UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS
CENTRO DE CIENCIAS GENÓMICAS

ANÁLISIS DE LA REGULACIÓN DE LA EXPRESIÓN GÉNICA DE *PHASEOLUS VULGARIS* A NIVEL TRANSCRIPCIONAL Y POST-TRANSCRIPCIONAL, ENFOCADO EN EL SPLICING ALTERNATIVO

TESIS
PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:
LUIS PEDRO IÑIGUEZ RÁBAGO

DRA. GEORGINA HERNÁNDEZ DELGADO
CENTRO DE CIENCIAS GENÓMICAS, UNAM

DR. ENRIQUE MERINO PÉREZ
INSTITUTO DE BIOTECNOLOGÍA, UNAM

DR. CEI ABREU GOODGER
LABORATORIO NACIONAL DE GENÓMICA PARA LA BIODIVERSIDAD, CINVESTAV

CUERNAVACA, MORELOS, MÉXICO, NOVIEMBRE 2017

Quiero dedicar esta tesis
a mi familia, a mis amigos
y todos lo que hicieron esto posible

Agradecimientos

Primero que nada quisiera agradecerle a Gina, la Dra. Georgina Hernández, por haberme recibido en su laboratorio, por confiar en mí y mandarme a Minnesota a colaborar en un proyecto del cual teníamos poca idea de cómo abordarlo. Gracias también por ser mi tutora en la tesis de licenciatura y aceptarme para realizar mis estudios doctorales en un proyecto bioinformático, del cual ambos sabíamos muy poco pero queríamos aprender mucho. Pero sobre todo muchísimas gracias por permitirme conocer en ti a una mujer tan admirable y aprender tanto de ti.

Muchas gracias a todos los miembros del laboratorio: Víctor, Maru, Sara y Alfonso; por recibirme y permitirme formar parte del grupo. Quisiera agradecerle también a Marlene por ser mi confidente y escucharme tanto en mis buenos como en mis malos momentos y a Mario Ramírez porque siempre que a lo lejos se escuchaba una voz ronca cantando una canción yo sabía que se avecinaba una buena plática sobre las Chivas, la Selección Mexicana o el Real Madrid. Gracias también a los miembros que formaron parte del laboratorio y que ahora forjan su camino en otros lados; Bárbara, Ana Belén, Sole y Sujay. Y también quisiera agradecerles a los nuevos miembros del grupo; Mario Serrano, Marta, Damien, Alexandre Tromas y Jose. Ustedes llegaron en el momento justo para reencender la curiosidad en mí. De Mario y Marta aprendí a no caer en el conformismo y buscar siempre trascender basándose en un grupo sólido y confiable. Gracias también a Damien por las pláticas científicas y las discusiones filosóficas dentro y fuera del laboratorio.

Estos logros también se deben a los diversos colaboradores de este proyecto. Scott Jackson y Brad Barbazuk que fomentaron en mi las ideas de evolución y conservación de distintos aspectos, esto cambió en mí la forma de ver la ciencia. Por otro lado Mario Ramírez, con quien hice una buena mancuerna entre la parte experimental, de su parte, y la parte bioinformática, de la mía, para concretar diversos proyectos. En este apartado no puedo olvidar al Dr. Federico Sánchez, quien fue un gran profesor para mí y fue el primero en confiar en mis conocimientos y enseñarme el amor por la ciencia. Quisiera prometerle que seguiré intentando transmitir mi pasión por la ciencia como él lo hizo conmigo.

Un punto fundamental ha sido mi familia y mi esposa principalmente. A Dany quisiera agradecerle el aguantarme de tantos años de estudio, alejado de ella, y los sermones motivacionales cuando los necesitaba para seguir adelante. Gracias también por mantenerse a mi lado, querer formar una familia conmigo y seguirme en mis aventuras, ya sea hacia el postdoc o hacia la vacación. De ti he aprendido muchísimas cosas que he aplicado en esta etapa de mi vida como el no darme por vencido, luchar por lo que uno desea y que siempre habrá pretextos para darse por vencidos pero queda en nosotros sobrepasarlos.

Gracias a mi mamá por ser comprensiva conmigo, por ayudarme y apoyarme en todo momento. Sé que sin tu ayuda nunca lo hubiera logrado. Gracias por darme una gran familia, ser la columna de ella, por enseñarme a adaptarme a todos los entornos y resolver todo aquello que tiene solución, que tú llamas problema. Papá mil gracias por impulsarme a trabajar con algo que aunque tenía poco experiencia tenía gran potencial. Gracias también por inculcar en mí tu "orden? de las cosas y por la perseverancia para no dejar inconcluso un proyecto. Gracias también a mi hermano por siempre cubrirme las espaldas y por las pláticas eternas y necias sobre temas que los demás consideran irrelevantes. Agus: tú me has plantado muchas más dudas que posibles respuestas y no me deja de sorprender tu valentía.

Mi familia va más allá de mis padres y hermano y no quisiera dejarlos afuera de estos agradecimientos ya que gracias a ellos hice esto posible. En especial quisiera agradecer a mi abuela adoptiva, mi tía Alicia, ya que ella no me dejó rendirme y siempre me impulsó a terminar esta etapa. Tía siempre has estado a mi lado y hemos compartido nuestra pasión por la ciencia. Me has enseñado a defenderme y que la familia es lo más importante. Gracias a toda mi familia que hacen que cada semana santa o navidad sea un motivo de alegría y diversión.

No quisiera dejar de agradecerles a mis amigos de toda la vida; Alain, Raúl, Alex, Elías y Marco. Gracias por aguantarme, mantenerme como su amigo y siempre lograr sacar de mí las risas, incluso en los momentos más tristes. Gracias también a mis compañeras y amigas del alma; Heid, Karen, Majo y Mafe porque con su cariño y su linda forma de ser me han apoyado en todo momento.

Gracias también a todo el equipo de los Bioperros; Checho, Caborca, Braulio, Chuy, Rafa y Ricardo, por motivarme a la competencia y generar una distracción fuera de los laboratorios. Ahora más que nunca vamos a la ?palapa? por una y ya, para festejar.

A Mitzy, Victor, Jazmin, Querétaro, Agus y Mario quisiera agradecerles por las pláticas exhaustivas sobre cualquier tema. Es increíble poder platicar con ustedes sobre lo que sea y discutirlo a fondo. Nunca olvidaré las maratónicas noches de dominó, papelitos o UNO. Orlando, gracias por llegar a ser parte del grupo, por adaptarte a lo que ya era la quinta y por compartir tantos gustos conmigo. Chiapas, sabes que te considero un gran amigo y te agradezco por todos esos momentos reflexivos donde tu forma de pensar cambiaba la perspectiva del tema, sólo tengo 2 palabras que decirte. Lauri y Daniela, muchas gracias por ser mis compañeras no sólo en la licenciatura, también en el doctorado. Gracias por todas esas tardes de Iker con pláticas que se extendían hasta la madrugada. Gracias porque juntos formamos un muy buen equipo y nos logramos comprender juntos. Estoy seguro de que esto no hubiera sido posible sin que estuvieran a mi lado.

Por último quisiera agradecerles a los demás que han hecho esto posible. Sra. Carolina; muchísimas gracias por permitirme hacer de su casa mi hogar los últimos 10 años. A Ian y Val por ser mis coaches en el crossfit y además de darme un hobby, enseñarme a no rendirme, a todos los que me faltaron y al Centro de Ciencias Genómicas y a la UNAM por darme la formación necesaria tanto en la licenciatura como en el doctorado para, espero, trascender en la vida y en la ciencia.

México, Pumas, Universidad
Goya, goya, cachun cachun ra ra, cachun cahun ra ra, goya, UNIVERSIDAD.

Abstract

The adequate transcription of the genes leads an organism to adapt to its environment and to develop itself in an appropriate way. This process is known as gene expression. Gene expression is regulated by different factors at different levels such as transcriptional regulation, which depends, for example, on the binding sites of transcription factors in the promoter regions of the genes. Another example may be post-transcriptional regulation that affects gene expression at the RNA level. The post-transcriptional regulation considers the interaction of other molecules independent of the messenger RNA as miRNA's to inhibit its translation or the proper messenger RNA can undergo several changes in its structure to generate multiple proteins. The latter process is known as Alternative Splicing. The different regulatory mechanisms orchestrate a series of processes that direct the cell to establish different patterns of gene expression dependent on the cellular environment. In this work different aspects of the regulation of gene expression, especially Alternative Splicing, are addressed, all this was studied using the legume *Phaseolus vulgaris*, better known as common bean, as a model specie.

Resumen

La adecuada transcripción de los genes conlleva a un organismo a adaptarse a su entorno y desarrollarse de forma apropiada. A este proceso se le conoce como expresión génica. La expresión génica está regulada por distintos factores a distintos niveles como puede ser una regulación transcripcional, que depende por ejemplo de los sitios de unión de factores de transcripción en las regiones promotoras de los genes. Otro ejemplo puede ser la regulación post-transcripcional que afecta la expresión génica a nivel de RNA. La regulación post-transcripcional considera la interacción de otras moléculas independientes del RNA mensajero como miRNA's para inhibir su traducción o el mensajero de RNA puede sufrir diversos cambios en su estructura para generar múltiples proteínas. A este último proceso se le conoce como Splicing Alternativo. Los diferentes mecanismos de regulación orquestan una serie de procesos que dirigen a la célula a establecer diferentes patrones de expresión génica dependientes del entorno celular. En este trabajo se abordan distintos aspectos de la regulación de la expresión génica en especial el Splicing Alternativo, todo esto se estudió utilizando a la leguminosa *Phaseolus vulgaris*, mejor conocida como frijol, como modelo.

Índice

1. Introducción

Phaseolus vulgaris, o frijol, es una especie sumamente importante en la dieta humana en países en vías de desarrollo debido a su alto contenido de proteínas y micronutrientes (Broughton et al., 2003). Dos independientes domesticaciones, una en Mesoamérica y otra en los Andes, han permitido el cultivo y cosecha de este alimento (Kwak et al., 2012). En México el frijol es el segundo producto más sembrado, después del maíz. *P. vulgaris* pertenece a la familia de las leguminosas y el estudio científico de esta familia ha sido considerablemente alto ya que cuenta con especies como la soya, la alfalfa, el cacahuate, entre otras, que son de suma importancia en la agricultura e industria.

A parte de su importancia comercial las leguminosas se ha estudiado de forma particular ya que éstas tienen la capacidad de establecer una relación simbiótica con bacterias fijadoras de nitrógeno (rhizobia) (Ferguson et al., 2010), proceso conocido como fijación simbiótica de nitrógeno (FSN). El nitrógeno es un elemento fundamental para la vida ya que es un elemento esencial de diversas moléculas, como aminoácidos o ácidos nucleicos, y su disposición en los suelos es escasa. Debido a esta relación las leguminosas pueden crecer en condiciones de estrés de nitrógeno. La simbiosis entre las leguminosas y las rhizobia se da en la parte subterránea de la planta y empieza con un diálogo molecular entre la planta y las bacterias. Una vez establecida la comunicación entre ambos organismos la planta forma un órgano especializado para establecer la simbiosis. Este órgano se llama nódulo y es ahí donde las bacterias se transforman en bacteroides y fijan nitrógeno atmosférico en forma de nitratos que son procesados por la planta y a cambio reciben productos fotosintéticos. A este proceso se le conoce como fijación de nitrógeno simbiótica (SNF, por sus siglas en inglés)(Ferguson et al., 2010).

P. vulgaris es un organismo diploide que cuenta con 11 cromosomas y aproximadamente 28,000 genes. La secuenciación masiva del ácido desoxirribonucleico (DNA) ha permitido obtener la secuencia genómica de dos cultivares de frijol, uno proveniente de la domesticación mesoamericana (Vlasova et al., 2016) y otro de la andina (Schmutz et al., 2014). La secuencia genómica nos permite conocer ciertos aspectos de la especie en estudio, como su

evolución, número de genes, vías metabólicas, identificar especies relacionadas evolutivamente, entre otras. El genoma de *P. vulgaris* ha permitido la realización de diversos estudios genómicos como la re-secuenciación del genoma de diversas accesiones de frijol que ha llevado a entender mejor el proceso de domesticación y evolución de esta especie (Rendon-Anaya et al., 2017). Otro de los estudios realizados a partir de la secuenciación masiva del DNA fue identificar las regiones deletadas de mutantes de *P. vulgaris* generadas a partir de la técnica de Fast Neutron (Li and Zhang, 2002),la cual consiste en radiar las células con neutrones de tal forma que se produzcan deleciones de doble cadena en el DNA. Estas mutantes fueron seleccionadas por un determinado fenotipo y a partir de la secuenciación se identificaron posibles blancos de estudio (O'Rourke et al., 2013).

El estudio del DNA no es suficiente para entender la biología de un organismo debido a que la información de los genes debe de ser transcrita para generar moléculas funcionales. La transcripción génica es el proceso por el cual la secuencia nucleotídica de un gene es copiada a una molécula de ácido ribonucleico (RNA) y esto constituye el primer paso en la expresión génica. La mayoría de los transcritos que son traducidos a proteínas son sintetizados por la RNA polimerasa II. Para iniciar la transcripción la polimerasa requiere de factores de transcripción, proteínas que reconocen secuencias de DNA (motivos) en la región río arriba de los genes. Los factores de transcripción pueden ser activadores o represores transcripcionales. La presencia, ausencia y diferencias de concentraciones de los factores de transcripción orquestan distintos perfiles de transcripción. Los perfiles de transcripción varían por tiempo, tejido, estadio de desarrollo y condiciones de estrés lo que conlleva al organismo a desarrollarse y adaptarse. El estudio de la transcripción ha permitido grandes avances en la biología, el desarrollo actual de la genómica permite conocer la totalidad de los transcritos producidos en una célula/tejido en cierta condición, a lo que se le ha denominado transcriptómica. Esta ciencia hace posible el estudio de diversos fenómenos biológicos de manera integral (Velculescu et al., 1997).

Existen diversas técnicas para estudiar la transcriptómica, uno de los primeros estudios que se utilizó fueron los microarreglos. Estos son chips que contienen secuencias de DNA y

miden la expresión génica a partir de la hibridación de cadenas sintetizadas de determinada muestra. Estas cadenas sintetizadas son marcadas con fluoróforos para que éstos sean liberados al hibridar con la cadena complementaria. Esta metodología fue de gran utilidad antes de la secuenciación masiva, pero cuenta con el problema de necesitar la secuencia nucleotídica a *priori*, es decir que solo se puede medir la expresión génica de secuencias conocidas. Sin embargo las técnicas de secuenciación masiva han permitido identificar transcritos nuevos y medir la expresión de éstas (RNA-seq)(Wilhelm and Landry, 2009).

La secuenciación del genoma del frijol en conjunto con RNA-seq han permitido estudiar diferentes aspectos de la transcriptómica de frijol. Recientemente se realizó un estudio sobre el atlas de expresión génica en frijol. En este trabajo se analizaron 24 muestras de RNA-seq de distintos tejidos de la planta en diferentes estadios de desarrollo, como raíz, nódulo, tallo, hoja, flor, vaina o semilla y además se tomaron muestras de diferentes condiciones nutricionales; plantas tratadas con fertilizante nitrogenado, plantas que fueron inoculadas con rhizobias fijadoras de nitrógeno y plantas que fueron inoculadas con una cepa de rhizobia que forma nódulos, pero no fija nitrógeno (*Rhizobium giardini*). Los diferentes patrones de expresión en esta leguminosa arrojaron resultados importantes para lograr, en un futuro, entender procesos de diferenciación, desarrollo y nutricionales (O'Rourke et al., 2014).

Los genes en organismos eucariotes están interrumpidos por secuencias no codificantes, llamadas intrones. Los intrones tiene que ser removidos del transcrito inicial (pre-mRNA) de forma que las secuencias codificantes, exones, formen una cadena madura (mRNA) que después será traducida en los ribosomas (Reddy, 2007). El origen de los intrones es incierto, aunque se cree que el ancestro común de los eucariotes haya tenido una densidad similar de intrones a los organismos actuales. A lo largo de la evolución han habido grandes pérdidas y ganancias de exones. Se cree que a corto plazo la pérdida de intrones se ve favorecida sin embargo en las mayores transiciones evolutivas se observa una gran ganancia de intrones (Rogozin et al., 2012). La pérdida de intrones puede suceder mediante una transcripción reversa del mRNA, es decir que el RNA sea leído por una transcriptasa reversa que codifica la información en DNA, y su producto sea insertado en el genoma. En cambio, la ganancia de

intrones sucede por una intronización o mediante elementos móviles del DNA. La intronización se da cuando se genera un codón de paro, donde no debería, y para evitar que se sintetice una proteína trunca se forma un intrón de forma que el mRNA no contenga el codón de paro (Catania and Lynch, 2008). Los elementos móviles del DNA que dan origen a los intrones son comúnmente llamados "introners", estos tienen características de elementos transponibles que se insertan en diversos sitios del genoma, tienen una rápida evolución y pierden las capacidades de transponerse (van der Burgt et al., 2012; Worden et al., 2009).

Por qué se mantienen los intrones es una pregunta que aún no se contesta, sin embargo, existen diversas teorías que explicarían la conservación de estos. Una de ellas tiene que ver con recombinación, ya que al tener intrones los genes se vuelven más grandes y hay una mayor posibilidad de recombinar exones (Patthy, 1999) lo cual generaría una mayor variabilidad genética. Por otro lado, se ha visto que los intrones pueden regular la expresión génica al alargar la secuencia génica en el DNA. Se ha visto que genes que están expresados altamente tienden a tener intrones pequeños mientras que genes con una expresión menor presentan intrones más largos (Castillo-Davis et al., 2002). Mediante la genética de poblaciones se explica la conservación de los intrones, ya que los organismos multicelulares eucariotes no llegan a tener una población efectiva grande y por lo tanto no existe una selección negativa hacia los intrones (Irimia and Roy, 2014). Otra explicación es que la presencia de intrones permite llevar a cabo splicing alternativo, lo cual amplía enormemente la variabilidad de mRNA y en consecuencia la aumenta la diversidad proteica.

Existen cuatro tipos diferentes de intrones en organismos eucariotes, el tipo I y tipo II no necesitan de una maquinaria y sufren auto-splicing, los intrones de RNA de transferencia y los intrones que requieren un complejo proteico llamado spliceosoma. Los intrones tipo I tienen una estructura terciaria específica y necesitan, comúnmente, una guanina libre que ataque el fosfato del sitio 5' del intrón para cortar la cadena de RNA. Luego el exón libera el intrón en su región 3' mediante un ataque nucleofílico y se une al otro exón. Los intrones tipo II tienen una estructura terciaria que permite trans-esterificación entre una adenina interna del intrón (branching point) y el sitio 5' del intrón que corta el RNA y forma una estructura de lazo.

Luego al igual que los intrones tipo I el exón 5' identifica el sitio 3' del intrón y mediante una esterificación libera el lazo y se da el splicing. La reacción de los intrones tipo II es la misma que sucede en los intrones que requieren spliceosoma, sin embargo, el spliceosoma tiene ribonucleoproteínas nucleares pequeñas (snRNP's, por sus siglas en inglés) que identifican los sitios 5', branching point y 3' de los intrones. Estos sitios están conservados en la mayoría de los intrones spliceosomales: donador o sitio de splicing 5'(GU) y aceptor o sitio de splicing 3' (AG). El otro tipo de intrones son de RNA de transferencia (tRNA) y estos dependen de endonucleasas, que corten el mRNA, y ligasas, que unen los exones (Irimia and Roy, 2014).

En organismos eucariotes la mayoría de los intrones son dependientes del spliceosoma aunque existe una gran diferencia entre el tamaño de estos entre animales y plantas. Los animales presentan intrones mucho más grandes y por lo tanto el splicing se da a partir del reconocimiento de exones, es decir la maquinaria reconoce el sitio 3' del intrón rio arriba y el 5' del intrón rio abajo. En cambio, en plantas la maquinaria reconoce al intrón (McGuire et al., 2008). Debido a que los sitios de splicing son motivos muy pequeños y muy comunes un gene puede presentar diversos patrones de splicing, conocidos como splicing alternativo (AS). Se piensa que el splicing alternativo evolucionó para evitar mutaciones deletéreas de forma que en un principio dos isoformas de un gene se transcribieran y se seleccionara la que no contara con la mutación en el mRNA (Rogozin et al., 2012). Sin embargo, con el transcurso de la evolución las diferentes isoformas, producto del AS, tienen funciones relevantes como la regulación de la velocidad de traducción, la regulación de las concentraciones de mRNA o el incremento de la diversidad proteica (Liu et al., 2017).

La relevancia evolutiva del AS ha sido relacionada con la complejidad de los organismo, respecto al número de diferentes tipos celulares. El número de genes en nematodos es muy similar al de humanos, sin embargo los diferentes tipos de células que forman a cada organismo varían significativamente (Hedges et al., 2004). El 98 % de los genes multi-exónicos presentan AS en humanos (Pan et al., 2008) mientras que en nematodos sólo el 20 % (Ramani et al., 2011). Chen et al. (2014) analizaron diversos organismos que varían en el número de diferentes tipos de células, como aproximación a complejidad de organismos, y encontraron

una fuerte correlación positiva entre el número de células diferentes y los niveles de AS. Es decir que los organismos con una mayor complejidad tienden a presentar niveles más altos de AS.

Se conoce que el splicing y por lo tanto el AS son regulados por diferentes factores. Las proteínas SR (proteínas ricas en serina y arginina) se unen al pre-mRNA mediante reconocimiento de motivos y promueven la identificación de sitios de splicing (Jeong, 2017). Por otro lado, están las ribonucleoproteínas nucleares (hnRNP) que son complejos formados por proteínas y otros RNA no codificantes (ncRNA) que inhiben la identificación de los sitios de splicing (Jean-Philippe et al., 2013). La expresión de estos diferentes factores establecen los patrones de splicing de los genes (Filichkin et al., 2015). Estos patrones de splicing se ha observado que cambian en condiciones de estrés, estadios de desarrollo y por tejido (Reddy et al., 2013).

Existen diversos eventos de AS que se generan por el reconocimiento de diferentes sitios de splicing. Puede ser que haya variación en el sitio 5', donador alternativo (AD, alternative donor), en el 3', aceptor alternativo (AA, alternative acceptor), o ambos pueden variar, sitios de splicing alternativo (ASS, alternative splicing sites). Otros eventos de AS es la omisión de un exón (ES, exon skipping) en comparación con el transcrito principal o la aparición de un exón dentro de un intrón (RE, retained exon). Por otro lado, los intrones pueden ser retenidos en (IR, intron retention) o generar uno nuevo dentro de un exón (NI, new intron). Los eventos de AA, AD, ES e IR son los más estudiados en los organismos ecuariotes. Los efectos de algunos eventos de AS pueden afectar la localización del mRNA al igual que la eficiencia de traducción (Reddy, 2007; Reddy et al., 2013). Una de las consecuencias más estudiadas es la generación de codones de paro debido a un cambio de fase, esto está asociado al decaimiento mediado por incoherencia (NMD, non-sense mediated decay) (Kalyna et al., 2012). Otros procesos relacionados con AS son el almacenaje de mRNA (Boothby2013), reconocimiento de blancos de miRNA (Yang et al., 2012) y la traducción de distintas proteínas a partir de un gene que se localizan en diferentes estructuras o presentan distintas funciones (Remy et al., 2013).

Los estudios a escala genómica de identificación de AS comenzaron a partir de la secuenciación de transcritos expresados (EST, expression sequence tags). En un principio se estimaba que alrededor del 20 % de los genes en humanos presentaban AS, sin embargo, a partir de la secuenciación masiva se ha descubierto que alrededor del 98 % de los genes multiexónicos en humanos presentan AS (Pan et al., 2008). En plantas se han observado resultados similares, ya que los primeros estudios arrojaban que el ˜15 % de los genes de *A. thaliana* presentaban AS (Iida et al., 2004) y en estudios recientes, utilizando RNA-seq, se estima que más del 60 % de los genes sufre distintos patrones de splicing (Marquez et al., 2012). Las diferencias entre estos dos organismos en cuanto a los porcentajes de genes afectados por el AS pudieran explicarse debido a la disparidad de estudios sobre el tema, ya que en humanos existen muchos análisis de identificación de AS y en plantas muy pocos, o debido a la complejidad antes mencionada. Sin embargo, los resultados de plantas y animales difieren en aspectos importantes. El evento más común en animales es ES mientras que en plantas es IR. Esto se correlaciona con el tamaño de los intrones y la forma de identificación de estos en cada especie (McGuire et al., 2008).

Se han realizado diversos estudios de identificación del AS en diversas plantas (Ner-Gaon et al., 2007; Zhou et al., 2011; Marquez et al., 2012; Li et al., 2014; Panahi et al., 2014; Shen et al., 2014; Thatcher et al., 2014; Vitulo et al., 2014; Xu et al., 2014; Mandadi and Scholthof, 2015), sin embargo, no se ha estudiado a detalle la conservación de eventos de AS entre especies. La conservación de eventos de AS es importante para identificar eventos que pudieran definir funciones relevantes. Un ejemplo claro es el TFIIIA que presenta un evento de ES, el exón omitido está altamente conservado en plantas y su función en generar un codón de paro para regular su expresiónHammond2009,Barbazuk2010. Hasta el momento se ha reportado un análisis de conservación de AS entre especies, donde se identifican eventos en 9 especies de angiospermas a partir de muestras de RNA-seq (Chamala et al., 2015). En este estudio se reportan eventos de AS en 8 y 10 muestras de RNA-seq de *P. vulgaris* y *G. max*, respectivamente (Chamala et al., 2015). Este es el único reporte, que conozco, donde se identifican eventos de AS a partir de RNA-seq y se analiza su conservación en leguminosas.

La identificación, caracterización y conservación de eventos de AS son temas de relevancia para entender mejor la complejidad transcipcional.

Existen otros reguladores post-transcripcionales como los microRNAs (miRNAs). Los miRNAs se han descrito como importantes reguladores negativos de la expresión génica en plantas y se les han atribuido funciones en el desarrollo, metabolismo, respuesta a estrés, defensa ante patógenos, integridad del genoma y simbiosis, entre otras cosas(Combier et al., 2006; Chiou et al., 2006; Navarro et al., 2006; Meng et al., 2010; Staiger et al., 2013). Estos miRNAs son fragmentos procesados de RNA con una longitud entre 19-23 nucleótidos (Voinnet, 2009). Los miRNAs se transcriben del genoma mediante la polimerasa II dando origen al pri-miRNA. Los pri-miRNA presentan una estructura terciaria de tallo-asa la cual es reconocida por una proteína RNAasa tipo III llamada DICER-LIKE1 (DCL1) que identifica esta estructura y forma al precursor de los miRNAs conocido como pre-miRNA en el núcleo de la célula. En plantas, a diferencia de animales, los pre-miRNA son procesados dentro del núcleo formando una estructura de doble cadena compuesta por el miRNA y su secuencia complementaria (miRNA*) mientras que en animales el procesamiento sucede en citoplasma. El duplex miRNA:miRNA* es identificado por un complejo protéico llamado RISC (RNA-Induced Silencign Complex) que facilita la identificación de los blancos del miRNA mediante el apareamiento casi perfecto de secuencias entre el miRNA y su mRNA blanco. Existen diversas formas en las que los miRNAs inhiben la expresión génica; pueden degradar al mRNA blanco, secuestrarlo o bloquear su traducción. En plantas, a diferencia de animales, es más frecuente que la inhibición suceda a partir de degradar el blanco mediante una proteína llamada ARGONAUTE (AGO1), un componente esencial de RISC (Jones-Rhoades et al., 2006).

La identificación de los miRNAs y sus respectivos blancos en *P. vulgaris* ha sido del interés científico y el primer análisis para identificarlos se realizó de manera *in silico*. En este trabajo se buscaron posibles precursores en los EST's reportados para frijol y a partir de estos se lograron identificar 24 miRNAs (Arenas-Huertero et al., 2009). Otra aproximación se realizó a partir de secuenciar pequeños RNAs de forma masiva (small RNA-seq) de hojas, raíces, plántulas y flores. Este análisis se realizó cuando aún no se contaba con el genoma

de *P. vulgaris* y por lo tanto el objetivo fue identificar cuáles miRNAs reportados en otras plantas estaban presentes en estos tejidos de frijol (Pelaez et al., 2012).

La secuenciación del genoma de *P. vulgaris* ha sido el comienzo de muchos otros análisis que han permitido analizar diferentes aspectos de su biología. El término de transcriptómica abarca distintos aspectos relacionados al RNA y es un estudio fundamental para entender cómo los organismos se adaptan a su entorno, se desarrollan y se diferencian. Por estos motivos mis estudios doctorales se han enfocado en analizar la expresión génica a partir de la regulación transcripcional y post-transcripcional y he ahondado más en el splicing alternativo de *P. vulgaris*.

2. Objetivos

- Objetivo General:

  Análizar la regulación de la expresión génica de *Phaseolus vulgaris* (frjiol) a distintos niveles como el transcripcional y el post-transcripcional, en específico sobre el AS y miRNAs.

- Objetivos Específicos:

  - Analizar los perfiles transcripcionales de *Phaseolus vulgaris*:
    - En nódulos bajo estrés oxidativo inducido por paraquat.
    - En nódulos de plantas inoculadas con una cepa modificada de *R. etli* que expresa la globina Vhb en condiciones control y bajo estrés oxidativo.

  - Analizar procesos del control transcripcional y post-transcripcional de la expresión génica de *Phaseolus vulgaris* a través de:
    - La identificación a nivel genómico de miRNAs y sus respectivos blancos.
    - La identificación de otros RNAs pequeños no codificantes conocidos como phasiRNAs.
    - El análisis de genes coexpresados con el factor de transcripción *AP2-1*, que es blanco de miRNA172c.
    - Identificar los sitios de pegado de factores de transcripción en las regiones promotoras de los genes coexpresados así como de los loci de miRNA172.
    - La anotación de la familia de factores de transcripción AGL y el análisis de su expresión en nódulos y raíces.

  - Analizar el Splicing Alternativo (AS) a través de:
    - La identificación de diferentes eventos de AS en las especies de leguminosas relacionadas; *P. vulgaris* y *G. max*.
    - La caracterización de los diferentes tipos de eventos de AS respecto a la región génica afectada por el evento, motivos de reconocimiento de splicing en regiones 5' y 3' de los intrones.
    - La identificación de eventos conservados en genes homólogos de *P. vulgaris* y *G. max*.
    - La caracterización de eventos de AS presentes en genes duplicados.

3. Perfiles Transcripcionales

La transcriptómica, como se menciona en la introducción, es el estudio de aquellas secuencias de RNA que fueron transcritas a partir del DNA. Uno de los principales objetivos de la transcriptómica es averiguar cuáles genes se están expresando en determinada condición y en que concentraciones. De esta forma al conocer el transcriptoma de dos o más condiciones contrastantes se podrían comparar las concentraciones y así determinar los genes inducidos o reprimidos en determinado entorno y se puede llegar a inferir su función.

3.1. Perfil transcripcional de nódulos en estrés oxidativo

3.1.1. Introducción

Las plantas, como organismos sésiles, tienen que contender contra los diferentes tipos de estrés de formas diferentes a otros organismos. Diferentes tipos de estrés como la sequía, salinidad, bajas temperaturas, alta luminosidad y toxicidad por metales, han demostrado que generan un estrés de tipo oxidativo en la planta. El estrés oxidativo sucede cuando hay un desequilibrio entre las reacciones de oxidación y reducción y se generan las llamadas especies reactivas de oxígeno (ROS, por sus siglas en inglés). Las ROS son moléculas como el anión superóxido ($\cdot O_2$-), el peróxido de hidrógeno (H_2O_2) o el radical hidroxilo ($\cdot OH$), entre otros, que se generan en todas las células con metabolismo aeróbico. Sin embargo, se ha visto una acumulación de ROS en organismos bajo diferentes estreses lo cual provoca un daño directo o indirecto a los ácidos nucleicos, lípidos, aminoácidos, o co-factores de proteínas y generalmente conlleva a una muerte celular.

Las ROS tienen un papel fundamental en los inicios de la relación simbiótica entre las leguminosas y *Rhizobium*, ya que son esenciales para un desarrollo óptimo del nódulo(Montiel et al., 2016). Sin embargo éstas, en exceso, pueden afectar la formación del nódulo y por lo tanto la SNF. La SNF se ve afectada ya que varias proteínas importantes para la fijación son sensibles a ROS, como nitrogenasas, ferrodoxinas, hidrogenasas y leghemoglobina(Matamoros et al., 2003). La alta tasa respiratoria de los nódulos incrementa la producción

de ROS, especialmente H_2O_2 y $\cdot O_2-$, y se necesita de un conjunto de enzimas tanto de la planta como del bacteroide para combatir las ROS sobre-acumuladas y contender contra el estrés oxidativo. Además, diferentes estreses abióticos como los ya mencionados pueden llevar a una sobre-acumulación de ROS y generan estrés oxidativo en los nódulos. Se conoce que en esas condiciones se producen cambios en el perfil transcripcional tanto del bacteroide como del nódulo, lo cual es un tema de interés en el grupo de investigación donde desarrollé mi tesis. Esto ha llevado al planteamiento y realización de proyectos orientados a identificar los cambios en la expresión génica de manera global (transcriptómica) de los nódulos de frijol en condiciones de estrés oxidativo generado por condiciones ambientales adversas. Durante mi doctorado yo participé en uno de estos proyectos que se reportó en el artículo: Ramírez M, Guillén G, Fuentes SI, **Iñiguez LP**, Aparicio-Fabre R, Zamorano-Sánchez D, Encarnación-Guevara S, Panzeri D, Castiglioni B, Cremonesi P, Strozzi F, Stella A, Girard L, Sparvoli F, Hernández G. (2013) TRANSCRIPT PROFILING OF THE COMMON BEAN NODULES UNDER OXIDATIVE STRESS. *Physiologia Plantarum*. 149: 389-407, que se muestra en el Anexo A. A continuación resumiré mi participación en este trabajo.

3.1.2. Resultados y Discusión

Para estudiar el estrés oxidativo en plantas de frijol en simbiosis con rhizobia se usó un herbicida llamado paraquat (PQ) (dicloruro de 1,1'-Dimetil-4,4'-bipiridilo o metil vilógeno) que induce la producción de ROS y genera estrés oxidativo en la planta de manera similar a estreses abióticos como sequía, salinidad o toxicidad por metales (Donahue et al., 1997). Este se aplicó a plantas inoculadas con *Rhizobium tropici* CIAT 899 a los 16 días post inoculación (dpi) y los nódulos fueron recolectados a 18 dpi, cuando su desarrollo se considera maduro y presentan una tasa alta de fijación de nitrógeno. Se extrajo el RNA tanto de nódulos bajo los efectos de PQ como su control, sin el herbicida, y se analizó el perfil transcripcional de ambas muestras.

El transcritoma de los nódulos se analizó a través de la hibridación de microarreglos. Se diseñó y preparó el microarreglo denominado "Bean Custom Array 90K". Para el diseño

de este microarreglo se utilizaron EST's reportadas para *Phaseolus vulgaris* y *Glycine max* (soya), una leguminosa mucho más estudiada y relacionada con el frijol. El chip contiene 30,150 sondas (secuencias) que incluyen un conjunto de 18,867 secuencia únicas de frijol y 11,205 de soya.

Al comparar los perfiles transcripcionales de los nódulos bajo el estrés y su control se encontraron 4280 EST's expresadas diferencialmente. De estas 2,218 EST's se reprimieron y 2,062 fueron inducidas en nódulos bajo el estrés oxidativo. Estos genes fueron analizados de diferentes maneras para entender el significado biológico de su expresión diferencial en los nódulos tratados con PQ. Una de los aspectos estudiados fueron las rutas metabólicas en las que pudieran estar participando. Se identificaron aquellos genes que participaran en las rutas metabólicas presentes en el microarreglo, de estos se identificaron los expresados diferencialmente y mediante una prueba de Fisher se identificaron aquellas vías que presentaran una significancia estadística. 5 vías metabólicas presentaron una disminución de expresión y 4 un aumento de expresión en nódulos bajo estrés oxidativo (Tabla 1).

Tabla 1: Vías metabólicas

Reprimidos	Inducidos
Fijación de carbono	Biosíntesis de isoflavonoides
Metabolismo de sulfatos	Metabolismo de metano
Metabolismo de purinas	Metabolismo de fármacos - citocromo P450
Metabolismo de almidón y sacarosa	Metabolismo de glicerolípidos
Fijación de CO_2	

Además de este análisis se realizó la identificación de ontologías génicas (GO, por sus siglas en inglés) sobre-representadas en las listas de EST's con expresión inducida o reprimida. Los resultados se muestran en la Figura 1.

Los resultados de ambos análisis muestran aspectos biológicos relevantes para entender lo que sucede en los nódulos sometidos a un estrés oxidativo. Los resultados de GO muestran de forma más general en qué están involucrados los EST's inducidos o reprimidos mientras que el análisis de vías metabólicas es un poco más específico.

A manera general los análisis reflejan que en los nódulos tratados con PQ se inducen pro-

Figura 1: GO de Procesos Biológicos en EST's diferencialmente expresados
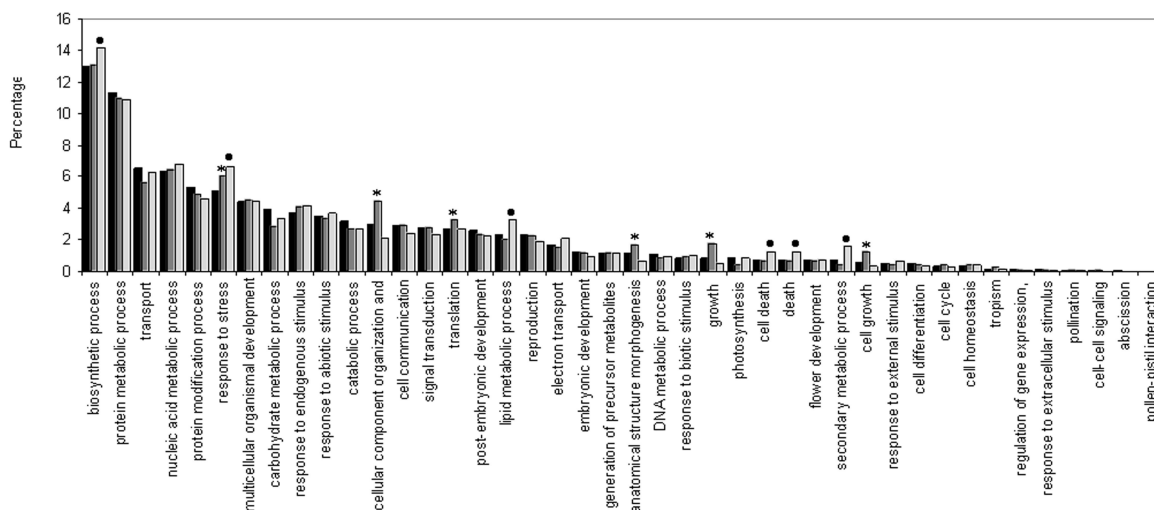
Se muestran los porcentajes de cada uno de los GO pertenecientes a procesos biológicos que se encontraron en las listas de EST's diferencialmente expresados. Las barras negras corresponden al total de cada categoría de los EST's probando en el microarreglo, las barras gris oscuras representan el porcentaje de cada proceso entre los EST's reprimidos y las barras gris claro el porcentaje entre los EST's inducidos bajo el estrés de PQ. Los asteriscos representan categorías que mostraron una sobre-representación en los EST's reprimidios mientras que los puntos negros una sobre-representación en los EST's inducidos.

cesos de muerte celular o respuestas a estrés. De forma más específica las vías metabólicas que se inducen incluyen el metabolismo de metano, de fármacos y la biosíntesis de isoflavonoides, éstas consideradas parte del metabolismo secundario. Dentro de la biosíntesis de isoflavonoides se encuentran enzimas involucradas en la biosíntesis de fitoalexinas, moléculas que tienen una función de antimicrobiana y son consideradas importantes para las defensas vegetales. De forma similar sucede con la biosíntesis de fármacos. Estos resultados muestran un estrés generalizado de los nódulos y una activación de los mecanismos de defensa vegetales.

El análisis general de las EST's que se encuentran reprimidas ante el estrés oxidativo indica que como resultado del estrés se inhiben procesos generalmente fundamentales como el crecimiento, organización de componentes celulares y la traducción. De manera más específica a través de las vías metabólicas sobre-representadas en los EST's inhibidos (Tabla 1) se observan vías esenciales para el funcionamiento del nódulo como las relacionadas con

el metabolismo de carbono (fijación de CO_2, metabolismo de almidón y sacarosa), ya que a partir de éstas la planta le brinda al bacteroide la energía (ATP's) necesarios para la fijación de nitrógeno. Además, se ven reprimidas vías del metabolismo de nitrógeno como la síntesis de purinas que interviene en la síntesis de uréidos. En frijol, el amonio aportado por el bacteroide es asimilado en las células del nódulo y ahí la planta utiliza el amonio para sintetizar uréidos. Los uréidos son los principales compuestos nitrogenados que se transportan desde el nódulo hasta las hojas como fuente de nitrógeno. La represión de estas vías afecta drásticamente la eficiencia de la fijación y asimilación del nitrógeno.

Estos resultados se discuten a más detalle en el Anexo A.

3.2. Perfil transcripcional de nódulos de frijol en simbiosis con *R. etli* expresando la hemoglobina de *Vitroescilla*, en condiciones de estrés oxidativo

3.2.1. Introducción

Aunque los rhizobia son organismos aeróbicos, la fijación SNF por los bacteroides en los nódulos de leguminosas ocurre en un ambiente microaeróbico. La nitrogenasa, complejo enzimático que cataliza la reducción de nitrógeno atmosférico a amonio, es sumamente sensible al O_2. La leghemoglobina producida por la planta en los nódulos es fundamental para proveerle oxígeno a la oxidasa terminal del bacteroide. Esta oxidasa opera de manera óptima a presiones bajas de oxígeno y genera el ATP requerido para el funcionamiento de la nitrogenasa.

En la naturaleza existen diversas proteínas con funciones de transporte y almacenamiento del oxígeno similares a las de la leghemoglobina. Una de éstas es la globina de la bacteria gram-negativa aeróbica: *Vitroescilla* sp. que se denomina Vhb. Esta proteína tiene una eficiencia más alta que sus homólogos en organismos multicelulares y se ha asociado a diversas funciones como catalizador de reacciones redox y protección celular ante estrés oxidativo (Hardison, 1998; Kaur et al., 2002; Geckil et al., 2003).

La alta afinidad de la VHb de *Vitroescilla* hacia el oxígeno ha generado un gran interés

de la comunidad y se ha utilizado para la biotecnología. Se ha estudiado el crecimiento de *Escherichia coli* que expresan esta hemoglobina y se ha visto que existe un mayor crecimiento y una mayor densidad celular en comparación con la cepa silvestre (Geckil et al., 2001; Khosla and Bailey, 1988). También se ha visto que la expresión de VHb en *E. coli* le confiere resistencia al estrés oxidativo ya que se correlaciona con una alta expresión de antioxidantes (Geckil et al., 2003; Anand et al., 2010). Estos hallazgos han permito plantear la hipótesis sobre si la expresión de Vhb en rhizobia durante la simbiosis con leguminosas (bacteroides) pudiera conferirle a la planta y al nódulo cierta resistencia al estrés oxidativo y de esta forma mejorar la SNF.

En trabajos pasados del laboratorio se estudiaron los efectos de Vhb, expresada bajo su promotor nativo, en *R. etli* tanto en vida libre como en simbiosis con *P. vulgaris* (Ramirez et al., 1999). Los resultados muestran que en condiciones bajas de oxígeno las bacterias de vida libre tienen una mayor tasa respiratoria, contenido de energía química y expresión de genes fijadores de nitrógeno comparado con *R. etli* silvestres que no expresan esta hemoglobina heteróloga. Además, las plantas inoculadas con las bacterias que VHb mostraron mayor biomasa, una temprana floración y formación de vainas, una mayor tasa de actividad de la nitrogenasa y un mayor contenido de nitrógeno en comparación con plantas inoculadas con *R. etli* silvestre (Ramirez et al., 1999).

Estos resultados fueron los antecedentes de un trabajo que se publicó en: Ramírez M, **Iñiguez LP**, Guerrero G, Sparvoli F, Hernández G. (2016) *Rhizobium etli* BACTEROIDS ENGINEERED FOR Vitreoscilla HEMOGLOBIN EXPRESSION ALLEVIATE OXIDATIVE STRESS IN COMMON BEAN NODULES THAT REPROGRAMME GLOBAL GENE EXPRESSION. *Plant Biotechnology Reports*, 10: 463-474. El objetivo general de este trabajo fue analizar el efecto de la cepa de *R.etli* expresando la VHb en simbiosis con plantas de frijol creciendo en condiciones de estrés oxidativo, generado por el herbicida PQ. Como se mencionó antes, la hipótesis era que la VHb en el bacteroide puede tener ventajas que resulten en un mejor desempeño de las plantas en SNF ante este estrés. El enfoque de este trabajo fue de corte genómico ya que se analizó de manera comparativa el transcriptoma de

los nódulos de frijol en cuatro diferentes simbiosis: frijol inoculado con *R. etli* silvestre en condición control (1) y en estrés por adición de PQ (2); *R. etli* / VHb en condición control (3) y en estrés por PQ (4). Para el análisis del transcriptoma se utilizaron los microarreglos descritos en la sección anterior. A continuación resumo este artículo que se presenta en el Anexo B.

3.2.2. Resultados y Discusión

Las plantas inoculadas con *R. etli* / Vhb mostraron una menor sensibilidad al tratamiento con PQ, ya que presentaron un menor decaimiento de la actividad de nitrogenasa y contenido de ureidos en comparación con las plantas en condiciones control. Estos resultados sugieren que VHb protege a los nódulos del estrés oxidativo generado por PQ, lo que resulta en una mejora en la eficiencia de la SNF.

Para analizar la expresión génica en los nódulos de las cuatro diferentes condiciones de simbiosis, se extrajo RNA de nódulos de 18 días y se analizaron utilizando el microarreglo descrito en la sección anterior. Se identificaron los EST's expresados diferencialmente entre las muestras (Anexo B), sin embargo, las 2 comparaciones en las que se enfocó el análisis en este trabajo fueron entre las condiciones +PQ vs control en los nódulos formados por la cepa de *R. etli* que expresa la Vhb y la cepa silvestre, es decir 1 vs 3 y 2 vs 4. Sólo 8 % de los EST's expresados diferencialmente (reprimidos o inducidos) en la condición de estrés oxidativo se compartían independientemente del inóculo. Esto indica que los patrones de expresión génica se ven afectados por la modificación de *R. etil* / VHb. Para analizar esto de manera más precisa se identificaron los GO (procesos biológicos) de los EST's diferencialmente expresados (Figura 2a). Entre las categorías sobre-representadas de los EST's inducidos en nódulos con la cepa silvestre de *R. etli* tratados con PQ se encuentran proceso involucrados en defensa e inducción a muerte celular y en los EST's reprimidos procesos derivados de la fijación de nitrógeno (Figura 2a). Estos resultados son similares a los discutidos en la sección anterior (Anexo A). A diferencia de los nódulos inoculados con la cepa silvestre, los nódulos formados por la cepa modificada que expresa la Vhb muestran una mayor cantidad de EST's

inducidos que reprimidos. Las categorías de GO que aparecen sobre-representadas entre los EST's altamente expresados en nódulos con PQ se refieren a estreses abióticos (Figura 2b). Esto muestra que la presencia de VHb está induciendo una respuesta para contender contra el estrés y por lo tanto le está confiriendo cierta resistencia a la planta.

Figura 2: GO de Procesos Biológicos en EST's diferencialmente expresados
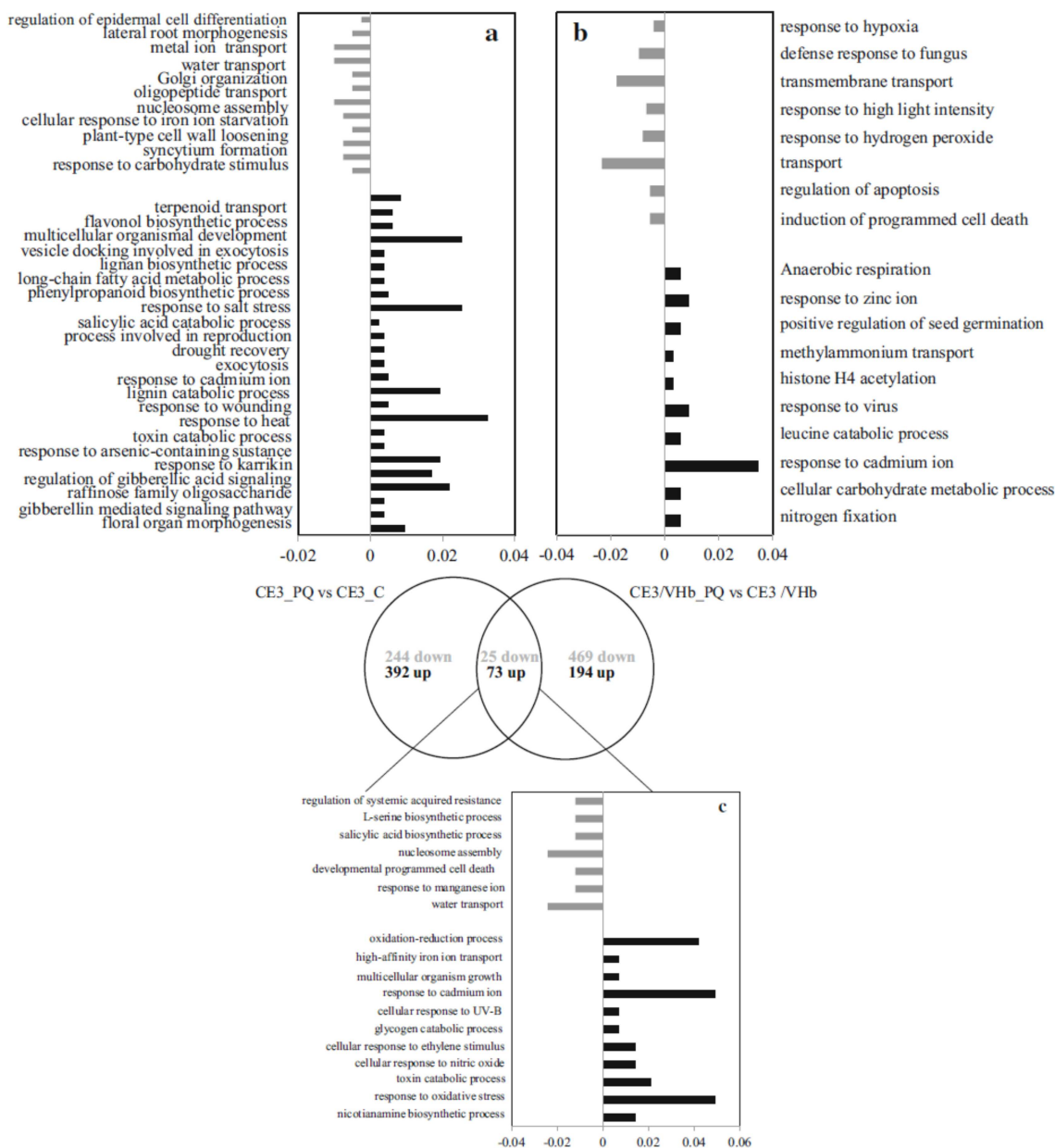
Se muestran los GO de procesos biológicos significantemente enriquecidos en el set de genes reprimidos (barras grises) y aumentados (barras negras). Procesos biológicos enriquecidos en EST's diferencialmente expresados de las comparaciones entre control vs PQ inoculados con la cepa silvestre (a) y con la cepa de *R. etli* que expresa Vhb (b) y los procesos biológicos de los EST's que presentan el mismo patrón de expresión en ambas comparaciones (c).

A pesar de las diferencias en los patrones de expresión entre ambas comparaciones, existe un grupo de EST's que mantiene la tendencia de estar inducidos o reprimidos ante la presencia de PQ. Estos EST's presentan una clara tendencia a ser respuestas comunes del estrés oxidativo. Entre los EST's inducidos en los nódulos tratados con PQ se encuentran procesos como respuesta a estrés oxidativo y procesos de oxidación-reducción entre otros (Figura 2c).

Los resultados de este trabajo muestran que la inclusión de VHb a *R. etli* beneficia a la planta no solo en condiciones control (Ramirez et al., 1999) sino que también en condiciones de estrés oxidativo, lo que sugiere explorar el uso de esta bacteria modificada como biofertilizante para fines biotecnológicos.

4. Control Transcripcional y Post-Transcripcional

El control transcripcional constituye uno de los modos más importantes de la expresión génica. Esta regulación está orquestada por factores de transcripción, regiones promotoras, interacciones proteicas que inhiben o estimulan la transcripción, entre otras. Tanto las diferentes concentraciones de factores de transcripción como la accesibilidad a las regiones promotoras y la presencia de distintas proteínas provocan distintos perfiles de transcripción, analizados y discutidos en la sección anterior. Sin embargo, para lograr entender la complejidad de los perfiles transcripcionales es fundamental entender las piezas que lo conforman, es decir entender cuáles factores de transcripción está presentes y estos a cuáles genes regulan. Los factores de transcripción están catalogados en familias, dependiendo de los motivos proteicos conservados/compartidos entre los miembros.

Además del control transcripcional existe la regulación post-transcripcional de la expresión. Este tipo de regulación ocurre posterior a la transcripción y sucede en los mRNA. La regulación post-transcripcional abarca,el splicing, el transporte de los mRNA producidos en el núcleo hacia el citoplasma y a los ribosomas, donde sería traducido, y la eficiencia de asociación entre el mRNA y los ribosomas para el inicio de la traducción. Los miRNAs, mencionados en la introducción, son parte de la regulación post-transcripcional en las plantas, estos son reguladores negativos que inhiben la asociación del mRNA con los ribosomas ya sea secuestrándolo, o degradándolo para así impedir su traducción.

4.1. Identificación de pequeños RNAs de *Phaseolus vulgaris*

4.1.1. Introducción

La regulación post-transcripcional a partir de los miRNAs descrita en la introducción no es la única función que tienen estos ya que también son encargados de la producción de otro tipo de RNAs pequeños no codificantes llamados phasiRNA (phased small interfering RNA). Ciertos transcritos largos, blancos de miRNA, reclutan polimerasas de RNA que generan una doble cadena de RNA. Una vez formada la doble cadena de RNA es procesada por DCL4 o

DCL3b que corta ambas cadenas produciendo pequeños RNAs de 21 o 24 nucleótidos. La característica más importante de estos sncRNA se da debido al procesamiento de las DCL4 o DCL3b ya que corta las cadenas a manera de fase, es decir en ventanas de 21 o 24 nucleótidos. Los phasiRNA resultantes de este proceso pueden actuar como miRNAs e inhibir de la misma forma ciertos genes a través del complejo RISC.(Jones-Rhoades et al., 2006; Voinnet, 2009)

La identificación de los miRNAs en *P. vulgaris* ha sido fundamental en los estudios que se han realizado en el laboratorio y en cuanto se hizo pública la secuencia genómica de *P. vulgaris* (Schmutz et al., 2014) realizamos un proyecto de identificación de miRNAs y sus respectivos blancos, así como la identificación de phasiRNAs. Para esto se utilizaron las muestras de small RNA-seq publicadas anteriormente (Pelaez et al., 2012) en conjunto con otras de nódulos y además secuencias de "degradoma". Estas últimas provienen de una técnica para identificar los blancos de miRNAs que son degradados por AGO1. Los miRNAs en el complejo RISC al identificar a su blanco y cortarlo mediante AGO1 dejan un fosfato libre en el extremo 5' que es único para los blancos digeridos por miRNAs/AGO1. Para secuenciar estos productos el protocolo establece que hay que seleccionar transcriptos con poly-adenilación en el 3' terminal, ligar un adaptador al 5' que contenga el fosfato libre y utilizar una enzima de restricción para seleccionar a aquellas secuencias que tengan el adaptador (German et al., 2008). En este proyecto participé específicamente en la identificación y análisis de phasiRNA, los cuales nunca habían sido descritos en *P. vulgaris*. Los resultados se publicaron en: Formey D, **Iñiguez LP**, Peláez P, Li Y-F, Sunkar R, Sánchez F, Reyes J, Hernández G. (2015) GENOME-WIDE IDENTIFICATION OF THE *Phaseolus vulgaris* sRNAome USING SMALL RNA AND DEGRADOME SEQUENCING. *BMC Genomics*, 16: 423. A continuación resumo este artículo que se presenta en el Anexo D.

4.1.2. Resultados y Discusión

En este análisis se identificaron 185 nuevos miRNAs provenientes de 307 pre-miRNAs distribuidos en 98 familias diferentes. De estos, 64 ya habían sido identificados en otras especies, 57 son nuevas isoformas y el resto (64) son específicos de frijol. Debido a la gran

cantidad de información que se obtiene de small RNA-seq se pudo asignar una expresión para cada miRNA en los diferentes tejidos analizados. Los miRNAs que ya habían sido identificados en otras especies presentaban una expresión en varios tejidos mientras que los específicos de *P. vulgaris* preferencialmente se expresaban en tejidos únicos. Estos resultados sugieren una funcionalidad ancestral y global de los miRNAs conservados y una regulación post-transcripcional especie-específica para los miRNAs nuevos.

Parte de estos análisis se enfocaron en miRNAs de nódulos. Para ello se infirieron redes de coexpresión en nódulos en las que se encontraron miRNAs conservados ya previamente identificados en procesos de simbiosis junto con miRNAs o isoformas nuevas. Esto da la pauta para caracterizar estos miRNAs no conservados e identificar su función en los nódulos.

Los datos del degradoma permitieron identificar 181 blancos de miRNAs, siendo la mayoría blancos de miRNAs conservados. Estos resultados sugieren, al igual que los de la expresión, una regulación a post-transcipcional de mRNAs expresados en diversos órganos por parte de los miRNAs conservados que los especie-específicos. Cabe resaltar que el degradoma sólo muestra aquellos blancos que están siendo degradados por miRNAs pero no se identifican a los blancos secuestrados o que su traducción está siendo bloqueada debido a la especificidad de la metodología del degradoma.

Para identificar los phasiRNAs se utilizaron los mapeos de las small RNA-seq al genoma y se calculó un score de fase (Zhai et al., 2011) para cada fragmento de 21 nucleótidos alineados al genoma. Para eliminar el ruido producido por fragmentos degradados se tomaron en cuenta las lecturas mapeadas en fase y las no en fase y se utilizó una prueba de Chi ($p < 0.01$). En total se identificaron 125 loci capaces de producir phasiRNAs, 13 de ellos en todas las muestras (Figura 3a). Para identificar si estos provenían de los miRNAs se buscaron cuáles pudieran ser blancos de los miRNAs. 47 loci de phasiRNAs se identificaron como posibles blancos para 31 miRNAs. Un aspecto interesante de los phasiRNAs es que sólo dos de ellos se encontraron conservados en otras leguminosas (Figura 3b). Esto pudiera deberse a dos situaciones, una sería la falta de estudio de este tipo de sncRNA o que los mecanismos biológicos de acción de estos sean específicos para cada especie (Apéndice D).

Figura 3: Diagramas de Venn de phasiRNA
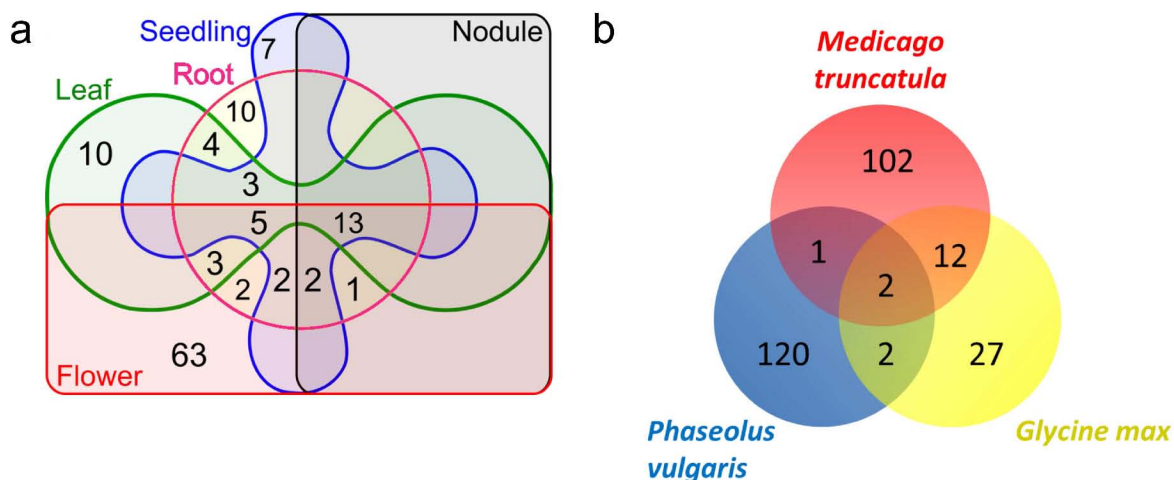
(a) Distribución de los phasiRNA genes en los 5 órganos de la planta. Los números en cada área del diagrama de Venn corresponden al número de loci de phasiRNA expresados en los distintos órganos. (b) Conservación de genes phasiRNA presentes en 3 especies de leguminosas

4.2. El nodo miRNA172c-*AP2* como regulador de la fijación de nitrógeno simbiótica en la relación *P. vulgaris-Rhizobium etli*

4.2.1. Introducción

Dado su papel fundamental como reguladores centrales, los miRNAs han sido de gran interés para el laboratorio. Además de los proyectos generales sobre identificación de miRNAs, que se mencionaron antes, se han analizado miRNAs específicos y se ha demostrado su papel en procesos relevantes para las plantas de frijol. Así se ha estudiado el papel de los miRNAs en *P. vulgaris* en condiciones como la deficiencia de fósforo (Valdes-Lopez et al., 2008; Ramirez et al., 2013), la deficiencia de nutrientes (Valdes-Lopez et al., 2010) y la toxicidad por metales (Mendoza-Soto et al., 2012; Naya et al., 2014; Mendoza-Soto et al., 2015). Uno de estos proyectos del grupo, en los que participé, fue sobre el nodo del miRNA172 y su blanco *AP2*. Los resultados de este proyecto se publicaron en: Nova-Franco B, **Iñiguez LP**, Valdés-López O, Alvarado-Affantranger X, Leija A, Fuentes SI, Ramírez M, Paul S, Reyes JL, Girard L, Hernández G (2015) THE microRNA-172c-APETALA2-1 NODE AS A KEY REGULATOR OF THE COMMON BEAN RHIZOBIA NITROGEN FIXATION SYMBIOSIS. *Plant Physiology*, 168: 273-290. Yo colaboré con análisis de tipo transcriptómicos relacionados a

este nodo y su participación en la simbiosis de frijol con *Rhizobium etli*.

El miR172 se identificó como promotor de la floración en *A. thaliana* (Aukerman and Sakai, 2003), ya que inhibe la traducción de un inhibidor de la floración llamado *APETALA2*. Sin embrago en los trabajos de identificación de miRNAs en *P. vulgaris* se detectó una isoforma del miRNA172 inducida en nódulos de frijol (Pelaez et al., 2012; Formey et al., 2015), lo que permitió plantear la hipótesis de que este miRNA y su gene blanco tenían alguna función biológica relevante en la simbiosis con rhizobia.

4.2.2. Resultados y Discusión

En este trabajo se identificaron 6 loci en el genoma que codifican para isoformas del miR-NA172, sin embargo, sólo tres diferentes isoformas (a-c) del miRNA maduro son detectables, ya que la isoforma a es idéntica a la e y f, y la isoforma c es igual a la d. Se identificó que el miRNA172c es el más expresado en los nódulos mientras que la isoforma miRNA172a está expresada mayoritariamente en flores. Otro aspecto importante en los estudios de miRNA es la identificación de su respectivo blanco. Arenas-Huertero et al. (2009) ya habían identificado y validado experimentalmente al factor de transcripción *AP2-1*, que muestra una expresión opuesta (correlación negativa) al nivel del miRNA172 en diferentes órganos y etapas de la simbiosis con rhizobia. Los resultados indican que la expresión del miRNA172c se da en los nódulos y regula post-transcripcionalmente a *AP2-1*. En ese trabajo se observó que el miR172c maduro se incrementa ante la infección con rhizobia y continúa incrementando hasta alcanzar su máximo nivel en nódulos maduros y disminuye en nódulos senescentes. La expresión del miRNA y por lo tanto su función como regulador post-transcripcional es dependiente del estadio de desarrollo del nódulo.

La caracterización funcional del miRNA172c continuó al analizar el fenotipo simbiótico de mutantes transitorias de frijol. Las mutantes transitorias, o plantas compuestas, se logran a partir transformar raíces y nódulos mediante *Agrobacterium rhizogenes* (Estrada-Navarrete et al., 2007) que sobre-expresan el miRNA172c o el *AP2-1* mutado en el sitio que reconoce el miRNA, esto le confiere insensibilidad a la presencia del miRNA172c. Las raíces sobre-

expresantes del miRNA172c muestran un incremento en su crecimiento, una mayor infección rhizobial, una mejor nodulación, una mayor tasa de fijación de nitrógeno y un aumento en la expresión de genes relacionados con los primeros pasos de la nodulación y la autoregulación de la misma. La sobre-expresión de *AP2-1* causó los efectos contrarios.

Para entender mejor los efectos de esta regulación se consideró que el factor de transcripción *AP2-1* tuviera un rol de activador transcripcional y se buscaron, en el atlas de expresión de *P. vulgaris*, los genes que tuvieran el mismo patrón de expresión que *AP2-1* (O'Rourke et al., 2014) en las bibliotecas de nódulos y raíz. Se identificaron 114 genes que presentaban un patrón similar al del blanco de miRNA172c (Figura 4). Se analizaron las regiones promotoras de estos genes para identificar los motivos de pegado de los factores de transcripción y se encontró una sobre-representación de sitios ERF2 y DREB1B, ambos pertenecientes a la familia de AP2. Estos resultados sugieren que los genes que están coexpresados con *AP2-1* pueden ser regulados de manera positiva por este factor de transcripción. Para confirmar esta hipótesis se verificó la expresión de algunos de los genes coexpresados en las raíces transgénicas que sobre-expresan *AP2-1* y los resultados muestran un incremento en su expresión. También se analizaron las categorías de GO sobre-representadas en los genes coexpresados y se encontró una categoría de proteínas cinasas que se han relacionado con la senescencia de los nódulos en *Medicago truncatula* (Van de Velde et al., 2006; Perez Guerra et al., 2010). En general, los resultados de este trabajo indican que la regulación post-transcripcional del AP2-1 por el miRNA172c es relevante para etapas iniciales de la simbiosis con rhizobia así como para la senescencia de los nódulos, por lo que el miRNA172c debe regular negativamente la expresión de *AP2-1* en los nódulos maduros. El artículo publicado en el Anexo C presenta todos los resultados de este trabajo, así como una discusión más amplia de los mismos.

Otro aspecto importante dentro de la regulación post-transcripcional del *AP2-1* es conocer cómo se regula el miRNA172c a nivel transcripcional. En la compleja red de regulación de la floración en *A. thaliana* la transcripción del miRNA se activa por un factor de transcripción llamado SPL (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE). Para identificar si esto mismo sucede en frijol se buscaron los sitios de unión de factores de transcripción en las

Figura 4: Perfil de Expresión de Genes Coexpresados con *AP2-1*
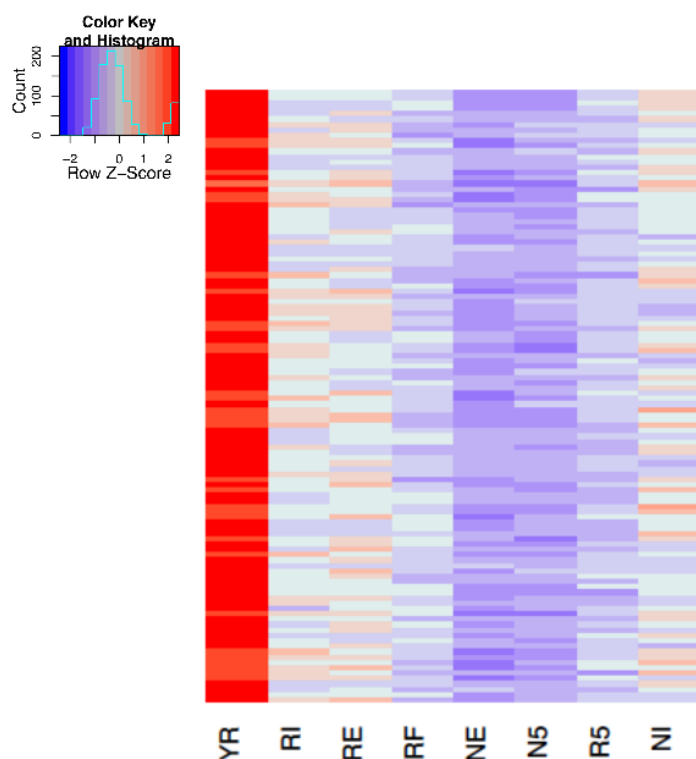
Se muestra el perfil de expresión de los 144 genes coexpresados con *AP2-1* (Z-scores: rojo=mayor expresión, blanco= menor expresión). Las muestras que aparecen en el mapa de calor son: YR=raíz joven, RI=raíz inoculada con nódulos que no fijan nitrógeno a 21 dpi, RE=raíz con nódulos fijadores de nitrógeno a 21 dpi, RF= raíz fertilizada, NE= nódulos fijadores de nitrógeno (21 dpi), N5= nódulos pre-fijadores (5 dpi), NI= nódulos no fijadores (21dpi)

regiones promotoras de las seis isoformas de miRNA172. Los resultados no muestran motivos para SPL, sin embargo se encontró una sobre-representación de motivos de AGL. Resultados similares se obtuvieron en genes altamente expresados en nódulos maduros (O'Rourke et al., 2014). Esto sugiere que los factores de transcripción de la familia AGAMOUS-LIKE PROTEIN (AGL) pueden estar involucrados en la regulación de la nodulación, no sólo a través de la expresión del miRNA172c sino que de otros genes importantes para este proceso.

El papel de ciertos factores de transcripción de la familia AGL se ha estudiado a partir de la regulación de la organogénesis de la flor (Jofuku et al., 1994). Algunos miembros de esta familia son indispensables para este proceso, el cual se regula por una compleja red en la que interaccionan los AGL con otros factores de transcripción, como los AP2 y con miRNAs como el miRNA172 (Azpeitia et al., 2014; Causier et al., 2010).

Por los resultados ya comentados en el Anexo C decidimos continuar el análisis de la familia AGL en frijol. Este trabajo fue publicado en el artículo: **Iñiguez LP**, Nova-Franco B, Hernández G (2015) NOVEL PLAYERS IN THE AP2-miR172 REGULATORY NETWORK FOR COMMON BEAN NODULATION. *Plant Signaling & Behavior*, 10 (10), e1062957, que se muestra a continuación. Encontramos que en el genoma de *P. vulgaris* hay 91 genes anotados como AGLs. Al analizar sus patrones de expresión en diferentes órganos encontramos que 21 AGLs muestran expresión en nódulos o raíces de frijol (O'Rourke et al., 2014) y 8 de ellos muestran una expresión diferencial entre las raíces o nódulos en comparación con otros tejidos. Los AGLs que mostraron una expresión en los tejidos subterráneos fueron anotados con mayor detenimiento utilizando a *A. thaliana* como referencia y buscando los genes ortólgos de soya.

