C_L_I_V_K_Q_P_*
 TCTAAGCTAGTAAATGAAGTGAGAACAGCCCCAGACAAACTAGGAGCTGTCAACAAACTCACAATAAGGT
 GGGTCCCGGGACACAACAACATCCCGGGAAATGAGCTAGCGGACAACCTAGCCAGGAAGGGGCCAGAGAA
 CCCTCTAATTGGGCCCGAACATTTCTGTGGTGTGGTCCACCACAGAGTACTGGGCTTACTTAAGTCAATAG
 CGGAAGAAAAACGTCTGTCTTCTGGGAACACCTACCAGGACTTAGGCAGTCTAAGATTTCTCTCCGTGA
 ATATAACCATAAAAAGTTCAAGACCTTAATGACATACGGGAAAAACACCGTGCGCATTTTACTGGCCTT
 CTTACGGGGCATTGCCGACTTACAGCCATTTGCACAAGATTGGCATTGAAGACTTCTGCTGCACGGAGG
 AGGAGAGCTCAGCACATATTATATGTGACTGCATGGCGCTTCAATCAAGAGGAACAGATTCCTCGGCATG
M
 TATGTAGTCCCACGGGAAACAATTACCCCAATAAAAATTCTGGCATTATACAGTGCATTGGACTTCAGGG
Y_V_V_P_R_E_T_I_T_P_I_K_F_W_H_L_Y_S_A_L_D_F_R
 GGATCTTTGAGTCATGAAGGGGTAGTACAATAGATCTGACTGGGTCGCAGTACACAGACAAAATCACTTA
G_I_F_E_S_*
 ATAATAATACTTCATAAGTCAAGCTAAGAGGCCTGCTTCACAAAGTTTCAGCTTAGGAACAGTAATGGTC
M_V
 TTTAATGGACCAACACGAGATTTTCGAGCAATATAACTAGCCGGAACGCTTTCTCCTTTGGACACAATGT
F_N_G_P_T_R_D_F_E_Q_Y_N_*
M_D_Q_H_E_I_S_S_N_I_T_S_R_N_A_F_L_L_W_T_Q_C
 AAACACAAGCACCATAAGCCTTAAGGCTAGCATCGCTAAACCCATGAATTACAGATCTGAATCTGAGTC
K_H_K_H_H_K_P_*
 CAGCACAAAACGTGGGAATTCATCGCTAAAGCTAATACAAAAGCTAATATCGAGCCATTTTCGTATTTTGA
 GACTCGGGTATGCTTTAATGCCAGGATAACTTCTCCTTACAGAGCTGTTGCAGAAAGATTTTACACTTTG
 CGATGACGTGTCCAAAGTCCAATCGTAAAACCTCGCAAGTGAGGATGCCTACCAGGAATAAACTGACCATG
M_T_C_P_K_S_N_R_K_T_S_Q_V_R_M_P_T_G_I_N_* M
 TATGAGTTTATAGATCAGGAGATCAGGAAATCCAAACGAAGGAATTATATAACCTTCTGGGAGGCAATAGTA
Y_E_F_R_S_G_D_Q_E_I_Q_T_K_E_L_Y_N_L_L_G_G_N_S
M_S_L_D_Q_E_I_R_K_S_K_R_R_N_Y_I_T_F_W_E_A_I_V
 AACACATGTGAAGGCGCTAGACTTCACTGAGCAATCTCCAAAGGAGTGACAGAATCAAATCAGGCACTAT
K_H_M_*
N_T_C_E_G_A_R_L_H_*
 GAAAGGAGGATAACTCTTATACTATAGGAAGTAAGGAGAACTTGAGCTACTACTGGCAACTCACTTTCC
 GGAAGCACGCTTCAAACGGATGGGAATACCCCGGTAAGTGCACAAAGTAGACCGCGTACTGATTGG

GCCAAAGCAAAATCACAAAGTAACAACAGAGAGATTAGGATAGGCAAATTGTACATTCCAGCCCTACAAGT
 CTCCTGGTCCAGATGACATTCAGCCAATCCTGCTTCAACGATCAGGAAAATACTGATCAGCAGCCTAGCT
 M T F S Q S C F N D Q E N T D Q Q P S
 TTAGGACTCATACCTACTGCCTGGAGTATAGCAAAAGTGGTCTTCATTCCGAAAGCTGGAAAGAAAGACA
 F R T H T Y C L E Y S K S G L H S E S W K E R H
 TGACATGACTGATCCGAAAGTCGTTTAGACCCATCAGCTTGATATCGTTTTTGCTAAAACCTTGGA
 D *
 M T D P K S F R P I S L I S F L L K P W K S *

