

Tesis Doctoral

Fosfatasas de proteínas con dominios Kelch en *Arabidopsis thaliana*. Estudios funcionales, estructurales y evolutivos

Maselli, Gustavo Ariel

2016-05-31

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Maselli, Gustavo Ariel. (2016-05-31). Fosfatasas de proteínas con dominios Kelch en *Arabidopsis thaliana*. Estudios funcionales, estructurales y evolutivos. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Maselli, Gustavo Ariel. "Fosfatasas de proteínas con dominios Kelch en *Arabidopsis thaliana*. Estudios funcionales, estructurales y evolutivos". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2016-05-31.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Fosfatasas de proteínas con dominios Kelch en *Arabidopsis thaliana*.

Estudios funcionales, estructurales y evolutivos.

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área de Ciencias Biológicas

Lic. Gustavo Ariel Maselli

Director de tesis: Dr. Santiago Mora García

Consejero de estudios: Dr. Pablo Cerdán

Lugar de trabajo: Fundación Instituto Leloir – IIBBA-CONICET

Buenos Aires, 2016.

Fosfatasas de proteínas con dominios Kelch en *Arabidopsis thaliana*. Estudios funcionales, estructurales y evolutivos.

Las fosfatasas de proteínas con dominios Kelch (PPKL) son miembros de la familia PPP de fosfatasas de serinas y treoninas que están presentes sólo en algas verdes, plantas y alveolados. Las PPKL son proteínas bimodulares con un dominio N-terminal de tipo turbina β y un dominio catalítico en el extremo C-terminal, separados por una extensa secuencia conectora sin homología con otras conocidas. En las plantas, las PPKLs han sido descritas como efectoras de la señalización por brassinosteroides y aparecen involucradas en el control de la proliferación celular.

En este trabajo se han reevaluado los roles de los cuatro miembros de la familia PPKL presentes en *Arabidopsis thaliana* (BSL1, BSL2, BSL3 y BSU1) mediante análisis filogenéticos, funcionales y genéticos. BSL1 y BSL2/BSL3 pertenecen a dos clados evolutivos antiguos que han sido altamente conservados en las plantas terrestres; en cambio, los genes de tipo BSU1 se encuentran exclusivamente en la familia Brassicaceae y muestran una notable divergencia de secuencia, incluso entre especies estrechamente emparentadas. La pérdida de función simultánea de dos parálogos cercanamente relacionados, BSL2 y BSL3, causa un conjunto de alteraciones fenotípicas peculiares y novedosas; la pérdida de función de BSL1 es, en cambio, fenotípicamente silenciosa. Sin embargo, los tres parálogos más importantes cumplen juntos un papel esencial en etapas tempranas del desarrollo, en el cual BSU1 no puede suplantarlos. Estos patrones evolutivos se ven reflejados en un comportamiento inusual de las PPKL que las diferencia del resto de las fosfatasas de tipo PPP: su capacidad de participar en interacciones homotípicas. BSL1, BSL2 y BSL3 se comportan in vivo como homodímeros, mientras que BSU1 sólo es capaz de realizar interacciones débiles y/o transitorias consigo misma. Curiosamente, se encontró que los determinantes de la interacción son diferentes entre los miembros de los clados principales. En BSL1, el dominio catalítico es el principal impulsor de la interacción; en contraste, el dominio catalítico de BSL2 es necesario, pero no suficiente, y la dimerización requiere de dos porciones de secuencia en el dominio conector. El dominio conector de BSL2 se comporta en solución como una proteína intrínsecamente desordenada propensa a formar elementos de estructura secundaria frente a perturbaciones leves del medio.

Resumen

La presencia de un miembro altamente divergente, en conjunto con la inusual capacidad de formar homodímeros y las diferencias presentes entre parálogos con alto grado de conservación, ponen de manifiesto la complejidad de esta pequeña y esencial familia génica, al tiempo que cuestionan los fundamentos de la asignación de sus roles funcionales.

Palabras clave: *Arabidopsis thaliana*, fosfatasa de proteínas, PPKL, relaciones evolutivas, oligomerización, estructura intrínsecamente desordenada.

Protein phosphatases with Kelch-like domains in *Arabidopsis thaliana*. Functional, structural and evolutionary studies.

Protein phosphatases with Kelch-like domains (PPKL) are members of the PPP family present in green algae, plants and alveolates. PPKL are bimodular proteins with a β -propeller domain at the N-terminus and a catalytic domain akin to PP1 at the C-terminus; both domains are tethered by an intermediate sequence of unknown affiliations. In plants, PPKLs have been described as effectors of brassinosteroid signaling and cell proliferation. We reappraised the roles of the four members of the family present in *Arabidopsis thaliana* -*BSL1*, *BSL2*, *BSL3* and *BSU1*- through phylogenetic, functional and genetic analyses. *BSL1* and *BSL2/BSL3* belong to two ancient evolutionary clades that have been highly conserved in land plants; in contrast, *BSU1*-type genes are exclusively found in the Brassicaceae and display a remarkable sequence divergence, even among closely related species. Simultaneous loss of function of the closely related paralogs *BSL2* and *BSL3* causes a peculiar array of phenotypic alterations; loss of function of *BSL1* is, in turn, phenotypically silent. However, the three major paralogs play together an essential role in early stages of development, that *BSU1* is unable to supplement. These evolutionary patterns mirror an unusual behaviour in PPP phosphatases: the ability of PPKL to form homotypic interactions. *BSL1*, *BSL2* and *BSL3* behave *in vivo* as homodimers, whereas *BSU1* is only able of transient interactions. The interaction determinants are, however, different among members of the main clades. In *BSL1*, the catalytic domain is the main driver of interaction; in contrast, the catalytic domain of *BSL2* is necessary but not sufficient, and dimerization requires two conserved sequence stretches in the linker domain. The linker domain of *BSL2* behaves in solution as an intrinsically disordered protein prone to form secondary structure elements upon mild perturbations.

The presence of a highly divergent member, together with the unusual ability to form homodimers and the existing differences between paralogs with a high degree of conservation, reveal the complexity of this small and essential gene family, while questioning the foundations of their functional role assignment.

Keywords: *Arabidopsis thaliana*, protein phosphatase, PPKL, evolutionary relationship, oligomerization, intrinsic disordered structure.

AGRADECIMIENTOS

Quiero comenzar agradeciendo al Dr. Santiago Mora García por todo el esfuerzo dedicado a dirigir mi tesis y a enseñarme tantas cosas a lo largo de los años, pero no menos importante es agradecerle por las risas y los buenos momentos compartidos.

Agradezco al Dr. Ricardo Wolosiuk el haberme permitido realizar la tesis en su laboratorio y por compartir su vasta experiencia.

A Marcelo Yanovsky y a Pablo Cerdán, les agradezco por los consejos recibidos, por dejarme usar su equipamiento, compartir recursos y aguantarme rondando en sus laboratorios prácticamente a diario.

Agradezco al CONICET por la beca doctoral que me permitió llevar a cabo esta tesis.

Durante la realización de esta tesis tuve la suerte de conocer a excelentes personas que me acompañaron en el día a día: los chicos y chicas que están o pasaron por los laboratorios 103 y 204. En particular no puedo dejar de nombrar y agradecer a Maty, Sabri, Marti, Steve y Majo por las incontables horas de compañía y consejo. Conocerlos fue una de las mejores cosas de mi estadía en el Instituto.

Gracias a Zek, Carla y Ale, compañeros de la facultad y grandísimos amigos, con quiénes compartí muchos buenos momentos y me ayudaron a salir de más de un pozo.

Le agradezco a mi familia por haberme acompañado a lo largo de la carrera.

El último y más importante agradecimiento es para Mari. Gracias por estar siempre a mi lado y haber compartido todo esto conmigo.

ÍNDICE

Abreviaturas.....	9
Capítulo I	11
Caracterización evolutiva y funcional de las fosfatasa con dominios Kelch en <i>Arabidopsis thaliana</i>	11
Introducción	12
La fosforilación de proteínas.....	12
Las Fosfatasa de Serinas/Treoninas.....	14
Las Fosfatasa PPKL.....	18
Objetivos	23
Resultados	24
Análisis filogenético de las fosfatasa PPKL en plantas.....	24
Localización subcelular de las fosfatasa PPKL	28
Interacción de las fosfatasa BSL con participantes en la vía de señalización por BRs.....	30
Papel esencial de BSL1, BSL2, y BSL3	33
Conclusiones	36
Capítulo II	42
Análisis de las interacciones entre las fosfatasa PPKL de <i>Arabidopsis thaliana</i>	42
Introducción	43
Relación entre estructura y regulación en las fosfatasa PPP	43
Dominios en las Fosfatasa PPKL.....	46
Objetivos	48
Resultados	49
Interacciones homotípicas en las fosfatasa BSL de <i>Arabidopsis thaliana</i>	49
BSL1 y BSL2 difieren en sus requerimientos para que la interacción tenga lugar	53
Características del dominio conector.....	58
Modo de acción del dominio conector	62
Análisis delecional del dominio conector.....	64
Modificaciones en la región 600-642 del dominio L de BSL2.....	68
Diferencias entre los dominios catalíticos de BSL1 y BSL2.....	70
Conclusiones	78
Capítulo III	85
Introducción	86

Proteínas intrínsecamente desordenadas	86
Regulación de la actividad de fosfatasa PPP por proteínas intrínsecamente desordenadas.....	87
Objetivos	90
Resultados	91
Predicción de la estructura del dominio conector de BSL2.	91
Purificación y caracterización del dominio conector de BSL2.....	93
Propiedades estructurales del dominio conector de BSL2.	96
Conclusiones	99
Conclusiones Generales	103
Procedimientos Experimentales	105
1) Material vegetal y condiciones de crecimiento	105
2) Metodologías generales.....	105
2.a) Preparación de ADN Plasmídico.....	105
2.b) Preparación de fragmentos de ADN	106
2.b.i) Amplificación de fragmentos de ADN (PCR)	106
2.b.ii) Purificación de fragmentos de ADN de geles de agarosa	107
2.b.iii) Digestión con enzimas de restricción.....	107
2.b.iv) Reacciones de ligación de ADN	107
2.b.v) Clonados por sistema <i>Gateway</i>	107
3) Cepas bacterianas y transformación	108
3.a) <i>Escherichia coli</i>	108
3.a.i) Cepas y condiciones de cultivo.....	108
3.a.ii) Preparación de células de <i>E. coli</i> DH5 α electro competentes	108
3.a.iii) Transformación de <i>E. coli</i> por electroporación	108
3.b) <i>Agrobacterium tumefaciens</i>	109
3.b.i) Cepas y condiciones de cultivo.....	109
3.b.ii) Preparación de células electro competentes de <i>A. tumefaciens</i>	109
3.b.iii) Transformación de <i>A. tumefaciens</i> por electroporación.	109
3.b.iv) Agroinfiltración.....	109
4) Análisis filogenético.....	111
5) Plásmidos y construcciones utilizadas	112
5.a) Diseño y obtención de las construcciones para Y2H.....	112
5.b) Diseño y obtención de las construcciones fusionadas a epítopos	112
5.c) Diseño y obtención de las construcciones para biFC	113

6) Ensayo de doble híbrido en levaduras (Y2H)	113
6.a) Transformación de levaduras.....	113
6.b) Medición de la actividad de los genes reporteros	114
6.b.i) β -Galactosidasa (β -Gal)	114
Actividad β -Gal sobre filtro	114
Actividad β -Gal en medio líquido.....	115
6.b.ii) <i>HIS3</i>	116
7) Extracción y detección de proteínas mediante <i>Western Blot</i>	116
7.a) Extracción de proteínas totales.....	116
<i>Saccharomyces cerevisiae</i>	116
<i>Nicotiana benthamiana</i>	116
7.b) Electroforesis de proteínas	116
7.c) Transferencia a membrana	117
7.d) Inmunodetección y revelado.....	117
Anti-epítotope HA.....	117
Anti-GFP, anti-BSL2, anti-BD y anti-AD.....	117
Revelado.....	118
8) Co-inmunoprecipitación.....	118
9) Microscopía confocal	118
10) Cromatografía de exclusión molecular	119
10.a) Precipitación con TCA.....	119
11) Expresión y purificación del dominio conector de BSL2	119
12) Técnicas biofísicas	120
12.a) Dicroísmo circular (DC).....	120
12.b) Static Light Scattering (SLS).....	120
12.c) Resonancia magnética nuclear (RMN).....	121
Referencias.....	122

Abreviaturas

3AT	3-amino-1,2,4-triazol
AD	Dominio de transactivación del factor de transcripción Gal4
ADN	Ácido desoxirribonucleico
Ala	Alanina
Arg	Arginina
Asn	Asparagina
Asp	Ácido aspártico
AtBSL1 o BSL1	<i>BSU-like 1</i> de <i>A. thaliana</i>
AtBSL2 o BSL2	<i>BSU-like 2</i> de <i>A. thaliana</i>
AtBSL3 o BSL3	<i>BSU-like 3</i> de <i>A. thaliana</i>
AtBSU1 o BSU1	<i>Bri-1 suppressor 1</i> de <i>A. thaliana</i>
BAK1	<i>BRI1 associated kinase 1</i>
BD	Dominio de unión al DNA del factor de transcripción Gal4
BES1	<i>BRI1-EMS Suppressor 1</i>
biFC	Complementación bimolecular de fluorescencia
BIN2	<i>Brassinosteroid Insensitive 2</i>
BR	Brassinosteroides
BRI1	<i>Brassinosteroid insensitive 1</i>
BSK	<i>Brassinosteroid-signaling kinase</i>
BSL	<i>BSU-like</i>
BSLxK	Dominio Kelch de la proteína BSLx
BSLxL	Dominio conector (linker) de la proteína BSLx
BSLxP	Dominio catalítico (phosphatase) de la proteína BSLx
BSU1	<i>Bri1 Suppressor 1</i>
BZR1	<i>Brassinazole resistant 1</i>
CaMV 35S	promotor 35S del virus del mosaico del coliflor
CEM	Cromatografía de exclusión molecular
coIP	Coinmunoprecipitación
Cys	Cisteína
DC	Dicroísmo circular
DO ^{600nm}	Densidad óptica medida a 600 nm
EST	(<i>Expressed sequence tag</i>) Marcador de secuencia expresada
Gal4-AD	Dominio de activación del factor de transcripción Gal4
Gal4-BD	Dominio de unión al ADN del factor de transcripción Gal4
GFP	(<i>Green Fluorescent Protein</i>) Proteína verde fluorescente
Gln	Glutamina
Glu	Ácido glutámico
Gly	Glicina
HA	Epítipo Hemaglutinina
His	Histidina
I2	<i>Inhibitor 2</i> de PP1
IDP	(<i>intrinsically disordered protein</i>) Proteínas intrínsecamente desordenadas
Ile	Isoleucina
Inh3	<i>Inhibitor 3</i> de PP1

Abreviaturas

Leu	Leucina
Lys	Lisina
Met	Metionina
MYPT1	<i>Myosin phosphatase targeting subunit 1</i>
NIPP1	<i>Nuclear inhibitor of PP1</i>
PAGE	(<i>polyacrylamide gel electrophoresis</i>) Electroforesis en gel de poliacrilamida
Phe	Fenilalanina
PNUTS	<i>PP1 nuclear targeting subunit</i>
PPKL	(<i>Protein Phosphatases with Kelch-Like domains</i>) Fosfatasas con dominios Kelch
PPM	(<i>Metallo-dependent Protein Phosphatase</i>) Fosfatasas dependientes de metal.
PPP	(<i>PhosphoProtein Phosphatase</i>) Fosfatasas de fosfoproteínas
Pro	Prolina
PTP	(<i>Protein Tyr Phosphatases</i>) Fosfatasas de Tirosinas
PVDF	(<i>Polyvinylidene difluoride</i>) Fluoruro de polivinilideno
QTL	<i>Quantitative trait locus/loci</i>
RLCK	<i>Receptor-like cytoplasmic kinase</i>
RMN	Resonancia magnética nuclear
SD	(<i>Synthetic defined</i>) Medio mínimo definido para cultivo de levaduras
SDS	(<i>Sodium Dodecyl Sulfate</i>) Dodecilsulfato sódico
Ser	Serina
SN	Sobrenadante de centrifugación
TFE	trifluoroetanol
Thr	Treonina
Trp	Triptofano
Tyr	Tirosina
UV	Ultravioleta
Val	Valina
WT	(<i>Wild type</i>) Estado silvestre
Y2H	(<i>Yeast two hybrid</i>) Doble híbrido en levaduras

Capítulo I

Caracterización evolutiva y funcional de las fosfatasas con dominios

Kelch en *Arabidopsis thaliana*

Introducción

La fosforilación de proteínas

La adición o remoción de grupos fosfatos sobre ciertos residuos aminoacídicos constituye un evento clave en la regulación de la actividad de innumerables proteínas. Es también uno de los sistemas más universalmente usados por las células para acoplar la percepción de estímulos con los mecanismos de respuesta, en el proceso denominado transducción de señales.

En todos los organismos existen familias de enzimas que median en estos procesos. Las quinasas catalizan la formación de enlaces entre grupos fosfato y aminoácidos de las proteínas; el grupo transferido proviene, en primera instancia, del fosfato γ del ATP. El sistema retorna a su estado inicial por la acción de fosfatasas, que hidrolizan estos enlaces mixtos.

Las principales modalidades de fosforilación que participan en las vías de transducción de señales son:

- a) la fosforilación de los átomos de Nitrógeno del anillo imidazólico de residuos Histidina (His), con formación de enlaces fosforamidato;
- b) la formación de fosfomonoésteres entre el grupo fosfato y los grupos hidroxilo de residuos Serina (Ser), Treonina (Thr) o Tirosina (Tyr).

El primer tipo es más prevalente en eubacterias, mientras que el segundo es el más usado por las células eucariotas, aunque en los dos grupos se encuentran ejemplos de ambos mecanismos (Hwang et al., 2002; Kennelly, 2002; Deutscher and Saier, 2005; Pereira et al., 2011).

Las fosfohistidinas, con una energía libre de hidrólisis grande, son especialmente inestables en el medio intracelular y por ello actúan como intermediarias, cediendo el grupo fosfato a otros aceptores (Puttick et al., 2008). Es así como en el sistema de señalización llamado de “dos componentes” la percepción del estímulo desencadena la autofosforilación de la proteína sensora en una His. El grupo fosfato es rápidamente transferido al carboxilo γ de un residuo Aspártico (Asp) en el dominio receptor de la proteína reguladora de la respuesta,

formando un enlace anhídrido que también es inestable y permite en ciertos casos continuar las reacciones de transferencia de fosfatos hacia otros aceptores. La tendencia a la hidrólisis espontánea de estos enlaces llevó por un tiempo a considerar a la defosforilación activa como de limitada relevancia en este contexto. Sin embargo, la defosforilación de la proteína aceptora es esencial para controlar los niveles de la forma activa de la misma y así controlar la intensidad de la señal. En muchos casos, es la misma quinasa de His la que opera la actividad de fosfatasa del residuo Asp aceptor, asegurando así la especificidad de la reacción (Klumpp and Krieglstein, 2009; Dworkin, 2015).

Los enlaces fosfomonoéster son, en cambio, mucho más estables en el medio intracelular, y las formas fosforiladas de las proteínas son fácilmente detectables y analizables. Por esta razón, es esencial la acción de fosfatasas que modulen la intensidad de la señal al permitir que el sistema retorne a su estado basal para comenzar un nuevo ciclo de señalización.

Hasta el momento se han descrito tres grandes grupos de hidrolasas de fosfomonoésteres: 1) las fosfatasas dependientes de metales que hidrolizan grupos fosfato asociados a residuos Ser y Thr, pertenecientes principalmente a las familias PPP (por *phosphoprotein phosphatase*) y PPM (por *metallo-dependent protein phosphatase*), 2) las denominadas PTP (por “protein Tyr phosphatases”), categoría que engloba enzimas muy diversas que actúan sobre una gran variedad de sustratos, incluyendo grupos fosfato asociados a residuos Tyr, y que tienen como característica común la presencia de un residuo Cisteína (Cys) reactivo en el centro activo, y 3) el grupo de fosfatasas que actúa sobre residuos Ser y Thr pero cuyo mecanismo catalítico está basado en un motivo conservado de residuos Asp. En varios de estos casos la especificidad de sustrato no es absoluta, lo que hace más complejo el panorama (Moorhead et al., 2009). En este trabajo nos centraremos en un grupo particular que pertenece a la superfamilia PPP de fosfatasas de Ser/Thr, y por lo tanto no discutiremos las propiedades de las fosfatasas de Tyr ni de las basadas en Asp.

Aunque el descubrimiento de la actividad fosfatasa precedió en una década al de la actividad quinasa, durante mucho tiempo se consideró a las fosfatasas la contraparte pasiva y sujeta a menos eventos regulatorios de los procesos de señalización, encargadas de balancear la actividad de las mucho más reguladas y específicas quinasas (Brautigam, 2013). Contribuyó a ello el hecho de que en ensayos *in vitro* las fosfatasas aisladas muestran poca selectividad

por sustratos, y que los dominios catalíticos de fosfatasa de Ser/Thr, especialmente las del grupo PPP, son siempre menos variables que las de quinasas, lo que sugeriría una menor especialización funcional. Esta última observación coincide con lo encontrado en general al enumerar los genes correspondientes en genomas completamente secuenciados. Por ejemplo, se estima que el genoma humano codifica unas 500 quinasas: unas 100 específicas para Tyr y unas 400 para Ser/Thr. Sin embargo, se cuentan menos de 150 genes para fosfatasa, con unas 100 específicas para Tyr (comparable al número de quinasas), pero menos de 50 específicas para Ser/Thr. En las plantas terrestres no se encuentran quinasas de Tyr del tipo encontrado en animales. Sin embargo, se detectan Tyr fosforiladas, posiblemente por quinasas con menor especificidad de sustrato o por efecto de proteínas homólogas a las llamadas en animales “quinasas de doble especificidad” (Ghelis, 2011). En *Arabidopsis thaliana* los genes de quinasas representan un 4% del genoma y aunque en conjunto hay más genes codificando fosfatasa específica para Ser/Thr en el genoma de *Arabidopsis* que en el humano, unas 70 PPM y unas 20 PPP, son escasas respecto del total de quinasas, que suman unas 1100 (Moorhead et al., 2009). Tanto en animales como en plantas, entonces, el número de genes que codifican dominios activos de fosfatasa de Ser/Thr representa alrededor de un 10% del total de quinasas con esa especificidad. La evidencia indica, sin embargo, que la complejidad de regulación de estas enzimas es en todo comparable al de las quinasas.

Las Fosfatasa de Serinas/Treoninas

La interpretación de las relaciones entre las fosfatasa de Ser/Thr ha sufrido cambios, desde clasificaciones iniciales basadas en las preferencias por ciertos sustratos o en la sensibilidad a inhibidores, hasta ordenamientos más recientes con mayor sentido filogenético basados en secuencias y estructuras. La nomenclatura de estas enzimas, en consecuencia, arrastra las marcas de este devenir. Por ejemplo, las PPM se describieron inicialmente como aquellas fosfatasa cuya actividad *in vitro* depende de la presencia de un metal divalente. Esta definición no es sin embargo discriminante ya que, dependiendo de las condiciones, las PPP también pueden requerir de la presencia de metales en el medio de reacción. Otro ejemplo es el de la clasificación inicial de las PPP en dos grupos, PP1 y PP2, por su especificidad de sustrato y sensibilidad a la inhibición por ciertos factores (Ingebritsen and Cohen, 1983). Hoy se acepta que el grupo 1 (PP1) es una entidad monofilética que agrupa a proteínas

relacionadas, mientras que el grupo 2 es polifilético e incluye proteínas de tipo PPP y algunas de tipo PPM (como las PP2C). El repertorio se hace más complejo a medida que se secuencian genomas de organismos no modelo y se encuentran formas ausentes en hongos y animales, los grupos sobre los que se basaron las clasificaciones iniciales. La clasificación actual se basa en comparaciones de secuencias y resulta más consistente ya que refleja las relaciones evolutivas dentro de la familia (Shi, 2009).

Las fosfatasa de tipo PPP están encargadas del grueso de eventos de defosforilación de Ser/Thr en las células eucarióticas (Brautigan, 2013), y por lo tanto participan en una enorme cantidad de procesos en diversas localizaciones celulares, desde el metabolismo del glucógeno donde fueron inicialmente descubiertas, hasta la dinámica del citoesqueleto, la regulación del ciclo celular, la modulación de vías de señalización y la actividad de factores de transcripción. Las PPP se caracterizan por un dominio catalítico mínimo de unos 280 aminoácidos, cuya secuencia está entre las más conservadas en eucariotas (Shi, 2009). Esta superfamilia incluye, en plantas terrestres, los siguientes grupos:

- PP1, que incluye a las PP1 típicas; el dominio catalítico de los miembros de esta familia no presenta adiciones y actúa asociándose de forma combinatoria con numerosas subunidades regulatorias (ver más abajo).
- PP2 (también llamadas PP2A), PP4 y PP6; el dominio catalítico en estas enzimas tampoco presenta adiciones, y la forma activa adopta la forma de un complejo trimérico entre una subunidad catalítica y una subunidad de andamiaje que la conecta con una de varias subunidades regulatorias.
- PP5, que en todos los genomas eucarióticos analizados es un gen de copia única, que codifica una proteína compuesta por dos dominios: un dominio N-terminal tetratricopéptido y un dominio catalítico C-terminal.
- Tres grupos (exclusivos de plantas): 1) PP7, cuyo dominio catalítico presenta dos adiciones características: una inserción rica en aminoácidos cargados y una extensión C-terminal; 2) las fosfatasa de tipo PPKL (por *Protein Phosphatases with Kelch-like domains*) compuestas por un dominio N-terminal con repeticiones de tipo Kelch y un dominio catalítico C-terminal, que serán el objeto de esta tesis, y 3) las fosfatasa de tipo SLP (por *Shewanella-like Phosphatases*), que presenta variaciones de secuencia

características en el dominio catalítico. Estas subfamilias de fosfatasa, debido a que no se encuentran en animales o levaduras, han sido poco explorados funcional y estructuralmente (Uhrig et al., 2013).

En la figura 1.1 se muestran las relaciones entre los representantes de los distintos grupos codificados en el genoma de la angiosperma *Arabidopsis thaliana*.

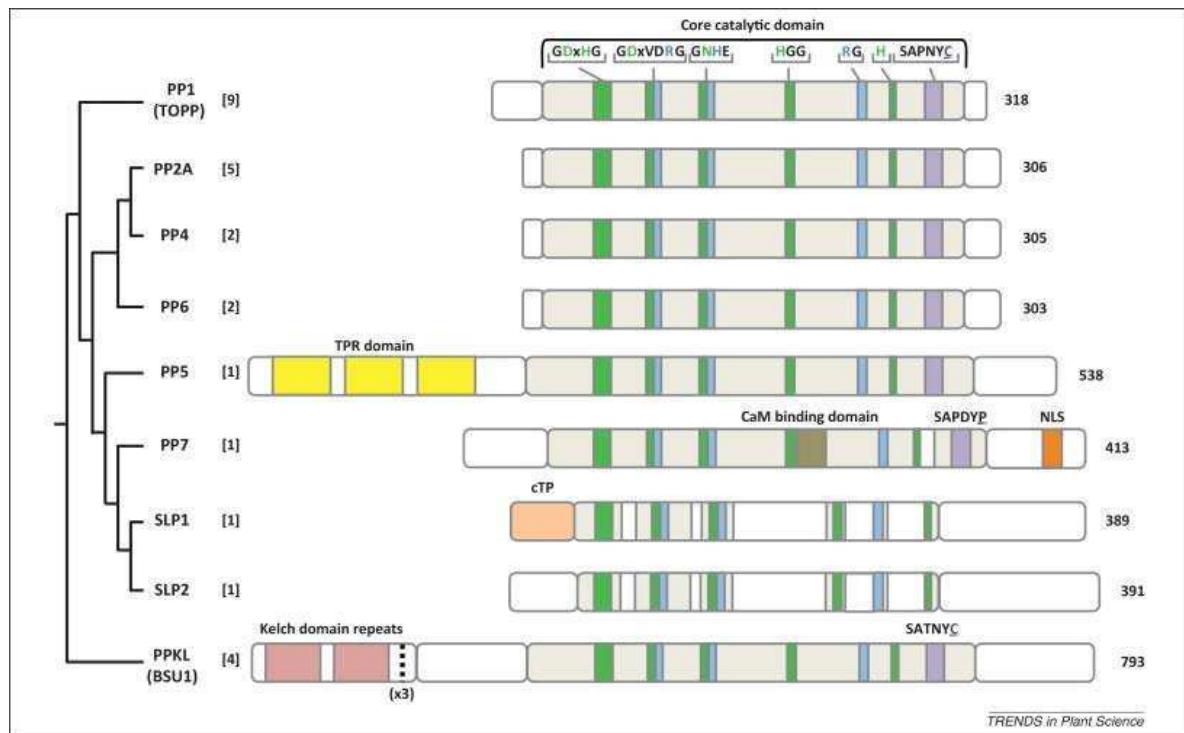


Figura 1.1 - Estructura y relaciones filogenéticas de los miembros de la familia PPP en *Arabidopsis thaliana*. El dominio catalítico conservado está señalado como una caja gris, en la que se indican los motivos de secuencia característicos de la familia; variaciones en estos motivos son diagnósticos de subfamilias. También se muestran otras características propias de cada subfamilia, como el dominio TPR (tetratricopéptido) en PP5, el dominio con repeticiones Kelch en PPKL (sólo se muestran dos de las seis repeticiones predichas), la señal de localización nuclear (NLS) en PP7 y el péptido de tránsito a cloroplastos (cTP) en SLP1. También se indica, en la segunda columna, el número de parálogos presentes en el genoma, y al final de cada representación, el número total de aminoácidos. En plantas no se encuentran secuencias de tipo PP2B/PP3 (calcineurina) canónicas. Tomado de (Uhrig et al., 2013)

Los miembros principales de las PPM son las PP2C y la fosfatasa de piruvato deshidrogenasa. Las PPM comparten un dominio catalítico de unos 290 aminoácidos con menor grado de conservación que el de las PPP, al que se agregan extensiones N- o C-terminales que constituyen dominios de interacción o regulatorios. Al contrario de la multiplicidad de acciones de las PPP, las PP2C parecen actuar principalmente terminando o atenuando

eventos de señalización asociados con respuestas a estrés (Schweighofer et al., 2004; Lammers and Lavi, 2007).

Aunque no parecen compartir un ancestro común, PPP y PPM presentan una estructura general y un mecanismo catalítico similares. La estructura básica del dominio catalítico de ambos grupos está organizada alrededor de un sandwich de dos láminas β rodeado por hélices α y, en ocasiones, láminas β adicionales. El centro activo, localizado en un borde del sandwich β , incorpora dos átomos metálicos coordinados por aminoácidos situados en las conexiones entre las hebras β y por moléculas de agua asociadas. Estos aminoácidos altamente conservados definen a cada familia. Así, las PPP se caracterizan por las secuencias GDxHG, GDxVDRG y GNHE, mientras que las PPM se caracterizan por una serie de residuos ácidos (D y E). Los iones metálicos han sido identificados como Mn^{2+} y Fe^{2+} en PP1 (Egloff et al., 1995), Mn^{2+} y Zn^{2+} en PP3 (Wozniak et al., 2000) y Mn^{2+} en PP2, PP5 y PP2C (Das et al., 1996).

La reacción de defosforilación procede en un solo paso y sin un intermediario covalente con la enzima. Se supone que procede por un mecanismo S_N2 en el que una molécula de agua lleva a cabo el ataque nucleofílico sobre el fósforo, de forma que el grupo saliente es la proteína que está siendo defosforilada. Los metales facilitan la reacción posicionando el grupo fosfato en la orientación correcta, aumentando la electrofilicidad del fósforo y estabilizando el anión hidroxilo atacante que provendría de la molécula de agua situada entre ambos átomos metálicos. La carga negativa del intermediario de reacción está estabilizada por una His en las PPP o por una Arg en las PPM (Egloff et al., 1995; Das et al., 1996). En la figura 1.2 se muestra un esquema de las estructuras de los centros activos de PPP y PPM y del mecanismo de reacción.

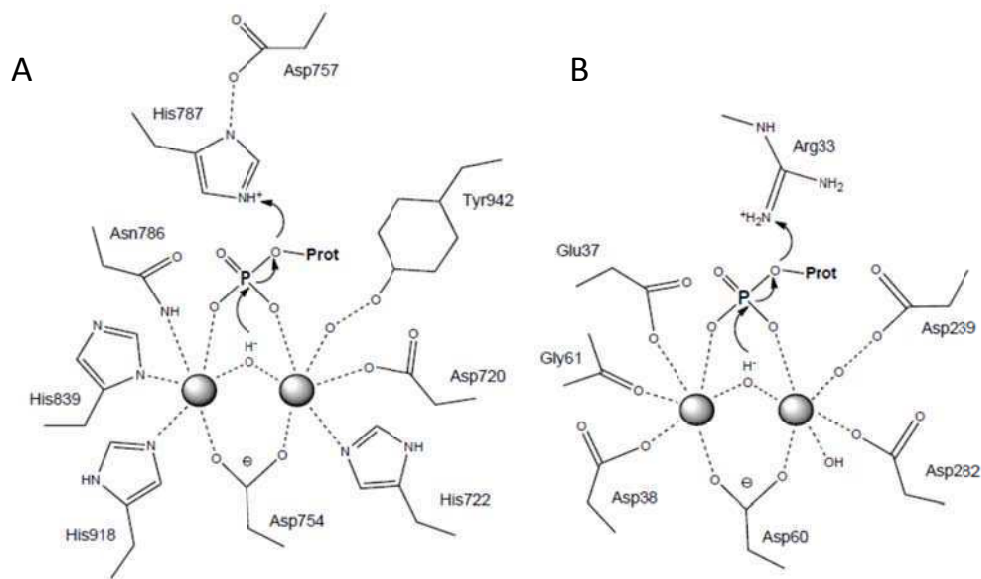


Figura 1.2 - Estructura del centro activo y mecanismo catalítico en las fosfatasa PPP y PPM. Se muestran los metales, coordinados por la proteína, y el mecanismo propuesto de ataque S_N2 sobre el grupo fosfato en A) PPP (numeración de AtBSL2) y B) PPM (numeración de PP2C α humana).

Las Fosfatasa PPKL

Dentro de los grupos de PPP ausentes en animales y hongos que surgen al extender el análisis a nuevos grupos de organismos, aparecen secuencias con peculiaridades tales que las hacen formar categorías propias. Entre éstas se encuentra un grupo caracterizado por dos dominios claramente definidos: un dominio catalítico C-terminal relacionado en secuencia con PP1, precedido por una larga extensión N-terminal que incluye un dominio de tipo turbina- β con repeticiones Kelch y, a modo de conector entre ambos dominios, una secuencia extensa y variable. Estas fosfatasa, como se mencionó anteriormente, han sido nombradas fosfatasa de proteínas con dominios de tipo Kelch (PPKL (Kutuzov and Andreeva, 2002; Moorhead et al., 2009; Uhrig et al., 2013)).

La primera descripción de una proteína de este tipo fue hecha en *Plasmodium falciparum* durante una búsqueda de quinasas y fosfatasa con el objeto de encontrar posibles blancos terapéuticos, dado que estas proteínas no se encuentran en los hospedadores animales. En este organismo sólo se caracterizó su expresión, que parece estar restringida a la fase sexual (Li and Baker, 1998). *Plasmodium* es un parásito del grupo Apicomplexa, que a su vez forma parte de uno de los grandes reinos que componen la diversidad de los eucariotas: el de los Alveolados. Los Alveolados son un grupo muy extendido y reconocidamente monofilético,

pero altamente diverso, de protistas de compleja organización celular. Como las PPKL se encuentran tanto en ciliados, una rama tempranamente divergente de este *phylum*, como en dinoflagelados, apicomplejos y grupos asociados como los perkinsidos, se asume que estas proteínas estaban presentes en el ancestro común del grupo (Bachvaroff et al., 2011).

Además de los Alveolados, secuencias de tipo PPKL se encuentran solamente en otro gran grupo de eucariotas: las Viridiplantas, definidas como los organismos fotosintéticos derivados de una endosimbiosis primaria y que acumulan clorofila *a* y *b* como pigmentos cloroplásticos principales. Secuencias de tipo PPKL se encuentran en los dos grandes grupos en que se organizan estos organismos: *Chlorophyta* (formado por organismos fotosintéticos verdes acuáticos marinos o de agua dulce, con nivel de organización unicelular a multicelular simple) y *Streptophyta* (organismos fotosintéticos acuáticos y terrestres, caracterizados por la formación de una estructura llamada fragmoplasto durante la citoquinesis, uni- a multicelulares con niveles complejos de organización, como el caso de las plantas terrestres).

El análisis de las secuencias PPKL en Alveolados y Viridiplantas no permite, sin embargo, establecer relaciones entre ellas, dado que no se encuentran evidencias convincentes de secuencias ancestrales o derivadas.

El primer miembro de la familia PPKL para el que se asignó una posible función se halló en *Arabidopsis thaliana*. *BSU1* (por *BRI1 Suppressor 1*) se identificó como un modulador positivo de la señalización por los compuestos esteroideos llamados brassinosteroides (BRs) (Mora-García et al., 2004). En este caso, se observó que la sobreexpresión de *BSU1* suprimía parcialmente el fenotipo enano de plantas deficientes en la percepción de la hormona.

Los BRs son esteroides polihidroxilados que actúan de forma aparentemente parácrina, en la modulación de la división y la elongación celular. Defectos en su producción y/o percepción acarrearán una serie de manifestaciones fenotípicas características: enanismo extremo, infertilidad, alteraciones del desarrollo vascular y en las respuestas a otros estímulos, como la luz y otras hormonas (Clouse, 2011). Un esquema simplificado de la vía de transducción de la señal se muestra en la Figura 1.3.

La señalización por BRs comienza con la percepción de la hormona por un complejo receptor de membrana formado por un receptor, *BRI1*, y un co-receptor, *BAK1*. La unión de la

hormona al dominio extracelular de BRI1 (Kinoshita et al., 2005; Hothorn et al., 2011) estimula la formación de un complejo entre ambas proteínas y desencadena una serie de reacciones de transfosforilación de sus dominios quinasa intracelulares (Wang et al., 2008). La fosforilación por parte de BAK1 de una región comprendida entre el paso transmembrana y el dominio quinasa propiamente dicho de BRI1 libera la proteína inhibitoria BKI1 para así activar la vía intracelular de señalización (Wang and Chory, 2006). El dominio quinasa de BRI1 así activado fosforilaría a una familia de quinasas de tipo receptor (RLCK, por *Receptor-like cytoplasmic kinase*) con dominios tetratricopéptido y asociadas a membranas llamadas BSK (por *Brassinosteroid-signaling kinase* (Tang et al., 2008)). La señal emanada del complejo receptor activado inhibe en última instancia a un grupo de quinasas de tipo GSK3, entre las que la primera identificada fue BIN2 (por *Brassinosteroid Insensitive 2* (Li et al., 2001; Yan et al., 2009). Estas GSK3, en ausencia del estímulo, fosforilan de forma constitutiva a una familia de factores de transcripción de tipo *basic helix-loop-helix* (bHLH) específicos de plantas de la familia BES1/BZR1 (Yin et al., 2002), que son los encargados de mediar la respuesta transcripcional asociada a la percepción de BRs. Estos factores de transcripción son fosforilados por BIN2 en varios residuos Ser/Thr, uno de los cuales está incluido en una secuencia de reconocimiento por proteínas 14-3-3. La unión a proteínas 14-3-3 en el estado fosforilado promueve la salida de BES1/BZR1 del núcleo y/o su retención en el citoplasma (Gampala et al., 2007). Por otro lado, se postula que la fosforilación de BES1/BZR1 las convierte en sustrato de ubiquitinación, con su consiguiente degradación por la vía del proteasoma. Cuando BIN2 es inhibida por la percepción de BRs, las formas defosforiladas se acumulan rápidamente en el núcleo por acción de PP2A asociada a subunidades regulatorias de tipo B' y así pueden ejercer su acción (Tang et al., 2011). De esta forma, la señalización por BRs procede por inhibición de un inhibidor de la vía.

Se encontró que la sobreexpresión de *BSU1* promueve la acumulación de las formas defosforiladas de BES1/BZR1, por lo que se postuló que contrarrestaría la acción de BIN2 (Mora-García et al., 2004). Sin embargo, recientemente se postuló que *BSU1* no defosforilaría directamente a BES1/BZR1 sino que inactivaría a BIN2 (Kim et al., 2009). Las quinasas GSK3 están constitutivamente fosforiladas en una Tyr en la zona de activación. La proteína se autofosforila en esa Tyr apenas terminada su síntesis en una reacción dependiente de chaperonas: una vez modificada, adopta su actividad “madura” de quinasa

intermolecular de Ser y Thr asociadas a Pro (Lochhead et al., 2006). En el modelo de señalización por BRs imperante, el estímulo lleva a la fosforilación de las kinasas BSK, que se asociarían (aunque no se demostró interacción directa) con BSU1 y quizás sus homólogos, activándolas (por un mecanismo no caracterizado pero que no involucraría fosforilación) para que defosforilen a BIN2 en esa Tyr conservada (Kim et al., 2009).

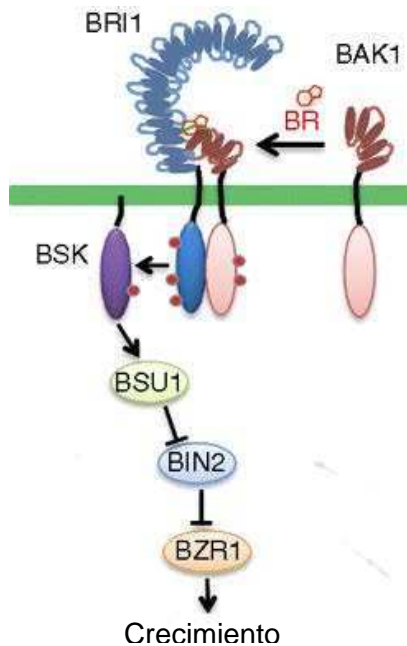
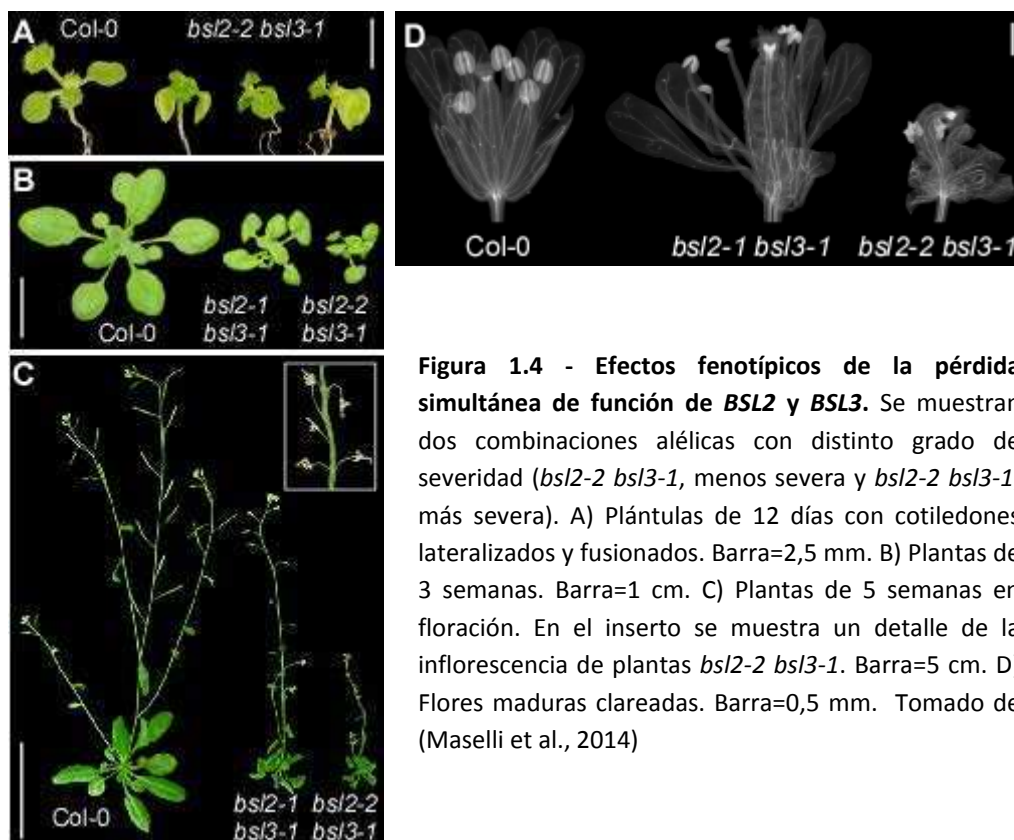


Figura 1.3 - Representación esquemática de la vía de transducción de señales de BRs. Se muestran algunos de los componentes descritos en el texto. Adaptada de (Wang, 2012)

BSU1 es uno de los cuatro genes que conforman la subfamilia PPKL en *Arabidopsis*; los otros homólogos se denominaron *BSL1* a *3* (por *BSU1-Like* (Mora-Garcia et al., 2004)). Por esta razón, denominaremos genéricamente a este grupo de fosfatasa como de tipo BSL cuando nos refiramos a los miembros de la familia en *Arabidopsis*. Los tres homólogos de *BSU1*, *BSL1*, *BSL2* y *BSL3*, han recibido comparativamente menos atención y se asumió que desempeñan roles redundantes en la vía (Mora-Garcia et al., 2004; Kim et al., 2009; Kim et al., 2011). De hecho, la pérdida de función simultánea de *BSU1* y de su parálogo más cercano, *BSL1*, en combinación con la disminución simultánea de expresión para *BSL2* y *BSL3* (llamado mutante *bsu-q*) produjo plantas enanas aparentemente defectuosas en la señalización por BR (Kim et al., 2009; Kim et al., 2011). Sin embargo, recientemente se obtuvieron en arroz (*Oryza sativa*) evidencias de actividades contrastantes entre los miembros de la familia. En esta especie se identificó un polimorfismo de secuencia en el dominio Kelch de un homólogo de *BSL2* (*GL3.1*) como un QTL (Quantitative trait locus) que impacta en la longitud del grano (Hu et al., 2012; Qi et al., 2012; Zhang et al., 2012). La

longitud del grano se incrementó en plantas con pérdida de función para *GL3.1* o su parólogo más cercano, pero decreció en plantas que no poseían el homólogo restante, relacionado con *BSL1* (Zhang et al., 2012).

Como parte de los análisis previos llevados a cabo en el laboratorio, se identificaron líneas de *Arabidopsis* con pérdida de expresión de estos genes por inserciones de T-DNA en ellos. De esta forma se encontró que la falta de expresión de los distintos homólogos, de forma individual, no produce efectos fenotípicos apreciables. Sin embargo, la pérdida simultánea de expresión de *BSL2* y *BSL3*, dos homólogos muy similares, produjo una serie de alteraciones novedosas. Las plántulas mutantes dobles *bsl2 bsl3* son pequeñas, con cotiledones epinásticos que sufren distintos grados de fusión lateral. Cuando adultas, las plantas son pequeñas y de hojas plegadas, con tallos de inflorescencias fasciados y fusiones generalizadas entre pecíolos florales y el tallo principal. Las flores se vieron particularmente afectadas, con alteraciones en la forma de los órganos, procesos de fusión de los filamentos de estambres y alteraciones en el desarrollo del gineceo. Estas mutantes dobles resultaron estériles (Figura 1.4).



Objetivos

Para profundizar el estudio de las funciones de este grupo de fosfatasas de proteínas en plantas y poner a prueba algunos de los supuestos del modelo de señalización por BRs, nos propusimos realizar un análisis detallado y comparativo de los miembros de la familia presentes en *Arabidopsis thaliana*.

Resultados

Análisis filogenético de las fosfatasas PPKL en plantas

Con el fin de clarificar las relaciones filogenéticas de las fosfatasas PPKL en plantas, se buscaron secuencias relacionadas en genomas completamente secuenciados y colecciones de secuencias EST de organismos verdes, incluyendo algas de los grupos Chlorophyta y Streptophyta. En todos los casos, las secuencias obtenidas se controlaron manualmente, ya que comprobamos que numerosas anotaciones automáticas de genomas generaban secuencias anómalas que, al ser analizadas en detalle, revelaban errores en la asignación de sitios de splicing; la alta conservación de secuencias de los dominios Kelch y catalítico facilitó esta tarea. Las secuencias proteicas usadas en los siguientes análisis se vuelcan en el Anexo I (adjunto en la versión digital).

La figura 1.5 muestra el árbol filogenético obtenido a partir de secuencias aminoacídicas de longitud completa y algunas secuencias derivadas de ESTs que resultaron ser lo suficientemente largas como para permitir una asignación robusta. También se construyó un árbol similar usando sólo el dominio catalítico C-terminal altamente conservado, para maximizar la astringencia de las comparaciones y para incluir algunos grupos taxonómicamente relevantes para los cuales sólo se encontraron secuencias provenientes del extremo 3' de ESTs (Figura 1.6).

Comprobamos que, en los organismos verdes, las PPKL se encuentran tanto en microalgas unicelulares como *Ostreococcus tauri* y *Micromonas pusilla*, que se cuentan entre los eucariotas de células más pequeñas conocidos (Vaulot et al., 2008), como en las plantas terrestres multicelulares. En las algas, aparentemente, estos genes serían de copia única. En cambio, en las plantas terrestres, pueden ser agrupados en dos clados claramente definidos (que denominaremos según los nombres de sus representantes en Arabidopsis): el clado BSL2/BSL3, con miembros desde hepáticas hasta angiospermas, y el clado BSL1, donde secuencias *bona fide* se detectan primero en *Cycas rumphii*, un miembro del grupo de gimnospermas aún existentes que se asume divergió más tempranamente (Brenner et al., 2003). La falta de representación en el muestreo de secuencias de ciertos grupos hace que sea difícil confirmar la presencia de secuencias de tipo BSL1 en ellos. Por ejemplo, el genoma

de la licofita *Selaginella moellendorffii* codifica para dos genes de tipo BSL: uno de ellos (Sm32g01790) claramente pertenece al grupo BSL2/BSL3, mientras que el otro (Sm3g04900) muestra una posición inestable en los árboles, pero tiende a agruparse dentro del clado BSL1. Las secuencias de ESTs disponibles del helecho *Pteridium aquilinum* [(Der et al., 2011) no incluidas en los árboles pero presentes en el Anexo I, ver versión digital] revelaron la presencia de secuencias de tipo BSL2/BSL3 (contig_13341 and contig_12782), pero no fue posible asignar inequívocamente ninguna al clado BSL1. Sin embargo, una búsqueda entre las secuencias disponibles de forma preliminar derivadas del proyecto OneKP de secuenciación de mRNA obtenido a partir de 1000 especies de plantas (<https://sites.google.com/a/ualberta.ca/onekp/>) mostró que en helechos en efecto se encuentran tanto secuencias de tipo BSL2/BSL3 como otras característicamente diferentes que podrían ser consideradas formas ancestrales de las de tipo BSL1. Estas secuencias no están aún disponibles de forma pública y por esa razón no fueron incluidas en el análisis. Las algas carofitas (Timme et al., 2012) aparecen como un grupo hermano de las plantas terrestres en el árbol construido con las secuencias del dominio C-terminal (Figura 4); sin embargo, la secuencia de EST más larga disponible (proveniente de *Spirogyra pratensis*) se asocia preferentemente con el clado BSL2/BSL3 en el árbol construido con las secuencias completas (Figura 1.5). En resumen, las secuencias de tipo BSL2/BSL3 mantendrían características ancestrales, mientras que las de tipo BSL1 serían una innovación de las plantas vasculares.

En conjunto, los clados BSL2/BSL3 y BSL1 se caracterizan por tener ramas cortas, que indican una conservación de secuencia robusta. Por el contrario, las secuencias de tipo BSU1 se destacan como parálogos divergentes de la subfamilia BSL1, con ramas muy largas y con variaciones importantes incluso entre especies cercanamente relacionadas, lo que sugiere una aceleración del ritmo evolutivo. Notablemente, secuencias de tipo BSU1 sólo se encuentran en miembros de la familia Brassicaceae, a la que pertenece *Arabidopsis*; en las búsquedas realizadas esta familia se presentó como la única entre las angiospermas que presenta un grupo divergente de proteínas BSL.

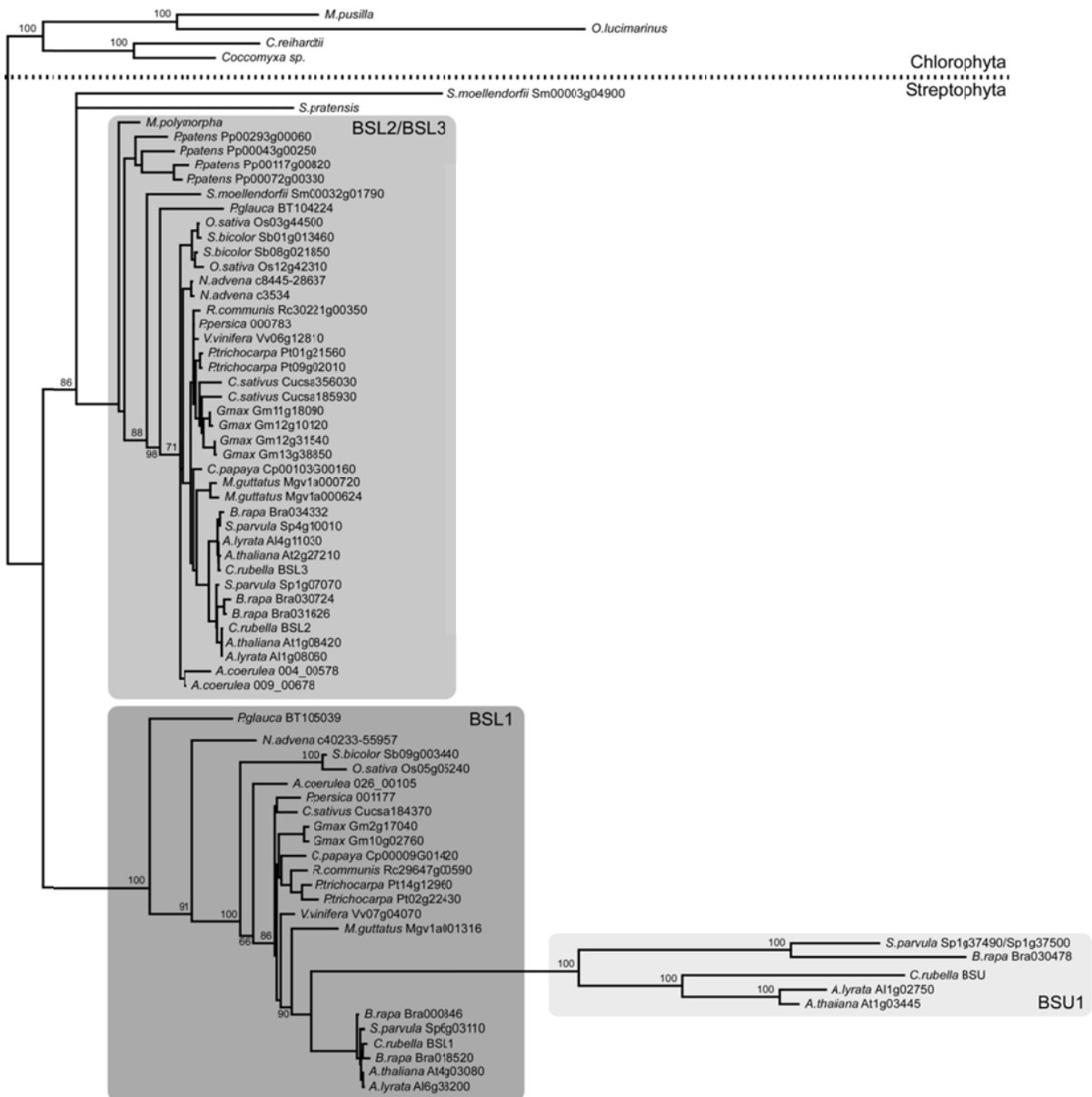


Figura 1.5 - Relaciones filogenéticas entre los genes BSL. Árbol filogenético de máxima verosimilitud construido con secuencias aminoacídicas completas de genes BSL de Viridiplantae; se usaron secuencias derivadas de genomas completamente secuenciados o las secuencias de ESTs más largas disponibles. Las secuencias se alinearon con Muscle, y las ramas del árbol se validaron con 100 repeticiones. Se muestran valores de Bootstrap mayores al 50%; valores mayores al 50% en nodos estrechamente agrupados se excluyeron para mayor claridad. Tomado de (Maselli et al., 2014).



Figura 1.6 - Relaciones filogenéticas entre los genes BSL. Árbol filogenético de máxima verosimilitud construido con secuencias aminoacídicas alineadas del dominio C-terminal de genes BSL de Viridiplantae. Las secuencias se alinearon con Muscle, y las ramas del árbol se validaron con 100 repeticiones. Se muestran valores de *Bootstrap* mayores al 50%; valores mayores al 50% en nodos estrechamente agrupados se excluyeron para mayor claridad. Tomado de (Maselli et al., 2014).

La divergencia de secuencia no es la única característica inusual de los genes de tipo BSU1. En las plantas terrestres, se ha observado que los patrones de *splicing* están muy conservados en genes con función relevante (Teich et al., 2007). De hecho, las posiciones y fases de los intrones en estas fosfatasa están conservadas desde las hepáticas (*Marchantia polymorpha* en este caso) hasta las angiospermas. La única excepción son los genes de tipo

BSU1: en ellos, el patrón de *splicing* que precede al dominio catalítico no sólo se encuentra rearrreglado respecto de sus homólogos conservados, sino que difiere aún entre especies de la familia Brassicaceae, como sucede en *Arabidopsis thaliana* y *Brassica rapa* (Figura 1.7).

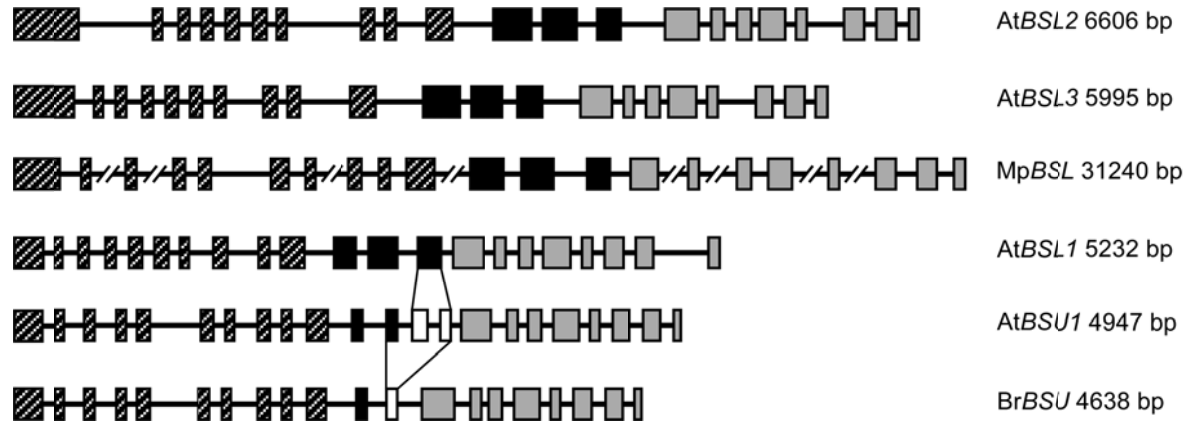


Figura 1.7 - Organización genómica de los genes BSL. Estructura genómica de la región codificante de genes BSL. At: *Arabidopsis thaliana*, Mp: *Marchantia polymorpha*, Br: *Brassica rapa*. Los exones están representados por cajas (rayadas: dominio Kelch, negras: dominio conector, grises: dominio catalítico) y los intrones por líneas. Los exones en blanco indican rearrreglos en los genes de tipo BSU1. Tomado de (Maselli et al., 2014).

Las fosfatasa PPKL de plantas terrestres se agrupan, entonces, en dos clados antiguos y altamente conservados. Las secuencias de tipo BSU1 de Brassicaceas se caracterizan, por el contrario, por un nivel inusualmente alto de divergencia en las secuencias.

Localización subcelular de las fosfatasa PPKL

Las diferencias filogenéticas observadas sugieren que los miembros de esta familia pueden no ser redundantes, como se supuso en un principio. De hecho, en el laboratorio se habían obtenido evidencias de que éste podía ser el caso. Cuando se comenzó con la caracterización de esta familia, se evaluaron los patrones de expresión de estos genes en *Arabidopsis* mediante fusiones de sus regiones promotoras a β -Glucuronidasa (GUS). Los resultados mostraron que en efecto estos genes no se expresan en forma totalmente solapada. Mientras que *BSL2* y *BSL3* mostraron patrones casi idénticos, con baja expresión limitada principalmente a zonas en crecimiento (Figura 1.8 C, D, E, F), *BSL1* mostró un patrón de expresión más intenso y extendido (Figura 1.8 A y B). Es particularmente notable el caso de los granos de polen, en los que *BSL2* y *BSL3* se expresan intensamente mientras que *BSL1* no se expresa en absoluto. Por su parte, la expresión de *BSU1* resultó indetectable, aun cuando

se evaluaron varias líneas transgénicas en diferentes condiciones de tinción (Figura 1.8 G y H).

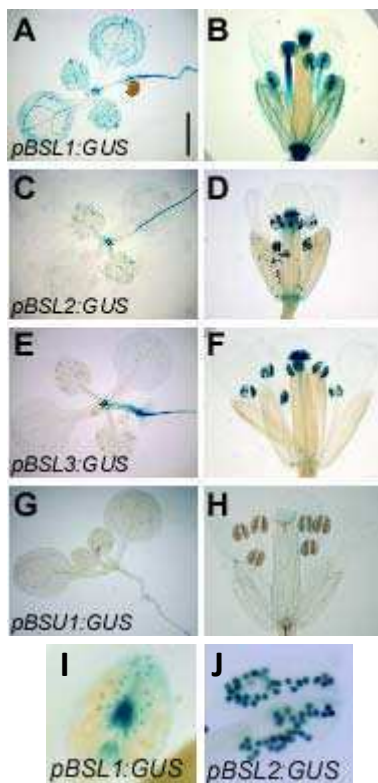


Figura 1.8 - Patrones de expresión de los genes BSL. La expresión se evaluó en plántulas de 10 días (A, C, E y G) y en flores maduras (B, D, F y H) de plantas transformadas con las construcciones indicadas. En I y J se muestra un detalle de las anteras de plantas *pBSL1:GUS* y *pBSL2:GUS*; sólo en el segundo caso se observa tinción en los granos de polen. Barra=2mm. Javier Bianchi, tesis de licenciatura. Tomado de (Maselli et al., 2014).

Para avanzar en la caracterización de estas proteínas, se evaluó su localización subcelular mediante fusiones de las mismas a GFP. Para hacerlo, se expresaron transitoriamente construcciones de ADNc de los cuatro genes homólogos bajo el control del promotor 35S (*p35S:BSL:GFP*) en hojas de *Nicotiana benthamiana* (Figura 1.9 A, C, E y G), y de forma estable en *Arabidopsis* (Figura 1.9 B, D, F, y H). En ambos casos los resultados fueron equivalentes: BSL2 y BSL3 se localizaron en el citoplasma y núcleo, mientras que BSL1 se localizó solamente en el citoplasma y fue excluida del núcleo. BSU1 mostró una localización fuerte en el núcleo y débil en el citoplasma. Cabe resaltar que el patrón observado fue el mismo cuando se utilizó una construcción que expresa BSL2 fusionada a GFP a partir de un fragmento genómico completo y que resulta funcional porque es capaz de complementar el fenotipo de las mutantes dobles *bs/2 bs/3* (Maselli et al., 2014). Patrones similares fueron reportados para BSU1 y BSL1 de *Arabidopsis* (Mora-Garcia et al., 2004; Kim et al., 2009), para un miembro del clado BSL2/BSL3 de arroz (Qi et al., 2012), y para BSL1 de tomate (*Solanum lycopersicum*) expresada en hojas de *N. benthamiana* (Saunders et al., 2012).

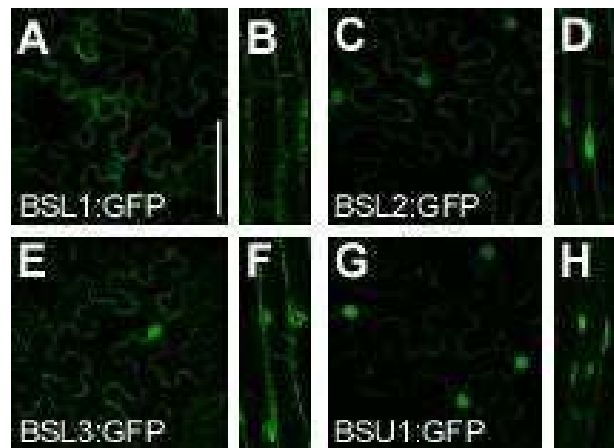


Figura 1.9 - Las fosfatasa BSL de *Arabidopsis thaliana* difieren en su localización subcelular. Se muestra la localización de las versiones fusionadas a GFP de BSL1 (A y B), BSL2 (C y D), BSL3 (E y F) y BSU1 (G y H), expresadas transitoriamente en hojas de *N. benthamiana* (A, C, E, G) o establemente en el hipocotilo de plántulas transgénicas de *Arabidopsis* de 4 días de edad etioladas (B, D, F, H). Barra = 100 μ m. Tomado de (Maselli et al., 2014).

En conclusión, y en concordancia con la evidencia filogenética, se comprobó que el patrón de expresión y la localización subcelular de estas proteínas muestran diferencias características entre los miembros de la familia, lo que podría traducirse en diferencias funcionales que abogarían en contra de la suposición de que las isoformas cumplen acciones solapadas.

Por ello, se buscó poner a prueba sus presuntos roles redundantes en la vía de señalización por BRs.

Interacción de las fosfatasa BSL con participantes en la vía de señalización por BRs.

