



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE QUÍMICA**

**DETERMINACIÓN *IN SILICO* DE ELEMENTOS  
REGULATORIOS -CIS EN GENES  
PERTENECIENTES A LA FAMILIA TRITHORAX  
EN PLANTAS.**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE  
QUÍMICO FARMACÉUTICO BIÓLOGO**

**PRESENTA:**

**JUAN LUIS RODRÍGUEZ NORIA**



**MÉXICO, D.F.**

**2011**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**JURADO ASIGNADO:**

**PRESIDENTE:** Profesor: JESUS FERNANDO MONTIEL AGUIRRE

**VOCAL:** Profesor: EDUARDO BONILLA ESPINOSA

**SECRETARIO:** Profesor: RAUL ALVAREZ VENEGAS

**1er. SUPLENTE:** Profesor: RAQUEL ORTEGA MUÑOZ

**2° SUPLENTE:** Profesor: JOSE IGNACIO PARAMO RAMIREZ

**SITIO DONDE SE DESARROLLÓ EL TEMA:** CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL I.P.N., UNIDAD IRAPUATO.

**ASESOR DEL TEMA:** DR. RAÚL ÁLVAREZ VENEGAS



**SUSTENTANTE :** JUAN LUIS RODRÍGUEZ NORIA



## DEDICATORIA

*A mis hijos, José Aarón y Abigali que son la fuerza impulsora de mi vida y a quienes dedico este trabajo, como muestra del fruto de su amor, apoyo y comprensión y como ejemplo de que se puede lograr lo que se quiere con el apoyo incondicional de la gente que realmente nos ama.*

*A mi esposa Lizbeth Jiménez Velásquez por su amor, comprensión y cariño, por acompañarme y apoyarme en todo momento, pues sin ella, su ayuda y su confianza no hubiera podido llegar hasta aquí.*

*A mis padres, Josefina Noria Valadez y José Rodríguez Conejo por su amor, confianza, comprensión y todo el apoyo que me brindaron cuando más lo necesite, ellos fueron y son, gracias a Dios, los maestros que me han impulsado a ser un hombre de valores y de bien.*

*A mis hermanos: María Elena, Oscar, Marta, José Nolasco, Carlos, Ramón, Sanjuana y María Guadalupe, por todo el respeto, apoyo y cariño que me han brindado lo cual nos ha mantenido juntos en las buenas y en las malas y a quienes debo mucho de lo que soy.*

*A mis abuelos, aunque físicamente ya no están, por haberme compartido sus experiencias y enseñanzas que enriquecieron mi vida y que sin ellos yo no estaría aquí.*

*A toda mi familia y conocidos que directa o indirectamente participaron en mi formación.*

## **AGRADECIMIENTO ESPECIAL.**

“Este proyecto fue realizado gracias al apoyo del CONACYT, proyecto de investigación CB-2006/55028, otorgado al Dr. Raúl Álvarez Venegas.”

Al Dr. Raúl Álvarez Venegas por su gran apoyo, confianza, paciencia y dedicación para conmigo y el proyecto, por permitirme ser parte del equipo del laboratorio de Cromatina y Epigenética.

## **AGRADECIMIENTOS.**

Al Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unida Irapuato, por el apoyo y la confianza de permitirme llevar a cabo este proyecto de tesis en sus instalaciones, el cual comparto con sus profesores, investigadores y estudiantes.

Al laboratorio de Cromatina y Epigenética, del Departamento de Ingeniería Genética del Cinvestav Unidad Irapuato por el apoyo y la confianza para la realización de este proyecto.

A los miembros actuales del laboratorio de Cromatina y Epigenética: Dr. Raúl Álvarez Venegas (Investigador Principal), M. en C. María del Pilar Valencia Morales (Estudiante de Doctorado), Dr. Francisco Luna Martínez (Posdoctorante).

A los miembros pasados del laboratorio de Cromatina y Epigenética: Dr. José Alberto Camas Reyes (Posdoctorado), María del Pilar Valencia Morales (Maestría), Juan Manuel Quiceno Rico (Maestría) y José Juan Torres Martínez (Tesis de Licenciatura). A todos ellos agradezco sus comentarios, observaciones y aportaciones para este proyecto.

## **AGRADECIMIENTOS.**

A los miembros del jurado asignado: Jesús Fernando Montiel Aguirre, Eduardo Bonilla Espinosa, Raúl Álvarez Venegas, Raquel Ortega Muñoz y José Ignacio Paramo Ramírez por sus observaciones y aportaciones a esta tesis, lo cual los hace miembros directos y distinguidos de ella.

A la QFB Norma Castellanos Chávez y a su personal por su apoyo y atenciones en la culminación de este proyecto.

A la Universidad Nacional Autónoma de México y en especial a la Facultad de Química por haberme permitido ser parte de esta familia tan especial y única, lo cual me enorgullece tanto...

A mis maestros y compañeros en la Facultad de Química por permitirme ser parte de su vida y por compartir sus conocimientos y experiencias.

*... y por sobre todas las cosas agradezco a DIOS por ser mi guía, por acompañarme siempre y por permitirme llegar hasta aquí, tener lo que tengo y ser lo que soy.*

## ÍNDICE

INTRODUCCIÓN .....	1
1. OBJETIVOS .....	3
2. GENERALIDADES .....	4
2.1. Genética versus Epigenética .....	6
2.2. La plantilla de la cromatina .....	8
2.3. Modificaciones de las histonas y el código de histonas .....	9
2.4. Dominio SET .....	10
2.5. Grupos Polycomb y Trithorax .....	12
2.6. Proteínas de dominio SET en plantas.....	13
2.7. Arabidopsis SET (AtSET) .....	15
2.8. Clase II-ash1 clase (H3K36 metilación).....	16
2.9. Clase III-trithorax clase (H3K4 metilación).....	17
2.10. Transcripción en eucariotas. Regulación y expresión.....	19
2.11. Elementos reguladores –cis.....	20
3. MATERIALES Y MÉTODOS .....	23
4. RESULTADOS .....	32

5. DISCUSIÓN DE RESULTADOS .....	57
5.1. Análisis de secuencias–cis en <i>Arabidopsis thaliana</i> .....	57
5.2. Análisis de los elementos –cis en genes de la familia trithorax en tres distintos genomas de plantas .....	59
6. CONCLUSIONES.....	70
6.1. Sugerencias.....	72
7. ANEXOS .....	73
Anexo 1. Gráficas que muestran la frecuencia y la presencia de los elementos – cis en los promotores de genes de <i>Arabidopsis thaliana</i> .....	74
Anexo 2. Resumen de secuencias con sus respectivos parámetros estadísticos utilizados para el criterio de selección (HSPs).....	86
Anexo 3.Tabla que concentra los elementos –cis encontrados por especie y por parejas.....	91
Anexo 4A y 4B.Tablas que muestra a los elementos -cis y su presencia- frecuencia en cada una de las secuencias de los promotores seleccionados para las tres especies estudiadas.....	95
Anexo 5. Diccionario de elementos –cis, sus secuencias y sus posibles funciones reportadas actualmente en las bases de datos consultadas .....	130
8. BIBLIOGRAFÍA .....	138

## INTRODUCCIÓN

No todos los genes son expresados en todo momento. Genes diferentes son requeridos para completar los procesos de desarrollo y maduración sexual de la planta. En general, dos tipos de elementos regulatorios de la expresión génica han sido descritos: *cis* y *trans*. Los elementos regulatorios-*cis*, o elementos –*cis*, son regiones del DNA o RNA que regulan la expresión de genes localizados en la misma molécula de DNA. Estos elementos son a menudo elementos de unión para uno o más factores “trans-acting”. Un elemento –*cis* puede estar localizado en la región 5’ de la secuencia codificante del gen que controla (e.g. promotor), en un intrón, o en la región 3’ de la secuencia codificante (tanto en la región no traducida, como en la región no transcrita). Por el contrario, los elementos trans-regulatorios son las secuencias de DNA que codifican para factores de transcripción (“trans-acting elements”), los cuales son factores difusibles (usualmente proteínas), que pueden modificar la expresión de genes distantes a la región de donde se transcribieron.

De acuerdo a los modelos actuales, cambios en las secuencias regulatorias –*cis* de genes duplicados pueden contribuir a cambios específicos en los patrones de expresión de genes parálogos. Nuevos sitios de unión a factores de transcripción o combinaciones de dominios en la región promotora pueden gobernar la expresión específica, espacio-temporal, de los distintos parálogos. Estos cambios en los patrones de expresión entre parálogos ha permitido la divergencia funcional de los mismos.

El objetivo principal del proyecto consiste en la determinación y análisis *in silico* de los elementos –*cis* presentes en la región genómica de factores epigenéticos en plantas. Particularmente, de secuencias presentes en genes que codifican para proteínas con el dominio SET [Su(var)3-9, E(z), Trithorax], con actividad de metil-transferasas de los grupos amino terminales de las histonas.

Los procesos que químicamente modifican la estructura de las histonas del nucleosoma y por consiguiente la cromatina, son consideradas como marcas epigenéticas que pueden activar o reprimir el proceso de la transcripción.

Los análisis *in silico* actualmente permiten la identificación de los promotores de genes y sus elementos –cis, entre otras de sus aplicaciones.

## 1. OBJETIVO GENERAL

Determinar los elementos-cis presentes en la región genómica de factores epigenéticos pertenecientes al grupo Trithorax en plantas.

### OBJETIVOS ESPECÍFICOS.

1.1. Determinación y análisis *in silico* de los elementos -cis presentes en promotores de genes de *Arabidopsis thaliana* que codifican para proteínas con el dominio SET [Su(var)3-9,E(z),Trithorax] de la familia Trithorax.

1.2. Comparación y cuantificación de los elementos -cis presentes en genes que codifican proteínas con el dominio SET de la familia Trithorax, en tres distintos genomas de plantas.

## 2. GENERALIDADES

La identificación de los elementos reguladores es uno de los problemas más relevantes en los estudios de Biología Molecular. En los últimos años grandes avances se han hecho desde el campo experimental, esto ha permitido el acercamiento a muchos de los aspectos de la regulación de la transcripción. Además, dada la importancia actual de los medios computacionales en el análisis de datos de bio-secuencias, se han desarrollado varias aproximaciones, *in silico*, para detectar motivos reguladores dentro de diferentes genomas, como herramienta para identificar genes, en grupos de genes co-regulados, y genes ortólogos entre otros (Mejía-Guerra, M. y Lareo, L., 2006).

El nivel más frecuente abordado para la predicción de promotores *in silico* es la búsqueda de elementos *cis*-reguladores organizados modularmente dentro de los promotores; sin embargo, los genes pueden ser regulados en muchas formas involucrando interacciones entre varios elementos reguladores (Hu, 2000).

La región promotora de un gen está constituida por un grupo de secuencias específicas (elementos *cis*-reguladores) que pueden interactuar con un conjunto de proteínas específicas llamadas factores de transcripción (elementos *trans*-reguladores), con el fin de regular la transcripción de un gen. Esta región se encuentra por lo general al extremo 5'-UTR del sitio de inicio de la transcripción y se ha dividido en varias partes, la parte que contiene al sitio de inicio y la que al parecer controla la transcripción a nivel basal (denominada región promotora núcleo), en esta se encuentran una serie de motivos de DNA más o menos conservado en todos los eucariontes a los que se unen distintos factores de transcripción (TFs) para lograr así la unión de la RNA polimerasa II.

Los motivos *cis*-reguladores componentes de los elementos promotores son motivos cortos, usualmente entre los 5 y los 20 nucleótidos de longitud (Wingeder, 1998). Las secuencias reguladoras están formadas por componentes modulares que otorgan un alto nivel de especificidad para un número relativamente pequeño de diferentes factores de transcripción.

Para el entendimiento adecuado de la función génica, se debe centrar la atención en los patrones de regulación de secuencias y no en los elementos individuales. La búsqueda de tales secuencias debería ser más efectiva en la anotación de nuevas secuencias en comparación con el reconocimiento de motivos individuales. En la actualidad existen bases de datos para los elementos cis-reguladores y los factores de transcripción (Mejía-Guerra, M. y Lareo, L., 2006).

Numerosos enfoques, tanto experimentales como *in silico*, se han desarrollado para descubrir los elementos cis-reguladores, por ejemplo: inmuno-precipitación de la cromatina junto con “microarrays / chip-on-chip” proporciona evidencia experimental directa de las interacciones proteína-ADN en la escala del genoma y es un enfoque poderoso. Sin embargo, el método chip-on-chip no es en la actualidad fácilmente aplicable en muchos eucariotas superiores. Otros métodos experimentales establecidos tales como deleciones de promotores escalonadas o huellas DNAsa I ofrecen puntos de vista de alta resolución de los promotores individuales, pero son inviables para el análisis a gran escala. Para superar las limitaciones experimentales, métodos computacionales se han desarrollado con el tiempo para complementar el descubrimiento de motivos de gran escala. Estos incluyen el mapeo de los conocidos, así como la detección de los motivos de novo, por ejemplo. En general, dos tipos de conjunto de datos se utilizan para las búsquedas de motivos como fuentes de información: primero, un grupo de una relación funcional conocida para que las secuencias del promotor de genes co-expresados de un grupo (especie) se complementen con las secuencias de los promotores ortólogos de una o más especies relacionadas. En el primer caso, los motivos candidatos se espera que sean enriquecidos (estadísticamente significativos) en comparación con un modelo de fondo estadístico. Por lo tanto, pueden ser detectados por su excesiva representación en el grupo de gen respectivo. En este último se espera que las regiones no funcionales sean mucho más diversificadas en comparación con elementos cis-funcionales. En un enfoque ampliamente aplicado, sitios candidatos aparecen como patrones de conservación o huellas filogenéticas de alineamientos (locales) entre secuencias evolutivas relacionadas. Además del uso de una sola fuente informativa (la

sobrerrepresentación o la co-expresión), se han desarrollado diversas herramientas que combinan la conservación de la información y la co-expresión (Wang, X., Haberer, G., y Mayer, K.FX., 2009).

En la última década, una serie de herramientas computacionales ha sido desarrollada para analizar a los promotores de diversos organismos. Estas herramientas se dividen en tres categorías. Las herramientas en la primera categoría descubren o identifican a los promotores de genes a partir de secuencias de nucleótidos. Las herramientas en la segunda categoría predicen los supuestos elementos-cis en el promotor de una familia de genes utilizando el descubrimiento de patrones y técnicas de coincidencia de patrones; estas herramientas describen y comparan tanto métodos enumerativos como métodos probabilísticos. Las herramientas en la tercera categoría modelan y analizan la presencia de combinaciones de elementos-cis en promotores de genes y el efecto de estas combinaciones en la regulación de la transcripción.

## 2.1. GENÉTICA VERSUS EPIGENÉTICA.

La determinación de los detalles estructurales del ADN de doble hélice se erige como uno de los descubrimientos destacados en la Biología. El ADN es la primera macromolécula que almacena información genética (Avery et al. 1944), y se propaga esta información almacenada para la próxima generación a través de la línea germinal. De los hallazgos de este y otros, surgió el "dogma central" de la Biología Moderna. Este dogma encapsula los procesos que intervienen en el mantenimiento y la traducción de la plantilla genética necesaria para la vida. Las etapas esenciales son: (1) la auto-reproducción del DNA por replicación semiconservativa, (2) la transcripción unidireccional 5' a 3', con plantilla por el código genético (ADN) con la generación de un intermediario de ARN mensajero (ARNm); (3) la traducción de ARNm para producir polipéptidos que consiste en cadenas de aminoácidos que son colineales con el orden 5' a 3' de ADN. En términos simples:  $ADN \leftrightarrow ARN \rightarrow \text{proteína}$  (Allis, et al., 2007).

El origen de la Epigenética se deriva desde hace mucho tiempo de estudios aparentemente anómalos (es decir, no mendelianos) y las distintas estructuras de la herencia en muchos organismos. La herencia mendeliana clásica de los rasgos fenotípicos (por ejemplo, el color del guisante, número de dígitos, o insuficiencia de la hemoglobina) son resultados de las diferencias alélicas causadas por mutaciones en la secuencia de ADN. En conjunto, las mutaciones definen la base de los rasgos fenotípicos, lo que contribuye a la determinación de los límites de las especies. Estos límites entonces se forman por las presiones de la selección natural, como es explicado por la teoría de la evolución de Darwin. Tales conceptos colocan estas mutaciones en el corazón de la Genética clásica. Por el contrario, la herencia no mendeliana (por ejemplo, la variación del crecimiento embrionario, coloración de la piel de mosaico, inactivación del cromosoma X [al azar], o paramutación en plantas) puede manifestarse, por ejemplo, con la expresión de un solo (de dos) alelo en el mismo entorno nuclear. Es importante destacar que, en estas circunstancias, la secuencia de ADN no es alterado. Esto es distinto al del otro comúnmente patrón de herencia no mendeliana que surge de la herencia materna de las mitocondrias (Birky 2001) citado por Allis, et al., (2007).

El aparato molecular ha evolucionado para controlar la expresión del gen. Los organismos eucariotas transportan una tremenda cantidad de información genética, solo para poner en código de 6000 a 100,000 proteínas necesarias para sostener la vida de levaduras a animales. Además, los genomas también contienen cantidades inmensas de reguladores-cis y reguladores-trans que contribuyen a la expresión del gen, y que son además, responsables de dirigir modelos espaciales y temporales de la expresión del gen en respuesta a requerimientos metabólicos, programas de desarrollo y plétora de estímulos externos. Para mantener semejante carga genética los eucariotas han organizado su ADN co-lineal en nucleosomas discretos, cada uno empaquetado en la cromatina donde el ADN se organiza y empaqueta en el núcleo de la célula eucariota.

## 2.2. LA PLANTILLA DE LA CROMATINA.

Recientemente, el estudio sobre las modificaciones de la cromatina en eucariotas se ha convertido en un tema de investigación importante. Esto se debe principalmente a la estructura de orden superior de la cromatina que tiene un papel importante en el control de la expresión génica y en el mantenimiento de la integridad del genoma. La cromatina es una estructura compleja construida a partir de unidades de repetición, los nucleosomas (Yu, et al., 2009).

Cada nucleosoma está compuesto por 147 pares de bases (pb) de ADN el cual se arregla dos veces alrededor del octámero de histonas compuesto por dos copias de cada histona H2A, H2B, H3 y H4. Ahora está claro que el montaje de la cromatina ejerce una gran influencia en la expresión de genes que afectan a la accesibilidad de la maquinaria transcripcional, incluyendo complejos de ARN polimerasa y factores de transcripción, en el ADN. Algunos estudios demostraron que el ADN organizado en el nucleosoma no es accesible a la maquinaria transcripcional. Este efecto se ve favorecido cuando las lisinas del extremo amino-terminal no sufren modificaciones, de hecho las cargas positivas de estos residuos permiten la interacción electrostática con las cargas negativas de los grupos fosfato del ADN, generando una organización más cerrada y compacta (Pontvianne et al., 2010).

Las interacciones internucleosomales se ven favorecidas por la quinta histona, la histona H1 y otras proteínas no histónicas como las proteínas de alta movilidad electroforética (HMGs). Así, se organiza el nivel superior de la cromatina generado por la reunión de varios nucleosoma, y formando una hebra de 30 nm de diámetro. Cada hebra se reúne entre sí varias veces para generar otra de mayor grosor y formar parte de una cromátida de un cromosoma que representa el mayor nivel de estructuración de la cromatina.

“Históricamente, la cromatina se ha clasificado como heterocromática o eucromática, derivados de la tinción nuclear con patrones de colorantes utilizados por citólogos para visualizar el ADN. La eucromatina es cromatina descondensada, aunque puede ser transcripcionalmente activa o inactiva. La heterocromatina en términos generales se puede definir como cromatina altamente compactada y silenciada. Puede existir como cromatina en forma permanente silenciada (heterocromatina constitutiva), donde los genes rara vez se puede expresar en cualquier tipo de célula del organismo, o reprimidos (heterocromatina facultativa) en algunas células durante un ciclo celular específico o estadio de desarrollo” (Allis, et al., 2007, p 31).

### 2.3. MODIFICACIONES DE LAS HISTONAS Y EL CÓDIGO DE HISTONAS.

“La comprensión de la naturaleza y la regulación de las enzimas responsables de la modificación de determinados residuos de aminoácidos en las histonas del nucleosoma es esencial para descifrar el código de la histona” (Ng, et al, 2007, p 317). Las histonas, como el núcleo estructural del nucleosoma, están sometidas a múltiples tipos de modificaciones covalentes, tales como acetilación, metilación, fosforilación, ubiquitinación, sumoilación, citrulinación y ADP-ribosilación. La gran variedad de modificaciones en las histonas, que potencialmente confieren información regulatoria, ha sido hipotetizado para constituir el llamado “código de histonas” (Junuwein y Allis, 2001; Kouzarides, 2007) en Pontvianne et al., (2010).

Estas modificaciones adicionales a las histonas representan información epigenética sobre la cromatina y se han propuesto para formar un "código de histonas" que pueden ser reconocidas por proteínas regulatorias específicas o complejos de proteínas que participan en la regulación de la actividad transcripcional de un gen incorporado (Yu, et al., 2009). La hipótesis del código de las histonas sostiene que la modificación covalente post-traducciona de las histonas son interpretadas por la célula para producir una rica combinación transcripcional. Así, descifrar la compleja combinatoria de modificaciones de histona es crucial para un entender más claro de la regulación transcripcional.

La metilación de la lisina de las histonas puede ocurrir en los residuos de lisina 4, 9, 27, 36 y 79 de la histona H3, y en la lisina 20 de la histona H4. Por otra parte, los residuos de lisina pueden ser mono-, di-, o trimetilada, que añaden otra capa de complejidad de información del código de las histonas. En general, la metilación en H3K4, K36 y K79 se correlaciona con la transcripción activa (esta última no ha sido reportada en plantas), mientras que la metilación de H3K9, K27, y asociados H4K20 se correlaciona con heterocromatinización y el silenciamiento de genes. En la actualidad, proteínas responsables de la metilación de las histonas de residuos específicos de lisina se han caracterizado. Aparte de DOT1 y DOT1L, que metilan H3K79 en la región globular de la histona, todas las metiltransferasas de histonas (HMTasas) contienen un dominio SET conservado de cerca de 130 aminoácidos de longitud (Yu, et al., 2009).

#### 2.4. DOMINIO SET.

Varios estudios indican que las proteínas de dominio SET tienen actividad de metiltransferasa de histonas (HMTasa) y están involucrados en la regulación de la expresión génica de forma positiva, o negativamente, por catalizar diferentes sustratos en la histona. El nombre del dominio SET, que fue caracterizado por Stephens et al. (1998), proviene de las tres proteínas fundadoras identificadas en *Drosophila*: Supresor de la variegación 3,9 [Su (var) 3-9], Enhancer of zeste [E (z)], y Trithorax (Trx): SET [Su(var)3-9, E(z), Trithorax].

El dominio SET, una secuencia que abarca de 130-160 aminoácidos que se cree ha evolucionado de un motivo antiguo encontrado en proteínas bacterianas (Álvarez-Venegas, et al., 2007), es encontrado en proteínas que son miembros del grupo Polycomb (PcG), grupo Trithorax (TrxG) y Su(var). Homólogos de estos tres genes han sido identificados en levaduras y mamíferos. Los primeros genes en plantas identificados que codifican proteínas con el dominio SET fueron homólogos E(z) (Baumbusch, et al, 2001). El descubrimiento del evolutivamente conservado dominio SET, en las mencionadas HKMTs, fue un gran paso hacia una completa comprensión de la regulación epigenética de la expresión del gen a través de metilación de las histonas.

Todas las HKMTs conocidas en plantas, particularmente en *Arabidopsis*, tienen el dominio SET que es responsable de la actividad catalítica de las enzimas. Por lo tanto, estas proteínas son miembros de una super-familia llamada SDG (Set Domain Group) (Gendler et al., 2008) (F Pontvianne et al., 2010).

El dominio SET posee un único pliegue dominado por 12  $\beta$ -hebras (Couture y Trievel, 2006). Los otros dos dominios, el pre-SET y el post-SET, a veces flanquean el dominio SET y puede facilitar las interacciones con los sustratos específicos de las histonas. El grupo hidroxilo de una tirosina altamente conservada en el dominio SET interactúa con el sustrato y transfiere un grupo metilo a la lisina utilizando S-adenosilmetionina (AdoMet) como donante del grupo metilo (Couture y Trievel, 2006; Rea et al, 2000; Xiao et al, 2003b) citado por F Pontvianne et al., (2010).

Mientras que en los mamíferos SUV39H1 fue la primera enzima identificada para demostrar que poseen actividad de histona metiltransferasa de lisina (HKMT) hacia H3K9, su homólogo, Su (var) 3-9, en *Drosophila* fue la primera que se identificó en un rastreo genético de un supresor del efecto de posición de variegación (PEV). El fenómeno de PEV (position effect variegation) se descubrió por HJ Muller en 1930 al describir el reordenamiento del gen de color de los ojos blancos, de una región eucromática a una heterocromática en regiones cromosómicas. Estos cambios conformacionales son el resultado de la variación de silenciamiento de la expresión del gen de célula a célula, expresión que lleva al fenotipo de color de ojos de mosaico en *Drosophila* (Ng et al., 2007).

En *Drosophila*, la represión de la transcripción o la activación de la expresión de genes homeóticos (o genes maestros) está mediada por dos distintos grupos antagonicos de los complejos de proteínas. La represión de la expresión de genes homeóticos está mediada por un grupo de proteínas complejas pertenecientes al Grupo Polycomb (PcG), por ejemplo: Enhancer de zeste [E(z)] y peines adicionales sexo (ESC, extra sex combs), que dirigen la metilación de H3K27 y H3K9. Por el contrario, el grupo de proteínas trithorax (trxG), que incluye a ASH1 y TRX, cuya función es la metilación de

H3K4, están involucradas en el mantenimiento de un estado activo de genes homeóticos.

## 2.5. GRUPOS POLYCOMB Y TRITHORAX.

“Aunque todas las células de un organismo en desarrollo tienen el mismo ADN, expresan diferentes genes y transmiten estos patrones de expresión génica de las células hijas a través de múltiples rondas de división celular. Esta memoria celular para los estados de expresión de los genes se mantiene por dos grupos de proteínas: las proteínas del grupo Polycomb (PcG), que establecen y mantienen estable el silenciamiento de genes y las proteínas del grupo trithorax (trxG), que contrarrestan el silenciamiento y permiten mantener la activación de genes” (Schuettengruber, et al., 2009, p 147).

El Grupo trithorax (trxG) es un conjunto de proteínas reguladoras de la cromatina que generalmente actúan para mantener la expresión génica. Este grupo de proteínas es bastante heterogénea y contiene los miembros que forman varios complejos multiprotéicos de la cromatina. Una actividad principal del complejo trithorax es para activar la transcripción mediante, por ejemplo, la inducción de trimetilación de lisina 4 de la histona H3 (H3K4me3) en determinados sitios de regulación de la cromatina en su objetivo. Este estado activo se ve reforzado por la acetilación de la histona H4. Además, algunas proteínas trxG tienen actividad en la remodelación de la cromatina, utilizando la energía del ATP para movilizar a los nucleosomas.

El Grupo Polycom (PcG) es la contraparte represiva de las proteínas del grupo trithorax en que también se unen a la cromatina en los genes Hox y actúan para inducir la trimetilación de la lisina 27 de la histona H3 (H3K27me3), para evitar la acetilación de la histona H4, y contrarrestar la remodelación del nucleosoma, manteniendo así los estados silenciados de la expresión génica.

Se desconoce cómo este equilibrio y cómo exactamente estas proteínas son reclutadas hacia las secuencias blanco. Mediante la cartografía de la distribución en todo el

genoma de los factores PcG y trxG y proteínas conocidas para reclutarlos a la cromatina, se encontró que los supuestos reclutadores no sólo localizan sitios de unión PcG, sino también uniones a muchas otras regiones genómicas que son en realidad los sitios de unión del complejo trithorax. Por último, se ha encontrado que los dos fragmentos de la proteína trithorax tienen diferentes distribuciones en los cromosomas, lo que sugiere que pueden tener distintas funciones nucleares. En *Drosophila*, las proteínas de PcG y TrxG son reclutadas a la cromatina por elementos regulatorios ADN llamados PcG y TrxG elementos de respuesta (PREs y TREs, respectivamente). Estos elementos demostraron dirigir la herencia epigenética de los estados de silenciamiento y activación de la cromatina en todo el desarrollo.

Se han identificado proteínas adicionales PcG / trxG que no forman parte del núcleo de estos complejos, pero son asociadas con ellos y, por tanto, pueden ser consideradas como proteínas asociadas PcG/trxG. Estas proteínas pueden existir en forma de moléculas individuales en la célula, pero también es posible que formen parte de complejos de proteínas que contienen otras adicionales, como proteínas PcG/trxG aún no identificadas. Los complejos PcG y trxG (excepto PhoRC) no se unen a sus secuencias DNA objetivo de una manera específica in vitro, pero son reclutados por secuencias PRE/TRE in vivo.

## 2.6. PROTEÍNAS DE DOMINIO SET EN PLANTAS.

Las plantas contienen un gran número de genes que codifican proteínas de dominio SET. De acuerdo con la anotación de las bases de datos Pfam y ChromDB, por lo menos 47, 35, y 37 proteínas del dominio SET están presentes en Arabidopsis, maíz y arroz, respectivamente. Además de los conocimientos actuales sobre las proteínas vegetales SET en Arabidopsis, maíz y arroz, existen pruebas recientes de que splicing alternativo puede ser relativamente común de genes provenientes de las plantas y considerar la posibilidad que tiene un papel funcional en la regulación de la metilación de la histona. Además, el posible papel del ARN antisentido en expresión de los genes de Arabidopsis SET deben discutirse (Ng, et al., 2007).

La finalización de la secuencia del genoma de *Arabidopsis thaliana* permite un análisis comparativo de los reguladores de la transcripción de los tres reinos eucarióticos. *Arabidopsis* dedica más del 5% de su genoma para codificar más de 1500 factores de transcripción, alrededor del 45% de los cuales provienen de familias específicas de plantas. Factores de transcripción de *Arabidopsis* que pertenecen a las familias comunes a todos los eucariotas no comparten similitud significativa con la de los otros reinos más allá de los dominios de unión de ADN conservado, muchos de los cuales han sido dispuestos en combinaciones específicas de cada linaje. La comparación del genoma revela la generación de la evolución de la diversidad en la regulación de la transcripción (Riechmann, et al., 2000).

Desde el punto de vista económico y agronómico el genoma de las plantas monocotiledóneas es de gran importancia, sin embargo, hasta hace poco sólo la secuencia del genoma del arroz estaba disponible. Estos enfoques limitados en genómica comparativa, y en particular a promotores de genes, para los que unos pocos cientos de ortólogos en monocotiledóneas se han descrito y analizado, había sido una limitante. Sin embargo, con la finalización y disponibilidad del genoma del sorgo, esta limitación ha sido superada y ahora estamos en condiciones de realizar estudios comparativos del genoma a gran escala entre genomas evolutivamente relacionados de monocotiledóneas. En la familia de las gramíneas de gran importancia económica, nuestro conocimiento de los sitios de unión del factor de transcripción y las redes de la transcripción es aún muy limitada. Con la finalización del genoma del sorgo y la secuencia del genoma de arroz disponible, los análisis comparativos de promotor ahora permiten la detección de elementos cis-conservados en el genoma (Wang et al, 2009).

## 2.7. Arabidopsis SET (AtSET).

“En *Arabidopsis thaliana*, 49 genes que codifican el dominio SET han sido identificados (www.chromDB.org; Baumbusch et al, 2001.; Ng et al., 2007). Del mismo modo, el genoma humano codifica para 50 proteínas SDG, incluyendo 24 HKMTs. Por el contrario, el genoma de la levadura *Saccharomyces cerevisiae* codifica sólo cuatro proteínas SDG (Allis et al., 2007). De las 49 proteínas SDG de *Arabidopsis thaliana*, 31 se piensa que tienen actividad HKMT y se pueden dividir en cinco clases (I a V), en función de la arquitecturas de su dominio y / o diferencias en la actividad enzimática (Baumbusch et al, 2001; Ng et al, 2007; Springer et al, 2003). Existe otra clase (VI) que contiene proteínas SDG con un dominio SET interrumpido, y una última clase VII, que incluye proteínas SDG que metila proteínas no histónicas (Ng et al, 2007). Los análisis filogenéticos de *Arabidopsis thaliana* y los genes de *Zea mays* indican que la mayor parte de la duplicación de genes y la diversificación funcional de eventos que dieron lugar a la familia de proteínas SDG se produjo antes de la divergencia de las monocotiledóneas y las plantas dicotiledóneas (Ng et al, 2007; Springer et al, 2003)” (Pontvianne et al., 2010, p 4).

Para Ng, et al., (2007) “En las plantas, las proteínas Arabidopsis SET (AtSET) son el mejor grupo anotado y caracterizado. Baumbusch, et al., (2001) fueron los primeros en clasificar a 37 genes putativos de Arabidopsis SET en cuatro clases diferentes: (1) Potenciador de zeste [E (z)] homólogos, (2) Ash1 homólogos y relacionados, (3) trithorax (trx) homólogos y asociados, y (4) supresor de variegación [Su (var)] homólogos y afines. Posteriormente, Springer et al (2003) clasificaron 22 proteínas de maíz y 32 de Arabidopsis con dominio SET en 19 grupos distribuidos en cinco clases de acuerdo a sus relaciones filogenéticas y la organización del dominio”. Por último, Ng, et al, (2007) agrupan las proteínas provenientes de las plantas con dominio SET en las clases I-VII: clase I-E (z), la clase II-ASH1; Clase III-trithorax; clase IV-proteínas con un SET y un dominio PHD; clase V-Su (var), la clase VI-proteínas con un dominio SET interrumpido; clase VII-no-histonas metiltransferasas y proteínas relacionadas. Excepto por la adición de dos clases (VI y VII), este análisis es consistente con la clasificación de Springer et al. Aunque aún queda mucho por verificar experimentalmente, en

general, parece que los grupos resultantes reflejan las especificidades de sustrato de los miembros.

El dominio SET, como ya se ha mencionado, es una secuencia de 130-160 aminoácidos, es encontrado en proteínas que son miembros de PcG, TrxG y Su(var) y fueron llamados después los genes Su(var)3-9, Enhancer of zeste [E(z)] y Trithorax (Trx). Homólogos de estos tres genes han sido identificados en levaduras y mamíferos. Los primeros genes en plantas identificados que codifican proteínas con el dominio SET fueron homólogos E(z). Este tipo de proteínas juegan un mayor rol en el mantenimiento de la integridad del cromosoma. Mientras que las proteínas codificadas por Su(var)3-9, presumiblemente tienen una función en el empaquetamiento de la heterocromatina y las proteínas Trx pueden estar involucradas en las interacciones proteína-proteína (Baumbusch, et al., 2007).

## 2.8. CLASE II-ASH1 clase (H3K36 metilación).

Todos los miembros de esta clase (cinco en total) tienen el dominio SET que es invariablemente precedido por un dominio AWS (associated with SET) y seguido por el dominio post-SET rico en cisteína. Como fue reportado por Springer et al, 2003, esta clase de proteínas pueden ser agrupadas en varios grupos ortólogos, basado en la posición del dominio SET (Ng, et al., 2007).

ASHH2 es el homólogo en *Arabidopsis* del gen en la levadura SET2 el cual ha sido mostrado que exhibe actividad de metiltransferasa en la di- y tri-metilación en la lisina 36 de la histona 3 (H3K36) en *Arabidopsis* y se asocia con el dominio fosforilado C-terminal (CTD) de la ARN polimerasa II (RNAPII), lo que implica que esta enzima (ASHH2) podría jugar un papel importante en el proceso de elongación de la transcripción. ASHH1, contienen un dominio SET N-terminal localizado mientras ASHH3, ASHH4, poseen una más céntrica. Comparado con otras proteínas de esta clase, ASHH2, es significativamente grande y puede ser colocado en un distinto grupo ortológico. Además de los dominios de la familia ASHH2 contiene un dominio CW

(cisteína y triptófano conservado) que puede ser encontrado en al menos otras cinco familias de plantas superiores.

En cuanto al grupo ortológico que contiene ASHR3, un elemento adicional PHD (plant homeodomain) es localizado río arriba del dominio AWS. Por lo tanto, es posible que durante la evolución en las proteínas que contienen este dominio SET, un dominio adicional fuera adquirido para activar las funciones adicionales.

Además, los estudios genéticos revelaron que mutantes de ASHH2 muestran un fenotipo de floración temprana que se acompaña de una disminución de la metilación H3K36 en el FLC (Flowering Locus C). Además del principio de floración, los mutantes muestran fenotipos pleiotrópicos adicionales como el tamaño de la planta y fertilidad reducida, lo que sugiere que la contribución de ASHH2 no se limita al control del tiempo de floración. Por otro lado mutantes de ASHH1 muestran un fenotipo de floración tardía, debido a la expresión regulada por FLC. El análisis de microarrays indica además que ASHH1 participa principalmente en el mantenimiento del estado de los genes reprimidos, mientras que ASHH2 está principalmente involucrado en el mantenimiento del estado de los genes activados. ASHR3, se expresa específicamente en los órganos florales y se involucra en las metilaciones de H3K4 y H3K36. Por otra parte, fue demostrado que ASHR3 se une al factor de transcripción con medidas de acompañamiento (microsporas abortado), por lo que se propone a la cromatina como objetivo y la regulación de la expresión génica por asociación con medidas de acompañamiento en el desarrollo de las flores (Ng, et al., 2007).

