



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas

INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS



Control Escolar de Licenciatura

VOTOS DE APROBATORIOS

Secretaria ejecutiva del Instituto de Investigación en Ciencias Básicas Aplicadas de la Universidad Autónoma del Estado de Morelos.

P r e s e n t e .

Por medio de la presente le informamos que después de revisar la versión escrita de la tesis que realizó la C. **BUCAREY RUÍZ FRANCISCA CATALINA ANASTASSIA** con número de matrícula **20154001194** cuyo título es:

"Caracterización de la actividad-espacio temporal del promotor sintético prom35S-MSA en raíces transgénicas de *Phaseolus vulgaris* inoculadas con rizobia"

Consideramos que **SI** reúne los méritos que son necesarios para continuar los trámites para obtener el título de **Licenciado Ciencias Área Terminal de Bioquímica y Biología Molecular**.

Cuernavaca, Mor a 09 de marzo del 2021

Atentamente
Por una universidad culta

Se adiciona página con la e-firma UAEM de los siguientes:

DR. RAMON ANTONIO GONZALEZ GARCIA CONDE
DRA VERÓNICA LIRA RUAN
DRA. ROSANA SÁNCHEZ LÓPEZ
DRA. MA. DEL RAYO SÁNCHEZ CARBENTE
DRA ANABEL ORTÍZ CALTEMPA

PRESIDENTE
SECRETARIO
VOCAL
PRIMER SUPLENTE
SEGUNDO SUPLENTE



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

Sello electrónico

ROSANA SÁNCHEZ LÓPEZ | Fecha:2021-05-13 14:24:00 | Firmante

A+A++2kEWr4iq+r3bhC/jLLv4gDAcsGmUK9zTusatzGV3top+UOacMODEYdOp3icW233x2Bx66ApH7Qbzuy29cYe3F1gmCsF4e1HEk8iqxeCNS5RS7Q14/6f7fOq5ZlxdQVA8Wie sBD9DzxZe/f7Q26CePK/Vg5wKY/ZytAmRCVvHvyCnLY5+ppETMdBO/eoFnbkFgbn2+GHIBApfwXhCJS50isiV1IKw2H4l6n66lv5kENK7+11xjaa6BomAbgpouTYCFr/BdPCKeZh XpDQoBv4BXnwdBecvoRwOnAp6FrWR+CWLYu8Q5yqM+RDBKmyOqAQpoi8tdvVyeS9X2DJA==

RAMON ANTONIO GONZALEZ GARCIA CONDE | Fecha:2021-05-13 15:31:44 | Firmante

nJZPMA3x+8YEf6uhp+CPf8coBusLviG6JF700r6xMwzWfIB/Kh7xMdk1730cEodiFqfGKKF82N8VPtVLGfOe+okr7ZJ1kN/tiwrHSL4NwVpABtLShWXZcGyLKaCkpWlssv7usQ62 rJy3nWP5Z15jo+jaX6fA49oMbhYgfp/cw5PQAeG3e56a7Kn9SVix7byYpAMAwR8pX5eenNO19GCHpAqS88l4dYWN4DA/bwOuC/KlmjzI3OaZXHwvTpvAQ6DphzlixhXA6z5cC T11/CS6bM4IBUYIX58Z7C+Kkh1SEtuOJ4Yayyn2X4xLjbN5WzUIYvfMOeXaDeGQil9W6A==

ANABEL ORTIZ CALTEMPA | Fecha:2021-05-13 21:29:45 | Firmante

G5ZSNZ74Jt5CJz/MORSC7kSV+nl+ie2bdte2TAHhFbyuOG/2Xs3Cylcx+ZB3PB257l4pHHWEjkdj116KXmjaOxk1PRzrSa3/+gf2yz8mo+DEn+1/yAgn85Eo09aq2z0D4ul1Eo76v3 U715fCtpc1S8cqP3Q+CkM4nJPiswsWDDeJjP8JiHNPILUYmonemsua+ni6jehB48VvzavThxfS+V5GeMRsXcbMmz5+QBnCRpvdIQTRx6Uo7ujlR43RF1ovh2l0lUupYbYiGEkbOR STd/8Bjiz8apbXJDiOhM/ReluJIXW+2ro2ctqah+EhFPH0RlcPugfR58Qw3O0Kxgng==

MARIA DEL RAYO SANCHEZ CARBENTE | Fecha:2021-05-13 22:33:37 | Firmante

POhwo4Voaaj5ifu5ii9Kq2rl2qSncfo0kNcm+Keil0JPkaZuRcNidqaAVho+lf08aH7guc2L1v4BQxLkkjw1EEHGvMSEvQeubOXmvoUYDIWrm4DcqdUnzWvxSxc7TJ50U0ioeRTM 9ViARedKkAdxNZ8DhljLtcNyb22S0LoPjYUW5+nbd5E80EYmLy7Wn7xnQNWgqFKuF+9eevizjT8v8Q+9ue+48Ts3srTLpx8QV8lhUgSYASn8ukaAeGxq7oJ+eb6n3UF1XrDI+Ku ASCJHbjA9R1++m2Z4SacrCWjOh4yM2OfSmWjgnvRCxhmM+1qiHG8yhryGbePpucwlcGg==

VERONICA LIRA RUAN | Fecha:2021-06-01 16:50:37 | Firmante

THCR0wrzmnlnHb5OLPULJe70v1QoFGFZdpDvXelPj3SXhj1MQ2MpKdvjxHXJ0SBnaccpLd63birjfmThL4bCSnFVMWcWsOCnqXzwY4W/UdKqitGfVnZBNCBa8E0SWwV7Rd pFqFaDBZ4UYU6eDL+CGGyMcRmXDIIF4G07PSCm0UIVo3RwmydN/8+sK7/HiahOFiERk1RYm5u/bCzmmY6pyFUw6eriC3Wvk00Ok2CtDfA6z/8BbqOeXJPWAG0anIEZYdlG O5p0PdxwSapVnRMsia1cgle/kHHYBzMfWLATCuPydl2q1qCPbKpKa6mPNBUx9dOKWSFZORzkTv2n5Q1A==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



HdDAZv

<https://efirma.uaem.mx/noRepudio/WUxX9AHtdS0c1Sf1x5Ve5fWU31YOstd>





UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS
CENTRO DE INVESTIGACIÓN EN DINÁMICA CELULAR

**Caracterización de la actividad espacio-temporal de promotor sintético
prom35S-MSA en raíces transgénicas *Phaseolus vulgaris* inoculados con rizobia**

TESIS

para obtener el título de:

**LICENCIADO EN CIENCIAS
ÁREA TERMINAL BIOQUÍMICA Y BIOLOGÍA MOLECULAR**

PRESENTA

FRANCISCA CATALINA ANASTASSIA BUCAREY RUIZ

DIRECTOR DE TESIS:

Dra. Rosana Sánchez López

- **Presidente.** Dr. Ramón González García Conde
- **Secretario.** Dra. Verónica Lira Ruan
- **Suplente 1.** Dra. Ma. Del Rayo Sánchez Carbente
- **Suplente 2.** Dra. Anabel Ortiz Caltempa

CUERNAVACA, MORELOS

AGOSTO 2021

Este trabajo se realizó en el Departamento de Biología Molecular de Plantas del Instituto de Biotecnología-UNAM, con financiamiento parcial de proyecto DGAPA-PAPIIT IN206118, cuya responsable es la Dra. Rosana Sánchez López

ABREVIATURAS

5'UTR: Región 5 prima no traducida (**5'** **U**ntranslated **R**egion, por sus siglas en inglés).

3'UTR: Región 3 prima no traducida (**3'** **U**ntranslated **R**egion, por sus siglas en inglés).

A. thaliana: *Arabidopsis thaliana*

AtBD_MSA: base de datos que contiene el análisis de los 86 genes de *A. thaliana* identificados como portadores de motivos MSA por las plataformas MatInspector, PlantCARE, MAST y MCAST.

AtBD_Menges: Lista de los 155 genes reportados por Menges y colaboradores (2005).

AtBD_prom-5'UTR: base de datos que contiene el análisis de los fragmentos constituidos por 2000 pb río arriba del sitio TSS y la secuencia del 5'UTR de los genes en AtBD_Menges.

AtDB_Up: base de datos “Upstream sequences” *Arabidopsis thaliana*.TAIR10.37 de meme-suite, que contiene el análisis de un conjunto de 285,977 secuencias con una longitud entre 201 y 1201 pb.

ch: cromátidas hermanas

G2: fase Gap 2.

M: Mitosis.

MSA: Motivo activador específico de la mitosis (**M**itosis **S**pecific **A**ctivator, por sus siglas en inglés).

mt: microtúbulos.

P. vulgaris: *Phaseolus vulgaris*

PPB: La banda Preprofase (**P**re**p**rophase **B**and por sus siglas en inglés)

pb: pares de bases

PvBD_MSA: base de datos que contiene el análisis de 73 genes de *P. vulgaris* identificados como portadores de motivos MSA por las plataformas MatInspector, PlantCARE, MAST y MCAST.

PvBD_prom-5'UTR: base de datos que contiene el análisis de los fragmentos constituídos por 2000 pb río arriba del sitio TSS y la secuencia del 5'UTR de los 187 genes de *P. vulgaris* identificados (159 homólogos a los genes de la base de datos AtBD_MSA y 28 genes identificados en el análisis de PvDB_Up).

PvDB_Up: base de datos “Upstream sequences” *Phaseolus vulgaris*.PhaVulg1 0.38 de meme- suite que contiene el análisis de un conjunto de 175,879 secuencias de ADN con una longitud entre 152 y 1201 pb.

SIT: **S**itio de **i**nicio de la **T**raducción

STT: Sitio de terminación de la Traducción (codón de paro)

TSS: sitio de inicio de la Transcripción (Transcription Start Site, por sus siglas en inglés).

ÍNDICE

1. INTRODUCCIÓN	9
1.1 Generalidades de la simbiosis leguminosa:rizobio	9
1.1.1 Rizobia, bacterias fijadoras de nitrógeno.....	9
1.1.2 La familia Fabaceae o Leguminosae	9
1.1.3 <i>Phaseolus vulgaris</i> como modelo de estudio.....	10
1.1.4 Generalidades del desarrollo del nódulo fijador de nitrógeno	10
1.2 Ciclo celular en células vegetales	12
1.2.1 Interfase.....	13
1.2.2 Mitosis	14
1.2.3 Citocinesis vegetal y formación del fragmoplasto	14
1.3 Estructura del promotor de un gen eucariótico	16
1.3.1 Características de los promotores de transcripción de genes eucarióticos	17
1.3.2 Tipos de promotores	18
1.3.3 Promotor constitutivo 35S	18
2. ANTECEDENTES.....	20
2.1 Proteínas de la familia SNARE.....	20
2.2 La syntaxina KNOLLE, específica de citocinesis en células vegetales	21
2.3 El promotor del gen <i>PvKNOLLE</i>	21
2.4 Motivo MSA	22
3. JUSTIFICACIÓN	23
4. HIPÓTESIS	24
5. OBJETIVO GENERAL	25
5.1 OBJETIVOS PARTICULARES.....	25
6. ESTRATEGIA EXPERIMENTAL	26
7. MATERIALES Y MÉTODOS	27
7.1 Plásmidos disponibles en el laboratorio	27
7.2 Reacciones con enzimas de restricción, fosfatasa y ligasa	27
7.3 Reacción de clonación por recombinación	28
7.4 Transformación en <i>E. coli</i> TOP10 quimiocompetentes	28
7.5 PCR en colonia.....	28
7.6 Transformación en <i>A. rhizogenes</i> K599 electrocompetentes	28
7.7 Preparación de <i>A. rhizogenes</i> K599 electrocompetentes.....	29
7.8 Herramientas bioinformáticas para la búsqueda de motivos MSA.....	29
7.9 Herramientas bioinformáticas para la búsqueda de posibles homólogos a genes de <i>A. thaliana</i> en <i>P. vulgaris</i>	33

8.1 Diseño del promotor sintético prom35S-MSA en fusión transcripcional con una secuencia 5'UTR.....	34
8.2 Construcción de los plásmidos pBGWFS7-prom35MSA::5'UTR_PvKNOLLE y pBGWFS7-prom35S-MSA::5'UTR_OmegaLeader	35
8.2.1 Construcción del plásmido pENTR-prom35S- MSA::5'UTR_PvKNOLLE	36
8.2.2 Construcción del plásmido pENTR-prom35S-MSA::5'UTR_OmegaLeader	37
8.2.3 Recombinación de prom35S-MSA::5'UTR_PvKNOLLE, prom35S-MSA::5'UTR_OmegaLeader y prom35S::5'UTR_OmegaLeader en pBGWFS7	38
8.2.4 Electroporación de los plásmidos pBGFWS7-prom35S-MSA::5'UTR_PvKNOLLE, pBGFWS7-prom35S-MSA::5'UTR_OmegaLeader o pBGFWS7-prom35S::5'UTR_OmegaLeader en <i>A. rhizogenes</i> K599.....	39
8.3 Búsqueda de genes del ciclo celular en el genoma de <i>P. vulgaris</i> , cuyos promotores presenten motivos MSA.....	40
8.3.1 Búsqueda de genes de ciclo celular en <i>A. thaliana</i> e identificación de motivos MSA en secuencias de promotor	40
8.3.2 Identificación <i>in silico</i> de la secuencia de genes de <i>A. thaliana</i>	41
8.3.3 Búsqueda de motivos MSA en secuencias promotoras en <i>A. thaliana</i>	42
8.3.4 Búsqueda de motivos MSA en genes de <i>P. vulgaris</i>	46
8.3.5	46
8.3.7 Identificación de los motivos MSA en los 187 genes de <i>P. vulgaris</i> utilizando las plataformas PlantCARE, MAST y MCAST.....	48
8.3.8 Análisis de los motivos MSA en genes de <i>A. thaliana</i> y <i>P. vulgaris</i>	49
8.4 Posible función de los genes de <i>P. vulgaris</i> portadores de motivos MSA.	51
9. DISCUSIÓN	53
10. CONCLUSIONES	58
11. PERSPECTIVAS	59
12. REFERENCIAS.....	60
13. ANEXOS.....	66
Tabla A1. Genes en AtBD_Menges.....	66
Tabla A2: AtBD_prom-5'UTR.....	71
Tabla A4: AtBD_MSA.....	76
Tabla A5: Genes de <i>P. vulgaris</i> homólogos en AtBD_MSA	79
Tabla A6. Genes en PvBD_Up.....	84
Tabla A7. Análisis de PvBD_prom-5'UTR	85
Tabla A8. Posición de los motivos MSA con respecto al TSS (+1).....	90
Tabla A9. Dominios Pfam proteínas genes PvDB-MSA	93

RESUMEN

La citocinesis en células vegetales depende de la formación de una maquinaria única de estas células, denominada fragmoplasto, cuya función es la reorganización de los microtúbulos para formar una red citocinética en el centro de la célula en división. Este andamiaje sirve de guía para que vesículas derivadas de Golgi se desplacen al punto medio en el plano de división y se fusionen para dar lugar a una estructura membranosa conocida como placa celular. La sintaxina KNOLLE es esencial para la fusión vesicular citocinética. La transcripción y traducción del gen KNOLLE tiene lugar durante las fases G2/M y M del ciclo celular, respectivamente. La transcripción dependiente de ciclo celular está regulada por la actividad del promotor KNOLLE que en *Arabidopsis thaliana* y *Phaseolus vulgaris* (*P. vulgaris*) se detecta específicamente en células en división. Se postula que la especificidad pudiera estar determinada por los motivos MSA (siglas en inglés de Mitosis Specific Activator) presentes en tandem en el extremo proximal del promotor de KNOLLE de *P. vulgaris* y *A. thaliana*. Con el propósito de aportar elementos que refuerzen esta hipótesis y disponer de un promotor específico de células en división, en esta tesis se plantea el diseño de un promotor sintético tipo prom35S-MSA, cuya actividad espacio-temporal sea específica de células mitóticas. Para demostrar que la especificidad funcional de los motivos MSA es independiente del 5'UTR de PvKNOLLE, se diseñaron dos versiones del promotor sintético: prom35S-MSA::5'UTR_PvKNOLLE y, como control, prom35S-MSA::5'UTR_Omega Leader, ya que la secuencia 5'UTR_Omega Leader no presenta elementos de regulación. Ambas versiones de promotor sintético fueron construidas, el análisis de su actividad será tema de un futuro proyecto.

Un segundo objetivo de esta tesis fue la búsqueda bioinformática de motivos MSA en genes de *A. thaliana* y de *P. vulgaris*. Las secuencias analizadas corresponden a fragmentos que contienen 2000 pb de la secuencia río arriba del sitio de inicio de la transcripción y la secuencia 5'UTR de cada gen. Las herramientas bioinformáticas utilizadas fueron PlantCARE, MAST y MCAST. Para facilitar su análisis, la información obtenida fue ordenada en una serie de bases de datos, lo que permitió identificar 54 y 73 genes de *A. thaliana* y *P. vulgaris*, respectivamente, portadores de motivos MSA. 70 de los 100 motivos MSA presentes en los 73 genes de *P. vulgaris* se localizan en la secuencia del promotor, de los cuales 52% están distribuidos en el extremo proximal (en un fragmento de 500 pb río arriba del sitio de inicio de la transcripción). Los resultados en un análisis bioinformático de búsqueda de dominios conservados presentes en la secuencia de las proteínas codificadas por los genes de *P. vulgaris* identificados sugiere que la mayoría de estos genes tienen una función relacionada con los procesos de ciclo celular.

ABSTRACT

Cytokinesis in plant cells depends on the formation of a unique machinery, called the phragmoplast, whose function is the reorganization of microtubules to form a cytotkinetic network at the center of the dividing cell. This scaffold serves as a guide for Golgi-derived vesicles to move to the midpoint in the plane of division and fuse, to give rise to the membranous structure known as the cell plate. The syntaxin KNOLLE is essential for the cytotkinetic vesicle fusion. Transcription and translation of the KNOLLE gene takes place during the G2/M and M phases of the cell cycle, respectively. Cell cycle-dependent transcription is regulated by the activity of the KNOLLE promoter, which in *Arabidopsis thaliana* and *Phaseolus vulgaris* is specifically detected in dividing cells. It is postulated that the specificity could be determined by the Mitosis Specific Activator (MSA) motifs present in tandem at the proximal end of the KNOLLE promoter in *P. vulgaris* and *A. thaliana*. In order to provide elements that reinforce this hypothesis and to have a specific promoter for dividing cells, this thesis the design of a synthetic promoter prom35S-MSA type is proposed, whose spatio-temporal activity is specific for mitotic cells. To demonstrate that the functional specificity of the MSA motifs is independent of the PvKNOLLE 5'UTR sequence, two versions of the synthetic promoter were designed: prom35S-MSA::5'UTR_PvKNOLLE and, as a control, prom35S-MSA::5'UTR_Omega Leader, since the 5'UTR_Omega Leader sequence has no regulatory elements. Both synthetic promoter versions were constructed, but the analysis of their activity will be the subject of a future project.

A second objective of this thesis was the bioinformatic search for *A. thaliana* and *P. vulgaris* genes containing MSA motifs. The sequences analyzed correspond to fragments of 2,000 pb located upstream of the transcription start site and the 5'UTR sequence of each gene. The bioinformatics tools used were PlantCARE, MAST and MCAST. To facilitate their analysis, the information obtained was sorted into a series of databases, which allowed the identification of 54 and 73 genes of *A. thaliana* and *P. vulgaris*, respectively, carrying MSA motifs. 70 out of the 100 MSA motifs present in the 73 genes from *P. vulgaris* are located in the promoter sequence, of which 52% are distributed at the proximal end (in a 500 bp fragment upstream of the transcription start site). Results from a bioinformatic analysis searching for conserved domains present in the peptide sequence encoded by the identified *P. vulgaris* genes suggests that most of these genes have a function related to cell cycle processes.

1. INTRODUCCIÓN

La relación simbiótica entre los rizobios y las leguminosas resulta en una estructura denominada nódulo, en cuyo entorno anaeróbico se albergan bacterias rizobiales responsables de la actividad del complejo multienzimático denominada nitrogenasa, encargado de la fijación de nitrógeno.

1.1 Generalidades de la simbiosis leguminosa:rizobio

1.1.1 Rizobia, bacterias fijadoras de nitrógeno

La mayoría de los organismos son incapaces de asimilar la principal fuente de nitrógeno en la tierra, el nitrógeno molecular (N_2), el cual constituye aproximadamente el 75% del aire atmosférico, así que deben incorporar el nitrógeno a partir de compuestos como el nitrato y el amonio, o de fuentes orgánicas. Los únicos seres vivos capaces de asimilar el nitrógeno molecular, es decir, realizan la fijación biológica de nitrógeno, son los microorganismos conocidos como diazótrofos.

Las bacterias diazótroficas conocidas como rizobio, pertenecientes a la familia Rhizobiaceae, son aerobias, Gram negativas, con forma de bacilo y presentan flagelos; pueden vivir de forma libre en suelo o en la cercanía de raíces (rizósfera). Otra propiedad de los rizobia es que pueden establecer una interacción simbiótica con leguminosas. Esta relación simbiótica es de tipo especie-específica, es decir, cada especie de rizobio se asocia con una especie de leguminosa. Por ejemplo, *Medicago truncatula* establece simbiosis con bacterias del género *Sinorhizobium*, mientras que el endosimbionte de *Glycine max* es *Bradyrhizobium meliloti*, el de *Lotus japonicus* es *Mesorhizobium loti*, y *Phaseolus vulgaris* tiene como simbionte principal a *Rhizobium etli*, aunque también puede hacer simbiosis con *R. tropici* (Martínez-Romero y cols., 1991). Recientemente, se ha reportado que bacterias no-rizobiales tipo Burkholderia (*Ralstonia taiwanensis*, hoy *Cupriavidus taiwanensis*) son capaces de generar nódulos fijadores de nitrógeno en la leguminosa Mimosa (Andrews y Andrews, 2017). Cabe mencionar que el tipo de nódulo que se desarrolla (determinado o indeterminado, ver inciso 1.1.4) depende del tipo de leguminosa.

1.1.2 La familia Fabaceae o Leguminosae

Las leguminosas pertenecen a la familia Fabaceae o Leguminosae, la tercera familia más grande de las angiospermas y la segunda en importancia agrícola y económica. Varias especies de leguminosas han sido domesticadas para el consumo humano, por ejemplo, lenteja (*Lens esculenta*), frijol común, (*P. vulgaris*); chícharo, (*Pisum sativum*), entre otras. Las características morfológicas de las leguminosas van desde tener hojas simples a compuestas siendo pinadas, bipinadas, unifoliadas,

trifoliadas, o como un simple zarcillo, con estípulas o desarrollando espinas. Tienen órganos florales tipo actinomorfo y zigomorfo, con un solo carpelo superior; pero la característica morfológica principal es su fruto: en vainas de dos a varias semillas (Cardoso y cols., 2009).

1.1.3 *Phaseolus vulgaris* como modelo de estudio

El frijol común corresponde a uno de los cultivos más antiguos y de mayor importancia económica y nutricional a nivel mundial, por lo que juega un papel importante en la dieta de los habitantes, sobre todo en México, Centro y Sudamérica (Jacinto y cols., 1996)

La caracterización de su cualidad nutricional ha demostrado que estas leguminosas tienen un alto contenido proteico, de carbohidratos y aminoácidos como L-triptófano y L-metionina, además de poseer fitoquímicos como polifenoles, fibra, inhibidores de tripsina y lectinas (Ulloa y cols., 2011).

Por su importancia dentro de la dieta humana se han optimizado las condiciones de cultivo y se ha trabajado en la obtención de mejores plantas en cuanto a período vegetativo, producción y tamaño de la semilla, resistencia al ambiente y mayor calidad nutricional. Por otro lado, el estudio genómico y transcripcional de algunas leguminosas ha sido un aporte en el desarrollo de plantas transgénicas.

