



**UNIVERSIDAD DEL PAPALOAPAN**

**INGENIERÍA EN BIOTECNOLOGÍA  
CAMPUS TUXTEPEC**

**“Estudio bioinformático de enzimas transglutaminasas vegetales  
y el aislamiento parcial de un gen *tg* en caña de azúcar  
(*Saccharum* sp. híbrido Mex 69-290)”**

**TESIS**

**QUE PARA OBTENER EL TÍTULO DE:**

**Ingeniero en Biotecnología**

**PRESENTA:**

**JOAN HAZAEL PÉREZ FIGUEROA**

**Director: Dr. Enrique Villalobos Amador**

**Co-Director: Dr. José Francisco Morales Domínguez**

**(Universidad Autónoma de Aguascalientes)**

**SAN JUAN BAUTISTA TUXTEPEC, OAXACA**

**2019**



# UNIVERSIDAD DEL PAPALOAPAN

## CAMPUS TUXTEPEC

### ACTA DE REVISIÓN DE TESIS

En la ciudad de San Juan Bautista Tuxtepec, Oaxaca, el día 23 de octubre de 2019 a las 12 hr, los miembros de la comisión revisora de tesis designada por la Jefatura de Carrera de la Ingeniería en Biotecnología se reunieron en la sala de juntas del Instituto de Biotecnología de la Universidad del Papaloapan, con la finalidad de examinar la tesis titulada "Estudio bioinformático de enzimas transglutaminasas vegetales y el aislamiento parcial de un gen *tg* en caña de azúcar (*Saccharum sp.* híbrido Mex 69-290" presentada por el alumno Joan Hazael Pérez Figueroa, con número de matrícula 10090020, aspirante al título de Licenciatura.


Después de intercambiar opiniones, los miembros de la comisión manifestaron que la tesis **satisface** los requisitos señalados por las disposiciones reglamentarias vigentes, otorgando su **aprobación** para que el aspirante pueda proceder con el proceso de titulación.

Tuxtepec, Oaxaca, a 23 de octubre de 2019

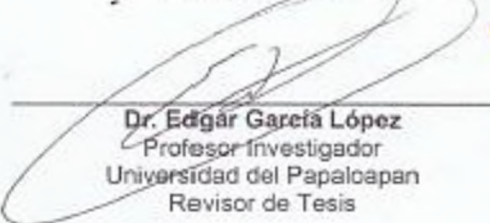
ATENTAMENTE  
LA COMISIÓN REVISORA


  
\_\_\_\_\_  
**Dr. Enrique Vilalobos Amador**  
Profesor Investigador Titular "B"  
Universidad del Papaloapan  
Director de Tesis

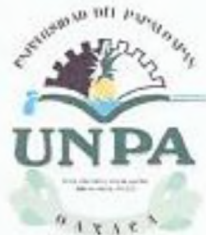
  
\_\_\_\_\_  
**Dra. Blanca Estela Barrera Figueroa**  
Profesor Investigador  
Universidad del Papaloapan  
Revisor de Tesis

  
\_\_\_\_\_  
**Dr. Miguel Angel Peña Rico**  
Profesor Investigador  
Universidad del Papaloapan  
Revisor de Tesis

  
\_\_\_\_\_  
**Dr. Edgar Baldemar Sepúlveda García**  
Profesor Investigador  
Universidad del Papaloapan  
Revisor de Tesis

  
\_\_\_\_\_  
**Dr. Edgar García López**  
Profesor Investigador  
Universidad del Papaloapan  
Revisor de Tesis

  
\_\_\_\_\_  
**Dr. Francisco Noé Mendoza Ambrosio**  
Profesor Investigador  
Universidad del Papaloapan  
Revisor de Tesis



# UNIVERSIDAD DEL PAPALOAPAN

## CAMPUS TUXTEPEC

Tuxtepec, Oaxaca, a 24 de octubre de 2019  
Oficio No. JCIB/007/10/2019

**Lic. Yesenia Barrientos Arenal**  
Jefe de Servicios Escolares  
Universidad del Papaloapan

Con base en el dictamen de la comisión revisora, se autoriza la impresión del trabajo de tesis del alumno **Joan Hazael Pérez Figueroa** titulado "**Estudio bioinformático de enzimas transglutaminasas vegetales y el aislamiento parcial de un gen *tg* en caña de azúcar (*Saccharum sp.* híbrido Mex 69-290)**". Para ser presentado como trabajo de tesis para obtener el título de Licenciado en Ingeniería en **Biología**, toda vez que cumple satisfactoriamente con la reglamentación establecida para tal fin.

El Jurado de Examen Profesional estará compuesto por los siguientes profesores:

Dra. Blanca Estela Barrera Figueroa (Presidente)  
Dr. Miguel Ángel Peña Rico (Vocal)  
Dr. Edgar Baldemar Sepúlveda García (Secretario)  
Dr. Edgar García López (Primer Suplente)  
Dr. Francisco Noé Mendoza Ambrosio (Segundo Suplente)

Sin más por el momento le envío un cordial saludo.

**Atentamente**  
*Terra uberrima, mens aperta*  
*Bou Lo-tama, chí jí jú*



UNPA  
CAMPUS TUXTEPEC

**Dr. Julián Mario Peña Castro**  
Jefe de Carrera de Ingeniería en  
Biología  
Universidad del Papaloapan



**Vo.Bo. M.C. Héctor López Arjona**  
Vice Rector Académico  
Universidad del Papaloapan  
VICE-RECTORIA  
ACADEMICA

c.c.p. Dr. Enrique Villalobos Amador, Director de tesis, Para su conocimiento  
c.c.p. Joan Hazael Pérez Figueroa, Alumno, Para su conocimiento  
c.c.p. Archivo

## HOJA DE ORIGINALIDAD

El presente trabajo no ha sido aceptado o empleado para el otorgamiento de título o grado diferente o adicional al actual. La tesis es resultado de las investigaciones del autor, excepto donde se indican las fuentes de información consultadas. El autor otorga su consentimiento a la **Universidad del Papaloapan** para la reproducción del documento con el fin del intercambio bibliotecario siempre y cuando se indique la fuente.

## **AGRADECIMIENTOS**

A mis padres y mis hermanos que en todo momento fueron mi apoyo incondicional y de los que me siento muy orgulloso, gracias por todo.

Al Dr. Enrique Villalobos Amador, director de esta tesis, no solo por la oportunidad brindada para realizar este trabajo sino también por las enseñanzas que me apoyaron en todo momento. Su dedicación y empeño mostrado para concluir con este proyecto.

Al Dr. José Francisco Morales Domínguez, co-director de esta tesis, por las facilidades otorgadas para llevar a cabo parte de mi trabajo en la Universidad Autónoma de Aguascalientes en la realización del análisis molecular. Además de mi admiración y de su paciencia otorgada, su buen sentido del humor que me brindó como persona, así como la dedicación y el esfuerzo que alentó en todo momento la pasión por esta área. A la Dra. Cristina Garcidueñas Piña por sus enseñanzas y la paciencia otorgada que me hicieron aprender de mis errores.

A todos y cada uno de los amigos que encontré a lo largo de mi estancia en el edificio 60. A Mariana y Vivian por inducirme en esta área y brindarme parte de sus conocimientos. A Daniel y al Dr. Josafat por corregirme y compartir conmigo sus conocimientos y experiencias en el laboratorio. A Lili, Carito, Fanny. Gracias por su amistad y los buenos momentos que compartieron conmigo y que fueron parte de este proyecto.

A los profesores: Dra. Blanca Estela, Dr. Miguel Ángel, Dr. Edgar Baldemar, Dr. Edgar García y Dr. Francisco Noé por el tiempo dedicado a la revisión de esta tesis ya que con sus aportes individuales enriquecieron el contenido de este proyecto.

**A MIS PADRES**

Por su apoyo incondicional, nada de esto sería posible sin ustedes.

**A MIS HERMANOS**

Jamin, Yarah y Sebas por ser parte de mi vida y un motivo de superación constante.

**A TI, MORENA**

## CONTENIDO

COPIA DE ACTA DE REVISIÓN DE TESIS .....	II
COPIA DE OFICIO DE AUTORIZACIÓN DE TESIS .....	III
HOJA DE ORIGINALIDAD .....	IV
AGRADECIMIENTOS .....	V
ÍNDICE DE FIGURAS.....	X
ÍNDICE DE TABLAS .....	XII
ABREVIATURAS .....	XIII
RESUMEN.....	XVI
ABSTRACT .....	XVII
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>2. MARCO TEÓRICO .....</b>	<b>3</b>
2.1 Bioinformática.....	3
2.2 Transglutaminasas .....	5
2.3 Transglutaminasas microbianas.....	8
2.4 La familia TG en mamíferos.....	10
2.5 Transglutaminasas vegetales .....	13
2.5.1 El gen <i>AtPng1p</i> de <i>Arabidopsis thaliana</i> .....	15
2.5.2 El gen <i>tgo</i> de arroz ( <i>Oryza sativa</i> ) .....	16
2.5.3 Antecedentes del gen <i>tgz15/tgz21</i> de maíz ( <i>Zea mays</i> ) .....	16
<b>3. JUSTIFICACIÓN .....</b>	<b>20</b>
<b>4. HIPÓTESIS .....</b>	<b>21</b>
<b>5. OBJETIVOS .....</b>	<b>22</b>
5.1 Objetivo principal.....	22
5.2 Objetivos secundarios.....	22
<b>6. MATERIALES Y MÉTODOS .....</b>	<b>23</b>
6.1 Búsqueda y distribución de secuencias biológicas anotadas funcionalmente como transglutaminasas en las bases de datos biológicas.....	23
6.2 Identificación y obtención de secuencias biológicas anotadas funcionalmente como transglutaminasas de maíz ( <i>Zea mays</i> ).....	24
6.3 Búsqueda de proteínas similares a las transglutaminasas de maíz ( <i>Zea mays</i> ).....	24
6.4 Manipulación y alineamiento de secuencias biológicas .....	24
6.5 Reconocimiento de patrones de conservación en regiones biológicas funcionales de las transglutaminasas de maíz ( <i>Zea mays</i> ).....	25
6.6 Construcción de árboles filogenéticos.....	25
6.7 Predicción de la estructura proteica mediante técnicas de enhebrado y <i>ab initio</i> .....	25

6.8	Análisis de enriquecimiento con anotaciones GO obtenidas desde Ffpred3 a partir de las TGs de maíz ( <i>Zea mays</i> ) .....	27
6.9	Análisis Molecular .....	30
6.9.1	Material vegetal .....	30
6.9.2	Cepa bacteriana y vector de clonación .....	30
6.9.3	Extracción de ADN de maíz ( <i>Zea mays</i> ) y caña de azúcar ( <i>Saccharum</i> sp. híbrido Mex 69-290) .....	30
6.9.4	Verificación de la integridad y cuantificación de la concentración de ADN obtenido .....	31
6.9.5	Diseño de oligonucleótidos para la identificación del gen <i>tg</i> en caña de azúcar ( <i>Saccharum</i> sp. híbrido Mex 69-290) .....	31
6.9.6	Amplificación por PCR del gen <i>tg</i> en caña de azúcar ( <i>Saccharum</i> sp. híbrido Mex 69-290) .....	32
6.9.7	Purificación y clonación de los fragmentos amplificados .....	33
6.9.8	Extracción de plásmidos y verificación de ligación por enzimas de restricción .....	33
6.9.9	Secuenciación y análisis de secuencias .....	34
<b>7.</b>	<b>RESULTADOS Y DISCUSIÓN.....</b>	<b>35</b>
7.1	Análisis comparativo de la distribución y organización de secuencias biológicas anotadas funcionalmente como transglutaminase en las bases de datos biológicas.....	35
7.1.1	Identificación de secuencias biológicas vegetales con anotación funcional transglutaminase. ....	38
7.2	Identificación y selección de proteínas transglutaminasas de maíz ( <i>Zea mays</i> ) .....	41
7.2.1	Análisis de secuencias seleccionadas de las transglutaminasas de maíz ( <i>Zea mays</i> ) .....	42
7.2.2	Reconocimiento de regiones funcionales dentro de las transglutaminasas de maíz ( <i>Zea mays</i> ) .....	43
7.3	Análisis de similitud con proteínas similares a las TGs de maíz.....	45
7.3.1	Alineamiento de la región N-terminal con proteínas similares a la TGZ15 y TGZi .....	45
7.3.2	Alineamiento de la región C-terminal con proteínas similares a la TGZ15 .....	47
7.3.3	Alineamiento de la región C-terminal con proteínas similares a la TGZi .....	51
7.3.4	Alineamiento de proteínas similares a TGZc .....	53
7.3.5	Alineamiento de proteínas similares a TGZc2 .....	54
7.3.6	Análisis del núcleo catalítico de la proteína TGZ15 con otras proteínas TGs.....	55
7.4	Análisis filogenético de las proteínas TGs de maíz.....	60
7.4.1	Árbol filogenético de la familia de proteínas similares a TGZ15 y TGZi .....	60
7.4.2	Árbol filogenético de proteínas similares a TGZc .....	62
7.4.3	Árbol filogenético de proteínas similares a TGZc2 .....	64
7.4.4	Árbol filogenético de la familia TG en diversas especies y su relación con la TGZ15.....	65
7.5	Predicción del modelo estructural de las proteínas TGs de maíz por la técnica de enhebrado y <i>ab initio</i> .....	67
7.5.1	Modelo estructural de la proteína TGZ15 y su isoforma inactiva TGZi .....	68
7.5.2	Modelo estructural de la proteína TGZc .....	73
7.5.3	Modelo estructural de la proteína TGZc2 .....	75
7.6	Análisis de enriquecimiento con anotaciones GO de las proteínas de interés.....	76
7.6.1	Procesos biológicos atribuidos por ontologías en las proteínas TGs de maíz.....	76

7.6.2	Funciones moleculares atribuidos por ontologías en las proteínas TGs de maíz .....	79
7.7	Análisis molecular .....	82
7.7.1	ADN extraído de hojas de maíz ( <i>Zea mays</i> ) y caña de azúcar ( <i>Saccharum</i> sp. híbrido Mex 69-290) .....	82
7.7.2	Amplificación del gen tgz15 de maíz ( <i>Zea mays</i> ) como control positivo .....	82
7.7.3	Aislamiento del gen tg de caña de azúcar ( <i>Saccharum</i> sp. híbrido Mex 69-290) mediante PCR .....	83
7.7.4	Clonación del gen tg de caña de azúcar ( <i>Saccharum</i> sp. híbrido Mex 69-290).....	84
7.7.5	Análisis del fragmento del gen tg de caña de azúcar ( <i>Saccharum</i> sp. híbrido Mex 69-290) obtenido por secuenciación .....	85
<b>8.</b>	<b>CONCLUSIONES .....</b>	<b>88</b>
<b>9.</b>	<b>PERSPECTIVAS Y RECOMENDACIONES .....</b>	<b>90</b>
<b>10.</b>	<b>REFERENCIAS.....</b>	<b>91</b>
<b>11.</b>	<b>ANEXOS.....</b>	<b>110</b>
11.1	Preparación de células competentes .....	110
11.2	Transformación de células competentes .....	110
11.3	Extracción de plásmidos (MINIPREP Bimboim Daily).....	110
11.4	Contexto genómico de los genes anotados con el término transglutaminase de maíz .....	112
11.5	Alineamiento pareado entre las proteínas TGZ15 y TGZi (B8A2N4 y COHGH1) .....	113
11.6	Diagramas tipo logos de TRs y péptido tránsito al cloroplasto.....	113
11.7	Secuencias de proteínas referencia utilizadas en el análisis bioinformático.....	114
11.8	Tabla de las secuencias homólogas a TGZ15 y TGZi.....	115
11.9	Tabla de las secuencias homólogas a TGZc.....	116
11.10	Tabla de las secuencias homólogas a TGZc2 .....	117
11.11	Alineamiento de motifs similares al patrón de expresión: H-[IT]-[TS]-[LTR]-[SGC]-P-[DTS]-[LS]-[SG]-H a partir de proteínas similares a TGZ15.....	118
11.12	Alineamiento de motifs similares al patrón de expresión: [HY]-I-S-R-[EGV]-[PLA]-[RS]-[LM]-[GE] proveniente de los TRs de la proteína TGZ15 .....	119
11.13	Alineamiento de motifs similares al patrón de expresión: Q-A-G-[AVL]-[VYA]-Q-[PQ]-G a partir de proteínas similares a TGZi.....	120
11.14	Frecuencia de aminoácidos presentes en las proteínas completa y el sitio activo de las enzimas.. .....	121
11.15	Tablas de las proteínas que presentan el sitio activo de la TG y la posición de los residuos involucrados en la tríada catalítica .....	122
11.16	Resultados de las anotaciones enriquecidas en las proteínas anotadas funcionalmente como TGs de maíz de acuerdo al criterio de “Procesos Biológicos” .....	124
11.17	Resultados de las anotaciones enriquecidas en las proteínas anotadas funcionalmente como TGs de maíz de acuerdo al criterio de “Función Molecular” .....	125
11.18	Análisis de similitud BLASTn en las bases de datos CIRAD del genoma de caña de azúcar .....	126

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Actividad de transamidación catalizada por la TG. ....	7
<b>Figura 2.</b> Estructura general de la TGm (PDB ID: 3IU0).....	10
<b>Figura 3.</b> Estructura general de la TG2h.....	12
<b>Figura 4.</b> Modelamiento por homología entre la enzima de 75 kDa de <i>H. tuberosus</i> y la TG2h.....	14
<b>Figura 5.</b> Homología de proteínas entre la familia <i>Poaceae</i> . ....	17
<b>Figura 6.</b> Alineamiento múltiple global de secuencias con el sitio catalítico de las TGs. ....	18
<b>Figura 7.</b> Distribución cuantitativa de secuencias de nucleótidos y aminoácidos con anotación transglutaminase en las bases de datos de la NCBI.....	35
<b>Figura 8.</b> Organización de secuencias proteicas con anotación transglutaminase en las bases de datos de la NCBI protein. ....	36
<b>Figura 9.</b> Organización de secuencias proteicas con anotación transglutaminase en la base de datos UniProtKB. ....	37
<b>Figura 10.</b> Distribución de secuencias de nucleótidos y aminoácidos de origen vegetal con anotación transglutaminase en las bases de datos de la NCBI gene y NCBI protein. ....	38
<b>Figura 11.</b> Representación porcentual del número de proteínas en familias vegetales con anotación funcional transglutaminase.....	40
<b>Figura 12.</b> Alineamiento de las secuencias codificantes <i>tg</i> de maíz delimitadas por la zona de mutación.....	42
<b>Figura 13.</b> Diferencia entre la composición de los aminoácidos de las secuencias TGZ15 activa, B8A2N4 y COHGH1 (TGZi) de maíz. ....	43
<b>Figura 14.</b> Diagrama esquemático de los elementos biológicos funcionales hallados en las proteínas anotadas como transglutaminasas en maíz.....	44
<b>Figura 15.</b> Alineamiento de la región N-terminal a partir de una secuencia consenso.....	47
<b>Figura 16.</b> Alineamiento de la región C-terminal e identificación de regiones conservadas y recurrentes de las proteínas homólogas a TGZ15.. ....	49
<b>Figura 17.</b> Identificación de logos consenso de TRs en el C-terminal de proteínas similares a TGZ15.. ....	50
<b>Figura 18.</b> Alineamiento de la región C-terminal e identificación de regiones conservadas y recurrentes de las proteínas homólogas a TGZi. ....	52
<b>Figura 19.</b> Identificación de logos consenso de TRs en el C-terminal de proteínas similares a TGZi. ....	53
<b>Figura 20.</b> Alineamiento de la proteína TGZc.....	54
<b>Figura 21.</b> Alineamiento de la proteína TGZc2. ....	55
<b>Figura 22.</b> Alineamiento del núcleo conservado de la TG .....	59
<b>Figura 23.</b> Árbol filogenético de las proteínas TGZ15 y TGZi en especies vegetales de la familia <i>Poaceae</i> . ....	62
<b>Figura 24.</b> Árbol filogenético de la proteína TGZc. ....	64
<b>Figura 25.</b> Árbol filogenético de la proteína TGZc2. ....	64
<b>Figura 26.</b> Árbol filogenético de proteínas con dominio TGc .....	66
<b>Figura 27.</b> Cobertura, confiabilidad e información de las secuencias templado utilizados en la construcción del modelo proteico de TGZ15 y TGZi.....	69
<b>Figura 28.</b> Perfil de confiabilidad con el nivel de desorden intrínseco de las proteínas TGZi y TGZ15. ....	70
<b>Figura 29.</b> Modelo de la estructura tridimensional de las proteínas TGZ15 y TGZi.....	71
<b>Figura 30.</b> Perfil Verify3D y gráfico de Ramachandran para el análisis de la confiabilidad de los modelos proteicos TGZ15 y TGZi. ....	73

<b>Figura 31.</b> Modelo parcial de la proteína TGZc.....	74
<b>Figura 32.</b> Perfil de confiabilidad con el nivel de desorden intrínseco de la proteína TGZc.....	75
<b>Figura 33.</b> Modelo parcial de la proteína TGZc2.....	75
<b>Figura 34.</b> Perfil de confiabilidad con el nivel de desorden intrínseco de la proteína TGZc2.....	76
<b>Figura 35.</b> Gráfico tipo mapa de calor representando las anotaciones de procesos biológicos relacionados a proteínas anotadas funcionalmente como TGs de maíz.....	78
<b>Figura 36.</b> Gráfico tipo mapa de calor representando las anotaciones de funciones moleculares relacionados a proteínas anotadas funcionalmente como TGs de maíz.....	80
<b>Figura 37.</b> Electroforesis en gel de agarosa al 1% de ADN a partir de tejido foliar de maíz y caña de azúcar diluidas a 100 ng.....	82
<b>Figura 38.</b> Amplificación y purificación del gen tgz15 en maíz por PCR.....	83
<b>Figura 39.</b> Amplificación y purificación del gen tg en caña de azúcar por PCR.....	84
<b>Figura 40.</b> Verificación de la ligación del fragmento del gen tg de caña de azúcar por la enzima de restricción EcoRI.....	85
<b>Figura 41.</b> Esquema de la secuencia obtenida del producto de PCR clonado del gen tg en caña de azúcar.....	86
<b>Figura 42.</b> Análisis comparativo de secuencias tg codificantes de <i>Zea mays</i> y <i>Saccharum</i> sp. híbrido Mex 69-290 provenientes de ADN y ARNm.....	87

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Clasificación de proteínas por tipos de evidencias registradas en la UniProtKB. ....	5
<b>Tabla 2.</b> Proteínas miembros de la familia TG en humanos. ....	11
<b>Tabla 3.</b> Lista de genes <i>tg</i> de origen vegetal secuenciados y caracterizados hasta el momento. ....	15
<b>Tabla 4.</b> Oligonucleótidos utilizados para la amplificación del gen <i>tg</i> en caña de azúcar. ....	32
<b>Tabla 5.</b> Componentes de la mezcla de reacción para la PCR. ....	32
<b>Tabla 6.</b> Componentes de reacción para ligación de productos de PCR. ....	33
<b>Tabla 7.</b> Proteínas seleccionadas .....	41
<b>Tabla 8.</b> Porcentaje de residuos encontrados en el C-terminal respecto al total de aminoácidos de la proteína completa de TGZ15 y TGZi. ....	51

## ABREVIATURAS

Cys (C). Cisteína  
His (H). Histidina  
Asp (D). Aspartato  
Asn (N). Asparagina  
FXIIIa. Factor 13a de Coagulación sanguínea  
TG2h. Transglutaminasa 2 humana  
TGm. Transglutaminasa microbiana  
kDa. Kilodaltones  
PDB. Protein Data Bank  
EMBL. European Molecular Biology Laboratory  
DDBJ. DNA Data Bank of Japan  
PIR. Protein Information Resource  
RefSeq. Reference Sequence Database  
ADN. Ácido Desoxirribonucleico  
Ca. Calcio  
Gln (Q). Glutamina  
Glu (E). Ácido Glutámico  
Trp (W). Triptófano  
PDI. Protein Disulfide Isomerase  
RMN. Resonancia Magnética Nuclear  
proTGm. proenzima de la Transglutaminasa microbiana  
Thr (T). Treonina  
EPB42. Erythrocyte membrane protein band 4.2  
TGM. Gen que codifica para la Transglutaminasa  
GTP. Guanosín Trifosfato  
ATP. Adenosín Trifosfato  
ADP. Adenosín Difosfato  
GDP. Guanosín Difosfato  
Phe (F). Fenilalanina

Arg (R). Arginina  
Val (V). Valina  
Tyr (Y). Tirosina  
PNGasa. Péptido N-Glicanasa  
WT. Wild-type (Organismo de tipo silvestre)  
KO. Gen Knockout (Inactivación génica)  
pb. pares de bases  
TR. Tándem Repeats  
ADNc. Ácido Desoxirribonucleico complementario  
Gly (G). Glicina  
ARNm. Ácido Ribonucleico mensajero.  
ADNg. Ácido Desoxirribonucleico genómico  
nr. Base de datos no redundante.  
CASP. Critical Assessment of protein Structure Prediction  
GO. Gene Ontology  
SVM. Support Vector Machines  
RIDs. Regiones Intrínsecamente Desordenadas  
MF. Molecular Function  
BP. Biological Process  
mM. miliMolar  
Tris-HCl. Tris-Hidroclorido  
M. Molar  
pH. potencial de Hidrógeno  
NaCl. Cloruro de sodio  
p/v. Relación porcentual entre el peso del soluto y el volumen de la solución.  
EDTA. Ácido etilendiaminotetraacético  
CTAB. Bromuro de hexadeciltrimetilamonio  
mg. miligramos  
ng. nanogramos  
mL. mililitros

PVP. Polivinilpirrolidona

μL. microlitros

°C. grados centigrados

min. minutos

rpm. revoluciones por minuto

V. Volts

nm. nanómetros

Tm. Temperatura de fusión

PCR. Reacción en cadena de la polimerasa

MgCl. Cloruro de Magnesio

ORF. Open Reading Frames

TGc. Dominio Transglut core

aa. aminoácidos

Ala (A). Alanina

Ile (I). Isoleucina

Leu (L). Leucina

Met (M). Metionina

Lys (K). Lisina

## RESUMEN

Las transglutaminasas (TGs) son una familia de enzimas multifuncionales que se encuentran presentes en diferentes órganos, tejidos y fluidos corporales. Son ampliamente localizadas en bacterias, plantas y animales. Se componen de una tríada catalítica (Cys/His/Asp ó Asn) que comparte un origen evolutivo con la superfamilia de enzimas proteasas-cisteínas tipo papaína, induciendo modificaciones post-traduccionales de proteínas. La principal biocatálisis de las TGs es la reticulación, donde se forman enlaces isopeptídicos con un producto final polimerizado y con un amplio potencial biotecnológico. En plantas, la investigación derivada de las TGs es aún escasa. No obstante, existe un gran número de secuencias proteicas anotadas funcionalmente como TGs en las bases de datos biológicas y que carecen de una certeza biológica acerca de su actividad. Las únicas proteínas vegetales caracterizadas y validadas funcionalmente son la TGZ15/TGZ21 de maíz y la TGO de arroz, ambas proteínas originadas de genes de copia única y con un dominio tipo TG no canónico.

Una búsqueda exhaustiva en las bases de datos proteicas encontró que las gramíneas son el grupo vegetal más extenso con secuencias proteicas TGs registradas. Entre ellas, se encontró una isoforma inactiva (TGZi) de la TGZ15, producto de splicing alternativo, que originó una divergencia funcional entre ambas proteínas mediado por un mecanismo evolutivo entre “parálogos internos” del mismo gen. El evento causal de la divergencia se originó no solo por un transcripto alternativo, sino por un evento de duplicación génica que hace más tolerante las mutaciones dentro de un genoma. La diferencia entre ambas proteínas se concentra en un C-terminal distinto y ausente de la tríada catalítica. Esta característica mutacional también se observó en el gen *tg* parcial de caña de azúcar, homólogo al gen *tgz15* de maíz, que se logró identificar mediante la técnica de PCR. Los análisis de secuencias revelaron una diversidad de proteínas homólogas a TGZi, sugiriendo una robustez mutacional en el genoma de las gramíneas que permitió la conservación de genes duplicados. En conclusión, TGZ15 y TGZi son similares por la presencia de RIDs en forma de TRs que se conservan en el proteoma eucariota. La predicción de sus estructuras y el análisis de enriquecimiento sugieren un desorden estructural atribuido a RIDs que tienen una implicación en los procesos biológicos que desempeñan las TGs. La TGZ15 de maíz es evolutivamente independiente del resto de TGs, con un dominio catalítico divergente, que plantea una analogía funcional como resultado de una evolución convergente.

## ABSTRACT

Transglutamines (TGs) are a family of multifunctional enzymes present in different organs, tissues and body fluids. They are widely located in bacteria, plants and animals. They consist of a catalytic triad (Cys/His/Asp or Asn) that shares an evolutionary origin with the papain-like protease-cysteine enzyme superfamily, inducing post-translational protein modifications. The main biocatalysis of TGs is cross-linking, where isopeptide bonds are formed with a polymerized final product and with a broad biotechnological potential. In plants, research derived from TGs is still deficient. However, there is a large number of protein sequences functionally annotated as TGs in biological databases and lacking a biological certainty about their activity. The only plant proteins characterized and functionally validated are the active corn TGZ15 and the rice TGO, both proteins originated from single copy genes and with a non-canon TG type domain.

An exhaustive search in protein databases found that grasses are the most extensive plant group with recorded TG protein sequences. Among them, an inactive isoform (TGZi) of TGZ15 was found, product of alternative splicing, which caused a functional divergence between both proteins by an evolutionary mechanism between “internal paralogs” of the same gene. The causal event of the divergence originated not only by an alternative transcript, but by a gene duplication event that makes mutations within a genome more tolerant. The difference between the two proteins is concentrated in a different C-terminal and absent from the catalytic triad. This mutational characteristic was also observed in the partial sugarcane *tg* gene, homologous to the corn *tgz15* gene, which was identified by the PCR technique. Sequence analysis revealed a variety of proteins homologous to TGZi, suggesting a mutational robustness in the grass genome that allowed the conservation of duplicate genes. In conclusion, TGZ15 and TGZi are similar because of the presence of RIDs in the form of TRs that are conserved in the eukaryotic proteome. The prediction of their structures and the enrichment analysis suggest a structural disorder attributed to RIDs that have an implication in the functional processes of TGs enzymes. The corn TGZ15 is evolutionarily independent of the rest of TGs, with an active site and structural form divergent, it raises a functional analogy resulting from a convergent evolution.

## 1. INTRODUCCIÓN

Las TGs (EC.2.3.2.13; R-glutaminil-péptido-amina- $\gamma$ -glutamil transferasas) son enzimas multifuncionales presentes en gran parte de tejidos, órganos y fluidos corporales ampliamente distribuidas en la naturaleza. A diferencia de otras enzimas industriales, las TGs inducen la polimerización de proteínas. La TG realiza la transferencia acil dependiente del  $\text{Ca}^{2+}$  entre grupos  $\gamma$ -carboxiamida de residuos de glutamina y  $\epsilon$ -amino de residuos de lisina u otros aminos primarios. Este grupo de enzimas forman reticulaciones mediante enlaces covalentes a nivel intra- o intermolecular entre proteínas, péptidos o grupos aminos primarios. La actividad catalítica reticulante es de interés biotecnológico, con amplias aplicaciones en la industria alimentaria, ya que mejora la estabilidad térmica, solubilidad, capacidad de retención de agua y valor nutricional de proteínas (Motokis & Kumazawa 2000; Griffin *et al.*, 2002; Lorand & Graham 2003; Li *et al.*, 2017).

Estudios bioinformáticos han establecido que las TGs han evolucionado a partir de la familia proteasas-cisteínas tipo papaína, donde comparten un dominio particular llamado tríada catalítica (Cys/His/Asp o Asn; Marakova *et al.*, 1999). La familia de TGs en mamíferos son las mejor estudiadas y caracterizadas en términos de propiedades tanto estructurales como catalíticas. Las TGs que destacan son el FXIIIa de coagulación sanguínea y la TG2h. La amplia cobertura de investigación de TGs en animales derivó en la producción enzimática a nivel industrial, sin embargo, la comercialización de la TG a partir de tejidos animales tiene un alto costo en el proceso de separación y purificación (Motoki & Kumazawa, 2000).

En la actualidad, la TG microbiana (TGM) lidera la industria alimentaria y por ello son obtenidas en grandes cantidades para el procesamiento de cárnicos y productos pesqueros. La TGM es una simple proteína monomérica de 38 kDa, con una estructura proteica diferente al ya caracterizado FXIIIa del humano; existe una relación entre las estructuras secundarias del sitio activo de la TGM y la TG-pep, lo que sugiere una relación evolutiva de convergencia entre las dos proteínas (Yokoyama *et al.*, 2004). La biosíntesis de TGs a partir de bacterias ofrece una estabilidad en un amplio rango de valores de pH y que no requieren de iones de calcio para su activación, estas propiedades la hacen deseables para la manipulación industrial (Langston *et al.*, 2006).

En plantas, la información sobre TGs se encuentra aún en una etapa joven, por lo que hasta el momento las únicas TGs expresadas de forma heteróloga han sido la AtPNG1p de *Arabidopsis thaliana* (Della Mea *et al.*, 2004a) y TGZ15/TGZ21 de *Zea mays* (Villalobos *et al.*, 2004). Sin embargo, la TGZ15 se diferencia del resto de TGs caracterizadas porque contiene un sitio activo ubicado en el C-terminal con un entorno catalítico diferente. Todas estas características también se contrastan con la ausencia de una estructura proteica, por lo que ofrece la oportunidad de analizar sus características tanto estructurales y funcionales respecto a otras TGs estudiadas, la TGZ15 es una enzima única con un dominio catalítico putativo y aún no comprobable (Serafini-Fracassini *et al.*, 2009).

Para realizar este estudio se utilizaron herramientas bioinformáticas tanto de uso local como en línea, sin embargo, en la actualidad nos enfrentamos a grandes volúmenes de datos biológicos almacenados en las bases de datos, originados a partir de experimentos previos. Muchas de las secuencias a analizar contienen una variedad de anotaciones curadas manual o automáticamente que infieren la función biológica de una secuencia génica o proteica en base a un umbral de similitud. En muchas ocasiones las anotaciones inferidas a una secuencia son erróneas y son transferidas a través de referencias cruzadas entre las bases de datos, la transferencia automática de la anotación funcional entre todos los grupos de genes ortólogos y parálogos se vuelve redundante, porque hay casos en que las funciones de los genes han divergido. En este estudio se hace una diferencia entre este tipo de anotaciones y la naturaleza de las bases de datos utilizadas.

## 2. MARCO TEÓRICO

### 2.1 Bioinformática

En la biología molecular actual, el desafío ya no es obtener datos, sino analizarlos e integrarlos de manera significativa e innovadora, obtener nuevos conocimientos y comprender el sistema biológico en su conjunto. Para esto, se desarrollaron técnicas informáticas derivadas de disciplinas como las matemáticas, la estadística y las ciencias computacionales que permitieran entender y organizar la información asociada con el sistema de información que se obtiene de la biología molecular a gran escala (Luscombe *et al.*, 2001).

El término bioinformática se definió por primera vez en la Universidad de Utrecht por Ben Hesper y Paulien Hogeweg como el estudio de los procesos informáticos en sistemas bióticos (Hesper & Hogeweg 1970). El desarrollo de herramientas computacionales y bases de datos, así como sus aplicaciones, hacen posible la generación de conocimiento biológico para entender mejor a los sistemas vivos. Estas herramientas son usadas en tres áreas de la investigación genómica y biología molecular como lo son el análisis de secuencias, de función y de estructura (Xiong 2006). En la actualidad, los bioinformáticos trabajan con datos moleculares como secuencias de genes porque sigue siendo el tipo de información más fácil de obtener. Sin embargo, la existencia de varios tipos de datos en otros campos de la biología está ganando cada vez más interés por parte de la comunidad bioinformática.

El incremento en la producción de datos de secuenciación ha resultado en la creación de bases de datos que almacenen la información y datos de las secuencias con la finalidad de inferir conocimiento, por lo tanto, es remarcable el crecimiento de datos proteicos manualmente curados o automáticamente anotados. Muchas veces los datos crudos, aquellos que son obtenidos directamente de los procesos experimentales necesitan ser analizados y procesados con el fin de facilitar la interpretación. La curación de los datos organiza e integra los datos coleccionados desde diversas fuentes e involucra la anotación de los datos sobre el tiempo que pueden llegar a permanecer, así como su valor obtenido. El procesamiento de los datos se realiza por etapas, donde el dato procesado resultante es el dato crudo del siguiente.

La calidad y revisión de los datos consiste en analizar, interpretar e integrar la información biológica que se encuentra en las bases de datos mediante un valor extra de la anotación y la interconexión de los datos dentro de un marco biológico común. Las personas encargadas de la revisión no solo anotan, sino también gestionan el cúmulo de datos mediante la incorporación y estandarización de estos. Todo este procesamiento de información se facilita mediante la incorporación de vocabularios controlados y ontologías que facilitan el manejo del conocimiento mediante un lenguaje natural en torno a un área específica. Los vocabularios controlados ofrecen un conjunto de términos formales. (Choudhuri 2014).

La existencia de las bases de datos biológicas se origina a partir del procesamiento que reciben sus datos almacenados, donde muchas de ellas coexisten en la misma área, por lo que es común encontrar referencias cruzadas (Yagüe *et al.*, 2014). Existen tres tipos de bases de datos que se clasifican de la siguiente manera: las bases de datos primarias que almacenan datos biológicos crudos y que son sometidos por la comunidad científica (GenBank, PDB, EMBL y DDBj) sin revisión exhaustiva, las bases de datos secundarias que contienen información procesada computacionalmente o curada manualmente por un experto a partir de información original de las bases de datos primarias, aquí se encuentran secuencias traducidas de proteínas que contienen anotaciones funcionales (Swiss-Prot, TrEMBL, PIR y RefSeq), finalmente, las bases de datos especializadas se centran en un interés de investigación particular (Xiong 2006).

La UniProtKB es una base de datos de alta calidad que posee una anotación completamente clasificada, enriquecida y precisa con una mínima redundancia que se complementa con referencias cruzadas extensivas. UniProtKB unifica su información a partir de dos tipos de revisiones de calidad proteica. Las secuencias proteicas anotadas computacionalmente y sin revisión, conocida como la base de datos TrEMBL, son en su mayoría traducciones de secuencias codificantes; el incremento exponencial de anotaciones sin revisión es registrada con frecuencia como “proteínas no caracterizadas” ya que carecen de información más descriptiva. Por otra parte, las secuencias con anotaciones revisadas manualmente por expertos (Swiss-Prot) contienen información extraída desde la literatura con predicciones computacionales evaluadas por un curador de datos (Schneider *et al.*, 2005). Además de clasificar su información en base a

anotaciones revisadas o no revisadas, una de sus características importantes que integra UniProtKB es la anotación de un criterio de existencia que se le atribuye a la proteína (ver Tabla 1).

**Tabla 1.** Clasificación de proteínas por tipos de evidencias registradas en la UniProtKB. La base de datos UniProtKB clasifica las secuencias proteicas en base a 5 tipos de evidencia que soportan la existencia de una proteína (UniProt Consortium 2007).

Tipo de evidencia	Características
Predichas	Es usada para entradas sin evidencia a niveles transcritos, proteicos o inferido por homología.
Experimental a nivel transcritos	Evidencia de proteína que no ha sido estrictamente probada sino por expresión de datos indicando la existencia del transcritos.
Experimental a nivel proteína	Existencia de la proteína por clara y estricta evidencia experimental.
Inferido por homología	Existencia de proteína probable por evidentes ortólogos existentes en especies cercanamente relacionados.
Proteína incierta	No existe certeza de que la proteína exista.

El incremento exponencial de los datos biológicos crea una redundancia de información que al momento de ser analizada resulta en inferencias de anotaciones funcionales erróneas, estos hechos dificultan la creación u organización de una base de datos sólida, libre de redundancia y completamente revisada. Por ejemplo, una misma secuencia de ADN puede ser atribuida a múltiples autores, ya que fue introducida a una base de datos primaria sin una revisión exhaustiva. Este impacto lleva a proteínas falsamente anotadas y con resultados en publicaciones de una investigación incierta (Brenner 1999; Nadzirin *et al.*, 2012; Percudani *et al.*, 2013; Liberal *et al.*, 2013).

## 2.2 Transglutaminasas

Las TGs son una familia de enzimas tiol transferasas que se localizan tanto intracelular como extracelularmente originando modificaciones post-traduccionales en proteínas. Las TGs han evolucionado a partir de la familia proteasas-cisteínas tipo papaína que comparten un dominio catalítico particular llamado tríada catalítica (Cys/His/Asp o Asn; Marakova *et al.*, 1999; Lorand & Graham 2003). El nombre transglutaminasa fue acuñado por primera vez por Waelsch en 1957 (Sarkar *et al.*, 1957), sin embargo, en algunas publicaciones (Lorand & Conrad 1984; Mehta & Eckert 2005) se menciona que el nombre transglutaminasa no es el más apropiado para esta

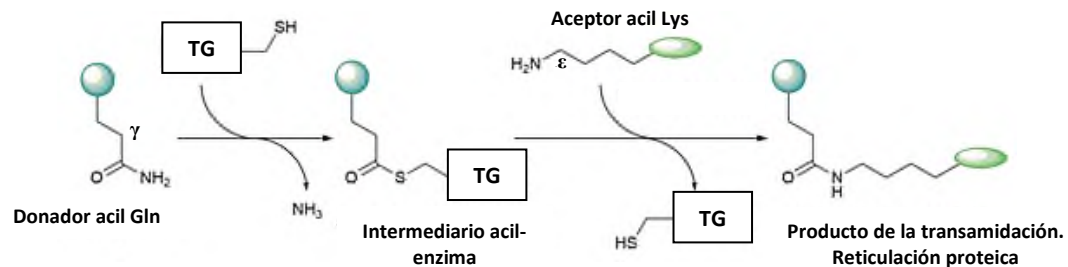
enzima, ya que no reaccionan con un aminoácido libre de glutamina, sino que realizan funciones carbonilamida en la cadena lateral de residuos de glutamina en proteínas sustrato. Estas reacciones forman un entrecruzamiento o reticulación llamado enlace isopeptídico  $\epsilon$ -( $\gamma$ -glutamil) lisil (Folk & Finlayson 1977; Griffin *et al.*, 2002).

Todo inició con la primer TG descrita en 1957, donde un equipo conformado por Sarkar, Clarke & Waelsh reportaron enzimas capaces de marcar proteínas por aminos presentes en extractos de cerebro y dependencia a iones de calcio, además de presentar mayor actividad en otros tejidos como el hígado. En el marco del descubrimiento de nuevas TGs, se halló actividad de reticulación proteica, así como la desaminación de residuos de glutamina en proteínas (Sarkar *et al.*, 1957; Mycek *et al.*, 1959; Beninati *et al.*, 2009). La TG está distribuida en plantas, animales, cuerpo humano y microorganismos (Piacentini *et al.*, 2004; Mirzaei 2011; Klöck & Khosla 2012; Del Duca *et al.*, 2014;), lo que la convierte en una enzima ubicua y multifuncional. Ejemplos de estos hechos es que la TG se encuentra involucrada en muchos procesos fisiológicos como en la coagulación, reacciones inmunes antibacteriales y en la fotosíntesis (Kashiwagi *et al.*, 2002).

La reticulación por TGs en células eucariotas ha mostrado ser activada por la elevación de  $\text{Ca}^{2+}$  intracelular que usualmente causa cambios permanentes de senescencia celular en la membrana y esqueleto de la célula (Lorand 2007). La actividad TG también es posible en la ausencia del  $\text{Ca}^{2+}$  que sólo ha sido descrita en bacterias (Williams-Ashman & Canellakis 1980; Elgavish *et al.*, 1984; Ho *et al.*, 2000; Barros *et al.*, 2003; Yokoyama *et al.*, 2004). Los efectos negativos influenciados por calcio provocan una reorganización significativa en la estructura de la TG (Fesüs *et al.*, 1977; Murthy *et al.*, 1999; Eckert *et al.*, 2014).

Las TGs son sintetizadas como zimógenos que requieren de su activación antes de exhibir actividad transamidasa. La catálisis de transamidación mediada por la enzima es parte de un conjunto de procesos post-traduccionales en las cuales las proteínas están sujetas y pueden incluirse a un número de estrategias regulatorias eficientes. Comparada con otras reacciones propias de la TG, la transamidación es la más estudiada y comprendida (Folk 1980; Beninati *et al.*, 1988) por su amplio potencial biotecnológico. La actividad enzimática TG dependiente del  $\text{Ca}^{2+}$  se

realiza con base en la transferencia acil de uniones amida covalente, participando un grupo  $\gamma$ -carboxiamida de un residuo glutamil de algunas proteínas o péptidos (amino aceptoras) y un grupo amino primario como el encontrado en la posición  $\epsilon$ -amino de un residuo de lisina o un grupo amino libre de una amina soluble como PAs (amino donadores; ver Figura 1).



**Figura 1.** Actividad de transamidación catalizada por la TG. La transamidación y la reacción entre los substratos se lleva en dos pasos, en la primera el residuo  $\gamma$ -carboxiamida se une al grupo -SH del residuo de Cys en el sitio activo de la TG, lo que da lugar a un complejo enzimático intermediario acil-enzima ( $\gamma$ -glutamiltioéster), inmediatamente se libera amonio ( $\text{NH}_3$ ). En el segundo paso el residuo  $\epsilon$ -amino se une al intermediario acil-enzima y reacciona con  $\gamma$ -glutamiltioéster para formar el enlace isopeptídico covalente y la regeneración de Cys en la TG. El enlace isopeptídico formado es resistente a la proteólisis y los productos reticulados son altamente resistentes a químicos, así como la degradación física y enzimática (Griffin *et al.*, 2002; Iisma *et al.*, 2009). Bajo condiciones fisiológicas las reacciones de reticulación por TGs son usualmente irreversibles (Gatta *et al.*, 2016; Bruns & Kilbinger 2016).

Si no existen aminas primarias disponibles, el agua puede funcionar como un grupo aceptor acil provocando una desaminación del residuo de Gln (Folk & Chung 1973; Lorand & Conrad 1984), el reemplazo de un residuo de Gln neutral por Glu con carga negativa importa consecuencias biológicas de gran importancia a partir de la cadena peptídica que van desde el cambio en la solubilidad, conformación, estabilidad de oligómeros o interacciones de las proteínas con otras macromoléculas o compuestos. Muchos de estos aminoácidos se encuentran naturalmente en proteínas de la matriz extracelular o membrana basal de células epiteliales.