## 2.9. Clase III-trithorax clase (H3K4 metilación).

En contraste a las proteínas PcG, las proteínas trxG (grupo trithorax) se identifican genéticamente como un regulador positivo de los genes homeóticos. En *Drosophila* las proteínas TRX (trithorax) Trx-G, que poseen actividad HMTasa H3K4 específica, puede regular la expresión de genes homeóticos través de un complejo de proteínas TAC1 (trithorax acetiltransferasa complex-1). Del mismo modo, el homólogo de la levadura SET1 cuya función es la activación de la transcripción a través del complejo de

proteínas COMPASS (proteínas complejas asociadas a SET1) (Ng et al., 2007). Las proteínas de *Arabidopsis* implicados en esta clase (siete en total) también se han denominado ATX (*Arabidopsis trithorax*) y ATXR (*Arabidopsis trithorax* relacionados).

La metilación de la lisina 4 y de la lisina 27 de la histona H3 (H3K4me3 y H3K27me3) son interpretados como marcas “activadoras” o “represivas”, respectivamente. “Los genes de la familia Trithorax codifican factores que pueden modular la estructura de la cromatina a través de su habilidad de metilar la parte N-terminal de la lisina 4 de la histona 3 (H3K4). Genes homólogos a Trithorax han sido identificados en plantas y animales, lo cual sugiere que mecanismos comunes de regulación epigenética derivaron de un ancestro común. De acuerdo a los modelos recientes, genes duplicados (parálogos) pueden haber permanecido con funciones redundantes o pueden haber adquirido funciones distintas” (Álvarez-Venegas, et al., 2008, p. 568). Si bien es lógico esperar que parálogos estructuralmente divergentes evolucionaron nuevas funciones, es imposible predecir aún las funciones de genes duplicados con secuencias codificantes grandemente conservadas.

Las proteínas de la clase III llevan varios dominios muy conservados [PWWP (dominio llamado en honor al motivo conservado Pro-Trp-Trp-Pro), FYRN (F/Y-rich N-terminal), FYRC (F/Y-rich C-terminal)] y el dominio PHD (a menudo se encuentran en proteínas asociadas a la cromatina) (Álvarez-Venegas y Avramova, 2001). Si bien todos estos dominios conservados se pueden encontrar en las proteínas dentro del subgrupo III-1 (ATX1 y ATX2), las proteínas del subgrupo III-2 (ATX3, ATX4 y ATX5) carecen de los dominios FYRN, y FYRC. En contraste los otros subgrupos de proteínas dentro de la misma clase, ATXR3, ATXR7 (subgrupos III-3 y 4) carecen de todos los dominios altamente conservados, y poseen sólo una C-terminal ubicado en el dominio SET y un dominio post SET.

Aunque la falta de caracterización bioquímica para otros miembros de la clase Trx de plantas (excepto ATX1 y ATX2), la presencia de varios dominios altamente conservados dentro de esta clase puede sugerir diversas funciones de estas proteínas que contiene el dominio SET.

Recientemente, una función específica, la trimetilación de ATX1 en H3K4 en el control de la activación de FLC ha sido informada. Aunque estructuralmente similar a ATX1, ATX2 codifica la proteína con funciones divergentes: ATX2 dimetila la lisina 4 de la histona 3 e influye en la expresión de los genes (Ng, et al., 2007).

## 2.10. TRANSCRIPCIÓN EN EUCARIOTAS. REGULACIÓN Y EXPRESIÓN.

En eucariotas superiores, la transcripción de genes es controlada por una variedad de mecanismos, tales como modificaciones de la cromatina o la degradación a través de miRNAs complementarias. Los promotores de genes y su composición de elementos reguladores -cis, sin embargo, son los puntos de control inicial para las actividades de la transcripción y definir la expresión potencial espacio-temporal de un gen. Entre otros aspectos, el conocimiento de las unidades funcionales elementales - sitios de unión al factor de transcripción - es un requisito previo para la comprensión de la regulación de los genes individuales y su incorporación en las redes de regulación. La regulación espacio-temporal de la expresión génica depende en gran medida de la presencia y la ausencia de sitios reguladores -cis en el promotor (Wang, et al., 2009).

La duplicación del gen es esencial para la evolución, proporcionando nuevas oportunidades para los organismos de diversificar en la función. Una vía fundamental de la diversificación funcional de las copias de los genes inicialmente redundantes o parálogos, es a través de alteraciones en sus patrones de expresión. Aunque los mecanismos subyacentes a la divergencia de expresión no comprenden la totalidad de los sitios de unión al factor de transcripción y la ocupación de nucleosomas se sabe que juega un papel importante en el proceso. Los intentos previos para detectar las variaciones genómicas mediante divergencia de expresión en ortólogos han tenido un éxito limitado por dos razones principales. En primer lugar, es inherentemente difícil de comparar las expresiones entre ortólogos debido a la variable de efectos de transacción y los estudios anteriores en segundo lugar, han cuantificado la divergencia de expresión en términos de una similitud general de los perfiles de expresión a través de varias muestras, lo que oscurece los cambios de expresión de la condición específica. Por otra parte, la correlación de expresiones intrínsecamente entre

homólogos presenta retos estadísticos, no se aborda adecuadamente en muchos estudios previos. El uso de rigurosas pruebas estadísticas, caracterizan la relación entre la divergencia de elemento -cis y la divergencia de expresión de la condición específica de los genes parálogos en *Saccharomyces cerevisiae*, esto presenta un novedoso medio de investigación de los potencialmente cambios en los reguladores -cis mediante divergencia de expresión en las familias de genes parálogos en condiciones específicas (Singh y Sridhar, 2009).

Recientemente, los avances en los proyectos genoma de eucariotas superiores, por ejemplo, vertebrados y el género *Drosophila*, ha impulsado el descubrimiento de motivos y la comprensión de las redes de regulación en estos organismos. En las plantas superiores, sin embargo, hasta el momento la falta de secuencias de genomas de plantas estrechamente relacionadas evolutivamente ha restringido el análisis a gran escala, principalmente a las plantas dicotiledóneas, como el sistema modelo *Arabidopsis thaliana* (Wang, et al., 2009).

## 2.11. ELEMENTOS REGULADORES CIS.

Un elemento cis-regulador o el cis-elemento es una secuencia de ADN o ARN que regula la transcripción o expresión de genes localizados en ese mismo filamento. La transcripción es por lo general río arriba de la región codificante. Este término se construye de Latino palabra *cis*, que significa “en el mismo lado que”.

Un cis-elemento se puede establecer en la región 5' del promotor al gene que controla, en un intrón, o en la región 3' sin traducir. Los bloques de construcción básicos de tales promotores son regiones de ADN-cis\_reguladoras, que en eucariotas a menudo forman parte de agrupaciones de elementos reguladores cis (CREs) (llamados motivos compuestos, o módulos) (Fauteux y Strömvik, 2009).

En cambio, los elementos trans-reguladores son las especies que pueden modificar la expresión de los genes distantes del gene que fue transcrito originalmente para crearlos. Para demostrar el concepto (esto no es un ejemplo específico), un factor de la

transcripción que regula un gen en el cromosoma 6 se ha transcrito en sí mismo de un gene encendido en el cromosoma 11. Este término se construye de raíz latina -trans, el cual significa “a través”.

Un gen consiste de una región de transcripción y otra de regulación. La región de transcripción es la parte del ADN que se transcribirá a una molécula de ARN complementaria. La región de regulación se puede dividir en elementos cis-reguladores (cis en el mismo lado) y elementos trans-reguladores (trans del otro lado). Los elementos cis-reguladores son sitios de unión de los factores de transcripción, que son las proteínas que al unirse con elementos cis-reguladores pueden afectar (aumentar o reprimir) la transcripción. Los elementos trans-reguladores son las secuencias de ADN que codifican los factores de transcripción.

Los elementos cis-reguladores pueden ser divididos en los siguientes cuatro tipos:

1. Promotor: Está compuesto por una secuencia específica de ADN localizada justo donde se encuentra el punto de inicio de la transcripción del ARN y contiene la información necesaria para regular la expresión (activar o desactivar) del gen que codifica. Los promotores facilitan la transcripción de un gen en particular y por lo general río arriba de la región codificante.
2. Potenciador (enhancer): Es una región corta del ADN que puede unirse con proteínas (factores de transcripción) para aumentar los niveles de transcripción de genes en un grupo de estos. Un potenciador no tiene por qué estar localizado cerca de los genes sobre los que actúa, ni siquiera en el mismo cromosoma; por tanto, puede tener efectos muy distantes de los niveles de transcripción de los genes.
3. Represor: Es una proteína de unión de ADN que regula la expresión de uno o más genes mediante la disminución de la tasa de transcripción. Los represores se unen a un segmento de ADN evitando que la ADN polimerasa cree el ARN mensajero.

4. Elemento de respuesta: El sitio capaz de ser reconocido por un factor de transcripción dado de manera más o menos específica.

Tanto los promotores como los elementos reguladores son comúnmente determinados mediante análisis experimental. Este proceso es extremadamente laborioso y tardado; por esta razón la predicción mediante métodos computacionales de promotores y elementos reguladores es especialmente interesante, pues permitiría remplazar costosos procesos de análisis experimental. En resumen, los elementos -cis están presentes en el mismo filamento del gene que él regula, mientras que los elementos - trans pueden regular los genes distantes al gene de el cual fueron transcritos (Dr. Eduardo A. Rodríguez Tello, 2010).

### 3. MATERIALES Y MÉTODOS

#### OBJETIVO ESPECIFICO 1.

The Arabidopsis Gene Regulatory Information Server (AGRIS) <http://arabidopsis.med.ohio-state.edu/> es un nuevo recurso de información para el análisis de las secuencias del promotor de *Arabidopsis*, los factores de transcripción y de sus genes blanco. AGRIS actualmente contiene tres bases de datos de las cuales AtcisDB consta de aproximadamente 33,000 regiones río arriba de los genes anotados de *Arabidopsis* (comunicado de TAIR9) con una descripción de validación experimental y elementos reguladores –cis predichos. AGRIS integra además una variedad de herramientas para determinar los factores de transcripción y sus sitios putativos vinculantes a todos los genes para reconstruir redes de regulación transcripcional en *Arabidopsis*.

AtcisDB fue la herramienta de consulta para recolección y ordenamiento de los elementos reguladores-cis en los promotores de genes de *Arabidopsis thaliana* de las clases II y III de la actual clasificación de Ng y colaboradores (2007), como se cito en el apartado 2.7 (*Arabidopsis* SET (AtSET)) de los antecedentes (ver tabla 3.1).

Para ampliar e integrar la información al respecto, la página The Arabidopsis Information Resource (TAIR) <http://www.arabidopsis.org/> que mantiene una base de datos de Genética y Biología Molecular de plantas superiores del modelo *Arabidopsis thaliana*, también ha sido una fuente de consulta.

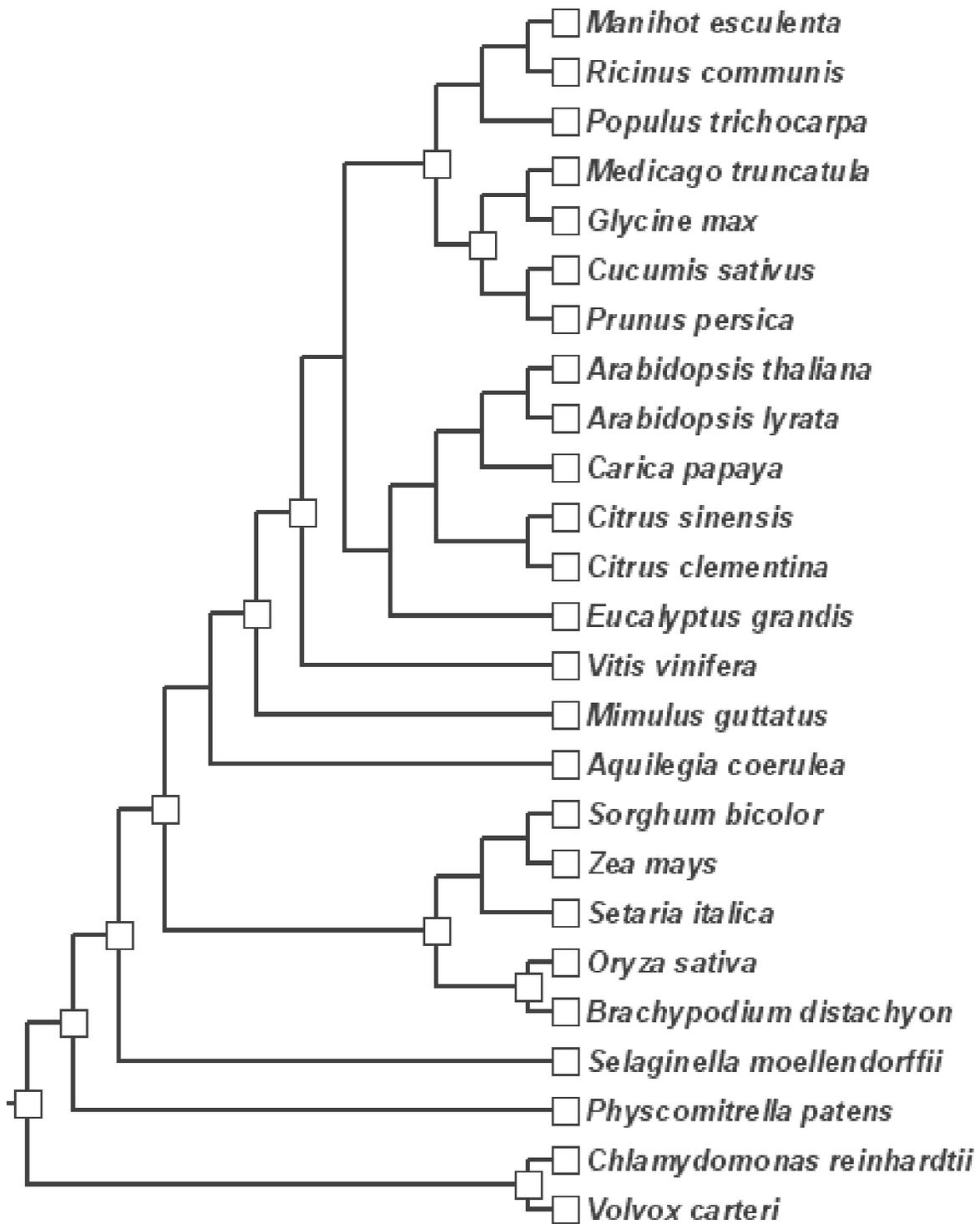
## OBJETIVO ESPECIFICO 2.

Para la comparación y cuantificación de los elementos -cis presentes en genes que codifican proteínas con el dominio SET de la familia Trithorax, en tres distintos genomas de plantas; del árbol filogenético de la página <http://www.phytozome.net/search.php>, se seleccionaron tres especies en relación a su linaje y a su ubicación dentro del género viridiplantae, las especies seleccionadas son: *Glycine max* (soya), *Zea mays* (maíz) y *Physcomitrella patens* (musgo) (figura 3.1). Para la obtención o descubrimiento de los elementos-cis, se realizó por separado en cada una de las tres especies de plantas un análisis con un modelo de fondo (consulta) sobre la base de el conjunto de los promotores de una especie representativa, en nuestro caso, *Arabidopsis thaliana*. Dado que nuestro objetivo es sobre los genes de la familia Trithorax que codifican proteínas con el dominio SET, tomamos los genes de las clases II y III de la actual clasificación de Ng y colaboradores (2007) (ver tabla 3.1).

Las secuencias de los promotores (3000 pares de bases (pb) río arriba de la región de inicio de la transcripción) de las tres especies, que al final de su recolección y depuración bajo el criterio de considerar solo aquellas con máxima similitud de acuerdo a la herramienta BLAST (The Basic Local Alignment Search Tool) (ver tabla 4.4), fueron: *Glycine max* (22 promotores), *Zea mays* (10 promotores) y *Physcomitrella patens* (12 promotores) los cuales se obtuvieron de secuencias de fuentes públicas de bases de datos (tabla 4.5).

BLAST encuentra las regiones de similitud entre las secuencias locales. El programa compara las secuencias de nucleótidos o proteínas en bases de datos de secuencias y calcula la significancia estadística de partidos. BLAST puede utilizarse para inferir las relaciones funcionales y evolutivas entre las secuencias, así como ayudar a identificar a los miembros de familias de genes.

Figura 3.1



Árbol filogenético de la actual página: [http:// www.phytozome.net/search.php](http://www.phytozome.net/search.php)

El método fue realizado de la siguiente manera: Utilizando los locus de las clases II y III de la clasificación de Ng, et al, (2007) (e.g. At2g31650), encontramos las secuencias de nucleótidos en la página: <http://www.ncbi.nlm.nih.gov/> de la cual tomamos la región CDS que contiene la identificación de la secuencia para la codificación de la proteína. Todas las secuencias obtenidas fueron convertidas al formato FASTA para ser utilizadas en la siguiente etapa de nuestra búsqueda. FASTA se refiere a un formato para una un ácido nucleico o una secuencia de proteína.

Seguido a la obtención de las secuencias de nucleótidos, éstas fueron utilizadas como consulta con la herramienta BLAST (BLASTP) de la página: <http://www.phytozome.net/search.php>, utilizando como blancos u objetivos y por separado cada una las tres especies en cuestión, de esta manera se obtuvieron centenas de genes (2279 en total) de los cuales 1743 contienen el dominio SET, separando de ellos, los de máxima similitud con respecto a la herramienta BLAST utilizada, 84 en total (ver tablas 4.6 y 4.7), extrayendo de éstos últimos las secuencias 3000 pb río arriba del sitio de inicio de la transcripción con la finalidad de incluir la región promotora de tales genes; de esta manera se obtuvo la particular base de datos sobre la cual ampliaríamos nuestro análisis, una vez depurada de acuerdo a un criterio de selección (ver discusión de resultados), eliminando aquellas secuencias de promotores comunes entre las especies de cada clase en la familia trithorax (ver tabla 4.5).

PlantCARE (Plant Cis-Acting Regulatory Elements) <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/> es una base de datos de los elementos -cis reguladores en plantas y un portal de herramientas para el análisis *in silico* de las secuencias del promotor. PlantCARE fue utilizado para identificar los elementos reguladores -cis en las secuencias de los promotores de los genes dentro de cada especie (ver tablas 4.8 y 4.9). Sin embargo, esta base de datos, actualmente, solo acepta secuencias con una longitud máxima de 1500 pb, por lo cual nuestras secuencias originales (de 3000 pb) fueron reducidas a las primeras 1500 pb a partir del sitio de iniciación de la transcripción.

PLACE (Plant cis-acting regulatory DNA elements) [http://www.dna.affrc.go.jp / htdocs / PLACE/](http://www.dna.affrc.go.jp/htdocs/PLACE/) es una base de datos de elementos -cis encontrados en las plantas, todos de reportes publicados con anterioridad. Cubre solo las plantas vasculares. Además de los motivos originalmente reportados, sus variaciones en otros genes o en otras especies de plantas reportados más tarde son también compiladas. Contiene también una breve descripción de cada motivo y literatura relevante, con números de identificación de PubMed. Este informe resume la situación actual de esta base de datos y herramientas disponibles. La disposición de PLACE permitió su consulta para la comparación de los elementos -cis encontrados en el presente análisis y su base de dato así como información complementaria.

Diagrama de flujo que resume la metodología para la determinación y cuantificación de los elementos –cis en los tres distintos genomas del presente estudio.

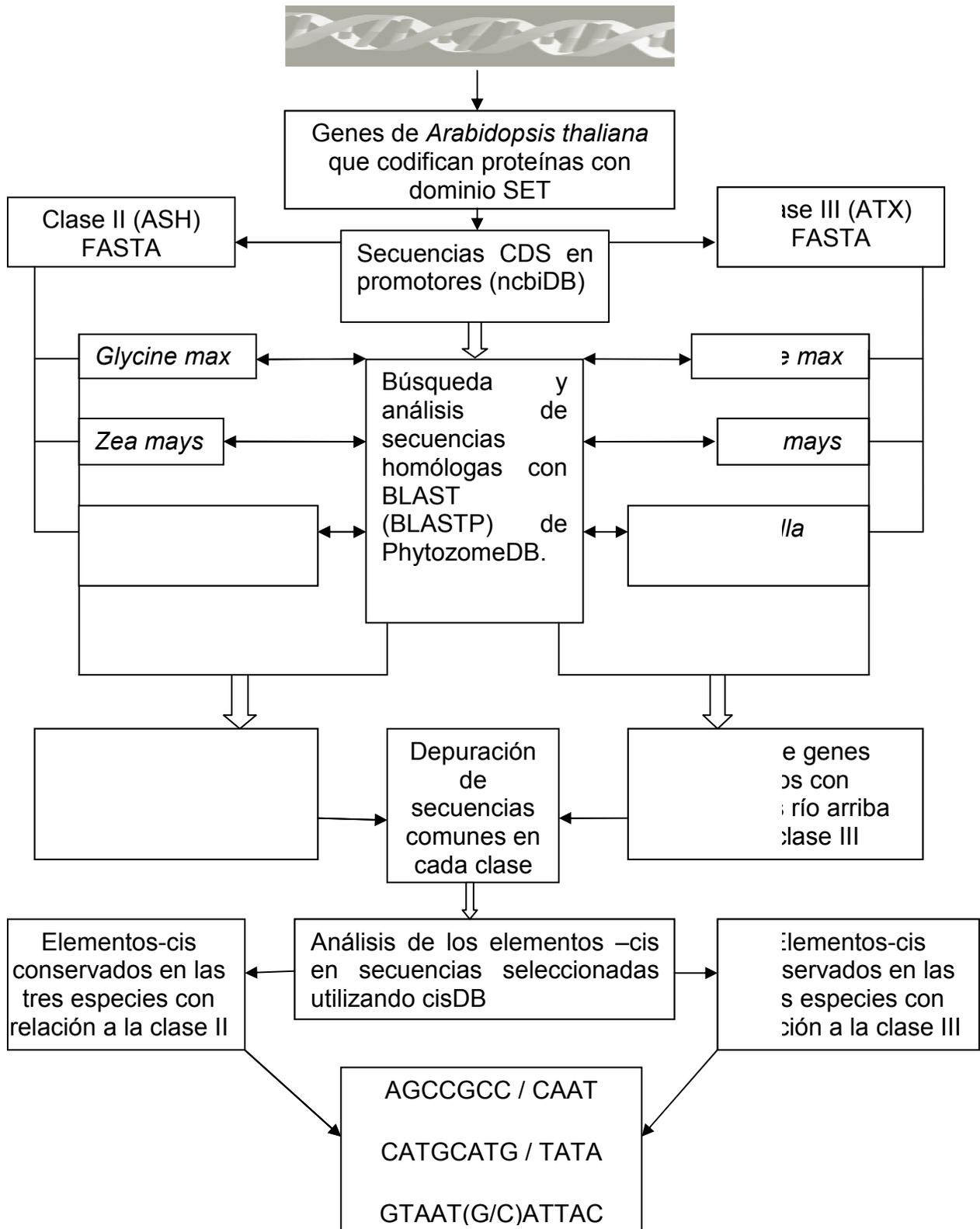


Tabla 3.1. Proteínas que contienen el dominio SET en Arabidopsis, maíz y arroz.

Clase	Grupo	<i>Arabidopsis thaliana</i>	<i>Zea mays</i>	<i>Oryza sativa</i> spp. <i>Japonica</i>
I	1	MEA (SDG5; At1g02580; 689aar)	–	–
	2	CLF (SDG1; At2g23380; 902aar)	MEZ1 (AF443596; 931aar)	SDG711 (Os06g16390; 896aar)
	3	SWN (SDG10; At4g02020; 856aar)	MEZ2 (AF443597; 894aar),	SDG718 (Os03g19480; 895aar)
			MEZ3(AF443598;895aar)	
II	1	ASHH3 (SDG7; At2g44150; 363aar)	SDG110 (AF545814; 342aar)	SDG724 (Os09g13740; 340aar)
		ASHH4 (SDG24; At3g59960; 352aar)	–	–
	2	ASHR3 (SDG4; At4g30860; 497aar)	–	SDG736 (Os02g39800; 492aar)
	3	ASHH1 (SDG26; At1g 76710; 492aar)	SDG102 (AY122273; 513aar)	SDG708 (Os04g34980; 518aar)
	4	ASHH2 (SDG8; At1g77300; 1759aar)	–	SDG725 (Os02g34850; 2133aar)
III	1	ATX1 (SDG27; At2g31650; 1062aar)	–	SDG723 (Os09g04890; 1022aar)
		ATX2 (SDG30; At1g05830; 1193aar)	SDG106 (DR826986; 385aar)*,	–
			SDG128 (DV534215; 296aar)*	
	2	ATX3 (SDG14; At3g61740; 902aar)	–	–
		ATX4 (SDG16; At4g27910; 912aar)	SDG115 (DN233805; 147aar)*	SDG721 (Os01g11950; 991aar)
		ATX5 (SDG29; At5g53430; 1043aar)	–	SDG705 (Os01g46700; 1057aar)
	3	ATXR3 (SDG2; At4g15180; 2351aar)	SDG108 (DV025562; 614aar)*	SDG701 (Os08g08210; 2257aar)
	4	ATXR7 (SDG25; At5g42400;1421aar)	SDG127 (DN204179; 141aar)*	SDG717 (Os12g41900; 1212aar)
IV		ATXR5 (SDG15; At5g09790; 352aar)	SDG139 (DV172958; 231aar)*	SDG720 (Os01g73460; 385aar)
		ATXR6 (SDG34; At5g24330; 349aar)	SDG129 (BM736459; 109aar)*	SDG730 (Os02g03030; 361aar)

Tabla 3.1. (continuación) Proteínas que contienen el dominio SET en Arabidopsis, maíz y arroz.

V	1	SUVH1 (SDG32; At5g04940; 670aar)	SDG101 (DV540845; 795aar)	SDG704 (Os11g38900; 813aar), SDG713 (Os03g20430; 627aar)
		SUVH3 (SDG19; At1g73100; 669aar)	SDG105 (AY093419; 678aar), SDG111 (AY187718; 486aar)	SDG728 (Os05g41170; 672aar)
		SUVH7 (SDG17; At1g17770; 693aar)	SDG113 (AF545813; 766aar)	SDG709 (Os01g59620; 736aar),
		SUVH8 (SDG21; At2g24740; 755aar)	–	SDG733 (Os11g03700; 663aar), SDG734 (Os12g03460; 663aar)
	2	SUVH4 (SDG33; At5g13960; 624aar)	SDG118 (AY122271; 696aar)	SDG714 (Os01g70220; 663aar)
		SUVH6 (SDG23; At2g22740; 790aar)	–	SDG727 (Os09g19830; 921aar)
	3	SUVH2 (SDG3; At2g33290; 651aar)	SDG136 (DN217108; 662aar)	SDG726 (Os07g25450; 684aar)
		SUVH9 (SDG22; At4g13460; 650aar)	SDG135 (CF049431; 449aar),*	SDG715 (Os08g45130; 594aar)
			SDG137 (DN232551; 685aar)	
	4	SUVR3 (SDG20; At3g03750; 338aar)	SDG116 (CO520549; 128aar)*	SDG729 (Os01g56540; 338aar)
	5	SUVH5 (SDG9; At2g35160; 794aar)	SDG103 (DN214510, 654aar)*, SDG104 (AY122272; 886aar), SDG119 (DV031465, 508aar)*	SDG703 (Os04g34990; 842aar), SDG710 (Os08g30910; 1173aar)
	6	SUVR1 (SDG13; At1g04050; 630aar)	–	–
		SUVR2 (SDG18; At5g43990; 717aar)	–	SDG712 (Os02g40770; 741aar)
		SUVR4 (SDG31; At3g04380; 492aar)	SDG107 (DV544186; 148aar)*	–
7	SUVR5 (SDG6; At2g23740; 203aar)	SDG117 (AY187719; 1198aar), SDG131 (DN228281; 131aar)*	SDG706 (Os02g47900; 1136aar)	

Tabla 3.1. (continuación) Proteínas que contienen el dominio SET en Arabidopsis, maíz y arroz.

VI	ASHR1 (SDG37; At2g17900; 480aar)	SDG130 (AAL75997; 410aar)	SDG716 (Os03g49730; 403aar)
	ASHR2 (SDG39; At2g19640; 398aar)	–	SDG740 (Os08g10470; 392aar)
	ATXR1 (SDG35; At1g26760; 969aar)	SDG122 (DY238779; 534aar)	SDG739 (Os03g07260; 536aar)
	ATXR2 (SDG36; At3g21820; 473aar)	SDG140 (DV505249; 217aar)*	SDG722 (Os04g53700; 517aar)
	ATXR4 (SDG38; At5g06620; 258aar)	SDG123 (AY172976; 303aar)	SDG741 (Os10g27060; 298aar)
VII	SDG40 (At5g17240; 491aar)		SDG731 (Os07g28840; 479aar)
	Putative RuBisCo SSMT (At3g07670; 504aar)		
	At2g18850 (543aar)		
	At5g14260 (514aar)		
	RuBisCo LSMT (At1g14030; 482aar)		
	At1g01920 (572aar)		
	At1g24610 (476aar)		
	At3g55080 (463aar)		
	At3g56570 (531aar)		
	At4g20130 (483aar)		

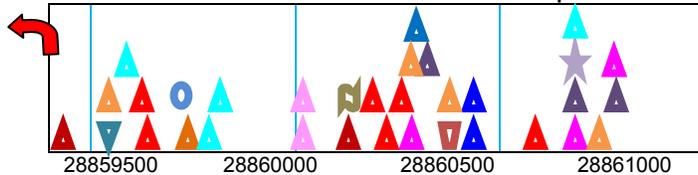
## 4. RESULTADOS

Ilustración 4.1. Código de colores para la identificación de los elementos -cis en las figuras.

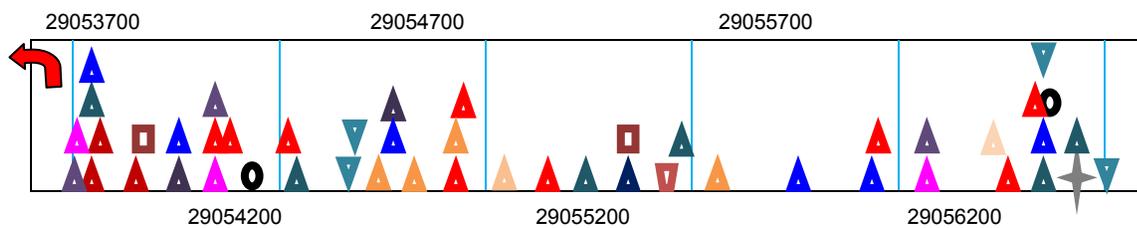
ABRE-like binding site motif(Abscisic acid Responsiveness Element)	
ARF binding site motif (Auxin Response Factor)	
ARF1 binding site motif (Auxin Response Factor-1)	
ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	
ATHB1 binding site motif Arabidopsis (A.T.) homeobox gene (ATHB)	
ATHB2 binding site motif (Arabidopsis thaliana homeobox)	
ATHB5 binding site motif	
AtMYC2 BS in RD22	
Bellringer/replumless/pennywise BS1 IN AG	
BoxII promoter motif	
CARG promoter motif	
CCA1 binding site motif	
DPBF1&2 binding site motif (Dc3 promoter-binding factor-1 and 2)	
DRE-like promoter motif	
E2F binding site motif	
E2F-varient binding site motif	
GATA promoter motif [LRE] (Light Responsive Element)	
G-box promoter motif [LRE]	
GCC-box promoter motif	
Hexamer promoter motif	
HSEs binding site motif (Heat Stress Element)	
Ibox promoter motif	
L1-box promoter motif	
LFY consensus binding site motif (LEAFY)	
MYB binding site promoter	
MYB1 binding site motif	
MYB3 binding site motif	
MYB4 binding site motif	
RAV1-A binding site motif	
RAV1-B binding site motif	
RY-repeat promoter motif	
SBP-box promoter motif	
SORLIP1 (Sequences Over-Represented in Light-Induced Promoters)	
SORLIP2	
SORLIP5	
SORLREP2 (Over-represented in light-induced cotyledon and root-specific genes)	
SORLREP3	
T-box promoter motif	
TELO-box promoter motif	
TGA1 binding site motif	
W-box promoter motif	

Las figuras muestran un esquema de los elementos -cis en el alineamiento de varios de los promotores de genes de *Arabidopsis thaliana* (las posiciones están en escala, pero aproximadamente). El sitio de inicio de transcripción TSS (transcription start site) y su sentido es indicado por la flecha roja.

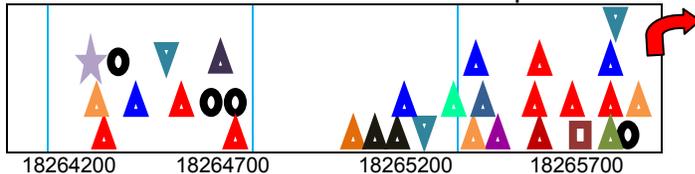
ASHH1. Cromosoma 1: 28859430 pb...28860898 pb. (R: 28859449-28860818: 1369 pb)



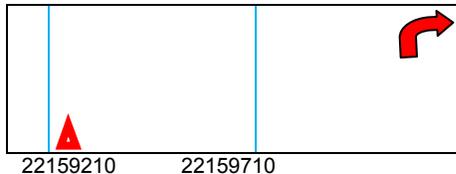
ASHH2. Cromosoma 1: 29053703pb...29056315 pb. (R: 29053724-29056218: 2494 pb)



ASHH3. Cromosoma 2: 18264239 pb...18265685 pb. (R: 18264306-18245660: 1354 pb)



ASHH4. Cromosoma 3: 22159209 pb...22159309 pb. (R: 22159267- 22159271: 4 pb)



ASHR3. Cromosoma 4: 15022282 pb...15024477 pb. (R: 15022516-15024467: 1951 pb)

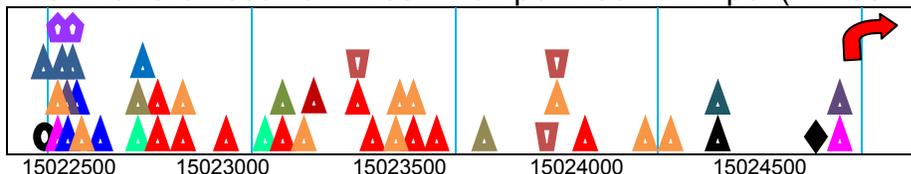


Tabla 4.1. Elementos-cis presentes en promotores de genes de *Arabidopsis thaliana* que codifican para proteínas con el dominio SET de la familia Trithorax [clases II (ASH)].

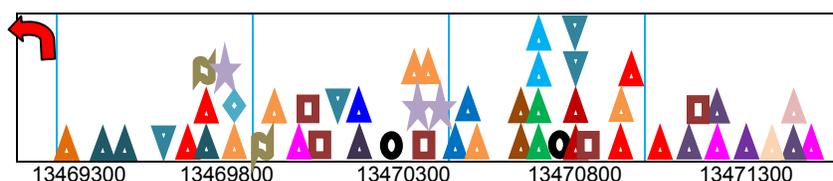
	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
ABRE-like binding site motif			x		x (2)	3
ARF binding site motif	x(3)	x (3)			x (2)	8
ARF1 binding site motif	x (3)	x (3)			x (2)	8
ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	x	x			x (3)	5
ATHB1 binding site moti						0
ATHB2 binding site motif			x		x (2)	3
ATHB5 binding site motif						0
AtMYC2 BS in RD22		x (2)	x			3
Bellringer/replumless/pennywise BS1 IN AG	x (2)	x (3)	x		x	7
BoxII promoter motif		x (2)	x			3
CArG promoter motif	x (2)					2
CCA1 binding site motif	x				x	2
DPBF1&2 binding site motif	x (2)	x (6)	x (4)		x (3)	15
DRE-like promoter motif	x					1
E2F binding site motif						0
E2F-variant binding site motif					x	1
GATA promoter motif [LRE]	x (4)	x (4)	x (3)		x (9)	20
GCC-box promoter motif						0
G-box promoter motif					x (2)	2
Hexamer promoter motif			x (2)			2
HSEs binding site motif						0
Ibox promoter motif	x		x			2
L1-box promoter motif		x				1
LFY consensus binding site motif	x	x (4)	x (3)			8
MYB binding site promoter		x				1
MYB1 binding site motif		x				1
MYB3 binding site motif					x	1
MYB4 binding site motif		x (6)			x (2)	8
RAV1-A binding site motif	x (6)	x (9)	x (7)	x	x (10)	33
RAV1-B binding site motif	x		x		x	3
RY-repeat promoter motif						0
SBP-box promoter motif						0
SORLIP1	x (3)					3
SORLIP2	x		x		x	3
SORLIP5						0
SORLREP2			x			1
SORLREP3						0
T-box promoter motif	x				x (2)	3

Tabla 4.1. (continuación)

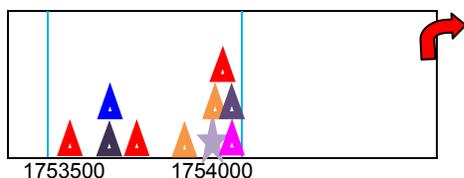
	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
TELO-box promoter motif		x				1
TGA1 binding site motif						0
W-box promoter motif		x (2)	x (4)		x	7
Total para cada promotor:	33	49	32	1	46	161

Notas tabla 4.1. Tabla que hace referencia a las figuras de los promotores de genes del grupo II ASH. El símbolo X indica la presencia del elemento –cis en el promotor y el número entre paréntesis su frecuencia. Las posiciones relativas de cada uno se indican en la figura respectiva.