1.1.4 Generalidades del desarrollo del nódulo fijador de nitrógeno

Una de las características evolutivas más interesantes de las leguminosas es el proceso adaptativo que dio lugar al programa de desarrollo de la estructura simbiótica denominada nódulo fijador de nitrógeno, el que se desarrolla *de novo* en la raíz de la leguminosa. Existen dos tipos de nódulos, los indeterminados, que se caracterizan por tener forma ovalada y actividad meristemática persistente en el ápice, y son típicos de la nodulación en la leguminosa modelo *Medicago truncatula*. Los nódulos determinados, típicos de la planta modelo *L. japonicus*, *G. max* (soya) y *P. vulgaris* (frijol común), presentan una forma esférica con actividad meristemática en la etapa de maduración (Ferguson y cols., 2010).

La formación del nódulo conlleva la coordinación de una serie de procesos celulares complejos que determinan la infección y la organogénesis del nódulo (Oldroyd y Downie, 2008), como se ilustra en la Figura 1 y se describe a continuación:

1. El inicio está orquestado por un diálogo molecular entre la bacteria y las células epidermales diferenciadas a pelos radicales, en el que la proteína bacteriana NodD percibe los flavonoides exudados por la raíz hacia la rizósfera, como consecuencia se activa la expresión de genes rizobiales *nod*, responsables de la síntesis y secreción de compuestos lipoquitolisacáridos conocidos como factores Nod (FN).

La especificidad en la relación simbiótica está determinada, por un lado, por el tipo de flavonoide reconocido por el rizobio endosimbionte, por ejemplo, el flavonoide naringenina exudado por *P. vulgaris* es el reconocido por *R. etli* (Martínez-Romero y cols., 1991; Liu y Murray, 2016)

Los factores Nod tienen un esqueleto de 4-6 grupos de N-acetilglucosamina en cuyo extremo no-reductor contienen un ácido graso de 16-20 carbonos e insaturación variable, así como la presencia de radicales tipo fucosil, arabinosa, metilo, grupos sulfato, carbamilo (carbamoilo), entre otros, que en su conjunto determinan la especificidad de los FN por la leguminosa compatible modificaciones dependiendo de la especie de rizobio (Hirsch y cols., 1992).

2. Los FN son reconocidos por los receptores tipo cinasa con dominios LysM (NFR1 y NFR5 en *L. japonicus* o LYK3 y NFP en *M. truncatula*) presentes en la membrana plasmática de pelos radicales en crecimiento (Figura 1a). La percepción de los FN desencadena una serie de cambios en el pelo radical como son el influjo y eflujo de calcio, despolarización de la membrana plasmática, alcalinización del pH intracelular, rearrreglos del citoesqueleto, expresión de genes de nodulinas tempranas, deformación y enroscamiento del ápice del pelo radical, en cuyo pliegue quedan atrapadas las bacterias (Cárdenas y cols., 2002). A este espacio se le denomina cámara de infección (Figura 1b).

3. En la cámara de infección, las bacterias forman una microcolonia e inducen la invaginación de la pared celular y membrana plasmática del pelo radical, dando lugar a una estructura tubular, apoplástica, transcelular, llamada hilo de infección (HI, por sus siglas). El HI se extiende hacia la base del pelo radical (Figura 1c) y es el conducto por donde las bacterias entran a la raíz y se dirigen hacia el córtex (Figura 1d).

4. Simultáneo a la formación y al avance del HI, se reactiva el ciclo celular de las células corticales en el sitio de infección. En los nódulos indeterminados son las células del periciclo y córtex interno las que comienzan la división celular, mientras que en los determinados las células corticales adyacentes al pelo radical son las que la reinician. La constante actividad meristemática da lugar a la formación de un primordio de nódulo (Figura 2).

5. El HI se ramifica y avanza hacia las células del primordio del nódulo. Las bacterias son vertidas desde el hilo de infección hacia el interior de las células del tejido central del nódulo en formación (nódulo joven, Figura 2) mediante un proceso parecido a la endocitosis. En este entorno, la bacteria está contenida en una estructura subcelular conocida como simbiosoma, en donde se diferencian a bacteroides, los cuales expresan el complejo nitrogenasa que les confiere la capacidad de fijar nitrógeno (Mora y cols., 2005)

6. El nódulo maduro presenta un sistema vascular que conecta al nódulo con la vasculatura de la raíz.

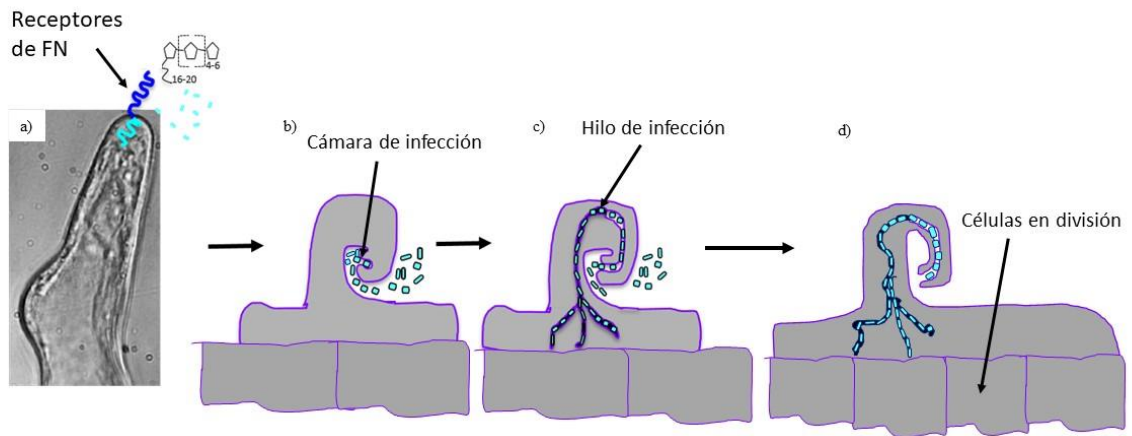


Figura 1. Esquema que ilustra las etapas en la formación del hilo de infección. a) Percepción de factores de nodulación (Modificado de Cárdenas y cols., 2006). b) Enroscamiento del pelo radical, formación de microcolonia en pliegue. c) formación de hilo de infección. d) reactivación del ciclo celular (Sánchez-López, R., esquema no publicado)

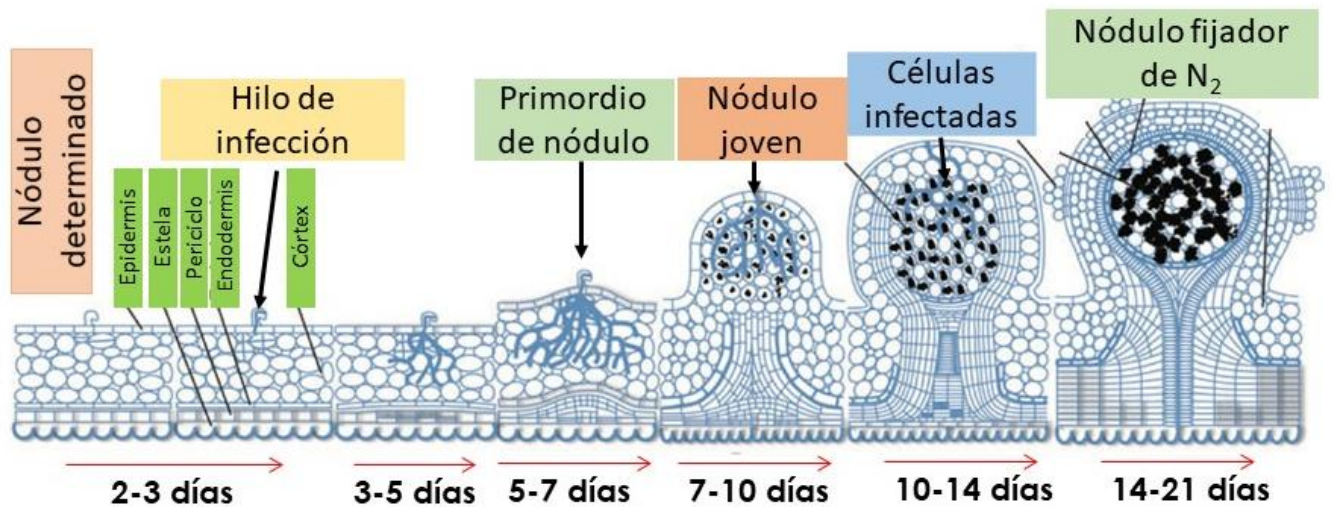


Figura 2. Esquema de formación de un nódulo determinado. Versión modificada de Ferguson y colaboradores (2010)

1.2 Ciclo celular en células vegetales

En organismos pluricelulares, como plantas y animales, la división celular lleva al crecimiento del organismo, y la reparación de tejidos y órganos (Raven y cols., 2005). En la raíz, la división celular se observa en la zona del meristemo apical (RAM, siglas en inglés de **R**oot **A**pical **M**eristem), así como durante la formación de una raíz lateral y del nódulo fijador de nitrógeno.

El ciclo celular consiste en la multiplicación de las células, al duplicar su contenido genético y dividirlo en dos células hijas, dando lugar a nuevas células similares en estructura y función a la célula parental (Alberts y cols., 2002) y se divide en etapas: interfase que comprende las fases G1, S y G2, etapa de mitosis (M), en la que hay división nuclear, y citocinesis, en la que tiene lugar la división citoplasmática (Figura 3).

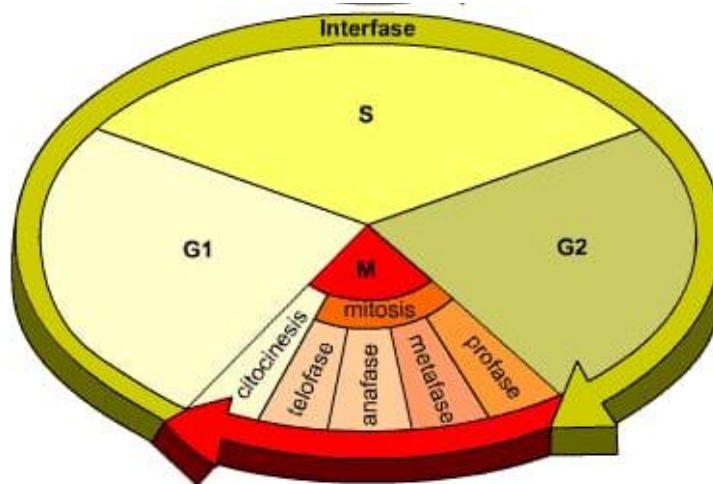


Figura 3. Esquema de las fases que constituyen el ciclo celular (Tomada de Alberts y cols., 2002).

1.2.1 Interfase

La interfase corresponde a la etapa de crecimiento celular y de replicación del material genético para la posterior división celular, y está organizada en 3 fases sucesivas: G1, S y G2. En la fase G1, la célula produce organelos y componentes citoplasmáticos que sean suficientes para las dos células hijas, es decir, la célula está metabólicamente activa. En esta fase tiene lugar uno de los puntos de control de la división (“checkpoint”). Durante la fase S ocurre la replicación del contenido genético y se sintetizan proteínas asociadas al ADN. La fase G2 es el segundo punto de control en el que la célula verifica que la síntesis de ADN haya concluido, que no exista daño en ADN y si fuese el caso lo repara. En esta fase sólo se transcriben y traducen los genes de la maquinaria responsable de la mitosis y citocinesis.

Tabla 1. La interfase se divide en tres fases sucesivas: G₁, S y G₂

Fase	Función
G ₁	La célula aumenta su tamaño, sintetiza más enzimas, ribosomas, organelos, moléculas citoplasmáticas y estructuras.
S	Replicación de ADN, tubulina y síntesis de histonas.
G ₂	Control de que no haya daño en el ADN, y eventualmente su reparación, y transcripción de genes de la maquinaria de mitosis y citocinesis.

1.2.2 Mitosis

La mitosis, que inicia al concluir la fase G₂, es el proceso de división nuclear y repartición equitativa del material genético en cada núcleo. Esta etapa se divide en 4 fases: profase, metafase, anafase y telofase, cuyas características se recapitulan en la Tabla 2, y se indican las propias de la división de células vegetales.

Tabla 2. Fases de la mitosis en células animales y vegetales

Etapa	Función
Profase	La cromatina se condensa en cromosomas, se desensambla la envoltura nuclear e inicia la reorganización de microtúbulos. En células vegetales los microtúbulos se ordenan paralelos al plano de la división celular y forman la banda de la pre-profase (PPB, por sus siglas en inglés)
Metafase	Se forma el huso mitótico y se reorganizan los microtúbulos (mt). En células vegetales los mt se organizan perpendiculares al plano de la división celular.
Anafase	Los cromosomas se direccionan hacia los polos opuestos. En células vegetales da inicio la formación del fragmoplasto.
Telofase	Se reensambla la envoltura nuclear. En células vegetales en esta etapa se inicia la formación de la placa celular.

1.2.3 Citocinesis vegetal y formación del fragmoplasto

En este proceso celular tiene lugar la distribución de organelos y componentes citoplasmáticos entre las dos células hijas.

La citocinesis en células vegetales se caracteriza por la organización de una maquinaria única de estas células, conocida como fragmoplasto, cuya función es el desarrollo coordinado de la placa celular en la zona central del plano ecuatorial de la célula en división. Otra característica de la citocinesis vegetal es que el retículo endoplásmico y el aparato de Golgi no se desensamblan y el tráfico vesicular se mantiene activo (Austin y cols., 2005)

Durante la fase G2, los microtúbulos pasan de tener una distribución cortical a un arreglo en forma de anillo, anclado a la membrana plasmática en el plano de la división celular, que rodea al núcleo, llamado banda de preprofase (PPB, **P**re**p**rophase **B**and por sus siglas en inglés). La PPB empieza a desaparecer cuando se establecen los husos mitóticos durante la metafase. Entre la anafase tardía y la telofase temprana, los microtúbulos forman una estructura tipo barril, denominada **fragmoplasto**, que se ubica entre los dos núcleos (Figura 4), y da soporte al inicio de la formación de la placa celular durante la citocinesis. El fragmoplasto está compuesto por un arreglo de microtúbulos antiparalelos, cuyos extremos más (“plus”) se orientan hacia la zona central y están organizados de manera perpendicular al plano de la división. El retículo endoplásmico, el aparato de Golgi y vesículas derivadas del Golgi que portan el material de la pared celular y membrana requeridos para la formación de la placa celular son reclutadas en la zona central del fragmoplasto. La formación de la placa celular se origina por la continua fusión de estas vesículas en el punto central del fragmoplasto, creando una estructura membranosa tipo sáculo que se extiende de manera radial hasta fusionarse con la membrana plasmática de la célula madre, dando lugar a dos células hijas independientes (Seguí-Simarro y cols. 2004). Entre los genes asociados a la formación de la placa celular se encuentran KEULE, Dynamin-related protein 1 y 2, las syntaxinas SNAP33 y KNOLLE.

En el grupo de investigación donde se desarrolló este proyecto, se utiliza KNOLLE como marcador molecular para visualizar la citocinesis y estudiar la dinámica entre la formación del HI y la división celular en el sitio de infección con rizobia. Para tal propósito se diseñó el cassette de expresión promPvKNOLLE::YFP-PvKNOLLE.

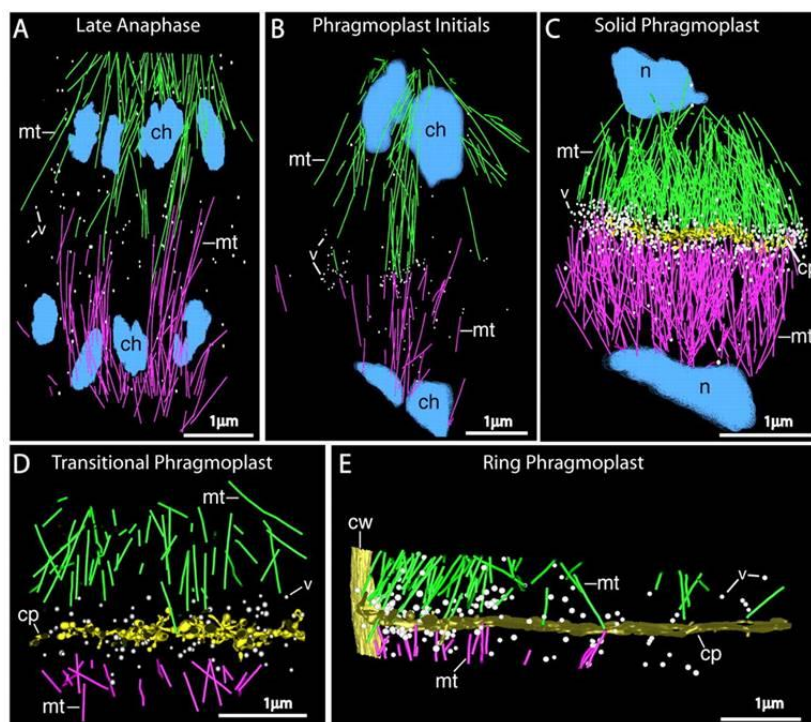


Figura 4. Reconstrucción a partir de tomografía por microscopía electrónica de la formación del fragmoplasto. a) Migración de las cromátidas hermanas (ch), y se alinean los microtúbulos (mt) perpendiculares al plano de la división celular. Se reclutan vesículas derivadas de Golgi (v). b) Se organiza el fragmoplasto en el plano de la división celular y la fusión de vesículas derivadas de Golgi, para iniciar la formación de la placa celular. c) Los núcleos (n) se ensamblan y separan hacia los polos opuestos al plano de la división, la placa celular (cp) crece desde el centro hacia la periferia de la célula en división. d) Los microtúbulos se desensamblan en el sitio de crecimiento de la placa. e) La placa celular se fusiona a la membrana parental (cw) para dar lugar a la separación de las células hijas. (Tomada de Austin y cols. 2005).

1.3 Estructura del promotor de un gen eucariótico

De manera general, un gen eucariótico que codifica para una proteína está organizado de la siguiente manera:

- la secuencia que se transcribe que consta de exones separados por intrones. El proceso de maduración del transcrito implica la eliminación de la secuencia de los intrones (“splicing”). El primer intrón contiene la secuencia 5’ no traducida o 5’UTR (por sus siglas en inglés de **U**n**t**r**a**n**s**l**a**t**e**d **R**egion) y el nucleótido en la posición 5’ corresponde al sitio de inicio de la transcripción del gen (TSS por siglas en inglés **T**r**a**n**s**cription **I**n**i**tiation **S**ite). Róo abajo, e inmediata al extremo 3’ de la región 5’UTR se encuentra el sitio de inicio de la traducción o SIT, y corresponde al primer codón AUG y suele estar en el primer exón. La secuencia entre el SIT y el sitio de terminación de la traducción (o STT) es la

que codifica para la secuencia aminoacídica. La secuencia que se encuentra río abajo del STT se conoce como la región 3' no traducida o 3' UTR

- la región río arriba del TSS comprende la secuencia promotora cuya función es la regulación de la transcripción del gen.

1.3.1 Características de los promotores de transcripción de genes eucarióticos

El promotor contiene las secuencias que actúan como sitios de unión de la ARN polimerasa y de factores de transcripción que activan o inhiben la transcripción de un gen de manera espacio-temporal (Andersson y Sandelin, 2019; Geertz y cols., 2010; Goñi, J. R., 2007)

En general, la secuencia promotora de genes eucarióticos tiene las siguientes partes (Guglielmo y cols., 2016, Haberle y cols. 2018):

1- Núcleo del promotor o promotor mínimo, que en células animales está localizado entre -35 y -10 pb del sitio de inicio de la transcripción (+1 o TSS, por sus siglas en inglés), incluye los elementos necesarios para el inicio de la transcripción, entre los cuales están el elemento BRE (siglas en inglés de **B** recognition element) que es reconocido por el factor de transcripción TFIIB, el elemento MTE (siglas en inglés de **M**otif **T**en **E**lement), la caja TATA o Hogness, la caja CCAAT, entre otros (Wang y cols., 2006). Algunos promotores pueden no contener la caja TATA, es decir son “TATA-less” (Zuo y cols., 2011).

2- Secuencias reguladoras proximales al núcleo del promotor, a las que se unen factores de transcripción que promueven el ensamblaje del complejo de iniciación de la transcripción.

3- Secuencias reguladoras distales, las cuales modulan la actividad del núcleo del promotor y pueden ser de dos tipos (Liu y Stewart Jr. 2016):

a) Potenciadores (del inglés “enhancers”), son secuencias de acción en *cis* o *trans* que pueden ser distales o proximales al núcleo del promotor, cuya función es aumentar la tasa de transcripción

b) Inhibidores (del inglés “inhibitors”), son secuencias de acción en *cis* o *trans* a las que se les unen proteínas cuya función es la represión o reducción del nivel de transcripción.

La caracterización de la actividad de promotores permitió la identificación de una serie de elementos de regulación, cuyas secuencias han sido registradas en bases de datos. Por lo que se han desarrollado herramientas bioinformáticas que utilizan estrategias de búsqueda de secuencias para facilitar la identificación de posibles elementos de regulación en secuencias de promotor (Bailey y cols., 2009).

1.3.2 Tipos de promotores

Dependiendo de su actividad, los promotores se clasifican en (Dey y cols., 2015):

- Promotores constitutivos, cuya actividad es continua en la mayoría de los tejidos del organismo e independiente de mecanismos de regulación de la transcripción. Dependiendo del promotor, su actividad puede ser alta o baja.
- Promotores inducibles, son activados en respuesta a estímulos específicos como las hormonas, glucosa, estrés, metales pesados, entre otros.
- Promotores específicos de tejido o estadio, promueven la transcripción en tejidos particulares o en ciertos estadios de desarrollo del organismo
- Promotores sintéticos, estos promotores no existen en la naturaleza, sino que son diseñados con el propósito de que la actividad del núcleo del promotor esté bajo el control de un elemento específico y acorde al tipo de expresión que se busca obtener en respuesta a estímulo o una condición espacio-temporal. (Guglielmo y cols., 2016). El diseño de promotores sintéticos puede incluir dos o más elementos de regulación con el fin de que sean activos en respuesta a una combinación de estímulos o condiciones, a estos promotores se les conoce como promotores quiméricos (Venter y cols., 2007).

1.3.3 Promotor constitutivo 35S

El promotor constitutivo conocido como 35S (prom35S) proviene del virus del mosaico de la coliflor o CaMV (por siglas en inglés de Cauliflower mosaic virus; Benfey y cols., 1989, 1990ab; Bhullar y cols., 2007), de la familia *Caulimoviridae*. El genoma de este pararetrovirus está compuesto de 8,024 pares de bases (pb) de ADN cadena doble circular, y codifica para 6 marcos abiertos de lectura (ORFs). La transcripción se lleva a cabo a partir de dos promotores localizados en regiones intergénicas, y genera un ARN mensajero monocistrónico (19S) y un policistrónico (35S) que comprende 5 de los ORFs (Covey y Hull, 1981). El promotor 35S corresponde al fragmento de 343 pb río arriba del SIT del ARN mensajero 35S (Odell y cols., 1985) y presenta una eficiente actividad promotora de la transcripción en embrión, tallo, raíz y flor de *A. thaliana*, tabaco, algodón, entre otras dicotiledóneas (Odell y cols., 1985; Benfey y cols., 1989, 1990a, 1990b; Bhullar y cols., 2007). La organización funcional de este promotor fue descifrada mediante análisis de la actividad de fragmentos generados por eliminación seriada de secuencias, iniciando por el extremo 5' (en inglés 5'-end deletions) del promotor. La actividad de cada fragmento fue estimada por fusión transcripcional con la secuencia codificante de los reporteros luciferasa (LUC) de luciérnaga, o beta-glucoronidasa (GUS) de *E. coli* (Benfey y cols., 1989, 1990a, 1990b; Bhullar y cols., 2007). El análisis mostró la organización

modular del promotor, que comprende la región A (-90 a +1) o núcleo del promotor, la cual contiene las cajas TATA y CCAAT, y la región B (-343 a -90), la cual a su vez está dividida en 5 subdominios: B1 (-108 a -90), B2 (-155 a -108), B3 (-208 a -155), B4 (-301 to -208) y B5 (-343 to -301). La combinación de los subdominios de la región B le confiere a la región A una actividad promotora sinérgica en la variedad de tejidos vegetales analizados (Benfey y cols., 1989, 1990a, 1990b; Bhullar y cols., 2007).

