

UNIVERSIDAD DE SALAMANCA

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE MICROBIOLOGÍA Y GENÉTICA

(ÁREA DE FISIOLOGÍA VEGETAL)

INSTITUTO HISPANOLUSO DE INVESTIGACIONES AGRARIAS (CIALE)



Caracterización de la familia ST en *Medicago truncatula*, un modelo para el estudio de esta nueva clase de proteínas vegetales

Lucía Albornos Llorente

2015

ÍNDICE

1	Introducción	2
2	Planteamiento del trabajo	8
3	Resultados.....	10
	Sección 1: Caracterización de los genes y proteínas de la familia ST	10
	Sección 2: Localización subcelular de las proteínas ST de <i>Medicago truncatula</i>	12
	Sección 3: Análisis de la actividad de los promotores ST de <i>Medicago truncatula</i>	13
	Sección 4: Estudio de la acumulación de los transcritos de los genes ST de <i>Medicago truncatula</i>	17
4	Discusión	19
5	Conclusiones	28
6	Bibliografía	30

1 Introducción

Las proteínas ST, son una familia de proteínas con secuencias repetidas en tándem que se han descrito en la familia de las fabáceas (De Vries y col., 1983; De Vries y col., 1985; Williams y col., 1990; Muñoz y col., 1997). La secuencia *PsaST2* de *Pisum sativum* fue el primer cDNA de tipo *ST* descrito en la bibliografía y se identificó como un transcrito específico de tallos (*stem specific*, lo que da nombre a la familia *ST*) (De Vries y col., 1983).

Estructuralmente, las proteínas *ST* están formadas por 3 zonas: una región N-terminal hidrofóbica que determina un péptido señal, lo que indica que entran en la vía de secreción; una zona de aminoácidos no repetida; y una región de mayor tamaño, con una secuencia de aminoácidos que se repite en tándem (De Vries y col., 1985; Williams y col., 1990; Muñoz y col., 1997). Dichas secuencias tienen 25 o 26 aminoácidos y se repiten un número variable de veces en las diferentes proteínas codificadas por los genes *ST*. Así, en las dos proteínas de *P. sativum* previamente descritas aparecen 4 veces en *PsaST2* y 2 en *PsaST3* (De Vries y col., 1985; Williams y col., 1990), mientras que en *Cicer arietinum* aparecen 7 veces en *CarST1* y 8 en *CarST2* (Muñoz y col., 1997). Las secuencias que se repiten presentan una elevada similitud entre sí y todas ellas contienen la secuencia de aminoácidos [-(D/E)FEP RP----Y-], apareciendo en algunos casos en la zona entre la segunda Pro y la Tyr un posible sitio de N-glicosilación (Asn-X-Ser) (De Vries y col., 1985; Williams y col., 1990; Muñoz y col., 1997).

La presencia de repeticiones en tándem es una de las características estructurales de estas proteínas *ST*. Actualmente se piensa que aproximadamente el 14 % de las proteínas

contienen secuencias repetidas, siendo más abundantes en eucariotas que en procariotas (Marcotte y col., 1999). Algunos autores sostienen que estas repeticiones pueden determinar la estructura y función de las proteínas (Katti y col., 2000; Kajava, 2012).

Las proteínas con repeticiones en tándem presentan características muy variables, tanto en la secuencia de la repetición como en el número y disposición de las repeticiones en la proteína, por lo que se han definido diferentes categorías que facilitan su estudio. Katti y col. (2000) establecieron tres grupos de proteínas con oligopéptidos repetidos atendiendo a la longitud del oligopéptido, si bien, más recientemente, como consecuencia del aumento de proteínas cuya estructura secundaria y terciaria se ha determinado, Kajava (2012) ha establecido 5 clases de proteínas con repeticiones en tándem en función de su conformación tridimensional.

En la Clase III de Kajava (2012) se encuentran proteínas con estructuras alargadas donde las unidades repetidas se necesitan unas a otras para mantener la conformación. En esta Clase, las proteínas presentan repeticiones de entre 5 y 40 aminoácidos que pueden adoptar plegamientos tipo solenoide o no solenoide. Las repeticiones se acumulan en un número indefinido y la superestructura que se forma va adquiriendo cierto grado de curvatura. Dentro de este grupo se encuentran las proteínas con repeticiones ankirina (Mosavi y col., 2002), armadillo (Peifer y col., 1994), LRR (ricas en leucina) (Kobe y Deisenhofer, 1994; Kajava, 1998), TPR (*tetratricopeptide repeat*) (Sikorski y col., 1990), PPR (*pentatricopeptide repeat*) (Small y Peeters, 2000), las proteínas anticongelantes con repeticiones IRI (Davies, 2014) y las EPR (*icosapentapeptide repeat*) (Archak y Nagaraju, 2006). Todas estas proteínas presentan una clara diversidad estructural, aunque lo más llamativo, es su enorme diversidad funcional, que frecuentemente deriva de la presencia de otros dominios además de las regiones repetidas y de la capacidad para establecer interacciones proteína-proteína.

Aunque la estructura general de las proteínas ST no se ha determinado por el momento, atendiendo a la longitud de las repeticiones de las de guisante y garbanzo (De Vries y col., 1985; Williams y col., 1990; Muñoz y col., 1997), estas proteínas podrían incluirse en esta Clase III. Sin embargo, las proteínas ST de guisante y garbanzo, a diferencia de las proteínas de esta Clase, muestran un elevado grado de conservación en la secuencia que se repite (De Vries y col., 1985; Muñoz y col., 1997). Así, por el grado de perfección de sus repeticiones, las ST podrían ser un tipo de proteínas intrínsecamente desestructuradas (IUP) ya que Jordá y col. (2010) demostraron que las proteínas con repeticiones en tándem perfectas o casi perfectas tienen tendencia a ser IUP. Como consecuencia de la falta de estructura

tridimensional definida tienen una flexibilidad y plasticidad estructural que constituye una gran ventaja funcional, permitiendo su unión con un amplio conjunto de ligandos como membranas, proteínas, ácidos nucleicos o pequeñas moléculas (Uversky, 2010). Algunas IUP participan en procesos tan relevantes como la transducción de señales (proteínas GRAS), la regulación de la transcripción (factores de transcripción de la familia NAC o bZip), la percepción de la luz (proteínas CRY) o la respuesta al estrés abiótico (proteínas LEA) (Sun y col., 2013).

Las proteínas ST tienen un péptido señal que indica su posible localización en la pared celular, por lo que podría tratarse de algún tipo de proteína estructural de la misma (Albornos y col., 2014), si bien estas pertenecen a la Clase II de Kajava (2012) y presentan repeticiones de 2 a 5 aminoácidos que no coinciden con las de las ST. El trabajo realizado por Albornos y col. (2014) determinó que las ST no pertenecen a ninguno de los siguientes grupos de proteínas de pared celular: ricas en hidroxiprolina, ricas en prolina, ricas en glicina, ni arabinogalactano proteínas. A pesar de ello, las proteínas ST de garbanzo sí se encuentran en la pared celular, bien unidas débilmente por enlaces iónicos o solubles en el apoplasto (Albornos y col., 2014), aunque también se han encontrado en el interior de la célula (Albornos, 2010; Albornos y col., 2014).

Se desconoce la función de las proteínas ST en el desarrollo de la planta aunque se han relacionado fundamentalmente con los procesos de crecimiento (De Vries y col., 1985; Williams y col., 1990; Muñoz y col., 1997; Hernández-Nistal y col., 2006) y con el desarrollo de frutos y semillas (Williams y col., 1990; Waters y col., 2005; Levi y col., 2006; Fernández y col., 2007; Wechter y col., 2008; Albornos y col., 2014). Además se ha establecido su intervención en la asociación de la planta con hongos micorrícicos (Wulf y col., 2003; Hohnjec y col., 2005; Liu y col., 2007).

La función de las proteínas ST en el crecimiento celular ha sido la más estudiada hasta el momento ya que, en un primer momento, el perfil de acumulación de los transcritos que codifican las proteínas CarST1 y CarST2 de garbanzo y PsaST2 de guisante, las relacionaba con este proceso en epicotilos y tallos aunque no en raíces (De Vries y col., 1985; Williams y col., 1990; Muñoz y col., 1997). Al profundizar en esta función se vio que los transcritos ST no se detectaban en los ejes embrionarios de guisante (De Vries y col., 1985) ni de garbanzo (Hernández-Nistal y col., 2006) durante la germinación en sentido estricto, apareciendo al comenzar la elongación del epicotilo y alcanzando su máxima acumulación en el momento en el que este órgano se encuentra en crecimiento exponencial (De Vries y col., 1985;

Muñoz y col., 1997; Hernández-Nistal y col., 2006). La relación entre la acumulación de los transcritos *ST* y la tasa de crecimiento no se encontró únicamente en los epicotilos de plántulas etioladas sino también en los tallos de plantas de más edad, al analizar la cantidad de transcritos que presentaban los diferentes entrenudos. Así, los transcritos *PsaST2* (Williams y col., 1990) y *CarST1* (Muñoz y col., 1997) eran más abundantes en los entrenudos más apicales, con mayor tasa de crecimiento, que en los entrenudos más basales, reforzando la relación de las proteínas *ST* con este proceso. Esta relación entre la abundancia de mRNA y la tasa de crecimiento no se pudo establecer para los transcritos del gen *CarST2* (Muñoz y col., 1997).

Aunque todas las hormonas vegetales se han asociado de un modo u otro con la regulación del crecimiento vegetal, solo auxinas, citoquininas (CK), giberelinas (GA) y brasinólidos (BL) se han considerado esenciales para este proceso (Depuydt y Hardtke, 2011). Otras hormonas, como el ácido abscísico (ABA), se han relacionado con la reducción en la tasa de crecimiento que se produce cuando la planta se enfrenta a algún tipo de estrés, como la sequía, la salinidad o las altas temperaturas. El análisis del efecto de estas hormonas en la acumulación de los transcritos *ST* de garbanzo indicó que su incremento durante el crecimiento no se encontraba mediado por auxinas o BL, ya que cuando se expusieron los epicotilos a estas hormonas se redujeron los niveles de los transcritos *ST* (Muñoz y col., 1997). Estos estudios se complementaron con el análisis de los transcritos en condiciones de inhibición del crecimiento del epicotilo provocada por ácido abscísico (Hernández-Nistal y col., 2010) o por diferentes condiciones de estrés abiótico (Muñoz y col., 1997; Hernández-Nistal y col., 2010), encontrándose en todos los casos que la inhibición del crecimiento disminuía los transcritos *ST* hasta hacerlos prácticamente indetectables en algunos casos (Muñoz y col., 1997; Hernández-Nistal y col., 2010), obteniéndose el mismo efecto que el provocado por las hormonas que inducen el crecimiento. Estos resultados no permitieron confirmar que las proteínas *ST* estén relacionadas con el proceso de elongación, si bien no se puede descartar que participen en otros aspectos del crecimiento.