Para poder definir el papel, si lo tienen, de las fosfatasa BSL en la vía de transducción de señales por BRs, en el laboratorio se combinaron mutaciones en los genes *BSL* con mutantes de esta vía de señalización. A modo de ejemplo, en la Figura 1.10 se muestran los efectos de la combinación de las mutaciones *bs2 bs3* con una mutación de efecto débil en el receptor BRI1 (*bri1-301* (Xu et al., 2008)). Como se observa en la Figura 1.10 A, la combinación de las tres mutaciones tuvo un efecto aditivo, pero no sinérgico, es decir, que el fenotipo resultante fue el producto de la combinación de los efectos fenotípicos de cada mutación, sin producir efectos adicionales, como se habría esperado si estos genes participaran directamente de la misma vía. El estado de fosforilación de BES1 es un indicador sensible de

la intensidad de la señal a través de la vía. La Figura 1.10 B muestra que los niveles de BES1 y el cambio en sus niveles de fosforilación en respuesta al tratamiento con epi-BL (eBL, un análogo activo de BL) no difirieron significativamente entre las mutantes dobles y las plantas silvestres, aunque en las mutantes dobles la proporción de isoformas con distintos grados de (de)fosforilación parece estar alterada: se reduce aparentemente la forma fosforilada de migración electroforética más lenta y se incrementa la intensidad de banda/s intermedia/s entre las formas fosforilada y defosforilada. A modo de control negativo del tratamiento de eBL sobre la vía de señalización, se realizó el mismo tratamiento sobre la mutante *bri1-301*, incapaz de percibir la señal de BR. En este caso, los niveles de BES1 se vieron disminuídos y no se observó respuesta al eBL. Por otro lado, las mutantes dobles *bsl2 bsl3* respondieron de la misma manera que las plantas silvestres al tratamiento con un inhibidor de la síntesis de BRs, Brassinazole (Maselli et al., 2014). Teniendo en cuenta estos resultados, y que la pérdida de función de *BSL1* no produjo alteraciones fenotípicas, no resulta evidente asignar una función a estas proteínas BSL en la vía de señalización por BRs.

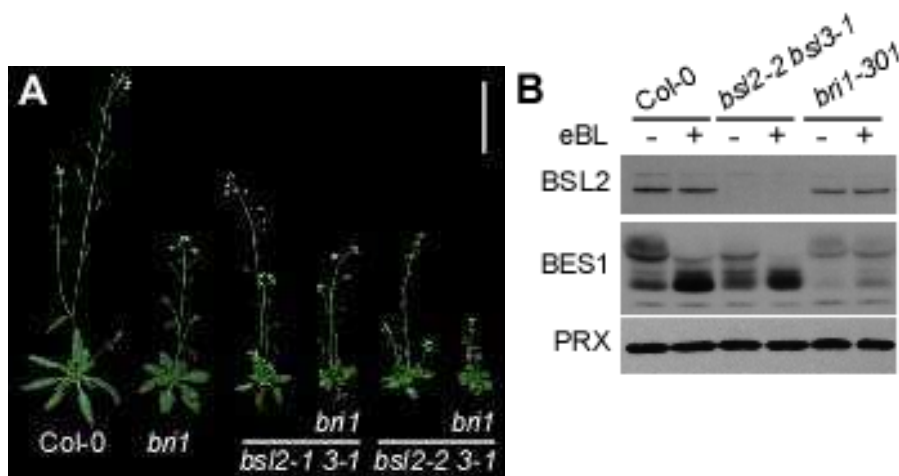


Figura 1.10 - Respuestas a BRs en las mutantes *bsl2 bsl3*. A) Las combinaciones alélicas *bsl2 bsl3* tuvieron un efecto aditivo, pero no sinérgico, sobre un sustrato mutante *bri1-301*. Barra=5 cm. Plantas de 40 días. *bri1* indica homocigosis para *bri1-301*. B) Evaluación de los niveles de fosforilación de BES1 luego de un tratamiento con eBL, en los transfondos genéticos indicados, utilizando *Western blot*. Plántulas de 12 días se trataron con 1 μ M eBL o con la cantidad equivalente de solvente por 2 horas en luz. Los extractos proteicos fueron separados y transferidos a membrana, que fue separada en 3 porciones: la superior (>80kDa) fue revelada con antisuero anti-BSL2 (cuya producción se comenta en el capítulo 3), la intermedia (entre 80 y 30 kDa) fue revelada con antisuero anti-BES1, donación del Dr. Yanhai Yin, Iowa State University, EEUU, y la inferior (<30 kDa) con antisuero anti-2-Cys Peroxirredoxina de cloroplastos (PRX), como control de carga. El círculo negro denota la forma fosforilada de BES1, y el blanco la defosforilada. Tomado de (Maselli et al., 2014).

En paralelo, se puso a prueba la capacidad de las proteínas BSL de interactuar con RLCKs que formarían parte de la cadena de señalización intracelular (Tang et al., 2008). Se postula que las formas activadas de BSKs interactuarían y modularían la actividad de BSU1 (Kim et al., 2009). Mediante la estrategia de doble híbrido en levaduras (yeast two hybrid, Y2H), se evaluó la interacción de las proteínas BSL con algunos miembros de la familia BSK. Se eligió BSK1 debido a que es fosforilada *in vivo* por BRI1 (Tang et al., 2008) y BSK3 porque su ausencia produce un fenotipo débil de deficiencia de señalización por BRs (Tang et al., 2008; Sreeramulu et al., 2013). Encontramos que BSL2 y BSL3 fueron capaces de interactuar con BSK1 y BSK3, mientras que BSL1 interactuó con BSK3, pero sólo débilmente con BSK1 (Figura 1.11 A y B). Sorprendentemente, BSU1 no interactuó con ninguna BSK en este sistema, aunque se comprobó que todas las proteínas alcanzaron niveles de expresión similares en las levaduras (Figura 1.11 C).

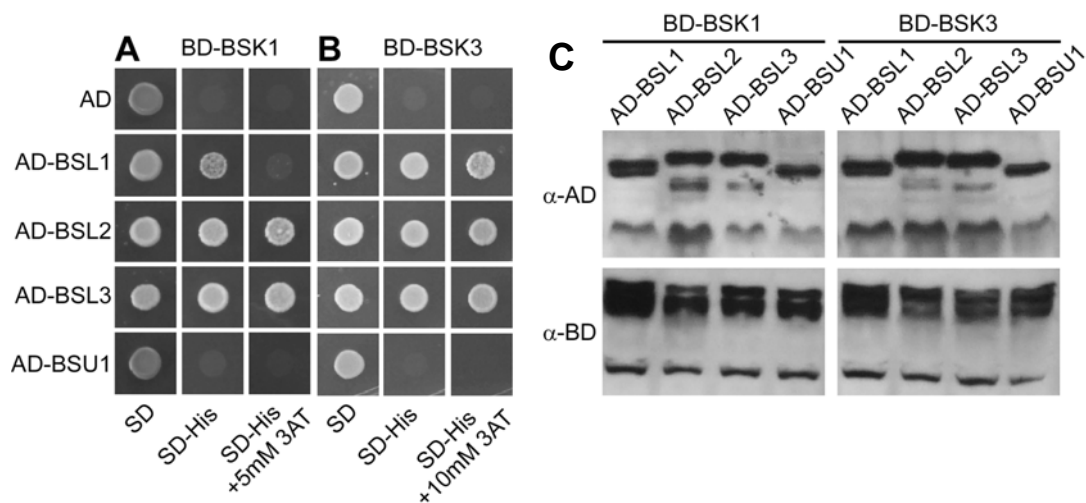


Figura 1.11 - Interacción de las fosfatasa BSL con intermediarios de la vía de señalización por BRs. A) y B) Ensayos de Y2H usando como reportero el crecimiento en ausencia de His. Se evalúa la interacción entre BSK1 (A) y BSK3 (B) con las fosfatasa BSL. SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de las concentraciones indicadas de 3-amino-1,2,4-triazol (3AT); AD: Dominio de trans-activación de Gal4; BD: Dominio de unión al DNA de Gal4. C) Inmunodetección de AD (panel superior) y BD (panel inferior) en extractos proteicos de levaduras expresando BSK1 o BSK3 (como “carnada”) en combinación con las proteínas BSL (como “presas”). Tomado de (Maselli et al., 2014).

En conjunto, estos resultados muestran que las fosfatasa BSL interactúan con intermediarios de la vía de señalización por BRs (Kim et al., 2009; Kim et al., 2011), pero al mismo tiempo muestran que las isoformas poseen diferente, e incluso contrastante, afinidad por las proteínas ensayadas. En particular, los resultados obtenidos se oponen a los de Kim

et al., que muestran la interacción de BSU1 con BSKs (aunque en ese caso no fue ensayada por Y2H) (Kim et al., 2009).

Papel esencial de BSL1, BSL2, y BSL3

Dada la distribución filogenética y las evidencias fenotípicas encontradas en las mutantes dobles *bsl2 bsl3*, se conjeturó que la anulación de la función de los tres genes más conservados de la familia contribuiría a definir mejor el papel de estas proteínas y a echar luz al mismo tiempo sobre la función de BSU1, el miembro más divergente de la familia y específico de Brassicaceae.

Para ello, se procedió a combinar dos alelos de pérdida de función de *BSL1* con las combinaciones *bsl2 bsl3* disponibles. La producción de progenie mutante es posible siempre y cuando el esporofito lleve al menos un alelo normal de *BSL2* o *BSL3*. Dicho en otras palabras, es posible obtener mutantes dobles *bsl2 bsl3* a partir de gametas deficientes para ambos genes, siempre que la planta que da origen a las gametas tenga al menos una copia funcional de alguno de los genes. Como se mencionó antes, cuando no hay copias funcionales de estos genes, tanto el polen como los óvulos de las mutantes dobles son totalmente estériles. De todas formas, se observó que existe un desvío en la segregación de mutantes dobles a partir de plantas homocigotas para un gen y heterocigotas para el otro. Por ejemplo, se obtuvieron 73 mutantes dobles en la progenie de 360 individuos de una planta homocigota para la forma mutante de *BSL2* y heterocigota para *BSL3* (*bsl2-2^{+/-} bsl3-1^{-/-}*), frente a los 90 esperados para cumplir con la relación 1:3 para un alelo recesivo simple. Esto sugiere un efecto de penetrancia sobre el desarrollo de las gametas y/o embriones en la configuración mutante.

En trasfondos genéticos sensibilizados (homocigota *bsl2*/heterocigota *bsl3* o heterocigota *bsl2*/homocigota *bsl3*, que se indican como *bsl2^{-/-} bsl3^{+/-}* o *bsl2^{+/-} bsl3^{-/-}*) la pérdida de función de *BSL1*, que es fenotípicamente silenciosa en un contexto normal para *BSL2* y *BSL3*, causa alteraciones fenotípicas relativamente leves: las plantas que portan sólo una copia funcional de *BSL2* o *BSL3* (además de *BSU1*) son algo más pequeñas, florecen más temprano, y producen inflorescencias más ramificadas que la silvestre (Figura 1.12 A y B). Es destacable que, aunque más pequeñas, estas plantas no se corresponden con el fenotipo típico de mutantes deficientes en la vía de BRs.

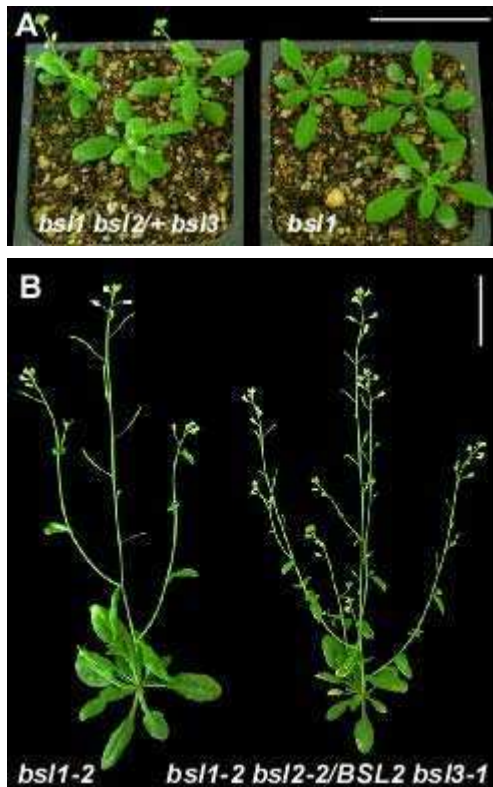


Figura 1.12 - BSL1, BSL2 y BSL3 en conjunto cumplen una función esencial en Arabidopsis. A) Comparación entre plantas de 3 semanas de edad *bsl1-2^{-/-} bsl2-2^{+/-} bsl3-1^{-/-}* (izquierda, indicadas como *bsl1 bsl2/+ bsl3*) y plantas homocigotas *bsl1-2* (derecha, indicadas como *bsl1*). Barra=5 mm. B) Fenotipo de plantas *bsl1-2^{-/-}* homocigotas y plantas *bsl1-2^{-/-} bsl2-2^{+/-} bsl3-1^{-/-}* de 30 días de edad. Barra=3 cm. Tomado de (Maselli et al., 2014).

Sin embargo, en la progenie de plantas *bsl1-1^{-/-} bsl2-2^{-/-} bsl3-2^{+/-}* o *bsl1-2^{-/-} bsl2-2^{+/-} bsl3-1^{-/-}* autofecundadas la relación entre individuos que llevan al menos una copia de T-DNA y aquellas que no llevan T-DNA, en el gen que en el parental se encontraba en heterocigosis fue cercana a 1:1, en vez del esperado 3:1, y no se pudieron recuperar mutantes triples (Tabla 1). No se observó letalidad a nivel de plántula, sino que las silicuas de plantas *bsl1-2^{-/-} bsl2-2^{+/-} bsl3-1^{-/-}* autofecundadas presentan faltantes de semillas desarrolladas (Figura 1.13).

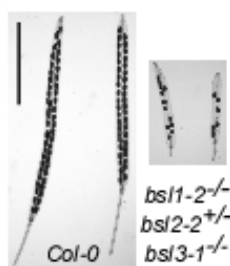


Tabla 1

<i>bsl1-1^{-/-} bsl2-2^{-/-} bsl3-2^{+/-}</i>
54 <i>bsl3-2</i> /106
<i>bsl1-2^{-/-} bsl2-2^{+/-} bsl3-1^{-/-}</i>
88 <i>bsl2-2</i> /168

Figura 1.13 y Tabla 1 - Alteración de frecuencias alélicas en la progenie segregante de plantas con pérdida de función en BSL1, BSL2 y BSL3. Tabla 1. La progenie de los genotipos indicados portando el alelo mutante segregante se desvió significativamente de la relación 3:1 esperada. Las frecuencias observadas son las siguientes: 51% para *bsl1-1^{-/-} bsl2-2^{-/-} bsl3-2^{+/-}* y 52% para *bsl1-2^{-/-} bsl2-2^{+/-} bsl3-1^{-/-}*. Figura 1.13. Silicuas clareadas de los genotipos indicados. Barra=5 mm. Tomado de (Maselli et al., 2014).

Aún cuando los 3 genes se localizan en cromosomas diferentes, se observaron desvíos significativos de una segregación Mendeliana típica en la progenie de plantas *bsl1-2^{+/-} bsl2-2^{+/-} bsl3-1^{-/-}* (Tabla 2). La segregación distorsionada podía deberse a la inviabilidad de una de las gametas o a letalidad embrionaria temprana. Tanto polen como óvulos que portaban los

3 alelos mutantes fueron capaces de producir progenie en cruzas recíprocas entre *bsl2-2^{+/-}* *bsl3-1^{-/-}* *bsl1-2^{-/-}* y plantas silvestres. Sin embargo, mientras que la transmisión del alelo *bsl2-2* vía polen no fue significativamente diferente de la esperada para el caso de viabilidad completa de las gametas masculinas, se recuperó sólo un tercio de los descendientes esperados de óvulos que llevaban el alelo *bsl2-2*, lo que sugiere que los gametofitos femeninos sufren cierto grado de letalidad, aunque con penetrancia incompleta. Corrigiendo las proporciones de óvulos viables que llevaban las diferentes combinaciones alélicas por la distorsión en la proporción de la progenie de plantas *bsl2 bsl3* comentada más arriba (Tabla 1), se logró un ajuste razonable a las frecuencias genotípicas observadas (Tabla 2).

Tabla 2

<i>bsl1-2^{+/-}</i> <i>bsl2-2^{+/-}</i> <i>bsl3-1^{-/-}</i>	F ^{obs}	F ^{exp}	F ^{exp corr}
<i>BSL1 BSL2</i>	0.716	0.5625	0.636
<i>bsl1 BSL2</i>	0.171	0.1875	0.194
<i>BSL1 bsl2</i>	0.112	0.1875	0.141
<i>bsl1 bsl2</i>	0	0.0625	0.029
n=134		$\chi^2_3=18.2$ P<0.001	$\chi^2_3=6.4$ P>0.05

Tabla 2 – Incidencia de la viabilidad gamética en la frecuencia alélica. La progenie de los genotipos indicados

se apartó significativamente de la proporción esperada 9:3:3:1 para dos loci no ligados y con gametas viables (F^{obs} vs F^{exp}). Cuando la proporción de gametas femeninas se corrige de la forma 0.33 (*BSL1 BSL2*): 0.33 (*bsl1 BSL2*): 0.225 (*BSL1 bsl2*): 0.115 (*bsl1 bsl2*), mientras que la proporción de gametas masculinas se mantiene en 0.25 para cada genotipo, las distribuciones de frecuencias obtenidas no resultan significativamente diferentes. *bsl1* y *bsl2* indican la presencia de las inserciones de T-DNA. Los datos fueron comparados con un test de Chi cuadrado. Tomado de (Maselli et al., 2014).

La ausencia completa de mutantes triples aparenta ser, entonces, el resultado combinado del funcionamiento parcialmente defectuoso del gametofito femenino como de letalidad embrionaria temprana. Estos resultados muestran que, en *Arabidopsis*, los homólogos conservados *BSL1* y *BSL2/BSL3*, más allá de sus diferencias en patrones de expresión y de historia evolutiva, cumplen en conjunto un papel esencial. Al mismo tiempo, y por oposición, indican que *BSU1*, que está presente en su forma normal en estas plantas, no es capaz de suplir la función de sus homólogos conservados.

Conclusiones

Utilizando evidencia filogenética, funcional y genética, en esta sección se reevaluaron las funciones de los genes *BSL* en *Arabidopsis*. Se puso especial énfasis en los miembros de la familia más conservados y extendidos, *BSL1*, *BSL2* y *BSL3*, que curiosamente, por razones históricas, han sido los menos estudiados. Estos tres genes están altamente conservados en todas las plantas terrestres; por el contrario, los genes de tipo *BSU1* sólo se encuentran en la familia Brassicaceae y poseen secuencias divergentes, lo que los diferencia entre los restantes miembros de esta familia muy conservativa.

Se encontraron correspondencias robustas entre la historia evolutiva, los patrones de expresión y localización subcelular y las funciones de los distintos homólogos. Los parálogos más cercanos, *BSL2* y *BSL3*, que además serían los más próximos a las formas ancestrales, mostraron comportamientos equivalentes, lo que se corrobora al comprobar que sólo la pérdida de función simultánea de ambos causa un peculiar conjunto de alteraciones fenotípicas. *BSL1*, por su parte, muestra patrones de expresión y localización diferentes a los de *BSL2* y *BSL3*, coincidente con el hecho de que *BSL1* sería una forma derivada presente sólo en plantas vasculares. Tomando una perspectiva histórica, se puede postular que en el origen de las plantas vasculares se produjo una duplicación de un gen *BSL* ancestral que dio origen a genes de tipo *BSL2/BSL3*, que se mantuvieron más cercanos a las características ancestrales, y los de tipo *BSL1*, más divergentes y para el que no se observaron correlatos fenotípicos evidentes por la pérdida de su función en *Arabidopsis*. Como consecuencia de esta duplicación se habrían producido divergencias en expresión, localización y funciones, en un proceso conocido como subfuncionalización, por el cual las funciones del gen ancestral se reparten entre los derivados (Duarte et al., 2006). Dentro de los dos clados principales se pueden distinguir ciertas tendencias. Se acepta que la mayoría de los linajes de angiospermas han sufrido uno o más eventos de multiplicación genómica después de un supuesto evento ancestral de duplicación hace unos 190 millones de años (Jiao et al., 2011). Exceptuando el caso de poliploides relativamente recientes, como soja, o de evolución lenta, como álamo, la rama *BSL1* parece sufrir una pérdida rápida de parálogos, de forma de mantener sólo una copia por genoma. Por el contrario, en la rama *BSL2/BSL3* tienden a encontrarse dos, o múltiplos de dos, parálogos. Estas duplicaciones no aparecen como una

característica ancestral, porque derivan de eventos de duplicación independientes. Esta tendencia se ve más claramente en las familias más conocidas de plantas con flores, es decir, las Poaceae, las Fabaceae y las Brassicaceae. En cada uno de estos grupos, la duplicación de genes de tipo *BSL2/BSL3* aparece como ancestral para cada familia. Las razones que llevan a la pérdida o retención preferencial de genes después de duplicaciones genómicas es un área de intenso análisis en estudios evolutivos. En *Arabidopsis*, se ha sugerido que la conservación de genes duplicados y funcionalmente solapados correlaciona con la severidad de los efectos fenotípicos que produce su pérdida, posiblemente como forma de mantener un buffer mutacional (Hanada et al., 2009). Este patrón podría aplicarse a la familia BSL. En el caso de *BSL2/BSL3*, la recurrencia de conservación de parálogos, los patrones de expresión solapados y el fuerte efecto fenotípico de la pérdida de función de ambos genes en *Arabidopsis* apoyan la idea de una función más importante o extendida que la de *BSL1*, gen que se encuentra en general en copia única por genoma y cuya pérdida de función no produce efectos fenotípicos evidentes, al menos en presencia de otros homólogos funcionales. De todas formas, sería necesario comprobar que estas mismas consideraciones son válidas en otras especies.

Aunque los genes de tipo *BSL2/BSL3* parecen haber permanecido más próximos en secuencia a las formas ancestrales, necesariamente deben haberse modificado en su funcionamiento en el contexto de la presencia de genes de tipo *BSL1*. Esta hipótesis se ve reforzada al encontrar que, temprano en el desarrollo, los tres miembros conservados de la familia en *Arabidopsis* juegan en conjunto un papel esencial. Recientemente se ha comprobado que, en efecto, en organismos que portan una sola copia de genes PPKL, éstos cumplen un papel relevante. En el apicomplejo *Plasmodium*, agente causal de la malaria, el único gen de tipo PPKL se expresa preferentemente en los estadios que tienen lugar en el mosquito, en particular en los gametocitos femeninos (los precursores de las gametas femeninas que se diferencian al entrar en el mosquito) y en las ooquinetas, los cigotos diploides móviles que invaden el intestino del mosquito para producir, por intermedio de una fase inmóvil donde se produce la meiosis, las formas haploides móviles que pasan al hospedador mamífero. Si bien la eliminación del gen no produce letalidad en las fases de vida haploides, causa severas alteraciones en el desarrollo y motilidad de las ooquinetas por alteraciones en el citoesqueleto, que impiden la colonización del intestino del mosquito

(Guttery et al., 2012; Philip et al., 2012). En una búsqueda de mutaciones que alteraran el ciclo celular en la clorofita unicelular *Chlamydomonas reinhardtii* se identificó una mutación en el único gen de tipo BSL en este organismo. Esta mutante, *div44-1*, tiene defectos en la replicación del DNA y en la formación de husos mitóticos, probablemente por alteraciones en la dinámica del citoesqueleto (Tulin and Cross, 2014). Una conexión entre genes *BSL* y el ciclo celular se observó recientemente en arroz. El gen de tipo *BSL2* en el que se identificó el QTL *GL3.1* interacciona con la ciclina T1;3: diferencias en la activación de esta ciclina por parte de OsBSL2 explicarían la variación en la longitud de los granos (Qi et al., 2012). En conjunto, estas evidencias indican que las fosfatasa PPKL participan en la regulación de la dinámica del citoesqueleto y/o (ambos procesos están muy relacionados) de la progresión del ciclo celular.

Precisar los sustratos sobre los que actúan las fosfatasa de Ser/Thr se ve usualmente complicado por varios factores. Por un lado, cumplen numerosas funciones en las células, por lo que la pérdida de su función conduce a perturbaciones severas. Por otro lado, en plantas, los principales grupos de PPP están formados por familias génicas de varios miembros, muchos de ellos con funciones posiblemente solapadas. Por ejemplo, el genoma de *Arabidopsis* codifica 9 isoformas de PP1; no es sorprendente que no se hayan descrito hasta el momento mutaciones en esta subfamilia, pero en cambio la pérdida de función de un regulador común a todas las PP1, Inhibitor-3, de copia única en el genoma de *Arabidopsis*, produce letalidad embrionaria (Takemiya et al., 2009). Mutaciones de pérdida de función de cada uno de los genes de la familia *BSL* en *Arabidopsis* no causan alteraciones visibles, pero cuando se combinan sí. Esta puede ser la causa por la que nunca se han encontrado mutaciones en estos genes en búsquedas de mutantes.

Todos los estudios conducentes al modelo que incluye a las BSL en la señalización por BRs se basan en la sobreexpresión de *BSU1* (Mora-Garcia et al., 2004; Ryu et al., 2007; Kim et al., 2009; Ryu et al., 2010; Kim et al., 2011). La pista original provino de una búsqueda de supresores del fenotipo de una mutante débil en el receptor BRI1, mediante la técnica de *activation tagging*. Como resultado, se obtuvo una línea, *bsu1-D*, que sobreexpresaba *BSU1* a niveles mucho más altos que los encontrados en plantas silvestres, con un fenotipo similar a plantas que sobreproducen BRs o sobreexpresan el receptor, y con un enriquecimiento en las formas defosforiladas de BES1. En un primer momento se postuló que *BSU1* podría

defosforilar directamente a BES1 contrarrestando la acción de BIN2 (Mora-Garcia et al., 2004; Ryu et al., 2007). Posteriormente se postuló en cambio que BSU1, activada por RLCKs que responden a la percepción de BRs, inactivaría a BIN2 (Kim et al., 2009; Kim et al., 2011). Extendiendo esta función a los restantes genes BSL, se postuló que el fenotipo extremadamente enano de la línea llamada *bsu-q* (que combina mutaciones en *BSL1* y *BSU1* con la reducción simultánea en los niveles de expresión de *BSL2* y *BSL3* por interferencia de RNA) resultaba de la insensibilidad de estas plantas a BRs, aunque tal respuesta fue medida sólo de forma indirecta a través de la expresión de ciertos genes (Kim et al., 2009). Los resultados presentados aquí y otros obtenidos en el laboratorio no permiten sin embargo establecer de forma definitiva el papel de los genes BSL en la vía de señalización por BRs. Por un lado, el fenotipo de las mutantes múltiples en estos genes no se corresponden con el típico síndrome de deficiencia en la respuesta a estas hormonas; de hecho, las respuestas a BRs en las mutantes *bs/2 bs/3* no se encuentran afectadas, aunque la distribución de formas fosforiladas de BES1 está alterada. Esta sería la primera vez en que modificaciones en el estado de este factor de transcripción no se reflejan en el fenotipo. A la luz de los resultados presentados, se puede interpretar el fenotipo de la línea *bsu-q* como el resultado de una disminución de la función de los tres genes *BSL* principales con efectos fenotípicos pleiotrópicos. Por otro lado, *BSL1*, *BSL2* y *BSL3* (pero no *BSU1*, en el ensayo de Y2H al menos) interactúan, con aparente distinta afinidad, con miembros de la familia BSK. Sin embargo, las proteínas BSK no participarían sólo en la señalización por BRs. Miembros de esta familia han sido identificados como participantes en la percepción de varios patógenos; en particular, *BSK1* es capaz de interactuar con el receptor de flagelina asociado a membranas, *FLS2* (Shi et al., 2013). Este hallazgo no es sorprendente, ya que resulta recurrente hallar elementos compartidos en las vías de señalización en plantas dependientes de la percepción de estímulos extracelulares. De esta forma, es posible que las proteínas BSL participen de forma genérica de la señalización asociada a múltiples receptores de membrana y/o que participen en otra vía que de alguna forma afecta a la vía de BRs, sin por eso actuar de forma directa sobre ella.

En cuanto a *BSU1*, sus niveles de expresión son tan bajos en la mayoría de los tejidos que es prácticamente imposible detectar su expresión mediante amplificación, como así tampoco es detectable en experimentos de microarreglos, o por secuenciación masiva de transcritos

(RNAseq). Notablemente, las Brassicaceas constituyen una anomalía entre las angiospermas por la presencia del grupo de genes de tipo *BSU1*. Estos genes surgieron en esta familia por duplicación genómica como parálogos de *BSL1*. Dado que, como comentamos, los genes de tipo *BSL1* tienden a preservarse en forma de copia única, los genes de tipo *BSU1* pueden haber evitado ser eliminados a través de un proceso de divergencia funcional conocido como neofuncionalización, por el cual uno de los duplicados emprende un camino evolutivo divergente adquiriendo nuevas funciones. De hecho, es evidente que las secuencias de tipo *BSU1* han estado sujetas a un régimen evolutivo mucho más relajado que sus homólogos altamente conservados. Analizando sólo la pareja *BSL1-BSU1* en *Arabidopsis*, Liu *et al.* (Liu and Adams, 2010) sugirieron que *BSU1* debió en efecto haber sufrido un proceso de neofuncionalización, acoplado a un proceso de subfuncionalización que restringió su expresión al polen y excluyó a *BSL1* del gametofito masculino. Los resultados mostrados aquí, relevando varias especies de Brassicaceas, sugieren en cambio que, si el patrón de expresión diferencial de los genes *BSL* en polen es debido a un proceso de subfuncionalización, éste debería aplicarse a los clados mayores, *BSL2/BSL3* y *BSL1*, que difieren claramente en su expresión en polen (Figura 1.8). La significación funcional de este patrón, sin embargo, es oscura, porque la pérdida de expresión de *BSL2/BSL3*, los parálogos de mayores niveles de expresión en polen, no parece afectar la viabilidad del polen. Una posible explicación sería que *BSU1*, cuya expresión en polen sería una característica secundaria y específica de las Brassicaceas, podría aportar esta función en polen, pero no en los tejidos vegetativos. Pero este razonamiento entraría en contradicción con la fuerte divergencia en secuencia y el argumento de posible neofuncionalización: si las proteínas de tipo *BSU1* cumplen en polen la misma función que las otras *BSLs*, ¿por qué son sus secuencias tan divergentes?

Los genes de tipo *BSU1* muestran, al analizarlos en detalle, varias características inusuales que no se observan en el resto de los genes *BSL*. En *AtBSU1* coexisten dos formas de mRNA por splicing alternativo: una de ellas retiene el 13^{er} intrón, que introduce un codón terminador al final del dominio Kelch. Los intentos en el laboratorio por detectar una proteína truncada o asignar una función al dominio Kelch aislado no han dado resultados claros. En la secuencia genómica reportada de *Capsella rubella*, falta un nucleótido en el exón 9, lo que fuerza un splicing anómalo en la forma anotada automáticamente para

mantener el marco de lectura. En nuestros análisis introdujimos el nucleótido faltante para restaurar el patrón de splicing conservado y la secuencia completa, pero se debería comprobar si esta mutación es real o un error de secuenciación. En la secuencia de *Schrenkiella parvula*, la región entre los dominios Kelch y catalítico es rica en secuencias repetitivas, lo que hace difícil determinar un posible patrón de splicing. El que decidimos postular para el análisis filogenético es similar al encontrado en *Brassica rapa*, dado que ambas especies pertenecen al mismo subgrupo dentro de las Brassicaceas. Notablemente, en los restantes genes *BSL* de Brassicaceas no encontramos anomalías semejantes, lo que sugiere que no se trataría de errores generales de las secuencias reportadas, sino que serían característicos de los genes de tipo *BSU1*. Se ha propuesto que, en *Arabidopsis*, genes con niveles muy bajos de expresión pueden estar sufriendo un proceso de pseudogenización (Yang et al., 2011). En el caso de los genes de tipo *BSU1*, sería pertinente plantearse si, dado el régimen selectivo relajado en que parecen encontrarse, están originando una nueva función, o bien se encuentran en proceso de transformarse en pseudogenes.

Los resultados aquí mostrados sugieren que realizar inferencias mecanísticas basadas en el miembro divergente y filogenéticamente restringido de una familia de proteínas puede ser engañoso, sobre todo en una familia con una importante conservación de secuencia y estructura. Los resultados presentados en el siguiente capítulo apoyan esta hipótesis.

Capítulo II

Análisis de las interacciones entre las fosfatasas PPKL de *Arabidopsis thaliana*

Introducción

Relación entre estructura y regulación en las fosfatasas PPP

Una característica saliente de las fosfatasas del grupo PPP, deducida del análisis de numerosas estructuras cristalinas, es que el plegamiento global del dominio catalítico está muy conservado, es muy resistente a cambios estructurales y no parece sujeto a regulación alostérica (Peti et al., 2013). En el caso de PP1, PP2, PP4 y PP6 las interacciones con otras proteínas se dan sobre el dominio catalítico mismo; en otras PPP, al dominio catalítico se agregan extensiones N- o C-terminales con posibles o comprobadas funciones regulatorias o de interacción. Una característica saliente del dominio catalítico de las fosfatasas PPP es la presencia de surcos poco profundos en su superficie a través de los cuales se da la interacción con otras proteínas. El centro activo se sitúa en el punto de convergencia de tres surcos que conforman una Y sobre la cara activa; cada brazo de la Y tiene características propias y se denominan, a partir de los caracterizados en PP1, surco hidrofóbico, ácido y C-terminal (Bollen et al., 2010). El dominio catalítico puede entonces considerarse una plataforma con múltiples zonas de interacción sobre la que se asocian otras proteínas, sin imponer interacciones estrechas de tipo llave-cerradura. Esta aparente labilidad sería de hecho la base de su mecanismo regulatorio, al permitir el intercambio de subunidades y sustratos.

En la figura 2.1 se muestran algunas de estas características, ejemplificadas sobre la estructura cristalina de PP1 humana.

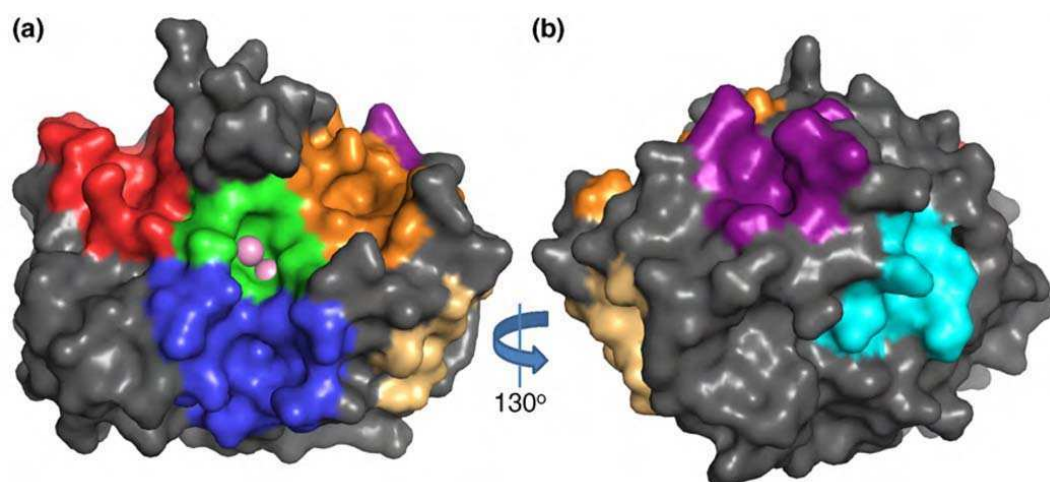


Figura 2.1 (página anterior) - **Aspectos estructurales de las fosfatasa de tipo PP1.** Representación del dominio catalítico de PP1 humana (PDB 3E7B). A la izquierda (a) la vista enfrenta la cara del centro activo, a la derecha (b) la cara opuesta. En (a) se indican los átomos metálicos en el centro activo (en verde) y los surcos superficiales que convergen sobre el sitio activo: en rojo el surco C-terminal, en azul el surco hidrofóbico y en naranja el surco ácido. En (b) se indica el surco RVxF de interacción con las subunidades regulatorias en violeta. Tomada de (Bollen et al., 2010).

En PP1, la unidad funcional es un heterodímero entre una subunidad catalítica y una variedad creciente de proteínas regulatorias. Por diversas técnicas, tanto comprobadas experimentalmente como por predicción o comprobadas experimentalmente, se han identificado al menos 200 posibles subunidades regulatorias en levaduras y animales (Meiselbach et al., 2006; Bollen et al., 2010). El análisis de numerosas subunidades regulatorias reveló la presencia de una secuencia consenso que se denominó el motivo RVxF (en su forma extendida [K/R][K/R][V/I]x[F/W]) (Bollen et al., 2010)), por el cual estas proteínas se asocian a PP1 a través de un surco superficial opuesto al centro activo, interaccionando por sus residuos básicos N-terminales con la región ácida que rodea al surco y por interacciones hidrofóbicas en su extremo C-terminal (Gibbons et al., 2005) (Figura 2.1 b). El motivo RVxF no es el único sitio de interacción de las subunidades regulatorias, sino sólo el que la mayoría de ellas comparten. De hecho, las subunidades regulatorias interaccionan de forma más extensa y a través de múltiples sitios sobre la superficie del dominio catalítico.

PP2 es la más abundante de las PPP en células humanas, representando hasta el 4% de la proteína total en ciertos tipos celulares. La forma activa mínima de PP2 es un heterotrímero formado por un complejo catalítico común y una subunidad regulatoria variable (Shi, 2009). El complejo catalítico común está formado, a su vez, por la subunidad catalítica C asociada con la subunidad de andamiaje A, encargada de poner en contacto las subunidades regulatorias con la catalítica (Xu et al., 2006). La subunidad C es el blanco principal de numerosas toxinas secretadas por organismos marinos tales como ácido okadaico, caliculina, microcistinas y nodularinas, que se usan *in vitro* para el estudio de las propiedades de estas proteínas. Estas toxinas se fijan a la proteína alojándose en el surco hidrofóbico y bloqueando el centro activo. La subunidad A está formada por 15 repeticiones de tipo HEAT (Huntingtin-Elongation-A subunit-TOR), un motivo de dos hélices α antiparalelas, que en conjunto conforman una estructura de tipo herradura de gran flexibilidad. La subunidad C interacciona con las repeticiones HEAT C-terminales a través de una zona de la cara opuesta

al centro activo cuya secuencia es particularmente divergente respecto de otras PPP (Xu et al., 2006). El complejo PP2A-C recluta a las subunidades regulatorias B, que imparten especificidad al mediar el reconocimiento de los sustratos. Hay cuatro familias de subunidades B en el genoma humano (y en el de *Arabidopsis*, aunque en este caso están menos caracterizadas), denominadas B (B55/PR55), B' (B56/PR61), B'' (PR72/PR70) y B''' (PR93/Striatin), con varios miembros en cada familia. Las subunidades regulatorias interaccionan con las repeticiones HEAT N-terminales de la subunidad A, de forma que quedan enfrentadas a la subunidad catalítica. En algunos casos, como en la subfamilia B', también se producen contactos entre las subunidades B y C, mientras que en otros, como las subunidades de la subfamilia B, no establecen casi contactos con la subunidad C (Xu et al., 2008). En ambos casos las subunidades B aportan superficies de interacción con sustratos en las cercanías del centro activo.

En PP5, los dominios catalítico y regulatorio están incluidos en el mismo polipéptido. Estas proteínas tienen una extensión N-terminal de repeticiones de tipo tetratricopéptido (TPR), que actúa como inhibidor de la actividad al bloquear el acceso al sitio activo. En animales, PP5 tiene un papel importante en la señalización por hormonas esteroideas y en respuestas a estrés, al interaccionar directamente, a través del dominio TPR, con el complejo entre Hsp90 y el receptor de glucocorticoides (Chen et al., 1996). La interacción del dominio TPR con Hsp90 o con ácidos grasos largos no saturados, como araquidónico, libera la autoinhibición (Chen and Cohen, 1997; Ramsey and Chinkers, 2002). Se desconoce si el dominio catalítico interacciona con otras proteínas regulatorias. Tanto el dominio TPR como la extensión C-terminal del dominio catalítico son necesarios para mantener el estado de autoinhibición. La remoción de cualquiera de ellos produce una proteína constitutivamente activa. La zona C-terminal interacciona con la cara opuesta del dominio TPR para estabilizar su unión con el dominio catalítico, y por esta razón pudo determinarse su estructura (Yang et al., 2005).

Sobre las restantes PPP no hay mayores datos estructurales.

En resumen, las fosfatasa PPP están formadas por un dominio catalítico muy conservado y estructuralmente compacto. Este dominio catalítico es capaz de interaccionar con otras proteínas a través de varias zonas de su superficie y puede tener extensiones N- o C-

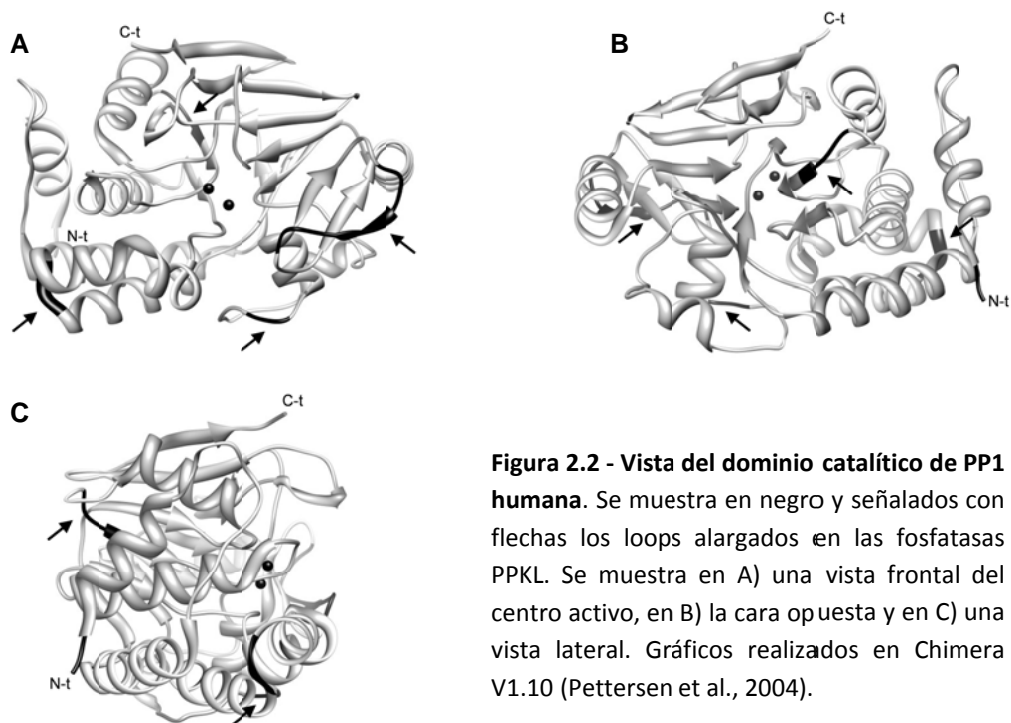
terminales con diversas funciones. Sin embargo, no existe registro sobre la formación de interacciones homotípicas en esta familia de fosfatasas de proteínas.

Dominios en las Fosfatasas PPKL

Tanto en plantas como en alveolados las fosfatasas PPKL están compuestas por tres dominios. El dominio N-terminal es de tipo turbina- β con repeticiones de tipo Kelch. Los dominios turbina- β son módulos estructurales presentes en numerosas proteínas. Existen dos superfamilias de turbinas- β : la de repeticiones WD40 y la de repeticiones Kelch. Aunque con toda probabilidad surgieron evolutivamente de forma independiente, guardan muchas similitudes en estructura y en funciones. En ambas, el dominio consiste en unidades repetitivas (en general 6 o 7) formadas por una lámina β antiparalela de cuatro hebras, unidas por regiones conectoras dispuestas alrededor de un eje central, adoptando una forma toroidal (Li et al., 2004). Su plegamiento se completa “mordiéndose la cola”: la lámina β que cierra el toro está formada por tres hebras de un extremo del dominio y una hebra del extremo opuesto. La torsión de las láminas β hace que la estructura se asemeje a la sección de un cono, con una cara (por convención, la superior) más estrecha que la otra. De estas caras protruyen las regiones conectoras entre las hebras β , constituyendo zonas de interacción con otras proteínas o, en algunos casos, donde se aloja el centro activo. Mientras que en las proteínas con dominios WD40 ambas caras suelen ser variables y participar en interacciones, en las Kelch sólo la cara superior tiende a ser variable (Hudson and Cooley, 2008). En el dominio Kelch de las proteínas BSL, que consta de 6 unidades repetitivas, se verifica esta regularidad.

El dominio catalítico de las proteínas PPKL está relacionado en secuencia con PP1, pero tiene varias peculiaridades que lo diferencian. Por un lado, presenta cambios conservados en ciertas posiciones diagnósticas. Una especialmente relevante es el reemplazo de una Pro, presente en todas las restantes PPP, por una Thr en el loop denominado β 12- β 13, que se encuentra entre los surcos C-terminal y ácido de la cara del centro activo y que en PP1 y PP2 participa en las interacciones con ciertas moléculas regulatorias. En el extremo C-terminal del dominio catalítico, por su parte, se encuentra una extensión sin mayor homología con otras PPP, con una región conservada rica en Pro y Arg.

Otra característica saliente del dominio catalítico de PPKL es la presencia de extensiones en la superficie, señaladas en la Figura 2.2 (Kutuzov and Andreeva, 2002). La más extensa se localiza en la cara opuesta a la activa, en la zona donde se concentran los residuos que median el contacto entre las subunidades C y A de PP2; posiblemente esta diferencia contribuye a segregar ambos tipos de fosfatasa. Otra inserción aparece en la conexión entre dos hélices contiguas, y las otras dos modifican loops que bordean los surcos hidrofóbico y ácido cerca del centro activo. Las extensiones son siempre más extensas en las proteínas de alveolados que en las de plantas.



En las fosfatasas PPKL se encuentra, finalmente, un tercer dominio que conecta los dominios Kelch y catalítico. Este dominio intermedio, que denominaremos en principio conector (L por Linker), tiene una longitud de entre 130 y 210 aminoácidos y no guarda similitud con otras secuencias conocidas fuera de la familia.

Objetivos

Puesto que las fosfatasa PPP actúan en forma de holoenzimas asociadas con múltiples proteínas que definen su especificidad, propiedades regulatorias y determinan la localización subcelular del dominio catalítico, se llevó a cabo, antes del comienzo de esta tesis, una búsqueda de proteínas asociadas para comprender mejor el funcionamiento de las fosfatasa PPKL en plantas. En particular, se enfrentó el dominio catalítico de BSL2 a una biblioteca de cDNA derivada de meristemas apicales de plántulas de Arabidopsis mediante la estrategia de Y2H. El límite N-terminal del dominio catalítico se definió como la zona a partir de la cual se encuentra homología considerable con otras PPP de plantas o de otros organismos. En esta búsqueda se recuperaron varios clones que, en una primera corroboración preliminar, cumplieron con los requisitos de expresión de genes reporteros para ser considerados interactores válidos. Curiosamente, entre los clones que demostraron interacciones más fuertes se recuperaron varios que codificaban el mismo dominio catalítico de BSL2, asociado con diferentes porciones del dominio conector. También, aunque en menor número, se recuperaron clones correspondientes a BSL1 y BSL3, con similar configuración. En estudios anteriores (Gustavo Maselli, tesis de licenciatura) se validó, usando el ensayo de Y2H, la interacción de BSL2 consigo misma y con otras BSL. A partir de estos datos, se decidió profundizar en la caracterización de la capacidad de interacción de las proteínas BSL consigo mismas.

Resultados

Interacciones homotípicas en las fosfatasa BSL de *Arabidopsis thaliana*.

En la búsqueda por Y2H original y en los estudios previos ya descritos, realizados en el laboratorio, se utilizó la cepa de levaduras MaV203, provista por el sistema ProQuest (Invitrogen). Se observó que la expresión de la forma completa de BSL2 en esta cepa producía perturbaciones en el crecimiento, de forma tal que las colonias que la portaban crecían lentamente y adoptaban una morfología “rugosa” característica. Este efecto era aún más marcado en cultivos líquidos, que crecían muy lentamente y agrumados. Estas particularidades dificultaban la comparación entre diferentes construcciones, en particular en ensayos que dependían de la determinación de la densidad óptica alcanzada por los cultivos como forma de normalizar la cantidad de células usada. Para subsanar estas dificultades, se ensayaron varias cepas diferentes, y se encontró que el crecimiento de las cepas AH109 y Y187, provistas con el sistema Matchmaker (Clontech), era menos sensible a la expresión de las proteínas BSL. Por esa razón, en las secciones siguientes, se utilizarán estas cepas junto con los plásmidos del sistema ProQuest, que hacen uso del sistema de clonado por recombinación Gateway.

Se comenzó verificando la interacción de cada proteína consigo misma, en su forma completa, mediante el sistema de Y2H (Figura 2.3 A). BSL1, BSL2 y BSL3 mostraron interacciones homotípicas. En el caso de BSU1 la interacción no pudo ser evaluada por este método, ya que al expresar la secuencia completa fusionada al dominio de unión al ADN del factor de transcripción Gal4 (BD), la proteína de fusión fue capaz de activar la expresión de los genes reporteros por sí sola (en particular, en la figura 2.3 A y en las siguientes se muestra la capacidad de crecer en ausencia de His).

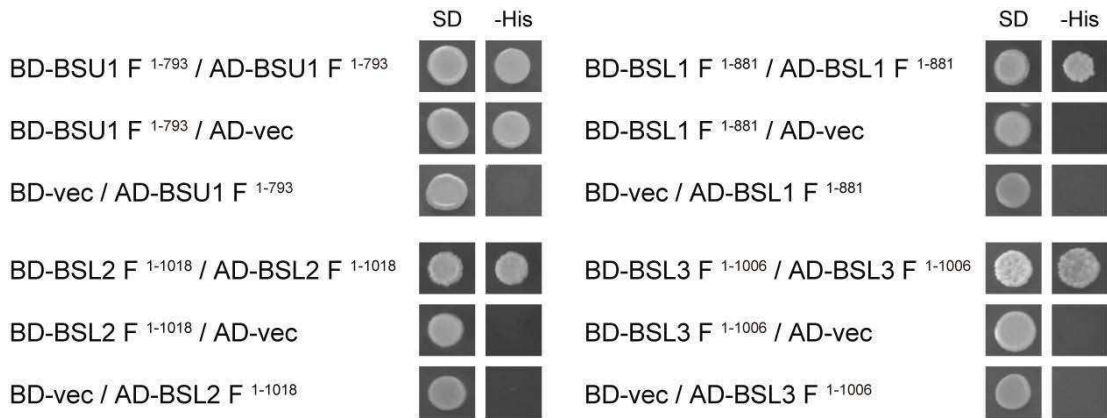
Para validar las interacciones encontradas, se utilizaron métodos independientes. En la Figura 2.3 B se observan los resultados de ensayos de coimmunoprecipitación (coIP) con las versiones completas de las proteínas, fusionadas a GFP y al epítipo de hemaglutinina (HA) y coexpresadas de forma transitoria en hojas de *Nicotiana benthamiana*. Tanto para BSL1, como para BSL2 y BSL3 se verifica que la inmunoprecipitación con anticuerpo anti-GFP arrastra a la forma con etiqueta HA. Significativamente, esto no sucede en el caso de BSU1.

La Figura 2.3 C muestra los resultados de cromatografía de exclusión molecular (CEM) de extractos de hojas de *N.benthamiana* que expresaban de forma transitoria las formas de las proteínas fusionadas a GFP. Nuevamente, se observa que BSL1, BSL2 y BSL3 son excluidas de la columna en un volumen que condice, aproximadamente, con el esperado para formas dimericas o de peso molecular aparente algo mayor, pero no para formas monoméricas. BSU1 mantiene sus propiedades anómalas, dado que eluye en la posición esperada de un monómero, aunque la elución muestra una cola extendida hacia los pesos moleculares aparentes mayores.

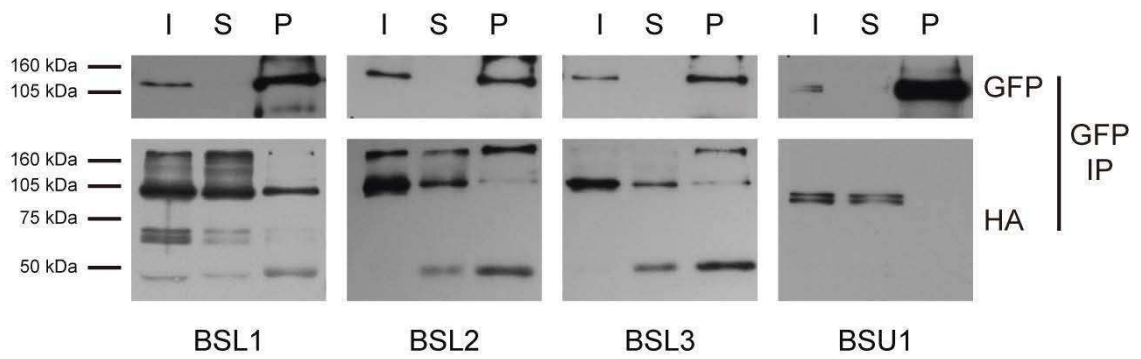
Finalmente, comprobamos la capacidad de interacción mediante complementación bimolecular de la fluorescencia (biFC, por sus siglas en inglés), nuevamente expresando de forma transitoria las proteínas en *N.benthamiana* (Figura 2.3 D). Se observó que, en este caso, las cuatro proteínas lograron reconstituir la señal fluorescente. El comportamiento de BSU1 parece, en este sentido, contrario a los resultados anteriores, aunque en realidad no representa necesariamente una contradicción. El ensayo de biFC es lo suficientemente sensible como para permitir el análisis de interacciones entre proteínas expresadas en niveles comparables a los endógenos. Sin embargo, en condiciones de sobreexpresión puede llegar a sobreestimar la intensidad de una interacción debido a que la reconstitución de la proteína fluorescente y la unión covalente de los aminoácidos que forman el fluoróforo estabiliza el complejo de forma irreversible (Ohad et al., 2007). El ensayo de biFC, por otra parte, no sólo muestra interacción, sino que también brinda información acerca de la localización subcelular de la misma. En este caso, la señal derivada de la interacción coincidió con la localización reportada de las proteínas fusionadas a GFP (ver figura 1.8 y (Maselli et al., 2014)).

Las cuatro técnicas muestran, entonces, que las proteínas BSL son capaces de presentar interacciones homotípicas. La interacción de BSU1 consigo misma, si existe, debe ser mucho más débil que la de sus homólogos, de forma que no resiste el pasaje por la columna de exclusión molecular o las condiciones de coIP y sólo es detectable por una técnica que irreversibiliza la interacción. La diferencia de comportamiento entre BSU1 y sus homólogos coincide, de hecho, con los análisis evolutivos descritos en el capítulo anterior.

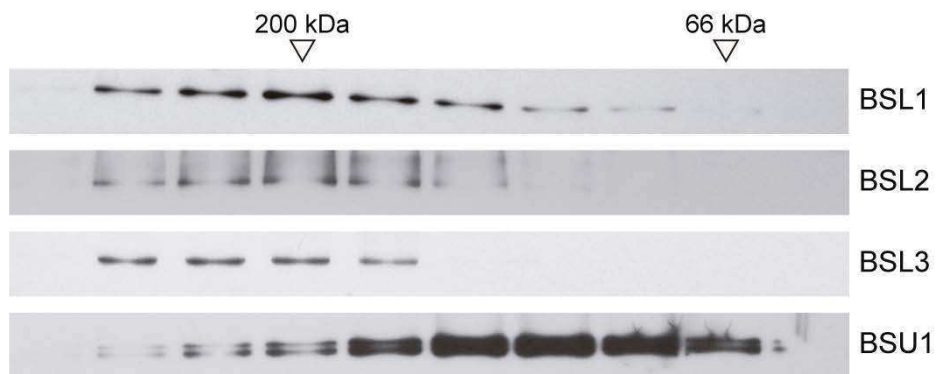
A



B



C



D

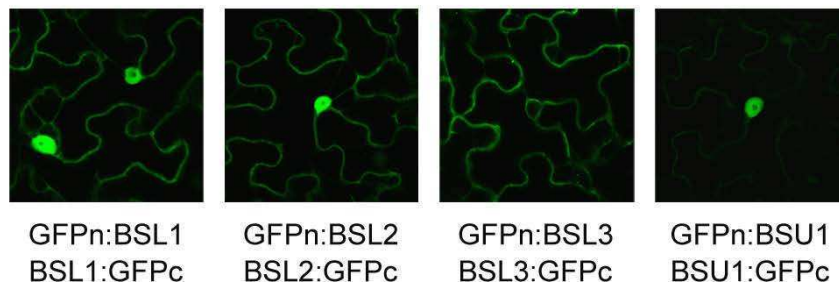


Figura 2.3 (página anterior) - **Interacciones homotípicas de las fosfatasas BSL.** A) Ensayo de Y2H usando como reportero el crecimiento en ausencia de Histidina. Se evaluaron las fusiones de las proteínas indicadas al dominio de activación del factor de transcripción Gal4 (AD) y al dominio de unión al DNA del factor de transcripción Gal4 (BD). Junto al nombre de cada construcción se indica el número de aminoácidos que la conforman. F denota que la construcción abarca la totalidad de la secuencia codificante del gen. B) Coimmunoprecipitación a partir de extractos de hojas de *N.benthamiana* que expresan cada homólogo fusionado a GFP y al epítoto HA. C) CEM de los cuatro homólogos fusionados a GFP. Se indican las fracciones en las que eluyen algunos de los marcadores de peso molecular utilizados. D) Ensayo de biFC de las fusiones de los cuatro homólogos al fragmento N-terminal (GFPn) o C-terminal (GFPc) de GFP. Se muestran imágenes de microscopía confocal de células de epidermis de hojas de *N.benthamiana* que expresaban ambas formas de las proteínas. BSU1 F: secuencia completa de BSU1; BSL1 F: secuencia completa de BSL1; BSL2 F: secuencia completa de BSL2; BSL3 F: secuencia completa de BSL3; SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$. Vec: vector sin inserto. I: Alícuota del extracto sin inmunoprecipitar; S: Alícuota del sobrenadante de la inmunoprecipitación; P: Alícuota del inmunoprecipitado.

Dado que la búsqueda inicial de interactores que dio origen a este proyecto se realizó mediante la técnica de Y2H utilizando al dominio catalítico de BSL2 como “carnada” (o sea, como fusión a BD), se comprobó si todos los homólogos eran capaces de interactuar con su correspondiente dominio catalítico aislado (Figura 2.4). En concordancia con lo mostrado en la figura 2.3, los dominios catalíticos de BSL1 (BSL1 P), BSL2 (BSL2 P) y BSL3 (BSL3 P) interactuaron con su correspondiente proteína completa (F), mientras que el dominio catalítico de BSU1 (BSU1 P) no lo hizo.

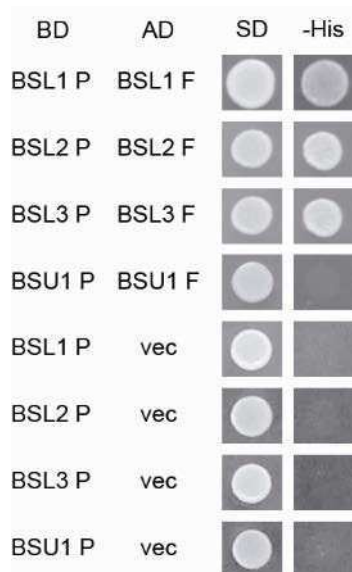


Figura 2.4 - Interacción de los dominios catalíticos de BSL1, BSL2 y BSL3 con sus correspondientes proteínas de longitud completa. Ensayo de Y2H usando como reportero el crecimiento en ausencia de Histidina. Se enfrentaron fusiones entre el dominio de unión al DNA del factor de transcripción Gal4 (BD) y los dominios catalíticos aislados (P) con fusiones del dominio de activación del factor de transcripción Gal4 (AD) y la proteína completa correspondiente (F). SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$. Vec: vector sin inserto.

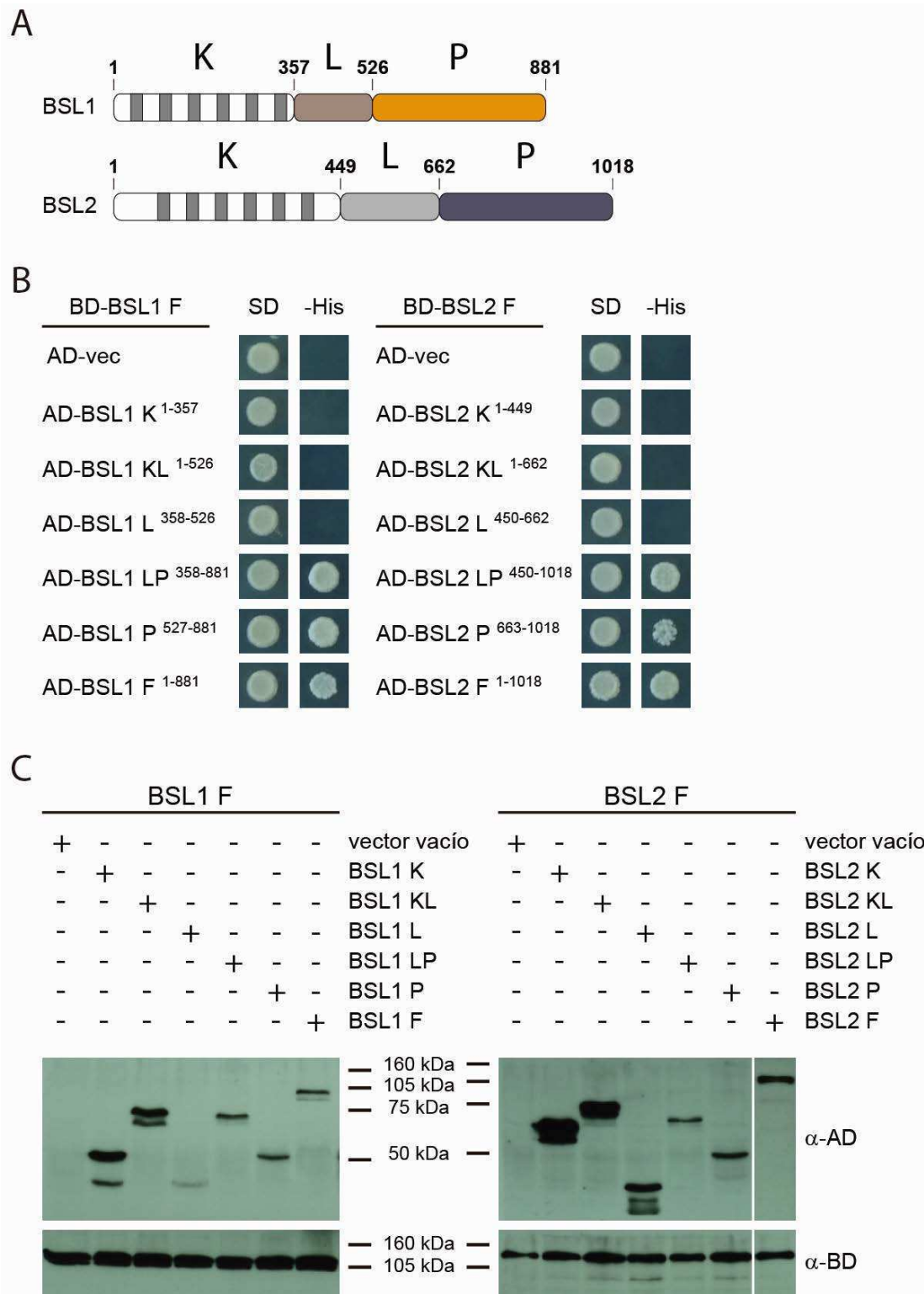
En base a estos resultados, se decidió continuar el análisis con BSL1 y BSL2. BSL2 y BSL3 son parálogos cercanos, presentan una identidad del 90%, parecen cumplir funciones solapadas

y se comportan de forma equivalente en los análisis mostrados hasta el momento. Por su parte, BSU1 no parece ser capaz de formar interacciones homotípicas o lo hace con una afinidad suficientemente baja como para no poder detectarse fielmente por los métodos empleados. Por otra parte, BSL2/BSL3 y BSL1 se encuentran en todas las plantas con flor, mientras que BSU1 está restringida a una sola familia.

BSL1 y BSL2 difieren en sus requerimientos para que la interacción tenga lugar

Una vez elegidas BSL1 y BSL2 como ejemplos de las dos subfamilias de fosfatasa BSL, se quiso investigar cuáles son los dominios involucrados en la interacción homotípica registrada. Para ello se generaron construcciones que expresaban los tres dominios aislados o combinados de a pares con su dominio adyacente, y se enfrentaron a la construcción que expresa la proteína completa en ensayos de Y2H. La interacción se verificó, en ambas proteínas, siempre y cuando estuvieran presentes dos dominios catalíticos (Figura 2.5).

Figura 2.5 (página siguiente) - **Requerimientos mínimos para que se verifique la interacción homotípica en BSL1 y BSL2.** A) Diagrama de BSL1 y BSL2. Se indican los dominios y el número de aminoácidos de cada uno de ellos. B) Ensayo de Y2H usando como reportero el crecimiento en ausencia de Histidina. Se evalúan fusiones al dominio de activación del factor de transcripción Gal4 (AD) de los dominios indicados, frente a fusiones de la correspondiente proteína completa con el dominio de unión al DNA del factor de transcripción Gal4 (BD). Junto al nombre de cada construcción se indican los aminoácidos que la conforman. C) Inmunodetección de AD (panel superior) y BD (panel inferior) en extractos proteicos de levaduras expresando BSL1 o BSL2 en combinación con los dominios aislados. K: Dominio Kelch; L: Dominio conector; P: Dominio catalítico; SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$. Vec: vector sin inserto.



En estos ensayos, el dominio Kelch no formó parte de ninguna combinación que mostrara interacción. Como los dominios turbina- β están asociados con interacciones proteína-proteína, se buscó una evidencia independiente que permitiera descartar la participación del mismo en la interacción. El laboratorio contaba con una línea de Arabidopsis mutante doble *bsl2 bsl3* transformada establemente con la construcción *pBSL2:BSL2-F:HA* (un fragmento de DNA genómico incluyendo las regiones codificante y promotora de BSL2 con una etiqueta HA en su extremo C-terminal), cuya expresión complementaba el fenotipo mutante, y otra línea de Arabidopsis que expresaba establemente la construcción *pBSL2:BSL2-KL:GFP* (el mismo fragmento de DNA genómico ingenierizado de forma de eliminar la región codificante del dominio catalítico, preservando la región 3' no traducida, generando una fusión traduccional a GFP en su extremo C-terminal). Se obtuvieron mediante cruzamiento plantas que expresaban ambas proteínas. Mediante coIP pudo verificarse, tal como se observó en levaduras, que el conjunto de los dominios Kelch y conector no participa, aún en su contexto fisiológico, en la interacción homotípica (Figura 2.6).

