



Universidad Autónoma Chapingo



Departamento de Fitotecnia

Instituto de Horticultura

IDENTIFICACIÓN Y CARACTERIZACIÓN *in silico* DE LAS FAMILIAS GÉNICAS TPS Y TPP DEL METABOLISMO DE LA TREHALOSA EN JITOMATE (*Solanum lycopersicum* L.)

TESIS

Que como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS EN BIOTECNOLOGÍA AGRÍCOLA

Presenta:

IRMA LÓPEZ GONZÁLEZ

Bajo la supervisión del: DR. JOSÉ OSCAR MASCORRO GALLARDO



APROBADA



Chapingo, Estado de México, junio de 2021



IDENTIFICACIÓN Y CARACTERIZACIÓN *in silico* DE LAS FAMILIAS
GÉNICAS TPS Y TPP DEL METABOLISMO DE LA TREHALOSA EN
JITOMATE (*Solanum lycopersicum* L.

Tesis realizada por **IRMA LÓPEZ GONZÁLEZ** bajo la supervisión del Comité
Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para
obtener el grado de:

MAESTRO EN CIENCIAS EN BIOTECNOLOGÍA AGRÍCOLA

Director:



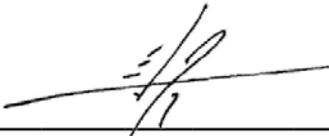
Dr. José Oscar Mascorro Gallardo

Asesor:



Dr. José Luis Rodríguez de la O

Asesor:



Dr. Juan Enrique Rodríguez Pérez

CONTENIDO

	Pág.
.....	i
LISTA DE CUADROS.....	vi
LISTA DE FIGURAS.....	vii
LISTA DE APÉNDICES.....	<u>ix</u>
DEDICATORIA.....	xi
AGRADECIMIENTOS.....	xii
DATOS BIOGRÁFICOS.....	<u>xiii</u>
RESUMEN.....	<u>xiv</u>
ABSTRACT.....	<u>xv</u>
I. INTRODUCCIÓN.....	1
II. OBJETIVOS.....	3
4.1 Objetivo general.....	3
4.2 Objetivo específico.....	3
III. HIPÓTESIS.....	3
IV. REVISIÓN DE LITERATURA.....	4
4.1 Trehalosa.....	4
4.1.1. Vías de biosíntesis de trehalosa.....	4
4.1.2. Papel de la trehalosa en el crecimiento y desarrollo.....	5
4.1.3. Papel de la trehalosa durante estrés abiótico.....	5
4.2 Métodos bioinformáticos utilizados.....	8
4.2.1. Herramientas para la identificación de genes en genoma completo ..	9

4.2.2. Árboles evolutivos.....	9
4.2.2.1. Construcción de un árbol de máxima verosimilitud (ML).....	10
4.2.3. Visualización de alineación de secuencia múltiple	11
4.2.4. Análisis de secuencias proteicas	11
4.2.4.1. Predicción de la localización subcelular de proteínas	11
4.2.4.2. Predicción de péptidos señal y NES en proteínas	11
4.2.4.3. Estructura genética	12
4.2.4.4. Dominios conservados	13
4.2.4.5. Motivos.....	13
4.2.5. Análisis de estructuras proteicas	14
4.2.5.1. Análisis de estructura primaria de proteínas.....	14
4.2.5.2. Análisis de estructura secundaria de proteínas	14
4.2.5.3. Predicción de la estructura terciaria de proteínas mediante modelos de homología.....	15
4.2.6. Análisis de duplicación genética	16
4.2.7. Análisis de elementos reguladores <i>cis</i>	16
4.2.8. Análisis de expresión génica.....	17
4.2.7.1. Heat map.....	17
4.2.7.1. Volcano Plot.....	17
V. MATERIALES Y MÉTODOS.....	18
5.1. Identificación de miembros de las familias TPS y TPP en jitomate	18
5.2. Análisis de dominios de proteínas TPS/TPP en jitomate y otras especies	
18	

5.3.	Análisis filogenético de las familias TPS y TPP en jitomate	19
5.4.	Análisis de características genéticas y localización subcelular de proteínas de las familias TPS y TPP en jitomate	20
5.5.	Localización en cromosoma y análisis evolutivo de las familias TPS/TPP en jitomate	20
5.6.	Distribución exón-intrón de genes de las familias TPS/TPP en jitomate	21
5.7.	Análisis de motivos proteicos y dominios conservados de las familias TPS/TPP en jitomate.....	21
5.8.	Análisis de la estructura 3D de las proteínas TPS/TPP en jitomate	22
5.9.	Análisis de elementos reguladores cis en promotores de genes <i>TPS/TPP</i> en jitomate.....	22
5.10.	Análisis de expresión <i>in silico</i> de genes <i>TPS/TPP</i> en jitomate.....	22
VI.	RESULTADOS Y DISCUSIÓN.....	24
6.1.	Identificación de los miembros de las familias TPS y TPP en jitomate	24
6.2.	Análisis de dominios de proteínas TPS/TPP en jitomate y otras especies	25
6.3.	Análisis filogenético de las familias <i>TPS/TPP</i> en jitomate.....	28
6.4.	Localización subcelular de proteínas de las familias TPS/TPP en jitomate	31
6.5.	Localización en cromosoma y análisis de duplicación genética de las familias <i>TPS/TPP</i> en jitomate.....	34
6.6.	Distribución exón-intrón de genes <i>TPS/TPP</i> en jitomate.....	37

6.7.	Análisis de dominios conservados y motivos proteicos de las familias TPS/TPP en jitomate.....	38
6.8.	Análisis de estructura 3D de proteínas TPS/TPP en jitomate	42
6.9.	Análisis de elementos reguladores <i>cis</i> en promotores de genes <i>TPS/TPP</i> en jitomate.....	45
6.9.1	Elementos reguladores <i>cis</i> relacionados con la luz	45
6.9.2	Elementos reguladores <i>cis</i> relacionados con el estrés	47
6.9.3	Elementos reguladores <i>cis</i> relacionados con regulaciones hormonales.....	50
6.9.4	Elementos reguladores <i>cis</i> relacionados con el desarrollo celular .	53
6.9.5	Elementos reguladores <i>cis</i> centrales ¡Error! Marcador no definido.	
6.9.6	Otros elementos reguladores <i>cis</i>	54
6.10.	Análisis de expresión <i>in silico</i> de genes TPS/TPP en jitomate.....	55
6.10.1	Niveles de expresión de los genes <i>SITPS/SITPP</i> en condiciones de sequía	56
6.10.2	Niveles de expresión de los genes <i>SITPS/SITPP</i> en condiciones de calor	60
6.10.3	Niveles de expresión de los genes <i>SITPS/SITPP</i> en condiciones de frío	63
6.10.4	Niveles de expresión de los genes <i>SITPS/SITPP</i> en condiciones de salinidad	65
VII.	CONCLUSIONES	68
VIII.	REFERENCIAS BIBLIOGRÁFICAS	69
IX.	APÉNDICES	84

LISTA DE CUADROS

Cuadro 1. Información y características de los miembros de las familias TPS y TPP en <i>Solanum lycopersicum</i>	27
Cuadro 2. Predicción de péptidos de tránsito al cloroplasto y/o mitocondria y secuencias NLS y NESs.	34
Cuadro 3. Estimaciones de las relaciones Ka/Ks de los genes <i>SITPS/SITPP</i> duplicados segmentadamente con su tiempo de divergencia en jitomate.	37

LISTA DE FIGURAS

Figura 1. Distribución de los dominios TPS y TPP en diferentes genes de la biosíntesis de la trehalosa..	28
Figura 2. Árbol filogenético circular de TPSs y TPPs.....	31
Figura 3. Árbol filogenético rectangular de SITPS/SITPP.....	31
Figura 4. Distribución de los 10 genes SITPS y 9 genes SITPP en los 12 cromosomas de jitomate.....	35
Figura 5. Análisis de sintenia de genes TPS/TPP entre jitomate y arabidopsis..	37
Figura 6. Representación esquemática de la distribución de exón-intrón, dominios y motivos de SITPS/SITPP..	39
Figura 7. Representación esquemática del alineamiento de todas las proteínas SITPSs.	41
Figura 8. Representación esquemática del alineamiento de todas las proteínas SITPPs..	41
Figura 9. Representación esquemática de las estructuras 3-D de las proteínas y la gráfica de Ramachandran para SITPSs.	42
Figura 10. Representación esquemática de las estructuras 3-D de las proteínas y la gráfica de Ramachandran para SITPPs..	44
Figura 11. Elementos reguladores cis relacionados con la luz en SITPS/SITPPs.	47
Figura 12. Elementos reguladores cis relacionados con el estrés en SITPS/SITPPs..	50
Figura 13. Elementos reguladores cis relacionados con regulaciones hormonales en SITPS/SITPPs.....	53
Figura 14. Elementos reguladores cis relacionados con el desarrollo celular en SITPS/SITPPs..	54

Figura 15. Otros elementos reguladores cis en SITPS/SITPPs.....	55
Figura 16. Patrón de expresión de genes SITPS y SITPP en respuesta a sequía.	59
Figura 17. Patrón de expresión de genes SITPS y SITPP en respuesta a calor..	62
Figura 18. Patrón de expresión de genes SITPS y SITPP en respuesta a frío..	64
Figura 19. Diagramas de volcán de expresión de SITPS/SITPP en respuesta a frío.....	64
Figura 20. Patrón de expresión de genes SITPS y SITPP en respuesta a salinidad.	66
Figura 21. Diagramas de volcán de expresión de SITPS/SITPP en respuesta a salinidad.....	66
Figura 22. Diagrama de Venn que muestra la expresión de genes SITPS bajo estrés por sequía, calor y salinidad.....	67
Figura 23. Diagrama de Venn que muestra la expresión de genes SITPP bajo estrés por sequía, calor y salinidad.....	68

LISTA DE APÉNDICES

Apéndice 1. Las secuencias proteicas TPS usadas para construir los árboles filogenéticos.....	84
Apéndice 2. Las secuencias proteicas TPP usados para construir los árboles filogenéticos.....	85
Apéndice 3. Identidad y porcentaje de cobertura entre SITPSs.	87
Apéndice 4. Identidad y porcentaje de cobertura entre SITPPs.	88
Apéndice 5. Identidad y porcentaje de cobertura entre AtTPS y SITPS.	89
Apéndice 6. Identidad y porcentaje de cobertura entre AtTPP y SITPP.	90
Apéndice 7. Motivos detectados en los genes SITPS/SITPP.	91
Apéndice 8. Detalles de los elementos de la estructura secundaria de SITPSs calculados por SOPMA.....	92
Apéndice 9. Validación de las estructuras de las proteínas SITPS.	93
Apéndice 10. Detalles de los elementos de la estructura secundaria de SITPPs calculados por SOPMA.....	93
Apéndice 11. Validación de las estructuras de las proteínas SITPP.	94
Apéndice 12. Elementos Regulatorios Cis (CREs) con secuencias y funciones en los promotores de SITPSs.	95
Apéndice 13. Elementos Regulatorios Cis (CREs) con secuencias y funciones en los promotores de SITPPs.	99
Apéndice 14. Elementos cis relacionados con la luz (LrCRE) encontrados en SITPSs.	103
Apéndice 15. Elementos cis relacionados con la luz (LrCRE) encontrados en SITPPs.	103
Apéndice 16. Elementos cis relacionados con el estrés (SrCRE) encontrados en SITPSs.	104

Apéndice 17. Elementos cis relacionados con el estrés (SrCRE) encontrados en SITPPs..	104
Apéndice 18. Elementos cis relacionados con las hormonas (HrCRE) encontrados en SITPPs.	105
Apéndice 19. Elementos cis relacionados con las hormonas (HrCRE) encontrados en SITPPs.	105
Apéndice 20. Elementos cis relacionados con el desarrollo celular (CdCRE) encontrados en SITPPs..	106
Apéndice 21. Elementos cis relacionados con el desarrollo celular (CdCRE) encontrados en SITPPs.	106
Apéndice 22. Otros elementos cis encontrados en SITPPs..	107
Apéndice 23. Otros elementos cis encontrados en SITPPs..	107

DEDICATORIA

A mis padres Natalia González Hernández y Antolín Antonio López Martínez por sus consejos y apoyo incondicional. Este logro también es de ustedes.

A mis hermanos Emma, Aracely y Jasiel por estar en los momentos más importantes de mi vida.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada durante mis estudios de maestría.

A la Universidad Autónoma Chapingo y al Instituto de Horticultura del Departamento de Fitotecnia por darme la oportunidad de cursar un posgrado.

A la Fundación Telmex-Telcel por la beca otorgada, la cual me da más ánimo para seguir adelante con mis estudios.

Al Dr. José Oscar Mascorro Gallardo por el tiempo, dedicación y paciencia en la elaboración de este documento.

A mis asesores Dr. José Luis Rodríguez de la O y Dr. Juan Enrique Rodríguez Pérez por su tiempo, recomendaciones y guía durante mis estudios.

Al Dr. Muhammad Zulfiqar Ahmad por su disponibilidad para compartir su conocimiento sobre estudios de genoma completo y por sus atentas respuestas a las inquietudes surgidas durante el desarrollo de este proyecto

Al Biól. César Ramiro Martínez González, por compartir sus conocimientos en bioinformática con nosotros.

A mis profesores de la maestría, en especial al Dr. Luis Fernando Contreras Cruz por sus consejos y por su esfuerzo y dedicación para enseñar.

A mis compañeros de la maestría por su apoyo y constancia, y por compartir horas de estudio.

A mis amigos Asu, Víctor Manuel, Tranqui y José Miguel que siempre estuvieron para ayudarme, escucharme y aconsejarme.

A todas las personas que ayudaron directa e indirectamente en la realización de este proyecto.

DATOS BIOGRÁFICOS

Datos personales

Nombre	Irma López González
Fecha de nacimiento	Octubre 12 1996
Lugar de nacimiento	San Juan Mixtepec, Miahuatlán, Oaxaca
CURP	LOGI961012MOCPNR08
Cédula profesional	

Desarrollo académico

Bachillerato	Preparatoria Agrícola, Universidad Autónoma Chapingo
Licenciatura	Ingeniera Agrónoma Especialista en Fitotecnia, Universidad Autónoma Chapingo, 2018

RESUMEN

IDENTIFICACIÓN Y CARACTERIZACIÓN *in silico* DE LAS FAMILIAS GÉNICAS TPS Y TPP DEL METABOLISMO DE LA TREHALOSA EN JITOMATE (*Solanum lycopersicum* L.

Los genes de biosíntesis de la trehalosa en plantas contienen dos familias, trehalosa-6-fosfato sintasas (TPSs) y trehalosa-6-fosfatasa (TPPs). Ambas familias juegan un papel importante en el crecimiento, desarrollo y tolerancia al estrés de las plantas. Estudios anteriores han hecho grandes esfuerzos en la identificación de los miembros de las dos familias en jitomate, pero todavía no ha habido un análisis enfocado al perfil de expresión de las familias de genes TPS/TPP en esta especie. En este estudio encontramos diez genes *SITPS* y nueve genes *SITPP* en el genoma del jitomate, cada familia se clasificó en dos clases. Los miembros clase I y clase II de *SITPS* difieren en estructuras genéticas, dominios y motivos conservados, lo que implica funciones genéticas variadas. El análisis de duplicación genética mostró que cuatro pares de genes *SITPS* y dos pares de genes *SITPP* están formados por eventos de duplicación segmental. El valor de Ka/Ks de los genes era inferior a 1, lo que sugiere selección purificadora en estas familias genéticas. Los elementos regulatorios *cis* mostraron que muchos miembros de las familias TPS y TPP pueden estar relacionados con el estrés, las hormonas y la luz. El análisis de los perfiles de expresión reveló que 9 genes *SITPS* y 8 genes *SITPP* se expresaron diferencialmente en condiciones de estrés abiótico. Estos hallazgos establecen las bases para futuros estudios experimentales y funcionales de estas familias de genes y su papel en el metabolismo, el desarrollo y la respuesta al estrés en jitomate.

Palabras clave: bioinformática, estrés abiótico, expresión génica, Ka/Ks , sintenia

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION *in silico* OF TPS AND TPP GENE FAMILIES OF METABOLISM OF TREHALOSE IN TOMATO (*Solanum lycopersicum* L.)

The trehalose biosynthesis genes in plants contain two gene families, trehalose-6P synthase (TPSs) and trehalose-6P phosphatase (TPPs). Both families play an important role on growth, development and stress tolerance of plants. Previous studies have made great efforts to identify the members of the two families in tomato, but there has not yet done an analysis focused on the expression profile of the TPS / TPP gene families in the specie. In this study we found ten *SITPS* genes and nine *SITPP* genes in the tomato genome, each family was classified into two classes. The class I and class II members of *SITPS* differ in genetic structures, domains, and conserved motifs, implying varied genetic functions. Genetic duplication analysis showed that four pairs of *SITPS* genes and two pairs of *SITPP* genes are formed by segmental duplication events. The K_a / K_s value of the genes was less than 1, suggesting purifying selection in these gene families. The *cis* regulatory elements showed that many members of the TPS and TPP families may be involved in responses to stress, hormones and light. The analysis of the expression profiles revealed that 9 *SITPS* genes and 8 *SITPP* genes were differentially expressed under abiotic stress conditions. These findings establish basis for future experimental and functional studies of these gene families and their role in metabolism, development and stress response in tomato.

Key words: abiotic stress, bioinformatics, gene expression, K_a/K_s , syteny

Thesis, Universidad Autónoma Chapingo
Author: Irma López González
Advisor: Dr. Jose Oscar Mascorro Gallardo

I. INTRODUCCIÓN

Las condiciones de estrés abiótico como sequía, calor, frío y salinidad reducen la calidad y el rendimiento de los cultivos (Lomeshwar *et al.*, 2020), por lo que son una grave preocupación para la seguridad alimentaria. Además, se prevé que las tensiones ambientales empeoren con el calentamiento global y el cambio climático previsto (Sallam *et al.*, 2019). Por lo tanto, se requieren cultivos tolerantes a estrés abiótico para satisfacer las demandas alimentarias proyectadas.

Las plantas han desarrollado varias estrategias para superar los desafíos de tensiones abióticas, ya sea adoptando un mecanismo adaptativo que les permita sobrevivir a las condiciones adversas o hábitos de crecimiento específicos para evitar condiciones de estrés (Parul *et al.*, 2016).

Un mecanismo efectivo para reducir el daño de estrés por sequía y salinidad es la producción y acumulación de altos niveles intracelulares de compuestos orgánicos, altamente solubles, de bajo peso molecular, eléctricamente neutros y no tóxicos, conocidos como osmolitos u osmoprotectores (Per *et al.*, 2017).

Los osmoprotectores ayudan en el aumento de la tolerancia al estrés al proteger las membranas biológicas, estabilizar la estructura de las proteínas y otras estructuras celulares, desintoxicar las ROS y mantener el equilibrio redox celular (Zulfiqar *et al.*, 2019). Según sus propiedades químicas, los osmoprotectores se clasifican en: aminoácidos (prolina, ectoína, ácido pipercolico, etc.), betaínas (glicina betaína, colina-O-sulfato, β -alanina betaína, etc.), alcoholes de azúcar y azúcares no reductores (trehalosa, sorbitol, inositol, manitol, etc.) (Slama *et al.*, 2015).

La trehalosa (α -D-glucopiranosil-1,1- α -D-Glucopiranosido), un disacárido no reductor compuesto por dos moléculas de glucosa (Lunn *et al.*, 2015), en las plantas funciona estabilizando las proteínas y las membranas lipídicas en condiciones de estrés por sequía, frío, calor y salinidad (Garg *et al.*, 2002).

Existen al menos cinco vías biosintéticas de trehalosa en diferentes organismos

(Avonce *et al.*, 2006), pero en las plantas superiores sólo existe una, la vía TPS-TPP. En esta ruta, la trehalosa se sintetiza mediante la catálisis de dos enzimas: trehalosa-6-fosfato sintasa (TPS) y trehalosa-6-fosfato fosfatasa (TPP). Primero TPS cataliza la unión de UDP-glucosa y glucosa 6-fosfato para formar trehalosa-6-fosfato (T6P) y UDP. Después TPP desfosforila la T6P para producir trehalosa libre. La trehalasa (TRE) es una enzima que hidroliza específicamente la trehalosa, produciendo glucosa (Chen y Haddad, 2004). En plantas, los genes TPS y TPP constituyen dos familias multigénicas, mientras que TRE está presente como un gen de copia única en la mayoría de los genomas de plantas secuenciadas (Lunn *et al.*, 2007).

Además de ser una vía para producir trehalosa, en plantas superiores se ha descubierto que actúan como moléculas de señalización que modulan diversos procesos metabólicos y de desarrollo en las plantas (Acosta-Pérez *et al.*, 2020). La perturbación de esta vía conduce a alteraciones en el metabolismo del carbono y nitrógeno, desarrollo de las hojas, tiempo de floración, ramificación de brotes, función estomatal y tolerancia a tensiones abióticas y bióticas en *Arabidopsis* y otras especies (Figueroa y Lunn, 2016).

El jitomate (*Solanum lycopersicum*) es un cultivo con gran importancia económica, que también funciona como un sistema modelo para plantas de frutos carnosos (Kimura y Sinha, 2008). Su crecimiento y desarrollo se ven fácilmente afectados por diversas tensiones ambientales, como la sequía, la salinidad y el calor (Yu *et al.*, 2019), lo que hace necesario implementar estrategias que minimicen las pérdidas causadas por el estrés en este cultivo.

Bajo este contexto, en el presente trabajo se identificaron los miembros de las familias TPS/TPP en jitomate y se caracterizaron en detalle utilizando métodos bioinformáticos, se analizó la ubicación de los genes en los cromosomas, relación evolutiva, estructura intrón/exón, motivos proteicos y dominios conservados, estructura 3D, elementos regulatorios *cis* y los patrones de expresión bajo diversos tipos de estrés. Al llevar a cabo este trabajo sobre las familias TPS y TPP se espera que apoye futuras investigaciones experimentales de estas familias de genes en jitomate, con la finalidad de mejorar la tolerancia de los

cultivos al estrés mediante transformación, edición genética y mejoramiento asistido por marcadores moleculares.

II. OBJETIVOS

4.1 Objetivo general

Identificar los genes TPS y TPP de biosíntesis de la trehalosa relacionados con el estrés abiótico en jitomate

4.2 Objetivo específico

- a) Identificar los genes TPS y TPP en el genoma completo de jitomate
- b) Describir las características genéticas de los genes SITPS y SITPP
- c) Analizar la evolución de los genes SITPS y SITPP
- d) Distinguir la estructura genética, motivos y dominios conservados de los SITPS y SITPP
- e) Mostrar la estructura 3D de las proteínas SITPS y SITPP
- f) Clasificar los elementos reguladores *cis* de los promotores de los genes SITPS y SITPP
- g) Comparar los perfiles de expresión de los genes SITPS y SITPP bajo diferentes tipos de estrés abiótico

III. HIPÓTESIS

El análisis *in silico* de los genes involucrados en la biosíntesis de trehalosa permitirá obtener importantes claves sobre su estructura, función y papel en la respuesta al estrés abiótico y en el crecimiento y desarrollo del jitomate

IV. REVISIÓN DE LITERATURA

4.1 Trehalosa

La trehalosa es un disacárido no reductor formado por dos monómeros de glucosa que están vinculados con un enlace 1,1-glicosídico. Este disacárido se encuentra en muchos organismos como bacterias, hongos, nematodos, artrópodos y plantas (Elbein *et al.*, 2003) pero no en vertebrados. Los niveles de trehalosa son insignificantes en la mayoría de las plantas superiores, a excepción de algunas especies de plantas de resurrección, por ejemplo, *S. lepidophylla*, donde su alto contenido de trehalosa se correlacionó con alta tolerancia al estrés (Iturriaga *et al.*, 2000).

La trehalosa y los intermedios en su biosíntesis tiene un papel importante en el metabolismo de las células vegetales para la tolerancia al estrés abiótico.

4.1.1. Vías de biosíntesis de trehalosa

En la naturaleza existen al menos cinco vías de biosíntesis de trehalosa (Avonce *et al.*, 2006). La vía más común es la TPS-TPP, se encuentra en una amplia gama de organismos a partir de eubacterias, arqueas, hongos, insectos y plantas. En esta vía, la trehalosa-6-fosfato sintasa (TPS) cataliza la transferencia de glucosa de UDP-glucosa a glucosa-6-fosfato (G6P), para producir el intermedio trehalosa-6-fosfato (T6P). La trehalosa-6-fosfatasa (TPP) desfosforila T6P a trehalosa (Elbein *et al.*, 2003). La vía TreZ-TreY se distribuye en arqueas termofílicas del género *Sulfolobus*, implica la conversión de maltooligosacáridos o almidón a trehalosa catalizada por la maltooligosyl trehalosa sintasa (TreY) y posteriormente maltooligosyl trehalosa trehalohidrolasa (TreZ) (Maruta *et al.*, 1996). La vía TreP, se encuentra en hongos como *Agaricus bisporus* y protista como *Euglena gracilis*, trehalosa fosforilasa (TreP) cataliza la síntesis reversible de trehalosa, a partir de glucosa-1-fosfato y glucosa (Wannet *et al.*, 1998; Belocopitow y Maréchal, 1970). En la vía TreS, la trehalosa sintasa (TreS) isomeriza el enlace α 1- α 4 de la maltosa al enlace α 1- α 1 de trehalosa. Esta vía se informó por primera vez en *Pimelobacter* sp. (Elbein *et al.*, 2003) y se ha encontrado sólo en bacterias.

La vía TreT implica trehalosa glicosiltransferasa (TreT) que cataliza la conversión reversible de ADP-glucosa y glucosa a trehalosa. Esta vía se produce en extremófilos como *Thermococcus litoralis* y *Thermotoga maritima* (Qu *et al.*, 2004).

La trehalosa se hidroliza en dos unidades de glucosa por trehalasa (TRE). Esto se ha observado en muchos organismos incluyendo bacterias, hongos, plantas e insectos (Elbein *et al.*, 2003).

4.1.2. Papel de la trehalosa en el crecimiento y desarrollo

T6P desempeña un papel importante en la síntesis de almidón a través de la actividad redox postraduccional de la pirofosforasa ADP-glucosa (Kolbe *et al.*, 2005). T6P también funciona como inhibidor de la actividad de la proteína quinasa 1 relacionada con SNF1 (SnRK1), que regula los genes clave implicados en la asignación (SWEETs) y el uso (biosíntesis de almidón) de sacarosa. SnRK1 además tiene un papel fundamental en las redes de transcripción del estrés vegetal, metabolismo energético, crecimiento, desarrollo y supervivencia de las plantas (Baena-Gonzalez *et al.*, 2007). La vía biosintética y los genes de crecimiento reprimidos por SnRK1 son activados por T6P, mientras que los genes de la vía catabólica y el estrés que son inducidos por SnRK1 son reprimidos por T6P (Paul *et al.*, 2018).

En *Arabidopsis thaliana*, *AtTPS1* cataliza la síntesis de la trehalosa-6-fosfato (Tre6P), la cual es una molécula de señalización de sacarosa, y es esencial para la embriogénesis, el crecimiento, desarrollo y floración normal (Fichtner *et al.*, 2020), prueba de ello es que el mutante de *AtTPS1* es un gen letal recesivo de la embriogénesis (Eastmond *et al.*, 2006). *AtTPS6* regula la arquitectura de la planta, forma celular y ramificación de los tricomas (Chary *et al.*, 2008). *AtTPS11* regula la expresión del gen PHYTOALEXIN DEFICIENT4, que es un modulador clave de las defensas contra el pulgón verde del melocotón (Singh *et al.*, 2011).

4.1.3. Papel de la trehalosa durante estrés abiótico

Muchos estudios han sugerido que los genes *TPS* desempeñan un papel vital en la tolerancia al estrés en plantas, por ejemplo, en *Selaginella lepidophylla*, *TPS1* participó en la respuesta al estrés por calor y salinidad (Zentella *et al.*, 1999). Los

niveles de expresión de *TPS* en el algodón aumentaron bajo estrés por sequía (Kosmas *et al.*, 2006). La sobreexpresión del gen *TaTPS11* del trigo mejoró la tolerancia al frío e influyó en el metabolismo de carbohidratos en plantas de *Arabidopsis* (Liu *et al.*, 2019). La sobreexpresión de *OsTPS1* mejoró la tolerancia de plántulas de arroz al frío, alta salinidad y sequía sin otros cambios fenotípicos significativos, al incrementar la cantidad de trehalosa y prolina, y regular la expresión de genes relacionados con el estrés (Li *et al.*, 2011). Los genes *TPS* en maíz estaban regulados positivamente en respuesta al estrés por salinidad y frío (Jiang *et al.*, 2010).

Se han llevado a cabo muchas investigaciones para estudiar la sobreexpresión de genes biosintéticos de la trehalosa para aumentar la tolerancia al estrés abiótico en plantas. En *Arabidopsis thaliana*, líneas transgénicas con una fusión traduccional quimérica de levadura trehalosa-6-fosfato sintasa y trehalosa-6-fosfato fosfatasa mostraron un aumento significativo en la tolerancia a la sequía, salinidad, congelación y calor (Miranda *et al.*, 2007).

Por otro lado, *AtTPS1* es el único gen de la clase I que posee una extensión N-terminal, que se ha demostrado que es un dominio inhibitorio que regula la actividad TPS. El truncamiento de esta región N-terminal en *AtTPS1* condujo a un aumento de 10-40 veces la actividad TPS de la planta, lo que sugiere que las enzimas TPS de la planta poseen actividad catalítica con alto potencial (Van Dijck *et al.*, 2002).

La sobreexpresión de *AtTPS1* en *Arabidopsis* usando el promotor 35S confirió tolerancia a la deshidratación, sin alteraciones morfológicas visibles, salvo floración tardía y además condujo a un pequeño aumento en los niveles de trehalosa y trehalosa-6-P (Avonce *et al.*, 2004).

Se ha informado que los miembros de clase I de TPS, a excepción de *AtTPS3*, poseen actividad enzimática TPS1 en *Arabidopsis*. Tanto *AtTPS2* como *AtTPS4* pueden rescatar el defecto de crecimiento de la glucosa cuando se expresa en levadura *tps1Δtps2Δ* doble mutante, y la expresión de estos genes en el doble mutante de levadura también resultó en la acumulación de altos niveles de T6P cuando se cultivó en glucosa, incluso mucho más que *AtTPS1* (Delorge *et al.*,

2015). Estas proteínas TPS de clase I truncadas sólo se encuentran en las Brassicaceae, cuyo significado es desconocido (Lunn *et al.*, 2007). *AtTPS3* es probablemente un pseudogen, ya que contiene un codón de parada traduccional prematuro (Vandesteene *et al.*, 2010). Las proteínas TPS de clase II (*AtTPS5-AtTPS11*) no poseen actividad catalítica de TPS1 (Delorge *et al.*, 2015). Los genes *AtTPS5* (Tian *et al.*, 2019) *AtTPS6* (Chary *et al.*, 2008), *OsTPS8* (Vishal *et al.*, 2019) y *TaTPS11* (Liu *et al.*, 2019) codifican enzimas multifuncionales que poseen actividad de sintasa y fosfatasa (Yang *et al.*, 2020). Por ejemplo, *AtTPS5* influye en la termotolerancia al interactuar con MBF1c, un activador transcripcional (Suzuki *et al.*, 2008).

Con respecto a las TPP, las plantas que sobreexpresaban *AtTPPD* eran más tolerantes al estrés por alta salinidad, lo que se correlacionó con altos niveles de almidón y una mayor acumulación de azúcares solubles (Krasensky *et al.*, 2014). La sobreexpresión de *AtTPPF* condujo a un aumento del contenido de trehalosa y azúcares solubles totales en condiciones de sequía y participó en la regulación de la transcripción de genes sensibles al estrés (Lin *et al.*, 2019). *AtTPPG* está involucrado en la regulación del cierre estomatal (Van Houtte *et al.*, 2013). La sobreexpresión de *AtTPPI* aumentó significativamente la tolerancia a la sequía y mejoró la eficiencia en el uso del agua (WUE) al disminuir las aperturas estomatales y mejorar la arquitectura de raíces (Lin *et al.*, 2020). Además, *AtTPPI* desempeña un papel en la respuesta al estrés por salinidad, pero no parece ser esencial para el estrés oxidativo y osmótico (Kataya *et al.*, 2020).

En el arroz, la sobreexpresión regulada de un gen de fusión de genes biosintéticos de trehalosa de *Escherichia coli* (*otsA* y *otsB*) mostró que líneas transgénicas exhibieron un crecimiento vegetal sostenido, menos daño fotooxidativo y un equilibrio mineral más favorable en condiciones de estrés por sal, sequía y baja temperatura. Además, indicaron que el efecto principal de la trehalosa no es como un soluto compatible, sino que el aumento de la acumulación de trehalosa se correlaciona con niveles más altos de carbohidratos solubles y una capacidad elevada para la fotosíntesis tanto en condiciones de estrés como sin estrés, de acuerdo con su papel sugerido en la modulación de la

sensibilidad al azúcar y el metabolismo de los carbohidratos (Garg *et al.*, 2002). OsTPS8 regula la tolerancia al estrés por salinidad en el arroz mediante el control de las concentraciones de azúcares solubles y la regulación de la expresión de genes implicados en la señalización SAPK9 mediada por ABA (Vishal *et al.*, 2019).

En un estudio en jitomate, generaron plantas transgénicas mediante la introducción de un gen que codifica una fusión bifuncional de los genes trehalosa-6-fosfato sintasa y trehalosa-6-fosfato fosfatasa de *Escherichia coli* bajo el control del promotor CaMV35S, estas plantas mostraron una mayor tolerancia a la sequía y la salinidad que las plantas de tipo silvestre y tenían patrones de crecimiento y apariencia normales (Lyu *et al.*, 2013). Otra investigación demostró que la expresión de *ScTPS1* también en jitomate, empleando el promotor CaMV35S, mejoró la tolerancia a sequía, salinidad y estrés oxidativo (Cortina y Culiañez-Macia, 2005).

La transcripción del gen *NtTPPL* de tabaco fue inducida significativamente por el estrés por calor y solo ligeramente por NaCl, PEG y tratamientos de baja temperatura (Wang *et al.*, 2005). La sobreexpresión de un gen que codifica una trehalosa-6-fosfato fosfatasa (TPP) de arroz en mazorcas de maíz en desarrollo utilizando un promotor floral redujo la concentración de trehalosa-6-fosfato (T6P) y aumentó la concentración de sacarosa en las espiguillas, además incrementó la formación de granos y el índice de cosecha en plantas bien irrigadas o en condiciones de sequía (Nuccio *et al.*, 2015). Por lo tanto, la modulación del metabolismo de la trehalosa endógena es una estrategia prometedora para mejorar la tolerancia al estrés en las plantas de cultivo.

4.2 Métodos bioinformáticos utilizados

Un estudio "*in silico*" se define como un análisis que se realiza utilizando computadoras o mediante simulaciones por computadora. Puede involucrar cualquier tipo de estrategia en el manejo, extracción, integración e interpretación de datos moleculares a nivel genómico, metabolómico, proteómico, filogenético, celular o de organismo completo. Los enfoques *in silico* en plantas no solo reducen la necesidad de costosos trabajos de laboratorio y pruebas de campo,

sino que también aceleran la tasa de descubrimiento genético. Una ventaja única del enfoque *in silico* es su accesibilidad en todo el mundo y, en algunos casos, basta con tener acceso a Internet y una computadora (Rehman *et al.*, 2017).

Mediante el uso de herramientas *in silico*, se pueden identificar varios elementos relacionados con el estrés que responden significativamente a los cambios ambientales y este conocimiento podría aprovecharse posteriormente para desarrollar plantas de cultivo mejoradas en términos de calidad y productividad, con un mayor nivel de tolerancia al estrés abiótico (Singh *et al.*, 2011).

4.2.1. Herramientas para la identificación de genes en genoma completo

Phytozome es un centro comparativo para el genoma vegetal, incluidas las plantas terrestres y las algas. Se pueden recuperar genes, familias y secuencias de genes, secuencias genómicas y secuencias promotoras. Las proteínas homólogas y la ascendencia genética también están disponibles para determinar la evolución de las familias de genes. Además, están disponibles la anotación funcional de genes y la presentación de la estructura de genes. Los genes han sido anotados por asignaciones de PFAM, KOG, KEGG, PANTHER y GO (Goodstein *et al.*, 2012).