Más recientemente, la inmunolocalización de la proteína *CarST1* en distintos órganos de la planta a lo largo del desarrollo también respalda la posible función de esta proteína en el crecimiento (Albornos y col., 2014), ya que esta proteína es más abundante en las secciones apicales de radículas y epicotilos de plántulas etioladas de 4 d que en las secciones centrales y basales, cuya tasa de elongación es menor (Albornos y col., 2014). La localización de esta proteína en las fibras perivasculares de los epicotilos, una zona de acumulación de reservas en tejidos jóvenes en crecimiento (Bowes, 1996), sugería una posible función como proteína

de reserva vegetativa (VSP) (Albornos y col., 2014) que pudiera servir como fuente de N₂ durante el crecimiento de la planta y que no estuviese directamente relacionada con el proceso de elongación. Las VSP son un grupo de proteínas muy heterogéneo que no muestran similitud en la secuencia de aminoácidos ni en la estructura de la proteína y no parecen estar relacionadas filogenéticamente entre sí (Cooke y Weih, 2005), lo que hace difícil determinar por comparación de secuencias si una proteína de función desconocida, como las ST, pertenece a este grupo. Algunas proteínas VSP presentan actividad enzimática y su función puede estar relacionada con la misma (Andrews y col., 1988; Spilatro y Anderson, 1989; Selengut, 2001; Meuriot y col., 2004; Tian y col., 2007). Las características de las VSP no permiten descartar que la proteína CarST1 de garbanzo pertenezca a este grupo de proteínas, ya que comparte algunas características con las VSP de soja (Albornos y col., 2014). Sin embargo, la CarST1 no parece tener ninguna actividad enzimática ni tampoco se ha estudiado si la acumulación de los transcritos y las proteínas responden al estatus de N₂ en los distintos órganos de la planta.

Las ST se han vinculado al desarrollo de frutos desde que se determinó que el transcrito *PsaST3* de guisante era específico de vainas (Williams y col., 1990). A diferencia de lo que ocurre en guisante, no se han detectado transcritos *ST* en las vainas de garbanzo durante su formación, pero sí se ha localizado la proteína CarST1 en el endocarpo de las vainas verdes (Albornos y col., 2014). Adicionalmente, diferentes estudios encaminados a esclarecer el desarrollo de este órgano en *Vitis vinifera* (Waters y col., 2005; Fernández y col., 2007) y *Citrullus lanatus* (Levi y col., 2006; Wechter y col., 2008) apoyaron la posible función de las ST en este proceso fisiológico. Los trabajos realizados en el desarrollo de frutos carnosos, como uvas (Waters y col., 2005; Fernández y col., 2007) y sandías (Levi y col., 2006; Wechter y col., 2008) indican que las proteínas ST participan en la morfogénesis temprana del fruto, aumentando hasta la maduración del mismo (Fernández y col., 2007) y disminuyendo en el fruto maduro (Wechter y col., 2008). Finalmente, el estudio de los mutantes *flb* (*fleshless berry*), que no presentan división ni determinación de las capas de células especializadas en el agrandamiento del ovario después de la fecundación, y de frutos sin semilla (Fernández y col., 2007) confirmaron la implicación de las proteínas ST en el desarrollo del fruto.

Las proteínas ST parecen participar también en el desarrollo de las semillas, al haberse localizado la proteína CarST1 en la testa y el parénquima de reserva de las misma (Albornos y col., 2014), aunque no se detectaron los transcritos que las codifican (Muñoz y col., 1997). Su presencia en este parénquima de reserva y su acumulación en los cotiledones de las semillas secas, que disminuye durante la germinación hasta desaparecer cuando estos

cotiledones se encuentran exhaustos (Albornos y col., 2014), podrían apuntar a que esta proteína fuera una proteína de reserva en la semilla. Las únicas proteínas de reserva de semillas con repeticiones en tándem son las prolaminas, características de las monocotiledóneas, al que podrían pertenecer las proteínas ST, si bien se diferenciarían de ellas, al igual que ocurre con las α -zeínas, y serían el equivalente de estas proteínas en dicotiledóneas. Sin embargo, las diferencias en la acumulación de los transcritos y las proteínas ST de garbanzo (Albornos, 2010; Albornos y col., 2014) no permite confirmar que su función en la semilla esté relacionada con la acumulación de reservas y podría tener otra función en las semillas independientemente de que pudiera usarse como fuente de N_2 durante la germinación (Albornos y col., 2014).

Finalmente, el estudio de las asociaciones biotróficas entre las plantas y los hongos formadores de micorrizas ha llevado a relacionar un transcrito *ST* de la leguminosa modelo *Medicago truncatula* (TC107197) (Wulf y col., 2003) con esta interacción simbiótica. Este transcrito, que hemos identificado como *MtrST6* en este trabajo, es específico de raíces que establecen micorrizas arbusculares (Wulf y col., 2003; Liu y col., 2007). De hecho, se considera que el gen que codifica esta *ST* podría formar parte del núcleo de genes que se inducen cuando la planta interactúa con hongos formadores de micorrizas (AMF) (Liu y col., 2007). La simbiosis con AMF es la más antigua y extendida de las simbiosis planta-microorganismo. El 80 % de las plantas terrestres son capaces de interactuar con hongos del phylum *Glomeromycota* (Schüßler y col., 2001). Esta interacción procura a la planta nutrientes minerales, especialmente fósforo, y mejora la absorción de agua. En contrapartida, la planta proporciona glucosa al hongo, que es un biotrofo obligado (Parniske, 2008). Se ha observado que el transcrito *ST*, que aumenta marcadamente en las etapas iniciales de la interacción con AMF, mantiene niveles elevados en etapas posteriores (Kuhn y col., 2009) y la expresión del gen *ST* también está regulada específicamente en las etapas tardías de la micorrización (Kuhn y col., 2009), acumulándose preferentemente en las células que albergan arbusculos (Gaude y col., 2012). Todos estos estudios indican que esta proteína *ST* de *M. truncatula* puede participar en el establecimiento de interacciones con AMF, estando la acumulación de sus transcritos regulada por la ruta común SYM de reconocimiento inicial, y también en las fases finales de la simbiosis. Estos estudios añaden otra posible función a la familia de proteínas *ST* o, al menos, a alguno de sus miembros.

2 Planteamiento del trabajo

Los estudios realizados hasta el momento en la familia génica *ST* en diferentes especies, sobre todo *C. arietinum* y *P. sativum*, pero también *V. vinifera*, *C. lanatus* y *M. truncatula*, han permitido relacionarla con diferentes funciones durante el desarrollo de la planta: crecimiento, formación de frutos y semillas e interacciones bióticas.

El objetivo general de este trabajo es definir y caracterizar la familia de proteínas *ST* y obtener una visión global de su posible participación en diferentes procesos del desarrollo vegetal. Con este fin, nos propusimos una serie de objetivos concretos que se detallan a continuación.

Objetivo 1: Definir las características de los genes de la familia *ST* y las proteínas que éstos codifican.

El estudio de una familia de proteínas de función desconocida requiere recopilar toda la información disponible para definir dicha familia en su conjunto y establecer las pautas que permitan profundizar en el conocimiento de la misma.

Para desarrollar este objetivo se realizará, en primer lugar, una búsqueda por similitud de secuencia utilizando las *ST* conocidas para encontrar cualquier otra secuencia que pudiera formar parte de esta familia. A continuación, se analizará la distribución de los genes *ST* en la naturaleza; se estudiarán las secuencias obtenidas en busca de patrones comunes a todas ellas para establecer sus características; y si es posible, se identificará el origen de las secuencias encontradas para determinar en qué órganos de la planta o situaciones ambientales se acumulan los transcritos *ST*.

Los resultados obtenidos al culminar este objetivo nos llevaron a centrarnos en la familia *ST* de *M. truncatula*, que con 6 miembros es la más numerosa de las halladas hasta la fecha. Esta planta se ha consolidado en la última década como leguminosa modelo, y tiene como ventaja que su genoma ha sido completamente secuenciado y se han desarrollado métodos eficaces para su transformación genética. Nuestra hipótesis de trabajo es que cada una de las proteínas *ST* de una determinada especie puede participar en diferentes procesos del desarrollo vegetal, bien de forma exclusiva o de forma principal, cubriendo así la mayoría de las funciones propuestas.

Por todas estas razones, los siguientes objetivos se llevaron a cabo con la familia *ST* de *M. truncatula*.

Objetivo 2: Determinar la localización subcelular de las distintas proteínas ST de *M. truncatula*.

El lugar en el que se localiza una proteína a nivel celular determina el rango de funciones en las que puede estar implicada. Así, es muy relevante conocer dónde se encuentran las proteínas ST en la célula, lo que nos permitirá acotar las posibles funciones en las que pueden participar.

Para realizar este objetivo se construirán plantas transgénicas de *A. thaliana* que produzcan cada una de las ST de *M. truncatula* fusionadas a la proteína verde fluorescente (GFP) y se analizarán por microscopía confocal para determinar la localización subcelular en la que se encuentra cada una de ellas.

Objetivo 3: Caracterizar y analizar la actividad de los promotores de los 6 genes ST de *M. truncatula*.

Los promotores constituyen uno de los puntos clave en la regulación de la expresión génica. El análisis de la secuencia de los promotores ST y la determinación de su actividad aportará una visión global de los momentos en los que se induce la transcripción de cada uno de los miembros de la familia ST de *M. truncatula* a lo largo de su desarrollo. Determinar dónde y cuándo se encuentra activo el promotor de un determinado gen ST podría permitir asociar la proteína que codifica con alguna de las funciones propuestas hasta el momento para esta familia.

Para desarrollar este objetivo se utilizarán dos aproximaciones complementarias. Por una parte, se examinarán las secuencias de los promotores utilizando herramientas *in silico* para detectar elementos *cis* reguladores (CRE) de la transcripción. Por otra parte, se construirían plantas transgénicas, tanto de *A. thaliana* como de *M. truncatula*, que produzcan la proteína *reporter* β -glucuronidasa (GUS) bajo el control de los promotores de los genes ST y se analizará la actividad de cada uno de ellos a lo largo del desarrollo de las plantas, localizando la actividad GUS.

Objetivo 4: Analizar la acumulación de los transcritos de la familia ST de *M. truncatula* en diferentes órganos de la planta y en distintas condiciones de crecimiento.

Conocer el perfil de acumulación de los transcritos ST a lo largo de la vida de la planta nos ayudará a relacionar cada miembro de la familia con los procesos en los que podría participar. Además, el análisis comparativo entre los niveles de acumulación de los

transcritos *ST* en diferentes órganos y situaciones, y la actividad del promotor correspondiente en esas mismas circunstancias (objetivo 3) puede indicar si existe regulación transcripcional a otros niveles o algún tipo de regulación post-transcripcional.

Para completar este objetivo se determinará por RT-sqPCR (transcripción reversa acoplada a PCR semicuantitativa) los niveles de cada transcrito en distintos órganos de plantas de *M. truncatula* desde su germinación hasta el final de su desarrollo. Por otra parte, se valorarán los cambios producidos en la acumulación de estos transcritos cuando la planta crece en presencia de distintos fitorreguladores u otras sustancias relacionadas con el crecimiento, las interacciones bióticas y el estrés abiótico.

3 Resultados

Sección 1: Caracterización de los genes y proteínas de la familia *ST*

Para poder caracterizar los genes y las proteínas *ST* se buscaron en las bases de datos las secuencias que presentaran similitud con las *ST* de guisante y garbanzo, dando como resultado 327 secuencias nucleotídicas que codificaban 136 proteínas diferentes.

Todas las secuencias *ST* pertenecían a especies de angiospermas, en concreto a las dicotiledóneas (65 géneros) repartidas mayoritariamente en la familia Fabaceae (24 %) y Asteraceae (19 %), sin encontrarse en familias como las Brassicaceae.

Todos los genes presentaban un intrón en la zona 5' de la ORF y codificaban proteínas altamente hidrofílicas de tamaño variable. En conjunto, el 36 % de las proteínas *ST* están formadas por menos de 150 aminoácidos, el 41 % contienen entre 150 y 250 y un 23 % presentan más de 250 aminoácidos.

En la región N-terminal, las proteínas *ST* presentaban un péptido señal que las dirige a la vía de secreción celular, sin que fuera posible predecir su localización subcelular realizando estudios *in silico*. Las modificaciones post-traduccionales encontradas fueron: N-glicosilación, fosforilación, N-miristoilación y amidación.