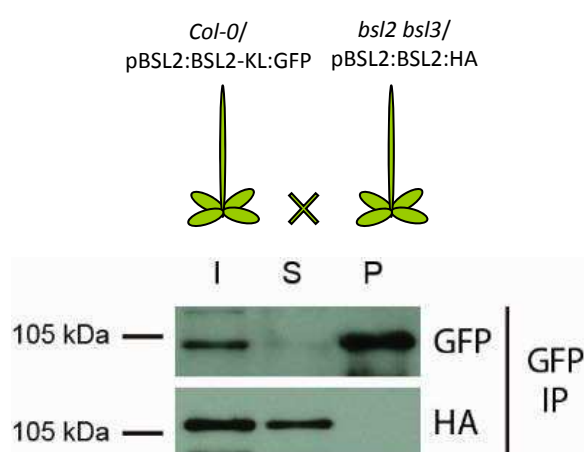


Figura 2.6 - La región N-terminal de BSL2 (dominios Kelch y conector) no participa en la interacción homotípica. Un extracto de plántulas F1 provenientes de la cruce entre una línea de Arabidopsis *bsl2bsl3/pBSL2:BSL2-F:HA* de fenotipo normal y otra línea que expresa establemente la construcción *pBSL2:BSL2-KL:GFP* fue sometido a inmunoprecipitación utilizando anticuerpo anti-GFP. I: Alícuota del extracto sin inmunoprecipitar; S: Alícuota del sobrenadante de la inmunoprecipitación; P: Alícuota del inmunoprecipitado.

A partir de los resultados mostrados en la Figura 2.5, se puede concluir que el dominio P es necesario para que tenga lugar la interacción con la proteína completa. Esta observación motivó la búsqueda de la mínima unidad de interacción. Dado que no se encontraron evidencias de una participación directa del dominio Kelch en la interacción, se analizó el papel del dominio L. Un dato relevante que se tuvo en cuenta en este análisis es que en la búsqueda original de interactores de BSL2, que dio origen a esta investigación, todos los clones de BSL2 recuperados incluían al dominio P y porciones del dominio L.

El ensayo de interacción por Y2H (Figura 2.7 A) mostró que los dominios P de BSL1 son capaces de interactuar consigo mismos, mientras que los de BSL2 no. Sin embargo, la presencia de al menos un dominio L restaura la capacidad de interacción en el caso de BSL2; con BSL1 también se verifica la interacción, pero en este caso el rol del dominio L queda enmascarado por la interacción entre los dos dominios P. La capacidad de los dominios P aislados de interactuar con otro dominio P en presencia o ausencia de su correspondiente dominio L se evaluó también mediante coIP y CEM. Los ensayos de coIP (Figura 2.7 B) corroboraron lo hallado por Y2H: la interacción homotípica de BSL2 requiere la presencia de al menos un dominio L unido a su correspondiente dominio P, mientras que los dominios P de BSL1 aislados son capaces de interactuar. Por otra parte, la CEM (Figura 2.7 C) permitió definir que la interacción de BSL2 probablemente da lugar a la formación de un dímero. En CEM, BSL1 se comportó de forma más compleja. La elución de BSL1 P:GFP correspondió a un tamaño intermedio entre un monómero (69,8 kDa) y un dímero (139,7 kDa). Por otro lado, la forma LP de BSL1 eluyó con un peso molecular aparente mayor que el correspondiente a un dímero. Estos resultados podrían explicarse de varias formas. Por un lado, la afinidad entre los dominios P aislados de BSL1 puede no ser tan fuerte como para mantener la interacción durante el pasaje a través de la columna, produciéndose una disociación parcial. Alternativamente, pueden disociarse durante la cromatografía y estar asociados a algún otro componente de las hojas de *N.benthamiana*. En el caso de BSL1-LP, ésta podría estar formando un complejo mayor que un dímero, o también puede estar interactuando con algún componente que aumenta su tamaño, o bien puede tener un radio hidrodinámico más amplio que lo que su tamaño hace suponer, de forma que eluye como una proteína de mayor tamaño.

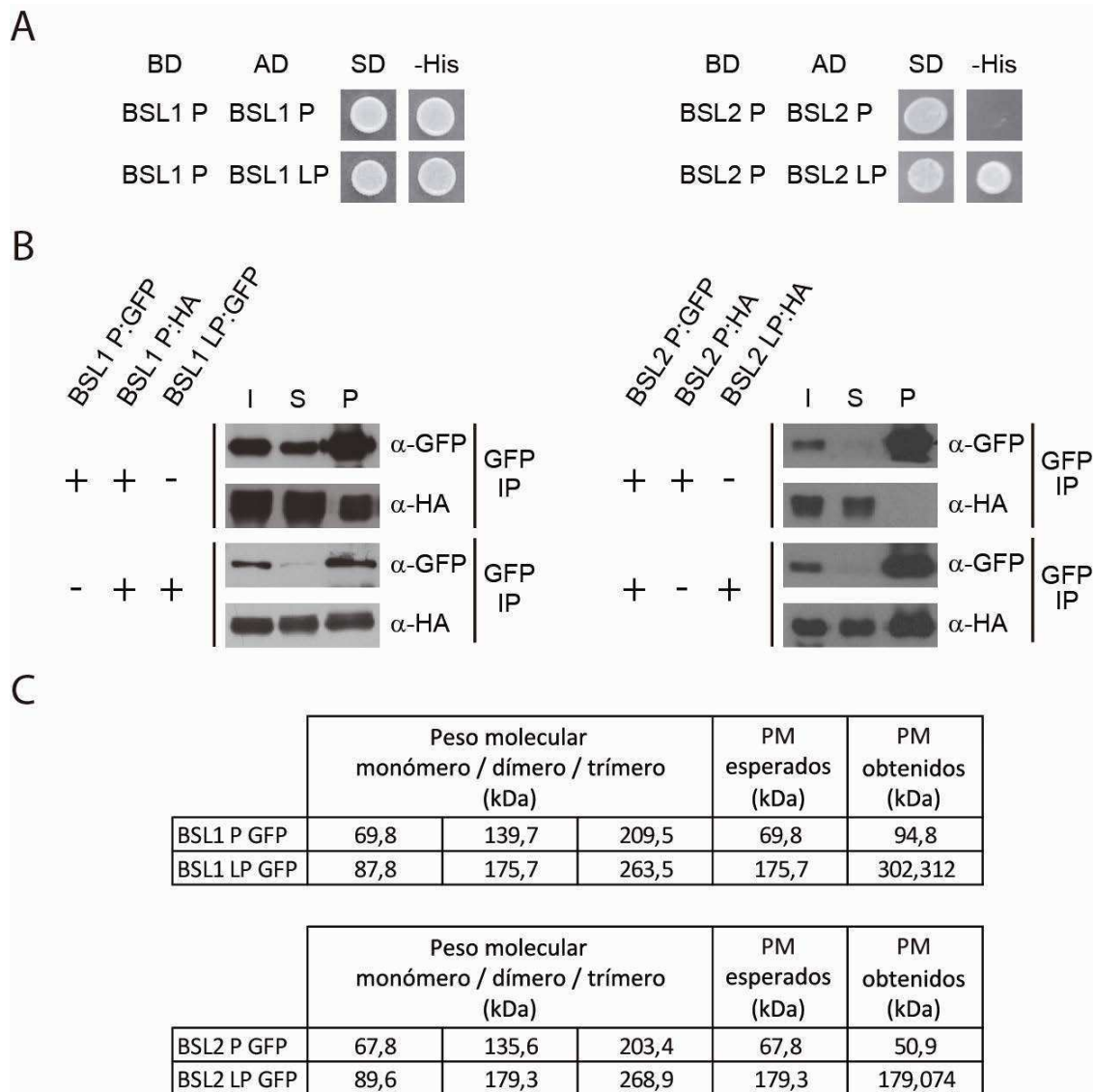


Figura 2.7 - BSL1 y BSL2 difieren en sus requerimientos para interactuar. A) Ensayo de Y2H donde se evalúan fusiones a Gal4-AD de los dominios P en presencia o ausencia de su correspondiente dominio L, enfrentándolas a fusiones entre Gal4-BD y los dominios P indicados. B) Coimmunoprecipitación a partir de extractos de hojas de *N.benthamiana* que expresan las formas indicadas fusionadas a GFP y al epítipo HA. C) CEM de las formas indicadas fusionadas a GFP. Se determinaron las fracciones de elución mediante *Western blot* con anticuerpos anti-GFP. La tabla muestra, en los paneles de la izquierda, los pesos moleculares teóricos para las formas monoméricas, diméricas y triméricas. En los paneles de la derecha se comparan los valores esperados según los resultados anteriores, con los obtenidos. SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$. I: Alícuota del extracto sin inmunoprecipitar; S: Alícuota del sobrenadante de la inmunoprecipitación; P: Alícuota del inmunoprecipitado.

El dominio L entonces, a pesar de lo que su nombre pueda sugerir, parece tener otras funciones más allá de simplemente conectar los dos dominios globulares que conforman a

las fosfatasa BSL. Sin embargo, lo único que se sabe sobre él es que comprende una región de 206 aminoácidos que no reviste homología con ninguna proteína conocida excepto con el mismo dominio en fosfatasa PPKL de otros organismos. Puede afirmarse que el dominio L es imprescindible para que la interacción en BSL2 tenga lugar, sin embargo su estructura o sus determinantes de interacción son incógnitas.

Estos resultados muestran que, a pesar de que las proteínas completas forman interacciones homotípicas, el mecanismo a través del cual lo hacen no parece estar sometido a los mismos requerimientos. Esto abrió dos vías de análisis para intentar contestar las siguientes preguntas:

- ¿Qué regiones del dominio L de BSL2 son necesarias para que tenga lugar la interacción?
- ¿Cuáles son las diferencias entre los dominios P de BSL1 y BSL2 que dan cuenta de su comportamiento?

Características del dominio conector

Para abordar el estudio del dominio L, se realizó una comparación de secuencias entre los dominios de la familia Brassicaceae (Figura 2.8 A). Se encontraron diferencias marcadas entre las distintas subfamilias. Los dominios de mayor longitud se encuentran en la subfamilia BSL2/3. Los dominios L en BSL1 son, comparados con éstos, más cortos, pero presentan alto grado de conservación entre sí. Por el contrario, los dominios L de la subfamilia BSU1 no sólo son aún más cortos sino que muestran poca conservación entre especies cercanamente relacionadas, una característica señalada en el capítulo I. En las dos especies de *Arabidopsis* y en *Capsella rubella* su tamaño es aproximadamente 50% del dominio L de BSL2/3, mientras que en *Brassica rapa* y *Schrenkiella parvula* este dominio se reduce de forma considerable, con un bajo porcentaje de identidad. La Figura 2.8 B refleja estas diferencias. En la misma puede observarse que un anticuerpo policlonal generado a partir del dominio L de BSL2 (ver capítulo III) reconoce de forma equivalente a BSL2 y BSL3, que su afinidad por BSL1 es débil y que no es capaz de reconocer a BSU1.

Como se mostró, BSU1 no interacciona consigo misma en las condiciones ensayadas, o lo hace con muy baja afinidad. En BSL1, la mínima unidad de interacción no incluye al dominio

L mientras que en BSL2 este dominio es completamente necesario para la interacción. Teniendo en cuenta esto, las diferencias en las secuencias aminoacídicas entre los dominios L de las proteínas BSL podrían responder a una diferencia en su función.

A)

BSU	At1g03445	350	-----TSQSSSPEPEEDNPNYMIIDDDY-----	-----MDEPKPLSSEPEASSFTMR-----	-----STSEIAMDRLAE-----	-----AHN-----	
	Allg02750	354	-----TSQSSSPELEEDNPNYMIIDDFI-----	-----MDEPKPLSSEPEAPSFIRR-----	-----SSSESAMERLAE-----	-----AYN-----	
	CrBSU	361	-----HLISPEPEEEDNPNHMYLDDDDP-----	-----ETSNPSFDPPEEVPVSLTIRR-----	-----SSSEGSLEKRTVE-----	-----EYN-----	
	BsBSU	354	-----TSQLSST-EPEDNPNYMIIDDFI-----	-----VAAASKPSSEPEEVPVFMRR-----	-----SSSESSERLAE-----	-----AYN-----	
	Br030478	337	-----EILSSGPTVPSFFDMNDGYKP-----	-----	-----	-----	
	SpBSU	348	-----DDRAQPSSSGPAPSLDDMNYRILP-----	-----	-----	-----	
BSL1	At4g03080	359	--NSTTQSDISSPLLA-SDRTQOSTPRFSVA-----	-----ARPPSGSEPT-SFS-----	-----MSEGLSLDENSLKLTASAAEAEBVASSVWRAAQ-----	-----IGA-GTLDDE-----	-----PSTSDASS-----
	Al6g38200	359	--NSTTQSDISSPLLA-SDRTQOSTPRFSVA-----	-----ARPPSGSEPT-TFA-----	-----MSEGLSLDENSLKLTASAAEAEBVASSVWRAAQ-----	-----IGS-GTLDDE-----	-----PSTSDASS-----
	CrBSL1	359	--NSTTQSDISSPLLA-SDRTQOSTPRFSVA-----	-----ARPPSGSEPT-AIP-----	-----MSEGLSLDENSLKLTASAAEAEBVASSVWRAAQ-----	-----IGS-GTLDDE-----	-----PSTSDANS-----
	BsBSL1	359	--NSTTQSDITSSLLA-SDRTQOSTPRFSVA-----	-----ARPPSGSEPT-AIP-----	-----MSEGLSLDENSLKLTASAAEAEBVASSVWRAAQ-----	-----IGS-GTLDDE-----	-----PSTSDATS-----
	Br018520	360	--NSTTQSDISSPLSASSDRTQOSTPRFSVA-----	-----ARPPSGSEPT-A-----	-----SEGLSLDENSLKLTASAAEAEBVASSVWRAAQ-----	-----SS-ASLDEE-----	-----PSASDGSS-----
	Br000846	359	--NSTTQSDISSPLSA---TQOSTPRFSVA-----	-----ARPPSGSEPT-AIS-----	-----LSEGLSLDENSLKLTASAAEAEBVASSVWRAAQ-----	-----DS-TSLDEE-----	-----PSASDGSS-----
	SpBSL1	359	--NSTTQSDISSPLLA-SDRTQOSTPRFSVA-----	-----ARPPSGSEPT-AIS-----	-----TSEGLSLDENSLKLTASAAEAEBVASSVWRAAQ-----	-----GGSASLDEE-----	-----PSASDGSS-----
BSL2	At1g08420	454	AAETTTAASHAA-AAAATNPPGRFLPGRYGFSDERNRELSAADGAVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	Allg08060	453	AAETTTAASHAA-AAAATNPPGRFLPGRYGFSDERNRELSAADGAVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	CrBSL2	454	AAETTTAASHAA-AAAATNPPGRFLPGRYGFSDERNRELSAADGAVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	BsBSL2	453	AAETTTAASHAA-AAAATNPPGRFLPGRYGFSDERNRELSAADGAVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	Br030724	434	AAETTTAASHAA-AAAATNPPGRFLPGRYGFSDERNRELSAADGAVVLGSPVAPPVNDLHNDISSENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	Br031626	432	AAETTTAASHAA-AAAATNPPGRFLPGRYGFSDERNRELSAADGAVVLGSPVAPPVNDMHTDISSENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	SpBSL2	445	AAETTTAASHAA-AAAATNPPGRFLPGRYGFSDERNRELSAADGAVVLGSPVAPPVNDMHTDISSENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
BSL3	At2g27210	443	AAETTTAASHAAAAAATNPPGRSPGRYGFSDERTGELPESAPD-AVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	Al4g11030	439	AAETTTAASHAAAAAATNPPGRSPGRYGFSDERTGELPESAPD-SVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	CrBSL3	441	AAETTTAASHAAAAAATNPPGRSPGRYGFSDERTGELPESAPD-AVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	BsBSL3	451	AAETTTAASHAAAAAATNPPGRSPGRYGFSDERTGELPESAPD-AVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	Br034332	436	AAETTTAASHAA-AAAATNPPGRSAGRYGFSDERTGELPESAPD-AVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
	SpBSL3	448	AAETTTAASHAAAAAATNPPGRSPGRYGFSDERTGELPESAPD-AVVLGSPVAPPVNDMHTDISPENALLPGTTRRTNKGVEYLVEASAAEAEBVASSVWRAAQ-----	-----	-----	-----	-----PDRDGGAEATPSGKPTFS-----
BSU	At1g03445	418	LPTINAFYDS-AIEEYVPLQH-----CAETVGNRGGVIRTVSLDRFANEDQQTNILMP-----	-----	-----	-----	-----QFIGNSTODL-----
	Allg02750	423	LPTINAFYDS-EIEEYVPLQH-----CAEVDGNRGGVIRTVSLDRFANEDQQTNILMP-----	-----	-----	-----	-----QFIGNSTODL-----
	CrBSU	439	LPSSLNAFYDCSETEECFLVLE-----TAMDVGNRPRDGRTPSIDHYLYQSQRD-----	-----	-----	-----	-----QFGQHNSSHPL-----
	BsBSU	422	LPTVETAFYDT-EIEEYVPLQEE-----RAETVGNRGGVIRTVSLDRFANESQNNILVT-----	-----	-----	-----	-----HFNEHNSSEEL-----
	Br030478	375	-----	-----	-----	-----	-----NFSREKLDL-----
	SpBSU	390	-----	-----	-----	-----	-----QFLSQRSTODL-----
BSL1	At4g03080	459	-PIVHSTT-DGTEANECDVRLHPRAVVVAKETVGSLLGGMVRQLSIDQFQNESSRRVPMNNSDVPQPTKKFTTRQKSPQGL	-----	-----	-----	-----
	Al6g38200	459	-PTVETTT-DDIANECDVRLHPRAVVVAKETVGSLLGGMVRQLSIDQFQNESSRRVPMNNSDVPQPTKKFTTRQKSPQGL	-----	-----	-----	-----
	CrBSL1	459	-PAVHSTT-DGTEANECDVRLHPRAVVVAKETVGSLLGGMVRQLSIDQFQNESSRRVPMNNSDVPQPTKKFTTRQKSPQGL	-----	-----	-----	-----
	BsBSL1	459	-PTVHSTA-DGTEANECDVRLHPRAVVVAKETVGSLLGGMVRQLSIDQFQNESSRRVPMNNSDVPQPTKKFTTRQKSPQGL	-----	-----	-----	-----
	Br018520	458	-PTVETIT-DYPETECDVRLHPRAVVVAKETVGSLLGGMVRQLSIDQFQNESSRRVPMNNSDVPQPTKKFTTRQKSPQGL	-----	-----	-----	-----
	Br000846	456	-ATVETTS-DGPETECDVRLHPRAVVVAKETVGSLLGGMVRQLSIDQFQNESSRRVPMNNSDVPQPTKKFTTRQKSPQGL	-----	-----	-----	-----
	SpBSL1	462	-PTICSTS-DGPETECDVRLHPRAVVVAKETVGSLLGGMVRQLSIDQFQNESSRRVPMNNSDVPQPTKKFTTRQKSPQGL	-----	-----	-----	-----
BSL2	At1g08420	587	LTKPDSMG-SMSVTPAGIRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	Allg08060	586	LTKPDSMG-SMSVTPAGIRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	CrBSL2	587	LTKPDSMG-SMSVTPAGIRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	BsBSL2	586	LTKPDSMG-SMSVTPAGIRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	Br030724	567	LMKPDMSG-S-----AGIRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	Br031626	565	LTKPDSMG-STGGTPAGIRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	SpBSL2	578	VFKPDSMG-SVGVTPAGIRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
BSL3	At2g27210	576	LTKPDSAV-PNSVIPAGVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	Al4g11030	572	LTKPDSAV-PNSVIPAGVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	CrBSL3	574	LTKPDSAV-PNSVIPAGVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	BsBSL3	584	LTKPDSAV-PNSVIPAGVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	Br034332	568	LTKPDSAV-STSVIPAGVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----
	SpBSL3	581	LTKPDSAV-STSVIPAGVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSV	-----	-----	-----	-----

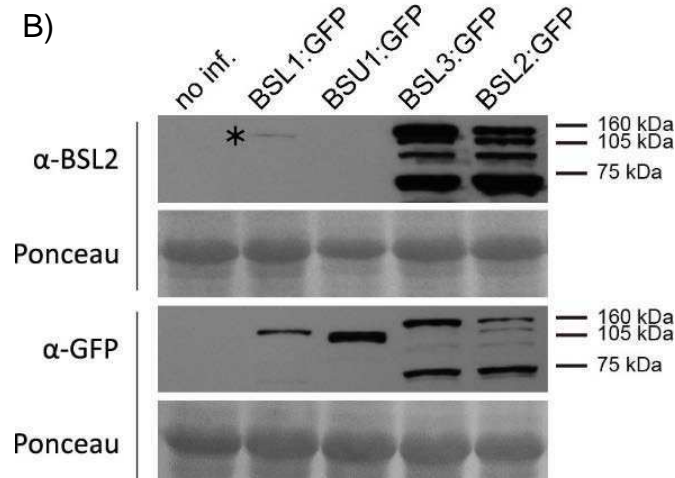


Figura 2.8 (página anterior) - **Comparación de secuencias de dominios L en Brassicaceae.** A) Alineamiento de secuencias del dominio L de proteínas BSL presentes en At: *Arabidopsis thaliana*, Al: *Arabidopsis lyrata*, Br: *Boechea stricta*, Cr: *Capsella rubella*, Br: *Brassica rapa* y Sp: *Schrenkiella parvula*. En negro se marcan los aminoácidos que no varían y en gris los cambios conservativos. El alineamiento se realizó con Muscle y se sombreó con BoxShade. B) Western Blot de extractos de hoja de *N.benthamiana* que expresan las proteínas BSL fusionadas a GFP, revelado con anticuerpos anti dominio L de BSL2 (panel superior) y anti GFP (panel inferior). El panel superior se sobreexpuso para revelar la débil afinidad del antisuero por BSL1 (señalada con un asterisco) y la ausencia de reactividad con BSU1.

Para evaluar esta posibilidad se generaron quimeras fusionando el dominio L de cada homólogo al dominio P de BSL2, y se evaluó la capacidad de mediar la interacción con el dominio P aislado de BSL2. Dado que el dominio P de BSL2 no es capaz de interactuar con otro dominio igual, puede utilizarse como un modelo para evaluar la afinidad de cada dominio L por el dominio catalítico. En la Figura 2.9 A y B se observa que el dominio L de BSL1, alcanzó para promover la interacción entre dos dominios P de BSL2, pero no presentó la misma afinidad que el dominio L de BSL2; el dominio L de BSU1, en cambio, no fue capaz de mediar interacción. La interacción de las mismas construcciones con la combinación de dominios LP de BSL2 hace las veces de control de las construcciones quiméricas: dado que en todas las combinaciones hay dos dominios P y un dominio L funcional, todas pueden activar al gen reportero.

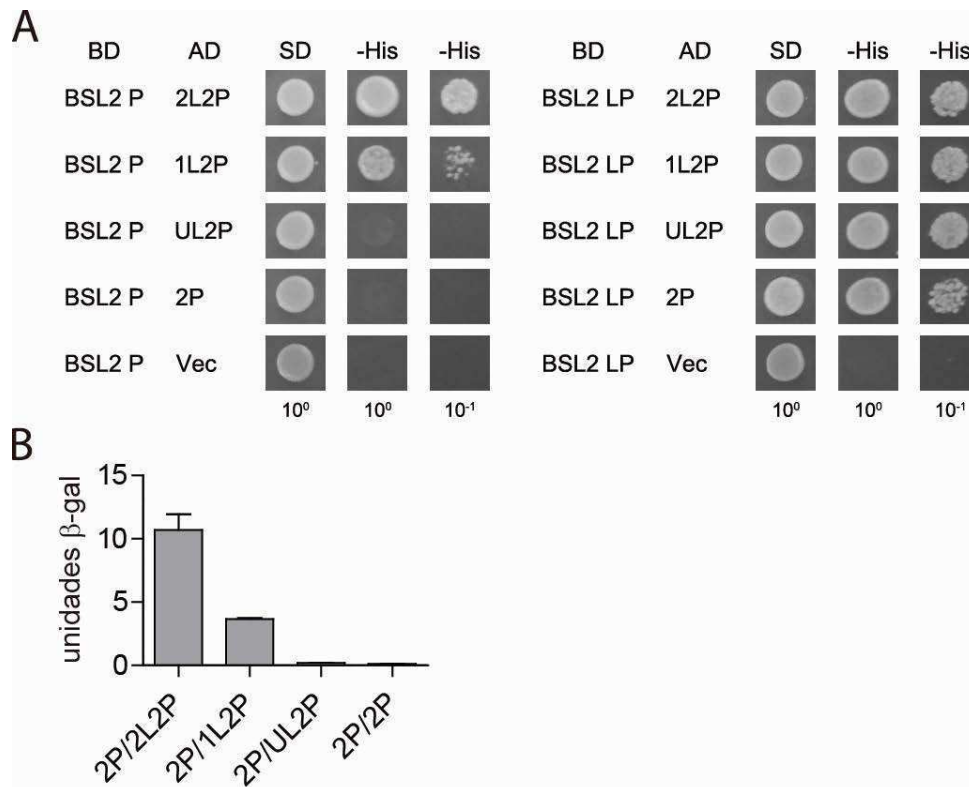


Figura 2.9 - Los dominios L difieren en su habilidad de mediar la interacción entre los dominios P de BSL2. Construcciones quiméricas de dominios L de proteínas BSL fusionados al dominio P de BSL2 se utilizaron en ensayos de Y2H usando como reportero el crecimiento en ausencia de Histidina para evaluar la interacción con otro dominio P (izquierda) y con la combinación de dominios LP (derecha) como control positivo. En ambos casos se muestran dos diluciones sucesivas ($DO^{600}=1$ y $DO^{600}=0,1$) B) Cuantificación de la actividad del reportero β-galactosidasa en las combinaciones mostradas en A. L: Dominio conector; P: Dominio catalítico; 1: BSL1; 2: BSL2; U: BSU1; xLyP: quimera entre el dominio L de la proteína x y el dominio P de la proteína y; SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de las diluciones seriadas de los respectivos cultivos, partiendo de $DO^{600nm}=1$ (10^0).

Modo de acción del dominio conector

Dado el requerimiento del dominio L, pero con la evidencia de que por sí mismo o en combinación con el dominio K (Figuras 2.5 y 2.6) es incapaz de mediar la interacción, surgió la pregunta de si el dominio L interacciona con el dominio P de la otra subunidad o si lo hace con el dominio P al que está unido, generando así una superficie de interacción con la que se asocia la otra subunidad. Para responder esta pregunta, se utilizó la fosfatasa TOPP1, una de las 9 isoformas de PP1 codificadas en el genoma de Arabidopsis. Como se describió, el dominio catalítico de PP1 actúa como un monómero al que se unen diversas subunidades regulatorias. A nivel de secuencia, las PP1 están cercanamente emparentadas con el dominio P de las fosfatasas BSL.

Los dominios catalíticos de BSL2 y TOPP1 no son capaces de interactuar en ensayos de Y2H (Figura 2.10 A). A partir de este dato, se generó una quimera entre el dominio L de BSL2 y TOPP1 y se ensayó la interacción con el dominio P de BSL2. El dominio L de BSL2 confirió a TOPP1 la capacidad de interactuar con el dominio P de BSL2. Sin embargo, cuando el dominio L está unido al dominio catalítico de BSL2 no se verifica interacción con TOPP1. Esto indicaría que la interacción del dominio L ocurre en *trans*, o sea con el dominio P de la otra subunidad: la construcción BSL2-LP no interactúa con TOPP1 (que presenta un nivel de expresión similar al de la quimera 2L:TOPP1, Figura 2.10 B) probablemente debido a que el dominio L de BSL2 no tiene afinidad por TOPP1, mientras que la construcción BSL2L:TOPP1 posee un dominio L con afinidad por el dominio catalítico de BSL2. Estos resultados parecen contradecir los mostrados en las Figuras 2.5 y 2.6, que indican que el dominio L, aislado o asociado al extremo C-terminal del dominio K, no podría interactuar de forma directa con el dominio P. Este resultado será evaluado a la luz de los resultados del Capítulo III.

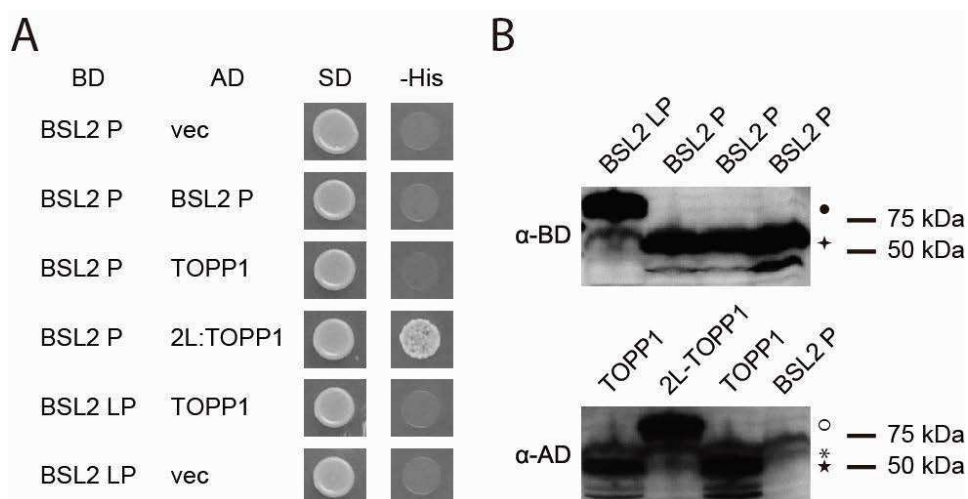


Figura 2.10 - Acción en trans del dominio L de BSL2. A) Ensayo de Y2H usando como reportero el crecimiento en ausencia de Histidina. Se evalúa la capacidad de interactuar de las construcciones indicadas. Vec: vector vacío; 2L:TOPP1: Quimera formada por el dominio L de BSL2 fusionado a la fosfatasa TOPP1. SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. B) Inmunodetección de BD (panel superior) y AD (panel inferior) en extractos proteicos de levaduras expresando las construcciones indicadas. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$. El círculo lleno (●) señala la banda correspondiente a BSL2 LP y la estrella de cuatro puntas (◆) señala la banda correspondiente a BSL2 P. El círculo vacío (○) indica la posición de la banda correspondiente al producto quimérico 2L-TOPP1, el asterisco (*) la de BSL2 P y la estrella de cinco puntas (★) se encuentra a la altura de la banda correspondiente a TOPP1.

Análisis delecional del dominio conector

Habiendo caracterizado las diferencias entre los dominios conectores de los distintos homólogos y determinado la necesidad de este dominio para la interacción entre dos dominios catalíticos de BSL2 surgió la pregunta de si todo el dominio es requerido o si algunas regiones son más importantes que otras.

Para responder esta pregunta se realizó un análisis delecional del dominio L de BSL2. Dado que este dominio no tiene similitudes aparentes con ninguna otra secuencia previamente caracterizada, para definir posibles zonas relevantes se estudió la conservación de secuencia aminoacídica y se realizó una predicción de estructura secundaria (Figura 2.11). Con la excepción de una zona de alrededor de 25 aminoácidos que se predice formaría una hélice α , la mayoría del dominio se predijo desestructurado (Figura 2.11 A). En cuanto a la conservación de secuencias entre proteínas de tipo BSL2 en angiospermas, el dominio L es una de las zonas menos conservadas en el contexto de la proteína completa, junto a los extremos N- y C-terminal. Dentro del dominio existe una alternancia de zonas con considerable grado de conservación con otras de alta variabilidad (Figura 2.11 B). Teniendo en cuenta estos datos, se definieron inicialmente cuatro regiones: 450-555, 556-599, 600-642 y 643-662. La región 556-599 se definió como la de menor conservación de secuencia y que se predice como más homogéneamente desestructurada; la región 600-642, por el contrario, es de las zonas más conservadas y presenta una predicción de estructura ambigua, pero con posibles elementos discernibles; las otras regiones se eligieron, *a priori*, por diferencia con éstas.

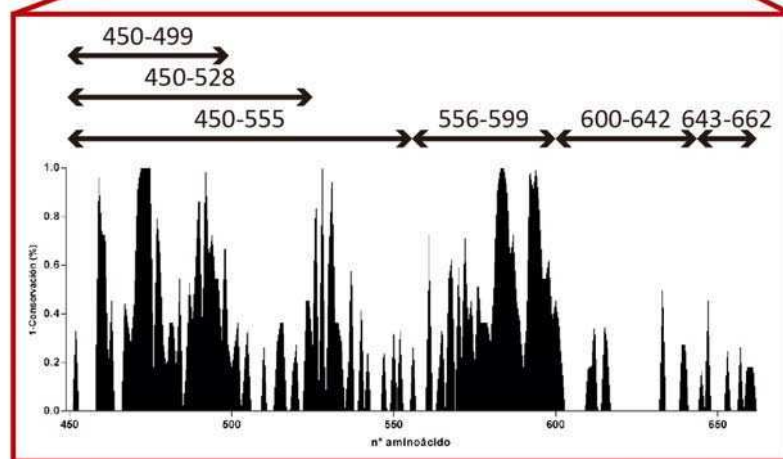
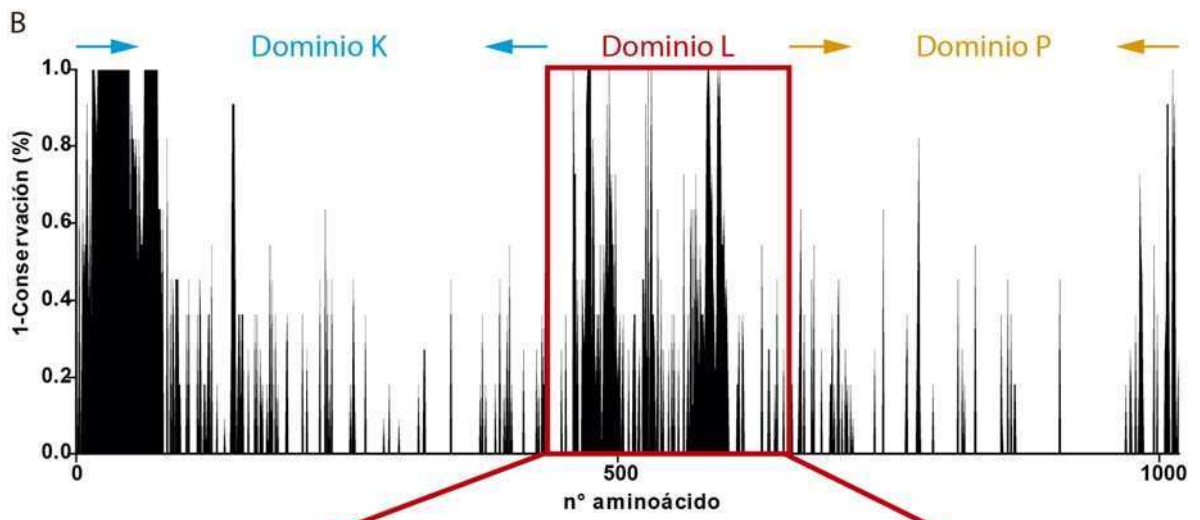
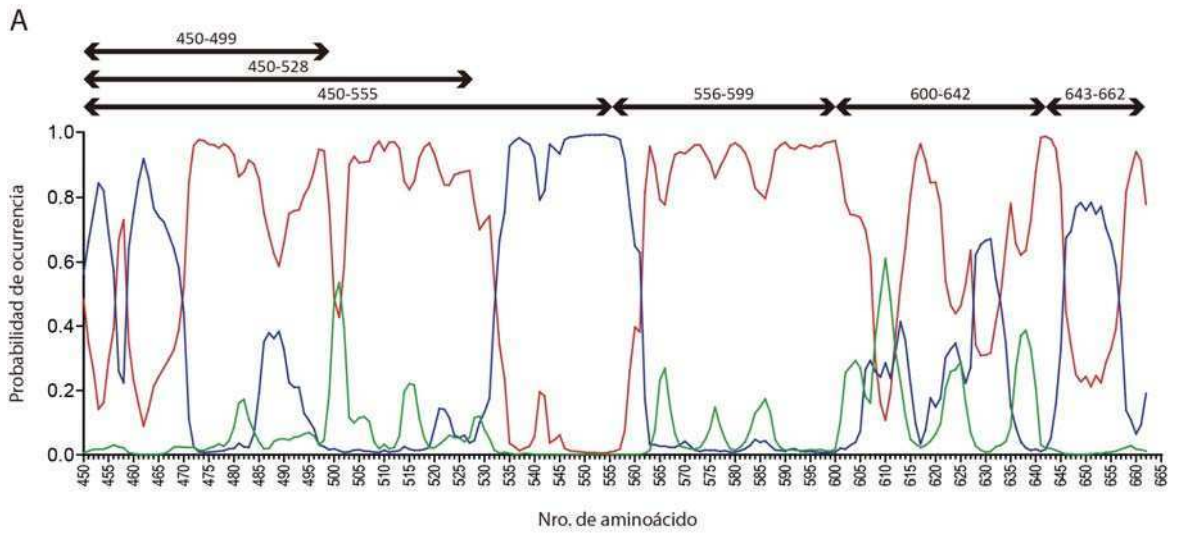


Figura 2.11 (página anterior) - **Predicción de estructura secundaria y conservación de secuencia del dominio L de BSL2.** A) Predicción de estructura secundaria. En rojo, probabilidad de desorden; en azul, probabilidad de hélice α y en verde, probabilidad de lámina β (PSIPRED v3.3 (Buchan et al., 2010)). B) Las secuencias de ortólogos de BSL2 en angiospermas con genomas completamente secuenciados se alinearon con MUSCLE, y se determinó el grado de conservación de cada posición. Los datos se presentan como grado de variabilidad [$1 - (\text{grado de conservación})$]. En A y en B se indican con flechas las zonas definidas para realizar las deleciones. Las secuencias utilizadas se incluyen en el Anexo I (digital) y corresponden a *A.trichopoda*, *N.advena*, *O.sativa*, *S.italica*, *M.acuminata*, *A.caerulea*, *V.vinifera*, *P.trichocarpa*, *E.grandis*, *L.usitatissimum*, *R.communis*, *C.papaya*, *G.raimondii*, *T.cacao*, *A.thaliana*, *BSL2*, *C.sinensis*, *C.sativus*, *M.truncatula*, *P.persica*, *U.gibba*, *N.benthamiana* y *M.guttatus*.

La capacidad de interactuar de las construcciones que presentan el dominio conector con las diferentes deleciones asociado al dominio catalítico P (LP $\Delta x-y$) se ensayaron en levaduras, en la forma de fusiones al dominio AD de Gal4, enfrentándolas a la secuencia de prueba, el dominio P aislado, y al control positivo de interacción, la secuencia dominio conector-dominio catalítico (LP) completa, estas últimas como fusiones al dominio BD de Gal4. Se evaluó la capacidad de crecimiento en ausencia de His y la actividad del reportero β -galactosidasa (Figura 2.12).

En una primera aproximación, se realizaron construcciones eliminando las tres zonas N-terminales adyacentes en forma aditiva, generando las deleciones $\Delta 450-555$, $\Delta 450-599$ y $\Delta 450-642$ (Figura 2.12). Se observó que la deleción $\Delta 450-555$ produjo una disminución en la expresión de β -galactosidasa respecto del dominio conector completo (Figura 2.12). La deleción del fragmento de mayor tamaño ($\Delta 450-642$) anuló la capacidad de interacción con el dominio P aislado. Para explorar la contribución de esta última zona se obtuvo la forma con la deleción $\Delta 600-642$, en el contexto del resto del dominio L; en este caso también se produjo una disminución, pero no anulación, de la expresión de los reporteros.

Dado que la ausencia de las regiones 450-555 y 600-642 provocó una disminución en la interacción, se decidió combinarlas, generando la forma $\Delta 450-555/\Delta 600-642$: estas deleciones conjuntas anularon por completo la interacción. Cabe destacar que la longitud del dominio L en esta última forma (62 aminoácidos) resultó equivalente a la de la deleción $\Delta 450-599$, que mostró capacidad, aunque disminuida, de interacción. De esta forma se pudo descartar que la anulación de la interacción se debiera a un efecto de la disminución de la longitud de la secuencia previa al dominio P que podía generar algún tipo de impedimento estérico con el dominio de Gal4.

Se lograron así definir dos regiones determinantes de la interacción, la secuencia 450-555 y la secuencia 600-642. Ambas parecen actuar de forma aditiva, lo que se comprueba por el efecto parcial de pérdida en la intensidad de la interacción al presentarse de forma individual, mientras que su ocurrencia simultánea anula por completo la interacción entre los dominios P. La región 450-555 es relativamente extensa, e incluye la única región del dominio L (residuos 529-555) que se predice con cierto grado de confianza que adopta algún tipo de estructura secundaria. Para lograr acotar la/s región/es determinante/s en esta zona, se realizaron dos deleciones parciales ($\Delta 450-499$ y $\Delta 450-528$, Figura 2.12) en el contexto sensibilizado de la proteína que lleva la deleción $\Delta 600-642$. Se pudo así acotar la zona mínima necesaria para que tenga lugar la interacción: las regiones 500-528 y 600-642. Estas regiones presentan un alto grado de conservación de secuencia entre los dominios L de angiospermas. Significativamente, la región 529-555, la única que se predice con alta probabilidad que podría adoptar estructura secundaria, no parece ser necesaria para la interacción.

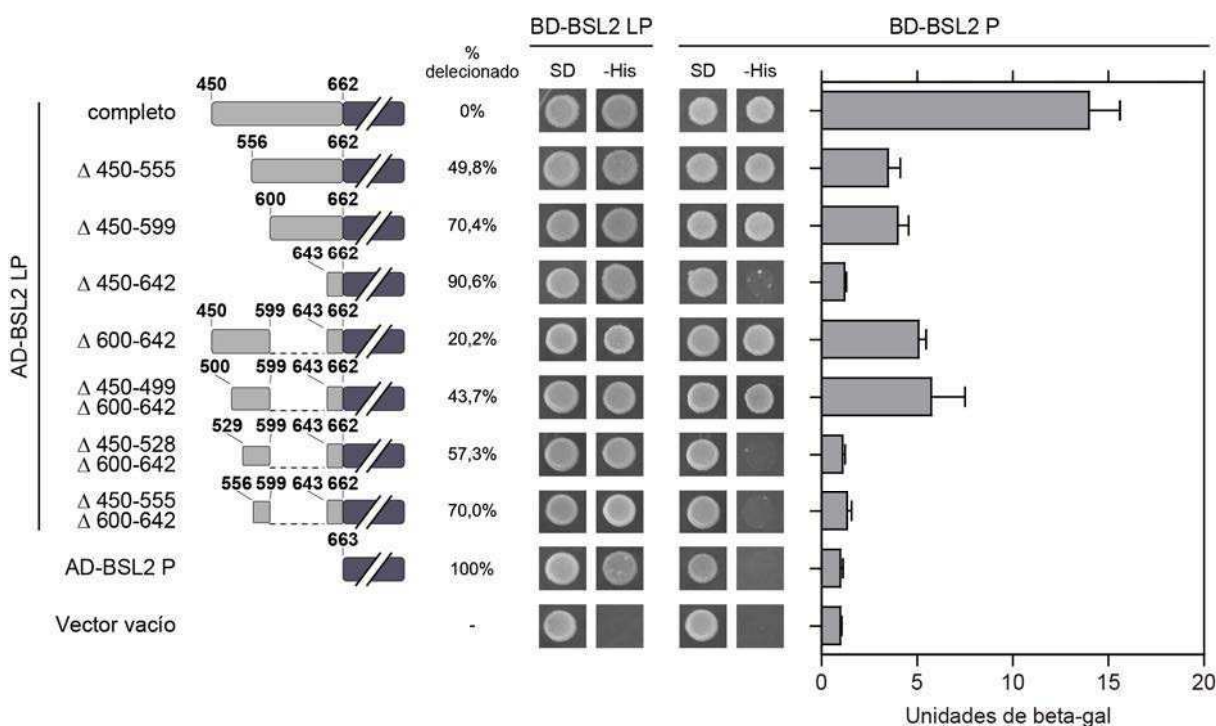


Figura 2.12 - Regiones del dominio L de BSL2 necesarias para la interacción. Construcciones que expresan dominios L y P (LP) con distintas deleciones en la región L se enfrentaron al dominio P o al dominio LP en ensayos de Y2H. Se indican los aminoácidos delecionados en cada construcción. En el esquema, la barra gris claro representa al dominio L y la gris oscuro al dominio P. SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$.

Modificaciones en la región 600-642 del dominio L de BSL2

La región 600-642 del dominio L de BSL2 es una de las dos regiones determinantes para la capacidad del mismo de mediar interacciones homotópicas. Es además una de las zonas con mayor conservación de secuencia entre diferentes especies, lo que sugiere un papel funcional y/o estructural importante. En la figura 2.13 se muestra un detalle de esta zona, donde se señalan dos características salientes: la Ser627 (en la numeración de AtBSL2) y la secuencia que recuerda al motivo RVxF en las proteínas de tipo BSL2.

	642	663
A.thaliana BSL2	MVRQLSIDQFENEGRRVSYGTPES--ATAARKLLDRQMSINSVPPKKVIAHLLKPRGWKPPVRRQFFL	
M.guttatus	MVRQLSIDQFENEGRRVSYGTPES--ATAARKLLDRQMSISSVPPKKVITHLLKPRGWKPPVRRQFFL	
V.vinifera	MVRQLSIDQFENEGRRVSYGTPES--ATAARKLLDRQMSINSVPPKKVIAHLLKPRGWKPPVRRQFFL	
P.trichocarpa	MVRQLSIDQFENEGRRVSYGTPES--ATAARKLLDRQMSINSVPPKKVIVHLLKPRGWKPPVRRQFFL	
A.caerulea	MVRQLSIDQFENEGRRVSYGTPES--ATAARKLLDRQMSINSVPPKKVIAHLLKPRGWKPPVRRQFFL	
O.sativa	MVRQLSIDQFENEGRRVSYGTPEN--ATAARKLLDRQMSINSVPPKKVIASLLKPRGWKPPVRRQFFL	
N.advena	MVRQLSIDQFENEGRRVSYGTPES--ANAARKLLDRQMSISSVPPKKVIAHLLKPRGWKPPVRRQFFL	
P.glauca	LVRQLSIDQFENEGRRVSYGTPES--AVAARKLLDRQMSINSIPKKVLAYLLKPRGWKPPVRRQFFL	
W.mirabilis	MVRQLSIDQFENEGRRVSYGTPES--ANAARKLLDRQMSINSVPPKKILTYLLKPRGWKPPVRRQFFL	
P.aquilinum	LVRQLSLDQFENEGRRVSYGNVDS--GIAARQLDRQMSTRGLHKKIISLILLPRGWKPPVRRQFFL	
S.moellendorffii	MVRQLSIDQFENEGRRVSYGTPES--ANAARKLLDRRTTSEIGVQKKVINYLKPRGWKPPVRRQFFL	
P.patens	LVRQLSIDQFENEGRRVSYGTPES--ASHARLLDRQTSVHGVQKKVLSYLLKSRGWKPPVRRQFFM	
M.polymorpha	LVRQLSIDQFENEGRRVSYGTPDS--ASAARKLLDRQMSISGVQKKVLGHLKPRGWKPPVRRQFFM	
C.orbicularis	RQLSIDHLENEGRRVSYGTPDGSSAAARKLLDRQMSIQGMHKKVLNYLLKPRGWKPPVRRQFFL	
K.flaccidum	MVRQLSIDQFENEGRRVSYGTPDSSHGKEVNKMLDRSMSLQGVHKKVLSALLKPRGWKPPVRRQFFL	
A.thaliana BSL1	MVRQLSLDQFQNESRRMVPMMNSDV-PQPTTK-FTRQKSPQGLHKKVIAALLRPRNWKPPGNRKFFL	
M.guttatus	MVRQLSLDQFENESRRMIPSHN-DL-SNPTTK-FARQRSPQGLHKKIISTLLRPRNWKPPVNRKFFL	
V.vinifera	MVRQLSLDQFENESRRMVPMMN-DL-SYPTTK-FTRQKSPQGLHKKIISTLLRPRNWKAPANRRFFL	
P.trichocarpa	LVRQLSLDQFENESRRMLPMNN-DA-SYPARK-FTRKKSPQGLHKKIISMLLRPRNWKAPANRRFFL	
A.caerulea	LVRQLSLDQFENESRRMSPLNS-DT-TYPTKFFTRQKSPQGLHKKIISALLRPRNWKAPTNRFFL	
O.sativa	LVRQLSLDQFENESRRMHPSN-DQ-SYPAKKALNRQSPQGLHKKVIFLLKPRNWRAPAERAFFL	
N.advena	LVRQLSLDQFENESRRMNPWNM-DQ-GYAAKRIFNKQKSPVGLHKKIISLLRPRNWSAPMVRKFFL	
P.glauca	LVRQLSLDQFENEGRRMSPWNL-DQ-AHAARKLLIRQKSPVALHKKVIFLLRPRNWKPPVNRGFFL	
C.revoluta	LVRQLSLDQFENEGRRMSPLNT-DN-AYGARKLLHRQKSPVGLHKKVIFLLRPRNWKPPVNRQFFL	
S.moellendorffii	LVRQLSLDRFENEGRRVTRTEI-----SRRVLERQISMYGVHKKVLAALLKPREWKPSLRNLFLL	

Figura 2.13 - Alineamiento de los aminoácidos que preceden al dominio catalítico en proteínas de tipo BSL2/BSL3 (parte superior) y BSL1 (parte inferior). El dominio catalítico tal como está definido en esta tesis se muestra sombreado en gris claro. La numeración corresponde a AtBSL2. En sombreado negro se indican Ser627 y la región con similitud con el motivo RVxF. Las especies mostradas son *Arabidopsis thaliana*, *Mimulus guttatus*, *Vitis vinifera*, *Populus trichocarpa*, *Aquilegia caerulea*, *Oryza sativa*, *Nufar advena* (Magnoliophyta), *Picea glauca* (Coniferophyta), *Welwitschia mirabilis* (Gnetophyta), *Cycas rumphii* (Cycadophyta), *Pteridium aquilinum* (Moniliformopses), *Coleochaete orbicularis* (Coleochaetophyceae) y *Klebsormidium flaccidum* (Klebsormidiophyceae).

Si se realiza una búsqueda en la base de datos de fosfopéptidos identificados por espectrometría de masas en *Arabidopsis* (PhosPhAt - <http://phosphat.uni-hohenheim.de/index.html>), se encuentra que el residuo Ser en posición 627 de la secuencia de BSL2 se halla fosforilado. Esta Ser se encuentra altamente conservada en todas las PPKL, tanto en viridiplantas como en alveolados, excepto en algunas isoformas de tipo BSU1 (ver Figura 2.8). Este residuo se encuentra fosforilado, además, en péptidos recuperados no sólo

de los homólogos de BSL2 en Arabidopsis, sino también de PPKL en otras especies, incluyendo alveolados (de la Fuente van Bentem et al., 2008; Treeck et al., 2011; Rose et al., 2012; Lv et al., 2014).

Por otro lado, el motivo de tipo RVxF en las proteínas de tipo BSL2 (RRVSY en AtBSL2) se encuentra altamente conservado aún en musgos y algas estreptofitas, mientras que está característicamente modificado en las de tipo BSL1 (RRMVP en AtBSL1).

Con la finalidad de evaluar el posible papel de la fosforilación en Ser627 en la interacción de BSL2 consigo misma, se reemplazó este residuo por ácido glutámico (S627E), generando una versión fosfomimética, o por alanina (S627A), alterando la naturaleza del aminoácido. El efecto de las mutaciones se evaluó mediante el sistema de Y2H, ensayando la afinidad de un dominio P por otro dominio P asociado al fragmento $\Delta 450-599$, es decir, la mínima porción que incluye la posición 627 y permite la interacción (ver Figura 2.12). Sin embargo, en este sistema ninguna de las versiones generadas afectó la capacidad de interacción (Figura 2.14).

BD	AD	SD	-His
BSL2 P	BSL2 LP ($\Delta 450-599$)		
BSL2 P	BSL2 LP ($\Delta 450-599$) - S627A		
BSL2 P	BSL2 LP ($\Delta 450-599$) - S627E		

Figura 2.14 - Mutaciones en la Ser627 conservada de la región 600-642 del dominio conector no afectan la capacidad de mediar la interacción de BSL2. Ensayo de Y2H donde se evalúa la capacidad de interaccionar de construcciones de BSL2 con cambios en Ser627 en el contexto de la mínima porción del dominio conector que permite la interacción ($\Delta 450-599$). SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$.

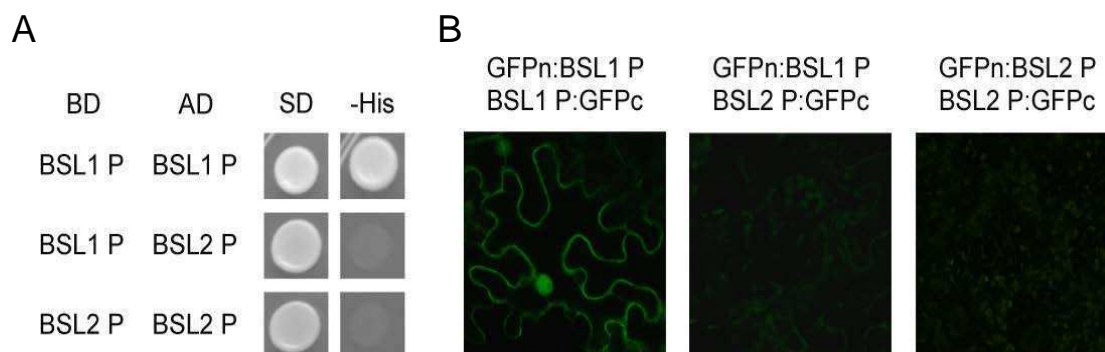
También se modificó la secuencia 636-640 RRVSY presente en BSL2 por RRMVP presente en BSL1, en el mismo contexto que en el caso anterior, es decir, en la forma $\Delta 450-599$ y se ensayó la capacidad de interacción de esta especie con un dominio P aislado por Y2H. Tampoco en este caso las mutaciones introducidas afectaron la capacidad de interacción (Figura 2.15).



Figura 2.15 - El reemplazo de la secuencia RRVSY conservada de la región 600-642 del dominio conector no afecta la capacidad de mediar la interacción de BSL2. Ensayo de Y2H donde se evalúa la capacidad de interaccionar de construcciones de BSL2 con la secuencia RRVSY de la subfamilia BSL2/3 modificada por la secuencia RRMVP de la subfamilia BSL1, en el contexto de la mínima porción del dominio conector que permite la interacción (Δ 450-599). SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$.

Diferencias entre los dominios catalíticos de BSL1 y BSL2

El hecho de que dos dominios catalíticos de BSL1 sean suficientes para mediar la interacción, mientras que en BSL2 sea requerido además un dominio L, permite inferir que existen diferencias relevantes entre los dominios catalíticos de estos homólogos, aunque guarden entre sí un 81% de identidad. Para encarar el estudio de estas diferencias, se analizó en primer lugar la capacidad de interacción entre dominios P de estas proteínas. Como se muestra en la Figura 2.16, estas diferencias determinan que no se observe interacción entre un dominio P de BSL1 y un dominio P de BSL2: el dominio P de BSL1, por lo tanto, parece ser capaz de discriminar entre secuencias propias y no propias (Figura 2.16).

**Figura 2.16**

Los dominios P de BSL1 y BSL2 no presentan afinidad entre sí. La capacidad de interactuar de las construcciones indicadas se ensayó mediante A) Ensayo de Y2H, B) biFC. La fluorescencia observada en el 2do y 3er panel (interacción mixta entre BSL1P y BSL2P o interacción de BSL2P consigo misma, respectivamente) no difiere del negativo en las condiciones evaluadas. SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de 10 mM 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de diluciones de los respectivos cultivos normalizadas a $DO^{600nm}=1$.

Seguidamente se realizó un alineamiento de secuencias aminoacídicas de los dominios catalíticos de BSL1 y BSL2 de diversas especies de plantas con semillas (organismos en los que se encuentran secuencias de tipo BSL1) con el fin de definir posibles residuos responsables del comportamiento diferencial (Figura 2.17). Este análisis reveló que los aminoácidos cuyos cambios no son conservativos pueden dividirse en dos grupos según su posición en el dominio: cambios que corresponden al extremo C-terminal y cambios que ocurren en el núcleo estructural.

El extremo C-terminal comprende los últimos 40-41 aminoácidos de la proteína. Esta zona presenta una menor conservación de secuencia respecto del resto del dominio y al analizar la proteína mediante programas de predicción de estructura secundaria, esta zona se predice como desestructurada (Ver capítulo III).

B1. *P. taeda* 546 HKKVISLLRPRNWRPPVNRGFFLDSEYVGLCYTAEQIFMHEPTTLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B1. *N. advena* 555 HKKVISLLRPRNWSAPVVRKFFLDSEYVGLCYAABQIFLQEPITVLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B1. *O. sativa* 537 HKKVISLLRPRNWRPAPAEARFFLDSEYVGLCYAABQIFMHEPTTLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B1. *A. caerulea* 531 HKKVISALLRPRNWKAPNRRFFLDSEYVGLCYAABQIFMHEPTTLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B1. *V. vinifera* 534 HKKVISLLRPRNWKAPANRRFFLDSEYVGLCYAABQIFMHEPTTLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B1. *R. communis* 531 HKKVISLLRPRNWKAPANRRFFLDSEYVGLCYAABQIFMHEPTTLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B1. *A. thaliana* 527 HKKVIAALLRPRNWKPPNRKFFLDSEYVGLCYAABQIFMHEPTTLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B1. *M. guttatus* 523 HKKVISLLRPRNWKPPVNRKFFLDSEYVGLCYAABQIFMHEPTTLQLKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *P. taeda* 650 PKKVIHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDITYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *N. advena* 654 PKKVIAHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDIAYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *O. sativa* 648 PKKVIAHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDIAYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *A. caerulea* 669 PKKVIAHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDIAYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *V. vinifera* 624 PKKVIAHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDIAYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *R. communis* 645 PKKVIAHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDIAYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *A. thaliana* 663 PKKVIAHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDIAYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL
 B2. *M. guttatus* 651 PKKVIHLLKPRGWKPPVRRQFFLDCEIAIDLCDSAEIRIFSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGFPPSTAGDIAYIDYFLFLGDYVDRGQHSLETITLLLALKVEYEPDNVHL

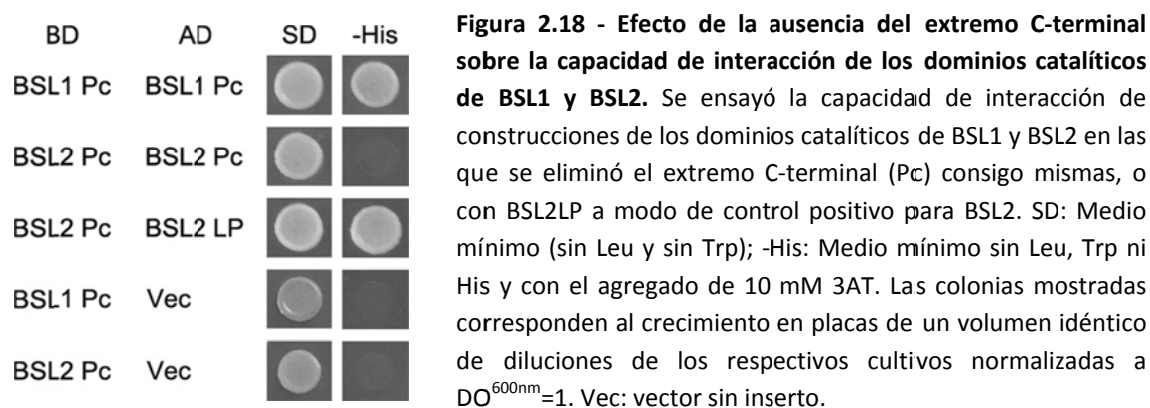
B1. *P. taeda* 667 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B1. *N. advena* 676 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B1. *O. sativa* 658 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B1. *A. caerulea* 652 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B1. *V. vinifera* 655 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B1. *R. communis* 652 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B1. *A. thaliana* 648 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B1. *M. guttatus* 644 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *P. taeda* 771 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *N. advena* 775 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *O. sativa* 769 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *A. caerulea* 790 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *V. vinifera* 745 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *R. communis* 766 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *A. thaliana* 784 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 B2. *M. guttatus* 772 IRGNHEADINALFGFRIECIERMGERDGIWAWTRFNQLFNWLPLAALIEKKIICMHGGIGRSINHVQEIKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV

B1. *P. taeda* 787 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQPA-HDR-NSLAWI
 B1. *N. advena* 796 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQPA-HDR-NSLAWI
 B1. *O. sativa* 778 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQSA-SDR-NSLAWI
 B1. *A. caerulea* 772 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQPD-HDR-NSLAWI
 B1. *V. vinifera* 775 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQPD-LDR-NSLAWI
 B1. *R. communis* 772 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQPD-LDR-NSLAWI
 B1. *A. thaliana* 768 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQPD-FDR-NSLAWI
 B1. *M. guttatus* 763 LDFCKKMKLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPPISPEPSPERHIEDTWMQELNQRPPPTTRGRPQPA-HDR-NSLAWI
 B2. *P. taeda* 891 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQAS-NDR-GSLAWI
 B2. *N. advena* 895 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQVA-NDR-GSLAWI
 B2. *O. sativa* 889 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQVA-NDR-GSLAWI
 B2. *A. caerulea* 910 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQVT-NDR-GSLAWI
 B2. *V. vinifera* 865 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQVT-NDR-GSLAWI
 B2. *R. communis* 886 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQVT-NDR-GSLAWI
 B2. *A. thaliana* 904 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQNS-NDR-GSLAWI
 B2. *M. guttatus* 892 MEFCCNNNDLQLIIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAISPETSPEPHIEDTWMQELNQRPPPTTRGRPQVA-NDR-GSLAWI

Figura 2.17 (página anterior) - **Comparación entre los dominios catalíticos de BSL1 y BSL2.** Alineamiento de secuencias del dominio P de fosfatasas de tipo BSL1 y BSL2 presentes en las especies indicadas. En negro se marcan los aminoácidos que no varían y en gris los cambios conservativos. Las estrellas negras señalan las posiciones variables principales. Los círculos blancos, los residuos que forman el surco RVxF. El extremo C-terminal se denota con una línea sobre las secuencias. B1: secuencias de tipo BSL1. B2: secuencias de tipo BSL2/BSL3.

El hecho de poseer un extremo C-terminal de estructura indefinida y variable es una característica usual dentro de la familia de fosfatasas PPP. De hecho, este extremo ha sido eliminado para poder determinar estructuras de PP1, dado que su naturaleza desordenada interfiere con el empaquetamiento del cristal (por ejemplo (Kelker et al., 2009)).

Con el fin de comprobar si las diferencias observadas en esta región contribuyen al comportamiento diferencial entre BSL1 y BSL2, se realizaron construcciones que expresaban el dominio catalítico de ambas fosfatasas sin su extremo C-terminal. El análisis evidenció que, en este ensayo, esta región no afectó el comportamiento de los dominios catalíticos, ya que las construcciones sin el extremo C-terminal se comportaron igual que las que expresan el dominio completo (Figura 2.18).



A partir de lo visto, la atención se centró en el análisis de las zonas diferenciales encontradas en el núcleo estructural del dominio. Si bien la estructura del domino catalítico de las fosfatasas BSL no ha sido resuelta, su alta homología con dominios catalíticos de fosfatasas de la familia PPP, cuya estructura es conocida, permitió realizar un modelado por homología (Figura 2.19).

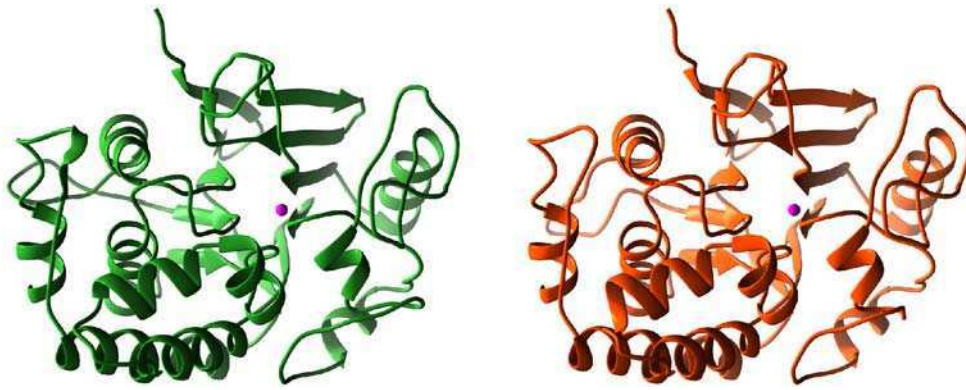


Figura 2.19 - Modelo de la estructura de los dominios catalíticos de BSL1 y BSL2. Modelado por homología de los dominios catalíticos de BSL1 (izquierda en verde) y BSL2 (derecha en naranja) tomando como base la estructura de PP1 de conejo (PDB 1fjm) y utilizando el programa Swiss Model. La bondad de ajuste de los modelos se evaluó por QMEAN (Benkert et al., 2008) a través del servidor QMEAN (<http://swissmodel.expasy.org/qmean/cgi/index.cgi>). Para BSL1 se obtuvo un valor de 0.752 con 92% de los residuos en regiones favorecidas y para BSL2 de 0.679 con 91% de los residuos en regiones favorecidas. Se muestra una vista de frente al sitio activo donde puede apreciarse uno de los átomos metálicos en color fucsia.

Al ubicar las diferencias de secuencia entre BSL1 y BSL2 en la estructura tridimensional simulada pudo observarse que la mayoría de los cambios no conservativos se ubican en dos zonas expuestas al solvente. Una es en las vueltas expuestas al solvente de dos hélices anfipáticas que se ubican paralelas sobre la superficie opuesta a la del centro activo. La otra es la región circundante al surco donde, en las fosfatasas de tipo PP1, se aloja la secuencia consenso RVxF presente en muchas de las subunidades regulatorias conocidas (Figura 2.20).