El uso del promotor 35S, en combinación con la generación de plantas transgénicas por transformación mediada por *A. tumefaciens* o *A. rhizogenes*, han abierto las puertas a la expresión heteróloga de proteínas en un sinnúmero de proyectos de investigación básica, en el diseño de promotores sintéticos específicos de plantas y en el desarrollo de la agrobiotecnología (Dey y cols., 2015; Liu y Stewart Jr, 2016; Aysha y cols., 2018; Amack y Antunes, 2020).

2. ANTECEDENTES

2.1 Proteínas de la familia SNARE

Las membranas biológicas tienen varias funciones, como: separar el contenido celular del exterior, permitir la compartimentalización dentro de la célula, además también están involucradas en varios eventos celulares como son la señalización celular, el transporte de iones, la endocitosis y exocitosis, etc. Algunos de estos eventos involucran etapas de tráfico y fusión de vesículas para el transporte de lípidos y proteínas, lo que implica fusión de membranas. Este es un proceso en el que dos bicapas lipídicas convergen en una, y se logra mediante la aproximación de la membrana donadora y la que acepta. No obstante, las fuerzas de repulsión electrostática entre dos membranas con la misma carga son energéticamente desfavorables, por lo que la fusión de membranas requiere la participación de una maquinaria de proteínas entre las que se encuentran las denominadas SNARE (**S**oluble NSF **S**ensitive **F**actor **A**ttachment **P**rotein) (Han y cols., 2017).

Las proteínas SNARE conforman una super familia que contiene un dominio evolutivamente conservado de 60 a 70 aminoácidos que consiste en repetidos de siete residuos, con estructura alfa-hélice (dominio SNARE) que favorece la interacción “coil-coil” entre proteínas SNARE. Algunas SNARE presentan un dominio transmembranal (TMD, por sus siglas en inglés). La interacción “coil-coil” entre las proteínas SNARE de la vesícula donadora y la vesícula aceptora facilita la actividad fusogénica entre las membranas. (Laufman y cols., 2013).

Las proteínas SNARE pueden ser clasificadas bajo dos criterios: su localización y sus dominios estructurales, y por la presencia de residuos conservados de arginina o glutamina:

- Proteínas v-SNARE y t-SNARE. Nomenclatura originalmente asignada a estas proteínas. v-SNARE hace referencia a la proteína localizada en la membrana de la vesícula donadora (“vesicle”), mientras que t-SNARE es la localizada en la membrana aceptora (“target”). (Bonifacino y cols., 2004)
- Proteínas Q-SNARE y R-SNARE. Es una reclasificación que toma en cuenta la organización del extremo N-terminal y la presencia de un residuo conservado que puede ser arginina (R) o glutamina (Q). Además, las Q-SNARE pueden ser divididas en tres subgrupos: Qa (syntaxina 1), Qb y Qc. Las SNARE del tipo syntaxina pertenecen al subgrupo Qa y se caracterizan por tener en el extremo N-terminal tres α -hélices (Habc) separadas del dominio SNARE por una secuencia conectora (Hong y cols., 2005)

2.2 La syntaxina KNOLLE, específica de citocinesis en células vegetales

La syntaxina tipo Q-SNARE, conocida como KNOLLE o SYP111 está codificada por el gen KNOLLE, vocablo que en alemán significa tubérculo, que describe el fenotipo de la plántula de la mutante de *Arabidopsis thaliana* (Mayer y cols., 1991). Las mutantes *knolle* presentan una citocinesis deficiente, con células alargadas y paredes celulares incompletas, así como la acumulación de vesículas sin fusionar (Mayer y cols., 1991; Touihri y cols, 2011).

Al identificar y caracterizar el gen *KNOLLE*, se determinó que este gen está involucrado en procesos de fusión de vesículas durante la formación de la placa celular en la citocinesis vegetal. Además, se encontró que el gen *KNOLLE* se transcribe de manera dependiente del ciclo celular, específicamente en la interfase G2/M, mientras que la proteína se traduce en la fase M. Así mismo se demostró que los anticuerpos anti-*AtKNOLLE* inmunodetectan esta syntaxina asociada a la placa celular, pero es indetectable al concluir la citocinesis (Lukowitz y cols, 1996; Lauber y cols, 1997; Touihri y cols, 2011).

En el grupo de investigación donde se desarrolló este proyecto, se utiliza *KNOLLE* como marcador molecular para visualizar la citocinesis y estudiar la dinámica entre la formación del HI y la división celular en el sitio de infección con rizobia (proyecto de doctorado de la M.C. Elizabeth Monroy Morales, en proceso). Para tal propósito se diseñó el cassette de expresión YFP-PvKNOLLE bajo el control del promotor endógeno *promPvKNOLLE*, previamente caracterizado (Tesis de licenciatura de Emmanuel Guzmán Ayala, 2015). Ya que este promotor es el tema de la presente tesis, a continuación, se recapitulan las características de su actividad.

2.3 El promotor del gen *PvKNOLLE*

El análisis bioinformático del genoma de *P. vulgaris* en la plataforma Phytozome (The Plant Genomics Resource) permitió identificar el gen *KNOLLE* (Phvul.004G077900) y obtener información sobre su organización y secuencias del transcrito maduro y la proteína *KNOLLE*. Si bien en Phytozome el TSS no está indicado, se asume que corresponde al primer nucleótido del 5'UTR.

Con el propósito de caracterizar el promotor de *PvKNOLLE*, se amplificó un fragmento de 1.8 kb río arriba del SIT del gen *PvKNOLLE*, es decir, incluye los 140 pb de la región 5'UTR, y se clonó en fusión transcripcional con los marcadores GFP-GUS para generar la construcción *pBGFWS7-promPvKNOLLE::5'UTR_PvKNOLLE-GFP-GUS*. El patrón espacio-temporal de la actividad del promotor fue caracterizado en raíces transgénicas de *P. vulgaris* pre- y post-inoculación con *R. tropici*. Este estudio demostró que el promotor es activo específicamente en células en división en el

meristemo apical de la raíz, así como en células meristemáticas durante la organogénesis del nódulo (Tesis de licenciatura Emmanuel Ayala Guzmán, 2015). Por otro lado, el análisis de la secuencia de promPvKNOLLE reveló la presencia de tres copias del elemento tipo MSA en un fragmento de 86 pb proximal al TSS.

2.4 Motivo MSA

El motivo MSA (siglas en inglés de Mitosis Specific Activator) ha sido reportado como un elemento de regulación en *cis*, específico de genes expresados en la fase G2/M del ciclo celular en células vegetales. Este motivo fue identificado al caracterizar la actividad del promotor del gen de la ciclina CycB1 (CYM) de *Catharanthus roseus* (Ito y cols., 1998). En el análisis se utilizaron versiones cortas del promotor, generadas por sustitución seriada de pares de bases o eliminación seriada de fragmentos de promotor. La actividad de cada versión de promotor fue determinada en muestras de células de tabaco BY2 cosechadas en las diferentes etapas del ciclo celular. Por un lado, este estudio demostró que el fragmento de 203 pb localizado en el extremo 3' del promotor CYM es suficiente para conferir la actividad de promotor específica de la fase G2/M en células vegetales (Ito y cols., 1998). Por otro lado, se encontró que este fragmento presenta una región con cuatro repetidos de 11 nt, cuya secuencia consenso es (T/C)C(T/C)AACGG(T/C)(T/C)A, que al ser insertados en tándem en el extremo 3' del promotor constitutivo 35S, le confiere una actividad específica en la fase G2/M, similar a la observada en el promotor CYM, de donde surgió la nomenclatura MSA (Ito y cols., 1998). La búsqueda sistemática de elementos de regulación en genes de ciclo celular ha demostrado que el motivo MSA está presente en el promotor de aproximadamente 80 genes asociados a ciclo celular en *A. thaliana* y *Nicotiana tabacum*, entre los que se encuentran los genes de ciclina tipo B, cinesinas, KNOLLE, proteína asociada a microtúbulos, cinasas tipo Aurora, entre otros (Ito y cols., 1998; Ito, 2005 Menges y cols., 2005; Kato y cols., 2009). Otro dato interesante es la similitud entre la secuencia del motivo MSA y el sitio de reconocimiento de los factores de transcripción tipo MYB (MYB3R1 y MYB3R4 en *A. thaliana* y NtmybA1, NtmybA2 y NtmybB en *N. bentamiana*) cuya secuencia consenso es (A/T/C)N(T/C)AACGG(T/C)(T/C)AN(A/T)(A/T)T, que conserva el pentámero invariable también encontrado en la secuencia MSA. (Haga y cols., 2011; Ito y cols., 2001). Se postula que los ortólogos de NtmybA1 y NtmybA2 son reguladores positivos de la función de los motivos MSA, mientras que NtmybB tiene actividad de represor (Ito y cols., 2001; Haga y cols., 2007; Haga y cols., 2011). El este trabajo de tesis se plantea el diseño de un promotor sintético cuyo único elemento de regulación en *cis* es el motivo MSA y la búsqueda bioinformática de genes cuyo promotor presente motivos MSA.

3. JUSTIFICACIÓN

Uno de los aspectos relevantes en la caracterización de un promotor es determinar la relación funcional entre los elementos de regulación presentes en su secuencia y su actividad espacio-temporal. En el caso de los promotores KNOLLE de *A. thaliana* y *P. vulgaris*, su actividad sólo ha sido caracterizada en el contexto de especificidad en células en división, la cual parece estar regulada por el elemento en *cis*, el motivo MSA. No se excluye que otras regiones del promotor regulen aspectos relacionados con la función del gen KNOLLE.

La especificidad funcional del motivo MSA es de particular interés si el propósito es dirigir la expresión heteróloga de marcadores moleculares o RNAi, específicamente en células mitóticas. Tal es el caso de proyectos de nuestro grupo, enfocados en el estudio de la dinámica entre la división celular y el avance del hilo de infección en raíces de *P. vulgaris* inoculadas con rizobia. En este contexto, un promotor sintético tipo prom35S-MSA pudiera ser la herramienta molecular que cumple con este requisito.

Por otro lado, el gen PvKNOLLE nos ha permitido estudiar la dinámica entre citocinesis y el avance del HI. Como siguiente paso, nos interesa extender este estudio a las etapas pre-citocinéticas del ciclo celular. Con este propósito, otra de las estrategias planteadas en este trabajo combinó el uso de herramientas bioinformáticas y la secuencia consenso del motivo MSA para identificar genes potencialmente específicos de la interfase G2/M del ciclo celular, cuya función pudiera favorecer su uso como marcadores moleculares específicos de las fases de la mitosis.

4. HIPÓTESIS

- 1) La inserción de los motivos MSA en el extremo 5' proximal del promotor constitutivo 35S le permitirá a este promotor sintético tener una actividad específica en células mitóticas.

- 2) La búsqueda bioinformática de motivos MSA en la secuencia río arriba del SIT en genes de *A. thaliana* y *P. vulgaris* favorecerá la identificación de genes específicos de la interfase G2/M del ciclo celular.

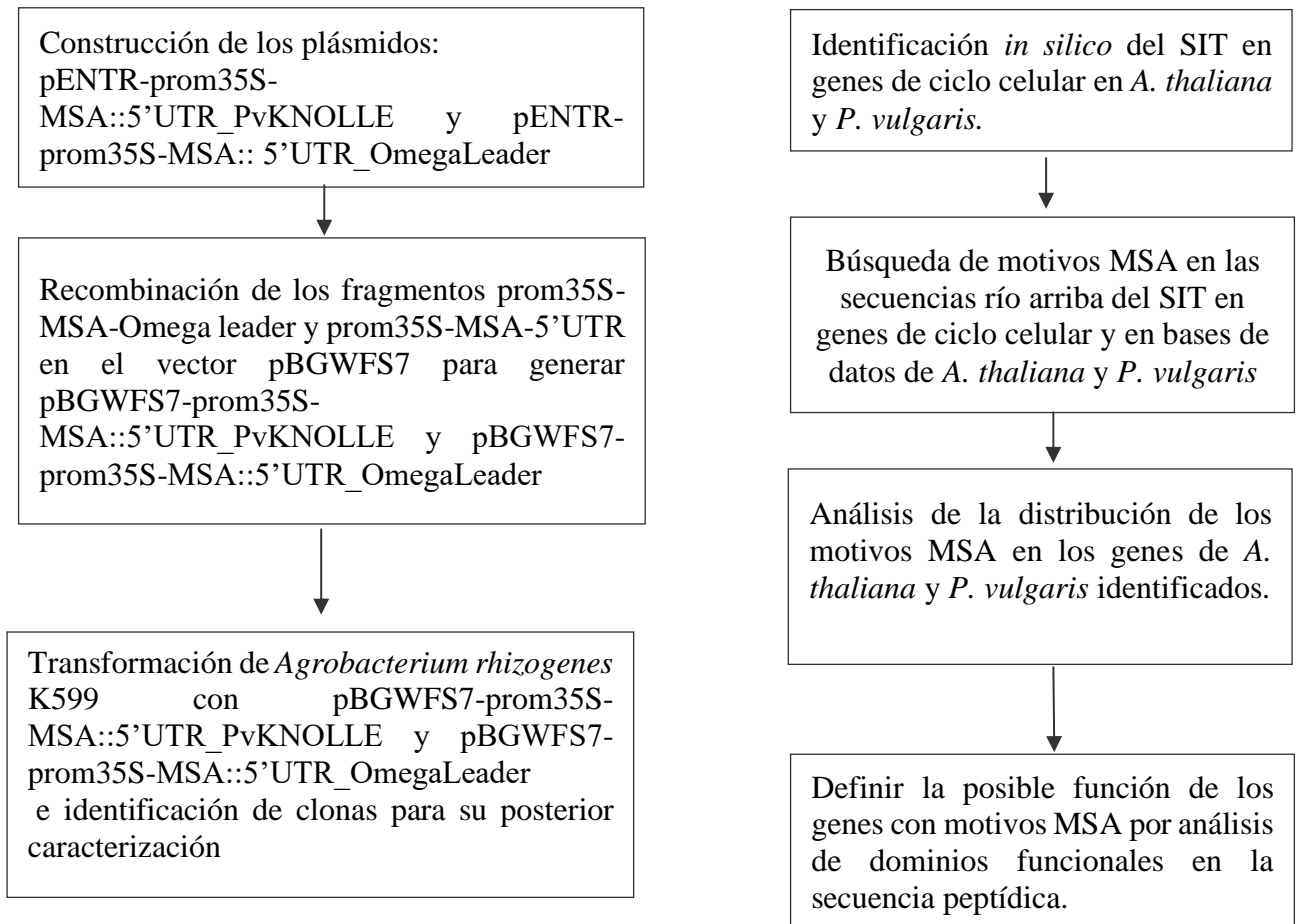
5. OBJETIVO GENERAL

Diseñar un promotor sintético tipo prom35S-MSA con actividad en células vegetales en mitosis e identificar genes de *A. thaliana* y *P. vulgaris* cuyo promotor tiene motivos MSA.

5.1 OBJETIVOS PARTICULARES

- 1.- Construir los cassettes prom35S-MSA::5'UTR_PvKNOLLE y prom35S-MSA::5'UTR_Omega Leader, y clonarlos en fusión transcripcional con el marcador quimérico GFP-GUS en el vector de expresión en plantas pBGWFS7, para generar los plásmidos pBGWFS7-prom35S-MSA::5'UTR_PvKNOLLE y pBGWFS7-prom35S-MSA::5'UTR_OmegaLeader.
2. Generar clonas de *A. rhizogenes* K599 que porten pBGWFS7-prom35S-MSA::5'UTR_PvKNOLLE o pBGWFS7-prom35S-MSA::5'UTR_OmegaLeader.
3. Identificar genes de *A. thaliana* y *P. vulgaris* cuyo promotor presenta motivos MSA
4. Analizar la distribución de los motivos MSA en los promotores identificados.
5. Definir la posible función de los genes identificados

6. ESTRATEGIA EXPERIMENTAL



7. MATERIALES Y MÉTODOS

7.1 Plásmidos disponibles en el laboratorio

Los plásmidos pENTR-prom35S-5'UTR_OmegaLeader (3,044 pb), pENTR-XhoI-MSA-XhoI (2,677 pb) y pENTR-XhoI-MSA-5'UTR_PvKNOLLE-AscI (2,831 pb) fueron construidos en un trabajo previo (Leopoldo Carrera y Rosana Sánchez, resultados no publicados). En pENTR-prom35S-5'UTR_OmegaLeader, la secuencia del promotor 35S está en fusión transcripcional con la secuencia 5' líder del virus del mosaico del tabaco (Gallie y cols., 1988). El plásmido pENTR-XhoI-MSA-XhoI porta un fragmento de 97 pb que corresponde a la región con tres repeticiones MSA del promotor PvKNOLLE (86 pb) flanqueada por sitios XhoI, cuya secuencia se incluyó en los oligonucleótidos específicos utilizados para su amplificación por PCR. En pENTR-XhoI-MSA-5'UTR_PvKNOLLE-AscI, fragmento MSA está en fusión transcripcional con la secuencia 5'UTR de PvKNOLLE. El vector pBGWFS7 permite la inserción y caracterización de la región promotora de un gen en fusión transcripcional con secuencia codificante de la proteína quimérica GFP-GUS. La actividad del promotor se revela mediante las actividades fluorescente (GFP) y enzimática de beta-glucoronidasa (GUS, que hidroliza sustrato cromogénico).

7.2 Reacciones con enzimas de restricción, fosfatasa y ligasa

Para la escisión de fragmentos y la preparación del plásmido aceptor (con base pENTR) en la clonación correspondiente, se utilizaron enzimas de restricción de la gama FastDigest (ThermoScientific) y la fosfatasa alcalina FastAP Thermosensitive Alkaline Phosphatase (ThermoScientific), cuyas condiciones de incubación son compatibles (buffer FastDigest y temperatura; ThermoScientific). Para la etapa de ligación, se utilizó el kit comercial Rapid DNA Ligation Kit (ThermoScientific). Para la mezcla de ligación se utilizó una relación molar 3:1 (inserto:vector) en un volumen final de 20 μ l y un tiempo de reacción de 5 min a 22 °C. Se tomaron 5-10 μ l de mezcla de ligación para transformar en *E. coli* TOP 10 quimiocompetentes. La identificación de clones de interés se realizó por ensayos de PCR en colonia.

7.3 Reacción de clonación por recombinación

La clonación de insertos en el vector pBGWFS7 se realizó por recombinación utilizando la mezcla enzimática LR Clonase II de la metodología Gateway (Invitrogen). La mezcla de recombinación fue transformada en *E. coli* TOP10 quimiocompetentes, con selección en medio LB con espectinomicina 200 µg/ml. La selección de clonas positivas se realizó por el método de PCR en colonia.

7.4 Transformación en *E. coli* TOP10 quimiocompetentes

Se mezcló 5-10 µl reacción de ligación con 50 µl de bacterias quimiocompetentes disponibles en el laboratorio. La mezcla se dejó 30 min en hielo y después se dio un choque térmico a 42°C/ por 45 segundos. Se agregaron 800 µl de medio LB líquido y se incubó a 37°C durante 1 hora. Se sembraron 100 o 900 µl de la suspensión bacteriana en cajas Petri con medio LB sólido con kanamicina (100 µg/ml). La incubación se realizó a 37°C durante 16 horas.

7.5 PCR en colonia

Las colonias analizadas fueron resuspendidas en 50 µl de agua tetradestilada estéril y lisadas mediante 3 ciclos de congelación (-80°C) y ebullición (100°C) seguidas de vórtex. El lisado se centrifugó a 14000 rpm y se recupera el sobrenadante. La mezcla de reacción de PCR contiene: dNTP's 0.2 mM, el par de oligonucleótidos gene-específicos (0.4 µM), buffer 5X, 1.5 U de Taq DNA Polimerasa Recombinante (Thermo Scientific) y 1 µl del sobrenadante (templado). Las condiciones de reacción son: un ciclo de 30 seg a 98°C, 35 ciclos a 98°C/10 seg, 55°C/15 seg, 72 °C/45-60 seg. Los productos de PCR fueron analizados por electroforesis en gel de agarosa 1%. Para la extracción de plásmido de las clonas identificadas como positivas se utilizó el kit comercial ZymoPure II Plasmid miniprep kit. Para confirmar la identidad del inserto, el plásmido purificado fue enviado a la Unidad de Síntesis y Secuenciación del IBT-UNAM para su secuenciación.

7.6 Transformación en *A. rhizogenes* K599 electrocompetentes

Para transformar *A. rhizogenes* K599, se mezcló 50-100 ng de plásmidos basados en pBgWFS7 con 50 µl de bacterias electrocompetentes. La mezcla se dejó 10 min en hielo, en seguida se pasó a la celda de electroporación y se sometió a un campo eléctrico utilizando un electroporador. Posteriormente se pasaron las bacterias a 900 µl de medio LB y se incubaron

a 30°C con agitación por 2 horas, para luego plaquear la suspensión en medio LB sólido con espectinomicina 200µg/mL e incubar a 30°C durante 48 horas. La identificación de clonas positivas se hizo por PCR en colonia y secuenciación.

7.7 Preparación de *A. rhizogenes* K599 electrocompetentes

Se partió de un precultivo de 30 ml en medio LB líquido (sin antibiótico) inoculando con una asada de *A. rhizogenes* K599 (cepa preservada a -70°C en glicerol al 40%) e incubando por aprox 16h a 30°C con agitación. Se tomaron 3 ml del precultivo para inocular 30 ml de medio LB líquido (sin antibiótico) y se incubó a 30°C con agitación hasta obtener una DO₆₀₀ de 0.5-0.6. Se pasó 1.4 ml de cultivo a un tubo Eppendorf de 1.5 ml estériles y se centrifugó a 9,000 rpm durante 1 min a 24 °C. Se descartó el sobrenadante y las bacterias se resuspendieron en 1ml de H₂O tetradestilada estéril (a 24 °C). La suspensión bacteriana se centrifugó a 9,000 rpm durante 1 min a 24 °C. Este paso se repitió una vez.

Por último, las bacterias se resuspendieron suavemente en 50 µl de glicerol al 10% estéril (a 24 °C), se etiqueta el tubo y guarda a -70°C.

7.8 Herramientas bioinformáticas para la búsqueda de motivos MSA

La búsqueda de secuencias de *A. thaliana* y *P. vulgaris* se realizó en el portal Phytozome. Las secuencias fueron convertidas al formato FASTA utilizando la herramienta en línea EMBOSS Seqret en el portal del Instituto Europeo de Bioinformática del Laboratorio Europeo de Biología Molecular (EMBL-EBI).

A continuación, se describen las plataformas que se utilizaron para la búsqueda de motivos MSA:

- **PlantCARE (Cis Acting Regulatory Elements)**, disponible en el sitio en línea <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/> (Lescot y cols., 2002).

La opción “Search for CARE” en este portal permite la búsqueda de motivos de regulación en *cis*, anotados en una base de datos del sitio y conformada por motivos reportados en la literatura, incluyendo el motivo MSA (reportado como MSA-like). Las secuencias a analizar son proporcionadas en formato FASTA vía una ventana de captura. La plataforma envía correo electrónico con los resultados en un archivo HTML en el que los motivos encontrados se ilustran en la Figura 5:

Motifs Found


























- + 
- +  A-box
- +  AAGAA-motif
- +  ABRE
- +  ABRE2
- +  AE-box
- +  ARE
- +  AT~TATA-box
- +  Box 4
- +  CAAT-box
- +  CCAAT-box
- +  CCGTCC motif
- +  CCGTCC-box
- +  CGTCA-motif
- +  DRE core
- +  E2Fb
- +  ERE
- +  G-Box
- +  G-box
- +  GATA-motif
- +  GT1-motif
- +  I-box
- +  LAMP-element
- +  MSA-like 

Figura 5. Representación de los resultados obtenidos por la plataforma PlantCARE.