BLAST proteína-proteína estándar (blastp) ayuda a identificar una secuencia de aminoácidos de consulta y a encontrar secuencias similares en bases de datos de proteínas al encontrar regiones locales de similitud. Cuando el programa detecta similitud de secuencia que abarca toda la secuencia, también informa un alineamiento global que es preferible en el caso de la identificación (Altschul *et al.*, 1990).

4.2.2. Árboles evolutivos

Los árboles evolutivos son gráficos bidimensionales que muestran la relación evolutiva entre un conjunto de elementos. Los elementos pueden ser organismos, genes o secuencias. Cada unidad está definida por una rama distinta. Para la construcción de árbol filogenético, primero se eligen los marcadores moleculares (ADN o secuencia de proteínas); en segundo lugar, se realiza la alineación de secuencia múltiple; luego se elige un modelo de evolución (Modelo Jukes-Cantor, Modelo Kimura, etc.); se determina un método de construcción de árboles

(fenético o cladístico); y finalmente se evalúa la confiabilidad del árbol (Bootstrapping, Jackknifing, etc.) (Rehman *et al.*, 2017).

ClustalW es un algoritmo de alineación, utiliza una interfaz de línea de comandos. Alinea primero las dos secuencias más estrechamente relacionadas. Clustal primero realiza todas las posibles alineaciones por pares entre cada par de secuencias y calcula la "distancia" entre cada par de secuencias basándose en estos alineamientos por pares aislados. Finalmente, Clustal genera una matriz de distancias. Las ventajas son la velocidad; y las desventajas son que no se mide ninguna función objetivo, no hay forma de cuantificar si la alineación es buena o no. Clustal es uno de los programas de alineación más utilizados. Si se introduce un error al principio del proceso de alineación, es imposible corregirlo más tarde. Clustal toma las secuencias como vienen, sin embargo, no es útil si todas las secuencias no pueden tratarse de la misma manera (Rehman *et al.*, 2017).

Los métodos cladísticos utilizan cada posición en la alineación como información evolutiva para la construcción de un árbol y el árbol resultante se conoce como cladograma. Aquí los árboles se generan considerando diferentes posibles vías de evolución. Los métodos cladísticos también se denominan métodos basados en caracteres (o métodos discretos) que se basan directamente en los caracteres de secuencia en lugar de en distancias por parejas. Cuentan los eventos mutacionales acumulados en las secuencias y, por lo tanto, puede evitar la pérdida de información cuando los caracteres se convierten a distancias. Los dos enfoques más populares son el método de máxima parsimonia (MP) y máxima verosimilitud (ML) (Rehman *et al.*, 2017),

MEGA (Análisis de Genética Evolutiva Molecular) es una herramienta integrada para realizar alineación automática y manual de secuencias, extraer bases de datos basadas en la web, inferir árboles filogenéticos, estimar tasas de evolución molecular y probar hipótesis evolutivas (Nicholas *et al.*, 1997).

4.2.2.1. Construcción de un árbol de máxima verosimilitud (ML)

El método de máxima verosimilitud busca los mejores valores de árbol y parámetros maximizando la probabilidad. Este método ofrece más flexibilidad, se puede considerar cualquier modelo de evolución de secuencia. Por lo tanto, para

cualquier modelo de evolución, se calcula la probabilidad de un árbol dado y se pueden usar procedimientos de optimización para inferir tanto la topología del árbol como los parámetros del modelo más probables. Es el más riguroso entre todos los enfoques, ya que ML usa la información de secuencia completa, no solo los sitios informativos (Baxevanis y Ouellette, 2001).

Bootstrapping es un método de evaluación de árbol de remuestreo que funciona con distancia, parsimonia, probabilidad, y casi cualquier otro método de derivación de árboles. Puede considerarse un proceso de dos pasos que comprende la generación de (muchos) nuevos conjuntos de datos del conjunto original y el cálculo de un número que da la proporción de veces que una rama en particular (por ejemplo, un taxón) apareció en el árbol. Ese número se conoce comúnmente como el valor de arranque (Baxevanis y Ouellette, 2001).

4.2.3. Visualización de alineación de secuencia múltiple

GeneDoc proporciona herramientas para visualizar, editar y analizar múltiples alineaciones de secuencias de proteínas y ácidos nucleicos, en un contexto explícitamente evolutivo, permite dividir las secuencias en grupos que reflejan la división de superfamilias de genes (y proteínas) en familias distintas. GeneDoc puede analizar y visualizar estos grupos por separado o juntos (Nicholas *et al.*, 1997).

4.2.4. Análisis de secuencias proteicas

4.2.4.1. Predicción de la localización subcelular de proteínas

YLoc es un servidor web interpretable, para predecir la localización subcelular, utiliza lenguaje natural para explicar por qué se hizo una predicción y qué propiedad biológica de la proteína fue la principal responsable de ella. Además, YLoc estima la confiabilidad de sus propias predicciones, por lo tanto, puede ayudar a comprender la localización de proteínas y a la ingeniería de ubicación de proteínas (Briesemeister *et al.*, 2010).

4.2.4.2. Predicción de péptidos señal y NES en proteínas

Las proteínas vegetales se translocan del citosol a orgánulos específicos mediante péptidos de tránsito N-terminal en el caso de cloroplastos y

mitocondrias, o señales de localización nuclear (NLS) en el caso de núcleos (Sperschneider *et al.*, 2017).

El servidor TargetP 1.1 predice la ubicación subcelular de proteínas eucariotas, comprobando la presencia de cualquiera de las presecuencias N-terminales: péptido de tránsito de cloroplasto (cTP), péptido de direccionamiento mitocondrial (mTP) o péptido señal de la vía secretora (SP). Si el servidor predice cualquier presecuencia N-terminal, también se puede predecir un sitio de escisión potencial. Las predicciones de TargetP se basan principalmente en residuos 130 N-terminal de cada secuencia de entrada. Por lo tanto, se recomienda tener el N-terminal para todas las secuencias a comprobar (Emanuelsson *et al.*, 2000).

ChloroP es un método basado en redes neuronales para identificar péptidos de tránsito de cloroplasto y sus sitios de escisión. Los sitios de escisión se predicen utilizando una matriz de puntuación derivada de un algoritmo automático de búsqueda de motivos (Emanuelsson *et al.*, 1999).

LOCALIZER aplica un enfoque de aprendizaje automático entrenado en proteínas vegetales con datos de localización verificados experimentalmente para predecir si una proteína vegetal o una proteína efectora se dirige a cloroplastos o mitocondrias, mediante el reconocimiento de péptidos de tránsito mitocondrial o cloroplasto y una búsqueda NLS para predecir la localización en el núcleo (Sperschneider *et al.*, 2017).

Las señales de exportación nuclear (NES) son reguladores extremadamente importantes de la ubicación subcelular de las proteínas. Esta regulación tiene un impacto en la transcripción y otros procesos nucleares, que son fundamentales para la viabilidad de la célula. El predictor de NES (NetNES) emplea redes neuronales y modelos de Markov ocultos en el algoritmo de predicción (la Cour *et al.*, 2004).

4.2.4.3. Estructura genética

La visualización de la estructura de los genes y las características anotadas ayuda a los biólogos a investigar su función y evolución de forma intuitiva. Gene Structure Display Server (GSDS) está diseñado para la visualización de características génicas, como la composición y posición de exones, intrones y

elementos conservados. La entrada podría ser secuencias, número de acceso de GenBank (o GI) o entidades en formatos BED/GTF/GFF3. Los usuarios pueden personalizar la forma y el color de la imagen. Se puede cargar un árbol filogenético y añadirlo en la figura (Hu *et al.*, 2015).

4.2.4.4. Dominios conservados

El CDD del NCBI, la Base de Datos de Dominios Conservados (CDD), es un recurso público para la anotación de proteínas con la localización de huellas de dominios conservadas. Se pueden hacer búsquedas en vivo de una sola proteína query y de grandes conjuntos de secuencias de proteínas queries. CDD contiene modelos de dominio y definiciones de dominio proporcionadas por varias colecciones disponibles en el dominio público, como Pfam, SMART, la colección de COGs, KOGs, TIGR, PRK, la colección de NCBI protein clusters, así como las producidas por un esfuerzo interno de curación de datos de NCBI (Marchler-Bauer *et al.*, 2015).

4.2.4.5. Motivos

Los motivos son secuencias de ADN o aminoácidos que se repiten y tienen funciones biológicas. Muestran sitios de unión específicos de secuencia para factores de transcripción (TF) y nucleasas y están asociados con el sitio de unión ribosómico, el procesamiento de ARNm y los procesos de terminación de la transcripción y la regulación de la expresión génica (Rehman *et al.*, 2017).

Multiple Expectation Maximizations for Motif Elicitation (MEME) es una herramienta para descubrir motivos en un grupo de secuencias de proteínas o ADN relacionadas, utiliza técnicas de modelado estadístico para elegir automáticamente el mejor ancho, número de ocurrencias y descripción para cada motivo. MEME toma como entrada un grupo de secuencias de ADN o proteínas y genera tantos motivos como se soliciten hasta un umbral de confianza estadística especificado por el usuario, representa motivos como matrices de probabilidad de letras dependientes de la posición que describen la probabilidad de cada letra posible en cada posición del patrón, para lo cual. Los motivos individuales de MEME no contienen espacios, por lo que divide los patrones con

espacios de longitud variable en dos o más motivos separados (Bailey y Gribskov, 1998).

4.2.5. Análisis de estructuras proteicas

4.2.5.1. Análisis de estructura primaria de proteínas

ProtParam es una herramienta que permite el cálculo de varios parámetros físicos y químicos para una proteína determinada almacenada en Swiss-Prot o TrEMBL o para una secuencia de proteínas ingresada por el usuario. Los parámetros calculados incluyen el peso molecular, el punto isoeléctrico, pI teórico, la composición de aminoácidos, la composición atómica, el coeficiente de extinción, la vida media estimada, el índice de inestabilidad, el índice alifático y el índice de hidropaticidad (GRAVY) (Gasteiger *et al.*, 2003).

El punto isoeléctrico (pI) es el pH en el que un aminoácido no se mueve en un campo eléctrico. Esto significa que su carga neta se vuelve cero (zwitterión). El pH del ambiente acuoso circundante determina la carga neta del aminoácido. $pI < 7$ indica que la proteína es ácida y $pI > 7$ indica que la proteína es básica. pI de 7 indica neutral. El punto isoeléctrico teórico (pI) es útil en el desarrollo de sistemas tampón para la purificación de proteínas (Kalibulla *et al.*, 2017)

GRAVY es la suma de los valores hidrofóbicos de todos los aminoácidos. El aumento de la puntuación positiva indica una mayor hidrofobicidad. Las proteínas con índice GRAVY muy bajo muestran una mejor interacción con el agua (Kalibulla *et al.*, 2017).

4.2.5.2. Análisis de estructura secundaria de proteínas

SOPMA (Método de predicción autooptimizado con alineación) es una mejora del método SOPM. Estos métodos se basan en el método homólogo. La mejora tiene lugar en el hecho de que SOPMA toma en cuenta información de un alineamiento de secuencias pertenecientes a la misma familia. Si no hay secuencias homólogas, la predicción de SOPMA es la de SOPM (Geourjon y Deléage, 1995).

4.2.5.3. Predicción de la estructura terciaria de proteínas mediante modelos de homología

La comprensión completa de las funciones biológicas de las proteínas requiere el conocimiento de sus estructuras. Esta información puede ser útil para el diseño racional de experimentos bioquímicos como mutagénesis dirigida al sitio, estabilidad de proteínas o análisis funcional. En resumen, el estudio de modelado ayuda a mejorar nuestra comprensión de las funciones de las proteínas. El método basado en el conocimiento predice las estructuras de las proteínas basándose en la información de la estructura de las proteínas que existe en las bases de datos. Es de dos tipos: el modelado de homología que construye modelos basados en una estructura determinada experimentalmente que está estrechamente relacionada a nivel de secuencia y el subproceso que identifica proteínas que son estructuralmente similares, con o sin similitudes de secuencia detectables (Kalibulla *et al.*, 2017).

El modelado de homología, también conocido como modelado comparativo, predice estructuras de proteínas de secuencia definida por el usuario (objetivo) basándose en la homología de secuencia con estructuras conocidas (plantillas). Se espera que, si dos proteínas comparten una similitud de secuencia lo suficientemente alta, es probable que tengan estructuras tridimensionales muy similares. Si una de las secuencias de proteínas tiene una estructura conocida, entonces la información estructural se puede copiar a la proteína desconocida con un alto grado de confianza. Como regla general, la proteína molde debe tener como mínimo un 30% de identidad de secuencia con la secuencia de consulta. Cuando se encuentran múltiples estructuras de bases de datos con similitud significativa como resultado de la búsqueda, se recomienda utilizar estructuras con el porcentaje más alto (secuencia) de identidad y la resolución más alta como plantilla (Kalibulla *et al.*, 2017).

SWISS-MODEL es una base de datos de modelos de homología generados automáticamente para organismos modelo relevantes e información de estructura experimental para todas las secuencias en UniProtKB. Para facilitar la exploración de la información disponible sobre una proteína diana determinada,

SWISS-MODEL proporciona enlaces cruzados a otros recursos y bases de datos (Waterhouse *et al.*, 2018).

4.2.6. Análisis de duplicación genética

Las duplicaciones genéticas se consideran de acuerdo con los siguientes criterios: > 70% de identidad de secuencia de proteínas y > 75% de identidad de longitud de secuencia y los resultados del análisis filogenético (Waterhouse *et al.*, 2018).

La duplicación en tándem y segmental ocurre cuando dos genes de la misma especie cercanamente relacionados se localizan en el mismo cromosoma o en diferentes cromosomas, respectivamente (Zulfiqar *et al.*, 2019).

Los genes homólogos son dos o más genes que derivan de un ancestro común. Los genes ortólogos y parálogos son diferentes tipos de genes homólogos. Los genes ortólogos son genes homólogos que divergieron después de un evento de especiación. Los genes parálogos son genes homólogos que han divergido dentro de una especie y surgen durante la duplicación de genes (Jensen, 2001).

4.2.7. Análisis de elementos reguladores *cis*

En eucariotas la regulación transcripcional de la expresión génica está controlada principalmente por la unión de factores de transcripción a elementos reguladores *cis* en la región promotora de los genes (Priest *et al.*, 2009). Los elementos *cis* son secuencias cortas de ADN no codificante. El patrón de ocurrencia de estos elementos *cis* descifra el mecanismo de regulación sensible al estrés de los genes aguas abajo o *downstream* (Sheshadri *et al.*, 2016).

PlantCARE es una base de datos de elementos reguladores, potenciadores y represores que actúan en *cis* en plantas. Los elementos regulatorios están representados por matrices posicionales, secuencias consenso y sitios individuales en secuencias promotoras particulares. Los datos sobre los sitios de transcripción se extraen principalmente de la literatura, complementados con los datos predichos *in silico*. Se muestra una descripción general para sitios específicos del factor de la transcripción, los niveles de confianza para la evidencia experimental, la información funcional y la posición en el promotor (Lescot *et al.*, 2002).

4.2.8. Análisis de expresión génica

GEO (Gene Expression Omnibus) (<http://www.ncbi.nlm.nih.gov/geo/>) es un repositorio público de datos de genómica funcional que admite envíos de datos compatibles con MIAME (información mínima sobre un experimento de microarreglos). Se pueden consultar y descargar experimentos y perfiles de expresión genética seleccionados.

Muchos de los métodos para visualizar e interpretar datos de expresión génica se pueden utilizar para experimentos de microarreglos y RNA-seq. A continuación, se describen dos métodos:

4.2.7.1. Heat map

Un heat map o mapa de calor es un método común para visualizar datos de expresión génica, también se puede combinar con métodos de agrupación que agrupan genes y / o muestras en función de la similitud de su patrón de expresión génica. Esto puede ser útil para identificar genes que están comúnmente regulados o firmas biológicas asociadas con una condición particular (por ejemplo, una condición ambiental). En los mapas de calor, los datos se muestran en una cuadrícula donde cada fila representa un gen y cada columna representa una muestra. El color y la intensidad de los cuadros se utilizan para representar cambios relativos (no valores absolutos) de la expresión génica (EMBL-EBI, 2021).

4.2.7.1. Volcano Plot

Los diagramas o gráficas de volcán se utilizan para mostrar los resultados de RNA-seq u otros experimentos ómicos. Un diagrama de volcán es un tipo de gráfica de dispersión que muestra significancia estadística (valor P) frente a la magnitud del cambio (Fold Change). Permite una identificación visual rápida de genes con grandes cambios que también son estadísticamente significativos. Estos pueden ser los genes biológicamente más significativos. En una gráfica de volcán, los genes más regulados al alza están hacia la derecha, los genes más regulados a la baja están hacia la izquierda y los genes más estadísticamente significativos estadísticamente están hacia la parte superior. Para generar una

gráfica de volcán de resultados de ARN-seq, se requiere un archivo de resultados expresados diferencialmente (Doyle, 2021).

V. MATERIALES Y MÉTODOS

5.1. Identificación de miembros de las familias TPS y TPP en jitomate

Se realizaron búsquedas BLAST en el genoma completo de *Solanum lycopersicum* iTAG2.4 de la base de datos Phytozome 12.0 (<https://phytozome.jgi.doe.gov/pz/portal.html>) usando las secuencias de aminoácidos de las proteínas TPS y TPP de *Arabidopsis thaliana* como plantillas para encontrar todas las secuencias disponibles de los genes TPS y TPP de jitomate, y los resultados se filtraron por valor E ($1e^{-5}$). Todas las secuencias de proteínas obtenidas fueron examinadas para detectar la presencia de dominios conservados utilizando la herramienta Motif Search (<https://www.genome.jp/tools/motif/>) y la herramienta analítica de dominios conservados en línea (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Las secuencias que no presentaron dominios no se tomaron en cuenta para análisis posteriores.

Los genes se clasificaron como *TPS* si contenían un dominio Glycosyl transferase family 20 y como *TPP* si sólo tenían el dominio fosfatasa. Sobre la base de los resultados de búsqueda BLAST, los diez genes TPS fueron nombrados *SITPS-1* a *SITPS-10* y los nueve genes TPP como *SITPP-1* a *SITPP-9*.

5.2. Análisis de dominios de proteínas TPS/TPP en jitomate y otras especies

Los dominios de genes de biosíntesis de trehalosa de *Escherichia coli*, *Saccharomyces cerevisiae*, *S. lepidophylla*, *A. thaliana* y *S. lycopersicum* se analizaron utilizando la herramienta de búsqueda de dominios conservados (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) y se visualizaron mediante el software TBtools. Se mostraron junto con un árbol filogenético de las secuencias de proteínas de longitud completa de dichas especies.

5.3. Análisis filogenético de las familias TPS y TPP en jitomate

Las secuencias proteicas de TPS de *Arabidopsis thaliana* (Leyman *et al.*, 2001), *Oryza sativa* (Zang *et al.*, 2011), *Selaginella lepidophylla* (Zentella *et al.*, 1999) se descargaron de NCBI (<http://www.ncbi.nlm.nih.gov/>), mientras que las secuencias proteicas de *Zea mays* (Zhou *et al.*, 2013) se obtuvieron de *Zea mays* Ensembl-18, *Solanum tuberosum* (Xu *et al.*, 2017) de *Solanum tuberosum* v4.03 en la base de datos Phytozome 12.0 (Apéndice 1). Las secuencias proteicas de TPP de *Arabidopsis thaliana* (Leyman *et al.*, 2001) de *Arabidopsis thaliana* TAIR10, *Brachypodium distachyon* (Wang *et al.*, 2019) de *Brachypodium distachyon* v3.1, *Oryza sativa* (Rahman *et al.*, 2021) de *Oryza sativa* v7_JGI, *Populus trichocarpa* (Rahman *et al.*, 2021) de *Populus trichocarpa* v3.1, *Zea mays* (Rahman *et al.*, 2021) de *Zea mays* Ensembl-18 de la base de datos Phytozome 12.0 (Apéndice 2).

Para clasificar las familias TPS y TPP en *S. lycopersicum* se construyeron dos árboles filogenéticos de cada familia. En el primero se utilizaron solamente las secuencias proteicas TPS/TPP de jitomate y en el segundo también se incluyeron secuencias TPS/TPP de otras especies. Para la construcción del árbol filogenético de SITPS y SITPP, se alinearon las secuencias proteicas de SITPS y SITPP mediante Clustal W en el programa MEGA 7.0, y las secuencias alineadas se utilizaron para construir un árbol filogenético con el método de máxima probabilidad (Kumar *et al.*, 2016), con métodos de corrección de Poisson y 500 réplicas de arranque. Además, se visualizó el alineamiento de las SITPSs y SITPPs en el programa GeneDoc (<http://www.psc.edu/biomed/genedoc>) (Nicholas *et al.*, 1997).

En lo que se refiere al árbol filogenético de la familia TPS de diferentes especies, se alinearon las múltiples secuencias de aminoácidos de *Solanum lycopersicum*, *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, *Solanum tuberosum* y *Selaginella lepidophylla*, y en el caso de la familia TPP se alinearon las secuencias de *Solanum lycopersicum*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Oryza sativa*, *Populus trichocarpa* y *Zea mays*, mediante Clustal W en el programa MEGA 7.0, y las secuencias alineadas se utilizaron para construir un árbol

filogenético con el método de máxima probabilidad (Kumar *et al*, 2016), con métodos de corrección de Poisson y 1000 réplicas de arranque.

5.4. Análisis de características genéticas y localización subcelular de proteínas de las familias TPS y TPP en jitomate

La información de los genes TPS y TPP de jitomate, incluidas las ubicaciones de los cromosomas, secuencias genómicas, secuencias completas de CDS, secuencias de proteínas y 1500 bp de las secuencias de nucleótidos aguas arriba (*upstream*) del codón de inicio de traducción se descargaron del sitio web Phytozome. El peso molecular (kDa), el punto isoeléctrico (pI) y el índice de hidropaticidad (GRAVY) de cada gen se calcularon utilizando la herramienta ExPASy (<https://web.expasy.org/protparam/>) (Gasteiger *et al*, 2003).

La localización subcelular de proteínas de las familias TPS y TPP se predijo utilizando YLoc Interpretable Subcellular Localization Prediction (<https://abi-services.informatik.uni-tuebingen.de/yloc/webloc.cgi>) (Briesemeister *et al*, 2010). Para la predicción de péptidos de tránsito se utilizó: TargetP 1.1 (<http://www.cbs.dtu.dk/services/TargetP-1.1/index.php>) (Emanuelsson *et al*, 2000); ChloroP 1.1 Server (<http://www.cbs.dtu.dk/services/ChloroP/>) (Emanuelsson *et al*, 1999) para la predicción de péptidos de tránsito al cloroplasto; LOCALIZER (<http://localizer.csiro.au/>) (Sperschneider *et al*, 2017) para la predicción de péptidos de tránsito al cloroplasto/mitocondria. Además, para la predicción de señales de exportación nuclear ricas en leucinas se utilizó NetNES 1.1 Server (<http://www.cbs.dtu.dk/services/NetNES/>) (la Cour *et al*, 2004) y para la predicción de las NLSs (señales de localización nuclear) en las proteínas se empleó LOCALIZER.

5.5. Localización en cromosoma y análisis evolutivo de las familias TPS/TPP en jitomate

El software TBtools (<http://www.tbtools.com/>) (Chen *et al*, 2018) se utilizó para generar el diagrama de localización de los genes SITPS y SITPP en cromosoma, de acuerdo con la información de posición en el cromosoma proporcionada en la base de datos de Phytozome. Las duplicaciones genéticas entre los genes *TPS/TPP* del jitomate podrían proveer una referencia significativa para el análisis

de evolución de los genes, por lo que se identificaron de acuerdo con los criterios de Wang *et al*, (2019).

Los valores de sustitución sinónima (K_s) y sustitución no sinónima (K_a) se calcularon usando el software TBtools. La relación K_a/K_s se ha calculado dividiendo los K_a entre valores K_s . El tiempo de divergencia (T) se calculó de acuerdo con la formula $T = K_s/2\lambda$ donde $\lambda = 6.56 \times 10^{-9}$ sustituciones por sitio por año para jitomate (YanJun *et al*, 2016).

La relación microsinténica de los genes *TPS* y *TPP* entre jitomate y arábidoopsis fue investigada utilizando un enfoque de búsqueda BLAST recíproco contra los genomas de estas especies. Los resultados se visualizaron utilizando el software TBtools. Las secuencias para las dos especies de plantas utilizadas para el análisis de microanálisis fueron recuperadas de Phytozome.

5.6. Distribución exón-intrón de genes de las familias TPS/TPP en jitomate

El sitio web GSDS 2.0 (Gene Structure Display Server) (<https://gsds.cbi.pku.edu.cn/>) (Hu *et al*, 2015) se utilizó para analizar las estructuras genéticas de las familias *TPS/TPP* del jitomate y trazar el diagrama de distribución exón-intrón.

5.7. Análisis de motivos proteicos y dominios conservados de las familias TPS/TPP en jitomate

Los dominios de los genes *SITPSs/SITPPs* se analizaron utilizando la herramienta de búsqueda de dominios conservados (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) y se visualizaron mediante el software TBtools.

El análisis de motivos en las secuencias proteicas de *SITPS* y *SITPP* se llevó a cabo utilizando la herramienta en línea MEME (<http://meme-suite.org/>) (Bailey y Gribskov, 1998) con un número máximo de 10 motivos, y demás parámetros predeterminados y se visualizaron mediante el software TBtools.

5.8. Análisis de la estructura 3D de las proteínas TPS/TPP en jitomate

El análisis de la estructura 3-D de SITPS/SITPPs se realizó mediante la herramienta en línea SWISS-MODEL (<https://swissmodel.expasy.org/interactive#sequence>) y SOPMA secondary structure prediction method (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) (Waterhouse *et al*, 2018; Combet *et al*, 2000; Geourjon y Deléage, 1995).

5.9. Análisis de elementos reguladores cis en promotores de genes TPS/TPP en jitomate

Se descargaron secuencias de 1500 bp aguas arriba del codón de inicio de la traducción de genes *SITPS/SITPP* de la base de datos Phytozome y se predijeron los elementos regulatorios *cis* mediante PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot *et al*, 2002) y se diagramó utilizando el software TBtools para investigar los elementos que actúan en *cis* en las regiones promotoras y estudiar el papel que juegan dichos genes en el estrés y otros estímulos.

5.10. Análisis de expresión *in silico* de genes TPS/TPP en jitomate

Para identificar el patrón de expresión de los genes *SITPS* y *SITPP* en diferentes tratamientos de estrés abiótico, se analizaron conjuntos de datos de expresión de Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>). Los datos de expresión normalizados se utilizaron para dibujar el mapa de calor utilizando el software TBTools.

En cuanto a la respuesta al estrés por sequía, se analizaron los perfiles de transcripción de plántulas de jitomate (*Solanum lycopersicum*) cv. M82, un cultivar sensible a la sequía, cuyos datos están disponibles con el número de accesión GSE151277 (Wang *et al*, 2020). Los tratamientos fueron D-0d = control, D-1d = estrés por sequía durante 1 día, D-2d = sequía durante 2 días, D-3d = sequía durante 3 días, D-4d = sequía durante 4 días, D-5d = sequía durante 5 días, D-Rec = sequía con recuperación. Así mismo, se analizaron los perfiles de ARNm de plantas 'Micro-TOM' bajo diferentes tratamientos: CK-0h = Micro-TOM control, CK-3h = Micro-TOM estrés por sequía durante 3 horas y CK-6h = Micro-

TOM estrés por sequía durante 6 horas, cuyo número de accesión es GSE148530 (Zhao *et al*, 2020).

Con respecto a la respuesta al estrés por calor, se analizaron los perfiles de transcripción de plántulas de jitomate (*Solanum lycopersicum*) cv. M82, un cultivar sensible al calor, disponible con el número de accesión GSE151277 (Wang *et al*, 2020). Los tratamientos fueron Heat-0h = control, Heat-2h = estrés por calor durante 2h, Heat-4h = calor durante 4h, Heat-12h = calor durante 12h, Heat-24h = calor durante 24h, Heat-Rec = calor con recuperación. Del mismo modo se analizó el transcriptoma de frutos de jitomate 'Micro-TOM' sometidos a estrés por calor, con número de accesión GSE174607 (Liu *et al*, 2021), en los siguientes tratamientos: MT-0h, MT-24h, MT-48 h y MT-96 h.

En relación con la respuesta al estrés por frío, se analizaron los perfiles de ARNm de hojas de plantas de jitomate 'Micro-TOM', disponible con el número de accesión GSE148887 (Zhao y Xu, 2020). Los tratamientos fueron CK-0h = control, CK-4h = estrés por frío durante 4h y CK-12h = frío durante 12h.

En lo que toca a la respuesta al estrés por salinidad, se analizaron los perfiles de ARNm de hojas de jitomate, disponible con el número de accesión GSE 148353 (Zhao y Xu, 2020). Los tratamientos fueron CK-0h = control, CK-2h = estrés por salinidad durante 2h y CK-8h = salinidad durante 8h.

VI. RESULTADOS Y DISCUSIÓN

6.1. Identificación de los miembros de las familias TPS y TPP en jitomate

La identificación en todo el genoma y caracterización de los genes de la familia TPS se ha realizado en varias plantas superiores, como *Arabidopsis thaliana* (Leyman *et al*, 2001), *Brachypodium distachyon* (Wang *et al*, 2019), *Capsicum annum* (Wei *et al*, 2016), *Cucumis melo* (Yang *et al*, 2017), *Cucumis sativus* (Dan *et al*, 2021), *Glycine max* (Xie *et al*, 2014), *Gossypium* (Mu *et al*, 2016), *Malus x domestica* (Du *et al*, 2017), *Medicago truncata* (Song *et al*, 2021), *Moringa oleifera* (Lin *et al*, 2018), *Nelumbo nucifera* (Jin *et al*, 2016), *Oryza sativa* (Zang *et al*, 2011), *Phaseolus vulgaris* (Barraza *et al*, 2016), *Populus trichocarpa* (Yang *et al*, 2012), *Prunus mume* (Yang *et al*, 2020), *Prunus persica* (Du *et al*, 2017), *Saccharum spp. hybrids* (Hu *et al*, 2020), *Selaginella lepidophylla* (Zentella *et al*, 1999), *Solanum tuberosum* (Xu *et al*, 2017), *Triticum aestivum* (Xie *et al*, 2015), *Vitis vinifera* (Morabito *et al*, 2021) y *Zea mays* (Zhou *et al*, 2013). En cambio, en la familia TPP se han realizado estudios en pocas plantas como: *Arabidopsis thaliana* (Yang *et al*, 2012), *Brachypodium distachyon* (Wang *et al*, 2019), *Oryza sativa* (Rahman *et al*, 2021), *Selaginella* (Pampurova *et al*, 2014) y *Zea mays* (Zhou *et al*, 2013).

Debido a la importancia económica del jitomate también se ha analizado su genoma en busca de genes *TPS* y *TPP*. Chen y Lou (2017) identificaron diez genes de la familia TPS de esta especie y estudiaron su expresión en plantas expuestas a diferentes espectros de luz. Zhang *et al*, (2016) identificaron 11 miembros de la familia TPS y 8 de TPP en el jitomate y evaluaron su participación en la resistencia a enfermedades. Sin embargo, no se ha realizado un análisis enfocado a la respuesta de las familias génicas TPS y TPP en respuesta al estrés abiótico en jitomate, así como su expresión por estímulos de fitohormonas.

En esta investigación identificamos y caracterizamos genes relacionados con la biosíntesis de la trehalosa en el genoma de jitomate (*S. lycopersicum* L.) mediante herramientas bioinformáticas y encontramos diez genes *SITPS* y

nueve *SITPP*, que denominamos *SITPS-1* a *SITPS-10* y *SITPP-1* a *SITPP-9*, respectivamente, con el prefijo *Sl* (*Solanum lycopersicum*) seguido con el nombre de la familia y la posición en el cromosoma de menor a mayor, como se observa en el Cuadro 1.

Las secuencias codificantes (CDS) de los genes *SITPSs* / *SITPPs* oscilaron entre 2556 pb (*SITPS-3*) a 2832 pb (*SITPS-2*) y de 345 pb (*SITPP-9*) a 1161 pb (*SITPP-8*), la longitud de la proteína varió de 851 aa (*SITPS-3* y *SITPS-5*) a 943 aa (*SITPS-2*) y de 114 aa (*SITPP-9*) a 386 aa (*SITPP-8*), el peso molecular varió de 96.012 KDa (*SITPS-5*) a 106.176 KDa (*SITPS-2*) y de 12.939 KDa (*SITPP-9*) a 43.63 KDa (*SITPP-8*), respectivamente (Cuadro 1). De acuerdo con el valor *pI* predicho, todos los genes *SITPSs* eran ácidos, mientras que cinco genes *TPP* (*SITPP-1*, 2, 5, 7 y 9) eran ácidos (<7) y cuatro genes (*SITPP-3*, 4, 6 y 8) eran básicos (> 7) (Cuadro 1). El GRAVY de los *SITPSs* varió de -0.374 a -0.163 y de los *SITPPs* de -0.482 a -0.231. Todos los *SITPS* y *SITPP* mostraron resultados valores negativos de GRAVY (Cuadro 1). El índice de hidropaticidad (GRAVY) brinda información sobre el carácter hidrofóbico o hidrofílico de una proteína con base a la composición de aminoácidos que presenta (Kyte y Doolittle, 1982). Un valor negativo del GRAVY indica que la proteína es hidrofílica y un valor positivo indica que la proteína es hidrófoba. Todas las proteínas tanto *SITPS* como *SITPP* eran hidrofílicas.

6.2. Análisis de dominios de proteínas TPS/TPP en jitomate y otras especies

Se compararon los dominios conservados de las proteínas TPS/TPP encontradas en jitomate con los dominios de los genes implicados en la vía de biosíntesis de la trehalosa en otras especies ampliamente estudiadas como *Escherichia coli*, *Saccaromyces cerevisiae*, *S. lepidophylla* y *A. thaliana*, y se visualizaron los dominios junto con un árbol filogenético (Figura 1), en donde la proteína *SITPP-9* no pudo incluirse debido a su corta longitud de tan sólo 345bp.

En *E. coli* las enzimas TPS y TPP están separadas. Las proteínas TPS de *A. thaliana* están codificadas por 11 genes (*AtTPS1–AtTPS11*) que se dividen en proteínas de clase I y clase II. Las proteínas de clase I (*AtTPS1–AtTPS4*) con

dominios TPS y TPP fusionados son más complejas que ScTPS1, pues tienen un dominio fosfatasa que la levadura no tiene. AtTPS1 tiene dos dominios, Glyco_transf_20 y Trehalose_PPase. Al igual que TPS1 de Selaginella. En jitomate dos proteínas, SITPS-2 y SITPS-8 se agruparon en la clase I. Las proteínas AtTPS3 y AtTPS4 poseen tres dominios: Glyco_transf_20, Trehalose_PPase y Glycos_transf_1, mientras que AtTPS2 presenta, además de los tres dominios mencionados, un dominio Glyco_trans_1_4 incompleto en la extremo N-terminal.

Las proteínas de clase II (AtTPS5-AtTPS11) son en su mayoría similares a ScTPS2 y presentan un dominio adicional Hydrolase_3, en AtTPS5, AtTPS7 y AtTPS9 este dominio está incompleto en la extremo N-terminal.

En jitomate se encontraron ocho proteínas de clase II, cuatro de las cuales (SITPS-3, SITPS-5, SITPS-7 y SITPS-9) también tienen un dominio Hydrolase_3 incompleto. AtTPS7 se distingue por la presencia de un dominio extra: Glycos_trans_1. De igual modo SITPS-7 y SITPS-10. Las proteínas TPP en Arabidopsis consisten en diez isoformas (AtTPPA–AtTPPJ), que contienen un dominio Trehalose_PPase. AtTPPB es la única proteína que tiene dos dominios conservados, Trehalose_PPase e Hydrolase_3. En jitomate se identificaron nueve proteínas SITPP.

Cuadro 1. Información y características de los miembros de las familias TPS y TPP en *Solanum lycopersicum*.

