



**INSTITUTO LATINOAMERICANO DE  
CIENCIAS DE LA VIDA Y LA NATURALEZA**

**PROGRAMA DE POSGRADO EN  
BIOCIENCIAS**

**ESTUDIO DE LA RESPUESTA GÉNICA *IN SILICO* AL ESTRÉS HÍDRICO EN  
*Pennisetum glaucum***

**LAURA BEATRIZ GARAY FARÍAS**

Foz do Iguazú  
2021



**INSTITUTO LATINOAMERICANO DE  
CIENCIAS DE LA VIDA Y LA NATURALEZA**

**PROGRAMA DE POSGRADO EN  
BIOCIENCIAS**

**ESTUDIO DE LA RESPUESTA GÉNICA *IN SILICO* AL ESTRÉS HÍDRICO EN  
*Pennisetum glaucum***

**LAURA BEATRIZ GARAY FARÍAS**

Tesis de maestría presentada al Programa de Posgrado en Biociencias, del Instituto Latinoamericano de Ciencias de la Vida y la Naturaleza, de la Universidad Federal de Integración Latinoamericana, como requisito parcial para la obtención del título de Maestría en Ciencias.

Orientador: Prof. Dr. Cristian Antonio Rojas

Foz do Iguazú  
2021

LAURA BEATRIZ GARAY FARÍAS

**ESTUDIO DE LA RESPUESTA GÉNICA *IN SILICO* AL ESTRÉS HÍDRICO EN  
*Pennisetum glaucum***

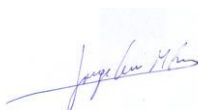
Tesis de maestría presentada al Programa de Posgrado en Biociencias, del Instituto Latinoamericano de Ciencias de la Vida y la Naturaleza, de la Universidad Federal de Integración Latinoamericana, como requisito parcial para la obtención del título de Maestría en Ciencias.

**BANCA EXAMINADORA**



---

Dr. Cristian Antonio Rojas  
Orientador  
UNILA



---

Dr. Jorge Luis Maria Ruiz  
UNILA



---

Dra. Clicia Grativol Gaspar De Matos  
UENF

Foz do Iguazú, 28 de Septiembre de 2021.

Catálogo elaborado pelo Setor de Tratamento da Informação  
Catálogo de Publicação na Fonte. UNILA - BIBLIOTECA LATINO

G212

Garay Farias, Laura Beatriz.

Estudio de la respuesta génica in silico al estrés hídrico en Pennisetum Glaucum / Laura Beatriz Garay Farias. - Foz do Iguaçu-PR, 2021.

86 f.: il.

Universidade Federal da Integração Latino-Americana. Instituto Latino-Americano de Ciências da Vida e da Natureza. Programa de Pós-Graduação em Biociências. Foz do Iguaçu-PR, 2021.

Orientador: Cristian Antonio Rojas.

1. Pennisetum Glaucum. 2. Stress hídrico. 3. Resposta genética. 4. Genética. 5. Proteínas. 6. Plantas - Reguladores. I. Rojas, Cristian Antonio. II. Título.

CDU 577.21

*A Dios, a mis Padres, mi esposo e  
hijos...*

## **AGRADECIMIENTO**

En primer lugar, agradezco a Dios por levantarme las veces que he caído y darme mi tan ansioso pedido de Fortaleza y sabiduría para enfrentar y afrontar los desafíos que fueron surgiendo durante el transcurso de este trabajo.

A mis padres por darme el legajo de Educación y hacerme entender lo importante que es la formación.

A mi orientador el Dr. Cristian Rojas, por su paciencia, y sus innúmeras horas de tiempo dedicado a compartir sus conocimientos.

A Sergio Litwiniuk, por su tiempo, paciencia y transmisión de conocimientos.

A Erwin, que con mucha paciencia me explico millones de veces las dudas que surgían durante el maestrado, por su disponibilidad y la energía positiva que me brindó en momentos complicados.

A mi esposo e hijos que tuvieron que soportar mis abandonos, ausencias y momentos de estrés.

A todas las personas que de alguna manera me ayudaron en la realización de este trabajo.

*«La educación es el arma más poderosa que puedes  
usar para cambiar el mundo»*

***Nelson Mandela***

FARÍAS, Laura Beatriz Garay. **ESTUDIO DE LA RESPUESTA GÉNICA *IN SILICO* AL ESTRÉS HÍDRICO EN *Pennisetum glaucum***. 86 páginas. Disertación de Maestría del Programa de Posgrado en Biociencias - Universidad Federal de Integración Latinoamericana, Foz do Iguazú, 2021.

## RESUMEN

Las literaturas revisadas ubican a *P. glaucum* como un cereal caracterizado por su calidad nutritiva y alta tolerancia al estrés por sequía. Sin embargo, es muy poco el conocimiento que se tiene sobre el mecanismo molecular que utiliza en respuesta al estrés hídrico. Para intentar esclarecer este punto realizamos un análisis de la modulación de la expresión de 2 tipos diferentes de genes a) PgVDAC, que codifica una proteína de voltaje de canal dependiente de aniones, como representante de genes estructurales y b) a la familia FT bZIP, como representante de genes reguladores por su baja exploración como candidatos potenciales para tolerancia al estrés hídrico. El hallazgo *in silico* de múltiples elementos ABREs, MYC, MYB-MBS y DRE en la región promotora de PgVDAC, sugieren su doble participación en el estrés hídrico, tanto en la vía dependiente como la vía independiente de ABA. Un cribado por todo el genoma de *P. glaucum* lanzó 52 probables FT bZIP, de los cuales 9 se mostraron diferencialmente expresos bajo condiciones de estrés hídrico, filtrados de datos de RNA-seq de un Transcriptoma depositado en el NCBI. Las regiones promotoras de estos genes bZIP presentaban múltiples elementos o motivos *cis* ABREs y DRE sugiriéndose así a su doble participación modulada en la respuesta lenta o adaptativa y en la respuesta rápida de este cereal al estrés hídrico. Los hallazgos de este estudio aportan datos complementarios para la comprensión del mecanismo detrás de la adaptación de *P. glaucum* bajo el estrés hídrico, pudiendo ser relevante para aplicaciones moleculares de cultivos potenciales.

**Palabras claves:** *P. glaucum*, estrés hídrico, PgVDAC, elementos *cis*, FT PgBZIP

FARÍAS, Laura Beatriz Garay. **IN SILICO STUDY OF GENE RESPONSE TO WATER STRESS IN *Pennisetum glaucum***. 86 pages. Master's thesis of the Post-Graduate Program in Biosciences – Federal University of Latin American Integration, Foz do Iguaçu, 2021.

### ABSTRACT

The reviewed literature places *P. glaucum* as a cereal characterized by its nutritional quality and high tolerance to drought stress. However, very little is known about the molecular mechanism used in response to water stress. To try to clarify this point, we carried out an analysis of the modulation of the expression of 2 different types of genes a) PgVDAC, which encodes an anion-dependent channel voltage protein, as a representative of structural genes and b) the FT bZIP family, as representative of regulatory genes due to their low screening as potential candidates for tolerance to water stress. The *in silico* finding of multiple ABREs, MYC, MYB-MBS and DRE elements in the promoter region of PgVDAC, suggest their double participation in water stress, both in the ABA-dependent and independent pathways. A full genome screening of *P. glaucum* released 52 putative FT bZIPs, of which 9 were differentially expressed under water stress conditions, filtered from RNA-seq data from a Transcriptome deposited at the NCBI. The promoter regions of these bZIP genes presented multiple elements or cis ABREs and DRE motifs, thus suggesting their double modulated participation in the slow or adaptive response and in the rapid response of this cereal to water stress. The findings of this study provide complementary data for the understanding of the mechanism behind the adaptation of *P. glaucum* under water stress, and may be relevant for molecular applications of potential crops.

**Keywords:** *P. glaucum*, water stress, PgVDAC, cis elements, FT PgBZIP

## LISTA DE FIGURAS

<b>Figura 1</b> – Cultivo de <i>P. glaucum</i> (mijo perla o millete) .....	14
<b>Figura 2</b> - Red de señalización en respuesta al estrés hídrico .....	18
<b>Figura 3</b> - Red de regulación transcripcional implicados en la respuesta al estrés. ....	26
<b>Figura 4</b> - Esquema <i>in silico</i> del gen funcional PgVDAC. ....	33
<b>Figura 5</b> - Esquema del experimento de estrés del Transcriptoma publicado .....	35
<b>Figura 6</b> - Esquema de trabajo para el análisis DEG .....	36
<b>Figura 7</b> – Construcción del Perfil HMM basados en el grado de conservación de la alineación de secuencias múltiples en <i>Arabidopsis</i> .....	42
<b>Figura 8</b> - Construcción del Perfil HMM basados en el grado de conservación de la alineación de secuencias múltiples en <i>Oryza</i> .....	43
<b>Figura 9</b> - Construcción del Perfil HMM basados en el grado de conservación de la alineación de secuencias en <i>S. Itálica</i> .....	44
<b>Figura 10</b> - Ontología Genética .....	46
<b>Figura 11</b> - Calidad de Lecturas .....	48
<b>Figura 12</b> - Visualización de Pg52.....	55
<b>Figura 13</b> - Visualización de Isoformas (Sashimi plots) de Pg5.....	56

## LISTA DE TABLAS

<b>Tabla 1</b> - Ubicación de secuencias inferidas de PgVDAC .....	32
<b>Tabla 2</b> - VDAC ortólogos .....	39
<b>Tabla 3</b> - Ubicación de VDAC ortólogas identificadas en <i>A. thaliana</i> y <i>O. sativa</i> .....	40
<b>Tabla 4</b> - Motivos <i>cis</i> presentes en VDAC de <i>A. thaliana</i> , <i>O. sativa</i> y <i>P. glaucum</i> .....	41
<b>Tabla 5</b> – Perfil HMM basados en el grado de conservación de dominio bZIP .....	41
<b>Tabla 6</b> – 52 probables FT bZIP identificados .....	45
<b>Tabla 7</b> - Mapeo y Ensamblaje de transcritos .....	49
<b>Tabla 8</b> - FT PgbZIP putativos seleccionados por su mayor DEG .....	50
<b>Tabla 9</b> - Valores de FC en el nivel de expresión de Pgbzip putativas seleccionadas ..	51
<b>Tabla 10</b> - Identidad de <i>P.glaucum</i> con modelos vegetales.....	52
<b>Tabla 11</b> - Funciones de bZIP ortólogas identificadas en <i>Arabidopsis</i> y <i>Oryza</i> .....	53
<b>Tabla 12</b> - Elementos reguladores involucrados en el estrés hídrico en PgbZIP seleccionados y ortólogos en <i>Oryza</i> y <i>Arabidopsis</i> .....	54
<b>Tabla 13</b> - Comparativos de expresión de isoformas .....	58
<b>Tabla 14</b> – Isoformas diferencialmente expresas en líneas endogámicas de <i>P. glaucum</i> .....	58

## LISTA DE ABREVIATURAS

ABA	Ácido abscísico
ABI	ABA insensible
ABRE	Elemento de Respuesta a ABA
Blast	Basic local alignment search tool
bZIP	Proteína con dominio básico y cremallera de leucina
DRE	Elemento de Respuesta a la Deshidratación
DEG	Diferencial de expresión génica
GO	Ontología genética
HMM	Modelo oculto de Markov
HSP	Proteínas estructurales de choque térmico
IGV	Genome Viewer browser
LEA	Proteínas abundantes de embriogénesis tardía
LogFC	Logaritmo en base 2 del cambio en la expresión
Pb	Pares de bases
Pg	Pennisetum glaucum
DHN	Deshidrinasa
HSP70	Chaperona térmica
GPx	Glutation peroxidasa
VDAC	Proteínas de canal de aniones dependiente de voltaje
TF	Factores de transcripción

## INDICE

<b>1</b>	<b>INTRODUCCIÓN</b> .....	14
<b>1.1</b>	<b>Revisión de Literatura</b> .....	16
1.1.1	Milleto ( <i>Pennisetum glaucum</i> ).....	16
1.1.2	Mecanismo de Señalización ante el Estrés hídrico.....	17
1.1.3	Respuestas Adaptativas ante el estrés hídrico.....	18
1.1.4	Genes que codifican para Proteínas Estructurales involucradas en la tolerancia a la sequía.....	19
1.1.5	Genes que codifican para Proteínas Regulatoras involucrados en la tolerancia a la sequía.....	22
1.1.6	Regiones promotoras.....	24
1.1.7	Elementos <i>cis</i> reguladores.....	25
1.2	Justificativa.....	29
<b>2</b>	<b>OBJETIVOS</b> .....	31
2.1	General.....	31
2.2	Específicos.....	31
<b>3</b>	<b>MATERIALES Y MÉTODOS</b> .....	32
3.1	Rastreo de la región promotora del gen estructural PgVDAC.....	32
3.2	Identificación y Análisis de la Región Promotora de VDAC ortólogos.....	32
3.3	Gen regulador de la familia bZIP.....	33
3.3.1	Cribado <i>in silico</i> en todo el genoma de <i>P. glaucum</i> de probables genes reguladores FT bZIP.....	33
3.4	Rastreo de la Región promotora FT bZIP putativos seleccionados.....	38
<b>4</b>	<b>RESULTADOS</b> .....	39
4.1	Análisis <i>in silico</i> de la Región promotora del gen estructural PgVDAC.....	39
4.2	Identificación y Análisis de las Regiones Promotoras de VDAC Ortólogos.....	39
4.3	Cribado <i>in silico</i> en todo el genoma de <i>P. glaucum</i> de probables genes reguladores de la familia bZIP.....	41
4.3.1	Anotación Funcional.....	46

4.3.2	Análisis de la Expresión Diferencial de genes bZIP putativos .....	47
<b>5</b>	<b>DISCUSIÓN</b> .....	<b>59</b>
<b>6</b>	<b>CONCLUSIONES</b> .....	<b>63</b>
<b>7</b>	<b>REFERENCIAS BIBLIOGRÁFICAS</b> .....	<b>65</b>
<b>8</b>	<b>ANEXOS</b> .....	<b>78</b>
8.1	Lista de bZIP referenciales utilizados.....	78
8.2	Alineamiento ATVDAC2 .....	80
8.3	Alineamiento OSVDAC2 .....	81
8.4	Términos ontológicos (GO) asociadas a la categoría de procesos biológicos y Funciones Moleculares.....	82
8.5	Ubicación de Región Promotora de probables Pg bZIP seleccionados y bZIP ortólogos.....	83
8.6	Herramientas del Galaxy .....	84
8.6.1	Hisat2 .....	84
8.6.2	StringTie .....	84
8.6.3	FeatureCounts .....	85
8.6.4	DESeq2.....	85
8.6.5	DESeq2 Normalizado .....	86
8.6.6	DESeq2 plots .....	86

## 1 INTRODUCCIÓN

El cambio climático inminente y las dificultades consecuentes de la producción de alimentos a nivel mundial es una realidad (ATKINSON *et al*, 2012).; esto lleva a la comunidad científica a innovar y entender los mecanismos genéticos de cultivos sugestivos que sirvan como herramientas para hacer frente al estrés abiótico, citando a uno de ellos “el estrés hídrico” que preocupa y crea una inseguridad alimentaria mundial.

*Pennisetum glaucum*, sinónimo *Cenchrus americanus*, conocido comúnmente como millete o mijo perla es un cereal (Figura 1) cultivado en tierras marginales de las regiones tropicales áridas y semiáridas del África e India; distinguido por su **alta tolerancia al estrés abiótico**, como temperatura elevada, sequía y pH elevado del suelo; apreciable por la biofortificación que aporta, y el bajo índice glicémico que presenta (SINGH P *et al*, 2017), características todas que lo hacen atrayente de estudiar.

**Figura 1** – Cultivo de *P. glaucum* (mijo perla o milleteo)



Recientemente, el consorcio internacional de secuenciación del genoma del mijo perla (IPMGSC) ha secuenciado el genoma de millete, cuyo borrador está disponible públicamente, y aunque se trate de una versión inicial podrá ayudar a comprender los mecanismos de tolerancia a la sequía de este cereal (VARSHNEY R *et al*, 2017).

El mecanismo de respuesta ante un estrés abiótico comienza con un estímulo ambiental que lleva a un cambio en la presión de turgencia celular, desencadenando una compleja ruta de señalización, la cual puede ser regulada o no por la presencia de una hormona vegetal de estrés llamada ácido abscísico (ABA). Se activa así una cadena de transducción que amplifica el estímulo inicial en la que participan mensajeros secundarios, y cuando llega esta señal al núcleo se produce la activación de genes específicos que codifican para proteínas funcionales o estructurales y genes que codifican para proteínas reguladoras (KUROMORI *et al*, 2014).

La regulación de la expresión de estos genes estructurales o reguladores activados en respuesta al estrés ocurre mayormente justo antes de la síntesis del transcripto (ARN mensajero), al formarse el complejo de proteínas de pre-iniciación sobre la región promotora del gen, cuyos elementos regulatorios *cis* podrían intervenir en el diálogo entre estas rutas de señalización (ATKINSON *et al*, 2012).

La bioinformática es una herramienta de investigación científica, siendo uno de sus pilares la recopilación de secuencias biológicas e información asociada en bases de datos, ya que a diario se generan grandes cantidades de información provenientes de secuenciación, datos de RNA-seq, vías metabólicas y otros tipos (BANERJEE A & ROYCHOUDHURY A, 2017). Estos datos biológicos nos permitirían dar uso a la gran cantidad de herramientas *in silico*, prediciendo por un análisis funcional de las regiones promotoras de genes estructurales y reguladores que pudieren estar involucrados en respuesta al estrés hídrico datos sobre el mecanismo de tolerancia en millete, lo cual permitiría diseñar estudios más detallados de la tolerancia a la sequía en esta especie, así como a desarrollar marcadores genéticos que promuevan al perfeccionamiento de los patrones de expresión de plantas de interés productivo modificadas genéticamente.

## 1.1 Revisión de Literatura

### 1.1.1 Milleto (*Pennisetum glaucum*)

El Mijo perla o millete (*Pennisetum glaucum*), sinónimo *Cenchrus americanus*, es un miembro de la familia de las gramíneas C4, diploide, con 7 cromosomas ( $2n=2x=14$ ), de polinización cruzada, ciclo corto, cultivado en tierras marginales de las regiones tropicales áridas y semiáridas del África e India. Sirve como uno de los alimentos básicos para millones de personas pobres y también se utiliza ampliamente como forraje y combustible (SEHGAL D, 2012). Frecuentemente es citado como un cultivo tolerante a ambientes hostiles como temperatura elevada, sequía y pH alto del suelo, posee un alto índice nutricional en comparación con el arroz, el trigo, el sorgo y el maíz, con un 8% a 19% de proteínas, bajo contenido de almidón, alto contenido de fibra y micronutrientes esenciales como hierro y zinc (TAKO E *et al*, 2015).

Con el fin de presentar el gran potencial de mijo perla DIAS-MARTINS (2018) describe a este cultivo como no tradicional, que viene siendo plantado en el Brasil durante al menos 50 años, pero con finalidad de cobertura y alimento para animales, sin valorar a sus granos por su alto potencial como alimento para los seres humanos por no contener gluten, índice glucémico bajo, con mayor fibra dietética que el arroz, similar en contenido de lípidos al maíz y mayor contenido de aminoácidos esenciales (leucina, isoleucina y lisina) que otros cereales tradicionales, como el trigo y el centeno. Además, citan los autores, este cultivo es menos susceptible a la contaminación por aflatoxinas en comparación con el maíz.

Estos atributos promueven al millete como un alimento alternativo para el control de peso y para reducir el riesgo de enfermedades crónicas, como la diabetes; y más aún por la alta tolerancia a la sequía y temperatura elevada, lo cual hace atractivo su producción para confrontar al cambio climático y la inseguridad alimentaria (SINGH P, 2017).

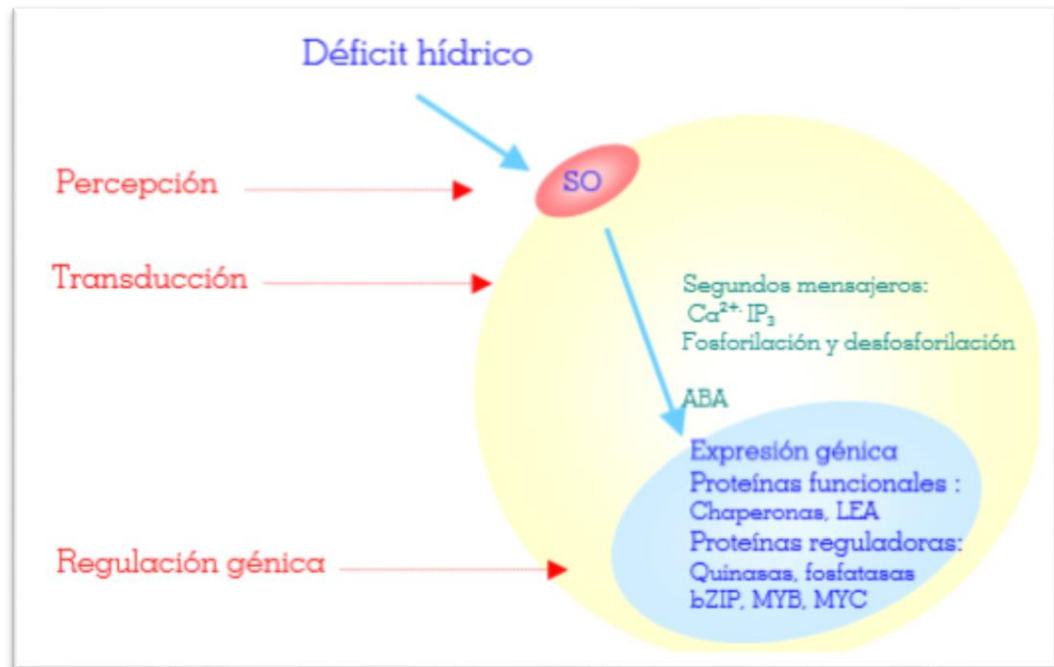
A pesar de poseer un genoma de gran tamaño (~2352 Mb), su tolerancia al estrés hídrico hace que se realicen estudios a fin de comprender la base molecular y genética de esta tolerancia, lo cual se facilita con la ayuda del borrador genómico de millete en su versión inicial, públicamente disponible. (SHIVHARE R *et al*, 2017).

### 1.1.2 Mecanismo de Señalización ante el Estrés hídrico

Como consecuencia de la limitación motriz, las plantas han desarrollado complejas respuestas bioquímicas a fin de protegerse de condiciones ambientales adversas. La combinación sinérgica o antagónica de los mecanismos fisiológicos, bioquímicos y moleculares conforman una cascada de señalización que permite la supervivencia de la planta ante las condiciones de estrés. Estudios que integren análisis en su transcriptómica, metabolómica y proteómica podrían ayudar a comprender estas respuestas (FUJITA, 2006).

La respuesta celular de las plantas a la falta de agua se da por la percepción de la señal de deshidratación, la transducción de la señal al núcleo y la regulación de la expresión génica, generando una respuesta bioquímica a partir de este estímulo físico. El mecanismo sensor del déficit hídrico no se conoce, aunque se postula que el cambio osmótico produce un cambio en la presión de turgencia celular, que es el desencadenante de la transducción de la señal, involucrando al receptor del ABA o Ácido abscísico (Figura 2) y a esta propia fitohormona. A partir de esta señal se activa una cadena de transducción que amplifica la señal inicial en la que participan mensajeros secundarios como el Calcio ( $\text{Ca}^{2+}$ ) o el Inositol trifosfato ( $\text{IP}_3$ ). Cuando la señal llega al núcleo se produce la activación de genes específicos que codifican para proteínas funcionales y genes que codifican para proteínas reguladoras (KUROMORI, 2014).

**Figura 2** - Red de señalización en respuesta al estrés hídrico



FUENTE: EXTRAÍDO DE BOZA C, 2003

### 1.1.3 Respuestas Adaptativas ante el estrés hídrico

El ABA es una fitohormona relacionada al estrés vegetal, ligada a respuestas adaptativas que involucran al estrés hídrico, el cual actúa como activador de diferentes rutas de señalización y metabólicas relacionadas con la respuesta al estrés abiótico, incluyendo a determinados y específicos procesos celulares tales como la traducción y transcripción de diversos factores de transcripción, en la activación de diversas proteínas involucradas en la fotosíntesis, mecanismos antioxidantes, respuestas a patógenos, señalización hormonal y síntesis de osmolitos compatibles (SUZUKI, 2014).