Novel players in the AP2-miR172 regulatory network for common bean nodulation

Luis P Íñiguez, Bárbara Nova-Franco, and Georgina Hernández*

Centro de Ciencias Genómicas; Universidad Nacional Autónoma de México; Cuernavaca, Mexico

The intricate regulatory network for floral organogenesis in plants that includes AP2/ERF, SPL and AGL transcription factors, miR172 and miR156 along with other components is well documented, though its complexity and size keep increasing. The miR172/AP2 node was recently proposed as essential regulator in the legume-rhizobia nitrogen-fixing symbiosis. Research from our group contributed to demonstrate the control of common bean (*Phaseolus vulgaris*) nodulation by miR172c/AP2-1, however no other components of such regulatory network have been reported. Here we propose AGLs as new protagonists in the regulation of common bean nodulation and discuss the relevance of future deeper analysis of the complex AP2 regulatory network for nodule organogenesis in legumes.

The transcription factor (TF) superfamily APETALA2/Ethylene Responsive Factor (AP2/ERF) is conserved among the plant kingdom. Members of this large TF family, that regulate plant development and response to stress, have the characteristic AP2/ERF domain (IPR001471, www.ebi.ac.uk/interpro/) that consists of 60-70 amino acids and is involved in DNA binding.[1,2] Since 1994, Jofuku et al.[3] reported the relevant role of AP2 in the control of flower development in Arabidopsis. The *AP2* gene is part of a complex regulatory network that includes the microRNA172 (miR172) which promotes flowering by repressing *AP2* through translation inhibition or mRNA cleavage.[4-6] Transcription of *MIR172* genes is activated by SPL (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE) TF, that are targets miR156.[7] In addition, both *MIR156* and *MIR172* promoters have binding sites for AP2 that acts as a dual transcriptional regulator activating *MIR156* and repressing *MIR172*, thus constituting a negative feedback loop in the complex and robust floral organogenesis networks.[8]

The key regulatory role of miR172/AP2 in legume nodulation during the rhizobia nitrogen-fixation symbiosis has been reported for soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*).[9-11] We have recently demonstrated that in common bean miR172c and its target gene *AP2-1* (Phvul.005G138300) are essential regulators for rhizobial infection and nodulation through the regulation of the expression of early nodulation and autoregulation-of-nodulation (AON) genes. In addition we proposed that AP2-1 is a positive regulator of nodule senescence genes and thus it is silenced, through miR172c-induced target cleavage, in mature nodules.[11] Though these studies,[9-11] suggest that miR172 and their corresponding AP2 targets are part of complex regulatory networks in legume nodules, it is essential to decipher other yet unknown network components. We have analyzed SPLs as possible transcription regulators of common bean *MIR172* genes, but only negative results were obtained.[11] In this work we propose other regulators involved in miR172/AP2-1 regulatory network for common bean nodulation and discuss the importance for their deeper future analysis

We used CLOVER,[12] to analyze the promoter sequence of each of the 6 *P. vulgaris MIR172* genes,[11] in order to identify additional transcriptional regulators. We found that AGL (AGAMOUS-LIKE PROTEIN) TF binding sites were statistically overrepresented in these promoters.

In addition, a similar analysis was performed to the promoters of genes previously identified as highly expressed in *P. vulgaris* mature nodules, these genes co-express with those from the purine and ureide biosynthetic pathways.[13] Ureides are the main product from fixed-nitrogen assimilation in warm season legumes and the main nitrogen supply for plant nutrition. Notably, among other TF binding sites, AGL sites were statistically over-represented in the promoters of common bean nodule-enhanced genes.

The AGL TF family, also known as MADS (MINICHROMOSOME MAINTENANCE1, AGAMOUS, DEFICIENS and SERUM RESPONSE FACTOR) in other eukaryotes, has been widely studied for the regulation of flower organogenesis.[14] Some members of this family, such as AP1, AP3, SOC, AG and FUL, are indispensable for the correct floral formation and may interact with members from the AP2/ERF superfamily, such as AP2 or TOE.[15,16] AGLs, such as AGL15 and SOC1, have been included as part of the complex AP2 network for floral transition and floral development in Arabidopsis.[8]

On this basis, we consider interesting to study AGLs from common bean as a new players of nodule organogenesis networks, something that, to our knowledge, has not been yet analyzed for legumes. Ninety-one genes were annotated as *AGL* or *MADS* TF genes in the *P. vulgaris* genome (www.phytozome.net/commonbean.php, v1.0).[17] According to the *P. vulgaris* Gene Expression Atlas (*Pv* GEA),[13] only 21 (out of 91) *AGL* genes were expressed in at least one root or nodule tissue sample; these are shown in **Figure. 1**. According to the common bean expression analysis among different tissues reported by O'Rourke et al.,[13] 8 of the *AGL* genes shown in **Figure. 1** were significantly up regulated in nodules and/or roots as compared to aerial tissues

This subset of 21 common bean *AGL* genes were carefully annotated based in their sequence similarity to Arabidopsis *AGL* genes, following the annotation from The Arabidopsis Information Resource (TAIR, www.arabidopsis.org) as shown in **Table 1**. In addition, **Table 1** includes information from TAIR about the expression in roots tissues of the Arabidopsis
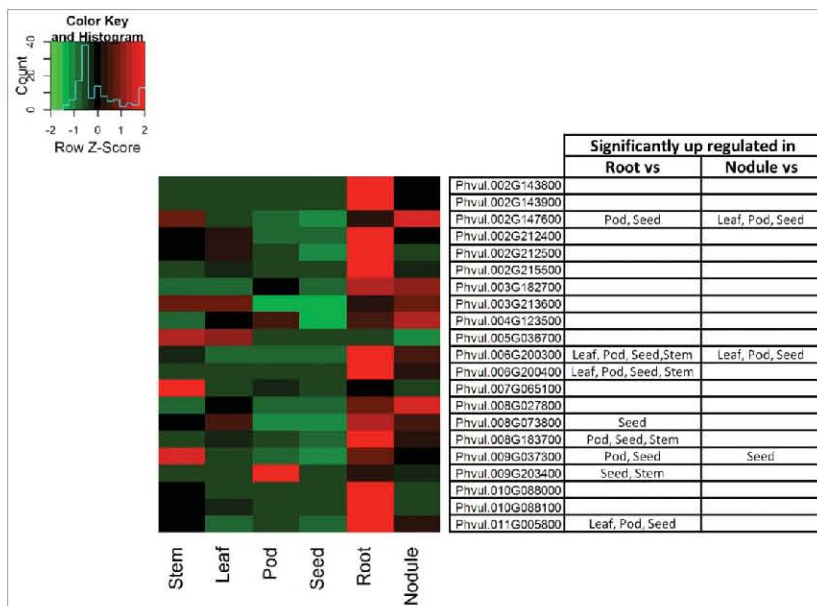
Figure 1. Heatmap of the *P. vulgaris* AGL genes that are expressed in root/nodule tissues. Values of expression were extracted from the *Pv* GEA, tissues were analyzed as reported in O'Rourke et al. [13]. Gene IDs are from www.phytozome.net/commonbean.php. Left: Expression patterns as Z-score. The color scale represents the expression distribution, red indicates high expression, and black indicates the mean of expression and green indicates low expression. Right: Differential expression (up regulation) in root and/or nodule as compared to other tissues, based in data from the Pv GEA.[13]

AGL genes that are similar to the root/nodule common bean *AGL* genes. Notably, most of these genes showed expression in one or more root-tissue samples from Arabidopsis (Table1), something that correlates with their tissue expression in common bean (**Fig. 1**). In order to gain insight into the potential roles of AGL TF in the control of the symbiotic interaction between legumes and nitrogen fixing bacteria we searched for the orthologous genes, to those analyzed from common bean, in soybean (*Glycine max*), a legume closely related to *P. vulgaris*.[17] As shown in Table 1, we identified 17 soybean genes orthologous to common bean AGL genes expressed in root or nodule. The data from the RNA-seq Atlas of *G. max*.[18] revealed that 13 ortholog AGL genes expressed highest in underground tissues (**Table 1**). Thus, the data on soybean gene expression, shown in **Table 1**, support the hypothesis of *AGL* genes as regulators of the legume /rhizobia symbiosis

AGL21 is expressed in several different Arabidopsis root tissues (Table1). This gene is crucial for lateral root

development; its expression is affected by phytohormones and in response to stresses. Arabidopsis plants overexpressing *AGL21* produce more and longer lateral roots. AGL21 has been implicated in auxin accumulation in lateral root primordia.[19] The common bean *AGL21* (Phvul.008G183700) is upregulated in roots vs pods, seeds and stems and it is also expressed in nodules (Fig.1). The soybean *AGL21* (Glyma.14G183800) shows its highest expression in roots (**Table 1**). It is known that the auxin/cytokinin ratio is strictly controlled and plays an important role in nodule development during the legume-rhizobia symbiosis. Auxins plays (at least) a dual role during nodulation: in the early stages auxin transport inhibition might result in a reduced auxin/cytokinin ratio to allow cell division to start and later divisions are inhibited by super optimal auxin levels.[20] Thus, it is conceivable that the root/nodule expression of common bean AGL21 is related to auxin regulation of the symbiotic process

XAL1 and *XAL2* are other AGL's important for Arabidopsis root

Table 1. Annotation of *P. vulgaris* root/nodule AGL genes and their expression in *A. thaliana* and *G. max* (soybean) tissues. The *P. vulgaris* gene annotation was based in that for Arabidopsis (TAIR, www.arabidopsis.org). The expression (TAIR) in one or more root tissue samples of the Arabidopsis orthologues of each AGL common bean gene, is indicated. The soybean *AGL* genes identified as orthologues of the common bean genes are indicated as well as the plant tissue where each gene presents its highest expression value, according to data from the RNA-seq Atlas of *G. max*.[18]

Phaseolus vulgaris	*Arabidopsis thaliana*		*Glycine max*	
Gene	Gene	Expression in root tissues	Gene	Highest expression in
Phvul.002G143800	AGL17	root epidermis, root, root cap, root elongation zone	Glyma.09G201700	root
Phvul.002G143900	ANR1	root	—	—
Phvul.002G147600	SVP	root	Glyma.02G041500	root
Phvul.002G212400	TT16	—	—	—
Phvul.002G212500	SVP	root	Glyma.08G068200	young leaf
Phvul.002G215500	ANR1	root	Glyma.08G065400	root
Phvul.003G182700	FUL	—	Glyma.05G018800	root
Phvul.003G213600	AGL19	root, lateral root cap	Glyma.17G132700	root
Phvul.004G123500	AGL13	root	Glyma.16G200700	nodule
Phvul.005G036700	AGL26	root	Glyma.08G152700	young leaf
Phvul.006G200300	AGL17	root epidermis, root, root cap, root elongation zone	Glyma.13G255200	root
Phvul.006G200400	ANR1	root	—	—
Phvul.007G065100	AGL13	root	Glyma.10G240900	root
Phvul.008G027800	CAL	root	Glyma.08G250800	root
Phvul.008G073800	SOC1	root	Glyma.09G266200	root
Phvul.008G183700	AGL21	lateral root primordium, primary root tip, central root cap of primary root, root endodermis, primary root apical meristem, root procambium, root, root stele	Glyma.14G183800	root
Phvul.009G037300	AGL24	—	Glyma.06G095700	root
Phvul.009G203400	FUL	—	Glyma.06G205800	pod shell 10DAF
Phvul.010G088000	SOC1	root	Glyma.07G080900	young leaf
Phvul.010G088100	XAL2	root	—	—
Phvul.011G005800	XAL1	non-hair root epidermal cell, root vascular system, primary root differentiation zone, root, root stele	Glyma.12G005000	root

development. Arabidopsis mutants affected in any of these genes present shorter roots. XAL1 is important for the regulation of cell cycle.[21] XAL2 controls auxin transport in the root and is important in the formation and maintenance of the root stem-cell niche.[22] In agreement, the soybean *XAL1* (Glyma.12G005000) gene shows its highest expression in roots.[18]

ANR1 (ARABIDOPSIS NITRATE REGULATED1) is expressed in Arabidopsis roots (**Table 1**), it is up-regulated upon nitrate starvation and in nitrate rich media it controls local proliferation of lateral roots.[23] Arabidopsis ANR1 is part of a nitrogen sensing network that include NLP7 (NIN-LIKE PROTEIN 7).[23,24] The NIN (NODULE INCEPTION) TF family was first identified in *Lotus japonicus*,[25] now it is known that NIN is crucial for the initiation of nodule formation in different legumes. Recently it was shown that *L. japonicus* NIN directly transcribes the CLE root signal genes, involved in the AON.[26] We identified 3 copies of *P. vulgaris ANR1* genes that are expressed in roots and nodules and one soybean ortholog (Glyma.08G065400) with highest expression in roots (**Table 1, Fig. 1**). Our previous report indicates a positive regulation of common bean NIN and AP2-1 to RIC1 (RHIZOBIUM-INDUCED CLE PEPTIDE1) and suggested the involvement of these regulators in the AON in common bean.[11] On this basis and considering the relevance of nitrogen metabolism in legume nodules we propose that ANR1 as new player in the regulation of legume/rhizobia symbiosis

The regulatory network for flower organogenesis, composed by AP2/ERF, AGL and miRNA members, is still growing in size and complexity. However, only few players have been described in

networks for nodule development. Recently, AP2 genes were described as protagonists in nodulation,[9-11] it seems most probable that these do not play alone. Here we propose AGL transcription regulators as another component of nodulation networks. Future work is required to demonstrate our proposal and to decipher the complete set of players from networks controlling the intricate nodule organogenesis during the symbiotic nitrogen fixation, a most relevant process for sustainable agriculture.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Licausi F, Ohme-Takagi M, Perata P. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytol 2013; 199:639-49; PMID:24010138; http://dx.doi.org/10.1111/nph.12291

2. Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol 2006; 140: 411-32; PMID:16407444; http://dx.doi.org/10.1104/pp.105.073783

3. Jofuku KD, den Boer BG, Van Montagu M, Okamuro JK. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. Plant Cell 1994; 6:1211-25; PMID:7919989; http://dx.doi.org/10.1105/tpc.6.9.1211

4. Aukerman MJ, Sakai H. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. Plant Cell 2003; 15:2730-41; PMID:14555699; http://dx.doi.org/10.1105/tpc.016238

5. Chen X. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 2004; 303:2022-5; PMID:12893888; http://dx.doi.org/10.1126/science.1088060

6. Jung JH, Seo YH, Seo PJ, Reyes JL, Yun J, Chua NH, Park CM. The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. Plant Cell 2007; 19:2736-48; PMID:17890372; http://dx.doi.org/10.1105/tpc.107.054528

7. Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. Cell 2009; 138:750-9; PMID:19703400; http://dx.doi.org/10.1016/j.cell.2009.06.031

8. Yant L, Mathieu J, Dinh TT, Ott F, Lanz C, Wollmann H, Chen X, Schmid M. Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. Plant Cell 2010; 22:2156-70; PMID:20675573; http://dx.doi.org/10.1105/tpc.110.075606

9. Yan Z, Hossain MS, Wang J, Valdés-López O, Liang Y, Libault M, Qiu L, Stacey G. miR172 regulates soybean nodulation. Mol Plant Microbe Interact 2013; 26:1371-7; PMID:23980625; http://dx.doi.org/10.1094/MPMI-04-13-0111-R

10. Wang Y, Wang L, Zou Y, Chen L, Cai Z, Zhang S, Zhao F, Tian Y, Jiang Q, Ferguson BJ, et al. Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate ENOD40 expression and regulate nodule initiation. Plant Cell 2014; 26:4782-801; PMID:25549672; http://dx.doi.org/10.1105/tpc.114.131607

11. Nova-Franco B, Íñiguez LP, Valdés-López O, Alvarado-Affantranger X, Leija A, Fuentes SI, Ramírez M, Paul S, Reyes JL, Girard L, Hernández G. The microRNA172c-APETALA2-1 node as a key regulator of the common bean-Rhizobium etli nitrogen fixation symbiosis. Plant Physiol 2015; 168:273-91; PMID:25739700; http://dx.doi.org/10.1104/pp.114.255547

12. Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res 2004; 32:1372-81; PMID:14988425; http://dx.doi.org/10.1093/nar/gkh299

13. O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, McClean PE, Li J, Dai X, Zhao PX, et al. An RNA-Seq based gene expression atlas of the common bean. BMC Genomics 2014; 15:866; PMID:25283805; http://dx.doi.org/10.1186/1471-2164-15-866

14. Gramzow L, Theißen G. Phylogenomics of MADS-box genes in plants - two opposing life styles in one gene family. Biology (Basel) 2013; 2:1150-64; PMID:24833059

15. Azpeitia E, Davila-Velderrain J, Villarreal C, Alvarez-Buylla ER. Gene regulatory network models for floral organ determination. Methods Mol Biol 2014; 1110:441-69; PMID:24395275; http://dx.doi.org/10.1007/978-1-4614-9408-9_26

16. Causier B, Schwarz-Sommer Z, Davies B. Floral organ identity: 20 years of ABCs. Semin Cell Dev Biol 2010; 21:73-9; PMID:19883777; http://dx.doi.org/10.1016/j.semcdb.2009.10.005 PMID:19883777; http://www.sciencedirect.com/science/article/pii/S1084952109001979

17. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, et al. A reference genome for common bean and genome-wide analysis for dual domestication. Nat Genet 2014; 46:707-713: PMID:24908249; http://dx.doi.org/10.1038/ng.3008

18. Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, et al. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biol 2010; 10:160; PMID:20687943; http://dx.doi.org/10.1186/1471-2229-10-160

19. Yu LH, Miao ZQ, Qi GF, Wu J, Cai XT, Mao JL, Xiang CB. MADS-box transcription factor AGL21 regulates lateral root development and responds to multiple external and physiological signals. Mol Plant 2014; 7:1653-69; PMID:25122697; http://dx.doi.org/10.1093/mp/ssu088

20. Ferguson BJ, Mathesius U. Signaling interactions during nodule development. J Plant Growth Regul 2003; 22:47-72; http://dx.doi.org/10.1007/s00344-003-0032-9 http://link.springer.com/article/10.1007%2Fs00344-003-0032-9

21. Tapia-López R, García-Ponce B, Dubrovsky JG, Garay-Arroyo A, Pérez-Ruíz RV, Kim SH, Acevedo F, Pelaz S, Alvarez-Buylla ER. An AGAMOUS-related MADS-box gene, XAL1 (AGL12), regulates root meristem cell proliferation and flowering transition in Arabidopsis. Plant Physiol 2008; 146:1182-92; PMID:18203871; http://dx.doi.org/10.1104/pp.107.108647

22. Garay-Arroyo A, Ortiz-Moreno E, de la Paz Sánchez M, Murphy AS, García-Ponce B, Marsch-Martínez N, de Folter S, Corvera-Poiré A, Jaimes-Miranda F, Pacheco-Escobedo MA, et al. The MADS transcription factor XAL2/AGL14 modulates auxin transport during Arabidopsis root development by regulating PIN expression. EMBO J 2013; 32:2884-95; PMID:24121311; http://dx.doi.org/10.1038/emboj.2013.216

23. Castaings L, Camargo A, Pocholle D, Gaudon V, Texier Y, Boutet-Mercey S, Taconnat L, Renou JP, Daniel-Vedele F, Fernandez E, et al. The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis. Plant J 2009; 57:426-35; PMID:18826430; http://dx.doi.org/10.1111/j.1365-313X.2008.03695.x

24. Gan Y, Bernreiter A, Filleur S, Abram B, Forde BG. Overexpressing the ANR1 MADS-box gene in transgenic plants provides new insights into its role in the nitrate regulation of root development. Plant Cell Physiol 2012; 53:1003-16; PMID:22523192; http://dx.doi.org/10.1093/pcp/pcs050

25. Schauser L, Roussis A, Stiller J, Stougaard J. A plant regulator controlling development of symbiotic root nodules. Nature 1999; 402:191-5; PMID:10647012; http://dx.doi.org/10.1038/46058

26. Soyano T, Hirakawa H, Sato S, Hayashi M, Kawaguchi M. Nodule inception creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. Proc Natl Acad Sci USA 2014; 111:14607-12; PMID:25246578; http://dx.doi.org/10.1073/pnas.1412716111

e1062957-4   Plant Signaling & Behavior   Volume 10 Issue 10

31

5. Splicing Alternativo

El AS como se explicó en la introducción es un proceso post-transcripcional que se refiere a la transcripción de diversas isoformas de un solo gen. El AS además de generar diversidad protéica funciona como un regulador de la expresión génica ya que a partir de éste se pueden generar proteínas truncas que se regulen por NMD, puede interferir en los sitios de reconocimiento de miRNAs o puede provocar marcos de lectura abiertos río arriba del inicio canónico de la traducción. Las diferentes funciones del AS apenas están siendo exploradas en algunos genes, sin embargo las características globales del AS pueden ayudar a esclarecer algunos aspectos del AS.

Por otra parte el AS tiene un aspecto importante en la evolución ya que algunas isoformas pudieran ser consideradas como parálogos internos. Esto quiere decir que se contaría con dos o más mRNAs de un mismo gen. Por lo tanto una de las preguntas evolutivas del AS es qué sucede con el AS cuando un gen se duplica.

5.1. AS en *Phaseolus vulgaris* y *Glycine max*

5.1.1. Introducción

El estudio del AS en plantas se ha centrado más en plantas modelo como *A. thaliana* u *O. sativa* (arroz). En leguminosas este aspecto se ha dejado a un lado ya que existen muy pocos reportes sobre AS en esta familia. A nivel de leguminosas se reportó que el evento de AS que más sucede es IR sin embargo este estudio se basó en secuencias de EST's por lo que la incidencia del AS fue bastante baja (~20 % de los genes con AS) (Wang et al., 2008). A pesar de esos resultados se lograron identificar eventos de AS conservados en las leguminosas analizadas. Los eventos de AS conservados, como se mencionó en la introducción, sugieren una funcionalidad y por ende una relevancia biológica.

La utilización de EST's para identificar AS es importante ya que mediante esta técnica se obtienen isoformas completas, pero debido a las limitaciones de la técnica sólo se alcanza a

apreciar una pequeña porción del AS. A partir de las metodologías de secuenciación masiva (RNA-seq) se pueden distinguir muchos más eventos de AS, pero no se cuenta con la información necesaria para obtener isoformas completas. A partir de RNA-seq se ha estudiado el AS en las leguminosas *Medicago truncatula* y en *Glycine max* (Mandadi and Scholthof, 2015; Shen et al., 2014; Wang et al., 2014). En estos estudios de ha establecido que el evento más común es el IR, consistente con lo reportado en *A. thaliana* (Marquez et al., 2012) sin embargo sólo se han enfocado en 4 diferentes tipos de eventos de AS; AA, AD, ES e IR.

Las recientes publicaciones del genoma (Schmutz et al., 2014; Vlasova et al., 2016) de *P. vulgaris* en conjunto con las bibliotecas públicas de RNA-seq de esta especie permitieron plantear el proyecto sobre la identificación de eventos de AS en frijol. Para identificar aquellos eventos biológicamente relevantes se utilizaron bibliotecas públicas de *G. max*, una especie relacionada filogenéticamente a frijol. Estos datos conllevan al mayor estudio de identificación, caracterización y conservación de eventos de AS con el que se cuenta hasta la fecha.

5.1.2. Resultados y Discusión

Para identificar los eventos de AS en *P. vulgaris* se utilizaron todas las muestras disponibles en los archivos SRA de RNA-seq de esta leguminosa. En total se utilizaron 157 bibliotecas pertenecientes a 107 muestras biológicas diferentes, la longitud de las secuencias estuvo en un rango de 36-101 nucleótidos en bibliotecas tanto single-end como pair-end. Se analizaron más de 11×10^9 secuencias. Además, se utilizaron 176,782 secuencias más largas reportadas como EST's. En cambio para *G. max* se utilizaron menos bibliotecas (84 de 77 muestras) de RNA-seq (3.5×10^9 secuencias) pero más EST's (1,461,723). La cantidad de muestras analizadas para cada especie y en total supera por mucho cualquier análisis de identificación de AS en alguna otra planta.

Para la identificación de los eventos a partir de las bibliotecas de RNA-seq se utilizaron dos diferentes algoritmos, cuyas estrategias difieren en la forma de identificar los eventos de AS. A partir de estos resultados se logró identificar que alrededor del 70 % de los genes

expresados presentan algún tipo de AS. Los eventos se catalogaron en los 7 diferentes tipos de eventos de AS e IR fue el evento más común en ambas leguminosas, seguido de AA y AD. La concordancia en los resultados de ambas plantas muestra la importancia del AS en las leguminosas y sobre todo sugiere mecanismos para seleccionar eventos de tipo IR, AA y AD.

Los diferentes tipos de eventos también fueron caracterizados dependiendo de la región génica que afectaban. Sorprendentemente los resultados de este análisis muestran que existe una sobre-representación de regiones no codificantes para eventos de tipo AA, AD, ASS y RE, mientras que IR sucede más en regiones CDS en ambas plantas. Otra caracterización de los diferentes tipos de eventos que se realizó fue respecto a su posición relativa dentro de un gen. Los eventos tanto de AA como AD tienden a estar en regiones 5' del gen mientras que IR tiende a presentarse hacia el extremo 3'. Al igual que los resultados de regiones codificantes ambas leguminosas muestran el mismo patrón.

Para identificar los eventos de AS que se conservaran entre especies se identificaron los genes homólogos. La conservación de eventos de AS de frijol en soya va desde el 22 % hasta el 37 % dependiendo de cuáles eventos se tomen en cuenta. Se observa una conservación del 22 % cuando se toma la totalidad de eventos de AS identificados en *P. vulgaris* mientras que el 37 % se refiere a los eventos conservados únicamente en genes homólogos. Mientras tanto los eventos conservados de *G. max* en *P. vulgaris* van desde el 18 % hasta el 35 %. Otro resultado interesante de la conservación surge a partir de analizar cada tipo de evento por separado ya que ambas leguminosas presentan el mismo patrón. IR es el evento más conservado seguido de AA y AD, los eventos más comunes y le siguen NI y RE. Esto muestra que estos últimos tipos de eventos pudieran tener una relevancia mayor que ASS y ES.

Dado el supuesto de que los eventos conservados tienen una función biológica relevante se realizó un análisis de regiones génicas (5'UTR, CDS y 3'UTR) donde se encontró un enriquecimiento de los eventos conservados en regiones 5'UTR mientras que una disminución notable en regiones 3'UTR. Esto pudiera significar que los eventos en las regiones 3'UTR se comportan de forma aleatoria y tienen una función poco relevante o nula.

Otro aspecto importante dentro de este trabajo fue la validación de eventos de AS. Se tomaron muestras de raíz y hojas a tempranas edades de crecimiento de frijol y soya y se probaron distintos eventos de AS en 9 genes. La metodología que se utilizó para la validación fue RT-PCR en punto final a partir oligonucléotidos diseñados para identificar el splicing constitutivo y el alternativo en una misma reacción.

El análisis detallado, los resultados y discusión de este trabajo fueron publicados en el artículo: **Iñiguez LP**, Ramírez M, Barbazuk WB, Hernández G (2017) IDENTIFICATION AND ANALYSIS OF ALTERNATIVE SPLICING EVENTS IN *PHASEOLUS VULGARIS* AND *GLYCINE MAX*. *BMC Genomics* 18:650 . El artículo se muestra a continuación:

RESEARCH ARTICLE

Identification and analysis of alternative splicing events in Phaseolus vulgaris and Glycine max

Luis P. Iñiguez[1*], Mario Ramírez[1], William B. Barbazuk[2] and Georgina Hernández[1*]

Abstract

Background: The vast diversification of proteins in eukaryotic cells has been related with multiple transcript isoforms from a single gene that result in alternative splicing (AS) of primary transcripts. Analysis of RNA sequencing data from expressed sequence tags and next generation RNA sequencing has been crucial for AS identification and genome-wide AS studies. For the identification of AS events from the related legume species Phaseolus vulgaris and Glycine max, 157 and 88 publicly available RNA-seq libraries, respectively, were analyzed.

Results: We identified 85,570 AS events from P. vulgaris in 72% of expressed genes and 134,316 AS events in 70% of expressed genes from G. max. These were categorized in seven AS event types with intron retention being the most abundant followed by alternative acceptor and alternative donor, representing ~75% of all AS events in both plants. Conservation of AS events in homologous genes between the two species was analyzed where an overrepresentation of AS affecting 5'UTR regions was observed for certain types of AS events. The conservation of AS events was experimentally validated for 8 selected genes, through RT-PCR analysis. The different types of AS events also varied by relative position in the genes. The results were consistent in both species.

Conclusions: The identification and analysis of AS events are first steps to understand their biological relevance. The results presented here from two related legume species reveal high conservation, over ~15–20 MY of divergence, and may point to the biological relevance of AS.

Keywords: Alternative splicing, Conservation of alternative splicing, RNA-seq, Legumes, Common bean, Soybean

Background

The majority of protein-coding genes from eukaryotic organisms contain introns, non-coding sequences that need to be spliced from the primary transcript to generate mature functional mRNAs. Although some introns can be self-spliced, most require a spliceosome, specialized splicing machinery. Spliceosomes are large ribonucleoprotein complexes that include small nuclear RNAs (snRNA) [1–3]. The spliceosome recognizes signals from common introns that allow their removal from the pre-mRNA. The U1 snRNA recognizes signals from the 5′ splice site, a GT dinucleotide. The U2 snRNA recognizes the 3′ splice site that includes an AG dinucleotide, an adenine which functions as a branching point upstream

of the 3′ splicing site and a polypyrimidine tract between the branching point and the 3′ splicing site [4]. Other snRNAs, such as U11 and U12, recognize different splice sites, although spliceosomal introns of this class represent a minority [3, 4]. Different proteins that facilitate the recognition of the motifs by the spliceosome also mediate splicing. Serine/arginine-rich (SR) proteins facilitate the splicing of the intron while heterogeneous nuclear ribonucleoproteins inhibit the recognition of splicing sites [5]. The intron motifs as well as splicing enhancers and inhibitors are commonly found at different sites within the intron and these juxtaposed signals can give rise to variation in splicing an intron from the pre-mRNA; this phenomenon is known as alterative splicing (AS) [6–8].

AS is a post-transcriptional regulatory process that affects the fate of the mRNA; it has been found in several tissues, stress conditions and developmental

---

* Correspondence: liniguez@lcg.unam.mx; gina@ccg.unam.mx
[1]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, Mexico
Full list of author information is available at the end of the article

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 2 of 17

stages of eukaryotic organisms [9]. AS can affect the localization of the mature mRNA and their translation efficiency [8]. Also, AS may produce alternative stop codons due to frame shifts in the mature mRNA sequence, thus regulating mRNA abundance by nonsense-mediated decay (NMD) [10]. Some other process linked to AS, such as mRNA storage or the target recognition of micro RNAs, have been reported [11, 12]. AS may also result in different protein isoforms derived from a single gene thus affecting protein localization or function [7, 13].

The evolutionary significance of AS has been related to organismal complexity. The number of genes in nematodes is very similar to that in humans although these organisms develop strikingly different cell types [14]. However, 98% of human multiple-exon genes undergo AS [15] in contrast to only 20% of nematode genes [16]. Chen et al. [17] analyzed several organisms that vary in their amount of different cell types –a proxy for organismal complexity- and found a strong positive correlation between the number of cell types and the level of AS. Organisms with higher complexity tend to have higher levels of AS.

AS genome-wide analyses, based on expression sequence tags (ESTs) and next generation RNA sequences (RNA-seq), have been reported for several plant species [18–27]. The first available resources for such analysis were EST databases that included sequences of large mRNA fragments often representing complete mRNA isoforms, but due to sampling the number of AS isoforms were likely underestimated. Next generation RNA-seq technologies produce a huge amount of sequences; however, these sequences are too short for complete isoform identification but can be used for characterization of AS events.

The recognition of splicing sites and the frequencies of AS types, vary between plants and animals. Animals tend to have very large introns and therefore the splicing machinery recognizes exons (exon definition) while in plants the spliceosome recognizes introns (intron definition) [28]. It has been proposed that a failure in the exon definition can lead to skipping an exon during splicing, while failure in intron definition results in intron retention [29]. The intron and exon definition models could explain the differences in the most common AS processes observed between plants and humans. Specifically, intron retention is the most abundant AS type observed in plants [30] while exon skipping is most common in animals [15].

Our research has focused on genome-wide analyses of transcriptional and post-transcriptional regulation in legume plants. Legumes are second only to Gramineae in their importance as crops; they are rich in protein content and have long been used for humans and animal consumption. Legumes are important contributors to

biological nitrogen due to their ability to establish symbiosis with nitrogen-fixing soil bacteria (rhizobia). This relationship allows the legumes to grow under low or non-nitrogen fertilized media. Symbiotic nitrogen fixation has been a focus of research due to its economic and environmental importance [31]. Common bean (*Phaseolus vulgaris*) and soybean (*Glycine max*) are the most important legume crops worldwide. Common bean is the most important legume for human consumption as a source of proteins and micronutrients for millions of people, especially in Latin America and Africa where beans are important components of traditional diets [32]. Soybean is important worldwide as the dominant source of protein for animal feed and cooking oil [33]. These legumes are closely related and their evolutionary history makes them ideal models for genomic studies. Both the *P. vulgaris* and *G. max* genomes have been sequenced [33–35]. These legume species are evolutionary closely related, having diverged only ~19.2 million years ago (MYA), and share a whole-genome duplication (WGD) event ~56.5 MYA. *G. max* experienced an independent WGD ~10 MYA [34]. Thus, they are also good models to analyze features related to polyploidization.

This work analyzes AS in *P. vulgaris* and *G. max* by identifying and characterizing seven different AS events types genome-wide. This includes the identification of the introns/exons affected by AS and their relative position in the gene. AS event conservation between these legumes helps to elucidate some important aspects of the different types of AS. This work increases our knowledge of AS in legumes and provides a platform for further investigation.

Results and discussion
AS identification
The characterization of AS events is a first step to understand the importance and prevalence of AS in plants. Four different types of AS events are the most frequently described in the literature: exon skipping (ES), where a whole exon is missed in comparison to the primary transcript; intron retention (IR), an intron is not spliced and is part of the mature mRNA; alternative donor (AD), the donor site, also known as 5′ splicing site, change in the mRNA isoform; and alternative acceptor (AA), where the 3′ splicing site is different. Based on a primary transcript three additional AS events can be described; alternative splicing sites (ASS), where both donor and acceptor sites change; new intron (NI), when a splicing site appears in a reported exon; and retained exon (RE), a new exon replaces a previously annotated intron in the mature mRNA. Schematic representations of these seven different type of AS events are presented in Fig. 1.

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 3 of 17

Fig. 1 AS events. The types of AS events identified in this work are show, all of these are based on a primary transcript. Transcripts affected by AS are highlighted with an asterisk on the right side of the drawing. Drawing by Illustrator for Biological Sequences [66]
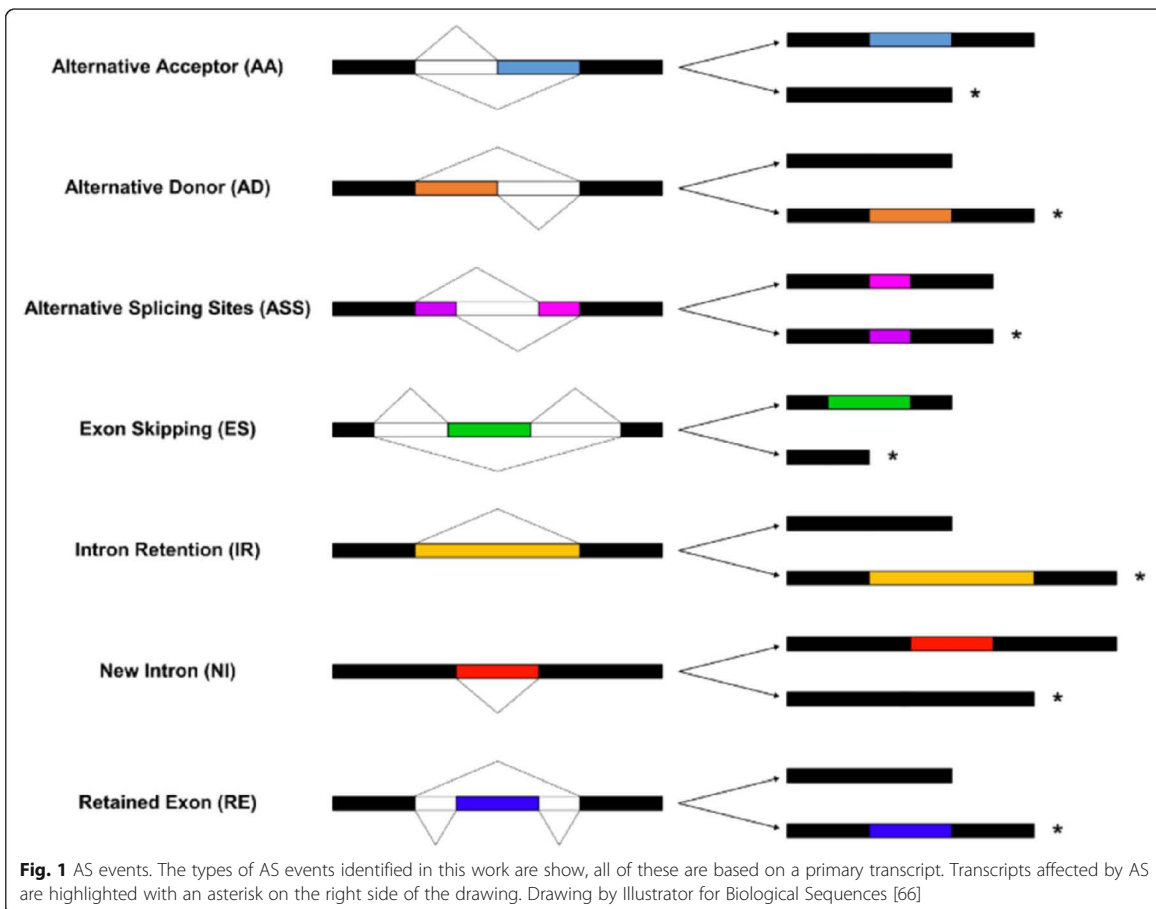
AS events were called when an exon-exon junction of a gene showed evidence for two or more splicing alternatives. The short length of RNA-seq sequences used for analysis allows the identification of AS events, but excludes the possibility of identifying whole isoforms; thus, alternative transcription start site, alternative polyadenylation site, translation start site and trans-splicing were not analysed.

Both the *P. vulgaris* and the *G. max* genome annotations [33, 34] include gene isoforms produced by AS in ~10 and ~23% of their genes, respectively. However, this is not consistent with reports that show higher proportions of AS genes in other plants [18–27, 36, 37]. The most frequent AS event reported for both legume genomes is AA, followed by AD and ES. In *P. vulgaris* the order of frequency following the first three events is IR, RE and NI and in *G. max* it is RE, NI and IR. The least frequent AS event reported for both plants is ASS (Additional file 1). Again, these results are not consistent with the AS profiles reported for other plants where IR the most common AS event, followed by AA

and AD, and ES being the least common [7, 8, 36, 37]. These results suggest that the current annotation lacks a comprehensive identification of AS isoforms in these two legume species.

Here, RNA-seq and EST's were analysed based on the methodology illustrated in Fig. 2, where two different mapping algorithms were used. The qualitative workflow allowed to identify AS events, based on the reported primary transcript. Thus AS events could be identified in any sample independently of the existence of the primary transcript in the same sample. This methodology (Fig. 2) led to the identification of AS events from *P. vulgaris* and *G. max* with a coverage of 88 and 72% of previously reported AS events, respectively. IR events were the most frequent for both species and ES the least frequent (Table 1). A total of 82,343 and 115,881 new AS events were identified for *P. vulgaris* and *G. max*, respectively (Table 1 and Additional file 2). The complete list of events covered 65% of *P. vulgaris* annotated genes and considering only the expressed genes included in the analysed RNA-seq samples (24,862), it

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 4 of 17

Fig. 2 Method for AS event identification. Workflow for the identification of AS events in a single species. RNA-seq reads were mapped to their respective genome by two different algorithms, each mapping file was processed with Cufflinks to obtain the gene models. All gene models were compared to the primary transcripts to identify AS events. EST's were also mapped and with both mapping files as well as the replicates all events were filtered. An expression value for each gene was calculated based on the mapping files

was 72% of genes (Table 2). For *G. max*, 55% of annotated genes were affected by AS, or 70% of genes included in analysed RNA-seq libraries (43,712) (Table 2).

Almost all (~98%) reported junctions in both legumes present the canonical intron motifs for the spliceosome recognition: the 5′ splicing site GT and the 3′ splicing site AG [4]. Most of the new junctions identified in this work, 78% for *P. vulgaris* and 85% for *G. max,* present the exact same motifs thus being considered substrates of the spliceosomal machinery (Additional file 3). The rest of the identified introns that presented non-canonical splicing sites, were also considered in our analysis based in previous knowledge about relevant regulatory roles of spliceosome-independent (self-splicing) introns from other organisms [38]. Nevertheless, the proportion of canonical splicing sites considered in this analysis –including reported and newly identified junctions- remains >93% for both legumes (Additional file 3).

The proportion of genes affected by AS are similar to those reported in other plants [18–27, 36, 37]. However, the distribution among types of AS events for both species differ from that reported in the genome annotations [33, 34] (Table 1 and Additional file 1). The three most frequent AS events identified for both plants were IR, AA, AD; corresponding to ~75% of all AS events (Table 1). Despite the different samples used for both species, the proportion of events as well as the number of genes affected by AS were similar (Tables 1 and 2). In mammals, the most common AS event is ES, ~50% of all events, this contrasts with plants where ES is generally less than 10% (Table 1) [9, 39]. Key processes have been implicated in the functionality of AS in plants for AA, AD and IR events. IR has been implicated in the process of NMD [10] due to the incorporation of stop codons. IR also plays an important role in *Mariselea vestita* in mRNA storage during it embryo development [12]. AA and AD are

Table 1 Genome-wide AS identification

AS event	*Phaseolus vulgaris*				*Glycine max*			
	Identified from genome annotation[a] [34]	New	Total	%	Identified from genome annotation[a] [33]	New	Total	%
AA	1076	16,997	18,073	21	5209	26,901	32,110	24
AD	814	12,464	13,278	15	4126	20,113	24,239	18
ASS	59	6001	6060	7	343	5300	5643	4
ES	579	4285	4864	6	2543	8351	10,894	8
IR	387	33,091	33,478	39	1980	41,016	42,996	32
NI	115	3819	3934	5	2325	5267	7592	6
RE	377	5686	6063	7	1909	8933	10,842	8

[a]Total AS events reported in the annotated genomes are shown in Additional file 1

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 5 of 17

Table 2 Genes affected by AS

AS event	Phaseolus vulgaris			Glycine max		
	Number	Percentage		Number	Percentage	
		All	Expressed		All	Expressed
AA	8820	32%	35%	16,092	29%	37%
AD	7339	27%	30%	13,872	25%	32%
ASS	4117	15%	17%	3893	7%	9%
ES	3396	12%	14%	7431	13%	17%
IR	14,162	52%	57%	22,058	39%	50%
NI	3650	13%	15%	6990	12%	16%
RE	4940	18%	20%	8967	16%	21%
Total	17,789	65%	72%	30,677	55%	70%

reported as consequences of small duplications in the splicing sites, allowing the incorporation or exclusion of one or several amino acids [40–43].

Experimental validation of identified AS events

The RT-PCR approach was used to experimentally validate AS events selected from those identified in ~70% expressed genes from *P. vulgaris* and *G. max* genomes (Tables 1 and 2). Nine *P. vulgaris* genes and their corresponding *G. max* homolog, expressed in the plant tissues analyzed (roots and leaves from seedings), were selected for RT-PCR analysis. The selected genes presented conserved AS within both species (see subsequent sections). The two RT-PCR reactions performed for each gene, with RNA samples from root and from leaf, showed similar results in every case; Fig. 3 shows the results from leaves samples except for panels a and f that show results from root samples. In every gene analyzed the amplified products (ranging from 166 to 1599 bp) corresponded to expected fragments from the primary transcript or from transcript isoforms derived from AS events, according to each gene model (Fig. 3). The different types of AS events (Fig.1) validated for the selected genes included: AA (Fig. 3a–g), ASS (Fig. 3f), IR (Fig. 3b–h) and RE (Fig. 3b, c). As expected, only one amplified product corresponding to the primary transcript, could be observed in the control gene selected (Fig. 3i). In Fig. 3g, h additional amplification products, from those predicted from the gene models, could be observed; we cannot rule out that these correspond to AS events not identified in our analysis something that could be related to restrictions in the methodology we used. Taken together, the experimental results (Fig. 3) do validate and increase the reliability of the bioinformatic data from this work.

AS in CDS-UTR regions

The seven event types analysed here can be divided into two classes depending on how the AS event modifies the reference transcript. AA, AD, ASS, IR and RE events modify reference exon-exon junctions, while NI and ES alter reference exons, either excluding it or introducing a new intron (Fig. 1). There are three main types of introns defined by the untranslated regions (UTR) and coding DNA sequences (CDS) of the primary transcript, and six types of exons. The intron classification is based on the types of exon they are delimiting: 5'UTR-5'UTR, CDS-CDS and 3'UTR-3'UTR. Exons, on the other hand, can be classified in seven different types, 5'UTR, CDS, 3'UTR, 5'UTR-CDS, when the translation start site is in that exon, CDS-3'UTR, when the stop codon is in this exon, and 5'UTR-3'UTR which are genes without introns.

In both *P. vulgaris* and *G. max*, most introns of the primary transcript are CDS-CDS (~94%) and the rest are 5'UTR-5'UTR (~4%) and 3'UTR-3'UTR (~2%) (Fig. 4a-c). In the case of exons, also in both genomes, the majority (69%) were CDS (Fig. 4e), while the other exon types, 5'UTR-CDS and CDS-3'UTR, correspond to 12% (Fig. 4g, h). The other 7% of exons were divided into 3% single exon genes (Fig. 4i), 3% 5'UTR and 1% 3'UTR exons (Fig. 4d, f).

The introns, or exon junctions, affected by the AS were analysed to determine if they were randomly distributed. The percentage of each type of intron/exon in the genome was compared to the percentage affected by AS. Both legumes showed an enrichment of AS events in UTR introns (Additional file 4), whereas, AS events in CDS junctions were under-represented, despite being the majority of affected junctions. Similar data were observed for each individual AS event (Fig. 4a-c). In the case of the exons affected by AS a decrease of CDS exons and an over-representation of every other type of exon were observed (Additional file 4). Interestingly in the case of exons, the under- and over-representation of each type of AS event varied among types of exons. CDS exons were underrepresented in NI events while overrepresented in ES events (Fig. 4e). In contrast, 5'UTR-CDS and CDS-3'UTR were enriched in NI and decreased in ES events (Fig. 4g, h). Common bean and soybean showed similar results (Fig. 4).

While the main effect of ES is to skip CDS exons, NI introduces introns in combined exons such as 5'UTR-CDS or CDS-3'UTR. Marquez et al. [44] reported the presence of NI in single exon genes and called them "exitrons". Together, these results point to a non-random distribution of AS and to the selection of AS in specific regions of the genes. The similar results obtained for both legumes suggest important aspects of specific AS events in these species.

An interesting example of AS in CDS-CDS regions was identified the CSN7 gene, this protein is one of the eight subunits of the COP9 signalosome (CSN) that is a key player in the DNA-damage response, cell-cycle
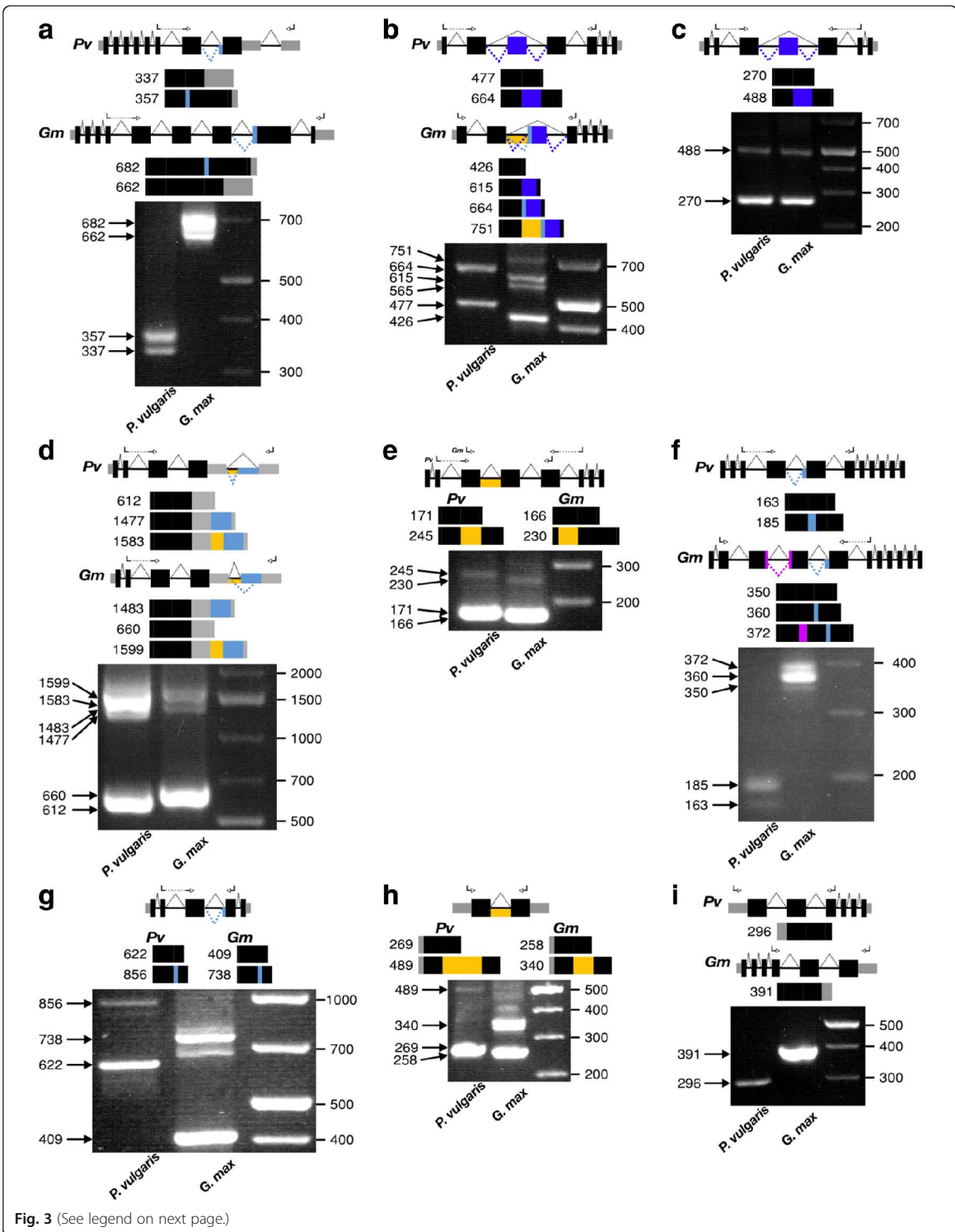
Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 6 of 17

Fig. 3 (See legend on next page.)

41

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 7 of 17

(See figure on previous page.)
Fig. 3 Experimental validation of AS events in *Phaseolus vulgaris* and *Glycine max* genes. Each selected gene, with conserved AS events in both species, is shown in a different panel: **a**) Phvul.011G190600, Glyma.13G187200; **b**) Phvul.008G270400, Glyma.02G293300; **c**) Phvul.007G262600, Glyma.09G104200; **d**) Phvul.005G127700, Glyma.12G181300; **e**) Phvul.003G077300, Glyma.07G259200; **f**) Phvul.002G298304, Glyma.08G023000; **g**) Phvul.001G014900, Glyma.14G075500; **h**) Phvul.006G191200; Glyma.13G245600; **i**) Phvul.008G013100, Glyma.18G289000. From top to bottom each panel includes: drawing of the gene model or of a different gene model for each species (not drawn to scale) with arrows indicating the position of the primers used for RT-PCR reactions and dotted lines indicating the splicing resulting in the primary transcript (above the gene model line) and the AS (below the line), the color code for different types of AS events is the same used in Fig. 1; drawings representing the amplification products expected for each transcript isoform, with its size (bp) indicated at the left; RT-PCR products resolved in 3% agarose gels, arrows indicate size (bp) of predicted fragments, the GeneRuler 1 kb Plus DNA ladder (Thermo Scientific, USA) was included for reference (third lane)

control and gene expression. CSN7, as well as other 5 subunits, contains a N-terminal PCI domain that is important for subunits interactions. In addition, CSN7 C-terminal tail is responsible for interactions with the non-PCI protein CSN6 as well as with other proteins such as the ribonucleotide reductase RNR2. The different protein-protein interactions of CSN7 C-terminal regulate the CSN complex assembly as well as the function of RNR2 [45]. As shown in Fig. 3a, the CSN7 gene from *P. vulgaris* (Phvul.011G190600) and *G. max* (Glyma.13G187200)

presents a conserved a AS event, of the AA type, in a CDS-CDS junction (intron 7) of the C-terminal region. The reported primary transcript differs among these species, being the *P. vulgaris* primary transcript similar to an alternative transcript isoform of *G. max* and viceversa (Fig. 3a). The AA transcript isoform from *P. vulgaris* presents a modification of the reading frame from exon 8 that shifts the stop codon to exon 9 (Fig. 3a). In *G. max* CDN7 the stop codon of the primary transcript is located in exon 9 and a similar AA event is present on intron 7 with the
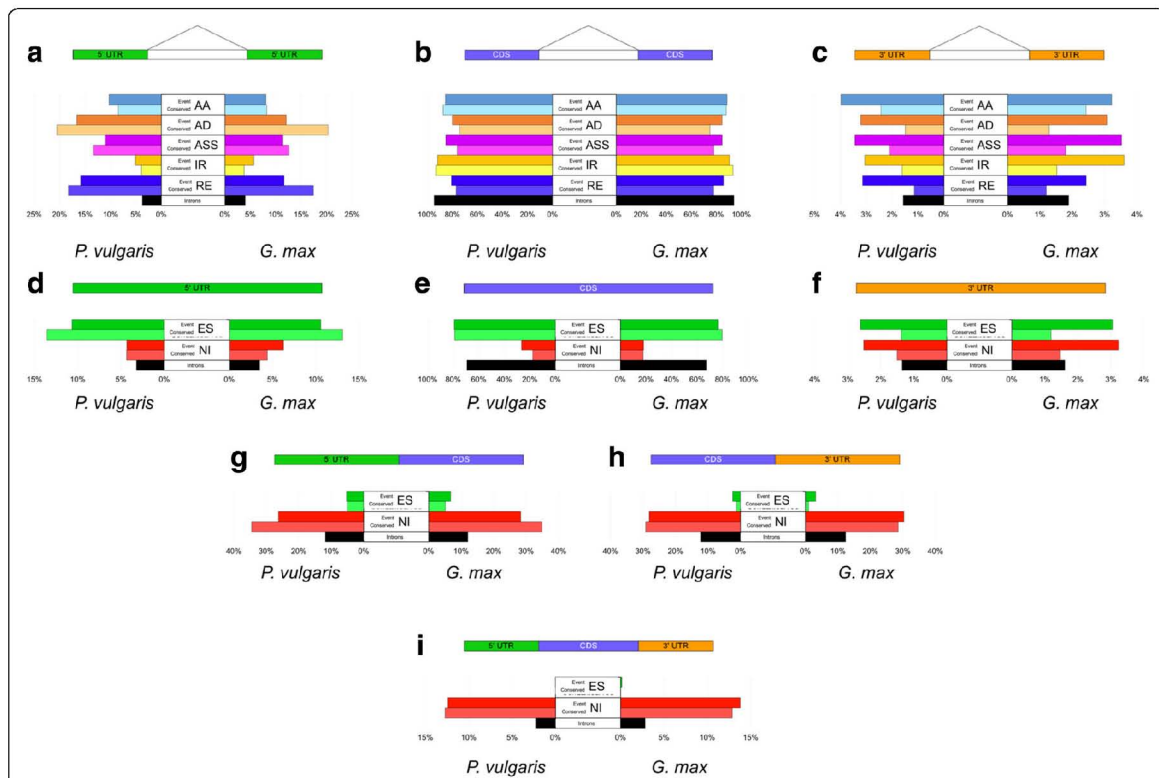
Fig. 4 Percentage of introns and exons from UTR or CDS regions affected by AS events. AS events are dissected in terms of the percentage of exons or introns they affect. **a**, **b** and **c** show the introns affected by AA, AD, ASS, IR and RE (top to bottom, colored bars) and the percentage of introns in the genome (black bar, bottom). Each AS event includes two bar plots: the upper bar (dark tone color) shows the percentage of the type of junction is affected by AS event and the bottom bar (light tone color) shows the percentage of the type of junction affected in conserved AS events. The left side of each graph corresponds to *Phaseolus vulgaris* and the right side to *Glycine max*. **d**, **e**, **f**, **g**, **h**) and **i**) follow the same structure as the above, showing the percentage of exons affected by ES and NI

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 8 of 17

stop codon in exon 8 (Fig. 3a). This AA event on both species modifies the C-terminal of CSN7, we propose this could regulate the interactions with different proteins thus affecting the functionality of CNS in these legume species.

Conservation of AS between *P. vulgaris* and *G. max*

Different approaches may be used to analyse the conservation of AS between homologous genes. Here, we used a junction conservation approach rather than the overly strict position conservation approach. The position conservation approach is based on the conservation of the event in exact positions while for junction conservation only the event and intron must be conserved [46]. An AS event was considered conserved if both homologous genes from *P. vulgaris* and *G. max* showed the same AS event at a specific junction. Redundant AS events were removed from identified AS events in both plants. Redundant AS events imply that the same type of event occurred in the same intron or exon but at a different position. Since a junction conservation approach was used for the AS event conservation analysis the exact position of the AS event was, for this study, irrelevant.

The first step to identify the conservation of AS events was to define *P. vulgaris* and *G. max* homologous genes with the same gene model. *P. vulgaris* and *G. max* have a relative short evolutionary distance of ~19.2 MY [34] and soybean experienced a recent WGD, ~10 MYA; therefore, a high proportion of common bean genes (51%) have two homologous genes in soybean (representing 50% of the total gene set), resulting in a 1:2 relationship. As shown in Fig. 5a, 13,962 *P. vulgaris* genes with 2 *G. max* homologs were identified. Of these 55% (7693) were expressed and had the same gene model in both *G. max* homologs. There were 7039 *P. vulgaris* genes with only one identifiable homolog in *G. max* (Fig. 5b) (including genes with a 1:2 relationship but where one did have the same gene structure or only 1 *G. max* homolog was expressed, plus those genes with only one expressed homolog with the same gene structure). In total 14,712 *P. vulgaris* genes and 22,405 *G. max* genes, representing more than 50% of expressed genes in both plants, were selected for analysis of AS conservation.
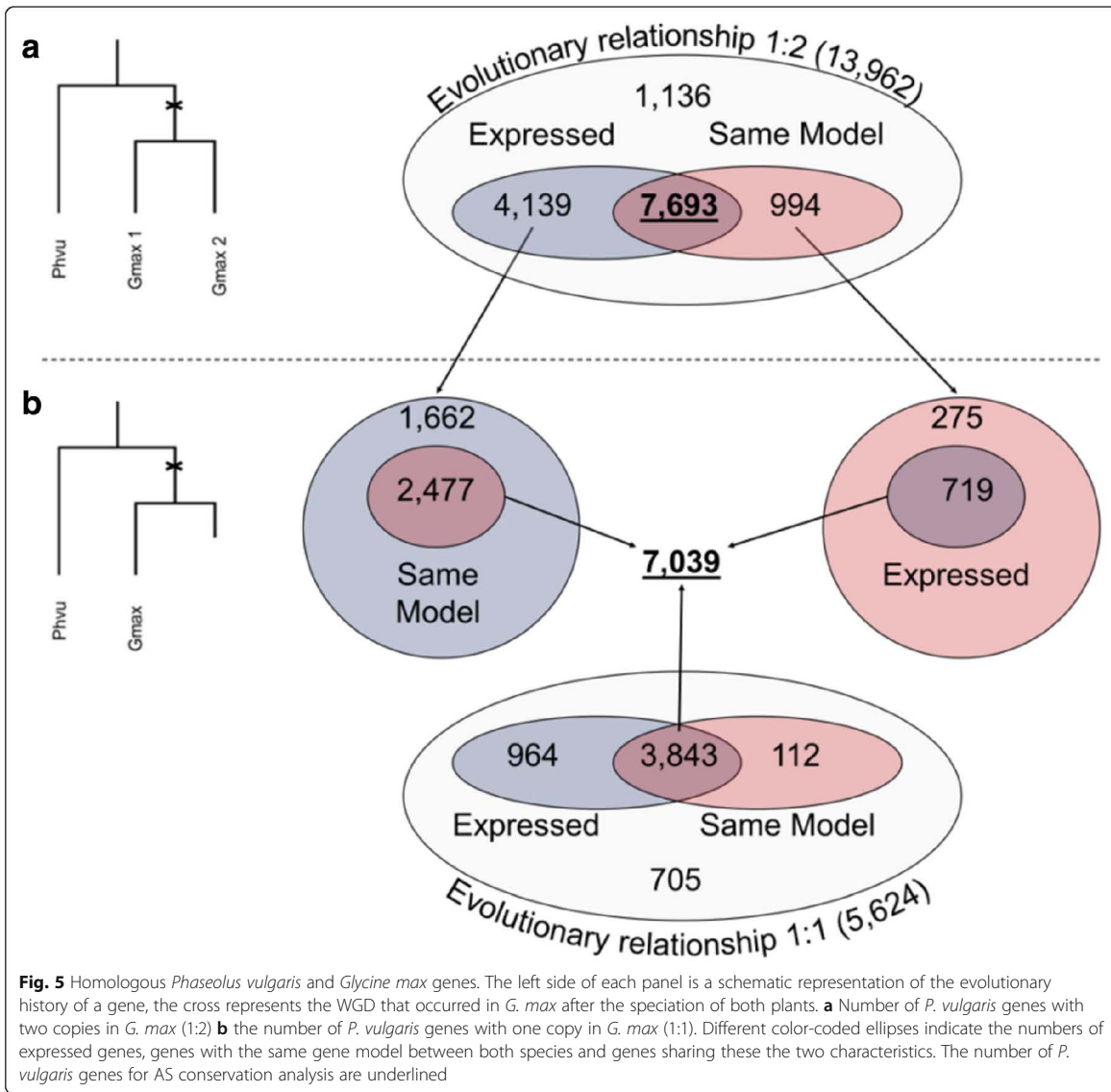
The junction conservation approach considers the type of AS event and the affected intron/exon. Therefore, if two events from the same type that coincide in a single intron/exon, even although they differ in positions, were considered redundant and were collapsed into nonredundant events. 61 and 51% of all non-redundant events belong to the genes described in Fig. 5 in *Phaseolus vulgaris* and *Glycine max*, respectively. The proportions of the types of AS in the non-redundant events is similar in genes with homology in the other species relative to the proportions of AS types in all expressed genes (Additional file 5).

Considering only homologous genes with the same model between *P. vulgaris* and *G. max* the rate of conservation of *P. vulgaris* AS events within *G. max* was 37%, here assumed as the maximum conservation rate. Since not all the homologous genes were considered, the minimum rate of AS conservation was 22% based on all non-redundant events in common bean (78,027) (Fig. 6a). On the other hand, the rate of AS conservation of *G. max* in *P. vulgaris* homologs with the same gene model was 35 and 18% of AS events for all soybean non-redundant events (121,133) (Fig. 6b). Interestingly, conservation differed depending on the type of AS event. The proportion of IR events from *P. vulgaris* conserved in *G. max* was 45 (Fig. 6a) and 53% of the soybean IR events were conserved in common bean homologous genes (Fig. 6b). Following the order of conservations, AA and AD stand after IR (Fig. 6a, b), coinciding with the order of abundance of AS types (Table 1).

These results support the suitability of using the junction approach for identifying AS conservation among homologous genes. Chamala et al. [46] reported that the proportion of conservation of AS events varies among the types of AS events. They analysed four types of events (AA, AD, ES and IR) and found IR to be the most conserved type, consistent with our data. The results of Chamala et al. [46] and this work indicate that IR, AA and AD are not only the most common events in both legume plants but these present the highest AS event conservation rate across angiosperms. Their evolutionary conservation indicates potential function. Although ASS and ES as well as NI and RE have similar proportions in both species, the first two are less conserved. This could be interpreted as ASS and ES being more species specific than NI and RE or their function is not conserved.

The data used for these analyses also enabled the identification of AS events conserved between soybean paralogs that arose during the last WGD. The paralogous genes set must have identical gene models in order to identify the conserved events by junction. A total of 40% AS events were conserved among *G. max* paralogous genes (Fig. 6c). Similar as seen for AS conservation between species, IR was the type of AS event with the highest conservation rate followed by AA and AD; while AS and ES showed the lowest percentage of conservation although this was higher than 20% (Fig. 6c). New AS events could arise after the WGD within either one or both paralogs, isoforms could have predated the WGD and still remain in both paralogs, or only one of the paralogs could have lost its isoform subsequent to the WGD event. To address this question further analyses on conserved AS within genes duplicated through WGD in another species need to be examined.

The conservation of AS events between these two legumes may be the result of their performing an essential

43

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 9 of 17

Fig. 5 Homologous *Phaseolus vulgaris* and *Glycine max* genes. The left side of each panel is a schematic representation of the evolutionary history of a gene, the cross represents the WGD that occurred in *G. max* after the speciation of both plants. **a** Number of *P. vulgaris* genes with two copies in *G. max* (1:2) **b** the number of *P. vulgaris* genes with one copy in *G. max* (1:1). Different color-coded ellipses indicate the numbers of expressed genes, genes with the same gene model between both species and genes sharing these the two characteristics. The number of *P. vulgaris* genes for AS conservation analysis are underlined

function, particularly since these events have been conserved over ~20 MY. The isoforms produced due to AS could be important for a specific tissue, condition or developmental stage. To understand the function, and/or temporal and spatial conditions under which these AS isoforms are expressed will require additional investigation. However, since the majority of the AS events are not conserved, the lack of conservation of AS events may not only be due to their biological function, but may reflect the divergence time between these species. Paralogous genes in *G. max* with shorter diversification time accounted for a majority of the gene specific AS events. A combination of functionality and

diversification time could lead to the percentage of AS conservation observed here.

Two examples of AS events conserved among plant species that indicate their important biological functions, are the following. The transcription factor TFIIIA is required for the synthesis of 5S–rRNA by RNA polymerase III. The third exon of this gene, with a structural element that mimics 5S rRNA, presents an ER event [47–49]. This TFIIIA exon, that is highly conserved in land plants [50], is referred to as a suicide exon because its absence produces a functional transcript whereas its retention results in a non-functional transcript containing premature termination codons (PTC) thus targeted

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 10 of 17

Fig. 6 Conservation of AS among *Phaseolus vulgaris* and *Glycine max*. Cumulative bar plots represent total AS events (black bars) and each AS event type (colored bars). In every bar plot the stripped bar (top) corresponds to AS events in non-homologous genes, the dark tone color bar (middle or top) to conserved AS events and the light tone color bar (bottom) to non-conserved AS events in homologous genes. The numbers over the bar indicate the percentage of conserved events over all AS events (top) and over AS events in homologous genes (bottom). AS event conservation of *P. vulgaris* in *G. max* (**a**), of *G. max* in *P. vulgaris* (**b**) and in paralogous *G. max* genes, originated by its recent WGD (**c**)

to NMD. The L5 ribosomal protein binds to the 5S rRNA mimic and controls the synthesis of 5S RNA by regulating AS of TFIIIA [48]. Fig. 3b shows that the ER event identified in the third exon of Arabidopsis TFIIIA is conserved in *P. vulgaris* (Phvul.008G270400) and *G. max* (Glyma.02G293300). The absence or presence of the exon is clearly observed in both species and *G. max* presents additional transcript isoforms varying in the length of the RE (Fig. 3b). As in Arabidopsis, PTC were identified in the RE in both legume species, thus indicating the conservation of AS and of regulation/function of TFIIIA in the legume species analysed in this work.

Another example is the SCL33, a protein from the SR family that regulate splicing by binding to splicing regulatory elements -found in exons or introns-, facilitating spliceosome assembly and enhancing splicing [51]. The extensive AS of the third intron of Arabidopsis SCL33, that includes a RE with PTC, results in potential targets of NMD and these have been implicated in autoregulation of AS by SCL33 [52]. This AS event is conserved in other plants such as *Brachypodium distachyon* [27]. Fig. 3c shows that an RE event in the third intron is also conserved in *P. vulgaris* (Phvul.007G262600) and *G.max* (Glyma.09G104200) SCL33; the RE presents PTC in both legume species. We propose that the AS-related function of the legumes' SCL33 gene is similar to that known for Arabidopsis, regarding its targeting to NMD and the auto-regulation of SCL33 protein content.

Conservation of AS in UTR-CDS regions

Conserved AS events between *P. vulgaris* and *G. max* constitute a small proportion relative to all AS events in each species (Fig. 6). However, an assumption we made

was that conserved AS events may have a biological function. The percentage of intron/exons of UTR and CDS regions affected in conserved AS events from homologous genes as compared to all AS events were analyzed to explore potential function (Fig. 4).

Regarding AS events in 5'UTR-5'UTR junctions (Fig. 4a), AD, ASS and RE showed a higher percentage of conserved events as compared to all AS events in both plants. These data indicate that AD, ASS and RE events are preferentially conserved in 5'UTR introns, which could imply a conserved function, though such function is yet unknown. AS occurring in 5'UTR regions has been implicated in upstream open reading frames (uORF) of small proteins, that in turn have been implicated in the mRNA stability through NMD and in translation efficiency [53]. Nevertheless, the amino acid conservation among uORF from different organisms suggests a possible translation of small proteins and a possible function of these [53]. Proteomic studies would be required for further studies on the existence and function of such small proteins in legumes.

IR, however, exhibits a different pattern since the percentage of conserved events in 5'UTR-5'UTR junctions was lower than the percentage of all IR events (Fig. 4a). Nevertheless, in both species the percentage of IR was higher than the percentage of all events from CDS-CDS junctions (Fig. 4b), indicating a possible role for this type of event in CDS introns. In contrast, a reduction in the percentage of conservation of AD, ASS and RE in CDS introns was observed in both plants (Fig. 4b). For AA the percentage of conserved events in the 5'UTR and CDS introns followed a different pattern between *P. vulgaris* and *G. max* (Fig. 4a, b), showing that 5'UTR or CDS regions do not affect clearly the conservation of

45

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 11 of 17

AA events. Interestingly these five types of AS events that affect introns showed a reduction in percentages of 3'UTR-3'UTR junctions in conserved AS events as compared to all events in both species (Fig. 4c).

A similar analysis was done for ES and NI events that affect exons (Fig. 4d-i). Despite the differences in conserved NI event proportions between these two species, this AS was enriched in conserved AS events, for 5'UTR-CDS, CDS-3'UTR exons and single exon genes (Fig. 4g–i). This indicates potential functional relevance in those regions. Notably, the 3'UTR exons showed a reduction in the percentage of conserved events for ES and NI compared to all events as well as for the five AS events affecting introns (Fig. 4f).

Taken together the results of conserved AS events in CDS and UTR regions suggest that the potential of AS to affect either introns or exons is greater on regions upstream from the 3'UTR region.