El análisis de la estructura de las proteínas *ST* permitió diferenciar 3 regiones dentro de las mismas: un péptido señal; una región no repetida, a continuación del péptido señal, que presenta una serie de características conservadas; y una región con una secuencia de aminoácidos muy conservada que se repite en tándem hasta el extremo C-terminal y que caracteriza a esta familia de proteínas

La región N-terminal no repetida de las proteínas ST se encuentra entre el péptido señal y el inicio de las repeticiones en tándem. Esta región comenzaba por una Arg que en el 63 % de los casos aparecía seguida de Lys, estando también muy conservadas las posiciones 7 y 8 (83 % de las proteínas) que eran Tyr y Trp, respectivamente, siendo éste generalmente (69 %) el único Trp encontrado en la proteína madura. Además, en esta región se encontraba frecuentemente (84 %) la única Cys que aparece en la proteína.

La región de las repeticiones en tándem empezaba generalmente con 1 o 2 repeticiones imperfectas, que podían tener diferente longitud que el resto y/o carecer de algunas de las características de las repeticiones canónicas, seguidas de una serie de repeticiones perfectas para, finalmente, terminar con una repetición truncada que solía tener 13 aminoácidos. Tanto el tamaño de la repetición como el número de las repeticiones fue relativamente variable, si bien el grado de perfección de las repeticiones de una misma proteína era muy alto.

Las repeticiones comenzaban con un hexapéptido altamente conservado que presentaba la siguiente secuencia: (E/D)FEPRP. A continuación, se encontraban 4 aminoácidos que constituyen la denominada región X₄, tetrapéptido cuya secuencia se ha utilizado como base para la clasificación de las proteínas ST. La siguiente posición (la 11), la ocupaba una Tyr totalmente conservada en todas las repeticiones de la familia ST. En el alineamiento de las secuencias de aminoácidos de 2 repeticiones canónicas de cada una de las proteínas ST se observa claramente que las primeras 11 posiciones de la repetición [(D/E)FEPRPX₄Y] eran las más conservadas.

En función de las características estructurales de las proteínas ST, se establecieron 3 tipos distintos y se confirmó que su diferencia más importante se encontraba en la secuencia X₄:

- Tipo I: se caracteriza porque el patrón X₄ presenta la secuencia consenso: SxTx (x = Lys, Val, Ala o Gly). En este tipo el tetrapéptido no determina un sitio de N-glicosilación.
- Tipo II: en este tipo el patrón X₄ presenta Asn en la primera posición e incluye dos subtipos. El subtipo IIa, con un patrón X₄: NxSx (x = Ile, Val o Ala) que determina un posible sitio de N-glicosilación y el subtipo IIb que se caracteriza por el tetrapéptido Nx(F/L)x (x, muy variable).
- Tipo III: Con un patrón X₄: Dx(S/T)x (x = Ile, Val o Ala).

El árbol filogenético realizado permitió demostrar que las secuencias se agrupaban principalmente por familias taxonómicas, aunque dentro de cada familia las proteínas se agrupaban de acuerdo a los dos tipos mayoritarios establecidos en este trabajo (tipo I y tipo II).

La mayoría de las secuencias *ST* encontradas, un 46 % (149 secuencias), se aislaron de genotecas de cDNA construidas a partir de mRNA procedente de raíces o radículas de 31 especies diferentes, mientras que el resto de las secuencias se hallaron en genotecas construidas a partir de mRNA de otros órganos. Del total de las secuencias, el 7,3 % procedía de semillas; el 6,7 % de hojas; el 5,2 % de órganos elongantes (hipocotilos, epicotilos y tallos); el 2,4 % en frutos; y el resto de flores, haustorios, yemas u otros, con porcentajes menores al 2 %. Con respecto a las condiciones de crecimiento de las plantas en las que se encontraron transcritos *ST*, 122 secuencias (37,3 %) se hallaron en plantas que habían crecido en condiciones de estrés o en presencia de algún tratamiento específico, siendo los más destacados los encontrados en genotecas de plantas sometidas a estrés hídrico o de plantas estableciendo interacciones micorrícicas.

Sección 2: Localización subcelular de las proteínas *ST* de *Medicago truncatula*

La producción de proteínas fusionadas a GFP es una metodología que permite la localización subcelular de las mismas. Con este objetivo se clonaron las ORF de los 6 genes *ST* de *M. truncatula* (números de acceso: LN827607, LN827608, LN827609, LN827610, LN827611 y LN827612, respectivamente para las ORF *ST1*, *ST2*, *ST3*, *ST4*, *ST5* y *ST6*), se construyeron los vectores binarios pK7FWG2.ST y se construyeron plantas transgénicas de *A. thaliana* por transformación con *Agrobacterium tumefaciens*. Estas plantas transgénicas p35S::ST-eGFP producen las proteínas *ST* de *M. truncatula* fusionadas a GFP en su extremo C-terminal, de modo que no interfiera con el procesamiento del péptido señal que se encuentra en el extremo N-terminal.

El análisis de la localización subcelular de las proteínas *ST* en las raíces de las plantas 35S::ST-GFP de *Arabidopsis* se realizó por microscopía confocal. Los resultados obtenidos cuando las raíces se montaron con agua (células turgentes) nos indicaron que las proteínas *ST* se encontraban localizadas en la pared celular. Sin embargo, cuando las células estaban turgentes fue difícil distinguir si las proteínas de fusión *ST*-GFP se encontraban en la pared, en la membrana o en el citoplasma.

Al aplicar manitol a las raíces y provocar la plasmólisis celular se observó que las proteínas ST se mantuvieron en la pared celular pero además 3 de ellas, en concreto las proteínas ST2, ST3 y ST6, también se encontraron en el citoplasma retraído, como se dedujo de la presencia de fluorescencia verde en el interior del citoplasma y que no colocalizó con la pared celular teñida con yoduro de propidio. Esta doble localización es muy clara en las plantas transgénicas p35S::ST3-GFP.

Sección 3: Análisis de la actividad de los promotores ST de *Medicago truncatula*

El análisis de la actividad de los promotores ST (*pST*) de *M. truncatula* se abordó de diferentes maneras: en primer lugar se hizo un estudio bioinformático para encontrar elementos *cis* reguladores (CRE) conocidos en la secuencia de los mismos y, posteriormente, se construyeron plantas transgénicas, tanto de *A. thaliana* como de *M. truncatula*, en las que cada *pST* dirige la expresión de un gen *reporter* y se estudió su actividad *in vivo*.

El análisis de promotores *in silico* se ha convertido en una fuente de información muy importante para el estudio de la regulación transcripcional de los genes. El software TSSP predijo la presencia de la caja TATA en todos los *pST*. El análisis de las secuencias de los 6 *pST* utilizando las bases de datos PLACE y PlantCARE permitió identificar 133 motivos diferentes que se clasificaron en 7 categorías, de acuerdo a la información funcional que proporcionaban las bases de datos. Estas 7 categorías son: específicos de órgano o de tejido, de respuesta a luz (luz), de respuesta a hormonas (hormonas), de respuesta a interacciones bióticas (interacciones bióticas), de respuesta a estrés abiótico (estrés abiótico), *enhancers* y otros elementos (otros). El análisis detallado de la distribución de CRE en la familia *pST* reveló que todos los promotores contenían al menos un 40 % de elementos para controlar la actividad específica en órganos y tejidos, siendo la más abundante la relacionada con la actividad en semillas (entre un 20 % y un 30 %), sobre todo en *pST1* (28 %) y *pST5* (30 %). La categoría de respuesta a hormonas constituyó aproximadamente un 10 % de los CRE en todos los promotores excepto en el *pST1*, que sólo contenía un 5 % de este tipo de elementos. Todos los *pST* contenían elementos relacionados con el establecimiento de interacciones bióticas (entre un 11 % y un 20 % en función del promotor) y de respuesta al estrés abiótico (10 % de los CRE encontrados en cada *pST*, alcanzando un 13 % en *pST3*). Los resultados obtenidos indicaron que cada uno de los promotores de los genes ST presentaba una composición característica de CRE que lo diferenciaba del resto, no solo en cuanto a los motivos concretos que presentaron, sino también en cuanto a los porcentajes de CRE de

cada categoría, excepto los promotores *pST4* y *pST6* que resultaron casi idénticos en ambos aspectos.

La producción de proteínas *reporter* como GUS o GFP bajo un promotor específico, permite determinar la actividad del mismo. Con este objetivo se clonaron los promotores de los 6 genes *ST* de *M. truncatula*, se construyeron los vectores binarios pKGWFS7.pST y se introdujeron por electroporación en *A. tumefaciens*. Con estas agrobacterias se transformaron plantas de *A. thaliana* por inmersión floral y plantas de *M. truncatula* por cocultivo seguido de embriogénesis somática, obteniéndose diferentes líneas transgénicas de cada promotor. Estas plantas transgénicas pST::eGFP-GUS (abreviadamente pST::GUS) expresan las proteínas GUS y eGFP bajo los promotores de los genes *ST* (*pST1*, *pST2*, *pST3*, *pST4* y *pST6*). La clonación del promotor *pST5* no se ha podido llevar a cabo hasta hace pocos meses y por ello no se recoge el análisis de las correspondientes plantas transgénicas.

Las plantas transgénicas pST::GUS de *A. thaliana* y de *M. truncatula* se utilizaron para estudiar la actividad de los promotores en los diferentes tejidos y estados de desarrollo en ambas plantas modelo, localizando la actividad β -glucuronidasa, codificada por el gen *GUS*, con el sustrato X-GlcA que produce tinción azul.

Los primeros análisis se realizaron en plántulas de *A. thaliana* de 3 y 10 d de edad y se estudiaron tanto plántulas crecidas en luz (verdes) como en oscuridad (etioladas). Las plántulas pST4::GUS no presentaron tinción azul ni a los 3 d creciendo en luz o en oscuridad, ni a los 10 d creciendo en oscuridad. En el resto de los promotores (*pST1*, *pST2*, *pST3* y *pST6*) se observó actividad GUS en los cotiledones y en el cilindro vascular de los distintos órganos, siendo la actividad en los cotiledones mayor en las plantas más jóvenes y en las que crecían en oscuridad, si bien el patrón y la intensidad de la tinción histoquímica variaron entre los distintos promotores.

El patrón de actividad de cada promotor *pST* se mantuvo invariable durante el desarrollo de los órganos vegetativos de las plantas adultas. En general, en todas las plantas pST::GUS la actividad predominaba en el cilindro vascular de la raíz, el hipocotilo y las hojas, aunque con patrones distintos, siendo más activos en la raíz que en la parte aérea. La actividad de los distintos *pST* fue muy diferente durante el desarrollo de flores y frutos, lo que confirmó la especificidad de estos promotores. La actividad del *pST1* se restringió al androceo, mientras que el *pST2* presentó actividad en el estilo de la flor desde el inicio de su formación, manteniéndose a lo largo del desarrollo del gineceo. El rasgo diferencial del *pST3* fue su actividad inicialmente localizada en los óvulos y posteriormente en las semillas

durante su desarrollo. El *pST4* apenas presentó actividad durante el desarrollo de flores y frutos. Finalmente, *pST6* se mostró activo, de forma casi exclusiva, en el septo del ovario.

El MeJA ejerció diferentes efectos sobre la actividad de los *pST*. Así, incrementó la actividad *pST3*, *pST4* y *pST6* sin modificar la de los promotores *pST1* y *pST2*.

La actividad de los promotores *ST* en *M. truncatula* se estudió en plántulas de 3, 6 y 10 d crecidas en oscuridad (etioladas) y en luz (verdes), y en plantas adultas en las que se analizó la actividad en hojas, flores y frutos.