El rol del intermediario enzimático es la estabilización de un residuo de Gln en papaína y un residuo de Trp en el caso de la TG (Iisma *et al.*, 2003). El papel del Trp en el sitio activo de la TG

es esencial para la catálisis donde estabiliza los estados de transición por enlaces de hidrógeno, interacciones ión-cuadropolo o por van der Waals. La conservación de un Trp241 en todas las TGs enzimáticamente activas de la TG2h, indica que la estabilización del estado de transición por este residuo es una característica general de catálisis en todas las TG eucariotas que van desde moho del limo hasta mamíferos (Iisma *et al.*, 2003; Murthy *et al.*, 2002). De acuerdo con las similitudes encontradas entre el FXIIIa y la papaína dentro del sitio activo, se propuso que la función del mecanismo catalítico en la estructura enzimática se presenta de la siguiente manera: la Cys314 actúa como nucleófilo y la His373 como el catalizador ácido/base; la Asp396 participa en un rol secundario, la cual es estabilizar la forma protonada de His373 que provee una orientación favorable de este residuo (Pedersen *et al.*, 1994).

Solo se han identificado tres familias de TGs caracterizadas estructuralmente, TGs similares a papaína (FXIIIa, TG-pep y TG2h-TG3h), las TGs tipo PDI similares (que tienen actividad isomerasa de enlace disulfuro además de su actividad TG) y por último las toxina bacteriana TG (como el factor citotóxico de *E. coli* y la neurotoxina *B. Bronchiseptica*) que son predominantemente enzimas desamidantes, pero que también parecen catalizar algunas reacciones de transamidación (Chen & Mehta 1999; Lorand & Graham 2003).

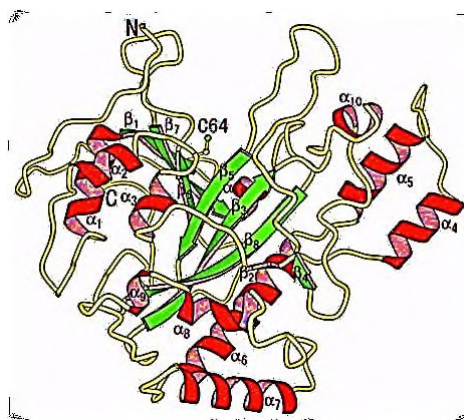
### **2.3 Transglutaminasas microbianas**

Las propiedades catalíticas de las TGs empezaron a aplicarse en procesos industriales alimenticios porque mejoraban las propiedades funcionales de varias proteínas, como las proteínas de carne, soja, miosina, globulina, caseína, cacahuete y suero (Martins *et al.*, 2014). La comercialización de la TG surgió a partir de tejidos animales; el problema radicó en el proceso de separación y purificación pues representaba un alto costo (Zhu *et al.*, 1999; Motoki & Kumazawa 2000). Se descubrió que las propiedades de las TGM son mucho más útiles ya que no requieren  $Ca^{2+}$  como cofactor para mostrar actividad catalítica, mostraron estabilidad en un amplio rango de pH, con una capacidad para reticular la mayoría de las proteínas de los alimentos y con menores costos de producción debido a la fácil producción y purificación de enzimas (Langston *et al.*, 2006; Mirzaei 2011; Martins *et al.*, 2014).

La primer TGM aislada fue la obtenida a partir de un medio de cultivo de *Streptovorticillium sp. S-8112* (Ando *et al.*, 1989). Actualmente la TGM de *Streptovorticillium morbaraense* es producida en grandes cantidades para la producción de productos cárnicos y pesqueros procesados (Carlos & Carreño 2007). La TGM de *Streptovorticillium morbaraense* tiene un peso molecular de 38 kDa (bajo peso molecular comparado con las TG en mamíferos) y una composición de 331 aminoácidos. Esta proteína posee un solo residuo de Cys, no forma enlaces disulfuro y mantiene una actividad catalítica independiente del calcio, propiedades que han sido ampliamente estudiadas (Ando *et al.*, 1989; Nonaka *et al.*, 1989; Takashiba & Sakurai 2016). En el 2002 la estructura cristal de la TGM fue hallada con una resolución de 2.4 Å (Kashiwagi *et al.*, 2002), revelando una estructura completamente diferente a la del FXIII y las TGs de mamíferos; además por medio de RMN se descubrió que la velocidad de reacción y la especificidad del substrato para el donador acil de la TGM son mayores y menores que las de las TGs similares al factor XIII (Shimba *et al.*, 2002).

La TGM es expresada como zimógeno (proTGM), con un corto segmento de 45 aminoácidos en la región N-terminal, representa la secuencia líder que posiblemente sea esencial para el plegamiento eficiente de la proteína, la secreción y la supresión de su actividad enzimática (Yang *et al.*, 2011; Takashiba & Sakurai 2016). La estructura general de la proTGM forma un dominio compacto que adopta una forma parecida a un disco con una hendidura profunda en el borde del disco (Yokoyama *et al.*, 2004). La estructura secundaria pertenece a la clase de plegamiento  $\alpha+\beta$ , contiene 11  $\alpha$ -hélices y 8  $\beta$ -laminas, una de las hojas  $\beta$  está rodeada por hélices  $\alpha$ , que se agrupan en tres regiones (ver Figura 2).

El núcleo catalítico encontrado entre la TGM y la TG-pep son similares. Se postula que estas TGs están relacionadas por una evolución convergente; por lo que la TGM ha desarrollado un nuevo mecanismo catalítico especializado para la reacción de reticulación (Yokoyama *et al.*, 2004). Estas proteínas microbianas tienen una tríada catalítica no convencional y no muestran homología de secuencia con ninguno de los otros dos grupos de proteínas bacterianas, o con las TGs similares a papaína.



**Figura 2.** Estructura general de la TGm (PDB ID: 3IU0). El modelo tipo cartoon de la TGm se visualiza desde la parte de arriba del disco. El residuo de Cys64 visualizado en forma de bolas y palillos se encuentra en un loop entre las hélices  $\alpha_2$  y  $\alpha_3$ . La lámina  $\beta$  central forma una estructura antiparalela de siete cadenas y está fuertemente retorcida entre las hebras  $\beta_5$  y  $\beta_6$ , y solo hay un enlace de hidrógeno entre las cadenas principales de estas hebras (Trp258 y Thr273; Yokoyama *et al.*, 2004).

#### 2.4 La familia TG en mamíferos

Las TGs son codificadas por una familia de genes cercanamente relacionados que poseen un alto grado de similitud de secuencia (Grenard *et al.*, 2001). Todos estos genes aislados hasta ahora parecen estar organizados de una manera similar. Los miembros de la familia TG pueden distinguirse en función de sus propiedades físicas, distribución tisular, localización y mecanismos de activación.

Existen 9 miembros de la familia TG en el humano (Venter *et al.*, 2001), pero solo 8 poseen actividad catalítica, desde la TG 1-7 y el FXIIIa, la novena es conocida por su ausencia del sitio catalítico, esta proteína es nombrada “Band-4.2” (ver Tabla 2; Ichinose *et al.*, 1990; Aeschlimann & Paulsson 1994; Griffin *et al.*, 2002). Cada miembro de la familia TG generalmente se caracteriza por la falta de enlaces de glicosilación y puentes disulfuro a pesar de la presencia de posibles sitios de N-glicosilación y residuos de Cys en estas proteínas.

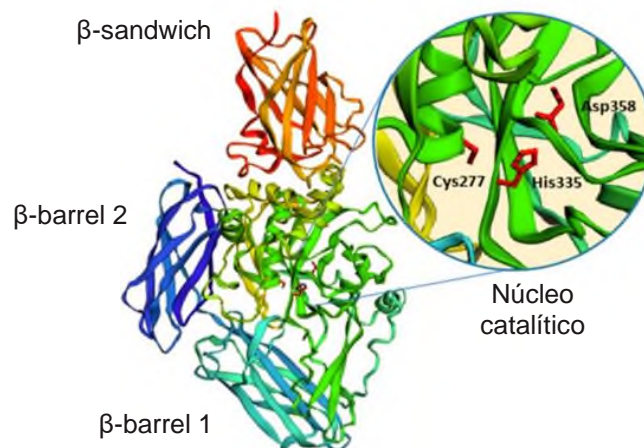
Todas las TGs incluyendo aquellas que son secretadas (Factor XIII) o asociada con la membrana (TG1) carecen de secuencia hidrófoba N-terminal (Mehta 2005). Todos los miembros de esta

superfamilia de enzimas poseen una tríada catalítica (Cys/His/Asp ó Asn) conservada, a excepción de la proteína inactiva Band 4.2 que presenta homología estructural a los demás miembros TG; la inactividad es debido a la sustitución de la Cys, lo que produce un mayor componente para la formación de eritrocitos del esqueleto membranal (Beninati *et al.*, 2009).

**Tabla 2.** Proteínas miembros de la familia TG en humanos. Las TGs identificadas en el ser humano son calcio-dependientes para su actividad catalítica (Lorand & Graham 2003).

Gen-Proteína	Funciones	Residuos (kDa)
TGM1-TG1	Formación de envoltura celular durante diferenciación de queratinocitos.	814(90)
TGM2-TG2	Apoptosis, adhesión celular, estabilización de la matriz, señalización a la supervivencia celular, etc.	686(78)
TGM3-TG3	Formación de la envoltura celular durante la diferenciación de queratinocitos.	692(77)
TGM4-TG4	Reproducción especialmente en roedores como resultado de la coagulación del semen.	684(77)
TGM5-TG5	Formación de la envoltura celular cornificada durante la diferenciación de queratinocito.	720(81)
TGM6-TG6	Se desconoce	625(70)
TGM7-TG7	Se desconoce	710(80)
F13A1-FXIIIa	Coagulación sanguínea, cicatrización de heridas y crecimiento óseo.	732(83)
EPB42-Band4.2	Componente especializado en redes de eritrocitos esqueléticos.	690(72)

Existen ortólogos de diversas TGs humanas en organismos mamíferos hasta invertebrados y presentan un alto grado de conservación en la estructura primaria y secundaria. Se caracterizan por poseer cuatro dominios en su estructura terciaria similares a los que presenta el FXIIIa, TG2h y TG3h, así como en la TG-pep (Yee *et al.*, 1994; Ahvazi *et al.*, 2002; ver Figura 3).



**Figura 3.** Estructura general de la TG2h. En el modelo tipo cartoon de la TG2h se visualiza la conformación cerrada de la proteína con los 4 dominios estructurales mostradas de diferentes formas cilíndricas. En esta conformación la proteína es inactiva debido al espacio dimensional de los dominios  $\beta$ -barrel que ocultan al empaque catalítico. El sitio catalítico está compuesto por Cys277/His335/Asp358, además de residuos clave en unión a GTP (Phe174, Arg478, Val479, Arg580, Tyr583; Stammaes *et al.*, 2010).

La TG2 sobresale por su multifuncionalidad y por ser el miembro mayormente estudiado de toda la familia de las TGs. La TG2 ha mostrado participar en la hidrólisis, unión a GTP y ATP (Bergamini *et al.*, 1987; Lee *et al.*, 1989; Lai *et al.*, 1998; Beninati *et al.*, 2017), como PDI, quinasa y destacando además su comportamiento como proteína andamio o enlazante. Esta proteína única sufre cambios conformacionales y exhibe diversas localizaciones (Gundemir *et al.* 2012). La unión ATP/ADP no mostró un efecto en la actividad de transamidación, por lo que la TG2 solo es activa cuando se enlaza con calcio e inactiva cuando se enlaza a GTP/GDP (Begg *et al.*, 2006; Eckert *et al.*, 2014). Las implicaciones de la TG2h en diversas enfermedades humanas están presentes en patologías como la inflamación, cáncer y fibrosis, incluso en enfermedades cardiovasculares, neurodegenerativas y celíacas, en las que desempeña un papel protector o inductor a la patogénesis. La modulación de las actividades biológicas de la TG2 en estas enfermedades tendrá un valor terapéutico en el futuro (Siegel & Khosla 2007).

La participación de la TG en el espacio extracelular está identificada en el proceso de hemostasia. Es conocido que el plasma sanguíneo forma un heterotetrámero de 320 kDa, donde 283 kDa subunidades A son TGs zimogénicas (Ichinose 2012). El FXIIIa zimógeno es activado por escisión proteolítica mediante trombina, participando en la coagulación de la sangre. Por lo que es la TG mejor caracterizada y la primera reportada con una estructura cristal para la familia TG, revelando similitudes a las proteasas-cisteína en el sitio activo. El FXIIIa es un modelo de comparación por homología y es la primera homóloga reportada con la TG del cerdo de Guinea (Ichinose *et al.*, 1990).

Recientemente Sánchez-Pulido y Ponting (2016) identificaron por análisis de secuencias, proteínas vasoinhibinas (encargadas de regular la angiogénesis, crecimiento de tumor y la metástasis) como nuevos miembros de la superfamilia proteasas-cisteína del tipo TG, pero con la diferencia de poseer una tríada catalítica no canónica encontrada como Cys/His/Ser ó Thr. Esta tríada no canónica no es exclusiva de esta familia de proteínas vasoinhibinas, puesto que existen miembros de la superfamilia proteasas-cisteína tipo TG con un núcleo catalítico (Pfam: P01841) no canónico (Cys/His/Thr) como la TG de *Syntrophobacter fumaroxidans*.

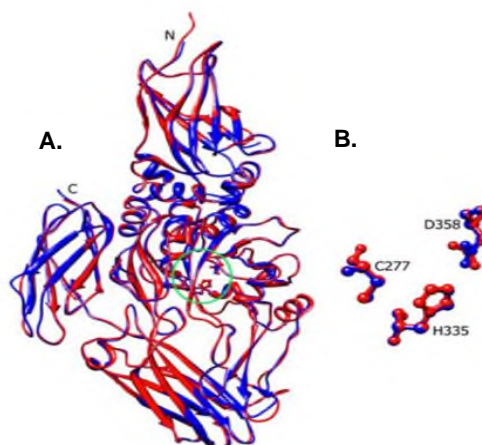
## 2.5 Transglutaminasas vegetales

En plantas, la modificación post-traducciona l de proteínas y PAs catalizadas por TGs ha sido estudiado desde 1987 (Mosseti 1987; Ickson & Apelbaum 1987). La actividad principal de las TGs en plantas son la conjugación de PAs y otros amino donadores (Del Duca & Serafini-Fracassini 2005; Serafini-Fracassini & Del Duca, 2008; Del Duca *et al.*, 2014). Las TGs en plantas han sido detectadas principalmente por la producción de derivados de PAs-glutamil, así como a través de características bioquímicas mediante la detección con anticuerpos y modificación de sustratos animales. Las a minas primarias, específicamente las PAs son los sustratos de las TGs en plantas; las PAs más estudiadas son la espermidina, putrescina y espermina, además de la biotina-cadaverina que es usada para detectar actividad enzimática *in vitro* (Kang & Cho 1996; Del Duca *et al.*, 2000a; Della Mea *et al.*, 2004b).

Las TGs se localizan en diferentes órganos de la planta y en varios compartimentos celulares tales como cloroplastos, mitocondria, citoplasma y paredes celulares. La ubicación de TGs de manera extracelular e intracelular ha sido ampliamente estudiado (Del Duca S., *et al.*, 2014; Serafini-Fracassini & Del Duca 2008). La actividad TG en plantas inferiores y superiores se encuentra involucrado en el crecimiento y procesos de diferenciación, así como relacionadas a la fertilización, conjugación de PAs, germinación del polen, estrés biótico y abiótico, además de senescencia y muerte celular programada (Dondini *et al.*, 2001; Del Duca & Serafini-Fracassini 2005; Del Duca *et al.*, 2007; Della Mea *et al.*, 2007; Iorio *et al.*, 2008; Sobieszczuk-Nowicka *et al.*, 2007, 2008; Serafini-Fracassini *et al.*, 2009).

La actividad tipo TG en plantas fue descrita por primera vez en el guisante (Icekson & Apelbaum 1987), posteriormente se halló actividad en *Helianthus tuberosus* (Serafini-Fracassini *et al.*, 1988), en hojas de acelga (*Beta vulgaris*; Signori *et al.*, 1991), hojas de soya (*Glycine max*; Kang & Chao 1996), en raíces y brotes de tejidos de plantas dicotiledóneas (como la pera y el haba), así como en monocotiledóneas (cebada y trigo; Lilley *et al.*, 1998). Se ha demostrado que proteínas sustrato como actina y tubulina son elegidos por la TG en plantas superiores a través del rápido crecimiento del tubo del polen que se sustenta en una reorganización dramática del citoesqueleto (Del Duca S. *et al.*, 1997). En plantas superiores, las TGs muestran una baja similitud de secuencia con las ya conocidas TGs de mamíferos. Existe la hipótesis que las proteínas codificadas desde el genoma plastidial (origen evolutivo procariota) podrían compartir similitud considerable con las cianobacterias, sin embargo, esto no puede ser verificado debido a que las secuencias de TGs no están disponibles en el genoma de las cianobacterias (Del Duca *et al.*, 2014).

Beninati *et al.*, (2013) identificaron en *Helianthus tuberosus* 3 bandas (58, 75 y 85 kDa) de TGs aisladas en brotes de ápices; se identificó actividad TG relevante y similitud con las TGs animales. En el estudio se analizó la composición de aminoácidos de la enzima de 75 kDa, que fue prácticamente idéntica (con una diferencia del 0.29%) a la proteína EPB42 de eritrocitos. Sabiendo la similitud encontrada entre estas proteínas, se pudo observar que también la enzima de 85 kDa se ajusta perfectamente con TG2h; en base a esto, se construyó un modelo 3D a partir de la homología encontrada (ver Figura 4).



**Figura 4.** Modelamiento por homología entre la enzima de 75 kDa de *H. tuberosus* y la TG2h. A. El modelo general por homología detalla en azul a la TG2h y en rojo la enzima de 75 kDa. B. El sitio activo se encuentra en las posiciones Cys277/His335/Asp358 (Beninati *et al.*, 2013).

En la actualidad, los genes caracterizados y con actividad funcional TG comprobable en plantas se restringe solo a 2 especies, arroz y maíz; por otra parte, aún existe la duda si la *AtPng1p* de *Arabidopsis thaliana* exhibe una actividad tipo TG o no (ver Tabla 3).

**Tabla 3.** Lista de genes tg de origen vegetal secuenciados y caracterizados hasta el momento.

Todos estos genes fueron aislados a partir de ARNm.

Organismo	Nombre del Gen	Tamaño (pb)	Referencia	ID EMBL Nucleótido
<i>Arabidopsis thaliana</i>	<i>AtPng1p</i>	2327	Della Mea <i>et al.</i> , 2004a.	AM745095.1
<i>Zea mays</i>	<i>tgz15</i> y <i>tgz21</i>	1748 y 1910	Villalobos <i>et al.</i> , 2004.	AJ21525.2 y AJ488103.1
<i>Oryza sativa</i>	<i>tgo</i>	1605	Campos <i>et al.</i> , 2013.	FR717829.1

### 2.5.1 El gen *AtPng1p* de *Arabidopsis thaliana*

El gen *AtPng1p* es un gen de copia única, ubicua, de baja expresión y que codifica para una proteína de 721 aminoácidos. El gen codificante fue sobreexpresado en *E. coli* para demostrar actividad TG, la proteína recombinante fue purificada con un peso de banda de 86 kDa (Della Mea *et al.*, 2004a). Años atrás ya se había identificado este mismo gen *AtPng1p* de *Arabidopsis thaliana*, señalándolo con una función putativa como péptido N-glicanasa y con un dominio catalítico TG (Suzuki *et al.*, 2001).

Más tarde, Diepold *et al.*, (2007) sostiene que la *AtPng1p* codifica a una PNGasa. La actividad PNGasa fue exhibida cuando fue ensayada en una cepa de *Saccharomyces* con actividad PNGasa deficiente. Con estos experimentos se sugirió que la *AtPng1p* no es una TG, debido a la baja actividad TG comparada con otras proteínas y la presencia de actividad TG similar en plantas WT y KO. Por lo tanto, se cuestiona que la actividad TG asociada a *AtPng1p* pueda ser debido a condiciones experimentales usadas para detectar actividad enzimática *in vitro*, pareciendo poco probable que refleje actividad fisiológica en plantas. Debido a lo anterior, Serafini-Fracassini *et al.*, (2009) explicó que la enzima *AtPng1p* pueda presentar actividad multifuncional, esto, debido a que la proteína comparte propiedades bioquímicas e inmunológicas con homólogos animales, que dan origen a derivados de glutamilo, característicos de la actividad transamidasa.

### **2.5.2 El gen *tgo* de arroz (*Oryza sativa*)**

El gen *tgo* de arroz (*Oryza sativa*) posee una región codificante de 1059 pb mientras que el gen completo tiene una longitud de 1605 pb que poseen dos intrones y tres exones, este gen codifica para una proteína de 352 aminoácidos con peso molecular de 40 kDa compartiendo un 52% de identidad con la proteína TGZ15 (Campos *et al.*, 2013). La TGO crea uniones proteicas supramoleculares reticuladas de manera similar a la función asignada para las TGs en células animales, células vegetales y microorganismos. Esto puede ser similar a la llamada formación del dímero D inducida por el Factor XIII, que aparece en el último paso de la coagulación de la sangre y los dominios D de la fibrina. La dependencia al calcio, la especificidad del sustrato e inhibidor, así como la regulación de la actividad de la luz por los sustratos de plantas, indican que la TG de arroz puede actuar como una TG auténtica.

A diferencia de la proteína TGZ15 de maíz, TGO no presentó por predicción computacional péptido tránsito al cloroplasto, pero experimentalmente se demostró que la proteína no solo se encuentra en la grana del cloroplasto sino también en vesículas cercanas a ella, llegando a concluir que existe un patrón diferencial de esta enzima en arroz. Esta proteína presenta dos regiones TRs y una tríada catalítica putativa (Cys254/His260/Asp295) junto con la presencia de un Trp252 anterior a la Cys (Campos *et al.*, 2013), el cual también presuntamente actúa como estabilizador catalítico a partir de lo descrito por Murthy *et al.*, (2002).

### **2.5.3 Antecedentes del gen *tgz15/tgz21* de maíz (*Zea mays*)**

En 2001 se identificó por primera vez una TG de maíz a partir de cloroplastos. Las células de cloroplastos estaban expuestas a la luz y la actividad enzimática incrementó significativamente en el ensayo enzimático (Villalobos *et al.*, 2001). Más tarde, se demostró actividad TG en maíz inhibida por GTP y otros componentes de bloqueo. Estos eventos descritos resultaron en el aislamiento, clonación, secuenciación y caracterización de tres ADNc que codifican para la TGZ activa (*tgz15*, *tgz21* y *tgz4*, nombrados de acuerdo con el número de TRs), estos genes probablemente sean de copia única en el genoma del maíz. La TGZ fue sobreexpresada en *E. coli* donde existe en cuerpos de inclusión y fue producida para aplicaciones industriales (Villalobos *et al.*, 2004; Carvajal *et al.*, 2006; Carvajal-Vallejos *et al.*, 2007; Carvajal *et al.*, 2011).



El dominio catalítico de la TGZ15 es próximo al C-terminal con una ubicación en la Cys439/His446/Asp471 pero que no muestran similitud y conservación alguna con otras TGs estudiadas a nivel global (ver Figura 6). Precisamente esta composición no convencional del sitio activo de la TGZ activa fue analizada. Sobieszczuk-Nowicka, *et al.*, (2015) encontró eventos mutacionales junto a otras especies vegetales de la familia *Poaceae*, en donde se confirma que existe una alta similitud de secuencia en la región codificante hasta la posición 747 del ADNc, donde la TGZ15 carece de una guanina conservada, lo que resulta en una delección y cambio de marco de lectura (GG-A por GGG) codificando para el mismo aminoácido.

```

TGM2_HUMAN      275 GGCWVFAAVA-CTVLRCLGIPTRVVTNYSAHQNSNLLIEYFRNEFGEIQGD-KSEMIW 307
TGM2_CAVCU      275 GGCWVFAAVA-CTVLRCLGIPTRVVTNFSAHQNSNLLIEYFRNESGEIEGN-KSEMIW 307
TGM2_MOUSE      275 GGCWVFAAVA-CTVLRCLGIPTRVVTNYSAHQNSNLLIEYFRNEFGELESN-KSEMIW 307
TGM2_BOVIN      275 GGCWVFAAVA-CTVLRCLGIPTRVVTNYSAHQNSNLLIEYFRNEFGEIQSD-KSEMIW 307
TGM3_HUMAN      272 GGCWVFAAGTL-NTALRSLGIPSRVITNFSAHDTDRNLSVDVYDPMGNPLDK-GSDSVW 303
CONSENSUS      275 GGCWVFAAVA-CTVLRCLGIPTRVVTNYSAHQNSNLLIEYFRNEFGEIQSN-KSEMIW 307
F13A_HUMAN      313 GGCWVFAAGVF-NTFLRCLGIPARIVTNYFSAHDNDANLQMDIFLEEDGNVNSKLTKDSVW 343
PNG1_ARATH      250 GRCSEWANCFTLYCRTFGYDSRLIMD-----F 276
Q6KF70_MAIZE    265 MQVHIPATPLHISREPRLGHSRVL----- 289
EPB42_HUMAN     266 GQAWVLAAVA-CTVLRCLGIPARVVTTFASAQGTGGRLLIDEYYNEEGLQNGEGQRRIW 324
      : * :* :*::

TGM2_HUMAN      308 NHHCWVESWMTR-PDLQPGYEGWQALDPTPQEKSEGTCCGFPVPVRAIKEGDLSTKYDAP 391
TGM2_CAVCU      308 NHHCWVESWMTR-PDLEPGYEGWQALDPTPQEKSEGTCCGFPVPVRAIKEGHLNWKYDAP 391
TGM2_MOUSE      308 NHHCWVESWMTR-PDLQPGYEGWQALDPTPQEKSEGTCCGFPVSVRAIKEGDLSTKYDAP 391
TGM2_BOVIN      308 NHHCWVESWMTR-PDLQPGYEGWQALDPTPQEKSEGTCCGFPVPVRAIKEGDLSTKYDAP 391
TGM3_HUMAN      304 NHHVWNEGWVFR-SDLGPSYGGWQVLDATPQERSQGVFCGPASVIGVREGDVQLNFDMP 387
CONSENSUS      308 NHHCWVESWMTR-PDLQPGYEGWQALDPTPQEKSEGTCCGFPVPVRAIKEGDLSTKYDAP 391
F13A_HUMAN      344 NYHCWNEAWMTR-PDLPVGFGGWQAVDSTPQENSDCMYRCGPASVQAIKHGHVCFQFDAP 430
PNG1_ARATH      277 TDHWTECYSH-----SLKRWIHIIDPCEG-----VY-----DKP 304
Q6KF70_MAIZE    290 ----RLGYISREPRL----GHISREPRLGHSRGR-MG-HISRGLRIGHISREPRLG 337
EPB42_HUMAN     325 IFQTSTECWMTR-PALPQGYDGWQILHPSAENGGVGLGSDLVFPVRAVKEGLGLTPAVS 383

```

**Figura 6.** Alineamiento múltiple global de secuencias con el sitio catalítico de las TGs. La familia de las TGs en mamíferos (TGM2\_HUMAN, TGM2\_CAVCU, TGM2\_MOUSE, TGM2\_BOVIN, TGM3\_HUMAN, F13A\_HUMAN) poseen la región conservada del sitio catalítico al igual que en *Arabidopsis thaliana* (PNG1\_ARATH). La TGZ15 de maíz (Q6KF70\_MAIZE; Villalobos *et al.*, 2004) muestra el sitio catalítico en el C-terminal (Beninati *et al.*, 2013).

En las bases de datos de la UniProtKB existen 31 secuencias de proteínas anotadas funcionalmente como transglutaminasas, no obstante, solo existen 4 genes identificados en la NCBI gene con origen en un transcritto, pero sin reporte del gen *tgz* activo (Villalobos *et al.*, 2004). El caos y redundancia de información a partir de anotaciones inferidas, se ejemplifica claramente con esta proteína. En la NCBI gene se registra un gen *tgz21* procesado por splicing alternativo, que codifica a una proteína de 513 aminoácidos con código de acceso [NP\\_001105158.2](https://www.ncbi.nlm.nih.gov/nuccore/NP_001105158.2) y referenciado a: Villalobos *et al.*, (2004). Pero en la UniProtKB, esta variante se registra con el ID de acceso a B8A2N4 y con referencia en la publicación de: Soderlund *et al.*, (2009), como parte de proyectos de secuenciación.

La actividad TG en maíz y su actividad catalítica detallada hasta el momento, describen un rol importante en estructuras vegetales. En el presente, la TG activa de maíz está siendo ampliamente estudiada dentro de la industria alimentaria y es una propuesta interesante como enzima industrial substituta a la TGm en la manufactura del yogurt (Li *et al.*, 2013, 2014); la optimización de esta enzima provee una alta consistencia, cohesividad, índice de viscosidad y aparente viscosidad que los productos no tratados con TGZ (Li *et al.*, 2017). Basado en las relaciones previamente descritas entre las TGs vegetales, se pretende analizar la información anotada en las bases de datos con ayuda de herramientas bioinformáticas, con el fin de explicar las relaciones funcionales y evolutivas que guardan las proteínas no caracterizadas anotadas funcionalmente como TGs y la ya caracterizada TGZ15 de maíz.

### 3. JUSTIFICACIÓN

En las bases de datos proteicas existen 31 secuencias de aminoácidos de maíz anotadas funcionalmente como transglutaminasas y sin evidencia de su actividad enzimática, traducidas a partir de secuencias codificantes e inferidas por un umbral de similitud con un dominio tipo TG o por la composición parcial de su estructura primaria. No obstante, solo 4 secuencias proteicas poseen información génica con evidencia en un transcripto (es decir, a partir de ARNm). Paradójicamente la única TG de maíz secuenciada, clonada y caracterizada con evidencia enzimática descrita es la TGZ15, con un dominio catalítico no canónico. La información redundante en las bases de datos muestra una referencia de la TGZ15 activa de maíz hacia una isoforma proteica inactiva y registrada como producto de splicing alternativo. La diferencia entre ambas isoformas radica en un C-terminal divergente que promueve la ausencia de la tríada catalítica y una composición de regiones de baja complejidad en forma de TRs. Estos eventos son derivados de una mutación sinónima con un cambio de marco de lectura río abajo en la secuencia codificante.

Las *Poaceae* o gramíneas son un grupo de plantas monocotiledóneas en donde secuencias codificantes como las de la cebada y el sorgo presentan el mismo fenómeno mutacional que promueve la ausencia de la tríada catalítica. Asimismo, se tienen antecedentes de la existencia de un homólogo de la TGZ15 de maíz en caña de azúcar (*Saccharum officinarum*) y que hasta el momento no ha sido caracterizada. Por lo tanto, analizar la frecuencia mutacional dentro de esta familia vegetal a partir de análisis bioinformáticos permitirá revelar si la acumulación de este fenómeno que inhabilita su dominio activo es exclusiva del genoma de las gramíneas; así como también las implicaciones evolutivas, funcionales y estructurales que existen alrededor de la TGZ15 y su isoforma inactiva. Existe la oportunidad de validar la homología de la TGZ15 de maíz en caña de azúcar y que pueda demostrar o no la presencia de la mutación. Con este trabajo se pretende llegar a obtener un mayor entendimiento de la TGZ15 de maíz como TG a partir de la generación de nuevos hallazgos que reafirmen su actividad biológica desde una relación estructura-función.

#### 4. HIPÓTESIS

- Los productos codificantes de transglutaminasas putativas en maíz, localizadas en las bases de datos, presentan una función tipo transglutaminasa en función de los dominios conservados identificados, su estructura y sus relaciones evolutivas.
- La caña de azúcar (*Saccharum* sp. híbrido Mex 69-290) posee al menos una transglutaminasa putativa homóloga al gen *tgz15* de maíz (*Zea mays*) que presenta el desplazamiento del codón por inserción puntual de una guanina.

## 5. OBJETIVOS

### 5.1 Objetivo principal

Analizar mediante herramientas bioinformáticas las variaciones funcionales, evolutivas y estructurales que comparten las TGs putativas de maíz y la TGZ15 con sus homólogos más cercanos en plantas e identificar experimentalmente la presencia del gen *tg* en caña de azúcar a partir de ADNg, para con ello determinar la presencia o ausencia en el exón del cambio de marco de lectura (GG-A por GGGGA) que inhabilita el dominio catalítico de la TG como sucede con la isoforma inactiva de maíz.

### 5.2 Objetivos secundarios

- Comparar la información almacenada respecto a anotaciones funcionales inferidas como transglutaminasas que existen en las bases de datos primarias y secundarias.
- Analizar las diferencias de las estructuras primarias anotadas funcionalmente como transglutaminasas homólogas a las de maíz y determinar su relación evolutiva, funcional y estructural mediante herramientas bioinformáticas.
- Aislar, clonar, secuenciar y caracterizar el gen transglutaminasa de caña de azúcar homólogo a la *tgz15* activa de maíz.

## 6. MATERIALES Y MÉTODOS

### 6.1 Búsqueda y distribución de secuencias biológicas anotadas funcionalmente como transglutaminasas en las bases de datos biológicas

Para la realización de este trabajo se accedió a las bases de datos de la NCBI gene, NCBI protein y la UniProtKB. En estas bases de datos se encuentran organizadas secuencias biológicas anotadas con información de carácter biológico que se presentan en descripciones textuales y que son actualizadas con frecuencia. Esta información depositada contiene palabras clave en base a un tema particular de interés o que forman parte de la investigación de otro objeto de estudio.

Se inició una búsqueda genérica de las secuencias biológicas que se encuentran en las bases de datos alrededor del criterio de búsqueda: transglutaminase, para este fin se utilizó la plataforma en línea NCBI “Entrez query” (<https://www.ncbi.nlm.nih.gov/search/>). Con esta búsqueda se hizo un registro de secuencias de genes, proteínas, nucleótidos y genes producto por eventos de splicing alternativo. Es importante mencionar que se hizo un hincapié en las secuencias proteicas, por lo que se proyectó una representación porcentual empleando la base de datos de proteínas de la NCBI.

Para comparar la diferencia de secuencias registradas y anotadas con el término transglutaminase, se observó la información almacenada entre las secuencias de proteínas o péptidos. Para comparar el sesgo de información con respecto a la base de datos primaria anterior, se realizó una segunda búsqueda, pero ahora teniendo como referencia una base de datos secundaria como la UniProtKB. La información recopilada se clasificó de acuerdo con el tipo de evidencia que soporta la existencia de la proteína, revisión que posee por parte de expertos en datos (revisada o no revisada), nombre del organismo, nombre de la proteína, fragmento (sí o no), longitud de la proteína y peso molecular. La información analizada se representó por medio de gráficas de barras y de tipo pastel, donde se proyectó la cantidad de datos anotados que están relacionados al término transglutaminase.

## **6.2 Identificación y obtención de secuencias biológicas anotadas funcionalmente como transglutaminasas de maíz (*Zea mays*)**

Para identificar las secuencias génicas y proteicas con anotación funcional tipo TG en plantas se siguió la misma estrategia anterior utilizando la base de datos de la NCBI protein y UniProtKB. La extracción de datos se llevó a cabo mediante el módulo de búsqueda avanzada, utilizando la palabra clave “transglutaminase” acompañado del criterio taxonómico “viridiplantae”. Se representó gráficamente el número de genes anotados (codificantes y no codificantes), secuencias de proteínas, nucleótidos y genes procesados por eventos de splicing alternativo. Igualmente se graficó la representación porcentual de la abundancia de secuencias de proteínas tipo TG por organismo vegetal de acuerdo a la familia vegetal que pertenece.

## **6.3 Búsqueda de proteínas similares a las transglutaminasas de maíz (*Zea mays*)**

Las secuencias de genes y proteínas de maíz fueron extraídas de las bases de datos UniProtKB y NCBI gene. La selección de estas secuencias se determinó por la existencia de información génica en la NCBI gene y el tipo de evidencia experimental descrita. Las secuencias de proteínas seleccionadas fueron inspeccionadas mediante la herramienta en línea HMMSCAN, que forma parte del software HMMER (Finn *et al.*, 2015), con este primer análisis se registró la existencia de dominios ajustados a las proteínas. En regiones conservadas con algún patrón de funcionalización o de señalización las secuencias fueron inspeccionadas por la base de datos TargetP 1.1 server (Emanuelsson *et al.*, 2007) en unión con ChloroP 1.1 server (Emanuelsson *et al.*, 1999). Esto se llevó a cabo con el fin de obtener una mejor comprensión del conjunto de datos a analizar.

## **6.4 Manipulación y alineamiento de secuencias biológicas**

Para determinar el grado de similitud que guardan las proteínas seleccionadas con anotación funcional transglutaminase en maíz, se compararon mediante alineamientos locales pareados con el algoritmo Smith-Waterman implementado en el paquete de software FASTA de uso local (Pearson & Lipman 1988). La búsqueda de similitud y establecimiento de homología se construyó con ayuda de BLASTp 2.6.0 (Altschul *et al.*, 1997) con lo cual se seleccionaron los homólogos cercanos, se utilizó como conjunto las bases de datos nr, Swiss-Prot y TrEMBL. Los parámetros utilizados en la búsqueda comprendieron un tamaño de palabra 3 y un e-value de 0.001.

## **6.5 Reconocimiento de patrones de conservación en regiones biológicas funcionales de las transglutaminasas de maíz (*Zea mays*)**

El alineamiento de las secuencias permitió identificar regiones conservadas de residuos característicos en dominios o motivos biológicos fundamentales que presentan algún tipo de relación funcional entre las proteínas alineadas. Estas regiones son más conservadas a lo largo del tiempo evolutivo, y evolucionan como unidades que se ganan o se pierden como un solo módulo (Xiong 2006). Identificar y determinar estas regiones conservadas permitió analizar la divergencia evolutiva de las TGs que existe entre las proteínas vegetales y la familia TG en mamíferos. Para confirmar la existencia de sitios conservados en los alineamientos, se analizó cada conjunto de secuencias relacionadas mediante la plataforma MEME suite (Bailey *et al.*, 2009). En las regiones identificadas con una conservación evidente se utilizó la plataforma WebLogo Creator (Crook *et al.*, 2004) para crear diagramas tipo logos de secuencias consenso.

## **6.6 Construcción de árboles filogenéticos**

Para la reconstrucción de la filogenia molecular, cada alineamiento significativo (en formato FASTA) fue tomado como entrada para ser evaluado mediante la herramienta IQ-TREE versión 1.6.6 (Kalyaanamoorthy *et al.*, 2017) y determinar el modelo evolutivo más verosímil. Los resultados arrojaron un archivo de texto en formato newick, el cual contenía la información de los datos de distancias evolutivas entre cada organismo. Estos parámetros del modelo fueron introducidos de una manera similar al programa en línea phylogeny.fr en modo “Advanced” y el resultado obtenido fue modelado con el programa FigTree (Descargado desde: <http://tree.bio.ed.ac.uk/software/figtree/>) para la realización del árbol filogenético.

## **6.7 Predicción de la estructura proteica mediante técnicas de enhebrado y *ab initio***

Para la predicción estructural de las proteínas anotadas con el término transglutaminase de maíz, se implementaron técnicas de reconocimiento de plegamiento o por enhebrado independientemente de la similitud de sus secuencias molde. El reconocimiento de plegamientos ayuda a asignar plegamientos a secuencias blanco que tienen muy poca identidad de secuencia (<30%) hacia estructuras ya conocidas. Mediante este método, cada secuencia blanco es comparado contra una librería de plegamientos potenciales en templados y usando potenciales

de energía, el mejor modelo estructural será aquel que minimice la energía libre de la secuencia consulta. El templado con la puntuación de energía más bajo o la puntuación de similitud más alto es asumido como el mejor ajuste del plegamiento en las proteínas blanco (McGuffin 2008). Seguido a este método de enhebrado se implementó el método *ab-initio* para las regiones que no pueden ser cubiertas por los templados seleccionados, para esto se implementó el algoritmo Poing, el cual simplifica un simulador del plegamiento proteico en zonas no cubiertas con el bombardeo de residuos hidrofóbicos, resortes pronosticados de estructuras secundarias analizadas para mantener las conformaciones de  $\alpha$ -hélices o  $\beta$ -laminas y la prevención del choque esférico. Estos conjuntos de métodos computacionales son implementados por Phyre2 (Kelley *et al.*, 2015), el cual determinó el modelo final de la estructura tridimensional.

Phyre2 es una plataforma en línea que predice y analiza las estructuras proteicas, así como su función y mutaciones. La principal diferencia entre Phyre2 y otros servidores predictivos es que ofrece una interfaz intuitiva y simple. En el margen de utilizar un método que favoreciera una mejor calidad del modelo proteico se decidió utilizar Phyre2 como método predictivo de las proteínas de interés en este estudio, el resultado de la predicción fue un producto final en archivo .pdb y el análisis de la estructura en formato HTML, el modelo predicho se manipuló con el uso del software Chimera versión 1.13.1 (Pettersen *et al.*, 2004). Para determinar la calidad del modelo predicho se accedió al servidor SAVES v5.0 (<http://servicesn.mbi.ucla.edu/SAVES/>), haciendo uso de las herramientas Verify3D (Eisenberg *et al.*, 1997) y PROCHECK (Laskowski *et al.*, 1993) que se encargaron de medir la aceptación del modelo en referencia a proteínas completas que ya han sido reveladas su estructura tridimensional con buena aceptación.

El servidor Verify3D determina la compatibilidad de un modelo atómico (3D) a partir de su propia secuencia de aminoácidos (1D) junto con la asignación de sus ubicaciones y entornos de una clase estructural que finalmente son comparados con una colección de estructuras proteicas resueltas de alta calidad y que son usadas como referencia. Verify3D asigna puntuaciones de coincidencia para cada combinación de residuo de aminoácido y clase de entorno, denominando score 3D-1D, y como la puntuación de coincidencia total de la proteína depende solo de las clases de entorno, la coincidencia es independiente de las similitudes de secuencia entre la proteína de consulta y

las proteínas blanco que se usaron para generar la base de datos de información. El puntaje total de validación Verify3D es la suma de los puntajes 3D-1D de todos los residuos dentro de la secuencia. Los residuos con un score por encima de 0.2 deben ser considerados como confiables, por lo tanto, las secuencias por debajo de tal score tienden a proyectar un estado pobre de calidad en el modelo (Eisenberg *et al.*, 1997).

PROCHECK hace su tarea analizando la distribución de los ángulos de torsión en la estructura de la proteína mediante la representación de un diagrama de Ramachandran, el cual ofrece una visión de las regiones de torsión que están permitidas y no permitidas, y por tanto de la flexibilidad de una región para adoptar un determinado plegamiento. Los ángulos de torsión o ángulos de Ramachandran describen las rotaciones del esqueleto de la cadena polipeptídica alrededor de los enlaces entre N y C $\alpha$ . El ángulo *Phi* ( $\phi$ ) es la rotación sobre el enlace N-C $\alpha$ , que se mide por el ángulo entre un plano virtual formado por C-N-C $\alpha$  y el plano virtual por N-C $\alpha$ -C. Y el ángulo *Psi* ( $\psi$ ) que es la rotación sobre el enlace C $\alpha$ -C, el cual es medido por un ángulo entre el plano virtual formado por N-C $\alpha$ -C y el plano virtual C $\alpha$ -C-N. Los ángulos de torsión son uno de los parámetros estructurales más importantes que controlan el plegamiento de una proteína. Si se es capaz de predecir los ángulos de torsión de Ramachandran de una proteína, existe la oportunidad de predecir su plegamiento y su estructura secundaria (Laskowski *et al.*, 1993; Yagüe *et al.*, 2014).

## **6.8 Análisis de enriquecimiento con anotaciones GO obtenidas desde Ffpred3 a partir de las TGs de maíz (*Zea mays*)**

Para entender mejor el rol funcional de las proteínas anotadas como transglutaminase y la TGZ15 de maíz, se realizó un análisis de enriquecimiento a partir de ontologías de genes con la entrada de las secuencias de proteínas correspondientes. Se utilizaron métodos de agrupamiento o clustering a través del procesamiento de datos en el software Rstudio y de las librerías del paquete Bioconductor (Seefeld & Linder 2007; Team R.C., 2013). El análisis se representó gráficamente mediante una matriz de color, mejor conocido como mapa de calor (o en inglés, *heatmap*), que contenía los datos o scores visualizados por un color a partir de agrupamientos jerárquicos. El agrupamiento o clúster jerárquico se encuentra implementado en un algoritmo

incluido en Rstudio que agrupa objetos similares dentro de grupos llamados clústeres. El resultado final es un conjunto de clústeres, que es distinto el uno del otro, y el objeto o dato dentro de cada clúster es generalmente similar al otro. El objetivo fue identificar patrones de expresión similares entre las proteínas estudiadas, por lo tanto, se buscó y analizó el conjunto de términos enriquecidos GO asociados a cada proteína. Las ontologías de genes crean un vocabulario estructurado y controlado sobre anotaciones de genes, proteínas y secuencias (Gene Ontology Consortium. 2004) mediante 3 tipos de categorías como la función molecular (MF por sus siglas en inglés), proceso biológico involucrado (BP por sus siglas en inglés) y del componente celular al que pertenecen (CC por sus siglas en inglés).

En primera instancia se recuperaron los términos GO mediante el programa en línea Ffpred3 (Cozzetto *et al.*, 2016) tomando como entrada las estructuras primarias de las proteínas a analizar. Ffpred3 es un predictor de función basado en secuencias proteicas desde la transferencia funcional con base en la homología. Por lo que la plataforma designa términos de ontologías de genes a secuencias de proteínas cuando la homología con proteínas caracterizadas por análisis convencionales no proporciona la suficiente información complementaria. Las predicciones las realiza con el escaneo de secuencias a través de una matriz, que codifica las características de la secuencia de aminoácidos y las analiza mediante métodos de clasificación y regresión con la ayuda de una técnica de aprendizaje supervisado con clasificadores de SVM. Las predicciones examinan la relación entre la función de una proteína y los atributos biofísicos que describen la estructura secundaria, hélices transmembranales, las RIDs, péptido señal y otras regiones o motivos. Por lo que las anotaciones GO se encuentran entrenados para reconocer patrones de características asociadas con una función particular. La salida de cada modelo clasificador es una decisión binaria (sí o no) para cada término GO con un puntaje de confianza asociado que oscila entre 0 y 1. Es importante resaltar que FFpred3 es una herramienta diseñada en función de las proteínas humanas, sin embargo, Cozzetto *et al.*, (2016) mencionó que la herramienta también puede ayudar a anotar la función de la proteína para otros organismos eucariotas.

Las secuencias proteicas Q6KF70, B8A2N4, COPED4 y AOA1D6EW30 de longitud completa fueron analizadas por separado. Posteriormente FFpred3 entregó un archivo de texto con los resultados clasificados de la siguiente manera:

- Score: Probabilidad posterior para la anotación del término.
- Término GO: Nombre de la etiqueta a la que pertenece el término.
- RL: Nivel de confiabilidad (Alto o bajo).
- Dominio: Es el nivel de clasificación o dominio (MF y BP).
- Descripción: Nombre del término GO de manera más descriptiva.