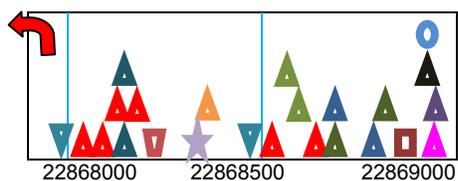
ATX1. Cromosoma 2: 13469257 pb...13471274 pb. (R: 13469323-13471233: 1910 pb)



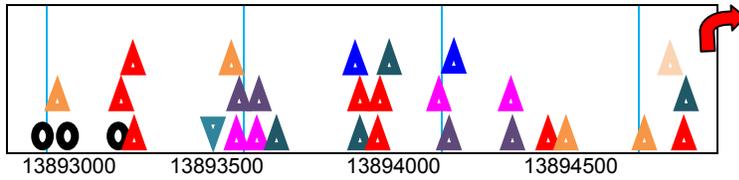
ATX2. Cromosoma 1: 1753764 pb...1754242 pb. (R: 1753774-1754163: 389 pb)



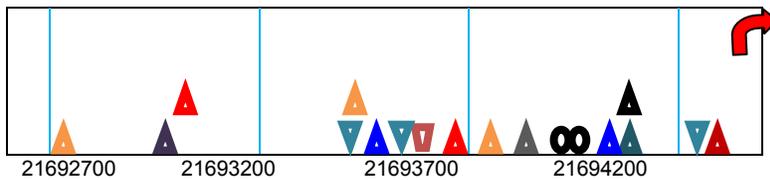
ATX3. Cromosoma 3: 22867946 pb...22868870 pb. (R: 22868007-22868864: 857 pb)



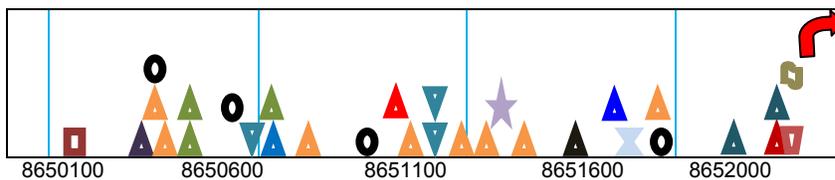
ATX4. Cromosoma 4:13892999 pb...13894698 pb. (R: 13893017-13894643: 1626 pb)



ATX5. Cromosoma 5: 21692693 pb...21694871 pb. (R: 21692761-21694346: 1585 pb)



ATXR3. Cromosoma 4: 8650081 pb...8651997 pb. (R: 8650170- 8651950: 1780 pb)



ATXR7. Cromosoma 4: 16977898 pb...16980899 pb. (R: 16977931-16980892: 2961 pb)

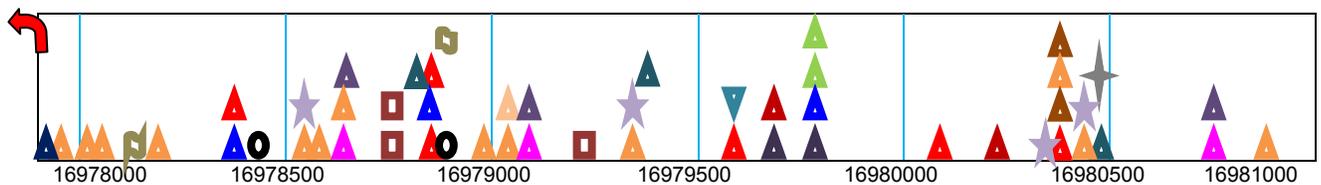


Tabla 4.2. Elementos-cis presentes en promotores de genes de *Arabidopsis thaliana* que codifican para proteínas con el dominio SET de la familia Trithorax [clases III (ATX)].

	ATX 1	ATX 2	ATX 3	ATX 4	ATX 5	ATXR 3	ATXR 7	TOTAL
ABRE-like binding site motif			x (2)					2
ARF binding site motif	x (3)	x	x	x (4)			x (3)	12
ARF1 binding site motif	x (3)	x	x	x (4)			x (3)	12
ATB2/AtbZIP53/AtbZIP44/GB F5 BS in ProDH			x		x	x		3
ATHB1 binding site motif	x (2)							2
ATHB2 binding site motif								0
ATHB5 binding site motif	x (2)							2
AtMYC2 BS in RD22	x	x			x (2)	x	x (3)	8
Bellringer/replumless/pennywise BS1 IN AG	x				x	x	x (2)	5
BoxII promoter motif	x (5)		x			x	x (3)	10
CARG promoter motif								0
CCA1 binding site motif	x (2)					x		3
DPBF1&2 binding site motif	x	x		x (2)	x (2)	x	x (3)	10
DRE-like promoter motif			x					1
E2F binding site motif	x							1
E2F-variant binding site motif								0
GATA promoter motif [LRE]	x (6)	x (2)		x (4)	x (3)	x (8)	x (13)	36
GCC-box promoter motif	x							1
Hexamer promoter motif			x			x		2
HSEs binding site motif	x (2)						x (2)	4
Ibox promoter motif	x (3)	x	x			x	x (4)	7
L1-box promoter motif							x	1
LFY consensus binding site motif	x (4)		x (2)	x	x (3)	x (3)	x	14
MYB binding site promoter							x	1
MYB1 binding site motif							x	1
MYB3 binding site motif					x			1
MYB4 binding site motif	x (3)		x (2)	x (4)	x	x (2)	x (3)	15
RAV1-A binding site motif	x (5)	x (3)	x (6)	x (8)	x	x	x (5)	29
RAV1-B binding site motif	x							1
RY-repeat promoter motif							x (2)	2
SBP-box promoter motif					x			1
SORLIP1								0
SORLIP2			x (2)			x (3)		5

Tabla 4.2. (continuación)

	ATX 1	ATX 2	ATX 3	ATX 4	ATX 5	ATXR 3	ATXR 7	TOTAL
SORLIP5	x							1
SORLREP2								0
SORLREP3						x		1
T-box promoter motif	x (2)					x	x (2)	5
TELO-box promoter motif	x			x				2
TGA1 binding site motif			x (2)					2
W-box promoter motif	x (2)			x (3)	x (2)	x (4)	x (2)	13
Total para cada promotor:	52	10	23	31	18	31	54	216

Notas tabla 4.2. Tabla que hace referencia a las figuras de los promotores de genes del grupo III ATX. El símbolo X indica la presencia del elemento –cis en el promotor y el número entre paréntesis indica su frecuencia. Las posiciones relativas de cada uno se indican en la figura respectiva.

En general, la longitud de los promotores de genes aquí reportados es del orden de 100 pb a 3000 pb y el rango (R) de ocupación dentro de éstos para los elementos reguladores –cis abarcan desde 4 pb para el promotor del gen ASHH4, un único elemento –cis encontrado (RAV1-A) hasta 2961 pb para el promotor del gen ATXR7, se indica además el número de cromosoma donde se ubica el gen. Este último promotor (ATXR7) fue considerado sobre los demás, por su mayor tamaño, como base para la elección de la longitud de las secuencias que se obtendrían para nuestro segundo objetivo.

Los elementos –cis presentes en los promotores de genes de *Arabidopsis thaliana* que codifican para proteínas con el dominio SET de la familia Trithorax, clases II (ASH) y III (TRX) de la actual clasificación de Ng, et al, (2007) fueron extraídos de AGRIS (AtcisDB), (ver materiales y métodos), para ser analizados y organizados como se muestra en las figuras correspondientes; el análisis básicamente se enfocó en la cuantificación, ubicación y comparación de los elementos –cis presentes a la fecha, en ambas clases. En el recuento total fueron: considerando su presencia, 41 elementos –cis en total y considerando su frecuencia, 377 elementos –cis totales (161 para la clase II y 216 para la clase III). Sin embargo, se observó que algunos están presentes en una clase y ausentes en la otra y además los que están presentes tienen un porcentaje bajo

en relación a su clase particular y en relación también con ambas clases (familia), como se observa en la tabla 4.3.

Tabla 4.3. Elementos-cis presentes en promotores de genes de *Arabidopsis thaliana* que codifican para proteínas con el dominio SET de la familia Trithorax [clases II (ASH) y III (ATX)].

	% por clase		% por familia	Secuencias
	ASH	TRX		
ABRE-like binding site motif	0.74 (1)	0.92 (2)	0.86	GACGTGGC/ TACGTGGA
ARF binding site motif	4.48 (6)	5.55 (12)	5.14	TGTCTC
ARF1 binding site motif	6.0 (8)	5.55 (12)	5.71	TGTCTC
ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	3.7 (5)	1.38 (3)	2.29	ACTCAT
ATHB1 binding site motif	0	0.92 (2)	0.57	TGATAA
ATHB2 binding site motif	2.2 (3)	0	0.86	AGATAG
ATHB5 binding site motif	0	0.92 (2)	0.57	CAATAATTG
AtMYC2 BS in RD22	2-2 (3)	3.70 (8)	3.14	CACATG
Bellringer/replumless/pennywise BS1 IN AG	5.2 (7)	2.3 (5)	3.43	AAATTAAA
BoxII promoter motif	2.2 (3)	4.62 (10)	3.71	GGTTAA
CARG promoter motif	1.5 (2)	0	0.57	CCTTTTTTGG / CCAAAAAAGG
CCA1 binding site motif	1.5 (2)	1.38 (3)	1.43	AAAAATCT
DPBF1&2 binding site motif	11.2 (15)	4.62 (10)	7.14	ACACATG/ ACACAGG
DRE-like promoter motif	0.74 (1)	0.46 (1)	0.57	AACCGACGA
E2F binding site motif	0	0.46 (1)	0.28	TTTCCCGC
E2F-varient binding site motif	0.74 (1)	0	0.28	TCTCCCGCC
GATA promoter motif [LRE]	8.2 (11)	16.66 (36)	13.43	AGATAA / TGATAA/ TGATAG
GCC-box promoter motif	0	0.46 (1)	0.28	GCCGCC
Hexamer promoter motif	1.5 (2)	0.92 (2)	1.14	CCGTCCG
LFY consensus binding site motif	6.0 (8)	6.48 (14)	6.29	CCAATG/ CCATTG/ CCAGTG
MYB binding site promoter	0.74 (1)	0.46 (1)	0.57	CACCTACC
MYB1 binding site motif	0.74 (1)	0.46 (1)	0.57	ATCCAACC
MYB4 binding site motif	6.0 (8)	6.9 (15)	6.57	AACTAAC/ AACAAAC/ ACCAACC

Tabla 4.3. (continuación)

	% por clase		% por familia	Secuencias
	ASH	TRX		
RAV1-A binding site motif	18.6 <b>(25)</b>	13.4 <b>(29)</b>	<b>15.43</b>	CAACA
RAV1-B binding site motif	1.5 (2)	0.46 (1)	0.86	CACCTG
RY-repeat promoter motif	0	0.92 (2)	0.57	CATGCATG
SBP-box promoter motif	0	0.46 (1)	0.28	TCCGTACAA
SORLIP1	2.2 (3)	0	0.86	AGCCAC
SORLIP2	1.5 (2)	2.3 (5)	2.00	GGGCC
SORLIP5	0	0.46 (1)	0.28	GAGTGAG
SORLREP2	0.74 (1)	0	0.28	ATAAAACGT
SORLREP3	0	0.46 (1)	0.28	TGTATATAT
T-box promoter motif	0.74 (1)	2.3 (5)	1.71	ACTTTG
TELO-box promoter motif	0.74 (1)	0.92 (2)	0.86	AAACCCTAA
TGA1 binding site motif	0	0.92 (2)	0.57	TGACGTGG
W-box promoter motif	5.2 (7)	<b>6.0 (13)</b>	5.71	TTGACT/ TTGACC

Notas para la tabla 4.3. El cero (0) indica la ausencia del elemento –cis y el número entre paréntesis el total, ambos en la clase correspondiente, como se indica en las tablas 1A y 1B. El porcentaje por familia fue calculado en base a la sumatoria de los elementos –cis de ambas clases.

**Gráfica 4.1 Elementos -cis presentes en promotores de genes de *Arabidopsis thaliana* de las clases II (ASH) y III (ATX) que codifican proteínas con el dominio SET.**

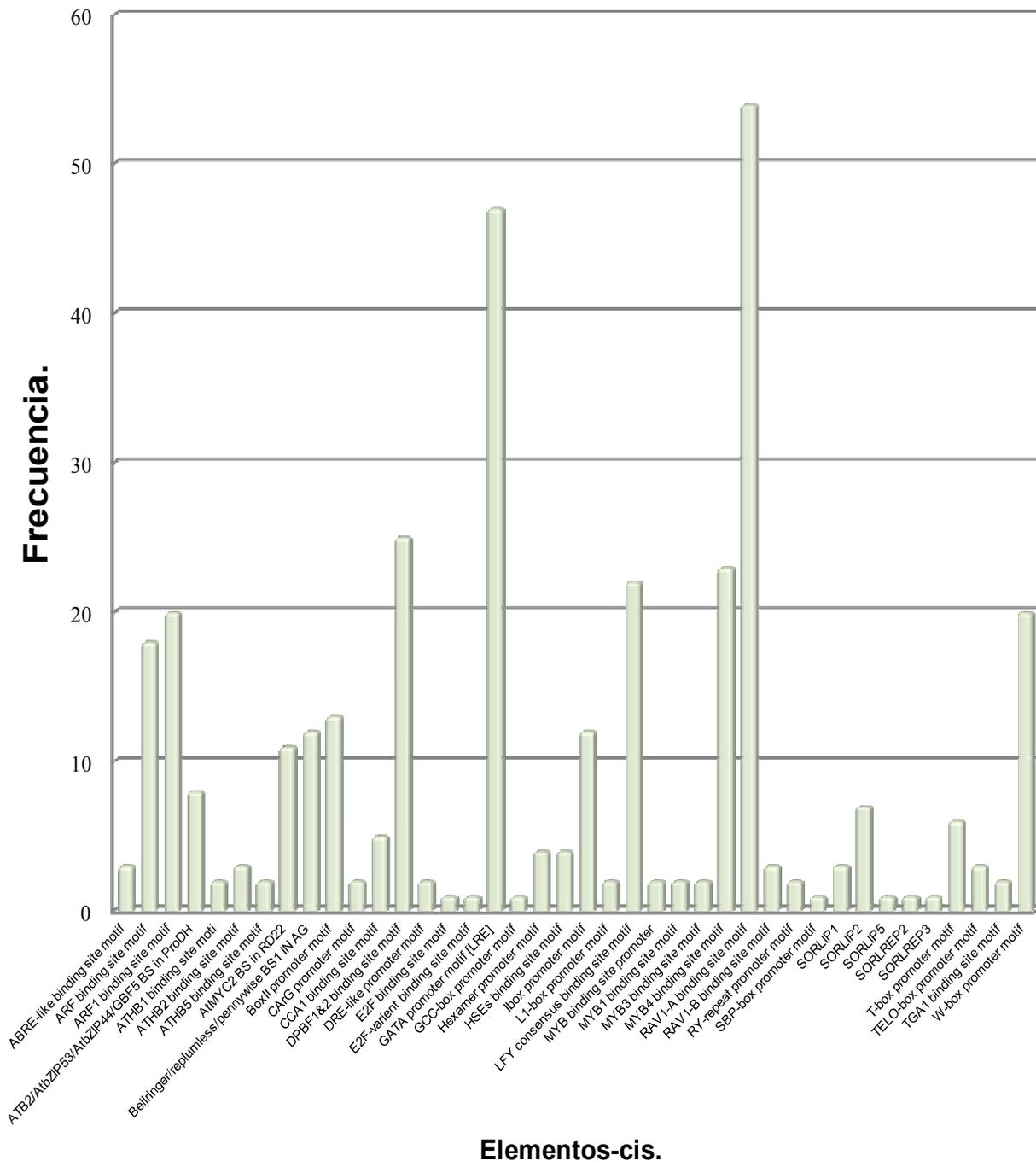


Tabla 4.4. Resumen de las secuencias obtenidas con máxima similitud utilizando la herramienta BLAST.

	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
Glyma08g29010.1	x	<u>x</u>											2
Glyma18g51890.1	x	<u>x</u>											2
Glyma07g06190.1			<u>x</u>	x	x								3
Glyma03g37370.1			<u>x</u>	x	x								3
Glyma19g39970.1			<u>x</u>	x	x								3
Glyma16g02800.1			<u>x</u>	x	x								3
Glyma04g41500.1			x	<u>x</u>	x								3
Glyma06g13330.1			x	<u>x</u>	x								3
Glyma19g17460.1 (.2)			x	x	<u>x</u>								3
Glyma12g11060.1						<u>x</u>							1
Glyma06g45740.1						<u>x</u>							1
Glyma13g38090.1						<u>x</u>							1
Glyma12g32290.1						<u>x</u>							1
Glyma09g05740.1							<u>x</u>						1
Glyma15g17030.1							<u>x</u>						1
Glyma10g36720.1								<u>x</u>					1
Glyma20g30870.1								<u>x</u>					1
Glyma04g42410.1									x				1
Glyma06g12390.1									x				1
Glyma16g33220.1 (.2)										<u>x</u>	x		2
Glyma09g28430.1 (.2)										<u>x</u>	x		2
Glyma06g29960.1												x	1
GRMZM2G013794_T02	<u>x</u>	x											2
GRMZM2G085266_T01			x	<u>x</u>	x								3

GRMZM2G149153_T01							x								1
GRMZM2G473138_T01								<u>x</u>							1
GRMZM2G147619_T03 (T05 y T04)									<u>x</u>						1
GRMZM2G352431_T02 (T01)										<u>x</u>					1
GRMZM2G033694_T03 (T02, T01)										<u>x</u>	x	x	x		3
GRMZM2G130910_T01 (T02)										x	<u>x</u>	x	x		3
GRMZM2G451374_T01													<u>x</u>		1
GRMZM2G067019_T01													<u>x</u>		1
Pp1s282_23V6.1	<u>x</u>	x													2
Pp1s93_144V6.1	<u>x</u>	x													2
Pp1s223_92V6.1			<u>x</u>	x	x										3
Pp1s47_35V6.1			<u>x</u>	x	x										3
Pp1s311_17V6.1							<u>x</u>								1
Pp1s123_132V6.1							<u>x</u>								1
Pp1s133_150V6.1								<u>x</u>							1
Pp1s101_4V6.1								<u>x</u>							1
Pp1s28_356V6.1									x	<u>x</u>		x	x		4
Pp1s183_22V6.1									x	<u>x</u>		x	x		4
Pp1s102_150V6.1											<u>x</u>	x	x		3
Pp1s76_183V6.2												x	x		2
	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3		84	
Total (máxima similitud)	3	2	6	3	1	7	5	3	5	4	1	4		44	

Notas tabla 4.4.

La equis (X) indica la presencia de las secuencias en el promotor respectivo con relación máxima. La equis ( X ) indica la secuencia que fue conservada de todas las comunes o la única, en su caso, de acuerdo a un criterio de selección y que se concentran en la tabla 4.5.

Tabla 4.5. Resumen de la depuración de las secuencias obtenidas con máxima similitud utilizando la herramienta BLAST y sus respectivos promotores.

	<i>Glycine max</i>	<i>Zea mays</i>	<i>Physcomitrella patens</i>
ATX1	-	GRMZM2G013794_T02	Pp1s282_23V6.1 Pp1s93_144V6.1
ATX2	Glyma08g29010.1 Glyma18g51890.1	-	-
ATX3	Glyma07g06190.1 Glyma03g37370.1 Glyma19g39970.1 Glyma16g02800.1	-	Pp1s223_92V6.1 Pp1s47_35V6.1
ATX4	Glyma04g41500.1 Glyma06g13330.1	GRMZM2G085266_T01	-
ATX5	Glyma19g17460.1 (.2)	-	-
ATXR3	Glyma12g11060.1 Glyma06g45740.1 Glyma13g38090.1 Glyma12g32290.1	GRMZM2G149153_T01	Pp1s311_17V6.1 Pp1s123_132V6.1
ATXR7	Glyma09g05740.1 Glyma15g17030.1	GRMZM2G473138_T01	Pp1s133_30V6.1 Pp1s101_4V6.1
ASHH1	Glyma10g36720.1 Glyma20g30870.1	GRMZM2G147619_T03 (T05 y T04)	-
ASHH2	Glyma04g42410.1 Glyma06g12390.1	GRMZM2G352431_T02 (T01)	Pp1s28_356V6.1 Pp1s183_22V6.1
ASHH3	Glyma16g33220.1 Glyma09g28430.1	GRMZM2G033694_T03	Pp1s102_150V6.1
ASHH4	-	GRMZM2G130910_T01 (T02)	-
ASHR3	Glyma06g29960.1	GRMZM2G451374_T01 GRMZM2G067019_T01	Pp1s76_183V6.2
TOTAL	22	10	12

Nota tabla 4.5. - Secuencias eliminadas durante la depuración, de acuerdo a criterios de selección.

Tabla 4.6. Resumen de genes obtenidos (BLAST) para las especies de plantas estudiadas, respecto a la clase III (Trx).

		ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
<i>Glycine max</i>	a	98	94	85	72	78	40	62	529
	b	57	58	61	61	59	40	62	398
	c	2	2	7	7	7	4	2	31
<i>Zea mays</i>	a	100	100	77	85	90	50	61	563
	b	51	48	64	48	50	28	61	350
	c	1	1	1	1	1	1	1	7
<i>Physcomitrella patens</i>	a	53	57	40	42	46	46	31	315
	b	28	33	30	32	30	30	31	214
	c	2	2	2	2	2	2	2	14

Tabla 4.7. Resumen de genes obtenidos (BLAST) para las especies de plantas estudiadas, respecto a la clase II (ASH).

		ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL	TOTAL*	
<i>Glycine max</i>	a	66	72	77	65	77	357		
	b	66	66	61	65	69	327		
	c	2	2	2	2	1	9		
<i>Zea mays</i>	a	71	68	66	64	76	345		
	b	71	48	059	64	52	294		
	c	1	1	2	2	4	10		
<i>Physcomitrella patens</i>	a	30	38	36	32	34	170		a) 2279
	b	30	32	33	32	33	160		b) 1743
	c	2	2	1	4	4	13		c) 84

Notas tablas 4.6 y 4.7. a) Total de genes identificados para cada especie.

b) Genes que contienen el dominio SET y otros dominios asociados.

c) Genes que contienen el dominio SET con máxima similitud respecto al BLAST.

\*Cuantificación total para ambas clases.

Tabla 4.8. Elementos –cis obtenidos de PlantCARE para las secuencias del promotor de las especies en estudio con respecto a la clase III (ATX).

<i>Glycine max</i> (ATX)		<i>Zea mays</i> (ATX)		<i>Physcomitrella patens</i> (ATX)	
Elemento -cis	Total	Elemento-cis	Total	Elemento-cis	Total
5UTR Py-rich stretch	24	5UTR Py-rich stretch	1	<b>3-AF1 binding site</b>	1
<b>3-AF3 binding site</b>	3	AAGAA-motif	2	<b>AAAC-motif</b>	1
AAGAA-motif	19	A-box	1	5UTR Py-rich stretch	5
A-box	2	ABRE	2	AAGAA- motif	7
ABRE	13	ACE	1	A-box	3
AC I	2	ACE II	1	ABRE	7
AC II	2	ARE	4	AC II	3
ACE	7	<b>AS-2-BOX</b>	1	ACE	5
AE-BOX	7	<b>ATGCAAAT motif</b>	1	AC-I	1
ARE	23	<b>AuxRR-core</b>	2	AE-BOX	4
<b>AT1-motif</b>	7	Box 4	3	ARE	8
ATCT-MOTIF	4	<b>box E</b>	1	<b>AS-2-BOX</b>	2
AT-RICH ELEMENT	3	<b>Box III</b>	2	<b>ATCC-motif</b>	1
<b>AuxRR-core</b>	3	<b>Box-W1</b>	4	<b>ATC-motif</b>	1
Box 4	35	CAAT-box	93	ATCT-MOTIF	3
<b>box E</b>	1	CATT-motif	2	AT-RICH ELEMENT	2
BOX I	14	CCAAT-box	1	Box 4	6
BOX II	1	CCGTCC-box	1	BOX I	7
<b>Box III</b>	4	CGTCA-motif	2	BOX II	1
BOX S	1	chs-CMA2a	1	CAAT-box	208
<b>Box-W1</b>	4	circadian	6	CAT-BOX	5
CAAT-box	505	E2Fa	1	<b>CATTAT- motif</b>	2
CAT-BOX	2	<b>EIRE</b>	1	CATT-motif	6
CATT-motif	7	GAG-motif	5	CCAAT-box	4
CCAAT-box	5	GA-motif	1	CCGTCC-box	4
CCGTCC-box	2	GARE-motif	1	CGTCA- motif	18
<b>CE3</b>	1	GATA-motif	1	chs-CMA2a	1
CGTCA-motif	8	G-Box	3	CHS-UNIT 1 M1	2
<b>chs-CMA1a</b>	2	G-box	7	circadian	14
circadian	19	GC-motif	9	CTAG-MOTIF	2
chs-CMA2a	4	GCN4 motif	3	E2Fa	1
CHS-UNIT1 M1	2	GT1-motif	2	ELI-BOX3	1
CTAG-MOTIF	1	HSE	3	GA-motif	1
E2Fa	2	I-box	5	GAG-motif	6
<b>EIRE</b>	1	LTR	3	GARE-motif	3
ELI-BOX3	2	MBS	8	GATA-motif	3
<b>ERE</b>	1	MNF1	3	GATT-MOTIF	1
<b>F-box</b>	2	<b>motif Ilb</b>	1	G-Box	8
GAG-motif	8	<b>NON-box</b>	1	G-box	20
GA-motif	7	O2-site	3	GC-motif	4

Tabla 4.8. (continuación)

<i>Glycine max</i> (ATX)		<i>Zea mays</i> (ATX)		<i>Physcomitrella patens</i> (ATX)	
Elemento -cis	Total	Elemento-cis	Total	Elemento-cis	Total
GARE-motif	17	<b>OCT</b>	1	GCN4_motif	3
GATA-motif	6	Skn-1_motif	12	GT1-motif	1
GATT-MOTIF	1	Sp1	14	<b>HD-Zip 3</b>	1
G-Box	13	TATA-box	112	HSE	6
G-box	28	TATCCAT/C-motif	1	I-box	7
GC-motif	5	TCA-element	3	LTR	6
GCN4_motif	16	<b>TCCACCT-motif</b>	2	MBS	7
GT1-motif	20	TC-rich repeat	1	MNF1	2
HSE	22	TCT-motif	2	MRE	2
I-box	3	TGACG-motif	2	O2-site	8
<b>LAMP-element</b>	1	TGA-element	2	P-BOX	3
<b>L-box</b>	2	Unnamed_1	13	Skn-1_motif	31
LTR	4	<b>UNNAMED_15</b>	1	Sp1	60
MBS	18	<b>UNNAMED_16</b>	2	TATA-box	92
MNF1	2	<b>Unnamed_17</b>	1	TATC-BOX	1
MRE	5	Unnamed_2	6	TATCCAT/C-motif	1
O2-site	4	Unnamed_3	13	TCA-element	7
<b>OBP-1 site</b>	1	Unnamed_4	59	TCCC-MOTIF	3
P-BOX	4	<b>W box</b>	4	TC-rich repeats	6
Skn-1_motif	35	TOTAL	444	TCT-motif	7
Sp1	14			TGACG-motif	18
<b>TA-rich region</b>	8			TGA-element	6
TATA-box	832			Unnamed_1	21
TATC-BOX	1			UNNAMED_11	1
TATCCAT/C-motif	3			<b>UNNAMED_15</b>	1
TCA_element	16			<b>UNNAMED_16</b>	1
<b>TCCACCT-motif</b>	2			Unnamed_2	2
TCCC-MOTIF	5			Unnamed_3	21
TC-rich repeats	11			Unnamed_4	136
TCT-motif	15			WUN-MOTIF	1
TGACG-motif	8			BOX S	1
TGA-element	5			<b>motif I</b>	1
Unnamed_1	24			TOTAL	846
UNNAMED_11	5				
Unnamed_2	6				
Unnamed_3	17				
Unnamed_4	115				
<b>Unnamed_6</b>	5				
<b>W box</b>	4				
WUN-MOTIF	3				
TOTAL	2068				

Notas Tabla 4.8. Elementos –cis comunes a las tres especies se muestran en arial, en **negritas cursivas** los que son únicos en cada especie, en **Impact** son comunes entre las parejas *Glycine max* y *Zea mays*, en **CASTELLAR** son comunes entre las parejas *Glycine max* y *Physcomitrella patens* y en **STENCIL** entre *Zea mays* y *Physcomitrella patens*.

De entre todos los elementos –cis encontrados para las tres especies (ver tablas 4.8 y 4.9) unos pocos son específicos para cada una de ellas (ver tablas 4.8e y 4.9e), otras más son comunes entre pares (ver tablas 4.8b,4.8c,4.8d y 4.9b,4.9c,4.9d) y la mayoría, comunes a las tres (ver tablas 4.8a y 4.9a).

Tabla 4.8a. Elementos -cis comunes a las tres especies de plantas estudiadas (*Glycine max*, *Zea mays* y *Physcomitrella patens*) y *Arabidopsis thaliana*, homólogas de ATX.

Elemento -cis	<i>Glycine max</i>	<i>Zea mays</i>	<i>Physcomitrella patens</i>	Total	<i>Arabidopsis thaliana</i>
5UTR Py-rich stretch	24	1	5	30	ABRE-like binding site motif
AAGAA-motif	19	2	7	28	ARF binding site motif
A-box	2	1	3	6	ARF1 binding site motif
ABRE	13	2	7	22	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH
AC II	2	1	3	6	ATHB1 binding site moti
ACE	7	1	5	13	ATHB2 binding site motif
ARE	23	4	8	35	ATHB5 binding site motif
Box 4	35	3	6	44	AtMYC2 BS in RD22
CAAT-box	505	93	208	806	Bellringer/replumless/pennywise BS1 IN AG
CATT-motif	7	2	6	15	BoxII promoter motif
CCAAT-box	5	1	4	10	CARG promoter motif
CCGTCC-box	2	1	4	7	CCA1 binding site motif
CGTCA-motif	8	2	18	28	DPBF1&2 binding site motif
circadian	19	6	14	39	DRE-like promoter motif
chs-CMA2a	4	1	1	6	E2F binding site motif
E2Fa	2	1	1	4	E2F-varient binding site motif
GA-motif	7	1	1	9	GATA promoter motif [LRE]
GAG-motif	8	5	6	19	GCC-box promoter motif
GARE-motif	17	1	3	21	Hexamer promoter motif
GATA-motif	6	1	3	10	HSEs binding site motif
G-Box	13	3	8	24	lbox promoter motif
G-box	28	7	20	55	L1-box promoter motif
GC-motif	5	9	4	18	LFY consensus binding site motif
GCN4_motif	16	3	3	22	MYB binding site promoter
GT1-motif	20	2	1	23	MYB1 binding site motif

Tabla 4.8a (continuación)

Elemento -cis	<i>Glycine max</i>	<i>Zea mays</i>	<i>Physcomitrella patens</i>	Total	<i>Arabidopsis thaliana</i>
HSE	22	3	6	31	MYB3 binding site motif
I-box	3	5	7	15	MYB4 binding site motif
LTR	4	3	6	13	RAV1-A binding site motif
MBS	18	8	7	33	RAV1-B binding site motif
MNF1	2	3	2	7	RY-repeat promoter motif
O2-site	4	3	8	15	SBP-box promoter motif
Skn-1_motif	35	12	31	78	SORLIP1
Sp1	14	14	60	88	SORLIP2
TATA-box	832	112	92	1036	SORLIP5
TATCCAT/C-motif	3	1	1	5	SORLREP2
TCA_element	16	3	7	26	SORLREP3
TC-rich repeats	11	1	6	18	T-box promoter motif
TCT-motif	15	2	7	24	TELO-box promoter motif
TGACG-motif	8	2	18	28	TGA1 binding site motif
TGA-element	5	2	6	13	W-box promoter motif
Unnamed_1	24	13	21	58	
Unnamed_2	6	6	2	14	
Unnamed_3	17	13	21	51	
Unnamed_4	115	59	136	310	

Tabla 4.8b. Elementos -cis comunes a *Glycine max* y *Zea mays* homólogos de ATX.

Elemento -cis	<i>Glycine max</i>	<i>Zea mays</i>	Total
AuxRR-core	3	2	5
Box III	4	2	6
Box-W1	4	4	8
EIRE	1	1	2
TCCACCT-motif	2	2	4
W box	4	4	8
box E	1	1	2

Tabla 4.8c. Elementos -cis comunes a *Glycine max* y *Physcomitrella patens* homólogos de ATX.

Elemento -cis	<i>Glycine max</i>	<i>Physcomitrella patens</i>	Total
AC I	2	1	3
AE box	7	4	11
ATCT-motif	4	3	7
AT-rich element	3	2	5
Box I	14	7	21
Box II	1	1	2
box S	1	1	2
CAT-box	2	5	7
chs -Unit1 m1	2	2	4
CTAG-motif	1	2	3
ELI-box3	2	1	3
GATT-motif	1	1	2
MRE	5	2	7
P-box	4	3	7
TATC-box	1	1	2
TCCC-motif	5	3	8
Unnamed_11	5	1	6
WUN-motif	3	1	4

Tabla 4.8d. Elementos -cis comunes a *Zea mays* y *Physcomitrella patens* homólogos de ATX.

Elemento -cis	<i>Zea mays</i>	<i>Physcomitrella patens</i>	Total
As-2-box	1	2	3
Unnamed_15	1	1	2
Unnamed_16	2	1	3

Tabla 4.8e. Elementos -cis ÚNICOS en cada una de las tres especies de plantas estudiadas (*Glycine max*, *Zea mays* y *Physcomitrella patens*) homólogas de ATX.

<i>Glycine max</i>		<i>Zea mays</i>		<i>Physcomitrella patens</i>	
Elemento -cis	Total	Elemento -cis	Total	Elemento -cis	Total
3-AF3 binding site	3	ATGCAAAT motif	1	AAAC-motif	1
AT1-motif	7	motif IIb	1	3-AF1 binding site	1
CE3	1	NON-box	1	ATCC-motif	1
chs-CMA1a	2	OCT	1	ATC-motif	1
ERE	1	Unnamed_17	1	CATTAT motif	2
F-box	2			HD-Zip 3	1
LAMP-element	1			motif I	1
L-box	2				
OBP-1 site	1				
TA-rich region	8				
Unnamed_6	5				

Tabla 4.9. Elementos –cis obtenidos de PlantCARE para las secuencias del promotor de las especies en estudio con respecto a la clase II (ASH).