Al presionar el signo “+” en el motivo de interés se despliega la información correspondiente, como se ilustra a continuación:



Site name	Organism	Position	Strand	Matrix score	Sequence	Function
MSA-like	<i>Catharanthus roseus</i>	1922	-	9	(T/C)C(T/C)AACGG(T/C) (T/C)A	Cis-acting element involved in cell cycle regulation

- MAST (Motif Alignment & Search Tool) disponible en el sitio en línea <http://memesuite.org/tools/mast> (Bailey y Gribskov, 1998)

Es una de las herramientas de búsqueda de motivos (“motif scanning”) incluidas en el portal MEME-suite (<http://meme-suite.org/>). MAST realiza la búsqueda de uno o más motivos (secuencia de nucleótidos o aminoácidos) en una secuencia de interés, o bien en una base de datos de secuencias. La plataforma permite que los motivos sean proporcionados en formato

tipo MEME Minimal Motif Format (formato texto en ASCII) y lo convierte en formato MEME Motif Format (matriz de probabilidad dependiente de posición), un requisito para procesar la búsqueda. MAST calcula la probabilidad de coincidencia de cada posición en la secuencia de interés respecto a la posición en un motivo (valor p), así como el valor de probabilidad p de coincidencia combinada en un conjunto de motivos que se describen colectivamente. Como referencia MAST también reporta la probabilidad de coincidencia de cada posición en un grupo de motivos encontrada en las secuencias de una base de datos (valor E) y corresponde al producto del valor p combinado (que considera la probabilidad conjunta de encontrar más de un motivo) por el número de secuencias en la base de datos. Los resultados se clasifican de acuerdo con el valor E , resultando en la identificación de las secuencias con la mejor coincidencia al motivo proporcionado. La interfase en la que se realiza la búsqueda es muy sencilla, se proporciona la secuencia del motivo a buscar (recuadro Input the motifs) y la secuencia problema (Input the sequences) o bien la base de datos donde se quiere buscar el motivo (seleccionar una de la opción que se despliegan en recuadro “Input the sequences”, y el genoma del organismo de interés). A continuación, se ilustran ambas opciones (Figura 6 y 7)

MAST
Motif Alignment & Search Tool
Version 5.1.1

MAST searches sequences for matches to a set of motifs, and sorts the sequences by the **best combined match** to all motifs (sample output for motifs and sequences). See this Manual for more information.

Data Submission Form

Find sequences that match a set of motifs.

Input the motifs
Enter motifs you wish to scan with.
Type in motifs: DNA [DNA ?]
[TCYAACGGYYA]
TCYAACGGYYA

Input the sequences
Enter sequences or select the database you want to scan for matches to motifs.
Upload sequences: Seleccionar archivo positivos1a33.txt [DNA ?]

Input job details
(Optional) Enter your email address. [Ath 2kb]
(Optional) Enter a job description.

Advanced options

Note: if the combined form inputs exceed 80MB the job will be rejected.

Start Search Clear Input

Figura 6. Representación de la plataforma MAST.

Además, tiene como opción el análisis de las secuencias río arriba del sitio TSS de todos los genes de un organismo. En nuestro caso se seleccionó la base de datos “Upstream Sequences: plants”, como se ilustra a continuación (Figura 7):

The screenshot shows the MAST web interface. A dropdown menu is open, listing various databases under the heading "Databases (select category)". The options include "Ensembl Ab Initio Predicted Proteins", "Ensembl Fungi Genomes and Proteins", "Ensembl Genomes and Proteins", "Ensembl Metazoa Genomes and Proteins", "Ensembl Plant Genomes and Proteins", "Ensembl Protists Genomes and Proteins", "GenBank Bacteria Genomes and Proteins", "GenBank Fungi Genomes and Proteins", "Other Databases", "Upstream Sequences: Fungal", "Upstream Sequences: Metazoan", "Upstream Sequences: Other", "Upstream Sequences: Plant" (highlighted with a red arrow), "Upstream Sequences: Prokaryotic", "Upstream Sequences: Protist", and "Upstream Sequences: Viral". Below the dropdown, there are input fields for "Upstream Sequences: Plant", "Aegilops tauschii.ASM34733v1.37", and "Jul 12, 2018". There are also checkboxes for "DNA" and "RNA". The "Input job details" section includes fields for "Optional) Enter your email address" and "Optional) Enter a job description". At the bottom, there is a "Start Search" button and a "Clear Input" button. A note at the bottom states: "Note: if the combined form inputs exceed 80MB the job will be rejected."

Figura 7. Representación de la base de datos “Upstream Sequences: plants”

La plataforma despliega los resultados que incluyen la posición y número de motivos encontrados, la probabilidad estadística de cada uno y en qué hebra (+ o -) del ADN se encuentran, así como una representación gráfica de la posición del motivo en la secuencia analizada. Los resultados pueden ser descargados y guardados. A continuación, se ilustra el despliegue gráfico de resultados (Figura 8).

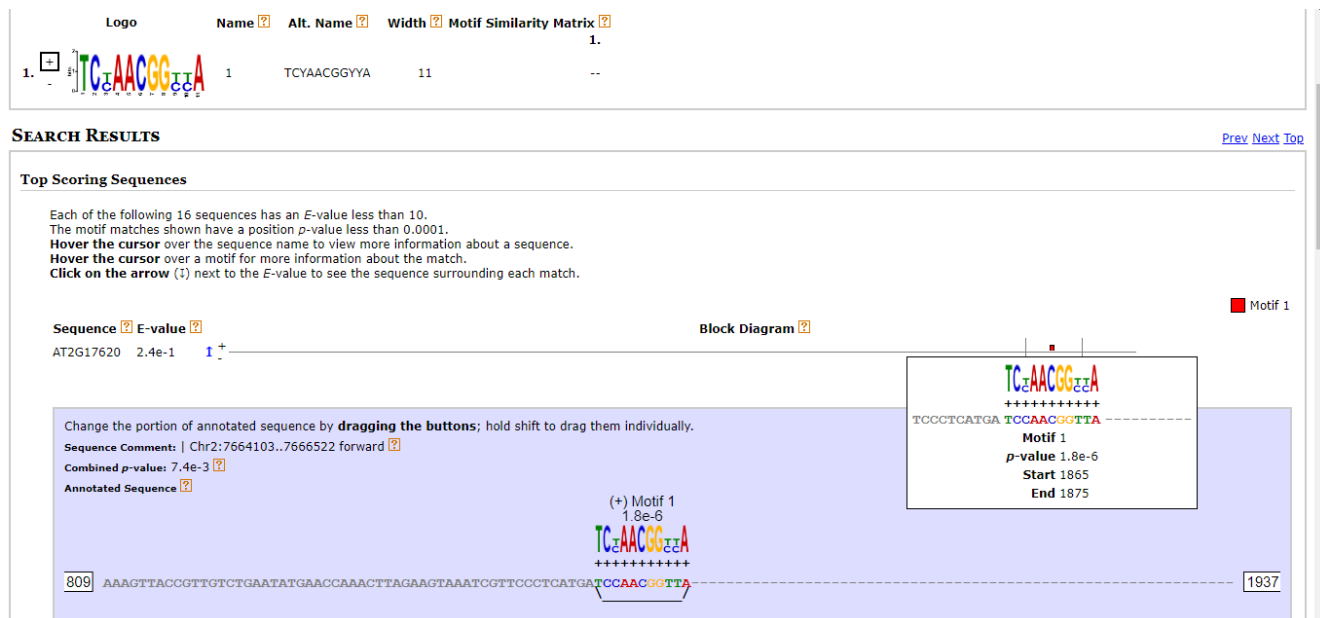


Figura 8. Representación gráfica de los resultados obtenidos por las plataformas MAST y MCAST.

- **MCAST (Motif Cluster Alignment and Search Tool)**, disponible en el sitio en línea <http://meme-suite.org/tools/mcast> (Bailey y Noble, 2003)

Es una herramienta similar a MAST, pero MCAST se caracteriza por la búsqueda de grupos de un mismo motivo (“Cluster Motif Scanning”) en una secuencia de interés, o bien en una base de datos de secuencias. Uno de los parámetros a definir en la búsqueda con MCAST es el largo máximo de separación entre cada motivo, a indicar en recuadro “Advanced options”. Los resultados son presentados en el mismo formato que en MAST.

7.9 Herramientas bioinformáticas para la búsqueda de posibles homólogos a genes de *A. thaliana* en *P. vulgaris*

Para esta búsqueda se utilizó la herramienta BLASTp (Basic Local Alignment Search Tool, por sus siglas en inglés) y la base de datos del proteoma de *P. vulgaris* en el portal Phytozome. Como template en la búsqueda (“query”) se utilizaron las secuencias en aminoácidos codificadas por los genes de *A. thaliana* identificados.

8. RESULTADOS

8.1 Diseño del promotor sintético prom35S-MSA en fusión transcripcional con una secuencia 5'UTR.

Para evaluar la función de los motivos MSA proximales al TSS de PvKNOLLE como reguladores de la actividad de promotor específica de células en división, en este trabajo se propone el diseño del promotor sintético prom35S-MSA, cuya actividad se espera sea similar a la de promPvKNOLLE reportada por Emmanuel Ayala Guzmán (2015). En este punto, es importante puntualizar que la actividad de promPvKNOLLE se caracterizó utilizando el cassette promPvKNOLLE::5'UTR_PvKNOLLE, lo que dejó abierta la posibilidad de que el 5'UTR de PvKNOLLE contribuya a la regulación de la expresión del gen. Por lo tanto, en el diseño de promotor sintético resulta pertinente realizar un análisis comparativo entre la actividad de prom35S-MSA en fusión transcripcional con la región 5'UTR de PvKNOLLE (prom35S-MSA::5'UTR_5'UTR-PvKNOLLE, véase Figura 9a), y la actividad del cassette prom35S-MSA::5'UTR_Omega-Leader (véase Figura 9b), en donde el 5'UTR_Omega-Leader corresponde al extremo 5'UTR del virus de ARN del mosaico de tabaco (TMV; Gallie y *cos.*, 1998), no contiene elementos de regulación y es suficiente para una eficiente traducción de ARNm (Gallie y *col.*, 2002). Las secuencias de prom35S y 5'UTR_Omega-Leader (367 y 68 pb, respectivamente) provienen del cassette prom35S::5'UTR_OmegaLeader presente en el vector de expresión en plantas pEarleyGate104. Como control en el análisis de especificidad de la actividad de promotor, se planteó utilizar el cassette prom35S::5'UTR_OmegaLeader, cuya actividad es constitutiva (Figura 9c).

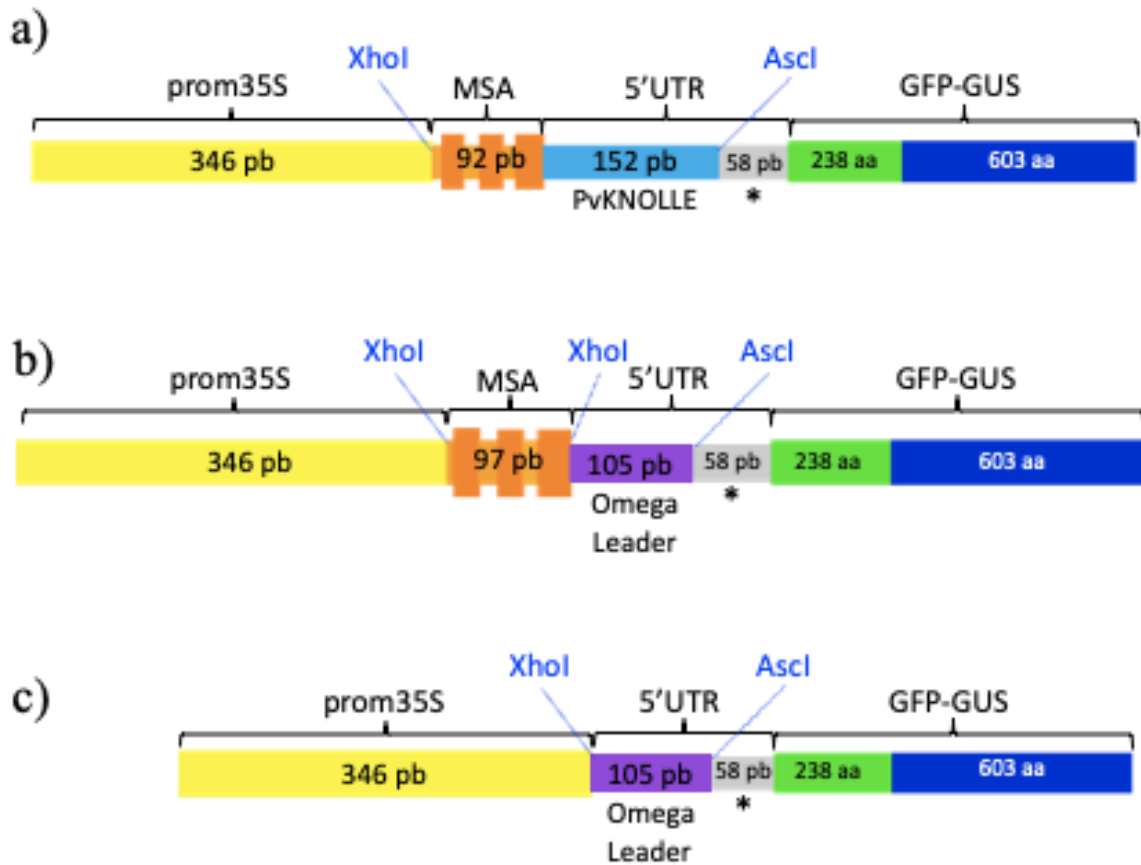


Figura 9. Representación gráfica de los cassettes diseñados en este trabajo clonados en fusión transcripcional con la secuencia codificante del reportero quimérico GFP-GUS en el vector pBGWFS7: a) prom35S-MSA::5'UTR_PvKNOLLE; c) prom35S-MSA::5'UTR_OmegaLeader y c) prom35S::5'UTR_OmegaLeader. * Inserción por reacción de clonación por recombinación

8.2 Construcción de los plásmidos pBGWFS7-prom35MSA::5'UTR_PvKNOLLE y pBGWFS7-prom35S-MSA::5'UTR_OmegaLeader

pBGWFS7 es un vector de expresión en plantas, diseñado para caracterizar la actividad del promotor de interés en fusión transcripcional con la secuencia codificante de la proteína quimérica reportera GFP-GUS. Es un vector del sistema de recombinación *in vitro* (Gateway®, Invitrogen) que utiliza la mezcla enzimática LR Clonase™ II (Invitrogen; véase Figura 10) e implica dos pasos de clonación, 1) clonación del fragmento de ADN en un vector de entrada, en este caso el vector pENTR™/D-TOPO® (Invitrogen), y 2) la recombinación

sitio específica del inserto clonado en el vector de entrada que se recombina con el vector destino, en este caso pBGWFS7.

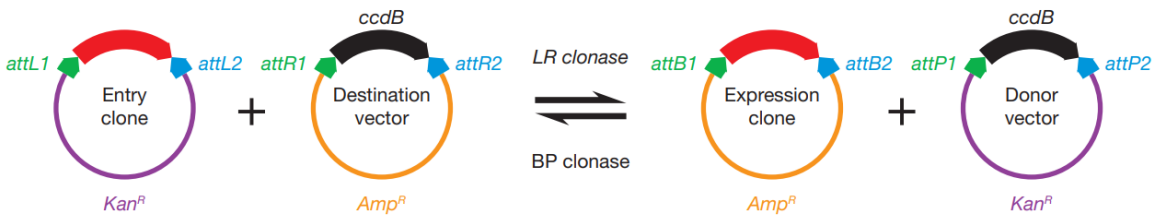


Figura 10. Representación de la recombinação *in vitro* sitio específica del sistema Gateway® (Invitrogen).

8.2.1 Construcción del plásmido pENTR-prom35S- MSA::5'UTR_PvKNOLLE

Para generar esta construcción, los plásmidos pENTR-XhoI-MSA-5'UTR_PvKNOLLE-AscI y pENTR-prom35S::5'UTR-OmegaLeader fueron digeridos con las enzimas de restricción XhoI y AscI, en presencia de fosfatasa alcalina (véase Materiales y Métodos y Figura 9). La digestión del plásmido pENTR-XhoI-MSA-5'UTR_PvKNOLLE-AscI generó dos fragmentos: uno de 2,572 y otro de 259 pb, mientras que el plásmido pENTR-prom35S::OmegaLeader produjo dos fragmentos: uno de 2,938 y otro de 106 pb. Los fragmentos de interés (259 pb y 2,938 pb, respectivamente) fueron purificados y ligados utilizando la enzima T4 ADN ligasa (Rapid DNA Ligation kit). La mezcla de ligación fue transformada en *E. coli* TOP10. Para la identificación de las clonas que portan pENTR-prom35S-MSA::5'UTR_PvKNOLLE se utilizó el método de PCR en colonia. Los resultados se presentan en la Figura 11. La secuenciación de los plásmidos de las clonas 11 y 12 confirmó la identidad de la construcción de pENTR-prom35S- MSA::5'UTR_PvKNOLLE. Para las siguientes etapas de clonación se utilizó a la clona 11.

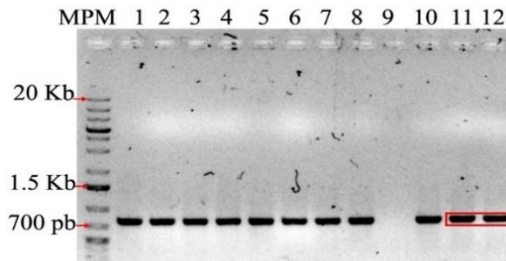


Figura 11. Análisis electroforético de PCR en colonia. Las reacciones de PCR se realizaron utilizando los oligonucleótidos pEarly 35S Fw y M13 Rv. El producto esperado es de 794 pb, que corresponde a la amplificación del fragmento prom35S MSA::5'UTR_PvKNOLLE. Carriles 1-12, producto de PCR en colonia. Los recuadros en rojo marcan las clonas seleccionadas, 11 y 12. MPM, marcador de peso molecular (GeneRuler™1 kb Plus DNA Ladder, Thermo Scientific). Como referencia se indica el tamaño de algunos fragmentos en pb.

8.2.2 Construcción del plásmido pENTR-prom35S-MSA::5'UTR_OmegaLeader

A partir del plásmido pENTR-XhoI-MSA-XhoI (véase Materiales y Métodos) se escindió el fragmento XhoI-MSA-XhoI de 97 pb por digestión con la enzima XhoI. En paralelo, el plásmido pENTR-prom35S::OmegaLeader se linearizó por digestión con XhoI en presencia de fosfatasa alcalina. Posteriormente, el fragmento XhoI-MSA-XhoI y el plásmido linearizado pENTR-prom35S::OmegaLeader fueron ligados utilizando la enzima T4 ADN ligasa (Rapid DNA Ligation kit), para generar el plásmido pENTR-prom35S-MSA::5'UTR_OmegaLeader. Para la identificación de las clonas que portan el plásmido generado se utilizó el método de PCR en colonia utilizando a los oligonucleótidos M13 Fw y M13, específicos de la secuencia pENTR™/D-TOPO®. El tamaño del producto esperado es de 880 pb. La Figura 12 muestra los resultados obtenidos. Se encontró que las clonas 2, 4, 5, 6 y 7 portan el fragmento esperado, por lo que enviaron a secuenciación. Los resultados mostraron que las clonas 2, 5 y 6 tienen el cassette prom35S-MSA::5'UTR_OmegaLeader, en el cual el fragmento MSA tiene la orientación correcta.

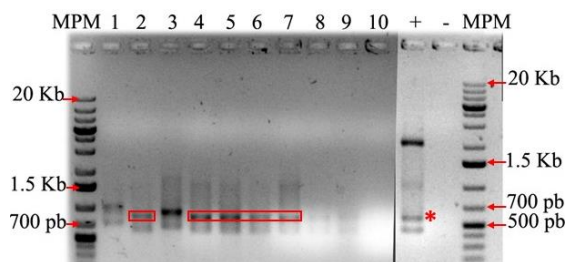


Figura 12. Análisis electroforético de productos de PCR en colonia para identificar las clonas que portan pENTR-prom35S-MSA::5'UTR_OmegaLeader. Carriles 1-10, productos de PCR en colonia de clonas 1-10. El tamaño del producto esperado es de 880 pb. Carril +, producto de PCR utilizando pENTR-XhoI-MSA-5'UTR_PvKNOLLE como templado (amplificación de fragmento de 575 pb, indicado con *). Todas las reacciones de PCR se realizaron utilizando los oligonucleótidos M13 Fw y M13 Rv. Los recuadros en rojo marcan las clonas seleccionadas y enviadas para secuenciación: 2, 4-7. MPM, marcador de peso molecular (GeneRuler™1 kb Plus DNA Ladder, Thermo Scientific), como referencia se indica el tamaño de algunos fragmentos en pb.

8.2.3 Recombinación de prom35S-MSA::5'UTR_PvKNOLLE, prom35S-MSA::5'UTR_OmegaLeader y prom35S::5'UTR_OmegaLeader en pBGWFS7

La mezcla de recombinación entre el inserto en el plásmido donador pENTR-prom35S-MSA::5'UTR_PvKNOLLE (clona 11), pENTR-prom35S-MSA::5'UTR_OmegaLeader (clona 5) o pENTR-prom35S::5'UTR_OmegaLeader (clona 3) y el plásmido aceptor pBGWFS7 fue transformada *E. coli* DH5 y las colonias resultantes fueron analizadas por PCR en colonia utilizando los pares oligonucleótidos específicos indicados en la Figura 13

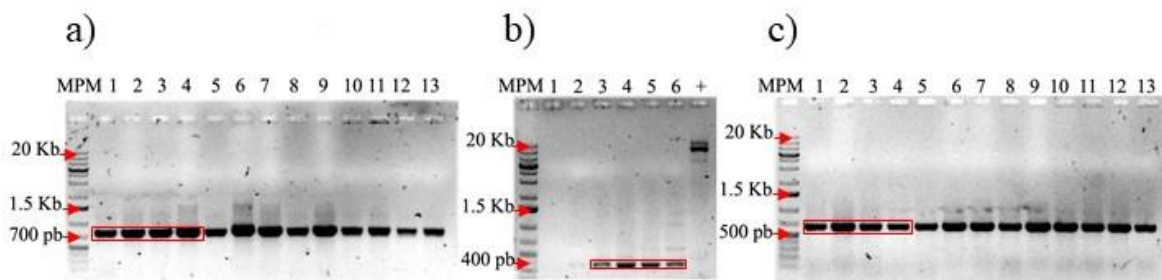


Figura 13. Análisis electroforético de PCR en colonia para identificar las clonas de *E. coli* que portan las construcciones **a**, pBGWFS7-prom35S-MSA::5'UTR_PvKNOLLE; **b**, pBGWFS7-prom35S-MSA::5'UTR_OmegaLeader; y **c**, pBGWFS7-prom35S::5'UTR_OmegaLeader. **a** y **c**) Carriles 1-13, productos de PCR en colonia. **b**) Carriles 1-7, productos de PCR en colonia. Los productos de amplificación esperados son de 803 pb (**a**), 371 pb (**b**) y 649 pb (**c**). Las reacciones de PCR se realizaron usando los pares de oligonucleótidos pEarley 35S Fw y GFP3 Rv (**a**, **c**) o MSA Fw y GFP3 Rv (**b**). Los recuadros en rojo marcan las clonas seleccionadas y secuenciadas. MPM, marcador de peso molecular (GeneRuler™1 kb Plus DNA Ladder, Thermo Scientific), como referencia se indica el tamaño de algunos fragmentos en pb.

El análisis por PCR en colonia mostró que la eficiencia de la recombinación fue exitosa (ver Figura 13) y la secuenciación confirmó que los plásmidos pBGFWS7-prom35S-MSA::5'UTR_PvKNOLLE (clonas 1-4), pBGFWS7-prom35S-MSA::5'UTR_OmegaLeader (clonas 3-4) y pBGFWS7-prom35S::5'UTR-OmegaLeader (clonas 1-4) tienen la inserción y la secuencia esperada.

8.2.4 Electroporación de los plásmidos pBGFWS7-prom35S-MSA::5'UTR_PvKNOLLE, pBGFWS7-prom35S-MSA::5'UTR_OmegaLeader o pBGFWS7-prom35S::5'UTR_OmegaLeader en *A. rhizogenes* K599

Los plásmidos electroporados fueron pBGFWS7-prom35S-MSA::5'UTR_PvKNOLLE (clonas 1 y 4), pBGFWS7-prom35S-MSA::5'UTR_OmegaLeader clonas 3 y 4 y pBGFWS7-prom35S::5'UTR_OmegaLeader (clona 1). La Figura 14 muestra el análisis de PCR en colonia de las clonas seleccionadas. La nomenclatura asignada a las clonas inicia con el nombre del plásmido, un primer dígito que corresponde al número de la clona *E. coli*, a partir de la que se aisló el plásmido a electroporar y un segundo dígito que indica el número de la clona en *A. rhizogenes* K599.