Mediante manipulación manual de los resultados se tomaron aquellas ontologías con niveles de confiabilidad altos y clasificados de acuerdo con el dominio MF y BP. Finalmente se construyó un archivo limpio con los términos de las ontologías clasificadas y el score acompañado correspondiente a cada proteína analizada. Estos archivos de texto con los resultados incluidos fueron manipulados mediante el software Rstudio e implementado el método de agrupamiento jerárquico. Para la realización de este análisis se implementó un script realizado por Thomas Girke de la UC Riverside ([http://manuals.bioinformatics.ucr.edu/home/R\\_BioCondManual](http://manuals.bioinformatics.ucr.edu/home/R_BioCondManual)).

## **6.9 Análisis Molecular**

### **6.9.1 Material vegetal**

Los análisis moleculares se realizaron a partir de la colecta de hojas jóvenes y sanas de caña de azúcar (*Saccharum* sp. híbrido Mex 69-290) y maíz (*Zea mays*). El proceso de la colecta de hojas se realizó en campos de cultivo en San Juan Bautista Tuxtepec, Oaxaca. El corte del tejido foliar fue manipulado con tijeras y el uso de guantes para no maltratarlo, inmediatamente fueron rotuladas y guardadas en bolsas de plástico. El proceso final consistió en transportar las muestras del tejido vegetal hasta la Universidad Autónoma de Aguascalientes (UAA).

### **6.9.2 Cepa bacteriana y vector de clonación**

Se utilizó la cepa DH5 $\alpha$  de *E. coli* en la transformación de células competentes por choque térmico para la clonación del gen de interés junto con el vector de clonación pGEM<sup>®</sup>-T Easy (Promega).

### **6.9.3 Extracción de ADN de maíz (*Zea mays*) y caña de azúcar (*Saccharum* sp. híbrido Mex 69-290)**

Para la obtención de ADN se utilizó un protocolo adaptado por Bermúdez-Guzmán *et al.* (2016). El primer paso consistió en la lisis celular para lo cual se utilizó detergente CTAB (100 mM de Tris-HCl pH 8, 50 mM de EDTA pH 8, 1.4 M de NaCl, 3% (p/v) de CTAB), un minuto previo a la maceración del tejido se agregaron en tubos de 1.5 mL 800  $\mu$ L de solución amortiguadora CTAB junto con 10  $\mu$ L de 2- $\beta$ -mercaptoetanol concentrado y 5  $\mu$ L de proteinasa K (20 mg/mL), esto con el fin de evitar la rápida descongelación del tejido una vez dentro del tubo. En un mortero se pulverizó el tejido foliar junto con 0.3 g de PVP y nitrógeno líquido. Una vez obtenido un polvo fino del pulverizado se colocó de 20 a 100 mg (aproximadamente a la marca de 0.1 del tubo) en los microtubos de 1.5 ml con la solución de lisis celular previamente adicionado. Inmediatamente se mezclaron por inversión y se incubaron las muestras en un termobloque durante 60 min a 65 °C (para hojas de caña de azúcar) y 57 °C (para hojas de maíz) mezclando por inversión cada 5 min. El segundo paso fue la eliminación de proteínas y purificación, por lo cual se adicionaron a las muestras 800  $\mu$ L de solución fenol:cloroformo:alcohol isoamílico (25:24:1), se mezcló vigorosamente usando un vórtex y se centrifugó a 14,000 rpm a 4 °C durante 10 min, este paso se realizó dos veces. El sobrenadante recuperado se colocó en tubos nuevos de 1.5 mL. El último

paso consistió en la precipitación y limpieza del ADN, con las muestras del sobrenadante ya recuperado se precipitaron con 0.6 volúmenes de isopropanol frío y 0.1 volúmenes de acetato de sodio 3 M pH 5.2 durante toda la noche a -20 °C. Al día siguiente las muestras se centrifugaron a 14,000 rpm a 4 °C durante 10 min. El sobrenadante se decantó y se agregó 500 µL de etanol al 70%, se agitó manualmente hasta desprender la pastilla de ADN y se centrifugaron los tubos a 14,000 rpm a 4 °C durante 5 min, se decantó la fase acuosa y se invirtieron sobre papel absorbente para retirar el exceso de etanol. La pastilla de ADN se secó a temperatura ambiente y se resuspendió en 40 µL de agua estéril, se mezcló con micropipeta, se incubó a temperatura ambiente por 5 minutos y se almacenó a -20 °C.

#### **6.9.4 Verificación de la integridad y cuantificación de la concentración de ADN obtenido**

La calidad e integridad del ADN extraído se verificó por electroforesis en gel de agarosa al 1% con un medio de buffer TAE 1X a 70 V. El gel fue teñido por 1 min en bromuro de etidio y añadiendo 15 minutos en agua para eliminar el exceso de bromuro. Posteriormente el gel fue visualizado bajo luz ultravioleta en el transiluminador DyNA Light Dual Intensity UV Transiluminator (de Labnet). Además, se cuantificó la concentración de ADN mediante el uso de un espectrofotómetro de microvolúmenes Colibri de TITERTEK BERTHOLD, utilizando densidad óptica a 260 nm, con una pureza valorada en la relación 260/280 nm.

#### **6.9.5 Diseño de oligonucleótidos para la identificación del gen *tg* en caña de azúcar (*Saccharum* sp. híbrido Mex 69-290)**

Los oligonucleótidos utilizados (ver Tabla 4) fueron diseñados a partir de la secuencia molde de ARNm del gen *tgz15* de maíz (Villalobos *et al.*, 2004) disponible en la base de datos ENA (por sus siglas en inglés, European Nucleotide Archive) con el código de acceso CAD13455.1. y sintetizados por la compañía T4 Oligo. El diseño de los oligonucleótidos específicos fue llevado a cabo por el programa DNASTar (DNASTAR, Inc.) utilizando la herramienta Primer Select 5.0. Los oligonucleótidos fueron resuspendidos en agua estéril con una concentración final de 100 µM. Para el uso posterior de los oligonucleótidos en la PCR, se prepararon viales en tubos de 1.5 mL, en diluciones 1:10 para cada par de oligonucleótidos con un volumen final de 50 µL (45 µL de H<sub>2</sub>O y 5 µL de oligonucleótido).

**Tabla 4.** Oligonucleótidos utilizados para la amplificación del gen *tg* en caña de azúcar.

Nombre	Gen	Secuencia	Tamaño esperado del amplicón	Tm °C
TG3F TG3R	<i>tgz15</i>	F- 5' GCCGCTCACCTTTGCAGTCTTCAT 3' R- 3' CTGGGCTCCTCTGGTTTGGTCATA 5'	1447 pb	63

#### 6.9.6 Amplificación por PCR del gen *tg* en caña de azúcar (*Saccharum* sp. híbrido Mex 69-290)

Para aislar el gen de interés en caña de azúcar, primero se realizó una PCR control del gen *tgz15* en maíz. Se utilizó un control positivo a partir de ADN extraído de maíz (diluido a 100, 50 y 20 ng) y un control negativo con una reacción que sólo contenía H<sub>2</sub>O estéril. Para la amplificación de fragmentos se empleó el kit comercial GoTaq® DNA Polymerase (Promega). La preparación de la mezcla de reacción fue llevada a un volumen final de 45.96 µL (ver Tabla 5), después fue dividida en 4 microtubos de 0.2 mL con un volumen de 11.49 µL y para finalmente añadir 1 µL del ADN diluido con un volumen final de 12.49 µL.

**Tabla 5.** Componentes de la mezcla de reacción para la PCR.

Componente	Cantidad (1 reacción)	Cantidad (4 reacciones)
Buffer	2.5 µL	10 µL
MgCl <sub>2</sub>	1.25 µL	5 µL
dNTPs	0.5 µL	2 µL
Oligonucleótido (→)	0.5 µL	2 µL
Oligonucleótido (←)	0.5 µL	2 µL
Taq polymerase	0.062 µL	0.248 µL
H <sub>2</sub> O	6.18 µL	24.72 µL
<b>Volumen Final</b>	<b>11.49 µL</b>	<b>45.96 µL</b>

Las condiciones de amplificación fueron: 1 ciclo de 95° C 2 min (pre-desnaturalización); 30 ciclos de 95° C por 0.5 min (desnaturalización), 63° C por 0.5 min (alineamiento) y 72° C por 1 min (extensión); por último 1 ciclo de 72° C por 1 min (extensión final). El proceso de la PCR fue llevado a cabo en un termociclador con gradientes SelectCycler™ II. Finalmente, se verificó por electroforesis la amplificación del fragmento obtenido junto a un marcador de peso molecular de 1 kb (Promega) en un gel de agarosa al 1% en un medio con buffer TAE 1x a 70 V. El gel fue teñido por 1 min en bromuro de etidio y añadiendo 15 minutos en agua para eliminar el exceso de bromuro.

### 6.9.7 Purificación y clonación de los fragmentos amplificados

Los fragmentos amplificados por PCR fueron purificados con el kit comercial Wizard® SV Gel and PCR Clean-Up System (Promega), siguiendo el protocolo descrito por el fabricante. Los fragmentos purificados fueron clonados en el vector pGEM®-T Easy (Promega), igualmente siguiendo las especificaciones del fabricante. En primera instancia se realizaron ligaciones (ver Tabla 6) con las siguientes proporciones inserto:vector 3:1, 1:3 y 1:1. Y se calculó las proporciones del inserto usando la siguiente ecuación:

$$\frac{(\text{ng de vector}) * (\text{tamaño kb inserto})}{(\text{tamaño kb vector})} * \frac{\text{proporcion molar inserto}}{\text{proporción molar vector}} = \text{ng inserto}$$

La reacción de ligación se incubó toda la noche a 37 °C. Posteriormente se transformaron células competentes de *E. coli* DH5α.

**Tabla 6.** Componentes de reacción para ligación de productos de PCR.

Elemento	Cantidad
Buffer de ligación 2x	5 µL
Vector pGem-T Easy (Promega)	1 µL
Ligasa T4	0.5 µL
H <sub>2</sub> O	X µL
Inserto	X µL
<b>Volumen total</b>	<b>10 µL</b>

Para la preparación de células competentes se siguió el protocolo descrito en el Anexo 11.1. La mezcla de ligación fue utilizada para transformar células competentes de *E. coli* DH5α por choque térmico (siguiendo el protocolo descrito en el Anexo 11.2), finalmente, las bacterias transformadas se sembraron en cajas de petri con medio LB (Luria-Bertini) sólido adicionado con 100 mg/L de ampicilina. El crecimiento de colonias en el medio de selección demuestra la presencia de bacterias posiblemente transformadas, así que se aislaron las clonas positivas en nuevas placas petri con medio LB sólido y en tubos de ensayo con medio LB líquido, ambos adicionados con ampicilina (100 mg/L).

### 6.9.8 Extracción de plásmidos y verificación de ligación por enzimas de restricción

La extracción de ADN plasmídico se llevó a cabo mediante el protocolo descrito por Birboim & Doly (1979) basado en la extracción de plásmidos por un método alcalino. Se realizó el

procedimiento de acuerdo a lo descrito en el Anexo 11.3. El escrutinio de las clonas positivas (vector más el inserto de interés) fue realizado mediante la digestión de los plásmidos extraídos con la enzima EcoRI (Invitrogen, No. Catálogo: 15202-013). Las condiciones de las digestiones realizadas fueron las siguientes: para cada reacción se colocaron 5 µL de ADN plasmídico, 1 µL de buffer para EcoRI 10X, 0.3 µL de enzima EcoRI, 0.3 µL de ARNasa, 0.15 µL de BSA y 3.05 µL de H<sub>2</sub>O para ajustar un volumen final de 10 µL. La digestión fue incubada a 37 °C durante toda la noche y posteriormente se verificó por electroforesis en gel de agarosa al 1%, en un medio con buffer TAE 1x a 70 V. El gel fue teñido por 1 min en bromuro de etidio y añadiendo 15 minutos en agua para eliminar el exceso de bromuro, para ser verificado en luz ultravioleta.

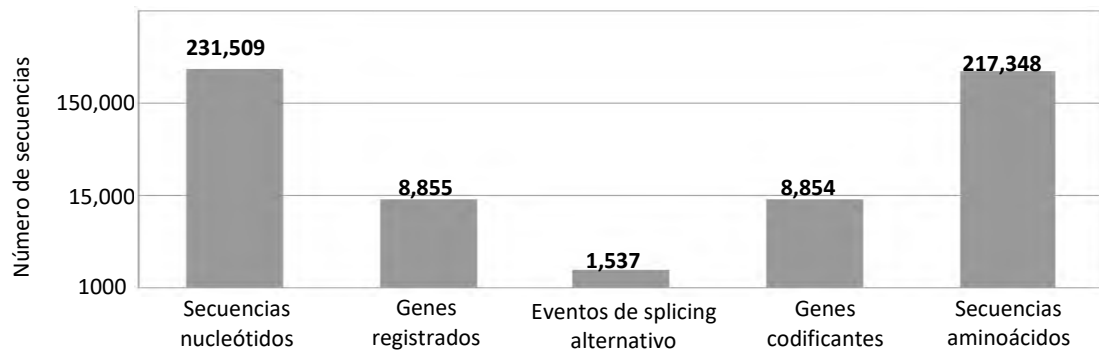
#### **6.9.9 Secuenciación y análisis de secuencias**

Las muestras de ADN plasmídico fueron enviadas al Instituto Potosino de Investigación Científica y Tecnológica (IPICYT) para su secuenciación por el método de Sanger. El resultado obtenido fue comparado por búsqueda de similitud a través de la herramienta BLASTn de la NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) para confirmar su homología. La traducción de la secuencia codificante a proteína a través de la plataforma en línea ExpASY translate tool (<https://web.expasy.org/translate/>) permitió revelar la composición de sus aminoácidos a partir de su marco de lectura. La secuencia de aminoácidos fue comparada por análisis de similitud con la base de datos del genoma de la caña de azúcar (<http://sugarcane-genome.cirad.fr>). Por último, se implementó un alineamiento múltiple con el programa ClustalW (Larkin *et al.*, 2007) a partir de la secuencia de nucleótidos obtenida con otras secuencias de maíz aisladas de ADNg y ARNm previamente reportadas por Villalobos (2007).

## 7. RESULTADOS Y DISCUSIÓN

### 7.1 Análisis comparativo de la distribución y organización de secuencias biológicas anotadas funcionalmente como transglutaminasa en las bases de datos biológicas

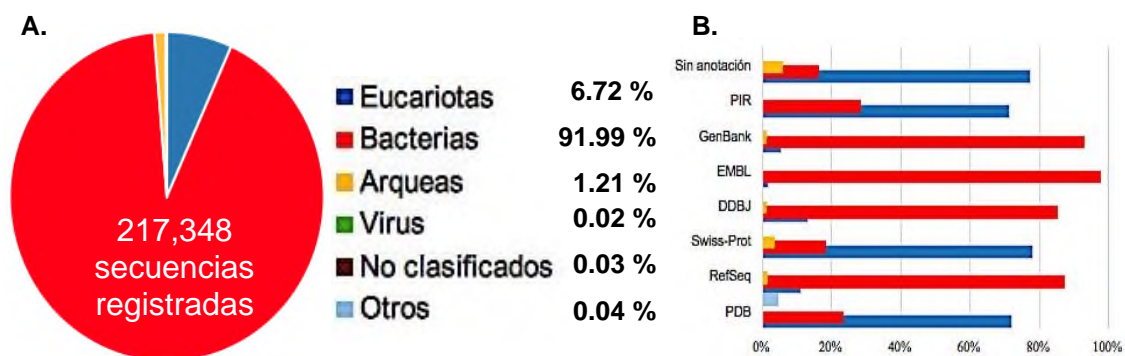
De las anotaciones almacenadas en las bases de datos de la NCBI gene, se encontraron 8,855 secuencias génicas anotadas con el término transglutaminase (ver Figura 7). De los genes registrados se observó un ARN no codificante (ARNnc) de origen eucariota que es generado de un modelo predicho encontrado en ácaros (*Galendromus occidentalis*), esta secuencia génica se encuentra registrada como LOC100906054 (NW\_003805109.1) y se anota como substrato de TG con un peso de 8.6 kDa; no obstante, su anotación funcional TG no demuestra actividad biológica comprobable. Además, dentro del total de genes registrados, se observaron 1,537 (17.35%) eventos de splicing alternativo. El splicing alternativo posee múltiples estudios donde se analiza la presencia de estas variantes en TGs, se piensa que la regulación diferencial de estas expresiones en TG tisular o TG2h contribuyen a las patologías de las enfermedades celíacas e incluso la de modular sus roles en el cáncer, neurodegeneración, inflamación y cicatrización de la herida en personas que la padecen, de manera que altera la funcionalidad proteica (Citron *et al.*, 2002; Lai & Greenberg 2013).



**Figura 7.** Distribución cuantitativa de secuencias de nucleótidos y aminoácidos con anotación transglutaminasa en las bases de datos de la NCBI. Con el criterio de búsqueda transglutaminase, se encontró el número de secuencias almacenadas y la anotación de genes con localización cromosómica. Nota: es importante recordar que las bases de datos son actualizadas con frecuencia y solo se pretende proyectar una relación porcentual de la cantidad de datos redundantes que pueden existir.

En cambio, en la NCBI protein, se observaron 217,348 secuencias de aminoácidos (ver Figura 8) las cuales surgen de diferentes bases de datos unificadas. Muchas de estas secuencias de

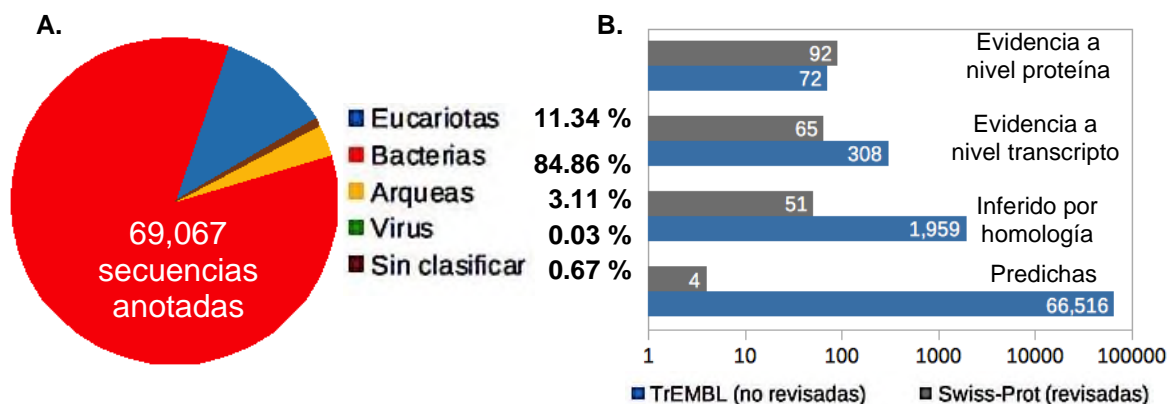
aminoácidos son traducciones a partir de ADN<sub>g</sub> por métodos computacionales y que se le infiere una anotación funcional TG por un umbral de similitud por composición de aminoácidos. Si bien la cantidad de secuencias proteicas no refleja la realidad de proteínas caracterizadas, esto es, debido a que gran parte de la información depositada proviene de secuencias interrumpidas o fragmentos, isoformas o variantes codificados por el mismo gen que se almacenan como entradas separadas, así como ORFs predichos de los cuales no se tiene certeza que codifiquen algún producto. Las bases de datos unificadas en la NCBI exhiben su naturaleza redundante debido a que son datos originales sin ningún procesamiento posterior que verifique su autenticidad, lo que puede ocasionar que una misma entrada este repetida varias veces para cada proyecto de secuenciación de nucleótidos obtenido por laboratorios diferentes.



**Figura 8.** Organización de secuencias proteicas con anotación transglutaminase en las bases de datos de la NCBI protein. **A.** De acuerdo al criterio taxonómico, las bacterias son el grupo más grande con 199,938 (91.99%) secuencias almacenadas. **B.** Con respecto a las bases de datos, existen 49 secuencias sin anotación y pertenecientes a arqueas, bacterias y eucariotas, de las cuales ninguna presenta vinculación con las bases de datos de la NCBI. Nota: las secuencias “No clasificados” son provenientes de metagenomas y “Otros” son las secuencias provenientes de construcciones sintéticas.

La mayoría de las veces, los datos redundantes pueden ser erróneos lo que ocasiona análisis inconsistentes (Yagüe *et al.*, 2014). A nivel estructural, las bases de datos vinculadas a la NCBI protein registran 153 secuencias depositadas en PDB y 7 secuencias en PIR, esta diferencia es evidente debido al procesamiento de su información que forma parte de la clasificación de las bases de datos biológicas. Como se puede observar, la cantidad de datos anotados en la NCBI protein es demasiado alto y sigue en aumento, el registro de información a partir de datos

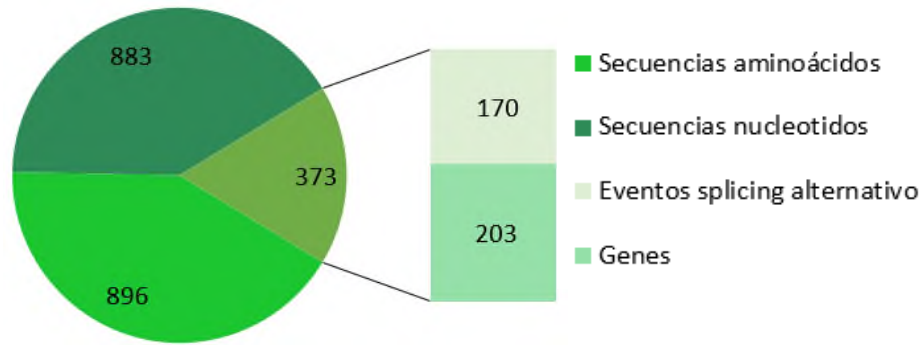
experimentales y sin revisión exhaustiva realizada por esta base de datos primaria ofrece redundancia de información y anotaciones muchas veces errónea (Choudhuri S. 2014). Para comparar el sesgo de información biológica en función de la cantidad de información anotada, se observó la cantidad de secuencias proteicas registradas en la UniProtKB, donde existen 69,067 secuencias registradas con el término de búsqueda transglutaminase, y además soportadas por un criterio de existencia (ver Figura 9). Solo 2,795 (4.05%) secuencias son fragmentos de proteínas que generalmente corresponden a productos de ORFs alternativos y de isoformas procesadas por splicing alternativo, la mayoría poseen un nivel de existencia predicha. El criterio de existencia a nivel predicha conforma la mayor parte del conjunto de secuencias registradas, con una cantidad de 66,520 secuencias de aminoácidos (96.31%) anotadas como transglutaminase, actualmente solo 4 de estas secuencias y además de origen bacteriano poseen una revisión completa. El criterio de existencia a nivel proteico cuenta con 164 (0.24%) secuencias, en su mayoría por la familia TG en mamíferos y bacterias, en este nivel de evidencia se encuentran vinculadas 29 estructuras terciarias ampliamente estudiadas.



**Figura 9.** Organización de secuencias proteicas con anotación transglutaminase en la base de datos UniProtKB. De acuerdo con la UniProtKB: **A.** Existen almacenadas 69,067 secuencias de proteínas con un registro mayoritario en bacterias (84.86%). **B.** Cada secuencia se encuentra anotada a un criterio de evidencia proteico que representa la calidad de su anotación y la revisión que obtiene por parte de los expertos Las secuencias predichas y no revisadas representan el mayor porcentaje de criterio de existencia proteica presente (96.30%) y el menor número con revisiones presentes (0.0057%). Nota: Las secuencias “sin clasificar” pertenecen a metagenomas. Es importantes recordar que estas cantidades solo reflejan una proyección porcentual que están sujetas a cambios continuos.

### 7.1.1 Identificación de secuencias biológicas vegetales con anotación funcional transglutaminase

Solo el 4.22% de los genes anotados con el término transglutaminase en las bases de datos NCBI gene son de origen vegetal, todos ellos codificantes. Sin embargo, una parte de estos genes codificantes en plantas son procesados de splicing alternativo (ver Figura 10).



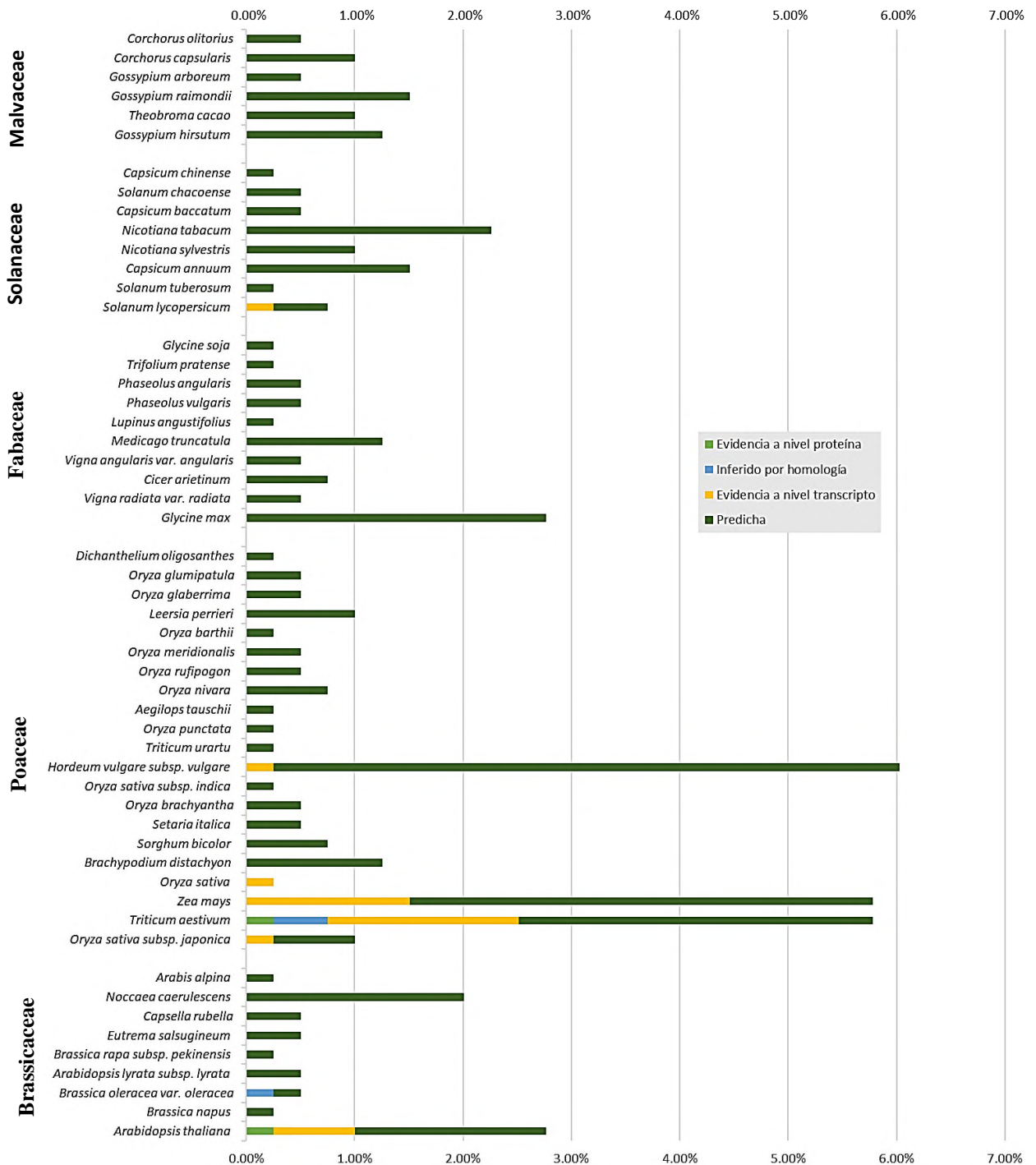
**Figura 10.** Distribución de secuencias de nucleótidos y aminoácidos de origen vegetal con anotación transglutaminase en las bases de datos de la NCBI gene y NCBI protein. Con el criterio de búsqueda: (transglutaminase) AND viridiplantae[Organism], fueron hallados 373 secuencias de genes codificantes putativos a TG. Nota: Las bases de datos son actualizadas con frecuencia y solo se pretende proyectar una relación porcentual de los datos.

En la NCBI protein no se encontró presencia de estructuras terciarias vinculadas a la base de datos PIR o PDB. Sin embargo, en las bases de datos de UniProtKB, la información almacenada mostró un escenario distinto. Se observaron 398 secuencias de proteínas anotadas funcionalmente como transglutaminase en plantas, donde solo 13 (3.47%) secuencias de aminoácidos presentan una revisión completa (Swiss-Prot) y 10 de ellas registran una estructura terciaria resuelta por el método de difracción de rayos X, las cuales son nombradas como Alpha/beta-gliadin MM1 de trigo (*Triticum aestivum*) pertenecientes a la familia gliadina/glutenina y que se anotan como substratos de TG. Las gliadinas son un tipo de proteínas prolaminas presentes en trigo y otros cereales del género *Triticum*, tanto las gliadinas como las gluteninas son dos componentes principales del gluten en la semilla de trigo. Está demostrado que la TG2h tiene una alta especificidad de sustrato por proteínas tipo gluten y sus péptidos derivados, debido a la abundante presencia del aminoácido glutamina en esas proteínas sustrato (Facchiano *et al.*, 2006). La desamidación de los péptidos de gliadina derivados del gluten por la TG2h es

responsable de la toxicidad inducida por la gliadina y la respuesta inmune en la mucosa del intestino delgado en humanos (Rauhavirta *et al.*, 2013).

Con estos métodos de búsqueda se encontró el registro de 31 secuencias de proteínas anotadas funcionalmente como transglutaminasas de maíz, donde se incluyen las ya caracterizadas TGZ15 y TGZ21 (Q6KF70 y Q6KF61), anotadas como proteínas sin revisión (TrEMBL). Ninguna de estas dos estructuras primarias presentó alguna anotación de dominio tipo TG descrita, a pesar de que existen evidencias experimentales de actividad enzimática vía conjugación de PAs y reticulación (Villalobos *et al.*, 2004; Li *et al.* 2013; Li *et al.* 2014). Del total de proteínas o péptidos con anotación TG en plantas, solo 28 (7.03%) secuencias son fragmentos o secuencias parciales. En general, son 113 especies vegetales que presentan secuencias primarias con anotación tipo transglutaminase.

En el gráfico de la Figura 11 se muestra la abundancia porcentual de secuencias registradas con respecto al total de secuencias de la base de datos UniProtKB, mostrándose sólo las familias botánicas con mayor representatividad. Las gramíneas representan la familia vegetal con mayor número de secuencias registradas (108), donde se incluyen las TGs activas de maíz (*tgz15/tgz21*; Villalobos *et al.*, 2004) y su homóloga en arroz (*tgo*; Campos *et al.*, 2013). La segunda familia vegetal más importante son las *Brassicaceae*, en este grupo se concentra la TG putativa de *Arabidopsis thaliana*, AtPNG1p (Della Mea *et al.*, 2004a), actividad TG que aún continúa en duda ya que se le ha atribuido mayor actividad como enzima péptido N-glicanasa (Diepold *et al.*, 2007). En esta búsqueda no se encontró ninguna proteína anotada tipo TG en caña de azúcar (*Saccharum officinarum*). El genoma de caña de azúcar posee una alta complejidad, es por eso que su secuenciación fue una tarea difícil. No fue sino hasta julio del 2018 que un equipo del CIRAD pudieron secuenciar el genoma de *Saccharum officinarum* (<http://sugarcane-genome.cirad.fr>) que cuenta con la presencia de 10 a 12 copias de cada cromosoma. La complejidad radica en la gran cantidad de copias de cromosomas (altamente poliploide), número variable de copias (aneuploidía), origen bioespecífico del cromosoma, diferencias estructurales y cromosómicas interespecíficos (Garsmeur *et al.*, 2018). Los investigadores se dieron a la tarea de añadir la predicción de 25,315 genes codificantes a proteínas.



**Figura 11.** Representación porcentual del número de proteínas en familias vegetales con anotación funcional transglutaminase. El porcentaje de abundancia de TGS vegetales a nivel familia taxonómica se determinó usando los registros de la UniProtKB. Se tomó como referencia el total de secuencias almacenadas y la proporción de anotaciones de la existencia de proteínas por organismo. Nota: Los datos registrados fluctúan constantemente con respecto al tiempo.

## 7.2 Identificación y selección de proteínas transglutaminasas de maíz (*Zea mays*)

Se seleccionaron 4 secuencias de proteínas que son registradas con la anotación funcional transglutaminase de maíz (ver Tabla 7). De acuerdo con sus descripciones, ninguna proteína cuenta con evidencia experimental descrita, y solo son respaldadas por la existencia de información genómica detallada en la NCBI gene (ver Anexo 11.4) junto con el registro de sus secuencias de aminoácidos depositados en la UniProtKB con evidencia a nivel transcripto (ARNm).

Los antecedentes demuestran que la TGZ15 es la única TG caracterizada de maíz con actividad experimental descrita y aplicaciones dentro de la industria alimentaria como substituta de la TGm en procesos alimentarios (Li *et al.*, 2013; Li *et al.*, 2014). La TGZ15 de maíz no posee información genómica reconocible en la NCBI gene ni tampoco un dominio característico TG, pero existe información acerca de su tríada catalítica putativa y actividad enzimática que se detalla ampliamente en el estudio de Villalobos *et al.*, (2004) y Villalobos (2007). La proteína TGZ15 fue referente de estudio para analizar la relación funcional y evolutiva que guarda con otras proteínas TGs.

**Tabla 7.** Proteínas seleccionadas. Cada proteína seleccionada contiene la información integrada de acceso a la secuencia de aminoácidos y al gen, así como de su localización en el cromosoma. El identificador del gen provee la información más representativa encontrada en las bases de datos de la NCBI gene. Nota: El símbolo de la proteína solo representa la identificación de la secuencia en este estudio. Los genes codificantes que pertenecen a las secuencias NM\_001111688.2 y NM\_001361416.1 son anotados como genes procesados por eventos de splicing alternativo.

Símbolo Proteína	ID NCBI (Nucleótido)	ID UniProtKB (Proteína)	ID NCBI Gen	Cromosoma	Referencia Publicación
TGZi	NM_001111688.2	B8A2N4	542049	2	Villalobos <i>et al.</i> , (2004)
TGZc2	NM_001320429.1	COPED4	100278776	3	Schnable <i>et al.</i> , (2009)
TGZi	NM_001254887.2	C0HGH1	100285057	10	Soderlund <i>et al.</i> , (2009)
TGZc	NM_001361416.1	A0A1D6EW30	100304380	2	Alexandrov <i>et al.</i> , (2009)
TGZ15	AJ421525.2	Q6KF70	Sin registro	2	Villalobos <i>et al.</i> , (2004)

### 7.2.1 Análisis de secuencias seleccionadas de las transglutaminasas de maíz (*Zea mays*)

Se realizaron alineamientos globales comparando el nivel de similitud-identidad en la composición de aminoácidos de las TGs putativas y la TGZ15 de maíz. Con el programa FASTA e implementando el algoritmo Needleman-Wunsch se observó que solo las secuencias B8A2N4 y COHGH1 presentaron un grado de identidad-similitud del 52.8-67% y 50.2-63.9% respecto a la TGZ15. Del alineamiento (ver Anexo 11.5) se observó una composición de aminoácidos altamente conservada en la región N-terminal y adyacente a ella (que va de la posición 1-249 y finaliza con una Gly conservada), posteriormente la composición de aminoácidos cambia drásticamente originando una divergencia hacia el C-terminal, caracterizado por la ausencia de residuos que participan en la tríada catalítica (Cys/His/Asp). Por lo tanto, se puede deducir que las secuencias B8A2N4 y COHGH1 son isoformas inactivas (nombrándolas TGZi) de la TGZ15.

Con el análisis de las secuencias codificantes (ver Figura 12) pertenecientes a las secuencias proteicas: TGZ15 (AJ421525.2), B8A2N4 (NM\_001111688.2) y COHGH1 (NM\_0012544887.1). El resultado demostró que la Gly249 entre las proteínas se encuentra conservada debido a que los codones sinónimos GGG y GGA codifican para el mismo aminoácido, evento mutacional que ya ha sido descrito por Sobieszczuk-Nowicka *et al.*, (2015).

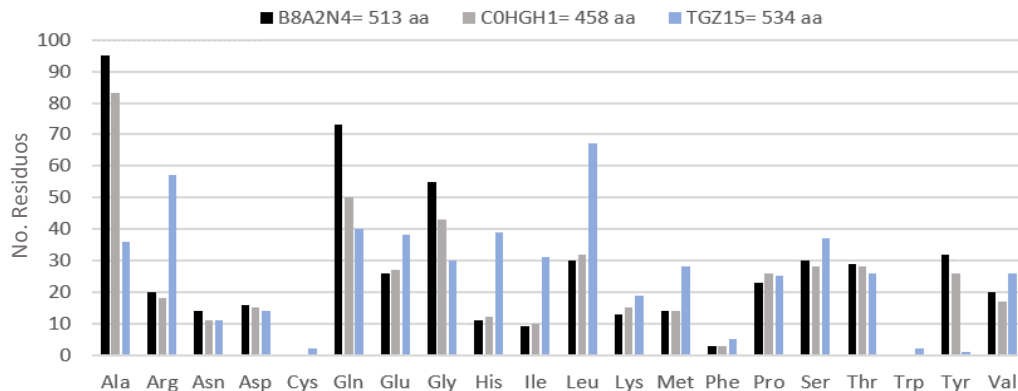
```

                                     739       749       759       769       779
NM_0012544887.1  730 ACAGCTTCACAGCCTGGGACAGCTCAAGCTACAGCGGTTCCAGCAG---C 1377
NM_001111688.2  730 ACAGCTTCGCAGCCTGGGACAGCTCAAGCTACCACGGTTTCAGCAGCAGC 1542
AJ421525.2       730 ACAGCTTCGCAGCCTGGGACAGCTCAAGCTACCACGGTTTCAGCAGCAGC 1605
```

**Figura 12.** Alineamiento de las secuencias codificantes *tg* de maíz delimitadas por la zona de mutación.

La isoforma TGZi es codificada por dos variantes génicas (NM\_001111688.2 y NM\_0012544887.1) localizadas en diferentes cromosomas (2 y 10) en el genoma de maíz (variedad B73). Solo la proteína B8A2N4 es generada a partir de splicing alternativo y se diferencia de COHGH1 por la longitud de su estructura primaria. Ambos concentran la misma composición de aminoácidos ricos en Gln y Ala (ver Figura 13). Estas isoformas generan una posible divergencia funcional respecto a la TGZ15, donde convergen dos mecanismos de variabilidad funcional dada por la duplicación de genes y el splicing alternativo. Al respecto, existen investigaciones que describen una diversidad proteica con un profundo impacto en la secuencia de aminoácidos, estructura y

función atribuido tanto por el splicing alternativo como por la duplicación génica (Lynch & Conery 2000; Graveley 2001; Talavera *et al.*, 2007).



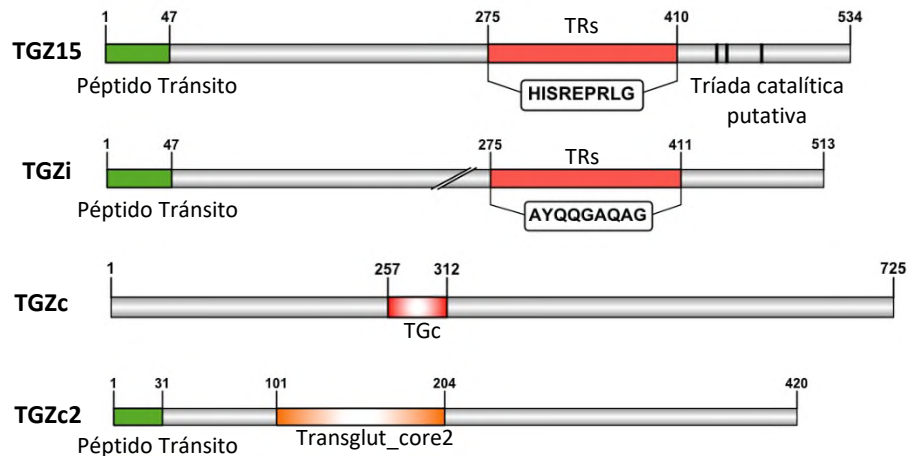
**Figura 13.** Diferencia entre la composición de los aminoácidos de las secuencias TGZ15 activa, B8A2N4 y COHGH1 (TGZi) de maíz.

Si se toma en cuenta que la proteína TGZ15 es codificado por un gen de copia única (Villalobos *et al.*, 2004), puede considerarse que actúa como un gen ancestral, por lo que la duplicación génica que derivó en la formación de una isoforma inactiva (TGZi) a partir de splicing alternativo obedece lo descrito por Bush *et al.*, (2017), señalando que las copias duplicadas se retienen y divergen para parecerse a isoformas de splicing alternativo producidas por el gen ancestral de una sola copia. El splicing alternativo puede promover la retención de genes duplicados contribuyendo a la generación de nuevas funciones en las isoformas o facilitando la subfuncionalización, donde las copias génicas adquieren mutaciones recíprocas para que las copias duplicadas cumplan la función genética del original, que evita que se conviertan en pseudogenes (Lynch & Force 2000; Su & Gu 2012).

### 7.2.2 Reconocimiento de regiones funcionales dentro de las transglutaminasas de maíz (*Zea mays*)

Con el fin de encontrar la relación de funcionalidad TG, se analizaron las secuencias de las proteínas putativas y la TGZ15 de maíz. Cada secuencia de aminoácidos se sometió a una búsqueda con el servidor ChloroP (Emanuelsson *et al.*, 1999), el cual, mostró la presencia de un péptido tránsito al cloroplasto en las proteínas: TGZi, TGZ15 y TGZc2 (ver Figura 14). Anteriormente Villalobos *et al.*, (2004) ya había reportado la presencia de un péptido tránsito a cloroplasto en TGZ15 con una longitud de 47 aminoácidos y que además fue demostrado por

Campos *et al.*, (2013). El segmento de señalización hacia el cloroplasto hallado en la proteína TGZc2 comprende 31 aminoácidos (ver Anexo 11.6) y es favorecido por un score mayor (0.558) que el de la enzima TGZ15, el cual mostró una señalización preferente hacia la mitocondria (0.623). No obstante, Villalobos (2007) demostró la localización con el uso de anticuerpos la presencia de la TGZ15 de maíz en cloroplastos, confirmando trabajos previos sobre el papel de las TGs en la fotosíntesis (Margosiak *et al.*, 1990; Del Duca *et al.*, 1994; Bernet *et al.*, 1999).



**Figura 14.** Diagrama esquemático de los elementos biológicos funcionales hallados en las proteínas anotadas como transglutaminasas en maíz. Con la información recopilada se construyeron bloques esquemáticos desde la plataforma IBS 1.0.3 (Liu *et al.*, 2015).

La identificación de dominios funcionales se determinó con la ayuda de la herramienta en línea HMMSCAN. La TGZ15 posee un dominio no canónico, por lo tanto, no presenta un dominio funcional registrado en las bases de datos. Además, no existe suficiente información a nivel estructural que demuestre su homología y aún no se ha confirmado la ubicación exacta de su tríada catalítica (Villalobos *et al.*, 2004; Villalobos 2007) pese a haber sido ya demostrada su actividad enzimática (Villalobos *et al.*, 2004, Li *et al.*, 2013, Li *et al.*, 2014, Li *et al.*, 2017). Por otra parte, las estructuras primarias de TGZc y TGZc2 presentaron los dominios TGc (SM00460) y Transglut\_core2 (PF13369) con una composición de aminoácidos muy diferentes entre sí (ver Anexo 11.7). En la actualidad, la anotación del dominio TG está representado por un perfil de secuencias consenso depositados en la base de datos PFAM con el identificador PF01841 y en la base de datos SMART con el identificador SM00460. En el caso del dominio Transglut\_core2 se identifica como PF13369, y el registro de su información biológica es nulo. Debido a esto, la

identificación del dominio funcional en la proteína TGZc2 fue inspeccionada a detalle, y se encontró que esta proteína presenta una anotación parcial del dominio.

La longitud completa del modelo HMM que contiene el perfil de secuencias PFAM con el dominio PF13369 es de 155 aa, mientras que la longitud que presenta la proteína con respecto a este alineamiento es de solo 60 aa. Los dominios parciales son en gran medida el resultado de artefactos de alineación y su anotación debe ser tomada con precaución. Así que la presencia de anotaciones parciales de dominios en las proteínas, plantea anotaciones inciertas al momento de que la predicción del gen de la proteína pueda ser incompleta (Triant & Pearson 2015). Existe una fuerte evidencia de que alineaciones parciales son el resultado de errores en la identificación de la secuencia de la proteína en el dominio truncado, y esto es particularmente frecuente en las secuencias eucariotas de especies mal anotadas (Kelley & Sternberg 2015).