<i>Glycine max</i> (ASH)		<i>Zea mays</i> (ASH)		<i>Physcomitrella patens</i> (ASH)	
Elemento -cis	Total	Elemento -cis	Total	Elemento -cis	Total
<b>3-AF1 binding site</b>	1	<b>3-AF1 binding site</b>	2	5UTR Py-rich stretch	4
<b>4cl-CMA2b</b>	1	<b>4cl-CMA2b</b>	1	AAGAA-motif	6
5UTR Py-rich stretch	8	5UTR Py-rich stretch	9	A-box	4
AAGAA-motif	6	<b>AAAC-motif</b>	1	ABRE	5
A-box	1	AAGAA-motif	2	AC-I	1
ABRE	2	A-box	2	ARE	6
<b>AC II</b>	2	ABRE	18	as-2-box	3
<b>ACE</b>	4	<b>ACE</b>	4	ATCT-motif	1
AC-I	1	AC-I	1	ATGCAAAT motif	1
<b>AE-box</b>	8	<b>AC-II</b>	1	<b>AuxRR-core</b>	2
ARE	19	<b>AE-box</b>	2	Box 4	2
as-2-box	2	ARE	9	Box I	1
<b>AT1-motif</b>	1	as-2-box	2	<b>box II</b>	1
ATCT-motif	1	<b>ATC-motif</b>	1	Box III	1
ATGCAAAT motif	2	ATCT-motif	2	BOX-WI	2
Box 4	15	ATGCAAAT-motif	2	CAAT-box	81
Box I	5	Box 4	3	CAT-box	3
Box III	2	Box I	5	CATT-motif	3
BOX-WI	2	Box III	1	CCAAT-box	2
CAAT-box	226	<b>box S</b>	2	CCGTCC-box	4
CAT-box	6	CAAT-box	126	CGTCA-motif	10
CATT-motif	8	CAT- box	4	circadian	3
CCAAT-box	2	CATT-motif	6	<b>E2F</b>	1
CCGTCC-box	1	<b>C-box</b>	1	<b>EIRE</b>	1
CGTCA-motif	7	CCAAT-box	3	GAG-motif	7
<b>chs-CMA1a</b>	1	CCGTCC-box	2	GA-motif	1
circadian	5	<b>CE1</b>	1	GARE-motif	2
<b>csH-Unit1 m1</b>	1	CGTCA-motif	14	G-Box	5
<b>CTAG-motif</b>	1	<b>chs-CMA2a</b>	1	G-box	11
<b>E2Fa</b>	1	circadian	4	GCN4_motif	2
<b>ERE</b>	2	<b>CTAG-motif</b>	1	GT1-motif	5
GAG-motif	11	<b>ERE</b>	3	HSE	3
GA-motif	1	<b>F-box</b>	1	I-box	2
GARE-motif	6	GAG-motif	5	LTR	2
<b>GATA-motif</b>	4	GA-motif	1	MBS	10
<b>GATT-motif</b>	1	GARE-motif	4	MNFI	1
G-Box	3	<b>GATA-motif</b>	2	MRE	2
G-box	11	G-Box	10	<b>MSA-like</b>	1
GCN4_motif	2	G-box	25	O2-site	2
GT1-motif	2	<b>GCC box</b>	1	<b>RY-element</b>	1
HSE	10	<b>GC-motif</b>	4	Skn-1_motif	8
I-box	5	GCN4_motif	2	Sp1	14

Tabla 4.9. (continuación)

<i>Glycine max</i> (ASH)		<i>Zea mays</i> (ASH)		<i>Physcomitrella patens</i> (ASH)	
Elemento -cis	Total	Elemento -cis	Total	Elemento -cis	Total
<b>LAMP-element</b>	1	GT1-motif	8	<b>TA-rich region</b>	11
<b>L-box</b>	1	HSE	5	TATA-box	123
LTR	4	I-box	5	<b>TATC-BOX</b>	1
MBS	14	<b>LAMP-element</b>	1	TCA-element	6
<b>MBSII</b>	1	<b>L-box</b>	1	TCCC-motif	4
MNF1	1	MBS	7	TC-rich repeats	5
MRE	2	<b>motif IIb</b>	2	TCT-motif	2
O2-site	4	MRE	1	TGACG-motif	10
<b>P-box</b>	2	O2-site	2	<b>TGA-ELEMENT</b>	3
<b>Pc-CMA2a</b>	1	<b>P-box</b>	1	<b>TGGCA-motif</b>	1
<b>sbp-CMA1a</b>	1	Skn-1_motif	12	Unnamed_1	5
Skn-1_motif	9	Sp1	18	<b>UNNAMED_10</b>	1
Sp1	37	TATA-box	271	<b>UNNAMED_12</b>	1
TATA-box	305	<b>TATC-BOX</b>	1	<b>Unnamed_13</b>	1
<b>TATCCAT/C-motif</b>	2	TCA-element	3	<b>UNNAMED_14</b>	1
TCA_element	11	<b>TCCACCT-motif</b>	1	<b>Unnamed_16</b>	1
TCCC-motif	2	TCCC-motif	1	Unnamed_3	5
TC-rich repeats	7	TC-rich repeats	5	Unnamed_4	76
TCT-motif	3	TCT-motif	4	<b>UNNAMED_8</b>	1
TGACG-motif	7	TGACG-motif	13	<b>UNNAMED_9</b>	1
Unnamed_1	6	<b>TGA-ELEMENT</b>	3	UNNAMED_2	4
<b>Unnamed_11</b>	1	Unnamed_1	8	W BOX	2
UNNAMED_2	1	<b>UNNAMED_10</b>	1	WUN-MOTIF	1
Unnamed_3	6	<b>Unnamed_11</b>	2	TOTAL	493
Unnamed_4	61	<b>UNNAMED_12</b>	1		
<b>Unnamed_6</b>	1	<b>UNNAMED_14</b>	1		
W BOX	2	<b>Unnamed_15</b>	2		
WUN-MOTIF	2	Unnamed_3	8		
TOTAL	895	Unnamed_4	79		
		<b>Unnamed_6</b>	1		
		<b>UNNAMED_8</b>	1		
		<b>UNNAMED_9</b>	1		
		TOTAL	761		

Notas Tabla 4.9. Elementos –cis comunes a las tres especies se muestran en arrial, en **negritas cursivas** los que son únicos en cada especie, en **Impact** son comunes entre las parejas *Glycine max* y *Zea mays*, en **CASTELLAR** son comunes entre las parejas *Glycine max* y *Physcomitrella patens* y en **STENCIL** entre *Zea mays* y *Physcomitrella patens*.

Tabla 4.9a. Elementos -cis comunes a las tres especies de plantas estudiadas (*Glycine max*, *Zea mays* y *Physcomitrella patens*) y *Arabidopsis thaliana*, homólogos de ASH.

Elemento -cis	<i>Glycine max</i>	<i>Zea mays</i>	<i>Physcomitrella patens</i>	Total	<i>Arabidopsis thaliana</i>
5UTR Py-rich stretch	8	9	4	21	ABRE-like binding site motif
AAGAA-motif	6	2	6	14	ARF binding site motif
A-box	1	2	4	7	ARF1 binding site motif
ABRE	2	18	5	25	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH
AC-I	1	1	1	3	ATHB1 binding site moti
ARE	19	9	6	34	ATHB2 binding site motif
as-2-box	2	2	3	7	ATHB5 binding site motif
ATCT-motif	1	2	1	4	AtMYC2 BS in RD22
ATGCAAAT motif	2	2	1	5	Bellringer/replumless/pennywise BS1 IN AG
Box 4	15	3	2	20	BoxII promoter motif
Box I	5	5	1	11	CArG promoter motif
Box III	2	1	1	4	CCA1 binding site motif
CAAT-box	226	126	81	433	DPBF1&2 binding site motif
CAT-box	6	4	3	13	DRE-like promoter motif
CATT-motif	8	6	3	17	E2F binding site motif
CCAAT-box	2	3	2	7	E2F-varient binding site motif
CCGTCC-box	1	2	4	7	GATA promoter motif [LRE]
CGTCA-motif	7	14	10	31	GCC-box promoter motif
circadian	5	4	3	12	Hexamer promoter motif
GAG-motif	11	5	7	23	HSEs binding site motif
GA-motif	1	1	1	3	Ibox promoter motif
GARE-motif	6	4	2	12	L1-box promoter motif
G-box	11	25	11	47	LFY consensus binding site motif
G-Box	3	10	5	18	MYB binding site promoter
GCN4_motif	2	2	2	6	MYB1 binding site motif
GT1-motif	2	8	5	15	MYB3 binding site motif
HSE	10	5	3	18	MYB4 binding site motif
I-box	5	5	2	12	RAV1-A binding site motif
MBS	14	7	10	31	RAV1-B binding site motif
MRE	2	1	2	5	RY-repeat promoter motif
O2-site	4	2	2	8	SBP-box promoter motif
Skn-1_motif	9	12	8	29	SORLIP1
Sp1	37	18	14	69	SORLIP2
TATA-box	305	271	123	699	SORLIP5
TCA_element	11	3	6	20	SORLREP2
TCCC-motif	2	1	4	7	SORLREP3
TC-rich repeats	7	5	5	17	T-box promoter motif
TCT-motif	3	4	2	9	TELO-box promoter motif
TGACG-motif	7	13	10	30	TGA1 binding site motif

Tabla 4.9a (continuación)

Elemento -cis	<i>Glycine max</i>	<i>Zea mays</i>	<i>Physcomitrella patens</i>	Total	<i>Arabidopsis thaliana</i>
Unnamed_1	6	8	5	19	W-box promoter motif
Unnamed_3	6	8	5	19	
Unnamed_4	61	79	76	216	

Tabla 4.9b. Elementos -cis comunes a *Glycine max* y *Zea mays* homólogos de ASH.

Elemento -cis	<i>Glycine max</i>	<i>Zea mays</i>	Total
3-AF1 binding site	1	2	3
4cl-CMA2b	1	1	2
AC II	2	1	3
ACE	4	4	8
AE-box	8	2	10
CTAG-motif	1	1	2
ERE	2	3	5
GATA-motif	4	2	6
LAMP-element	1	1	2
L-box	1	1	2
P-box	2	1	3
Unnamed_6	1	1	2
Unnamed_11	1	2	3

Tabla 4.9c. Elementos -cis comunes a *Glycine max* y *Physcomitrella patens* homólogos de ASH.

Elemento -cis	<i>Glycine max</i>	<i>Physcomitrella patens</i>	Total
Box-W1	2	2	4
LTR	4	2	6
MNF1	1	1	2
Unnamed_2	1	4	5
W box	2	2	4
WUN-motif	2	1	3

Tabla 4.9d. Elementos -cis comunes a *Zea mays* y *Physcomitrella patens* homólogos de ASH.

	<i>Zea mays</i>	<i>Physcomitrella patens</i>	
Elemento -cis			Total
TATC-box	1	1	2
TGA-element	3	3	6
Unnamed_10	1	1	2
Unnamed_12	1	1	2
Unnamed_14	1	1	2
Unnamed_8	1	1	2
Unnamed_9	1	1	2

Tabla 4.9e. Elementos -cis ÚNICOS en cada una de las tres especies de plantas estudiadas (*Glycine max*, *Zea mays* y *Physcomitrella patens*) homólogos de ASH.

<i>Glycine max</i>		<i>Zea mays</i>		<i>Physcomitrella patens</i>	
Elemento -cis	Total	Elemento -cis	Total	Elemento -cis	Total
AT1-motif	1	AAAC-motif	1	AuxRR-core	2
chs-CMA1a	1	ATC-motif	1	box II	1
chs-Unit1 m1	1	box S	2	E2F	1
E2Fa	1	C-box	1	EIRE	1
GATT-motif	1	CE1	1	MSA-like	1
MBSII	1	chs-CMA2a	1	RY-element	1
Pc-CMA2a	1	F-box	1	TA-rich region	11
sbp-CMA1a	1	GCC box	1	TGGCA-motif	1
TATCCAT/C-motif	2	GC-motif	4	Unnamed_13	1
		motif IIb	2	Unnamed_16	1
		TCCACCT-motif	1		
		Unnamed_15	2		

## 5. DISCUSIÓN DE RESULTADOS

### 5.1 Análisis de secuencias–cis en *Arabidopsis thaliana*.

En nuestro análisis para los promotores de genes de *Arabidopsis thaliana*, gráfica y numéricamente se destacan dos motivos en la familia: RAV1-A binding site motif CAACA y GATA promoter motif [LRE] AGATAA / TGATAA/ TGATAG (gráfica 4.1).

RAV1-A, en la clase II representa la mayoría pero es segundo en la clase III y para GATA la situación es opuesta, solo superado por DPBF1&2 binding site motif ACACATG/ ACACAGG en la clase II que también está presente en la clase III seguidos de los motivos MYB4 binding site motif AACTAAC/ AACAAAC/ ACCAACC, LFY (LEAFY) consensus binding site motif CCAATG/ CCATTG/ CCAGTG, ARF1 (Auxin Response Factor 1) binding site motif TGTCTC, W-box promoter motif TTGACT/ TTGACC, ARF (Auxin Response Factor) binding site motif TGTCTC y BoxII promoter motif GGTTAA que también se encuentran en un porcentaje considerable en esta clase III y en la familia (ver tabla 4.3).

RAV1-A, es una secuencia de reconocimiento bipartito (junto con RAV1-B) a través de dos distintos dominios de unión (AP2-like y B3-like) del factor de transcripción RAV1. Los dominios AP2-like y B3-like reconocen los motivos CAACA (RAV1-A) y CACCTG (RAV1-B) respectivamente. Se ha reportado que los niveles de expresión de RAV1 fueron relativamente altas en hojas, roseta y raíces; y que únicamente se encuentran en las plantas superiores.

El motivo GATA en el promotor participa en factores: de organización, de expresión y de unión al ADN. Es necesario para regular un alto nivel de luz, y la expresión específica de tejido, por lo que se considera un elemento de respuesta a la luz.

DPBF1&2 binding site motif (Dc3 promoter-binding factor-1 and 2) es una secuencia central vinculante; interactúan con los elementos que responden a ABA. La expresión de Dc3 es normalmente embrión específico, y también puede ser inducida por ABA, y regula un subconjunto de genes al final de la embriogénesis.

En relación a “MYB4 binding site motif”, el perfil de expresión de los genes de *Arabidopsis* que responden a estrés hídrico, sugiere su función putativa es en respuesta al estrés ambiental. Participa en la regulación de los genes que es inducida por el estrés y la deshidratación.

LFY (LEAFY) consensus binding site motif es una secuencia blanco en el intrón del gen *AGAMOUS* en *Arabidopsis*; gen que se expresa específicamente en las células centrales en el meristemo radicular.

El motivo ARF1 y ARF participa en la dimerización y unión al ADN de los factores de respuesta a auxinas.

W-box promoter motif se encuentra en el promotor del gen *NPR1* de *Arabidopsis thaliana* y existe evidencia de un papel importante de las proteínas del WRKY vinculante al ADN en la regulación de la expresión génica de *NPR1*. Otras funciones que se le han asignado son la respuesta al estrés ambiental y la defensa contra patógenos.

BoxII promoter motif, participa en la activación de la transcripción de *Arabidopsis* GT-1 que puede ser a través de la interacción con el complejo TFIIA-TBP-TATA.

Las funciones para estos elementos representativos en *Arabidopsis thaliana* nos indican que el control de la transcripción se realiza básicamente en función de las condiciones ambientales, ello sugiere, generalizando, que en las plantas los niveles de transcripción pueden variar y aun modificarse. Por experiencia sabemos, que cualquier planta en condiciones adecuadas se desarrolla mejor que aquellas con condiciones inadecuadas lo que es sinónimo de una eficiente o deficiente transcripción. Como veremos más adelante, solo algunos de los elementos –cis presentes en los promotores de genes de *Arabidopsis thaliana* son comunes en los promotores de las especies estudiadas (tablas 4.8a y 4.9a), por ejemplo, GCC-box (AGCCGCC) únicamente aparece en la especie *Zea mays* de la clase II (GRMZM 2G147619\_ T03 homólogo de ASHH1) y en *Arabidopsis thaliana* de la clase III (ATX1) y no es encontrado en *Glycine max* y

*Physcomitrella patens* en ambas clases. Por otro lado, E2Fa (TTTCCCGC) se encuentra en las tres especies (homólogas de ATX) y en *Glycine max* (homólogo de ASH) y no aparece en *Zea mays* de la clase II, mientras que en *Physcomitrella patens* (Pp1s183\_22V6.1 homólogo de ASHH1) aparece la variante E2F (AGTGGCGGNNNNNTTTGAA) al igual que en *Arabidopsis thaliana* en el promotor de ASHR3. GCC box tiene la capacidad de responder a heridas y a patógenos mientras que E2Fa (nombre del sitio) y E2F binding site (nombre largo) participan en la transición del ciclo celular. En su oportunidad se ampliara esta comparación de los elementos –cis presentes en las cuatro especies en cuanto a su conservación y su presencia/ausencia vista desde el punto de vista de sus distancias filogenéticas.

## **5.2. Análisis de los elementos –cis en genes de la familia trithorax en tres distintos genomas de plantas.**

Para nuestro segundo objetivo específico, realizamos el descubrimiento de los elementos –cis en 44 promotores de genes de *Glycine max*, *Zea mays* y *Physcomitrella patens* utilizando un enfoque de identificación de los motivos (secuencias cortas) que son estadísticamente significativos (enriquecidos) en este conjunto de secuencias de promotor en comparación con el fondo de un conjunto de secuencias conocidas.

Estas 44 secuencias de promotor de genes fueron seleccionadas de las 88 que contienen el dominio SET (tablas 4.4 y 4.5) bajo el criterio de su similitud en base a sus valores de E (E-value, parámetro que describe el número de aciertos (hits) que se puede "esperar" ver por casualidad, al buscar en una base de datos de un tamaño determinado), la tabla 4.5 que reúne estas secuencias es consistente con la de Ng y colaboradores (tabla 3.1; Ng et al.,2007) en cuanto a *Zea mays*, sin embargo, aparecen diferencias debido a que los criterios de selección no son similares, ellos utilizaron las relaciones filogenéticas entre secuencias homólogas y la arquitectura de sus dominios.

El valor E da una idea de la significación estadística de un alineamiento dado por parejas y refleja el tamaño de la base de datos y el sistema de puntuación utilizado (S, score). Cuanto menor sea el valor E, es más significativo el hit (E=0). Un alineamiento

de secuencias que tiene un E-valor de 0.05, por ejemplo, significa que esta similitud tiene un 5 en 100 (1 en 20) probabilidades de que ocurra por casualidad (la literatura indica que un valor de  $E \leq 1e-04$  puede considerarse homólogo o relacionado). A pesar de que un estadístico podría considerar que esto es importante, todavía no puede representar un resultado biológicamente significativo, y el análisis experimental de las alineaciones es necesario para determinar la significancia "biológica" de las secuencias encontradas.

Para un mejor entendimiento de la clasificación en la tabla 4.5, se consideran las secuencias de ATX1 y de ATX2. La alta puntuación en parejas de segmentos HSPs (high-scoring segment pairs) fue considerada, la cual refleja el criterio de ubicación en general dentro de la misma (ver anexo 2). Se observa como para los dos promotores se obtuvieron las mismas secuencias (comunes) y en el mismo orden, además, todas con un valor de  $E=0$  lo cual permite considerarlas como homólogas o relacionadas a las secuencias utilizadas como fondo de consulta (ATX1 y ATX2 en este caso). Sin embargo, los otros parámetros considerados: identidad y similitud, muestran que Glyma08g29010.1 y Glyma18g51890.1 en ATX2 presentan una homología mayor en cuanto a su porcentaje de similitud que aquellos en ATX1, razón por la cual se han ubicado en la tabla 4.5 en el apartado de ATX2.

Del mismo modo, GRMZM2G013794\_T02 en ATX1 presenta ligeramente una mayor homología que en ATX2, lo mismo ocurre con Pp1s282\_23V6.1 y Pp1s93\_144V6.1 y por lo tanto, todas ellas se ubican en el apartado de ATX1. Esta es la razón por la cual aparecen celdas vacías en la tabla 4.5 (Para una mayor comprensión ver anexo 2).

Tabla que muestra parte del anexo 2.

	Orden	Secuencias	S (score) bits	E (E-value)	Identidad (%)	Similitud (%)
ATX1	1o	. Glyma08g29010.1	1120.9	0	56.2	69.7
	2o	Glyma18g51890.1	1104.7	0	55.6	68.9
	1o	GRMZM2G013794_T02	936.8	0	54.7	69.7
	1o	Pp1s282_23V6.1	784.3	0	45.7	59.8
	2o	Pp1s93_144V6.1	775.8	0	44.9	59.2
ATX2	1o	Glyma08g29010.1	1147.9	0	61.3	74.6
	2o	Glyma18g51890.1	1142.1	0	61.3	74.1
	1o	GRMZM2G013794_T02	948.7	0	53.9	68.6
	1o	Pp1s282_23V6.1	775.0	0	44.8	59.2
	2o	Pp1s93_144V6.1	765.0	0	44.0	59.2

Por el contrario mientras que para ASHH2, ASHH4 y ASHR3 Ng y colaboradores (2007) no reportan secuencias para *Zea mays*, en la tabla 4.5 se observa que sí aparecen estas. Nuevamente, el criterio de selección es diferente y más aun la clasificación de los primeros. Actualmente la base de datos pudiera haberse modificado debido a que las actualizaciones en este campo son muy frecuentes lo cual podría ser una razón más para estas diferencias, incluso aun, que nuestra metodología requiera que se consideren otros aspectos o parámetros, sin pasar por alto que el nuestro es un análisis *in silico*, sin embargo, nuevamente, ambos análisis son consistentes al menos en este punto.

Las regiones reguladoras del genoma codifican la información de regulación transcripcional utilizando elementos regulatorios incrustados en las secuencias de fondo y de este modo podemos ver las regiones reguladoras de los genes de interés que contienen los elementos –cis con nuestras secuencias de fondo. Los elementos –cis encontrados son típicamente más conservados y estadísticamente más sobre representados que aquellos en la secuencia de fondo. Esto es particularmente cierto para las secuencias genómicas reguladoras, donde un pequeño número de factores de transcripción regulan una gran cantidad de genes, haciendo de los elementos -cis funcionales más representados.

Estudios previos sobre la comparación de secuencias han demostrado que genes similares unos a otros en el nivel de nucleótidos pueden estar relacionados entre sí por diferentes historias evolutivas. Aunque algunos pocos ejemplos de evolución convergente (dos genes previamente no relacionados, que se transformaron similares a medida que adquirieron nuevas funciones relacionadas) se han reportado, la mayoría de los enfoques genómicos comparativos investigan las secuencias que están relacionadas por la evolución divergente de un ancestro común que es nuestro caso. Genes derivados de un gen ancestral común son homólogos, y el nivel de similitud en sus secuencias a menudo refleja el tiempo transcurrido desde que se separaron (Frazer et al., 2003).

En nuestro análisis, se ha observado que la similitud en la homología entre las especies es diferente, esto es, las secuencias de *Glycine max* y *Zea mays* son similares en porcentajes superiores al 50%, mientras que entre *Zea mays* y *Physcomitrella patens* son inferiores al 50% (entre un 30% y un 48%); por lo tanto, entre *Glycine max* y *Physcomitrella patens*, también presentan diferencias, por lo general, inferior como sucede con *Zea mays*. Esto podría interpretarse como una razón más para su ubicación en el árbol filogenético que muestra la separación de las especies, es claro que la divergencia espacio temporal se presenta entre las especies estudiadas dado que la similitud en la homología de sus secuencias indica la proximidad relativa entre *Glycine max* y *Zea mays* y una separación relativa entre estas y *Physcomitrella patens*. La cantidad de secuencias cortas (motivos) altamente conservadas disminuyó con el aumento de la distancia filogenética. Las tablas 4.8, 4.9 y derivadas dan muestra de lo anterior.

De los 99 diferentes elementos –cis encontrados para los promotores del grupo II (ASH) el 42.42% (42/99) es compartido por las tres especies, *Glycine max* y *Zea mays* (*Gm-Zm*) comparten el 13.13% (13/99), *Glycine max* y *Physcomitrella patens* (*Gm-Pp*) el 6.06% (6/99) y *Zea mays* y *Physcomitrella patens* (*Zm-Pp*) el 7.07% (7/99). Mientras el 9.09% (9/99) es específico para *Glycine max* (*Gm*), el 12.12% (12/99) lo es para *Zea mays* (*Zm*) y el 10.10% (10/99) para *Physcomitrella patens* (*Pp*). En lo siguiente, nos

referiremos a las especies y parejas de ellas como se ha indicado arriba, *Gm-Zm*, *Gm-Pp*, *Zm-Pp*, etc.

Para los 95 elementos –cis en los promotores del grupo III (ATX) el 46.32% (44/95) es compartido por las tres especies, *Gm-Zm* comparten el 7.37% (7/95), *Gm-Pp* el 18.94% (18/95) y *Zm-Pp* el 3.16% (3/95). Mientras el 11.58% (11/95) es específico para *Gm*, el 5.26% (5/95) lo es para *Zm* y el 7.37% (7/95) para *Pp*.

Los porcentajes anteriores reflejan que, por lo menos en este análisis, la mayoría de los elementos –cis son conservados en las tres especies y además existen elementos específicos para cada una de ellas que pudieron haberse diversificado funcionalmente de otros también conservados como muestran sus porcentajes relativamente bajos. Además, los porcentajes entre parejas nos dan un indicio de su separación como especies en la evolución, sin embargo, debe hacerse notar que esto está basado en la información actual, es decir, que las diferentes especies de plantas, incluidas estas, hayan sido estudiadas unas más que otras o que aun no se han reportado nuevos elementos –cis que pudieran hacer alguna diferencia.

Un análisis más profundo de las tablas 4.8, 4.9 y derivadas indica que los elementos –cis comunes presentes entre pares de especies son encontrados también en el conjunto de elementos –cis comunes a las tres especies pero del otro grupo, sin embargo, siguen existiendo -aunque pocos- comunes o específicos entre parejas como se observa en la tabla 5.1, también se incluyen sus secuencias donde es interesante notar que la mayoría contienen 9 o más miembros (nucleótidos).

El elemento 4cl-CMA2b (TCTCACCAACCCCA) es un elemento regulador –cis que forma parte del “4cl conserved DNA module array (4cl-CMA2)” involucrado con la respuesta a luz (eventualmente UV), también llamado L-box (CTCACCTACCAA) por ser similares, no idénticos, asociado con A-box (CCGTCC) y P-box (GCCTTTTGAGT) se involucran en inducir la actividad transcripcional, pero no suficiente para conferir. 4cl-CMA2b y L-box se ubican en la pareja *Gm-Zm* de ASH con una frecuencia de 1/1 y 1/1 respectivamente, incluso en las mismas secuencias (Glyma04g42410.1 de ASHH2 y

RMZM2G033694\_T03 de ASHR3) pero L-box también se encuentra como un elemento único en *Gm* de ATX con una frecuencia de 2 (Glyma13g38090.1 y Glyma12g32290.1, ambas de ATXR3) como puede verse en el anexo 3, lo que nos permite deducir que este sitio es más específico para *Gm*, sin embargo, dado que se encuentran en la pareja con idéntica función y sus secuencias son muy similares, podríamos tener un ejemplo de co-conservación y con ello la especiación.

Como se observa en ambas tablas (5.1 y 5.2) los elementos –cis de *Gm-Pp* (ASH) y *Zm-Pp* (ATX) no son incluidos por encontrarse en los grupos de ATX y ASH, respectivamente, de las tres especies o en los elementos específicos de cada una, en la tabla siguiente se muestra esta información y para su interpretación se cita el caso de box S en *Gm-Pp* de ATX que también se encuentra en *Zm* de ASH, chs- Unit1 m1 y GATT-motif de *Gm-Pp* de ATX que son encontrados en *Gm* de ASH. Por otro lado, en *Gm-Zm* de ASH se encuentran los motivos AE-box, CTAG-motif, P-box, Unnamed\_11 y también en *Gm-Pp* de ATX y W box de *Gm-Zm* de ATX también está presente en *Gm-Pp* de ASH; mientras que en *Gm-Pp* de ATX y en *Zm-Pp* de ASH es común el motivo TATC-box y por último WUN-motif se encuentra solo en *Gm-Pp* de ambos grupos.

La clasificación de la tabla (anexo 3) básicamente se realizó con la finalidad de detectar a los elementos más específicos de las especies que se concentran en la tabla 5.3, dado que genes homólogos pueden ser generados por la especiación, que produce pares de ortólogos (genes de diferentes especies que se derivan del mismo gen en la última especie ancestral común, y por lo tanto suelen tener funciones similares) (Frazer et al., 2003) en los cuales se esperaría encontrar por lo tanto elementos -cis también funcionalmente similares como se cito más arriba para el caso particular de 4cl-CMA2b y L-box.

Tabla 5.1. Elementos –cis específicos entre parejas de especies del grupo II (ASH).

Elemento –cis	<i>Glycine max</i>	<i>Zea mays</i>	Secuencias	Función
4cl-CMA2b	1	1	TCTCACCAACCCCA	*
	<i>Glycine max</i>	<i>Physcomitrella patens</i>		
	<i>Zea mays</i>	<i>Physcomitrella patens</i>		
Unnamed_8	1	1	TCCACGTAGA	*
Unnamed_9	1	1	TCCACGTAGA	*
Unnamed_10	1	1	TCCACGTAGA	*
Unnamed_12	1	1	TCCACGTAGA	*
Unnamed_14	1	1	TCCACGTAGA	*

Nota tabla 5.1 \*Función no reportada.

Tabla 5.2. Elementos –cis específicos entre parejas de especies del grupo III (ATX).

Elemento –cis	<i>Glycine max</i>	<i>Zea mays</i>	Secuencias	Función
box E	1	1	ACCCATCAAG	Elemento de inducción al desencadenamiento de hongos.
	<i>Zea mays</i>	<i>Physcomitrella patens</i>		
	<i>Glycine max</i>	<i>Physcomitrella patens</i>		
AT-rich element	3	2	ATAGAAATCAA	Sitio de union al AT-rich DNA binding protein (ATBP-1).
Box II	1	1	TGGTAATAA	Parte de un elemento de respuesta de luz.
ELI-box3	2	1	AAACCAATT	Elemento responsable de iniciación.

Tabla 5.3. Elementos específicos por especie.

Elemento - cis	Frecuencia	Secuencia	Función
<b>Gm ASH</b>			
MBSII	1	AAAAGTTAGTTA	Implicados en la regulación de genes para la biosíntesis de flavonoides. Potenciador de la función.
Pc-CMA2a	1	CAACCAATGAAAA	Parte de un elemento de respuesta de luz.
sbp-CMA1a	1	CTCTATCTTCTCCGG	Parte de un elemento de respuesta de luz.
<b>Gm ATX</b>			
3-AF3 binding site	3	CACTATCTAAC	
CE3	1	GACGCGTGTC	Elemento involucrado en la respuesta a ABA y VP1.
<b>Zm ASH</b>			
OBP-1 site	1	TACTACTTTTGG	Incrementa significativamente la habilidad de las proteínas OBF para unirse al elemento ocs.
C-box	1	CTGACGTCAG	Elemento de respuesta a luz.
CE1	1	TGCCACCGG	Elemento de acoplamiento con un ABRE (el más cercano) involucrado con la respuesta a ABA.
GCC box	1	AGCCGCC	Respuesta a heridas y a patógenos.
<b>Zm ATX</b>			
NON-box	1	AGATCGACG	Elemento relacionado con la activación específica del meristemo.
OCT	1	CGCGGATC	Elemento relacionado con la activación específica del meristemo.
Unnamed_17	1	TAGGAGCAGCT	

Tabla 5.3. (continuación) Elementos específicos por especie.

Elemento - cis	Frecuencia	Secuencia	Función
<b>Pp ASH</b>			
box II	1	TCCACGTGGC	Parte de un elemento de respuesta de luz.
E2F	1	AGTGGCGGNNNNNTTTGAA	Transición del ciclo celular
MSA-like	1	TCCAACGGT	Implicado en la regulación del ciclo celular
RY-element	1	CATGCATG	Implicado en la regulación específica de la semilla.
TGGCA-motif	1	GATGGAAGTGGCA	Parte de un elemento de respuesta de luz.
Unnamed_13	1	TCCAAGTATA	
<b>Pp ATX</b>			
ATCC-motif	1	CAATCCTC	Parte de un elemento de respuesta de luz.
CATTAT-motif	2	GTCATTATCGG	
HD-Zip 3	1	GTAAT(G/C)ATTAC	Sitio de unión a proteínas.
motif I	1	gGTACGTGGCG	Activador de la transcripción con la especificidad de las raíces podría estar implicado en la activación del desarrollo en las diferentes etapas.

Por otro lado, refiriéndonos a los elementos que son comunes a las tres especies se observa que ABRE (ACGT-containing ABA Response Element) (GACACGTACGT), es compartido con *Arabidopsis thaliana* (ABRE-like binding site motif) aunque con una frecuencia muy baja (uno en el promotor de ASHH3 y dos en el de ATX3) mientras que en Gm de ATX su frecuencia es de 13 y en Zm de ASH es de 18 (tablas 4.8a y 4.9a). Dado que este elemento –cis según su función está involucrado en la capacidad de respuesta al ácido abscísico (ABA, abscisic acid), existen otros dos elementos –cis relacionados con esta función y que son únicos de Gm en ATX: CE3 (Coupling Element 3) GACGCGTGTC y CE1 (Coupling Element 1) TGCCACCGG en Zm de ASH que no

están presentes en *Pp*. Nuevamente encontramos otro ejemplo de co-conservación y podría pensarse que ABRE es un elemento que se ha especializado funcionalmente con la evolución.

Otros elementos presentes en la cuatro especies son HSE e I-box en los dos grupos (ASH y ATX), E2F-binding site, TGA1 binding site motif y W-box promoter motif en ATX. TGA1 binding site motif (TGACGTGG) de *Arabidopsis thaliana* es similar a TGACG-motif (TGACG) encontrado con mayor frecuencia (18) en *Pp* de ATX y en *Zm* y *Pp* de ASH con frecuencias de 13 y 10 respectivamente, cuya función es la intervención en la respuesta a MeJA (binding site of Arabidopsis bZIP protein TGA1a) asociado con su complementario CGTCA-motif (CGTCA) con frecuencias prácticamente iguales que TGACG-motif.

De los elementos –cis MYB binding site motif (MYB, MYB1, MYB3 y MYB4) presentes en *Arabidopsis thaliana* solo MYB3 (TAACTAAC) tiene relación con el elemento MBS (MYB binding site) ((T/C)AACTG) encontrado en las tres especies en ambos grupos con las siguientes frecuencias: para el grupo ASH *Gm* 14, *Zm* 7, *Pp* 10 y para ATX *Gm* 18, *Zm* 8, *Pp* 7; implicados en la inducibilidad por sequía.

GCC-box (AGCCGCC) y RY-element (CATGCATG) de *Arabidopsis thaliana* solo se encontraron en el grupo ASH en *Zm* y *Pp* como elementos específicos respectivamente.

De los elementos -cis ATHB binding site motif (ATHB1, ATHB2 y ATHB5) presentes en *Arabidopsis thaliana*, ATHB5 binding site motif CAATAATTG es muy similar en su secuencia con HD-Zip 3 (Homeodomain-leucine zipper 3) GTAAT(G/C)ATTAC que es un sitio de unión para proteína, encontrado como elemento único en *Pp* de ATX.

Por último, de los elementos –cis encontrados con mayor frecuencia en *Arabidopsis thaliana* solo BoxII promoter motif (GGTTAA), GATA promoter motif [LRE] y W-box promoter motif son encontrados en los promotores de las especies estudiadas. BoxII es similar a GT1-motif (GGTTAA) encontrado en las tres especies, GATA se localiza en

todos excepto en *Pp* de ASH. y *W*-box solo en los promotores de *Gm-Pp* de ASH y en *Gm-Zm* de ATX.

Debe notarse nuevamente que los elementos –cis reportados (plantCARE) corresponden a las primeras 1500 pb a partir de sitio de inicio de la transcripción y que fueron considerados en ambos sentidos esta es una razón por la cual la frecuencia de algunos de ellos incluyendo el número de promotores analizados en conjunto fue elevada, sobre todo en aquellos que están presentes en todos los promotores como es el caso de CAAT-box y TATA-box, por ejemplo (ver anexo 4). Por otro lado, los elementos –cis encontrados en los promotores en estudio fueron más sobrerrepresentados que aquellos en los promotores de *Arabidopsis thaliana*, que utilizamos como modelo de fondo de consulta o de comparación.

## 6. CONCLUSIONES

Esta tesis presenta otra opción para la búsqueda y determinación de los elementos reguladores –cis presentes en los promotores de genes de plantas superiores, tiene importancia porque tales elementos están implicados en el proceso de la transcripción en organismos eucariotas y es una herramienta de apoyo a la investigación que combina la Biología, la informática y la computación. En nuestro análisis detectamos la presencia de los elementos –cis en los promotores de tres especies de plantas utilizando un set de promotores de la planta modelo *Arabidopsis thaliana* (pequeña planta con flores que se utiliza ampliamente como un organismo modelo en biología vegetal). *Arabidopsis* es un miembro de la familia de la mostaza (Brassicaceae), que incluye las especies cultivadas como la col y el rábano. *Arabidopsis* no es de mayor significancia agronómica, pero ofrece ventajas importantes para la investigación básica en Genética y Biología Molecular.

Nuestra metodología ha demostrado que no solo se puede utilizar en plantas sino que en cualquier otra especie (animales por ejemplo) para la determinación de los elementos –cis presentes en los promotores de tales genes utilizando las bases de datos adecuadas para tal fin que actualmente se encuentran públicamente disponibles. Además no requiere de entrenamiento por lo cual puede ser ampliamente aplicado y presenta la opción de encontrar nuevos elementos reguladores compartidos entre especies que presenten una distancia evolutiva relativamente cercana.

**Conclusión referida a nuestro objetivo específico 1.** En cuanto a la determinación de los elementos –cis presentes en los promotores de genes de *Arabidopsis thaliana* de la familia Trithorax (clase II-ASH y clase III-ATX) se encontró que la mayoría de ellos están presentes en ambos grupos aunque con frecuencias relativamente diferentes, si bien unos pocos no están presentes en un grupo si lo están en el otro, por lo tanto, ambos grupos se complementan en cuanto a la presencia de los elementos –cis en estos promotores de genes.

De acuerdo al marco teórico de esta investigación, los elementos –cis presentes en el promotor de un gen se encuentran más o menos conservados y que los cambios en las

secuencias de estos pueden dar lugar a nuevas funciones lo que implica que tal variación da origen a genes parálogos por lo cual se concluye que las diferencias básicas entre los promotores de genes de *Arabidopsis thaliana* tiene que ver con la presencia o la ausencia de uno o un conjunto de elementos –cis en cada uno de ellos con relación a los otros en cuanto a su función.

**Conclusión referida a nuestro objetivo específico 2.** De acuerdo a la cuantificación de los elementos -cis encontrados en los promotores de las especies estudiadas, la mayoría (con frecuencias relativamente altas) son comunes o conservados en ellas lo cual es concordante con que genes de diferentes especies (genes ortólogos) tienen funciones similares y por lo tanto elementos –cis también funcionalmente similares. De la misma manera aquellos elementos -cis que son diferentes entre las especies sugieren que cada una de ellas se ha diversificado en la especiación.

La comparación entre los elementos –cis presentes en los promotores de genes de estas especies han evidenciado casos de co-conservación entre ellas. En el caso de los elementos específicos (con frecuencias relativamente bajas) para cada uno de ellos la mayoría forman parte de módulos de ADN conservados que participan en la respuesta a estímulos externos u otra función biológica específica, de aquí que los elementos –cis funcionalmente participan en conjunto más que de manera individual.