El promotor del gen *ST1* se activó en todos los órganos de las plántulas etioladas y verdes durante los 10 primeros días de crecimiento. La actividad GUS se asoció, principalmente, con los tejidos vasculares siendo más intensa en la raíz y en el hipocotilo que en los cotiledones y en las hojas, al menos en plántulas verdes. En las plantas adultas, el *pST1* se activó en la fase reproductora únicamente en los granos de polen y en la vaina cuando alcanzó la morfología característica de barril, momento en que tanto el funículo como el hilio de la semilla presentaron coloración azul. En cortes transversales de las raíces etioladas de plántulas de 6 d, la actividad *pST1* se observó en todo el cilindro central, siendo más intensa en el periciclo, lo que también ocurrió en las raíces de plántulas verdes. En estas últimas, la actividad GUS también fue detectada en el córtex. Finalmente, en cortes transversales se observó que el *pST1* se encuentra activo además en los haces vasculares de los frutos.

En las plántulas de todas las líneas transgénicas *pST2::GUS* el promotor fue muy activo en los haces vasculares de todos los órganos, decreciendo la actividad con la edad de la planta como se confirmó en las raíces secundarias, cuya base apareció claramente azul, y en la región más apical de la raíz principal. Este promotor fue activo en el tallo y las hojas de plantas adultas así como en flores y frutos. Durante el desarrollo de las flores, el *pST2* se activó en los estambres y en la base de los sépalos, mientras que en las primeras etapas de la formación del fruto se encontró actividad *pST2* en el pedúnculo, así como en las semillas que se estaban formando. En etapas más avanzadas del desarrollo del fruto, cuando la vaina presentaba espinas y comenzaba a endurecerse, la tinción azul se mantuvo en la zona de unión del fruto al pedúnculo y se encontró además en los haces vasculares externos. Los cortes histológicos realizados en la raíz de plántulas de 6 d revelaron una fuerte actividad del *pST2* en el periciclo, independientemente de que las plantas crecieran en oscuridad o en luz. La actividad del *pST2* en las secciones del tallo y de hojas confirmó que la tinción se asociaba principalmente a los tejidos conductores, aunque en las hojas también se localizó en la epidermis superior y en el parénquima empalizada.

En plántulas y órganos vegetativos de plantas adultas el patrón de actividad GUS controlado por *pST3* presenta ciertas similitudes con el encontrado en las plantas transgénicas *pST2::GUS*. El patrón de actividad GUS en las flores y en los frutos fue, sin embargo, diferente. Durante la formación de las flores, el promotor *pST3* se activó en los estambres de forma transitoria y las semillas aparecieron marcadas desde el inicio de su desarrollo hasta que el fruto alcanzó su morfología definitiva, sin que se detectara marcaje en la etapa final de su formación. Los frutos aparecieron marcados a lo largo de su desarrollo. En los cortes histológicos se confirmó claramente que la actividad *pST3* en la raíz fue mucho menor en plántulas etioladas que en verdes, mientras que los frutos de las plantas transgénicas *pST3::GUS* en distintos estados de desarrollo mostraron una actividad muy específica de este promotor.

Al igual que ocurrió al analizar la actividad de otros *pST*, la tinción GUS en plántulas *pST4::GUS* se asoció a los haces vasculares y fue decreciendo con la edad. El promotor *pST4* fue el único que se activó en los botones florales en etapas muy tempranas, si bien su actividad fue decreciendo rápidamente según avanzaba su desarrollo. Una vez producida la fecundación, el fruto mostró inicialmente actividad del promotor en la zona apical y posteriormente a lo largo de todo el órgano. Este marcaje se apreció claramente en el funículo pero no se extendió a las semillas. El promotor se activó fuertemente en los tejidos vasculares y en el endocarpo, como se observó en la sección transversal de la vaina. Finalmente, el *pST4* tenía una fuerte actividad en el periciclo de raíces de plántulas de 6 d etioladas y verdes.

El *pST6* fue claramente activo en los cotiledones de plántulas de 3 y 6 d, decreciendo su actividad en plántulas de 10 d. En las plántulas, la tinción en la raíz fue también muy alta y en el caso de las plántulas verdes, el color azul no solo abarcó los haces vasculares sino que se extendió alrededor de los mismos. La actividad del *pST6* fue muy alta en raíces laterales desde el inicio de su formación, como se observó tanto en las raíces de plántulas etioladas como verdes, y a lo largo del crecimiento inicial. En las flores en desarrollo el *pST6* se activó en la base de las anteras y en los óvulos y se encontró posteriormente en las semillas. Los cortes histológicos de las raíces de plántulas *pST6::GUS* de 6 d demostraron la alta actividad de este promotor en este órgano y mientras en raíces de plántulas etioladas la actividad se restringió al cilindro central, en las de plántulas crecidas en luz la tinción se extendió a todos los tejidos de la misma. La actividad GUS en los cortes histológicos de frutos jóvenes determinó que la coloración azul se encontraba principalmente en la región de la nucela de las semillas.

Sección 4: Estudio de la acumulación de los transcritos de los genes *ST* de *Medicago truncatula*

En esta sección se recoge el estudio de la acumulación de los transcritos de los genes de la familia *ST* de *M. truncatula* en distintas partes de la planta a lo largo de su desarrollo, así como en respuesta a distintos fitorreguladores u otras condiciones de interés.

Debido a la elevada similitud de secuencia que presentan los genes *ST* entre sí fue necesario verificar que los cebadores diseñados reconocían exclusivamente a cada miembro de la familia. Esto se realizó por PCR utilizando los pares de oligonucleótidos diseñados utilizando como molde las ORF clonadas, gDNA y mRNA.

Se estudió en primer lugar la acumulación de los transcritos de los 6 genes *ST* de *M. truncatula* en los diferentes órganos a lo largo del desarrollo de plántulas crecidas en oscuridad (etioladas) y en condiciones normales de luz.

Cada uno de los transcritos *ST* se acumulaba de forma diferente durante el desarrollo de las plántulas etioladas, tanto en el perfil como en la cantidad, si bien se encontraron algunos patrones comunes. En estas plántulas etioladas se acumularon casi exclusivamente los transcritos *ST1*, *ST2* y *ST3*, sobre todo este último, tanto en la parte aérea como en la raíz. La cantidad de transcrito fue siempre menor en la parte aérea que en la raíz y decreció con la edad de la plántula en ambos órganos.

Al igual que en plántulas etioladas, en las verdes se encontraron mayoritariamente los transcritos de los genes *ST1*, *ST2* y *ST3*, siendo este último el más abundante en todos los órganos. En plántulas verdes también se detectaron los transcritos *ST6* que no se habían detectado en las etioladas. Cada uno de los transcritos presentó un perfil diferente, aunque al igual que en plántulas etioladas, los transcritos *ST1*, *ST2* y *ST3* presentaron un máximo de acumulación en raíces de plántulas de 3 d, mientras que el *ST6* presentó el máximo en la parte aérea de estas mismas plántulas.

Una vez analizados los transcritos de la familia *ST* de *M. truncatula* en plántulas de hasta 10 d, se estudió como se acumulaban estos transcritos en raíces y hojas de plantas adultas. Se compararon 3 estados diferentes de hojas atendiendo a su grado de desarrollo. Todos los transcritos, salvo *ST3*, fueron mucho más abundantes en la raíz que en las hojas y por primera vez en este estudio se detectaron transcritos del gen *ST4*, si bien únicamente en muy pequeña cantidad en raíces.

Para estudiar los niveles de los transcritos *ST* durante la etapa reproductora, se establecieron 3 estados de desarrollo de las flores y 3 de desarrollo de los frutos (1 a 3 de menor a mayor). Además se analizó la presencia de transcritos en la semilla verde. A lo largo de las distintas etapas estudiadas se detectaron todos los transcritos de la familia *ST*, excepto *ST5*, que no se encontró en ningún caso, y *ST4*, que solo se halló en los estados 1 y 2 de flores. Los transcritos *ST1*, *ST2* y *ST3* se acumularon sobre todo en el estado 2 de flores y en el 3^{er} estado de desarrollo de frutos. Finalmente, los transcritos *ST6* se detectaron en todos los materiales estudiados, aunque su perfil de acumulación difiere del encontrado en *ST1*, *ST2* y *ST3*.

Después de estudiar los niveles de cada transcrito *ST* en los distintos órganos a lo largo del desarrollo de la planta, se analizó su acumulación en plántulas de 7 d expuestas a distintos tratamientos, que se agruparon en diferentes categorías: crecimiento y desarrollo, estrés abiótico e interacciones bióticas.

Los tratamientos con auxinas, CK, GA, BL y estrigolactonas (SL) modificaron los niveles de los transcritos de la familia *ST*, excepto los transcritos *ST4* y *ST5* que no se detectaron en ningún caso. Se pueden establecer 3 patrones diferentes en la acumulación de transcritos: el *ST1* y el *ST6* se comportaron de forma diferente, mientras que el *ST2* y el *ST3* compartían el mismo perfil. En general, las CK aumentaron los transcritos *ST1*, *ST2* y *ST3*, que se diferenciaban en que las auxinas solo estimularon el *ST1*, que también estimuló el *ST6*.

Se estudió el impacto que producían sobre la acumulación de los transcritos *ST* el ABA, el estrés salino (NaCl) y osmótico (manitol), la carencia de nutrientes (N_2 y Pi) y la exposición a temperaturas desfavorables para el crecimiento (4°C y 37°C). Ninguno de estos tratamientos alteró la acumulación de los transcritos *ST4* y *ST5*, que no fueron detectados. Cuando las plantas se mantuvieron en un medio con ABA, NaCl o manitol los niveles de los transcritos *ST1* disminuyeron de forma drástica, mientras que los de los transcritos *ST2*, *ST3* y *ST6* aumentaron notablemente. Asimismo las carencias nutricionales afectaron más claramente a los transcritos *ST1*. Solo las altas temperaturas tuvieron un claro efecto inhibitorio de la acumulación de transcritos *ST*.

Para completar este estudio, se analizó el efecto que causaron sobre la acumulación de los transcritos *ST* las principales hormonas que regulan las respuestas de la planta a la interacción con otros organismos (metil jasmonato, MeJA; etileno, ET; y ácido salicílico, SA), algunas combinaciones de ellas y la herida mecánica. Estos compuestos produjeron cambios importantes en los niveles de los transcritos de todos los miembros de la familia *ST*,

incluidos en algunos casos los mRNA *ST4* y *ST5*. De forma general y con la excepción de los transcritos *ST6*, que se acumulan de forma diferente, la aplicación de ET y de SA disminuye los transcritos de los genes *ST* y el metil jasmonato los induce.

4 Discusión

Las proteínas *ST*, objeto de estudio de este trabajo, se describieron por primera vez en guisante (De Vries y col., 1983; 1985; Williams y col., 1990) y posteriormente se caracterizaron en garbanzo (Muñoz y col., 1997). La función en el desarrollo vegetal de estas proteínas, que presentan una estructura característica de secuencias de 25 o 26 aminoácidos que se repiten consecutivamente en tándem (De Vries y col., 1985; Williams y col., 1990; Muñoz y col., 1997), no se conoce.

Con el objetivo de definir la familia *ST* en su conjunto se han buscado todas las especies que codifican proteínas *ST* (94 especies) y se ha visto que todas ellas pertenecen al Reino Vegetal, dentro del cual no se distribuyen de forma general sino que se restringen a algunas familias de dicotiledóneas, siendo más abundantes en las subclases Asteridae y Rosidae. La distribución restringida de las proteínas *ST* dentro del Reino Vegetal apunta a que su función estaría relacionada con las características diferenciales de estas familias de plantas.