### **7.3 Análisis de similitud con proteínas similares a las TGs de maíz**

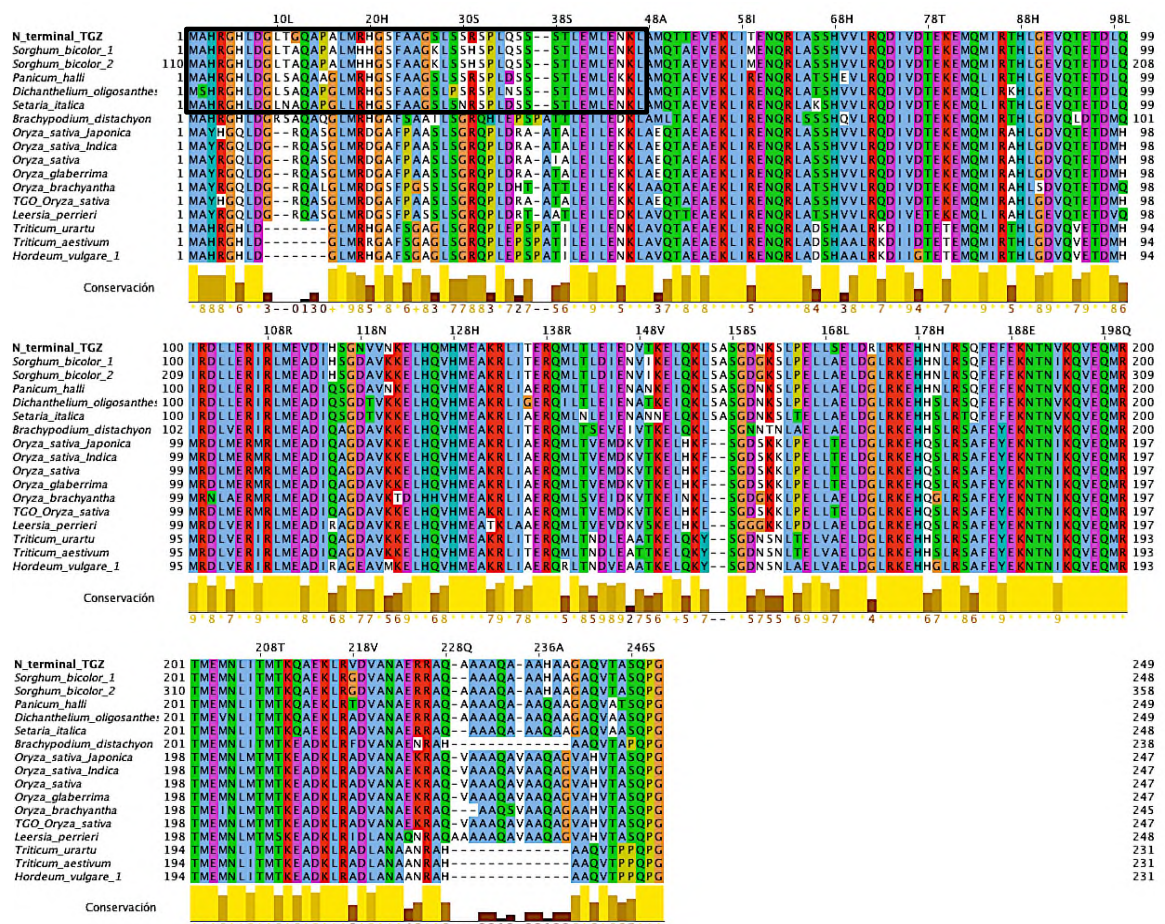
Por medio de la herramienta BLASTp se realizó una búsqueda de similitud en las bases de datos nr, SwissProt y TrEMBL donde se tomó en cuenta la observación de los alineamientos locales y los valores esperados (e-value) con un umbral de 0.001 para obtener la identificación de los homólogos más cercanos a las proteínas de estudio.

#### **7.3.1 Alineamiento de la región N-terminal con proteínas similares a la TGZ15 y TGZi**

La proteína TGZ15 y su isoforma inactiva TGZi comparten rasgos comunes en el extremo amino de su estructura primaria. Determinar si estas características son conservadas en otras especies ayudaría a esclarecer la funcionalidad de proteínas homólogas a ellas. El primer paso fue construir una secuencia consenso (N\_terminal\_TGZ) obtenida de la región N-terminal de las proteínas TGZ15 y TGZi, que va de la posición 1-249. Esta secuencia consenso sirvió como referencia para realizar una búsqueda de similitud con otras proteínas vegetales.

En la Figura 15, se observa el alineamiento de 16 proteínas predichas de diferentes organismos dentro de la familia *Poaceae*. La región N-terminal y adyacente a ella presenta una conservación entre las proteínas alineadas con un grado de similitud del 75-90%. La zona del péptido tránsito (ver Figura 15, remarcada en un cuadro negro) dentro del extremo amino es conservado solo en

4 especies diferentes (*Panicum halli*, *Dichanthelium oligosanthes*, *Setaria italica* y *Sorghum bicolor*). El núcleo del péptido tránsito contiene un largo segmento de aminoácidos hidrofóbicos (marcados en azul) que corresponden al 36% y otro 14% de aminoácidos con carga positiva (marcados en rojo), cantidades porcentuales respecto al total de aminoácidos encontrados en el péptido. Estos rasgos en la composición de aminoácidos son característicos de un péptido de señalización y que han sido ampliamente descritos por Bhushan *et al.*, (2006) y Käll *et al.* (2004). El segmento de esta región termina con una Leu (L) conservada en todas las secuencias de la posición 47 del alineamiento. También existen residuos de Arg (R) conservados y ubicados en las posiciones 3, 19 y 31, además de una Lys (K) en la posición 46, estos son aminoácidos cargados positivamente lo que puede ayudar a reforzar la topología adecuada del polipéptido durante la translocación (Von Heune & Gavel 1988; Fujita *et al.*, 2011). Al final del péptido tránsito se reconoce un corte por una señal peptidasa que se denomina sitio de escisión, un sitio de corte en la posición 35 de las secuencias.



**Figura 15.** Alineamiento de la región N-terminal a partir de una secuencia consenso. La secuencia N-terminal\_TGZ se tomó como referencia en el alineamiento para mostrar la conservación de caracteres entre el N-terminal de la TG de maíz y proteínas TG putativas de especies vegetales de la familia *Poaceae*.

El resto de las proteínas homólogas no presentaron estas propiedades características del péptido tránsito, entre las que destaca la proteína TGO de *Oryza sativa*. Este hecho es de gran relevancia biológica ya que pueden no estar localizadas en cloroplastos y por lo tanto cumplir otras funciones distintas relacionadas a la fotosíntesis. Sin embargo, Campos *et al.*, (2013) logró comprobar experimentalmente que la proteína TGO se ubica en este organelo plastidial y por lo tanto su implicación en la fotosíntesis. Sería de gran utilidad estudiar la localización subcelular a nivel experimental del resto de homólogos y por ende determinar sus posibles implicaciones en procesos celulares.

### **7.3.2 Alineamiento de la región C-terminal con proteínas similares a la TGZ15**

La región C-terminal en la proteína TGZ15 y su isoforma TGZi difieren en la composición de sus aminoácidos. A partir de este hecho, se realizó una búsqueda de similitud con el C-terminal de ambas proteínas por separado. A pesar de la divergencia de esta zona entre ambos grupos de proteínas, comparten rasgos comunes, como lo son la presencia de TRs muy particulares. Estos fragmentos repetitivos son secuencias de aminoácidos que contienen repeticiones de aminoácidos simples o motivos de aminoácidos cortos con poca diversidad en su composición, por lo tanto, forman regiones de baja complejidad (Coletta *et al.*, 2010). Estos son extremadamente abundantes en el proteoma eucariota. En el ADN, los TRs se encuentran en regiones que codifican proteínas, estas regiones mutan mucho más rápido por inserciones o deleciones (Wilkins *et al.*, 2018).

Se construyó (ver Figura 16A) un primer alineamiento con las proteínas similares al C-terminal de la proteína TGZ15, donde se observaron TRs y la conservación parcial de los residuos involucrados en la tríada catalítica. La proteína no caracterizada en sorgo (*Sorghum bicolor*; XP\_021319478.1) presentó la mayor similitud (54%) respecto a la TGZ15, esta proteína posee una Cys439/Asp471 conservados, pero con la ausencia de un residuo His. En tanto, la proteína TGO de arroz (*Oryza*

*sativa*) contiene un núcleo catalítico en las posiciones: Cys254/His260/Asp295 y un presumible Trp252 (Campos *et al.*, 2013), que es anterior a las TRs de la TGZ15. La proteína no caracterizada en cebada (*Hordeum vulgare*; F2D7X6) contiene una Cys137/His164 con la ausencia del residuo Asp; resultaría interesante en un futuro comprobar su actividad biológica dada su corta longitud de aminoácidos (168 aa) que ofrecería ventajas biotecnológicas. Además de comprobarse su actividad, no sería la única proteína monomérica ya que también existe una TG de origen microbiano ampliamente estudiada, como lo es TGM de *Sreptovercillium sp.* con 331 aa y de la cual se proyecta una posible diada catalítica entre Cys-Asp (Kashiwagi *et al.*, 2002).

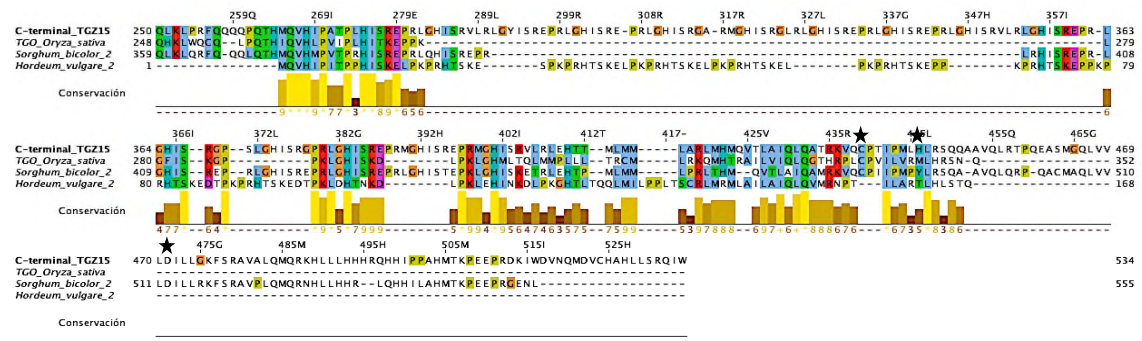
Es sugerido aislar y sobreexpresar como proteínas recombinantes aquellas proteínas no caracterizadas que no se ajustan a la tríada catalítica particular, para demostrar si poseen actividad enzimática tipo TG o no. Como lo demostrado por Hongbo Li *et al.*, (2013), donde realizó la expresión de una proteína recombinante de TGZ15 expresada en *Pichia pastoris*, optimizando la secuencia de codones del gen *tgz15* a partir de la eliminación de 14 de los 15 TRs, los cuales resultaron no ser necesarios para la actividad enzimática, el resultado mostró que estas regiones no afectan a los codones codificantes para la tríada catalítica.

A pesar de la divergencia entre estas proteínas similares al C-terminal de TGZ15, existen bloques conservados que se observan a lo largo del alineamiento, para obtener una identificación más confiable de estas regiones, se recurrió a utilizar la herramienta GLAM2 integrada en la plataforma MEME suite. El objetivo fue identificar regiones conservadas y recurrentes de longitud fija y con gaps que proyectaran una identidad de esta familia de proteínas respecto al otro grupo de proteínas homólogas, permitiendo indeles que reflejen los cambios evolutivos. Los parámetros utilizados fueron: un rango de columnas a alinear de 2-20, 10 réplicas de alineamientos y 2000 iteraciones realizadas

Los TRs de TGZ15 se encuentran ubicados a partir de la posición 275-410, con 15 repeticiones de 9 aminoácidos. En cuanto a la composición de estas regiones en las proteínas homólogas, la proteína no caracterizada de sorgo (*Sorghum bicolor*) presenta 6 TRs que van de la posición 274-330 con una composición de 9 aminoácidos, en cebada (*Hordeum vulgare*) se tienen 8 TRs con

una composición de 10 aminoácidos que van de la posición 13-96, y por último la TGO en arroz (*Oryza sativa*) presenta 2 TRs con una composición de 10 aminoácidos que van de la posición 276-300. Anteriormente Villalobos *et al.*, (2004) reportó la diferencia en cuanto al número de TRs en TGs y otras proteínas homólogas remotas; un caso ejemplar es el factor V de coagulación en vertebrados, cuyo número de TRs es significativamente distinto entre mamíferos y ausente en vertebrados inferiores como pueden ser aves y anfibios (Camire & Bos 2009).

**A. Alineamiento de proteínas homólogas al C-terminal de TGZ15.**



**B. Regiones recurrentes y semiconservadas.**



**Figura 16.** Alineamiento de la región C-terminal e identificación de regiones conservadas y recurrentes de las proteínas homólogas a TGZ15. **A.** Solo se identificaron 2 secuencias proteicas no caracterizadas y predichas de la familia *Poaceae*: *Sorghum bicolor* y *Hordeum vulgare* que presentan la conservación parcial de los residuos involucrados en la tríada catalítica. El fragmento alineado va de la posición 259-534. **B.** Por medio de la herramienta GLAM2 se identificaron 2 regiones biológicas encontradas a partir del alineamiento previo. La región 1 es la huella inicial conservada entre estas 4 proteínas alineadas, la región 2 es recurrente a lo largo de las proteínas y representa los TRs que forman parte de este grupo de proteínas.

La homología existente entre la familia de proteínas vegetales similares a TGZ15 y las proteínas animales se encuentran relacionadas en torno a la región 2 (ver Figura 16B) encontrada en el C-terminal. Estas regiones presentan una similitud baja del 25-30% con secuencias proteicas con evidencia a nivel proteína, respecto a proteínas como el factor V de coagulación sanguínea en cerdo (*Sus scrofa*; Q9GLP1, Grimm *et al.*, 2001) y hombre (*Homo sapiens*; P12259, Cripe *et al.*, 1992). A partir de esta observación se obtuvo un logo consenso de las regiones similares entre las

proteínas (ver Figura 17A), dicha región cumple con la siguiente expresión regular del motivo: H-[IT]-[TS]-[LTR]-[SGC]-P-[DTS]-[LS]-[SG]-H, la composición de los aminoácidos presentes son comparables con los TRs de 9 aminoácidos encontrados en TGZ15 (Ver Figura 17B).



**Figura 17.** Identificación de logos consenso de TRs en el C-terminal de proteínas similares a TGZ15. A partir del programa WebLogo creator se construyó un logo consenso de las regiones similares que contienen los TRs encontrados en proteínas homólogas animales similares al grupo de proteínas vegetales y los TRs encontrados en TGZ15.

Para obtener una confirmación de la homología existente entre los TRs del factor V de coagulación de animales y los TRs de las proteínas vegetales, se implementó una búsqueda exhaustiva de homólogos remotos en otras especies, en base a los patrones encontrados a partir de su expresión regular del motivo con la ayuda de la herramienta ScanProsite Tool (De Castro *et al.*, 2006). Se tomó como consulta la base de datos UniProtKB, donde se incluyeron las secuencias de proteínas a partir de TrEMBL y Swiss-Prot, así como variantes de splicing.

Las regiones repetitivas o TRs (ver Figura 17A) encontradas en otras proteínas confirmaron una similitud (ver Anexo 11.11) con el factor V de coagulación sanguínea en otros organismos animales como el de gorila (*Gorilla gorilla*, G3R571), chimpancé (*Pan troglodytes*, H2QZV8) y mono verde (*Chlorocebus sabaues*, A0A0D9RFF4). También se encontró una similitud con la proteína Cytochrome O ubiquinol oxidase de la bacteria *Bradyrhizobium sp.* (A0A1Q5SAS4 y A0A1Q5SJM9), además de una variedad de secuencias provenientes de proteínas no caracterizadas en organismos animales. Para los TRs propios de la TGZ15 (ver Figura 17B) que cumplen con la expresión regular: [HY]-I-S-R-[EGV]-[PLA]-[RS]-[LM]-[GE], se encontró similitud recurrente con secuencias descritas como subunidades pequeñas del ARN ribosomal metiltransferasa provenientes de bacterias (ver Anexo 11.12).

### 7.3.3 Alineamiento de la región C-terminal con proteínas similares a la TGZi

En el segundo alineamiento (ver Figura 18A) la región del C-terminal de la proteína TGZi comparte un grado de similitud del 35-70% con otras 14 proteínas predichas de la familia *Poaceae*, que van de la posición 249-421. Esta región es rica en residuos de carácter polar (N, Q, S y T; color verde) organizados en regiones altamente repetitivas. Se observó además un incremento significativo de ciertos residuos entre las proteínas TGZ15 y TGZi, la mayoría presentes en el C-terminal. De acuerdo con la composición de aminoácidos en la proteína completa, se observó el incremento significativo de 3 residuos que determinan la identidad del C-terminal en la isoforma TGZi (Ver Tabla 8) respecto a la proteína TGZ15.

**Tabla 8.** Porcentaje de residuos encontrados en el C-terminal respecto al total de aminoácidos de la proteína completa de TGZ15 y TGZi.

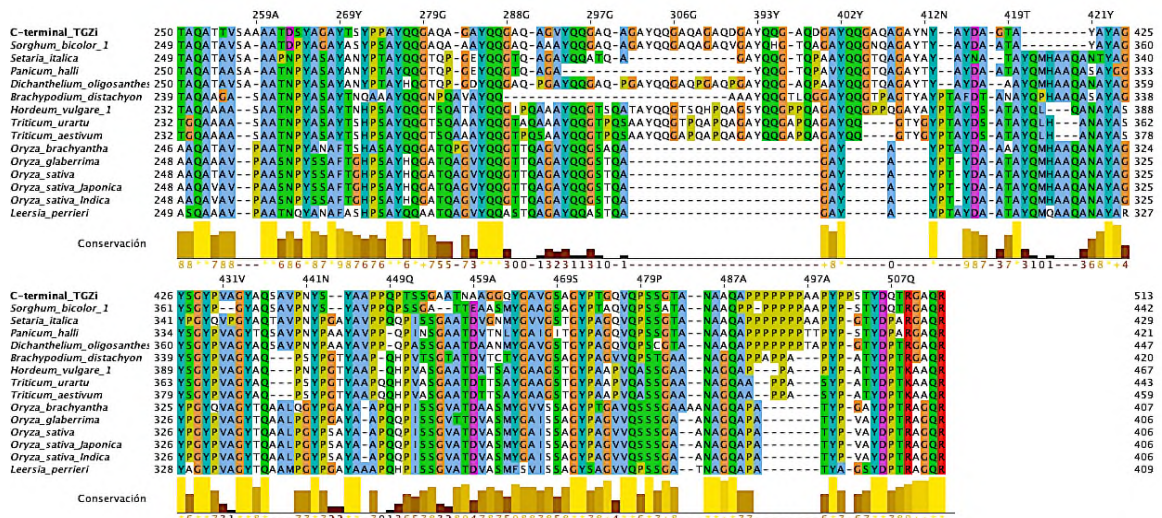
Aminoácidos	TGZ15 (% residuos)	TGZi (% residuos)
Alanina (A)	12/36= 33.3%	71/95= 74.7%
Glutamina (Q)	21/40= 52.5%	54/73= 73.9%
Glicina (G)	20/30= 66.6%	45/55= 81.8 %

La isoforma TGZi contiene regiones repetitivas ubicadas en el C-terminal rica en residuos de Gln y Ala. Los estudios acerca de proteínas con regiones ricas en glutamina y alanina (también conocidas como PolyQ y PolyA) son descritas en la participación de un gran número de pacientes con enfermedades genéticas como Huntington y una variedad de ataxias; estas proteínas son codificadas por trinucleótidos consecutivos y repetitivos del tipo CAG. Las proteínas tipo PolyQ son estructuralmente inestables y están asociadas a una composición bias que es conservado entre los proteomas eucariotas; y pueden ser encontrados en una variedad de familias proteicas, las cuales no parecen estar relacionadas (Ramazzotti *et al.*, 2012). En plantas, hasta el momento, las proteínas tipo PolyQ y PolyA se han relacionado a segmentos estrechos que son enriquecidos en factores de transcripción (Kottenhagen *et al.*, 2012).

El C-terminal de las proteínas homólogas a TGZi se caracteriza por una mayor conservación de los residuos que lo componen, a diferencia de la proteína TGZ15. No obstante, no dejan de tener una composición de baja complejidad al poseer regiones repetitivas y de poca variación de

aminoácidos. Esta región del alineamiento con respecto a todas las secuencias, fueron sometidas al servidor GLAM2 con el objetivo de identificar la región recurrente y mayormente representativa que es propia de esta familia de proteínas. Para esto se utilizaron los mismos parámetros anteriores, se eligió un rango de columnas a alinear de 2-20, 10 réplicas de alineamientos y 2000 iteraciones realizadas. Se obtuvo un logo consenso de una región rica de Gln y Ala (ver Figura 18B) compuesto por 20 residuos donde se puede observar la recurrencia de los TRs en TGZi.

**A. Alineamiento de proteínas homólogas al C-terminal de TGZi.**



**B. Región recurrente y conservada.**



**Figura 18.** Alineamiento de la región C-terminal e identificación de regiones conservadas y recurrentes de las proteínas homólogas a TGZi. **A.** Se alinearon 14 secuencias de proteínas predichas de la familia *Poaceae* en plantas. **B.** La poca diversidad en la composición de aminoácidos crea zonas altamente repetitivas o también llamados regiones de baja complejidad, se hace hincapié de la secuencia consenso AYQQGGAQAG presentes en la posición del alineamiento 275-411 propia de la proteína TGZi.

La región recurrente encontrada en el C-terminal de proteínas homólogas a TGZi proyectaron un 40% de similitud con proteínas ricas en Glutamina de organismos animales como el toro (*Bos taurus*, F1MJ07), caballo (*Equus caballus*, F7CG37) y oveja (*Ovis aries*, W5PCK5); a partir de esta

relación de similitud se construyó un logo consenso que representó los residuos similares entre estas proteínas obteniendo una región de 8 residuos conservados en ambos grupos proteicos, el consenso de estos bloques (ver Figura 19A) está representado por la siguiente expresión regular: Q-A-G-[AVL]-[VYA]-Q-[PQ]-G.

A. Logo consenso de TRs en otras proteínas.



B. Logo consenso de TRs en la proteína TGZi.



**Figura 19.** Identificación de logos consenso de TRs en el C-terminal de proteínas similares a TGZi.

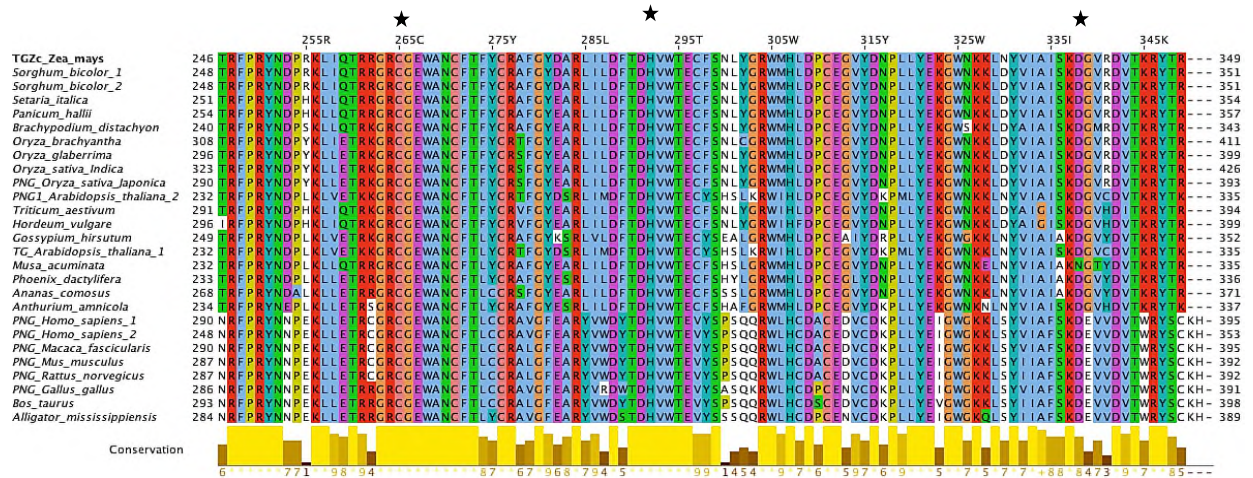
A partir del programa WebLogo creator se construyó un logo consenso de las regiones similares que contienen los TRs encontrados en proteínas homólogas animales similares al grupo de proteínas vegetales y los TRs encontrados en TGZi.

Utilizando la misma metodología anterior, se recurrió al servidor en línea de ScanProsite Tool (De Castro *et al.*, 2006), donde a partir de la expresión regular de la región consultada se implementó una segunda búsqueda más exhaustiva de homólogos remotos en otras especies, en base a patrones que pueden encontrarse en las secuencias proteicas encontradas en las bases de datos TrEMBL y Swiss-Prot, así como variantes de splicing. La región consenso de TRs encontradas en otras proteínas (ver Figura 19A) presentó una similitud con secuencias de proteínas muy variada que van desde proteínas nombradas como “ricas en Glutamina”, un bloque similar al factor VIII de coagulación sanguínea del ave *Ficedula albicollis* (U3KAV8), aciltransferasas de la bacteria *Alcanivorax sp.* (IDs: A0A2E4T5S3, A0A358F6G5, A0A348WXN2, A0A359KYI2, A0A356Q7G9, A0A356PLP3, A0A357JN83, A0A2D5RGH5), y entre otras proteínas no caracterizadas (ver Anexo 11.13). Sin embargo, la expresión regular de los TRs propios de la proteína TGZi (ver Figura 19B), [AV]-Y-Q-Q-G-A-Q-[ADS]-G, son exclusivos del grupo de proteínas vegetales TGs putativas y homólogas al C-terminal de la proteína TGZi.

#### 7.3.4 Alineamiento de proteínas similares a TGZc

En el caso de la proteína TGZc se encontraron homólogos en 27 especies animales y vegetales. Se destacan 7 estructuras primarias ampliamente revisadas en la UniProtKB. Este grupo de proteínas se anotan funcionalmente como péptido N-glicanasas; entre éstas se encuentran las PNG de

*Arabidopsis thaliana* (Della Mea *et al.*, 2004a), arroz (*Oryza sativa subsp. Japonica*), humano (*Homo sapiens*), macaco (*Macaca fascicularis*), ratón (*Mus musculus*), rata (*Rattus norvegicus*) y gallo (*Gallus gallus*). El alineamiento entre estas proteínas homólogas muestra una región altamente conservada que coincide con el dominio TGc (SM00460) con un grado de similitud del 87-100% en plantas y del 72-78% en animales respecto a la TGZc. Estas proteínas homólogas a TGZc comparten la tríada catalítica propia de las TGs (ver Figura 20, resaltando los residuos activos de la tríada catalítica con forma de estrellas). La relación entre la tríada catalítica conservada tanto en TGs como PNGasas sigue el tipo de reacción que desempeñan; en levaduras se realizaron experimentos a partir de mutagénesis dirigida, donde se confirmó que los residuos de la tríada catalítica de las PNGasas son esenciales para su actividad de desglucosilación (Katiyar *et al.*, 2002).

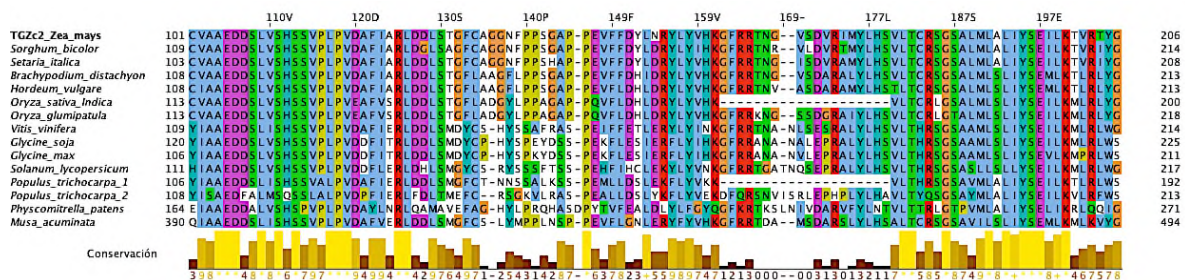


**Figura 20.** Alineamiento de la proteína TGZc. Se alinearon 27 secuencias de proteínas que pertenecen a organismos vegetales y animales. El fragmento alineado corresponde al dominio TGc a partir de las posiciones 246-349, tomando como secuencia referente a TGZc. La tríada catalítica conservada se ubica de la siguiente manera: Cys265, His292 y Asp309, además de un hipotético Trp268 como residuo estabilizador.

### 7.3.5 Alineamiento de proteínas similares a TGZc2

El alineamiento de la proteína TGZc2 (ver Figura 21) mostró un grado de similitud del 60-94% con respecto a 14 secuencias homólogas que se encuentran distribuidas únicamente en organismos vegetales. La característica de este alineamiento se basa en proteínas con la conservación de una región denominada Transglut\_core2 (PF13369).

No obstante, la existencia de estas proteínas se limita a una evidencia del tipo predichas, es decir a partir de la traducción automática por ordenador de secuencias codificantes desde ADN; la TGZc2 es la única proteína vegetal de su clase que tiene un criterio de evidencia experimental a nivel transcripto. La escasa información registrada en las bases de datos que considera a TGZc2 con el dominio Transglut\_core2, la describe como miembro del clan Peptidasa\_CA (CL0125). Sin embargo, debe ser tomado con precaución puesto que solo se ajusta parcialmente con el dominio mencionado y no presenta similitud con otras proteínas que contengan el dominio mencionado.



**Figura 21.** Alineamiento de la proteína TGZc2. Se alinearon 14 proteínas de origen vegetal, tomando como referencia a la proteína TGZc2. El fragmento alineado forma parte del dominio Transglut\_core2 (PF13369) que se extendió desde la posición 101-205 con referencia a la TGZc2.

### 7.3.6 Análisis del núcleo catalítico de la proteína TGZ15 con otras proteínas TGs

A partir de la búsqueda de similitud en las bases de datos y la construcción de los alineamientos múltiples, se observó que la proteína TGZ15 (Q6KF70) no presentó similitud en la configuración de su sitio activo con respecto a otras TGs. La tríada catalítica de la TGZ15 se organiza de una manera diferente localizándose en el C-terminal. Debido a este problema, se realizaron alineamientos múltiples delimitando la zona que comprende el núcleo activo de la TG. A partir del alineamiento, se analizó la naturaleza de los residuos encontrados en el núcleo catalítico conservado entre las diferentes especies.

Para realizar los alineamientos de esta zona, se descargó desde la base de datos SMART protein los fragmentos de proteínas con dominio TGc (SM00460) de acuerdo al siguiente criterio taxonómico: mammals, fungi, viridiplantae y bacterial. Las secuencias integradas fueron revisadas con las bases de datos UniProtKB. El alineamiento se construyó con el software MAFFT implementando un método iterativo refinado muy lento L-INS-i y una matriz BLOSUM45, el cual

alinea un conjunto de fragmentos que contienen secuencias flanqueantes alrededor de un dominio alineable, las secuencias de flanqueo se ignoran en la alineación por parejas mediante el algoritmo Smith-Waterman. En la Figura 22 se muestra la conservación de sitios funcionales específicos propios de la superfamilia de enzimas tipo TG que se centran alrededor de los residuos activos de la tríada catalítica (Cys/His/Asp) previamente estudiada por Marakova *et al.*, (1999). Sin embargo, en este alineamiento se agregan dos nuevos grupos, las TGs viridiplantae y fungi, de las cuales se destacan las secuencias AtPNG1p de *Arabidopsis thaliana*, TGO de *Oryza sativa* y TGZ15 de *Zea mays*, con el objetivo de encontrar regiones más similares entre el grupo vegetal.

La principal problemática en debate es que la proteína TGZ15 no se incluye o no se ajusta al dominio canónico de la TG como lo describen Beninati *et al.*, (2013) y Lorand & Graham (2003). En este alineamiento se observa una nueva ubicación putativa del nicho catalítico que involucra a las proteínas TGZ15 y TGO, con una baja pero considerable similitud en regiones específicas de otras secuencias TGs. Villalobos *et al.*, (2004) propuso una ubicación de la tríada catalítica (Cys439/His446/Asp471) basada en la homología encontrada con una secuencia de ADNg de arroz (*Oryza sativa* subps. Japonica; AL606595), con la que compartió el 87% de identidad y una tríada catalítica perfectamente ajustada. Sin embargo, la similitud encontrada fue a partir de un alineamiento local desde la deducción de aminoácidos con fuente en ADNg, por lo que no se tomó en cuenta las expresiones diferenciales que pudo llegar a tener el producto codificante, en las que codifica a una proteína de 406 aminoácidos (Q7XK24). Las posiciones de la tríada catalítica en el alineamiento de la TGZ15 ubican una conservación de la Cys439, pero con una reorganización de la His498 y Asp522 que ajustan perfectamente con los homólogos en mamíferos y bacterias (ver Figura 22). Campos *et al.*, (2013) analizó la proteína TGO (M1VT12), donde consideró una tríada catalítica putativa en las posiciones Cys254/His260/Asp295. No obstante, en el alineamiento realizado se observa que únicamente la Cys254 es conservada con las demás especies, con una marcada diferencia de los residuos posteriores, con una His301 conservada y la ausencia del Asp.

La identificación de 3 motivos que comprenden los residuos catalíticos en la superfamilia de enzimas TGs (Marakova *et al.*, 1999) se encuentran posicionados de la siguiente manera: El

motivo 1 representa la región que contiene a la Cys catalítica y es distinguible por la presencia de residuos pequeños conservados río arriba de la Cys, uno de estos es la Gly localizada en todas las especies a excepción de las proteínas TGO en *Oryza sativa* y TGZ15 de *Zea mays*, además de la proteína CYK3 de *Saccharomyces cerevisiae*. Este mismo evento existe en las proteínas tipo papaína y el FXIIIa como lo describe Lorand & Graham (2003). Una característica particular del grupo viridiplantae es la conservación de residuos polares (K o R) con carga positiva 4 posiciones arriba de la Cys y que es comparable con la parcial conservación observada en las demás especies, como el Factor XIIIa y la TGM2 (Lorand & Graham 2003). La región de transición entre el motivo 1 y el motivo 2 que contiene a la His activa, es una zona rica en residuos hidrofóbicos comparables en todas las especies alineadas. El motivo 2 se caracteriza por la presencia de residuos aromáticos río abajo de la His, mediado por la transición de regiones ricas en residuos hidrofóbicos (Lorand & Graham, 2003). Finalmente, el motivo 3 con el Asp activo se encuentra en una posición acompañado por un Trp conservado río arriba en todas las especies a excepción de la TGO en *Oryza sativa* (ver Figura 22; Lorand & Graham 2003).

La composición de los aminoácidos dentro del sitio activo de las enzimas son los responsables directos de la actividad catalítica. Una de las propiedades catalíticas dentro del sitio activo de las TGs es la exclusión de agua alrededor de un nicho hidrofóbico. La exclusión del agua del sitio catalítico implica que tanto el sustrato donador que contiene Gln como el nucleófilo (conocido como amina o sustrato aceptor) deban acceder a la Cys a través de un canal rodeado de residuos no polares mediante unos conductos con propiedades específicas. Por el contrario, la probabilidad de que aparezcan moléculas de agua en los nichos hidrofóbicos de la enzima provocaría consecuencias en el intermediario acil-enzima dada por una inestabilidad catalítica (Folk & Finlayson 1977; Nemes *et al.*, 2005).

Como se observa en el alineamiento de la Figura 22, la preferencia de ciertos residuos se localiza más frecuentemente distribuidos a lo largo del núcleo catalítico. Por lo tanto, se analizó la distribución de los aminoácidos en el sitio activo en comparación con la proteína completa (ver Anexo 11.14).



**Figura 22.** Alineamiento del núcleo conservado de la TG. Los fragmentos de secuencias con el dominio TGc de cada proteína fueron agrupados de acuerdo al grupo taxonómico de la siguiente manera: 7 secuencias de bacterias, 16 secuencias de mamíferos con referencia en el Factor XIIIa y Band 4.2 de humano, 9 secuencias viridiplantae y 5 secuencias de origen fungi. La selección de estas secuencias se basa en la anotación del criterio de evidencia y la revisión que posee por parte de la UniProtKB. Los bloques conservados entre las distintas especies afines al núcleo catalítico muestran las zonas más similares entre las secuencias, cada número marca la posición donde inicia cada fragmento de secuencia. En la parte superior del alineamiento se muestra con una estrella los residuos activos conservados de la tríada catalítica, estos presentan un grado de conservación con la Cys (C; marcado con color rosa) del 89%, His (H; marcado con color cyan) del 94% y el Asp (D; marcado con color magenta) del 92%. Los residuos hidrofóbicos son marcados en color azul (A, I, L, M, F, W, V) si presentan un grado de conservación mayor del 60%. Los residuos de Gly (G) de naturaleza pequeña son marcados en color anaranjado independientemente de su conservación. Por otra parte, la parcial conservación de residuos con carga positiva (K, y R) son marcados en color rojo si presentan una conservación mayor del 60%.

Dentro del sitio activo se observó una alta recurrencia de cadenas laterales hidrofóbicas en TGs bacterianas (A, M y V). En el caso de mamíferos, los residuos hidrofóbicos con mayor presencia son las Ala (A), no obstante, residuos como Val (V) y Leu (L) tienen una presencia casi proporcional entre la proteína completa y el sitio activo. En las TGs viridiplantae y fungi disminuye la variabilidad de residuos hidrofóbicos, pero no pierden su propiedad hidrofóbica al contar con una alta frecuencia de Ile (I) y Val (V). En el caso especial de las proteínas vegetales TGZ15 maíz y TGO de arroz, existe una recurrencia de aminoácidos hidrofóbicos como Leu (L), Val (V), Ala (A) y Met (M) que se contrastan con la presencia de regiones ricas en Gln (Q), His (H), Arg (K) y Asp (D). Villalobos *et al.*, (2004) resaltó la importancia de una zona de transición entre una región hidrofóbica y una región hidrofílica ya cerca del C-terminal. La proteína TGO concentra una alta composición de Ile en el sitio activo en comparación de la proteína completa, pero que se contrastan igualmente con la abundancia de Gln (Q), His (H) y Arg (K).

Todas las TGs poseen un rasgo muy similar, y es que presentan grupos aromáticos (Y, W y F) con mayor frecuencia en su núcleo catalítico que en la proteína completa. Particularmente se puede observar una composición frecuente de Gly en los grupos de bacterias, mamíferos y viridiplantae. En casi todas las secuencias alineadas, la His es más común en el sitio activo que en la proteína completa, a excepción del grupo bacteriano. Las cisteínas por otra parte son encontradas en menor número en bacterias (2 residuos), TGZ15 (2 residuos) y TGO (3 residuos).

#### **7.4 Análisis filogenético de las proteínas TGs de maíz**

La reconstrucción filogenética involucró la generación de 4 árboles filogenéticos a partir de la construcción de los alineamientos realizados anteriormente que representan a:

- La familia de proteínas vegetales con referencia a TGZ15 y TGZi.
- La familia de proteínas vegetales y animales PNG con referencia a TGZc que conservan el núcleo catalítico de la TG (TGc).
- La familia de proteínas vegetales similares a TGZc2.
- La familia de TGs que conservan el núcleo catalítico TGc (SM00460) en bacterias, fungi, mamíferos y plantas.

Cada alineamiento (en formato fasta) fue ingresado por separado en el programa IQ-TREE versión 1.6.6 (Kalyaanamoorthy *et al.*, 2017). Se implementó la matriz de sustitución JTT como modelo ideal a utilizar con los alineamientos para la reconstrucción del árbol filogenético basado en el principio de MV, se obtuvo la información de las distancias evolutivas en formato newick, posteriormente los parámetros del modelo utilizados fueron implementados en phylogeny.fr donde el resultado fue visualizado y editado en el programa FigTree.

##### **7.4.1 Árbol filogenético de la familia de proteínas similares a TGZ15 y TGZi**

TGZ15 y TGZi (B8A2N4 y COHGH1) son proteínas que comparten un grado de similitud elevada con el 53.7% a nivel global, las cuales poseen una región N-terminal altamente conservada y un extremo carboxilo muy divergente con proteínas putativas TG no caracterizadas pertenecientes a la familia *Poaceae*. A partir de estos eventos, se construyó un árbol filogenético con el alineamiento de 20 secuencias completas, con lo cual, se consideraron sitios no homólogos en todas las proteínas; el objetivo fue que las zonas N-terminal y C-terminal más similares a TGZ15 y

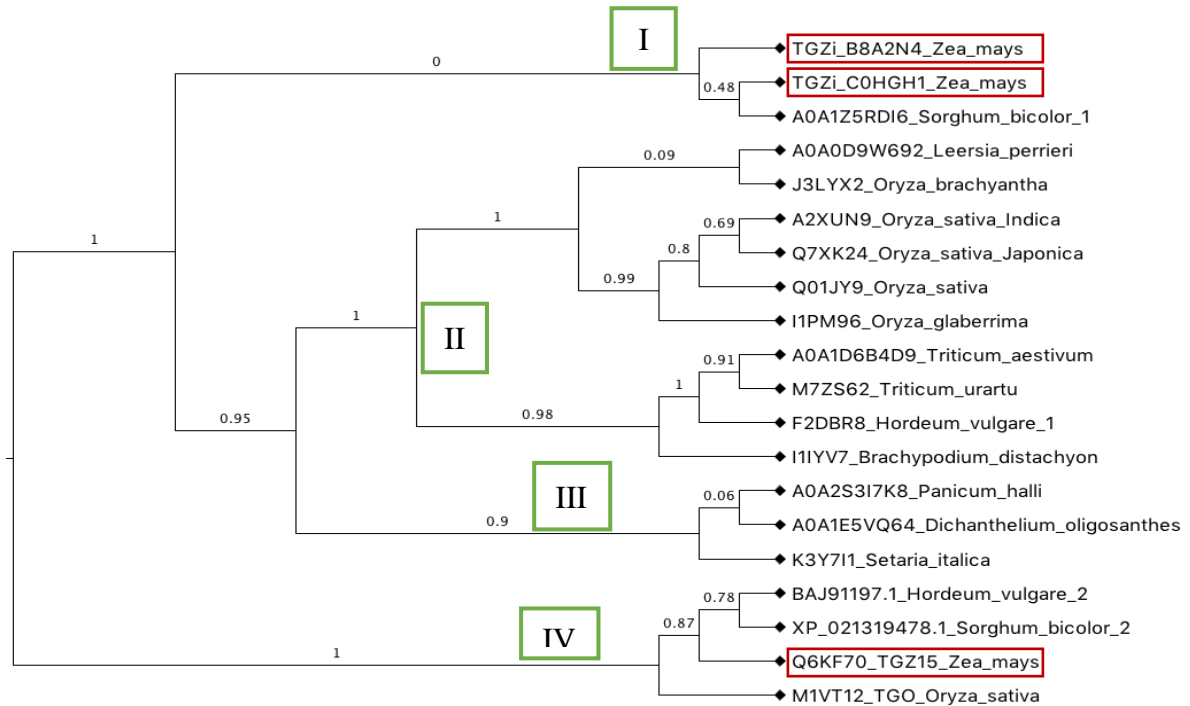
TGZi homólogas a otros subgrupos de proteínas vegetales sirvieran como referencia para el establecimiento de clados proteicos con relación filogenética.

En la Figura 23 se observa que las proteínas TGZ15 y TGZi de maíz se separan en diferentes nodos agrupándose en subgrupos de proteínas de diversas especies monocotiledóneas. La relación evolutiva demuestra que proteínas homólogas a TGZi comparten una estrecha relación evolutiva caracterizada por eventos de duplicación que difieren de la función original, por lo cual se organizan en tres grupos proteicos (I, II y III), conservando un C-terminal con regiones ricas en Gln, Ala y Gly. La aparición de estos genes duplicados dentro del genoma de las gramíneas les confiere nuevas funcionalidades que les brindan adaptación e innovación biológica para cada organismo.

La nueva funcionalidad de cada proteína que difiere de la función TG original, se basa en que las duplicaciones de regiones génicas pequeñas establecen más interacciones genéticas siendo funcionalmente más divergentes. Por lo tanto, cada uno de los genes duplicados codificantes podrá adquirir nuevas funciones con mayor facilidad. Esta abundancia de genes duplicados es una particularidad de las especies *Poaceae* angiospermas, que, en comparación con otros grupos taxonómicos secuenciados, son uno de los grupos más diversos tanto evolutivamente como fenotípicamente dentro de organismos eucariotas (Zhang 2003; Fligel & Wendel 2009). Utilizando genómica comparativa, investigadores han propuesto la hipótesis de que esta extraordinaria diversidad es principalmente un producto de eventos de duplicación génica (Zhang 2003; Fligel & Wendel 2009; Van de Peer et al., 2009; Jiang, & Assis 2019).

El grupo IV sobresale de un brazo que diverge del resto de proteínas agrupadas al nodo principal, este grupo está representado por dos proteínas caracterizadas, TGO de arroz (Campos *et al.*, 2013) y TGZ15 de maíz (Villalobos *et al.*, 2004), ambas codificadas por genes de copia única, con un 70% de similitud y con evidencia proteica a nivel transcripto. A las que se les une dos proteínas no caracterizadas, provenientes de sorgo (*Sorghum bicolor*) y cebada (*Hordeum vulgare*). En el caso más simple, se puede considerar que tanto TGZ15 como TGO representan las copias ancestrales de un gen de copia única con una función enzimática. La conservación de su función ancestral unido a la adquisición de una nueva función a través de sus copias duplicadas solo puede

ser explicado mediante un mecanismo de neofuncionalización. La hipótesis de la neofuncionalización ocurre como resultado de una selección positiva que actúa sobre mutaciones beneficiosas que surgen en una copia (Ohno 2013). Otra explicación de esta divergencia funcional es explicada por Todd *et al.* (2002a), en la cual se sostiene que precursores catalíticos ancestrales evolucionan a enzimas inactivas ocasionando la pérdida de residuos catalíticos.



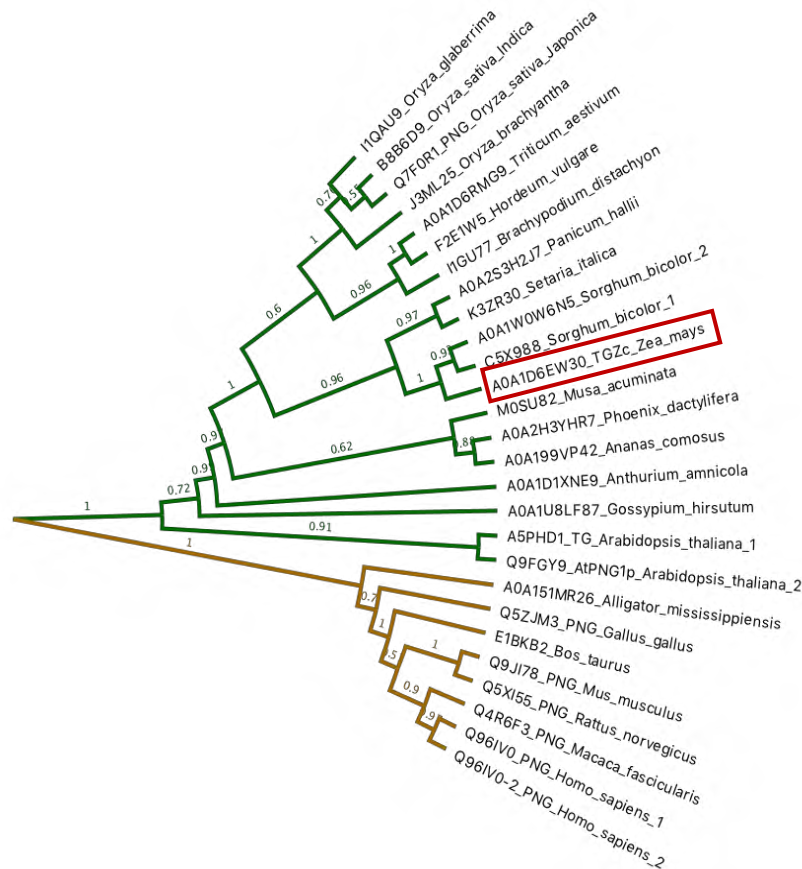
**Figura 23.** Árbol filogenético de las proteínas TGZ15 y TGZi en especies vegetales de la familia *Poaceae*. El árbol filogenético de las proteínas TGZ15 y TGZi se realizó con la herramienta computacional IQ-TREE y phylogeny.fr, el alineamiento realizado para este árbol se construyó en Clustal Omega. Fueron seleccionadas 20 proteínas completas, donde la mayor parte de ellas son proteínas no caracterizadas de diferentes especies; la secuencia F2D7X6 de *Hordeum vulgare* fue alineada y ajustada manualmente debido a su corta longitud con respecto a las demás secuencias.