**Conclusión general.** La determinación de los elementos-cis presentes en la región genómica de factores epigenéticos pertenecientes al grupo Trithorax en las plantas estudiadas fueron más sobrerrepresentados que aquellos en los promotores de *Arabidopsis thaliana*. Lo anterior y otros aspectos como se señaló en la discusión de los resultados son consecuencia de la información obtenida aplicando la metodología *in silico* para la identificación de bio-secuencias, ello puede significar que puede aplicarse a más tres especies en estudios comparativos y particularmente en especies evolutivamente relacionadas (secuencias ortólogas).

## 6.1 SUGERENCIAS

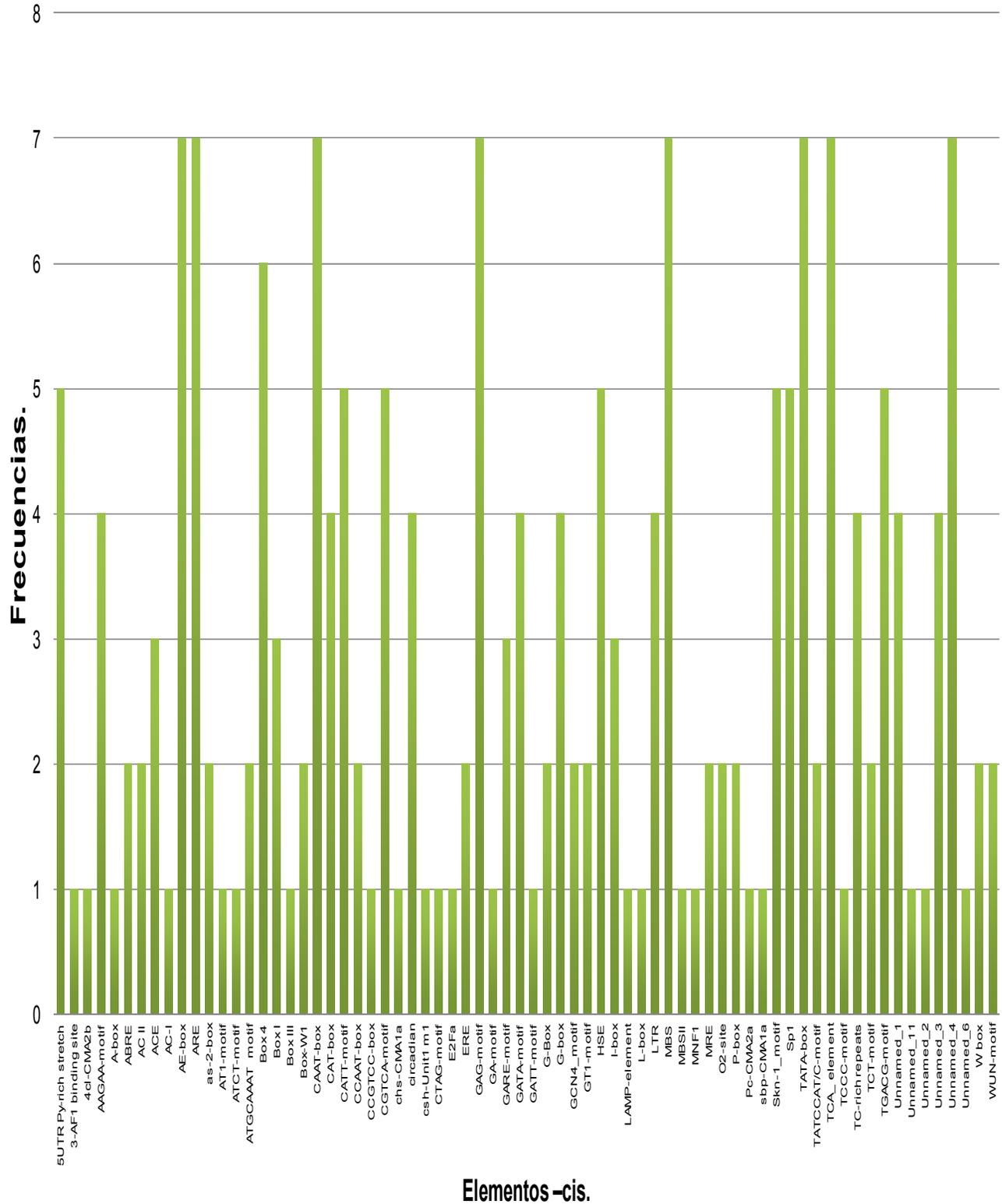
Los elementos –cis aquí reportados no son infalibles pero reflejan la información actualmente reportada por las bases de datos consultadas, sin embargo, el objetivo principal que básicamente fue un análisis cualitativo, nos llevo a los resultados que ya hemos presentado y discutido. Un análisis cuantitativo se podrá realizar como consecuencia de este u otro estudio que se quisiera emprender con este método, incluso la verificación experimental de tales elementos –cis encontrados podría llevarse a cabo. Se podría considerar delimitar la investigación a elementos –cis específicos, a grupos de ellos o a la cantidad de nucleótidos presentes en tales elementos utilizando filtros o programas (algoritmos) adecuados para tal fin, incluso se podría modificar la longitud de los promotores o enfocar el estudio a una región específica de este o aun más modificar los parámetros utilizados para determinar la similitud entre ellos.

No se debe olvidar que un análisis *in silico* es una herramienta poderosa en la Biología actual, que por encima de otros ahorra gastos y tiempo, sobre todo en estudios previos a una investigación más profunda de sus resultados o en las búsquedas de cualquier posible candidato (biosecuencia) de interés a ser estudiado con fines agroeconómicos o en el campo de la salud donde también es aplicado.

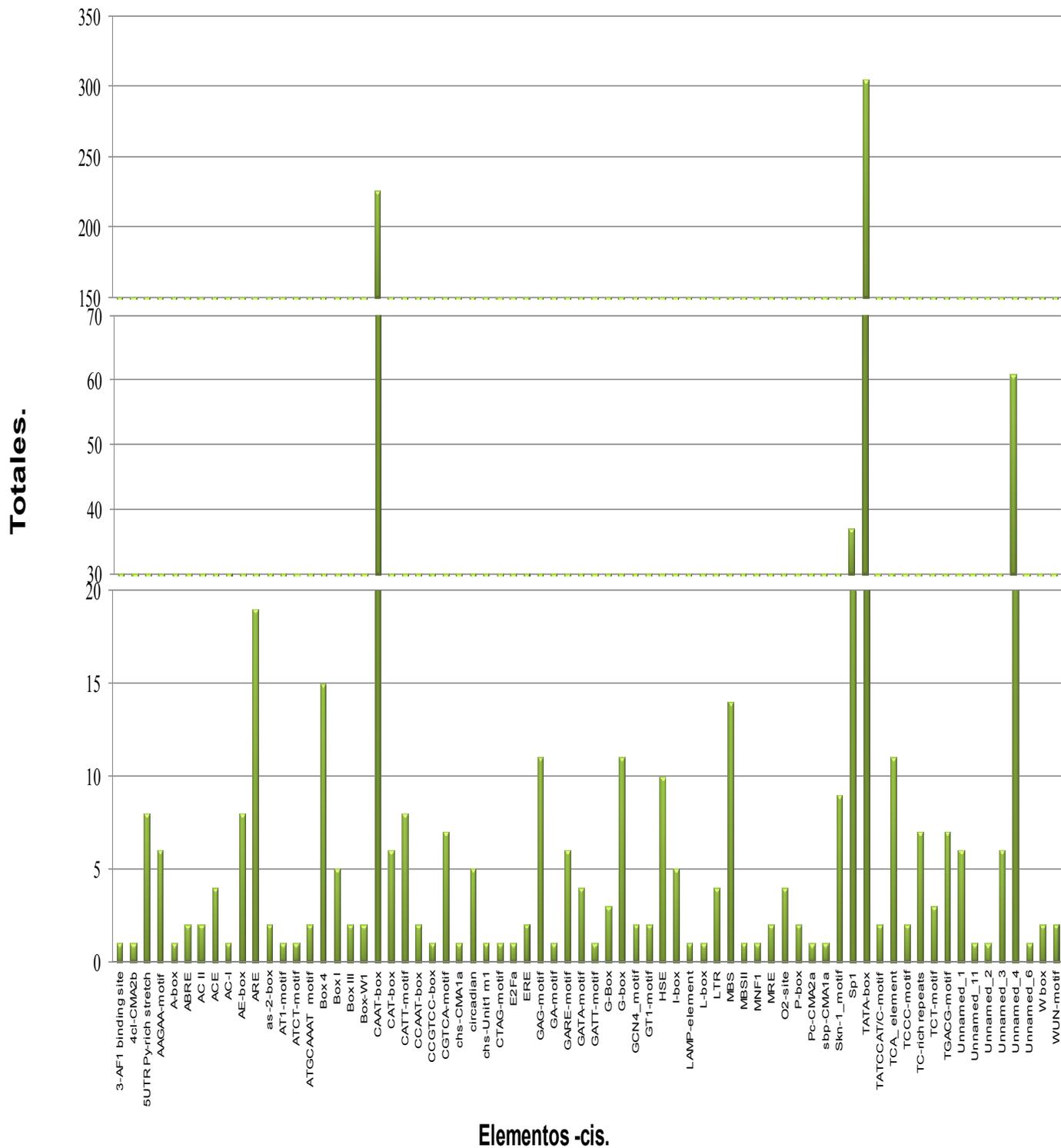
## 7. ANEXOS

Anexo 1. Gráficas que muestran la frecuencia y la presencia de los elementos –cis en los promotores de genes de *Arabidopsis thaliana*.

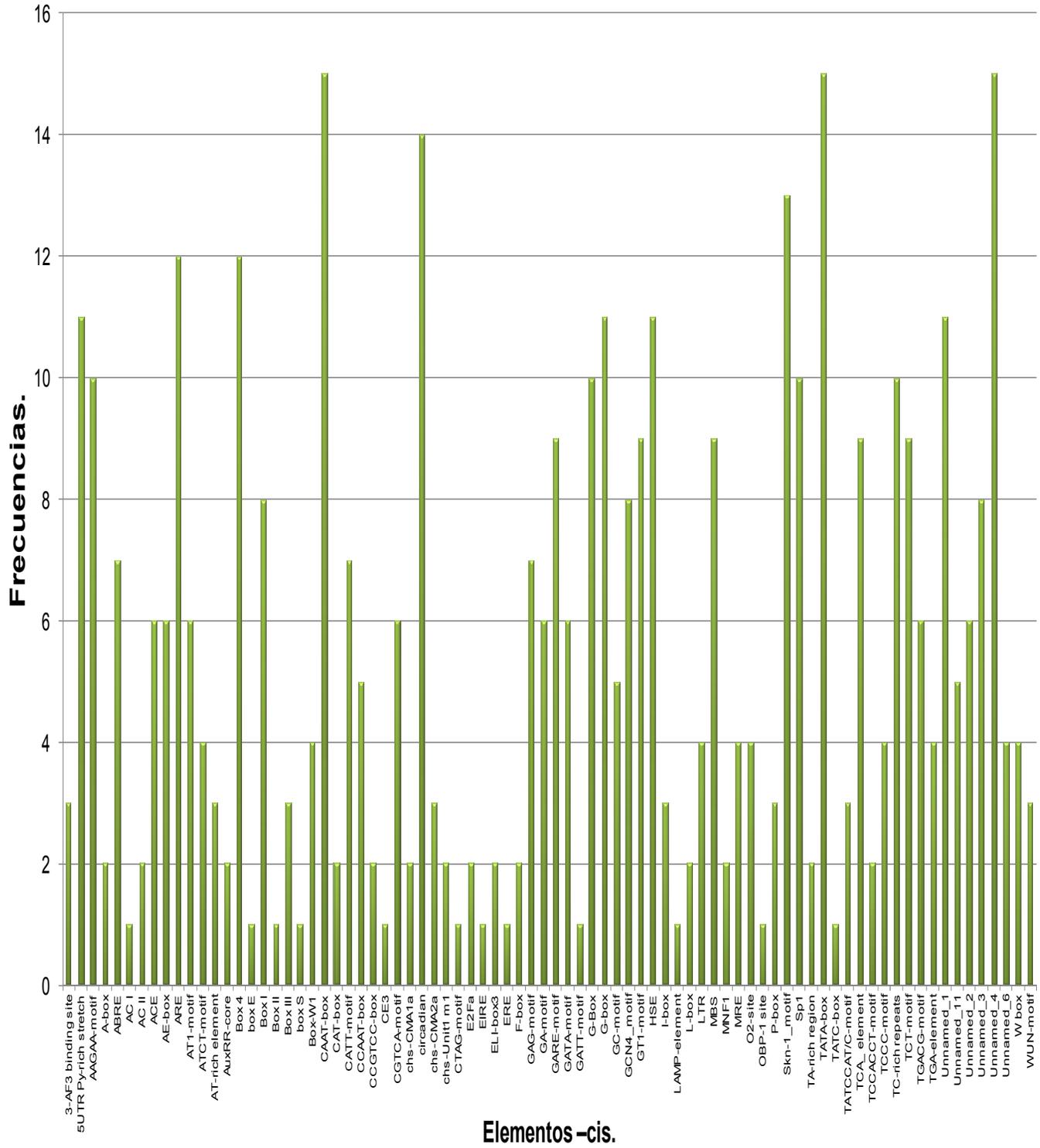
## Frecuencia de elementos -cis en los promotores de *Glycine max* de la clase II (ASH).



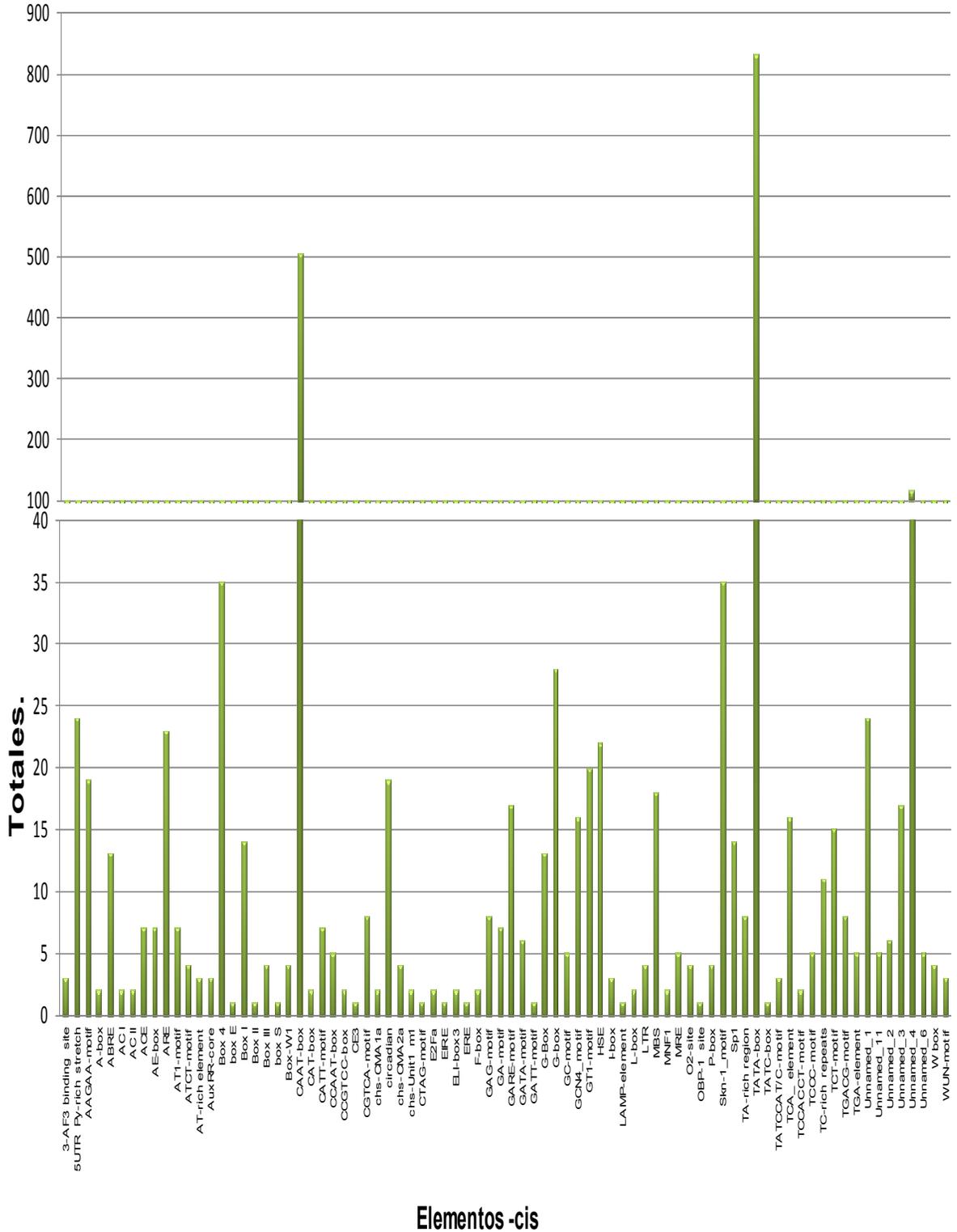
## Presencia de elementos *-cis* en los promotores de *Glycine max* del grupo II (ASH).



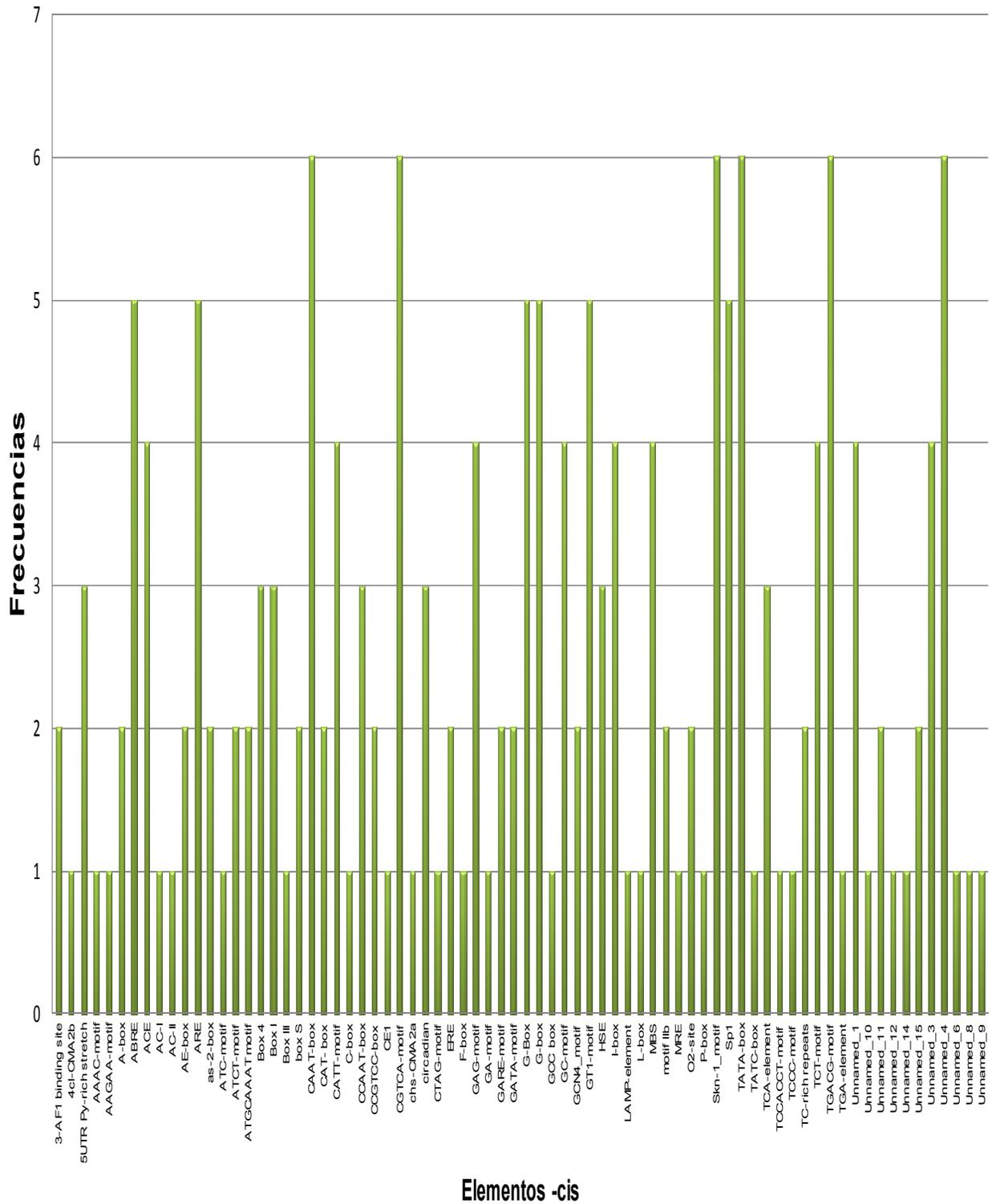
## Frecuencia de elementos -cis en los promotores de *Glycine max* del grupo III (ATX).



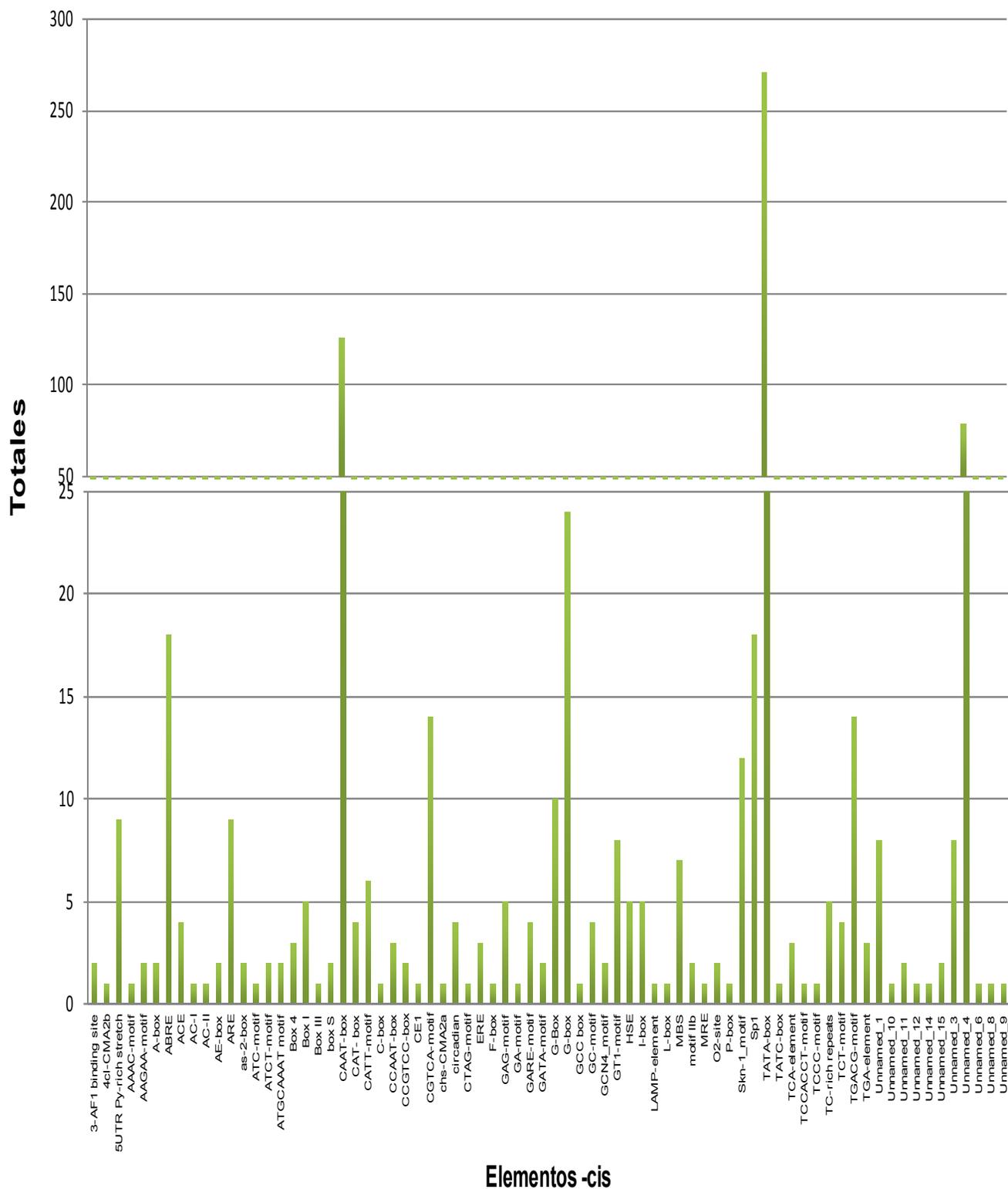
## Presencia de elementos -cis en los promotores de *Glycine max* del grupo III (ATX).



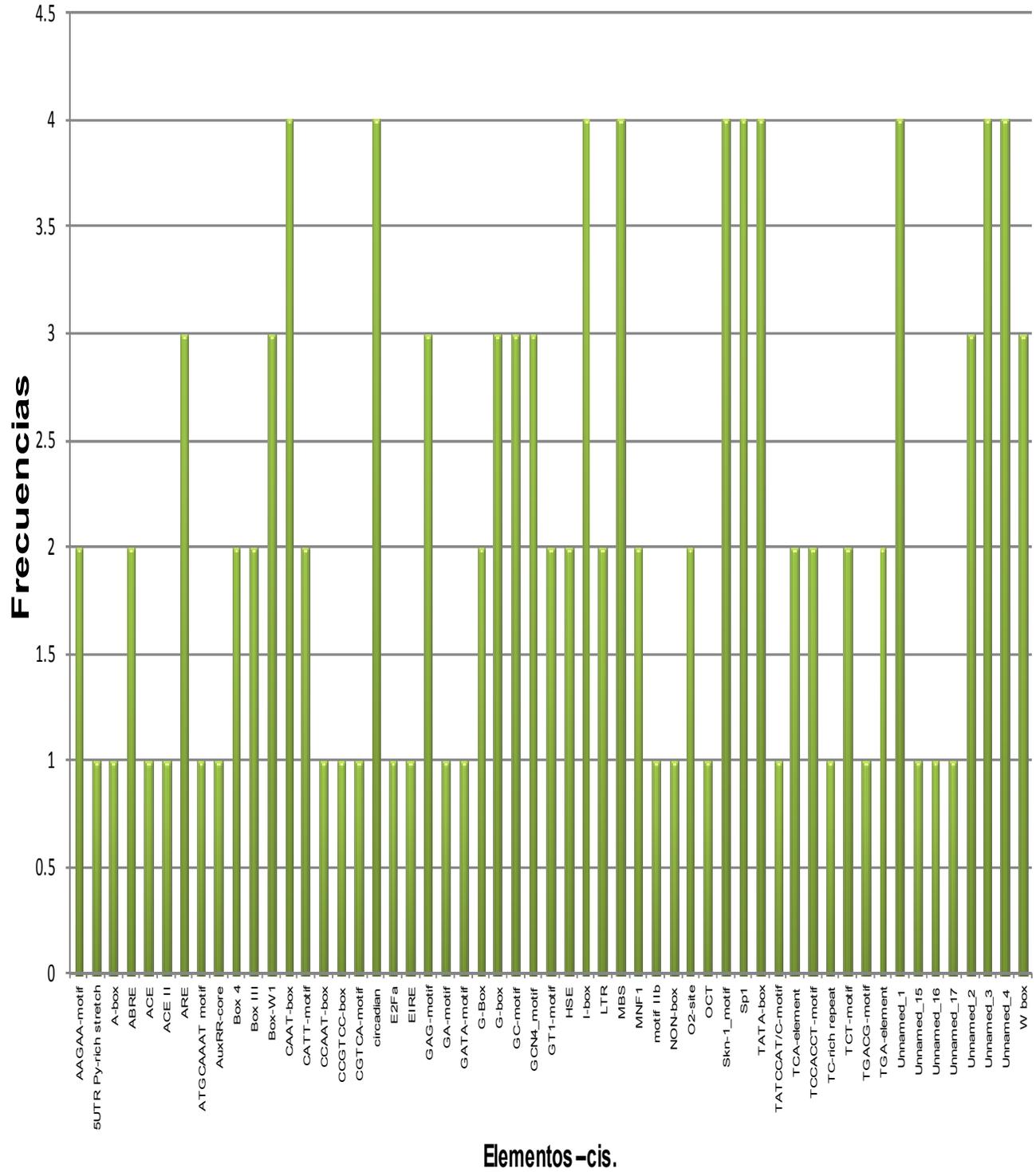
## Frecuencia de elementos -cis en los promotores de *Zea mays* del grupo II (ASH)



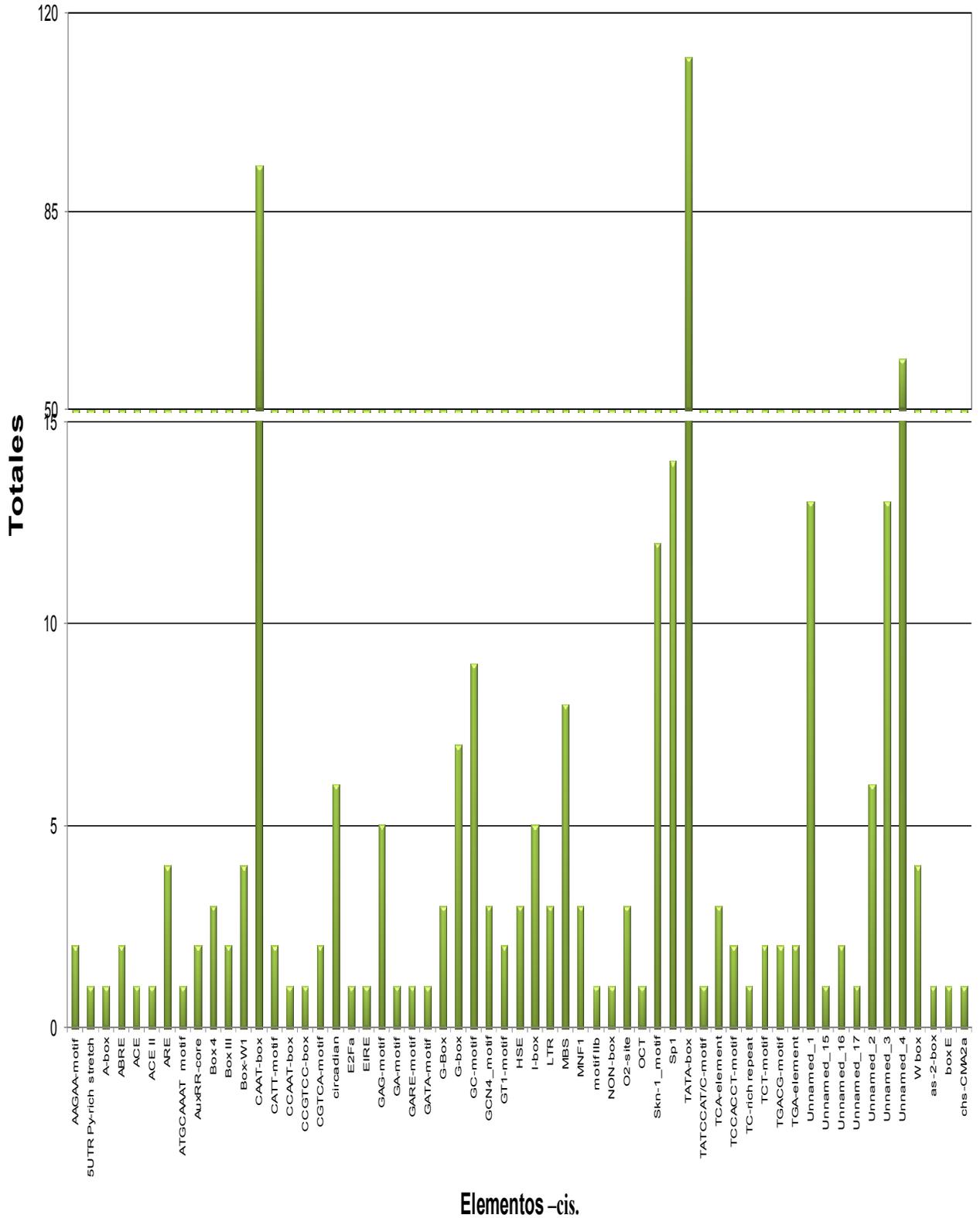
## Presencia de elementos -cis en los promotores de Zea mays del grupo II (ASH).



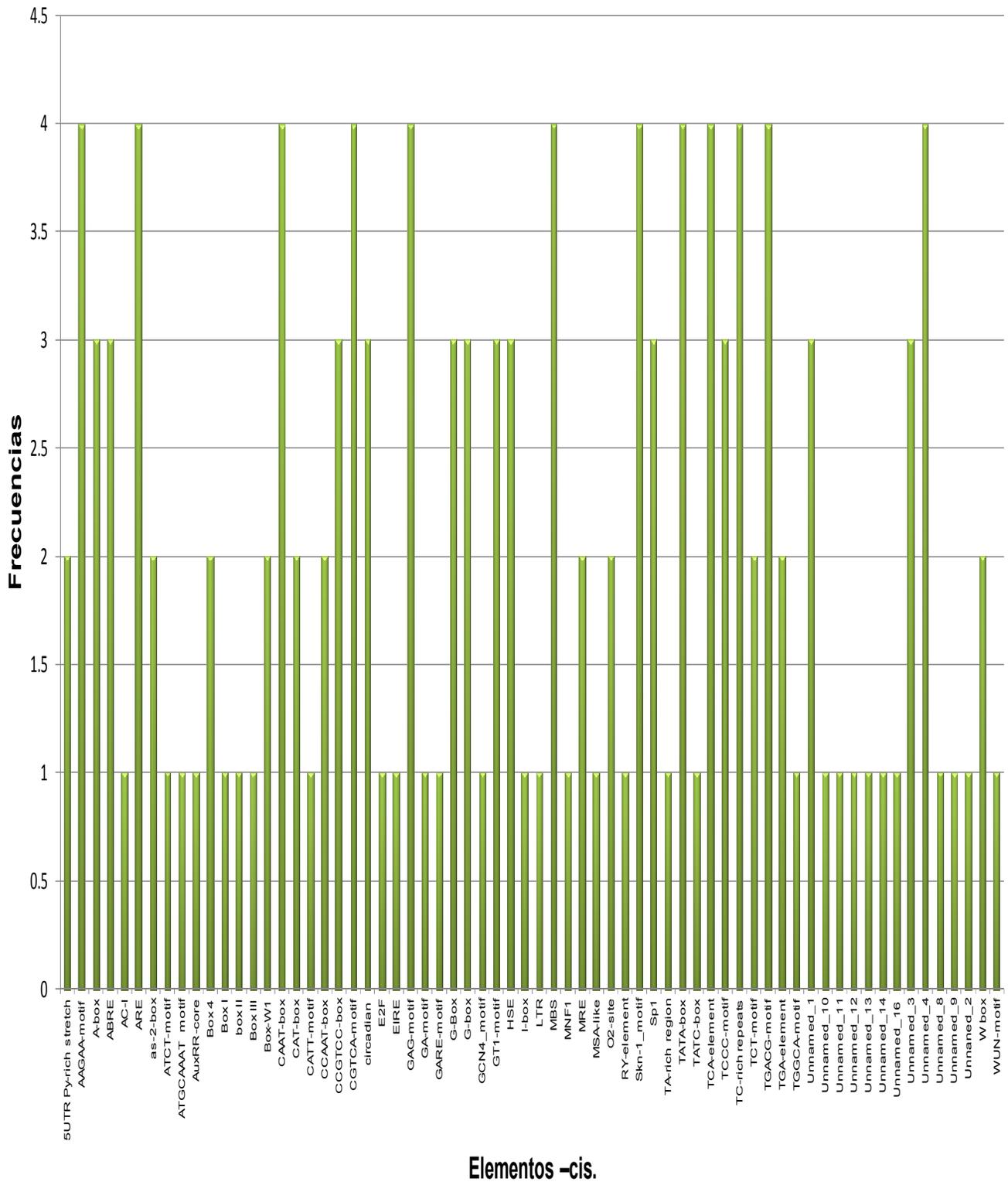
## Frecuencia de elementos -cis en los promotores de *Zea mays* del grupo III (ATX)



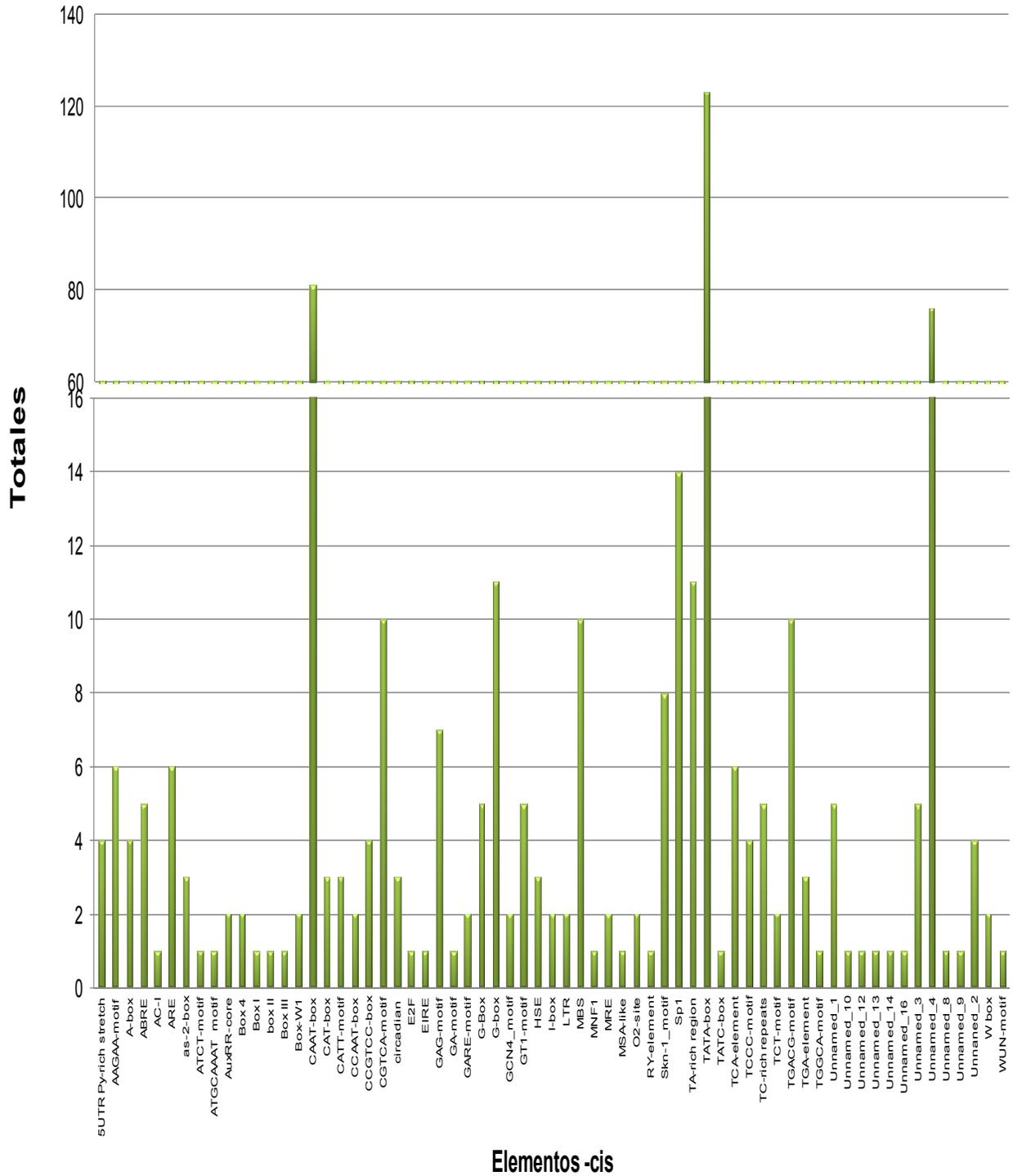
## Presencia de elementos -cis en los promotores de *Zea mays* del grupo III (ATX).