Nombre Gen	ID Gen	PC	CDS (bp)	Localización en el cromosoma			Intron / Exon	Long. (AA)	MW (kDa)	pI	GRAVY	SLoc
				N°	St	Start End						
SITPS-1	Solyc01g005210.2	II-1	2574	1	+	165134 169351	2:3	857	97.268	5.61	-0.213	C, N
SITPS-2	Solyc02g071590.1	I-1	2832	2	-	40987454 41004753	17:18	943	106.176	5.91	-0.301	C, N
SITPS-3	Solyc02g072150.2	II-2	2556	2	-	41431871 41435950	2:3	851	96.204	5.58	-0.217	C
SITPS-4	Solyc04g025940.2	II-3	2574	4	-	20337858 20341062	2:3	857	96.732	5.59	-0.212	C
SITPS-5	Solyc05g005750.2	II-3	2556	5	+	574848 578018	2:3	851	96.012	6.2	-0.224	C
SITPS-6	Solyc07g006500.2	II-4	2631	7	+	1301353 1304574	2:3	876	98.645	6.39	-0.234	C
SITPS-7	Solyc07g055300.2	II-2	2577	7	+	63384523 63390604	2:3	858	97.038	5.59	-0.264	C
SITPS-8	Solyc07g062140.2	I-1	2781	7	-	64940052 64953466	16:17	926	104.513	6.23	-0.374	C, N
SITPS-9	Solyc08g076650.2	II-1	2589	8	+	60585195 60590640	2:3	862	97.23	5.79	-0.163	C
SITPS-10	Solyc10g007950.2	II-2	2574	10	-	2108916 2114253	2:3	857	97.458	5.63	-0.3	C, M
SITPP-1	Solyc03g007290.2	A2	1011	3	+	1859030 1861212	7:8	336	37.771	5.44	-0.322	C, N
SITPP-2	Solyc03g083960.2	A1	1104	3	-	53896143 53898914	9:10	367	41.022	6.41	-0.351	C, N, P
SITPP-3	Solyc04g054930.2	A4	954	4	-	53438494 53441329	8:9	317	36.335	9.14	-0.447	N, C
SITPP-4	Solyc04g072920.2	A4	1098	4	+	59909039 59911234	8:9	365	41.317	9.33	-0.464	N, C
SITPP-5	Solyc04g082550.1	A2	882	4	-	66187054 66189383	8:9	293	33.288	5.58	-0.231	C, N
SITPP-6	Solyc05g051880.2	A4	843	5	-	62283483 62288578	7:8	280	32.075	9.62	-0.482	N, C
SITPP-7	Solyc06g060600.1	A1	957	6	-	38600897 38602858	10:11	318	35.925	6.47	-0.356	C, N
SITPP-8	Solyc08g079060.2	A1	1161	8	+	62710550 62713243	10:11	386	43.63	7.05	-0.451	N, C
SITPP-9	Solyc10g046770.1	A2	345	10	+	36882331 36885697	1:2	114	12.939	4.88	-0.273	C, N, EE

PC Clado filogenético, bp pares de base, No número, St, Cadena, MW Peso Molecular, AA aminoácido, KDa Kilo Dalton, pI Punto Iso-eléctrico, GRAVY Promedio de hidropatía, SLoc Localización sub-celular, C Citoplasma, N Núcleo, P Peroxisoma, M Mitocondría, EE Espacio extracelular.

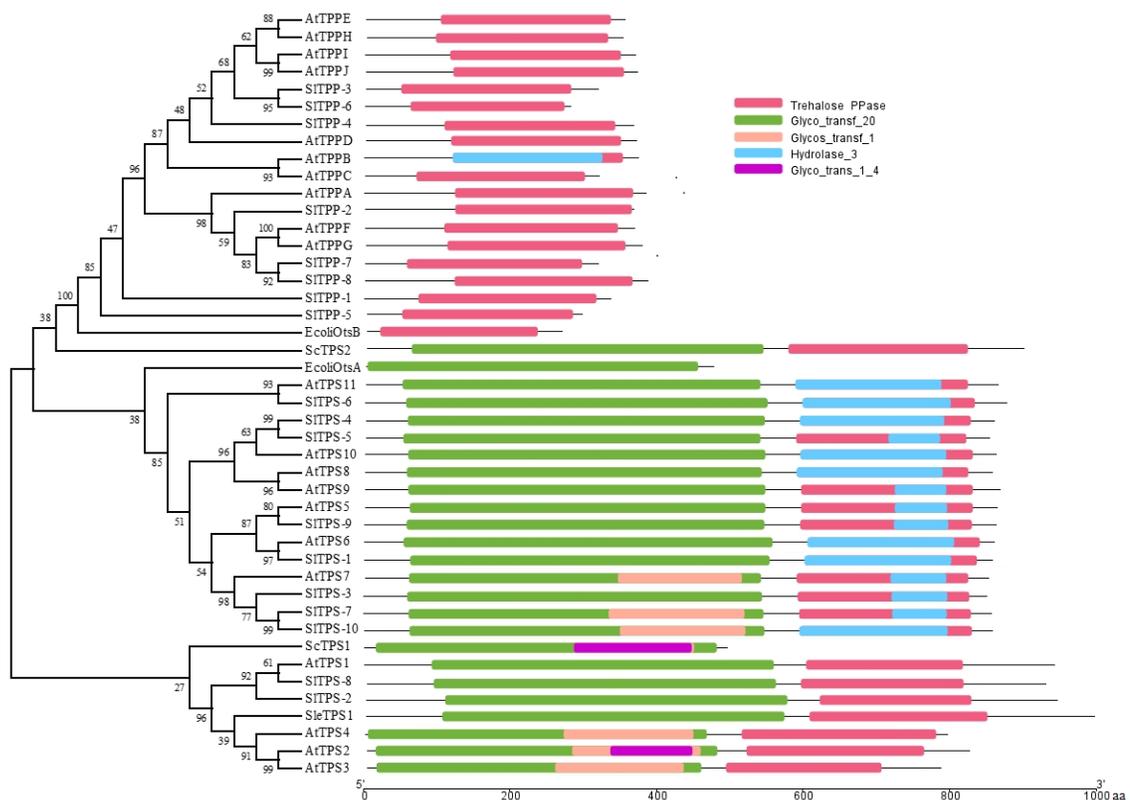


Figura 1. Distribución de los dominios TPS y TPP en diferentes genes de la biosíntesis de la trehalosa. *E. coli* OtsA y OtsB, *S. cerevisiae* (Sc) TPS1 y TPS2, *S. lepidophylla* (Sle) TPS1, *Arabidopsis thaliana* (At) Clase I: TPS1-5, Clase II: TPS5-11; Clase III: TPPA. El alineamiento es de secuencias proteicas.

6.3. Análisis filogenético de las familias TPS/TPP en jitomate

Estudios previos identificaron 1 gen *TPS* en el genoma de Selaginella (Zentella *et al*, 1999), 8 en papa (Xu *et al*, 2017), 11 en cada uno de los genomas de *Arabidopsis* y arroz (Yang *et al*, 2012) y 15 en maíz (Zhou *et al*, 2013).

Arabidopsis y arroz contienen cuatro (*AtTPS1*, *AtTPS2*, *AtTPS3* y *AtTPS4*) (*Leyman et al*, 2001) y uno (*OsTPS1*) genes *TPS* clase I, respectivamente, y los genes *TPS* restantes son genes *TPS* clase II. En algunos estudios han dividido a los genes clase I en dos subclados: B1 y B2 (Yang *et al*, 2012); I-1 y I-2 (Xie *et al*, 2015) y en otros corresponden a un sólo grupo: E (Xin *et al*, 2020); I-1 (Xu *et al*, 2017). Los genes de la clase II se han dividido en cuatro subclados: II-1, II-2, II-3, II-4 (Xie *et al*, 2015); A, B, C, D (Xin *et al*, 2020) y en cinco: II-1, II-2, II-3, II-4, II-5 (Xu *et al*, 2017); A1, A2, A3, A4, A5 (Yang *et al*, 2012). En este estudio, alineamos las 56 secuencias de proteínas de longitud completa de Selaginella,

papa, Arabidopsis, arroz, maíz y jitomate (Apéndice 1) y reconstruimos un árbol filogenético, el cual mostró que las 56 proteínas TPS se dividieron en dos clados distintos (clados I y II) con soporte de arranque 100% (Figura 2A). El clado I contenía sólo 10 miembros proteicos de clase I (17.86% del total de proteínas), mientras que el clado II contenía 46 miembros proteicos de clase II (82.14% del total de proteínas). Con el fin de describir las relaciones paralogas y ortologas entre esta familia, las diez proteínas TPS de clase I (Clado I) se dividieron aún más en dos subclados (subclados I-1 y I-2) con soporte de arranque 100%, lo que sugiere que las proteínas de cada subgrupo podrían compartir un origen similar. El subclado I-1 contenía siete proteínas TPS de clase I (AtTPS1, OsTPS1, SleTPS1, ZmTPS11.1, ZmTPS14.1, SITPS-2, SITPS-8) de Arabidopsis, arroz, Selaginella, maíz y jitomate, respectivamente, mientras que el subclado I-2 contenía sólo tres proteínas TPS de clase Arabidopsis (AtTPS2, AtTPS3 y AtTPS4).

Según la alineación de secuencias múltiples y el análisis filogenético, las diez SITPS se agruparon en dos grupos, similar a la división de TPS en Arabidopsis (Yang *et al*, 2012) y arroz (Zang *et al*, 2011). Nuestros resultados sugieren una pérdida de genes TPS en jitomate en comparación con la familia de genes *TPS* en arabidopsis (11), arroz (11), álamo (12), maíz (15), manzana (13), trigo (12) y soya (20). A diferencia de los cuatro genes clase I encontrados en el genoma de arabidopsis, en jitomate se han identificado sólo dos miembros, *SITPS-8* y *SITPS-2*, pertenecientes al grupo 1. Se tiene la hipótesis que los *SITPS* clase I experimentaron velocidades de duplicación más lentas que los *SITPS* clase II en el transcurso de la evolución del genoma, de ahí que haya pocos genes de clase I. *SITPS-8* y *SITPS-2* mostraron una alta identidad (79 y 74%) con AtTPS1 (Apéndice 5). *SITPS-8* está estrechamente relacionado filogenéticamente con AtTPS1 (Figura 1), que se demostró que es esencial para el crecimiento vegetativo (Van Dijck *et al*, 2002), la embriogénesis y la floración (Eastmond *et al*, 2006), y en el análisis microsinténico se encontró que *SITPS-8* y AtTPS-1 son ortólogos (Figura 5), por lo que es probable que las funciones de *SITPS-8* podrían

parecerse a las de AtTPS1. Sin embargo, sus funciones exactas deben investigarse más a fondo en jitomate.

Además, ocho de las proteínas de la familia TPS de jitomate estaban contenidas en pares con un gen *TPS* de papa, lo que sugiere la existencia de ortólogos en estas especies. En la mayoría de los grupos hay algún gen *SITPS*, excepto en I-2 (Figura 2A). Las proteínas SITPS se colocaron en cinco grupos: I-1 (SITPS-2, SITPS-8), II-1 (SITPS-1, SITPS-9), II-2 (SITPS-3, SITPS-7, SITPS-10), II-3 (SITPS-3, SITPS-4) y II-4 (SITPS-6).

Para clasificar la familia SITPP, construimos un árbol filogenético con 5 especies más (Figura 2B y Apéndice 2), alineamos un total de 60 secuencias proteicas de longitud completa de 6 especies: *Arabidopsis thaliana* (10), arroz (12), maíz (11), *Brachypodium distachyon* (10), *Populus trichocarpa* (8) y jitomate (9). El árbol filogenético de esta clase III, se dividió principalmente en dos clados, clado A y clado B. El clado A se dividió además en cuatro subclados, A1, A2, A3 y A4; y clado B en dos subclados, a saber, B1 y B2. El clado A contenía 39 miembros proteicos (65% del total de proteínas), mientras que el clado B contenía sólo 21 miembros proteicos (35% del total de proteínas). La clasificación de las proteínas TPP coincide parcialmente con un trabajo realizado por Rahman *et al*, (2021). Las plantas dicotiledóneas y monocotiledóneas se agrupaban en diferentes subclados: A3, B1 y B2 pertenecían a plantas monocotiledóneas, mientras que el grupo A4 pertenecía a las dicotiledóneas. En comparación con *Arabidopsis* (10), arroz (12), *Brachypodium* (10) y maíz (15), las proteínas SITPP mostraron una disminución. Las proteínas SITPP se colocaron solamente en tres grupos del clado A: A1, A2, A4, que contenían tres proteínas cada uno (SITPP-2, SITPP-7, SITPP-8), (SITPP-1, SITPP-5, SITPP-9) y (SITPP-3, SITPP-4, SITPP-6), respectivamente.

Posteriormente reconstruimos un árbol filogenético con los miembros de las familias SITPS/SITPP en jitomate (Figura 3), donde observamos sorprendentemente que SITPP-9, que anteriormente habíamos clasificado como TPP debido a que contenía solamente un dominio trehalosa incompleto, apareció relacionado cercanamente con la familia TPS.

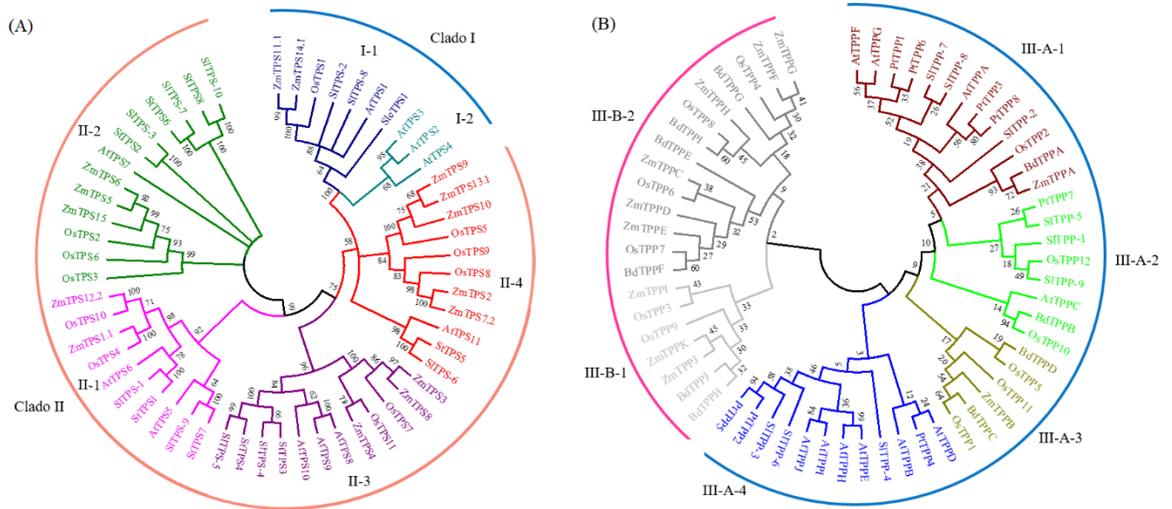


Figura 2. Árbol filogenético circular de TPSs y TPPs. (A) Árbol filogenético de los miembros proteicos de la familia TPS. (B) Árbol filogenético de los miembros proteicos de la familia TPP.

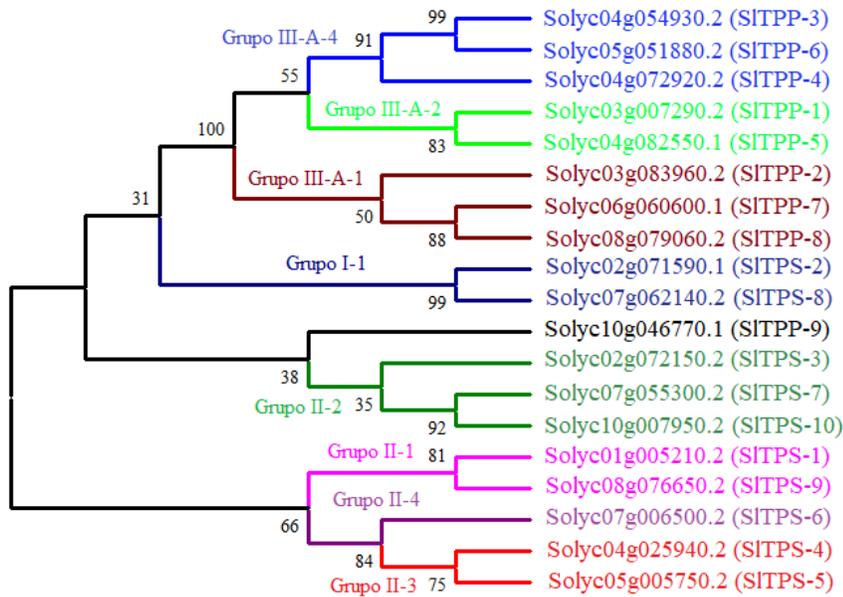


Figura 3. Árbol filogenético rectangular de SITPS/SITPP.

6.4. Localización subcelular de proteínas de las familias TPS/TPP en jitomate

La localización de una proteína a menudo se correlaciona con su función molecular. Por lo tanto, el conocimiento de la localización de proteínas puede ayudar a los biólogos a inferir la función de una proteína (Briesemeister *et al*, 2010).

Al predecir la localización subcelular de las proteínas en YLoc, encontramos que seis TPS (SITPS-3, SITPS-4, SITPS-5, SITPS-6, SITPS-7 y SITPS-9) se ubicaron específicamente en el citoplasma, tres proteínas (SITPS-1, SITPS-2 y SITPS-8) se ubicaron en el citoplasma y núcleo, y una proteína (SITPS-10) se ubicó en el citoplasma y mitocondria. En papa (Xu *et al*, 2017) y algodón (Mu *et al*, 2016) la mayoría de las proteínas TPS podrían localizarse en el citoplasma, sólo algunas en la membrana plasmática y núcleo. El dominio N-terminal que posee AtTPS1 regula su distribución entre núcleo y citosol (Vandesteene *et al*, 2010).

La localización subcelular de siete proteínas TPP (SITPP-1, SITPP-3, SITPP-4, SITPP-5, SITPP-6, SITPP-7 y SITPP-8) se predijo en el citoplasma y núcleo, SITPP-2 en el citoplasma, núcleo y peroxisoma, y SITPP-9 se encontró en citoplasma, núcleo y espacio extracelular (Cuadro 1). En arabidopsis se observaron diferentes tipos de localizaciones: AtTPPG, AtTPPI y AtTPPJ se detectaron en el núcleo. AtTPPA, AtTPPB, AtTPPC, AtTPPF y AtTPPH se mostraron en el citosol y núcleo. AtTPPD y AtTPPE estaban localizados en cloroplasto (Krasensky *et al*, 2014).

Recientemente se encontró una señal de orientación peroxisomal putativa y no canónica tipo 1 (PTS1) en AtTPPI, e identificaron a esta proteína como multi-dirigida a peroxisoma, núcleo/nucleolo y plastidios (Kataya *et al*, 2020). AtTPPI muestra una identidad de 67% con SITPP-3 (Apéndice 6) y según el análisis de sintenia AtTPPI y SITPP-3 son regiones conservadas en el genoma de arabidopsis y jitomate. Sin embargo, no se detectó esta señal en la proteína SITPP-3 utilizando la herramienta web PredPlantPTS1 (Reumann y Lingner, 2012).

Para la búsqueda de péptidos de tránsito al cloroplasto y/o mitocondria tanto en las secuencias de las proteínas SITPS como SITPP se emplearon tres programas web: TargetP 1.1, ChloroP 1.1 Server y LOCALIZER. En las proteínas SITPS, TargetP 1.1 encontró péptidos de tránsito al cloroplasto en SITPS-1, SITPS-2 y SITPS-7. En cambio, ChloroP 1.1 Server no detectó péptidos en ninguna de las proteínas. En LOCALIZER se predijeron péptidos en las proteínas SITPS-1 y SITPS-10. Para SITPP, utilizando TargetP 1.1 se hallaron péptidos de tránsito al

cloroplasto en SITPP-2, SITPP- 4, SITPP-6 y SITPP-8; y ChloroP 1.1 Server los detectó en SITPP-2, SITPP-4 y SITPP-8. LOCALIZER sólo predijo este tipo de péptido en SITPP-9 (Cuadro 2).

La predicción de péptidos de tránsito a la mitocondria se realizó en TargetP 1.1 y LOCALIZER para las proteínas SITPS y SITPP. Para SITPS, en TargetP se encontró el péptido en SITPS-10, y en LOCALIZER no se predijo ningún péptido para las proteínas TPS (Cuadro 2). En SITPP no se detectaron péptidos.

Algunas proteínas contienen secuencias llamadas señales de localización nuclear o NLSs, que les permiten entrar en el núcleo unidas a otras proteínas transportadoras. Mientras que, otras secuencias de aminoácidos denominadas señales de exportación nuclear o NESs, les permiten salir del núcleo al citoplasma formando complejos con una proteína transportadora llamada CRM1 (Cautain *et al*, 2015), por lo tanto, se predijeron las NLSs y NESs en las proteínas SITPS y SITPP mediante las herramientas LOCALIZER y NetNES. Las proteínas SITPS-3, SITPS-4, SITPS-5, SITPS-7, SITPS-9 y SITPS-10 contenían secuencias NLS y NES. SITPS-1 y SITPS-6 únicamente presentaron NES. SITPS-8 sólo tenía NLS. Las proteínas TPP que presentaron secuencias NLS y NES fueron SITPP-3, SITPP-5 y SITPP-6, mientras que SITPP-2 y SITPP-4 sólo contenían NES (Cuadro 2).

Cabe señalar que algunas proteínas pueden ser transportadas dentro y fuera del núcleo a través de la interacción con otras proteínas que contengan señales de transporte funcionales, a pesar de que ellas no tengan ni NLS ni NES (Cautain *et al.*, 2015), por lo que algunas proteínas que se predijeron como localizadas en el núcleo en YLoc y que en sus secuencias no se encontraron NLS ni NES es probable que interactúen con otras proteínas para ser transportadas al núcleo.

Cuadro 2. Predicción de péptidos de tránsito al cloroplasto y/o mitocondria y secuencias NLS y NESs.

	TargetP	ChloroP	NetNES	LOCALIZER		
				Cloroplasto	Mitocondria	NLS
SITPS-1	C	-	NES	Y (0.883 1-32)	-	-
SITPS-2	C	-	-	-	-	-
SITPS-3	-	-	NES	-	-	Y (KRRS, KKLSIDTIVSAYSRAKNR)
SITPS-4	-	-	NES	-	-	Y (RKLSLDHIVSSYRRTQRR)
SITPS-5	-	-	NES	-	-	Y (RKLSIDHIVSSYRRTQRR)
SITPS-6	-	-	NES	-	-	-
SITPS-7	C	-	NES	-	-	Y (KRRP, RKLSIDDIVNAYIKSKSR)
SITPS-8	-	-	-	-	-	Y (KRHR)
SITPS-9	-	-	NES	-	-	Y (RKLSVEHIVSAYKRTKHR)
SITPS-10	M	-	NES	Y (0.722 1-41)	-	Y (KRRP, RKLSIDDIESAYIKSKSR)
SITPP-1	-	-	-	-	-	-
SITPP-2	C	Y	NES	-	-	-
SITPP-3	-	-	NES	-	-	Y (RKLARYFPTAIVSGRCRDK, RRLVEWKKLSLRRQFRIRR)
SITPP-4	C	Y	NES	-	-	-
SITPP-5	-	-	NES	-	-	Y (LGEV)
SITPP-6	C	-	NES	-	-	Y (RKLARYFPTAIVSGRCRDK)
SITPP-7	-	-	-	-	-	-
SITPP-8	C	Y	-	-	-	-
SITPP-9	-	-	-	Y (0.901 1-51)	-	-

6.5. Localización en cromosoma y análisis de duplicación genética de las familias *TPS/TPP* en jitomate

Generamos un mapa que muestra las posiciones cromosómicas de todos los genes *SITPS/SITPP* (Figura 4); nueve de los doce cromosomas de jitomate albergaban al menos un gen *SITPS* o *SITPP*, y la mayoría de los genes residían cerca de extremos cromosómicos. Los cromosomas 4 y 7 tenían la mayor cantidad de genes *SITPS/SITPP* (4 y 3, respectivamente). Los cromosomas 2, 3, 5, 8 y 10 poseían cada uno dos genes *SITPS/SITPP*, mientras que los cromosomas 1 y 6 contenían un solo gen *TPS* o *TPP*.

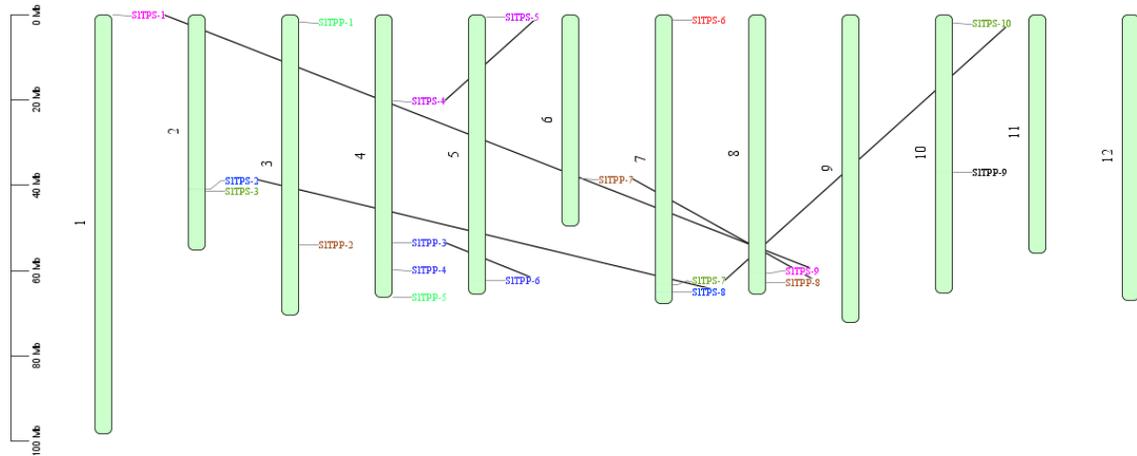


Figura 4. Distribución de los 10 genes *SITPS* y 9 genes *SITPP* en los 12 cromosomas de jitomate. Los genes duplicados en el genoma están conectados por las líneas de color negro. Los números cromosómicos se indican al lado izquierdo de cada cromosoma. Se estimaron tamaños cromosómicos y ubicaciones genéticas utilizando la escala millones de pares de base (Mbp) a la izquierda de la figura.

La duplicación genética es la principal vía para la expansión de una familia de genes y que resulta en la divergencia funcional, incluye duplicación de genoma completo, duplicaciones segmentales y duplicaciones en tándem (Kong *et al*, 2007). Un par genético se considera duplicado en tándem si está separado por cinco o menos genes que residen dentro de una ventana de 100 kbp en el mismo cromosoma (Wai *et al*, 2021). Basado en los resultados del alineamiento de secuencia, % cobertura de consulta, % de identidad de secuencia (Apéndice 3 y Apéndice 4) y el árbol filogenético (Figura 3), se identificaron 4 pares de genes *SITPS* (*SITPS-5/SITPS-4*, *SITPS-2/SITPS-8*, *SITPS-1/SITPS-9* y *SITPS-7/SITPS-10*) y 2 pares de genes *SITPP* (*SITPP-3/SITPP-6* y *SITPP-7/SITPP-8*) probablemente originados por duplicación segmental (Cuadro 3). No se detectaron genes *SITPS* ni *SITPP* duplicados en tándem. Estos resultados sugieren que la duplicación segmental puede haber contribuido en mayor medida a la expansión de la familia de genes *SITPS* y *SITPP*. En un estudio anterior encontraron que el genoma de arabis contenía 2 pares de genes duplicados específicos de esta especie (*AtTPS2/4* y *AtTPS7/8*). *AtTPS7* y *AtTPS8* fueron creados por un evento reciente de duplicación del genoma completo de

Arabidopsis (Yang *et al*, 2012). *AtTPS2* y *AtTPS3* probablemente surgieron de la duplicación segmental de la región del gen *AtTPS1*, seguida de una duplicación en tándem que dio lugar al par de genes *AtTPS2* y *AtTPS3*. En *Populus* se encontraron cuatro pares duplicados por un evento de duplicación del genoma completo y un par se formó mediante un evento de duplicación segmentaria (Yang *et al*, 2012). En el arroz se detectaron cuatro pares de genes duplicados, tres de los cuales se formaron por un evento de duplicación segmentaria (Yang *et al*, 2012). Otro análisis mostró que cuatro pares de genes *BdTPP* estaban formados por eventos de duplicación (Wang *et al*, 2019). Por otro lado, en *Arabidopsis* cuatro pares de genes *AtTPP* se originaron a partir de la duplicación del genoma, a excepción de *AtTPPA* y *AtTPPD* (Vandesteene *et al*, 2012).

Para evaluar el tipo de presión selectiva impuesta a los miembros de las familias de genes *TPS* y *TPP* duplicados segmentalmente, calculamos la relación Ka/Ks para cada par de genes paralogos. La tasa de sustitución no sinónima (Ka), la tasa de sustitución sinónima (Ks) y la relación Ka/Ks se utilizan para comprender la dirección de la evolución y su fuerza selectiva en una secuencia codificada. $Ka/Ks > 1$ indica una selección positiva, $Ka/Ks < 1$ indica una selección negativa y $Ka/Ks \approx 1$ indica una evolución neutra (Li *et al*, 2009). Las proporciones Ka / Ks de los genes duplicados eran inferiores a 1, lo que sugiere que estos genes experimentaron una fuerte selección negativa o purificadora (Hurst, 2002). El rango de tiempo de divergencia de los genes *SITPS* fue hace 43 a 306 millones de años (MYA), lo que indica que los pares de genes *SITPS-5* y *SITPS-4* se duplicaron recientemente (43 MYA). En el caso de los genes *SITPP* fue hace 38 a 54 MYA, los genes más recientemente duplicados son *SITPP-7* y *SITPP-8* (Cuadro 3).

Además, construimos un mapa microsinténico de genes de la familia *TPS/TPP* del genoma de jitomate y el de *Arabidopsis*, para encontrar las regiones conservadas evolutivamente de los genes presentes en el genoma de ambas especies. Identificamos 8 pares ortólogos en cada familia de genes (Figura 5).

Cuadro 3. Estimaciones de las relaciones Ka/Ks de los genes SITPS/SITPP duplicados segmentadamente con su tiempo de divergencia en jitomate.

S. N°	Genes duplicados		Query Cover (%)	Ident (%)	Ka	Ks	Ka/Ks	Time (MYA)
1	SITPS-1	SITPS-9	99%	76.64%	0.1613	4.0201	0.0401	306.4112
2	SITPS-2	SITPS-8	84%	75.49%	0.1594	1.8591	0.0857	141.6984
3	SITPS-5	SITPS-4	100%	88.45%	0.0609	0.5661	0.1075	43.149
4	SITPS-7	SITPS-10	99%	89.20%	0.0618	0.6621	0.0933	50.465
5	SITPP-3	SITPP-6	90%	70.73%	0.2358	0.7122	0.3311	54.286
6	SITPP-7	SITPP-8	98%	70.68%	0.1253	0.51	0.2456	38.870

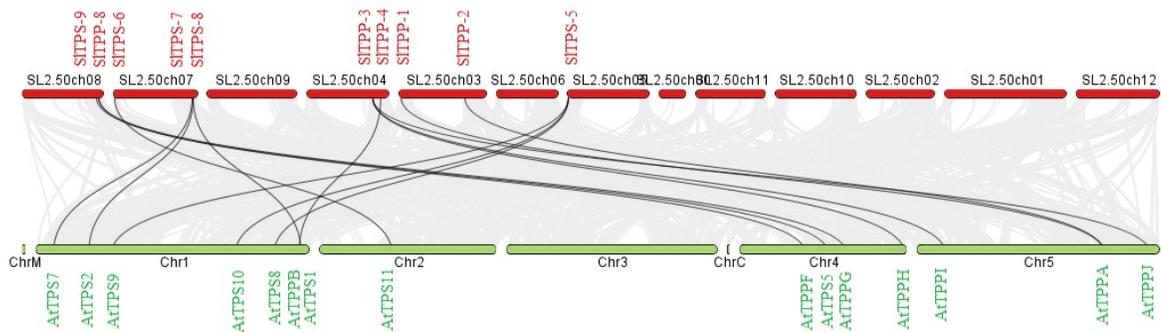


Figura 5. Análisis de sintenia de genes TPS/TPP entre jitomate y Arabidopsis. Los cromosomas de las dos especies están representados en diferentes colores: jitomate, rojo; y Arabidopsis, verde.

6.6. Distribución exón-intrón de genes *TPS/TPP* en jitomate

Entre los *SITPS*, la mayoría de los genes tienen 2 intrones, excepto *SITPS-2* y *SITPS-8* de clase I que tienen 17 y 18 intrones, respectivamente (Figura 6A), notablemente mayor en contraste con dos exones en los genes *SITPS* clase II. En Arabidopsis y el arroz, los genes de la clase I poseían 16 intrones y los genes de la clase II tenían dos intrones (Yang *et al*, 2012). Entre los *SITPPs*, el número de intrones fue más variable, desde *SITPP-9* con un intrón; *SITPP-1* y *SITPP-6* con 7; *SITPP-3*, *SITPP-4* y *SITPP-5*, cada uno de ellos con 8 intrones; *SITPP-2* con 9 intrones; dos genes, *SITPP-7* y *SITPP-8*, cada uno con 10 intrones (Figura 6A). Según un informe anterior (Hu, 2016) la tasa de ganancia de intrones es más baja que la de pérdida de intrones después de la duplicación segmentaria en el

arroz. Por lo que se puede concluir que *SITPS-2*, *SITPS-8*, *SITPP-7* y *SITPP-8* podrían ser los genes ancestrales junto con otros genes derivados de ellos.

6.7. Análisis de dominios conservados y motivos proteicos de las familias TPS/TPP en jitomate

La alineación de múltiples secuencias de aminoácidos y el análisis de dominios conservados mostró que las proteínas de la familia TPS de jitomate consistían en dos dominios conservados comunes, incluyendo un dominio típico de la familia Glycosyltransferase 20 (Glyco_transf_20) ubicado en la región N-terminal y un dominio de trehalosa fosfatasa (Trehalose_PPase) ubicado en la región C-terminal de la proteína TPS (Figura 6B), lo que sugiere que las estructuras de los dos dominios se fundaron básicamente antes de la diferenciación de estos genes, lo cual estuvo de acuerdo con las observaciones anteriores (Yang *et al*, 2012; Lin *et al*, 2018; Wang *et al*, 2019; Yang *et al*, 2020; Du *et al*, 2017). En la clase I, se observaron solamente estos dos dominios, mientras que en la clase II también se encontraron los dominios Hydrolase_3 y Glyco_transf_1, este último en los genes *SITPS-7* y *SITPS-10* pertenecientes al grupo II-2.

En un estudio preliminar encontraron un dominio Trehalose_PPase corto para OsTPP11 y un dominio truncado para OsTPP12 (Rahman *et al*, 2021). Anteriormente OsTPP11 y OsTPP12 se informaron como pseudogenes putativos (Ge *et al*, 2008). A partir de nuestros análisis, encontramos un dominio Trehalose_PPase que se conservaba en la mayoría de las proteínas SITPPs, excepto en SITPP-9, donde estaba incompleto (Figura 6B). Chen y Lou (2017) predijeron SITPP-9 como una TPS incompleta, pues sólo contiene un dominio parcial TPP y la denominaron SITPS11. En el análisis filogenético (Figura 2B) se observó que OsTPP12 y SITPP-9 se encuentran cercanamente relacionados, lo que apoya nuestra clasificación de SITPP-9 como TPP y no como TPS. En el análisis de motivos proteicos, se encontraron dos motivos conservados (motivo 1 y 6) en ocho SITPPs mientras que en SITPP-9 sólo se halló el motivo 6.