Para el gen 2R_96, el programa Augustus Gene Prediction predijo una estructura compuesta por 13 exones de 425 pares de bases el primero, 175 pares de bases el segundo, 186 pares de bases el tercero, 212 pares de bases el cuarto, 175 pares de bases el quinto, 217 pares de bases el sexto, 88 pares de bases el séptimo, 175 pares de bases el octavo, 217 pares de bases el noveno, 88 pares de bases el décimo, 175 pares de bases el decimoprimer, 107 pares de bases el decimosegundo, y 620 pares de bases el decimotercero, y 12 intrones de 2023pares de bases el primero, 507 pares de bases el segundo, 2317 pares de bases el tercero, 2023 pares de bases el cuarto, 1756 pares de bases el quinto, 1626 pares de bases el sexto, 113 pares de bases el séptimo, 1756 pares de bases el octavo, 1626 pares de bases el noveno, 113 pares de bases el décimo, 153 pares de bases el decimoprimer, y 53 pares de bases el decimosegundo.

Se encontró acuerdo con espectros de masas para el péptido DQITTGITTVSTTR con una probabilidad de $\rho=0.042$ y un valor $E=0.000102560808637565$; para el péptido TLIYCLIVK con una probabilidad de $\rho=0.058$ y un valor $E=0.000731119522471778$; para el péptido SLDQEIRK con una probabilidad de $\rho=0.002$ y un valor $E=0.000191118927747795$; para el péptido MSITQLINAPEHFSR con una probabilidad de $\rho=0.035$ y un valor $E=9.18080278409815e-05$; para el péptido SITQLINAPEHFSR con una probabilidad de $\rho=0.016$ y un valor $E=9.66773360525004e-05$; para el péptido ELYNLLGGNSKHM con una probabilidad de $\rho=0.007$ y un valor $E=0.000134877430120634$; para el péptido RQPADGSNHYWHYSIHHK con una probabilidad de $\rho=0.006$ y un valor $E=0.000917922108684765$; para el péptido SPSQGTQLQSSPTKK con una probabilidad de $\rho=0.054$ y un valor $E=0.000734166497450426$.

En esta región genómica se observa la pauta de sORFs separados por secuencias pequeñas (<2 kbp) como las que se esperan para mensajeros policistrónicos basados en *tal*. En este caso se observan una serie de 121 sORFs. Se pueden observar repeticiones. No se observa presencia de ESTs, con lo cual no se halla evidencia de transcripción. La probable traducción se basa en la presencia en bases de datos de 8 péptidos definidos por espectrometría de masas que cumplirían con los requisitos mínimos establecidos en este trabajo ($E < 0,001$) para su validación.

La falta de evidencia de traducción de los demás sORFs no significa que no sean traducidos. Para validarlos, sería necesaria la realización de un análisis extensivo por espectrometría de masas que excede el objetivo de este trabajo final de licenciatura, pero que forma parte del plan de tesis que se llevará a cabo más adelante. Por lo tanto se considera que esta es una posible región expresada del genoma y aún no anotada.

DISCUSIÓN Y CONCLUSIONES

Los resultados de este trabajo demuestran que ORFs no canónicos (< 100 codones= sORFs) efectivamente existen en el cromosoma 2 de *Drosophila melanogaster*, codificando pequeños péptidos funcionales. Los sORFs analizados en el presente estudio están codificados por ARNs mensajeros muy posiblemente policistrónicos, separados por secuencias pequeñas (<2 kpb), en donde la disposición de genes en un grupo funcional constituye un operón porque opera como una unidad a partir de un único promotor, al igual que los operones encontrados en organismos procariotas. De todos modos no se puede descartar *a priori* que dichos sORFs puedan formar parte de transcritos independientes. De todos modos la disposición en clusters sugiere fuertemente una estructura similar a la descrita para *tal*.