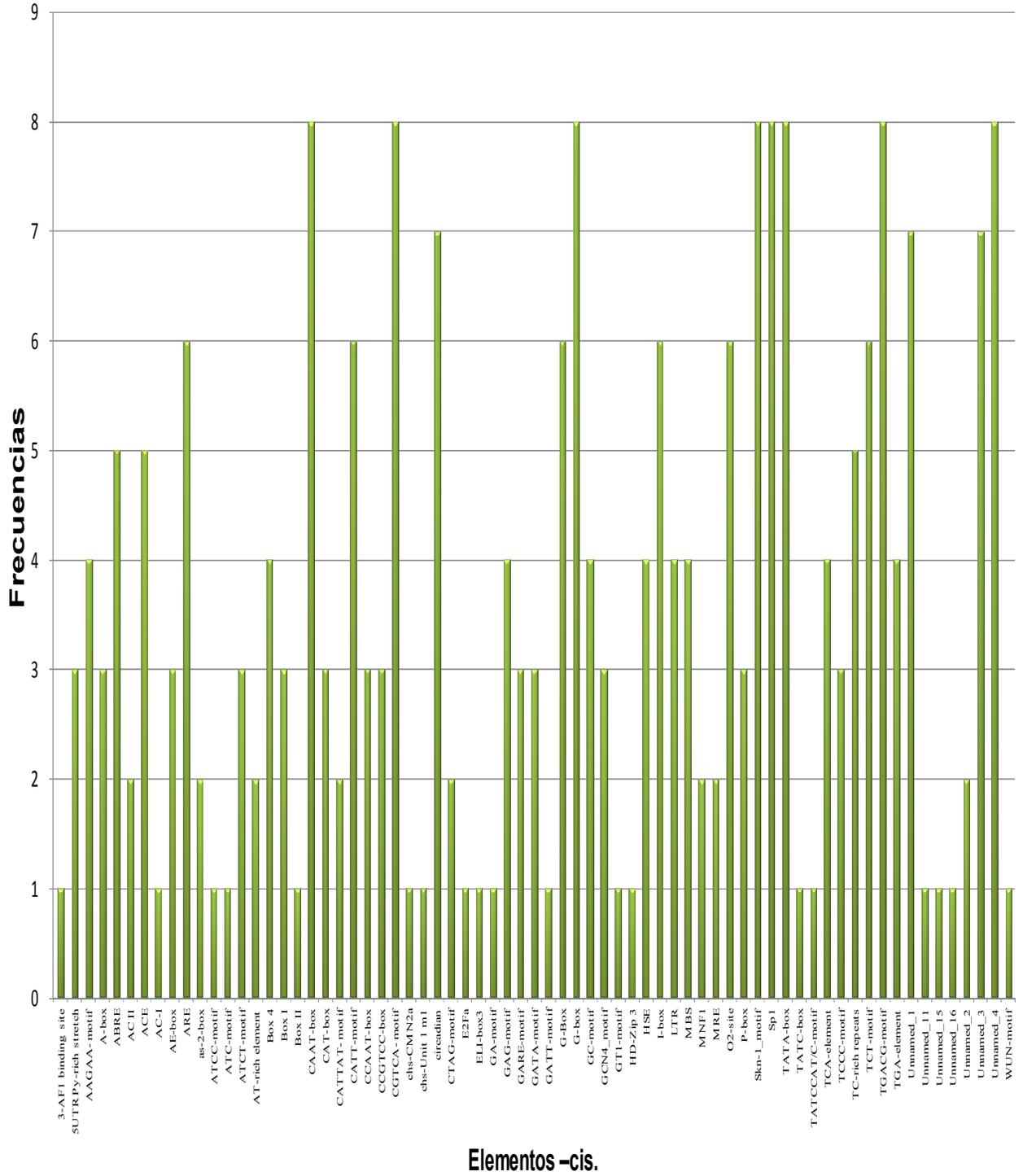
## Frecuencia de elementos -cis en los promotores de *Physcomitrella patens* del grupo II (ASH)



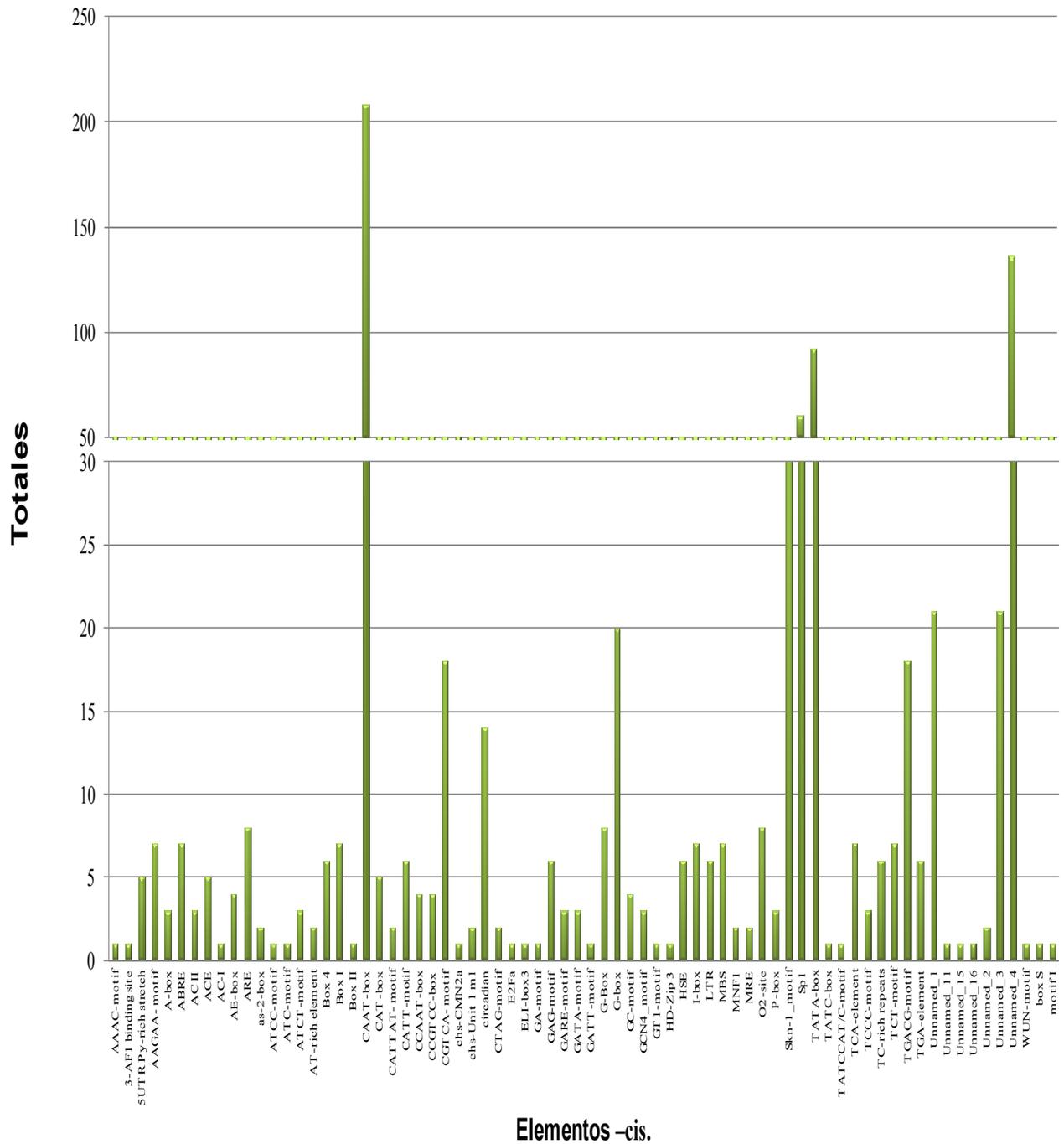
## Presencia de elementos -cis en los promotores de *Physcomitrella patens* del grupo II (ASH).



## Frecuencia de elementos -cis en los promotores de *Physcomitrella patens* del grupo III (ATX).



## Presencia de elementos -cis en los promotores de *Physcomitrella patens* del grupo III (ATX).



Anexo 2. Resumen de secuencias con sus respectivos parámetros estadísticos utilizados para el criterio de selección (HSPs). NOTA. En fondo gris se muestra las secuencias que fueron seleccionadas de entre las demás por su mayor homología.

Promotor	Orden	Secuencias	HSP#	S (score bits)	E (E-value)	Identidad (%)	Similitud (%)
ATX1	1o	Glyma08g29010.1	1	1120.9 (2898.0)	0	56.2 (611/1087)	69.7 (758/1087)
	2o	Glyma18g51890.1	1	1104.7 (2856.0)	0	55.6 (603/1085)	68.9 (748/1085)
	1o	GRMZM2G013794_T02	1	936.8 (2420.0)	0	54.7 (464/848)	69.7 (591/848)
	1o	Pp1s282_23V6.1	1	784.3 (2024.0)	0	45.7 (417/912)	59.8 (545/912)
	2o	Pp1s93_144V6.1	1	775.8 (2002.0)	0	44.9 (411/916)	59.2 (542/916)
ATX2	1o	Glyma08g29010.1	1	1147.9 (2968.0)	0	61.3 (590/963)	74.6 (718/963)
	2o	Glyma18g51890.1	1	1142.1 (2953.0)	0	61.3 (590/962)	74.1 (713/962)
	1o	GRMZM2G013794_T02	1	948.7 (2451.0)	0	53.9 (486/902)	68.6 (619/902)
	1o	Pp1s282_23V6.1	1	775.0 (2000.0)	0	44.8 (415/926)	59.2 (548/926)
	2o	Pp1s93_144V6.1	1	765.0 (1974.0)	0	44.0 (412/936)	59.2 (554/936)
ATX3	1°	Glyma07g06190.1	1	1102.4 (2850)	0	57.7 (574/995)	70.9 (705/995)
	2°	Glyma04g41500.1	1	992.3 (2564)	0	53.0 (501/946)	69.3 (656/946)
	3°	Glyma06g13330.1	1	997.6 (2526)	0	55.6 (478/859)	71.5 (614/859)
	4°	Glyma03g37370.1	1	889.8 (2298.0)	0	51.2 (500/977)	64.2 (627/977)
	5°	Glyma19g39970.1	1	883.2 (2281.0)	0	55.4 (456/823)	68.3 (562/823)
	6°	Glyma16g02800.1	1	865.5 (2235.0)	0	62.5 (429/686)	75.4 (517/686)
			2	190.3 (482.0)	7.7e-48	72.3 (81/112)	89.3 (100/112)
	7°	Glyma19g17460.1	1	670.2 (1728.0)	0	63.4 (322/508)	78.3 (398/508)
	1o	GRMZM2G085266_T01	1	837.8 bits (2163.0)	0	51.7 (431/833)	67.6 (563/833)
	1o	Pp1s223_92V6.1	1	718.0 (1852.0)	0	43.2 (423/980)	56.5 (554/980)
2o	Pp1s47_35V6.1	1	692.6 (1786.0)	0	41.6 (432/1039)	54.9 (570/1039)	

Anexo 2. (continuación) Resumen de secuencias con sus respectivos parámetros estadísticos utilizados para el criterio de selección (HSPs). NOTA. En fondo gris se muestra las secuencias que fueron seleccionadas de entre las demás por su mayor homología.

Promotor	Orden	Secuencias	HSP#	S (score bits)	E (E-value)	Identidad (%)	Similitud (%)
ATX4	1o	Glyma04g41500.1	1	1399.8 (3622.0)	0	66.6 (704/1057)	76.5 (809/1057)
	2o	Glyma06g13330.1	1	1369.0 (3542.0)	0	63.5 (702/1105)	73.8 (816/1105)
	3o	Glyma07g06190.1	1	1039.3 (2686.0)	0	52.5 (544/1037)	65.2 (676/1037)
	4o	Glyma16g02800.1	1	941.4 (2432.0)	0	52.9 (481/910)	66.0 (601/910)
	5o	Glyma03g37370.1	1	856.3 (2211.0)	0	44.3 (481/1085)	58.7 (637/1085)
	6o	Glyma19g39970.1	1	827.0 (2135.0)	0	49.4 (415/840)	65.1 (547/840)
	7o	Glyma19g17460.1	1	827.0 (2135.0)	0	49.4 (415/840)	65.1 (547/840)
	1o	GRMZM2G085266_T01	1	892.5 (2305.0)	0	53.9 (449/833)	69.9 (582/833)
	1o	Pp1s223_92V6.1	1	737.3 (1902.0)	0	41.4 (413/997)	58.6 (584/997)
	2o	Pp1s47_35V6.1	1	717.6 (1851.0)	0	40.2 (421/1046)	56.6 (592/1046)
ATX5	1o	Glyma04g41500.1	1	1414.1 (3659.0)	0	65.7 (702/1069)	77.9 (833/1069)
	2o	Glyma06g13330.1	1	1392.9 (3604.0)	0	63.4 (702/1107)	75.8 (839/1107)
	3o	Glyma07g06190.1	1	1048.1 (2709.0)	0	54.9 (526/958)	69.3 (664/958)
	4o	Glyma16g02800.1	1	951.0 (2457.0)	0	52.0 (484/930)	65.7 (611/930)
	5o	Glyma03g37370.1	1	859.0 bits (2218.0)	0	43.8 (484/1106)	58.0 (642/1106)
	6o	Glyma19g39970.1	1	838.6 (2165.0)	0	49.3 (414/840)	65.0 (546/840)
	7o	Glyma19g17460.1	1	818.1 (2112.0)	0	76.1 (386/507)	87.6 (444/507)
	1o	GRMZM2G085266_T01	1	875.9 (2262.0)	0	52.5 (439/836)	68.4 (572/836)
	1o	Pp1s223_92V6.1	1	741.9 (1914.0)	0	42.3 (421/996)	57.5 (573/996)
	2o	Pp1s47_35V6.	1	736.1 (1899.0)	0	41.1 (420/1023)	57.4 (587/1023)

Anexo 2. (continuación) Resumen de secuencias con sus respectivos parámetros estadísticos utilizados para el criterio de selección (HSPs). NOTA. En fondo gris se muestra las secuencias que fueron seleccionadas de entre las demás por su mayor homología.

Promotor	Orden	Secuencias	HSP#	S (score bits)	E (E-value)	Identidad (%)	Similitud (%)
ATXR3	1o	Glyma12g11060.1	1	2358.2 (6110.0)	0	58.6 (1340/2288)	71.9 (1645/2288)
	2o	Glyma06g45740.1	1	2335.1 (6050.0)	0	58.3 (1320/2263)	72.1 (1631/2263)
	3o	Glyma13g38090.1	1	2316.2 (6001.0)	0	57.1 (1331/2331)	71.3 (1661/2331)
			2	43.5 (101.0)	2.83e-3	69.0 (20/29)	75.9 (22/29)
	4o	Glyma12g32290.1	1	2261.5 (5859.0)	0	54.4 (1349/2480)	68.0 (1687/2480)
	1o	GRMZM2G149153_T01	1	1605.9 (4157.0)	0	48.3 (896/1855)	63.7 (1182/1855)
	1o	Pp1s123_132V6.1	1	750.7 (1937.0)	0	52.2 (388/744)	70.2 (522/744)
			2	98.6 (244.0)	4.8e-20	32.1 (69/215)	46.0 (99/215)
			3	82.8 (203.0)	2.7e-15	30.9 (59/191)	45.5 (87/191)
	2o	Pp1s311_17V6.1	1	749.2 (1933.0)	0	51.0 (398/781)	68.2 (533/781)
2			134.8 (338.0)	5e-31	24.4 (151/620)	40.6 (252/620)	
ATXR7	1o	Glyma09g05740.1	1	342.8 (878.0)	1.3e-93	63.2 (194/307)	75.2 (231/307)
			2	129.0 (323.0)	3.1e-29	54.1 (59/109)	66.1 (72/109)
			3	79.3 (194.0)	2.6e-14	32.8 (77/235)	44.7 (105/235)
	2o	Glyma15g17030.1	1	340.9 (873.0)	4.8e-93	65.2 (193/296)	75.3 (223/296)
			2	300.4 (768.0)	8.1e-81	32.0 (252/788)	42.4 (334/788)
	1o	GRMZM2G473138_T01	1	295.0 (754.0)	2.3e-79	70.3 (154/219)	83.6% (183/219)
			2	66.6 (161.0)	1.7e-10	34.6 (53/153)	50.3 (77/153)
	1o	Pp1s133_30V6.1	1	264.2 (674.0)	3.6e-70	60.1% (137/228)	74.1% (169/228)
	2o	Pp1s101_4V6.1	1	244.6 (623.0)	2.7e-64	52.8 (122/231)	71.4 (165/231)
			2	49.7 (117.0)	1.4e-5	20.8 (62/298)	36.6 (109/298)

Anexo 2. (continuación) Resumen de secuencias con sus respectivos parámetros estadísticos utilizados para el criterio de selección (HSPs). NOTA. En fondo gris se muestra las secuencias que fueron seleccionadas de entre las demás por su mayor homología.

Promotor	Orden	Secuencias	HSP#	S (score bits)	E (E-value)	Identidad (%)	Similitud (%)
ASHH1	1o	Glyma10g36720.1	1	599.7 (1545.0)	1.7e-171	61.7 (300/486)	74.1 (360/486)
	2o	Glyma20g30870.1	1	595.9 (1535.0)	2.5e-170	62.3 (303/486)	74.5 (362/486)
	1o	GRMZM2G147619_T03	1	549.3 (1414.0)	2.7e-156	55.6 (279/502)	70.3 (353/502)
	1o	Pp1s28_356V6.1	1	208.4 (529.0)	6.3e-54	43.7 (107/245)	57.1 (140/245)
	2o	Pp1s183_22V6.1	1	200.7 (509.0)	1.6e-51	45.1 (97/215)	59.5 (128/215)
ASHH2	1o	Glyma04g42410.1	1	602.4 (1552.0)	1.2e-171	40.2 (417/1037)	54.4 (564/1037)
			2	104.4 (259.0)	1.1e-21	37.0 (104/281)	50.5 (142/281)
	2o	Glyma06g12390.1	1	580.1 (1494.0)	6.3e-165	48.3 (333/689)	61.2 (422/689)
			2	106.7 (265.0)	2.0e-22	37.4 (70/187)	49.7 (93/187)
			3	44.3 (103.0)	1.26e-3	31.4 (33/105)	48.6 (51/105)
	1o	GRMZM2G352431_T02	1	428.7 (1101.0)	1.8e-119	59.2 (200/338)	72.8 (246/338)
			2	140.6 (353.0)	9.2e-33	30.1 (129/429)	41.3 (177/429)
	1o	Pp1s28_356V6.1	1	401.0 (1029.0)	2.8e-111	45.5 (225/495)	59.4 (294/495)
			2	99.4 (246.0)	2e-20	31.5 (76/241)	43.6 (105/241)
	2o	Pp1s183_22V6.1	1	396.4 (1017.0)	6.7e-110	51.3 (196/382)	67.0 (256/382)
			2	77.8 (190.0)	5.6e-14	58.1 (43/74)	74.3 (55/74)
	ASHH3	1o	Glyma16g33220.1	1	404.8 (1039.0)	5.8e-113	65.4 (238/364)
2o		Glyma09g28430.1	1	332.4 (851.0)	3.3e-91	69.4 (179/258)	81.0 (209/258)
			2	87.0 (214.0)	2.8e-17	60.3 (41/68)	75.0 (51/68)
1o		GRMZM2G033694_T03	1	233.8 (595.0)	1.6e-61	45.7 (170/372)	59.4 (221/372)
2o		GRMZM2G130910_T01	1	223.0 (567.0)	2.6e-58	46.2 (165/357)	60.2 (215/357)
1o	Pp1s102_150V6.1	1	216.1 (549.0)	2.1e-56	49.0 (147/300)	64.7 (194/300)	

Anexo 2. (continuación) Resumen de secuencias con sus respectivos parámetros estadísticos utilizados para el criterio de selección (HSPs). NOTA. En fondo gris se muestra las secuencias que fueron seleccionadas de entre las demás por su mayor homología.

Promotor	Orden	Secuencias	HSP#	S (score bits)	E (E-value)	Identidad (%)	Similitud (%)
ASHH4	1o	Glyma16g33220.1	1	338.6 (867.0)	5.5e-93	55.1 (200/363)	66.9 (243/363)
	2o	Glyma09g28430.1	1	271.2 (692.0)	1.0e-72	56.8 (147/259)	67.6 (175/259)
			2	53.5 (127.0)	3.3e-7	43.0 (34/79)	58.2 (46/79)
	1o	GRMZM2G033694_T03	1	249.6 (636.0)	2.5e-66	60.4 (128/212)	76.4 (162/212)
	2o	GRMZM2G130910_T01	1	246.9 (629.0)	1.9e-65	60.6 (126/208)	75.5 (157/208)
	1o	Pp1s102_150V6.1	1	213.4 (542.0)	1.4e-55	42.3 (129/305)	56.7 (173/305)
	2o	Pp1s76_183V6.2	1	179.5 (454.0)	2.0e-45	49.5 (96/194)	66.5 (129/194)
	3o	Pp1s28_356V6.1	1	152.9 (385.0)	2.1e-37	40.6 (89/219)	57.5 (126/219)
	4o	Pp1s183_22V6.1	1	150.2 (378.0)	1.4e-36	40.6 (89/219)	57.5 (126/219)
ASHR3	1o	Glyma06g29960.1	1	428.3 (1100.0)	8.2e-120	55.8 (235/421)	67.2 (283/421)
	1o	GRMZM2G451374_T01	1	238.4 (607.0)	8.7e-63	50.8 (128/252)	63.9 (161/252)
	2o	GRMZM2G067019_T01	1	224.9 (572.0)	9.8e-59	71.2 (99/139)	83.5 (116/139)
			2	74.3 (181.0)	2.3e-13	32.5 (40/123)	46.3 (57/123)
	3o	GRMZM2G130910_T01	1	204.1 (518.0)	1.9e-52	47.9 (102/213)	62.9 (134/213)
	4o	GRMZM2G033694_T03	1	203.4 (516.0)	3.2e-52	48.4 (105/217)	63.6 (138/217)
	1o	Pp1s76_183V6.2	1	265.8 (678.0)	3.3e-71	54.1 (131/242)	70.7 (171/242)
	2o	Pp1s102_150V6.1	1	197.2 (500.0)	1.6e-50	53.8 (93/173)	71.1 (123/173)
			2	55.1 (131.0)	1.1e-7	26.2 (49/187)	41.7 (78/187)
	3o	Pp1s28_356V6.1	1	137.1 (344.0)	2.2e-32	34.5 (76/220)	51.4 (113/220)
4o	Pp1s183_22V6.1	1	136.3 (342.0)	3.9e-32	34.5 (76/220)	51.8 (114/220)	

Anexo 3. Tabla que concentra los elementos –cis encontrados por especie y por parejas. Los elementos –cis en **negritas** son comunes entre parejas o entre especies en sus respectivas columnas y en *cursiva* son comunes entre parejas y especies. Los elementos en **CASTELLAR** corresponden a los elementos de las tablas 8 y 9 y en **STENCIL** los elementos de la tabla 10. En **Impact** se muestran los elementos únicos y de parejas que son comunes en las tres especies pero del otro grupo.

COLUMNA 1 (C-1)		COLUMNA 2 (C-2)		ASH		ATX	
<i>Gm</i> ASH		<i>Gm-Zm</i> ASH		<i>Gm/Zm/Pp</i>		<i>Gm/Zm/Pp</i>	
<i>Elemento -cis</i>		<i>Elemento -cis</i>	<i>Gm/Zm</i>	C-1	C-2	C-1	C-2
<b>AT1-motif</b>	1	<i>3-AF1 binding site</i>	1/2				
<b>chs-CMA1a</b>	1	<i>4CL-CMA2B</i>	1/1				
<i>chs-Unit1 m1</i>	1	<b>AC II</b>	<b>2/1</b>				<b>2/1/3</b>
<b>E2Fa</b>	<b>1</b>	<b>ACE</b>	<b>4/4</b>			<b>2/1/1</b>	<b>7/1/5</b>
<i>GATT-motif</i>	1	<b>AE-box</b>	8/2				
<b>MBSII</b>	<b>1</b>	<b>CTAG-motif</b>	1/1				
<b>PC-CMA2A</b>	<b>1</b>	<i>ERE</i>	2/3				
<b>SBP-CMA1A</b>	<b>1</b>	<b>GATA-motif</b>	<b>4/2</b>				<b>6/1/3</b>
<b>TATCCAT/C-motif</b>	<b>2</b>	<i>LAMP-element</i>	1/1			<b>3/1/1</b>	
<i>Gm</i> ATX		<i>L-box</i>	1/1				
<b>3-AF3 BINDING SITE</b>	<b>3</b>	<b>P-box</b>	2/1				
<b>AT1-motif</b>	7	<i>Unnamed_6</i>	1/1				
<b>CE3</b>	<b>1</b>	<i>Unnamed_11</i>	1/2				
<b>chs-CMA1a</b>	2	<i>Gm-Zm</i> ATX					

<i>ERE</i>	1	<i>AuxRR-core</i>	3/2				
<b>F-box</b>	2	<b>Box III</b>	<b>4/2</b>		<b>2/1/1</b>		
<i>LAMP-element</i>	1	<b>Box-W1</b>	4/4				
<i>L-box</i>	2	<i>EIRE</i>	1/1				
<b>OBP-1 SITE</b>	<b>1</b>	<i>TCCACCT motif</i>	2/2				
<b>TA-rich region</b>	8	<b>W box</b>	4/4				
<i>Unnamed_6</i>	5	BOX E	1/1				
<i>Zm ASH</i>		<i>Gm-Pp ASH</i>					
<b>AAAC-motif</b>	1	<b>Box-W1</b>	2/2				
<b>ATC-motif</b>	1	<b>LTR</b>	<b>4/2</b>				<b>4/3/6</b>
<i>box S</i>	2	<b>MNF1</b>	<b>1/1</b>				<b>2/3/2</b>
<b>C-BOX</b>	<b>1</b>	<b>Unnamed_2</b>	<b>1/4</b>				<b>6/6/2</b>
<b>CE1</b>	<b>1</b>	<b>W box</b>	2/2				
<b>chs-CMA2a</b>	<b>1</b>	<b>WUN-motif</b>	2/1			<b>4/1/1</b>	
<b>F-box</b>	1	<i>Gm-Pp ATX</i>					
<b>GCC BOX</b>	<b>1</b>	<b>AC I</b>	<b>2/1</b>		<b>1/1/1</b>		
<b>GC-motif</b>	<b>4</b>	<b>AE box</b>	7/4			<b>5/9/4</b>	
<b>motif IIb</b>	2	<b>ATCT-motif</b>	<b>4/3</b>		<b>1/2/1</b>		
<i>TCCACCT-motif</i>	1	AT-RICH ELEMENT	3/2				
<i>Unnamed_15</i>	2	<b>Box I</b>	<b>14/7</b>		<b>5/5/1</b>		

<i>Zm</i> ATX		BOX II	1/1				
<b>ATGCAAAT motif</b>	<b>1</b>	<i>box S</i>	1/1	<b>2/2/1</b>			
motif IIb	1	<b>CAT-box</b>	<b>2/5</b>		<b>6/4/3</b>		
<b>NON-BOX</b>	<b>1</b>	<i>chs -Unit1 m1</i>	2/2				
<b>OCT</b>	<b>1</b>	<b>CTAG-motif</b>	1/2				
<b>UNNAMED_17</b>	<b>1</b>	ELI-BOX3	2/1				
<i>Pp</i> ASH		<i>GATT-motif</i>	1/1				
<i>AuxRR-core</i>	2	<b>MRE</b>	<b>5/2</b>	<b>2/1/2</b>			
<b>BOX II</b>	<b>1</b>	<b>P-box</b>	4/3				
<b>E2F</b>	<b>1</b>	<b>TATC-box</b>	1/1				
<i>EIRE</i>	1	<b>TCCC-motif</b>	<b>5/3</b>	<b>2/1/4</b>			
<b>MSA-LIKE</b>	<b>1</b>	Unnamed_11	5/1				
<b>RY-ELEMENT</b>	<b>1</b>	<b>WUN-motif</b>	3/1				
<b>TA-rich region</b>	11	<i>Zm-Pp</i> ASH					
<b>TGGCA-MOTIF</b>	<b>1</b>	<b>TATC-box</b>	1/1				
<b>UNNAMED_13</b>	<b>1</b>	<b>TGA-element</b>	<b>3/3</b>				<b>5/2/6</b>
<i>Unnamed_16</i>	1	UNNAMED_10	1/1				
<i>Pp</i> ATX		UNNAMED_12	1/1				
<b>AAAC-motif</b>	1	UNNAMED_14	1/1				
<i>3-AF1 binding site</i>	1	UNNAMED_8	1/1				

<b>ATCC-MOTIF</b>	<b>1</b>	UNNAMED_9	1/1				
<b>ATC-motif</b>	1	<i>Zm-Pp ATX</i>					
<b>CATTAT- MOTIF</b>	<b>2</b>	<b>As-2-box</b>	<b>1/2</b>		<b>2/2/3</b>		
<b>HD-ZIP 3</b>	<b>1</b>	<i>Unnamed_15</i>	1/1				
<b>MOTIF I</b>	<b>1</b>	<i>Unnamed_16</i>	1/1				

Anexo 4A. Tabla que muestra a los elementos –cis y su presencia-frecuencia en cada una de las secuencias de los promotores seleccionados de la clase III (ATX) para las tres especies estudiadas. El número entre paréntesis muestra su frecuencia.

