



UNIVERSIDAD NACIONAL DE LA PLATA  
FACULTAD DE CIENCIAS EXACTAS  
DEPARTAMENTO DE CIENCIAS BIOLÓGICAS



Trabajo de Tesis Doctoral:  
Identificación y caracterización de pequeños ARNs  
regulatorios en la interacción simbiótica entre  
*Phaseolus vulgaris* y *Rhizobium etli*.

Tesista : Méliisse Castaingts  
Directora : María Eugenia Zanetti  
Año : 2018



**UNIVERSIDAD NACIONAL DE LA PLATA**

**FACULTAD DE CIENCIAS EXACTAS**

**DEPARTAMENTO DE CIENCIAS BIOLÓGICAS**

---

**Trabajo de Tesis Doctoral:**

***Identificación y caracterización de pequeños ARNs  
regulatorios en la interacción simbiótica entre  
Phaseolus vulgaris y Rhizobium etli.***

---

**Tesista: *Mélisse Castaingts***

**Directora: *María Eugenia Zanetti***

**Año: *2018***

**El presente trabajo de Tesis para optar al grado de Doctor de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata ha sido realizado en el Instituto de Biotecnología y Biología Molecular, en el grupo Laboratorio de Biología de la Raíz, bajo la dirección de la Dra. María Eugenia Zanetti.**

Durante el transcurso de esta tesis doctoral tuve la desgracia de perder a un amigo querido en mi país. Dedico este manuscrito a Alexy, y recuerdo la suerte que tenemos todos aquí presentes: que nadie nos haya quitado el derecho de vivir.

*Nous nous étions rencontrés dans les couloirs de la cité universitaire, peut-être bien que j'avais frappé à ta porte pour te demander un briquet ; au fond de moi l'espoir naissant de faire une belle rencontre. Avec ton grand sourire sans doute, tu m'as ouvert les portes de ta chambre et de ton cœur. De voisins nous sommes très vite passés à amis et confidents, vadrouilleurs et aventuriers. Nous avons rapidement établi un rituel basé sur le sport et la détente. Autour de ta théière sacrée fusaient des discussions philosophiques où nos idées se croisaient et nos opinions s'entremêlaient. Toujours à la recherche d'un monde meilleur, d'une vie saine et plaisante, dans le respect de la nature. Tu savais profiter du moment présent et tu envoyais en toutes circonstances ton énergie positive. Nous n'avons qu'une vie, et tu le savais très bien. La vie, ce jeu où personne ne nous a demandé notre avis, n'est pas toujours juste. Comme partout, il y en a qui trichent : les menteurs, les voleurs et les tueurs. Nous nous sommes menti lors de notre dernier au revoir. Ton assassin nous a volé nos retrouvailles.*

*Mais pourquoi se dire adieu si tu es toujours présent dans mon cœur ?*

À Alexy

# *Agradecimientos*

*A Eugenia. La historia con Eugenia empezó desde lejos, en el momento en que me dio su confianza al hacerme un lugar en su laboratorio para realizar una pasantía, brindándome su tiempo y sus conocimientos para enseñarme a trabajar en la Biotecnología de plantas. Me formó, me permitió conocer una cultura distinta, en otro continente, con valores diferentes; los cuales son los que me hacen feliz ahora. Volví más grande a mi país, tanto en saber cómo en madurez, en experiencia laboral y de vida. Claramente cuatro meses fueron pocos y pasaron demasiado rápido. Sentí que en Argentina mi vida recién empezaba. Entonces juntas nos presentamos a una beca doctoral. Ambas sabíamos que esto sería difícil, tal vez más que para cualquier otro estudiante de doctorado; tantos trámites, otro idioma y diferencias culturales... pero es frente a la dificultad que uno crece y se hace más fuerte.*

*Eligio un proyecto de doctorado en total armonía con mi personalidad, como si me hubiese conocido más que mi familia o yo a mí misma. Un plan de trabajo novedoso, ambicioso, complejo, diverso y a su vez divertido. El proyecto era como una puerta de madera cerrada, no se veía lo que podría ocultar. Algo de miedo me generó, pero ella sabía que me gustaban los desafíos y lo desconocido. Tenía que encontrar la llave, algo que no me fue muy sencillo... ¡De mi impaciencia hubiera derrumbado la puerta! Pero ella me acercó un manojo de llaves, y una vez la llave correcta en la mano, pudimos abrir la puerta. Ahí encontramos un espacio, oscuro pero brillante, sin límite, flotando entre las estrellas. Con mi carácter totalmente disperso me dejó divertirme un rato con esta sucesión de cuatro letras que son las secuencias. Sin sentido de la orientación ni de razón me hubiese quedado ahí a la deriva para siempre, recorriendo un planeta tras otro, quemándome con los rayos del sol o asfixiándome por la falta de aire. Ella siempre con un cable a tierra me observaba curiosa mientras yo buscaba la puerta de salida. Claramente tuvo que dibujarme un camino para que este juego llegue a su fin. Un fin feliz y luminoso, como es ella y como tendríamos que ser todos.*

*Además de su dirección, su enseñanza y su paciencia, sentí desde el principio que me dio su amistad y su afecto. Tal vez se encariñó por tener bajo su dirección a una persona que estaba lejos de su familia, o porque en algunos casos me vio un poco perdida, buscando mi camino como en una carrera, el de la Ciencia o el de la vida, en este hermoso país que es Argentina. También supo encontrar la medida justa entre mostrarme el camino correcto, acompañarme en éste y dejarme recorrerlo sola. Obviamente, hubo varias caídas. Pero siempre encontré su mano para ayudarme a levantarme y seguir para adelante. Una vez más me demostró su confianza. No era autónoma y no quería serlo, pero ella me obligó y sé que le costó. Lo hizo por mí. Siempre fue muy generosa conmigo, nunca me negó vacaciones para ir a conocer una provincia, o para competir en una carrera. Supo retarme, darme ánimo, reconfortarme y felicitarme cuando hizo falta. Siempre me cocinó mis postres preferidos en las reuniones del trabajo o para mis cumpleaños. No se comportó como una directora, sino como una amiga cuando compartimos un entrenamiento de pileta o una carrera de aguas abiertas, y como una mamá me recibió en su casa, me alojó y me llevó a una carrera un día de lluvia; además del cariño que me dio cada día. Eugenia es simplemente todo. Y con los sentimientos los más sinceros, por eso le agradezco.*

Al Doctor Antonio Lagares y al Doctor Mario Aguilar, al Instituto de Biotecnología y Biología Molecular (IBBM), a la Facultad de Ciencias Exactas y a la Universidad Nacional de La Plata (UNLP) por brindarme el espacio físico, equipos e instalaciones necesarios para el desarrollo de este trabajo.

A la CIC y CONICET por el otorgamiento de becas que posibilitaron la realización de la tesis. A Migraciones por el otorgamiento de las residencias temporarias y permanente. Al país por la belleza de su naturaleza, y a sus residentes por sus valores.

A mi grupo de trabajo del Laboratorio de Biología de la Raíz: a Eugenia y Flavio por la dirección, la profesionalidad, las recomendaciones y consejos, la amistad, el apoyo y las correcciones del manuscrito. A mis compañeros: Virginia, Carolina, Mauricio, Joaquín, Priscila, Efrén, Sole, Ana, Marcos, Carla, Claudio, Karen, Caro, Julieta, Mili, Cris, Mica, que me soportaron y acompañaron diariamente. Agradezco muchísimo a Mica B. por su paciencia, su dedicación y nuestra amistad.

A todos los investigadores y becarios del IBBM, especialmente a Florcita, Yuli, Carla, Mati, Fran, Agustina, Leti, Flor, la Peti, Facundo, Belén, Luli, Esteban, Vale, Laura, José... A Moni por su exotismo y su mente abierta, a lle por escucharme, aconsejarme y retarme. A Emy, por su apoyo y su cariño diario. Siempre recordaré a Gabo por su costillar al asador para el festejo de mi Residencia, y a todos los que participaron. Agradezco muchísimo a Eduardo por haberme incentivado a participar en mi primera carrera de aventura.

A todo el personal y los CPAs del IBBM: particularmente a Claudio, Paula, Silvana, Abel, Andrea, Gisela, Juliana y Ulises.

A mis amistades en Argentina y sus familias: a mis amigas de la pensión en 2011, en particular a Majo y a Eva. A mis amigas del Cindefi: Yani, Flor y Andrea. A Lucas por su presencia, dedicación y amistad, a Martín, Maguie, Seba, Emiliana, Ariel, Lina, Sebastián, Marcela, Sil y Néstor. A ellos que estuvieron dándome cariño, compañía y diversión.

A mis amigos de carreras y deportes varios: a Los Ángeles por los entrenamientos de fútbol y competencias en los Juegos de Conicet. A Fernando, Maxi y Ale por la generosidad de prestarme sus bicis. A Luis, Juan y Jime para la compañía en los entrenamientos y aventuras. Un gran cariño igualmente para mis competidores, que carrera tras carrera fueron compañeros y amigos: Nadia, Cande, Marquitos, Rebecca, Lucas, Dani, María Laura, Nelly, José, Diego, Cristian, Julio, Mirian, Lucha, Agustín, Gonza, Emilia, Araceli, Felix, Julia, Sole, Ceci y Facundo. Un cariño muy grande para Lorena, los hermanos Pozzi y Estani. Agradezco a los organizadores de carreras que me hicieron divertirse, exigir y crecer: Matías y Kayak Platense, Marcos y Andar Extremo, Eduardo Cresta, Damián Esteban, Hugo y el CAB, todos por ser profesionales apasionados. A mi profe de natación Ariel, al Hogar Social, a Lorenzo de Poseidón, a Gabo y al equipo Moebius de Ultimate. A los guardavidas de Berisso y Ensenada, a los de la pileta del Nacional, al Rocódromo.

A mis amigos de inglés por las lindas juntadas, a mi profesor Fede, a Ana y Francy, y a Hannah por su forma didáctica de ayudarme en la comprensión del mundo.

A todos los que me ayudaron durante los viajes, por su amistad, el alojamiento y el transporte, por compartir conmigo una cena, un entrenamiento o una carrera. Especialmente a Jorge, a la gente de Jujuy, al Padre Joaquín de La Cumbre y a su escuela perdida en la sierra.

A mis amigos de Mar del Plata y a los que conocí en esa hermosa ciudad: Ane, Priscila, Maguie, María Inés, Silvia, Leo, Hernán, Niko.

A los artistas que me emocionaron con su arte y su personalidad: Leo Gauna, Luxor, Simón Jatip, Gustako Cornejo, Frank Zarate y Cristian. Agradezco especialmente a Sofía de "Sin Título Dimensión Autor" por organizar hermosos eventos, y a los muralistas que pusieron amor y color en la ciudad de La Plata. A los músicos Vicente de "Sudor Marika" y a Yuli de "Cachitas Now" por hacerme soñar y bailar. A Barfuss, Cabra y Juan Merino por brindarnos con pasión las mejores cervezas artesanales de La Plata.

Agradezco a los que tomaron el papel de familia en Argentina: a mi familia de Buenos Aires por su amor, por compartir fines de semana, cumpleaños y asados. A mi amiga Yamila, a su hermana y hermano, a los sobrinos, a Papi Rubén, y a Mami Isabel. A César, Yola, Gaby, Pati y al Principito Santino, por compartir conmigo su casa, su comida y el ambiente caluroso de su familia.

Y del otro lado del océano agradezco al estado francés y sus tres lemas de Libertad, Igualdad y Fraternidad por los valores que me brindó y con los cuales crecí. Al sistema educativo francés por su calidad y rigurosidad y a las becas

otorgadas que me permitieron estudiar. A la facultad de Ciencias y Técnicas de la Universidad de Montpellier (UM2), especialmente al Master Biología Funcional de Plantas. A sus profesores por alimentar mi pasión por la ciencia, especialmente a Guilhem Desbrosses, Fabrice Varoquaux, Eric Lacombe y Sergio Swistoonoff, quienes me hicieron llegar hasta Argentina; y a los estudiantes que me acompañaron. Agradezco a mi tío Louis Briquet, quien incentivó mi pasión por la investigación desde muy joven, y me apoyó desde el comienzo de mis estudios universitarios hasta el día de hoy.

Agradezco también a mis profesores de MTB de la facultad, Daniel y Jean François Marie, quienes tienen un papel clave en el sueño que pude cumplir aquí en Argentina: ser triatleta. Agradezco a mis amistades del período universitario: Alexy, Anastasia, Benoit, Chloé, Edith, Elodie, Emilie, Hélène, Pauline...

Sin embargo la pasión por la Biología me fue transmitida por los profesores del Lyceo Déodat de Séverac en Céret: la pareja Megevand, y el profesor y amigo François Masson, a quien agradezco muchísimo por todas las aventuras extra-escolares que pudimos compartir y compartiremos nuevamente. Sin duda mi llegada a Argentina fue incentivada también por los profesores de castellano durante mi escolaridad, especialmente por Marie Christine Marty.

Quiero agradecer a mis padres Brigitte y Pedro, por darme la oportunidad de vivir y crecer en un ambiente hermoso rodeado de naturaleza. Les agradezco infinitamente porque si no me dieron la vida que yo esperaba (cosa muy difícil porque era una utopía), sí que supieron proporcionarme las herramientas para que yo fuera a buscarla. Les agradezco por su apertura de espíritu y por haberme transmitido buenos valores más que códigos sociales superficiales. También por haber sido buenos y malos sin mala intención, en un equilibrio imperfecto y desinteresado. Dejaron nacer y crecer en mí el deseo de irme y volver, partir de nuevo para luego regresar, en una ida y vuelta sin fin.

Agradezco a mis hermanas, a quienes quiero más que todo: Cyrille y Florie, cada una con su particularidad, en nuestra complicidad y amor. Les agradezco por haber compartido conmigo momentos de mi vida en Argentina. Agradezco a mis primas y primos que también fueron amigos y sostén, particularmente a Fanny y Aurélien, a Dédé y Marine, a Sophie y Thomas y a Maïe y Grégoire. Una mención especial para Louis y Rosie; Véra, Laurent y Zeilani, quienes me visitaron en Argentina. Agradezco a todos los primos del país vasco, a los primitos hermosos que me llenaron de amor y esperanza de ambos lados de los Pirineos y todo el resto de mi familia catalana-vasca. Agradezco a mis amigos franceses: Brice, Olivier, Cédric, Agathe y Pauline. Con mucho cariño agradezco a «Les Chattes»: Martha, Marlène, Maéva, Christelle, Sophie, Eva, Marion y Alicia, con quienes he compartido momentos inolvidables. Agradezco a los que viajaron hasta Argentina para conocer su encanto.

Para terminar agradezco a Mika, mi actual y futuro compañero de carreras y aventuras, recién llegado en mi vida y a su vez tanto tiempo esperado. Gracias por tu comprensión y tu apoyo, tu amor incondicional, y por compartir nuestra “non-vie fausse-vraie vie”. Gracias por creer en nuestro proyecto Entre deux Mers; la culminación de esta pasión que nos consume y enloquece, tanto como nos mantiene vivos.

**Como los pequeños ARNs, todos ustedes son individuos sencillos y numerosos, cada uno con papeles claves en mi vida. Se interconectaron en redes de regulaciones complejas a través de su apoyo, presencia, ayuda, amistad, amor, tiempo, dedicación y momentos compartidos. Influyeron mi humor, mi bienestar, mis pasiones, y la felicidad que pude sentir durante estos seis años en Argentina. Esto sin dudas se ha reflejado en la finalización de mi trabajo, y desde entonces el éxito de esta estadía a nivel profesional y personal. Aprendí de todos ustedes, sobre la ciencia, la investigación, el deporte, el idioma, la cultura, la historia, la gastronomía, las relaciones sociales, los valores y sobre mí. Me acompañaron en el descubrimiento de mi personalidad confusa y dispersa y me ayudaron a fortalecerla y enriquecerla. Me identificaron y aceptaron tal cual soy, “auténtica y completa, exigente e insatisfecha, insoportable y adorable”. Desde el corazón les agradezco.**

# Índice general

<b>Resumen</b>	<b>1</b>
<b>Introducción</b>	<b>3</b>
<b>Hipótesis y Objetivos</b>	<b>32</b>
<b>Capítulo I: Resultados</b>	<b>34</b>
<b>Capítulo I: Discusión</b>	<b>70</b>
<b>Capítulo II: Resultados</b>	<b>78</b>
<b>Capítulo II: Discusión</b>	<b>91</b>
<b>Capítulo III: Resultados</b>	<b>96</b>
<b>Capítulo III: Discusión</b>	<b>109</b>
<b>Conclusiones y Perspectivas</b>	<b>116</b>
<b>Materiales y Métodos</b>	<b>118</b>
<b>Anexos</b>	<b>148</b>
<b>Bibliografía</b>	<b>163</b>



# Índice temático

<b>Resumen</b>	<b>1</b>
<b>Introducción</b>	<b>3</b>
<b>1. La simbiosis fijadora de nitrógeno y su importancia en la agricultura</b>	<b>3</b>
1.1 Importancia de las leguminosas en la agricultura	3
1.2 Importancia del poroto en la alimentación y salud humana	4
<b>2. Interacción simbiótica entre leguminosas y rizobios</b>	<b>6</b>
2.1 Formación de nódulos funcionales: de la infección bacteriana a la fijación del nitrógeno	6
2.2 Simbiosis y regulación hormonal	7
2.3 Plantas modelos, secuenciación y aplicación	8
2.4 Simbiosis: diálogo molecular inicial	10
2.5 Señalización temprana en la simbiosis fijadora de nitrógeno	12
2.6 Simbiosis y evolución	15
<b>3. Relación entre simbiosis y respuestas de defensa</b>	<b>17</b>
<b>4. La interacción entre <i>Phaseolus vulgaris</i> y <i>Rhizobium etli</i></b>	<b>19</b>
4.1 Asociación preferencial entre variedades de <i>P. vulgaris</i> y su simbionte <i>R. etli</i>	19
4.2 Competitividad de los rizobios del suelo	21
4.3 Genes de <i>P. vulgaris</i> implicados en la preferencia de cepa	22
<b>5. Pequeños ARNs no codificantes</b>	<b>23</b>
5.1 Clasificación de los sARNs no codificantes	23
5.2 Biogénesis y modo de acción de los miARNs	25
5.3 Diferencias entre miARNs de plantas y animales	27
5.4 miARNs implicados en la simbiosis	29
<b>Hipótesis y Objetivos</b>	<b>32</b>
<b>Capítulo I: Bibliotecas de pequeños ARNs</b>	<b>34</b>
<b>1. Construcción de las bibliotecas de pequeños ARNs</b>	<b>34</b>
1.1 Generación de las muestras para las bibliotecas	34
1.2 Construcción de las bibliotecas de pequeños ARNs	37
1.3 Validación cuantificación y normalización de las bibliotecas	39
<b>2. Secuenciación de las bibliotecas y procesamiento de los datos</b>	<b>41</b>
<b>3. Clasificación de los pequeños ARNs por tamaño</b>	<b>46</b>
<b>4. Caracterización de los miARNs</b>	<b>47</b>
4.1 Identificación de miARNs utilizando el algoritmo miRCat	47
4.2 Distribución por tamaño de los miARNs	49
4.3 Cuantificación de miARNs	49
4.4 Identificación de miARNs con acumulación diferencial	51
4.5 Análisis de miARNs utilizando los genomas de referencia mesoamericano y andino	54
4.6 Análisis y cuantificación de miARNs conservados	55
4.7 Conservación de un miARN previamente asociado a la simbiosis, el miR171	57
<b>5. Caracterización de los phasiARNs</b>	<b>64</b>

<b>Capítulo I: Discusión</b>	<b>70</b>
Asignación de la función de los sARNs identificados	70
El miR167, un miARN conservado con acumulación diferencial	71
El miR1514a, una nueva isoforma identificada en <i>P. vulgaris</i>	72
El miR171a, un miARN de la respuesta simbiótica	73
Identificación de nuevos miARNs por secuenciación masiva	74
<b>Capítulo II: El miR390b y su vía de regulación en <i>P. vulgaris</i></b>	<b>78</b>
1. La vía de regulación del miR390b a 24 hpi	78
2. Caracterización funcional del miR390b	83
<b>Capítulo II: Discusión</b>	<b>91</b>
Conservación y evolución del miR390 en plantas	91
El miR390 como mediador de la regulación por auxinas en la nodulación	91
Nuevas funciones del miR390 en leguminosas	92
<b>Capítulo III: Caracterización de un nuevo miARN involucrado en la preferencia de <i>P. vulgaris</i> por cepas más eficientes en la nodulación</b>	<b>96</b>
1. Caracterización de la expresión del miR5924 maduro y de su precursor	96
2. Identificación <i>in silico</i> de transcritos blancos putativos del miR5924	98
3. Caracterización funcional del miR5924	102
<b>Capítulo III: Discusión</b>	<b>109</b>
El miR5924 un nuevo miARN de <i>P. vulgaris</i> procesado a partir de <i>DCL1</i>	109
El miR5924 y su función en el establecimiento de la simbiosis	111
El miR5924, y su posible función como modulador de las respuestas de defensa	112
<b>Conclusiones y Perspectivas</b>	<b>116</b>

<b>Material y Métodos</b>	<b>118</b>
<b>1. Medios de cultivos y soluciones</b>	<b>118</b>
1.1 Medios de cultivo	118
1.2 Soluciones	119
<b>2. Material biológico</b>	<b>121</b>
2.1 Semillas de poroto	121
2.2 Cepas bacterianas	121
2.3 Transformación bacteriana	122
<b>3. Condiciones de cultivo, transformación de plantas de <i>P. vulgaris</i> e inoculación con <i>R. etli</i></b>	<b>123</b>
3.1 Esterilización y germinación de las semillas	123
3.2 Crecimiento de plántulas	123
3.3 Crecimiento de <i>A. rhizogenes</i> K599	123
3.4 Obtención de plantas compuestas	123
3.5 Crecimiento de plantas compuestas	123
3.6 Observación de las raíces transgénicas	124
3.7 Inoculación de plantas de poroto con rizobios	125
3.8 Recolección de tejido de raíz e otros órganos	125
<b>4. Caracterización fenotípica de las plantas compuestas</b>	<b>126</b>
4.1 Cinéticas de nodulación	126
4.2 Determinación del tamaño de los nódulos	126
4.3 Ensayos de competencia	126
4.4 Análisis del polimorfismo del gen <i>nodC</i>	126
4.5 Caracterización de los eventos tempranos de infección rizobiana	127
<b>5. Oligonucleótidos y clonados</b>	<b>127</b>
5.1 Diseño de oligonucleótidos	127
5.2 Construcción de vectores mediante el sistema GATEWAY	129
<b>6. Técnicas comunes de biología molecular</b>	<b>130</b>
6.1 Reacciones de PCR	130
6.2 PCR de colonias bacterianas	131
6.3 Electroforesis de ADN y ARN en geles de agarosa	131
6.4 Purificación de ADN a partir de gel de agarosa	132
6.5 Purificación de ADN plasmídico	132
6.6 Extracción de ARN total de tejido vegetal	132
6.7 Tratamiento del ARN con DNAsa	132
6.8 Síntesis de ADN complementario	132
<b>7. Análisis de expresión</b>	<b>133</b>
7.1 Reacciones de PCR semicuantitativas	133
7.2 PCR cuantitativa en tiempo real (qRT-PCR)	133
<b>8. Técnica de <i>Stem-loop</i> PCR</b>	<b>133</b>
8.1 Diseño de oligonucleótidos de <i>stem-loop</i>	133
8.2 Transcripción reversa para los miARNs	134
8.3 Medición de la acumulación de los miARNs por qPCR	135

<b>9. Construcción de bibliotecas de pequeños ARNs</b>	<b>135</b>
9.1 Material biológico	135
9.2 Extracción y separación de ácidos ribonucleicos	136
9.3 Generación de bibliotecas de pequeños ARNs	138
9.4 Validación y secuenciación de las bibliotecas	142
<b>10. Análisis bioinformáticos</b>	<b>142</b>
10.1 Bases de datos de secuencias genómicas	142
10.2 Análisis de datos de secuenciación masiva	143
10.3 Análisis de secuencias nucleotídicas y proteicas	145
<b>Anexos</b>	<b>148</b>
<b>1. Figuras Suplementarias</b>	<b>148</b>
<b>2. Tablas Suplementarias</b>	<b>149</b>
<b>3. Secuencias de miARN, miARN* y precursores</b>	<b>158</b>
<b>4. Oligonucleótidos</b>	<b>161</b>
<b>Bibliografía</b>	<b>163</b>

# Resumen

*Phaseolus vulgaris* es una leguminosa de grano de amplio interés económico y social, la cual establece una interacción simbiótica con la bacteria *Rhizobium etli* que resulta en la formación de nódulos fijadores de nitrógeno. En esta interacción, las plantas de origen mesoamericano son noduladas preferencial y más eficientemente por cepas de rizobios que son predominantes en la misma región geográfica. Nuestro grupo de investigación ha identificado y caracterizado varios genes involucrados en esta asociación preferencial. En la presente Tesis doctoral nos planteamos como objetivo general identificar y caracterizar a nivel funcional aquellos pequeños ARNs (sARNs) que desempeñan una función en el establecimiento de esta simbiosis eficiente. Se construyeron y secuenciaron bibliotecas de sARNs de raíces inoculadas con cepas de rizobio de alta y baja eficiencia en la nodulación o con el medio de cultivo control utilizando la tecnología Illumina. Se obtuvieron en promedio 28 millones de lecturas por biblioteca, de las cuales aproximadamente el 90 % mapearon al genoma de *P. vulgaris*. A partir del análisis *in silico*, se identificaron microARNs (miARNs) conservados y nuevos que muestran cambios en su perfiles de acumulación en respuesta al rizobio. Se seleccionó un miARN conservado, el miR390b, el cual se acumula a mayores niveles en respuesta a la cepa más eficiente. En paralelo, se seleccionó y caracterizó funcionalmente un nuevo miARN diferencial, denominado miR5924, el cual es procesado a partir de la región no traducida del transcripto que codifica para *Dicer-like 1 (DCL1)*. Los resultados obtenidos sugieren que este nuevo miARN cumple una función clave en el establecimiento de la simbiosis. Los resultados obtenidos han permitido avanzar y profundizar el conocimiento acerca de las bases moleculares de la nodulación eficiente. A su vez, los miARNs caracterizados proveen nuevos candidatos a ser utilizados en futuros programas de mejoramiento genético que contribuyan a mejorar la fijación biológica de nitrógeno y la productividad del cultivo de *P. vulgaris*.

# *Introducción*

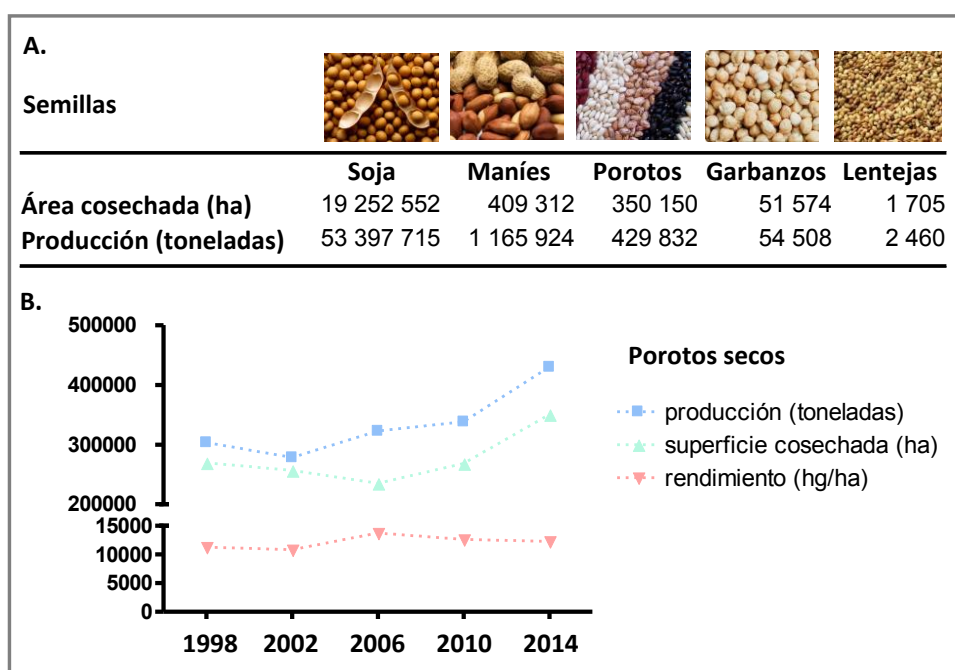
# 1. La simbiosis fijadora de nitrógeno y su importancia en la agricultura

## 1.1 Importancia de las leguminosas en la agricultura

La adquisición de nutrientes es una de las necesidades más importantes para el crecimiento y desarrollo de las plantas, lo que adquiere una dimensión importante en el rendimiento de los cultivos en sistemas agronómicos. En particular, la disponibilidad de macro-nutrientes como el nitrógeno y el fósforo es uno de los factores que más limita la producción agrícola después del agua. Para suplir el déficit de nitrógeno de los suelos se desarrolló el proceso de Haber-Bosh para la producción industrial de grandes cantidades de amonio ( $\text{NH}_4^+$ ), el cual es utilizado como fertilizante nitrogenado en los sistemas agrícolas. Este proceso consiste en la reacción de nitrógeno e hidrógeno gaseosos a alta presión y temperatura para producir amoníaco. La aplicación de fertilizantes sintéticos en los sistemas agrícolas generó un importante incremento en la producción de los cultivos, sin embargo este incremento fue acompañado de altos costos económicos y ambientales. Se estima que el 1% de la energía mundial está destinada al proceso de Haber-Bosh. Uno de los desafíos mayores de la biotecnología moderna aplicada a la agricultura es el de aumentar la productividad de los cultivos requerida para alimentar a la población humana en permanente crecimiento de manera sustentable, reduciendo a su vez los costos económicos y ambientales asociados (FAO, 2009, *How to Feed the World in 2050*). Esto implica la aplicación de estrategias alternativas que permitan utilizar al máximo los recursos agronómicos disponibles y optimizar el rendimiento de los cultivos ('la nueva Revolución Verde', Jez et al., 2016).

El nitrógeno atmosférico ( $\text{N}_2$ ) es el gas más abundante en la atmósfera terrestre; sin embargo pocos organismos pueden incorporarlo directamente a su metabolismo. Son los casos de ciertas bacterias, cianobacterias, algas cianofíceas y actinomicetos. La fijación biológica de nitrógeno (FBN) consiste en la reducción de  $\text{N}_2$  en amonio ( $\text{NH}_3$ ) por estos microorganismos. El uso agronómico de la FBN es una alternativa sustentable a la síntesis de fertilizantes nitrogenados, y consecuentemente, una forma de disminuir la contaminación ambiental que producen. La asociación simbiótica entre plantas leguminosas y bacteria del género *Rhizobium* da cuenta de un porcentaje significativo del nitrógeno incorporado a los suelos (200 millones de toneladas fijadas por año (Ferguson et al., 2010). Además de proveer una fertilización natural a los suelos para los subsecuentes cultivos, el crecimiento de leguminosas en la rotación de cultivos promueve la diversidad y reduce significativamente las poblaciones de patógenos vegetales (Brockwell et al., 1995). Por otra parte, aproximadamente el 33% de las necesidades nutricionales de nitrógeno en las dietas humanas provienen de las leguminosas de grano, siendo la base proteica de las dietas de numerosos países en vías de desarrollo (Graham and Vance, 2003). De ahí que las leguminosas resulten indispensables para el desarrollo de sistemas de agricultura sustentable, preservando la integridad de los ambientes ecológicos y a su vez representen un valor económico y social para aquellas regiones dedicadas a la producción de leguminosas.

Las leguminosas de grano (*Fabaceae*, utilizadas para el consumo humano y alimentación del ganado) y las forrajeras incluyen las especies de mayor interés comercial, tanto a nivel global como local. Como ejemplos pueden mencionarse la soja (*Glycine max*), el poroto (*Phaseolus vulgaris*), el maní (*Arachis hypogaeae*), el garbanzo (*Cicer arietinum*), la lenteja (*Lens culinaris*), la arveja (*Pisum sativum*), y la alfalfa (*Medicago sativa*). Entre 12 y 15 % de la superficie de la tierra es utilizada para la siembra de leguminosas (Graham and Vance, 2003) y representan el 25 % de la producción primaria de cultivos del mundo (Ferguson et al., 2010). En Argentina, aproximadamente el 7 % de la superficie total del país está destinada al cultivo de la soja (Figura 1A, FAO stat.), siendo el tercer país productor después de Estados Unidos y Brasil. El 84% de la de soja producida en Argentina está destinada a la exportación como grano, harina, aceite o biodiesel (Bolsa de Comercio de Rosario, 2016), mientras que el resto de la producción local está destinada esencialmente a la alimentación del ganado.



**Figura 1: Leguminosas comestibles en Argentina.** A. El consumo de las leguminosas de grano, o legumbres es importante a nivel mundial. Los cultivos de soja (*Glycine max*), maní (*Arachis hypogaeae*), poroto (*Phaseolus vulgaris*), garbanzo (*Cicer arietinum*) y lenteja (*Lens culinaris*) tienen gran importancia en la agricultura argentina. Los datos representados de la área cosechada en hectáreas y la producción en toneladas provienen de la FAO (*Food and Agriculture Organization of the United Nations*), del último año disponible (2014). B. Evolución del cultivo de poroto en Argentina. La superficie cosechada y la producción de porotos secos han aumentado en los últimos años, sin embargo no se refleja en el rendimiento (expresado en hectogramos por hectárea). <http://www.fao.org/faostat>

## 1.2 Importancia del poroto en la alimentación y salud humana

El poroto común representa la mitad de las legumbres de grano de consumo mundial humano, aportando más de 70 % del contenido proteico de las dietas de las









poblaciones pobres de América Latina y África Oriental (Broughton et al., 2003). Su composición rica en minerales, especialmente hierro y zinc, y en vitaminas (Tabla 1) (Beebe et al., 1999; Ganesan and Xu, 2017) son esenciales para prevenir la anemia, la gastroenteritis, las infecciones respiratorias y la malnutrición, especialmente en niños (Bouis, 2003). Las semillas de poroto son fuentes importantes de hidrato de carbono (>50 % para la mayoría de las variedades) y de proteínas (entre 6 y 22 % según la variedad considerada). En cuanto a los minerales, son buena fuente de calcio, hierro, potasio y magnesio.

Particularmente, el poroto negro, de origen mesoamericano, es una variedad que tiene los mayores aportes energéticos, con una composición importante en proteínas (14 %), lípidos (21 %), carbohidratos (57 %), y que contiene altos niveles de minerales y vitaminas. Las semillas de poroto también representan una fuente de compuestos polifenólicos con numerosas propiedades para la salud humana (Ganesan and Xu, 2017). Los polifenoles son metabolitos de plantas con propiedades antioxidantes que reducen los efectos del estrés oxidativo generado en enfermedades graves. También poseen propiedades que contribuyen a prevenir la diabetes, la obesidad, las inflamaciones y el cáncer.

**Tabla 1: Variedades de porotos y aportes nutricionales.**

La composición nutricional está expresada por cada 100 g de una porción comestible. Esta tabla se realizó a partir de los datos proviniendo de Ganesan and Xu, 2017.

Variedades de poroto		Negro	Blanco	Kidney	Colorado	Pinto	Cranberry
Fenotipos de las semillas							
<b>Nutrientes</b>	<b>Unidades</b>						
Energía	kcal	464	92	92	167	500	257
Proteína	g	14.29	6.15	5.38	22.22	10.71	22.86
Lípidos totales	g	21.43	0	0.38	0	21.43	0
Carbohidratos	g	57.14	16.15	20.77	63.89	60.71	60
Fibras totales	g	14.3	4.6	5.4	44.4	10.7	25.7
Azúcares totales	g	3.57	0	0.77	2.78	3.57	2.86
Almidón	g	1.7	4.2	2	3.8	1.9	2.6
<b>Minerales</b>							
Calcio	mg	143	62	46	167	71	114
Hierro	mg	3.86	1.38	1.38	7.29	2.57	5.1
Potasio	mg	279	300	268	222	96	265
Magnesio	mg	60	48	37	44	43	39
Sodio	mg	286	108	8	69	286	10
<b>Vitaminas</b>							
Vitamina C	mg	0	0.9	0.9	0	0	0
Folate	µg	128	127	115	140	147	124
<b>Lípidos</b>							
Ácidos grasos saturados totales	g	1.79	0	0	0	1.79	0
Ácidos grasos mono-insaturados totales	g	14.29	0	0	0	14.29	0
Ácidos grasos poli-insaturados totales	g	5.36	0	0	0	5.36	0
Polifenoles	equiv/g	12.6	12.47	14.14	13.68	12.52	11.73
Flavonoides	equiv/g	1.28	1.78	2.59	1.55	0.98	1.65

América Latina contiene la área más grande del mundo de cultivo de poroto (Beebe et al., 2013). En Argentina, este cultivo ocupa la segunda posición en la producción en leguminosas consumidas por los humanos (Figura 1A). Mientras la superficie cosechada y la producción de porotos secos han aumentado constantemente en los últimos años en Argentina, el rendimiento de este cultivo no ha sido mejorado considerablemente (Figura 1B, FAO stat.). La actividad agrícola asociada a la

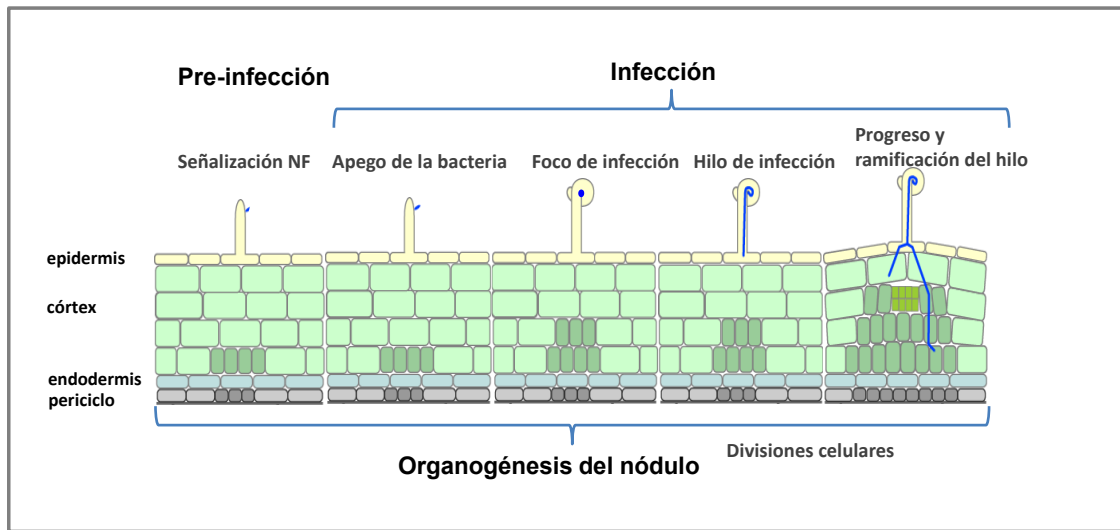
producción de poroto posee una importancia económica y social para la región del Noroeste del país; sin embargo, la producción alcanza solamente el 55 % de su potencial en cuanto al rendimiento, lo que provee amplio margen para su mejoramiento (Blair et al., 2009). Los experimentos a campo efectuados a nivel mundial mostraron que las tasas de nodulación y de fijación del nitrógeno son aún muy bajas (Graham, 1981; Hardarson, 1993). Por eso, es esencial estudiar la compatibilidad entre diferentes variedades de poroto y cepas de rizobios para lograr mejorar la fijación biológica de nitrógeno, y, en consecuencia, el rendimiento de los cultivos y la calidad de las semillas.

## **2. Interacción simbiótica entre leguminosas y rizobios**

### **2.1 Formación de nódulos funcionales: de la infección bacteriana a la fijación del nitrógeno**

Las plantas son capaces de establecer relaciones endosimbióticas con microorganismos del suelo como hongos y bacterias, las cuales proveen beneficios para ambos organismos. Una de las simbiosis de mayor importancia ecológica y agronómica es la simbiosis fijadora de nitrógeno, que se establece entre plantas leguminosas y bacterias del suelo conocidas comúnmente como rizobios. Esta asociación conduce a la formación de un nuevo órgano en la raíz de la planta denominado nódulo. En el interior del mismo se produce la conversión del nitrógeno atmosférico ( $N_2$ ) a formas reducidas que pueden ser utilizadas por el metabolismo de la planta, gracias a la actividad de un complejo enzimático bacteriano, la nitrogenasa (Thorneley et al., 1992). La bacteria recibe, como contrapartida, compuestos orgánicos derivados de la fotosíntesis que puede utilizar como fuente de energía y un nicho microaeróbico requerido para la reducción del  $N_2$ .

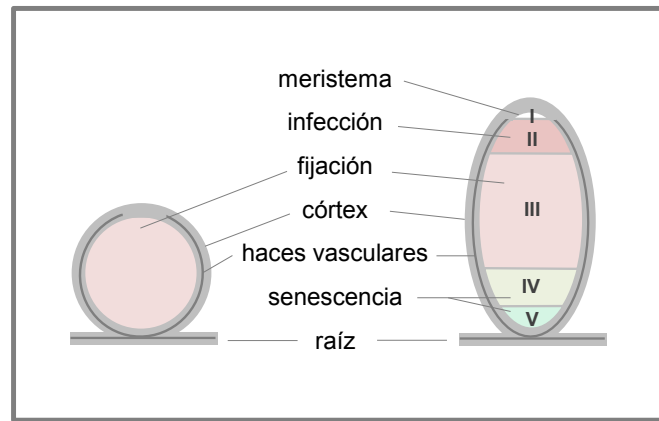
La formación de nódulos fijadores de nitrógeno requiere la activación simultánea de dos programas genéticos finamente coordinados: la infección bacteriana y la reiniciación de las divisiones corticales para formar el primordio del nódulo (Madsen et al., 2010). En el modo de infección más común de las simbiosis entre rizobios y leguminosas, la bacteria penetra la epidermis y las diferentes capas corticales hasta el nódulo emergente por una vía de infección muy sofisticada que involucra al pelo radical (Figura 2). La bacteria se adhiere a la superficie del pelo radical, el cual se curva atrapando a la bacteria. Dentro del hilo curvado la bacteria se multiplica formando una microcolonia. A partir de esta curvatura se produce la invaginación de la membrana plasmática que da lugar a la formación de una estructura tubular denominada hilo de infección. El hilo de infección se alarga y ramifica hacia las capas internas del córtex. Al mismo tiempo, se producen divisiones celulares en el córtex de la raíz que dan lugar a la formación del primordio del nódulo. Una vez que el hilo de infección alcanza el primordio nodular, las bacterias son internalizadas y rodeadas por una membrana derivada de la planta llamada membrana del simbiosoma, formando estructuras subcelulares de tipo organelas denominadas simbiosomas. Dentro de los simbiosomas, las bacterias diferencian a bacteroides (Haag et al., 2012), y utilizan la energía provista por la planta para convertir el nitrógeno atmosférico ( $N_2$ ) en amonio ( $NH_3$ ) (revisado en Gage, 2004).



**Figura 2: Las etapas de la infección bacteriana en la raíz.** El proceso de infección bacteriana se inicia por invasión de los pelos radicales. La bacteria (en azul) se adhiere al pelo radical, que se deforma y encierra la microcolonia, formando el foco de infección. Las bacterias progresan hacia el interior de la raíz a través del hilo de infección, una estructura tubular formada por invaginación de la membrana plásmatica, colonizando primero el pelo y luego las capas celulares internas corticales. En simultáneo, las células internas del córtex y del periciclo empiezan a dividirse (células más oscuras en el dibujo) para dar origen al primordio del nódulo.

## 2.2 Simbiosis y regulación hormonal

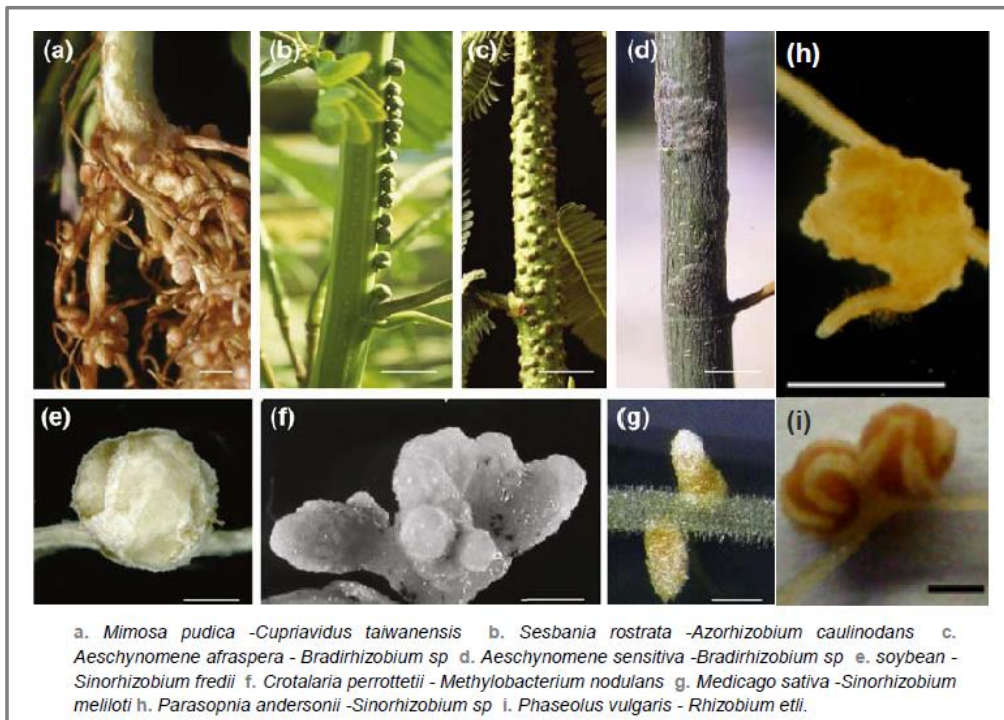
La organogénesis del nódulo es dependiente de una señalización hormonal que involucra una vía de citoquinina y la inhibición localizada del transporte de auxina en los sitios de infección (revisado en Oldroyd et al., 2011). Esta modulación por hormonas es diferente para los dos tipos de nódulos que se forman en la simbiosis con rizobios, los nódulos determinados (de los géneros *Lotus*, *Glycine*, *Vigna* y *Phaseolus*), y nódulos indeterminados (grupos *Medicago* y *Pisum*). Los dos se diferencian por su forma en el estado maduro, redonda o alargada para determinado e indeterminado respectivamente (Figura 3). Los nódulos de tipo determinado pierden la actividad meristemática rápidamente, después de la iniciación y el tamaño del nódulo crece debido al crecimiento celular. Al contrario, los nódulos indeterminados mantienen un meristema apical activo, lo que produce la formación de nuevas células a lo largo de la maduración del órgano. Los nódulos indeterminados son estructurados en zonas, las cuales dependen de su desarrollo. En el ápex se encuentra el meristema persistente (zona I), seguido de la zona de infección (zona II), y de la zona III, donde los bacteroides diferenciados fijan el nitrógeno atmosférico. En los estadios más tardíos se forma una cuarta zona en la parte basal donde los bacteroides y las células vegetales son degradadas (zona IV de senescencia) (Patriarca et al., 2002). En los estadios maduros avanzados, la parte basal del nódulo de la zona de senescencia evoluciona en zona V, caracterizada por la presencia de bacterias saprófitas (Dupont et al., 2012).



**Figura 3: Arquitectura de los nódulos.** Los nódulos de tipo determinado crecen por expansión celular, resultando en nódulos maduros de forma esférica (ej. *P. vulgaris* a la izquierda). Los nódulos indeterminados (ej. *M. truncatula* a la derecha), por su meristema continuamente activo son característicos por su forma cilíndrica en el estado maduro. Se dividen en zonas correspondiendo a las diferentes etapas de su desarrollo. La zona I situada en el ápex es la zona meristemática, la zona II es la zona de infección donde los rizobios son liberados desde los hilos de infección. La zona III es la zona de fijación del nitrógeno atmosférico. La zona IV es la zona en la cual las células entran en senescencia. En los nódulos maduros se distingue una zona V de senescencia avanzada donde las bacterias no diferenciadas se multiplican en un modo de vida saprófito e invaden las células. Los nódulos fijadores de nitrógeno son de color rosado debido al funcionamiento de la leghemoglobina. Cuando las células entran en senescencia, esta proteína es degradada dando un color verde. Los nódulos son protegidos por tejidos externos corticales y están conectados al sistema de conducción de la raíz a través de haces vasculares que permiten el intercambio de solutos y agua.

### 2.3 Plantas modelos, secuenciación y aplicación

La familia de las leguminosas (*Fabaceae*) es la tercera familia más amplia de las angiospermas, con más de 18000 especies ya descritas. La familia incluye hierbas, arbustos, árboles y vides repartidos a través el mundo en diferentes hábitats y ecosistemas. Dos especies pertenecientes a la subfamilia *Faboideae*, *Lotus japonicus* y *Medicago truncatula*, han sido seleccionadas como modelos de leguminosas y han facilitado los estudios genéticos y funcionales a partir del desarrollo de herramientas genéticas, como métodos de transformación, bancos de mutantes, datos de expresión a gran escala y la secuenciación completa de sus genomas. *L. japonicus* y *M. truncatula* presentan tipos de nodulación determinada e indeterminada respectivamente. Estas plantas modelos de leguminosas permitieron dilucidar muchos de los procesos moleculares y fisiológicos de la simbiosis leguminosas-rizobios. Sin embargo, los nódulos formados por los rizobios presentan una gran diversidad fenotípica, donde la forma, el color, la cantidad y la morfología no son parámetros universales (Figura 4). Semejante diversidad apareció como consecuencia del tiempo de divergencia y de diferencias en el estado de ploidia entre especies de *Fabaceae*, involucrando mutaciones de genes y selección basada en las condiciones ambientales.



**Figura 4: Diversidad en la formación de nódulos en la simbiosis entre rizobios y leguminosas.**

La forma de los nódulos, su densidad y su localización es muy variable según la especie de leguminosas. Los nódulos son generalmente formados en la raíz, en la zona de pelos radicales en crecimiento, aunque en algunos casos (ej. b-c-d) los rizobios pueden penetrar por el tallo de las plantas, formando nódulos extra-radicales. Escalas: (a-d) 1 cm, (e-h) 1 m, (i) 5 mm. Figura adaptada a partir de Masson-Boivin et al., 2009 (a-g); Op den Camp et al., 2011 (h); y Zanetti et al., 2010 (i).

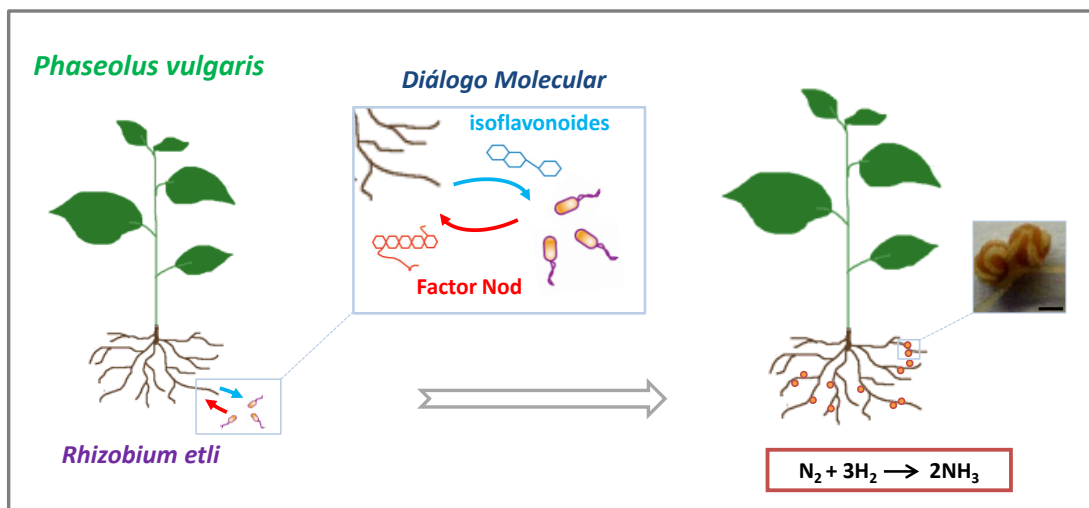
Por otra parte, la subfamilia *Papilionoideae* contiene varias leguminosas de gran importancia agronómica como la soja (*Glycine max*) y el poroto (*Phaseolus vulgaris*). Los genomas completos de estas plantas han sido secuenciados utilizando la estrategia de shotgun (Schmutz et al., 2010, 2014, respectivamente). En primer lugar se secuenció el genoma de una variedad andina de *P. vulgaris* (cultivar Race Peru), y, recientemente con una estrategia de secuenciación híbrida involucrando las tecnologías de 454, SOLiD y Sanger se obtuvo la secuenciación del genoma de la variedad mesoamericana BAT93 (Vlasova et al., 2016). Recientemente se ha desarrollado una base de datos de factores de transcripción de poroto (PvTFDB) con más de 2000 modelos génicos putativos disponibles (Bhawna et al., 2016).

La disponibilidad de los genomas representantes de los dos *pools* genéticos de *P. vulgaris* y su importancia agronómica posicionan al poroto como sistema modelo de leguminosas de grano utilizadas para la alimentación humana (Castro-Guerrero et al., 2016; Bitocchi et al., 2017). La diversidad natural visible en el fenotipo de sus semillas (forma, tamaño y color, Tabla 1), relacionado a su historia evolutiva y el desarrollo de nuevas herramientas genéticas, incentivó nuevamente el interés de su estudio. Algunas de las orientaciones en las líneas de investigación principales actualmente son las caracterizaciones de las respuestas adaptativas de plantas creciendo en suelo con deficiencia de nutrientes (Castro-Guerrero et al., 2016), la comprensión en la evolución del cultivo (revisado en Bitocchi et al., 2017), o los estudios de

caracterización y determinación de eficiencia simbiótica con rizobios aislados a partir de nódulos (de Sousa Cardoso et al., 2016). Una de las líneas de investigación con aplicación directa en el campo se enfoca en el mejoramiento a través de la selección de caracteres de interés agronómico utilizando variedades silvestres (resistencia a patógenos, tolerancia a estreses abióticos, etc.) para su transferencia a variedades domesticadas, permitiendo aumentar la biomasa y el rendimiento. Aunque el mejoramiento de los cultivos fue desde siempre una prioridad para la agricultura, surgió nuevamente la necesidad de acceder a los recursos genéticos presentes en la biodiversidad con el fin de responder a problemas ecológicos o climáticos actuales, como la destrucción de hábitats, la aparición de especies invasivas, la urbanización, la contaminación ambiental y los cambios climáticos.

## 2.4 Simbiosis: diálogo molecular inicial

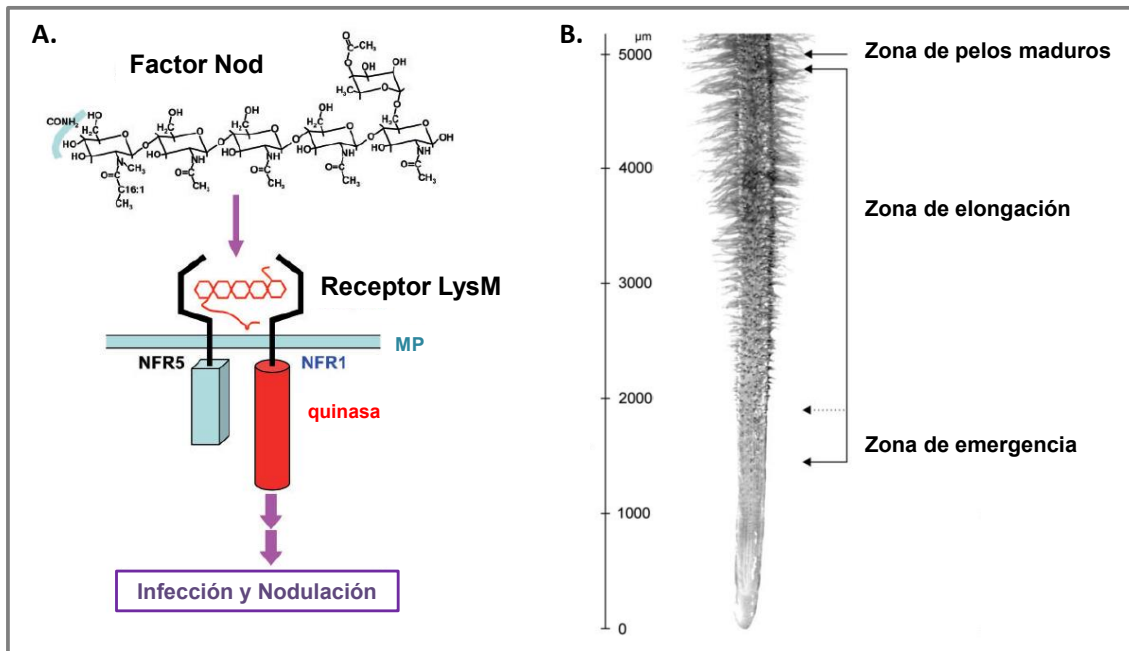
La interacción simbiótica entre leguminosas y rizobios se inicia con un intercambio de señales en la interfase entre la raíz y el suelo (Figura 5). En primer lugar, la bacteria percibe metabolitos exudados por la raíz derivados de la ruta de los fenilpropanoides, principalmente flavonoides e isoflavonoides (Redmond et al., 1986; Hassan and Mathesius, 2012). Estos compuestos son percibidos por el regulador proteico bacteriano NodD, que activa los genes bacterianos de nodulación o genes *nod*, cuyos productos proteicos participan de la síntesis de los factores Nod (NF). Estas moléculas de naturaleza lipochito-oligosacárida actúan activando muchas de las respuestas requeridas para la nodulación y son determinantes de la especificidad de especies leguminosas que los rizobios son capaces de nodular (Oldroyd and Downie, 2008).



**Figura 5: Simbiosis fijadora de nitrógeno.** La interacción empieza en la rizósfera a través del intercambio de moléculas entre los dos simbioses. En respuesta a la percepción de los isoflavonoides secretados por la leguminosa, se activan genes del rizobio que participan en la síntesis de los factores Nod. Estas moléculas de naturaleza lipochito-oligosacárida dan inicio a los procesos de infección bacteriana y de organogénesis del nódulo. En este órgano, las bacterias diferenciadas en bacteroides fijan el nitrógeno atmosférico ( $N_2$ ) convirtiéndolo en amonio ( $NH_4^+$ ). La disponibilidad de nitrógeno en formas asimilables por el metabolismo vegetal tiene un impacto positivo en su crecimiento.

La percepción de los NF en la epidermis de la raíz es mediada por receptores tipo quinasa (RLKs, *Receptor like kinases*) con dominios extracelulares de tipo LysM (Figura 6A). En *L. japonicus*, los receptores NFR1 y NFR5 (*Nod Factor Receptor 1* y 5) fueron identificados mediante estudios funcionales de mutantes con pérdida de función (Madsen et al., 2003; Radutoiu et al., 2007). Plantas mutantes *Ljnfr1* y *Ljnfr5* exhiben el mismo fenotipo no nodulante, sugiriendo que ambos receptores podrían formar un receptor heterodimérico. Estudios posteriores presentaron evidencias adicionales de la heterodimerización de estos receptores (Broghammer et al., 2012). Por otra parte, se ha demostrado que los dominios LysM de NFR5 funcionan como responsables de la especificidad de la percepción del NF (Radutoiu et al., 2007). De manera similar, en *M. truncatula*, el receptor NFP (por *Nod Factor Perception*) fue identificado mediante el análisis de mutantes deficientes en la nodulación (Amor et al., 2003). NFP es un RLK con dominios LysM, sin embargo, su dominio quinasa no muestra actividad de autofosforilación, en contraposición al receptor codificado por el gen *LYK3* (*LysM receptor-like kinase 3*), ortólogo a *NFR1* (Arrighi, 2006). LYK3 es considerado como receptor de entrada (Limpens et al., 2003; Smit et al., 2007), implicado en el reconocimiento de los sustituyentes estructurales específicos de los NF (grupos lipídicos y acetatos) (revisado en Jones et al., 2009). Este reconocimiento y el posterior desarrollo del programa de nodulación está circunscripto a la región de pelos radicales en crecimiento activo, conocido como zona II (Figura 6B) (Hadri and Bisseling, 1998). En *Medicago*, el reconocimiento temprano de los NF por NFP a nivel de los pelos radicales inicia una vía de señalización que involucra oscilaciones de calcio, denominadas *calcium spiking*, las cuales son esenciales para la activación de genes de nodulación o nodulinas tempranas. En una segunda etapa, NFP y LYK3 asociados a la membrana son necesarios para la activación de la vía de señalización asociada a la infección.

Este reconocimiento de los NF activa respuestas diversas en el citoplasma, como influjos de iones y oscilaciones de calcio, que son responsables de los cambios morfológicos y moleculares en las células epidérmicas, y al mismo tiempo, son necesarias para los eventos moleculares que permiten la organogénesis del nódulo (Oldroyd and Downie, 2008; Oldroyd et al., 2011).