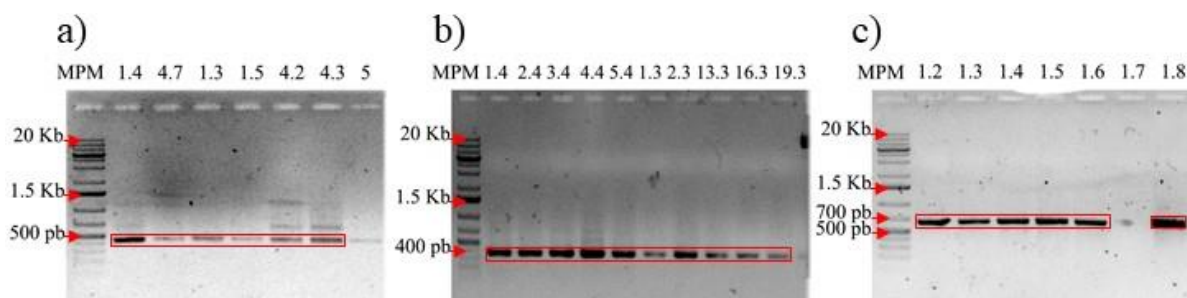


Figura 14. Análisis electroforético de PCR en colonia para identificar las clonas de *A. rhizogenes* K599 que portan: a) plásmido pBGFWS7-prom35S-MSA::5'UTR_PvKNOLLE; b) pBGFWS7-prom35S-MSA::5'UTR_OmegaLeader; y c) pBGFWS7-prom35S::5'UTR_OmegaLeader. Las reacciones de PCR se realizaron usando los pares de oligonucleótidos MSA Fw GFP3 Rv (a, b), pEarly 35S Fw y GFP3 Rv (c). Los productos de amplificación esperados son de 427 pb (a), 371 pb (b) y 649 pb (c). Carriles: clonas analizadas. (+) Control positivo, reacción de PCR utilizando como templado el plásmido pBGFWS7 35 S MSA 5'UTR_PvKNOLLE (a), pBGFWS7-prom35S-MSA::5'UTR_OmegaLeader (b) o pBGFWS7-prom35S::5'UTR_OmegaLeader (c). (-) Control negativo: reacción en ausencia de templado. MPM, marcador de peso molecular (GeneRuler™ 1 kb Plus DNA Ladder, Thermo Scientific). Como referencia se indica el tamaño de algunos fragmentos en pb.

Para determinar cuáles de las clonas en *A. rhizogenes* K599 tienen la capacidad de generar raíces transgénicas, se seleccionaron las siguientes:

- *A. rhizogenes* 599 pBGFWS7-prom35S::MSA:5'UTR_PvKNOLLE clonas 1.2, 1.3, 1.4, 1.5, 4.1, 4.2, 4.3 y 4.7
- *A. rhizogenes* 599 pBGFWS7-prom35S::MSA:5'UTR_OmegaLeader clonas 3.1, 3.2, 3.13, 3.16, 3.19, 4.1, 4.2, 4.3, 4.4 y 4.5
- *A. rhizogenes* 599 pBGFWS7-prom35S::5'UTR_OmegaLeader clonas 1.2, 1.3, 1.4, 1.5, 1.6 y 1.8.

8.3 Búsqueda de genes del ciclo celular en el genoma de *P. vulgaris*, cuyos promotores presenten motivos MSA

8.3.1 Búsqueda de genes de ciclo celular en *A. thaliana* e identificación de motivos MSA en secuencias de promotor

Para identificar motivos MSA en los promotores de los genes de *P. vulgaris*, se utilizó como referencia los 155 genes descritos por Menges y colaboradores (2005). Estos autores describen un análisis por microarreglos de muestras de ARN aislado a partir de células de *A. thaliana* cosechadas a diferentes tiempos post-sincronización de afidicolina, un inhibidor de las subunidades α y δ de la ADN polimerasa, o en ayuno de sacarosa. El análisis reportó 155 genes relacionados con el ciclo celular. A este conjunto de genes lo hemos denominado AtBD_Menges (Base de datos de genes de A. thaliana en Menges). Menges y colaboradores (2005) utilizaron la plataforma en línea MatInspector y la secuencia de búsqueda o “query” YCYAACGGYY para determinar cuáles de los 155 genes presentan motivos MSA en los 500 pb de secuencia río arriba del SIT. La Tabla 3 presenta algunos de los genes reportados por Menges y colaboradores (2005) y el número de motivos MSA encontrados. La lista completa de AtBD_Menges se puede consultar en el Anexo Tabla A1. Los autores concluyeron que sólo un 41% (64 de 155) de los genes analizados contenían motivos MSA.

Tabla 3. Genes en la base de datos AtBD_Menges

<i>A. thaliana</i> (ID)	Sinónimo	MSA
At1g08560	syntaxin-related protein KNOLLE	3
At3g11520	CYCB1;3	3
At2g17620	CYCB2;1	3
At2g26760	CYCB1;4	4
At5g06150	CYCB1;2	1
At5g64960	CDKC;2	–
At1g73690	CDKD;1	–
At1g18040	CDKD;3	–
At5g63370	CDKG;1	–

Se indican el número de acceso o identificación (ID) y sinónimo de los genes de *A. thaliana* reportados por Menges y colaboradores (2005). La columna MSA indica el número de motivos MSA identificados por los autores utilizando la plataforma en línea MatInspector; el signo “-“ significa que no se encontró el motivo MSA. El sinónimo o nombre del gen fue asignado de acuerdo con la nomenclatura descrita en Vandepoele y cols. (2002). El Anexo Tabla A1 incluye la lista de los 155 genes reportados en Menges y colaboradores (2005).

Debido a que la plataforma MatInspector no es de acceso libre y con el propósito de familiarizarnos con los criterios de búsqueda de genes y motivos de secuencia, se estableció una ruta de análisis (“pipeline”) tomando como referencia los genes de *A. thaliana*. La estrategia fue la siguiente: 1) identificación de genes, 2) identificación de la secuencia río arriba del sitio de inicio de la traducción y 3) búsqueda de motivos MSA utilizando diversas plataformas bioinformáticas.

8.3.2 Identificación *in silico* de la secuencia de genes de *A. thaliana*

Para obtener las secuencias de los 155 genes en AtBD_Menges se utilizó el ID de cada gen como criterio de búsqueda en Phytozome (siglas en inglés de **P**lant **C**omparative **G**enomics portal of **E**nergy’s **J**oint **G**enome **I**nstitute; <https://phytozome.jgi.doe.gov/pz/portal.html#>), un portal que contiene las secuencias genómicas de una variedad de plantas incluyendo *A. thaliana* y *P. vulgaris*. Además, permite la identificación de secuencias río arriba del codón de inicio de la traducción.

De acuerdo a Chung y colaboradores (2006), en *A. thaliana* el tamaño de la secuencia 5'UTR oscila entre 50-500 pb. Por lo que en este proyecto se optó por crear una base de datos en formato FASTA, la cual denominamos AtBD_prom-5'UTR, que contiene la secuencia de un fragmento (prom-5'UTR) constituido por 2000 pb río arriba del sitio TSS y la del 5'UTR de cada uno de los genes en AtBD_Menges. En AtBD_prom-5'UTR, el promedio de prom-5'UTR es de 2,232 pb.

8.3.3 Búsqueda de motivos MSA en secuencias promotoras en *A. thaliana*

Se exploró una gama de plataformas en línea diseñadas con el propósito de buscar secuencias de motivos o elementos de regulación, así como sitios de unión a factores de transcripción. Las plataformas exploradas fueron:

PlantCARE (**C**is **A**cting **R**egulatory **E**lements)

(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>),

PLACE (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>),

Plantpan (http://plantpan2.itps.ncku.edu.tw/TFBSinfo.php?matrix=TFmatrixID_0357)

Integbio (<https://integbio.jp/dbcatalog/en/record/nbdc00168>),

Softberry Motif Finder

(<http://www.softberry.com/berry.phtml?topic=multirun&group=programs&subgroup=motifexp&advanced=on>),

MAST (<http://memesuite.org/tools/mast>),

MCAST (<http://meme-suite.org/tools/mcast>),

Plantpromoterdb (<http://ppdb.agr.gifu.ac.jp/ppdb/cgi-bin/index.cgi>)

Con el fin de explorar cuál de estas plataformas nos permitiría identificar motivos MSA, se utilizó la secuencia de AtKNOLLE (ID At1g08560) incluida en AtBD_prom-5'UTR. De acuerdo a lo reportado por Menges y cols. (2005) este gen contiene 3 motivos MSA, como se indica en la Tabla 3. Las plataformas que detectaron la presencia de motivos MSA en la secuencia prom-5'UTR de AtKNOLLE (At1g08560) fueron PlantCARE, MAST y MCAST. Por lo que se decidió que estas plataformas serían utilizadas para el análisis de los 155 fragmentos en AtBD_prom-5'UTR. La Tabla 4 ilustra algunos de los resultados obtenidos.

Tabla 4. Análisis de promotores en AtBD_prom-5'UTR utilizando las plataformas PlantCARE, MAST y MCAST

<i>A.thaliana</i> (ID)	MatInspector	PlantCARE	MAST	MCAST
At1g08560	3	1	1	1
At3g11520	3	1	1	1
At2g17620	3	1	1	1
At2g26760	4	1	2	2
At5g06150	5	3	3	3
At5g64960	-	-	-	-
At1g18040	-	1	1	1
At5g63370	-	1	2	2

El número indicado en cada columna corresponde al número de motivos MSA encontrados por cada plataforma. El signo “-“ significa que no se encontró el motivo MSA. El Anexo Tabla A2 incluye todos los resultados obtenidos en este análisis. Como referencia se incluyen los resultados reportados por Menges y colaboradores (2005), utilizando MatInspector.

Este análisis permitió identificar un total de 46 genes portadores de motivos MSA que se desglosan de la siguiente manera:

- De los 64 genes reportados como portadores de MSA por Menges y colaboradores (2005), PlantCARE identificó 25, los cuales también están reportados por MAST y MCAST en una lista de 33 genes. Cabe mencionar que MAST y MCAST identifican los mismos fragmentos en AtBD_prom-5'UTR como portadoras de motivos MSA.
- Las tres plataformas fueron capaces de reportar la presencia de motivos MSA en 13 nuevos genes (Tabla 5), es decir, no detectados por MatInspector (Menges y cols., 2005). De los cuales PlantCARE permitió localizar 8, mientras que MAST y MCAST identificaron PlantCARE, es decir, hay un gen (At4g10010) que presenta motivo MSA e identificado por PlantCARE que no fue detectado por MAST y MCAST.

Tabla 5. Nuevos genes de *A. thaliana* con motivos MSA identificados en AtBD_prom-5'UTR

<i>A. thaliana</i> (ID)	MatInspector	PlantCARE	MAST	MCAST
At4g10010	-	1	-	-
At1g73690	-	1	1	1
At1g18040	-	1	1	1
At5g63370	-	1	2	2
At5g11300	-	1	1	1
At1g18370	-	1	1	1
At5g13840	-	4	2	2
At2g33560	-	1	1	1
At3g50630	-	-	1	1
At1g49620	-	-	1	1
At5g67270	-	-	1	1
At1g75920	-	-	1	1
At1g61450	-	-	1	1

El número indicado en cada columna corresponde al número de motivos MSA encontrados por cada plataforma. El signo “-“ significa que no se encontró el motivo MSA. El Anexo Tabla A2 incluye estos resultados.

La diferencia entre los resultados reportados por Menges y colaboradores (2005) utilizando MatInspector y los encontrados con PlantCARE, MAST y MCAST posiblemente se deba a que la longitud y origen de las secuencias analizadas con MatInspector fue de 500 pb (río arriba del SIT) mientras que en el análisis con PlantCARE, MAST y MCAST se utilizaron fragmentos prom_5'UTR con un promedio de 2,232 pb. Otra posible explicación está asociada a los algoritmos y las bases de datos utilizados en cada plataforma.

Es de mencionar que mientras que PlantCARE identifica cuales de los elementos de regulación en *cis* reportados en la literatura (incluyendo el motivo MSA) están presentes en la secuencia analizada, las plataformas MAST y MCAST ofrecen la posibilidad de realizar la búsqueda de un motivo en particular, tanto en la secuencia de interés como en las secuencias en bases de datos a las que tienen acceso estas plataformas. Entre estas bases de datos se encuentran “Upstream Sequences”. Estas comprenden secuencias río arriba del SIT

de los genes presentes en el genoma de un organismo, por ejemplo, en el de *A. thaliana*, *P. vulgaris*, entre otros. La base de datos “Upstream Sequences, Arabidopsis thaliana.TAIR10.37”, es un conjunto de 285,977 secuencias con una longitud entre 201 y 1201 pb (con un promedio de 453 pb), la cual en este trabajo denominamos como AtDB_Up. La base de datos “Upstream sequences, Phaseolus vulgaris.PhaVulg1 0.38” es un conjunto de 175,879 secuencias con una longitud entre 152 y 1201 pb (con un promedio de 662 pb), la cual en este trabajo denominamos como PvDB_Up.

El análisis de la base de datos AtDB_Up utilizando MAST reportó la presencia de motivos MSA en 2 genes que no se encontraban en AtBD_prom-5'UTR. Al realizar un análisis similar con MCAST se encontraron 7 genes con motivos MSA de los cuales 6 no estaban presentes en AtBD_prom-5'UTR, como se muestra en la Tabla 6. Por lo que estos genes fueron añadidos a una base de datos que denominamos AtBD_MSA (véase Tabla A4 en el Anexo) que contiene un total de 86 genes, incluyendo los 64 genes con motivos MSA reportados por Menges y colaboradores (2005), y los 54 genes de *A. thaliana* en los que las plataformas PlantCARE, MAST y MCAST han encontrado un motivo MSA.

Tabla 6. Genes con motivos MSA identificados en AtBD_Up utilizando MAST y MCAST

<i>A.thaliana</i> (ID)	MatInspector	MCAST	MAST
At1g33930	–	0	2
At1g76320	–	3	4
At1g76310	2	3	3
At1g77660	–	2	3
At1g79340	–	2	2
At5g65580	–	2	2
At3g05110	–	3	3
At3g51150	–	1	0
At3g47750	–	1	0

El número indica la presencia de motivos MSA. El Anexo Tabla A4 incluye estos resultados. Como referencia se incluyen los resultados con MatInspector en Menges y colaboradores (2005).

En total, PlantCARE identificó 38 genes con motivos MSA, mientras que fueron 52 y 51 los genes detectados por MAST y MCAST, respectivamente. Sólo uno de los 38 genes identificados por PlantCARE no fue reconocido por las plataformas MAST y MCAST. A su vez, de los 52 genes detectados por MAST, 50 fueron identificados por MCAST. Es de mencionar que los 2 genes portadores de motivos MSA no identificados por MCAST, tampoco fueron reconocidos por PlantCARE. Por otro lado, uno de los 51 genes detectados por MCAST no fue identificado por MAST, ni por PlantCARE.

8.3.4 Búsqueda de motivos MSA en genes de *P. vulgaris*

Para identificar los genes en *P. vulgaris* que portan motivos MSA se aplicó una estrategia que combina los siguientes pasos: 1) Búsqueda de genes homólogos a los genes descritos en la base de datos AtBD_MSA, 2) Búsqueda de genes con motivos MSA en la base de datos PvBD_Up utilizando MAST y MCAST, y 3) Confirmación de la presencia de los motivos MSA en los genes identificados, utilizando las plataformas PlantCARE, MAST y MCAST.

8.3.5 Búsqueda de genes de *P. vulgaris* homólogos a los genes descritos en la base de datos AtBD_MSA

Para identificar los genes de *P. vulgaris* homólogos a los de *A. thaliana* que presentan motivos MSA, se hizo una búsqueda tipo BLASTp (**B**asic **L**ocal **A**lignment **S**earch **T**ool, por sus siglas en inglés) en el proteoma de *P. vulgaris* en Phytozome, utilizando como secuencia de búsqueda (“query”) la secuencia peptídica codificada por cada uno de los genes reportados en AtBD_MSA. Hay que tener en cuenta que algunos de los genes en esta base de datos pertenecen a familias o subfamilias génicas en las que el número de miembros puede variar según el organismo. Por lo que en la búsqueda del posible homólogo de cada uno de los genes en AtBD_MSA se identificaron varias proteínas de *P. vulgaris* que pertenecen a la misma subfamilia. La Tabla 7 incluye algunos resultados representativos de los genes de *P. vulgaris* identificados en esta búsqueda, entre los cuales se resaltan genes de *P. vulgaris* identificados como homólogos de genes de *A. thaliana* que pertenecen a una subfamilia, estos genes (IDs) de *P. vulgaris* están repetidos. De donde, Se prestó especial atención en depurar la lista de genes identificados en *P. vulgaris* para evitar repeticiones. Por otro lado, algunas de las proteínas codificadas por genes en AtBD_MSA no presentaron homología con proteínas de

P. vulgaris. En el Anexo Tabla A5 se recapitula la lista de los 159 genes de *P. vulgaris* identificados.

Tabla 7. Genes de *P. vulgaris* identificados como posibles homólogos de genes en AtBD_MSA

<i>A. thaliana</i>		<i>P. vulgaris</i>
ID	Sinónimo	ID
At1g08560	KNOLLE	Phvul.004G077900
At5g06150	CYCB1;2	Phvul.001G119700 ¹
		Phvul.008G203300 ²
		Phvul.001G119800 ³
At3g11520	CYCB1;3	Phvul.001G119700 ¹
		Phvul.008G203300 ²
		Phvul.001G119800 ³
At2g26760	CYCB1;4	Phvul.001G119700 ¹
		Phvul.008G203300 ²
At2g17620	CYCB2;1	Phvul.002G115100
		Phvul.009G115900
At1g73690	CDKD;1	Phvul.004G152000 ⁴
		Phvul.002G326800
At1g18040	CDKD;3	Phvul.004G152000 ⁴
At5g63370	CDKG;1	Phvul.002G175700
		Phvul.003G209300

Se indican el número de acceso ID, número de identificación de cada gen; Sinónimo, nombre del gen de acuerdo con la nomenclatura descrita en Vandepoele y cols. (2002). ^{1,2,3,4}, ID de genes repetidos. En el Anexo Tabla A5 se recapitula la lista de los 159 genes de *P. vulgaris* identificados.

8.3.6 Búsqueda de genes con motivos MSA en la base de datos PvBD_Up utilizando MAST y MCAST

Por otro lado, se analizó la base de datos denominada PvDB_Up con MAST y MCAST. Ambas plataformas detectan los mismos 34 genes con motivos MSA, de los cuales 6 ya

habían sido identificados en el análisis reportado en el inciso 8.3.5. Los 28 genes nuevos están enlistados en el Anexo Tabla A6. Con estas dos estrategias se obtuvo un total de 187 genes de *P. vulgaris*, los cuales se enlistan en el Anexo Tabla A7.

8.3.7 Identificación de los motivos MSA en los 187 genes de *P. vulgaris* utilizando las plataformas PlantCARE, MAST y MCAST

Con el propósito de realizar un análisis similar al que se llevó a cabo en las bases de datos de *A. thaliana* generadas en este trabajo, se identificó la secuencia de 2000 pb río arriba del SIT de cada uno de los 187 genes de *P. vulgaris* encontrados como portadores de motivos MSA. Los resultados obtenidos constituyen la base de datos denominada PvBD_prom-5'UTR (Tabla A7), en la cual la longitud promedio de las secuencias 5'UTR es de 232 pb.

El análisis de la base de datos PvBD_prom-5'UTR en la plataforma PlantCARE permitió identificar 42 genes con motivos MSA, mientras que fueron 63 y 57 los detectados por MAST y MCAST, respectivamente. 41 de los genes identificados por PlantCARE también están en la lista de los detectados por MAST, al igual que 35 genes de los reportados por MCAST (véase Tabla A7). De los 57 genes detectados por MCAST, 48 también fueron identificados por MAST por lo tanto hay 9 genes identificados por MCAST los cuales no fueron identificados por PlantCARE ni por MAST. De los 63 genes identificados por MAST, sólo uno no fue identificado por PlantCARE ni MCAST.

En resumen, en la Tabla A7 se indican los 73 genes de *P. vulgaris* que contienen motivos MSA, los cuales constituyen la base de datos PvBD_MSA.

Tabla 8. Análisis de PvBD_prom-5'UTR en las plataformas PlantCARE, MAST y MCAST

<i>P. vulgaris</i> (ID)	PlantCARE	MAST	MCAST
Phvul.004G077900 KNOLLE	1	2	2
Phvul.001G119700	1	2	3
Phvul.008G203300	2	1	4
Phvul.001G119800	2	2	2
Phvul.002G115100	1	1	2
Phvul.009G115900	-	-	-
Phvul.004G152000	-	-	-
Phvul.002G326800	-	-	-
Phvul.002G175700	-	-	-
Phvul.003G209300	-	-	-

Se indican el número de acceso ID, número de identificación de cada gen, el número indicado en cada columna corresponde al número de motivos MSA encontrados por cada plataforma. En el Anexo Tabla A7 se recapitula la lista de los 187 genes de *P. vulgaris* y los motivos MSA identificados. El signo “-“ significa que no se encontró el motivo MSA

8.3.8 Análisis de los motivos MSA en genes de *A. thaliana* y *P. vulgaris*

El análisis de las bases de datos AtBD_MSA y PvBD_MSA mostró que PlantCARE, MAST y MCAST identifican el mismo número de motivos MSA en 20 (32%) o 14 (19%) de los genes con motivos MSA, respectivamente, es decir, en *A. thaliana* las tres plataformas identifican 13 genes con 1 motivo MSA, 6 con dos y 1 con tres (Tabla 8). Mientras que en *P. vulgaris* las tres plataformas identifican 9 genes con 1 motivo, 5 con dos y 1 con tres (Tabla 8). Por otro lado, cuando un motivo es identificado por más de una plataforma su posición siempre coincide, por lo tanto, detectan el mismo motivo.

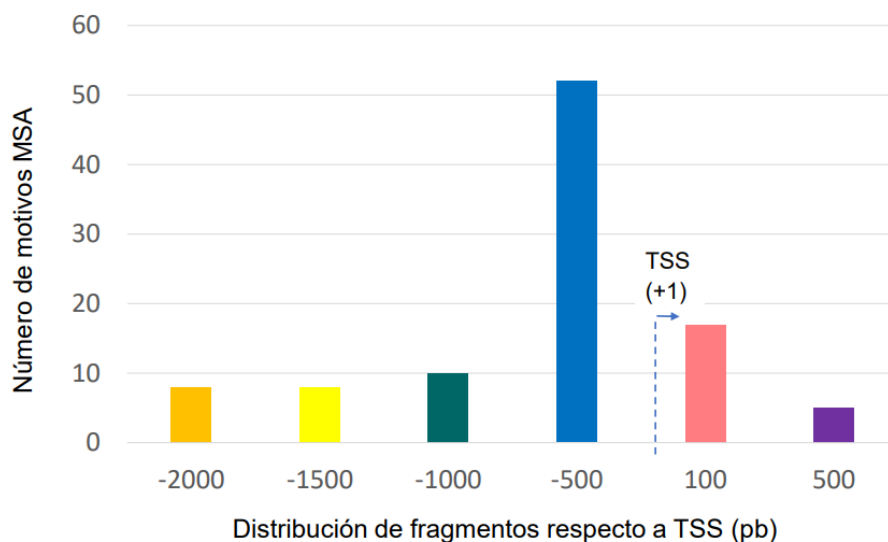
Estos resultados llevaron a investigar la distribución de la posición de los motivos MSA en AtBD_MSA y PvBD_MSA respecto al sitio TSS indicado en los transcriptomas de *A. thaliana* y *P. vulgaris* en Phytozome. Las Tablas A9 y A10 recapitulan la posición de cada motivo MSA en genes de *A. thaliana* y *P. vulgaris* en AtBD_MSA y PvBD_MSA, respectivamente. Para determinar la distribución de frecuencia de la posición de los motivos

MSA en la secuencia de promotor, se utilizó un rango de 500 pb, es decir, fragmentos ubicados entre los nucleótidos -1 y -500, -501 y -1000, -1001 y -1500, -1501 y -2000 respecto al sitio TSS (+1). Así mismo, para el análisis de distribución de motivos MSA en la secuencia anotada como 5'UTR se analizaron los fragmentos entre las posiciones +1 a +100 y +101 a +500. Como se ilustra en la Figura 15, de los 100 motivos MSA presentes en los 73 genes de la base de datos PvBD_MSA, 52 de los motivos se encuentran en el fragmento de 500 pb río arriba al TSS y hay 8 y 10 motivos en los fragmentos más distales. Por otro lado, se encontraron 17 motivos ubicados en el fragmento de 100 pb proximal al nucleótido +1 del 5'UTR de algunos genes y 5 adicionales en el fragmento distal (100-500 pb río abajo del TSS). Estos datos demuestran que el 70% de los los motivos MSA están presentes en el promotor del gen, en su mayoría (52%) en una posición proximal al TSS.