En la mayoría de las ocasiones la respuesta de la planta en la naturaleza integra las diferentes vías de respuesta; un estrés osmótico causado por sequía en verano se acompaña frecuentemente de estrés por altas temperaturas y las bajas temperaturas se relacionan con el estrés osmótico al dificultar la absorción y el transporte del agua. El ABA es el mayor mediador de respuestas a nivel de planta entera en situaciones de sequía o deshidratación. El papel principal del ABA durante el crecimiento vegetativo es optimizarlo en las condiciones adversas, mediante el mantenimiento de la

homeostasis osmótica, y para ello promueve respuestas de desarrollo (como la disminución del crecimiento), adaptaciones morfológicas (como el cierre de estomas) y cambios fisiológicos (como la disminución de la transpiración). Las raíces son probablemente los órganos que primero perciben el estrés hídrico y sintetizan el ABA, que posteriormente será transportado a otras partes de la planta, como las hojas, donde producirá el cierre de los estomas (KUROMORI, 2014).

#### **1.1.4 Genes que codifican para Proteínas Estructurales involucradas en la tolerancia a la sequía**

Varios genes se activan en respuesta al estrés abiótico a nivel transcripcional, y se contempla que sus productos proporcionan tolerancia al estrés mediante la producción de proteínas metabólicas vitales y en la regulación de los genes (LATA, 2011).

Amplios análisis de transcriptomas han revelado que estos productos génicos pueden clasificarse en dos grupos. El primer grupo se compone de genes que codifican proteínas que defienden a las células de los efectos del déficit hídrico e incluyen proteínas que participan en el metabolismo de osmoprotectores (osmolitos), proteínas de transporte (canales iónicos, acuaporinas), proteínas involucradas en procesos de reparación y degradación (chaperonas, proteasas), enzimas detoxificantes (ascorbato peroxidasa), así como otras proteínas de función desconocida, como las LEA (late embryogenesis abundant) que posiblemente estén involucradas en la protección de otras proteínas y de las membranas celulares (HASHIMOTO Y, 2002).

Otro grupo de genes estimulados por estrés abiótico incluye proteínas reguladoras que regulan la transducción de la señal de estrés y alteran la expresión génica; comprendiendo a varios factores de transcripción (FT) que enfatizan el papel de varios mecanismos reguladores de la transcripción en las vías de transducción de señales de estrés; como las proteínas quinasas, fosfatasa y proteasas implicadas en la regulación de la señalización del estrés y la expresión génica (LATA, 2011).

En *P. glaucum*, la PgGPx (Glutathion peroxidasa en *P. glaucum*) se presenta como un gen estructural que codifica para una enzima peróxida glutathion peroxidasa, involucrada en diversas vías de adaptación al estrés biótico y abiótico. En un experimento realizado, sugirieron que la transcripción de PgGPx está altamente regulada en respuesta

a las tensiones de salinidad y sequía; cuyo análisis de expresión dependiente de la transcripción de PgGPx, mostró una regulación positiva significativa con la sequía (ISLAM T, 2015).

Las proteínas estructurales de choque térmico (HSP) actúan como chaperonas moleculares, con funciones variadas, pero más específicamente asociadas con la tolerancia al calor adquirida, ya que son responsables de la síntesis, maduración y degradación de proteínas bajo condiciones normales, y bajo condiciones de estrés, participan en la protección de proteínas, estabilización de membranas y asisten al plegamiento de las proteínas (BALOJI G, 2019).

Algunos estudios caracterizan a PgHSP70 (o HSC70; chaperona térmica expresa en *P. glaucum*) como una chaperona citosólica expresa en respuesta al estrés por calor. El análisis *in silico* de la secuencia promotora del gen PgHSC70 condujo a la identificación de una serie de supuestos elementos *cis* reguladores como HSE (Elemento de Shock térmico), ABRE (Elemento de respuesta al ácido abscísico) para la expresión sensible al ABA, entre otros. La comparación de la secuencia del promotor PgHsc70 con otros promotores de otras chaperonas de plantas reveló un motivo común con alta homología con el elemento HSE. Sin embargo, la presencia de otros elementos reguladores y su acción concertada en asociación con sus correspondientes factores de transcripción pueden regular la expresión del transcrito de PgHsc70 en diversas condiciones ambientales (REDDY, 2010).

Las deshidrinas (DHN) o las proteínas LEA del grupo II juegan un papel protector en las plantas bajo diferentes condiciones de estrés abiótico como sequía, salinidad, estrés por frío y calor. Un ensayo realizado manifiesta que el gen PgDHN (gen que codifica para Dehidrina en *P. glaucum*) de millete es expreso bajo condiciones de estrés hídrico, acumulándose en las hojas. El análisis de secuencia *in silico* de 0,817 kb del promotor del gen PgDHN identificó elementos reguladores putativos que actúan en *cis* como el motivo MBS inducible por sequía, elementos ABRE para la expresión de respuesta ABA y otros varios motivos que no solo se expresan durante el estrés hídrico sino también durante otros estreses ambientales para proteger a las plantas (SINGH, 2015).

Los niveles de expresión de la transcripción de PgDREB2A fueron estudiados por AGARWAL P *et al* (2007) en respuesta a la deshidratación, baja temperatura (4 ° C) y estrés salino por análisis de transferencia Northern. La transcripción de PgDREB2A se reguló fuertemente en respuesta al tratamiento con frío en comparación con sal o tratamientos de sequía. En el caso de *Arabidopsis* y arroz, el gen DREB2A está regulado solo durante la deshidratación y el estrés de salinidad, pero no durante baja temperatura. El árbol filogenético reveló que PgDREB2A está más cerca de DREBs aislados de monocotiledóneas, aunque forma una rama independiente. Sus datos indican que la inducción de estrés de los genes podría ocurrir a través de la modificación postraduccional por fosforilación de DREB2A.

La membrana interna de las mitocondrias y los cloroplastos contienen numerosas proteínas transportadoras (porinas) que facilitan el movimiento selectivo de iones y metabolitos a través de él. En bajo potencial de membrana estas porinas están en un estado completamente abierto y permite el paso de aniones, por ello también se conocen como Proteínas de canal de aniones dependiente de voltaje (VDAC) que se encuentran en las membranas mitocondriales externas de todos los organismos superiores, actuando como puertas controladoras para el movimiento de diferentes metabolitos conduciendo a un rápido cambio en el potencial de membrana como una señal para reajustar la homeostasis celular ante algún tipo de estrés. La mayoría de las VDAC de plantas han sido estudiadas por sus propiedades electrofisiológicas; sin embargo, la clonación y el análisis de secuencia de genes que codifican para VDAC de plantas se han reportado solo en pocas especies vegetales, con niveles de transcripción diferencial en diferentes tejidos. El número de publicaciones dedicadas específicamente a las VDAC de plantas ha crecido más lentamente que las que tratan de VDAC de otros organismos. A pesar de la relativa escasez de información sobre algunas características de las VDAC vegetales, estos canales tienen sus propias especificidades y desempeñan un papel clave en la fisiología de las plantas (HOMBLÉ F, 2012).

Se ha destacado que la clonación y caracterización de un gen que codifica VDAC en *P. glaucum* (PgVDAC) está regulado de forma diferencial en respuesta a diversas tensiones abióticas en una manera independiente del ácido abscísico (ABA); los niveles de expresión de la transcripción de PgVDAC en respuesta a tensiones como la

deshidratación y la baja temperatura, mostraron una significativa sobre regulación de la transcripción de PgVDAC; destacándose así que las condiciones estresantes abióticas como baja temperatura, desecación y alta salinidad pueden alterar la homeostasis en el potencial hídrico y la distribución de iones que conduce a daños moleculares, detención del crecimiento e incluso los sistemas de transporte dependientes de energía. Los canales (VDAC) en estas membranas pueden mediar grandes flujos pasivos de aniones en los compartimientos inter y/o intracelular, desencadenando un cambio rápido en el potencial de membrana, emitiendo una señal para reajustar la homeostasis celular que se interrumpe durante tensiones ambientales (DESAI, 2005).

### **1.1.5 Genes que codifican para Proteínas Reguladoras involucrados en la tolerancia a la sequía**

La regulación de la expresión que se da en diferentes tipos de tejidos y órganos, en diferentes etapas de crecimiento y desarrollo del organismo, la cual puede ser afectada por estímulos ambientales, está regulada a nivel transcripcional, post-transcripcional y pos-traducciona; siendo la tasa transcripcional del gen la estrategia de regulación más precisa (ALLISON, 2011). Las proteínas reguladoras están involucradas en el control de la expresión génica durante la deshidratación e incluyen factores de transcripción (bZIP, MYC, MYB, etc.), proteínas quinasas y fosfatasas (ATKINSON, 2012).

Es así que las plantas se han adaptado para escapar de las tensiones ambientales a través del empleo de varios factores transcripcionales integrados. Los factores de transcripción específicos (FT, proteínas reguladoras) se unen a respectivos elementos *cis* afines presentes en la región promotora de sus genes diana y modulan el nivel de expresión de los genes en condiciones de estrés. Estas interacciones se manifiestan significativamente para controlar la supervivencia de la planta en condiciones ambientales estresantes, presentando efecto estabilizador o represor sobre el complejo de pre-iniciación de la transcripción, por interacciones directas con los componentes del complejo o interacciones indirectas con alguno de los co-reguladores asociados (SHRESTHA A, 2018).

Un factor de transcripción está compuesto por un dominio de unión al ADN, un dominio de dimerización y dominio regulador, categorizados en distintas familias de acuerdo a las propiedades estructurales conservadas de los dominios de sitios de unión al ADN (LINDEMOSE *et al*, 2013).

Se destaca la regulación transcripcional en eucariotas controlada especialmente por los factores de transcripción (TF), que interactúan con la región promotora de los genes para regular la función de los genes hacia diversos estreses abióticos / bióticos y activan objetivos posteriores para mejorar la tolerancia al estrés de las plantas (MA Y, 2015). Se han reportado más de 1500 FT en plantas, cada uno con la capacidad de controlar la expresión de pocos o muchos genes por medio de complejas vías de señalización (Gerhardt *et al*, 2011); a saber, de unión a ABRE (ABF) / unión de elemento sensible a ABA (AREB), factores de unión de elemento sensible al etileno (ERF), DREB (Proteína de unión al elemento sensible a la deshidratación), NAC, AP2 / ERF, WRKY, MYB, MYC, bZIP, etc (CHANWALA J, 2020).

Las proteínas bZIP son una familia de factores de transcripción relacionadas a la tolerancia de estrés abiótico, caracterizadas por presentar un motivo básico hélice-alfa continua (conformada de alrededor de 20 aminoácidos con seis argininas y cuatro residuos de lisina) responsable de la unión específica al ADN; la cual muchas veces contiene la secuencia RKQS, importantes para la transducción de señal; y un motivo de dimerización (constituida de alrededor de 41 aminoácidos con residuos de leucina en cada séptima posición, pudiendo ser homo o heterodímera), denominada zipper de leucina dada su conformación estructural (GATZ, 2012).

Los FT bZIP, así como otros factores de transcripción se unen al ADN en puntos de reconocimiento específicos. Estos motivos consisten en elementos *cis* característicos de cajas promotoras que son reconocidas y pareadas por el factor de transcripción. Cada subfamilia bZIP reconoce cajas promotoras específicas, los de mayor preferencia son la caja A (TACGTA), caja C (GACGTC), caja G (CACGTG), caja H (CCTACC), motivo TGACG, motivos ABREs. Múltiples copias de ABRE presentes en las regiones cadena arriba de genes inducibles por estrés pueden iniciar una transcripción dependiente de ABA (ROYCHOUDHURY. 2013).

Estudios de factores de transcripción bZIP en *Arabidopsis*, tomate (*Solanum lycopersicum*), soja (*Glycine máx*) y arroz (*Oryza sativa*) han demostrado una mayor expresión ante un estrés por sequía, pero pocos factores de transcripción bZIP han sido explorados como candidatos potenciales para aplicaciones de mejoramiento genético para tolerancia a sequía en cultivos (LINDEMOSE *et al*, 2013).

### 1.1.6 Regiones promotoras

Como fue destacado, la regulación transcripcional tiene un importante rol en la activación y supresión de la expresión génica, controlada en gran parte por los promotores de los genes diana, así como de la relación de estos con los factores de transcripción y de sus interacciones con secuencias regulatorias (FINER, 2014). Un promotor puede tener tamaños variables de 1-2 kb o más (Dutt *et al*, 2014), constituidos de secuencias de DNA, ubicado en la región río arriba (*upstream*) del sitio de inicio de la transcripción; cuya función está determinada por la combinación de varios elementos que lo conforman, reguladores en *cis* y activadores, además de la interacción de proteínas específicas (VEDEL, 2011).

Por lo general la región promotora se divide en tres partes: el **promotor central** de 50 a 100 pb, adyacente al inicio de la transcripción, con dos elementos claves, la caja TATA y la región iniciadora; el **promotor proximal**, ubicado a unos cuantos cientos de pares de bases cadena arriba del codón de inicio, relacionada directamente sobre la eficiencia de la actividad transcripcional, con varios elementos reguladores, como la caja AGGA; la caja GC, que intensifica la expresión del gen (BIŁAS , 2016); y por último el **promotor distal**, región que contiene los potenciadores y silenciadores de genes, típicamente encontrada a miles de pares de bases corriente arriba del inicio de la transcripción (LIU , 2016). Pueden clasificarse en **constitutivos**, las que dirigen la expresión a niveles constantes en todos los tejidos y en todo momento; promotores **tejido-específico**, proveen un mayor control de la expresión del gen nativo, la expresión se restringe a ciertas células, tejidos, órganos o etapas de desarrollo; **promotores inducibles**, inducidos por factores físicos como factores bióticos y abióticos, así como por agentes químicos (BIŁAS, 2016).

Son varios los promotores que se han aislado de especies agrícolas importantes, así como también se han usado para el mejoramiento genético de otros cultivos (NAQVI, 2016).

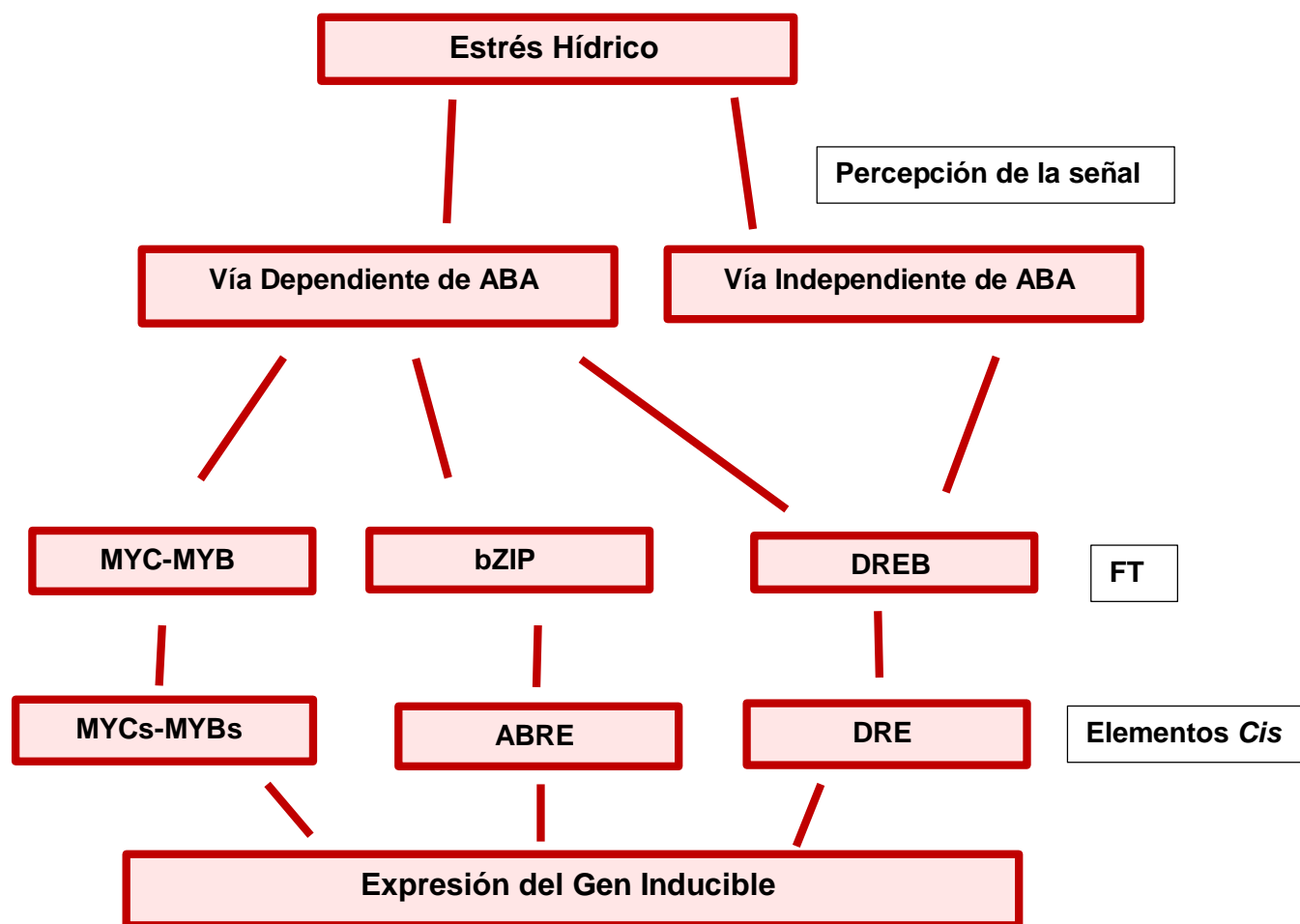
### 1.1.7 Elementos *cis* reguladores

Son secuencias cortas de DNA funcionales, encontradas tanto en el promotor proximal como distal, en el que su distribución y presencia contribuye a un patrón de expresión temporal y espacial de un gen en particular. Los factores de transcripción se unen a estas secuencias, por lo que se les conoce como sitios de unión de factores de transcripción, TFBS, del inglés *transcription factor binding sites*, también llamados motivos de unión al DNA o elementos reguladores en *cis* (Vedel *et al*, 2011); siendo su identificación y entendimiento de sus funciones importantes para la elucidación de los mecanismos en los cuales las células perciben y responden correctamente a su entorno (MOCKLER, 2009).

En la regulación transcripcional dependiente de ABA, los genes que son inducibles por deshidratación y que son dependientes de ABA, contienen elementos de respuesta a ABA, conocidos como ABREs (Elementos de respuesta al ácido abscísico). Estos elementos de ADN tienen una secuencia conservada de al menos 8-10 pares de bases con una secuencia núcleo central ACGT, conocido como caja G encontradas en la región de los promotores (ROCK C, 2000). Por tanto, la biosíntesis de ABA se induce en situación de deshidratación y activa la expresión de genes mediante dos vías (Figura 3). En la primera, las proteínas de la familia bZIP son activadas e inducen la expresión de genes que contienen potenciales elementos de respuesta a ABA (ABREs) en sus promotores. La segunda vía requiere la biosíntesis de otros factores de transcripción como MYC y MYB, sintetizados en respuesta al ABA y actúan regulando de manera cooperativa la expresión de genes que contienen elementos *cis* en sus promotores que son reconocidos por estos factores. Estas dos vías parecen actuar en la respuesta lenta y adaptativa de las plantas a la deshidratación. Una tercera vía independiente de ABA, media una respuesta rápida al estrés por deshidratación a través de la activación del

elemento *cis* DRE (GCCGAC) o Elemento de Respuesta a la Deshidratación (YOSHIDA *et al*, 2015).

**Figura 3** - Red de regulación transcripcional implicados en la respuesta al estrés.



FUENTE: ADAPTADO DE LATAN *et al*, 2011.

### 1.1.8 Análisis *in silico*

El análisis *in silico* designa un tipo de experimento que no se hace directamente en el ser vivo o invitro, sino directamente en un ordenador, consideradas como simulaciones informáticas de procesos biológicos; es así que un diseño *in silico* con la aplicación de la tecnología de computadores a la gestión y análisis de datos biológicos de genes semi-sintéticos, constructos genéticos o casetes de expresión, se convierte en herramienta versátil que permite al investigador manejar fácilmente variables en la secuencia de ADN y de los constructos que con el uso de técnicas moleculares convencionales sería muy costoso (LI , 2013).

Una estrategia para el análisis de promotores implica el uso de herramientas bioinformáticas para explorar su secuencia, basada en la búsqueda de secuencias conservadas de DNA del banco de datos. De esta manera, se pueden detectar motivos o elementos ya descritos con una función determinada, debido a la información recopilada en las bases de datos (YAMAMOTO, 2007).

Varias herramientas *in silico* fueron utilizadas a fin de identificar FT de la familia WRKY de la base de datos completa del proteoma de *P. glaucum*, prediciendo su participación en diversos procesos biológicos, moleculares y como componente celular; su distribución cromosómica y perfil transcripcional, por la presencia de elementos *cis* específicos en regiones promotoras inducibles a un tipo estrés abiótico, cuyos hallazgos proporcionaron una base para una mayor caracterización funcional e identificación del mecanismo regulador de PgWRKY (CHANWALA J , 2020). A continuación, se citan algunos ejemplos:

- Con el auxilio de herramientas bioinformáticas se analizó el rol de varios FT en plantas en respuesta a un estrés biótico dado por bacterias, identificando elementos *cis* regulatorios relacionados con la regulación de los miRNAs (PÉREZ & LÓPEZ 2013)
- Otro trabajo realizado con miRNAs, identificó computacionalmente pequeñas secuencias de RNA y sus blancos en *M. esculenta*, encontrando elementos *cis* sensibles al estrés por sequía y motivos

asociados a fitohormonas en sus regiones promotoras (PATANUN, 2013).

- Otros autores analizaron *in silico* elementos reguladores en familia de genes encargados del transporte de sacarosa en arroz y *Arabidopsis*, usando las bases de datos informáticos; augurando los elementos reguladores que actúan en *cis*, por medio de un escaneo de 1,5kpb de la región 5' de los genes transportadores de sacarosa desde el codón de inicio de la traducción. Los resultados efectivamente mostraron la identificación de elementos reguladores en *cis* relacionados con varias vías metabólicas y de señalización (IBRAHEEM, 2010).
- Fue descrito que la proteína quinasa interactiva CBL (CIPK6) estaba involucrada en la respuesta de la planta al estrés salino y al ABA por rastreo *in silico* e identificación potencial de elementos *cis* de su región promotora (CHEN, 2013).
- Fueron caracterizadas familias génicas de lectina (proteína importante en la defensa de las plantas y la señalización celular) a partir del análisis de la región promotora, encontrando que la mayoría de las familias respondían a condiciones de luz y presentaban una expresión específica según el órgano o tejido de la planta. La presencia de elementos *cis* asociados con hormonas, como el ácido jasmónico, en las regiones promotoras de todos los genes de lectina indicaron su importancia en la respuesta a la señalización y estrés del cultivo (HA & WYCKHUYS, 2017).
- Un análisis *in silico* de regiones promotoras de genes inducidos por frío en arroz y *Arabidopsis* determinó que la mejor manera para estudiar la regulación de procesos celulares debe tener como objetivo principal la identificación de combinación de motivos, y, como segundo, buscar sitios de unión individuales que presenten sobreexpresión (Lindlöf et al., 2009); destacándose a los promotores inducibles por factores bióticos y abióticos, tejido específico o inducibles como un sistema atractivo para la elaboración de promotores sintéticos (ZHANG, 2013).

## 1.2 Justificativa

Las particularidades nutricionales y la rusticidad presentada por un cultivo no tradicional ni modelo como lo es *P. glaucum*, lo hacen atractivo para comprender su base genética en respuesta a condiciones de estrés hídrico, factor que afecta a la producción alimentaria en general.

El análisis *in silico* que designa a un tipo de experimento bioinformático nos permite rastrear y analizar los genes expresos como mecanismo de tolerancia al estrés hídrico en millete, volcándose esta como una estrategia complementaria al uso de técnicas moleculares convencionales onerosas a fin de indagar en las regiones promotoras de genes estructurales, que promueven a una defensa constitutiva; y genes reguladores que pueden alterar la expresión de un gen.