#### 7.4.2 Árbol filogenético de proteínas similares a TGZc

El árbol filogenético de la proteína TGZc (ver Figura 24) está formado por dos clados que representan a proteínas anotadas funcionalmente como péptido N-glicanasas, estas conservan el núcleo catalítico de la TG tanto en animales como en plantas. En el primer clado (color verde) se relacionaron 19 proteínas ortólogas de origen vegetal, en las cuales se encuentra una estrecha

relación entre las proteínas ampliamente revisadas y con origen en un transcrito, una PNG putativa de arroz (*Oryza sativa*) y la AtPNG1p de *Arabidopsis thaliana*.

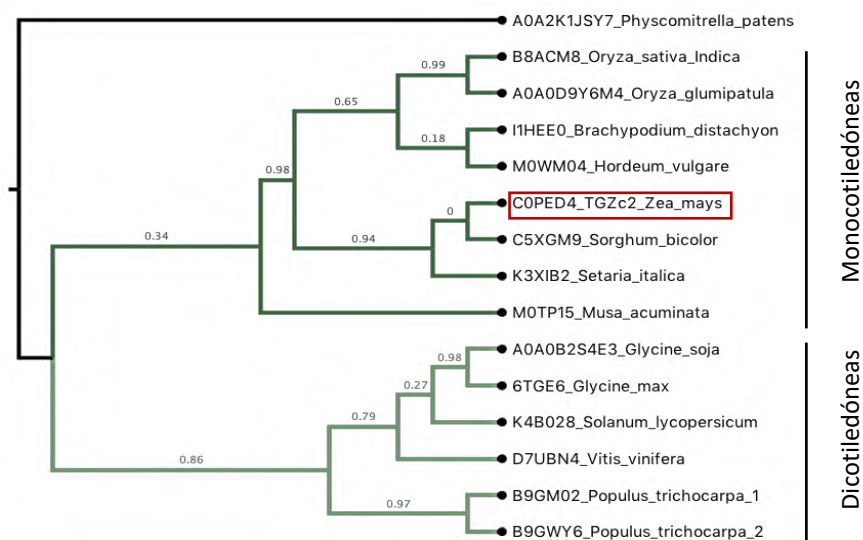
En el segundo clado (color café) se ubican 8 proteínas de origen animal anotadas como PNG, con una amplia revisión a nivel proteico, tomando como referencia a las secuencias Q96IV0 de humano, Q4R6F3 del macaco, Q5XI55 de rata, Q9JI78 de ratón, Q5ZJM3 de gallo. Por lo tanto, se presume que la TGZc es una enzima PNG agrupándose con otras proteínas predichas no caracterizadas, como *Sorghum bicolor*, *Setaria italica* y *Panicum halli*. En el segundo clado formado por PNG animales la distribución de las especies en el cladograma muestra a la TGZc como una proteína evolutivamente más alejada a las 9 proteínas PNG animales, evento que se contrasta con la AtPNG1p que mantiene una mayor similitud evolutiva con las proteínas animales. Las PNG animales son ampliamente revisadas y poseen un criterio de existencia a nivel transcrito y proteico. Diepold *et al.*, (2007) menciona que las PNG tienen una estrecha relación evolutiva con las TGs, por la similitud del núcleo catalítico, aumentando las posibilidades de una multifuncionalidad como lo explica Serafini-Fracassini *et al.*, (2009).



**Figura 24.** Árbol filogenético de la proteína TGZc. El árbol filogenético de la proteína TGZc se realizó con la herramienta computacional IQ-TREE y phylogeny.fr, el alineamiento realizado para este árbol se construyó en Clustal Omega. Fueron alineadas 27 proteínas completas que comparten el núcleo catalítico de la TG (SM00460). En color verde se representa a las proteínas de origen vegetal y en amarillo las de origen animal.

### 7.4.3 Árbol filogenético de proteínas similares a TGZc2

Las proteínas alineadas con el dominio Transglut\_core2 homólogas a TGZc2 se distribuyen en el árbol filogenético (ver Figura 25), entre las que sólo se encuentran secuencias proteicas de origen vegetal con un score de anotación pobre, sin revisión y con evidencia a nivel predicha. En el primer grupo se observan a proteínas vegetales monocotiledóneas de la familia Poaceae, de este mismo grupo sobresale un brazo que pertenece a una proteína no caracterizada de *Musa acuminata* de la familia vegetal *Musaceae*. El segundo grupo pertenece a proteínas vegetales dicotiledóneas, del mismo nodo principal sobresale un brazo distante que pertenece a una secuencia vegetal de musgo (*Physcomitrella patens*). No se encontraron mayores informes reportados acerca de la anotación de este dominio registrado. Por lo tanto, hace falta mayores estudios relevantes sobre este tipo de proteínas y sus homólogos para entender la funcionalidad biológica.



**Figura 25.** Árbol filogenético de la proteína TGZc2. El árbol filogenético de la proteína TGZc se realizó con la herramienta computacional IQ-TREE y phylogeny.fr, el alineamiento realizado para este árbol se construyó en Clustal omega. Fueron alineados 15 proteínas completas que comparten similitud con el dominio identificado.

#### 7.4.4 Árbol filogenético de la familia TG en diversas especies y su relación con la TGZ15

El núcleo catalítico de las TGs presenta una composición canónica ampliamente estudiada en mamíferos. El árbol filogenético (ver Figura 26) muestra la distribución de los fragmentos de proteínas delimitadas por el dominio catalítico de la TG de acuerdo al tipo de especie al que pertenecen, para esto se puede observar la formación de 2 grupos principales y un brazo independiente que sobresale y agrupa a las proteínas TGZ15 y TGO.

El nodo principal se divide en subnodos internos que se componen de especies viridiplantae (brazos en color verde), fungi (brazos en color amarillo) y bacterias (brazos en color azul). El primer grupo está compuesto por 9 proteínas vegetales (viridiplantae) muy cercanas evolutivamente, en su mayoría conteniendo proteínas tipo PNG, en las que se encuentran la AtPGN1 de *Arabidopsis thaliana*, PNG1 de *Oryza sativa* y la TGZc de maíz. Por otra parte, la combinación de grupos taxonómicos refleja una cercanía evolutiva evidente entre fungi y viridiplantae, donde se relaciona con otras 3 proteínas PNG de organismos fungi de *Schizosaccharomyces pombe*, *Neurospora crassa* y *Saccharomyces cerevisiae*. El segundo grupo de proteínas lo componen 6 TGs bacterianas y 2 TGs fungi (*Schizosaccharomyces pombe* y *Saccharomyces cerevisiae*) que se encuentran más estrechamente relacionadas entre ellas, estas proteínas bacterianas son del género *Mycobacterium*, *Bacillus*, *Haemophilus* y *Pseudomonas*. Otro grupo de proteínas TGs surge de un nodo que se mantiene alejado de las demás TGs, aquí se agrupan las TGs de mamíferos mayormente estudiadas y caracterizadas a nivel proteico (brazos en color café) con un dominio catalítico conservado.

La divergencia evolutiva de las proteínas TGZ15 y TGO (con brazos en color rojo) es remarcada porque sobresalen de un brazo del nodo principal separándolas del resto de proteínas TGs. Ambas proteínas contienen un dominio catalítico no conservado y con una ubicación diferente en la configuración de sus residuos catalíticos. Las tríadas catalíticas han mostrado los mejores ejemplos de evolución convergente; las restricciones químicas en la catálisis han llevado a que la misma solución catalítica evolucione independientemente en al menos 23 superfamilias separadas, cada una de estas superfamilias es el resultado de una evolución convergente con una



## 7.5 Predicción del modelo estructural de las proteínas TGs de maíz por la técnica de enhebrado y *ab initio*

A partir de la metodología descrita, se implementó Phyre2 para la predicción de un modelo proteico. Posteriormente se validó la calidad del modelo utilizando las herramientas que ofrece el servidor SAVES v5.0. La predicción se realizó en conjunto con el análisis de las estructuras secundarias de cada proteína de interés, para esto se utilizó el servidor en línea PSIPRED (McGuffin *et al.*, 2000).

El análisis de la estructura secundaria consistió en la identificación de las  $\beta$ -laminas,  $\alpha$ -hélices y coils existentes. Incluso se analizó la presencia de RIDs con ayuda del programa DISORPRED3 (Jones & Cozzetto 2014), donde se determinó la existencia de regiones desordenadas dentro de las proteínas y que dificultan la calidad de la predicción. Con los análisis anteriores se pudieron observar regiones de alta complejidad o con poca variación de aminoácidos (Coletta *et al.*, 2010) propios de los TRs encontrados en C-terminal de las proteínas TGZ15 y TGZi.

Algunas proteínas tienen regiones que son nativamente desordenadas, y generalmente son flexibles, dinámicas y pueden ser parcial o completamente extendidas en solución. El estado nativo de una proteína puede existir como estructuras globulares, random coils extendidos e incluso como glóbulos fundidos. Recientemente se ha establecido que RIDs en proteínas son también elementos cruciales para una funcionalidad proteica, olvidando el concepto tradicional de una estructura para una función (Latysheva *et al.*, 2015). Se ha revisado que proteínas con RIDs o completamente desordenadas se encuentran involucrados en procesos regulatorios importantes como la transcripción y la citoquinesis, así como en procesos de señalización que involucran proteínas quinasas y pequeñas GTPasas (Ward *et al.*, 2004).

Las regiones desordenadas son usualmente caracterizadas por la baja complejidad de secuencia, una composición bias de aminoácidos hacia residuos aromáticos e hidrofílicos con una alta flexibilidad (Ward *et al.*, 2004). Identificar estas regiones es vital para entender la funcionalidad de estas en las proteínas. En plantas, el desorden se encuentra involucrado en el control del ciclo celular, metabolismo, señalización hormonal y regulación de la expresión génica, así como el

desarrollo y respuesta al estrés. Es importante describir que los proteomas de *Zea mays* y *Glicine max* contienen un desorden significativamente alto en una proporción similar al 50% (Covarrubias *et al.*, 2017).

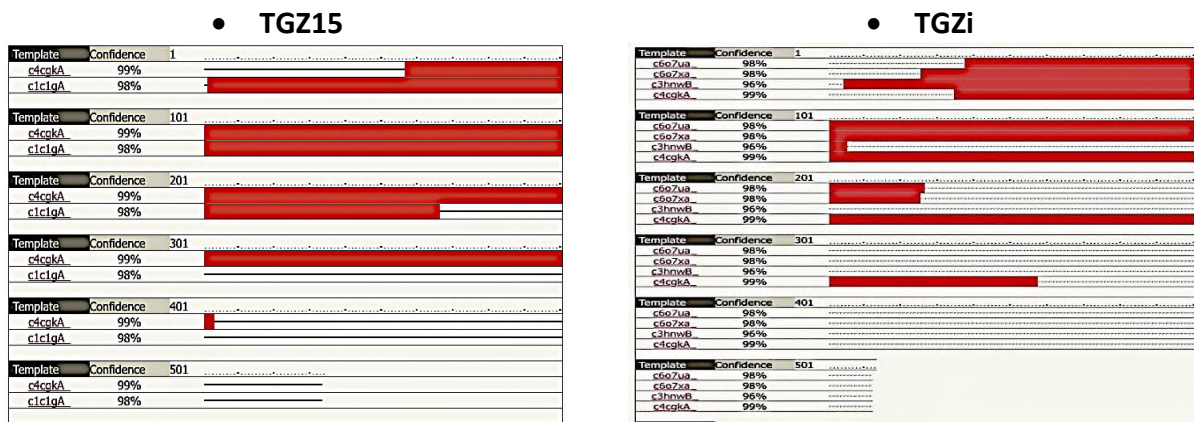
### **7.5.1 Modelo estructural de la proteína TGZ15 y su isoforma inactiva TGZi**

Con las secuencias de las proteínas TGZ15 (Q6KF70) y TGZi (B8A2N4) se obtuvieron las posibles conformaciones que adoptan sus estructuras secundarias dentro de una estructura tridimensional. La proteína TGZ15 fue resuelta en un 75% (400 aa) de los residuos modelados mediante enhebrado con el 90% de confiabilidad utilizando 2 secuencias templado (ver Figura 27), que solo permitieron una cobertura parcial de la proteína hasta la posición 403. Por otro lado, la proteína variante TGZi presentó el 68% (349 aa) de los residuos modelados con un grado de confiabilidad del 90% con el soporte de 4 secuencias templado (ver Figura 27) que igualmente permitieron solo una cobertura parcial de la proteína hasta la posición 357.

A pesar que ambas proteínas comparten un alto grado de identidad hasta la posición 249, Phyre2 ofreció una diferencia notoria en la utilidad de los templados seleccionados, ya que el único templado que es común entre ambas proteínas es la perteneciente a c4cgkA. Tanto en TGZ15 como en TGZi el resto de los aminoácidos que se encuentran mayormente en la zona de los TRs y el dominio catalítico adyacente al C-terminal no presentaron algún tipo de ajuste con las secuencias templado seleccionadas, así que fueron modelados por el método *ab initio*. Cabe resaltar que estas zonas ubicadas en el C-terminal predicen un alto grado de desorden, por lo que no pueden ser significativamente fiables al momento de construir el modelo, así que el modelamiento por *ab initio* debe ser tomado con precaución. Siendo que la isoforma TGZi es producto de splicing alternativo, la expresión del gen tiene más probabilidades de codificar regiones desordenadas, es decir, con múltiples configuraciones estables. Estas regiones desordenadas podrían experimentar una selección más fuerte para funciones relacionadas con la actividad de unión, estableciendo vínculos entre el splicing alternativo y la evolución de redes regulatorias en la expresión del gen (Romero *et al.*, 2006; Blencowe 2006; Jangi & Sharp 2014).

La predicción de las estructuras secundarias en TGZ15 y TGZi encontraron un enriquecimiento significativo de  $\alpha$ -helices. En TGZ15 las  $\alpha$ -helices lo componen el 44% junto con un mínimo 2% de  $\beta$ -laminas. La diferencia no es tan remarcada con la proteína TGZi, ya que posee estructuras de  $\alpha$ -helices en un 35% y un 2% de  $\beta$ -laminas. Ambas proteínas son contrastadas por la presencia significativa de regiones flexibles que son atribuidas por las RIDs, en este caso de los TRs.

#### A. Cobertura y confiabilidad de las secuencias templado.



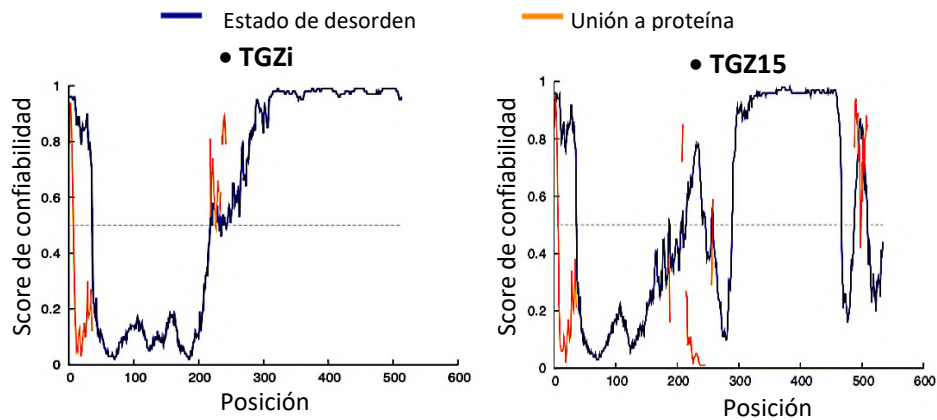
#### B. Información de las secuencias templado.

ID librería plegamientos	ID PDB	Descripción	Organismo
c4cgkA	4CGK	Estructura cristal de la proteína pcsb	<i>Streptococcus pneumoniae</i>
C1c1gA	1C1G	Estructura cristal de la tropomiosina	<i>Sus scrofa</i>
C6o7ua	6O7U	Proteína membrana V-ATPasa STV-VO	<i>Saccharomyces cerevisiae</i>
C6o7xa	6O7X	Proteína membrana V-ATPasa STV-V1VO	<i>Saccharomyces cerevisiae</i>
C3hnmwB	3HNW	Estructura cristal de una proteína coiled-coil de función desconocida	<i>Eubacterium eligens</i>

**Figura 27.** Cobertura, confiabilidad e información de las secuencias templado utilizados en la construcción del modelo proteico de TGZ15 y TGZi. **A.** En rojo se muestra los ajustes de residuos entre los templados y la proteína consulta, las regiones vacías sin cobertura por algún templado fueron modelados por *ab initio*. **B.** En la tabla se describen las secuencias templado con la información respecto a su descripción y organismo.

La proteína TGZ15 fue predicha con un 59% de desorden, en tanto, la TGZi obtuvo una predicción del 56% de regiones desordenadas. La presencia de RIDs significativas están ampliamente ubicadas a lo largo del C-terminal y que además son característicos de los TRs. Los análisis revelaron que la TGZ15 confirmó una alta confiabilidad de RIDs a lo largo de su estructura

primaria, estas regiones repetitivas van de la posición 275-410 (ver Figura 28). Al final del C-terminal, la confiabilidad de RIDs disminuyó, pero encontrando a la par sitios de unión a proteína que se relacionan con una parte del núcleo catalítico de la proteína. En la proteína TGZi, el alto grado de desorden se mantiene presente a lo largo del C-terminal (ver Figura 28), esto se debe a la composición de regiones ricas en residuos polares como Gln, Asp, Ser y Thr.

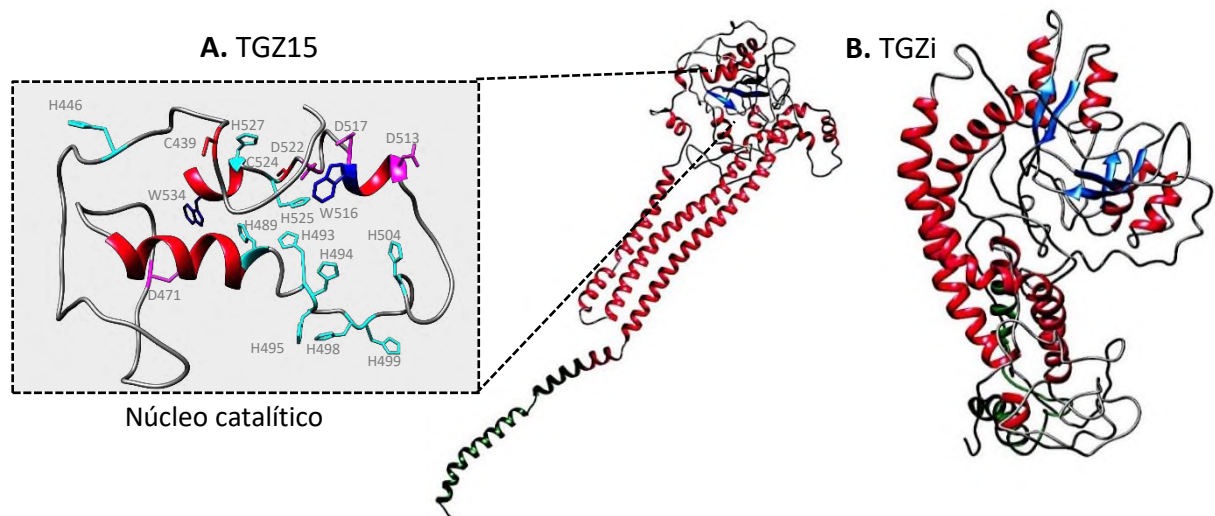


**Figura 28.** Perfil de confiabilidad con el nivel de desorden intrínseco de las proteínas TGZi y TGZ15.

En el resultado final de la predicción, Phyre2 resolvió un modelo en formato pdb que fue visualizado y editado a través del software Chimera. En la Figura 29 se observa el modelo final de la estructura terciaria para la proteína TGZ15 y TGZi. Se resaltaron las formas de las estructuras secundarias presentes. A pesar de que Villalobos *et al.*, (2004) propuso una tríada catalítica ubicada como: Cys439, His446 y Asp471, aún no ha sido posible confirmar la posición espacial de estos, debido a la escasa similitud con otras TGs estudiadas. Sin embargo, con este estudio se demostró la posible presencia de otros residuos posteriores a la Cys439 con similitud importante a las demás TGs caracterizadas, además de una extensión del núcleo catalítico que contiene características importantes como una Cys más en la posición 524.

No obstante, el núcleo catalítico de la TGZ15 se encuentra dentro de una zona medianamente desordenada, por lo que esclarecer la ubicación y comportamiento espacial de los residuos catalíticos dentro de la estructura no es confiable, ya que la flexibilidad que se encuentra dentro de esta región puede tomar múltiples configuraciones. A pesar de lo anterior, se observó que la tríada catalítica propuesta por Villalobos *et al.*, (2004) se encuentran de manera más cercana

debido al plegamiento tomado por la proteína en esa región. Pero no hay que descartar los posibles aminoácidos siguientes que forman un plegamiento adyacente a la Cys439, como la His489, His498 y Asp522. Es importante resaltar la presencia de los únicos dos Trp (posición 516 y 534) presentes en la TGZ15 y ubicados precisamente en el C-terminal; y a pesar de que no presentan la misma similitud con otras TG caracterizadas, es un residuo que participa en el estado de transición por medio de su cadena lateral con un grupo -indol (Iisma *et al.*, 2003; Murthy *et al.*, 2002). La pérdida del núcleo catalítico en la isoforma TGZi y el alto grado de desorden permanente a lo largo del C-terminal resultó en una conformación espacial diferente a la TGZ15 y altamente flexible, careciendo de una estructura estable y que se pliega de manera azarosa (Cuevas-Velázquez & Covarrubias-Robles 2011). Los TRs en las proteínas contienen desde su naturaleza nativa un comportamiento intrínsecamente desestructurado o desordenado (Tompa 2003). La abundancia de TRs desordenados está correlacionado con el nivel de la perfección de la repetición, Kajava *et al.*, (2012) describe que estos eventos respaldan la hipótesis de que, en general, la perfección de la repetición es un signo de eventos evolutivos recientes. Además de que las repeticiones recientemente emergidas se reparen durante la evolución y permitan una respuesta rápida a cambios ambientales (Tompa 2003).



**Figura 29.** Modelo de la estructura tridimensional de las proteínas TGZ15 y TGZi. La visualización tipo *cartoon* muestra en rojo las  $\alpha$ -helices (el péptido tránsito en verde), las  $\beta$ -laminas en azul y en gris las regiones coils. **A.** La región punteada muestra los residuos que han sido propuestos como parte del núcleo catalítico de la TGZ15, ubicados en una zona de gran flexibilidad. **B.** La proteína TGZi se presenta un comportamiento mayormente flexible o azarosa visualizada como random coils en gris.

TGZ15 es una proteína con 534 aminoácidos, su estructura tridimensional fue obtenida mediante métodos de reconocimiento de plegamiento (75% modelado) y *ab initio* (25% modelado). Para comprobar la calidad del modelo proteico predicho, el modelo en formato pdb fue sometido en el servidor Verify3D y PROCHECK. La calidad del modelo obtenido para la proteína TGZ15 resultó con un nivel de aceptación pobre y que es esperado debido a que gran parte de su secuencia fue modelado por *ab initio*. El perfil de score promedio 3D-1D de Verify3D (Ver Figura 30A) mostró que solo el 23.22% se ubican por encima del 0.2, lo que resulta en un modelo deficiente. Por otra parte, el gráfico de Ramachandran (Ver Figura 30A) demostró que la mayor parte de los residuos (representados por un punto negro) que componen a la proteína TGZ15 se encuentran dentro de zonas de color rojo que representan el 82% de aminoácidos posicionados en las regiones más favorables. Además, se encontró el 12.8% de residuos en regiones permitidas adicionales (zonas en color amarillo), un 3.1% de residuos en regiones generosamente permitidas (zonas de color beige) y el 2.1% de residuos en regiones no permitidas (zonas en color blanco). La estereoquímica de los ángulos de torsión propios de los residuos en las regiones permitidas es cercana a un buen modelo de calidad que es esperado a ser por encima del 90% (Ramachandran *et al.*, 1963; Morris *et al.*, 1992).

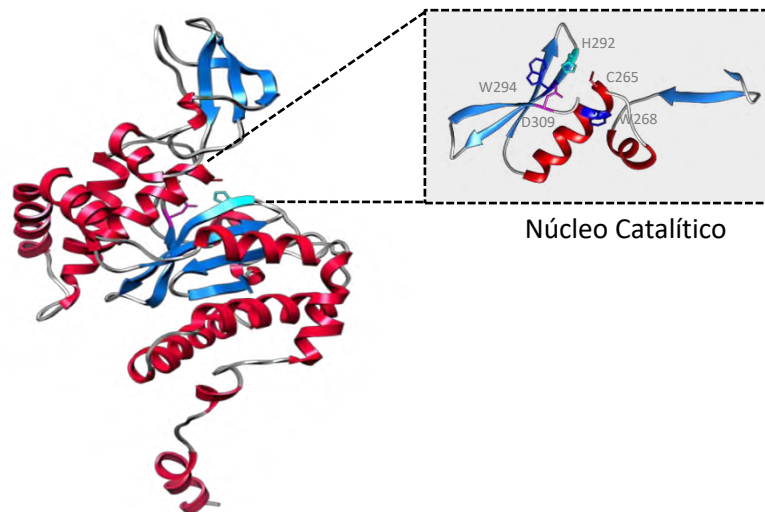
En el caso del modelo 3D de la proteína TGZi, con una composición de 513 aminoácidos, se logró predecir mediante reconocimiento de plegamiento (68% modelado) y *ab initio* (32% modelado). Por lo cual el gráfico de Ramachandran (ver Figura 30B) ubicó al 75.1% de residuos en regiones más favorables (zonas en color rojo). Además, de un 17.8% de residuos en regiones adicionales permitidas (zonas en color amarillo), el 6.2% de residuos en regiones generosamente permitidas (zonas en color beige) y el 0.9% de residuos en regiones no permitidas (zonas en color blanco). Los resultados de Verify3D señalaron una proporción mínima del 20% por encima del 0.2 cuando fue comparado con otros modelos proteicos de calidad.

Los análisis que muestran la calidad de los modelos proteicos, sugieren que ninguno es confiable ya que no mostraron resultados óptimos. Por otra parte, ofrecieron un panorama sobre el comportamiento de las regiones flexibles con respecto a la alta frecuencia de RIDs dentro de las

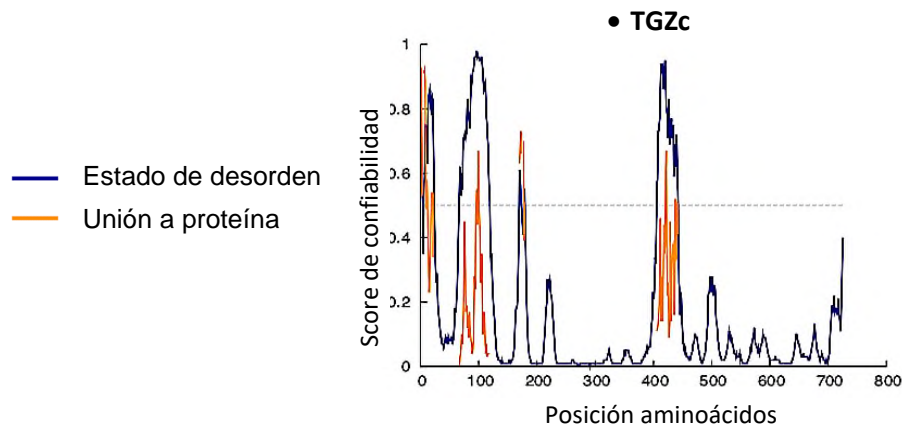


Phyre2 utilizó solo un templado con ID de registro: d1x3za1, que pertenece a un complejo proteico N-glicanasa-Rad23 de *Saccharomyces cerevisiae*. Por lo tanto, pertenecen a la superfamilia de proteasas cisteína por el tipo de plegamiento tomado con el núcleo catalítico típico de las TGs. Con esto se relaciona su identidad como proteína péptido N-glicanasa y su entorno catalítico TG. El dominio TGc está representado por una  $\alpha$ -helice y una  $\beta$ -lamina que comprende el motivo 1 de la Cys265, dos  $\beta$ -laminas que acompañan al motivo 2 y 3 donde residen la His292 y Asp309. El modelo no fue sometido a validación ya que la proteína no fue predicha completamente.

La proteína TGZc homóloga a las proteínas con anotación funcional tipo PNG fue predicha con un grado de desorden (27%) mínimo y un entorno catalítico más ordenado, comparado a las proteínas TGZ15 y TGZi (ver Figura 32) que presentaron una estructura desordenada. Este modelo parcial de la proteína TGZc contiene estructuralmente un 33% en  $\alpha$ -helices y un 24% de  $\beta$ -laminas.



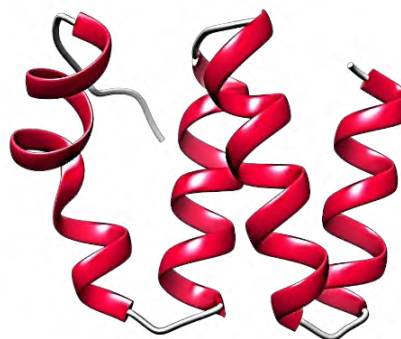
**Figura 31.** Modelo parcial de la proteína TGZc. La visualización tipo *cartoon* muestra en rojo las  $\alpha$ -helices, las  $\beta$ -laminas en azul y en gris las regiones coils. La región punteada muestra los residuos que forman parte del núcleo catalítico de la TG. Esta proteína adopta un plegamiento propio de la superfamilia proteasas cisteína.



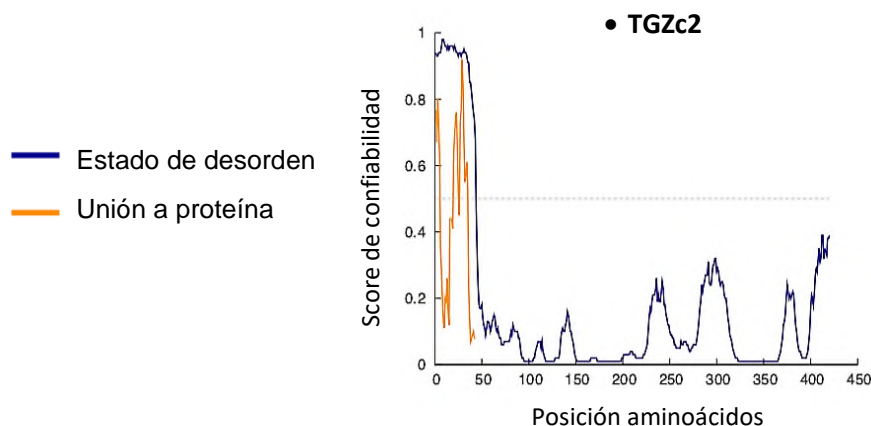
**Figura 32.** Perfil de confiabilidad con el nivel de desorden intrínseco de la proteína TGZc.

### 7.5.3 Modelo estructural de la proteína TGZc2

La estructura terciaria de la proteína TGZc2 (COPED4) fue predicha con el modo normal del servidor Phyre2. Sin embargo, solo se obtuvo una cobertura parcial del 15% (65 residuos) de una longitud de 420 aminoácidos, la confiabilidad modelada fue del 96.3% debido a la alta puntuación del templado que pertenece a una estructura cristal del complejo APC3-APC16 de *Homo sapiens* y clasificada como proteína de unión (ID: c4rg6B; ver Figura 33). El alineamiento estructural con el templado va de la posición 307-378. No obstante, la predicción de la estructura secundaria reflejó una composición del 58% de  $\alpha$ -hélices y una nula composición de  $\beta$ -laminas; se encontró un mediano grado de desorden del 41% (ver Figura 34), las posiciones de los residuos con un alto grado de confiabilidad que fomentan el desorden se encuentran en el N-terminal.



**Figura 33.** Modelo parcial de la proteína TGZc2. La visualización tipo *cartoon* muestra en rojo las  $\alpha$ -hélices y en gris las regiones coils. A pesar de la falta de una cobertura amplia en la predicción de una estructura completa, se identificó un fragmento proteico que pertenece a un templado de una proteína de unión, con un ID de la librería de plegamiento c4rg6B.



**Figura 34.** Perfil de confiabilidad con el nivel de desorden intrínseco de la proteína TGZc2.

## 7.6 Análisis de enriquecimiento con anotaciones GO de las proteínas de interés

Para comprender de manera más detallada la relación funcional existente entre las proteínas putativas TGs y la proteína TGZ15 de maíz, se realizó un análisis de enriquecimiento con la plataforma en línea Ffpred3 donde cada ontología atribuida a la proteína era calificada por un score de acuerdo a la categoría perteneciente. La interacción de ontologías de las diferentes proteínas TG de maíz son estadísticamente enriquecidos tomando en cuenta solo relaciones únicas dentro de las categorías de procesos biológicos y función molecular.

### 7.6.1 Procesos biológicos atribuidos por ontologías en las proteínas TGs de maíz

El análisis de enriquecimiento de procesos biológicos (ver Figura 35) mostró que tanto la proteína TGZ15 ya caracterizada como las proteínas no caracterizadas TGZc, TGZc2 y la isoforma inactiva TGZi son proteínas que se encargan de procesos metabólicos (GO:0019222, GO:0051171). Uno de los principales procesos metabólicos que realizan las TGs en plantas, animales y microorganismos es que actúan como reguladoras de modificaciones en proteínas para protecciones fisiológicas contra diversos ambientes de estrés (Zhong *et al.*, 2019; Rao & Mehta 2004).

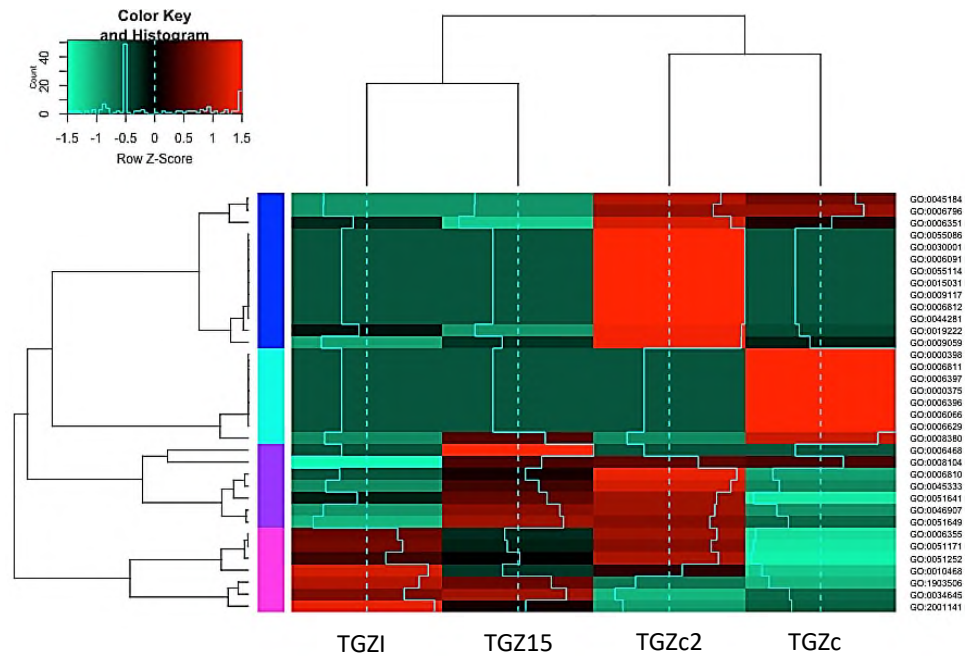
Anotaciones inferidas a partir de procesos biosintéticos (GO:0009059 y GO:0034645) en TGs es respaldado por evidencias que demuestran su participación en el procesamiento de la biosíntesis de las PAs, así como la protección de proteínas en el cloroplasto (GO:0008104). La inactividad tipo TG por parte de la isoforma TGZi no resalta en estos procesos ya que posee una nula actividad

catalítica. El conjunto de propiedades biosintéticas y de protección aumentan la supervivencia de plantas bajo estrés.

La existencia significativa con procesos referentes a la expresión génica (GO:0010468) y procesos transcripcionales (GO:0006355, GO:1903506, GO:0006351) de TGs en plantas han sido encontrado en estudios realizados por Villalobos *et al.*, (2004), Campos *et al.*, (2013) y Della Mea *et al.*, (2004b). Actualmente, Zhong *et al.*, (2019) demostró que el nivel de transcritos de la TG es gradualmente inducido por luz. Por lo que se obtiene información acerca de que la luz juega un rol vital para la regulación de la expresión de la TG en plantas.

Las PAs son moléculas catiónicas que sirven como sustratos de las TGs vegetales y como reguladoras del crecimiento, también se ha mencionado un rol protector del aparato fotosintético a través de la regulación de proteínas fotosintéticas. El concepto general en plantas es que la acumulación de PAs está relacionada con el estrés salino prolongado (Maiale *et al.*, 2004), debido a que su concentración se eleva durante condiciones de estrés por sales. La síntesis y oxidación de PAs toma lugar en los cloroplastos por lo que la adición de PAs ayuda a inhibir la destrucción de los tilacoides, así como también previene la pérdida de pigmentos durante el estrés salino (Dondini *et al.*, 2003). Sin embargo, experimentos realizados en pepino y tomate demostraron que el contenido de PAs disminuía bajo condiciones de estrés, resultando en un daño en los órganos fotosintéticos (Duan *et al.*, 2008; Sang *et al.*, 2017).

El hecho es que tanto PAs como TGs son vinculadas dentro de los cloroplastos, donde estas enzimas son un elemento crucial de los tilacoides (Serafini-Fracassini & Del Duca 2008). Al respecto, Zhong *et al.*, (2019) reveló que las TGs en plantas juegan roles positivos en la acumulación de PAs y que además inducen la transcripción de genes relacionados al proceso fotosintético en cloroplastos, con lo que se establece un rol crucial en la regulación de la estabilidad de órganos fotosintéticos (Zhong *et al.*, 2019). Estos roles ya habían sido sugeridos anteriormente por Villalobos *et al.*, (2004) y Della Mea *et al.*, (2004b). Así que la TG en plantas participa en la mejora de ambientes de estrés mediante la modulación de proteínas fotosintéticas.



**Figura 35.** Gráfico tipo mapa de calor representando las anotaciones de procesos biológicos relacionados a proteínas anotadas funcionalmente como TGs de maíz. Los valores de expresión son comparados entre las distintas proteínas, donde cada valor de score se encuentra representado mediante una tonalidad, las celdas con un color rojo claro representan valores de score significantes y el tono color verde claro representa a valores de score no significantes. Sin embargo la transición de tonalidades en una misma fila denota la existencia de valores semejantes entre las proteínas. Las barras de colores representan los clusters formados de acuerdo a scores similares dentro de una anotación GO.

Otro factor probable a la tolerancia o resistencia a ambientes de estrés, así como la estabilidad proteica y afinidad por substratos, tiene que ver con la composición de las regiones repetitivas (TRs) encontrados en la TGZ15. Existe la hipótesis que muchos de los TRs encontrados en una diversidad de proteínas vegetales, son enriquecidos con anotaciones en procesos metabólicos, respuestas a estímulos de estrés biótico y abiótico, señales de transducción, modificación macromolecular y procesos biosintéticos celulares. Los TRs juegan un rol en los andamios estructurales, promoviendo la formación de complejos moleculares, así como interacciones. Han demostrado que la alta conservación de TRs son enriquecidas con anotaciones funcionales de unión, donde un cambio en la forma que posean los TRs puede llegar a ser perjudicial y determinante en los efectos funcionales utilizados selectivamente en contra a lo largo de la

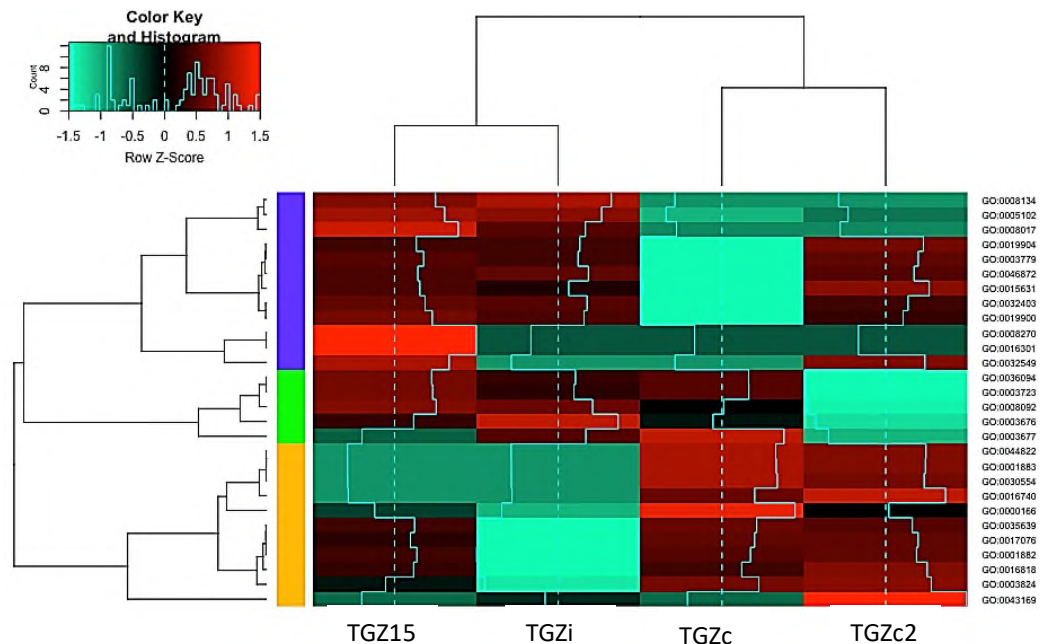
evolución. Los TRs altamente mutables pueden proporcionar una fuente de variación para una rápida adaptación (Schaper & Anisimova 2015).

### **7.6.2 Funciones moleculares atribuidos por ontologías en las proteínas TGs de maíz**

Las funciones moleculares (ver Figura 36) que fueron inferidos en estas proteínas, demostraron el enriquecimiento de la actividad catalítica (GO:0003824) con mayor significancia en las proteínas TGZc y TGZc2. La inferencia de este enriquecimiento es de acuerdo a la similitud con núcleos catalíticos previamente estudiados, por lo que hasta el momento no existía análisis bioinformático que vinculara un núcleo catalítico TG similar a las TGs animales en TGZ15, sin embargo, FFpred3 demostró una pequeña significancia con la proteína caracterizada de Villalobos *et al.*, (2004). Considerando que la naturaleza enzimática de estas proteínas es del tipo transferasa, la anotación para este tipo de actividad transferasa (GO:0016740), solo fue inferida significativamente en las proteínas TGZc y TGZc2.

Las propiedades de una estructura estable y ordenada son frecuentes en proteínas involucradas en actividades de catálisis, unión de pequeños ligandos, así como proteínas de la membrana (Covarrubias *et al.*, 2017; Radivojac *et al.*, 2013). El desorden estructural propiciado por RIDs dentro del núcleo catalítico de la TGZ15 y de la enzima inactiva TGZi (ver Figura 28) es significativamente alto comparado a un nicho hidrofóbico y altamente ordenado como el que muestra la TG2h (ver Figura 3). Interesantemente, existe una dicotomía con la descripción de algunas proteínas enzimáticas que contienen RIDs en formas estructurales de loops o tails. Estos sitios desestructurados y flexibles participan en la modificación conformacional de la proteína más allá de la unión a substratos, exponiendo los residuos catalíticos que contribuyen a la actividad enzimática (McElheny *et al.*, 2005; Schulenburg & Hilvert 2013). Las RIDs dentro de TGZ15 le han conferido ciertos roles y anotaciones únicas respecto a las otras, donde ciertos sitios actúan como modulador de modificaciones post-traduccionales, como la fosforilación de proteínas (GO:0006468), es sugerido que estos sitios, generalmente actúan como switches de la activación o inactivación y/o incluso como moduladores de su actividad (Covarrubias *et al.*, 2017).

La especificidad de sustratos con sitios de unión a tubulina (GO:0015632) y actina (GO:0003779) se encuentra ampliamente reportado en plantas superiores, muchos de estos sitios de unión son inferidos por la similitud ya reportada en datos previos obtenidos a partir de estudios en células animales (descrito ampliamente en: Serafini-Fracassini *et al.*, 2009) y que guardan una gran relación con sitios de unión característicos en TGZ15, TGZc2 y la isoforma inactiva TGZi. Tanto actina como tubulina son proteínas propias del citoesqueleto, el cual también guarda un sitio de unión (GO:0008092) presente en las 4 TGs de maíz.



**Figura 36.** Gráfico tipo mapa de calor representando las anotaciones de funciones moleculares relacionados a proteínas anotadas funcionalmente como TGs de maíz. Los valores de expresión son comparados entre las distintas proteínas, donde cada valor de score se encuentra representado mediante una tonalidad, las celdas con un color rojo claro representan valores de score significantes y el tono color verde claro representa a valores de score no significantes. Sin embargo la transición de tonalidades en una misma fila denota la existencia de valores semejantes entre las proteínas. Las barras de colores representan los clusters formados de acuerdo a scores similares dentro de una anotación GO.

Todos estos sitios de unión ligados a la especificidad, control y regulación con proteínas sustrato en el citoesqueleto tienen un respaldo bibliográfico que describe el rápido crecimiento del tubo del polen a partir de una reorganización dramática del citoesqueleto, que es probablemente

causada por la actividad TG en el polen y que ha sido sugerida a participar en la regulación de actina y tubulina. Así que cualquier arreglo en el motor citoesquelético genera cambios críticos en la motilidad del organelo y el crecimiento del tubo del polen (Del Duca et al., 2014).

También se encontró una anotación de sitios de unión de microtúbulos (GO:0008017) con gran relevancia en la proteína TGZ15. Interesantemente la actividad reticulante de la TG2 involucrada en la organización del citoesqueleto se encuentra mediado por proteínas substrato como tau microtúbulos, b-tubulin, actina, miosina, espectrina, timosin b, troponin T y vimentin. Por lo que se ha demostrado que la TG2 está involucrada en la estabilización de los microtúbulos mediante la poliaminación de tubulinas (Piacentini *et al.*, 2014).