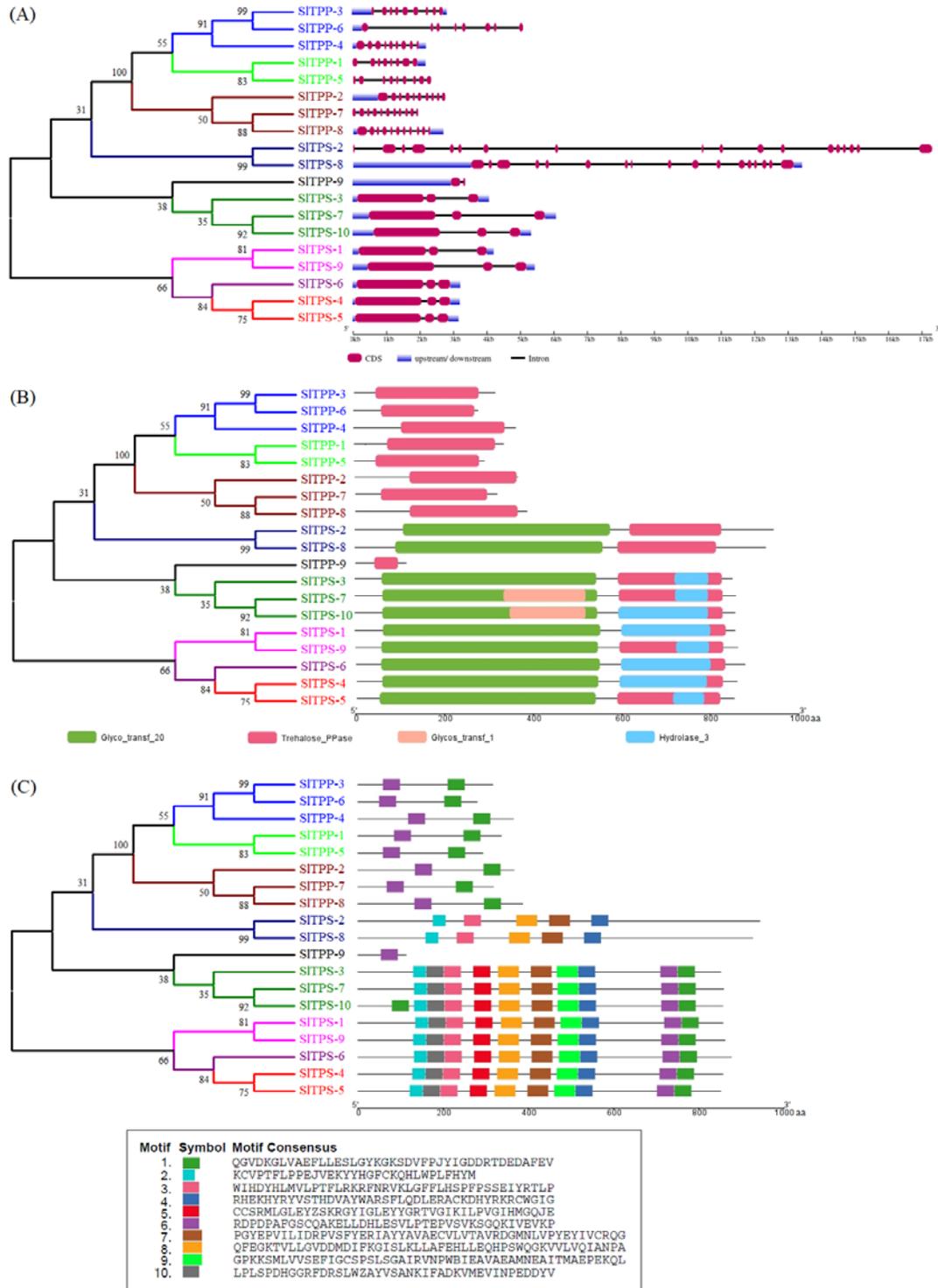


Figura 6. Representación esquemática de la distribución de exón-intrón, dominios y motivos de SITPP/SITPS. (A) Distribución exón-intrón. Cajas rojas representan los exones, las líneas negras representan los intrones. Regiones no-traducidas (UTRs) se indican con cajas de color azul. (B) Dominios. (C) Motivos conservados dibujados en software TBtools.

Leyman *et al.*, (2001) describió dos firmas de fosfatasa: LDYD|GD|T|LM y GDDRSD en plantas, pertenecientes al motivo 1 y 6, respectivamente. La primera firma se presentó como LDYDGTLS en las siete proteínas SITPPs (SITPP-1, SITPP-2, SITPP-3, SITPP-4, SITPP-5, SITPP-7, SITPP-8), en SITPP-9 se conservaron sólo tres aminoácidos y en SITPP-6 no. La segunda firma GDDRSD se conservó en la mayoría de las proteínas SITPPs excepto en SITPP-9.

Analizamos secuencias de motivos para los SITPS/SITPP utilizando la herramienta MEME e identificamos diez motivos significativos (motivo 1–10) (Figura 6C). Las proteínas TPS de jitomate clasificadas en el mismo clado en el árbol filogenético presentaron una composición de motivo idéntica, lo cual coincide con Dan *et al.*, (2021). Por ejemplo, todos los miembros de la clase II contenían los diez motivos altamente conservados. SITPS-10 presentó dos motivos 1, pero el motivo que aparece al inicio de la secuencia está menos conservado como se puede observar en el alineamiento múltiple (Figura 7). Las proteínas del grupo I-2 (SITPS-2 y SITPS-8) contenían solamente cuatro motivos: 2, 3, 7 y 8, lo que sugiere que estas diferencias entre las dos clases pueden estar involucradas en diversas funciones. Los motivos 1, 6 corresponden al dominio de trehalosa fosfatasa y el resto de los motivos al dominio de la familia glycosyltransferase 20 (Apéndice 7).

La mayoría las proteínas SITPPs presentaron los motivos 1 y 6, a excepción de SITPP-9 que solamente contenía el motivo 6. Esta presencia y ausencia de motivos en las proteínas indican su acción y función específicas. Los pares paralogos también mostraron un número similar de motivos (Figura 6C). También identificamos los motivos en la alineación de secuencias múltiples (Figura 7 y Figura 8).

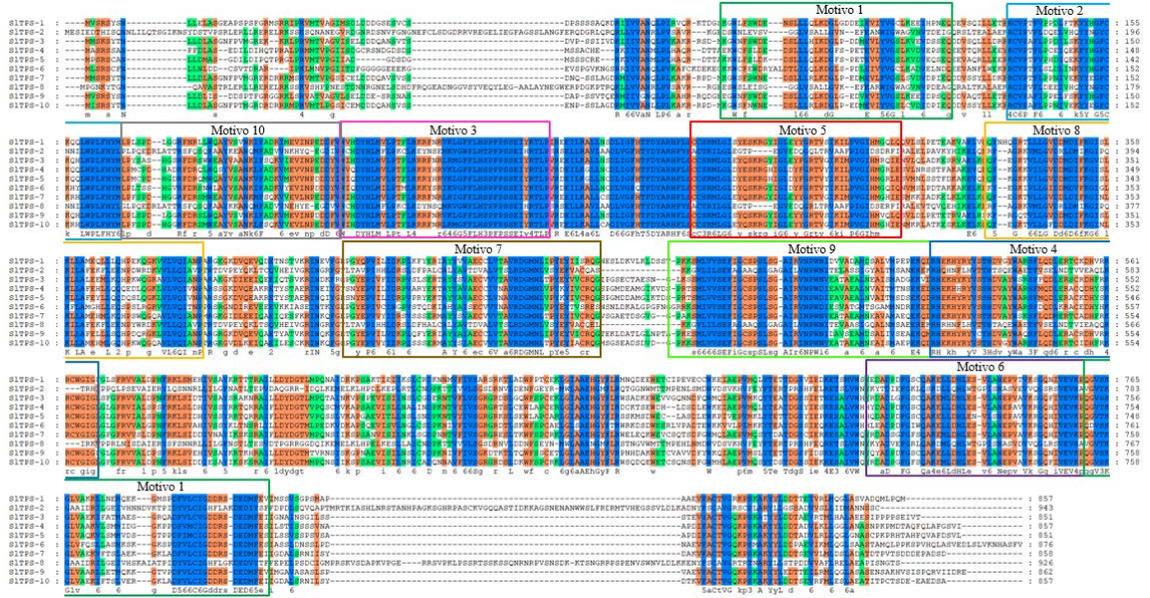


Figura 7. Representación esquemática del alineamiento de todas las proteínas SITPPs.

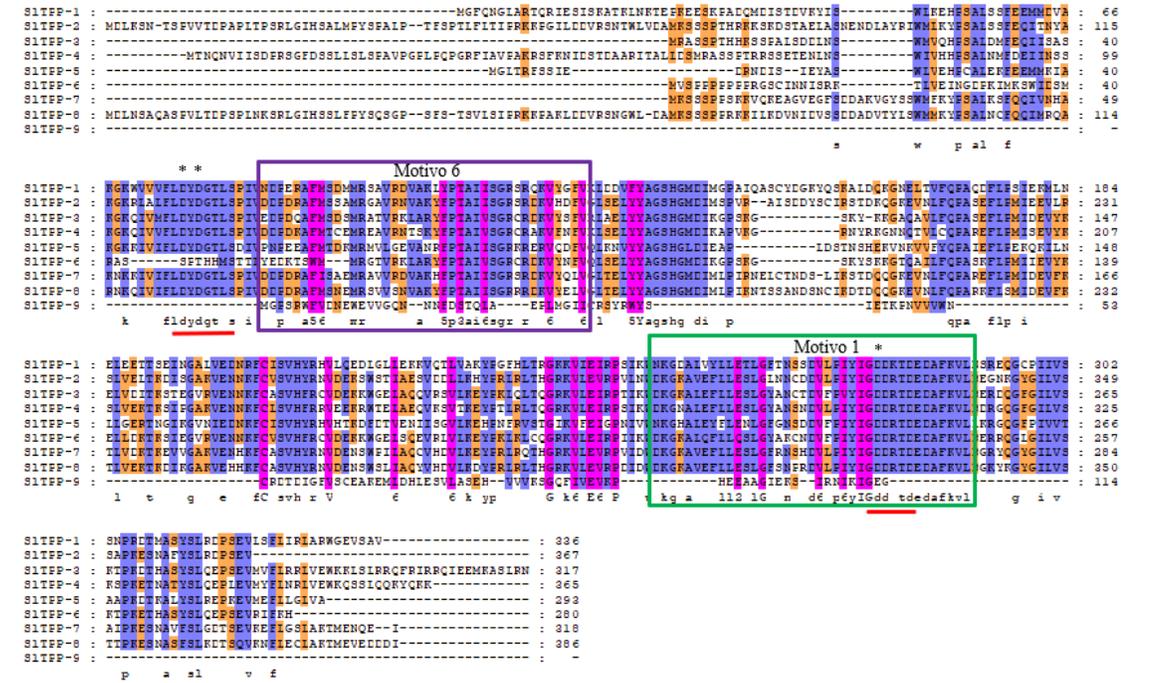


Figura 8. Representación esquemática del alineamiento de todas las proteínas SITPPs. Las firmas descritas por Leyman et al, 2011 están subrayadas de color rojo. Los sitios conservados se indican con asteriscos (*).

6.8. Análisis de estructura 3D de proteínas TPS/TPP en jitomate

Conocer la estructura tridimensional de una proteína es fundamental para comprender su función biológica a nivel molecular (Bordoli y Schwede, 2012). Además, la información sobre ligandos, cofactores y sitios de unión a menudo proporciona pistas cruciales sobre su función (Gallo *et al*, 2014).

En SWISS-MODEL (Waterhouse *et al*, 2018) se identificaron estructuras 3D de las diez proteínas SITPSs (Figura 9A): para el análisis de todas las SITPSs se utilizó la plantilla "5hut.1. A". El análisis de la estructura secundaria en Sopma mostró que las proteínas SITPS estaban compuestas por un 42% - 43.71% hélice α , un 14.25% - 17.39% de hebra extendida, un 4.55% - 5.82% giro β y un 34.43% - 36.69% espiral aleatoria (Apéndice 8). El estado del oligo para todos los SITPSs fue un monómero. Ninguno de ellos mostró ligandos. Se utilizó Structure Assessment de SWISS-MODEL para validar las estructuras SITPS (Figura 9B y Apéndice 9). Las gráficas Ramachandran generadas ocupaban una media del 93.3% de la región favorecida y el 1.53% de la región exterior. La identidad media de secuencia fue del 40.69%, mientras que la similitud de secuencias fue del 41.20%, y una cobertura de 53.50% del total de la secuencia de consulta encontrada por método de rayos X en 1.90 Å (Apéndice 9).

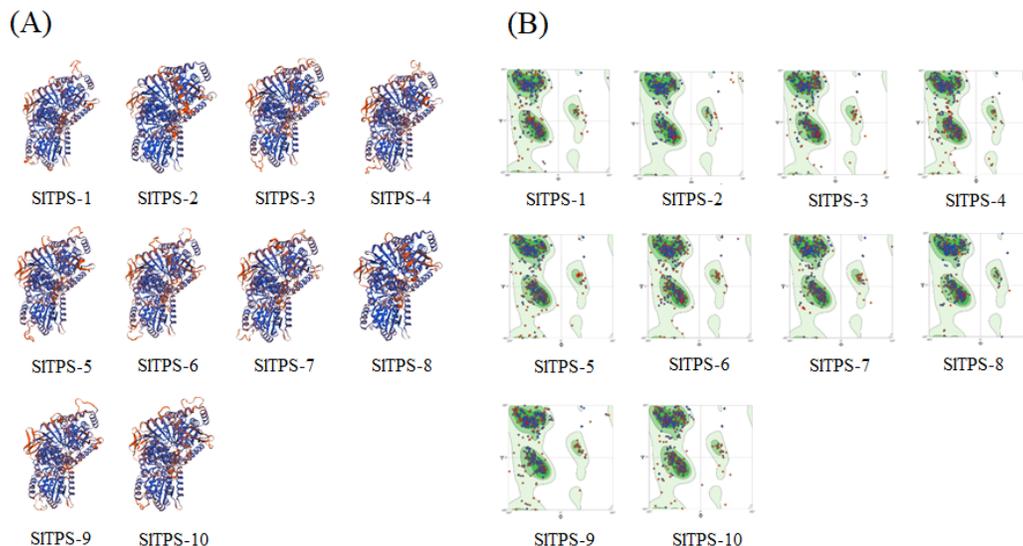


Figura 9. Representación esquemática de las estructuras 3-D de las proteínas y la gráfica de Ramachandran para SITPSs (A) Estructuras 3-D de las proteínas SITPSs predichas. (B) Gráficas Ramachandran para SITPSs.

Al modelar la estructura 3D de SITPPs encontramos tres residuos conservados, que actuaron como ligandos interactuando entre la cadena A y el ión de magnesio (Mg^{2+}) dentro de 4 Å en 7 de las 9 proteínas, lo que reveló que los SITPPs tienen funciones catalíticas específicas.

En SWISS-MODEL (Waterhouse *et al*, 2018) se identificaron estructuras 3D de las nueve proteínas SITPPs (Figura 10A): las primeras ocho SITPPs se analizaron con la plantilla "5gvx.1. A" y SITPP-9 con "5dx9.1. A". En Sopma se calcularon elementos de estructura secundaria de secuencias proteicas (Apéndice 10). Las proteínas SITPP estaban compuestas por un 32.70% - 46.06% hélice α , un 15.26% - 21.05% de hebra extendida, un 4.39% - 10.43% giro β y un 29.87% - 44.69% una espiral aleatoria.

En el extremo -C, se observó una estructura en forma de bobina enrollada en los primeros ocho SITPPs (SITPP-1 a SITPP-8). El estado del oligo para todos los SITPPs fue un homo-dímero, siete de ellos mostraron ligandos de iones de magnesio (Mg^{2+}), pero no SITPP-6 y SITPP-9 (Figura 10A). Utilizamos Structure Assessment de SWISS-MODEL para validar las estructuras SITPP (Figura 10B y Apéndice 11). Las gráficas Ramachandran generadas ocupaban una media del 92.63% de la región favorecida y el 1.42% de la región exterior (Apéndice 11). La identidad media de secuencia fue del 33.95%, mientras que la similitud de secuencias fue del 37.11%, y una cobertura de 78.11% del total de la secuencia de consulta encontradas por método de rayos X en 2.6 Å (Apéndice 11).

La fuente de información de interacción proteína-ligando (PLIP) confirmó la interacción del ligando entre la cadena A y Mg^{2+} , y se encontró que el sitio del residuo estaba altamente conservado (Apéndice 11). Además, analizamos estos residuos conservados en todos los alineamientos de la secuencia de la proteína SITPP y encontramos que estos tenían ácido aspártico (D / Asp) (Figura 8).

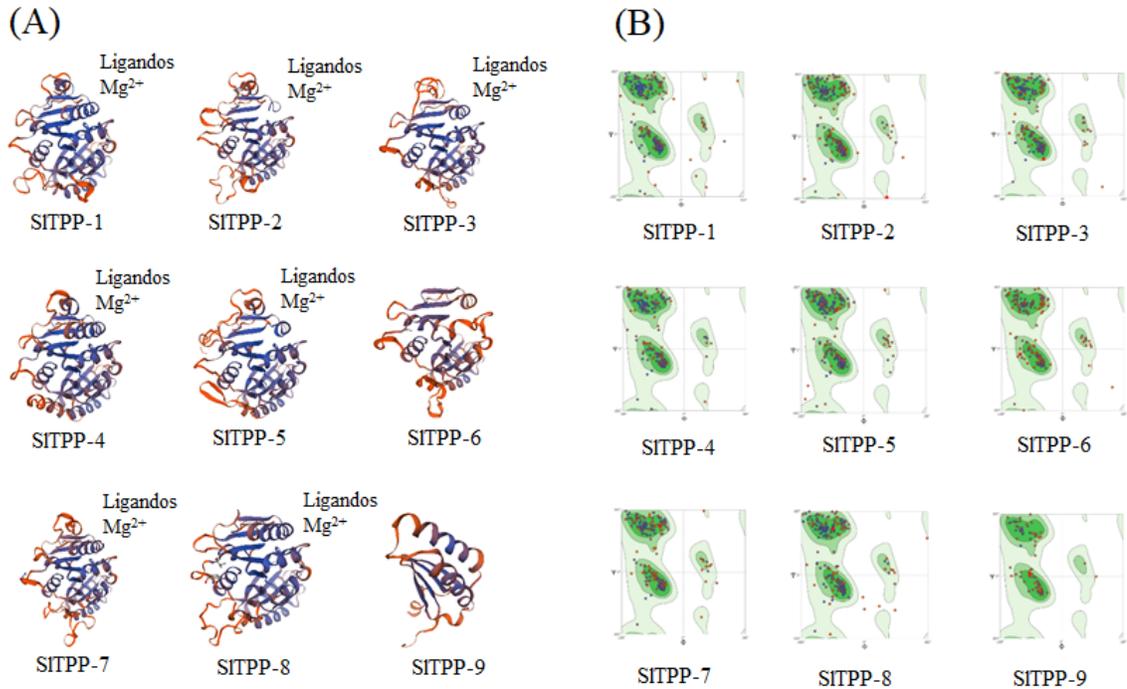


Figura 10. Representación esquemática de las estructuras 3-D de las proteínas y la gráfica de Ramachandran para SITPPs. (A) Estructuras 3-D de las proteínas SITPPs predichas. (B) Gráficas Ramachandran para SITPPs.

Según el análisis de la estructura 3D, en las proteínas SITPS no se detectaron ligandos mientras que, en la mayoría de las SITPP, excepto SITPP-6 y SITPP-9, si se observaron sitios del ligando Mg²⁺ y una estructura más conservada que SITPP-6 y SITPP-9. Lo anterior se asemeja al modelado 3D en arroz, pues de las doce OstPP, once mostraron ligandos de iones de magnesio, siendo OstTPP12 el único que no. OstTPP1 a OstTPP10 mostraron una estructura más conservada que OstTPP11 y OstTPP12 (Rahman *et al*, 2021). El análisis de la estructura secundaria en Sopma mostró que las proteínas SITPS y SITPP estaban compuestas por hélice α , hebra extendida, giro β y espiral aleatoria, aunque el porcentaje de estas estructuras varió un poco.

6.9. Análisis de elementos reguladores *cis* en promotores de genes *TPS/TPP* en jitomate

Analizamos regiones promotoras de 1.5 kb inmediatamente aguas arriba (“upstream”) del codón de inicio (ATG) de todos los genes *SITPS* / *SITPP*, descargadas de Phytozome, utilizando la herramienta en línea PlantCARE (Lescot *et al*, 2002) y se visualizaron en TBtools (Chen *et al*, 2018), identificamos y clasificamos los CRE (Cis Regulatory Element) en cinco grupos: relacionados con la luz, el estrés, las hormonas, el crecimiento y desarrollo, y otros CRE (Apéndice 12 y Apéndice 13).

6.9.1 Elementos reguladores *cis* relacionados con la luz

Los promotores de los genes *SITPS*s contenían un total de 18 CRE relacionadas con la luz (LrCRE). Cada uno de los genes *TPS* tenían de 3 (*SITPS-1* y *SITPS-2*) a 10 (*SITPS-7*) LrCRE diferentes (Figura 11). Todos los LrCRE en los promotores de los *SITPS*s se visualizaron en el Apéndice 14.

Los promotores de los genes *SITPP*s contenía un total de 19 CRE relacionadas con la luz (LrCRE). Se detectaron de 4 (*SITPP-4*) a 8 (*SITPP-3*) LrCRE en la mayoría de los *SITPP*s, excepto en *SITPP-5* donde no hubo (Figura 11). Todos los LrCRE en los promotores de los *SITPP*s se visualizaron en el Apéndice 15.

Los elementos regulatorios *cis* ABRE3a y ABRE4 relacionados con la respuesta a la luz (Rahman *et al*, 2021) se encontraron en tres genes *TPS* (*SITPS-7*, *SITPS-8*, y *SITPS-10*) y en cinco genes *TPP* (*SITPP-1*, *SITPP-3*, *SITPP-6*, *SITPP-7* y *SITPP-8*); ACE también relacionado con la luz (Kaur *et al*, 2017) en *SITPS-10* y en tres *TPP* (*SITPP-1*, *SITPP-3* y *SITPP-7*).

G-box (CACGTG) que está involucrado en la respuesta a la luz, el ácido abscísico, el metil-jasmonato, la anaerobiosis y tiene un papel en la inducción de etileno, así como en la expresión específica en semillas (Kaur *et al*, 2017), se encontró en ocho promotores *TPS* (todos menos *SITPS-1* y *SITPS-9*), y en siete *TPP* (todos menos *SITPP-5* y *SITPP-9*), fue el elemento LrCRE más frecuente en las dos familias génicas.

Box 4, un módulo conservado de ADN involucrado en respuesta a luz (Sadau *et al*, 2021) se encontró en la mayoría de los genes *TPS* excepto *SITPS-4*, *SITPS-*

5 y *SITPS-6*, y en siete *TPP* (todos menos *SITPP-5* y *SITPP-9*), siendo otro de los elementos más abundantes.

MRE, un sitio de unión MYB involucrado en la respuesta a luz (Rahman *et al*, 2021) se localizó en *SITPS-5* y *SITPS-7*; y en *SITPP-3* y *SITPP-9*.

El motivo ACA, que es parte de *gapA* relacionado con la luz (Sadau *et al*, 2021) se halló en *SITPS-1*. Los elementos de respuesta a luz: 3-AF1 binding site y Sp1 (Mongkolsirawatana *et al*, 2009) se encontraron en *SITPP-9* y *SITPP-7*, respectivamente; mientras que GT1-motif se localizó en 6 *TPS* (*SITPS-4*, *SITPS-6*, *SITPS-7*, *SITPS-8*, *SITPS-9* y *SITPS-10*) y en 4 promotores *TPP* (*SITPP-2*, *SITPP-4*, *SITPP-8* y *SITPP-9*). GT1-motif se ha encontrado en muchos promotores *rbcS*, sirve como sitio de unión de un factor nuclear, denominado GT-1 y el análisis de mutantes específicos de sitio del promotor *rbcS-3A* de chícharo demostró que la unión de GT-1 *in vitro* está correlacionada con la expresión sensible a la luz del promotor *rbcS* en plantas transgénicas (Lam y Chua, 1990). AT1-motif se halló en *SITPP-6*; AE-box en tres promotores *TPS* (*SITPS-3*, *SITPS-5* y *SITPS-9*) y cuatro *TPP* (*SITPP-1*, *SITPP-3*, *SITPP-8* y *SITPP-9*), respectivamente. Estos dos elementos son parte de un módulo de respuesta a luz (Mongkolsirawatana *et al*, 2009).

Asimismo, se encontraron varios CREs como Box-II, *chs-CMA1a*, GA-motif, Gap-box, GATA-motif, I-box, L-box, TCCC-motif y TCT-motif que forman parte de un elemento relacionado con la luz (Kaur *et al*, 2017) en diferentes genes a excepción del gen *SITPS-10* (Figura 11).

Box II, *chs-CMA1a*, *chs-CMA2a*, Gap-box, I-box y LAMP-element que son parte de un elemento de respuesta a luz (Mongkolsirawatana *et al*, 2009) se hallaron solamente en un promotor cada uno, a saber, *SITPP-9*, *SITPP-1*, *SITPP-2*, *SITPP-8*, *SITPP-8* y *SITPP-2*, respectivamente. GATA-motif y TCT-motif que también son parte de un elemento de respuesta a luz (Kaur *et al*, 2017) se encontraron en 4 y 3 promotores respectivamente.

	SITPS-1	SITPS-2	SITPS-3	SITPS-4	SITPS-5	SITPS-6	SITPS-7	SITPS-8	SITPS-9	SITPS-10		SITPP-1	SITPP-2	SITPP-3	SITPP-4	SITPP-5	SITPP-6	SITPP-7	SITPP-8	SITPP-9	
ABRE3a								1	1	2	3-AF1 binding site										1
ABRE4								1	1	2	ABRE3a	1	1				1	1	3		
ACA-motif	1										ABRE4	1	1				1	1	3		
ACE										1	ACE	1	1						1		
AE-box			1		2					2	AE-box	1	1							2	1
Box II			1			2					AT1-motif						1				
Box 4	2	1	1				6	4	2	3	Box II										1
chs-CMA1a							1				Box 4	5	1	2	1		7	2	2		
G-Box		1	10	2	2	3	2	4		3	chs-CMA1a	1									
GA-motif	1						1		1		chs-CMA2a		1								
Gap-box						1					G-Box	2	2	3	1		1	1	12		
GATA-motif			1	1		1	1	2	1		Gap-box						1				
GT1-motif			2			1	2	2	1	1	GATA-motif		1	1							1
I-box			2								GT1-motif		1		1					2	2
L-box			1								I-box									2	
MRE				1		1					LAMP-element		1								
TCCC-motif								1	1		MRE			1							1
TCT-motif		3	1				1	2	1		Sp1								1		
											TCT-motif		1		1		2	1			

Figura 11. Elementos reguladores *cis* relacionados con la luz en *SITPS/SITPPs*. Los números en las cajas de colores indican las repeticiones.

6.9.2 Elementos reguladores *cis* relacionados con el estrés

Los promotores de los genes *SITPSs* contenían un total de 21 CRE relacionadas con el estrés (SrCRE). Los genes poseían de 5 (*SITPS-1*) a 13 (*SITPS-3* y *SITPS-8*) SrCRE (Figura 12). Todos los SrCRE en los promotores de los *SITPSs* se visualizaron en el Apéndice 16.

Los promotores de los genes *SITPPs* contenía un total de 20 CRE relacionadas con el estrés (SrCRE). Los genes tenían de 3 (*SITPP-5*) a 14 (*SITPP-9*) SrCRE (Figura 12). Todos los CRE relacionados con el estrés en los promotores de los *SITPPs* se visualizaron en el Apéndice 17.

MYC y MYB funcionan para la activación de la expresión de genes en respuesta a estrés hídrico, uniéndose a los elementos de respuesta MYCR y MYBR (MYC, MYB recognition sites), respectivamente (Abe *et al*, 2003). Myb-binding site se ubicó en *SITPS-5*, *SITPP-1*, *SITPP-5* y *SITPP-9*. MYC, elemento de respuesta a sequía (Pradhan *et al*, 2021) se halló en todos los genes *SITPS* y *SITPP*, el mayor número de este elemento estuvo presente en *SITPS-4*, *SITPS-5*, *SITPS-6*, *SITPP-1* y *SITPP-9*. MBS es un sitio de unión para los factores de transcripción MYB, involucrado en la inducción por sequía (Rahman *et al*, 2021) se encontró

en seis *TPS* (*SITPS-2*, *SITPS-3*, *SITPS-4*, *SITPS-5*, *SITPS-6* y *SITPS-8*) y en cuatro *TPP* (*SITPP-2*, *SITPP-7*, *SITPP-8* y *SITPP-9*). En el análisis de expresión de los genes *SITPS-4*, *SITPS-5*, *SITPS-6*, *SITPP-1* y *SITPP-9* se comprobó que efectivamente están regulados en condiciones de sequía, y algunos también responden al calor (Figura 16 y Figura 17).

Myb-like sequence, MYB recognition site, Myb, Myc y Unnamed__1, todos ellos elementos de respuesta a estrés (Rahman *et al*, 2021), se encontraron en algunos promotores de *SITPS* y *SITPP*.

STRE, un elemento de respuesta a estrés, que se une a las proteínas Msn2 y Msn4 (Gao *et al*, 2013) se encontró en todos los promotores de *SITPS* y en 8 *SITPP* (todos menos *SITPP-5*). Otro elemento de respuesta a estrés, MYB (Rahman *et al*, 2021) en 9 promotores *SITPS* y 7 *SITPP*.

El elemento sensible a la deshidratación (DRE), que contiene la secuencia central A / GCCGAC, está involucrado principalmente en la regulación de genes sensibles al frío, sequía y salinidad (Maruyama *et al*, 2012). Por ejemplo, DREB1A podría unirse al elemento DRE/CRT dentro del promotor *AtTPPF* y activar la expresión de *AtTPPF* (Lin *et al*, 2019).

DRE-core, elemento de respuesta al frío y a la deshidratación (Pradhan *et al*, 2021; Maruyama *et al*, 2012), se encontró en *SITPS-3* y *SITPP-4*. DRE1, un elemento de respuesta a sequía (Baranwal *et al*, 2017) se encontró anteriormente en el promotor del gen *Rab17* y respondió al aumento del ácido abscísico (ABA) y a los tratamientos de sequía. En este análisis, DRE1 se halló en *SITPS-5*, y su patrón de expresión mostró que aumenta su nivel de expresión en condiciones de sequía y salinidad (Figura 16 y Figura 20)

LTR, elemento involucrado con la respuesta a bajas temperaturas (Becerra *et al*, 2016; Pradhan *et al*, 2021) se encontró en *SITPS-4*, *SITPS-7*, *SITPS-8*, *SITPS-10* y *SITPP-2*. Sin embargo, de todos los genes *SITPS* y *SITPP*, únicamente *SITPP-4* mostró una regulación positiva al estrés por frío (Figura 18).

Además, se encontraron algunos elementos de respuesta a defensa como TC-rich repeats, CCAAT-box, box S, as-1, W-box (Rahman *et al*, 2021) en diferentes genes, excluyendo *SITPS-1*, *SITPS-9* y *SITPP-5*.

TC-rich repeats participa en la defensa de la planta y la respuesta al estrés (Wai *et al*, 2021; Wang *et al*, 2014). CCAAT-box es un sitio de unión de MYBHv1, involucrado en la activación de genes de defensa y estrés en plantas (Kaur *et al*, 2017). Box fue identificada como una nueva región reguladora de acción independiente en perejil, en los promotores de dos miembros de la familia de genes ELI7, ELI7.1 y ELI7.2, fuertemente inducidos de forma transcripcional después del tratamiento con un elicitor derivado del fitopatógeno *Phytophthora sojae* (Kaur *et al*, 2017), lo cual indica su papel en la elicitación. Además de que, box S participa en la respuesta a patógenos y heridas (Muthusamy *et al*, 2017; Li, 2014; Gao *et al*, 2014) Los promotores de genes inducibles por patógenos en plantas tienen múltiples elementos que actúan en *cis*, incluyendo W box, GCC y as-1 (Kim y Zhang, 2004). En este análisis solamente se detectaron as-1 y W box. Al elemento de respuesta a estrés as-1 se unen factores de transcripción de la familia TGA y se ha caracterizado como un elemento de respuesta a defensa, inducible por patógenos (Kong *et al*, 2018) mientras que, W-box interactúa con factores de transcripción WRKY y está involucrado en la respuesta a patógenos (Gao *et al*, 2014) por ejemplo, regula la expresión de genes relacionados con la defensa como PR10, proteínas relacionados con la patogénesis en arábidopsis (Kaur *et al*, 2017); y tiene un papel en el estrés biótico y abiótico como sequía y ABA; (Wang *et al*, 2009) latencia de la semilla y senescencia (Kaur *et al*, 2017). WRE3 (Wu *et al*, 2019) se encontró en 6 TPS y 3 promotores TPP (SITPP-1, SITPP-7 y SITPP-9) y WUN-motif en 3 TPS (SITPS-3, SITPS-8 y SITPS-10) y 4 TPP (SITPP-2, SITPP-6, SITPP-7 y SITPP-9), ambos involucrados en la respuesta a estrés por heridas (Jiang *et al*, 2014). ARE, elementos reguladores esenciales para la inducción anaeróbica (Li, 2014; Zhou *et al*, 2017) en todos los promotores, excepto SITPS-2, SITPS-10, SITPP-1 y SITPP-5.

	SITPS-1	SITPS-2	SITPS-3	SITPS-4	SITPS-5	SITPS-6	SITPS-7	SITPS-8	SITPS-9	SITPS-10		SITPP-1	SITPP-2	SITPP-3	SITPP-4	SITPP-5	SITPP-6	SITPP-7	SITPP-8	SITPP-9
ARE	1		1	3	1	4	1	1	2		ARE		2	1	1		1	2	3	2
as-1		4	4	1	1		1	2		1	as-1		2	1						1
box S						1				1	box S	1							1	
CCAAT-box		1		1						1	CCAAT-box	1	1							1
DRE core			1								DRE core			1						
DRE1					1						LTR		1							
LTR				3			1	1		1	MBS		1					2	6	1
MBS		1	1	2	1	2		1			MYB	3	3		1	1		3	1	7
MYB	2	1	2	2	5		1	2	2	1	MYB-like sequence				1			1		4
MYB-like sequence		1		2	1		1	1	1		MYB recognition site	1		1						1
MYB recognition site		1		1						1	Myb		2				2	2	7	4
Myb	1	1	1	3	5	2	1	1	1		Myb-binding site	1				1				2
Myb-binding site					3						MYC	5	2	3	4	1	3	4	1	5
MYC	1	3	1	4	3	3	1	2	2	1	Myc			1				1		
Myc			2				1	2			STRE	1	2	2	2		2	3	1	1
STRE	1	2	1	3	2	8	3	2	1	3	TC-rich repeats		1	1	1		1		2	
TC-rich repeats								1			Unnamed_1	1	2	2				1		
Unnamed_1		1	5			1	1			2	W box		1	1	1			1		2
W box			1		1	2				1	WRE3	2						2		1
WRE3			2		1	1		2	2	3	WUN-motif		1				4	2		1
WUN-motif			1					2		1										

Figura 12. Elementos reguladores *cis* relacionados con el estrés en *SITPS/SITPPs*. Los números en las cajas de colores indican las repeticiones.

6.9.3 Elementos reguladores *cis* relacionados con regulaciones hormonales

Los promotores de los genes *SITPSs* contenían un total de 14 CREs relacionados con regulaciones hormonales (HrCREs). Cada promotor contenía de 3 (*SITPS-1* y *SITPS-9*) a 8 (*SITPS-3* y *SITPS-4*) HrCREs (Figura 13) Todos los HrCREs encontrados en los promotores de los *SITPSs* se visualizan en el Apéndice 18.

Los promotores de los genes *SITPPs* contenían un total de 14 CREs relacionados con regulaciones hormonales (HrCREs). Cada promotor contenía 1 (*SITPP-5*) a 6 (*SITPP-9*) HrCREs (Figura 13). Todos los CRE relacionados con hormonas encontrados en los promotores de los *SITPPs* fueron visualizados en el Apéndice 19.

En un estudio anterior se encontró que en presencia de ABA o en condiciones de estrés, se activa la señalización ABA, por consiguiente, el factor de transcripción bZIP ABF2 se une directamente a la región del promotor de *AtTPPE* que contiene el elemento de respuesta al ácido abscísico (ABRE), y activa la expresión de *AtTPPE*, que conduce a la acumulación de trehalosa. La trehalosa induce la

acumulación de ROS en raíces y estomas, lo que mejora el efecto del ABA en la inhibición del crecimiento de las raíces y la promoción del cierre estomatal, permitiendo a la planta responder rápidamente a las condiciones de estrés (Wang *et al*, 2020). Además, ABRE, que contiene la secuencia central ACGTGG / T, regula la expresión génica en la deshidratación y la alta salinidad en Arabidopsis y arroz (Kazuo y Yamaguchi-Shinozaki, 2007). El elemento ABRE se encontró en la mayoría de los promotores de *SITPSs* y *SITPPs*, excepto en *SITPS-1*, *SITPS-9*, *SITPP-5* y *SITPP-9*, siendo *SITPP-8*, *SITPS-3* y *SITPS-6* los que presentaron mayor número de este elemento. En el caso de *SITPS-6* en el análisis de los patrones de expresión se mostró regulado positivamente en respuesta a estrés por sequía, calor y salinidad, mientras que *SITPS-3* y *SITPP-8* se regularon negativamente en condiciones de sequía. *SITPP-8* se reguló al alza en respuesta a salinidad (Figura 20 y Figura 21).