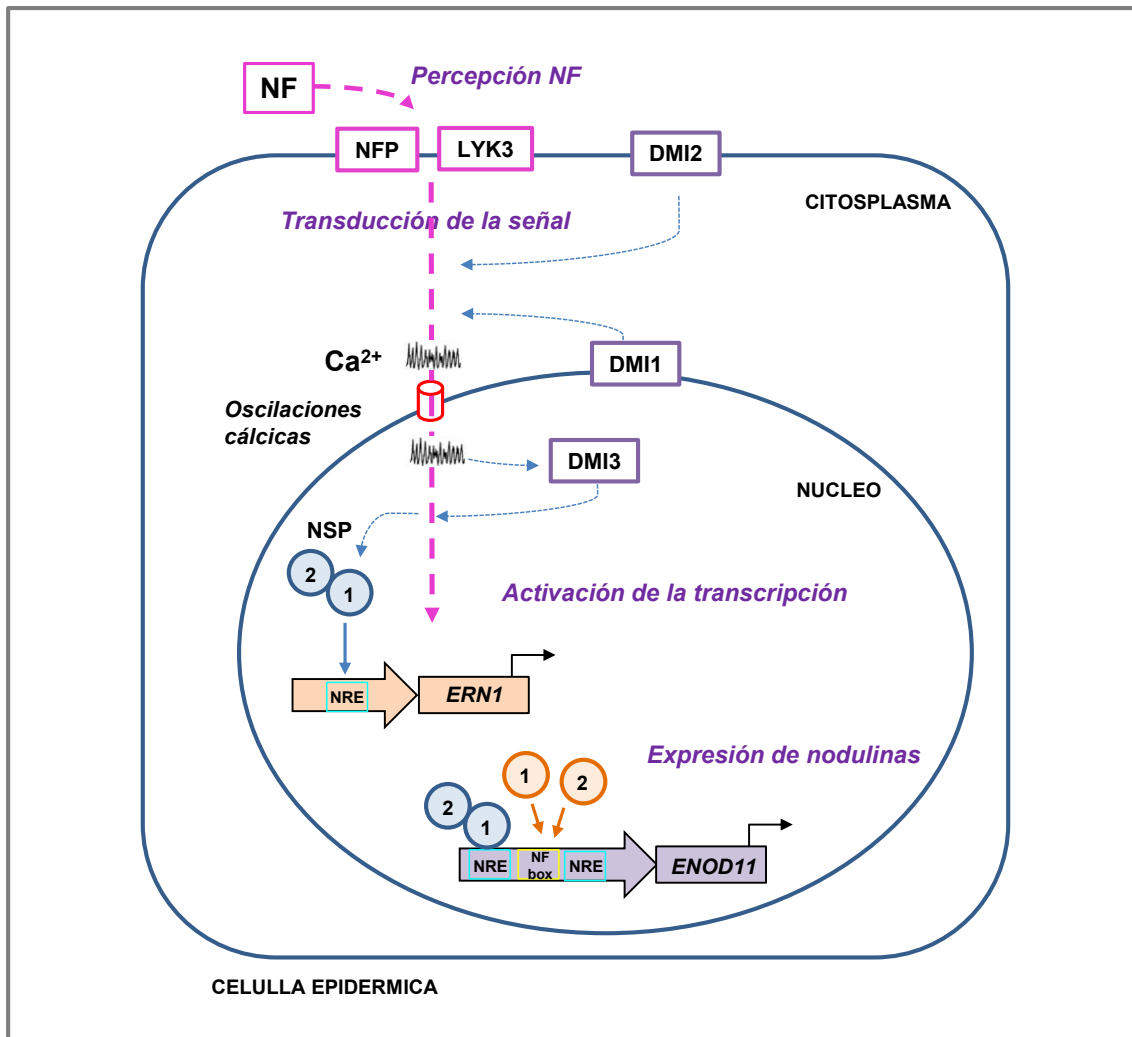
**Figura 6: Percepción del Factor Nod en la epidermis de los pelos radicales.** A. Reconocimiento del Factor Nod por NFR1 y NFR5 en *L. japonicus*. Los receptores putativos de Factor Nod fueron identificados en varias especies de leguminosas. Su estructura está constituida por un dominio extracelular con motivos lisina (LysM), un dominio transmembrana (MP por membrana plasmática), y un dominio quinasa intracelular. Los dominios LysM extracelulares serían responsables de la unión del Nod Factor al receptor, y la quinasa de NFR1 de la transducción de la señal conduciendo a la infección bacteriana y la nodulación. El dominio quinasa de NFR5 contiene aminoácidos no conservados en su sitio catalítico, por lo que ha sido catalogado como una pseudoquinasa. B. Zonas del desarrollo del pelo radical en *L. japonicus*. La infección rizobiana ocurre en la zona de elongación de los pelos radicales. Figura adaptada a partir de Kouchi et al., 2010 (A) y Karas et al., 2005 (B).

## 2.5 Señalización temprana en la simbiosis fijadora de nitrógeno

En estos últimos 20 años se han realizado enormes progresos en la caracterización de la respuesta a los NF a través de estudios genéticos y funcionales, especialmente en las dos leguminosas modelos. En los pelos radicales, la activación de la cascada de señalización dependiente de NF conduce a oscilaciones de calcio específicas necesarias para la transcripción de genes de respuesta temprana, llamados nodulinas (Figura 7). Por ejemplo, *ENOD11* es un gen de *M. truncatula* que codifica para una proteína de la pared de tipo con dominios repetitivos ricos en prolina (RPRP, *Repetitive Proline Rich Protein*), el cual es fuertemente expresado en respuesta a la bacteria simbiótica o por tratamiento de la raíz con NF purificados (Journet et al., 2001; Charron et al., 2004). Comúnmente se utiliza el transcripto de *ENOD11* como marcador molecular de la activación de esta vía de señalización en estudios de las etapas tempranas de la interacción. Luego de la percepción de los NF por los receptores LysM-RLK, el canal iónico localizado en la membrana nuclear codificado por *DMI1* (*Doesn't Make Infection 1*; *CASTOR* y *POLLUX* en *L. japonicus*) y el receptor



quinasa con dominio LRR (*Leucine Rich Repeat*) localizado en la membrana plasmática codificado por *DMI2* (*SYMRK* en *L. japonicus*), son actores esenciales a la transmisión de la señal anterior a las oscilaciones de calcio (Riely et al., 2007). Recientemente se identificaron canales de nucleótidos cíclicos requeridos para las oscilaciones de calcio. Estos CNGCs (*nuclear-localized cyclic nucleotide-gated channels*) de la membrana nuclear interactúan con el canal de potasio *DMI1* para modular la liberación de calcio necesaria a las respuestas simbióticas (Charpentier et al., 2016). La quinasa dependiente de calcio y calmodulina *CCaMk* (*Calcium-Calmodulin-dependent protein kinase*) codificada por *DMI3* decodificaría las oscilaciones de calcio en el núcleo. En *M. truncatula* *IPD3* (o *CYCLOPS* en *L. japonicus*) codifica por una proteína específica de plantas con dominio NLS (*Nuclear Localization Signal*) y un dominio super enrollado coiled-coil. En el núcleo, esta proteína conservada en angiospermas interactúa y es fosforilada por *DMI3*. En *L. japonicus* se mostró que *CYCLOPS* es requerido para la activación transcripcional de *NIN* y *ENOD40* indispensable para la organogénesis del nódulo (Yano et al., 2008). La nodulina *NIN* (*Nodule Inception*) tiene un papel central como regulador de la nodulación, participando en la coordinación del programa de desarrollo durante la simbiosis. En las células epidérmicas es necesario para la iniciación de la infección bacteriana, mientras que también promueve la organogénesis del nódulo en las células corticales (Schauser et al., 1999; Marsh et al., 2007; Soyano et al., 2013). *NIN* actúa además como regulador negativo inhibiendo las repuestas al NF después de la activación inicial de la señalización en la epidermis (Vernié et al., 2015). Así, actuando en conjunción con las auxinas y citoquinas (Ferguson and Mathesius, 2014), *NIN* tiene un papel central en la regulación de la formación de nódulos, denominado autorregulación de la nodulación (Marsh et al., 2007; Soyano and Hayashi, 2014; Soyano et al., 2014; Yoro et al., 2014). En las células epidérmicas, *NIN* restringe la expresión de *ENOD11* a través de la inhibición competitiva de *ERN1*. En el córtex, *NIN* es suficiente para promover la señalización de citoquinas a través de la activación del receptor de citoquinina *CRE1* (*Cytokinin Response 1*), necesario para la organogénesis del nódulo (Vernié et al., 2015).



**Figura 7: Señalización iniciada por NF en la epidermis de la raíz.** Los principales determinantes moleculares implicados en la respuesta a NF conocidos han sido caracterizados en las leguminosas modelos. Los actores involucrados en la percepción del NF y la transducción de señales que lleva a la organogénesis del nódulo participan del denominado *Nod Pathway*. En este esquema están representados algunos de sus componentes validados por genética funcional en *Medicago truncatula* y que han sido implicados en las etapas tempranas de la interacción. El receptor *like-kinase* NFP asociado a LYK3 percibe los factores Nod en la membrana de los pelos radicales. El receptor quinasa DMI2 en la membrana citoplasmática y el canal catiónico DMI1 de la membrana nuclear son esenciales para la inducción de señales de calcio dentro y en la periferia del núcleo entrando a través de canales de calcio (representado por el cilindro rojo). La proteína quinasa DMI3 es la que respondería a las oscilaciones de calcio en el núcleo, siendo necesaria para la activación de los factores de transcripción NSP1 y NSP2 (representados por los círculos azules), que actúan como heterodímeros. NSP1 reconoce los motivos NRE presentes en los promotores de *ERN1* y *ENOD11*. *ERN1* y *ERN2* (representados por los círculos naranja) son factores de transcripción que reconocen la NF box del promotor de *ENOD11*, activando su expresión.

Aguas abajo de *DMI3*, se ha mostrado la importancia de factores de transcripción de tipo GRAS (NSP) y ERF (ERN) en la transmisión de la señal: estos factores son esenciales para la expresión de *ENOD11*. En primer lugar *NSP1* y *NSP2* (por

*Nodulation Signalling Pathway*) (Smit et al., 2005; Kaló et al., 2005), forman un heterodímero que se une al ADN en los motivos NRE (por *Nodulation Responsive Element*) de los promotores de *ERN1* y *ENOD11*, activando la expresión de dichos genes (Hirsch et al., 2009). Además, el promotor de *ENOD11* contiene una secuencia de 33 bp (el NF-box) necesaria y suficiente para su expresión mediada por los NF en los pelos radicales (Boisson-Dernier et al., 2005; Andriankaja et al., 2007). La NF-box es el blanco de los factores de transcripción ERN1 y ERN2 (por *ERF Required for Nodulation*) para activar la transcripción de *ENOD11* (Andriankaja et al., 2007). ERN1 es requerido para activar la expresión de *ENOD11* dependiendo de los NF exclusivamente durante los tiempos tempranos de la pre-infección (antes y durante la progresión del hilo de infección), mientras que NSP1 y NSP2, no tienen un papel directo sobre la regulación de *ENOD11* durante la señalización a pesar de actuar aguas arriba de ERN1. A su vez, ERN2 está implicado (pero no es esencial) en algunas etapas de la infección rizobiana más tardías, como durante la penetración del hilo de infección en las células corticales y los nódulos infectados (Cerri et al., 2012).

Además de la señal contenida en los NF, existen moléculas de señales secundarias presentes en la superficie de la pared bacteriana o secretadas al medio, tales como los exopolisacáridos (EPS) y lipopolisacáridos (LPS), que contribuyen a la selección de la cepa de rizobios, suprimiendo las respuestas de defensa durante la iniciación y el desarrollo de la infección (Aslam et al., 2008; Scheidle et al., 2004; revisado en Downie, 2010). La inoculación de raíces con cepas de bacterias mutantes afectadas en la biosíntesis de EPS no resulta en la producción de hilos de infección ni nódulos (Cheng and Walker, 1998; Jones and Walker, 2008). Recientemente en *L. japonicus* se ha identificado *EPR3*, un receptor de tipo RLK que reconoce el EPS de *M. loti* (Kawaharada et al., 2015). Mientras la percepción de los NF es determinante para la inducción de la expresión de *EPR3*, el reconocimiento de los EPS tendría un rol en la discriminación de bacterias compatibles e incompatibles. Por otra parte, la Inoculación con mutantes bacterianos afectados en la biosíntesis de LPS llevan a anomalías en el proceso de infección (Perotto et al., 1994; Gao et al., 2001). Recientemente se realizó un análisis de transcriptómica a tiempos tempranos de la simbiosis (24 horas post inoculación) en raíces de *P. vulgaris* inoculadas con cepas de *R. etli* deficientes en las síntesis de NF, EPS o LPS (Dalla Via et al., 2015). Más de 2600 genes fueron identificados como diferenciales de acuerdo a su expresión, incluyendo factores de transcripción y receptores. Estos resultados revelan una respuesta molecular compleja de la planta frente a las diferentes moléculas de señalización del rizobio. El análisis más profundo de estos datos ómicos permitirá identificar nuevos actores directos o indirectos de la señalización generada por estas moléculas de superficie o secretadas por los rizobios. Esta comunicación química compleja permite el reconocimiento y la colonización de cepas de bacterias específicas, a la vez que inhibe las respuestas de defensa prolongadas en las raíces de la planta.

## 2.6 Simbiosis y evolución

Dada la naturaleza sésil de las plantas, la optimización en la adquisición de nutrientes es crucial para su supervivencia. Para superar esta limitación, las plantas han desarrollado mecanismos que les permiten establecer interacciones simbióticas con

microorganismos benéficos que facilitan la captura de estos nutrientes. La simbiosis fijadora de nitrógeno involucra más de 18000 especies de leguminosas en asociación con  $\alpha$ - y  $\beta$ -proteobacterias, de los géneros *Rhizobium*, *Sinorhizobium*, *Mesorhizobium* y *Bradyrhizobium*. Desde el punto de vista evolutivo se ha propuesto que la capacidad de establecer una interacción simbiótica fijadora de nitrógeno apareció varias veces independientemente, ya que se encuentra presente no sólo en las leguminosas, sino también en especies muy distante como por ejemplo en *Parasponia* y en las plantas actinorizias (revisado en Doyle, 2011). El mecanismo de señalización es común entre las especies de leguminosas (Fabaceae) y *Parasponia* (Cannabaceae), y evolucionó probablemente desde una señalización más ancestral que permite a la mayor parte de las plantas terrestres reconocer los hongos que forman micorrizas arbusculares (AM) (Gherbi et al., 2008; Op den Camp et al., 2011; Maillet et al., 2011; Harrison, 2005; Froussart et al., 2016).

La simbiosis AM es una interacción mutualista en la cual el microsimbionte facilita la captura de minerales, esencialmente fosfato, a través de la formación de estructuras especializadas en la raíz, denominadas arbusculos, donde los hongos se alojan (Strack et al., 2003; Parniske, 2008). Ambos tipos de simbiosis, la simbiosis fijadora de nitrógeno (SFN) y la AM, requieren del establecimiento de un diálogo molecular complejo entre ambos organismos (Gough and Cullimore, 2011; Oldroyd, 2013). En el caso de la SFN, ésta se inicia por la exudación de flavonoides por parte de la raíz, mientras que la AM se inicia por la liberación de estrigolactonas (Kretzschmar et al., 2012), señales que serán reconocidos por los rizobios u hongos AM, respectivamente. Estas señales de la planta inducen la producción de lipochitoligosacáridos específicos, NF para los rizobios, y Myc-LCOs para los hongos AM (Dénarié et al., 1996; Maillet et al., 2011). La percepción de estas moléculas por las células de la raíz conduce a una vía de señalización común denominada *Common Symbiotic Signaling Pathway* (Bapaume and Reinhardt, 2012; Singh and Parniske, 2012; Harrison, 2012; Haage and Parniske, 2013; Oldroyd, 2013). Varios componentes de estas vías de señalización son compartidos entre ambas simbiosis y se pueden clasificar según su función en las diferentes etapas del establecimiento de la interacción (Gutjahr and Parniske, 2013). Las etapas de infección del microorganismo (Catoira et al., 2000; Kistner, 2005) y la percepción simbiótica en la membrana plasmática de la raíz son mediadas por receptores que ya han sido caracterizados en la SFN, pero aún no se han identificado los receptores de los Myc-LCOs (Gobbato, 2015). La transducción de señales mediada por oscilaciones de calcio en el núcleo y la activación de otras proteínas quinasas, junto con SYMRK (Markmann et al., 2008; Stracke et al., 2002) y la CCaMK (Singh et al., 2014) están también conservados entre las dos simbiosis. La nodulina *ENOD11*, que tiene un papel clave en diferentes etapas de la SFN es altamente expresada en las células corticales internas durante la colonización por hongo micorrízicos (Journet et al., 2001). El papel de nucleoporinas como NUP133, NUP85 y NENA (Kanamori et al., 2006; Saito et al., 2007; Groth et al., 2010) también ha sido demostrado en ambas simbiosis. Genes de esta vía de señalización están también funcionalmente conservados en arroz, lo que sugiere que su existencia es anterior a la diversificación de las Angiospermas (Markmann et al., 2008; Gutjahr et al., 2008; Chen et al., 2009; Imaizumi-Anraku et al., 2005; Banba et al., 2008).

Más recientemente se estudió y comparó la implicación de las hormonas en el desarrollo de ambas simbiosis (Foo et al., 2016). Estos procesos están regulados por las auxinas, citoquininas, estrigolactonas, giberelinas, ácido jasmónico, etileno y ácido abscísico, a través de redes complejas, que dependen de estímulos del medio ambiente (Gutjahr, 2014; Ferguson and Mathesius, 2014). También se estableció la función de las proteínas DELLA, que son reguladoras centrales de la señalización del ácido giberélico y su conexión con factores de transcripción del *Common Symbiotic Signaling Pathway* (Jin et al., 2016). Otras de las características que tienen en común estas simbiosis son la estrategia de colonización de la raíz y la supresión de las respuestas de defensa de la planta (Martin et al., 2017).

### 3. Relación entre simbiosis y respuestas de defensa

Las plantas han desarrollado sofisticados mecanismos de defensas que les permiten protegerse del ataque de patógenos (bacterias, hongos, virus, insectos). La primera línea de defensa es una respuesta inmune innata (revisado en Cao et al., 2017), basada en el reconocimiento de moléculas o patrones moleculares asociadas a microbios que se encuentran conservados en la evolución: los MAMPs o PAMPs (por *microbial- o pathogen-associated molecular patterns*). Este reconocimiento requiere receptores localizados en la membrana plasmática de las células, las PRRs (*pattern recognition receptors*) e inicia la respuesta inmune que corresponde a la PTI (*PAMP-triggered immunity*). Los patógenos competentes han adquirido la capacidad de superar la PTI a través de la producción de efectores que ingresan al citoplasma del huésped e interfieren con la respuesta de defensa, la ETS (*effector-triggered susceptibility*). Con la evolución, las plantas resistentes desarrollaron otra línea de defensa llamada ETI (*effector-trigger immunity*), para lidiar contra los patógenos virulentos. La ETI depende del reconocimiento específico de un efector por una proteína de resistencia codificada por un gen de resistencia (R). La evolución del sistema inmune de las plantas en respuesta a los patógenos ha sido ilustrado por el Modelo Zigzag (Jones and Dangl, 2006). Este representa el aumento de la amplitud de la defensa que avanza con la complejidad de las repuestas de virulencia del patógeno y de la resistencia del huésped. La mayoría de los genes R codifican receptores intracelulares con dominios de unión a nucleótido y dominios ricos en leucina, los denominados NB-LRR (con dominios *nucleotide binding and leucine rich repeat*) (Dangl, 2013; y revisado en Elmore et al., 2012; Takken and Govers, 2012; Hacquard et al., 2017). Como en la simbiosis, la PTI y ETI disparan una cascada de señalización donde participan influjos de iones, activación de proteínas quinasas y modificación de la síntesis de hormonas. En el núcleo, más particularmente la ETI está asociada a la regulación transcripcional de la expresión de genes de defensa.

Los receptores que detectan las moléculas de patógenos contienen diferentes dominios de reconocimiento. Los principales son los LRR (revisado en Bella et al., 2008) y los dominios de unión a lectina, que desencadenan una transducción de la señal a través de la reprogramación de la expresión de genes (revisado en Clúa et al., 2018). Estos dominios extracelulares son capaces de reconocer proteínas, péptidos o carbohidratos secretados por los patógenos o productos de la acción de enzimas hidrolíticas de la pared celular vegetal. Por ejemplo, las moléculas derivadas de quitina

son importantes en la respuesta de señalización al hongo patógeno de cual las paredes celulares son degradadas por la acción de quitinasas de la planta. Los chito-oligosacáridos (CO) y lipochito-oligosacáridos (LCO) son igualmente moléculas importantes en las señalizaciones de las simbiosis a micorrizas (Maillet et al., 2011; Genre et al., 2013) y formadores de nódulos (Dénarié and Cullimore, 1993). Los receptores de los factores Nod y los exopolisacáridos (EPS) rizobianos contienen dominios LysM extracelulares (Amor et al., 2003; Arrighi, 2006; Kawaharada et al., 2015; Limpens et al., 2003), que son motivos comunes a las PRR (revisado en Desaki et al., 2017). En un trabajo reciente, se ha mostrado que diferentes miembros de los receptores LysM de *L. japonicus* y *M. truncatula* están implicados en la percepción de CO y LCO, ambos iniciando las respuestas de defensa o simbióticas (Bozsoki et al., 2017). Estos resultados sugieren que las vías de transducción de señales son independientes y pueden ejercer un efecto negativo entre ellos.

Durante la simbiosis entre las leguminosas y los rizobios, el éxito de la infección rizobiana depende de una efectiva supresión de las respuestas de defensa de las plantas. Experimentos de transcriptómica hechos en soja (*G. max*, Libault et al., 2010), en *M. truncatula* (Lohar et al., 2006) y *L. japonicus* (Kouchi et al., 2004) sustentan esta hipótesis. Estos estudios mostraron que las etapas tempranas de la interacción con el rizobio o etapas de pre-infección están asociadas a la inducción de genes de defensa, mientras que en etapas más tardías estas respuestas se reprimen para dar paso al inicio de los eventos de infección. La relación entre la simbiosis y defensa ha sido demostrada mediante estudios funcionales. Por ejemplo, las raíces de *L. japonicus* tratadas tanto con el MAMP flg22 o una suspensión de células de *M. loti* muestran la inducción de respuestas de defensas como aumento en la producción de etileno, la activación de la MAPK y la inducción de genes de defensa canónicos. El pre-tratamiento de flg22 genera un retraso de la nodulación y una reducción del número de nódulos. Por el contrario, el tratamiento con flg22 después de la inoculación con el rizobio no afecta la cantidad de nódulos formados (Lopez-Gomez et al., 2012), sugiriendo que el rizobio ha superado la respuesta de defensa. En *M. truncatula* también se ha mostrado que la inoculación de plantas con la cepa DC3000 de la bacteria patógena *Pseudomonas syringae* pv. Tomato o la cepa Sm2011 del simbiote *S. meliloti* activa las respuestas de defensa. Sin embargo, cuando las dos cepas fueron co-inoculadas, el rizobio suprime las respuestas de defensa generadas por la bacteria patógena, y esta última reduce el número de hilos de infección y de nódulos formados por el simbiote, disminuyendo también la expresión de nodulinas como NIN (Chen et al., 2017). Los receptores NB-LRR son conocidos por su papel en las respuestas de defensa, en particular en la ETI mediante el reconocimiento de factores específicos. Sin embargo, estos receptores pueden también ejercer una función en la simbiosis fijadora de nitrógeno. Por ejemplo Yang et al., 2010 han demostrado en soja que *Rj2* y *Rfg1* son genes alélicos que codifican proteínas TIR-NB-LRR (*Toll-Interleukine NB-LRR Receptor*) que participan en el control de la especificidad de huésped, en la infección y la nodulación.

En conjunto, los antecedentes mencionados sugieren que existe un mecanismo de reconocimiento común entre las interacciones simbióticas y patogénicas (revisado en Zipfel and Oldroyd, 2017). Durante la simbiosis, luego de una inducción transitoria de las respuestas de defensas, estas son reprimidas para permitir la infección rizobiana.

Sin embargo, la interconexión entre las dos vías de señalización no ha sido aún dilucidada. La existencia de componentes y/o comunes o compartidos en las vías de señalización de las interacciones simbióticas o patogénicas sugieren que la maquinaria de reconocimiento y señalización de la simbiosis rizobiana podría haber evolucionado a partir de aquellas existentes previamente en las interacciones patogénicas (Deakin and Broughton, 2009).

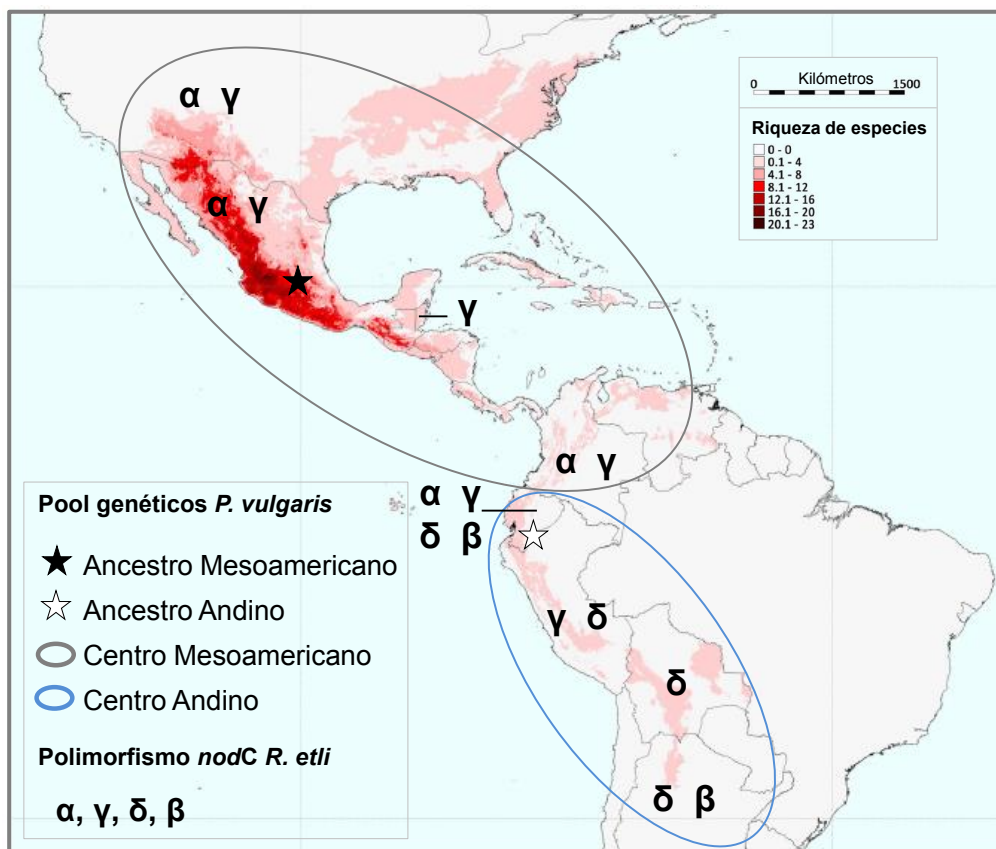
## **4. La interacción entre *Phaseolus vulgaris* y *Rhizobium etli***

### **4.1 Asociación preferencial entre variedades de *P. vulgaris* y su simbiote *R. etli***

Dentro de la familia de las leguminosas, el género *Phaseolus* incluye unas setenta especies, de las cuales cinco han sido domesticadas: *Phaseolus vulgaris*, *P. coccineus*, *P. dumosus*, *P. acutifolius*, y *P. lunatus* (Bitocchi et al., 2017). *Phaseolus vulgaris* es nativo de América y la distribución actual de su cultivo cubre una amplia zona geográfica, desde el norte de México hasta el noroeste Argentino. Durante muchos años se consideró que su origen fue en los actuales territorios de Ecuador y de Perú (Kami et al., 1995). Sin embargo, estudios más recientes de AFLP y variaciones de nucleótidos simples sugieren un origen mesoamericano (Rossi et al., 2009; Bitocchi et al., 2012). La domesticación del poroto habría ocurrido entre 4000 y 8000 años atrás (Martínez-Romero, 2003; Castro-Guerrero et al., 2016) en dos centros mayores de diversificación genética: la región Mesoamericana y la región Andina (Becerra-Velásquez and Gepts, 1994) (Figura 8). El poroto es considerado como un hospedador promiscuo, ya que puede ser nodulado por una amplia variedad de rizobios (Kaschuk et al., 2006). En Brasil, por ejemplo, existen reportes sobre la capacidad de *P. vulgaris* de establecer simbiosis con *Rhizobium tropici*, *R. etli*, *R. leguminosarum* y *R. giardinii*, con bacterias del género *Mesorhizobium* y *Sinorhizobium*, y con otras bacterias que podrían representar nuevas especies (Mostasso et al.; Grange and Hungria, 2004). Sin embargo, *R. etli* biovar *phaseoli* es la especie predominante que se encuentra asociada tanto a las variedades de poroto salvajes como a los cultivares agronómicos en México, Colombia y la parte sur de los Andes (Amarger, 2001). *R. etli* es una bacteria que contiene un cromosoma circular y seis plásmidos totalmente secuenciados (Gonzalez et al., 2006). La mayoría de los genes relacionados a la simbiosis se ubican sobre los plásmidos o en islas cromosómicas, probablemente adquiridos por transferencia horizontal (González et al., 2003). *R. etli* posee cerca de 200 familias de ADN repetitivo, las cuales han recombinado dando lugar a rearrreglos genómicos. Se especula que esta flexibilidad genómica podría estar involucrada en el potencial adaptativo de esta bacteria creando variabilidad. La rápida adaptación de las bacterias a condiciones de un medio ambiente cambiante es esencial para colonizar exitosamente nuevos nichos ecológicos. Ya que la adaptación local bacteriana es considerada como prueba de evolución, ha sido propuesta la coevolución entre el poroto y su rizobio huésped (Aguilar et al., 2004).

La asociación preferencial a través el reconocimiento específico influencia la eficiencia de nodulación, directa o indirectamente. Se ha demostrado que las plantas que han

coevolucionado con sus simbioses en un ambiente natural muestran una preferencia por aquellas cepas más eficientes, sugiriendo que existen mecanismos de selección que permiten evitar a los simbioses de menor calidad, determinando de esta forma la competencia de las bacterias en la rizósfera. La coevolución es considerada una de las mayores fuentes de creación de biodiversidad en el planeta, conectando los genomas de las especies que interactúan. El poroto, frijol o judía es una de las leguminosas en las que este fenómeno ha sido descrito en detalle (Brockwell et al., 1995). Se mostró que las plantas que coevolucionaron con su bacteria simbiótica muestra preferencia por la cepa más eficiente (Aguilar et al., 2004; Meschini et al., 2008). Eso sugiere la existencia de un mecanismo que permite discriminar y seleccionar entre variedades genéticas con diferente eficiencia en la nodulación.



**Figura 8: Evolución de *P. vulgaris* y *R. etli* en América Central y América del Sur.** *P. vulgaris* es nativo de América, más precisamente de la región centroamericana. Dos centros mayores de diversificación génica han sido caracterizados: el Centro Mesoamericano (poroto negro), que incluye los actuales territorios de México, América central y Colombia; y el Centro Andino (poroto blanco) que incluyen Ecuador, Perú y el norte de Argentina. Los análisis del polimorfismo del gen *nodC* de *R. etli* mostraron que las cepas con el alelo *nodC* tipo α son predominantes en los suelos mesoamericanos, mientras las cepas que portan los alelos tipo δ son predominantes de la región andina. Los colores rojos del mapa indican la cantidad de especies de *Phaseolus* actuales (resultados de Ramírez-Villegas et al., 2010). Figura adaptada a partir de Bitocchi et al., 2017, realizada con los datos publicados en Aguilar et al., 2004.

El uso de marcadores moleculares, como las phaseolinas (Blair et al., 2009; Llaca and Gepts, 1996; Burle et al., 2010) ha incrementado la comprensión del proceso de



domesticación del poroto común y su coevolución con los rizobios asociados. Las poblaciones de rizobios provenientes de diferentes regiones geográficas presentan un polimorfismo en el gen de nodulación *nodC* (Aguilar et al., 2004). Análisis de las comunidades bacterianas presentes en los suelos mostraron que las cepas aisladas a partir de suelos mesoamericanos llevan el alelo *nodC* tipo alfa (*nodC*  $\alpha$ ), mientras que el alelo *nodC* delta (*nodC*  $\delta$ ) es predominante en las cepas provenientes de suelos andinos (Figura 8).

El estudio de la eficiencia de nodulación de diferentes variedades de poroto, inoculadas con representantes de las cepas de *R. etli nodC*  $\alpha$  y  $\delta$  (SC15 y 55N1 respectivamente) mostró que las variedades mesoamericanas, tanto salvajes como domesticadas, son noduladas más eficientemente con la cepa SC15 *nodC*  $\alpha$ . Por el contrario, las variedades andinas muestran una mayor cantidad de nódulos con 55N1, una cepa *nodC*  $\delta$ . La preferencia de cepa podría entonces ser directamente relacionada con una nodulación más eficiente. Por otra parte, accesiones salvajes y cultivadas de poroto representantes de los centros de diversificación genética mesoamericano y andino fueron analizadas en ensayos de co-inoculación con cepas de *R. etli* que contienen los alelos *nodC*  $\alpha$  y  $\delta$ , provenientes de las diferentes zonas geográficas. Los resultados de estos experimentos de competencia mostraron una asociación preferencial de las variedades mesoamericanas de *P. vulgaris* con cepas de *R. etli* que predominan en los suelos de la misma región geográfica (Aguilar et al., 2004; Zanetti et al., 2010). Esta preferencia sugiere que las plantas provenientes de Mesoamérica pueden reconocer y seleccionar aquellos rizobios que son predominantes en la misma región geográfica.

### 4.2 Competitividad de los rizobios del suelo

En los ecosistemas naturales o en cultivos agrícolas, las plantas encuentran en los suelos diversas comunidades de microorganismos. La simbiosis fijadora de nitrógeno es una interacción altamente específica, ya que pocas combinaciones de especie y cepas hospedador-huésped llevan a una colonización exitosa. La eficiencia de la FBN de una especie leguminosa puede variar más de 10 veces, proveyendo un enorme potencial para la optimización del proceso. En el cultivo de leguminosas se utilizan cepas de rizobios altamente eficientes en cuanto a la fijación del nitrógeno, las que se agregan al suelo como inoculantes (Den Herder and Parniske, 2009). Sin embargo, estas cepas suelen ser poco competitivas y más del 90% de los nódulos formados en las plantas cultivadas son ocupados principalmente por las bacterias presentes naturalmente en el suelo, menos eficientes, pero mejores competidoras (Thies et al., 1991b, 1991a). Esto disminuye la efectividad del uso de bio-inoculantes en agricultura (revisado en Triplett and Sadowsky, 1992). La existencia de cepas con menor capacidad de competir en el suelo, pero que son capaces de establecer una interacción funcional, sugiere la existencia de señales que permitirían a las plantas discriminar entre cepas con menor o mayor grado de eficacia. La discriminación entre diferentes especies de rizobios comienza a establecerse en la rizósfera (Simms, 2002; Simms and Taylor, 2002), donde la planta tiene un rol activo en seleccionar y establecer simbiosis más eficientes que han sido favorecidas por coevolución (Aguilar et al., 2004). Mientras el reconocimiento de especies es determinada por los NF

(Radutoiu et al., 2007), la selección de una cepa de rizobio particular por la planta tiene un impacto significativo en la eficiencia de nodulación, la competitividad a campo y consecuentemente en la cantidad de nitrógeno fijado. Muchos de los genes involucrados en el establecimiento de la simbiosis han sido caracterizados y son comunes a las diferentes especies de leguminosas. Sin embargo, se conoce mucho menos acerca de los genes de la planta involucrados en la selección y/o preferencia de cepas.

### 4.3 Genes de *P. vulgaris* implicados en la preferencia de cepa

Estudios previos en nuestro laboratorio utilizando una estrategia de hibridación sustractiva permitieron identificar 41 genes que se expresan diferencial y específicamente en respuesta a la cepa SC15 a tiempos tempranos de la interacción (Meschini et al., 2008). La clasificación funcional de dichos genes indicó que los mismos corresponden a diferentes categorías funcionales, siendo las categorías más representadas las de síntesis y procesamiento de proteínas (20%), metabolismo primario (15%) y transducción de señales (12%). Uno de estos genes codifica por una pequeña GTPasa que pertenece a la familia Rab; la cual participa en la formación del crecimiento polar del pelo radical y es requerido para la reorientación del mismo en respuesta a *R. etli* en los tiempos tempranos de la interacción (Blanco et al., 2009). Los experimentos de silenciamiento del gen *RabA2* mostraron que las raíces de plantas que expresan un ARN interferencia (ARNi) *RabA2* no son capaces de formar nódulos, independientemente de la cepa de rizobio inoculada. En las plantas silenciadas se ve afectada la inducción de nodulinas a tiempos temprano, tales como ENOD40 y ERN. Estos resultados indicaron que *RabA2* es necesaria para las repuestas moleculares tempranas tanto en el epidermis que en las células corticales (Blanco et al., 2009). Recientemente se estudió con más detalles la función de esta GTPasa monomérica utilizando versiones constitutivamente activas y dominantes negativas, demostrando que se requiere de una *RabA2* funcional para la progresión del hilo de infección y la integridad de su membrana (Dalla Via et al., 2017). Otro gen caracterizado fue *NF-YC1* que codifica para un factor de transcripción con dominio de unión al ADN tipo CCAAT. Esta proteína reguladora interacciona con los promotores mediando la activación o inhibición transcripcional de genes, permitiendo a las células respondan a cambios ambientales. Se ha mostrado el papel de *NF-YC1* en la determinación de la preferencia de los cultivares mesoamericanos para cepas de rizobios que compiten y fijan el nitrógeno más eficientemente (Zanetti et al., 2010). Más precisamente, la reducción de los niveles de transcriptos de *NF-YC1* por ARNi en plantas mesoamericanas detiene el desarrollo del nódulo y provoca defectos en el proceso de infección con las cepas más y menos eficientes. Además, la sobreexpresión de este gen tiene un efecto positivo en la eficiencia de nodulación y la selección de cepas de *R. etli* que son normalmente menos eficaces y malos competidores. A su vez *NF-YC1* regula, directa o indirectamente, la expresión de genes del ciclo celular durante la formación del primordio del nódulo. Posteriormente, se demostró que *NF-YC1* se induce específicamente en las variedades mesoamericanas pero no en las andinas; sugiriendo que existen diferentes mecanismos de regulación dependiendo de la variedad. Análisis funcionales de *NF-YC1* indicaron que el silenciamiento de *NF-YC1* reduce la nodulación en la variedad andina Alubia (Mazziotta et al., 2013), de manera semejante a lo observado

previamente en el cultivar mesoamericano NAG12 (Zanetti et al., 2010). Sin embargo, contrariamente a lo observado en el cultivar mesoamericano, la sobreexpresión de *NF-YC1* tiene un impacto negativo en la nodulación del cultivar andino, independientemente de la cepa utilizada (Mazziotta et al., 2013). Estos resultados indican que las variaciones transcripcionales y funcionales de *NF-YC1* en diferentes variedades de *P. vulgaris* pueden modular positivamente o negativamente la formación de nódulos y/o la preferencia de cepa. La comparación genómica de las secuencias de las ORF de *NF-YC1* entre dos cultivares mesoamericano (NAG12) y andino (Alubia) no mostró diferencias nucleotídicas, por lo que las diferencias funcionales podrían encontrarse a nivel de los promotores o de las regiones transcritas no traducidas.

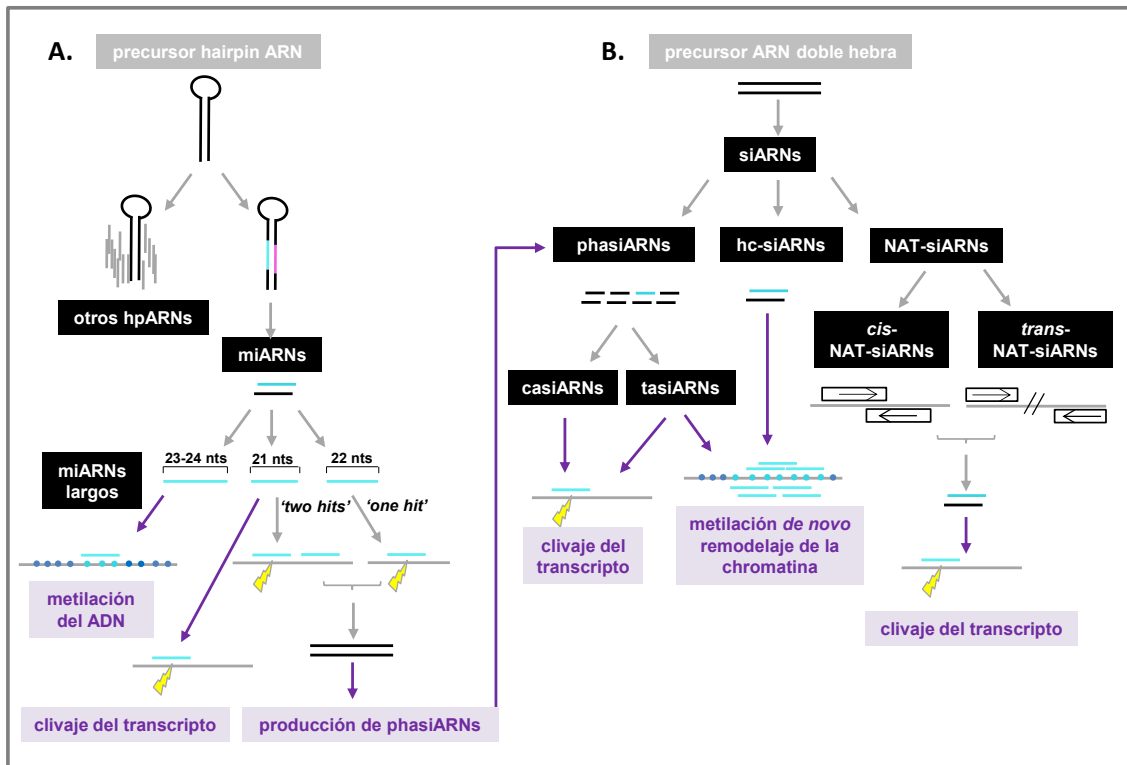
## 5. Pequeños ARNs no codificantes

### 5.1 Clasificación de los sARNs no codificantes

La regulación génica es un continuo que incluye diversos mecanismos que van desde la activación transcripcional hasta la degradación de los productos funcionales codificados en los genes. En la última década, los pequeños ARNs o sARNs (*small* ARNs) han emergido como reguladores claves de la expresión génica tanto a nivel transcripcional como post-transcripcional (Bartel, 2009; Megraw et al., 2016). Estas moléculas en su forma madura poseen una longitud de 20 a 24 nucleótidos (nts) y son generados a partir de *loci* genómicos diferentes mediante vías que pueden ser separadas genéticamente (Axtell, 2013). La primera distinción que se puede hacer entre las diferentes clases de sARNs es según el precursor que les da origen (Figura 9). Los sARNs derivan de un transcripto en simple hebra con una estructura de hairpin (caso de los miARNs y hpARNs) o a partir de un ARN doble hebra (los siARNs).

Si bien la biogénesis de los distintos sARNs es diferente (Bologna and Voinnet, 2014), todos ellos son incorporados en complejos inductores del silenciamiento a ARN (RISC, *RNA induced-silencing complexes*) que contiene un miembro de la familia de Argonauta (AGO) como componente principal (Chapman and Carrington, 2007; Fang and Qi, 2016). Los sARNs guían a las proteínas AGO hasta el ARN blanco usando la complementariedad de base como herramienta de búsqueda. A su vez, los complejos RISC regulan la expresión de los genes a nivel transcripcional y post transcripcional (Voinnet, 2009).

Los miARNs (de 20 a 24 nts) son procesados a partir del precursor de manera precisa para dar uno o pocos sARNs funcionales. Los sARNs que provienen del procesamiento impreciso de una hairpin no son considerados miARNs. Los miARNs largos (23 y 24 nts) han sido asociados a los mecanismos de metilación del ADN, depositando marcas represivas sobre la cromatina. Los miARNs de 21 y 22 nts son reguladores post-transcripcionales de la expresión génica, y actúan regulando la estabilidad o traducibilidad de los transcriptos blancos.



**Figura 9: Clasificación de los pequeños ARNs endógenos en plantas.** Los sARNs no codificantes en planta pueden ser diferenciados según su biogénesis, su tamaño o su función. A. sARNs procesados a partir de un precursor de ARN cuya estructura secundaria genera una estructura de *hairpin* (los hpARNs y los miARNs). B. Los siARNs son sARNs generados a partir de una doble hebra de ARN. Los phasiARNs son sARNs secundarios ya que son procesados a partir de un ARN doble hebra generado luego del corte de un miARN sobre un transcripto. Figura realizada a partir de Axtell, 2013.

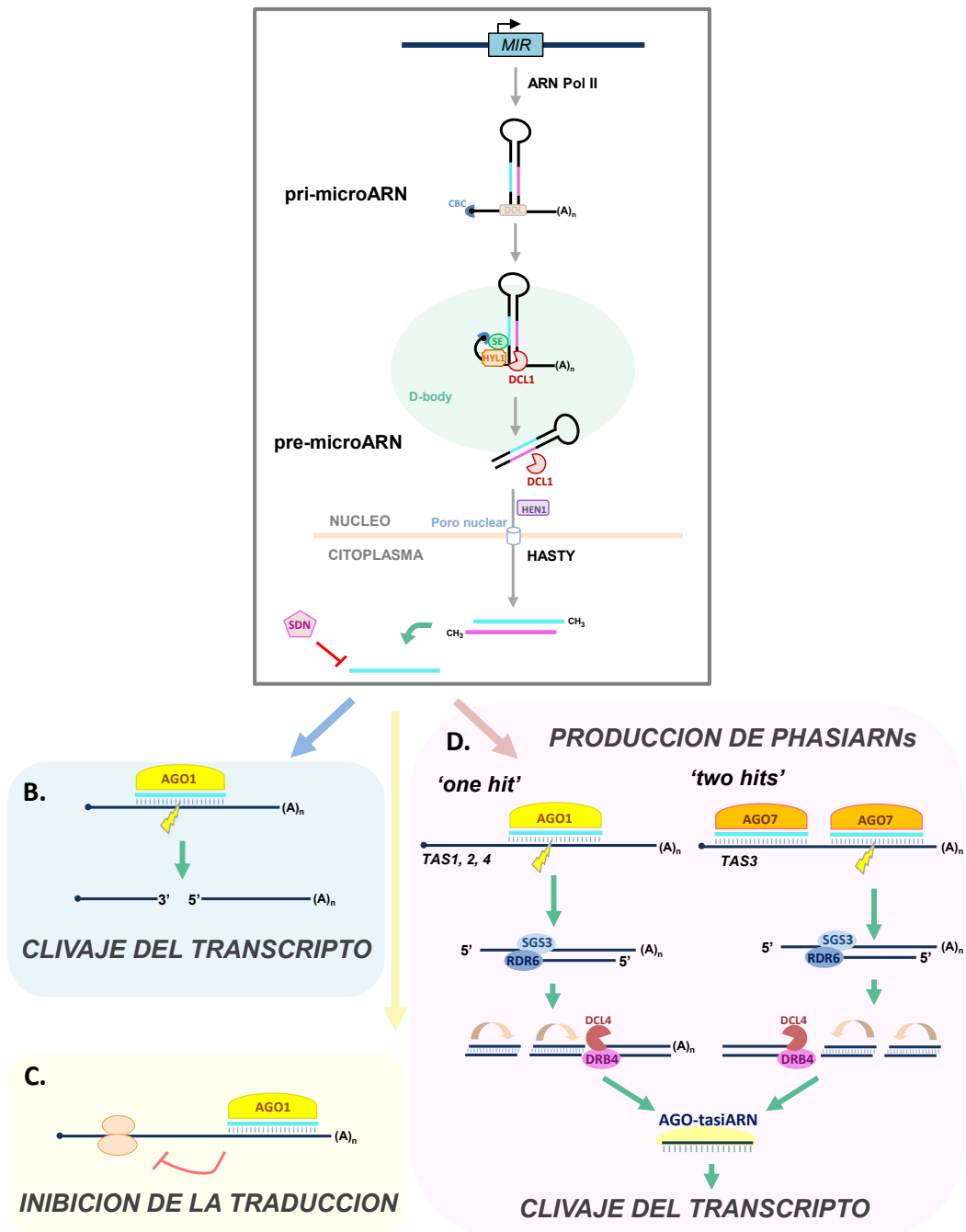
Otra gran categoría de sARNs es la que corresponde a los pequeños ARNs de interferencia o siARNs (*small interference RNAs*). Los siRNAs se generan a partir de ARN doble cadena, en su forma madura poseen una longitud de 21 a 24 nts y funcionan principalmente como supresores de la actividad de elementos móviles (transposones) y manteniendo la estabilidad del genoma vía metilación del ADN o modificaciones de la cromatina (Matzke et al., 2015; Deleris et al., 2016; Richard and Takken, 2017). Los siARNs endógenos provienen de un ARN largo en doble hebra, generado a partir del producto de una polimerasa dependiente de ARN (RdRP), una transcripción convergente o repeticiones de transcritos. Los siARNs participan generalmente a su propio silenciamiento, en el cual actúan sobre blancos de ADN o ARN correspondiendo a los *loci* de donde han sido generados. Los siARNs de heterocromatina (hc-siARNs), de 23 o 24 nts son producidos a partir de regiones intergénicas o repetitivas. Estos siARNs participan a la vía canónica de metilación del ADN dirigida por ARN (RdDM, *RNA directed DNA methylation*) involucrada en la metilación *de novo* del ADN. Los NAT-siARNs son producidos a partir de un precursor ARN de doble hebra formado por la hibridación de transcritos independientes y complementarios. Se diferencian los cis-NAT-siARNs de los trans-NAT-siARNs por el origen de su precursor, los dos actúan como represores transcripcional. Los precursores de los cis-NAT-siARNs son transcritos a partir de genes que se

superponen en polaridad opuesta. Los precursores de los trans-NAT-siARNs provienen de mensajeros de *loci* diferentes con complementariedad.

Los miARNs pueden actuar como disparadores de la producción de sARNs secundarios, los phasiARNs. Los phasiARNs son siARNs (para *phased small interference RNA*) que son procesados en fase a partir de un ARN doble hebra generado luego del corte de un miARN sobre un transcripto. Se distinguen los casiARNs (*cis-acting* siARNs) de los tasiARNs (*trans-acting* siARNs) según actúen sobre el mismo transcripto que les dio origen o sobre un transcripto blanco diferente, es decir en *cis* o en *trans*, respectivamente. Algunos tasiARNs están implicando en la vía no-canónica de RdDM, en la cual serán incorporados con una proteína AGO mediando la metilación *de novo* de genes *TAS*.

## 5.2 Biogénesis y modo de acción de los miARNs

Los sARNs mejor caracterizados son los miARNs. En plantas, estos son pequeños ARNs que poseen una longitud de 20 a 23 nts y funcionan principalmente como reguladores post-transcripcionales regulando negativamente la expresión de sus genes blanco. En la Figura 10 se muestra un esquema de la biogénesis y el modo de acción de los miARNs en plantas. Los transcriptos primarios de miARNs (los pri-miARNs) son sintetizados por la ARN polimerasa II a partir de la expresión de los genes *MIR*, que suelen ser unidades transcripcionales independientes (Figura 10A). Los miARNs primarios son procesados mediante la adición de una cola de poliA en el extremo 3' y un cap en el extremo 5'. La proteína nuclear DDL de unión a ARN (por *DAWDLE*) estabiliza el transcripto. Los pri-miARNs son convertidos en precursores de miARNs (premiARNs) en los centros de procesamientos de los miARNs (*Dicing-bodies*) (Fang and Spector, 2007) por un complejo ribonucleoproteico orquestado por una RNAsa de tipo III denominada DCL1 (*Dicer Like 1*). DCL1 contiene dos NLS putativos (*Nuclear Localization Signal*), por lo que es generalmente aceptado que las diferentes etapas de procesamientos de los miARNs de plantas ocurren todas en el núcleo (Schauer et al., 2002). La proteína HYL1 (por *HYPONASTIC LEAVES1*) reconoce las estructuras de ARN en doble hebra, y junto a SE (por *SERRATE*) una proteína *zinc-finger*, interactúa con DCL1. HEN1 (*Hua Enhancer 1*) metila el extremo 3' del dúplex del miARN, protegiéndolo de la degradación de la exonucleasa SDN (*Small RNA Degrading Nuclease*). El dúplex del miARN con su ARN complementario, denominado miARN asterisco (miARN\*) es exportado al citoplasma a través del poro nuclear que contiene a HASTY, una proteína de planta homóloga a la exportin5 de animales (Park et al., 2005). El miARN maduro se asocia a alguna de las proteínas AGO y el complejo resultante actúa reprimiendo la expresión génica por clivaje del transcripto blanco (Figura 10B) o de inhibición de la traducción (Figura 10C). En el caso del clivaje de transcripto, el corte se produce en la región central de sitio de unión entre el miARN y su ARN blanco, generalmente entre las posiciones 10 y 11 a partir del extremo 5' del miARN (Llave et al., 2002; Tang et al., 2003; Kasschau et al., 2003).



**Figura 10: Biogénesis y modo de acción de los miARNs.** A. La biogénesis de los miARNs ocurre en el núcleo a partir de la expresión de un gen *MIR*. La ARN polimerasa II produce los miARNs primarios con una cola poliA en el 3' y una cap en el 5', al cual se asocia el complejo de unión al cap (CBC *por Cap Binding Complex*). DCL1 orquesta la conversión del miARN primario al precursor de miARN en el *D-body* (área circular verde). El dúplex (miARN maduro en celeste, asociado al miARN\* en violeta) está exportado al citoplasma por el poro nuclear mediante un mecanismo donde participa la proteína HASTY. El miARN maduro incorporado a la proteína AGO lleva a cabo las reacciones de silenciamiento de represión de la expresión de los genes por inhibición transcripcional (en B); o de inhibición de la traducción (en C). D. Los miARNs son igualmente responsables de la biogénesis de pequeños ARNs secundarios: los phasiARNs. Dos vías son actualmente reportadas: el modelo "one hit" (a la izquierda), iniciada por un miARN de 22 nts y el modelo "two hits" (a la derecha), por un miARN de 21 nts.

Como se mencionó en la sección anterior, otra función específica de algunos miARNs de 21 o 22 nts es promover la biogénesis de pequeños ARNs secundarios de 21 nts procesados en fase de a 21 nts a partir del sitio de clivaje del miARN, denominados los phasiARNs (Figura 10D). Dentro de estos se han identificado cuatro phasiARNs que actúan en trans, los cuales son procesados a partir de los transcritos *TAS1*, *TAS2*, *TAS3* y *TAS4*. Se han propuesto dos modelos para la biogénesis de los tasiARNs. En el modelo “*one hit*”, un miARN de 22 nts asociado a la proteína AGO1 se une al sitio de unión de un mensajero non-codificante (ej. *TAS1*, 2, 3) y lo corta. RDR6 (por *RNA-dependant RNA polymerase 6*) interactúa con SGS3 (*suppressor of gene silencing 3*) y sintetiza el fragmento de ARN en doble hebra (dsARN). El dsARN es el sustrato de DCL4 (*Dicer Like 4*), lo cual interactúa con DRB4 (*DICER RNA binding factor4*) para procesar los siARNs (*small interference*) (phasi o tasiARNs) en fase de a 21 nts respecto al sitio de corte. En el modelo “*two hits*”, el miARN de 21 nts asociado a la proteína AGO7 reconoce dos sitios blancos. El transcripto (ej. *TAS3*) es clivado en el sitio más próximo al extremo 3' del transcripto. El producto 5' de clivaje es copiado por RDR6 con ayuda de SGS3. EL ARN doble cadena es procesado en fase de a 21 nts a partir del sitio de clivaje del miARN. En el modelo “*one hit*” el procesamiento se efectúa en la dirección 3' respecto del sitio de clivaje (*downstream*), mientras en el modelo “*two hits*” se realiza hacia el 5' del sitio de clivaje (*upstream*). El siARN secundario es cargado con la proteína AGO para reconocer y cortar un transcripto complementario generalmente en *trans*.

### 5.3 Diferencias entre miARNs de plantas y animales.

Tanto en animales como en plantas, los genes que dan lugar a los miARNs o genes *MIR* son genes endógenos, presentes en los genomas, y son transcritos por la ARN polimerasa II dando lugar al transcripto primario de miARN (pri-miARN) que forman estructuras secundarias imperfectas (Lee et al., 2004). Los precursores primarios de miARNs de plantas son en su mayoría generados a partir de una unidad transcripcional independiente (Xie, 2005; Szarynska et al., 2009); mientras los de animales provienen de intrones, zonas no traducidas de regiones codificantes (UTR por *Untranslated Region*) o precursores en tándem (Kim et al., 2009). Tanto en plantas como en animales los pri-miARNs llevan una cap en 5' y son poliadenilados. La proteína DDL (por *DAWDLE*) estabiliza el transcripto en el núcleo (Yu et al., 2008). Como se mencionó anteriormente en la sección 5.2 de esta Introducción, los pri-miARNs de plantas son procesados por DCL1 en un precursor de miARN y luego nuevamente por DCL1 para dar el miARN maduro (Kurihara and Watanabe, 2004). En la biogénesis de miARNs de animales el primer corte del pri-miARN al premiARN es catalizada por la RNAsa Drosha que no es similar a Dicer en su secuencia pero si en su funcionalidad (Lee et al., 2003). En animales contrariamente a la biogénesis de miARNs de plantas, el procesamiento esta compartimentalizado (Axtell et al., 2011; Bologna et al., 2013). Luego del primer clivaje del pri-miARN en el núcleo por Drosha, el factor de exportación nuclear (Exp5) exporta el premiARN al citoplasma, a donde estará procesado por Dicer en dúplex de miARN. En *Drosophila* un solo producto del procesamiento de Dicer se acumula a partir de una sola hebra del dsARN (Hutvagner et al., 2001) mientras que el miARN\* es rápidamente degradado (Lau et al., 2001; Mourelatos et al., 2002). Al contrario el procesamiento de los premiARNs de plantas a

partir de los cortes catalizados por DCL1 produce una molécula de ARN doble cadena con complementariedad perfecta y dos nucleótidos protuberantes en los extremos 3' (Elbashir et al., 2001). En plantas cobra relevancia el control de la estabilidad de los miARNs por acción de las exonucleasas SDN (por *Small RNA Degrading Nuclease*) (Ramachandran and Chen, 2008). La supresión simultánea de los 3 miembros de la familia génica *SDN* en *Arabidopsis* causa defectos severos en el desarrollo, ocasionados por la sobre acumulación miARNs. Las proteínas SND son altamente conservadas en eucariotas, y sus homólogos en animales también podrían participar en la degradación de los sARNs, particularmente de los miARNs. Sin embargo las nucleasas Eri-1 (Enhanced RNAi-1) identificadas originalmente en *C. elegans* (Kennedy et al., 2004) difieren de las SDNs de plantas (Ramachandran and Chen, 2008), ya que a diferencia de las SDN las proteínas Eri-1 degradan específicamente los dúplex de siARNs. En animales los siARNs y piwiARNs serían O-metilados por el ortólogo de HEN1, sin embargo, no se han identificado miARNs de animales que lleven esta modificación en su extremo 3' (Horwich et al., 2007; Aravin et al., 2007).

Los precursores de miARNs suelen divergir considerablemente en las regiones afuera de la secuencia del miARN. La presión de selección se evidencia esencialmente en el dúplex miARN/miARN\*, resultando en mutaciones de muy pocas bases en este segmento, conservando la total tendencia a formar la doble hebra de ARN (Reinhart et al., 2002). El tamaño de las estructuras de tallo y horquilla de premiARNs de *Arabidopsis* son más variables que las de los premiARNs de animales y son generalmente mucho más largos. Por ejemplo, el miR169 es uno de los miARNs conservados más largos en *Arabidopsis* con  $\approx 190$  nts, mientras que en *C. elegans*, los premiARNs poseen sólo  $\approx 70$  nts con el miARN maduro ubicado a  $\approx 2-10$  bp del bucle terminal de la estructura de tallo y horquilla (Lau et al., 2001).

Los miARNs de plantas poseen una pequeña cantidad de transcriptos blancos generalmente miembros duplicados de una familia multigénica (Voinnet, 2009), uniéndose en la mayoría de los casos a un único sitio que exhibe alta complementariedad de bases (no más de 5 *mismatches*) con el miARN (Fahlgren and Carrington, 2010). Estos sitios de unión a miARNs de plantas se ubican generalmente en las regiones codificantes de los transcriptos blancos (Mallory et al., 2004; Schwab et al., 2005). Por el contrario, los miARNs de animales presentan una complementariedad con sus blancos limitada a la zona central (6 nts) y cada transcripto puede contener sitios de unión para varios miARNs (Rajewsky, 2006). Las diferencias observadas entre los miARNs de plantas y de animales sugieren una aparición independiente entre los dos linajes (Axtell et al., 2011; Moran et al., 2017). En plantas, los miARNs conservados a lo largo de la evolución son esencialmente aquellos que regulan factores de transcripción implicados en el desarrollo del embrión, vegetativo o floral (Chen et al., 2005; Jones-Rhoades et al., 2006; Mallory and Vaucheret, 2006). En estos casos, no solamente el miARN se encuentra conservado en múltiples especies sino también el mecanismo de regulación del miARN sobre su ARN blanco (Jones-Rhoades and Bartel, 2004).