El análisis bioinformático utilizado, el cual es novedoso, permitió descubrir 33 loci aún no anotados y con una probable estructura génica coherente. En cada una de esas regiones existen diferentes sORFs codificantes de pequeños péptidos, muchos de los cuales poseen evidencia de expresión de acuerdo a lo encontrado en las bases de datos de espectros de masas y, por otro lado, en bases de datos de ESTs. La evidencia en espectros de masas establece la probable traducción de dichos sORFs, y la evidencia de ESTs hace referencia a la evidencia de transcripción. La ausencia de evidencia de traducción de los demás sORFs no significa que no sean traducidos, pero con el criterio de selección estricta usada en este trabajo, no fueron analizados.

Estos resultados están en concordancia con los descritos para el gen *tarsal-less* (8) descubierto en el genoma de *Drosophila melanogaster*. La función de *tal* es mediada por varios ORFs de 33 nucleótidos, siendo de esta manera un sORF policistrónico.

Cuatro de los cinco péptidos codificados por *tal* poseen una secuencia muy similar, tal como observamos en nuestros genes, conteniendo un motivo proteico conservado

(LDPTGXY). La existencia de sólo uno de estos cuatro péptidos alcanza para proveer la función completa de *tal*. La presencia de estos sORFs repetidos es una estrategia para asegurar una traducción suficiente de péptidos que contengan el motivo LDPTGXY. En el presente trabajo se observan también repeticiones de sORFs que podrían cumplir una función similar a la de los péptidos repetidos en *tal*. De esta manera, se demuestra que los pequeños péptidos codificados por sORFs efectivamente existen en el genoma de *Drosophila melanogaster*, pudiendo cumplir funciones biológicas de vital importancia. En este trabajo se sientan las bases para el análisis funcional de cada uno de estos genes. Todos los genomas secuenciados y anotados contienen genes y transcritos sin función conocida. En particular, un número creciente de transcritos están siendo clasificados como no codificantes (ncRNA) sobre la base de no tener ORFs evidentes. Sin embargo esto se basa en la incapacidad de los sistemas de anotación automática de detectar dichos ORFs. En este sentido, este trabajo es una importante contribución para el descubrimiento de nuevos genes con métodos de predicción novedosos. Es destacable que los genomas contienen cientos de miles de sORFs potenciales que son sistemáticamente eliminados de las anotaciones génicas por razones estadísticas. Bibliotecas de ADN complementario (ADNc) y colecciones de ESTs también discriminan contra pequeños ADN complementarios, perdiendo así muchos transcritos potenciales. Es por eso, por ejemplo, que los micro RNAs solo fueron descubiertos muy recientemente y, tal como en el caso de los sORFs su predicción es tan compleja. Los pequeños péptidos conocidos cumplen importantes funciones biológicas, mayormente en la comunicación neural y endócrina, pero en todos los casos descriptos esos péptidos son productos maduros, clivados de un ORF mayor. En otros casos, como el de la insulina o histonas, para dar dos ejemplos, la anotación se basó en el conocimiento previo de las proteínas, antes de la era de la genómica. La originalidad de

el trabajo presentado aquí radica en la comprobación de la existencia de péptidos menores a 100 aminoácidos en el genoma de *Drosophila melanogaster*, y en que tales péptidos son codificados por ARN mensajeros policistrónicos, donde varios ORFs se traducirían a partir de una única molécula de ARN mensajero.

Este trabajo demuestra que pequeños péptidos codificados por sORFs pueden ser descubiertos y analizados utilizando herramientas bioinformáticas, con una metodología novedosa y menos restrictiva para la búsqueda de genes.

Ha sido demostrado que sORFs existen en *Drosophila* y en *S. cerevisiae*, lo que abre la posibilidad de la existencia de estos genes en vertebrados superiores. El número de genes codificantes de pequeños péptidos en el genoma de *Drosophila* es elevado, lo cual demuestra la probabilidad de la presencia de muchos más genes que los descubiertos en este trabajo. En los procesos de anotación existe una cantidad creciente de transcritos clasificados como no codificantes, y por lo tanto, son pasados por alto, sin ser analizada su posible traducción. Una búsqueda sistemática de péptidos codificados por sORFs permitirá el descubrimiento de nuevos genes. Esta fase del trabajo excede el objetivo de un trabajo final de licenciatura por su complejidad y extensión, pero a la vez, este trabajo permite iniciar estudios funcionales complejos en este y otros organismos.