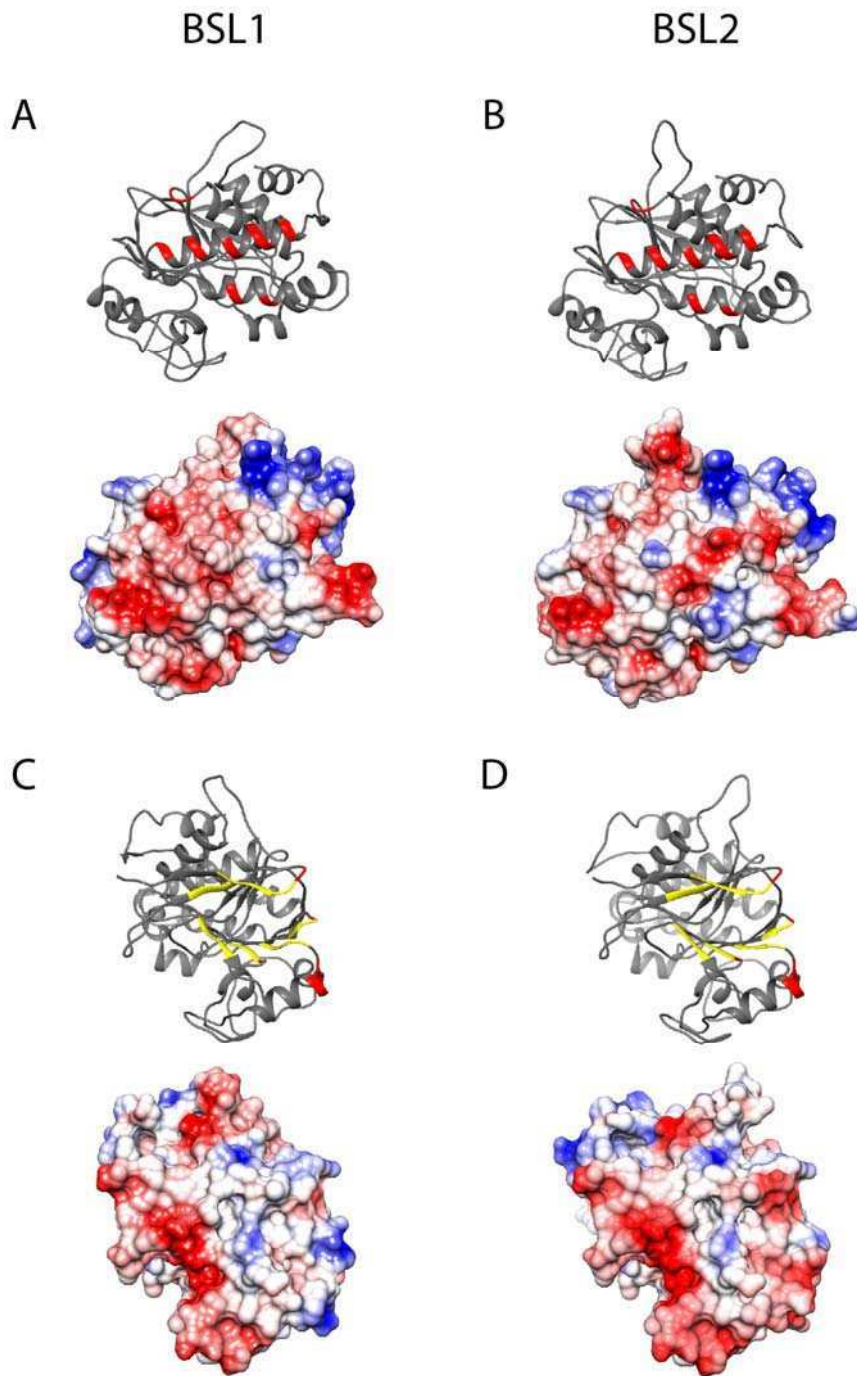


Figura 2.20 - Localización de los cambios no conservativos en la estructura tridimensional simulada de los dominios catalíticos de BSL1 y BSL2. Cada panel presenta la misma vista de la estructura en representación de cintas (arriba) y en forma de superficie accesible coloreada según su potencial superficial Coulombico (abajo). Los paneles A y B muestran la cara de la proteína opuesta al sitio activo; en las representaciones en cinta, en rojo se indican los residuos diferenciales localizados en las caras expuestas al solvente de las dos hélices paralelas y en un loop localizado sobre la misma cara. En los paneles C y D los residuos que bordean el surco de interacción con la secuencia consenso RVxF se muestran en amarillo; en rojo se señalan los aminoácidos diferenciales. Coloreado de superficie Coulombica realizado con los parámetros preestablecidos por Chimera V1.10.

Las quimeras tenían como objetivo identificar determinantes de secuencia que posibilitan la afinidad entre dos dominios P de BSL1 que no están presentes en el dominio P de BSL2. Sin embargo, al evaluar la interacción de las quimeras con el dominio P de BSL1, la interacción observada fue mucho más débil que entre los mismos dominios P silvestres (Figura 2.22), lo que imposibilitó extraer conclusiones a partir de estas construcciones.

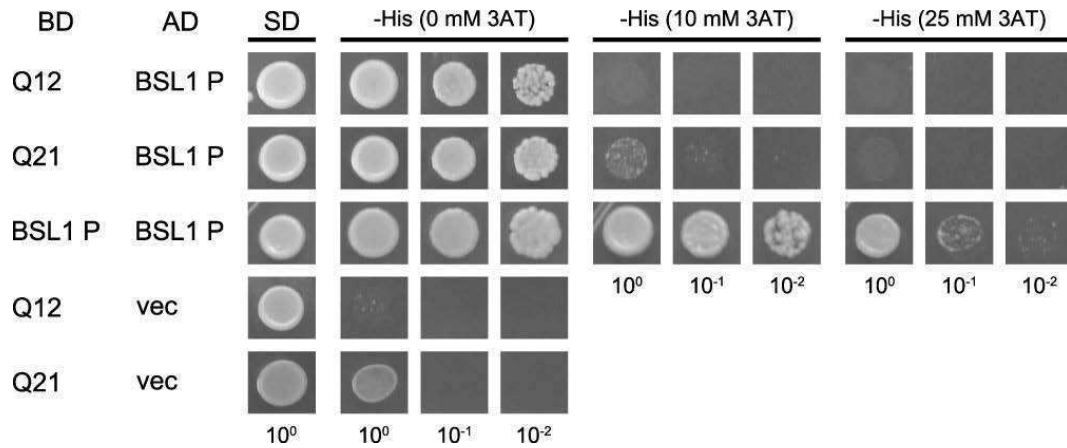


Figura 2.22 - Interacción de las formas quiméricas del dominio catalítico con BSL1. Las quimeras resultantes de combinar fragmentos de dominios P de BSL1 y BSL2 (Q12: quimera entre el extremo N-terminal de BSL1 y el C-terminal de BSL2, Q21: caso complementario) fueron evaluadas por su capacidad de interactuar con el dominio P de BSL1. Se muestran diluciones sucesivas con distintas concentraciones de 3AT como forma de evaluar afinidad. SD: Medio mínimo (sin Leu y sin Trp); -His: Medio mínimo sin Leu, Trp ni His y con el agregado de las concentraciones indicadas de 3AT. Las colonias mostradas corresponden al crecimiento en placas de un volumen idéntico de las diluciones seriadas de los respectivos cultivos, partiendo de $DO^{600nm}=1$ (10^0).

Conclusiones

Las fosfatasa de la familia PPP se caracterizan por interactuar con un gran número de subunidades con propiedades regulatorias o que otorgan especificidad de sustrato y/o localización subcelular al dominio catalítico. Para avanzar en el estudio de las proteínas asociadas a las fosfatasa BSL, y como forma de comprender sus funciones, se llevó a cabo una búsqueda de proteínas asociadas mediante Y2H usando el dominio catalítico de BSL2 como carnada. Se encontró que muchos de los clones recuperados codificaban para el mismo dominio asociado a porciones de secuencia N-terminal adyacente. En trabajos anteriores (Gustavo Maselli, tesis de licenciatura) se validó lo encontrado en la búsqueda por Y2H, y en este capítulo se amplió la caracterización de estas interacciones.

La principal conclusión a la que se llegó en esta sección es que estas proteínas forman interacciones homotípicas, probablemente en forma de dímeros, y se determinó que los representantes de los dos grupos principales de proteínas PPKL presentes en *Arabidopsis* alcanzan esta conformación a través de diferentes caminos.

La interacción está mediada por el dominio catalítico. En particular, el dominio catalítico de las proteínas de tipo BSL1 es necesario y suficiente para la asociación, mientras que en BSL2 los dominios catalíticos son necesarios, pero no suficientes y requieren del dominio conector. Esta particularidad explica los resultados obtenidos en la búsqueda de Y2H: todos los clones recuperados incluían porciones del dominio conector. Coincidiendo con las evidencias filogenéticas y funcionales mostradas en el Capítulo I, BSU1, el miembro divergente de la familia, no se comporta de la misma forma que sus homólogos conservados, o si lo hace es con características diferentes.

En el caso de BSL2, se determinó que para que la interacción tenga lugar se requieren al menos dos dominios P y un dominio L. Los ensayos de CEM establecieron, además, que BSL2 se presenta en forma de dímero. El dominio L aislado o asociado al extremo C-terminal del dominio K no fue capaz de mediar asociación (Figuras 2.5 y 2.6), lo que sugiere que la configuración dominio L-dominio P es importante, probablemente por requerimientos conformacionales en el extremo C-terminal del dominio L derivados de su unión con el dominio P. El dominio L parece actuar en *trans* sobre la otra subunidad P, de acuerdo al

resultado obtenido con la proteína quimérica formada por el dominio L de BSL2 unido a la PP1 TOPP1 enfrentada al dominio P aislado de BSL2 (Figura 2.10). En este caso, a pesar de que los dominios catalíticos no son capaces de interactuar entre sí, la presencia del dominio L hace posible la interacción, que es más débil que en el caso de BSL2 consigo misma. Esta última observación sugiere que los dominios P también participan de la interacción, no sólo como plataformas para la unión del dominio L, sino posiblemente formando, en la interfase entre ambos, alguno de los puntos de unión del dominio L.

La insuficiencia de los dominios P de BSL2 para mediar la interacción permitió llevar a cabo un análisis detallado del dominio L. Se determinó que para que la interacción tenga lugar son necesarias dos regiones no contiguas de este dominio, las secuencias 500-528 y 600-642, según la numeración de BSL2. Dentro de un dominio que posee bajo nivel de conservación de secuencia en el contexto de la proteína completa, ambas zonas contienen regiones conservadas entre las proteínas de tipo BSL2. En la medida en que el ensayo de interacción usando el sistema Y2H refleje la afinidad de asociación, ambas regiones actuarían de forma cooperativa, ya que la eliminación de una u otra región redujo la capacidad aparente de asociación. Este comportamiento es usual en proteínas que se unen a PP1, en las que motivos cortos de asociación actúan de forma sinérgica, como por ejemplo en Inhibitor-3 (Inh3 (Zhang et al., 2008)) o PNUTS (Choy et al., 2014). En Inh3, por ejemplo, mutaciones en el sitio RVxF (KKVEW en este caso) aumentaron unas 1000 veces la IC_{50} (concentración inhibitoria al 50%), mientras que mutaciones en un segundo sitio localizado a unos 25 residuos del anterior la hizo aumentar unas 200 veces (Zhang et al., 2008). Esta proteína sirve de ejemplo de la mayoría de las subunidades regulatorias de PP1 estudiadas a nivel estructural, en las que las zonas de interacción están separadas por relativamente pocos aminoácidos. Un caso diferente es el de Inhibitor-2 (I-2), una proteína que, con 205 aminoácidos (en la proteína humana), tiene una longitud similar al dominio L de BSL2. En I-2 dos de las tres regiones de interacción con PP1 están separadas por unos 70 aminoácidos y la zona intermedia se mantiene desestructurada (Hurley et al., 2007). Las regiones de asociación determinadas en el dominio L de BSL2 también están separadas por unos 70 o más residuos; la zona intermedia incluye a la única región del dominio que se predice adoptaría con alta probabilidad una estructura secundaria de tipo hélice α . Esta región no

parece participar de la interacción y probablemente cumpla otras funciones. Resta determinar si, en las dos regiones identificadas, de 29 y 42 residuos, todos ellos o sólo algunos participan de la interacción.

En el caso de BSL1, los dominios catalíticos son suficientes para que la interacción tenga lugar, por lo que en el contexto de esta proteína sólo puede especularse sobre la función del dominio L. Un posible papel emergería durante la interacción de BSL1 con alguna/s subunidad/es regulatoria/s, del modo análogo a lo que ocurre en la interacción entre PP1, GADD34 e Inhibitor-1, donde el complejo es estabilizado por interacciones entre las dos subunidades regulatorias (Connor et al., 2001). Mediante el uso de quimeras entre el dominio P de BSL2 y los dominios L de BSL1 y BSU1 se pudo comprobar que, comparados con el dominio L de BSL2, el de BSL1 es menos efectivo y el de BSU1 prácticamente incapaz de mediar interacción. Se pueden plantear dos posibles argumentos, no necesariamente excluyentes. Por un lado, los dominios L podrían mostrar especificidad por sus propios dominios P a través de cambios de aminoácidos correlacionados en ambas regiones. Por otro, analizando la Figura 2.8 en el contexto de las deleciones del dominio L de BSL2, se observa que la secuencia equivalente a 600-642 en BSL2 está conservada en BSL1 y hasta cierto punto en BSU1, mientras que la secuencia equivalente a 500-528 tiene un grado de conservación, si es que puede describirse así dadas las diferencias que existen, mucho menor en BSL1 y no existe como tal en BSU1. Las diferencias aparentes de afinidad de los dominios L en el contexto de la interacción entre dominios P de BSL2 podrían deberse, entonces, a la ausencia de una de las dos regiones que actúan de forma cooperativa.

Los perfiles de elución por CEM de BSL1P y LP no coincidieron con los esperados para dímeros. En el caso del dominio P aislado, que eluye en una forma intermedia entre monómero y dímero, puede que sufra disociación durante la cromatografía. Alternativamente, es posible que los dominios P de BSL1 se asocien con otra/s proteína/s en hojas de *N.benthamiana*, de forma que aumenta su tamaño aparente. En ese caso, el dominio P aislado estaría eluyendo como monómero. No es inusual que proteínas que forman complejos en solución se disocien en sus formas monoméricas en CEM. Un ejemplo de éstos es la 2-Cys Peroxirredoxina que, especialmente en el estado oxidado, forma un pentámero de dímeros en solución que se disocia en sus dímeros constituyentes en la

cromatografía (Konig et al., 2013). Otra posibilidad es que la forma LP adopte una conformación más abierta que BSL2, de forma que su radio hidrodinámico es más grande. Esta hipótesis tendría sustento a la luz de lo discutido en el párrafo anterior, si el dominio L de BSL1 se uniera a su dominio P a través de una sola de las zonas de interacción definidas en BSL2.

La evidencia de interacciones homotípicas lleva a la pregunta de cuál es la relevancia funcional de este comportamiento. Por más que las proteínas de tipo BSL1 y BSL2 utilicen mecanismos diferentes, en ambas el resultado final es la formación del complejo. No hay evidencias concretas de que otras fosfatasa de tipo PPP se asocien consigo mismas de forma estable, por lo que este comportamiento debe estar asociado a mecanismos regulatorios. El dominio L puede ser un elemento necesario sólo para la interacción, o actuar como una porción regulatoria interna. Por otra parte, los complejos pueden no ser permanentes, agregando así otra dimensión regulatoria a estas proteínas. Por otra parte, en trabajos previos (Gustavo Maselli, tesis de licenciatura), se determinó que las distintas isoformas de las proteínas BSL pueden formar asociaciones mixtas, es decir, combinando subunidades de tipos diferentes. Aunque es difícil abordar el estudio de estos casos y evaluar su relevancia fisiológica (puede tratarse de efectos artificiales de la sobreexpresión), podría constituir otro nivel regulatorio, si estos complejos mixtos presentan propiedades diferenciales.

Para responder estas preguntas se deberían hacer ensayos comparativos de actividad *in vitro* del dominio P aislado o asociado a sus dominios L. También se podrían llevar a cabo estudios detallados de las dos zonas del dominio L de BSL2 que se determinaron importantes para la interacción, con el fin de identificar los residuos involucrados. Una vez definidos, se podría mutagenizarlos sobre la forma completa de la proteína y, habiendo comprobado que anulan la interacción, introducir esta forma en el sustrato mutante *bs12 bs13*, para comprobar si es capaz de restaurar el fenotipo silvestre.

De hecho, se intentó, aunque sin éxito, definir algunos de estos elementos. En la región 600-642 se encuentran dos características potencialmente interesantes. Una de ellas es la presencia de un residuo Ser (627 en BSL2) conservado que se identifica fosforilado en trabajos de fosfoproteómica global en diferentes especies e isoformas (incluyendo BSL1 y en

alveolados). La introducción de un residuo fosfomimético en esta posición no afectó sin embargo la capacidad de interacción, aún en el contexto de la forma sensibilizada $\Delta 450-599$. En los relevamientos de fosfopéptidos mencionados se encuentran otros residuos fosforilados en las proteínas PPKL, en particular, en la extensión C-terminal variable de estas proteínas. No se puede descartar que estas modificaciones, en combinación con el residuo Ser en cuestión, puedan tener un efecto sobre la asociación. Sin embargo, no se obtuvieron evidencias de que la extensión C-terminal cumpla un papel en la interacción entre dominios P en BSL1 (Figura 2.18). Por otro lado, no en todos los casos las mutaciones fosfomiméticas simulan adecuadamente la presencia de grupos fosfato. Por ejemplo, la proteína 4E-BP2 se une al factor de iniciación de la traducción eucariótico eIF4E, impidiendo el reclutamiento de ribosomas dependiente del cap del mRNA. Cuando 4E-BP2 es fosforilada en varias posiciones sufre un cambio conformacional y se disocia de eIF4E. En estudios de interacción *in vitro* se observó que la sustitución de dos residuos Thr que son fosforilados por variantes fosfomiméticas no fue capaz de recapitular el comportamiento de la proteína fosforilada (Bah et al., 2015), por lo que no se puede descartar que éste sea el caso en BSL2.

La otra característica de la porción 600-642 del dominio L es la secuencia con similitud al motivo RVxF de las proteínas que interaccionan con PP1. Mutaciones en esta secuencia afectan severamente la capacidad de unión de estas proteínas a PP1 (por ejemplo, en el caso de PNUTS (Choy et al., 2014)). Las proteínas PPKL tienen alteraciones características en los residuos que bordean al surco RVxF respecto de otras fosfatasas PPP. En las subunidades regulatorias de PP1, el aminoácido aromático terminal del consenso (Trp o Phe) se aloja en la cavidad formada en PP1 por los anillos de dos Phe enfrentadas y una Leu, y cubierta por una Arg expuesta al solvente {Gibbons, 2005 #777}. Si bien buena parte de los aminoácidos hidrofóbicos que bordean el surco RVxF están conservados, al menos en naturaleza, en las fosfatasas PPKL, la dos Phe enfrentadas y la Arg están reemplazadas por Arg, Pro y Gly respectivamente. Esto altera significativamente las características de la zona, quizás como un mecanismo para excluir a las subunidades regulatorias de PP1. Esto llevó a suponer que la secuencia RRVS Ψ , altamente conservada en las proteínas de tipo BSL2, pero no en BSL1, podría ser un motivo de tipo RVxF, donde el residuo aromático está reemplazado por Tyr, quizás respondiendo a la naturaleza más polar del surco. Sin embargo, tampoco se pudo

comprobar que esta secuencia tenga un papel en la interacción. Hay que tener en cuenta que estos ensayos sólo se llevaron a cabo en levaduras y que habría que corroborarlos en otros sistemas. En todo caso, por el momento serían resultados “en negativo”, que identifican residuos que *no* participan de la interacción.

Otra pregunta pendiente es entender la razón por la que los dominios catalíticos de BSL1 son capaces de interactuar entre sí, mientras que los de BSL2 no. En ambas claramente hay elementos diferenciadores, ya que un dominio P de BSL1 no es capaz de interactuar con uno de BSL2. Analizando las secuencias de ambos tipos de proteínas se definieron dos posibles zonas que concentran la mayoría de los cambios de aminoácidos: en la región N-terminal, la superficie opuesta al sitio activo que incluye varias hélices α anfipáticas, y una zona adyacente al surco RVxF en la región C-terminal. A partir de estas observaciones y de las predicciones de estructura del dominio se generaron proteínas quiméricas entre mitades recíprocas de BSL1 y BSL2. Para evitar en lo posible disrupciones conformacionales, se decidió ensamblar las secuencias en una región idéntica en ambas proteínas que separa lo más claramente posible las regiones variables y que incluye, en la mitad N-terminal, las hélices superficiales, así como también una hélice más interna sobre la que las primeras se empaquetan. Estas proteínas quiméricas no arrojaron, sin embargo, resultados clarificadores, ya que ambas mostraron menor afinidad por BSL1 que BSL1 por sí misma. Dos residuos que bordean el surco RVxF y que son de naturaleza diferente entre BSL1 y BSL2 quedaron incluidos en la porción N-terminal de las quimeras. Sin embargo, esta diferencia de secuencia sólo se da en las proteínas de Brassicaceas, mientras que en el resto de las angiospermas analizadas la secuencia es tal como se encuentra en proteínas de tipo BSL2. De hecho, aunque en esta tesis sólo se analizaron las proteínas de Arabidopsis, en el laboratorio se comprobó que el dominio P de BSL1 de tomate (cuya secuencia en esta zona en particular es como la de BSL2) también es capaz de asociarse consigo mismo. Por todo lo expuesto, no es posible emitir por el momento una hipótesis sobre las razones del fracaso de estas quimeras, si no es por una alteración conformacional. Esta puede deberse tanto a cambios en el núcleo hidrofóbico como a alteraciones en residuos superficiales que afectan la interacción o por algún tipo de influencia del extremo C-terminal (que también es variable entre ambas proteínas) sobre el resto de la proteína. Una posible estrategia sería

mutagenizar residuos candidatos, particularmente superficiales, y ensayar su efecto sobre la interacción.

Aunque en este trabajo no ahondamos en las funciones del dominio Kelch, se pueden mencionar algunos datos. Los dominios de tipo turbina β usualmente actúan como plataformas de anclaje de otras proteínas. Es interesante destacar que se comprobó, mediante aproximaciones independientes, que el dominio K no participa en las interacciones homotípicas de las fosfatasas BSL. Datos de la bibliografía indican que este dominio interaccionaría con proteína/s involucrada/s en el control del ciclo celular. En arroz, las variantes alélicas en el gen GL3.1, que codifica para una de las proteínas de tipo BSL2/BSL3 y que definen el QTL para longitud de grano, se localizan en el dominio K, y se encontró que este dominio interacciona con la ciclina T1;3 (Qi et al., 2012). En *Chlamydomonas*, la mutante *div44-1*, que tiene afectada la progresión del ciclo celular, presenta una mutación en el dominio K (Tulin and Cross, 2014). Por otro lado, la localización subcelular de BSL1 estaría determinada, al menos parcialmente, por el dominio K. En formas fusionadas a GFP, la proteína completa se localiza en la periferia celular, pero es excluida del núcleo, mientras que los dominios P aislados se encuentran tanto en citoplasma como en núcleo (comparar Figuras 1.9 y 2.16). Por ende, el dominio K interaccionaría con proteínas sustrato y contribuiría a mantener la localización subcelular de estas proteínas.

Capítulo III

Análisis estructural del dominio conector de BSL2

Introducción

Proteínas intrínsecamente desordenadas

Un dogma de la bioquímica es que la función de una proteína correlaciona directamente con su estructura tridimensional. Sin embargo, en los últimos años se ha descubierto que aproximadamente 25% de todas las proteínas eucarióticas contienen regiones desordenadas de significativa extensión. Estas regiones pueden abarcar la proteína entera (en cuyo caso se denominan Intrinsically Disordered Proteins, IDPs) o ciertas porciones de la proteína, en la forma de regiones adyacentes a uno o múltiples dominios estructurados o conectores flexibles entre dominios con estructura. Las regiones desestructuradas pueden ser identificadas usando herramientas bioinformáticas ya que están enriquecidas en ciertos aminoácidos, particularmente Asn, Gln, Ser, Thr, Gly, Ala y Pro. Además, típicamente contienen pocos residuos hidrofóbicos, como Trp, Val, Leu, Ile, Phe, y Tyr, que constituyen, generalmente, el núcleo de las proteínas plegadas. La naturaleza extremadamente dinámica de las IDPs impide su descripción como una estructura única y rígida, y en cambio se describen como conjuntos conformacionales probabilísticos (Uversky, 2011).

Las IDPs tienen múltiples propiedades diferenciales sobre las proteínas estructuradas. Por un lado, pueden asociarse con sus interactores, a menudo proteínas estructuradas, en una conformación ampliamente extendida. Estos eventos de unión pueden resultar en nuevas, y a veces inesperadas, superficies de interacción, que son más extensas que lo que comúnmente se ve en las interacciones entre proteínas plegadas. Por eso, las IDPs requieren menos residuos de unión que las proteínas estructuradas, con un área de superficie idéntica. Por otro, su flexibilidad intrínseca permite a una sola IDP unirse a múltiples proteínas, ya que le permite adaptarse a diferentes superficies. Finalmente, la conformación desplegada otorga a las IDPs un extenso radio de captura de sus objetivos, permitiéndoles iniciar interacciones proteína-proteína más eficientemente que sus contrapartes plegadas (Mittag et al., 2010).

Las IDPs pueden asociarse con sus interactores de diferentes modos. En algunas instancias la reacción de unión está acoplada al plegamiento de la IDP sobre la superficie de la proteína a

la que se une, en un mecanismo denominado de “ajuste inducido” (*induced fit* en inglés). Alternativamente, la reacción de unión puede proceder a través de “selección conformacional”, por la cual ciertos conformeros de la IDP presentes en solución, que contienen elementos estructurales que se asemejan a los del estado unido, son seleccionados durante la interacción (Figura 3.1) (Oldfield and Dunker, 2014; Arai et al., 2015). Con todo, las IDPs pueden mantener niveles altos de flexibilidad aun estando unidas a una o múltiples proteínas diana, permitiéndoles ejercer funciones de nucleadores de complejos multiproteicos.

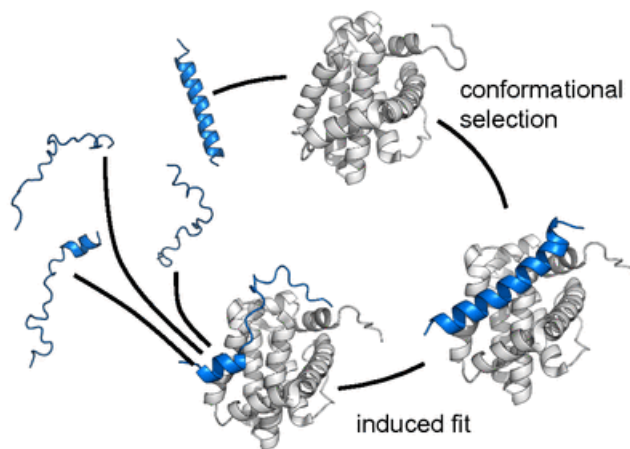


Figura 3.1 - Ejemplo de los posibles mecanismos de unión entre una IDP y una proteína estructurada. Tomado de (Rogers et al., 2014)

Regulación de la actividad de fosfatasa PPP por proteínas intrínsecamente desordenadas

La diversidad funcional de las fosfatasa de Ser y Thr de tipo PPP no se logra a través de procesos de duplicación génica seguidos de divergencia, como sucede en el caso de las quinasas de Ser y Thr, sino por su habilidad para interactuar con numerosas subunidades regulatorias. Se estima que PP1 de mamíferos, la fosfatasa de tipo PPP más estudiada, sería capaz de asociarse con más de 200 proteínas (Heroes et al., 2013). A pesar de que sólo se ha analizado en detalle una minoría de estos posibles complejos, han surgido algunos temas recurrentes. En particular, predicciones y estudios estructurales permiten suponer que aproximadamente dos tercios de estas subunidades regulatorias serían desordenadas en al menos parte de su secuencia, y en particular, en la región de unión a PP1 (Bollen et al., 2010).

Hasta el momento se han obtenido las estructuras cristalográficas de cinco complejos entre PP1 y otras tantas subunidades regulatorias (Figura 3.2). PNUTS (PP1 nuclear targeting subunit) (Choy et al., 2014) y NIPP1 (nuclear inhibitor of PP1) (O'Connell et al., 2012) son componentes importantes de los complejos que PP1 forma en el núcleo, donde regula la descondensación de cromatina y la transcripción y controla el estado de fosforilación de componentes centrales del ciclo celular. Spinofilina es una proteína con funciones de andamiaje que dirige a PP1 a las terminales post-sinápticas, donde regula la dinámica del citoesqueleto y la actividad de receptores de glutamato (Ragusa et al., 2010). MYPT1 (myosin phosphatase targeting subunit 1) forma junto con PP1 el complejo “fosfatasa de miosina” que regula los niveles de fosforilación de miosina II y con ello la contracción del músculo liso y la dinámica del citoesqueleto (Terrak et al., 2004). Estas son proteínas con varios dominios, entre los que está el de unión a PP1, que en todas resulta ser una secuencia de tipo intrínsecamente desordenada. Finalmente, I-2, uno de los principales reguladores de la actividad de PP1, es una proteína pequeña sin dominios discernibles, caracterizada por su alta termoestabilidad y amplio radio hidrodinámico en solución (Hurley et al., 2007).

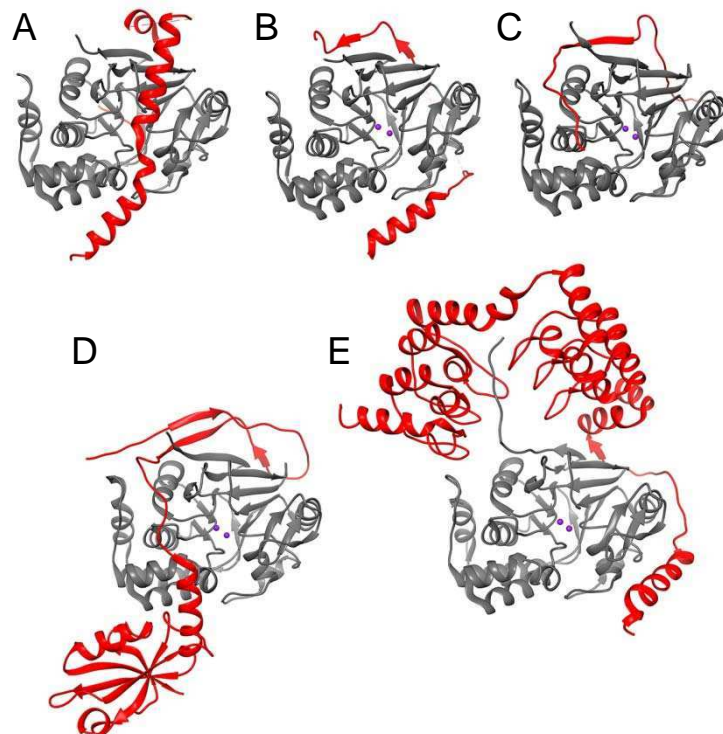


Figura 3.2 (página anterior) - Complejos entre PP1 y subunidades regulatorias. En cada complejo PP1 se muestra en gris, siempre en la misma orientación, con la cara del centro activo al frente. En rojo, se muestran las subunidades regulatorias: A) I-2 (PDB 2O8A), B) NIPP1 (PDB 3V4Y), C) PNUTS (PDB 4MOY), D) spinofilina (PDB 3EGG) y E) MYPT1 (PDB 1S7O). Figura hecha con Chimera V1.10.

En todos estos casos se observa que la subunidad regulatoria establece varios puntos de contacto a lo largo de la superficie de PP1, de forma que ésta es “abrazada” por su interactora. Estos motivos de anclaje son diversos y forman secuencias consenso laxas, por lo que las distintas variantes muestran considerables diferencias en su afinidad por PP1. Todas estas proteínas comparten el consenso RVxF, pero cada subunidad se asocia, además, con diferentes regiones de PP1 y a través de distintos determinantes, y con ello ejercen su acción de distinta manera. PNUTS y spinofilina ocupan el surco C-terminal, mientras que NIPP1 y MYPT1 se alojan en el surco hidrofóbico. En estos casos, si bien el centro activo no resulta directamente bloqueado, la ocupación de los surcos y el cambio de las propiedades electrostáticas alrededor del centro activo modulan la accesibilidad y afinidad por sustratos. I-2, en cambio, bloquea directamente el sitio activo. Las regiones de interacción de estas subunidades con PP1 rondan los 60 aminoácidos y cuando son estudiadas en solución se observa que forman, con diferente estabilidad, estructuras secundarias transitorias equivalentes a las encontradas en el complejo con PP1. Por otra parte, mientras que las zonas de unión de NIPP1 y MYPT1 se encuentran fijas en el complejo, la de spinofilina mantiene mayores niveles de flexibilidad, y en I-2 sólo un cuarto de la secuencia muestra una estructura definida, mientras que el resto permanece indeterminado. La región que permanece desestructurada modula, sin embargo, las propiedades de la interacción: la fosforilación en una Thr en esa zona modifica la conformación de la zona plegada de I-2 unida a PP1 de forma que se libera el bloqueo del sitio activo sin que por ello se disocie el complejo (Cannon, 2013). La flexibilidad de estas proteínas jugaría, además, un papel más allá de propiciar superficies extensas de interacción con PP1: las regiones flexibles residuales podrían formar sitios adicionales de unión para otras proteínas (Choy et al., 2012). Por otro lado, la ocupación de distintos sitios sobre la superficie de PP1 permite que se formen complejos ternarios, como el que se observa entre PP1, spinofilina e I-2 (Dancheck et al., 2011).

Objetivos

En el capítulo anterior se realizó un análisis del dominio conector de las PPKL en relación a su capacidad de mediar las interacciones homotípicas que presentan estas proteínas. Para ahondar en el estudio de esta región, se abordó la caracterización estructural del dominio conector de BSL2.

Resultados

Predicción de la estructura del dominio conector de BSL2.

Para abordar la caracterización estructural del dominio conector de BSL2, se analizó en primera instancia la información que puede deducirse a partir de su composición aminoacídica. Los análisis basados en composición son simples y generales: se basan en las propiedades de cada aminoácido y en las frecuencias de los mismos en proteínas de conformación conocida. En la Figura 3.3 los aminoácidos se ordenaron según su frecuencia de aparición en regiones ordenadas o desordenadas del dominio, y se evaluó la diferencia de abundancia de cada uno de ellos respecto de bases de datos. La región conectora tiene baja representación de aminoácidos promotores de orden (en especial aromáticos), mientras que está enriquecida en residuos promotores de desorden, en particular aquellos más frecuentes en zonas desordenadas como Ser, Pro y Asp (los del extremo derecho de la escala).

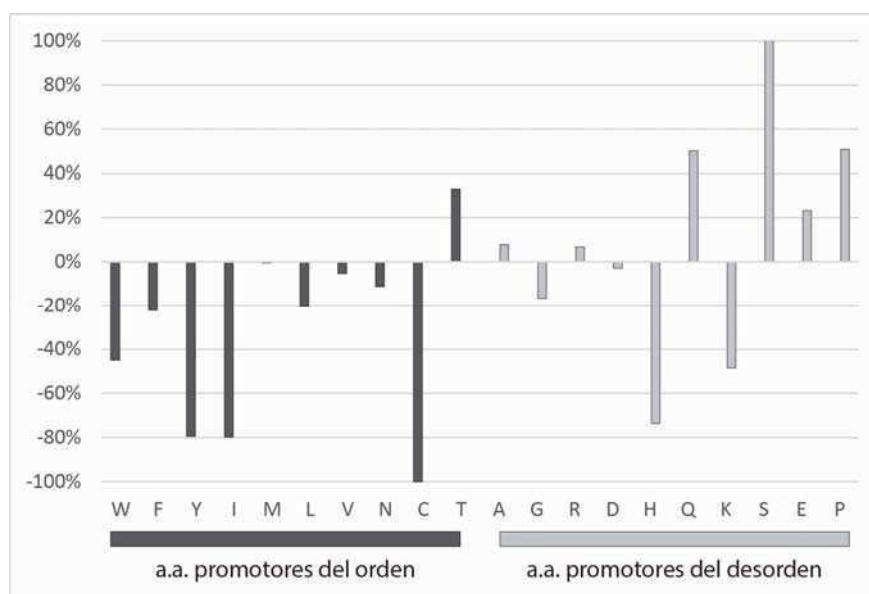


Figura 3.3 - Análisis de composición aminoacídica del dominio conector. Desvío de la composición aminoacídica del dominio conector frente a la composición aminoacídica de la base de datos de Swiss-PROT. Se muestra el enriquecimiento relativo de los aminoácidos (a.a.) promotores de orden y de desorden. Los residuos se ordenaron en el eje x según el índice de flexibilidad TOP-IDP descrito en (Campen et al., 2008; Habchi et al., 2010).

En la Figura 3.4, las proteínas se agrupan según la relación entre su carga neta y su hidrofobicidad: las proteínas ordenadas tienden a tener núcleos hidrofóbicos y menor carga

neta, mientras que proteínas de conformación desordenada son más polares y cargadas. El dominio L de BSL2 se dispuso en la región enriquecida en proteínas ordenadas.

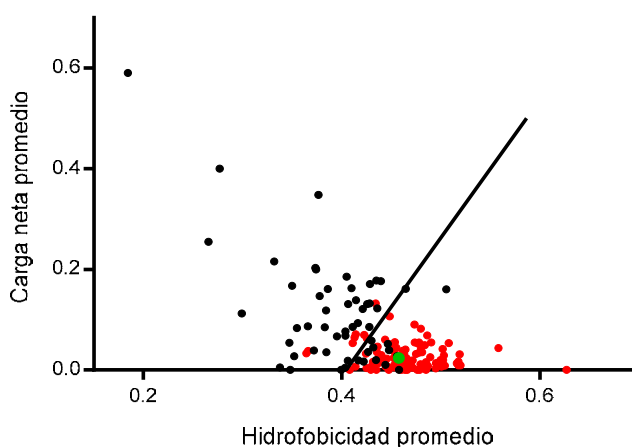


Figura 3.4 - Gráfico de carga vs hidrofobicidad (CH plot). Proteínas de estructura conocida se disponen según la relación entre su carga neta y su hidrofobicidad, computada siguiendo la escala de Kite-Doolittle. Los puntos negros corresponden a proteínas desordenadas, y los rojos a proteínas ordenadas; la diagonal es el límite empírico entre ambas clases (Uversky, 2002). El punto verde corresponde al dominio conector de BSL2. Calculado usando PONDR® (<http://www.pondr.com>).

En paralelo, se determinó la probabilidad de regiones desordenadas en la secuencia completa de BSL2 mediante tres métodos independientes: CSpritz (<http://protein.bio.unipd.it/cspritz>) (Walsh et al., 2011), MFDp2 (<http://biomine-ws.ece.ualberta.ca/MFDp2>) (Mizianty et al., 2014) y DISOPRED3 (<http://bioinf.cs.ucl.ac.uk/psipred>) (Jones and Cozzetto, 2015). Estos métodos usan diferentes combinaciones de información proporcionada por la composición aminoacídica y por la predicción de estructuras secundarias, usando bases de datos de proteínas desordenadas como referencia, para inferir las probabilidades de desorden de la proteína en cuestión (Figura 3.5).

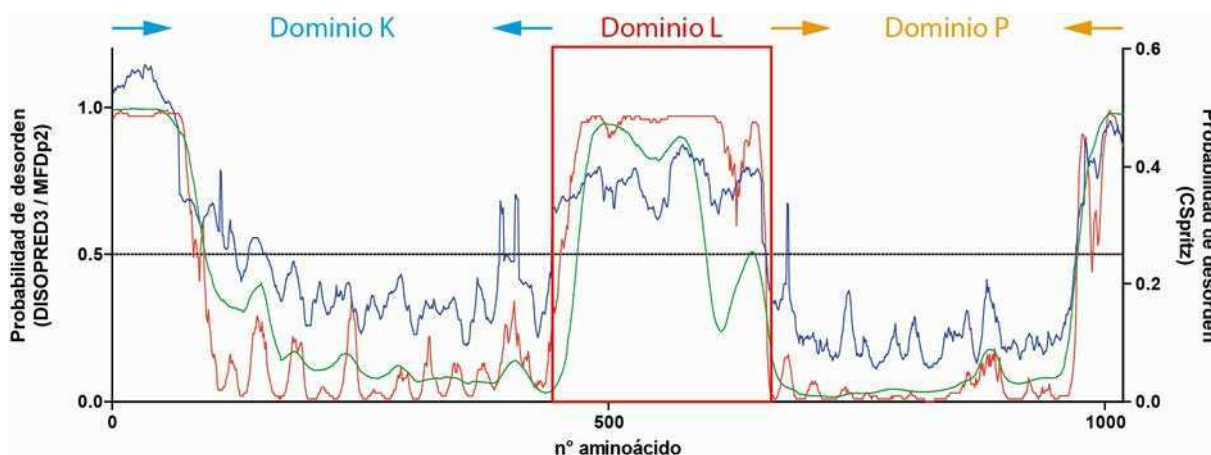


Figura 3.5 - El dominio conector de BSL2 se predice como una región intrínsecamente desordenada.

Gráfico de probabilidad de desorden en la secuencia completa de BSL2 calculada con los métodos DISOPRED3 (línea roja), CSpritz (línea azul) y MFDp2 (línea verde). Las regiones del gráfico correspondientes a los dominios están delimitadas por flechas. Los ejes de ordenadas tienen distintas escalas con el fin de hacer coincidir los valores de umbral que separa regiones estructuradas de regiones desordenadas en los diferentes métodos (línea horizontal).

Como se observa, la información estructural inferida a partir de la secuencia predice que, además de los extremos N- y C-terminal, el dominio conector adoptaría con alta probabilidad una conformación de tipo desordenada.

Purificación y caracterización del dominio conector de BSL2.

Para poder llevar a cabo el análisis estructural del dominio conector, se introdujo la secuencia del mismo (aminoácidos 460 a 662, 606 pb) en un vector para su expresión de forma recombinante en E.coli. En primera instancia se introdujo en el vector pHis8 (Jez et al., 2000) para generar una proteína fusionada a una secuencia de poli-His en su extremo N-terminal. La proteína así expresada se recuperó, bajo diferentes condiciones de inducción, de forma soluble, pero como un conjunto de fragmentos producto de degradación. Por el contrario, cuando se introdujo la secuencia en el vector pET22b+ (Novagen) para generar una proteína de fusión con poli-His en su extremo C-terminal, se logró expresar completa y de forma soluble. Sin embargo, se observó en ensayos preliminares que, en extractos crudos o parcialmente purificados, la proteína tenía una gran tendencia a ser degradada en múltiples fragmentos. Para evitar este problema, se diseñó un protocolo de purificación rápido en presencia de inhibidores de proteasas para evitar incubaciones prolongadas, como diálisis de varias horas.

Una característica de las IDPs es su termoestabilidad, debido a la ausencia de estructuras secundarias y núcleos hidrofóbicos definidos. Se encontró que al tratar el sobrenadante del lisado proveniente de células inducidas por 10 minutos a 80°C en presencia de inhibidores de proteasas, el dominio conector permanecía en la fracción soluble. A continuación, esta fracción se sometió a una columna HiTrap-IMAC con Ni²⁺ inmovilizado. Las fracciones de elución de la proteína fueron inmediatamente desaladas a través de columnas de exclusión molecular. El resultado del procedimiento de purificación se muestra en la Figura 3.6 A. La proteína así aislada resultó estable después de una incubación a 4°C durante toda la noche.

Dado que el dominio posee un residuo Cys aislado, existía la posibilidad de que esta Cys formara puentes disulfuro e impidiera una correcta caracterización. Cuando se analizó la proteína purificada por SDS-PAGE en presencia o ausencia de reductor se comprobó que, en efecto, una fracción de la misma forma dímeros (Figura 3.6 B). Para evitarlo, se trató a la proteína aislada con 10mM ditioneol, se bloquearon las Cys libres con iodoacetamida y ésta se eliminó a través de una columna de filtración molecular equilibrada en buffer de almacenamiento; la forma tratada con iodoacetamida no formó dímeros (Figura 3.6 B).

A continuación, se determinó la masa de la proteína purificada mediante Static Light Scattering (SLS), una técnica que evalúa la intensidad de la luz dispersada a partir de la excitación en distintos ángulos para deducir características de las partículas dispersantes, en particular su masa molecular. Al asociar la determinación con CEM, se observó que el dominio conector eluyó como un único pico, lo que indica que no se produjo agregación. La proteína presentó una masa molecular promedio (Mw) de 21.1 kDa, que coincidió con la esperada para la forma monomérica (22.04 kDa, Figura 3.6 C).

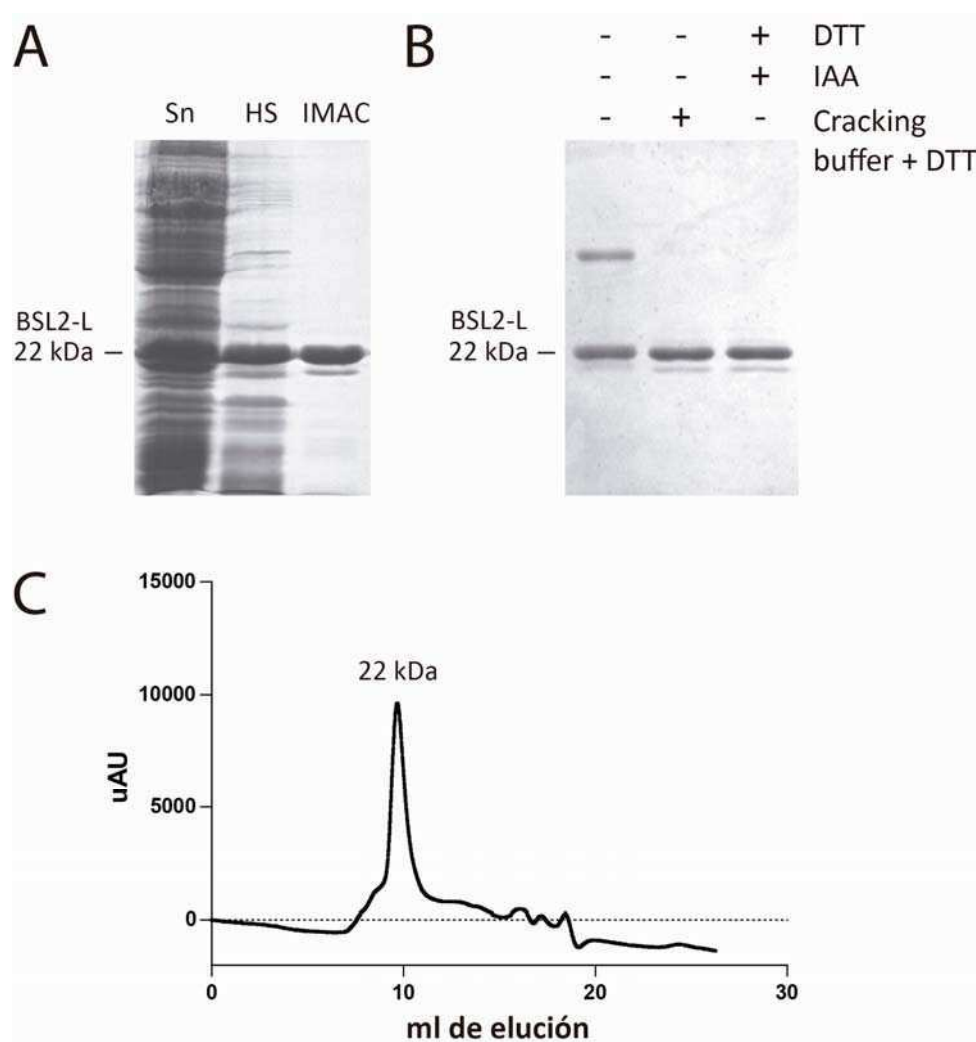


Figura 3.6 - Purificación y caracterización del dominio conector de BSL2. A) Gel de poliacrilamida (SDS-PAGE) 10% mostrando el estado de la proteína a lo largo del protocolo de purificación. Se muestran la fracción sobrenadante del lisado total (Sn), el sobrenadante resultante de someter esta misma fracción a 80°C por 10 minutos (HS), y las fracciones eluidas de la columna de IMAC. Tinción mediante Coomassie blue. B) Reducción y bloqueo de Cys libres. La proteína aislada fue tratada con ditioneitol (DTT) y con iodoacetamida (IAA) según se indica. Las proteínas fueron sometidas a electroforesis con buffer de carga en presencia o ausencia de DTT, según se indica. C) Patrón de elución en CEM y cálculo de la masa molecular por SLS.

También se determinó su radio hidrodinámico (Rh) mediante RMN (Diffusion Ordered Spectroscopy, DOSY) una técnica que evalúa el coeficiente de difusión en solución de una proteína, que depende de su tamaño, y lo relaciona con su masa molecular. Se comprobó así que, en solución, esta proteína adopta una conformación que se corresponde con una forma parcialmente desplegada (Tabla 3.1).

	Radio hidrodinámico
BSL2 L en solución	28.5 ± 0.6 Å
BSL2 L 100% plegada (teórico)	22.4 Å
BSL2 L 100% desplegada (teórico)	46.7 Å

Tabla 3.1 - Radio hidrodinámico (Rh) del dominio conector de BSL2 determinado por RMN y valores teóricos para una proteína de la misma cantidad de aminoácidos de conformación globular o totalmente desplegada, según (Wilkins et al., 1999). Rh de una proteína 100 % plegada: $4.75 * N^{0.29} = 22.39 \text{ \AA}$; Rh de una proteína 100 % desplegada: $2.21 * N^{0.57} = 46.67 \text{ \AA}$; N (N° aa) = 202.

De esta forma, se logró obtener el dominio conector de BSL2 de forma soluble y aislado. El hecho de que la proteína sea termotolerante pero, al mismo tiempo, altamente susceptible a la presencia de proteasas sugiere que en efecto se trata de una IDP. Apoya esta hipótesis la determinación de un radio hidrodinámico más extendido que el esperable para una proteína globular del mismo tamaño.

Propiedades estructurales del dominio conector de BSL2.

Una vez aislado el dominio conector de BSL2, se procedió a una caracterización estructural más detallada. Para ello se analizaron sus propiedades por Dicroísmo Circular (DC) en el UV lejano. La figura 3.7 A muestra el espectro de DC de la proteína aislada a 20°C. La característica más notable del espectro es que presentó un mínimo principal en la zona de 200 nm, lo que indica que en efecto se trata de un polipéptido sin componentes importantes de estructura secundaria. Con todo, se observó un mínimo local en la zona de 220 nm, que sugiere la existencia de cierto componente de estructura secundaria. Sin embargo, la intensidad de la señal a 222 nm (que se asume es proporcional al contenido de hélices α) no se modificó con un cambio gradual de la temperatura entre 20 y 80°C, lo que confirma la termotolerancia de esta proteína (Figura 3.7 B).

Para analizar el comportamiento del dominio conector frente a perturbaciones del medio, se determinaron los espectros de DC en presencia de concentraciones crecientes de trifluoroetanol (TFE, Figura 3.7 C). El agregado del cosolvente TFE induce la formación de estructuras secundarias, en general de tipo hélice α , en proteínas o péptidos en solución, probablemente por un efecto caotrópico (Buck, 1998). Se observó que concentraciones bajas (entre 0 y 15%) de TFE produjeron un cambio significativo en las características del

espectro, en particular, la aparición de un mínimo a 210 nm mientras que se mantuvo el mínimo de 220 nm, lo que sugiere la formación de estructuras de tipo hélice α . Esta sola observación no permite, sin embargo, afirmar que el dominio conector adopte estructuras helicoidales, sino sólo que su estructura es afectada por concentraciones bajas de TFE. Las concentraciones de TFE usadas se encuentran en el rango bajo de las usualmente empleadas en el estudio de regiones desordenadas, pero es importante considerar que muchos estudios se centran en secuencias lineales de unas pocas decenas de aminoácidos (como por ejemplo (Noval et al., 2013)), mientras que el dominio conector usado tiene 202 residuos. De hecho, proteínas completas como α -sinucleína, una IDP de 140 aminoácidos que en presencia de membranas o detergentes adopta una estructura helicoidal que se supone es la causa de patologías como Parkinson, muestran cambios espectroscópicos en el rango de concentraciones bajas de TFE (Anderson et al., 2010). Al mismo tiempo, se observó que estas concentraciones de TFE producen un cierto grado de precipitación en la proteína. Es necesario tener en cuenta que el dominio conector se trata de un polipéptido significativamente extenso con propiedades mixtas. Las IDPs pueden clasificarse en subtipos según la relación entre sus señales dicróicas a 200 nm y 222 nm; en particular, se destacan aquellas mayormente desplegadas con señal pronunciada a 200 nm y menos intensa a 222 nm (de tipo *random coil*) y aquellas que presentan cierta compactación y estructuras secundarias transitorias, con mayor señal a 222 nm y relativamente menor a 200 nm (de tipo *pre-molten globule*). El dominio conector de BSL2 se agrupa con estas últimas (Figura 3.7 D), coincidiendo con la evidencia obtenida de su composición aminoacídica y de su sensibilidad al tratamiento con TFE.

Si bien esta caracterización preliminar no permite extraer conclusiones definitivas sobre la estructura del dominio, puede afirmarse que es una proteína de tipo desordenada, con alta estabilidad térmica, pero que al mismo tiempo muestra alta sensibilidad frente a cambios relativamente sutiles del medio ambiente, lo que sugiere que puede estar sujeta a procesos de plegamiento local transitorios o inducidos por interacción con otras proteínas, en especial, con la misma BSL2.

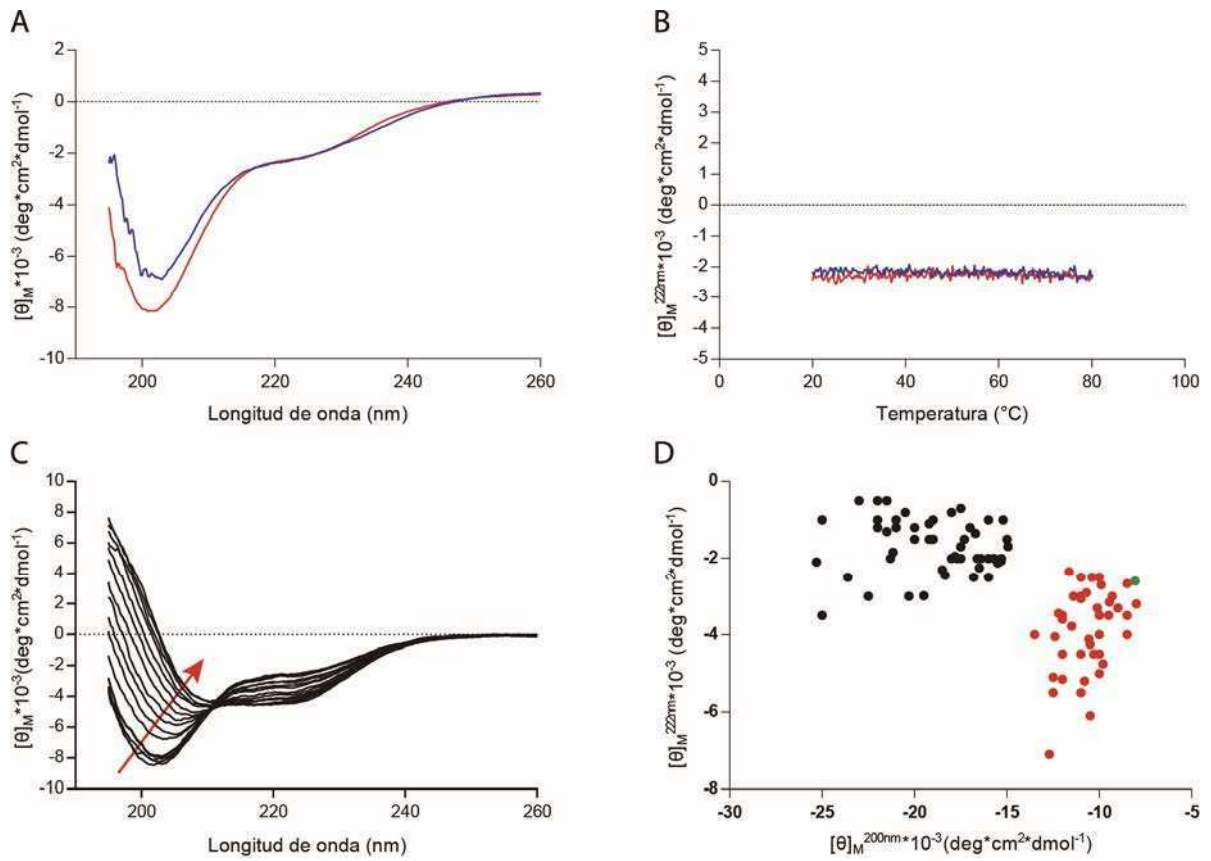


Figura 3.7 - Propiedades estructurales del dominio conector de BSL2. A) Espectro de DC en la región de UV lejano. En rojo, espectro a 20°C, en azul, a 80°C. B) Efecto de cambios graduales de temperatura en la señal dicroica a 222nm. En rojo, señal dicroica entre 20 y 80°C, en azul, reversión del gradiente entre 80 y 20°C. C) Cambios en el espectro de DC en presencia de cantidades crecientes de TFE entre 0 y 15% (según indica la flecha). D) Diagrama de elipticidad 200 nm vs. 222 nm. Los puntos negros corresponden a proteínas con conformación desplegada (*random coil*), los puntos rojos a proteínas parcialmente desplegadas con características de *pre-molten globule*. El punto verde corresponde al dominio conector de BSL2 (Uversky, 2002).

Conclusiones

En los últimos años se ha tornado evidente que proteínas o dominios sin estructura definida en solución participan y cumplen funciones centrales en redes de interacciones entre proteínas y, en particular, en complejos de señalización (Fong and Panchenko, 2010). Como ejemplo, la fosfatasa PP1, que participa en multitud de procesos en células eucarióticas, se asocia a gran cantidad de subunidades regulatorias cuya característica más saliente es que adoptan, o se predice que adoptan, conformaciones en solución de tipo intrínsecamente desordenadas que les permiten interactuar de forma extendida con el dominio catalítico (Choy et al., 2012). El dominio conector de BSL2, necesario para que la interacción homotípica en esta proteína tenga lugar, constituye una región que cumple con las condiciones para suponer que adopta una conformación de tipo desplegada. En efecto, este dominio:

- 1) es una región de secuencia variable que conecta dos dominios altamente conservados (Figura 2.11);
- 2) presenta una composición aminoacídica enriquecida en aminoácidos promotores de desorden;
- 3) es probable que adopte una conformación desordenada, a partir de lo predicho por distintos algoritmos de predicción de regiones de este tipo (Figura 3.5);
- 4) es una proteína termoestable y sensible a proteólisis, características de proteínas desplegadas sin núcleos hidrofóbicos colapsados (Figura 3.6);
- 5) tiene características espectroscópicas que sugieren que no presenta, aislado y en solución, estructuras secundarias estables significativas (Figura 3.7).

En este contexto, es necesario tener en cuenta que las proteínas PPKL presentan otras regiones que se predicen desordenadas: los extremos N- y C-terminales (Figura 3.5). Aunque hasta el momento no se pudo determinar un papel para ellos, hay evidencias indirectas que hacen suponer que cumplirían alguna función. El dominio N-terminal que precede al dominio Kelch propiamente dicho es particularmente extenso en proteínas de tipo BSL2 y BSL3 (86 aa en BSL2, asumiendo un comienzo tentativo del dominio Kelch) y difiere característicamente de las de tipo BSL1, donde no es reconocible una región con estas características. En

estudios de fosfoproteómica se identificaron residuos de Ser o Thr fosforilados en el extremo C-terminal de proteínas de tipo BSL1 y BSL2 (de la Fuente van Bentem et al., 2008; Rose et al., 2012), lo que sugiere un posible punto de regulación. De hecho, el extremo C-terminal de PP1 también es de tipo desordenado y sólo pudo determinarse su estructura en complejo con MYPT1 dado que esta proteína lo fija a través de su dominio TTP aun cuando mantiene su conformación extendida (ver figura 3.2 E) (Terrak et al., 2004). El extremo C-terminal participa en la regulación de la actividad de PP1. Al final de la mitosis se desencadena una cascada de defosforilaciones que llevan a la terminación del proceso de división celular. Estas defosforilaciones son mediadas por PP1 y ciertas isoformas de PP2 que durante la mitosis se mantienen reprimidas tanto por la presencia de subunidades inhibitorias como por fosforilaciones mediadas por quinasas dependientes de ciclinas. La fosforilación de una Thr en el extremo C-terminal de PP1 produce un efecto autoinhibitorio; cuando los niveles de ciclinas disminuyen, PP1 es capaz de auto-reactivarse defosforilando ese residuo para así iniciar la fase de terminación de la mitosis (Grallert et al., 2015). El extremo C-terminal de la subunidad catalítica de PP2 también adopta una conformación extendida; distintas modificaciones post-traduccionales en esta zona (en particular, metilación del extremo C-terminal y fosforilación en ciertas Thr y Tyr conservadas) generan un “código” que es importante para el ensamblado de la holoenzima y el reclutamiento de distintas subunidades regulatorias (Janssens et al., 2008). En general, las evidencias se acumulan indicando que los extremos flexibles de muchas proteínas son cruciales para la funcionalidad de las mismas (Uversky, 2013).

Las regiones conectoras entre dominios de proteínas multifuncionales tienden a adoptar también una conformación característicamente desplegada (Fong and Panchenko, 2010). Las evidencias mostradas sugieren que el dominio conector de BSL2 es una proteína de este tipo que, aunque presenta en solución características de una proteína intrínsecamente desordenada, es altamente sensible a perturbaciones ligeras del medio (Figura 3.7). Esta característica sugiere que podría adoptar estructuras plegadas al interactuar con el dominio catalítico u otras proteínas. Los cambios espectroscópicos asociados al tratamiento con concentraciones crecientes de TFE no permiten inferir qué tipo particular de estructuras secundarias se forman, sino sólo indican que se produce un cambio. Curiosamente, el *CH-*

Plot no mostró una fuerte tendencia al desorden (Figura 3.4) y dos de los métodos usados para predecir regiones desordenadas (Disopred3 y MFDp2, Figura 3.4) mostraron un descenso en la probabilidad de desorden hacia el final del dominio conector, significativamente, en la región que rodea a Ser627. Esta observación refuerza la idea de que la fosforilación de este residuo puede tener un papel estructural y/o funcional, a pesar de que, como se describe en el capítulo precedente, no se haya encontrado por el momento. Finalmente, los valores de elipticidad a 200 nm y 222 nm del dominio conector (Figura 3.7 D) muestran que éste presenta características similares a IDPs que adoptan una conformación de tipo *pre-molten globule*, es decir, conformaciones parcialmente colapsadas que no llegan a formar un núcleo hidrofóbico estable, con elementos de estructura secundaria transitorios (Chaffotte et al., 1997).

Una observación indirecta sobre las características de la región C-terminal del dominio conector es que la forma de la proteína con poli-His en su extremo C-terminal resultó más estable frente a degradación que cuando la etiqueta se encontraba en el extremo N-terminal. Esto sugiere que residuos adyacentes al extremo C-terminal podrían (por efectos de carga o por simple presencia) inducir algún tipo de cambio conformacional que aumentaría su estabilidad. Esta hipótesis se relaciona con la evidencia obtenida en el Capítulo II (Figuras 2.5, 2.6 y 2.10): el dominio L parece requerir un dominio proteico en su extremo C-terminal para ser efectivo. Sin embargo, no cualquier proteína tiene efecto ya que, como se muestra en la Figura 2.6, la fusión C-terminal de GFP a la porción KL no es capaz de mediar interacción, mientras que sí lo es el dominio L de BSL2 unido a una proteína relacionada con el dominio P como TOPP1 (Figura 2.10). En resumen, aunque la evidencia sugiere que el dominio L actúa en *trans* asociándose al dominio P de la otra subunidad, sería necesario un efecto conformacional en *cis* fruto de su asociación y/o interacción con la superficie del dominio de tipo P al que está unido por su extremo C-terminal. La combinación de ambas interacciones podría volver al dominio L competente para la asociación con el dominio P mediante un mecanismo de *induced fit*.

En la Figura 3.8 se muestra un modelo de BSL2 (sin los extremos N- y C-terminal de secuencia variable y probablemente desordenados), en el que el dominio conector se representa de forma extendida con el fin de enfatizar su tamaño en relación a los dominios

plegados. Aun cuando porciones del dominio conector se asocien y plieguen parcialmente sobre el dominio catalítico, a semejanza de I-2 sobre PP1, su longitud es suficientemente grande como para suponer que parte del mismo permanecerá libre, aunque constreñido por efecto de la asociación, y potencialmente disponible para otras interacciones. Recordando que la búsqueda de interactores que dio origen a este trabajo se llevó a cabo con el dominio catalítico de BSL2 aislado, y dada la evidencia de que el dominio conector es necesario para la interacción de esta proteína consigo misma y que este comportamiento puede extrapolarse a interacciones con otras proteínas, sería informativo llevar a cabo una nueva búsqueda incorporando este dominio.

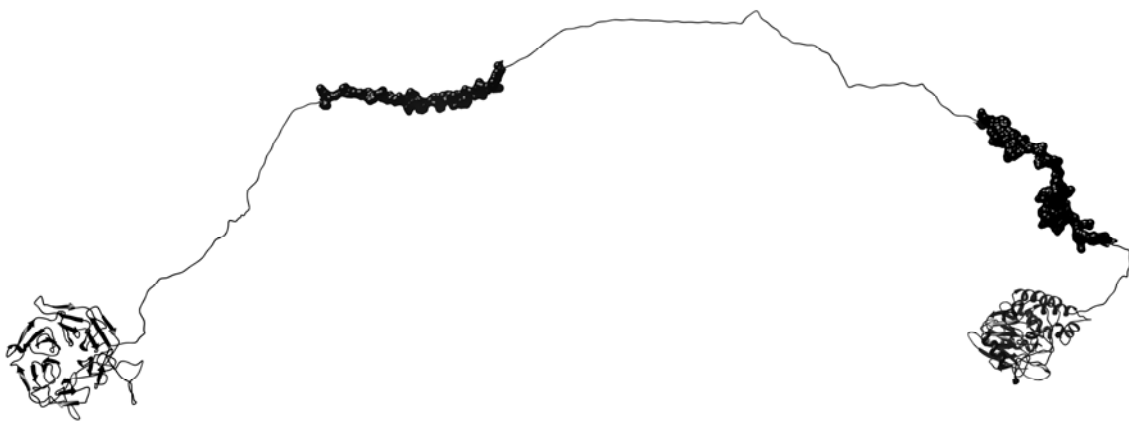


Figura 3.8 - Representación de la posible estructura de BSL2 con su dominio conector totalmente desplegado. El dominio Kelch (a la izquierda, sin el extremo N-terminal) y el dominio catalítico (a la derecha, sin el extremo C-terminal) fueron modelados en SwissModel tomando como base las estructuras de Keap1 murina (1x2r) y PP1- γ humana (1fjm). El dominio conector fue modelado por IntFOLD2 (http://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD2_form.html); los residuos de las dos regiones que se determinó son importantes para la asociación se representan en forma de esferas.