#### 5.4. miARNs implicados en la simbiosis

Se han descritos miARNs conservados en la simbiosis entre leguminosas y rizobios, otros involucrados preferencialmente en un tipo de nodulación (determinada o indeterminada), o específicos de especies. Algunos miARNs de leguminosas han sido involucrados en el organogénesis y desarrollo del nódulo tales como miR169 y miR166 de *M. truncatula* (Combier et al., 2006; Boualem et al., 2008) y los miR482, miR1512, y miR1515 de soja (Li et al., 2010). Muchos de los genes blancos de estos miARNs son factores de transcripción relacionados con el establecimiento de la simbiosis. Dentro de este grupo, uno de los primeros miARN identificado de la simbiosis en *M. truncatula* es el miR169, regulador de NF-YA1, un factor de transcripción de la familia de los NF-Ys (Baudin et al., 2015). En *M. truncatula*, NF-YA1 (*HAP2-1*) codifica para un factor de transcripción simbiótico que actúa durante el desarrollo del nódulo controlando la persistencia del meristemo (Combier et al., 2006). El miR169 restringe la expresión de NF-YA1 en la zona meristemática de los nódulos indeterminados en los eventos tempranos de la simbiosis. El miR169 presenta alto nivel de conservación evolutiva entre las especies de plantas (Li et al., 2008; Pant et al., 2009). NF-YA pertenece a una familia multigénica y actúa en complejos heterotrimericos con las subunidades B y C, involucrados en varios procesos de desarrollo y respuestas a estímulos del medio ambiente (Zanetti et al., 2017). En *P. vulgaris*, la familia NF-Y y la regulación de la subunidad NF-YA1 por el miR169 estaría conservada (Rípodas et al., 2015). Como NF-YC1 (ver sección 4.3), NF-YA1 contribuye a la selección de cepas de rizobio que son más eficientes en la formación de nódulos (Rípodas, Tesis Doctoral).

En *M. truncatula* se identificó también el miR166, con una función en el desarrollo de nódulos y de las raíces laterales a través de la regulación de los HD-ZIP III (por *class-III homeodomain-leucine zipper*) (Boualem et al., 2008). En soja, se ha reportado que la sobreexpresión de los miARNs miR482, miR1512 y miR1515 incrementa significativamente el número de nódulos respecto a las plantas salvajes (Li et al., 2010). Específicamente, el miR482 regula genes asociados a la resistencia primaria, y podría estar involucrado en la modulación de los mecanismos de defensa de la planta para permitir el establecimiento de la simbiosis. El miR1512 regula proteínas de unión de tipo calmodulina implicados en la señalización celular y en el tráfico de membranas. Finalmente, el miR1515 regula una proteína homóloga a DCL2 de *A. thaliana*, la cual interviene en el control de virus de ARN y silenciamiento (Diaz-Pendon et al., 2007; Moissiard et al., 2007). Sin embargo se desconoce su función en la interacción simbiótica.

A otros miARNs se les han asignados funciones en el desarrollo de los nódulos relacionados a la regulación de los niveles de las fitohormonas, como es el caso de las auxinas. Un miARN relacionado con las auxinas es el miR390. El miR390 es un miARN conservado en plantas cuyo blanco es un mRNA no codante, denominado TAS3. El clivaje de TAS3 por el miR390 genera la producción de los tasiARFs. Estos tasiARNs regulan mediante una regulación transcripcional los factores de transcripción regulados por auxinas, los ARFs (Montgomery et al., 2008). En *M. truncatula* los niveles estacionarios de los tasiARFs disminuyen en respuesta a la inoculación con *S. meliloti* a tiempos tempranos, mientras sus blancos, los ARF2, ARF3 y ARF4 aumentan significativamente (Reynoso et al., 2013). A su vez, la activación de la vía miR390/TAS3, mediante la sobreexpresión ectópica del miR390, regula positivamente

el crecimiento de las raíces laterales y negativamente la nodulación (menos cantidad de nódulos formados y nódulos más pequeños) y la infección bacteriana (menos hilos de infección). A su vez, las raíces que sobreexpresan el miR390 presentan una disminución significativa en los niveles de expresión de los transcritos *ARF2/3/4* (Hobecker et al., 2017).

Otro ejemplo de miARNs regulados por hormonas es el miR160, un regulador negativo de la organogénesis nodular, el cual posee a los *ARF10*, *ARF16* y *ARF17* como mARNs blancos. El miR160 controla no solamente la iniciación de los nódulos determinados e indeterminados sino que también el desarrollo de los nódulos en estadios más avanzados (Bustos-Sanmamed et al., 2013; Turner et al., 2013b). Al contrario, el miR164, el primer miARN asociado a la nodulación relacionado a las auxinas tendría una función en la formación de los nódulos indeterminados de *M. truncatula* pero no afectaría la nodulación de los nódulos determinados formados en soja (D'haeseleer et al., 2011; Mao et al., 2013). Otro ejemplo es el miR393, el cual tendría funciones duales en las simbiosis. Este miARN actúa regulando al receptor de auxinas en *M. truncatula* y sería requerido para la formación de los arbuscúlos de la simbiosis endomicorrizas (Etemadi et al., 2014). En *G. max*, el miR393j-3p formado a partir de la otra hebra del precursor (o miR393 asterisco) tendría una función en el control de la actividad del meristema independiente de las auxinas (Yan et al., 2015).

Un caso interesante es el del miR172 que presenta múltiples funciones en el desarrollo y en la simbiosis a través de la represión de los factores de transcripción APETALA2 (AP2) (Aukerman and Sakai, 2003; Huijser and Schmid, 2011). En leguminosas el miR172 pertenece a una familia multigénica (entre 4 y 12 miembros según la especie en la base de datos de miRBase); y 5, 6 y 10 genes de *AP2* presentan sitios de unión al miR172 en *G. max*, *L. japonicus*, y *P. vulgaris*, respectivamente (Song et al., 2011; Holt et al., 2015; Nova-Franco et al., 2015). El miR172 es un regulador positivo de la simbiosis rizobiana, que interviene en diferentes procesos de la simbiosis. La sobreexpresión de los precursores del miR172 resulta en la hiper-deformación de los pelos radicales e hiper-infección en *G. max* y *P. vulgaris* (Wang et al., 2014; Nova-Franco et al., 2015); pero no en *L. japonicus* (Holt et al., 2015), lo que sugiere funciones específicas de especie. Sin embargo, en todas las leguminosas estudiadas, el miR172 controla la cantidad de nódulos formados y la eficiencia de fijación del nitrógeno. Es el caso de *P. vulgaris* para el cual, este fenotipo (aumento de la cantidad de nódulos formados y de la fijación de nitrógeno) es causado por la expresión ectópica del miR172c, y correlacionado con la represión del gen *PvAP2.1* (Nova-Franco et al., 2015). Finalmente, se ha demostrado en varias especies de leguminosas que la expresión de las isoformas de miR172 que controlan la nodulación son dependientes de la señalización activada por NF (Wang et al., 2014; Holt et al., 2015; Nova-Franco et al., 2015). Es decir que existen evidencias sustanciales que soportan la hipótesis que la regulación mediada por miARNs constituye un nivel adicional de regulación de la expresión génica que lleva al establecimiento y desarrollo de la simbiosis.

# *Hipótesis y Objetivos*

## Hipótesis

En los últimos años se ha avanzado significativamente en el conocimiento de los mecanismos moleculares y fisiológicos de la nodulación (revisado en Oldroyd et al., 2011). Sin embargo, la disección de las rutas de señalización que llevan a la infección y la formación del nódulo no ha redundado en un mejoramiento de la eficiencia agronómica de las interacciones rizobio-leguminosa. La agricultura continúa favoreciendo el uso de fertilizantes sintéticos, que son caros y poco seguros para el ambiente. Esta situación nos obliga replantear el enfoque con el que se ha encarado el estudio de los aspectos moleculares de la asociación entre leguminosas y rizobios.

Teniendo en cuenta los sARNs actúan como reguladores maestros del desarrollo y la respuesta a estímulos en plantas y que la simbiosis involucra tanto un proceso de organogénesis del nódulo como el reconocimiento específico de los micro-simbiontes que permitan suprimir y promover la infección rizobiana, nuestra hipótesis de trabajo es que estos sARNs participarían de los mecanismos moleculares que les permiten a la planta entre cepas más o menos eficientes en la formación de nódulos. La identificación de los sARNs y sus blancos putativos permitiría dilucidar las redes regulatorias que actúan a nivel transcripcional y/o post-transcripcional durante la selección de cepas.

## Objetivo general

El objetivo general del trabajo es identificar y caracterizar a nivel funcional sARNs (incluyendo siARNs y miARNs) regulados diferencialmente durante el proceso de reconocimiento e infección de cepas de alta y baja eficiencia en la nodulación en el sistema *P. vulgaris*-*R. etli*.

## Objetivos específicos:

Para cumplimentar con este objetivo general se plantearon los siguientes objetivos específicos:

1. Generar, secuenciar y analizar bibliotecas para secuenciación masiva (Illumina) de sARNs involucrados en la interacción entre *P. vulgaris* y cepas de *R. etli* de alta y baja eficiencia en la nodulación. En este objetivo se buscará identificar sARNs, incluyendo tanto siARNs como miARNs conservados y nuevos, que se acumulan diferencialmente en las raíces de *P. vulgaris* en respuesta a cada cepa.
2. Identificar los transcritos blancos putativos de los sARNs candidatos mediante análisis bioinformático y caracterizar su perfil de expresión mediante qRT-PCR.
3. Caracterizar la función de un grupo de miARNs nuevos o conservados seleccionados en la asociación preferencial entre plantas de origen Mesoamericano por cepas de *R. etli* altamente eficientes en la formación de nódulos.

En su conjunto, estos objetivos contribuirán a dilucidar las bases moleculares responsables del reconocimiento por parte de las plantas leguminosas de cepas de rizobio que son altamente eficientes y competitivas respecto de otras menos eficientes.

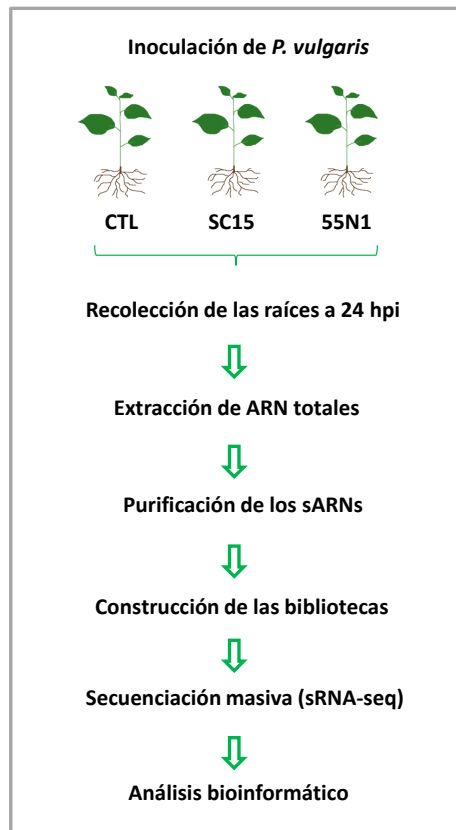
# *Capítulo I*

# Bibliotecas de pequeños ARNs

## 1. Construcción de las bibliotecas de pequeños ARNs

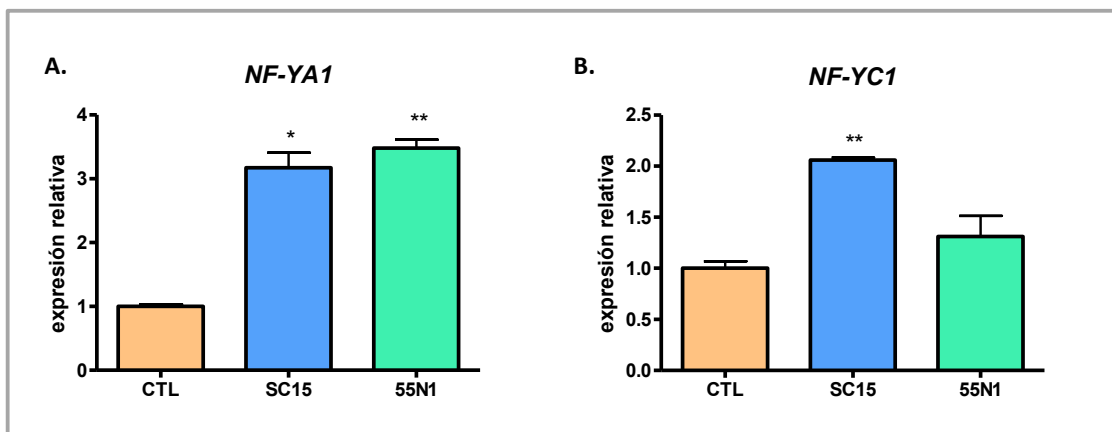
### 1.1 Generación de las muestras

Con el objetivo de caracterizar los cambios en las poblaciones de pequeños ARNs en raíces de *P. vulgaris* frente a la inoculación con cepas de *R. etli* de alta y baja eficiencia en la nodulación se construyeron bibliotecas de pequeños ARNs para secuenciación masiva utilizando la tecnología de la compañía Illumina. En primer lugar se generó tejido de raíces salvajes de *P. vulgaris* de la variedad mesoamericana NAG12 colectado a las 24 horas después de la inoculación (24 hpi por horas post inoculación) con dos cepas de *R. etli* o con el medio de crecimiento de los rizobios como condición control (Figura 11). Para la inoculación se utilizaron la cepa más eficiente SC15, portadora del alelo *nodC* tipo  $\alpha$ , y la cepa menos eficiente 55N1, portadora del alelo *nodC* tipo  $\delta$ . A las 24 hpi, los pelos radicales comienzan a deformarse en respuesta al rizobio, aunque los hilos de infección no son visibles todavía. A este tiempo, muchas de las respuestas moleculares asociadas a la infección y la organogénesis del nódulo como la inducción de nodulinas tempranas ya se encuentran activadas en las raíces de *P. vulgaris* (Meschini et al., 2008). Se obtuvieron dos réplicas biológicas independientes a partir de tejido de la zona de pelos radicales en crecimiento (zona II) utilizando un mínimo de 5 plantas por condición (ver Figura 64 de la sección Materiales y Métodos).



**Figura 11: Esquema que ilustra la estrategia utilizada para la identificación de pequeños ARNs de las etapas tempranas de la simbiosis eficiente entre *P. vulgaris* y *R. etli*.** Las bibliotecas de pequeños ARNs (sARNs) fueron construidas a partir de raíces de plantas mesoamericanas inoculadas con la cepa de *R. etli* SC15 (cepa más eficiente), 55N1 (cepa menos eficiente) o con el medio de cultivo YEM libre de bacteria como control (CTL) y recolectadas 24 horas después de la inoculación (hpi).

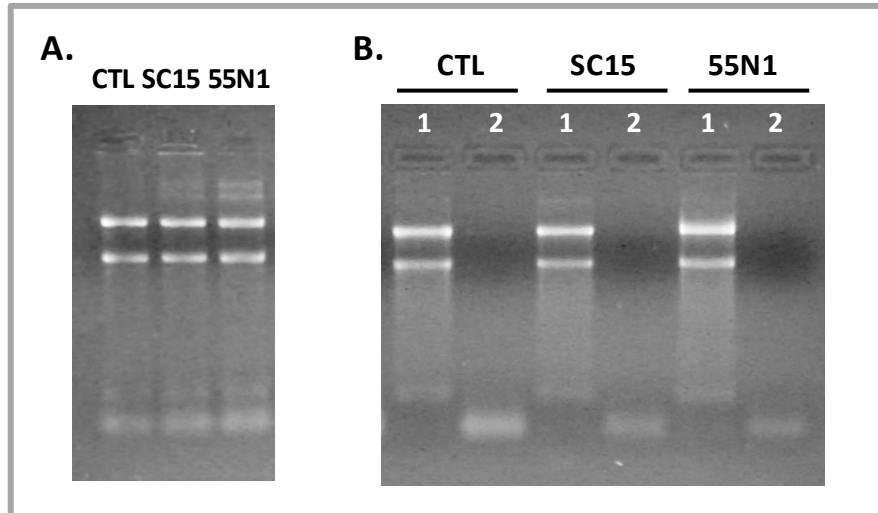
Las dos réplicas biológicas fueron obtenidas independientemente en diferentes épocas del año. Más precisamente, la segunda réplica biológica fue realizada a partir de tejido generado previamente por la Dra. Dalla Vía durante el transcurso de su Tesis doctoral para el análisis transcriptómico por RNA-seq de raíces de *P. vulgaris* a tiempos tempranos de la interacción con cepas de *R. etli* (Dalla Via et al., 2015). En esta réplica biológica se habían determinado por transcripción reversa seguida de PCR cuantitativa (RT-qPCR), los niveles de acumulación de las nodulinas *ENOD40*, *NIN* y *ERN1*, las cuales fueron significativamente inducidas en respuesta al rizobio. Para validar la eficiencia de la infección bacteriana en la réplica biológica generada durante el transcurso de esta Tesis, se evaluaron los niveles de expresión de dos genes de la familia de los *NF-Ys*, los cuales se inducen en respuesta al rizobio a las 24 hpi (Rípodas et al., 2015). En esta réplica biológica, *NF-YA1* se indujo aproximadamente 3 veces en comparación al control independientemente de la cepa de rizobio utilizada para la inoculación (Figura 12A). Esta inducción fue estadísticamente significativa en un test *t-Student* para ambas cepas. Por otra parte, *NF-YC1* presentó una inducción significativa de 2 veces a las 24 hpi, únicamente con la cepa más eficiente SC15 (Figura 12B). Este resultado coincide con los datos de expresión reportados previamente en Zanetti et al. (2010), donde se mostró que *NF-YC1* desempeña una función relevante en la preferencia de cepa observada en la variedad mesoamericana de *P. vulgaris*,



**Figura 12: Análisis de la expresión de *NF-Ys* a tiempos tempranos de la infección rizobiana.** A. Niveles relativos de acumulación del mRNA que codifica *NF-YA1*. B. Niveles relativos de acumulación del mRNA que codifica *NF-YC1*. Los niveles de transcritos fueron cuantificados mediante RT-qPCR en muestras de raíces de *P. vulgaris* inoculadas durante 2 horas con el medio de crecimiento bacteriano (CTL), con la cepa de *R. etli* SC15 o con la cepa 55N1. Los valores muestran la expresión relativa al gen de referencia *EF1α* y fueron normalizados por los valores de la muestra control (CTL). Los asteriscos indican diferencias significativas del control respecto las condiciones de inoculación en un t-test (\*,  $p < 0,05$ ; \*\*,  $p < 0,01$ ).

A partir de estas muestras de tejido se asilaron los ARNs totales utilizando el método de extracción con Trizol para conservar los pequeños ARNs. Para cada muestra se realizaron 4 extracciones de ARN totales para obtener la cantidad de material requerida para la ulterior purificación de pequeños ARNs. La calidad e integridad del ARN total se analizó mediante

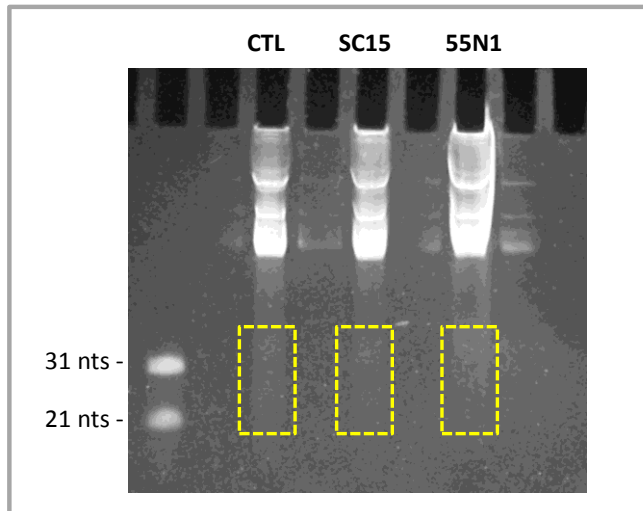
electroforesis en un gel de agarosa (Figura 13A) donde los ARN ribosomales se visualizaron por tinción con bromuro de etidio. Los ARNs totales de cada muestra fueron *poolados* y utilizados como material de partida para la separación de ARNs de alto y bajo peso molecular (ver el protocolo de la Figura 65 de la sección Materiales y Métodos). Para la separación de los pequeños ARNs, se utilizó polietilenglicol (PEG) siguiendo el protocolo de Lu et al., 2007 según se detalla en la sección 9 “Construcción de bibliotecas de pequeños ARNs” de Materiales y Métodos. La efectividad de la separación de los ARNs de alto y bajo peso molecular fue evaluada mediante electroforesis en un gel de agarosa (Figura 13B).



**Figura 13: Obtención de ARNs de bajo peso molecular.** A. La extracción del ARN total con Trizol incluye la población de pequeños ARNs visibles en la región inferior del gel. La obtención de las bandas de ARN ribosómico (ARNr 25S y 18S) bien definidas, con una banda del ARNr 25S más intensa que la 18S indican una buena calidad del ARN obtenido. B. Los ARNs de alto peso molecular (señalados en 1) fueron precipitados con PEG y NaCl, mientras que los ARNs de bajo peso molecular quedaron en solución (2). La integridad de los ARNs (totales, de alto y bajo peso molecular) fue validada por electroforesis en gels de agarosa de 1,5%. Se sembraron 0,5  $\mu$ l de ARN en cada calle. Los ARNs se revelaron por tinción con bromuro de etidio y visualización bajo luz UV.

Si bien los ARNs de alto peso molecular fueron eliminados de las muestras, los ARNs de bajo peso molecular constituyen una población diversa con una gran amplitud de tamaño. De todos ellos, los pequeños ARN de 20 a 24 nts constituyen la población de interés para la construcción de las bibliotecas. Para separar y purificar los ARNs de 21 a 24 nts se llevó a cabo una electroforesis en un gel no desnaturizante de poliacrilamida/urea (Figura 14). Para determinar el rango de tamaño de los pequeños ARNs de interés se utilizaron oligonucleótidos simple hebra de 21 y 31 nts como patrón de tamaño molecular. Los ácidos nucleicos se tiñeron con bromuro de etidio y se visualizaron bajo luz UV, permitiendo identificar las poblaciones de pequeños ARNs en el rango definido por el patrón de tamaño molecular. Con un bisturí se escindieron y colectaron las regiones de gel correspondientes al tamaño de interés ( $\approx$ 35 a 18 nts) para cada muestra. Esta población de pequeños ARNs fue purificada del gel y utilizada para la generación de bibliotecas de pequeños ARNs para secuenciación masiva (sRNA-seq).

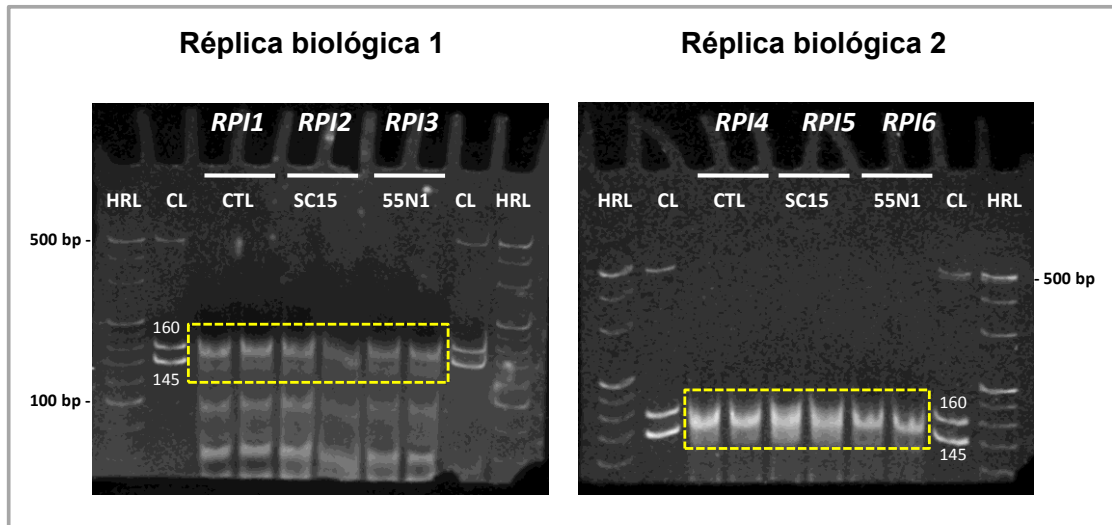




**Figura 14: Purificación de los pequeños ARNs.** Los ARNs de bajo peso molecular (10  $\mu$ l de cada muestra) fueron sembrados en un gel de poliacrilamida/urea 15% (p/v). Los pequeños ARNs fueron seleccionados por su tamaño ( $\approx$ 35 a 18 nts, recuadro amarillo) mediante el uso de oligonucleótidos de 21 y 31 nts.

## 1.2 Construcción de las bibliotecas de pequeños ARNs

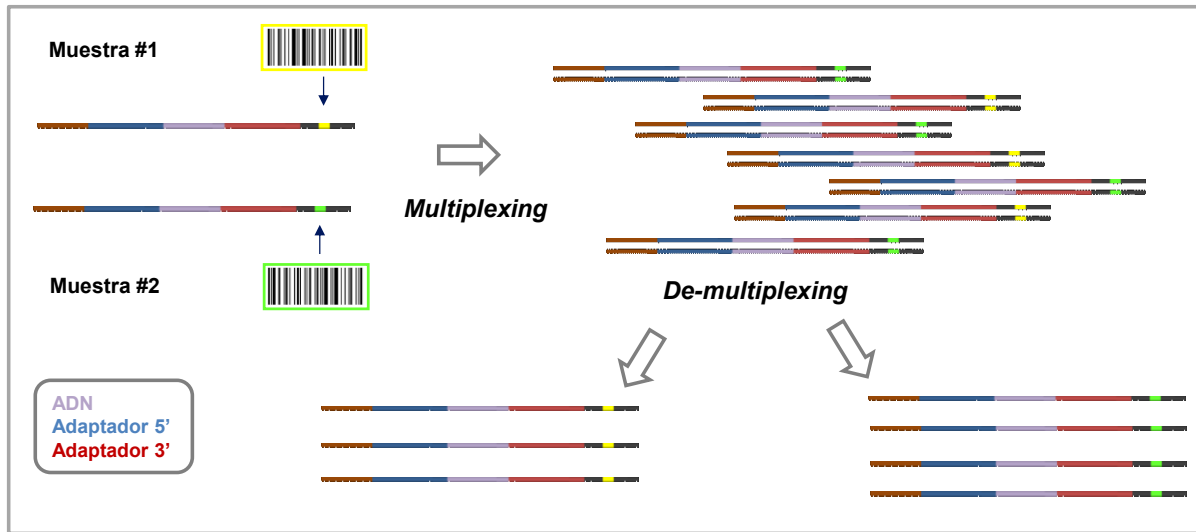
Para la generación de las bibliotecas de pequeños ARNs se utilizó el kit *TruSeq Small RNA preparation* de Illumina (según se detalla en la sección Materiales y Métodos). Este protocolo está basado en la particularidad que las moléculas de pequeños ARNs poseen un grupo fosfato en el extremo 5' y un hidroxilo en el extremo 3' del RNA a los cuales los adaptadores de Illumina se ligan específicamente. Las instrucciones del proveedor indican que las bibliotecas pueden ser generadas a partir de ARN total o de pequeños ARNs purificados. Usando el ARN total como muestra de partida para la construcción de las bibliotecas no se obtuvo el resultado esperado, ya que los productos correspondieron esencialmente al tamaño de adaptadores libres y dímeros u oligómeros de adaptadores. Aunque las concentraciones de ARN totales habían sido elevadas, los pequeños ARNs se encontraban muy diluidos en comparación a los ARNs de mayor peso molecular. Los dímeros/oligómeros de adaptadores compitieron probablemente en la reacción de RT-PCR. La puesta a punto de la purificación de los pequeños ARNs en las muestras de raíces de *P. vulgaris* fue entonces indispensable para la realización de las bibliotecas de pequeños ARNs. Para lidiar con la baja eficiencia de la ligación de los pequeños ARNs y los adaptadores observada, se purificaron las poblaciones de pequeños ARNs como se detalla en la sección anterior de Resultados. Luego de la ligación de los adaptadores de ARN, la primera hebra de ADN complementario (ADNc) fue sintetizada por transcripción reversa (RT) y el producto amplificado por PCR. Los productos de amplificación por PCR fueron visualizados en geles de poliacrilamida 15 % (p/v) (Figura 15). Estos productos de amplificación fueron escindidos del gel, purificados y almacenados a -20 °C.



**Figura 15: Purificación y precipitación del ADN.** Los productos de la PCR fueron sembrados en un gel de poliacrilamida 6 % (p/v), para seleccionar los ADNc de interés. Estas bandas de amplificación de ~140 a 160 nts (recuadro amarillo) corresponden a los pequeños ARNs ligados a los adaptadores retrotranscriptos y amplificados con oligonucleótidos específicos. Las porciones de gel recolectados fueron fragmentados para eluirlos y el ADN precipitado con etanol. (RPI= índices usados para la construcción de las bibliotecas, CL=*Custom Ladder* (Illumina), HRL= *High Resolution Ladder* (Illumina)).

Las bibliotecas de pequeños ARNs tienen un tamaño de alrededor de 150 pares de bases (bp), las cuales corresponden a 20 a 30 nts de la secuencia de ADNc de los pequeños ARNs más 120 nts de la secuencia de ADNc de los adaptadores. Esa etapa de selección de tamaño es muy importante dado a la presencia de dímeros de adaptadores formados durante la ligación y también de productos de peso molecular mayores generados a partir de la ligación de otros pequeños ARNs no deseados que contienen al igual que los pequeños ARNs grupos fosfatos en el 5' (ej. ARN de transferencia y ARNs pequeños nucleolares).

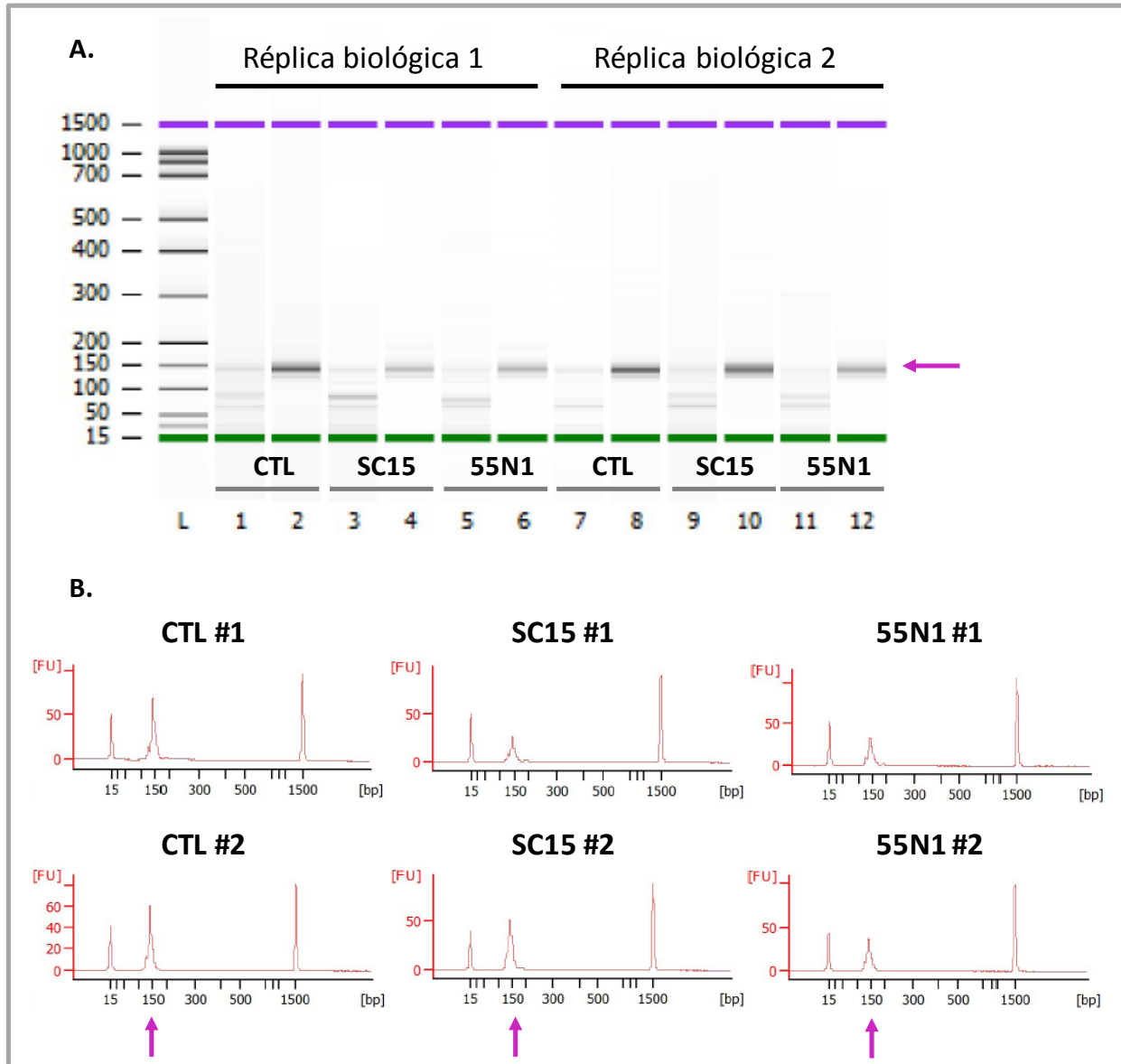
Se generaron un total de seis bibliotecas que representan tres condiciones de inoculación (CTL, SC15 y 55N1) y dos réplicas biológicas. Cabe mencionar que durante la etapa de amplificación por PCR se usaron oligonucleótidos con índices diferentes para cada condición (RPI<sub>x</sub>). Estos oligonucleótidos contienen una secuencia de 6 nucleótidos únicos que permiten su identificación (Figura 16). El uso de índices diferentes para cada condición y/o réplica biológica durante la construcción de las bibliotecas permitió *pool*ear las bibliotecas antes de secuenciarlas. La estrategia de *Multiplexing* permite secuenciar varias bibliotecas en una misma línea de secuenciación, reduciendo los costos sin afectar la representatividad de las muestras, maximizando así la adquisición de datos (Morozova and Marra, 2008; Buermans and den Dunnen, 2014; Wong et al., 2013).



**Figura 16: Multiplexing de las bibliotecas.** En la etapa de amplificación por PCR de la construcción de las bibliotecas usando la Tecnología de Illumina, cada muestra fue identificada por un “código de barra” o índices dados por una secuencia de 6 nucleótidos contenidos en los oligonucleótidos RPIx. Las bibliotecas son normalizadas y *pooleadas* (*Multiplexing*) antes de ser secuenciadas en la misma línea del secuenciador automático. Las secuencias índice (verde y amarillo) permite asociar cada lectura a su muestra de origen (*De-multiplexing*). Se ilustra un esquema para 2 muestras. En nuestro caso se usaron 6 índices diferentes, uno para cada muestra.

### 1.3 Validación, cuantificación y normalización de las bibliotecas

La concentración e integridad de las bibliotecas fueron analizadas mediante electroforesis capilar utilizando un Agilent Technologies 2100 Bioanalyzer (Figura 17). Los productos obtenidos antes y después de la purificación en el gel de poliacrilamida (ver Figura 14) fueron controlados en una corrida de electroforesis capilar usando el kit DNA-1000, siguiendo las recomendaciones del proveedor (Agilent). La observación de bandas únicas luego de la etapa de purificación del ADNc en el gel confirmó su efectividad (Figura 17A, líneas pares). El análisis de los electroferogramas obtenidos mediante el Bioanalyzer reveló la presencia de un único pico ubicado al alrededor de 150 bp, sin productos de degradación evidentes (Figura 17B).



**Figura 17: Electroforesis capilar de las muestras de ADN de las bibliotecas de pequeños ARNs.** A. Las 6 muestras de ADN de las bibliotecas fueron separadas mediante electroforesis capilar en un Agilent 2100 Bioanalyzer. Los números impares corresponden al producto de PCR (visualización de varias bandas), mientras los números pares provienen de las muestras después de la selección y purificación de las bandas del tamaño seleccionado a partir del gel de poliacrilamida (indicado por las flechas violetas a ~150 bp). B. Electroferogramas de la electroforesis capilar para el producto final de las bibliotecas (luego de la última etapa de purificación). Los tamaños (eje x en pares de bases, bp) y las concentraciones de las muestras (eje y en unidades de fluorescencia FU) fueron estimados por el programa mediante el uso de marcadores moleculares de 15 y 1500 bp incluidos en el kit.

La determinación de la concentración (en ng/μl y nmol/l) para cada muestra establecida por el programa asociado al Bioanalyzer se muestra en la Tabla 2. Las electroforesis capilares indicaron una buena calidad de las bibliotecas obtenidas, con altas concentraciones en ADN,

permitiendo su subsecuente secuenciación. Además de informar sobre la calidad y la pureza del ADN, la electroforesis capilar permitió conocer precisamente el tamaño de los productos de amplificación (de 140 a 142 bp) y su concentración (Tabla 2). Usando esta información se pudieron normalizar las bibliotecas ajustando cada una de ellas a la misma concentración, para evitar errores de sobrerepresentación durante la secuenciación y posterior análisis de los datos obtenidos. Con este fin, las muestras con concentraciones más altas fueron diluidas con agua miliQ estéril para obtener la concentración de la muestra la más baja (24,9 nM), (Tabla 2).

**Tabla 2: Tamaño y concentración de las muestras de ADNc.**

Los tamaños y concentración de los fragmentos de ADN de las bibliotecas fueron estimadas en un Agilent Bioanalyzer 2100 utilizando el kit de Agilent DNA100.

<b>Muestra</b>	<b>Tamaño (bp)</b>	<b>Concentración (ng/μl)</b>	<b>Molaridad (nmol/l)</b>
<b>CTL #1</b>	142	5,78	61,7
<b>SC15 #1</b>	141	2,33	24,9
<b>55N1 #1</b>	142	2,83	30,1
<b>CTL #2</b>	140	5,39	58,3
<b>SC15 #2</b>	141	6,26	67,1
<b>55N1 #2</b>	140	3,10	33,4

## 2. Secuenciación de las bibliotecas y procesamiento de los datos

Las bibliotecas se *poolearon* y se secuenciaron en un Genome Analyzer II utilizando el servicio HTS de la Universidad de California Riverside (UCR), USA. Previamente a su secuenciación, la calidad y concentración de las muestras fueron analizadas nuevamente por el proveedor del servicio. Para la secuenciación, se realizaron corridas que arrojan lecturas simples (*single reads*) de 51 pares de bases, en corridas de 36 ciclos. Luego de la secuenciación se realizó el *de-multiplex* según el índice que contienen asociando cada lectura a la muestra de origen. Los datos denominados “Small YS5” fueron cargados en la plataforma de secuenciación de Illumina de UCR (<http://illumina.ucr.edu/ht/>). Se obtuvieron más de 170 millones de lecturas redundantes y 33 millones no redundantes repartidas en las 6 bibliotecas, con un mínimo de 25 millones de lecturas por biblioteca (Tabla 3). Más del 97% de las lecturas exhibieron valores de Q superiores a 30, lo cual indica una alta calidad de la secuenciación.

**Tabla 3: Número de lecturas de las bibliotecas.** Cantidad de lecturas totales (redundantes) y únicas (no redundantes) para cada biblioteca.

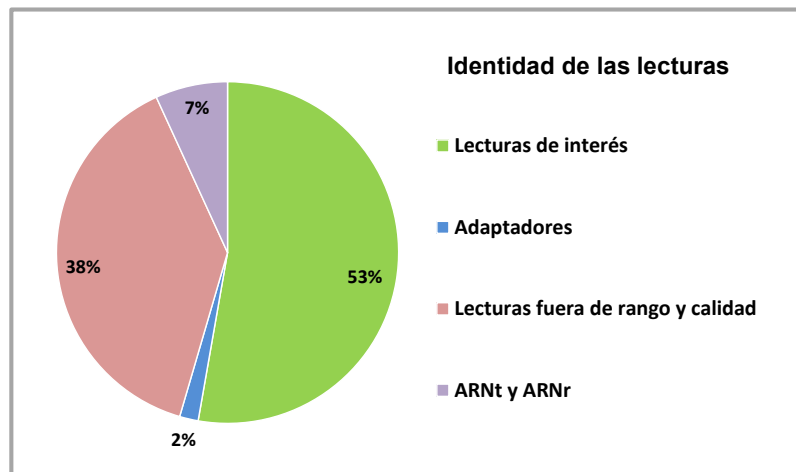
Biblioteca	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2	TOTAL
<b>Lecturas totales</b>	28 821 434	32 521 580	28 569 207	28 418 340	29 084 409	25 053 599	172 468 569
<b>Lecturas únicas</b>	6 961 601	7 258 370	6 552 488	6 697 383	6 130 675	6 340 575	33 941 092

Los archivos entregados por el servicio de secuenciación de Illumina poseen un formato *fastq*. El formato *fastq*, además de tener las secuencias en formato *fasta*, contiene información adicional como la calidad de lecturas de cada base. Este formato no suele ser compatible con los programas que llevan adelante el análisis posterior de los datos. Es por ello que se llevó a cabo la conversión de los archivos conteniendo los datos crudos en *fastq* a *fastq Sanger* utilizando la herramienta *NGS: QC Manipulation* disponible en la plataforma Galaxy (Blankenberg et al., 2010a). Se procedió al análisis de la calidad de las lecturas con *NGS: QC Manipulation > FastQC*, que indicó buena calidad de secuenciación. El siguiente paso fue filtrar las secuencias usando el programa UEA Workbench, The small RNA Workbench Versión 3.2 (Stocks et al., 2012; Beckers et al., 2017; Mohorianu et al., 2017). En primer lugar se aplicó un filtro para remover las secuencias correspondientes a los adaptadores libres. Estas secuencias correspondieron a un total de 3040381 lecturas, que representaron menos de 2% del total (Tabla 4). Luego se aplicó un filtro de tamaño, seleccionando únicamente las secuencias de 18 a 26 nucleótidos, y un filtro de calidad eliminando la baja complejidad y las secuencias inválidas. En este paso se eliminaron un total de 66655041 lecturas (38%). Finalmente se eliminaron las secuencias correspondientes al ARN de transferencia (ARNt) y ARN ribosomales (ARNr), las cuales representaron un 7 % del total (11787915 lecturas). La representación de las secuencias según su identidad fue semejante entre las muestras: de 1,3 % a 2,2 % para los adaptadores y de 5,0 % a 8,7 % para los ARNt/r. En el caso de las bibliotecas individuales, más de la mitad de las secuencias correspondieron a lecturas de interés (>50 %), con la excepción de la biblioteca que corresponde a la muestra de raíces inoculadas con SC15 de la réplica biológica #2, en la cual las lecturas de interés representaron el 42,6 %. Esta muestra contiene un mayor número de lecturas fuera de rango y calidad (49,5 %) respecto de las demás muestras (entre 34 y 38 %) (Tabla 4).

**Tabla 4: Identidad de las lecturas obtenidas luego de la secuenciación de las bibliotecas de pequeños ARNs.** Número y porcentaje de las lecturas de las 6 bibliotecas según su identidad.

Biblioteca	CTL #1	%	SC15 #1	%	55N1 #1	%	CTL #2	%	SC15 #2	%	55N1 #2	%	TOTAL	%
<b>Lecturas totales</b>	28 821 434	100	32 521 580	100	28 569 207	100	28 418 340	100	29 084 409	100	25 053 599	100	172 468 569	100
<b>Adaptadores</b>	566 083	2.0	699 532	2.2	622 301	2.2	421 764	1.5	374 273	1.3	356 428	1.42	3 040 381	1.76
<b>Lecturas fuera de rango y calidad</b>	10 965 559	38.0	11 712 744	36.0	10 615 813	37.2	9 833 437	34.6	14 389 108	49.5	9 138 380	36.48	66 655 041	38.65
<b>ARNt y ARNr</b>	2 496 444	8.7	1 852 091	5.7	2 395 498	8.4	1 872 677	6.6	1 931 086	6.6	1 240 119	4.95	11 787 915	6.83
<b>Lecturas de interés</b>	14 793 348	51.3	18 257 213	56.1	14 935 595	52.3	16 290 462	57.3	12 389 942	42.6	14 318 672	57.15	90 985 232	52.75

Una tabla con los datos de la cantidad de lecturas totales (no redundantes) e individuales (redundantes) luego de aplicar los filtros se presenta en la sección Anexos (Tabla S1). Luego de aplicar estos filtros se obtuvo un total de 90.985.232 lecturas remanentes entre las 6 bibliotecas, lo que representa cerca de 53 % del número total obtenido (Figura 18), con un promedio de 15 millones de secuencias redundantes restantes para cada biblioteca (Tabla 4), más de 90 millones totales (Tablas 4 y 5), y más de 28 millones de secuencias únicas por biblioteca (Tabla 5).



**Figura 18: Distribución de lecturas en las bibliotecas de pequeños ARNs.** Identidad de las lecturas redundantes totales provenientes de las 6 bibliotecas expresadas en porcentaje.

**Tabla 5: Número de lecturas filtradas.** Se presentan las secuencias redundantes (totales) y no redundantes (únicas) expresadas en cantidad de lecturas y porcentajes.

<b>Bibliotecas totales</b>	<b>totales</b>	<b>%</b>	<b>únicas</b>	<b>%</b>
Lecturas totales	172 468 569	100	39 941 092	100
Filtradas por adaptadores	169 428 188	98.2	36 196 510	90.6
Filtradas por tamaño (18-26)	102 808 961	59.6	28 869 530	72.3
Filtradas por calidad	102 773 147	59.6	28 842 847	72.2
Filtradas por ARNt y ARNr	90 985 232	52.8	28 622 813	71.7

En el proceso de filtrado de los datos de secuenciación, se eliminaron aquellas lecturas que no alineaban al genoma de *P. vulgaris*. Para esto, se utilizó la base de datos correspondiente al genoma de *P. vulgaris* v2.1 de la variedad andina Race Peru G19833 (*P.vulgaris\_218.fa*, Schmutz et al., 2014) disponible públicamente en *Phytozome* v12. (<https://phytozome.jgi.doe.gov/pz/portal.html>). El genoma mesoamericano de la variedad BAT93 fue secuenciado recientemente (Vlasova et al., 2016) y se encuentra depositado con libre de acceso en el National Center for Biotechnology Information (NCBI). Las lecturas fueron alineadas independientemente contra cada uno de los genomas de *P. vulgaris* disponibles utilizando la herramienta *Bowtie2* de Galaxy (Langmead et al., 2009; Langmead, 2013). El uso de esta herramienta en la plataforma Galaxy tiene la ventaja de combinar los resultados obtenidos con un análisis estadístico que permite comparar los alineamientos de ambos genomas (*Collect Alignment Summary Metrics*). Los resultados completos de este análisis se presentan en la Tabla 6. Para cada genoma se detallan los números de lecturas y sus valores en porcentajes correspondientes a las secuencias no alineadas, alineadas una vez o más veces a los genomas mesoamericano y andino. En el caso del genoma mesoamericano, entre 34 y 38 % de las lecturas alinearon una sola vez contra el genoma, mientras que entre 50 y 53 % de las mismas alinearon varias veces. En el caso del genoma andino entre 26 y 30 % de las secuencias se alinearon una única vez, mientras que entre el 48 y 51 % de las secuencias lo hicieron varias veces. Los resultados de la Tabla 6 fueron graficados sumando el total de las lecturas alineadas provenientes de las dos réplicas biológicas para cada condición (Figura 19).

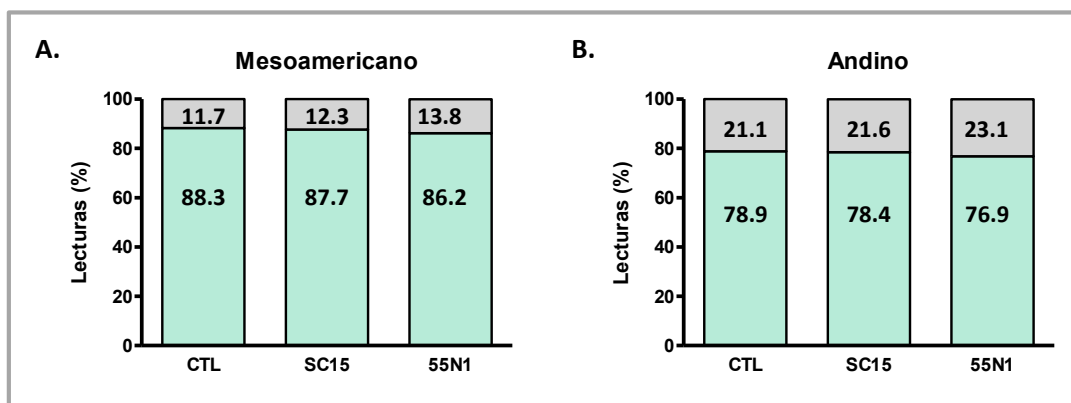


**Tabla 6: Alineamiento de las lecturas contra los genomas de referencia de *P. vulgaris*.** Las secuencias redundantes libres de adaptadores y filtradas por tamaño, calidad y sin las secuencias de ARNt y ARNr fueron alienadas contra los genomas de referencia de *P. vulgaris* de una variedad mesoamericana (BAT93) y una variedad andina (Race Peru G19833) utilizando la plataforma Galaxy.

Genoma MESOAMERICANO	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2
Lecturas totales	14793348	18257213	14935595	16290462	12389942	14318672
Lecturas no alineadas	1746119	1931310	1931473	1885687	1826558	2093191
Lecturas no alineadas (%)	11.8	10.6	12.9	11.6	14.7	14.6
Lecturas alineadas = 1	5631246	6893950	5289831	5738304	4374063	4852568
Lecturas alineadas = 1 (%)	38.1	37.8	35.4	35.2	35.3	33.9
Lecturas alineadas > 1	7415983	9431953	7714291	8666471	6189321	7372913
Lecturas alineadas > 1 (%)	50.1	51.7	51.7	53.2	50.0	51.5
Total lecturas alineadas	13047229	16325903	13004122	14404775	10563384	12225481
Total lecturas alineadas (%)	88.2	89.4	87.1	88.4	85.3	85.4

Genoma ANDINO	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2
Lecturas totales	14793348	18257213	14935595	16290462	12389942	14318672
Lecturas no alineadas	3233738	3660246	3373018	3335183	2948918	3386697
Lecturas no alineadas (%)	21.9	20.1	22.6	20.5	23.8	23.7
Lecturas alineadas = 1	4416487	5472080	4105257	4552468	3460999	3796083
Lecturas alineadas = 1 (%)	29.9	30.0	27.5	28.0	27.9	26.5
Lecturas alineadas > 1	7143123	9124887	7457320	8402811	5980025	7135892
Lecturas alineadas > 1 (%)	48.3	50.0	49.9	51.6	48.3	49.8
Total lecturas alineadas	11559610	14596967	11562577	12955279	9441024	10931975
Total lecturas alineadas (%)	78.1	80.0	77.4	79.5	76.2	76.4

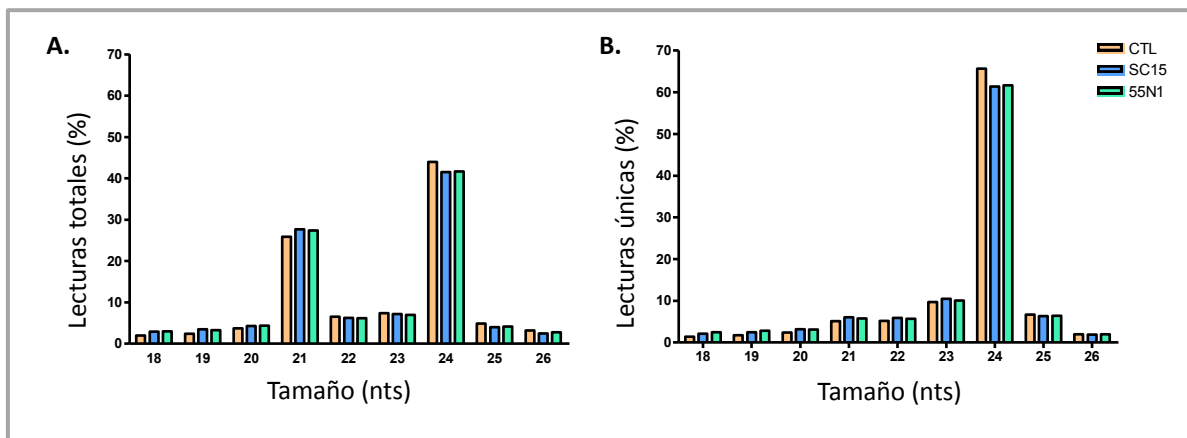


**Figura 19: Alineamiento de las lecturas de sRNA-seq a los genomas de referencia de *P. vulgaris*.** A. Porcentaje de lecturas alineadas contra el genoma de referencia de *P. vulgaris* de la variedad Mesoamericana BAT93. B. Porcentaje de lecturas alineadas al genoma de la variedad andina Race Peru G19833. En verde están representadas la cantidad de lecturas alineadas y en gris las lecturas que no mapearon contra los genomas de referencia (valores expresadas en porcentajes).

El mapeo de la secuencias contra los genomas de referencia mostró que el porcentaje de lecturas alineadas al genoma mesoamericano fue un 10% mayor que aquellas que alinearon al genoma andino en las tres condiciones de inoculación analizadas (Figura 19). Este resultado era de esperarse dado que la variedad NAG12 utilizada para la construcción de las bibliotecas es de origen mesoamericano. Se puede destacar una leve diferencia entre las diferentes condiciones. En comparación al control, 0.5 % menos de las lecturas alinearon contra el genoma en la condición de inoculación con la cepa más eficiente SC15, mientras que 2 % menos lo hizo en el caso de la inoculación con la cepa 55N1, independientemente del genoma utilizado como referencia. Una parte de las secuencias que no mapearon a ninguno de los genomas de *P. vulgaris* podrían corresponder a fragmentos amplificados a partir de ARNs bacterianos o contaminaciones de la muestra.

### 3. Clasificación de los pequeños ARNs por tamaño

Los pequeños ARNs pueden ser clasificados según su identidad en miARNs, siARNs o phasiARNs, la cual está asociada tanto a su biogénesis como también a su tamaño y función (ver Figura 10 de la sección Introducción). Por ejemplo los phasiARNs son generalmente de 21 nts, los miARNs pueden tener entre 20 y 22 nts y la población de ARNs de 24 nts se compone mayoritariamente de siARNs de heterocromatina (hc-siARNs). Para evaluar la representación de la cantidad de lecturas según el tamaño, las secuencias totales y únicas (redundantes y no redundantes respectivamente) libres de adaptadores y filtradas por tamaño y calidad fueron clasificadas por tamaño (de 18 a 26 nts) para cada condición. La distribución de las lecturas de las dos réplicas biológicas combinadas según su tamaño es presentada en porcentaje en la Figura 20. El mismo gráfico con los valores expresados en cantidad de lecturas se encuentra en la Figura S1 de la sección de Anexos.



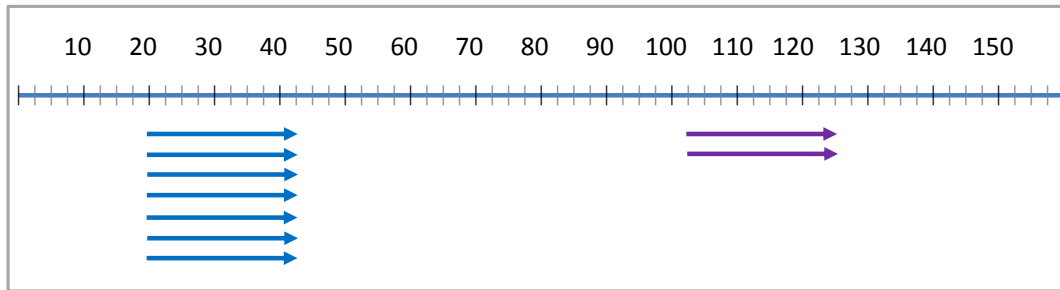
**Figura 20: Distribución de las lecturas según el tamaño.** A. Las secuencias totales (redundantes) libres de adaptadores y filtradas por tamaño y calidad fueron clasificadas según su tamaño (nucleótidos, nts) para cada una de las condiciones. B. Distribución de secuencias únicas (no redundantes) en relación al tamaño de los ARNs. Las dos réplicas biológicas (#1 y #2) fueron graficadas juntas. Los datos se presentan el porcentaje de lecturas versus su tamaño en nucleótidos (nts).

No se observaron diferencias significativas en la distribución de las lecturas por tamaño entre las 3 condiciones ensayadas (CTL, SC15 y 55N1), con la excepción de las lecturas de 24 nts que fueron levemente más abundantes en la condición control. Las secuencias de 21 y 24 nts fueron las más representadas en las lecturas totales, alcanzando más del 25 % y 40 % del total, respectivamente (Figura 20A). En las lecturas no redundantes (Figura 20B), las secuencias de 24 nts fueron las más representadas, con más de 60 % del total; sin embargo, las secuencias de 21 nts representaron menos de 10 %, un valor muy parecido a las de 22 nts y levemente inferior al de las de 23 nts. Estos resultados indican que los pequeños ARNs de 24 nts poseen una gran diversidad respecto de los de 21 a 23 nts, que constituyen una población con menor complejidad (i.e. cantidad de secuencias únicas). Los ARNs de 24 nts, que corresponden principalmente a miARNs largos o hc-siARNs, ambos con función de metilación del ADN (Axtell, 2013), los cuales no fueron caracterizados en detalle en este trabajo de Tesis.

## 4. Caracterización de los miARNs

### 4.1 Identificación de miARNs utilizando el algoritmo miRCat

La herramienta miRCat de UEA Workbench (Moxon et al., 2008) predice los miARNs a partir de datos de secuenciación masiva sin requerir *a priori* de las secuencias de los precursores putativos, las cuáles serán identificados por el programa a partir del genoma de referencia utilizado. Con este fin, las secuencias de pequeños ARNs fueron alineadas al genoma de referencia usando PatMaN (*Pattern Matching in Nucleotide Databases*, Prüfer et al., 2008), una de las funciones del paquete de miRCat. Una vez que las secuencias fueron alineadas, los *loci* que contienen las lecturas de pequeños ARNs fueron identificados como posibles *loci* de miARNs por el algoritmo miRCat según los siguientes criterios: 1) Los *loci* deben contener no más de 4 pequeños ARNs superpuestos (*no overlapping*); 2) cada pequeño ARN asociado a un *locus* no puede estar alejado más de 200 nts de su *locus* vecino más cercano; 3) al menos el 90 % de los pequeños ARNs asociados a un mismo *locus* tienen que tener la misma orientación. En una segunda etapa, el programa analiza cada uno de los *loci* putativos generadores de miARNs, de manera de identificar los mejores candidatos. La secuencia más abundante para un *locus* dado es definida por el programa como el miARN más probable. Posteriormente, para identificar los precursores de miARNs, las secuencias genómicas ‘flanqueantes’ al miARN candidato son analizadas por miRCat utilizando diferentes tamaños de ventanas y la estructura secundaria de cada miARN putativo es predicha utilizando RNAfold. De todos los posibles miARNs candidatos identificados por el programa, únicamente aquellos pequeños ARNs que pasan los siguientes filtros manuales fueron categorizados como miARNs: 1) El miARN y el miARN asterisco (miARN\*) no pueden tener más de 3 *mismatches* consecutivos; 2) el complejo miARN/miARN\* tiene que tener por lo menos 17 nts apareados sobre los 25 nts centrados alrededor del miARN putativo; 3) la región de tallo-rulo formada tiene que tener un mínimo de 75 nts de largo (Figura 21). El porcentaje de bases apareadas en la región de tallo-rulo putativa debe ser al menos de 50%.



**Figura 21: Representación de las lecturas identificadas como miARNs.** Las lecturas de pequeños ARNs de 20 a 22 nts fueron alineadas al genoma de referencia usando el programa miRCat de UEA Workbench. Las lecturas del miARN (en azul) y del miARN\* (en violeta) que forman la región de tallo-rulo tienen que mapear dentro de una ventana genómica de al menos 75 bp (valor que puede ser ajustado y que fue definido por el programa para precursores de miARNs de planta).

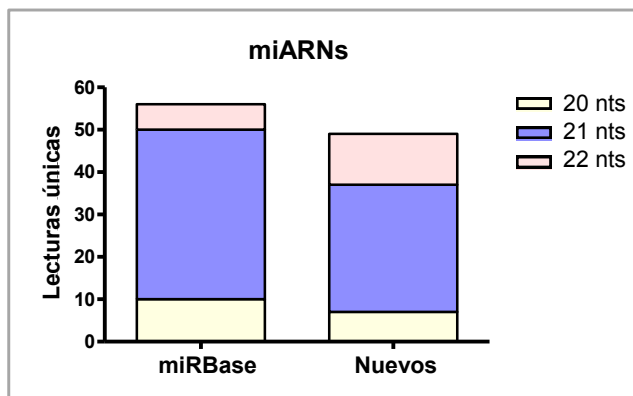
Luego, el programa define la región de tallo-rulo con el mínimo de energía libre (*Minimum Free Energy*, MFE) como el precursor más probable del miARN candidato (pre-miARN). Finalmente el pre-miARN fue testado por el algoritmo Randfold (*cutoff default = 0.1*) que calcula la probabilidad de que, para una secuencia de ARN dada, la MFE de su estructura secundaria sea más baja que la de secuencias de ARNs elegidas al azar (Bonnet et al., 2004). La información obtenida a partir del algoritmo miRCat incluye la lista de los miARNs maduros potenciales y su ubicación en el genoma, la orientación en la hebra, su tamaño, abundancia y la cantidad de veces que la secuencia mapea contra el genoma. Aquellos miARNs que mapean en más de un *locus* en el genoma se denominan miARNs '*multi-loci*'. También se presenta la presencia del miARN\* y su secuencia. La información sobre la región de tallo-rulo predicha incluye su secuencia y ubicación, el porcentaje GC, las energías libres mínimas (MFE y MFE ajustada según la longitud del tallo-rulo) y la probabilidad (*p-value*) según Randfold. Finalmente se analizó la existencia de los miARNs putativos depositados en la base a los datos miRBase (miRBase ID, Griffiths-Jones, 2010). Las secuencias de miARNs que presentaban una identificación en miRBase, designados como miARNs miRBase, se separaron manualmente de aquellas que no poseen identificación usando el genoma mesoamericano de referencia y se los clasificó como miARNs nuevos. Los miARNs nuevos para los cuales no fue asignado un miARN\* fueron clasificados en la categoría "otros". Las lecturas clasificadas de esta manera se presentan en la Tabla 7.