An example of conserved AS in the 3'UTR region was identified for U2AF35, that is a component of the U2AF (U2 snRNP auxiliary factor) heterodimer, an essential pre-mRNA splicing factor. U2AF35 plays critical roles in the recognition of the 3′-splicing [54]. In addition human U2AF35 is implicated in the determination of mRNAs 3'UTR-length; mutated U2AF35 results in longer 3'UTR of certain genes [55]. AS in 3'UTR has been associated with the regulation of protein expression, by yet unidentified mechanisms [56]. Figure 3d shows the conserved 3'UTR AA event in *P. vulgaris* (Phvul.005G127700) and *G. max* (Glyma.12G181300) resulting in mature U2AF35 mRNAs varying in their 3'UTR length (Fig. 3d). This is another example of a gene with different primary transcript in both species; the primary transcript of one is similar to the alternative transcript isoform of the other (Fig. 3a, d). Additionally, a conserved IR event was validated in both species (Fig. 3d). Based in previous knowledge [55], we hypothesize that protein expression of U2AF35a in both legume species could be self-regulated through AS.
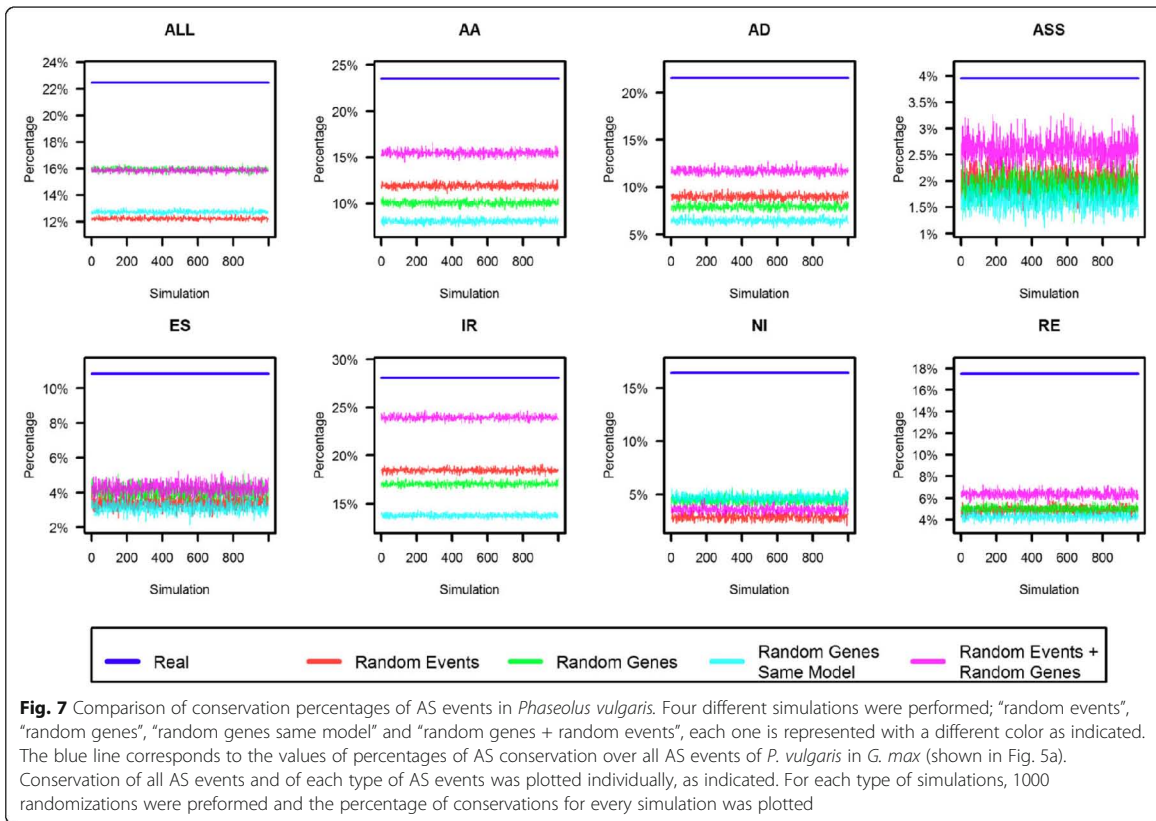
AS simulation

An important question in this work was if the 22 and 18% of AS conservation (Fig. 6) is a significant result from the *P. vulgaris* and *G. max* comparison, or if such values are just been obtained by chance. To answer this question four different simulations for AS event conservation were performed (Additional file 6). In the first simulation ("random events") all the non-redundant AS events in the expressed genes were shuffled, thus any intron/exon could be alternative spliced and the real homologous genes were maintained. In the second simulation ("random genes") the genes were shuffled to consider different homologs with their AS events maintained, the distribution of the number of exons per gene

(Additional file 7) was not considered. The third simulation ("random genes same model") was like the second, maintaining real AS events, but the exon distribution was that of the real homologs (Additional file 7). The objective of "random genes" and "random genes same model" simulations was to explore if the homology between *P. vulgaris* and *G. max* genes was important for the AS event conservation. The last simulation ("random genes + random event") where homologous genes and AS events were shuffled was a combination of the first and second simulations (Additional file 6).

The data on percentages of AS conservation (Fig. 6) were compared to the data obtained for each of the simulations; data of *P. vulgaris* in *G. max* are represented in Fig. 7 and *G. max* in *P. vulgaris* in Additional file 8. All simulations resulted in a value for the percentage of AS conservation that is lower than observed in our analysis for both overall and individual AS events, thus indicating that these results are neither random or artefacts (Fig. 6, Additional file 8).

The analysis of conservation of each type of AS derived from the simulations revealed interesting features such as a correlation of exon number with the conservation of AS events. For every type of AS event, except NI, the "random genes + random event" simulation showed the highest percentage of conserved AS (Fig. 7). One interpretation is that having more exons increases the probability of having AS and therefore the AS event conservation is also more likely. NI does not follow this rule as gene structure plays a major role for this type of AS event. This is consistent with the results of NI percentage in the UTR-CDS regions, where a high percentage was observed in the single exon genes (Fig. 4i), similar to a report by Marquez et al. [44]. For NI the highest percentage of conservation was observed in "random genes" and "random genes same model" simulations (Fig. 7). This suggests that higher NI occurrence is also related to specific genes.

The simulations also provided insights into the conservation of AS events. The position of the AS events within the gene tends to influence AS event conservation. If the positions of the AS conserved events within the gene were random, the percentage of conservation from "random events" and "random genes same model" would be similar. However, "random events" presented higher percentage of conservation in AA, AD and IR events (Fig. 7), something that could indicate that some homologous genes tend to present AS event conservation in specific introns or exons. The most common AS event types (AA, AD and IR) showed a bias for conserving AS events relative to position within the gene, as well as the number of exons, as described above. This tendency added to the exon number tendency were observed in the "random genes same model" simulation with the lowest conservation of AS events in AA, AD and
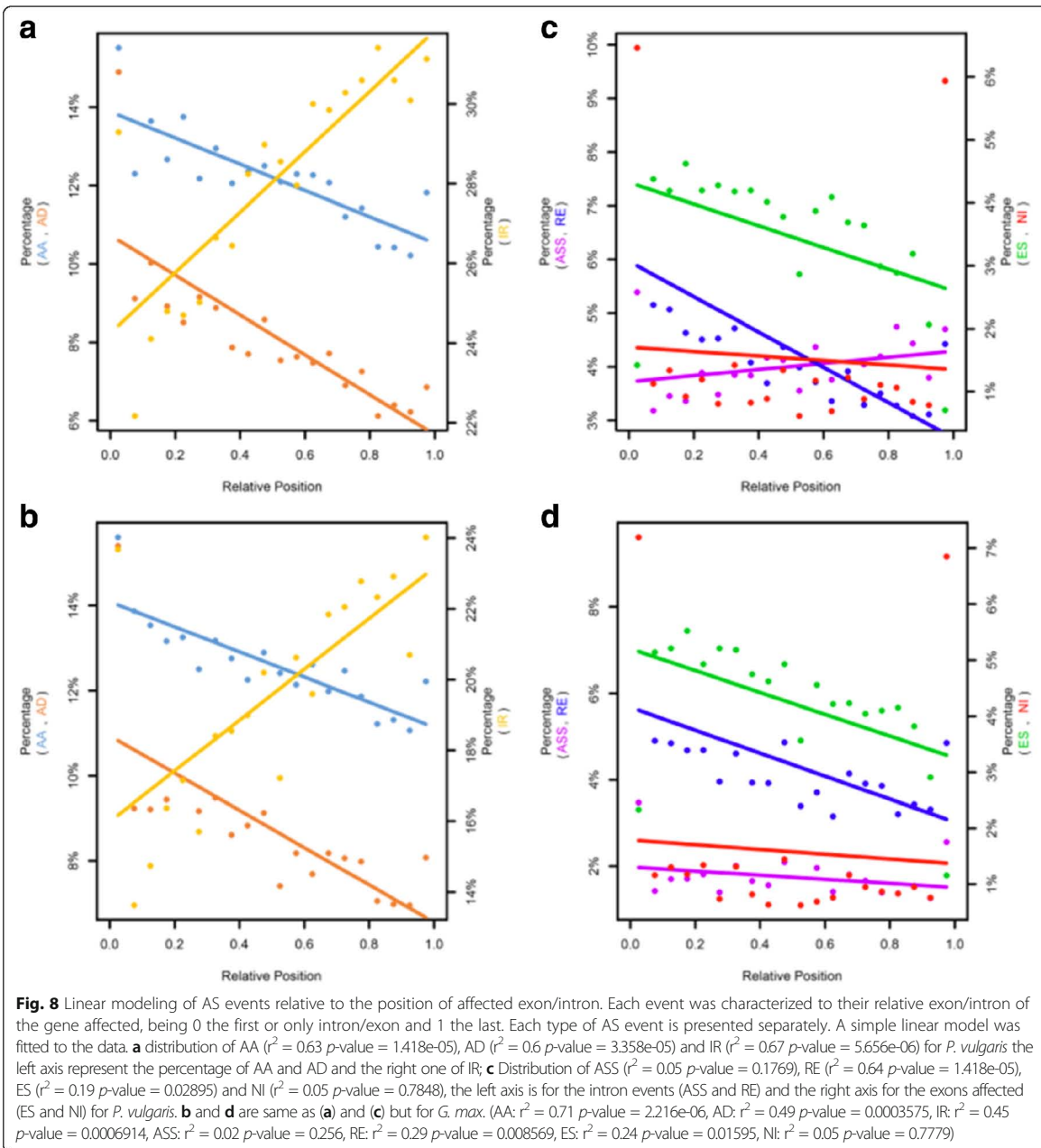
Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 12 of 17

Fig. 7 Comparison of conservation percentages of AS events in *Phaseolus vulgaris*. Four different simulations were performed; "random events", "random genes", "random genes same model" and "random genes + random events", each one is represented with a different color as indicated. The blue line corresponds to the values of percentages of AS conservation over all AS events of *P. vulgaris* in *G. max* (shown in Fig. 5a). Conservation of all AS events and of each type of AS events was plotted individually, as indicated. For each type of simulations, 1000 randomizations were preformed and the percentage of conservations for every simulation was plotted

IR (Fig. 7). A similar phenomena was observed for "random genes" in comparison with "random genes + random events" simulations, where besides having more exons the percentage of AS event conservation was lower in "random gene" simulation than when randomizing events ("random events + random genes") (Fig. 7, Additional file 7). This indicates that there is a bias for certain genes to conserve these types of events in certain introns. With these results, it is not possible to determine if RE, ASS and ES types have a bias in the position, as the percentages of conserved AS events in the simulations were similar due to the low number of non-redundant AS events (Fig. 7 and Additional file 8).

AS position within the gene
For an in-depth analysis of the position of AS event within a gene the affected introns/exons were catalogued with respect to their relative position in the gene, with the first or single exon/intron designated as 0, and the terminal exon designated as 1. The percentages of affected exons/introns for each relative position were calculated (Fig. 8). A linear regression was calculated for each type of event looking for a bias in the positions of

AS events within the genes. In agreement with the results from the simulations, AA, AD and IR presented a positional bias (Figs. 7, 8). AA and AD were enriched in initial introns, in contrast to IR where there was a bias for the terminal introns (Fig. 8a, b). RE and ES did not show a clearly position bias in the conservation studies, while this analysis uncovers a tendency of these events to affect initial introns in both plants (Fig. 8c, d). This position preference was similar to that seen for AA and AD. These results were consistent in both species (Fig. 8) and highlight the importance of the position of the AS event. The position bias of AS events within the gene observed for different AS events provide insights into how and in which gene positions each AS event is regulated.

It has been observed that the splicing occurs co-transcriptionally and sometimes it depends on the rate of the RNA-polymerase reaction [57]. The transcription rate also plays an important role in the formation of the secondary structure of the nascent RNA. The secondary RNA structure has been implicated in AS, there are some structures that prevent splicing site recognition and others that facilitate it [58]. The inhibition of splicing due to secondary RNA structure has been related

47

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 13 of 17

Fig. 8 Linear modeling of AS events relative to the position of affected exon/intron. Each event was characterized to their relative exon/intron of the gene affected, being 0 the first or only intron/exon and 1 the last. Each type of AS event is presented separately. A simple linear model was fitted to the data. **a** distribution of AA ($r^2 = 0.63$ *p*-value = 1.418e-05), AD ($r^2 = 0.6$ *p*-value = 3.358e-05) and IR ($r^2 = 0.67$ *p*-value = 5.656e-06) for *P. vulgaris* the left axis represent the percentage of AA and AD and the right one of IR; **c** Distribution of ASS ($r^2 = 0.05$ *p*-value = 0.1769), RE ($r^2 = 0.64$ *p*-value = 1.418e-05), ES ($r^2 = 0.19$ *p*-value = 0.02895) and NI ($r^2 = 0.05$ *p*-value = 0.7848), the left axis is for the intron events (ASS and RE) and the right axis for the exons affected (ES and NI) for *P. vulgaris*. **b** and **d** are same as (**a**) and (**c**) but for *G. max*. (AA: $r^2 = 0.71$ *p*-value = 2.216e-06, AD: $r^2 = 0.49$ *p*-value = 0.0003575, IR: $r^2 = 0.45$ *p*-value = 0.0006914, ASS: $r^2 = 0.02$ *p*-value = 0.256, RE: $r^2 = 0.29$ *p*-value = 0.008569, ES: $r^2 = 0.24$ *p*-value = 0.01595, NI: $r^2 = 0.05$ *p*-value = 0.7779)

with a competition between the structure and the site recognition of splicing factors. Therefore the splicing efficiency is directly correlated with the stability of the secondary RNA structure. On the other hand RNA structures could bring important splicing signals into closer proximity enhancing the splicing [59]. Based on this information we hypothesize that at transcription initiation, the nascent RNA still lacks a stable secondary structure thus facilitating the splicing site recognition and producing AS events such as AA, AD, ES or RE. IR events presented a tendency for terminal introns from already transcribed RNA with defined secondary RNA structures that could inhibit the recognition of splicing sites thus resulting in an IR. Nevertheless, these hypotheses need to be tested in order to better understand the regulation of the different AS events.

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 14 of 17

Conclusions

The algorithms and methodology used in this work allowed the identification and analysis of seven different types of AS events from two agronomically important legume species: *P. vulgaris* (common bean) and *G. max* (soybean). While the number of AS events was highly underestimated in their respective genome sequence annotations [33, 34], here it was shown that ~60% of all genes and ~70% of expressed genes from both species may undergo AS. Each type of AS event affected a different proportion of genes, with IR being the most frequent AS event followed by AA and AD in both plants.

The AS events were characterized in terms of the region they affected, and their relative position within the gene. The results of this characterization exposed different patterns for each of the AS events, such as preference for single exon genes from NI events or the contrasting result for position preferences between AA, AD and IR. These results were similar for both species, highlighting global aspects of these AS events in these legumes.

The conservation of AS events in two evolutionary related legume species was analysed considering those *P. vulgaris* genes with two (evolutionary relationship 1:2) or one (1:1) homologous gene in *G. max.* A significant proportion (ranging from 18 to 37%) of AS events were conserved between species. The conserved AS events are key to further research since they have been conserved for ~20 MY and they may provide insights into the functional role of these AS events in legumes. The conservation of AS events was experimentally validated for 8 selected genes, through RT-PCR analysis, something that enhances the reliability of bioinfomatic data from this work. The proposed function and biological significance of some of the validated conserved AS events was discussed, nevertheless they need to be further studied. The percentage of AS event conservation varies among the different AS types, with IR, AA and AD the most highly conserved in both species. Conserved events were also characterized in terms of the gene region they affect. The results threw a significant tendency to conserve events upstream 3'UTR regions, which indicates that events in 3'UTR are preferentially specie specific.

This work increases the knowledge of the yet almost unexplored process of AS. Tissue specific analysis, as well as isoform analyses, need to be performed to understand this relevant process for genome expression/function in eukaryotes.

Methods

RNA-seq libraries and expressed sequenced tags (ESTs)

All RNA-seq from *Phaseolus vulgaris* available (February 2016) in the Sequence Read Archive were downloaded and small RNA-seq were filtered. A total of 157 libraries belonging to 105 different samples were selected for the AS analysis in common bean. Eighty four libraries belonging to 77 different samples from *Glycine max* were selected based on project or tissue similarity with those from *P. vulgaris.* (Additional file 9). All available EST from NCBI and TGI database [60] from common bean and soybean were analysed. EST's from both databases were sequenced form multiple tissues and conditions. In total 176,782 and 1,461,723 EST sequences from common bean and soybean, respectively, were analysed.

Mapping and AS annotation

All RNA-seq libraries were mapped to their respective genome without gene annotation information, alignments must be unique and perfect. For the mapping two different approaches were used; a seed and extend approach, where reads are sliced into short seeds, which are mapped to the genome, allowing the identification of splicing sites; and an exon first approach, where complete reads are mapped at first and based on that mapping an exon-exon junctions database is created in order to align the unmapped reads later [61]. TopHat2 [62] was used for the exon first approach and gsnap [63] for the seed extended (Fig. 2). Each mapping result, one for gsnap and one for TopHat2, was the input for a gene prediction modelling performed with Cufflinks [64], this was carried out also without any gene annotation information as well (Fig. 2). EST's were mapped to their respective genome with Gmap [63] (Fig. 2). The gene models were filtered to avoid chimeras based on the coordinates of the primary transcript of each gene in the genome annotation. Models in zones where genes overlap or models that are part of two or more genes were removed. Each gene model from each mapping algorithm of each RNA-seq library was compared to their corresponding annotated primary transcript in the genome with an in-house perl script (Fig. 2). This algorithm identifies the seven different AS events (AA, AD, ASS, ES, IR, NI and RE) by comparing the genome coordinates of the out coming gene models to all primary transcripts (Fig. 1). AS events present in EST's, or in both mapping results for a particular library, or in at least half of the sample replicates from a type of mapping were selected for further analysis (Fig. 2 and Additional file 2).

Gene expression

The expression value for each gene was calculated using all FPKM values from each model that belong to each gene (Fig. 2). The FPKM from each model was multiplied by the length of the model, the sum of all products from each gene was then divided by the length of the primary transcript resulting in a normalized value of expression per mapping algorithm. The mean of both normalized expression value, one for gsnap and other for TopHat2, was the library expression. The sample expression was

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 15 of 17

calculated by the median of the replicates libraries expressions values. A gene was considered expressed in a tissue if the sample expression was above one.

RT-PCR analysis

For RNA isolation surface-sterilized seeds from *P. vulgaris* and *G. max* were germinated over moist paper, in sterile conditions, for 2 days. The root and cotyledonary leaves from germinated seedlings were cut, frozen in liquid nitrogen and stored separately at −80 C until used. Total RNA was isolated from 200 to 400 mg frozen tissue using Trizol reagent (Life Technologies, California, USA), as reported [65]. Absence of genomic DNA contamination was subsequently confirmed for each sample by PCR amplification using primers for the ACR9 (ACT-domain containing protein) gene (Phvul.008G013100, Glyma.18G289000). To validate the presence of different transcript isoforms identified through bioinformatics analysis, two-step RT-PCR was performed following the manufacturer's directions (Thermo Scientific, USA) using poly-thymine deoxynucleotide primer. Eight genes with from *P. vulgaris* and their corresponding *G. max* homologs were selected for AS events validation and the ACR9 gene that did not present AS was included as a control. For each selected gene, a pair of oligonucleotide primers was designed to amplify products specific for the primary transcript or for transcript isoforms derived from AS events; primer sequences as well as genes IDs and annotation are shown in Additional file 10. For RT-PCR reactions the thermocycler was set to: 60 / 68 °C for annealing / extension and 35−40 cycles and a High Fidelity DNA Polymerase (Jena Bioscience, Germany) was used. Amplification products were resolved in a 3% agarose gel in 1xTAE and EtBr stained for visualization.

AS simulation

Simulation of AS were performed based on number of exons/introns of expressed genes. Four different types of simulations were performed: "random events", "random genes", "random genes same model" and "random genes + random event". One thousand independent simulations were performed for each type of simulation.

AS event conservation

Phaseolus vulgaris and *Glycine max* orthologous genes were pulled out from Schmutz et al. [34]. There were 13,962 common bean genes with two orthologous genes in soybean, resulting from the recent whole genome duplication in this legume, and 5624 orthologs with only one copy in soybean. The AS conservation was based on junction conservation and not in position conservation. Due to this fact, genes should have the same gene model (same number of exons) and been expressed in at least one sample (7´692 common bean genes with two orthologous genes in soybean and 7´039 with one gene).

Additional files

Additional file 1: AS events reported in the *Phaseolus vulgaris* [34] and *Glycine max* [33] annotated genomes. (XLSX 11 kb)

Additional file 2: AS event. (XLSX 8631 kb)

Additional file 3: Splicing sites. Percentage of splicing sites motifs reported in the genome (inner circle), in the new junctions (middle cirlce) and the genome with the new junctions (outer circle). U2 motifs (gray), U12 motifs (black) and non-canonical splicing sites (striped). Panel a show the results from *P. vulgaris* while b from *G. max*. (TIFF 816 kb)

Additional file 4: Introns and exons from CDS and UTR regions affected by AS events. Percentage of *P. vulgaris* and *G. max* introns (a) and exons (b) affected by AS compared to their total proportion in each genome. Proportions of common bean as well as soybean are plotted. (TIFF 159 kb)

Additional file 5: Percentage of non-redundant AS event types in homologous genes. (XLSX 11 kb)

Additional file 6: Four AS conservation simulations. Four different simulations for AS event conservation percentage testing were performed. "random events": randomize the AS events in the expressed genes maintaining homologous genes; "random genes": randomize homologous genes, the gene model was not taken into account but the events remained as the real data; "random genes same model": same as "random genes" but the gene model stays equal and "random events + random genes": AS events as well as homologous genes, ignoring real gene models, were randomized. (TIFF 232 kb)

Additional file 7: Exon distribution. Proportions of number of exons per gene in the annotated *P. vulgaris* and *G. max* genomes, homologous genes with an evolutionary relationship of with evolutionary relationship 1:2 and 1:1 and pseudo-homologous genes resulted from "random genes" simulation. (TIFF 115 kb)

Additional file 8: Comparison of conservation percentages of AS events in *Glycine max*. Data from each performed simulation are plotted with a different color while the blue line corresponds to the values of percentage of AS conservation shown in Fig. 6b. The percentage of AS conservation of *G. max* in *P. vulgaris* considered over all AS events in soybean were analyzed. For description of each plot see legend to Fig. 7. (TIFF 136 kb)

Additional file 9: Sample used for the AS event identification. (XLSX 41 kb)

Additional file 10: *P. vulgaris* and homologous *G. max* genes selected for RT-PCR analysis. (XLSX 9 kb)

Abbreviations
AA: Alternative acceptor; AD: Alternative donor; AS: Alternative splicing; ASS: Alternative splicing sites; CDS: Coding DNA sequences; DN: Nonsynonymous mutations; DS: Synonymous mutations; ES: Exon skipping; EST: Expressed sequence tags; IR: Intron retention; MYA: Millions years ago; NI: New intron; NMD: Nonsense-mediated decay; RE: Retained exon; RNA-seq: Next generation RNA-sequencing; snRNA: Small nuclear RNAs; SR: Serine/arginine-rich proteins; uORF: Upstream open reading frame; UTR: Untranslated region; WGD: Whole genome duplication

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 16 of 17

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and its additional files.

Authors' contributions
LPI, WBB and GH designed the research and discussed the data. LPI performed the bioinformatic research and analyzed the data. MR designed and performed the experimental research. LPI and GH wrote the manuscript. All authors have read and approved the manuscript for publication.

Ethics approval and consent to participate
Plant materials were developed by Mario Ramírez, plants were grown in greenhouses from the Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México at Cuernavaca, México. The seed samples from *Phaseolus vulgaris* cv. negro jamapa and from *Glycine max* cv. amsoy are publicly available for research institutions from Mexico and abroad upon reasonable request.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, Mexico. [2]Department of Biology, University of Florida, Gainesville, FL, USA.

References
1. Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. Mol Cell. 2003;12:5–14.
2. Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? BioEssays. 2003;25:1147–9.
3. Rogozin IB, Carmel L, Csuros M, Koonin EV. Origin and evolution of spliceosomal introns. Biol Direct Biology Direct. 2012;7:1.
4. Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol. 2014;6
5. Jean-Philippe J, Paz S, Caputi M. hnRNP A1: the Swiss Army knife of gene expression. Int J Mol Sci. 2013;14:18999–9024.
6. Simpson CG, Manthri S, Raczynska KD, Kalyna M, Lewandowska D, Kusenda B, et al. Regulation of plant gene expression by alternative splicing. Biochem Soc Trans. 2010;38:667–71.
7. Reddy ASN. Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol. 2007;58:267–94.
8. Zhiguo E, Wang L, Zhou J. Splicing and alternative splicing in rice and humans. BMB Rep. 2013;46:439–47.
9. Petrillo E, Godoy Herz MA, Barta A, Kalyna M, Kornblihtt AR. Let there be light: regulation of gene expression in plants. RNA Biol. 2014;11:1215–20.
10. Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, et al. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res. 2012;40:2454–69.
11. Boothby TC, Zipper RS, Van der Weele CM, Wolniak SM. Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea Vestita*. Dev Cell Elsevier Inc. 2013;24:517–29.
12. Yang X, Zhang H, Li L. Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis. Plant J. 2012;70:421–31.
13. Remy E, Cabrito TR, Baster P, Batista RA, Teixeira MC, Friml J, et al. A major facilitator superfamily transporter plays a dual role in polar auxin transport and drought stress tolerance in Arabidopsis. Plant Cell. 2013;25:901–26.
14. Hedges SB, Blair JE, Venturi ML, Shoe JL. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol. 2004;4:2.
15. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40:1413–5.
16. Ramani A, Calarco J, Pan Q. Genome-wide analysis of alternative splicing in *Caenorhabditis Elegans*. Genome Res. 2011;21:342–8.
17. Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. Mol Biol Evol. 2014;31:1402–13.
18. Ner-Gaon H, Leviatan N, Rubin E, Fluhr R. Comparative cross-species alternative splicing in plants. Plant Physiol. 2007;144:1632–41.
19. Zhou R, Moshgabadi N, Adams KL. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proc Natl Acad Sci. 2011;108:16122–7.
20. Marquez Y, Brown JWS, Simpson C, Barta A, Kalyna M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res. 2012;22:1184–95.
21. Li Q, Xiao G, Zhu YX. Single-nucleotide resolution mapping of the *Gossypium raimondii* transcriptome reveals a new mechanism for alternative splicing of introns. Mol Plant. 2014;2014(7):829–40. The Authors
22. Panahi B, Abbaszadeh B, Taghizadeghan M, Ebrahimie E. Genome-wide survey of alternative splicing in *Sorghum bicolor*. Physiol Mol Biol Plants. 2014;20:323–9.
23. Shen Y, Zhou Z, Wang Z, Li W, Fang C, Wu M, et al. Global dissection of alternative splicing in Paleopolyploid soybean. Plant Cell. 2014;26:996–1008.
24. Thatcher SR, Zhou W, Leonard A, Wang B-B, Beatty M, Zastrow-Hayes G, et al. Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. Plant Cell Online. 2014;26:3472–87.
25. Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, D'Angelo M, et al. A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biol. 2014;14:99.
26. Xu P, Kong Y, Song D, Huang C, Li X, Li L. Conservation and functional influence of alternative splicing in wood formation of Populus and eucalyptus. BMC Genomics. 2014;15:780.
27. Mandadi KK, Scholthof K-BG. Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in Brachypodium Distachyon. Plant Cell. 2015;27:71–85.
28. McGuire AM, Pearson MD, Neafsey DE, Galagan JE. Cross-kingdom patterns of alternative splicing and splice recognition. Genome Biol. 2008;9:R50.
29. Ast G. How did alternative splicing evolve? Nat Rev Genet. 2004;5:773–82.
30. Reddy ASN, Marquez Y, Kalyna M, Barta A. Complexity of the alternative splicing landscape in plants. Plant Cell. 2013;25:3657–83.
31. Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, et al. Molecular analysis of legume nodule development and autoregulation. J Integr Plant Biol. 2010;52:61–76.
32. Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P. Beans (*Phaseolus* spp.) - model food legumes. Plant Soil. 2003;252:55–128.
33. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature Nature Publishing Group. 2010;463:178–83.
34. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet Nature Publishing Group. 2014;46:707–13.
35. Vlasova A, Capella-Gutiérrez S, Rendón-Anaya M, Hernández-Oñate M, Minoche AE, Erb I, et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. Genome Biol Genome Biology. 2016;17:1–18.
36. Zhang C, Yang H, Yang H. Evolutionary character of alternative splicing in plants. Bioinform Biol Insights. 2015;9:47–52.
37. Simpson CG, Lewandowska D, Fuller J, Maronova M, Kalyna M, Davidson D, et al. Alternative splicing in plants. Biochem Soc Trans. 2008;36:508–10.
38. Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. Nat. Rev. Genet. Nature Publishing Group. 2016;17:407–21.

Iñiguez *et al. BMC Genomics* (2017) 18:650

Page 17 of 17

39. Florea L, Song L, Salzberg SL. Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Research. 2013;2:188.

40. Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, et al. Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. Nat Genet. 2004;36:1255–7.

41. Iida K, Shionyu M, Suso Y. Alternative splicing at NAGNAG acceptor sites shares common properties in land plants and mammals. Mol Biol Evol. 2008;25:709–18.

42. Schindler S, Szafranski K, Hiller M, Ali GS, Palusa SG, Backofen R, et al. Alternative splicing at NAGNAG acceptors in *Arabidopsis thaliana* SR and SR-related protein-coding genes. BMC Genomics. 2008;9:159.

43. Bradley RK, Merkin J, Lambert NJ, Burge CB. Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. PLoS Biol. 2012;10

44. Marquez Y, Höpfler M, Ayatollahi Z, Barta A, Kalyna M. Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. Genome Res. 2015;25:995–1007.

45. Dessau M, Halimi Y, Erez T, Chomsky-Hecht O, Chamovitz DA, Hirsch JA. The Arabidopsis COP9 Signalosome subunit 7 is a model PCI domain protein with subdomains involved in COP9 Signalosome assembly. Plant Cell Online. 2008;20:2815–34.

46. Chamala S, Feng G, Chavarro C, Barbazuk WB. Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. Front Bioeng Biotechnol. 2015;3:33.

47. Fu Y, Bannach O, Chen H, Teune JH, Schmitz A, Steger G, et al. Alternative splicing of anciently exonized 5S rRNA regulates plant transcription factor TFIIIA. Genome Res. 2009;19:913–21.

48. Hammond MC, Wachter A, Breaker RR. A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. Nat Struct Mol Biol. 2009;16:541–9.

49. Layat E, Cotterell S, Vaillant I, Yukawa Y, Tutois S, Tourmente S. Transcript levels, alternative splicing and proteolytic cleavage of TFIIIA control 5S rRNA accumulation during *Arabidopsis thaliana* development. Plant J. 2012;71:35–44.

50. Barbazuk WB. A conserved alternative splicing event in plants reveals an ancient exonization of 5S rRNA that regulates TFIIIA. RNA Biol. 2010;7:397–402.

51. Thomas J, Palusa SG, Prasad KVSK, Ali GS, Surabhi GK, Ben-Hur A, et al. Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33 pre-mRNA. Plant J. 2012;72:935–46.

52. Palusa SG, Reddy ASN. Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. New Phytol. 2010;185:83–9.

53. Hu Q, Merchante C, Stepanova A, Alonso J, Heber S. Genome-wide search for translated upstream open reading frames in *Arabidopsis thaliana*. IEEE Trans Nanobioscience. 2016;1241:1–1.

54. Pacheco TR, Gomes AQ, Barbosa-Morais NL, Benes V, Ansorge W, Wollerton M, et al. Diversity of vertebrate splicing factor U2AF35: identification of alternatively spliced U2AF1 mRNAs. J Biol Chem. 2004;279:27039–49.

55. Park SM, Ou J, Chamberlain L, Simone TM, Yang H, Virbasius CM, et al. U2AF35(S34F) promotes transformation by directing aberrant ATG7 pre-mRNA 3′ end formation. Mol Cell Elsevier Inc. 2016;62:479–90.

56. Mockenhaupt S, Makeyev EV. Non-coding functions of alternative pre-mRNA splicing in development. Semin Cell Dev Biol Academic Press. 2015;47–48:32–9.

57. Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet Nature Publishing Group. 2014;15:163–75.

58. Warf MB, Berglund JA. The role of RNA structure in regulating pre-mRNA splicing. Trends Biochem Sci. 2010;35:169–78.

59. Jin Y, Yang Y, Zhang P. New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. RNA Biol. 2011;8:450–7.

60. Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, et al. The TIGR gene indices: clustering and assembling EST and know genes and integration with eukaryotic genomes. Nucleic Acids Res. 2005;33:71–4.

61. Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods Nature Publishing Group. 2011;8:469–77.

62. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol BioMed Central Ltd. 2013;14:R36.

63. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010;26:873–81.

64. Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nat. Biotechnol. 2011;28:511–5.

65. Ramirez M, Graham MA, Blanco-Lopez L, Silvente S, Medrano-Soto A, Blair MW, et al. Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol. 2005;137:1211–27.

66. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, et al. IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics. 2015;31:3359–61.

5.2. El Papel del Splicing Alternativo en la Duplicación Génica

El Splicing Alternativo es un tema en constante desarrollo, durante los últimos años se ha incrementado la cantidad de información disponible sobre este tipo de regulación post-transcripcional. Cada vez son más los trabajos que prueban la importancia de este tema y las implicaciones funcionales de este proceso dentro de diversos organismos. Por otra parte la duplicación génica, aunado a mutaciones en los genes duplicados, es uno de los procesos por el cual los organismos adquieren nuevos genes. La evolución de estos dos aspectos es de mucho interés ya que a partir de entender su historia evolutiva se podrá comprender un poco más sobre los procesos y fuerzas que llevan hacia los distintos destinos de los genes duplicados. La conservación y divergencia del Splicing Alternativo se ha estudiado en diferentes especies y esta sección se enfocará en el splicing alternativo y su relación con las duplicaciones génicas.

5.2.1. Introducción

La diversidad proteica que existe en los organismos hoy en día es el resultado de varios procesos evolutivos. Se conoce que la duplicación génica en conjunto con la acumulación de mutaciones son responsables, junto con otros factores, del incremento en el número de diferentes proteínas. Otro de los factores que promueven la diversidad proteica es el AS, discutido en la sección anterior, por el cual se pueden transcribir diferentes isoformas de mRNA a partir de un solo gene. Estos dos mecanismos por los cuales se incremente la diversidad de proteínas varían enormemente y pueden influir en el otro. Una pregunta importante es cómo se ven afectados los patrones de AS una vez que los genes se hayan duplicado. La relación evolutiva entre la duplicación génica y el AS ha sido un tema de interés y discusión por lo que decidimos publicar una revisión del tema en el siguiente artículo.

The Evolutionary Relationship between Alternative Splicing and Gene Duplication

Luis P. Iñiguez* and Georgina Hernández*

Programa de Genómica Funcional de Eucariotes, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, México

The protein diversity that exists today has resulted from various evolutionary processes. It is well known that gene duplication (GD) along with the accumulation of mutations are responsible, among other factors, for an increase in the number of different proteins. The gene structure in eukaryotes requires the removal of non-coding sequences, introns, to produce mature mRNAs. This process, known as *cis*-splicing, referred to here as splicing, is regulated by several factors which can lead to numerous splicing arrangements, commonly designated as alternative splicing (AS). AS, producing several transcripts isoforms form a single gene, also increases the protein diversity. However, the evolution and manner for increasing protein variation differs between AS and GD. An important question is how are patterns of AS affected after a GD event. Here, we review the current knowledge of AS and GD, focusing on their evolutionary relationship. These two processes are now considered the main contributors to the increasing protein diversity and therefore their relationship is a relevant, yet understudied, area of evolutionary study.

Keywords: alternative splicing, evolution, gene duplication, polyploidy, mRNA isoforms

INTRODUCTION

Organismal protein diversity has increased through evolution. This diversity has allowed organisms to adapt to different environments, to develop and to differentiate specialized tissues. GD, along with mutations, has been an important process to increase the protein diversity. The genome sequence of a variety of organisms has allowed the identification of paralogous genes produced by GD. Nevertheless, the vast amount of proteins cannot be solely explained by this process. Before the human genome was sequenced it was widely believed that the human genome would encode around 100,000 genes based on an estimate of the number of different proteins existing in human cells (Fields et al., 1994). Surprisingly only ~21,000 coding genes were annotated in the genome sequence (Auton et al., 2015). Almost all protein coding genes from eukaryotes contain non-coding sequences, known as introns, flanked by coding sequences, or exons. During transcription introns need to be removed in order to form a mature mRNA, this process is

Abbreviations: AA, alternative acceptor; AD, alternative donor; AS, alternative splicing; ES, exon skipping; GD, gene duplication; IR, intron retention; MEE, mutually exclusive exon; SSD, short segmental duplication; WGD, whole genome duplication.

called splicing. Gilbert (1978) proposed that alterations on the splicing of a gene could form multiple isoforms from a single gene. AS is the process through which a single gene can produce different mRNA isoforms and those in turn, if translated, could lead to multiple proteins. The AS process could explain the discrepancy between the estimates of number of genes and number of proteins.

GENE DUPLICATION

The vast number of genes in eukaryotic organisms is in large part due to GD. Several processes can occur in the cell that can lead to the duplication of genes. There are two GD classifications; SSD and WGD. SSD duplicates one or several genes while WGD increases significantly the offspring's gene count compared to the parent. SSDs events may result from unequal cross-over (DNA-dependent) or retrotransposition (RNA-dependent). Both processes give rise to different gene structures. Retrotransposed genes become single exon genes while DNA-dependent duplications inherit gene structure and regulatory sequences (reviewed by Van de Peer et al., 2009). A WGD is caused by autopolyploidy or allopolyploidy (Van de Peer et al., 2009). It has been hypothesized that WGDs provide organisms with certain defenses against extinction, because individuals can accumulate mutations in duplicated genes that may enhance their adaptation to stress or environmental conditions (Innan and Kondrashov, 2010).

Four models have been proposed to describe the evolution of duplicated, or paralogous genes. The model of neofunctionalization establish that one copy of the gene retains the ancestral function while the function of the other diverges into a new one. Subfunctionalization, or duplication-divergence-complementation propose that the ancestral gene function is partitioned between paralogs. Subfunctionalization was tested by Kito et al. (2016) in several yeast species, some of which contained several SSD. Using proteomics, they found that in species that experienced SSD, the sum of paralogous proteins was similar to the amount of the non-duplicated homologous proteins in other yeast species. Another evolutionary model is that one paralog retains the ancestral function while the other paralog devolves into a pseudogene (Panchy et al., 2016). Finally, the function of both paralogs can remain similar if an increased production of the protein is advantageous or if a dosage balance occurs in conjunction with other gene products (Innan and Kondrashov, 2010; Magadum et al., 2013). Wang et al. (2011, 2012) studied the spatiotemporal expression profiles of duplicated genes to identify the different evolutionary models based on the gene expression profiles of paralogous genes. They concluded that the divergence in expression depends on the process of duplication. To complement these studies, Lan and Pritchard (2016) analyzed SSD and found that the genomic distance, the type of duplication and the time since duplication all influence the fate of paralogous genes. These evolutionary models have been studied and reviewed by

Cañestro et al. (2013) where examples of each model are described.
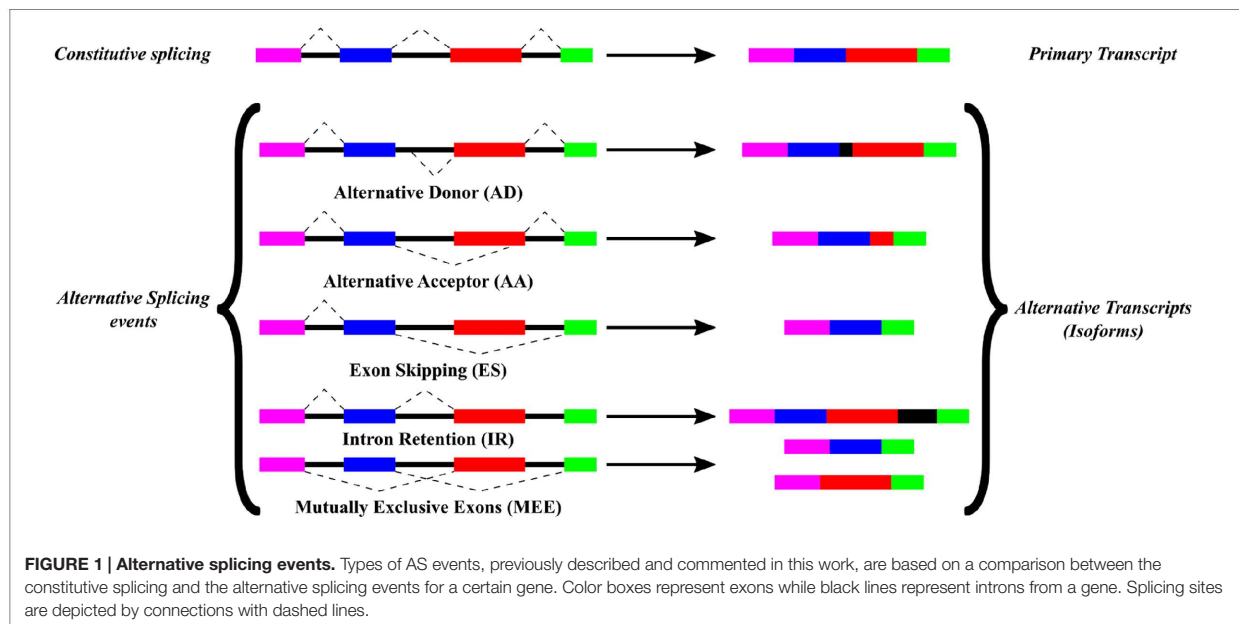
ALTERNATIVE SPLICING

Alternative splicing is a post-transcriptional process that occurs in different stresses, developmental conditions and in different cell types (reviewed by Wang et al., 2008; Staiger and Brown, 2013; Li et al., 2016). AS affects the localization of the mature mRNA and its translation efficiency (Zhiguo et al., 2013). It can also produce alternative stop codons and can regulate protein expression by non-sense-mediated decay (Kalyna et al., 2012). AS may result in different protein isoforms derived from a single gene and these isoforms alter their cellular localization or function with respect to the primary transcript (Remy et al., 2013). AS can also influence in protein–protein interactions. These interactions have been associated with intrinsically disordered protein domains, which are more susceptible to AS (Niklas et al., 2015; Yang et al., 2016). Five main AS events lead to the production of different isoform transcripts as shown in **Figure 1**: ES, where a complete exon is absent from the primary transcript; AA, where the 3′ end of an intron changes; AD, where the 5′ splice site of the intron is different; IR, where a reported intron is not spliced and is part of the mature mRNA and MEE, where one of two exons is retained in a given isoform but not both exons. These AS events vary in their frequency among different eukaryotic organisms. In animals, ES is the most common AS event, which represents around 50% of all AS events, while in plants IR is the most frequent AS event (reviewed Vélez-Bermúdez and Schmidt, 2014; Zhang et al., 2015).

Nematodes and humans show high variation in cell types and both genomes code for a similar number of genes. However, 98% of human multiple-exon genes exhibit AS (Pan et al., 2008) while AS is present in only 25% of nematode genes (Ramani et al., 2011). Chen et al. (2014) analyzed several organisms which vary in their number of different cell types –referred here as organism complexity- and found a strong positive correlation between the number of cell types and the number of AS events. Organisms with more complexity tend to have higher AS. In addition to AS, it has been found that non-coding RNA's have a correlation with organismal complexity (Liu et al., 2013). There is evidence that proteome size, structural disorder of proteins, protein–protein interactions and AS are all part of a fine tuning of a complex network to ensure organism complexity (Schad et al., 2011; Dunker et al., 2015).

THE EVOLUTION OF AS IS CLOSELY LINKED TO GD EVENTS

Transcript isoforms resulting from AS events can be viewed as having "internal-paralogs" in the same gene (Kopelman et al., 2005). These "internal-paralogs" may have different functions, similar to the neofunctionalization model of gene evolution. For these reasons the comprehension and analysis of the relationship between AS and GD is an interesting topic in the evolutionary
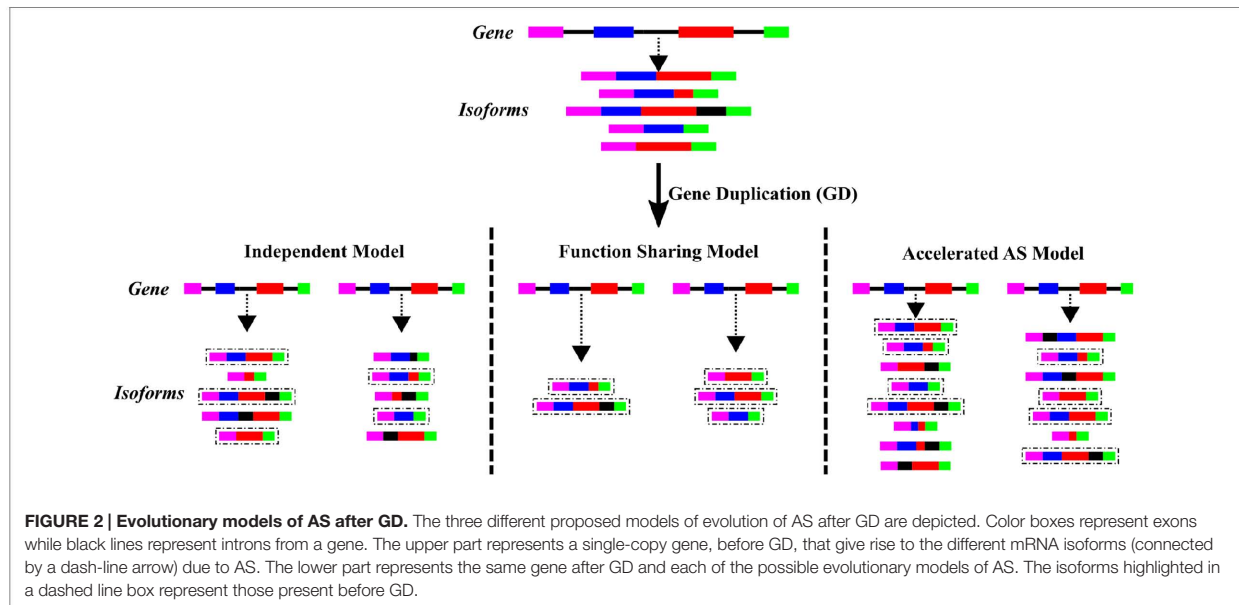
55

FIGURE 1 | **Alternative splicing events.** Types of AS events, previously described and commented in this work, are based on a comparison between the constitutive splicing and the alternative splicing events for a certain gene. Color boxes represent exons while black lines represent introns from a gene. Splicing sites are depicted by connections with dashed lines.

field. It has been observed that mutating a single intronic nucleotide can provoke changes in gene splicing patters (Hsiao et al., 2016), which would facilitate a fast evolution on patterns of AS in paralogus genes. Reddy et al. (2013) reviewed three different models to explain this relationship, these models are summarized in **Figure 2**. The independent model establishes the lack of relationship between GD and AS and therefore the number of isoforms in paralogs vs. non-duplicated genes would be similar. The functional sharing model illustrates the subfunctionalization of the paralogous genes, where one paralog would adopt certain number of ancestral AS events and the other paralog adopts the rest of the ancient isoforms. Therefore, in the functional sharing model the number of AS events per gene decreases in comparison with the non-duplicated genes. The last model is the accelerated AS model, where an increase in the number of AS events per gene results from a relaxed selection pressure for each paralogous gene.

The most common AS event analyzed in paralogous genes is MEE. This AS event accounts for the particularity where given the possibility of two exons, one is maintained in one duplicated gene and lost in the other, and vice versa in the paralog. The first example found to have this pattern was the microphthalmia-associated transcription factor in *Danio rerio* (Altschmied et al., 2002). This is a single copy gene with and at least two mRNA isoforms in mammals. The isoforms from this gene vary in the 3′-end of the mature mRNA and are expressed in different tissues. Altschmied et al. (2002) analyzed this gene in zebrafish, a species from the teleost which have presented several GD events in their evolutionary history. They found two paralogs, one that had one exon while the paralog had the other exon. Both paralogs were expressed in different tissues, thus confirming that GD had replaced AS. Several reports have confirmed the model of function sharing for MEE events in a few genes (Wei-Ping et al.,

2003; Pacheco et al., 2004; Cusack and Wolfe, 2007; Hultman et al., 2007; Marshall et al., 2013). To generalize this model, Abascal et al. (2015) analyzed ∼90 human genes exhibiting MEE events. They identified the duplicated orthologous genes in five different fish species, including zebrafish. While several cases were identified that fit the function sharing model, one paralog containing one exon, while the other paralog contained the other exon, not all orthologous genes fit the function sharing model. In this report they also found duplicated genes with the same MEE event observed in humans. This report suggested that a single model could not be generalized, instead each gene possesses a unique AS evolutionary model.

The first attempt to generalize an evolutionary model for AS after GD was proposed by Kopelman et al. (2005). They divided the human genes in families depending on the number of paralogs and determined the proportion of genes affected by AS. They reported that larger gene families tend to have fewer genes affected by AS in comparison with single-copy genes (singletons). They also searched for homologous genes in mouse and classified them into duplicated and non-duplicated genes. They found that duplicated genes were less affected by AS in both organisms. These results were further confirmed by Su et al. (2006), who proposed the function sharing model as the main evolutionary model of AS after GD. In this study they report that AS events are lost in recent gene duplicates but novel AS events are gained in ancient paralogs. They also investigated the symmetry of AS in paralogs, where the number of AS events in the two paralogs are equal. They argued that 16 to 34% of the paralogous genes analyzed exhibit patterns of asymmetric AS. Jin et al. (2008) analyzed the proportion of genes affected by AS and the average number of AS events per family. They found that larger families have a smaller proportion of genes affected by AS and that the number of AS events per gene is lower in

FIGURE 2 | Evolutionary models of AS after GD. The three different proposed models of evolution of AS after GD are depicted. Color boxes represent exons while black lines represent introns from a gene. The upper part represents a single-copy gene, before GD, that give rise to the different mRNA isoforms (connected by a dash-line arrow) due to AS. The lower part represents the same gene after GD and each of the possible evolutionary models of AS. The isoforms highlighted in a dashed line box represent those present before GD.

larger gene families. Another important aspect to consider in the relationship between GD and AS is the expression patterns of the paralogous genes and their isoforms. Talavera et al. (2007) studied the correlation of expression of paralogs with and without AS events in a set of tissues. They found that duplicated genes without AS have more similar expression patterns compared to the expression pattern between duplicated genes with AS. This result suggests a relationship between gene expression patterns and AS in duplicated genes. Nevertheless, it is not clear if AS controls gene expression or vice versa.

Hughes and Friedman (2008) analysis of the AS and GD relationship in *Caenorhabditis elegans*, confirmed findings from previous analyses in mouse and human indicating that larger gene families had fewer AS events. However, these authors focused their work on an interaction network that consisted of ∼900 genes; they found a negative correlation between AS events and the connectivity in the network. This means that a gene with multiple connections to other genes, also known as a hub, had fewer AS events than genes connected to one or two genes. Further, they argued that duplication of hub genes was an uncommon phenomenon. Therefore AS, interpreted as an internal paralog, was unlikely to occur in hub genes.

Plants are organisms with a tendency to have undergone WGDs throughout their evolutionary history; and are therefore particularly important in the study of AS and GD. Lin et al. (2008) studied families of paralogous genes in *Arabidopsis thaliana* and *Oryza sativa*. They analyzed several factors and characteristics of singletons vs. duplicated genes and found that singletons were less affected by AS than paralogs in both plant species; which contrasts with findings from animal studies. In agreement with these results, Roux and Robinson-Rechavi (2011) found that paralogous families containing exactly two members had more AS events and a higher proportion of these genes are affected

by AS than the rest of the other gene families in humans. This is also in agreement with previous reports indicating that larger families have fewer AS events and are less affected than singletons genes. They classified the duplicated genes according to their time of appearance in the evolutionary history, observing a positive correlation between the number of AS events and the time since duplication indicating that duplicated genes acquire AS events over time. They also observed a negative correlation between selective pressure and AS, meaning that paralogous genes under strong positive selection tend to have fewer AS events than paralogs under weak selection. They also searched for orthologous genes and AS events in mouse and found that genes duplicated in human but not in mouse had fewer AS events than genes that have not undergone duplication. They argue that the model best explains AS after GD and, based on the comparison with mouse, genes with fewer AS events tend to duplicate more frequently. Chen et al. (2011) analyzed duplicated genes in human and mouse and observed, based on protein similarity, that the time of duplication is positively correlated with AS event acquisition. This accounts for why ancient paralogs tend to have more AS than recently duplicated genes, which is consistent with the findings described above for plants.

The two contrastingly AS evolutionary models, function sharing and independent, were discussed more recently by Su and Gu (2012). They argued against the acquisition of AS through time and a predisposition of duplication based on their AS events mentioned by Roux and Robinson-Rechavi (2011). These authors also argued the AS evolution after GD is most influenced by the different types of GD. Thus, GD arising from SSD tend to accumulate more mutations than WGD. Such mutations could be a replacement of AS events and therefore AS events are lost quickly. Tack et al. (2014) studied different WGD and tandem duplications in *A. thaliana* looking for the qualitative

and quantitative conservation of AS events in paralogs. Such conservation was higher in paralogous genes resulting from tandem duplications than from WGD-resulting paralogs. They found that IR, the most frequent AS event in plants, was also the event with highest conservation in *A. thaliana*. These results suggest that the fate of AS events after GD depends on the type of AS event and the type of GD.

Lambert et al. (2014) analyzed paralogous and singleton genes from three important gene ontology categories in zebrafish to better understand the relationship between AS and GD. The authors investigated the conservation of exon structure in paralogous genes by cataloging genes into paralogs with the same exon structure and paralogs with different exon structures. They found that paralogs with a change in its exon structure tend to have fewer AS events than genes with the same exon structure. They also looked for orthologous genes and their isoforms in human. They found that the percentage of human genes affected by AS and the number of AS events was less in homologous genes with altered exon structure compared to genes with the same exon structure. For this work authors pooled together retrogenes with DNA-dependent SSD which could influence the conclusions. These results refute the hypothesis that genes lacking AS are predisposed to GD events (Roux and Robinson-Rechavi, 2011). Rather, there is the possibility that exon structure of paralogs is predisposed to change if the ancient gene encoded only a low number of AS events. There was no evidence of AS differences between duplicated genes with the same exon structure and singletons. These findings were confirmed by Lambert et al. (2015), where only DNA-dependent SSD were analyzed. This study compared three gene sets from human, mouse and zebrafish: duplicates with same exon structure, duplicates with different exon structure and singletons. They found that AS was less frequent in paralogous genes with different exon structure and these genes exhibited tissue-specific expression. They concluded that paralogs with altered exon structure are subfunctionalized because the expression of the two paralogs occurs in different tissues.

The relationship between GD and AS is far from being understood. Several models of this relationship have been proposed and examples of each have been demonstrated. The analysis of these processes is complex and therefore a generalization of an evolutionary model is a difficult task. The development and utilization of new technologies has allowed the identification of gene isoforms expressed in a tissue-specific or even a cell-specific manner in more model organisms (Conesa et al., 2016). More studies in this field need to be performed to more fully understand this relationship.

PERSPECTIVES

The correct identification and classification of AS and GD is fundamental to improve the understanding of the evolution of both processes. AS events must be classified in terms of how they modify the primary transcript and the expression of unique isoforms in specific conditions, tissues and developmental stages. Future studies should also consider the percentage of affected genes and that the frequency of different AS event types varies between plants and animals. The classification of AS events should also be complemented by identifying and classifying GD events. The time elapsed after a GD event is an important factor in paralogous gene AS. In addition, this could be complemented with the mutation rates for each gene. These could give insights of the evolution of the paralogous genes. Besides classifying GD events as either SSD or WGD, researchers must also determine the way they were produced and the time since the duplication is necessary. For this classification, the comparison between species is important. There are a variety of model species, each uniquely qualifies as a study organism, for the different GD processes. Plants and teleost species in the animal kingdom, for example, are good models for the classification of WGD and SSD, respectively. Several organisms and a variety of tissues and conditions need to be analyzed with a characterization and classification of type of AS and GD in order to identify an evolutionary model of AS events after GD. These analyses could be complemented with other genome-wide analysis including isoforms quantification and proteomic studies (Payne, 2015). A single evolutionary model of AS after GD may not be solely responsible for AS events, instead a combination of multiple models is more likely. Identifying the mechanisms governing which models are utilized in specific genes will improve our understanding of the evolutionary relationship between GD and AS.

REFERENCES

Abascal, F., Tress, M. L., and Valencia, A. (2015). The evolutionary fate of alternatively spliced homologous exons after gene duplication. *Genome Biol.* 7, 1392–1403. doi: 10.1093/gbe/evv076

Altschmied, J., Delfgaauw, J., Wilde, B., Duschl, J., Bouneau, L., Volff, J. N., et al. (2002). Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. *Genetics* 161, 259–267.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Cañestro, C., Albalat, R., Irimia, M., and Garcia-Fernàndez, J. (2013). Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin. Cell Dev. Biol.* 24, 83–94. doi: 10.1016/j.semcdb.2012.12.008

Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A., and Urrutia, A. O. (2014). Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol. Biol. Evol.* 31, 1402–1413. doi: 10.1093/molbev/msu083

Chen, T.-W., Wu, T. H., Ng, W. V., and Lin, W.-C. (2011). Interrogation of alternative splicing events in duplicated genes during evolution. *BMC Genomics* 12:S16. doi: 10.1186/1471-2164-12-S3-S16

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13. doi: 10.1186/s13059-016-0881-8

Cusack, B. P., and Wolfe, K. H. (2007). When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet.* 23, 270–272. doi: 10.1016/j.tig.2007.03.010

Dunker, A. K., Bondos, S. E., Huang, F., and Oldfield, C. J. (2015). Intrinsically disordered proteins and multicellular organisms. *Semin. Cell Dev. Biol.* 37, 44–55. doi: 10.1016/j.semcdb.2014.09.025

Fields, C., Adams, M., White, O., and Venter, J. (1994). How many genes in the human genome? *Nat. Genet.* 7, 345–346. doi: 10.1038/ng0794-345

Gilbert, W. (1978). Why genes in pieces? *Nature* 271:501. doi: 10.1038/271501a0

Hsiao, Y. E., Bahn, J. H., Lin, X., Chan, T., and Drive, C. E. Y. (2016). Alternative splicing modulated by genetic variants demonstrates accelerated evolution regulated by highly conserved proteins. *Genome Res.* 10, 440–450. doi: 10.1101/gr.193359.115.440

Hughes, A. L., and Friedman, R. (2008). Alternative splicing, gene duplication and connectivity in the genetic interaction network of the nematode worm *Caenorhabditis elegans*. *Genetica* 134, 181–186. doi: 10.1007/s10709-007-9223-9

Hultman, K. A., Bahary, N., Zon, L. I., and Johnson, S. L. (2007). Gene Duplication of the zebrafish *kit ligand* and partitioning of melanocyte development functions to *kit ligand* a. *PLoS Genet.* 3:e17. doi: 10.1371/journal.pgen.0030017

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689

Jin, L., Kryukov, K., Clemente, J. C., Komiyama, T., Suzuki, Y., Imanishi, T., et al. (2008). The evolutionary relationship between gene duplication and alternative splicing. *Gene* 427, 19–31. doi: 10.1016/j.gene.2008.09.002

Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* 40, 2454–2469. doi: 10.1093/nar/gkr932

Kito, K., Ito, H., Nohara, T., Ohnishi, M., Ishibashi, Y., and Takeda, D. (2016). Yeast interspecies comparative proteomics reveals divergence in expression profiles and provides insights into proteome resource allocation and evolutionary roles of gene duplication. *Mol. Cell. Proteomics* 15, 218–235. doi: 10.1074/mcp.M115.051854

Kopelman, N. M., Lancet, D., and Yanai, I. (2005). Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat. Genet.* 37, 588–589. doi: 10.1038/ng1575

Lambert, M. J., Cochran, W. O., Wilde, B. M., Olsen, K. G., and Cooper, C. D. (2015). Evidence for widespread subfunctionalization of splice forms in vertebrate genomes. *Genome Res.* 125, 624–632. doi: 10.1101/gr.184473.114

Lambert, M. J., Olsen, K. G., and Cooper, C. D. (2014). Gene duplication followed by exon structure divergence substitutes for alternative splicing in zebrafish. *Gene* 546, 271–276. doi: 10.1016/j.gene.2014.05.068

Lan, X., and Pritchard, J. K. (2016). Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science* 352, 1009–1013. doi: 10.1126/science.aad8411

Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P. N. (2016). High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. *Dev. Cell* 39, 508–522. doi: 10.1016/j.devcel.2016.10.012

Lin, H., Ouyang, S., Egan, A., Nobuta, K., Haas, B. J., Zhu, W., et al. (2008). Characterization of paralogous protein families in rice. *BMC Plant Biol.* 8:18. doi: 10.1186/1471-2229-8-18

Liu, G., Mattick, J. S., and Taft, R. J. (2013). A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle* 12, 2061–2072. doi: 10.4161/cc.25134

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., and Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *J. Genet.* 92, 155–161. doi: 10.1007/s12041-013-0212-8

Marshall, A. N., Montealegre, M. C., Jiménez-López, C., Lorenz, M. C., and van Hoof, A. (2013). Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. *PLoS Genet.* 9:e1003376. doi: 10.1371/journal.pgen.1003376

Niklas, K. J., Bondos, S. E., Dunker, A. K., and Newman, S. A. (2015). Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front. Cell Dev. Biol.* 3:8. doi: 10.3389/fcell.2015.00008

Pacheco, T. R., Gomes, A. Q., Barbosa-Morais, N. L., Benes, V., Ansorge, W., Wollerton, M., et al. (2004). Diversity of vertebrate splicing factor U2AF35: identification of alternatively spliced U2AF1 mRNAs. *J. Biol. Chem.* 279, 27039–27049. doi: 10.1074/jbc.M402136200

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415. doi: 10.1038/ng.259

Panchy, N., Lehti-Shiu, M. D., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523

Payne, S. H. (2015). The utility of protein and mRNA correlation. *Trends Biochem. Sci.* 40, 1–3. doi: 10.1016/j.tibs.2014.10.010

Ramani, A., Calarco, J., and Pan, Q. (2011). Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res.* 21, 342–348. doi: 10.1101/gr.114645.110.Freely

Reddy, A. S. N., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523

Remy, E., Cabrito, T. R., Baster, P., Batista, R. A., Teixeira, M. C., Friml, J., et al. (2013). A major facilitator superfamily transporter plays a dual role in polar auxin transport and drought stress tolerance in *Arabidopsis*. *Plant Cell* 25, 901–926. doi: 10.1105/tpc.113.110353

Roux, J., and Robinson-Rechavi, M. (2011). Age-dependent gain of alternative splice forms and biased duplication explain the relation between splicing and duplication. *Genome Res.* 21, 357–363. doi: 10.1101/gr.113803.110

Schad, E., Tompa, P., and Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.* 12:R120. doi: 10.1186/gb-2011-12-12-r120

Staiger, D., and Brown, J. W. S. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803

Su, Z., and Gu, X. (2012). Revisit on the evolutionary relationship between alternative splicing and gene duplication. *Gene* 504, 102–106. doi: 10.1016/j.gene.2012.05.012

Su, Z., Wang, J., Yu, J., Huang, X., and Gu, X. (2006). Evolution of alternative splicing after gene duplication. *Genome Res.* 16, 182–189. doi: 10.1101/gr.4197006

Tack, D. C., Pitchers, W. R., and Adams, K. L. (2014). Transcriptome analysis indicates considerable divergence in alternative splicing between duplicated genes in *Arabidopsis thaliana*. *Genetics* 198, 1473–1481. doi: 10.1534/genetics.114.169466

Talavera, D., Vogel, C., Orozco, M., Teichmann, S. A., and De La Cruz, X. (2007). The (In)dependence of alternative splicing and gene duplication. *PLoS Comput. Biol.* 3:e33. doi: 10.1371/journal.pcbi.0030033

59

Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600

Vélez-Bermúdez, I. C., and Schmidt, W. (2014). The conundrum of discordant protein and mRNA expression. Are plants special? *Front. Plant Sci.* 5:619. doi: 10.3389/fpls.2014.00619

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509

Wang, Y., Wang, X., and Paterson, A. H. (2012). Genome and gene duplications and gene expression divergence: a view from plants. *Ann. N. Y. Acad. Sci.* 1256, 1–14. doi: 10.1111/j.1749-6632.2011.06384.x

Wang, Y., Wang, X., Tang, H., Tan, X., Ficklin, S. P., Feltus, F. A., et al. (2011). Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. *PLoS ONE* 6:e28150. doi: 10.1371/journal.pone.0028150

Wei-Ping, Y., Sydney, B., and Byrappa, V. (2003). Duplication, degeneration and subfunctionalization of the nested synapsin – Timp genes in Fugu. *Trends Genet.* 19, 180–183. doi: 10.1016/S0168-9525(03)00041-6

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G. M., Hao, T., Richardson, A., et al. (2016). Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* 164, 805–817. doi: 10.1016/j.cell.2016.01.029

Zhang, C., Yang, H., and Yang, H. (2015). Evolutionary character of alternative splicing in plants. *Bioinform. Biol. Insights* 9, 47–52. doi: 10.4137/BBI.S33716

Zhiguo, E., Wang, L., and Zhou, J. (2013). Splicing and alternative splicing in rice and humans. *BMB Rep.* 46, 439–447. doi: 10.5483/BMBRep.2013.46.9.161

60

5.2.2. Número de eventos en parálogos recientes de *Glycine max*

Existen diversas teorías sobre lo que sucede con los eventos de AS una vez que los genes se han duplicado que se discuten en la introducción de esta sección. Estas teorías se basan en las observaciones donde encuentran un bajo número de eventos de AS en genes duplicados. Para explorar este fenómeno las plantas son modelos ideales ya que cuentan con una gran variedad de duplicaciones. Todas las leguminosas, como se comentó en el artículo de la sección anterior (Iniguez et al., 2017), cuentan con una duplicación genómica completa ancestral (˜56 millones de años) y una fracción de los genes presentan un parálogo hasta hoy en día. Además de esta característica de las leguminosas, *G. max* presenta una duplicación reciente (10 millones de años) y muchos de los genes presentan un parálogo. Esta duplicación reciente conlleva a que para un gen de *P. vulgaris* hay dos genes homólogos en *G. max*. Esta característica de ambas especies permite el análisis evolutivo del AS en genes duplicados a partir de una duplicación genómica completa.

En la sección anterior se analizó la conservación de eventos entre soya y frijol y para determinar si la conservación se pudiera dar de manera aleatoria se realizaron diversas simulaciones. Estas simulaciones arrojaron resultados sobre el número de eventos de AS que se esperaría en los parálogos recientes de soya donde el número de eventos esperados era mayor al identificado en las muestras analizadas. Por otro lado para los genes de *G.max* que no presentaban un duplicado de acuerdo a los criterios de selección de parálogos (falta de expresión de un parálogo o diferencias en el número de exones), el número de eventos de AS era similar en las simulaciones comparado con el número de eventos identificados (Figura 5). Estos resultados fueron consistentes con los reportados en humanos y ratones donde el número de eventos de AS en genes duplicados es menor al número de eventos de AS en genes sin copias (Su et al., 2006; Chen et al., 2011).

Por otro lado los genes de *P. vulgaris* que presentaban homología con parálogos recientes de soya ("relación evolutiva 1:2") también mostraron en general un menor número de eventos de AS al esperado al azar, exceptuando los eventos de IR y NI (Figura 5). Y aquellos genes de frijol con una "relación evolutivarespecto a soya de 1:1 mostraron un número similar de

eventos a los esperados al azar. Para investigar este fenómeno se analizaron diversos compo-

nentes de los genes pertenecientes a ambas "relaciones evolutivas"(1:1 y 1:2), tanto de frijol

como de soya.

Figura 5: Número de eventos esperados al azar para genes con "relación evolutiva 1:2 y 1:1"
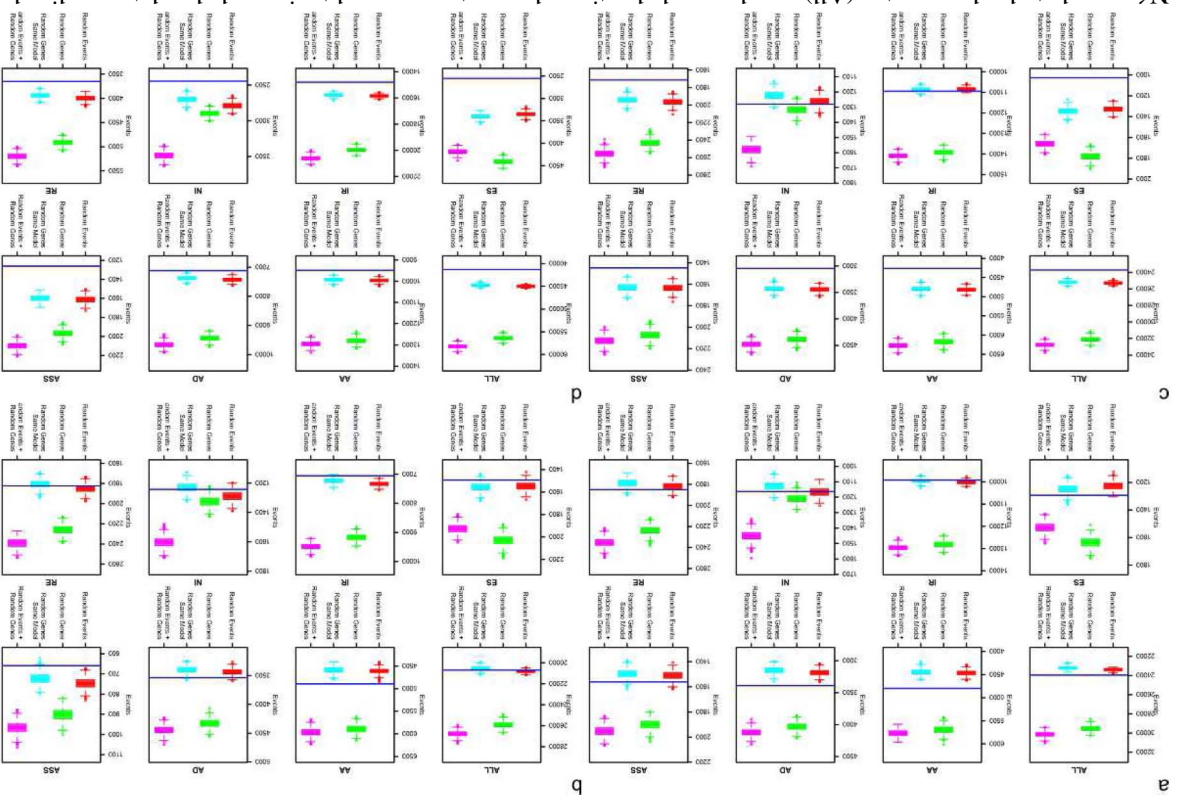
Número de todos los eventos (All) y cada uno de los tipos de eventos que se obtuvieron de los datos analizados (línea azul) y la distribución de eventos obtenida en las simulaciones. (a) Genes de *P. vulgaris* con un solo homólogo con la misma estructura en *G. max* (1:1), los homólogos de soya están graficados en (b). (c) y (d) son iguales a (a) y (b) pero para genes con una "relación evolutiva 1:2".