Entre los genes estructurales de millete podemos citar a las porinas, las Proteínas de canal de aniones dependiente de voltaje (PgVDAC), involucradas en el mantenimiento de homeostasis celular como respuesta al estrés hídrico y salino (Desai M *et al*, 2006); sin embargo. el análisis de secuencia de genes que codifican para VDAC de plantas se han reportado solo en pocas especies, con un bajo número de publicaciones (HOMBLÉ F, 2012).

Referente a los genes reguladores, los factores de transcripción bZIP en *Arabidopsis*, tomate (*Solanum lycopersicum*), soja (*Glycine máx*) y arroz (*Oryza sativa*) han demostrado una mayor expresión ante un estrés por sequía, pero pocos factores de transcripción bZIP han sido explorados como candidatos potenciales en aplicaciones de mejoramiento genético para tolerancia a sequía en cultivos (LINDEMOSE *et al*, 2013).

Es por tanto que, bajo lo citado, el análisis de VDAC como gen estructural y los factores de Transcripción bZIP como genes reguladores fundamentados en el estudio funcional de sus regiones promotoras se muestra promisorio en detectar motivos o elementos *cis* que revelen estrategias de respuesta peculiares, sustentándose en la hipótesis: “existen elementos *cis* involucrados en vías de respuesta al estrés hídrico que modulan la expresión de PgVDAC como gen estructural y de los FT bZIP como gen regulador”. bajo las siguientes interrogantes;

¿Existen elementos *cis* que modulan la expresión de PgVDAC como gen estructural y de los FT bZIP como gen regulador en *P. glaucum* involucrados en vías de respuesta al estrés hídrico?

¿Los elementos *cis* afines presentes en la región promotora del gen estructural PgVDAC regulan su expresión bajo condiciones de estrés hídrico?

¿Están estos elementos identificados en PgVDAC presentes e involucrados en la misma vía de respuesta de VDAC ortólogos en plantas modelos?

¿Los elementos *cis* encontrados en las regiones promotoras de la familia de genes reguladores FT bZIP modulan su expresión bajo condiciones de estrés hídrico?

¿Estos elementos *cis* identificados en FT bZIP de *P. glaucum* están presentes e involucrados en la misma vía de respuesta de bZIP ortólogos en plantas modelos?

## 2 OBJETIVOS

### 2.1 General

Analizar *in silico* la estructura de regiones reguladoras y expresión de PgVDAC (gen estructural) y FTbZIP (genes reguladores) en *P. glaucum* bajo condiciones de estrés hídrico.

### 2.2 Específicos

➤ Explorar *in silico* la regulación de expresión del gen estructural VDAC, identificando los elementos *cis* afines presentes en la región promotora involucrados en vías de respuesta al estrés hídrico.

➤ Identificar e indagar en la región promotora de VDAC ortólogos en plantas modelos razonando su participación en el estrés hídrico.

➤ Cribar en todo el genoma de *P. glaucum* probables genes reguladores de la familia bZIP utilizando secuencias de referencias de plantas modelos, y rastrear en la región promotora de las bZIP diferencialmente expresas filtradas por análisis *in silico* de un Transcriptoma publicado de estrés hídrico en *P. glaucum*.

➤ Examinar en la región promotora de bZIP ortólogos identificados en plantas modelos razonando su participación en vías de respuesta al estrés hídrico.

### 3 MATERIALES Y MÉTODOS

#### 3.1 Rastreo de la región promotora del gen estructural PgVDAC

El gen PgVDAC accedido del Banco de datos del NCBI (<https://www.ncbi.nlm.nih.gov/>) abarca 2967 pb, consta de cinco intrones y seis exones; un marco de lectura abierto (ORF) de 834 pb, ubicación y acceso disponible en la Tabla 1, la cual codifica para una porina de 277 aminoácidos.

Se seleccionó la región de ~1500 pb secuencia río arriba (*upstream*) del marco abierto de lectura del gen estructural PgVDAC (nucleótido -1), posición 144.404.454 a 144.405.954 del cromosoma 5, para la identificación de los elementos involucrados en el estrés hídrico de la región promotora PgVDAC mediante el uso del programa plantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)

PlantCARE se presenta como una base de datos de elementos reguladores que actúan en *cis* de plantas y un portal a herramientas para el análisis *in silico* de secuencias promotoras (LESCOT M *et al*, 2002).

**Tabla 1** - Ubicación de secuencias inferidas de PgVDAC

Nombre	Ubicación	Acceso
PgVDAC	Cromosoma 5. 144.405.954-144.408.919	AY823551
Región Promotora PgVDAC	Cromosoma 5. 144.404.454-144.405.954	

FUENTE: EL AUTOR, 2021

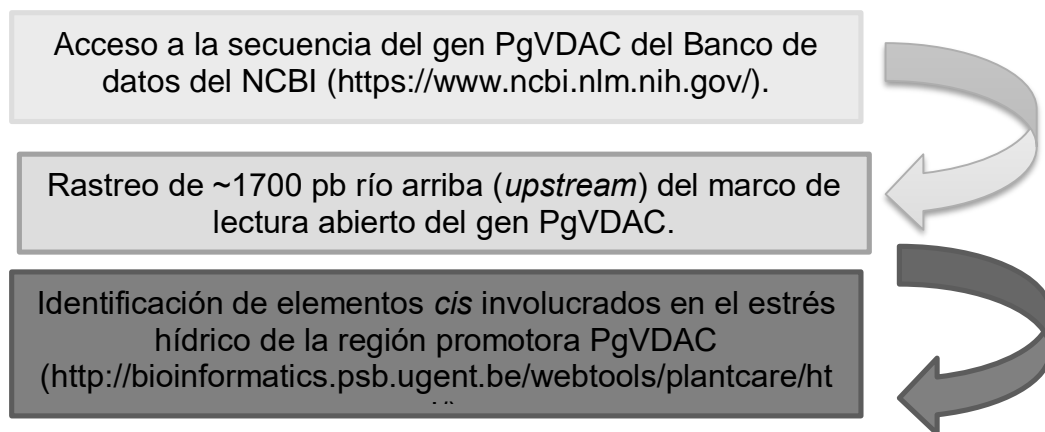
#### 3.2 Identificación y Análisis de la Región Promotora de VDAC ortólogos

La secuencia del gen PgVDAC fue sometida al alineamiento local de secuencias con la utilización de la herramienta BLAST a fin de hallar VDAC ortólogos contra la base de datos de *Oryza sativa* (<https://phytozome.jgi.doe.gov>) y *Arabidopsis thaliana* (<https://www.arabidopsis.org/cgi-bin/Blast/TAIRblast.pl>). BLAST, de la abreviatura en inglés *Basic Local Alignment Search Tool* es una herramienta de comparación de secuencias que rápidamente es capaz de encontrar secuencias de

nucleótidos o de proteínas similares a la de entrada en formato FASTA. Puede ser utilizada para encontrar genes en un genoma, predecir la función de proteínas, identificar los homólogos, entre otros (ALTSCHUL S *et al*, 1997).

Se rastreó secuencia arriba de 1500-1700 pb de los genes VDAC ortólogos identificados a fin de indagar en la región *upstream* verificándose los elementos reguladores en *cis* involucrados en vías de respuestas al estrés hídrico. El rastreo diferencial de 1500 a 1700 pb secuencia arriba -1 se debe por la seguridad de encontrar el máximo de elementos disponibles y algunos suelen estar a grandes distancias. En la figura 4 se detalla el esquema bioinformático utilizado para examinar la región promotora de PgVDAC, el mismo esquema fue utilizado una vez identificados los VDAC ortólogos en modelos vegetales de arroz y *Arabidopsis*.

**Figura 4** - Esquema *in silico* del gen funcional PgVDAC.



FUENTE: EL AUTOR, 2021

### 3.3 Gen regulador de la familia bZIP

#### 3.3.1 Cribado *in silico* en todo el genoma de *P. glaucum* de probables genes reguladores FT bZIP

**3.3.1.1** La búsqueda de secuencias de FT bZIP en *P. glaucum* se inició con la construcción del Perfil HMM (Hidden Markov Model). Para ello, las secuencias proteicas de FT bZIP involucradas en el estrés hídrico se usaron como queries y fueron descargadas de uniprot (<https://www.uniprot.org/>) un total de 101 secuencias de referencia (Anexo 1) de *Oryza sativa* (18 bZIP), *Arabidopsis thaliana* (25 bZIP) y *Setaria*

*itálica* (58 bZIP). Una vez descargadas las secuencias bZIP referenciales de *O. sativa*, *S. itálica* y *A. thaliana* se alinearon con el programa Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) basado en la alineación de secuencias múltiples; obteniendo la secuencia conservada con el programa Jalview (<https://www.jalview.org>) y utilizada para generar el perfil HMM (FINN R. *et al*, 2011) con la herramienta HMMER (<https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan>). La secuencia conservada generada se sometió al tBlastn contra la base de datos del Genoma de *P. glaucum* (*Cenchrus americanus*) utilizando la herramienta del NCBI, considerando un ID $\geq$ 45% y un E-value de 0.02 a valores  $\leq$ 1.

La herramienta bioinformática HMMER se utiliza en conjunto con una base de datos de perfiles HMM previamente preparada para analizar secuencias biológicas. HMM, de las siglas en inglés Hidden Markov Model o Modelo oculto de Markov permite determinar la probabilidad con que cada secuencia de una base de datos haya sido generada a partir de él. De esta manera pueden detectarse miembros de una familia de proteínas, o secuencias que contengan ciertos dominios o motivos funcionales, así como la predicción de genes, entre otros. Algunas de las bases más importantes de familias y dominios, como Pfam o Interpro, usan HMMER como base para la construcción de sus propios perfiles HMM (EDDY, 2003).

**3.3.1.2** Los términos de ontología génica (GO) para las proteínas o FT bZIP putativas reconocidas se identificaron utilizando parámetros predeterminados con la herramienta Blast2GO v5 prediciéndose los procesos biológicos y moleculares en los que participan.

Blast2GO es una herramienta bioinformática que permite la caracterización funcional *in silico* de una secuencia a través de técnicas de minería de datos. Utiliza la herramienta Blast contra la BD (Base de datos) no redundante (nr) del NCBI para encontrar los homólogos, mapea para recuperar términos GO y anota para seleccionar funciones confiables (CONESA A. *et al*, 2005). Agrupa las ontologías en categorías de función molecular, asociada con la actividad bioquímica del producto génico, como “actividad transportadora” o “regulación enzimática”; proceso biológico, en relación al objetivo biológico que lleva a cabo el producto génico, como por ejemplo el término

“transducción de señales” o “fotosíntesis”; y la ontología componente celular correspondiente al lugar donde el producto génico lleva a cabo su función, como por ejemplo la “membrana nuclear” o el “ribosoma” (ASHBURNER *et al*, 2000).

**3.3.1.3** El conjunto de datos de un Transcriptoma publicado en el NCBI como BioProject ID: PRJNA419859, fue utilizado para el análisis de expresión génica diferencial. El experimento esquematizado en la figura 5 utilizó como material biológico a dos líneas endogámicas de millete tolerantes a la sequía ICMB843 e ICMB863 cuyas plántulas con 21 días después de la siembra (DDS) fueron sometidas a estrés hídrico por cese de riego total por 5 días y tomadas muestras de tejido de raíz con sus respectivos controles a los 26 días después de la siembra fueron utilizadas para el análisis de expresión diferencial génica (DUDHATE A *et al*, 2018).

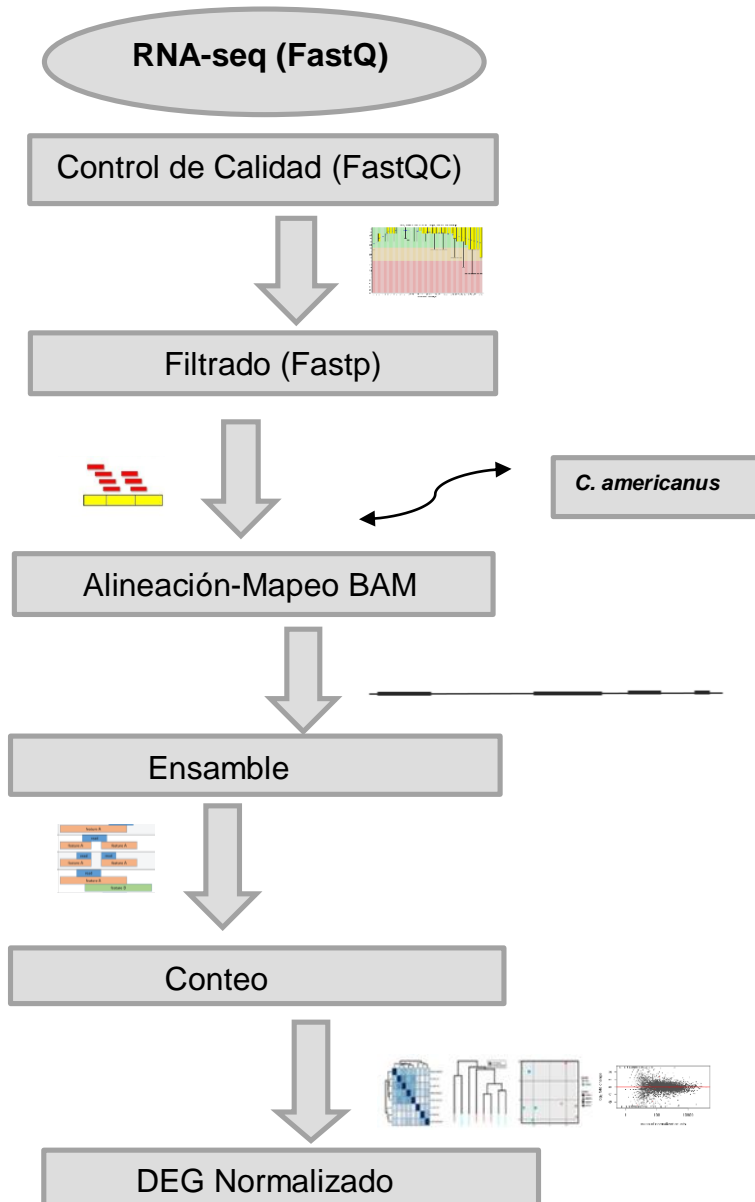
**Figura 5** - Esquema del experimento de estrés del Transcriptoma publicado



FUENTE: EXTRAÍDO Y ADAPTADO DE DUDHATE A *et al*, 2018

El análisis transcriptómico se realizó siguiendo el esquema de la Figura 6, utilizando herramientas bioinformáticas del Galaxy (AFGAN E et al, 2018).

**Figura 6** - Esquema de trabajo para el análisis DEG



FUENTE: EL AUTOR, 2021

El secuenciamiento de RNA (RNA-Seq) genera lecturas de DNA a partir del RNA (transcriptos) de un organismo. Es así, que el análisis de un transcriptoma requiere de eficientes herramientas bioinformáticas a fin analizar la cantidad de información de lecturas generadas que permiten la identificación de genes, sus isoformas alternativas y la variación de sus perfiles de expresión entre tipos celulares, tejidos y órganos, frente a una determinada condición externa, lo cual lleva a entender la complejidad de procesos y funciones biológicas de cualquier organismo (FANG, 2012).

El programa informático FastQC permite realizar controles de calidad en archivos de secuencias, el contenido de G/C, si hay bases catalogadas como N o si hay secuencias que se encuentren sobre representadas (ANDREWS, S. sf). La herramienta Fastp retira los adaptadores, fragmentos de ADN de doble cadena con secuencias consenso donde se alinearon los cebadores para las etapas de amplificación del templado y los cebadores para la secuenciación propiamente dicha; y filtra las lecturas incorrectas de baja calidad, cortas o con muchas N (CHEN S, 2018).

Una vez ya alzado el genoma de referencia de *P. glaucum* (*C. americanus*) y anotación del transcriptoma se procedió al mapeo con el uso de HISAT2, el cual se presenta como un programa de alineación sensible para mapear lecturas de RNA-seq; permitiendo así que las reads o lecturas se alineen contra el genoma de referencia fidedignamente (KIM D, 2015).

StringTie toma como entrada el archivo BAM de lecturas de RNA-seq de la salida de HISAT2, actuando como un ensamblador rápido y altamente eficiente. Se realizó la fusión de los ensamblados generados con la anotación del transcriptoma de referencia utilizando StrinTie Merge (KOVAKA S, 2019). Para el conteo se utilizó la herramienta FeatureCounts, la cual produce una tabla que contiene lecturas contadas por gen (LIAO Y, 2013).

El análisis de conteos de lecturas por gen se realizó con la ayuda de DESeq2 (LOVE M *et al*, 2014) a fin de conocer el diferencial de expresión génica (DEG) bajo condiciones de estrés hídrico y control (sin estrés hídrico) tanto para las líneas endogámicas ICMB843 como ICMB863. Una vez confirmada la presencia de los genes PgbZIP putativos en el transcriptoma analizado, se seleccionaron aquellos con mayor DEG. Posteriormente, los valores de expresión fueron transformados a su logaritmo

(base 2), comparando los cambios de niveles de expresión (LogFC) o fold changes (Log<sub>2</sub>) que no son más que cocientes de los valores de expresión; es decir la expresión media del gen FT bZIP putativo seleccionado en condiciones de estrés hídrico versus su expresión media bajo condiciones controladas (sin estrés hídrico).

### 3.4 Rastreo de la Región promotora FT bZIP putativos seleccionados

3.3.1 Se dio acceso a las secuencias bZIP putativas seleccionadas según su expresión diferencial deducida del Banco de datos del NCBI (<https://www.ncbi.nlm.nih.gov/>).

3.3.2 Los bZIP reguladores seleccionados se exploraron con el uso de NCBI Genome Data Viewer, a fin de rastrear ~1700 pb secuencia río arriba (*upstream*) del gen regulador.

3.3.3 Los elementos *cis* involucrados en el estrés hídrico de la región promotora de los genes reguladores bZIP seleccionados se identificaron con el uso de plantCARE.

3.3.4 Se procedió al rastreo y análisis de las regiones promotoras (~1700 pb *upstream*) de bZIP ortólogos en *A. thaliana* y *O. sativa*.

A fin de analizar las estructuras de transcripción de las probables PgbZIP diferencialmente expresas y con elementos *cis* involucrados en el estrés hídrico que resultaron sugestivas fueron visualizadas a través de IGV Sashimi Plot (KATZ et al, 2015), cuya representación gráfica indica a las alineaciones de exones como densidades de lectura, la unión de empalme como arcos que conectan un par de exones.

## 4 RESULTADOS

### 4.1 Análisis *in silico* de la Región promotora del gen estructural PgVDAC

Los elementos reguladores en *cis* funcionales involucrados en las vías de respuesta al estrés hídrico encontrados en la región *upstream* (río arriba) del gen PgVDAC fueron los motivos ABREs, MBS-MYB, MYC y DRE.

### 4.2 Identificación y Análisis de las Regiones Promotoras de VDAC Ortólogos

Mediante la utilización de la herramienta Blast se identificó a ATVDAC2 (AT5G67500) en *Arabidopsis thaliana* y a OsVDAC2 (OsJ\_19349) en *Oryza sativa*, cuyos porcentajes (%) de identidad pueden verse en la Tabla 2 (Anexos 2 y 3).

**Tabla 2** - VDAC ortólogos

Gen	E-value	Identidad (%)	Acceso
<b>ATVDAC2</b>	6e <sup>-60</sup>	47	NM_001345833
<b>OsVDAC2</b>	2e <sup>-49</sup>	45	AJ251562.1

FUENTE: EL AUTOR, 2021

La estructura de la proteína VDAC y las propiedades del canal se han conservado bien durante la evolución, aunque los análisis *in silico* revelaron características moleculares específicas en algunas secuencias de VDAC en *A thaliana*, ya que se pudo identificar isoformas diferentes de VDAC. Para la determinación de los sitios funcionales de los cuatro genes (isoformas) VDAC se generaron plantas transgénicas de *Arabidopsis* que llevaban el promotor VDAC respectivo para cada isoforma, en la cual la actividad del promotor de ATVDAC2 se elevó constitutivamente en las partes aéreas de la planta y en las puntas de las raíces, destacándose su importancia en el desarrollo de hojas, polen y el potencial de la membrana mitocondrial en estado estable; pudiendo estar involucrado en el desarrollo temprano de plántulas mediado por ABA y en la resistencia a enfermedades. El análisis fenotipo adicional de las plantas transgénicas ATVDAC2 estables confirmó que esta participó en la señalización de ABA

durante la germinación de la semilla y el desarrollo de cotiledones verdes (CHIKA TATEDA, 2011).

Las experimentaciones mencionadas por AL BITAR F *et al* (2003) destacan la expresión de OsVDAC2 de forma general, en raíces, tallos, hojas y polen, con una elevada expresión durante los primeros días después de la germinación y luego disminuye con el tiempo; y no es inducido por estrés osmótico. Sin embargo, destacan una expresión mejorada rápidamente de OsVDAC2 en las raíces durante el período de recuperación después de la liberación del estrés. Esta expresión mejorada no se correlaciona con la cantidad de proteína VDAC detectada en las raíces, lo que sugiere una regulación pos-transcripcional.

En la Tabla 3 se especifican las regiones promotoras seleccionadas de las ATVVDAC2 y OsVDAC2 de *Arabidopsis* y arroz respectivamente.

**Tabla 3** - Ubicación de VDAC ortólogas identificadas en *A. thaliana* y *O. sativa*

Nombre	Ubicación	Acceso
<b>ATVVDAC2</b>	Cromosoma 5: 26.935.223-26.937.123	NC_003076.8
<b>Región Promotora ATVVDAC2</b>	Cromosoma 5: 26.933.575-26.935.264	
<b>OsVDAC2</b>	Cromosoma 5: 26.628.154-26.629.343	NC_029260.1
<b>Región Promotora OsVDAC2</b>	Cromosoma 5. 26.626.654 a 26.628.334	

FUENTE: EL AUTOR, 2021

Los resultados resumidos en la Tabla 4 indican la presencia de motivos ABREs, MYB (MBS) y MYC en los 3 organismos.

En PgVDAC las secuencias de los elementos *cis* fueron las siguientes; ABREs ACGTG/AACCGG/TACGTG; DRE GCCGAC; MYC CAATTG/CATTTG/CATGTG; MBS CAACTG; MYB CAACCA/TAACCA. En ATVVDAC2; ABREs ACGTG/GACACGTGGC; DRE GCCGAC; MYC CAATTG/CATGTG/TCTCTTA; MBS CAACTG; MYB CAACAG/TAACCA/CCGTTG; mientras que en OSVDAC2 ABREs ACGTG/CACGTG; MYC CATTTG; MYB CAACAG/TAACCA/CAACCA/TAACCTG.

**Tabla 4** - Motivos *cis* presentes en VDAC de *A. thaliana*, *O. sativa* y *P. glaucum*

	ABREs	MYC	MYB	MBS	DRE
ATVDAC2	4	3	5	1	1
OsVDAC2	2	1	4	0	0
PgVDAC	4	5	2	3	1

FUENTE: EL AUTOR, 2021

#### 4.3 Cribado *in silico* en todo el genoma de *P. glaucum* de probables genes reguladores de la familia bZIP

La construcción del Perfil HMM basado en la alineación de secuencias múltiples generó una secuencia conservada para los organismos modelos vegetales utilizados *Arabidopsis*, *Oryza* y *Setaria*, visualizada en la figura 7, 8 y 9 respectivamente, atribuido al dominio bZIP conservado que utiliza PFAM (PF00170) descritas en la Tabla 5 (SONNHAMMER E *et al*, 1998).