**Tabla 7: miARNs identificados utilizando miRCat.** Cantidad de miARNs identificados por miRCat que mapearon contra el genoma mesoamericano *P. vulgaris*. En cada biblioteca, los miARNs fueron clasificados en miARNs conocidos (presentes en miRBase), o nuevos (presentando por lo menos una lectura del miARN\*). La categoría “otros” corresponde a miARNs putativos para los cuales el programa no detectó la presencia del miARN\* en los datos correspondientes a esa biblioteca.

MESOAMERICANO	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2
miARNs_miRBase	66	76	74	69	71	69
miARNs_nuevos	14	21	19	20	20	19
Otros	3425	3396	3415	3297	3685	3184

#### 4.2 Distribución por tamaño de los miARNs

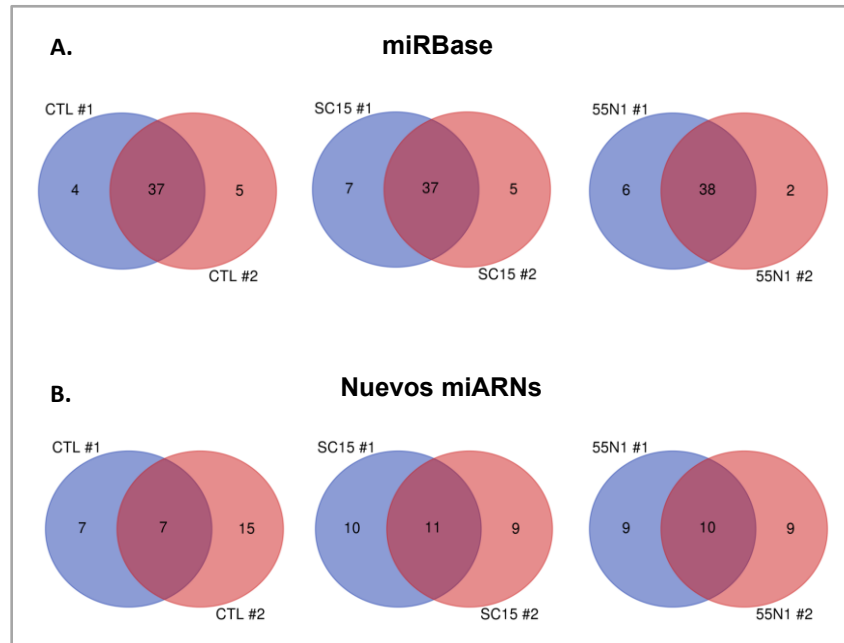
A partir de los miARNs identificados se determinaron las secuencias únicas totales en las 6 bibliotecas. Se identificaron un total de 56 miARNs conocidos (depositados en miRBase) y 48 nuevos miARNs que presentan un miARN\*. Los miARNs identificados fueron clasificados por tamaño (Figura 22). Los miARNs de 21 nts fueron los más representados en ambas categorías (conocidos y nuevos, 80 y 60 % respectivamente) seguidos por los de 20 nts (>15 %) en el caso de los miARNs conocidos, mientras que los más representados para los miARNs nuevos luego de los de 21 nts fueron los miARNs de 22 nts (>25 %).



**Figura 22: Distribución de los miARNs según el tamaño.** Los miARNs identificados por miRCat, conocidos (presentes en miRBase) o nuevos fueron clasificados según su tamaño en nucleótidos (nts). Los valores están expresados en cantidad de miARNs únicos.

#### 4.3 Cuantificación de miARNs

Las secuencias únicas de los miARNs que mapean contra el genoma mesoamericano fueron comparadas entre las dos réplicas biológicas para cada condición. Los resultados obtenidos representaron mediante los diagramas de Venn presentados en la Figura 23 (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

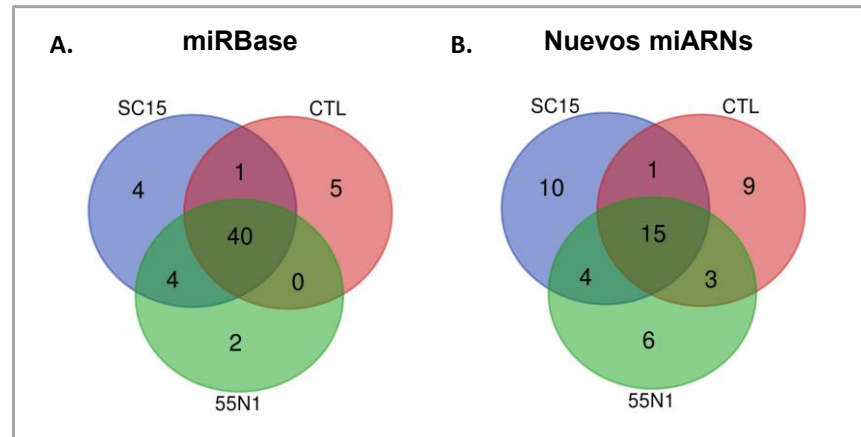


**Figura 23: Comparación de los miARNs identificados entre ambas réplicas biológicas.** Diagramas de Venn que representan la cantidad de secuencias de miARNs identificados en cada una de las dos réplicas biológicas. A. miARNs clasificados en miRBase. B. miARNs nuevos que presentan por lo menos una secuencia del miARN\*.

La mayoría de los miARNs reportados en miRBase estuvieron presentes en ambas réplicas biológicas con un porcentaje de superposición que osciló entre el 75 y el 82 % según la condición (Figura 23A). En el caso de los nuevos miARNs el grado de superposición entre ambas réplicas biológicas varió entre el 24 y el 37% (Figura 23B). Las listas de las secuencias presentes en ambas réplicas biológicas para cada condición se presentan en las Tablas S2 y S3 de la sección Anexos para los miARNs reportados en miRBase y los nuevos miARNs, respectivamente.

Posteriormente, las secuencias únicas de miARNs identificadas en cada una de las réplicas biológicas fueron comparadas entre las distintas condiciones, distinguiendo en el análisis los miARNs presentes en miRBase de los nuevos miARNs (Figura 24). Más del 70 % de las secuencias de miARNs presentes en miRBase fueron compartidas entre las 3 condiciones (Figura 24A), mientras que para los nuevos miARNs ese porcentaje fue mucho menor (30 %, Figura 24B). Solamente 4 secuencias de miARNs fueron compartidas entre las dos condiciones correspondientes a la inoculación con las cepas de *R. etli* (SC15 y 55N1); lo que corresponde sólo al 7 % para los miARNs conocidos, y al 8 % para los nuevos miARNs. Un bajo número de secuencias únicas fueron identificadas específicamente en una de las condiciones para los miARNs reportados en miRBase. La proporción de secuencias únicas presentes en una sola condición fue más alta para los nuevos miARNs que para los miARNs conservados. Sin embargo, las secuencias identificadas en una sola condición o compartidas de a pares, no presentan consistencia entre las dos réplicas biológicas, tanto para los miARNs conocidos como para los nuevos identificados. Únicamente una isoforma del miR399, un miARN de 21

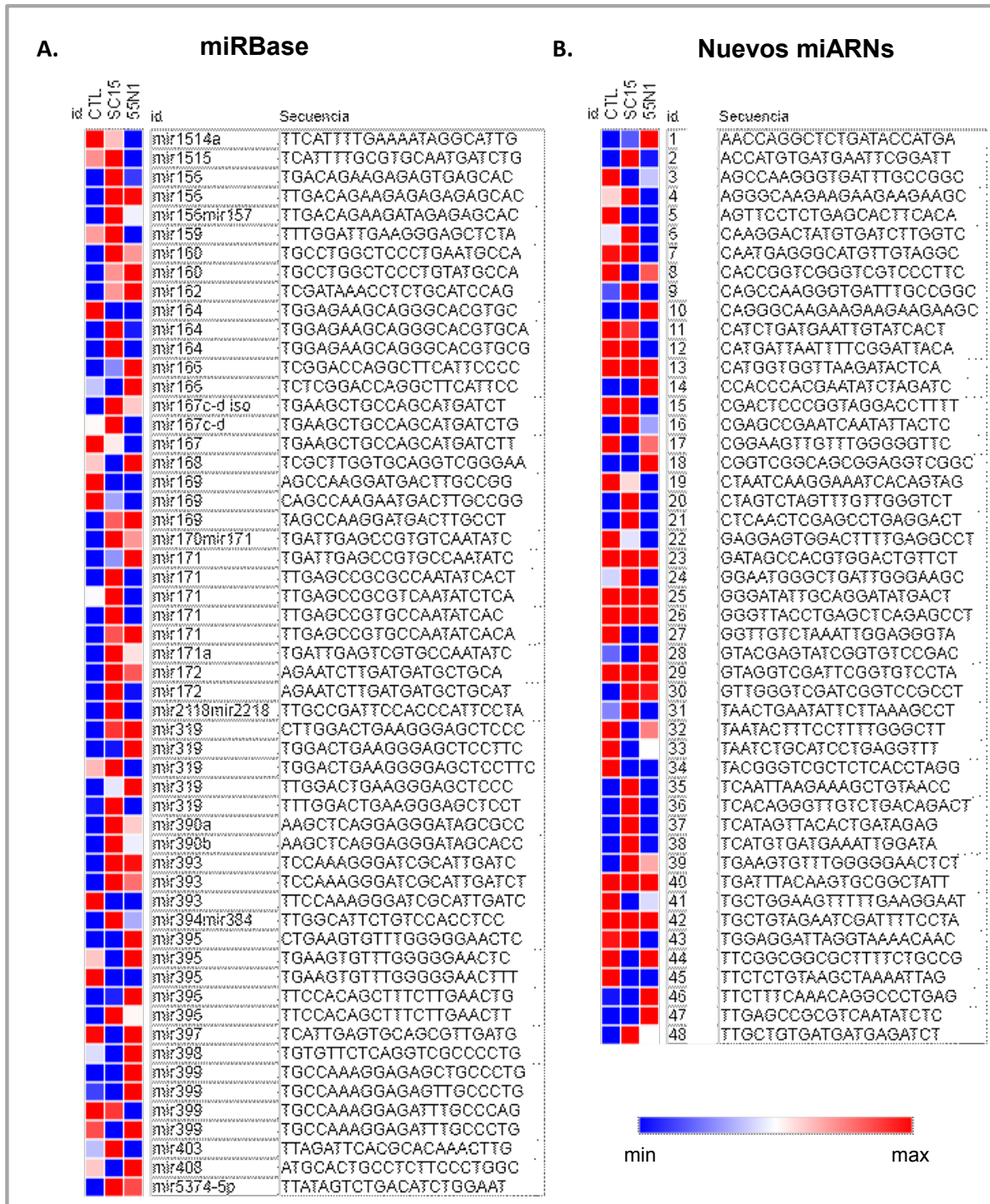
nts, se encontró presente en las dos réplicas biológicas en la condición de inoculación con 55N1, aunque sus niveles estacionarios fueron muy bajos (Tabla S4 de la sección Anexos).



**Figura 24: miARNs identificados en el genoma mesoamericano.** Diagramas de Venn representando la cantidad de secuencias únicas de miARNs identificadas por miRCat (Workbench) usando el genoma mesoamericano como referencia. A. Análisis de los miARNs clasificados en miRBase. B. Análisis de los nuevos miARNs, los cuales presentan por lo menos una secuencia del miARN\*.

#### 4.4 Identificación de miARNs con acumulación diferencial

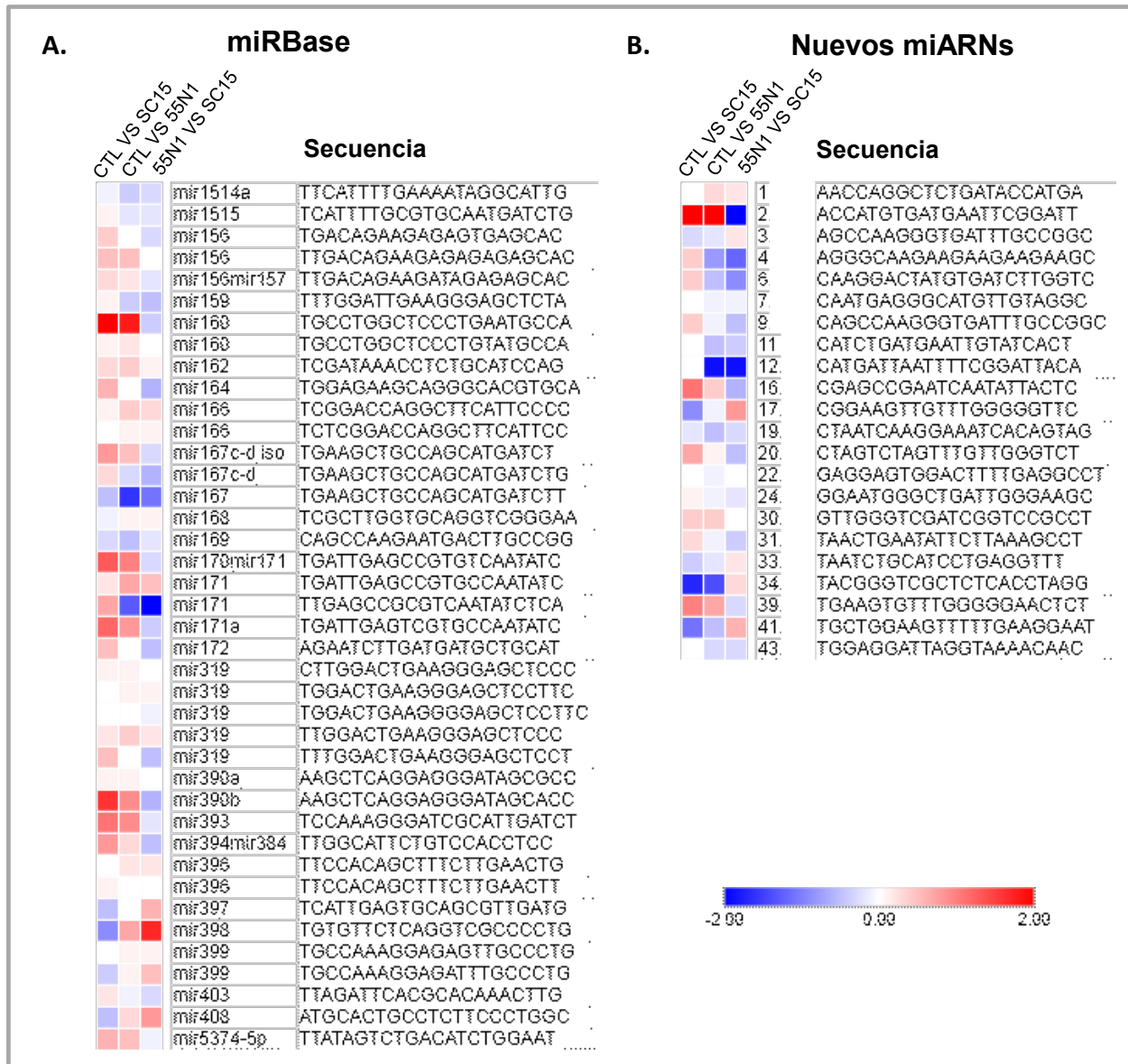
Para comparar la acumulación de los miARNs identificados entre cada condición, se calcularon las abundancias normalizadas de los miARNs para cada de las secuencias expresadas en lecturas por millón de lecturas (RPM):  $[(\text{abundancia}/\text{número total de lecturas filtradas}) \times 10^6]$ . Se comparó la abundancia normalizada de los miARNs conocidos en cada muestra tomando los valores de cada réplica biológica. A los miARNs que no fueron detectados en una condición se les asignó un valor nulo de abundancia. Las listas completas de los miARNs identificados usando la variedad mesoamericana del genoma de *P. vulgaris* (56 miARNs de miRBase y 48 miARNs nuevos) asociados a sus valores de abundancias se presentan en las Tablas S4 y S5 de la sección Anexos. Los miARNs que presentan mayores niveles estacionarios en nuestros datos de secuenciación corresponden a miARNs conservados en distintas especies de plantas (Tabla S4). Es el caso de los miR166, miR159, miR396, miR319 y miR403, que en su mayoría están involucrados en procesos de desarrollo (Jones-Rhoades et al., 2006). La mayoría de los nuevos miARNs identificados presentan niveles bajos (comparando con los valores promedios obtenidos en nuestra data de secuenciación), con la excepción de 8 miARNs (16 %) que presentan valores superiores a 50 RPM (Tabla S5). Para comparar la acumulación de los miARNs entre las distintas muestras se promediaron los valores normalizados de las dos réplicas biológicas para cada condición (Tablas S6 y S7 de la sección Anexos). Los valores se muestran en un mapa de calor (*heat map*) donde se muestra el patrón de acumulación de los miARNs identificados en miRBase (Figura 15A) y los nuevos miARNs (Figura 25B).



**Figura 25: Patrón de acumulación de los miARNs identificados.** Los valores de abundancias normalizadas (en RPM) obtenido del promedio de las dos réplicas biológicas se representan a través de un mapa de calor (*heat map*) obtenido mediante el programa Morpheus. A. Secuencias de los miARNs conocidos (presentes en miRBase). B. Secuencias de los nuevos miARNs que presentan por lo menos una secuencia del miARN\*. Se muestra la escala de referencia determinada por el programa a partir de los valores máximo y mínimo de cada fila.



La identificación de los miARNs con perfil diferencial se realizó determinando el *fold change* en comparaciones de a pares (CTL vs SC15, CTL vs 55N1 y 55N1 vs SC15) y calculando el logaritmo en base 2 del promedio de los valores de abundancias normalizadas de cada réplica. Estos valores se presentan en las Tablas S6 y S7 de la sección Anexos. Los miARNs sin lecturas (con valor infinito en la tabla) fueron retirados del análisis, conservando la numeración de los nuevos miARNs estudiados. Los patrones de acumulación diferencial (valores de *fold change*) de los miARNs fueron graficados en el *heat map* de la Figura 26.

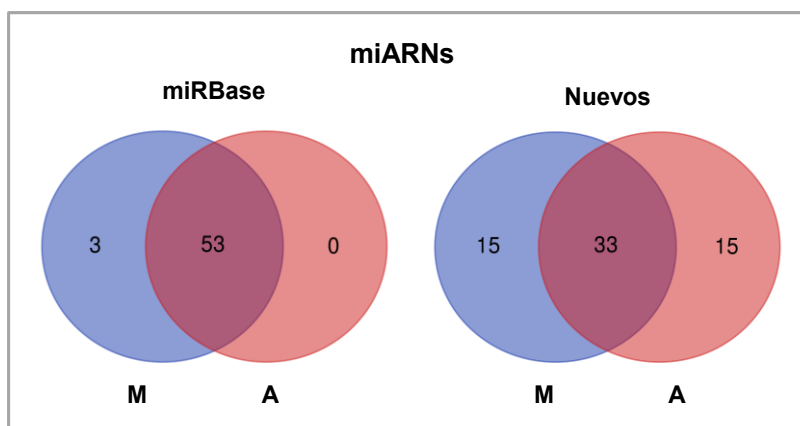


**Figura 26: Patrón diferencial de acumulación de los miARNs identificados.** Los valores de *fold change* (obtenidos a partir del cálculo del logaritmo en base 2 de los promedios de los valores de abundancias normalizadas) fueron analizadas en un mapa de calor (*heat map*) usando el programa Morpheus. A. Secuencias de los miARNs conocidos (presentes en miRBase). B. Secuencias de los nuevos miARNs (con miARN\*). La escala fue ajustada manualmente.

A partir de los resultados de los *heat maps* de las Figuras 25 y 26 se identificaron miARNs conservados y nuevos con perfiles de acumulación diferencial entre las 3 condiciones. Dentro de este grupo se priorizaron aquellos que tenían perfiles diferenciales en respuesta al rizobio en comparación al control (Tablas S6 y S7 de la sección Anexos). Se descartaron los miARNs que presentaron abundancias muy bajas en las 3 condiciones (cerca del límite de detección) y los que no tenían consistencia entre las dos réplicas biológicas (Tablas S4 y S5). Dentro de los miARNs conservados (Figuras 25A y 26A), se seleccionó una isoforma del miR390 (el miR390b) que presenta un perfil diferencial en respuesta al rizobio, más marcado con SC15, la bacteria que forma nódulos más eficientemente en el cultivar mesoamericano (Aguilar et al., 2004; Meschini et al., 2008). El análisis de este miARN se desarrolla en el Capítulo II de la presente Tesis Doctoral. En el caso de los nuevos miARNs (Figuras 25B y 26B), se destacó un nuevo miARN que presenta un perfil diferencial similar a miR390b (el miARN identificado con el número 4 de la Tabla S5 de la sección Anexos), es decir, con mayores niveles estacionarios en raíces inoculadas con SC15 en comparación a las otras dos condiciones. Este miARN fue denominado miR5924 y su estudio se presenta en el Capítulo III de esta Tesis.

#### 4.5 Análisis de miARNs utilizando los genomas de referencia mesoamericano y andino

Teniendo en cuenta la disponibilidad de bases de datos genómicas obtenidas a partir de una variedad andina y otra mesoamericana de *P. vulgaris*, se procedió a realizar un análisis comparativo de las poblaciones de miARNs. Este análisis permitiría identificar miARNs específicos de la variedad mesoamericana. Para ello, se compararon los resultados obtenidos a partir del análisis realizado con el algoritmo miRCat utilizando el genoma de referencia mesoamericano y andino. Las secuencias únicas de miARNs identificadas entre las 6 bibliotecas fueron graficadas en los diagramas de Venn de la Figura 28, separando los miARNs presentes en miRBase de los nuevos. Se compararon los números de secuencias no redundantes de miARNs identificados según hayan alineado al genoma mesoamericano, al andino o a ambos.



**Figura 28: miARNs identificados según el genoma de referencia.** Los diagramas representan el número de secuencias individuales de los miARNs identificados (conocidos en miRBase o nuevos con miARN\*). Se muestran las secuencias comunes o que aparecen en sólo uno de los dos genomas de referencia utilizados (M: mesoamericano; A: andino).

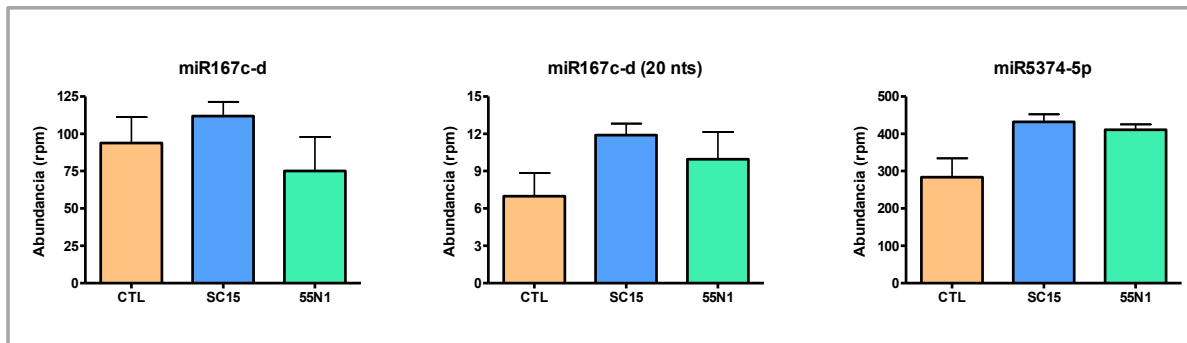
En cuanto a los miARNs ya reportados en miRBase, de los 56 miARNs identificados por miRCat en nuestra data de secuenciación de pequeños ARNs de la variedad mesoamericana, la mayoría (95%) fue encontrada también en el genoma andino (53 miARNs). Solamente 3 miARNs fueron identificados únicamente en el alineamiento al genoma mesoamericano. De los 48 nuevos miARNs identificados en el análisis anterior usando el genoma mesoamericano, 33 miARNs fueron identificados también en el genoma andino, mientras que 15 nuevos miARNs fueron identificados específicamente en el genoma mesoamericano. Además se identificaron 15 miARNs que alinearon sólo al genoma andino (Figura 28). Ya que las bibliotecas se realizaron a partir de muestras de raíces de plantas de variedad mesoamericana, los miARNs encontrados específicamente en el análisis utilizando el genoma andino como referencia sugieren que la secuenciación y la anotación del genoma mesoamericano son incompletas. Los miARNs que fueron caracterizados y se presentan en los Capítulos II y III (miR390b y miR5924) se alinean a los genomas de ambas variedades.

#### 4.6 Análisis y cuantificación de miARNs conservados

A continuación se describen algunos de los miARNs conservados que presentaron altos niveles de acumulación y consistencia entre las dos replicas biológicas (Tabla S4 de la sección Anexos). Estos miARNs presentan perfiles diferenciales interesantes entre las diferentes condiciones (Figura 26A y Tabla S6), los cuales sugieren que podrían tener una relación directa o indirecta con el establecimiento de la simbiosis. Los 56 miARNs conservados en otras especies vegetales (Figuras 22 y 24A y Tabla S4) están distribuidos en 28 familias de miARNs, de las cuales 15 presentan varias isoformas, mientras 13 tienen una sola secuencia. La mayoría de estos miARNs ha sido asociada a respuestas frente a situaciones de estreses abióticos, defensa a patógenos, o bien son conocidos por desempeñar funciones importantes en el desarrollo o en la nodulación. Uno de ellos, miR167, es miembro de una familia conservada con varios isoformas. En *Arabidopsis*, los mensajeros blancos putativos serían los factores de transcripción *ARF6* y *ARF8* (*Auxin Response Factors*) (Jones-Rhoades and Bartel, 2004). Más recientemente, dos isoformas de 21 nts del miR167 fueron secuenciadas en bibliotecas de pequeños ARNs de *M. truncatula* (Zhai et al., 2011) y *G. max* (Subramanian et al., 2008), los cuales se acumulan principalmente en hojas y nódulos, respectivamente. Por otro lado, la isoforma de 20 nts (TGAAGCTGCCAGCATGATCT), menos caracterizada, fue identificada en datos de secuenciación de pequeños ARNs de flores de *A. thaliana* (Kasschau et al., 2007). Se ha reportado igualmente una función del miR167 de 20 nts en respuesta a la herbivoría en hojas de *Nicotiana attenuata*, donde los blancos putativos de esta familia serían los genes relacionados a la señalización y biosíntesis de las fitohormonas (Pandey et al., 2008).

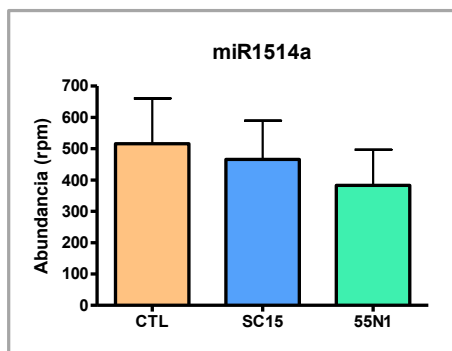
A su vez, el miR5374 (TTATAGTCTGACATCTGGAAT) es un miARN de 21 nts que fue identificado en bibliotecas de pequeños ARNs de *G. max* (Zhai et al., 2011) y que presenta niveles relativamente bajos en comparación al gma-miR167, con una acumulación mayor en raíces respecto de otros tejidos. También ha sido descrito en el trabajo de Kulcheski et al., 2011, donde se describen nuevos miARNs involucrados en respuestas a estreses bióticos y abióticos. Los blancos probables de este miARN son genes (*Glyma17g20860*, *Glyma05g09440*) que codifican proteínas con dominios LRR (*Leucine Rich Repeat*).

En nuestras bibliotecas, los miARNs miR167c-d de 21 y 20 nts, así como también el miR5374-5p se acumulan a mayores niveles en raíces inoculadas con la cepa SC15 en comparación con las demás condiciones (Figura 31, Figura 25A). Los precursores de estos miARNs se encontraron presentes tanto en el genoma mesoamericano como en el andino. El miR167c-d de 21 nts presenta una acumulación mayor en raíces inoculadas con SC15 que en aquellas inoculadas con 55N1. La isoforma de 20 nts del miR167c-d y el miR5374-5p presentaron acumulaciones mayores en la condición de inoculación con SC15 y 55N1, respecto al control.



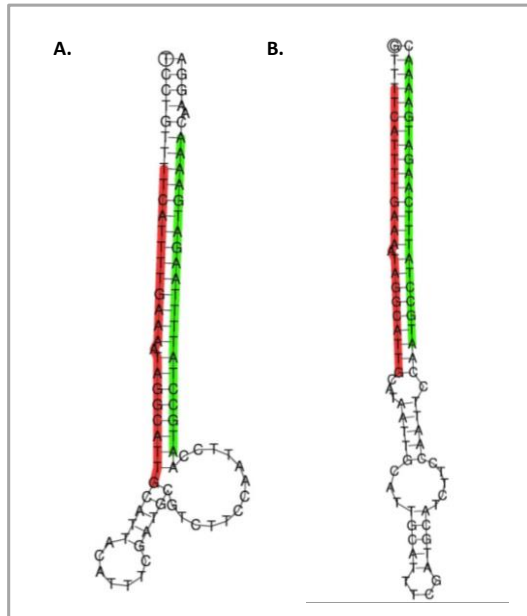
**Figura 31: Niveles de miARNs conservados en plantas en respuesta a rizobios.** miARNs de *P. vulgaris* identificados utilizando la herramienta miRCat. Los valores normalizados del promedio de las dos réplicas biológicas están expresados en lecturas por millón (RPM). Las secuencias de miR167c-d se discriminaron según su longitud en lecturas de 21 nts y 20 nts.

El miR1514 pertenece a una familia que contiene tres isoformas de 20, 21 y 22 nts. Las isoformas de 20 y 21 nts han sido descritas en bibliotecas de raíces de *G. max* inoculadas con *Bradyrhizobium japonicum* (Subramanian et al., 2008), en plántulas (Song et al., 2011) y en los cotiledones (Goettel et al., 2014). El miR1514a de 22 nts está presente en bibliotecas de plantas sometidas a estrés (sequía en hojas y plántulas tratadas por el ácido abscísico) y en plántulas inoculadas con rizobio (Arenas-Huertero et al., 2009). En nuestros datos de secuenciación se obtuvieron solamente lecturas de 22 nts (TTCATTTTGAAAATAGGCATTG) para este miARN. El miR1514a se acumula en las 3 condiciones ensayadas, con abundancias levemente menores en la condición de inoculación con SC15 y 55N1 respecto de la condición control (Figura 32, Figura 25A). En ambas réplicas biológicas se observa la misma tendencia, con una abundancia todavía menor para las plantas inoculadas con la bacteria menos eficiente 55N1 (Tabla S4).



**Figura 32: Niveles estacionarios de miR1514a en respuesta a rizobios.** Se muestran los valores normalizados del promedio de las dos réplicas biológicas expresados en lecturas por millón (RPM).

El miR1514a es un miARN denominado ‘*multi-loci*’, ya que el mismo podría ser procesado a partir de dos precursores con estructuras tallo-rulo independientes, una generada a partir de un gen presente en el cromosoma 3 y el otro en el cromosoma 7 (Figura 33, ver lista de secuencias en la sección Anexos). Ambos generarían estructuras estables, con valores de MFE similares.



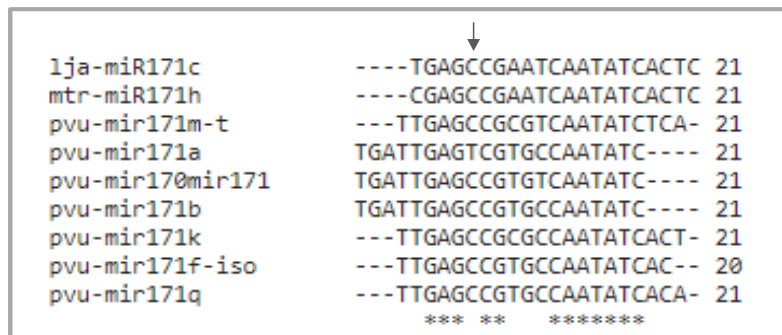
**Figura 33: Estructuras secundarias de los precursores del miR1514a.** A. Precursor del miR1514a generado a partir del gen ubicado en el cromosoma 3. MFE= -31,11. B. Precursor del miR1514a generado a partir del gen ubicado en el cromosoma 7. MFE= -35,10. La secuencia del miARN maduro de 22 nts está resaltado en rojo. Las estructuras secundarias fueron obtenidas mediante la herramienta de visualización de UEA Workbench, utilizando los miARNs\* (resaltados en verde) más abundantes presentes en la data de secuenciación.

#### 4.7 Conservación de un miARN previamente asociado a la simbiosis, el miR171

El miR171 pertenece a una familia multigénica conservada en una gran diversidad de plantas (Dezulian et al., 2005) con diversas funciones en el desarrollo (Zhou et al., 2008; Wong et al., 2011), en las respuestas a estreses abióticos como la tolerancia a metales pesados (Ding and Zhu, 2009) y a estreses bióticos (Lelandais-Briere et al., 2009; Laressergues et al., 2012; Zhai et al., 2011; Lelandais-Brière et al., 2016). En *A. thaliana* los blancos putativos del miR171a,d son los factores de transcripción Scarecrow (SCR, *At2g45160*, *At3g60630*, *At4g00150*, Jones-Rhoades and Bartel, 2004). En plantas leguminosas, una de las isoformas del miR171, el miR171c, fue asociado al proceso de infección del nódulo en *Lotus japonicus* (De Luis et al., 2012). Otra isoforma, el miR171h ha sido reportado como regulador de la simbiosis endomicorrízica con hongos arbusculares (AM) en *M. truncatula* (Laressergues et al., 2012). En *P. vulgaris* se han identificado 14 miembros pertenecientes a esta familia de miARNs (Formey et al., 2016). En nuestros datos de secuenciación se detectaron 7 isoformas del miR171 (Tabla S4). Las secuencias de estas isoformas fueron obtenidas mediante búsqueda en la base de datos miRBase para asignarles manualmente una identificación. En esta base de datos ninguna isoforma del miR171 ha sido reportada para la especie de *P. vulgaris*. Sin embargo, esto puede deberse a un retraso en la actualización de la base de datos de miRBase o a restricciones en la incorporación de miARNs putativos que no han sido validados. Por ejemplo, en *P. vulgaris* se identificó una isoforma de 21 nts del miR171: el pvu-miR171b. Esta

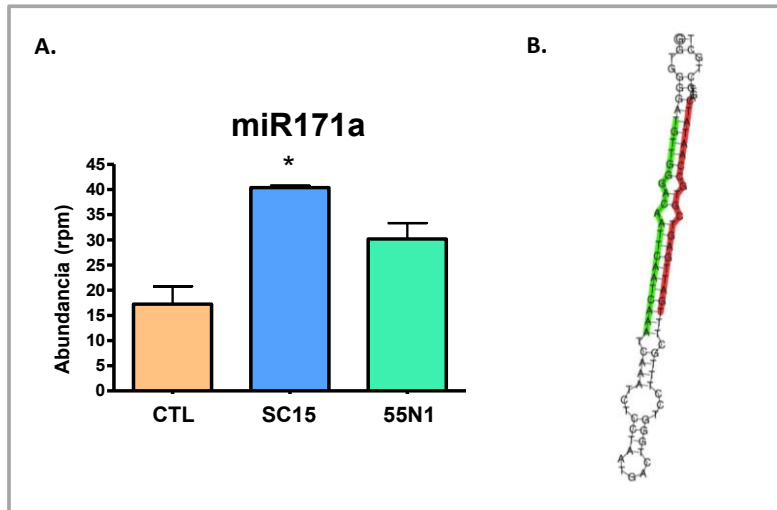
isoforma fue clonada y validada por *RNA Blot* a partir de bibliotecas de respuesta a estreses y actuarial también sobre el factor de transcripción SCR (Arenas-Huertero et al., 2009). Sin embargo su secuencia no se encuentra depositada en miRBase.

Las 7 isoformas de *P. vulgaris* presentes en nuestros datos de secuenciación (Figura 34) corresponden a isoformas identificadas en otras especies reportadas en miRBase. Se utilizó la nomenclatura de las especies más cercanas (*M. truncatula* y *G. max*), para la anotación de las isoformas de *P. vulgaris* según la homología de las secuencias. Sin embargo, la nomenclatura de las isoformas no muestra siempre correspondencia entre las diferentes especies en las cuales han sido descritas. Por ejemplo el miR171b de *P. vulgaris* (Arenas-Huertero et al., 2009) corresponde al miR171d de *M. truncatula*. Para este caso se respetó entonces la nomenclatura reportada previamente en *P. vulgaris*.



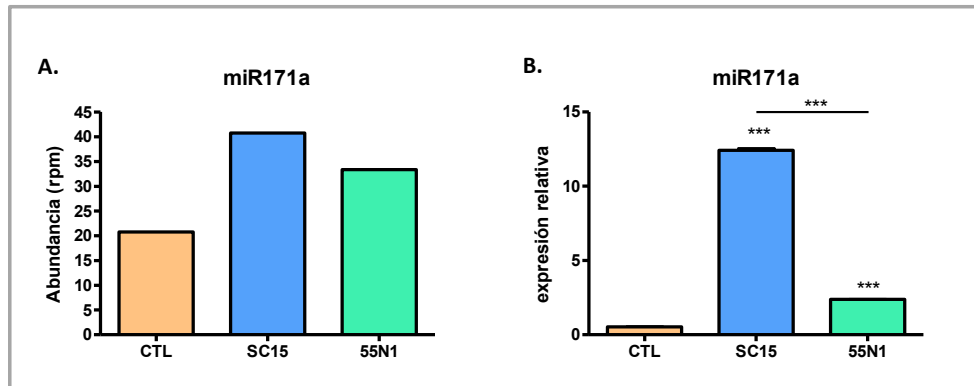
**Figura 34: Alineamiento múltiple de las isoformas de miR171 de *P. vulgaris*.** Comparación de las secuencias de las 7 isoformas del miR171 de *P. vulgaris* identificadas en nuestra datos de secuenciación y dos isoformas de *L. japonicus* y *M. truncatula* asociadas a la simbiosis (lja-miR171c y mtr-miR171h). pvu = *P. vulgaris*, lja = *L. japonicus*, mtr = *M. truncatula*. Los alineamientos múltiples de secuencias se realizaron con Clustal Omega. Los asteriscos representan nucleótidos conservados. Le flecha indica el nucleótido variable de la isoforma del miR171a en la posición 9. El tamaño de las isoformas (en nts) está indicado a la derecha.

El alineamiento de estas secuencias con las isoformas del miR171c y miR171h de las leguminosas *L. japonicus* y *M. truncatula* que han sido involucrados en las simbiosis, mostraron que estos dos miembros de la familia del miR171 no son idénticos a ninguna de las isoformas identificadas en nuestras bibliotecas de *P. vulgaris* (Figura 34). Con excepción de una isoforma de 20 nts (que tiene una sola lectura), todas las isoformas identificadas de miR171 son de 21 nts. Sin embargo, 6 de las 7 isoformas presentes en nuestras bibliotecas presentan valores de abundancias muy bajas y/o sin consistencia entre las dos réplicas biológicas (Tabla S4). Por el contrario, la isoforma miR171a presenta valores altos de FPM y un perfil de acumulación diferencial en respuesta al rizobio (Figura 35 y Figura 25A). La acumulación diferencial del miR171a fue estadísticamente significativa únicamente en las raíces inoculadas con SC15 respecto de las raíces control, mientras que aquellas inoculadas con 55N1 no fueron significativas (Figura 35A). Aunque estas diferencias fueron detectadas en ambas réplicas biológicas, las mismas fueron más notorias en la segunda réplica (Tabla S4).



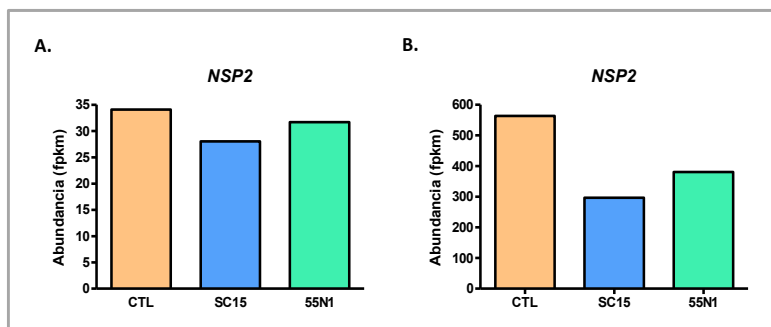
**Figura 35: Niveles de miR171a en respuesta a la inoculación con rizobios.** A. Valores de acumulación normalizados expresados en lecturas por millón (RPM). Los datos representan el promedio y desvío estándar de las dos réplicas biológicas B. Estructura secundaria del precursor del miR171a obtenida mediante la herramienta de visualización de UEA Workbench. El miR171a maduro aparece resaltado en rojo y su miARN\* en verde. MFE= -31,20. El asterisco indica diferencias significativas respecto al control en un t-test con  $p < 0,05$ .

Con el objetivo de validar el patrón de acumulación diferencial se cuantificaron los niveles de acumulación del miR171a maduro en las tres condiciones ensayadas (Figura 36A) utilizando la técnica de *stem-loop* RT-qPCR (Chen et al., 2005; Varkonyi-Gasic et al., 2007) siguiendo el protocolo descrito en la sección Materiales y Métodos. Los resultados obtenidos indicaron que la acumulación del miR171a fue mayor en aquellas raíces inoculadas con SC15 (aproximadamente 13 veces), que en aquellas inoculadas con 55N1, en donde si bien la inducción es significativa, es de sólo de 3 veces (Figura 36B). Este resultado permitió verificar las tendencias observadas en los datos de secuenciación utilizando un método independiente, indicando que el sRNA-seq es un método robusto para identificar miARNs maduros que se regulan diferencialmente durante la simbiosis.



**Figura 36: Comparación de la abundancia del miR171a en sRNA-seq y PCR cuantitativo.** A. Valores normalizados en RPM obtenidos a partir de la réplica biológica #1 mediante sRNA-seq. B. Medición de la abundancia del miR171a en la réplica biológica #1 mediante la técnica *Stem-Loop* PCR. Los asteriscos indican diferencias significativas respecto al control en un t-test (\*\*\*,  $p < 0,0001$ ).

A partir del análisis de posibles genes blancos del miR171a usando el algoritmo psRNATarget (Dai and Zhao, 2011), se identificaron el transcripto *Phvul.008G143400*, ortólogo putativo de *NSP2* de *M. truncatula*, y los transcriptos *Phvul.010G072700.1* y *Phvul.010G072900.1*, homólogos a *SCR* de Arabidopsis. Haciendo uso de datos transcriptómicos obtenidos en nuestro laboratorio a partir de raíces inoculadas y colectadas a 24 hpi con las dos cepas de *R. etli* estudiadas (Dalla Via et al., 2015; Clúa et al., no publicado) se analizó la expresión de *NSP2* y *SCR*. Los valores de expresión de *SCR* fueron similares entre cada condición en ambas bibliotecas (datos no mostrados). En la Figura 37 se presentan los valores de expresión para *NSP2* obtenidos a partir de una de las réplicas biológicas utilizadas para construir las bibliotecas de sARNs (Figura 37A) y los valores promediados de tres réplicas biológicas generados durante el trabajo de Tesis Doctoral del Dr. Clúa (Facultad de Ciencias Exactas, Universidad Nacional de La Plata) (Figura 37B). Aunque estos transcriptos no mostraron diferencias estadísticamente significativas ( $p\text{-value} > 0,2$ ), en ambas bibliotecas las expresiones de *NSP2* se correlacionan inversamente con la acumulación del miR171a maduro observada en nuestras bibliotecas.

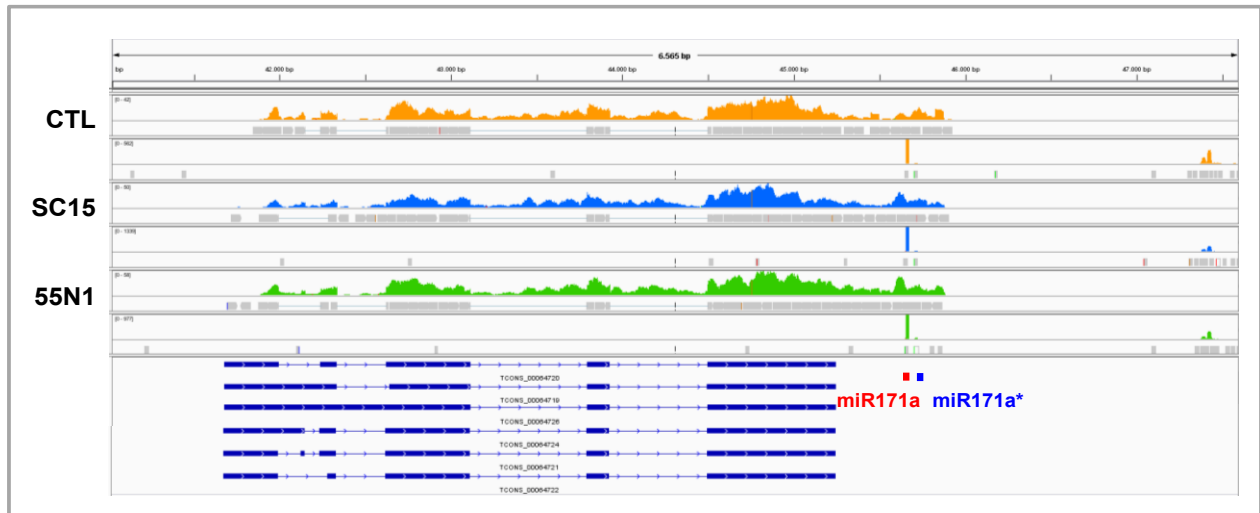


**Figura 37: Niveles de expresión del blanco putativo *NSP2* de miR171a.** Valores de expresión normalizadas en FPKM de *NSP2* (*Phvul.008G143400.1*) obtenidas a partir de los datos de RNA-seq. A. Datos obtenidos a partir de los datos de RNA-seq reportados por Dalla Vía et al., 2015. B. Datos no publicados

obtenidos a partir de tres réplicas biológicas (Joaquín Clúa, Tesis Doctoral).



Adicionalmente se analizó la ubicación del miR171a en el genoma mesoamericano. Usando la herramienta de miRCat de UEA Workbench, las lecturas del miR171a fueron asignadas por el programa a 3 regiones del genoma de *P. vulgaris*. Además de alinear a una región del cromosoma 2, el miR171a alineó a regiones de los cromosomas 1 y 3. Sin embargo, las secuencias del miR171a\* obtenidas en nuestros datos de secuenciación alinearon únicamente con la región genómica del cromosoma 2. La visualización de las lecturas del miR171a y su miARN\* alineadas con esta última usando *Integrative Genome Viewer* (IGV) se presentan en la Figura 38.

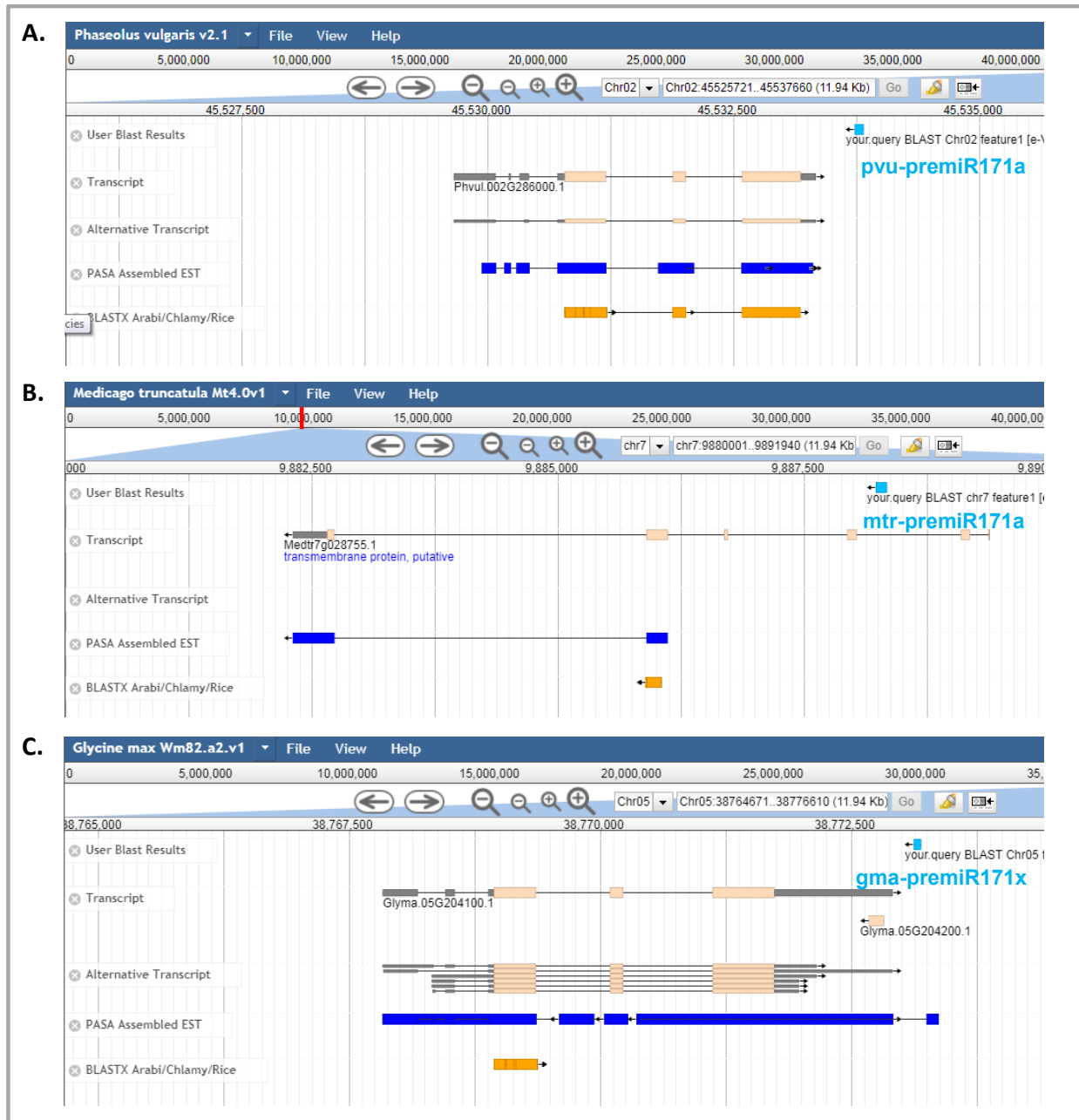


**Figura 38: Visualización del locus de *P. vulgaris* al cual alinea pvu-miR171a utilizando el programa IGV.** Lecturas del miR171a alineadas al genoma de *P. vulgaris* mesoamericano (cromosoma 2). Se presentan los datos de secuenciación de sRNA-seq (panel inferior) y RNA-seq (panel superior) obtenidos por Dalla Via et al., 2015 para cada condición (CTL, SC15, 55N1). Los datos representan las lecturas provenientes de ambas réplicas biológicas. Los diferentes modelos génicos propuestos están indicados abajo en azul.

La región tallo-rulo identificada por miRCat (91 nts) que contiene el miR171a y el miR171a\* (Figura 35B) se encuentra próxima a un transcripto que no posee anotación funcional (Figura 38). Los modelos génicos disponibles corresponden parcialmente a la cobertura de lecturas observada (parte 3' terminal con lecturas sin anotación). Sin embargo, la presencia de lecturas sin interrupción entre el transcripto anotado y el precursor ponen en duda la correcta anotación de esta región del genoma.

Según los datos disponibles en miRBase, la secuencia del miR171a maduro de *P. vulgaris* es idéntica a la del miR171a de *M. truncatula*. En esta base de datos esta isoforma no fue reportada para ninguna otra especie vegetal. Para determinar si los precursores del miR171a están conservados entre estas dos leguminosas, se realizó una búsqueda por BLAST en Phytozome usando la secuencia del precursor del premiR171a de *P. vulgaris*. Se integró al análisis *G. max*, la leguminosa modelo más cercana a *P. vulgaris*, que comparte la nodulación de tipo determinado, a diferencia de *M. truncatula*, que presenta nódulos indeterminados. Los resultados del BLAST indicaron que el premiR171a de *P. vulgaris* presenta 78,7% y 90,7 % de identidad con los premiARNs correspondientes de *M. truncatula* y *G. max*, respectivamente. La

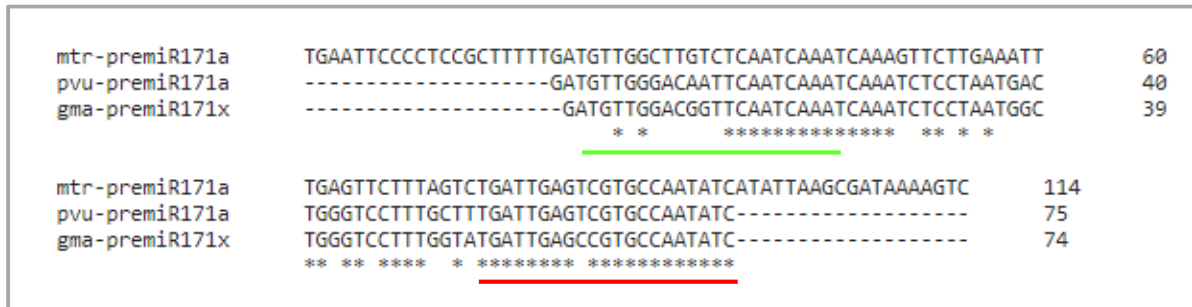
región identificada en el genoma de *M. truncatula* corresponde a la secuencia del precursor del mtr-miR171a depositado en miRBase. Los alineamientos de los premiR171a de *P. vulgaris*, *M. truncatula* y *G. max* a sus genomas respectivos están presentados en la Figura 39. Al igual que lo observado en el genoma mesoamericano (Figura 38), el premiR171a de *P. vulgaris* está ubicado en una región cercana a un transcripto (*Phvul.002G286000.1*) del cromosoma 2 del genoma andino (Figura 39A). La secuencia del pvu-premiR171a se alinea por fuera del 3' UTR de este transcripto. En el caso de *M. truncatula*, el precursor del miR171a se encuentra en el intrón de otro transcripto (*Medtr7G028755.1*), como lo muestra la Figura 39B. En *G. max*, el precursor correspondiente al premiR171a de *P. vulgaris* (denominado gma-premiR171x por no corresponder a ningún precursor disponible en miRBase para esta especie) corresponde a una región de alta sintenia, ubicándose en el 3' terminal de un transcripto homólogo (*Glyma.05G204100.1*) al de *P. vulgaris*. El análisis de las secuencias mostró que ambos codificarían proteínas con dominios conservados (PTHR31251:SF20 - SQUAMOSA PROMOTER-BINDING-LIKE PROTEIN 13A-RELATED).



**Figura 39: Ubicación del premiR171a en los genomas de las leguminosas.** A. Alineamiento del premiR171a al genoma andino de *P. vulgaris*. B. Alineamiento del premiR171a al genoma de *M. truncatula*. C. Alineamiento del premiR171x del genoma de *G. max*. Su secuencia fue obtenida a través de una búsqueda por homología de secuencia utilizando la herramienta BLAST.

A partir de las secuencias de los precursores del miR171a se realizó un alineamiento múltiple con las secuencias de las tres leguminosas (Figura 40). Los precursores se encuentran conservados entre las tres especies en la totalidad de la región tallo-rulo, y más específicamente en la zona correspondiente al miARN maduro. En nuestros datos de

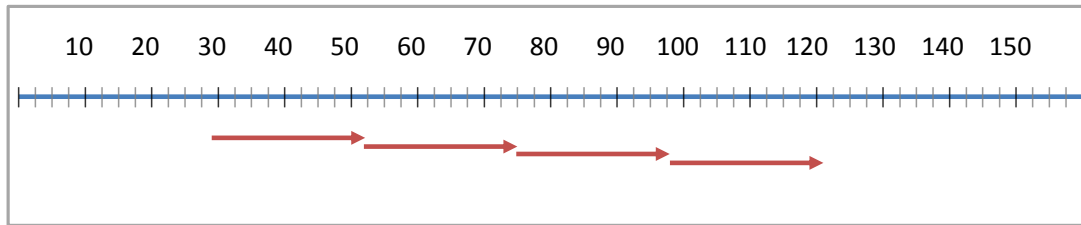
secuenciación se identificaron dos secuencias del miR171a\*, una de 21 nts (63 lecturas) y una isoforma de 22 nts (24 lecturas), ambas con la misma secuencia, sugiriendo que pueden originarse mediante diferentes procesamiento de un mismo precursor. La secuencia del miR171a\* se encuentra poco conservada entre las tres especies (6 *mismatches*).



**Figura 40: Precursores del miR171a de leguminosas.** Alineamientos de las secuencias de los precursores de los miR171a de *P. vulgaris* (pvu), *M. truncatula* (mtr) y *G. max* (gma-premiR171x). Los alineamientos de secuencias múltiples se realizaron con Clustal Omega. Los asteriscos representan nucleótidos conservados. En rojo está indicada la región correspondiente a los miR171 maduros y en verde el pvu-miR171a\*. Los tamaños de los precursores (en nts) están indicados a la derecha.

## 5. Caracterización de los phasiARNs

Los phasiARNs son moléculas de 21 nts específicas de plantas generadas a partir de ARNs doble hebra (Allen and Howell, 2010; Yoshikawa, 2013). Los alineamientos de los phasiARNs sobre su *locus* de origen muestra un patrón característico en fase a partir del cual pueden ser identificados *in silico* (Axtell et al., 2006). La herramienta de predicción de phasiARNs de la plataforma Workbench (*TA-SI Prediction* de UEA Workbench) permite identificar pequeños ARNs en fase de 21 nts que son característicos de *loci* que dan lugar a la producción de phasiARNs (Figura 41). Para ello, el programa usa un algoritmo descrito por Chen et al., 2007, el cual calcula la probabilidad de que la periodicidad de pequeños ARNs de 21 nts sea significativa según una distribución híper-geométrica. Este algoritmo informa los *loci* que dan lugar a la producción de phasiARNs, con su ubicación en el genoma (número de cromosoma, la posición de inicio y terminación del *locus*), la cantidad de secuencias de pequeños ARNs únicas que mapean en cada *locus*, la cantidad de pequeños ARNs únicos en fase de a 21 nts y la probabilidad para que el período repetitivo de 21 nts de las secuencias mapeadas en el *locus* ocurra por azar (*p-value*). También genera una lista de las secuencias de los pequeños ARNs identificados para cada *locus*, sus abundancias y coordenadas genómicas (cromosoma, posición y orientación).



**Figura 41: Representación de las lecturas identificadas como phasiARNs.** Usando el programa *TA-SI Prediction* de UEA Workbench, las lecturas de pequeños ARNs de 21 nts son alineadas al genoma de referencia. Las lecturas correspondiendo a phasiARNs (en rojo) tienen que aparecer en fase de 21 nts en relación al genoma.

Si bien la herramienta *TA-SI Prediction* del programa UEA Workbench nombra las secuencias ‘tasiARNs’, aquellas corresponden a phasiARNs ya que no informa sobre el modo de acción de las secuencias identificadas. Esta herramienta computa únicamente los pequeños ARNs de 21 nts con una abundancia mínima de 2 lecturas entran en el análisis (Lott, 2012). La representación de los phasiARNs en fase o fuera de fase sobre los *loci* de origen está explicitada en Stocks et al., 2012; los phasiARNs en fase siguen un patrón de 21 nts alineados al genoma (secuencias cada 21 nts), mientras que los se encuentran fuera de fase son secuencias que alinean a un *locus* particular, pero que no entran en la regularidad del patrón. Esto está contemplado en el programa, ya que los phasiARNs son identificados por grupos de fase más que por secuencias individuales (Lott, 2012).

Usando dicha herramienta y considerando sólo los phasiARNs en fase, pudieron identificarse aproximadamente 180 a 1000 phasiARNs en cada biblioteca de pequeños ARNs usando el genoma mesoamericano como referencia (Tabla 9). Con excepción del caso de la muestra control de la primera réplica biológica (CTL #1) que posee un 87%, el resto de las muestras presentaron aproximadamente un 50 % de los phasiARNs en fase de 21 nts.