En general, los genes de ambas especies con una "relación evolutiva 1:2"presentaron menos eventos de AS, en especial AA, AS, ASS ES y RE al compararlos con los genes en "relación evolutiva 1:1". Los eventos de IR y NI, en *P. vulgaris*, no presentaron diferencias entre ambas listas génicas. Los eventos de ES fueron los que más diferencias mostraron entre ambas relaciones en las dos especies, lo cual concuerda con los resultados de Chen et al. (2011) en los que ven que los duplicados de humanos y ratones tienen menos eventos de AS y dado que ES es el evento más común en estos organismos el resultado se refleja de manera global en los eventos de AS.

Se analizaron tanto el número de exones como la longitud de los mismos y los tipos de exones, ya fuera 5'UTR, CDS o 3'UTR, pertenecientes a ambas relaciones evolutivas. La distribución del número de exones por gene y exones de tipo CDS resultaron similares entre ambas relaciones evolutivas en las dos especies. El número de exones combinados (5'UTR-CDS y CDS-3'UTR) fue menor en los genes cuya "relación evolutiva es de 1:1.ª aquellos con una "relación evolutiva 1:2". Estos tipos de exones se vieron enriquecidos en los eventos de NI (Iniguez et al., 2017) pero no influyeron de manera positiva en los genes con "relación evolutiva 1:2". La distribución del número de exones y longitud de exones no codificantes también fue diferente entre ambas relaciones evolutivas; genes con una "relación evolutiva 1:1"presentaron más exones de estos tipos (3'UTR y 5'UTR) pero más pequeños que aquellos en una "relación evolutiva 1:2". Esto quiere decir que los genes con "relación evolutiva 1:1"presentan más exones no codificantes pero de longitud pequeña. Estos resultados son importantes porque pueden explicar el bajo número de algunos eventos de AS en genes con "relación evolutiva 1:2ʑa que como se explicó en el artículo de Iniguez et al. (2017) algunos tipos de eventos de AS tienden a estar en intrones de exones no codificantes y por lo tanto al haber menos exones no codificantes en los genes con "relación evolutiva 1:2resulta en una disminución de eventos de AA, AD, ASS y RE. Los eventos de IR no mostraron una preferencia por intrones de exones no codificantes como los demás eventos (Iniguez et al., 2017), y como el número de exones codificantes fue similar en ambas listas el número de eventos de IR esperado en ambas listas sería similar. El último aspecto evaluado fue la tasa evolutiva utilizando la tasa de sustituciones no sinónimas sobre sustituciones sinónimas (dN/dS) de ambas listas génicas al ser comparadas con su homólogo en la otra especie. La tasa evolutiva en general fue menor en genes con "relación evolutiva 1:2", esto implica que estos genes están bajo una selección más fuerte y esto puede influenciar a la adquisición de nuevos eventos de AS restringiendo la producción de isoformas.

Es difícil identificar el mecanismo que da origen a las observaciones donde los genes con una "relación evolutiva 1:2"presentan menos eventos que aquellos con 1:1 pero los resultados que se muestran en esta sección guían hacia ciertos experimentos y datos que se necesitan

analizar. Los resultados mostrados aquí indican una tendencia a tener pocos eventos de AS en genes duplicados, pero el bajo número de eventos de AS en genes recientemente duplicados de *G. max* se ve también en los genes homólogos de *P. vulgaris*. Estos resultados concuerdan con los de (Lambert et al., 2014, 2015) donde se sugiere que los eventos de AS no se pierden en los duplicados recientes sino que los genes que tienden a mantener su estructura génica, uno de los criterios utilizados para establecer ortólogos, tienen un menor número de eventos de AS.
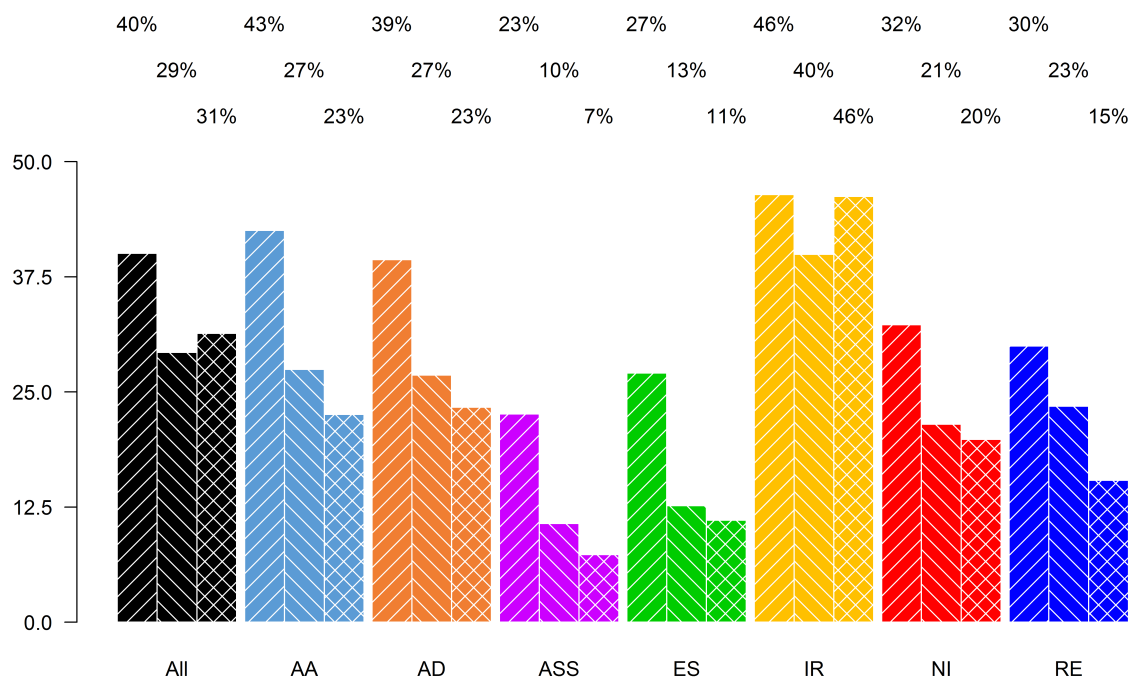
5.2.3. Conservación de eventos entre parálogos

Una de los modelos sobre el destino del AS al duplicarse los genes es la de función compartida, donde las isoformas que se producían antes de la duplicación se dividen en los parálogos. Si esto sucediera la conservación se perdería y el número de isoformas se reduciría. Por otro lado un modelo opuesto es el de independencia donde cada parálogo es independiente y la transcripción de isoformas a partir del AS mantiene cierta libertad mientras las funciones ancestrales se conserven. Con el modelo de independencia el número de isoformas se mantiene y puede llegar a conservar eventos de AS.

El número de eventos se ha visto que es bajo en los genes duplicados, discutido en la sección anterior. Mientras que la conservación de eventos en genes parálogos es una pregunta que queda pendiente para analizar. Debido a las dos duplicaciones que presenta *G. max*, la reciente y la ancestral compartida con *P. vulgaris* es posible identificar los eventos conservados entre parálogos recientes y ancestrales. Para analizar la conservación de parálogos recientes se analizaron 9,058 parejas de parálogos de *G. max* y se observó una conservación general del 40 % de los eventos, mientras que para cada evento en específico la conservación varió entre un 23 % para ASS a 46 % para IR (Figura 6. En *P. vulgaris* la conservación en parálogos ancestrales (2114 pares de parálogos) se redujo a un 29 % de manera general y para cada evento el porcentaje de conservación fue entre 10 % y 40 %. Para identificar la conservación en parálogos ancestrales de soya se colapsaron los eventos en los pares de genes que presentaban duplicaciones recientes. Esto permitió comparar entre pares de parálogos ancestrales.

2,296 pares de genes colapsados presentaron una conservación ancestral general del 31 % y el rango de conservación entre los eventos específicos fue de 7 %-46 % (Figura 6).

Figura 6: Conservación de eventos entre parálogos

En la gráfica se muestran los porcentajes de conservación entre los parálogos de manera general (color negro) y para cada evento en específico en las duplicaciones ancestrales de *P. vulgaris* y *G. max* y en la reciente para soya. Las barras con líneas diagonales hacia la izquierda corresponden a *G. max* y las barras con líneas diagonales hacia la derecha a la duplicación ancestral. Los porcentajes de la parte superior corresponden a la conservación de eventos de AS en duplicados recientes de soya, duplicados ancestrales de frijol y duplicados ancestrales de soya, de manera descendente.

Los resultados de la conservación muestran que existe una mayor conservación entre parálogos recientes que ancestrales para casi todos los eventos sin embargo la conservación de IR es similar entre las duplicaciones ancestrales y recientes. Esto quiere decir que la funcionalidad o regulación de estos eventos de AS no es dependiente del tiempo de divergencia entre parálogos y por lo tanto el modelo de función compartida no aplica y se ajustaría mejor el modelo de independencia. Para el resto de los eventos la conservación se reduce considerablemente con el tiempo y es constante entre ambas especies pero se mantiene una conservación significativa. Con estos resultados se puede concluir que un único modelo evolutivo

no se puede aplicar para todos los eventos, sino que depende del tipo de evento y puede ser que cada gen presente algún modelo. Para investigar esto a más detalle se necesitan realizar más experimentos.

6.   Discusión General y Perspectivas

La transcriptómica es un enfoque fundamental para entender la biología, no solo la vegetal sino la de todos los organismos, ya que permite identificar los genes expresados ante situaciones, condiciones específicas o durante el desarrollo. El control de la transcripción regula los mRNA que se expresan y en qué concentraciones pero no es la única regulación por la cual se definen las concetraciones protéicas ya que existen diversos factores post-transcripcionales que también la regulan. Entre éstos se encuentran los miRNAs y el AS.

Los distintos perfiles transcripcionales que se estudiaron a lo largo de la tesis muestran distintos patrones de expresión en determinadas circunstancias y aunque los resultados muestran a gran escala lo que sucede a nivel transcriptómico un estudio más a detalle sobre los genes, las vías o las ontologías diferencialmente expresadas puede resultar en contribuciones de mayor relevancia para lograr esclarecer la respuesta celular ante el estrés oxidativo. Además de esto una mejor anotación genómica pudiera funcionar para identificar a mejor detalle diversos aspectos relacionados con el estrés.

La identificación de miRNAs en *P. vulgaris* en conjunto con su expresión determinaron una variedad de redes de coexpresión donde además de encontrarse miRNAs estudiados se identificaron nuevos. Esto sugiere una posible funcionalidad de estos inéditos RNA's no codificantes. La integración de datos experimentales junto con resultados bioinformáticos puede dar origen a nuevos aspectos biológicos, incluyendo aspectos de interés sobre los nódulos. Este fenómeno se demostró con el trabajo del miRNA172c y su blanco *AP2* y actualmente se están realizando análisis experimentales y bioinformáticos para determinar relaciones funcionales entre ciertos miRNAs y sus respectivos blancos.

A partir del análisis de genes co-expresados con el factor de transcripción AP2 se logró obtener genes candidatos relacionados con el desarrollo y muerte de los nódulos. De igual forma al analizar los sitios de unión de factores de transcripción en las regiones promotoras de los genes de *MIRNA172* se identificó una familia de factores de transcripción (AGL). Esta familia se ha caracterizado durante la organogénesis de flor y puede ser importante para el

desarrollo de nódulo. A partir de esto ya se están realizando diversos trabajos en el laboratorio para caracterizar la función de algunos de estos factores de transcripción.

El estudio a escala genómica del AS en *P. vulgaris* y *G. max* arrojó una gran cantidad de datos sobre AS. Una de las preguntas que quedan por resolver es la funcionalidad de los eventos de AS conservados. Para esto ya se realizó una búsqueda de eventos presentes en nódulos o raíces pero no en ambos tejidos y también aquellos presentes en diversas muestras provenientes de plantas que fueron crecidas en diferentes fuentes de nitrógeno, como fertilización vs SNF. Por el momento se cuenta con una lista de genes candidatos con eventos de AS que indican una función biológica diferente al mRNA principal y los experimentos pertinentes se están comenzado a realizar.

La relación que existe entre el AS y la duplicación génica es un aspecto importante para la evolución y algunos análisis presentados en esta tesis muestran resultados interesantes. Sin embargo estos no son del todo concluyentes y más análisis se necesitan realizar. La duplicación génica puede provocar un desequilibrio en la cantidad de AS que presenten los parálogos y esto pudiera tener alguna relación con la expresión génica o con la funcionalidad biológica. Este tipo de preguntas quedan pendientes de resolverse.

Por último queda mencionar que la integración de todos los datos es fundamental para la comprensión de los organismos. A partir de la secuenciación masiva se pueden plantear una serie de experimentos a diferentes escalas que conlleven a un mejor entendimiento de la biología en general. Y con base en esos resultados se deben de plantear los experimentos y análisis precisos que orienten a entender cada una de los mecanismos que conforman a los organismos.

Referencias

Anand,A., Duk,B. T., Singh,S., Akbas,M. Y., Webster,D. A., Stark,B. C., and Dikshit,K. L. Redox-mediated interactions of VHb (Vitreoscilla haemoglobin) with OxyR: novel regulation of VHb biosynthesis under oxidative stress. *Biochem. J.*, 426(3):271–280, Feb 2010.

Arenas-Huertero,C., Perez,B., Rabanal,F., Blanco-Melo,D., De la Rosa,C., Estrada-Navarrete,G., Sanchez,F., Covarrubias,A. A., and Reyes,J. L. Conserved and novel miRNAs in the legume Phaseolus vulgaris in response to stress. *Plant Mol. Biol.*, 70(4):385–401, Jul 2009.

Aukerman,M. J. and Sakai,H. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell*, 15(11):2730–2741, Nov 2003.

Azpeitia,E., Davila-Velderrain,J., Villarreal,C., and Alvarez-Buylla,E. R. Gene regulatory network models for floral organ determination. *Methods Mol. Biol.*, 1110:441–469, 2014.

Broughton,W. J., Hernández,G., Blair,M., Beebe,S., Gepts,P., and Vanderleyden,J. Beans (*Phaseolus* spp.); model food legumes. *Plant and Soil*, 252(1):55–128, 2003.

Castillo-Davis,C. I., Mekhedov,S. L., Hartl,D. L., Koonin,E. V., and Kondrashov,F. A. Selection for short introns in highly expressed genes. *Nat. Genet.*, 31(4):415–418, Aug 2002.

Catania,F. and Lynch,M. Where do introns come from? *PLoS Biol.*, 6(11):e283, Nov 2008.

Causier,B., Schwarz-Sommer,Z., and Davies,B. Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.*, 21(1):73–79, Feb 2010.

Chamala,S., Feng,G., Chavarro,C., and Barbazuk,W. B. Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front Bioeng Biotechnol*, 3:33, 2015.

Chen,L., Bush,S. J., Tovar-Corona,J. M., Castillo-Morales,A., and Urrutia,A. O. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol. Biol. Evol.*, 31(6):1402–1413, Jun 2014.

Chen,T. W., Wu,T. H., Ng,W. V., and Lin,W. C. Interrogation of alternative splicing events in duplicated genes during evolution. *BMC Genomics*, 12 Suppl 3:S16, Nov 2011.

Chiou,T. J., Aung,K., Lin,S. I., Wu,C. C., Chiang,S. F., and Su,C. L. Regulation of phosphate homeostasis by MicroRNA in Arabidopsis. *Plant Cell*, 18(2):412–421, Feb 2006.

Combier,J. P., Frugier,F., de Billy,F., Boualem,A., El-Yahyaoui,F., Moreau,S., Vernie,T., Ott,T., Gamas,P., Crespi,M., and Niebel,A. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes Dev.*, 20(22):3084–3088, Nov 2006.

Donahue,J. L., Okpodu,C. M., Cramer,C. L., Grabau,E. A., and Alscher,R. G. Responses of Antioxidants to Paraquat in Pea Leaves (Relationships to Resistance). *Plant Physiol.*, 113 (1):249–257, Jan 1997.

Estrada-Navarrete,G., Alvarado-Affantranger,X., Olivares,J. E., Guillen,G., Diaz-Camino,C., Campos,F., Quinto,C., Gresshoff,P. M., and Sanchez,F. Fast, efficient and reproducible genetic transformation of Phaseolus spp. by Agrobacterium rhizogenes. *Nat Protoc*, 2(7): 1819–1824, 2007.

Ferguson,B. J., Indrasumunar,A., Hayashi,S., Lin,M. H., Lin,Y. H., Reid,D. E., and Gresshoff,P. M. Molecular analysis of legume nodule development and autoregulation. *J Integr Plant Biol*, 52(1):61–76, Jan 2010.

Filichkin,S., Priest,H. D., Megraw,M., and Mockler,T. C. Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.*, 24:125–135, Apr 2015.

Formey,D., Iniguez,L. P., Pelaez,P., Li,Y. F., Sunkar,R., Sanchez,F., Reyes,J. L., and Hernández,G. Genome-wide identification of the Phaseolus vulgaris sRNAome using small RNA and degradome sequencing. *BMC Genomics*, 16:423, Jun 2015.

Geckil,H., Stark,B. C., and Webster,D. A. Cell growth and oxygen uptake of Escherichia coli and Pseudomonas aeruginosa are differently effected by the genetically engineered Vitreoscilla hemoglobin gene. *J. Biotechnol.*, 85(1):57–66, Jan 2001.

Geckil,H., Gencer,S., Kahraman,H., and Erenler,S. O. Genetic engineering of Enterobacter aerogenes with the Vitreoscilla hemoglobin gene: cell growth, survival, and antioxidant enzyme status under oxidative stress. *Res. Microbiol.*, 154(6):425–431, 2003.

German,M. A., Pillay,M., Jeong,D. H., Hetawal,A., Luo,S., Janardhanan,P., Kannan,V., Rymarquis,L. A., Nobuta,K., German,R., De Paoli,E., Lu,C., Schroth,G., Meyers,B. C., and Green,P. J. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.*, 26(8):941–946, Aug 2008.

Hardison,R. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J. Exp. Biol.*, 201(Pt 8):1099–1117, Apr 1998.

Hedges,S. B., Blair,J. E., Venturi,M. L., and Shoe,J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.*, 4:2, Jan 2004.

Iida,K., Seki,M., Sakurai,T., Satou,M., Akiyama,K., Toyoda,T., Konagaya,A., and Shinozaki,K. Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. *Nucleic Acids Res.*, 32(17):5096–5103, 2004.

Iniguez,L. P., Ramirez,M., Barbazuk,W. B., and Hernandez,G. Identification and analysis of alternative splicing events in Phaseolus vulgaris and Glycine max. *BMC Genomics*, 18(1): 650, Aug 2017.

Irimia,M. and Roy,S. W. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb Perspect Biol*, 6(6), Jun 2014.

Jean-Philippe,J., Paz,S., and Caputi,M. hnRNP A1: the Swiss army knife of gene expression. *Int J Mol Sci*, 14(9):18999–19024, Sep 2013.

Jeong,S. SR Proteins: Binders, Regulators, and Connectors of RNA. *Mol. Cells*, 40(1):1–9, Jan 2017.

Jofuku,K. D., den Boer,B. G., Van Montagu,M., and Okamuro,J. K. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell*, 6(9):1211–1225, Sep 1994.

Jones-Rhoades,M. W., Bartel,D. P., and Bartel,B. MicroRNAS and their regulatory roles in plants. *Annu Rev Plant Biol*, 57:19–53, 2006.

Kalyna,M., Simpson,C. G., Syed,N. H., Lewandowska,D., Marquez,Y., Kusenda,B., Marshall,J., Fuller,J., Cardle,L., McNicol,J., Dinh,H. Q., Barta,A., and Brown,J. W. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.*, 40(6):2454–2469, Mar 2012.

Kaur,R., Pathania,R., Sharma,V., Mande,S. C., and Dikshit,K. L. Chimeric Vitreoscilla hemoglobin (VHb) carrying a flavoreductase domain relieves nitrosative stress in Escherichia coli: new insight into the functional role of VHb. *Appl. Environ. Microbiol.*, 68(1):152–160, Jan 2002.

Khosla,C. and Bailey,J. E. Heterologous expression of a bacterial haemoglobin improves the growth properties of recombinant Escherichia coli. *Nature*, 331(6157):633–635, Feb 1988.

Kwak,M., Toro,O., Debouck,D. G., and Gepts,P. Multiple origins of the determinate growth habit in domesticated common bean *(Phaseolus vulgaris)*. *Ann. Bot.*, 110(8):1573–1580, Dec 2012.

Lambert,M. J., Olsen,K. G., and Cooper,C. D. Gene duplication followed by exon structure divergence substitutes for alternative splicing in zebrafish. *Gene*, 546(2):271–276, Aug 2014.

Lambert,M. J., Cochran,W. O., Wilde,B. M., Olsen,K. G., and Cooper,C. D. Evidence for widespread subfunctionalization of splice forms in vertebrate genomes. *Genome Res.*, 25 (5):624–632, May 2015.

Li,Q., Xiao,G., and Zhu,Y. X. Single-nucleotide resolution mapping of the Gossypium raimondii transcriptome reveals a new mechanism for alternative splicing of introns. *Mol Plant*, 7(5):829–840, May 2014.

Li,X. and Zhang,Y. Reverse genetics by fast neutron mutagenesis in higher plants. *Funct. Integr. Genomics*, 2(6):254–258, Nov 2002.

Liu,Y., Gonzalez-Porta,M., Santos,S., Brazma,A., Marioni,J. C., Aebersold,R., Venkitaraman,A. R., and Wickramasinghe,V. O. Impact of Alternative Splicing on the Human Proteome. *Cell Rep*, 20(5):1229–1241, Aug 2017.

Mandadi,K. K. and Scholthof,K. B. Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in Brachypodium distachyon. *Plant Cell*, 27(1):71–85, Jan 2015.

Marquez,Y., Brown,J. W., Simpson,C., Barta,A., and Kalyna,M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.*, 22(6):1184–1195, Jun 2012.

Matamoros,M. A., Dalton,D. A., Ramos,J., Clemente,M. R., Rubio,M. C., and Becana,M. Biochemistry and molecular biology of antioxidants in the rhizobia-legume symbiosis. *Plant Physiol.*, 133(2):499–509, Oct 2003.

McGuire,A. M., Pearson,M. D., Neafsey,D. E., and Galagan,J. E. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol.*, 9(3):R50, 2008.

Mendoza-Soto,A. B., Sanchez,F., and Hernández,G. MicroRNAs as regulators in plant metal toxicity response. *Front Plant Sci*, 3:105, 2012.

Mendoza-Soto,A. B., Naya,L., Leija,A., and Hernández,G. Responses of symbiotic nitrogen-fixing common bean to aluminum toxicity and delineation of nodule responsive microRNAs. *Front Plant Sci*, 6:587, 2015.

Meng,Y., Ma,X., Chen,D., Wu,P., and Chen,M. MicroRNA-mediated signaling involved in plant root development. *Biochem. Biophys. Res. Commun.*, 393(3):345–349, Mar 2010.

Montiel,J., Arthikala,M. K., Cardenas,L., and Quinto,C. Legume NADPH Oxidases Have Crucial Roles at Different Stages of Nodulation. *Int J Mol Sci*, 17(5), May 2016.

Navarro,L., Dunoyer,P., Jay,F., Arnold,B., Dharmasiri,N., Estelle,M., Voinnet,O., and Jones,J. D. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science*, 312(5772):436–439, Apr 2006.

Naya,L., Paul,S., Valdes-Lopez,O., Mendoza-Soto,A. B., Nova-Franco,B., Sosa-Valencia,G., Reyes,J. L., and Hernández,G. Regulation of copper homeostasis and biotic interactions by microRNA 398b in common bean. *PLoS ONE*, 9(1):e84416, 2014.

Ner-Gaon,H., Leviatan,N., Rubin,E., and Fluhr,R. Comparative cross-species alternative splicing in plants. *Plant Physiol.*, 144(3):1632–1641, Jul 2007.

O'Rourke,J. A., Iniguez,L. P., Bucciarelli,B., Roessler,J., Schmutz,J., McClean,P. E., Jackson,S. A., Hernández,G., Graham,M. A., Stupar,R. M., and Vance,C. P. A re-sequencing based assessment of genomic heterogeneity and fast neutron-induced deletions in a common bean cultivar. *Front Plant Sci*, 4:210, 2013.

O'Rourke,J. A., Iniguez,L. P., Fu,F., Bucciarelli,B., Miller,S. S., Jackson,S. A., McClean,P. E., Li,J., Dai,X., Zhao,P. X., Hernández,G., and Vance,C. P. An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics*, 15:866, Oct 2014.

Pan,Q., Shai,O., Lee,L. J., Frey,B. J., and Blencowe,B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, 40(12):1413–1415, Dec 2008.

Panahi,B., Abbaszadeh,B., Taghizadeghan,M., and Ebrahimie,E. Genome-wide survey of Alternative Splicing in Sorghum Bicolor. *Physiol Mol Biol Plants*, 20(3):323–329, Jul 2014.

Patthy,L. Genome evolution and the evolution of exon-shuffling–a review. *Gene*, 238(1): 103–114, Sep 1999.

Pelaez,P., Trejo,M. S., Iniguez,L. P., Estrada-Navarrete,G., Covarrubias,A. A., Reyes,J. L., and Sanchez,F. Identification and characterization of microRNAs in Phaseolus vulgaris by high-throughput sequencing. *BMC Genomics*, 13:83, Mar 2012.

Perez Guerra,J. C., Coussens,G., De Keyser,A., De Rycke,R., De Bodt,S., Van De Velde,W., Goormachtig,S., and Holsters,M. Comparison of developmental and stress-induced nodule senescence in Medicago truncatula. *Plant Physiol.*, 152(3):1574–1584, Mar 2010.

Ramani,A. K., Calarco,J. A., Pan,Q., Mavandadi,S., Wang,Y., Nelson,A. C., Lee,L. J., Morris,Q., Blencowe,B. J., Zhen,M., and Fraser,A. G. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. *Genome Res.*, 21(2):342–348, Feb 2011.

Ramirez,M., Valderrama,B., Arredondo-Peter,R., Soberon,M., Mora,J., and Hernández,G. Rhizobium etli genetically engineered for the heterologous expression of Vitreoscilla sp hemoglobin: Effects on free- living and symbiosis. *Molecular Plant-Microbe Interactions*, 12(11):1008–1015, 1999.

Ramirez,M., Flores-Pacheco,G., Reyes,J. L., Luzlvarez,A., Drevon,J. J., Girard,L., and Hernández,G. Two common bean genotypes with contrasting response to phosphorus deficiency show variations in the microRNA 399-mediated PvPHO2 regulation within the PvPHR1 signaling pathway. *Int J Mol Sci*, 14(4):8328–8344, Apr 2013.

Reddy,A. S. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol*, 58:267–294, 2007.

Reddy,A. S., Marquez,Y., Kalyna,M., and Barta,A. Complexity of the alternative splicing landscape in plants. *Plant Cell*, 25(10):3657–3683, Oct 2013.

Remy,E., Cabrito,T. R., Baster,P., Batista,R. A., Teixeira,M. C., Friml,J., Sa-Correia,I., and Duque,P. A major facilitator superfamily transporter plays a dual role in polar auxin transport and drought stress tolerance in Arabidopsis. *Plant Cell*, 25(3):901–926, Mar 2013.

Rendon-Anaya,M., Montero-Vargas,J. M., Saburido-Alvarez,S., Vlasova,A., Capella-Gutierrez,S., Ordaz-Ortiz,J. J., Aguilar,O. M., Vianello-Brondani,R. P., Santalla,M., Delaye,L., Gabaldon,T., Gepts,P., Winkler,R., Guigo,R., Delgado-Salinas,A., and Herrera-Estrella,A. Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.*, 18(1):60, Mar 2017.

Rogozin,I. B., Carmel,L., Csuros,M., and Koonin,E. V. Origin and evolution of spliceosomal introns. *Biol. Direct*, 7:11, Apr 2012.

Schmutz,J., McClean,P. E., Mamidi,S., Wu,G. A., Cannon,S. B., Grimwood,J., Jenkins,J., Shu,S., Song,Q., Chavarro,C., Torres-Torres,M., Geffroy,V., Moghaddam,S. M., Gao,D., Abernathy,B., Barry,K., Blair,M., Brick,M. A., Chovatia,M., Gepts,P., Goodstein,D. M., Gonzales,M., Hellsten,U., Hyten,D. L., Jia,G., Kelly,J. D., Kudrna,D., Lee,R., Richard,M. M., Miklas,P. N., Osorno,J. M., Rodrigues,J., Thareau,V., Urrea,C. A., Wang,M., Yu,Y., Zhang,M., Wing,R. A., Cregan,P. B., Rokhsar,D. S., and Jackson,S. A. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.*, 46(7):707–713, Jul 2014.

Shen,Y., Zhou,Z., Wang,Z., Li,W., Fang,C., Wu,M., Ma,Y., Liu,T., Kong,L. A., Peng,D. L., and Tian,Z. Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell*, 26(3):996–1008, Mar 2014.

Staiger,D., Korneli,C., Lummer,M., and Navarro,L. Emerging role for RNA-based regulation in plant immunity. *New Phytol.*, 197(2):394–404, Jan 2013.

Su,Z., Wang,J., Yu,J., Huang,X., and Gu,X. Evolution of alternative splicing after gene duplication. *Genome Res.*, 16(2):182–189, Feb 2006.

Thatcher,S. R., Zhou,W., Leonard,A., Wang,B. B., Beatty,M., Zastrow-Hayes,G., Zhao,X., Baumgarten,A., and Li,B. Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation. *Plant Cell*, 26(9):3472–3487, Sep 2014.

Valdes-Lopez,O., Arenas-Huertero,C., Ramirez,M., Girard,L., Sanchez,F., Vance,C. P., Luis Reyes,J., and Hernández,G. Essential role of MYB transcription factor: PvPHR1 and microRNA: PvmiR399 in phosphorus-deficiency signalling in common bean roots. *Plant Cell Environ.*, 31(12):1834–1843, Dec 2008.

Valdes-Lopez,O., Yang,S. S., Aparicio-Fabre,R., Graham,P. H., Reyes,J. L., Vance,C. P., and Hernández,G. MicroRNA expression profile in common bean (Phaseolus vulgaris) under nutrient deficiency stresses and manganese toxicity. *New Phytol.*, 187(3):805–818, Aug 2010.

Van de Velde,W., Guerra,J. C., De Keyser,A., De Rycke,R., Rombauts,S., Maunoury,N., Mergaert,P., Kondorosi,E., Holsters,M., and Goormachtig,S. Aging in legume symbiosis. A molecular view on nodule senescence in Medicago truncatula. *Plant Physiol.*, 141(2): 711–720, Jun 2006.

van der Burgt,A., Severing,E., de Wit,P. J., and Collemare,J. Birth of new spliceosomal introns in fungi by multiplication of introner-like elements. *Curr. Biol.*, 22(13):1260–1265, Jul 2012.

Velculescu,V. E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M. A., Bassett,D. E., Hieter,P., Vogelstein,B., and Kinzler,K. W. Characterization of the yeast transcriptome. *Cell*, 88(2): 243–251, Jan 1997.

Vitulo,N., Forcato,C., Carpinelli,E. C., Telatin,A., Campagna,D., D'Angelo,M., Zimbello,R., Corso,M., Vannozzi,A., Bonghi,C., Lucchin,M., and Valle,G. A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biol.*, 14:99, Apr 2014.

Vlasova,A., Capella-Gutierrez,S., Rendon-Anaya,M., Hernández-Onate,M., Minoche,A. E., Erb,I., Camara,F., Prieto-Barja,P., Corvelo,A., Sanseverino,W., Westergaard,G., Dohm,J. C., Pappas,G. J., Saburido-Alvarez,S., Kedra,D., Gonzalez,I., Cozzuto,L., Gomez-Garrido,J., Aguilar-Moron,M. A., Andreu,N., Aguilar,O. M., Garcia-Mas,J., Zehnsdorf,M., Vazquez,M. P., Delgado-Salinas,A., Delaye,L., Lowy,E., Mentaberry,A., Vianello-Brondani,R. P., Garcia,J. L., Alioto,T., Sanchez,F., Himmelbauer,H., Santalla,M., Notredame,C., Gabaldon,T., Herrera-Estrella,A., and Guigo,R. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biol.*, 17:32, Feb 2016.

Voinnet,O. Origin, biogenesis, and activity of plant microRNAs. *Cell*, 136(4):669–687, Feb 2009.

Wang,B. B., O'Toole,M., Brendel,V., and Young,N. D. Cross-species EST alignments reveal novel and conserved alternative splicing events in legumes. *BMC Plant Biol.*, 8:17, Feb 2008.

Wang,L., Cao,C., Ma,Q., Zeng,Q., Wang,H., Cheng,Z., Zhu,G., Qi,J., Ma,H., Nian,H., and Wang,Y. RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol.*, 14:169, Jun 2014.

Wilhelm,B. T. and Landry,J. R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3):249–257, Jul 2009.

Worden,A. Z., Lee,J. H., Mock,T., Rouze,P., Simmons,M. P., Aerts,A. L., Allen,A. E., Cuvelier,M. L., Derelle,E., Everett,M. V., Foulon,E., Grimwood,J., Gundlach,H., Henrissat,B., Napoli,C., McDonald,S. M., Parker,M. S., Rombauts,S., Salamov,A., Von Dassow,P., Badger,J. H., Coutinho,P. M., Demir,E., Dubchak,I., Gentemann,C., Eikrem,W., Gready,J. E., John,U., Lanier,W., Lindquist,E. A., Lucas,S., Mayer,K. F., Moreau,H., Not,F., Otillar,R., Panaud,O., Pangilinan,J., Paulsen,I., Piegu,B., Poliakov,A., Robbens,S., Schmutz,J., Toulza,E., Wyss,T., Zelensky,A., Zhou,K., Armbrust,E. V., Bhattacharya,D., Goodenough,U. W., Van de Peer,Y., and Grigoriev,I. V. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. *Science*, 324 (5924):268–272, Apr 2009.

Xu,P., Kong,Y., Song,D., Huang,C., Li,X., and Li,L. Conservation and functional influence of alternative splicing in wood formation of Populus and Eucalyptus. *BMC Genomics*, 15: 780, Sep 2014.

Yang,X., Zhang,H., and Li,L. Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis. *Plant J.*, 70(3):421–431, May 2012.

Zhai,J., Jeong,D. H., De Paoli,E., Park,S., Rosen,B. D., Li,Y., Gonzalez,A. J., Yan,Z., Kitto,S. L., Grusak,M. A., Jackson,S. A., Stacey,G., Cook,D. R., Green,P. J., Sherrier,D. J., and Meyers,B. C. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.*, 25(23):2540–2553, Dec 2011.

Zhou,R., Moshgabadi,N., and Adams,K. L. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc. Natl. Acad. Sci. U.S.A.*, 108(38):16122–16127, Sep 2011.

A. Anexo

Transcript profiling of common bean nodules subjected to oxidative stress

Mario Ramírez[a,*], Gabriel Guillén[b], Sara I. Fuentes[a], Luis P. Íñiguez[a], Rosaura Aparicio-Fabre[a], David Zamorano-Sánchez[a], Sergio Encarnación-Guevara[a], Dario Panzeri[c], Bianca Castiglioni[c], Paola Cremonesi[c], Francesco Strozzi[d], Alessandra Stella[c,d], Lourdes Girard[a], Francesca Sparvoli[c] and Georgina Hernández[a]

[a]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad 1001, Cuernavaca, Mor. 62209, Mexico
[b]Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Cuernavaca, Mor. 62209, Mexico
[c]Istituto di Biologia e Biotecnologia Agraria, CNR, Via Bassini 15, 20133, Milano, Italy
[d]Parco Tecnologico Padano, Statistical Genomics and Bioinformatics Unit, Via Einstein Loc. Cascina Codazza, 26900, Lodi, Italy

Several environmental stresses generate high amounts of reactive oxygen species (ROS) in plant cells, resulting in oxidative stress. Symbiotic nitrogen fixation (SNF) in the legume–rhizobia symbiosis is sensitive to damage from oxidative stress. Active nodules of the common bean (*Phaseolus vulgaris*) exposed to the herbicide paraquat (1,1′-dimethyl-4,4′-bipyridinium dichloride hydrate), which stimulates ROS accumulation, exhibited reduced nitrogenase activity and ureide content. We analyzed the global gene response of nodules subjected to oxidative stress using the Bean Custom Array 90K, which includes probes from 30 000 expressed sequence tags (ESTs). A total of 4280 ESTs were differentially expressed in stressed bean nodules; of these, 2218 were repressed. Based on Gene Ontology analysis, these genes were grouped into 42 different biological process categories. Analysis with the PathExpress bioinformatic tool, adapted for bean, identified five significantly repressed metabolic pathways related to carbon/nitrogen metabolism, which is crucial for nodule function. Quantitative reverse transcription (qRT)-PCR analysis of transcription factor (TF) gene expression showed that 67 TF genes were differentially expressed in nodules exposed to oxidative stress. Putative *cis*-elements recognized by highly responsive TF were detected in promoter regions of oxidative stress regulated genes. The expression of oxidative stress responsive genes and of genes important for SNF in bacteroids analyzed in stressed nodules revealed that these conditions elicited a transcriptional response.

Introduction

Reactive oxygen species (ROS), such as singlet oxygen ($_1O_2$), superoxide anion radical ($O_2{}^-$), hydroxyl radical (OH[·]) and hydrogen peroxide (H_2O_2), are generated as inevitable by-products of aerobic metabolism. The dynamic balance between reactions that generate ROS existing in every aerobic cell, stimulates nonspecific oxidations and antioxidants. A change in this balance in favor of oxidative reactions is known as oxidative stress (Sies 1991). Because ROS react with a wide range of biomolecules, such as nucleic acids, proteins and lipids, that are essential for maintenance of the integrity of cellular structures, they may cause

---

Abbreviations – APX, ascorbate peroxidase; EST, expressed sequence tag; MDA, malondialdehyde; ROS, reactive oxygen species; SNF, symbiotic nitrogen fixation; SOD, superoxide dismutase; TF, transcription factor.

irreversible damage that could subsequently lead to tissue necrosis and eventually kill the organism (Op den Camp et al. 2003, Apel and Hirt 2004). A large variety of environmental stresses, such as drought, salinity, chilling, high light intensity and metal toxicity, can cause an increase in the generation of ROS. In the majority of aerobic organisms, there is a need to effectively remove the ROS generated as a result of environmental stress. To control the amount of ROS with the purpose of protecting the cells from oxidative damage, plants have developed a complex antioxidant defense system to scavenge ROS. This antioxidant system includes various enzymes and nonenzymatic metabolites that destroy ROS that are produced in excess from those normally required for metabolism (Inze and Van Montagu 1995, Vranova et al. 2002, Minchin et al. 2008). Nonenzymatic antioxidants include mainly ascorbate and glutathione (γ-glutamylcysteinylglycine), as well as tocopherol, flavonoids, alkaloids and carotenoids. Plant ROS scavenging enzymes include superoxide dismutase (SOD), ascorbate peroxidase (APX), glutathione peroxidase and catalase. Although ROS were initially recognized as hazardous derivatives of aerobic metabolism that need to be removed, it subsequently became apparent that they are key actors in regulating development of the plant, responses to stress and programmed cell death. It is now generally accepted that ROS plays an important signaling in diverse cellular mechanisms (Neill et al. 2002).

Legumes can establish symbioses with bacteria of the *Rhizobiaceae* family to capture atmospheric nitrogen directly and thereby to support plant growth. The symbiotic nitrogen fixation (SNF) occurring in the legume–rhizobia symbiosis is the main source of nitrogen in agro-ecosystems. SNF takes place in specialized rhizobia-induced legume root nodules and involves extensive recognition and a tight association between the two symbionts. The involvement of ROS in the initial stages of the *Rhizobium*–legume symbiosis has been outlined; ROS is essential for the development of an optimal symbiosis, pointing to a signaling role for ROS (Mandon et al. 2009). On the other hand, the root nodules of legumes are sites of strong biochemical activity and therefore are at increased risk of harm from ROS generation (Dalton et al. 2009, Becana et al. 2010). O_2^- and H_2O_2 are generated in nodules by high respiratory rates necessary to support SNF, the autoxidation of the oxygenated form of leghemoglobin, and the oxidation of various proteins with strong reducing potential (e.g. nitrogenase, ferredoxin and hydrogenase). To cope with oxidative stress during the symbiotic interaction, both the host plant and rhizobia have at their disposal a set of antioxidant tools

(Becana et al. 2000, Matamoros et al. 2003). However, modification of redox homeostasis and a decline in antioxidative defense, which are both triggered by ROS accumulation, greatly influence the metabolism of the nodule and SNF. Drought, salinity, high temperature and metal toxicity are common abiotic stress conditions of the tropical semiarid regions in which legume crops are grown and stimulate changes in redox homeostasis and a decline in antioxidative defense. In this work, we sought to further our understanding of the global response to oxidative stress in the common bean (*Phaseolus vulgaris*)–*Rhizobium tropici* symbiosis.

The most important legume for direct human consumption worldwide is common bean. SNF and bean crop production are affected by environmental factors, such as drought, salinity and nutrient stress that may lead to oxidative stress (Broughton et al. 2003). Our understanding of how common bean plants respond to and cope with abiotic stresses, such as drought and salinity, is improving due to the research of several groups (Serraj et al. 1998, Verdoy et al. 2004). However, much less is known about the response of common bean to metal toxicity. Several areas in which bean is grown are characterized by acidic soils and low drainage, which favor the increased availability of certain metals, including aluminum (Al), copper (Cu) and manganese (Mn). Mn toxicity stimulates ROS production and generates oxidative stress; it negatively affects photosynthesis, respiration and the homeostasis of Cu, iron (Fe) and zinc (Zn), which are vital nutrients in the legume–rhizobia symbiosis (González and Lynch 1997, Yang et al. 2009). We recently identified microRNAs that are induced or repressed in nodules of bean plants exposed to Mn stress (Valdés-López et al. 2010). In this work, we proposed to identify the transcription factors (TFs) that are regulated in common bean nodules under Mn toxicity.

In this work, we analyzed the response of mature common bean nodules exposed to the herbicide paraquat (PQ), a ROS-inducing agent that mimics the oxidative stress resulting from abiotic stresses. PQ is a widely used herbicide generally applied to shoots; its mode of action is strengthened by light during the process of photosynthesis in plants (Donahue et al. 1997, Iturbe-Ormaetxe et al. 1998). Inside leaf cells, PQ is reduced by chloroplasts to its radical cation and upon auto-oxidation H_2O_2 and several oxygen radicals are formed. These radicals are phytotoxic, causing chlorophyll destruction and lipid peroxidation. However, to analyze the in vivo effect of oxidative stress on nodule metabolism, we applied PQ directly to the nodulated roots of common bean plants. This approach is similar to the one reported by Dalton (1992) for *Glycine max* (soybean) nodules and Marino et al. (2006) for *Pisum sativum* (pea) nodules.

77

Most of the antioxidants, both enzymatic and nonenzymatic, in legume nodules are similar to those in other plant tissues or organs; however, concentrations are often higher, which suggests an important link between SNF and antioxidants (Dalton 1995, Becana et al. 2000, Matamoros et al. 2003, Minchin et al. 2008, Becana et al. 2010). On the other hand, rhizobia, like other bacteria, possess certain ROS-scavenging enzymes, including catalase, SOD, peroxidases and peroxiredoxin (Crockford et al. 1995, Ohwada et al. 1999, Santos et al. 2000, Ampe et al. 2003, Djordjevic et al. 2003, Vargas et al. 2003, Dombrecht et al. 2005). There are few reports on the effect of oxidative stress on rhizobia transcript profiles and these mainly correspond to free-living conditions. These include the genome-wide transcript profiles of *B. japonicum* exposed to H_2O_2 or PQ (Donati et al. 2011, Jeon et al. 2011). Martinez-Salazar et al. (2009) reported a transcriptional profile analysis of *R. etli* that suggested that RpoE4, an ECF σ factor, is an important general regulator involved in the responses to several stresses.

Plants exhibit specific transcriptomic responses to oxidative stress. Several studies have indicated that according to the type of ROS or its subcellular site of production (plastid, cytosol, peroxisome or apoplast), a different physiological, biochemical and molecular response is caused and consequently results in a clear reprogramming of the plant's transcriptome (Mittler et al. 2004, Gadjev et al. 2006). In *Arabidopsis thaliana*, genome-wide microarrays have provided the means to analyze how expression profiling are affected by rising ROS levels (Op den Camp et al. 2003, Gechev et al. 2004, 2005, Vandarauwera et al. 2005). Transcriptomic footprints have been produced for plants with compromised levels of ROS-scavenging enzymes, such as chloroplastic Cu/Zn SOD, APX cytosolic and peroxisomal catalase mitochondrial alternative oxidase (Davletova et al. 2005b, Umbach et al. 2005, Vandarauwera et al. 2005), as well as for plants treated with agents that generate ROS (Gechev et al. 2004, 2005). While transcriptomic approaches, through macroarray and microarray analyses, have allowed the identification of a great number of genes that are differentially expressed in nodules of different legumes such as *Medicago truncatula, Lotus japonicus*, soybean and common bean (Colebatch et al. 2002, 2004, El Yahyaoui et al. 2004, Kouchi et al. 2004, Lee et al. 2004, Asamizu et al. 2005, Ramírez et al. 2005, Starker et al. 2006, Brechenmacher et al. 2008, Hernández et al. 2009), there is no information on the transcriptional profile of legume nodules exposed to oxidative stress.

On this basis, we hypothesized that the global response of common bean nodules to oxidative stress

occurs at the transcriptional level and aimed to decipher the extent of transcript changes derived from ROS accumulation in both symbionts. To this end, we carried out a microarray-based transcript profiling of PQ-treated common bean nodules, elicited by *Rhizobium tropici*, using the Bean Custom Array 90K designed by our group. In addition, we analyzed the expression of rhizobial genes that had previously been reported to be differentially expressed under stress conditions and of transcriptional regulators and enzymes that control nitrogen fixation.

The metabolomic and transcriptomic profile of *Arabidopsis* cells exposed to oxidative stress revealed that this treatment had a profound effect on central metabolic pathways, such as the tricarboxylic acid cycle and amino acid metabolism (Baxter et al. 2007). In the case of legumes, the decrease in SNF caused by environmental stresses is intimately related to cellular carbon/nitrogen metabolism, which is key to nodule function and to efficient SNF. Therefore, we conducted a transcriptomic analysis of PQ-stressed nodules to identify metabolic pathways that might be affected by oxidative stress. MapMan (Thimm et al. 2004), PathExpress (Goffard and Weiller 2007, Goffard et al. 2009), and Gene Ontology (The Gene Ontology Consortium et al. 2000) bioinformatics tools adapted to common bean were used to interpret the microarray gene expression data. PathExpress allowed us to identify differentially expressed genes that were assigned EC numbers and were thus associated with the relevant metabolic pathways that operate in nodules under oxidative stress. Furthermore, we hypothesized that TFs function as global regulators of the nodule's response to ROS accumulation. Specific TFs and other regulatory elements of the transcriptomic response of Arabidopsis to oxidative stress are beginning to be uncovered (Davletova et al. 2005a, Baxter et al. 2007). We have identified TFs that respond to phosphorus deficiency in common bean nodules (Hernández et al. 2009). In this work, we used our qRT-PCR-based TF profile platform (Hernández et al. 2007) to identify TFs that are involved in the regulation of gene expression in bean nodules subjected to oxidative stress. In addition we analyzed the sequences of promoter regions of selected PQ responsive genes to identify putative TF-binding cis-elements.

Materials and methods

Plant and bacterial growth conditions

Surface sterilization of *Phaseolus vulgaris* cv. Negro Jamapa 81 seeds was performed with a sodium hypochlorite solution (10%) for 10 min and seedlings were germinated at 30°C on moist sterile filter paper.

Three-days-old seedlings were planted in pots with vermiculite, as substrate, and then inoculated with 1 ml bacterial suspension (approximately 10^7 cells) per plant. The inoculum was prepared from an overnight culture of *Rhizobium tropici* CIAT 899 grown in PY medium at $30°C$. Pots were watered 3 days per week with a nitrogen-free nutrient plant solution (Summerfield et al. 1977). Nodulated plants were grown in a naturally lit greenhouse with a controlled temperature ($26–28°C$) for 18 days, at which point the nodules reached maturity. Different stress treatments known to generate oxidative stress were tested in mature common bean nodules. For PQ treatment, at 16 days post-inoculation (dpi) plants were watered with 1 mM 1,1′-dimethyl-4,4′-bipyridinium dichloride hydrate, the herbicide PQ, for 48 h. For chilling stress, plants (at 18 dpi) were incubated at $4°C$ for 4 h and then the nodules were harvested. For salinity stress, plants (16 dpi) were watered with nutrient solution supplemented with 150 mM of NaCl at 0 and 12 h and were harvested at 12 and 24 h. Bean plants were subjected to drought stress for 1 week by stopping watering at 15 dpi. To generate Mn toxicity conditions, plants were inoculated and watered with nutrient plant solution supplemented with 400 μM MnCl for 18 days. Mature nodules from stressed and control plants were harvested and immediately frozen in liquid nitrogen and stored at $-80°C$ until used.

Lipid peroxidation, ROS detection, nitrogenase activity and ureide content

Lipid peroxidation, an indicator of cell membranes damage by high concentrations of ROS, was assessed from 18-dpi nodules harvested from stressed or control bean plants in three independent experiments (three replicates per experiment) as reported (Heath and Packer 1968, Du and Bramlage 1992). MDA equivalents were calculated using an extinction coefficient of 155 M^{-1} cm^{-1}. ROS accumulation in PQ-treated nodules, as compared to control nodules, was assessed by fluorescence activity assays. Nodules (16 dpi) treated with PQ for 24 h and nontreated nodules were further incubated in 10 μM 2′,7′-dichlorodihydro-fluorescein diacetate (DFCH-DA) (Invitrogen, Carlsbad, CA) for 6 h. After a brief wash in distilled water, the nodules were observed using a Carl Zeiss AXIOSKOPZ fluorescence optical microscope. All fluorescence images were captured using a Canon Power Shot G5 digital camera. DFCH-ROS complexes present in the nodules of bean plants were quantified based on fluorescence intensity using the NIH IMAGEJ software program (http://rsbweb.nih.gov/ij/). Student's *t*-tests were performed to detect significant differences in transcript levels ($P < 0.05$).

Nitrogenase activity was determined by measuring the acetylene reduction activity of 18-dpi-nodulated roots from 15 plants per condition and three independent experiments. Ethylene production was determined by gas chromatography in a Variant model 3300 chromatograph as reported (Ramírez et al. 1999). Specific activity is expressed as nmol ethylene h^{-1} g^{-1} nodule DW.

The ureide content was analyzed in nodules and in xylem exudates from 18-dpi PQ-treated and control bean plants. For nodule assays, 200 mg FW of tissue was used and three independent experiments with three biological replicates each were performed. Xylem exudates were collected for 1 h from the cut stems of 10–15 plants in three independent experiments. Stem collection was done from 12:00 to 13:00 h, and plants were watered 1 h before samples were collected. After recording the volume of the exudates, the collected sap was stored at $-20°C$ until used for analysis. The ureide content was determined using the differential analysis reported by Vogels and Van Der Drift (1970).

Bean microarray design and hybridization

The bean microarray was printed using a platform with a custom 90K array layout at the Plant Functional Genomic Center of the University of Verona, Italy. To define the layout, an in-house bioinformatics pipeline was created to collect, compare and filter bean, and soybean RNA sequences available from the DFCI Bean Gene Index (http://compbio.dfci.harvard.edu/tgi, version 3.0) with 21 497 total unique sequences and 33 001 reference soybean genes from the NCBI UniGene build 38.0 (http://www.ncbi.nlm.nih.gov/unigene). Because the bean sequences could not be considered as a complete set in terms of representation of genes and transcripts, these were used as references and the soybean UniGenes were added to generate a larger dataset. Duplicated sequences were then removed using a homology search pipeline with BlastN, to define minimally redundant datasets of transcripts. After this step, all the sequences in the dataset were processed to design probes of 35–40 nucleotides, according to the OligoArray 2.0 parameters (Rouillard et al. 2003). These probes were subsequently filtered to avoid cross-hybridizations. The final layout accounted for 18 867 unique bean sequence probes and 11 205 soybean UniGene probes, along with positive and negative controls to yield a total of 30 150 different probes on the microarray. Each probe was printed in triplicate to ensure the presence of internal replicates and to provide a good statistical representation of each transcript on the array. The microarray was named Bean Custom Array 90K.

Total RNA from 18-dpi PQ-treated and control bean nodules was isolated using Trizol reagent (Life Technologies, Carlsbad, CA) as reported (Hernández et al. 2007). Total RNA (1 µg) was used as template to synthesize antisense RNA (aRNA) with Cy5-ULS, using the RNA Amplification and Labelling Kit from CombiMatrix (ampULSe, Kreatech Biotechnology, Amsterdam, the Netherlands), according to the manufacturer's instructions. Prehybridization was made by incubating the arrays with prehybridization solution (6X SSPE, 0.05% Tween-20, 20 mM EDTA, 5× Denhardt's solution, 100 ng µl^{-1} salmon sperm DNA, 0.05% SDS) for 30 min at 45°C. Labeled aRNA (4 µg) was fragmented by incubation with 5.2 µl of fragmentation solution (200 mM Tris-acetate pH 8.1, 500 mM KOAc, 150 mM MgOAc) for 20 min at 95°C.

Hybridization was done at 45°C for 16 h in hybridization solution (6X SSPE, 0.05% Tween-20, 20 mM EDTA, 25% diformamide, 100 ng µl^{-1} salmon sperm DNA, 0.04% SDS). After hybridization and washing, the microarray was dipped in imaging solution, covered with LifterSlip™, and then scanned using a GenePix 4000B microarray scanner (Axon, Toronto, Canada) and the accompanying acquisition software (CombiMatrix Microarray Imager Software). Each hybridization was scanned multiple times at different photomultiplier (PMT) settings. Arrays were stripped and re-hybridized using the CustomArray Stripping Kit for 90K CombiMatrix (ampULSe, Kreatech Biotechnology, Amsterdam, the Netherlands), following the protocols of the manufacturer. Each array was used up to four times without showing a deterioration in signal or increase in background.

Bean custom array 90K data analysis

The raw intensity data were first processed using COMBIMATRIX MICROARRAY IMAGER software. This program allows the visual inspection of the entire microarray slide and is used to check the quality of each spot and, if needed, to perform possible corrections (i.e. for dust or scratches on the surface). Intensity data were then exported into the Feature and Probe format of CombiMatrix, where the actual raw intensity per probe and per spot is stored. Data were loaded into R (http://www.r-project.org) and analyzed using the Limma package (Smyth 2004). The median value of each spot on the array was determined and the probes were filtered to remove quality and factory controls. Within-array probe replicates were defined as technical replicates, and the mean intensity across different probes was calculated. The probe intensity values of each biological condition were then normalized using the quantile function of Limma. The values present in the expression matrix were transformed in log$_2$ and a design matrix was defined to describe the biological samples. The expression matrix was then used to fit a linear model using the design matrix and the functions of Limma. A set of contrast matrices was defined to describe the comparisons among samples in the experiment and a second linear fit was performed for each contrast. Errors were corrected using the Bayesian functions of Limma and the list of differentially expressed genes was generated for each contrast after correcting for multiple testing using the Benjamin-Hochberg method and setting 0.05 as the adjusted P-value cutoff.

For the purpose of interpreting the biological significance of gene expression data three bioinformatics-based approaches were used for analyses. First, the statistically significant array data (Table S1) were organized in functional categories according to Gene Ontology (GO) guidelines (The Gene Ontology Consortium et al. 2000). Fisher's exact test (Routledge 1998) was applied to determine which GO categories were statistically over-represented in each set of differentially expressed ESTs that was induced or repressed under oxidative stress. Second, an expression data analysis was performed using MAPMAN software version 3.5.1 (Thimm et al. 2004, http://gabi.rzpd.de/projects/MapMan/). To extend MAPMAN to common bean, a *Phaseolus vulgaris* map was developed and uploaded to MAPMAN. The fold change PQ/control of the induced or repressed common bean genes, expressed in log$_2$, was used to visualize the expression data. We used the PathExpress web-based tool (Goffard and Weiller 2007, Goffard et al. 2009), to identify the most relevant metabolic pathways related to the subsets of differentially expressed bean genes. To adapt PathExpress to common bean, we defined the orthologous gene from soybean and its Affymetrix ID for each differentially expressed common bean EST (Table S1) and used the Affymetrix Soybean Genome Array (*Glycine max*) data set from the PathExpress web-based tool. ESTs associated with metabolic pathways or subpathways that were statistically over-represented ($P \leq 0.05$) were detected amongst the differentially expressed sets of ESTs.

Quantitative RT-PCR (qRT-PCR) analysis

A qRT-PCR approach was used to validate the microarray data and to generate a transcript profile of previously reported bean TF genes (Hernández et al. 2007). Total RNA was isolated from 200 mg of frozen nodules as described above, for qRT-PCR. Three biological replicates were carried out for each condition, using PQ-treated and control nodules, and RNA was extracted from

differing sets of plants grown under similar conditions. A NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) was used to measure RNA concentration. Genomic DNA degradation, cDNA synthesis and quality verification for qRT-PCR were done as reported (Hernández et al. 2007). Reactions were done in a 96-well format with the 7300 Real-Time PCR System and 7300 System Software (Applied Biosystems, Foster City, CA). In each qRT-PCR run, the APX (TC32379) coding sequence was included as a marker gene of the PQ response in bean nodules. APX is an integral component of the glutathione-ascorbate cycle, and its transcript level is remarkably increased in response to the presence of ROS (Yoshimura et al. 2000). The gene encoding APX was induced in nodules exposed to oxidative stress (ca. three average expression ratio), thus confirming the oxidative stress status of the nodules under study. UBC9 (TC34057), which was constant in the tested conditions, was included for normalization in every qRT-PCR run. Average expression ratios (PQ/C) were calculated with the $\Delta\Delta C_T$ method as reported (Hernández et al. 2007) and the fold change value ($\log_2$) was calculated. Student's t-test was performed with a P-value cutoff of 0.05.

To validate the microarray data, 30 genes were selected that had been identified by microarray analysis as being induced or repressed in nodules exposed to oxidative stress. For each qRT-PCR reaction, 100 ng of RNA free of genomic DNA were used as template. Table S5 shows the sequences of the qRT-PCR primer pairs used. Relative transcript levels of selected genes were quantified using the Power SYBR Green RNA – to C_T 1 Step Kit (Applied Biosystems Foster City, CA).

TF profiling, based on qRT-PCR, was performed to determine the differential expression of TF genes in nodules exposed to oxidative stress. The profile included a set of 372 bean TF genes previously identified (Hernández et al. 2007). Primer pairs and qRT-PCR conditions used have been reported (Hernández et al. 2007).

Gene expression of bacteroids from PQ-stressed and control nodules was analyzed by qRT-PCR. Total RNA isolated from bean nodules was used to synthesize bacteroid cDNA using the RevertAid™ H Minus First Strand cDNA Synthesis Kit and random hexamer primers (Fermentas, Copenhagen, Denmark), following the manufacturer's instructions. Specific primers of selected genes known to be related to the bacterial response to oxidative stress, such as peroxidases, superoxide dismutase and glutathione peroxidase, and of some key genes involved in SNF were designed using the genome sequence of *R. etli* CFN42 (González et al. 2006; Table S6). The equipment and conditions used for bacteroid qRT-PCR analysis were similar to those used for plants, as described above.

In silico analysis of *cis*-elements in the promoter regions of responsive genes

The sequences of 63 genes corresponding to the promoter regions, which were over-represented in different metabolic pathways, were obtained by performing a BLAST search against the bean genome deposited in Phytozome (www.phytozome.net, Goodstein et al. 2012). Putative *cis*-elements present in the promoter sequences, defined as the 2-kb upstream region, of selected genes were examined using the Plant Promoter Analysis Navigator (PlantPAN), which identifies transcription factor binding sites in a group of gene promoters (Chang et al. 2008).

Results

Symbiotic and oxidative stress response phenotype

In this work, we analyzed the response of common bean nodules to oxidative stress after exposure to the herbicide known as PQ. PQ has been widely used to evaluate the effects of ROS in plants as it mimics the action of environmental stresses. The selected PQ concentration (1 mM) is similar to that reported by Dalton (1992) used to generate oxidative stress in soybean nodules. We measured lipid peroxidation by monitoring malondialdehyde (MDA) content, which is among the most widely accepted assay for oxidative damage in plants (Shulaev and Oliver 2006). Lipid peroxidation in effective common bean nodules (18 dpi) exposed to PQ for 6–48 h increased over time and peaked at 48 h, when the MDA content was threefold that in nodules of plants grown under control conditions (Fig. 1A). This result was confirmed by analyzing ROS content in nodules incubated with 2′,7′-dichlorodihydro-fluorescein diacetate and subsequently observed by fluorescence microscopy. A twofold increase ($P \leq 0.05$) in fluorescence intensity was observed in nodules exposed to PQ for 48 h as compared to control nodules (data not shown) (Fig. 1B). We tested if PQ treatment had the same effect as actual oxidative stress arising from environmental stresses by comparing the extent of ROS production in bean nodules under chilling, salinity, drought and Mn toxicity stresses. The levels of MDA observed in nodules exposed to environmental stresses ranged from 2.7- to 3.5-fold with respect to those in the control nodules (Fig. 1A). These values were similar to those obtained in nodules exposed to PQ for 48 h (Fig. 1A). Therefore, we reasoned that a 48-h PQ (1 mM) treatment mimics the

81
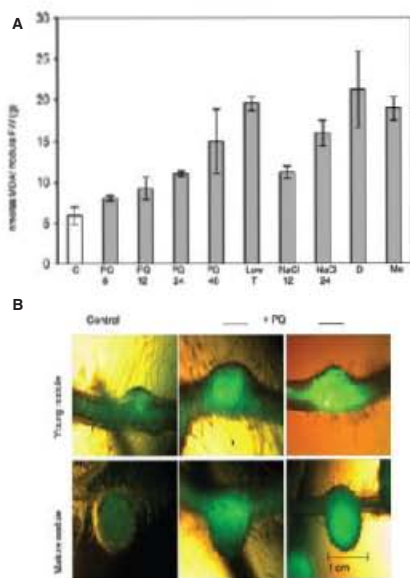
Fig. 1. Lipid peroxidation and ROS detection in common bean nodules. (A) Lipid peroxidation, in terms of malondialdehyde (MDA) content, in mature common bean nodules exposed to various stress treatments known to induce high levels of ROS production, with respect to that in control nodules. PQ (6, 12, 24 and 48 h), Low T (4°C, 4 h), NaCl (150 mM of NaCl, 12 and 24 h), D (drought, water withheld for 1 week), and Mn (400 μM MnCl for 18 days). Values are mean ± SE of three independent experiments. (B) ROS detection in PQ-treated bean nodules (24 h) using 2′,7′ dichlorodihydrofluorescein-diacetate (DCFH-DA).


Fig. 2. The effect of oxidative stress (PQ treatment) on bean plants in symbiosis with *Rhizobium tropici*. Nitrogenase activity (A) and ureide content (B, C) were determined in 16-dpi plants treated with PQ for 48 h (gray bars) and in 18-dpi control (nontreated) plants (white bars). (A) Specific nitrogenase activity, as determined by an acetylene reduction assay. Ureide content in xylem exudates (B) and mature nodules (C). Values are mean ± SE of three independent experiments.

actual oxidative stress achieved in common bean plants growing under environmental stress conditions similar to those occurring in nature and so this treatment regimen was chosen for this work.

The symbiotic phenotype was assessed in 18-dpi bean plants under PQ treatment, by determining the nitrogenase activity and ureide content of the plants. The specific nitrogenase activity of PQ-treated nodules was 48% lower than that of control nodules (Fig. 2A). Ureides are the main N-compounds transported from bean nodules to the aerial parts of the plant (Schubert 1981, Boldt and Zrenner 2003). In the xylem sap of control plants, the ureide content was 2.5-fold more than in the xylem of plants grown under conditions of oxidative stress (Fig. 2B). In the nodules of plants with increased ROS production, the ureide content was around 1.5-fold lower (Fig. 2C). Overall, these results indicate that SNF is negatively affected in PQ-exposed

nodules and this directly prevents optimum ureide synthesis and transport within the plant.

Microarray analysis of the nodule's response to PQ-induced oxidative stress

In contrast to our previous work using macroarrays printed with ca. 2000–3000 bean ESTs (Ramírez et al. 2005, Hernández et al. 2009), we now used, for the first time, a microarray approach for bean nodule transcriptome analysis to obtain global insight into the response of the expression of genes of nodules subjected to oxidative stress as compared to control nodules. For this analysis, we designed a Bean Custom Array 90K, which includes a 30K unigene set from common bean (ca. 18 000 ESTs) and nonredundant soybean (*Glycine max*) genes (ca. 11 000 ESTs). The data reported in this paper were depositeds in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are available through GEO Series accession number GSE37638 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37638). To verify that the microarray data were reliable, the normalized median signal intensities resulting from control samples were compared through all the repetitions (i.e. two biological replicates and three technical replicates of each biological sample). A high correlation coefficient, ranging from 0.94 to 0.96, was observed between control samples that indicates a low technical and biological variation among repetitions. Following statistical analysis of the data, a total of 4280 ESTs were defined as being differentially expressed (corrected P values < 0.05; signal intensity ≥ 1.5-fold) in response to ROS accumulation induced by PQ. From among these genes, 34% corresponded to *G. max* ESTs included in
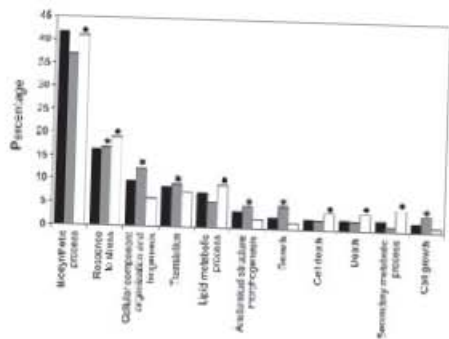
Fig. 3. Biological processes significantly over-represented in the set of induced (●) or repressed (*) ESTs (Fisher exact test $P < 0.05$), according to Gene Ontology analysis (The Gene Ontology Consortium et al. 2000). Black bars, percentage of ESTs corresponding to each biological process within the microarray.

the microarray (see Table S1). More ESTs were repressed (2218) than induced (2062).

The induced or repressed ESTs as determined by microarray analysis were organized in functional categories according to Gene Ontology (GO) guidelines (The Gene Ontology Consortium et al. 2000). Of the 4280 probe sets differentially expressed upon PQ treatment, we assigned at least one GO term to 3822 probe sets based on sequence similarity. Differentially expressed ESTs were classified into 42 different biological processes (see Fig. S1). Fig. 3 shows the six GO categories statistically over-represented within the induced EST data set and the six GO categories that were over-represented in the repressed EST data set.

To further analyze the gene expression data, we also used the PathExpress (Goffard and Weiller 2007, Goffard et al. 2009) bioinformatic tool adapted to common bean to identify the differentially expressed genes associated with the relevant metabolic pathways operating in bean nodules under oxidative stress. Table 1 shows the pathways significantly induced or repressed in PQ-treated bean nodules. Several of significantly induced pathways and their corresponding enzymes have been implicated in defense processes for different types of abiotic stresses. Most of the significantly repressed pathways were related to carbon/nitrogen metabolism, which are relevant pathways for the optimal functioning of legume nodules (see Table S4). Fig. 4 illustrates three relevant pathways that were significantly repressed in bean nodules under oxidative stress: carbon fixation, purine metabolism and starch and sucrose metabolism. Our microarray analysis showed that several enzymes from each of these pathways were significantly repressed.

Validation of microarray results by real-time qRT-PCR

To confirm the results obtained from microarray analysis, we performed quantitative reverse transcription (qRT)-PCR on three biological replicates and three technical replicates of each biological sample. Thirty genes showing various degrees of repression or induction in the microarray analysis were chosen for validation (Table 2; see Table S5). We found a good correlation ($r^2 = 0.86$) and a similar pattern between the microarray and qRT-PCR fold-change values for each gene, indicating the reliability of our microarray data. The variation in the fold-change values between the microarray and qRT-PCR gene expression data is probably due to the different sensitivities of the two methods.

Some of the genes selected for the validation correspond to enzymes belonging to metabolic pathways that were repressed in nodules exposed to oxidative stress, as revealed by PathExpress analysis. For all of these genes, the repressed fold-change value from microarrays was confirmed by qRT-PCR (Table 2, Fig. 4).

qRT-PCR-based TF transcript profiling

We used the reported qRT-PCR platform for common bean TF expression profiling (Hernández et al. 2007) to identify the differential expression of 372 bean TF genes in nodules subjected to oxidative stress. Table 3 shows that 67 TF genes from 26 TF gene families (Riechmann 2002) were differentially expressed, twofold or more ($P \leq 0.05$), in nodules treated with PQ. Among these, 18 TF genes from 13 families were repressed and 49 TFs corresponding to 22 families were induced. The qRT-PCR expression analysis validated the microarray data of 10 TF genes (see Table S2); the remaining 57 responsive TF genes (Table 3) showed no significant differential expression upon PQ treatment after our statistical analysis of microarray data. The latter finding may be related to the higher sensitivity and accuracy of the qRT-PCR platform for TF expression profiles. The TF gene families that had more members that were differentially expressed in nodules exposed to oxidative stressed are as follows: the MYB superfamily (seven members); the AP2/EREBP family (seven members); and the GRAS, C2C2(Zn), and C2H2(Zn) families (four members each). In addition, we performed a TF expression profile analysis in common bean nodules subjected to Mn toxicity, as this condition also generates ROS accumulation (Fig. 1). We found that 28 of 372 TF genes were differentially expressed in bean nodules exposed to Mn toxicity, and 21 of these were induced (see Table S3).

Pathway/subpathway	No. of enzymes included in bean EST sequences	No. of differentially expressed enzymes (TC or Gene Bank Accession)	P
Repressed			
Carbon fixation	22	10 (BQ481899, NP7938839, TC14212, TC16003, TC8682, TC10742, CB541583, FG232722, BQ481541, BQ481565)	3.69e-03
Sulfur metabolism	9	5 (20TC8860, FD789961, TC18230, TC16782, TC13963)	3.12e-02
Purine metabolism	46	16 (TC13266, TC10993, TC12224, CV535717, FG988543, TC12194, BI970858, TC12218, TC8904, TC13149, CV538017, TC8860, NP7927042, TC16782, TC14212, TC15930)	3.39e-02
Starch and sucrose metabolism	29	11 (BQ481932, TC15905, TC9852, TC13149, FG987269, TC15649, TC10682, TC15408, FE676015, FD797285, CX129651)	4.09e-02
Reductive carboxylate cycle (CO$_2$ fixation)	7	4 (NP7938839, EG948449, TC8682, TC19131)	4.92e-02
Induced			
Isoflavonoid biosynthesis	3	3 (TC24029,CV542400, GW908631)	5.58e-03
Methane metabolism	8	4 (NP7938634, TC14371, TC14097, TC13495)	3.70e-02
Drug metabolism-cytochrome P450	5	4 (TC9537, TC12978, TC9580, FG229649)	4.22e-02
Glycerolipid metabolism	16	6 (TC9580, TC8317, TC10848, FD796300, TC8984, TC15381)	4.88e-02

In silico analysis of *cis*-elements in the promoter regions of responsive genes

To determine whether there is a link between oxidative stress-regulated genes and TF that respond to this type of stress, we did an in silico search for *cis*-elements corresponding to putative for TF binding sites in the promoter regions of the genes over-represented in metabolic pathways (Table 3). The analysis revealed that the group of genes analyzed contained *cis*-motifs matching those of the MYB superfamily, AP2/EREBP, ARF, WRKY, bZIP, CCAAT-box, ZF-HD or bHLH motifs (Table 4). These represent TF families reported in the present work as being differentially expressed in bean nodules exposed to oxidative stress (Table 3). As expected, given that this analysis corresponds to genes expressed in nodules, most of the analyzed genes contained *cis*-elements that correspond to TF binding sites that regulate the nodule gene expression.