**Tabla 5** – Perfil HMM basados en el grado de conservación de dominio bZIP

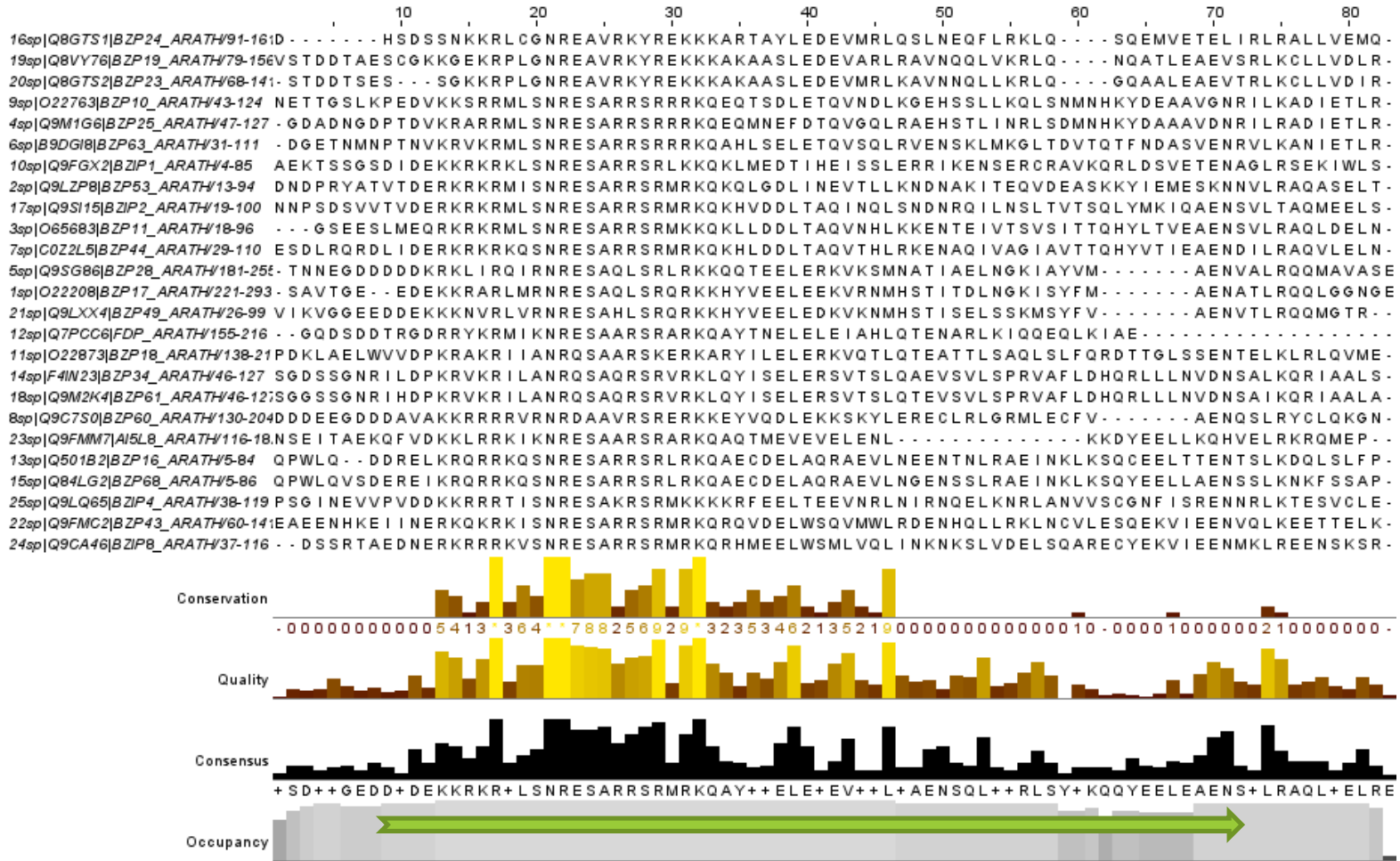
Organismo	PFAM <sup>1</sup>	E-value
<i>A. thaliana</i>	PF00170	2.8-e <sup>-21</sup>
<i>O. sativa</i>	PF00170	4.6e <sup>-15</sup>
<i>S. itálica</i>	PF00170	1.6e <sup>-14</sup>

<sup>1</sup>Familia proteica PF00170 atribuida a dominio bZIP

FUENTE: EL AUTOR, 2021

Al someter cada perfil bZIP generado de *Arabidopsis*, *Oryza* y *Setaria* a la herramienta tBlastn contra la base genómica de *P. glaucum* identificó 52 probables secuencias bZIP representadas en la Tabla 6.

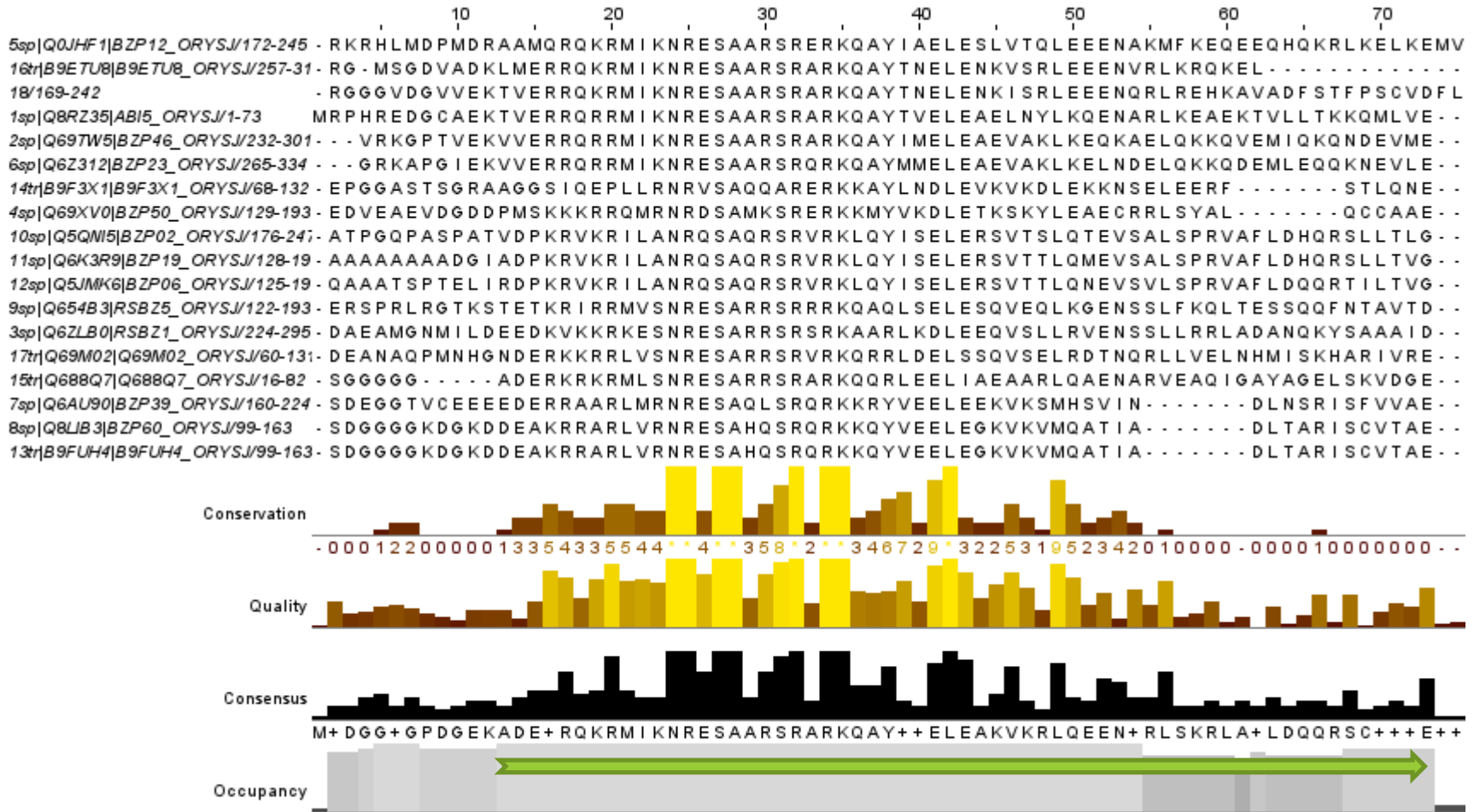
**Figura 7 – Construcción del Perfil HMM basados en el grado de conservación de la alineación de secuencias múltiples en Arabidopsis**



→ DDEKKRKRKLSNRESARRSRMRKQAYTDELEDEVNQLQAENSQLLARLSYLKQQYEELEAENS, secuencia conservada (495-558) atribuida a la PFAM bZIP

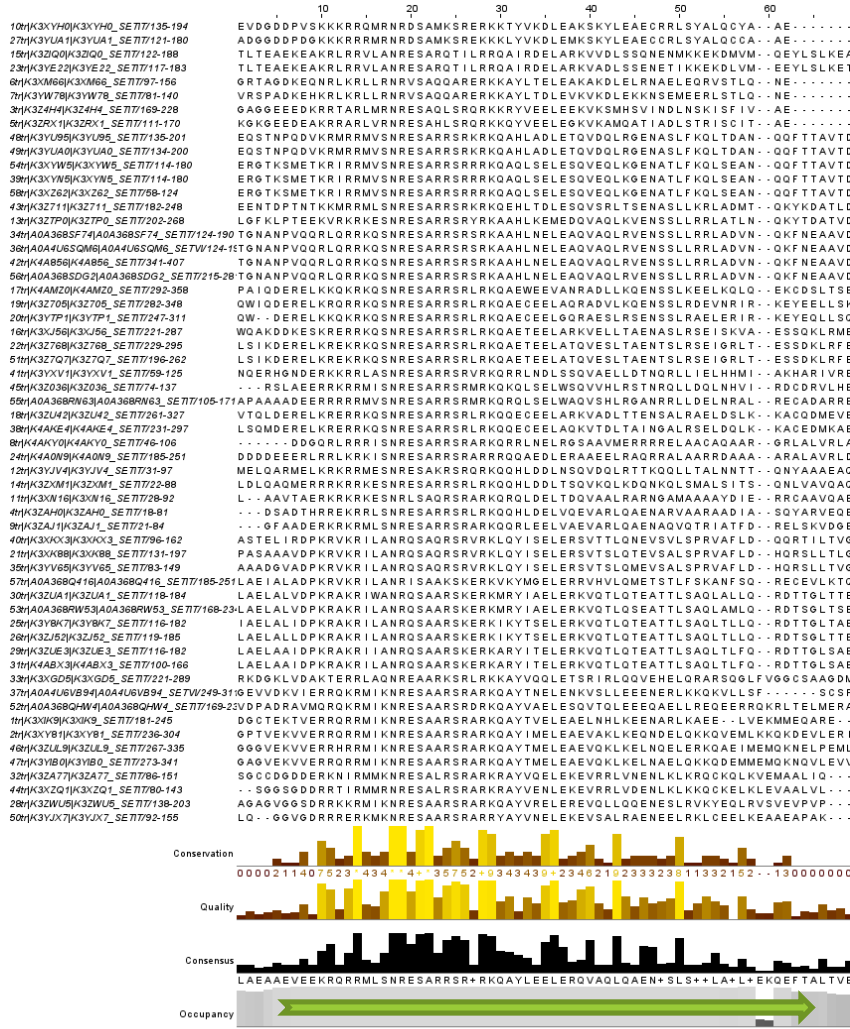
FUENTE: EL AUTOR, 2021

**Figura 8 - Construcción del Perfil HMM basados en el grado de conservación de la alineación de secuencias múltiples en *Oryza***



—————> ADEKRQKRM IKNRESAARS RARKQAY IEELEAKVKRLQEENARLSKRLADLDQQR SCLTAEF, secuencia conservada (424-486) atribuida a la PFAM bZIP FUENTE: EL AUTOR, 2021

Figura 9 - Construcción del Perfil HMM basados en el grado de conservación de la alineación de secuencias en *S. Itálica*



→ EVEEKQRRLMSNRESARRSRARKQAYLEELERQVAGLQAEENASLSAQLADLNEKQEFTA, secuencia conservada (633-693)

atribuida a la PFAM bZIP FUENTE: EL AUTOR, 2021

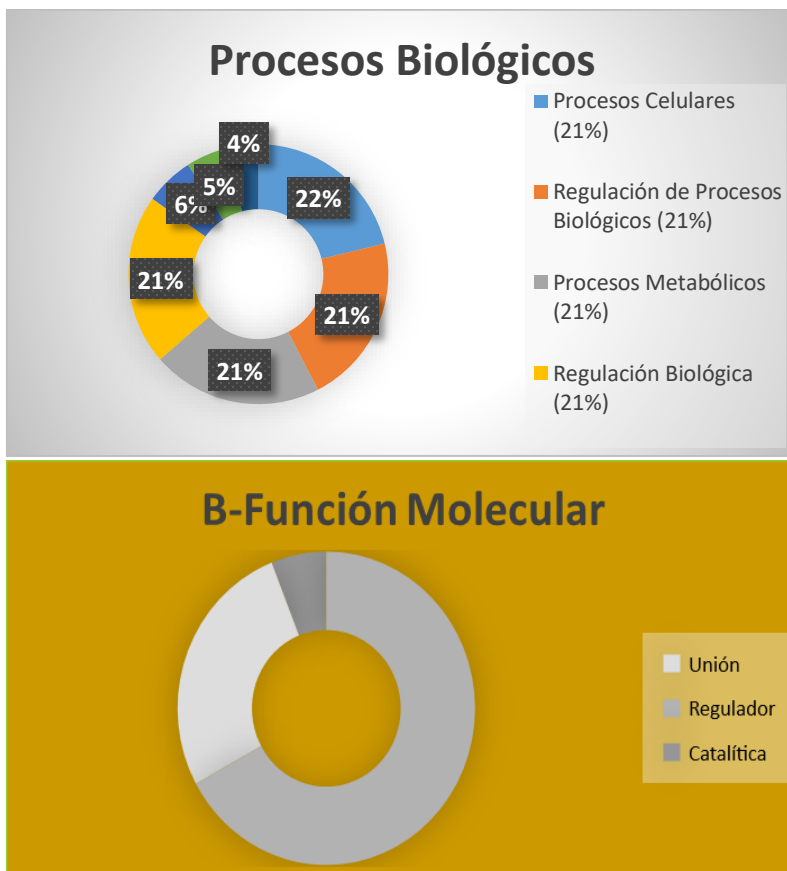
**Tabla 6 – 52 probables FT bZIP identificados**

<b>*PgbZIP</b>	<b>Acceso</b>	<b>%Id</b>	<b>*PgbZIP</b>	<b>Acceso</b>	<b>%Id</b>	<b>*PgbZIP</b>	<b>Acceso</b>	<b>%Id</b>
Pg1	SETCM007983.2	86.36	Pg19	ARACM007987.2	76	Pg37	ORCM007982.2	88.12
Pg2	ARACM007983.2	75.44	Pg20	ARACM007986.2	84	Pg38	ORCM007982.2	77.78
Pg3	ARACM007983.2	73.68	Pg21	ORCM007988.2	77.42	Pg39	ORCM007982.2	88.89
Pg4	ARACM007983.2	82.14	Pg22	ORCM007988.2	72.5	Pg40	ORCM007982.2	74.19
Pg5	ARACM007983.2	65.71	Pg23	ORCM007988.2	77.78	Pg41	ORCM007982.2	75.86
Pg6	ARACM007983.2	72.41	Pg24	ORCM007988.2	52	Pg42	ORCM007982.2	67.57
Pg7	ARACM007983.2	65.08	Pg25	ORCM007988.2	66.67	Pg43	ORCM007985.2	47.62
Pg8	ARACM007983.2	80	Pg26	ORCM007987.2	63.89	Pg44	ORCM007985.2	73.33
Pg9	ARACM007984.2	70.97	Pg27	ORCM007987.2	75	Pg45	ORCM007985.2	77
Pg10	ARACM007982.2	69.39	Pg28	ORCM007987.2	65	Pg46	ORCM007984.2	80
Pg11	ARACM007982.2	68	Pg29	ORCM007987.2	74.19	Pg47	ORCM007984.2	72.34
Pg12	ARACM007982.2	46	Pg30	ORCM007983.2	74.36	Pg48	ORCM007984.2	97.73
Pg13	ARACM007982.2	68.89	Pg31	ORCM007983.2	71.11	Pg49	ORCM007986.2	81.08
Pg14	ARACM007982.2	68.89	Pg32	ORCM007983.2	81.08	Pg50	ORCM007986.2	72.34
Pg15	ARACM007982.2	53.12	Pg33	ORCM007983.2	62.26	Pg51	ORCM007986.2	76.32
Pg16	ARACM007985.2	68.89	Pg34	ORCM007983.2	71.11	Pg52	ORCM007986.2	80
Pg17	ARACM007987.2	68.57	Pg35	ORCM007982.2	77.14	*Probables 52 FT bZIP FUENTE: EL AUTOR, 2021		
Pg18	ARACM007987.2	68.42	Pg36	ORCM007982.2	86.67			

### 4.3.1 Anotación Funcional

Mediante el programa Blast2GO se predijeron los términos ontológicos (GO) (Anexo 4) de las 52 probables PgbZIP, en la categoría de función biológica con una participación prevista mayor del 21 % en la regulación de los procesos biológicos, celulares, metabólicos y regulación biológica mientras que la categoría de función molecular representada sugiere 67% de actividad reguladora, visualizadas en la Figura 10, coincidente, ya que se trata de FT que regulan la expresión de un gen. Al unirse a un elemento en *cis* del ADN estas proteínas bZIP (FT) facilitan la unión de otros componentes de la maquinaria transcripcional como la ARN polimerasa y otros FT, regulando la expresión de genes de respuesta (BANERJEE & ROYCHOUDHURY, 2015).

**Figura 10 - Ontología Genética**



FUENTE: EL AUTOR, 2021

### 4.3.2 Análisis de la Expresión Diferencial de genes bZIP putativos

Los datos de lecturas del secuenciamiento de las líneas ICM843 control y estresada, ICMB863 control y estresada en formato FastQ fueron sometidos al análisis de calidad FastQC (Figura 11), generando más de 260 millones de lecturas.

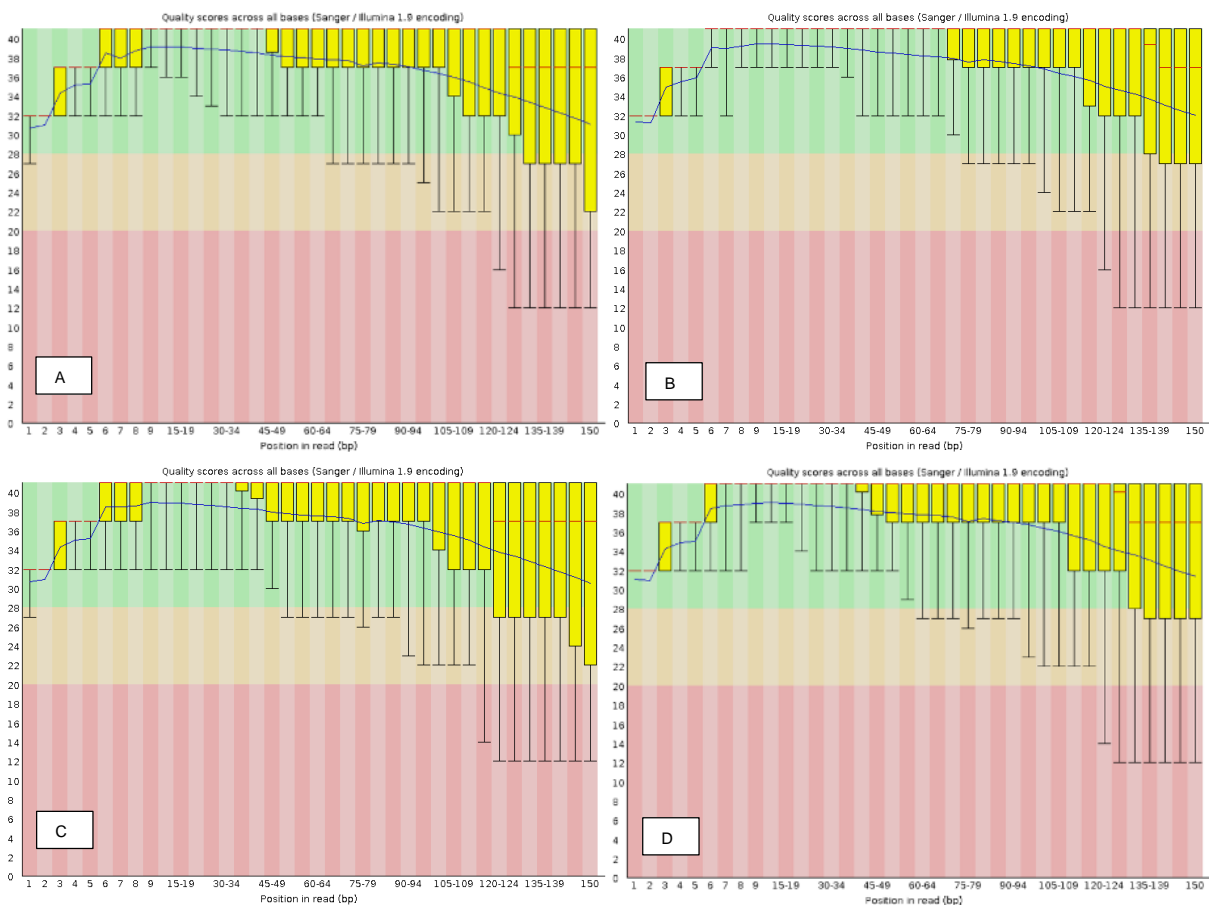
La calidad del proceso de identificación de cada nucleótido durante la secuenciación se mide en escala Phred. Un valor Q en esta escala está relacionado logarítmicamente con la probabilidad P de que se haya cometido un error al identificar al nucleótido en cuestión; donde un valor de  $Q = 10$  se corresponde con un valor  $P = 0,1$ , es decir que la probabilidad de que el fragmento original NO haya tenido la base reportada en dicha posición es 1 en 10; equivalentemente, indica que la exactitud en la estimación del nucleótido es del 90 %; considerando aquí un valor de phred  $\geq 20$  como buena calidad (VAN-VERK *et al*, 2013), seguido del filtrado fastp y alineación o mapeo.

El análisis *in silico* de los datos del transcriptoma pueden verse resumidos en la Tabla 7, donde la línea endogámica ICMB863 sometida al estrés hídrico obtuvo 72% de mapeado, y su control ICMB863 sin estrés hídrico 76,68%; mientras que la ICMB843 sometida al estrés 80% y su control o testigo 77% de lecturas mapeadas.

El porcentaje de lecturas mapeadas en el transcriptoma publicado asignadas fue menor en las muestras de ICMB 843 estresado (13,2%) por sequía comparativamente con las otras muestras; coincidente con lo planteado en el trabajo realizado por DUDHATE A *et al*, (2018) argumentando que esto podría deberse a que el estrés por sequía aumentaría la variación en las transcripciones en ICMB 843, ya que esta línea endogámica bajo condiciones de invernadero se muestra más tolerante a la sequía que ICMB863.

El conjunto de datos del Transcriptoma analizado utilizando herramientas bioinformáticas del Galaxy pueden ser accedidos al url: <https://usegalaxy.org/u/laura.farias.2019/h/in-silico-analysis-of-the-entire-p-glaucum-genome-identifies-regulatory-genes-of-the-bzip-family-modulated-in-response-pathways-to-water-stress>.

**Figura 11 - Calidad de Lecturas**



Valores de Phred de calidad para A- ICMB863 Control (3 réplicas) B- ICMB863 Estresado (3 réplicas) C- ICMB843 Control (3 réplicas) D- ICMB843 Estresado (3 réplicas).