Bibliografía citada

- (1) Lacadena Juan Ramón, Citogenética, 1ª Edición, marzo 1996, Editorial complutense.
- (2) Lacadena Juan Ramón, Citogenética, 1ª Edición, marzo 1996, Editorial complutense.
- (3) Crollius, H. R., et al, 2000, Estimate of Human Gene Number Provided by Genome-Wide Analysis Using *Tetraodon nigroviridis* DNA Sequence, *Nature Genetics* **25**, 235-38; Ewing, B. and P. Green, 2000, Analysis of Expressed Sequence Tags Indicates 35,000 Human Genes, *Nature Genetics* **25**, 232-34; Liang, F., et al, 2000, Gene Index Analysis of the Human Genome Estimates Approximately 120,000 Gene, *Nature Genetics* **25**, 239-40.
- (4) Kastenmayer James P., Ni Li Chu Angela, Kitchen Lauren E., Au Wei-Chun, Yang Hui, Carter Carole D., Wheeler David, Davis Ronald W., Boeke Jef D., Snyder Michael A. and Basrai Munira A., 2006, Functional genomics of genes with small open reading frames (sORFs) in *S. Cerevisiae*, 16:365–373 ISSN 1088-9051/06.
- (5) Yang Xiaohan, J. Tschaplinski Timothy, Hurst Gregory B., Jawdy Sara, Hettich Robert L., Lindquist Erika, Kalluri Udaya C., Gunter Lee E., Pennacchio Christa and Abraham Paul E., Lankford Patricia K., Adams Rachel M., Shah Manesh B., Tuskan Gerald A., 2011, Discovery and annotation of small proteins using genomics, proteomics, and computational approaches, *Genome Research* 21:000–000 ISSN 1088-9051/11.
- (6) Galindo Máximo Ibo, Pueyo José Ignacio, Fouix Sylvaine, Bishop Sarah Anne and Couso Juan Pablo, 2007, Peptides Encoded by Short ORFs Control Development and Define a New Eukaryotic Gene Family, *PLoS Biology* 5(5): e106. doi:10.1371/journal.pbio.0050106.
- (7) Pueyo Jose Ignacio, Couso Juan Pablo, The 11-aminoacid long Tarsal-less peptides trigger a cell signal in *Drosophila* leg development, 2008, *Developmental Biology* 324 192-201 doi:10.1016/j.ydbio.2008.08.025.
- (8) Kondo T., Plaza S., Zanet J., Benrabah E., Valenti P., Hashimoto Y., Kobayashi S., Payre F., Kageyama Y., Small Peptides Switch the Transcriptional Activity of Shavenbaby During *Drosophila* Embryogenesis, 2010, *Science* Vol. 329 336-339.

RESUMEN

El número de genes que codifican proteínas en eucariotas superiores identificados por proyectos de genoma es más bajo que lo esperado, mientras que ha sido detectada una considerable cantidad de transcritos putativamente no codificantes. Esto refleja, nuevamente, las limitaciones de los sistemas predictivos. Pero en caso de ser correctas, ¿cómo es posible que la enorme variabilidad fenotípica de los organismos tenga como origen un pequeño número de genes mayormente conservados? Varias respuestas han sido propuestas: interacciones génicas complejas, splicing alternativo, y genes no codificantes. Otra posibilidad menos discutida podría ser muchos genes no canónicos y por lo tanto no identificados, escapen a la detección de los programas de anotación génica.

Algunos métodos de detección génica y anotación permiten encontrar genes con sORFs (pequeños marcos abiertos de lectura, del inglés *small open reading frames*) que codificarían pequeños péptidos, los cuales esperan caracterización. En el presente trabajo se analizó el genoma del cromosoma 2 de *Drosophila melanogaster* en busca de sORFs codificados en ARNs policistrónicos, descubriéndose que los sORFs efectivamente existen en el genoma de *Drosophila*. Se utilizaron métodos bioinformáticos novedosos que permitieron detectar dichos genes.

ANEXOS

Abreviatura de aminoácidos utilizada:

A Alanina	C Cisteína	D Aspartato
E Glutamato	F Fenilalanina	G Glicina
H Histidina	I Isoleucina	K Lisina
L Leucina	M Metionina	N Asparagina
P Prolina	Q Glutamina	R Arginina
S Serina	T Treonina	V Valina
W Triptofano	Y Tirosina	* Stop