Gene expression analysis of the bacteroid response to oxidative stress

We aimed to determine if the expression of transcriptional regulators of genes essential for SNF and of selected genes known to respond in oxidative stress in other rhizobia was significantly affected in *R. tropici* bacteroids from PQ stressed bean nodules (Table 4). Our analysis revealed that a putative *sodB* gene (superoxide dismutase) was significantly overexpressed (with a PQ/control ratio of 95) after PQ treatment. On the other hand, the expression of *katG* and *oxyR* and a putative hydroperoxide reductase, which participate in the response to hydrogen peroxide in different bacteria, was only moderately induced (Table 4; Table S6). Interestingly, an *fnrN*-like gene was overexpressed after PQ treatment. These transcriptional regulators activate the expression of the symbiotic terminal oxidase (*fixNOQP* operon) in rhizobia (Schlüter et al. 1997, Lopez et al. 2001). We evaluated the transcription of a *fixN*-like

84
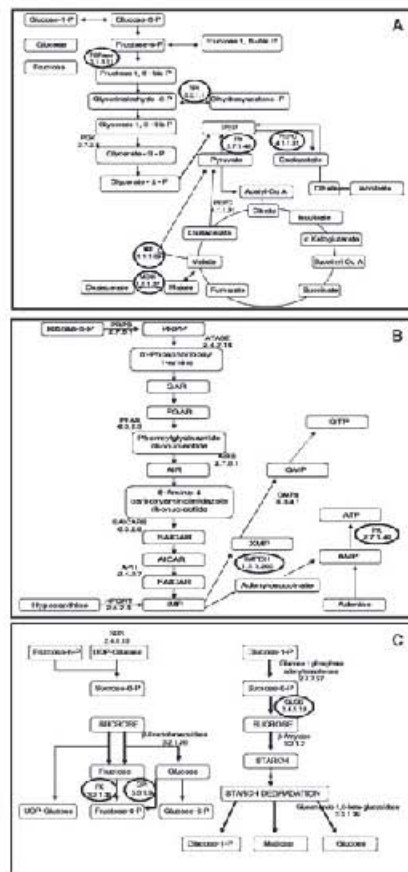
Fig. 4. Selected metabolic pathways identified by PathExpress (Goffard et al. 2009) as being repressed in nodules under oxidative stress. Graphical representations were produced using MapMan. The metabolites are shown in boxes. Those enzymes significantly repressed in each pathway are indicated by their name or abbreviation corresponding EC according to the International Union of Biochemistry Molecular Biology. Circles highlight enzymes with repressed expression ratio values confirmed by qRT-PCR (Table 2). (A) Carbon fixation. (B) Purine metabolism. (C) Starch sucrose metabolism.

gene of *R. tropici* and found that it was only moderately induced. In *R. etli*, it was reported that *stoRd* is a negative regulator of *fixN* and of SNF (Granados-Baeza et al. 2007). Our results revealed the presence of a similar *stoR* gene in *R. tropici*, the expression of which was induced after PQ treatment. Overexpression of this regulator might at least partially account for the drop in SNF

and the reduction in *fixN* expression under these conditions. We found that the transcription of *nifH*, encoding for the iron–protein subunit of the nitrogenase reductase complex, was negatively affected in PQ-treated bacteroids, it was only 6% that in untreated bacteroids. Under our experimental conditions, the expression of NifA, the activator of *nifH* in nitrogen fixing bacteria, was only moderately reduced. We propose that the reduced transcription level of *nifH* might be the result of post-translational modifications in NifA that render the regulator inactive.

Discussion

In this work, we analyzed the response at the transcriptomic level of common bean nodules to oxidative stress, generated by the addition of PQ. The PQ treatment used resulted in a similar degree of ROS generation as that achieved in common bean plants growing under environmental stresses such as salinity, drought, low temperature and Mn toxicity (Fig. 1). The effect on SNF, nodule cell metabolism and whole plant physiology depends on the concentration and time of exposure to PQ (Marino et al. 2006). When applied to the root, PQ translocation to the shoot depends on the capacity of root cells to sequester PQ by accumulating it in the vacuole and then expelling it across the tonoplast and plasma membrane back into the external solution (Lasat et al. 1997). Marino et al. (2006) reported that only a high dose of PQ applied to SNF pea roots affected photosynthesis and reduced chlorophyll content after 48–72 h, suggesting that PQ was translocated to the shoot. Similar as reported by Dalton (1992) for soybean, we applied a low dose of PQ for a short period of time to SNF common bean roots and observed that the chlorophyll content in leaves was similar in stress and control conditions (data not shown). Therefore, we assumed that the translocation of PQ to the shoots was minimal, if occurring at all, and so the treatment used allowed us to characterize the response to ROS production directly in root nodules before any PQ could travel to the shoots, deplete the level of photoassimilates and disrupt nodule function. In addition, the treatment used mimics the action of environmental stresses, but avoids the complex effects that these might have, such as disrupting long distance transport or resulting in a deficiency of nutrients. The deleterious effects detected in PQ treated common bean nodules are in agreement with those reported for other rhizobia–legume symbiosis (Dalton 1992, Minchin et al. 2008, Becana et al. 2010).

Bacteroids from PQ-treated common bean nodules clearly showed a repression of the nitrogenase reductase (*nifH*) gene as well as an induction of a negative regulator

Table 2. Validation of microarray data by qRT-PCR analysis. Selected genes identified as being induced or repressed by microarray analysis in nodules under oxidative stress. Fold change (log$_2$) values are the average of three biological replicates. Normalized values from qRT-PCR and microarray analyses of each gene are shown. Enzymes of significantly over-represented metabolic pathways identified by PathExpress analysis are in bold.

GenBank Accession/TC	Annotation	Fold change	
		qRT-PCR	Microarray
Repressed			
BQ481932	**Fructokinase**	−1.95190115	−0.49457439
CV535717	Formylglycinamide ribonucleotide amidotransferase	−2.46443725	−0.83182409
TC39206	**Phosphoenolpyruvate Carboxylase**	−0.85218044	−0.68288539
TC14212	**Pyruvate kinase**	−1.31417487	−0.4300321
TC32699	**Malate dehydrogenase**	−1.68215314	−0.62119766
FG232722	**Triosephosphate isomerase**	−2.6916894	−0.69059881
TC34973	Aspartate aminotransferase	−0.91494788	−0.5883288
BQ481565	**Fructose 1,6-bisphosphatase**	−3.1841602	−1.88680387
EG948449	ATP-citrate synthase	−1.63412326	−0.86573177
TC32974	Guanine deaminase	−1.25120767	−0.90464125
TC35238	**Malic enzyme**	−1.63218785	−0.77583495
CB541583	Alanine amino transferase	−1.19974818	−0.79701952
TC36935	**Inosine monophosphate dehydrogenase**	−0.85273768	−0.74852143
TC35171	1,4 alpha glucan-branching Enzyme	−0.81570607	−1.19458157
TC41168	**Glucose phosphate isomerase**	−0.43660897	−0.57051891
Induced			
TC33445	Glutathione *S*-transferase	3.25481869	1.83184673
TC40887	Isoflavone 7-*O*-methyltransferase	1.12824712	0.93360918
TC39388	Ferredoxin hydrogenase	1.01432508	0.60352121
TC33776	Trehalose 6-phosphate synthase	1.49464108	0.5152012
FD788437	Heat shock protein	2.541844	1.17823886
TC44951	Heme oxygenase	1.22478938	1.3024448
TC41871	No symbiotic hemoglobin	1.8394217	0.92409108
TC39083	Methylenetetrahydrofolate reductase	1.098865	0.61120015
TC39733	Mannitol dehydrogenase	3.47421458	1.98544248
TC33800	Alcohol dehydrogenase	1.02455883	1.15918073
TC40120	Dihydroflavanol 4-reductase	0.92110651	1.04529227
TC43445	Cys peroxiredoxin	1.80433923	3.98401971
NP7938634	Peroxidase	1.64124339	0.59225671
CB543559	Chalcone synthase	1.85203812	1.0767921
TC38753	Oxophytodienoate reductase	3.23363093	2.58904155

of SNF (*stoR*) (Table 5). The activator of *nifH* in nitrogen fixing bacteria is NifA, we observed that the expression of this regulator was only moderately reduced. It has been suggested that, in rhizobia, this regulator might need to interact with a metal cofactor in order to activate transcription, and that the resulting cluster functions as an oxygen sensor (Fischer et al. 1989, Souza et al. 1999). Thus, we suggest that the reduced transcription of *nifH* might be due to the metal-catalyzed oxidation of NifA, which renders the regulator inactive. Furthermore, the ATP pool required for nitrogenase activity might be affected by an impairment of carbon supply from the shoot, resulting from the downregulation of enzymes involved in carbon metabolic pathways (Table 1, Fig. 4) and/or from the inhibitory effect of PQ on photosynthesis.

Carbon/nitrogen metabolism are important processes in effective legume nodules, because they provide bacteroids with the carbon compounds used to generate ATP and to assimilate ammonium from SNF into organic N-compounds. Environmental stress conditions commonly affect carbon/nitrogen metabolism in the nodules (Vance and Heichel 1991). In PQ treated nodules showed a profound inhibitory effect on the expression of genes involved in the central metabolic pathways, such as tricarboxylic acid metabolism, nitrogen assimilation, glycolysis, sucrose metabolism, amino acid metabolism and purine biosynthesis, such as sucrose synthase (CX12965), malic enzyme (TC10742), pyruvate kinase (TC14212), alanine aminotransferase (CB541583), cysteine synthase (TC18230), serine acetyltransferase (TC13963) and aspartate aminotransferase (AAT, TC16003) genes (Table 1, Fig. 4). These and other enzymes that participate in the tricarboxylic acid cycle they are known to be susceptible indicators of oxidative

86

Table 3. TF families significantly expressed in bean nodules under oxidative stress identified by real-time RT-PCR (see Table S2). TFs that were also found to be significantly induced or repressed in the microarray analysis are underlined. TFs that were also induced or repressed in nodules subjected to Mn toxicity are in bold (see Table S3).

TF family or domain	Induced	Repressed	Total
MYB superfamily	7 (TC45841,**BQ481567**,TC37047,TC32577,CV537484,TC43761,TC34453)	0	7
AP2/EREBP	5 (TC32865,**TC38814**,TC43751,CV536700,CV537333)	2 (TC34846,TC35571)	7
C2H2(Zn)	4 (TC44148,CV534889,CV535367,TC36863)	2 (**TC41414**,TC43607)	6
C2C2(Zn)	4 (TC32961,**BQ481590**,**BQ481651**,TC40362)	1 (TC36074)	5
GRAS	2 (**TC42487**,TC36190)	2 (TC43723,TC43983)	4
Aux/IAA	2 (TC33102,TC42047)	1 (TC39030)	3
WRKY(Zn)	0	3 (TC33590,TC36992,TC43110)	3
CCAAT	2 (**TC35458**,CV533732)	1 (**TC39639**)	3
C3H-type 1(Zn)	3 (CV537094,TC33834,CV536281)	0	3
ARF	3 (BQ481892,**CV534368**,CV539480)	0	3
ZIM	2 (TC37352,TC45159)	0	2
YABBY C2C2(Zn)	2 (TC38230,TC33137)	0	2
TAZ	1 (CV538139)	1 (TC45225)	2
Response regulator receiver	1 (CV542253)	1 (TC38359)	2
Dof-type C2C2(Zn)	2 (TC38448,TC34332)	0	2
bZIP	1 (TC35580)	1 (TC36609)	2
bHLH	2 (CV535953,TC32288)	0	2
Znf_CCHC	1 (CV537577)	0	1
ZF-HD	0	1 (**CB543438**)	1
TUB	0	1 (TC34864)	1
K-box	1 (CV541709)	0	1
HSF	1 (TC32970)	0	1
CBF/NF-Y	1 (TC38784)	0	1
AS2	0	1 (TC41522)	1
SR-rich splicing factor	1 (TC32718)	0	1
Alfin-like	1 (CB542396)	0	1

stress (Verniquet et al. 1991, Baxter et al. 2007). Interestingly, the negative effect of oxidative stress on carbon metabolic pathways is not exclusive to legume nodules. Arabidopsis plants subjected to oxidative stress showed a rearrangement in carbon metabolism that reflects a short-term survival mechanism; genes encoding enzymes involved in starch and sucrose biosynthetic pathways were downregulated (Scarpeci and Valle 2008). Similar results showing that oxidative stress had a negative impact on carbon metabolism in alfalfa nodules were reported by Naya et al. (2007); enzymes involved in carbon metabolism, mainly sucrose synthase, which is essential for SNF in legume nodules, were downregulated.

87

Table 4. Enriched putative *cis*-elements for binding of PQ-responsive TFs (Table 3) identified in the promoter regions of genes from over-represented metabolic pathways (Table 1). Carbon fixation (CF); sulfur metabolism (SM); purine metabolism (PM); starch and sucrose metabolism (SSM); reductive carboxylate cycle (RCC); isoflavonoid biosynthesis (IB); methane metabolism (MM); drug metabolism-cytochrome P450 (DM) and glycerolipid metabolism (GM). Bullets indicate the pathways in which the putative cis-elements are present.

Putative *cis*-elements	Responsive TF	Genes over-represented in metabolic pathways								
		CM	SM	PM	SSM	RCC	IB	MM	DM	GM
WBOXATNPR1	WRKY	•	•	•	•	•	•	•	•	•
RAV1AAT	AP2/ERBF	•		•	•	•	•	•	•	•
SURECOREATSULTR11	ARF	•	•				•			•
CCAATBOX1	CCAAT	•			•		•		•	
Core consensus binding sequence	ZF-HD		•	•						
MYBCORE	MYB				•	•	•	•		
MYB B4	MYB									•
MYB1AT	bHLH				•		•	•	•	
MYB2CONSENSUSAT	bHLH					•				
PREATPRODH	bZIP					•				
GBF5	bZIP					•	•	•		
DPBFCOREDCDC3	bZIP							•	•	

Table 5. *Rhizobium tropici* bacteroids gene expression after PQ treatment. Fold change (log$_2$) values are the average of three biological replicates (Table S6).

Gene	Annotation	Fold change (log$_2$)	P
sodB	Superoxide dismutase protein	6.572	0.012
fnrN	Transcriptional regulator protein, Fnr/CRP family	2.231	0.002
hpr	Hydroperoxide reductase protein	1.056	0.000
fixN	cbb3-type cytochrome c oxidase	0.578	0.027
oxyR	Hydrogen peroxide sensing transcriptional regulator protein	0.686	0.014
katG	Catalase protein	0.614	0.020
stoR	Probable transcriptional regulator (activator) protein, CRP family	3.356	0.003
nifH	Nitrogenase, iron protein	−1.435	0.000
nifA	Transcriptional regulator NifA protein	−0.285	0.009

Ureides, the main nitrogen compound product of SNF, are exported from the nodule to the shoot of common bean plants (Schubert 1981, Boldt and Zrenner 2003). The reduction in ureide content of nodule and xylem sap form PQ-stressed common bean plants (Fig. 2B, C), indicates that the observed reduction in SNF decreased the assimilation of ammonium into nitrogen compounds in the nodule cells, which led to a decrease in ureide transport through the xylem (Alamillo et al. 2010). Genes (Kim et al. 1995) were repressed in PQ-treated bean nodules. The repression of genes encoding key enzymes that regulate the biosynthesis of purines, precursors of ureides (Table 1, Fig. 4), would reduce the activity of enzymes from relevant metabolic pathways thus significantly decreasing the ureide concentration in the nodules.

On the other hand, isoflavone biosynthesis (related to secondary metabolism), methane metabolism, drug metabolism-cytochrome P450 and glycerolipid metabolism pathways were significantly induced in bean nodules treated with PQ (Table 1). Genes induced in the isoflavonoid biosynthesis pathway include those encoding enzymes that are involved in the biosynthesis of phytoalexines, low molecular mass secondary metabolites with antimicrobial activity that are induced in the stress response and are important components of the plant defense repertoire (Ahuja et al. 2012). Representative phytoalexins of *Phaseolus* spp. are isoflavones, isoflavans, pterocarpans (e.g. phaseolin), coumestans and isoflavanones (e.g. kievitone) (Durango et al. 2002). Induced genes from the methane and drug metabolism-cytochrome P450 pathways included: glutathione-*S*-transferases GST (TC9373, TC12978, TC9537), peroxidase (NP7938634), ferredoxin hydrogenase (TC14371), and a *S*-(hydroxymethyl) glutathione dehydrogenase (TC13495), which are involved in the antioxidant response and are as closely related to the oxidative stress response as to that occurring in PQ-treated nodules. GSTs, which form a large gene family in plants are involved in xenobiotic detoxification and ROS scavenging and may remove peroxidized lipids (Becana et al. 2010). The importance of GST in nitrogen fixing nodules was observed in soybean nodules which contain at least 14 isoforms of GST with variable but significant expression levels, moreover silencing the most predominant form (GST9) results in a substantial reduction of nitrogenase activity (Dalton et al. 2009). The induction of glycerolipid metabolism in PQ-stressed nodules (Table 1) might be related to the general response of plants to stress, which involves remodeling of membrane fluidity and the release of α-linolenic acid (18:3) from membrane lipids. The fluidity in membranes is regulated

by changes in unsaturated fatty acid levels, partly by the controlled activity of fatty acid denaturizing enzymes. Adjusting membrane fluidity, through the biosynthesis of glycerolipids and the production of polyunsaturated fatty acids, maintains an appropriate setting for the function of essential integral proteins during stressful conditions (Upchurch 2008). Interestingly, common bean genes of the glycerolipid pathway that are induced upon exposure to PQ (Table 1) include those encoding enzymes responsible for the synthesis of chloroplast membrane lipids, i.e. monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG). In Arabidopsis, an increase in the DGDG:MGDG ratio and fatty acid unsaturation led to an uncharacterized cellular adaptation capacity to respond to drought stress (Gigon et al. 2004). Furthermore, DGDG biosynthesis is stimulated by phosphorus deficiency revealing a biochemical machinery by which plants conserve phosphate by replacing phospholipids with nonphosphorous galactolipid, a process required for survival under phosphate deprivation (Härtel et al. 2000).

Little is known about the involvement of TFs in legume nodule development and/or in the response to environmental stress (Libault et al. 2009). Because the level of variation in the expression of TF genes in response to environmental stress is low, the use of qRT-PCR is a more sensitive and reliable approach, as compared to microarray analysis, to analyze global TF expression profiles in plants (Czechowski et al. 2004, Hernández et al. 2007). We report 67 TF genes differentially expressed in bean nodules exposed to PQ or to Mn toxicity (Table 3, see Table S2). The transcriptomic responses are known to partially overlap among different kinds of stress (Kreps et al. 2002, Fujita et al. 2006). In agreement we found that the expression pattern of 10 TFs overlapped in nodules subjected to Mn toxicity and PQ, probably because both treatments resulted in ROS production. The TF families that exhibited the strongest responses to oxidative stress were MYB, AP2/EREBP, C2H2(Zn), C2C2(Zn), GRAS, CCAAT, ARF and ZnF-CCHC (Table 3, see Tables S2 and S3). The stress-modulated expression of these TFs suggests that they may be important in the oxidative stress response and that they may constitute a core group of TFs relevant for the response of bean nodules to oxidative stress. Three members of the WRKY TF gene family were repressed in PQ-treated bean nodules (Table 3); TFs from this family respond to oxidative stress in Arabidopsis (Scarpeci et al. 2008). Four members of the GRAS family responded to oxidative stress in bean nodules, including TC42487, which was induced by up to 13-fold (Table 3). Members of the GRAS TF family are involved in development and other processes, such as rhizobial

Nod-factor induction and drought stress response; however, little is known about their physiological role under oxidative stress (Ge et al. 2010). Seven members of the MYB family, which is involved in biotic and abiotic stress responses as well as in developmental processes (Yanhui et al. 2006), were differentially expressed (from three- to sevenfold) in PQ-treated nodules. TFs from the AP2-EREBP family were both induced and repressed in bean nodules subjected to oxidative stress; members of this family are known have important roles in the development and growth of the plant and particularly in response to different stresses and hormonal regulation (Riechmann and Meyerowitz 1998).

A mechanism for the regulation of gene expression is the interaction among TFs and *cis*-regulatory DNA sequences, it is an essential functional connection between gene regulatory networks. The different types of *cis*-elements detected in our in silico analysis have been reported to be involved in responses to abiotic and biotic stresses. Most of the genes that were regulated by oxidative stress, according to our PathExpress analysis, contained at least three of these *cis*-elements (Table 4). The W-box and W-box-like motifs, which correspond to the binding sites of TF-type WRKY, were present in all groups of genes analyzed, and motifs were also found in the *cis*-elements of these genes that facilitate binding to the AP2/ERBF TF family (Table 4). A more detailed analysis of the regulatory elements present in the promoter regions of genes that are responsive to oxidative stress can be undertaken to further our understanding of gene regulatory networks in beans exposed to oxidative stress conditions.

Our current knowledge of bean oxidative stress responsive genes is limited. Here, we performed a comprehensive transcriptome profiling of common bean nodules exposed to oxidative stress. Nearly 15% of the tested genes, which participate in key biological processes, such as secondary metabolism, response to stress, translation, cell death and carbon/nitrogen metabolism, responded to oxidative stress, thus indicating that the accumulation of ROS within the nodule induced transcriptome reprogramming in this organ. It would be of great interest to pinpoint differences between the specific response to oxidative stress in nodules and the response to reduced SNF and hence diminished nitrogen assimilation. The identification of highly oxidative stress-responsive TFs is of crucial importance, especially in the case of legumes, as less that 1% of the total TFs present in the genome have been characterized (Udvardi et al. 2007). Several of the TFs identified here may represent interesting candidates for the genetic engineering of bean plants with enhanced resistance to oxidative stress. The results of our study

may provide a basis for future research that aims to improve the common bean germplasm through genetic or biotechnological approaches.

References

Ahuja I, Kissen R, Bones AM (2012) Phytoalexins in defense against pathogens. Trends Plant Sci 17: 73–90

Alamillo JM, Diaz-Leal JL, Sanchez-Moran MV, Pineda M (2010) Molecular analysis of ureide accumulation under drought stress in *Phaseolus vulgaris* L. Plant Cell Environ 33: 1828–1837

Ampe F, Kiss E, Sabourdy F, Batut J (2003) Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. Genome Biol 4: R15

Apel K, Hirt H (2004) Reactive oxygen species: metabolism oxidative stress signal transduction. Annu Rev Plant Biol 55: 373–399

Asamizu E, Nakamura Y, Sato S, Tabata S (2005) Comparison of the transcript profiles from the root the nodulating root of the model legume *Lotus japonicus* by serial analysis of gene expression. Mol Plant Microbe Interact 18: 487–498

Baxter CJ, Redestig H, Schauer N, Repsilber D, Patil KR, Nielsen J, Selbig J, Liu J, Fernie AR, Sweetlove LJ (2007) The metabolic response of heterotrophic Arabidopsis cells to oxidative stress. Plant Physiol 143: 312–325

Becana M, Dalton DA, Moran JF, Iturbe-Ormaetxe I, Matamoros MA, Rubio MC (2000) Reactive oxygen species antioxidants in legume nodules. Physiol Plant 109: 372–381

Becana M, Matamoros MA, Udvardi M, Dalton DA (2010) Recent insights into antioxidant defenses of legume root nodules. New Phytol 188: 960–976

Boldt R, Zrenner R (2003) Purine pyrimidine biosynthesis in higher plants. Physiol Plant 117: 297–304

Brechenmacher L, Kim MY, Benitez M, Li M, Joshi T, Calla B, Lee MP, Libault M, Vodkin LO, Xu D (2008) Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. Mol Plant Microbe Interact 21: 631–645

Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Verleyden J (2003) Beans (*Phaseolus* spp): model food legume. Plant Soil 252: 55–128

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. BMC Genomics 9: 561

Colebatch G, Kloska S, Trevaskis B, Freund S, Altmann T, Udvardi MK (2002) Novel aspects of symbiotic nitrogen fixation uncovered by transcript profiling with cDNA arrays. Mol Plant Microbe Interact 15: 411–420

Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. Plant J 39: 487–512

Crockford AJ, Davis GA, Williams HD (1995) Evidence for cell-density-dependent regulation of catalase activity in *Rhizobium leguminosarum bv phaseoli*. Microbiology 141: 843–851

Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time RT PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root-shoot-specific genes. Plant J 38: 366–379

Dalton DA (1992) Effects of paraquat on the oxygen free radical biology of soybean root nodules. Bull Environ Contam Toxicol 48: 721–726

Dalton DA (1995) Antioxidant defenses of plants fungi. In: Ahmad S (ed) Oxidative Stress Antioxidant Defenses in Biology. Chapman & Hall, New York, pp 298–355

Dalton DA, Boniface C, Turner Z, Lindahl A, Kim HJ, Jelinek L, Govindarajulu M, Finger RE, Taylor CG (2009) Physiological roles of glutathione *S*-transferases in soybean root nodules. Plant Physiol 150: 521–530

Davletova S, Rizhsky L, Liang H, Zhong S, Oliver DJ, Coutu J, Shulaev V, Schlauch K, Mittler R (2005a) Cytosolic ascorbate peroxidase1 is a central component of the reactive oxygen gene network of *Arabidopsis*. Plant Cell 17: 268–281

Davletova S, Schlauch K, Coutu J, Mittler R (2005b) The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in Arabidopsis. Plant Physiol 139: 847–856

Djordjevic MA, Chen HC, Natera S, Van Noorden G, Menzel C, Taylor S (2003) A global analysis of protein expression profiles in *Sinorhizobium meliloti*: discovery of new genes for nodule occupancy stress adaptation. Mol Plant Microbe Interact 16: 508–524

Dombrecht B, Heusdens C, Beullens S, Verreth C, Mulkers E, Proost P, Verleyden J, Michiels J (2005) Defence of *Rhizobium etli* bacteroids against oxidative stress involves a complexly regulated atypical 2-Cys peroxiredoxin. Mol Microbiol 55: 1207–1221

Donahue JL, Okpodu CM, Cramer CL, Grabau EA, Alscher RG (1997) Responses of antioxidants to paraquat in pea leaves. Plant Physiol 113: 249–257

Donati A, Jeon J, Sangurdeka D, So J, Chang W (2011) Genome-wide transcriptional physiological responses of *Bradyrhizobium japonicum* to paraquat-mediated oxidative stress. Appl Environ Microbiol 77: 3633–3643

Du Z, Bramlage WJ (1992) Modified thiobarbituric acid assay for measuring lipid oxidation in sugar rich plant tissue extracts. J Agric Food Chem 40: 1566–1570

Durango D, Quinones W, Torres F, Rosero Y, Gil J, Echeverri F (2002) Phytoalexin accumulation in Colombian bean varieties and aminosugars as elicitors. Molecules 7: 817–832

Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207–210

El Yahyaoui F, Kuster H, Ben Amor B, Hohnjec N, Puhler A, Becker A, Gouzy J, Vernie T, Gough C, Niebel A (2004) Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation including many potential regulators of the symbiotic program. Plant Physiol 136: 3159–3176

Fischer H, Fritsche S, Herzog B, Hennek H (1989) Critical spacing between two essential cysteine residues in the interdomain linker of the *Bradyrhizobium japonicum* NifA protein. FEBS Lett 255: 167–171

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol 9: 436–442

Gadjev I, Verauwera S, Gechev TS, Laloi C, Minkov IN, Shulaev V, Apel K, Inzé D, Mittler R, Van Breusegem F (2006) Transcriptomic footprints disclose specificity of reactive oxygen species signaling in *Arabidopsis*. Plant Physiol 141: 436–445

Ge Y, Li Y, Zhu YM, Bai X, Lu DK, Guo DJ, Ji W, Cai H (2010) Global transcriptome profiling of wild soybean (*Glycine soja*) roots under NaHCO$_3$ treatment. BMC Plant Biol 10: 153

Gechev TS, Gadjev IZ, Hille J (2004) An extensive microarray analysis of AAL-toxin-induced cell death in *Arabidopsis thaliana* brings new insights into the complexity of programmed cell death in plants. Cell Mol Life Sci 61: 1185–1197

Gechev TS, Minkov IN, Hille J (2005) Hydrogen peroxide-induced cell death in *Arabidopsis*: transcriptional mutant analysis reveals a role of an oxoglutarate-dependent dioxygenase gene in the cell death process. IUBMB Life 57: 181–188

Gigon A, Matos A, Laffray D, Zuily-fodil Y, Pham-Thi A (2004) Effect of drought stress on lipid metabolism in the leaves of *Arabidopsis thaliana* (Ecotype Columbia). Ann Bot 94: 345–351

Goffard N, Weiller G (2007) PathExpress: a Web-based tool to identify relevant pathways in gene expression data. Nucleic Acids Res 35: W176–W181

Goffard N, Frickey T, Weiller G (2009) PathExpress update: the enzyme neighbourhood method of associating gene-expression data with metabolic pathways. Nucleic Acids Res 37: W335–W339

González A, Lynch JP (1997) Effects of manganese toxicity on leaf CO$_2$ assimilation of contrasting common bean genotypes. Physiol Plant 101: 872–880

González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G (2006) The partitioned *Rhizobium etli* genome: genetic metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci USA 103: 3834–3839

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40: 1178–1186

Granados-Baeza MJ, Gómez-Hernández N, Mora Y, Delgado MJ, Romero D, Girard L (2007) Novel reiterated Fnr-type proteins control the production of the symbiotic terminal oxidase *cbb3* in *Rhizobium etli* CFN42. Mol Plant Microbe Interact 20: 1241–1249

Härtel H, Dörmann P, Benning C (2000) DGD1-independent biosynthesis of extraplastidic galactolipids after phosphate deprivation in Arabidopsis. Proc Natl Acad Sci USA 97: 10649–10654

Heath RL, Packer L (1968) Photoperoxidation in isolated chloroplasts. I. Kinetics and stoichiometry of fatty acid peroxidation. Arch Biochem 125: 189–198

Hernández G, Ramírez M, Valdes-López O, Tesfaye M, Graham MA, Czechowski T, Schlereth A, Wrey M, Erban A, Cheung F, Wu H, Lara M, Town CD, Kopka J, Udvardi MK, Vance CP (2007) Phosphorus stress in common bean: Root transcript metabolic responses. Plant Physiol 144: 752–767

Hernández G, Valdes-López O, Ramírez M, Goffard N, Weiller G, Aparicio-Fabre R, Fuentes SI, Erban A, Kopka J, Udvardi MK, Vance CP (2009) Global changes in the transcript metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. Plant Physiol 151: 1221–1238

Inze D, Van Montagu M (1995) Oxidative stress in plants. Curr Opin Biotechnol 6: 153–158

Iturbe-Ormaetxe I, Escuredo PR, Arrese-Igor C, Becana M (1998) Oxidative damage in pea plants exposed to water deficit or paraquat. Plant Physiol 116: 173–181

Jeon J, Lee H, Donatim A, So J, Emerich D, Chang W (2011) Whole-genome expression profiling of *Bradyrhizobium japonicum* in response to hydrogen peroxide. Mol Plant Microbe Interact 24: 1472–1481

Kim JH, Delauney AJ, Verma DPS (1995) Control of *de novo* purine biosynthesis genes in ureide-producing legumes: induction of glutamine

phosphoribosylpyrophosphate amidotransferase gene and characterization of its cDNA from soybean and Vigna. Plant J 7: 77–86

Kouchi H, Shimomura K, Hata S, Hirota A, Wu GJ, Kumagai H, Tajima S, Suganuma N, Suzuki A, Aoki T (2004) Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume *Lotus japonicus*. DNA Res 11: 263–274

Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF (2002) Transcriptome changes in Arabidopsis in response to salt, osmotic and cold stress. Plant Physiol 130: 2129–2141

Lasat MM, DiTomaso JM, Hart JJ, Kochian LV (1997) Evidence for vacuolar sequestration of paraquat in roots of a paraquat-resistant *Hordeum glaucum* biotype. Physiol Plant 99: 255–262

Lee HI, Hur CG, Oh CJ, Kim HB, Park SY, An CS (2004) Analysis of the root nodule–enhanced transcriptome in soybean. Mol Cell 18: 53–62

Libault M, Joshi T, Benedito VA, Xu D, Udvardi MK, Stacey G (2009) Legume transcription factor genes: what makes legumes so special? Plant Physiol 151: 991–1001

Lopez O, Morera C, Miranda-Rios J, Girard L, Romero D, Soberón M (2001) Regulation of gene expression in response to oxygen in *Rhizobium etli*: role of FnrN in *fixNOQP* expression and in symbiotic nitrogen fixation. J Bacteriol 183: 6999–7006

Mandon K, Pauly N, Boscari A, Brouquisse R, Frendo P, Demple B, Puppo A (2009) ROS in the legume-*Rhizobium* symbiosis. In: Del Río LA, Puppo A (eds) Reactive Oxygen Species in Plant Signaling. Signaling and Communication in Plants. Springer-Verlag, Berlin, Heidelberg, pp 135–147

Marino D, González EM, Arrese-Igor C (2006) Drought effects on carbon and nitrogen metabolism of pea nodules can be mimicked by paraquat: evidence for the occurrence of two regulation pathways under oxidative stresses. J Exp Bot 57: 665–673

Martinez-Salazar JM, Salazar E, Encarnacion S, Ramirez-Romero MA, Rivera J (2009) Role of the extracytoplasmic function sigma factor RpoE4 in oxidative osmotic stress responses in *Rhizobium etli*. J Bacteriol 191: 4122–4132

Matamoros MA, Dalton DA, Ramos J, Clemente MR, Rubio MC, Becana M (2003) Biochemistry molecular biology of antioxidants in the rhizobia-legume symbiosis. Plant Physiol 133: 499–509

Minchin FR, James EK, Becana M (2008) Oxygen diffusion production of reactive oxygen nitrogen species antioxidants in legume nodules. In: Dilworth MJ, James EK, Sprent JI, Newton WE (eds) Nitrogen-Fixing Symbioses. Springer, New York, pp 321–362

Mittler R, Verauwera S, Gollery M, Van Breusegem F (2004) Reactive oxygen gene network in plants. Trends Plant Sci 9: 490–498

Naya L, Ladrera R, Ramos J, Gonza'lez EM, Arrese-Igor C, Minchin FR, Becana M (2007) The response of carbon metabolism and antioxidant defenses of alfalfa nodules to drought stress and to the subsequent recovery of plants. Plant Physiol 144: 1104–1114

Neill SJ, Desikan R, Clarke A, Hurst RD, Hancock JT (2002) Hydrogen peroxide and nitric oxide as signaling molecules in plants. J Exp Bot 53: 1237–1247

Ohwada T, Shirakawa Y, Kusumoto M, Masuda H, Sato T (1999) Susceptibility to hydrogen peroxide catalase activity of root nodule bacteria. Biosci Biotechnol Biochem 63: 457–462

Op den Camp RG, Przybyla D, Ochsenbein C, Laloi C, Kim C, Danon A, Wagner D, Hideg E, Göbel C, Feussner I, Nater M, Apel K (2003) Rapid induction of distinct stress responses after the release of singlet oxygen in *Arabidopsis*. Plant Cell 15: 2320–2332

Ramírez M, Valderrama B, Arrendondo-Peter R, Sobéron M, Mora J, Hernández G (1999) *Rhizobium etli* genetically engineered for the heterologous expression of *Vitreoscilla* sp hemoglobin: Effect on free-living symbiosis. Mol Plant Microbe Interact 12: 1008–1015

Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol 137: 1211–1227

Riechmann JL (2002) Transcriptional regulation: a genomic overview. In: Somerville CR, Meyerowitz EM (eds) The Arabidopsis Book. American Society of Plant Biologists, Rockville, MD, http://www.aspborg/publications/Arabidopsis

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biol Chem 379: 633–646

Rouillard JM, Zuker M, Gulari E (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res 31: 3057–3062

Routledge R (1998) Fisher's exact test. In: Armitage P, Colton T (eds) Encyclopedia of Biostatistics, Vol. 2. John Wiley & Sons, New York, NY, pp 1519–1523

Santos R, Hérouart D, Puppo A, Touati D (2000) Critical protective role of bacterial superoxide dismutase in *Rhizobium*–legume symbiosis. Mol Microbiol 38: 750–759

Scarpeci TE, Valle EM (2008) Rearrangement of carbon metabolism in *Arabidopsis thaliana* subjected to oxidative stress condition: an emergency survival strategy. Plant Growth Regul 54: 133–142

Scarpeci TE, Zanor MI, Carrillo N, Mueller-Roeber B, Valle EM (2008) Generation of superoxide anion in chloroplasts of *Arabidopsis thaliana* during active

photosynthesis: a focus on rapidly induced genes. Plant Mol Biol 66: 361–378

Schlüter A, Patschkowski T, Quandt J, Selinger LB, Weidner S, Krämer M, Zhou L, Hynes MF, Priefer UB (1997) Functional and regulatory analysis of the two copies of the *fixNOQP* operon of *Rhizobium leguminosarum* strain VF39. Mol Plant Microbe Interact 10: 605–616

Schubert KR (1981) Enzymes of purine biosynthesis catabolism in *Glycine max*: I. Comparison of activities with N(2) fixation composition of xylem exudate during nodule development. Plant Physiol 68: 1115–1122

Serraj R, Vasquez-Diaz H, Drevon J (1998) Effects of salt stress on nitrogen fixation, oxygen diffusion, and ion distribution in soybean, common bean, and alfalfa. J Plant Nutr 21: 475–488

Shulaev V, Oliver DJ (2006) Metabolic and proteomic markers for oxidative stress. New tools for reactive oxygen species research. Plant Physiol 141: 367–372

Sies H (1991) Oxidative stress: introduction. In: Sies H (ed) Oxidative Stress: Oxidants Antioxidants. Academic Press, London, UK, pp xv–xxii

Smyth GK (2004) Linear models empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article 3

Souza EM, Pedrosa FO, Drummmond M, Rigo LU, Yates MG (1999) Control of *Herbaspirillum seropedicae* nifA activity by ammonium ions and oxygen. J Bacteriol 181: 681–684

Starker CG, Parra-Colmenares AL, Smith L, Mitra RM, Long SR (2006) Nitrogen fixation mutants of *Medicago truncatula* fail to support plant bacterial symbiotic gene expression. Plant Physiol 140: 671–680

Summerfield RJ, Huxley PA, Minchin FR (1977) Plant husbandry management techniques for growing grain legumes under simulated tropical conditions in controlled environments. Exp Agric 13: 113–121

The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25: 25–29

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) MapMan: a user-driven tool to display genomic data sets onto diagrams of metabolic pathways other biological processes. Plant J 37: 914–939

Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, Zhang JY, Benedito V, Hofer JM, Chueng F, Town CD, (2007) Legume transcription factors: global regulators of plant development and response to the environment. Plant Physiol 144: 538–549

Umbach AL, Fiorani F, Siedow JN (2005) Characterization of transformed *Arabidopsis* with altered alternative oxidase levels and analysis of effects on reactive oxygen species in tissue. Plant Physiol 139: 1806–1820

Upchurch RG (2008) Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. Biotechnol Lett 30: 967–977

Valdés-López O, Yang SS, Aparicio-Fabre R, Graham PH, Reyes JL, Vance CP, Hernández G (2010) MicroRNA expression profile in common bean (*Phaseolus vulgaris*) under nutrient deficiency stresses and manganese toxicity. New Phytol 187: 805–818

Vance CP, Heichel GH (1991) Carbon in N_2 fixation: limitation or exquisite adaptation. Annu Rev Plant Physiol Plant Mol Biol 42: 373–392

Vandarauwera S, Zimmermann P, Rombauts S, Venabeele S, Langebartels C, Gruissem W, Inzé D, Van Breusegem F (2005) Genome-wide analysis of hydrogen peroxide-regulated gene expression in *Arabidopsis* reveals a high light-induced transcriptional cluster involved in anthocyanin biosynthesis. Plant Physiol 139: 806–821

Vargas MC, Encarnación S, Dávalos A, Reyes-Pérez A, Mora Y, García-de los Santos A, Brom S, Mora J (2003) Only one catalase *katG* is detectable in *Rhizobium etli*, and is encoded along with the regulator OxyR on a plasmid replicon. Microbiology 149: 1165–1176

Verdoy D, Lucas M, Manrique E, Covarrubias A, De Felipe M, Pueyo J (2004) Differential organ specific response to salt stress and water deficit in nodulated bean (*Phaseolus vulgaris*). Plant Cell Environ 27: 757–767

Verniquet F, Gaillard J, Neuburger M, Douce R (1991) Rapid inactivation of plant aconitase by hydrogen peroxide. Biochem J 276: 643–648

Vogels GD, Van Der Drift C (1970) Differential analyses of glyoxylate derivatives. Anal Biochem 33: 143–157

Vranova E, Inzé D, Van Breusegem F (2002) Signal transduction during oxidative stress. J Exp Bot 53: 1227–1236

Yang ZB, You JF, Xu MY, Yang ZM (2009) Interaction between aluminum toxicity and manganese toxicity in soybean. Plant Soil 319: 277–289

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol 60: 107–124

Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S (2000) Expression of spinach ascorbate peroxidase isoenzymes in response to oxidative stresses. Plant Physiol 123: 223–234

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Gene ontology (GO) biological processes of bean nodule ESTs differentially expressed under oxidative stress.

Table S1. Gene expression significantly affected by PQ in common bean nodules identified by microarray analysis.

Table S2. TF genes with significant changes in expression in common bean nodules under PQ treatment, as identified by qRT-PCR analysis.

Table S3. TF genes significantly expressed in common bean nodules exposed to Mn toxicity, as identified by qRT-PCR analysis.

Table S4. Metabolic pathways and enzymes significantly over-represented in common bean nodules under PQ treatment.

Table S5. Primers used for qRT-PCR expression analysis to validate microarray gene expression data.

Table S6. Primers used for qRT-PCR expression analysis of bacteroids subjected to PQ treatment.

94

B. Anexo

CrossMark

ORIGINAL ARTICLE

Rhizobium etli bacteroids engineered for *Vitreoscilla* hemoglobin expression alleviate oxidative stress in common bean nodules that reprogramme global gene expression

Mario Ramírez[1] · Luis P. Íñiguez[1] · Gabriela Guerrero[1] · Francesca Sparvoli[2] · Georgina Hernández[1]

Abstract The common bean (*Phaseolus vulgaris* L.)–*Rhizobium etli* symbiosis and crop productivity are highly affected by adverse environmental conditions that cause oxidative stress. Based on the improved symbiosis of common bean inoculated with engineered *R. etli* expressing the *Vitreoscilla* hemoglobin (VHb) (Ramírez et al., Mol Plant Microbe Interact 12:1008–1015, 1999), in this work we analyzed the effect of this strain in plants exposed to the herbicide paraquat (PQ) which generates oxidative stress. PQ-treated plants inoculated with the engineered (VHb) *R. etli* strain showed higher nitrogenase activity and ureide content than plants inoculated with the wild type strain. We performed microarray transcriptomic analysis to identify PQ-responsive genes in nodules elicited by engineered vs wild type strains. An evident reprogramming of the transcriptional profile was observed in PQ-treated nodules, and the global changes in gene expression were different between nodules elicited with each strain. The most relevant difference was the increased number of up-regulated PQ-responsive genes in wild type strain nodules as compared to VHb-expressing nodules. The majority of these genes were classified into biological processes/functional categories related to defense, response to abiotic stress or signaling, as revealed by Gene Ontology and MapMan analysis. Taken together our analysis suggests that the expression of VHb in *R. etli* bacteroids contributes to buffering the damage caused by increased reactive oxygen species, and this is reflected in nodule cells that showed decreased sensitivity to oxidative stress and response of stress-related genes. Biotechnological applications of VHb-expressing rhizobia inoculants could be further explored.

Keywords Symbiotic nitrogen fixation · *Phaseolus vulgaris–Rhizobium etli* symbiosis · Oxidative stress · *Vitreoscilla* hemoglobin · Rhizobia engineered strain · Nodule transcriptomics

Introduction

Legumes are the second most important crop for humans, accounting for 27% of crop production worldwide (Graham and Vance 2003). A hallmark trait of legumes is their ability to establish symbiotic relations with N_2-fixing soil bacteria collectively known as rhizobia. Infection of legume roots by compatible rhizobia results in the formation of a new organ, the root nodule, where bacteroids carry out the symbiotic nitrogen fixation (SNF) in a microaerobic environment. Leghemoglobin is synthesized in high amounts by nodule cells and plays a key role in delivering oxygen to the bacteroid terminal oxidase system, which operates optimally at very low oxygen pressure and generates ATP required for nitrogenase (Wittenberg 2012). SNF occurring in the legume-rhizobia symbiosis is the main source of nitrogen in agro-systems. The advantages of legume crop production using SNF, or bio-fertilization, is the reduced costs for producers and the prevention of soil and water pollution by excessive use of chemical

✉ Mario Ramírez
mario@ccg.unam.mx

1 Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad 1001, 62209 Cuernavaca, Morelos, Mexico

2 Istituto di Biologia e Biotecnologia Agraria, CNR, Via Bassini 15, 20133 Milan, Italy

🐋 Springer

fertilizers. Thus, research aiming to improve rhizobia-based bio-fertilizers has been going on for several years, for example, through the discovery of rhizobia strains with increased SNF capacity to be used as better inoculant. For this, approaches based in rhizobial genetic modification/engineering are in progress.

Common bean (*Phaseolus vulgaris* L.) is the main food protein crop for direct human consumption worldwide. It is a staple food in developing countries, such as Mexico, where beans are the main dietary supply of plant proteins (Broughton et al. 2003; Graham and Vance 2003). Its N_2-fixing capacity is one of the main advantages of this crop because it reduces the nitrogen fertilizer demand, thus having a positive impact on the economy of developing countries. Research from the National University of Mexico has led to modified *Rhizobium etli* strains, the prominent rhizobial species for common bean nodulation (Segovia et al. 1993), with better SNF capacity. Examples of modifications in improved *R. etli* strains are: the expression of the *Bradyrhizobium japonicum fixNOQP* operon (Soberón et al. 1999) or the mutation in *StoR*d, a negative regulator of ccb_3 (Granados-Baeza et al. 2007), both resulting in the overproduction of the bacteroid ccb_3 symbiotic high oxygen affinity terminal oxidase; the expression of the heterologous *Vitreoscilla* hemoglobin VHb as described below (Ramírez et al. 1999); and increased nitrogenase production obtained through the expression of a chimeric *nifHDK* operon regulated by the strong *nifHc* promoter. The latter engineered strain was produced by the Mexican bio-fertilizer industry and successfully used as an inoculant for common bean (Peralta et al. 2004).

Common bean SNF and crop productivity is highly affected by disease and insect pressure and also by edaphic constraints. SNF in the legume-rhizobia symbiosis is sensitive to oxidative stress resulting from the excessive production of reactive oxygen species (ROS)—superoxide (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radicals (OH^*)—that are generated by abiotic stresses, such as high and low temperatures, particularly in combination with high light intensities, drought, exposure to air pollutants, ultraviolet light, toxic metals in the soil, and exposure to herbicides such as paraquat (PQ) (Vadez et al. 2012). Using 'omics and other approaches, our group has analyzed the response of common bean plants to some of these abiotic stresses, including aluminum and copper toxicity as well as PQ treatment (Ramírez et al. 2013; Naya et al. 2014; Mendoza-Soto et al. 2015). Ramírez et al. (2013) showed that the effect of PQ (1,1′-dimethyl-4,4′-bipyridinium dichloride hydrate) treatment in SNF common bean plants was similar to that resulting from oxidative stress

generated by environmental constraints, such as low temperature, manganese toxicity, salinity and drought. *Rhizobium tropici*-nodulated common bean plants exposed to PQ for short periods were drastically affected, showing reduced nitrogenase activity and ureide content that correlated with nodule transcriptome reprogramming. The main transcriptomic response of PQ-exposed nodules included the repression of genes for carbon/nitrogen metabolism that are crucial for nodule function (Ramírez et al. 2013).

A major goal for agriculture is to improve stress tolerance and crop yield in plants. Research in this area with legumes includes, among others, obtaining stress-tolerant rhizobia strains that when used as inoculants would reduce the sensitivity of SNF-legumes to abiotic stress. Interestingly, there are reports showing that the inoculation of common bean plants with modified *R. etli* strains alleviate the negative effects of abiotic stresses to SNF and plant yield. For example, both a *R. etli* strain that over-expresses the trehalose-6-phosphate synthase gene (Suárez et al. 2008) as well as a strain that over-expresses the bacteroid oxidase cbb_3 (Talbi et al. 2012) decrease drought sensitivity of SNF common bean plants inoculated with either of these engineered bacteria. To this end, in this work we investigated the effect of an engineered *R. etli* strain expressing VHb on the performance of SNF common bean plants under oxidative stress generated by PQ addition.

VHb is a hemoprotein produced by *Vitreoscilla* sp., a gram-negative strictly aerobic bacteria found in hypoxic environments (Joshi et al. 1998; Wakabayashi et al. 1986). VHb has higher chemical reactivity than its hemoglobin homologues from multicellular organisms and is associated with diverse functions, such as catalyzing redox reactions and protecting cells against oxidative stress, besides its function of oxygen storage and transport (Geckil et al. 2003; Hardison 1998; Kaur et al. 2002). There are numerous reports indicating that different microorganisms expressing VHb show improved growth or metabolite production. *Escherichia coli* expressing VHb showed higher oxygen uptake rates as well as better growth and higher cell density as compared to control strains (Geckil et al. 2001; Khosla and Bailey 1988). Interestingly, the expression of VHb both in *E. coli* and in *Enterobacter aerogenes* confers significant protection against oxidative stress that is correlated with a higher expression of antioxidant enzymes (Anand et al. 2010; Geckil et al. 2003).

We have reported the effect VHb expression, driven by its native promoter, in *R. etli*, both under free-living condition and in symbiosis with common bean (Ramírez et al. 1999). The presence of VHb transcript and protein in the engineered *R. etli* strain was demonstrated. Free-living *R. etli* expressing VHb grown under limiting oxygen concentrations showed increased respiratory activity, chemical

energy content, and expression of the nitrogen fixation gene *nifHc*. Phenotypic analysis of common bean plants inoculated with the engineered *R. etli* (VHb) strain as compared plants inoculated with the wild type strain showed increased foliage dry weight, earlier pod formation and flowering, higher nitrogenase activity and total nitrogen content. Our results led us to propose that VHb production in *R. etli* could improve oxygen delivery to the symbiotic terminal oxidase, thus increasing the respiratory efficiency and nitrogen fixation in bacteroids that result in increased nitrogen assimilation by the plant and overall growth improvement.

In this work, we explore if the presence of VHb in engineered *R. etli* in symbiosis with common bean plants would decrease sensitivity to oxidative stress generated by the addition of PQ. Our data show that PQ treatment plants inoculated with the engineered *R. etli* strain (CE3/VHb) showed higher nitrogenase activity and ureide content than plants inoculated with the wild type strain (CE3). Comparative transcriptomic analyses were performed to assess nodule global gene expression in plants inoculated with CE3/VHb or CE3 strains grown in control or PQ treatments. Beside the transcriptome reprogramming in oxidative stressed nodules (Ramírez et al. 2013), a clear difference in transcript profile was observed in nodules with engineered bacteroids. Nodule differentially expressed genes (DEG) in PQ treatment were analyzed using the Gene Ontology and MapMan bioinformatics tools (Ashburner et al. 2000; Thimm et al. 2004) adapted to common bean. The DEG analysis led to interpretations about their role in the better performance of CE3/VHb-inoculated common bean plants subjected to oxidative stress. Our work sets the basis for exploring bio-fertilizers made up of VHb-engineered rhizobial strains with the aim of contributing to improvement of common bean yield, especially when grown in deleterious environmental conditions resulting in oxidative stress.

Materials and Methods

Plant material and growth conditions

The common bean (*Phaseolus vulgaris*) Mesoamerican cv. Negro Jamapa 81 was used in this study. The growth conditions and tissues collected were similar to those reported by Ramírez et al. (2013). Briefly, surface-sterilized seeds were germinated at 30 °C on sterile, wet filter paper in darkness for 3 days. Germinated seedlings were planted in pots with vermiculite and inoculated with 1 ml of saturated *Rhizobium etli* liquid culture. The *R. etli* strains used in this study were the wild type CE3 strain and CE3

bearing the stable plasmid pMR4 (Ramírez et al. 1999) containing the *Vitreoscilla* sp. *vgb* gene (the gene encoding VHb), hereby denominated strain CE3/VHb. The expression of VHb, driven by its natural promoter, as well as the presence of VHb protein in the CE3/Vhb strain was demonstrated previously (Ramírez et al. 1999). Pots were watered 3 days per week with N-free Summerfield plant nutrient solution (Summerfield et al. 1977). Plants were grown under controlled environmental conditions (26–28 °C, 70% humidity, 16-h photoperiod) for 18 days. To induce oxidative stress, at 16 days post-inoculation (dpi) plants with active nodules were watered with nutrient solution supplemented with 1 mM PQ (1,1′-dimethyl-4,4′-bipyridinium dichloride) for 48 h. At 18 dpi, plants from control condition (C) or PQ treatment were harvested for analysis; mature nodules were detached, frozen in liquid nitrogen and stored at −80 °C until used.

Phenotypic analysis

Nitrogenase activity was determined by the acetylene reduction assay (Hardy et al. 1968) in detached nodulated roots from ten plants from each treatment (C or PQ) from three independent experiments. Ethylene production was determined by gas chromatography in a Variant model 3300 chromatograph as reported (Ramírez et al. 1999). Specific activity is expressed as nmol ethylene h^{-1} g^{-1} nodule DW.

The ureide content was analyzed in xylem exudates from 18-dpi PQ-treated and control bean plants as described previously (Ramírez et al. 2013). In brief, xylem exudates were collected from the cut stems of 10–15 plants from three independent experiments. After recording the volume of the exudates, the collected sap was stored at −20 °C until used for analysis. The ureide content was determined using the differential analysis reported by Vogels and Van Der Drift (1970).

Transcriptomic analysis by microarray hybridization

The global gene responses of nodules elicited by CE3 or CE3/Vhb strains from plants grown under C and PQ treatments were analyzed using the Bean Custom Array 90 K described previously (Aparicio-Fabre et al. 2013; Ramírez et al. 2013). A total of 30,150 different features were available on the microarray, including 18,867 unique common bean sequence probes (based in the EST sequences available at The Gene Index Project. http://compbio.dfci.harvard.edu/tgi/plant.html, v3.0) and 11,205 soybean UniGene probes (based in NCBI UniGene build 38.0) along with positive and negative controls. Each probe was printed in triplicate to ensure the presence of internal

replicates and to have a good statistical representation of each transcript on the array.

Total RNA from 18-dpi PQ-treated and C nodules from common bean plants inoculated with the different *R. etli* strains was isolated using Trizol reagent (Life Technologies, Carlsbad, CA) as reported (Hernández et al. 2007). Total RNA (1 μg) was used as template to synthesize antisense RNA (aRNA) with Cy5-ULS, using the RNA Amplification and Labelling Kit from CombiMatrix (ampULSe, Kreatech Biotechnology, Amsterdam, The Netherlands), according to the manufacturer's instructions. Prehybridization and hybridization conditions were described previously (Ramírez et al. 2013). After hybridization and washing, the microarray was dipped in imaging solution, covered with LifterSlip^{TM}, and then scanned using a GenePix 4000B microarray scanner (Axon) and the accompanying acquisition software (CombiMatrix Microarray Imager Software). Multiple time scans at different photomultiplier (PMTs) were provided for each hybridization.

Microarray data analysis

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE 83324 (http://www.ncbi.nlm.nih.gov/geo/query/acc. cgi?acc=GSE83324).

Statistically significant signal intensities from the microarray probes were obtained as reported previously (Aparicio-Fabre et al. 2013, Ramírez et al. 2013). Briefly, raw intensity data were first processed using the CombiMatrix Microarray Imager software and then exported into the Feature and Probe format of CombiMatrix, where the actual raw intensity per probe and per spot is stored. Data were loaded into R and analyzed using the Limma package (Smyth 2004). Within-array probe replicates were defined as technical replicates, by calculating the mean intensity across the different probes. The values in the expression matrix were transformed into log$_2$ and a design matrix was defined to describe the biological samples. The expression matrix was used to fit a linear model using the design matrix and the functions of Limma. A set of contrast matrices was defined to describe the comparisons among samples and a second linear fitting was performed for each contrast. The data were then error corrected using the Bayesian functions of Limma and a list of DEG was generated for each contrast, after correcting for multiple testing using the Benjamin–Hochberg method and setting 0.05 as the adjusted P value cutoff.

We used Gene Ontology (GO) (Ashburner et al. 2000) and MapMan (Thimm et al. 2004) bioinformatics-based approaches for analyses aimed at interpreting the biological significance of gene expression data. The ESTs corresponding to the microarray probe sets were organized in functional categories according to GO guidelines (Ashburner et al. 2000). The Fisher's exact test (Routledge 1998) was applied to determine which GO categories were statistically overrepresented within each set of differentially expressed ESTs analyzed ($P \leq 0.01$, corrected by Bonferroni adjustment). A second approach for expression data analysis was based on the MapMan software version 3.5.1 (Thimm et al. 2004; http://gabi.rzpd.de/projects/MapMan/). The *P. vulgaris* map previously developed (Aparicio-Fabre et al. 2013; Ramírez et al. 2013) was used to extend MapMan to common bean. The change in expression ratio of each gene was calculated as the log$_2$-fold change to generate the MapMan experimental file. The MapMan software was used to visualize the amplitudes of the changes in expression of individual genes in diagrams of metabolic pathways or cellular processes.

qRT-PCR analysis

A quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) approach was used to validate the microarray data of selected genes. Total RNA was isolated from 200 mg of frozen nodules as described. Nodules were detached from different sets of plants growing in conditions similar to those described for C and PQ treatments. Genomic DNA removal, cDNA synthesis, and quality verification for qRT-PCR were performed as reported (Hernández et al. 2007; Ramírez et al. 2013). Three biological replicates were carried out for each condition. Reactions were done in a 96-well format with the 7300 Real-Time PCR System and 7300 System Software (Applied Biosystems, Foster City, CA). In each qRT-PCR run, the ascorbate peroxidase gene (APX, Phvul.009G126500) was included as a marker gene (Ramírez et al. 2013); in all the experiments this gene was induced (ca. 3-fold) in PQ-treated nodules, thus confirming the oxidative nature of the stress. Ubiquitin Conjugating Enzyme 9 (UBC9, Phvul.006G110100), which was constant in the tested conditions, was included for normalization in every qRT-PCR run. The sequences of specific oligonucleotide primers used for qRT-PCR amplification for each gene have been reported (Hernández et al. 2007; Ramírez et al. 2013). Relative expression for each sample was calculated with the comparative Ct method as reported (Hernández et al. 2007) and the fold change value (log$_2$) was calculated. Student's t test was performed with a P value cutoff of 0.05.

Results

Symbiosis and oxidative stress response phenotype

We characterized the oxidative stress response of SNF common bean plants inoculated with the VHb-expressing

R. etli strain CE3/VHb as compared to plants inoculated with the CE3 wild type strain. At 16 dpi plants with active nodules were watered with nutrient solution (control condition) or with solution supplemented with PQ for 2 days. In agreement with Ramírez et al. (1999), when grown in the control condition common bean plants inoculated with CE3/VHb showed increased length (20–25%) and foliage dry weight (ca. 40%), as well as higher nitrogenase activity (28%) as compared to CE3-inoculated plants (Table 1). In common bean nodule cells, ammonia from N_2-fixation provided by the bacteroids is assimilated to generate ureides that are the main N-compounds transported from the nodules to aerial parts of the plant (Boldt and Zrenner 2003). As shown in Table 1, CE3/VHb-inoculated plants showed 33% higher ureide content in xylem sap than CE3 inoculated plants. Similar to the results obtained with *R. tropici*-inoculated common bean plants (Ramírez et al. 2013), the PQ treatment had a negative effect on *R. etli* CE3-inoculated plants, which showed decreased nitrogenase activity (66%) and ureide content (71%) (Table 1). However, CE3/VHb-inoculated plants showed less sensitivity to PQ as their nitrogenase activity and ureide content decreased only 46 and 48%, respectively, as compared to plants inoculated with the same strain under control conditions (Table 1). The comparison between control CE3-inoculated plants and CE3/VHb-inoculated plants under PQ treatment revealed a lower decrease in nitrogenase activity (32%) and uriede content (23%) (Table 1). Our results (Table 1) indicate that the presence of VHb in the nodule bacteroids has a protective role against the oxidative stress generated by PQ that results in a more efficient SNF and plant growth under PQ treatment.

Transcriptomics: experimental design and validation

We performed transcriptomic analysis to gain insight into global gene expression in nodules elicited by CE3 or CE3/VHb strains grown under control and PQ-stressed conditions, as described above. Our approach was based on hybridization using the Bean Custom Array 90 K (Aparicio-Fabre et al. 2013; Ramírez et al. 2013). Since the *P. vulgaris* genome has been sequenced, we mapped the 30,150 common bean EST printed on the microarray to the genome (Schmutz et al. 2014; http://www.phytozome.net/commonbean v1.0) and could assign a gene ID and annotation to 25,442 of the printed ESTs.
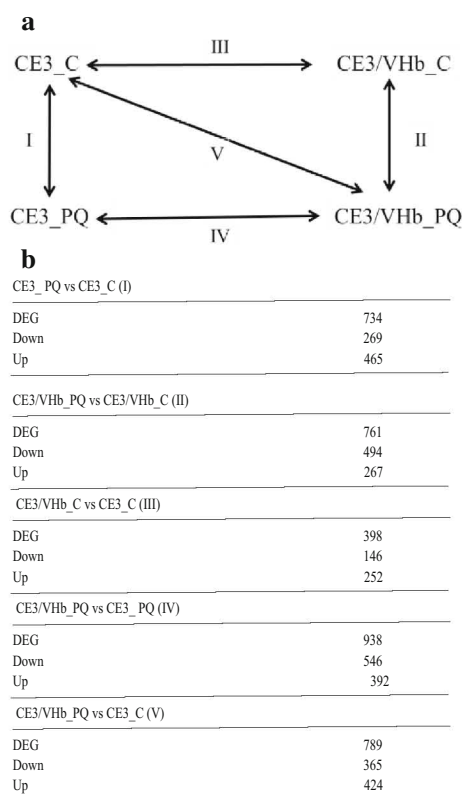
The experimental design, based on circular hybridizations, included four conditions as shown in Fig. 1a, with two independent biological replicates and three technical replicates for each conditions that allowed five possible comparisons (I–V, Fig. 1a, b). A total of 2418 DEG were identified among the different combinations shown in Fig. 1a. The lists of DEG (log_2-fold change $\geq 1.0, P \leq 0.01$) from each comparison (I–V) are presented in Supplemental Table s1 and their numbers in Fig. 1b. Though the number of DEG from comparison I (CE3_PQ vs CE3_C) is very similar to that of comparison II (CE3/VHb_PQ vs CE3/VHb_C), the majority of genes from comparison I are up-regulated while in comparison II most were down-regulated (Fig. 1b). On the other hand, comparison III that includes nodules elicited by different strains from control plants (CE3/VHb_C vs CE3_C) showed the lowest DEG number, while comparison IV that includes nodules from stressed plants (CE3/VHb_PQ vs CE3_PQ) showed the highest DEG number (Fig. 1b).

The results from microarray analysis were validated by qRT-PCR expression analysis. In our previous work (Ramírez et al. 2013), we identified 67 transcription factor (TF) genes that responded to PQ treatment in common bean nodules belonging to TF families known to participate in plant stress responses. On this basis, in this work we selected 13 TF genes that were differentially expressed in CE3/VHb_PQ vs CE3_PQ nodules according to the microarray data for qRT-PCR expression analysis. We found a good correlation ($r^2 = 0.89$) and a similar pattern (up- or down-regulation) between the microarray and the qRT-PCR fold change values, indicating the reliability of the microarray data (Table 2). The variation between the microarray and qRT-PCR expression data is probably due to different sensitivities of the two methods. Table 2 shows that WRKY and bZIP were the most abundant TF families showing response to PQ: members from these families are known to respond to oxidative and drought/salinity stress in Arabidopsis (Malhotra and Sowdhamini 2014). In common bean, an enrichment of putative cis-elements for WRKY and bZIP TF was identified in the promoter regions of genes differentially expressed in PQ-treated nodules (Ramírez et al. 2013). Our data indicate that beside the previously observed response of

Table 1 The effect of VHb expression in *Rhizobium etli* during symbiosis with bean plants under paraquat treatment

	CE3		CE3/VHb	
	C	PQ	C	PQ
Nitrogenase activity (nmol $C_2H_2/min^{-1}/g^{-1}$ DW)	628 ($\pm$48)	277 ($\pm$43)	800 ($\pm$95)	426 ($\pm$32)
Ureide content in xylem sap (μmol/ml/h)	235 ($\pm$44)	68 ($\pm$13)	355 ($\pm$43)	182 ($\pm$38)

C control, *PQ* paraquat

a

b

CE3_PQ vs CE3_C (I)

DEG	734
Down	269
Up	465

CE3/VHb_PQ vs CE3/VHb_C (II)

DEG	761
Down	494
Up	267

CE3/VHb_C vs CE3_C (III)

DEG	398
Down	146
Up	252

CE3/VHb_PQ vs CE3_PQ (IV)

DEG	938
Down	546
Up	392

CE3/VHb_PQ vs CE3_C (V)

DEG	789
Down	365
Up	424

Fig. 1 Experimental design based on circular hybridization of Bean Custom Array 90 K used for transcriptomic analysis to identify differentially expressed genes (DEG) from nodules elicited by *R. etli* strains CE3 or CE3/VHb on plants under PQ treatment or control condition. **a** Scheme representing the circular hybridization and possible comparisons among analyzed conditions. For each condition, two biological replicates and three technical replicates of each biological sample were analyzed. **b** Number of DEG (log$_2$-fold change ≥ 1.0, $P \leq 0.01$) identified among different comparisons numbered as shown in **a**

TF genes to PQ treatment in nodules elicited by a wild type strain (Ramírez et al. 2013), an additional response of some of these genes to VHb expression occurs in nodules of PQ-treated plants (Table 2).

Gene Ontology and MapMan analysis of DEG from PQ-stressed nodules elicited by VHb-expressing or wild type *R. etli*

Gene Ontology (GO) analysis was performed to interpret the biological significance of DEG from nodules under PQ

treatment. Our main interest was to compare the PQ-responsive genes in nodules elicited by VHb-expressing strain with those from nodules elicited by the wild type strain lacking this protein, therefore we focused on comparisons I and II (Fig. 1a). From the total DEG from comparisons I and II (Figs. 1b, Supplemental Table s1), 1397 could be organized in biological processes according to GO guidelines (Ashburner et al. 2000): these are shown in Supplemental Table s2. Fisher's exact test (Routledge 1998) was applied to determine the statistically overrepresented GO categories, down- or up-regulated in the comparisons. Besides GO enrichment analysis from the DEG exclusive to comparisons I and II (Fig. 2a, b), we also analyzed DEG that are commonly down- or up-regulated between these two comparisons, which we refer to as second level comparison (Fig. 2c).

The GO enrichment analysis of DEG exclusive to CE3-elicited nodules in PQ treatment, comparison I, revealed 11 down-regulated and 23 up-regulated biological processes (Fig. 2a). The majority of up-regulated DEG were classified in processes related to defense (GO:0080167 response to karrikin, GO:0009408 response to heat, GO:0009611 response to wounding, GO:0009819 drought recovery, GO:0009651 response to salt stress) and to central metabolism (GO:0010325 raffinose family oligosaccharide biosynthetic process, GO:0001676 long-chain fatty acid metabolic process, GO:0009807 lignan biosynthetic process, GO:0046274 lignin catabolic process) (Fig. 2a). Likewise, processes related to secondary metabolism (GO:0051555 flavonol biosynthetic process, GO:0009699 phenylpropanoid biosynthetic process, GO:0046865 terpenoid transport), known to participate in cell protection against biotic and abiotic stresses, were overrepresented in the up-regulated DEG of these nodules (Fig. 2a).

Similar analysis for DEG exclusive to VHb-expressing nodules in PQ treatment, comparison II, showed less overrepresented GO biological processes; 8 were down-regulated and 10 were up-regulated (Fig. 2b). In contrast to results from comparison I (Fig. 2a), fewer enriched up-regulated processes were related to defense against abiotic stress (GO:0009615 response to virus, GO:0043967 histone H4 acetylation, GO:0010043 response to zinc ion, GO:0046686 response to cadmium ion). Other processes up-regulated in CE3/VHb PQ-treated nodules were related to nitrogen fixation and carbon metabolism (GO: 0009399 nitrogen fixation, GO:0044262 cellular carbohydrate metabolic process) that are basic metabolic pathways for adequate nodule function (Fig. 2b).

The second level comparison between I and II allowed identifying DEG that commonly respond to PQ treatment in nodules elicited by different strains (Fig. 2c). All the commonly enriched up-regulated GO biological processes (11) indicate their general roles in nodule responses to

Table 2 Validation of the microarray data of selected transcription factor genes identified as induced or repressed in nodules elicited by CE3_C vs CE3/VHb under PQ treatment (comparison IV, Fig. 1b)

Gene ID (EST)	Annotation	Fold change	
		Microarray	qRT-PCR
Phvul.005G097800 (BQ253715)	bZIP transcription factor family protein	−0.44737305	−2.84215718
Phvul.004G053600 (CV540001)	Myb domain protein 52	−0.43988486	−0.63550938
Phvul.006G101700 (TC18231)	bZIP transcription factor family protein	−1.3695091	−2.00631975
Phvul.001G119300 (CV523295)	bZIP transcription factor family protein	0.70484463	1.95048633
Phvul.005G153900 (GD956937)	bZIP transcription factor family protein	1.42611818	2.107039974
Phvul.006G021900 (BM094171)	bZIP transcription factor family protein	0.99095394	2.108264949
Phvul.006G078500 (TC11890)	bZIP transcription factor family protein	0.83543646	0.753812
Phvul.009G094200 (TC11318)	bZIP transcription factor family protein	1.1649796	1.48192373
Phvul.007G241600 (TC8812)	AP2/ERF transcription factor family protein	0.56769841	1.6488426
Phvul.008G286100 (TC20065)	WRKY family transcription factor	1.05326109	1.69716774
Phvul.002G297100 (TC11713)	WRKY family transcription factor	0.772599831	1.785165967
Phvul.010G111900 (TC12249)	WRKY family transcription factor	0.501022063	2.213934243

Fold change ($\log_2$) values are the average of three biological replicates. Normalized values from qRT-PCR and microarray analyses of each gene are shown

stress; these were related to defense (GO:0071369 cellular response to ethylene stimulus, GO:0071493 cellular response to UV-B, GO:0046686 response to cadmium ion) and to oxidation–reduction processes (GO:0055114 oxidation–reduction process) that participate in oxidative and nitric stresses (GO:0006979 response to oxidative stress, GO: 0071732 cellular response to nitric oxide) (Fig. 2c). Interestingly, the genes included in these processes showed higher induction in PQ-treated nodules elicited by strain CE3 as compared to nodules expressing VHb (Supplemental Table s2).