FUENTE: EL AUTOR, 2021

**Tabla 7** - Mapeo y Ensamblaje de transcritos

Muestra	Lecturas crudas <sup>1</sup>	Lecturas limpias <sup>2</sup>	Mapeado <sup>3</sup>	Conteo de Lecturas Mapeadas asignadas <sup>4</sup>
ICMB863C1	18594038	18231103	14333939(78.62%)	5092594(21.4%)
ICMB863C2	20338348	19892384	15363561(77.23%)	4345211(18%)
ICMB863C3	29140416	28664162	21421521(74.22%)	8019088(25.6%)
<b>ICMB863C</b>	<b>22690934</b>	<b>22262550</b>	<b>17039674(76.68%)</b>	<b>5818964(21.67%)</b>
ICMB863E1	31315258	30747767	22109729(71.91%)	4802334(20.6%)
ICMB863E2	20499542	20126304	14193331(70.52%)	7836030(24.1%)
ICMB863E3	18684304	18258200	13364531(73.2%)	4435584(21.84%)
<b>ICMB863E</b>	<b>23499701</b>	<b>23044090</b>	<b>16555864(72%)</b>	<b>5691316(22.35%)</b>
ICMB843C1	22376874	21807411	16718535(76.66%)	6006279(25.7%)
ICMB843C2	22140452	21654523	16446374(75.95%)	5965605(24.1%)
ICMB843C3	22680506	22168239	17190017(77.54%)	5924834 (23.9%)
<b>ICMB843C</b>	<b>22399277</b>	<b>21876724</b>	<b>16784975(77%)</b>	<b>5965572(24.57%)</b>
ICMB843E1	14867570	14536585	11356011(78.12%)	3004666(14.6%)
ICMB843E2	22764616	22261198	17249123(77.49%)	4342146(10.3%)
ICMB843E3	25133462	24752802	20881585(84.36%)	4593698(14.7%)
<b>ICMB843E</b>	<b>20921883</b>	<b>20516862</b>	<b>16495573(80%)</b>	<b>3980170(13.2%)</b>
<b>TOTAL</b>	<b>268535386</b>	<b>263100678</b>		

<sup>1</sup>FastQC, <sup>2</sup>Fastp, <sup>3</sup>HISAT2, <sup>4</sup>FeatureCounts

FUENTE: EL AUTOR, 2021

Se confirmó la presencia de FT PgbZIP putativos identificados en el Transcriptoma analizado, seleccionándose 9 (Tabla 8) de estos por haber presentado mayor diferencial de expresión génica (DEG) entre las 52 probables PgbZIP.

**Tabla 8 - FT PgbZIP putativos seleccionados por su mayor DEG**

PgbZIP	Gen	Descripción	E63	E43	C63	C43
Pg4	Pgl_GLEAN_10027837	bZIP9	2997	5300	4163	6214
Pg1	Pgl_GLEAN_10027141	bZIP25	669	1634	608	1613
Pg21	Pgl_GLEAN_10022793	bZIP60	575	1115	483	595
Pg52	Pgl_GLEAN_10021506	ABI 5-2X1	500	79	358	219
Pg37	Pgl_GLEAN_10033760	bZIP39	487	570	333	580
Pg27	Pgl_GLEAN_10017352	HY5	370	679	408	343
Pg47	Pgl_GLEAN_10002524	Desconocida	335	215	232	79
Pg5	Pgl_GLEAN_10038005	bZIP61	146	48	326	226
Pg46	Pgl_GLEAN_10026094	RF2a	70	36	84	73

<sup>1</sup>Expresión en ICMB863 bajo condición estresada, <sup>2</sup>Expresión en ICMB843 bajo condición estresada, <sup>3</sup>Expresión en ICMB863 bajo condición normal, <sup>4</sup>Expresión en ICMB843 bajo condición normal.

FUENTE: EL AUTOR, 2021

Las comparaciones de niveles de expresión o *fold changes* ( $\text{Log}_2$ ) de probables FT bZIP seleccionados pueden verse en la Tabla 9.

Los cambios de niveles de expresión  $\text{Log}_2$  (*Fold change*) muestran que en ICMB863 4 PgbZIP fueron regulados negativamente (*downregulation*) y 5 positivamente (*upregulation*); mientras que en ICMB843, 6 fueron regulados negativamente y 3 positivamente bajo condiciones de estrés por sequía. El análisis transcriptómico realizado por DUDHATE *et al*, (2018) destacaba mayor cantidad de genes regulados negativamente en ICMB843 bajo condiciones de sequía en comparación con ICMB863.

**Tabla 9** - Valores de FC en el nivel de expresión de Pgbzip putativas seleccionadas

PgbZIP	C63	E63	E63/C63	LogFC	P-value	P-adj
Pg4	4163	2997	0.719874579	-0.47418252	0.218709344	0.627308929
Pg1	608	669	1.101364725	0.139292307	0.781392798	0.947285664
Pg21	483	575	1.189393141	0.250225661	0.60623343	0.885686569
Pg27	408	370	0.90779043	-0.13956882	0.770784207	0.942666236
Pg52	358	500	1.397456049	0.482802909	0.531375794	0.853089672
Pg37	333	487	1.465057385	0.550957175	0.159580299	0.550636426
Pg5	326	146	0.448577041	-1.15657231	0.010038211	0.12242559
Pg47	232	335	1.447590749	0.533653793	0.049853253	0.318470983
Pg46	84	70	0.827344916	-0.27343919	0.404627493	0.784872261
PgbZIP	C43	E43	E43/C43	LogFC	P-value	P-adj
Pg4	6214	5300	0.85286301	-0.22961407	0.117318573	0.254932183
Pg1	1634	1613	0.987148103	-0.01866154	0.888117254	0.942630491
Pg21	595	1115	1.872911542	0.905282762	3.12E <sup>-06</sup>	2.75E <sup>-05</sup>
Pg27	343	679	1.981097485	0.986299874	2.85E <sup>-05</sup>	0.00020567
Pg52	219	79	0.360000618	-1.47392871	2.49E <sup>-07</sup>	2.60E <sup>-06</sup>
Pg37	580	557	0.960480627	-0.05817158	0.761651321	0.868983062
Pg5	226	48	0.211604169	-2.24056004	1.22E <sup>-08</sup>	1.56E <sup>-07</sup>
Pg47	79	215	2.721015348	1.444145094	2.69E <sup>-07</sup>	2.80E <sup>-06</sup>
Pg46	73	36	0.489716255	-1.02998201	0.025061909	0.076141676

FUENTE: EL AUTOR, 2021

En la Tabla 10 se puede ver la identidad presentada de estos 9 FT bZIP seleccionados de *P. glaucum* con bZIP de *modelos vegetales*; donde Pg5 (bZIP61) y Pg27 (HY5) mostraron una identidad de 58 y 66% respectivamente con *A. thaliana*; mientras que las 9 PgbZIP seleccionadas en millete mostraron una identidad del 59 a 92% con bZIP de *O. sativa*, lo cual concuerda en vista que ambos cultivos son monocotiledóneas.

**Tabla 10 - Identidad de *P.glaucum* con modelos vegetales**

<i>P. glaucum</i>	Descripción	<i>A. thaliana</i>	%Id	E-value	Acceso
Pg1			35	9.88e <sup>-38</sup>	
Pg4			33	2.70e <sup>-38</sup>	
Pg5	bZIP50	ATbZIP61	58	3e <sup>-13</sup>	NC_003070.9
Pg21			38	1e <sup>-71</sup>	
Pg27	HY5	ATHY5	66	2e <sup>-18</sup>	NC_003076.8
Pg37			40	3e <sup>-44</sup>	
Pg46			39	9e <sup>-79</sup>	
Pg47			70	3e <sup>-31</sup>	
Pg52	ABI-X1	ATAREB1	59	1e <sup>-05</sup>	NM_001333260.1
<i>P. glaucum</i>	Descripción	<i>O. sativa</i>	%Id	E-value	Acceso
Pg1	RISBZ2	OsRISBZ-2	60	2e <sup>-155</sup>	XM_015776133.2
Pg4	RISBZ4	OsRISBZ-4	59	3e-111	XM_015767766.2
Pg5	bZIP19	Os bZIP19	92	8e-73	XM_015767304.2
Pg21	bZIP60	Os bZIP60	78	0.0	XM_015792260.2
Pg27	HY5	OsHY5	66	5e-46	XM_015762010.2
Pg37	bZIP39	Os bZIP39	73	0.0	XM_015784599.2
Pg46	RF2a	OsRF2a	75	4e-58	XM_015757206.2
Pg47	bZIP-23	OJ1004_A11.20	73	2e-63	XM_015770367.2
Pg52	ABI-X1	OsABI	67	3e-65	XM_015787607.2

FUENTE: EL AUTOR, 2021

Se establecieron las funciones de las bZIP ortólogas (Tabla 11) encontradas y fue rastreada la región promotora (Tabla12) de las bZIP de *Oryza* y *Arabidopsis* que mostraron identidad con millete.

**Tabla 11** - Funciones de bZIP ortólogas identificadas en *Arabidopsis* y *Oryza*

<b>Gen</b>	<b>Función</b>
<b>ATbZIP61</b>	Interacción con ATbZIP34, la cual actúa en la funcionalidad del polen. GIBALOVÁ A <i>et al</i> , 2017
<b>ATHY5</b>	Respuesta transcripcional de las plantas a los cambios en las condiciones de luz locales. BURKO Y <i>et al</i> , 2020
<b>ATAREB1</b>	Respuesta al estrés por sequía y adaptación fisiológica. PAES DE MELO B <i>et al</i> , 2020
<b>OsZIP19</b>	Inducido por luz, expresando en brotes y raíces. NIJHAWAN A <i>et al</i> , 2008.
<b>OsZIP23</b>	Respuesta al estrés hídrico. YANG S <i>et al</i> , 2019.
<b>OsZIP39</b>	Estrés al RE (Respuesta a la Proteína desplegada-UPR). POÓR P <i>et al</i> , 2019.
<b>OsRISBZ4</b>	Almacenamiento en semillas. ONODERA Y <i>et al</i> , 2001.
<b>OsRISBZ2</b>	Almacenamiento en semillas. ONODERA Y <i>et al</i> , 2001.
<b>OsZIP60</b>	Estrés al RE (Respuesta a la Proteína desplegada-UPR). POÓR P <i>et al</i> , 2019.
<b>OsHY5</b>	Promueve varios genes involucrados en la biosíntesis de clorofila y la biogénesis del cloroplasto. Pigmentación de semillas durante la maduración. KIM D <i>et al</i> , 2018.
<b>OsRF2a</b>	Tolerancia a enfermedad vírica. Z.G.E <i>et al</i> , 2014.
<b>OsABI</b>	Inducida en alta salinidad, y regulada negativamente por sequía y frío (4 °C) en plántulas. Z.G.E <i>et al</i> , 2014.

FUENTE: EL AUTOR, 2021

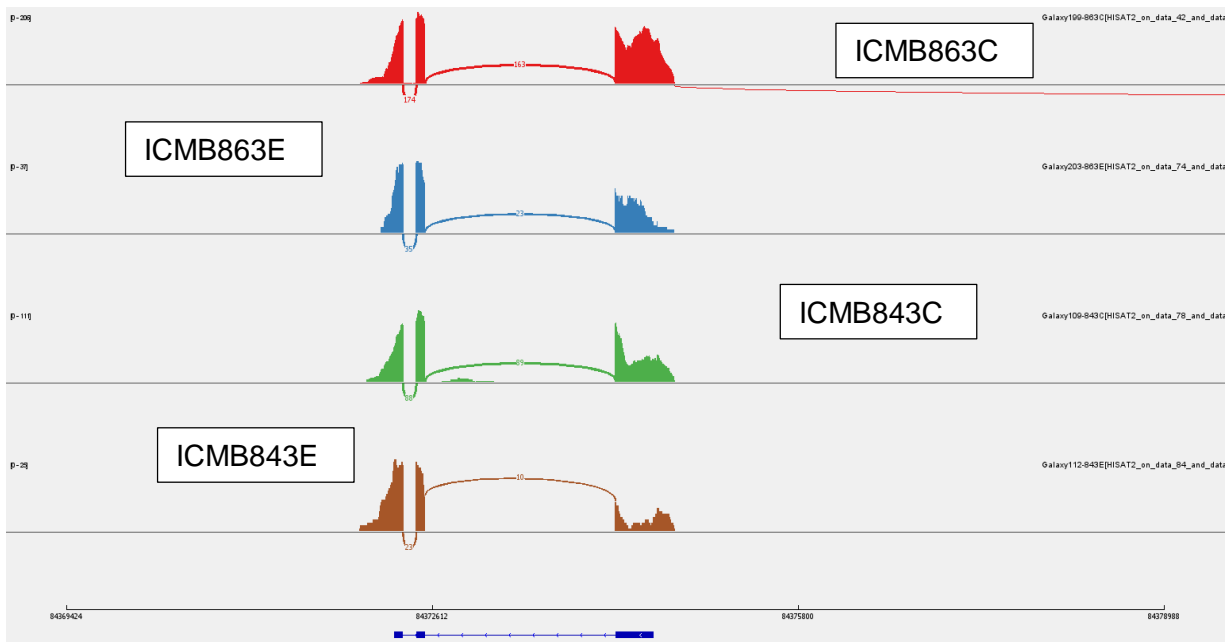
**Tabla 12** - Elementos reguladores involucrados en el estrés hídrico en PgbZIP seleccionados y ortólogos en *Oryza* y *Arabidopsis*

<i>P.glaucum</i>	ABREs	DRE	MYB	MBS	MYC	<i>A.thaliana</i>	ABREs	DRE	MYB	MBS	MYC	<i>O.sativa</i>	ABREs	DRE	MYB	MBS	MYC
<b>Pg1</b>	8	2	2	0	5							<b>OsRISBZ2</b>	5	1	3	0	1
<b>Pg4</b>	1	2	5	1	4							<b>OsRISBZ4</b>	4	2	3	0	7
<b>Pg5</b>	1	1	5	0	4	<b>ATbZIP61</b>	1	0	1	1	13	<b>OsZIP19</b>	2	1	3	1	8
<b>Pg21</b>	5	2	9	1	6							<b>OsZIP60</b>	1	2	4	2	4
<b>Pg27</b>	2	0	1	1	4	<b>ATHY5</b>	1	0	3	0	4	<b>OsHY5</b>	2	0	3	0	2
<b>Pg37</b>	6	0	4	0	3							<b>OsZIP39</b>	2	0	3	0	1
<b>Pg46</b>	2	1	4	0	3							<b>OsRF2a</b>	3	0	3	1	5
<b>Pg47</b>	6	1	2	0	1							<b>OsZIP23</b>	5	2	7	1	2
<b>Pg52</b>	0	1	4	1	1	<b>ATAREB1 (ABF2)</b>	4	0	7	0	6	<b>OsABI</b>	0	0	9	0	1

FUENTE: EL AUTOR, 2021

El análisis de la estructura de las PgbZIP sugestivas evidenció isoformas en una de ellas. En la Figura 12 la Pg52 diferencialmente expresa, con un valor de Fold Change  $\log_2 < -1$  para la línea ICMB843 visualiza través de IGV Sashimi Plot (KATZ *et al*, 2015) una única isoforma.

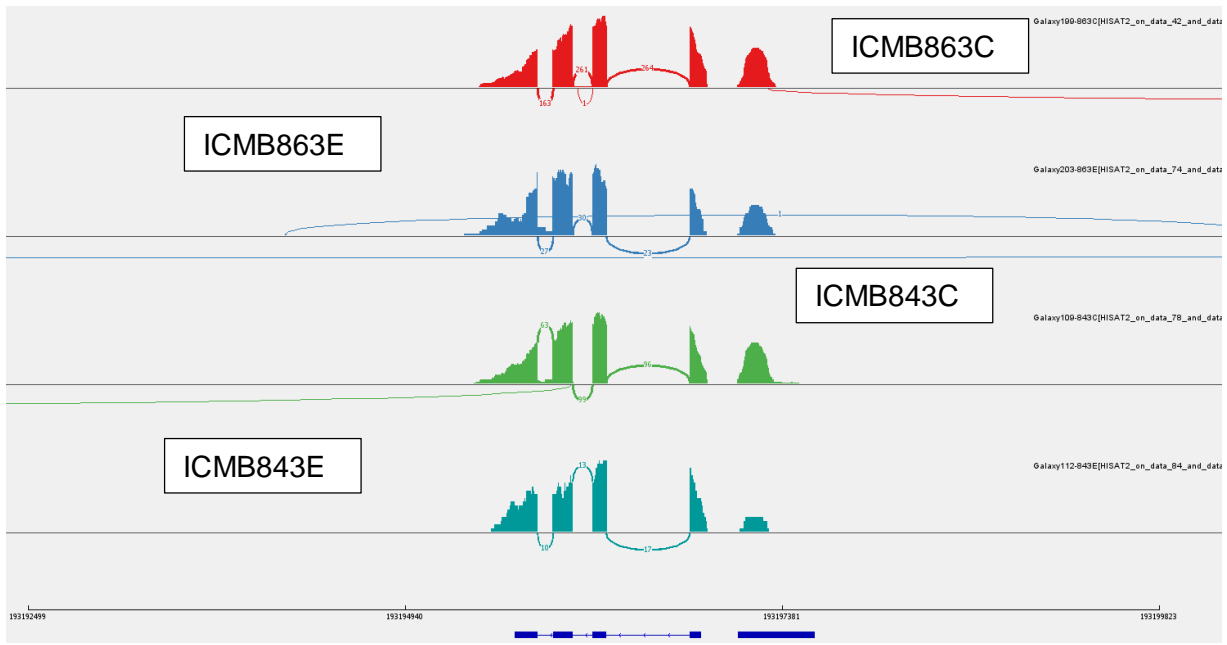
**Figura 12** - Visualización de Pg52



C (Control), E (Estrés hídrico). Las alineaciones en exones se representan como densidades de lectura y las uniones de empalme se dibujan como arcos que conectan un par de exones. Las isoformas de ARNm se muestran en la parte inferior (exones en azul, e intrones como líneas con puntas de flecha).

FUENTE: EL AUTOR, 2021

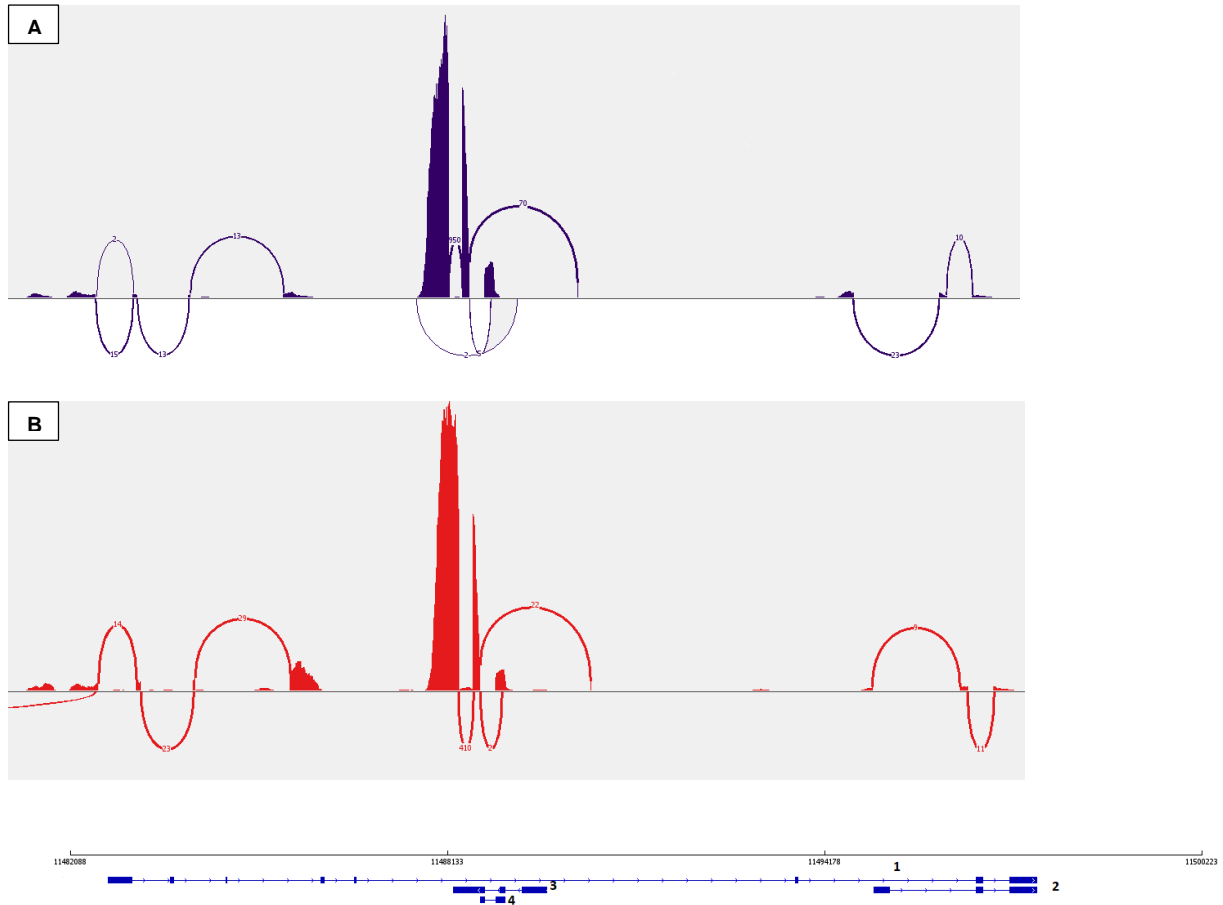
**Figura 13** - Visualización de Isoformas (Sashimi plots) de Pg5



C (Control), E (Estrés hídrico).  
FUENTE: EL AUTOR, 2021

La Pg5 diferencialmente expresa y regulada negativamente ( $\log_2 -2,24$ ) en ICMB843 también indicó una única isoforma según lo establecido en la Figura 13; mientras que la Pg47 diferencialmente expresa y regulada positivamente ( $\log_2 1.44$ ) en ICMB843, visualizó cuatro isoformas en todas las condiciones, según lo indicado en la Figura 14. Las isoformas encontradas fueron confirmadas en presencia y expresión en el transcriptoma analizado.

**Figura 14** - Visualización de Isoformas (Sashimi plots) de Pg47 (ICMB843)



A- Control B- Estrés hídrico. Las isoformas de ARNm se muestran en la parte inferior (exones en azul, intrones como líneas con puntas de flecha).

FUENTE: EL AUTOR, 2021

Al analizar los valores de expresión diferencial en ICMB843 (Tabla 13) la isoforma 3 fue la más expresada inclusive superior a la isoforma 1; sin embargo su Fold change fue -0.99, regulado negativamente y con un valor por encima de -1 y por debajo de 1. Comparativamente las líneas endogámicas tienen una diferencia de expresión en la isoforma 4.

**Tabla 13 - Comparativos de expresión de isoformas**

Id	Pg47	C43	E43	E43/C43	Log <sub>2</sub>
Pgl_GLEAN_10002524	Isoforma 1	79	215	2.72	1.44
MSTRG.11370.2	Isoforma 2	13.04	6.57	0.5	-0.99
MSTRG.11371.1	Isoforma 3	947.56	1244	1.31	0.39
Pgl_GLEAN_10002525	Isoforma 4	0.93	5.42	5.84	2.54
Id	Pg47	C63	E63	E63/C63	Log <sub>2</sub>
Pgl_GLEAN_10002524	Isoforma 1	232	335	1.45	0.53
MSTRG.11370.2	Isoforma 2	8.54	17.62	2.06	1.04
MSTRG.11371.1	Isoforma 3	1207	1030	0.85	-0.23
Pgl_GLEAN_10002525	Isoforma 4	0	0	0	0

FUENTE: EL AUTOR, 2021

En la Tabla 14 se puede ver que la isoforma 3 fue la más expresa en la línea endogámica ICMB843 estresada e ICMB863 control.

**Tabla 14 – Isoformas diferencialmente expresadas en líneas endogámicas de *P. glaucum***

	Pgl_GLEAN_10002524	MSTRG.11370.2	MSTRG.11371.1	Pgl_GLEAN_1000252
<b>ICMB843C</b>	79	13	948	1
<b>ICMB843E</b>	215	7	1244	5
<b>ICMB863C</b>	232	9	1207	0
<b>ICMB863E</b>	335	18	1030	0

FUENTE: EL AUTOR, 2021

## 5 DISCUSIÓN

La revisión bibliográfica realizada perfila a *P. glaucum* como un cereal domesticado nutritivo y tolerante al estrés abiótico.