En maíz, la expresión del gen ABP9, que codifica un factor de transcripción bZIP capaz de unirse al motivo ABRE2 en el promotor Cat1, es inducida por ABA, H₂O₂, sequía y sal. La expresión constitutiva de ABP9 en Arabidopsis transgénica conduce a una tolerancia notablemente mejorada a múltiples estreses que incluyen sequía, alto contenido de sal, temperatura de congelación y estrés oxidativo (Zhang *et al*, 2011). ABRE2 se localizó solamente en *SITPS-3*.

Siete promotores *SITPS* y tres promotores *SITPP* (*SITPP-2*, *SITPP-3* y *SITPP-9*) contenían motivos CGTCA y TGACG involucrados en la capacidad de respuesta a MeJA (Liu *et al*, 2017). El motivo TGACG se ha identificado previamente como un sitio de unión para los factores transactivadores de bZIP. La mutagénesis dirigida al sitio en estos motivos TGACG abolió la expresión que responde a MeJA, identificándolos claramente como elementos que responden a MeJA (Rouster *et al*, 1997). En *SITPS-2* y *SITPS-3* hay múltiples motivos CGTCA y motivos TGACG, sugiriendo que estos genes podrían mediar la señal MeJA para regular el estrés abiótico. En caña de azúcar se han identificado motivos CGTCA y motivos TGACG en 8 de los 9 *ScTPS* (Xin *et al*, 2020).

Dos promotores *TPS* (*SITPS-7* y *SITPS-8*) y dos promotores *TPP* (*SITPP-1* y *SITPP-3*) contenían TGA-element y dos (*SITPS-4* y *SITPP-2*) presentaron

AuxRR-core, que podrían funcionar como un elemento de respuesta a las auxinas (Li, 2014).

Dos genes (*SITPS-6* y *SITPS-7*) tenían CARE; dos genes (*SITPP-1* y *SITPP-9*) tenían GARE-motif; ocho *TPS* y dos *TPP* (*SITPP-4* y *SITPP-7*) tenían P-box; dos genes *TPS* (*SITPS-4* y *SITPS-10*) y tres *TPP* (*SITPP-4*, *SITPP-7* y *SITPP-9*) tenían TATC-box, que son elementos de respuesta a giberelinas (Xu *et al*, 2018; Sutoh y Yamauchi, 2003).

Cinco promotores *TPS* (*SITPS-4*, *SITPS-5*, *SITPS-6*, *SITPS-9* y *SITPS-10*) y un *TPP* (*SITPP-8*) presentaron TCA-element; tres *TPS* (*SITPS-1*, *SITPS-2*, y *SITPS-10*) y cinco promotores *TPP* (*SITPP-2*, *SITPP-5*, *SITPP-6*, *SITPP-8* y *SITPP-9*) tenían TCA, elementos involucrados en la capacidad de respuesta del ácido salicílico (Fan *et al*, 2019).

Siete genes *TPS* (todos menos *SITPS-4*, *SITPS-5* y *SITPP-10*) y tres *TPP* (*SITPP-1*, *SITPP-8* y *SITPP-9*) tenían ERE, un elemento de respuesta al etileno. En tabaco transgénico transformado con construcciones de promotor SmCP-GUS encontraron que la región promotora que contiene el elemento sensible al etileno, ERE, confirió la inducción de etileno de la expresión de GUS, demostrando que cada ERE se une a proteínas nucleares de plántulas de 5 semanas tratadas con etefón y sin tratar, lo cual sugiere que diferentes factores de transcripción se unen a cada ERE en diferentes condiciones fisiológicas (Rawat *et al*, 2005).

O2-site, involucrado en la regulación del metabolismo de la zeína (Kaur *et al*, 2017) se localizó en *SITPS-3*, *SITPS-4* y *SITPP-8* (Figura 13).

La presencia de elementos de respuesta a las hormonas (ácido abscísico, jasmonato de metilo, ácido salicílico y etileno) en la mayoría de los promotores de *SITPS* y *SITPP*, indica que posiblemente podrían ensamblar diferentes vías de transducción de señal en la respuesta al estrés biótico y abiótico (Wolters y Jürgens, 2009). En particular, *SITPS-3*, *SITPS-2*, *SITPS-8* y *SITPP-8* contenían el mayor número de elementos de respuesta a hormonas.

	SITPS-1	SITPS-2	SITPS-3	SITPS-4	SITPS-5	SITPS-6	SITPS-7	SITPS-8	SITPS-9	SITPS-10		SITPP-1	SITPP-2	SITPP-3	SITPP-4	SITPP-5	SITPP-6	SITPP-7	SITPP-8	SITPP-9
ABRE		1	8	1	1	5	2	3		2	ABRE	1	2	1	1		1	1	10	
ABRE2			1								AT-rich element						1			
AT-rich element			1								AT-rich sequence	1								
AuxRR-core				1							AuxRR-core		1							
CARE						1	1				CGTCA-motif		2	1						1
CGTCA-motif		4	4	1	1		1	2		1	GARE-motif	1								1
ERE	2	3	3			2	2	4	1		ERE	4							1	1
O2-site			2	1							O2-site								1	
P-box	3		1	1	1	1		1	1	2	P-box			1				1		
TATC-box				1						1	TATC-box			1				1		1
TCA	1	1								1	TCA		1			1	1		1	1
TCA-element				1	1	1			2	1	TCA-element								1	
TGA-element							1	1			TGA-element	1		1						
TGACG-motif		4	4	1	1		1	2		1	TGACG-motif		2	1						1

Figura 13. Elementos reguladores *cis* relacionados con regulaciones hormonales en *SITPS/SITPPs*. Los números en las cajas de colores indican las repeticiones.

6.9.4 Elementos reguladores *cis* relacionados con el desarrollo celular

Los promotores de los genes *SITPSs* contenían un total de 8 CRE relacionadas con el desarrollo celular (CdCRE). Cada promotor contenía de 1 (*SITPS-1*, *SITPS-3* y *SITPS-8*) a 5 (*SITPS-6*) CdCRE (Figura 14). Todos los CdCRE encontrados en los promotores de los *SITPSs* se visualizaron en el Apéndice 20. Los promotores de los genes *SITPSs* contenían un total de 6 CRE relacionadas con el desarrollo celular (CdCRE). Cada promotor contenía 1 (*SITPP-4*, *SITPP-5* y *SITPP-8*) a 4 (*SITPP-2*, *SITPP-3* y *SITPP-7*) CdCRE (Figura 14). Todos los CRE relacionados con el desarrollo celular encontrados en los promotores de los *SITPPs* se visualizaron en el Apéndice 21.

Entre ellos, solo Unnamed__4, que puede funcionar con especificidad espacial y temporal en la expresión de DYT1 (Anther) (Zhou *et al*, 2017), se encuentra en todos los promotores de genes *SITPS* y *SITPP*, contribuyó con el 78.4% y 72.94% del total de CdCRE de cada familia. Cinco promotores de genes *TPS* (*SITPS-5*, *SITPS-6*, *SITPS-7*, *SITPS-9* y *SITPS-10*) y cinco *TPP* (*SITPP-2*, *SITPP-3*, *SITPP-6*, *SITPP-7* y *SITPP-9*) tenían un motivo AAGAA involucrado en la expresión específica de la semilla (Jaiswal *et al*, 2007).

GCN4-motif se halló en *SITPS-5*, *SITPS-7*, *SITPP-1*, *SITPP-3* y *SITPP-7*. GCN4-motif está altamente conservado en los promotores de genes de proteínas de

almacenamiento de semillas de cereales y juega un papel importante en el control de la expresión específica del endospermo. Este motivo es el sitio de reconocimiento de un factor transcripcional de zipper de leucina básico que pertenece al grupo de proteínas similares al Opaque-2 (O2) del maíz (Onodera *et al*, 2001).

Un elemento del ciclo circadiano con secuencia CAAAGATATC (Rahman *et al*, 2021) se encontró en dos promotores (*SITPS-2* y *SITPS-5*) y en tres promotores (*SITPP-2*, *SITPP-3* y *SITPP-7*).

AC-I y AC-II se localizaron en *SITPS-6*. Los factores de transcripción MYB. se unen a los cis de AC (AC-I y AC-II), regulan la expresión de genes biosintéticos de lignina (Zhang *et al*, 2018), y tienen funciones importantes en sus expresiones específicas de tejido (Osakabe *et al*, 2009).

CAT-box se localizó en los promotores de *SITPS-4*, *SITPS-10*, *SITPP-1* y *SITPP-9*. CCGTCC-box, en *SITPP-2*; dOCT, en *SITPS-6*. CAT-box, CCGTCC-box y dOCT, son elementos relacionado con la activación específica de meristemo (Zhou *et al*, 2017; Basyuni *et al*, 2018).

	SITPS-1	SITPS-2	SITPS-3	SITPS-4	SITPS-5	SITPS-6	SITPS-7	SITPS-8	SITPS-9	SITPS-10		SITPP-1	SITPP-2	SITPP-3	SITPP-4	SITPP-5	SITPP-6	SITPP-7	SITPP-8	SITPP-9
AAGAA-motif					1	1	1		5	1	AAGAA-motif		4	1			2	1		2
AC-I						1					CAT-box	1								3
AC-II						1					CCGTCC-box		2							
CAT-box				1						2	circadian		1	2				1		
circadian		1			1						GCN4 motif	1		1				1		
dOCT						1					Unnamed_4	8	12	7	1	3	5	9	4	13
GCN4 motif					1		1													
Unnamed_4	3	7	7	8	8	6	6	15	2	7										

Figura 14. Elementos reguladores *cis* relacionados con el desarrollo celular en *SITPS/SITPPs*. Los números en las cajas de colores indican las repeticiones.

6.9.5 Otros elementos reguladores *cis*

Los promotores de los genes *SITPSs* y *SITPPs* contenían contenía dos y nueve elementos del grupo de otros CREs (Figura 15). Todos los otros CREs en los *SITPSs* y *SITPPs* se visualizaron en los Apéndice 22 y Apéndice 23.

Unnamed_1, que presenta la secuencia GATTTAATTAA, se localizó en tres promotores *TPS* (*SITPS-5*, *SITPS-7* y *SITPS-9*) y en *SITPP-1*, donde podría

funcionar como un sitio de unión de proteína 60K (Xu *et al*, 2018). Por otra parte, *SITPP-8* contenía Unnamed_1, con la secuencia GGATTTTACAGT, un elemento implicado en la regulación negativa de fitocromos (Inaba *et al*, 1999).

A-box es un elemento *cis* que en perejil parece ser necesario, pero no suficiente para la activación del gen PAL (fenilalanina amoniaco-liasa) mediada por un elicitor o por la luz (Logemann *et al*, 1995) y se encontró en *SITPP-2*.

AP-1, elemento de respuesta a estrés (Peng *et al*, 2020) se localizó en *SITPP-9*; Box III, un sitio de unión a proteínas (Xiao *et al*, 2015) en *SITPP-2*; Unnamed_5, sitio de unión del factor SEF1 (Yu *et al*, 2014) en *SITPP-6*.

Unnamed_6, con secuencia taTAAATATct, se encontró en tres promotores *TPS* (*SITPS-5*, *SITPS-7* y *SITPS-9*) y en dos *TPP* (*SITPP-5* y *SITPS-6*); Unnamed_2 con secuencia AACCTAACCT y CTAG-motif con secuencia ACTAGCAGAA en *SITPP-9*, todos esos elementos de función desconocida (Lescot *et al*, 2002).

	SITPS-1	SITPS-2	SITPS-3	SITPS-4	SITPS-5	SITPS-6	SITPS-7	SITPS-8	SITPS-9	SITPS-10		SITPP-1	SITPP-2	SITPP-3	SITPP-4	SITPP-5	SITPP-6	SITPP-7	SITPP-8	SITPP-9	
Unnamed_6		1	1						1		A-box		1								
Unnamed_1					1		1		1		AP-1										1
											Box III		1								
											CTAG-motif										1
											Unnamed_2										1
											Unnamed_5						1				
											Unnamed_6					1	2				
											Unnamed_1	1									
											Unnamed_1										1

Figura 15. Otros elementos reguladores *cis* en *SITPS/SITPPs*. Los números en las cajas de colores indican las repeticiones.

Algunos elementos relacionados con el estrés (MYB, MYC, Myb, STRE, ARE), hormonas (ABRE, ERE) y respuesta a la luz (Box 4, G-box) existían ampliamente en las regiones promotoras de la mayoría de los genes *SITPS* y *SITPP*, lo que indica que los genes *TPS* y *TPP* pueden estar involucrados en las respuestas al estrés, las hormonas y la luz. Estos elementos también se localizaron en los genes *TPS* de la papa (Xu *et al.*, 2017).

6.10. Análisis de expresión *in silico* de genes *TPS/TPP* en jitomate

Se sabe que los genes *TPS* y *TPP* son importantes en la respuesta de las plantas a las tensiones ambientales. En este estudio aprovechamos los datos disponibles

de perfiles de expresión genética de jitomate en GEO (Gene Expression Omnibus), para analizar el conjunto completo de genes *SITPS/SITPP* bajo diferentes tensiones abióticas como sequía, calor, frío y salinidad.

Los valores FPKM de la expresión de cada uno de los genes en las diferentes condiciones se obtuvieron del GEO, se normalizaron y se utilizó el software TBtools para visualizar los niveles de expresión en un heat map. En los experimentos de estrés por frío y por salinidad, al contar con los datos de log₂ y Pvalue, se pudieron construir gráficas de Volcán en TBtools.

6.10.1 Niveles de expresión de los genes *SITPS/SITPP* en condiciones de sequía

En el análisis de expresión en respuesta a sequía se estudiaron dos conjuntos de datos, en los que los genes *SITPSs* y *SITPPs* demostraron tres tipos de niveles de expresión, incluyendo baja expresión, expresión media y media a alta.

Se analizó el nivel de expresión de *SITPS* en plántulas de jitomate cv. M82 en respuesta a estrés por sequía (Wang *et al*, 2020), sometidas a diferentes condiciones: control (CK), estrés por sequía durante 1 día (D-1d), 2 días (D-2d), 3 días (D-3d), 4 días (D-4d), 5 días (D-5d) y sequía con recuperación (D-Rec), se encontró que *SITPS-2* y *SITPS-10* tuvieron una baja expresión y *SITPS-7* presentó una expresión media en todas las condiciones, es decir, no cambiaban en respuesta a sequía. *SITPS-1* y *SITPS-9* mostraron una expresión alta en comparación con los demás genes. *SITPS-1* reveló su expresión máxima un día después del inicio de las condiciones de estrés. *SITPS-9* se expresó más en el tratamiento de estrés por sequía durante 4 días. *SITPS-4*, *SITPS-5* y *SITPS-6* se expresaron de la siguiente manera: en CK se expresaron muy poco, iban incrementando su expresión con el aumento de sequía, alcanzando un máximo en el día 2 o día 3 y finalmente en el tratamiento de recuperación mostraron un nivel menor al del CK. *SITPS-3* y *SITPS-8* se caracterizaron porque su expresión disminuyó conforme aumentaron los días de sequía y finalmente la expresión incrementó en el tratamiento de sequía con recuperación. (Figura 16A). En las hojas de plantas 'Micro-TOM' (Zhao *et al*, 2020) evaluadas en diferentes condiciones: control (CK), estrés por sequía durante 3h (CK-3h) y 6h (CK-6h), se

encontró que *SITPS-2* y *SITPS-9* revelaron una baja expresión y similar en todos los tratamientos. De la misma forma *SITPS-7* y *SITPS-10* presentaron una baja expresión en la condición control, pero que incrementó ligeramente en los tratamientos de sequía. *SITPS-1* y *SITPS-5* por su parte mostraron expresión media en las tres condiciones. *SITPS-4* y *SITPS-6* tuvieron su expresión máxima en la condición de 3h de sequía y a las 8h disminuyó su expresión. *SITPS-3* y *SITPS-8* demostraron un patrón de expresión similar, se expresaron más en la condición control en comparación con los tratamientos de sequía (Figura 16B). En resumen, *SITPS-1* y *SITPS-9*, genes paralogos pertenecientes al grupo II-1, incrementan fuertemente su expresión, pero tardan días por eso no aumentaron en el experimento con plantas 'Micro-TOM'. *SITPS-5* también aumenta su expresión después de un día de sequía. En cambio, *SITPS-4* y *SITPS-6* incrementaron rápidamente su expresión durante las primeras horas de estrés por sequía y disminuyó después de 4 días. Por su parte, *SITPS-3* y *SITPS-8* disminuyeron su expresión en las primeras horas, siguió disminuyendo con el paso de los días y finalmente en el tratamiento de sequía con recuperación aumentó. *SITPS-2*, *SITPS-7* y *SITPS-10* no cambiaron en respuesta a la sequía. En lo que toca a los genes SITPPs, en plántulas de jitomate cv. M82 (Wang *et al*, 2020), se encontró que tres genes (*SITPP-5*, *SITPP-6* y *SITPP-7*) mostraron una baja expresión en la mayoría de las condiciones evaluadas, e incluso nula en algunas condiciones en el caso de *SITPP-5* y *SITPP-7*. Dos genes (*SITPP-2* y *SITPP-4*) revelaron un nivel de expresión similar en la condición control, disminuyeron su expresión a medida que aumentaban los días de sequía, con la diferencia de que *SITPP-2* se expresó más en la condición de sequía con recuperación. *SITPP-3* mostró la expresión más alta en comparación con todos los demás genes *SITPP*, aumentó en los primeros dos días de sequía, donde alcanzó su máximo, de ahí disminuyó su expresión. *SITPP-8* también demostró regulación negativa y aumentó su expresión en el tratamiento de sequía con recuperación. Por otro lado, *SITPP-1* y *SITPP-9* estaban regulados al alza, con la diferencia de que *SITPP-9* incrementó su expresión en sequía con recuperación mientras que *SITPP-1* no (Figura 16C). En las plantas 'Micro-TOM'

(Zhao *et al*, 2020), *SITPP-5* no se expresó, *SITPP-6* y *SITPP-7* mostraron baja expresión y no cambiaban en respuesta a sequía. Tres genes (*SITPP-2*, *SITPP-4* y *SITPP-8*) revelaron regulación descendente, los primeros dos con expresión media a alta, mientras que *SITPP-8* reveló baja expresión. *SITPP-1*, *SITPP-3* y *SITPP-9* estaban regulados al alza (Figura 16D).

En resumen, los dos experimentos coincidieron en que tres genes (*SITPP-5*, *SITPP-6* y *SITPP-7*) demostraron una baja o nula expresión y no cambiaban en condiciones de estrés por sequía, los genes *SITPP-2*, *SITPP-4* y *SITPP-8* estaban regulados negativamente, y tres genes (*SITPP-1*, *SITPP-3* y *SITPP-9*) se mostraron regulados positivamente.

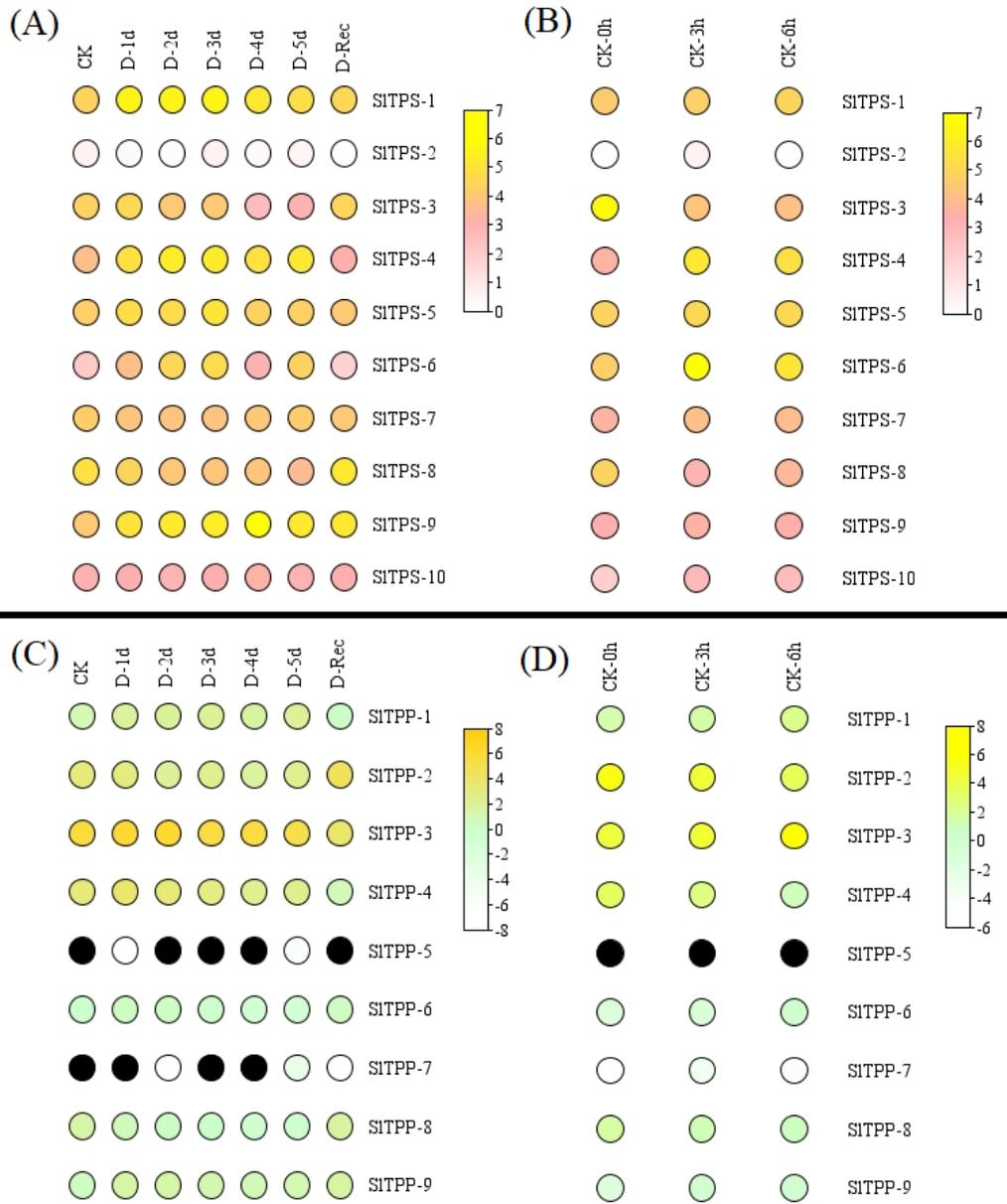


Figura 16. Patrón de expresión de genes *SITPS* y *SITPP* en respuesta a sequía. (A) *SITPS* en plántulas de jitomate cv. M82, (B) *SITPS* en plántulas de jitomate Micro TOM, (C) *SITPP* en plántulas de jitomate cv. M82, (D) *SITPP* en plántulas de jitomate Micro TOM. La clave de color representa la abundancia de transcripción relativa de los genes. El color negro representa nula expresión.

6.10.2 Niveles de expresión de los genes *SITPS/SITPP* en condiciones de calor

En el análisis de expresión en respuesta a calor se estudiaron dos conjuntos de datos. En plántulas de jitomate cv. M82 (Wang *et al*, 2020), expuestas a diferentes condiciones: control (Heat-0h), estrés por calor durante 2h (Heat-2h), 4h (Heat-4h), 12h (Heat-12h), 24h (Heat-24h) y estrés por calor con recuperación (Heat-Rec), se encontró que *SITPS-5* no cambió en ninguna condición y reveló una expresión media. *SITPS-2* y *SITPS-10* mostraron un nivel bajo y medio, respectivamente, estaban regulados negativamente por el calor. En *SITPS-7* con un nivel de expresión medio, la regulación negativa se notó un poco más. De igual manera en *SITPS-1* y *SITPS-8* se encontró disminución de la expresión a medida que aumentaba la duración del estrés por calor y un ligero aumento en el calor con recuperación. Mientras tanto en *SITPS-9* la regulación negativa se mostró hasta las 24h y en el calor con recuperación incrementó un poco la expresión. *SITPS-3* no mostró cambios en el estrés por calor, pero disminuyó gravemente su expresión en el calor con recuperación. Los niveles de expresión de *SITPS-4* y *SITPS-6* estaban regulados positivamente, mostraron su máxima expresión a las 24h y la mínima en el tratamiento de calor con recuperación (Figura 17A). En los frutos de jitomate 'Micro-TOM' (Liu *et al*, 2021), bajo las siguientes condiciones: control (MT-0H), estrés por calor durante 24h (MT-24h), 48h (MT-24h), y 96h (MT-24h), se observó que cuatro genes (*SITPS-1*, *SITPS-5*, *SITPS-7* y *SITPS-8*) demostraron un patrón similar, una expresión media a alta y que no cambiaba en las condiciones de calor. *SITPS-2* y *SITPS-9* se mostraron regulados negativamente en el estrés por calor. La expresión de *SITPS-1* y *SITPS-3* disminuyó en los tratamientos de 24h y 48 h comparada con la del control (0h) e incrementó un poco a las 96h. *SITPS-3* destaca por su alta expresión en la condición control. *SITPS-4*, *SITPS-6* y *SITPS-10* parecen estar regulados positivamente en condiciones de estrés por calor (Figura 17B). De los análisis anteriores se infiere que, *SITPS-2* y *SITPS-9* están regulados negativamente por el calor, mientras que *SITPS-4*, *SITPS-6* y *SITPS-10* positivamente.

Los niveles de expresión de *SITPS-1*, *SITPS-7* y *SITPS-8* se mostraron regulados negativamente en el experimento con plántulas de jitomate cv. M82, mientras que en jitomate Micro-TOM no, esto podría deberse a que son genotipos diferentes. Lo mismo podría explicar porque *SITPS-10* disminuyó su expresión durante las primeras horas en jitomate Micro-TOM, en cambio en jitomate cv. M82 estuvo regulado positivamente.

De igual modo se analizó la expresión de los genes *SITPPs* en las condiciones de calor, en plántulas de jitomate cv. M82 (Wang *et al*, 2020), se encontró que *SITPP-5* no cambió en los tratamientos de calor y mostró una baja e incluso nula expresión en algunos tratamientos. *SITPS-7* aumentó un poco su expresión en las primeras 4h, después disminuyó hasta que ya no se detectó. *SITPP-3* demostró una expresión media, que no cambió en respuesta a calor. *SITPP-1*, *SITPP-2*, *SITPP-6* y *SITPP-8* estaban regulados negativamente, *SITPP-2* reveló la máxima expresión durante las primeras 4h de calor. *SITPP-4* y *SITPS-9* estaban regulados positivamente, *SITPS-4* alcanzó su máxima expresión a las 24h y disminuyó en el tratamiento de calor con recuperación (Figura 17C).

En los frutos de jitomate 'Micro-TOM' (Liu *et al*, 2021) no se observó expresión de *SITPP-5*, mientras que *SITPP-2* no cambió en los diferentes tratamientos de estrés por calor y reveló la expresión más alta. La expresión de *SITPP-1* y *SITPP-4* disminuyó ligeramente al aumentar las horas de calor. *SITPP-6* y *SITPP-7* demostraron una baja expresión y estaban regulados a la baja. *SITPP-1* y *SITPP-4* también disminuyeron su expresión y revelaron un nivel de expresión medio, *SITPP-3* mostró un nivel de expresión media a alta y estuvo regulada negativamente después de 96h. *SITPP-8* aumentó su expresión a las 24h y luego disminuyó. *SITPP-9* aumentó su expresión a las 24h. (Figura 17D).

Por lo tanto, de los dos análisis podemos concluir que se encontraron grandes diferencias entre los patrones de expresión en los dos genotipos de jitomate, como en M82 donde *SITPP-4* aumentó su expresión durante las primeras 24h mientras que en Micro TOM disminuyó. Así mismo, en M82 *SITPP-2* estaba regulado negativamente y en Micro TOM no cambió su expresión. Además, *SITPP-8* estuvo regulado a la baja en M82 mientras que en Micro TOM aumentó

su expresión a las 24h y luego disminuyó. Por consiguiente, sólo se tomaron en cuenta los patrones de expresión que coinciden en ambos experimentos: *SITPP-1*, *SITPP-3* y *SITPS-6* estaban regulados negativamente, y *SITPP-9* mostró regulación positiva.

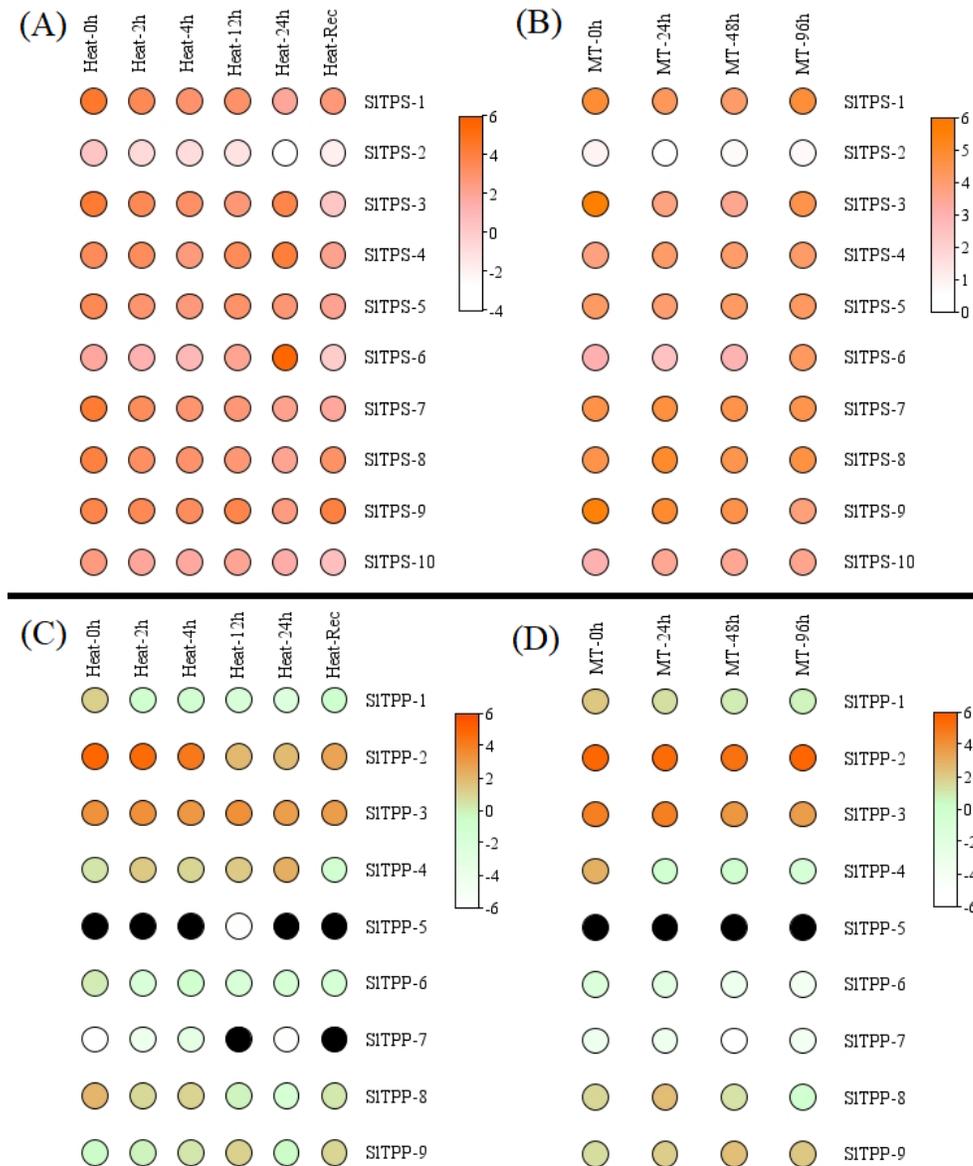


Figura 17. Patrón de expresión de genes *SITPS* y *SITPP* en respuesta a calor. (A) *SITPS* en plántulas de jitomate cv. M82, (B) *SITPS* en plántulas de jitomate Micro TOM, (C) *SITPP* en plántulas de jitomate cv. M82, (D) *SITPP* en plántulas de jitomate Micro TOM. La clave de color representa la abundancia de transcripción relativa de los genes. El color negro representa nula expresión.

6.10.3 Niveles de expresión de los genes *SITPS/SITPP* en condiciones de frío

En el análisis de expresión en respuesta a frío de hojas de plantas de jitomate 'Micro-TOM' (Zhao y Xu, 2020), expuestas a diferentes condiciones: control (CK0), estrés por frío durante 4h (CK4) y 12h (CK12), se encontró que *SITPS-2* y *SITPS-10* demostraron una baja expresión; *SITPS-4*, *SITPS-7* y *SITPS-9*, una expresión media y los demás genes una expresión alta. Sin embargo, notamos que la mayoría de los genes no cambian su expresión en tratamientos de frío. *SITPS-4*, *SITPS-6*, *SITPS-9* y *SITPS-10* disminuyen ligeramente su expresión conforme se incrementan las horas de frío. *SITPS-8* es el único que mostró una ligera regulación positiva (Figura 18A).

Para confirmar si el patrón de expresión observado en el mapa de calor es significativo se realizaron gráficas de volcán (Figura 19). Observamos que ningún gen *SITPS* cambiaba en respuesta al frío (Figura 19A-C) y que *SITPS-8* fue el gen que estuvo más cerca de ser regulado positivamente, pero no fue significativo estadísticamente hablando.

Analizando la expresión de *SITPPs* en el mismo conjunto de datos (Zhao y Xu, 2020) se encontró que *SITPP-5* no se expresa en ninguna condición evaluada, *SITPP-7* tampoco se detectó en el tratamiento de estrés por frío durante 4 horas (CK4), mientras que en el testigo y en el estrés por frío durante 12 horas (CK12) presentó una baja expresión. Los genes *SITPP-1* y *SITPP-6* demostraron una baja expresión. *SITPP-4*, *SITPP-8* y *SITPP-9* tuvieron una expresión media. *SITPP-2* y *SITPP-3* revelaron una expresión alta (Figura 18B). Los diagramas de volcán mostraron que comparando CK0-CK4, *SITPP-4* y *SITPP-6* estaban regulados positivamente (Figura 19D), mientras que en la comparación de CK4-CK12, *SITPP-6* estuvo regulado negativamente (Figura 19E), y comparando CK0-CK12, *SITPP-4* se reguló positivamente (Figura 19F), por lo que en conclusión *SITPP-4* estaba regulado al alza.

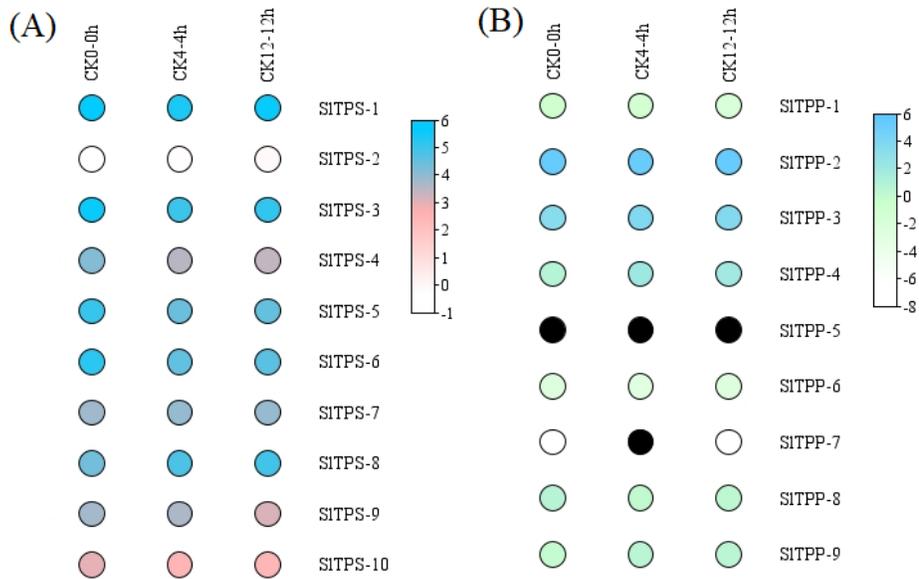


Figura 18. Patrón de expresión de genes *SITPS* y *SITPP* en respuesta a frío. (A) *SITPS* en plántulas de jitomate Micro TOM, (B) *SITPP* en plántulas de jitomate Micro TOM. La clave de color representa la abundancia de transcripción relativa de los genes. El color negro representa nula expresión.