MapMan and GO analyses are counterpart for comparative gene expression analyses (Klie and Nikoloski, 2012). Therefore, to further analyze our gene expression data we used MapMan software (Thimm et al. 2004; http://gabi.rzpd.de/projects/MapMan/ v.3.5.1) to validate GO enrichment data and to classify DEG in more detail. As for GO analysis, our MapMan analysis focused on comparisons I and II (Fig. 1A). Using the *P. vulgaris* map previously developed and up-loaded to MapMan (Aparicio-Fabre et al. 2013; Ramírez et al. 2013), 1140 from the total DEG from comparison I and II (Figs. 1b, Supplemental Table s3), could be assigned to a MapMan functional category and were considered for analysis. These DEG were classified into 32 tree-structured Bins from MapMan, the DEG classified as unknown (Bin 35) was not further analyzed.
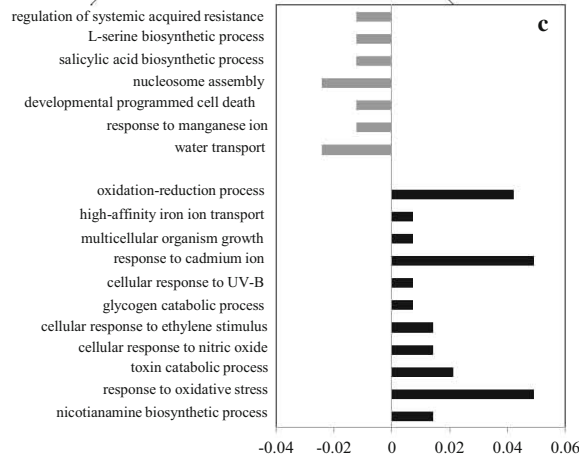
The majority (77%) of analyzed DEG for comparison I were up-regulated while 48% were up-regulated for comparison II. Figure 3 shows the five most abundant functional categories (Bins) from both comparisons: protein,

transport, RNA, miscellaneous enzyme families and stress. A higher up-regulation in CE3 nodules under PQ treatment was evident (Fig. 3). Additionally, Fig. 4 depicts representative maps related to stress response and signaling, obtained with MapMan visualization tool, showing different response to PQ in nodules with CE3 bacteroids as compared to nodules with CE3/VHb bacteroids. In contrast to CE/VHb-elicited nodules, nodules with CE3 bacteroids showed a higher number of PQ-responsive genes with increased expression level (Fig. 4a, b). This was evident for stress response related maps (abiotic stress, secondary metabolism, heat shock proteins, glutathione-S-transferase) and for signaling related maps (transcription factors, hormone signaling). These results suggest that nodules with bacteroids lacking VHb are facing a more aggressive oxidative stress condition, something that induces higher gene expression of processes that help cope with the stress.

A comparative analysis on the number of DEG up- or down-regulated in CE3- and CE3/VHb-elicited nodules under PQ stress is presented as a Venn diagram in Fig. 5. Few DEG were common in both comparisons; from these 10% were up-regulated and 3% were down-regulated genes. Notably, 13 genes overlapped between CE3 and CE3/VHb PQ-treated nodules albeit with a different response, being up-regulated in CE_PQ nodules but down-regulated in CE/VHb_PQ nodules (Fig. 5).

In general, our results showed good correspondence among both the GO term and the MapMan enrichment analyses, highlighting functional categories/biological processes commonly observed in the response of plant cells to abiotic stress.

a

regulation of epidermal cell differentiation
lateral root morphogenesis
metal ion transport
water transport
Golgi organization
oligopeptide transport
nucleosome assembly
cellular response to iron ion starvation
plant-type cell wall loosening
syncytium formation
response to carbohydrate stimulus

terpenoid transport
flavonol biosynthetic process
multicellular organismal development
vesicle docking involved in exocytosis
lignan biosynthetic process
long-chain fatty acid metabolic process
phenylpropanoid biosynthetic process
response to salt stress
salicylic acid catabolic process
process involved in reproduction
drought recovery
exocytosis
response to cadmium ion
lignin catabolic process
response to wounding
response to heat
toxin catabolic process
response to arsenic-containing sustance
response to karrikin
regulation of gibberellic acid signaling
raffinose family oligosaccharide
gibberellin mediated signaling pathway
floral organ morphogenesis

-0.02   0   0.02   0.04

b

response to hypoxia
defense response to fungus
transmembrane transport
response to high light intensity
response to hydrogen peroxide
transport
regulation of apoptosis
induction of programmed cell death

Anaerobic respiration
response to zinc ion
positive regulation of seed germination
methylammonium transport
histone H4 acetylation
response to virus
leucine catabolic process
response to cadmium ion
cellular carbohydrate metabolic process
nitrogen fixation

-0.04 -0.02   0   0.02   0.04

CE3_PQ vs CE3_C   CE3/VHb_PQ vs CE3 /VHb

244 down   25 down   469 down
392 up   73 up   194 up

c

regulation of systemic acquired resistance
L-serine biosynthetic process
salicylic acid biosynthetic process
nucleosome assembly
developmental programmed cell death
response to manganese ion
water transport

oxidation-reduction process
high-affinity iron ion transport
multicellular organism growth
response to cadmium ion
cellular response to UV-B
glycogen catabolic process
cellular response to ethylene stimulus
cellular response to nitric oxide
toxin catabolic process
response to oxidative stress
nicotianamine biosynthetic process

-0.04 -0.02   0   0.02   0.04   0.06

Fig. 2 Biological processes significantly enriched in the set of down-regulated (*gray bars*) and up-regulated (*black bars*) DEG according to Gene Ontology analysis. Biological processes enriched exclusively in nodules from comparisons I: CE3-PQ vs CE3_C (**a**), from comparison II:CE3/VHb_PQ vs CE3/VHb_C (**b**) or those commonly enriched in a second level comparison of I vs II (**c**). Fisher's exact test $P \leq 0.01$, FDR ≤ 0.05
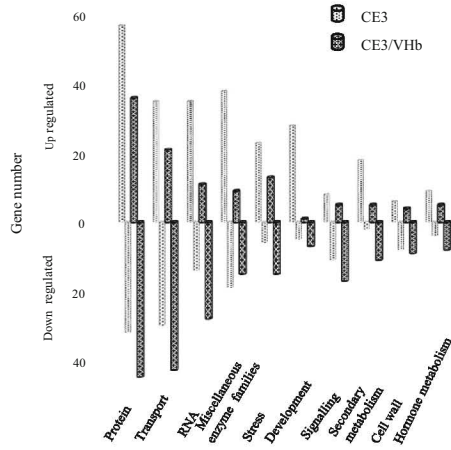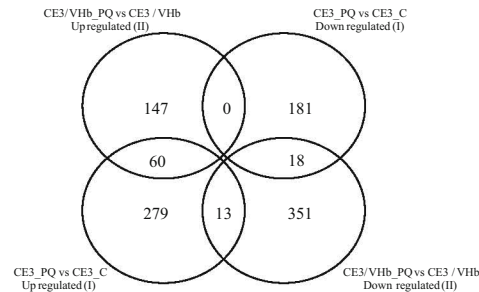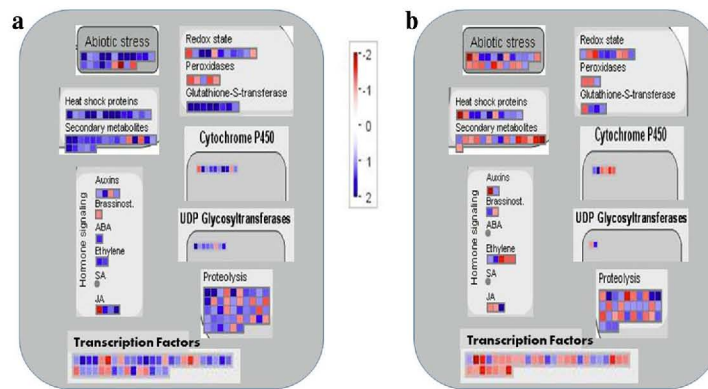

Fig. 3 MapMan functional categories of DEG from comparison I (*light bars*) and II (*dark bars*). The histograms represent the number of up- or down-regulated genes assigned to each category or Bin, designated as indicated below

Fig. 5 Venn diagram representing DEG analyzed by Mapman in PQ-treated nodules elicited by CE3 or CE3/VHb. Number of up- and down-regulated DEG from comparison I or II are shown in different *circles* as indicated

Discussion

This work shows the decreased sensitivity to oxidative stress, generated by PQ treatment, of SNF common bean plants inoculated with an engineered *R. etli* strain expressing the heterologous *Vitreoscilla* hemoglobin. After

PQ treatment plants inoculated with CE3/VHb showed a decrease in nitrogenase activity and xylem sap ureide content of ca. 47% while the decrease observed in CE3 inoculated plants was ca. 68%. These results are in agreement with and add information relevant to the improvement of SNF in common bean inoculated with VHb-engineered *R. etli* grown in control condition, as shown here and in a previous report (Ramírez et al. 1999). While positive effects of VHb expression in various non-symbiotic bacterial species are well documented (Stark et al. 2015), to our knowledge there are no other studies on VHb-engineered rhizobia species in symbiosis with their host legume, besides our previous report (Ramírez et al. 1999). Positive effects of VHb expression in transgenic plants, i.e., Arabidopsis, tobacco or cabbage, growing in control or in stress conditions have also been reported (Holmberg et al. 1997; Li et al. 2005; Wang et al. 2009).

The expression of *vgb* gene as well as the presence of VHb protein was demonstrated for the engineered *R. etli*

Fig. 4 Selected MapMan maps depicting DEG profiles from **a** comparison I (CE3_PQ vs CE3_C) and **b** comparison II (CE3/VHb_PQ vs CE3/VHb_C). Up- or down-regulated genes are false color-coded with increasing *blue* or *red*, respectively, saturating at amplitude of 0.2 (log₂ value) as indicated in the *bar* in the middle

Springer

103

strain used in this work (Ramírez et al. 1999). The improvement in respiration, chemical energy content and nitrogen fixation gene expression in *R. etli* CE3/VHb grown in limited oxygen concentration indicates that VHb expression in bacteroids contributes to better oxygen delivery to the symbiotic terminal oxidase resulting in increased respiratory capacity and N_2 fixation that in turn provides more nitrogen for a better growth of common bean (Ramírez et al. 1999). This is in agreement with the reports on improved SNF common bean performance resulting form inoculation with modified *R. etli* strains showing increased expression of the symbiotic terminal oxidase (Granados-Baeza et al. 2007; Soberón et al. 1999).

The control of redox homeostasis in the legume-rhizobia symbiosis is a key feature for efficient SNF, taking into account the essential signaling role of ROS in the initial steps of the symbiosis as opposed to the increased risk of harm from ROS generation because of the strong biochemical activity in nodules (Becana et al. 2010; Dalton et al. 2009; Mandon et al. 2009). The PQ treatment used in this work increases ROS production and generates oxidative stress at a similar level as that generated by abiotic stresses, such as metal toxicity, salinity or drought (Ramírez et al. 2013). Expression of VHb in *E. coli* and *Enterobacter aerogenes* protects against oxidative stress (Anand et al. 2010; Geckil et al. 2003). We interpret that VHb expression in *R. etli* bacteroids from nodules of plants under oxidative stress, besides improving respiration and N_2 fixation, contributes to buffering the damage caused by increased ROS and controlling redox homeostasis, and this is reflected in nodule cells that showed reprogramming of global gene expression.

Aiming to understand the molecular response to oxidative stress of nodule cells derived from the expression of VHb in the bacteroids, we performed comparative transcriptomic analyses based in hybridization of the Bean Custom Array 90K. This has proven as a reliable tool for analyzing global gene expression in common bean organs, such as nodules under oxidative stress (Ramírez et al. 2013) and roots under nutritional stress (Aparicio-Fabre et al. 2013). The DEG analysis conducted to identify biological processes or functional categories was based in the GO (Ashburner et al. 2000) and MapMan (Thimm et al. 2004) bioinformatics tools adapted to common bean (Ramírez et al. 2013). For this we focused in comparisons I and II that include DEG in oxidative stressed vs control nodules with the engineered or wild type bacteroids.

It is known that nodules from common bean plants under oxidative stress, after PQ exposure, reprogram their transcript profile (Ramírez et al. 2013); data from this work are in agreement with this finding. However, the transcriptomic PQ-response from nodules elicited with CE3 was notably different from that of nodules with VHb-expressing bacteroids. Only ca. 8% of DEG were commonly responsive in both type of nodules under similar oxidative stress. Enriched down-regulated biological processes in CE3_PQ nodules were related to carbon/nitrogen primary metabolism, which are key to nodule function (Ramírez et al. 2013), while these processes were less affected or even up-regulated in CE3/VHb_PQ nodules. Nevertheless, the most evident difference derived from our transcriptome analysis of comparisons I and II was that the response to PQ of CE3-elicited nodules included an increased number of up-regulated DEG as compared to CE3/VHb-elicited nodules. The majority of up-regulated DEG from CE3_PQ nodules were classified into biological processes/functional categories related to defense, response to abiotic stress or signaling.

Examples of up-regulated GO biological processes enriched in CE3_PQ nodules include synthesis of raffinose oligosaccharide and of secondary metabolites. Raffinose oligosaccharides accumulate under drought stress, functioning as osmolytes to maintain cell turgor and to stabilize cell proteins (ElSayed et al. 2014). They may also act as antioxidants to counteract the accumulation of reactive oxygen species (ROS) under stress conditions (Peshev et al. 2013). Phenylpropanoids and flavonoids are widely known to participate in abiotic stress responses in plants; flavonoids have the capacity to absorb the most energetic solar wavelengths (i.e., UV-B and UV-A), to inhibit the generation of ROS and to quench these once they are formed (Ferrer et al. 2008; Brunetti et al. 2013).

The MapMan analysis showed that the Transport (Bin 27) functional category included a large number of DEG; again CE3_PQ nodules showed more up-regulated genes from this category. DEG assigned to this category encode protein, sugar and cation transporters and also transporters from the large ABC (ATP-binding cassette) family, which participate in export or import of various substrates across biological membranes and have an important role in response to abiotic stress (Moon and Jung 2014). The abundant Signaling (Bin 30) functional category included oxidative stress responsive kinases from several families, such as the Mitogen-activated protein kinase (MAPK) family that has been identified as important components in stress signal transduction in response to ABA and antioxidant defense (Danquah et al. 2014). The Hormone metabolism (Bin 17) also showed high number of DEG in PQ-treated nodules, including genes from signaling pathways known to be involved in environmental stresses, such as ethylene, abscisic acid and jasmonic acid (Verma et al. 2016). Hormone signaling involves TF from different families, such as the AP2 family (that includes ERF), WRKY, and bZIP known to be relevant for plant response to abiotic stress (Pieterse et al. 2012). We observed that CE3_PQ nodules highly up-regulate different TF genes, in

agreement with previous report (Ramírez et al. 2013). The TF response in CE3/VHb_PQ nodules was the opposite, showing mostly down-regulation of these genes.

Taken as a whole, data from this work indicate that VHb-expressing *R. etli* bacteroids alleviate the oxidative stress level in nodules of common bean plants exposed to PQ by a major transcriptomic reprogramming. Our transcriptomic datasets provide a basis for deeper analysis, based on biotechnological approaches, of common bean genes that could play key roles in response to oxidative stress. From previous analyses (Ramírez et al. 2013), we predict that similar positive effects would be obtained in CE3/VHb-inoculated common bean plants subjected to drought, salinity or metal toxicity, which are serious constraints in acidic soils from Latin America where this crop is widely grown. This work provides a foundation for exploring inoculants based in VHb-engineered rhizobia that would be relevant for increasing growth and crop production of legumes grown under oxidative stress condition.

References

Anand A, Duk BT, Singh S, Akbas MY, Webster DA, Stark BC, Dikshit KL (2010) Redox-mediated interactions of VHb (*Vitreoscilla* haemoglobin) with OxyR: novel regulation of VHb biosynthesis under oxidative stress. Biochem J 426:271–280. doi:10.1042/BJ20091417

Aparicio-Fabre R, Guillén G, Loredo M, Arellano J, Valdés-López O, Ramírez M, Íñiguez LP, Panzeri D, Castiglioni B, Cremonesi P, Strozzi F, Stella A, Girard L, Sparvoli F, Hernández G (2013) Common bean (*Phaseolus vulgaris* L.) PvTIFY orchestrates global changes in transcript profile response to jasmonate and phosphorus deficiency. BMC Plant Biol 13:1. doi:10.1186/1471-2229-13-26

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet 25:25–29. doi:10.1038/75556

Becana M, Matamoros MA, Udvardi M, Dalton DA (2010) Recent insights into antioxidant defenses of legume root nodules. New Phytol 188:960–976. doi:10.1111/j.14698137.2010.03512.x

Boldt R, Zrenner R (2003) Purine pyrimidine biosynthesis in higher plants. Physiol Plant 117:297–304

Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp): model food legume. Plant Soil 252:55–128. doi:10.1023/A:1024146710611

Brunetti C, Di Ferdinando M, Fini A, Pollastri S, Tattini M (2013) Flavonoids as antioxidants and developmental regulators: relative significance in plants and humans. Int J Mol Sci 14:3540–3555. doi:10.3390/ijms14023540

Dalton DA, Boniface C, Turner Z, Lindahl A, Kim HJ, Jelinek L, Govindarajulu M, Finger RE, Taylor CG (2009) Physiological roles of glutathione S-transferases in soybean root nodules. Plant Physiol 150:521–530. doi:10.1104/pp.109.136630

Danquah A, de Zelicourt A, Colcombet J, Hirt H (2014) The role of ABA and MAPK signaling pathways in plant abiotic stress responses. Biotechnol Adv 32:40–52. doi:10.1016/j.biotechadv.2013.09.006

Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression hybridization array data repository. Nucleic Acids Res 30:207–210. doi:10.1093/nar/30.1.207

ElSayed AI, Rafudeen MS, Golldack D (2014) Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. Plant Biol 16:1–8. doi:10.1111/plb.12053

Ferrer JL, Austin MB, Stewart C Jr, Noel JP (2008) Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiol Biochem 46:356–370. doi:10.1016/j.plaphy.2007.12.009

Geckil H, Stark BC, Webster DA (2001) Cell growth and oxygen uptake of *Escherichia coli* and *Pseudomonas aeruginosa* are differently effected by the genetically engineered *Vitreoscilla* hemoglobin gene. J Biotechnol 85:57–66. doi:10.1016/S0168-1656(00)00384-9

Geckil H, Gencer S, Kahraman H, Erenler SO (2003) Genetic engineering of *Enterobacter aerogenes* with the *Vitreoscilla* hemoglobin gene: cell growth, survival, and antioxidant enzyme status under oxidative stress. Res Microbiol 154:425–431. doi:10.1016/S0923-2508(03)00083-4

Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. Plant Physiol 131:872–877. doi:10.1104/pp.017004

Granados-Baeza MJ, Gómez-Hernández N, Mora Y, Delgado MJ, Romero D, Girard L (2007) Novel reiterated Fnr-type proteins control the production of the symbiotic terminal oxidase *cbb₃* in *Rhizobium etli* CFN42. Mol Plant Microbe Interact 20:1241–1249. doi:10.1094/MPMI-20-10-1241

Hardison R (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. J Exp Biol 201:1099–1117

Hardy RW, Holsten RD, Jackson EK, Burns RC (1968) The acetylene–ethylene assay for N_2 fixation: laboratory and field evaluation. Plant Physiol 43:1185–1207

Hernández G, Ramírez M, Valdes-López O, Tesfaye M, Graham MA, Czechowski T, Schlereth A, Wandrey M, Erban A, Cheung F, Wu H, Lara M, Town CD, Kopka J, Udvardi MK, Vance CP (2007) Phosphorus stress in common bean: root transcript and metabolic responses. Plant Physiol 144:752–767. doi:10.1104/pp.107.096958

Holmberg N, Lilius G, Bailey JE, Bülow L (1997) Transgenic tobacco expressing *Vitreoscilla* hemoglobin exhibits enhanced growth and altered metabolite production. Nat Biotechnol 15:244–247. doi:10.1038/nbt0397-244

Joshi M, Mande S, Dikshi KL (1998) Hemoglobin biosynthesis in *Vitreoscilla stercoraria* DW: cloning, expression, and characterization of a new homolog of a bacterial globin gene. Appl Environ Microbiol 64:2220–2228

Kaur R, Pathania R, Sharma V, Mande SC, Dikshit KL (2002) Chimeric *Vitreoscilla* hemoglobin (VHb) carrying a flavoreductase domain relieves nitrosative stress in *Escherichia coli*: new insight into the functional role of VHb. Appl Environ Microbiol 68:152–160. doi:10.1128/AEM.68.1.152-160.2002

Khosla C, Bailey JE (1988) Heterologous expression of a bacterial haemoglobin improves the growth properties of recombinant *Escherichia coli*. Nature 18:633–635. doi:10.1038/331633a0

Klie S, Nikoloski Z (2012) The choice between MapMan and Gene Ontology for automated gene function prediction in plant science. Front Genet 3:115. doi:10.3389/fgene.2012.00115

Li X, Peng RH, Fan HQ, Xiong AS, Yao QH, Cheng ZM (2005) *Vitreoscilla* hemoglobin overexpression increases submergence tolerance in cabbage. Plant Cell Rep 23:710–715. doi:10.1186/s12870-016-0728-1

Malhotra S, Sowdhamini R (2014) Interactions among plant transcription factors regulating expression of stress-responsive genes. Bioinform Biol Insights 8:193–198. doi:10.4137/BBI.S16313

Mandon K, Pauly N, Boscari A, Brouquisse R, Frendo P, Demple B, Puppo A (2009) ROS in the legume-Rhizobium symbiosis. In: Del Río LA, Puppo A (eds) Reactive oxygen species in plant signaling. Springer, Berlin, pp 135–147

Mendoza-Soto AB, Naya L, Leija A, Hernández G (2015) Responses of symbiotic nitrogen-fixing common bean to aluminum toxicity and delineation of nodule responsive microRNAs. Front Plant Sci 6. doi:10.3389/fpls.2015.00587

Moon S, Jung KH (2014) Genome-wide expression analysis of rice ABC transporter family across spatio-temporal samples and in response to abiotic stresses. J Plant Physiol 171:1276–1288. doi:10.1016/j.jplph.2014.05.006

Naya L, Paul S, Valdés-López O, Mendoza-Soto AB, Nova-Franco B, Sosa-Valencia G, Reyes JL, Hernández G (2014) Regulation of copper homeostasis and biotic interactions by microRNA 398b in common bean. PLoS One 9:e84416. doi:10.1371/journal.pone.0084416

Peralta H, Mora Y, Salazar E, Encarnación S, Palacios R, Mora J (2004) Engineering the nifH promoter region and abolishing poly-β-hydroxybutyrate accumulation in *Rhizobium etli* enhance nitrogen fixation in symbiosis with *Phaseolus vulgaris*. Appl Environ Microbiol 70:3272–3281. doi:10.1128/AEM.70.6.3272-3281.2004

Peshev D, Vergauwen R, Moglia A, Hideg É, Van den Ende W (2013) Towards understanding vacuolar antioxidant mechanisms: a role for fructans? J Exp Bot 64:1025–1038. doi:10.1093/jxb/ers377

Pieterse CM, Van der Does D, Zamioudis C, Leon-Reyes A, Van Wees SC (2012) Hormonal modulation of plant immunity. Annu Rev Cell Dev Biol 28:489–521. doi:10.1146/annurev-cellbio-092910-154055

Ramírez M, Valderrama B, Arrendondo-Peter R, Sobéron M, Mora J, Hernández G (1999) *Rhizobium etli* genetically engineered for the heterologous expression of *Vitreoscilla* sp. hemoglobin: effect on free-living symbiosis. Mol Plant Microbe Interact 12:1008–1015. doi:10.1094/MPMI.1999.12.11.1008

Ramírez M, Guillén G, Fuentes SI, Iñiguez LP, Aparicio-Fabre R, Zamorano-Sánchez D, Encarnación-Guevara S, Panzeri D, Castiglioni B, Cremonesi P, Strozzi F, Stella A, Girard L, Sparvoli F, Hernández G (2013) Transcript profiling of common bean nodules subjected to oxidative stress. Physiol Plant 149:389–407. doi:10.1111/ppl.12040

Routledge R (1998) Fisher's exact test. In: Armitage P, Colton T (eds) Encyclopedia of biostatistics, vol 2. Wiley, New York, pp 1519–1523

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MM, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, Jackson SA (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet 46:707–713. doi:10.1038/ng.3008

Segovia L, Young JPW, Martínez-Romero E (1993) Reclassification of American *Rhizobium leguminosarum* biovar phaseoli type I strains as *Rhizobium etli* sp. nov. Int J Syst Evol Microbiol 43:374–377. doi:10.1099/00207713-43-2-374

Smyth GK (2004) Linear models empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3:1–26

Soberón M, López O, Morera C, Girard ML, Tabche M, Miranda J (1999) Enhanced nitrogen fixation in a *Rhizobium etli* ntrC mutant that overproduces the *Bradyrhizobium japonicum* symbiotic terminal oxidase cbb_3. Appl Environ Microbiol 65:2015–2019

Stark BC, Pagilla KR, Dikshit KL (2015) Recent applications of *Vitreoscilla* hemoglobin technology in bioproduct synthesis and bioremediation. Appl Microbiol Biotechnol 99:1627–1636. doi:10.1007/s00253-014-6350-y

Suárez R, Wong A, Ramírez M, Barraza A, Orozco MDC, Cevallos MA, Lara M, Hernández G, Iturriaga G (2008) Improvement of drought tolerance and grain yield in common bean by overexpressing trehalose 6-phosphate synthase in rhizobia. Mol Plant Microbe Interact 21:958–966. doi:10.1094/MPMI-21-7-0958

Summerfield RJ, Huxley PA, Minchin FR (1977) Plant husbandry management techniques for growing grain legumes under simulated tropical conditions in controlled environments. Exp Agric 13:113–121. doi:10.1017/S0014479700007638

Talbi C, Sánchez C, Hidalgo-Garcia A, González EM, Arrese-Igor C, Girard L, Delgado MJ (2012) Enhanced expression of *Rhizobium etli* cbb_3 oxidase improves drought tolerance of common bean symbiotic nitrogen fixation. J Exp Bot 63:5035–5043. doi:10.1093/jxb/ers101

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) MapMan: a user-driven tool to display genomic data sets onto diagrams of metabolic pathways other biological processes. Plant J 37:914–939. doi:10.1111/j.1365-313X.2004.02016.x

Vadez V, Berger JD, Warkentin T, Asseng S, Kumar PR, Rao KPC (2012) Adaptation of grain legumes to climate change: a review. Agron Sustain Dev 32:31–44. doi:10.1007/s13593-011-0020-6

Verma V, Ravindran P, Kumar PP (2016) Plant hormone-mediated regulation of stress responses. BMC Plant Biol 16:1. doi:10.1186/s12870-016-0771-y

Vogels GD, Van Der Drift C (1970) Differential analyses of glyoxylate derivatives. Anal Biochem 33:143–157

Wakabayashi S, Matsubara H, Webster DA (1986) Primary sequence of a dimeric bacterial haemoglobin from *Vitreoscilla*. Nature 322:481–483. doi:10.1038/322481a0

Wang Z, Xiao Y, Chen W, Tang K, Zhang L (2009) Functional expression of *Vitreoscilla* hemoglobin (VHb) in Arabidopsis relieves submergence, nitrosative, photo-oxidative stress and enhances antioxidants metabolism. Plant Sci 176:66–77. doi:10.1186/s12870-016-0728-1

Wittenberg JB (2012) Reactivity and function of leghemoglobin. In: Caughey W (ed) Biochemical and clinical aspects of oxygen. Academic Press, London, pp 53–68

106

C. Anexo

The Micro-RNA172c-APETALA2-1 Node as a Key Regulator of the Common Bean-*Rhizobium etli* Nitrogen Fixation Symbiosis[1][OPEN]

Bárbara Nova-Franco, Luis P. Íñiguez, Oswaldo Valdés-López, Xochitl Alvarado-Affantranger, Alfonso Leija, Sara I. Fuentes, Mario Ramírez, Sujay Paul, José L. Reyes, Lourdes Girard, and Georgina Hernández*

Centro de Ciencias Genómicas (B.N.-F., L.P.I., A.L., S.I.F., M.R., S.P., L.G., G.H.), Laboratorio Nacional de Microscopía Avanzada (X.A.-A.), and Departamento de Biología Molecular de Plantas (J.L.R.), Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, Mexico; and Laboratorio de Genómica Funcional de Leguminosas, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlalnepantla, Edo de Mexico 54090, Mexico (O.V.-L.)

ORCID IDs: 0000-0002-2481-7715 (X.A.-A.); 0000-0001-5129-9741 (J.L.R.); 0000-0003-0004-9106 (L.G.); 0000-0001-6214-311X (G.H.).

Micro-RNAs are recognized as important posttranscriptional regulators in plants. The relevance of micro-RNAs as regulators of the legume-rhizobia nitrogen-fixing symbiosis is emerging. The objective of this work was to functionally characterize the role of micro-RNA172 (miR172) and its conserved target APETALA2 (AP2) transcription factor in the common bean (*Phaseolus vulgaris*)-*Rhizobium etli* symbiosis. Our expression analysis revealed that mature miR172c increased upon rhizobial infection and continued increasing during nodule development, reaching its maximum in mature nodules and decaying in senescent nodules. The expression of *AP2-1* target showed a negative correlation with miR172c expression. A drastic decrease in miR172c and high *AP2-1* mRNA levels were observed in ineffective nodules. Phenotypic analysis of composite bean plants with transgenic roots overexpressing miR172c or a mutated *AP2-1* insensitive to miR172c cleavage demonstrated the pivotal regulatory role of the miR172 node in the common bean-rhizobia symbiosis. Increased miR172 resulted in improved root growth, increased rhizobial infection, increased expression of early nodulation and autoregulation of nodulation genes, and improved nodulation and nitrogen fixation. In addition, these plants showed decreased sensitivity to nitrate inhibition of nodulation. Through transcriptome analysis, we identified 114 common bean genes that coexpressed with *AP2-1* and proposed these as being targets for transcriptional activation by *AP2-1*. Several of these genes are related to nodule senescence, and we propose that they have to be silenced, through miR172c-induced *AP2-1* cleavage, in active mature nodules. Our work sets the basis for exploring the miR172-mediated improvement of symbiotic nitrogen fixation in common bean, the most important grain legume for human consumption.

The symbiotic nitrogen fixation (SNF) occurring in the legume-rhizobia symbiosis takes place in root-developed specialized organs called nodules. Nodulation is a complex process that involves communication between rhizobia and legumes through molecular signals, including rhizobial lipochitin-oligosaccharide symbiotic signals known as nodulation factors (NFs), that triggers a root-signaling cascade essential for rhizobia infection (for review, see Crespi and Frugier, 2008; Oldroyd and Downie, 2008; Kouchi et al., 2010; Murray, 2011; Oldroyd, 2013).

Nuclear Ca^{2+} oscillations, or calcium spiking, is one of the earliest NF-induced responses in legume root hairs. Perception and transduction of the calcium-spiking signal involves Ca^{2+}/CALMODULIN-DEPENDENT PROTEIN KINASE (CCaMK), which interacts with the nuclear protein CYCLOPS, and other downstream components, such as the transcriptional regulators NODULATION SIGNALING PATHWAY (NSP1)/NSP2, NUCLEAR FACTOR YA1 (NF-YA1)/YA2, ETHYLENE-RESPONSIVE FACTOR REQUIRED FOR NODULATION1, and NODULE INCEPTION (NIN), which, in turn, control the expression of early nodulation genes.

Legumes strictly regulate the number of developing nodules in response to internal and external cues. An important internal cue is the systemic feedback regulatory mechanism called autoregulation of nodulation (AON), which consists of root-derived and shoot-derived long-distance signals. AON is initiated in response to rhizobial NF during nodule primordium formation by the root production of CLAVATA3/Embryo-Surrounding Region Protein-related (CLE) peptides (Reid et al., 2011a). Some CLE peptides are predicted, although not proven, to act as the ligand for a shoot CLAVATA1-like Leu-rich

repeat receptor kinase (Okamoto et al., 2009). Activation of this receptor is proposed to initiate the production of a shoot-derived inhibitor that is transported to the root, where it inhibits further nodule formation (for review, see Magori and Kawaguchi, 2009; Ferguson et al., 2010; Kouchi et al., 2010; Reid et al., 2011b). Soil nitrogen is an important external cue for the control of nodulation (Streeter and Wong, 1988). Recent work indicates that nitrate inhibition of nodulation may function via an up-regulation of a nitrate-induced CLE peptide that is perceived by a Leu-rich repeat receptor kinase in the root (Okamoto et al., 2009; Reid et al., 2011a).

In recent years, microRNAs (miRNAs), a class of noncoding RNA 21 to 24 nucleotides in length, have been identified as central regulators of gene expression in plants, controlling fundamental processes such as stress response, phytohormone regulation, organ morphogenesis, and development (Rogers and Chen, 2013). The plant miRNA precursors, generally transcribed by RNA polymerase II, adopt stem-loop structures that are processed by several enzymes and generate mature miRNAs that are exported to the cytosol. The role of miRNAs in posttranscriptional regulation is mediated by the almost perfect complementarity with their target mRNAs, thereby causing their degradation or their translational inhibition (Zhang et al., 2006; Rogers and Chen, 2013).

Progress in high-throughput sequencing technologies has facilitated the genome-wide identification of large miRNA populations and their target mRNAs in different legumes (for review, see Simon et al., 2009; Bazin et al., 2012; Bustos-Sanmamed et al., 2013). Conserved and legume-specific miRNA families differentially expressed during nodule organogenesis have been reported for *Medicago truncatula*, soybean (*Glycine max*), and *Lotus japonicus* (Subramanian et al., 2008; Lelandais-Brière et al., 2009; De Luis et al., 2012; Turner et al., 2012; Dong et al., 2013). Recently, Formey et al. (2014) identified miRNAs from *M. truncatula* roots that respond to treatments with purified NF. However, evidence for the functional involvement of miRNAs in rhizobial infection and the functionality of nodules has only been obtained for a small number of candidates. The involvement of *M. truncatula* microRNA166 (miR166), miR169, and miR164 in nodule development has been reported. miR169 controls nodule meristem maintenance through the repression of *NF-YA1* (previously called *HAEM ACTI-VATOR PROTEIN2-1*), a nodule-responsive transcription factor (TF; Combier et al., 2006), while miR166 and its target gene, *HOMEODOMAIN-LEUCINE ZIPPER protein of class III* TF, regulate meristem activity and vascular differentiation in roots and nodules (Boualem et al., 2008). The overexpression of miR164, a conserved miRNA targeting *NAC1* (for no apical meristem [NAM], Arabidopsis transcription activation factor [ATAF1-2], and cup-shaped cotyledon [CUC2] domain1) TF in roots, affected nodule organogenesis presumably through the deregulation of auxin responses (D'haeseleer et al., 2011). In soybean, the overexpression of miR482, miR1512, and miR1515 results in increased nodule numbers without affecting root development or the number of nodule primordia (Li et al., 2010). Recently, Turner et al. (2013) reported that the overexpression of soybean miR160, which targets a set of repressor auxin response factors, resulted in an enhanced sensitivity to auxin and inhibition of nodule development, apparently through a reduction in cytokinin sensitivity. Likewise, the overexpression of *M. truncatula* miR160 affected root gravitropism and nodule number (Bustos-Sanmamed et al., 2013). Specific variants of *L. japonicus* and *M. truncatula* miR171 target the GRAS-family *NSP2* TF, a key regulator of the common symbiotic pathway for rhizobial and arbuscular mychorrizal symbioses (Ariel et al., 2012; De Luis et al., 2012; Lauressergues at al., 2012). *M. truncatula* roots over-expressing miR171h showed deceased arbuscular mychorrizal colonization (Lauressergues at al., 2012), while in *L. japonicus*, miR171c regulates the maintenance and establishment of the nodule but not the bacterial infection (De Luis et al., 2012). In addition, the role of *L. japonicus* miR397 in nodule copper homeostasis, through the regulation of a member of the laccase copper protein family, has been documented (De Luis et al., 2012).

Common bean (*Phaseolus vulgaris*) is the most important crop legume for human consumption and the main source of proteins for people in African and Central/South American countries (Broughton et al., 2003). Our research is focused on identifying and functionally characterizing common bean miRNAs. High-throughput sequencing of small RNAs generated from different organs of common bean let us identify more than 100 conserved miRNAs and to predict novel miRNAs (Peláez et al., 2012). Common bean miRNAs that respond to drought, salinity, nutrient deficiencies, or metal toxicity stresses have been identified, and their target genes have been predicted or validated (Arenas-Huertero et al., 2009; Valdés-López et al., 2010; Contreras-Cubas et al., 2012). The roles of miR399 in the common bean root response to phosphorus deficiency (Valdés-López et al., 2008) and of miR398 in the regulation of copper homeostasis and response to biotic interactions (Naya et al., 2014) have been demonstrated. In this work, we analyzed the role of miR172 in common bean roots and nodules.

miR172 is conserved in all angiosperm lineages; its conserved targets are TFs from the *APETALA2* (*AP2*) family. The miR172 node that involves the miR156 node is one of the best-understood networks that regulate developmental timing in Arabidopsis (*Arabidopsis thaliana*) and other plants (Rubio-Somoza and Weigel, 2011). Aukerman and Sakai (2003) first described that miR172 promotes flowering by repressing *AP2* genes, primarily through translation inhibition (Chen, 2004) but also through mRNA cleavage (Kasschau et al., 2003; Jung et al., 2007). In addition, the miR172 node regulates the juvenile-to-adult phase transition during shoot development (Wu et al., 2009; Huijser and Schmid, 2011). Such developmental transitions are coordinated by the antagonistic activities of the miR156 and miR172 nodes. miR156 targets a subset of SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE (SPL) TFs that bind to the miR712 promoter and directly promote its transcription, resulting in *AP2* silencing (Wu et al., 2009).

108

miR172 has been identified in the legumes *M. truncatula*, *L. japonicus*, soybean, and common bean; it is highly accumulated in mature nodules relative to other plant tissues (Lelandais-Brière et al., 2009; Wang et al., 2009; Valdes-López et al., 2010; De Luis et al., 2012). Yan et al. (2013) reported the regulation of soybean nodulation by miR172 that involves a complex regulatory circuit in which miR156 regulates miR172 expression, which, in turn, controls the level of its *AP2* target gene. They propose that *AP2*, directly or indirectly, controls the expression of nonsymbiotic hemoglobin, which is essential for regulating the levels of nodulation and nitrogenase activity (Yan et al., 2013). Very recently, Wang et al. (2014) demonstrated that soybean miR172c modulates rhizobial infection and nodule organogenesis. They showed that miR172c regulates nodule formation by repressing its target gene *NODULE NUMBER CONTROL1* (*NNC1*), an AP2 TF, which directly targets and represses the early nodulin gene *ENOD40* that plays a key role in nodulation. However, it is not clear whether miR172c controls early responses that are critical to establish a functional symbiosis between legumes and rhizobia.

The aim of this work was to analyze the role of the miR172 node in the common bean-rhizobia symbiosis. We determined an increased expression of miR172c upon rhizobial infection and during nodulation, showing a negative correlation with *AP2-1* expression. We achieved the overexpression of miR172c and of a mutagenized *AP2-1* insensitive to miR172 cleavage in composite common bean plants. Common bean plants with increased miR172c levels showed an improved symbiotic phenotype as well as lower sensitivity to nitrate inhibition of nodulation. We explored the possible role of AP2-1 as a transcriptional activator and/or repressor. Candidate target genes for downstream transcriptional activation by AP2-1 were identified; these could be relevant in the nodule senescence process. Our work extends the knowledge of miR172 function in the nodulation of common bean, an agronomically important legume.

RESULTS

Common Bean miR172 Isoforms and Target Genes

The Arabidopsis genome contains five loci that generate miR172 isoforms miR172a to miR172e, while 12 miR172 isoforms are reported for soybean (www.mirbase.org, version 20). The high-throughput small RNA sequencing analysis by Peláez et al. (2012) led us to identify four isoforms of common bean miR172.

In this work, we analyzed the recently published (Schmutz et al., 2014; www.phytozome.net/commonbean.php, v1.0) common bean genome sequence and identified six *MIR172* loci that map in different common bean chromosomes. The most stable secondary structure of the miR172 precursors was predicted, and these showed the expected stem-loop structure. The six isoforms of mature common bean miR172, 20 or 21 nucleotides long, were designated miR172a to miR172f (Supplemental Fig. S1). The nucleotide sequences of miR172a, miR172b, and

miR172c isoforms differ. However, although encoded by different loci, the sequences of miR172e and miR172f are identical to miR172a, while miR172d is identical to miR172c (Supplemental Fig. S1).
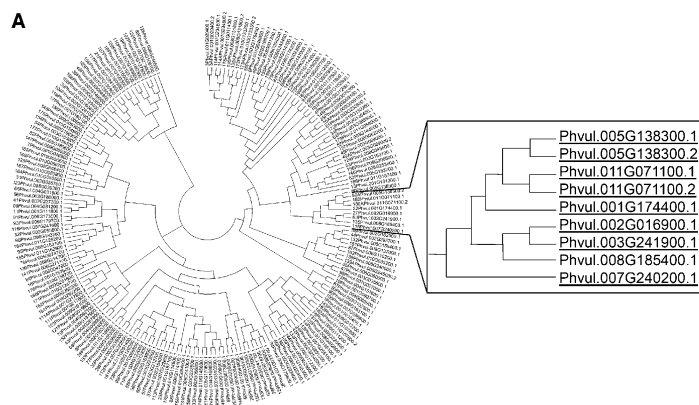
The conserved targets for miR172 in different plants are transcripts that encode TFs from the *AP2* superfamily. In Arabidopsis, six *AP2* TF genes, *AP2*, *TARGET OF EARLY ACTIVATION TAGGED1* (*TOE1*), *TOE2*, *TOE3*, *SCHLAFMUTZE*, and *SCHNARCHZAPFEN*, which act as floral repressors, are targeted by miR172 and silenced through translation inhibition or cleavage (Aukerman and Sakai, 2003; Schmid et al., 2003; Chen, 2004; Jung et al., 2007). In soybean, 10 *AP2* TF genes have been proposed as miR172 targets (Song et al., 2011). A recent analysis of the common bean transcriptome by RNA sequencing (RNA-seq) combined with available gene calls (O'Rourke et al., 2014) identified 202 transcripts encoding AP2-type TFs. The phylogenetic tree generated using sequence alignments of the common bean AP2 proteins is depicted in Figure 1A. From the whole set (202), we identified eight *AP2* transcripts, encoded by six loci, with putative miR172 binding sites within their coding regions. The six predicted *AP2* target genes showed base pairing with the three different miR172 isoforms (Fig. 1B). In each case, the penalty score for miRNA:mRNA (Jones-Rhoades and Bartel, 2004) was low; the highest score was observed in Phvul.007G240200, with the three miR172 isoforms. From the predicted *AP2* targets, Phvul.005G138300, hereafter denominated as *AP2-1*, has been experimentally validated as a target of common bean miR172 (Arenas-Huertero et al., 2009). In addition, Phvul.011G071100 was identified as a target in a common bean degradome analysis (D. Formey, L.P. Íñiguez, P. Peláez, Y.F. Li, R. Sunkar, F. Sánchez, J.L. Reyes, and G. Hernández, unpublished data). Interestingly, the transcripts of *AP2* predicted targets were organized in a single clade of the phylogenetic tree (Fig. 1A). However, this clade also includes the Phvul.008G185400.1 transcript, which has an AP2 domain (http://www.phytozome.net/commonbean.php) but lacks a detectable miR172 binding site and, therefore, is not proposed as a target (Fig. 1A). The recently published *Phaseolus vulgaris* Gene Expression Atlas (*Pv* GEA; O'Rourke et al., 2014) showed a very low expression of this *AP2* gene in all the tissues reported (reads per kilobase per million = 6, highest values in leaves and pods). Therefore, Phvul.008G185400 *AP2* is perhaps highly expressed in tissue-, development-, or environment-specific conditions not yet analyzed and its transcript level could be regulated by factors other than miR172. In addition, we could not detect a Phvul.008G185400.1 ortholog in the soybean genome sequence, perhaps indicating that it could be a pseudogene.

Differential Expression of miR172, Predicted *AP2* Target Genes, miR156, and *SPL6* in Plant Tissues

The differential expression of miR172 isoforms in plant organs/tissues at different developmental stages has been reported for Arabidopsis and soybean (Aukerman

Figure 1. Common bean *AP2* transcripts with predicted miR172 binding sites. A, Neighbor-joining tree of AP2 proteins retrieved from the common bean genome sequence (http://www.phytozome.net/commonbean.php, v1.0). The clade including *AP2* transcripts with miR172 binding sites (underlined) is shown in the inset. B, Pairing of the three different miR172 isoforms (a–c) with the predicted binding sites of the six different AP2 transcripts highlighted in A. Watson-Crick base pairing is indicated by lines, G:U base pairing is indicated by circles, and dashes indicate mismatches. Penalty scores, shown in parentheses, were calculated as described by Jones-Rhoades and Bartel (2004).

A

Phvul.005G138300.1
Phvul.005G138300.2
Phvul.011G071100.1
Phvul.011G071100.2
Phvul.001G174400.1
Phvul.002G016900.1
Phvul.003G241900.1
Phvul.008G185400.1
Phvul.007G240200.1

| **B**	miR172a	miR172b	miR172c
	miRNA:mRNA pairing (penalty score)		
Phvul.005G138300	5'lllllllllll°lllll3' (0.5)	5'lllllllllll°lllll3' (1.0)	5'lllllllllll°lllll°3' (1.0)
Phvul.011G071100	5'lllllllllll°lllll3' (0.5)	5'lllllllllll°lllll3' (1.0)	5'lllllllllll°lllll°3' (1.0)
Phvul.001G174400	5'lllllllllll°lllll-3' (1.5)	5'lllllllllll°lllll-3' (1.5)	5'lllllllllll°lllll3' (0.5)
Phvul.002G016900	5'lllllllllll°lllll-3' (1.5)	5'lllllllllll°lllll-3' (1.5)	5'lllllllllll°lllll3' (0.5)
Phvul.003G241900	5'lllllllllll°lllll-3' (1.5)	5'lllllllllll°lllll-3' (1.5)	5'lllllllllll°lllll3' (0.5)
Phvul.007G240200	5'-lllllllllll°lllll-3' (2.5)	5'-lllllllllll°lllll-3' (2.5)	5'-lllllllllll°lllll-3' (2.5)

and Sakai, 2003; Wu et al., 2009; Yan et al., 2013; Wang et al., 2014). In this work, we performed expression analyses of the miR172 isoforms and their putative *AP2* target genes (Fig. 1) in different tissues of SNF common bean plants 18 d post inoculation (dpi) with *Rhizobium etli* (Fig. 2).

Northern-blot analysis revealed that mature miR172 transcripts were most abundant in nodules followed by flowers (Fig. 2A). The miR172a probe was used for blot hybridization, but the quantified signals might reflect the combined levels of miR172 isoforms whose sequences differ only in two nucleotides (Supplemental Fig. S1). For real-time quantitative reverse transcription (qRT)-PCR expression analysis, specific primers were synthesized for each of the miR172 isoforms (a, b, and c; Supplemental Fig. S1; Supplemental Table S1). The data obtained by northern-blot and qRT-PCR expression analyses showed similar trends regarding the highest levels in nodules and flowers and the lowest levels in roots and leaves (Fig. 2, A and B). The observed variation in the levels of cumulative miR172 expression for each tissue may be attributable to the different sensitivities of the two methods. In addition, qRT-PCR analysis revealed differential expression of the miR172 isoforms among tissues, especially in those tissues with higher cumulative levels. Nodules showed the highest level of miR172c and very low amounts of miR172a and miR172b, while flowers showed the highest level of miR172b, followed by miR172c and a low amount of miR172a (Fig. 2B).

The transcript levels of each of the *AP2* TF genes proposed as miR172 targets (Fig. 1) were determined by qRT-PCR (Fig. 2C) in tissues from SNF bean plants (Fig. 2C). Cumulative *AP2* transcript levels were very high in roots and very low in nodules, thus showing a negative correlation with cumulative miR172 levels in these tissues (Fig. 2, B and C). In roots, the most highly expressed of the *AP2* genes were Phvul.005G138300 (*AP2-1*) and Phvul.011G071100; *AP2-1* was also highly expressed in embryonic leaves (Fig. 2C). However, a distinct pattern was observed in flowers, where the expression of these two genes was negligible and Phvul.001G174400, Phvul.003G241900, and Phvul.002G16900 were highly expressed.

The miR156 node has been implicated in upstream negative regulation of the miR172 node (Rubio-Somoza and Weigel, 2011). Arabidopsis miR156 represses miR172 expression by targeting members of the *SPL* family of TFs that directly bind to the *MIR172* promoter and positively regulate its expression (Wu et al., 2009). Here,
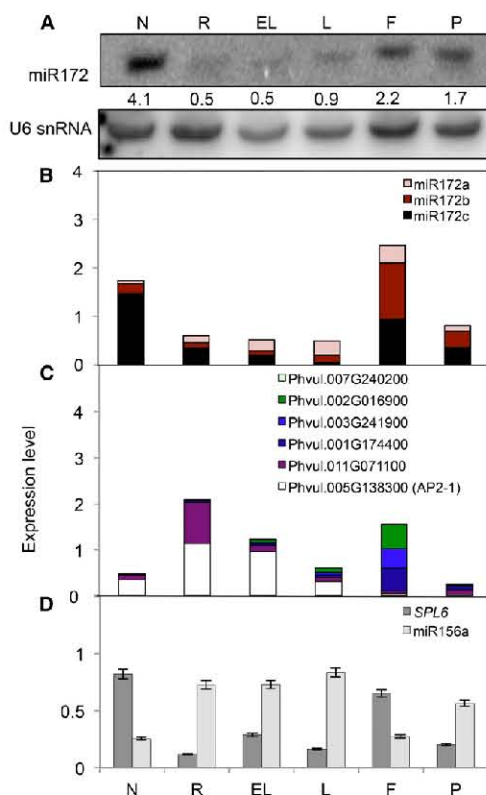
Figure 2. Expression analysis of miR172, *AP2* target genes, miR156a, and its *SPL* transcription factor target gene in different tissues of *R. etli* CE3-inoculated common bean plants (18 dpi). A, Mature miR172 levels detected by northern-blot analysis using U6 small nuclear RNA (snRNA) as a loading control for normalization. Signal intensities of the miR172 and U6 hybridization bands for each tissue were determined to calculate normalized expression levels. Numbers in each lane indicate normalized values of miR172 signal intensity. N, Nodules; R, roots; EL, embryonic leaves; L, leaves; F, flowers; P, pods. B to D, Transcript levels of mature miR172 isoforms (B; Supplemental Fig. S1), predicted *AP2* target genes (C; Fig. 1), and mature miR156a and its target gene *SPL6* (D) determined by qRT-PCR. Expression level refers to gene expression, based on threshold cycle (C_t) value, normalized with the expression of the housekeeping miR159 or *UBIQUITIN CONJUGATING ENZYME9* (*UBC9*) gene.

we determined the levels of mature miR156a (Peláez et al., 2012) in common bean tissues. The expression of miR156a was elevated in roots and leaves but low in nodules and flowers, showing an opposite trend of the cumulative expression of miR172 (Fig. 2D). We identified 32 *SPL* genes in the common bean genome, and 14 of these showed putative miR156 binding sites, including Phvul.009G165100, which was validated as a common bean miR156a target (Arenas-Huertero et al., 2009).

Comparative sequence analysis with the Arabidopsis *SPL* gene family indicated that the common bean Phvul.009G165100 *SPL* gene is an ortholog to Arabidopsis *SPL6*. We analyzed the expression of the validated miR156a target *SPL6* gene in different common bean organs. As shown in Figure 2D, common bean *SPL6* was highly expressed in nodules and flowers while it was decreased in roots and leaves, thus showing a negative correlation with miR156 expression. We also searched for SPL transcription factor binding sites (TFBS) in the 5′ (promoter) region of each of the six *MIR172* loci mapped in the genome, but we could not identify any, while 35 other TFBS were present in one or more of these loci (Supplemental Table S2).

Our data (Fig. 2) showed miR172c as the isoform with the highest expression in nodules and low expression in roots. Its expression pattern is opposite that of *AP2-1* (Phvul.005G138300), the experimentally validated target (Arenas-Huertero et al., 2009) that showed the highest expression in roots. Therefore, we then focused our analysis on miR172c and *AP2-1* in common bean plants interacting with *R. etli*.

Expression Analysis of miR172 and *AP2-1* during Symbiosis

To assess a possible role of miR172c/*AP2-1* in the common bean-*R. etli* symbiosis, we determined their expression in inoculated roots (Fig. 3) and in effective nodules at different developmental stages (Fig. 4A).

To analyze miR172c/*AP2-1* regulation at early stages of the symbiosis, common bean plantlets were grown in plastic square bioassay dishes and *R. etli* inoculum was applied directly to the roots. For gene expression analysis, only the responsive root zone, where initial bacteria-host recognition takes place, was collected at the initial time (0 h) and at 3, 6, 12, 24, and 48 h post inoculation (hpi). The mature miR172c level increased significantly after 6 h, whereas the high expression level of the *AP2-1* target gene decreased significantly after 12 h; both transcript levels persisted until 48 h (Fig. 3A).

To investigate if miR172c up-regulation correlated with relevant events in the rhizobial infection process, we determined the expression of early nodulation genes in inoculated roots (Fig. 3, B–G). O'Rourke et al. (2014) identified common bean nodulation genes that were highly expressed in young and/or mature nodules and that are homologous to cognate nodulation genes previously identified in other legume species. From these, we selected six early nodulation genes for expression analysis: the TF genes *NF-YA1* (Phvul.001G196800.1), *NSP2* (Phvul.009G122700.1), and *NIN* (Phvul.009G115800); *CYCLOPS* (Phvul.002G128600.1), coding for a nuclear protein that interacts with CCaMK; *FLOTILLIN-LIKE2* (*FLOT2*; Phvul.009G090700.1), coding for a lipid raft component; and *ENOD40* (Phvul.008G291800), which lacks an open reading frame but encodes two small peptides and may function as a cell-cell signaling molecule for nodulation (Crespi and Frugier, 2008; Oldroyd and

111

Figure 3. Increased expression of miR172c and of early nodulation genes upon rhizobial infection. Expression levels of mature miR172c, *AP2-1* (Phvul.005G138300; A), and early nodulation genes (B–G) were determined in roots inoculated with *R. etli* CE3 at the initial time (0) and after the indicated hpi. The common bean early nodulation genes were identified in the *Pv* GEA (O'Rourke et al., 2014): *NF-YA1*, Phvul.001G196800 (B); *NSP2*, Phvul.009G122700 (C); *CYCLOPS,* Phvul.002G128600 (D); *NIN*, Phvul.009G115800 (E); *FLOT2*, Phvul.009G090700 (F); and *ENOD40,* Phvul.002G064200 (G). Values represent averages ± sᴅ from three biological replicates and two technical replicates each. Expression level refers to gene expression, based on C_t value, normalized with the expression of the housekeeping miR159 or *UBC9* gene. Different lowercase letters indicate statistically different groups (ANOVA, $P < 0.001$); in A, lowercase and uppercase letters were used for *AP2-1* and miR172c values, respectively.
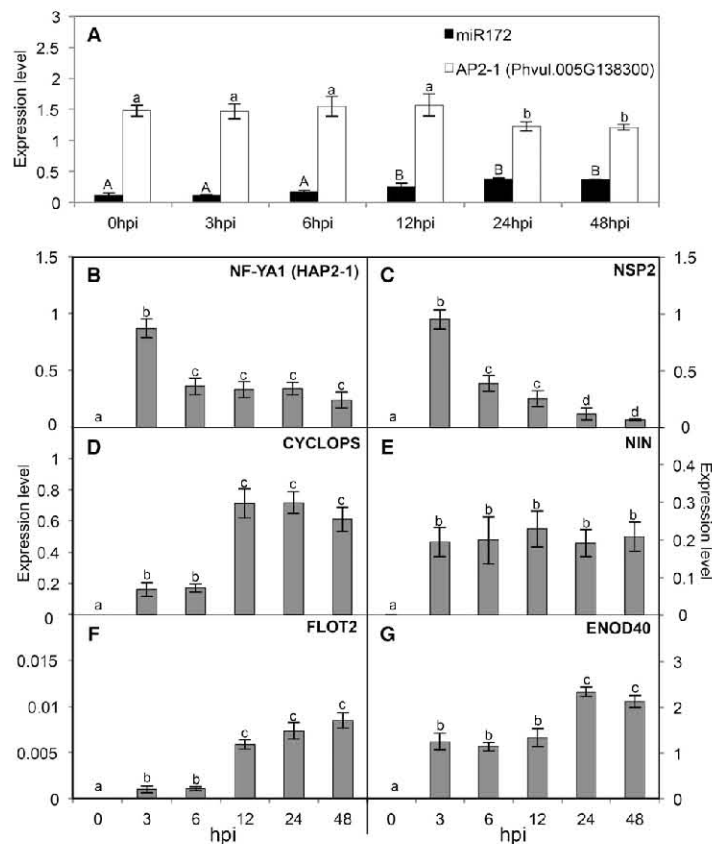
Downie, 2008; Kouchi et al., 2010; Murray, 2011; Oldroyd, 2013). Figure 3, B to G, shows the expression levels of the early nodulation genes in the responsive zone of *R. etli*-inoculated roots. The expression of all the genes tested increased significantly after the initial time (3 h). Highest levels of *NF-YA1* and *NSP2* decreased gradually after 3 h. Increased *NIN* expression persisted, whereas *CYCLOPS* and *FLOT2* transcripts increased further after 6 h and persisted until 48 h. The *ENOD40* transcript level increased significantly after 12 h and persisted until 48 h.

We then analyzed the regulation of the miR172 node during the development of effective nodules elicited by the *R. etli* CE3 wild-type strain (Fig. 4A). Nodules from inoculated common bean plants were harvested at different developmental stages, as defined by the differential expression of nodule development marker genes (Ramírez et al., 2005; Van de Velde et al., 2006; Supplemental Table S3). Immature, prefixing, 13-dpi nodules showed the highest *ENOD55* expression and low (19.6%) nitrogenase activity. At 18 dpi, the nodules were fully developed and showed the highest nitrogenase activity and expression of the *PHOSPHOENOLPYRUVATE CARBOXYLASE* (*PEPc*) gene, essential for carbon assimilation in mature nodules (Ramírez et al., 2005). By 35 dpi, nodules had low nitrogenase activity (11%) and high *CYSTEINE PROTEINASE* (*CP*) gene expression, described as being specific for nodule senescence (Van de Velde et al., 2006; Supplemental Table S3). As shown in Figure 4A, the increased expression level of miR172c observed at 2 dpi (or 48 h in Fig. 3A) persisted in immature, prefixing nodules (13 dpi). In contrast, miR172c increased significantly to its highest level in mature, fully active nodules (18 dpi). Afterward, a drastic decrease in miR172c level was observed; it remained barely detectable until nodule senescence (35 dpi). Slightly decreased levels of *AP2-1* transcripts persisted in immature nodules (13 dpi), a further decrease was observed in mature 18-dpi nodules, and afterward, the level of *AP2-1* transcripts gradually increased. The lowest level of *AP2-1* correlated with the highest level of miR172c in mature nodules (18 dpi).
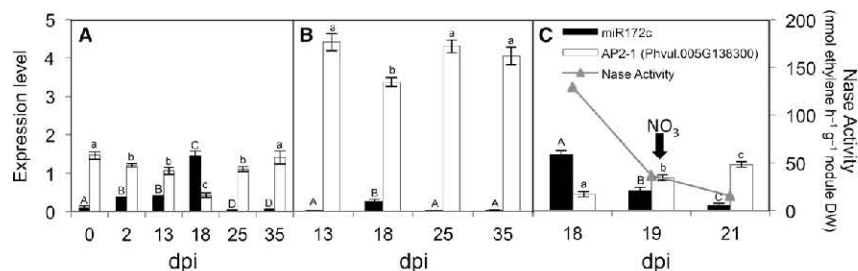
Figure 4. miR172c and *AP2-1* are differentially regulated in effective versus ineffective *R. etli* symbioses. Transcript levels were determined by qRT-PCR in inoculated roots or nodules harvested at the indicated dpi. A, Plants inoculated with the CE3 wild-type strain; determinations at 0 and 2 dpi were done in inoculated roots. B, Plants inoculated with the *fix⁻ R. etli nifA⁻* mutant. C, Plants inoculated with the CE3 wild-type strain were grown for 18 d and watered with nitrogen-free nutrient solution. Subsequently, these plants were watered with nutrient solution supplemented with 10 mM KNO₃ (black arrow). Nitrogenase (Nase) activity and transcript levels were determined at 18 dpi and at 1 or 3 d after nitrate addition (19 or 21 dpi). Values represent averages ± SD from three biological replicates and two technical replicates each. Expression level refers to gene expression, based on C_t value, normalized with the expression of the housekeeping miR159 or *UBC9* gene. Different letters indicate statistically different groups (ANOVA, *P* < 0.001); lowercase and uppercase letters were used for *AP2-1* and miR172c values, respectively.

Taken together, these data suggest that miR172c is involved in rhizobial infection and nodule development/function.

Altered Expression of miR172c and *AP2-1* in Ineffective Symbioses

Because the maximum level of miR172c expression and *AP2-1* silencing correlated with the peak of SNF (Fig. 4A; Supplemental Table S2), we assessed the regulation of the miR172 node in ineffective, nonfixing common bean-*R. etli* symbioses.

The nitrogen fixation genes regulator A (NifA)/RNA polymerase sigma factor complex is a master regulator of the N₂ fixation genes in rhizobia. Transcriptional analysis of the *R. etli nifA⁻* (CFNX247) mutant strain demonstrated the *nifA* dependency of symbiotic genes on the symbiotic plasmid (Girard et al., 1996). The symbiotic phenotype of common bean plants inoculated with the *R. etli nifA⁻* mutant strain was drastically altered, as evidenced by a diminished amount of early-senescent nodules with few infected cells having bacteroids and devoid of nitrogenase activity and with symptoms characteristic of nitrogen deprivation in the leaves (Supplemental Fig. S2; Supplemental Table S3). As shown in Figure 4B, the ineffective nodules elicited by *R. etli nifA⁻* had nearly undetectable levels of miR172c, although a minor, but significant, increase was observed in 18-dpi ineffective nodules. Meanwhile, the *AP2-1* target gene was highly induced at the different developmental stages of ineffective nodules; these values were even higher (approximately 2-fold) than those observed during effective symbiosis (Fig. 4, A and B). A slight but significant decrease in *AP2-1* transcript was observed in 18-dpi ineffective nodules, when the miR172c level increased (Fig. 4B).

A similar effect was observed when the abolishment of SNF was achieved by adding nitrate to effective

R. etli-elicited nodules (Fig. 4C), a well-known phenomenon in the legume-rhizobia symbiosis (Streeter and Wong, 1988). A short time (1 and 3 d) after nitrate addition, nitrogenase activity decreased drastically and nodules senesced (Fig. 4C; Supplemental Table S3). The latter correlated with the drastic decrease in mature miR172c and a concomitant increase of *AP2-1* transcript level in the ineffective nodules (Fig. 4C).

Taken together, these data indicate a contrasting regulation of miR172c/*AP2-1* expression in effective versus ineffective symbioses.

Effect of miR172c Overexpression on Root Development and Rhizobial SNF

To further investigate the role of miR172c and its target gene *AP2-1* in SNF, we aimed to modulate their expression in common bean composite plants with transgenic roots and untransformed aerial organs, generated through *Agrobacterium rhizogenes*-mediated genetic transformation. This protocol has been used as an alternative method for stable transformation in common bean and other recalcitrant species (Estrada-Navarrete et al., 2007). The construct for miR172 overexpression (OE172) contained the 35S cauliflower mosaic virus promoter fused to the miR172c precursor. The OEAP2m plasmid contained a mutagenized *AP2-1* gene that is insensitive to miR172 cleavage due to nucleotide substitutions in the miR172 binding site. Both constructs as well as the control empty vector (EV) contain the tdTomato (red fluorescent protein) reporter gene (Supplemental Fig. S3). We obtained several composite plants and determined the level of transgene expression for each plant (Supplemental Fig. S4). The OE172 composite plants showed very high levels of mature miR172c in both nodules and roots as well as a decreased level of *AP2-1* transcript. Roots and nodules of OEAP2m plants showed very high levels of *AP2-1*. The

variation in the degree of overexpression between individual transgenic roots is because each results from an independent transformation event.

We first assessed if miR172 overexpression affected the root phenotype of fertilized (noninoculated) common bean plants as compared with those inoculated with *R. etli*. As shown in Figure 5, roots with high miR172c showed increased biomass and density of secondary roots, both in fertilized and SNF composite plants. The opposite phenotype was observed in OEAP2m composite plants. These data indicate that miR172 had a positive effect on root biomass/architecture independent of the presence of rhizobia.

To analyze if the positive effect of miR172 on root development (Fig. 5) also affects rhizobial infection and SNF, we investigated the response of composite plants altered in miR172 content to *R. etli* infection, early symbiotic stages, and nodule development/function.

Figure 6 shows data for the analysis of rhizobial infection and early nodulation gene expression. For these experiments, the plastic square bioassay dish system was used for the inoculation and growth of OE172, EV, or OEAP2m composite plants. Notably, the amount of deformed root hairs was significantly higher in 48-hpi inoculated roots that overexpress miR172, while the opposite effect was observed in OEAP2m roots (Fig. 6A; Supplemental Fig. S5). A correlation of altered root hair deformation and the expression of early nodulation genes essential for rhizobial infection was observed after determining the transcript level of selected genes (*NF-YA1*, *NSP2*, *CYCLOPS*, *ENOD40*, *FLOT2*, and *NIN*) in the responsive root zone from 0 to 48 hpi (Fig. 6, B–G). All the genes tested showed increased expression in OE172 inoculated roots; *NIN* expression was increased only as compared with OEAP2m inoculated roots (Fig. 6, B–G).

Nodule number and nitrogenase activity as well as histological analysis of nodules stained with SYTO13 (a fluorescent dye binding nucleic acids) from composite plants overexpressing miR172 or *AP2m* and control (EV) plants are presented in Figures 7 and 8. At 14 and 21 dpi, OE172 plants showed increased nodule number and nitrogenase activity that correlated with higher *PEPc* and reduced *CP* expression (Fig. 7; Supplemental Table S4). In addition, OE172 plants showed accelerated nodule development: nodule primordia and well-formed nodules (approximately 50 per root) were easily observed at 5 and 7 dpi, respectively, while only some unorganized primordia and very few tiny nodules were seen in EV plants. In addition, young (7 and 14 dpi) nodules from OE172 plants were larger (increased perimeter; Fig. 8, A and B). In contrast, OEAP2m plants at 21 dpi had fewer nodules with diminished nitrogenase activity and higher *CP* expression (Fig. 7; Supplemental Table S4). At all time points analyzed, the OEAP2m nodules had significantly reduced perimeters (Fig. 8, A and B). Similar values for SYTO13 intensity per nodule area were obtained for EV, OE172, and OEAP2m nodules, indicating similar bacteroid densities in infected cells (Fig. 8C).
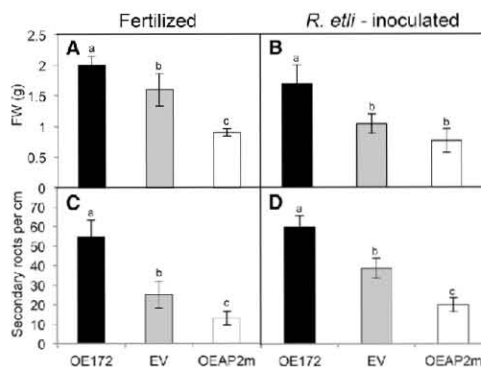
Figure 5. miR172c and *AP2-1* control the root development of fertilized and *R. etli*-inoculated common bean plants. Root fresh weight (FW; A and B) and density of secondary roots (C and D) were determined in composite bean plants with transgenic roots overexpressing miR172c (OE172) or a mutated insensitive *AP2-1* target gene (OEAP2m) as compared with control EV transformed roots. A set of composite plants was fertilized with full-nutrient solution for 10 d (A and C), and another set was inoculated with *R. etli* and watered with nitrogen-free nutrient solution for 21 dpi (B and D). Values represent averages ± SD from roots of eight independent composite plants each. Different lowercase letters indicate statistically different groups (ANOVA, *P* < 0.001).

We assessed if increased nodulation in OE172 common bean plants could be related to alterations in the AON. In soybean, AON involves long-distance signaling requiring the interaction of RHIZOBIA-INDUCED CLE peptides (RIC1/RIC2), with NODULE AUTOREGULATION RECEPTOR KINASE (NARK) in the leaf and the subsequent inhibition of nodulation via the production of a shoot-derived inhibitor. For local nitrate inhibition, the nitrate-induced CLE peptide (NIC1) interacts with NARK in the root, leading to a nitrate-induced inhibitor (Reid et al., 2011a). The homologous common bean *RIC* and *NIC* genes *RIC1* (Phvul.005G096900), *RIC2* (Phvul.011G135900), and *NIC1* (Phvul.005G097000) were identified from the *Pv* GEA (O'Rourke et al., 2014). As in soybean (Reid et al., 2011a), the common bean *RIC1* genes were expressed in inoculated common roots at early stages of rhizobial infection, while *RIC2* was expressed at later time points in prefixing and mature nodules. Figure 6, H and I, shows the expression levels of *RIC1* and *NIC1* genes in 48-hpi inoculated roots from OE172, EV, and OEAP2m plants. *RIC1* expression in control (EV) transgenic roots indicates the rhizobial induction of CLE-derived peptides for AON. Interestingly, the level of *RIC1* was decreased significantly in OE172 inoculated roots that showed increased nodulation, while it was increased significantly in OEAP2m roots with diminished nodulation (Fig. 6H). As expected, the expression of *NIC1* was low in the transgenic inoculated roots (Fig. 6I) under nitrogen-free conditions. These data point to the involvement of a common bean AON mechanism in the miR172 control of nodulation.
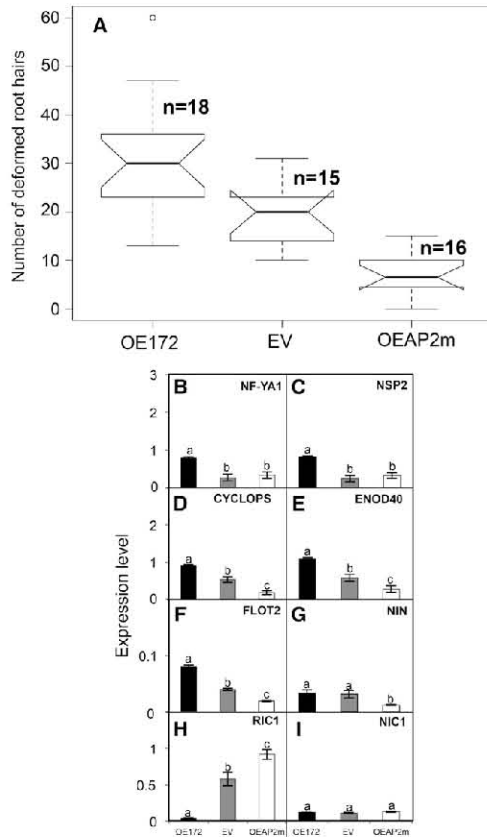
Figure 6. miR172c and *AP2-1* control the rhizobial infection of common bean roots. Roots from OE172, EV, or OEAP2m composite plants were inoculated with *R. etli* CE3, and at 48 hpi, the root responsive zones were harvested for analysis. A, Quantification of the number of deformed root hairs (branched and swollen root hair) per 0.5 cm; each box plot indicates the number of the transgenic roots analyzed for each construct. B to I, Expression analysis of selected early nodulation genes. Most gene identifiers are indicated in the legend to Figure 3; *RIC1*, Phvul.005G096900; and *NIC1*, Phvul005G097000. Values represent averages ± SD from three biological replicates and two technical replicates each. Expression level refers to gene expression, based on C_t value, normalized with the expression of the housekeeping *UBC9* gene. Different lowercase letters indicate statistically different groups (ANOVA, $P < 0.001$).

miR172c Overexpression Decreases the Sensitivity to Nitrate Inhibition of Rhizobial Symbiosis

Nitrogen (nitrate or ammonia) in the soil perceived by legume plants is an important external stimulus that inhibits nodulation as part of the AON mechanism. Considering the improved rhizobial infection and nodule development/function in common bean plants with

increased miR172c (Figs. 6–8), we assessed if these alterations could be related to a decreased sensitivity to external nitrate inhibition of *R. etli* nodulation (Fig. 9). For this experiment, we applied a low nitrate concentration (1 mM) to *R. etli*-inoculated OE172, EV, and OEAP2m composite plants. The nodulation of inoculated OEAP2m plants, with low miR172c, was completely abolished when low nitrate was added. For this reason, we analyzed plants overexpressing miR172c as compared with control (EV) plants (Fig. 9). As expected for nitrate inhibition of the rhizobial infection process, the expression level of most early nodulation genes was reduced in EV-inoculated plants in the presence of nitrate (Fig. 9, A–F) as compared with the nitrogen-free condition (Fig. 6, B–G). Notably, in the presence of nitrate, the early nodulation genes *NF-YA1*, *NSP2*, *CYCLOPS*, *ENOD40*, and *FLOT2* showed significantly increased expression in OE172 inoculated roots (Fig. 9, A–F). In fact, the expression level of early nodulation genes in OE172 inoculated roots was similar in the absence (Fig. 6, B–G) or presence (Fig. 9, A–F) of nitrate. As expected, while the *RIC1* gene was barely detectable, *NIC1* was expressed in the responsive root zone of EV plants inoculated in the presence of nitrate, and it was increased in OE172 roots (Fig. 9, G–H).