Contar con un procedimiento rápido y eficiente para obtener el dominio conector en estado puro abre las puertas para análisis estructurales más detallados, como la determinación de su conformación en solución mediante RMN y el estudio de las características espectroscópicas de las formas delecionadas o mutagenizadas descritas en el Capítulo 2. Una aproximación similar podría seguirse con el dominio conector de BSL1, para el que hasta el momento no se dispone de datos funcionales más allá de su aparente capacidad reducida de mediar asociación en el contexto de BSL2 (como se muestra en la Figura 2.9).

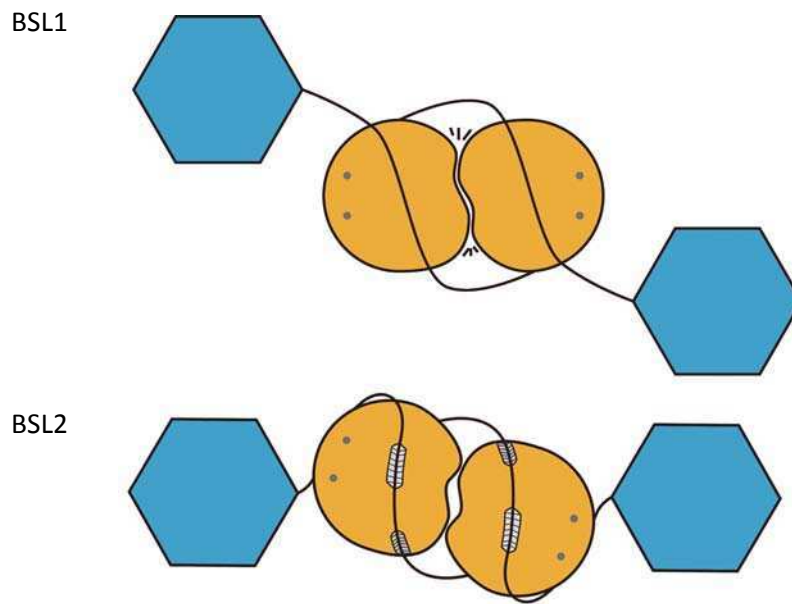
Conclusiones Generales

En conjunto, el trabajo presentado muestra que:

- 1) las fosfatasa de tipo PPKL cumplen en plantas una función esencial, que resta determinar;
- 2) en plantas terrestres existen dos tipos principales de fosfatasa PPKL que difieren en caracteres diagnósticos de secuencia y en sus patrones de expresión y localización subcelular;
- 3) estas fosfatasa forman asociaciones homotípicas, probablemente en forma de dímeros, un comportamiento no reportado en otras fosfatasa de la familia PPP;
- 4) la asociación depende de la presencia de dominios catalíticos y, condicionalmente, del dominio conector adyacente; el dominio Kelch no participa de la asociación;
- 5) cuando el dominio conector es necesario para que tenga lugar la interacción, se requieren dos regiones conservadas no contiguas del mismo;
- 6) este dominio conector es una región de tipo intrínsecamente desordenada que presenta tendencia a cambios conformacionales ante perturbaciones débiles del medio.
- 7) lo que a lo largo de esta tesis se llamó dominio conector, nombre que en cierta forma implica un papel pasivo de simplemente unir dos dominios plegados, es con derecho propio un dominio funcional adicional que puede tener un papel clave en la regulación de la actividad de estas proteínas.

Las conclusiones de tipo estructural pueden resumirse en el siguiente modelo:

Conclusiones Generales



Modelo de la interacción homotípica de las proteínas BSL

El dominio Kelch se representa en turquesa y el catalítico en naranja; la posición relativa de los dominios y la disposición del dominio conector de BSL1 cumplen sólo fines de representación. Los dos puntos grises indican los átomos metálicos del centro activo y las zonas blancas rayadas sobre los dominios catalíticos de BSL2 representan zonas de interacción entre los dominios conectores y los dominios catalíticos.

Procedimientos Experimentales

1) Material vegetal y condiciones de crecimiento

Las líneas de inserción de T-DNA de *Arabidopsis thaliana* se obtuvieron de ABRC (<http://www.abrc.osu.edu>). Todas las líneas de T-DNA se retrocruzaron dos veces con el sustrato silvestre Col-0.

Las semillas de *Arabidopsis* se esterilizaron con etanol 70% / Triton X-100 0,05%, se sembraron en placas de medio MS 0,5x con 0,8% de agar, se estratificaron por 3 días a 4°C y se indujo la germinación con un pulso de 1 hora de luz blanca. Plántulas de 10 días se transfirieron a macetas conteniendo una mezcla 1:1:1 de vermiculita, perlita y turba, y se crecieron bajo condiciones de día largo (16 horas de luz: 8 horas de oscuridad) a 22°C.

Las plantas de *Nicotiana benthamiana* se crecieron en macetas conteniendo una mezcla 1:1:1 de vermiculita, perlita y turba, en condiciones de día largo a 24°C fertilizando cada 15 días.

2) Metodologías generales

2.a) Preparación de ADN Plasmídico

Se partió de 5 ml de cultivos crecidos toda la noche a 37°C en medio YEP líquido (bacto-triptona 1% p/v, extracto de levadura 1% p/v, NaCl 0,5% p/v). Las células se cosecharon por centrifugación a 7000 rpm por 10 minutos. El pellet se resuspendió en 250 µl de solución P1 (Tris-HCl 25 mM pH 8, EDTA 10 mM) y se traspasó a tubos eppendorf. Se agregaron 300 µl de solución P2 (NaOH 0,2 N, SDS 1% p/v) preparada en el momento, se mezcló suavemente por inversión, se agregaron 350 µl de solución P3 (Acetato de potasio 3M pH 5,2). Se mezcló enérgicamente cada tubo y se los colocó en hielo por aproximadamente 5 minutos. Luego se centrifugaron a 10000 rpm por 10 minutos en micro centrífuga, se tomaron aproximadamente 850 µl del sobrenadante y se transfirieron a un nuevo tubo con 650 µl de isopropanol. Tras dejar reposar 5 minutos a temperatura ambiente se centrifugó a 10000 rpm por 10 minutos, se realizó un lavado del pellet con 500 µl de etanol 70%, se descartó el sobrenadante y se dejó secar. Una vez seco, el pellet se resuspendió en 50 µl de buffer P1 0,5X + 0,1 mg/ml de RNAsa A para eliminar restos de RNA.

Para obtener una purificación de mayor grado la preparación de plásmido se diluyó 1:5 en buffer PB (cloruro de guanidinio 5M, Tris-HCl 20 mM pH 6,6, etanol 38% v/v) y se depositó en una columna que contiene una matriz de sílica comercial (EconoColumn, Epoch Life Science). Se centrifugó a 10000 r.p.m. por 1min, y se lavó con 750 µl de buffer PE (Tris-HCl 10 mM pH 8,0, NaCl 100 mM, etanol 80% v/v). Se centrifugó a 10000 r.p.m. por 1 minuto, se descartó el eluido y se realizó una nueva centrifugación para remover el etanol residual. Se ubicó la columna que contiene la matriz de sílica en un tubo eppendorf y se agregaron 50 µl de buffer EB (Tris-HCl 10 mM pH 8,5). Se centrifugó a 10000 r.p.m. para obtener el plásmido eluido.

2.b) Preparación de fragmentos de ADN

2.b.i) Amplificación de fragmentos de ADN (PCR)

La amplificación de fragmentos de ADN que se utilizaron para los clonados se realizó utilizando las ADN polimerasas de alta fidelidad Phusion o Q5 (ambas de New England Biolabs), según el protocolo sugerido y con los buffers suministrados por el fabricante. Los ciclos de variación de temperatura también se realizaron según recomendaciones provistas en la hoja de datos del producto. Típicamente se utilizó el siguiente termociclado:

Desnaturalización inicial	98 °C	30 segundos
25-35 ciclos (desnaturalización, hibridación y extensión)	98 °C	5-10 segundos
	50-72 °C	10-30 segundos
	72 °C	20-30 segundos/kb
Extensión final	72 °C	2 minutos

Para el cálculo de la temperatura de hibridación se utilizó la herramienta online <http://tmcalculator.neb.com> sugerida y provista por New England Biolabs.

Los productos de las reacciones se analizaron por electroforesis en gel de agarosa 1-2% p/v, dependiendo del tamaño de los fragmentos a resolver, en buffer TBE 0,5x o TAE 1x, conteniendo bromuro de etidio.

2.b.ii) Purificación de fragmentos de ADN de geles de agarosa

Los productos de PCR de interés se identificaron por electroforesis en geles de agarosa. Se cortaron del gel las bandas correspondientes a los productos de interés y se purificaron utilizando matrices de sílica comerciales (EconoColumn, Epoch Life Science). Para ello se determinó el peso del taco de agarosa y se lo solubilizó en buffer QG (tiocianato de guanidinio 5.5 M, Tris-HCl pH 6.6 20 mM) en una relación de 3000 µl/g de agarosa. La solución se colocó dentro de una columna que contiene una matriz de sílica comercial y a partir de este paso se continuó igual que en la purificación de ADN plasmídico.

2.b.iii) Digestión con enzimas de restricción

Las digestiones de plásmidos e insertos para clonado y los análisis de patrones de restricción se realizaron con enzimas comerciales siguiendo las recomendaciones del fabricante. Las reacciones se incubaron entre 2 y 5 horas a la temperatura óptima para cada enzima y se utilizaron de 2 a 6 unidades de enzima por µg de ADN. La eficiencia de la digestión se visualizó por electroforesis en geles de agarosa y en el caso de las digestiones para su posterior clonado se repurificó el ADN digerido como se describió anteriormente.

2.b.iv) Reacciones de ligación de ADN

Las reacciones de ligación de plásmidos e insertos se realizaron con la enzima T4 ADN ligasa con el buffer suministrado y siguiendo las recomendaciones del fabricante (New England Biolabs). Se utilizaron 25-50 ng de plásmido, relaciones molares de inserto:plásmido 3:1, un volumen final de 10 µl y 10 unidades de enzima por reacción. Las reacciones se incubaron a 16°C-18°C durante toda la noche. El producto de ligación se utilizó para transformar células *E. coli* DH5α.

2.b.v) Clonados por sistema Gateway

Este sistema se basa en el clonado de secuencias mediante recombinación en vectores de tipo pDONR para generar un *Entry Clone*. Este nuevo vector permite pasar las secuencias a otros plásmidos compatibles con el sistema Gateway mediante sucesivas recombinaciones.

En este trabajo se clonaron los fragmentos de ADN amplificados por PCR portando las secuencias de recombinación flanqueantes attB en el vector pDONR201 usando la recombinasa BP (Invitrogen) siguiendo las recomendaciones del fabricante. Este vector

posee un “*casette*” letal que es reemplazado por el inserto al recombinar. Una vez transformado el producto de la recombinación se seleccionaron los clones que portan el plásmido recombinante en un medio de selección. Se realizó una preparación de ADN plasmídico de el/los clon/es seleccionados, se comprobó que el patrón de restricción fuera el esperado y que la secuencia no tuviera errores. Con estos plásmidos seleccionados se procedió a transferir el fragmento a los vectores de destino por recombinación usando la recombinasa LR (Invitrogen).

3) Cepas bacterianas y transformación

3.a) *Escherichia coli*

3.a.i) Cepas y condiciones de cultivo

Se utilizó la cepa DH5 α para realizar las construcciones genéticas. Las cepas se cultivaron en medio líquido YEP a 37°C con agitación a 250 r.p.m. Los cultivos sólidos se efectuaron en medio YEP suplementado con agar 1,5% p/v a 37°C. De acuerdo a la resistencia otorgada por cada plásmido, los medios se suplementaron con ampicilina 100 μ g/ml, kanamicina 50 μ g/ml, espectinomicina 100 μ g/ml, gentamicina 20 μ g/ml o cloranfenicol 35 μ g/ml.

3.a.ii) Preparación de células de *E. coli* DH5 α electro competentes

La obtención de células de *E. coli* DH5 α competentes para ser transformadas por electroporación se realizó de la siguiente manera. Se sembró una colonia fresca y aislada de bacterias en 10 ml de medio de cultivo YEP y se cultivó toda la noche a 37°C con agitación (250 r.p.m.). Se utilizó ese cultivo para inocular 1Lt de medio de cultivo YEP y se cultivaron las bacterias a 28°C hasta alcanzar una densidad óptica a 600nm de 0.5-0.7. En condiciones de esterilidad, se lavaron dos veces en frío con H₂O bidestilada. Finalmente, se resuspendieron en glicerol 10% frío. Las células se alicuotaron en fracciones de 50 μ l y se congelaron en nitrógeno líquido inmediatamente. Fueron almacenadas a -80°C hasta ser usadas.

3.a.iii) Transformación de *E. coli* por electroporación

Se agregaron 1-2 μ l del producto de ligación o 0,1-10 ng de plásmido purificado a 50 μ l de células competentes mantenidas en hielo. La mezcla se traspasó a una cubeta de electroporación y se electroporó en un electroporador Biorad Gene Pulser (1.25 kV/cm,

capacitancia: 25 μ F y resistencia: 200 Ohms). Se comprobó que la constante de tiempo fuera mayor o igual a 4,5. Se agregaron 500 μ l de medio YEP y la suspensión bacteriana se transfirió a un tubo donde se incubó a 37°C durante 1 hora con agitación para permitir la recuperación del shock eléctrico. Para seleccionar las células transformadas, se sembraron 100-300 μ l de la suspensión bacteriana en medio YEP sólido suplementado con el antibiótico correspondiente.

3.b) *Agrobacterium tumefaciens*

3.b.i) Cepas y condiciones de cultivo

Se utilizó la cepa GV3101 de *A. tumefaciens* que posee resistencia plasmídica (pTiC58) a gentamicina. Las células se cultivaron en medio YEP suplementado con gentamicina y el antibiótico correspondiente con la resistencia otorgada por cada plásmido binario a 28°C con agitación a 250 r.p.m. Los cultivos sólidos se efectuaron a 28°C en medio YEP suplementado con agar 1,5% p/v.

3.b.ii) Preparación de células electro competentes de *A. tumefaciens*.

La obtención de células de *A. tumefaciens* GV3101 competentes para ser transformadas por electroporación se realizó de la misma manera que con las células de *E. coli* DH5 α , con la salvedad de que se suplementaron los medios de cultivo con gentamicina.

3.b.iii) Transformación de *A. tumefaciens* por electroporación.

La transformación de células de *A. tumefaciens* GV3101 competentes se realizó de la misma manera que con las células de *E. coli* DH5 α , con la salvedad de que la recuperación se realizó incubando a 28°C.

3.b.iv) Agroinfiltración

Se sembró una colonia fresca y aislada de bacterias en 5 ml de medio de cultivo YEP suplementado con los antibióticos correspondientes y se cultivó toda la noche a 28°C con agitación (250 r.p.m.). Una vez saturado el cultivo, se agregó acetosiringona de manera de obtener una concentración final de 200 μ M y se incubó por otras 3 hs a 28°C y 250 rpm. Luego de ese tiempo, se colectaron las bacterias por centrifugación y se lavaron con medio de infiltración (MI) 1X. Se resuspendieron las bacterias en MI 1X de forma que alcanzaran una densidad óptica a 600nm de 0.5 y se incubaron 1 h a 28°C y 200 rpm. La suspensión de

Procedimientos Experimentales

bacterias se infiltró en la cara abaxial de hojas de *N.benthamiana* de unas 4-5 semanas con jeringa de insulina o tuberculina. Pasados 3 o 4 días se procesaron las hojas según las características de las proteínas expresadas.

Medio de Infiltración 5X

250 mM	Mes
10 mM	NaH ₂ PO ₄
pH 5,6 con NaOH	

Medio de Infiltración (MI) 1X (para 40 ml)

200µM	Acetosiringona
0,1 µg/ml	ácido 2,4 diclorofenoxiacético
200 mg	Sacarosa
8 ml	Medio de Infiltración 5X
32 ml	Volumen final

4) Análisis filogenético

Las secuencias usadas se obtuvieron de varias fuentes utilizando tblastn (Altschul et al., 1997), con la secuencia aminoacídica de los genes BSL de Arabidopsis o arroz como secuencia “*query*”. Siempre que fue posible, los resultados de BLAST se utilizaron como punto de partida para recuperar las secuencias genómicas completas, las cuales fueron sometidas a “*splicing*” manualmente. Las secuencias nucleotídicas curadas fueron traducidas a sus correspondientes secuencias aminoacídicas (Geneious 5.6.5, Biomatters, www.biomatters.com); la lista de secuencias usadas se muestra en el Anexo I en la versión digital.

Las secuencias aminoacídicas se alinearon usando el algoritmo MAFFT implementado en Genious (Kato et al., 2002). Las secuencias alineadas se inspeccionaron y editaron manualmente con eBioX v1.5.1 (ebioinformatics.org) con el fin de refinar el alineamiento y excluir regiones hipervariables ambiguamente alineadas. Se generó un conjunto de datos de secuencia completa y un alineamiento conteniendo sólo la región C-terminal. Para cada conjunto de datos se realizó un análisis filogenético usando el método RAXML versión 2.2.3 (Stamatakis, 2006) implementado en la interfaz gráfica TOPALi 2.5 (Milne et al., 2004). El análisis filogenético fue llevado a cabo por el Dr. Claudio Slamovits, Dalhousie University, Halifax, Canada.

5) Plásmidos y construcciones utilizadas

Los clones de cDNA de secuencia completa de Arabidopsis para BSL1 (RAFL09-11-J01), BSL2 (RAFL16-56-H16) y BSL3 (RAFL16-39-G23) se obtuvieron de RIKEN (www.brc.riken.jp/lab/epd/catalog/cdnaclone.html). Los clones de cDNA de BSL1 y BSL3 fueron editados antes de usarlos: el clon de BSL1 porta una mutación en un residuo altamente conservado del dominio fosfatasa (Q794P) y el clon de BSL3 es una forma errónea de splicing alternativo que retiene parte del intrón número 13, introduciendo un codón de terminación prematuro. En ambos casos, las secuencias alrededor de estas posiciones se reemplazaron por su respectivo “*cassette*” WT amplificado de RNA retrotranscrito de Col-0. El cDNA de secuencia completa de BSU1 se amplificó de RNA retrotranscrito derivado de plantas bsu1-D (Mora-Garcia et al., 2004). Los cDNA de secuencia completa de BSK1 y BSK3 se amplificaron a partir de RNA retrotranscrito de Col-0.

5.a) Diseño y obtención de las construcciones para Y2H

Se amplificó el ADNc de los genes o dominios seleccionados con oligonucleótidos específicos (ver listado en el Anexo II en la versión digital), con el propósito de obtener construcciones flanqueadas por sitios de recombinación del sistema Gateway (Invitrogen). Los mismos fueron transferidos al vector pDONR201, y de ahí a los plásmidos de destino del sistema Proquest Two-Hybrid (Invitrogen), pDEST32 y pDEST22, para obtener fusiones al dominio de unión al ADN y al dominio de activación del factor de transcripción Gal4, respectivamente. Los marcadores de selección son Gentamicina y Leu para pDEST32 y Ampicilina y Trp para pDEST22, en E.coli y levaduras respectivamente.

5.b) Diseño y obtención de las construcciones fusionadas a epítopes

Se amplificó el ADNc de los genes o dominios seleccionados con codón iniciador flanqueado por una secuencia de tipo Kozak y sin codón terminador, con primers específicos (ver listado de primers, Tabla Anexo.2), con el propósito de obtener construcciones flanqueadas por sitios de recombinación del sistema Gateway (Invitrogen). Las construcciones se recombinaron en pEarley103 (Earley et al., 2006) y en pK7m34GW, en combinación con pEN-R2-3xHA-L3 y pDONRP4-P1_p35S (Karimi et al., 2007), para obtener las fusiones a GFP o al epítipo HA, respectivamente. Ambas construcciones fusionadas se encuentran bajo el control del promotor 35S.

5.c) Diseño y obtención de las construcciones para biFC

Se amplificaron dos versiones del ADNc de cada gen seleccionado con primers específicos (ver listado de primers), con el propósito de obtener construcciones flanqueadas por sitios de recombinación del sistema Gateway (Invitrogen): una de las formas con codón iniciador, pero sin codón terminador, y la otra sin codón iniciador, pero con codón terminador. Las primeras se recombinaron en el vector pK7m34GW, en combinación con pDONR-P4-P1_p35S, y pDONR-P2R-P3-heGFP o pDONR-P2R-P3-teGFP. De esta manera se obtuvieron las fusiones de las mitades N- y C-terminal de GFP en el extremo C-terminal de las construcciones, respectivamente. Las otras formas se recombinaron en el vector pK7m24GW2, en combinación con pDONR-P4-P1R-heGFP o pDONR-P4-P1R-teGFP para obtener las fusiones a las mitades N- y C-terminal de GFP respectivamente en el extremo N-terminal de las construcciones. Todas las construcciones obtenidas de esta manera se encuentran bajo el control del promotor 35S.

6) Ensayo de doble híbrido en levaduras (Y2H)

Los clones de ADNc recombinados en pDEST22 o pDEST32, se introdujeron en levaduras de la cepa AH109 (Clontech, www.clontech.com).

6.a) Transformación de levaduras

Los plásmidos pDEST32 y pDEST22, recombinados o vacíos como control, se introdujeron simultáneamente en la cepa AH109. Se partió de 1 ml de cultivo líquido incubado toda la noche a 28°C en medio YPD (Tryptona 20 g/lit, extracto de levadura 10 g/lit, glucosa 20 g/lit, Hemisulfato de Adenina 120 mg/lit). Se concentraron las células centrifugando por 1 minuto a 6000 rpm, se descartó el sobrenadante, se lavaron con agua, se resuspendieron en la mezcla de transformación y se incubaron a 42°C por 60-90 minutos. Pasado este tiempo, las células se concentraron centrifugando por 1 minuto a 6000 r.p.m., se descartó el sobrenadante, se resuspendieron en 100 µl de agua y se depositaron 30 µl en placas de medio selectivo sin Leu ni Trp (llamado de aquí en más -L-W).

Mezcla de transformación

PEG 3350 50%	120 μ l
Acetato de Litio 1 M	18 μ l
ADN de esperma de arenque (10 μ g/ μ l)	5 μ l
Plásmido (0,1 a 1 μ g) + H ₂ O	37 μ l

Medio selectivo (pH 5,6)

Difco Yeast Nitrogen Base	6.9 g
Glucosa	20 g
Synthetic Complete Drop Out mix	Variable según la selección
H ₂ O	c.s.p. 1 lt

Una vez que se obtuvieron las colonias transformantes, se reestriaron 6 colonias independientes en medio fresco -L-W para ensayar la actividad de los reporteros. Se ensayó primero la actividad β -Galactosidasa sobre estrías de las colonias seleccionadas, y luego la habilidad de las mismas para crecer en medio selectivo “*drop out*” -Leu -Trp -His con la adición de diferentes concentraciones de 3-amino-1,2,4-triazol (3AT, Sigma-Aldrich).

6.b) Medición de la actividad de los genes reporteros

6.b.i) β -Galactosidasa (β -Gal)

Actividad β -Gal sobre filtro

Las colonias elegidas se reestriaron sobre papel de filtro Whatman N°1 depositado sobre la superficie de una placa de medio -L-W. Se incubaron a 28°C hasta que la estría se volvió visible (típicamente unas 16 horas). El filtro se sumergió en N₂ líquido y se retiró del mismo por 1 minuto. Este procedimiento se repitió 3 veces para romper las células. El filtro se depositó sobre papel Whatman 3MM embebido en la solución de revelado. Se incubó a 37°C hasta que el color azul se hizo visible (típicamente entre 1 y 2 horas).

Solución de revelado

Buffer Z	100 ml
β-Mercaptoetanol	0.27 ml
X-Gal 2%	1.67 ml

Buffer Z

Buffer Fosfato de Sodio pH 7.0	100 mM
Sulfato de Magnesio	1 mM
Cloruro de Potasio	10 mM

Actividad β-Gal en medio líquido

Por cada colonia a ensayar, se incubaron las células a 28 °C con agitación a 250 rpm en medio líquido -L-W durante por lo menos 16 horas. Se midió la densidad óptica a 600 nm y se diluyeron en 5 ml de medio YPD para obtener una densidad óptica de 0,8. Se incubaron a 28 °C hasta alcanzar una densidad óptica entre 1 y 1.5. Se centrifugó 1 ml de cultivo en un tubo eppendorf de 1,5 ml a 6000 rpm por 30 seg y se descartó el sobrenadante. Se resuspendió el pellet en 1,5 ml de buffer Z y se volvió a centrifugar a 6000 rpm por 30 seg. Se resuspendió el pellet en 200 µl de buffer Z y se separaron en 2 tubos con 100 µl cada uno. Se sumergieron los tubos en N₂ líquido hasta que la suspensión se congeló (10 seg) y se transfirieron rápidamente a un baño a 37 °C por 90 seg. Se repitió este procedimiento dos veces más. Se agregaron a los 100 µl en cada tubo 700 µl de buffer Z + 1,9 µl de β-mercaptoetanol, y 160 µl de buffer Z + o-Nitrofenilgalactósido (ONPG en concentración 4 mg/ml). Se incubaron a 30°C, comenzando a contar el tiempo desde el agregado de ONPG. Cuando la coloración amarilla se volvió notable se detuvo la reacción agregando 400 µl de una solución de Na₂CO₃ 1M y los tubos se centrifugaron a 10000 rpm por 5 minutos. Se midió la absorbancia de los sobrenadantes a 420 nm y la actividad β-Gal se calculó como

$$\text{Unidades de } \beta\text{-Gal} = 1000 \times \text{OD}^{420} / t \times V \times \text{OD}^{600}$$

donde V= volumen inicial de cultivo usado

t = tiempo de reacción

6.b.ii) HIS3

La expresión del gen reportero *HIS3* se evaluó depositando las colonias seleccionadas sobre medio -L-W (como control de transformación) y sobre medio -L-W-H con el agregado de distintas concentraciones de 3AT, un inhibidor competitivo de la enzima imidazolglicerol-fosfato dehidratasa. Para ello las células se crecieron en medio líquido -L-W durante toda la noche, se midió la densidad óptica alcanzada y se hicieron diluciones con agua hasta llegar a una DO^{600nm} de 1,0, o diluciones sucesivas.

7) Extracción y detección de proteínas mediante Western Blot

7.a) Extracción de proteínas totales

Saccharomyces cerevisiae

Las extracciones de proteínas se realizaron a partir de cultivos de levaduras en fase exponencial de crecimiento, siguiendo las indicaciones del “*Yeast Protocols Handbook*” (Clontech). Brevemente, cultivos en medio selectivo se diluyeron en medio rico fresco y se dejaron crecer hasta alcanzar una DO^{600} entre 0.4 y 0.6. Se colectaron por centrifugación células equivalentes a 4.5 UDO, que se conservaron en N_2 líquido. Las células se rompieron con cuentas de vidrio con el agregado de buffer de carga caliente y finalmente el lisado se hirvió durante 5 minutos. El sobrenadante se depositó en el gel.

Nicotiana benthamiana

Fragmentos de unos 50 mg de hojas de *N.benthamiana* infiltradas con *Agrobacterium* llevando las construcciones correspondientes se congelaron en N_2 líquido y se molieron, se agregaron 2 volúmenes de buffer de carga 2x (Tris-HCl 125 mM pH 7,5, SDS 4%, Urea 8M, DTT 0,2 M, Azul de Bromofenol 0,02% p/v) por unidad de masa colectada y se hirvieron por 5 min. Se centrifugó a 10000 r.p.m. en microcentrífuga por 10 minutos y una fracción del sobrenadante se depositó en los geles.

7.b) Electroforesis de proteínas

Los extractos proteicos fueron separados en geles desnaturalizantes de poliacrilamida (SDS-PAGE, *Sodium dodecyl sulfate-Polyacrylamide gel electrophoresis*) de 10 cm x 10 cm y de 0.75 mm de grosor de acuerdo con el método de Laemmli (Laemmli, 1970). La electroforesis

se realizó a 80 Volts (voltaje constante) para el gel concentrador y a 100 Volts para el separador hasta que el frente alcanzó el final del gel. Como buffer de corrida se utilizó Tris-glicina-SDS 1x (Tris 25 mM, glicina 192 mM, SDS 0,1% p/v).

7.c) Transferencia a membrana

Las proteínas separadas por SDS-PAGE se electrotransfirieron a una membrana de PVDF (difluoruro de polivinilideno, Bio-Rad) usando un equipo de transferencia húmeda (Bio Rad) a 90 Volts por 1 hora, en cuarto frío (8-12°C) y con agitación constante del buffer de transferencia (Tris base 3% p/v, glicina 14,4% p/v, metanol 20% v/v). Luego de la transferencia la membrana se tiñó brevemente con rojo Ponceau para verificar que la transferencia fuera exitosa.

7.d) Inmunodetección y revelado

La membrana se bloqueó con leche descremada al 5% en TBS (Tris-HCl 50mM pH 7,5, NaCl 150mM) durante 2 horas a temperatura ambiente. Transcurrida la incubación, se eliminó la solución de bloqueo y se incubó con los anticuerpos correspondientes.

Anti-epítotope HA

Se incubó un anticuerpo anti-HA monoclonal hecho en rata (Roche) en una dilución 1:500 en leche descremada al 2% en TBS por 1 hora a temperatura ambiente y luego toda la noche a 4°C con agitación. Se lavó 2 veces por 5 minutos con TBST (TBS con 0,05% Tween20) a 4°C. Se añadió un anticuerpo secundario anti-IgG de rata unido a peroxidasa hecho en cabra (Sigma) en una dilución 1:5000 en leche descremada al 2% en TBS. Se incubó por 2 horas a temperatura ambiente. Se lavó 3 veces por 6 minutos con TBS a 4°C.

Anti-GFP, anti-BSL2, anti-BD y anti-AD

El suero policlonal anti-GFP obtenido de conejo (Invitrogen) se usó en una dilución 1:1500, el suero policlonal anti-BSL2 obtenido en ratones en una dilución 1:500, los sueros monoclonales anti-BD y anti-AD obtenidos en ratones (Santa Cruz Biotechnology) en una dilución 1:400, todos en leche descremada al 2% en TBS. Se incubaron 1 hora a temperatura ambiente y luego toda la noche a 4°C con agitación. Se lavó 3 veces por 10 minutos con TBS a temperatura ambiente. Se añadieron los anticuerpos secundarios conjugados a peroxidasa

(Sigma) en una dilución 1:5000 en leche descremada al 2% en TBS. Se incubó por 2 horas a temperatura ambiente. Se lavó 3 veces por 10 minutos con TBS a temperatura ambiente.

Revelado

Transcurrida la incubación con el anticuerpo secundario y los lavados pertinentes de acuerdo a cada anticuerpo, se reveló la actividad peroxidasa mediante el agregado de un sustrato que genera una señal quimioluminiscente al ser modificado por la enzima peroxidasa (ECL, GE Healthcare). Para visualizar la señal, las membranas se expusieron a una placa radiográfica (Kodak BioMax).

8) Co-inmunoprecipitación

Se partió de 300 mg de hojas que fueron congelados inmediatamente después de cosechados. Se homogeneizó el tejido en mortero en presencia de N₂ líquido, hasta obtener un polvo fino. El mismo se transfirió a un eppendorf de 2ml, donde se le adicionó 1,8 ml de buffer de extracción (Tris-HCl 50 mM pH 7,5, NaCl 150 mM, Glicerol 10%, NP40 0,1%) y se resuspendió mediante vortex para luego incubar por 10 minutos en hielo. Una vez finalizada la incubación, se centrifugó a 10000 r.p.m. en microcentrífuga por 10 minutos. Esta acción se realizó por duplicado tomando el sobrenadante cada vez. Se tomó 1ml de la extracción proteica y se agregó con anticuerpo anti-GFP (Invitrogen) en una dilución 1:200 por 2,5 hs a en frío (8-12°C) con agitación. A continuación, se añadieron 5 mg de Sepharosa unida a proteína A resuspendida en 100 µl de buffer de extracción y se incubó nuevamente por 1,5 hs en frío con agitación. Al concluir, el extracto incubado (*input*), se centrifugó por 5 minutos a 4000 r.p.m. y se extrajo el sobrenadante (SN) con cuidado de no arrastrar Sepharosa (PP). Se lavó la Sepharosa 3 veces con buffer de extracción; finalmente se eliminó el resto con jeringa de tuberculina (aguja 27G ½). Se resuspendió la Sepharosa en 50 µl de buffer de extracción y 50 µl de buffer de siembra desnaturizante 2x y se incubó durante 10 minutos a 95°C y se procedió a realizar un Western blot como se describió.

9) Microscopía confocal

La presencia de GFP se reveló en un microscopio confocal Zeiss LSM5 Pascal usando una longitud de onda de excitación de 488 nm provista por un laser de argón; la fluorescencia se detectó usando un filtro BP 505-530 nm. Los ensayos de expresión transitoria (de las

proteínas completas o sus dominios, y los ensayos de biFC) se realizaron infiltrando hojas de *N.benthamiana* con cepas de *Agrobacterium* que expresaban las construcciones correspondientes. Porciones de la cara abaxial de las hojas fueron montadas en agua sobre un portaobjetos, 86 hs post-infiltración.

10) Cromatografía de exclusión molecular

Se utilizó una columna Superdex-200 10/300 GL (GE Healthcare). La columna fue equilibrada en buffer de cromatografía (Tris-HCl 50 mM pH 7,5, NaCl 150 mM, Glicerol 10%). Primero, se inyectó una solución de marcadores de peso molecular conocido y se registraron los volúmenes de elución de cada marcador. Con estos datos se realizó una curva de calibración. A continuación, se realizó la extracción, en buffer de cromatografía, de hojas de *N.benthamiana*, tal como se describió para la coimmunoprecipitación. Se inyectaron 500 μ l del extracto en la columna y se colectaron fracciones de 500 μ l con un colector automático. Las fracciones elegidas se precipitaron con ácido tricloroacético (TCA). Se realizó un Western blot de las fracciones precipitadas para revelar el perfil de elución de la proteína.

10.a) Precipitación con TCA

A un volumen de extracto se agregó un volumen de TCA 20% y se incubó por 20 minutos en hielo. A continuación, se centrifugó a 10000 r.p.m. en microcentrífuga refrigerada (4°C) por 10 min. Se descartó el sobrenadante y se lavó dos veces con 500ul de Acetona pura fría (-20°C). Se dejó secar y se resuspendió en 12,5 ul de Urea 7M. Se incubó toda la noche a temperatura ambiente y al día siguiente se agregaron 12,5 μ l de buffer de carga desnaturizante 2x. Se incubó durante 10 minutos a 95°C y se sembró en SDS-PAGE.

11) Expresión y purificación del dominio conector de BSL2

La secuencia de la región linker de BSL2 (642 pb, aminoácidos 449 a 662) se clonó en el vector pET22b+ para generar una proteína de fusión con poli-histidinas en su extremo C-terminal. La construcción se expresó en E.coli BSL21(DE3)/pRIL. Se partió de 5 ml de un cultivo crecido toda la noche a 37°C en medio YEP líquido. Se realizó una dilución 1/100 en medio YEP fresco y se incubó a 28°C con agitación hasta alcanzar una $DO^{600nm} = 0,4 - 0,5$. A esa OD se realizó la inducción de la expresión de la proteína recombinante con 0,1 mM de isopropil- β -D-1-tiogalactopiranosido (IPTG) por 3 hs a 28°C con agitación. Se centrifugó el

cultivo y se resuspendió el pellet en 10 ml de Buffer de lisis (Tris-HCl 20 mM, pH 7,5; NaCl 100 mM; mezcla de inhibidores de proteasas cComplete™, EDTA-free, Roche). La lisis se realizó con una French Press, repitiendo el proceso 3 veces. El lisado se centrifugó y el sobrenadante se calentó por 10 minutos a 80°C (HS). Una vez centrifugado, el nuevo sobrenadante se filtró a través de una columna HiTrap-IMAC equilibrada en Tris-HCl 20 mM, pH 7,5; NaCl 100 mM; 10 mM imidazol. La proteína fue eluída con un gradiente de imidazol entre 10 mM y 500 mM. Las fracciones de elución, a las que se agregó mezcla de inhibidores fresca a medida que se colectaban, fueron inmediatamente desaladas (columnas PD-10, GEHealthcare) y almacenadas en buffer Tris-HCl 20 mM, pH 7,5; NaCl 100 mM. Para evitar la formación de dímeros, la proteína purificada se trató con DTT 10 mM, se bloquearon las cisteínas libres con iodoacetamida 15 mM y se desaló nuevamente a través de una columna equilibrada en buffer de almacenamiento. La proteína se cuantificó por absorbancia a 280 nm, estimando una $\epsilon^{280\text{nm}}=4470 \text{ M}^{-1} \text{ cm}^{-1}$.

12) Técnicas biofísicas

12.a) Dicroísmo circular (DC)

Las mediciones de DC se realizaron en un espectropolarímetro Jasco J-810 con una cubeta de cuarzo 0.2 cm de paso óptico. Los espectros se tomaron entre 195 nm y 260 nm a una velocidad de barrido de 50 nm/min con un tiempo de respuesta de 8 s y ancho de banda de 2 nm. Se usaron concentraciones de proteína de alrededor de 5 μM y los espectros se tomaron a 20°C, a menos que se indique lo contrario. Los datos se convirtieron a elipticidad molar de enlace mediante la fórmula:

$$[\theta]_M = \theta / (c * l * fd)$$

donde θ es el valor experimental de elipticidad en grados, l el paso óptico de la cubeta en cm, c la concentración en dmol/cm^3 de enlaces, resultante de multiplicar la concentración de proteína por el número de enlaces (número de residuos-1) y fd es el factor de dilución, para el caso del agregado de TFE.

12.b) Static Light Scattering (SLS)

La masa molecular promedio se determinó por SLS con un instrumento Precision PD2080 conectado a un equipo de cromatografía líquida y a un refractómetro LKB 2142. La proteína

se inyectó en una columna analítica Superdex 75 y se eluyó con buffer 20 mM Tris-HCl / 200 mM NaCl (pH 8). La dispersión de luz, la absorbancia a 280 nm y el índice de refracción se analizaron con el software asociado (Discovery32). El detector se calibró con seroalbúmina bovina (66.5 kDa).

12.c) Resonancia magnética nuclear (RMN)

La determinación del Radio Hidrodinámico (Rh) se llevó a cabo en un espectrómetro Bruker 600 MHz Avance III equipado con criosonda. Se usó dioxano como estandar interno. Los tiempos de difusión fueron 15 ms para dioxano y 180 ms para el dominio conector. El Rh se calculó como

$$R_h = (D_{\text{diox}}/D_{\text{prot}}) \times R_{h,\text{diox}}$$

donde D_{diox} y D_{prot} son los coeficientes de difusión del dioxano y la proteína, y $R_{h,\text{diox}}$ es el radio hidrodinámico del dioxano, que se considera 2.12 Å.

Referencias

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402
- Anderson VL, Ramlall TF, Rospigliosi CC, Webb WW, Eliezer D** (2010) Identification of a helical intermediate in trifluoroethanol-induced alpha-synuclein aggregation. *Proc Natl Acad Sci U S A* **107**: 18850-18855
- Arai M, Sugase K, Dyson HJ, Wright PE** (2015) Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc Natl Acad Sci U S A* **112**: 9614-9619
- Bachvaroff TR, Handy SM, Place AR, Delwiche CF** (2011) Alveolate phylogeny inferred using concatenated ribosomal proteins. *J Eukaryot Microbiol* **58**: 223-233
- Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE, Forman-Kay JD** (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**: 106-109
- Benkert P, Tosatto SC, Schomburg D** (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* **71**: 261-277
- Bollen M, Peti W, Ragusa MJ, Beullens M** (2010) The extended PP1 toolkit: designed to create specificity. *Trends Biochem Sci* **35**: 450-458
- Brautigan DL** (2013) Protein Ser/Thr phosphatases--the ugly ducklings of cell signalling. *FEBS J* **280**: 324-345
- Brenner ED, Stevenson DW, McCombie RW, Katari MS, Rudd SA, Mayer KF, Palenchar PM, Runko SJ, Twigg RW, Dai G, Martienssen RA, Benfey PN, Coruzzi GM** (2003) Expressed sequence tag analysis in *Cycas*, the most primitive living seed plant. *Genome Biol* **4**: R78
- Buck M** (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys* **31**: 297-355
- Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT** (2010) Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38**: W563-568
- Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK** (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett* **15**: 956-963
- Cannon JF** (2013) How phosphorylation activates the protein phosphatase-1 * inhibitor-2 complex. *Biochim Biophys Acta* **1834**: 71-86
- Clouse SD** (2011) Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development. *Plant Cell* **23**: 1219-1230
- Connor JH, Weiser DC, Li S, Hallenbeck JM, Shenolikar S** (2001) Growth arrest and DNA damage-inducible protein GADD34 assembles a novel signaling complex containing protein phosphatase 1 and inhibitor 1. *Mol Cell Biol* **21**: 6841-6850
- Chaffotte AF, Guijarro JI, Guillou Y, Delepierre M, Goldberg ME** (1997) The "pre-molten globule," a new intermediate in protein folding. *J Protein Chem* **16**: 433-439
- Chen MS, Silverstein AM, Pratt WB, Chinkers M** (1996) The tetratricopeptide repeat domain of protein phosphatase 5 mediates binding to glucocorticoid receptor heterocomplexes and acts as a dominant negative mutant. *J Biol Chem* **271**: 32315-32320
- Chen MX, Cohen PT** (1997) Activation of protein phosphatase 5 by limited proteolysis or the binding of polyunsaturated fatty acids to the TPR domain. *FEBS Lett* **400**: 136-140
- Choy MS, Hieke M, Kumar GS, Lewis GR, Gonzalez-DeWhitt KR, Kessler RP, Stein BJ, Hessenberger M, Nairn AC, Peti W, Page R** (2014) Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code. *Proc Natl Acad Sci U S A* **111**: 4097-4102

- Choy MS, Page R, Peti W** (2012) Regulation of protein phosphatase 1 by intrinsically disordered proteins. *Biochem Soc Trans* **40**: 969-974
- Dancheck B, Ragusa MJ, Allaire M, Nairn AC, Page R, Peti W** (2011) Molecular investigations of the structure and function of the protein phosphatase 1-spinophilin-inhibitor 2 heterotrimeric complex. *Biochemistry* **50**: 1238-1246
- Das AK, Helps NR, Cohen PT, Barford D** (1996) Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *Embo J* **15**: 6798-6809
- de la Fuente van Bentem S, Anrather D, Dohnal I, Roitinger E, Csaszar E, Joore J, Buijnink J, Carreri A, Forzani C, Lorkovic ZJ, Barta A, Lecourieux D, Verhounig A, Jonak C, Hirt H** (2008) Site-specific phosphorylation profiling of Arabidopsis proteins by mass spectrometry and peptide chip analysis. *J Proteome Res* **7**: 2458-2470
- Der JP, Barker MS, Wickett NJ, dePamphilis CW, Wolf PG** (2011) De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* **12**: 99
- Deutscher J, Saier MH, Jr.** (2005) Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established. *J Mol Microbiol Biotechnol* **9**: 125-131
- Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW** (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. *Mol Biol Evol* **23**: 469-478
- Dworkin J** (2015) Ser/Thr phosphorylation as a regulatory mechanism in bacteria. *Curr Opin Microbiol* **24**: 47-52
- Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS** (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* **45**: 616-629
- Egloff MP, Cohen PT, Reinemer P, Barford D** (1995) Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *J Mol Biol* **254**: 942-959
- Fong JH, Panchenko AR** (2010) Intrinsic disorder and protein multibinding in domain, terminal, and linker regions. *Mol Biosyst* **6**: 1821-1828
- Gampala SS, Kim TW, He JX, Tang W, Deng Z, Bai MY, Guan S, Lalonde S, Sun Y, Gendron JM, Chen H, Shibagaki N, Ferl RJ, Ehrhardt D, Chong K, Burlingame AL, Wang ZY** (2007) An essential role for 14-3-3 proteins in brassinosteroid signal transduction in Arabidopsis. *Dev Cell* **13**: 177-189
- Ghelis T** (2011) Signal processing by protein tyrosine phosphorylation in plants. *Plant Signal Behav* **6**: 942-951
- Gibbons JA, Weiser DC, Shenolikar S** (2005) Importance of a surface hydrophobic pocket on protein phosphatase-1 catalytic subunit in recognizing cellular regulators. *J Biol Chem* **280**: 15903-15911
- Grallert A, Boke E, Hagting A, Hodgson B, Connolly Y, Griffiths JR, Smith DL, Pines J, Hagan IM** (2015) A PP1-PP2A phosphatase relay controls mitotic progression. *Nature* **517**: 94-98
- Guttery DS, Poulin B, Ferguson DJ, Szoor B, Wickstead B, Carroll PL, Ramakrishnan C, Brady D, Patzewitz EM, Straschil U, Solyakov L, Green JL, Sinden RE, Tobin AB, Holder AA, Tewari R** (2012) A unique protein phosphatase with kelch-like domains (PPKL) in *Plasmodium* modulates ookinete differentiation, motility and invasion. *PLoS Pathog* **8**: e1002948
- Habchi J, Mamelli L, Darbon H, Longhi S** (2010) Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS One* **5**: e11684
- Hanada K, Kuromori T, Myouga F, Toyoda T, Li WH, Shinozaki K** (2009) Evolutionary persistence of functional compensation by duplicate genes in Arabidopsis. *Genome Biol Evol* **1**: 409-414
- Heroes E, Lesage B, Gornemann J, Beullens M, Van Meervelt L, Bollen M** (2013) The PP1 binding code: a molecular-lego strategy that governs specificity. *FEBS J* **280**: 584-595
- Hothorn M, Belkhadir Y, Dreux M, Dabi T, Noel JP, Wilson IA, Chory J** (2011) Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474**: 467-471

- Hu Z, He H, Zhang S, Sun F, Xin X, Wang W, Qian X, Yang J, Luo X** (2012) A Kelch motif-containing serine/threonine protein phosphatase determines the large grain QTL trait in rice. *J Integr Plant Biol* **54**: 979-990
- Hudson AM, Cooley L** (2008) Phylogenetic, structural and functional relationships between WD- and Kelch-repeat proteins. *Subcell Biochem* **48**: 6-19
- Hurley TD, Yang J, Zhang L, Goodwin KD, Zou Q, Cortese M, Dunker AK, DePaoli-Roach AA** (2007) Structural basis for regulation of protein phosphatase 1 by inhibitor-2. *J Biol Chem* **282**: 28874-28883
- Hwang I, Chen HC, Sheen J** (2002) Two-component signal transduction pathways in Arabidopsis. *Plant Physiol* **129**: 500-515
- Ingebritsen TS, Cohen P** (1983) The protein phosphatases involved in cellular regulation. 1. Classification and substrate specificities. *Eur J Biochem* **132**: 255-261
- Janssens V, Longin S, Goris J** (2008) PP2A holoenzyme assembly: in cauda venenum (the sting is in the tail). *Trends Biochem Sci* **33**: 113-121
- Jez JM, Ferrer JL, Bowman ME, Dixon RA, Noel JP** (2000) Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**: 890-902
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW** (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97-100
- Jones DT, Cozzetto D** (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**: 857-863
- Karimi M, Depicker A, Hilson P** (2007) Recombinational cloning with plant gateway vectors. *Plant Physiol* **145**: 1144-1154
- Katoh K, Misawa K, Kuma K, Miyata T** (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059-3066
- Kelker MS, Page R, Peti W** (2009) Crystal structures of protein phosphatase-1 bound to nodularin-R and tautomycin: a novel scaffold for structure-based drug design of serine/threonine phosphatase inhibitors. *J Mol Biol* **385**: 11-21
- Kennelly PJ** (2002) Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS Microbiol Lett* **206**: 1-8
- Kim TW, Guan S, Burlingame AL, Wang ZY** (2011) The CDG1 kinase mediates brassinosteroid signal transduction from BRI1 receptor kinase to BSU1 phosphatase and GSK3-like kinase BIN2. *Mol Cell* **43**: 561-571
- Kim TW, Guan S, Sun Y, Deng Z, Tang W, Shang JX, Sun Y, Burlingame AL, Wang ZY** (2009) Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nat Cell Biol* **11**: 1254-1260
- Kinoshita T, Cano-Delgado A, Seto H, Hiranuma S, Fujioka S, Yoshida S, Chory J** (2005) Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**: 167-171
- Klumpp S, Krieglstein J** (2009) Reversible phosphorylation of histidine residues in proteins from vertebrates. *Sci Signal* **2**: pe13
- Konig J, Galliardt H, Jutte P, Schaper S, Dittmann L, Dietz KJ** (2013) The conformational bases for the two functionalities of 2-cysteine peroxiredoxins as peroxidase and chaperone. *J Exp Bot* **64**: 3483-3497
- Kutuzov MA, Andreeva AV** (2002) Protein Ser/Thr phosphatases with kelch-like repeat domains. *Cell Signal* **14**: 745-750
- Laemmli UK** (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685

- Lammers T, Lavi S** (2007) Role of type 2C protein phosphatases in growth regulation and in cellular stress signaling. *Crit Rev Biochem Mol Biol* **42**: 437-461
- Li J, Nam KH, Vafeados D, Chory J** (2001) BIN2, a new brassinosteroid-insensitive locus in Arabidopsis. *Plant Physiol* **127**: 14-22
- Li JL, Baker DA** (1998) A putative protein serine/threonine phosphatase from *Plasmodium falciparum* contains a large N-terminal extension and five unique inserts in the catalytic domain. *Mol Biochem Parasitol* **95**: 287-295
- Li X, Zhang D, Hannink M, Beamer LJ** (2004) Crystallization and initial crystallographic analysis of the Kelch domain from human Keap1. *Acta Crystallogr D Biol Crystallogr* **60**: 2346-2348
- Liu SL, Adams KL** (2010) Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the Brassicaceae. *Mol Biol Evol* **27**: 2817-2828
- Lochhead PA, Kinstry R, Sibbet G, Rawjee T, Morrice N, Cleghon V** (2006) A chaperone-dependent GSK3beta transitional intermediate mediates activation-loop autophosphorylation. *Mol Cell* **24**: 627-633
- Lv DW, Li X, Zhang M, Gu AQ, Zhen SM, Wang C, Li XH, Yan YM** (2014) Large-scale phosphoproteome analysis in seedling leaves of *Brachypodium distachyon* L. *BMC Genomics* **15**: 375
- Maselli GA, Slamovits CH, Bianchi JI, Vilarrasa-Blasi J, Cano-Delgado AI, Mora-Garcia S** (2014) Revisiting the evolutionary history and roles of protein phosphatases with Kelch-like domains in plants. *Plant Physiol* **164**: 1527-1541
- Meiselbach H, Sticht H, Enz R** (2006) Structural analysis of the protein phosphatase 1 docking motif: molecular description of binding specificities identifies interacting proteins. *Chem Biol* **13**: 49-59
- Milne I, Wright F, Rowe G, Marshall DF, Husmeier D, McGuire G** (2004) TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* **20**: 1806-1807
- Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD** (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **18**: 494-506
- Mizianty MJ, Uversky V, Kurgan L** (2014) Prediction of intrinsic disorder in proteins using MFDp2. *Methods Mol Biol* **1137**: 147-162
- Moorhead GB, De Wever V, Templeton G, Kerk D** (2009) Evolution of protein phosphatases in plants and animals. *Biochem J* **417**: 401-409
- Mora-Garcia S, Vert G, Yin Y, Cano-Delgado A, Cheong H, Chory J** (2004) Nuclear protein phosphatases with Kelch-repeat domains modulate the response to brassinosteroids in Arabidopsis. *Genes Dev* **18**: 448-460
- Noval MG, Gallo M, Perrone S, Salvay AG, Chemes LB, de Prat-Gay G** (2013) Conformational dissection of a viral intrinsically disordered domain involved in cellular transformation. *PLoS One* **8**: e72760
- O'Connell N, Nichols SR, Heroes E, Beullens M, Bollen M, Peti W, Page R** (2012) The molecular basis for substrate specificity of the nuclear NIPP1:PP1 holoenzyme. *Structure* **20**: 1746-1756
- Ohad N, Shichrur K, Yalovsky S** (2007) The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation. *Plant Physiol* **145**: 1090-1099
- Oldfield CJ, Dunker AK** (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**: 553-584
- Pereira SF, Goss L, Dworkin J** (2011) Eukaryote-like serine/threonine kinases and phosphatases in bacteria. *Microbiol Mol Biol Rev* **75**: 192-212
- Peti W, Nairn AC, Page R** (2013) Structural basis for protein phosphatase 1 regulation and specificity. *FEBS J* **280**: 596-611

- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE** (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612
- Philip N, Vaikkinen HJ, Tetley L, Waters AP** (2012) A unique Kelch domain phosphatase in *Plasmodium* regulates ookinete morphology, motility and invasion. *PLoS One* **7**: e44617
- Puttick J, Baker EN, Delbaere LT** (2008) Histidine phosphorylation in biological systems. *Biochim Biophys Acta* **1784**: 100-105
- Qi P, Lin YS, Song XJ, Shen JB, Huang W, Shan JX, Zhu MZ, Jiang L, Gao JP, Lin HX** (2012) The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res* **22**: 1666-1680
- Ragusa MJ, Dancheck B, Critton DA, Nairn AC, Page R, Peti W** (2010) Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nat Struct Mol Biol* **17**: 459-464
- Ramsey AJ, Chinkers M** (2002) Identification of potential physiological activators of protein phosphatase 5. *Biochemistry* **41**: 5625-5632
- Rogers JM, Oleinikovas V, Shammas SL, Wong CT, De Sancho D, Baker CM, Clarke J** (2014) Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proc Natl Acad Sci U S A* **111**: 15420-15425
- Rose CM, Venkateswaran M, Grimsrud PA, Westphall MS, Sussman MR, Coon JJ, Ane JM** (2012) Medicago PhosphoProtein Database: a repository for *Medicago truncatula* phosphoprotein data. *Front Plant Sci* **3**: 122
- Ryu H, Kim K, Cho H, Hwang I** (2010) Predominant actions of cytosolic BSU1 and nuclear BIN2 regulate subcellular localization of BES1 in brassinosteroid signaling. *Mol Cells* **29**: 291-296
- Ryu H, Kim K, Cho H, Park J, Choe S, Hwang I** (2007) Nucleocytoplasmic shuttling of BZR1 mediated by phosphorylation is essential in *Arabidopsis* brassinosteroid signaling. *Plant Cell* **19**: 2749-2762
- Saunders DG, Breen S, Win J, Schornack S, Hein I, Bozkurt TO, Champouret N, Vleeshouwers VG, Birch PR, Gilroy EM, Kamoun S** (2012) Host protein BSL1 associates with *Phytophthora infestans* RXLR effector AVR2 and the *Solanum demissum* Immune receptor R2 to mediate disease resistance. *Plant Cell* **24**: 3420-3434
- Schweighofer A, Hirt H, Meskiene I** (2004) Plant PP2C phosphatases: emerging functions in stress signaling. *Trends Plant Sci* **9**: 236-243
- Shi H, Yan H, Li J, Tang D** (2013) BSK1, a receptor-like cytoplasmic kinase, involved in both BR signaling and innate immunity in *Arabidopsis*. *Plant Signal Behav* **8**
- Shi Y** (2009) Assembly and structure of protein phosphatase 2A. *Sci China C Life Sci* **52**: 135-146
- Shi Y** (2009) Serine/threonine phosphatases: mechanism through structure. *Cell* **139**: 468-484
- Sreeramulu S, Mostizky Y, Sunitha S, Shani E, Nahum H, Salomon D, Hayun LB, Gruetter C, Rauh D, Ori N, Sessa G** (2013) BSKs are partially redundant positive regulators of brassinosteroid signaling in *Arabidopsis*. *Plant J* **74**: 905-919
- Stamatakis A** (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690
- Takemiya A, Ariyoshi C, Shimazaki K** (2009) Identification and functional characterization of inhibitor-3, a regulatory subunit of protein phosphatase 1 in plants. *Plant Physiol* **150**: 144-156
- Tang W, Kim TW, Osés-Prieto JA, Sun Y, Deng Z, Zhu S, Wang R, Burlingame AL, Wang ZY** (2008) BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. *Science* **321**: 557-560
- Tang W, Yuan M, Wang R, Yang Y, Wang C, Osés-Prieto JA, Kim TW, Zhou HW, Deng Z, Gampala SS, Gendron JM, Jonassen EM, Lillo C, DeLong A, Burlingame AL, Sun Y, Wang ZY** (2011) PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1. *Nat Cell Biol* **13**: 124-131

- Teich R, Grauvogel C, Petersen J** (2007) Intron distribution in Plantae: 500 million years of stasis during land plant evolution. *Gene* **394**: 96-104
- Terrak M, Kerff F, Langsetmo K, Tao T, Dominguez R** (2004) Structural basis of protein phosphatase 1 regulation. *Nature* **429**: 780-784
- Timme RE, Bachvaroff TR, Delwiche CF** (2012) Broad phylogenomic sampling and the sister lineage of land plants. *PLoS One* **7**: e29696
- Trecek M, Sanders JL, Elias JE, Boothroyd JC** (2011) The phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* **10**: 410-419
- Tulin F, Cross FR** (2014) A microbial avenue to cell cycle control in the plant superkingdom. *Plant Cell* **26**: 4019-4038
- Uhrig RG, Labandera AM, Moorhead GB** (2013) Arabidopsis PPP family of serine/threonine protein phosphatases: many targets but few engines. *Trends Plant Sci* **18**: 505-513
- Uversky VN** (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**: 739-756
- Uversky VN** (2011) Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* **43**: 1090-1103
- Uversky VN** (2013) The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS Lett* **587**: 1891-1901
- Vaulot D, Eikrem W, Viprey M, Moreau H** (2008) The diversity of small eukaryotic phytoplankton (< or =3 microm) in marine ecosystems. *FEMS Microbiol Rev* **32**: 795-820
- Walsh I, Martin AJ, Di Domenico T, Vullo A, Pollastri G, Tosatto SC** (2011) CSpritz: accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs. *Nucleic Acids Res* **39**: W190-196
- Wang X, Chory J** (2006) Brassinosteroids regulate dissociation of BKI1, a negative regulator of BRI1 signaling, from the plasma membrane. *Science* **313**: 1118-1122
- Wang X, Kota U, He K, Blackburn K, Li J, Goshe MB, Huber SC, Clouse SD** (2008) Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. *Dev Cell* **15**: 220-235
- Wang ZY** (2012) Brassinosteroids modulate plant immunity at multiple levels. *Proc Natl Acad Sci U S A* **109**: 7-8
- Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ** (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**: 16424-16431
- Wozniak E, Oldziej S, Ciarkowski J** (2000) Molecular modeling of the catalytic domain of serine/threonine phosphatase-1 with the Zn²⁺ and Mn²⁺ di-nuclear ion centers in the active site. *Comput Chem* **24**: 381-390
- Xu W, Huang J, Li B, Li J, Wang Y** (2008) Is kinase activity essential for biological functions of BRI1? *Cell Res* **18**: 472-478
- Xu Y, Chen Y, Zhang P, Jeffrey PD, Shi Y** (2008) Structure of a protein phosphatase 2A holoenzyme: insights into B55-mediated Tau dephosphorylation. *Mol Cell* **31**: 873-885
- Xu Y, Xing Y, Chen Y, Chao Y, Lin Z, Fan E, Yu JW, Strack S, Jeffrey PD, Shi Y** (2006) Structure of the protein phosphatase 2A holoenzyme. *Cell* **127**: 1239-1251
- Yan Z, Zhao J, Peng P, Chihara RK, Li J** (2009) BIN2 functions redundantly with other Arabidopsis GSK3-like kinases to regulate brassinosteroid signaling. *Plant Physiol* **150**: 710-721
- Yang J, Roe SM, Cliff MJ, Williams MA, Ladbury JE, Cohen PT, Barford D** (2005) Molecular basis for TPR domain-mediated regulation of protein phosphatase 5. *Embo J* **24**: 1-10
- Yang L, Takuno S, Waters ER, Gaut BS** (2011) Lowly expressed genes in *Arabidopsis thaliana* bear the signature of possible pseudogenization by promoter degradation. *Mol Biol Evol* **28**: 1193-1203

Referencias

- Yin Y, Wang ZY, Mora-Garcia S, Li J, Yoshida S, Asami T, Chory J** (2002) BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* **109**: 181-191
- Zhang L, Qi Z, Gao Y, Lee EY** (2008) Identification of the interaction sites of Inhibitor-3 for protein phosphatase-1. *Biochem Biophys Res Commun* **377**: 710-713
- Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, Wu Y, Tang H, Qian Q, Li J, Zhang H** (2012) Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc Natl Acad Sci U S A* **109**: 21534-21539

Fosfatasa de proteínas con dominios Kelch en *Arabidopsis thaliana*. Estudios funcionales, estructurales y evolutivos.

Lic. Gustavo Ariel Maselli

Anexo I

Secuencias de fosfatasa BSL usadas en esta tesis. Se usaron las siguientes bases de datos:

Aquilegia caerulea, *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Capsella rubella*, *Carica papaya*,
Chlamydomonas reinhardtii, *Cucumis sativus*, *Glycine max*, *Mimulus guttatus*, *Oryza sativa*,
Populus trichocarpa, *Prunus persica*, *Ricinus communis*, *Selaginella moellendorffii*, *Sorghum bicolor*,
Vitis vinifera: <http://www.phytozome.net/>

Brassica rapa: <http://brassicadb.org/> y <http://www.phytozome.net/>

Coccomyxa sp. C-169: <http://www.jgi.doe.gov/>

Cycas rumphii: contig entre las entradas de GeneBank EX926969, EX925683, EX925320 y EX928531

Marchantia polymorpha: GeneBank AP009096

Nuphar advena: <http://ancangio.uga.edu/>

Ostreococcus taurii y *Micromonas sp. RCC299*: <http://bioinformatics.psb.ugent.be/plaza/>

Physcomitrella patens: <http://www.cosmoss.org/>

Picea glauca: GeneBank BT104224 and BT105039

Pteridium aquilinum: (Der et al., 2011)

Spirogyra pratensis, *Coleochaete orbicularis* y *Klebsormidium flaccidum*: (Timme et al., 2012) y
http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/

Schrenkiella parvula: <http://thellungiella.org/>

Welwitschia mirabilis: contig entre las entradas de GeneBank DT603568 y DT584820

Las X indican huecos en las secuencias.