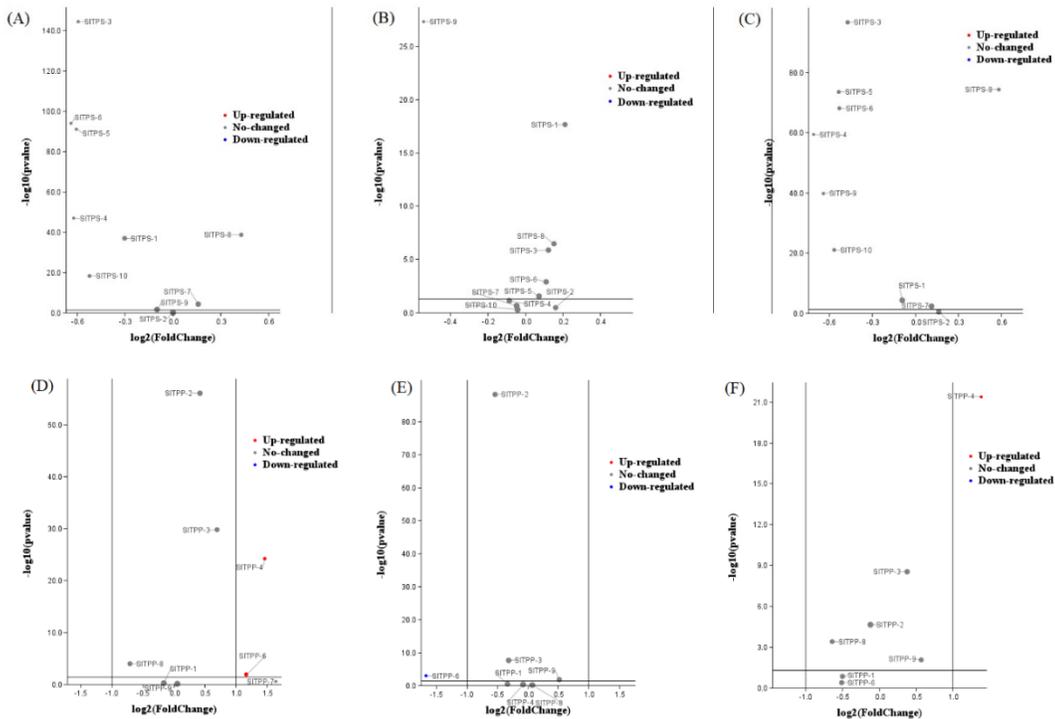


Figura 19. Diagramas de volcán de expresión de *SITPS/SITPP* en respuesta a frío. (A) *SITPS* en CK0-CK4, (B) *SITPS* en CK4-CK12, (C) *SITPS* en CK0-CK12, (D) *SITPP* en CK0-CK4, (E) *SITPP* en CK4-CK12, (F) *SITPP* en CK0-CK12. En X se muestra la magnitud del cambio (Fold Change) y en Y la significancia estadística (P value).

6.10.4 Niveles de expresión de los genes *SITPS/SITPP* en condiciones de salinidad

En el análisis de perfiles de expresión de ARNm en hojas de jitomate (Zhao y Xu, 2020) bajo las siguientes condiciones: testigo (CK0), estrés por salinidad durante 2h (CK2) y salinidad durante 8h (CK8) observamos que *SITPS-2* y *SITPS-9* demostraron una baja expresión; *SITPS-10*, *SITPS-7* y *SITPS-8*, una expresión media; y el resto de los genes revelaron una expresión de media a alta. Además, se encontró que los genes *SITPS-7* y *SITPS-10* no cambiaban su expresión. *SITPS-1*, *SITPS-2*, *SITPS-4*, *SITPS-5* y *SITPS-6* estaban regulados positivamente, siendo este último el que mostró la más alta expresión. *SITPS-8* se mostró regulado negativamente al estrés por salinidad (Figura 20A). Estas observaciones se confirmaron al dibujar los diagramas de volcán (Figura 21A-C), que revelaron que en comparación con los niveles de expresión del CK0, en CK2 *SITPS-6* aumentó su expresión, mientras que *SITPS-9* estuvo regulado a la baja. Comparando los niveles de expresión de CK2 y CK8, los genes *SITPS-1*, *SITPS-4*, *SITPS-6* y *SITPS-9* estaban regulados al alza, mientras que *SITPS-3* y *SITPS-8* disminuyeron su expresión. Haciendo la comparación de la expresión CK0 y CK8, observamos que cinco genes (*SITPS-1*, *SITPS-2*, *SITPS-4*, *SITPS-5* y *SITPS-6*) estaban regulados positivamente y *SITPS-8*, negativamente.

También se analizó la expresión de *SITPPs* en el mismo conjunto de datos (Zhao y Xu, 2020), no se detectó la expresión de *SITPP-5*. Por su parte *SITPP-6*, *SITPP-7* y *SITPP-9* demostraron baja expresión; *SITPP-8*, una media; y *SITPP-1*, *SITPP-2*, *SITPP-3* y *SITPP-4* una alta expresión (Figura 20B). Se dibujaron los diagramas de volcán para los *SITPPs*, encontramos que comparando los niveles de expresión de CK0 y CK2, varios genes (*SITPP-1*, *SITPP-2*, *SITPP-3*, *SITPP-7* y *SITPP-8*) estaban regulados al alza (Figura 21D), siendo casi los mismos genes los que se revelaron regulados negativamente comparando CK2 y CK8, a excepción de *SITPP-8* que ya no cambio, y a los demás genes se sumó *SITPP-4* que parece estar regulado negativamente también (Figura 21E). Contrastando el nivel de expresión en CK0 con el de CK8, notamos que los tres genes (*SITPP-*

1, *SITPP-3* y *SITPP-8*) estaban regulados significativamente al alza y *SITPP-4* se encontró regulado a la baja (Figura 21F).

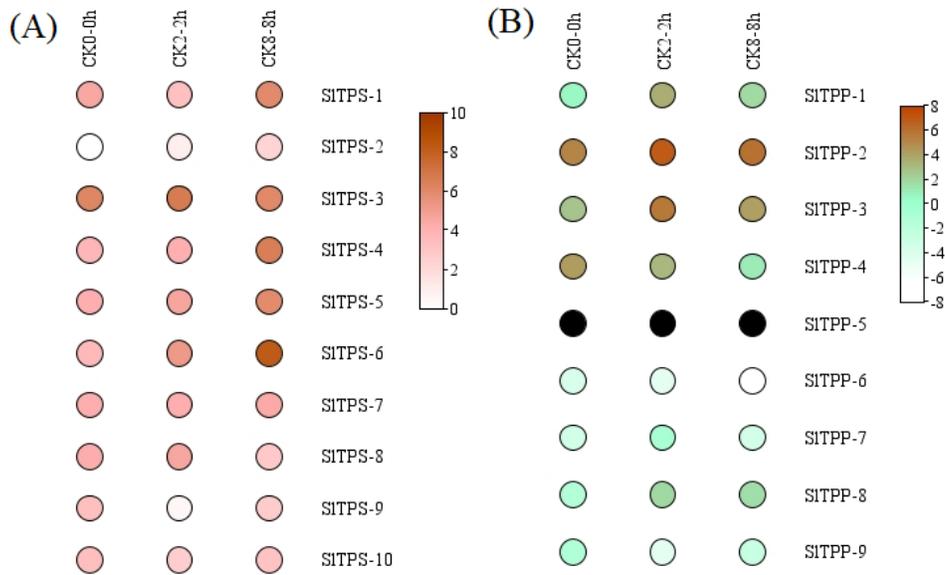


Figura 20. Patrón de expresión de genes *SITPS* y *SITPP* en respuesta a salinidad. (A) *SITPS* en hojas de jitomate, (B) *SITPP* en hojas de jitomate. La clave de color representa la abundancia de transcripción relativa de los genes. El color negro representa nula expresión.

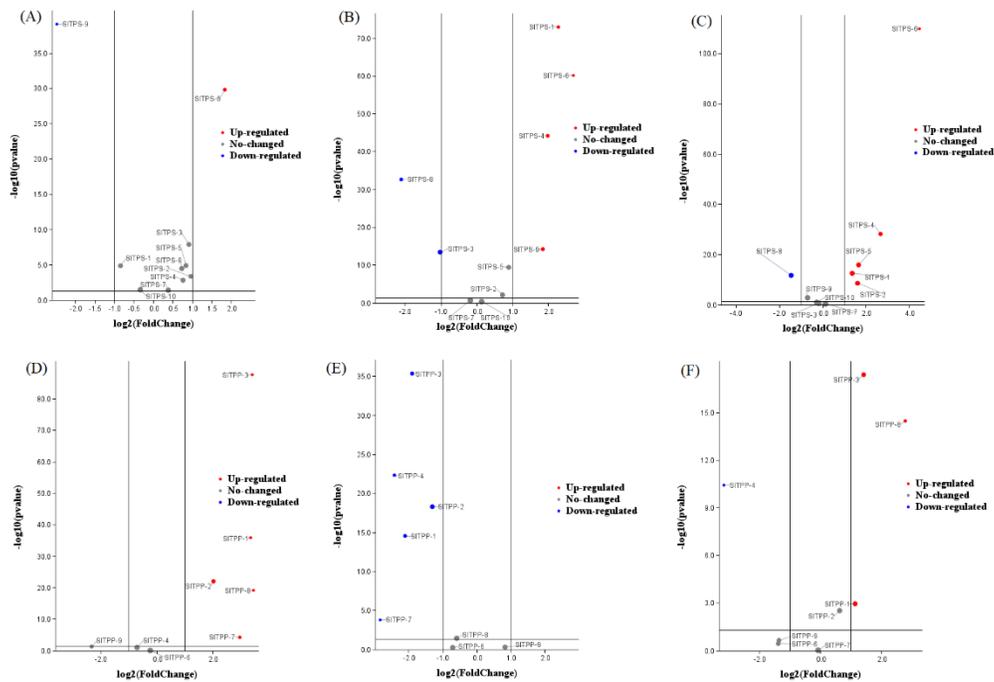


Figura 21. Diagramas de volcán de expresión de *SITPS/SITPP* en respuesta a salinidad. (A) *SITPS* en CK0-CK2, (B) *SITPS* en CK2-CK8, (C) *SITPS* en CK0-CK8, (D) *SITPP* en CK2-CK8, (E) *SITPP* en CK0-CK8, (F) *SITPP* en CK0-CK2.

Como se ha mencionado anteriormente, en diversos estudios han demostrado que los genes *TPS* y *TPP* desempeñan un papel importante en las plantas en respuesta a diversos tipos de estrés como la sequía, el calor, el frío y la salinidad. En este estudio, los perfiles de expresión de los genes *TPS* y *TPP* bajo diferentes tipos de estrés revelaron que los genes *TPS* del jitomate estaban ampliamente involucrados en la respuesta a estrés por sequía, calor y salinidad (Figura 22), mientras que los genes *TPP* estaban involucrados mayormente en la respuesta a estrés por sequía y salinidad, y en menor medida en la respuesta a calor (Figura 23) y frío. Es interesante notar que cinco de los genes *SITPS* estuvieron fuertemente regulados al alza durante el estrés por sequía, lo que indica que esos genes podrían estar involucrados principalmente en la vía biológica del estrés por sequía. Dos genes *SITPS* (*SITPS-4* y *SITPS-6*) mostraron el mismo patrón de expresión bajo el tratamiento de sequía, calor y salinidad, lo que sugiere que estos genes *SITPS* se coexpresaron en respuesta al estrés por sequía, calor y salinidad. Sin embargo, los perfiles de expresión génica de genes *SITPP* que respondían a las diferentes condiciones de estrés por lo general tendían a ser diferentes: algunos genes regulados por sequía se regularon negativamente en el estrés por calor, lo que indicó que dos conjuntos de genes *SITPP* estaban involucrados en la respuesta a la sequía y el estrés por calor, respectivamente, y solamente *SITPP-9* estaba regulado positivamente por la sequía y el calor. Estos resultados también indican que las vías de señalización en las respuestas al estrés abiótico de las plantas son sistemas muy complicados.

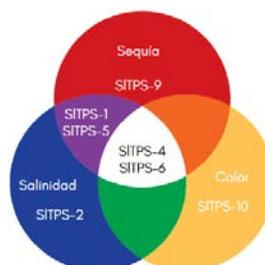


Figura 22. Diagrama de Venn que muestra la expresión de genes *SITPS* bajo estrés por sequía, calor y salinidad. Entre ellos, se detectaron cinco genes *SITPS* regulados positivamente por sequía, cinco por alta salinidad y tres por calor, y se observó que dos genes *SITPS* respondían notablemente a las tres tensiones.

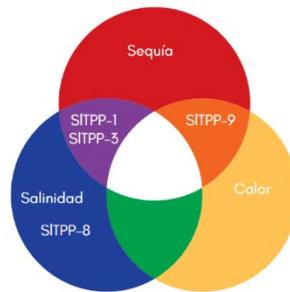


Figura 23. Diagrama de Venn que muestra la expresión de genes *SITPP* bajo estrés por sequía, calor y salinidad. Entre ellos, se detectaron tres genes *SITPP* regulados positivamente por sequía, tres por alta salinidad sequía y dos por calor.

VII. CONCLUSIONES

En este análisis *in silico* identificamos diez genes *SITPS* y nueve *SITPP* en el genoma de jitomate y los caracterizamos sistemáticamente utilizando diferentes herramientas bioinformáticas para explorar su papel en la respuesta al estrés abiótico, principalmente. El análisis de la estructura genética mostró que *SITPS-2* y *SITPS-8*, genes de clase I, podrían ser los genes ancestrales. Todos los genes *TPS* y *TPP* han evolucionado principalmente por selección purificadora. Además, cuatro pares de genes *SITPS* y dos pares de genes *SITPP* se expandieron por duplicación segmental. El análisis de elementos reguladores *cis* mostró que los genes *SITPS* y *SITPP* podrían desempeñar un papel importante en la respuesta al estrés, las hormonas, el desarrollo y la fotosíntesis. El análisis de datos de RNAseq reveló que la mayoría de los genes *TPS* podrían ser inducidos por diferentes tipos de estrés, como la sequía, la salinidad y el calor, y aproximadamente la mitad de los genes *TPP* pueden estar involucrados en la tolerancia al estrés. En particular, *SITPS-4*, *SITPS-6*, *SITPS-1*, *SITPS-1*, *SITPS-5*, *SITPP-1*, *SITPP-3* y *SITPP-9* podría ser genes de resistencia al estrés abiótico de amplio espectro. En conjunto, la información obtenida en este estudio proporcionará una mejor comprensión de las estructuras y funciones de los *SITPSs* y *SITPPs*, muchos de los cuales pueden ser genes candidatos potenciales para el desarrollo de variedades de jitomate con mayor tolerancia al estrés a través del mejoramiento asistido por marcadores moleculares, transformación o edición genética.

VIII. REFERENCIAS BIBLIOGRÁFICAS

- Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell*, 15, 63-78.
- Acosta-Pérez, P., Camacho Zamora, B. D., Espinosa-Sánchez, E. A., Gutiérrez-Soto, G., Zavala-García, F., Abraham-Juárez, M. J., Sinagawa-García, S. R. (2020). Characterization of Trehalose-6-phosphate Synthase and Trehalose-6-phosphate Phosphatase Genes and Analysis of its Differential Expression in Maize (*Zea mays*) Seedlings under Drought Stress. *Plants*, 9, 315.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol.*, 215, 403–410.
- Avonce, N., Leyman, B., Mascorro-Gallardo, J., Van Dijck, P., Thevelein, J. M., Iturriaga, G. (2004). The Arabidopsis Trehalose-6-P Synthase AtTPS1 Gene Is a Regulator of Glucose, Abscisic Acid, and Stress Signaling. *Plant Physiology*, 136(3), 3649-3659.
- Avonce, N., Mendoza-Vargas, A., Morett, E., Iturriaga, G. (2006). Insights on the Evolution of Trehalose Biosynthesis. *BMC Evol. Biol.*, 6, 109.
- Baena-Gonzalez, E., Rolland, F., Thevelein, J., Sheen, J. (2007). A central integrator of transcription networks in plant stress and energy signalling. *Nature*, 448, 938–942.
- Bailey, T. L. y Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, 14, 48-54.
- Baranwal, V., Negi, N., Khurana, P. (2017). Auxin Response Factor Genes Repertoire in Mulberry: Identification, and Structural, Functional and Evolutionary Analyses. *Genes*, 8, 202.
- Barraza, A., Contreras-Cubas, C., Estrada-Navarrete, G., Reyes, J., Juárez-Verdayes, M., Avonce, N., Sánchez, F. (2016). The Class II Trehalose 6-phosphate Synthase Gene PvTPS9 Modulates Trehalose Metabolism in Phaseolus vulgaris Nodules. *Front. Plant Sci.*, 7, 1589.
- Basyuni, M., Wati, R., Baba, S., Oku, H. (2018). Characterization of cycloartenol synthase KcCAS promoter region from mangrove plant (*Kandelia obovata* Sheue, H. Y. Liu, J. Yong). *AIP Conference Proceedings*, 2002, 020040.
- Baxevanis, A. D. y Ouellette, F. (2001). Bioinformatics. A Practical Guide to the Analysis of Genes and Proteins. *Wiley Interscience*, 341.

- Becerra, B. J., Narváez-Vásquez, j., Pillitteri, L. J., Orozco-Cárdenas, M., Tang, L., Lovatt, C. J. (2016). Identification of cis-regulatory elements related to water-deficit and low-temperature stress within the promoter of *Citrus sinensis* APETALA1. *Citrus Research & Technology*, 37(2), 138-146.
- Belocopitow, E. y Maréchal, L. R. (1970). Trehalose phosphorylase from *Euglena gracilis*. *Biochim Biophys Acta*, 198(1), 151–154.
- Biłas, R., Szafran, K., Hnatuszko-Konka, K. *et al.* (2016). Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tiss Organ Cult*, 127, 269–287.
- Bordoli, L. y Schwede, T. (2012). Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods in molecular biology (Clifton, N. J.)*. 857, 107–136.
- Briesemeister, S., Rahnenführer, J., Kohlbacher, O. (2010). YLoc – an interpretable web server for predicting subcellular localization. *Nucleic Acids Research*, 38, W497-W502.
- Cautain, B., Hill, R., de Pedro, N., Link, W. (2015). Components and regulation of nuclear transport processes. *The FEBS Journal*, 282(3), 445-462.
- Chary, S. N., Hicks, G. R., Choi, Y. G., Carter, D., Raikhel, N. V. (2008). Trehalose-6-Phosphate Synthase/Phosphatase Regulates Cell Shape and Plant Architecture in Arabidopsis. *Plant Physiology*, 146(1), 97-107.
- Chen, C., X. R., Chen, H., He, Y. (2018). TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv*, 289660.
- Chen, O. y Haddad, G. G. (2004). Role of trehalose phosphate synthase and trehalose during hypoxia: from flies to mammals. *The Journal of Experimental Biology*, 207, 3125-3129.
- Chen, Z. y Lou, J. (2017). Identification and expression of the trehalose-6-phosphate synthase gene family members in tomato exposed to different light spectra. *Arch Biol Sci.*, 69(1), 93-101.
- Combet, C. B., Geourjon, C., Deléage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem Sci*, 25, 147–150.
- Cortina, C. y Culiñez-Macia, F. A. (2005). Tomato abiotic stress enhanced tolerance by trehalose biosynthesis. *Plant Science*, 169, 75–82.
- Dan, Y., Niu, Y., Wang, C., Yan, M., Liao, W. (2021). Genome-wide identification and expression analysis of the trehalose-6-phosphate synthase (TPS) gene family in cucumber (*Cucumis sativus* L.). *PeerJ*, 9, e11398.

- Delorge, I., Figueroa, C., Feil, R., Lunn, J., Van Dijck, P. (2015). Trehalose-6-phosphate synthase 1 is not the only active TPS in *Arabidopsis thaliana*. *Biochem. J.*, 466, 283–290.
- Doyle, M. (2021). Visualization of RNA-Seq results with Volcano Plot. Obtenido de Galaxy Training Materials: <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/rna-seq-viz-with-volcanoplot/tutorial.html>
- Du, L., Qi, S., Ma, J., Xing, L., Fan, S., Zhang, S., Han, M. (2017). Identification of TPS family members in apple (*Malus x domestica* Borkh.) and the effect of sucrose sprays on TPS expression and floral induction. *Plant Physiol Biochem.*, Nov (120), 10-23.
- Eastmond, P., van Dijken, A., Spielman, M., Kerr, A., Tissier, A., Dickinson, H., Graham, I. (2006). Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation. *Plant J.*, 29(2), 225-235.
- Elbein, A. D., Pan, Y. T., Pastuszak, I., Carroll, D. (2003). New insights on trehalose: a multifunctional molecule. *Glycobiology*, 13(4), 17–27.
- Emanuelsson, O., Nielsen, H., von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.*, 8(5), 978-984.
- Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.*, 300, 1005-1016.
- EMBL-EBI. (2021). Biological interpretation of gene expression data. Obtenido de Functional Genomics II: <https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/biological-interpretation-of-gene-expression-data-2/>
- Fan, S., Liu, A., Zhang, Z., Zou, X., Jiang, X., Huang, J., Shang, H. (2019). Genome-Wide Identification and Expression Analysis of the Metacaspase Gene Family in *Gossypium* Species. *Genes*, 10(7), 527.
- Fichtner, F., Olas, J. J., Feil, R., Watanabe, M., Krause, U., Hoefgen, R., Lunn, J. E. (2020). Functional Features of TREHALOSE-6-PHOSPHATE SYNTHASE1, an Essential Enzyme in Arabidopsis. *Plant Cell*, 32(6), 1949-1972.
- Figueroa, C. y Lunn, J. (2016). A tale of two sugars: Trehalose 6-phosphate and sucrose. *Plant Physiol.*, 172, 7–27.

- Gallo, C. T., Bordoli, L., Schwede, T. (2014). Assessment of ligand binding site predictions in CASP10. *Proteins*, 2, 154–163.
- Gao, Y., Zan, X. L., Wu, X. F., Yao, L., Chen, Y. L., Jia, S. W., Zhao, K. J. (2014). Identification of fungus-responsive cis-acting element in the promoter of *Brassica juncea* chitinase gene, BjCHI1. *Plant Science*, 215, 190-198.
- Gao, Z., Zhao, R., Ruan, J. (2013). A genome-wide cis-regulatory element discovery method based on promoter sequences and gene co-expression networks. *BMC genomics*, 14 Suppl 1 (Suppl 1), S4.
- Garg, A., Kim, J., Owens, T., Ranwala, A., Choi, Y., Kochian, L., Wu, R. (2002). Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. *Proc. Natl. Acad. Sci.*, 99, 15898–15903.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucl. Acids Res.*, 31, 3784–3788.
- Ge, L., Chao, D., Shi, M., Zhu, M., Gao, J., Lin, H. (2008). Overexpression of the trehalose-6-phosphate phosphatase gene OsTPP1 confers stress tolerance in rice and results in the activation of stress responsive genes. *Planta*, 228, 191–201.
- Geourjon, C. y Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*, 11, 681–684.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*, 40, D1178–D1186.
- Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 31(8), 1296-1297.
- Hu, W. (2016). Genome-wide characterization and analysis of bZIP transcription factor gene family related to abiotic stress in cassava. *Sci. Rep.*, 6, 22783.
- Hu, X., Wu, Z. D., Luo, Z. Y., Burner, D. M., Pan, Y. B., Wu, C. W. (2020). Genome-Wide Analysis of the Trehalose-6-Phosphate Synthase (TPS) Gene Family and Expression Profiling of ScTPS Genes in Sugarcane. *Agronomy*, 10, 969.
- Hurst, L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet*, 18, 486–487.

- Inaba, T., Nagano, Y., Sakakibara, T., Sasaki, Y. (1999). Identification of a cis-regulatory element involved in phytochrome down-regulated expression of the pea small GTPase gene *pra2*. *Plant physiology*, 120(2), 491–500.
- Iturriaga, G., Gaff, D. F., Zentella, R. (2000). New desiccation-tolerant plants, including a grass, in the central high-lands of Mexico, accumulate trehalose. *Aust J Bot*, 48, 153–158.
- Jaiswal, R., Nain, V., Abdin, M. Z., Kumar, P. A. (2007). Isolation of pigeon pea (*Cajanus cajan* L.) legumin gene promoter and identification of conserved regulatory elements using tools of bioinformatics. *Indian Journal of Biotechnology*, 6, 495-503.
- Jensen, R. A. (2001). Orthologs and paralogs - we need to get it right. *Genome Biol*, 2, interactions1002.1.
- Jiang, W., Fu, F. L., Zhang, S. Z., Wu, L., Li, W. C. (2010). Cloning and characterization of functional trehalose-6-phosphate synthase gene in maize. *J Plant Biol.*, 53(2), 134–141.
- Jiang, Y., Duan, Y., Yin, J., Ye, S., Zhu, J., Zhang, F., Luo, K. (2014). Genome-wide identification and characterization of the Populus WRKY transcription factor family and analysis of their expression in response to biotic and abiotic stresses. *Journal of Experimental Botany*, 65(22), 6629–6644.
- Jin, Q., Hu, X., Li, X., Wang, B., Wang, Y., Jiang, H., Xu, Y. (2016). Genome-Wide Identification and Evolution Analysis of Trehalose-6-Phosphate Synthase Gene Family in *Nelumbo nucifera*. *Front. Plant Sci.*, 7, 1445–145.
- Kalibulla, S. I., Guruswami, G., Zothansanga, Ravi, P. Y., Nachimuthu, S. K., Shunmugiah, K. P., Surender, M. (2017). Bioinformatics-A Student's Companion. Singapore: Springer.
- Kataya, A. R., Elshobaky, A., Heidari, B., Dugassa, N. F., Thelen, J. L. (2020). Multi-targeted trehalose-6-phosphate phosphatase I harbors a novel peroxisomal targeting signal 1 and is essential for flowering and development. *Planta*, 251(98).
- Kaur, A., Pati, P. K., Pati, A. M., Nagpal, A. K. (2017). In-silico analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. *PloS one*, 12(9), e0184523.
- Kazuo, S. y Yamaguchi-Shinozaki, K. (2007). Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany*, 58(2), 221–227.

- Kim, C. Y. y Zhang, S. (2004). Activation of a mitogen-activated protein kinase cascade induces WRKY family of transcription factors and defense genes in tobacco. *The plant journal*, 38(1), 142-151.
- Kimura, S. y Sinha, N. (2008). Tomato (*Solanum lycopersicum*): A model fruit-bearing crop. *Cold Spring. Harb. Protoc.*, 3(11), 1-8.
- Kirsch, C., Takamiya-Wik, M., Schmelzer, E., Hahlbrock, K., Somssich, I. (2000). A novel regulatory element involved in rapid activation of parsley ELI7 gene family members by fungal elicitor or pathogen infection. *Mol Plant Pathol*, 1(4), 243-251.
- Kolbe, A., Tiessen, A., Schluepmann, H., Paul, M., Ulrich, S., Geigenberger, P. (2005). Trehalose 6-phosphate regulates starch synthesis via posttranslational redox activation of ADP-glucose pyrophosphorylase. *Proc. Natl. Acad. Sci. USA*, 102, 11118–11123.
- Kong, H., Landherr, L., Frohlich, M., Leebens-Mack, J., Ma, H., DePamphilis, C. (2007). Patterns of gene duplication in the plant SKP1 gene family in angiosperms: Evidence for multiple mechanisms of rapid gene birth. *Plant J.*, 50, 873–885.
- Kong, W., Ding, L., Cheng, J., Wang, B. (2018). Identification and expression analysis of genes with pathogen-inducible cis-regulatory elements in the promoter regions in *Oryza sativa*. *Rice (New York, N.Y.)*, 11(1), 52.
- Kosmas, S. A., Argyrokastritis, A., Loukas, M. G., Eliopoulos, E., Tsakas, S., Kaltsikes, P. J. (2006). Isolation and characterization of drought-related trehalose 6-phosphatesynthase gene from cultivated cotton (*Gossypium hirsutum* L.). *Planta*, 223(2), 329–339.
- Krasensky, J., Broyart, C., Rabanal, F. A., Jonak, C. (2014). The Redox-Sensitive Chloroplast Trehalose-6-Phosphate Phosphatase AtTPPD Regulates Salt Stress Tolerance. *Antioxidants & Redox Signaling*, 21(9), 1289-1304.
- Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, 33, 1870–1874.
- Kyte, J., y Doolittle, R. (1982). A simple method for displaying the hydrophathic character of a protein. *J Mol Biol*, 157(1), 105-132.
- la Cour, T., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K., Brunak, S. (2004). Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.*, 17(6), 527-36.
- Lam, E., y Chua, N. H. (1990). GT-1 binding site confers light responsive expression in transgenic tobacco. *Science*, 248(4954), 471-4.

- Lescot, M., Déhais, P., Moreau, Y., De Moor, B., Rouzé, P., Rombauts, S. (2002). PlantCARE: a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res., Database issue*, 30(1), 325-327.
- Leyman, B., Van Dijck, P., Thevelein, J. M. (2001). An unexpected plethora of trehalose biosynthesis genes in *Arabidopsis thaliana*. *Trends Plant Sci*, 6(11), 510–513.
- Li, J., Zhang, Z., Vang, S. *et al.* (2009). Correlation Between Ka/Ks and Ks is Related to Substitution Model and Evolutionary Lineage. *J Mol Evol*, 68, 414–423
- Li, H. W., Zang, B. S., Deng, X. W. (2011). Overexpression of the trehalose-6-phosphate synthase gene *OsTPS1* enhances abiotic stress tolerance in rice. *Planta*, 234, 1007–1018.
- Li, W. Y. (2014). Genome-wide analysis of the NADK gene family in plants. *PloS one*, 9(6), e101051.
- Lin, M., Jia, R., Li, J., Zhang, M., Chen, H., Zhang, D., X., C. (2018). Evolution and expression patterns of the trehalose-6-phosphate synthase gene family in drumstick tree (*Moringa oleifera* Lam.). *Planta*, 248, 999–1015.
- Lin, Q., Wang, S., Dao, Y., Wang, J., Wang, K. (2020). *Arabidopsis thaliana* trehalose-6-phosphate phosphatase gene *TPPI* enhances drought tolerance by regulating stomatal apertures. *Journal of Experimental Botany*, 71(14), 4285–4297.
- Lin, Q., Yang, J., Wang, Q., Zhu, H., Chen, Z. D., Wang, K. (2019). Overexpression of the trehalose-6-phosphate phosphatase family gene *AtTPPF* improves the drought tolerance of *Arabidopsis thaliana*. *BMC Plant Biol*, 19, 381.
- Liu, L., Zhang, X., Chen, F., Mahi, A. A., Wu, X., Chen, Q., Fu, Y. F. (2017). Analysis of promoter activity reveals that *GmFTL2* expression differs from that of the known Flowering Locus T genes in soybean. *The Crop Journal*, 5(5), 438-448.
- Liu, X., Fu, L., Qin, P., Sun, Y., Liu, J., Wang, X. (2019). Overexpression of the wheat trehalose 6-phosphate synthase 11 gene enhances cold tolerance in *Arabidopsis thaliana*. *Gene*, 710, 210-217.
- Liu, Y., Zhang, Y., Liu, J. (19 de Mayo de 2021). The expression profile of tomato in response to heat stress. Obtenido de GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174607>
- Logemann, E., Parniske, M., Hahlbrock, K. (1995). Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. *Proceedings of the National Academy of Sciences of the United States of America*, 92(13), 5905-5909.

- Lomeshwar, S., Manu, P., Neeru, K., Kalpna, B., Shikha, C., Om, P. D., Harsh, N. (2020). Plant growth-regulating molecules as thermoprotectants: functional relevance and prospects for improving heat tolerance in food crops. *Journal of Experimental Botany*, 71(2), 569-594.
- Lunn, Delorge, I., Figueroa, C. M., Van Dijck, P., Stitt, M. (2007). Gene families and evolution of trehalose metabolism in plants. *Funct Plant Biol*, 34, 550–563.
- Lunn, J., Delorge, I., Van Dijck, P. (2015). Trehalose metabolism in plants. *Plant J. Cell Mol. Biol.*, 79, 544–567.
- Lyu, J., Min, S., Lee, J., Lim, Y., Kim, J., Bae, C., Liu, J. (2013). Overexpression of a trehalose-6-phosphate synthase/phosphatase fusion gene enhances tolerance and photosynthesis during drought and salt stress without growth aberrations in tomato. *Plant Cell Tissue Organ Cult.*, 112, 257-262.
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y. *et al.* (2015). CDD: NCBI's conserved domain database. *Nucleic acids research*, 43(Database issue), D222–D226.
- Maruta, K., Mitsuzumi, H., Nakada, T., Kubota, M., Chaen, H., Fukuda, S. *et al.* (1996). Cloning and sequencing of a cluster of genes encoding novel enzymes of trehalose biosynthesis from thermophilic archaeobacterium *Sulfolobus acidocaldarius*. *Biochim Biophys Acta*, 1291, 177–181.
- Maruyama, K., Todaka, D., Mizoi, J., Yoshida, T., Kidokoro, S., Matsukura, S. *et al.* (2012). Identification of Cis-Acting Promoter Elements in Cold- and Dehydration-Induced Transcriptional Pathways in Arabidopsis, Rice, and Soybean. *DNA Research*, 19(1), 37-49.
- Miranda, J., Avonce, N., Suárez, R., Thevelein, J., Van Dijck, P., Iturriaga, G. (2007). A Bifunctional TPS–TPP Enzyme from Yeast Confers Tolerance to Multiple and Extreme Abiotic-Stress Conditions in Transgenic Arabidopsis. *Planta*, 226, 1411–1421.
- Mongkolsiriwatana, C., Pongtongkam, P., Peyachoknagul, S. (2009). *In silico* Promoter Analysis of Photoperiod-Responsive Genes Identified by DNA Microarray in Rice (*Oryza sativa* L.). *Kasetsart Journal - Natural Science*, 43(1), 164 - 177.
- Morabito, C., Secchi, F., Schubert, A. (2021). Grapevine TPS (trehalose-6-phosphate synthase) family genes are differentially regulated during development, upon sugar treatment and drought stress. *Plant Physiology and Biochemistry*, 164, 54-62.

- Mu, M., Lu, X. K., Wang, J. J., Wang, D. L., Yin, Z. J., Wang, S. *et al.* (2016). Genome-wide Identification and analysis of the stress-resistance function of the TPS (Trehalose-6-Phosphate Synthase) gene family in cotton. *BMC Genet*, 17, 54.
- Muthusamy, S., Sivalingam, P., Sridhar, J., Singh, D., Haldhar, S., Kaushal, P. (2017). Biotic stress inducible promoters in crop plants- a review. *Journal of Agriculture and Ecology*, 4, 14-24.
- Nicholas, K., Nicholas, H. J., Deerfield, D. I. (1997). GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS*, 4, 14.
- Nuccio, M., Wu, J., Mowers, R., Zhou, H.-P., Meghji, M., Primavesi, L. *et al.* (2015). Expression of Trehalose-6-Phosphate Phosphatase in Maize Ears Improves Yield in Well-Watered and Drought Conditions. *Nat. Biotechnol.*, 33, 862-869.
- Onodera, Y., Suzuki, A., Wu, C. Y., Washida, H., Takaiwa, F. (2001). A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. *J Biol Chem*, 276(17), 14139-14152.
- Osakabe, Y., Osakabe, K., Chiang, V. L. (2009). Characterization of the tissue-specific expression of phenylalanine ammonia-lyase gene promoter from loblolly pine (*Pinus taeda*) in *Nicotiana tabacum*. *Plant Cell Rep*, 28, 1309–1317.
- Pampurova, S., Verschooten, K., Avonce, N., Van Dijck, P. (2014). Functional screening of a cDNA library from the desiccation-tolerant plant *Selaginella lepidophylla* in yeast mutants identifies trehalose biosynthesis genes of plant and microbial origin. *J Plant Res*, 127, 803-813.
- Parul, S., Arif, T. J., Mudsser, A., Qazi, M. R. (2016). Plant abiotic stress: a prospective strategy of exploiting promoters as alternative to overcome the escalating burden. *Frontiers in Life Science*, 9(1), 52-63.
- Paul, M. J., Gonzalez-Uriarte, A., Griffiths, C., Hassani-Pak, K. (2018). The Role of Trehalose 6-Phosphate in Crop Yield and Resilience. *Plant Physiology*, 177, 12–23.
- Peng, Q., Wang, L., Ogutu, C., Liu, J., Liu, L., Mollah, M. D., Han, Y. (2020). Functional Analysis Reveals the Regulatory Role of PpTST1 Encoding Tonoplast Sugar Transporter in Sugar Accumulation of Peach Fruit. *Int. J. Mol. Sci.*, 21(3), 1112.
- Per, T. S., Khan, N. A., Reddy, P. S., Masood, A., Hasanuzzaman, M., Khan, M., Anjum, A. N. (2017). Approaches in modulating proline metabolism in plants for salt and drought stress tolerance: phytohormones, mineral nutrients and transgenics. *Plant Physiol Biochem*, 115, 126–140.