A pesar de que *P. glaucum* sea un cultivo interesante por su rusticidad y con su genoma disponible pocos son los estudios de genes reguladores registrados en este cultivo. Con el objetivo de revelar la regulación molecular dinámica temprana de *P. glaucum* sometido a estrés por sequía y dilucidar los genes claves que son responsables de la tolerancia a este tipo de estrés, Zhang *et al* (2021) analizaron datos de RNA-Seq a fin de detectar el patrón de expresión de genes en las raíces en etapa de plántula en fases tempranas durante tres puntos de tiempo (1 h, 3 h y 7 h) de estrés por sequía. El número de DEG disminuyó de 1 h a 7 h, sugiriendo que tras la exposición al estrés por sequía *P. glaucum* mostró inmediatamente respuestas debido a su alta capacidad de tolerancia a la sequía, y ya luego las células regresaron a un nivel relativamente equilibrado. Esto podría ser una señal para las plantas de que ya estaban en condiciones normales, por lo que en poco tiempo no era esencial expresar muchos genes, sino que simplemente crecían normalmente. En la investigación realizada, 12 genes exhibieron una regulación positiva en los tres puntos temporales, lo que podría ilustrar que juegan algunos roles importantes en respuesta al estrés por sequía. Para estos genes regulados positivamente, algunos se asociaron a proteínas dedos de zinc tipo CCCH, proteínas FtsH de metaloproteasa de zinc y alcohol deshidrogenasa (gen ADH1) y para cuatro de los genes sus funciones no estaban claras, pero su expresión bajo tratamiento de sequía fue 20 veces más alta que aquellos bajo condiciones normales.

A pesar de las pocas publicaciones dedicadas a VDAC, un análisis *Southern blot* había sugerido que PgVDAC, como gen estructural codificante para una porina o Proteínas de canal de aniones dependiente de voltaje podría existir como un gen de copia única en *P. glaucum*; regulado de forma diferencial en respuesta a diversas tensiones abióticas (DESAI M, 2006). En el presente trabajo la presencia *in silico* de múltiples elementos ABREs, MYC, MYB-MBS y DRE en la región promotora de PgVDAC, podrían sugerir su doble participación en el estrés hídrico, tanto en la vía dependiente (respuesta lenta y adaptativa) como la vía independiente de ABA (respuesta rápida).

Las regiones promotoras pueden contener varios elementos reguladores en *cis* que sirven como sitios de unión para las proteínas reguladoras de genes; desempeñando un papel importante en la especificidad cualitativa de la expresión génica (DUTT 2014). A nivel molecular, se tiene documentado que el estrés por deshidratación dispara la biosíntesis de ABA, el cual es una respuesta imprescindible de tolerancia a este estrés. Los motivos ABREs son importantes para la actividad del promotor en respuesta al estrés hídrico. Múltiples copias de ABRE (ABREs) confieren inducibilidad de ABA a un promotor heterólogo, mientras que una sola copia puede no ser suficiente para su expresión (ROYCHOUDHURY *et al*, 2013). Los elementos MYC y MYB (MBS) también están implicados en la respuesta a la deshidratación. El sistema MYB puede funcionar en un proceso de respuesta al estrés lento y adaptativo (PANDEY S, 2015).

Otro elemento *cis* que ha sido reportado para la regulación del promotor en un estrés hídrico es el motivo DRE (GCCGAC), el cual se sabe que participa en la regulación ABA independiente, mediando una respuesta a la sequía rápida (YOSHIDA *et al*, 2015). Los elementos ABREs y DRE involucrados en el estrés hídrico estuvieron presentes en la región promotora de PgVDAC y su ortólogo ATV DAC2 de *A. thaliana*. CHIKA TATEDA, 2011 describe una posible participación de ATV DAC2 en el desarrollo temprano de plántulas mediado por ABA y en la resistencia a enfermedades.

Aquí en este trabajo, considerando a los factores de transcripción vitales en la regulación de genes en condiciones de estrés identificamos probables FT bZIP a través del análisis *in silico* de un transcriptoma publicado en millete, el cual aportó probables datos interesantes que pudieren ayudar a entender el mecanismo de tolerancia que presenta *P. glaucum* al estrés hídrico.

Los elementos reguladores en *cis* involucrados en vías de respuesta al estrés hídrico de los PgbZIP seleccionados por su mayor DEG evidencian *in silico* la presencia de múltiples elementos ABREs, MYC, MYB-MBS, y DRE, concordando con ZHANG *et al*, (2013) quienes destacan que la función de un promotor normalmente es determinada por la acción combinatoria de múltiples elementos reguladores, que modulan el patrón de expresión espacio-temporal de los genes.

De los 9 FT bZIP diferencialmente expresos; Pg1, Pg4, Pg5, Pg21, Pg46, Pg47 y Pg52 presentan en sus regiones promotoras elementos o motivos *cis* ABREs y DRE;

sugiriéndose así doble su participación modulada en la respuesta lenta o adaptativa y en la respuesta rápida de milleteo al estrés hídrico. Los resultados comparativos *in silico* encontrados en el presente trabajo entre los promotores de los ortólogos exhiben regiones muy conservadas, ya que se tratan de secuencias que compartirían una red de regulación similar, en relación a ABREs, MYB (MBS) y MYC, coordinadas en las vías de respuesta a ABA, una respuesta lenta y adaptativa de las plantas a la deshidratación.

Se sabe que una gran parte de los genes que codifican los Factores de Transcripción presentan isoformas (SEVERING *et al*, 2012); sin embargo, dado que las isoformas de un gen comparten muchos de sus exones, su cuantificación es uno de los procesos más complejos (GARBER *et al*, 2011). Cuando las proporciones de las isoformas cambian en una condición con respecto de otra, como consecuencia de una alteración en el mecanismo empalme alternativo, podría ser relevante en términos biológicos (SONESON *et al*, 2016). SHANKAR *et al*, (2016) indican que el empalme alternativo (EA) juega un papel importante en las respuestas al estrés; señalando que hay estudios que demuestran que el control pos-transcripcional del metabolismo del ARN juega un rol importante.

En nuestro análisis *in silico* las PgbZIP seleccionadas Pg52 y Pg5 diferencialmente expresas y reguladas negativamente presentaron una única isoforma, mientras que la Pg47 diferencialmente expresa y regulada positivamente visualizó cuatro isoformas bajo condición de estrés hídrico y control sin estrés. Más allá de lo poco que se sabe sobre la participación de los factores de transcripción bZIP en respuesta al estrés hídrico, existen algunos reportes que demuestran que diferentes estreses abióticos tal como salinidad, sequía y ABA exógeno causaron regulación diferencial de formas de empalme alternativo en arroz (GUPTA *et al*, 2017). Así mismo, en un análisis transcriptómico de maíz realizado por SONG K *et al*, (2017) encontraron genes diferencialmente expresos en respuesta al estrés hídrico con valores de Fold change  $\log_2 > 1$ , de los cuales aproximadamente el 60% estaban regulados positivamente y 40 % negativamente, en el cual la mayoría de estos genes tenían una única isoforma expresada diferencialmente, pero seis genes regulados positivamente y un gen regulado negativamente tenían dos isoformas expresadas diferencialmente.

La isoforma 2 de Pg47 mostró identidad con la proteína DLC cloroplástica de *Oryza sativa*. En un intento de establecer la función de DCL, se constató que la desregulación de la expresión del gen DCL en *Arabidopsis* provocó múltiples fenotipos como clorosis, flores estériles y desarrollo anormal de cotiledones, lo que sugiere que este gen es necesario en diferentes órganos; además de estar involucrado en la maduración del ARNr de los plástidos (BELLAOUI M *et al*, 2004). En tanto que tanto la isoforma 3 como la isoforma 4 estaban presentes en cultivos como *Zea mayz*, *Oryza sativa* y *Setaria itálica* pero con función desconocida.

Por los resultados aquí alcanzados la PgbZIP47 se torna interesante; no mostró identidad en *Arabidopsis* pero sí 73% con bZIP23 de *Oryza*. OsbZIP23 desempeña una actividad funcional en respuesta al estrés salino e hídrico según lo indicado por ZHU M. *et al*, (2018). Al analizar la región promotora de Pg47 muestra múltiples elementos ABREs y DRE lo que indica su doble participación en vías de respuesta al estrés hídrico.

## 6 CONCLUSIONES

- Para el objetivo “*Explorar **in silico** la regulación de expresión del gen estructural VDAC, identificando los elementos cis afines presentes en la región promotora involucrados en vías de respuesta al estrés hídrico*”; los resultados indicaron la presencia de múltiples elementos *cis* que modulan la expresión de PgVDAC como gen estructural en *P. glaucum* indicando su probable participación en vías de repuesta al estrés hídrico tanto dependiente como independiente del ABA.
- Para el objetivo “*Identificar e indagar en la región promotora de VDAC ortólogos en plantas modelos razonando su participación en el estrés hídrico*”, se encontró mayor similitud con su ortólogo ATVDAC2 en *A. thaliana*, aunque este no es correlacionado con función en el estrés hídrico.
- Para el objetivo “*Cribar en todo el genoma de P. glaucum probables genes reguladores de la familia bZIP utilizando secuencias de referencias de plantas modelos, y rastrear en la región promotora de las bZIP diferencialmente expresas filtradas por análisis **in silico** de un Transcriptoma publicado de estrés hídrico en P. glaucum*”, se pudieron identificar 9 FT PgbZIP diferencialmente expresos bajo condiciones de estrés hídrico, lo que sugiere que esta familia pueda ser uno de los factores claves para la tolerancia al estrés hídrico en esta especie vegetal; y la presencia de múltiples elementos *cis* que modulan la expresión de los FT bZIP como genes reguladores en *P. glaucum* indicaron su probable participación en vías de repuesta al estrés hídrico tanto dependiente como independiente del ABA.
- Para el objetivo “*Examinar en la región promotora de bZIP ortólogos identificados en plantas modelos razonando su participación en vías de respuesta al estrés hídrico*”, exhibieron regiones muy conservadas con mayor similitud con arroz, coordinadas en las vías de repuesta al estrés hídrico ABA dependiente e independiente. La participación de los

Factores de Transcripción bZIP en respuesta al estrés hídrico puede ser considerado como elemento clave para la comprensión de las bases moleculares que modulan las respuestas bioquímicas y fisiológicas de las plantas.

La acumulación de datos aumenta gracias al desarrollo de las herramientas bioinformáticas, permitiendo dar una idea de la composición y de la funcionalidad de los elementos regulatorios de los promotores, por lo cual se propone continuar con la caracterización de los motivos, dándole continuidad al procesamiento y análisis *in silico* de repositorio de RNA-seq de experimentos de estrés depositados en el Archivo de lecturas de secuencias (SRA).

Nuestros hallazgos proporcionan una base para una mayor caracterización funcional e identificación del mecanismo regulador en *P. glaucum* en las respuestas al estrés hídrico. Las técnicas de mejoramiento molecular y herramientas de edición del genoma aliadas a la bioinformática podrían tornar a esta especie un componente potencial a emplear en mejoras de producción agrícola y garantizar la seguridad alimentaria en el futuro.

## 7 REFERENCIAS BIBLIOGRÁFICAS

AFGAN, E.; BAKER, D.; BATUT, B.; VAN DEN BEEK, M.; BOUVIER, D.; ČECH, M.; *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses. **Nucleic Acids Research**. v 46, p537-544, 2018.

AL BITAR, F.; ROOSENS, N.; SMEYERS, M.; VAUTERIN, M.; VAN BOXTEL, J.; JACOBS, M.; HOMBLÉ, F. Sequence analysis, transcriptional and posttranscriptional regulation of the rice vdac family. **Biochim Biophys Acta**. 1625 (1): 43-5, 2003.

ALLISON, L. Transcription in eukaryotes. **Fundamental Molecular Biology**. 672, 2011.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W., AND LIPMAN, D. J. Gapped blast and psiblast: a new generation of protein database search programs. **Nucleic Acids Research** 25, 3389:3402, 1997.

ANDREWS, S. FastQC A Quality Control tool for High Throughput Sequence Data. s.f.

ASHBURNER M, BALL CA, BLAKE JA, BOTSTEIN D, BUTLER H, CHERRY JM, DAVIS AP, DOLINSKI K, DWIGHT SS, EPPIG JT, HARRIS MA, HILL DP, ISSEL-TARVER L, KASARSKIS A, LEWIS S, MATESE JC, RICHARDSON JE, RINGWALD M, RUBIN GM, SHERLOCK G. Gene Ontology: tool for the unification of biology. **Nature Genetics**, 25: 25-30 2000.

ATKINSON, N.&URWIN, P. The interaction of plant biotic and abiotic stresses: from genes to the field. **Journal of Experimental Botany**. 63: 3523-3543, 2012.

BALOJI G., PASHAM1 S., MAHANKALI V., GARLADINNE M., ANKANAGARI S. Insights from the molecular docking analysis of phytohormone reveal brassinolide interaction with HSC70 from Pennisetum glaucum. **Bioinformatics** 15 (2): 131-138, 2019.

BANERJEE, A., AND ROYCHOUDHURY, A. WRKY proteins: signaling and regulation of expression during abiotic stress responses. **Sci. World J.** 1–17, 2015.

BELLAOUI M., GRUISSEM W.; BELLAOUI, M.; GRUISSEM, W. Altered expression of the Arabidopsis ortholog of DCL affects normal plant development. **Planta.** 219 (5): 819-26, 2004.

BIŁAS R.; SZAFRAN K.; HNATUSZKO-KONKA K. *Cis* -regulatory elements used to control gene expression in plants. **Plant Cell, Tissue and Organ Culture (PCTOC)**. 2016.

BURKO Y.; SELUZICKI A.; ZANDER M.; PEDMALE UV.; ECKER, J.; CHORY, J. Chimeric Activators and Repressors Define HY5 Activity and Reveal a Light-Regulated Feedback Mechanism. **Plant Cell.** 32 (4): 967-983, 2020.

CHANWALA, J.; SATPATI, S.; DIXITI, A.; PARIDA, A.; KUMAR, M.; DEY, N. Genome-wide identification and expression analysis of WRKY transcription factors in pearl millet (*Pennisetum glaucum*) under dehydration and salinity stress. **BMC Genomics.** 21 (231): 1-16, 2020.

CHEN, L.; WANG, Q; ZHOU, L.; REN, F.; LI, D & LI, X. B. Arabidopsis CBLinteracting protein kinase (CIPK6) is involved in plant response to salt/osmotic stress and ABA. **Molecular biology reports**,.40 (8); 4759-4767, 2013.

CHEN, S.; ZHOU, Y.; CHEN, Y. & GU, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. 2018.

CHIKA, T.; KANAKO, W.; TOMONOBU, K.; YOSHIHIRO, T. Molecular and genetic characterization of the gene family encoding the voltage-dependent anion channel. **Journal of Experimental Botany.** v. 62, 4773–4785, 2011.

CONESA, A.; GOTZ, S.; GARCIA-GOMEZ, J.M.; TEROL, J.; TALON, M.; ROBLES, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**. 21(18): 3674–6, 2005.

DESAI, M.K.; MISHRA, R.N; VERMA, D.; NAIR, S.; SOPORY, S.K.; REDDYM, K. Structural and functional analysis of a salt stress inducible gene encoding voltage dependent anion channel (VDAC) from pearl millet (*Pennisetum glaucum*). **Plant Physiol Biochem**.44 (7-9): 483-93, 2006.

DIAS-MARTINS, A.M.; PESSANHA, K.L.F.; PACHECO, S.; RODRIGUES, JAS.; CARVALHO, CWP. Potential, use.; of pearl millet, (*Pennisetum glaucum* (L.) R. BR.). **Brazil: Food security, processing, health benefits and nutritional products. Food Res Int**. 109:175-186, 2018.

DUDHATE, A.; SHINDE, H.; TSUGAMA, D.; LIU, S.; TAKANO, T. Transcriptomic analysis reveals the differentially expressed genes and pathways involved in drought tolerance in pearl millet [*Pennisetum glaucum* (L.) R. Br]]. **PLoS ONE** 13 (4), 2018.

DUTT, M.; DHEKNEY, S.; SORIANO, L.; KANDEL R.; GROSSE, J. Temporal and spatial control of gene expression. **horticultural crops**. 1: 14047, 2014.

EDDY, S. The HMMER User's Guide (biological sequence analysis using profile hidden Markov models). Howard Hughes Medical Institute and Dept. of Genetics Washington University School of Medicine, 660 South Euclid Avenue, Box 8232 Saint Louis, Missouri 63110, USA, version 2.3.2 edition. [hmmerr.wustl.edu](http://hmmerr.wustl.edu), 2003.

EDGAR RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Res**. 32 (5): 1792–7, 2004.

HOMBLÉ, F.; KRAMMER, E.; PRÉVOST, M. Plant VDAC: Facts and speculations, *Biochimica et Biophysica Acta (BBA)*. **Biomembranes**, v 1818, p. 1486-1501, 2012.

FANG, Z.; MARTIN, J.; WANG, Z. Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. **Cell & bioscience**. 2(1): 26, 2012.

FINN, R.D.; CLEMENTS, J.; EDDY, S.R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Res**. 39, 2011.

FINER, J.; HERNANDEZ-GARCÍA C. Identification and validation of promoters and cisacting regulatory elements. **Plants Science**. 217-218: 109-119, 2014.

FUJITA, Y.; FUJITA, M.; SATOH, R.; MARUYAMA, K.; PARVEZ, M. M.; SEKI, M.; HIRATSU, K.; OHME-TAKAGI, M.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. **Plant Cell**. 17, 2005.

GARBER, M.; GRABHERR, M. G.; GUTTMAN, M.; TRAPNELL, C. Computational methods for transcriptome annotation and quantification using RNA-seq. **Nature methods**. 8 (6): 469–477, 2011.

GATZ, C. From Pioneers to Team Players: TGA Transcription Factors Provide a Molecular Link Between Different Stress Pathways. **Molecular Plant-Microbe Interactions**. 26, 2012.

GERHARDT, G. J.; ECHEVERRIGARAY, S. Rules extraction from neural networks applied to the prediction and recognition of prokaryotic promoters. **Genetics and molecular biology**. 34 (2): 353-360, 2011.

GIBALOVÁ, A.; STEINBACHOVÁ, L.; HAFIDH, S.; BLÁHOVÁ, V.; GADIOU, Z.; MICHAILEDIS, C.; MÜLLER, K.; PLESKOT, R.; DUPLÁKOVÁ, N.; HONYS, D. Characterization of pollen-expressed bZIP protein interactions and the role of ATbZIP18 in the male gametophyte. **Plant Reprod**. 30 (1):1-17, 2017.

GUPTA, P.; NUTAN, K.; SINGLA-PAREEK, S. L. & PAREEK, A. Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice. **Frontiers in plant science**. 8, 1944, 2017.

HA, C. D & WYCKHUYS, K. A. *In silico* identification and characterization of the lectin gene families in cassava (*Manihot esculenta* Crantz). **TAP CHI SINH HOC**, 39(3), 2017.

HASHIMOTO, Y.; ZHANG, C.; KAWAUCHI, J.; IMOTO, I.; ADACHI, M.T.; INAZAWA, J.; AMAGASA, T.; HAI, T.; KITAJIMA, S. An alternatively spliced isoform of transcriptional repressor ATF3 and its induction by stress stimuli. **Nuc Ac Res**. 30: 2398-2406, 2002.

HOMBLÉ, F.; KRAMMER, E.; PRÉVOST, M. Plant VDAC: Facts and speculations, *Biochimica et Biophysica Acta (BBA)*. **Biomembranes**, v 1818, p. 1486-1501, 2012.

IBRAHEEM, O.; BOTHA, C. E. & BRADLEY, G. *In silico* analysis of *cis*-acting regulatory elements in 5' regulatory regions of sucrose transporter gene families in rice (*Oryza sativa* Japonica) and *Arabidopsis thaliana*. **Computational biology and chemistry**. 34(5-6), 268-283, 2010.

IDROVO, ESPÍN; PERAZA-ECHEVERRÍA, S.; FUENTES, G.; SANTAMARÍA, J.M. *In silico* cloning and characterization of the TGA (TGACG MOTIF-BINDING FACTOR) transcription factors subfamily in *Carica papaya*. **Plant Physiology and Biochemistry**. 54: 113-122, 2012.

ISLAM, T.; MANNA, M.; REDDY, M.K. Glutathione Peroxidase of *Pennisetum glaucum* (PgGPx) Is a Functional Cd<sup>2+</sup> Dependent Peroxiredoxin that Enhances Tolerance against salinity and Drought Stress. **PLoS ONE**. 10 (11), 2015.

JAISWAL, S.; ANTALA, T.J.; MANDAVIA, M.K.; MEENU, CH.; RAHUL, S.; RUKAM, S.; JASHMINKUMAR, K.; UB, A.; MA, I.; BA, G.; ANIL, R.; DINESH, K. Transcriptomic

signature of drought response in pearl millet (*Pennisetum glaucum* (L.) and development of web-genomic resources. **Sci Rep.** 8, 3382, 2018.

JAKOBY, M.; WEISSHAAR, B.; DRÖGE-LASER, W.; VICENTE-CARBAJOSA, J.; TIEDEMANN; J.; KROJ, T.; PARCY, F. bZIP transcription factors in Arabidopsis. Trends in **Plant Science.** 7: 106-111, 2002.

KATZ, Y; WANG, E; SILTERRA, J.; SCHWARTZ, S.; WONG, B.; MESIROV, J.; AIROLDI, E; BURGE, C. Sashimi plots: quantitative visualization of RNA sequencing read alignments. **Bioinformatics.** 31(14): 2400-2402, 2015.

KIM, D; PARK, S; LEE, J; HA, S; LEE, J; LIM, S. A Rice B-Box Protein, OsBBX14, Finely Regulates Anthocyanin Biosynthesis in Rice. **Int J Mol Sci.** 19 (8): 2190, 2018.

KIM, D.; LANGMEAD, B. & SALZBERG, S. L. HISAT: a fast spliced aligner with low memory requirements. **Nature Methods.** 12 (4): 357-360, 2015.

KIM, W.C.; KIM, J.Y.; KO, J.H.; KANG, H.; HAN, K.H. Identification of direct targets of transcription factor MYB46 provides insights into the transcriptional regulation of secondary wall biosynthesis. **Plant Mol. Biol.** 85, 589–599, 2014.

KOVAKA, S.; ZIMIN, A.; PERTEA, G.; RAZAGHI, R.; SALZBERG, S. L. & PERTEA, M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. **Genome Biology.** 20 (1), 2019.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Mol Biol Evol.** 33(7):1870–4, 2016.

KUROMORI, T.; MIZOI, J.; UMEZAWA, T.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Drought Stress Signaling Network. **Molecular Biology.** Springer. pp 383-409, 2014.

LATA, C.; YADAV, A.; PRASSAD, M. Role of plant transcription factor in abiotic stress tolerance. **Abiotic Stress Response Plants**. 10, 269–296, 2011.

LESCOT, M.; DÉHAIS, P.; THIJS, G.; MARCHAL, K.; MOREAU, Y.; VAN DE PEER, Y.; ROUZÉ, P.; ROMBAUTS S. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. **Nucleic Acids Res**. 30(1):325–7, 2002.

LI, X.; LI, S.; LANG, Z.; ZHANG, J.; ZHU, L. & HUANG, D. Chloroplast-targeted expression of the codon-optimized truncated cry1Ah gene in transgenic tobacco confers a high level of protection against insects. **Plant Cell Reports**, 32(8), 1299-1308, 2013.

LIAO, Y.; SMYTH, G. K. & SHI, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**. 30(7), 923-930, 2013.

LINDEMOSE, S.; SHEA, C.; JENSEN, M.; SKRIVER, K. Structure, Function and Networks of Transcription Factors Involved in Abiotic Stress Responses. **International Journal of Molecular Sciences**. 14: 5842, 2013.

LINDLÖF, A.; BRÄUTIGAM, M.; CHAWADE, A.; OLSSON, O. & OLSSON, B. *In silico* analysis of promoter regions from cold-induced genes in rice (*Oryza sativa* L.) and *Arabidopsis thaliana* reveals the importance of combinatorial control. **Bioinformatics**. 25 (11), 2009.

LIU W & STEWART C. Plant synthetic promoters and transcription factors. **Biotechnology**. 37: 36–44, 2016.

LOVE, M.; HUBER, W. & ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology** 15 (12), 2014.

MA, Y.; DAI, X.; XU, Y.; LUO, W.; ZHENG, X.; ZENG, D.; PAN, Y.; LIN, X.; LIU, H.; ZHANG, D., *et al.* *COLD1* confers chilling tolerance in rice. **Cell**. 160:1209–21, 2015.

MALIK, S. Pearl millet-nutritional value and medicinal use. *Int.J.Adv.Res. Innov. Ideas Educ.* 1, 414–418, 2015.

MOCKLER, T.; PRIEST, H.; FILICHKIN, S. *Cis*-regulatory elements in plant cell signaling. **Plant Biology**. 12: 643-649, 2009.