Amborella trichopoda

>gnl|Ambo_Trinity|comp8149_c0

MDVDSTMGTENEQQQPRDHSSSPAVDEHQQAQEQEHEQQEQQTPSLPSSSAAAAAPSSPQLQAAASVMGPRPAPTYSVVN
AIIEKKEDGPGPRCGHTLTAVAAVGEETPGYIGPRLILFGGATALEGNSSSTAPSGPPSSGGNAGIRLAGATADVHCYD
VLSNKWSRLTPLGEQPSRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGGPRYGHVMALV
GQRFLLAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDDLGLLLCGGRDANSVPLASAYGLAKH
RDGRWEWAIAPGVSPSPRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTTTPRTGRYSADAAG
GDASVELTRRCRHAAAAGVDLIFIYGGLRGGVLLDLDLVAEDLAAEETTTAASHAAAAAASNAQAGRSPGRYVYADERA
RHVIPEAVPDGAVVLGSPVAPPINGDMYTDISTENAVLQGRRLSKGVEESLVEASAAEAEAI SAALAAAKARQSNGEVEQL
PDRDRGAEATPSGKQLSAPAKSPDSVITNNVPPTGVRLLHHRVVAEETGGALGGMVRQLSIDQFENEGRRVSYGTPESA
NAARKLLDRQMSINSVPKVI AHLLKPRGWKPPVRRQFFLDCNEIADLCTAERIFSCEPSVLQIKAPVKIFGDLHGQFG
DLMRLFDEYGSPTAGDITYIDYFLGDYVDRGQHSLETITLLLALKVEYPHNVHLIRGNHEAADINALFGFRIEIERM
GERDGIWVWHRINRLFNLPLAALIEKKIICMHGGIGRSINHVEQIENIQRPISMEAGSVVLMDDLWSDPTENDSVEGLR
PNARGPLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLLITLFSATNYCGTANNAGAILVLGRDLVVVVKLI
HPVPPAITSPEHHEI EDTWMQELNAHRPPTPTRGRPQSGTDRGSLAWI

gnl|Ambo_Trinity|compl6037_c0

MGSKLRLCPAPLPASAYRVVDTSWETDEDESPGPRCGHTLTPVSTPSNSSSSSSGSHHYVGPRLILFGGATAIEMSSTGAGI
 RLAGVTNSVHSYDVLTRKWTRLQPSGEPSPRAAHAAAAGVTMVVIQGGIGPAGHSTDDIYVLDITSAKSRWHRVWVQGP
 GPGPRYGHAMDLVQRYLVAVSGNDVFWGALFLGRRVLSDAWVLDTTQKPYQWQRLNPEGDRPFARMYATVSARS DGML
 LLCGRDSSCLPLADAYGLLKHHRDGRWEWTLAPGIPSPRYQHAAAFVVGARLHVSGGVLRRGGRAVEGEATVSVLDTAAGV
 WVDNRGIVSSSRSNKSSVETDPMDDLHRCRHASASVGLRVYVYGGLRGDVLDDLLVAENSPVMTEVSSLIINTSEIPSG
 LRNRDLQNLKEPQTLASQPTTTGQSDIAASSGTSLDDKSLEKLV EASTAEAEAVSSVFEAAKAAAAAVHLEGTSDGQNP
 VVQTDDKSQVSHCSLDSSDEDALESVRLHPRAVVAKETVGGGLGGLVRLQSLDQFENESRRMSPWNMDQGYGAKRVYNK
 QKSPLGLHRKVISLRLRPNWKPPTNRSFFLDSYEVGELCYAAEQIFLQEPVTLQLRAPIKIFGDLHGQFGDLMRLFDEY
 GYPSTAGDITYIDYFLGDYVDRGQHSLETITLLLALKIEYEPHVHLIRGNHEAADINALFGFRIECIERMGENGDIWAW
 TRFNQLFNTPLAALIIEKKIICMHGGIGRSINFVEQIEKLERPITMDAGSIIILMDLLWSDPTENDSIEGLRPNARGPGLV
 TFGPDRVLDVDFCKKNRLQLIRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILVVGRDLVVVVKLIHPLPPPFQS
 PESSPDHALEETWMQELNIQRPPPTPTRGRPQPAHDRNSLAFI

Aquilegia coerulea

>Ac_009_00678

MEVDSSMMGPETTTTTINHQQDQRGHQNEHVSSTPSISISSDELEQSNSTPSTSPSTSVQONNNNNNQLQQQQAVGSGSAV
 VVAGPRPAPTYSIVNAVIEKKEDGPGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATALEGNSAAAGPPSPAGSAGIR
 LAGATADVHCYDVL SNKWSRLTPLGEQSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVWVQGP
 GPGPRYGHVMAVQRFLLAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGDGPPPCMYATASARS DGLLLLCGGRDANS
 VPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHAAAFVFNARLHVSGGALGGGRMVEDSSSIAVLDTAAGVWCDTKSVVTS
 PRTGRYSADAAGGDAAVELTRRCRHAASIGDLFIYGGGLRGGVLLDDLLVAEDLAAAETSAASQAAAVAASNNVQAGR
 LQGRYGFADDRTRQTATDAVSDGAVVLGSPVAPPVNGDMYTDISTENAVLQGPRLSKGVEYLVEAAAAEAEAINATLAA
 AKARQVNGDVEQLPDRDRGAEATPSGKQISTLKTTPDSAFSSITPPPGVRLHHRVVAEETGGALGGMVRLS IDQFENE
 GRRVSYGTPE SATAARKLLDRQMSINSVPKVI AHLLKPRGWKPPVRRQFFLDCNEIADLCDSTERIFASEPSVLQIKAP
 VKIFGDLHGQFGDLMRLFDEY GSPSTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEYPHNVHLIRGNHEADINA
 LFGFRIECIERMGERDGIWAWHRINRLFNWPLAALIIEKKIICMHGGIGRSINHVEQIESLQRPITMEAGSVVLMDDLWS
 DPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILV
 LGRDLVVVVKLIHPLPPAITSPEASPERHI EDTWMQELNANRPPTPTRGRPQVTNDRGSLAWI

>Ac_004_00578

MEIDPTSSSSTITTTATNSDQSLQLPDSEHLDDSSSSTLQPVLPPEPASPTPVI AVSGGIGPRPAPTYTVVKS VIEKKED
 GPGCRGHTLTAVSPVGEETPGYIGPRLILFGGATALEGNSSSSAAAASGGGSSSGASSPTGPPSSPVAGSAGIRLAGA
 TADVHCYDVL SNKWSRLPLGEQSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVWVQGP GPGPR
 YGHVMAVQRFLLAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGDGPPPCMYATASARS DGLLLLSGGRDANSVPLS
 SAYGLAKHRDGRWEWAIAPGVSPSPRYQHAAAFVVSARLHVSGGALGGGRMVEDSSSIAVLDTAAGVWCDTKSVVTTPTG
 RYSADAAGGDAAVELTRRCRHAASIGDLVFIYGGGLRGGILLDDLLVAEDLAAAETATHAAVTS DVQLGRPLGRYAI FEE
 RTRELIPEPVADGSSVVLGSPVARSMNGDIYTDINTGNAMLQGPGRHSGKVESLVEASAAEAEIASAALAAKARQVNGEV
 DQLPERDRGAEASPSGKQSTMTKMPDLATSNMPPPGVRLHHRVIAHLLKPRGWKPPVCRQFFLDCNEIADLCDSAER
 IFASEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEY GSPSTAGDISYIDYFLGDYVDRGQHSLETIVLLLALKVEYPHNV
 HLIRGNHEAADINALFGFRIECIERMGERDGIWAWHRINRLFNWPLAAVIEKKIICMHGGIGRSINHVEQIESLQRP
 IA MEAGSVVLMDDLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHILITLFSAT
 NYCGTANNAGAILVGRDLVVVVKLIHPLPPALTSPE TSPERHI EDSSWMQELNANRPPTPTRGRPQI INDRGSLAWI

>Ac_026_00105

MGSKPWLYPAPTYHSVETSWDIDDDAPGPRCGHSLTAVAATKSHGPRLILFGGATAIEGGASNSTAGIRLAGVTNSVYCY
 DVLTRKWTRIRPAGDPPSPRAAHAAASAVGTMVVFQGGIGPAGHSTDDLYVLDMTNERFKWHRVWVQDPGPGPRYGHAMDL
 VAQRYLVTVSGNDGKRVLSDAWALDTAQKPYRWLKNPEGDRPSPRMYATASARS DGMFLLCGGRDTAGTPLSDAYGLLM
 HRNGQEWETLAPGVSPSPRYQHAAAFVVGARLHVTTGSLRGRAIEGEAAIAVLDTAAGVWLDNRGIVTSSRTNKVPSDLD
 GSLELMRRCRHASASVGVRIYIYGGGLRGDVLDDLIVAENSPFQSEVGSPI LNSERTPTTSPFKTNQASSSSYAPPNSDE
 RPEIPGSGDMSVDKSI NTLVEASAAEAEAVSAVWQA AASTDPEEGTSGSDKRLLPETSSDGSDAEADVRLHSRAVVVA
 KETIGNLGLVRLQSLDQFENESRRMSPLNSDTTYPTKFFTRQKSPQGLHKKIISALLRPNWKAPTNRFFFLDSYEVG
 ELCYAAEQIFMHEPTVLQKAPIKVFGDLHGQFGDLMRLFDEY GPFSTAGDITYIDYFLGDYVDRGQHSLETITLLLAL
 KIEYPENVHLIRGNHEAADINALFGFRIECIERMGESDGIWAWTRFNQLFNWPLAALIIEKKIICMHGGIGRSIHSVEQI
 EKLERPITMDAGSIIILMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTDVDFCKKNLQLIRAHECVMDGFERFAQGG
 LITLFSATNYCGTANNAGAILVVGRGLVVVVKLIHPLPPPIQSPETSPERVMDDTWMQELNIQRPPPTPTRGRPQPDHRS
 SLAYI

Arabidopsis lyrata

>A16g38200

MGSKPWLHPAPQYKTLTFWDEDDAPGPRCAHTLTAVAATKTHGPRLILFGGATAIEGGSSSVPGIRLAGVTNSVHSYD
 VLTRKWTRLKPAGEPPSPRAAHAAAAGVTMVVFQGGIGPAGHSTDDLYVLDMTNDKFKWHRVWVQDGP GPRYGHVMDLV
 SQRYLVTVTGNDGKRALS DAWALDTAQKPYVWQRLNPDGDRPSARMYASGSARS DGMFLLCGGRDTLGAPLGDAYGLLMH
 RNGQEWETLAPGVAPSPRYQHAAAFVVGARLHVSGGVLRRGGRVIDAEASVAVLDTAAGVWLDNRGQVTSARGIKGQIDQDP

SFELMRRCRHGAASVGIRIYVHGGLRGDVLDDFLVAENSTFQSDISSPLLASDRTQSSSTPRFSYAARPPSGSEPTFAM
 SEGLSLDENSLEKLTEASAAEA EVASSVWRAAQLGSGTLDEEPSTSDASSPTVETTTDDIANEGDVRHLHPRAVVVAKETV
 GSLGGMVRQLSLDQFQNESRRMVP MNNSDVPQPTKKFTRQKSPQGLHKKVIAALLRPRNWKPPGNRKFLLDSYEVGELCY
 AAEQIFMHEQTVLQLKAPIKVFGLHGQFGDLMLRFDEYGF PSTAGDITYIDYFLG DYVDRGQHSLETITLLLALKIEY
 PENVHLIRGNHEAADINALFGFRLECIERMGENDIWAWTRFNQLFNYLPLAALIENKIIICMHGGIGRSISTVEQIEKIE
 RPITMDAGSLVMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTEFCKKNKLQLIIRAHECVMDGFERFAQQQLITL
 FSATNYCGTANNAGAILVGRGLVIVPKLIHPLPPPILSPENSPEHSGDDAWMQELNIQRPTPTRGRPQPDFDRSSLAY
 I

>A14g11030

MDLDSMVPENDQDPVATPENHQSPMEKEASEQSTEETGSESESASLTPPSPPSQQQPQVTTTVGPRCAPTYSVNNAI
 EKKEDGPGPRCGHTLTAVPAVGEEGTSYIGPRLILFGGATALEGNSGGTGTPT SAGSAGIRLAGATADVHCYDVL TNKW
 SRLTPYGEPPSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTDQQRPRWHRVVVQGP GPGPRYGHVMAVQRYLM
 AIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLCGGRDANSVPLASAYGLAKHRDGRWE
 WAIAPGVSPSARYQHA AVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTS PRGTGRYSADAAGGDASVE
 LTRRCRHA AAAVGD LFIYGGLRGGVLLD LLLVAEDLAAAETTSAASHAAAAAATNSPPGRSPGRYGFSDERTGELPES
 APDSVVLGSPVAPPVNGDMYTDISTENAMVPGTRRTSKGVEYLVEASAAEA EI SATLAAAKARQVNGEVELPDRDRGAE
 ATPSGKPSLSLIKPD SAVPNSVIPAGVRLHHRAVVAAETGGALGGMVRQLS IDQFENEGRRVSYGTPESATAARKLLDR
 QMSINSVPKVVVAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIF SSEPTVLQLKAPIKIFGDLHGQFGDLMLRFDEY
 GSPSTAGDISYIDYFLG DYVDRGQHSLETITLLLALKVEYQHNVHLIRGNHEAADINALFGFRIECIERMGERDGIWVW
 HRINRLFNWLP LAALIENKIIICMHGGIGRSINHVEQIENIQR PITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLV
 TFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGH LITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPPAITS
 PETS PERHI EDTWMQELNANRPPTPTRGRPQNPNDRGS LAWI

>A11g08060

MDEDSSMVA DNDQDREFQSLDGGQSPSPMERETPQQMNDQSPPEGGSVPTPPP SDPNPTSSPQQA AAVVQEQQPAQVVG
 PRCAPTYSVVDAMMDKEDGPGPRCGHTLTAVPAVGDEGTPGYIGPRLVLF GGATALEGNSGGTGTPT SAGSAGIRLAGA
 TADVHCYDVL SNKWTR LTPFGEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTDQQRPRWHRVVVQGP GPGPR
 YGHVMAVQRYLMA IGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLCGGRDANSVPLA
 SAYGLAKHRDGRWEA IAPGVSPSSRYQHA AVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTS PRGTG
 RYSADAAGGDASVELTRRCRHA AAAVGD LFIYGGLRGGVLLD LLLVAEDLAAAETTAAASHAAAAAATNSPPGR LPGRY
 GFSDERNRELS ESATDGA VVLGSPVAPPVNGDMHTDISPENAF LPGTRRTNKGVEYLVEASAAEA EI SATLAAAKARQV
 NGEVELPDRDCGAEATPSGKPTFSLIKPDSMGMSVTPAGIRLHHRAVVAAETGGALGGMVRQLS IDQFENEGRRVSYG
 TPESATAARKLLDRQMSINSVPKVI AHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFASEPTVLQLKAPIKIFGDL
 HGQFGDLMLRFDEY GSPSTAGDISYIDYFLG DYVDRGQHSLETISLLLALKVEYQHNVHLIRGNHEAADINALFGFRIE
 CIERMGERDGIWVWHRINRLFNWLP LAASIEKKIIICMHGGIGRSINHVEQIENIQR PITMEAGSIVLMDLLWSDPTENDS
 VEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGH LITLFSATNYCGTANNAGAILVLRDLV
 VPKLIHPLPPALSSPETS PERHI EDTWMQELNANRPATPTRGRPQNSNDRGGS LAW

>A11g02750

MASQSYSHQYPSPSYESIQTFYDTEGWPGPRCGHTLTGVFVNNSHQLILFGGSSTAVTTHDSSLPEISLEGVTNSVHS
 FDVLRKWLTRLNPDGDVPS PRACHAAALYGTLLLIQGGIGPSGSDGDVYMLDITNNKWLKLLVGG EAPSPRYGHVMDIA
 AQRWLVI FSGNNGNEILADT WALDTRGSFTWDR LNP SGNQPSGRMYASGTSREDGIFLLCGGIDHSGVTLGDTYGLKMDS
 VSQCAWTPVPVGVAPSPRYQHTAVFGGPKMHVIGGVLNRARLIYGEAVAAVLDTETGEWLDADRPETSMGAYRQNQYQLM
 RRYHHAAASFGTHMYVHGGIRDDVLLD LLLVAETSQSSSPELEEDNPENYMLLDDFLMDEIPKPLGSEPEAPSFIRRSSS
 ESAMERLAEAYNLP TIENAFYDSEIEGYVPLQHGAE SDGNRGGVLRVTVSLDRFANEDQQTNI LMPQF IQNSTQDLHKKV
 ISTLLRPKTWMP PANRDFFLSYLEV KYLCDEVEKIEFKHEPTLLQLKVP I KVF GDIHGQYGD LMLRFHEYGYP SVEGDITH
 IDYFLG DYVDRGRHSLETIMLLFALKIEYPKNIHLIRGNHESLAMNRIYGF LKECKERMGDS DGF EAWL KINQVFDHLP
 LAALLEKRVLCVHGGIGRADTVEQIEKIKRPVFPDTTESIVLKDILWSDPTRNDTVLGIVENPRGEGVVSFGPDIVKAFL
 ERNGLEMILRAHECVIDGFERFAQGR LITVFSATNYCGLAQNAGAILVIGRDMVIYPKMIHPNPPPISSSEEDSPDKAGM
 QELNIEKPPTPARGESSE

Arabidopsis thaliana

>At4g03080

MGSKPWLHPAPQYKTLETFWDD EDDAPGPRCAHTLTAVAATKTHGPR LILFGGATAIEGGSSVPGIRLAGVTNTVHSYD
 ILTRKWTR LKPAGEPPSPRAHAAA AVGTMVVFQGGIGPAGHSTDDLYVLDMTNDKFKWHRVVVQGDGPGPRYGHVMDLV
 SQRYLVTVTGNDGKRALS DAWALDTAQKPYVWQRLNPDGDRPSARMYASGSARS DGMFLLCGGRD TLGAPLGDAYGLLMH
 RNGQWEWTLAPGVAPSPRYQHA AVFVGARLHVSGGVL RGGRVIDAEASVAVLDTAAGVWLD RNGQVTSARGSKGQIDQDP
 SFELMRRCRHGAASVGIRIYVHGGLRGDVLDDFLVAENSTFQSDISSPLLASDRTQSSSTPRFSYAARPPSGSEPSFSM
 SEGLSLDENSLEKLTEASAAEA EVASSVWRAAQLGAGTLDEEPSTSDASSPIVETTTDGTANEGDVRHLHPRAVVVAKETV
 GSLGGMVRQLSLDQFQNESRRMVP MNNSDVPQPTKKFTRQKSPQGLHKKVIAALLRPRNWKPPGNRKFLLDSYEVGELCY
 AAEQIFMHEQTVLQLKAPIKVFGLHGQFGDLMLRFDEYGF PSTAGDITYIDYFLG DYVDRGQHSLETITLLLALKIEY
 PENVHLIRGNHEAADINALFGFRLECIERMGENDIWAWTRFNQLFNYLPLAALIENKIIICMHGGIGRSISTVEQIEKIE
 RPITMDAGSLVMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTEFCKRNKLQLIIRAHECVMDGFERFAQQQLITL

FSATNYCGTANNAGAILVVRGLVIVPKLIHPLPPPILSPENSPEHSGDDAWMQELNIQRPTPTRGRPQPDFDRSSLAY
I

>At1g08420

MDESSMVADNDQDREFQSLDGGQSPSPMERETPQQMNDQSPPEGGSVPTPPSPDPNPATSQQQAAAVVGQEQQPALVV
GPRCAPTYSVVDAMMDKKEDGPGPRCGHTLTAVPAVGEDETPGYIGPRLVLFGGATALEGNSGGTGTPTSAGSAGIRLAG
ATADVHCYDVL SNKWTRLTTPFGEPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGP
RYGHVMALVGQRYLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCCGGRDANSVPL
ASAYGLAKHRDGRWEWA IAPGVSPSSRYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSPT
GRYSADAAGGDASVELTRRCRHA AAAVGD LIF IYGGLRGGVLLD LLLVAEDLAAAETT YAASHAAAAAATNSPPGRLPGR
YGFSDERNRELSESAADGAVVLGSPVAPPVNGDMHTDISPENALLPGTRRTNKGVEYLVEASAAEAEAI SATLAAAKARQ
VNGEVELPDRDCGAEATPSGKPTFSLIKPDSMGSM SVTPAGIRLHHRVAVVAAETGGALGGMVRQLSIDQFENEGRVSY
GTPESATAARKLLDRQMSINSVPK VIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFASEPTVLQKAPIKIFGD
LHGQFGDLMLRFDEYGSPTAGDISYIDYFLG DYVDRGQHSLETISLLLALKVEYQHNVHLIRGNHEAADINALFGFRI
ECIERMGERDGIWVWHRINRLFNWPLAASIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVLMDLLWSDPTEND
SVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLV
VVPKLIHPLPALSSPETS PERHIEDTWMQELNANRPATPTRGRPQNSNDRGGS LAW M

>At2g27210

MDLDSMVPENDQDPIATSENQSPMEEKEEASEQQTGSESESASLTPLSPPPSQQQQQQQQPQVTA VVGPRCAPTYSVV
NAIIEKKEDGPGPRCGHTLTAVPAVGEETSSYIGPRLILFGGATALEGNSGGTGTPTSAGSAGIRLAGATADVHCYDVL
SNKWSRLTPYGEPPSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPGRYGHVMALVGQ
RYLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCCGGRDANSVPLASAYGLAKHRD
GRWEWA IAPGVSPSARYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSPTGRYSADAAGGD
ASVELTRRCRHA AAAVGD LIF IYGGLRGGVLLD LLLVAEDLAAAETT SAASHAAAAAATNTPPGRSPGRYGFSDERTGE
LPESAPDAVVLGSPVAPPVNGDMYTDISTENAMVPGIRRTSKGVEYLVEASAAEAEAI SATLAAAKARQVNGEVELPDRD
RGAEATPSGKPSLSLIKPD SAVPNSVIPAGVRLHHRVAVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARK
LLDRQMSINSVPK VIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPTVLQKAPIKIFGDLHGQFGDLMLR
LDEYGSPTAGDISYIDYFLG DYVDRGQHSLETITLLALKVEYQHNVHLIRGNHEAADINALFGFRIECIERMGERDGI
WVWHRINRLFNWPLAALI EKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARG
PGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPP
AITSPETS PERHIEDTWMQELNVNRPPTPTRGRPQNPNDRGS LAW I

>At1g03445

MAPDQSYQYSPSYESIQTIFYDDEDWPGPRCGHTLTAVFVNNSHQLILFGGSTTAVANHNSSLPEISLDGVTNSVHSFD
VLTRKWTRLNPIGDVPSPRACHAAALYGTLLI LIQGGIGPSGSDGDVYMLDMTNNKWKIFLVGGETPSPRYGHVMDIAAQ
RWLVIFSGNNGNEI LDDT WALDTRGPF SWDR LNPSGNQPSGRMYASGSSREDGIFLLCGGIDHSGVTLGDTYGLKMDSDN
VWTPVPAVAPSPRYQHTAVFGGSKLHVIGGILNRARLIDGEAVVAVLDTETGEWVDTNQPETSASGANRQNQYQLMRRCH
HAAASFGSHLYVHGGIREDVLLD LLLVAETSQSSSPEPEEDNPDNYMLLDDYLMDEPKPLSSEPEASSFIMRSTSEIAMD
RLAEAHNLP TIENAFYDS AIEGYVPLQHGAETVGNRGLVRTASLDQSTQDLHKKVI STLLRPKTTWTPPANRDFFLSYLE
VKHLCDVEKIFMNEPTLLQLKVP IKVFGDIHGQYGDMLRFHEYGHPSVEGDITHIDYFLG DYVDRGQHSLEIIMLLF
ALKIEYPKNIHLIRGNHESLAMNRIYGFLECEERMGESYGF EAWLKINQVFDYLP LAALLEKKVLCVHGGIGRAVTIEE
IENIERPAFPDTGSMVLKDI LWSPTMNDTVLGI VDNARGEVVSFGPDIVKAFLE RNGLMILRAHECVIDGFERFADG
RLITVFSATNYCGTAQ NAGAILVIGRDMVIYPKLIHPPHPPISSEEDYTDKAWMQELNIEMPPTPARGESSE

Brassica rapa

>Bra031626

MDESSMLPETDQDRDSRSPDEAQSPLPMEREANRDESDSGSAP EAASSQE QPPQVVGPRCAPTYSVVDALMDKKEDGPG
PRCGHTLTAVPAVGEETPGYIGPRLVLFGGATALEGNSGGTGTPTSATPSAGIRLAGATADVHCYDVL SNKWTRLTTPFG
EPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPGRYGHVMALVGQRYLMAIGGNDG
KRPLADVWALDTAAEPYEWRKLEPEGEGPPPCMYATASARSDGLLLCCGGRDANSVPLASAYGLAKHRDGRWEWA IAPGV
SPSARYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSPTGRYSADAAGGDASVELTRRCR
AAAAVGD LIF IYGGLRGGVLLDGLLVAEDLSAAETT YAASHAAAAAATNSPPGRLPGRYGFSDERNRELSESAADGAVVL
GSPVAPPVNGDMYTDISSENAVLPGTRRTSKGVEYLVEASAAEAEAI SDTLAAAKARQVNGEVELPDSGGETSPSGTPTF
SLKPD SMGSTGGTPAGIRLHHRVAVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLLDRQMSINSVPK
VIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSNEPTVLQKAPIKIFGDLHGQFGDLMLRFDEYGSPTAGDIS
YIDYFLG DYVDRGQHSLEVISLLLALKVEYQHNVHLIRGNHEAADINALFGFRIECIERMGERDGIWVWHRINRLFNW
PLAASIEKKIICMHGGIGRSINHVEQIENIQRPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEF
CNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPPAISSTETPEGHTED
TWMQELNANRPATPTRGRPQTANDKGGSLAW M

>Bra030724

MDESSMVPEHHDH QHDEAQSQRDASGDGGGSM PAFSSPEPSQQEAVGGGEAQSAQVVVGPRCAPPYSVVDALMDKKEDG
PGPRCGHTLTAVPAVGEETPGYIGPRLVLFGGATALEDNSGGTGTPTSAA PSAGIRLAGATADVHCYDVL SNKWTRLTTP
FGEPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPGRYGHVMALVGQRYLMAIGGN
DGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCCGGRDANSVPLASAYGLAKHRDGRWEWA IAP

GVSPSARYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDDTKSVVTSPTGRYSADAAGGDASVELTRRC
 RHAAAAGVGLIF IYGGLRGGVLLDGLLVAEDLAAAETTYAASHPAAAAATNSAPGRLPGRYGFSDERNRELSDSAADGAV
 VLESVPVAPPVNGDLHNDISSENAINPGIRRTSKGVEYLVEASAAEAEAIRATLAAAKARQVNGEVELPDRDCGGEASPSG
 TPTFSLMKPDSMGSAGIRLHHRVVAEETGGALGGMARQFSIDQFENEGRRVSYGTPESATAARKLLDRQMSISSVPPK
 VIAHLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSNEPTVLQKAPIKIFGDLHGQFGDLMRLFDEYGSPTAGDIS
 YIDYFLGDYVDRGQHSLETICLLLALKVEYQHNHVLIRGNHEAADINALFGFRLECIERMGERDGIWVWHRINRLFNL
 PLAASIEKKILCMHGGIGRSINHVEQIESIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEF
 CNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPPALSSTEASHPDDTW
 MQELNANRPATPTIGRPQAPNDKGGSLAWM

>Bra034332

MDSDSMAPEENDQDPASLNHQPPPQDSSSESPAQTGSESATAFTPTPAAASAQQQPPPVAGPRCAPPYSVNVNAILEKKE
 DGPGRPCGHTLTAVPAVGEERSANYIGPRLILFGGATALEGNSGGTGTPTSAGSAGIRLAGATADVHCYDVLTKNWSRL
 TPHGEPSPRAAHVAVGTMMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMAVVGQRYLMAIG
 GNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPTCMYATASARSDDLCCGGRDANSVPLASAYGLAKHRDGRWEWAI
 APGVSPSARYQHAAVFVNARLHVSGGALGGGRMVEDSSIAVLDTAAGVWCDDTKSVVTSPTGRYSADAAGGDASGELTR
 RCRHAAAAGVGLIF IYGGLRGGVLLDGLLVAEDLAAAETTSASHAAAAAATNSPPGRSAGRYGFSDERTGELPESAPDA
 VVLGSPVAPPVNGDMYTDISTENAMPLGTRRTSKGVEYLVEASAAEAEAISATLAAAKARQVNGEVEQPDRDRGAEATPS
 GKPSSSLIKPDSAVSTSVIPAGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRRVGYGTPESATAARKLLDRQOSI
 NSVPPKVVSHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPTVLQLRAPIKIFGDLHGQFGDLMRLFDEYGS
 TAGDISYIDYFLGDYVDRGQHSLETITLLLALKVEYQHNHVLIRGNHEAADINALFGFRLECIERMGERDGIWVWHRIN
 RLFNLPLAALIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGP
 DRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPPAITSPTS
 PERHIDDTWMQELNANRPPTPTRGRPQTPDDRGSRAWI

>Bra000846

MAKPWLHPAPQYKTLTFWDEEDDAPGPRCAHTLTAVAATKTHGPRLILFGGATAIEGGNSVPGIRLAGVTNTVHSYD
 VLTRKWTRLKPAGEPPSPRAAHAAAAGVGMVVFQGGIGPAGHSTDDLYVLDMTNDKYKWHRVVVQEGGPGPRYGHVMDLV
 SQRYLVTVTGNDGKRALSADAWALDTAQKPYVWQRLNPDGDRPSARMYASGSARSDDGMFLLCGGRDTLGLAPLGDAYGLLMH
 RNGQWETLAPGVAPSPRYQHAAVFVGARLHVSGGVLRGGRMIDAEATVAVLDTAAGVWLDNRNGQVTSARGSKGQVDQDP
 SFELMRRCRHGAASVGIRIYVHGGLRGDVLDDFLVAENSTFQSDISSPLSATQQSSTPRFSYAARPPSGSEPALSLSEG
 LRLDENSMEKLTASAAEAQVASSVWRAAQLDSTSLDEEPSASDGSSATVETSDGPETEGDVRLHPRAVVGAKEGGS
 GGMGRQLSLDQFQNESRRMVPMMNSDVPQPTKFKTRQKSPQGLHKKVISALLRPNWKPKNRKFLLDSYEVGELCYAAE
 QIFMHEQTVLQKAPIKVFGLHGQFGDLMRLFDEYGFSTAGDITYIDYLFWDYVARGQHSLETITLLLALKIEYPEN
 FHLIRGNHEAADINALFGFRLECIERMGESDGIWAWTRFNQLFNHLPLAALIENKIICMHGGIGRSISTVEQIEKIERPI
 TMDAGSLVMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTEFCCKKNLQLIIRAHECVMDGFERFAQGQLITLFS
 ATNYCGTANNAGAILVGRGLVIVPKLIHPLPPPILSPENSPEHSGDDAWMQELNIQRPPPTPTRGRPQPDFDRSSLAYI

>Bra018520

MGTKPFVHPAPQYKTIETFWDEEDDAPGPRCAHTLTAVAATKTHGPRLILFGGATAIEGGNSVPGIRLAGVTNSVHSY
 DVLTRKWTRLKPAGEPPSPRAAHAAAAGVGMVVFQGGIGPAGHSTDDLYVLDMTNDKYKWHRVVVQEGGPGPRYGHVMDL
 VSQRYLVTVTGNDGKRALSADAWALDTAQKPYAWQRLNPDGDRPSARMYASGSARSDDGMFLLCGGRDTLGLVPPQGDAYGLLM
 HRNGQWETLAPGVAPSPRYQHAAVFVGARLHVSGGVLRGGRVDAEACVAVLDTAAGVWLDKNGQVTSARGSKVQMDQD
 PSFELMRRCRHGAASVGIRIYVHGGLRGDVLDDFLVAENSTFQSDISSPLSASSDRTQQSSTPRFSYAARPPSGSEPAS
 EGMSLDENSLEKLTASAAEAQVASSVWRAAQLSSASLDEEPSASDGSSPTVETITDYPETEEDVRLHPRAVVVGKETVG
 SLGGMVRQLSLDQFQNESRRMVPMMNSDVPQPTKFKTRQKSPQGLHKKVISALLRPNWKPKNRKFLLDSYEVGELCYA
 AEQIFMHEQTVLQKAPIKVFGLHGQFGDLMRLFDEYGFSTAGDVITYIDYLFGLGDYVDRGQHSLETITLLLALKIEYP
 ENVHLIRGNHEAADINALFGFRLECIERMGENGDIWAWTRFNQLFNHLPLAALIENKIICMHGGIGRSISSVEQIEKIER
 PITMDAGSLIIMDLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTEFCCKKNLQLIIRAHECVMDGFERFAQGQLITLFS
 ATNYCGTANNAGAILVGRGLVIVPKLIHPLPPPILSPENSPEHSGDDAWMQELNIQRPPPTPTRGRPQPDFDRSSLAYI

>Bra030478

MADLNLQRFPFPYATSTTQTLFNPQEDLPGPRCGHTLTAVYHRLILFGGVTLVPSGGSDNSLQCLTNSVYCLNSLTKKW
 TRVYPEGEPSPRTNHAAVSCGDGVI IQGGIGPSGICNGDLHVLDMSDFTFKWKKVVVQEGMAPCPRYGHVMDLAGEKLV
 IFGGINGNLVLDLWALNTTRHPNVWEILYPYGLPCGRVYASASSTRDHRYFMLNGGRDQHGMPGLDITYKLEPFSSSGF
 WWTRSPHLDSLKRQHAAVFVNRLHVFGGALSNTHLVDAEEAVSVLDTSTGVVVDTSNEVMRRSLHAAASIGSRIYVY
 GGIREGVLLDGLLISGELLSSGPTVPSFFDMNDGYKPNFSREKLHDLVNVKVIITLLRPQTWEPPVDRKFFLSFPELAELC
 FAAKQIEQEPTVLQLYAPIKVFGLHGQFGDLMRLFYEYGYPSRQGDIAIDYLFGLGDYVDRGKHSLETIILLLALKVE
 YPKNIHLIRGNHEAEATNTVYGFLEDECIERIECNDRAFKLINDFFSHLPLAALIEKKIFCVHGGIGSSVFTMEQITNI
 KRPVMDMCEYNYKVVKDLLWSDPTAHDSILGIGVNERGSHIVSFGPDRVNAFCERNIDMIIRGHECVLDGFERFAQGKL
 ITVFSATNYCGKFNAGAILVIGRNLVVPKLIHPLPPPILSPENVPDKAWIEVDRERPPPTIRGQRKPDFEQGSTS

Capsella rubella

>Carubv10012319m

MASRRSYPPSPYAYGSSRTFPSSVAACPGPRCGHTLTAAIANVNGSLLLVLFGGSTTAVAAHGLPNISLEGVTNSVHC
 FDVLRSWTRINPTGDVPSPRACHAATSLGTMILVQGGIGPSGSDGDVYTLDLTRNNLKWHRMLMVSGAAGPGRYGHVMD

VAAQRWLVI FGGRSGNTILSDTWALD TTGPVAVNRLNPNYGSRPSARIYASAKARE DGVFFLCGGTDYSGMPLGD TYALEI
 QSVARCGWTRVPGVAPSPRYQH TVVFE GQRMHVI GVS LINGARLIDGDAIVAVLDTGDSKWL DITGPETSSTSGSNRQIQY
 QFMRRFHHASASVGSYMYVHGGLKEDILLDDLQVAGTSQLHLISPEPEEPEDNPENHMLVDDDDPETSNSPSSFDPEPEVPS
 LTRRSSSEGSLEKRTVEEYNLPSLENAFYDCSETEGCFVLVLETAMDVGNPRDLGRTPSLDHLYLQSQRDQFGQHNSHPL
 HKRVISKLLKPKDWEAPADRTFFLSLLEVE TLCAVHQIFIQEPTVLQKVPKVF GDIHGQFGDLMRIFREYGSPSLRG
 DITHIDYLF LGDYVDRGTHSLETIMLLLCLKIEYPENIHLIRGNHESSTNRLYGFLHECQQRMGEADGIQAWYRINKLF
 GYLPLGAVIEKKILCVHGGIGRTVSVVEEIASIQRPAPFDSGSTVLKDTLWSDPTVGD SVMGITSNTRGEGVISFGPDIVK
 EFCKRNKLQMIIRAHECVYNGFETFAQGQLITVFSATNYCGTVKNAGAILVIGKDLVIYPKVIHPDPPSISSESSPEDN
 AWMQELNMRPPTPPREDF

>Carubv10008182m

MDDSSMVADNDQDREFQSLDGSQSPSPMERETPPQMNQSPPEGGSVPPPPSSDPNPTSSQPQPAAPVVGQEQQPPQVV
 GPRCAPTYSVVDALMDKEDGPGPRCGHTLTAVPAVGE GTPSYIGPRLVLFGGATALEGNSGGTGTPTSAGSAGIRLAG
 ATADVHCYDVL SNKWSRLTPFGEPTPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGP
 RYGHVMALVGQRYLMAIGGNDGKRPLADVWALD TAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLLCGGRDANSVPL
 ASAYGLAKHRDGRWEWAIAPGVSPSSRYQHA AVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTTPT
 GRYSDAAGGDASVELTRRCRHA AAAVGD LIF IYGG LRGVLLDDLLVAEDLAAAETTYAASHAAAAAATNSPPGRLPGR
 YGFSDERNRELSESASDGA VVLSPVAPPVNGDMHTDI SENALLPGTRRTNKGVEYLVEASAAEAEAI SATLAAAKARQ
 VNGEVELSDRDGGA EATPSGKPTFSLIKPDSMG SMSVSPAGIRLHHRVVVA AETGGALGGMVRQLSIDQFENEGRVSY
 GTPESATAARKLLDRQMSINSVPK VIAHLLKPRGWKPPVRRQFFLDCNEIADLCDSAERIFASEPTVLQKAPIKIFGD
 LHGQFGDLMLR LDFEY GSPSTAGDISYIDYLF LGDYVDRGQHSLETISLLLALKVEYQHNVHLIRGNHEAADINALFGFRI
 ECIERMGERDGIWVHRINR LFNWLP LAASIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVLMDLLWSDPTEND
 SVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLV
 VVVKLIHPLPPALSSPETS PERHIEDTWMQELNANRPATPTRGRPQNANDRGGSLAWM

>Carubv10003201m

MGSKPWLHPAPQYK TLETFWDEEEDAPGPRCAHTLTAVAATKTHGPR LILFGGATAIEGGSSSVPGIRLAGVTNSVHSYD
 VLTRKWTRLKPAGEPPSPRAAHAAA AVGTMVVFQGGIGPAGHSTDDLYVLDMTNDKFKWHRVVVQEGGPGPRYGHVMDLV
 SQRYLVTVTGNDGKRALSDAWALDTAQKPYVWQRLNPDGDRPSARMYATGSARS DGMFLLCGGRDTV GAPLGDAYGLLMH
 RNGQWEWTLAPGVAPSPRYQHA AVFVGARLHVSGGVLRGGRVIDAEASVAVLDTAAGVWLDNRNGQVT SARGNKGQIDQDP
 SFELMRRCRHGAASVGIRIYVHGG LRGDVLDDDFLVAENSTFQSDISSPLASDRTQSSSTPRFSYAARPPSGAEP AIPM
 SEGLSLDENSLEKLTEASAAEAEVASSVWRAAQLGSGTLDEEPSTSDANSVPAVESTTDTGTEGTVRLHPRAVVAKET
 VGS LGGMVRQLSLDQFQNESRRMVP MNNSDVLQPTK KFTROKSPQGLHKKVI SALLRPRNWKPPGNRKFFLDSYEVGELC
 YAAEQIFMHEQTVLQKAPIKVF GDLHGQFGDLMLR LDFEY GYPSTAGDITYIDYLF LGDYVDRGQHSLETITLLALKIE
 YPENVHLIRGNHEAADINALFGFRLECIERMGENDGIWAWTRFNQLFNYLPLAALIENKIICMHGGIGRSISSVEQIEKI
 ERPITMDAGSLVMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTEFCCKNKLQLIIRAHECVMDGFERFAQGQLIT
 LFSATNYCGTANNAGAILVVRGLVIVPKLIHPLPPPILSPENSPESHGDDAWMQELNIQRPTPTRGRPQPDFDRSSLA
 YI

>Carubv10024614m

MDLSSMVSENDQDPVTTTASSPMEKEASEQSTEETGSESESASLTPSSPSSQGGGQQQQVTPVVGPRCAPTYSV VNA
 VIEKKEDGPGPRCGHTLTAVPAVGE EGTSCYIGPRLILFGGATALEGNSGGTGTPTSAGSAGIRLAGATADVHCYDVL TN
 KWSRLTPYGEPPSPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPGRYGHVMALVGQRY
 LMAIGGNDGKRPLADVWALD TAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLLCGGRDANSVPLASAYGLAKHRDGR
 WEWAIAPGVSPSPRYQHA AVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSPTGRYSADAPGGDAS
 VELTRRCRHA AAAVGD LIF IYGG LRGVLLDDLLVAEDLAAAETTYAASHAAAAAATNSQPGRSPGRYGFSDERTGEP
 ESAPDAVVLGSPVAPPVNGDMYTDISTENALVPGTRRTSKGVEYLVEASAAEAEAI SATLAAAKARQYNGVEVELPDRDRG
 AEATPSGKASLSLIKPD SAVPNSVIPAGVRLHHRVVVA AETGGALGGMVRQLSIDQFENEGRVSYGTPESATAARKLL
 DRQMSINSVPKVV AHLKPRGWKPPVRRQFFLDCNEIADLCDSAERIFSSEPTVLQLRAPIKIFGDLHGQFGDLMLR LDF
 EY GSPSTAGDISYIDYLF LGDYVDRGQHSLETITLLALKVEYQHNVHLIRGNHEAADINALFGFRLECIERMGERDGIW
 VVHRINR LFNWLP LAALIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGP
 LVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAI
 TSPETS PERHIEDTWMQELNANRPPTPTRGRPQNPNDRGSLAWI

Carica papaya

>Cp00103G00160

XXXVVGPRCAPTYSV VNAVIEKKEDGPGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATALEGNSAASGTPSSAGSAG
 IRLAGATADVHCYDVL TNKWSRITPYGEPTPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQ
 PGPGRYGHVMALVGQRYLMAIGGNDGKRPLADVWALD TAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLLCGGRDA
 SSVPLASAYGLAKHRDGRWEWAIAPGVSPSARYQHA AVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVI
 TSPRTGRYSADAAGGDASVELTRRCRHA AAAVGD LIF IYGG LRGVLLDDLLVAEDLAAAETTYAASHAAAAAATNVAP
 GRLPGRYGFIDERTRQMPDAAPDGA VVVGAVVGNPVAPPVNGDMYTDISTENAMLQGSRRLSKGVEYLVEASAAEAEAI SATLA
 AAKARQVNGEVELPDRDRGAEATPSGKQITSLIKPD SAVSNNVAPPVRLHHRVVVA AETGGALGGMVRQLSIDQFENE
 GRRVSYGTPESATAARKLLDRQMSINSVPK VIAHLLKPRGWKPPVRRQFFLDCNEIADLCDNAERIFASEPSVLQLRAP
 IKIFGDLHGQFGDLMLR LDFEY GAPSTAGDIAYIDYLF LGDYVDRGQHSLETIALLLALKVEYPHNVHLIRGNHEAADINA

LFGRFRIECIERMGERDGIWAWHRINRLFNLPLAALIEKKIICMHGGIGRSINHVEQIENLQSPITMEAGSIVLMDLLWS
 DPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILV
 LGRDLVVVVKLIHPLPPAISSPETSPPERHIEDTWMQELNANRPPTPTRGRPQAANDRGSRAWI

>Cp00009G01420

MGSKSWLHPAPTYRPLETYWDTDEDAPGPRCGHTLTAVAATKTHGPRLLIFGGATAIEGGASASAPGIRLAGVTNSVHSY
 CVLTRKWTRIRPAGDPPSPRAAHAAGAVGTMVVFQGGIGPAGHSTDDLYVLDLTNEKCLKWHRVVVQGGPGPRYGHAMD
 VAQRYLTVSGNDGKRVLSDTWALDTTQKPYAWQKLNPEGDRPSARMYATASARSDGMFMLCGGRDSSGTPLSDAYGLLL
 HRNGQWETLAPGVAPSPRYQHAAXXDTAAGVWLDNRGLVTSRGGKGHAGHADQDPSLELMRRCRHASASVGVRIYVS
 GGLRGDVLDDDFLVAENSPFQSDINSPILTSEAPVTNPKINQYSLSPYTTPPSDSGQEALPSENMSMDKSSLEKLREA
 SAAEAEASAVWQAAQASASISAEETSVDSSSTAETSDGTDAEVDVRLHPRAVVVAKEAMGNLGGMVRQLSLDQFEN
 ESRRMVPVNNDLSPYTKKFSRQKSPQGLHKKVIISTLLRPNWKAPANRRFFLDSYEVGELCYAAEQLFMHEPTVLQKAP
 VKVFGDLHGQFGDLMRLFDEYGFSTAGDITYIDYFLGDIYVDRGQHSLETITLLLALKIEYPENVHLIRGNHEAADINA
 LFGRFRIECIERMGENDGIWAWTRFNQLFNYPPLAALVEKKIICMHGGIGRSIHSVEQIEKLERPITMDAGSIVLMDLLWS
 DPTENDSVEGLRPNARGPGLVTFGPDRVTFCKKNKLQLIIRAHECVMDGFERFAQQQLITLFSATNYCGTANNAGAILV
 IGRGLVVVVKLIHPLPPPLRSPETSPERIIDDAWMQELNIQRPTPTRGRPQPDLDRSSLAYI

Chlamydomonas reinhardtii

>Cr01g10530

MVDMSDLRPSVPRPRVFPVVGTAAGPRCGHTLTAVSIPGEGDLSKAKLVLFGGATALEGSSGKSESTQSTPGSAASGIR
 LAGATNDVHIFDVRTGKWEKTTTPAGEPPSPRAAHAAGAVGTMVVFQGGIGPAGLASEDLHVLDFDTPDRPRWHRVLSVGP
 GPSARYAHTLSLVANRFLVAMGGNDGKSTLGDALDWTSEKPYAWRKITDAGEMPSPRMYATAAARSDGLLLCGGRDIS
 GTPLDGAYGFARHRDGRWEWHSAPGAMPTGRYQHGAVFVGNRLHISGGAVGGGRMVDESTSTVVDLTTTHGCWVTPVGGAT
 GEDFTRRCRHAVASVGPVFIYGGKGSQLLDDLLVADDSNGTELSIFDPRSSAWQQYMETMHGNAAASRMLEKAAAEAA
 AAAAAANVRRVSAMDDLRCVDEQLDGTGAGGSRTSDSPSTNRSSSPDVSLGKDHIPGTTPTPDVKLWHRVAVVHQENSL
 RGLVRQLSIDQLDNEGRVSIYENGGGGGAKEKSPLFTRSMSVTNDGVEGVHKRVINELLRPTWKAPDDRRFLLNAQE
 IEELCNNAERIFREEPTVLDVRAPIKIFGDLHGQFGDLMRLFEEYGTPTSTAGDITYIDYFLGDIYVDRGSHSLETICLLL
 ALKIEHPRSVHLIRGNHEADINALFGFRLECLERLGDPEGVFVWRRINELFNYPPLAALIEGKILCMHGGIGRCLHKKID
 QISELKRPIITMEDGGPVLMDLLWSDPTTNDVSVQGVQPSRGPGLVTFGPDRVKEFCNNNLQMIVRAHECVMDGFERFAQ
 GHLITLFSATNYCGTAGNAGAILVGRDLVMVVKLIHPLPPNTPHTPDSSGGLDEDPTPHANDTWMQSIINDERPPTPP
 RGRPHMGSSLEYF

Citrus sinensis

>orange1.lg002713m

MNAKPWLHPAPSYRTLETYWDTDEDAPGPRCGHTLTAVAATKTTGPRLLIFGGATAIEGGATSSAPGIRLAGVTNSVHLY
 DVLTRKWTRIRPAGEPPSPRAAHAAGAVGTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGGPGPRYGHVMDL
 VSQRYLVSVSGNDGKRVLSDAWALDQAQKPYVWQRLNPEGDRPSARMYATASARSDGMFLLCGGRDASGAPLADAYGLLM
 HRNGQWETLAPGVAPSPRYQHAAVFVGARLHVTTGGALRGGRAIEGEAAVAVLDTAAGVWLDNRGLVTSRRTSKGHGEHD
 PSLELMRRCRHASASVGVRIYIYGGKGDILLDDFLVAENSPFQSDVNSPLTSEAPVTNPKINQYSLSPYTTPPSDSGQEALP
 LHSPFSRGLSMDKNSMEKLRASAAEAEANAVWQAAQASAVPPEETSVDSSSTAETSDGTDAEVDVRLHPRAVVVA
 AKEAVGNLGGMVRQLSLDQFENESRRMIPLNNELSPYTKKFTSRQKSPQGLHKKVIISTLLRPNWKAPANRRFFLDSYEVG
 ELCYAAEQIFMQEPTVQLRAPVKVFGDLHGQFGDLMRLFDEYGFSTAGDITYIDYFLGDIYVDRGQHSLETITLLLAL
 KIEYPENVHLIRGNHEADINALFGFRLECIERMGENDGIWAWTRFNQLFNCLPLAALIEKKIICMHGGIGRSIHSVEQI
 EKLERPITMDAGSIIIMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVSDFCRKNKLQLIIRAHECVMDGFERFAQQQ
 LITLFSATNYCGTANNAGAILVGRGLVVVVKLIHPLPPPLQSPETSPEVVIDMWMQELNIQRPTPTRGRPQPDLDRN
 SLAYI

>orange1.lg001895m

MDVDSNMVPEADHDPAVEKENTSEESMEREQLADTEQPSGSPASTTSQSQQQPQQHQQQQTQQTTPVVGPRCAPTYSVVNA
 VIEKKEDGPGPRCGHTLTAVAAGVEEGTPEGYIGPRLIFGGATALEGNSAASGTPSSAGSAGIRLAGATADVHCYDVLTN
 KWSRITPFGEPTPRAAHVATAVGTMVVFQGGIGPAGLSAEDLHVLDTLQQRPRWHRVVVQGGPGPRYGHVMAVGVQRY
 LMAIGGNDGKRPLADVWALDQAQKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCGGRDASSVPLASAYGLAKHRDGR
 WEWAIAPGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVDESSESVAVLDTAAGVWCDTKSVVTSRPTGRYSADAAGGDA
 VELTRRCRHAAGVGDLIYIYGGLRGGVLLDDLLVAEDLAAAEETTTAASHAAAAAASNVQSRLPGRYGFVDERTRQTI
 EAAPDGSVVLGNPVAPPINGDMYTDISTENAMLQGTTRRLTKGVEVSLVQASAAEAEIASATLAAVKARQVNGEVELPDRDR
 GAEATPSGKQMIKPSAGSINSIAPAGVRLHHRVAVVAEETGGALGGMVRQLSIDQFENEGRRVSYGTPESATAARKLLDR
 QMSINSVPKVIHLLKPRGWKPPVRRQFFLDCNEIADLCSAERIFSSPEVQLKAPIKIFGDLHGQFGDLMRLFDEY
 GSPSTAGDIAYIDYFLGDIYVDRGQHSLETITLLLALKIEYPENVHLIRGNHEADINALFGFRLECIERMGERDGIWAW
 HRINRLFNLPLAALIEKKIICMHGGIGRSINHVEQIENLQSPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLV
 TFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVGRDLVVVVKLIHPLPPAISS
 PETSPPERHIEDTWMQELNANRPPTPTRGRPQNDRVAFM

Coccomyxa spp

>Coccomyxa_spC-169

MEPDQAPPAPAPRPRPFPCVGNPPGPRCGHTLTAISGPEGEFSAAKLVFMFGGATALEGSQRTDGGPPASPGAPSGSGIRL
 AGATSDVHIMDVRSGKWEQVVPQGDPPSPRAAHAAAAVGSVMVIQGGIGPAGLASEDLHVLDFTEFDRPRWHRVMVQGGP
 PSARYAHTLALVANRFLVAMGGNDGKQTLADAWALDTSKPYQWRKITDAGDTPNARMYATAAARS DGLLLCGGRDAEG
 VPLGDAYGLARHRDGRWEWAAAPAGNMP TARYQHGA VFGARLHISGGAVGGGRMVDDQSSIVVLDTAAGTWCTQAEP
 MNGVADDGADDWARRRCRHA VAVGQYVFTYGGGLRGSALLDDFLLADDSSGSELSICDPRSGAWRQWLDAAHGTTAGAEM
 LAAAAEEAAAAALGVRRVGAVEDLRCLDENLDSVGRMAGRPEEDLSNSRGGSPSLSLHQRLDPRTPPTPDVRLYHR
 AVVVAQATVLLQEGTGSRLRGLVRLS IDQFENEGRRVMGAESAVKLPADGRTPIFTHSMDPGSIHKTVLRLLNPRQWKP
 PEDRSFILDGDIINTLCDQAERVFQEQEASVLRRLRAPVKIFGDLHGQFGDLMRLFEYEGTPTSTAGDITYIDYFLGDIYVDR
 GAHSLETICLLALKIEHPKSVHLIRGNHEAADINALFGFRLECIERMGDQQLWAWQRLNSLFWNLPLAALIEDRIICM
 HGGIGRSISRVDQIDVLRPLSMEDGGVVLMDLLWSDPTTNDGVEGVQPSPRGPGLVTFGPDRVKEFCCKMNLQMI VRAH
 ECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVMVVKLIHPLPPNTPHTPKSSDADDVEEENPTPHMPTDT
 WMAINEERPPTPRGRPSNGPNLAFFE

Coleochaete orbicularis

>Corbicularis_GB_J0251979

RQLSIDHLENEGRVSYGTPDGSSSAAAKRLLDRQMSIQGMHKKVLYLLKPRGWKPPVKRQFFLDAHDIIDLCEHAEGF
 FSQEPSVLQIRAPVKIFGDLHGQFGDLMRLFDEYGSPTSTAGDITYIDYFLGDIYVDRGQHSLETIALLLALKIEYPKNVH
 LIRGNHEAADINALFGFRLECIERMGEAAGIWAQRINQLFNWLPLAALIEKKIVCMHGGIGRSIHSIKQIEALQRPLTM
 EAGSVILMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCCKNNDLQLIVRAHECVMDGFERFAQGHLITLFSATN
 YCGTANNAGAILVLRDLVVPKLIHPLPPAGHSPHSSPDSQHEDTWMQELNVQRPPTPTRGRPQSESNRNSLAWI

Cucumis sativus

>Cucsa.184370

MASKPWLYPAPSYRALETYWDTDDAPGPRCGHTLTAVAATKSFGRLLIFFGGATAIEGGASSAAPGIRLAGVTNSVHSY
 DVLTRKWTRIRPAGEPPSPRAAHAAAAVGMTMVVFQGGIGPAGHSTDDLFLVLDLTNDKFKWHRVVVQGGPGPRYGHVMDL
 VAQRYLVS VSGNDGKRVLSDAWALD TAQKPYAWQRLNPEGRPSARMYATASARSDGMFLLCGGRDSTGAPLADAYGLLM
 HRNGQWETLAPGVSPSPRYQHA AVFVGARLHVTTGGTLRGGRAIEGEASIAVLDTAAGVWLDNRGLVTSTRGNKGHADYD
 SSLELMRRCRHAASVGVRIYIYGGLRGDVLLDDFLVAENSSFQSEISSPVITSERAPIITSSKVNQTNITPSTILPSSD
 GLENTSAGGTSMDKSSMEKLEASAAEAEASAVWHAVQAASVEDTSVSDDNSRAAETSDGSDYEADVRLHPRAVVVAK
 EAIGNLGMVRQLSLDQFENESRRMIPSNLDSYPTKKTFRQKSPQGLHKKVISILLRPNWKAPANRRFFLDSYEVGELC
 YAAEQIFMHEQTVLQLKAPVKVFGDLHGQFGDLMRLFDEYGFPTSTAGDITYIDYFLGDIYVDRGQHSLETITLLALKIE
 YPDNVHLIRGNHEAADINALFGFRLECIERMGESDGIWA WTRFNQLFNYLPLAALIEKKIICMHGGIGRSIQTVEQIEKL
 ERPITMDAGSIIIMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTEFCCKRNKLQLIRAHECVMDGFERFAQGLIT
 LFSATNYCGTANNAGAILVIGRGLVVPKLIHPLPPPLQSPETSPERMIEDTWMQELNIQRPPTPTRGRPQPDLDSSLA
 YI

>Cucsa.185930

MEVDSSMVPEPDHDPVAVQNHSTVSSTAAGVDREQPQAGGTSPOQMSLPPQQOQTTVQMOTTVVGPRLAPAYSVVNAIID
 KKEDGPGPRCGHTLTAVGSGVEEGTPGYIGLRLILFGGATALEGNSAASGTPSSAGSAGIRLAGATADVHCDFVLANKWS
 RITPLGEPPTPRAAHVATAVGMTMVIQGGIGPAGLSAEDLHVLDTLQQRPRWHRVVVQGGPGPRYGHVMALVGQRYLMA
 IGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCGGRDANSVPLASAYGLAKHRDGRWEW
 AIAPGVSPSPRYQHA AVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDIKSVVTTPTRTGRYSADAAGGDASVEL
 TRRCRHA AAAVGD LIF IYGGLRGGVLLDDLLVAEDMAAAET TSAASHAAAAAASVQPGRLPPRYGFTDERARQTMPEAA
 PDGSSVVLGNPVAPPVNGDMYTDISTENAMLQGGRRMSKGV EYLVEASAAEAEIASATFAAKARQVNGEVELPDRDRGAE
 ATPSGKQISSLIKPDSTGSNSIAPAGVRLHHRAVVAAETGGALGGMVRQLS IDQFENEGRRVSYGTPENATAARKLLDR
 QMSINSVPKKVIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAEKIFSSEPSVLQLKAPIKIFGDLHGQFGDLMRLFDEY
 GAPSTAGDIAYIDYFLGDIYVDRGQHSLETITLLALKVEYQQNVHLIRGNHEAADINALFGFRLECIERMGERDGIWA W
 HRINRFLFNWLPLAALIEKKIICMHGGIGRSINHVVQIENIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLV
 TFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPPAMTS
 PEASPERHLEDTWMQELNANRPPTPTRGRPQVTNDRGSLAWI

>Cucsa.356030

MEVDSSMLSEVDHDLAQDQT TASPPAEMEGDQDGEQSPGGNGSPVQSPATSQPQQSSTPQVQQISVVGPRHAPTYSVVNA
 LMDKKEDGPGPRCGHTLTAVSAVGEDGTPGYSGRLLIFFGGATALEGNSAAAGTPT SAGNAGIRLAGATADVHCYDVLTN
 KWTRVTPLEPPTPRAAHVATAVGMTMVIQGGIGPAGLSS EDLHVLDTLQQRPRWHRVVVQGGPGPRYGHVMALVGQRY
 LMVVGNDGKRPLTDVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCGGRDANSVPLASAYGLAKHRDGR
 WEWAIAPGVSPSARYQHA AVFNARLHVSGGALGGGRMVEDSSSIAVLDTAAGIWCDTKSVVTSPTGRYSADAAGGDAA
 VELTRRCRHA AAAVGD LIF IYGGLRGGVLLDDLLVAEDQAAEET TSAASHAAAAAATSCIQSGRLAGKCYGFNDRSRHAI
 HVVAPDGA VLVGNPVAPPVNGDMYTDISTENALLQGGRRMKNKGV EYLVEASAAEAEIASATLAAAKARQE QDNGEVELPD
 RDRGAEATPSGTQISSLIKPDSDALNNIVPSGVRLLHHRAVVIAAETSGALGGMVRQLS IDQFENEGRRVSYGTPESATAA
 RKLLDRQMSINSVPKKVIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAEKIFSSEPSVLELKAPIKIFGDLHGQFGDLM
 RLFDEY GAPSTAGDIAYIDYFLGDIYVDRGQHSLETITLLALKVEYPNNVHLIRGNHEAADINALFGFRLECIERMGER
 DGIWTVHRINRFLFNWLPLAALIEKKIICMHGGIGRSINHVEQIESLQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPN
 ARGPLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPI
 PPAISSPETS PERHMETWMQELNANRPPTPTRGRPQVANDRGSLAWM

Cycas rumphii

>Crumphii_contig_GB_EX925683_GB_EX928531_GB_EX925320_GB_EX926969
 AGQTEKAERLQGEQKHMDEASKKSEAPGPSLTSVDDTLETVDRLHHRVVIKAEKPVGSLGGLVLRQLSLDQFENEGRMS
 PLNTDNAYGARKLLHRQKSPIGLHKKVISCLLRPRNWKPPVNRQFFLDSYEVGEVCYAEQIFMQESTVIQLKAPVKIFG
 DLHGQFQGLMRLFDYGFSTAGDITYIDYFLGQYVDRGQHSLETITILLLALKIEYPEHVHMIRGNHEAADINALFGFR
 LECIERMGERDGIWAWTRFNQLFNWLPLAALIIEKKIICMHGGIGRSINFVEQIESLQRPITMDAGSIVLMDLLWSDPTEN
 DSVEGLRPNARGPLVTFGPDRVLDVDFCKKKNLQLIIRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGALLVLRDL
 VVVPKLIHPLPPPFQSSSESSPEHHSEETWMLQELNIQRPPPTPRERPHSAHDRSSLAFI

Eucalyptus grandis

>Egrandis_v1_0.002312m.g
 MGSKPWLYPAPTYRNLETYWDADAPLPRVGHLLTAVAAATKSHGPRLLIFGGATAIEGGAASSAPGIRLAGVTNSVHSY
 DILTKKWTRMKPAGDPPSPRAAHAAAAGVTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVLVQGGPGPRYGHVMDL
 VAQRVYLVSVSGNDGKRVLSDAWVLDTAQKPYVWRRNLNPEGDRPSARMYATACARS DGMFLLCGGRDSSGAPLADAYGLLM
 HRNGQWEWTLAPGVSPSPRYQHAAVFVGARLHVTTGGVLRGGRVGEAAIAVLDTAAGVWLDNRGLVTSRSTSKDHSQYD
 PTLLELMRRCRHAAAAGVRIYVYGGLRGDIILLDDLLVAENSLFQSDGSDPMLTSEASPISSPKLNAPRVVSATSPTS
 ALDSSSGMGMDKNSMEKLRASAAEAEANAVWQAVQTASSVPAEETSVEENSPAETASDGSDTEGDVRLHPRVAVV
 AKEAVGNLGGMVRQLSLDQFENESSRMVPMNELSYPTKFFTRQKSPQGLHKKI IATLLRPRNWKAPANRRFFLDSYEVG
 ELCYAAEQLFMHEPTVLQLKAPVKVFGDLHGQFQGLMRLFDYGFSTAGDITYIDYFLGQYVDRGQHSLETITLLLLAL
 KIEYPESVHLIRGNHEAADINALFGFRLECIERMGENGGIWAWTRFNQLFNYLPLAALIIEKKIICMHGGIGRSIHSVEQI
 EKLERPITMDAGSIILMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTDVDFCKKKNLQLIIRAHECVMDGFERFAQGG
 LITLFSATNYCGTANNAGAILVVRGLVIVPKLIHPLPPPLQSPETSPERVIEDTWMQELNIQRPPPTPRGRPQPDLD
 SLAYI

>Egrandis_v1_0.001476m.g
 MEVDSSMATDADHDPVQNHNNGPSPPPAEQPPAGGGSPAAPPQPPPSQQQAGVQQSPVVGPRPAPTYSVVNAILEKKED
 GPGPRCGHTLTAVAAVGEETPGYIGPRLLIFGGATALEGNSAASGAPSSAGNAGIRLAGATADVHCYDVLTKNWSRITP
 FGEPTPRAAHVACAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVWVQGGPGPRYGHVMALVGQRYLMAIGGN
 DGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLCGGRDASSVPLASAYGLAKHRDGRWEWAIAP
 GVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSSAVLDTAAGVWCDTKSVVTSRPTGRYSADAAGGDAAVELTRC
 RHAAAAGDLIFIYGGLRGGVLLDDLLVAEDLAAAETTTAASHAAAAAASSNSQSPRLTGRYGYVDERTRQMI EAVPDGA
 VVLGNPVAPPVNGDMYTDISTENAMLQGSRRMAKSVEDLVKASAAEAEAI SATLAAAKARQSNGEVESPRDRGAEATPS
 GKQLSSSVKIPDSAGPNI IAPAGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRVSYGTPESASAARKLLDRQMS
 INNVPKVI AHLKPRGWKPPVRRQFFLDCNEIADLDCSAEKIFSSEPSVLQLKAPIKIFGDLHGQFQGLMRLFDYEGAP
 STAGDIAYIDYFLGQYVDRGQHSLETITLLLLALKVEYPSNVHLIRGNHEAADINALFGFRIECIERMGERDGIWAWHRI
 NRLFNLPLAALIIEKKIICMHGGIGRSINHVEQIESIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFG
 PDRVMEFCNINDLQLIIRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPPAISSPET
 SPERHIEDTWMQELNANRPPTPRGRPQVATDRGSLAWI

Egrandis_v1_0.001931m.g
 MQGDSAMPPEADRDPPAENLSCDWPPPAVAAAAGGERAEQWRGGGSAPGTPRLHQLPLPQPVGPRHAPRYAVVDAI I
 EEEEDGPGQRCGHTLTAVAAVGDGSPGYMGPRLLIFGGATSLEGNPAASGPPSSPGGGGAGIRLAGAAADVHCYDILT
 NTKWSKITPLGEPSPRAAHVATAVGTMVVIQGGIGPDGLSAEDLHVLDLTSHRPRWQRVATQGGPGPRYGHVMALVGQRY
 LMAIGGTDGKRILADVWALDTAAKPYEWRKLEPEGEGPPPCTYATASARS DGLLLCGGRDASGVPLASAYGLAKHRDGR
 WEWAIAPGVSPSPRYQHAAVFVNARLHVSGGSLGGGRMVEDSSSAVLDTAAGVWCDTMSVVTSPRTGRYAVDAPGGDAA
 VHLMRRCRHAAAAGDLIYVYGGLRGGVLLDDLLVAEDLAAADTATAVSQAAAATENDDMYTDKTDHAMLSGSWRLNKG
 VEYLVEASAAEAEITAALAAAKARQANGVELPSRDHGVEATSIEKQVVLPAEAKALGSTVKGYPIDQFEKDGRI SVN
 TNAARKLFDVQMSKSNVKKVIAHLKPRGWKPPVCRQFFLDCNEIEDLDCDAERIFSAEPSVLQLKAPVKIFGDLHGQF
 GDLMRLFDYGSPSAAGDIAYIDYFLGQYVDRGQHSLETIGLLLALKVEYDPNIHLIRGNHEAADINALFGFRIECIER
 LGERDGIWAWHRI NRLFNLPLAALIENKILCMHGGIGRSINYVEQIENIPRITMEAGSILLMDLLWSDPTENDGVEGL
 RPNARGPGLVTFGPDRVMEFCNNDLELIVRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILVLRDLVVVPKLI
 HPLPPASSFPEASAECQAEDTWMQELNANRPPTPRGRPQVTKDRGSVAWI

Glycine max

>Gm2g17040
 MGSKPWLYPAPTYRTIETFWVTDEDAPGPRCGHTLTAVAAATKTQGPRLIFGGATAIEGGSSSAPGIRLAGVTNTVHSYD
 VQSRKWTSIKPAGEPPSPRAAHAAAAGVTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVWVQGGPGPRYGHVMDLV
 AQRVYLVTVSGNDGKRVVSDAWAFDTAQKPYVWQKLNPEGDRPSARMYATASARS DGMFLLCGGRDSSGAPLADAYGLLMH
 RNGQWEWTLAPGVSPSPRYQHAAVFVGARLHVTTGGVLRGGRSVEGEASIAVLDTAAGVWLDNRGIVSSSRNKGHDYDPS
 LELMRRCRHAAAAGVHIF IYGGLRGDIILLDDFLVAENSLQPDINS PERASPVTSSKQNSNLNYNVTTPNLDGGPDIP
 SSGSGMDKNSLEKLRASAAEAEASAVWQAVQAISSNPAEETSVDSDNSQAVETVSDGSDTEGDVRLHPRVAVVAKA
 LGNLGGMVRQLSLDQFENESSRMI PVNNDLPYPTKFFTRQKSPQGLHKKI IATLLRPRNWKAPANRRFFLDSYEVGELCY
 AAEQIFMHEPTVLQLKAPVKVFGDLHGQFQGLMRLFDYGFSTAGDITYIDYFLGQYVDRGQHSLETITLLLLALKIEY
 PENVHLIRGNHEAADINALFGFRIECIERMGENGGIWAWTRFNQLFNYLPLAALIIEKKIICMHGGIGRSINSVEQIEKLE

RPITMDTGSITLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTEFCKKNKLQLIIRAHECVMDGFERFAQQQLITL
FSATNYCGTANNAGAILVVGRGLVVVPKLIHPIPPPLQSPETSPERVMDETMQELNIQRPTPTRGRPPQDLDRGSLAY
I

>Gm10g02760

MGSKPWLYPAPTYRTIETVWDTDEDAPGPRCGHTLTAVAATKTLGPRLILFGGATAIEGGSSSAPGIRLAGVTNTVHAYD
VQSRKWTSVKPAGEPPSPRAHAAAAVGTMVVFQGGIGPAGHSTDDLVLVLDLTNDKYKWHRVVVVQGGPGPRYGHAMD
LV AQRYLVTVSGNDGKR VVSDAWALDTAQKPYVWQKLNPEGDRPSARMYATASARS DGMFLLCGGRDSSGAPLADAYGLLMH
RNGQWEWTLAPGVSPSPRYQHAAV FVGARLHVTTGGVLRGGRSVEGEASIAVLDTAAGVWLDNRNGIVSSSRSNKGDYDPP
LELMRRCRHAAAAVGVHIF IYGGLRGDILLDDFLVAENSPLQPDINSSERASPVTSKQNSLTYNVTAPNLDGGPDTP
SFGHSGMDKNSLEKLEASAAEAEASAVWQAVQSISSNPAEETSLSDENSQAQETVSDGSDTEGDVRLHPRAVVVAKEA
VGNLGGMVRQLSLDQFENESRRMIPVNNDLPYPTKKFTRQKSPQGLHKKIISTLLRPRNWKAPANRRFFLDSEVGE
LICY AAEQIFMHEPTVLQLKAPVKVFGDLHGQFGDLMLRFDEYGF PSTAGDITYIDYFLGDYVDRGQHSLETITLLL
LALKIEY PENVHLIRGNHEAADINALFGFRIECIERMGENDWIWA WTRFNQLFNYLPLAALIEKKIICMHGGIGRS
INSVEDIEKLRPITMDAGSIIIMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTEFCKKNKLQLIIRAHECVMDG
FERFAQQQLITL FSATNYCGTANNAGAILVVGRGLVVVPKLIHPIPPPLQSPETSPERGMDETMQELNIQRPTPTR
GRPPQDLDRGSLAY I