- Pradhan, S., Shyamli, P., Suranjika, S., Parida, A. (2021). Genome Wide Identification and Analysis of the R2R3-MYB Transcription Factor Gene Family in the Mangrove *Avicennia marina*. *Agronomy*, 11, 123.
- Priest, H. D., Filichkin, S. A., Mockler, T. C. (2009). Cis-regulatory elements in plant cell signaling. *Curr Opin Plant Biol.*, 12(5), 643-649.
- Qu, Q., Lee, S. J., Boss, W. (2004). TreT, a novel trehalose glycosyltransferring synthase of the hyperthermophilic archeon *Thermococcus litoralis*. *J Biol Chem*, 279, 47890–47897.
- Rahman, M., Rahman, M., Eom, J. S., Jeon, J. S. R., Rahman, M., Eom, J. S., Jeon, J. S. (2021). Genome wide Identification, Expression Profiling and Promoter Analysis of Trehalose 6 Phosphate Phosphatase Gene Family in Rice. *Journal of Plant Biology*, 64, 55-71.
- Rawat, R., Xu, Z., Yao, K. *et al.* (2005). Identification of cis-elements for ethylene and circadian regulation of the *Solanum melongena* gene encoding cysteine proteinase. *Plant Mol Biol*, 57, 629–643.
- Rehman, H. K., Malik, A., Vardar-Sukan, F., Ozturk, M. (2017). Plant Bioinformatics. Decoding the Phyta. Cham, Switzerland: Springer International Publishing.
- Reumann, S. B., y Lingner, T. (2012). PredPlantPTS1: A Web Server for the Prediction of Plant Peroxisomal Proteins. *Frontiers in Plant Science*, 3, 194.
- Rouster, J., Leah, R., Mundy, J., Cameron-Mills, V. (1997). Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. *The plant journal*, 11(3), 513-523.
- Sadau, S. B., Ahmad, A., Tajo, S. M., Ibrahim, S., Kazeem, B. B., Wei, H., Yu, S. (2021). Overexpression of GhMPK3 from Cotton Enhances Cold, Drought, and Salt Stress in Arabidopsis. *Agronomy*, 11(6), 1049.
- Sallam, A., Alqudah, A. M., Dawood, M., Baenziger, P. S., Börner, A. (2019). Drought Stress Tolerance in Wheat and Barley: Advances in Physiology, Breeding and Genetics Research. *International Journal of Molecular Sciences*, 20(13), 3137.
- Sheshadri, S. A., Nishanth, M. J., Bindu, S. (2016). Stress-Mediated cis-Element Transcription Factor Interactions Interconnecting Primary and Specialized Metabolism in planta. *Frontiers in Plant Science*, 7, 1725.
- Singh, V. K., Singh, A. K., Chand, R., Kushwaha, C. (2011). Role of bioinformatics in agriculture and sustainable development. *Int J Bioinforma Res*, 3(2), 221–226.

- Singh, V., Louis, J., Ayre, B. J., Reese, J. C., Shah, J. (2011). TREHALOSE PHOSPHATE SYNTHASE11-dependent trehalose metabolism promotes *Arabidopsis thaliana* defense against the phloem-feeding insect *Myzus persicae*. *The Plant Journal*, 67(1), 94-104.
- Slama, I., Abdelly, C., Bouchereau, A., Flowers, T., Saviouré, A. (2015). Diversity, distribution and roles of osmoprotective compounds accumulated in halophytes under abiotic stress. *Annals of Botany*, 115(3), 433-447.
- Song, J., Mao, H., Cheng, J., Zhou, Y., Chen, R., Zeng, L., et al. (2021). Identification of the trehalose-6-phosphate synthase gene family in *Medicago truncatula* and expression analysis under abiotic stresses. *Gene*, 787, 145641.
- Sperschneider, J., Catanzariti, A. M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., et al. (2017). LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports*, 7, 44598.
- Sutoh, K., y Yamauchi, D. (2003). Two cis-acting elements necessary and sufficient for gibberellin-upregulated proteinase expression in rice seeds. *The plant journal*, 34(5), 635-645.
- Suzuki, N., Bajad, S., Shuman, J., Shulaev, V., Mittler, R. (2008). The transcriptional co activator MBF1c is a key regulator of thermotolerance in *Arabidopsis thaliana*. *Journal of Biological Chemistry*, 283 (14), 9269-9275.
- Tian, L., Xie, Z., Lu, C., Hao, X., Wu, S., Huang, Y. et al. (2019). The trehalose-6-phosphate synthase TPS5 negatively regulates ABA signaling in *Arabidopsis thaliana*. *Plant Cell Rep.*, 38, 869–882.
- Van Dijck, P., Mascorro-Gallardo, J. O., De Bus, M., Royackers, K., Iturriaga, G., Thevelein, J. M. (2002). Truncation of *Arabidopsis thaliana* and *Selaginella lepidophylla* trehalose-6-phosphate synthase unlocks high catalytic activity and supports high trehalose levels on expression in yeast. *Biochem J.*, 366, 63-71.
- Van Houtte, H., Lopez-Galvis, L., Vandesteene, L., Beeckman, T., Van Dijck, P. (2013). Redundant and non-redundant roles of the trehalose-6-phosphate phosphatases in leaf growth, root hair specification and energy-responses in *Arabidopsis*. *Plant Signal Behav.*, e23209.
- Vandesteene, L., López-Galvis, L., Vanneste, K., Feil, R., Maere, S., Lammens, W. et al. (2012). Expansive evolution of the trehalose-6-phosphate phosphatase gene family in *Arabidopsis*. *Plant Physiol.*, 160, 884-896.
- Vandesteene, L., Ramon, M., Le Roy, K., Van Dijck, P., Rolland, F. (2010). A single active trehalose-6-P synthase (TPS) and a family of putative regulatory TPS-like proteins in *Arabidopsis*. *Mol Plant*, 3, 406–419.

- Vishal, B., Krishnamurthy, P., Ramamoorthy, R., Kumar, P. P. (2019). *OsTPS8* controls yield-related traits and confers salt stress tolerance in rice by enhancing suberin deposition. *New Phytologist*, 221, 1369–1386.
- Wai, A., Waseem, M., Khan, A., Nath, U., Lee, D., Kim, S. *et al.* (2021). Genome-Wide Identification and Expression Profiling of the PDI Gene Family Reveals Their Probable Involvement in Abiotic Stress Tolerance in Tomato (*Solanum lycopersicum* L.). *Genes*, 12, 23.
- Wang, Q., Hu, T., Zhan, X. (30 de Diciembre de 2020). Transcription profiles of tomato (*Solanum lycopersicum*) seedlings response to drought and heat stress. Obtenido de GEO Gen Expression Omnibus: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151277>
- Wang, S., Ouyang, K., Wang, K. (2019). Genome-Wide Identification, Evolution, and Expression Analysis of TPS and TPP Gene Families in *Brachypodium distachyon*. *Plants (Basel, Switzerland)*, 8(10), 362.
- Wang, W., Chen, Q., Xu, S., Liu, W. C., Zhu, X., Song, C. P. (2020). Trehalose-6-phosphate phosphatase E modulates ABA-controlled root growth and stomatal movement in Arabidopsis. *Journal of Integrative Plant Biology*, 62(10), 1518–1534.
- Wang, Y. J., Hao, Y. J., Zhang, Z. G., Chen, T., Zhang, J. S., Chen, S. Y. (2005). Isolation of Trehalose-6-Phosphate Phosphatase Gene from Tobacco and Its Functional Analysis in Yeast Cells. *J. Plant Physiol.*, 162, 215–223.
- Wang, Z., Rashotte, A. M., Moss, A. G. (2014). Two NAC transcription factors from *Citrullus colocynthis*, CcNAC1, CcNAC2 implicated in multiple stress responses. *Acta Physiol Plant*, 36, 621–634.
- Wang, Z., Zhu, Y., Wang, L., Liu, X., Liu, Y., Phillips, J. *et al.* (2009). A WRKY transcription factor participates in dehydration tolerance in *Boea hygrometrica* by binding to the W-box elements of the galactinol synthase (BhGolS1) promoter. *Planta*, 230, 1155-1166.
- Wannet, W. J., Op den Camp, H. J., Wisselink, H. W., van der Drift, C., Van Griensven, L. J., Vogels, G. D. (1998). Purification and characterization of trehalose phosphorylase from the commercial mushroom *Agaricus bisporus*. *Biochim Biophys Acta*, 1425, 177–188.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R. *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, 46(W1), W296–W303.

- Wei, B. Q., Wang, L. L., Zhang, R., Chen, L. Z. Z. L., Zhang, J. N. (2016). Identification of *CaTPS* Gene Family and Expression Analysis of *CaTPS1* in Hot Pepper. *Acta Horticulturae Sinica*, 43(8), 1504-1512.
- Wolters, H. y Jürgens, G. (2009). Survival of the flexible: hormonal growth control and adaptation in plant development. *Nat Rev Genet*, 10, 305–317.
- Wu, A., Hao, P., Wei, H., Sun, H., Cheng, S., Chen, P. *et al.* (2019). Genome-Wide Identification and Characterization of Glycosyltransferase Family 47 in Cotton. *Frontiers in Genetics*, 10, 824.
- Xiao, W., Chang, H., Zhou, P. *et al.* (2015). Genome-wide identification, classification and expression analysis of *GHMP* genes family in *Arabidopsis thaliana*. *Plant Syst Evol*, 301, 2125–2140.
- Xie, D. W., Wang, X. N., Fu, L. S., Sun, J., Zheng, W., Li, Z. F. (2015). Identification of the trehalose-6-phosphate synthase gene family in winter wheat and expression analysis under conditions of freezing stress. *Journal of Genetics*, 94(1), 55-65.
- Xie, L., Wang, Z.-X., Huang, B. (2014). Genome-wide identification classification and expression of *TPS* family genes in soybean. *Chinese Journal of Oil Sciences*, 36(2), 160-167.
- Xin, H., Zhuan-Di, W., Zheng-Ying, L., David, M. B., Yong-Bao, P., Cai-Wen, W. (2020). Genome-Wide Analysis of the Trehalose-6-Phosphate Synthase (TPS) Gene Family and Expression Profiling of *ScTPS* Genes in Sugarcane. *Agronomy*, 10(969), 1-15.
- Xu, Y., Wang, Y., Mattson, N., Yang, L., Jin, Q. (2017). Genome-wide analysis of the *Solanum tuberosum* (potato) trehalose-6-phosphate synthase (TPS) gene family: Evolution and diferential expression during development and stress. *BMC Genomics*, 18, 926.
- Xu, Z., Sun, M., Jiang, X., Sun, H., Dang, X., Cong, H., Qiao, F. (2018). Glycinebetaine Biosynthesis in Response to Osmotic Stress Depends on Jasmonate Signaling in Watermelon Suspension Cells. *Frontiers in Plant Science*, 9, 1469.
- Yang, H. L., Liu, Y. J., Wang, C. L., Zeng, Q. Y. (2012). Molecular evolution of trehalose-6-phosphate synthase (TPS) gene family in *Populus*, *Arabidopsis* and rice. *PLoS One*, 7(8), e42438.
- Yang, Y. C., Zhang, H. F., Yang, X. Z., Wang, Z. Y., Wei, C. H., Zhang, X. (2017). Expression analysis and genome-wide identification of trehalose-6-phosphate synthase gene family in melon (*Cucumis melo* L.). *Acta Botanica Boreali-Occidentalia Sinica*, 37(6), 1066-1072.

- Yang, Y., Ma, K., Zhang, T., Li, L., Wang, J., Chend, T., Zhang, Q. (2020). Characteristics and Expression Analyses of Trehalose-6-Phosphate Synthase Family in *Prunus mume* Reveal Genes Involved in Trehalose Biosynthesis and Drought Response. *Biomolecules*, 10(10), 1358.
- Yanjun, L., Xue, L., Lei, Y., Changtian, P., Lifei, C., Tao, Z., Gang, L. (2016). Genome-Wide Identification and Expression Analysis of Two-Component System Genome Tomato. *International Journal of Molecular Sciences*, 17(1204), 1-21.
- Yu, J., Zhang, Z., Wei, J. *et al.* (2014). SFGD: a comprehensive platform for mining functional information from soybean transcriptome data and its use in identifying acyl-lipid metabolism. *BMC Genomics*, 15, 271.
- Yu, W., Wang, L., Zhao, R., Sheng, J., Zhang, S., Li, R., Shen, L. (2019). Knockout of SIMAPK3 enhances tolerance to heat stress involving ROS homeostasis in tomato plants. *BMC Plant Biology*, 19(354).
- Zang, B., Li, H., Li, W., Deng, X. W., Wang, X. (2011). Analysis of trehalose-6-phosphate synthase (TPS) gene family suggests the formation of TPS complexes in rice. *Plant Mol Biol*, 76, 507–522.
- Zentella, R., Mascorro-Gallardo, J. O., Van Dijck, P., Folch-Mallol, J., Bonini, B., Van Vaeck, C., *et al.* (1999). A *Selaginella lepidophylla* trehalose-6-phosphate synthase complements growth and stress-tolerance defects in a yeast *tps1* mutant. *Plant Physiology*, 119, 1473–1482.
- Zhang, H., Hong, Y., Huang, L., Liu, S., Tian, L., Dai, Y. *et al.* (2016). Virus-Induced Gene Silencing-Based Functional Analyses Revealed the Involvement of Several Putative Trehalose-6-Phosphate. *Front. Plant Sci*, 7, 1176.
- Zhang, J., Xie, M., Tuskan, G. A., Muchero, W., Chen, J. G. (2018). Recent Advances in the Transcriptional Regulation of Secondary Cell Wall Biosynthesis in the Woody Plants. *Frontiers in plant science*, 9, 1535.
- Zhang, X., Wang, L., Meng, H., Wen, H., Fan, Y., Zha, J. (2011). Maize ABP9 enhances tolerance to multiple stresses in transgenic *Arabidopsis* by modulating ABA signaling and cellular levels of reactive oxygen species. *Plant Mol Biol*, 75, 365–378.
- Zhao, T. y Xu, X. (19 de Abril de 2020). Comparative transcriptome analysis between two tomato materials under cold stress. Obtenido de GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148887>
- Zhao, T. y Xu, X. (10 de Abril de 2020). Comparative transcriptome analysis between two tomato materials under salt stress. Obtenido de GEO: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148353>

- Zhao, T., Wu, T., Zhang, J., Wang, Z., Pei, T., Yang, H. *et al.* (2020). Genome-Wide Analyses of the Genetic Screening of C2H2-Type Zinc Finger Transcription Factors and Abiotic and Biotic Stress Responses in Tomato (*Solanum lycopersicum*) Based on RNA-Seq Data. *Frontiers in genetics*, 11, 540.
- Zhou, B., Fang, Y. F. *et al.* (2017). Expressional characterization of two class I trehalose-6-phosphate synthase genes in *Hevea brasiliensis* (para rubber tree) suggests a role in rubber production. *New Forests*, 48, 513–526.
- Zhou, M. L., Zhang, Q., Sun, Z. M., Chen, L.-H., Liu, B. X., Zang, K. X. *et al.* (2013). Trehalose Metabolism-Related Genes in Maize. *J Plant Growth Regul*, 33, 256–271.
- Zhou, S., Zhang, H., Li, R., Hong, Q., Li, Y., Xia, Q., Zhang, W. (2017). Function identification of the nucleotides in key cis-element of Dysfunctional Tapetum1 (DYT1) promoter. *Front Plant Sci*, 8, 153.
- Zhou, Y., Hu, L., Wu, H., Jiang, L., Liu, S. (2017). Genome-Wide Identification and Transcriptional Expression Analysis of Cucumber Superoxide Dismutase (SOD) Family in Response to Various Abiotic Stresses. *International journal of genomics*, 2017, 7243973.
- Zulfiqar, A. M., Sana, A., Jamil, A., Abdul, N. J., Ahmed, S., Uzair, H. M., Abdullah. (2019). A genome wide approach to the comprehensive analysis of GASA gene family in Glycine max. *Plant Molecular Biology*, 100, 607–620.
- Zulfiqar, F., Akram, N. A., Ashraf, M. (2019). Osmoprotection in plants under abiotic stresses: new insights into a classical phenomenon. *Planta*, 251(3), 2020.

IX. APÉNDICES

Apéndice 1. Las secuencias proteicas TPS usadas para construir los árboles filogenéticos.

Organismo	Secuencia	Gene identifier	Database
<i>Arabidopsis thaliana</i>	AtTPS1	Q9SYM4.1	NCBI
	AtTPS2	Q9FZ57.1	
	AtTPS3	Q9SHG0.3	
	AtTPS4	Q9T079.1	
	AtTPS5	O23617.2	
	AtTPS6	Q94AH8.2	
	AtTPS7	Q9LMI0.1	
	AtTPS8	Q0WUI9.1	
	AtTPS9	Q9LRA7.1	
	AtTPS10	O80738.1	
	AtTPS11	Q9ZV48.1	
<i>Oryza sativa</i>	OsTPS1	AEB53177.1	NCBI
	OsTPS2	AEB53178.1	
	OsTPS3	AEB53179.1	
	OsTPS4	AEB53180.1	
	OsTPS5	AEB53181.1	
	OsTPS6	AEB53187.1	
	OsTPS7	AEB53182.1	
	OsTPS8	AEB53183.1	
	OsTPS9	AEB53184.1	
	OsTPS10	AEB53185.1	
	OsTPS11	AEB53186.1	
<i>S. lepidophylla</i>	SleTPS1	AAD00829	NCBI
<i>Zea mays</i>	ZmTPS1.1	GRMZM2G099860_T01	Phytozome
	ZmTPS2	GRMZM2G079928_T01	
	ZmTPS3	GRMZM2G008226_T01	
	ZmTPS4	GRMZM2G527891_T01	
	ZmTPS5	GRMZM2G304274_T01	
	ZmTPS6	GRMZM2G123277_T01	
	ZmTPS7.2	GRMZM2G007736_T02	
	ZmTPS8	GRMZM2G366659_T01	
	ZmTPS9	GRMZM2G312521_T01	
	ZmTPS10	GRMZM2G122231_T01	
	ZmTPS11.1	GRMZM2G001304_T01	
	ZmTPS12.2	GRMZM2G019183_T02	
	ZmTPS13.1	GRMZM2G416836_T01	
	ZmTPS14.1	GRMZM2G068943_T02	

<i>Solanum tuberosum</i>	ZmTPS15	GRMZM2G118462_T01	Phytozome
	StTPS1	PGSC0003DMT400058627	
	StTPS2	PGSC0003DMT400073263	
	StTPS3	PGSC0003DMT400061642	
	StTPS4	PGSC0003DMT400010526	
	StTPS5	PGSC0003DMT400070613	
	StTPS6	PGSC0003DMT400044502	
	StTPS7	PGSC0003DMT400045229	
	StTPS8	PGSC0003DMT400027375	
<i>Solanum lycopersicum</i>	SITPS-1	Solyc01g005210.2	Phytozome
	SITPS-2	Solyc02g071590.1	
	SITPS-3	Solyc02g072150.2	
	SITPS-4	Solyc04g025940.2	
	SITPS-5	Solyc05g005750.2	
	SITPS-6	Solyc07g006500.2	
	SITPS-7	Solyc07g055300.2	
	SITPS-8	Solyc07g062140.2	
	SITPS-9	Solyc08g076650.2	
	SITPS-10	Solyc10g007950.2	

Apéndice 2. Las secuencias proteicas TPP usados para construir los árboles filogenéticos.

Organismo	Secuencia	Gene identifiier	Database
<i>Arabidopsis thaliana</i>	AtTPPA	AT5G51460	Phytozome
	AtTPPB	AT1G78090	
	AtTPPC	AT1G22210	
	AtTPPD	AT1G35910	
	AtTPPE	AT2G22190	
	AtTPPF	AT4G12430	
	AtTPPG	AT4G22590	
	AtTPPH	AT4G39770	
	AtTPPI	AT5G10100	
	AtTPPJ	AT5G65140	
<i>Brachypodium distachyon</i>	BdTPPA	Bradi3g32970.1	Phytozome
	BdTPPB	Bradi1g27470.1	
	BdTPPC	Bradi3g50810.1	
	BdTPPD	Bradi5g17890.1	
	BdTPPE	Bradi3g35590.1	
	BdTPPF	Bradi4g29030.1	
	BdTPPG	Bradi3g58960.1	
	BdTPPH	Bradi1g60950.1	

<i>Oryza sativa</i>	BdTPPI	Bradi1g45480.1	Phytozome
	BdTPPJ	Bradi1g21420.1	
	OsTPP1	LOC_Os02g44230.1	
	OsTPP2	LOC_Os10g40550.1	
	OsTPP3	LOC_Os07g43160.1	
	OsTPP4	LOC_Os02g51680.1	
	OsTPP5	LOC_Os04g46760.1	
	OsTPP6	LOC_Os08g31630.1	
	OsTPP7	LOC_Os09g20390.1	
	OsTPP8	LOC_Os06g11840.1	
	OsTPP9	LOC_Os03g26910.1	
	OsTPP10	LOC_Os07g30160.1	
	OsTPP11	LOC_Os12g09060.1	
OsTPP12	LOC_Os08g25410.1		
<i>Populus trichocarpa</i>	PtTPP1	Potri.003G112400.1	Phytozome
	PtTPP2	Potri.007G090900.1	
	PtTPP3	Potri.015G126900.1	
	PtTPP4	Potri.002G094500.1	
	PtTPP5	Potri.005G077200.1	
	PtTPP6	Potri.001G120500.1	
	PtTPP7	Potri.012G001000.1	
	PtTPP8	Potri.012G126100.1	
<i>Zea mays</i>	ZmTPPA	GRMZM2G178546	Phytozome
	ZmTPPB	GRMZM2G112830	
	ZmTPPC	GRMZM2G347280	
	ZmTPPD	GRMZM2G140078	
	ZmTPPE	GRMZM2G174396	
	ZmTPPF	GRMZM2G055150	
	ZmTPPG	GRMZM2G151044	
	ZmTPPH	GRMZM2G080354	
	ZmTPPI	GRMZM2G014729	
	ZmTPPJ	GRMZM2G117564	
	ZmTPPK	GRMZM5G840145	
ZmTPPA	GRMZM2G178546		
<i>Solanum lycopersicum</i>	SITPP-1	Solyc03g007290.2	Phytozome
	SITPP-2	Solyc03g083960.2	
	SITPP-3	Solyc04g054930.2	
	SITPP-4	Solyc04g072920.2	
	SITPP-5	Solyc04g082550.1	
	SITPP-6	Solyc05g051880.2	
	SITPP-7	Solyc06g060600.1	
	SITPP-8	Solyc08g079060.2	
	SITPP-9	Solyc10g046770.1	

Apéndice 3. Identidad y porcentaje de cobertura entre SITPSs.

	SITPS-1		SITPS-2		SITPS-3		SITPS-4	
	Query Cover (%)	Ident (%)						
SITPS-1	100%	100%	87%	35.68%	98%	66.47%	99%	64.20%
SITPS-2	77%	35.68%	100%	100.00%	75%	33.56%	75%	33.99%
SITPS-3	98%	66.47%	85%	33.56%	100%	100%	99%	62.71%
SITPS-4	98%	63.77%	85%	33.99%	98%	62.71%	100%	100%
SITPS-5	97%	64.82%	87%	34.36%	98%	62.81%	100%	88.45%
SITPS-6	97%	59.56%	87%	33.70%	97%	59.35%	96%	62.68%
SITPS-7	98%	64.83%	90%	33.82%	99%	76.35%	98%	62.06%
SITPS-8	77%	35.04%	96%	75.76%	77%	34.83%	78%	33.81%
SITPS-9	98%	76.17%	85%	35.57%	97%	68.01%	97%	65.68%
SITPS-10	98%	65.64%	86%	34.73%	99%	76.59%	98%	61.18%

	SITPS-5		SITPS-6		SITPS-7		SITPS-8	
	Query Cover (%)	Ident (%)						
SITPS-1	99%	64.82%	99%	59.56%	99%	64.83%	86%	35.04%
SITPS-2	75%	34.17%	80%	33.70%	81%	33.58%	94%	75.49%
SITPS-3	99%	62.81%	99%	59.35%	99%	76.24%	85%	34.83%
SITPS-4	100%	88.45%	98%	62.56%	98%	62.06%	86%	33.81%
SITPS-5	100%	100%	97%	63.73%	97%	62.49%	86%	33.86%
SITPS-6	96%	63.88%	100%	100%	96%	58.93%	86%	33.55%
SITPS-7	98%	62.49%	98%	58.92%	100%	100%	86%	34.95%
SITPS-8	77%	33.86%	77%	33.77%	78%	34.95%	100%	100%
SITPS-9	98%	65.69%	97%	60.59%	98%	64.86%	85%	36.10%
SITPS-10	98%	61.42%	98%	59.13%	99%	89.20%	85%	35.89%

	SITPS-9		SITPS-10	
	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)
SITPS-1	99%	76.64%	98%	65.64%
SITPS-2	76%	35.57%	76%	34.73%
SITPS-3	98%	68.01%	99%	76.59%
SITPS-4	97%	65.68%	97%	61.18%
SITPS-5	98%	65.69%	97%	61.42%
SITPS-6	96%	60.68%	96%	58.62%
SITPS-7	98%	65.21%	99%	89.20%
SITPS-8	77%	36.10%	77%	35.89%
SITPS-9	100%	100%	97%	65.48%
SITPS-10	98%	65.60%	100%	100%

Apéndice 4. Identidad y porcentaje de cobertura entre SITPPs.

	SITPP-1		SITPP-2		SITPP-3		SITPP-4	
	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)
SITPP-1	100%	100%	97	50.3	85	54.17	85	54.67
SITPP-2	75%	56.07%	100%	100%	93%	56.81%	98%	55.71%
SITPP-3	87%	54.17%	93%	60.95%	100%	100%	95%	72.37%
SITPP-4	76%	54.67%	97%	56.02%	98%	65.59%	100%	100%
SITPP-5	96%	54.08%	95%	50.17%	94%	52.52%	98%	52.78%
SITPP-6	77%	50.88%	78%	62.72%	97%	69.47%	97%	55.59%
SITPP-7	90%	54.14%	97%	69.90%	97%	59.81%	97%	57.56%
SITPP-8	77%	52.16%	97%	69.50%	90%	54.08%	95%	52.14%

	SITPP-5		SITPP-6		SITPP-7		SITPP-8	
	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)
SITPP-1	87	53.54	67	50.88	92	52.08	88	52.16
SITPP-2	76%	49.12%	61%	63.11%	87%	71.21%	100%	69.73%
SITPP-3	87%	51.96%	90%	70.73%	92%	59.81%	92%	57.69%
SITPP-4	78%	52.23%	94%	53.30%	82%	56.83%	94%	52.14%
SITPP-5	100%	100%	74%	49.31%	95%	48.44%	95%	46.55%
SITPP-6	77%	49.31%	100%	100%	77%	60.18%	77%	59.03%
SITPP-7	90%	48.29%	72%	59.83%	100%	100%	98%	79.68%
SITPP-8	75%	46.08%	58%	59.03%	86%	80.30%	100%	100%

Apéndice 5. Identidad y porcentaje de cobertura entre AtTPS y SITPS.

	AtTPS1		AtTPS2		AtTPS3		AtTPS4	
	Query Cover (%)	Ident (%)						
SITPS-1	77%	35.26%	93%	34.84%	97%	32.69%	99%	33.45%
SITPS-2	96%	74.70%	93%	65.80%	97%	57.26%	98%	59.05%
SITPS-3	75%	33.95%	92%	33.71%	96%	32.19%	98%	31.66%
SITPS-4	75%	33.86%	93%	34.28%	98%	32.33%	99%	32.53%
SITPS-5	75%	33.16%	92%	34.62%	98%	32.57%	99%	32.69%
SITPS-6	76%	34.21%	93%	32.84%	91%	31.97%	99%	31.12%
SITPS-7	80%	34.05%	95%	34.47%	98%	33.13%	99%	32.81%
SITPS-8	99%	79.39%	93%	62.23%	97%	59.88%	99%	61.98%
SITPS-9	76%	35.36%	93%	34.73%	97%	33.37%	98%	33.37%
SITPS-10	83%	33.90%	94%	34.44%	98%	32.41%	99%	33.54%

	AtTPS5		AtTPS6		AtTPS7		AtTPS8	
	Query Cover (%)	Ident (%)						
SITPS-1	98%	72.31%	99%	81.45%	98%	64.29%	98%	62.46%
SITPS-2	85%	35.32%	88%	35.22%	85%	34.31%	86%	33.50%
SITPS-3	97%	66.23%	98%	64.76%	98%	75.03%	98%	60.33%
SITPS-4	97%	62.84%	99%	62.63%	98%	61.33%	100%	73.11%
SITPS-5	97%	63.85%	99%	63.58%	98%	62.51%	100%	73.34%
SITPS-6	98%	58.62%	99%	59.13%	98%	60.02%	99%	61.27%
SITPS-7	99%	63.29%	100%	62.41%	98%	74.56%	98%	60.68%
SITPS-8	85%	35.75%	87%	34.28%	85%	34.04%	87%	33.68%
SITPS-9	97%	78.82%	99%	72.94%	98%	66.00%	100%	61.48%
SITPS-10	98%	63.56%	99%	62.63%	98%	75.09%	97%	60.85%

	AtTPS9		AtTPS10		AtTPS11	
	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)
SITPS-1	97%	63.32%	97%	62.12%	96%	57.89%
SITPS-2	86%	33.97%	87%	34.44%	85%	34.99%
SITPS-3	97%	61.01%	99%	60.61%	96%	57.62%
SITPS-4	100%	74.19%	100%	71.54%	98%	60.23%
SITPS-5	100%	74.16%	100%	72.71%	96%	62.01%
SITPS-6	97%	60.49%	99%	58.39%	95%	65.36%
SITPS-7	97%	61.16%	97%	58.46%	91%	59.34%
SITPS-8	86%	33.85%	84%	34.53%	85%	34.60%
SITPS-9	97%	64.50%	100%	61.53%	97%	59.51%
SITPS-10	97%	60.95%	97%	58.98%	91%	58.91%

Apéndice 6. Identidad y porcentaje de cobertura entre AtTPP y SITPP.

	AtTPPA		AtTPPB		AtTPPC		AtTPPD	
	Query Cover (%)	Ident (%)						
SITPP-1	75%	54.27%	72%	54.23%	84%	52.46%	75%	54.64%
SITPP-2	99%	69.51%	98%	50.77%	96%	54.35%	93%	56.46%
SITPP-3	82%	58.62%	98%	60.21%	97%	62.42%	99%	64.02%
SITPP-4	84%	57.06%	100%	61.64%	97%	61.59%	100%	62.07%
SITPP-5	74%	51.38%	75%	51.03%	85%	48.92%	74%	55.40%
SITPP-6	59%	59.57%	79%	53.16%	66%	64.06%	86%	54.06%
SITPP-7	85%	65.96%	79%	57.01%	84%	59.36%	87%	55.79%
SITPP-8	97%	61.94%	95%	50.79%	93%	55.70%	88%	55.75%

	AtTPPE		AtTPPF		AtTPPG		AtTPPH	
	Query Cover (%)	Ident (%)						
SITPP-1	77%	53.85%	76%	53.31%	76%	53.63%	81%	53.40%
SITPP-2	84%	58.20%	97%	67.72%	98%	65.89%	85%	59.06%
SITPP-3	84%	75.50%	86%	56.88%	90%	53.64%	91%	73.46%
SITPP-4	84%	73.42%	96%	51.49%	89%	53.08%	86%	71.19%
SITPP-5	75%	51.87%	75%	48.03%	74%	47.72%	79%	51.06%
SITPP-6	83%	60.07%	60%	57.92%	60%	56.83%	85%	59.60%
SITPP-7	82%	55.87%	87%	72.56%	88%	69.37%	85%	57.73%
SITPP-8	82%	56.01%	98%	68.93%	98%	65.80%	85%	56.60%

	AtTPPI		AtTPPJ	
	Query Cover (%)	Ident (%)	Query Cover (%)	Ident (%)
SITPP-1	77%	51.53%	77%	53.22%
SITPP-2	89%	54.60%	95%	52.43%
SITPP-3	97%	67.74%	99%	69.25%
SITPP-4	89%	66.16%	100%	63.30%
SITPP-5	78%	48.97%	76%	50.68%
SITPP-6	86%	54.69%	85%	56.15%
SITPP-7	83%	54.63%	83%	55.38%
SITPP-8	95%	52.72%	93%	52.75%

Apéndice 7. Motivos detectados en los genes SITPS/SITPP.

Motif #	Sequence	Width	Description
1	QGVDKGLVAEFLLES LGYK GKSDVFPJY IGDDRTDEDAFEV	41	Trehalose- phosphatase
2	KCVPTFLPPEJVEKYYHG FCKQHLWPL FHYM	31	Glycosyltransferase family 20
3	WIHDYHLMVLPTFLRKR FN RVKLGFFLH SPFPSSEIYRTL P	41	Glycosyltransferase family 20
4	RHEKHRYVSTHDVAYWARSFLQDLER ACKDHYRKRCWGIG	41	Glycosyltransferase family 20
5	CCSRMLGLEYZSKRGYIGLEYYGRTVGI KILPVG IHMGQJE	41	Glycosyltransferase family 20
6	RDPDPAFGSCQAKELLDHLESVLPTEP VSVKSGQKIVEVKP	41	Trehalose- phosphatase
7	PGYEPVILIDRPVSFYERIAYYAVAECVL VTAVRDGMNLVPYEYIVCRQG	50	Glycosyltransferase family 20
8	QFEGKTVLLGVDDMDIFKGISLKLLAFEH LLEQHPSWQGKVV LVQIANPA	50	Glycosyltransferase family 20
9	GPKKSMLVVSEFIGCSPSLSGAIRVNPW BIEAVAEAMNEAITMAEPEKQL	50	Glycosyltransferase family 20
10	LPLSPDHGGFRFDRSLWZAYVSANKIFA DKVMEVINPEDDYV	41	Glycosyltransferase family 20

Apéndice 8. Detalles de los elementos de la estructura secundaria de SITPSs calculados por SOPMA.

Secondary structure element	SITPS-1	SITPS-2	SITPS-3	SITPS-4	SITPS-5
Alpha helix	42.59%	42.89%	43.24%	42.82%	43.71%
310 helix	0.00%	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%	0.00%
Extended strand	16.80%	15.80%	16.57%	17.39%	17.16%
Beta turn	4.90%	5.62%	4.58%	4.78%	4.70%
Bend region	0.00%	0.00%	0.00%	0.00%	0.00%
Random coil	35.71%	36.69%	35.61%	35.01%	34.43%
Ambiguous states	0.00%	0.00%	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%	0.00%	0.00%

Secondary structure element	SITPS-6	SITPS-7	SITPS-8	SITPS-9	SITPS-10
Alpha helix	42.01%	42.07%	42.98%	42.00%	42.82%
310 helix	0.00%	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%	0.00%
Extended strand	16.55%	17.25%	14.25%	16.71%	16.69%
Beta turn	5.82%	4.55%	5.62%	5.22%	4.90%
Bend region	0.00%	0.00%	0.00%	0.00%	0.00%
Random coil	35.62%	36.13%	37.15%	36.08%	35.59%
Ambiguous states	0.00%	0.00%	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%	0.00%	0.00%

Apéndice 9. Validación de las estructuras de las proteínas SITPS.

Protein Name	Favored region	Outlier region	Sequence Identity	Sequence similarity	Coverage
SITPS-1	93.00%	1.83%	38.17%	40%	55%
SITPS-2	95.42%	1.70%	49.68%	44%	49%
SITPS-3	93.53%	1.44%	37.47%	40%	55%
SITPS-4	92.02%	1.23%	39.23%	41%	55%
SITPS-5	92.23%	1.84%	37.95%	40%	55%
SITPS-6	93.84%	1.62%	35.82%	39%	54%
SITPS-7	92.71%	1.54%	38.46%	41%	55%
SITPS-8	96.11%	0.54%	50.98%	45%	49%
SITPS-9	92.06%	1.55%	40.60%	41%	54%
SITPS-10	92.09%	2.05%	38.54%	41%	54%

Apéndice 10. Detalles de los elementos de la estructura secundaria de SITPPs calculados por SOPMA.