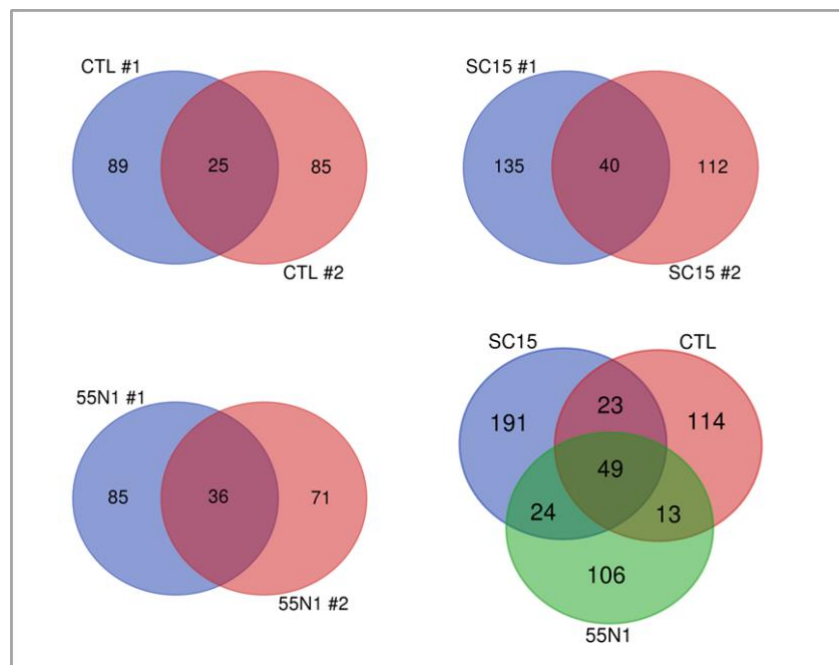
	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2
Total	1112	947	512	336	696	579
Sin fase	145	533	269	155	325	261
En fase	967	414	243	181	371	318
En fase (%)	87	44	47	54	53	55
<i>Loci</i> en fase	235	92	52	42	87	72

**Tabla 9: Identificación de phasiARNs.** Cantidad de secuencias de phasiARNs identificados con el *TA-SI Prediction* de Workbench, mapeando las secuencias contra el genoma mesoamericano. Tomando las secuencias filtradas de cada biblioteca se identificaron phasiARNs que fueron categorizados en phasiARNs en fase o fuera de fase. Se muestra el porcentaje de phasiARNs en fase respecto del total y la cantidad de *loci* totales de las secuencias en fase.

Las secuencias únicas de los phasiARNs alineadas en fase al genoma mesoamericano fueron comparadas entre cada réplica biológica utilizando los diagramas de Venn presentados en la Figura 42. Aunque la consistencia entre replicas biológicas es baja para cada condición (menos

de 20 % de secuencias de phasiARNs presentes en ambas réplicas), no se observan diferencias entre las tres condiciones (12,5 % para la condición CTL, 14 % para SC15 y 18,7 % para 55N1). La lista de las secuencias de phasiARNs que están presentes en las dos réplicas biológicas para cada condición se encuentra en la Tabla S8.

Las secuencias únicas de los phasiARNs en fase para cada condición (seleccionando las secuencias presentes en las dos replicas biológicas) fueron analizadas para determinar las secuencias compartidas entre dos o tres condiciones e identificar las secuencias específicas de una condición. La mayoría de las secuencias están presentes en una sola de las condiciones (79 %), con más secuencias específicas únicas en la muestra inoculada con SC15. Menos del 10 % de las secuencias de phasiARNs en fase fueron compartidas entre las tres condiciones (Figura 42, Tabla S8).



**Figura 42: phasiARNs identificados a partir de las bibliotecas de sARNs.** Los diagramas de Venn representan la cantidad de phasiARNs en fase identificados por *TA-SI Prediction* usando el genoma mesoamericano como referencia. El programa permitió identificar las secuencias de phasiARNs compartidas entre réplicas biológicas y las secuencias únicas de phasiARNs en fase identificadas para cada condición (abajo a la derecha).

Para cada grupo de secuencias de phasiARNs en fase identificado se analizaron las que tienen consistencia entre las dos réplicas biológicas para cada condición. Las secuencias presentes en una sola réplica fueron retiradas del análisis aunque se encuentren en las dos réplicas de otra condición. Aplicando este filtro se identificaron 18 phasiARNs en fase (denominados de 1 a 18), de los cuales 10 están presentes en las tres condiciones, dos son específicos de la inoculación con rizobios (SC15 y 55N1) y 6 son específicos de una condición (uno de la

condición control, dos de SC15 y tres de 55N1). La abundancia normalizada de los phasiARNs para cada réplica biológica en RPM [(abundancia/núm. total de secuencias filtradas)x10<sup>6</sup>], así como su localización genómica son presentados en la Tabla 10. La mayoría de los phasiARNs identificados presentan abundancias bajas, a excepción de los phasi#3, #6, y #7 (abundancias >20 RPM). Cinco de los 18 phasiARNs se encuentran repetidos en varios *loci* del genoma. Más precisamente estos phasiARNs pueden estar asociados a la misma región, como los phasi#4 y #7 (*scaffold00968*); o en regiones diferentes, como por ejemplo el phasi#11. Este último podría ser generado a partir de 10 *loci* diferentes, sin embargo sus abundancias son muy bajas.

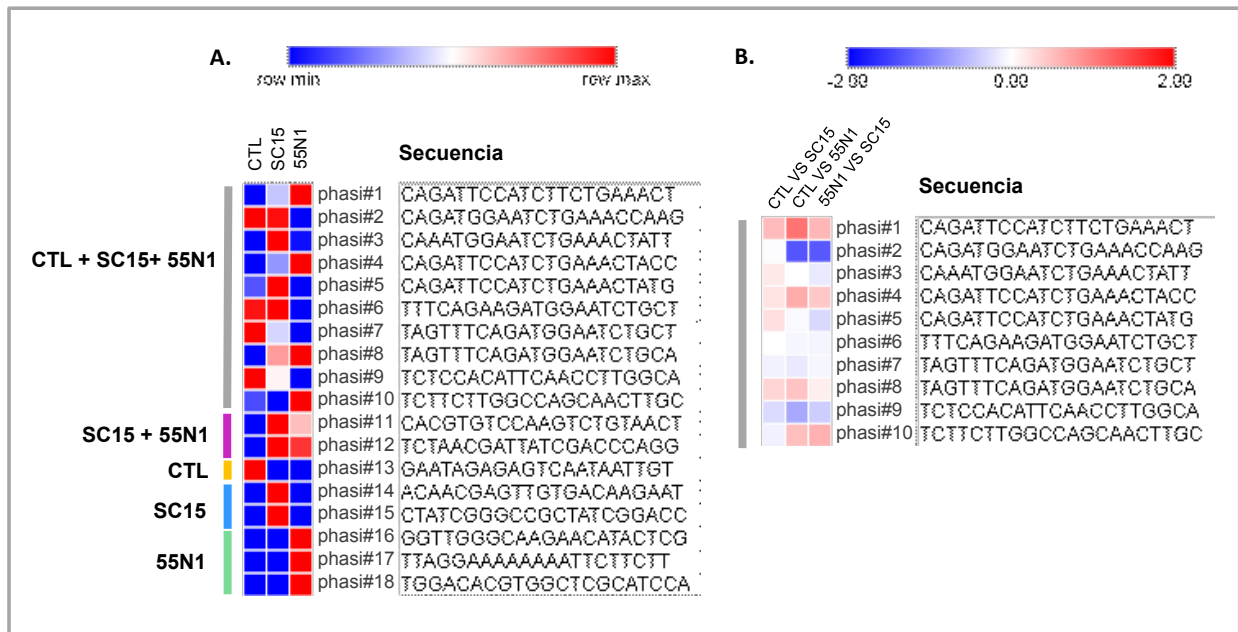
**Tabla 10: Identificación de phasiARNs en fase.** Valores normalizadas (en RPM) de los phasiARNs en fase específicos de una condición o compartidos entre distintas condiciones. Únicamente fueron considerados los phasiARNs que se encuentran representados en las dos replicas biológicas. Se indican las ubicaciones en el genoma Mesoamericano de *P. vulgaris*, y si estas secuencias fueron identificadas como phasiARNs en fase usando el genoma andino (por lo menos en una condición).

	phasiARN	Secuencia	Ubicación Mesoamericano	hebra	Andino	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2
CTL + SC15 + 55N1	phasi#1	CAGATTCATCTTCTGAAACT	scaffold00968 (75671-75691)	-	si	0.1	0.2	0.3	0.2	0.2	0.4
	phasi#2	CAGATGGAATCTGAAACCAAG	scaffold00968 (75692-75712)	-	si	0.1	0.3	0.1	0.6	0.4	0.1
	phasi#3	CAAAATGGAATCTGAAACTATT	scaffold00968 (75713-75733)	-	no	25.3	49.2	28.9	54.6	41.2	51.3
	phasi#4	CAGATTCATCTGAAACTACC	scaffold00968 (75776-75796)	-	si	0.3	0.5	0.5	0.4	0.3	0.6
			scaffold00968 (75797-75817)	-		0.3	0.5	0.5	0.4	0.3	0.6
	phasi#5	CAGATTCATCTGAAACTATG	scaffold00968 (75839-75859)	-	no	0.4	0.7	0.4	0.7	0.6	0.7
	phasi#6	TTCAGAAGATGGAATCTGCT	scaffold00968 (75673-75693)	+	si	29.9	48.6	38.2	47.9	29.4	35.8
	phasi#7	TAGTTTCAGATGGAATCTGCT	scaffold00968 (75736-75756)	+	no	24.8	39.5	32.8	42.8	23.7	27.4
			scaffold00968 (75799-75819)	+		24.8	39.5	32.8	42.8	23.7	27.4
	phasi#8	TAGTTTCAGATGGAATCTGCA	scaffold00968 (75862-75882)	+	si	0.7	1.8	1.8	1.7	1.4	1.6
phasi#9	TCTCCACATTCAACCTTGGCA	scaffold00019 (1028798-1028818)	+	no	1.0	0.7	0.5	0.7	0.7	0.6	
phasi#10	TCTTCTTGCCAGCAACTGTC	scaffold00019 (1028819-1028839)	+	si	1.7	1.5	3.1	6.0	5.6	7.7	
SC15 + 55N1	phasi#11	CACGTGTCCAAGTCTGTAAC	contig106903 (471-491)	+	si	0.0	0.2	0.3	0.0	0.5	0.1
			contig107413 (438-458)	+		0.0	0.2	0.3	0.0	0.5	0.1
			contig121238 (97-117)	+		0.0	0.2	0.3	0.0	0.5	0.1
			contig95704 (117-137)	+		0.0	0.2	0.3	0.0	0.5	0.1
			scaffold00109 (872525-872545)	+		0.0	0.2	0.3	0.0	0.5	0.1
			scaffold00202 (113458-113478)	+		0.0	0.2	0.3	0.0	0.5	0.1
			scaffold00689 (1250-1270)	+		0.0	0.2	0.3	0.0	0.5	0.1
			scaffold01142 (3311-3331)	+		0.0	0.2	0.3	0.0	0.5	0.1
			scaffold01983 (11700-11720)	-		0.0	0.2	0.3	0.0	0.5	0.1
			scaffold02699 (24-44)	+		0.0	0.2	0.3	0.0	0.5	0.1
phasi#12	TCTAACGATTATCGACCCAGG	scaffold00012 (647124-647144)	-	si	0.0	0.2	0.1	0.0	0.2	0.3	
CTL #1 #2	phasi#13	GAATAGAGAGTCAATAATTGT	scaffold00587 (56138-56158)	+	si	2.4	0.0	0.0	2.6	0.0	0.0
			scaffold00587 (108594-108614)	+		2.4	0.0	0.0	2.6	0.0	0.0
			scaffold00587 (202204-202224)	-		2.4	0.0	0.0	2.6	0.0	0.0
SC15 #1 #2	phasi#14	ACAACGAGTTGTGACAAGAAT	scaffold00029 (974152-974172)	-	si	0.0	0.2	0.0	0.0	0.4	0.0
			scaffold00353 (360894-360914)	-		0.0	0.3	0.0	0.0	0.3	0.0
	phasi#15	CTATCGGGCCGCTATCGGACC	scaffold00010 (988861-988881)	+	si	0.0	0.3	0.0	0.0	0.3	0.0
55N1 #1 #2	phasi#16	GGTTGGGCAAGAACATACTCG	scaffold00029 (973921-973941)	+	si	0.0	0.0	0.1	0.0	0.0	0.1
	phasi#17	TTAGGAAAAAAATCTTCTT	scaffold00029 (974047-974067)	+	si	0.0	0.0	0.1	0.0	0.0	0.1
	phasi#18	TGGACACGTGGCTCGCATCCA	scaffold05704 (2219-2239)	+	si	0.0	0.0	0.3	0.0	0.0	0.1

Las ubicaciones de los phasi#1 hasta el phasi#8 indican que son secuencias en fase de 21 nts, ubicados de ambos lados de la hebra del ADN. Llamativamente, los phasi#4 y #7 comparten la ubicación en el genoma, lo que sugiere que podrían provenir del mismo transcripto. Además, cabe destacar que los phasiARNs orientados sobre la misma hebra de ADN presentan alta homología de secuencias sugiriendo que podrían haber sido generados a partir de una zona repetitiva del genoma. Estas secuencias de phasiARNs fueron detectadas en las 6 bibliotecas con valores de abundancia semejantes entre las diferentes condiciones.

Por otra parte, se analizó si los 18 phasiARNs eran detectados en el genoma andino de *P. vulgaris*. Solamente 4 secuencias no fueron identificadas como phasiARNs en fase por el programa, de los cuales tres provendrían del mismo *locus* generador de phasiARN. Las secuencias de 21 nts generadas a partir de un mismo *locus* tendrían que estar en fase independientemente del genoma considerado, salvo que los genomas sean divergentes en ese *locus*. Estas diferencias podrían deberse a la historia evolutiva de los genomas o a una secuenciación incompleta.

Para representar los niveles de acumulación de los phasiARNs se realizaron *heat maps* a partir del promedio de las abundancias normalizadas (Figura 43A). A partir de estos valores se identificaron los phasiARNs con acumulación diferencial (calculando el *fold change*) (Tabla S9). Debido a que los valores nulos dan resultados que ser representados en este *heat map* (infinito), únicamente los phasiARNs presentes en las 3 condiciones fueron representados en este gráfico (Figura 43B).



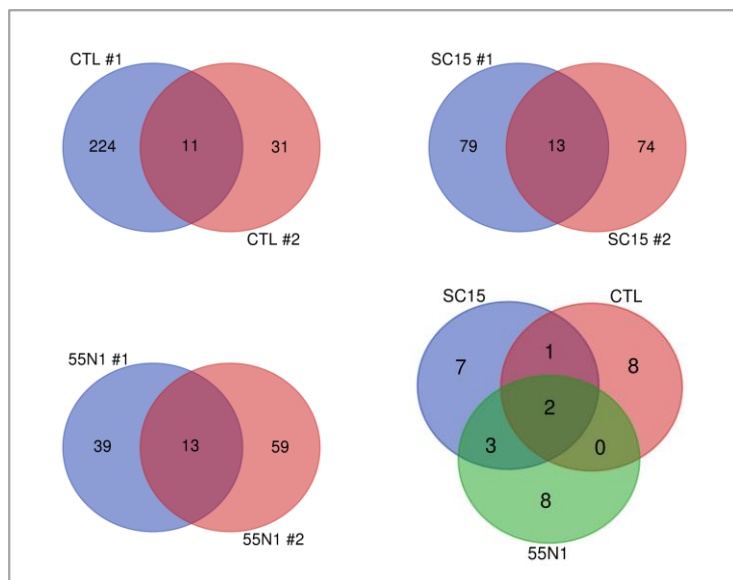
**Figura 43: Patrón diferencial de acumulación de los phasiARNs.** A. Los valores de abundancias normalizadas (en RPM) del promedio de las dos replicas biológicas de los phasiARNs identificados en fase con el genoma mesoamericano fueron analizadas en un mapa de calor (*heat map*) usando el programa Morpheus. La escala de colores fue determinada por el programa a partir de los valores máximo y mínimo de cada fila. En el análisis se consideraron únicamente los phasiARNs que presentaron lecturas en las dos replicas biológicas. B. *Heat map* representando los valores de *fold change* de los phasiARNs en fase (obtenidos a partir del cálculo del logaritmo en base 2 de los promedios de los valores de abundancias normalizadas).

Como se mencionó anteriormente, los phasiARNs más abundantes (phasi#3, #6, y #7) no presentaron un patrón diferencial muy marcado (Figura 43B y Tabla S9). Si se observó un perfil diferencial para el caso del phasi#2 el cual disminuye en respuesta a la cepa 55N1 pero no



frente a SC15, o bien para el phasi#10, el cual se incrementa en la condición de inoculación con la cepa 55N1 (Tabla 10).

Los análisis anteriores de los phasiARNs identificados revelaron una baja reproducibilidad entre las diferentes condiciones al nivel de secuencias individuales. Los valores de abundancias muy bajos que presentan la mayoría de los phasiARNs sugieren que son secuencias poco estables. Sin embargo, esto no significa que los mismos no desempeñen una función importante en el establecimiento de la simbiosis. Debido a que los phasiARNs son identificados por grupos de fase más que por secuencias individuales (Lott, 2012), y con el fin de identificar regiones diferencialmente reguladas entre las diferentes condiciones ensayadas, se analizaron los *loci* generadores de phasiARNs identificados en fase por el programa (Tabla 9). Se analizó la reproducibilidad entre réplicas biológicas para cada condición que fue representada en los diagramas de Venn (Figura 44). Como para los phasiARNs, la mayoría de los *loci* fue identificada en una sola réplica biológica, para cada una de las condiciones. Los *loci* con consistencia entre las dos réplicas biológicas de cada condición fueron analizados y se determinaron los que podrían ser específicos de una condición o compartidos entre varias. La lista de los *loci* identificados se encuentra en la Tabla S10.



**Figura 44: PhasiARNs regulados durante la simbiosis entre *P. vulgaris* y cepas de *R. etli*.** Diagramas de Venn representando los phasiARNs identificados en las bibliotecas de pequeños ARNs. Se muestran los phasiARNs compartidos entre réplicas biológicas y los que se encuentran compartidos entre las diferentes condiciones (abajo a la derecha).

En la lista de los *loci* compartidos entre las tres condiciones se encuentran dos de los listados en la Tabla 10. Para los tres *loci* compartidos entre las dos condiciones de inoculación (SC15+55N1), uno solo ha sido identificado en el análisis de los phasiARNs de secuencias individuales (*chromo:Scaffold00012*). Estos resultados sugieren que los phasiARNs generados a partir de los dos otros *loci*, aunque que no hayan sido identificados en el análisis de secuencias individuales, podrían ser específicos de la simbiosis. Este nuevo análisis a escala global permitió identificar *loci* generadores de phasiARNs con patrones interesantes. Sin embargo, se requieren estudios funcionales que permitirán dilucidar si dichos *loci* y los phasiARNs producidos a partir de los mismos desempeñan funciones relevantes en la simbiosis.

# Discusión

## Asignación de la función de los sARNs identificados

En este capítulo de trabajo de Tesis, hemos identificado pequeños ARNs, incluyendo siARNs y miARNs, regulados diferencialmente durante el proceso de reconocimiento e infección de cepas de alta y baja eficiencia en las etapas tempranas de la interacción entre *P. vulgaris* y *R. etli*. En particular, en *P. vulgaris*, las 24 hpi con *R. etli* corresponde al momento en que los pelos radicales comienzan a deformarse en respuesta a los factores Nod producidos por los rizobios. A partir de esta comunicación simbiótica, se inducen importantes cambios en la expresión génica asociados a los procesos de infección y organogénesis del nódulo (Peltzer-Meschini et al., 2008; Dalla Via et al., 2015), dónde los sARNs podrían tener una función regulatoria relevante.

Nuestros datos indicaron que los pequeños ARNs mayoritarios corresponden a ARNs de 21 y 24 nts, siendo los de 24 nts más diversos que los de 21 nts. Estos datos proveen información acerca de la naturaleza de estos pequeños ARNs, siendo los siARNs de 24 nts muy numerosos y poco abundantes, mientras que los sARNs de 21 nts (los cuales incluyen tanto los miARNs, siARNs y phasiARNs) son menos numerosos pero acumulados a mayores niveles en las células de la raíz de *P. vulgaris*. Estos resultados son consistentes con la diversidad y abundancia de sARNs encontrados previamente en otras especies vegetales, incluyendo Arabidopsis, maíz, tomate, soja, *M. truncatula* y maní, entre otras (Rajagopalan et al., 2006; Lelandais-Briere et al., 2009; Zhai et al., 2011; Chi et al., 2011; Li et al., 2013; Arikiti et al., 2014). Los miARNs y phasiARNs son reguladores post transcripcionales con funciones claves en el desarrollo de las plantas. En asociación con AGO1, tanto los miARNs como los phasiARNs son capaces de regular negativamente la expresión de los genes, clivando un ARNm blanco.

Tal como se mencionó en la Introducción de esta Tesis, los miARNs son pequeñas moléculas de 20 a 24 nts que regulan la expresión de los genes en eucariotas (Axtell et al., 2011; Moran et al., 2017). Estas moléculas usan la complementariedad del apareamiento de bases para guiar la regulación post transcripcional de la expresión de los ARNm. En animales el reconocimiento del mensajero blanco por el miARN y su apareamiento imperfecto resultan en una inhibición traduccional (Hutvagner and Simard, 2008; Ameres and Zamore, 2013). En plantas además de presentar esta función, la acción de los miARNs sobre el mensajero resulta más comúnmente en el clivaje endonucleolítico de este transcripto blanco (Bologna and Voinnet, 2014; Megraw et al., 2016). Muchos estudios destinados a identificar los miARNs y sus funciones se realizaron en la planta modelo *A. thaliana* (Jones-Rhoades et al., 2006). A partir de las secuencias de miARNs de *A. thaliana*, se identificaron posteriormente numerosos miARNs conservados en la Angiospermas, tanto en monocotiledóneas como en di-cotiledóneas, con funciones en desarrollo, respuestas a estreses y autorregulación de los miARNs. A pesar de que muchos miARNs han sido identificados en plantas leguminosas en esta última década (Dezulian et al., 2005; Zhang et al., 2006; Zhou et al., 2008; Szitty et al., 2008; Lelandais-Briere et al., 2009;

Zhai et al., 2011; Yan et al., 2016) pocos han sido caracterizados en profundidad. Sin embargo, algunos de ellos han sido caracterizados funcionalmente y validados por análisis genéticos en las plantas modelas *M. truncatula*, *L. japonicus* o *G. max*. Varios de estos miARNs han sido asociados a funciones importantes en el establecimiento de la simbiosis entre leguminosas y rizobios. El primer miARN asociado a la simbiosis fue el miR169a de *M. truncatula* (Combiér et al., 2006). Este miARN tiene como blanco de acción los transcritos que codifican varios miembros de la familia de factores de transcripción NF-YA. Estudios de genética reversa demostraron que el mismo regula la diferenciación del meristema de nódulos indeterminados restringiendo la expresión de *MtNF-YA1* a la región meristemática del nódulo de *M. truncatula*. La familia de miR169 constituye una de las familias de miARNs más numerosas en plantas y todos ellos cumplen funciones importantes en el desarrollo. Estos miARNs no fueron detectados a altos niveles de acumulación en nuestros datos de secuenciación de pequeños ARNs de raíces de *P. vulgaris*. Este resultado puede deberse a un efecto de dilución de estos miARNs cuando se colectan la raíz completa, dado que los mismos suelen presentar una expresión tejido-específica. Otros miARNs asociados a la simbiosis son el miR160, miR164, miR166, miR167, miR171, miR390, algunos de los cuales se discuten en las secciones siguientes de esta Discusión.

### **El miR167, una miARN conservado con acumulación diferencial**

En particular, el miR167 de 21 nts identificado en nuestros datos de secuenciación de sARNs de *P. vulgaris* fue descrito originalmente en Jones-Rhoades and Bartel, 2004 en un estudio de comparación computacional de miARNs conservados entre *Arabidopsis* y arroz. En *Arabidopsis*, los mARNs blancos putativos de miR167 son los factores de transcripción *ARF6* y *ARF8* (*At1g30330* y *AT5g37020*, respectivamente), los cuáles se encuentran también conservados en gimnospermas. En otro trabajo más reciente, dos isoformas de 21 nts del miR167, el miR167a y miR167c, fueron secuenciadas en bibliotecas de *M. truncatula* (Zhai et al., 2011). Ambas isoformas fueron detectadas en las bibliotecas de todos los órganos y condiciones analizadas en dicho estudio (semillas y plántulas, hojas, raíces, flores, nódulos, raíces micorrizadas o infectadas con nematodos). Los niveles de acumulación de las isoformas miR167a y miR167c fueron similares en nódulos y en raíces micorrizadas. Entre otros, el *ARF8* ha sido propuesto blanco putativo en esta especie. En este trabajo los autores presentan también resultados de la comparación de los datos de secuenciación entre varias leguminosas como entre *M. truncatula* y *G. max*. En *G. max*, solamente la isoforma del miR167c está presente, con abundancias relativas más elevadas en la biblioteca correspondiente al nódulo (aunque los datos de la biblioteca de hoja no están disponibles para esta especie). Interesantemente, en esta especie se observó un aumento en la acumulación del miR167c en pelos radicales en respuesta a la inoculación con rizobio (Zhai et al., 2011). Aunque el tejido no sea el mismo que el analizado en nuestros datos de secuenciación (raíces inoculadas contra pelos radicales inoculados respectivamente), los resultados obtenidos estarían de acuerdo con lo reportado en soja dado que en nuestros datos el miR167 muestra un incremento de su acumulación en respuesta a la bacteria más eficiente en raíces a 24 hpi (Figura 31). En el trabajo de Zhai et al. (2011), los autores declaran tener datos de bibliotecas de *P. vulgaris* de raíz, flores, nódulos y plántulas, pero estos datos no

fueron analizados y descritos en detalle. Según los autores, en ese momento no estaban disponibles los genomas de *P. vulgaris* y pocos miARNs estaban anotados en miRBase para poder llevar adelante un estudio comparativo. En el trabajo de Arenas-Huertero et al. (2009), la isoforma del miR167d de 22 nts fue identificada en *P. vulgaris*, con abundancias mayores en las muestras de hojas sometidas a un tratamiento de sequía.

### **El miR1514a, una nueva isoforma identificada en *P. vulgaris***

El miR1514a de 22 nts identificado en este trabajo de Tesis representa un caso interesante, ya que el mismo podría ser una isoforma específica de *P. vulgaris*. Una isoforma de 21 nts similar al miR1514a de 22 nts de *P. vulgaris* fue detectada en una hibridación por *Northern Blot* en plántulas de *G. max* (Arenas-Huertero et al., 2009). Sin embargo, la secuencia de los miARNs maduros de *P. vulgaris* (TTCATTTTGAAAATAGGCATTG) varía en dos nucleótidos respecto de la de *G. max* (TTCATTTTIAAAATAGGCATT) (Subramanian et al., 2008; Arenas-Huertero et al., 2009; Song et al., 2011). El miR1514a no fue detectado hasta el momento en plantas no leguminosas como *A. thaliana* ni en monocotiledóneas como *Oriza sativa*, *Zea mays*, ni tampoco en las otras leguminosas como *L. japonicus*, *M. truncatula* y *Pisum sativum* (Arenas-Huertero et al., 2009). Los alineamientos de secuencias de los precursores putativos del miR1514a reportados en Arenas-Huertero et al., 2009 muestran una conservación relativamente alta entre las tres especies *P. vulgaris*, *G. max* y *M. truncatula*. Contrariamente a los premiR1514a de *G. max* y *P. vulgaris* que fueron encontrados en las bases de datos de EST (*Expressed Sequence Tag*), los autores no indican que sea el caso del premiR1514a predicho para de *M. truncatula*. Aunque las secuencias de los precursores putativos del miR1514a estén conservadas entre *P. vulgaris* y *M. truncatula*, el miARN maduro no pudo ser detectado por *Northern Blot* en muestras de *M. truncatula*. Además la secuencia del miARN maduro de *M. truncatula* presenta dos *mismatches* en comparación a la secuencia de *P. vulgaris* (TTCATTTTCIAAAATAGGCATTG) (Arenas-Huertero et al., 2009).

El miR1514a identificado en nuestros datos de sRNA-seq presenta acumulaciones del miARN maduro muy elevadas en las tres condiciones ensayadas (Figura 32). Este es un miARN '*multi-loci*', y las estructuras de tallo-rulo formadas por los dos precursores putativos presentan estructuras compatibles y estabilidad similares a la de otros precursores de miARNs según las reglas definidas previamente por Rajagopalan et al., 2006b; Axtell et al., 2007; Axtell y Meyers, 2018. Es por ello que las acumulaciones elevadas observadas en las tres condiciones ensayadas podrían ser el resultado de la suma del procesamiento de ambos precursores. Cabe destacar que los niveles de acumulación del miR1514a fueron más bajos en las condiciones de inoculaciones en comparación a la del control, una tendencia que se mantuvo en las dos replicas biológicas. Ha sido reportado que en *P. vulgaris* el miR1514a se acumula diferencialmente en respuesta a estrés por sequía y a la aplicación exógena de ácido abscísico (ABA) (Arenas-Huertero et al., 2009). Ha sido reportado que el tratamiento de las raíces de leguminosas con ABA inhibe la formación de nódulos, sugiriendo que el ABA desempeña una función negativa sobre la nodulación (Ferguson and Mathesius, 2014). En nuestros datos de sRNA-seq, los niveles de acumulación de este

miARN fueron más bajos en las condiciones de inoculación con rizobios, sugiriendo que el miR1514a podría actuar como un regulador negativo de la nodulación mediando la respuesta a ABA.

En *G. max*, el miR1514 presenta dos isoformas, una de 21 nts (miR1514a) y otra de 20 nts (miR1514b), las cuales fueron identificadas en bibliotecas de raíces inoculadas por *B. japonicum* (Subramanian et al., 2008). El miR1514a de *G. max* tendría como blancos putativos mARNs que codifican proteínas de pegamiento de tipo NSF y factores de transcripción de la familia NAC (Song et al., 2011). El análisis *in silico* del miR1514a de *P. vulgaris* usando psRNATarget (Dai and Zhao, 2011) permitió identificar como blancos posibles varios transcritos que no tienen una anotación funcional disponible. A través de un análisis comparativo entre plantas de varios géneros el grupo de Zhai et al., 2011 ha mostrado la predominancia de miARNs de 22 nts en las leguminosas, especialmente para los que son generadores de phasiARNs. Teniendo en cuenta la fuerte acumulación de este miARNs de 22 nts en nuestra datos de secuenciación, sería interesante analizar si los *loci* de estos blancos putativos están asociados a la producción de phasiARNs de 21 nts en el lugar de corte para poder asignar una función a esta nueva isoforma del miR1514a en raíces de *P. vulgaris*.

### **El miR171a, un miARN de la respuesta simbiótica**

En la simbiosis fijadora de nitrógeno, el miR171c de *L. japonicus* ha sido asociado al proceso de infección del nódulo (Høgslund et al., 2009), y los experimentos de 5' RACE mostraron que actúa directamente sobre el factor de transcripción NSP2 mediante un mecanismo de clivaje endonucleolítico (De Luis et al., 2012). Por otra parte, el miR171h, ha sido reportado como regulador de la simbiosis endomycorrízica con hongos arbusculares (AM) en *M. truncatula*. En nuestros datos de secuenciación, se identificaron 7 isoformas de la familia del miR171 (Figura 34). Salvo para la isoforma miR171a, las abundancias de las otras isoformas fueron muy bajas en todas las condiciones analizadas en este trabajo (Tabla S4); al contrario del miR171a que presentó una acumulación diferencial significativa con respecto al control en las raíces inoculadas con la cepa SC15. Por otro lado, la medición del miR171a maduro por técnica de *Stem-loop* PCR mostró una acumulación diferencial entre cada condición verificando la tendencia observada en la datos de secuenciación masiva de sARNs (Figura 36). Este miR171a también fue identificado en bibliotecas de *P. vulgaris* realizadas a partir de muestras de pelos radicales tratados por factor Nod colectados a tiempos tempranos (6 hpi) (Formey et al., 2016). La acumulación diferencial del miR171a maduro fue validada en *P. vulgaris* por Formey et al., (2016) mediante qRT-PCR en muestras de pelos radicales a 6 hpi con el NF, donde aumento 3 veces respecto del control, y en muestras de raíces inoculadas con rizobio donde se incrementó 2 veces respecto de las raíces no inoculadas. A tiempos más tardíos de la interacción (raíces de 10 dpi con rizobio), la acumulación de este miARN fue mucho menor en comparación al control (Formey et al., 2016).

El blanco del miR171h en *M. truncatula* es *NSP2* (Devers et al., 2011; Laressergues et al., 2012; Ariel et al., 2012), un factor de transcripción responsable de la expresión de nodulinas tempranas (Kaló et al., 2005). En las condiciones de Formey et al., 2016,

así como en los datos de RNA-seq generados en nuestro grupo (Dalla Via et al., 2015; Clúa et al., datos no publicados) los perfiles de acumulación de *NSP2* de *P. vulgaris* (*phvul.008G143400*) fueron inversamente proporcionales a la acumulación del miR171a, sugiriendo que el miR171a en *P. vulgaris* podría tener el mismo blanco que el miR171h de Medicago, aunque sus secuencias sean muy diferentes (solamente 50 % de homología para las secuencias de los miR171 maduros, Figura 34). Por otra parte, el factor de transcripción de tipo GRAS SCR (*Phvul.010G072700.1*) fue identificado en los datos de degradoma generados por Formey et al., 2016. Sin embargo, este target putativo no mostró diferencias en nuestros datos de RNA-seq (Dalla via et al., 2015). Estos resultados sugieren que, en nuestras condiciones experimentales, el miR171a de *P. vulgaris* podría actuar sobre el mRNA de *NSP2* reprimiendo su expresión mediante un mecanismo de clivaje, mientras que SCR no sería un blanco de acción de dicho miARN o podría ser regulado a nivel de inhibición traduccional. En este sentido resultaría interesante investigar los patrones de expresión espaciales del miR171a y de sus target putativos a fin de determinar si los mismos se expresan en los mismos tejidos y/o tipos celulares.

Los datos de bibliografía indican que las isoformas miR171c y miR171h, dos isoformas muy relacionadas en secuencia (una única base de diferencia en el 5'), poseen funciones relevantes en las simbiosis. En la simbiosis rizobiana, el miR171c está involucrado a tiempos tardíos de la interacción en *L. japonicus* (De Luis et al., 2012). Nuestros datos de expresión sumados a los de Formey et al., 2016 sugieren que la isoforma del miR171a de *P. vulgaris* podría desempeñar una función importante a tiempos tempranos de la interacción simbiótica, en particular en el establecimiento de una simbiosis más eficiente. Sin embargo, experimentos de genética funcional permitirían validar su función y determinar si dicho miARN está implicado en la preferencia de cepa observada en las variedades Mesoamericanas de *P. vulgaris*.

### **Identificación de nuevos miARNs por secuenciación masiva**

La secuenciación de nueva generación (NGS, *Next Generation Sequencing*) es una de las tecnologías más potentes para caracterizar pequeños ARNs a una escala genómica funcional e identificar nuevos miARNs dado que la misma ofrece la sensibilidad, especificidad y facilidad necesarias para maximizar la obtención de datos (profundidad), minimizando a su vez los costos generados (Morozova and Marra, 2008; Buermans and den Dunnen, 2014; Wong et al., 2013). La facilidad de acceso a estas nuevas tecnologías ha obligado a los investigadores a ser más cautelosos y meticulosos al momento de llevar adelante el análisis de los datos (Head et al., 2014). La inspección manual de los mismos evita la generación de datos contradictorios para la comunidad científica ya que la producción de datos a gran escala es liberada en bases de datos públicas y de libre de acceso que no está sujeta a un control de calidad sobre el significado biológico. El mayor desafío a la hora de generar una biblioteca para secuenciación masiva es evitar lo máximo de sesgos posibles. Los sesgos se definen como la distorsión sistemática de la data dada por el diseño experimental. Para eso las estrategias a adoptar para eliminar las fuentes de sesgos experimentales son primero informarse bien sobre la técnica que se va a utilizar a fin de conocer los sesgos frecuentes y tratar de minimizarlos. Segundo, para los sesgos

que son inevitables, se debe ajustar el diseño experimental de manera que este tenga un impacto mínimo en el análisis (Head et al., 2014). Los sesgos sistemáticos conocidos en la generación de datos de secuenciación masiva son los que resultan de manipulaciones moleculares incorrectas (*batch effect*, por los experimentos realizados en tanda). Aquellos resultan de la variabilidad del procesamiento de la muestra según las condiciones de reacción, los reactivos, la precisión del pipeteo, y el operador (Head et al., 2014). En un laboratorio de biología molecular se ve comúnmente en resultados variables de PCR de rutina, y en el caso de secuenciación masiva, se ha visto variabilidad entre corridas de secuenciación o entre diferentes líneas de un secuenciador. Un sesgo que se encuentra frecuentemente en la construcción de bibliotecas de secuenciación masiva es introducido por un alto número de ciclos de PCR. Las amplificaciones extensivas suelen introducir sesgos sobre representando secuencias con amplificación preferencial. En este trabajo, la etapa de amplificación por PCR del protocolo de la construcción de las bibliotecas (TruSeq Small RNA Sample Preparation Kit de Illumina) se realizó utilizando sólo 11 ciclos, que corresponde al programa estándar optimizado y recomendado por los proveedores, por lo que consideramos que este bajo número de ciclos no debe haber afectado significativamente la representatividad de los pequeños ARNs en cada una de las bibliotecas. Tanto en la etapa de transcripción reversa como en la amplificación por PCR (dado particularmente por el contenido en GC) muestran preferencia de templado, favoreciendo la amplificación de pequeños ARNs en comparación a otros. También la eficiencia de la síntesis de ADNc puede estar fuertemente afectada por el uso de adaptadores y códigos de barras diferentes, las estructuras complejas del ARN y sus modificaciones. Al contrario, se ha mostrado que la elección de la plataforma de secuenciación tiende a tener un efecto mínimo (Raabe et al., 2014).

En la sección de Resultados del presente capítulo se han mencionados algunas pocas inconsistencias entre réplicas biológicas, particularmente en los análisis de poblaciones de los nuevos miARNs putativos y los phasiARNs. La variabilidad observada entre ambas réplicas puede deberse a que aquellas fueron obtenidas en diferentes momentos del año, por distintos operadores o bien a partir de diferentes lotes de semillas.

En el caso de la identificación de nuevos miARNs, si bien la presencia del miARN\* es una condición requerida, no es un criterio suficiente para poder anotar dichas secuencias como miARNs nuevos (Meyers et al., 2008). Una validación robusta consiste en combinar la evidencia de expresión de un miARN (acumulación por *RNA Blot* o secuenciación de sARNs) con evidencia de su biogénesis a partir de la presencia de la región tallo-rulo, conservación filogenética del miARN, y/o cambio en su abundancia en un mutante de alguna de los genes *dicer* (Ambros et al., 2003). Recientemente, la necesidad de minimizar los falsos positivos llevaron a Axtell y Meyers (2018) a publicar una actualización acerca de las reglas para la anotación de nuevos miARNs. En dicha actualización aclaran que la detección de un pequeño ARN en un *RNA Blot* no permite distinguir entre miARNs, siARNs endógenos u otros fragmentos de ARN parcialmente degradados. A su vez, agregan que la conservación filogenética no tendría que ser un criterio válido para la anotación. En plantas, la disminución o acumulación de un sARN en mutantes de *dicer* no es definitiva, ya que *DCL* pertenece a una familia multigénica con funciones redundantes en la biogénesis

de los miARNs y siARNs. Por esto, este criterio sólo permite validar que el sARN medido no sería un producto de degradación sino de procesamiento por algunas DCL. Otra diferencia que se puede destacar de esta nueva recomendación para la anotación de nuevos miARNs en comparación a la anterior es la necesidad de que los criterios sean válidos por lo menos en dos réplicas biológicas de sRNA-seq. En nuestro análisis, seleccionamos las secuencias que presentan un miARN\* al menos en una de las réplicas biológicas, ya que al momento de abordar el presente trabajo de Tesis las recomendaciones para la anotación de miARNs propuestas por Meyers et al. (2008) aceptaban como este criterio como válido. Es por ello que, se planea en un futuro cercano reanalizar los datos a la luz de estos nuevos criterios propuestos por Axtell y Meyers (2018).

Utilizando los criterios recomendados por Meyer et al., 2008, en este trabajo se identificaron 48 miARNs que tuvieron valores de RPM diferentes de 0 en las raíces de *P. vulgaris* en las tres condiciones ensayadas (CTL, SC15 y 55N1) y en ambas réplicas biológicas. De estos 48, 10 fueron regulados, ya sea positiva o negativamente, específicamente frente a la inoculación con la cepa que establece una interacción más eficiente SC15. 9 se regularon específicamente en respuesta a la cepa menos eficiente 55N1 y sólo uno presentó fue regulado en respuesta a ambas cepas. Este resultado sugieren que los miARNs nuevos participarían de las respuestas específicas que son activadas en la variedad mesoamericana por cepas provenientes del mismo o de diferente centro de domesticación de *P. vulgaris*. Estos nuevos miARNs con una respuesta específica de cepa podría ser parte de los mecanismos moleculares que son determinantes para la selección de cepa y por lo tanto son los candidatos más interesantes a estudiar en mayor profundidad mediante análisis funcionales y genéticos. En la presente trabajo de Tesis se seleccionó uno de ellos para profundizar su estudio, el miARN#4/miR5924, y los resultados de caracterización funcional y análisis de sus blancos putativos se presenta en el Capítulo 3 de esta Tesis Doctoral.

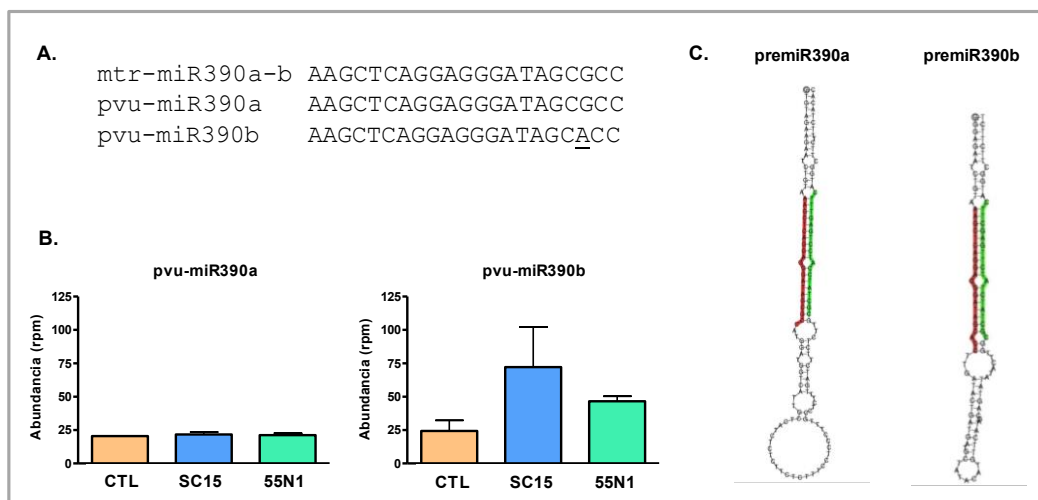


# *Capítulo II*

# El miR390b y su vía de regulación en *P. vulgaris*

## 1. La vía de regulación del miR390b a 24 hpi

En la lista de los miARNs conservados identificados en el Capítulo I de esta Tesis, el miR390b presentó un perfil de expresión interesante ya que el mismo se acumuló a mayores niveles en las raíces inoculadas con *R. etli* respecto a aquellas de la condición control inoculadas con medio YEM (Figura 25 del Capítulo I). En nuestras bibliotecas de sRNA-seq se secuenciaron dos isoformas del miR390 de 21 nts, miR390a y miR390b, las cuales difieren únicamente en una base en la posición 19 (Figura 45A). En *M. truncatula*, dos precursores diferentes son procesados para dar lugar a una única isoforma madura que corresponde al miR390a-b, la cual es idéntica a la isoforma miR390a de *P. vulgaris* (Figura 45A). Fue interesante observar que las dos isoformas presentan niveles de acumulación diferentes del miR390 maduro en raíces de 24 horas (Figura 45B). Además de tener menores niveles de acumulación relativa que el miR390b, el miR390a no presentó diferencias de acumulación entre las tres condiciones ensayadas en nuestro estudio. Por el contrario, el miR390b presenta mayores niveles en las raíces inoculadas con SC15 respecto de las plantas control, y en menor medida de aquellas inoculadas con 55N1, la bacteria menos eficiente. Aunque estas diferencias fueron marcadas en la réplica biológica #1, las dos replicas siguen las mismas tendencias (Tabla S4 de la sección Anexos).



**Figura 45: Los miR390 de *P. vulgaris* a 24 hpi.** A. Secuencias de las dos isoformas del miR390 de *P. vulgaris* alineadas con la de *M. truncatula*. El miR390b se diferencia por la base subrayada. B. Abundancias normalizadas en lecturas por millón (RPM) de los miR390a y miR390b en la data de secuenciación de sRNA-seq. Se graficaron los promedios de las dos replicas biológicas para cada isoforma. C. Estructura secundaria del precursor del miR390a y miR390b de *P. vulgaris*. Los miARNs maduros están resaltados en rojo, asociados a los miARNs\* (resaltados en verde). Las estructuras tallo-rulo fueron obtenidas usando la herramienta de visualización de UEA Workbench. MFE(pvu-premiR390a)= -54,17; MFE(pvu-premiR390b)= -46,40.

El miR390 corresponde a una familia de miARNs conservada en plantas. En *P. vulgaris*, el miR390a alineó con una región de la hebra negativa del cromosoma 10 (12464336-12464356) según la anotación del genoma mesoamericano. El miR390a podría provenir de dos precursores: uno transcrito a partir de la región del cromosoma 10 y otra región del cromosoma 6. Sin embargo el miR390a no presentó lecturas del miARN\* asociadas a este último locus en nuestros datos de secuenciación. Por otra parte, el miR390b alineó con una sola región del cromosoma 8 (56694789-56694809) de la hebra negativa, sugiriendo que el miR390b de *P. vulgaris* sería procesado a partir de un único precursor. Las estructuras secundarias predichas para los precursores del miR390a y 390b de *P. vulgaris* presentan sólo uno o dos nucleótidos desapareados en la región miARN/miARN\* y dos nucleótidos 3' protuberantes (Figura 45C).

Para determinar si los precursores de los miR390a y miR390b de *P. vulgaris* se encuentran conservados con los de *M. truncatula*, se realizó un alineamiento múltiple de las secuencias obtenidas en nuestra data de secuenciación y en miRBase para los precursores de *P. vulgaris* y *M. truncatula*, respectivamente (Figura 46). Resultó interesante constatar que tanto los premiR390b como los premiR390a de las dos especies presentaban alta identidad de secuencia, sugiriendo un origen evolutivo común. Comparando entre los diferentes precursores, las secuencias de los miARNs\* presentan 100% de identidad, con la excepción del premiR390b de *P. vulgaris* que presenta 3 mismatches. El segundo precursor del miR390a de *P. vulgaris* (ubicado en el cromosoma 6) fue el que presentó mayores diferencias en comparación a las otras secuencias por lo cual no fue incluido en el alineamiento de la figura 2. Su divergencia sumado a la ausencia de secuencia del miARN\* asociado a este precursor en nuestras bibliotecas podría sugerir que el mismo haya perdido funcionalidad.



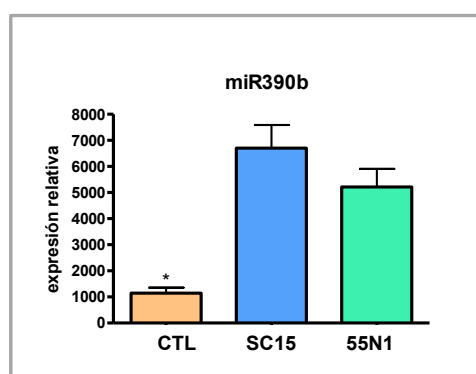
**Figura 46: Precursores de los miR390 de leguminosas.** Alineamientos de las secuencias de los precursores de los miR390a y miR390b de *P. vulgaris* (pvu) y de *M. truncatula* (mtr). Los alineamientos de secuencias múltiples se realizaron con Clustal Omega. Los asteriscos representan nucleótidos conservados. En rojo están indicados los miR390 maduros y en verde el pvu-miR390\*. Los tamaños de los precursores (en nts) están indicados a la derecha.

Resultó interesante constatar que los precursores del miR390a entre *P. vulgaris* y *M. truncatula* fueron los que presentaron el porcentaje de identidad más elevado con 84

%, mientras los precursores del miR390b presentaron solamente 67,4 % de identidad entre las dos especies. Los premiR390a y premiR390b de *M. truncatula* presentan una identidad fuerte con 77,5 %; al contrario de los premiR390a y premiR390b de *P. vulgaris* que son los más divergentes con 64 % de identidad. Por otra parte, el premiR390b de *M. truncatula* presenta más porcentaje de identidad con el premiR390a de *P. vulgaris* (71,2 %) que con el premiR390b de *P. vulgaris* (67,4 %). Finalmente el premiR390b de *P. vulgaris* presenta un porcentaje de identidad similar tanto con miR390b de *M. truncatula* que el mtr-miR390a de (66,7 %).

Las divergencias de las secuencias del precursor del miR390b de *P. vulgaris* y de su miARN\*, sumado al perfil diferencial del miR390b maduro a 24 horas después de la inoculación de rizobios sugirieron que el miR390b podría ejercer una función específica en esta leguminosa, lo que nos incentivó a profundizar su análisis.

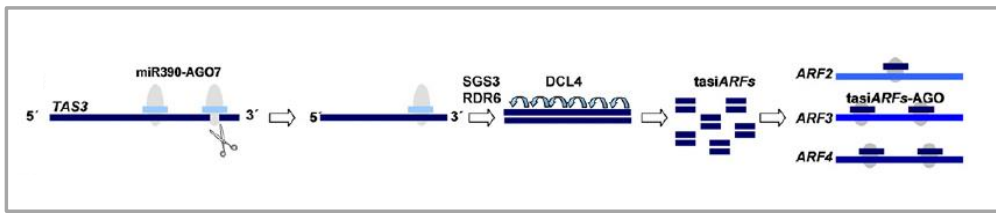
Con el objetivo de validar la acumulación diferencial observada en las bibliotecas, se evaluaron los niveles de acumulación del miR390b maduro en las tres condiciones ensayadas utilizando la técnica de *stem-loop* RT-PCR (Chen et al., 2005; Varkonyi-Gasic et al., 2007). Los resultados obtenidos indicaron que la acumulación del miR390b respecto al control es significativamente mayor en aquellas raíces inoculadas con ambas cepas. La acumulación del miR390b respecto al control es mayor con SC15 (aproximadamente 7 veces), que en aquellas inoculadas con 55N1 (5 veces) (Figura 47). Este resultado es coincidente con los resultados obtenidos previamente por sRNA-seq.



**Figura 47: Abundancia del miR390b por real time.** Medición de la abundancia del miR390b por la técnica *Stem-Loop* PCR en muestra de raíces de 24 hpi. El asterisco indica diferencias significativas del control respecto de las condiciones de inoculación en un *t*-test con  $p < 0,05$ .

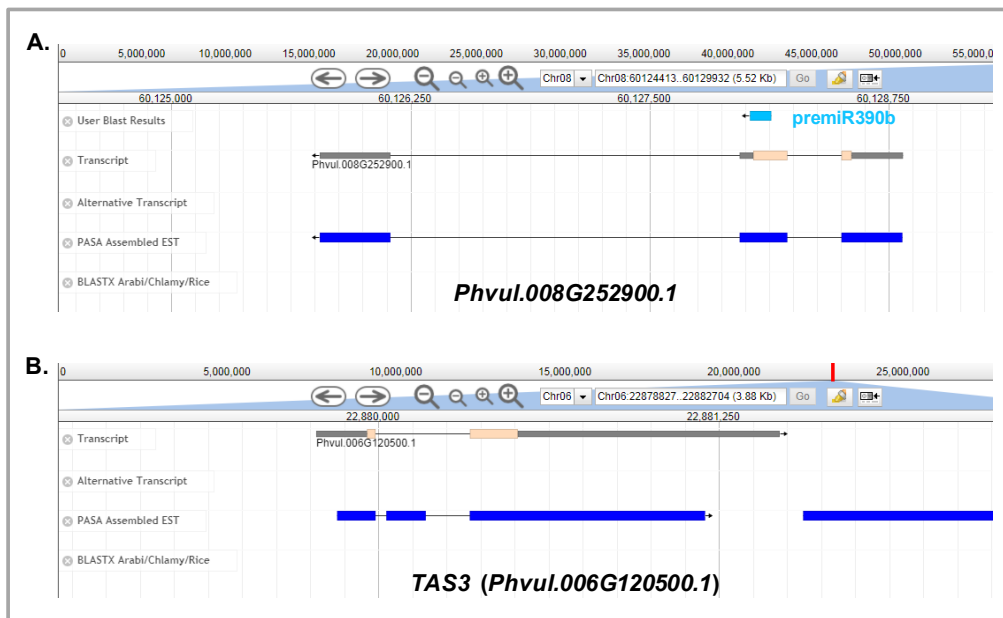
El miR390 tiene como ARNm blanco un transcripto no-codante, denominado *TAS3* (por *trans-acting small interference RNA3*), a partir de cual se generan tasiARNs (Allen et al., 2005; Axtell et al., 2006; Jagadeeswaran et al., 2009; Marin et al., 2010), (Figura 48). El miR390 se asocia específicamente a la proteína AGO7 y se une a dos sitios complementarios de *TAS3*. El complejo AGO7/miR390 cliva el transcripto *TAS3* en el sitio blanco más próximo al extremo 3' terminal. El producto de clivaje es convertido en ARN doble cadena (dsRNA) por la acción del complejo formado por RDR6 y SGS3. La RNAsa DCL4 procesa el dsRNA en pequeños ARNs de 21 nts: los tasiARFs; los cuales son producidos en fase de a 21 nts a partir del sitio de clivaje de miR390. Los tasiARFs actúan como reguladores post-transcripcionales negativos de los ARNms *ARF2*, *ARF3* y *ARF4*, uniéndose en uno o dos sitios complementarios (Montgomery et al., 2008). Los ARFs (por *Auxine Response Factors*) están involucrados en la señalización y respuestas a las auxinas, regulando positivamente el crecimiento lateral

de las raíces y ejerciendo una regulación negativa sobre la infección rizobiana y la organogénesis del nódulo en *M. truncatula* (Hobecker et al., 2017).



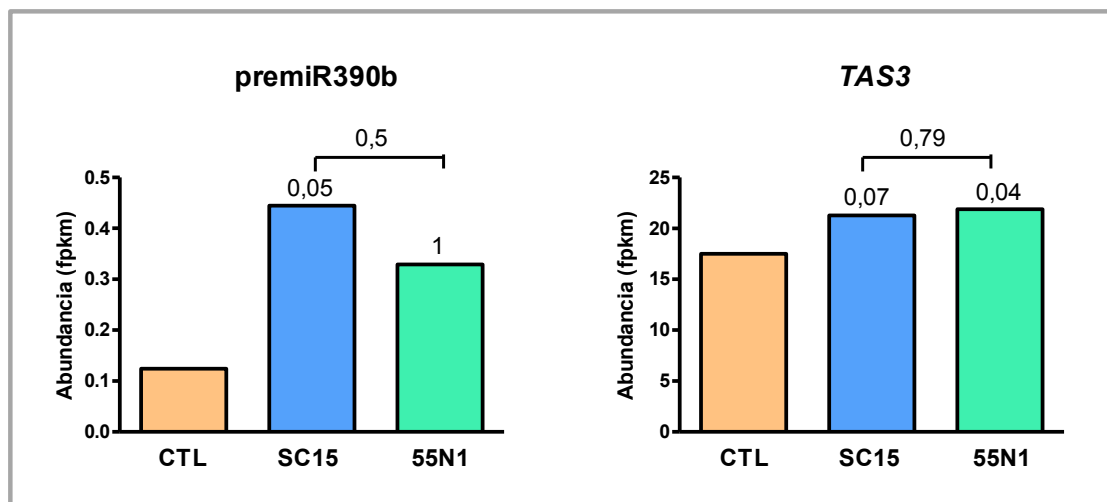
**Figura 48: Vía de regulación del miR390.** El complejo AGO7/miR390 cliva *TAS3* al segundo sitio de unión. RDR6 y SGS3 convierte el producto de clivaje en ARN doble cadena, el cual es procesado por DCL4 y convertido en tasiARNs. Los tasiARNs regulan los mARNs *ARF2*, *ARF3* y *ARF4*. Esquema obtenido de Hobecker et al., 2017.

La secuencia del precursor del miR390b fue alineada al genoma andino de *P. vulgaris* disponible en Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>), una plataforma *online* de base de datos genómica y herramientas para el análisis de secuencias en plantas. Resultó interesante observar que el precursor de este miR390b alineó con 100% de identidad con el *locus* que da lugar a transcripto *Phvul.008G252900.1*, el cual no posee anotación funcional (Figura 49A). Los precursores del miR390a de *P. vulgaris* (de los cromosomas 10 y 6), alinearon en zonas intergénicas (datos no mostrados). Por otra parte, el transcripto no codificante *TAS3* (*Phvul.006G120500.1*) fue identificado en una región del cromosoma 6 de *P. vulgaris*, que tampoco presenta anotación funcional (Figura 49B).



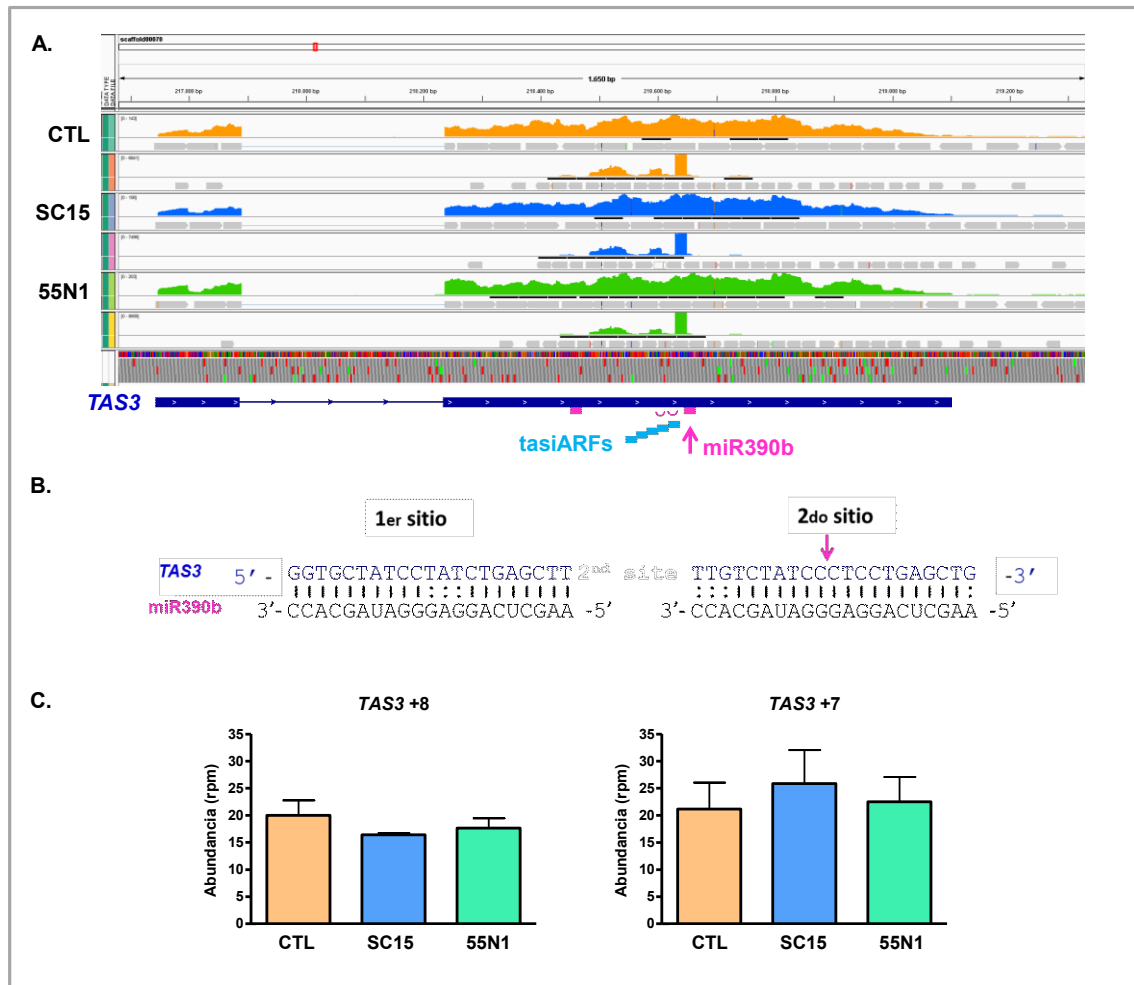
**Figura 49: Alineamientos del precursor y del blanco del miR390b en *P. vulgaris*.** A. Alineamiento del premiR390b (*Phvul.008G252900.1*) al genoma de *P. vulgaris* en Phytozome v2.1. B. Alineamiento de *TAS3* (*Phvul.006G120500.1*). En el modelo génico propuesto las líneas corresponden a intrones y las cajas a exones. Las cajas naranjas corresponden a regiones codificantes y las grises a las 5' y 3' UTRs. La región correspondiente al premiR390b se muestra en celeste.

Posteriormente, se analizaron los niveles de expresión del precursor y de *TAS3* haciendo uso de los datos de RNA-seq de raíces a 24 hpi generados en nuestro grupo de investigación. Los valores de expresión del premiR390b obtenidos a partir de la data de RNA-seq mostraron una correlación directa con la acumulación del miR390b maduro que resultó ser significativa para la condición SC15 respecto al control, indicando que tanto el precursor como el miARN maduro son modulados por el rizobio en estadios tempranos de la interacción simbiótica. Por otro lado, el transcripto *TAS3* no mostró diferencias importantes en respuesta a ninguna de las dos cepas rizobianas (Figura 50), de manera consistente con lo que se había observado previamente en la leguminosa modelo *M. truncatula* (Hobecker et al., 2017).



**Figura 50: Niveles de expresión del precursor y del blanco del miR390b a 24 hpi.** Valores de expresión normalizados de los transcritos *Phvul.008G252900.1* (premiR390b) y *Phvul.006G120500.1* (*TAS3*) en FPKM obtenidos a partir de los datos de RNA-seq de tres réplicas biológicas disponibles en el laboratorio (Tesis doctoral de Joaquín Clúa). Los valores de probabilidad de expresión diferencial (*p-value*) de las condiciones de inoculación contra el control (SC15 vs CTL o 55N1 vs CTL) se indican arriba de las barras, y de SC15 vs 55N1 se señala arriba del corchete.