En aquellos casos en los que se ha podido estudiar la organización de los genes *ST*, hemos observado que la mayoría de ellos presentan un intrón en una región cercana al codón de inicio, como ya se había determinado en *P. sativum* (Williams y col., 1990), indicando que los genes *ST* comparten una misma estructura. Estos genes forman familias multigénicas de número variable en función de la especie (entre 2 y 6). En *Medicago truncatula* hemos encontrado la familia multigénica más numerosa, con 6 genes *ST* lo que, junto al hecho de que sea una planta leguminosa modelo, nos ha llevado a centrar nuestro trabajo en la caracterización de la familia *ST* en esta especie. Los genes *ST* de *M. truncatula* tienen una distribución característica en el genoma que sugiere que algunos de los genes *MtrST* se pueden haber generado por duplicación, que es un mecanismo frecuente de evolución y pueden haber sufrido fenómenos de subfuncionalización o neofuncionalización (revisado en Koonin, 2005).

Sin duda, la característica más relevante de las proteínas *ST* es la presencia de repeticiones en tándem con un elevado grado de conservación en su secuencia. De acuerdo a la longitud de sus repeticiones se engloban en la Clase III (repeticiones de 5 a 40 residuos) de la clasificación establecida por Kajava (2012), que se ha adoptado en este trabajo. El

análisis de las 136 secuencias ST diferentes encontradas en las búsquedas, 72 de las cuales codificaban proteínas completas, nos ha permitido definir la estructura común de estas proteínas. Constan de 3 zonas bien diferenciadas: un péptido señal, que comentaremos más adelante, una región no repetida y una región de secuencias repetidas en tándem.

Se ha observado que la región no repetida de las proteínas ST maduras presenta algunos aminoácidos muy conservados, especialmente una Tyr y un Trp consecutivos y una Cys, que puede ser objeto de numerosas modificaciones post-traduccionales que son importantes en la regulación de la actividad de las proteínas (Beltrao y col., 2013).

La secuencia consenso de la repetición comienza por un hexapéptido muy conservado (D/EFÉPRP), seguido de 4 aminoácidos (región X₄) que determinan la clasificación de las proteínas ST, y una Tyr en posición 11 que forman el núcleo de la repetición ST. Frecuentemente las proteínas con repeticiones en tándem se han asociado con el establecimiento de uniones proteína-proteína o con otros ligandos. La variabilidad de la región X₄, ha permitido establecer 3 tipos diferenciados dentro de la familia y esta zona podría dar lugar a un bolsillo hidrofóbico de reconocimiento de ligandos. La abundancia de Tyr en las proteínas ST nos llevó a considerar que quizá fueran objeto de fosforilación por Tyr kinasas pero la mayoría de las Tyr de las repeticiones no estarían sujetas a fosforilación y únicamente MtrST1 y MtrST6 tienen un posible sitio de modificación por Tyr kinasas. Al contrario que en el conjunto de la familia, dentro de una misma proteína ST las repeticiones están muy conservadas en toda su longitud, tal y como se ha comprobado mediante el algoritmo T-REKS (Jordá y Kajava, 2009), y se ha relacionado el grado de perfección de las repeticiones con la tendencia de las proteínas a ser intrínsecamente desestructuradas (IUP) (Jordá y col., 2010). La posibilidad de que las ST sean IUP no se puede descartar por el momento, ya que presentan características propias de estas proteínas (Mathura y col., 2001). La familia ST se ha asociado al dominio de función desconocida 2775 (DUF2775) (Marchler-Bauer y col., 2011) que comprende una parte de la región no repetida de las proteínas y 3 repeticiones en tándem pero que a la vista de nuestros resultados quizá debiera reconsiderarse su definición.

La presencia de un péptido señal en la secuencia de las proteínas ST indica que estas proteínas se dirigen a la vía de secreción celular y por lo tanto, podrían encontrarse en distintos compartimentos en el interior de la célula vegetal, así como en la pared celular de la misma. Los algoritmos bioinformáticos utilizados en este trabajo no proporcionan resultados concluyentes. Sin embargo, los resultados obtenidos *in vivo* utilizando plantas

transgénicas p35S::ST-eGFP de *Arabidopsis* en las que se expresan las proteínas ST de *M. truncatula* sí nos indican el lugar en el que se encuentran. En estas plantas las proteínas ST se acumulan en la pared celular, y 3 de ellas, MtrST2, MtrST3 y MtrST6, presentan una doble localización, ya que también se han detectado en el interior de la célula, al igual que ocurría con las proteínas CarST1 y CarST2 de garbanzo (Albornos 2010; Albornos y col., 2014). Se ha descartado que las proteínas ST de garbanzo sean proteínas estructurales de la misma (Iglesias, 2007) y se ha establecido que CarST1 no está unida a la matriz polisacáridica por enlaces covalentes o iónicos, indicando que estas proteínas permanecen solubles en el apoplasto (Albornos y col., 2014), donde quizá pudieran tener un papel en la defensa (Soares y col., 2007).

En el estudio de la función de las proteínas de la familia ST, nuestra hipótesis de partida fue que las proteínas ST de *M. truncatula* podrían tener funciones específicas y, por lo tanto, los genes ST podrían estar regulados de forma diferencial durante el desarrollo de la planta y/o responder a distintos factores ambientales. En base a esta hipótesis, hemos estudiado cómo está regulada la expresión de cada uno de los genes de la familia ST de *M. truncatula*, analizando los promotores, *in silico* e *in vivo*, y estableciendo la acumulación de sus transcritos. Considerando de forma general el conjunto de resultados obtenidos en este trabajo las proteínas MtrST de *M. truncatula* (que denominaremos simplemente ST en adelante) se pueden agrupar en 3 grupos funcionales que curiosamente coinciden con su localización en el genoma: el formado por la proteína ST1, cuyo gen se halla en el cromosoma 4; el que constituyen las proteínas ST2 y ST3, codificadas por 2 genes consecutivos en el cromosoma 3; y el que engloba a las proteínas ST4, ST5 y ST6, cuyos genes también se encuentran próximos en el cromosoma 3. Esta especialización dentro de los miembros de una misma familia de proteínas, se corresponde con una posible evolución por duplicación, donde cada nueva copia del gen habría adquirido diferencias en la regulación de su expresión y/o en la secuencia codificante.

Independientemente de las divergencias que nos han llevado a establecer los 3 grupos funcionales dentro de la familia ST de *M. truncatula*, hemos encontrado aspectos comunes a todos ellos, que detallaremos en primer lugar.

El primer aspecto común está relacionado con la estructura de los *pST*. El análisis bioinformático de estos promotores indica que tienen una caja TATA para determinar el inicio de la transcripción. Aunque se ha establecido que es más habitual encontrar promotores sin caja TATA en los genes de plantas (Morton y col., 2014), se ha visto que su

presencia es más frecuente en genes duplicados (Zhou y col., 2011), como podrían ser los *ST* de *M. truncatula*. Además, todos los *pST* contienen elementos *cis* de regulación (CRE) por factores endógenos (específicos de órganos o tejidos, hormonas y *enhancer*) y exógenos (luz, estrés abiótico e interacciones bióticas), aunque se diferencian en la combinación exacta de los diferentes CRE y en su número, excepto *pST4* y *pST6* que son prácticamente iguales. El número de CRE encontrado en cada promotor *ST* es muy alto, lo que es habitual en el análisis *in silico* de promotores aunque únicamente un porcentaje muy bajo de los CRE encontrados *in silico*, en torno al 5 %, constituirá un sitio real de unión de un factor de transcripción (TF) y será funcional (Hardison y Taylor, 2012).

Dentro de las características comunes en la actividad de todos los *pST* durante el desarrollo vegetativo destacamos: su actividad en luz y oscuridad; su mayor actividad en etapas tempranas del desarrollo; su actividad en órganos vegetativos, principalmente en raíces; y su actividad en el cilindro vascular y los nervios de las hojas. Por otra parte, el análisis de la acumulación de transcritos confirma en numerosas ocasiones los resultados del estudio de la actividad de promotores y permite matizar algunos resultados.

Los *pST* de *M. truncatula* se activan en plántulas creciendo en oscuridad y en luz, lo que está de acuerdo con lo establecido en guisante y garbanzo en los que los transcritos *ST* se acumulan tanto en plántulas etioladas como verdes (De Vries y col., 1983; Muñoz y col., 1997). Sin embargo, mientras generalmente las actividades de los promotores son más altas en plántulas etioladas, sobre todo en los cotiledones y especialmente en *arabidopsis*, el análisis de los niveles de transcritos indica lo contrario. No es fácil explicar estos resultados en base a la composición de CRE de los *pST* ya que ningún elemento por sí mismo es capaz de dirigir la regulación de un promotor en respuesta a la luz, sino que es la combinación de varios CRE lo que determina finalmente el efecto de ésta (Terzaghi y Cashmore, 1995; Argüello-Astorga y Herrera-Estrella, 1996).

La comparación de plántulas etioladas y verdes, además de para estudiar la regulación por luz, es un buen sistema para establecer la relación de una proteína con el crecimiento en elongación, una de las posibles funciones postuladas para las *ST* y la más estudiada hasta el momento, sobre todo en garbanzo (Muñoz y col., 1997; Hernández-Nistal y col., 2006; Hernández-Nistal y col., 2010; Albornos y col., 2014). Sin embargo, nuestros resultados nos llevan a considerar que, en general, las proteínas *ST* no están relacionadas con la elongación. Aunque determinados datos nos hacían pensar que la proteína *ST1* pudiera estar involucrada en el proceso de elongación: la baja acumulación de los mismos en plántulas

etioladas con un crecimiento muy rápido, la localización de la actividad *pST1* asociada principalmente a los haces vasculares, el efecto contradictorio de GA y BL y la fuerte inducción de la acumulación del transcrito causada por MeJA, un regulador negativo del crecimiento (Yamane y col., 1980; Ueda y Kato, 1982), nos hace descartar esta función también para la proteína ST1.

Cuando consideramos el estado de desarrollo de la planta, la actividad de todos los *pST* y la acumulación de los transcritos en los órganos vegetativos es siempre mayor en los estados de menor desarrollo. Esta mayor acumulación de los transcritos *ST* en órganos jóvenes concuerda con la presencia de CRE asociados a la expresión específica en zonas meristemáticas. El elemento más abundante en relación con la expresión en meristemas es NTBBF1ARROLB. De hecho, en secciones transversales de raíces, se vio que todos los *pST* tenían en común una fuerte actividad en las células del periciclo y explica la actividad de los *pST* en las raíces laterales desde que estas comienzan a emerger. Por tanto, podemos indicar que las *ST* tienen una función en las primeras etapas del desarrollo de la raíz.

Aunque los transcritos *ST* pueden acumularse tanto en las raíces como en las partes aéreas de la planta, independientemente de su estado de madurez, los niveles son mayores en las raíces con pocas excepciones, lo que coincide con la actividad de los *pST* y confirma el perfil EST. Asociado a esta actividad mayoritaria en raíces encontramos un elevado número de CRE relacionados con el desarrollo de las mismas.

Otra característica común de la actividad de los *pST* en los órganos vegetativos es que se encuentra fuertemente asociada al cilindro vascular, tanto en la parte aérea como en la raíz. Los cortes histológicos de raíces revelaron que había diferencias a nivel tisular en la actividad de cada promotor en la raíz. Estas diferencias en la expresión de los *pST* a nivel tisular reflejan diferencias en la regulación de estos genes que podrían implicar además diferencias funcionales. La asociación entre la actividad de los *pST* y el cilindro vascular, también relaciona las proteínas *ST* con células próximas a los haces vasculares como son las células de transferencia, que facilitan la llegada de nutrientes. Las células de transferencia, se desarrollan, por ejemplo, en los ápices de plántulas jóvenes (Pate y col., 1970), en las venas menores de las hojas (Talbot y col., 2001) o en zonas de ramificación (Royo y col., 2007) mientras las conexiones vasculares se establecen adecuadamente. Además estas células son características de las zonas de intercambio maternofilial durante la formación de semillas. Precisamente, en muchas de estas regiones se ha encontrado una fuerte actividad de los *pST*, lo que indica su relación con células de transferencia que facilitan la llegada de

nutrientes a los nuevos órganos en desarrollo y apoyaría una función de las ST en el almacenamiento y/o movilización de reservas (Albornos y col., 2014).