NAQVI, R.; MUBEEN, H.; RAZA, S. Role of plant promoters and their *cis* regulatory elements in gene expression regulation. **European Journal of Pharmaceutical and Medicinal Research**. 1: 347-352, 2016.

NIJHAWAN, A.; JAIN, M.; TYAGI, A.K.; KHURANA, .JP. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. **Plant Physiol**. 146(2):333-50, 2008.

ONODERA, Y.; SUZUKI, A.; WU, C.Y.; WASHIDA, H.; TAKAIWA, F. A rice functional transcriptional activator, *RISBZ1*, responsible for endosperm-specific expression of storage protein genes through *GCN4* motif. **Biol Chem**. 276 (17): 14139-52, 2001.

PAES DE MELO, B.P.; LOURENÇO-TESSUTTI, I.T.; PAIXÃO, J.F.R.; NORIEGA, D.D.; SILVA, M.; DE ALMEIDA-ENGLER, J.; FONTES, E.; GROSSI-DE-AS, M.F. Transcriptional modulation of *AREB-1* by CRISPRa improves plant physiological performance under severe water deficit. **Sci Rep**. (1):16231, 2020.

PANDEY, S.; REDDY, C.; YAQOO, U.; KUMAR, Y.; ARORA, S.; KAUL, T. *In silico* analysis of *cis* acting regulatory elements CAREs in *upstream* regions of ascorbate glutathione pathway genes from *Oryza sativa*. **Biochemistry and Physiology**. 4 (2): 1-7, 2015.

PATANUN, O.; LERTPANYA SAMPATHA, M.; SOJIKUL, P.; VIBOONJUN, U. & NARANGAJAVANA, J. Computational identification of microRNAs and their targets in cassava (*Manihot esculenta* Crantz.). **Molecular biotechnology**. 53 (3), 257-269, 2013.

PÉREZ-QUINTERO, Á. & LÓPEZ, C. Identificación de elementos *cis*-regulatorios y predicción bioinformática de factores de transcripción involucrados en la regulación de miARNs en plantas. *Acta Biológica Colombiana*, 18 (1), 2013.

POÓR, P.; CZÉKUS, Z.; TARI, I.; ÖRDÖG, A. The Multifaceted Roles of Plant Hormone Salicylic Acid in Endoplasmic Reticulum Stress and Unfolded Protein Response. **Int J Mol Sci**. 20 (23):5842, 2019.

REDDY, P. S.; MALLIKARJUNA, G.; THIROLOGACHANDAR, V.; VAISHNAVI, C.S.; AAKRATI, A. Molecular characterization and expression of a gene encoding cytosolic Hsp90 from *Pennisetum glaucum* and its role in abiotic stress adaptation. **Gene**. 474 (2011) 29–38, 2010.

ROCK C. Pathways to abscisic acid-regulated gene expression. **New Phytol**. 148: 357-396, 2000.

ROYCHOUDHURY, A.; PAUL, S.; BASU, S. Interferencia entre las vías dependientes del ácido abscísico y las independientes del ácido abscísico durante el estrés abiótico. **Plant Cell**. 32: 985–1006, 2013.

SEHGAL, D.; RAJARAM, V.; ARMSTEAD, I.; VADEZ, V.; YADAV, Y.; HASH, C.; YADAV, R. Integration of gene-based markers in a pearl millet genetic map for identification of candidate genes underlying drought tolerance quantitative trait loci. **BMC Plant Biol**. 12 (1): 9, 2012.

SEVERING, E.; VAN DIJK, A.; MORABITO, G.; BUSSCHER-LANGE, J.; IMMINK, R. G. & VAN HAM, R. C. Predicting the impact of alternative splicing on plant MADS domain protein function. **PLoS One**. 7 (1), 2012.

SHANKAR, R.; BHATTACHARJEE, A. & JAIN, M. Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. **Sci. Rep.** 2016

SHRESTHA, A.; KHAN, A.; DEY, N. *Cis*–*trans* engineering: advances and perspectives on customized transcriptional regulation in plants. **Mol Plant**. 11 (7): 886–98, 2018.

SHIVHARE, R. & LATA, C. Exploración de recursos genéticos y genómicos para la tolerancia al estrés abiótico y biótico en mijo perla. **Frontiers in Plant Science**. 7: 2069, 2017.

SHIVHARE, R.; LAKHWANI, D.; ASIF, M.H.; PUNNET, S. C.; CHARU, L. De novo assembly and comparative transcriptome analysis of contrasting pearl millet (*Pennisetum glaucum* L.) genotypes under terminal drought stress using illumina sequencing. **Nucleus**. 2020.

SINGH, J.; REDDY, P.S.; REDDY, C.S.; REDDY, M.K. Molecular cloning and characterization of salt inducible dehydrin gene from the C4 plant *Pennisetum glaucum*. **Plant Gene**. 4, 55–63, 2015.

SINGH, P.; BOOTE, K.J.; KADIYALA, M.; NEDUMARAN, S.; GUPTA, S.; SRINIVAS, K.; BANTILAN, M. An assessment of yield gains under climate change due to genetic modification of pearl millet. **Sci Total Environ**. 601-602:1226–37, 2017.

SONESON, C.; MATTHES, K. L.; NOWICKA, M.; LAW, C. W.; AND ROBINSON, M. D. Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage. **Genome biology**. 17(1):12, 2016.

SONG, L.; MA, Q.; ZOU, Z.; SUN, K.; YAO, Y.; TAO, J.; KALERI N.; LI, X. Molecular Link between Leaf Coloration and Gene Expression of Flavonoid and Carotenoid Biosynthesis in *Camellia sinensis* Cultivar 'Huangjinya'. **Frontiers in plant science**. 8, 803 2017.

SONNHAMMER, E.; EDDY, S.; BIRNEY, E.; BATEMAN, A.; DURBIN, R. Pfam: multiple sequence alignments and HMM-profiles of protein domains. **Nucleic Acids Res**. 26 (1): 320–2, 1998.

SUZUKI, N.; RIVERO, R.; SHULAEV, V.; BLUMWALD, E.; MITTLER, R. Abiotic and biotic stress activity from guava fruit extracts. **Journal of Food Composition and Analysis**. 19: 669-675, 2014.

TAKO, E.; REED, S.; BUDIMAN, J.; HART, J.; GLAHN, R. Higher iron pearl millet (*Pennisetum glaucum* L.) provides more absorbable iron that is limited by increased polyphenolic content. **Nutr J**. 14 (1):11, 2015.

TIESSEN, A.; GÓMEZ, F.; TREJO, L.; LÓPEZ, A.; PADILLA, D.; VARGAS, E.; PALACIOS, N. Fundamentos y metodologías innovadoras para el mejoramiento genético de maíz. 1ra Edición. Capítulo: 6. Factores de transcripción. Fundación Ciencia Activa. Editor: Axel Tiessen. pp.153-188, 2009.

VADEZ, V.; HASH, T.; BIDINGER, F. & KHOLOVA, J. Phenotyping pearl millet for adaptation todrought. **Front Physiol**. 2012.

VAN VERK, M. C.; HICKMAN, R.; PIETERSE, C. M.; VAN WEES, S. RNA-Seq: revelation of the messengers. **Trends in plant science**. v. 18, n. 4, p. 175-179, 2013.

VARSHNEY, R.K.; SHI, C.; THUDI, M.; MARIAC, C.; WALLACE, J.; Qi, P., *et al*. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. **Nat Biotechnol**. 35(10):969–76, 2017.

VEDEL, V. & SCOTTI, I. Promoting the promoter. **Plant Science**. 180 (2): 182–189, 2011.

YAMAMOTO, Y.; ICHIDA, H.; MATSUI, M.; OBOKATA, J.; SAKURAI, T.; SATOU, M.; SEKI, M.; SHINOZAKI, K.; ABE, T. Identification of plant promoter constituents by analysis of local distribution of short sequences. **BMC Genomics**. 8: 1-23, 2007.

YANG, S.; XU, K.; CHEN, S.; LI, T.; XIA, H.; CHEN, L.; LIU, H.; LUO, L. A stress-responsive bZIP transcription factor OsbZIP62 improves drought and oxidative tolerance in rice. **BMC Plant Biol**. 17;19 (1):260, 2019.

YOSHIDA, T.; FUJITA, Y.; MARUYAMA, K.; MOGAMI, J.; TODAKA, D.; SHINOZAKI, K.; YAMAGUCHI, K. Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. **Plant, Cell and Environment**. 38: 35–49, 2015.

ZGE; ZHANG, Y.P.; ZHOU, J.H.; WANG, L. Mini review roles of the bZIP gene family in rice. **Genet Mol Res**. 16; 13 (2): 3025-36, 2014.

ZHANG, S.; LIAN, Y.; LIU, Y.; WANG, X.; LIU, X.; WANG, G. Characterization of a Maize Wip1 Promoter in Transgenic Plants. **International Journal of Molecular Science**. 14, 2013.

ZHANG, A.; JI, Y.; SUN, M.; LIN, C.; ZHOU, P.; REN, J.; LUO, D.; WANG, X., MA, C.; ZHANG, X.; FENG, G.; NIE, G.; HUANG L. Research on the drought tolerance mechanism of *Pennisetum glaucum* (L.) in the root during the seedling stage. **BMC Genomics**. 23; 22 (1): 568, 2021.

ZHONG, L.; CHEN, D.; MIN, D.; LI, W.; XU, Z.; ZHOU, Y.; LI, L.; CHEN, M.; MA, Y. AtTGA4, a bZIP transcription factor, confers drought resistance by enhancing nitrate transport and assimilation in Arabidopsis thaliana. **Biochemical and Biophysical Research Communications**. 457: 433-439, 2015.

ZHU, J.; Z.; XUEWEN, W.; LINGXIAO, G.; QINGSHAN, X.; SHIQI, Z.; FANGDONG, L.; XIAOMEI, Y.; SHENGRUI, L.; CHAOLING, W. Characterization and Alternative Splicing Profiles of the Lipoxygenase Gene Family in Tea Plant (*Camellia sinensis*), **Plant and Cell Physiology**, v 59, 9: 1765–1781, 2018.

## 8 ANEXOS

### 8.1 Lista de bZIP referenciales utilizados

<i>Oryza Sativa</i>	
1	>sp Q8RZ35 ABI5_ORYSJ bZIP transcription factor ABI5 homolog OS=Oryza sativa subsp. japonica OX=39947 GN=ABI5 PE=1 SV=1
2	>sp Q69TW5 BZP46_ORYSJ bZIP transcription factor 46 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP46 PE=1 SV=1
3	>sp Q6ZLB0 RSBZ1_ORYSJ bZIP transcription factor RISBZ1 OS=Oryza sativa subsp. japonica OX=39947 GN=RISBZ1 PE=1 SV=1
4	>sp Q69XV0 BZP50_ORYSJ bZIP transcription factor 50 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP50 PE=2 SV=1
5	>sp Q0JHF1 BZP12_ORYSJ bZIP transcription factor 12 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP12 PE=2 SV=1
6	>sp Q6Z312 BZP23_ORYSJ bZIP transcription factor 23 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP23 PE=2 SV=1
7	>sp Q6AU90 BZP39_ORYSJ bZIP transcription factor 39 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP39 PE=2 SV=1
8	>sp Q8LIB3 BZP60_ORYSJ bZIP transcription factor 60 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP60 PE=2 SV=1
9	>sp Q654B3 RSBZ5_ORYSJ bZIP transcription factor RISBZ5 OS=Oryza sativa subsp. japonica OX=39947 GN=RISBZ5 PE=2 SV=1
10	>sp Q5QNI5 BZP02_ORYSJ Basic leucine zipper 2 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP02 PE=2 SV=1
11	>sp Q6K3R9 BZP19_ORYSJ Basic leucine zipper 19 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP19 PE=2 SV=1
12	>sp Q5JMK6 BZP06_ORYSJ Basic leucine zipper 6 OS=Oryza sativa subsp. japonica OX=39947 GN=BZIP06 PE=2 SV=1
13	>tr B9FUH4 B9FUH4_ORYSJ BZIP domain-containing protein OS=Oryza sativa subsp. japonica OX=39947 GN=OsJ_25329 PE=4 SV=1
14	>tr B9F3X1 B9F3X1_ORYSJ BZIP domain-containing protein OS=Oryza sativa subsp. japonica OX=39947 GN=OsJ_05807 PE=4 SV=1
15	>tr Q688Q7 Q688Q7_ORYSJ BZIP transcription factor OS=Oryza sativa subsp. japonica OX=39947 GN=Os05g0129300 PE=2 SV=1
16	>tr B9ETU8 B9ETU8_ORYSJ BZIP domain-containing protein OS=Oryza sativa subsp. japonica OX=39947 GN=OsJ_03842 PE=4 SV=1
17	>tr Q69M02 Q69M02_ORYSJ BZIP domain-containing protein OS=Oryza sativa subsp. japonica OX=39947 GN=OsJ_05709 PE=4 SV=1
18	>osbzip62-tr Q0D3H8 Q0D3H8_ORYSJ Os07g0686100 protein OS=Oryza sativa subsp. japonica OX=39947 GN=Os07g0686100 PE=2 SV=1
<i>Arabidopsis thaliana</i>	
1	>sp O22208 BZP17_ARATH bZIP transcription factor 17 OS=Arabidopsis thaliana OX=3702 GN=BZIP17 PE=1 SV=2
2	>sp Q9LZP8 BZP53_ARATH bZIP transcription factor 53 OS=Arabidopsis thaliana OX=3702 GN=BZIP53 PE=1 SV=1
3	>sp O65683 BZP11_ARATH bZIP transcription factor 11 OS=Arabidopsis thaliana OX=3702 GN=BZIP11 PE=1 SV=1
4	>sp Q9M1G6 BZP25_ARATH Basic leucine zipper 25 OS=Arabidopsis thaliana OX=3702 GN=BZIP25 PE=1 SV=1
5	>sp Q9SG86 BZP28_ARATH bZIP transcription factor 28 OS=Arabidopsis thaliana OX=3702 GN=BZIP28 PE=1 SV=1
6	>sp B9DGI8 BZP63_ARATH Basic leucine zipper 63 OS=Arabidopsis thaliana OX=3702 GN=BZIP63 PE=1 SV=1
7	>sp C022L5 BZP44_ARATH bZIP transcription factor 44 OS=Arabidopsis thaliana OX=3702 GN=BZIP44 PE=1 SV=1
8	>sp Q9C7S0 BZP60_ARATH bZIP transcription factor 60 OS=Arabidopsis thaliana OX=3702 GN=BZIP60 PE=1 SV=1
9	>sp O22763 BZP10_ARATH Basic leucine zipper 10 OS=Arabidopsis thaliana OX=3702 GN=BZIP10 PE=1 SV=2
10	>sp Q9FGX2 BZP1_ARATH Basic leucine zipper 1 OS=Arabidopsis thaliana OX=3702 GN=BZIP1 PE=1 SV=1
11	>sp O22873 BZP18_ARATH bZIP transcription factor 18 OS=Arabidopsis thaliana OX=3702 GN=BZIP18 PE=1 SV=1
12	>sp Q7PCC6 FDP_ARATH bZIP transcription factor 27 OS=Arabidopsis thaliana OX=3702 GN=FDP PE=1 SV=1
13	>sp Q501B2 BZP16_ARATH bZIP transcription factor 16 OS=Arabidopsis thaliana OX=3702 GN=BZIP16 PE=1 SV=1
14	>sp F4IN23 BZP34_ARATH Basic leucine zipper 34 OS=Arabidopsis thaliana OX=3702 GN=BZIP34 PE=1 SV=1
15	>sp Q84LG2 BZP68_ARATH bZIP transcription factor 68 OS=Arabidopsis thaliana OX=3702 GN=BZIP68 PE=1 SV=1
16	>sp Q8GTS1 BZP24_ARATH Basic leucine zipper 24 OS=Arabidopsis thaliana OX=3702 GN=BZIP24 PE=1 SV=1
17	>sp Q9SI15 BZP2_ARATH bZIP transcription factor 2 OS=Arabidopsis thaliana OX=3702 GN=BZIP2 PE=1 SV=1
18	>sp Q9M2K4 BZP61_ARATH Basic leucine zipper 61 OS=Arabidopsis thaliana OX=3702 GN=BZIP61 PE=1 SV=1
19	>sp Q8VY76 BZP19_ARATH Basic leucine zipper 19 OS=Arabidopsis thaliana OX=3702 GN=BZIP19 PE=2 SV=1
20	>sp Q8GTS2 BZP23_ARATH Basic leucine zipper 23 OS=Arabidopsis thaliana OX=3702 GN=BZIP23 PE=2 SV=1
21	>sp Q9LXX4 BZP49_ARATH bZIP transcription factor 49 OS=Arabidopsis thaliana OX=3702 GN=BZIP49 PE=1 SV=1
22	>sp Q9FMC2 BZP43_ARATH Basic leucine zipper 43 OS=Arabidopsis thaliana OX=3702 GN=BZIP43 PE=1 SV=1
23	>sp Q9FMM7 AI5L8_ARATH ABSCISIC ACID-INSENSITIVE 5-like protein 8 OS=Arabidopsis thaliana OX=3702 GN=BZIP15 PE=1 SV=1
24	>sp Q9CA46 BZP8_ARATH Basic leucine zipper 8 OS=Arabidopsis thaliana OX=3702 GN=BZIP8 PE=1 SV=1
25	>sp Q9LQ65 BZP4_ARATH Basic leucine zipper 4 OS=Arabidopsis thaliana OX=3702 GN=BZIP4 PE=1 SV=1

<i>Setaria Italica</i>	
1	>tr K3XIK9 K3XIK9_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101772577 PE=2 SV=1
2	>tr K3XY81 K3XY81_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101776070 PE=4 SV=1
3	>tr K3Z4H4 K3Z4H4_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101778763 PE=4 SV=1
4	>tr K3ZAH0 K3ZAH0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101769858 PE=4 SV=1
5	>tr K3ZRX1 K3ZRX1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_2G411200v2 PE=4 SV=1
6	>tr K3XM66 K3XM66_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101778144 PE=4 SV=1
7	>tr K3YW78 K3YW78_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101760850 PE=4 SV=1
8	>tr K4AKY0 K4AKY0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 PE=4 SV=1
9	>tr K3ZAJ1 K3ZAJ1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101754036 PE=4 SV=1
10	>tr K3XYH0 K3XYH0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101763662 PE=4 SV=1
11	>tr K3XN16 K3XN16_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101759214 PE=4 SV=1
12	>tr K3YJV4 K3YJV4_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101766549 PE=4 SV=1
13	>tr K3ZIP0 K3ZIP0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101771944 PE=4 SV=1
14	>tr K3ZXM1 K3ZXM1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101761819 PE=4 SV=1
15	>tr K3ZIQ0 K3ZIQ0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_8G036300v2 PE=4 SV=1
16	>tr K3XJ56 K3XJ56_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101779496 PE=4 SV=1
17	>tr K4AMZ0 K4AMZ0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101784853 PE=4 SV=1
18	>tr K3ZU42 K3ZU42_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101761017 PE=4 SV=1
19	>tr K3Z705 K3Z705_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101753891 PE=4 SV=1
20	>tr K3YTP1 K3YTP1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101756519 PE=4 SV=1
21	>tr K3XK88 K3XK88_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101770717 PE=4 SV=1
22	>tr K3Z768 K3Z768_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101755376 PE=4 SV=1
23	>tr K3YE22 K3YE22_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101757132 PE=4 SV=1
24	>tr K4A0N9 K4A0N9_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 PE=4 SV=1
25	>tr K3Y8K7 K3Y8K7_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101780472 PE=4 SV=1
26	>tr K3ZJ52 K3ZJ52_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101777244 PE=4 SV=1
27	>tr K3YUA1 K3YUA1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101753963 PE=4 SV=1
28	>tr K3ZWU5 K3ZWU5_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101786916 PE=4 SV=1
29	>tr K3ZUE3 K3ZUE3_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101783447 PE=4 SV=1
30	>tr K3ZUA1 K3ZUA1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101761959 PE=4 SV=1
31	>tr K4ABX3 K4ABX3_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101756896 PE=4 SV=1
32	>tr K3ZA77 K3ZA77_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101766883 PE=4 SV=1
33	>tr K3XGD5 K3XGD5_SETIT Uncharacterized protein OS=Setaria italica OX=4555 GN=101771249 PE=4 SV=1
34	>tr A0A368SF74 A0A368SF74_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_9G055600v2 PE=4 SV=1
35	>tr K3YV65 K3YV65_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101782719 PE=4 SV=1
36	>tr A0A4U6SQM6 A0A4U6SQM6_SETVI BZIP domain-containing protein OS=Setaria viridis OX=4556 GN=SEVIR_9G055100v2 PE=4 SV=1
37	>tr A0A4U6VB94 A0A4U6VB94_SETVI BZIP domain-containing protein OS=Setaria viridis OX=4556 GN=SEVIR_3G203000v2 PE=4 SV=1
38	>tr K4AKE4 K4AKE4_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 PE=4 SV=1
39	>tr K3XYN5 K3XYN5_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101779340 PE=4 SV=1
40	>tr K3XKX3 K3XKX3_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101755973 PE=4 SV=1
41	>tr K3YXV1 K3YXV1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101775585 PE=4 SV=1
42	>tr K4A856 K4A856_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 PE=4 SV=1
43	>tr K3Z711 K3Z711_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_3G379700v2 PE=4 SV=1
44	>tr K3XZQ1 K3XZQ1_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101767052 PE=4 SV=1
45	>tr K3Z036 K3Z036_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101782330 PE=4 SV=1
46	>tr K3ZUL9 K3ZUL9_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101758999 PE=4 SV=1
47	>tr K3YIB0 K3YIB0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101765591 PE=4 SV=1
48	>tr K3YU95 K3YU95_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101778683 PE=4 SV=1
49	>tr K3YUA0 K3YUA0_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101778683 PE=4 SV=1
50	>tr K3YJX7 K3YJX7_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 PE=4 SV=1
51	>tr K3Z7Q7 K3Z7Q7_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101755376 PE=4 SV=1
52	>tr A0A368QHW4 A0A368QHW4_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_3G228000v2 PE=4 SV=1
53	>tr A0A368RW53 A0A368RW53_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_7G151800v2 PE=4 SV=1
54	>tr K3XYW5 K3XYW5_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101779340 PE=4 SV=1
55	>tr A0A368RN63 A0A368RN63_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_6G190600v2 PE=4 SV=1
56	>tr A0A368SDG2 A0A368SDG2_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_9G055600v2 PE=4 SV=1
57	>tr A0A368Q416 A0A368Q416_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=SETIT_2G278200v2 PE=4 SV=1
58	>tr K3XZ62 K3XZ62_SETIT BZIP domain-containing protein OS=Setaria italica OX=4555 GN=101779340 PE=4 SV=1

## 8.2 Alineamiento ATVDAC2

>Arabidopsis thaliana chromosome 5 sequence

Sequence ID: NC\_003076.8 Length: 26975502

Range 1: 26936521 to 26937120

Score:189 bits (479), Expect:6<sup>e-60</sup>,

Method:Compositional matrix adjust.,

Identities:94/200(47%), Positives:132/200(66%), Gaps:28/200(14%)

```

Query 106      KVELHYFHENASLATVVGTKPSPVVELSGTVGAQGVTFGAEAGYDTASGKFTKYTAAIGL 165
                ++E+ YFH++A++      K +P+++++ T+G+  ++FGAEAGYDT S  FTKY A I +
Sbjct 26936521 QLEVQYFHDHATVTAATAALKQNPLIDITATLGSPVISFGAEAGYDTTSKTFTKYNAGISV 26936700

Query 166      TKPDYHAAFIL-----ADKGDITIKVSGVYHLDEKQKA 197
                TKPD  + IL                      DKGD++K S ++H DE ++
Sbjct 26936701 TKPDAACLSIIL*EYKFQNKAITSSRCGLCGSHLSFFCSRGDKGDCLKASYLHHFDEFKRT 26936880

Query 198      SAVAELTRRLSTNVNLTLVGGLYKIDPQTAVKARLNNTGTLAALLQHELKPKSLLTISGE 257
                +AV E+ R+ STN NT+TVGGLY ID  TAVKA+LNN GTL ALLQHE+ P+SL+T+S E
Sbjct 26936881 AAVGEVYRKfstnEntITVGGLYAIDHSTAVKAKLNNHGTLGALLQHEVLPRLVTVSSE 26937060

Query 258      FDTKALDRAPKFGLALALKP 277
                DTKAL++ P+FGL+LALKP
Sbjct 26937061 IDTKALEKHPRFGLSLALKP 26937120