Usando la plataforma Integrative Genomic Viewer (IGV), se buscó la región cromosómica correspondiente a *TAS3* en el genoma mesoamericano de *P. vulgaris*. Se cargaron los datos de RNA-seq de nuestros datos de sRNA-seq provenientes de las dos réplicas biológicas y se ajustaron las escalas de manera a que estén normalizadas entre sí. En el modelo génico para el transcripto de *TAS3* se ubicaron los dos sitios de unión del miR390b. Resultó interesante observar la acumulación de lecturas de pequeños ARNs entre los dos sitios de unión. El análisis de las secuencias de estas lecturas confirmó que aquellas corresponden a pequeños ARNs de 21 nts en fase a partir del sitio de clivaje del miR390b (Figura 51A y B), es decir los tasiARFs. En la Figura 51C se graficaron los dos tasiARFs con valores de abundancias mayores en nuestros datos de secuenciación, *TAS3+7* y *TAS3+8*, los cuales corresponden a los tasiARNs más estables reportados previamente en otras especies vegetales (Jagadeeswaran et al., 2009; Hobecker et al., 2017).



**Figura 51: Visualización de las lecturas de RNA-seq y sRNA-seq alineadas al locus *TAS3*.** A. Lecturas de *TAS3* (*Phvul.006G120500.1*) alineadas al genoma de *P. vulgaris* con datos de secuenciación de RNaseq de Dalla Vía (panel superior, y lecturas de sARNs con datos de sRNAseq (panel inferior) para cada condición (CTL, SC15, 55N1). Los datos de las dos replicas biológicas fueron combinados para cada condición. El modelo génico del transcripto se indica abajo en azul, los sitios del mi390b en fucsia y la producción de tasiARFs en celeste. B. Detalles del apareamiento del miR390b a los dos sitios blancos de *TAS3*. La flecha rosa indica la zona de corte. C. Abundancias normalizadas en lecturas por millón (RPM) de los tasiARFs (*TAS3* +8, *TAS3* +7) obtenidas a partir de sRNA-seq. Se graficaron los promedios de las dos replicas biológicas.

## 2. Caracterización funcional del miR390b

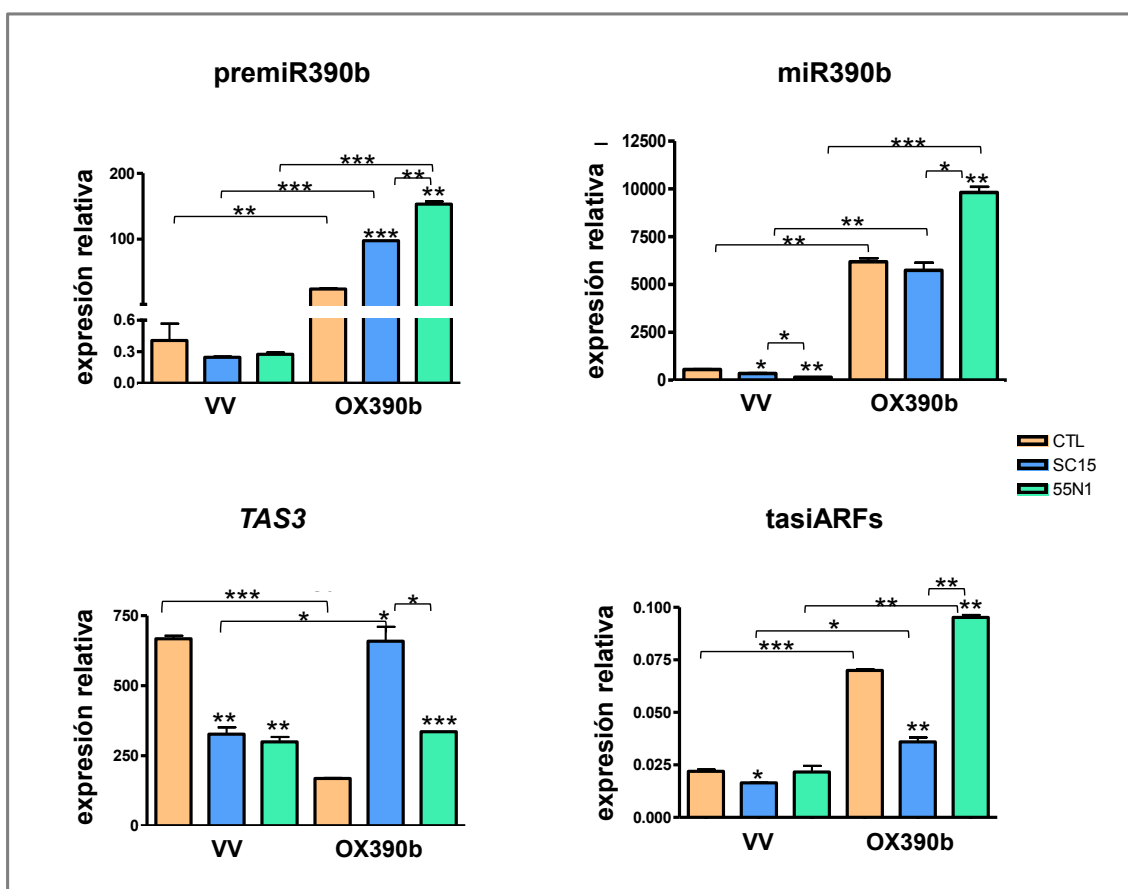
Los datos de expresión del miR390b a 24 hpi sugieren que el mismo podría desempeñar una función en el establecimiento de la simbiosis rizobiana de *P. vulgaris*. Para determinar si el miR390b está involucrado en la eficiencia de la nodulación y la preferencia de cepa observada en las variedades mesoamericanas de *P. vulgaris*, se decidió caracterizar mediante genética reversa la función de dicho miARN en la formación de nódulos, la infección rizobiana y en la ocupación de los nódulos por

cepas más eficientes. Para ello, se llevó a cabo la sobreexpresión del precursor del miR390b en raíces de *P. vulgaris*. La secuencia correspondiente al precursor fue amplificada a partir de ADNc utilizando dos oligonucleótidos específicos que amplifican un fragmento de 217 nts. Los fragmentos obtenidos fueron clonados en el vector de expresión pk7WG2D,1 (Figura 72 de la Sección Materiales y Métodos), el cual dirige la expresión del premiARN bajo el control del promotor del ARN 35S del virus del mosaico del coliflor (CaMV35S). Esta construcción se denominó p35S:premiR390b. El vector de expresión utilizado posee el gen reportero que codifica la proteína verde fluorescente GFP bajo el promotor constitutivo *rolD* de *Agrobacterium tumefaciens*, lo que permite visualizar el tejido transformado (Figura 69 de la Sección Materiales y Métodos). La construcción para la sobreexpresión del precursor del miR390b (premiR390b) fue introducida en la cepa de *Agrobacterium rhizogenes* K599, y dicha cepa utilizada para generar raíces transgénicas en plantas compuestas de *P. vulgaris*. Estas plantas compuestas consisten en una parte aérea salvaje (no transgénica) y su sistema radical transgénico en donde cada raíz representa un evento de transformación independiente. Un grupo de plantas fue transformado con el vector vacío como control. Una vez que emergieron las raíces transgénicas, las plantas fueron crecidas durante una semana en cajas conteniendo un medio Fahraeus sin fuente de nitrógeno y luego visualizadas bajo lupa de fluorescencia para identificar aquellas raíces que expresan GFP. Aproximadamente el 90 % de las raíces presentaron expresión de la proteína GFP. Las raíces que no presentaban fluorescencia fueron escindidas.

Para evaluar los niveles de acumulación del premiR390b, se extrajeron los ARNs totales a partir de tejido de raíz de plantas individuales de plantas transgénicas llevando la construcción con el vector vacío (VV) y plantas de sobreexpresión (OX390b). Una planta para cada condición de inoculación realizada en nuestro estudio fue evaluada para cada una de las condiciones ensayadas: inoculadas con YEM (CTL), SC15 (cepa de *R. etli* más eficiente) y 55N1 (cepa de *R. etli* menos eficiente) y colectadas a las 24 horas luego de inoculación. A partir del ARN obtenido se llevó a cabo la síntesis de ADNc mediante transcripción reversa (RT) y se realizaron reacciones de qPCR utilizando oligonucleótidos específicos para el premiR390b (Tablas S12 de la Sección Anexos). Las plantas evaluadas llevando la construcción de sobreexpresión 35S:premiR390b (OX390b) mostraron valores de acumulación importantes del precursor en comparación a las plantas control transformadas con el vector vacío, para las tres condiciones de inoculación enunciadas (Figura 52). Para determinar si la sobreexpresión del precursor del miR390b se correlaciona con la acumulación del miARN maduro, se cuantificaron los niveles de acumulación del miR390b en aquellas mismas muestras utilizando la técnica de *stem-loop* RT-PCR (Chen et al., 2005; Varkonyi-Gasic et al., 2007). Los resultados obtenidos indicaron que la acumulación del miR390b es significativamente mayor en aquellas raíces que llevan la construcción de sobreexpresión que en sus controles respectivos en las tres condiciones ensayadas (Figura 52). Para cada condición, los perfiles de acumulación del premiR390b y del miR390b maduro siguen correlaciones similares. En las mismas muestras se evaluaron los valores de expresión de *TAS3* (Figura 52). Salvo para la condición control (CTL), los valores de expresión de *TAS3* no siguen una correlación inversa con los valores de acumulaciones del miR390b. Sin embargo, en los datos de bibliotecas a 24 hpi tampoco se había notado una disminución del blanco del miR390b



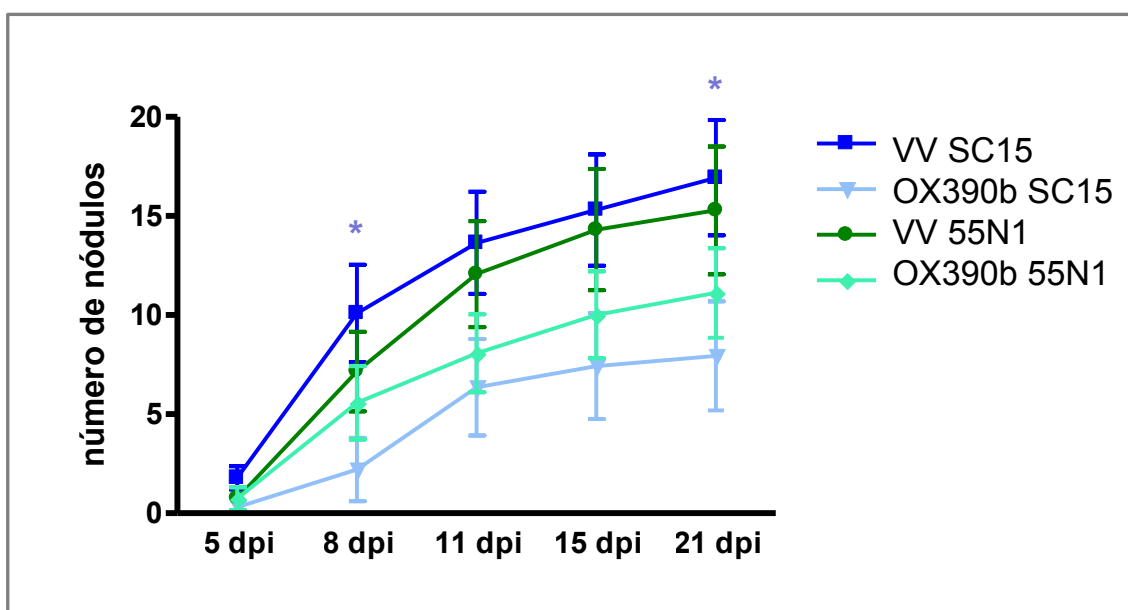
en muestras en cuales el miR390b (Figura 45B) y su precursor (Figura 50) se encuentran aumentados. Para la totalidad de las muestras analizadas, el transcrito *TAS3* ha siempre mostrado una expresión muy alta, lo que podría explicar que la acción del miR390b tendría un efecto menor sobre los niveles estacionarios de *TAS3*, sugiriendo que el miR390b más que *TAS3* es el factor limitante para la producción de tasiARFs. Finalmente, en estas muestras se midieron las acumulaciones de los tasiARFs. Los *TAS3+7* y *TAS3+8* presentan una única diferencia en el 3' terminal de sus secuencias nucleotídicas (TTCTTGACCTTGTAAGACCTC, TTCTTGACCTTGTAAGACCTI, respectivamente) que no permite distinguir las mediante la técnica de *stem-loop* RT-PCR. La medición de los tasiARFs en estas muestras reveló una correlación directa entre las acumulaciones del miR390b y los tasiARFs en las diferentes condiciones ensayadas (Figura 52).



**Figura 52: Niveles de expresión de la vía del miR390b a 24 hpi.** Medición de la expresión del premiR390b y *TAS3* por real time; y de la abundancia del miR390b y de los tasiARFs por la técnica *Stem-Loop* PCR en muestra de raíces control (VV) y de raíces de sobreexpresión del miR390b (OX390b) a 24 hpi. Los asteriscos indican diferencias significativas respecto al control y entre las diferentes muestras señaladas por los corchetes en un *t*-test (con \*,  $p < 0,05$ ; \*\*,  $p < 0,01$ ; \*\*\*,  $p < 0,0001$ ).

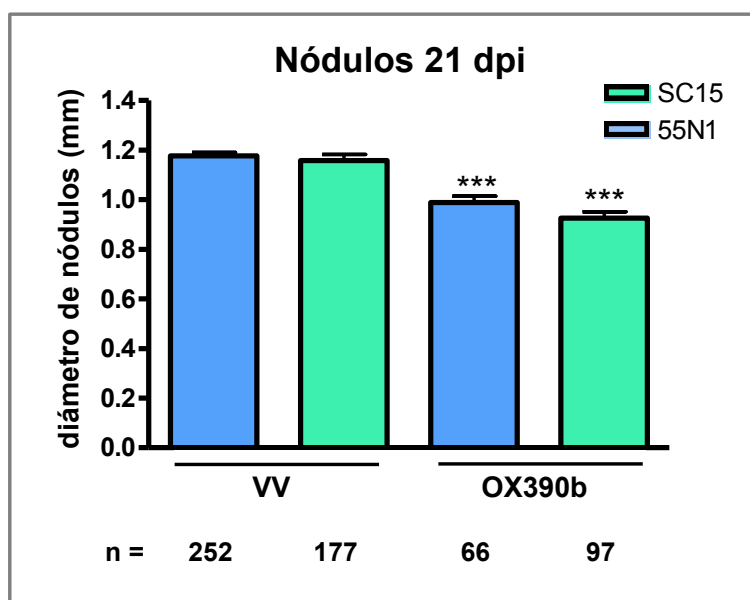
Habiendo verificado la sobreexpresión de los precursores del miR390b y la acumulación consecuente de su miARN maduro así como una mayor producción de tasiARFs en las raíces OX390b, se abordó la caracterización fenotípica de dichas plantas, la cual incluyó la cuantificación de los nódulos por raíz en función del tiempo

(cinética de nodulación) y la determinación del tamaño de los nódulos formados luego de la inoculación con las cepas de *R. etli* SC15 y 55N1. Las plantas compuestas de variedad mesoamericana NAG12 que llevan la construcción p35S:premiR390b (OX390b) o el vector vacío (VV) fueron transferidas a cajas conteniendo medio Fahreaus sin fuente de nitrógeno. Se utilizó un mínimo de 10 plantas por construcción y por condición, las cuales fueron inoculadas con la cepa más eficiente SC15 o con la menos eficiente 55N1. Los nódulos formados fueron cuantificados por raíz dado que cada raíz representa un evento de transformación independiente. La recolección de los datos comenzó a partir de los 5 días post inoculación (dpi), momento en el cual los nódulos o primordios de nódulos son visibles a simple vista. Los resultados de la cinética de nodulación mostraron que las plantas OX390b desarrollaron una cantidad de nódulos significativamente menor que las control (VV) cuando las mismas fueron inoculadas con SC15. Esta disminución fue significativa al largo del tiempo estudiado (i.e 8 y 21 dpi). Interesantemente, las diferencias observadas entre las plantas OX390b y control no fueron significativas cuando se utilizó la cepa menos eficiente 55N1 (Figura 53). Estos resultados indican que la sobreexpresión del premiR390b afectaría negativamente la formación de nódulos en la simbiosis de *P. vulgaris* y que esta alteración es más pronunciada en la interacción con la cepa más eficiente SC15. Esto sugiere que el miR390b actuaría como un regulador negativo de la nodulación.



**Figura 53: Cinética de nodulación del miR390b.** Número de nódulos formados por raíz transgénica en plantas control (VV) y plantas que sobreexpresan el precursor del miR390b (OX390b), frente a la inoculación con las cepas de *R. etli* SC15 (*nodC*  $\alpha$ ) y 55N1 (*nodC*  $\delta$ ). Las barras de error representan los desvíos estándar del conteo de la totalidad de las raíces GFP de un mínimo de 10 plantas por condición. Se graficaron juntos los resultados de conteo de raíces transgénicas generadas por 2 colonias de *A. rhizogenes* llevando la misma construcción. Las diferencias fueron significativas en un *t*-test con  $p < 0,05$  en los tiempos señalados por los asteriscos en las muestras de inoculación con SC15. El experimento es representativo de tres réplicas biológicas independientes.

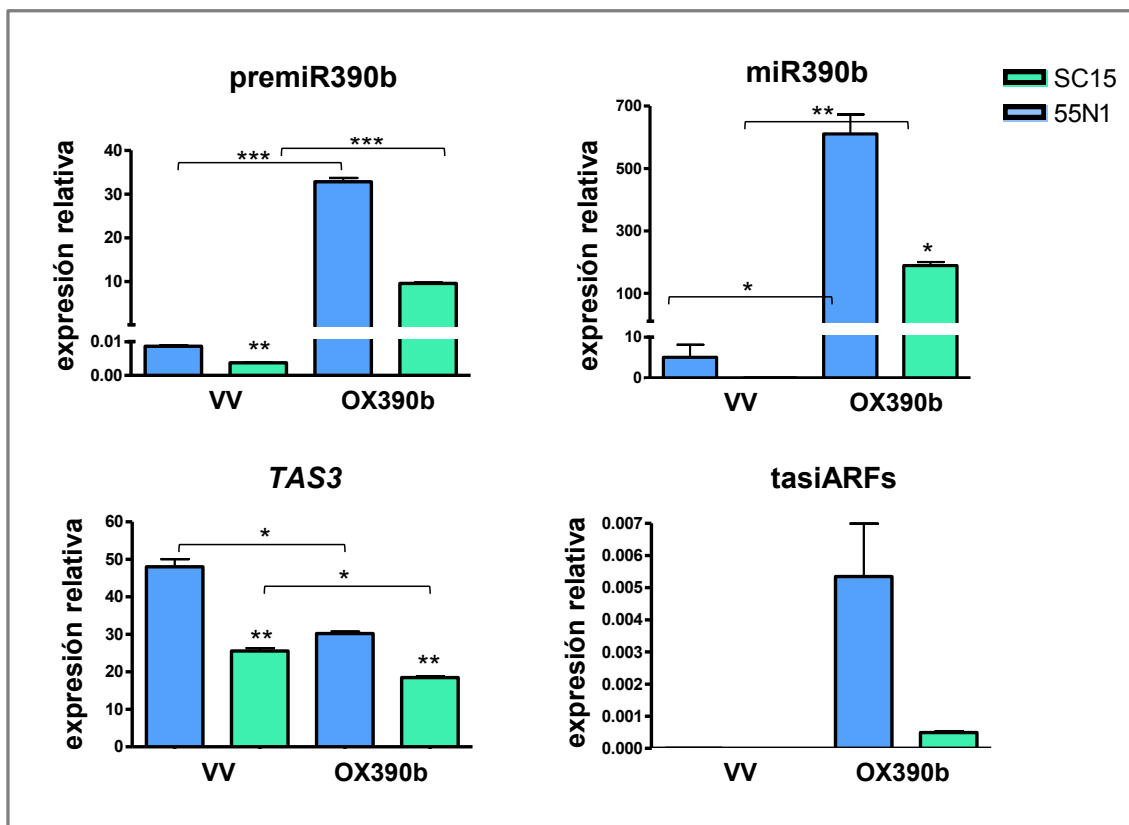
Al finalizar de la cinética de nodulación (a los 21 dpi), se midieron los diámetros de los nódulos de las plantas VV y OX390b en las condiciones ensayadas. Los resultados presentados en la Figura 54 mostraron que los nódulos formados tanto por la bacteria más eficiente (SC15) como la menos eficiente (55N1) fueron significativamente más pequeños en las raíces OX390b que en las raíces VV, indican que la sobreexpresión del premiR390b afecta también el desarrollo de los nódulos formados las cepas de *R. etli*.



**Figura 54: Tamaño de nódulos de 21 dpi del miR390b.** Medición del tamaño de nódulos de 21 dpi formados en las raíces transgénicas de plantas control (VV) y plantas que sobreexpresan el precursor del miR390b (OX390b) con SC15 o 55N1. Las barras de error representan los desvíos estándar de la medición de la totalidad de los nódulos GFP (ver *n*, número de nódulos entrando en el análisis para cada condición). Los asteriscos indican diferencias significativas de las muestras de sobreexpresión con sus controles respectivos en un *t*-test con  $p < 0,0001$ .

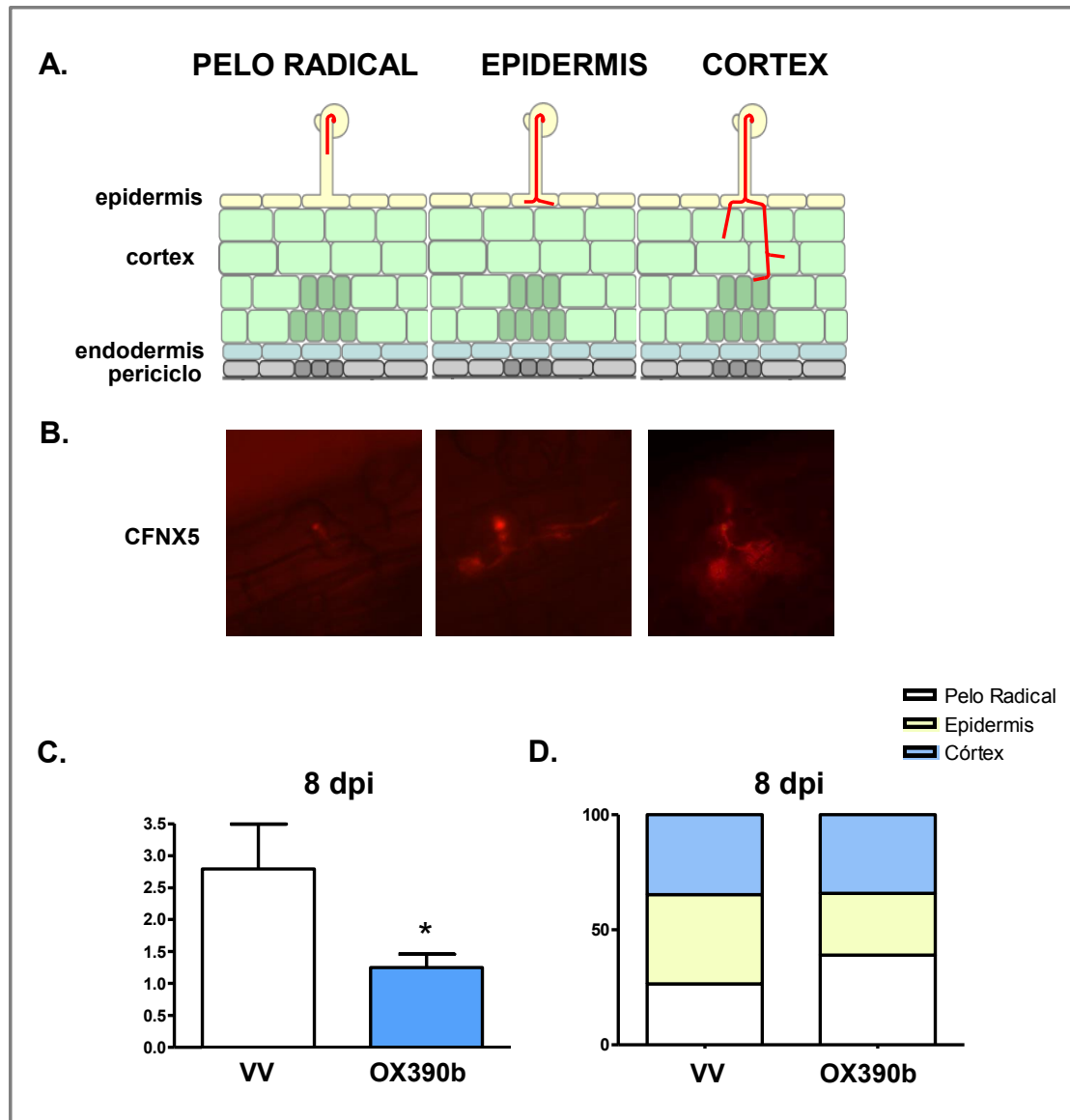
Para validar que los fenotipos observados fueron debidos a los niveles más altos del miR390b, se evaluaron los niveles de expresión en raíces de plantas individuales provenientes de estos experimentos funcionales al finalizar el recuento de los nódulos. Se colectó un *pool* de raíces noduladas de plantas transformadas con el vector vacío (VV) y raíces de plantas llevando la construcción de sobreexpresión (OX390b), cada una inoculada con SC15 o 55N1. Se extrajo el ARN de cada muestra y se llevó a cabo la síntesis de ADNc y qPCR para determinar los niveles del premiR390b, el miR390b maduro, el transcripto *TAS3* y los tasiARFs. Las raíces inoculadas con SC15 mostraron una acumulación del precursor >3000 veces mayor en comparación a las plantas control transformadas con el vector vacío; contra >200 veces para las raíces inoculadas con 55N1. Los niveles de acumulación del miR390b medidos por la técnica de *stem-loop* RT-PCR en aquellas mismas muestras indicaron una correlación significativa directa entre la acumulación del premiR390b y del miARN maduro (Figura 55). En las mismas muestras se evaluaron los valores de expresión de *TAS3*, el blanco del miR390b. Los valores de expresión de *TAS3* son significativamente más bajos en las plantas que acumulan el miR390b en comparación a sus controles respectivos, en las dos condiciones de inoculaciones ensayadas. Finalmente, las

acumulaciones de los tasiARFs, *TAS3+7* y *TAS3+8*, fueron detectables únicamente en las raíces de plantas que presentan altos niveles del miR390b (Figura 55).



**Figura 55: Niveles de expresión de la vía del miR390b en raíces noduladas.** Medición de expresión del premiR390b y *TAS3* por real time y de la abundancia del miR390b y de los tasiARFs por la técnica *Stem-Loop* PCR en muestra de raíces control (VV) y de raíces de sobreexpresión del miR390b (OX390b) al final de la cinética de nodulación (21 dpi). Los asteriscos indican diferencias significativas respecto al control y entre las diferentes muestras señaladas por los corchetes en un *t*-test (con \*,  $p < 0,05$ ; \*\*,  $p < 0,01$ ; \*\*\*,  $p < 0,0001$ ).

Para determinar si la sobreexpresión del premiR390b altera la infección bacteriana, se inocularon raíces transgénicas con una cepa de *R. etli* marcada con una proteína fluorescente. Se estudió la formación de hilos de infección, su desarrollo y su progresión hasta la ocupación de los nódulos formados. A tiempos tempranos se cuantificaron los hilos de infección formados por centímetros de raíz transgénicas en las raíces OX390b y VV. A su vez, para cada hilo de infección observado se determinó su estado de progresión, clasificando los mismos como hilos que se encuentran en el pelo radical, hilos que alcanzaron las células de la epidermis e hilos que alcanzaron el córtex. Los diferentes estados de progresión están representados en el esquema de la Figura 56A e ilustrado en Figura 56B mediante adquisiciones fotográficas de raíces inoculadas con la cepa de *R. etli nodC* tipo  $\alpha$  CFNX5 que expresa la proteína fluorescente IDsRed (Battaglia et al., 2014).

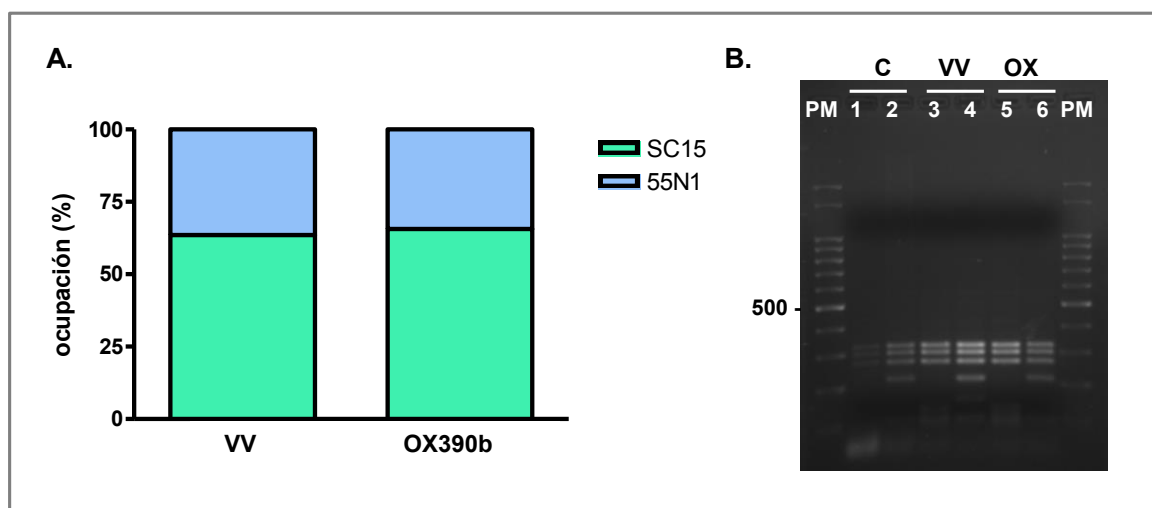


**Figura 56: Estudio de la infección bacteriana.** A. Esquema de los diferentes estados de caracterización de los hilos usados en este estudio: pelo radical, epidermis y córtex. B. Fotos de microscopía invertida ilustrando la progresión de los hilos de infección con la cepa de *R. etli* CFNX5 marcada con IDsRed. C. Cantidad de hilos de infección contabilizados a 8 dpi en raíces transgénicas de sobreexpresión del miR390b (OX390b) y del control (VV), valores expresados en cm de raíz, inoculación con CFNX5 marcada con IDsRed. El asterisco indica diferencias significativas en un *t*-test con  $p < 0,05$ . D. Repartición de la cantidad de hilos de infección según el estadio de progresión: pelo radical, epidermis y córtex (valores expresados en porcentaje). El experimento es representativo de dos réplicas biológicas independientes.

Los resultados del análisis de la infección bacteriana a 8 dpi muestran una disminución significativa en la densidad de los hilos de infección formados por la cepa CFNX5 de *R. etli* (*nodC* tipo  $\alpha$ ) en las raíces de sobreexpresión del miR390b (OX390b) en comparación a la condición control (Figura 56C). Además, el estudio de la progresión del hilo de infección desde el pelo radical hasta las capas corticales reveló un mayor porcentaje de hilos de infección en los pelos radicales en las raíces OX390b en comparación a las raíces control. Concomitantemente, se observó menor proporción

de hilos de infección que avanzaron hasta la epidermis, mientras que no se notaron diferencias para los hilos que alcanzaron el córtex. Estos resultados sugieren que la sobreexpresión del miR390b altera tanto la iniciación como la progresión de los eventos de infección en la simbiosis entre *P. vulgaris* y *R. etli*.

Para determinar si el miR390b desempeña una función en la preferencia de cepa observada en la simbiosis entre variedades mesoamericanas de *P. vulgaris* y cepas de *R. etli* que portan el alelo *nodC* tipo- $\alpha$ , se realizaron ensayos de competencia en plantas compuestas que llevan la construcción p35S-premiR390b y en plantas control (transformadas con el vector vacío). Se inocularon plantas crecidas en potes sin fuente de nitrógeno con una mezcla equicelular de cultivos de SC15 y 55N1 en el mismo estadio de crecimiento. A los 27 dpi, los nódulos formados fueron recolectados individualmente, y se analizó su ocupación, es decir la identidad de la cepa que ocupa cada nódulo (Figura 57A). La identidad de la cepa de *R. etli* se determinó mediante caracterización fenotípica de las cepas y su validación genotípicamente, mediante el análisis del polimorfismo del gen *nodC* (Figura 57B). Se purificó el ADN bacteriano, se amplificó el gen *nodC* usando los oligonucleótidos específicos (listados en la Tabla S12 de la sección Anexos) y se analizó el perfil de restricción de los fragmentos amplificados luego de la digestión enzimática con *Hinfl*. No se observaron de diferencia de ocupación de los nódulos formados entre las plantas que llevan la construcción de sobreexpresión del miR390b de las plantas control (Figura 57A). Estos resultados sugieren que el miR390b si bien actúa como regulador de la nodulación, no participaría de los mecanismos moleculares que determinan la preferencia de cepa observada de esta simbiosis.



**Figura 57: Ensayo de competencia en raíces de sobreexpresión del premiR390b.** A. Análisis de la ocupación de los nódulos de raíces transgénicas co-inoculadas llevando el plásmido control (VV) o llevando la construcción de sobreexpresión (OX390b). El experimento es representativo de dos réplicas biológicas independientes. B. Perfil de restricción del gen *nodC* de las cepas de *R. etli* luego de la digestión con la enzima *Hinfl* (PM, peso molecular en pares de bases). Análisis de las cepas de *R. etli* proviniendo de nódulos de raíces transgénicas co-inoculadas llevando el plásmido control (VV) o llevando la construcción de sobreexpresión (OX). El patrón de restricción está indicado en C (cepas controles) para SC15 (*nodC*  $\alpha$ , números impares) y 55N1 (*nodC*  $\delta$ , números pares).

# Discusión

## Conservación y evolución del miR390 en plantas

Como se ha mencionado en la Introducción, los miARNs participan en la regulación de la expresión génica durante el desarrollo de las hojas y la formación de órganos post-embrionarios en las plantas, como los nódulos y las raíces laterales. La formación de las raíces laterales se inicia en respuesta a una concentración local máxima de auxinas (Dubrovsky et al., 2006; Oldroyd et al., 2011). Un estudio previo llevado a cabo en *A. thaliana* demostró que el miR390, los tasiARFs y los factores ARFs integran una red regulatoria en respuesta a auxina que controla el crecimiento de las raíces laterales (Marin et al., 2010). En base a este trabajo se propuso un modelo en el cual los tasiARFs, producidos por el clivaje de *TAS3* mediado por miR390, actuarían reprimiendo la expresión de *ARF2*, *ARF3* y *ARF4*, promoviendo el crecimiento de las raíces laterales. Dada su función en el desarrollo de hojas y raíces, la familia del miR390, es un miARN muy conservado en plantas (Xia et al., 2017). El análisis de las secuencias de los genes *MIR* de esta familia mostró que los mismos representan un clado basal separado de grupos filogenéticos de otros genes *MIR*. A su vez, el miR390 comparte identidad con otros miARNs como el miR7122 o el miR4376, a partir de lo cual se propuso que los genes *MIR* del miR390 serían los ancestros de otras superfamilia de miARNs (Xia et al., 2013). No solamente el miR390 maduro ha evolucionado con una presión de selección sobre su secuencia, sino también a nivel del precursor en donde se observó que la secuencia debajo de la región del tallo-rulo, crítica para el procesamiento de DCL1, se encuentra conservada en plantas angiospermas. Esta conservación se observa también a nivel funcional a través de la vía regulatoria miR390-*TAS3*-ARFs. Estudios recientes indican que los genes componentes de esta vía se encuentran conservados en miles de especies de plantas, desde las briofitas hasta las angiospermas (Xia et al., 2017). En nuestros datos de secuenciación se encontraron dos isoformas del miR390 maduro, el miR390a y el miR390b, los cuales también se encuentran conservados a nivel de secuencia y mecanismo de acción.

## El miR390 como mediador de la regulación por auxinas en la nodulación

Las auxinas son hormonas con una función clave en la nodulación. El inicio de la formación de nódulos indeterminados requiere una disminución en los niveles de auxinas, para luego dar lugar a una acumulación de auxinas en las células del periciclo y córtex en división del primordio nodular. Una vez formado el nódulo, las auxinas se restringen a la zona meristemática del nódulo y al tejido vascular (Mathesius, 2008). Su concentración es finamente regulada ya que a bajas concentraciones incrementa la nodulación mientras que a concentraciones altas la inhibe. Para eso la vía del miR390 y la producción de tasiARFs tiene un papel clave a través de la regulación de los

niveles de mARNs que codifican los factores de transcripción regulados por auxinas: *ARF2*, *ARF3* y *ARF4* (Montgomery et al., 2008). Los ARFs interactúan específicamente con las proteínas Aux/IAA que son codificadas por los genes de respuesta temprana a auxina. Los ARFs se unen a elementos de respuesta a auxinas (ARE) en los promotores de genes primarios sensibles a auxinas, mediando su transcripción. Las Aux/IAAs se unen a los ARFs, inhibiendo la transcripción de genes mediada por ARE.

En *M. truncatula* se observó que los niveles de tasiARFs decrecen significativamente en respuesta a la infección rizobiana mientras que los niveles de los blancos de los tasiARFs, *ARF2*, *ARF3* y *ARF4*, se incrementan en etapas tempranas de la infección (Reynoso et al., 2013). Los niveles de miR390a/b se reducen drásticamente (~ 80%) en raíces a las 48 horas post inoculación con *S. meliloti* en comparación con las raíces inoculadas con agua. La activación de la vía miR390/TAS3, mediante la sobreexpresión ectópica del miR390 regula negativamente la nodulación y la infección bacteriana de *M. truncatula* (Hobecker et al., 2017). Nuestros resultados funcionales en *P. vulgaris*, una planta con nodulación de tipo determinado mostró el mismo fenotipo que la leguminosa de tipo indeterminado: menor cantidad de nódulos formados, los cuales presentaron un tamaño más pequeño, y menor densidad de hilos de infección (Figura 53, 54 y 56C). Por otra parte, la medición de los niveles estacionarios de los tasiARFs y de los *ARFs* en raíces de *P. vulgaris* de sobreexpresión del miR390b, presentan una aumentación de los tasiARFs correlativa con la disminución significativa de los niveles de expresión de los *ARFs* (Figura 55), como en *M. truncatula* (Hobecker et al., 2017). Aunque la isoforma estudiada en *P. vulgaris* en este trabajo se diferencia en una base con la de *Medicago* (Figura 45A), los resultados de expresión asociado a la historia evolutiva del miR390 y la conservación de su función en plantas (Xia et al., 2017) sugieren que el pvu-miR390b podría tener una redundancia de función con el miR390a/b de *Medicago* a través de la vía reguladora miR390/TAS3.

## Nuevas funciones del miR390 en leguminosas

Al contrario de *M. truncatula*, en *P. vulgaris* los datos de secuenciación mostraron que el miR390b se acumula en respuesta a la inoculación (24 hpi), con mayores niveles con la bacteria más eficiente (Figura 45B). Los resultados de los alineamientos de los precursores (premiR390a y b) entre estas dos especies mostraron que el premiR390b de *P. vulgaris* es el más divergente en su secuencia (Figura 46). Los análisis funcionales en la simbiosis con cepas más y menos eficiente de *R. etli* no indicarían que este podría cumplir un papel en la preferencia de cepa. Aunque el fenotipo de la sobreexpresión observado en la cinética de nodulación sea más marcado con SC15 la bacteria más eficiente (Figura 53), la cantidad de nódulos formados por 55N1, la bacteria menos eficiente se ve también disminuida, aunque en menor proporción. El miR390b conservaría la función de regulador negativo descrito en *M. truncatula*, independientemente de la cepa considerada. Sin embargo, los datos obtenidos a partir de las bibliotecas generadas podrían sugerir que el miR390b tiene un nivel de regulación espacio temporal diferente al de *M. truncatula*, cuya regulación podría ser dada a nivel de la divergencia del precursor y/o ser modulada en respuesta a los



niveles locales de auxinas requeridos para dar inicio a la organogénesis de un nódulo determinado, la cual difiere sustancialmente de la de un nódulo indeterminado.

En *M. truncatula* se caracterizó la expresión temporal de los genes *MIR390a* y *MIR390b*, que dan lugar al mismo miR390 maduro. La actividad del promotor de *MIR390a* fue detectada en las últimas etapas de la interacción en los haces vasculares periféricos, y en las zonas meristemática y de infección distal de los nódulos. Por otro lado, la actividad del promotor de *MIR390b* fue detectada a 6 dpi, cuando los hilos de infección se ramifican y alcanzan el córtex. Su expresión está localizada en el pelo radicular, en las células epidérmicas adyacentes alcanzadas por los hilos de infección y en las células corticales divisorias debajo del sitio de la infección. En las últimas etapas de la interacción simbiótica se encontró actividad del promotor de *MIR390b* en aquellas células de la zona de infección a las que alcanzan los hilos de infección y en las células no infectadas de la zona de fijación. La actividad del promotor *MIR390b* también se vio en meristema del nódulo en etapas tardías. Estos resultados indicaron que el *MIR390b* es un gen sensible a rizobios y su expresión está asociada a células de raíces y nódulos jóvenes que se alcanzan mediante el crecimiento activo de los hilos de infección, mientras que en los nódulos maduros su expresión se restringe al meristema (Hobecker et al., 2017). Estudios de las actividades de los promotores de los genes que dan lugar a los miR390a y miR390b en *P. vulgaris* nos podría informar acerca de la redundancia de función de la isoforma pvu-miR390b o de una subfuncionalización de la misma (Zhang et al., 2006).

En *L. japonicus*, dos mutantes han sido caracterizados genéticamente: *rel1* y *rel3* (por *reduced leaflet*), ambos presentando defectos en los órganos vegetativos y reproductivos (Yan et al., 2010). *REL1* y *REL3* son genes homólogos *SGS3* y *AGO7* de *Arabidopsis*, dos componentes claves de la biogénesis de los tasiARNs (Figura 48). Los mutantes con pérdida de función *rel3* (*ago7*) presentan defectos en la acumulación de los phasiARNs producidos a partir de *TAS3* y los niveles de expresión de sus blancos los *ARFs* son más altos. El fenotipo asociado a la nodulación de las plantas *rel3* presenta menor cantidad de nódulos fijadores y menos eventos de infección y de iniciación de la nodulación que las plantas salvajes (Li et al., 2014). Los autores mostraron que los mutantes *rel3* presentan respuestas alteradas tanto a las auxinas como al etileno. El fenotipo de nodulación *rel3* se ve restaurado por la inhibición de la síntesis de etileno (Li et al., 2014). A través de la alteración de un componente de la biogénesis de los tasiARNs, se mostró la importancia de la regulación mediada por estos tasiARNs y su integración en la señalización hormonal para la coordinación de la infección rizobiana y la iniciación de la nodulación en *L. japonicus*.

Finalmente, en *P. vulgaris* el miR390a fue detectado en bibliotecas provenientes de órganos vegetativos; particularmente en hojas bajo condición de estrés hídrico (sequía) y en plántulas tratadas con ABA (Arenas-Huertero et al., 2009). Teniendo en cuenta el papel negativo del ABA sobre la nodulación (Ferguson and Mathesius, 2014), resultaría interesante medir la acumulación del miR390a a diferentes tiempos de la interacción entre *P. vulgaris* y rizobios para determinar si esta isoforma podría cumplir también un rol de regulador negativo como en otras leguminosas, o si esta función es restringida a la isoforma del miR390b en *P. vulgaris*.

## Capítulo II: Discusión

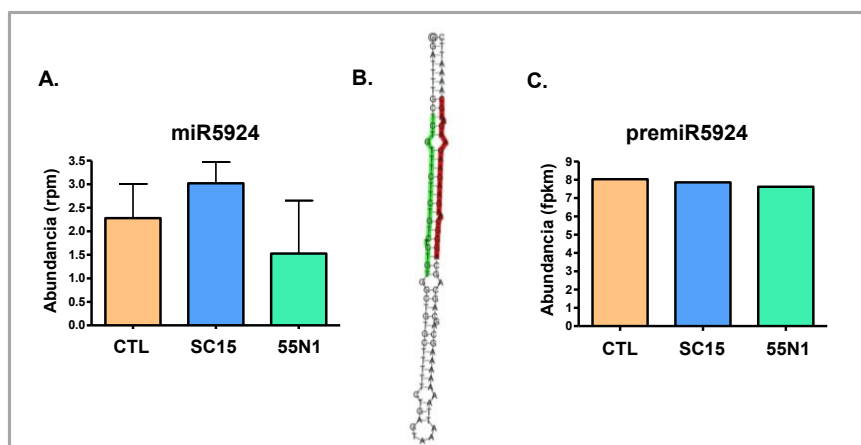
En conjunto, la regulación del desarrollo de las hojas y la formación de órganos post-embriónicos en las plantas, no pasaría únicamente por la modulación de los niveles de auxinas sino también involucraría una red regulatoria más compleja con la participación de otras fitohormonas como el etileno y el ABA.

# *Capítulo III*

## Caracterización de un nuevo miARN involucrado en la preferencia de *P. vulgaris* por cepas más eficientes en la nodulación.

### 1. Caracterización de la expresión del miR5924 maduro y de su precursor

A partir del análisis de los datos de secuenciación presentados en el capítulo I de esta Tesis, se identificaron entre 14 y 21 nuevos miARNs en cada una de las bibliotecas de sARNs (Tabla 7). Teniendo en cuenta las tres condiciones ensayadas, las dos réplicas biológicas y el alineamiento de las secuencias a ambos genomas de *P. vulgaris* se identificaron un total de 64 nuevos miARNs (Figura 28). A partir de esta lista (Tablas S5 y S7 de la sección Anexos) se buscaron los miARNs con perfil de acumulación diferencial entre raíces inoculadas con la cepa más eficiente y competitiva SC15 y aquellas inoculadas con cepa menos eficiente 55N1, o con medio YEM. Uno de estos miARNs (AGGGCAAGAAGAAGAAGC), el cual presentó niveles de acumulación mayores en las raíces inoculadas con la SC15 respecto de aquellas inoculadas con 55N1 (Figura 58A), fue el miARN número 4 presentado en la Tabla S5 de la sección Anexos. A este miARN se lo denominó miR5924. La secuencia de este miR5924 alineó con una región del cromosoma 9 (37237592-37237612) según la anotación del genoma andino, que corresponde al *scaffold00316* (98643-98663) del genoma mesoamericano.



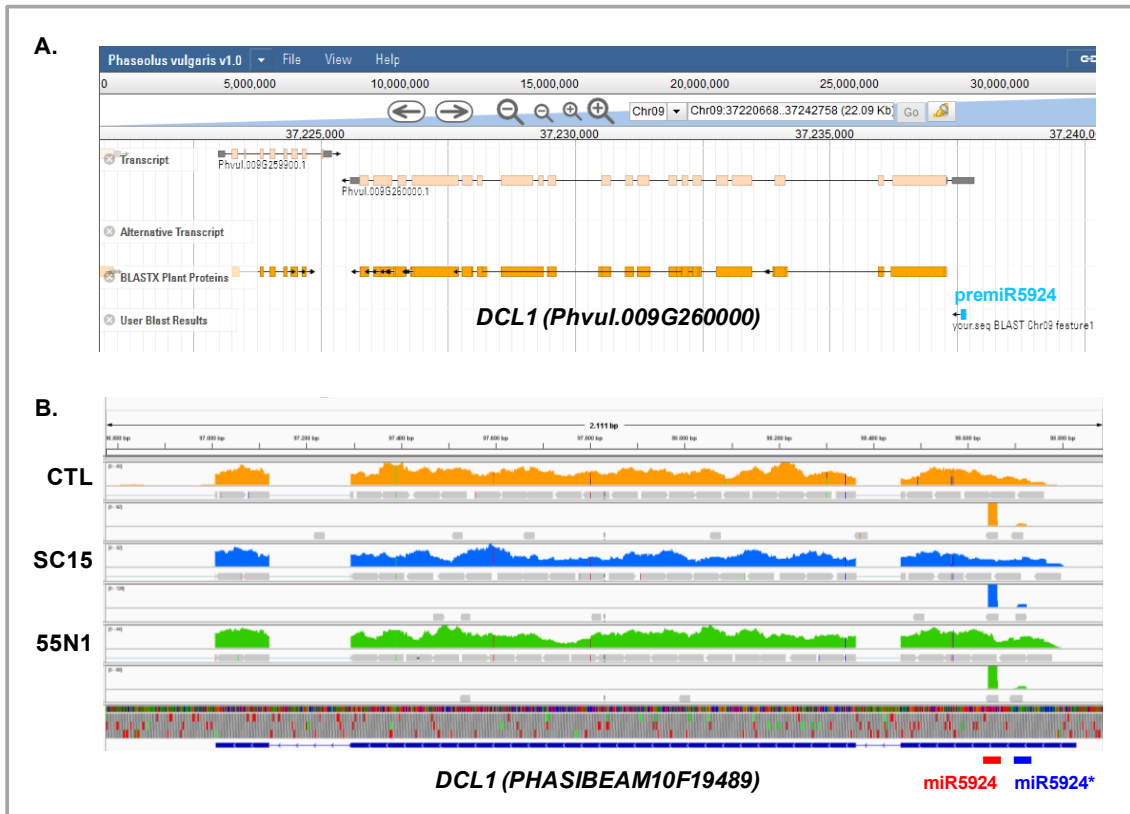
**Figura 58: Niveles de acumulación del miR5924 y estructura secundaria del premiR5924.** A. Valores de expresión normalizados del promedio de las dos replicas biológicas expresados en lecturas por millón de lecturas (rpm, por *read per million*). B. Estructura secundaria del premiR5924 que da lugar al miR5924 de 21 nts (resaltado en rojo) y su miARN\* (resaltado en verde). MFE (*Minimum Free Energy*)= -45,30. C. Niveles de expresión del premiR5924 expresadas en FPKM (*fragments per kilobase per million*) obtenidas a partir de los datos de RNA-seq de Dalla Via et al., 2015.

Durante la ejecución de esta Tesis, la secuencia correspondiente al miR5924 fue reportada en dos estudios en los cuales se caracterizaron la población de los pequeños ARNs de diferentes tejidos de *P. vulgaris* (Formey et al., 2015, 2016). En el primero de ellos, se identificaron tres isoformas (20, 21 y 22 nts) de este miARN que provienen del mismo precursor, las cuales se acumularon a mayores niveles en flores respecto de los otros tejidos analizados (Formey et al., 2015), mientras que el segundo enfocado en la caracterización de sARNs de pelos radicales, se identificó únicamente la isoforma de 21 nts (Formey et al., 2016). Los datos de expresión para cada isoforma obtenidos a partir de los datos reportados en estos dos trabajos se presentan en la Tabla S5 de la sección Anexos. Cabe destacar que este miARN no fue validado ni caracterizado a nivel funcional en ninguno de los estudios mencionados. Por otra parte, en nuestros datos de secuenciación masiva, se detectó principalmente la isoforma de 21 nts del miR5924, sin embargo en la misma región cromosómica se detectaron algunas lecturas correspondientes a una isoforma de 22 nts (CAGGGCAAGAAGAAGAAGC) que provienen del mismo precursor (Figura 58B). Esta isoforma corresponde al miARN número 10 de la Tabla S5 de la sección Anexos, donde se presentan también los valores de expresión de cada isoforma.

La estructura de tallo-rulo predicha para el precursor del miR5924 (premiR5924) por el programa *RNA folding* de Workbench fue coherente con las reglas definidas para miARNs de plantas (Meyers et al., 2008; Taylor et al., 2014; Axtell and Meyers, 2018). Como se puede observar en la Figura 1B la estructura del precursor del miR5924 posee los dos nucleótidos protuberantes del dúplex miARN/miARN\*. Dado que el miARN presenta una acumulación diferencial en las raíces inoculadas con SC15 que en aquellas inoculadas con 55N1 o YEM, se evaluaron los niveles de acumulación la expresión de su precursor, el premiR5924, utilizando los datos de secuenciación masiva de ARN (RNA-seq) generados previamente en el laboratorio a partir de ARN total (Dalla Via et al., 2015). Los valores de acumulación observados para el precursor no variaron en las tres condiciones ensayadas (Figura 58C). Esto sugiere que las variaciones observadas en los niveles del miARN maduro podrían deberse a un procesamiento diferencial del precursor o a una diferente estabilidad del miARN maduro en cada una de las condiciones ensayadas.

La secuencia del premiR5924 fue alineada al genoma de la variedad andina de *P. vulgaris* disponible en Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>), una plataforma *online* con bases de datos genómicas y herramientas para el análisis de secuencias de plantas. Resultó interesante observar que el precursor de este miARN alineó con un 100% de identidad a la región 5' no traducida (5'UTR) del gen *Phvul.009G260000*, el cual codifica un polipéptido anotado como *Dicer like 1 (DCL1)* (Figura 59A). Este gen sería el ortólogo putativo de *DCL1* de *Arabidopsis thaliana*. Usando la plataforma Integrative Genomic Viewer (IGV), se buscó la región cromosómica correspondiente al premiR5924 en el genoma de la variedad mesoamericana de *P. vulgaris*. Se cargaron los datos de RNA-seq de Dalla Via et al., 2015 y los datos de sRNA-seq provenientes de las dos réplicas biológicas generados en este trabajo de Tesis y se ajustaron las escalas de manera que estén normalizadas

entre sí. Se observó que las secuencias del miR5924 maduro y el miR5924\* también alinearon al 5' UTR del gen anotado como *DCL1* (*PHASIBEAM10F19489*) del genoma Mesoamericano (Figura 59B)



**Figura 59: Alineamientos de los precursores de los nuevos miARNs a los genomas andino y mesoamericano de *P. vulgaris*.** A. Secuencia del premiR5924 identificado por miRCat alineada al 5' UTR del gen *Phvul.009G260000*, anotado como *DCL1*, utilizando la plataforma Phytozome. En el modelo génico propuesto las líneas corresponden a intrones y las cajas a exones. Las cajas naranjas corresponden a regiones codificantes y las grises a las 5' y 3' UTRs. La región correspondiente al premiR5924 se muestra en celeste. B. Secuencia del premiR5924 alineada al transcripto *PHASIBEAM10F19489* del genoma mesoamericano de *P. vulgaris*. Se presentan las lecturas obtenidas a partir de los datos de secuenciación de RNA-seq de Dalla Vía (panel superior) y sRNA-seq (panel inferior) para cada condición (CTL, SC15, 55N1). El modelo génico se indica abajo en azul, el miR5924 en rojo y su miARN\* en azul.

## 2. Identificación *in silico* de transcriptos blancos putativos del miR5924

La predicción de los blancos o “*target*” putativos del miR5924 se realizó utilizando el programa psRNATarget (versión 2011) (Dai and Zhao, 2011) y la base de datos de transcriptos de *P. vulgaris* (JGI genomic Project). El algoritmo psRNATarget reporta los valores de máxima probabilidad (*E*, para *Maximum expectation* según Zhang, 2005) y

la accesibilidad del mensajero al miARN (*UPE* de *Target accessibility*), definido en Kertesz et al., 2007, según el programa RNAup de Mückstein et al.(2006). En este algoritmo, valores más bajos de *E* y más altos de *UPE* corresponden a una mayor probabilidad de ser blanco de un miARN. La búsqueda de los blancos putativos del miR5924 utilizando este algoritmo arrojó una extensa lista de transcritos. En la Tabla 11 se presentan sólo los primeros 7 transcritos blancos más probables ordenados en base a la probabilidad de ser target de la isoforma de 21 nts del miR5924. Para todos los blancos putativos identificados, las probabilidades de los blancos respectivos fueron más altas para el miR5924 de 21 nts, en comparación a la isoforma de 22 nts, a excepción del transcripto *Phvul.003G154300.1*.

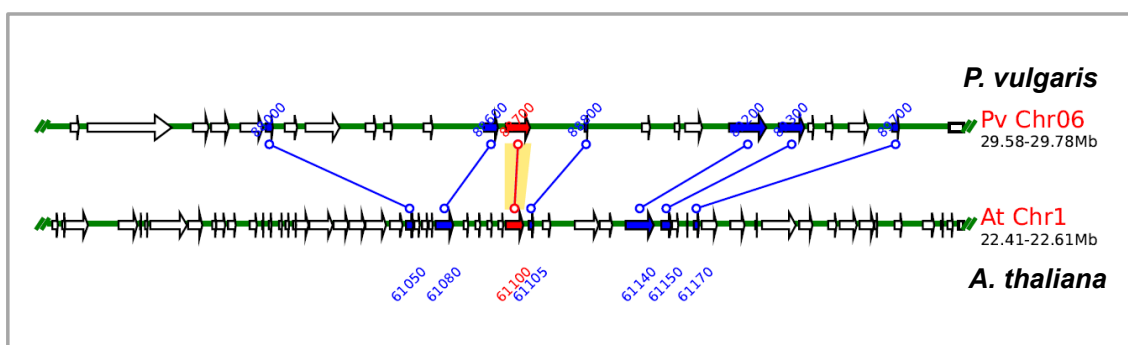
**Tabla 11: Blancos putativos identificados con psRNATarget.** Transcritos con probabilidad más alta identificados en el genoma de *P. vulgaris* mediante psRNATarget para el miR5924 según la isoforma considerada (21 o 22 nts). El primer número corresponde a la posición en el orden de probabilidad, mientras los valores (*E*) y (*UPE*) indicados corresponden a la probabilidad máxima y la accesibilidad del mensajero al miARN respectivamente. Los datos de anotación funcional provienen de Phytozome v.12.

Transcripto	Probabilidad	21 nts	22 nts	Anotación
<i>Phvul.006G188700.1</i>	Posición	1	3	PTHR31008:SF6 - CSL1, disease resistance protein (TIR class), putative
	<i>E</i>	1	2	
	<i>UPE</i>	4.944	4.929	
<i>Phvul.008G056400.1</i>	Posición	2	4	K15747 - beta-ring hydroxylase (LUT5, CYP97A3)
	<i>E</i>	1	2	
	<i>UPE</i>	4.839	6.284	
<i>Phvul.010G041600.1</i>	Posición	3	5	PTHR34954:SF2 - EXPRESSED PROTEIN
	<i>E</i>	1	2	
	<i>UPE</i>	3.374	3.374	
<i>Phvul.011G022000.2</i>	Posición	4	13	K03859 - phosphatidylinositol glycan, class C (PIGC, GPI2)
	<i>E</i>	1	2	
	<i>UPE</i>	3.203	3.268	
<i>Phvul.003G154300.1</i>	Posición	5	1	PTHR31807:SF2 - HAUS AUGMIN-LIKE COMPLEX SUBUNIT 8
	<i>E</i>	1.5	0.5	
	<i>UPE</i>	10.1	9.624	
<i>Phvul.006G121000.1</i>	Posición	6	20	PTHR15239:SF6 - NUCLEAR EXPORT MEDIATOR FACTOR NEMF
	<i>E</i>	1.5	2.5	
	<i>UPE</i>	2.454	9.48	
<i>Phvul.002G044500.1</i>	Posición	7	26	PF01535//PF13041//PF14432 - PPR repeat (PPR) // PPR repeat family (PPR_2) // DYW family of nucleic acid deaminases (DYW_deaminase)
	<i>E</i>	1.5	2.5	
	<i>UPE</i>	6.948	7.115	

Los datos de anotación indicados fueron obtenidos a través de la plataforma Phytozome versión 12 usando la base de datos disponibles de *P. vulgaris* (v2.1, variedad andina). El blanco más probable del miR5924 fue el transcripto

*Phvul.006G188700.1*, el cual codificaría una proteína de resistencia a enfermedades ocasionadas por patógenos, de tipo TIR (*Toll-interleukine 1 Receptor*) de *M. truncatula* (*Medtr2g012990.1*, Phytozome). Según los datos de expresión del atlas de RNA-seq de *P. vulgaris* reportados por O'Rourke et al., 2014 y disponibles públicamente en la plataforma Phytozome, este transcripto provendría de un *locus* altamente expresado en tejido vegetativos incluyendo hojas, tallo, raíces y nódulos, y muy bajos niveles en tejido reproductivo como flores, botones florales y vainas jóvenes. El segundo blanco putativo fue un transcripto (*Phvul.008G056400.1*) que codificaría una  $\beta$ -ring hidrolasa, el cual presenta una alta expresión en hojas y baja en tallos, raíces y nódulos. Otro blanco putativo identificado en nuestra búsqueda con psRNATarget, aunque con menor probabilidad, fue el transcripto *Phvul.002G044500.1*, el cual codifica una proteína con repeticiones de pentatricopéptido (PRR) asociada a las respuestas de defensa (Cao et al., 2017). Este transcripto se acumula a mayores niveles en nódulos que en raíces, y en parte aérea presenta mayor acumulación en tejido reproductivo que en vegetativo.

Para corroborar la anotación del blanco putativo más probable, *Phvul.006G188700.1*, se analizó la sintenia del *locus* que da lugar al transcripto, utilizando la herramienta *Locus search* de la plataforma *Plant Genome Duplication Database* (PGDD), (<http://chibba.agtec.uga.edu/duplication/index/locus>). Resultó interesante encontrar el *locus Phvul.006G188700* del cromosoma 6 de *P. vulgaris* en sintenia con el gen *AT1G61100* del cromosoma 1 de *A. thaliana* (Figura 60), el cual codifica una proteína tipo TIR de la membrana plásmica implicada en la defensa frente a bacterias y hongos (<http://www.arabidopsis.org/servlets/Tair>) reforzando así la anotación de este transcripto de *P. vulgaris*.