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
3-AF1	1	Pp1s93_1 44V6.1 (1)							1
3-AF3	3		Glyma08g2901 0.1 (1)		Glyma16g0 2800.1 (1)			Glyma15g1 7030.1 (1)	3
5UTR Py-rich stretch	1			Glyma07g0 6190.1 (1) Glyma03g3 7370.1 (1) Glyma16g0 2800.1 (1)	Glyma04g4 1500.1 (5)	Glyma19g1 7460.1 (2)	Glyma12g1 1060.1 (4) Glyma06g4 5740.1 (2) Glyma13g3 8090.1 (1) Glyma12g3 2290.1 (2)	Glyma09g0 5740.1 (2) Glyma15g1 7030.1 (3)	2 4
	1						GRMZM2G 149153_T0 1(1)		1
	3	Pp1s282_23V6.1 (2)		Pp1s47_35 V6.1 (1)				Pp1s101_4 V6.1 (2)	5
mot	1							Pp1s133_3 0V6.1 (1)	1
AAGAA-motif	10		Glyma08g2901 0.1 (1)	Glyma07g0 6190.1 (1) Glyma03g3 7370.1 (2) Glyma16g0 2800.1 (3)	Glyma06g1 3330.1 (1)	Glyma19g1 7460.1 (3)	Glyma12g1 1060.1 (2) Glyma06g4 5740.1 (4)	Glyma09g0 5740.1 (1) Glyma15g1 7030.1 (1)	1 9
	2	GRMZM2 G013794 _T01(1)					GRMZM2G 149153_T0 1(1)		2
	4	Pp1s282_23V6.1 (3) Pp1s93_1 44V6.1 (2)		Pp1s47_35 V6.1 (1)				Pp1s133_3 0V6.1 (1)	7

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
ABRE	7			Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (1)	Glyma04g4 1500.1 (1)		Glyma12g1 1060.1 (2) Glyma06g4 5740.1 (3) Glyma13g3 8090.1 (2) Glyma12g3 2290.1 (3)		1 3
	2	GRMZM2 G013794 _T01(1)			GRMZM2G 085266_T0 1(1)				2
	5	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (1)		Pp1s223_9 2V6.1 (2) Pp1s47_35 V6.1 (2)				Pp1s101_4 V6.1 (1)	7
A-box	2			Glyma07g0 6190.1 (1)	Glyma04g4 1500.1(1)				2
	3	Pp1s93_1 44V6.1 (1)					Pp1s123_1 32V6.1(1)	Pp1s101_4 V6.1 (1)	3
ACE	6		Glyma0 8g2901 0.1 (1)	Glyma07g0 6190.1 (1) Glyma19g3 9970.1 (1)	Glyma04g4 1500.1 (1) Glyma06g1 3330.1 (2)			Glyma15g1 7030.1 (1)	7
	1				GRMZM2G 085266_T0 1(1)				1
	5	Pp1s282_ 23V6.1 (1)		Pp1s223_9 2V6.1 (1)			Pp1s311_1 7V6.1 (1) Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (1)	5
AC I	1				Glyma04g4 1500.1 (2)				2
	1						Pp1s123_1 32V6.1 (1)		1
AC II	2		Glyma1 8g5189 0.1 (1)	Glyma07g0 6190.1 (1)					2
	1	GRMZM2 G013794 _T01(1)							1
	2	Pp1s282_ 23V6.1 (1)						Pp1s133_3 0V6.1 (2)	3

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
AE-box	6		Glyma1 8g5189 0.1 (1)	Glyma07g0 6190.1 (1) Glyma03g3 7370.1 (1)			Glyma12g1 1060.1 (2) Glyma06g4 5740.1 (1)	Glyma09g0 5740.1 (1)	7
	1						GRMZM2G 149153_T0 1(1)		1
	3	Pp1s93_1 44V6.1 (2)						Pp1s133_3 0V6.1(1) Pp1s101_4 V6.1 (1)	4
ARE	1 2		Glyma0 8g2901 0.1 (3) Glyma1 8g5189 0.1	Glyma07g0 6190.1 (2) Glyma19g3 9970.1 (2) Glyma16g0 2800.1 (1)	Glyma04g4 1500.1 (1)	Glyma19g1 7460.1 (2)	Glyma12g1 1060.1 (2) Glyma06g4 5740.1 Glyma13g3 8090.1 (2) Glyma12g3 2290.1 (1)	Glyma15g1 7030.1 (1)	2 3
	3				GRMZM2G 085266_T0 1(1)		GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1(1)	4
	6	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (1)		Pp1s223_9 2V6.1 (2)			Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (2) Pp1s101_4 V6.1 (1)	8
	1							GRMZM2G 473138_T0 1(1)	1
as-2-box	2	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (1)							2
AT1-motif	6		Glyma1 8g5189 0.1 (1)	Glyma19g3 9970.1 (2)			Glyma06g4 5740.1 (1) Glyma13g3 8090.1 (1) Glyma12g3 2290.1 (1)	Glyma09g0 5740.1 (1)	7
C- mot	1						Pp1s311_1 7V6.1 (1)		1
ATCC- motif	1	Pp1s282_ 23V6.1 (1)							1

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
ATCT .motif	3		Glyma1 8g5189 0.1 (2)		Glyma04g4 1500.1 (1) Glyma06g1 3330.1 (1)				4
	3	Pp1s93_1 44V6.1 (1)		Pp1s223_9 2V6.1 (1)				Pp1s133_3 0V6.1 (1)	3
AT-rich element	3			Glyma07g0 6190.1 (1) Glyma03g3 7370.1 (1)		Glyma19g1 7460.1 (1)			3
	2							Pp1s133_3 0V6.1 (1) Pp1s101_4 V6.1 (1)	2
AAAT motif	1	GRMZM2 G013794 _T01(1)							1
AuxRR-core	2						Glyma13g3 8090.1 (2) Glyma12g3 2290.1 (1)		3
	1				GRMZM2G 085266_T0 1 (2)				2
Box 4	1		Glyma0 8g2901 0.1 (4) Glyma1 8g5189 0.1 (5)	Glyma07g0 6190.1 (3) Glyma03g3 7370.1(6) Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (5)	Glyma06g1 3330.1 (1)		Glyma12g1 1060.1 (2) Glyma06g4 5740.1 (4) Glyma13g3 8090.1 (2) Glyma12g3 2290.1 (1)	Glyma09g0 5740.1 (1)	3
	2				GRMZM2G 085266_T0 1(2)		GRMZM2G 149153_T0 1(1)		3
	4	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (2)		Pp1s47_35 V6.1 (2)				Pp1s133_3 0V6.1 (1)	6
box E	1						Glyma12g3 2290.1 (1)		1
	1							GRMZM2G 473138_T0 1(1)	1

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
Box I	8		Glyma1 8g5189 0.1 (1)	Glyma07g0 6190.1 (2) Glyma03g3 7370.1 (1) Glyma19g3 9970.1 (3) Glyma16g0 2800.1 (2)		Glyma19g1 7460.1 (1)	Glyma06g4 5740.1 (1)	Glyma09g0 5740.1(3)	1 4
	3			Pp1s223_9 2V6.1 (1) Pp1s47_35 V6.1 (3)				Pp1s133_3 0V6.1 (3)	7
Box II	1						Glyma06g4 5740.1 (1)		1
	1	Pp1s282_ 23V6.1 (1)							1
Box III	3			Glyma19g3 9970.1 (1)			Glyma06g4 5740.1 (2) Glyma13g3 8090.1 (1)		4
	2						GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1(1)	2
box S	1						Glyma06g4 5740.1 (1)		1
	1							Pp1s133_3 0V6.1 (1)	1
Box W1	4		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)		Glyma06g1 3330.1 (1)			1 Glyma15g1 7030.1 (1)	4
	3	GRMZM2 G013794 _T01(2)					GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1(1)	4

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
CAAT-box	15		Glyma08g29010.1(32) Glyma18g51890.1(43)	Glyma07g06190.1(25) Glyma03g37370.1(42) Glyma19g39970.1(33) Glyma16g02800.1(34)	Glyma04g41500.1(33) Glyma06g13330.1 (24)	Glyma19g17460.1 (50)	Glyma12g11060.1(33) Glyma06g45740.1(28) Glyma13g38090.1(32) Glyma12g32290.1(36)	Glyma09g05740.1(31) Glyma15g17030.1(29)	505
	4	GRMZM2G013794_T01 (18)			GRMZM2G085266_T01 (32)		GRMZM2G149153_T01 (23)	GRMZM2G473138_T01 (20)	93
	8	Pp1s282_23V6.1(28) Pp1s93_144V6.1(26)		Pp1s223_92V6.1(28) Pp1s47_35V6.1(34)			Pp1s311_17V6.1(15) Pp1s123_132V6.1(23)	Pp1s133_30V6.1(25) Pp1s101_4V6.1(29)	208
CAT-box	2				Glyma04g41500.1 (1)		Glyma13g38090.1 (1)		2
	3			Pp1s223_92V6.1(2) Pp1s47_35V6.1(2)			Pp1s311_17V6.1 (1)		5
CATT-motif	7			Glyma03g37370.1 (1)	Glyma04g41500.1 (1) Glyma06g13330.1 (1)		lyma12g11060.1 (1) Glyma06g45740.1 (1) Glyma12g32290.1 (1)	Glyma09g05740.1 (1)	7
	2	GRMZM2G013794_T01 (1)			GRMZM2G085266_T01(1)				2
	6	Pp1s282_23V6.1 (1) Pp1s93_144V6.1 (1)		Pp1s223_92V6.1 (1)			Pp1s311_17V6.1 (1) Pp1s123_132V6.1 (1)	Pp1s133_30V6.1 (1)	6
CATTAT-motif	2	Pp1s282_23V6.1 (1)					Pp1s311_17V6.1 (1)	2	

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
CCAAT-box	5			Glyma07g0 6190.1 (1) Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (1)			Glyma12g1 1060.1 (1) Glyma06g4 5740.1 (1)		5
	1						GRMZM2G 149153_T0 1(1)		1
	3	Pp1s93_1 44V6.1 (2)					Pp1s311_1 7V6.1 (1) Pp1s123_1 32V6.1 (1)		4
CCGTCC-box	2			Glyma07g0 6190.1 (1)	Glyma04g4 1500.1 (1)				2
	1						GRMZM2G 149153_T0 1(1)		1
	3	Pp1s93_1 44V6.1 (2)					Pp1s123_1 32V6.1 (1)	Pp1s101_4 V6.1 (1)	4
CE3	1					Glyma12g1 1060.1 (1)		1	
CGTCA-motif	6	Glyma08g 29010.1 (1) Glyma18g 51890.1 (1)		Glyma16g0 2800.1 (1)			Glyma12g1 1060.1 (3) Glyma06g4 5740.1 (1)	Glyma09g0 5740.1 (1)	8
	1						GRMZM2G 149153_T0 1(2)		2
	8	Pp1s282_ 23V6.1 (5) Pp1s93_1 44V6.1 (1)		Pp1s223_9 2V6.1 (1) Pp1s47_35 V6.1 (1)			Pp1s311_1 7V6.1 (3) Pp1s123_1 32V6.1 (3)	Pp1s133_3 0V6.1 (1) Pp1s101_4 V6.1 (3)	1 8
chs-CMA1a	2		Glyma0 8g2901 0.1 (1)	Glyma07g0 6190.1 (1)					2

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
chs-CMA2a	3			Glyma16g0 2800.1 (1)			Glyma12g1 1060.1 (2) Glyma06g4 5740.1 (1)		4
	1							GRMZM2G 473138_T0 1(1)	1
	1						Pp1s123_1 32V6.1 (1)		1
circadian	1 4		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)	Glyma03g3 7370.1 (3) Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (2)	Glyma04g4 1500.1 (2) Glyma06g1 3330.1 (1)	Glyma19g1 7460.1 (2)	Glyma12g1 1060.1 (1) Glyma06g4 5740.1 (1) Glyma13g3 8090.1 (1) Glyma12g3 2290.1 (1)	Glyma09g0 5740.1 (1) Glyma15g1 7030.1 (1)	1 9
	4	GRMZM2 G013794 _T01(1)			GRMZM2G 085266_T0 1(1)		GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1 (3)	6
	8	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (3)		Pp1s223_9 2V6.1 (1)			Pp1s311_1 7V6.1 (1) Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (5) Pp1s101_4 V6.1 (2)	1 4
chs-unit1 m1	2				Glyma06g1 3330.1 (1)			Glyma09g0 5740.1 (1)	2
	1	Pp1s282_ 23V6.1 (2)							2
CTAG-motif	1						Glyma12g3 2290.1 (1)		1
	2	Pp1s93_1 44V6.1 (1)					Pp1s123_1 32V6.1 (1)		2
E2Fa	2		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)						2
	1	GRMZM2 G013794 _T01(1)							1
	1							Pp1s101_4 V6.1 (1)	1

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
EIRE	1				Glyma04g4 1500.1 (1)				1
	1				GRMZM2G 085266_T0 1(1)				1
ELI-box3	2			Glyma19g3 9970.1 (1)			Glyma06g4 5740.1 (1)		2
	1							Pp1s101_4 V6.1 (2)	1
ERE	1			Glyma19g3 9970.1 (1)					1
F-box	2		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)						2
GA-motif	6			Glyma19g3 9970.1 (1)		Glyma19g1 7460.1 (1)	Glyma13g3 8090.1 (2) Glyma12g3 2290.1 (1)	Glyma09g0 5740.1 (1) Glyma15g1 7030.1 (1)	7
	1						GRMZM2G 149153_T0 1(1)		1
	1							Pp1s133_3 0V6.1 (1)	1
GAG-motif	7		Glyma0 8g2901 0.1 (1)	Glyma07g0 6190.1 (1) Glyma19g3 9970.1 (2)	Glyma04g4 1500.1 (1)	Glyma19g1 7460.1 (1)	Glyma12g1 1060.1 (1)	Glyma09g0 5740.1 (1)	8
	4	GRMZM2 G013794 _T01(2)			GRMZM2G 085266_T0 1(1)		GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1(1)	5
	4	Pp1s282_ 23V6.1 (3) Pp1s93_1 44V6.1 (1)					Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (1)	6

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
GARE-motif	9		Glyma1 8g5189 0.1 (1)		Glyma04g4 1500.1 (3) Glyma06g1 3330.1 (2)	Glyma19g1 7460.1 (4)	Glyma12g1 1060.1 (1) Glyma13g3 8090.1 (1) Glyma12g3 2290.1(1)	Glyma09g0 5740.1(3) Glyma15g1 7030.1 (1)	1 7
	1							GRMZM2G 473138_T0 1 (2)	1
	3	Pp1s93_1 44V6.1 (1)		Pp1s47_35 V6.1 (1)			Pp1s311_1 7V6.1 (1)		3
GATA-motif	6		Glyma0 8g2901 0.1 (1)	Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (1)	Glyma04g4 1500.1 (1)		Glyma12g1 1060.1 (1) Glyma06g4 5740.1 (1)		6
	1						GRMZM2G 149153_T0 1(1)		1
	3			Pp1s223_9 2V6.1 (1)			Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (1)	3
GATT-motif	1				Glyma04g4 1500.1 (1)				1
	1							Pp1s101_4 V6.1 (1)	1
G-Box	1 0		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)	Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (1)	Glyma04g4 1500.1 (1) Glyma06g1 3330.1 (1)		Glyma12g1 1060.1 (2) Glyma06g4 5740.1 (2) Glyma13g3 8090.1(1) Glyma12g3 2290.1 (2)		1 3
	2				GRMZM2G 085266_T0 1(2)		GRMZM2G 149153_T0 1(1)		3
	6	Pp1s282_ 23V6.1 (2)		Pp1s223_9 2V6.1 (2) Pp1s47_35 V6.1 (1)			Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (1) Pp1s101_4 V6.1 (1)	8

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
G-box	1 1		Glyma08g2901 0.1 (1) Glyma18g5189 0.1 (1)	Glyma19g39970.1 (4) Glyma16g02800.1 (2)	Glyma04g41500.1(4) Glyma06g13330.1 (3)	Glyma19g17460.1 (2)	Glyma12g11060.1 (4) Glyma06g45740.1 (3) Glyma13g38090.1(1) Glyma12g32290.1 (3)		2 8
	3	GRMZM2 G013794 _T01(1)			GRMZM2G 085266_T0 1(5)		GRMZM2G 149153_T0 1 (1)		7
	8	Pp1s282_23V6.1 (5) Pp1s93_144V6.1 (2)		Pp1s223_92V6.1 (2) Pp1s47_35V6.1 (3)			Pp1s311_17V6.1 (3) Pp1s123_132V6.1 (2)	Pp1s133_30V6.1 (2) Pp1s101_4V6.1 (1)	2 0
GC-motif	5			Glyma07g06190.1 (1)	Glyma04g41500.1 (1)		Glyma12g11060.1 (1) Glyma06g45740.1 (1) Glyma13g38090.1 (1)		5
	3	GRMZM2 G013794 _T01(2)			GRMZM2G 085266_T0 1(4)		GRMZM2G 149153_T0 1(3)		9
	4	Pp1s282_23V6.1 (1) Pp1s93_144V6.1 (1)		Pp1s223_92V6.1 (1)				Pp1s101_4V6.1 (1)	4
GCN4_motif	8			Glyma07g06190.1(1)	Glyma04g41500.1 (3) Glyma06g13330.1 (2)	Glyma19g17460.1 (2)	Glyma12g11060.1 (4) Glyma06g45740.1 (1) Glyma12g32290.1 (2)	Glyma15g17030.1 (1)	1 6
	3	GRMZM2 G013794 _T01(1)			GRMZM2G 085266_T0 1(1)			GRMZM2G 473138_T0 1(1)	3
	1	Pp1s93_144V6.1(1)		Pp1s223_92V6.1 (1)				Pp1s101_4V6.1 (1)	3

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
GT1-motif	9		Glyma08g29010.1 (4) Glyma18g51890.1 (2)	Glyma03g37370.1(1) Glyma16g02800.1 (2)			Glyma12g11060.1 (4) Glyma06g45740.1 (3) Glyma12g32290.1 (1)	Glyma09g05740.1 (2) Glyma15g17030.1 (1)	20
	2						GRMZM2G149153_T01(1)	GRMZM2G473138_T01(1)	2
	1	Pp1s93_144V6.1 (1)							1
- Zip	1							Pp1s133_30V6.1 (1)	1
HSE	11		Glyma08g29010.1 (1) Glyma18g51890.1 (3)	Glyma07g06190.1 (2) Glyma16g02800.1 (3)	Glyma04g41500.1 (2)		Glyma12g11060.1 (1) Glyma06g45740.1 (3) Glyma13g38090.1 (2) Glyma12g32290.1 (1)	Glyma09g05740.1 (1) Glyma15g17030.1 (3)	22
	2				GRMZM2G085266_T01(2)		GRMZM2G149153_T01(1)		3
	4	Pp1s282_23V6.1 (1) Pp1s93_144V6.1 (1)		Pp1s223_92V6.1 (2)				Pp1s133_30V6.1 (2)	6
I-box	3			Glyma03g37370.1(1)	Glyma06g13330.1 (1)	Glyma19g17460.1 (1)			3
	4	GRMZM2G013794_T01(1)			GRMZM2G085266_T01(2)		GRMZM2G149153_T01(1)	GRMZM2G473138_T01(1)	5
	6	Pp1s282_23V6.1 (1) Pp1s93_144V6.1 (1)		Pp1s223_92V6.1 (1)			Pp1s311_17V6.1 (1) Pp1s123_132V6.1 (1)	Pp1s133_30V6.1 (2)	7
LAMP-element	1						Glyma09g05740.1 (1)	1	

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
L-box	2						Glyma13g3 8090.1 (1) Glyma12g3 2290.1 (1)		2
LTR	4			Glyma07g0 6190.1 (1) Glyma03g3 7370.1 (1)			Glyma13g3 8090.1 (1)	Glyma09g0 5740.1 (1)	4
	2				GRMZM2G 085266_T0 1(2)		GRMZM2G 149153_T0 1(1)		3
	4	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (3)		Pp1s223_9 2V6.1 (1) Pp1s47_35 V6.1 (1)					6
MBS	9		Glyma0 8g2901 0.1 (1)	Glyma03g3 7370.1 (1) Glyma19g3 9970.1 (4)	Glyma04g4 1500.1 (2) Glyma06g1 3330.1(2)	Glyma19g1 7460.1 (5)	Glyma06g4 5740.1(1)	Glyma09g0 5740.1 (1) Glyma15g1 7030.1 (1)	1 8
	4	GRMZM2 G013794 _T01(2)			GRMZM2G 085266_T0 1(2)		GRMZM2G 149153_T0 1 (3)	GRMZM2G 473138_T0 1(1)	8
	4	Pp1s282_ 23V6.1 (3) Pp1s93_1 44V6.1 (1)					Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (2)	7
MNF1	2				Glyma04g4 1500.1 (1)			Glyma09g0 5740.1 (1)	2
	2						GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1 (2)	3
	2	Pp1s282_ 23V6.1 (1)						Pp1s133_3 0V6.1 (1)	2
motif I	1						Pp1s133_3 0V6.1 (1)	1	
motif IIb	1				GRMZM2G 085266_T0 1(1)				1

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
MRE	4		Glyma1 8g5189 0.1 (1)	Glyma03g3 7370.1 (1) Glyma19g3 9970.1 (1)	Glyma06g1 3330.1 (2)				5
	2						Pp1s311_1 7V6.1 (1)	Pp1s101_4 V6.1 (1)	2
NON-box	1	GRMZM2 G013794 _T01(1)							1
O2-site	4		Glyma1 8g5189 0.1 (1)			Glyma19g1 7460.1 (1)	Glyma12g1 1060.1 (1)	Glyma15g1 7030.1(1)	4
	6	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (2)		Pp1s47_35 V6.1 (1)			Pp1s311_1 7V6.1 (1) Pp1s123_1 32V6.1 (1)	Pp1s133_3 0V6.1 (2)	8
OBP-1 site	1						Glyma12g1 1060.1 (1)		1
OCT	1						GRMZM2G 149153_T0 1(1)		1
P-box	3		Glyma0 8g2901 0.1 (2)	Glyma07g0 6190.1 (1) Glyma16g0 2800.1 (1)					4
	3			Pp1s223_9 2V6.1 (1) Pp1s47_35 V6.1 (1)				Pp1s101_4 V6.1 (1)	3

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
Skn-1_ motif	1 3		Glyma08g2901 0.1 (3)	Glyma07g06190.1 (2) Glyma03g37370.1 (3) Glyma19g39970.1 (1) Glyma16g02800.1 (3)	Glyma06g13330.1 (1)	Glyma19g17460.1 (1)	Glyma12g11060.1 (5) Glyma06g45740.1 (2) Glyma13g38090.1 (2) Glyma12g32290.1 (2)	Glyma09g05740.1 (6) Glyma15g17030.1 (4)	3 5
	4	GRMZM2 G013794 _T01(2)			GRMZM2G 085266_ TO 1(5)		GRMZM2G 149153_ TO 1 (4)	GRMZM2G 473138_ TO 1(1)	1 2
	8	Pp1s282_23V6.1 (3) Pp1s93_144V6.1 (2)		Pp1s223_92V6.1 (3) Pp1s47_35V6.1 (3)			Pp1s311_17V6.1 (6) Pp1s123_132V6.1 (7)	Pp1s133_30V6.1 (3) Pp1s101_4V6.1 (4)	3 1
Sp1	1 0		Glyma08g2901 0.1 (1) Glyma18g5189 0.1 (3)	Glyma03g37370.1 (2) Glyma19g39970.1 (1)	Glyma04g41500.1 (1) Glyma06g13330.1 (1)		Glyma06g45740.1 (1) Glyma12g32290.1 (2)	Glyma09g05740.1 (1) Glyma15g17030.1 (1)	1 4
	4	GRMZM2 G013794 _T01(2)			GRMZM2G 085266_ TO 1(7)		GRMZM2G 149153_ TO 1(3)	GRMZM2G 473138_ TO 1(2)	1 4
	8	Pp1s282_23V6.1(8) Pp1s93_144V6.1 (5)		Pp1s223_92V6.1(14) Pp1s47_35V6.1 (2)			Pp1s311_17V6.1 (5) Pp1s123_132V6.1 (6)	Pp1s133_30V6.1 (4) Pp1s101_4V6.1 (16)	6 0
TA-rich region	2			Glyma07g06190.1 (4)			Glyma12g11060.1 (4)		8

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
TATA-box	15		Glyma08g29010.1 (35) Glyma18g51890.1(74)	Glyma07g06190.1 (108) Glyma03g37370.1 (97) Glyma19g39970.1 (88) Glyma16g02800.1 (84)	Glyma04g41500.1 (14) Glyma06g13330.1 (19)	Glyma19g17460.1(18)	Glyma12g11060.1 (63) Glyma06g45740.1 (44) Glyma13g38090.1 (53) Glyma12g32290.1 (33)	Glyma09g05740.1 (48) Glyma15g17030.1 (54)	832
	4	GRMZM2G013794_T01(23)			GRMZM2G085266_T01(39)		GRMZM2G149153_T01(33)	GRMZM2G473138_T01(17)	112
	8	Pp1s282_23V6.1 (7) Pp1s93_144V6.1 (6)		Pp1s223_92V6.1 (6) Pp1s47_35V6.1 (10)			Pp1s311_17V6.1 (18) Pp1s123_132V6.1 (12)	Pp1s133_30V6.1 (24) Pp1s101_4V6.1 (9)	92
TATC-box	1			Glyma16g02800.1 (1)					1
	1			Pp1s223_92V6.1 (1)					1
TATCCAT/C-motif	3		Glyma18g51890.1 (1)		Glyma04g41500.1 (1)	Glyma19g17460.1 (1)			3
	1	GRMZM2G013794_T01(1)							1
	1						Pp1s311_17V6.1 (1)		1
TCA-element	9		Glyma08g29010.1 (1) Glyma18g51890.1 (1)	Glyma19g39970.1 (4)	Glyma04g41500.1 (3) Glyma06g13330.1 (1)		Glyma13g38090.1(2) Glyma12g32290.1 (1)	Glyma09g05740.1 (2) Glyma15g17030.1 (1)	16
	3	GRMZM2G013794_T01(1)			GRMZM2G085266_T01 (1)			GRMZM2G473138_T01 (1)	3
	4	Pp1s282_23V6.1 (1)					Pp1s311_17V6.1 (2) Pp1s123_132V6.1 (1)	Pp1s133_30V6.1 (3)	7

Anexo 4A. (continuación)

ELEMNTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
TCCACCT-motif	2		Glyma0 8g2901 0.1 (1)	Glyma16g0 2800.1 (1)					2
	2						GRMZM2G 149153_T0 1 (1)	GRMZM2G 473138_T0 1(1)	2
TCCC-motif	4		Glyma0 8g2901 0.1 (2) Glyma1 8g5189 0.1(1)				Glyma13g3 8090.1 (1)	Glyma15g1 7030.1 (1)	5
	3	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (1)						Pp1s101_4 V6.1 (1)	3
TC-rich repeats	10			Glyma03g3 7370.1 (1) Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (1)	Glyma04g4 1500.1 (1) Glyma06g1 3330.1 (1)	Glyma19g1 7460.1 (1)	Glyma12g1 1060.1 (1) Glyma06g4 5740.1 (1) Glyma13g3 8090.1 (1)	Glyma15g1 7030.1 (2)	11
	1						GRMZM2G 149153_T0 1(1)		1
	5	Pp1s282_ 23V6.1 (1) Pp1s93_1 44V6.1 (1)		Pp1s223_9 2V6.1(1)				Pp1s123_1 32V6.1 (1) Pp1s133_3 0V6.1 (2)	6
TCT-motif	9		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)	Glyma19g3 9970.1 (1) Glyma16g0 2800.1 (2)	Glyma04g4 1500.1 (2) Glyma06g1 3330.1 (3)		Glyma12g1 1060.1 (2) Glyma06g4 5740.1 (1)	Glyma15g1 7030.1 (2)	15
	2						GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1(1)	2
	6	Pp1s282_ 23V6.1(1) Pp1s93_1 44V6.1 (1)		Pp1s223_9 2V6.1 (1) Pp1s47_35 V6.1(1)				Pp1s311_1 7V6.1 (2) Pp1s101_4 V6.1 (1)	7

Anexo 4A. (continuación)

ELEMENTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
TGA-element	4					Glyma19g1 7460.1 (1)	Glyma06g4 5740.1 (2) Glyma13g3 8090.1 (1)	Glyma09g0 5740.1 (1)	5
	2	GRMZM2 G013794 _T01(1)					GRMZM2G 149153_T0 1(1)		2
	4			Pp1s47_35 V6.1 (1)			Pp1s311_1 7V6.1 (2) Pp1s123_1 32V6.1 (1)	Pp1s101_4 V6.1 (2)	6
TGACG-motif	6		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)	Glyma16g0 2800.1 (1)			Glyma12g1 1060.1 (3) Glyma06g4 5740.1 (1)	Glyma15g1 7030.1 (1)	8
	1						GRMZM2G 149153_T0 1(2)		2
	8	Pp1s282_ 23V6.1 (5) Pp1s93_1 44V6.1 (1)		Pp1s223_9 2V6.1 (1) Pp1s47_35 V6.1 (1)			Pp1s311_1 7V6.1 (3) Pp1s123_1 32V6.1 (3)	Pp1s133_3 0V6.1 (1) Pp1s101_4 V6.1 (3)	1 8
Unnamed_1	1 1		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (2)	Glyma07g0 6190.1 (1) Glyma03g3 7370.1 (1) Glyma19g3 9970.1 (2)	Glyma04g4 1500.1 (3) Glyma06g1 3330.1 (3)	Glyma19g1 7460.1 (5)	Glyma12g1 1060.1(2) Glyma06g4 5740.1 (3) Glyma12g3 2290.1 (1)		2 4
	4	GRMZM2 G013794 _T01(3)			GRMZM2G 085266_T0 1(2)		GRMZM2G 149153_T0 1(5)	GRMZM2G 473138_T0 1(3)	1 3
	8	Pp1s282_ 23V6.1 (4) Pp1s93_1 44V6.1 (3)		Pp1s223_9 2V6.1 (4) Pp1s47_35 V6.1 (2)			Pp1s311_1 7V6.1 (6) Pp1s123_1 32V6.1 (1)	Pp1s101_4 V6.1 (1)	2 1

Anexo 4A. (continuación)

ELEMNTTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
Unnamed_2	6			Glyma19g3 9970.1 (1)	Glyma06g1 3330.1 (1)		Glyma12g1 1060.1 (1) Glyma06g4 5740.1 (1) Glyma13g3 8090.1 (1)	Glyma09g0 5740.1 (1)	6
	3				GRMZM2G 085266_TO 1(2)		GRMZM2G 149153_TO 1(3)	GRMZM2G 473138_TO 1(1)	6
	2			Pp1s223_9 2V6.1 (1)			Pp1s311_1 7V6.1 (1)		2
Unnamed_3	8		Glyma0 8g2901 0.1 (1)	Glyma07g0 6190.1 (1) Glyma19g3 9970.1 (2)	Glyma04g4 1500.1 (3) Glyma06g1 3330.1 (3)	Glyma19g1 7460.1 (5)	Glyma12g1 1060.1 (1) Glyma06g4 5740.1 (1)		1 7
	4	GRMZM2 G013794 _T01(3)			GRMZM2G 085266_TO 1 (2)		GRMZM2G 149153_TO 1(5)	GRMZM2G 473138_TO 1(3)	1 3
	8	Pp1s282_ 23V6.1 (4) Pp1s93_1 44V6.1 (3)		Pp1s223_9 2V6.1 (4) Pp1s47_35 V6.1 (2)			Pp1s311_1 7V6.1 (6) Pp1s123_1 32V6.1 (1)	Pp1s101_4 V6.1 (1)	2 1
Unnamed_4	1 5		Glyma0 8g2901 0.1 (11) Glyma1 8g5189 0.1 (8)	Glyma07g0 6190.1 (5) Glyma03g3 7370.1 (6) Glyma19g3 9970.1 (11) Glyma16g0 2800.1 (4)	Glyma04g4 1500.1 (4) Glyma06g1 3330.1 (9)	Glyma19g1 7460.1 (7)	Glyma12g1 1060.1 (8) Glyma06g4 5740.1 (8) Glyma13g3 8090.1 (8) Glyma12g3 2290.1 (12)	Glyma09g0 5740.1 (7) Glyma15g1 7030.1 (7)	1 1 5
	4	GRMZM2 G013794 _T01(15)			GRMZM2G 085266_TO 1(15)		GRMZM2G 149153_TO 1(22)	GRMZM2G 473138_TO 1(7)	5 9
	8	Pp1s282_ 23V6.1 (29) Pp1s93_1 44V6.1 (20)		Pp1s223_9 2V6.1 (15) Pp1s47_35 V6.1 (17)			Pp1s311_1 7V6.1 (16) Pp1s123_1 32V6.1 (15)	Pp1s133_3 0V6.1 (13) Pp1s101_4 V6.1 (11)	1 3 6
Unnamed_6	4		Glyma1 8g5189 0.1 (1)	Glyma07g0 6190.1 (1) Glyma19g3 9970.1 (2)	Glyma06g1 3330.1 (1)				5

Anexo 4A. (continuación)

ELEMNTO -CIS	Frecuencia	ATX1	ATX2	ATX3	ATX4	ATX5	ATXR3	ATXR7	TOTAL
Unnamed_11	5			Glyma07g0 6190.1 (1) Glyma03g3 7370.1 (1) Glyma16g0 2800.1 (1)			Glyma12g1 1060.1 (1) Glyma06g4 5740.1 (1)		5
	1	Pp1s282_ 23V6.1 (1)							1
Unnamed_15	1						GRMZM2G 149153_T0 1(1)		1
Unnamed_16	1				GRMZM2G 085266_T0 1(2)				2
Unnamed_17	1						GRMZM2G 149153_T0 1(1)		1
W box	4		Glyma0 8g2901 0.1 (1) Glyma1 8g5189 0.1 (1)		Glyma06g1 3330.1 (1)			Glyma15g1 7030.1 (1)	4
	3	GRMZM2 G013794 _T01(2)					GRMZM2G 149153_T0 1(1)	GRMZM2G 473138_T0 1(1)	4
WUN-motif	3			Glyma19g3 9970.1(1)	Glyma04g4 1500.1 (1)	Glyma19g1 7460.1 (1)			3
	1							Pp1s101_4 V6.1 (1)	1

Anexo 4B. Tabla que muestra a los elementos –cis y su presencia-frecuencia en cada una de las secuencias de los promotores seleccionados de la clase II (ASH) para las tres especies estudiadas. El número entre paréntesis muestra su frecuencia.

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
3-AF1	2	GRMZM2G147619_T03 (1)				GRMZM2G451374_T01 (1)	2
3-AF3	1			Glyma09g28430.1 (1)			1
4cl-CMA2b	1		Glyma04g42410.1 (1)				1
	1			GRMZM2G033694_T03 (1)			1
5UTR Py-rich stretch	5	Glyma20g30870.1 (1)	Glyma04g42410.1 (3)	Glyma16g33220.1 (1) Glyma09g28430.1 (2)		Glyma06g299600.1 (1)	8
	3		GRMZM2G352431_T02 (7)	GRMZM2G033694_T03 (1)		GRMZM2G451374_T01 (1)	9
	2		Pp1s28_356V6.1 (3)	Pp1s102_150V6.1 (1)			4
AAAC-motif	1		GRMZM2G352431_T02 (1)				1
AAGAA-motif	4	Glyma10g36720.1 (1)	Glyma04g42410.1 (3)	Glyma16g33220.1 (1)		Glyma06g299600.1 (1)	6
	1					GRMZM2G067019_T01 (2)	2
	4		Pp1s28_356V6.1 (2) Pp1s183_22V6.1 (2)	Pp1s102_150V6.1 (1)		Pp1s76_183V6.2 (1)	6
A-box	1	Glyma10g36720.1 (1)					1
	2	GRMZM2G147619_T03 (1)				GRMZM2G451374_T01(1)	2
	3		Pp1s28_356V6.1 (1) Pp1s183_22V6.1 (2)			Pp1s76_183V6.2 (1)	4

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
ABRE	2		Glyma04g4241 0.1 (1)			Glyma06g299600 .1 (1)	2
	5		GRMZM2G352 431_T02 (3)	GRMZM2G033 694_T03 (10)	GRMZM2G130 910_T01 (3)	GRMZM2G45137 4_T01 (1) GRMZM2G06701 9_T01 (1)	1 8
	3		Pp1s183_22V6. 1 (2)	Pp1s102_150V 6.1 (2)		Pp1s76_183V6.2 (1)	5
ACE	3	Glyma10g3672 0.1 (2) Glyma20g3087 0.1 (1)	Glyma04g4241 0.1 (1)				4
	4	GRMZM2G147 619_T03 (1)	GRMZM2G352 431_T02 (1)	GRMZM2G033 694_T03 (1)		GRMZM2G06701 9_T01(1)	4
AC I	1		Glyma04g4241 0.1 (1)				1
	1			GRMZM2G033 694_T03 (1)			1
	1		Pp1s183_22V6. 1 (1)				1
AC II	2		Glyma04g4241 0.1 (1) Glyma06g1239 0.1 (1)				2
	1			GRMZM2G033 694_T03 (1)			1
AE-box	7	Glyma10g3672 0.1 (1) Glyma20g3087 0.1 (1)	Glyma04g4241 0.1 (1) Glyma06g1239 0.1 (1)	Glyma16g3322 0.1 (1) Glyma09g2843 0.1 (2)		Glyma06g299600 .1 (1)	8
	2	GRMZM2G147 619_T03 (1)				GRMZM2G45137 4_T01 (1)	2

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
ARE	7	Glyma10g36720 .1 (2) Glyma20g30870 .1 (2)	Glyma04g42410 .1 (6) Glyma06g12390 .1 (2)	Glyma16g33220 .1 (3) Glyma09g28430 .1 (3)		1 Glyma06g29960 0.1 (1)	1 9
	5	GRMZM2G1476 19_T03 (1)		GRMZM2G0336 94_T03 (2)	GRMZM2G1309 10_T01 (3)	GRMZM2G4513 74_T01 (1) GRMZM2G0670 19_T01 (2)	9
	4		Pp1s28_356V6. 1 (1) Pp1s183_22V6. 1 (1)	Pp1s102_150V 6.1 (2)		Pp1s76_183V6. 2 (2)	6
as-2-box	2	Glyma20g30870 .1 (1)	Glyma06g12390 .1 (1)				2
	2			GRMZM2G0336 94_T03 (1)		GRMZM2G4513 74_T01 (1)	2
	2		Pp1s28_356V6. 1 (2)			Pp1s76_183V6. 2 (1)	3
AT1- motif	1		Glyma16g33220 .1 (1)			1	
C- motif	1		GRMZM2G3524 31_T02 (1)				1
ATCT-motif	1					Glyma06g29960 0.1 (1)	1
	2	GRMZM2G1476 19_T03 (1)				GRMZM2G4513 74_T01 (1)	2
	1					Pp1s76_183V6. 2 (1)	1
ATGCAAT-motif	2	Glyma20g30870 .1 (1)		Glyma09g28430 .1 (1)			2
	2		GRMZM2G3524 31_T02 (1)			GRMZM2G0670 19_T01 (1)	2
	1			Pp1s102_150V 6.1 (1)			1
AuxRR -core	1		Pp1s28_356V6. 1 (2)				2

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
Box 4	6	Glyma10g3672 0.1 (5) Glyma20g3087 0.1 (3)	Glyma04g4241 0.1 (1) Glyma06g1239 0.1 (1)	Glyma16g3322 0.1 (3) Glyma09g2843 0.1 (2)			15
	2		GRMZM2G352 431_T02 (1)			GRMZM2G451 374_T01 (2)	3
	2		Pp1s28_356V6. 1 (1)			Pp1s76_183V6. 2 (1)	2
Box I	3	Glyma20g3087 0.1 (2)		Glyma16g3322 0.1 (2)		Glyma06g2996 00.1 (1)	5
	3			GRMZM2G033 694_T03 (2)	GRMZM2G130 910_T01 (2)	GRMZM2G451 374_T01 (1)	5
	1		Pp1s183_22V6. 1 (1)				1
Box III	1					Glyma06g2996 00.1 (2)	2
	1		GRMZM2G352 431_T02 (1)				1
	1					Pp1s76_183V6. 2 (1)	1
Box W1	2			Glyma09g2843 0.1 (1)		Glyma06g2996 00.1 (1)	2
	2		Pp1s28_356V6. 1 (1)			Pp1s76_183V6. 2 (1)	2
box =	1			Pp1s102_150V 6.1 (1)			1
box S	2			GRMZM2G033 694_T03 (1)		GRMZM2G451 374_T01 (1)	2
CAAT-box	7	Glyma10g3672 0.1 (34) Glyma20g3087 0.1 (34)	Glyma04g4241 0.1 (23) Glyma06g1239 0.1 (23)	Glyma16g3322 0.1 (41) Glyma09g2843 0.1(33)		Glyma06g2996 00.1 (38)	22 6
	6	GRMZM2G147 619_T03 (16)	GRMZM2G352 431_T02 (6)	GRMZM2G033 694_T03 (28)	GRMZM2G130 910_T01 (24)	GRMZM2G451 374_T01 (18) GRMZM2G067 019_T01 (34)	12 6
	4		Pp1s28_356V6. 1 (20) Pp1s183_22V6. 1 (16)	Pp1s102_150V 6.1 (16)		Pp1s76_183V6. 2 (29)	81

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
CAT-box	3	Glyma20g3087 0.1 (1)	Glyma04g4241 0.1 (2) Glyma06g1239 0.1 (2)			Glyma06g2996 00.1 (1)	6
	2			GRMZM2G033 694_T03 (1)		GRMZM2G451 374_T01 (3)	4
	2		Pp1s28_356V6. 1 (1) Pp1s183_22V6. 1 (2)				3
CATT-motif	5	Glyma10g3672 0.1 (2) Glyma20g3087 0.1 (2)	Glyma04g4241 0.1 (1) Glyma06g1239 0.1 (1)	Glyma16g3322 0.1 (2)			8
	4		GRMZM2G352 431_T02 (1)	GRMZM2G033 694_T03 (1)		GRMZM2G451 374_T01 (1) GRMZM2G067 019_T01 (3)	6
	1		Pp1s183_22V6. 1 (3)				3
C-box	1					GRMZM2G067 019_T01 (1)	1
CCAAT-box	1	Glyma10g3672 0.1 (1) Glyma20g3087 0.1 (1)					2
	3	GRMZM2G147 619_T03 (1)			GRMZM2G13091 0_T01(1)	GRMZM2G067 019_T01 (1)	3
	2		Pp1s28_356V6. 1 (1)	Pp1s102_150V 6.1 (1)			2
CCGTCC-box	1	Glyma10g3672 0.1 (1)					1
	2	GRMZM2G147 619_T03 (1)				GRMZM2G067 019_T01 (1)	2
	3		Pp1s28_356V6. 1 (1) Pp1s183_22V6. 1 (2)			Pp1s76_183V6. 2 (1)	4
CE1	1	GRMZM2G147 619_T03 (1)				1	