Tabla 9. Resultados ilustrativos del análisis de la posición de los motivos MSA con respecto al sitio TSS (+1)

<i>P. vulgaris</i> (ID)	Número de MSA	Posición con respecto a TSS (+1)
Phvul.001G061200	1	-112
Phvul.001G075000	1	-404
Phvul.001G119700	2	-93
		-17
Phvul.001G119800	2	42
		-1821
Phvul.002G115100	1	-53
Phvul.004G034600	2	24
		65
Phvul.004G077900	2	-80
		-52
Phvul.007G041900	1	7
Phvul.008G076000	1	-619
Phvul.008G203300	1	-68
Phvul.011G125100	1	-112

Se muestra el ID de cada gen, el número indicado en cada columna corresponde al número de motivos MSA encontrados y la posición de cada uno con respecto al sitio TSS, los signos (-) y (+) indican la posición río arriba o río abajo del TSS, respectivamente. En el Anexo Tabla A8 se recapitula la posición de los motivos MSA encontrados en los 73 genes de *P. vulgaris*.



Gráfica 1. Distribución de frecuencia de las posiciones de los motivos MSA en genes de *A. thaliana* (A) y *P. vulgaris* (B), con respecto al sitio TSS (+1), en los 73 genes de *P. vulgaris*. La Tabla A8 en el Anexo recapitula la posición de los motivos MSA encontrados.

8.4 Posible función de los genes de *P. vulgaris* portadores de motivos MSA.

Como punto de partida, se planteó la identificación de dominios en la secuencia peptídica codificada por los 73 genes en PvDB_MSA utilizando la opción de búsqueda en la base de datos de familias de proteínas Pfam.

La Tabla 10 muestra algunos de los resultados obtenidos, entre los que se encuentran los dominios SNARE o syntaxina, “Myb-like DNA-binding”, ciclina N-terminal y C-terminal, entre otros. El Anexo A9 contiene la lista completa de los dominios de los 73 genes de *P. vulgaris*. La interpretación de los resultados en Pfam requiere un análisis más exhaustivo.

Tabla 10. Resultados ilustrativos del análisis de los dominios encontrados por Pfam en cada uno de los genes de *P. vulgaris* identificados como portadores de MSA

<i>P. vulgaris</i> (ID)	Dominio Pfam
Phvul.004G077900 KNOLLE	Syntaxin / SNARE domain
Phvul.001G119700	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
Phvul.008G203300	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
Phvul.001G119800	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
Phvul.002G115100	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
Phvul.004G034600	Targeting protein for Xklp2 (TPX2)
Phvul.007G041900	B3 DNA binding domain
Phvul.008G076000	Microtubule associated protein (MAP65/ASE1 family)
Phvul.011G125100	Reticulon
Phvul.001G075000	alpha/beta hydrolase fold
Phvul.001G061200	Myb-like DNA-binding domain

El Anexo Tabla A9 contiene la lista de los 73 genes de *P. vulgaris* y el dominio Pfam identificado en la secuencia peptídica correspondiente.

9. DISCUSIÓN

Durante la organogénesis del nódulo existe una reactivación del ciclo celular de las células corticales en el sitio de infección con rizobia. Esta constante actividad meristemática da lugar a la formación de un primordio de nódulo que madura hasta convertirse en un nódulo fijador de nitrógeno. En este contexto, la identificación de genes de ciclo celular y su caracterización funcional brinda la oportunidad de generar herramientas moleculares para el estudio de la dinámica de la infección y organogénesis del nódulo, y otros temas de biología del desarrollo que involucran etapas de división celular. El gen KNOLLE, esencial para la formación de la placa celular durante la etapa citocinética de la división celular en células vegetales, es una de estas herramientas.

Este proyecto derivó de la caracterización de la actividad del promotor del gen KNOLLE de *P. vulgaris* (promPvKNOLLE) en raíces pre- y post-inoculadas con rizobia, lo que permitió confirmar el potencial de este gen para visualizar células en división, específicamente durante la citocinesis (Tesis de licenciatura de Emmanuel Ayala Guzmán, 2015). El análisis de la secuencia de promPvKNOLLE permitió constatar la presencia de motivos MSA, previamente identificados en el promotor del gen ortólogo KNOLLE en *A. thaliana* y en otros genes de ciclo celular en células vegetales como CycB1 de *C. roseus*, AtNACK1/HK, AtCycB1;1, AtCycB1;112, CDC20.1, entre otros (Ito y cols. 1998; Haga y cols. 2007; Kato y cols. 2009). La caracterización funcional de los motivos MSA sugiere que estos motivos confieren un actividad espacio-temporal específica de las fases G2/M. Esta especificidad ofrece un gran potencial como herramienta molecular para la expresión de proteínas de interés en células en división y la identificación de genes que pudieran estar involucrados en ciclo celular. Con estas directrices se planteó el diseño de un promotor sintético específico de células en división y la identificación de genes de mitosis.

Diseño de un promotor sintético específico de células en división

Una de las estrategias fundamentales en la caracterización funcional de genes ha sido la identificación y análisis de los elementos que regulan la actividad del promotor del gen de interés, en *cis* y *trans* que contribuyen a regular la actividad del promotor del gen de interés. Si bien, el promotor de un gen puede presentar la secuencia consenso de uno o más elementos de regulación, la actividad funcional de cada motivo debe ser demostrada experimentalmente.

Entre las estrategias aplicadas se encuentra la caracterización de versiones del promotor generadas por eliminación seriada de fragmentos que contengan el motivo en estudio, por mutagénesis seriada o sitio dirigida, por inserción seriada de secuencias (en inglés linker-scanning), eliminación del motivo de interés, entre otras.

Otra alternativa es analizar la actividad de un promotor sintético, es decir un promotor constitutivo en fusión con la secuencia de motivo o elemento de regulación de interés. Con esta metodología se ha validado la función del elemento de respuesta a auxina (AuxRE, siglas en inglés) al caracterizar la actividad del promotor sintético prom35S-(AuxRE)8X, mejor conocido como promotor DR5, uno de los más ampliamente utilizados en plantas.

Un promotor sintético es un promotor quimérico en el que la secuencia de un promotor mínimo o núcleo de promotor con actividad constitutiva se combina con la secuencia de un elemento de regulación en *cis* que le proporciona una cierta especificidad (Liu y Stewart Jr, 2016; Aysha y cols., 2018). Uno de los núcleos de promotor más utilizado en plantas es el fragmento -340 a +1 del promotor 35S, el cual presenta una fuerte actividad constitutiva en raíz, hoja, tallo y flor (Benfey y cols., 1990). Sobre esta base se han construido promotores sintéticos específicos de plantas, cuyo único elemento regulador es una secuencia de respuesta a hormona, de respuesta a estrés, o inducible por algún otro factor de estimulación (“enhancer”), entre los que se encuentran los promotores sintéticos DR5, TCS, 4X RSRE (inducible por estrés en general, siglas en inglés de rapid stress response element), 4X GCC (inducible por ácido jasmónico) y 4X ROSE 1 (inducible por especies de oxígeno reactivo, siglas en inglés de ROS inducible element), entre otros (Dey y cols., 2015; Liu y Stewart Jr, 2016; Aysha y cols., 2018).

Tomando ventaja de la especificidad de los motivos MSA, en este trabajo de tesis se propuso el diseño del promotor sintético (prom35S-MSA), cuya actividad se espera sea específica de células en división y así disponer de una herramienta molecular para identificar células mitóticas en células vegetales. En prom35S-MSA, el fragmento -340 a -1 del prom35S está en fusión con los tres repetidos del motivo MSA del promotor PvKNOLLE. Así mismo, con el propósito de investigar si la secuencia 5'UTR de PvKNOLLE tiene un papel en la regulación de la expresión de este gen, se construyeron dos cassettes en los que prom35S-MSA está en fusión transcripcional con secuencia 5'UTR de PvKNOLLE (prom35S-MSA::5'UTR_PvKNOLLE) o con la secuencia 5'UTR conocida como OmegaLeader, para

generar prom35S-MSA::OmegaLeader. El fragmento OmegaLeader, proviene del virus del mosaico de tabaco y promueve una traducción eficiente pero no contiene elementos de regulación en *cis* (Gallie y cols., 1988, 2002). Para visualizar la especificidad de la actividad de los cassettes diseñados fueron clonados en fusión transcripcional con la secuencia codificante de GUS. Las construcciones correspondientes han sido transformadas en *A. rhizogenes* K599 y quedan en espera de ser utilizadas para generar raíces transgénicas y caracterizar la actividad de los promotores sintéticos diseñados.

Identificación de genes de *P. vulgaris*, cuyos promotores presenten motivos MSA

Por otro lado, dada la especificidad que confieren los motivos MSA a la actividad de promotor en células en la fase G2/M del ciclo celular (Ito y cols., 1998), era de esperar que, al igual que en el genoma *A. thaliana* (Menges y cols., 2005), el motivo MSA estuviera presente en el promotor de un grupo de genes asociados al ciclo celular en *P. vulgaris*.

El estudio realizado por Menges y colaboradores (2005) reportó la identificación de 155 genes de *A. thaliana* cuya expresión aumenta durante la fase G2/M del ciclo celular. El análisis bioinformático (en la plataforma MatInspector) de 500 pb de secuencia río arriba del SIT de estos genes demostró que 64 de estos genes portan repetidos del elemento de regulación en *cis* MSA (Ito y cols., 1998).

Para identificar los genes de *P. vulgaris* que portan motivos MSA, y dado que la plataforma MatInspector no es una herramienta de acceso abierto, en esta tesis se exploró el uso de las plataformas PlantCARE, MAST y MCAST. Para validar estas herramientas, se optó por analizar la secuencia de fragmentos constituidos por 2000 pb río arriba del TSS y el 5'UTR de los 155 genes reportados por Menges y colaboradores (2005). PlantCARE, MAST y MCAST identificaron como portadores de MSA a 46 (30%) de los 155 genes, de los cuales 13 no habían sido identificados por MatInspector. Una posible explicación a la diferencia entre lo reportado por MatInspector y lo encontrado con PlantCARE, MAST y MCAST pudiera estar asociada, por un lado, a la longitud de la secuencia analizada (500 pb versus más de 2000 pb). Por otro lado, en esta tesis se utilizaron como “query” los 11 nucleótidos de la secuencia consenso del motivo MSA (YCYAACGGYYA), mientras en el análisis con MatInspector la secuencia usada es de 10 nucleótidos y omite la adenina invariable en el extremo 3' del motivo (Menges y cols., 2005), lo que quizás explica la identificación de un

mayor número de genes. Es de resaltar que 5 de los 11 nucleótidos del motivo MSA (nucleótidos subrayados) constituyen el “core” del motivo MSA, pero también corresponden a la secuencia consenso de unión de algunos factores de transcripción tipo Myb (Ito y cols., 1998). La búsqueda la realizaron en una base de datos con secuencias de unión a factores de transcripción, por lo que quizás no discrimina entre motivos MSA y secuencias reconocidas por Myb. Mientras que el análisis en la plataforma PlantCARE reporta de forma individual la identificación de motivos presentes en la secuencia de interés y distingue entre motivos Myb y MSA (Lescot y cols., 2002). Una de las características de MAST y MCAST es que no permiten sustituciones inespecíficas (“mismatches”) ni la inclusión de espacios o “gaps” (Bailey y Gribskov, 1998). Una diferencia entre estas dos plataformas es que MCAST incluye otro parámetro, la búsqueda de grupos (“clusters”) del motivo de interés, separados por una distancia máxima. En el caso de secuencias nucleotídicas, la plataforma MCAST define esta distancia por “default” como igual o menor a 50 nt, o bien el usuario la establece. Este parámetro resulta relevante si se toma en cuenta que en análisis previos se llegó a la conclusión de que los motivos MSA pueden estar repetidos hasta en 4-5 copias y suelen estar en tándem (Ito y cols. 1998; Kato y cols., 2009). El conjunto de estas características permite pensar que los algoritmos de búsqueda de PlantCARE, MAST y MCAST pudieran ser más robustos que los de MatInspector en la identificación de motivos MSA. De hecho, PlantCARE, MAST y MCAST han sido ampliamente utilizadas para la búsqueda global de elementos de regulación en *cis* en genes y genomas de *A. thaliana*, *N. tabacum*, *M. truncatula*, entre otros. (Baruah y cols., 2020; Gao y cols., 2019; Liu y Stewart Jr, 2016; Liu y cols., 2019; Aysha y cols., 2018; Pen y cols., 2016)

También se aprovechó la opción de MAST y MCAST que utiliza la base de datos “Upstream Sequences” (meme-suite.org), constituida por las secuencias río arriba del codón de inicio en los genes de un organismo en particular. En este ejercicio se identificaron 8 nuevos genes de *A. thaliana* portadores de MSA, no identificados por Menges y cols., (2005). En un análisis similar con 187 genes de *P. vulgaris*, PlantCARE, MAST y MCAST identificaron 73 genes como portadores de MSA, 43 genes homólogos de los genes de *A. thaliana* descritos en este trabajo y 28 genes adicionales incluidos en la base de datos “Upstream sequences” de *P. vulgaris*.

Por otro lado, uno de los elementos que valida el uso de PlantCARE, MAST y MCAST para buscar genes con motivos MSA se refiere a que, en algunos genes, las tres plataformas identifican el mismo número de estos elementos, con una distribución en tándem y en su mayoría (52 %) se ubican en el fragmento -500 a -1 pb proximal al TSS, un patrón similar al de motivos MSA en genes de *N. tabacum* cuya expresión es regulada por el factor de transcripción NtmybA2 (Kato y cols., 2009).

En un intento por definir la posible relación funcional entre estos genes y el ciclo celular en células vegetales, se analizó la arquitectura y organización de los dominios en la estructura primaria de las proteínas producto de los genes de *P. vulgaris* identificados. Este análisis sugiere que estas pertenecen a diferentes subfamilias de proteínas en las que algunos miembros tienen funciones relacionadas con ciclo celular: ciclinas (Murray, 2004), cinesinas (Audett y Maresca, 2018), cinasas, fosfatasa (proteína fosfatasa tipo 2C y Tirosin-fosfatasa; Fuchs y cols., 2012; Shankar y cols., 2015), inhibidor de cinasa dependiente de ciclina (Kumar y Larkin, 2017), factores de transcripción (Myb y bZip; Dubos y cols., 2010; Dröge-Laser y cols., 2018), proteínas asociadas a la maquinaria de traducción, componentes de cinetocoro (CENP-T y Mad3/BUB1; Nishino y cols., 2012; Leontiou y cols., 2019), proteínas reguladoras del ensamblaje del huso (MAP65/ASE1 y HORMA domain; Rosenberg y Corbett, 2015; She y cols., 2019), proteína asociada a microtúbulo TPX2 (Wadsworth, 2015), entre otras. Si bien es relevante confirmar si la expresión de estos genes y sus productos es específica en la fase G2/M del ciclo celular, así como la especificidad de la actividad de promotor, la función de los genes identificados refuerza la propuesta de que la presencia de dominios MSA en el promotor es una herramienta para distinguir genes potencialmente relacionados con ciclo celular.

10. CONCLUSIONES

- Se generaron dos construcciones que permitirán evaluar la función de los motivos MSA en la actividad del promotor sintético
- La lista de genes de *A. thaliana* cuyos promotores tienen motivos MSA es más extensa que lo reportado en Menges y cols., (2005). En este trabajo se describe la identificación de 54 genes portadores de MSA, de los cuales 21 genes adicionales, no reportados en el análisis de Menges y cols., (2005).
- En *P. vulgaris* se identificaron 73 de 187 genes homólogos a los de *A. thaliana* como portadores de motivos MSA.
- Los resultados presentados en la presente investigación confirman que las plataformas PlantCARE, MAST y MCAST son herramientas útiles en la búsqueda de motivos MSA.
- La plataforma MAST es la plataforma más adecuada para la búsqueda de motivos MSA de acuerdo con los resultados obtenidos en este proyecto; sin embargo, una búsqueda con las demás plataformas pudiera enriquecer los resultados.
- Los motivos MSA de los genes de *P. vulgaris* se encuentran principalmente en la región promotora y ocupan posiciones proximales al sitio TSS a encontrarse en la región promotora, en las regiones proximales al sitio TSS, es decir en la región promotora.

11. PERSPECTIVAS

El siguiente reto será demostrar que los genes identificados son específicos y, quizás esenciales para que la célula inicie y concluya las etapas de mitosis y citocinesis del ciclo celular.

Igualmente, sería interesante aplicar por métodos bioquímicos para identificar qué factores de transcripción de *P. vulgaris* se unen a los motivos MSA e investigar su posible función durante la división celular.

12. REFERENCIAS

- Alberts, B.,** Bray, D., Wilson, J. H., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (2002). *Molecular Biology of the Cell* (Vol. 4). New York: Garland Science.
- Amack, S. C.,** & Antunes, M. S. (2020). CaMV35S promoter—a plant biology and biotechnology workhorse in the era of synthetic biology. *Current Plant Biology*, 24, 100179.
- Andersson R y Sandelin A.** (2019). *Nature Reviews Genetics*, 1, 20-45.
- Andrews, M.,** & Andrews, M. E. (2017). Specificity in legume-rhizobia symbioses. *International Journal of Molecular Sciences*, 18(4), 705.
- Audett, M. R.,** & Maresca, T. J. (2018). Cell Division: Here Comes the Kinesin Cavalry. *Current Biology*, 28(17), R943-R946.
- Austin, J. R.,** Seguí-Simarro, J. M., White, E. A., & Staehelin, L. A. (2004). Electron tomographic analysis of somatic cell plate formation in meristematic cells of Arabidopsis preserved by high-pressure freezing. *The Plant Cell*, 16(4), 836-856.
- Aysha, J.,** Noman, M., Wang, F., Liu, W., Zhou, Y., Li, H., & Li, X. (2018). Synthetic promoters: designing the cis regulatory modules for controlled gene expression. *Molecular Biotechnology*, 60(8), 608-620.
- Bailey, T. L.,** & Noble, W. S. (2003). Searching for statistically significant regulatory modules. *Bioinformatics*, 19(suppl_2), ii16-ii25.
- Bailey, T. L.,** Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren J., Li W. W., & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids Research*, 37(suppl_2), W202-W208.
- Bailey, T. L.,** & Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics (Oxford, England)*, 14(1), 48-54.
- Benfey, P. N.,** Ren, L., & Chua, N. H. (1989). The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *The EMBO Journal*, 8(8), 2195-2202.
- Benfey, P. N.,** & Chua, N. H. (1990). The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science*, 250(4983), 959-966.
- Benfey, P. N.,** Ren, L., & Chua, N. H. (1990). Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *The EMBO Journal*, 9(6), 1677-1684.

- Bonifacino, J. S.,** & Glick, B. S. (2004). The mechanisms of vesicle budding and fusion. *Cell*, *116*(2), 153-166.
- Cárdenas, L., Feijó, J. A.,** Kunkel, J. G., Sánchez, F., Holdaway-Clarke, T., Hepler, P. K., & Quinto, C. (1999). Rhizobium Nod factors induce increases in intracellular free calcium and extracellular calcium influxes in bean root hairs. *The Plant Journal*, *19*(3), 347-352.
- Cárdenas, L.,** Alemán, E., Nava, N., Santana, O., Sánchez, F., & Quinto, C. (2006). Early responses to Nod factors and mycorrhizal colonization in a non-nodulating *Phaseolus vulgaris* mutant. *Planta*, *223*(4), 746-754.
- Chung, B. Y.,** Simons, C., Firth, A. E., Brown, C. M., & Hellens, R. P. (2006). Effect of 5'UTR introns on gene expression in *Arabidopsis thaliana*. *BMC Genomics*, *7*(1), 120.
- Covey, S. N.,** & Hull, R. (1981). Transcription of cauliflower mosaic virus DNA. Detection of transcripts, properties, and location of the gene encoding the virus inclusion body protein. *Virology*, *111*(2), 463-474.
- Dey, N.,** Sarkar, S., Acharya, S., & Maiti, I. B. (2015). Synthetic promoters in planta. *Planta*, *242*(5), 1077-1094.
- Dröge-Laser, W.,** Snoek, B. L., Snel, B., & Weiste, C. (2018). The Arabidopsis bZIP transcription factor family -an update. *Current Opinion in Plant Biology*, *45*, 36-49.
- Dubos, C.,** Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. *Trends in Plant Science*, *15*(10), 573-581.
- Ferguson, B. J.,** Indrasumunar, A., Hayashi, S., Lin, M. H., Lin, Y. H., Reid, D. E., & Gresshoff, P. M. (2010). Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology*, *52*(1), 61-76.
- Fuchs, S.,** Grill, E., Meskiene, I., & Schweighofer, A. (2013). Type 2C protein phosphatases in plants. *The FEBS Journal*, *280*(2), 681-693.
- Gallie, D. R.** (2002). The 5'-leader of tobacco mosaic virus promotes translation through enhanced recruitment of eIF4F. *Nucleic Acids Research*, *30*(15), 3401-3411.
- Geertz, M.,** & Maerkl, S. J. (2010). Experimental strategies for studying transcription factor–DNA binding specificities. *Briefings in Functional Genomics*, *9*(5-6), 362-373.
- Goñi, J. R.,** Pérez, A., Torrents, D., & Orozco, M. (2007). Determining promoter location based on DNA structure first-principles calculations. *Genome Biology*, *8*(12), R263.

- Guglielmo C, Z. M., & Fernandez Da Silva, R.** (2016). Principales promotores utilizados en la transformación genética de plantas. *Revista Colombiana de Biotecnología*, *18*(2), 119-128.
- Haberle, V., & Stark, A.** (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, *19*(10), 621-637.
- Haga, N., Kato, K., Murase, M., Araki, S., Kubo, M., Demura, T., Suzuki K., Müller I., Voß U., Jürgens G., & Ito, M.** (2007). R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in *Arabidopsis thaliana*. *Development*, *134*(6), 1101-1110.
- Han, J., Pluhackova, K., & Böckmann, R. A.** (2017). The multifaceted role of SNARE proteins in membrane fusion. *Frontiers in Physiology*, *8*, 5. <https://doi.org/10.3389/fphys.2017.00005>
- Hirsch, A. M.** (1992). Developmental biology of legume nodulation. *New Phytologist*, *122*(2), 211-237. Hong, W. (2005). SNAREs and traffic. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1744*(2), 120-144.
- Ito, M., Iwase, M., Kodama, H., Lavisse, P., Komamine, A., Nishihama, R., Mashida Y., & Watanabe, A.** (1998). A novel cis-acting element in promoters of plant B-type cyclin genes activates M phase-specific transcription. *The Plant Cell*, *10*(3), 331-341.
- Ito, M., Araki, S., Matsunaga, S., Itoh, T., Nishihama, R., Machida, Y., Doonan Y. H., & Watanabe, A.** (2001). G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors. *The Plant Cell*, *13*(8), 1891-1905.
- Jacinto-Hernández, C., Carrillo-Castañeda, G., & Campos-Escuerdo, A.** (1996). Caracterización de cuatro variedades de frijol (*Phaseolus vulgaris* L.) por sus cualidades físicas y nutricionales. I. Aspectos metodológicos. *Agronomía Mesoamericana*, 37-41.
- Kato, K., Gális, I., Suzuki, S., Araki, S., Demura, T., Criqui, M. C., Potuschak T., Genschik P., Fukuda H., Matsuoka K., & Ito, M.** (2009). Preferential up-regulation of G2/M phase-specific genes by overexpression of the hyperactive form of NtmybA2 lacking its negative regulation domain in tobacco BY-2 cells. *Plant Physiology*, *149*(4), 1945-1957.
- Kumar, N., & Larkin, J. C.** (2017). Why do plants need so many cyclin-dependent kinase inhibitors?. *Plant Signaling & Behavior*, *12*(2), e1282021.