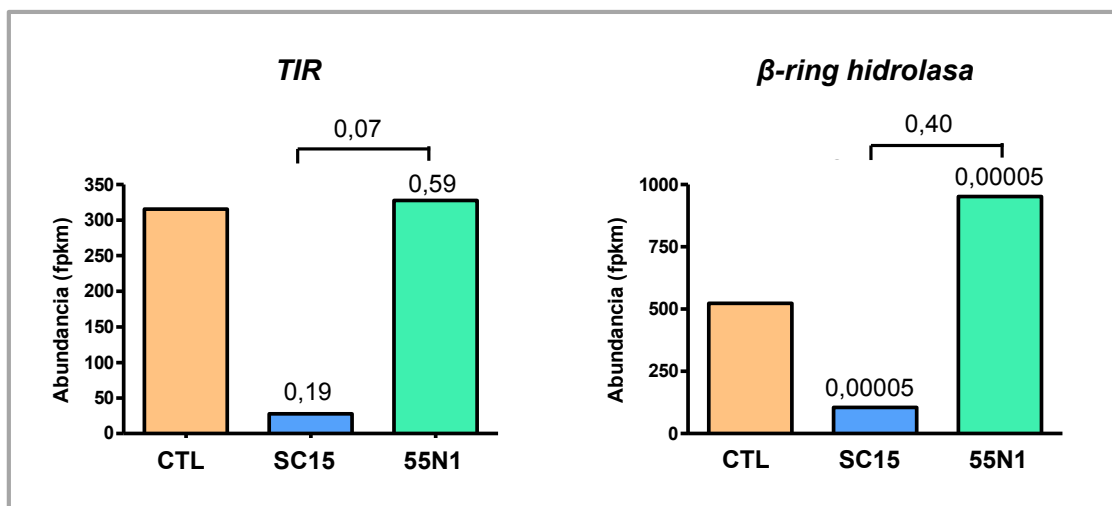


**Figura 60: Estudio de sintenia del *Phvul.006G188700.1*.** *Locus* en sintenia entre *P. vulgaris* y *A. thaliana*. En rojo el gen correspondiendo al transcripto *Phvul.006G188700* en sintenia con el *AT1G61100* de *A. thaliana* (*Score*= 700; *E*= 3e-80) usando *Plant Genome Duplication Database*. En azul están marcados otros genes conservados.

Posteriormente, se analizaron los niveles de expresión de los dos transcriptos blancos más probables utilizando los datos de RNA-seq de raíces a 24 hpi generados por nuestro grupo de investigación: el transcripto *Phvul.006G188700.1* que codifica una proteína TIR y el *Phvul.008G056400.1*, el cual codifica una  $\beta$ -ring hidrolasa. Los valores de expresión obtenidos a partir de los datos de RNA-seq muestran una



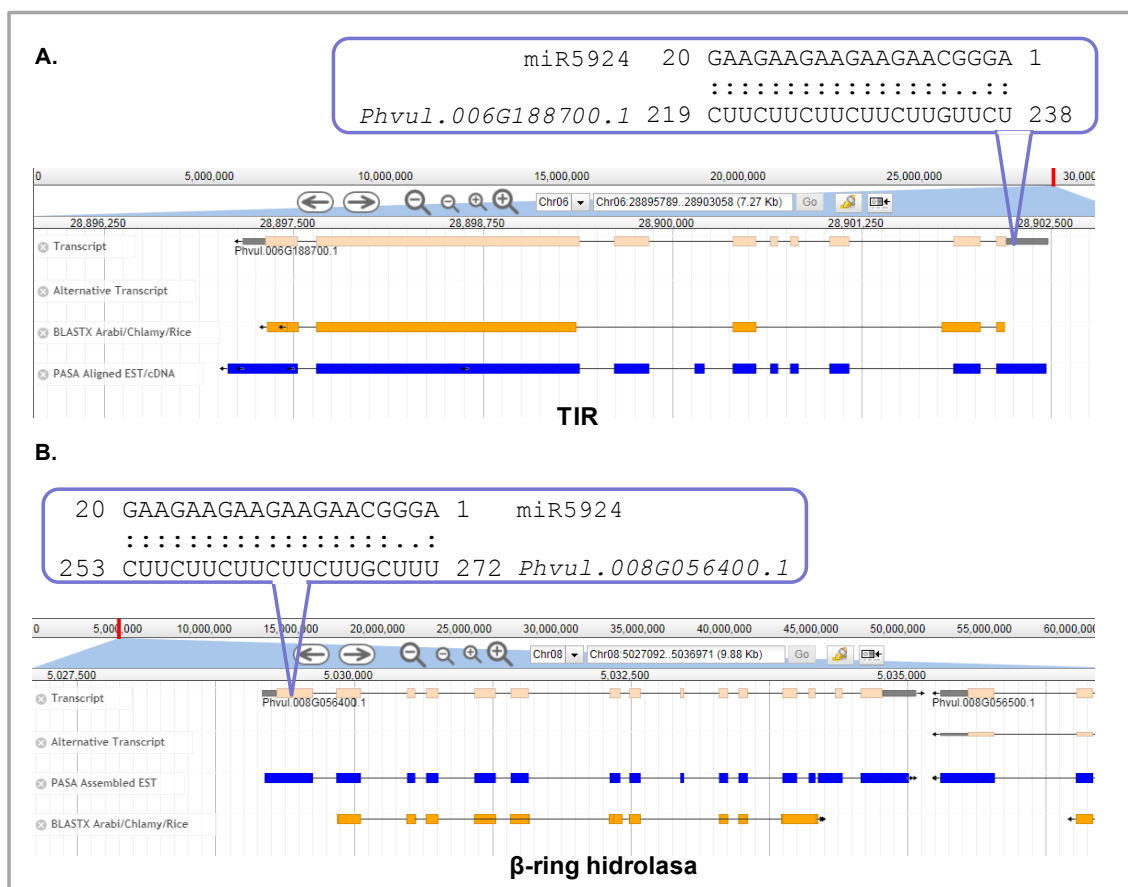
correlación inversa con los niveles de acumulación del miR5924 maduro observados en las mismas condiciones ensayadas (Figura 61). Ambos transcritos se reducen en respuesta a la cepa SC15 y se mantienen altos o aumentan en respuesta a 55N1, presentando diferencias significativas con el control (inoculado con YEM) sólo en el caso del transcripto que codifica la  $\beta$ -ring hidrolasa.



**Figura 61: Niveles de expresión de los blancos putativos del miR5924.** Valores de expresión normalizadas de los transcritos *Phvul.006G188700.1* (*TIR*) y *Phvul.008G056400.1* ( $\beta$ -ring hidrolasa) en FPKM obtenidas a partir de los datos de RNA-seq de tres réplicas biológicas disponibles en el laboratorio (Tesis doctoral de Joaquín Clúa). Los valores de probabilidad de expresión diferencial (*p-value*) de las condiciones de inoculación contra el control (SC15 vs CTL o 55N1 vs CTL) se indican arriba de las barras, y de SC15 vs 55N1 se señala arriba del corchete.

Posteriormente, se identificaron las zonas de apareamiento del miR5924 a los dos blancos más probables: *TIR* (*Phvul.006G188700.1*) y  $\beta$ -ring hidrolasa, (*Phvul.008G056400.1*), utilizando los datos de predicción de blancos de psRNATarget. Las secuencias de 21 nts alineadas a los transcritos blancos ubicados en el genoma de *P. vulgaris* se presentan en la Figura 62. El miR5924 se uniría a una región del 5' UTR del transcripto *TIR* (Figura 562), mientras este mismo miARN se uniría a la región codificante dentro del primer exón del transcripto  $\beta$ -ring hidrolasa (Figura 62B).

Los resultados presentados en esta sección en base al análisis *in silico* y a los datos de expresión indican que tanto el transcripto que codifica el receptor de tipo *TIR* como el que codifica la  $\beta$ -ring hidrolasa podrían ser blancos de acción del miR5924. Futuros experimentos de 5'RACE y/o degradoma contribuirán a validar estos blancos putativos.

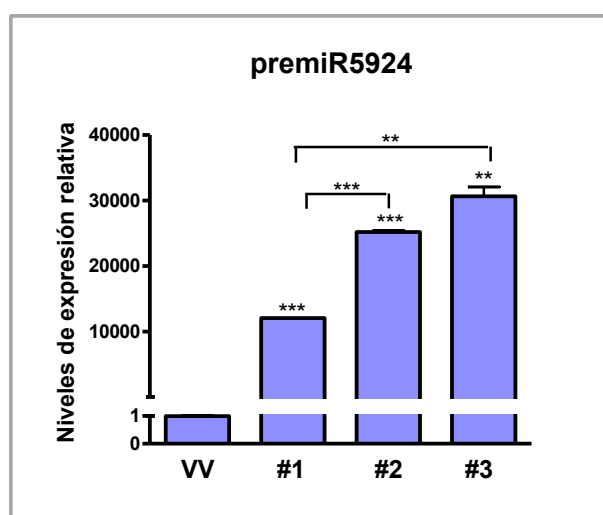


**Figura 62: Ubicación de los blancos del miR5924.** A. Alineamiento del miR5924 al 5'UTR del transcripto *Phvul.006G188700.1* (TIR) en el genoma de *P. vulgaris* en Phytozome. B. Alineamiento del miR5924 a la secuencia codificante del transcripto *Phvul.008G056400.1* ( $\beta$ -ring hidrolasa).

### 3. Caracterización funcional del miR5924

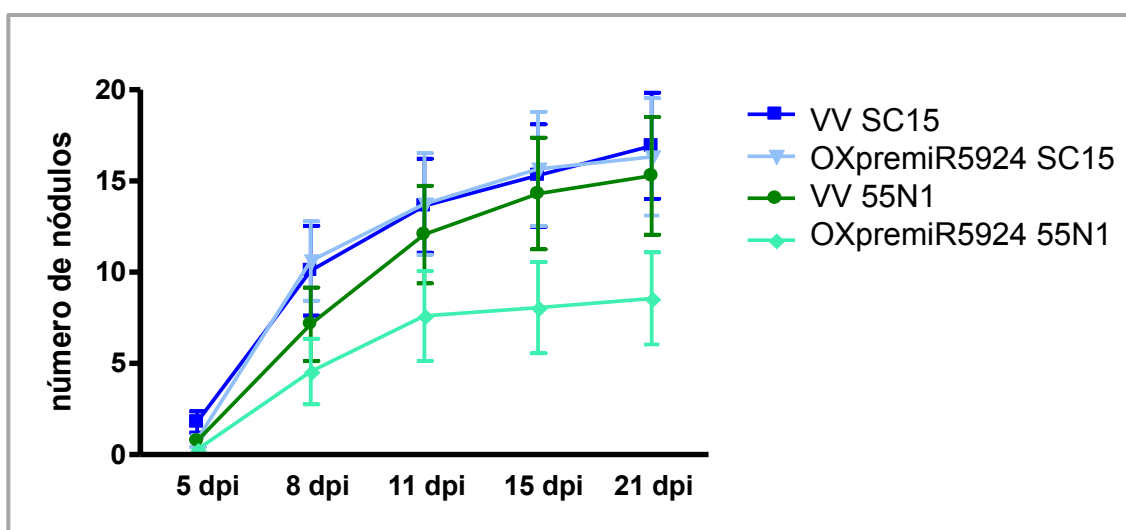
Los datos de expresión del miR5924 sugieren que el mismo podría desempeñar una función importante en la eficiencia de la nodulación y la preferencia de cepa observada en las variedades mesoamericanas de *P. vulgaris*. En base a ello, se decidió caracterizar mediante genética reversa la función de dicho miARN en la formación de nódulos, la infección rizobiana y la ocupación de los nódulos por cepas más eficientes. Para ello, se llevó a cabo la sobreexpresión del precursor del miR5924 en raíces de *P. vulgaris*. La secuencia correspondiente al precursor fue amplificada a partir de ADNc utilizando dos oligonucleótidos específicos que amplifican un fragmento de 192 nts (Figura 71 de la Sección Materiales y Métodos). Los fragmentos obtenidos fueron clonados en el vector de expresión pk7WG2D,1 (Figura 72 de la Sección Materiales y Métodos), el cual dirige la expresión del prem5924 bajo el control del promotor del ARN 35S del virus del mosaico del coliflor (CaMV35S). Esta construcción se denominó p35S:premiR5924. El vector de expresión utilizado posee además el gen reportero que

codifica la proteína verde fluorescente GFP bajo el control del promotor constitutivo *rolD* de *Agrobacterium tumefaciens*, lo que permite visualizar el tejido transformado bajo luz UV (Figura 69 de la Sección Materiales y Métodos). La construcción para la sobreexpresión del premiARN fue introducida en la cepa de *Agrobacterium rhizogenes* K599, y la cepa obtenida fue utilizada para generar raíces transgénicas en plantas compuestas de *P. vulgaris*. Un grupo de plantas fue transformado con el vector vacío como control. Una vez que emergieron las raíces transgénicas, las plantas fueron crecidas durante una semana en cajas conteniendo un medio Fahraeus suplementado con nitrógeno y luego visualizadas bajo lupa de fluorescencia para identificar aquellas raíces que expresan la GFP. Aproximadamente el 70 % de las raíces presentaron expresión de la proteína GFP y aquellas raíces que no presentaban fluorescencia fueron escindidas. Para evaluar los niveles de acumulación del precursor del miR5924, se extrajeron los ARNs totales a partir de tejido de raíz de tres plantas individuales que llevan la construcción p35S:premiR5924 (OXpremiR5924) y una transformada con el vector vacío (VV) sin inocular y se llevó a cabo la síntesis de ADNc mediante transcripción reversa (RT). A partir de estas muestras de ADNc se realizaron reacciones de qPCR utilizando oligonucleótidos específicos para el premiR5924 (Figura 63) y para el transcripto que codifica el factor de elongación de la traducción en eucariontes *EF1 $\alpha$*  como gen de referencia. Las 3 plantas evaluadas (#1, #2, #3) mostraron una acumulación del precursor >10000 veces en comparación a las plantas control transformadas con el vector vacío (VV).



**Figura 63: Expresión de los precursores en raíces transgénicas medida por RT-qPCR.** Medición de la expresión del premiR5924 de raíces transgénicas llevando un plásmido control (VV), y raíces transgénicas provenientes de 3 plantas individuales que portan la construcción p35S:premiR5924 (#1, #2, #3). Los ADNc fueron obtenidos a partir de tejido de raíces sin inocular. Los valores de expresión fueron normalizados por el gen de referencia *EF1 $\alpha$*  y expresados en relación al valor del CTL al cual se le otorgó un valor de 1. Los asteriscos indican diferencias significativas respecto al control en un *t*-test (\*\*,  $p < 0,01$ ; \*\*\*,  $p < 0,0001$ ).

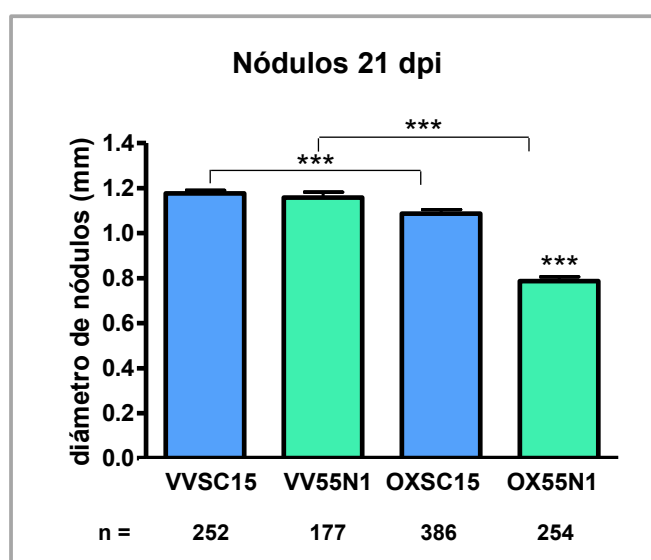
Habiendo verificado la sobreacumulación del precursor del miR5924 se abordó la caracterización fenotípica de dichas plantas, la cual incluyó la cuantificación del número de nódulos por raíz en función del tiempo (cinética de nodulación) y la determinación del tamaño de los nódulos formados luego de la inoculación con las cepas de *R. etli* SC15 y 55N1. Las plantas compuestas de variedad mesoamericana NAG12 que llevan la construcción p35S:premiR5924 (OXpremiR5924) o el vector vacío (VV) fueron transferidas a cajas conteniendo medio Fahreaus sin una fuente de nitrógeno. Un mínimo de 10 plantas por construcción fueron inoculadas con la cepa más eficiente SC15, o con la menos eficiente 55N1. Los nódulos formados fueron cuantificados a partir de los 5 días post inoculación (dpi). Los resultados de la cinética de nodulación no mostraron diferencias entre las raíces VV y la OXpremiR5924 cuando las mismas fueron inoculadas con SC15. Por el contrario, cuando se utilizó la cepa 55N1, las plantas OXpremiR5924 desarrollaron una menor cantidad de nódulos que las VV (Figura 64). Esta disminución en el número de nódulos fue más significativa a tiempos más largos luego de la inoculación (i.e 15 y 21 dpi). Estos resultados indican que la sobreexpresión de premiR5924 afectaría negativamente la formación de nódulos en la interacción de *P. vulgaris* con la cepa 55N1.



**Figura 64: Cinética de nodulación de plantas OXpremiR5924.** Número de nódulos formados por raíz transgénica en plantas control (VV) y plantas que sobreexpresan el precursor del miR5924 (OXpremiR5924) luego de la inoculación con las cepas de *R. etli* SC15 (*nodC*  $\alpha$ ) y 55N1 (*nodC*  $\delta$ ). Las barras de error representan los desvíos estándar del número de nódulos de la totalidad de las raíces transgénicas provenientes de un mínimo de 10 plantas por condición (aprox. 50-60 raíces transgénicas). El experimento es representativo de dos réplicas biológicas independientes.

Al finalizar la cinética de nodulación (a los 21 dpi), se midieron los diámetros de los nódulos de las plantas VV y OXpremiR5924 en las condiciones ensayadas. Los nódulos formados por la bacteria más eficiente (SC15) fueron levemente más pequeños en las raíces OXpremiR5924 que en las raíces VV, mientras que los nódulos formados por 55N1 en las plantas OXpremiR5924 mostraron una reducción

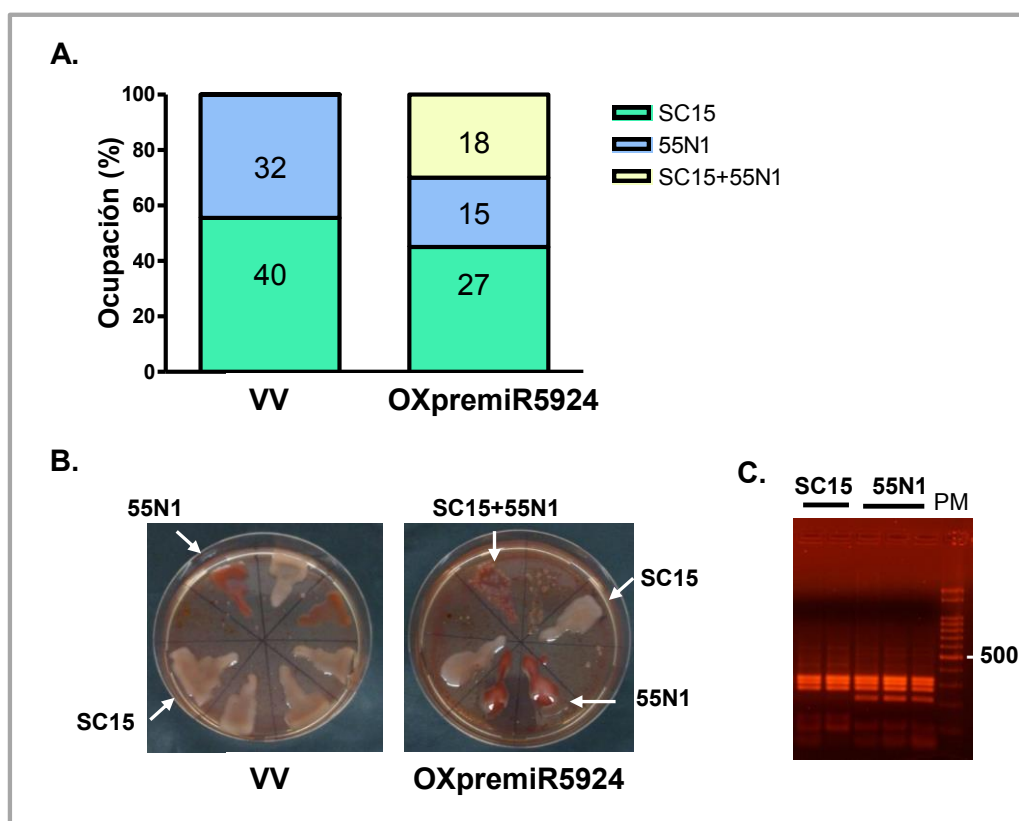
de aproximadamente un 35 % en comparación a los nódulos de las raíces VV (Figura 65). Estos resultados, los cuales son representativos de otro experimento independiente, indican que la sobreexpresión del premiR5924 afecta no sólo el número de nódulos formados por la cepa 55N1 sino también el desarrollo de los mismos.



**Figura 65: Tamaño de nódulos de 21 dpi del premiR5924.** Medición del tamaño de nódulos de 21 dpi formados en las raíces transgénicas de plantas control (VV) y plantas que sobreexpresan el precursor del miR5924 (OX) con SC15 o 55N1. Las barras de error representan los desvíos estándar de la medición de la totalidad de los nódulos GFP (con  $n$ , número de nódulos entrando en el análisis para cada condición). Los asteriscos indican diferencias significativas respecto al control en un  $t$ -test con  $p < 0,05$  (\*\*\*,  $p < 0,0001$ ).

Con el objetivo de investigar si el miR5924 desempeña una función en la preferencia de cepa observada en la simbiosis entre *P. vulgaris* y *R. etli* se realizaron ensayos de competencia en plantas compuestas que llevan la construcción p35S-premiR5924 y en plantas control (transformadas con el vector vacío). Se inocularon plantas crecidas en potes sin fuente de nitrógeno con una mezcla equicelular de cultivos de SC15 y 55N1 en el mismo estadio de crecimiento. Luego de 27 dpi, los nódulos formados fueron recolectados individualmente, y se analizó su ocupación, es decir la identidad de la cepa que ocupa cada nódulo (Figura 66A). La identidad de la cepa de *R. etli* se determinó mediante caracterización fenotípica y genotípica de las cepas. La cepa SC15 (*nodC*  $\alpha$ ) posee un color rosa mientras que la cepa 55N1 (*nodC*  $\delta$ ) exhibe un color rojo y un crecimiento mucoso debido a la alta cantidad de exopolisacáridos (Figura 66B). La validación genotípicamente de las cepas se realizó analizando el polimorfismo del gen *nodC* (Figura 66C). Se purificó el ADN bacteriano, se amplificó el gen *nodC* usando los oligonucleótidos específicos (Tabla S12) y se analizó el perfil de restricción de los fragmentos amplificados luego de la digestión enzimática con *Hinfl*. En este experimento, el 60 % de los nódulos transgénicos formados en las raíces control (transformadas con el vector vacío) fueron ocupados por la cepa SC15 (Figura

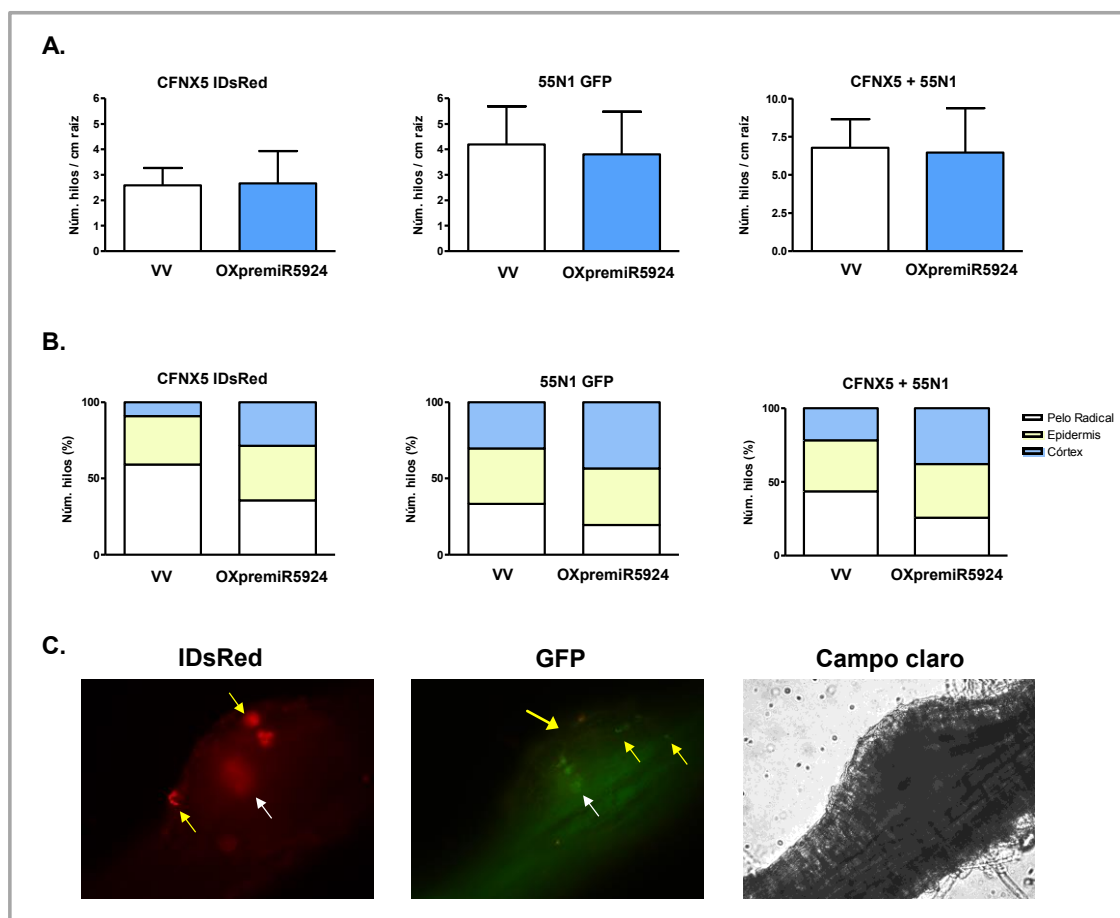
66A). En el caso de las raíces que llevan la construcción de sobreexpresión del precursor del miR5924, el 45 % de los nódulos formados fueron ocupados por la cepa SC15 y un 30 % ocupado por 55N1. Notoriamente, el 25 % de los nódulos de las plantas OXpremiR5924 presentaron co-ocupación, es decir que los nódulos fueron ocupados por ambas cepas bacterianas (Figura 66B). Esta co-ocupación no fue observada en ninguno de los nódulos provenientes de las plantas control. Estos resultados muestran que la sobreexpresión de premiR5924 es capaz de alterar la ocupación de los nódulos, resultando en un aumento en la proporción de nódulos co-ocupados.



**Figura 66: Ensayo de competencia en raíces de sobreexpresión del premiR5924.** A. Análisis de la ocupación de los nódulos formados en raíces transgénicas llevando el plásmido control (VV) o llevando la construcción de sobreexpresión del premiR5924 (OXpremiR5924) co-inoculadas con una mezcla equicelular de las cepas SC15 y 55N1. Los valores del eje y del grafico están expresados en porcentajes y los números corresponden a la cantidad de nódulos analizados. B. Fenotipos del crecimiento de las cepas de *R. etli* en placas de Petri conteniendo medio YEM-rojo Congo. Las cepas de *R. etli* SC15 (*nodC*  $\alpha$ ) y 55N1 (*nodC*  $\delta$ ) se diferencian fenotípicamente por el color (rosa para SC15 y rojo para 55N1). C. Las cepas de *R. etli* presentan un perfil de restricción diferente del gen *nodC* luego de la digestión con la enzima *HinfI* (PM, peso molecular en pares de bases).

Para determinar si la sobreexpresión del premiR5924 altera la infección bacteriana, se inocularon raíces transgénicas con cepas de *R. etli* marcadas con proteínas

fluorescentes. Como es explicado en el capítulo anterior, se estudió la formación de hilos de infección, su desarrollo y su progresión, así como la ocupación de los nódulos formados. En nuestros ensayos de infección, las raíces OXpremiR5924 y control fueron co-inoculadas con la cepa CFNX5 (*nodC*  $\alpha$ ) marcada con la proteína fluorescente DsRed, y con la cepa 55N1 (*nodC*  $\delta$ ) marcada con la proteína fluorescente GFP (Figura 67).



**Figura 67: Análisis de la infección bacteriana en raíces de sobreexpresión del premiR5924.** A. Densidad de hilos de infección observados en raíces transgénicas de sobreexpresión del premiR5924 (OXpremiR5924) y del control (VV) (valores expresados en cm de raíz) a los 14 dpi con CFNX5 marcada con DsRed, con 55N1 marcada con GFP, y los hilos de infección totales (CFNX5 + 55N1). B. Distribución de los hilos de infección según el estado de progresión: pelo radical, epidermis y córtex (valores expresados en porcentaje). C. Visualización de nódulos co-ocupados provenientes de raíces OXpremiR5924, a cabo de 11 dpi con CFNX5 (IDsRed) y 55N1 (GFP). Las flechas amarillas indican bacterias de hilos de infección y las blancas a dentro del nódulo.

Los resultados del análisis de la infección a 14 dpi no muestran diferencias en la densidad de los hilos de infección formados en las raíces de sobreexpresión del

premiR5924 en comparación a la condición control para ninguna de las dos cepas de *R. etli* (Figura 67A). Sin embargo, el estudio de la progresión del hilo de infección desde el pelo radical hasta las capas corticales reveló un mayor porcentaje de hilos de infección que alcanzan el córtex en las raíces OXpremiR5924 en comparación a las raíces control. Si bien esta observación se verificó con ambas cepas ensayadas, el fenotipo fue más marcado con 55N1, donde el porcentaje de los hilos que alcanzaron el córtex fue del 40% en las raíces OXpremiR5924 comparado con un 25 % en las raíces control (Figura 67B). Concomitantemente, se observó una menor proporción de hilos de infección en el estadio de pelo radical, mientras que no se notaron diferencias para los hilos que avanzaron hasta la epidermis. Finalmente en este ensayo de co-inoculación se analizó la ocupación de los nódulos formados. Se observaron casos de ocupación de nódulos con ambas bacterias en nódulos que llevan la construcción p35S:premiR5924, lo cual no fue observado en la condición control (Figura 67C). Además se observó la presencia de hilos de infección formados independientemente por las dos bacterias en un mismo nódulo. Este resultado es coincidente con lo presentado en la Figura 66A, donde se analizó la identidad de las cepas que ocuparon los nódulos mediante análisis fenotípico de las mismas y del polimorfismo del gen *nodC*. Ambos resultados sugieren que la desregulación de los niveles del premiR5924 altera la capacidad de las plantas de origen mesoamericano de discriminar y seleccionar aquellas cepas que son más eficientes en la formación de nódulos.



# Discusión

## **El miR5924 un nuevo miARN de *P. vulgaris* procesado a partir de *DCL1***

Los miARNs nuevos o no conservados, pueden ser activos aun a bajos niveles de expresión, o también pueden ser finamente regulados en ciertas condiciones ambientales o bien ser expresados específicamente en un tejido o tipo celular en particular (Rajagopalan et al., 2006). Esta característica asociada a los miARNs no conservados fue observada tanto para el miR5924 como para otros miARNs nuevos identificados en el Capítulo I de este trabajo de Tesis, los cuales presentaron niveles de acumulación muy bajos en comparación a los miARNs conservados cuando se analizan las raíces completas o se acumularon sólo en una de las tres condiciones analizadas (Tabla 8). El mi5924 presentó niveles de acumulación relativa que estuvieron uno, dos o incluso tres órdenes de magnitud por debajo de los miARN conservados como el miR390, miR156 o miR159, respectivamente (Tablas S4 y S5). El miARN\* identificado por el programa para este nuevo miR5924 fue secuenciado una sola vez en nuestras bibliotecas (abundancia = 1 lectura). Este resultado no es sorprendente ya que el mismo miR5924 también presenta baja abundancia en las condiciones experimentales ensayadas. A su vez, los miARNs\* suelen ser pocos estables y rápidamente degradados, por lo que se han reportado ejemplos en los cuales miARNs que pertenecen a familias conservadas con abundancias elevadas no presentaban secuencia de sus miARNs\* respectivos en los datos de secuenciación. Como es el caso del miR397 en bibliotecas de *A. thaliana* (Rajagopalan et al., 2006). Por otra parte, el miR5924 fue también identificado en el trabajo reportado por Formey et al. (2016), en donde se secuenció y analizó una biblioteca de pequeños ARNs de pelos radicales luego del tratamiento con el factor Nod purificado de la cepa CE3 (*nodC*  $\alpha$ ) de *R. etli*. Este miARN, designado por Formey et al. (2016) como miRNNov250, no presentó una acumulación diferencial en respuesta al tratamiento con el factor Nod. En este trabajo los autores usaron el genotipo G19833 de *P. vulgaris*, una variedad andina. En nuestro caso, el miR5924 fue identificado a partir de la secuenciación de raíces de una variedad mesoamericana de *P. vulgaris* (NAG12). Estos datos indican que dicho miARN sería común a ambas variedades. Dado que, al menos en pelos radicales a tiempos temprano luego del tratamiento con el factor Nod purificado este miARN no presenta un comportamiento diferencial en la variedad andina (Formey et al., 2016), es posible que la respuesta diferencial frente a cepas de tipo *nodC*  $\alpha$  como SC15 sea específica de la variedad Mesoamericana. Para verificar si el comportamiento diferencial del miR5924 frente a la infección con cepas de *R. etli* se haya conservado o difiere entre ambas variedades de *P. vulgaris*, resultaría interesante analizar la acumulación de dicho miARN en raíces de variedades andinas y mesoamericanas luego de la infección con varias cepas *nodC*  $\alpha$  y *nodC*  $\delta$ , utilizando la técnica de *Stem-loop* RT-qPCR (Chen et al., 2005; Varkonyi-Gasic et al., 2007). Esta técnica se intentó utilizar para corroborar el patrón de expresión del miR5924 en la variedad Mesoamericana NAG12. Sin embargo, hasta el momento, esta técnica no nos ha permitido cuantificar la acumulación de este nuevo miARN. La visualización de los productos de amplificación de estas muestras en un gel de agarosa mostró bandas de diferentes tamaños, sugiriendo cierta falta de

especificidad en la reacción de *Stem-loop* RT-PCR. El miR5924 tiene una secuencia poco común para un miARN ya que contiene 5 repeticiones de AAG consecutivas (AGGGCAAGAAGAAGAAGC). Esta particularidad podría interferir en el método de *Stem-loop* RT-PCR, ya sea por un problema del apareamiento al miARN maduro y/o por el apareamiento a otros posibles transcriptos. Para mejorar la detección de los miARNs se podría realizar una amplificación por *Stem-loop* PCR a partir de sARNs purificados, o validar la acumulación de estos mediante Northern *Blot*; aunque este método no permitiría distinguir poblaciones de sARNs con diferencias de uno o dos nucleótidos (Axtell and Meyers, 2018). Otra alternativa sería usar el sistema miRScript de la compañía Qiagen (miScript Plant RT Kit), el cual se basa en la ligación de un adaptador preadenilado el extremo 3' de los miARNs, seguida de la transcripción reversa usando *oligo dT* ligado a una secuencia adaptadora denominada universal y luego un qPCR utilizando un primer universal y uno específico del miARN.

Si bien los precursores de miARNs de plantas son por la mayoría generados a partir de unidad transcripcional no codificante independiente (Xie, 2005; Szarzynska et al., 2009), existen ejemplos de transcriptos en cuales los premiARNs se procesan a partir de precursores en tandem (Chuck et al., 2007; Boualem et al., 2008), o ubicados en las regiones 5' y 3' UTR de transcriptos que codifican proteínas (Rajagopalan et al., 2006), como ocurre para los miARNs de animales. La estructura secundaria del premiR5924 predicha reunió las exigencias requeridas para la identificación de precursores de nuevos miARNs (Rajagopalan et al., 2006; Axtell et al., 2007; Axtell and Meyers, 2018). El premiR5924 forma una estructura de tallo-rulo en la cual el dúplex miARN/miARN\* se aparea formando el brazo del tallo. Dentro de este brazo, se observaron pocos nucleótidos no apareados (Llave et al., 2002; Tang et al., 2003; Kasschau et al., 2003) y dos nucleótidos protuberantes en el 3' para el miARN y su miARN\*, siendo este último un criterio sólido para predecir que el miARN/miARN\* es generado a partir del 5'UTR del transcripto *DCL1* (Elbashir et al., 2001; Park et al., 2002) (Figura 58). El premiR5924 se ubica en la región 5'UTR del transcripto *DCL1* (Figura 59A y Figura 59B). Como detallado en la Introducción, la proteína DCL1 tiene una función central en la biogénesis de los miARNs ya que la misma interviene en el procesamiento del pri-miARN y del premiARN para producir el dúplex miARN:miARN\* (Kurihara and Watanabe, 2004). Dada la importancia de la regulación post-transcripcional de la expresión génica mediada por miARNs en el desarrollo de la planta, no es sorprendente que las plantas mutadas en los genes que codifican las proteínas que participan en la biogénesis de miARNs presenten fenotipos pleiotrópicos severos (Schauer et al., 2002; Reinhart et al., 2002; Kurihara and Watanabe, 2004). Por ejemplo, plantas de *Arabidopsis* mutantes en *DCL1* presentan severos defectos en la formación de embriones (revisado en Schauer et al., 2002). A su vez, la expresión de las proteínas involucradas en la biogénesis de miARNs es finamente regulada durante el desarrollo. En particular, la expresión de *DCL1* es finamente regulada para procesar correctamente los transcriptos no codificantes y generar las poblaciones de miARNs en diferentes etapas del desarrollo (Reinhart et al., 2002). Es así que el mRNA de *DCL1* está sujeto a una retro-regulación negativa a través de la acción del miR162 (Xie et al., 2003). En nuestros estudios de RNA-seq en *P. vulgaris*, se encontró que los niveles de acumulación del transcripto *DCL1* se mantuvieron constante en las tres condiciones ensayadas (Figura 58C). Teniendo en cuenta

la ubicación del premiR5924 en un transcripto esencial para la biogénesis de miARNs, la acumulación diferencial del miR5924 observada podría estar dada por la regulación diferencial del procesamiento del precursor o bien por un cambio en la estabilidad del miARN maduro. En el caso de un procesamiento diferencial esto podría ser por la acción misma de DCL1 por su papel clave en la biogénesis de los miARNs, por una proteína asociada, o por otro regulador sensible al contexto biológico que permita responder a la presencia de la bacteria simbiótica.

### **El miR5924 y su función en el establecimiento de la simbiosis**

Los resultados obtenidos en este trabajo indicaron que la sobreexpresión del precursor del miR5924 altera el número de nódulos formados por la bacteria menos eficiente 55N1 (*nodC*  $\delta$ ) en raíces transgénicas de la variedad mesoamericana (Figura 64). A su vez, se observó un leve retraso en la nodulación en aquellas raíces que sobreexpresan el premiR5924, donde no se observaron nódulos a 5 dpi mientras si se observaron en las plantas control. A 21 dpi, el tiempo final analizado, se alcanzó un plató en la condición de inoculación de 55N1 en las 3 réplicas biológicas realizadas, algo que no ocurrió en las raíces control. Por otra parte, la sobreexpresión del premiR5924 no afectó la formación de nódulos cuando las raíces fueron inoculadas con la bacteria más eficiente (SC15). Sin embargo, los nódulos formados en estas plantas presentaron un tamaño inferior a los del control, tanto luego de la inoculación con SC15 como con 55N1 (Figura 65). Esto datos sugieren que el miR5924 podría afectar no sólo la iniciación de los nódulos, sino también su desarrollo. Las observaciones fenotípicas de los nódulos formados no indicaron diferencias mayores sobre el color de los mismos, los cuales fueron de color rosado tanto en las raíces control como en las de sobreexpresión del miR5924, revelando que los mismos presentan acumulación de leghemoglobina, un indicador de una fijación de nitrógeno eficiente (Dupont et al., 2012).

La organogénesis del nódulo y la infección son dos procesos independientes que han podido separarse genéticamente; sin embargo ambos procesos deben estar coordinados para que estos órganos sean funcionales, es decir que fijen nitrógeno activamente (Oldroyd and Downie, 2008). Para determinar si la sobreexpresión del premiR5924 afecta la infección bacteriana se realizaron inoculaciones con bacterias marcadas con proteínas fluorescentes (Figura 67). La cuantificación de los hilos de infección no mostró diferencias sobre el número de hilos formados (Figura 67A) pero sí sobre la progresión de los mismos (Figura 67B). Las raíces que sobreexpresan el precursor del miR5924 presentan una mayor proporción de hilos que alcanzan el córtex. Si bien, esto se observó para las dos cepas utilizadas la predominancia de los hilos de infección en el córtex fue mayor en la condición de inoculación con 55N1 que en el caso de CFNX5 (Figura 67B). Relacionando estos resultados con la menor cantidad de nódulos formados con 55N1 en las raíces de sobreexpresión, se podría sugerir que la sobreexpresión de miR5924 afecta la coordinación entre la infección rizobiana y la formación de nódulos, en particular en el caso de la bacteria menos eficiente 55N1. Una posibilidad es que si bien las bacterias alcanzan el córtex, las mismas no sean liberadas ya sea porque no encuentran el primordio nodular o bien debido a defectos en el proceso de internalización y formación de los simbiosomas. Esta última

hipótesis estaría de acuerdo con el retraso de la nodulación observado en estas raíces transgénicas que sobreacumulan premiR5924. El tamaño de los nódulos también podría estar relacionado con defectos en la liberación de las bacterias en los nódulos, dando una menor ocupación de los mismos y por lo tanto nódulos más pequeños. Los ensayos de competencia revelaron información interesante acerca de la función que podría cumplir el miR5924 sobre la colonización bacteriana (Figura 66). La sobreexpresión del premiR5924 no modificó la proporción de nódulos formados por las cepas SC15 y 55N1 pero sí la co-ocupación de los mismos. Cerca del 25 % de los nódulos formados fueron co-ocupados por ambas bacterias en las plantas que sobreexpresan el premiR5924, mientras que el porcentaje de co-ocupación fue nulo en las raíces control (Figura 66A). Aunque no haya sido cuantificado, este fenotipo de co-ocupación fue observado nuevamente en un experimento de co-inoculación con bacterias marcadas con proteínas fluorescentes. La observación microscópica en detalles de estos nódulos reveló la presencia de hilos de infección pertenecientes a las dos bacterias asociados al mismo nódulo (Figura 67C). En este experimento no se pudo discriminar entre aquellos hilos que alcanzaban la región central del nódulo y los hilos que eran abortivos. Tampoco pudimos distinguir claramente si los hilos de infección eran monoclonales o si había hilos co-ocupados por ambas cepas. En este sentido, observaciones al microscopio confocal permitirían visualizar las distintas capas celulares y contribuir a contestar estas cuestiones que no han podido ser resueltas mediante epifluorescencia. A su vez, el experimento presentado en la Figura 67, los hilos de infección fueron cuantificados sólo a un tiempo tardío (14 dpi), cuando ya se han formado los nódulos. Para determinar en qué momento la sobreexpresión del premiR5924 altera la progresión de los eventos de infección en las raíces de *P. vulgaris* deberían realizarse nuevos experimentos que cuantifiquen y clasifiquen los hilos de infección a tiempos más tempranos.

### **El miR5924, y su posible función como modulador de las respuestas de defensa**

La predicción de los blancos putativos del miR5924 realizada con el programa psRNATarget (versión 2011) reveló una lista extensa de mARNs (Tabla 11). La vasta cantidad de posibles mARNs blancos podría estar dada por la peculiaridad de la secuencia de este miARN, el cual posee cinco repeticiones de la secuencia AAG en un total de 21 nts. El blanco más probable del miR5924 identificado fue el transcripto *Phvul.006G188700.1*, el cual sería el ortólogo del gen *Medtr2g0112990* de *M. truncatula* y *AT1G61100* de *A. thaliana*. Estos genes han sido anotados como genes codificantes para una proteína de defensa a patógenos de tipo TIR (Toll/interleukin-1 receptor). A su vez, este mismo blanco fue identificado también para el miR5924/miRNov250 por Formey et al., (2015 y 2016). A partir de estos estudios, se dedujo que las tres isoformas de este miARN podrían actuar sobre este mismo mARN blanco, las probabilidades son generalmente más altas para el miR5924 de 21 nts, tanto considerando los valores de máxima expectación (*E*) como los de la accesibilidad del mensajero al miARN (*UPE*). Otros blancos identificados comúnmente con el grupo de Formey fueron los transcriptos *Phvul.011G185300.1* que codificarían por un proteína de la familia de las helicasas a ARN de tipo DEA(D/H)-box (identificado en

posiciones 16 y 4 para las isoformas 21 y 20 nts respectivamente; y los transcriptos *Phvul.006G216000* (1, 2, 3 y 4) codificando para una proteína serina hydroxymethyltransferasa, con probabilidades más bajas (entre la posición 6 y 23 según la isoforma). Los transcriptos *Phvul.005G066000* y *Phvul.003G125700.1* identificados en los datos de degradoma de las bibliotecas de Formey et al., 2016 y Formey et al., 2015 respectivamente, no corresponden a blancos probables predichos por psRNAtarget, (Tabla S11).

En nuestros datos de secuenciación, el miR5924 se acumula a mayores niveles a las 24 hpi con la bacteria más eficiente SC15, mientras que se reprime en respuesta a la inoculación con 55N1 respecto de la condición control (YEM). De manera inversa, los blancos putativos de este miARN son reprimidos por la bacteria SC15, pero no así por la 55N1. A tiempos tempranos, las bacterias simbióticas usan estrategias moleculares similares a las de los patógenos para suprimir las respuestas de defensas y entrar en una simbiosis exitosa (Hacquard et al., 2017). La acumulación diferencial del miR5924 disparada por la bacteria más eficiente SC15 podría actuar como un mecanismo que regula negativamente o suprime las respuestas de defensa generales (receptor PRR) o específicas (receptor NB-LRRs de tipo TIR). Esto permitiría a la bacteria más eficiente colonizar de manera más rápida el primordio del nódulo. Por el contrario, la represión del miR5924 con 55N1, podría aumentar la presencia de receptores que disparen una respuesta de defensa activa, que limite el ingreso de la bacteria menos eficiente 55N1. La presencia de niveles basales del miR5924 en la condición control podría contribuir a los mecanismos generales de defensa disparados por patrones moleculares asociados a microorganismos (MAMPs) frente a los microorganismos patogénicos y simbióticos presentes en la rizósfera. Para sustentar esta hipótesis sería interesante medir la expresión de *Phvul.006G188700.1*, en las raíces que sobreexpresan el precursor del miR5924; así como también llevar adelante la sobreexpresión de versiones resistentes al miR5924 de este blanco putativo evaluando luego la formación de nódulos y la iniciación y progresión de hilos de infección utilizando cepas de *R. etli nodC* tipo  $\alpha$  y  $\delta$ . Podrían evaluarse también la expresión de genes que son marcadores moleculares de las respuestas de defensa frente a ambos tipos de cepas, tanto en plantas control como OXpremiR5924, ya sea mediante determinación de los niveles de mRNA por RT-qPCR o de los niveles de proteínas de defensa por *Western blot*.

Un estudio comparativo de los miARNs entre plantas Angiospermas y Gimnospermas mostró que los miARNs de 22 nts son más numerosos y abundantes en leguminosas, (Zhai et al., 2011). Como se menciona en la Introducción, los miARNs de 22 nts tienen la capacidad de generar la producción de phasiARNs (Chen et al., 2010; Cuperus et al., 2010), los cuales a su vez activan diferentes cascadas de regulación de la expresión génica (Fei et al., 2013). Los NB-LRRs son los blancos de múltiples familias de miARNs de 22 nts independientes. Este es el caso del miR2109 de *M. truncatula*, un miARN de 22 nts que dispara la producción de phasiARNs, a partir del clivaje de mRNA que codifica un NB-LRR (Zhai et al., 2011). El nucleótido 3'-terminal de los miARNs de 22 nts tiene una importancia mayor en la producción de phasiARNs ya que se ha demostrado que las mutaciones en el extremo 3' disminuye drásticamente la producción de phasiARNs (Fei et al., 2015). Sin embargo, en *G. max* la isoforma mayoritaria del miR2109 es de 21 nts (Zhai et al., 2011). La

diversificación del tamaño por la alteración de la proporción de variantes de miARNs de 21 a 22 nts podría proveer cierta flexibilidad en el grado de silenciamiento de los genes blancos debidos a las diferencias de la habilidad que tienen de producir phasiARNs, moléculas amplificadoras del PTGS (Zhai et al., 2011). En nuestros datos de secuenciación la proporción de miARNs de 22 nts es mayor en el caso de los miARNs nuevos, ya sea porque no se hayan depositados en el miRBase y/o que no han sido estudiados previamente. Los resultados preliminares del análisis funcional del miR5924, un miARN con varias isoformas incluyendo una de 22 nts, sugieren una función en el establecimiento de la simbiosis, tal vez mediante la modulación de los genes de defensa. La profundización en los estudios del miR5924 permitiría dilucidar aspectos mecanísticos y funcionales de este miARN en la regulación de la simbiosis.

*Conclusiones  
Y Perspectivas*

# Conclusiones y Perspectivas

Los resultados obtenidos a partir del análisis de los datos de secuenciación de pequeños ARNs generados durante el transcurso de esta Tesis Doctoral constituyen una fuente de información valiosa para identificar nuevos reguladores involucrados en las etapas tempranas de la nodulación. En este trabajo nos hemos enfocado en el estudio de las poblaciones de miARNs y phasiARNs que podrían participar en el establecimiento de una interacción eficiente. Sin embargo, un análisis profundo de los pequeños ARNs de 24 nucleótidos, los cuales fueron los más numerosos en nuestros datos de secuenciación, permitiría identificar nuevos reguladores que actúan a otro nivel de regulación génica, como es el caso de los siARNs de heterocromatina (hc-siARNs) o los miARNs largos con funciones en la metilación del ADN y regulación del estado de condensación de la cromatina.

Particularmente, en este trabajo se analizó la familia del miR171, cuyos miembros se destacan por su participación en la simbiosis. La acumulación diferencial observada para el miR171a en raíces de *P. vulgaris* en respuesta a la cepa más eficiente de *R. etli* permitió documentar no solamente una nueva isoforma del miR171 involucrada en la simbiosis, sino también un nuevo candidato a cumplir funciones en esta asociación preferencial. Su caracterización funcional así como la validación de sus transcritos blancos permitirían profundizar nuestro conocimiento sobre este miembro de la familia miR171 y su función en el establecimiento de la simbiosis preferencial entre *P. vulgaris* y *R. etli*.

La caracterización funcional del miR390b en *P. vulgaris*, permitió verificar que el mismo actúa como un regulador negativo de la simbiosis en una leguminosa que forma nódulos de tipo determinado. El análisis de la vía regulatoria miR390/TAS3/tasiARFs mostró que los componentes de esta vía, se encuentran conservados en *P. vulgaris*. La acumulación diferencial de la isoforma miR390b, pero no de la miR390a, combinada con el estudio comparativo de las secuencias de sus precursores, sugieren nuevas funciones para este miARN conservado en la evolución.

Finalmente, el desarrollo de esta Tesis doctoral permitió identificar un nuevo miARN, el miR5924, implicado en la preferencia de plantas mesoamericanas por cepas más eficientes en la nodulación. El análisis funcional del miR5924 mediante genética reversa y caracterización fenotípica demostró que el mismo cumple una función no sólo en el desarrollo de nódulos fijadores de nitrógeno, sino también en la selección de cepa. Además, la particularidad de su precursor, el cual mapea en la región 5'UTR de *Dicer like 1*, un componente principal de la biogénesis de los miARNs, indica que el miR5924 es generado a partir del procesamiento de una unidad transcripcional de un transcripto codificante. Un estudio detallado de su procesamiento podrían revelar aspectos mecánicos y biológicamente relevantes acerca de la biogénesis de las diferentes isoformas detectadas, y a su vez identificar aquellos componentes asociados a su regulación. Por último, la validación de los blancos putativos identificados, como por ejemplo el mRNA que codifica una proteína TIR asociada a las respuestas de defensa a patógeno, permitiría precisar la función de este miARN en la simbiosis.



# *Materiales y Métodos*

# Materiales y Métodos

## 1. Medios de cultivos y soluciones

### 1.1 Medios de cultivo

#### Medio de cultivo para *Escherichia coli* y *Agrobacterium rhizogenes*

---

##### LB Luria-Bertani

---

Extracto de levadura	5 g/l
NaCl	10 g/l
Triptona	10 g/l

---

Se agregaron 15 g/l de agar para obtener el medio LB-sólido. El medio LB líquido o sólido fue suplementado con el antibiótico adecuado para la selección de cada cepa.

#### Medios de cultivo para rizobios

---

##### Medio YEM

---

Extracto de levadura	0,4 g/l
Manitol	7,0 g/l
NaCl	0,1 g/l
MgSO <sub>4</sub> .7H <sub>2</sub> O	0,2 g/l
K <sub>2</sub> HPO <sub>4</sub>	0,5 g/l

---

Para el medio de cultivo YEM sólido, el medio fue suplementado con 10 ml/l de una solución acuosa de Rojo Congo 2,5 g/l y 15 g/l de agar.

#### Medio Fahraeus

Solución mineral nutritiva para riego de las plantas crecidas en macetas y para preparar las cajas conteniendo el medio sólido, para el cual se agregó 10 g/l de agar bacteriológico (Britania).

Soluciones	Concentración 1000X	
CaCl <sub>2</sub>	114 g/l	
MgSO <sub>4</sub> .7H <sub>2</sub> O	120 g/l	
Citrato férrico	5 g/l	
KH <sub>2</sub> PO <sub>4</sub>	100 g/l	
Na <sub>2</sub> HPO <sub>4</sub>	150 g/l	
<b>Micronutrientes</b>		Los
KCl	3,73 g/l	micronu
H <sub>3</sub> BO <sub>3</sub>	1,55 g/l	trientes
MnSO <sub>4</sub> .H <sub>2</sub> O	0,85 g/l	se
CuSO <sub>4</sub> .5H <sub>2</sub> O	0,13 g/l	prepara
ZnSO <sub>4</sub> .7H <sub>2</sub> O	0,58 g/l	ron en
Na <sub>2</sub> MoSO <sub>4</sub> .2H <sub>2</sub> O	0,018 g/l	una
		solución
		stock
		concent

rada 1000X, la que se esterilizó por filtración a través de un filtro de 0,45 µm de poro.

En el caso de Fahraeus suplementado con fuente de nitrógeno se adicionó KNO<sub>3</sub> hasta una concentración final de 8 mM.

## 1.2 Soluciones

### Antibióticos

Concentración de los antibióticos usados (µg/ml), para los cultivos bacterianos en medios líquidos y sólidos.

Antibiótico	<i>E. coli</i>	<i>A. rhizogenes</i>
Espectinomicina	100	200
Kanamicina	50	100

Antibiótico	<i>R. etli</i>
Tetraciclina	4

### Soluciones para la construcción de las bibliotecas de RNA-seq

Las soluciones se preparan con agua miliQ.

---

**TBE 5X**

Tris base (Invitrogen)	54 g/l
Ácido Bórico	27,5 g/l
EDTA 0,5 M (pH 8.0)	2 % (v/v)

---



---

**Buffer de carga para electroforesis de ácidos nucleídos en geles de agarosa 6X (buffer de tipo III)**

Azul de bromofenol	0,25 % (v/v)
Xilene cyanol FF	0,25 % (v/v)
Glycerol hidratado	30 % (v/v)

---



---

**Buffer de carga para electroforesis de ácidos nucleídos en geles de agarosa 2X**

Buffer de carga 6X	16,6 % (v/v)
TBE 10X	10 % (v/v)
Formamida	73,3 % (v/v)

---



---

**Soluciones**

**Para 100 ml**

Cloruro de Sodio 5 M	29,27 g
Poliétileno glicol 50 % (PEG MW8000)	50 g

---



---

**Policrilamida**

**Para 100 ml**

Acrilamida 40 %	40 g
Bis-acrilamida 2 %	2 g

---

La solución de poliacrilamida se esterilizó a través de un filtro de 0,45 µm de poro y se conservó a 4 °C.

---

**Persulfato de amonio (APS)**

**Para 1 ml**

APS 10 %	10 mg
----------	-------

---

La solución de APS se preparó en el momento y se conservó a -20°C.

## 2. Material biológico

### 2.1 Semillas de poroto

Las semillas de *P. vulgaris* cultivar NAG12 (Mesoamericano) fueron provistas por Susana García Medina (INTA, Estación Experimental Agropecuaria Salta, Cerrillos, Argentina).

### 2.2 Cepas bacterianas

#### Cepas de *E. coli*

Cepa	Aplicación	Referencia
DH5 $\alpha$	Cepa utilizada para la multiplicación y mantenimiento de plásmidos.	(Taylor et al., 1993)
<i>ccdB survival 2 TR1</i>	Cepa provista por la empresa INVITROGEN para la multiplicación y mantenimiento de los vectores destino del sistema GATEWAY portadores del gen <i>ccdB</i> .	Invitrogen

Las cepas de *E. coli* DH5 $\alpha$  y *ccdB survival 2 TR1* fueron crecidas en medio LB líquido a 37 °C y 250 rpm de agitación, o en cultivos sólidos en placas de Petri de 10 cm de diámetro con medio LB-agar.

#### Cepas de *Rhizobium etli* y *Agrobacterium*

Especie	Cepa	Referencia
<i>R. etli</i>	55N1 (polimorfismo <i>nodC</i> $\delta$ )	(Aguilar et al., 2004)
<i>R. etli</i>	SC15 (polimorfismo <i>nodC</i> $\alpha$ )	(Aguilar et al., 2004)
<i>R. etli</i>	CFNX5 DsRed ( <i>nodC</i> $\alpha$ )	(Battaglia et al., 2014)
<i>R. etli</i>	55N1 DsRed ( <i>nodC</i> $\delta$ )	Beker, Tesis Doctoral 2011
<i>R. etli</i>	55N1 GFP ( <i>nodC</i> $\delta$ )	(Blanco et al., 2009)
<i>A. rhizogenes</i>	K599	(Bond and Gresshoff, 1993)

Las cepas de *R. etli* se cultivaron inicialmente en medio YEM sólido con Rojo Congo y posteriormente en medio YEM líquido a 28 °C con agitación a 250 rpm.

Las cepas de *A. rhizogenes* fueron crecidas en medio LB sólido suplementado con los antibióticos adecuados a 28 °C y luego en medio LB líquido selectivo a 28 °C con agitación a 250 rpm. La cepa de *A. rhizogenes* K599 fue donada por el Dr. Federico Sánchez (Universidad Nacional Autónoma de México, Cuernavaca, México).

## 2.3 Transformación bacteriana

### Transformación de células de *E. coli* competentes

A partir de 100 µl de *E. coli* DH5α competentes mantenidas a -80 °C se realizó el protocolo de transformación descrito por Sambrook and Russel, (2001). Uno o 2 µl del plásmido de interés o de una mezcla de reacción se agregaron a las células competentes mantenidas en hielo. La mezcla fue incubada en hielo durante 40 min antes de realizar el tratamiento térmico a 42 °C por 45 seg. Inmediatamente el tubo de reacción fue colocado en hielo durante 5 min. Luego, se adicionó 1 ml de medio LB estéril y las bacterias se cultivaron a 37 °C con agitación (250 rpm) durante 1 hora. El cultivo fue centrifugado a 5000 g durante 2 min. Después de descartar el sobrenadante, el precipitado fue resuspendido con el medio de cultivo restante (~40 µl) y plaqueado en medio LB-agar suplementado con el antibiótico adecuado para la selección del plásmido de interés.

### Transformación de células de *A. rhizogenes* electrocompetentes

Las células competentes se transformaron por electroporación usando el equipo *BioRad Gene Pulser* (BioRad). Uno o 2 µl de plásmido de interés fueron agregados a 40 µl de bacterias competentes *A. rhizogenes* K599 mantenidas a -80 °C. La mezcla se colocó en las cubetas de electroporación (2 mm de espesor) estériles y frías para realizar la electroporación sometiéndole 1 pulso a 2,5 kV. La suspensión bacteriana fue recuperada de la cubeta inmediatamente y colocada en un tubo conteniendo 1 ml de medio LB. El tubo bacteriano se incubó a 28 °C con agitación (250 rpm) durante 1 hora. Luego, la suspensión de células fue centrifugada a 5000 g durante 2 min. Después de eliminar el exceso de medio sobrenadante, las bacterias fueron resuspendidas y plaqueadas en medio LB sólido conteniendo el antibiótico adecuado.

### Medios de conservación

Las cepas bacterianas fueron conservadas a largo plazo a -80 °C en glicerol. Para ello, los cultivos líquidos de las bacterias crecidas hasta fase exponencial tardía (una noche para *E. coli* y dos noches para *R. etli* y *A. rhizogenes*) fueron suplementados con glicerol estéril hasta alcanzar una concentración final de 15 % (v/v) para *E. coli*, o 10 % (v/v) para *R. etli* y *A. rhizogenes*.