Además de las características comunes que hemos encontrado en la actividad de los *pST*, los resultados obtenidos en este trabajo, basados sobre todo en el nivel de los transcritos de cada uno de los genes *ST* y el efecto de los diferentes tratamientos sobre su acumulación, han sido determinantes para establecer los 3 grupos funcionales en las ST de *M. truncatula* antes indicados. A pesar de que en nuestro estudio la regulación de la expresión del gen *ST1* tiene características comunes con *ST2* y *ST3*, muestra algunas diferencias fundamentales con estos, en especial con respecto a la acumulación de transcritos en los diferentes tratamientos analizados. Es por ello que consideramos que debe constituir un grupo aparte. El hecho de que solo los transcritos de los genes *ST1*, *ST2* y *ST3* sean abundantes en casi todos los momentos del desarrollo analizados, sobre todo *ST2* y *ST3*, puede reflejar que algunas proteínas ST realizan su función durante el desarrollo y otras actúan en procesos que tienen lugar en respuesta a condiciones muy determinadas.

La proteína ST1 es muy similar a CarST1 a nivel de secuencia, y dado que el tipo I dentro de la familia ST es un grupo bien diferenciado y las especies que presentan ST de tipo I contienen solo una, es posible que todas las proteínas ST1 tengan una misma función en el desarrollo vegetal, al contrario que ocurre con las ST de los otros tipos. Existen evidencias experimentales a favor y en contra de que la proteína ST1 de *M. truncatula* puede ser un tipo de proteína de almacenamiento vegetativo (VSP), si bien consideramos más determinantes aquellas que se postulan en contra. Por tanto, el conjunto de resultados obtenidos nos han llevado a concluir que la proteína ST1 no es una VSP, pero sí que actúa en el metabolismo del N₂. La posible regulación del *pST1* por el estatus de N₂ en la planta, se apoya en el hecho de que este promotor es el único que contiene un CRE característico de genes relacionados con el metabolismo de N₂ (AMMORESIIUDUNIA). Dado que también es el único promotor que presenta el CRE denominado P1BS, relacionado con la activación transcripcional de transportadores de Pi, es posible que la proteína ST1 tenga una función más amplia en el estatus nutricional de la planta y no solo en el del N₂. Estos resultados, podrían relacionar la proteína ST1 con el establecimiento de la simbiosis con bacterias fijadoras de N₂ y con hongos formadores de micorrizas, lo que vendría apoyado por el elevado número de transcritos encontrados en genotecas realizadas a partir de plantas en las que se están estableciendo este tipo de simbiosis. Existen más datos, algunos circunstanciales, que permiten pensar que la proteína ST1 puede tener una función en las interacciones de la planta con otros organismos, sean simbióticas o patógenas. Estos datos

incluyen la localización de ST1 en el apoplasto; la inducción por MeJA, que tiene un papel reconocido en la defensa de la planta; la inducción por IAA y CK, que regulan positivamente y de forma coordinada la organogénesis de los nódulos (Oldroyd y col., 2011); y el hecho de que las hormonas que regulan negativamente la nodulación, como ABA (Ding y col., 2008), ET (Penmetsa y Cook, 1997; Oldroyd y col., 2001) y SA (Martinez-Abarca y col., 1998), disminuyen drásticamente los transcritos *ST1*.

El siguiente grupo de nuestra clasificación funcional lo integran ST2 y ST3. La actividad de los promotores *pST2* y *pST3* así como la alta acumulación de los correspondientes transcritos indican una representación significativa de las proteínas en la planta, independientemente del estado de desarrollo de la misma. En este caso la actividad de los promotores *pST2* y *pST3* se refleja en los niveles de transcritos encontrados en los distintos órganos lo que indica que, en este caso, la regulación de la expresión de los genes depende en gran medida de la actividad del promotor.

Aparte del hecho de que los transcritos *ST2* y *ST3* son los más representados de la familia de *M. truncatula* en prácticamente todas las condiciones analizadas, una de las características más destacadas de estos 2 genes es la fuerte acumulación de sus transcritos en las semillas verdes, lo que está en consonancia con la elevada cantidad de CRE relacionados con la expresión génica en semillas, muchos de ellos específicos de promotores de genes de proteínas de reserva de semillas (SSP). Se ha establecido que las ST podrían comportarse como SSP. Sin embargo, se ha comprobado que *pST2* y *pST3* no presentan ninguna de las combinaciones de CRE características de los promotores de genes SSP (revisado en Vicente-Carbajosa y Carbonero, 2005). El hecho de que los transcritos *ST2* y *ST3*, además de inducirse fuertemente en la etapa final de formación de la semilla, experimenten un aumento considerable en presencia de ABA, NaCl y manitol nos llevó a considerar que estas proteínas pudieran ser un grupo de proteínas LEA (*late embryogenesis abundant*). Las proteínas ST2 y ST3 comparten las características comunes de las LEA como su alta hidrofiliidad; un alto contenido de Gly y de aminoácidos pequeños, como Ala o Ser; un alto número de aminoácidos cargados, y la falta de Trp y Cys (Garay-Arroyo y col., 2000). Otra característica común entre las LEA y las proteínas ST es que, al menos 3 de los 7 grupos de proteínas LEA establecidos, contienen repeticiones en tándem de una secuencia conservada (revisado en Battaglia y col., 2008) y finalmente ambos tipos de proteínas podrían ser IUP, y así se ha confirmado en algunas LEA (Lisse y col., 1996; Thalhammer y col., 2010; Boucher y col., 2010). Según el patrón de activación de los promotores y el perfil de acumulación de los transcritos las proteínas ST2 y ST3 se podrían localizar en los mismos

momentos del desarrollo en los que se ha descrito la presencia de las proteínas LEA (Nylander y col., 2001; Rorat y col., 2004; Hundermart y Hicha, 2008).

Al igual que las ST, las proteínas LEA se han detectado en diferentes etapas del desarrollo de las plantas en respuesta a condiciones de déficit hídrico ya sean fisiológicas, como la maduración de la semilla, o bien ocasionadas por estreses ambientales derivados de la sequía, la salinidad o las bajas temperaturas (revisado en Battaglia y col., 2008). De hecho, por su localización subcelular la proteína ST3 podría contribuir a la estabilidad de las membranas en condiciones de choque osmótico. Esta función en la deshidratación se ve apoyada por la presencia de diferentes sitios CRE en *pST2* y *pST3*. Los factores de transcripción que modulan la expresión génica en situaciones de estrés abiótico se unen a sitios ABRE (ABA-responsive element) (DRE2COREZMRAB17, ABRELATERD1 y ABRERATCAL) en las respuestas dependientes de ABA. Los promotores *pST2* y *pST3* contienen sitios ABRE (3 y 4 respectivamente) y al igual que muchos promotores inducibles por ABA contienen también elementos G-box (CACGTGMOTIF), que son elementos auxiliares necesarios para que ocurra la inducción mediada por otros CRE de respuesta a diferentes estímulos, como los ABRE (Donald y Cashmore, 1990; Kim y col., 1992; Shen y Ho, 1995). Además, estos promotores contienen sitios LTRE (*low temperature responsive element*) y los sitios de unión a TF MYB y MYC (Shinozaki y col., 2003; Yamaguchi-Shinozaki y Shinozaki, 2006), aunque solo *pST3* presenta sitios de unión de TF DRE/CTR (dehydration-responsive element/C repeat) (DRECRTCOREAT). En relación con el estrés abiótico, también es interesante la activación de *pST3* por el tratamiento con MeJA y la acumulación de los transcritos *ST2* y *ST3*. Los TF WRKY, presentes en ambos promotores, que tradicionalmente se han relacionado con el estrés biótico, tienen un papel relevante en la regulación del estrés abiótico (Qiu y Yu, 2009). Además de las características de la regulación de los genes *ST2* y *ST3* que hemos expuesto hasta ahora, la relación de las ST con el estrés abiótico se refleja en el perfil EST, donde un alto porcentaje de secuencias se han encontrado en plantas sometidas a algún estrés de este tipo, especialmente sequía y salinidad.

Finalmente, los resultados obtenidos en este trabajo nos llevan a incluir las proteínas ST4, ST5 y ST6 en un mismo grupo funcional. Los niveles de sus transcritos son mucho menores que el resto de la familia *ST*, e incluso indetectables durante el desarrollo vegetativo, lo que implica que pudieran tener funciones muy específicas. De acuerdo con la composición en CRE de los promotores, el perfil EST y el efecto de las hormonas sobre la acumulación de sus transcritos proponemos que estas proteínas desempeñen sus funciones

en la interacción con otros organismos, aunque no descartamos que también otros miembros de la familia puedan participar en estos procesos.

La combinación de CRE encontrada en cada uno de los 6 *pST* de *M. truncatula* fue diferente, excepto en *pST4* y *pST6*, que tenían los mismos elementos en un número muy similar. Este hecho apuntaría a que ambos genes tienen una regulación semejante y a que las proteínas que codifican podrían intervenir en los mismos procesos, llevando a cabo funciones similares y/o complementarias. Tanto la actividad de los promotores como los niveles de acumulación de los transcritos en plántulas indican que la acumulación del *ST6* es mayor que la del *ST4*, cuyos transcritos no se detectaron en ningún estado ni órgano de plántulas. De hecho, los niveles de *ST6* son mucho menores que los de los transcritos *ST1*, *ST2* y *ST3* pero no así la actividad de su promotor que fue muy similar. Esta diferencia puede deberse a la presencia de algún silenciador distal en el genoma que condicione la actividad del promotor *pST6* en su entorno nativo o a la existencia de algún mecanismo de regulación post-transcripcional que afecte a la estabilidad de los mRNA *ST6* en la célula.

Las proteínas *ST4* y *ST6* se han relacionado con el establecimiento de interacciones bióticas por diferentes razones. El estudio *in silico* de los promotores *pST4* y *pST6* indica que contienen una mayor proporción de CRE asociados con las respuestas de defensa que los otros *pST*. Concretamente, *pST4* y *pST6* presentan un elevado número de sitios W-box (WRKY71OS) que es el núcleo de la secuencia de unión de los TF WRKY, considerados un elemento central en la regulación de la defensa (Eulgem y Somssich, 2007) y sitios de unión para los TF TGA (ASF1MOTIFCAMV) (Niggeweg y col., 2000) relacionados con la inducción mediada por SA de numerosos genes PR (*pathogenesis related*) (Strompen y col., 1998, Grüner y col., 2003). En cambio, *pST4* y *pST6* carecen de sitios GCC-like que son los CRE a los que se unen los TF AP2/ERF que median las respuestas de defensa dirigidas por ET (Eulgem y Somssich, 2007). Otro aspecto que apoya la participación de *ST4* y *ST6* en las relaciones bióticas, es el hecho de que las principales hormonas de defensa: MeJA, SA y ET, modifican significativamente los niveles de acumulación de sus transcritos. Sin embargo, pese a la similitud de sus promotores, nuestros resultados indican que la expresión de los genes *ST4* y *ST6* se regula a través de rutas diferentes en la señalización de defensa: JA y SA, respectivamente.

El perfil de acumulación de transcritos *ST4* es típico de genes regulados en la ruta de defensa coordinada por JA y, además, se han encontrado estos transcritos en hojas de plantas atacadas por *Spodoptera exigua*, un insecto herbívoro frente al cual se activa la ruta

del JA (Howe y Jander, 2008), respaldando nuestra idea de que la proteína ST4 tiene un papel en la defensa de la planta frente a herbívoros.