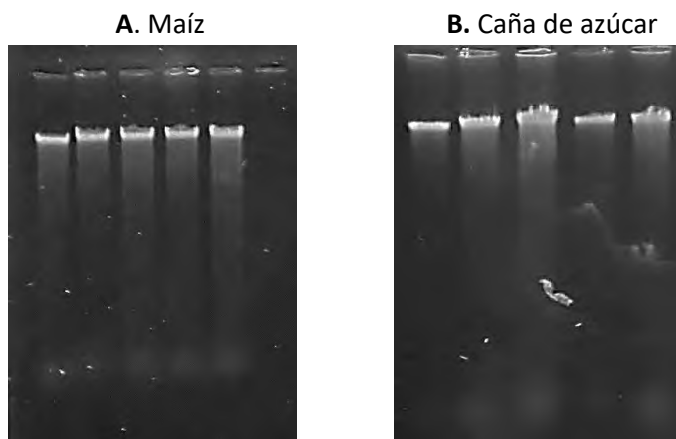
Descripciones de actividad quinasa (GO:0016301) encontrada en TG2 es también hallada significativamente en TGZ15, así como también sitios de unión a quinasa (GO:0019900), que, como se revisó anteriormente, son sitios ubicados en las RIDS (Ward *et al.*, 2004). Los sitios de unión a ácidos nucleicos, nucleósidos, ADN y ARN (GO:0003676, GO:0000166, GO:0017076, GO:0001882, GO:0035639, GO:0003677, GO:0032549, GO:0003723, GO:0008270) son hallados con un alto grado de score en TGZ15 y que además son funciones análogas en otras TGs de animales (Liu *et al.*, 2002; Zanetti *et al.*, 2004; Gundemir *et al.*, 2012).

Aunque no se tiene confirmada la actividad que puede desempeñar la TGZi dentro de estos procesos, la posible participación por parte de procesos transcripcionales encontrados por regiones ricas en glutamina y alanina (o también llamados polyQ y polyA) como lo describe Kottenhagen *et al.*, (2012), sería una posible sugerencia de actividad TGZi, enriquecido por este tipo de anotaciones en la participación de unión a factores de transcripción (GO:0008134).

## 7.7 Análisis molecular

### 7.7.1 ADN extraído de hojas de maíz (*Zea mays*) y caña de azúcar (*Saccharum sp.* híbrido Mex 69-290)

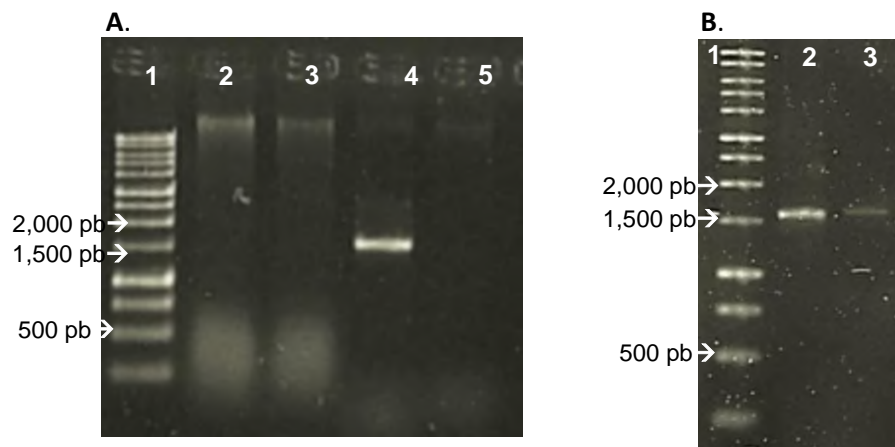
La extracción de ADN a partir de tejido foliar de caña azúcar (*Saccharum sp.* híbrido Mex 69-290) y maíz (*Zea mays*), fue posible de acuerdo al protocolo de Bermúdez-Guzmán *et al.*, (2016). En maíz se obtuvo una concentración de 1389 ng/ $\mu$ L con una pureza (relación 260/280) de 2.00 a 2.09 (ver Figura 37A) y en caña de azúcar se obtuvo una concentración de 1262 ng/ $\mu$ L con una pureza (relación 260/280) de 1.80 a 2.07 (ver Figura 37B). Cada muestra fue diluida a 100 ng.



**Figura 37.** Electroforesis en gel de agarosa al 1% de ADN a partir de tejido foliar de maíz y caña de azúcar diluidas a 100 ng.

### 7.7.2 Amplificación del gen *tgz15* de maíz (*Zea mays*) como control positivo

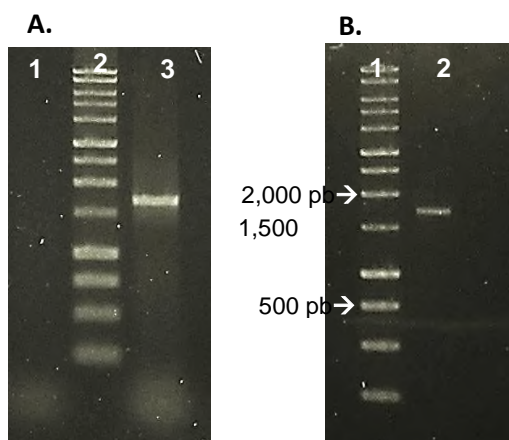
Para demostrar la eficiencia de amplificación de los oligonucleótidos diseñados, se utilizó un control positivo y la verificación de la calidad del ADN de maíz. Previamente a la preparación de la PCR, se utilizaron muestras de ADN a 100 ng, 50 ng y 20 ng. Cada una de estas muestras diluidas fueron utilizadas en la PCR junto con los oligos TG3F y TG3R. Con la verificación del gel de agarosa por electroforesis se obtuvo una banda de aproximadamente 1600 pb (ver Figura 38A), considerando un margen de error por la apreciación. Con estos resultados se demostró la eficiencia del proceso de PCR con un ADN molde diluido a 20 ng. El producto amplificado de aproximadamente 1600 pb fue purificado por el kit comercial Wizard® SV Gel and PCR Clean-Up System obteniendo una concentración de 72.06 ng/ $\mu$ L con una pureza (relación 260/280) de 1.94 (ver Figura 38B) y un rendimiento de 3.60  $\mu$ g.



**Figura 38.** Amplificación y purificación del gen *tgz15* en maíz por PCR. **A.** Amplificación del gen *tgz15* de maíz. En el carril 1) se muestra el marcador de peso molecular de 1 Kb, carril 2) PCR con un ADN molde a una concentración de 100 ng, carril 3) PCR con un ADN molde a una concentración de 50 ng, carril 4) PCR con un ADN molde a una concentración de 20 ng y carril 5) control negativo (sin ADN molde). **B.** Purificación del producto amplificado. En el carril 1) se muestra el marcador de peso molecular de 1kb, el carril 2) la purificación del producto amplificado mediante la escisión del fragmento del gel y el carril 3) purificación del producto amplificado directamente de PCR.

### 7.7.3 Aislamiento del gen *tg* de caña de azúcar (*Saccharum* sp. híbrido Mex 69-290) mediante PCR

El aislamiento del gen *tg* en caña de azúcar mediante la técnica de PCR de tiempo final con los oligonucleótidos TG3F y TG3R amplificó una banda de aproximadamente 1800 pb. (ver Figura 39A). Este gen *tg* en caña de azúcar es altamente similar al gen *tgz15* de maíz y que hasta el momento no ha sido aislado.

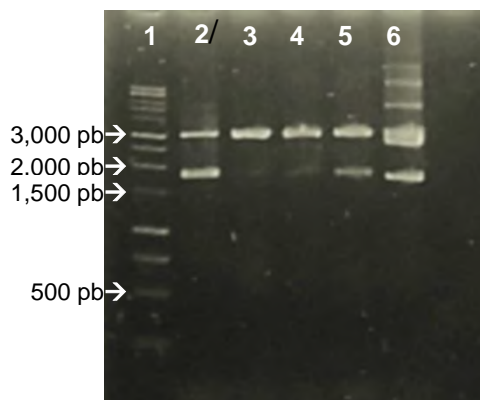


**Figura 39.** Amplificación y purificación del gen *tg* en caña de azúcar por PCR. En **A.** Se muestran 3 carriles, el carril 1) control negativo sin ADN molde, carril 2) marcador de peso molecular de 1 kb y carril 3) producto amplificado con el oligonucleótido TG3F y TG3R tiene un tamaño aproximado de 1800pb. **B.** El producto amplificado es purificado mediante la escisión de la banda del gel de agarosa al 1%.

#### 7.7.4 Clonación del gen *tg* de caña de azúcar (*Saccharum* sp. híbrido Mex 69-290)

A partir del ADN extraído de hojas de caña de azúcar se amplificó el gen *tg* con los oligos específicos del gen *tgz15* de maíz utilizados como templado. Los productos de amplificación fueron clonados en el vector pGEM-T Easy. Se realizó la transformación de células competentes de *E. coli* DH5 $\alpha$  junto con una mezcla de ligación. Las colonias transformadas fueron aisladas y seleccionadas al azar para crecer y extraer ADN plasmídico.

Por medio de lisis alcalina se realizaron minipreparaciones para extraer y purificar el ADN plasmídico que contenía el gen amplificado. Se obtuvieron diversas concentraciones de plásmido purificado a partir de las colonias seleccionadas. Para verificar el estado de los plásmidos se realizaron digestiones con la enzima de restricción EcoRI y se verificó el tamaño del inserto de aproximadamente 1600 pb (ver Figura 40). Con la verificación de la digestión en el gel de agarosa al 1% (ver Figura 40), se seleccionaron 2 minipreps, donde se seleccionó el purificado plasmídico del carril 2 con una concentración de 1804.96 ng/ $\mu$ L y pureza (relación 260/280) 1.80 y del carril 6 con una concentración 2223.83 ng/ $\mu$ L y pureza (relación 260/280) 1.93.



**Figura 40.** Verificación de la ligación del fragmento del gen *tg* de caña de azúcar por la enzima de restricción EcoRI. La digestión con enzimas EcoRI en el plásmido pGem-T Easy del gen *tg* en caña de azúcar se llevó a cabo a partir de la selección al azar de 5 colonias aisladas. En el carril 1) se muestra el marcador de peso molecular 1kb, en los demás carriles 2,3,4,5 y 6) el vector presenta un tamaño de 3 kb y el inserto de aproximadamente 1800 pb.

#### **7.7.5 Análisis del fragmento del gen *tg* de caña de azúcar (*Saccharum* sp. híbrido Mex 69-290) obtenido por secuenciación**

Por secuenciación del método Sanger se obtuvo una secuencia parcial de un gen codificante *tg* en caña de azúcar (ver Figura 41), obtenido con el fragmento de PCR amplificado y clonado en el vector pGem-T Easy, este producto génico posee una longitud de 103 pb. Río arriba de la secuencia codificante se encontró un intrón de 905 pb., sumando una longitud total de 1008 pb (ver Anexo 11.18, recientemente ingresada a la NCBI con el código de acceso: BANKit2210131 transglutaminase MK748471).

Mediante una búsqueda de similitud en las bases de datos de la NCBI por BLASTn, la secuencia codificante amplificada presentó un alto grado de identidad (94-95%) con fragmentos de secuencias génicas en *tgs* de maíz, entre las que destacan *tgz15* (AJ421525.1), *tgz21* (AJ488103.1) y la variante inactiva *tgzi* (NM\_001111688.2). El intrón no resultó con similitud alguna respecto a otra secuencia genómica existente en las bases de datos de la NCBI. No obstante, con la actual secuenciación del genoma de *Saccharum officinarum* (Garsmeur *et al.*, 2018) se accedió a la base de datos: <http://sugarcane-genome.cirad.fr>, donde se obtuvo una similitud considerable con la secuencia "Sh\_205O16\_g000010 Transglutaminase" (ver Anexo 11.18) que es similar desde la posición 13-1001, registrando un umbral esperado de 0.0 y una identidad del 97%, demostrando así la identidad parcial del gen codificante *tg* en esta especie.

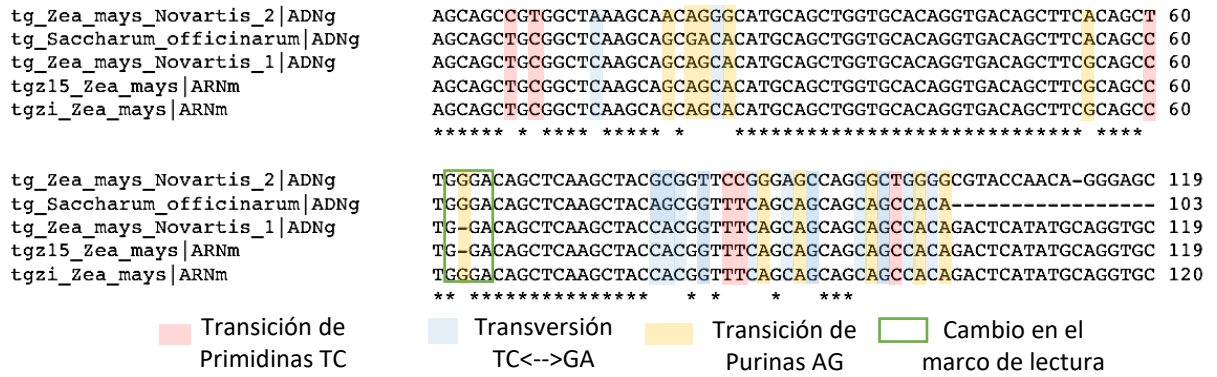
La secuencia parcial codificante en caña de azúcar se comparó con secuencias provenientes de ARNm como *tgz15* (Villalobos *et al.*, 2004) y *tgzi* de la variedad B73, así como con secuencias provenientes de ADNg de la variedad Novartis de maíz (ver Figura 42). Se encontraron diferentes substituciones nucleotídicas que ocurren en la primera y tercera posición del codón, las cuales

terminaban en mutaciones sinónimas, características propias de la redundancia del código genético que lo hace degenerado. En diferentes organismos eucariotas existen codones sinónimos que son preferidos sobre otros codones sinónimos, lo que ocasiona una frecuencia de aparición respecto a codones sinónimos dentro de un genoma. Este fenómeno es conocido como



**Figura 41.** Esquema de la secuencia obtenida del producto de PCR clonado del gen *tg* en caña de azúcar.

La razón de porqué ocurre con gran frecuencia el uso de codones sinónimos también se encuentra en función del contenido de G+C dentro del genoma. En estudios hechos a partir de especies monocotiledóneas y dicotiledóneas se descubrió que el sesgo en el uso del codón se correlacionó con el contenido de G+C en la tercera posición del codón. Los contenidos de G+C fueron más altos en especies de monocotiledóneas que en dicotiledóneas en todas las posiciones de codones. Los altos contenidos de G+C de las especies de monocotiledóneas pueden ser el resultado de sesgos mutacionales relativamente fuertes que ocurrieron en el linaje de las especies *Poaceae* (Mittra *et al.*, 2016; Kawabe & Miyashita 2003).



**Figura 42.** Análisis comparativo de secuencias *tg* codificantes de *Zea mays* y *Saccharum sp.* híbrido *Mex 69-290* provenientes de ADNg y ARNm. Se compararon secuencias codificantes provenientes de ARNm de *tgz15* (AJ421525.2) y *tgzi* (NM\_001111688.2) de *Zea mays* de la variedad B73. Además, se incluyeron secuencias codificantes *tg* provenientes de ADNg de dos bandas amplificadas de *Zea mays* de la variedad Novartis (Novartis\_1 y Novartis\_2) estudiadas en la tesis de Villalobos (2001) y la secuencia parcial *tg* de *Saccharum sp.* híbrido *Mex 69-290*.

A partir de lo descrito, se confirma la presencia del evento mutacional en la secuencia parcial *tg* de *Saccharum sp.* híbrido *Mex 69-290* existente en otras secuencias codificantes homólogas a TGZ15, donde la inserción de una guanina fuera del marco de lectura en el codón G-GA → GGG resulta en una mutación sinónima que codifica para Glicina (G), lo que ocasiona un cambio de marco de lectura río abajo de la mutación. Este comportamiento mutacional es una característica que ocurre cuando la región de codificación de un gen se altera mediante la adición o eliminación de una cantidad de pares de bases que no es un múltiplo de tres. La ocurrencia de una eliminación frente a un tipo de inserción de desplazamiento de marco depende de la naturaleza de la estructura intermedia transitoria formada durante la síntesis de ADN (Ripley 1990).

## 8. CONCLUSIONES

A partir de una búsqueda exhaustiva de la distribución y organización de elementos biológicos anotados funcionalmente en las bases de datos (primarias y secundarias) como transglutaminase, se obtuvo una tendencia cuantitativa (de acuerdo al número de secuencias depositadas) y cualitativa (de acuerdo a la confiabilidad que ofrecía cada secuencia) que provee información acerca de la calidad de los datos registrados entre diferentes bases de datos. La redundancia de información para este análisis puede ser una guía para entender la variabilidad de secuencias biológicas tipo TG encontradas en diversos organismos.

Los análisis comparativos de secuencias demostraron que la isoforma TGZ<sub>i</sub> de maíz pertenece a un grupo de proteínas homólogas no enzimas, anotadas funcionalmente como TGs, que divergen por un proceso de duplicación génica que las hace más tolerantes a las mutaciones y conservativas en el proteoma de las gramíneas. Las proteínas TGZ<sub>15</sub> de maíz y TGO de arroz son enzimas codificados por genes de copia única y que actúan como los precursores catalíticos ancestrales homólogos.

La inferencia de la anotación funcional TG en las proteínas TGZ<sub>c</sub> y TGZ<sub>c2</sub> solo está relacionada por la similitud que guardan con dominios estructurales tipo TG (PF01841 y PF13369) que conservan la tríada catalítica. La proteína TGZ<sub>c</sub> debe ser considerada una PNGasa por su cercana relación evolutiva con homólogos en animales y plantas, su predicción estructural encaja perfectamente con un complejo proteico N-glicanasa-Rad23 de *Saccharomyces cerevisiae* (ID: d1x3za1). En el caso de la proteína TGZ<sub>c2</sub>, no debe ser inferida como TG ya que su relación con el dominio funcional "Transglut\_core2" está limitado por una identificación parcial del dominio sin homología con otras proteínas caracterizadas que den certeza a su funcionalidad.

Se postula que el desorden estructural provisto por RIDs en las proteínas TGZ<sub>15</sub> y TGZ<sub>i</sub> tienen un impacto en la estabilidad proteica y afinidad de substratos, esto es apoyado por las anotaciones enriquecidos en funciones moleculares y procesos biológicos que describen actividades de unión, interacción en procesos metabólicos, señales de transducción, modificación macromolecular, procesos biosintéticos celulares, así como procesos de adaptación como respuesta al estrés.

La enzima TGZ15 contiene un núcleo catalítico obtenido de un origen evolutivo diferente que conserva sus propiedades fisicoquímicas, la obtención de una estructura análoga que diverge del resto de TGs caracterizadas solo es producto de una evolución convergente.

El gen *tgz15* de maíz presentó un gen homólogo en caña de azúcar (*Saccharum* sp. híbrido Mex 69-290) que contiene el desplazamiento del codón por inserción puntual de una guanina que ocasiona el cambio de marco de lectura río abajo y posteriormente la pérdida de los residuos involucrados en la tríada catalítica. Este hallazgo ofrece una evidencia más que la robustez mutacional en el genoma de las gramíneas permite incrementar la tolerancia de genes duplicados a las mutaciones.

## 9. PERSPECTIVAS Y RECOMENDACIONES

Con base a los resultados y discusiones realizados en este trabajo de investigación, y con el propósito de mejorar la calidad de los mismos, se sugiere lo siguiente:

- Discriminar las anotaciones funcionales de las proteínas con base en un análisis exhaustivo antes de confiar en el registro de su función biológica depositado en las bases de datos.
- Caracterizar y clonar la longitud completa del gen tanto del *tgz15* como su homólogo en caña de azúcar y estudiar los mecanismos de su regulación génica, con el objetivo de identificar elementos reguladores en su secuencia y factores de transcripción responsables de la expresión de ambos genes en los diferentes tejidos de las plantas, así como también su respuesta ante tensores ambientales.
- Seguir estudiando el proceso de divergencia funcional que adquiere la isoforma inactiva de la TGZ15 de maíz.
- Hallar nuevos productos génicos de TGs en otras gramíneas, y con ello realizar un análisis evolutivo que permita obtener un nuevo precedente sobre cambios evolutivos en dicha familia vegetal.
- Implementar el uso de nuevos algoritmos que nos permita optimizar y encontrar las similitudes fisicoquímicas del sitio activo de la TGZ15 y otras TGs ya estudiadas.
- Buscar una estrategia más eficiente que permita modelar la estructura proteica de la TGZ15 (Q6KF70) y TGZi (B8A2N4).

## 10. REFERENCIAS

1. Motokis, M., & Kumazawa, Y. (2000). Recent research trends in transglutaminase technology for food processing. *Food Science and Technology Research*, 6(3), 151-16.
2. Griffin, M., Casadio, R., & Bergamini, C.M. (2002). Transglutaminases: Nature's biological glues. *Biochemical Journal*, 368(2), 377-396.
3. Lorand, L., & Graham, R.M. (2003). Transglutaminases: crosslinking enzymes with pleiotropic functions. *Nature Reviews. Molecular Cell Biology*, 4(2), 140-56.
4. Li, H., Cui, Y., Zhang, L., Zhang, L., Liu, H., & Yu, J. (2017). Optimization of recombinant *Zea mays* transglutaminase production and its influence on the functional properties of yogurt. *Food Science and Biotechnology*, 26(3), 723-730.
5. Marakova, K.S., Aravind, L., & Koonin, E.V. (1999). A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein Science*, 8(8), 1714-1719.
6. Yokoyama, K., Nio, N., & Kikuchi, Y. (2004). Properties and applications of microbial transglutaminase. *Applied Microbiology and Biotechnology*, 64(4), 447-454.
7. Langston, J., Blinkovsky, A., Byun, T., Terribilini, M., Ransbarger, D., & Xu, F. (2006). Substrate specificity of *Streptomyces* transglutaminase. *Applied Biochemistry and Biotechnology*, 136(3), 291-308.
8. Della Mea, M., Caparrós-Ruiz, D., Claparols, I., Serafini-Fracassini, D., & Rigau, J. (2004a). AtPng1p. The first plant transglutaminase. *Plant Physiology*, 135(4), 2046-2054.
9. Villalobos, E., Santos, M., Talavera, D., Rodríguez-Falcón, M., & Torné, J.M. (2004). Molecular cloning and characterization of a maize transglutaminase complementary DNA. *Gene*, 336(1), 93-104.
10. Serafini-Fracassini, D., Della Mea, M., Tasco, G., Casadio, R., & Del Duca, S. (2009) Plant and animal transglutaminases: ¿do similar functions imply similar structures? *Amino Acids*, 36(4), 643-657.
11. Luscombe, N.M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? An introduction and overview. *Yearbook of Medical Informatics*, 10(1), 83-100.
12. Hesper, B., & Hogeweg, P. (1970). Bioinformatica: een werkconcept. *Kameleon*, 1(6), 28-29.
13. Xiong, J. (2006). *Essential bioinformatics*. Cambridge University Press.

14. Choudhuri, S. (2014). *Bioinformatics for beginners: genes, genomes, molecular evolution, databases and analytical tools*. Elsevier.
15. Sebastián-Yagüe, Á., Pascual-García, A., Abascal, F., Aguirre, J., Bajic, D., Baú, D., ... & Santos, H.G.D. (2014). "Bioinformática con Ñ v1. 0": a collaborative project of young Spanish scientists to write a complete book about Bioinformatics.
16. Schneider, M., Bairoch, A., Wu, C.H., & Apweiler, R. (2005). Plant protein annotation in the UniProt Knowledgebase. *Plant Physiology*, 138(1), 59-66.
17. UniProt Consortium. (2007). The universal protein resource (UniProt). *Nucleic Acids Research*, 36(1), D190-D195.
18. Brenner, S.E. (1999). Errors in genome annotation. *Trends in Genetics*, 15(4), 132-133.
19. Nadzirin, N., & Firdaus-Raih, M. (2012). Proteins of unknown function in the protein data bank (pdb): An inventory of true uncharacterized proteins and computational tools for their analysis. *International Journal of Molecular Sciences*, 13(10), 12761-12772.
20. Percudani, R., Carnevali, D., & Puggioni, V. (2013). Ureidoglycolate hydrolase, amidohydrolase, lyase: how errors in biological databases are incorporated in scientific papers and vice versa. *Database*, 2013, 1-9.
21. Liberal, R., & Pinney, J.W. (2013). Simple topological properties predict functional misannotations in a metabolic network. *Bioinformatics*, 29(13), i154-i161.
22. Sarkar, N.k., Clarke, D.D., & Waelsch, H. (1957). An enzymically catalyzed incorporation of amines into proteins. *Biochimica et Biophysica Acta*, 25(2), 451-452.
23. Lorand, L., & Conrad, S.M. (1984). Transglutaminases. *Molecular and Cellular Biochemistry*, 58(1-2), 9-35.
24. Mehta, K., & Eckert, R. (2005). *Transglutaminases: Family of enzymes with diverse functions*. Karger Medical and Scientific Publishers.
25. Folk, J.E., & Finlayson, J.S. (1977). The epsilon-(gamma-glutamyl)lysine crosslink and the catalytic role of transglutaminases. In *Advances in Protein Chemistry*, 31, 1-33.
26. Mycek, M.J., Clarke, D.D., Neidle, A., & Waelsch, H. (1959). Amine incorporation into insulin as catalyzed by transglutaminase. *Archives of Biochemistry Biophysics*, 84(2), 528-540.
27. Beninati, S., Bergamini, C.M., Piacentini, M. (2009). An overview of the first 50 years of Transglutaminase research. *Amino Acids*, 36(4), 591-598.

28. Piacentini, M., Rodolfo, C., Farrace, M.G., & Autuori, F., (2004). Tissue transglutaminase in animal development. *International Journal of Developmental Biology*, 44(6), 655-662.
29. Mirzaei, M. (2011). Microbial Transglutaminase application in food industry. *International Conference on Food Engineering and Biotechnology*, 9, 267-271.
30. Klöck, C., & Khosla, C. (2012). Regulation of the activities of the mammalian transglutaminase family of enzymes. *Protein Science*, 21(12), 1781-1791.
31. Del Duca, S., Verderio, E., Serafini-Fracassini, D., Iorio, R., & Cai, G. (2014). The plant extracellular transglutaminase: what mammal analogues tell. *Amino Acids*, 46(3), 777-792.
32. Kashiwagi, T., Yokoyama, K., Ishikawa, K., Ono, K., Ejima, D., Matui, H., & Suzuki, E. (2002). Crystal structure of microbial transglutaminase from *Streptovercillium mobaraense*. *Journal of Biological Chemistry*, 277(46), 44252-44260.
33. Lorand, L. (2007). Crosslinks in blood: transglutaminases and beyond. *The FASEB Journal*, 21(8), 1627.
34. Williams-Ashman, H.G., & Canellakis, Z.N. (1980). Transglutaminase-media covalent attachment of polyamines to proteins: mechanisms and potential physiological significance. *Physiological Chemistry and Physics*, 12(5), 457-472.
35. Elgavish, A., Wallace, R.W., Pillion, D.J., & Meezan, E. (1984). Polyamines stimulate d- glucose transport in isolated renal brush-border membrane vesicles. *Biochimica et Biophysica Acta*, 777(1), 1-8.
36. Ho, M.L., Leu, S.Z., Hsieh, J.F., & Jiang, S.T. (2000). Technical approach to simplify the purification method and characterization of microbial transglutaminase produced form *Streptovercillium ladakanu*. *Journal of Food Science*, 65(1), 76-80.
37. Barros, S.L.H., Assmann, F., & Zachia, A.M.A. (2003). Purification and properties of a transglutaminase produced by a *Baciilus circulans* strain isolated from Amazon environment. *Biotechnology and Applied Biochemistry*, 37(3), 295-299.
38. Fesüs, L., & Laki, K. (1977). Two antigenic sites of tissue transglutaminase. *Biochemistry*, 16(18), 4061-4066.
39. Murthy, S.N., Lomasney, J.W., Mak, E.C., & Lorand, L. (1999). Interactions of G(h)/transglutaminase with phospholipase Cdelta1 and with GTP. *Proceedings of the National Academy of Sciences*, 96(21), 11815-11819.

40. Eckert, R.L., Kaartinen, M.T., Nurminskaya, M., Belkin, A.M., Colak, G., Johnson, G.V. & Mehta, K. (2014). Transglutaminase regulation of cell function. *Physiological Reviews*, 94(2), 383-417.
41. Folk, J.E., Park, M.H., Chung, S.I., Schrode, J., Lester, E.P., Cooper, H.L. (1980). Polyamines as physiological substrates for transglutaminases. *Journal of Biological Chemistry*, 255(8), 3695-3700.
42. Beninati, S., Piacentini, M., Cocuzzi, E.T., Autuori, F., & Folk, J.E. (1988). Covalent incorporation of polyamines as gamma-glutamyl derivatives into CHO cell protein. *Biochimica et Biophysica Acta*, 952(3), 325-333.
43. Iismaa, S.E., Mearns, B.M., Lorand, L., & Graham, R.M. (2009). Transglutaminases and disease: lessons from genetically engineered mouse models and inherited disorders. *Physiological Reviews*, 89(3), 991-1023.
44. Gatta, N.G., Cammarota, G., & Gentile, V. (2016). Possible roles of transglutaminase in molecular mechanisms responsible for human neurodegenerative diseases. *AIMS Biophysics*, 3(4), 529-545.
45. Bruns, N., & Kilbinger, A.F. (2016). *Bio-inspired Polymers*. Royal Society of Chemistry.
46. Folk, J.E., & Chung, S.I. (1973). Molecular and catalytic properties of transglutaminase. *Advances in Enzymology and Related Areas Molecular Biology*, 38, 109-191.
47. Iisma, S.E., Holman, S., Wouter, M.A., Lorand, L., Graham, R.M. & Husain, A. (2003). Evolutionary specialization of a tryptophan indole group for transition-state stabilization by eukaryotic transglutaminase. *Proceedings of National Academy Sciences*, 100(22), 12636-12641.
48. Murthy, S.N., Iisma, S., Begg, G., Freymann, D.M., Graham, R.M., & Lorand, L. (2002) Conserved tryptophan in the core domain of transglutaminase is essential for catalytic activity. *Proceedings of National Academy Sciences*, 99(5), 2738-2742.
49. Pedersen, L.C., Yee, V.C., Bishop, P.D., Trong, I.L., Teller, D.C., & Stenkamp, R.E. (1994). Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Science*, 3(7), 1131-1135.
50. Chen, J.S., & Mehta, K. (1999). Tissue transglutaminase: an enzyme with a Split personality. *The International Journal of Biochemistry and Cell Biology*, 31(8), 817-836.

51. Martins, I.M., Matos, M., Costa, R., Silva, F., Pascoal, A., Estevinho, L.M., & Choupina, A.B. (2014). Transglutaminases: recent achievements and new sources. *Applied Microbiology and Biotechnology*, 98(16), 6957-6964.
52. Zhu, Y., Bol, J., Rinzema, A., Tramper, J., & Wijngaards, G. (1999). Transglutaminase as a potential tool in developing novel protein foods. *Agro Food Industry Hi-Tech*, 1(10), 8-10.
53. Ando, H., Adachi, M., Umeda, K., Matsuura, A., Nonaka, M., Uchio, R., ... & Motoki, M. (1989). Purification and characteristics of a novel transglutaminase derived from microorganisms. *Agricultural and Biological Chemistry*, 53(10), 2613-2617.
54. Carlos, H.P., & Carreño, F.G. (2007). Transglutaminasas: de importante papel fisiológico en los seres vivos al desarrollo novedoso de tecnología de alimentos. *Revista de la Sociedad Mexicana de Biotecnología y Bioingeniería*, 11(1), 21-27.
55. Nonaka, M., Tanaka, H., Okayama, A., Motoki, M., Ando, H., Umeda, K. & Matsuura, A., (1989). Polymerization of several proteins by calcium-independent transglutaminase derived from microorganisms. *Agricultural and Biological Chemistry*, 53(10), 2619-2623.
56. Takashiba, K., & Sakurai, K. (2016). Expression of proenzyme of microbial transglutaminase and its pressure-dependent unfolding behavior. *Memoirs of Institute of Advanced Technology, Kindai University*, 21, 11-22.
57. Shimba, N., Yokoyama, K.I., & Suzuki, E.I. (2002). NMR-based screening method for transglutaminases: rapid analysis of their substrate specificities and reaction rates. *Journal of Agricultural and Food Chemistry*, 50(6), 1330-1334.
58. Yang, M.T., Chang, C.H., Wang, J.M., Wu, T.K., Wang, Y.K., Chang, C.Y. & Li, T.T. (2011). Crystal structure and inhibition studies of transglutaminase from *Streptomyces mobaraense*. *Journal of Biological Chemistry*, 286(9), 7301-7307.
59. Grenard, P., Bates, M.K., & Aeschlimann, D. (2001). Evolution of transglutaminase genes: identification of a transglutaminase gene cluster on human chromosome 15q15. Structure of the gene encoding transglutaminase X and a novel gene family member, transglutaminase Z. *Journal of Biological Chemistry*, 276(35), 33066-33078.
60. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., .... & Gocayne, J.D. (2001). The sequence of the human genome. *Science*, 291(2207), 1304-1351.

61. Ichinose, A., Bottenus, R.E., & Davie, E.W. (1990). Structure of transglutaminase. *Journal of Biological Chemistry*, 265(23), 13411-13414.
62. Aeschlimann, D., & Paulsson, M. (1994). Transglutaminases: protein cross-linking enzymes in tissues and body fluids. *Thrombosis and Haemostasis*, 72(4), 402-15.
63. Mehta, K. (2005). Mammalian transglutaminases: a family portrait. *Progress in Experimental Tumor Research*, 38, 1-18.
64. Yee, V.C., Pedersen, L.C., Le, Trong., Bishop, P.D., Stenkamp, R.E., & Teller, D.C. (1994). Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII. *Proceedings the National Academy of Sciences*, 91(15), 7296-73000.
65. Ahvazi, B., Kim, H.C., Kee, S.H., Nemes, Z. & Steinert, P.M. (2002). Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. *The EMBO Journal*, 21(9), 2055-2067.
66. Stamnaes, J., Pinkas, D.M., Fleckenstein, B., Khosla, C., & Sollid, L.M. (2010). Redox regulation of transglutaminase 2 activity. *Journal of Biological Chemistry*, 285(33), 25402-25409.
67. Bergamini, C.M., Signorini, M., & Poltronieri, L. (1987). Inhibition of erythrocyte transglutaminase by GTP. *Biochimica et Biophysica Acta*, 916(1), 149-151.
68. Lee, K.N., Birckbichler, P.J., & Patterson, M.K. Jr. (1989). GTP hydrolysis by guinea pig liver transglutaminase. *Biochemical and Biophysical Research Communications*, 162(3), 1370-1375.
69. Lai, T.S., Slaughter, T.F., Peoples, K.A., Hettasch, J.M., & Greenberg, C.S. (1998). Regulation of human tissue transglutaminase function by magnesium-nucleotide complexes. Identification of distinct binding sites for Mg-GTP and Mg-ATP. *Journal of Biological Chemistry*, 273(3), 1776-1781.
70. Beninati, S., Piacentini, M., & Bergamini, C.M. (2017). Transglutaminase 2, a double face enzyme. *Amino Acids*, 49(3), 415-423.
71. Gundemir, S., Colak, G., Tucholski, J. & Johnson, G.V. (2012). Transglutaminase 2: a molecular Swiss army knife. *Biochimica et Biophysica Acta*, 1823(2), 406-419.
72. Begg, G.E., Carrington, L., Stokes, P.H., Matthews, J.M., Wouters, M.A., Husain, A., Lorand, L., Iismaa, S.E., & Graham, R.M. (2006). Mechanism of allosteric regulation of transglutaminase 2 by GTP. *Proceedings of National Academy of Sciences*, 103(52), 19683-19688.

73. Siegel, M., & Khosla, C. (2007). Transglutaminase 2 inhibitors and their therapeutic role in disease states. *Pharmacology & Therapeutics*, 115(2), 232-245.
74. Ichinose, A. (2012). Factor XIII is a key molecule at the intersection of coagulation and fibrinolysis as well as inflammation and infection control. *International Journal of Hematology*, 95(4), 362-370.
75. Sanchez-Pulido, L., & Ponting, C.P. (2016). Vaoshibins: new transglutaminase-like cysteine proteases possessing non-canonical Cys-His-Ser catalytic triad. *Bioinformatics Advance*, 32(10), 1441-1445.
76. Mossetti, U. (1987). Conjugated polyamines during dormancy and activation of tuber of Jerusalem artichoke. In *Conjugated Plant Hormones: Structure, Metabolism and Function*, 369-375.
77. Icekson, I., & Apelbaum, A. (1987). Evidence for transglutaminase activity in plant tissue. *Plant Physiology*, 84(4), 972-974.
78. Del Duca, S., & Serafini-Fracassini, D. (2005). Transglutaminases of higher, lower plants and fungi. *Progress in Experimental Tumor Research*, 38, 223-247.
79. Serafini-Fracassini, D., & Del Duca, S. (2008). Transglutaminases: widespread cross-linking enzymes in plants. *Annals of Botany*, 102(2), 145-152.
80. Kang, H., & Cho, Y.D. (1996). Purification and properties of transglutaminase from soybean (*Glycine max*) leaves. *Biochemical and Biophysical Research Communications*, 223(2), 288-292.
81. Del Duca, S., Creus, J.A., D'Orazi, D., Dondini, L., Bregoli, A.M., & Serafini-Fracassini, D., (2000a). Tuber vegetative stages and cell cycle in *Helianthus tuberosus*: protein pattern and their modification by spermidine. *Journal of Plant Physiology*, 156(1), 17-25.
82. Della Mea, M., Di Sandro, A., Dondini, L., Del Duca, S., Vantini, F., Bergamini, C., Bassi, R., & Serafini-Fracassini, D. (2004b). A *Zea mays* 39 kDa thylakoidal transglutaminase catalyses the modification by polyamines of light harvesting complex II in a light-dependent way. *Planta*, 219(5), 754-764.
83. Dondini, L., Bonazzi, S., Del Duca, S., Bregoli, A.M., & Serafini-Fracassini, D. (2001). Acclimation of chloroplast transglutaminase to high NaCl concentration in a polyamine-

- deficient variant strain of *Dunaliella salina* and in its wild type. *Journal of Plant Physiology* 158(2), 185-197.
84. Del Duca, S., Betti, L., Trebbi, G., Serafini-Fracassini, D., & Torrigiani, P. (2007). Transglutaminase activity changes during the hypersensitive reaction (HR), a typical defence response of tobacco NN plants to TMV. *Physiologia Plantarum*, 131(2), 241-250.
  85. Della Mea, M., De Filippis, F., Genovesi, V., Serafini-Fracassini, D., & Del Duca, S. (2007). The acropetal wave of developmental cell death (DCD) of tobacco corolla is preceded by activation of transglutaminase in different cell compartments. *Plant Physiology*, 144(2), 1211-1222.
  86. Iorio, R.A., Di Sandro, A., Scarpellini, A., Del Duca, S., Serafini-Fracassini, D., & Verderio, E. (2008). Visualisation of transglutaminase-mediated cross-linking activity in germinating pollen by laser confocal microscopy. *Plant Biosystems*, 142(2), 360-365.
  87. Sobieszczuk-Nowicka, E., Di Sandro, A., Del Duca, S., Serafini-Fracassini, D., & Legocka, J. (2007). Plastid membrane associated polyamines and thylakoid transglutaminases during etioplast to chloroplast transformation stimulated by kinetin. *Physiologia Plantarum* 130(4), 590-600.
  88. Sobieszczuk-Nowicka, E., Krzesłowska, M., & Legocka, J. (2008). Transglutaminases and their substrates in kinetin stimulated etioplast to chloroplast transformation in cucumber cotyledons. *Protoplasma*, 233(3-4), 187-194.
  89. Serafini-Fracassini, D., Del Duca, S., & D'Orazi, D. (1988). First evidence for polyamine conjugation mediated by an enzymic activity in plants. *Plant Physiology*, 87(3), 757-761.
  90. Signorini, M., Beninati, S., & Bergamini, C.M. (1991). Identification of transglutaminase activity in the leaves of Silver Beet (*Beta vulgaris* L.). *Journal of Plant Physiology*, 137(5), 547-552.
  91. Lilley, G.R., Skill, J., Griffin, M., & Bonner, P.L. (1998). Detection of Ca<sup>2+</sup>- dependent transglutaminase activity in root and leaf tissue of monocotyledonous and dicotyledonous plants. *Plant Physiology*, 117(3), 1115-1123.
  92. Del Duca, S., Bregoli, A.M., Bergamini, C., & Serafini-Fracassini, D. (1997). Transglutaminase catalyzed modification of cytoskeletal proteins by polyamines during the germination of *Malus domestica* pollen. *Sexual Plant Reproduction*, 10(2), 89-95.

93. Beninati, S., Iorio, R.A., Tasco, G., Serafini-Fracassini, D., Casadio, R., & Del Duca, S. (2013). Expression of different forms of transglutaminases by immature cells of *Helianthus tuberosus* sprout apices. *Amino Acids*, 44(1), 271-283.
94. Campos, N., Castañón, S., Urreta, I., Santos, M., & Torné, J.M. (2013). Rice transglutaminase gene: identification, protein expression, functionality, light dependence and specific cell location. *Plant Science*, 205, 97-110.
95. Suzuki, T., Park, H., Till, E.A., & Lennarz, W.J. (2001). The PUB domain: a putative protein-protein interaction domain implicated in the ubiquitin-proteasome pathway. *Biochemical and Biophysical Research Communications*, 287(5), 1083-1087.
96. Diepold, A., Li, G., Lennarz, W.J., Nurnberger, T., & Brunner, F. (2007). The Arabidopsis AtPNG1 gene encodes a peptide: N-glycanase. *The Plant Journal*, 52(1), 94-104.
97. Villalobos, E., Torné, J.M., Rigau, J., Ollés, I., Claparols, I., & Santos, M. (2001). Immunogold localization of a transglutaminase related to grana development in different maize cell types. *Protoplasma*, 216(3-4), 155-63.
98. Carvajal, P., Villalobos, E., Campos, A., Torne, J.M., Barbera, E., & Santos, M. (2006). Zea mays L. transglutaminase expression in *Escherichia coli*. *Microbial Cell Factories*, 5(1), P8.
99. Carvajal-Vallejos, P.K., Campos, A., Fuentes-Prior, P., Villalobos, E., Almeida, A.M., Barberá E., Torné, J.M. & Santos, M. (2007). Purification and in vitro refolding of maize chloroplast transglutaminase over-expressed in *Escherichia coli*. *Biotechnology Letters*, 29(8), 1255-1262.
100. Carvajal, P., Gibert, J., Campos, N., Lopera, O., Barberá, E., Torné, J.M., & Santos, M. (2011). Activity of maize transglutaminase over-expressed in *Escherichia coli* inclusion bodies: An alternative to protein refolding. *Biotechnology Progress*, 27(1), 232-240.
101. Cubiró, J.M.T., Lozano, M.A.S., Baro, D.T., Amador, E.V., & Lloveras, J.R. (2007). U.S. Patent No. 7,262,057. Washington, DC: U.S. Patent and Trademark Office.
102. Della Mea, M., Di Sandro, A., Dondini, L., Del Duca, S., Vantini, F., Bergamini, C., Bassi, R., & Serafini-Fracassini, D. (2004c). A Zea mays 39-kDa thylakoid transglutaminase catalyses the modification by polyamines of light-harvesting complex II in a light-dependent way. *Planta*, 219(5), 754-764.

103. Ioannidis, N.E., Lopera, O., Santos, M., Torné, J.M., & Kotzabasis, K. (2012). Role of plastid transglutaminase in LHCII polyamination and thylakoid electron and proton flow. *PLoS One*, 7(7), e41979.
104. Ioannidis, N.E., Ortigosa, S.M., Veramendi, J., Pintó-Marijuan, M., Fleck, I., Carvajal, P., Kotzabasis, K., Santos, M., & Torné, J.M. (2009). Remodelling of tobacco thylakoids by over-expression of maize plastidial transglutaminase. *Biochimica et Biophysica Acta* 1787(10), 1215-1222.
105. Campos, A., Carvajal-Vallejos, P.K., Villalobos, E., Franco, C.F., Almeida, A.M., Coelho, A.V., Torné, J.M. & Santos, M. (2010). Characterization of *Zea mays* L. plastidial transglutaminase: Interactions with thylakoid membrane proteins. *Plant Biology*, 12(5), 712-716.
106. Campos, N., Villalobos, E., Fontanet, P., Torné, J.M. & Santos, M. (2012). A peptide of 17 aminoacids from the N-terminal region of maize plastidial transglutaminase is essential for chloroplast targeting. *American Journal of Molecular Biology.*, 2(3), 245-257.
107. Villalobos, E. (2007). Study of maize (*Zea mays* L.) plant transglutaminases. Ph.D. Thesis, University of Barcelona, Barcelona.
108. Sobieszczuk-Nowicka, E., Zmienko, A., Samelak-Czajka, A., Łuczak, M., Pietrowska-Borek, M., Iorio, R., ... & Legocka, J. (2015). Dark-induced senescence of barley leaves involves activation of plastid transglutaminases. *Amino Acids*, 47(4), 825-838.
109. Soderlund, C., Descour, A., Kudrna, D., Bomhoff, M., Boyd, L., Currie, J., ... & Morrow, D. (2009). Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. *PLoS Genetics*, 5(11), e1000740.
110. Li, H., Zhang, L., Cui, Y., Luo, X., Xue, C., & Wang, S. (2013). Expression of soluble recombinant transglutaminase from *Zea mays* in *Pichia pastoris*. *World Journal of Microbiology and Biotechnology*, 29(5), 939-947.
111. Li, H., Cui, Y., Zhang, L., Yi, H., Han, X., Jiao, Y., ... & Zhang, S. (2014). Heterologous expression and purification of *Zea mays* transglutaminase in *Pichia pastoris*. *Food Science and Biotechnology*, 23(5), 1507-1513.
112. Finn, R.D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., & Eddy, S. R. (2015). HMMER web server: 2015 update. *Nucleic Acids Research*, 43(W1), W30-W38.

113. Emanuelsson, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2(4), 953.
114. Emanuelsson, O., Nielsen, H., & Von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*, 8(5), 978-984.
115. Pearson, W.R., & Lipman, D.J. (1988). Improved tools for biological sequence analysis. *Proceedings of National Academy Sciences*, 85(2444), 8.
116. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389-3402.
117. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., ... & Thompson, J.D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539.
118. Katoh, K., Misawa, K., Kuma, K.I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059-3066.
119. Letunic, I., Doerks, T., & Bork, P. (2014). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Research*, 43(D1), D257-D260.
120. Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., ... & Salazar, G.A. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 44(D1), D279-D285.
121. Pais, F.S.M., de Cássia Ruy, P., Oliveira, G., & Coimbra, R.S. (2014). Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology*, 9(1), 4.
122. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., & Barton, G.J. (2009). Jalview Version 2, a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191.
123. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., ... & Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37(suppl\_2), W202-W208.