Nitrate inhibition of nodulation was evident in control (EV) plants that presented delayed and diminished nodulation (Fig. 9I). These plants also showed decreased nitrogenase activity and *PEPc* expression and increased *ENOD55* (Fig. 9J; Supplemental Table S4) as compared with nitrogen-free inoculated plants (Fig. 7B). Notably, inoculated OE172 plants in the presence of nitrate showed few active nodules at 14 dpi, while at 21 dpi they had a similar number of mature nodules with slightly higher nitrogenase activity as compared with plants inoculated without nitrogen (Figs. 7 and 9, I and J; Supplemental Table S4). In addition, we observed that a higher nitrate concentration (3 mM) totally blocked the nodulation of EV plants, while OE172 plants were able to form active nodules (approximately 100 per root) at 21 dpi.
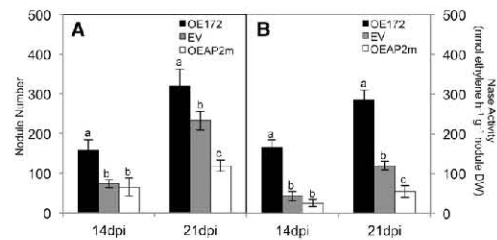
Figure 7. miR172c and *AP2-1* control nodule number (A) and nitrogenase (Nase) activity (B) of SNF common bean. Plants were inoculated with *R. etli* CE3 for the indicated dpi. Values represent averages ± SD from five replicate samples per time point. Different lowercase letters from each set of values at different dpi indicate statistically different groups (ANOVA, $P < 0.001$).

Figure 8. Alterations in the nodule development of OE172 and OEAP2m *R. etli*-inoculated composite bean plants. A, Fluorescent (F; left) and corresponding dark-field (DF; right) micrographs of central sections of nodules harvested at the indicated dpi. Red fluorescence from the tdTomato reporter gene expressed in *A. rhizogenes* transformed roots and green fluorescence from SYTO13 staining were observed. Magnification = 5×. B and C, Nodule perimeter (B) and SYTO13 fluorescence intensity per infection area (C) were calculated using the ImageJ program. Values represent averages ± SD from 10 replicate nodule images per condition. Different lowercase letters in B indicate statistically different groups (ANOVA, *P* < 0.001); values from C were not statistically different.
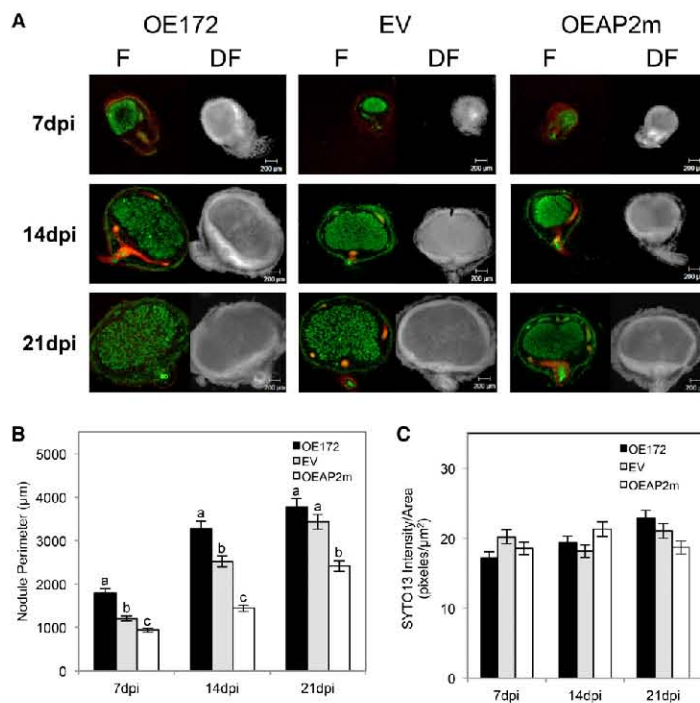

Exploring the Downstream *AP2-1* Regulation in SNF Plants

TFs from the *AP2* superfamily are widespread in plants and control diverse developmental programs and stress responses. Different *AP2* family members have been classified as activators or repressors of specific target genes (Licausi et al., 2013). In this work, we aimed to predict target genes for AP2-1 transcriptional activation or repression by analyzing data from the root and nodule libraries reported in the *Pv* GEA (O'Rourke et al., 2014), especially those from young roots and mature effective nodules that were derived from common bean tissues similar to those analyzed in this work. However, a caveat of this analysis is that the *Pv* GEA does not include libraries from roots inoculated with rhizobia for short periods, so we could not predict AP2-1 targets that would be regulated during the rhizobial infection process.

As shown in Figures 2 and 4, *AP2-1* showed high expression in common bean roots as opposed to mature nodules. We hypothesized that genes with an expression pattern similar to *AP2-1* are likely to be involved in root function/development and to be positively regulated by *AP2-1*, thus providing information on a possible mechanism of action of miR172/*AP2-1*. We identified 114 genes that had an expression pattern similar to *AP2-1*, designated *AP2-1* coexpressed genes (Supplemental Fig. S6). In

order to support the latter contention, we searched for TFBS in the 5′ promoter region of *AP2-1* coexpressed genes. Besides WRKY, the most statistically overrepresented (*P* = 0 and 0.001) TFBS were ETHYLENE-RESPONSIVE FACTOR2 (ERF2) and DEHYDRATION-RESPONSIVE ELEMENT BINDING 1B (DREB1B), which belong to the AP2 superfamily. These TFBS were identified in 82% of *AP2-1* coexpressed genes (Supplemental Table S5).

Sixty-seven of the 114 *AP2-1* coexpressed genes could be assigned to a Gene Ontology (GO) category (Table I; Supplemental Table S5). The most statistically significant (*P* = 0.006) assigned GO category, which included 19 coexpressed genes, is GO:0004672, associated with protein kinase activity (Table I). We validated by qRT-PCR the expression of protein kinase activity genes in roots and nodules of control (EV) plants and also of plants overexpressing *AP2-1*. We hypothesized that those genes positively targeted by *AP2-1*, with high expression in roots and low expression in nodules from wild-type or control EV plants (Supplemental Fig. S6), would show a different expression pattern, higher and/or similar in roots and nodules, in OEAP2m plants that have constitutively enhanced expression of the *AP2-1* transcriptional regulator. Table II shows six *AP2-1* coexpressed genes assigned to the protein kinase activity category (GO:0004672) whose expression levels agree with our hypothesis. In control

116
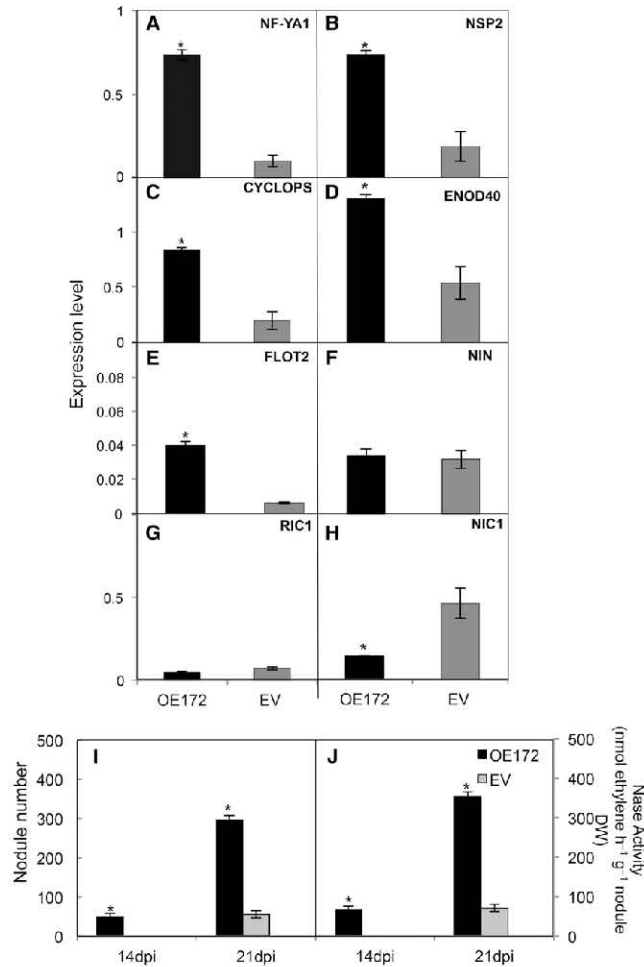
Figure 9. miR172c controls the sensitivity to nitrate inhibition of rhizobial symbiosis. To analyze the effect of nitrate in the symbiosis of *R. etli*-inoculated OE172 or EV composite plants, 1 mM KNO_3 was added to the nutrient solution used to water each set of plants daily. A to H, Expression analysis of selected early nodulation genes determined at 48 hpi in the root responsive zone. Gene identifiers are indicated in the legends to Figures 3 and 6. Values represent averages ± SD from three biological replicates and two technical replicates each. Expression level refers to gene expression, based on C_t value, normalized with the expression of the housekeeping *UBC9* gene. I and J, Nodules were counted (I) and nitrogenase (Nase) activity (J) was determined at the indicated dpi. Values represent averages ± SD from five replicate samples per time point. Asterisks indicate the level of statistically significant difference, if any, among values from OE172 and EV roots (Student's *t* test, *P* < 0.01).

(EV) plants, these genes were highly expressed in roots as compared with nodules, similar to *AP2-1*, thus validating the *Pv* GEA data (O'Rourke et al., 2014; Supplemental Fig. S6), while they showed an increased and/or similar expression in both tissues from OEAP2m plants (Table II). Interestingly, these genes shared the DREB, ERF, or both TFBS in their promoter regions (Table II; Supplemental Table S5). Transcriptomic analyses of *M. truncatula* nodule senescence have shown that protein kinases genes are one of the highly induced gene classes at different stages of this process (Van de Velde et al., 2006; Pérez Guerra et al., 2010). On this basis, we asked if the some of the *AP2-1* coexpressed genes from the protein kinase activity GO category might be related to nodule senescence in common bean. As shown in Table II,

four of the protein kinase activity genes analyzed (Phvul.002G326600, Phvul.007G049500, Phvul.007G049000, and Phvul.006G174500) were induced in senescent (35 dpi) as compared with mature (18 dpi) nodules, thus indicating their possible relation to common bean nodule senescence.

The root/nodule expression pattern of *AP2-1* was opposite that of miR172c, which showed highest expression in mature effective nodules and very low expression in young roots (Figs. 2 and 4). O'Rourke et al. (2014) have described a set of 402 common bean genes highly expressed in mature effective nodules as compared with all other tissues analyzed in the *Pv* GEA. These genes are likely involved in the establishment of symbiosis and SNF as supported by their assigned GO categories (O'Rourke et al., 2014). Here, we hypothesized

Table I. *GO categories statistically overrepresented for genes coexpressed with AP2-1 in roots*

GO Identifier	Description (Molecular Function)	P
GO:0004672	Protein kinase activity	0.006
GO:0008703	5-Amino-6-(5-phosphoribosylamino)uracil reductase activity	0.016
GO:0005516	Calmodulin binding	0.018
GO:0005471	ATP:ADP antiporter activity	0.024
GO:0004674	Protein Ser/Thr kinase activity	0.036
GO:0004351	Glu decarboxylase activity	0.055
GO:0004435	Phosphatidylinositol phospholipase C activity	0.055

that these nodule-enhanced genes, lowly expressed in young roots with increased *AP2-1* (Figs. 2 and 4), are candidates for AP2-1 negative regulation. To this end, we searched for TFBS in the 5′ promoter region of nodule-enhanced genes, but we did not find a statistical overrepresentation of AP2 TFBS.

DISCUSSION

The key role of the miR172 node in Arabidopsis flowering time and phase transition is well known; similar roles have also been documented in maize (*Zea mays*), rice (*Oryza sativa*), and barley (*Hordeum vulgare*; Zhu and Helliwell, 2011). Besides conserved roles, specialized/particular species-specific functions of miR172, such as the induction of tuberization in potato (*Solanum tuberosum*), have been reported (Martin et al., 2009). In legumes, conserved roles of the miR172 node have been documented for *L. japonicus* (control of flowering time; Yamashino et al., 2013) and soybean (control of juvenile-to-adult phase transition; Yoshikawa et al., 2013). In addition, the control of nodulation during the rhizobia symbiosis has been proposed as a family-specific acquired function of miR172 in

different legumes and has been demonstrated for soybean (Yan et al., 2013; Wang et al., 2014). In this work, we identified the miR172 node as a relevant regulator of rhizobial infection and nodulation in common bean.

We propose that different miR172 isoforms regulate different processes: miR172b is involved in flowering, while miR172c mainly regulates nodulation. Our data indicate that these miR172 isoforms exert their effects by silencing different target genes from the AP2 TF superfamily. Transcripts from two *AP2* genes (Phvul.005G138300 and Phvul.011G071100) are likely to function in roots and are cleaved by miR172c in nodules. Three other *AP2* genes (Phvul.003G241900, Phvul.002G16900, and Phvul.001G174400) are likely to function in young flowers, which showed a high level of these transcripts as well as of miR172, thus suggesting that, in flowers, the *AP2* target genes are silenced by miR172-induced translation inhibition, similar to Arabidopsis (Chen, 2004). Our work focused on the analysis of the miR172c/*AP2-1* (Phvul.005G138300) node in the common bean-rhizobia nitrogen-fixing symbiosis, and our proposed regulatory model is summarized in Figure 10.

In Arabidopsis, miR156 represses miR172 expression by targeting *SPL* TFs that directly bind to the *MIR172*

Table II. *Selected genes coexpressed with AP2-1 from the statistically overrepresented GO:0004672 category: protein kinase activity*

Expression level was determined by qRT-PCR from 21-dpi mature nodules (N) and roots (R) of EV and OEAP2m *R. etli*-inoculated composite plants and from 18 dpi mature or 35 dpi senescent nodules (N) from *R. etli*-inoculated wild-type plants. Values represent averages ± SD from three independent biological replicates and two technical replicates. TFBS for ERF and DREB (subfamilies of the AP2 TF family) were identified as statistically overrepresented in the promoter sequence of each gene as indicated.

Gene Identifier[a]	Annotation[a]	TFBS	21 dpi				Wild Type	
			EV		OEAP2m		18 dpi	35 dpi
			R	N	R	N	N	N
Phvul.002G326600	Aminocyclopropane carboxylate oxidase	DREB	0.76 ± 0.08	0.35 ± 0.13	1.2 ± 0.06	1.03 ± 0.07	0.11 ± 0.01	1.03 ± 0.07
Phvul.007G049500	Ser/Thr protein kinase	ERF	0.61 ± 0.09	0.25 ± 0.03	0.93 ± 0.08	0.97 ± 0.09	0.10 ± 0.02	0.53 ± 0.05
Phvul.007G049000	Ser/Thr protein kinase	DREB	0.10 ± 0.007	0.043 ± 0.01	0.02 ± 0.003	0.03 ± 0.004	0.07 ± 0.009	0.21 ± 0.03
Phvul.006G174500	Glycogen synthase kinase-3α	DREB	0.45 ± 0.06	0.24 ± 0.03	0.71 ± 0.07	0.57 ± 0.06	0.16 ± 0.02	0.37 ± 0.03
Phvul.011G169600	Ser/Thr protein kinase	DREB	0.35 ± 0.02	0.18 ± 0.02	0.52 ± 0.03	0.47 ± 0.03	0.33 ± 0.03	0.36 ± 0.02
Phvul.008G263900	Ser/Thr protein kinase	ERF and DREB	1.2 ± 0.06	0.15 ± 0.04	0.57 ± 0.02	0.50 ± 0.09	0.0023 ± 0.0003	0.0019 ± 0.0002

[a]From http://www.phytozome.net/commonbean.php.

promoter and positively regulate its expression (Wu et al., 2009). Transgenic soybean roots overexpressing miR156 showed a reduction in nodulation, decreased miR172 level, and decreased expression of two *SPL* genes proposed as miR156 targets, although evidence for the binding of SPL to *MIR172* promoters for transcription activation was not provided (Yan et al., 2013). Recently, Wang et al. (2015) reported that the overexpression of miR156 affects several aspects of plant architecture in *L. japonicus*, including underdeveloped roots and reduced nodulation, which correlate with the repression of several early symbiotic genes. However, the authors did not analyze a possible regulation of miR172 by miR156, which may be related to the miR156 effects in nodulation that they showed (Wang et al., 2015). In common bean, we observed opposite levels of mature miR156a as compared with miR172 and also a negative correlation between miR156a and its validated target *SPL6* gene (Phvul.009G165100). However, we could not identify SPL TFBS in any of the promoter regions of the six *MIR172* loci from the common bean genome, so it is difficult to propose SPL as a direct transcriptional regulator of miR172. However, the binding of SPL proteins to yet unknown sequence motifs present in *MIR172* promoters cannot be ruled out. Alternatively, miR156a may exert its negative regulation over common bean miR172 through other target genes not yet identified. For example, transcripts coding for tryptophan-aspartic acid repeats protein domain proteins, which may be involved in microtubule organization, protein-protein and protein-DNA interactions, or chromatin conformation, have been validated as miR156 targets in *M. truncatula* and *L. japonicus*, but their specific regulatory function has not been analyzed (Naya et al., 2010; Wang et al., 2015).

In this work, we showed that miR172c has a positive effect on root development independent of rhizobium infection. In addition, miR172c is relevant for the control of rhizobial infection. This miRNA increased after 6 h in *R. etli*-inoculated roots, when infection threads are formed, and this is related to the increase in root hair deformation observed in plants that overexpress miR172c. Preliminary data indicate that the roots overexpressing AP2m induce irregular infection threads (B. Nova-Franco, O. Valdés-López, and G. Hernández, unpublished data). Together, these data indicate a regulatory/signaling role of miR172c in the rhizobial infection of common bean (Fig. 10). Up-regulation of miR172c was concomitant with that of early nodulation genes, mainly expressed in the cortical cells, that are involved in infection thread initiation/progression (i.e. *FLOT2* and *ENOD40*) and act downstream of *NSP2*, *NIN*, and *NF-YA1* (Murray, 2011; Oldroyd, 2013), whose expression was highest after 3 h of rhizobial inoculation of common bean roots. Therefore, we propose that miR172c-mediated control of rhizobial infection is exerted at the level of cortical cell division downstream of NF perception, Ca^{2+} spiking, CCaMK, NSP2, and NIN (Fig. 10). In addition, our data on the repression of *RIC1* in roots overexpressing miR172c indicate the involvement of this miRNA in the AON at early stages of the common

bean symbiosis. Soyano et al. (2014) reported that the AON *L. japonicus CLE* root signal genes *CLE-RS1* and *CLE-RS2*, which are orthologous to soybean *RIC1* and *RIC2*, are directly transcribed by NIN, the essential inducer for nodule primordium formation. This constitutes a complex regulatory circuit with a long-distance feedback loop involved in the homeostatic regulation of nodule organ production in *L. japonicus* (Soyano et al., 2014). In soybean, Wang et al. (2014a) recently reported that NARK negatively regulates miR172 transcription during nodule primordium formation to prevent excess nodulation. In this work, we showed that both *NIN* and *RIC1* are expressed at early stages of rhizobial infection in common bean roots and that OEAP2 roots with decreased nodulation showed increased *RIC1* levels. Taken together, these data would indicate a positive regulation of *NIN* and *AP2-1* to *RIC1*, thus leading to reduced nodulation through AON in common bean (Fig. 10). Sequence analysis of the *RIC1* promoter region led us to identify NIN- and DREB/ERF-enriched regions, something that supports the latter contention. However, further work is required to fully demonstrate which TFs activate *RIC1* expression in common bean.

The regulation of nodulation through AON signaling is also relevant for the inhibition of nodulation that occurs when nitrate is present in the rhizosphere. For local nitrate inhibition, the nitrate-induced CLE peptide in soybean (NIC) interacts with NARK in the root, leading to a nitrate-induced inhibition of nodulation in soybean (Reid et al., 2011a). Our data indicate that common bean miR172c is a signaling component of the nitrate-induced AON (Fig. 10). In the presence of nitrate, rhizobia-inoculated roots that overexpress miR172c developed more active nodules and showed very low expression of *NIC1* that correlates with the up-regulation of *NF-YA1*, *NSP2*, *CYCLOPS*, *ENOD40*, and *FLOT2* early symbiotic genes. The expression of *NIN* was similar in EV and OE172 roots inoculated in the absence or presence of nitrate, which is in agreement with data from *L. japonicus* reported by Soyano et al. (2015). The legume-rhizobia symbiosis with increased resistance to soil nitrate is relevant for improving plant growth and crop production. Our better understanding of the elements involved in the control of this phenomenon, such as miR172c in common bean, opens the possibility to exploit it for the future improvement of symbiosis.

The effect of miR172c in rhizobial infection and nodulation of common bean is likely to be directly exerted by its target gene, the AP2-1 transcriptional regulator. TFs from the AP2 superfamily may function as repressors or as activators of transcription (Licausi et al., 2013). Work recently published by Yan et al. (2013) and Wang et al. (2014) about the mechanism of action of AP2 in soybean nodulation points to the repressor role of AP2. In this work, we explored the possible role of *AP2-1* as a transcriptional activator and/or repressor of genes relevant for common bean rhizobium infection and nodulation. Recently, Soyano et al. (2015) reported that in *L. japonicus*, the NIN TF could repress or activate transcription in different scenarios of rhizobial infection in the presence or
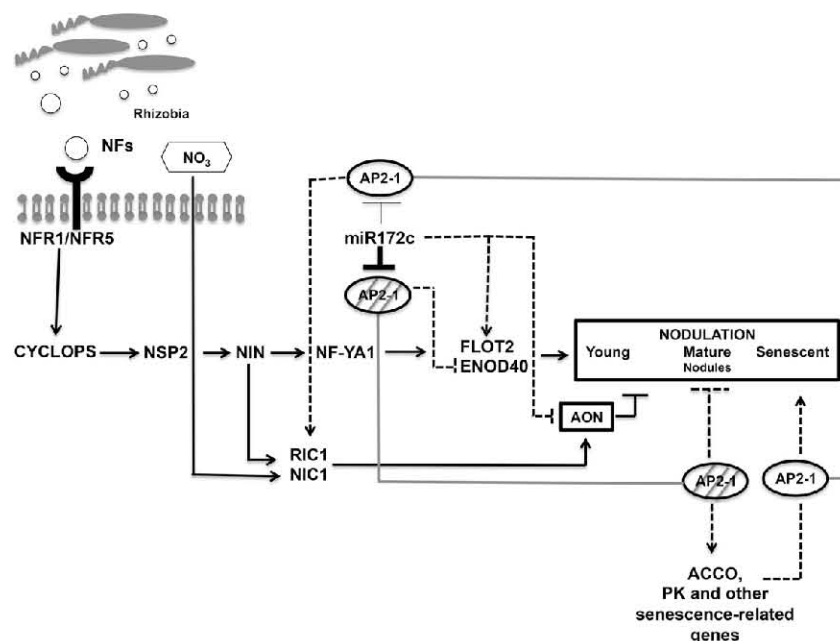
Figure 10. Model of miR172 node regulation in common bean-rhizobia symbiosis. Positive regulation is represented with arrows and negative regulation with lines. The root signaling cascade, triggered by rhizobial NF, is essential for rhizobia infection and nodule development in different legumes. Regulation of common bean rhizobial infection and nodulation by miR172c and AP2-1 are represented by dashed arrows or lines. A high level of miR172c (thick line) induces AP2-1 degradation (hatched circle), while active AP2-1 (circle) is present when the miR172c level is very low (thin line). miR172c positively regulates early nodulation gene expression and rhizobial infection, silencing AP2-1 that may repress *ENOD40* expression. Nodule number is positively regulated by miR172c; AON decreases through low RIC1/NIC1 expression, perhaps regulated by AP2-1. In mature nodules, abundant miR172c silences AP2-1, an activator of senescence-related genes that are further required during nodule senescence when AP2-1 levels are recovered. ACCO, Aminocyclopropane carboxylase oxidase.

absence of nitrate. They postulated that such dual regulatory functions might depend on specific coactivator or corepressor molecules that would interact with the same NIN TF in different scenarios.

We identified genes that significantly coexpress with *AP2-1* and are candidates for transcriptional activation by this TF (Fig. 10). Several of these genes were assigned to the protein kinase activity GO category. The best candidates are four protein kinases that are repressed in mature nodules and induced in roots and in senescent nodules of control plants, whereas they show high and/or constitutive expression in roots and nodules of AP2m overexpressing plants. This group includes the Phvul.007G049500 gene annotated (www.phytozome.net/commonbean.php) as a Ser/Thr protein kinase with a Cys-rich receptor-like protein kinase, Domain of Unknown Function26 (transmembrane), and Ser/Thr protein kinase domains. We found that this common bean kinase gene is similar (48%) to the Arabidopsis *CYSTEINE-RICH RECEPTOR-LIKE PROTEIN*

KINASE29 (*CRK29*) gene and to *M. truncatula SymCRK* (52% similarity) and has domains characteristic of the Cys-rich kinase family. Several members of this family are induced during *M. truncatula* nodule senescence (Van de Velde et al., 2006; Pérez Guerra et al., 2010). Specifically, SymCRK is involved in senescence and defense-like reactions during the *M. truncatula-Sinorhizobium meliloti* symbiosis (Berrabah et al., 2014). Another gene from this group encodes the aminocyclopropane carboxylase oxidase, the ethylene-forming enzyme. This and other genes encoding enzymes from the ethylene biosynthetic pathway are up-regulated during *M. truncatula* nodule senescence (Van de Velde et al., 2006). Ethylene plays a positive role in nodule senescence and also a significant inhibitory role in rhizobial infection and nodule formation (Van de Velde et al., 2006; Murray, 2011). Our interpretation of these results is that *AP2-1* transcriptional regulation is important in common bean roots but that this TF needs to be silenced for an adequate nodule function (SNF), something that is achieved

120

by posttranscriptional target cleavage mediated by miR172c. *AP2-1* silencing in effective mature nodules would maintain functionality by avoiding senescence through the down-regulation of nodule senescence genes activated by this TF (Fig. 10). Ineffective nodules where miR172c is not induced and AP2-1 remains elevated, as well as OEAP2 nodules, showed early senescence. Alternatively, other protein kinases, proposed as *AP2-1* targets, may participate in signaling pathways important for root development or nodule senescence. Nodule-specific protein kinases essential for signaling pathways during initial stages of nodulation are known (Oldroyd and Downie, 2008; Kouchi et al., 2010; Murray, 2011; Oldroyd, 2013).

Regarding AP2 transcriptional repression, Wang et al. (2014) recently reported that soybean NNC1, a miR172 target, represses *ENOD40* expression, which results in negative regulation of early stages of rhizobial symbiosis. In this work, we showed that the expression of common bean *ENOD40* is decreased in rhizobia-inoculated roots that overexpress AP2m. In addition, we identified ERF- and DREB-enriched regions in the *ENOD40* 5′ promoter region. So it is conceivable that, as in soybean, common bean *ENOD40* expression is repressed by AP2-1 (Fig. 10). However, it is important to consider that soybean *NNC1* is not the ortholog of common bean *AP2-1*. We identified *NNC1* (glyma12g07800) as the ortholog (84% similarity) of the Phvul.011G071100 *AP2* gene, identified as a miR172 target in a common bean degradome analysis (D. Formey, L.P. Íñiguez, P. Peláez, Y.F. Li, R. Sunkar, F. Sánchez, J.L. Reyes, and G. Hernández, unpublished data) but not further analyzed in this work; while common bean *AP2-1* was identified as the ortholog (93% similarity) of the soybean *AP2* gene glyma15g04930 that was predicted but not validated as the miR172 target (Wang et al., 2014). Therefore, different miR172 target genes from the AP2 family analyzed in soybean and in common bean may have different mechanisms for transcriptional regulation.

In addition, Yan et al. (2013) postulated that the regulation of soybean nodulation by miR172 is explained by the *AP2* repression of nonsymbiotic hemoglobin (*Hb*) gene expression that is essential for regulating the level of nodulation; however, the authors did not provide evidence demonstrating AP2 binding and direct transcriptional repression of *Hb* genes. To explore if this circuit is functioning in common bean, we first identified *Hb* genes encoded by the common bean genome: five symbiotic leg-hemoglobin (*Lb*) genes having greatly increased expression in effective nodules and four nonsymbiotic *Hb* genes (O'Rourke et al., 2014; Supplemental Fig. S7). From the latter, Phvul011G048600 and Phvul.011G048700 were identified as orthologs of the nonsymbiotic *Hb-1* and *Hb-2* genes in soybean, respectively. In common bean, these *Hb* genes showed similar low expression in roots and nodules of wild-type plants and also in composite plants that overexpress *AP2-1* or that have very low AP2-1 resulting from miR172 overexpression (Supplemental Fig. S7). Therefore, our data differ from those of Yan et al. (2013) and lead us to

conclude that *AP2-1* repression of *Hb-1* genes is not relevant for common bean nodulation. Our exploration of other common bean symbiotic genes repressed by *AP2-1* included the identification of TFBS statistically overrepresented in the promoter regions of 402 common bean genes reported by O'Rourke et al. (2014) as nodule-enhanced genes. The expression pattern of these genes is similar to that of miR172c and opposite to that of *AP2-1*, which shows low expression in mature nodules and high expression in roots, suggesting these as candidates for transcriptional repression by AP2-1. However, AP2 (ERF and DREB) TFBS were not overrepresented in these genes. The latter is different from our data from *AP2-1* coexpressed genes proposed as being activated by this TF; these genes did show overrepresentation of ERF/DREB TFBS in their promoter regions. Further work is required to demonstrate the direct transcription repression, if any, of the miR172c target gene *AP2-1* in common bean.

Legume crops with increased nodulation/decreased nitrate inhibition of nodulation would be relevant for sustainable agriculture. This work sets the basis for further exploration, through genetic/genomic approaches, for common bean cultivars with improved traits resulting from increased miR172 in roots and nodules.

MATERIALS AND METHODS

Identification and Analysis of miR172 Precursor Genes, Isoforms, and Target Genes

The common bean (*Phaseolus vulgaris*) genome sequence recently published (Schmutz et al., 2014; http://www.phytozome.net/commonbean.php, v1.0) was analyzed, and six miR172 isoforms (a–f) were identified. Of these, four isoforms were described previously through RNA-seq analysis of common bean small RNAs by Peláez et al. (2012). The secondary RNA structure of each miR172 isoform was predicted using mfold software (Zuker, 2003) available at http://mfold.ma.albany.edu, and only the lowest energy structure generated for each sequence was chosen (Supplemental Fig. S1).

Since the only targets identified for miR172 from different plant species belong to the *AP2*-type TF family, we focused our analysis on identifying common bean miR172 targets within this gene family. We performed target prediction analysis for all the common bean *AP2* gene transcripts identified in the *Pv* GEA (O'Rourke, et al., 2014) using the Web server psRNATarget (http://plantgrn.noble.org/psRNATarget/; Dai and Zhao, 2011). Stringent criteria were used to predict targets; that is, an alignment spanning at least 18 bp with maximum penalty score of 3. Score calculation considered 0.5 points for each G:U wobble, one point for each non-G:U mismatch, and two points for each bulged nucleotide in either RNA strand (Jones-Rhoades and Bartel., 2004). In addition, we constructed a phylogenetic cladogram from the amino acid sequences of common bean *AP2* genes; these were aligned using ClustalX version 2.1 (Larkin et al., 2007). The sequence-aligned file was used to construct the bootstrapped neighbor-joining tree using the NJ clustering algorithm and Phylip output format (.phb). The reliability of the phylogenetic analysis was estimated from 1,000 bootstrap resamplings, and the tree was viewed using the program MEGA version 5.2.1 (Tamura et al., 2011).

Plant Material and Growth Conditions

Common bean seeds from the Mesoamerican cv Negro Jamapa 81 were surface sterilized and germinated for 2 d at 26°C to 28°C in darkness. Plants were grown in a hydroponic system under controlled environmental conditions as described previously (Valdés-López et al., 2010). The hydroponic trays contained 8 L of Franco and Munns (1982) nutrient solution. The volume and pH (6.5) of the trays were controlled throughout the experiment. For SNF conditions, plantlets adapted by growth for 7 d in the hydroponic system with

nitrogen-free nutrient solution were inoculated with 200 mL of a saturated liquid culture of the *Rhizobium etli* CE3 wild-type strain or the *R. etli fix⁻* mutant strain CFNX247 (Δ*nifA*::ΩSp/Sm; Girard et al., 1996). Plants were harvested at different times (dpi) for analysis; tissues for RNA isolation were collected directly into liquid nitrogen and stored at −80°C.

To analyze the initial events of rhizobial infection, 2-d-old seedlings were placed in plastic square bioassay dishes (24 × 24 cm; Corning) with solidified nitrogen-free Fåhreus medium (Vincent, 1970). Plates containing common bean seedlings were incubated in a growth chamber at 25°C with a 16-h photoperiod. After 2 d, seedlings were inoculated by applying 1 mL of *R. etli* CE3 saturated liquid culture directly to the root and were further incubated at various times (3–48 h). After specific incubation times, the root responsive zones were detached, frozen in liquid nitrogen, and stored at −80°C until used.

Plasmid Construction, Plant Transformation, and Production of Composite Plants

To generate a plasmid to overexpress the pre-miR172 in common bean transgenic roots, a 217-bp PCR product was obtained using common bean nodule complementary DNA (cDNA) as template and the specific primers Fw-pre172 (5′-CACCCCAGTCACTGTTTGCCGGTGGAG-3′) and R-pre172 (5′-AAAAACCTCCTTTGCTCTGAGCGT-3′), based on the Phvul.001G233200 sequence. The PCR product was cloned by T-A annealing into pCR 2.1-TOPO (Invitrogen) and sequenced. To construct the OE172 plasmid, the pre-miR172 region was excised using the *Xho*I and *Bam*HI sites of the vector and cloned into the pTDTO plasmid that carries the reporter tdTomato (red fluorescent protein) gene (Aparicio-Fabre et al., 2013; Supplemental Fig. S3).

The complete cDNA clone of common bean *AP2-1* was obtained by PCR amplification using cDNA from roots and the specific primers Fw-AP2 (5′-CAGCTACCTTTCCGCCAAATGC-3′) and Rv-AP2 (5′-TAGGCTGG-GATGGTGTCTGCAG-3′), based on the Phvul.005G138300 sequence. The 1,127-bp product was cloned by T-A annealing into pCR 2.1 TOPO (Invitrogen) and analyzed by sequencing. Mutations of the putative miR172 cleavage site of *AP2-1* were introduced using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). The *Pst*I site present in the wild-type miR172 cleavage site (5′-CTGCAGCATCATCAGGATCT-3′) was eliminated by changing the A to a C at the 3′ position of the recognition site for this enzyme. Additionally, the nucleotides at positions 9, 10, and 11 of the miR172 cleavage site were modified by introducing a *Bgl*II site. The primers used were forward (5′-TTCTCTACTGCCGCCAGATCTGGATTCTCAATT-3′) and reverse (5′-AATTGAGAATCCAGATCTGGCGGCAGTAGAGAA-3′). The changes were checked by sequencing, and the modified AP2 was cloned into plasmid pTDTO to obtain plasmid OEAP2m. The nucleotide changes in AP2m introduced an amino acid substitution (Arg for Ser), but this does not seem to affect AP2-1 function (Supplemental Fig. S3).

Common bean composite plants with transformed root system and untransformed aerial system were generated as described (Estrada-Navarrete et al., 2007; Aparicio-Fabre et al., 2013). For plant transformation, *Agrobacterium rhizogenes* K599 strains bearing the EV, OE172, or OEAP2m plasmids were used. Selected composite plants were grown under controlled environmental conditions in pots with vermiculite and watered daily with B&D nutrient solution (Broughton and Dilworth, 1971), either nitrogen free for the symbiotic condition or with 10 mM potassium nitrate for the full-nutrient condition. SNF plants were adapted by growing in pots for 7 d and then inoculated with *R. etli* CE3. For experiments designed to analyze the effect of nitrate on symbiosis, B&D nutrient solution supplemented with 1 or 3 mM KNO₃ was used to water the inoculated plants daily from 1 dpi. Composite plants were analyzed phenotypically at different dpi; transgenic roots and nodules were collected in liquid nitrogen and stored at −80°C.

To analyze the initial events of rhizobial infection in transgenic roots, plastic square bioassay dishes were used to grow selected composite plants under the same conditions described above. Plates containing composite plants were sealed with Parafilm, and the root zone was covered with aluminum foil. After 2 d, composite plants were inoculated by applying 1 mL of *R. etli* CE3 saturated liquid culture directly to the root. At 48 hpi, the root responsive zones were detached and stored at −80°C until used or collected into phosphate-buffered saline (PBS) buffer for microscopic analyses.

Phenotypic Analysis

Nitrogenase activity was determined in detached nodulated roots by the acetylene reduction assay essentially as described by Hardy et al. (1968). Specific activity is expressed as nmol ethylene h⁻¹ g⁻¹ nodule dry weight.

The root fresh weight and the number of secondary roots per plant were determined in composite plants grown under full-nutrient (10 d) or symbiotic (21 dpi) conditions.

Microscopic analysis was performed on transgenic nodules at different developmental stages from EV, OE172, and OEAP2m composite plants. The protocol described by Haynes et al. (2004) was used for tissue staining with the nucleic acid-binding dye SYTO13. Nodule sections were stained with SYTO13 (1 μL mL⁻¹) in 80 mM PIPES, pH 7, for 5 min, then mounted in 1% (v/v) PBS/50% (v/v) glycerol and analyzed. Images were obtained using the Zeiss LSM 510 laser scanning microscope attached to an Axiovert 200 M. SYTO13 excitation was obtained at 488 nm using an argon laser and an HFT UV 488/543/633-nm dual dichroic excitation mirror with an LP 560 emission filter for detection. Sequentially, red fluorescence from the reporter gene was observed by exciting at 543 nm with a helium/neon laser, with the same dual dichroic excitation mirror and a BP 500-530 IR emission filter. Images were processed using the LSM 510 version 4.2 5P1 software (Carl Zeiss Micro-Imaging). For the determination of nodule perimeter and SYTO13 intensity per infected area, 10 images from individual nodule replicates from each condition were analyzed using the ImageJ program.

Statistical analyses of symbiotic parameters (root biomass/architecture, nodulation, and nitrogenase activity) were performed using one-way ANOVA and multiple paired Student's *t* tests ($P < 0.001$).

For analyses of root hair deformation and infection thread induction by rhizobial infection, the root responsive zones from inoculated composite plants grown in plastic square bioassay dishes, as described above, were collected at 48 hpi into PBS buffer. Responsive zone root samples were stained with 0.01% (w/v) Methylene Blue for 1 h and washed three times with double-distilled water; infection events were observed in an optical microscope.

RNA Isolation and Analysis

Total RNA was isolated from 100 mg of frozen nodules, 250 mg of frozen roots, or 200 mg of other frozen tissues from wild-type or composite plants grown under similar conditions, using Trizol reagent (Life Technologies) following the manufacturer's instructions. These samples were preserved at −80°C until tested. Genomic DNA removal, cDNA synthesis, and quality verification for qRT-PCR were performed as reported (Hernández et al., 2007).

RNA preparations were used to detect mature miR172 in different plant tissues by low-molecular-weight RNA-gel hybridization using 15 μg of total RNA, as reported (Naya et al., 2014). Synthetic DNA oligonucleotides with antisense sequence corresponding to miR172 (5′-ATGCAGCATCATCAA-GATTCT-3′) and to U6 snRNA (5′-CCAATTTTATCGGATGTCCCCG-3′) were used as probes after radioactive labeling. Hybridization of U6 snRNA was used as a loading control. The signal intensities of miR172 and U6 hybridization bands were determined using ImageQuant 5.2 software (Molecular Dynamics). Normalized miR172 expression levels were calculated related to U6 snRNA.

For the quantification of transcript levels of mature miRNAs, cDNA was synthesized from 1 μg of total RNA using the NCode miRNA First-Strand cDNA Synthesis Kit (Invitrogen) or the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas) for transcripts of selected genes. Resulting cDNAs were then diluted and used to perform qRT-PCR assays using SYBR Green PCR Master Mix (Applied Biosystems), following the manufacturer's instructions. The sequences of oligonucleotide primers used for qRT-PCR amplification of each gene are provided in Supplemental Table S1. Reactions were analyzed in a real-time thermocycler (Eco Illumina Real-Time PCR System; Illumina) with settings of 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 57°C for 60 s. Relative expression for each sample was calculated with the comparative C_t method. The C_t value obtained after each reaction was normalized with the C_t value of miR159 for miRNA levels or with the C_t value of *UBC* (Phvul.006G110100) for expression levels of miRNAs and transcripts, respectively.

Statistical analyses of gene expression (miR172c, *AP2-1*, and early symbiotic genes) from wild-type and composite SNF plants were performed using one-way ANOVA and multiple paired Student's *t* tests ($P < 0.001$).

Identification of Root-Enhanced Genes Coexpressed with *AP2-1*

To identify common bean genes with an expression pattern similar to that of the *AP2-1* target gene, the Euclidian distance between Z scores for each gene was determined in RNA-seq samples from the *Pv* GEA (O'Rourke et al., 2014). The tissue samples analyzed were as follows: young roots, prefixing effective

(5 dpi) nodules, effective (21 dpi) nodules, ineffective (21 dpi) nodules, roots from nonsymbiotic plants grown in full-nutrient solution, nodule-detached roots (5 dpi), effective nodule-detached roots (21 dpi), and ineffective nodule-detached roots (21 dpi). A threshold Euclidian distance of 0.9 was established as significant. A total of 114 genes within the threshold were identified as genes coexpressed with *AP2-1* (Supplemental Table S5).

Gene Sequence Analysis for the Identification of TFBS

The CLOVER program (Frith et al., 2004) was used to identify TFBS in 5′ promoter regions of *AP2-1* coexpressed genes (Supplemental Table S5) and of genes highly expressed in mature effective nodules described previously by O'Rourke et al. (2014). For this analysis, a 2,000-bp sequence from the region immediately upstream of the transcription start site of each gene was retrieved from the common bean genome sequence (Schmutz et al., 2014; http://www.phytozome.net/commonbean.php, v1.0).

Promoter regions (2,000-bp sequence) of selected early nodulation genes were tested for DREB/ERF or NIN binding sites using http://plants.rsat.eu/. A Markov order of 2 was used for predicting cis-regulatory element-enriched regions with default parameters. The cis-regulatory element-enriched regions were also searched in the promoter region of each locus encoding an miR172 isoform; 1,000 bp upstream of the transcription start site of isoforms a and c and 1,500 bp upstream of the precursors of isoforms b, d, e, and f were analyzed (Supplemental Table S2).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. miR172 isoforms encoded in the common bean genome and the most stable secondary structures predicted for their precursors.

Supplemental Figure S2. Symbiotic phenotypes of common bean plants inoculated with the *R. etli nifA⁻* mutant strain as compared with the CE3 wild-type strain.

Supplemental Figure S3. Schematic representation of plasmids used for miR172c or *AP2-1m* overexpression.

Supplemental Figure S4. Overexpression of miR172c and *AP2m* in transgenic roots and nodules of composite bean plants.

Supplemental Figure S5. miR172 and *AP2-1* control rhizobial infection in common bean roots.

Supplemental Figure S6. Expression pattern of *AP2-1* coexpressed genes.

Supplemental Figure S7. Expression analysis of common bean symbiotic (*Lb*) and nonsymbiotic (*Hb*) hemoglobin genes.

Supplemental Table S1. Primer sequences for qRT-PCR.

Supplemental Table S2. TFBS identified in the 5′ promoter region of each *MIR172* gene.

Supplemental Table S3. Nitrogenase activity and expression analysis of marker genes for nodule development.

Supplemental Table S4. Expression analysis of marker genes for nodule development in transgenic nodules of OE172, EV, or OEAP2m plants.

Supplemental Table S5. *AP2-1* coexpressed genes: assigned GO categories and identified TFBS.

LITERATURE CITED

Aparicio-Fabre R, Guillén G, Loredo M, Arellano J, Valdés-López O, Ramírez M, Iñiguez LP, Panzeri D, Castiglioni B, Cremonesi P, et al (2013) Common bean (*Phaseolus vulgaris* L.) PvTIFY orchestrates global changes in transcript profile response to jasmonate and phosphorus deficiency. BMC Plant Biol **13:** 26

Arenas-Huertero C, Pérez B, Rabanal F, Blanco-Melo D, De la Rosa C, Estrada-Navarrete G, Sánchez F, Covarrubias AA, Reyes JL (2009) Conserved and novel miRNAs in the legume *Phaseolus vulgaris* in response to stress. Plant Mol Biol **70:** 385–401

Ariel F, Brault-Hernandez M, Laffont C, Huault E, Brault M, Plet J, Moison M, Blanchet S, Ichanté JL, Chabaud M, et al (2012) Two direct targets of cytokinin signaling regulate symbiotic nodulation in *Medicago truncatula*. Plant Cell **24:** 3838–3852

Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. Plant Cell **15:** 2730–2741

Bazin J, Bustos-Sanmamed P, Hartmann C, Lelandais-Brière C, Crespi M (2012) Complexity of miRNA-dependent regulation in root symbiosis. Philos Trans R Soc Lond B Biol Sci **367:** 1570–1579

Berrabah F, Bourcy M, Eschstruth A, Cayrel A, Guefrachi I, Mergaert P, Wen J, Jean V, Mysore KS, Gourion B, et al (2014) A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. New Phytol **203:** 1305–1314

Boualem A, Laporte P, Jovanovic M, Laffont C, Plet J, Combier JP, Niebel A, Crespi M, Frugier F (2008) MicroRNA166 controls root and nodule development in *Medicago truncatula*. Plant J **54:** 876–887

Broughton WJ, Dilworth MJ (1971) Control of leghaemoglobin synthesis in snake beans. Biochem J **125:** 1075–1080

Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.): model food legumes. Plant Soil **252:** 55–128

Bustos-Sanmamed P, Mao G, Deng Y, Elouet M, Khan GA, Bazin J, Lelandais-Brière C (2013) Overexpression of miR160 affects root growth and nitrogen-fixing nodule number in *Medicago truncatula*. Funct Plant Biol **40:** 1208–1220

Chen X (2004) A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. Science **303:** 2022–2025

Combier JP, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, et al (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. Genes Dev **20:** 3084–3088

Contreras-Cubas C, Rabanal FA, Arenas-Huertero C, Ortiz MA, Covarrubias AA, Reyes JL (2012) The Phaseolus vulgaris miR159a precursor encodes a second differentially expressed microRNA. Plant Mol Biol **80:** 103–115

Crespi M, Frugier F (2008) De novo organ formation from differentiated cells: root nodule organogenesis. Sci Signal **1:** re11

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res **39:** W155–W159

De Luis A, Markmann K, Cognat V, Holt DB, Charpentier M, Parniske M, Stougaard J, Voinnet O (2012) Two microRNAs linked to nodule infection and nitrogen-fixing ability in the legume *Lotus japonicus*. Plant Physiol **160:** 2137–2154

D'haeseleer K, Den Herder G, Laffont C, Plet J, Mortier V, Lelandais-Brière C, De Bodt S, De Keyser A, Crespi M, Holsters M, et al (2011) Transcriptional and post-transcriptional regulation of a NAC1 transcription factor *in Medicago truncatula* roots. New Phytol **191:** 647–661

Dong Z, Shi L, Wang Y, Chen L, Cai Z, Wang Y, Jin J, Li X (2013) Identification and dynamic regulation of microRNAs involved in salt stress responses in functional soybean nodules by high-throughput sequencing. Int J Mol Sci **14:** 2717–2738

Estrada-Navarrete G, Alvarado-Affantranger X, Olivares JE, Guillén G, Díaz-Camino C, Campos F, Quinto C, Gresshoff PM, Sanchez F (2007) Fast, efficient and reproducible genetic transformation of *Phaseolus* spp. by *Agrobacterium rhizogenes*. Nat Protoc **2:** 1819–1824

Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM (2010) Molecular analysis of legume nodule development and autoregulation. J Integr Plant Biol **52:** 61–76

Formey D, Sallet E, Lelandais-Brière C, Ben C, Bustos-Sanmamed P, Niebel A, Frugier F, Combier JP, Debellé F, Hartmann C, et al (2014) The small RNA diversity from *Medicago truncatula* roots under biotic

interactions evidences the environmental plasticity of the miRNAome. Genome Biol 15: 457

Franco AA, Munns DN (1982) Nodulation and growth of *Phaseolus vulgaris* in solution culture. Plant Soil 66: 149–160

Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z (2004) Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res 32: 1372–1381

Girard L, Valderrama B, Palacios R, Romero D, Dávila G (1996) Transcriptional activity of the symbiotic plasmid of *Rhizobium etli* is affected by different environmental conditions. Microbiology 142: 2847–2856

Hardy RW, Holsten RD, Jackson EK, Burns RC (1968) The acetylene-ethylene assay for N_2 fixation: laboratory and field evaluation. Plant Physiol 43: 1185–1207

Haynes JG, Czymmek KJ, Carlson CA, Veereshlingam H, Dickstein R, Sherrier DJ (2004) Rapid analysis of legume root nodule development using confocal microscopy. New Phytol 163: 661–668

Hernández G, Ramírez M, Valdés-López O, Tesfaye M, Graham MA, Czechowski T, Schlereth A, Wandrey M, Erban A, Cheung F, et al (2007) Phosphorus stress in common bean: root transcript and metabolic responses. Plant Physiol 144: 752–767

Huijser P, Schmid M (2011) The control of developmental phase transitions in plants. Development 138: 4117–4129

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14: 787–799

Jung JH, Seo YH, Seo PJ, Reyes JL, Yun J, Chua NH, Park CM (2007) The *GIGANTEA*-regulated microRNA172 mediates photoperiodic flowering independent of *CONSTANS* in *Arabidopsis*. Plant Cell 19: 2736–2748

Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. Dev Cell 4: 205–217

Kouchi H, Imaizumi-Anraku H, Hayashi M, Hakoyama T, Nakagawa T, Umehara Y, Suganuma N, Kawaguchi M (2010) How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. Plant Cell Physiol 51: 1381–1397

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948

Lauressergues D, Delaux PM, Formey D, Lelandais-Brière C, Fort S, Cottaz S, Bécard G, Niebel A, Roux C, Combier JP (2012) The microRNA miR171h modulates arbuscular mycorrhizal colonization of *Medicago truncatula* by targeting NSP2. Plant J 72: 512–522

Lelandais-Brière C, Naya L, Sallet E, Calenge F, Frugier F, Hartmann C, Gouzy J, Crespi M (2009) Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. Plant Cell 21: 2780–2796

Li H, Deng Y, Wu T, Subramanian S, Yu O (2010) Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. Plant Physiol 153: 1759–1770

Licausi F, Ohme-Takagi M, Perata P (2013) APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytol 199: 639–649

Magori S, Kawaguchi M (2009) Long-distance control of nodulation: molecules and models. Mol Cells 27: 129–134

Martin A, Adam H, Díaz-Mendoza M, Zurczak M, González-Schain ND, Suárez-López P (2009) Graft-transmissible induction of potato tuberization by the microRNA miR172. Development 136: 2873–2881

Murray JD (2011) Invasion by invitation: rhizobial infection in legumes. Mol Plant Microbe Interact 24: 631–639

Naya L, Khan GA, Sorin C, Hartmann C, Crespi M, Lelandais-Brière C (2010) Cleavage of a non-conserved target by a specific miR156 isoform in root apexes of *Medicago truncatula*. Plant Signal Behav 5: 328–331

Naya L, Paul S, Valdés-López O, Mendoza-Soto AB, Nova-Franco B, Sosa-Valencia G, Reyes JL, Hernández G (2014) Regulation of copper homeostasis and biotic interactions by microRNA 398b in common bean. PLoS ONE 9: e84416

Okamoto S, Ohnishi E, Sato S, Takahashi H, Nakazono M, Tabata S, Kawaguchi M (2009) Nod factor/nitrate-induced *CLE* genes that drive HAR1-mediated systemic regulation of nodulation. Plant Cell Physiol 50: 67–77

Oldroyd GED (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol 11: 252–263

Oldroyd GED, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol 59: 519–546

O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, McClean PE, Li J, Dai X, Zhao PX, et al (2014) An RNA-Seq based gene expression atlas of the common bean. BMC Genomics 15: 866

Peláez P, Trejo MS, Iñiguez LP, Estrada-Navarrete G, Covarrubias AA, Reyes JL, Sánchez F (2012) Identification and characterization of microRNAs in *Phaseolus vulgaris* by high-throughput sequencing. BMC Genomics 13: 83

Pérez Guerra JC, Coussens G, De Keyser A, De Rycke R, De Bodt S, Van De Velde W, Goormachtig S, Holsters M (2010) Comparison of developmental and stress-induced nodule senescence in *Medicago truncatula*. Plant Physiol 152: 1574–1584

Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs: building a foundation for functional genomics. Plant Physiol 137: 1211–1227

Reid DE, Ferguson BJ, Gresshoff PM (2011a) Inoculation- and nitrate-induced CLE peptides of soybean control NARK-dependent nodule formation. Mol Plant Microbe Interact 24: 606–618

Reid DE, Ferguson BJ, Hayashi S, Lin YH, Gresshoff PM (2011b) Molecular mechanisms controlling legume autoregulation of nodulation. Ann Bot (Lond) 108: 789–795

Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell 25: 2383–2399

Rubio-Somoza I, Weigel D (2011) MicroRNA networks and developmental plasticity in plants. Trends Plant Sci 16: 258–264

Schmid M, Uhlenhaut NH, Godard F, Demar M, Bressan R, Weigel D, Lohmann JU (2003) Dissection of floral induction pathways using global expression analysis. Development 130: 6001–6012

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, et al (2014) A reference genome for common bean and genome-wide analysis of dual domestication. Nat Genet 46: 707–713

Simon SA, Meyers BC, Sherrier DJ (2009) MicroRNAs in the rhizobia legume symbiosis. Plant Physiol 151: 1002–1008

Song QX, Liu YF, Hu XY, Zhang WK, Ma B, Chen SY, Zhang JS (2011) Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol 11: 5

Soyano T, Hirakawa H, Sato S, Hayashi M, Kawaguchi M (2014) Nodule inception creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. Proc Natl Acad Sci USA 111: 14607–14612

Soyano T, Shimoda Y, Hayashi M (2015) NODULE INCEPTION antagonistically regulates gene expression with nitrate in *Lotus japonicus*. Plant Cell Physiol 56: 368–376

Streeter J, Wong PP (1988) Inhibition of legume nodule formation and N_2 fixation by nitrate. Crit Rev Plant Sci 7: 1–23

Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, Yu O (2008) Novel and nodulation-regulated microRNAs in soybean roots. BMC Genomics 9: 160

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739

Turner M, Nizampatnam NR, Baron M, Coppin S, Damodaran S, Adhikari S, Arunachalam SP, Yu O, Subramanian S (2013) Ectopic expression of miR160 results in auxin hypersensitivity, cytokinin hyposensitivity, and inhibition of symbiotic nodule development in soybean. Plant Physiol 162: 2042–2055

Turner M, Yu O, Subramanian S (2012) Genome organization and characteristics of soybean microRNAs. BMC Genomics 13: 169

Valdés-López O, Arenas-Huertero C, Ramírez M, Girard L, Sánchez F, Vance CP, Luis Reyes J, Hernández G (2008) Essential role of MYB transcription factor: PvPHR1 and microRNA: PvmiR399 in phosphorus-deficiency signalling in common bean roots. Plant Cell Environ 31: 1834–1843

Valdés-López O, Yang SS, Aparicio-Fabre R, Graham PH, Reyes JL, Vance CP, Hernández G (2010) MicroRNA expression profile in common bean (Phaseolus vulgaris) under nutrient deficiency stresses and manganese toxicity. New Phytol 187: 805–818

Van de Velde W, Guerra JCP, De Keyser A, De Rycke R, Rombauts S, Maunoury N, Mergaert P, Kondorosi E, Holsters M, Goormachtig S (2006) Aging in legume symbiosis: a molecular view on nodule senescence in *Medicago truncatula*. Plant Physiol 141: 711–720

124

Vincent JMA (1970) Manual for the Practical Study of Root Nodule Bacteria. IBP Handbook No. 15, Blackwell Scientific Publications, Edinburgh

Wang Y, Li P, Cao X, Wang X, Zhang A, Li X (2009) Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. Biochem Biophys Res Commun **378:** 799–803

Wang Y, Wang L, Zou Y, Chen L, Cai Z, Zhang S, Zhao F, Tian Y, Jiang Q, Ferguson BJ, et al (2014) Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate *ENOD40* expression and regulate nodule initiation. Plant Cell **26:** 4782–4801

Wang Y, Wang Z, Amyot L, Tian L, Xu Z, Gruber MY, Hannoufa A (2015) Ectopic expression of miR156 represses nodulation and causes morphological and developmental changes in *Lotus japonicus*. Mol Genet Genomics **290:** 471–478

Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS (2009) The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. Cell **138:** 750–759

Yamashino T, Yamawaki S, Hagui E, Ueoka-Nakanishi H, Nakamichi N, Ito S, Mizuno T (2013) Clock-controlled and FLOWERING LOCUS T (FT)-dependent photoperiodic pathway in *Lotus japonicus*. I. verification of the flowering-associated function of an FT homolog. Biosci Biotechnol Biochem **77:** 747–753

Yan Z, Hossain MS, Wang J, Valdés-López O, Liang Y, Libault M, Qiu L, Stacey G (2013) miR172 regulates soybean nodulation. Mol Plant Microbe Interact **26:** 1371–1377

Yoshikawa T, Ozawa S, Sentoku N, Itoh J, Nagato Y, Yokoi S (2013) Change of shoot architecture during juvenile-to-adult phase transition in soybean. Planta **238:** 229–237

Zhang B, Pan X, Cobb GP, Anderson TA (2006) Plant microRNA: a small regulatory molecule with big impact. Dev Biol **289:** 3–16

Zhu QH, Helliwell CA (2011) Regulation of flowering time and floral patterning by miR172. J Exp Bot **62:** 487–495

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res **31:** 3406–3415

125

BMC Genomics

Genome-wide identification of the *Phaseolus vulgaris* sRNAome using small RNA and degradome sequencing

Damien Formey[1], Luis Pedro Iñiguez[1], Pablo Peláez[2], Yong-Fang Li[3], Ramanjulu Sunkar[3], Federico Sánchez[2], José Luis Reyes[2*] and Georgina Hernández[1*]

Abstract

Background: MiRNAs and phasiRNAs are negative regulators of gene expression. These small RNAs have been extensively studied in plant model species but only 10 mature microRNAs are present in miRBase version 21, the most used miRNA database, and no phasiRNAs have been identified for the model legume *Phaseolus vulgaris*. Thanks to the recent availability of the first version of the common bean genome, degradome data and small RNA libraries, we are able to present here a catalog of the microRNAs and phasiRNAs for this organism and, particularly, we suggest new protagonists in the symbiotic nodulation events.

Results: We identified a set of 185 mature miRNAs, including 121 previously unpublished sequences, encoded by 307 precursors and distributed in 98 families. Degradome data allowed us to identify a total of 181 targets for these miRNAs. We reveal two regulatory networks involving conserved miRNAs: those known to play crucial roles in the establishment of nodules, and novel miRNAs present only in common bean, suggesting a specific role for these sequences. In addition, we identified 125 loci that potentially produce phased small RNAs, with 47 of them having all the characteristics of being triggered by a total of 31 miRNAs, including 14 new miRNAs identified in this study.

Conclusions: We provide here a set of new small RNAs that contribute to the broader knowledge of the sRNAome of *Phaseolus vulgaris*. Thanks to the identification of the miRNA targets from degradome analysis and the construction of regulatory networks between the mature microRNAs, we present here the probable functional regulation associated with the sRNAome and, particularly, in N_2-fixing symbiotic nodules.

Keywords: *Phaseolus vulgaris*, microRNAs, phasiRNAs, Degradome, Nodules, Legumes, Common bean

Background

Phaseolus vulgaris, known as common bean, is the most important legume for human consumption. This crop is the principal source of protein for hundreds of millions of people and more than 18 million tonnes of dry common bean are produced annually [1]. As a legume, *P. vulgaris* is also a model species for the study of symbiosis in association with nitrogen-fixing bacteria in the genus *Rhizobium*. The recent release of the common bean genome sequence [2] allows the research community to extend their analyses and acquire needed knowledge about this organism. In recent years, a number of studies have focused on gene expression in common bean and its role in a broad range of processes [3] including the response to biotic [4, 5] or abiotic [6, 7] stresses. Gene regulation has particularly been investigated at the post-transcriptional level with the action of small regulating elements called microRNAs (miRNAs) [8–11].

The regulatory processes performed by miRNAs are widely conserved in plants, animals, protists and fungi [12–15], highlighting the significant influence that these regulators can have on the evolution of gene expression. The miRNAs are small non-coding RNA sequences of ~ 22 nt that negatively regulate gene expression, usually, post-transcriptionally by base-pairing to complementary

* Correspondence: jlreyes@ibt.unam.mx; gina@ccg.unam.mx
[2]Departamento de Biología Molecular de Plantas, Instituto de Biotecnología (UNAM), Av. Universidad 2001, Cuernavaca 62210, Morelos, Mexico
[1]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México (UNAM), Av. Universidad 1001, Cuernavaca 62210, Morelos, Mexico
Full list of author information is available at the end of the article

Formey *et al. BMC Genomics* (2015) 16:423

Page 2 of 17

transcripts. In plants, these small RNAs are processed from longer hairpin-shaped precursors encoded in the genome and almost all of them are transcribed by RNA polymerase II. This pri-miRNA, for primary miRNA transcript, is first processed into a precursor, the pre-miRNA, and then excised as a RNA duplex by the endonuclease enzyme Dicer-like 1 (DCL1). The resulting duplex sequence, composed by the guide miRNA and the complementary miRNA* destined for degradation, is then exported to the cytoplasm by diverse factors, including the HASTY protein. Current knowledge is lacking about the spatio-temporal separation of the two strands but, once performed, the guide miRNA is loaded on to a member of the AGO protein family to assemble the RNA-induced silencing complex (RISC). The miRNA can then play its regulatory role by specifically binding a target transcript based on sequence complementarity. To date, three types of action have been identified for plant miRNAs: cleavage, leading to the degradation of the corresponding transcripts; translational inhibition, disrupting protein production; and in some cases DNA methylation, preventing the transcription of the corresponding genomic locus. *In fine*, the miRNA action leads to the loss of function of the gene by inhibiting protein production (see Voinnet [16] for review).

Plant miRNAs are involved in most, if not all, biological processes and have been found in all the organs where they have been searched for [17]. These sequences are key regulators in important processes such as hormone regulation, nutrient homeostasis, development and interaction with pathogens and symbionts [18–22]. As part of some of these mechanisms and processes, miRNAs can act indirectly on gene regulation *via* triggering the production of other small RNAs. These molecules are called phasiRNAs, for phased small interfering RNAs, including the tasiRNAs (*trans*-acting siRNAs) and other phased small RNAs that require cleavage by miRNAs [23]. Particular transcripts cleaved by micro-RNAs are recruited by the RNA-dependent RNA polymerase RDR6 and SGS3 to generate double-stranded RNA molecules [24] that are subsequently processed by DCL4 or DCL5 into phasiRNAs of 21 or 24 nt, respectively [25, 26]. Similar to what occurs with miRNAs, these molecules can be loaded on to AGO protein-containing complexes and can direct disruption of protein production by transcript targets, including transcripts distinct from their own production source, but still exhibiting sequence complementarity [27]. Most of the phasiRNAs are produced from protein-coding transcripts and, if not, they are derived from long non-coding mRNAs. Many phasiRNAs target, and are derived from, large protein families such as NB-LRR, MYB and PPR proteins. One recently hypothesized role of phasiRNAs is to target NB-LRR proteins that broadly regulate plant defenses

and beneficial microbial interactions [27], which are important features for model legumes such as common bean.

To date, around 10000 miRNAs have been identified in all the Viridiplantae organisms reported in miRBase version 21 [28]. Despite several studies on this topic [29–31], only 8 precursors of miRNAs, generating 10 mature sequences, are referenced for *P. vulgaris* in this database while more than 500 are present for other model legumes such as *Medicago truncatula* or *Glycine max*. The phasiRNAs have not been studied in this organism. Here, we use the recently released genome of common bean [2], 5 small RNA libraries obtained from 5 plant organs, and degradome sequencing to identify a high confidence genome-wide common bean miRNA dataset, the associated target transcripts, and the first *P. vulgaris* phasiRNA catalog ever published.

Results and discussion

Overview of the sequencing data

In this study, we used four published high-throughput sequencing libraries of small RNAs [29] obtained from flowers, leaves, seedlings and roots of *Phaseolus vulgaris* and a novel library acquired from symbiotic nodules of the same organism obtained by infection with *Rhizobium tropici*. To identify the whole set of miRNAs and phasiR-NAs, we performed this analysis with the recently described genome of *Phaseolus vulgaris* (Phaseolus vulgaris v1.0, DOE-JGI and USDA-NIFA, http://www.phytozome. net/commonbean) [2] as a reference. For the plant organ libraries and the symbiotic nodules, we obtained averages of 3,649,274 and 2,810,685 reads, respectively (Table 1). From these sets, an average of 51 and 37 % of reads were matched to the common bean genome, respectively (Table 1), and 8.3 % of the sequences from the nodule library matched with the corresponding bacterial genome (*Rhizobium tropici CIAT899*, [32]). The lower percentage of reads mapping to the plant genome in the nodule library was partly due to the presence of bacterial sequences and to an increased abundance of ligated adaptor sequences in this particular sample.

Identification and organ distribution of miRNAs

To identify the already referenced and novel miRNAs of *Phaseolus vulgaris*, we used the miRDeep-P pipeline [33] with each library as a set of small RNA candidates and the *P. vulgaris* genome as the corresponding reference.

Table 1 Statistics of the 5 small RNA library genomic mapping

	Filower	Leave	Root	Seedling	Nodule
Raw reads	3356817	3321526	3963174	3955581	2810685
Mapped reads	1756842	1663979	1950627	2138258	1032364
% of mapped reads	52 %	5 %	49 %	54 %	37 %

Formey *et al. BMC Genomics* (2015) 16:423

Page 3 of 17

We identified a total of 307 precursors that fulfilled our criteria (see methods) producing 185 unique mature miRNAs (Table 2). 64 of them are already referenced as mature miRNAs in other plant species (miRBase database ver. 21 [28]): these are produced by 111 precursors and distributed throughout 27 families. In this work, we refer to those sequences as "known" miRNAs. Additionally, 57 miRNAs are new isoforms (or family members) [34] of already referenced miRNAs in the miRBase database: these are generated by 98 precursors and distributed in 25 families (Additional file 1: Table S1). We refer to them as "new isomiRs". Finally, we identified 64 novel miRNAs that are not members of previously described miRNA families. These novel miRNAs are encoded by 98 precursors, grouped in 59 families.

In summary, we identified a total of 185 mature miRNAs encoded by 307 precursors and distributed in 98 families. These microRNAs included 64 already known miRNAs, 57 novel isoforms belonging to known miRNA families and a last set of 64 novel miRNAs not identified before. Among the 40 families already registered in miRBase 21 that we have identified (Additional file 1: Table S1), we retrieved all those conserved within the angiosperm genomes, which are the miRNAs from miR156 to miR408 [35]. As expected, the highly conserved miR482, together with the more restricted miR1512 and miR1515, which have been reported as positively regulating nodule number [36], are also present in our dataset. Part of the other identified miRNA families belong to the so-called "legume" miRNAs [37] such as the miR2111, miR2118 and miR2119. Other families retrieved in our libraries had only been identified, up to now, in *G. max* (miR4415, miR4416 and miR5786), *M. truncatula* (miR2597) or both (miR5037). Thanks to the identification of these miRNAs in *P. vulgaris,* we observe that they are now shared between at least two legume species and can properly be called "legume" miRNAs [37]. Interestingly, we also found 5 members of the miR1862 family and 1 member of the
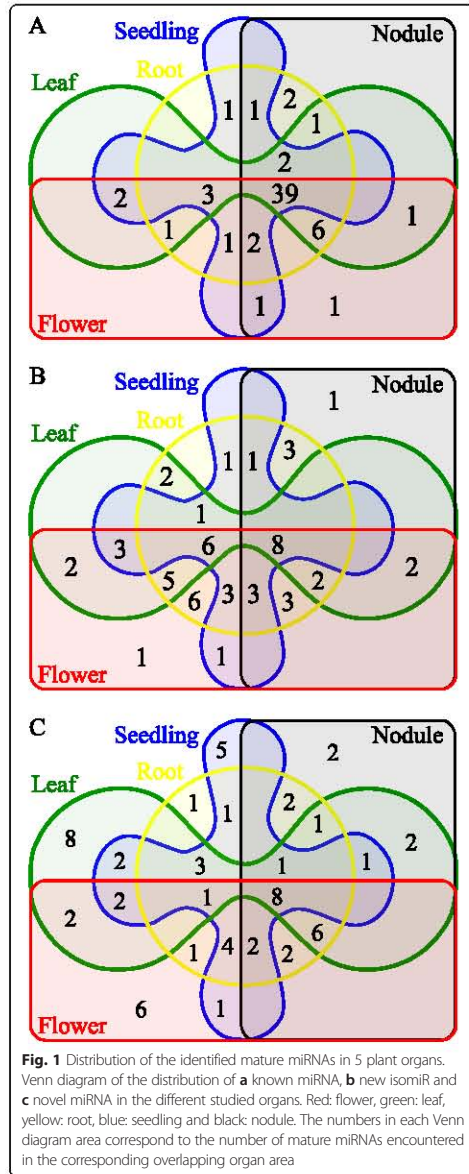
miR6175 that so far have only been identified in rice [38] and rubber tree [39], respectively. The members of these two families are sparsely expressed in our libraries and we can imagine that these miRNAs, and others characterized by exhibiting low expression, are highly spatially- or temporally-specific sequences. It is possible that this type of "specific" miRNA is, in fact, present in several organisms and, with the increase in the depth of high-throughput sequencing technologies, will begin to emerge in investigations focused on the identification of miRNAs in less-studied organisms. The miR4376 family may be an example of these. It has been shown to be a super-family derived from the miR390 and was probably present in the common ancestor of Spermatophyta [40], suggesting that this miRNA is conserved in most of the seed plants, but currently only identified in five species: soybean, tomato, ginseng, potato and, now, common bean (miRBase 21).