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
CGTCA-motif	5	Glyma10g36720 .1 (2) Glyma20g30870 .1 (1)	Glyma04g42410 .1 (2) Glyma06g12390 .1 (1)	Glyma09g28430 .1 (1)			7
	6	GRMZM2G1476 19_T03 (1)	GRMZM2G3524 31_T02 (1)	GRMZM2G0336 94_T03 (7)	GRMZM2G1309 10_T01 (2)	GRMZM2G4513 74_T01 (2) GRMZM2G0670 19_T01 (1)	1 4
	4		Pp1s28_356V6. 1 (3) Pp1s183_22V6. 1 (4)	Pp1s102_150V 6.1 (2)		Pp1s76_183V6. 2 (1)	1 0
CM A1	1	Glyma20g30870 .1 (1)					1
CM A2	1			GRMZM2G0336 94_T03 (1)			1
Uni t.1	1		Glyma04g42410 .1 (1)				1
circadian	4	Glyma10g36720 .1 (2)	Glyma04g42410 .1(1)	Glyma09g28430 .1 (1)		Glyma06g29960 0.1 (1)	5
	3	GRMZM2G1476 19_T03 (1)				GRMZM2G4513 74_T01 (1) GRMZM2G0670 19_T01 (2)	4
	3		Pp1s28_356V6. 1 (1) Pp1s183_22V6. 1 (1)	Pp1s102_150V 6.1 (1)			3
CTAG- motif	1		Glyma04g42410 .1 (1)				1
	1				GRMZM2G1309 10_T01 (1)		1
E2F	1		Pp1s183_22V6. 1 (1)				1
E2 Fa	1					Glyma06g29960 0.1 (1)	1
EIR E	1			Pp1s102_150V 6.1 (1)			1
ERE	2	Glyma20g30870 .1 (1)		Glyma16g33220 .1 (1)			2
	2			GRMZM2G0336 94_T03 (2)	GRMZM2G1309 10_T01 (1)		3

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
F-box	1					GRMZM2G451374_T01 (1)	1
GA-motif	1					Glyma06g299600.1 (1)	1
	1	GRMZM2G147619_T03 (1)					1
	1					Pp1s76_183V6.2 (1)	1
GAG-motif	7	Glyma10g36720 .1 (1) Glyma20g30870 .1 (1)	Glyma04g42410 .1 (4) Glyma06g12390 .1 (1)	Glyma16g33220 .1 (1) Glyma09g28430 .1 (2)		Glyma06g299600.1 (1)	1 1
	4		GRMZM2G352431_T02 (1)	GRMZM2G033694_T03 (1)	GRMZM2G130910_T01 (2)	GRMZM2G451374_T01 (1)	5
	4		Pp1s28_356V6.1 (2) Pp1s183_22V6.1 (1)	Pp1s102_150V6.1 (1)		Pp1s76_183V6.2 (3)	7
GARE-motif	3		Glyma06g12390 .1 (2)	Glyma16g33220 .1 (2) Glyma09g28430 .1 (2)			6
	2		GRMZM2G352431_T02 (3)			GRMZM2G067019_T01 (1)	4
	1			Pp1s102_150V6.1 (2)			2
GATA-motif	4	Glyma10g36720 .1 (1) Glyma20g30870 .1 (1)	Glyma04g42410 .1 (1)	Glyma16g33220 .1 (1)			4
	2		GRMZM2G352431_T02 (1)			GRMZM2G451374_T01 (1)	2
GATT-motif	1		Glyma06g12390 .1 (1)				1
G-Box	2		Glyma04g42410 .1 (2)			Glyma06g299600.1 (1)	3
	5	GRMZM2G147619_T03 (1)	GRMZM2G352431_T02 (1)	GRMZM2G033694_T03 (5)	GRMZM2G130910_T01 (2)	GRMZM2G451374_T01 (1)	1 0
	3		Pp1s183_22V6.1 (1)	Pp1s102_150V6.1(2)		Pp1s76_183V6.2 (2)	5

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
G-box	4	Glyma20g3087 0.1 (1)	Glyma04g4241 0.1 (4) Glyma06g1239 0.1 (5)			Glyma06g29960 0.1 (1)	1 1
	5	GRMZM2G147 619_T03 (2)	GRMZM2G352 431_T02 (4)	GRMZM2G033 694_T03 (10)	GRMZM2G130 910_T01 (3)	RMZM2G06701 9_T01(5)	2 4
	3		Pp1s183_22V6. 1 (5)	Pp1s102_150V 6.1 (3)		Pp1s76_183V6. 2 (3)	1 1
GC-motif	4		GRMZM2G352 431_T02 (1)		GRMZM2G130 910_T01 (1)	GRMZM2G4513 74_T01 (1) GRMZM2G0670 19_T01 (1)	4
GCC-box	1	GRMZM2G147 619_T03 (1)					1
GCN4_motif	2		Glyma04g4241 0.1 (1)			Glyma06g29960 0.1 (1)	2
	2			GRMZM2G033 694_T03 (1)		GRMZM2G0670 19_T01 (1)	2
	1					2 Pp1s76_183V6. 2	2
GT1_motif	2			1 Glyma16g3322 0.1 (1)		1 Glyma06g29960 0.1 (1)	2
	5		GRMZM2G352 431_T02 (1)	GRMZM2G033 694_T03 (1)	GRMZM2G130 910_T01 (2)	GRMZM2G4513 74_T01 (2) GRMZM2G0670 19_T01 (2)	8
	3		Pp1s28_356V6. 1 (2) Pp1s183_22V6. 1 (1)			Pp1s76_183V6. 2 (1)	5
HSE	5	Glyma10g3672 0.1 (2) Glyma20g3087 0.1 (3)		Glyma16g3322 0.1 (1) Glyma09g2843 0.1 (2)		Glyma06g29960 0.1 (2)	1 0
	3	GRMZM2G147 619_T03 (2)	GRMZM2G352 431_T02 (1)		GRMZM2G130 910_T01 (2)		5
	3		Pp1s183_22V6. 1 (1)	Pp1s102_150V 6.1 (1)		Pp1s76_183V6. 2 (1)	3

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
I-box	3	Glyma10g3672 0.1 (2) Glyma20g3087 0.1 (1)	Glyma04g4241 0.1 (2)				5
	4	GRMZM2G147 619_T03 (1)	GRMZM2G352 431_T02 (2)			GRMZM2G451 374_T01 (1) GRMZM2G067 019_T01 (1)	5
	1					Pp1s76_183V6. 2 (2)	2
LAMP- element	1		Glyma04g4241 0.1 (1)				1
	1					GRMZM2G067 019_T01 (1)	1
L-box	1		Glyma04g4241 0.1 (1)				1
	1			GRMZM2G03369 4_T03 (1)			1
LTR	4	Glyma10g3672 0.1 (1) Glyma20g3087 0.1(1)	Glyma06g1239 0.1 (1)			Glyma06g2996 00.1 (1)	4
	1			Pp1s102_150V6. 1 (2)			2
MBS	7	Glyma10g3672 0.1 (2) Glyma20g3087 0.1 (2)	Glyma04g4241 0.1 (1) Glyma06g1239 0.1 (4)	Glyma16g33220. 1 (1) Glyma09g28430. 1 (3)		Glyma06g2996 00.1 (1)	1 4
	4	GRMZM2G147 619_T03 (3)		GRMZM2G03369 4_T03 (1)	GRMZM2G130 910_T01 (1)	GRMZM2G451 374_T01 (2)	7
	4		Pp1s28_356V6. 1 (3) Pp1s183_22V6. 1 (2)	Pp1s102_150V6. 1 (1)		Pp1s76_183V6. 2 (4)	1 0
MB S II	1	Glyma10g3672 0.1 (1)					1
MNF1	1		Glyma06g1239 0.1 (1)				1
	1		Pp1s183_22V6. 1 (1)				1
motif lib	2		GRMZM2G352 431_T02 (1)	GRMZM2G03369 4_T03(1)			2

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
MRE	2	Glyma20g30870.1 (1)		Glyma09g28430.1 (1)			2
	1	GRMZM2G147619_ T03 (1)					1
	2		Pp1s28_356V6 .1 (1) Pp1s183_22V6 .1 (1)				2
MASA-like	1		Pp1s28_356V6 .1 (1)				1
O2-site	2		Glyma04g4241 0.1 (1) Glyma06g1239 0.1 (3)				4
	2			GRMZM2G033694_ T03 (1)	GRMZM2G130910_T 01(1)		2
	2		Pp1s28_356V6 .1 (1) Pp1s183_22V6 .1 (1)				2
P-box	2	Glyma10g36720.1 (1)		Glyma09g28430.1 (1)			2
	1				RMZM2G130910_T01 (1)		1
Pc-CMA2a	1	Glyma20g30870.1 (1)					1
RY-element	1		Pp1s28_356V6 .1 (1)				1
sbp-CMA1a	1	Glyma20g30870.1 (1)					1

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
Skn-1_motif	5	Glyma10g3672 0.1 (3) Glyma20g3087 0.1(1)	Glyma06g1239 0.1 (1)	Glyma16g33220. 1 (2)		Glyma06g29960 0.1(2)	9
	6	GRMZM2G147 619_T03 (2)	GRMZM2G352 431_T02 (2)	GRMZM2G0336 94_T03(2)	GRMZM2G1309 10_T01 (2)	GRMZM2G4513 74_T01 (2) GRMZM2G0670 19_T01 (2)	1 2
	4		Pp1s28_356V6 .1 (3) Pp1s183_22V6 .1 (1)	Pp1s102_150V6 .1 (1)		Pp1s76_183V6. 2 (3)	8
Sp1	5	Glyma10g3672 0.1 (1) Glyma20g3087 0.1 (1)		Glyma16g33220. 1 (22) Glyma09g28430. 1 (11)		Glyma06g29960 0.1 (2)	3 7
	6	GRMZM2G147 619_T03 (2)	GRMZM2G352 431_T02 (5)	GRMZM2G0336 94_T03 (8)	GRMZM2G1309 10_T01(2)	GRMZM2G4513 74_T01(1)	1 8
	3		Pp1s28_356V6 .1 (3) Pp1s183_22V6 .1 (9)	Pp1s102_150V6 .1 (2)			1 4
TA-rich region	1			Pp1s102_150V6 .1(11)			1 1
TATA-box	7	Glyma10g3672 0.1 (63) Glyma20g3087 0.1 (51)	Glyma04g4241 0.1 (28) Glyma06g1239 0.1 (5)	Glyma16g33220. 1 (60) Glyma09g28430. 1 (40)		Glyma06g29960 0.1 (58)	3 0 5
	6	GRMZM2G147 619_T03 (58)	GRMZM2G352 431_T02 (79)	GRMZM2G0336 94_T03 (30)	GRMZM2G1309 10_T01 (38)	GRMZM2G4513 74_T01 (35) GRMZM2G0670 19_T01 (31)	2 7 1
	4		Pp1s28_356V6 .1 (7) Pp1s183_22V6 .1 (10)	Pp1s102_150V6 .1 (87)		Pp1s76_183V6. 2 (19)	1 2 3

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
TATC-box	1			GRMZM2G033694_T03 (1)			1
	1					Pp1s76_183V6.2 (1)	1
CA T/C	2		Glyma06g12390.1 (1)	Glyma16g33220.1 (1)			2
TCA-element	7	Glyma10g36720.1 (1) Glyma20g30870.1 (1)	Glyma04g42410.1 (3) Glyma06g12390.1 (1)	Glyma16g33220.1 (1) Glyma09g28430.1 (2)		Glyma06g299600.1 (2)	1 1
	3		GRMZM2G352431_T02 (1)	GRMZM2G033694_T03 (1)		GRMZM2G067019_T01 (1)	3
	4		Pp1s28_356V6.1 (2) Pp1s183_22V6.1 (2)	Pp1s102_150V6.1 (1)		Pp1s76_183V6.2 (1)	6
CC T-	1		GRMZM2G352431_T02 (1)				1
TCCC-motif	1			Glyma16g33220.1 (2)			2
	1					GRMZM2G067019_T01 (1)	1
	3		Pp1s28_356V6.1 (1) Pp1s183_22V6.1 (1)	Pp1s102_150V6.1 (2)			4
TC-rich repeats	4	Glyma10g36720.1 (1) Glyma20g30870.1 (4)	Glyma06g12390.1 (1)	Glyma16g33220.1 (1)			7
	2	GRMZM2G147619_T03 (3)				GRMZM2G451374_T01 (2)	5
	4		Pp1s28_356V6.1 (1) Pp1s183_22V6.1 (2)	Pp1s102_150V6.1 (1)		Pp1s76_183V6.2 (1)	5

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
TCT-motif	2	Glyma20g3087 0.1 (1)				Glyma06g299600 .1 (2)	3
	4	GRMZM2G147 619_T03 (1)	GRMZM2G352 431_T02 (1)	GRMZM2G033 694_T03 (1)	GRMZM2G130 910_T01 (1)		4
	2		Pp1s28_356V6. 1 (1)	Pp1s102_150V 6.1 (1)			2
TGACG-motif	5	Glyma10g3672 0.1(2) Glyma20g3087 0.1(1)	Glyma04g4241 0.1(2) Glyma06g1239 0.1(1)	Glyma09g2843 0.1(1)			7
	6	GRMZM2G147 619_T03 (1)	GRMZM2G352 431_T02 (1)	GRMZM2G033 694_T03 (7)	GRMZM2G130 910_T01 (2)	GRMZM2G45137 4_T01 (2) GRMZM2G06701 9_T01 (1)	1 4
	4		Pp1s28_356V6. 1 (3) Pp1s183_22V6. 1 (4)	Pp1s102_150V 6.1(2)		Pp1s76_183V6.2 (1)	1 0
TGA-elemt	1					GRMZM2G06701 9_T01(3)	3
	2		Pp1s183_22V6. 1 (2)	Pp1s102_150V 6.1 (1)			3
TGGC A-motif	1		Pp1s28_356V6. 1 (1)				1
Unnamed_1	4		Glyma04g4241 0.1 (2) Glyma06g1239 0.1 (2)	Glyma16g3322 0.1 (1)		Glyma06g299600 .1 (1)	6
	4		GRMZM2G352 431_T02 (1)	GRMZM2G033 694_T03 (3)	GRMZM2G130 910_T01 (1)	GRMZM2G06701 9_T01(3)	8
	3		Pp1s183_22V6. 1 (2)	Pp1s102_150V 6.1 (1)		Pp1s76_183V6.2 (2)	5
Unnamed_2	1		Glyma04g4241 0.1 (1)				1
	1		Pp1s183_22V6. 1 (4)				4

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
Unnamed_3	4		Glyma04g4241 0.1 (2) Glyma06g1239 0.1 (2)	Glyma16g3322 0.1 (1)		Glyma06g29960 0.1 (1)	6
	4		GRMZM2G352 431_T02 (1)	GRMZM2G033 694_T03 (3)	GRMZM2G13091 0_T01 (1)	GRMZM2G0670 19_T01(3)	8
	3		Pp1s183_22V6 .1(2)	Pp1s102_150V 6.1 (1)		Pp1s76_183V6.2 (2)	5
Unnamed_4	7	Glyma10g3672 0.1(8) Glyma20g3087 0.1 (7)	Glyma04g4241 0.1 (9) Glyma06g1239 0.1 (15)	Glyma16g3322 0.1 (10) Glyma09g2843 0.1 (2)		Glyma06g29960 0.1 (10)	6 1
	6	GRMZM2G147 619_T03 (13)	GRMZM2G352 431_T02 (17)	GRMZM2G033 694_T03 (14)	GRMZM2G13091 0_T01(15)	GRMZM2G4513 74_T01 (6) GRMZM2G0670 19_T01 (14)	7 9
	4		Pp1s28_356V6 .1 (15) Pp1s183_22V6 .1 (22)	Pp1s102_150V 6.1 (24)		Pp1s76_183V6.2 (15)	7 6
Unnamed_6	1	Glyma20g3087 0.1(1)					1
	1		GRMZM2G352 431_T02 (1)				1
Unnamed_8	1					GRMZM2G0670 19_T01 (1)	1
	1			Pp1s102_150V 6.1 (1)			1
Unnamed_9	1					GRMZM2G0670 19_T01 (1)	1
	1			Pp1s102_150V 6.1 (1)			1
Unnamed_10	1					GRMZM2G0670 19_T01 (1)	1
	1			Pp1s102_150V 6.1 (1)			1

Anexo 4B. (continuación)

ELEMENTO -CIS	Frecuencia	ASHH1	ASHH2	ASHH3	ASHH4	ASHR3	TOTAL
Unnamed_11	1			Glyma16g33220 .1 (1)			1
	2	GRMZM2G14 7619_T03 (1)	GRMZM2G35 2431_T02 (1)				2
Unnamed_12	1					GRMZM2G0670 19_T01 (1)	1
	1			Pp1s102_150V6 .1 (1)			1
Unnamed_13	1			Pp1s102_150V6 .1 (1)			1
Unnamed_14	1					GRMZM2G0670 19_T01 (1)	1
	1			Pp1s102_150V6 .1 (1)			1
Unnamed_15	2				GRMZM2G13091 0_T01 (1)	GRMZM2G4513 74_T01 (1)	2
Unnamed_16	1			Pp1s102_150V6 .1 (1)			1
W box	2			Glyma09g28430 .1 (1)		Glyma06g29960 0.1 (1)	2
	2		Pp1s28_356V 6.1 (1)			Pp1s76_183V6. 2 (1)	2
WUN-motif	2	Glyma10g367 20.1 (1) Glyma20g308 70.1 (1)					2
	1		Pp1s183_22V 6.1 (1)				1

Anexo 5. Diccionario de elementos –cis, sus secuencias y sus posibles funciones reportados actualmente en las bases de datos consultadas en esta investigación.

**3-AF1 binding site:** TAAGAGAGGAA. Elemento de respuesta a la luz.

**3-AF3 binding site:** CACTATCTAAC. Parte de un modulo de DNA conservado (CMA3).

**4cl-CMA2b:** TCTCACCAACCCCA. Forma parte del “4cl conserved DNA module array (4cl-CMA2)”, involucrado con la respuesta a luz (eventualmente UV), también llamado L-box (CTCACCTACCAA) por ser similares, no idénticos. Asociado con A- box (CCGTCC) y P-box (GCCTTTTGAGT) se involucran en inducir la actividad transcripcional, pero no suficiente para conferir.

**5UTR Py-rich stretch:** TTTCTCTCTCTCTC. Elemento que confiere elevados niveles de transcripción.

**AAAC-motif:** CAATCAAACCT. Parte del “atpCD conserved DNA module array (AtpCD-CMA)” involucrado con la respuesta a luz.

**AAGAA-motif:** GAAAGAA.

**A-box:** CCGTCC. Elemento asociado con P-box y L-box involucrado en inducir la actividad transcripcional pero no suficiente para conferir la capacidad de respuesta a.

**ABRE (ACGT-containing ABA Response Element):** GACACGTACGT. Elemento involucrado en la capacidad de respuesta al ácido abscisico.

**AC I (AC-I element)(Pv):** CCCACCTACC. Elemento que confiere una mayor expresión del xilema y reprime la expresión del floema en tabaco transgénico; refuerza la expresión de pétalos.

**AC II (AC II element):** CCACCAACCCCC. Repetición de AC II en el promotor confiere expresión específica del xilema. El factor de transcripción ACBF se une al elemento AC II.

**ACE (ACGT-Containing Element):** ACGTGGA. Elemento involucrado en la respuesta a luz interactúa in vitro e in vivo con factores reguladores comunes de plantas CPRFs (common plant regulatory factors) del tipo b-ZIP.

Anexo 5. (continuación)

**AE-box (Activating Element):** AGAAACAT. Parte de un sistema compuesto de 3 a 4 cajas huecas y 2 cajas AE, que confieren la capacidad de respuesta de luz (parte de un módulo para la respuesta de la luz).

**ARE (Anaerobic Response Element):** AAACCA. Elemento esencial para la inducción anaeróbica.

**As-2-box:** GATAatGATG. Participa en disparar la expresión específica y la capacidad de respuesta a luz.

**AT1-motif:** AATTATTTTTTATT. Parte de un módulo de respuesta de luz.

**ATCC-motif:** CAATCCTC. Parte del “Pc conserved DNA module array (Pc-CMA1)” involucrado con la respuesta a luz.

**ATC-motif:** AGTAATCT. Parte del “rbcA conserved DNA module array (rbcA-CMA1)” implicado con la respuesta a luz.

**ATCT-motif (LAMP-like motif):** AATCTGATCG. Parte del “gapA conserved DNA module array (gapA-CMA1)” implicado con la respuesta a luz.

**ATGCAAAT motif (Octameric binding site):** ATACAAAT. Elemento asociado al TGAGTCA motif (GCN4).

**AT-rich element (AT-rich sequence):** ATAGAAATCAA. Sitio de unión al “AT-rich DNA binding protein (ATBP-1)”.

**AuxRR-core (Core of the Auxin Response Region):** GGTCCAT. Elemento involucrado en la respuesta a auxinas.

**Box 4:** ATTAAT. Parte de un módulo de ADN conservado que participan en la respuesta de la luz.

**box E (fungal elicitor):** ACCCATCAAG. Elemento de inducción al desencadenamiento de hongos.

**Box I:** TTTCAA. Elemento de respuesta a luz.

**box II:** TCCACGTGGC. Parte del Unit-1 motif, Implicado con la respuesta a luz.

**Box II:** TGGTAATAA. Parte del “Ls conserved DNA module array (Ls-CMA1)” involucrado con la respuesta a luz.

Anexo 5. (continuación)

**Box III:** atCATTTTCACT/ACCCATCAAG. atCATTTTCACT. Sitio de unión a proteínas: identificada como una caja candidato sobre la base de homología con los genes rbcS (no demostrado ser activa) /CATTTTCACT involucrados en la respuesta de la luz.

**box S:** AGCCACC. Desencadena la respuesta a heridas y patógenos.

**Box-W1 ((T)TGAC(C)-like element):** TTGACC. Elemento provocador de respuesta a hongos. Sitio de unión a las proteínas del WRKY1 (U48831).

**CAAT-box:** CAAT. Elemento común en el promotor y regiones potenciadoras.

**CAT-box:** GCCACT. Elemento relacionado con la expresión del meristemo.

**CATTAT- motif:** GTCATTATCGG.

**CATT-motif:** GCATTC. Parte de un elemento de respuesta a luz.

**C-box:** CTGACGTCAG. Elemento de respuesta a luz; motivo palíndromo.

**CE1 (Coupling Element 1):** TGCCACCGG. Elemento acoplado con un elemento ABRE (el más cercano) implicado en la respuesta a ABA (abscisic acid).

**CE3 (Coupling Element 3):** GACGCGTGTC. Elemento involucrado en la respuesta a ABA y VP1.

**CGTCA-motif (complementary motif of TGACG):** CGTCA. Elemento involucrado en la respuesta al MeJA.

**chs-CMA1a (Chs Conserved DNA Module Array 2(part a)):** TTACTIONTAA. Parte del "chs conserved DNA module array (chs-CMA2)" involucrados con la respuesta de luz.

**chs-CMA2a:** TCACTIONTGA. Parte del "chs conserved DNA module array (chs-CMA2)" implicado con la respuesta a luz.

**chs-Unit1 m1 (Chs-unit 1 module 1):** ACCTAACCCGG. Parte del "chs-Unit 1 motif" involucrado con la respuesta a luz.

**Circadian (Circadian clock):** CAAAGATATC. Elemento implicado en el control del circadian.

**CTAG-motif:** ACTAGCAGAA.

**E2F (E2F-binding site):** AGTGCGGNNNNNTTTGAA. G1-M transición del ciclo celular.

**E2Fa (E2F-binding site):** TTTCCCGC. G1-M transición de ciclo celular.

Anexo 5. (continuación)

**EIRE (Elicitor Responsive Element):** TTCGACC. Necesario para provocar respuesta.

**ELI-box3 (Elicitor box 3):** AAACCAATT. Elemento desencadenante de respuesta.

**ERE (Ethylene-Responsive Element):** ATTTCAAA. Probablemente implicados en el envejecimiento de los pétalos.

**F-box:** CTATTCTCATT.

**GAG-motif:** AGAGAGT. Parte del “rbcA conserved DNA module array (rbcA-CMA1)” implicado en la respuesta a luz. (At)

**GA-motif:** AAAGATGA. Parte del “LRE rbcS-(I-G) Unit” en el gen rbcS.

**GARE-motif:** AACAG A/TCTGTTG. Elemento de respuesta a giberelinas.

**GATA-motif:** AAGGATAAGG / AAGATAAGATT. AAGGATAAGG Elemento de respuesta a luz. AAGATAAGATT. Parte del “fed conserved DNA module array (fed-CMA1)” implicado en la respuesta a luz. Solo muestra una mayor actividad en las plántulas crecidas a la luz, unido con G-motif o con Z-motif da de 4 a 8 veces mayor actividad en la luz; la luz inhibe la actividad en las raíces pero la promueve en los cotiledones. GT1-motif con G.

**GATT-motif (LS7-like motif (long name)):** CTGCAGATTTCT. Parte del “cab conserved DNA module array (cab-CMA5)” involucrado con la respuesta a luz.

**G-Box:** CACGTG. Elemento implicado con la respuesta a luz. (Ps)

**G-box:** GACATGTGGT/ CACGTT/ CACGTG. GACATGTGGT. Parte del “LRE rbcS-(I-G) Unit” en el gen rbcS. CACGTT/ CACGTG. Elemento implicado con la respuesta a luz.

**GCC box:** AGCCGCC. Desencadena la respuesta a heridas y patógenos.

**GC-motif:** CCACGGGG/ CCCCCG. CCACGGGG parte del “rbcS conserved DNA module array (rbcS-CMA7)” implicado con la respuesta a luz / CCCCCG. Elemento potenciador; implicado en la inducibilidad anóxica específica.

**GCN4-motif:** TGAGTCA. Elemento involucrado en la expresión del endospermo.

Anexo 5. (continuación)

**GT1-motif:** GGTTAA. Light responsive element (LRE); No responde a la luz si un segundo GT1-motif no está presente en el promotor CaMV 35S proporcionando mayor actividad en la luz. Dos GT1-motifs presents en el promotor proporcionan mayor actividad en la oscuridad que en la luz. Sitio de unión para la proteína nuclear del factor de GT1.

**HD-Zip 3 (Homeodomain-leucine zipper 3):** GTAAT(G/C)ATTAC. Sitio de unión a proteína.

**HSE (Heat Shock Element):** AGAAAATTCG. cis elemento involucrado en la respuesta de estrés por calor.

**I-box:** GATAAGATT. Parte del “LRE rbcS-(I-G) Unit” en el gen rbcS.

**LAMP-element:** CCAAACCA. Parte del “AtpCD conserved DNA module array (AtpCD-CMA)” involucrado con la respuesta a luz.

**L-box:** CTCACCTACCAA. Parte del “pal conserved DNA module array (pal-CMA1)” involucrado con la respuesta a luz. Asociado con A-box and P-box se involucran en inducir la actividad transcripcional pero no suficiente para conferir respuesta a.

**LTR (Low Temperature Response):** CCGAAA. Elemento involucrado en respuesta a bajas temperaturas.

**MBS (MYB binding site):** (T/C)AACTG. Implicado en la inducibilidad por sequía.

**MBSII (MYB DNA Binding Site II):** AAAAGTTAGTTA. Involucrado en la regulación de genes para la biosíntesis de flavonoides. Potenciador de la función.

**MNF1:** GTGCCC(A/T)(A/T). Elemento de respuesta a luz.

**motif I:** gGTACGTGGCG. Elemento específico de la raíz. Activador de la transcripción con la especificidad de las raíces podría estar implicado en la activación del desarrollo en las diferentes etapas.

**motif IIb:** CCGCCGCGCT. Elemento de respuesta al ácido abscisico.

**MRE (MYB Recognition Element):** AACCTAA. Sitio de unión de MYB implicado en la respuesta a luz.

**MSA-like:** TCCAACGGT. Elemento implicado en la regulación del ciclo celular.

Anexo 5. (continuación)

**NON-box (Nonamer-box):** AGATCGACG. Elemento relacionado con la activación específica del meristemo.

**O2-site (Opaque-2 site):** GATGA(C/T)(A/G)TG(A/G). Elementos reguladores que intervienen en la regulación del metabolismo zeína (zein).

**OBP-1 site (Site of the OBF Binding Protein):** TACA(TTTT)GG. Aumenta significativamente la capacidad de las proteínas OBF de enlazar los elementos ocs.

**OCT (Octamer):** CGCGGATC. Elemento relacionado con la activación específica del meristemo.

**P-box (Pyrimidine box):** GCCTTTTGAGT. Elemento de respuesta a giberelina.

**Pc-CMA2a (Pc Conserved DNA Module Array (part a)):** CAACCAATGAAAA. Parte del “Pc conserved DNA module array (Pc-CMA2)” involucrado con la respuesta a luz.

**RY-element:** CATGCATG. Elemento involucrado en la regulación específica de la semilla.

**sbp-CMA1a (Sbp Conserved DNA Module Array 1(part a):** CTCTATCTTCTCCGG. Parte del “sbp conserved DNA module array (sbp-CMA1)” involucrado en respuesta a luz.

**Skn-1\_motif:** GTCAT. Elemento necesario para la expresión del endospermo. En interacción cooperativa con otros motivos (AACAA; GCN4; ACGT) es necesario para mayores niveles de expresión del endospermo.

**Sp1:** CC(G/A)CCC. Elemento de respuesta a luz.

**TATA-box:** TATAAA. Elemento en el núcleo del promotor en torno a -30 del inicio de la transcripción. Controla el nivel de la transcripción y determina la ubicación y el número de sitios de inicio de transcripción.

**TA-rich región:** TATATATATATATATATATA. Potenciador de transcripción.

**TATC-box:** TATCCA. Involucrado en la respuesta a giberelina.

**TATCCAT/C-motif (Antisense GATA-motif):** TATCCAT. Asociado con G-box like motif; involucrado en la respuesta de la represión al azúcar.

**TCA-element:** CAGAAAAGGA/TCAGAAGAGG/GAGAAGAATA. Elemento involucrado en la respuesta al ácido salicílico.

Anexo 5. (continuación)

**TCCACCT-motif:** TCCACCT.

**TCCC-motif:** TCTCCCT. Parte del “AtpCD conserved DNA module array (AtpCD-CMA)” implicado con la respuesta a luz.

**TC-rich repeats:** ATTCTCTAAC/ ATTTTCTTCA/ GTTTTCTTAC/ ATTTTCTCCA. Elemento involucrados en la respuesta de defensa y del estrés.

**TCT-motif:** TCTTAC. Parte del “gapB conserved DNA module array (gapB-CMA1)” implicado con la respuesta a luz.

**TGACG-motif:** TGACG. Elemento regulador que interviene en la respuesta a MeJA- (sitio de unión de Arabidopsis bZIP a la proteína TGA1a), asociado con su complementario CGTCA-motif.

**TGA-element:** AACGAC. Elemento de respuesta a auxinas.

**TGGCA-motif (G-box like motif):** GATGGAAGTGGCA. Parte del “cab conserved DNA module array (cab-CMA5)” implicado con la respuesta a luz.

**Unnamed\_1:** CGTGG. 60K Sitio de unión a proteína.

**Unnamed\_2:** AACCTAACCT.

**Unnamed\_3:** CGTGG.

**Unnamed\_4:** CTCC.

**Unnamed\_6:** taTAAATATct.

**Unnamed\_8:** TCCACGTAGA.

**Unnamed\_9:** TCCACGTAGA.

**Unnamed\_10:** TCCACGTAGA.

**Unnamed\_11:** TCCACATAGA.

**Unnamed\_12:** TCCACGTAGA.

**Unnamed\_13:** TCCAAGTATA.

**Unnamed\_14:** TCCACGTAGA.

**Unnamed\_15:** CCTCTCCCGTC.

**Unnamed\_16:** GCTGCCCGTC.

**Unnamed\_17:** TAGGAGCAGCT.

Anexo 5. (continuación)

**W box (wounding and pathogen respons):** TTGACC. Desencadena respuesta a heridas y patógenos. Unión con factores de transcripción tipo WRKY.

**WUN-motif:** TCATTACGAA/ AAATTCCT. Elemento de respuesta a heridas.

## BIBLIOGRAFÍA:

1. Allis, C.D., Jenuwein, T., y Reinberg, D. (2007). Overview and Concepts, en: Allis, C.D., Jenuwein, T., Reinberg, D., Caparros M. (Eds.). *Epigenetics*, Cold Spring Harbor, pp. 25–56.
2. Alvarez-Venegas, R., Sadler, M., Tikhonov, A., y Avramova, Z. (2007). Origin of the Bacterial SET Domain Genes: Vertical or Horizontal? *Mol. Biol. Evol.* 24(2):482–497.
3. Alvarez-Venegas, R., y Avramova, Z. (2002). The SET-domain proteins of the Su(var)3-9, E(z) and Trithorax families. *Gene*. 285: 25-37
4. Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schiltz, I., Reuter, G., y Aalen, R.B. (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET-domain proteins that can be assigned to four evolutionary conserved classes. *Nucleic Acid Res*; 29:4319-4333.
5. Fauteux, F. y Strömvik, M.V. (2009). Seed storage protein gene promoters contain conserved DNA motifs in *Brassicaceae*, *Fabaceae* and *Poaceae*. *BMC Plant Biology*. (en línea).
6. Frazer, K.A., Elnitski, L., Church, D.M., Dubchak, I., y Hardison, R.C. (2003). Cross-Species Sequence Comparisons: A Review of Methods and Available Resources. *Genome Res.* 13: 1-12
7. Kenichi Higo, Yoshihiro Ugawa, Masao Iwamoto y Tomoko Korenaga. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucl. Acids Res.* (1999) 27 (1):297-300.
8. Mejía-Guerra, M. y Lareo, L. Aproximaciones computacionales para la identificación de promotores eucariotas tipo II. *Universitas Scientiarum*, ene-jun, año/vol. 11, número especial, pp 35-47.
9. Ng, D.W., Wang, T., Chandrasekharan, M.B., Aramayo, R., Kertbundit, S., Hall, T.C. (2007). Plant SET domain-containing proteins: Structure, function and regulation. *Biochim et Biophys Acta*, 1769(5–6):316–329.
10. Pontvianne, F., Blevins, T., y Pikaard, C.S. (2010). *Arabidopsis* Histone Lysine Methyltransferases. *Advances in Botanical Research*, 53: 1-22

11. Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., Yu, G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, 15;290 (5499):2105-10.
12. Saleh A, Álvarez-Venegas R, Yilmaz M, Oahn-Le, Hou G, Sadler M, Al-Abdallat A, Xia Y, Lu G, Ladunga I, y Avramova Z. The Highly Similar Arabidopsis Homologs of Trithorax ATX1 and ATX2 Encode Proteins with Divergent Biochemical Functions. *The Plant Cell*, 2008; 20: 568–579.
13. Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, Lohuizen Mv, Tanay A, Cavalli G. Functional Anatomy of Polycomb and Trithorax Chromatin Landscapes in *Drosophila* Embryos. *PLoS Biology* 2009; 7:146-163.
14. Singh, L.N., y Hannenhalli, S. (2009). Correlated changes between regulatory cis elements and condition-specific expression in paralogous gene families. *Nucleic Acids Research*, 1–12
15. Springer, N.M., Napoli, C.A., Selinger, D.A., Pandey, R., Cone, K.C., Chandler, V.L., Kaeppler, H.F., y Kaeppler, S.M. (2003). Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiol* 132: 907-925.
16. Wang X, Haberer G. y Mayer K, FX. (2009). Discovery of *cis*-elements between *sorghum* and rice using co-expression and evolutionary conservation. *BMC Genomics*. (en línea).
17. Watson, Baker, Bell, Gann, Levine y Losick. 2004. *Molecular Biology of the Gene*. 5<sup>th</sup> edition. Pearson Education, San Francisco.
18. Yu, Y., Bu, Z., Shen, W-H., Dong, A. (2009) An update on histone lysine methylation in plants. *Progress in Natural Science*, 19: 407–413.

Enlaces:

<http://www.arabidopsis.org/> [consulta: febrero de 2011]

<http://www.arabidopsis.med.ohio-state.edu/AtcisDB/> [consulta: marzo de 2011]

<http://www.tamps.cinvestav.mx/~ertello/bioinfo/sesion10.pdf> [consulta: marzo de 2011]

<http://www.ncbi.nlm.nih.gov/guide/> [consulta: marzo de 2011]

<http://www.phytozome.net/search.php> [consulta: abril de 2011]

<http://http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html> [consulta: abril de 2011]

<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/> [consulta: abril de 2011]

<http://www.dna.affrc.go.jp/htdocs/PLACE/> [consulta: mayo de 2011]

<http://www.web-books.com/MoBio/Free/Ch4C.htm> [consulta: mayo de 2011]

Glossary – The NCBI Handbook - NCBI Bookshelf [consulta: junio de 2011]

<http://www.ncbi.nlm.nih.gov/books/NBK21106/> [consulta: junio de 2011]

[http://dieumsnh.qfb.umich.mx/bioqui2/cap2b.htm#\\_end5](http://dieumsnh.qfb.umich.mx/bioqui2/cap2b.htm#_end5) [consulta: mayo de 2011]