```

### 8.3 Alineamiento OSVDAC2

>Oryza sativa mRNA for voltage-dependent anion channel (VDAC2 gene)

Sequence ID: AJ251562.1 Length: 1189

Range 1: 70 to 879

Score:206 bits (523), Expect:1e<sup>-62</sup>,

Method:Compositional matrix adjust.,

Identities:124/273(45%), Positives:165/273(60%), Gaps:5/273(1%)

```

Query   7      GLFSDIGKKAKDLLTRDYTYDQKltvstvsssgvgltstAVKKGGLYTLDVSSVYKYKNT  66
          GL+SDIGKKA+DLL RDY  DQK T++T +++G  +T   KK      ++ S  K  N
Sbjct  70      GLYSDIGKKARDLLYRDYHTDQKFLLTTYYAANGAAITVAGTKKNESIFSEIQSQVKNNNV  249

Query   67      VVDIKVDTESNISTTLTVLD-ALPSTKLVTSVKLPDYNSGKVELHYFHENASLATVVGTK  125
          VD+K  ++S  + TT TV D   P  K  + S+  P   S  K  E+ Y  H  +A  L  +VG
Sbjct  250      SVDVKATSDSKLITTTFTVHDLGTPGLKGILSIPFPYQKSAKAEVQYLHPHAGLNAIVGLN  429

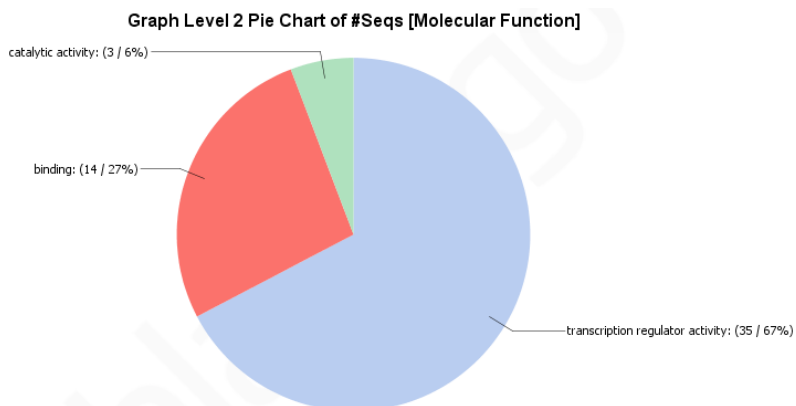
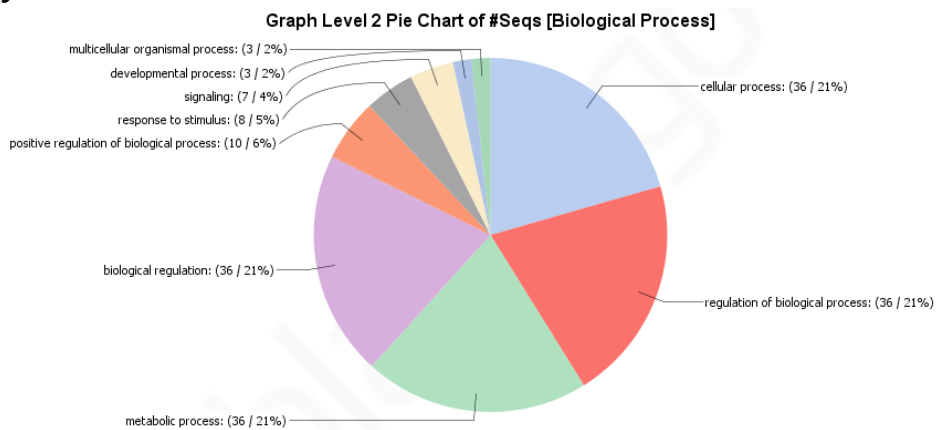
Query   126     PSPVVELSGTVGAQGVTFGAEAGYDTASGKFTKYTAAIGLTKPDYHAAFILADKGDTIKV  185
          +P+V  SG  G          G  +  +DTA+G FTKY  A  +  T  PD  AA  +L  +KGD++
Sbjct  430      ANPLVSFSGVFGTGAFVAGTDVAFDTATGDFTKYNAGLSHTTTPDLTAALLLNKGDSLAA  609

Query   186     SGVYHLDEKQKASAV-AELTRRLSTNVNTLTVGGLYKIDPQTAVKARLNNTGTLAALLQH  244
          S  YH  +  K  SAV  AEL      S+N  NTLT  G  +  +D  T  VKAR  NN  G  +AL+QH
Sbjct  610      S-YYH--KVSKTSAVGAELAHSFSSNENTLTFGTQHALDELTTVKARFNNFGMASALIQH  780

Query   245     ELKPKSLLTISGEFDTKALDRAPKFGALALALKP  277
          E  +PKSL+TIS  E  DTKA+D++  K  GL+L  LKP
Sbjct  781      EFRPKSLVTISTEVDTKAIDKSSKVGLSLVLKP  879

```

## 8.4 Términos ontológicos (GO) asociadas a la categoría de procesos biológicos y Funciones Moleculares



## 8.5 Ubicación de Región Promotora de probables Pg bZIP seleccionados y bZIP ortólogos

<b>bZIP</b>	<b>Región Promotora seleccionada</b>
<b>Pg1</b>	Cr2:3978000-3981000
<b>Pg4</b>	Cr2:222689500-22692000
<b>Pg5</b>	Cr2:193196802-93198357
<b>Pg21</b>	Cr7:3989000-3991500
<b>Pg27</b>	Cr6:228475698-28477448
<b>Pg37</b>	Cr1:176831019-76832601
<b>Pg46</b>	Cr3:249786695-49789500
<b>Pg47</b>	Cr3:11481000-11483000
<b>Pg52</b>	Cr5:84373500-84376000
<b>ATbZIP61</b>	Cr3:c21524787-21522243
<b>ATHY5</b>	Cr5:3596000-3593500
<b>ATAREB1</b>	Cr1:17169500-17167000
<b>Os bZIP19</b>	Cr2:8322488-8324100
<b>Os bZIP23</b>	Cr2:c32283000-32281000
<b>Os bZIP39</b>	Cr5:20118098-20116200
<b>OsRISBZ4</b>	Cr2:4108000-4110743
<b>OsRISBZ2</b>	Cr3:c33193500-33191000
<b>Os bZIP60</b>	Cr7:26807783-26811000
<b>OsHY5</b>	Cr1:c3813362-3811900
<b>OsRF2a</b>	Cr9:c20099800-20096675
<b>OsABI</b>	Cr6:30562500-30565000

## 8.6 Herramientas del Galaxy

### 8.6.1 Hisat2

The screenshot shows the Galaxy interface with the Hisat2 tool output. The main panel displays a table of sequencing reads with columns for QNAME, FLAG, RNAME, POS, MAPQ, CIGAR, MRNM, MPOS, ISIZE, and SEQ. The left sidebar contains various tool categories like 'Get Data', 'Collection Operations', and 'GENOMIC FILE MANIPULATION'. The right sidebar shows a 'History' panel with a search bar and a list of recent jobs.

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	MRNM	MPOS	ISIZE	SEQ
@HD VN:1.0 SO:coordinate									
@SQ SN:chr1 LN:275469374									
@SQ SN:chr2 LN:243054348									
@SQ SN:chr3 LN:300907625									
@SQ SN:chr4 LN:191830058									
@SQ SN:chr5 LN:158694342									
@SQ SN:chr6 LN:240570548									
@SQ SN:chr7 LN:154011256									
@PG ID:hisat2 PN:hisat2 VN:2.1.0 CL:/cvmfs/main.galaxyproject.org/deps/_conda/envs/mulled-v1-def0c4b1f9c0afb8225988d7892604b8cd4d0b7e8df6963									
@PG ID:samtools PN:samtools PP:hisat2 VN:1.11 CL:samtools sort -l 0 -T /galaxy-repl/main/jobdir/034/031/34031954/_job_tmp -O bam									
@PG ID:samtools.1 PN:samtools PP:samtools VN:1.11 CL:samtools view -O bam -@ 6 -o /galaxy-repl/main/files/053/033/dataset_53033214.dat									
8152393/1	0	chr1	5880	60	150M	*	0	0	TTGCAGCCATTGAATCTTCATCGGATATTT
1916570/2	0	chr1	5891	60	150M	*	0	0	GAATCTTCATCGGATATTTTATACCTACTC
1916571/2	0	chr1	5891	60	150M	*	0	0	GAATCTTCATCGGATATTTTATACCTACTC
717698/1	0	chr1	5895	60	150M	*	0	0	CTTCATCGGATATTTTATACCTACTCACC
984378/1	0	chr1	5896	60	25148M	*	0	0	TTTTCATGGATATTTTATACCTACTCACC
988670/1	0	chr1	5896	60	25148M	*	0	0	TTTTCATGGATATTTTATACCTACTCACC
676815/2	0	chr1	5901	60	150M	*	0	0	GGGATATTTTATACCTACTCAGCGAAGCA
7055315/1	0	chr1	5901	60	150M	*	0	0	GGGATATTTTATACCTACTCAGCGAAGCA
8282207/2	0	chr1	5901	60	150M	*	0	0	GGGATATTTTATACCTACTCAGCGAAGCA
717698/2	16	chr1	5917	60	150M	*	0	0	TACTCAGCGAAGCAAGCAAAATAGTCAGC
7055315/2	16	chr1	5926	60	150M	*	0	0	AACGACACAAATAGTTCAGCTGAAATAAA
8152393/2	16	chr1	5926	60	150M	*	0	0	AACGTCATACATAGTTCAGCTGAAATAAA
984378/2	16	chr1	5938	60	150M	*	0	0	AGTTCAGCTGAAATAAAAGCAAACTAG
988670/2	16	chr1	5938	60	150M	*	0	0	AGTTCAGCTGAAATAAAAGCAAACTAG
676815/1	16	chr1	5959	60	150M	*	0	0	ACAACATGATGATGATTTGGCAACACCG
7470088/2	0	chr1	5960	60	146M4S	*	0	0	CAACATGATGATGATTTGGCAACACGTA
7470088/1	16	chr1	5960	60	45146M	*	0	0	ATCTCAACATGATGATGATTTGGCAACAA

### 8.6.2 StringTie

The screenshot shows the Galaxy interface with the StringTie tool output. The main panel displays a table of transcript annotations with columns for Seqname, Source, Feature, Start, End, Score, Strand, Frame, and Attributes. The left sidebar contains various tool categories like 'Get Data', 'Collection Operations', and 'GENOMIC FILE MANIPULATION'. The right sidebar shows a 'History' panel with a search bar and a list of recent jobs.

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
# stringtie /galaxy-repl/main/files/053/033/dataset_53033214.dat -o /galaxy-repl/main/files/053/111/dataset_53111464.dat -p 6 -G guide.gff -b /special_o								
# StringTie version 2.1.1								
chr1	StringTie	transcript	38662	39102	1000	+	.	gene_id "STRG.1"; transcript_id "STRG.1.1"; cov "3.014706"; FPK
chr1	StringTie	exon	38662	38848	1000	+	.	gene_id "STRG.1"; transcript_id "STRG.1.1"; exon_number "1"; c
chr1	StringTie	exon	38950	39102	1000	+	.	gene_id "STRG.1"; transcript_id "STRG.1.1"; exon_number "2"; c
chr1	StringTie	transcript	5880	8229	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.1"; cov "17.216969"; FP
chr1	StringTie	exon	5880	6767	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.1"; exon_number "1"; c
chr1	StringTie	exon	6871	7332	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.1"; exon_number "2"; c
chr1	StringTie	exon	7708	8021	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.1"; exon_number "3"; c
chr1	StringTie	exon	8106	8229	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.1"; exon_number "4"; c
chr1	StringTie	transcript	6132	7287	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.2"; reference_id "Pgl_Gl
chr1	StringTie	exon	6132	6767	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.2"; exon_number "1"; r
chr1	StringTie	exon	6871	7287	1000	-	.	gene_id "STRG.2"; transcript_id "STRG.2.2"; exon_number "2"; r
chr1	StringTie	transcript	53857	55620	1000	+	.	gene_id "STRG.3"; transcript_id "STRG.3.1"; reference_id "Pgl_Gl
chr1	StringTie	exon	53857	55620	1000	+	.	gene_id "STRG.3"; transcript_id "STRG.3.1"; exon_number "1"; r
chr1	StringTie	transcript	55526	61915	1000	-	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; cov "7.489635"; FPK
chr1	StringTie	exon	55526	55662	1000	-	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; exon_number "1"; c
chr1	StringTie	exon	59265	59533	1000	-	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; exon_number "2"; c
chr1	StringTie	exon	60449	60513	1000	-	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; exon_number "3"; c
chr1	StringTie	exon	60623	60763	1000	-	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; exon_number "4"; c
chr1	StringTie	exon	61509	61915	1000	-	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; exon_number "5"; c
chr1	StringTie	transcript	56986	58494	1000	+	.	gene_id "STRG.5"; transcript_id "STRG.5.1"; reference_id "Pgl_Gl
chr1	StringTie	exon	56986	58494	1000	+	.	gene_id "STRG.5"; transcript_id "STRG.5.1"; exon_number "1"; r
chr1	StringTie	transcript	175353	175948	1000	-	.	gene_id "STRG.6"; transcript_id "STRG.6.1"; cov "8.018456"; FPK
chr1	StringTie	exon	175353	175948	1000	-	.	gene_id "STRG.6"; transcript_id "STRG.6.1"; exon_number "1"; c
chr1	StringTie	transcript	177226	177552	1000	-	.	gene_id "STRG.7"; transcript_id "STRG.7.1"; cov "8.027523"; FPK
chr1	StringTie	exon	177226	177552	1000	-	.	gene_id "STRG.7"; transcript_id "STRG.7.1"; exon_number "1"; c
chr1	StringTie	transcript	198150	198353	1000	-	.	gene_id "STRG.8"; transcript_id "STRG.8.1"; cov "5.808824"; FPK

## 8.6.3 FeatureCounts

**Galaxy** Workflow Visualize Shared Data Help User Using 70%

**Tools** search tools Upload Data

**Get Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

**Filter and Sort**

**Join, Subtract and Group**

**Datamash**

**GENOMIC FILE MANIPULATION**

**FASTA/FASTQ**

**FASTQ Quality Control**

**SAM/BAM**

**BED**

**VCF/BCF**

**Nanopore**

**Convert Formats**

**Lift-Over**

**COMMON GENOMICS TOOLS**

Geneid	HISAT2 on data 40 and data 62: aligned reads (BAM)
Geneid	HISAT2 on data 40 and data 62: aligned reads (BAM)
MSTRG.1.1	14
MSTRG.2.1	4
Pgl_GLEAN_10002239	8
MSTRG.4.1	88
Pgl_GLEAN_10000167	0
Pgl_GLEAN_10002237	90
MSTRG.6.1	37
Pgl_GLEAN_10002238	0
Pgl_GLEAN_10002241	0
Pgl_GLEAN_10002240	100
MSTRG.10.1	0
Pgl_GLEAN_10002242	0
Pgl_GLEAN_10002243	4
MSTRG.12.1	372
MSTRG.13.1	21
Pgl_GLEAN_10002244	862
MSTRG.15.1	0
MSTRG.15.2	0
MSTRG.15.3	0
MSTRG.16.1	0
MSTRG.17.1	5
MSTRG.18.1	24
Pgl_GLEAN_10002245	41
MSTRG.20.1	18
Pgl_GLEAN_10002246	1
MSTRG.22.1	5
MSTRG.23.1	5

**History** search datasets

**Unnamed history** 240 shown, 295 deleted, 16 hidden 174.35 GB

- ata 136 and data 104: Counts
- 140: featureCounts on data 136 and data 103: Summary
- 139: featureCounts on data 136 and data 103: Counts
- 138: featureCounts on data 136 and data 102: Summary
- 137: 63C(851)featureCounts on data 136 and data 102: Counts
- 136: StringTie merge on data 119, data 135, and others
- 135: StringTie on data 119 and data 118: Assembled transcripts

## 8.6.4 DESeq2

**Galaxy** Workflow Visualize Shared Data Help User Using 70%

**Tools** search tools Upload Data

**Get Data**

**Collection Operations**

**GENERAL TEXT TOOLS**

**Text Manipulation**

**Filter and Sort**

**Join, Subtract and Group**

**Datamash**

**GENOMIC FILE MANIPULATION**

**FASTA/FASTQ**

**FASTQ Quality Control**

**SAM/BAM**

**BED**

**VCF/BCF**

**Nanopore**

**Convert Formats**

**Lift-Over**

**COMMON GENOMICS TOOLS**

GeneID	Base mean	log2(FC)	StdErr	Wald-Stats	P-value	P-adj
Pgl_GLEAN_10011583	174.700538289098	-7.22663745923078	0.582537137221133	-12.4054536569186	2.44126145491443e-35	4.26708e-35
Pgl_GLEAN_10030544	909.485630416985	-3.47164527521709	0.338804751325965	-10.246743180638	1.22398343555966e-24	1.06970e-24
Pgl_GLEAN_10030642	459.97823932331	-2.69546246052434	0.266654102397961	-10.1084604972684	5.06742798312266e-24	2.95245e-24
MSTRG.205.1.1	267.267445718002	-3.9589834149945	0.400875532926051	-9.87584197543132	5.29860853379943e-23	2.31535e-23
Pgl_GLEAN_10027239	168.083635424856	-4.87167881173521	0.505980820753727	-9.62818868208914	6.07915796942984e-22	2.12515e-22
Pgl_GLEAN_10035117	1053.38375911971	-2.52612588880864	0.265254088485433	-9.52341923637254	1.67572742251859e-21	4.88167e-21
Pgl_GLEAN_10016428	204.655420040401	-4.73863174789667	0.534429421321966	-8.86671197138657	7.5337453639312e-19	1.88117e-19
Pgl_GLEAN_10014390	253.05472903479	3.02730416627477	0.356950301199521	8.48102426612783	2.23220050235603e-17	4.87707e-17
MSTRG.3549.1	32.7141254238082	-5.51943144442559	0.704311979362569	-7.83662866194737	4.62804108319212e-15	8.98817e-15
Pgl_GLEAN_10033402	8702.66394298714	2.02343930350294	0.2726661870575	7.42093959408409	1.16292350235603e-13	2.0326e-13
Pgl_GLEAN_10031164	55.2409167921907	-3.45610196523162	0.481142112380466	-7.1831209039933	6.81377927413304e-13	1.08270e-13
Pgl_GLEAN_10018139	178.909113170927	-3.14732598362579	0.449631751531696	-6.99978587567325	2.56353995266106e-12	3.73400e-12
Pgl_GLEAN_10029195	143.474410028238	-2.39139711705353	0.343022923489609	-6.9715373326237	3.13496101280369e-12	3.91399e-12
MSTRG.17235.1	378.494708999701	-2.66045705941471	0.381248225476771	-6.97828050501133	2.98814600854454e-12	3.91399e-12
MSTRG.302.1	153.420482198397	-2.4652870700918	0.357255121627957	-6.90063464690965	5.1770724328042e-12	6.03266e-12
Pgl_GLEAN_10005834	407.902859733698	-2.01905417457581	0.294295783356469	-6.86062896161244	6.85580201847266e-12	6.65736e-12
Pgl_GLEAN_10011581	529.450205777543	-3.44963837866006	0.502707111221213	-6.86212369321759	6.78442769695555e-12	6.65736e-12
Pgl_GLEAN_10035147	2758.55189761659	-1.88563237134871	0.27462790937774	-6.86613525777928	6.59645339019794e-12	6.65736e-12
Pgl_GLEAN_10038486	91.5005386900699	-2.54197425414172	0.376149947459101	-6.75787480847146	1.4003044061389e-11	1.28820e-11
Pgl_GLEAN_10029441	359.609121809179	3.38416812445488	0.504440231918679	6.70875935407238	1.96286056857776e-11	1.65715e-11
MSTRG.13712.1	242.848395451953	-2.57783373765491	0.384367883029478	-6.70668349638673	1.9909744546571e-11	1.65715e-11
Pgl_GLEAN_10034848	63.6389790424697	3.17821636032721	0.490521380633645	6.47926163018961	9.2125575534642e-11	7.32310e-11
MSTRG.17236.1	188.940659643781	-2.87689312231493	0.453667878263496	-6.3414080214954	2.2767467004844e-10	1.73022e-10
Pgl_GLEAN_10019694	66.3434121981824	2.96330497705047	0.469049835552708	6.31767618798679	2.65525642655575e-10	1.93380e-10
Pgl_GLEAN_10026856	1989.81232970333	-2.2239023431703	0.353603072159411	-6.28926188546928	3.18978965028098e-10	2.23017e-10
Pgl_GLEAN_10034700	2033.66598909679	-2.21954597198078	0.364146631753341	-6.09519841305087	1.09301795315678e-09	7.34802e-09
Pgl_GLEAN_10036648	48.8903391104911	3.65830864392749	0.60307824039841	6.0660597370092	1.31086386360883e-09	8.48614e-09
Pgl_GLEAN_10001980	3978.12475775682	-2.4865882529268	0.411740397770414	-6.03921370453749	1.54867026075107e-09	9.66757e-09

**History** search datasets

**Unnamed history** 240 shown, 295 deleted, 16 hidden 174.35 GB

- DESeq2 results on data 141, data 139, and others
- 433: DESeq2 plots on data 141, data 139, and others
- 432: DESeq2 result file on data 141, data 139, and others
- 412: Cuffmerge on data 62, data 119, and data 411: merged transcripts
- 411: Cuffcompare on data 62, data 119, and others: combined transcripts
- 410: Cuffcompare on data 62, data 119, and others: transcript tracking
- 409: Cuffcompare on data 62, data 119, and others

## 8.6.5 DESeq2 Normalizado

Galaxy

Workflow Visualize Shared Data Help User Using 70%

Tools

search tools

Upload Data

Get Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

Lift-Over

COMMON GENOMICS TOOLS

featureCounts on data 136 and data 109: Counts	featureCounts on data 136 and data 110: Counts	featureCounts on data 136 and data 111: Counts
MSTRG.1.1	20.3520683816731	44.2953703148969
MSTRG.2.1	0	0
Pgl_GLEAN_10002239	0	3.75384494194041
MSTRG.4.1	118.041996613704	93.8461235485103
Pgl_GLEAN_10000167	0	0
Pgl_GLEAN_10002237	35.2769185282333	38.2892184077922
MSTRG.6.1	26.457688896175	16.5169177445378
Pgl_GLEAN_10002238	0	0
Pgl_GLEAN_10002241	0	0
Pgl_GLEAN_10002240	23.7440797786186	22.5230696516425
MSTRG.10.1	0	0
Pgl_GLEAN_10002242	0	0
Pgl_GLEAN_10002243	1.3568045587782	0
MSTRG.12.1	142.464478671711	184.689171143468
MSTRG.13.1	18.9952638228949	10.5107658374332
Pgl_GLEAN_10002244	908.380652102007	881.402792367609
MSTRG.15.1	0	0
MSTRG.15.2	0	0
MSTRG.15.3	0	0
MSTRG.16.1	16.2816547053384	11.2615348258212
MSTRG.17.1	9.49763191144743	8.25845887226891
MSTRG.18.1	0	0
Pgl_GLEAN_10002245	50.8801709541826	43.5446013265088
MSTRG.20.1	0	0
Pgl_GLEAN_10002246	0	2.25230696516425
MSTRG.22.1	0.678402279389102	2.25230696516425
MSTRG.23.1	21.7088729404513	23.2738386400306
Pgl_GLEAN_10012377	253.044050212135	271.027604808098

History

search datasets

Unnamed history

219 shown, 316 deleted, 16 hidden

174.35 GB

- 437: Normalized counts file on data 153, data 151, and others
- 436: DESeq2 plots on data 153, data 151, and others
- 435: DESeq2 result file on data 153, data 151, and others
- 434: Normalized counts file on data 141, data 139, and others
- 433: DESeq2 plots on data 141, data 139, and others
- 432: DESeq2 result file on data 141, data 139, and others
- 412: Cuffmerge on data

## 8.6.6 DESeq2 plots