Una característica muy peculiar en la regulación de la acumulación de los transcritos *ST6* es que tanto el SA como el ET promueven su aumento. La acción conjunta de estas hormonas no ha sido apenas descrita en el estudio de la defensa, pero se ha encontrado que frente a algunos patógenos hemibiotrofos la planta activa tanto la vía del SA como la del ET (Sasek y col., 2012), lo que puede explicar que este transcrito se induzca por ambas hormonas. La composición en CRE del promotor *pST6* y el efecto de las hormonas de defensa sobre la acumulación de los transcritos apuntan a un papel de la proteína ST6 en la defensa frente a patógenos biotrofos. Numerosos autores han señalado que la acumulación de los transcritos *ST6* se incrementa notablemente en plantas que establecen micorrizas arbusculares (AM) (Wulf y col., 2003; Hohnjec y col., 2005; Liu y col., 2007), aunque todavía no se conoce en profundidad el papel que juegan las hormonas en el establecimiento de la AM (Gutjahr y Paszkowski, 2009; Gutjahr, 2014). Con estos datos parece que la proteína ST6 tendría una función durante el establecimiento de interacciones bióticas, beneficiosas o no, para el cual la expresión del gen *ST6* está específicamente regulada y en el que posiblemente no participarían otras proteínas ST. Queda por determinar cuál es su papel en el establecimiento de AM.

Finalmente, dentro de este grupo de ST de *M. truncatula*, las situaciones en las que se ha acumulado el transcrito *ST5*, similares a las de la defensina PDF1.22 (Penninckx y col., 1998), relacionan a la proteína por ellos codificada con la respuesta de defensa a patógenos necrotrofos.

Los resultados obtenidos en este trabajo ponen de manifiesto la especificidad de la regulación de la expresión génica en la familia ST de *M. truncatula*. Sin duda, esto debe traducirse en una cierta especialización funcional de las proteínas como proponemos en esta discusión.

5 Conclusiones

1. Las proteínas ST son un nuevo tipo de proteínas exclusivas de algunas familias de dicotiledóneas dentro del Reino Vegetal. Se caracterizan por la presencia del dominio DUF2775 y están codificadas por familias multigénicas de hasta 6 miembros, como en *Medicago truncatula*. Los genes *ST* mantienen una estructura conservada con un intrón próximo al extremo 5'. Las proteínas ST tienen 3 regiones bien diferenciadas:

un péptido señal, una zona N-terminal no repetida y una zona de repeticiones en tándem. El núcleo de las repeticiones consiste en un hexapéptido muy conservado, seguido de un tetrapéptido variable que denominamos X₄ y una Tyr en posición 11 totalmente conservada. Según la secuencia del tetrapéptido X₄ se pueden dividir en tres tipos estructurales. A pesar de que las proteínas ST presentan ciertas similitudes con las proteínas con repeticiones en tándem de la Clase III, tienen características únicas que no permiten incluirlas claramente en ninguno de sus grupos, pudiendo ser proteínas intrínsecamente desestructuradas.

2. Las proteínas ST tienen un péptido señal que las introduce en la vía de secreción, si bien el análisis bioinformático no permitió determinar su localización subcelular. Sin embargo, los resultados obtenidos *in vivo* indican que las ST de *Medicago truncatula* se encuentran localizadas en la pared celular y, tres de ellas, ST2, ST3 y ST6 se sitúan también en el interior de la célula. Esto indica que la doble localización subcelular no es una característica común de todas las proteínas ST.
3. Los promotores de los genes que codifican las 6 proteínas ST de *Medicago truncatula* se activan fundamentalmente en las raíces, sobre todo en los primeros momentos de su desarrollo, y en las zonas más jóvenes de plántulas tanto etioladas como verdes. Su actividad no depende de la luz y se asocia al cilindro vascular. Además, todos ellos se activan en diferentes etapas y/o tejidos durante el desarrollo de flores, frutos y semillas. El estudio de los elementos *cis* reguladores y la actividad de los promotores, así como el análisis de la acumulación de los transcritos, revela diferencias significativas entre los 6 miembros de la familia multigénica de *Medicago truncatula*, lo que nos ha llevado a establecer 3 grupos funcionales, confirmando nuestra hipótesis de partida de que cada ST puede tener funciones específicas.
4. La proteína ST1 podría estar relacionada con el estatus nutricional de la planta, en concreto con el metabolismo y movilización del nitrógeno y el fosfato inorgánico.
5. Las proteínas ST2 y ST3, cuyos transcritos son los más representados, forman un grupo funcional y parecen relacionadas con la protección frente a la desecación, de forma similar a las proteínas abundantes en la embriogénesis tardía o proteínas LEA.
6. Las proteínas ST4, ST5 y ST6 de la familia ST de *Medicago truncatula* parecen actuar principalmente en las interacciones bióticas, si bien cada una de ellas está asociada a una ruta diferente de señalización.

6 Bibliografía

Albornos, L. (2010). Plantas transgénicas de *Arabidopsis thaliana* como sistema modelo para el estudio de las proteínas ST de *Cicer arietinum*. Tesina de Licenciatura. Universidad de Salamanca.

Albornos, L.; Cabrera, J.; Hernández-Nistal, J.; Martín, I.; Labrador, E.; Dopico, B. (2014). Organ accumulation and subcellular location of *Cicer arietinum* ST1 protein. *Plant Sci.* 224: 44-53.

Andrews, D.; Beames, B.; Summers, M.; Park, W. (1988). Characterization of the lipid acyl hydrolase activity of the major potato (*Solanum tuberosum*) tuber protein, patatin, by cloning and abundant expression in a baculovirus vector. *Biochem. J.* 252: 199-206.

Archak, S.; Nagaraju, J. (2006). Eicosapentapeptide repeats (EPRs): Novel repeat proteins specific to flowering plants. *Bioinformatics* 22: 2455-2458.

Arguello-Astorga, G. R.; Herrera-Estrella, L. R. (1996). Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways. *Plant Physiol.* 112: 1151-1166.

Battaglia, M.; Olvera-Carrillo, Y.; Garcarrubio, A.; Campos, F.; Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* 148: 6-24.

Beltrao, P.; Bork, P.; Krogan, N. J.; Noort, V. (2013). Evolution and functional cross-talk of protein post-translational modifications. *Mol. Syst. Biol.* 9: 714.

Boucher, V.; Buitink, J.; Lin, X.; Boudet, J.; Hoekstra, F. A.; Hundertmark, M.; Renard, D.; Leprince, O. (2010). MtPM25 is an atypical hydrophobic late embryogenesis-abundant protein that dissociates cold and desiccation-aggregated proteins. *Plant Cell Environ.* 33: 418-430.

Bowes, B. G. (1996) *A colour atlas of plant structure*. Manson Publishing Ltd.

Cooke, J. E.; Weih, M. (2005). Nitrogen storage and seasonal nitrogen cycling in populus: Bridging molecular physiology and ecophysiology. *New Phytol.* 167: 19-30.

Davies, P. L. (2014). Ice-binding proteins: A remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem. Sci.* 39: 548-555.

De Vries, S. C.; Harmsen, M. C.; Kuiper, M. T.; Dons, H. J.; Wessels, J. G. (1983). Molecular cloning of pea mRNAs encoding a shoot-specific polypeptide and light-induced polypeptides. *Plant Mol. Biol.* 2: 295-303.

De Vries, S. C.; de Vos, W. M.; Harmsen, M. C.; Wessels, J. G. (1985). A shoot-specific mRNA from pea: Nucleotide sequence and regulation as compared to light-induced mRNAs. *Plant Mol. Biol.* 4: 95-102.

Depuydt, S.; Hardtke, C. S. (2011). Hormone signalling crosstalk in plant growth regulation. *Current Biol.* 21: R365-R373.

Ding, Y.; Kalo, P.; Yendrek, C.; Sun, J.; Liang, Y.; Marsh, J. F.; Harris, J. M.; Oldroyd, G. E. (2008). Abscisic acid coordinates nod factor and cytokinin signaling during the regulation of nodulation in *Medicago truncatula*. *Plant Cell* 20: 2681-2695.

Donald, R. G.; Cashmore, A. R. (1990). Mutation of either G box or I box sequences profoundly affects expression from the arabidopsis rbcS-1A promoter. *EMBO J.* 9: 1717-1726.

Eulgem, T.; Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* 10: 366-371.

Fernández, L.; Torregrosa, L.; Terrier, N.; Sreekantan, L.; Grimplet, J.; Davies, C.; Thomas, M. R.; Romieu, C.; Ageorges, A. (2007). Identification of genes associated with flesh morphogenesis during grapevine fruit development. *Plant Mol. Biol.* 63: 307-323.

- Garay-Arroyo, A.; Colmenero-Flores, J. M.; Garcarrubio, A.; Covarrubias, A. A. (2000).** Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* 275: 5668-5674.
- Gaude, N.; Bortfeld, S.; Duensing, N.; Lohse, M.; Krajinski, F. (2012).** Arbuscule-containing and non-colonized cortical cells of mycorrhizal roots undergo extensive and specific reprogramming during arbuscular mycorrhizal development. *Plant J.* 69: 510-528.
- Grüner, R.; Strompen, G.; Pfitzner, A. J.; Pfitzner, U. M. (2003).** Salicylic acid and the hypersensitive response initiate distinct signal transduction pathways in tobacco that converge on the as-1-like element of the PR-1a promoter. *Eur. J. Biochem.* 270: 4876-4886.
- Gutjahr, C. (2014).** Phytohormone signalling in arbuscular mycorrhiza development. *Curr. Op. Plant Biol.* 20: 26-34.
- Gutjahr, C.; Paszkowski, U. (2009).** Weights in the balance: Jasmonic acid and salicylic acid signaling in root-biotroph interactions. *Mol. Plant-Microbe Interact.* 22: 763-772.
- Hardison, R. C.; Taylor, J. (2012).** Genomic approaches towards finding *cis*-regulatory modules in animals. *Nat. Rev. Gen.* 13: 469-483.
- Hernández-Nistal, J.; Labrador, E.; Martín, I.; Jiménez, T.; Dopico, B. (2006).** Transcriptional profiling of cell wall protein genes in chickpea embryonic axes during germination and growth. *Plant Physiol. Biochem.* 44: 684-692.
- Hernández-Nistal, J.; Martín, I.; Esteban, R.; Dopico, B.; Labrador, E. (2010).** Abscisic acid delays chickpea germination by inhibiting water uptake and down-regulating genes encoding cell wall remodelling proteins. *Plant Growth Reg.* 61: 175-183.
- Hohnjec, N.; Vieweg, M. F.; Pühler, A.; Becker, A.; Küster, H. (2005).** Overlaps in the transcriptional profiles of *Medicago truncatula* roots inoculated with two different *Glomus* fungi provide insights into the genetic program activated during arbuscular mycorrhiza. *Plant Physiol.* 137: 1283-1301.
- Howe, G. A.; Jander, G. (2008).** Plant immunity to insect herbivores. *Annu. Rev. Plant Biol.* 59: 41-66.
- Hundertmark, M.; Hinch, D. K. (2008).** LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Gen.* 9: 118.
- Iglesias, R. (2007).** ST-1 y ST-2 ¿Nuevas proteínas de pared celular?. Tesina de Licenciatura. Universidad de Salamanca.
- Jordá, J.; Kajava, A. V. (2009).** T-REKS: Identification of tandem REpeats in sequences with a K-meanS based algorithm. *Bioinformatics* 25: 2632-2638.
- Jordá, J.; Xue, B.; Uversky, V. N.; Kajava, A. V. (2010).** Protein tandem repeats—the more perfect, the less structured. *FEBS journal.* 277: 2673-2682.
- Kajava, A. V. (1998).** Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* 277: 519-527.
- Kajava, A. V. (2012).** Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.* 179: 279-288.
- Katti, M. V.; Sami-Subbu, R.; Ranjekar, P. K.; Gupta, V. S. (2000).** Amino acid repeat patterns in protein sequences: Their diversity and structural-functional implications. *Prot. Sci.* 9: 1203-1209.
- Kim, S. R.; Choi, J. L.; Costa, M. A.; An, G. (1992).** Identification of G-box sequence as an essential element for methyl jasmonate response of potato proteinase inhibitor II promoter. *Plant Physiol.* 99: 627-631.
- Kobe, B.; Deisenhofer, J. (1993).** Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats.