- Lauber, M. H.**, Waizenegger, I., Steinmann, T., Schwarz, H., Mayer, U., Hwang, I., Lukowitz W., & **Jürgens, G.** (1997). The Arabidopsis KNOLLE protein is a cytokinesis-specific syntaxin. *The Journal of Cell Biology*, 139(6), 1485-1493.
- Laufman, O.**, Hong, W., & Lev, S. (2013). The COG complex interacts with multiple Golgi SNAREs and enhances fusogenic assembly of SNARE complexes. *Journal of Cell Science*, 126(6), 1506-1516.
- Leontiou, I.**, London, N., May, K. M., Ma, Y., Grzesiak, L., Medina-Pritchard, B., Amin P., Jeyaprakash A. A., Biggins S., & **Hardwick, K. G.** (2019). The bub1-tpr domain interacts directly with mad3 to generate robust spindle checkpoint arrest. *Current Biology*, 29(14), 2407-2414.
- Lescot, M.**, Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé P., & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, 30(1), 325-327.
- Liu, C. W.**, & Murray, J. D. (2016). The role of flavonoids in nodulation host-range specificity: An update. *Plants*, 5(3), 33.
- Liu, W.**, & Stewart Jr, C. N. (2016). Plant synthetic promoters and transcription factors. *Current Opinion in Biotechnology*, 37, 36-44.
- Lukowitz, W.**, Mayer, U., & Jürgens, G. (1996). Cytokinesis in the Arabidopsis embryo involves the syntaxin-related KNOLLE gene product. *Cell*, 84(1), 61-71.
- Martínez-Romero, E.**, Segovia, L., Mercante, F. M., Franco, A. A., Graham, P., & Pardo, M. A. (1991). *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L. beans and *Leucaena* sp. trees. *International Journal of Systematic and Evolutionary Microbiology*, 41(3), 417-426.
- Menges, M.**, De Jager, S. M., Gruissem, W., & Murray, J. A. (2005). Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *The Plant Journal*, 41(4), 546-566.
- Mora de Zayas, L.** (2005). Caracterización a nivel genético y molecular de la producción de bacteriocina por *Rhizobium leguminosarum* bv. *viciae* CEPA Z25. Tesis doctoral. Universidad de Granada, España. <https://digibug.ugr.es/handle/10481/679>.

- Nagata, T.**, Sakamoto, K., & Shimizu, T. (2004). Tobacco BY-2 cells: the present and beyond. *In Vitro Cellular & Developmental Biology Plant*, 40(2), 163-166.
- Nishino, T.**, Takeuchi, K., Gascoigne, K. E., Suzuki, A., Hori, T., Oyama, T., Morikawa K., Cheeseman I. M., & Fukagawa, T. (2012). CENP-TWSX forms a unique centromeric chromatin structure with a histone-like fold. *Cell*, 148(3), 487-501.
- Odell, J. T.**, Nagy, F., & Chua, N. H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature*, 313(6005), 810-812.
- Oldroyd, G. E.**, & Downie, J. A. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annual Review of Plant Biology*, 59, 519-546.
- Ow, D. W.**, Jacobs, J. D., & Howell, S. H. (1987). Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. *Proceedings of the National Academy of Sciences of the United States of America*, 84(14), 4870-4874.
- Quinto, C.**, & Cárdenas, L. (2007). Diálogo para ganar: interacción simbiótica entre una bacteria del suelo y el frijol. *Biotechnología*, 14, 273-280.
- Raven, P. H.**, Evert, R. F., & Curtis, H. (1981). In: *Biology of plants*. 3rd ed. New York, N.Y.: Worth Publishers. Chicago / Turabian.
- Rosenberg, S. C.**, & Corbett, K. D. (2015). The multifaceted roles of the HORMA domain in cellular signaling. *Journal of Cell Biology*, 211(4), 745-755.
- Shankar, A.**, Agrawal, N., Sharma, M., Pandey, A., & K Pandey, G. (2015). Role of protein tyrosine phosphatases in plants. *Current Genomics*, 16(4), 224-236.
- She, Z. Y.**, Wei, Y. L., Lin, Y., Li, Y. L., & Lu, M. H. (2019). Mechanisms of the Ase1/PRC1/MAP65 family in central spindle assembly. *Biological Reviews*, 94(6), 2033-2048.
- Srivastava, R.**, Rai, K. M., Srivastava, M., Kumar, V., Pandey, B., Singh, S. P., Bag K. S., Singh D. B., Tuli R., & **Sawant, S. V.** (2014). Distinct role of core promoter architecture in regulation of light-mediated responses in plant genes. *Molecular Plant*, 7(4), 626-641.
- Touihri, S.**, Knöll, C., Stierhof, Y. D., Müller, I., Mayer, U., & Jürgens, G. (2011). Functional anatomy of the Arabidopsis cytokinesis-specific syntaxin KNOLLE. *The Plant Journal*, 68(5), 755-764.

- Ulloa, J. A.,** Rosas Ulloa, P., Ramírez Ramírez, J. C., & Ulloa Rangel, B. E. (2011). El frijol (*Phaseolus vulgaris*): su importancia nutricional y como fuente de fitoquímicos. Repositorio Institucional Aramara. Universidad Autónoma de Nayarit. <http://dspace.uan.mx:8080/jspui/handle/123456789/582>.
- Venter, M.** (2007). Synthetic promoters: genetic control through cis engineering. *Trends in Plant Science*, 12(3), 118-124.
- Wadsworth, P.** (2015). Tpx2. *Current Biology*, 25(24), R1156-R1158.
- Wang, E. H.** (2006). Structure and Function of Core Promoter Elements in RNA Polymerase II Transcription. In: *Gene Expression and Regulation* (pp. 135-144). Springer, New York, NY.
- Zuo, Y. C., & Li, Q. Z.** (2011). Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. *Genomics*, 97(2), 112-120.

13. ANEXOS

Tabla A1. Genes en AtBD_Menges

	<i>A. thaliana</i> (ID)	Sinónimo	MSA
1	At1g02730	cellulose synthase catalytic subunit, putative	6
2	At5g15510	TPX2 (targeting protein for Xklp2) protein family	5
3	At1g03780	TPX2/ TARGETING PROTEIN FOR XKLP2	4
4	At2g25060	early nodulin-like predicted GPI-anchored protein	4
5	At2g26760	CYCB1;4	4
6	At3g02640	Transmembrane protein	4
7	At3g51280	Tetratricopeptide repeat (TPR)-like superfamily protein	4
8	At1g08560	syntaxin-related protein KNOLLE	3
9	At1g44110	CYCA1;1	3
10	At1g50490	UBC20/ UBIQUITIN-CONJUGATING ENZYME 20	3
11	At1g59540	kinesin	3
12	At1g72250	kinesin	3
13	At2g17620	CYCB2;1	3
14	At3g03090	sugar transporter-like	3
15	At3g11520	CYCB1;3	3
16	At3g19590	BUB3.1/ BUB (BUDDING UNINHIBITED BY BENZYMIDAZOL) 3.1	3
17	At3g23670	kinesin	3
18	At4g35620	CYCB2;2	3
19	At5g02370	kinesin	3
20	At5g11510	AtMYB3R4	3
21	At5g23910	kinesin	3
22	At5g36710	Transmembrane protein	3
23	At5g51600	PLE/ PLEIADE	3
24	At5g55830	L-Lectin kinase	3
25	At5g66230	Chalcone-flavanone isomerase family protein	3
26	At1g34355	PS1/ PARALLEL SPINDLE 1	2
27	At1g76310	CYCB2;4	2
28	At2g16270	Transmembrane protein	2
29	At2g20590	RTN17/ RETICULON 17	2
30	At2g22610	kinesin	2
31	At2g38160	BPP4/ BASIC PROLINE-RICH PROTEIN4	2
32	At2g44190	EDE1/ ENDOSPERM DEFECTIVE 1	2

33	At2g47500	kinesin	2
34	At3g12870	Transmembrane protein	2
35	At3g23890	topoisomerase II	2
36	At3g25980	putative mitotic checkpoint protein; MAD2-like	2
37	At3g27330	zinc finger (C3HC4-type RING finger) family protein	2
38	At3g55660	ROPGEF6/ ROP	2
39	At4g05520	EHD2/ EPS15 HOMOLOGY	2
40	At4g11080	3xHMG-box1/ 3XHIGH MOBILITY GROUP-BOX1	2
41	At4g15830	expressed protein	2
42	At4g23800	98b like protein, HMG; ABF2-like	2
43	At4g28430	RTN18/ RETICULON 18	2
44	At4g33260	WD-repeat protein; CDC20-like	2
45	At4g33400	Dem -like protein	2
46	At5g45700	Haloacid dehalogenase-like hydrolase (HAD)	2
47	At5g48310	RASD1/ RESPONSIVENESS TO ABA SALT AND DROUGHT 1	2
48	At1g20930	CDKB2;2	1
49	At1g23790	dicer-like protein (DUF936)	1
50	At1g33940	Serine/Threonine-kinase ULK4-like protein	1
51	At1g69400	mitotic checkpoint protein, putative	1
52	At2g28620	kinesin	1
53	At3g03130	lisH domain-like protein	1
54	At3g14190	expressed protein	1
55	At3g20150	kinesin	1
56	At3g26050	TPX2 (targeting protein for Xklp2) protein family	1
57	At3g60840	microtubule-associated protein; AtMAP65-4	1
58	At4g02800	GRIP/coiled-coil protein;	1
59	AT4G22860	TPXL3/ TPX2-LIKE PROTEIN 3	1
60	At4g26660	kinesin-like	1
61	At4g32830	Aurora-like kinase 2; ALK2	1
62	At5g06150	CYCB1;2	1
63	At5g55520	kinesin-like	1
64	At5g60930	kinesin	1
65	At1g02970	WEE1	-
66	At1g03740	CKL6	-
67	At1g09600	CKL11	-
68	At1g14750	SDS	-
69	At1g15570	CYCA2;3	-
70	At1g16330	CYCB3;1	-
71	At1g18040	CDKD;3	-

72	At1g18370	kinesin (AtNACK1); HINKEL	-
73	At1g18670	CKL3	-
74	At1g20590	CYCB2;5	-
75	At1g20610	CYCB2;3	-
76	At1g27630	CYCT;1	-
77	At1g33770	CKL14	-
78	At1g34460	CYCB1;5	-
79	At1g47210	CYCA3;2	-
80	At1g47220	CYCA3;3	-
81	At1g47230	CYCA3;4	-
82	At1g47870	E2Fc	-
83	At1g49620	KRP7	-
84	At1g53050	CKL15	-
85	At1g54610	CKL9	-
86	At1g57700	CKL10	-
87	At1g61450	unknown protein	-
88	At1g66750	CDKD;2	-
89	At1g67580	CDKG;2	-
90	At1g70210	CYCD1;1	-
91	At1g71530	CKL12	-
92	At1g73690	CDKD;1	-
93	At1g74330	CKL2	-
94	At1g75920	family II lipase EXL5 (EXL5)	-
95	At1g76540	CDKB2;1	-
96	At1g77390	CYCA1;2	-
97	At1g80370	CYCA2;4	-
98	At2g22490	CYCD2;1	-
99	At2g23430	KRP1	-
100	At2g25880	Aurora-like kinase 1; ALK1	-
101	At2g26430	CYCL;1	-
102	At2g27960	CKS1	-
103	At2g27970	CKS2	-
104	At2g32710	KRP4	-
105	At2g33560	expressed protein; BUBR1-like?	-
106	At2g36010	E2Fa	-
107	At2g38620	CDKB1;2	-
108	At3g01330	DEL3	-
109	At3g05050	CKL8	-
110	At3g12280	Rb	-

111	At3g24810	KRP5	-
112	At3g48160	DEL1	-
113	At3g50070	CYCD3;3	-
114	At3g50630	KRP2	-
115	At3g54180	CDKB1;1	-
116	At4g01730	DHHC-type zinc finger family protein;	-
117	At4g03270	CYCD6;1	-
118	At4g05190	kinesin	-
119	At4g10010	CKL13	-
120	At4g22860	TPXL3/ TPX2-LIKE PROTEIN 3	-
121	At4g22940	CKL4	-
122	At4g28230	Q9M0I4-1	-
123	At4g28980	CDKF;1	-
124	At4g34090	CYL1	-
125	At4g34160	CYCD3;1	-
126	At4g37490	CYCB1;1	-
127	At4g37630	CYCD5;1	-
128	At5g02110	CYCD7;1	-
129	At5g02470	DPa	-
130	At5g03415	DPb	-
131	At5g03455	CDC25	-
132	At5g10270	CDKC;1	-
133	At5g10440	CYCD4;2	-
134	At5g11300	CYCA2;2	-
135	At5g13840	CCS52b; cell cycle switch protein	-
136	At5g14960	DEL2	-
137	At5g17160	aspartic/glutamic acid-rich protein	-
138	At5g22220	E2Fb	-
139	At5g25380	CYCA2;1	-
140	At5g27620	CYCH;1	-
141	At5g33300	chromosome-associated kinesin-like protein	-
142	At5g39420	CKL1	-
143	At5g43080	CYCA3;1	-
144	At5g44290	CKL5	-
145	At5g48630	CYCC1;2	-
146	At5g48640	CYCC1;1	-
147	At5g48820	KRP3	-
148	At5g50860	CKL7	-
149	At5g62550	unknown protein	-

150	At5g63370	CDKG;1	-
151	At5g63610	CDKE;1	-
152	At5g64960	CDKC;2	-
153	At5g65420	CYCD4;1	-
154	At5g67260	CYCD3;2	-
155	At5g67270	microtubule-associated protein EB1-like protein	-

Tabla A2: AtBD_prom-5'UTR

	<i>A.thaliana</i> (ID)	MatInspector	PlantCARE	MAST	MCAST
1	At5g13840	-	4	2	2
2	At5g06150	5	3	3	3
3	At5g15510	5	2	3	3
4	At1g76310	2	2	3	3
5	At1g03780	4	2	2	2
6	At2g25060	4	2	2	2
7	At3g26050	1	2	2	2
8	At4g28230	1	2	2	2
9	At1g44110	3	2	1	1
10	At3g23670	3	2	1	1
11	At4g35620	3	2	1	1
12	At4g15830	2	2	1	1
13	At1g23790	1	2	1	1
14	At2g26760	4	1	2	2
15	At1g50490	3	1	2	2
16	At2g20590	2	1	2	2
17	At5g45700	2	1	2	2
18	At4g02800	1	1	2	2
19	At5g63370	-	1	2	2
20	At1g08560	3	1	1	1
21	At2g17620	3	1	1	1
22	At3g11520	3	1	1	1
23	At5g02370	3	1	1	1
24	At5g66230	3	1	1	1
25	At2g38160	2	1	1	1
26	At3g14190	1	1	1	1
27	At5g55520	1	1	1	1
28	At1g18040	-	1	1	1
29	At1g18370	-	1	1	1
30	At1g73690	-	1	1	1
31	At2g33560	-	1	1	1
32	At5g11300	-	1	1	1
33	At4g10010	-	1	-	-
34	At3g19590	3	-	2	2

35	At1g02730	6	-	1	1
36	At3g02640	4	-	1	1
37	At2g16270	2	-	1	1
38	At3g23890	2	-	1	1
39	At3g25980	2	-	1	1
40	At3g55660	2	-	1	1
41	At3g03130	1	-	1	1
42	At1g49620	-	-	1	1
43	At1g61450	-	-	1	1
44	At1g75920	-	-	1	1
45	At3g50630	-	-	1	1
46	At5g67270	-	-	1	1
47	At3g51280	4	-	-	-
48	At1g59540	3	-	-	-
49	At1g72250	3	-	-	-
50	At3g03090	3	-	-	-
51	At5g11510	3	-	-	-
52	At5g23910	3	-	-	-
53	At5g36710	3	-	-	-
54	At5g51600	3	-	-	-
55	At5g55830	3	-	-	-
56	At1g34355	2	-	-	-
57	At2g22610	2	-	-	-
58	At2g44190	2	-	-	-
59	At2g47500	2	-	-	-
60	At3g12870	2	-	-	-
61	At3g27330	2	-	-	-
62	At4g05520	2	-	-	-
63	At4g11080	2	-	-	-
64	At4g23800	2	-	-	-
65	At4g28430	2	-	-	-
66	At4g33260	2	-	-	-
67	At4g33400	2	-	-	-
68	At5g48310	2	-	-	-
69	At1g20930	1	-	-	-
70	At1g33940	1	-	-	-
71	At1g69400	1	-	-	-
72	At2g28620	1	-	-	-
73	At3g20150	1	-	-	-

74	At3g60840	1	-	-	-
75	AT4g22860	1	-	-	-
76	At4g26660	1	-	-	-
77	At4g32830	1	-	-	-
78	At5g60930	1	-	-	-
79	At1g02970	-	-	-	-
80	At1g03740	-	-	-	-
81	At1g09600	-	-	-	-
82	At1g14750	-	-	-	-
83	At1g15570	-	-	-	-
84	At1g16330	-	-	-	-
85	At1g18670	-	-	-	-
86	At1g20590	-	-	-	-
87	At1g20610	-	-	-	-
88	At1g27630	-	-	-	-
89	At1g33770	-	-	-	-
90	At1g34460	-	-	-	-
91	At1g47210	-	-	-	-
92	At1g47220	-	-	-	-
93	At1g47230	-	-	-	-
94	At1g47870	-	-	-	-
95	At1g53050	-	-	-	-
96	At1g54610	-	-	-	-
97	At1g57700	-	-	-	-
98	At1g66750	-	-	-	-
99	At1g67580	-	-	-	-
100	At1g70210	-	-	-	-
101	At1g71530	-	-	-	-
102	At1g74330	-	-	-	-
103	At1g76540	-	-	-	-
104	At1g77390	-	-	-	-
105	At1g80370	-	-	-	-
106	At2g22490	-	-	-	-
107	At2g23430	-	-	-	-
108	At2g25880	-	-	-	-
109	At2g26430	-	-	-	-
110	At2g27960	-	-	-	-
111	At2g27970	-	-	-	-
112	At2g32710	-	-	-	-

113	At2g36010	-	-	-	-
114	At2g38620	-	-	-	-
115	At3g01330	-	-	-	-
116	At3g05050	-	-	-	-
117	At3g12280	-	-	-	-
118	At3g24810	-	-	-	-
119	At3g48160	-	-	-	-
120	At3g50070	-	-	-	-
121	At3g54180	-	-	-	-
122	At4g01730	-	-	-	-
123	At4g03270	-	-	-	-
124	At4g05190	-	-	-	-
125	At4g22860	-	-	-	-
126	At4g22940	-	-	-	-
127	At4g28980	-	-	-	-
128	At4g34090	-	-	-	-
129	At4g34160	-	-	-	-
130	At4g37490	-	-	-	-
131	At4g37630	-	-	-	-
132	At5g02110	-	-	-	-
133	At5g02470	-	-	-	-
134	At5g03415	-	-	-	-
135	At5g03455	-	-	-	-
136	At5g10270	-	-	-	-
137	At5g10440	-	-	-	-
138	At5g14960	-	-	-	-
139	At5g17160	-	-	-	-
140	At5g22220	-	-	-	-
141	At5g25380	-	-	-	-
142	At5g27620	-	-	-	-
143	At5g33300	-	-	-	-
144	At5g39420	-	-	-	-
145	At5g43080	-	-	-	-
146	At5g44290	-	-	-	-
147	At5g48630	-	-	-	-
148	At5g48640	-	-	-	-
149	At5g48820	-	-	-	-
150	At5g50860	-	-	-	-
151	At5g62550	-	-	-	-

152	At5g63610	-	-	-	-
153	At5g64960	-	-	-	-
154	At5g65420	-	-	-	-
155	At5g67260	-	-	-	-

Tabla A4: AtBD_MSA

	<i>A.thaliana</i> (ID)	MatInspector	PlantCARE	MAST	MCAST
1	At5g13840	-	4	2	2
2	At1g76320	-	3	3	4
3	At5g06150	5	3	3	3
4	At5g15510	5	2	3	3
5	At1g76310	2	2	3	3
6	At3g05110	-	2	3	3
7	At1g77660	-	2	2	3
8	At2g25060	4	2	2	2
9	At1g03780	4	2	2	2
10	At3g26050	1	2	2	2
11	At4g28230	1	2	2	2
12	At1g79340	-	2	2	2
13	At5g65580	-	2	2	2
14	At1g44110	3	2	1	1
15	At4g35620	3	2	1	1
16	At3g23670	3	2	1	1
17	At4g15830	2	2	1	1
18	At1g23790	1	2	1	1
19	At2g26760	4	1	2	2
20	At1g50490	3	1	2	2
21	At2g20590	2	1	2	2
22	At5g45700	2	1	2	2
23	At4g02800	1	1	2	2
24	At5g63370	-	1	2	2
25	At3g11520	3	1	1	1
26	At2g17620	3	1	1	1
27	At1g08560	3	1	1	1
28	At5g02370	3	1	1	1
29	At5g66230	3	1	1	1
30	At2g38160	2	1	1	1
31	At5g55520	1	1	1	1
32	At3g14190	1	1	1	1
33	At1g73690	-	1	1	1

34	At1g18040	-	1	1	1
35	At5g11300	-	1	1	1
36	At1g18370	-	1	1	1
37	At2g33560	-	1	1	1
38	At4g10010	-	1		
39	At3g19590	3	-	2	2
40	At1g02730	6	-	1	1
41	At3g02640	4	-	1	1
42	At3g25980	2	-	1	1
43	At3g23890	2	-	1	1
44	At3g55660	2	-	1	1
45	At2g16270	2	-	1	1
46	At3g03130	1	-	1	1
47	At3g50630	-	-	1	1
48	At1g49620	-	-	1	1
49	At5g67270	-	-	1	1
50	At1g75920	-	-	1	1
51	At1g61450	-	-	1	1
52	At3g51150	-	-	1	-
53	At3g47750	-	-	1	-
54	At1g33930	-	-	-	2
55	At3g51280	4	-	-	-
56	At5g55830	3	-	-	-
57	At5g11510	3	-	-	-
58	At5g23910	3	-	-	-
59	At1g59540	3	-	-	-
60	At1g72250	3	-	-	-
61	At5g51600	3	-	-	-
62	At3g03090	3	-	-	-
63	At5g36710	3	-	-	-
64	At2g22610	2	-	-	-
65	At2g47500	2	-	-	-
66	At4g11080	2	-	-	-
67	At4g23800	2	-	-	-
68	At4g33400	2	-	-	-
69	At4g28430	2	-	-	-
70	At2g44190	2	-	-	-
71	At1g34355	2	-	-	-
72	At3g27330	2	-	-	-

73	At4g33260	2	-	-	-
74	At3g12870	2	-	-	-
75	At5g48310	2	-	-	-
76	At4g05520	2	-	-	-
77	At1g20930	1	-	-	-
78	At3g20150	1	-	-	-
79	At5g60930	1	-	-	-
80	At2g28620	1	-	-	-
81	At4g26660	1	-	-	-
82	At4g32830	1	-	-	-
83	At3g60840	1	-	-	-
84	At1g69400	1	-	-	-
85	At1g33940	1	-	-	-
86	At4g22860	1	-	-	-