124. Crooks, G.E., Hon, G., Chandonia, J.M., & Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Research*, 14(6), 1188-1190.
125. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A., & Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587.
126. McGuffin, L.J. (2008). Protein fold recognition and threading. In *Computational Structural Biology: Methods and Applications*. World Scientific London, 37-60.
127. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., & Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845.
128. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., & Ferrin, T.E. (2004). UCSF Chimera: a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612.
129. Eisenberg, D., Lüthy, R., & Bowie, J.U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. In *Methods in Enzymology*. Academic Press, 277(6364), 396-404.
130. Laskowski, R.A., MacArthur, M.W., Moss, D.S., & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283-291.
131. Seefeld, K., & Linder, E. (2007). *Statistics using R with biological examples*. Durham: University of New Hampshire.
132. Team, R. C. (2013). *R: A language and environment for statistical computing*. ISO 690.
133. Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(suppl\_1), D258-D261.
134. Cozzetto, D., Minneci, F., Carrant, H., & Jones, D.T. (2016). Ffpred 3: feature-based function prediction for all gene ontology domains. *Scientific Reports*, 6, 31865.
135. Bermúdez-Guzmán, M.D.J., Guzmán-González, S., Orozco-Santos, M., Velázquez-Monreal, J.J., Buenrostro-Nava, M.T., & Michel-López, C.Y. (2016). Optimización de un protocolo para aislamiento de DNA de hojas de *Saccharum officinarum*. *Revista Mexicana de Ciencias Agrícolas*, 7(4), 897-910.
136. Bimboim, H.C., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6), 1513-1523.

137. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Thompson, J.D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947-2948.
138. Citron, B.A., Suo, Z., SantaCruz, K., Davies, P.J., Qin, F., & Festoff, B.W. (2002). Protein crosslinking, tissue transglutaminase, alternative splicing and neurodegeneration. *Neurochemistry International*, 40(1), 69-78.
139. Lai, T.S., & Greenberg, C.S. (2013). Histaminylation of fibrinogen by tissue transglutaminase-2 (TGM-2): potential role in modulating inflammation. *Amino Acids*, 45(4), 857-864.
140. Facchiano, F., Facchiano, A., & Facchiano, A.M. (2006). The role of transglutaminase-2 and its substrates in human diseases. *Frontiers Bioscience*, 11(1758), 73.
141. Rauhavirta, T., Oittinen, M., Kivistö, R., Männistö, P.T., Garcia-Horsman, J.A., Wang, Z., & Lindfors, K. (2013). Are transglutaminase 2 inhibitors able to reduce gliadin-induced toxicity related to celiac disease? A proof-of-concept study. *Journal of Clinical Immunology*, 33(1), 134-142.
142. Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., ... & Costet, L. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications*, 9(1), 2638.
143. Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., ... & Minx, P. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science*, 326(5956), 1112-1115.
144. Alexandrov, N.N., Brover, V.V., Freidin, S., Troukhan, M.E., Tatarinova, T.V., Zhang, H., ... & Feldmann, K.A. (2009). Insights into corn genes derived from large-scale cDNA sequencing. *Plant Molecular Biology*, 69(1-2), 179.
145. Lynch, M., & Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, 290(5494), 1151-1155.
146. Graveley, B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics*, 17(2), 100-107.
147. Talavera, D., Vogel, C., Orozco, M., Teichmann, S.A., & De La Cruz, X. (2007). The (in) dependence of alternative splicing and gene duplication. *PLoS Computational Biology*, 3(3), e33.

148. Bush, S.J., Chen, L., Tovar-Corona, J.M., & Urrutia, A.O. (2017). Alternative splicing and the evolution of phenotypic novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1713), 20150474.
149. Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154(1), 459-473.
150. Su, Z., & Gu, X. (2012). Revisit on the evolutionary relationship between alternative splicing and gene duplication. *Gene*, 504(1), 102-106.
151. Margosiak, S.A., Dharma, A., Bruce-Carver, M.R., Gonzales, A.P., Louie, D., and Kuehn, G.D. (1990). Identification of the large subunit of ribulose 1, 5-bisphosphate carboxylase/oxygenase as a substrate for transglutaminase in *Medicago sativa* L.(alfalfa). *Plant Physiology*, 92(1), 88-96.
152. Del Duca, S., Tidu, V., Bassi, R., Serafini-Fracassini, D., & Esposito, C. (1994). Identification of transglutaminase activity and its substrates in isolated chloroplast of *Helianthus tuberosus*. *Planta*, 193, 283-289.
153. Bernet, E., Claparols, I., Dondini, L., Santos, M.A., Serafini-Fracassini, D., & Torné, J.M. (1999). Changes in polyamine content, arginine and ornithine decarboxylases and transglutaminase activities during light/dark phases (of initial differentiation) in maize calluses and their chloroplasts. *Plant Physiology and Biochemistry*, 37(12), 899-909.
154. Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., ... & Xue, Y. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics*, 31(20), 3359-3361.
155. Triant, D.A., & Pearson, W.R. (2015). Most partial domains in proteins are alignment and annotation artifacts. *Genome Biology*, 16(1), 99.
156. Kelley, L.A., & Sternberg, M.J. (2015). Partial protein domains: Evolutionary insights and bioinformatics challenges. *Genome biology*, 16(1), 100.
157. Bhushan, S., Kuhn, C., Berglund, A.K., Roth, C., & Glaser, E. (2006). The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. *FEBS Letters*, 580(16), 3966-3972.
158. Käll, L., Krogh, A., & Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5), 1027-1036.

159. Von Heune, G., & Gavel Y. (1988). Topogenic signals in integral membrane proteins. *European Journal of Biochemistry*, 174(4), 671-678.
160. Fujita, H., Yamagishi, M., Kida, Y., & Sakaguchi, M. (2011). Positive charges on the translocating polypeptide chain arrest movement through the translocon. *Journal of Cell Science*, 124(24), 4184-4193.
161. Coletta, A., Pinney, J.W., Solís, D.Y.W., Marsh, J., Pettifer, S.R., & Attwood, T.K. (2010). Low-complexity regions within protein sequences have position-dependent roles. *BMC Systems Biology*, 4(1), 43.
162. Wilkins, M., Zhang, N., & Schmid, J. (2018). Biological Roles of Protein-Coding Tandem Repeats in the Yeast *Candida Albicans*. *Journal of Fungi*, 4(3), 78.
163. Li, H., Zhang, L., Cui, Y., Luo, X., Xue, C., & Wang, S. (2013). Expression of soluble recombinant transglutaminase from *Zea mays* in *Pichia pastoris*. *World Journal of Microbiology and Biotechnology*, 29(5), 939-947.
164. Camire, R.M., & Bos, M.H.A. (2009). The molecular basis of factor V and VIII procofactor activation. *Journal of Thrombosis and Haemostasis*, 7(12), 1951-19.
165. Grimm, D.R., Colter, M.B., Braunschweig, M., Alexander, L.J., Neame, P.J., & Kim, H.K.W. (2001). Porcine factor V: cDNA cloning, gene mapping, three-dimensional protein modeling of membrane binding sites and comparative anatomy of domains. *Cellular and Molecular Life Sciences CMLS*, 58(1), 148-159.
166. Cripe, L.D., Moore, K.D., & Kane, W.H. (1992). Structure of the gene for human coagulation factor V. *Biochemistry*, 31(15), 3777-3785.
167. De Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., ... & Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, 34(suppl\_2), W362-W365.
168. Ramazzotti, M., Monsellier, E., Kamoun, C., Degl'Innocenti, D., & Melki, R. (2012). Polyglutamine repeats are associated to specific sequence biases that are conserved among eukaryotes. *PLoS One*, 7(2), e30824.

- 169.Kottenhagen, N., Gramzow, L., Horn, F., Pohl, M., & Theißen, G. (2012). Polyglutamine and polyalanine tracts are enriched in transcription factors of plants. In German Conference on Bioinformatics 2012. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 93-107.
- 170.Katiyar, S., Suzuki, T., Balgobin, B.J., & Lennarz, W.J. (2002). Site-directed mutagenesis study of yeast peptide: N-glycanase Insight into the reaction mechanism of deglycosylation. *Journal of Biological Chemistry*, 277(15), 12953-12959.
- 171.Nemes, Z., Csoz, E., Petrovski, G., & Fesus, L. (2005). Structure-function relationships of transglutaminases--a contemporary view. *Progress in Experimental Tumor Research*, 38, 19-36.
- 172.Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18(6), 292-298.
- 173.Flagel, L.E., & Wendel, J.F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*, 183(3), 557-564.
- 174.Van de Peer, Y., Maere, S., & Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, 10(10), 725.
- 175.Jiang, X., & Assis, R. (2019). Rapid functional divergence after small-scale gene duplication in grasses. *BMC Evolutionary Biology*, 19(1), 97.
- 176.Ohno, S. (2013). *Evolution by gene duplication*. Springer Science & Business Media.
- 177.Todd, A.E., Orengo, C.A., & Thornton, J.M. (2002a). Sequence and structural differences between enzyme and nonenzyme homologs. *Structure*, 10(10), 1435-1451.
- 178.Buller, A.R., & Townsend, C.A. (2013). Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proceedings of the National Academy of Sciences*, 110(8), E653-E661.
- 179.Todd, A.E., Orengo, C.A., & Thornton, J.M. (2002b). Plasticity of enzyme active sites. *Trends in Biochemical Sciences*, 27(8), 419-426.
- 180.McGuffin, L.J., Bryson, K., & Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404-405.
- 181.Jones, D.T., & Cozzetto, D. (2014). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*, 31(6), 857-863.

182. Latysheva, N.S., Flock, T., Weatheritt, R.J., Chavali, S., & Babu, M.M. (2015). How do disordered regions achieve comparable functions to structured domains?. *Protein Science*, 24(6), 909-922.
183. Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., & Jones, D.T. (2004). The DISORPRED server for the prediction of protein disorder. *Bioinformatics*, 20(13), 2138-2139.
184. Covarrubias, A.A., Cuevas-Velazquez, C.L., Romero-Pérez, P.S., Rendón-Luna, D.F., & Chater, C.C. (2017). Structural disorder in plant proteins: where plasticity meets sessility. *Cellular and Molecular Life Sciences*, 74(17), 3119-3147.
185. Romero, P.R., Zaidi, S., Fang, Y.Y., Uversky, V.N., Radivojac, P., Oldfield, C.J., ... & Dunker, A.K. (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proceedings of the National Academy of Sciences*, 103(22), 8390-8395.
186. Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. *Cell*, 126(1), 37-47.
187. Jangi, M., & Sharp, P.A. (2014). Building robust transcriptomes with master splicing factors. *Cell*, 159(3), 487-498.
188. Cuevas-Velázquez, C.L., & Covarrubias-Robles, A.A. (2011). Las proteínas desordenadas y su función: una nueva forma de ver la estructura de las proteínas y la respuesta de las plantas al estrés. *TIP. Revista Especializada en Ciencias Químico-Biológicas*, 14(2), 97-105.
189. Tompa, P. (2003). Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*, 25(9), 847-855.
190. Kajava, A.V. (2012). Tandem repeats in proteins: from sequence to structure. *Journal of Structural Biology*, 179(3), 279-288.
191. Ramachandran, G.T., & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. In *Advances in Protein Chemistry*. Academic Press, 23, 283-437.
192. Morris, A.L., MacArthur, M.W., Hutchinson, E.G., & Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Bioinformatics*, 12(4), 345-364.

193. Zhong, M., Wang, Y., Zhang, Y., Shu, S., Sun, J., & Guo, S. (2019). Overexpression of Transglutaminase from Cucumber in Tobacco Increases Salt Tolerance through Regulation of Photosynthesis. *International Journal of Molecular Sciences*, 20(4), 894.
194. Rao, R.U., & Mehta, K. (2004). Transglutaminases, thioredoxins and protein disulphide isomerase: diverse enzymes with a common goal of cross-linking proteins in lower organisms. *Indian Journal of Experimental Biology*, 42(3), 235-243.
195. Maiale, S., Sánchez, D.H., Guirado, A., Vidal, A., & Ruiz, O.A. (2004). Spermine accumulation under salt stress. *Journal of Plant Physiology*, 161(1), 35-42.
196. Dondini, L., Del Duca, S., Dall'Agata, L., Bassi, R., Gastaldelli, M., Della Mea, M., ... & Serafini-Fracassini, D. (2003). Suborganellar localisation and effect of light on *Helianthus tuberosus* chloroplast transglutaminases and their substrates. *Planta*, 217(1), 84-95.
197. Duan, J., Li, J., Guo, S., & Kang, Y. (2008). Exogenous spermidine affects polyamine metabolism in salinity-stressed *Cucumis sativus* roots and enhances short-term salinity tolerance. *Journal of Plant Physiology*, 165(15), 1620-1635.
198. Sang, Q.; Shan, X.; An, Y.; Shu, S.; Sun, J.; Guo, S. (2017). Proteomic analysis reveals the positive effect of exogenous spermidine in tomato seedlings' response to high-temperature stress. *Frontiers of Plant Sciences*, 8, 120.
199. Schaper, E., & Anisimova, M. (2015). The evolution and function of protein tandem repeats in plants. *New Phytologist*, 206(1), 397-410.
200. Radivojac, P., Clark, W.T., Oron, T.R., Schnoes, A.M., Wittkop, T., Sokolov, A., ... & Pandey, G. (2013). A large-scale evaluation of computational protein function prediction. *Nature Methods*, 10(3), 221.
201. McElheny, D., Schnell, J.R., Lansing, J.C., Dyson, H.J., & Wright, P.E. (2005). Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proceedings of National Academy of Science*, 102(14), 5032-5037.
202. Schulenburg, C., & Hilvert, D. (2013). Protein conformational disorder and enzyme catalysis. *Dynamics in Enzyme Catalysis*, 337, 41-67.
203. Piacentini, M., D'Eletto, M., Farrace, M.G., Rodolfo, C., Del Nonno, F., Ippolito, G., & Falasca, L. (2014). Characterization of distinct sub-cellular location of transglutaminase type II:

- changes in intracellular distribution in physiological and pathological states. *Cell and Tissue Research*, 358(3), 793-805.
204. Liu, S., Cerione, R.A., & Clardy, J. (2002). Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. *Proceedings of the National Academy of Sciences*, 99(5), 2743-2747.
205. Zanetti, L., Ristoratore, F., Bertoni, A., & Cariello, L. (2004). Characterization of sea urchin transglutaminase, a protein regulated by guanine/adenine nucleotides. *Journal of Biological Chemistry*, 279(47), 49289-49297.
206. Mitra, S., Ray, S.K., & Banerjee, R. (2016). Synonymous codons influencing gene expression in organisms. *Research and Reports in Biochemistry*, 6, 57-65.
207. Kawabe, A., & Miyashita, N.T. (2003). Patterns of codon usage bias in three dicot and four monocot plant species. *Genes & Genetic Systems*, 78(5), 343-352.
208. Ripley, L.S. (1990). Frameshift mutation: determinants of specificity. *Annual Review of Genetics*, 24(1), 189-211.

## 11. ANEXOS

### 11.1 Preparación de células competentes

**Preparación de Medio LB:** 0.3% Extracto de levadura, 1% Peptona, 0.5% NaCl, 1.3% Agar bacteriológico, 0.1 M CaCl<sub>2</sub> estéril, 0.1 M CaCl<sub>2</sub> con 10% de glicerol.

**Procedimiento:** Se cultivó *E. coli* (DH5 $\alpha$ ) en medio LB sólido a 37 °C durante toda la noche. Se tomó una colonia aislada y se sembró en 3 mL de medio de LB líquido, se incubó con agitación a 37 °C durante toda la noche. Posteriormente se sembró 1 mL del cultivo crecido en 10 mL de medio LB líquido y se incubó con agitación a 37 °C durante 3 h aproximadamente (NOTA: Observar cierta turbidez en el medio). En seguida se dividió el cultivo en 10 microtubos de 1.5 mL y se incubó en hielo durante 10 min, después se centrifugaron a 12 000 rpm durante 15 s (12k/5s) y se eliminó el sobrenadante en solución con cloro. La pastilla se resuspendió muy suavemente con 1 mL de CaCl<sub>2</sub> estéril y frío. Los tubos fueron incubados en hielo por 10 min y centrifugados a 12k/15s, nuevamente se eliminó el sobrenadante y se resuspendió la pastilla muy suavemente con 30  $\mu$ L de CaCl<sub>2</sub>. Finalmente se almacenó hasta su uso a -70 °C.

### 11.2 Transformación de células competentes

**Procedimiento:** Para la transformación de células competentes se añadieron a 30  $\mu$ L de células competentes 3  $\mu$ L de plásmido y se mezclaron suavemente. La muestra fue incubada en hielo durante 20 min y en seguida se calentó a 42 °C por 90 s, rápidamente se transfirieron a hielo para una incubación durante 5 min. Después se le añadieron 200  $\mu$ L de LB líquido y se incubaron a 37 °C durante 45 min con agitación. Por último, se esparció la mezcla obtenida con las bacterias en una caja de medio sólido LB, adicionándole antibiótico según el plásmido (ampicilina 50 mg/L) y permitiendo que el líquido fuera absorbido. Las cajas son invertidas e incubadas a 37 °C toda la noche.

### 11.3 Extracción de plásmidos (MINIPREP Bimboim Daily)

#### Soluciones:

- Bimboim I. 25mM Tris-HCl pH 8, 10mM EDTA pH 8, 50mM glucosa (esterilizar y almacenar a 4 °C).

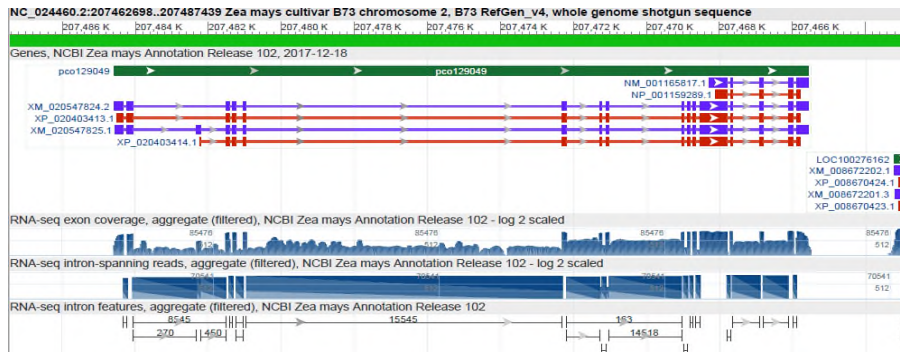
- Bimboim II. NaOH 0.2N, SDS 1%. Preparar al momento o guardar hasta una semana en temperatura ambiente. Para 1 mL de solución, mezclar 20  $\mu$ L de 10 N NaOH con 880  $\mu$ L de H<sub>2</sub>O y añadir 100  $\mu$ L de 10% SDS.
- Bimboim III. 60 mL 5M Acetato de potasio, 11.5 mL Ácido acético glacial, 28.5 mL H<sub>2</sub>O.
- Fenol-Cloroformo (1:1), Cloroformo-Alcohol Isoamílico (24:1), Isopropanol, Etanol 70%, H<sub>2</sub>O destilada estéril, ARNasa a 10 mg/ml.

**Procedimiento:** La extracción de ADN plasmídico se realizó mediante el protocolo propuesto por Bimboim y Doly (1979). En primera instancia se cultivaron bacterias *E. coli* previamente transformadas a 37 °C durante 16-20 hrs en medio LB líquido con ampicilina. Se tomaron 1.5 mL de cultivo bacteriano en microtubos y se centrifugan a 12 000 rpm durante 2 min. (12 K/2 min), se eliminó el sobrenadante en solución de hipoclorito de sodio y se añadieron 100  $\mu$ L de solución Bimboim I a la pastilla bacteriana, se resuspendió con ayuda de la micropipeta y se incubó en hielo por 5 min. Después se añadieron 200  $\mu$ L de solución Bimboim II, se mezcló perfectamente por inversión y se incubó a temperatura ambiente por 5 min, enseguida se agregaron 150  $\mu$ L de solución Bimboim III y se mezcló vigorosamente para después ser incubadas en hielo durante 5 min. Posteriormente se centrifugó a 12 K/5 min y se transfirió el sobrenadante a un tubo limpio, se le añadió 200  $\mu$ L de fenol:cloroformo, se mezcló vigorosamente hasta emulsificar y se centrifugó a 12 000 rpm durante 5 min (12 K / 5 min). Luego se recuperó la fase acuosa (superior) en un tubo limpio y se le adicionó 200  $\mu$ L de cloroformo-alcohol isoamílico, se agitó hasta emulsificar y se centrifugó a 12 K/ 5 min. Se recuperó la fase acuosa en un tubo limpio y se le agregó un volumen igual de isopropanol. Se mezcló por inversión durante 1 min. y se incubó en hielo por 1 min. Posteriormente se recuperó la pastilla por centrifugación a 12 K/5 min y se eliminó el sobrenadante. En seguida le fue agregado 200  $\mu$ L de etanol al 70% y se mezcló por inversión. De nueva cuenta se centrifugó a 12 K/ 5 min, se eliminó el sobrenadante y se puso a sacar la pastilla. Finalmente fue resuspendida en 50  $\mu$ L de H<sub>2</sub>O destilada estéril y se almacenó a 4 °C.

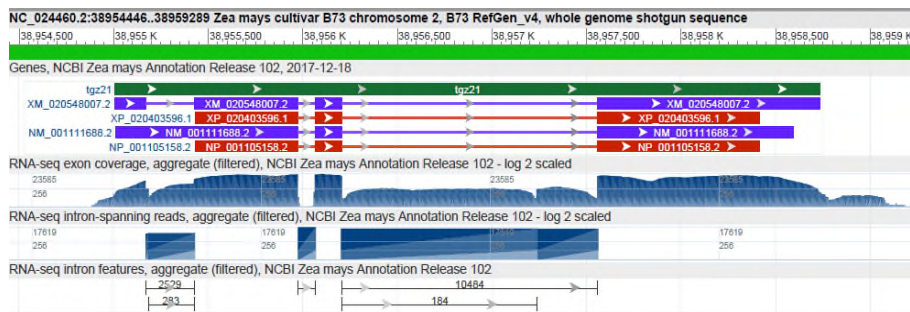
## 11.4 Contexto genómico de los genes anotados con el término transglutaminase de maíz

Los siguientes diagramas muestran secciones del gen y secuencias similares a ellas, junto con la distribución de los intrones y exones del gen. Las secuencias del tipo RefSeq son utilizadas para representar productos de ARN y sus regiones codificantes.

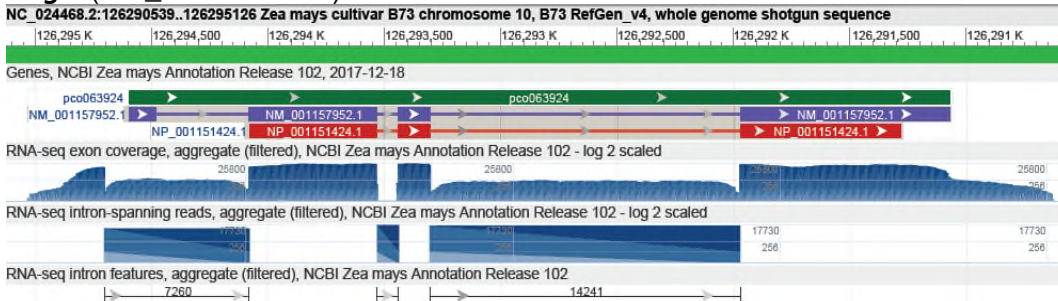
### >*tgzc* (NM\_001361416.1)



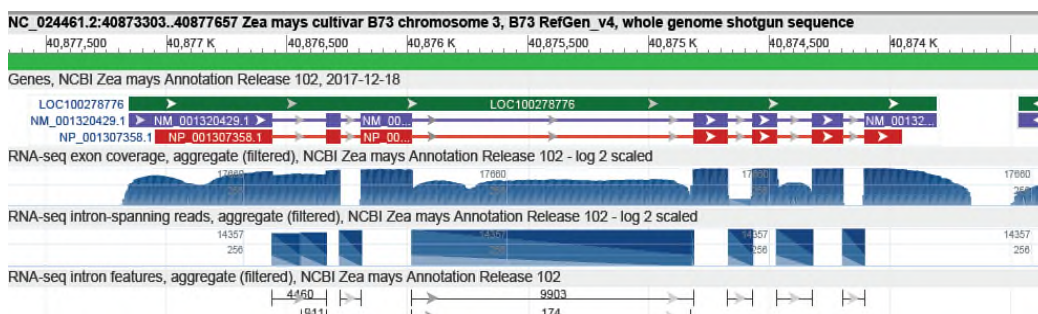
### >*tgzi* (NM\_001111688.2)



### >*tgzi* (NM\_001254887.2)



### >*tgzc2* (NM\_001320429.1)





Logo consenso TRs en TGZ15 (Q6KF70)  
15 copias ubicadas de la pos. 275-410



Logo consenso TRs de sorgo (XP\_021319478.1)  
8 copias ubicadas de la pos. 274-330



Logo consenso TRs en TGO (M1VT12)  
2 copias ubicadas de la pos. 276-300



Logo consenso TRs en *Hordeum v.* (BAJ91197.1)  
8 copias ubicadas de la pos. 13-96



## 11.7 Secuencias de proteínas referencia utilizadas en el análisis bioinformático

### >Consenso\_N-terminal/1-249\_(B8A2N4\_C0HGH1\_Q6KF70)

MAHRGHL DGLTGQAPALMRHGSFAAGSLSSRSPLQSSSTLEMLENKLAMQTTEVEKLITENQRLASSHVLR  
 QDIVDTEKEMQMIRTHLGEVQTETDLQIRDLLERIRLMEVDIHSNGNVNKKELHQMHEAKRLITERQMLTLE  
 IEDVTKELQKLSASGDNKSLPELLSELDRLRKEHHNLRSQFEFEKNTNVKQVEQMRTMEMNLITMTKQAEKL  
 RVDVANAERRAQAAAAQAAHAAGAQTASQPG

### >Cterminal\_TGZi/250-513

TAQATAVPAAAATDPYAGAYASYPAYQQGAQAGAYQQGAQAGTYQQGAGTGAYQQGAQAGAYQQGG  
 QDGAYQQGAQAGGYQQGAQAGAYQQGAGAGAYQQGAQAGAYQQGAQAGAYQQGAQAGAYQQGA  
 QAGTYQQGNQAGAYQQGAQDGAYQQGAQAGAYPYAYDAATAYTYAGYSGYPVAGYAQS AVPNYSYAAP  
 PQPISSGAATDAASGLYGAVGSAGYPTAQVQPSSGPANAAQAPPPPPPAAPYPPSTYDQTRGAQR

### >Cterminal\_TGZ15/250-534\_(Q6KF70)

QLKLPRFQQQPQTHMQVHIPATPLHISREPRLGHISRVLRLGYSREPRLGHISRPRLGHISRGARMGHISR  
 GLRLGHISRPRLGHISRPRLGHISRVLRLGHSREPRLGHISRGPSLGHISRGPRLGHISRPRMGHISRPRM  
 GHISRVLRLLEHTTMLMMLARLMHMVTLAIQLQATRQVQCPTIPMLHLRSQQA AVQLRTPQEASMGQLV  
 VLDILLGKFSRAVALQMQRKHLHHRQHIPPAMTKPEPRDKIWDVNQMDVCHALLSRQIW

### >A0A1D6EW30\_TGZc/1-725

MVARRFVVRHGAAGGSAEAAEEQQHEVEYDTEHGLDVLRLQIFSLTAVPPDLQKIVVEADGSVDDGTDL  
 EAVSERLRLLAIGEEGEDDGAARAQEKSD EEFARMLQAE EEAALLQQYNIQNDGGDVFRQRVEPYMHQVL  
 MYEDSVGQEA AQKTPICELEEKALVSLAKEGNFNPSKDEEKHAFLLQLLFWFKQSFRWVNAAPCDSCGRET  
 SNVGMGTPLTSEIKFGASRVEMYRCNHCSITRFPYNDPRKLIQTRRGRCGEWANCFTFYCRAFGYDARLIL  
 DFTDHVWTECFSNLYGRWMHLDPCGVYDNPLLYEKGWNNKLNIVIAISKDGV RDVTKRYTRKWHEVLSR

RIITSEETVSVLSSITGKYRTGLSPDALSVIENREKEESEELRKSAYLQVETALSLPGRLSGSVEWRKARSELGQA  
 DSLSCSSCPIRKCVDPGHVSNIYDALSFLSHFFDKKIPKERIIEAFEALKTLMMLTKSSNFKRSATLDQKAHHLFE  
 EIFPSIERLLSAISLKAELGTAGYQSVTVSGNPVHSSLALPVALDAVNEILSNYKNNTFFTEGNHFPRGNRVCSGS  
 VLASREQLPIGIATAAFDGIHSSKWEEDGAKGCWLIYTMQDGKSCELESYDLMSANDAPERDPMDWVLEG  
 SADGGSTWNILDARNSEMFESRFLRKTFTVDKRYKANAFRFRFLRVRDSSHNSNPRFQIGSIDLYGKTI

**>COPED4\_TGZc2/1-420**

MNCAVAWSPSSALVHRPGRVRRWRSSVRPCAAAVEQEGGAAGEAPRLVLHDSLDSVGVATAHARAARE  
 GFVAQVGRLTRVSAGSSIAISRGPDLARAALCVAEEDSLVSHSSVPLPVDAFIARLDDLSTGFCAGGFPPSGA  
 PPEVFFDYLNRYLYVHKGFRRRTNGVSDVRIMYLHSLVLTCSRGSALMLALIYSEILKTVRTYGLLDFDEIFFPNDLN  
 SLPRGYDKQKSKLVDEPHIMTSKLLVETLRTLKCAFWPQSDQSSSLFLNAVAANQRGPGTGDNQARSYGN  
 TSAIEMAAAKAAQHRLMRGVWTNARFGDMRRALAACERLILLHHPHELRYAALLYCGYYEDCLHYLSLY  
 QTAKSGKSPTNRLEILEDEAVNTLKARLMLILAEDGWSRCRPVARYWTKNSEPW

**>TGZc2\_Transglut\_core2/101-204\_PF13369**

CVAEEDSLVSHSSVPLPVDAFIARLDDLSTGFCAGGNFPPSGAPPEVFFDYLNRYLYVHKGFRRRTNGVSDVRI  
 MYLHSLVLTCSRGSALMLALIYSEILKTVRT

**>TGZc\_TGc/257-312\_SM00460**

LIQTRRGRCGEWANCFTFYCRAFGYDARLILDFTDHWTECFSNLYGRWMHLDPC

**>TGZ15\_núcleo\_extendido/431-534**

LQATRQVQCPTIPMLHLRSQQAQVLRTPQEASMGQLVVDILLGKFSRAVALQMQRKHLHHRQHHP  
 PAHMTKPEEPRDKIWDVNQMDVCH

**11.8 Tabla de las secuencias homólogas a TGZ15 y TGZi**

Entrada	Estado	Organismo	Longitud (aa)	Existencia Proteína
I1IYV7	no revisado	<i>Brachypodium distachyon</i>	420	Predicha
F2DBR8	no revisado	<i>Hordeum vulgare</i> subsp. vulgare	467	A nivel Transcripto
M7ZS62	no revisado	<i>Triticum urartu</i>	443	Predicha
AOA1D6B4D9	no revisado	<i>Triticum aestivum</i>	459	Predicha
AOA0D9W692	no revisado	<i>Leersia perrieri</i>	409	Predicha
J3LYX2	no revisado	<i>Oryza brachyantha</i>	407	Predicha
I1PM96	no revisado	<i>Oryza glaberrima</i>	406	Predicha
Q01JY9	no revisado	<i>Oryza sativa</i>	406	Predicha
Q7XK24	no revisado	<i>Oryza sativa</i> subsp. japonica	406	A nivel Transcripto

Entrada	Estado	Organismo	Longitud (aa)	Existencia Proteína
A2XUN9	no revisado	<i>Oryza sativa</i> subsp. indica	406	Predicha
K3Y711	no revisado	<i>Setaria italica</i>	429	Predicha
A0A2S317K8	no revisado	<i>Panicum hallii</i>	421	Predicha
B8A2N4	no revisado	<i>Zea mays</i>	513	A nivel Transcripto
A0A1Z5RDI6	no revisado	<i>Sorghum bicolor</i>	442	Predicha
C0HGH1	no revisado	<i>Zea mays</i>	458	A nivel Transcripto
A0A1E5VQ64	no revisado	<i>Dichanthelium oligosanthos</i>	447	Predicha
M1VT12	no revisado	<i>Oryza sativa</i>	352	A nivel Transcripto
Q6KF70	no revisado	<i>Zea mays</i>	534	A nivel Transcripto

### 11.9 Tabla de las secuencias homólogas a TGZc

Entrada	Estado	Organismo	Longitud	Existencia Proteína
A0A1D6EW30	no revisado	<i>Zea mays</i>	725	Evidencia a nivel transcripto
C5X988	no revisado	<i>Sorghum bicolor</i>	727	Predicha
A0A1W0W6N5	no revisado	<i>Sorghum bicolor</i>	622	Predicha
K3ZR30	no revisado	<i>Setaria italica</i>	730	Predicha
A0A2S3H2J7	no revisado	<i>Panicum hallii</i>	733	Predicha
J3ML25	no revisado	<i>Oryza brachyantha</i>	787	Predicha
I1QAU9	no revisado	<i>Oryza glaberrima</i>	775	Predicha
B8B6D9	no revisado	<i>Oryza sativa</i> subsp. indica	802	Predicha
Q7FOR1	revisado	<i>Oryza sativa</i> subsp. japonica	447	Evidencia a nivel transcripto
Q9FGY9	revisado	<i>Arabidopsis thaliana</i>	721	Evidencia a nivel transcripto
F2E1W5	no revisado	<i>Hordeum vulgare</i> subsp. vulgare	777	Evidencia a nivel transcripto
A0A1U8LF87	no revisado	<i>Gossypium hirsutum</i>	740	Predicha
A5PHD1	no revisado	<i>Arabidopsis thaliana</i>	721	Evidencia a nivel transcripto
M0SU82	no revisado	<i>Musa acuminata</i> subsp. malaccensis	713	Predicha
A0A2H3YHR7	no revisado	<i>Phoenix dactylifera</i>	714	Predicha
A0A199VP42	no revisado	<i>Ananas comosus</i>	1205	Predicha
A0A0D9WZ36	no revisado	<i>Leersia perrieri</i>	703	Predicha
A0A1D1XNE9	no revisado	<i>Anthurium amnicola</i>	716	Predicha
Q96IV0	revisado	<i>Homo sapiens</i>	654	Evidencia a nivel proteína
Q4R6F3	revisado	<i>Macaca fascicularis</i>	654	Evidencia a nivel transcripto
Q9JI78	revisado	<i>Mus musculus</i>	651	Evidencia a nivel proteína
Q5XI55	revisado	<i>Rattus norvegicus</i>	651	Evidencia a nivel transcripto
Q5ZJM3	revisado	<i>Gallus gallus</i>	651	Evidencia a nivel transcripto
E1BKB2	no revisado	<i>Bos taurus</i>	657	Predicha
A0A151MR26	no revisado	<i>Alligator mississippiensis</i>	649	Predicha
I1GU77	no revisado	<i>Brachypodium distachyon</i>	719	Predicha

### 11.10 Tabla de las secuencias homólogas a TGZc2

Entrada	Estado	Organismo	Longitud	Existencia Proteína
C0PED4	no revisado	<i>Zea mays</i>	420	Evidencia a nivel transcripto
C5XGM9	no revisado	<i>Sorghum bicolor</i>	428	Predicha
K3XIB2	no revisado	<i>Setaria italica</i>	413	Predicha
I1HEE0	no revisado	<i>Brachypodium distachyon</i>	427	Predicha
M0WM04	no revisado	<i>Hordeum vulgare</i> subsp. vulgare	427	Predicha
B8ACM8	no revisado	<i>Oryza sativa</i> subsp. indica	415	Predicha
A0A0D9Y6M4	no revisado	<i>Oryza glumipatula</i>	430	Predicha
D7UBN4	no revisado	<i>Vitis vinifera</i>	423	Predicha
A0A0B2S4E3	no revisado	<i>Glycine soja</i>	436	Predicha
K4B028	no revisado	<i>Solanum lycopersicum</i>	423	Predicha
B9GM02	no revisado	<i>Populus trichocarpa</i>	426	Predicha
B9GWY6	no revisado	<i>Populus trichocarpa</i>	423	Predicha
A0A2K1JSY7	no revisado	<i>Physcomitrella patens</i> subsp. patens	501	Predicha
M0TP15	no revisado	<i>Musa acuminata</i> subsp. malaccensis	540	Predicha

11.11 Alineamiento de motivos similares al patrón de expresión: H-[IT]-[TS]-[LTR]-[SGC]-P-[DTS]-[LS]-[SG]-H a partir de proteínas similares a TGZ15



tr F1RPW2 F1RPW2	H T T L S P D L G H	tr A0A214CXP0 A0A214CXP0	H I S T C P T S G H
tr F1RPW2 F1RPW2_2	H T T L S P D L S H	tr A0A214CXP0 A0A214CXP0_2	H I S T C P T S G H
tr F1RPW2 F1RPW2_3	H T T L S P D L G H	tr A0A151ZEZ0 A0A151ZEZ0	H T T T G P S S S H
tr F1RPW2 F1RPW2_4	H T T L S P D L G H	tr A0A151ZEZ0 A0A151ZEZ0_2	H T T T G P S S S H
tr F1RPW2 F1RPW2_5	H T T L S P D L G H	tr A0A1Q5SJM9 A0A1Q5SJM9	H I T T G P D S S H
tr F1RPW2 F1RPW2_6	H T T L S P D L G H	tr A0A214CXX8 A0A214CXX8	H I S T C P T S G H
tr F1RPW2 F1RPW2_7	H T T L S P D L S H	tr A0A214CXX8 A0A214CXX8_2	H I S T C P T S G H
tr F1RPW2 F1RPW2_8	H T T L S P D L S H	tr A0A060XNQ3 A0A060XNQ3	H T T T S P T S S H
tr F1RPW2 F1RPW2_9	H T T L S P D L G H	tr A0A1Q5SAS4 A0A1Q5SAS4	H I T T G P D S S H
tr F1RPW2 F1RPW2_10	H T T L S P D L G H	tr A0A2K5TJA4 A0A2K5TJA4	H T T L S P D L S H
tr F1RPW2 F1RPW2_11	H T T L S P D L G H	tr A0A2K5TJA4 A0A2K5TJA4_2	H T T L S P D L S H
tr A0A2K5M1W6 A0A2K5M1W6	H T T L S P D L S H	tr A0A2K5TJA4 A0A2K5TJA4_3	H T T L S P D L S H
sp Q9GLP1 FA5_PIG	H T T L S P D L G H	tr A0A2K5TJA4 A0A2K5TJA4_4	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_2	H T T L S P D L S H	tr A0A2J8SAU8 A0A2J8SAU8	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_3	H T T L S P D L G H	tr A0A2J8SAU8 A0A2J8SAU8_2	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_4	H T T L S P D L S H	tr A0A2K5M1S0 A0A2K5M1S0	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_5	H T T L S P D L G H	tr A0A213MVE4 A0A213MVE4	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_6	H T T L S P D L G H	tr A0A213MVE4 A0A213MVE4_2	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_7	H T T L S P D L G H	tr A0A213MVE4 A0A213MVE4_3	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_8	H T T L S P D L S H	tr A0A213MVE4 A0A213MVE4_4	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_9	H T T L S P D L S H	tr A0A213MVE4 A0A213MVE4_5	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_10	H T T L S P D L G H	tr A0A213MVE4 A0A213MVE4_6	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_11	H T T L S P D L G H	tr A0A1D5QJP4 A0A1D5QJP4	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_12	H T T L S P D L S H	tr A0A1D5QJP4 A0A1D5QJP4_2	H T T L S P D L S H
sp Q9GLP1 FA5_PIG_13	H T T L S P D L G H	tr A0A1D5QJP4 A0A1D5QJP4_3	H T T L S P D L S H
tr A0A2K6Q6W6 A0A2K6Q6W6	H T T L S P D L S H	tr A0A1D5QJP4 A0A1D5QJP4_4	H T T L S P D L S H
tr G1RYA9 G1RYA9	H T T L S P D L S H	tr H2N4S8 H2N4S8	H T T L S P D L S H
tr G1RYA9 G1RYA9_2	H T T L S P D L S H	tr H2N4S8 H2N4S8_2	H T T L S P D L S H
tr G1RYA9 G1RYA9_3	H T T L S P D L S H	tr A0A2J8LLQ9 A0A2J8LLQ9	H T T L S P D L S H
tr G1RYA9 G1RYA9_4	H T T L S P D L S H	tr A0A2K6Q6W9 A0A2K6Q6W9	H T T L S P D L S H
tr G1RYA9 G1RYA9_5	H T T L S P D L S H	tr A0A2K6Q6W9 A0A2K6Q6W9_2	H T T L S P D L S H
tr G3R571 G3R571	H T T L S P D L S H	tr A0A2K6Q6W9 A0A2K6Q6W9_3	H T T L S P D L S H
tr G3R571 G3R571_2	H T T L S P D L S H	tr A0A2K6Q6W9 A0A2K6Q6W9_4	H T T L S P D L S H
tr G3R571 G3R571_3	H T T L S P D L S H	tr A0A2K6Q6W9 A0A2K6Q6W9_5	H T T L S P D L S H
tr G3R571 G3R571_4	H T T L S P D L S H	tr A0A2J8LLR6 A0A2J8LLR6	H T T L S P D L S H
tr A0A2K6C912 A0A2K6C912	H T T L S P D L S H	tr G7MFA4 G7MFA4	H T T L S P D L S H
tr A0A096NPX1 A0A096NPX1	H T T L S P D L S H	tr G7MFA4 G7MFA4_2	H T T L S P D L S H
tr A0A096NPX1 A0A096NPX1_2	H T T L S P D L S H	tr G7MFA4 G7MFA4_3	H T T L S P D L S H
tr A0A096NPX1 A0A096NPX1_3	H T T L S P D L S H	tr G7MFA4 G7MFA4_4	H T T L S P D L S H
tr A0A096NPX1 A0A096NPX1_4	H T T L S P D L S H	tr G7MFA4 G7MFA4_5	H T T L S P D L S H
tr A0A096NPX1 A0A096NPX1_5	H T S L S P D L S H	tr A0A0D9RFF4 A0A0D9RFF4	H T T L S P D L S H
tr A0A096NPX1 A0A096NPX1_6	H T T L S P D L S H	tr A0A2K6KJU6 A0A2K6KJU6	H T T L S P D L S H
tr A0A2K5YD51 A0A2K5YD51	H T T L S P D L S H	tr A0A2K6KJU6 A0A2K6KJU6_2	H T T L S P D L S H
tr A0A2K5YD51 A0A2K5YD51_2	H T T L S P D L S H	tr A0A2K6KJU6 A0A2K6KJU6_3	H T T L S P D L S H
tr A0A287Bj70 A0A287Bj70	H T T L S P D L G H	sp P12259 FA5_HUMAN	H T T L S P D L S H
tr A0A287Bj70 A0A287Bj70_2	H T T L S P D L S H	tr H2QZV8 H2QZV8	H T T L S P D L S H
tr A0A287Bj70 A0A287Bj70_3	H T T L S P D L G H	tr A0A0A0MRJ7 A0A0A0MRJ7	H T T L S P D L S H
tr A0A287Bj70 A0A287Bj70_4	H T T L S P D L G H		
tr A0A287Bj70 A0A287Bj70_5	H T T L S P D L G H		
tr A0A287Bj70 A0A287Bj70_6	H T T L S P D L G H		
tr A0A287Bj70 A0A287Bj70_7	H T T L S P D L S H		
tr A0A287Bj70 A0A287Bj70_8	H T T L S P D L S H		
tr A0A287Bj70 A0A287Bj70_9	H T T L S P D L G H		
tr A0A287Bj70 A0A287Bj70_10	H T T L S P D L G H		
tr A0A287Bj70 A0A287Bj70_11	H T T L S P D L G H		
tr A0A2R98MQ2 A0A2R98MQ2	H T T L S P D L S H		
tr A0A2J8SAV1 A0A2J8SAV1	H T T L S P D L S H		
tr A0A2J8SAV1 A0A2J8SAV1_2	H T T L S P D L S H		
tr Q6KF70 Q6KF70	H I S R G P S L G H		
tr Q6KF61 Q6KF61	H I S R G P S L G H		

11.12 Alineamiento de motivos similares al patrón de expresión: [HY]-I-S-R-[EGV]-[PLA]-[RS]-[LM]-[GE] proveniente de los TRs de la proteína TGZ15