- Koonin, E. V. (2005).** Orthologs, paralogs, and evolutionary genomics 1. *Annu. Rev. Genet.* 39: 309-338.
- Kuhn, H.; Küster, H.; Requena, N. (2010).** Membrane steroid-binding protein 1 induced by a diffusible fungal signal is critical for mycorrhization in *Medicago truncatula*. *New Phytol.* 185: 716-733.
- Levi, A.; Davis, A.; Hernandez, A.; Wechter, P.; Thimmapuram, J.; Trebitsh, T.; Tadmor, Y.; Katzir, N.; Portnoy, V.; King, S. (2006).** Genes expressed during the development and ripening of watermelon fruit. *Plant Cell Rep.* 25: 1233-1245.
- Lisse, T.; Bartels, D.; Kalbitzer, H. R.; Jaenicke, R. (1996).** The recombinant dehydrin-like desiccation stress protein from the resurrection plant *Craterostigma plantagineum* displays no defined three-dimensional structure in its native state. *Biol. Chem.* 377: 555-562.
- Liu, J.; Maldonado-Mendoza, I.; Lopez-Meyer, M.; Cheung, F.; Town, C. D.; Harrison, M. J. (2007).** Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *Plant J.* 50: 529-544.
- Marchler-Bauer, A.; Lu, S.; Anderson, J. B.; Chitsaz, F.; Derbyshire, M. K.; DeWeese-Scott, C.; Fong, J. H.; Geer, L. Y.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Jackson, J. D.; Ke, Z.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Mullokandov, M.; Omelchenko, M. V.; Robertson, C. L.; Song, J. S.; Thanki, N.; Yamashita, R. A.; Zhang, D.; Zhang, N.; Zheng, C.; Bryant, S. H. (2011).** CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39: D225-9.
- Marcotte, E. M.; Pellegrini, M.; Yeates, T. O.; Eisenberg, D. (1999).** A census of protein repeats. *J. Mol. Biol.* 293: 151-160.
- Martínez-Abarca, F.; Herrera-Cervera, J.; Bueno, P.; Sanjuan, J.; Bisseling, T.; Olivares, J. (1998).** Involvement of salicylic acid in the establishment of the *Rhizobium meliloti*-alfalfa symbiosis. *Mol. Plant-Microbe Interact.* 11: 153-155.
- Meuriot, F.; Noquet, C.; Avice, J.; Volenec, J. J.; Cunningham, S. M.; Sors, T. G.; Caillet, S.; Ourry, A. (2004).** Methyl jasmonate alters N partitioning, N reserves accumulation and induces gene expression of a 32-kDa vegetative storage protein that possesses chitinase activity in *Medicago sativa* taproots. *Physiol. Plant.* 120: 113-123.
- Morton, T.; Petricka, J.; Corcoran, D. L.; Li, S.; Winter, C. M.; Carda, A.; Benfey, P. N.; Ohler, U.; Megraw, M. (2014).** Paired-end analysis of transcription start sites in arabidopsis reveals plant-specific promoter signatures. *Plant Cell* 26: 2746-2760.
- Mosavi, L. K.; Minor, D. L., Jr; Peng, Z. Y. (2002).** Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. USA* 99: 16029-16034.
- Muñoz, F. J.; Dopico, B.; Labrador, E. (1997).** Two growth-related organ-specific cDNAs from *Cicer arietinum* epicotyls. *Plant Mol. Biol.* 35: 433-442.
- Niggeweg, R.; Thurow, C.; Kegler, C.; Gatz, C. (2000).** Tobacco transcription factor TGA2.2 is the main component of as-1-binding factor ASF-1 and is involved in salicylic acid- and auxin-inducible expression of as-1-containing target promoters. *J. Biol. Chem.* 275: 19897-19905.
- Nylander, M.; Svensson, J.; Palva, E. T.; Welin, B. V. (2001).** Stress-induced accumulation and tissue-specific localization of dehydrins in *Arabidopsis thaliana*. *Plant Mol. Biol.* 45: 263-279.
- Oldroyd, G. E.; Engstrom, E. M.; Long, S. R. (2001).** Ethylene inhibits the nod factor signal transduction pathway of *Medicago truncatula*. *Plant Cell* 13: 1835-1849.
- Oldroyd, G. E.; Murray, J. D.; Poole, P. S.; Downie, J. A. (2011).** The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* 45: 119-144.
- Parniske, M. (2008).** Arbuscular mycorrhiza: The mother of plant root endosymbioses. *Nature Reviews Microbiology.* 6: 763-775.

- Pate, J.; Gunning, B.; Milliken, F. (1970).** Function of transfer cells in the nodal regions of stems, particularly in relation to the nutrition of young seedlings. *Protoplasma* 71: 313-334.
- Peifer, M.; Berg, S.; Reynolds, A. B. (1994).** A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell* 76: 789-791.
- Penmetsa, R. V.; Cook, D. R. (1997).** A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. *Science* 275: 527-530.
- Penninckx, I.A.; Thomma, B. P.; Buchala, A.; Mettraux, J. P.; Broekaert, W. F. (1998).** Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in arabidopsis. *Plant Cell* 10: 2103-2113.
- Qiu, Y.; Yu, D. (2009).** Over-expression of the stress-induced OsWRKY45 enhances disease resistance and drought tolerance in arabidopsis. *Environ. Exp. Bot.* 65: 35-47.
- Rorat, T.; Grygorowicz, W.; Irzykowski, W.; Rey, P. (2004).** Expression of KS-type dehydrins is primarily regulated by factors related to organ type and leaf developmental stage during vegetative growth. *Planta* 218: 878-885.
- Royo, J.; Gómez, E.; Hueros, G. (2007).** Transfer cells. En "Endosperm" (O.-A. Olsen, ed.) pp 73-89. Springer.
- Šašek, V.; Nováková, M.; Jindřichová, B.; Bóka, K.; Valentová, O.; Burketová, L. (2012).** Recognition of avirulence gene AvrLm1 from hemibiotrophic ascomycete *Leptosphaeria maculans* triggers salicylic acid and ethylene signaling in *Brassica napus*. *Mol. Plant-Microbe Interact.* 25: 1238-1250.
- Schübler, A.; Schwarzott, D.; Walker, C. (2001).** A new fungal phylum, the glomeromycota: Phylogeny and evolution. *Mycol. Res.* 105: 1413-1421.
- Selengut, J. D. (2001).** MDP-1 is a new and distinct member of the haloacid dehalogenase family of aspartate-dependent phosphohydrolases. *Biochem.* 40: 12704-12711.
- Shen, Q.; Ho, T. H. (1995).** Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel cis-acting element. *Plant Cell* 7: 295-307.
- Shinozaki, K.; Yamaguchi-Shinozaki, K.; Seki, M. (2003).** Regulatory network of gene expression in the drought and cold stress responses. *Curr. Opin. Plant Biol.* 6: 410-417.
- Sikorski, R. S.; Boguski, M. S.; Goebel, M.; Hieter, P. (1990).** A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell* 60: 307-317.
- Small, I. D.; Peeters, N. (2000).** The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* 25: 45-47.
- Soares, N. C.; Francisco, R.; Ricardo, C. P.; Jackson, P. A. (2007).** Proteomics of ionically bound and soluble extracellular proteins in *Medicago truncatula* leaves. *Proteomics* 7: 2070-2082.
- Spilatro, S. R.; Anderson, J. M. (1989).** Characterization of a soybean leaf protein that is related to the seed lectin and is increased with pod removal. *Plant Physiol.* 90: 1387-1393.
- Strompen, G.; Grüner, R.; Pfitzner, U. M. (1998).** An as-1-like motif controls the level of expression of the gene for the pathogenesis-related protein 1a from tobacco. *Plant Mol. Biol.* 37: 871-883.
- Sun, X.; Rikkerink, E. H.; Jones, W. T.; Uversky, V. N. (2013).** Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* 25: 38-55.
- Talbot, M. J.; Franceschi, V. R.; McCurdy, D. W., Offler, C. E. (2001).** Wall ingrowth architecture in epidermal transfer cell of *Vicia faba* cotyledons. *Protoplasma* 215: 191-203.

- Terzaghi, W. B.; Cashmore, A. R. (1995).** Light-regulated transcription. *Annu. Rev. Plant Biol.* 46: 445-474.
- Thalhammer, A.; Hundertmark, M.; Popova, A. V.; Seckler, R.; Hinch, D. K. (2010).** Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state. *Biochim. Biophys. Acta-Biomembranes.* 1798: 1812-1820.
- Tian, W. M.; Peng, S. Q.; Wang, X. C.; Shi, M. J.; Chen, Y. Y.; Hu, Z. H. (2007).** Vegetative storage protein in litchi chinensis, a subtropical evergreen fruit tree, possesses trypsin inhibitor activity. *Ann. Bot.* 100: 1199-1208.
- Ueda, J.; Kato, J. (1982).** Inhibition of cytokinin-induced plant growth by jasmonic acid and its methyl ester. *Physiol. Plant.* 54: 249-252.
- Uversky, V. N. (2010).** The mysterious unfoldome: Structureless, underappreciated, yet vital part of any given proteome. *J. Biomed. Biotechnol.* 2010: 568068.
- Vicente-Carbajosa, J.; Carbonero, P. (2005).** Seed maturation: Developing an intrusive phase to accomplish a quiescent state. *Int. J. Dev. Biol.* 49: 645.
- Waters, D. L.; Holton, T. A.; Ablett, E. M.; Lee, L. S.; Henry, R. J. (2005).** cDNA microarray analysis of developing grape (*Vitis vinifera* cv. shiraz) berry skin. *Funct. Int. Gen.* 5: 40-58.
- Wechter, W. P.; Levi, A.; Harris, K. R.; Davis, A. R.; Fei, Z.; Katzir, N.; Giovannoni, J. J.; Salman-Minkov, A.; Hernandez, A.; Thimmapuram, J.; Tadmor, Y.; Portnoy, V.; Trebitsh, T. (2008).** Gene expression in developing watermelon fruit. *BMC Genomics* 9: 275.
- Williams, M. E.; Mundy, J.; Kay, S. A.; Chua, N. (1990).** Differential expression of two related organ-specific genes in pea. *Plant Mol. Biol.* 14: 765-774.
- Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.; Takayama, S.; Brown, C. J.; Dunker, A. K. (2001).** The protein non-folding problem: Amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.* 6: 89-100.
- Wulf, A.; Manthey, K.; Doll, J.; Perlick, A. M.; Linke, B.; Bekel, T.; Meyer, F.; Franken, P.; Küster, H.; Krajinski, F. (2003).** Transcriptional changes in response to arbuscular mycorrhiza development in the model plant *Medicago truncatula*. *Mol. Plant-Microbe Interact.* 16: 306-314.
- Yamaguchi-Shinozaki, K.; Shinozaki, K. (2006).** Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* 57: 781-803.
- Yamane, H.; Sugawara, J.; Suzuki, Y.; Shimamura, E.; Takahashi, N. (1980).** Syntheses of jasmonic acid related compounds and their structure-activity relationships on the growth of rice seedlings. *Agric. Biol. Chem.* 44: 2857-2864.
- Zou, Y.; Huang, W.; Gu, Z.; Gu, X. (2011).** Predominant gain of promoter TATA box after gene duplication associated with stress responses. *Mol. Biol. Evol.* 28: 2893-2904.