>Gm11g18090

MDVDSMVPEPDHDPVQVQNHAEERDQLGDSPPSPSLSEGGGSLAQSPQQQQQQQQPPATAAQVHQPNLSLVGPRLA
PTYS VVNAILEKKEDGPGPRCGHTLTAVA AVGEEGTSGYIGPRLILFGGATALEGNSAASGTPSSAGNAGIRLAGATAD
VHCYD VINSKWSRITPIGEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVGVPGPGPRYGH
VMALV GQRYLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLLCGGRDANSVPLSSAY
GLAKH RDGRWEWA IAPGVSPSPRYQHAAV FVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTS
PRGTGRYSADAAG GDAAVELTRRCRHA AAAIGDLIF IYGGLRGGVLLDDLLVAEDLAAAETTTAASHAAAAAASNVQAG
RLPGRYGFIDDRTRQTMP EAAASDGSVVLGNPVAPPVNGDMYTDISTENALLPGTRRTSKGVEYLVEASAAEAEAI
SATLAAAKARQVNGEVEL PDRDRGAEATPSGKQISSLIKPDSAGSNSIPGGVRLHHRVAVVAAETGGALGGMVRQLS
IDQFENEGRVSVYGTPE SAT AARKLLDRQMSINSVPKKVVAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPS
VLQLRAPIKIFGDLHGQFGDLMLRFDEYGFAPSTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEYPNNVHLIRGN
HEAADINALFGFRIECIERMERDGIWTRHINRFLFNWPLAALIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVL
MDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQQHLITLFSATNYCGTAN
NAGAILVLRDRLVVVPKLIHPLPPAISSPDTSPERHIEDTMQELNANRPPTPTRGRPQVTNDRGSLAWI

>Gm12g10120

MDVDSMVPEPDHDPVQVQNHAEERDQLGESPPSPSEGGGSPAQQHQPLPPTAAQVQPNPLVGPRLAPTYSVVNAILEK
KEDGPGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATALEGNSAVSGTPSSAGNAGIRLAGATADVHCYDVI
TNKWSRITPIGEPSSRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVGVPGPGPRYGHVMALV
GQRYLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLLCGGRDANSVPLSSAYGLAK
HRDGRWEWA IAPGVSPSPRYQHAAV FVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTS
PRGTGRYSADAAGGDAAVELTRRCRHA AAAIGDLIF IYGGLRGGVLLDDLLVAEDLAAAETTTAASHAAAAAASNVQAG
RLPGRYGFIDDRTRQTMPETD DGSVVLGNPVAPPVNGDMYTDISTENALLPGSRRTSKGVEYLVEASAAEAEAI
SATLAAAKARQVNGEVELPDRDRGAEATPSGKQISSLIKPDSAGSNSIPGGVRLHHRVAVVAAETGGALGGMVRQLS
IDQFENEGRVSVYGTPESTTAAARKLLDRQMSINSVPKKVVAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPS
VLQLRAPIKIFGDLHGQFGDLMLRFDEYGFAPSTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEYPNNVHLIRGN
HEAADINALFGFRIECIERMERDGIWTRHINRFLFNWPLAALIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVL
MDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQQHLITLFSATNYCGTAN
NAGAILVLRDRLVVVPKLIHPLPPAISSPETS PERHIEDTMQELNANRPPTPTRGRPQVTNDRGSLAWI

>Gm12g31540

MDVDSMVPEPDHDPVQVQNHAAADDPADAAEAGEQLGEPSSGGGSPEQTQPAPASAPQSPVVGPRLAPSYTVVNAILEK
KEDGPGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATALEGNSAATGTPSSAGNAGIRLAGATADVHCYDVL
TNKWSRITPFGEPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQWRWHRVSVQGGPGPRYGHVMALV
GQRYLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLLCGGRDANSVPLSSAYGLAK
HRDGRWEWA IAPGVSPSPRYQHAAV FVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDIKSVVTS
SRTGRYSADAAGGDAAVELTRRCRHA AAAVGD LIF IYGGLRGGVLLDDLLVAEDLAAAETTTAASHAAAAAASASNVQAG
RLPGRYGFADDRTRTRETMP EAAADGSVVLGNPVAPPVNGDIYTDISTENAMLQGPRTAKGVEYLVEASAAEAEAI
SAALAAAKARQENGEVELPDRDRGAEATPSGKQTS SLIKPDSAGSNITPGGVRLHHRVAVVAAETGGALGGMVRQLS
IDQFENEGRVSVYGTPE NATAARKLLDRQMSINSVPKKVIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPS
VLQLRAPIKIFGDLHGQFGDLMLRFDEYGFAPSTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEYPNNVHLIRGN
HEAADINALFGFRIECIERMERDGIWTRHINRFLFNWPLAALIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVL
MDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQQHLITLFSATNYCGTAN
NAGAILVLRDRLVVVPKLIHPLPPAISSPETS PERHIEDTMQELNANRPPTPTRGRPPVTNDRGSLAWI

>Gm13g38850

MDVDSMVPEPDHDPAAQNHAAAAPADAAEAGEQLAEPPSSGDGSPEQAAASAPQSPVVGPRLAPTYTVVNAILEKKEDG
PGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATALEGNSAASGTPSSAGNAGIRLAGATADVHCYDVL
TNKWSRITPF

GEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLDTQQWPRWHRVSVQGGPGPRYGHVMALVGQRYLMAIGGND
 GKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLCGGRDANSVPLSSAYGLAKHRDGRWEWAIAPG
 VSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSPTGRYSADAAGGDAAVELTRRCR
 HAAAAGDLIFIIYGGGLRGGVLLDDLLVAEDLAAAETTAASNVQAGRLPERYGFADDRTRETMPAAAADGSVVLGNPVAPP
 MNGDIYTDISTENAMLQGPRTARGVEYLVEASAAEAEAI SAALAAAKARQENGEVELPDRDRGAEATPSGKETSSLIK
 DSAGSNNITPGGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRVGYGTPEENATAARKLLDRQMSINSVPKVI
 IAHLLKPRGWKPPVRRQFFLDCNEIADLCEAERIFSSSEPSVLQLRAPIKIFGDLHGQFGDLMRLFDEYGFAPSTAGDI
 AYIDY LFLGDYVDRGQHSLETITLLLALKVEYPPNNVHLIRGNHEAADINALFGFRIECIERMGERDGIW
 TWHRINKLFWNLPLAALIEKKIICMHGGIGRSINHVEQIENIQRPIPMEAGSIVLMDLLWSDPTENDSVEGLRPN
 ARGPLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAHQHLITLFSATNYCGTANNAGAILVLRDLV
 VVVKLIHPLPPAISSPETSPPERHIEDTWMQELNANRPPTPTRGRPPVTNDRGSLAWI

Gossypium raimondii

>Gorai.004G054500.1

MGSKLLYPAPTYRLETYWDSDDEDAPGPRCGHTLTAVAATKTHGPRLILFGGATAIEGGAASSAPGIRLAGVTNSVHSY
 DVLTRKWTRIRPAGEPPSPRAAHAAAAGVTMVVFQGGIGPAGHSTDDLYVLDLTDNDFKWHRRVVVQGGPGPRYGHVMDL
 VAQRYLTVSGNDGKRVLSDAWALDTGQKPYVWQRLNPEGDRPSARMYATASARS DGMFLLCGGRDSSGAPLADAYGLLM
 HRNGQWEWTLAPGVSPSPRYQHAAVFVGARLHVTTGGTLRGGRAIEGEATVAVLDTAAGVWLDNRGLVTSRNTKGV
 HEYDPSLELMRRCRHASASVGVRIYIYGGGLRGDMLLDDFLVAENSPFQSDMSSPILSPDRASSVSSPRFNQPNLSS
 FGTSPSTSDGQEFSSSGSMSMDKNSMEKLEASAAEAEASAVWQAAQASAVPAEETSVDNDNSPAAETASDGS
 DNEGDVRLHPRAVVVAKEAVGNLGGMVRQLSLDQFENESRRMIPSNNDLSYPTKKFARQSPQGLHKKVISTLLR
 PRNWKAPVNRFFLDSYEVGELCYAAEQLFMQEPTVLQLKAPVKVFGDLHGQFGDLMRLFDEYGFSTAGDITYIDY
 LFLGDYVDRGQHSLETITLLLALKIEYENVHLIRGNHEAADINALFGFRIECIERMGENDGIWAWTRFNQLFNYL
 PLAALIEKKIICMHGGIGRSIHSVEQIEKLERPITMDAGSIIIMDLLWSDPTENDSIEGLRPNARGPGLVTFGP
 DRVTFDFCKKNLQLIIRAHECVMDGFERFAHQGLITLFSATNYCGTANNAGAILVVRGLVIVPKLIHPLPPP
 IQSPETSPEVGDWTWMQELNIQRPPPTPTRGRPPQPDHRSSLAYI

>Gorai.007G132700.1

MGSKPWQSPAPTYRLETFWDTDEDAPGPRCGHTLTAVAATKTQGPRLILFGGATAIEGGASSSSPGIRLAGVTNSVHSY
 DVLTKKWTRIRPADPPSPRAHAATVVGTMVVFQGGIGPAGHSTDDLYVLDLTDNYKFKWHRRVVVQGGPGPRYGHVMDL
 VSQRYLITVSGNDGKRVLSDAWALDTGQKPIWQKLNTEGDRPSARMYATANARS DGMFLLCGGRDSSGTPMADAYGLLM
 HRNGHWEWTLAPGAAPSPRYQHAAVFVGVLHVTTGGALRGGRAVEGEAAIAVLDTAAGEWLDNRGLAKGHVEYDAS
 LEMLRRCRHASASVGARIVYHGGYLRGDTLLDDFLVAEYSSQVSDINSVLTPERVLNFSPPRFNNESSFTSFETT
 SDDVDPDFPSAGGMIMDEYSMERLNEASAAEAEANAVWQAAQASAVSSVASVDENSHAAETASDASEIEADVRL
 HVRVVAKEAVGNLGGMVRQLSLDQFENESRRMIPSSDLSYPAKKFYRQKSPQGLHKKVIYTLRPRNWKAPANRR
 FLDSYEVGELCFAAEQIFMQEPSVLQLKAPIKVFGLHGQFGDLMRLFDEYGFSTAGDITYIDY LFLGDYVDRGQHS
 LETITLLLALKIEHPENVHLIRGNHEASDINALFGFRIECIERMGENDGIWAWTRFNQLFNCIPLAALIEKKIICMH
 GGIGRSIHAVEQIEKLERPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVADFCCKNRLQLI
 IRAHECVMDGFERFAHQGLITLFSATNYCGTANNAGAILVVRGLVIVPKLIHPLPPP IHSPEASPERVGD
 DAWMQLNIQRPPPTPTRGRPPQPEVDRNSLAYI

>Gorai.004G011600.1

MDVDSMVPETGQNHSGGSQLPAPMEREQQQNNQQLNENQDSSSGPGGSLSSPGQQQAVPAASDGGQQAQT
 TVAGPRCAPPYSVNAIIEKKEDGPGPRCGHTLTAVAAGVEEGTSGYIGPRLILFGGATALEGNSAASGTPSSAGS
 SAGIRLAGATADVHCYDVLTKWRSRITPFGEPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLDTQPRPR
 WHRVVVQGGPGPRYGHVMAIVGQRFVMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS
 DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSSV
 AVLDTAAGVWCDTKSVVTSPTGRYSADAAGGDASVELTRRCRHAAGDLIFIIYGGGLRGGVLLDDLLVAEDL
 AAAETTTAASHAAAAAATDVHPGRLPGRYGFVDETRRETMPAAADGAVLGNPVAPPVNGDMHTDISTENAMVQ
 GSRMSKGEYLVEASAAEAEAI SATLAAAKARQVNGEVELPDRDRGAEATPSGKQTSTLIKMPDSSGSNDVAP
 PGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRVSYGTPEENATAARKLLDRQMSISSVPKVI
 AHLKPRGWKPPVRRQFFLDCNEIADLCEAERIFAAEPTVLQLKAPIKIFGDLHGQFGDLMRLFDEYGFSTAG
 DIAYIDY LFLGDYVDRGQHSLETISLLLALKIEYPPNNVHLIRGNHEAADINALFGFRIECIERMGERDGI
 WAWHRINRFLFWNLPLAALIEKKIICMHGGIGRSINHVEQIENLQRPITMEAGSIVLMDLLWSDPTENDSVEG
 LRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAHQHLITLFSATNYCGTANNAGAILVLRDL
 VVVVKLIHPLPPAMSSPEASSERHIEDTWMQELNANRPPTPTRGRPPQVTNDRGSLAWI

>Gorai.009G082300.1

MDVDSMVSSETDHDQNHGAMQSPAPMEREQQQPPQQNNQVENQASTPGPGSPSSAQQAAPAAQGPQAPVAG
 PRCAPTYSVNAIIEKKEDGPGPRCGHTLTAVAAGVEEGTSPSYIGPRLILFGGATALEGNSASAGTPSSAGS
 SAGIRLAGATADVHCYDVLTKWRSRITPFGEPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLDTQQRPR
 WHRVVVQGGPGPRYGHVMALVGQRYLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS
 DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSSV
 AVLDTAAGVWCDTKSVVTSPTGRYSADAAGGDASVELTRRCRHAAGDLIFIIYGGGLRGGVLLDDLLVAEDL
 AAAETTTAASHAAAAATATSNAARLPGRYGFVDERARQTMAEAVPDGSVVLGNPVAPPVNGDMYADISTENV
 MLQGSRRMNKGVLYLVEASAAEAEAI SATLAAAKARQVNGEVELPDRDRGAEATPSGKHMSTLIKMPDSTG
 SNSVAPAGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRVSYGTPEENATAARKLLDRQMSINSVPKVI
 AHLKPRGWKPPVRRQFFLDCNEIADLCEAERIFTVEPTVLQLKAPIKIFGDL

HGQFGDLMRLFDEYGAPOSTAGDIAYIDYFLG DYVDRGQHSLETITLLLALKVEYPNNVHLIRGNHEAADINALFGFRIE
 CIERMGERDGIWVWHRINRLFNWPLAALI EKKI ICMHGGIGRSINHVEQIENLQRPITMEAGSIVLMDLLWSDPTENDS
 VEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVV
 VPKLIHPLPPAMSSPEASPERHI EDTW MQELNANRPPTPTRGRPQVANDRGLAWI

>Gorai.006G116600.1

MDVDSSMVPETDHDQNQNGASQSSAPMEREQQQQNQLENQASPPGPVGSAPTSAAQQQSAAPPATQGGPQTPVVGPRWA
 PTYSVVAIIEKKEDGPGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATALEGNSAASGTPSSAGSAGIRLAGATADV
 HCYDVL SNKWSRITPFGEPPTPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHV
 MALVGQRYLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCGGRDANSVPLASAYG
 LAKHRDGRWEWAIAPGVSPSPRYQHA AVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSPTRTGRYS
 DAAGDASVELTRRCRHA AAAVGD LIF IYGG LRGGVLLDDLLVAEDLAAAETTAASHAAAAAASNVNAGR LPGRYGF
 DERERQTMPEAAPDGAVVLGSPVAPPVNGDMYTDISTENAMLQGSRRMSKGEYLV EAAAAEAEI SATLAAAKARQVNG
 EVELPDRDRGAEATPSGKQMS TLIKIPDSSGSNNVTPAGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRVSYGT
 PESATAARKLLDRQMSINSVPKVI AHLLKPRGWKPPVRRQFFLDCNEIADLCESAEKIFA AEPTVLQKAPIKIFGDLH
 GQFGDLMRLFDEYGAPOSTAGDIAYIDYFLG DYVDRGQHSLETITLLLALKVEYPNNVHLIRGNHEAADINALFGFRIE
 CIERMGERDGIWVWHRINRLFNWPLAALI EKKI ICMHGGIGRSINHVEQIENLQRPITMEAGSIVLMDLLWSDPTENDS
 VEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVV
 VPKLIHPLPPAMSSPEASPERHI EDTW MQELNANRPPTPTRGRPQVANDRGLAWI

Klebsormidium flaccidum

>Kflaccidum_GB_J0254070

SVLMEEPVAAEFEGDANMDISTENAVTHHLEQGDV SPLVQNAEEAAAASAILQAKTKPEEEMEMAAEEPNRETSPEVKQ
 TATGRAAEAAAARAVATTTQRNPDRDVTDPVRLFHRAVVVAGGTLGGLVRQLSIDQFENEGRVSVFGT PDS SHGKEVNKML
 DRMSLQGVHKKVLSALLKPRGWKPP TNRKFFLEAPEIAELCEKAEQHFQQEPSVLKLHAPIKLFGLDHGQFGDLMRLF
 EYGAPOSTAGDITYIDYFLG DYVDRGQHSLETICLLLALKIEYPRSVHLIRGNHEAADINALFGFRIE CIERMGERD
 GIPGIYAWQRINQLFNWPLAAMIENKI ICMHGGIGRSIEKVEQIEALQRPLTMDAGTDRERVVLMDDLLWSDPTENDS
 VEGLRPNARGPGLVTFGPDRVDFCKENDLQLIVRAHECVMDGFERFAMGHLITLFSATNYCGTANNAGAILVLRDLVV
 VPKLIHPLPPANNDSDSEDTW MQELNVQRPTPTRGRP

Linum usitatissimum

>Lus10024773

MGSKPWLYPAPAYRSLETYWDTDEDAPGPRCGHTLT SVAASKTHGPRLILFGGATAIEGGASSAPGIRLAGVTNSVHSY
 DVLTRKWTRIQPAGDPPSPRAAHAAA AVGT MVVFQGGIGPAGHSTDDLYVLDLTNDKYKWHRVVVQGGPGPRYGHVMDL
 VGQRYLVTVSGNDGKRVLSDAWTLDTAQKPYAWQRLNPEGDKPSARMYATASARSDGMFLLCGGRDISGVPLGDAYGLL
 HRNGQWEWTLAPGVSPSPRYQHA AVFVGARLHVTTGGVLRGGGRGVEGEAAIAVLDTAAGVWLDNRGLVTS SSKTGRGLADYD
 PSLELMRRCRHASASVGVRIYVYGG LKGDVLLDDFLVAENSALQADGSSPILTSEKASPVTS PRNLHNSGLSFGAE SPP
 FGSVSMDKISMEKLEASAAEAQAASAVWQAAQASVHSTEVTSVSDENSQAAEATSDGSDVEADVRLHPRAVVVAKEAIG
 NLGVVAKLGGMVRQLSLDQFENESRRMVPTTNDTTPNPAKRFTTRQKSPQGLHKKVISTLLRPRNWKAPTNRFFLDSY
 EIGE LCYAAEQIFMHEPTVLQKAPVKVFGDLHGQFGDLMRLFDEYGF PSTAGDITYIDYFLG DYVDRGQHSLETITLLLALK
 IEYPENVHLIRGNHEAADINALFGFRIE CIERMGESDGIWAWTRFNQLFNYLPLAALVEKKI ICMHGGIGRSITSVEQIE
 KLERPITMDAGSII LMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVDFCKKNKLQLIIRAHECVMDGFERFAQQQL
 ITLFSATNYCGTANNAGAILVGRGLVVVPKLIHPLPPPLQSPETS PERIAEETW MQELNIQRPTPTRGRPQPDLDNRNS
 LAYI

>Lus10009783

MGSKPWLYPAPAYRSLETYWDTDDDAPGPRCGHTLT SVAASKTHGPRLILFGGATAIEGGASSAPGIRLAGVTNSVHSY
 DVLTRKWTRIRPAGDPPSPRAAHAAA AVGT MVVFQGGIGPAGHSTDDLYVLDLTNDKYKWHRVVVQGGPGPRYGHAMD
 LVGQRYLVTVSGNDGKRVLSDAWTLDTAQKPYAWQRLNPEGDKPSARMYATASARSDGMFLLCGGRDSSGVPLGDAYGLL
 HRNGQWEWTLAPGVSPSPRYQHA AVFVGARLHVTTGGVLRGGGRGVEGEAAIAVLDTAAGVWLDNRGLVTS SSKTGRGLADYD
 PSLELMRRCRHASASVGVRIYVYGG LKGDVLLDDFLVAENSALQADGSSPILTSEKASPVTS PRNLHNSGLSFGAE SPP
 FGSISMDKISMEKLEASAAEAQAASAVWQAAQASVNSTEVTSVSDDNSQAAEATTDGSDVEADVRLHPRAVVVAKEAIG
 NLGGMVRQLSLDQFENESRRMVPTTNDTTPNPAKRFTTRQKSPQGLHKKVISTLLRPRNWKAPTNRFFLDSY EIGELCYAA
 EQIFMHEPTVLQKAPVKVFGDLHGQFGDLMRLFDEYGF PSTAGDITYIDYFLG DYVDRGQHSLETITLLLALKIEYPE
 NVHLIRGNHEAADINALFGFRIE CIERMGESDGIWAWTRFNQLFNYLPLAALI EKKI ICMHGGIGRSITSVEQIEKLERP
 ITMDAGSII LMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVDFCKKNKLQLIIRAHECVMDGFERFAQQQLITLFS
 ATNYCGTANNAGAILVGRGLVVVPKLIHPLPPPLQSPETS PERITDETW MQELNIQRPTPTRGRPQPDLDNRNSLAYI

>Lus10013357

MRIRWRSPIRPAPTYSVVAIIEKKEDGSGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATGLEGNAGGTPSSAG
 GAGIRLSGATADVHCYDVL TNKWSRLTPIGEPPTPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVV
 VQGGPGPRYGHV MALVGQRYLMAIGGNDGKRPLSDVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCGG
 RDANSVPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHA AVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWC
 DTKSVITSPRTGRYSADAAGDAVELTRRCRHA AAAVGD LIF IYGG LRGGVLLDDLLVAEDLAAAETTTAASLAAAAANLP
 GARMGGRYGFIDDRSQ TMSDTS PDGSSVVLGNPVAPPVNGDMYTDISTENAMLQGSRRTSKGEYLV EASAAEAEI TAT
 LAAAKARQVNGEVELPDRERTAEATPSGKQLLTMIKPDSIGLNHI PPAGVRLHHRVVAEETGGALGGMVRQLSIDQFE

NEGRRVSYGTPESATAARKLLDRQMSINSVPKVKVIAHLLKPRGWKPPVRRQFFLDCNEIADLCDSAERIFSSEPSVLQLK
 APIKIFGDLHGQFGDLMLRFDEYGAPOSTAGDISYIDYFLFGDYVDRGQHSLETIALLLLALKVEYPNNVHLIRGNHEAADI
 NALFGFRIECIERMGERDGIWVWHRINRLFNWLPLAALIEKKIICMHGGIGRSINHVEQIETIQRPITMEAGSIVLMDLL
 WSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAI
 LVLGRDLVVVVKLIHPLPPAISSPEASPERHIEDTWMQELNANRPPTPTRGRPQVANDKGLAWI

>Lus10008369

MDVDSSMVSADADQDTTAQNVSSSAAAREEQQLSEDPQPVASLSPPPSPPAQQSTQDQAQVQQQSTVVGPRPAPTYSVV
 NVIIIEKKEDGPGSRCGHTLTAVAAVGEETPGYIGPRLILFGGATGLEGNSTSTPTSAGGAGIRLSGATADVHCYDVLV
 NKWSRISPIGEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVWVQGGPGPRYGHVMAVVGQR
 FLMAIGGNDGKRPLSDVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSAGLGLLLCGGRDANSVPLASAYGLAKHRDG
 RWEWAIAAGVSPSSRYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDDTKSVITSPRTGRYSADAAGGDA
 AVELTRRCRHAHAAGVLDLIFITYGGLRGGVLLDDLLVAEDLAAEETSAASHAAAAAASSNVPGMRMSGRYGFIDERSRKT
 IPDASPDGSVILGNPVAPPVNGDMYTDISTENTMLPGSRRTTKGVEYLVEASAAEAEITAALAAAKARQENGEVELPDR
 ERTAEATPSGKQLSTMIKPDSTGSHSIPPTGVRLHHRVAVIAAETGGALGGMVRQLSIDQFENEGRRVSYGTPESATAAR
 KLLDRQMSINSVPKVKVIAHLLKPRGWKPLVRRQFFLDCNEIADLCDSAERIFSSEPSVLQLRAPIKIFGDLHGQFGDLMR
 LFDEYGPSTAGDIAYIDYFLFGDYVDRGQHSLETIALLLLALKVEYPNNVHLIRGNHEAADINALFGFRIECIERMGERD
 GIWVWHRINRLFNWLPLAALIEKKIICMHGGIGRSINHVEQIESIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNAR
 GPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILLVLGRDLVVVVKLIHPLP
 PAISSPETSUPERHIEDTWMQELNANRPPTPTRGRPQAANDRGLAWI

>Lus10004100

MDEDSSMASEPDQPPAQNDSSSADREQVRENQPPFGSLAATAAASSPPSPTLQQPVTAEVQVQQQSPVAGPRPAPTYSV
 VNAIIIEKKEDGSGPRLILFGGATGLEGNAGTPSSAGGAGIRLSGATADVHCYDVLVTKWSRLTPIGEPPTPRAAHVAT
 AVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVWVQGGPGPRYGHVMAVVGQRYLMAIGGNDGKRPLSDVWALDT
 AAKPYEWRKLEPEGEGPPPCMYATASARSAGLGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAIAAGVSPSPRYQHAAVF
 VNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDDTKSVITSPRTGRYSADAAGGDAVELTRRCRHAHAAGVLDLIFITY
 GGLRGGVLLDDLLVAEDLAAEETTTAASLAAAAANLPGARMGGRYGFIIDRSRQMTSDASPDGSSVLLGNPVAPPVNGDM
 YTDISTENAMLQGSRRITKSGVEYLVEASAAEAEITATLAAAKARQVNGEVELPDRERTAEATPSGKQLLTMIKPDSIGL
 NHIPPAGVRLHHRVAVIAAETGGALGGMVRQLSIDQFENEGRRVSYGTPESATAARKLLDRQMSINSVPKVKVIAHLLKPR
 GWKPPVRRQFFLDCNEIADLCDSAERIFSSEPSVLQLKAPIKIFGDLHGQFGDLMLRFDEYGAPOSTAGDISYIDYFLFGD
 YVDRGQHSLETIALLLLALKVEYPNNVHLIRGNHEAADINALFGFRIECIERMGERDGIWVWHRINRLFNWLPLAALIEKK
 IICMHGGIGRSINHVEQIETIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLI
 VRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILLVLGRDLVVVVKLIHPLPPAISSPEASPERHIEDTWMQELNA
 NRPPTPTRGRPQVANDKGLAWI

Marchantia polymorpha

>Mp_M408G1.2f (GB_AB288009)

MDVDAGMAGENDQSRPPLSEQPEQTQSPVHSGPRPAPTRFIVSAILDKKEDGPGPRCGHTLTAVAAVGGDGSYSYIGPR
 LILFGGATALEGNSNAAGPQTSSGAGIRLAGATADVHCYDVQSNKWARLTPVGDPPSPRAAHAATAVGTMVVIQGGIGPA
 GLSTDDLHVLDLTQARPRWHRVWVQGGPGPRYGHVMSLVAQRFLLSISGNDGKRPLADVWALDTAAKPYEWRKLEPEGD
 GPPPCMYATASARSAGLGLLLCGGRDASSVPLASAYGLAKHRDGRWEWALAPGVSPSPRYQHAAVFVNARLHVSGGALGGG
 RMVEDASSVAVLDTAAGVWCDDTKSVITSPRTGRYSADAAGGEASVELTRRCRHAHAAGVLDLIFMYGGLRGGVLLDDLLVAE
 DLAAEETTTAALQAAVASSAIGQVSSYHPVTSASAASAKSVPSEGRAKQLSPVSSLDGAVVMGNPVATPANGDTSMDI
 STENALSYGNRGAGKGVESLVEASAAEAEISLALAAAAKAKARESMGEVESVERDRGAEATSSGKPPVTASAVSNITIT
 KSSSVPVNYSPTGVRLHHRVAVVAAESGGALGGLVLRQLSIDQFENEGRRVSYGTPDSASAARKLLDRQMSISGVQKKV
 LGHLLKPRGWKPPVVKRQFFMDCNEVAELCETAERLFAREPSVLQIRAPVKIFGDLHGQFGDLMLRFDEYGPSTAGDITY
 IDYFLFGDYVDRGQHSLETITLALLALKVEYPRNVHLIRGNHEAADINALFGFRIECIERMGEEDGIWVWQRINQLFNWLP
 LAALIEKKIICMHGGIGRSINRVEQIEALQRPITMEAGSVVLMMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFC
 KNNDLQLIVRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILLVLGRDLVVVVKLIHPLPPAISSPEASPERHI
 EDTWMQELNVQRPPPTPTRGRPHAASDRGLAWI

Medicago truncatula

>Medtr1g089370.1

MGSKPWLYPAPTYKPIESFWDTDEDAPGPRCGHTLTAVAAATKSHGPRILILFGGATAIEGGSTAPGIRLAGVTNSVHAYD
 VDSKKWTRIKPAGDPPSPRAAHAATAVGTMVVFQGGIGPAGHSTDDLVLVLDLTNDKYKWHRVVWVQGGPGPRYGHVMDLV
 AQRYLVTVSGNDGKRVSDAWTLDTAQKPYAWQKLNPEGDRPSARMYATASARSAGLGLLLCGGRDSSGTPQADAYGLLMH
 RNGTWETLAPGVSPSPRYQHASVVFVGARLHVTTGGVLRGGRAVEGEPISAVLDTAAGVWLDNRNIVSSRSNKGHDYDHS
 LELMRRCRHAHAAGVQVYIYGGLRGDLLGLDFLVAENSPLQPVINERGSPLTSPKHNQSNFNYNKTPSMDGGVEIPSS
 GSGGLDKVSLKLEKEASAAEAEASAVWQSVQAISSPADETFVSDNNSHAADTVSDGSDTEGDVRLHPRVAVAKEAVG
 NLGGLVLRQLSLDQFENESRRMLP INNDSPTYPTKFKTRQKSPQGLHKKIISNLLRPRNWKAPVNRFFLDSYEVGELCYAA
 EQIFMHEPTVLQKAPVKVFGDLHGQFGDLMLRFDEYGFPOSTAGDITYIDYFLFGDYVDRGQHSLETITLALLALKIEYDP
 NVHLIRGNHEAADINALFGFRIECIERMGENDGIWAWTRFNQLFNHPLAALIEKKIICMHGGIGRSIHSVEQIEKIERP
 ITMDAGSIIIMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTEFCKKNLQLIIRAHECVMDGFERFAQGLITLFS
 ATNYCGTANNAGAILLVGRGLVVVVKLIHPLPPQSPESPVRVMDTWMQELNIQRPPPTPTRGRPQPDLDRLGLAYI

>Medtr2g087530.1

MEVDS SMLSEPGHNHGHHDHQP GAATAEAVVEGGEVTEAPAVVVGGGSPVEEVQHS AVVGPRLAPNYTVVNAI LEKKEDG
 PGP RCGHTLTAVAAV GEEGTPGYIGPRLILFGGATALEGNSAASGTPSSAGNAGIRLAGATADVHCYDVL TNKWSRLTPF
 GEPPTPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQWPRWHRVSVQGGPGPSRYGHVMALV GQRYLMAIGGND
 GKRPLADVWALDTAAKPYEWRKLEPEGEGPPP C MYATASARS DGLLLLCGGRDANSVPLASAYGLAKHRDGRWEWA IAPG
 VSPSPRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWC DTKSVVTTPTRTGRYSADAAGGDA AVELTRRCR
 HAAA AVGDLIF IYGGLRGGVLLD DLLVAEDLAAAETT SAASHAAAAAASNVQAGRLPGRYGFDDRPRQ TMTEVAADGSV
 LGNPVAPP MNGDIYTDI STENAILQGP RR TAKVEYLVEASAAEAEAI SAALAAAKARQENGEVELPDRDRGAEATPSGK
 HTSSLIKPD PVANNIASGGVRLHHRAVVIAAETGGALGGMVRQLSIDQFENEGRRVSYGTPENATAARKLLDRQMSINSV
 PKKVI AHLLKPRGWKPPVRRQFFLDCNEIADL CDSAERIFSSSEPSVLQLSAPIKIFGDLHGQFGDLMRLFDEY GAPTSTAG
 DIAYIDYLF LGDYVDRGQHSLETITLLLALKVEYPTNVHLIRGNHEAADINALFGFRIECIERMGERDGIW TWHRINRLF
 NWLPLAALIEKKIICMHGGIGRSINHIEQIENIQRPIMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRV
 MEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPPA ISSPETS PER
 HIEDTWMQELNANRPPTPTRGRPPVTNDRGSLAWI

>Medtr4g050540.1

MDVDS SSMASENDNDSKVAKIEEGEQ LGGNSIPAPSPSSSGGSPAEEQQSQQQGAAAGQVQHSSVVG PRLAPS YTVVDAILE
 KKEDGPGPRCGHTLTAVPAV GEEGTPGYIGPRLILFGGATALEGNSAASGTPPTSAGNAGIRLAGATADVHCYDVL TNKWS
 RITPFGEPPTPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVGVPGPGPGRYGHVMALV GQRYLMA
 IGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPP C MYATASARS DGLLLLCGGRDASSVPLASAYGLAKHRDGRWEW
 AIAPGVSPSSRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWC DTKSVVTTPTRTGRYSADAAGGDASVEL
 VRRCRHAAA AVGDLIF IYGGLRGGI LDDLLVAEDLAAAETT SAASHAAAAAASDVQMGRVPGRYGFVDDRTGQPMSDA
 AADGAVVLGNPVAPPVNGDMYTDI STENALLQGSRRTSK GVEYLVEASAAEAEAI SATLAAAKARQENGGVES PDRDRGA
 EATPSGKQISSMIKPEPGGPNSSGGVRLHHRAVVIAAETGGALGGMVRQLSIDQFENEGRRVSYGTPESATAARKLLDRQM
 SINSV PKKVI AHLLKPRGWKPPVRRQFFLDCNEIADL CDSAERIFSSSEPTVIQLRAPIKIFGDLHGQFGDLMRLFDEYGS
 PSTAGDIAYIDYLF LGDYVDRGQHSLETISLLLALKVEY PNNIHLIRGNHEAADINALFGFRIECIERMGERDGIW TWHR
 INRLFNW LPLAALIEKKIICMHGGIGRSINHLEQIENIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTF
 GPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPPALSPET
 SPERHIEDTWMQELNANRPPTPTRGRPPVANDRGSLAWI

Micromonas pusilla

>Micromonas_RCC29902g07380

MEGDGMLCAQFHV DTRSDDPGPRCGHTLTCTVPADGGGQRLIIFGGATALEGDGPSTSGIRLAGATS DVHVSFDVRSVGVW
 TKLDPTGEGPSPRAAHSA AAVGNM VVVQGGIGPSGLASEDLHVLDLQGAPR WHRVSVKGP GPGQRYAHVIFVVAQRFLV
 HGGNDGAKPLGDSWCLDTT SKPYEWFKMNPA GDVPPPRMYAAAAPRADGLLLLCGGRGADSSPLSDAFGLARHRDGRWEW
 AAAPGEAPTARYQH AVAFVGT RLHVSGGALGGGSMVDDHLSLAVLNTSAGSAAAGWSVAEIERAGGSSAAAATASRRCRHT
 TAGVGPLVFCGGLRGGNLLGDMYVMEESPA AQGLTRTEATAVLAEHVDLTAPAWRRWRDRAGLLEEA AAFLSANPDHSG
 GSVNSDHGGSVKPGFATPGGTNLGSLKFVGG SAGSLKFDHGPSGADGSPENNIGANNEAMTKLSSEEQA AVALAKAAH
 VDIDSPPE SHGSGKSSVHTPSPQSTARGGARGARGTASTARKPPSSEVRLHHRAVVVAAGSHDGPASGLGSLVRQLSIDQ
 FESESKRIGPDMAAFGTPMVKMGSLDPGSMT PDGRGVLT PDHEKPTVPRKVLNELLHPRQWEPDPV GASRGF ILNVDEI
 DELCTLAEEKFQAE STVLR LRAPTKIFGDLHGQYGDLMRLFAEY GSPSTAGDIAYIDYLF LGDYVDRGAYSLETMSLLLA
 LKIEHPDNIHLLRGNHEEADINALFGFRTECIERLGE EAGVQAWTRFNLFQWLPLAAVIEDRVACMHGGIGRSITHISQ
 LEELERPLTIESG GIMLMDVLWSDPTESDRVEGLRPNARGPGLVTFGPDRVRKFCEDNDLQMIIRAHECVMDGFERFAGG
 LLITVFSATNYCGTANNAGAILVLRDLKLYPKLIHPLPPSALYGDGGEDEDMEGGSWQLALNKDRPPTPRGRAGDQKS
 RIGSGGDI F

Mimulus guttatus

>Mgv1a000720

MDVDSAMAAEPDHPDSEQPTTANGADHPT ESESLLPQQHQQQEQQSNNNIIDNSNSSGVGGPGA AVAGPRCAPMYTVVN
 ALMDKKEDGPGPRCGHTLTAVPAV GEEGSPGYIGPRLILFGGATALEGNSGASGTPSSAGSAGIRLAGATADVHCYDVL T
 NKWSRITP IGEPTPRAAHVATTVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPGRYGHVMALV GQRY
 YLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPP C MYATASARS DGLLLLCGGRDANSVPLASAYGLAKHRD
 RWEWA IAPGVSPSSRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWC DTKSVVTSPTRTGRYSADAAGGDA
 AVELTRRCRHAASAVGDLIF IYGGLRGGVLLD DLLVAEDLAAAETT SAASHAAAAAASDVHQGR LQGRYGF TDDRNGAT
 VPDAADGAVVGNVAPPVNGDIYTDI INTENATFQGSRLSE GVEYLVKASAAEAEAI TAALAAAKARQVNGEVELPDR
 DRGAEATPSGKQISTLIKPDSSNNSIPAGVRLHHRAVVVA AETGGALGGMVRQLSIDQFENEGRRVSYGTPESATAARKL
 LDRQMSISSVPKKVITHLLKPRGWKPPVRRQFFLDCNEIADL CDSAERIFASEPSVLQLKAPIKIFGDLHGQFGDLMRLF
 DEYGSPTAGDIAYIDYLF LGDYVDRGQHSLETITLLLALKVEHPHQVHLIRGNHEAADINALFGFRIECIERMGERDGI
 WTWHRINRLFNW LPLAALIEKKIICMHGGIGRSINHVEQIENIQRP IAMEAGSIVLMDLLWSDPTENDSVEGLRPNARGP
 GLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPPA
 ISSPETS PERHIDDTWMQELNANRPPTPTRGRPQVANDRGSLAWI

>Mgv1a000624

MDVDSATTAEADHPDSEEPVAANGAVRLSESSLPQQQQENSDMNSDSGDGGGGGGGGVAGPRCAPSYSVVNALMD
 KKEDGPGPRCGHTLTAVPAV GEEGNPGYIGPRLILFGGATALEGNSAASGTPPTSAGSAGIRLAGATADVHCYDAL TNKWT

RVTPIGEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVHGPGPGPRYGHVMALVQRYLIA
 IGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCXXXPLASAYGLAKHRDGRWEWAIAPGVSPSSRYQHAAVFNARL
 RLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSRPTGRYSADAAGGDASVELTRRCRHA AAAVGD LFIYGG
 RGGVLLDDLLVAEDLAAAE'TTSAASHAAAAAASNLQGRQLQGRYGF'TDDRTRASIPDPAADGAVVVGSPVAPPVNGDIY
 TDISTENAMLQGSRALSKGVEYLVEASAAEAEAITAALAAAKARRVNGEVELPDRDRGAEATSSGKQISTQIKPDSAGPN
 NSVPAGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRRVSYGTPESATAARKLLDRQMSIHSVAKKVI A HLLKPRG
 WKPPVRRQFFLDCNEIADLDCSAERIFASEPSVLQKAPIKIFGDLHGQFGDLMRLFDEYGPSTAGDIAYIDYFLGDY
 VDRGQHSMETISLLLALKVEHPHQVHLIRGNHEAADINALFGFRIECIERMGERDGIWTHRINRFLFNWPLAALIEKKI
 ICMHGGIGRSINHVEQIENIQRP IAMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIV
 RAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPPAISSPETS PERLI EDTW MQELNAN
 RPPTPTRGRPQVTNDRGSLAWI

>Mgv1a001316

MGSKPWLYPAPAYRNLESFWD'TDDDAPGPRCGHTLTAVAAKTQGPRLILFGGATAIESGSGGAPGIRLAGVTNCIHSYD
 VLTRKWTRLRPAGEPPSPRAHAAA AVGSMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGGPGPRYGHVMDLV
 AQRYLTVSGNDGKRVLSDAWVLDTAQKPYVWQRLNPEGDRPSARMYATASARS DGMFLLCGGRDASGTPLADAYGLLMH
 RNDLWEWTLAPGVSPSPRYQHAAVFGARLHVTGGALRGGRAVEGDAAIAVLDTAAGVWLDNRGLVTSRVNKVPSDDPSL
 ESMRRCRHAASVGVRIYVHGGLKGDILLDDFLVAENSQFQTDVNP SALMSGSPRSNPNLSPFSSTSTPDDRSETPASG
 FLGMDKDSMEKLAESAAEAKVANAVWQSAQAASASPEETSLSDGNSQSAETTSEGS DTEADVRLHPRAVVAKETVGNL
 GGMVRQLSLDQFENESRRMIPSHNDLSNPTKKFARQRSPQGLHKKIISTLLRPNRWKPPVNRKFFLDSYEVGELCYAAEQ
 IFMHES'TVLQKAPIKVFGLHGQFGDLMRLFDEYGF PSTAGDITYIDYFLGDYVDRGQHSLETITLLLALKIEYDPNV
 HLIRGNHEAADINALFGFRIECIERMGESDGIWAWTRFNQLFNYLPLAALIEKKIICMHGGIGRSINSVEQIEKLERPIS
 MDAGTLVMDLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTFCKRNKQLIIRAHECVMDGFERFAQGQLITLFSATN
 YCGTANNAGAILVVGRELIVVPKLIHPLPPPFQSPETS PERVTD ETW MQELNIQRPPTPTRGRPQPELDRSSLAYI

Musa acuminata

>Ma11T25570

MGSKPWLIPAPNYRRLDSSWET'DDDAPGPRCGHSLTAVDASKSRGPRLILFGGATAIEGGASSAVPGIRLAGVTNSVHSY
 DVNTRKWTRIHVPEPPSPRAHAAA AVGTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGGPGPRYGHVMDLV
 VAQRYLVTISGNDGKRVLSDAWALDTAQKPYRWQKLNPEGDRPSARMYATASARS DGMFLLCGGRDSSGMSLSDAYGLLM
 HTNGQWETLAPGVSPSPRYQHSVAVFVGARLHVTGGALKGGRNIEGEGAIIVLDTAAGVWLDNRGIVTSARTQKSPTDRD
 GSLELLRRCRHASASVGTQIYVYGGLRGDMLLGDFLVAENSPLQSEISSMYNSDRVPAVNQKPNQSVNSDSQSVEQELD
 NKSGMASSASLSDVQKSIDLLTEASAAEAEAVSAVWRAAKEASIESLSDKESSESQVNFESNAEDSSDVAYSLEP'DVRL
 HSRVAVVAKETVSGSLGSLIRQLSLDQFENESRRVHP'TNGEQTYSAKFLNRQSPQGLHKKIISFLLRPRCWAYKAPATRRF
 FLDSYEVGELCHAAEQIFRQEP'TVLQKAPIKVFGLHGQFGDLMRLFDEYGF PSTAGDITYIDYFLGDYVDRGQHSLE
 TITLLLALKIEY PENVHLIRGNHEAADINALFGFRLECIERMGENDGIWAWTRFNQLFNYLPLAALIEKKIICMHGGIGR
 SINSVEQIEKLERPITMDVGSII LMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTFE'CKRNKQLIIRAHECVMDG
 FERFAQGQLITLFSATNYCGTANNAGAILVIGRGLVVPKLIHPLPPPLQSPESP EHAVEDTWMQELNIQRPPTPTRGR
 PQSANDRSSLAYI

>Ma9T25950

MDVDAAMSTNSDHDVLVLANHHDGSGEPSSPPTPAAGQGPVLAGPRPAPTYRVVNAIEKKEDGAGRCRCGHTLTAVPPVG
 EEGTPGYIGPRLILFGGATALEGNSAAPPSPAGTAGIRLAGATADVHCYDMLSNKWTRLTPLGEPSPRAAHVATAVGTM
 VVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPGPRYGHVMALVQRFLLTIGGNDGKRPLADVWALDTAAKPY
 EWRKLEPEGEGPPPCMYATASARS DGLLLCGGRDANSVPLSSAYGLAKHRDGRWEWAIAPGVSPSPRYQHAAVFNARL
 HVSGGALGGGRMVDDSSSIAVLDTAAGVWCDTKSVVTSRPTGRYSADAAGGDASVELTRRCRHA AAAVGD LFIYGG
 RGGVLLDDLLAAEDLAAAE'TTSAASHAAAAAATDVQVGRAPGRYLYSDERSRQDIPETAPDGAIMVGT'PVAPPVNGDIYTD
 ISPENAMLLGPRELNKGV EYLVEASAAEAEAISATLAAVKARQVNGEVEQLLDRDRGSEATPSGKMISSMVKVPDPSAAN
 NTPPTGVR LHHRVVAEETGGALGGMVRQLSIDQFENEGRRVSHGTPENATAARKLLDRQMSINSVPKKVIAHLLKPRG
 WKPPVRRQFFLDCNEIADLDCSAERIFTSEPTVLQIKAPVKIFGDLHGQFGDLMRLFDEYGPSTAGDIAYIDYFLGDY
 VDRGQHSLETITLLLALKVEYPHNVHLIRGNHEAADINALFGFRIECIERMGERDGIWAWHRFNRLFNWPLAALIEKKI
 ICMHGGIGRSINHVEQIEDLQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIV
 RAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPPAINSP EASPERHTEDTWMQELNAN
 RPPTPTRGRPQTSNDRGSLAWI

>Ma2T12930

MEVDSTMSTESGHDPAAGNHTGNPNEPSSSPSAPGAVPPAEAQVPAAGPRPAPGYSVVNAVIERKEDGPGRCRCGHTLTAV
 AAVGEEGTPGYIGPRLILFGGATALEGNSAAPASPTGSAGIRLAGATADVHCYDVL SNKWTRV'TPLGEPSPRAAHVATA
 VGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGGARYGHVMALVQRFLLTIGGNDGKRPLADVWALDTA
 AKPYEWRKLEPEGEGPPPCMYATASARS DGLLLCGGRDANGVPLSSAYGLAKHRDGRWEWAIAPGVSPSPRYQHAAVFN
 NARLHVSGGALGGGRMVEDSSSIAVLDTAAGVWCDTKSVVTSRPRGRYSVDAAGGDASVELTRRCRHA AAAVGD LFIY
 GLRGGVLLDDLLVAEDLAAAE'TTNSASHAAAVAAASNVQVGRSAGRYMFS'DERSRQSSPEAVPDGAVALGTPVAPPVNGD
 MFADISTENALFQGSRRLSKGVEYLVEASAAEAEISAAALAAAKARQVNGEVEQLPDQAHGSEATLGGKQVCNLSKASDS
 SLPNNGTPTGVR LHHRVVAEETGGALGGMVRQLSIDQFENEGRRVSYGTPESATAARKLLDRQMSINSVPKKVIAHLL
 KPRGWKPPVRRQFFLDCNEIADLDCSAERIFTSEPSVLQIKAPVKIFGDLHGQFGDLMRLFDEYGPSTAGDIAYIDYLF

LGDYVDRGQHSLETIALLLALKVEHPHNVHLIRGNHEAADINALFGFRTECIERMGERDGIWTHRINRLFNWLPLAALI
 EKKIICMHGGIGRSINHVEQIENLQRPITMETGSVVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDL
 QLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLGRDLVVVVKLIHPIPPAVSSPEASPEHHIEDTWMQE
 LNANRPPTPTRGRPQVASDRGSLAWI

Nicotiana benthamiana

>Nb00007375

MGSKPWLHPAPTYRILETFWDTDDAPGPRCGHTLTAVAATKTHGPRILIFGGATAIEGGSGGAPGIRLAGVTNCVHSYD
 VLTRKWTRVRPAGEPPSPRAHAAAAGVTMVVFQGGIGPAGHSTDDLYVLDMTNDKFKWHRVVVQGGPGPRYGHVMDLV
 AQRYLVTVSGNDGKRVLSDAWALDTAQKPYVWQRLNPEGDKPSARMYATASARTDGMFLLCGGRDASGTPLTDAYGLLMH
 RNGQWEWTLAPGVAPSPRYQHAAVFVGARLHVTTGGVLRGGRAVEGEAAVAVLDTAAGVWLDNRGLVTASRGSKAHTEQDP
 SLELMRRCRHAAAASVGVRIYIYGGLRGTLLEDECLIAENSPLHSDINSPVLTSEVPTIGSPTSNNNLLTPDGKPEHPH
 GGLSMDKDSMEKLTDFAAEAEANAVWQLAQAASTTPEETSVSDESSQVPTDGSSESDSESAVRLHTRAVVAKETVGN
 LGGMVRQLSLDQFENESRRMLPSYNDISNPTKKFTRQKSPQGLHKKIISTLLRPRNWKAPVNRKFFLDSYEVGELCYAAE
 QIFMHEPTVLQKAPVKVFGDLHGQFGDLMLRFLDEYGFSTAGDITYIDYFLGDYVDRGQHSLETITLLLALKIEYPEN
 VHLIRGNHEAADINALFGFRIECIERMGESDGIWAWTRFNQLFNPLAALVEKKIICMHGGIGRSINSVEQIEKIERPI
 TMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTFCKKNKLELIRAHECVMDGFERFAQGQLITLFS
 ATNYCGTANNAGAILVGRGLVIVPKLIHPLPPPLQSPETSPERVMEDEDTWMQELNIQRPPPTPTRGRPQPDHRS
 SLAYI

>Nb000027297

MNVDSMRSESSEDHDQNNGASSEQLNGQSSAGGCPPETPQNHQHPATSQQQQDSTPVAGPRCAPAYSVVHVAVLQKKEDGPG
 PRCGHTLTAVPAVGEESANYIGPRLILFGGATALEGNSAASGTPSSAGSAGIRLAGATADVHCYDVLTKNWSRMTPIGE
 PPTPRAAHVSTAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMAVVGQRYLMAIGGNDGK
 RPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDDLGLLCCGGRDANSVPLASAYGLAKHRDGRWEWA
 IAPGVS
 PSSRYQHAAVFVNARLHVSGGALGGGRMVEDSSSIAVLDTAAGVWCCTKSVVTSPTGRYSADAAGGDAAVELTRRCRHA
 AAAGDLIFIYGGLRGGVLLDLDLVAEDLATAETTSAAASHAAATDAGRIAGRYGFGDERARQTDPEVVNDGSVVLGTPVA
 PAVNGDMYSIDISTENATLQGSRRLIKGVYELVEASAAEAEITATLAAAKARQQNGEVELPDRDRGAEATPSGKQAPTL
 IKPDSILSNNVSPAGVRLHHRVVAEETGGALGGMVRQLSIDQFENEGRRVGYGTENATAARKLLDRQMSINSVPKKV
 IAHLKPRGWKPPVRRQFFLDCNEMADLCSAERIFASEPSVLQRLAPIKIFGDLHGQFGDLMLRFEYEGSPSTAGDISY
 IDYFLGDYVDRGQHSLETITLLLALKVEYPLNVHLIRGNHEADINALFGFRIECIERMGERDGIWAWHRFNRLFNWLP
 LAALIEKKIICMHGGIGRSINHVEQIENIQRPITMDAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVIEFC
 NNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLGRDLVVVVKLIHPLPPATSPETSPERHIEDT
 WMQELNANRPPTPTRGRPQVANDRGALAWI

>Nb00004040

MDVSSMPSEPDQPNNGGPTLEQPKDTSAPGESPVTTAPVAGPRCAPTYTVVQAVMEKKEDGPGPRCGHTLTAVPAVGE
 EGSSNYIGPRLILFGGATALEGNSAASAPAPSSPGSAGIRLAGATADVHCYDVLTKNWSRLTPIGEPTPRAAHVATAVG
 TMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMAVVGQRYLMAIGGNDGKRPLADVWALDTAA
 KPYEWRKLEPEGEGPPPCMYATASARSDDLGLLCCGGRDANSVPLASAYGLAKHRDGRWEWA
 IAPGVSPSPRYQHAAVFVN
 ARLHVSGGALGGGRMVEDSSSIAVLDTAAGVWCCTKSVVTSPTGRYSADAAGGDAAVELTRRCRHA
 AAAGDLVFIYGG
 LRGGVLLDLDLVAEDLAAAETTSAASHAAAAAANMQEGRITGRYGYGDERTRETVEAVNDGSVVLGNPVAPPVNGDM
 FTDISTENAVLQGSRKLGKGVYELVEASAAEAEITAALAAAKARQQNGEVELPDRDRGAEATPSGEQASTLIKPD
 SNN
 STPTGVRHHRVVAEETGGALGGMVRQLSIDQFENEGRRVSYGTENATAARKLLDRQMSVNSVPKKVIAHLKPRGW
 KPPVRRQFFLDCNEIADLCSAERIFSSEPSVLQRLAPIKIFGDLHGQFGDLMLRFLDEYEGSPSTAGDIAYIDYFLGDYV
 DRGQHSLETITLLLALKVEYPHNVHLIRGNHEAADINALFGFRIECIERMGERDGIWAWHRINRLFNWLPLAALIEKKI
 I
 CMHGGIGRSINHVEQIENIQRPITMEAGSIIIMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVR
 AHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLGRDLVVVVKLIHPLPPASPENSPERMEDTWMQELNANRPPT
 PTRGRPQTANDRGSLAWI

Nuphar advena

>Nad3_c40233-55957

MGSKTGLSPLSPPPPAPAPTYRWVETSWEREEDAPGPRXXXAHTLTAVAAGVEGTSDYVGPRLVLFGGATAIEGPLNA
 GIRLAGVTNLVHIYDVLTRKWSRIHPAGDPPSPRAHAAAAGVTMVVIQGGIGPAGHSTDDLYVLDLTHSKLKWHRVVVR
 GPGPGPRYGHAMDLVGQRYLVTVSGNDGMRVLSDAWALDTSQKPYEWRKLSPEGDHPSARMYATASARSDDLGLLCCGGRD
 SSGVPQADAYGLLKHRSRWETLAPGVSPSPRYQHAAVFVGNRLHVTTGGTLKGGRTVEGEATIAVLDTAAGVWLDNRNGI
 VASSRTNKSSLENPSELELLCRCHRHAAAAGVNRVYIYGGLRGDVLLDLDLFAENSTVQTVAMIPINASERAI SPTKIDQS
 IMTGSSYIQTMSDGLIETATSGASSMDRNSLEKLVESAAEAQAVSAVWEAAKAAASINQDNMLDGLQSSSDSVKDKSQAS
 EISSDSDGDILETDVRLHPRAVVVAKEPVGNLGLVLRQLSLDQFENESRRMNPWNMDQGYAAKRIFNKQKSPVGLHKKI
 ISLLLRPRNWSAPMVRKFFLDSYEVGELCYAAEQIFLQEPVTLQKAPVKIFGDLHGQFGDLMLRFLDEYGFSTAGDITY
 IDYFLGDYVDRGQHSLETITLLLALKIEYPEHVHLIRGNHEAADINALFGFRLECIERMGENDGIWAWTRFNQLFNALP
 LAALIEKKIICMHGGIGRSINYVDQIEKLERPITMDAGSIVLMDLLWSDPTENDTVEGLRPNARGPGLVTFGPDRVIDFC
 KKNKLQLIRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGALLVIGRDLVVVVKLIHPLPPFPQSPECSPDHTMEL
 TWMQELNIQRPPPTPTRGRPQPAHDRSSLAFI

>Nad3_c3534

MDADAAMVTEGDAQDQSQSTSDSMSQDQHIHHHGEHNEQLPQQQEQQQQQPPHSAPPVVGPRPAPAYSVVNA
 IIEKKEDGPGPCGHTLTAVAAVGEETPGYIGPRLILFGGATALEGNSAVSGPPSSGASTGIRLAGATADIHICYDVL
 SNKWSRLLPLGEPSPRAAHVATAVGTMVVIQGGIGPTGLSAEDLHVLDLTQQRPRWHRVVVQGGPGGPRYGHV
 MALVQGRFLLAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS
 DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHA
 AVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTS
 SPRPGRYSAADAAGGDASVELTRRCRHA
 AAAVGDLIIFYGGLRGGVLLDDLLVAEDLAAAETTSAA
 SHAAAAAASANVQVGKFPGRYVFNDDRSRQAA
 PELAPDGA
 VVFGSPVAPPVNGDMYTDISTENAMLQGGRRLSK
 GVEYLVEASAAEAEIASAALAAAKARQSN
 GEMEQLPDGDHGAEDTPSGKPLPTPKVSESA
 VPNNVPTTVRLHHRAVVAAETGGALGGMVR
 QLSIDQFENEGRVVSYGTPESANAA
 RKLLDRQMSISSVPKKVIAHLLKPRG
 WKPPVRRQFFLDCNEIADLCE
 SAERIFASEPSVLQIKAPL
 KIFGDLHGQFGDMLR
 LDFEYGS
 PSTAGDIAYIDYLFGLDYVDRGQHSLETIS
 LLLLALKVEYPHNVHLIRGNHEA
 ADINALFGFRIECIERMGER
 DGIWAWHRINR
 LFNWPLAALIEKKICMHGGIGRSINHVEQ
 IENLQRPITMETGSV
 VLMDDLWSDPTENDSVEGLR
 PNRGPGLVTFGPDRVMEFCNNNDLQ
 LIVRAHECVMDGFERFAQGH
 LITLFSATNYCGTANNAGAILV
 LGRDLVVVVKLIHPLPPAVSSPETS
 PPHIEDTWMQELNAHRPPTPTRGR
 PQVANDRGPLAWI

>Nad3_c8445-28637

MDVDGTMVTEGDPQDQSQSTSGSMSQDQHLHHHGESNEQLPQQQQQQQPEQQPEQQEQQQPHSAPPVVGPRHAPTYSVA
 NAIIEKKEDGPGPCGHTLTAVAAVGEETPGYIGPRLILFGGATALEGNSAVSGPPTXXXGASTGIRLAGATADVHCYD
 VLANKWSRLLPLGEPSPRAAHVATAVGTMVVIQGGIGPTGLSAEDLHVLDLTQQRPRWHRVVVQGGPGGPRYGHV
 MALVQGRFLLAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS
 DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHA
 AVFVNARLHVSGGALGGGRMVEDSSSIAVLDTAAGVWCDTKSVVTS
 PRPTGRYSADAAGGDASVELTRRCRHA
 AAAVGDLIIFYGGLRGGVLLDDLLVAEDLAAAETTSAA
 LAAAAAANVQVGKLPGRYAFNDDRS
 RQAAPESAPDGA
 VVVGSPVAPPVNGDMYTDISTENAMLQGGRRLSK
 GVEYLVEASAAEAEIASAALAAAKARQSN
 GEMEQLPDQDHGAETPSGKLLSTPKVSESA
 VPNNVPTTVRLHHRAVVAAETGGALGGMVR
 QLSIDQFENEGRVVSYGTPESATAARKLLDRQMS
 ISSVPKKVIAHLLKPRGWKPPVRRQFFLDCNEIADLCE
 SAERIFASEPSVLQIKAPVKIFGDLHGQF
 GDLMLR
 LDFEYGS
 PSTAGDIAYIDYLFGLDYVDRGQHSLETIS
 LLLLALKVEYSHNVHLIRGNHEA
 ADINALFGFRIECIERMGERDGIWAWHRINR
 LFNWPLAALIEKKICMHGGIGRSINHVEQ
 IENLQRPITMETGSV
 VLMDDLWSDPTENDSVEGLRPNARGP
 GLVTFGPDRVMEFCNNNDLQ
 LIVRAHECVMDGFERFAQGH
 LITLFSATNYCGTANNAGAILV
 LGRDLVVVVKLIHPLPPAVSSPETS
 PEPHIEDTWMQELNAHRPPTPTRGR
 PQVANDRGPLAWI

Oryza sativa

>Os03g44500

MDVDSRMTTESDSDSAAAATAAASASVAAQGLASETSSSSASAPSTPGTPTVAPAPAAAGATGPRPAPGYTAVSAVIE
 KKEDGPGCRCGHTLTAVPAVGEETPGYIGPRLILFGGATALEGNSATPPSSAGSAGIRLAGATADVHCYDVL
 SNKWSRLLTPQGEPPSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGGPRYGHV
 MALVQGRFLLTIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS
 DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHA
 AVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTPRIGRYSADAAGGDA
 AVELTRRCRHA
 AAAVGDQIFIFYGGLRGGVLLDDLLVAEDLAAAETTTAANHAAA
 SAAATNVQSGRTPGRYAYNDERARQTAPESAQDGSV
 VLGTVPVAPPVNGDMYTDISPENAVLQGGRRLSK
 GVDYLVEASAAEAEIASATLAAVKARQVNGEMEQL
 PDKEQSPDSASTSKHSSLIKPD
 SILSNMTPPPGVRLLHHRAVVAAETGGALGGMVR
 QLSIDQFENEGRVVSYGTPENATAARKLLDRQMS
 INSVPKKVIASLLKPRGWKPPVRRQFFLDCNEIADL
 CDSAERIFSSEPSVLQKAPVKIFGDLHGQF
 GDLMLR
 LDFEYGAPSTAGDIAYIDYLFGLDYVDRGQHSLET
 MTLALLALKVEYPQNVHLIRGNHEA
 ADINALFGFRIECIERMGERDGIWAWHRINR
 LFNWPLAALIEKKICMHGGIGRSINHVEQ
 IENLQRPITMEAGSV
 VLMDDLWSDPTENDSVEGLRPNARGP
 GLVTFGPDRVMEFCNNNDLQ
 LIVRAHECVMDGFERFAQGH
 LITLFSATNYCGTANNAGAILV
 LGRDLVVVVKLIHPLPPAITS
 PETSPEHHIEDTWMQELNANR
 PPTPTRGRPQVAANDRGSLAWI

>Os12g42310

MDVDSRMTTESDSDSAAAQGGGGGFGSETSSASPSAPGTPTAMGAGGAAPIAAAAIAAAA
 SAVVAGPRPAPGYTVVNAAMEKKEDGPGCRCGHTLTAVPAVGEETPGYIGPRLILFGGATALEGNSATPPSSAGSAGIRLAGATADVHCYD
 VSSN
 KWSRLLTPVGEPPSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGGPRYGHV
 MALVQGRFLLTIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS
 DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAIAPGVSPSPRYQHA
 AVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTPRTGRYSADAAGGDAS
 VELTRRCRHA
 AAAVGDMIYVYGGLRGGVLLDDLLVAEDLAAAETTTAANQAAAIAAASDIQAGRE
 PGRYAYNDEQTGQPA
 TITSPDGA
 VVLTGPVAPPVNGDMYTDISPENAVIQQRRMSKGV
 DYLVESAAEAEIASATLAAVKARQVNGEAEHSPDR
 EQSPDATPSVKQNASLIKPDYALSNNSTPPGVRLLHHRAVVAAETGGALGGMVR
 QLSIDQFENEGRVVIYGTPE
 SATAARKLLDRQMSIN
 SVPKKVIASLLKPRGWKPPVRRQFFLDCNEIADL
 CDSAERIFSSEPSVLQKAPIKIFGDLHGQF
 GDLMLR
 LDFEYGAPSTAGDIAYIDYLFGLDYVDRGQHSLETIT
 LLLLALKVEYPLNVHLIRGNHEA
 ADINALFGFRIECIERMGERDGIWAWHRINR
 LFNWPLAALIEKKICMHGGIGRSINHVEQ
 IENLQRPITMEAGSV
 VLMDDLWSDPTENDSVEGLRPNARGP
 GLVTFGPDRVMEFCNNNDLQ
 LIVRAHECVMDGFERFAQGH
 LITLFSATNYCGTANNAGAILV
 LGRDLVVVVKLIHPLPPAITS
 PETSPEHHLEDTWMQELNANR
 PPTPTRGRPQAANNDRGSLAWI

>Os05g05240

MGTAGKAWVVPAPAYREVEGWEGAGDSDSPGFRGCHSLTVVAPT
 KGHGPRILILFGGATAIEAGASSGMPGIRLAGVTNSV
 HSVDVTRRWR
 LHPAGEPPSPRAAHAAA
 AVGTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQ
 GAGPGPRYGHCMDLVAQR
 YLVTVSGNDGKRVLS
 DAWALDTAQKPYRWQKLN
 PDGDRPSARMYATASARTD
 GMLLLCGGRDASGMP
 LSDAYGLLMHTSGQEW
 ETLAPGVSPSPRYQHA
 AVFVGARLHVTTGGVLRGGRAIEGEG
 AIAVLDTAAGVWLD
 RNRGIVTSR
 TLKSSHD

HDASSDLLRRCRHAAASVGTQIYIYGGLRGDILLDDFLVADNAPIQSEFTSSMYDRVPRAENQNRNHNFNNSDSPTTNNRH
 EIASGFSTDKSIDMLTQASAAEAEAVSAVWRAAQEASHASSEDLSLSEIGISESPLSETSPMPEDLDDGGSLEPDVKLHS
 RAVVVSKAEVGDGLGCLVLRQLSLDQFENESRRMHPSSNDQSYPAKKALNRQRSPQGLHKKVIFSLLKPRNWRAPAERAFFL
 DSYEVGELCYAAEQIFMQEPTVLQKAPIKVFGDLHGQFGDLMRLFDEYGYSTAGDITYIDYFLFLGDYVDRGQHSLETI
 TLLLALKIEYPENVHLIRGNHEAADINALFGFRLECIERMGESDGIWAWTRFNQLFNYLPLAAMIEKKIICMHGGIGRSI
 NTIEQIEKLERPITMDVGSIIILMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTEFCKRNRLQLIIRAHECVMDGFE
 RFAHQQLITLFSATNYCGTANNAGAILLVGRGLVIVPKLIHPLPPPVSPESSPERAMDATWMQELNIQRPPPTTRGRPQ
 SASDRNSLAWI