Tabla A5: Genes de *P. vulgaris* homólogos en AtBD_MSA

	<i>P. vulgaris</i> (ID)	PlantCARE	MAST	MCAST
1	Phvul.003g072900	3	3	3
2	Phvul.001g119700	2	2	2
3	Phvul.003g244700	2	2	2
4	Phvul.004g165000	2	2	2
5	Phvul.008g200100	2	1	4
6	Phvul.001g226700	2	1	1
7	Phvul.009g115900	2	1	1
8	Phvul.001g071600	1	2	3
9	Phvul.002g119900	1	2	2
10	Phvul.004g034700	1	2	2
11	Phvul.001g241800	1	2	1
12	Phvul.005g071900	1	2	1
13	Phvul.002g106300	1	1	2
14	Phvul.004g113600	1	1	2
15	Phvul.003g002000	1	1	1
16	Phvul.004g152000	1	1	1
17	Phvul.006g087700	1	1	1
18	Phvul.008g142300	1	1	1
19	Phvul.011g020700	1	1	1
20	Phvul.001g055400	1	1	-
21	Phvul.002g015300	1	1	-
22	Phvul.006g059300	1	1	-
23	Phvul.007g263100	1	1	-
24	Phvul.008g281800	1	1	-
25	Phvul.008g289400	1	1	-
26	Phvul.007g044700	1	-	-
27	Phvul.003g295400	-	2	2
28	Phvul.011g125100	-	2	2
29	Phvul.001g130300	-	2	1
30	Phvul.004g034600	-	2	1
31	Phvul.001g232800	-	1	1
32	Phvul.003g082600	-	1	1

33	Phvul.003g265600	-	1	1
34	Phvul.004g084400	-	1	1
35	Phvul.009g097500	-	1	1
36	Phvul.011g101700	-	1	1
37	Phvul.001g258100	-	1	-
38	Phvul.001g261200	-	1	-
39	Phvul.003g262100	-	1	-
40	Phvul.004g077900	-	1	-
41	Phvul.004g080000	-	1	-
42	Phvul.006g052700	-	1	-
43	Phvul.006g171200	-	1	-
44	Phvul.008g062600	-	1	-
45	Phvul.008g272600	-	1	-
46	Phvul.001g000500	-	-	-
47	Phvul.001g004200	-	-	-
48	Phvul.001g004300	-	-	-
49	Phvul.001g024200	-	-	-
50	Phvul.001g029600	-	-	-
51	Phvul.001g052600	-	-	-
52	Phvul.001g061200	-	-	-
53	Phvul.001g119800	-	-	-
54	Phvul.001g151000	-	-	-
55	Phvul.001g191500	-	-	-
56	Phvul.001g198300	-	-	-
57	Phvul.001g210500	-	-	-
58	Phvul.001g234200	-	-	-
59	Phvul.001g238500	-	-	-
60	Phvul.001g258800	-	-	-
61	Phvul.001g264500	-	-	-
62	Phvul.002g019500	-	-	-
63	Phvul.002g034800	-	-	-
64	Phvul.002g037700	-	-	-
65	Phvul.002g071800	-	-	-
66	Phvul.002g078300	-	-	-
67	Phvul.002g115100	-	-	-
68	Phvul.002g145000	-	-	-
69	Phvul.002g152300	-	-	-
70	Phvul.002g153000	-	-	-
71	Phvul.002g175700	-	-	-

72	Phvul.002g214900	-	-	-
73	Phvul.002g289700	-	-	-
74	Phvul.002g296500	-	-	-
75	Phvul.002g296700	-	-	-
76	Phvul.002g326800	-	-	-
77	Phvul.003g009000	-	-	-
78	Phvul.003g078900	-	-	-
79	Phvul.003g119700	-	-	-
80	Phvul.003g127400	-	-	-
81	Phvul.003g209300	-	-	-
82	Phvul.003g213700	-	-	-
83	Phvul.003g216700	-	-	-
84	Phvul.003g241500	-	-	-
85	Phvul.003g264900	-	-	-
86	Phvul.003g293500	-	-	-
87	Phvul.004g113500	-	-	-
88	Phvul.004g127100	-	-	-
89	Phvul.004g154300	-	-	-
90	Phvul.004g157000	-	-	-
91	Phvul.005g017100	-	-	-
92	Phvul.005g024200	-	-	-
93	Phvul.005g046100	-	-	-
94	Phvul.005g054500	-	-	-
95	Phvul.005g101200	-	-	-
96	Phvul.005g168700	-	-	-
97	Phvul.006g048500	-	-	-
98	Phvul.006g057900	-	-	-
99	Phvul.006g134011	-	-	-
100	Phvul.006g144500	-	-	-
101	Phvul.006g160500	-	-	-
102	Phvul.006g220500	-	-	-
103	Phvul.007g044000	-	-	-
104	Phvul.007g078200	-	-	-
105	Phvul.007g107300	-	-	-
106	Phvul.007g111500	-	-	-
107	Phvul.007g159100	-	-	-
108	Phvul.007g179900	-	-	-
109	Phvul.007g183600	-	-	-
110	Phvul.007g199000	-	-	-

111	Phvul.007g217400	-	-	-
112	Phvul.007g262900	-	-	-
113	Phvul.008g051600	-	-	-
114	Phvul.008g076000	-	-	-
115	Phvul.008g113800	-	-	-
116	Phvul.008g117700	-	-	-
117	Phvul.008g117800	-	-	-
118	Phvul.008g125600	-	-	-
119	Phvul.008g181800	-	-	-
120	Phvul.008g203300	-	-	-
121	Phvul.008g228400	-	-	-
122	Phvul.008g239600	-	-	-
123	Phvul.008g253500	-	-	-
124	Phvul.008g271300	-	-	-
125	Phvul.008g274700	-	-	-
126	Phvul.008g279300	-	-	-
127	Phvul.009g033300	-	-	-
128	Phvul.009g065000	-	-	-
129	Phvul.009g067100	-	-	-
130	Phvul.009g086700	-	-	-
131	Phvul.009g090100	-	-	-
132	Phvul.009g094200	-	-	-
133	Phvul.009g106700	-	-	-
134	Phvul.009g112600	-	-	-
135	Phvul.009g128200	-	-	-
136	Phvul.009g205100	-	-	-
137	Phvul.009g242700	-	-	-
138	Phvul.009g248300	-	-	-
139	Phvul.010g005400	-	-	-
140	Phvul.010g015800	-	-	-
141	Phvul.010g048600	-	-	-
142	Phvul.010g098700	-	-	-
143	Phvul.010g148900	-	-	-
144	Phvul.010g151900	-	-	-
145	Phvul.011g012300	-	-	-
146	Phvul.011g020100	-	-	-
147	Phvul.011g021300	-	-	-
148	Phvul.011g023800	-	-	-
149	Phvul.011g025500	-	-	-

150	Phvul.011g061800	-	-	-
151	Phvul.011g070000	-	-	-
152	Phvul.011g101500	-	-	-
153	Phvul.011g125500	-	-	-
154	Phvul.011g180300	-	-	-
155	Phvul.011g196700	-	-	-
156	Phvul.011g211500	-	-	-
157	Phvul.005g139100	-	-	-
158	Phvul.007g133100	-	-	-
159	Phvul.005g149300	-	-	-

Tabla A6. Genes en PvBD_Up

	<i>P. vulgaris</i> (ID)	MAST	MCAST
1	Phvul.008g006700	4	5
2	Phvul.003g123000	2	3
3	Phvul.005g172300	2	3
4	Phvul.002g132500	2	2
5	Phvul.002g268100	2	2
6	Phvul.005g086300	2	2
7	Phvul.009g123000	2	2
8	Phvul.009g236800	2	2
9	Phvul.010g143900	2	2
10	Phvul.006g029200	2	1
11	Phvul.001g096900	1	2
12	Phvul.002g094500	1	2
13	Phvul.003g145900	1	2
14	Phvul.007g0419001	1	2
15	Phvul.001g075000	1	1
16	Phvul.001g183900	1	1
17	Phvul.005g179400	1	1
18	Phvul.007g096300	1	1
19	Phvul.010g133100	1	1
20	Phvul.006g161200	-	3
21	Phvul.008g171000	-	2
22	Phvul.001g061300	-	1
23	Phvul.002g010100	-	1
24	Phvul.003g082500	-	1
25	Phvul.003g237900	-	1
26	Phvul.003g289800	-	1
27	Phvul.007g060700	-	1
28	Phvul.007g175000	-	1

Tabla A7. Análisis de PvBD_prom-5'UTR

	<i>P. vulgaris</i> (ID)	PlantCARE	MAST	MCAST
1	Phvul.003g072900	3	3	3
2	Phvul.009g236800	3	2	2
3	Phvul.003g123000	2	2	3
4	Phvul.001g119700	2	2	2
5	Phvul.003g244700	2	2	2
6	Phvul.004g165000	2	2	2
7	Phvul.002g132500	2	2	2
8	Phvul.009g123000	2	2	2
9	Phvul.006g029200	2	2	1
10	Phvul.008g200100	2	1	4
11	Phvul.002g094500	2	1	2
12	Phvul.001g226700	2	1	1
13	Phvul.009g115900	2	1	1
14	Phvul.001g071600	1	2	3
15	Phvul.002g119900	1	2	2
16	Phvul.004g034700	1	2	2
17	Phvul.002g268100	1	2	2
18	Phvul.005g086300	1	2	2
19	Phvul.010g143900	1	2	2
20	Phvul.001g241800	1	2	1
21	Phvul.005g071900	1	2	1
22	Phvul.002g106300	1	1	2
23	Phvul.004g113600	1	1	2
24	Phvul.001g096900	1	1	2
25	Phvul.003g145900	1	1	2
26	Phvul.007g0419001	1	1	2
27	Phvul.003g002000	1	1	1
28	Phvul.004g152000	1	1	1
29	Phvul.006g087700	1	1	1
30	Phvul.008g142300	1	1	1

31	Phvul.011g020700	1	1	1
32	Phvul.001g183900	1	1	1
33	Phvul.005g179400	1	1	1
34	Phvul.007g096300	1	1	1
35	Phvul.010g133100	1	1	1
36	Phvul.001g055400	1	1	-
37	Phvul.002g015300	1	1	-
38	Phvul.006g059300	1	1	-
39	Phvul.007g263100	1	1	-
40	Phvul.008g281800	1	1	-
41	Phvul.008g289400	1	1	-
42	Phvul.007g044700	1	-	-
43	Phvul.008g006700	-	4	5
44	Phvul.005g172300	-	2	3
45	Phvul.003g295400	-	2	2
46	Phvul.011g125100	-	2	2
47	Phvul.001g130300	-	2	1
48	Phvul.004g034600	-	2	1
49	Phvul.001g232800	-	1	1
50	Phvul.003g082600	-	1	1
51	Phvul.003g265600	-	1	1
52	Phvul.004g084400	-	1	1
53	Phvul.009g097500	-	1	1
54	Phvul.011g101700	-	1	1
55	Phvul.001g075000	-	1	1
56	Phvul.001g258100	-	1	-
57	Phvul.001g261200	-	1	-
58	Phvul.003g262100	-	1	-
59	Phvul.004g077900	-	1	-
60	Phvul.004g080000	-	1	-
61	Phvul.006g052700	-	1	-
62	Phvul.006g171200	-	1	-
63	Phvul.008g062600	-	1	-
64	Phvul.008g272600	-	1	-
65	Phvul.006g161200	-	-	3
66	Phvul.008g171000	-	-	2
67	Phvul.001g061300	-	-	1
68	Phvul.002g010100	-	-	1
69	Phvul.003g082500	-	-	1

70	Phvul.003g237900	-	-	1
71	Phvul.003g289800	-	-	1
72	Phvul.007g060700	-	-	1
73	Phvul.007g175000	-	-	1
74	Phvul.001g000500	-	-	-
75	Phvul.001g004200	-	-	-
76	Phvul.001g004300	-	-	-
77	Phvul.001g024200	-	-	-
78	Phvul.001g029600	-	-	-
79	Phvul.001g052600	-	-	-
80	Phvul.001g061200	-	-	-
81	Phvul.001g119800	-	-	-
82	Phvul.001g151000	-	-	-
83	Phvul.001g191500	-	-	-
84	Phvul.001g198300	-	-	-
85	Phvul.001g210500	-	-	-
86	Phvul.001g234200	-	-	-
87	Phvul.001g238500	-	-	-
88	Phvul.001g258800	-	-	-
89	Phvul.001g264500	-	-	-
90	Phvul.002g019500	-	-	-
91	Phvul.002g034800	-	-	-
92	Phvul.002g037700	-	-	-
93	Phvul.002g071800	-	-	-
94	Phvul.002g078300	-	-	-
95	Phvul.002g115100	-	-	-
96	Phvul.002g145000	-	-	-
97	Phvul.002g152300	-	-	-
98	Phvul.002g153000	-	-	-
99	Phvul.002g175700	-	-	-
100	Phvul.002g214900	-	-	-
101	Phvul.002g289700	-	-	-
102	Phvul.002g296500	-	-	-
103	Phvul.002g296700	-	-	-
104	Phvul.002g326800	-	-	-
105	Phvul.003g009000	-	-	-
106	Phvul.003g078900	-	-	-
107	Phvul.003g119700	-	-	-
108	Phvul.003g127400	-	-	-

109	Phvul.003g209300	-	-	-
110	Phvul.003g213700	-	-	-
111	Phvul.003g216700	-	-	-
112	Phvul.003g241500	-	-	-
113	Phvul.003g264900	-	-	-
114	Phvul.003g293500	-	-	-
115	Phvul.004g113500	-	-	-
116	Phvul.004g127100	-	-	-
117	Phvul.004g154300	-	-	-
118	Phvul.004g157000	-	-	-
119	Phvul.005g017100	-	-	-
120	Phvul.005g024200	-	-	-
121	Phvul.005g046100	-	-	-
122	Phvul.005g054500	-	-	-
123	Phvul.005g101200	-	-	-
124	Phvul.005g168700	-	-	-
125	Phvul.006g048500	-	-	-
126	Phvul.006g057900	-	-	-
127	Phvul.006g134011	-	-	-
128	Phvul.006g144500	-	-	-
129	Phvul.006g160500	-	-	-
130	Phvul.006g220500	-	-	-
131	Phvul.007g044000	-	-	-
132	Phvul.007g078200	-	-	-
133	Phvul.007g107300	-	-	-
134	Phvul.007g111500	-	-	-
135	Phvul.007g159100	-	-	-
136	Phvul.007g179900	-	-	-
137	Phvul.007g183600	-	-	-
138	Phvul.007g199000	-	-	-
139	Phvul.007g217400	-	-	-
140	Phvul.007g262900	-	-	-
141	Phvul.008g051600	-	-	-
142	Phvul.008g076000	-	-	-
143	Phvul.008g113800	-	-	-
144	Phvul.008g117700	-	-	-
145	Phvul.008g117800	-	-	-
146	Phvul.008g125600	-	-	-
147	Phvul.008g181800	-	-	-

148	Phvul.008g203300	-	-	-
149	Phvul.008g228400	-	-	-
150	Phvul.008g239600	-	-	-
151	Phvul.008g253500	-	-	-
152	Phvul.008g271300	-	-	-
153	Phvul.008g274700	-	-	-
154	Phvul.008g279300	-	-	-
155	Phvul.009g033300	-	-	-
156	Phvul.009g065000	-	-	-
157	Phvul.009g067100	-	-	-
158	Phvul.009g086700	-	-	-
159	Phvul.009g090100	-	-	-
160	Phvul.009g094200	-	-	-
161	Phvul.009g106700	-	-	-
162	Phvul.009g112600	-	-	-
163	Phvul.009g128200	-	-	-
164	Phvul.009g205100	-	-	-
165	Phvul.009g242700	-	-	-
166	Phvul.009g248300	-	-	-
167	Phvul.010g005400	-	-	-
168	Phvul.010g015800	-	-	-
169	Phvul.010g048600	-	-	-
170	Phvul.010g098700	-	-	-
171	Phvul.010g148900	-	-	-
172	Phvul.010g151900	-	-	-
173	Phvul.011g012300	-	-	-
174	Phvul.011g020100	-	-	-
175	Phvul.011g021300	-	-	-
176	Phvul.011g023800	-	-	-
177	Phvul.011g025500	-	-	-
178	Phvul.011g061800	-	-	-
179	Phvul.011g070000	-	-	-
180	Phvul.011g101500	-	-	-
181	Phvul.011g125500	-	-	-
182	Phvul.011g180300	-	-	-
183	Phvul.011g196700	-	-	-
184	Phvul.011g211500	-	-	-
185	Phvul.L002157	-	-	-
186	Phvul.L002172	-	-	-

187	Phvul.L010943	-	-	-
-----	---------------	---	---	---

Tabla A8. Posición de los motivos MSA con respecto al TSS (+1)

ID <i>P. vulgaris</i>	Número de MSA	Posición con respecto a TSS (+1)
Phvul.003g009000	1	-14
Phvul.004g154300	1	-71
Phvul.001g096900	1	420
Phvul.001g183900	1	-37
Phvul.002g094500	2	-263
		-265
Phvul.002g132500	2	-710
		-507
Phvul.002g268100	1	-101
Phvul.003g123000	2	-123
		-180
Phvul.003g145900	1	-57
Phvul.005g086300	1	-157
Phvul.005g172300	1	-23
Phvul.005g179400	1	1
Phvul.006g029200	2	50
		-63
Phvul.007g041900	1	7
Phvul.007g096300	1	-102
Phvul.009g123000	2	-57
		-93
Phvul.009g236800	3	54
		56
		30
Phvul.010g133100	1	-972
Phvul.010g143900	1	-122

Phvul.001g119800	2	42
		-1821
Phvul.002g145000	2	-39
		-13
Phvul.004g077900	2	-80
		-52
Phvul.002g115100	1	-53
Phvul.001g119700	2	-93
		-17
Phvul.008g203300	1	-68
Phvul.008g181800	1	-128
Phvul.009g128200	1	-282
Phvul.001g061200	1	-112
Phvul.006g087700	1	-302
Phvul.006g057900	1	-1341
Phvul.011g021300	1	-61
Phvul.005g101200	2	-1177
		-17
Phvul.001g232800	1	-68
Phvul.001g258100	2	-817
		-452
Phvul.005g017100	2	-199
		-138
Phvul.008g076000	1	-619
Phvul.009g033300	1	-34
Phvul.011g125100	1	-112
Phvul.004g080000	1	-361
Phvul.001g151000	2	-1575
		-1295
Phvul.004g084400	1	-22
Phvul.001g258800	1	-109
Phvul.003g119700	1	-560
Phvul.009g106700	1	246
Phvul.001g234200	1	-1462
Phvul.003g262100	2	-86
		-27
Phvul.006g134011	1	1
Phvul.001g264500	1	38
Phvul.002g019500	1	-136
Phvul.003g078900	3	-31
		-3

		16
Phvul.004g127100	1	20
Phvul.004g034700	2	-1313
		-100
Phvul.004g113500	1	2
Phvul.003g264900	1	-40
Phvul.004g034600	2	24
		65
Phvul.003g293500	1	-142
Phvul.008g274700	1	-140
Phvul.008g051600	1	-114
Phvul.008g289400	1	-75
Phvul.006g220500	1	95
Phvul.011g125500	2	12
		27
Phvul.001g075000	1	-404
Phvul.008g006700	4	129
		155
		181
		233
Phvul.001g061300	1	-1906
Phvul.002g010100	1	-1670
Phvul.003g082500	1	-984
Phvul.003g237900	1	-1241
Phvul.003g289800	1	-957
Phvul.006g161200	3	-1980
		-1967
		-1915
Phvul.007g060700	1	-998
Phvul.007g175000	1	-949
Phvul.008g171000	2	-1091
		-1071
Phvul.007g078200	1	-1481

Tabla A9. Dominios Pfam proteínas genes PvDB-MSA

	<i>P. vulgaris</i> (ID)	Pfam
1	Phvul.001g061200	Myb-like DNA-binding domain
2	Phvul.001g061300	Protein phosphatase 2C
3	Phvul.001g075000	alpha/beta hydrolase fold
4	Phvul.001g096900	GDA1/CD39 (nucleoside phosphatase) family
5	Phvul.001g119700	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
6	Phvul.001g119800	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
7	Phvul.001g151000	UNCHARACTERIZED
8	Phvul.001g183900	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)/ Poly-adenylate binding protein, unique domain
9	Phvul.001g232800	Eukaryotic translation initiation factor eIF2A
10	Phvul.001g234200	Protein kinase domain/ Legume lectin domain
11	Phvul.001g258100	kinesin-like protein
12	Phvul.001g258800	HORMA domain
13	Phvul.001g264500	Kinesin motor domain/ Kinesin-related
14	Phvul.002g010100	Uncharacterized conserved protein (DUF2215)
15	Phvul.002g019500	Cyclin-dependent kinase inhibitor
16	Phvul.002g094500	Cytochrome P450
17	Phvul.002g115100	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
18	Phvul.002g132500	Myb-like DNA-binding domain/ MYB-CC type transfactor, LHEQLE motif
19	Phvul.002g145000	Kinesin motor domain/ Di-glucose binding within endoplasmic reticulum
20	Phvul.002g268100	UNCHARACTERIZED
21	Phvul.003g009000	Mad3/BUB1 homology region 1
22	Phvul.003g078900	UNCHARACTERIZED
23	Phvul.003g082500	Tyrosine phosphatase family/Tyrosine phosphatase family
24	Phvul.003g119700	Kinesin motor domain/ Kinesin motor

25	Phvul.003g123000	UNCHARACTERIZED
26	Phvul.003g145900	Centromere kinetochore component CENP-T histone fold
27	Phvul.003g237900	Alpha/beta hydrolase family
28	Phvul.003g262100	UNCHARACTERIZED
29	Phvul.003g264900	TPR repeat
30	Phvul.003g289800	GRAM domain
31	Phvul.003g293500	Kinesin motor domain/ Kinesin motor
32	Phvul.004g034600	Targeting protein for Xklp2 (TPX2)
33	Phvul.004g034700	Targeting protein for Xklp2 (TPX2)
34	Phvul.004g077900 KNOLLE	Syntaxin/ SNARE domain
35	Phvul.004g080000	WD domain, G-beta repeat
36	Phvul.004g084400	FHA domain/ PIN domain
37	Phvul.004g113500	Kinesin motor domain/ Helix-hairpin-helix motif
38	Phvul.004g127100	Microtubule associated protein (MAP65/ASE1 family)
39	Phvul.004g154300	Kinesin motor domain/ Domain of unknown function (DUF3490)
40	Phvul.005g017100	UNCHARACTERIZED
41	Phvul.005g086300	Cotton fibre expressed protein/ Domain of unknown function (DUF4408)
42	Phvul.005g101200	Reticulon
43	Phvul.005g172300	Glycosyl transferases group 1/ Glycosyl-transferase family 4
44	Phvul.005g179400	Thioredoxin-like [2Fe-2S] ferredoxin
45	Phvul.006g029200	bZIP transcription factor/ Zinc-binding RING-finger
46	Phvul.006057900	Protein kinase domain
47	Phvul.006g087700	Protein kinase domain/ Legume lectin domain
48	Phvul.006g134011	Family of unknown function (DUF566)
49	Phvul.006g161200	Glycosyl hydrolase family 14
50	Phvul.006g220500.2	Protein kinase domain
51	Phvul.007g0419001	B3 DNA binding domain
52	Phvul.007g060700	Ribosomal S3Ae family
53	Phvul.007g078200	Protein kinase domain/ Legume lectin domain
54	Phvul.007g096300	Aromatic amino acid lyase
55	Phvul.007g175000	Cupin superfamily protein
56	Phvul.008g006700	Ribosomal RNA-processing protein 7 (RRP7)
57	Phvul.008g051600	UNCHARACTERIZED
58	Phvul.008g076000	Microtubule associated protein (MAP65/ASE1 family)
59	Phvul.008g171000	UNCHARACTERIZED
60	Phvul.008g181800	Kinesin motor domain
61	Phvul.008g203300	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
62	Phvul.008g274700	Kinesin motor domain/ Kinesin-related
63	Phvul.008g289400	UNCHARACTERIZED

64	Phvul.009g033300	UNCHARACTERIZED
65	Phvul.009g106700	Myb-like DNA-binding domain
66	Phvul.009g123000	MORN repeat
67	Phvul.009g128200	Cyclin, N-terminal domain/ Cyclin, C-terminal domain
68	Phvul.009g236800	Subtilase family/ PA domain/ Peptidase inhibitor I9
69	Phvul.010g133100	TGS domain
70	Phvul.010g143900	Phosphofructokinase
71	Phvul.011g021300	UNCHARACTERIZED
72	Phvul.011g125100	Reticulon
73	Phvul.011g125500	Targeting protein for Xklp2 (TPX2)/ Cell cycle regulated microtubule associated protein