Secondary structure element	SITPP-1	SITPP-2	SITPP-3	SITPP-4	SITPP-5
Alpha helix	38.99%	32.70%	46.06%	43.84%	38.91%
310 helix	0.00%	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%	0.00%
Extended strand	15.77%	15.26%	16.09%	16.16%	18.77%
Beta turn	10.42%	7.36%	7.89%	6.58%	8.53%
Bend region	0.00%	0.00%	0.00%	0.00%	0.00%
Random coil	34.82%	44.69%	29.97%	33.42%	33.79%
Ambiguous states	0.00%	0.00%	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%	0.00%	0.00%

Secondary structure element	SITPP-6	SITPP-7	SITPP-8	SITPP-9
Alpha helix	34.64%	42.77%	37.05%	33.33%
310 helix	0.00%	0.00%	0.00%	0.00%
Pi helix	0.00%	0.00%	0.00%	0.00%
Beta bridge	0.00%	0.00%	0.00%	0.00%
Extended strand	20.71%	17.30%	16.32%	21.05%
Beta turn	9.64%	10.06%	7.25%	4.39%
Bend region	0.00%	0.00%	0.00%	0.00%
Random coil	35.00%	29.87%	39.38%	41.23%
Ambiguous states	0.00%	0.00%	0.00%	0.00%
Other states	0.00%	0.00%	0.00%	0.00%

Apéndice 11. Validación de las estructuras de las proteínas SITPP.

Protein Name	Favored region	Outlier region	Sequence Identity	Sequence similarity	Coverage
SITPP-1	91.20%	2.11%	35.36%	38%	78%
SITPP-2	92.57%	1.86%	34.80%	38%	68%
SITPP-3	92.11%	1.13%	38.34%	39%	80%
SITPP-4	94.55%	0.00%	34.85%	38%	72%
SITPP-5	92.06%	2.53%	34.59%	37%	91%
SITPP-6	92.98%	0.88%	34.86%	38%	78%
SITPP-7	94.35%	1.06%	35.09%	38%	83%
SITPP-8	89.93%	2.24%	35.89%	38%	64%
SITPP-9	93.94%	1.01%	21.78%	30%	89%

Protein Name	Residues within 4 Å°	PLIP interactions with chain A
SITPP-1	Chain A: D.76, D.78, D.279	Metal complexes: A:D.78, A:D.279
SITPP-2	Chain A: D.125, D.127, D.326	Metal complexes: A:D.125, A:D.127, A:D.326
SITPP-3	Chain A: D.50, D.52, D.242	Metal complexes: A:D.52, A:D.242
SITPP-4	Chain A: D.109, D.111, D.302	Metal complexes: A:D.109, A:D.111, A:D.302
SITPP-5	Chain A: D.50, D.52, D.243	Metal complexes: A:D.50, A:D.52, A:D.243
SITPP-6	N/A	N/A
SITPP-7	Chain A: D.59, D.61, D.261	Metal complexes: A:D.59, A:D.61, A:D.261
SITPP-8	Chain A: D.124, D.126, D.327	Metal complexes: A:D.124, A:D.126, A:D.327
SITPP-9	N/A	N/A

Apéndice 12. Elementos Regulatorios *Cis* (CREs) con secuencias y funciones en los promotores de *SITPSs*.

Grupo	Elementos <i>Cis</i>	Secuencia	Funciones
LrCRE	ABRE3a	TACGTG	<i>cis</i> -acting regulatory element involved in light responsiveness
	ABRE4	CACGTA	<i>cis</i> -acting regulatory element involved in light responsiveness
	ACA-motif	AATCACAACCATA	part of <i>gapA</i> in (<i>gapA</i> -CMA1) involved with light responsiveness
	ACE	CTAACGTATT	<i>cis</i> -acting element involved in light responsiveness
	AE-box	AGAAACAA/AGAAACTT	part of a module for light response
	Box II	ACACGTTGT/CCACGTGGC	part of a light responsive element
	Box 4	ATTAAT	part of a conserved DNA module involved in light responsiveness
	chs-CMA1a	TTACTTAA	part of a light responsive element
	G-box	CACGAC/CACGTT/CACGTG/ CCACGTAA/TACGTG/TCCA CGTGGCATT/TCCACATGGC A/ACACGTGGC/tgACACGTG GCA/GCCACGTGGA	<i>cis</i> -acting regulatory element involved in light responsiveness
	GA-motif	ATAGATAA	part of a light responsive element
	Gap-box	CAAATGAA(A/G)A	part of a light responsive element
	GATA-motif	GATAGGA/GATAGGG/AAGA TAAGATT/AAGGATAAGG	part of a light responsive element
	GT1-motif	GGTTAAT/GGTTAA	light responsive element
	I-box	ccttaccct/atGATAAGGTC	part of a light responsive element
	L-box	ATCCCACCTAC	part of a light responsive element
	MRE	AACCTAA	MYB binding site involved in light responsiveness
	TCCC-motif	TCTCCCT	part of a light responsive element

	TCT-motif	TCTTAC	part of a light responsive element
SrCRE	ARE	AAACCA	<i>cis</i> -acting regulatory element essential for the anaerobic induction
	as-1	TGACG	Defense response
	box S	AGCCACC	elicitation; wounding and pathogen responsiveness
	CCAAT-box	CAACGG	MYBHv1 binding site (involved in the activation of stress and defense genes in plants)
	DRE core	GCCGAC	<i>cis</i> -acting element that regulates cold- and dehydration-responsive gene expression
	DRE1	ACCGAGA	drought responsive element 1
	LTR	CCGAAA	<i>cis</i> -acting element involved in low-temperature responsiveness
	MBS	CAACTG	MYB binding site involved in drought-inducibility
	MYB	CAACCA/TAACCA/CAACAG	stress response element
	MYB-like sequence	TAACCA	stress response element
	MYB recognition site	CCGTTG	stress response element
	Myb	TAACTG/CAACTG	stress response element
	Myb-binding site	CAACAG	stress response element
	MYC	CATGTG/CATTTG/CAATTG	<i>cis</i> -acting element involved in drought responsiveness
	Myc	TCTCTTA	stress response element
	STRE	AGGGG	stress response element
	TC-rich repeats	GTTTTCTTAC	<i>cis</i> -acting element involved in defense and defense and stress responsiveness
	Unnamed__1	CGTGG/GCCACGTGGC	stress response element
W box	TTGACC	fungal elicitor responsive element	
WRE3	CCACCT	stress response element	

	WUN-motif	AAATTACT/TTATTACAT	stress response element
HrCRE	ABRE	ACGTG/CACGTG/GCAACGT GTC	<i>cis</i> -acting element involved in the abscisic acid responsiveness
	ABRE2	CCACGTGG	<i>cis</i> -acting element involved in the abscisic acid responsiveness
	AT-rich element	ATAGAAATCAA	binding site of AT-rich DNA binding protein (ATBP-1)
	AuxRR-core CARE	GGTCCAT CAACTCAC	<i>cis</i> -acting regulatory element involved in auxin responsiveness gibberellin-responsiveness
	CGTCA-motif ERE	CGTCA ATTTTAAA/ATTTTCATA	<i>cis</i> -acting regulatory element involved in the MeJA-responsiveness Ethylene-responsive element
	O2-site	GATGACATGG/GATGA(C/T)(A/G)TG(A/G)	<i>cis</i> -acting regulatory element involved in zein metabolism regulation
	P-box	CCTTTTG	gibberellin-responsive element
	TATC-box TCA	TATCCCA TCATCTTCAT	<i>cis</i> -acting element involved in gibberellin-responsiveness salicylic acid responsive element
	TCA-element TGA-element	CCATCTTTTT/TCAGAAGAG G AACGAC	<i>cis</i> -acting element involved in salicylic acid responsiveness auxin-responsive element
	TGACG-motif	TGACG	<i>cis</i> -acting regulatory element involved in the MeJA-responsiveness
CdCRE	AAGAA-motif	GAAAGAA/gGTAAAGAAA (T/C)C(T/C)(C/T)ACC(T/C)AC	<i>cis</i> -element involved in seed specific expression
	AC-I	C	Cellular Development
	AC-II	CCACCAACCCCC	AC-II repeat promoter confer xylem specific expression. ACBF binds to AC-II
	CAT-box	GCCACT	<i>cis</i> -acting regulatory element related to meristem expression

	circadian	CAAAGATATC	<i>cis</i> -acting regulatory element involved in circadian control
	dOCT	CACGGATC	the <i>cis</i> -acting regulatory element related to meristem-specific activation
	GCN4_motif	TGAGTCA	<i>cis</i> -regulatory element involved in endosperm expression
	Unnamed__4	CTCC	temporal and spatial specificity of DYT1(Anther) expression
Otros			
CRE	Unnamed__6	taTAAATATct	Unknown
	Unnamed__1	GAATTTAATTAA	60K protein binding site

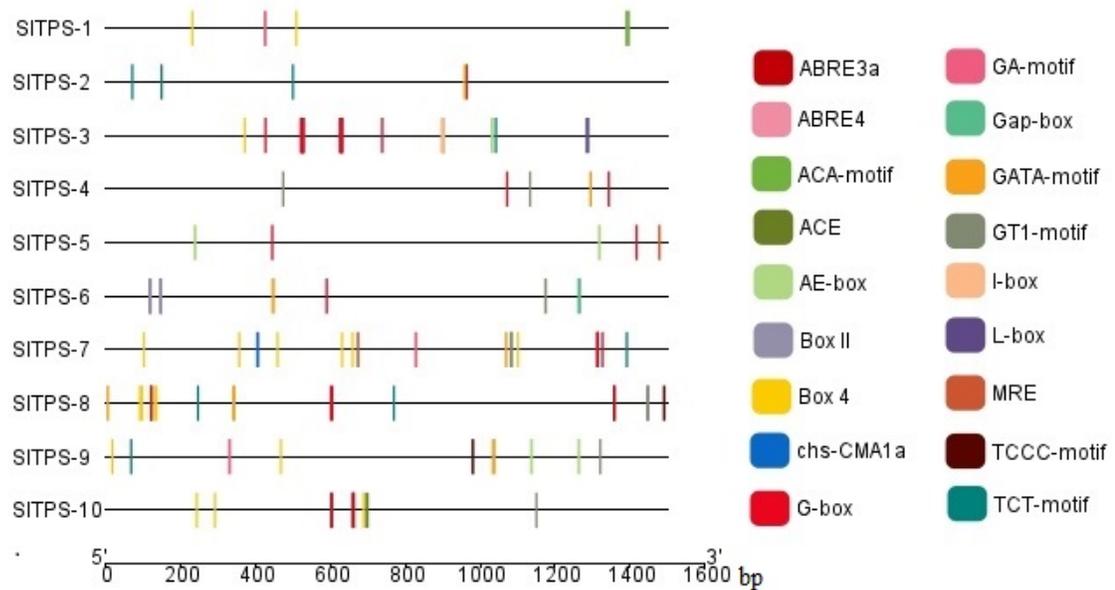
Apéndice 13. Elementos Regulatorios *Cis* (CREs) con secuencias y funciones en los promotores de SITPPs.

Grupo	Elementos <i>Cis</i>	Secuencia	Funciones
LrCRE	3-AF1 binding site	TAAGAGAGGAA	light responsive element
	ABRE3a	TACGTG/CACGTA	<i>cis</i> -acting regulatory element involved in light responsiveness
	ABRE4	CACGTA	<i>cis</i> -acting regulatory element involved in light responsiveness
	ACE	CTAACGTATT/GACACGTAT	<i>cis</i> -acting element involved in light responsiveness
	AE-box	G	
	AT1-motif	AGAAACAA/AGAAACTT	part of a module for light response
	Box I	AATTATTTTTTATT	part of a light responsive module
	Box II	TGGTAATAA	part of a light responsive element
	Box 4	ATTAAT	part of a conserved DNA module involved in light responsiveness
	chs-CMA1a	TACTTAA	part of a light responsive element
	chs-CMA2a	TCACTTGA	part of a light responsive element
	G-box	CACGAC/CACGTT/CACGTC/ CACGTG/CCACGTAA/TACG TG/ACACGTGT/CAGACGTG GCA	<i>cis</i> -acting regulatory element involved in light responsiveness
	Gap-box	CAAATGAA(A/G)A	part of a light responsive element
	GATA-motif	GATAGGA	part of a light responsive element
	GT1-motif	GGTTAAT/GGTTAA	light responsive element
	I-box	gGATAAGGTG	part of a light responsive element
	LAMP-element	CTTTATCA	part of a light responsive element
MRE	AACCTAA	MYB binding site involved in light responsiveness	
Sp1	GGGCGG	light responsive element	
TCT-motif	TCTTAC	part of a light responsive element	
SrCRE	ARE	AAACCA	<i>cis</i> -acting regulatory element essential for the anaerobic induction

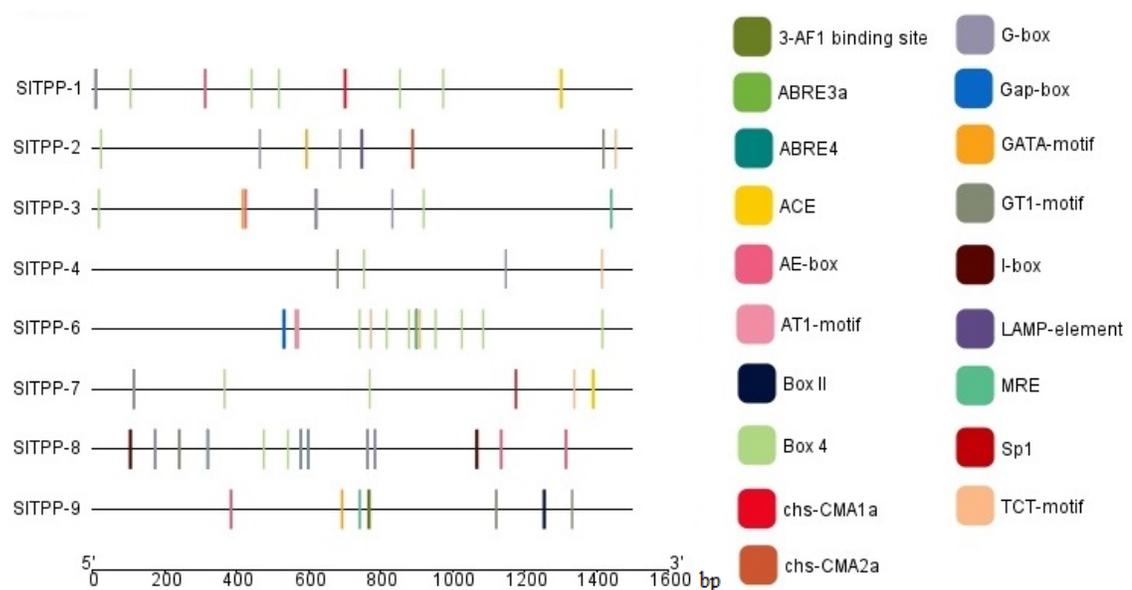
as-1	TGACG	Defense response
box S	AGCCACC	elicitation; wounding and pathogen responsiveness
CCAAT-box	CAACGG	MYBHv1 binding site (involved in the activation of stress and defense genes in plants)
DRE core	GCCGAC	<i>cis</i> -acting element that regulates cold- and dehydration-responsive gene expression
LTR	CCGAAA	<i>cis</i> -acting element involved in low-temperature responsiveness
MBS	CAACTG	MYB binding site involved in drought-inducibility
MYB	CAACCA/TAACCA/CAACAG	stress response element
MYB-like sequence	TAACCA	stress response element
MYB recognition site	CCGTTG	stress response element
Myb	TAACTG/CAACTG	stress response element
Myb-binding site	CAACAG	stress response element
MYC	CATGTG/CATTTG/CAATTG	<i>cis</i> -acting element involved in drought responsiveness
Myc	TCTCTTA	stress response element
STRE	AGGGG	stress response element
TC-rich repeats	GTTTTCTTAC/ATTCTCTAAC	<i>cis</i> -acting element involved in defense and stress responsiveness
Unnamed__1	CGTG	stress response element
W box	TTGACC	fungus elicitor responsive element
WRE3	CCACCT	stress response element

	WUN-motif	AAATTACT/AAATTTCTT/CC ATTTCAA	stress response element
HrCRE	ABRE	ACGTG/CACGTG/CGTACGT GCA	<i>cis</i> -acting element involved in the abscisic acid responsiveness binding site of AT-rich DNA binding protein (ATBP-1) element for maximal elicitor-mediated activation (2copies)
	AT-rich element	ATAGAAATCAA	<i>cis</i> -acting regulatory element involved in auxin responsiveness
	AT-rich sequence	TAAAATACT	<i>cis</i> -acting regulatory element involved in the MeJA-responsiveness
	AuxRR-core	GGTCCAT	gibberellin-responsive element
	CGTCA-motif	CGTCA	Ethylene-responsive element
	GARE-motif	TCTGTTG	<i>cis</i> -acting regulatory element involved in zein metabolism regulation
	ERE	ATTTTAAA/ATTTCATA	gibberellin-responsive element
	O2-site	GATGACATGG	<i>cis</i> -acting element involved in gibberellin-responsiveness
	P-box	CCTTTTG	salicylic acid responsive element
	TATC-box	TATCCCA	<i>cis</i> -acting element involved in salicylic acid responsiveness
	TCA	TCATCTTCAT	auxin-responsive element
	TCA-element	CCATCTTTTT	<i>cis</i> -acting regulatory element involved in the MeJA-responsiveness
	TGA-element	AACGAC	<i>cis</i> -element involved in seed specific expression
TGACG-motif	TGACG	<i>cis</i> -acting regulatory element related to meristem expression	
CdCRE	AAGAA-motif	GAAAGAA/gGTAAAGAAA	<i>cis</i> -acting regulatory element related to meristem specific activation
	CAT-box	GCCACT	
	CCGTCC-box		

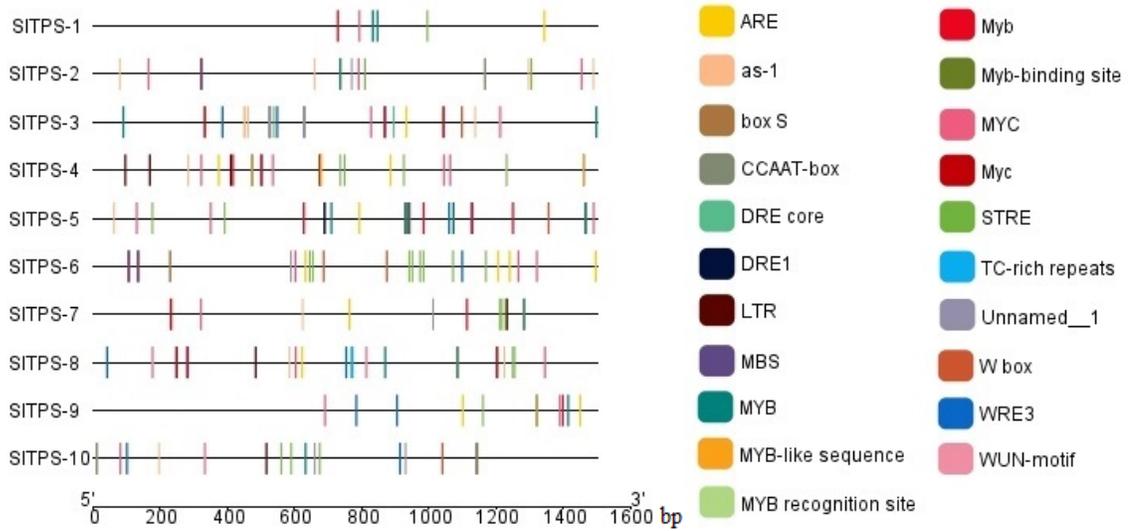
	circadian	CAAAGATATC	<i>cis</i> -acting regulatory element involved in circadian control
	GCN4_motif	TGAGTCA	<i>cis</i> -regulatory element involved in endosperm expression
	Unnamed__4	CTCC	temporal and spatial specificity of DYT1(Anther) expression
Otros CRE	A-box	CCGTCC	<i>cis</i> -acting regulatory element
	AP-1	TGAGTTAG	Stress response element
	Box III	atCATTTTCACt	protein binding site
	CTAG-motif	ACTAGCAGAA	Unknown
	Unnamed__2	AACCTAACCT	Unknow
	Unnamed__5	TGTAATAATATATTTATATT	SEF1 factor binding site
	Unnamed_6	taTAAATATct	Unknown
	Unnamed__1	GAATTTAATTAA	60K protein binding site
	Unnamed__1	GGATTTTACAGT	<i>cis</i> -acting element involved in phytochrome down-regulation expression



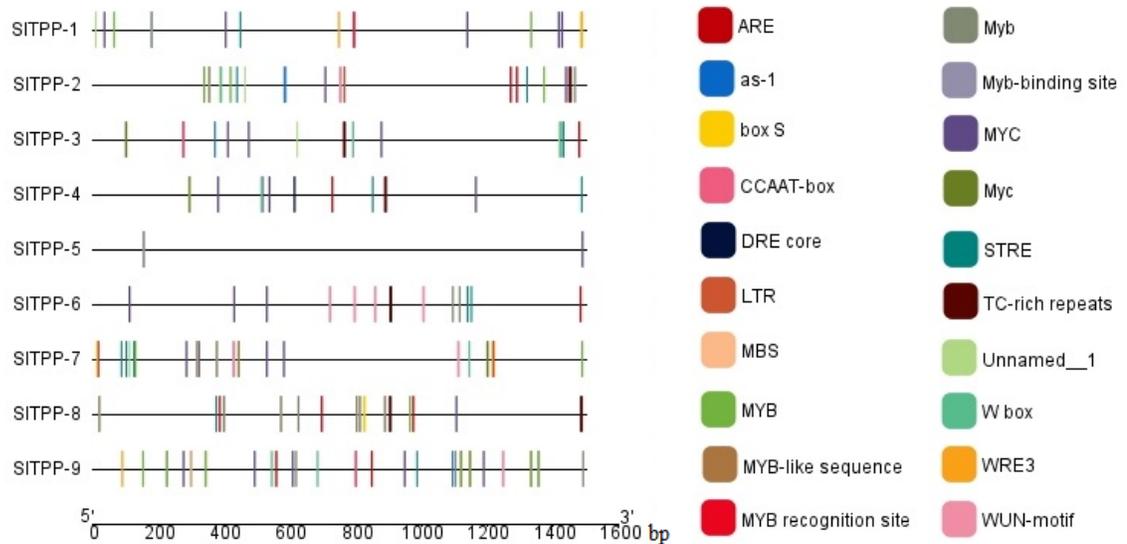
Apéndice 14. Elementos *cis* relacionados con la luz (LrCRE) encontrados en *SITPSs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



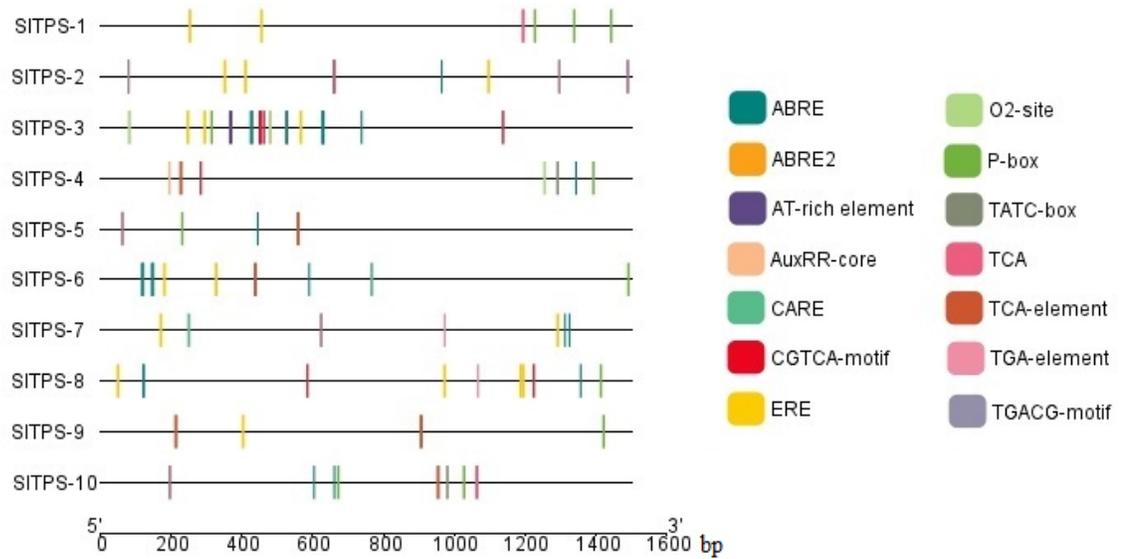
Apéndice 15. Elementos *cis* relacionados con la luz (LrCRE) encontrados en *SITPPs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



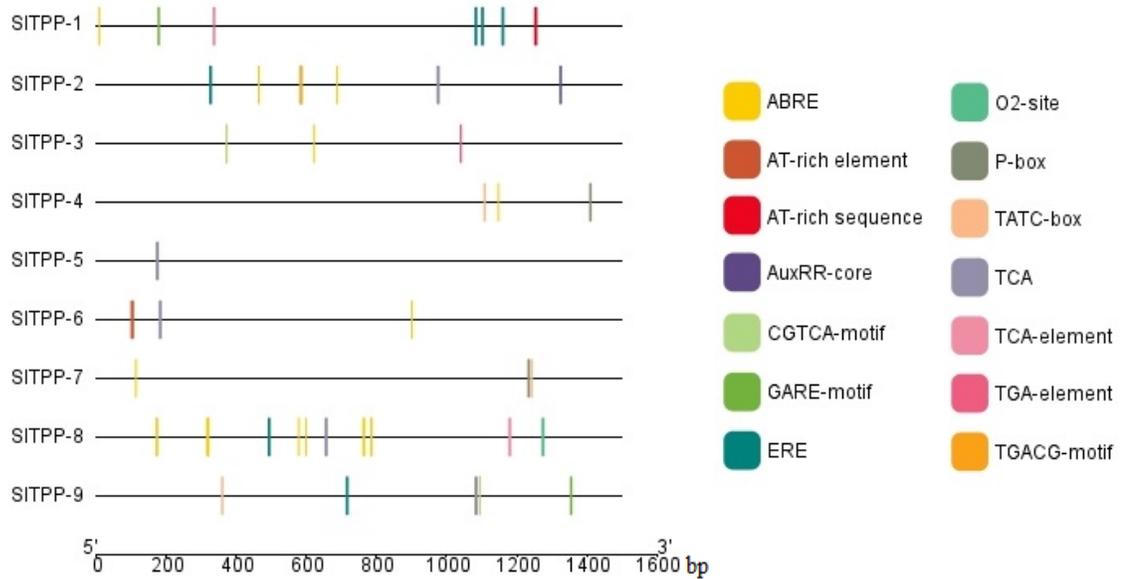
Apéndice 16. Elementos *cis* relacionados con el estrés (SrCRE) encontrados en *SITPSs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



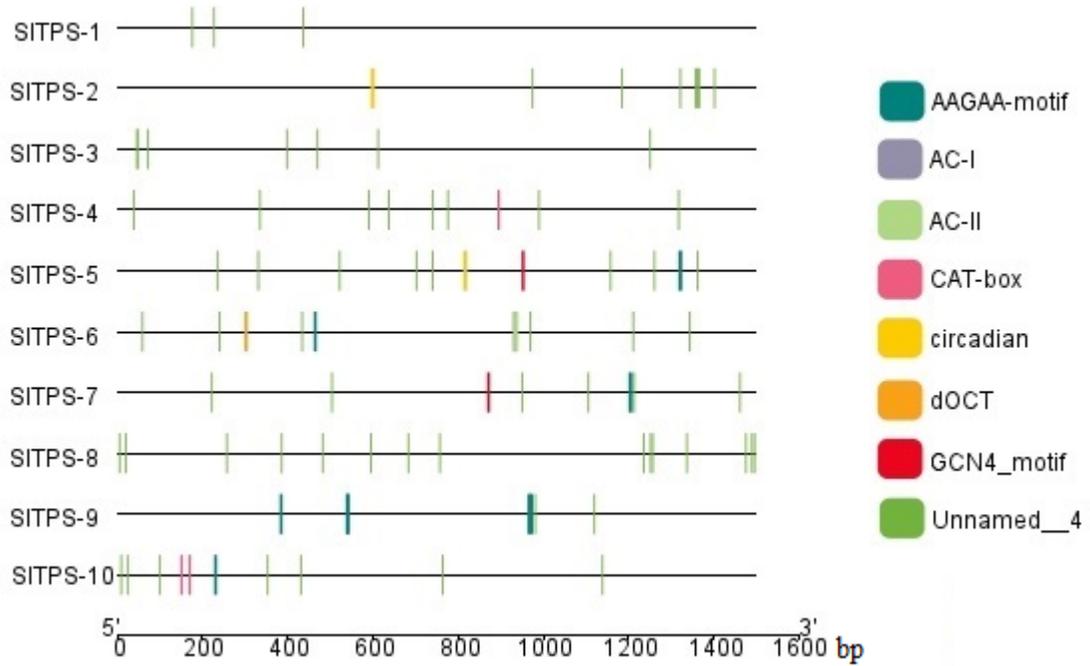
Apéndice 17. Elementos *cis* relacionados con el estrés (SrCRE) encontrados en *SITPPs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



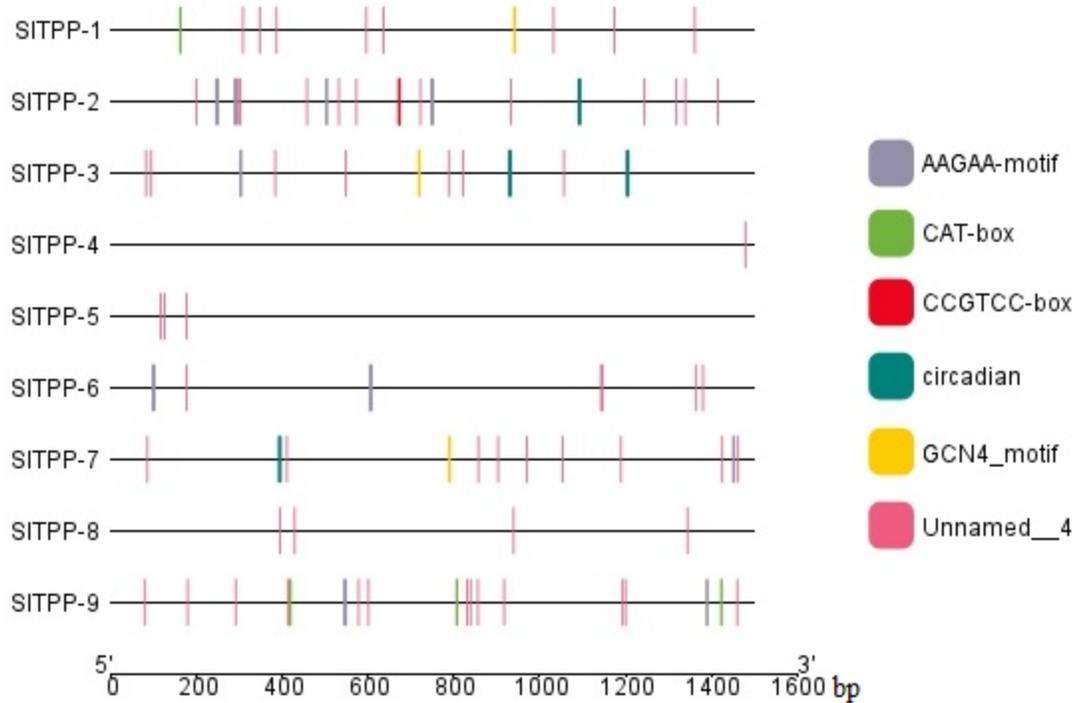
Apéndice 18. Elementos *cis* relacionados con las hormonas (HrCRE) encontrados en *SITPSs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



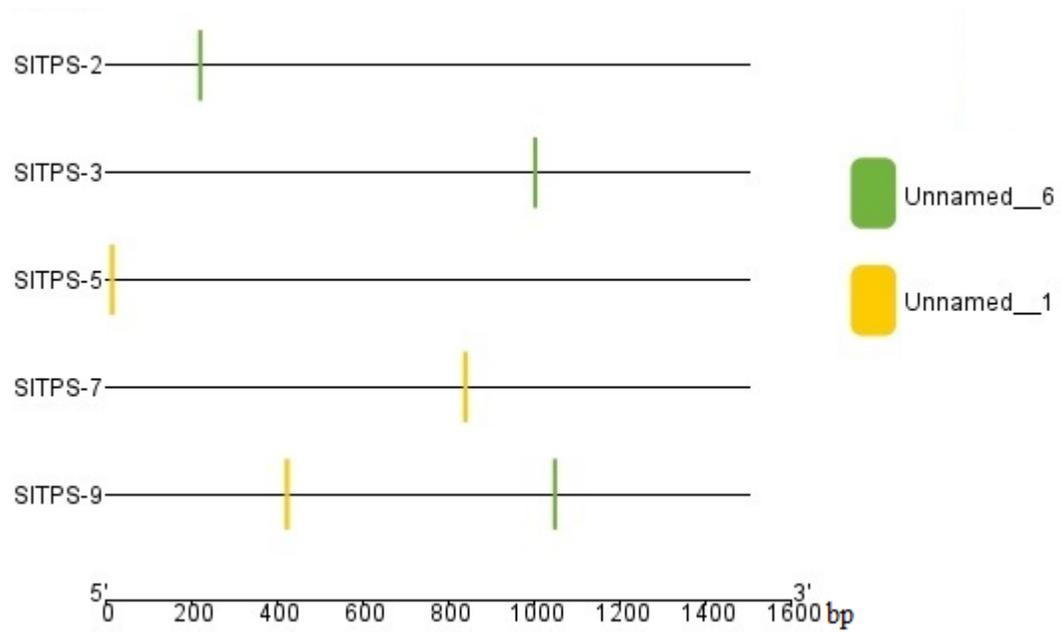
Apéndice 19. Elementos *cis* relacionados con las hormonas (HrCRE) encontrados en *SITPPs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



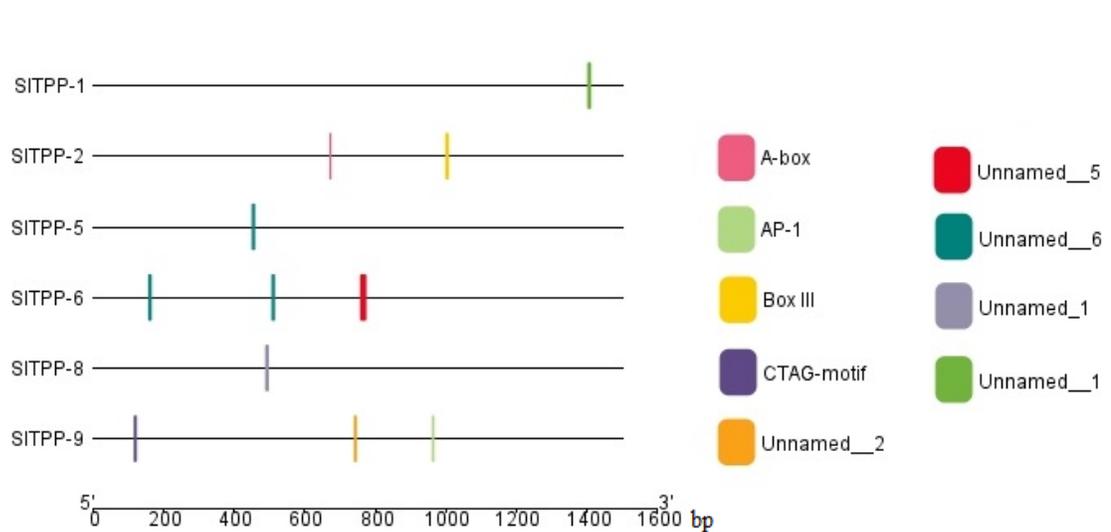
Apéndice 20. Elementos *cis* relacionados con el desarrollo celular (CdCRE) encontrados en *SITPSs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



Apéndice 21. Elementos *cis* relacionados con el desarrollo celular (CdCRE) encontrados en *SITPPs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



Apéndice 22. Otros elementos *cis* encontrados en *SITPSs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.



Apéndice 23. Otros elementos *cis* encontrados en *SITPPs*. Se analizaron 1500 bp aguas arriba del codón de inicio. El nucleótido -1 está en 1500.