Ostreococcus lucimarinus

>O110g01240

MLAAPFDATR SRAGFADATNNASTFANDSKTFATSSAPGARC GHTMTALRWNQTKIVTFGGATELEGGSGANASANVGL
 SPQAGR DAGAWVKLSGATNELHVLDPFNGEWGKLECGGDVPS PRAAHGAATVGGMLVVLGGIGPDGLADEDLVLDLATR
 DPKWHRVHVRGQGPGRYAHVLSFVAQRFLVIGGTDGSKCLDDTWTWLDTTTTKPYEWTKCAPSGPVP SARTYASASTRSD
 GLLLLCGGRGADGFALNDAYGLARHRDGRWEAEAPGKAPTRRYQHATAFVDTRLRHITGGASGGQLVTNEATMSMLDTS
 SGGSTGWRDAKSSEGKAQLSYDASSLVGRRCRHASVAYGPFIFVHGGLRNGTLLDDLVVLEEPPQESGSTRATERLAT
 LIDPNSLAWRKWLGETGLSADILGMHSPRNVRDSFAINTQRDATYSNGSFSSPSSPEGIALRPLGGSPGSDSPDNMNA
 AEKELRAASAQEAALAVARRKFSLGESDGSPPGSSVHTPSPGFARFGSPESALRTPASEVRLHHRVVAAAPPDSG
 AKSTPRGVASMRQLSIDQFENEARRIGTPGVDMLTPGDTPAKMARARRAAELGAQPVHKA VLTHLLHPHTWEPSTDRRF
 FLSASAINELCDAAEHCFKNEETVLTVKGPAKIFGDLHGQFGDLMRLFAYEGAPSTAGDIAYIDYVFLGDYVDRGAYSLE
 TISLLLALKIEHPQAVHLLRGNHEESDINALFGFRIE CVERLGE SAGDAVWRRFNELFEWLPLAAVIEDRICMHGGIGR
 SLTHISQINELKRPLNMENGGVELMDILWSDPTENDGIEGLRPNARGPGLVTFGPDRVKAF C ETNGIQMIIRAHECVMDG
 FERFAQQQLITVFSATNYCGTANNAGAILVLRDLTLYPKLIHPLPIAMESLSPSDRIDNWLQDVNRDRPPTPRGR
 LGQTQPIGLINPM

Physcomitrella patens

>Pp00117g00820

MEVDAAGMVENRGTGQEHQHYPHQLPGPRLAPAYKAITAICERKDDGPGPRCGHTLTAVAAVGEEGKLAYVGPRLILFG
 GATALEGNSSQAGPQMAVAGAGIRLAGATADVHCYDIHSNKWSRLIPVGDPPSPRAAHAATTVGTMVVIQGGIGPAGLST
 EELHVLDLTQAKPRWHRVVVQGTGPGPRYGHVMSLVGQRFLLCIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGDGPPP
 CMYATACARSDGLLLCGGRDANSAPIDSAYGLAKHRDGRWEWALAPGISPSPRYQHA AVFVNARLHVSGGALGGGRMVE
 DASSVAVLDTAAGMWC DRKGVVNTPRIGRYSNDAAGGEASNELTRRCRHA AAAVGDHIFMYGGLRGGVLLDDLLVAGDPS
 AVETFSVASHLPASQTNKQDVRKPSQTELSSTDGASVAIGNPVAAPFNGHTSMDTSTENSVSHGHRTERRGVESLVEAS
 AAEEAISAALAASAAKALREQNEEIEKQTPERDRGVESTSTAKPPQASSAALMALNRRASMAATVTHTPPTGVRLLHRA
 VVVAE SGGALGGLVRLS IDQFENEGRVSYGTPESASHARLLDRQTSVHGVQKKVLSYLLKSRGWKPPVKRQFFMDC
 NEIAELCNTAERLFAQEPSVLQIRAPVKIFGDLHGQFGDLMRLFDEYGYSTAGDITYIDYFLFLGDYVDRGQHSLETITL
 LLALKIEYPTNVHLIRGNHEASDINALFGFRIE CVERMGERDGIWAWQRINQLFNWLPLAALIEKKIICMHGGIGRSINH
 VHQIEALQRPITMDAGSVVLMDDLWSDPTENDSVEGLRPNARGPGLVTYGPDRVMEFCKNNDLQLIVRAHECVMDGFERF
 AQGHILITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPPAITSPESSPERHLEDTWMQDLNVARPPTTRGRPQPT
 SDRNSLAWI

>Pp00072g00330

MEVDVGGMVENGTTGQEQHHQHYPHQLAGPRPAPAFKAITAICERKDDGPGPRCGHTLTAVAAVGEEGSPAYIGPRLI
 LFGGATALEGNSSQAGPQTAVAGAGIRLAGATADVHCYDAHTNKWSRLIPVGDPPSPRAAHAATAVGTMVVIQGGIGPAG
 LSTEDLHVLDLTQAKPRWHRVVVQGTGPGPRYGHVMSLVGQRFLLCIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGDG
 PPPCMYATACARSDGLLLCGGRDANSAPIDSAYGLAKHRDGRWEWAVAPGISPSPRYQHA AVFVNARLHVSGGALGGGR
 MVEDASSVAVLDTAAGVWC DRKGVVNSPRIGRYSNDAAGGEAANELTRRCRHA AAAVGDHIFIYGGLRGGDILDDLLVAG
 DPSAVETTPAASQIAAVSATSTVSPARSNKQEQEMRKPSQTELSADRASVAMGSPVAAPVNGDTSMDISTENAVSHGHR
 ERRGVESLVEASAAEAEISAALAASAAKAMREQNGESDKQALERDHGAETPLSGKPLPTSSAALMALNRRASMTNAATH
 TPPTGVRLLHRAVVVAE SGGALGGLVRLS IDQFENEGRVSCGTPESASHARLLDRQTSVHGVQKKVLSYLLKPRGW
 KPPVKRQFFMDCNEIAELCNTAERLFAQEPSVLQIRAPVKIFGDLHGQFGDLMRLFDEYGYSTAGDITYIDYFLFLGDYV
 DRGQHSLETIALLLLALKIEYPTNVHLIRGNHEASDINALFGFRIE CIERMGERDGLWAWQRINQLFNWLPLAALIEKKIIC
 CMHGGIGRSINH VHQIEALQRPITMDAGSVVLMDDLWSDPTENDSVEGLRPNARGPGLVTYGPDRVMEFCKNNDLQLIVR
 AHECVMDGFERFAQQGHILITLFSATNYCGTANNAGAILVLRDLVVVPKLIHPLPPAITSPESSPERHLEDTWMQDLNVAR
 PPTTRGRPQPTSDRNSLAWI

>Pp00293g00060

MEVDASGGMQEVEVERHGEQEVQTI EQDGEKENKLEQVKKGEDEEVDKQTEALGTANVVSEPRSGDTWGGEAVTSP
 AVASVEQASQQAAPSAPVQNSHPQHQLSGPRPAPTYRVVNSVIEKRDDGPGPRCGHTLTAVASVNEEGSPGYIGPRLI
 LFGGATALEGSSAAGPQAASSGAGIRLAGATADVHCYDVQSNQWTRITPVGDPPSPRAAHAATAVGTMVVIQGGIGPAGL
 STDDLHVLDLTQAKPRWHRVVVQGPGRYGHVMSLVGQRFLLSISGNDGKRPLADVWALDTAAKPYEWRKLEPEGDG
 PPCMYATACARSDGLLLSGGRDVHSMPLDSAYGLAKHRDGRWEWAVAPGISPSPRYQHA AVFVNARLHVSGGALGGGR
 VEDSSVAVLDTAAGVWC DRRGVVTSPRTGRYSNDAAGDAINELTRRCRHA AAAVGDVLFMYGGLRGGVLLDDLLVAED
 LAAVETTPAALHAAAASHHSFSPHKPQEVKVPHPDLSGADGASVLMANPVAAPVNGDTSMDISTANALSHGHRNERRGV
 ESLVEASTAEAEISAALAAAEAKFRDQNGEIEPHSSDKDREAEATPSGKPPSSPSRSLMSLNRPSIPATPVHTPPAGV

RLHHRVAVVAAESGGALGGLVLRQLSIDQFENEGRVRSYGTDPNVSYARQLLDRQMSISGVQKKVLSYLLKPRGWKPPTKR
 HFFLDCEIADLCNTAERLFAQEPSVLQIRAPVKIFGDLHGQFGDLMRLFDEYGSPSTAGDITYIDYLFGLDGYVDRGQHS
 LETITLLLALKIEYPTNVHLIRGNHEAADINALFGFRIECIERMGERDGIWAWQRINQLFNWLPLAALIEKKIICMHGGI
 GRSINYVRQIEDLQRPITMEAGSVVLMDDLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCKNNDLQLIVRAHECVM
 DGFERFAQGHILITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPAITSPEHSPEHLLDDIWMQELNEHRPPTPTR
 GRPQPAHDRGSLAWI

>Pp00043g00250

MEAGAGMGRDAAVSKSEANGKSSSSSANGSANGEAVGGENIEEMPSSSSPSAAQAVAGRQOVAMATLPRPAPAFRMSVAVI
 EKKEDGPGPRCGHTLTAVAAVGEETPGYVGPRLILFGGATALEGNSSAAGGPQTAAAGAGIRLAGATADVHCYDVTSRK
 WSRLIPVGDPPSPRAAHAATAVGTMVVIQGGIGPAGLSTDDLHVLDLTQAKPRWHRVVVQGAGPGPRYGHVMSLVGQRF
 LSISGNDGKRPLADVWALDTAAKPYEWRKLDPEGEGPPPCMYATACARSDGLLLCGGRDANSVPLDTAYGLAKHRDGRW
 EWAVAPGIAPSARYQHAAVFVNARLHVSGGALGGGRMVEDESSVAVLDTAAGVWCDKKAVVNSPRSGRFSQDVAGGDAI
 ELTRRCRHAASVGDHIYLYGGLRGGVLLDDLVAEDLAAEETSAALHQALHSSAPSSQPNSISSAKAVGDGRQKTAD
 GSVMMGSPVAAPAAGDATMDISTENALAHGHRSEKRGVESLVEASAAEAEASNALKAAEAKAAACRDTNEDVEDGAPDR
 DRGAEATPSGESPMTPAAGSMGLNKLADGALTPPTGVRLHHRVAVVAAESGGALGGLVLRQLSIDQFENEGRVRSYGT
 TPNSAMLARRLLDRQMSMASVQKKVLAYLLKPRGWKPPVKRQFFMDCNEIAELCDAERLFSQESSVLQIRAPVKIFGDL
 HGQFGDLMRLFDEYGSPSTAGDITYIDYLFGLDGYVDRGQHSLETITLLLALKIEYPNNVHLIRGNHEAADINALFGFRIE
 CIERMGERDGIWAWQRINQLFNWLPLAALIEKKIICMHGGIGRSINHLHQIESLQRPITMEAGSVVLMDDLWSDPTENDS
 VEGLRPNARGPGLVTFGPDRVMEFCKNNDLQLIVRAHECVM DGFERFAQGHILITLFSATNYCGTANNAGAILVLRDLVV
 VPKLIHPPIPAITSPESSPERHLEDTWMQELNVQRPTPTRGRPQAGNDRGSLAWI

Picea glauca

>Pg_GB_BT105039

RGENLLSDLLGSEDTASRRGTEVNYTLEQPHKAANRAPTVDGPVDPVNSPSVSIIMDNKGLTLEKASAAEAEAVSPYF
 KAAKEAAAMDTEGEERPKEEESVHEESENVAASVSNLTSDDPVESDVRLHHRVAVVAKEPVGSLGGLVLRQLSLDQFEN
 EGRMSPWNLDQAHAAKRLIRQKSPVALHKKVISSLLRPRNWKPLNRGFFLDSYEVGELCYTAEQIFMQEPTVLQK
 PVKIFGDLHGQFGDLMRLFDEYGFSTAGDITYIDYLFGLDGYVDRGQHSLETITLLLALKIEYDPHVHLIRGNHEAADIN
 ALFGFRLECIERMGRDRGIWAWTRFNQLFNWLPLSALIERKIICMHGGIGRSINFLEQIESLRRPISMDAGSIVLMDLLW
 SDPTENDSVEGLRPNARGPGLVTFGPDRVLDVFCCKNKLQLIRAHECVM DGFERFAQGHILITLFSATNYCGTANNAGALL
 VLGRELVVIPKLIHPLPPPIRSPESPEHHEETWMQELNIQRPTPTRGRPQPAHDRSLSLAFI

>Pg_GB_BT104224

AEASIKTLAKARESNTEVEQSRDQDRGADSVQSGDLFNPGRSPNVVLPNDPLLTGVRLLHHRVAVVAAETGGALGGLVLRQ
 LSIDQFENEGRVRSYGTPESSAVAARKLLDRQMSINSIPKKVLAYLLKPRGWKPPVRRQFILDCEIAELCESAERIFARE
 PSQLQIKAPVKIFGDLHGQFGDLMRLFDEYGFSTAGDITYIDYLFGLDGYVDRGQHSLETIALLLALKVEHPYVNHVIRG
 NHEAADINALFGFRIECIERLGEKEGIWVWQRINQLFNWLPLAALIEKKIICMHGGIGRSINHVEQIENLQRPITMETGS
 VVLMDDLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCKNNDLQLIRARECVM DGFERFAQGHILITLFSATNYCGT
 ANNAGAILVLRDLVVVVKLIHPLPPAIASAEASPEGHMEDTWMQELNVQRPTPTRGRPQASNDRGSLAWI

Populus trichocarpa

>Pt09g02010

MDVEASLGATDHDPVQDGTSSPTEMDGEQIGEQSPQLAGSQAASTAEVTGSSTTTTAPQOVVQAQQQSPVVGPRHAPTY
 SVVNAIEKKEDGPGPRCGHTLTAVIAVGEETPGYIGPRLILFGGATALEGNSASTGTPSSAGSAGIRLAGATADVHCY
 DVLTKNWSRITPFGPEPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMA
 VGQRYLMAIGGNDGKRPLSDVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDGLLLCGGRDANSVPLASAYGLAK
 HRDGRWEWAIPGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSVAVLDTAAGVWCDTKSVVTSPTRTGRYSADAA
 GGDAAVELTRRCRHAASVSDLIFYGGLRGGVLLDDLVAEDLAAEETSAASHAAAVAAASNVQAGRVPGRYGFVDER
 NRQTMPEAAPDGSVVLGSPVAPPVNGDMYTDISTENAMLPGRRTNKGVEYLVEASAAEAEASATLAAAKARQVNGEVE
 LPDRDRGAEATPSGKQISTLIKPDAGSNNIVPAGVRLHHRVAVVAAETGGALGGMVRQLSIDQFENEGRVRSYGTPESSA
 TAARKLLDRQMSINSVPKVVVHLLKPRGWKPPVRRQFFLDCNEIADLCDSAERIFSSEPSVLQKAPIKIFGDLHGQFG
 DLMRLFDEYGSPSTAGDIAYIDYLFGLDGYVDRGQHSLETITLLLALKVEYVNNVHLIRGNHEAADINALFGFRIECIERM
 GERDGIWVWHRINRLFNWLPLAALIEKKIICMHGGIGRSINHVEQIESLQRPITMEAGSIVLMDLLWSDPTENDSVEGLR
 PNARGPGLVTFGPDRVMEFCKNNDLQLIVRAHECVM DGFERFAQGHILITLFSATNYCGTANNAGAILVLRDLVVVVKLI
 HPLPPAISSPEASPERHIEDTWMQELNANRPPTPTRGRPQVITNDRGSLAWI

>Pt02g22430

MGSKPWLHPAPAYRALETYWSSDDAPGPRCSHTLTAVAATKSHGPRILIFGGVTAIEGGASSAPGISYRLDGATNSVHS
 YDVLTRKWTRIQAAAGEPPSPRAAHAATAVGTMVVFQGGIGPAGHSTDDLVLVDLSTDKFKWHRVVVQGGPGPRYGHVMD
 LVAQRYLIVSNGDKRALSDAVLDTAQKPYAWQRLNPEGDRPSARMYATASARSDFMFLLCGGRDSFGTALGDAYGLL
 MHRNGQEWETLAPGVSPSTRYQHAAVFVGARLHVTTGGALKGGRLVEGEAAVAVLDTAAGVWLDNRNGIVTSSKTSKGHA
 EYDPSLELMRRCRHASASVGVRIYVYGGKGDVLDLVAENSPFQSDMNSPILTSERASTITSPRSHYLNLSFGTTTTPD
 GGSEIPLSGGISMDKNSMEKLEASAAEAEANAVWQAAQAASSNPAEETSVDSDNSQVAEATS DGS DNEADVRLHPRAV
 VVAKEAVGNLGGVLRQLSLDQFENESRRMLPMNNDASYPARKFTRKKSPOGLHKKIISMLLRPNWKAPANRRFFLDSYE
 VGE LCYAAEQIFMQEPTVLQKAPIKVFGDLHGQFGDLMRLFDEYGFSTAGDITYIDYLFGLDGYVDRGQHSLETITLLL
 ALKIENPENVHLIRGNHEAADINALFGFRIECIERMGE SDGIWAWTRFNQLFNCLPLAALIEKKIICMHGGIGRSIHSVE

QIEKLERPITMDAGSIIIMDLLWSDPTENDSIEGLRPNARGPGLVTFPGPDRVSDFCCKNKLQLIIRAHECVMDGFERFAQ
 GQLITLFSATNYCGTANNAGAILVVGRGLVVVVKLIHPLPPLQSPETSPERVIDDAWMQELNIQRPTPTRGRPQPDLD
 RNSLAYI

>Pt01g21560

MDLDASMAATDHDPAVQSASSAPAEMDGEQIGEQSQQPAGSQDTSTAAVAGSSTPTTAPQOVAQAQQQSPVVGPRHAPTY
 SVVNNAMDKKEDGPGPRCGHTLTNVAAVGEEGTPGYIGPRILIFGGATALEGNSASTGTPSSAGSAGIRLAGATADVHCY
 DVLTKWTRITPFGEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMAL
 VGQRYLMAIGGNDGKRPLSDVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDDLGLLCCGGRDANSVPLASAYGLAK
 HRDGRWEWAIAPGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSPTGRYSADAA
 GDAAVELTRRCRHAATAAVGDLIFIIYGGLRGGVLLDDLLVAEDLAAEETTSAASHAAAAAASNVQAGRALGRYGFVDER
 TRQTMPEAAPDGSVVLGNPVAPPVNGDMYTDISTENAMLSGPRRTNKGVEYLVEAAAAEAQAI SATLAAAKARQVNGEVE
 LPDRDRGAEATPSGKQISTLIKPDSSAGSNNIAPAGVRLHHRVAVVAAETGGALGGMVRQLSIDQFENEGRRVSYGTPESA
 TAARKLLDRQMSINSVPKVVVAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPSVLQKAPIKIFGDLHGQFG
 DLMRLFDEYGSSTAGDIEYIDYFLGDIYVDRGQHSLETMTLLLALKVEYPNNVHLIRGNHEAADINALFGFRIECIERM
 GERDGIWVWHRINRLFNLPLAALIEKKIICMHGGIGRSINHVEQIENLQRPITMEAGSIVLMDLLWSDPTENDSVEGLR
 PNARGPGLVTFPGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHILITLFSATNYCGTANNAGAILVLRDLVVVVKLI
 HPLPPAISSPEASPERHIEDTWMQELNANRPPTPTRGRPQVTNDRGSLAWI

>Pt14g12960

MGSKPWLHPAPTYRTRLETYWDSDDDAPGSRCAHTLTAVAATKSQGPRLILIFGGATAIEGGASSAPGIRLAGVTNSVHSYD
 VLTRKWTRVQPAGEPPSPRAHAAAAVGTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGGPGPRYGHAMDLV
 AQRYLVTVSGNDGKRVLSDAWALDTAQKPYVWQRLNPEGDRPSARVYATASARSDDGMFLCCGGRDSSGTPLGDAYGLLMH
 RNGQWEWTLAPGVSPSSRYQHAAVFVGARLHVTTGGSLKGGRLVEGEAAVAVLDTAAGAWLDRNGLVTSSTKSKGHAEYDS
 SFELMRRCRHASASVGVRIYVYGGKGDVLDLVAENSPFPQDINSPTLTSERASTITSPRLNHSNLNSFGTSTPDGG
 PEIPLSGGISMDRNSMEKLEASAAEAEANAVWQAAQAASTNSAEETSVSDDNSQAAEATSDGSDNEADVRLHPRVVV
 AKETVGNLGGMVRQLSLDQFENESRRMVPNNNDASNPAAKFFTRQRSPQGLHKKIISTLLRPRNWKAPANRRFFLDSYEVG
 ELCYAAEQIFMQEPTVLQRLAPIKVFGDLHGQFGDLMRLFDEYGFPSSTAGDITYIDYFLGDIYVDRGQHSLETMTLLLAL
 KIEYPNVHLIRGNHEAADINALFGFRIECIERMGEADGIWAWTRFNQLFNPLAALIEKKIICMHGGIGRSIHSVEQI
 EKLERPITMDAGSIIIMDLLWSDPTENDSIEGLRPNARGPGLVTFPGPDRVTFDFCKNKLQLIIRAHECVMDGFERFAQGG
 LITLFSATNYCGTANNAGAILVVGRGLVVVVKLIHPLPPLQSPETSPEHVIDDTWMQELNIQRPTPTRGRPQPDHDS
 SLAYI

Pteridium aquilinum

>Contig_13341

EANKTQGDTDQSMDEEERDSITDINGNDALGLPSSDVRLLHRAVVVAQEPGGALGGLVRLQLSLDQFENEGRRVSYGNVDS
 GIAARQLLDRQMSRGLHKKIISLILLPRGWKPPTRRQFFLDGNDVAELCDLVQQIFQREPTVLKPKAPLKIIFGDLHGQF
 GDLMLRLFDEYGYPSSTAGDITYIDYLYLR

>Contig_12782

RFEYPEHIHLIRGNHEAADINALFGFRIECIERMGEKDIWAWHRFNQLFNPLAALIEQRIICMHGGIGRSINYVSQI
 ESLQRPTSMAGSVLMDLLWSDPTENDSVEGLRPNARGPGLVTFPGPDRVIAFCNNNLQLIVRAHECVMDGFERFAQGH
 LITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPLQSPETSPEHVIDDTWMQELNSQRPTPTRGRPHTANDRS
 SLAWI

Prunus persica

>Ppa001177

MGSKPWLYPAPTYRTRLETYWDTDGPGPRCGHTLTAVAATKTHGPRILIFGGATAIEGGAASSAPGIRLAGVTNSVHSY
 DVLTRKWTRMRPAGEPPSPRAHAAAAVGTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGGPGPRYGHVMDL
 VAQRYLVTVSGNDGKRVLSDAWALDTAQKPYAWARLNPEGDRPSARMYATASARSDDGMFLCCGGRDISGAPLADAYGLLM
 HRNGQWEWTLAPGVSPSPRYQHAAVFVGARLHVTTGGALRGGRIEAGEGAI AVLDTAAGVWLDNRNGLVTSSTRTGKGQNDYD
 PPELMLRRCRHAASLGVRIYIYGGKGDILLDDFLVAENSSFQSEIGSPVLTSESRPTVTSRPFPHANTSTFGTSPSSD
 GEPESPPSRGLSMDKNSMEKLEASAAEAEASAVWQAVQATSATPAEETSVSDDNSHVAETLSDGSDTEADVRLHHRV
 VVAKEAVGNLGGMVRQLSLDQFENESRRMIPMNNDLSPNKKFFTRQKSPQGLHKKIISTLLRPRNWKAPVNRKFFLDSYE
 VGELCYAAEQIFMQEPTVLQKAPVKVFGDLHGQFGDLMRLFDEYGFPSAAGDITYIDYFLGDIYVDRGQHSLETITLLL
 ALKIEYDPNVHLIRGNHEAADINALFGFRIECIERMGENDGIWAWTRFNQLFNPLAALIEKKIICMHGGIGRSITSVE
 QIEKLERPITMDAGSIIIMDLLWSDPTENDSVEGLRPNARGPGLVTFPGPDRVTFDFCKRKNKLQLIIRAHECVMDGFERFAQ
 GQLITLFSATNYCGTANNAGAILVVGRGLVVVVKLIHPLPPLQSPETSPERVIDTWMQELNIQRPTPTRGRPQPDLD
 RNSLAYI

>Ppa022807

Tipo BSL1, probablemente un pseudogen

>Ppa000783

MDVDSMMVPETDHDPAVPNHSTTSSQAAVEGEQQLGEQSQPPTGSPAQQPQAAAAAAPQAQQTTPVAGPRHAPTYSVVN
 AILEKKEDGPGPRCGHTLTAVAAVGEEGTPGYIGPRILIFGGATALEGNSAASGTPSSAGSAGIRLAGATADVHCYDVL
 NKWSRITPFGEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMALVGQR
 YLMAIGGNDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARSDDLGLLCCGGRDANSVPLASAYGLAKHRDG

RWEWAIAPGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDDTKSVVTSPTGRYSADAAGGDA
 AVELTRRCRHHAAAVGDLIFLYGGLRGGVLLDDLLVAEDLAAAETTSAAASHAAAAAAQSVQGRLPGRYGFIDDRTRQTMP
 EAAPDGAUVLGNPVAPPVNGDMYTDISTENALLQGSRRMSKGEYLVESAAEAEAI SATLAAAKARQVNGEVELPDRDR
 GSEATPSGKQISTLIKPDSSAGSNNIP SAGVRLHHRVAVVAAETGGALGGMVRQLSIDQFENEGRRVSYGTPE SATAARKL
 LDRQMSINSVPKIVIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFCSEPSVLQLKAPIKIFGDLHGQFGDLMLRF
 DEYGSPTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEYAHNVHLIRGNHEAADINALFGFRIECIERMGERDGI
 WAWHRINRFLFNWPLAALI EKKI ICMHGGIGRSINHVEQIENLQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPNARGP
 GLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPA
 ISSPETS PERHIEDTWMQELNANRPPTPTRGRPQVANDRGLAWI

Ricinus communis

>Rc30221g00350

MDVDSTMVPTDQDQSPVQDASPPTAMDREQLRDQQQSAGSPTSVAAA VAVSPPAQQQTSPQAQVQQQSVVGP RHAPTYS
 VVNAIIEKKEDGPGPCGHTLTAVA AVGEEGSPGIYIPRLILFGGATALEGNSAAXXXLAGATADVHCYDVL TNKWSRIT
 PFGEPPTPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMAVVGQRYLMAIGG
 NDGKRPLSDVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWAI A
 PGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDDTKSVVTSPTGRYSADAAGGDA AVELTRR
 CRHAAAVGDLIFLYGGLRGGVLLDDLLVAEDLAAAETTTAASHAAAAAASNHVHAGRIPGSGGRYFGDERSRQTMPESV
 PDGSSVVLGNPVAPPINGDMYTDISTENAVLPGRRMSKGEYLVESAAEAEAI SATLAAAKARQVNGEVELPDRDRGAE
 ATPSGKQISTLIKPE SAGSNNIAPAGVRLHHRVAVVAAETGGALGGMVRQLSIDQFENEGRRVSYGTPE SATAARKLLDR
 QMSINSVPKIVIAHLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFCSEPSVLQLKAPIKIFGDLHGQFGDLMLRFDEY
 GAPSTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEYPHNVHLIRGNHEAADINALFGFRIECIERMGERDGIWVW
 HRINRFLFNWPLAALI EKKI ICMHGGIGRSINHVEQIENLQRPISMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLV
 TFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLLITLFSATNYCGTANNAGAILVLRDLVVVVKLIHPLPPA ISS
 AEASPERHIEDTWMQELNANRPPTPTRGRPQVTNDRGLAWI

>Rc29647g00590

MGSKPWLHPAPTYRPLETYWDTDEDAPGPRCGHTLTAVAATNTHGPRILIFGGATAIEGGGSSSAPGIRLAGVTNSVHSY
 DVLTRKWTRMRPAGEPPSPRAAHAAAVGTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGGPGPRYGHVMDL
 VAQRFLVTVSGNDGKRVLSDAWALDTAQKPYAWQRLNPEGRPTARMYATASARS DGMFLLCGGRDSSGMALGDAYGLLM
 HRNGQWETLAPGVSPSQRYQHAAVFVGARLHVTTGGVLKGGRSVESEAAVAVLDTAAGVWLDKNGLVTS PKTSKGHVEYD
 SSLEQMRRCRHHAAASVGVRIYVYGGLRGDVLLDDFLVAENSQFQSDINSSVLT SERASTITSARINHSSLNPHYDTITDGG
 LEFPSSGGISMDKNSMEKLR EASAAEAEVANAVWQAQAASTNSAEETSASDDNSQAAEATS DGSDEADVRLLSRAVVV
 AKEAVGSLGGLVRLQSLDQFENESRRMI PMNTDTPYPARKFARQKSPQGLHKKIISTLLRPNWKAPANRRFFLDSYEVG
 ELCYAAEQIFMHEPTVLQKAPVKVFGDLHGQFGDLMLRFDEYGF PSTAGDITYIDYFLGDYVDRGQHSLETITLLLAL
 KIEYPENVHLIRGNHEAADINALFGFRLECIERMGESDGIWATR FNQLFNFLPLAALI EKKI ICMHGGIGRSIHSVEQI
 EKIERPITMDAGSI ILMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTD FCKKNKLQLIIRAHECVMDGFERFAQGG
 LITLFSATNYCGTANNAGAILVGRGLVVVVKLIHPLPPPLQSPETS PERVMDDTWMQELNIQRPPPTPTRGRPQPD LDRS
 SLAYI

Selaginella moellendorffii

>Sm00003g04900

MEVEADAPVQESSESEAPKEAEVMESQDAVSSYRVVDAVAESEGDWPGPRCGHTLTALAPIGKEGTAVYIAHRLVLFGGA
 TALEGASGTGPPSPSSGDPGIRIAEVTANVHCFDTASKKWQKLNPSGIAPSPRAAHAAAVGPMVVFQGGIGPAGLSSD
 DLHVLDLSQPEAKWHRVVVKEGPRPGPRYGHVMAVVAQFLILVSGNDGVRLFSVWVLDTTKKPYQWEKVEAGGDGPPPC
 MYAAASTRKDGVLFCGGRDSIGVPVPGIYRLVRNQIGHWHWDRLPGTSPSPRYQHAAAFADARFHI SGGAQGRGHIVDD
 ASSFAVFDTISGVWYEAKSTGASDESALTKRCRHA AVGLKGTMYIYGGKSGVLFSDLLVADDEKLEAACRRNESRATI
 RRDNDLHLLVEASAAEAAAITAIMGKVQSQNGGVAVESVSPSSDL PMPNAMPGSDVRLHHRIVATEPGGGTLRSLVRQ
 LSLDRFENEGLGVTREISKRVLERQISMYGVHKKVLALLKPREWKPSLRNFLFEFNEVLELCDSAENIFRAEPNVLKL
 KAPVKIFGDIHGQFGDLMLRFNFYGSPTAGDISYIDYFLGDYVDRGQHSLETIFLLLALKIEHPEHVHMIRGNHEAQD
 INALFGFRTECVERMGEGVWAWHRMNQLFNWPLAAVIEEKI ICMHGGIGKSIHISQIDNLVLRPITMDTGSNVLM DL
 LWSPTENDSIQGIQPNARGPGLVTFGPDRVMDFCNNNDLDIRAHECVMDGFERFAQGHLLITLFSATNYCGTANNAGA
 LLVLRDLVVVVKLIHPLPPFDEDEMDVTWMQELNAQRPPPTPTRGRPPSVADRSSLAYI

>Sm00032g01790

MLEQPQI SSQPPPPRPAAPVVSQDEQN PASQQAPVPGTPDQAQQQSQQA AAFIRPPRPAPSYRVVNAIIEKKEDGP
 GPRCGHTLTSVA AVGEEQGTAGYIPRLILFGGATALEGNSAAAPT PPPAGNAGIRLAGATADVHCFDIISNKWSRLNAV
 GEPSPRAAHAATAVGS MVVIQGGIGPAGVSS EDLHVLDLTQARPRWHRVVVQGGPGPRYGHVMAVVGQRYLMAIGG
 KRPLADVWALDTAAKPYEWRKLEPEGDGPPPCMYATASARS DGLLLCGGRDASSVPIANAFGLAKHRDGRWEWAIAPGV
 SPSPRYQHAAVFVNARLHVSGGALGGGRMVEDGSSVAVLDTAAGVWC DRKSISSPRTGRYSADAAGGDASSELTRRCR
 HAAAVGDLIFLYGGLRGGVLLDDLLVAEDLAAAETTPAALQAAAIAATAAPNASVPAK PATDGQK PAPSDDGSSVVMEE
 PVAGPVNGDMTMDISSENAVADHR TSKGVE SLVESAAEAEAI SATMAKVAALRQANGEVGS PERDGGADASTSGASPVT
 KGAKRESSTMNHTPQTGVRLYHRAVVA AESSGNLGMVRQLSIDQFENEGRRVSYGTPE SANAARKLLDRTTSEIGVQK
 KVINYLLKPRGWKPPVRRQFFLDVNEILELCEKAERVFQRENTVLQVRAPVKIFGDLHGQFGDLMLRFDEYGSPTAGDI
 TYIDYFLGDYVDRGQHSLETISLLLALKIDYPRNVHLIRGNHEAADINALFGFRIECIERMGEREGIWTWQRINQLFNW

LPLAALIEDKIIICMHGGIGRSINHVEQIENLQRPISMDSGSVILMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVME
 FCKNNNLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLGRDLVVVPKLIHPLPPAVTSPESPPEHQL
 EDTWMHELNVQRPPPTPTRGRPHPAGNERSVAAWI

Setaria italica

>Si021168m

MGTAGKNWVVPAPAYREVEGWEGVGEDAPGFRCGHSLTAVAPTCKGHGPRLLIFGGATAIEAGASSGLPGIRLAGVTNSV
 HSYDVEKRRWTRLHPAGEPPSPRAAHSAAAVGTMVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGAGPGPRYGHC
 MDLVAQRYLVTVSGNDGKRVLSDAWALDTAQKPYRWQKLSPEGDRPSARMYATASSRSDGMLLLLCGGRDASGTPLSDAYG
 LLMHTNGQWEWTLAPGVSPSPRYQHAAVFGARLHVTTGGVLRGGRAIEGEGAI AVLDTAAGVWLDNRNGIVTSRTLKSSNE
 HDASSDLLRRCRHAAASVGSQIYIYGGLRGDILLDDFLVAENAPFQSEITSSMYGADRVPRGETQNRNHNYYSDSPVQQS
 SNNSTDKKSIDMLIEASTAEAEAVSAVWRAAKEASAASSEDLSSEGIGSESPLETSMPMEDFDDGGSLPEPDKLHSRAV
 VVAKEAVGDLGCLVRQLSLDQFENESRRMHPSSNDQSYPGRRALNRQSPQGLHKKIISFLLKPRNWRAPADRTFFLDSY
 EVGELCYAAEQIFMQEPTVLQKAPIKVFGLDHGQFGDLMRLFDEYGFSTAGDITYIDYFLFGDYVDRGQHSLETITLL
 LALKIEYPENVHLIRGNHEAADINALFGFRLECIERMGESDGIWAWTRFNQLFNYLPLAAMI EKKIICMHGGIGRSINTV
 EQIEKLERPITMDVGSIIILMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVTEFCRKNKLQLIRAHECVMDGFERFA
 HGQLITLFSATNYCGTANNAGAILVVGRLVVVPKLIHPLPPVNSPESPSPERGDATWMQELNIQRPPPTTRGRPQAAGD
 RNSLAYI

>Si034069m

MDVDSRMTTESDSDSDAAAAAASAGGSGSGSETSSSSAPSTPGTPTGTPAAAPNPAAVGAAGPRPAPGYTVVNAVIEKKED
 GPGCRGHTLTAVPAVGEETPGYIGPRLIFGGATALEGNSATPPSSAGSAGIRLAGATADVHCYDVSSNKWSRLTPLG
 EPPSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMALVGQRFLLTIGGNDG
 KRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLLLCGGRDANSVPLSSAYGLAKHRDGRWEWAIAIPGV
 SPSRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTTPTRTGRYSADAAGGDAVELTRCRH
 AAAVNDLIFIYGGLRGGVLLDDLLVAEDLAAAETTNAANHAAASAAATNLQAGRAPGRYAYNDERTRQTATESGPDGAV
 VLGTPVAPPVNGDMYTDISPENAVLQQRRLNKGVDYLVEASAAEAEIASATLAAVKARQVNGEVEQLPDKEQSPDSTSS
 SKHSSLIKPDIVLSNNMAAPPVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPENATAARKLLDRQMSI
 NSVPKVIASLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPSVLQKAPVKIFGDLHGQFGDLMRLFDEYGAPS
 TAGDIAYIDYFLFGDYVDRGQHSLETITLLLALKVEYQNVHLIRGNHEAADINALFGFRLECIERMGERDGIWTHRMN
 RLFNWLPLAALIEKKIICMHGGIGRSINHVEQIENLQRPITMEAGSVVLMDDLLWSDPTENDSVEGLRPNARGPGLVTFGP
 DRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLGRDLVVVPKLIHPLPPAITSPEAS
 PDHIEDTWMQELNANRPPPTTRGRPQAVANDRGS LAWI

>Si021021m

MKHPTKQPIRSTKVVSHCTPPSPSVVSASSPALFFHSTPRPPPHSHSHFQPPPLPRASPLTSPPRDPIAPRLDPTQVA
 TRRGAPPPPPPPPPRLAAAGERGSLGLARGDGGRAARRQLDRLLVQRLSEAELVRARGGSPCSARETLGGAAGGEAWR
 GLAAMEVDARMATESDSDSDARSGGGGGGGGGSGSGSETSPSAPPSPAGTPTATAAATSPGPVAGPRPAPGYTVVDAAMDK
 KEDGPGCRGHTLTAVPAVGEESPGYVGPRLIFGGATALEGNSATPPSSAGSAGIRLAGATADVHCYDVLSNKWTRLT
 PLGEPSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMALVGQRFLLTIGG
 NDGKRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLLLCGGRDANSVPLSSAYGLAKHRDGRWEWAIA
 PGVSPSPRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTTPTRTGRYSADAAGGDASVELTR
 CRHAAAAGDLIFVYGGLRGGVLLDDLLVAEDLAAAETTSAANHAAVAASANVQREPGRYAYNDEQSGQTVTVSSPDGA
 VVLGTPVAPPVNGDMYTDISPENAVLQQRRLSKGVDYLVEASAAEAEIASATLAAVKARQVNGEMEHS PDREQSPDAAS
 SGKQTSLLIKPDPALLNNTPPPGVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVVIYGTENATAARKLLDRQM
 SINSVPKVIASLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPSVLQKAPVKIFGDLHGQFGDLMRLFDEYGA
 PSTAGDIAYIDYFLFGDYVDRGQHSLETITLLLALKVEYPHNVHLIRGNHEAADINALFGFRLECIERMGERDGIWTHR
 VNRLFNWLPLAALIEKKIICMHGGIGRSINHVEQIENLQRPITMEAGSVVLMDDLLWSDPTENDSVEGLRPNARGPGLVTF
 GPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLGRDLVVVPKLIHPLPPAITSPE
 TSPEHHIEDTWMQELNANRPPPTTRGRPQAPNDRGSLAWI

Sorghum bicolor

>Sb01g013460

MDVDSRMTTESDSDSDAAAAAAGGSGSGSETSSSSAPSTPGTPTGTPAAAPNPAAVGAAGPRPAPGYTVVNAVIEKKED
 GPGCRGHTLTAVPAVGEETPGYIGPRLIFGGATALEGNSATPPSSAGSAGIRLAGATADVHCYDVSSNKWSRLTPLG
 EPPSPRAAHVATAVGTMVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPRYGHVMALVGQRFLLTIGGNDG
 KRPLADVWALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLLLCGGRDANSVPLSSAYGLAKHRDGRWEWAIAIPGV
 SPSRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTTPTRTGRYSADAAGGDAVELTRCRH
 AAAVNDLIFIYGGLRGGVLLDDLLVAEDLAAAETTNAANHAAASAAATNLQAGRAPGKYAYNDERTRQTATESGPDGAV
 VLGTPVAPPVNGDMYTDISPENAVLQQRRLSKGVDYLVEASAAEAEIASATLAAVKARQVNGEMEQLSDKEQSPDSTST
 TKHSSLIKPDSTLSNNMAVAPVRLHHRVVVAAETGGALGGMVRQLSIDQFENEGRVSYGTPENATAARKLLDRQMSI
 NSVPKVIASLLKPRGWKPPVRRQFFLDCNEIADLDCSAERIFSSEPSVLRKAPVKIFGDLHGQFGDLMRLFDEYGAPS
 TAGDIAYIDYFLFGDYVDRGQHSLETITLLLALKVEYQNVHLIRGNHEAADINALFGFRLECIERMGERDGIWTHRMN
 RLFNWLPLAALIEKKIICMHGGIGRSINHVEQIENLQRPITMEAGSVVLMDDLLWSDPTENDSVEGLRPNARGPGLVTFGP

GDASVELTRRCRHA AAAVGD LFIYGG LRGGVLLD DDLVAEDLAA AETTYAASHAAAAATTNSLPSRLPGRYGFS DERNR
 ELADSAADGAVVLASPVVPSVNRDMYTDI SSEDAI LPVTRRTSKGVEYLVEASAAEAEAI SATLAAAKARQVNGEVELPD
 RDCGAEATPSGTAASSVFKPDSMGVSVGVT PAGIRLHHRVVA AETGGALGGMVRQLSIDQFENEGRVSYGTPESATAA
 RKLDRQMSINSVPKVI AHLLKPRGWKPPVRRQFFLDCNEIADL CDSAERIFSS EPTVLQKAPIKIFGDLHGQFGDLM
 RLDFEYGS PSTAGDISYIDYFLGDYVDRGQHSLETIS LLLLALKVEYQHNVHLIRGNHEAADINALFGFRIECIERMGER
 DGIWVHRINR LFNWLP LAASIEKKIICMHGGIGRSINHVEQIENIQRPITMEAGSIVLMDLLWSDPTENDSVEGLRPN
 RGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPL
 PPAISSSEASPERDTWMQELNANRPATPTRGRPQTANEKGGSLAWM

>Sp6g03110

MGTKPWLHPAPQYKLTETFW DDEDDAPGPRCAHTLTAVAATKTHGPRLILFGGATAIEGGSSSVPGIRLAGVTNTVHSYD
 VLTRKWTRLKPAGEPPSPRAHAAAAVGT MVVFQGGIGPAGHSTDDLYVLDMTNDKYKWHRVVQEGEGPGRYGHVMDLV
 SQRYLVTVTGNDGKRALS DAWALDTAQKPYVWQRLNPDGDRPSARMYATGSARS DGMFLLCGGRDVTGAPLGDAYGLLMH
 RNGQWEWTLAPGVAPSPRYQHAAVFGARLHVSGGVLGRGRVIDAEASVAVLDTAAGVWLDNRNGQVTSVRGSKGQIDQDP
 SFELMRRCRHGAASV GIRIYVHGGLRGDVLDDFLVAENSTFQSDISSPLASDRTQSSSTPRFSYAARPPSGSEPALST
 SEGSLDENSLKLEASAAEAVASSVWRAAQLGGSASLDEEPSASDGSSTIGSTSDGPETEEDVRLHPRAVVAKET
 VGS LGGMVRQLSLDQFQNESRRMVP MNNSDVPQPTKFKTRQKSPQGLHKKVISALLRPRNWKPPGNRKFFLDSYEVGELC
 YAAEQIFMHEQTVLQKAPIKVF GDLHGQFGDLMRLDFEYGF PSTAGDITYIDYFLGDYVDRGQHSLETITLLLALKIE
 YPENVHLIRGNHEAADINALFGFRLECIERMGENDGIWAWTRFNQLFNYLPLAAI IENKIIICMHGGIGRSISSVEQIEKI
 ERPI TMDAGSLVMDLLWSDPTENDTIEGLRPNARGPGLVTFGPDRVTEFCCKNKLQLIIRAHECVMDGFERFAQGQLIT
 LFSATNYCGTANNAGAILVVRGLVIVPKLIHPLPPP IISPEHSPEHSGDDAWMQELNIQRPTPTRGRPQPDFDRSSLA
 YI

>Sp1g37500/Sp1g37490

MADFDSWQFPFPAYETARTYFNTQDDLPGPRCGHTLTAILHRLILFGGFTSVGNDASEFSFDAVTNSVHCLNVFTRKWTR
 VYPEGEPSPRAGHA AVSCGSAVLIQGGIGPSGLCNGDLYILDMTTDTFKWKKVVVREGLAPGPRYGHVIGLADNCIVIF
 GGNDGNETLADT WALNTTNKPNVWERLYPYGDLPSGRMYASASTRDYRHFMLCGGRDNLGMPLGDAYKLEPLSGGCWLWT
 RGPGFDPSPRYQHAAV FVKNR LHVIGGSLSTQLVDAEAVVSLD TTNGVWVDTRDGPVSVQLDSNGRYFNQVMQRS
 GHA AASVGSRIYVYGGFKEGVLLD LASKDDRAQSSSGPATPSLLDMNYRYTPQFLSQRSTQDLVKKVISTLLRPQTWEPP
 VDRKFFLSYDEVDLCAA AKEIIVKEPTVLQLNAPVKVFGDLHGQFRDLMKLFYEYGYPSKNGDILYIDYFLGDYVDRG
 THSLETFLLLLALKVEY PKNIH LIRGNHESKAMNLIYGFHAECKERMGWNDGENAFALITELFNHLPLAALIEKKIFCMH
 GGIGSSVMTVEQVANIKRPVLLDYEAESRVKDLLWSDPTSSDDILGIVVNPRGPGIVSFGPDRVKEFCEKNNIELIIRA
 HECVLDGFERFAQGKLI TVFSATNYCGNMN NAGAILVIGRDLAIVPKLIHPLPPP IASENVSDKAWMEVDSERPPTPVR
 GGPRSYFERGSTS

Theobroma cacao

>Thecc1EG004502t1

MGSKPWLYPAPTYRTLETY WDSDE DAPGPRCGHTLTAVAATKTHGPRLILFGGATAIEGGASSAPGIRLAGVTNSVHSY
 DVLTRKWTRIRPAGEPPSPRAHAAAAVGT MVVFQGGIGPAGHSTDDLYVLDLTNDKFKWHRVVVQGGGPGPRYGHVMDL
 VAQRYLVTVSGNDGKRVLS DAWAFDTAQKPYVWQRLSPEGDRPSARMYATASARS DGMFLLCGGRDSSGAPLADAYGLLM
 HRNGQWEWTLAPGVSPSPRYQQA AVFVGARLHVTTGGTLRGRAIEGEAAVAVLDTAAGVWLDNRGLVTSRNTKGHAEYD
 PSLELMRRCR HASASVGR IYIYGG LRGDMLLDDFLVAENSPFQSDMSSPILTSERTSNLSSPRFNQPNLSSFGTPTSD
 DGPEFPSSGMSMDKNSMEKLR EASAAEAEANAVWQAQAASAVPAEETSLSDDNSPAAETASDASDNEADVRLHPRAV
 VVAKEAVGNLGGMVRQLSLDQFENESRRMIPSNNDLSYPTKFKSRQKSPQGLHKKVISTLLRPRNWKAPVNRFFFLDSYE
 VGELCYAAEQIFMQEPTVLQKAPVKVFGDLHGQFGDLMRLDFEYGF PSTAGDITYIDYFLGDYVDRGQHSLETITLLL
 ALKIEYPENVHLIRGNHEAADINALFGFRIECIERMGENDGIWAWTRFNQLFNYLPLAALIEKKIICMHGGIGRSIHSVE
 QIEKLERPI TMDAGS IILMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTFCKKNKLQLIIRAHECVMDGFERFAQ
 GQLITLFSATNYCGTANNAGAILVVRGLVIVPKLIHPLPPP IQSPETSPERVADDTWMQELNIQRPTPTRGRPQPDLD
 RSSLAYI

>Thecc1EG041219t1

MDVDSMVPETDHDQNHGGASQSPPPMEREQQQQQQQQSQQLENQASPPGPGGSPPPPAQQQAAAAAATTQVQPQTPVVG
 PRCAPTYSVVNAIIEKKEDGPGPRCGHTLTAVA AVGEEGTPGYIGPRLILFGGATALEGNSAASGTPSSAGSAGIRLAGA
 TADVHCYDVL TNKWSRISPFGEPTPRAAHVATAVGT MVVIQGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGGPGPR
 YGHVMALVQRYLMAIGGNDGKRPLADV WALDTAAKPYEWRKLEPEGEGPPPCMYATASARS DGLLLLLCGGRDANSVPLA
 SAYGLAKHRDGRWEWAIAPGVSPSPRYQHAAVFNARLHVSGGALGGGRMVEDSSSVAVLDTAAGVWCDTKSVVTSRPTG
 RYSADAAGGDASVELTRRCRHA AAAVGD LFIYGG LRGGVLLD DDLVAEDLAA AETTTAASHAAAAAASNVHAGRLPGR
 YAFVDERTRQTMPEAVPDG SVVLGNPVAPPVNGDMYTDISTENAMLQGSRRMSKGVEYLVEASAAEAEAI SATLAAAKAR
 QVNGEVELPDRDRGAEATPSGKQMS TLIKMPDSAGSNSVAPAGVRLHHRVVA AETGGALGGMVRQLSIDQFENEGRV
 SYGTPESATAARKLLDRQMSINSVPKVI AHLLKPRGWKPPVRRQFFLDCNEIADL CDSAERIFSTEPSVLQKAPIKIF
 GDLHGQFGDLMRLDFEY GAPSTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEY PNNVHLIRGNHEAADINALFGF
 RIECIERMGERDGIWVHRINR LFNWLP LAALIEKKIICMHGGIGRSINHVEQIENLQRPITMEAGSIVLMDLLWSDPTE
 NDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRD
 LVVVPKLIHPLPPAMSSPEASPERHI EDTWMQELNANRPPTPTRGRPQVTNDRGSLAWI

Utricularia gibba

>Ug00227_g12007

MDADSTMASEADYDPSARPIPDNQADNNNPTESEASTPQQKQQQQQPSNNFNCSNPGAGGFGAALPGPRCAPTYSVVN
 AIEKKEDGPGPRCGHTLTAVPAVGEESPGYIGPRLILFGGATALEGNSAGSGTPTSAGSAGIRLAGATADAHSYDVLT
 NKWSRITP IGEPTTPRAAHVATAVGTMVVIQGGIGPSGLSAEDLHVLDLTQQRPRWHRVVVNGLPGPPRYGHVMAVVGQR
 YLMAIGGNDGKRPLADVWALD TAAKPYEWRKIEPEGEGPPPCMYANASARSDGLLLCGGRDANSVPLAGAYGLAKHRDG
 RWEWA IAPGVSPSPRYQHAAVFVNARLHVSGGALGGGRMVEDCSSVAVLDTAAGVWCDTKSVVTSPIGRYSADAAGGDA
 AVELTRRCRHA AAAVGD LIFVYGG LRGVLLDDLLVAEDLAAAETTSAASHAAAAASMP SMQQGRDQSR YGFADDKNGVS
 VNDGLCNGAVVVGSPVAPPINGDVYTDISTENAILHASRSSEGVEYLVEAAAAEAEAI CATLAAAKARQTNNGDAELSDKD
 SSTEASPGAKSAPT VSKPDSTLPDASVPAGVRLHHRVAVLTAEAGGALGGMVRQLSIDQFEHEGRRVSYGTPE SAVAARK
 LFDRQMSITSVPKVLHLKPRGWKPPVQRQFFLDNEIADLCNAERIFASEPTVLQKAPIKLFGLHGQFGDLMLRL
 FDEYGSPTAGDIAYIDYFLGDYVDRGQHSLETITLLLALKVEHPHQVHLIRGNHEAAEINALFGFRIECIERMGERDG
 IWVWHRMNRFLFNWLPLAALIEKKIICMHGGIGRSINHVEQIESIQRP IAMEAGSIVLMDLLWSDPTENDSVEGLRPNARG
 PGLVTFGPDRVMEFCNNNDLQLIVRAHECVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPP
 AMSSPETTS PERHI ED TWMQELNANRPPTPTRGR TQVAGDRGSLAW

>Ug169_g11381

MESKSWLYPAPAYRPLESFWDTDDDAPGPRCGHTLTAVAATKTHGPRILIFGGATAIEGGNGGAPGIRLAGVTNSVHSYD
 VLTRKWTRLR PAGEPPSPRAAHTAAAVGTMVVFQGGIGPAGHSTDDLYVLDMSSNNKFKWHRTGTGTQIWRPHGPGCSEIS
 CYSRRLSDAWALDTAQKPYVWQRLNPEGDRPSARMYATGCARS DGMFLLCGGRDASGAPLADAYGLLMHRS GMWEWTLA
 PGVSPSPRYQQASV FVGARLHVTTGGVVRGGRTVESDAAITAGVWLD RNLVTNRGNKGQSDVDPSSSEAMRRRCRHAASV
 IRIYVYGG LRGD TLLDDFLVADNSQFQSEFI PDRPLSSTIPRMEGVYVAPSCTSSSPVDGLE SPLSGSLGMDKSSMEK
 LAKASAAEAEVANAMWQSVQGASAFPEEAVVSDGNSQSVESTSEGS DSDADVRLHPKAVV VAKETVGNLGGMLRQLSLDQ
 FDHESRRVIPLSNDLSNPTK KFI RQSPQGLHKKVISTLLRPNWKPVNRFKFLDSYIEGELCYAAEQIFVNEPTVLQL
 KAPVKECPENIHLIRGNHEADINALFGFRIECIERMGE SDGLWAWTRVNQLFNFLPLAALIEKKIICMHGGIGRSINS
 IEQIEKLERPITMDAGSIVLMDLLWSDPTENDSVQGLRPNRGPGLVTFGPDRVTEFCKRNKLQLIIRAHECVMDGFERF
 AQGQLITLFSATNYCGTANNAGAILVIGRGLVIVPKLIHPLPPVQSPETSPERVADEAWMQELNIQRPTPTRGRPQPE
 YDRNSLAYI

Vitis vinifera

>Vv07g04070

MAGSKPWLYPAPTYRTLESYWD TDDDAPGPRCGHTLTAIAPT KTHGPRILIFGGATAIEGGASSAPGIRLAGVTNSVHS
 YDVLTRKWTRLR PAGEPPSVRAHAAAAVGT MVVFQGGIGPAGHSTDDLYVLDTNDKYKWHRVV VQGGPGPRYGHVID
 LVAQRYLVTSGNDGKRVLSDAWALDTAQKPYAWQRLNPEGDRPSARMYATASARS DGMFLLCGGRDSSGAPLADAYGLL
 MHRNQWETVTLAPGVSPSPRYQHAAVFVGARLHVTTGGALRGGRAIEGEAAVAVLDTAAGVWLD RNLVTSPRINKGHTY
 DPSELMLMRRCRHAASVGVRIYIYGG LKGDMLDDFLIAENSAFQSDINSPILTS DRAATITSPKLNQSNANLYGTTPTP
 DGRLEVPSSGAASMDKSSMEKLAESA AEA EAVSAVWQAMQAASAVPSDETSVSDDNSQA AETTSDGSDTEADVRLHPRA
 VVVAKEAVGNLGGMVRQLSLDQFENESRRMVP MNNDLSYPTKFT RQKSPQGLHKKIISTLLRPNWKAPANRRFFLDSY
 EVGELCYAAEQIFMHEPTVLQKAPIKVFGDLHGQFGDLMLR L FDEYGFSTAGDITYIDYFLGDYVDRGQHSLETITLL
 LALKIEYPENVHLIRGNHEADINALFGFRLECIERMGE SDGIWAWTRFNQLFNFLPLAALIEKKIICMHGGIGRSISSV
 EQIEKLERPITMDAGSIIIMDLLWSDPTENDSIEGLRPNARGPGLVTFGPDRVTD FCKKNKLQLIIRAHECVMDGFERFA
 QGQLITLFSATNYCGTANNAGAILVVRGLVVPKLIHPLPPPLQSPETSPERTVDDTWMQELNIQRPTPTRGRPQPD
 DRNSLAYI

>Vv06g12810

MEREQQLGAAEQSAPAQPPAQQQQQQQQAQVQQSPVAGPRHAPTYSVVNAVIEKKEDGPGPRCGHTLTAVAAVGEETP
 GYIGPRLILFGGATALEGNSAASGTPSSAGSAGIRLAGATADVHCYDVITNKWSRITPFGEPPTPRAAHVATAVGTMVVI
 QGGIGPAGLSAEDLHVLDLTQQRPRWHRVVVQGGPGPPRYGHVMAVVGQRYLMAIGGNDGKRPLADVWALD TAAKPYEWR
 KLEPEGEGPPPCMYATASARS DGLLLCGGRDANSVPLASAYGLAKHRDGRWEWA IAPGVSPSPRYQHAAVFVNARLHVS
 GGALGGGRMVEDSSVAVLDTAAGVWCDTKSVVTSPTRTGKYSADAAGGDAVELTRRCRHA AAAVGD LIFVYGG LRGVLL
 LDDLLVAEDLAAAETTSAASHAAAAAASNLQAGRLPGRYGFIDERTRQTIPEAAPDGSVVLGTPVAPPVNGDMYTDIST
 ENAMLQGSRRLSKGV EYLVEASAAEAEAI SATLAAAKARQVNGEVELPDRDRGAEATPSGKQISTLIKMPDSAGSNNIAP
 AGVRLHHRVAVVAETGGALGGMVRQLSIDQFENEGRRVSYGTPE SATAARKLLDRQMSINSVPKVI AHLKPRGWKPP
 VRRQFFLDCNEIADLCDSAEERIFSSPE SVLQIKAPVKIFGDLHGQFGDLMLR L FDEYGSPTAGDIAYIDYFLGDYVDRG
 QHSLETITLLLALKVEYPLNVHLIRGNHEADINALFGFRIECIERMGERDGIWAWHRINRLFNWLPLAALIEKKIICMH
 GGIGRSINHVEQIENLQRPI TMEAGSIVLMDLLWSDPTENDSVEGLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVRAHE
 CVMDGFERFAQGHLITLFSATNYCGTANNAGAILVLRDLVVPKLIHPLPPAISSPETSPERHLED TWMQELNANRPPT
 PTRGRPQVTNDRGSLAWI

Welwitschia mirabilis

YQHAAVFVNARLHVSGGALGGGRMVEDCSSVAVLDTAAGVWCDTKGVVTS SAHTGRYSADAAGGDASSELTRRCRHA AAAV
 GD LIFVYGG LRGVLLDDLLVAEDLAAAETTSAASHAAAVAAAAA IANAHPGGNAGKNVYSDGPSRHISSDSGDGGSVVM
 GSPVAPPINGDMFRDISTENALVPGHRR LGRGVE SLVEASAAEAEIAAALAAAKARQSNGEADQHPDRDRGAEATPSGK
 QPSPLSKTMDSVMTNYSSPSGVRLHHRVAVVAETGGALGGMVRSLSIDQFENEGRRVSYGTPE SANAARKLLDRQMSIN
 SVPKKILTYLLKPRGWKPPVRRQFFLDCNEISELCDGAERIFSQEPSVLQIKAPVKIFGDLHGQFGDLMLR L FDEYGA
 PSTAGDISYIGYFLGDYVDRGQHSLETISLLL

Anexo II

Lista de oligonucleótidos usados

Clonado Gateway (genérico)

attB1 GGGGACAAGTTTGTACAAAAAAGCAGGCT
attB2 GGGGACCACTTTGTACAAGAAAGCTGGGT

Genotipado de líneas de inserción en genes BSL

LB1 GGCAATCAGCTGTTGCCCGTCTCACTGGTG
p745 CGTCCGCAATGTGTTATTAAGTTGTC
S_147279-RB GCTTGCATTGAAGGCAAGACGGT
S_147279-LB CTCTGGAGATAAGATGGGA
S_051383-RB CCTGGCAAAAGTACAGAGCAG
S_051383-LB TCTCATGCGTTGGACATATTGG
S_032895-RB GTCTATGGTGGCTGACAACGATC
S_032895-LB CGTTCCTTCATCACCAACAGCAG
S_055335-RB GTACATTAGCAAAGTTCTGCCAG
S_055335-LB GTTCCAGAGCAGATGGAGATGT
DsLx245-RB CGACCAAAACATTAAGCAGAC
DsLx245-LB AGACTCGCTTAGTGATTCTGGAGT
S_072437-RB CTATGTTGTTTCCGTGATCCTTG
S_072437-LB GCAAAACGCTCAAATCCGTCCATC
S_151962-RB GGTGGCGAGTGAAACTGCTTAC
S_151962-LB GCCAACGTATTCGGTTGTGAATG

Construcciones etiquetadas de los genes BSL

BSL1-FW AAAAAAGCAGGCTTTCCAATGGGGCTCGAAGCC
BSL1-RV AGAAAAGCTGGGTAGATGTATGCAAGCGAGCTTC
BSL2-FW AAAAAAGCAGGCTTTTTGATAATGGATGAAG
BSL2-RV AGAAAAGCTGGGTACATCCAAGCCAGAGAACCCT
BSL3-FW AAAAAAGCAGGCTTGAAAAATATGGATTTGGATTTC
BSL3-RV AGAAAAGCTGGGTATATCCAAGCAAGAGAGCC
BSU1-FW AAAAAAGCAGGCTGTGTTATGGCTCCTGATCA
BSU1-RV AGAAAAGCTGGGTATTCACCTTGACTCCCCTCG
BSL2Latt1ATG AAAAAAGCAGGCTTTTTGATAATGGAAGATCTTGCTGCTGC
BSL2Patt1ATG AAAAAAGCAGGCTTTTTGATAATGGAACCAAAGAAGGTTATAGC
BSL1Latt1ATG AAAAAAGCAGGCTTTTGATAATGGGTGAAAATTCACATTC
BSL1Patt1ATG AAAAAAGCAGGCTTTTGATAATGGAACATAAAAAAGGTCATTGC

Construcciones para doble híbrido

BSL1-FW AAAAAAGCAGGCTCGGGCTCGAAGCCTTGGCT
BSL1-RV AGAAAAGCTGGGTTTCAGATGTATGCAAGCGA
BSL2-FW AAAAAAGCAGGCTCCGAATTCGATGAAGATTCG
BSL2-RV AGAAAAGCTGGGTGCGACTCACATCCAAGCCAGAG
BSL3-FW AAAAAAGCAGGCTCGTCTAGAGATTTGGATTCCTTC
BSL3-RV AGAAAAGCTGGGTGCGACTTATATCCAAGCAAGAG
BSU1-FW AAAAAAGCAGGCTCGTCTCCTGATCAATCTTATCA
BSU1-RV AGAAAAGCTGGGTGCGACCTTTATTCACCTTGACTC
BSK1-FW AAAAAAGCAGGCTCCGAATTCGGTTGTTGTCAATC
BSK1-RV AGAAAAGCTGGGTGCGACTCAAGATCCTCTGCCGCC
BSK3-FW AAAAAAGCAGGCTCCGAATTCGGAGGTCAATGCTC
BSK3-RV AGAAAAGCTGGGTGCGACTTACTTCACTCGGGGAAC
BSL1P-5' AAAAAAGCAGGCTCCAGATCTCATAAAAAAGGTCATTGCT
BSL1-3' STOP AGAAAAGCTGGGTGTCGACTCAGATGTATGCAAGCGAGC
BSL2P-5' AAAAAAGCAGGCTCCGAATTCCTCAAAGAAGGTTATAGCT
BSL2-3' STOP AGAAAAGCTGGGTGCGACTCACATCCAAGCCAGAGAAC
BSL3P-5' AAAAAAGCAGGCTCCGAATTCCTCAAAGAAGGTTGTAGC
BSL3-3' STOP AGAAAAGCTGGGTGCGACTTATATCCAAGCAAGAGAG
BSU1P-5' AAAAAAGCAGGCTCCGAATTCCTCAAAGAAGGTCATTTC
BSU1-3' STOP AGAAAAGCTGGGTGCGACCTTTATTCACCTTGACTCC
BSL2K-3' STOP AGAAAAGCTGGGTGCGACTCAGCAGCAAGATCTTCAGCA

Deleciones del dominio conector de BSL2

BSL2 Δ 450-555 AAAAAAGCAGGCTCCGAATTCGCCAAGGCAAGACAAGTCAATG
 BSL2 Δ 450-599 AAAAAAGCAGGCTCCGAATTCGCAGGGATTCGTCTTCATCATA
 BSL2 Δ 450-642 AAAAAAGCAGGCTCCGAATTCGCTCCAGAAAGTGCAACAGCAG
 BSL2 Δ 450-499 AAAAAAGCAGGCTCCGAATTCGCTGTGGTGCTGGGAAGC
 BSL2 Δ 450-528 AAAAAAGCAGGCTCCGAATTCAGAACGAATAAGGGTGTG

Quimeras entre BSL1 y BSL2

QuimeraB2B1-fw GCTGCATCGATTGAGAACAAAATTATATGTATGCATG
 QuimeraB2B1-rv CATGCATACATATAAATTTTGTTCCTCAATCGATGCAGC
 QuimeraB1B2-fw GCTGCACTAATCGAGAAGAAAAATTATCTGTATGCATG
 QuimeraB1B2-rv CATGCATACAGATAAATTTTCTTCTCGATTAGTGCAGC
 TOPP15' GAGAATTCGCGGAGAAAGCCGGCGCA
 TOPP13' GACGTCGACTTACAGAAAAGGCGACTTCT
 BSL1L5'Eco GACGAATTCGAAAATTCAACATTCCAG
 BSL1L3'Eco GACGAATTCAGGCCTTGCGGAGACTTC
 BSU1L5'Eco GACGAATTCGAAACATCCCAGTCTTC
 BSU1L3'Eco GACGAATTCGAAGTCTTGAGTAGATTG

Mutagénesis sobre BSL2

BSL2S627E-fw GAATGGTTAGACAGTTAGAGATTGATCAATTTGAAAATG
 BSL2S627E-rv CATTTTCAAATTGATCAATCTCTAACTGTCTAACCATTTC
 BSL2S627A-fw GAATGGTTAGACAGTTAGCTATTGATCAATTTGAAAATG
 BSL2S627A-rv CATTTTCAAATTGATCAATAGCTAACTGTCTAACCATTTC
 BSL2RVxF-fw GAAAAATGAGGGACGGCGTATGGTACCCATGAACCCAGAAAGTGCAACAGC
 BSL2RVxF-rv GCTGTTGCACTTTCTGGGTTTCATGGGTACCATACGCCGTCCCTCATTTTC

Expresión del dominio conector de BSL2

BSL2lk5' Nde GGCCATATGGCTGCTTCTCATGCTGC
 BSL2lk3' Xho CGACTCGAGACTGTTGATTGACATTTG