By sequencing five small RNA libraries corresponding to five different plant organs, *i.e.*, flower, leaf, nodule, root and seedling, we are able to indicate the distribution of the identified miRNAs in the whole plant based on their expression (see methods). As expected, we found about 61 % of the known miRNAs (39/64) present in all of the organs while only about 14 % and 12 % of the new isomiRs (8/57) and the novel miRNAs (8/64), respectively, are in this class (Fig. 1). In contrast, 22 novel miRNAs are organ-specific whereas none of the known miRNA is. This suggests that the known miRNAs, composed of a majority of conserved miRNAs, play a role at the whole-plant level, as expected from their implication in regulatory networks [41]. In contrast, the newly identified miRNAs, considered as recently emerged and, probably having a more specific action (or no action at all) are characterized by a specific spatial distribution [42]. Alternatively, the characteristic lower expression of isomiRs and novel miRNAs may preclude accurate detection in all organs tested, thus limiting this analysis.

The analysis of miRNA expression and distribution in different organs has been used to identify nodule-specific novel miRNAs in *G. max* and *M. truncatula* [43, 44]. Similar approaches used in this work led us to identify three nodule-specific miRNAs in *P. vulgaris*. These include the new isoform of a conserved miRNA family, the miRNov627 from the mtr-miR399 family and two newly identified miRNAs: miRNov153 and miRNov494. Their putative targets are involved in the regulation of plant-microorganism interactions, which will be discussed below. Other miRNAs with organ-specific expression patterns were detected in our samples and could be also regulating specific processes in other plant tissues (Additional file 2: Table S2).

Table 2 Statistics of the identified precursors and their corresponding mature miRNAs

	Known	New isomiR	Novel	Total
Identified precursors	111	98	98	307
Identified miRNAs	64	57	64	185
Relative position				
Intergenic	97	90	78	265
Intron	1	5	16	22
Exon	11	3	4	18
3'UTR	2	0	0	2
5'UTR	0	0	0	0

Formey *et al. BMC Genomics* (2015) 16:423

Page 4 of 17

Fig. 1 Distribution of the identified mature miRNAs in 5 plant organs. Venn diagram of the distribution of **a** known miRNA, **b** new isomiR and **c** novel miRNA in the different studied organs. Red: flower, green: leaf, yellow: root, blue: seedling and black: nodule. The numbers in each Venn diagram area correspond to the number of mature miRNAs encountered in the corresponding overlapping organ area
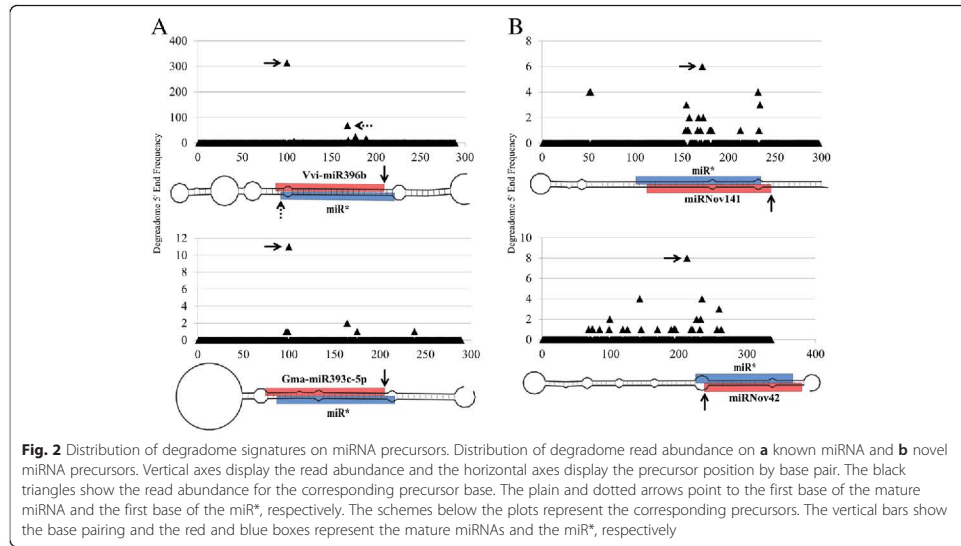
Validation of the predicted precursors

The recent release of the *P. vulgaris* gene atlas [45] allowed us to verify that the predicted miRNAs have a corresponding transcript in the 24 transcriptome data samples collected from seven distinct tissues of common bean at developmentally important time-points, including plants inoculated with either effective or ineffective *Rhizobium*. We found that about 65, 26 and 24 % of the known miRNA, new isomiR and novel miRNA precursors, respectively, have a complete transcript in at least one of the 24 common bean expression libraries. These validated precursors produce 84, 47 and 35 % of the known, new isomiR and novel mature sequences. To differentiate miRNAs from other small RNAs, we used degradome data to identify the specific signatures of DCL slicing during the miRNA precursor processing. These signatures are characterized by the finding of a significant processing event at the mature miRNA and/ or miRNA* boundaries within the precursor [46]. With this method, we validated the correct processing of 41 % of the known miRNA precursors, 30 % of the new isomiR precursors and 22 % of the novel miRNA precursors, thus constituting the validation of 63 % of the known mature miRNAs, 44 % of the new mature isomiRs and 33 % of the novel mature miRNAs (Fig. 2). The failure to validate some miRNA precursors could be due to undesirable RNA degradation producing random degradome signatures, or the presence of degradome signatures originating from other loci that mask the expected significant signatures [47]. It is also possible that some precursors are not functional and mature miRNAs only originate from a subset of potential precursors within the same MIR family. Compared with the precursors encoding well-conserved miRNAs, a lower proportion of precursors for novel miRNAs allowed degradome validation. A general problem encountered with these sequences is their lower expression levels and corresponding lower degradome signatures, the latter being obscured by other gene degradation products.

Finally, if we combine the results of the expression data with those of the degradome analysis, we obtain supporting data for 78, 47 and 35 % of the known, new isomiRs and novel miRNA precursors we encountered, respectively, representing 92, 54 and 44 % of the corresponding mature miRNAs originally identified. Overall, 64 % of the identified mature microRNAs show evidence of expression for at least one of their proposed precursors by at least one of the two methods examined.

miRNA identification aided by genomic sequence data

The present study profited from the use of the recently released complete genomic sequence for *P. vulgaris* [2] to identify miRNAs. Compared to the previous study of Peláez *et al.* [29], this new approach allowed us to confirm the presence of 81 of the 113 known miRNAs identified by Peláez *et al.* in *P. vulgaris*. 25 out of the 32 miRNAs previously identified by Peláez *et al.* and not encountered in our study are actually not present in the

Formey *et al. BMC Genomics* (2015) 16:423

Page 5 of 17

Fig. 2 Distribution of degradome signatures on miRNA precursors. Distribution of degradome read abundance on **a** known miRNA and **b** novel miRNA precursors. Vertical axes display the read abundance and the horizontal axes display the precursor position by base pair. The black triangles show the read abundance for the corresponding precursor base. The plain and dotted arrows point to the first base of the mature miRNA and the first base of the miR*, respectively. The schemes below the plots represent the corresponding precursors. The vertical bars show the base pairing and the red and blue boxes represent the mature miRNAs and the miR*, respectively

current version of the genome and the remaining seven were more precisely defined in our study. Four of the seven non-identified miRNAs have been detected in the libraries but we designated other more abundant isoforms or novel miRNAs as mature miRNAs in the corresponding precursors. The 3 other miRNAs were not identified by miRDeep-P because of potential splicing events in the hairpin sequence or because of a non-conventional folding of the stem-loop. In contrast, our study revealed 4 already identified mature miRNAs in the miRBase catalog of other species and 34 new isomiRs in addition to the set proposed by Peláez *et al.* Concerning the novel miRNAs, none of those identified by Peláez *et al.* fulfilled our selection criteria. We identified only one isomiR of a novel miRNA previously described by Peláez *et al.*, 21 could not be mapped to the genome and 17 were present in more than 40 locations in the genome and thus were not further considered (see methods). The other 19 miRNAs were not selected because they do not fulfill our folding, splicing or expression criteria. Conversely, the availability of genomic sequence data allowed us to identify 64 high-confidence novel miRNAs that satisfy all our criteria and all those currently accepted for microRNA annotation [48]. Means of 2.5 and 6.9 hits in genome for each known and novel miRNA were found, respectively. We can ascribe this difference to a lower selection pressure on the novel miRNAs [49], compared to conserved ones, and the fact that most of these newly identified miRNAs

are young miRNAs and, perhaps, their selection process is still occurring.

In summary, the new whole-genome investigation of the miRNA candidates allowed us to identify 307 genuine precursors of already known and new miRNAs and 121 high confidence unpublished mature sequences.

MiRNA precursor genomic localization

In plants, most of the miRNA genes are encoded in intergenic regions [16, 50] but some are present in introns, exons, or UTRs [51]. In our study, we localized the different precursors of miRNAs in the genome and determined their position relative to annotated genes. As expected, about 90 % of the known and new isomiR are located in intergenic regions (Table 2). One of the known miRNAs, pvu-miR1514a, has a precursor encoded in an intron and two, bna-miR167d and gma-miR167a, are located in 3'UTRs. There are 11 other known miRNAs located in exons. The proportion of known miRNAs located in exons of annotated genes is high compared to that of miRNAs present in other gene locations; however, due to the lack of curation in the current annotation of the *P. vulgaris* genome this could be an overestimation. These sequences are located in exons of putative small proteins lacking known domains or homologs in other organisms and it is very likely that these precursors are, in fact, intergenic. For the novel miRNAs compared to known ones and new isomiRs, we encountered fewer precursors in

Formey *et al. BMC Genomics* (2015) 16:423

Page 6 of 17

intergenic regions (~80 %, Table 2) but more in introns (~16 %). Since we found a higher proportion of novel miRNAs in introns, we can imagine that some of the younger miRNA genes could originate from introns. There are different hypotheses on the origin of the miRNAs: novel miRNA genes may originate from duplication of other miRNAs, from inverted terminal repeats of transposable elements or from random stem-loop structures emerging from intergenic or intronic regions (see Zhuo *et al.* for review [50]). In this sense, the birth of a miRNA gene is less costly if it derives from a ready-to-use transcription unit that an intron can confer indirectly [52]. In our results, 15 of the 16 precursors of novel miRNAs identified in introns are located in the same orientation as the corresponding gene, supporting this hypothesis. On the one hand, the expression of various miRNAs located in introns is dependent on the expression of the corresponding host gene [53] and we can think that the evolutionary selection of this type of transcription mechanism is fixed. On the other hand, we can hypothesize that some of the miRNAs present in introns are young miRNAs and the use of the transcriptional unit of the host gene is a first step in the evolution and, thereafter, they may acquire their own transcriptional unit and evolve as canonical independent miRNAs thanks to duplication events or exon shuffling [54].

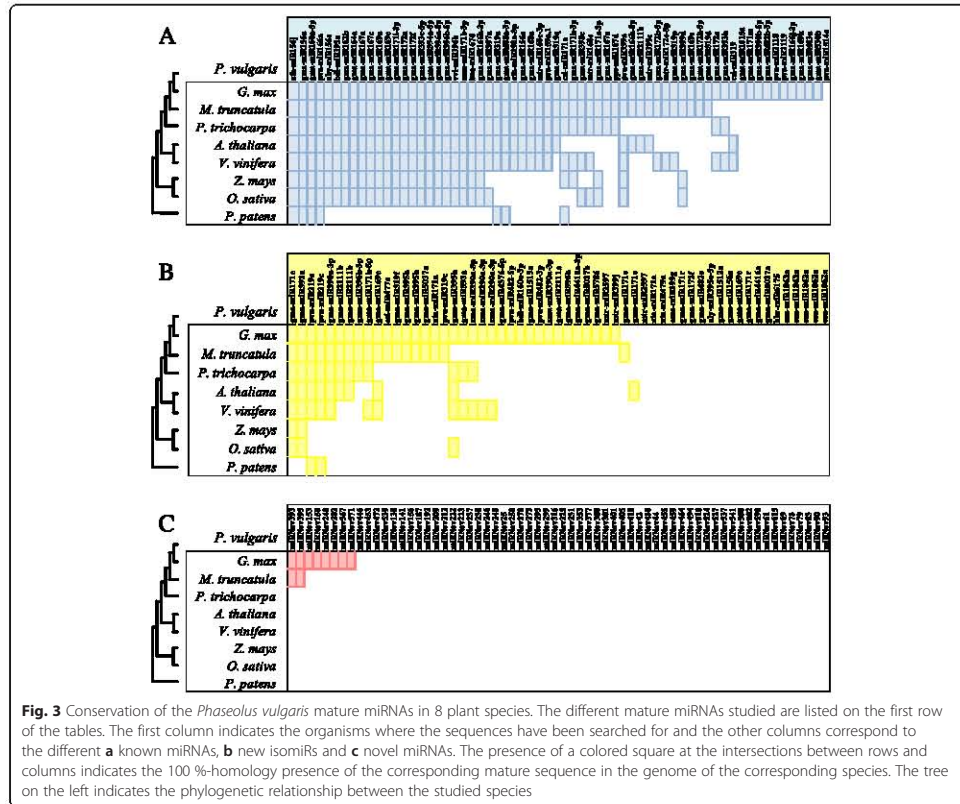### Conservation of the identified miRNAs in plants

We investigated the conservation of the identified miRNAs in 7 vascular plants and 1 moss. We chose two legumes (*Medicago truncatula* and *Glycine max*), three non-legume eudicots (*Vitis vinifera*, *Populus trichocarpa* and *Arabidopsis thaliana*), two monocots (*Oryza sativa* and *Zea mays*) and the moss *Physcomitrella patens*. A miRNA is considered as conserved in a given species when it is present in its full length without any mismatches. As expected, in a global view, the known miRNAs are the most conserved with 4 miRNAs (ath-miR156j, gma-miR156a, gma-miR160a and pvu-miR166a) present in all the selected genomes. We found 21 miRNAs, including 19 known and 2 new isomiRs (pvu-miR319c), conserved in all the plants except the moss (Fig. 3). A total of 118 miRNAs seem to be legume-specific: 22 % of the known miRNAs, 73 % of the new isomiRs and 100 % of the novel miRNAs. Among them, 77 are specific to *P. vulgaris*: 1 known miRNA (pvu-miR1514a), 20 new isomiRs and 56 novel miRNAs.

The proportion of species-specific miRNAs is variable in plants. For example, in *Arabidopsis thaliana* and *A. lyrata*, 35 % of the identified miRNAs are species-specific [55]. This is also the case of 23 % of the *Medicago truncatula* miRNAs [44], 37 % in *Populus* species

[56], 41 % in wheat [57] and 56 % in apple [58]. In our results we found that *P. vulgaris* is in the range of these other plants with around 42 % of the miRNAs specific to the common bean. Naturally, these numbers may fluctuate depending on the depth of sequence obtained and the methods used for the identification and the prediction of the microRNAs. The lack of conservation for the specific miRNAs suggests that they emerged recently during evolution and can thus be considered as young miRNAs.

Prediction and identification of miRNAs targets

Using degradome data and the prediction of putative targets, we are able to present a set of targets for the identified miRNAs. The degradome data reported here were obtained using a method based on the 5'-rapid amplification of cDNA ends (5'RACE) and further adapted for high-throughput sequencing. It allows the experimental identification of the target cleavage sites associated with miRNA cleavage on a genomic scale [59]. The regulation performed by miRNAs does not necessary include a cleavage of the target transcript. To complete the degradome data, we used a target prediction tool, called psRNATarget [60], based on reverse complementary matching between miRNA and target transcript. A mean number of one degradome target and 6.5 predicted targets per miRNA candidate were found (Additional file 3: Table S3). The known miRNAs possess significantly more degradome targets (1.9/miRNA) compared to the novel ones (0.3/miRNA) and also have more predicted targets (8.8/known miRNA *vs.* 5.8/novel miRNA). Several of the target transcripts identified here for known miRNAs correspond to previously defined targets found in other plant species, thus validating our analysis. As discussed in previous studies, the lack of conventional targets for the young miRNAs is not unusual. Most of the young miRNAs, in contrast to the conserved ones, are not involved in complex regulatory networks [41] and their evolution has sometimes been considered as neutral [49]. However, the experimental demonstration of an interaction with its predicted target is considered as "the most powerful method of validating a predicted miRNA" [61]. Although 73 % of the conserved miRNAs exhibit a degradome target in our data, only around 28 % of the novel miRNAs also have a target identified by degradome analysis (Fig. 4). Here we found that a significant portion of novel miRNAs has functional evidence, as revealed by degradome data, for a role in regulating gene expression, thus suggesting their relevance in different biological processes. Their degradome targets are for the most part involved in metabolic processes, biosynthesis, binding processes and various functions (Fig. 5). Compared to the conserved miRNAs that preferentially target genes involved in complex

Fig. 3 Conservation of the *Phaseolus vulgaris* mature miRNAs in 8 plant species. The different mature miRNAs studied are listed on the first row of the tables. The first column indicates the organisms where the sequences have been searched for and the other columns correspond to the different **a** known miRNAs, **b** new isomiRs and **c** novel miRNAs. The presence of a colored square at the intersections between rows and columns indicates the 100 %-homology presence of the corresponding mature sequence in the genome of the corresponding species. The tree on the left indicates the phylogenetic relationship between the studied species
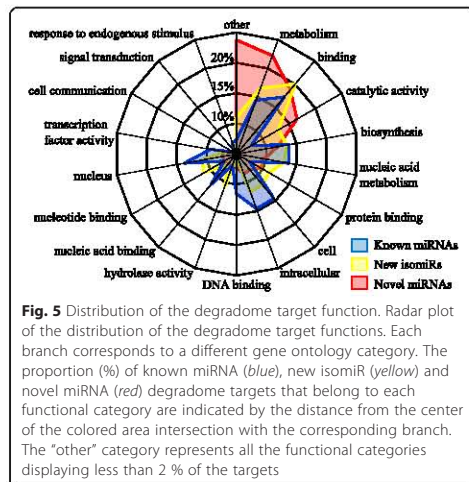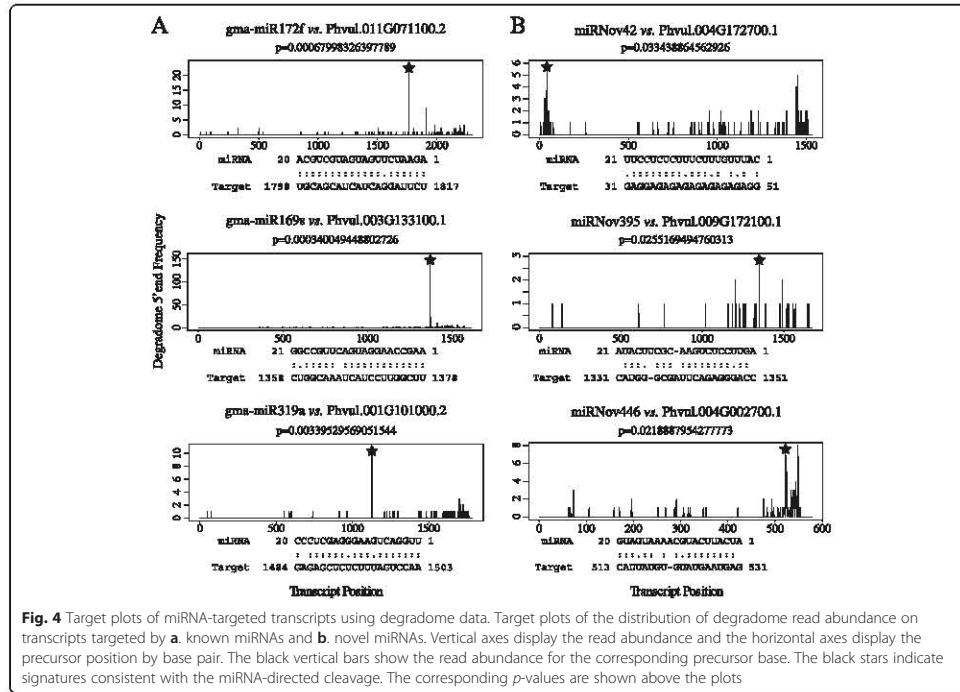
regulatory networks such as transcription factors [62], young miRNAs tend to target more precise and diversified functions. However, one of them, miRNov138, seems to target a transcript coding for a Homeobox-leucine zipper family protein. This transcription factor family is known to play an important role during the growth and development of plants by modulating phytohormone-signaling networks [63] and also in the interaction with microorganisms, especially during nodulation [64]. This young miRNA is specific to *P. vulgaris* and we hypothesize that it has been selected to perform a more precise role in common bean, like adaptation to a specific environment [63] or interaction with *Rhizobium*.

Analysis of microRNA co-expression networks in nodules

To understand the role of the newly identified miRNAs in nodules, we constructed weighted correlation networks of miRNAs. These networks describe the pairwise relationship among miRNAs that differentiate the nodule library from other libraries based on the miRNA expression patterns [65]. Using this approach, we identified 2 networks including pairwise relationships between novel miRNAs, new isomiRs and conserved miRNAs.

One of these networks is composed by 4 novel miRNAs (miRNov153, miRNov155, miR156 and miRNov494), two new members of the miR399 family, and gma-miR319d suggesting that they could act jointly (Fig. 6A). All these miRNAs have the same expression pattern, with increased expression in nodules (Additional file 1: Table S1). As shown in Fig. 7A, increased expression of pre-miR319d, pre-miRNov153/155 and pre-miRNov494 in nodules was experimentally validated by qRT-PCR. More precisely, the new isoform of the mtr-miR399j was specifically identified in nodule (Additional file 2: Table S2). This miRNA family is involved in phosphate homeostasis [66] but, furthermore, has been identified as being repressed by N-starvation [67]. Nodulation efficiency and functionality

Formey *et al. BMC Genomics* (2015) 16:423

Page 8 of 17

Fig. 4 Target plots of miRNA-targeted transcripts using degradome data. Target plots of the distribution of degradome read abundance on transcripts targeted by **a**. known miRNAs and **b**. novel miRNAs. Vertical axes display the read abundance and the horizontal axes display the precursor position by base pair. The black vertical bars show the read abundance for the corresponding precursor base. The black stars indicate signatures consistent with the miRNA-directed cleavage. The corresponding *p*-values are shown above the plots

Fig. 5 Distribution of the degradome target function. Radar plot of the distribution of the degradome target functions. Each branch corresponds to a different gene ontology category. The proportion (%) of known miRNA (*blue*), new isomiR (*yellow*) and novel miRNA (*red*) degradome targets that belong to each functional category are indicated by the distance from the center of the colored area intersection with the corresponding branch. The "other" category represents all the functional categories displaying less than 2 % of the targets

are related to these two nutrients [68, 69] and we hypothesize that this new isoform of miR399 could indirectly regulate the normal establishment of nodulation in *P. vulgaris*. Indeed, this miRNA family is known to target a transcript encoding PHO2, an ubiquitin-conjugating enzyme crucial for acquisition and translocation of phosphate [70]. In our data, we predicted another target for the new isoform of miR399 which could regulate a NB-ARC domain-containing protein, one of several factors known to be involved in the plant resistance and activation of innate immune responses [71]. The organ-specific accumulation of this miRNA and the function of the corresponding target suggest a role in the regulation of nodule-specific defense mechanisms. Two of the novel miRNAs from this network; miRNov153 and miRNov494, are specific to the nodule library. The miRNov494 is predicted to target two transcripts in our degradome data (Additional file 3: Table S3). qRT-PCR expression analysis showed that in nodules the expression of miRNov494 precursor increased while the expression of one of the predicted targets, an aldehyde dehydrogenase (Phvul.004G162200.1) decreased (Fig. 7A). The members of this protein family are known to change their expression in response to a wide variety of stresses and are important in supporting environmental adaptability

Formey *et al. BMC Genomics* (2015) 16:423

Page 9 of 17

Fig. 6 Weighted correlation network analysis of microRNAs from nodules versus other organs. Representation of two regulatory networks **a** & **b** between the miRNAs. The nodes correspond to each miRNA encountered in the corresponding network. The edges correspond to the direct connection and relation between the nodes. *Blue, yellow* and *red* tags indicate the known miRNAs, new isomiRs and novel miRNAs, respectively
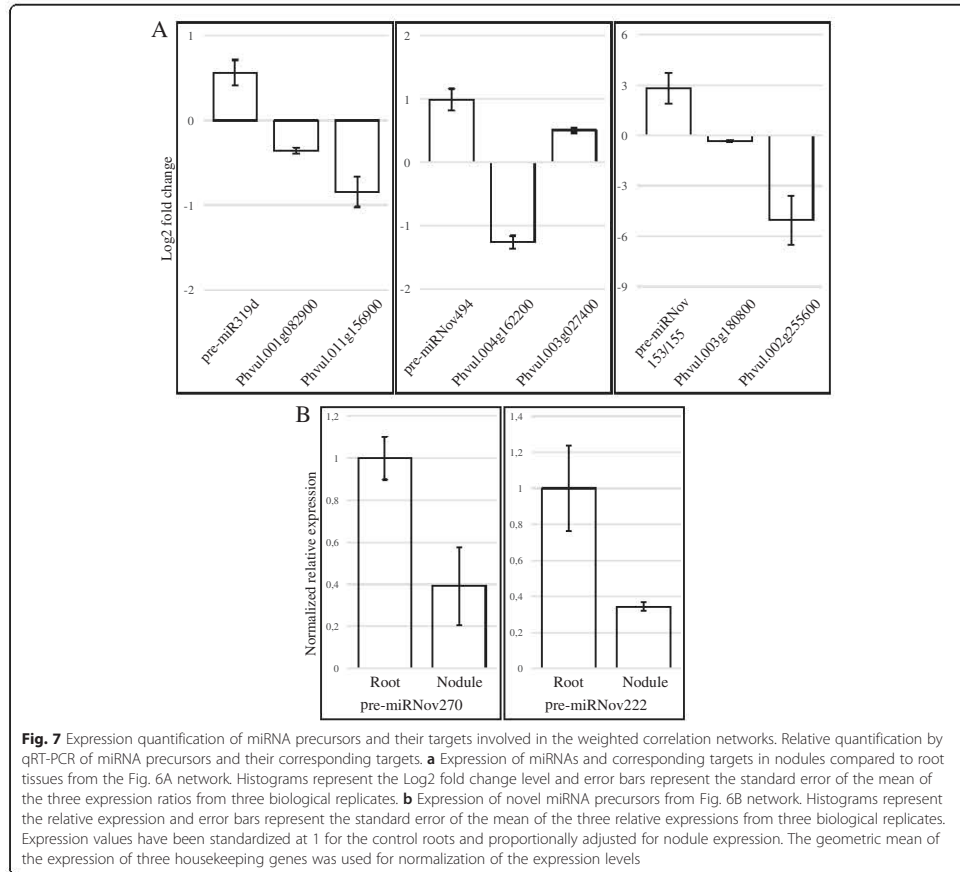
[72]. Our data suggest that the miRNov494 can regulate a member of the aldehyde dehydrogenase family in nodules and may help the plant to control the life cycle of this symbiotic organ. MiRNov153 is ~18 times more highly expressed than the miRNov494 and is encoded in an intron of an F-box protein gene (Additional file 1: Table S1). Part of this protein family plays important role in root symbioses [73] and particularly in the auto-regulation of nodulation [74]. This novel miRNA, miRNov153, is predicted to target an uridine kinase (Phvul.003 g180800), for which no expression variation has been measured in nodule compared to root tissue, and a transcript coding for a hypothetical protein (Phvul.002 g255600), for which we have observed an expression inversely correlated to its regulator miRNA in nodule (Fig. 7A). Although this target does not guide us to an obvious functionality, the fact that this miRNA is encoded in an intron of an F-box protein gene, in the same orientation, would permit us to envisage that this miRNA is co-expressed with the host gene and could act in the regulation of the systemic negative feedback control of nodulation. Among the 8 species for which we searched for the presence of the miRNov153, only the soybean genome encodes this miRNA and displays a *bona fide* hairpin-like precursor (Additional file 4: Figure S1A). To check if this locus can produce a miRNA corresponding to the mature miR-Nov153 in soybean, we mapped public nodule small RNA sequences from Arikit *et al.* [75] to the miRNov153 soybean precursor (Additional file 4: Figure S1B). Tens of

sequences mapped exactly to miRNov153 and the corresponding miRNov153* and only three sequences mapped to other coordinates, suggesting that nodules from soybean also produce the miRNov153. The available public small RNA libraries from nodule represent 5 time points from 10 dpi to 30 dpi. In these libraries, we observe an increased expression of the miRNov153 during nodule development. Additionally, this miRNA is absent from *Medicago truncatula,* which develops indeterminate nodules, and is conserved in common bean and soybean, two species developing determinate nodules. These results tend to show a role of miRNov153, *via* its targets, in the loss of meristematic activity or the senescence of determinate nodules.

To our knowledge, no link has been published before between miR319 and nodulation. This miRNA is a regulator of the plant stress responses against salinity, drought or cold [76, 77] but it is also implicated in the regulation of cell proliferation *via* the control of its targets, members of the TCP transcription factor family [78]. Here, we have validated the increase of miR319d precursor expression in nodules and the decrease of one of its predicted targets, a TCP transcription factor family member (Phvul.011 g156900), in the same tissue (Fig. 7A). Because this miRNA is present in a network that distinguishes the nodule library from the others, and because it is connected with miRNAs described in previous studies as being related to nodules, we hypothesize that miR319 and its targets are strong candidates for which a role in nodule development must be investigated.

We identified a second network involving 7 miRNAs: 2 novel ones, 2 new isomiRs and 3 already known miR-NAs (Fig. 6B). In this network, we retrieved members of miRNA families already described as regulating nodule development. Specific variants of *L. japonicus* and *M. truncatula* miR171 target the GRAS-family *NSP2* TF, a key regulator of the common symbiotic pathway for rhizobial and arbuscular mychorrizal symbioses [79–81]. MiR166 and its target gene, a HD-ZIP III transcription factor, regulate meristem activity and vascular differentiation in roots and nodules [64]. In soybean, the overexpression of miR482, targeting a GSK-3-like protein MsK4, resulted in an increase in nodule number without affecting root development or the number of nodule primordia [36]. The crucial role of miR172 and its target gene, a transcription factor from the AP2 family, have been described for soybean- and common bean-rhizobia symbiosis [82–84]. For soybean, Yan *et al.* [82] postulated that miR172 regulation of nodulation is explained by the *AP2* repression of non-symbiotic hemoglobin (*Hb*) gene expression that regulates the level of nodulation; while Wang *et al.* [84] recently reported that soybean NNC1, a AP2 transcription factor target of

Formey *et al. BMC Genomics* (2015) 16:423

Page 10 of 17

Fig. 7 Expression quantification of miRNA precursors and their targets involved in the weighted correlation networks. Relative quantification by qRT-PCR of miRNA precursors and their corresponding targets. **a** Expression of miRNAs and corresponding targets in nodules compared to root tissues from the Fig. 6A network. Histograms represent the Log2 fold change level and error bars represent the standard error of the mean of the three expression ratios from three biological replicates. **b** Expression of novel miRNA precursors from Fig. 6B network. Histograms represent the relative expression and error bars represent the standard error of the mean of the three relative expressions from three biological replicates. Expression values have been standardized at 1 for the control roots and proportionally adjusted for nodule expression. The geometric mean of the expression of three housekeeping genes was used for normalization of the expression levels

miR172, represses *ENOD40* expression that results in negative regulation of early symbiotic stages. For common bean, we recently demonstrated [83] that miR172c, which has the transcription factor AP2-1 as target gene, indirectly regulates the expression of the transcription factors NF-YA1, NSP2 and CYCLOPS as well as the gene *FLOT2*, all of which are essential regulators of early stages of the symbiosis. In addition, we postulated that miR172-induced AP2-1 silencing in mature common bean nodules is involved in down-regulating expression of genes related to nodule senescence postulated as targets of AP2-1 transcription activation [83]. We also found a new isomiR of a family already described as regulating the development of nodules: the miR2111 targeting a Kelch-related proteins [85]. Finally, the last two miRNAs encountered in this network are new miRNAs: miRNov222 and miRNov270. Although these miRNAs have low expression and present no processing evidences in degradome data or transcription evidences in transcriptome data (Additional file 1: Table S1), we have detected the corresponding precursors by qRT-PCR in roots and confirmed their expression decrease in nodules (Fig. 7B). As elements in a network containing miRNAs previously described as nodulation regulators, these novel miRNAs are also probably involved in the normal establishment and functioning of the nodulation process *via* regulation of their targets. These two miRNAs putatively target a protein kinase and a NB-ARC domain-containing disease resistance protein, respectively (Additional file 3: Table S3). Although the protein kinases belong to a large protein family, various members are known to be involved in signalling during

Formey *et al. BMC Genomics* (2015) 16:423

Page 11 of 17

nodulation [86] and, as described above for the putative target of the new isoform of the miR399, the NB-ARC domain-containing disease resistance proteins are known to be involved in the plant resistance and activation of innate immune responses [71] and could play a specific role in the nodule and allow the proper functioning of the nitrogen-fixing organ. These novel miRNAs are directly related to regulators of nodulation and it is conceivable that these *Phaseolus*-specific miRNAs could act in the same way as the connected conserved miRNAs and play a role during nodule establishment in a more spatio-temporal specific manner. Like the miR319 family, these novel miRNAs must be considered as major candidates for deciphering the specific regulation of nodulation in *P. vulgaris*.

Identification of phasi-RNAs and their associated targets

In all the sequenced libraries, a large number of reads do not correspond to miRNAs or other previously identified non-coding RNAs of common bean, suggesting they could belong to another class of small RNAs. We decided to focus on phasiRNAs and, based on the alignment of the reads with the genome and the phasing score for each locus, we predicted 125 statistically significant loci producing 21 nt-phased small RNAs (Additional file 5: Table S4, see example in Fig. 8A). Among the identified *PHAS* genes, only two are conserved

in the three legumes studied, *i.e. M. truncatula*, *G. max* and *P. vulgaris* (Additional file 6: Figure S2). One of these *PHAS* genes is known as the *TAS3* gene targeted by the miR390 and producing small RNAs targeting transcripts that encode AUXIN RESPONSE FACTORs (ARF3 and ARF4) [87]. In our data, we also identified miR390 and its regulatory action on the *TAS3* gene (Additional file 5: Table S4), providing a positive control for our methodology. The second *PHAS* gene codes for a member of the disease resistance protein (TIR-NBS-LRR class) family known to be targeted by miR2118 in soybean but not yet in *M. truncatula* [71]. In our analysis, we encountered that miR2118 potentially targets this *PHAS* gene but not in the expected phase. However, for this gene, we have identified a potential in-phase slicing by a novel miRNA, the miRNov212 (Additional file 5: Table S4). Among the 125 loci identified, 47 are predicted to be targeted by 31 different miRNAs with a cleavage site in phase with the initiation site of the corresponding phasiRNAs. In recent years, loci producing 21 nt-phasiRNAs (*PHAS* genes) have been identified in different species and their number in each case varies widely: 50 in *Arabidopsis*, 114 in *M. truncatula*, 157 in apple, 353 in peach and 864 loci in rice [40, 71, 88–90]. Not all *PHAS* genes are associated to a miRNA triggering phased small RNA, and the proportion of *PHAS* genes targeted by a miRNA also varies from species to species. For example, 26
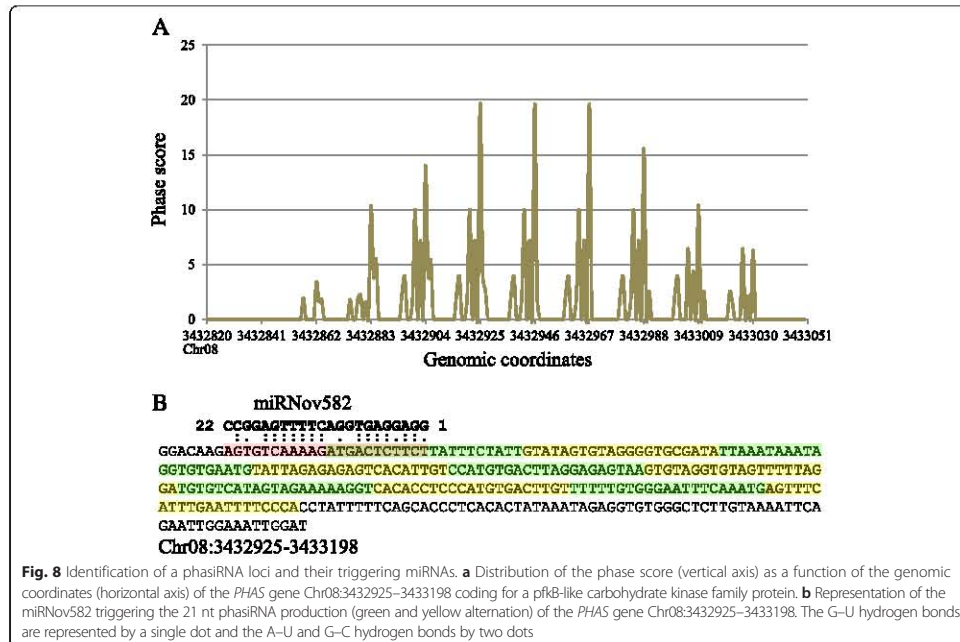
Fig. 8 Identification of a phasiRNA loci and their triggering miRNAs. **a** Distribution of the phase score (vertical axis) as a function of the genomic coordinates (horizontal axis) of the *PHAS* gene Chr08:3432925–3433198 coding for a pfkB-like carbohydrate kinase family protein. **b** Representation of the miRNov582 triggering the 21 nt phasiRNA production (green and yellow alternation) of the *PHAS* gene Chr08:3432925–3433198. The G–U hydrogen bonds are represented by a single dot and the A–U and G–C hydrogen bonds by two dots

Formey *et al. BMC Genomics* (2015) 16:423

Page 12 of 17

of the 157 apple *PHAS* genes and 160 of the 864 rice *PHAS* genes are potentially targeted by miRNAs that can trigger phasiRNA production. Here, we identified 13 known miR-NAs, 4 new isomiRs and 14 novel miRNAs that potentially trigger the production of phasiRNAs (an example is shown in Fig. 8B). Most of the 17 previously identified miRNAs are known to be involved in the production of phasiR-NAs in other species [27, 91]. The phasiRNAs produced by these known miRNAs originate from transcripts encoding proteins involved in a wide range of functions like the MYB genes targeted by miR159, NB-LRR genes targeted by miR482 and miR2118, Ca^{2+}-ATPases targeted by miR4376 or TIR/AFB targeted by miR393 [27]. Uncharacteristically, we identified one *PHAS* gene (Chr11: 2402498–2402857) that possesses all the features to produce two sets of phasiRNAs derived from two distinct phases that are triggered by two novel miRNAs (miRNov141 and miRNov316). The resulting phasiRNAs are only expressed in the flower and the corresponding miRNAs are almost uniquely expressed in the same organ (Additional file 2: Table S2). These phasiRNAs are produced from a transcript coding for a putative serine carboxypeptidase-like 35, a member of a large family involved in multiple cellular processes. In rice, a member of this family has been identified as playing a role in defense against pathogens and oxidative stress [92]. The mechanism of double-phasing by two different miRNAs allows production of two overlapping sets of phasiRNAs from the same sequence and we can imagine this confers an additional layer of small RNA production allowing the more precise regulation of defense reactions of the organism. The complete set of *PHAS* genes targeted by microRNAs must be larger than the number reported here, since we may not have a complete list of miRNAs yet. In addition, other miRNAs may generate phasiRNAs by cleaving target transcripts at an out-of-phase position by a phenomenon called phase-drift, caused by a DCL slippage [93], and, identification of these cleavage events becomes challenging.

All of the studied organs presented transcripts producing 21-nt phasiRNAs but only 13 of the 125 *PHAS* genes (~10 %) were identified in the five different small RNA libraries (Additional file 7: Figure S3). Moreover, most of the phasiRNAs are organ-specific (~73 %) and flowers present the largest set of identified phasiRNAs expressed, with around 73 % of the loci, including ~69 % of organ-specific expressed loci. Conversely, nodules had lower expressed phasiRNAs, with ~13 % of the identified loci. 86 of the 125 *PHAS* loci are localized in a predicted transcript, including 46 with an associated putative function (Additional file 5: Table S4) suggesting that the corresponding phasiRNAs can also target these transcripts or other members of the corresponding gene families.

Among the phasiRNA loci localized in transcripts with a putative function, two are expressed in all the organs and are derived from an Auxin signaling F-box 2 and a NAC transcription factor-like 9. These two proteins are related to auxin signaling and lateral root formation [94] and the corresponding families are known to be common targets of phasiRNAs [40]. The presence of the corresponding phasiRNAs in all the studied organs reflects the important role in regulation of these small RNAs at the organism level. MiR2118 is predicted to target two *PHAS* genes (Chr04:3128306–3129229 and Chr04: 9380782–9381299) coding for proteins involved in disease resistance: a NB-ARC domain-containing disease resistance protein and a Disease resistance protein (TIR-NBS-LRR class) family member. These loci are expressed in all the organs except nodules (Additional file 5: Table S4). Although miR2118 has been retrieved in all the organs studied, including nodules, we can hypothesize that miR2118 regulates nodule growth *via* the control of phasiRNA production derived from the NB-ARC and TIR-NBS-LRR transcripts. The investigation of phasiRNAs is quite recent and we envisage that, in coming years, the number of *PHAS* genes will increase, as it has for miRNAs in recent years, in terms of loci per species and number of studied plant species.

Conclusions

Thanks to the genome-wide identification of miRNAs and phasiRNAs from *P. vulgaris*, we provide a set of sequences allowing the extension of the sRNAome of common bean. The investigation of the miRNA degradome targets, the miRNA correlation networks and the *PHAS* gene function permitted us to propose a role for the identified small RNAs and provide genuine sequence candidates to be studied in plant-microorganism interactions and specifically in the root-nodule symbiosis.

Methods

Plant materials and sequencing

Plant materials

Flower, leaf, root and seedling raw sequences were obtained from the study of Peláez *et al.* [29]. Briefly, *Phaseolus vulgaris* L. cv. Negro Jamapa and cv. Pinto Villa were grown and roots (15 day old) and flower buds (35–40 day old) were collected in liquid N_2 and stored at - 80 °C. Whole 1–4 days old seedlings were collected in liquid N_2 and stored at –80 °C. A pool of leaves from 10 and 20 day-old plants was harvested for RNA purification.

Nodulation conditions

P. vulgaris L. cv. Negro Jamapa seeds were surface sterilized and germinated under sterile conditions for two days and then planted in pots with sterile vermiculite. A

fresh inoculum of *R. tropici* CIAT899 grown on PY liquid medium supplemented with CaCl$_2$ (7 μM), rifampicin (50 μg/ml), and nalidixic acid (20 μg/ml) at 30 °C to a cell density of 5×10^8 ml^{-1} was prepared. Immediately after transferring bean plants into pots with fresh sterile vermiculite, each plant was inoculated by adding 1 ml of the *R. tropici* culture directly to the root. Plants were grown in a greenhouse with a controlled environment (25–28 °C, 16 h light/8 h dark) and were watered with nitrogen-free B&D nutrient solution [95] every 2 days. Root nodules were collected 18 and 27 days after inoculation.

Library preparation and sequencing
Total RNA was isolated from frozen samples using the Trizol reagent according to the manufacturer's instructions (Invitrogen, Carlsbad, CA) and 10 μg of each sample was prepared for Deep Sequencing following Illumina's Small RNA alternative sample preparation protocol v1.5. Complementary DNA (cDNA) libraries were separately prepared and Single Read-sequenced using the Genome Analyzer IIx (GAIIx) (36 bp) and the Illumina Cluster Station (Illumina Inc, USA) at the Instituto de Biotecnología (Universidad Nacional Autónoma de México) and raw reads from the *Illumina Pipeline 1.4* for the small RNA libraries were purged of sequence adapters, low quality tags and small sequences (<16 nt long) [29]. The contamination of the nodule library by bacterial sequences was investigated by BLASTn of the whole nodule small RNA set against the *R. tropici* CIAT899 genome as a reference. Raw reads for the small RNA sequencing are available in the Gene Expression Omnibus database under accession number GSE67409.

Degradome library construction
Plant material
Seeds of *P. vulgaris*, var. Pinto Villa were surface sterilized in 50 % (v/v) sodium hypochlorite and 0.5 % (v/v) Tween-20 for 3 min, rinsed with distilled water for 10 min and germinated in wet paper towels in the dark at 24 °C for 3 days. At this point seedlings of similar size were transferred to moist vermiculite watered to substrate capacity and maintained at 24 °C for 48 h with 16 h/8 h light/dark day cycles. Seedlings were collected, ground to a fine powder under liquid N$_2$ and preserved at –80 °C until the material was used for RNA preparation.

Library construction for high throughput sequencing
The protocol used to obtain cDNA libraries for degradome analysis was carried out essentially as previously described [59, 96]. Total RNA from seedlings was obtained using the Trizol reagent (Life Technologies) according to the manufacturer's directions. Subsequently,

poly-A+ RNA was ligated to a 5'adapter containing a MmeI site at its 3'-end. The ligated products were used for cDNA production and amplified by PCR for five cycles. The PCR products were digested with MmeI and the resulting fragments ligated to a second double-stranded oligonucleotide, purified and amplified for another ten PCR cycles. The final product was purified and subjected to high throughput sequencing as described below.

High Throughput sequencing
DNA libraries were subjected to Single Read-sequencing (36 bp) using a Genome Analyzer IIx (GAIIx) and the Illumina Cluster Station (Illumina Inc, USA) at the UNAM Sequencing facility (Unidad Universitaria de Secuenciación Masiva de DNA, Universidad Nacional Autónoma de México). Raw reads for the degradome sequencing are available in the Gene Expression Omnibus database under accession number GSE67432.

Bioinformatics analysis of sequencing data
miRNA identification
The identification of the miRNA precursors was performed using the miRDeep-P pipeline [33] with a maximal identification window size of 250 nt, no mismatch allowed and a number of hits in the genome lower than 40. The precursors with a small RNA overlapping an ncRNA (tRNAs, rRNAs and other ncRNAs) were discarded by the software. We recovered only the precursors with a size of mature miRNA between 18 nt and 25 nt wherein the mature miRNA lies within the top 5 % of the most expressed sequences, in a given library. The selected mature miRNAs were compared with all the Viridiplantae miRNAs of the miRBase version 21 [28] using the NCBI BLASTn program [97], allowing no mismatches to identify the already known miRNAs. To identify the miRNAs families and the new isoforms of already known miRNAs, we used CD-HIT (Cluster Database at High Identity with Tolerance) [98] with at least 84.2 % of identity. For the novel miRNAs, a more stringent selection was performed and only those mature sequences for which the best precursor miRDeep score is greater than 2.2 (the threshold for the top 5 % of the precursor miRDeep scores) have been kept. To list all the precursors encoding for the final set of novel genuine mature miRNAs, we recovered all the corresponding precursors that have been identified by miRDeep-P and passed all our filters without a selection on the attributed score.

Target identification
Two programs were used to identify the targets of the selected miRNAs. CleaveLand ver.4 [99] was used to identify the putative target sites from degradome data

Formey *et al. BMC Genomics* (2015) 16:423

Page 14 of 17

(see Degradome library construction section). The transcripts from the version 1 of the *P. vulgaris* genome [2] were used as target templates, the total set of identified miRNAs has been used as the small RNA candidates and only targets with a p-value lower or equal to 0.05 were selected. psRNATarget [60] was used to predict putative miRNA targets on the same transcript dataset. Default parameters were used and only the targets with an expectation value lower than 3 were retained.

To identify the functional distribution of these sets of targets, we used the corresponding Gene Ontology Annotation [100] and classified the corresponding GO terms with CateGOrizer [101] using the GO_slim2 classification with the "accumulative all occurrences" count method.

Organ distribution and conservation analyses

The organ distribution was performed manually: a miRNA is considered as present in an organ when the corresponding mature sequence is in the top 5 % of the most expressed mapped sequences of the corresponding organ library. The "five sets" Venn diagrams were inspired by Edwards' Venn diagrams [102].

Conservation of mature miRNAs was investigated using the NCBI BLASTn program [97] against the genomes of 8 species: two Fabaceae (*Medicago truncatula* 4.0 and *Glycine max* Wm82.a2.v1), three non-legume eudicots (*Vitis vinifera* Genoscope.12X, *Populus trichocarpa* 3.0 and *Arabidopsis thaliana* TAIR10), two monocots (*Oryza sativa* 7.0 and *Zea mays* 6a) and the moss *Physcomitrella patens* 3.0. No mismatches were allowed in identifying the corresponding homologous sequences.

Weighted correlation network analysis of microRNAs

The weighted correlation networks were constructed with the WGCNA package for R [65, 103] following the automated one-step protocol and with the default parameters except the minimum module size of 3 miRNAs in order to discover the smallest networks. We used normalized expression data of the 185 mature miRNAs according to the formula: (miRNA read number * 1.000.000)/total mapped reads per library. The eigengene value was calculated for each module and networks in order to test the significant association with the nodulation trait (the difference between the nodule library and the other libraries). Then, the networks were drawn using Cytoscape [104].

PhasiRNA identification

Bowtie alignments [105] without mismatch were used to predict the phasiRNAs. Phasing score [71] was calculated for each 21 nt read mapped to the genome with a threshold of 15 as mentioned in Zhai *et al.*, [71]. To eliminate the contribution of random fragments generated from RNA

degradation products a chi-test was performed per locus, taking into account the number of reads in phase and the reads not in phase ($p < 0.01$). Finally, a locus was predicted as a phasiRNA if it passed the above filters and had 3 or more 21 nt windows with a number of reads in the most abundant 5 % and genome mapped reads in at least one library. All loci located in transposable elements were removed. Once the final phasiRNA-generating loci were determined, they were tested for the identification of phase-triggering microRNAs/phasiRNA-targeted transcripts using psRNATarget [60] with default parameters and a limit of expectation of 5.

Sample preparation and expression analysis

Total RNA was isolated from 100 mg of frozen roots and nodules (21 dpi) from three biological replicates using Trizol reagent (Life Technologies) following the manufacturer's instructions. cDNA was synthesized from 2 μg of total RNA using RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas). Resulting cDNAs were then diluted and used to perform qRT-PCR assays using SYBR Green PCR Master Mix (Applied Biosystems), following the manufacturer's instructions. The sequences of oligonucleotide primers used are provided in Additional file 8: Table S5. Reactions were analyzed in a real-time thermocycler Applied Biosystem 7300. Two technical replicates were performed for each reaction. Relative expression was calculated with the "comparative Ct method" and normalized with the geometrical mean of three housekeeping genes (HSP, MDH & Ubiquitin9) [106].

Availability of supporting data

The RNA-seq data sets supporting the results of this study are available in the Gene Expression Omnibus (GEO) database, under accession GSE67433.

Additional files

Additional file 1: Table S1. List of the identified microRNA precursors and their corresponding mature forms. The blue, yellow and red lines represent the known miRNAs, the new isomiRs and the novel miRNAs, respectively. The "All precursors" sheet lists all the precursors identified that fulfill our criteria. The "All matures nr" sheet lists all the non-redundant mature miRNAs resulting from the "All precursors" list. The "Families" sheet lists all the miRNA families identified in our study and the corresponding number of members found.

Additional file 2: Table S2. List of the miRNAs and the corresponding number of reads present in the different organs. The blue, yellow and red lines represent the known miRNAs, the new isomiRs and the novel miRNAs, respectively. The absence or presence of a "1" in the presence columns indicate the absence or the presence of the corresponding miRNA (lines) in the corresponding library (rows).

Additional file 3: Table S3. List of miRNA degradome and psRNATarget-predicted targets. The blue, yellow and red lines represent the targets of the known miRNAs, the new isomiRs and the novel miRNAs, respectively. The "Degradome targets", "Degradome targets rejected miRs", "psRNATarget-predicted targets" and "psRNATarget target reject" sheets list the

targets predicted by degradome data for the selected and rejected miRNAs, and the targets predicted by psRNATarget for the selected and rejected miRNAs, respectively.

Additional file 4: Figure S1. Distribution of sequencing reads from soybean nodule libraries mapped on the soybean precursor of miRNov153. (A) Minimum free energy structure prediction of miRNov153 precursor of soybean. Heat colors represent the base-pair probabilities for each nucleotide (Blue = 0; Red = 1). The brackets show the position of miRNov153 mature and star. (B) Visualization of the mapped read distribution on the miRNov153 soybean precursor. miRNA reference sequence is at the top. The oriented grey bars represent the mapped reds. The brackets show the position of miRNov153 mature and star.

Additional file 5: Table S4. List of the identified PHAS genes and their distribution in 5 plant organs. The presence of a "0" or "1" in the libraries columns indicates the absence or presence of the corresponding PHAS locus (lines) in the corresponding library (rows), respectively. The conservation column indicates the initials of the legume species in which the PHAS locus is conserved (Gma = *Glycine max*, Mtr = *Medicago truncatula*).

Additional file 6: Figure S2. Venn diagram of the conservation of the PHAS genes in 3 legume species. Venn diagram of the distribution of PHAS genes based on their presence in three legumes: *Medicago truncatula* (red), *Glycine max* (yellow) and *Phaseolus vulgaris* (blue). The numbers in each Venn diagram area correspond to the numbers of *PHAS* genes encountered in the corresponding overlapping species area.

Additional file 7: Figure S3. Distribution of the identified PHAS genes in 5 plant organs. Venn diagram of the distribution of PHAS gene expression in the different studied organs. Red: flower, green: leaf, yellow: root, blue: seedling and black: nodule. The numbers in each Venn diagram area correspond to the numbers of expressed phasiRNA loci encountered in the corresponding overlapping organ area.

Additional file 8: Table S5. List of oligonucleotide sequences used in quantification experiments.

Abbreviations
5'RACE: 5'-rapid amplification of cDNA ends; AGO: Argonaute; bp: Base pair; Cv: Cultivar; dsRNA: Double-strand RNA; GO: Gene Ontology; kcal: Kilocalorie; miRNA: microRNA; mol: Molar; MYB: Myeloblastosis; NB-LRR: Nucleotide Binding domain (NB) and a Leucine Rich Repeat (LRR) domain; ncRNA: Non-coding RNA; NSP: Nodulation Signaling Pathway; nt: Nucleotide; PPR: Pentatrico peptide repeat; RISC: RNA-induced silencing complex; rRNA: Ribosomal RNA; TasiRNA: Trans-acting siRNAs; tRNA: Transfer RNA; UTR: Untranslated region.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
DF performed the miRNA analyses, supervised all the analyses, the data interpretation, and wrote the manuscript. LPI performed the phasiRNA analyses, participated in data interpretation and contributed to the drafting of the manuscript. FS and PP produced the sRNA libraries data. JLR, YFL and RS produced the degradome data and JLR contributed to the drafting of the manuscript. GH conceived and designed the whole project, contributed to the drafting of the manuscript and gave final approval of the version to be published. All authors read and approved the final manuscript.

Author details
[1]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México (UNAM), Av. Universidad 1001, Cuernavaca 62210, Morelos, Mexico. [2]Departamento de Biología Molecular de Plantas, Instituto de Biotecnología (UNAM), Av. Universidad 2001, Cuernavaca 62210, Morelos, Mexico. [3]Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK 74078, USA.

References
1. Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J. Beans (Phaseolus spp.) – model food legumes. Plant Soil. 2003;252:55–128.
2. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46:707–13.
3. Kalavacharla V, Liu Z, Meyers BC, Thimmapuram J, Melmaiee K. Identification and analysis of common bean (Phaseolus vulgaris L.) transcriptomes by massively parallel pyrosequencing. BMC Plant Biol. 2011;11:135.
4. Liu Z, Crampton M, Todd A, Kalavacharla V. Identification of expressed resistance gene-like sequences by data mining in 454-derived transcriptomic sequences of common bean (Phaseolus vulgaris L.). BMC Plant Biol. 2012;12:42.
5. Hernández G, Valdés-López O, Ramírez M, Goffard N, Weiller G, Aparicio-Fabre R, et al. Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. Plant Physiol. 2009;151:1221–38.
6. Hiz MC, Canher B, Niron H, Turet M. Transcriptome analysis of salt tolerant common bean (Phaseolus vulgaris L.) under saline conditions. PLoS One. 2014;9:e92598.
7. Hernández G, Ramírez M, Valdés-López O, Tesfaye M, Graham MA, Czechowski T, et al. Phosphorus stress in common bean: root transcript and metabolic responses. Plant Physiol. 2007;144:752–67.
8. Valdés-López O, Yang SS, Aparicio-Fabre R, Graham PH, Vance CP, et al. MicroRNA expression profile in common bean (Phaseolus vulgaris) under nutrient deficiency stresses and manganese toxicity. New Phytol. 2010;187:805–18.
9. Ramírez M, Flores-Pacheco G, Reyes JL, Luzlvarez A, Drevon JJ, Girard L, et al. Two Common Bean Genotypes with Contrasting Response to Phosphorus Deficiency Show Variations in the microRNA 399-Mediated PvPHO2 Regulation within the PvPHR1 Signaling Pathway. Int J Mol Sci. 2013;14:8328–44.
10. Naya L, Paul S, Valdés-López O, Mendoza-Soto AB, Nova-Franco B, Sosa-Valencia G, et al. Regulation of copper homeostasis and biotic interactions by microRNA 398b in common bean. PLoS One. 2014;9:e84416.
11. Contreras-Cubas C, Rabanal FA, Arenas-Huertero C, Ortiz MA, Covarrubias AA, Reyes JL. The Phaseolus vulgaris miR159a precursor encodes a second differentially expressed microRNA. Plant Mol Biol. 2012;80:103–15.
12. Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol. 2006;57:19–53.
13. Li S-C, Chan W-C, Hu L-Y, Lai C-H, Hsu C-N, Lin W. Identification of homologous microRNAs in 56 animal genomes. Genomics. 2010;96:1–9.
14. Lin W-C, Li S-C, Lin W-C, Shin J-W, Hu S-N, Yu X-M, et al. Identification of microRNA in the protist Trichomonas vaginalis. Genomics. 2009;93:487–93.
15. Zhang Y-Q, Chen D-L, Tian H-F, Zhang B-H, Wen J-F. Genome-wide computational identification of microRNAs and their targets in the deep-branching eukaryote Giardia lamblia. Comput Biol Chem. 2009;33:391–6.
16. Voinnet O. Origin, biogenesis, and activity of plant microRNAs. Cell. 2009;136:669–87.
17. Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell. 2004;116:281–97.
18. Combier J-P, Frugier F, de Billy FF, Boualem A, El-Yahyaoui F, Moreau S, et al. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. Genes Dev. 2006;20:3084–8.
19. Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL. Regulation of phosphate homeostasis by microRNA in Arabidopsis. Plant Cell. 2006;18:412–21.
20. Meng Y, Ma X, Chen D, Wu P, Chen M. MicroRNA-mediated signaling involved in plant root development. Biochem Biophys Res Commun. 2010;393:345–9.

Formey *et al. BMC Genomics* (2015) 16:423

Page 16 of 17

21. Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, et al. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science (80). 2006;312:436–9.

22. Staiger D, Korneli C, Lummer M, Navarro L. Emerging role for RNA-based regulation in plant immunity. New Phytol. 2013;197:394–404.

23. Eckardt NA. The plant cell reviews aspects of microRNA and PhasiRNA regulatory function. Plant Cell. 2013;25:2382.

24. Allen E, Xie ZX, Gustafson AM, Carrington JC. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell. 2005;121:207–21.

25. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, et al. Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell. 2004;16:69–79.

26. Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, et al. Roles of DCL4 and DCL3b in rice phased small RNA biogenesis. Plant J. 2012;69:462–74.

27. Fei Q, Xia R, Meyers BC. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. Plant Cell. 2013;25:2400–15.

28. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68–73.

29. Peláez P, Trejo MS, Iñiguez LP, Estrada-Navarrete G, Covarrubias AA, Reyes JL, et al. Identification and characterization of microRNAs in Phaseolus vulgaris by high-throughput sequencing. BMC Genomics. 2012;13:83.

30. Han J, Xie H, Kong ML, Sun QP, Li RZ, Pan JB. Computational identification of miRNAs and their targets in Phaseolus vulgaris. Genet Mol Res. 2014;13:310–22.

31. Arenas-Huertero C, Pérez B, Rabanal F, Blanco-Melo D, De la Rosa C, Estrada-Navarrete G, et al. Conserved and novel miRNAs in the legume Phaseolus vulgaris in response to stress. Plant Mol Biol. 2009;70:385–401.

32. Ormeño-Orrillo E, Menna P, Almeida LGP, Ollero FJ, Nicolás MF, Pains Rodrigues E, et al. Genomic basis of broad host range and environmental adaptability of Rhizobium tropici CIAT 899 and Rhizobium sp. PRF 81 which are used in inoculants for common bean (Phaseolus vulgaris L.). BMC Genomics. 2012;13:735.

33. Yang X, Li L. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics. 2011;27:2614–5.

34. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A-L, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 2008;18:610–21.

35. Jones-Rhoades MW. Conservation and divergence in plant microRNAs. Plant Mol Biol. 2012;80:3–16.

36. Li H, Deng Y, Wu T, Subramanian S, Yu O. Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. Plant Physiol. 2010;153:1759–70.

37. Bustos-Sanmamed P, Bazin J, Hartmann C, Crespi M, Lelandais-Briere C, Lelandais-Brière C. Small RNA pathways and diversity in model legumes: lessons from genomics. Front Plant Sci. 2013;4:236.

38. Zhu Q-H, Spriggs A, Matthew L, Fan L, Kennedy G, Gubler F, et al. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res. 2008;18:1456–65.

39. Lertpanyasampatha M, Gao L, Kongsawadworakul P, Viboonjun U, Christin H, Liu R, et al. Genome-wide analysis of microRNAs in rubber tree (Hevea brasiliensis L.) using high-throughput sequencing. Planta. 2012;236:437–45.

40. Xia R, Meyers BC, Liu ZZ, Beers EP, Ye S, Liu ZZ. MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA Biogenesis in Eudicots. Plant Cell. 2013;25:1555–72.

41. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, et al. High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One. 2007;2:e219.

42. Cuperus JT, Fahlgren N, Carrington JC. Evolution and functional diversification of MIRNA genes. Plant Cell. 2011;23:431–42.

43. Turner M, Yu O, Subramanian S. Genome organization and characteristics of soybean microRNAs. BMC Genomics. 2012;13:169.

44. Formey D, Sallet E, Lelandais-Briere C, Ben CC, Bustos-Sanmamed P, Niebel A, et al. The small RNA diversity from Medicago truncatula roots under biotic interactions evidences the environmental plasticity of the miRNAome. Genome Biol. 2014;15:457.

45. O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, et al. An RNA-Seq based gene expression atlas of the common bean. BMC Genomics. 2014;15:866.

46. Addo-Quaye C, Snyder JA, Park YB, Li Y-F, Sunkar R, Axtell MJ. Sliced micro-RNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. RNA. 2009;15:2112–21.

47. Meng Y, Gou L, Chen D, Wu P, Chen M. High-throughput degradome sequencing can be used to gain insights into microRNA precursor metabolism. J Exp Bot. 2010;61:3833–7.

48. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, et al. Criteria for annotation of plant MicroRNAs. Plant Cell. 2008;20:3186–90.

49. Axtell MJ. Evolution of microRNAs and their targets: are all microRNAs biologically relevant? Biochim Biophys Acta. 2008;1779:725–34.

50. Zhuo Y, Gao G, Shi JA, Zhou X, Wang X. miRNAs: biogenesis, origin and evolution, functions on virus-host interaction. Cell Physiol Biochem. 2013;32:499–510.

51. Colaiacovo M, Lamontanara A, Bernardo L, Alberici R, Crosatti C, Giusti L, et al. On the complexity of miRNA-mediated regulation in plants: novel insights into the genomic organization of plant miRNAs. Biol Direct. 2012;7:15.

52. Axtell MJ, Westholm JO, Lai EC. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 2011;12:221.

53. Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA. 2005;11:241–7.

54. Elrouby N, Bureau TE. Bs1, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize. Plant Physiol. 2010;153:1413–24.

55. Ma Z, Coruh C, Axtell MJ. Arabidopsis lyrata small RNAs: transient MIRNA and small interfering RNA loci within the Arabidopsis genus. Plant Cell. 2010;22:1090–103.

56. Li B, Duan H, Li J, Deng XW, Yin W, Xia X. Global identification of miRNAs and targets in Populus euphratica under salt stress. Plant Mol Biol. 2013;81:525–39.

57. Sun F, Guo G, Du J, Guo W, Peng H, Ni Z, et al. Whole-genome discovery of miRNAs and their targets in wheat (Triticum aestivum L.). BMC Plant Biol. 2014;14:142.

58. Xia R, Zhu H, An Y-Q, Beers EP, Liu Z. Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol. 2012;13:R47.

59. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol. 2008;18:758–62.

60. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011;39(Web Server issue):W155–9.

61. Meyers BC, Green PJ. Plant MicroRNAs: Methods and Protocols, Humana pre. John M. Walker: Hatfield; 2010.

62. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP. Prediction of plant microRNA targets. Cell. 2002;110:513–20.

63. Brandt R, Cabedo M, Xie Y, Wenkel S. Homeodomain leucine-zipper proteins and their role in synchronizing growth and development with the environment. J Integr Plant Biol. 2014;56:518–26.

64. Boualem A, Laporte P, Jovanovic M, Laffont C, Plet J, Combier J-P, et al. MicroRNA166 controls root and nodule development in Medicago truncatula. Plant J. 2008;54:876–87.

65. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

66. Valdés-López O, Arenas-Huertero C, Ramírez M, Girard L, Sánchez F, Vance CP, et al. Essential role of MYB transcription factor: PvPHR1 and microRNA: PvmiR399 in phosphorus-deficiency signalling in common bean roots. Plant Cell Environ. 2008;31:1834–43.

67. Liang G, He H, Yu D. Identification of nitrogen starvation-responsive microRNAs in Arabidopsis thaliana. PLoS One. 2012;7:e48951.

68. Olivera M, Tejera N, Iribarne C, Ocaña A, Lluch C. Effect of phosphorous on nodulation and nitrogen fixation by Phaseolus vulgaris. In: Velázquez E, Rodríguez-Barrueco C, editors. First Int Meet Microb Phosphate Solubilization, vol. Volume 102. Dordrecht: Springer Netherlands; 2007. p. 157–60. Developments in Plant and Soil Sciences.

69. Awonaike KO, Lea PJ, Day JM, Roughley RJ, Miflin BJ. Effects of Combined Nitrogen on Nodulation and Growth of Phaseolus vulgaris. Exp Agric. 2008;16:303.

70. Bari R, Pant BD, Stitt M, Scheible W-R. PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiol. 2006;141:988–99.

71. Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, Li Y, et al. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev. 2011;25:2540–53.

72. Brocker C, Vasiliou M, Carpenter S, Carpenter C, Zhang Y, Wang X, et al. Aldehyde dehydrogenase (ALDH) superfamily in plants: gene nomenclature and comparative genomics. Planta. 2013;237:189–210.

Formey *et al. BMC Genomics* (2015) 16:423

Page 17 of 17

73. Yoshida S, Kameoka H, Tempo M, Akiyama K, Umehara M, Yamaguchi S, et al. The D3 F-box protein is a key component in host strigolactone responses essential for arbuscular mycorrhizal symbiosis. New Phytol. 2012;196:1208–16.

74. Takahara M, Magori S, Soyano T, Okamoto S, Yoshida C, Yano K, et al. Too much love, a novel Kelch repeat-containing F-box protein, functions in the long-distance regulation of the legume-Rhizobium symbiosis. Plant Cell Physiol. 2013;54:433–47.

75. Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, et al. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell. 2014;26:4584–601.

76. Wang S, Sun X, Hoshino Y, Yu Y, Jia B, Sun Z, et al. MicroRNA319 positively regulates cold tolerance by targeting OsPCF6 and OsTCP21 in rice (Oryza sativa L.). PLoS One. 2014;9:e91357.

77. Zhou M, Luo H. Role of microRNA319 in creeping bentgrass salinity and drought stress response. Plant Signal Behav. 2014;9.

78. Schommer C, Bresso EG, Spinelli SV, Palatnik JF. Role of MicroRNA miR319 in Plant Development. In: Sunkar R, editor. MicroRNAs Plant Dev Stress Responses, vol. Volume 15. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 29–47. Signaling and Communication in Plants.

79. De Luis A, Markmann K, Cognat V, Holt DB, Charpentier M, Parniske M, et al. Two microRNAs linked to nodule infection and nitrogen-fixing ability in the legume Lotus japonicus. Plant Physiol. 2012;160:2137–54.

80. Ariel F, Brault-Hernandez M, Laffont C, Huault E, Brault M, Plet J, et al. Two direct targets of cytokinin signaling regulate symbiotic nodulation in Medicago truncatula. Plant Cell. 2012;24:3838–52.

81. Lauressergues D, Delaux P-MM, Formey D, Lelandais-Brière C, Fort S, Cottaz S, et al. The microRNA miR171h modulates arbuscular mycorrhizal colonization of Medicago truncatula by targeting NSP2. Plant J. 2012;72:512–22.

82. Yan Z, Hossain MS, Wang J, Valdés-López O, Liang Y, Libault M, et al. miR172 regulates soybean nodulation. Mol Plant Microbe Interact. 2013;26:1371–7.

83. Nova-Franco B, Íñiguez LP, Valdés-López O, Alvarado-Affantranger X, Leija A, Fuentes SI, et al. The miR172c-AP2-1 Node as a Key Regulator of the Common Bean - Rhizobia Nitrogen Fixation Symbiosis. Plant Physiol. 2015;114:255547.

84. Wang Y, Wang L, Zou Y, Chen L, Cai Z, Zhang S, et al. Soybean miR172c Targets the Repressive AP2 Transcription Factor NNC1 to Activate ENOD40 Expression and Regulate Nodule Initiation. Plant Cell. 2014;26:4782–801.

85. Zhang S, Wang Y, Li K, Zou Y, Chen L, Li X. Identification of Cold-Responsive miRNAs and Their Target Genes in Nitrogen-Fixing Nodules of Soybean. Int J Mol Sci. 2014;15:13596–614.

86. Oldroyd GED, Downie JA. Calcium, kinases and nodulation signalling in legumes. Nat Rev Mol Cell Biol. 2004;5:566–76.

87. Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, et al. Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell. 2008;133:128–41.

88. Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, et al. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/ DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell. 2007;19:926–42.

89. Visser M, van der Walt AP, Maree HJ, Rees DJG, Burger JT. Extending the sRNAome of apple by next-generation sequencing. PLoS One. 2014;9:e95782.

90. Liu Y, Wang Y, Zhu Q-H, Fan L. Identification of phasiRNAs in wild rice (Oryza rufipogon). Plant Signal Behav. 2013;8.

91. Manavella PA, Koenig D, Weigel D. Plant secondary siRNA production determined by microRNA-duplex structure. Proc Natl Acad Sci U S A. 2012;109:2461–6.

92. Liu H, Wang X, Zhang H, Yang Y, Ge X, Song F. A rice serine carboxypeptidase-like gene OsBISCPL1 is involved in regulation of defense responses against biotic and oxidative stress. Gene. 2008;420:57–65.

93. De Paoli E, Dorantes-Acosta A, Zhai J, Accerbi M, Jeong D-H, Park S, et al. Distinct extremely abundant siRNAs associated with cosuppression in petunia. RNA. 2009;15:1965–70.

94. Xie Q, Frugis G, Colgan D, Chua NH. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. 2000;14:3024–36.

95. Broughton WJ, Dilworth MJ. Control of leghaemoglobin synthesis in snake beans. 1971.

96. German MA, Luo S, Schroth G, Meyers BC, Green PJ. Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc. 2009;4:356–62.

97. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36:W5–9.

98. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.

99. Addo-Quaye C, Miller W, Axtell MJ. CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics. 2009;25:130–1.

100. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.

101. Zhi-Liang H, Bao J, Reecy J. CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. Online J Bioinforma. 2008;9:108–12.

102. Edwards AWF. Cogwheels of the Mind: The Story of Venn Diagrams. JHU: JHU Press; 2004.

103. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2009.

104. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

105. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

106. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002, 3.