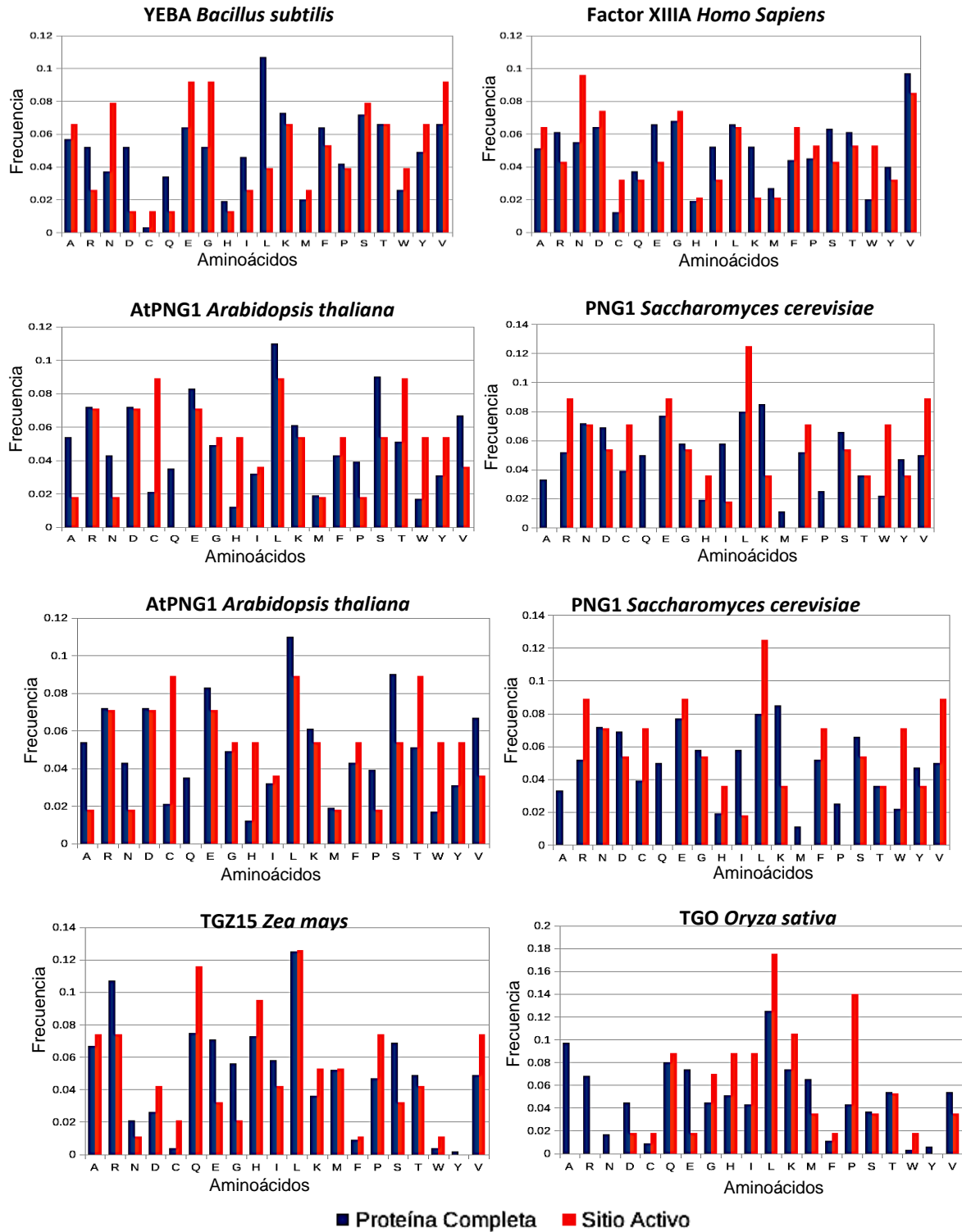
tr A0A1B5D0X9 A0A1B5D0X9	Y	I	S	R	V	L	R	M	G	tr A0A2U0WAY6 A0A2U0WAY6	Y	I	S	R	V	L	R	M	G
tr A0A1D6E8F5 A0A1D6E8F5	H	I	S	R	E	P	R	L	E	tr A0A219AA10 A0A219AA10	Y	I	S	R	V	L	R	M	G
tr A0A370XXG5 A0A370XXG5	Y	I	S	R	V	L	R	M	G	tr A0A183CZV4 A0A183CZV4	Y	I	S	R	V	L	R	M	G
tr A0A1B7LFH2 A0A1B7LFH2	H	I	S	R	V	L	R	L	E	tr A0A0W0HDR8 A0A0W0HDR8	Y	I	S	R	V	L	R	M	G
tr A0A023CC77 A0A023CC77	Y	I	S	R	V	L	R	M	G	tr A0A2T6LK03 A0A2T6LK03	Y	I	S	R	V	L	R	M	G
tr A0A1C5T7K9 A0A1C5T7K9	H	I	S	R	V	L	R	L	E	tr A0A1G4VRL1 A0A1G4VRL1	Y	I	S	R	V	L	R	M	G
tr A0A178BX04 A0A178BX04	H	I	S	R	G	P	S	L	E	tr A0A0Q8CTM8 A0A0Q8CTM8	Y	I	S	R	V	L	R	M	G
tr A0A2M9GEL9 A0A2M9GEL9	Y	I	S	R	V	L	R	M	G	tr A0A1D3K5J1 A0A1D3K5J1	Y	I	S	R	V	L	R	M	G
tr A0A1B5FF86 A0A1B5FF86	Y	I	S	R	V	L	R	M	G	tr A0A223VJ83 A0A223VJ83	Y	I	S	R	V	L	R	M	G
tr A0A1I0YUW6 A0A1I0YUW6	Y	I	S	R	V	L	R	M	G	tr A0A2N7Z8Y2 A0A2N7Z8Y2	Y	I	S	R	V	L	R	M	G
tr A0A373QCL1 A0A373QCL1	H	I	S	R	V	L	R	L	E	tr R6MGG4 R6MGG4	H	I	S	R	V	L	R	L	E
tr A0A2TOHHA6 A0A2TOHHA6	Y	I	S	R	V	L	R	M	G	tr A0A0R3AJC6 A0A0R3AJC6	Y	I	S	R	V	L	R	M	G
tr A0A2D5K0H4 A0A2D5K0H4	Y	I	S	R	E	L	S	L	E	tr A0A1S2RWG7 A0A1S2RWG7	H	I	S	R	V	L	R	M	G
tr A0A1I717A9 A0A1I717A9	Y	I	S	R	V	L	R	M	G	tr A0A327GH15 A0A327GH15	Y	I	S	R	V	L	R	M	G
tr A0A1F7RUF0 A0A1F7RUF0	H	I	S	R	V	L	S	L	E	tr A0A2A9DK29 A0A2A9DK29	Y	I	S	R	V	L	R	M	G
tr A0A1J5DI80 A0A1J5DI80	H	I	S	R	V	L	R	L	E	tr A0A2U0F1P4 A0A2U0F1P4	Y	I	S	R	V	L	R	M	G
tr A0A1H1MSW6 A0A1H1MSW6	Y	I	S	R	V	L	R	M	G	tr UI1TAH5 UI1TAH5	Y	I	S	R	V	L	R	M	G
tr A0A2G4DLQ2 A0A2G4DLQ2	Y	I	S	R	V	L	R	M	G	tr A0A0D0T503 A0A0D0T503	Y	I	S	R	V	L	R	M	G
tr A0A1X7FKP7 A0A1X7FKP7	Y	I	S	R	V	L	R	M	G	tr A0A1M6AAZ5 A0A1M6AAZ5	H	I	S	R	V	L	R	L	E
tr A0A2G5LPW3 A0A2G5LPW3	Y	I	S	R	V	L	R	M	G	tr A0A0W0MXT1 A0A0W0MXT1	Y	I	S	R	V	L	R	M	G
tr A0A0D2CQF5 A0A0D2CQF5	H	I	S	R	G	P	S	L	E	tr A0A0U2W6H8 A0A0U2W6H8	H	I	S	R	V	L	R	L	E
tr A0A2N8CNY6 A0A2N8CNY6	Y	I	S	R	V	L	R	M	G	tr C3K3H4 C3K3H4	Y	I	S	R	V	L	R	M	G
tr A0A1V0B1M0 A0A1V0B1M0	Y	I	S	R	V	L	R	L	E	tr A0A2E5HYC8 A0A2E5HYC8	H	I	S	R	V	L	S	L	E
tr A0A1C5WAY1 A0A1C5WAY1	H	I	S	R	V	L	R	M	G	tr A0A1M3MDJ1 A0A1M3MDJ1	Y	I	S	R	V	A	R	M	G
tr Q6KF70 Q6KF70	H	I	S	R	E	P	R	L	E	tr A0A0D2JYC8 A0A0D2JYC8	H	I	S	R	G	P	S	L	E
tr Q6KF70 Q6KF70_2	Y	I	S	R	V	L	R	L	E	tr A0A2G2MUG3 A0A2G2MUG3	Y	I	S	R	V	L	R	M	G
tr Q6KF70 Q6KF70_3	Y	I	S	R	E	P	R	L	E	tr A0A1D6E8F6 A0A1D6E8F6	H	I	S	R	E	P	R	M	G
tr Q6KF70 Q6KF70_4	H	I	S	R	E	P	R	L	E	tr A0A1D6E8F6 A0A1D6E8F6_2	H	I	S	R	V	L	R	L	E
tr Q6KF70 Q6KF70_5	H	I	S	R	G	A	R	M	G	tr Q6KF61 Q6KF61	H	I	S	R	E	P	R	L	E
tr Q6KF70 Q6KF70_6	H	I	S	R	G	L	R	L	E	tr Q6KF61 Q6KF61_2	H	I	S	R	V	L	R	L	E
tr Q6KF70 Q6KF70_7	H	I	S	R	E	P	R	L	E	tr Q6KF61 Q6KF61_3	Y	I	S	R	E	P	R	L	E
tr Q6KF70 Q6KF70_8	H	I	S	R	E	P	R	L	E	tr Q6KF61 Q6KF61_4	H	I	S	R	E	P	R	L	E
tr Q6KF70 Q6KF70_9	H	I	S	R	V	L	R	L	E	tr Q6KF61 Q6KF61_5	H	I	S	R	G	A	R	M	G
tr Q6KF70 Q6KF70_10	H	I	S	R	E	P	R	L	E	tr Q6KF61 Q6KF61_6	H	I	S	R	G	L	R	L	E
tr Q6KF70 Q6KF70_11	H	I	S	R	G	P	S	L	E	tr Q6KF61 Q6KF61_7	H	I	S	R	E	P	R	M	G
tr Q6KF70 Q6KF70_12	H	I	S	R	G	P	R	L	E	tr Q6KF61 Q6KF61_8	H	I	S	R	E	P	R	L	E
tr Q6KF70 Q6KF70_13	H	I	S	R	E	P	R	M	G	tr Q6KF61 Q6KF61_9	H	I	S	R	V	L	R	L	E
tr Q6KF70 Q6KF70_14	H	I	S	R	E	P	R	M	G	tr Q6KF61 Q6KF61_10	H	I	S	R	V	L	R	L	E
tr Q6KF70 Q6KF70_15	H	I	S	R	V	L	R	L	E	tr Q6KF61 Q6KF61_11	H	I	S	R	E	P	R	L	E
tr A0A1H0S455 A0A1H0S455	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_12	H	I	S	R	E	P	R	M	G
tr A0A2A9GE39 A0A2A9GE39	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_13	H	I	S	R	G	A	R	M	G
tr A0A2T6MHM2 A0A2T6MHM2	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_14	H	I	S	R	G	L	R	L	E
tr A0A2S9EL05 A0A2S9EL05	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_15	H	I	S	R	E	P	R	L	E
tr A0A1C5Z3T0 A0A1C5Z3T0	H	I	S	R	V	L	R	L	E	tr Q6KF61 Q6KF61_16	H	I	S	R	E	P	R	L	E
tr A0A0Q4EZB0 A0A0Q4EZB0	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_17	H	I	S	R	G	P	S	L	E
tr A0A1H1EN63 A0A1H1EN63	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_18	H	I	S	R	G	P	R	L	E
tr A0A2V5BCD5 A0A2V5BCD5	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_19	H	I	S	R	E	P	R	M	G
tr A0A2V4Y916 A0A2V4Y916	Y	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_20	H	I	S	R	E	P	R	M	G
tr A0A2T9MFG3 A0A2T9MFG3	H	I	S	R	V	L	R	M	G	tr Q6KF61 Q6KF61_21	H	I	S	R	V	L	R	L	E
tr A0A1G3F3Q6 A0A1G3F3Q6	Y	I	S	R	V	L	R	M	G	tr A0A0E3KC87 A0A0E3KC87	Y	I	S	R	V	L	R	M	G
tr A0A2E7QQG0 A0A2E7QQG0	H	I	S	R	V	L	S	L	E	tr A0A0W0HT77 A0A0W0HT77	Y	I	S	R	V	L	R	M	G
tr A0A1L8PXT2 A0A1L8PXT2	Y	I	S	R	V	L	S	M	G	tr A0A0H5ATD8 A0A0H5ATD8	Y	I	S	R	V	L	R	M	G
tr A0A1H5AQ86 A0A1H5AQ86	Y	I	S	R	V	L	R	M	G	tr A0A1C9WF61 A0A1C9WF61	Y	I	S	R	V	L	R	M	G
tr A0A1C3GT58 A0A1C3GT58	Y	I	S	R	V	L	R	M	G	tr A0A1N7UDL9 A0A1N7UDL9	Y	I	S	R	V	L	R	M	G
tr A0A031LZW1 A0A031LZW1	Y	I	S	R	V	L	R	L	E	tr A0A2N5EW93 A0A2N5EW93	Y	I	S	R	V	L	R	M	G
tr A0A2K9CWX6 A0A2K9CWX6	Y	I	S	R	V	L	R	M	G	tr A0A078MMP6 A0A078MMP6	Y	I	S	R	V	L	R	L	E
tr A0A1E4XG00 A0A1E4XG00	Y	I	S	R	V	L	R	M	G	tr A0A0R2Y1P5 A0A0R2Y1P5	Y	I	S	R	V	L	R	M	G
tr A0A0A1YU87 A0A0A1YU87	Y	I	S	R	V	L	R	M	G	tr I4K9J5 I4K9J5	Y	I	S	R	V	L	R	M	G
tr A0A1H2POC5 A0A1H2POC5	Y	I	S	R	V	L	R	M	G	tr A0A1B5FHZ6 A0A1B5FHZ6	Y	I	S	R	V	L	R	M	G
tr A0A0R2ZR65 A0A0R2ZR65	Y	I	S	R	V	L	R	M	G	tr A0A355M5G3 A0A355M5G3	Y	I	S	R	V	A	R	L	E
tr A0A1C1WEE4 A0A1C1WEE4	Y	I	S	R	V	L	R	M	G	tr A4J7F0 A4J7F0	H	I	S	R	V	L	R	M	E
tr A0A1T2XS29 A0A1T2XS29	Y	I	S	R	V	L	R	M	G	tr A0A177X6Q5 A0A177X6Q5	Y	I	S	R	V	L	R	M	G
tr A0A1G0WBV5 A0A1G0WBV5	Y	I	S	R	V	A	R	L	E	tr A0A1Q5X0J9 A0A1Q5X0J9	Y	I	S	R	V	L	R	M	G
tr F6CMW4 F6CMW4	H	I	S	R	V	L	R	L	E	tr A0A0J8G117 A0A0J8G117	Y	I	S	R	V	L	R	M	G
tr A0A2H2Z098 A0A2H2Z098	Y	I	S	R	V	L	R	M	G	tr A0A1V4C2N8 A0A1V4C2N8	Y	I	S	R	E	L	S	L	E
tr A0A1B5CR44 A0A1B5CR44	Y	I	S	R	V	L	R	M	G	tr I4L5D5 I4L5D5	Y	I	S	R	V	L	R	M	G
tr A0A2A2FJR7 A0A2A2FJR7	Y	I	S	R	V	L	R	L	E	tr A0A1H4KAM0 A0A1H4KAM0	Y	I	S	R	V	L	R	M	G
tr A0A1H3K5B4 A0A1H3K5B4	Y	I	S	R	V	L	R	M	G	tr A0A2S8XRZ4 A0A2S8XRZ4	Y	I	S	R	V	L	R	M	G
tr A0A1H9WKK3 A0A1H9WKK3	Y	I	S	R	V	L	R	L	E	tr A0A1C5TN36 A0A1C5TN36	H	I	S	R	V	L	R	M	G
tr A0A2R3I2R0 A0A2R3I2R0	Y	I	S	R	V	L	R	M	G	tr A0A1V2JK71 A0A1V2JK71	Y	I	S	R	V	L	R	M	G
tr E2XZM8 E2XZM8	Y	I	S	R	V	L	R	M	G	tr A0A0R3B9G6 A0A0R3B9G6	Y	I	S	R	V	L	R	M	G
tr A0A359B4P4 A0A359B4P4	H	I	S	R	V	L	R	M	G	tr A0A257E459 A0A257E459	Y	I	S	R	V	L	R	M	G
tr A0A1B3EE22 A0A1B3EE22	Y	I	S	R	V	L	R	M	G	tr A0A172Z8E0 A0A172Z8E0	Y	I	S	R	V	L	R	M	G
tr A0A183BA7 A0A183BA7	Y	I	S	R	E	L	R	L	E										
tr A0A2L2XAH1 A0A2L2XAH1	H	I	S	R	V	L	R	L	E										
tr A0A1U9Q6B8 A0A1U9Q6B8	Y	I	S	R	V	L	R	M	G										
tr A0A1H8CK14 A0A1H8CK14	H	I	S	R	V	L	R	M	G										
tr A0A2N7YNC2 A0A2N7YNC2	Y	I	S	R	V	L	R	M	G										
tr A0A2G5N1Y9 A0A2G5N1Y9	Y	I	S	R	V	L	R	M	G										
tr A0A319EW70 A0A319EW70	H	I	S	R	G	A	S	L	E										
tr Q6AQE9 Q6AQE9	H	I	S	R	V	L	R	L	E										
tr A0A2E9AC05 A0A2E9AC05	H	I	S	R	V	L	S	L	E										
tr A0A0R2Z4N8 A0A0R2Z4N8	Y	I	S	R	V	L	R	M	G										
tr A0A1H3QTJ9 A0A1H3QTJ9	Y	I	S	R	V	L	R	M	G										
tr A0A2N8A3I6 A0A2N8A3I6	Y	I	S	R	V	L	R	M	G										
tr A0A127HTC2 A0A127HTC2	Y	I	S	R	V	L	R	M	G										

### 11.13 Alineamiento de motivos similares al patrón de expresión: Q-A-G-[AVL]-[VYA]-Q-[PQ]-G a partir de proteínas similares a TGZ1



tr A0A2T7CWE5 A0A2T7CWE5	Q Q A Y V Q G	tr A0A2E4T5S3 A0A2E4T5S3	Q A G L V Q P G	tr A0A384CB74 A0A384CB74	Q A G L V Q P G	tr A0A1C5A1B2 A0A1C5A1B2	Q A G A A Q P G
tr A0A182JR41 A0A182JR41	Q A G V A Q P G	tr A0A1V4KTD9 A0A1V4KTD9	Q A G V A Q P G	tr A0A0H2S6A6 A0A0H2S6A6	Q A G V A Q P G	tr A0A1V4KT87 A0A1V4KT87	Q A G V A Q P G
tr W5PCK5 W5PCK5	Q A G V V Q P G	tr A2XUN9 A2XUN9	Q A G V V Q P G	tr A0A0Q4FR7 A0A0Q4FR7	Q A G V V Q P G	tr A0A2D5RCH5 A0A2D5RCH5	Q A G L V Q P G
tr W5PCK5 W5PCK5_2	Q A G V V Q P G	tr A2XUN9 A2XUN9_2	Q A G V V Q P G	tr U3KAV8 U3KAV8	Q A G V V Q P G	tr A0A2P8BCZ4 A0A2P8BCZ4	Q A G L V Q P G
tr W5PCK5 W5PCK5_3	Q A G V V Q P G	tr U1ZWK2 U1ZWK2	Q A G A A Q P G	tr A0A285CN10 A0A285CN10	Q A G V V Q P G	tr A0A1U9JNA0 A0A1U9JNA0	Q A G L V Q P G
tr W5PCK5 W5PCK5_4	Q A G V V Q P G	tr F2DBR8 F2DBR8	Q A G A A Q P G	tr A0A0E0BWR6 A0A0E0BWR6	Q A G V V Q P G	tr A0A1D6J8M0 A0A1D6J8M0	Q A G A Y Q P G
tr W5PCK5 W5PCK5_5	Q A G V V Q P G	tr A0A2A6ERU9 A0A2A6ERU9	Q A G L A Q P G	tr A0A0E0BWR6 A0A0E0BWR6_2	Q A G V V Q P G	tr A0A1D6J8M0 A0A1D6J8M0_2	Q A G A Y Q P G
tr W5PCK5 W5PCK5_6	Q A G V V Q P G	tr U3JNZ7 U3JNZ7	Q A G V A Q P G	tr A0A229THW3 A0A229THW3	Q A G A A Q P G	tr A0A1D6J8M0 A0A1D6J8M0_3	Q A G A Y Q P G
tr W5PCK5 W5PCK5_7	Q A G V V Q P G	tr A0A2P6VK25 A0A2P6VK25	Q A G V A Q P G	tr A0A1H5D703 A0A1H5D703	Q A G A A Q P G	tr A0A1D6J8M0 A0A1D6J8M0_4	Q A G A Y Q P G
tr W5PCK5 W5PCK5_8	Q A G V A Q P G	tr G3WJ1 G3WJ1	Q A G V A Q P G	tr V5F132 V5F132	Q A G A A Q P G	tr V6UCL6 V6UCL6	Q A G L A Q P G
tr W5PCK5 W5PCK5_9	Q A G V A Q P G	tr A0A2V5CEX6 A0A2V5CEX6	Q A G V A Q P G	tr C0HGH1 C0HGH1	Q A G A A Q P G	tr K3Y7H0 K3Y7H0	Q A G A Y Q P G
tr A0A2P7QH96 A0A2P7QH96	Q A G A A Q P G	tr K3ZU98 K3ZU98	Q A G L V Q P G	tr C0HGH1 C0HGH1_2	Q A G V Q P G	tr A0A341XZF6 A0A341XZF6	Q A G A Y Q P G
tr A0A358C5C5 A0A358C5C5	Q A G A A Q P G	tr A0A358F6G5 A0A358F6G5	Q A G L V Q P G	tr C0HGH1 C0HGH1_3	Q A G V Q P G	tr A0A341XZF6 A0A341XZF6_2	Q A G A Y Q P G
tr A0A1N5P9G7 A0A1N5P9G7	Q A G V V Q P G	tr K3Y7I1 K3Y7I1	Q A G A A Q P G	tr C0HGH1 C0HGH1_4	Q A G V Q P G	tr V8NEL6 V8NEL6	Q A G A A Q P G
tr A0A2T7L857 A0A2T7L857	Q A G L A Q P G	tr A0A117B3K3 A0A117B3K3	Q A G A A Q P G	tr A0A1H0EFE2 A0A1H0EFE2	Q A G A A Q P G	tr A0A1C8LKN2 A0A1C8LKN2	Q A G V A Q P G
tr A6WAQ1 A6WAQ1	Q A G A Y Q P G	tr A0A1D9GY18 A0A1D9GY18	Q A G L A Q P G	tr A0A1S3AKU1 A0A1S3AKU1	Q A G L V Q P G	tr Q0K345 Q0K345	Q A G A A Q P G
tr A0A2P7TUY5 A0A2P7TUY5	Q A G L A Q P G	tr A0A0A0M6H3 A0A0A0M6H3	Q A G V V Q P G	tr A0A1S3AKU1 A0A1S3AKU1_2	Q A G L V Q P G	tr A0A1L6PMF8 A0A1L6PMF8	Q A G L A Q P G
tr A0A366L5E6 A0A366L5E6	Q A G L A Q P G	tr A0A328N1D2 A0A328N1D2	Q A G L V Q P G	tr A0A1S3AKU1 A0A1S3AKU1_3	Q A G L V Q P G	tr A0A218UF37 A0A218UF37	Q A G V A Q P G
tr A0A1Q4VTM6 A0A1Q4VTM6	Q A G L V Q P G	tr A0A341YW72 A0A341YW72	Q A G L A Q P G	tr A0A1S3AKU1 A0A1S3AKU1_4	Q A G L V Q P G	tr A0A1Q3YDU3 A0A1Q3YDU3	Q A G A Y Q P G
tr A0A367PF69 A0A367PF69	Q A G A A Q P G	tr K8DZR6 K8DZR6	Q A G V V Q P G	tr A0A1S3AKU1 A0A1S3AKU1_5	Q A G L V Q P G	tr A0A2X0J57 A0A2X0J57	Q A G A A Q P G
tr A0A0C3PG2 A0A0C3PG2	Q A G V A Q P G	tr A0A2D7Z4M1 A0A2D7Z4M1	Q A G L V Q P G	tr A0A1S3AKU1 A0A1S3AKU1_6	Q A G L V Q P G	tr Q0J1Y9 Q0J1Y9	Q A G V V Q P G
tr A0A293MG26 A0A293MG26	Q A G A A Q P G	tr A0A0N0KYW3 A0A0N0KYW3	Q A G L A Q P G	tr A0A1S3AKU1 A0A1S3AKU1_7	Q A G L V Q P G	tr Q1J0Y1 Q1J0Y1	Q A G A Y Q P G
tr S4XW84 S4XW84	Q A G A A Q P G	tr A0A111TX62 A0A111TX62	Q A G A A Q P G	tr G3WJ2 G3WJ2	Q A G L V Q P G	tr Q0J1Y9 Q0J1Y9_2	Q A G A Y Q P G
tr A0A178E4G5 A0A178E4G5	Q A G V V Q P G	tr A0A348WKN2 A0A348WKN2	Q A G L V Q P G	tr A0A1G1SR85 A0A1G1SR85	Q A G A A Q P G	tr W9FS05 W9FS05	Q A G L A Q P G
tr A0A140LY9 A0A140LY9	Q A G V V Q P G	tr A0A359KY2 A0A359KY2	Q A G L V Q P G	tr A0A209BV44 A0A209BV44	Q A G V V Q P G	tr A0A001DTV6 A0A001DTV6	Q A G L A Q P G
tr A0A140LY9 A0A140LY9_2	Q A G V V Q P G	tr A0A176YNH4 A0A176YNH4	Q A G A A Q P G	tr A0A0Q7F499 A0A0Q7F499	Q A G A A Q P G	tr A0A287X4Q3 A0A287X4Q3	Q A G A A Q P G
tr A0A140LY9 A0A140LY9_3	Q A G A A Q P G	tr A0A1D5P8H8 A0A1D5P8H8	tr B8A2N4 B8A2N4	tr A0A002S03 A0A002S03	Q A G A A Q P G	tr A0A1C4XT81 A0A1C4XT81	Q A G L V Q P G
tr A0A140LY9 A0A140LY9_4	Q A G V V Q P G	tr B8A2N4 B8A2N4_2	tr B8A2N4 B8A2N4_2	tr J3Y2Y2 J3Y2Y2	Q A G V V Q P G	tr A0A2P6V3U7 A0A2P6V3U7	Q A G A Y Q P G
tr A0A140LY9 A0A140LY9_5	Q A G V V Q P G	tr B8A2N4 B8A2N4_3	tr B8A2N4 B8A2N4_3	tr A0A0356Q7G9 A0A0356Q7G9	Q A G V V Q P G	tr A0A372DPU1 A0A372DPU1	Q A G A A Q P G
tr A0A140LY9 A0A140LY9_6	Q A G V V Q P G	tr B8A2N4 B8A2N4_4	tr B8A2N4 B8A2N4_4	tr A0A002CCP7 A0A002CCP7	Q A G V V Q P G	tr A0A1F3NY74 A0A1F3NY74	Q A G L V Q P G
tr A0A140LY9 A0A140LY9_7	Q A G V V Q P G	tr B8A2N4 B8A2N4_5	tr B8A2N4 B8A2N4_5	tr A0A2A4P5P8 A0A2A4P5P8	Q A G A A Q P G	tr A0A2D4GLR5 A0A2D4GLR5	Q A G L A Q P G
tr I1PM96 I1PM96	Q A G V V Q P G	tr B8A2N4 B8A2N4_6	tr B8A2N4 B8A2N4_6	tr A0A0C19ACD6 A0A0C19ACD6	Q A G A A Q P G	tr A0A254V8AR6 A0A254V8AR6	Q A G L A Q P G
tr I1PM96 I1PM96_2	Q A G V V Q P G	tr B8A2N4 B8A2N4_7	tr B8A2N4 B8A2N4_7	tr A0A1D5PDD0 A0A1D5PDD0	Q A G V V Q P G		
tr A0A0N5DXW6 A0A0N5DXW6	Q A G V V Q P G	tr B8A2N4 B8A2N4_8	tr B8A2N4 B8A2N4_8	tr A0A0D3FX29 A0A0D3FX29	Q A G V V Q P G		
tr A0A1A9D491 A0A1A9D491	Q A G L A Q P G	tr B8A2N4 B8A2N4_9	tr B8A2N4 B8A2N4_9	tr A0A0D3FX29 A0A0D3FX29_2	Q A G V V Q P G		
tr A0A316KNV2 A0A316KNV2	Q A G A A Q P G	tr B8A2N4 B8A2N4_10	tr B8A2N4 B8A2N4_10	tr A0A028FXZ4 A0A028FXZ4	Q A G V V Q P G		
tr A0A1N7PY85 A0A1N7PY85	Q A G A A Q P G	tr A0A293M526 A0A293M526	tr A0A293M526 A0A293M526	tr A0A0D9ZM57 A0A0D9ZM57	Q A G V V Q P G		
tr D6A19 D6A19	Q A G L A Q P G	tr F1MJO7 F1MJO7	tr F1MJO7 F1MJO7	tr A0A026P456 A0A026P456	Q A G V V Q P G		
tr A0A210KCS2 A0A210KCS2	Q A G V V Q P G	tr F1MJO7 F1MJO7_2	tr F1MJO7 F1MJO7_2	tr A0A0E0PAJ9 A0A0E0PAJ9	Q A G V V Q P G		
tr A0A1V4KTD7 A0A1V4KTD7	Q A G V A Q P G	tr F1MJO7 F1MJO7_3	tr F1MJO7 F1MJO7_3	tr A0A0E0PAJ9 A0A0E0PAJ9_2	Q A G V V Q P G		
tr A0A0Q5M2U9 A0A0Q5M2U9	Q A G A A Q P G	tr F1MJO7 F1MJO7_4	tr F1MJO7 F1MJO7_4	tr A0A356PLP3 A0A356PLP3	Q A G V V Q P G		
tr V8N156 V8N156	Q A G A A Q P G	tr F1MJO7 F1MJO7_5	tr F1MJO7 F1MJO7_5	tr A0A317WKH5 A0A317WKH5	Q A G V V Q P G		
tr A0A2S3I7K8 A0A2S3I7K8	Q A G A A Q P G	tr F1MJO7 F1MJO7_6	tr F1MJO7 F1MJO7_6	tr K7TUE5 K7TUE5	Q A G V V Q P G		
tr A0A171K423 A0A171K423	Q A G A A Q P G	tr F1MJO7 F1MJO7_7	tr F1MJO7 F1MJO7_7	tr K7TUE5 K7TUE5_2	Q A G V V Q P G		
tr A0A285YGC8 A0A285YGC8	Q A G A A Q P G	tr F1MJO7 F1MJO7_8	tr F1MJO7 F1MJO7_8	tr K7TUE5 K7TUE5_3	Q A G V V Q P G		
tr A0A0D1WH16 A0A0D1WH16	Q A G A A Q P G	tr F1MJO7 F1MJO7_9	tr F1MJO7 F1MJO7_9	tr K7TUE5 K7TUE5_4	Q A G V V Q P G		
tr L1MBA7 L1MBA7	Q A G A A Q P G	tr F1MJO7 F1MJO7_10	tr F1MJO7 F1MJO7_10	tr K7TUE5 K7TUE5_5	Q A G V V Q P G		
tr A0A2T0SFE6 A0A2T0SFE6	Q A G V V Q P G	tr F1MJO7 F1MJO7_11	tr F1MJO7 F1MJO7_11	tr M7ZS62 M7ZS62	Q A G V V Q P G		
tr A0A0E0H258 A0A0E0H258	Q A G V V Q P G	tr F1MJO7 F1MJO7_12	tr F1MJO7 F1MJO7_12	tr M7ZS62 M7ZS62_2	Q A G V V Q P G		
tr A0A0E0H258 A0A0E0H258_2	Q A G V V Q P G	tr F1MJO7 F1MJO7_13	tr F1MJO7 F1MJO7_13	tr A0A0Q4KKN2 A0A0Q4KKN2	Q A G V V Q P G		
tr A0A341XQ9 A0A341XQ9	Q A G V V Q P G	tr F1MJO7 F1MJO7_14	tr F1MJO7 F1MJO7_14	tr W7VMV5 W7VMV5	Q A G V V Q P G		
tr A0A341XQ9 A0A341XQ9_2	Q A G V V Q P G	tr A0A210M4E3 A0A210M4E3	tr A0A210M4E3 A0A210M4E3	tr A0A1Y2N880 A0A1Y2N880	Q A G V V Q P G		
tr A0A341XQ9 A0A341XQ9_3	Q A G V V Q P G	tr A0A2P6VEX1 A0A2P6VEX1	tr A0A2P6VEX1 A0A2P6VEX1	tr A0A1D6J8M1 A0A1D6J8M1	Q A G V V Q P G		
tr A0A2K3CUH8 A0A2K3CUH8	Q A G A A Q P G	tr A0A1ZSRD6G A0A1ZSRD6G	tr A0A1ZSRD6G A0A1ZSRD6G	tr A0A1D6J8M1 A0A1D6J8M1_2	Q A G V V Q P G		
tr A0A1F8R3X1 A0A1F8R3X1	Q A G V V Q P G	tr A0A1ZSRD6G A0A1ZSRD6G_2	tr A0A1ZSRD6G A0A1ZSRD6G_2	tr A0A1D6J8M1 A0A1D6J8M1_3	Q A G V V Q P G		
tr B6U0E3 B6U0E3	Q A G V V Q P G	tr A0A1ZSRD6G A0A1ZSRD6G_3	tr A0A1ZSRD6G A0A1ZSRD6G_3	tr A0A2Z2LSP3 A0A2Z2LSP3	Q A G V V Q P G		
tr B6U0E3 B6U0E3_2	Q A G V V Q P G	tr A0A1ZSRD6G A0A1ZSRD6G_4	tr A0A1ZSRD6G A0A1ZSRD6G_4	tr U1H131 U1H131	Q A G V V Q P G		
tr B6U0E3 B6U0E3_3	Q A G V V Q P G	tr A0A0D1XU7 A0A0D1XU7	tr A0A0D1XU7 A0A0D1XU7	tr A0A1D6E8F4 A0A1D6E8F4	Q A G V V Q P G		
tr B6U0E3 B6U0E3_4	Q A G V V Q P G	tr A0A2E5PAS0 A0A2E5PAS0	tr A0A2E5PAS0 A0A2E5PAS0	tr A0A1D6E8F4 A0A1D6E8F4_2	Q A G V V Q P G		
tr A0A1F4MRK8 A0A1F4MRK8	Q A G A A Q P G	tr A0A226MYC1 A0A226MYC1	tr A0A226MYC1 A0A226MYC1	tr A0A1D6E8F4 A0A1D6E8F4_3	Q A G V V Q P G		
tr A0A0F9LUD5 A0A0F9LUD5	Q A G V V Q P G	tr A0A1Q5ULH8 A0A1Q5ULH8	tr A0A1Q5ULH8 A0A1Q5ULH8	tr A0A1D6E8F4 A0A1D6E8F4_4	Q A G V V Q P G		
tr COP655 COP655	Q A G A Y Q P G	tr A0A2S6IEM9 A0A2S6IEM9	tr A0A2S6IEM9 A0A2S6IEM9	tr A0A0D2LTS8 A0A0D2LTS8	Q A G V V Q P G		
tr COP655 COP655_2	Q A G A Y Q P G	tr D4YYY1 D4YYY1	tr D4YYY1 D4YYY1	tr A0A2T2WVF9 A0A2T2WVF9	Q A G L V Q P G		
tr COP655 COP655_3	Q A G A Y Q P G	tr Q7XK24 Q7XK24	tr Q7XK24 Q7XK24	tr A0A0U5NKQ2 A0A0U5NKQ2	Q A G L V Q P G		
tr COP655 COP655_4	Q A G A Y Q P G	tr Q7XK24 Q7XK24_2	tr Q7XK24 Q7XK24_2	tr A0A1D2SH20 A0A1D2SH20	Q A G V V Q P G		
tr COP655 COP655_5	Q A G A Y Q P G	tr A0A018M2L2 A0A018M2L2	tr A0A018M2L2 A0A018M2L2	tr U3J7Q2 U3J7Q2	Q A G V V Q P G		
tr A0A2X3MRP8 A0A2X3MRP8	Q A G A A Q P G	tr A0A1E4PIF6 A0A1E4PIF6	tr A0A1E4PIF6 A0A1E4PIF6	tr F7CC37 F7CC37	Q A G L A Q P G		
tr A0A2U3Q9Y7 A0A2U3Q9Y7	Q A G A A Q P G	tr A0A212D9M5 A0A212D9M5	tr A0A212D9M5 A0A212D9M5	tr F7CC37 F7CC37_2	Q A G L A Q P G		
tr A0A182PMN1 A0A182PMN1	Q A G V V Q P G	tr A0A212D9M5 A0A212D9M5_2	tr A0A212D9M5 A0A212D9M5_2	tr R4MKQ7 R4MKQ7	Q A G L A Q P G		
tr R0LHZ3 R0LHZ3	Q A G V V Q P G	tr A0A212D9M5 A0A212D9M5_3	tr A0A212D9M5 A0A212D9M5_3				

## 11.14 Frecuencia de aminoácidos presentes en las proteínas completa y el sitio activo de las enzimas



### 11.15 Tablas de las proteínas que presentan el sitio activo de la TG y la posición de los residuos involucrados en la tríada catalítica

De acuerdo a la base de datos UniProtKB. El tipo de evidencia que respalda la existencia de la proteína no proporciona información sobre la precisión o corrección de las secuencias mostradas. Si bien proporciona información sobre la existencia de una proteína, puede suceder que la secuencia difiera ligeramente de las secuencias genómicas, especialmente para las secuencias derivadas de las predicciones del modelo genético.

Las secuencias de las proteínas seleccionadas fueron halladas a partir de una búsqueda avanzada que contenía el código del dominio sm00460 acompañado del origen taxonómico.

- Criterio de búsqueda: sm00460 taxonomy: "Viridiplantae [33090]"

ID UniProtKB	Descripción	Cys	His	Asp	Longitud (aa)	Evidencia
Q7F0R1	PNG1_Oryza_sativa_japonica	267	294	311	447	Nivel Transcripto
A0A1D6EW30	TGZc_Zea_mays	265	292	309	725	Nivel Transcripto
F2E1W5	Hordeum_vulgare	315	342	359	777	Nivel Transcripto
Q9FGY9	AtPNG1p_Arabidopsis_thaliana	251	278	295	721	Nivel Transcripto
A5PHD1	Arabidopsis_thaliana	251	278	295	721	Nivel Transcripto
F8QR42	Solanum_lycopersicum	256	283	300	725	Nivel Transcripto
B8LLV8	Picea_sitchensis	182	209	226	657	Nivel Transcripto
Q6KF70	TGZ15_Zea_mays (Sitio activo extendido)	439	498	522	534	Nivel Transcripto
Q6KF70	TGZ15_Zea_mays (Villalobos <i>et al.</i> , 2004)	439	446	471	534	Nivel Transcripto
M1VT12	TGO_Oryza_sativa (Campos <i>et al.</i> , 2013.)	254	260	295	352	Nivel Transcripto
M1VT12	TGO_Oryza_sativa (Sitio activo extendido)	254	271	295	352	Nivel Transcripto

- Criterio de búsqueda: sm00460 taxonomy: "Fungi [4751]"

ID UniProtKB	Descripción	Cys	His	Asp	Longitud (aa)	Evidencia
O74739	PNG1_Schizosaccharomyces_pombe	163	190	207	333	Nivel Transcripto
Q02890	PNG1_Saccharomyces_cerevisiae	191	218	235	363	Nivel Proteína
Q07533	CYK3_Saccharomyces_cerevisiae	524	563	578	885	Nivel Proteína
O14302	CYK3_Schizosaccharomyces_pombe	536	577	592	886	Nivel Proteína
D1MY48	Neurospora_crassa	208	235	252	412	Nivel Transcripto

- Criterio de búsqueda: sm00460 taxonomy:"Bacteria [2]"

ID UniProtKB	Descripción	Cys	His	Asp	Longitud (aa)	Evidencia
Q9HZX3	TGPA_Pseudomonas_aeruginosa	404	448	464	668	Nivel Proteína
P94476	YEBA_Bacillus_subtilis	484	532	548	737	Predicha
P9WL93	Y2569_Mycobacterium_tuberculosis	196	239	259	314	Nivel Proteína
P0A5G6	Y2599_Mycobacterium_bovis	196	239	259	314	Predicha
P9WL92	Y2569_Mycobacterium_tuberculosis	196	239	259	314	Predicha
Q49757	Y607_Mycobacterium_leprae	175	217	231	279	Predicha
P44103	Y1048_Haemophilus_influenzae	210	266	282	369	Nivel Proteína

- Criterio de búsqueda: sm00460 taxonomy:"Mammalia [40674]"

ID UniProtKB	Descripción	Cys	His	Asp	Longitud (aa)	Evidencia
P00488	FactorXIIIA_Homo_sapiens	315	374	397	732	Nivel Proteína
O08619	FactorXIIIA_Rattus_norvegicus	315	374	397	732	Nivel Transcripto
Q99041	TGM4_Rattus_norvegicus	256	315	338	667	Nivel Proteína
P49221	TGM4_Homo_sapiens	268	327	350	684	Nivel Proteína
P22735	TGM1_Homo_sapiens	377	436	459	817	Nivel Proteína
P22758	TGM1_Oryctolagus_cuniculus	397	456	479	836	Nivel Transcripto
O95932	TGM3L_Homo_sapiens	274	333	356	706	Nivel Proteína
F1M4T7	TGM6_Rattus_norvegicus	273	332	355	705	Predicha
O43548	TGM5_Homo_sapiens	278	337	360	720	Nivel Proteína
Q96PF1	TGM7_Homo_sapiens	279	338	361	710	Nivel Transcripto
Q08188	TGM3_Homo_sapiens	273	331	354	693	Nivel Proteína
P21981	TGM2_Mus_musculus	277	335	358	686	Nivel Proteína
P21980	TGM2_Homo_sapiens	277	335	358	687	Nivel Proteína
P08587	TGM2_Cavia_cutleri	277	335	358	690	Nivel Proteína
P16452	EPB42_Homo_sapiens	-	-	-	691	Nivel Proteína

**11.16 Resultados de las anotaciones enriquecidas en las proteínas anotadas funcionalmente como TGs de maíz de acuerdo al criterio de “Procesos Biológicos”**

GO	TGZc2	TGZc	TGZi	TGZ15	Descripción
GO:0045333	0.895	0	0	0.539	Respiración celular
GO:0019222	0.868	0.791	0.802	0.773	Regulación del proceso metabólico
GO:0006810	0.859	0.769	0.785	0.812	Transporte
GO:0009059	0.855	0.763	0.724	0.757	Proceso biosintético de macromoléculas
GO:0051171	0.837	0.697	0.829	0.769	Regulación del compuesto metabólico nitrogenado
GO:0051252	0.827	0.599	0.769	0.736	Regulación del proceso metabólico del ARN
GO:0030001	0.819	0	0	0	Transporte de iones metálicos
GO:0006355	0.814	0.62	0.785	0.709	Regulación de la transcripción, con molde de ADN
GO:0006796	0.77	0.764	0	0	Compuesto metabólico que contiene fosfato
GO:0034645	0.763	0.778	0.822	0.829	Proceso biosintético macromolécula celular
GO:1903506	0.719	0.706	0.782	0.768	Regulación de la transcripción con molde de ácido nucleico
GO:2001141	0.703	0.718	0.8	0.749	Regulación del proceso biosintético del ARN
GO:0006351	0.675	0.606	0.572	0.51	Transcripción, con molde de ADN
GO:0051641	0.672	0.554	0.613	0.653	Localización celular
GO:0045184	0.649	0.536	0	0	Establecimiento de localización de proteínas
GO:0055086	0.642	0	0	0	Molécula metabólica que contiene nucleobases pequeña proceso metabólico
GO:0051649	0.64	0.566	0.544	0.645	Establecimiento de localización en la célula
GO:0010468	0.62	0.562	0.659	0.597	Regulación de la expresión génica
GO:0006091	0.579	0	0	0	Generación de metabolitos precursores y energía
GO:0055114	0.576	0	0	0	Proceso de oxidación-reducción
GO:0015031	0.575	0	0	0	Transporte de proteínas
GO:0046907	0.555	0	0	0.531	Transporte intracelular
GO:0008104	0.554	0.521	0	0.525	Localización de proteínas
GO:0009117	0.527	0	0	0	Proceso metabólico de los nucleótidos
GO:0006812	0.519	0	0	0	Transporte de cationes
GO:0044281	0.512	0	0	0	Proceso metabólico de moléculas pequeñas
GO:0008380	0	0.929	0	0.647	Splicing ARN
GO:0006468	0	0	0	0.554	Fosforilación de proteínas
GO:0006811	0	0.899	0	0	Transporte de iones
GO:0000398	0	0.87	0	0	Splicing de ARNm, a través de spliceosome
GO:0006397	0	0.795	0	0	Procesamiento de ARNm
GO:0000375	0	0.773	0	0	Splicing de ARN, mediante reacciones de transesterificación
GO:0006396	0	0.74	0	0	Procesamiento de ARN
GO:0006066	0	0.603	0	0	Metabolismo del alcohol
GO:0006629	0	0.534	0	0	Metabolismo lipídico

**11.17 Resultados de las anotaciones enriquecidas en las proteínas anotadas funcionalmente como TGs de maíz de acuerdo al criterio de “Función Molecular”**

GO	TGZ15	TGZi	TGZc	TGZc2	Descripción
GO:0003676	0.909	0.941	0.887	0.839	Unión a ácidos nucleicos
GO:0008092	0.886	0.865	0.771	0.543	Unión a proteínas del citoesqueleto
GO:0008017	0.809	0.538	0	0	Unión a microtúbulos
GO:0003723	0.781	0.634	0.707	0	Unión a ARN
GO:0016301	0.719	0	0	0	Actividad quinasa
GO:0019900	0.698	0.627	0	0.578	Unión a quinasa
GO:0000166	0.681	0.627	0.818	0.715	Unión de nucleótidos
GO:0008270	0.642	0	0	0	Unión de iones de zinc
GO:0003779	0.642	0.588	0	0.662	Unión a actina
GO:0015631	0.635	0.545	0	0.745	Unión a tubulina
GO:0016818	0.615	0	0.619	0.754	Actividad hidrolasa
GO:0032549	0.604	0	0	0.513	Unión de ribonucleósidos
GO:0003677	0.555	0.616	0.651	0.53	Unión a ADN
GO:0017076	0.545	0	0.639	0.638	Unión a nucleótidos de purina
GO:0001882	0.543	0	0.575	0.587	Unión de nucleósidos
GO:0035639	0.537	0	0.615	0.572	Unión de trifosfato de ribonucleósido de purina
GO:0003824	0.527	0	0.901	0.973	Actividad catalítica
GO:0046872	0.931	0.954	0.578	0.948	Unión de iones metálicos
GO:0036094	0.859	0.821	0.839	0.633	Unión de molécula pequeña
GO:0005102	0.814	0.794	0.503	0.557	Unión al receptor
GO:0032403	0.788	0.776	0	0.731	Unión al complejo proteico
GO:0043169	0.605	0.649	0.601	0.858	Enlace catiónico
GO:0019904	0.583	0.582	0	0.673	Unión específica del dominio de proteínas
GO:0008134	0.501	0.57	0	0	Unión a factor de transcripción
GO:0001883	0	0	0.773	0.662	Unión de nucleósidos de purina
GO:0030554	0	0	0.682	0.569	Unión de nucleótidos de adenilo
GO:0044822	0	0	0.668	0.579	Unión a poli (A) de ARN