### **3. Condiciones de cultivo, transformación de plantas de *P. vulgaris* e inoculación con *R. etli***

#### **3.1 Esterilización y germinación de las semillas**

Las semillas del cultivar mesoamericano NAG12 de *P. vulgaris* se esterilizaron superficialmente con alcohol 96° (v/v) durante 30 seg y fueron transferidas en una solución de hipoclorito de sodio 15 % (v/v) con agitación suave durante 5 min. Luego de 6 lavados con H<sub>2</sub>O destilada estéril, las semillas se colocaron en cajas de petri sobre un papel húmedo y se incubaron durante dos días a 26 °C en oscuridad para su germinación.

#### **3.2 Crecimiento de plántulas**

Las semillas germinadas fueron transferidas a macetas de plástico conteniendo una mezcla húmeda de perlita y arena (3:1) previamente esterilizada. Las plántulas se mantuvieron entre en una cámara de cultivo Percival Intellus Control System I41LLX a 26 °C, 80 % de humedad y un ciclo luz/oscuridad de 16/8 horas. Las plantas fueron regadas con medio Fahraeus suplementado con microelementos y KNO<sub>3</sub> a una concentración final de 8 mM.

#### **3.3 Crecimiento de *A. rhizogenes* K599**

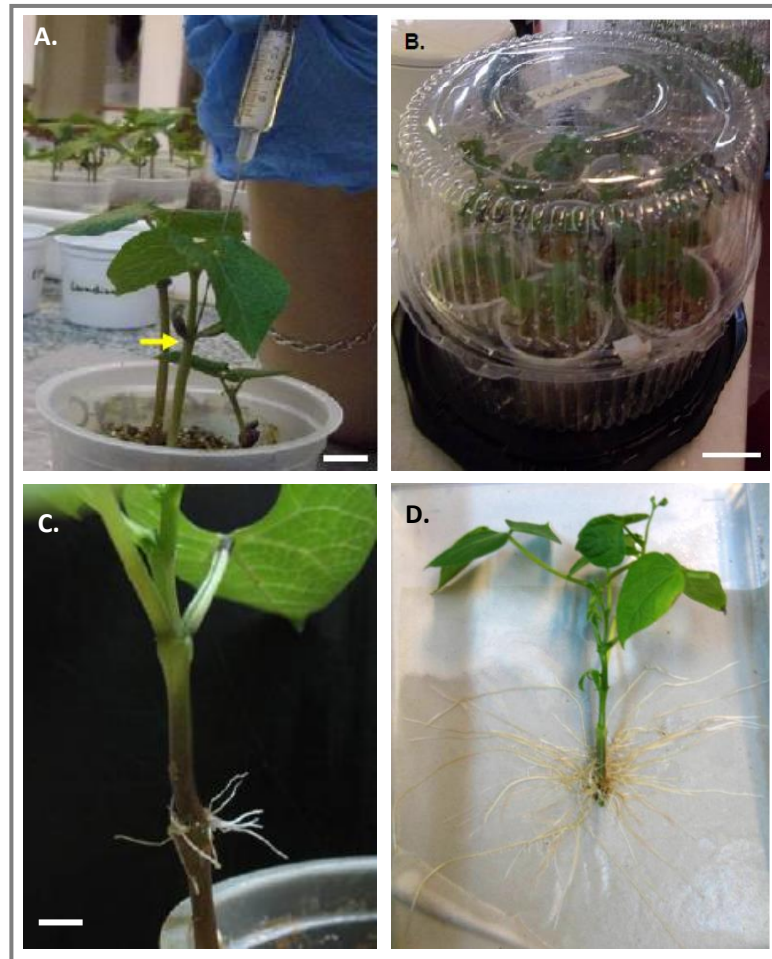
Se cultivaron las cepas de *A. rhizogenes* K599 llevando el plásmido de interés en medio LB-agar suplementado con el antibiótico correspondiente durante dos noches. Las células bacterianas fueron resuspendidas en 1 ml de agua destilada estéril para la transformación de plantas, obteniendo una solución saturada.

#### **3.4 Obtención de plantas compuestas**

Las plantas de *P. vulgaris* crecidas en macetas durante ~6 días fueron transformadas siguiendo la técnica descrita por Estrada-Navarrete et al., (2006, 2007) (Figura 68). Los tallos de las plantas fueron inoculados en la zona de los cotiledones con una suspensión de *A. rhizogenes* K599 utilizando una jeringa Hamilton. Se realizaron 3 inyecciones por planta (Figura 68A). Las plantas se colocaron dentro de recipientes plásticos transparentes a fin de mantener condiciones de humedad saturante (Figura 68B) y se incubaron en cámara de cultivo durante 15 días hasta que las raíces transgénicas salieran de los callos (Figura 68C).

#### **3.5 Crecimiento de plantas compuestas**

Una vez desarrolladas las raíces transgénicas en la zona de inoculación, el tallo fue cortado 1 cm por debajo del sitio de inyección y las raíces *wild type* fueron descartadas. Las plantas compuestas fueron transferidas a cajas de acrílico (17 cm x 22 cm x 4.5 cm) conteniendo medio agar Fahraeus inclinado sin KNO<sub>3</sub> (Figura 68D). Sobre el agar se depositó una hoja de papel estéril para prevenir que las raíces atravesaran el medio y facilitar el recuento de nódulos y la determinación del largo y número de raíces principales y secundarias. Las plantas se mantuvieron en la cámara de cultivo, en las condiciones descritas previamente.



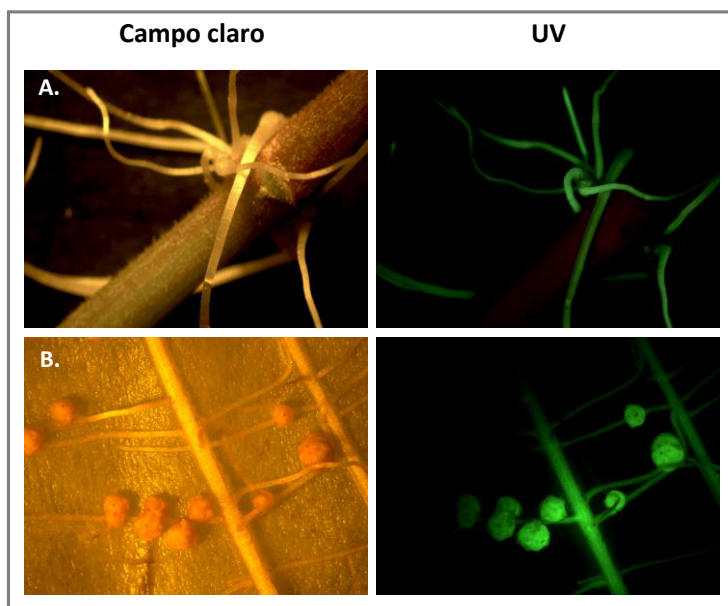
**Figura 68: Transformación de plantas de *P. vulgaris*.** A. Con una jeringa se inyectó una solución saturada de *A. rhizogenes* en la zona de los cotiledones de plantas de porotos de 7 días. B. Los pots conteniendo las plantas transformadas con *A. rhizogenes* fueron colocados en un dispositivo manteniendo una humedad saturada necesaria al desarrollo de raíces transgénicas. Los domos de plásticos se usaron igualmente para el crecimiento de plantas en los ensayos de competencia. C. Se obtuvieron raíces transgénicas o *hairy-roots* en la zona de inyección después de 15 días de crecimiento de las plantas en domos de plástico. D. 5 días después de la aparición de las raíces transgénicas, los tallos de las plantas fueron seccionados aproximadamente 1 cm bajo del sitio de inyección y las plantas compuestas obtenidas fueron crecidas en cajas acrílica sobre un medio sólido conteniendo una solución de Fahraeus. Escala: 1 cm (A), 5 cm (B), 5 mm (C).

### 3.6 Observación de las raíces transgénicas

La observación del sistema radical para identificar las raíces transformadas que expresan la proteína fluorescente verde GFP (*Green Fluorescent Protein*) se realizó utilizando una lupa de fluorescencia Leica MZFLIII (Leica Microsystems, Wetzlar, Germany), (Figura 69A y B). Antes de realizar los estudios fenotípicos o los experimentos de inoculación, las raíces transgénicas y no transgénicas fueron cuantificadas para determinar el porcentaje de transformación para cada construcción. Aquellas raíces no transgénicas, es decir que no presentaban fluorescencia fueron finalmente seccionadas. En cada transformación, varias colonias (entre 3-5) de *A.*



*rhizogenes* que llevan el mismo plásmido fueron testeadas para determinar la eficiencia de transformación de cada una de ellas. Las dos colonias con mayor eficiencia de transformación fueron seleccionadas para los posteriores análisis funcionales.



**Figura 69: Observación de raíces transgénicas.** A. Las raíces transgénicas expresan la *Green Fluorescent Protein* (GFP) cuya fluorescencia es visible con UV. La GFP sirve de gen reportero para la transformación. Las raíces no transgénicas fueron seccionadas antes la inoculación con el rizobio. B. En los nódulos también se puede observar la expresión fuerte de la GFP (nódulos formados por SC15 de *R. etli* a 21 dpi).

Imágenes adquiridas con la lupa Leica MZFLIII y el programa Image-Pro Plus 5.1.

### 3.7 Inoculación de plantas de poroto con rizobios

Las cepas de *R. etli* mantenidas a  $-80\text{ }^{\circ}\text{C}$  fueron plaqueadas en placas de petri conteniendo medio sólido YEM-Rojo Congo y cultivadas durante 2 días a  $28\text{ }^{\circ}\text{C}$ . Las cepas de *R. etli* CFNX5 y 55N1 que expresan las proteínas fluorescentes roja o verde (DsRED y GFP respectivamente) fueron crecidas en medios YEM sólido o líquido suplementados de la Tetraciclina a  $4\text{ }\mu\text{g/ml}$ . A partir de colonias aisladas se realizaron precultivos saturados de *R. etli* resuspendiendo las células en 3 ml del medio YEM líquido, e incubandolas a  $28\text{ }^{\circ}\text{C}$  con agitación a 250 rpm durante 48 horas. Un alícuota de los precultivos fue inoculado en erlenmeyers conteniendo medio YEM en una dilución 1/100. Los cultivos fueron incubados a  $28\text{ }^{\circ}\text{C}$  con agitación hasta alcanzar una densidad óptica a 600 nanómetros ( $\text{DO}_{600\text{ nm}}$ ) de 0,8 (fase exponencial tardía). Las plantas de *P. vulgaris* crecidas en cajas acrílicas durante 6 a 8 días fueron inoculadas con 5 ml de cultivo de *R. etli* a 0,8 o con medio YEM como control en condiciones de esterilidad.

### 3.8 Recolección de tejido de raíz e otros órganos

El tejido de las raíces transgénicas fue colectado a las 24 hpi o a tiempos más tardíos, por ejemplo el de raíces noduladas a 21 o 28 dpi, (Figura 69B) e inmediatamente congelado en nitrógeno líquido para la posterior extracción de ARN. El tejido fue pulverizado con un mortero de porcelana y se almacenó a  $-80\text{ }^{\circ}\text{C}$  hasta su utilización.

El tejido de hoja, tallo, raíz y nódulo de plantas de poroto no transgénicas crecidas en cajas acrílicas durante 7 días post-germinación fueron colectados por separado y

congelados en nitrógeno líquido. El tejido pulverizado, proveniente de 3 a 5 plantas por órgano, fue almacenado a -80 °C hasta su utilización.

## **4. Caracterización fenotípica de las plantas compuestas**

### **4.1 Cinéticas de nodulación**

Las plantas compuestas de poroto crecidas en caja acrílica sin fuente de nitrógeno e inoculadas por cepas de *R. etli* del genotipo *nodCa* o *nodCδ* fueron monitoreadas a diferentes tiempos luego de la inoculación. Teniendo en cuenta que cada raíz transgénica representa un evento de transformación independiente, los nódulos se cuantificaron por raíz a los 5, 8, 11, 15 y 21 dpi. Las plantas inoculadas con el medio YEM sin rizobio fueron observadas ocasionalmente durante el experimento para verificar que no hubo contaminación con otras cepas de rizobio. Se analizaron entre 7 y 10 plantas por cada construcción y condición, lo que correspondió a aproximadamente 50 raíces transgénicas por condición. Se realizaron tres experimentos independientes llevados a cabo en diferentes momentos del año.

### **4.2 Determinación del tamaño de los nódulos**

Las plantas noduladas fueron fotografiadas al final de la cinética (21 dpi) para determinar el tamaño de los nódulos maduros. El diámetro de los nódulos de la totalidad de las plantas inoculadas fue medido usando la herramienta de medición del programa Adobe Photoshop 7.0.

### **4.3 Ensayos de competencia**

Las plantas compuestas portando las construcciones deseadas fueron transferidas a macetas con una mezcla perlita/arena y regadas cada dos días con el medio Fahraeus sin KNO<sub>3</sub>. Las plantas fueron incubadas en domos de plástico durante 8 a 9 días de acuerdo al protocolo estandarizado para ensayos de coinoculación de raíces de *P. vulgaris* en cámara de cultivo (Aguilar et al., 2004). Se realizaron precultivos y cultivos de las cepas SC15 y 55N1 de *R. etli* por separado como se ha detallado en la sección 3.7 de Materiales y Métodos. Los cultivos de ambas cepas crecidos a una DO<sub>600</sub> de 0,8 fueron mezclados en una relación 1:1. Cinco ml de la mezcla de cultivos bacterianos fueron utilizados para inocular un mínimo de 5 plantas por construcción. Veintiséis días después de la coinoculación, se colectaron con un bisturí la totalidad de los nódulos transgénicos de cada construcción. La fluorescencia verde de la GFP fue monitoreada con un iluminador a LED, (UltraSlim MaestroGeen SLB-01W). Los nódulos transgénicos se procesaron para la extracción de ADN bacteriano y posterior análisis del polimorfismo del gen *nodC* como se indica en la sección 4.4 de Materiales y Métodos.

### **4.4 Análisis del polimorfismo del gen *nodC***

Los nódulos provenientes de los experimentos de coinoculación se esterilizaron con alcohol 96 % (v/v) durante 30 seg, seguido de una incubación en H<sub>2</sub>O<sub>2</sub> 30 vol por 4 min

y de 6 lavados con H<sub>2</sub>O estéril. Posteriormente, los nódulos fueron macerados individualmente en 10 µl de PBS 1X estéril. Finalmente, la solución de bacterias se sembró en una placa de Petri conteniendo medio YEM- Rojo Congo. Luego de la incubación durante 2 días a 28 °C, las placas bacterianas fueron transferidas a 4°C. Luego, se tomó una muestra de bacterias provenientes de algunos nódulos para determinar la identidad de la cepa. Las muestras de bacterias se resuspendieron en 200 µl de NaCl 1M, se precipitaron por centrifugación a 17300 xg durante 2 min, y se lavaron con 200 µl de H<sub>2</sub>O estéril. Para obtener el precipitado se centrifugaron nuevamente a 17000 xg por 2 min. El precipitado de bacterias se resuspendió en 200 µl de resina Chelex 6 % p/v (Bio Rad) y se incubó a 56 °C por 8 min seguido de una incubación a 99 °C por 10 min. Las muestras fueron posteriormente centrifugadas a 17300 xg por 2 min y el sobrenadante conteniendo el ADN bacteriano se utilizó para el análisis del polimorfismo del gen *nodC* mediante amplificación por PCR y análisis de restricción con la enzima de restricción *Hinfl* (Promega) según el protocolo descrito previamente (Aguilar et al., 2004). Los oligonucleótidos utilizados se encuentran en la Tabla 1 de oligonucleótidos en la sección Material Suplementario de los Anexos. Los productos de amplificación del gen *nodC* y de la digestión con la enzima *Hinfl* se analizaron por electroforesis en geles de agarosa no desnaturalizantes de 1 % y 1,5 % (p/v) respectivamente.

Los rizobios proviniendo de las placas de medio YEM- Rojo Congo para los ensayos de inoculación fueron regularmente testeados para el análisis del polimorfismo del gen *nodC* para verificar que no hubo contaminación cruzada de los cultivos.

#### **4.5 Caracterización de los eventos tempranos de infección rizobiana**

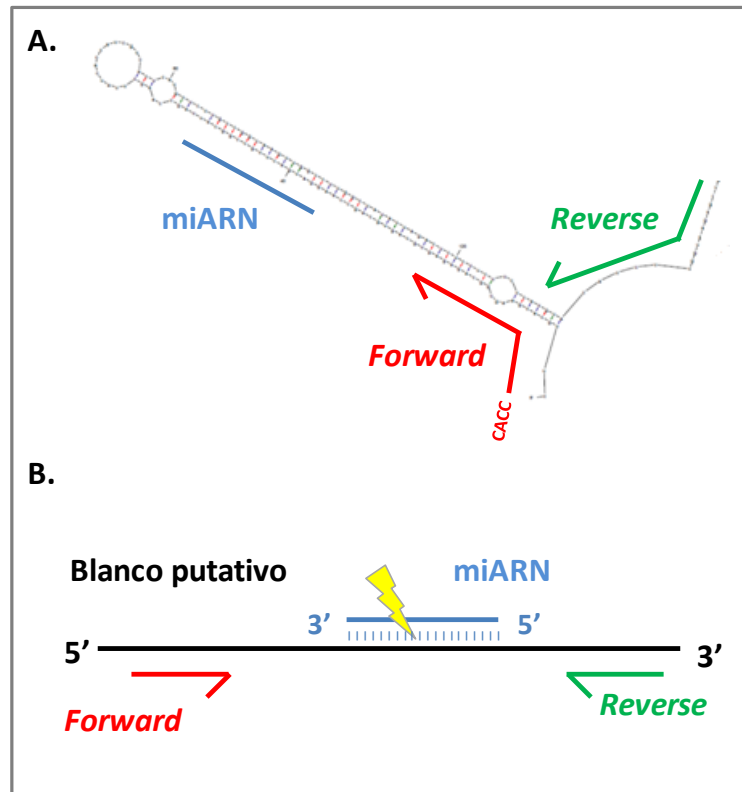
Las raíces transgénicas de *P. vulgaris* de 7 días crecidas en cajas acrílicas en un medio YEM carente de fuente de nitrógeno fueron inoculadas con cepas de *R. etli* que expresan las proteínas fluorescentes DSRed o GFP, las cuales se encontraban disponibles en el laboratorio (Zanetti et al., 2010; Battaglia et al., 2014). Cinco ml del cultivo líquido de las cepas CFNX5 IDsRed, 55N1 IDsRed o 55N1 GFP crecido a DO<sub>600</sub>=0,8 fueron utilizados para la inoculación de las plantas. La formación de hilos de infección (ITs) a 8 y/o 10 dpi y la progresión, y la ocupación de nódulos (hasta 14 dpi) fueron monitoreadas en un microscopio invertido Olympus IX51 utilizando luz blanca y UV con filtros apropiados (Olympus Corporation, Shinjuku Tokio, Japón) tal como se describe en (Battaglia et al., 2014). Las imágenes fueron capturadas usando una cámara Q-Color3 de alta resolución (Olympus Corporation).

## **5. Oligonucleótidos y clonados**

### **5.1 Diseño de oligonucleótidos**

Para el caso específico de los precursores de miARNs, se diseñaron oligonucleótidos directos y reversos que flanquean la región de tallo-rulo del precursor de miARN (Figura 70A). Además, a los oligonucleótidos directos destinados a realizar los clonados usando el sistema GATEWAY se agregaron las bases CACC en el extremo

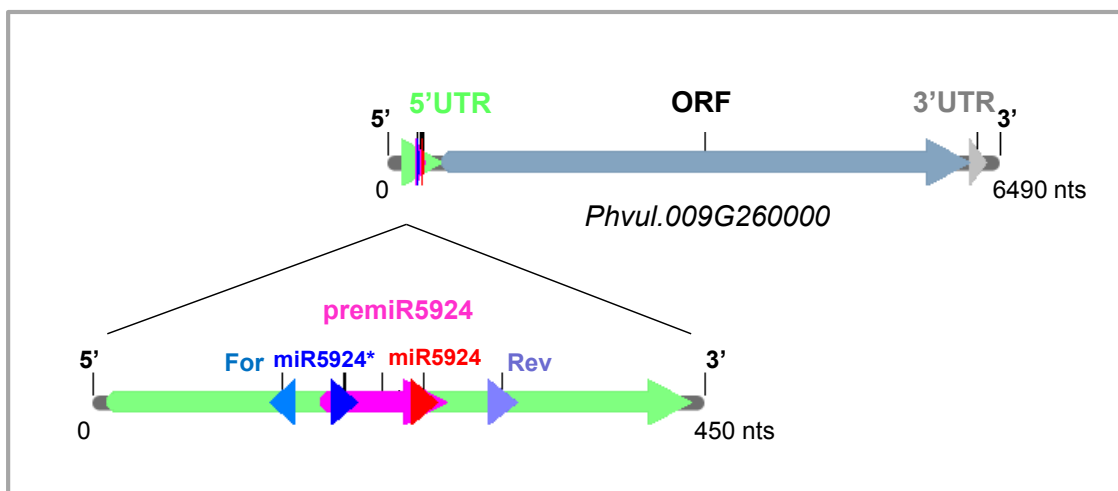
5' como indicado por el proveedor del kit pENTR/D-TOPO (Invitrogen). A su vez, el tamaño de amplificación es compatible con los estudios de expresión por PCR cuantitativa en tiempo real (RT-qPCR). Por otro lado, para los estudios de expresión mediante RT-qPCR de los genes blancos de miARNs, los oligonucleótidos directos y reversos fueron diseñados específicamente a cada uno de los dos lados de la zona de unión y de corte putativo del miARN (Figura 70B).



**Figura 70: Diseño de oligonucleótidos.** A. Los oligonucleótidos diseñados para los precursores de miARNs amplifican ~200 bp. Los mismos oligonucleótidos permiten clonar los precursores de miARNs y medir su expresión por qRT-PCR. Las bases CACC que no pertenecen a la secuencia de los precursores fueron inicialmente agregadas al 5' del oligonucleótido directo (*forward*) para clonar direccionalmente el producto de PCR mediante el sistema GATEWAY. Los oligonucleótidos directos y reversa (*reverse*) se diseñan afuera de la región tallo-rulo, en zonas a donde el precursor no formaría estructuras secundarias aparentadas (predicción de la estructura con mfold). B. Los oligonucleótidos destinados a medir la expresión de los blancos putativos de los miARNs estudiados fueron diseñados de cada lado de la zona de unión del miARN a su transcripto blanco. Eso permitiría dar cuenta del eventual clivaje (ralla amarilla) en los niveles de expresión medidos.

En los casos particulares del clonado del precursor del nuevo miARN caracterizado, el miR5924, los oligonucleótidos *Forward* y *Reverse* fueron diseñados para amplificar una parte del ARN primario (pri-miARN) ubicada en el 5'UTRs del transcripto de *DCL1* (*Phvul.009G260000*) (Figura 71), a su vez el tamaño del producto de amplificación fue compatible con estudios de expresión por PCR cuantitativa en tiempo real (RT-qPCR). El plegamiento de ácidos nucleicos y la predicción de hibridación del precursor se

analizaron usando la herramienta RNA Folding Form de mfold (Zuker, 2003) para diseñar los oligonucleótidos afuera de la región tallo-rulo para poder aparearse fácilmente en las zonas libres.



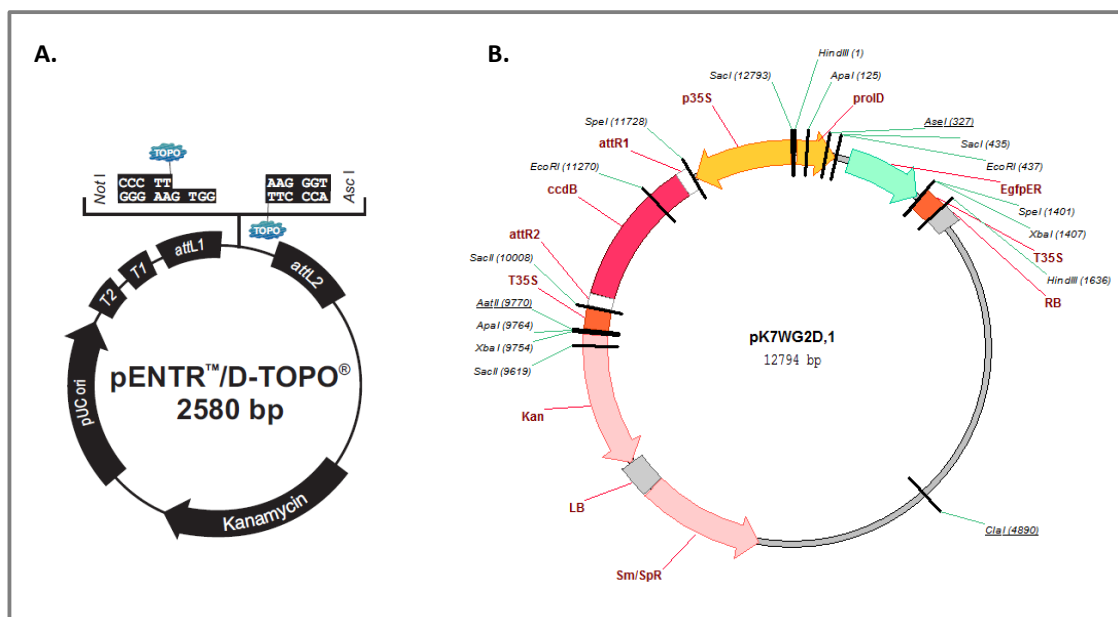
**Figura 71: Posición del precursor del miR5924 y diseño del clonado.** El premiR5924 se ubica en el 5'UTR del transcrito *Phvul.009G260000*. Los oligonucleótidos amplifican un amplicon de 192 nts que incluye la región tallo-rulo del precursor. Este esquema fue realizado con el programa Serial Cloner. En verde y en gris están representados los 5'UTR y 3'UTR respectivamente, en gris oscuro el marco abierto de lectura (ORF), en azul y violeta los oligonucleótidos *Forward* (For) y *Reverse* (Rev) respectivamente y en rojo los miARNs y azul oscuro los miARNs\*.

En todos los casos, la temperatura de *melting*, el porcentaje de GC, la formación eventual de homodimeros, heterodimeros, y de *hairpin* para cada oligonucleotido se determinó utilizando la herramienta OligoAnalyzer 3.1 de Integrated DNA Technologies (IDT, Coralville, IA, <https://www.idtdna.com/calc/analyzer>). Los oligonucleótidos utilizados se encuentran listados en la Tabla 1 en la sección Material Suplementario de los Anexos.

## 5.2 Construcción de vectores mediante el sistema GATEWAY

Para generar las construcciones de sobreexpresión, los precursores de miARNs fueron amplificados mediante PCR usando los oligonucleótidos específicos indicados en la Tabla 1. Los productos de amplificación fueron clonados en el vector pENTR/D-TOPO (Figura 72A) siguiendo las instrucciones del proveedor (Invitrogen, Carlsbad CA, USA). Posteriormente, los vectores pENTR/D-TOPO conteniendo los fragmentos de ADN fueron recombinados con el vector de destino GATEWAY pk7WG2D,1 (Figura 72B) utilizando la enzima LR clonasa de acuerdo a las instrucciones provistas por el fabricante (Invitrogen). Este vector permite la expresión de fragmentos de ADN bajo el control del promotor cuasi-constitutivo y fuerte 35S del virus del mosaico de coliflor (CaMV35S) y el terminador 35S (T35S). Además el pk2WG2D,1 contiene el gen que codifica la proteína de la GFP que sirve de gen reportero para visualización bajo luz UV de las raíces y nódulos transgénicos (ver Figura 69). Como control de los ensayos de sobreexpresión se utilizó el vector vacío previo a la recombinación. Los vectores

fueron multiplicados por transformación de *E. coli* y transferidos en *A. rhizogenes* K599 mediante electroporación.



**Figura 72: Vectores usados para los clonados de los precursores de miARNs.** A. Vector de entrada: pENTRY. El fragmento purificado fue clonado direccionalmente entre los dos sitios *attL* del vector de entrada mediante las bases CACC inicialmente agregadas al 5' del oligonucleótido directo. Estas bases permiten al fragmento unirse por complementariedad de base e insertarse en el vector. Este plásmido contiene el origen de replicación (pUC ori) que permite la fuerte replicación del plásmido y su mantenimiento en *E. coli* y las secuencias de terminación de la transcripción (T1 y T2). El pENTRY/D-TOPO dispone del gen de resistencia a la Kanamicina, lo que permite seleccionar las bacterias transformadas. B. Vector de destino: pK7WG2D,1. La recombinación entre el plásmido de entrada y el de destino es posible por los sitios de recombinación *attL* y *attR* respectivamente que posee cada plásmido. La construcción deseada se insertó en el lugar del gen *ccdB* y se encuentra bajo el promotor fuerte p35S, con su terminador T35S. Este plásmido posee el gen de resistencia Spectinomycin/Spectromycin (Sm/SpR) permitiendo su selección en bacterias. Este vector de destino contiene también el gen de selección de Kanamicina (Kan) que se usa generalmente para la selección de las raíces transgénicas en plantas. Sin embargo en el caso de *P. vulgaris* el crecimiento de aquellas no permite su selección con antibióticos, para eso este plásmido contiene el gen de la GFP (EgfpER).

## 6. Técnicas comunes de biología molecular

### 6.1 Reacciones de PCR

Las amplificaciones se realizaron en un volumen final de 20 µl. La concentración final fue 2,5 µM de cada oligonucleótido, 0,25 mM de cada dNTP, 1,5 mM de MgSO<sub>4</sub>, 1 X *buffer* comercial para Taq polimerasa y 0,5 U de la enzima Taq polimerasa de ADN

(Productos Bio-Logicos). Las reacciones se llevaron a cabo en un ciclador térmico de 96 posillos Applied Biosystems Veriti. El tiempo de elongación y la temperatura de apareamiento de los oligonucleótidos utilizados fue variable para cada reacción. El programa estándar consistió en una primera etapa a 95 °C durante 5 min, seguido de 35 ciclos de 30 seg a 95 °C, 30 seg a 52 °C (o T<sub>m</sub> indicado) y 1 min por kb a 72 °C. Se incluyó una etapa final de elongación de 5 min a 72 °C. Para la amplificación de los precursores de miARNs y su clonado en vectores binarios se usó la enzima Pfu (Promega) en un volumen final de 25 µl con el programa siguiente: 95 °C durante 1 min 30 seg, seguido de 40 ciclos de 15 seg a 95 °C, 30 seg a 52 °C (o T<sub>m</sub> indicado), 1 min a 68 °C, y 2 min finales a 68 °C.

## 6.2 PCR de colonias bacterianas

Las colonias bacterianas obtenidas sobre placas selectivas luego de la transformación fueron testeadas por PCR. Para cada clonado, un mínimo de 10 colonias aisladas fueron picadas con escarbadiantes estériles y transferidas a tubos de PCR de 0,2 ml estériles. Las muestras conteniendo el ADN de la bacteria y eventualmente el plásmido de interés fueron amplificadas con oligonucleótidos mixtos: un *primer* del plásmido (oligonucleótido M13 para el vector pENTR/D-TOPO o los oligonucleótidos pKan o p35S para el vector de destino según la etapa del clonado), y un oligonucleótido específico para el fragmento de ADN de interés. Estas reacciones de PCR permiten validar a su vez la presencia del fragmento y su direccionalidad. Las PCR fueron llevadas a cabo en un ciclador térmico Applied Biosystems con el programa siguiente: 10 min a 95 °C, seguido de 35 ciclos de 30 seg a 95 °C, 30 seg a 52 °C (o T<sub>m</sub> indicado) y 1 min por kb a 72 °C. Se incluyó una etapa final de 5 min a 72 °C. Las temperaturas y tiempos de *anealing* y elongación variaron según el clonado y los oligonucleótidos utilizados.

## 6.3 Electroforesis de ADN y ARN en geles de agarosa

La integridad y la concentración aproximada del ADN plasmídico purificado fueron analizados en geles de agarosa en un rango de 0,8 - 2 % (p/v) según el tamaño de los fragmentos de ADN a separar. En todos los casos, se utilizó el marcador de peso molecular adecuado para determinar el tamaño del amplicon (100 bp, 400 bp o 1 kb ladder de la empresa Productos Biológicos). La integridad y la concentración del ARN fue estimada en geles de agarosa 1,2 % (p/v). Para la preparación de los geles, la agarosa se fundió en un *buffer* TBE 0,5X (Tris base 45 mM, H<sub>3</sub>BO<sub>3</sub> 45 mM y EDTA 1 mM) por calentamiento. A esta solución se agregaron unas gotas de una solución de bromuro de etidio (BrEt) 10% (p/v). Las muestras de ADN y ARN fueron mezcladas con *buffer* de ADN 6X (Tris-HCl 20 mM pH 8,0, glicerol 50 % (p/v), azul de bromofenol 0,5 % (p/v), EDTA 2 mM pH 8,0) para obtener una concentración final 1X y sembradas en los pocillos de los geles. Las electroforesis se llevaron a cabo en *buffer* TBE 0,5 X con un voltaje constante (usando 10 V por cada cm lineal de gel). La visualización de las bandas de ADN o ARN se realizó utilizando un transiluminador de luz UV Uvitec Cambridge y la captura de imágenes se realizó utilizando el sistema Uvisave HD5 del Uvitec Cambridge.

#### **6.4 Purificación de ADN a partir de gel de agarosa**

Las bandas correspondientes a los productos de amplificación fueron escindidas del gel de agarosa con la ayuda de un bisturí. Para la purificación del ADN se usó el kit *Nucleo Spin Gel and PCR Clean-up DNA, RNA and protein purification* de Macherey - Nagel siguiendo las recomendaciones del proveedor. El ADN fue eluído con 25  $\mu$ l de *buffer* de elución. El rendimiento de la purificación del ADN fue analizado sembrando una alícuota de 3 o 4  $\mu$ l del producto obtenido en un gel de agarosa.

#### **6.5 Purificación de ADN plasmídico**

A partir de las colonias bacteriana de *E. coli* que resultaron positivas en los ensayos de PCR sobre colonias bacteriana se establecieron cultivos líquidos conteniendo 3 ml de medio LB y el antibiótico correspondiente al plásmido de interés. El cultivo fue incubado a 37 °C con agitación toda la noche. Se obtuvo un precipitado de células de bacterias transformadas por centrifugación a 5000 g durante 2 min. A partir del precipitado se procedió a la separación y purificación de ADN plasmídico utilizando el *kit Wizard Plus SV Minipreps* (Promega), siguiendo las instrucciones del proveedor. El ADN plasmídico se resuspendió en 50  $\mu$ l de agua miliQ estéril.

#### **6.6 Extracción de ARN total de tejido vegetal**

El material vegetal se pulverizó en mortero con nitrógeno líquido hasta obtener un polvo fino. Este tejido fue almacenado a -80 °C o procesado inmediatamente. La extracción de ARN se realizó con TRIZOL siguiendo las instrucciones del proveedor (Invitrogen). La resuspensión del ARN se realizó con 50  $\mu$ l de agua estéril y se almacenó a -80 °C hasta su utilización. La concentración del ARN se determinó en un espectrofotómetro Nanodrop2000 (Thermo Scientific) utilizando un volumen de 1  $\mu$ l. La concentración de la muestra se calculó a partir de la  $A_{260}$  y se estimó su pureza a través de la relación  $A_{260}/A_{280}$ , la cual fue cercana a 2. La integridad del ARN se analizó mediante una electroforesis no desnaturizante en geles de agarosa 1,2 % (p/v), visualizándose las bandas correspondientes al ARN ribosomal. Además, la visualización de ARNs en la zona de bajo peso molecular indicó la presencia de los ARNs pequeños.

#### **6.7 Tratamiento del ARN con DNAsa**

Para eliminar el ADN genómico que podría contener la muestra, 1  $\mu$ g de ARN se digirió con 1 U de la enzima RQ1 DNasa libre de RNasa (Promega) utilizando el *buffer* suministrado por el fabricante, en un volumen final de 10  $\mu$ l. Luego de incubar durante una hora y media a 37 °C se agregó 1  $\mu$ l de *DNasa Stop Solution* y se incubó durante 10 min a 65 °C para inactivar la enzima.

#### **6.8 Síntesis de ADN complementario**

La síntesis de la primera cadena de ADNc se realizó con 1  $\mu$ g de ARN tratado con DNasa como molde. Por cada  $\mu$ g de ARN se usaron 0,5  $\mu$ g de oligo dT<sub>15</sub> como cebador. La mezcla se incubó durante 5 min a 70 °C para eliminar la formación de estructuras secundarias del ARN y luego se mantuvo en hielo. A continuación se agregó una mezcla conteniendo 200 U/ $\mu$ l de transcriptasa reversa (MMLV RT, Promega), dNTPs 0,5 mM (Promega) y *buffer* 5X provisto con la enzima en un volumen final de 25  $\mu$ l. La mezcla de reacción se incubó durante 60 min a 42 °C,



temperatura a la cual tiene lugar la síntesis de ADNc, y finalmente 5 min a 95 °C para inactivar la enzima.

## 7. Análisis de expresión

### 7.1 Reacciones de PCR semicuantitativas

Las amplificaciones se realizaron en un volumen final de 20 µl. La concentración final fue 2,5 µM de cada *primer*, 0,25 mM de cada dNTP, 1,5 mM de MgSO<sub>4</sub>, 1 X *buffer* comercial para Taq polimerasa y 0,5 U de la enzima *Taq* polimerasa (Productos Bio-Logicos). Las reacciones se llevaron a cabo en un ciclador térmico (Applied Biosystems). El tiempo de elongación y la temperatura de apareamiento de los oligonucleótidos utilizados fue variable para cada reacción. El programa estándar para los genes de referencia y otros transcritos consistió en una primera etapa a 95 °C durante 5 min, seguido de 25 a 28 ciclos de 30 seg a 95 °C, 30 seg a 52 °C (o T<sub>m</sub> indicado) y 1 min por kb a 72 °C. Se incluyó una etapa final de 5 min a 72 °C. Para los precursores de miARNs las qPCR se realizaron a 35 ciclos (contra 40 ciclos para una PCR saturada).

### 7.2 PCR cuantitativa en tiempo real (qRT-PCR)

Cinco µl de una dilución 1/10 de la síntesis de ADNc fueron utilizados como molde en una reacción de PCR cuantitativa. La concentración de cada *primer* fue de 0,5 µM en un volumen final de 20 µl. Se utilizó la mezcla de reacción comercial *iQ SYBR Green Supermix* de BioRad. La reacción se llevó a cabo en un termociclador Bio-Rad CFX96 Real Time System (BioRad). El protocolo de amplificación fue el siguiente: una primera etapa a 95 °C durante 5 min, seguido de 45 ciclos de 30 seg a 95 °C, 30 seg a 52 °C o 51 °C según los *primers* usados y 20 seg a 72 °C. Se realizaron dos réplicas técnicas, un control negativo sin ADNc para evaluar la contaminación y una curva de calibración con diluciones seriadas del ADNc para cada par de oligonucleótidos utilizados. La posible dimerización de los oligonucleótidos o la amplificación de otros fragmentos fueron analizadas a partir de los resultados de las curvas de *melting*. La obtención de un único producto de amplificación del tamaño esperado también se verificó por corrida electroforética en geles de agarosa. La normalización de los datos de expresión se realizó con el transcrito correspondiente al factor de elongación 1α (*EF1α*) de *P. vulgaris*, según fuera descrito previamente por (Meschini et al., 2008). Los datos fueron analizados y graficados utilizando el programa GraphPad Prism V5.0 (GraphPad Prism Software).

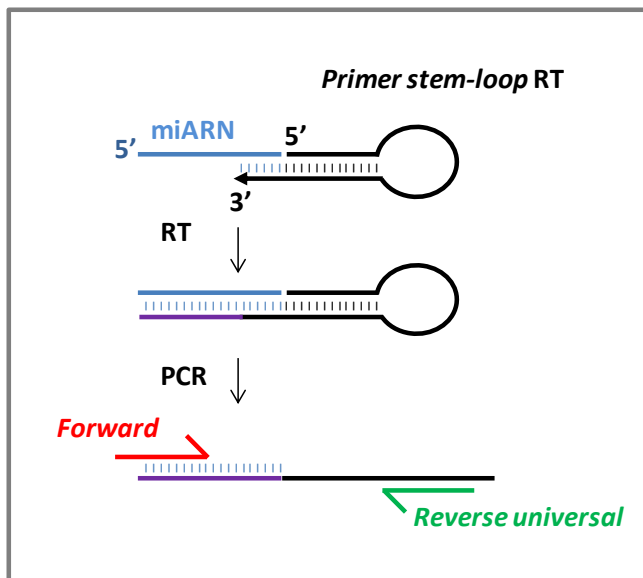
## 8. Técnica de *Stem-loop* PCR

### 8.1 Diseño de oligonucleótidos de *stem-loop*

Los oligonucleótidos de *stem-loop* para realizar la transcripción reversa (*primers stem-loop* RT) fueron diseñados según las recomendaciones de (Chen et al., 2005) y más detallado en (Kramer, 2011). La región común de los oligonucleótidos *stem-loop* RT

consiste de 44 nucleótidos (5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGAC -3'), mientras que la especificidad para cada miARN individual está dada por la extensión de los 6 nucleótidos en el extremo 3'. Esta extensión corresponde a la secuencia reversa y complementaria a los 6 últimos nucleótidos del extremo 3' del miARN, tal como es ilustrado en la Figura 73.

El oligonucleótido denominado universal corresponde a una secuencia de 16 nucleótidos de la región común de los *primers stem-loop* RT (5'-GTGCAGGGTCCGAGGT -3'). Los oligonucleótidos directos son específicos de la secuencia del miARN, pero excluyen los últimos 6 nucleótidos finales en la extremidad 3' del miRNA. Tal como ha sido recomendado por (Varkonyi-Gasic et al., 2007), una extensión de 5 a 7 nucleótidos generalmente enriquecida en GC fue agregada al extremo 5' de cada oligonucleótido directo para incrementar la temperatura de *melting*. Estas secuencias de nucleótidos fueron elegidas al azar. Estos oligonucleótidos fueron analizados con OligoAnalyzer 3.1 de IDT como explicitado previamente en la sección "diseño de oligonucleótidos".



**Figura 73: Técnica de *stem-loop*.** La parte 3' constituida de 6 nucleótidos de los *primers stem-loop* RT se hibridan a la extremidad 3' de un miARN. A partir de ARN total se realiza la Transcripción reversa (RT) que sintetiza la hebra complementaria del miARN. La amplificación del ADN complementario y su cuantificación por PCR y qRT-PCR se realiza usando los oligonucleótidos *Forward* que son específicos a cada miARN y el *Reverse universal* que se hibrida con la parte común de los *primers stem-loop* RT.

## 8.2 Transcripción reversa para los miARNs

Para realizar la transcripción reversa se preparó previamente una mezcla conteniendo la totalidad de los oligonucleótidos *stem-loop* RT correspondiendo a los miARNs de interés (Tabla S12), con una concentración final a 1  $\mu$ M. Esta mezcla se denominó *primer RT Super Mix*. La reacción de transcripción reversa se realizó siguiendo el protocolo descrito por (Varkonyi-Gasic et al., 2007) con las siguientes modificaciones. Se usaron entre 1 y 1,5  $\mu$ g de ARN total extraído previamente tratados con DNAsa, a los cuales se agregaron 1  $\mu$ l de *primer RT Super Mix* 1  $\mu$ M y 1  $\mu$ l de dNTPs 10 mM. La mezcla se incubó durante 5 min a 65  $^{\circ}$ C para eliminar la formación de estructuras secundarias del ARN y luego se mantuvo en hielo. A continuación se agregó una mezcla conteniendo 1  $\mu$ l de dTT a 0,1 M, 1  $\mu$ l de agua

miliQ, 1 µl de transcriptasa reversa (SuperScript III, Invitrogen), y 4 µl de *buffer* 5X provisto con la enzima en un volumen final de 20 µl. La mezcla de reacción se incubó durante 60 min a 50 °C, temperatura a la cual tiene lugar la síntesis la hebra complementaria de los miARNs, y finalmente 15 min a 70 °C para inactivar la enzima. Las reacciones se llevaron a cabo en un ciclador térmico (BioMetra UNO-Thermoblock).

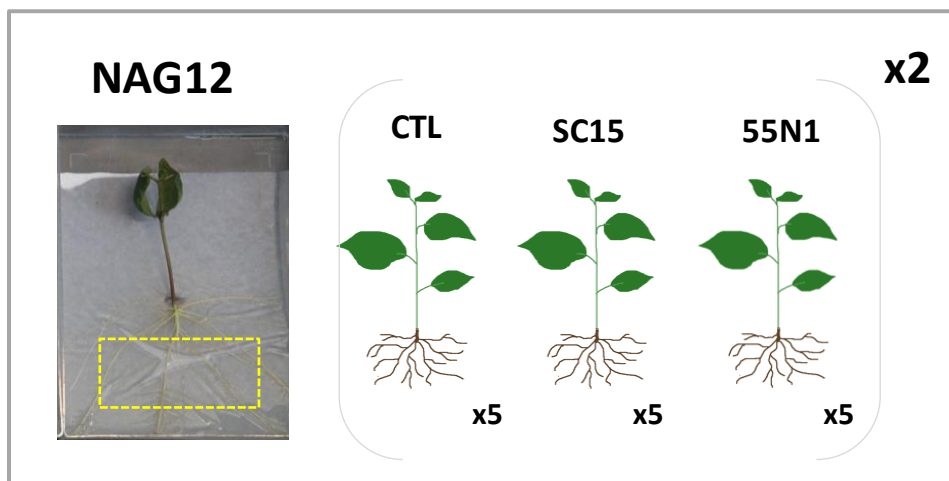
### 8.3 Medición de la acumulación de los miARNs por qPCR

Cinco µl de una dilución 1/5 de la reacción de *stem-loop* RT fueron utilizados como molde en una reacción de PCR cuantitativa en un volumen final de 20 µl, con dos réplicas técnicas por cada muestra. El protocolo de amplificación consistió en una primera etapa a 95 °C durante 5 min, seguido de 45 ciclos de 5 seg a 95 °C, 10 seg a 60 °C o 62 °C según los *primers* utilizados y 1 seg a 72 °C. La normalización de los datos de expresión se realizó con el pequeño ARN U6 según fuera descrito previamente (Peltier and Latham, 2008; Turner et al., 2013a; D'haene et al., 2012). Los datos fueron analizados y graficados utilizando el programa GraphPad Prism V5.0 (GraphPad Prism Software).

## 9. Construcción de bibliotecas de pequeños ARNs

### 9.1 Material biológico

Para la realización de bibliotecas de pequeños ARNs se generó tejido de raíces de *P. vulgaris* de variedad mesoamericana NAG12 de 5 plantas por condición. Las plantulas fueron crecidas en cajas de acrílico (17 cm x 22 cm x 4.5 cm) conteniendo medio agar Fahraeus inclinado sin KNO<sub>3</sub>. Sobre el agar se depositó una hoja de papel para prevenir que las raíces atravesasen el medio. Las plántulas se mantuvieron en una cámara de cultivo Percival Intellus Control System a 26 °C, 80 % de humedad y un ciclo luz/oscuridad de 16/8 horas. A los 7 días las raíces fueron inoculadas por SC15, una cepa alfa de *R. etli*; 55N1, una cepa delta; o YEM, el medio de cultivo rizobiano libre de bacterias (CTL). Cinco ml de cultivo o de solución YEM fueron aplicados sobre la zona II de las raíces, la cual corresponde a la zona con pelos radicales en activo crecimiento (Figura 74). Las cajas permanecieron 1h en posición horizontal y luego se colocaron en posición vertical en la cámara de cultivo. Las raíces fueron colectadas 24 horas luego de la inoculación con las cepas SC15 y 55N1 de *R. etli*. El material vegetal se pulverizó en mortero con nitrógeno líquido hasta obtener un polvo fino y almacenado a -80°C y procesado rápidamente. Se generaron dos replicas biológicas independientes.



**Figura 74: Diseño experimental para la construcción de bibliotecas de pequeños ARNs.** Las plantas de *P. vulgaris* mesoamericanas de la variedad NAG12 crecidas en cajas acrílicas durante 7 días sin fuente de nitrógeno fueron inoculadas con las cepas de *R. etli*, SC15 (cepa NodC  $\alpha$ ), 55N1 (cepa NodC  $\delta$ ), o la condición control (YEM, medio de cultivo bacteriano sin rizobios). A las 24 horas después de la inoculación, la zona de las raíces en crecimiento (cuadro amarillo) fue colectada para la extracción de ARN. Una muestra corresponde a la recolección de las raíces de 5 plantas. Luego se repitió el experimento, obteniendo de esta manera una segunda réplica biológica independiente.

## 9.2 Extracción y separación de ácidos ribonucleicos

### Extracción de ARN total

A partir de dicho tejido se realizó la extracción de ARN total con TRIZOL (Invitrogen) siguiendo las instrucciones del proveedor. Con el fin de aumentar la concentración del ARN, se realizaron 4 extracciones de ARN en paralelo, utilizando 100 mg de tejido en cada extracción. La resuspensión del ARN se realizó con 25  $\mu$ l de agua estéril. Después de colocar los tubos conteniendo la solución de ARN en baño seco a 60°C durante 10-15 min, los mismos fueron conservados toda la noche a 4 °C para su total disolución. La integridad del ARN se analizó mediante una electroforesis no desnaturante en geles de agarosa 1,5 % (p/v), visualizándose las bandas correspondientes al ARN ribosómico y los pequeños ARNs. Las 4 extracciones simultáneas de ARN provenientes del mismo tejido fueron *pool*eadas para obtener un volumen de 100  $\mu$ l de ARN total concentrado.

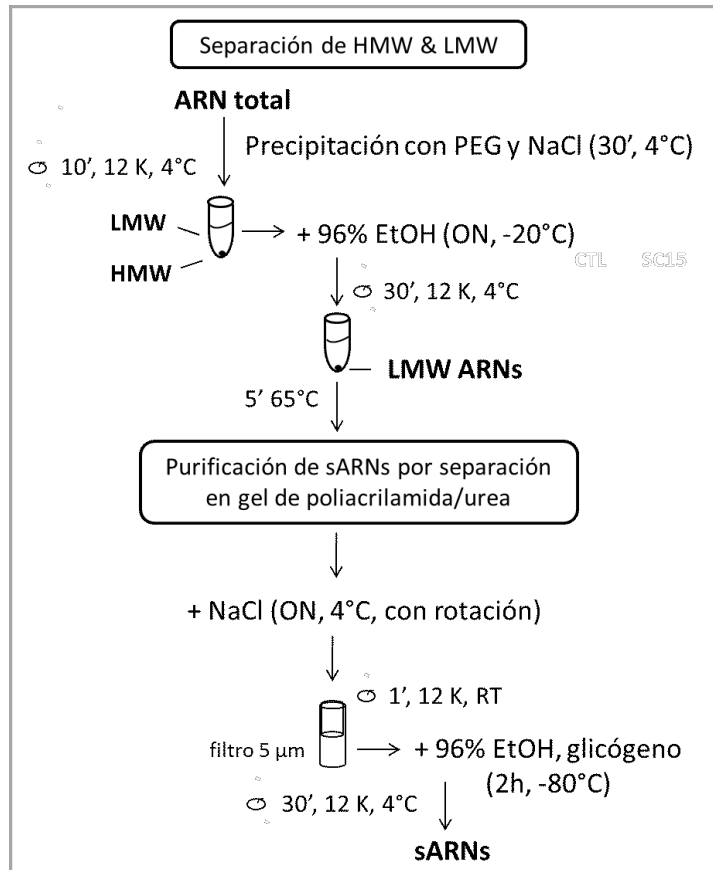
### Separación de ARN de alto y bajo peso molecular

La separación de los ARNs de alto (HMW) y bajo peso molecular (LMW) se llevó a cabo siguiendo el protocolo de precipitación con Polietileno glicol (PEG) descrito por Lu et al., 2007, el cual se esquematizó en la Figura 10. Los ARNs mensajeros y ribosómicos (de alto peso molecular) se precipitaron mediante el agregado de PEG y NaCl a concentraciones finales de 5 % y 0,5 M, respectivamente. La mezcla se mantuvo en hielo durante 30 min y se centrifugó a 13000 rpm durante 10 min a 4 °C. El sobrenadante conteniendo los ARN pequeños se transfirió a un nuevo tubo, se adicionaron 2,5 volúmenes de etanol 96 % (v/v), se mezcló por inversión y se incubó

toda la noche a -20 °C. Luego, de una centrifugación a 13000 rpm durante 30 min a 4 °C, el precipitado se lavó con etanol 80 %. Después de una centrifugación a 13000 rpm a 4 °C durante 15 min, se dejó secar el precipitado por 20 min y finalmente los ARN de bajo peso molecular fueron resuspendidos en 10 µl de agua miliQ estéril e incubados a 4°C toda la noche para asegurar su total disolución. La integridad de los ARNs se analizó mediante electroforesis no desnaturizante en geles de agarosa 1,5 % (p/v).

### **Purificación de los ARNs de 20 a 31 nts**

Los pequeños ARNs se separaron en un gel de poliacrilamida/urea a 15 % (p/v). Para la preparación del gel se mezclaron 4,8 g de urea, 3,75 ml de acrilamida 40 %-bisacrilamida 2 %, 1 ml de TBE 10X en un volumen final de 10 ml completando al volumen con agua miliQ esteril. La mezcla se calentó a 37 °C hasta disolver la urea (~20 min), y luego se adicionaron 60 µl de APS y 4,6 µl de TEMED. Se mezcló por inversión y la solución se volcó entre los vidrios del armador de geles utilizando sistema Mini-PROTEAN 3 (Bio-Rad). El gel de acrilamida se dejó solidificar durante 30 min a temperatura ambiente. Antes de sembrar las muestras, se hizo una pre-corrída del gel durante 30 min a 200 V. A los 10 µl de solución de ARNs de bajo peso molecular se agregó el mismo volumen del buffer de carga de ARN 2X. Las muestras de ARNs fueron calentadas a 65 °C durante 5 min para desasociar las estructuras secundarias y colocadas en hielo inmediatamente. Para determinar el tamaño de los pequeños ARNs en el gel se utilizaron fragmentos de ADN simple cadena de un tamaño de 21 nts y 31 nts. La electroforesis se realizó a 200 V constantes durante 1 h. El gel se tiñó durante 20 min en una solución de bromuro de etidio a 1 µg/ml. La población de pequeños ARNs contenida entre 20 y 31 nts, conteniendo los ARNs de 21 y 24 nts se escindieron del gel y se colocaron en tubos de 1,5 ml *Non-stick*, libres de ARNasa y estériles para proceder a su purificación (Figura 75). Se agregaron 250 µl de NaCl 0,3 M en cada tubo antes de moler el gel con un pequeño mortero limpio de plástico. El ARN fue eluido por rotación suave a 4 °C toda la noche. El eluido y los fragmentos de gel fueron transferidos a una columna de filtración (Spin-X filter, 5 µm), la cual se centrifugó a máxima velocidad durante 1 min a temperatura ambiente. Al filtrado se le agregaron 500 µl de etanol 96 % (v/v) frío y 3 µl de glicógeno y se incubó a -80 °C un mínimo de 2 horas para precipitar los ARNs pequeños. Las muestras se centrifugaron a máxima velocidad a 4 °C durante media hora, se lavaron con 200 µl de etanol frío a 80 % y se centrifugaron nuevamente a máxima velocidad a 4 °C durante 10 min. Se descartó el sobrenadante y el precipitado se dejó secar a temperatura ambiente durante 15 min. Los ARNs de 20 a 30 nts se resuspendieron en 10 µl de agua miliQ estéril.

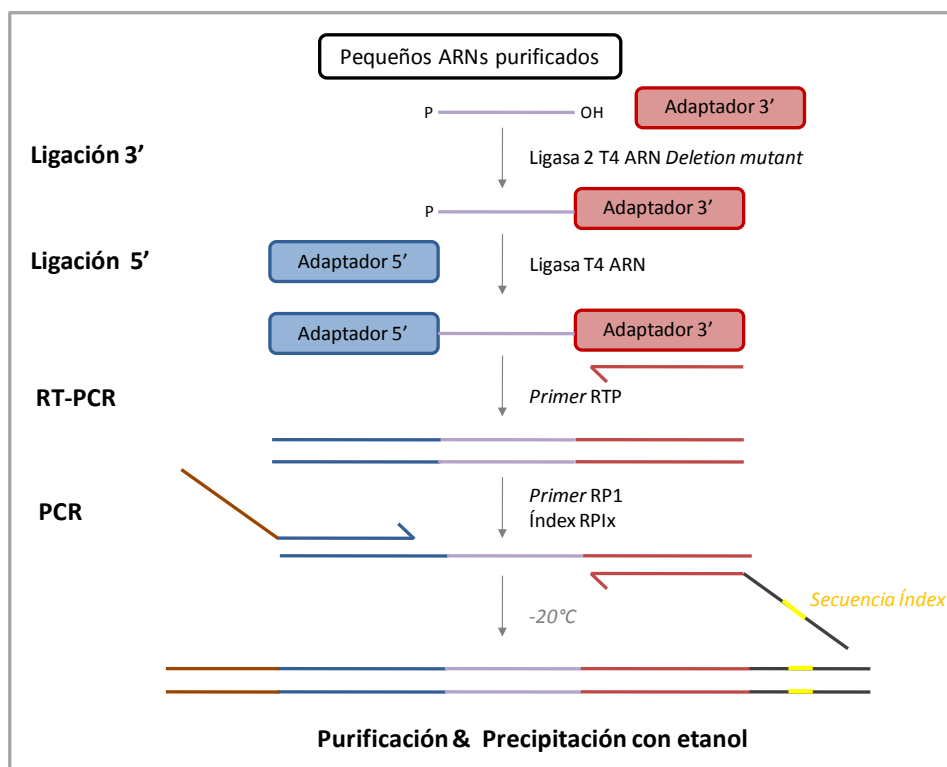


**Figura 75: Aislamiento de pequeños ARNs (sARNs) y purificación.** La extracción del ARN con Trizol permite obtener el ARN total. Los ARNs de alto peso molecular (*HMW* por *High Molecular Weight*) precipitan con polietileno glicol (PEG) y cloruro de sodio (NaCl), mientras los ARNs de bajo peso molecular (*LMW* por *Low Molecular Weight*) quedan en solución. La integridad de los ARNs (totales, *HMW* y *LMW*) es validada por corridas electroforesis en geles de agarosa 1,5%. Los pequeños ARNs son seleccionados por su tamaño (~35 a 18 nt) a partir de la corrida de los ARNs de bajo peso molecular en un gel de poliacrilamida/urea 15% y purificados a través de un filtro de 5 µl. ON: *Overnight* - RT: *Room temperature*

### 9.3 Generación de bibliotecas de pequeños ARNs

#### Input para las bibliotecas: pequeños ARNs purificados

Las muestras de pequeños ARNs purificados sirvieron de material de partida para la construcción de bibliotecas usando el kit TruSeq Small RNA preparation kit, siguiendo las recomendaciones del proveedor (Illumina), las cuales se esquematizan en la Figura 76.



**Figura 76: Construcción de bibliotecas de pequeños ARNs con Truseq sRNA Preparation Kit de Illumina.** Los adaptadores de ARN son específicamente ligados a las extremidades 3' y 5' de los pequeños ARNs purificados. Luego la primera hebra de ADN es sintetizado por transcripción reversa (RT-PCR) y el producto amplificado por PCR. Los *primers* RPIx son específicos de cada condición y contienen una secuencia de 6 nucleótidos únicos que permiten su identificación. Los productos de PCR pueden ser guardados a  $-20^{\circ}\text{C}$  antes realizar la etapa de purificación de los productos en el gel y su precipitación.

El protocolo de Illumina está basado sobre la presencia del grupo fosfato en el extremo 5' terminal de los miARNs. Esta particularidad permite el uso de ligasas que dirigen la ligación específicamente a los miARNs. En la primera etapa de la construcción de bibliotecas de pequeños ARNs con Truseq sRNA Preparation Kit de Illumina, el adaptador adenilado con la extremidad 3' bloqueada es ligada a la muestra de ARN con la ARN Ligasa 2 T4 *Deletion mutant*. Esta enzima es modificada para requerir el adaptador 3' como sustrato para ser adenilada. Eso permite que solamente los oligonucleótidos pre-adenilados van a ser ligados con el ARN que tiene el 3' libre. Una vez que el adaptador 3' es bloqueado, ese mismo no puede servir de sustrato y ser auto-ligado. En la segunda etapa el adaptador 5' es agregado con ATP y la ligasa a ARN 1. Únicamente las moléculas de ARN con la extremidad 5' fosforiladas serán sustratos efectivos para la ligación. Luego se procede a la transcripción reversa con un oligonucleótido RT que se hibrida al adaptador 3'.

El kit de Illumina: TruSeq Small RNA Sample Preparation Kit / Part #15004197/ Core Solution (Ref.15016911, lot.8206572), Indices A (Ref. 15016912, lot.8324069); contiene todos los reactivos necesarios para su realización, excepto la T4 ARN Ligase 2 Deletion Mutant y el kit de *SuperScript II* que fueron proveidas por Invitrogen. Las soluciones necesarias fueron preparadas en el momento. Las reacciones fueron realizadas en tubos eppendorf de 200  $\mu\text{l}$  esteriles y libres de nucleasa. Las sucesivas

incubaciones fueron realizadas en un termociclador Applied Biosystems Veriti. En las siguientes subsecciones se detalla el protocolo seguido para la construcción de las bibliotecas.

### **Ligación de los adaptadores 3'**

Un  $\mu\text{l}$  del adaptador 3' (RA3) fue agregado a 5  $\mu\text{l}$  de la solución de pequeños ARNs purificados. La mezcla de reacción fue incubada a 70 °C durante 2 min, y colocada en hielo inmediatamente. En un tubo separado se preparó la mezcla de reacción (multiplicado por el número de muestra) con 2  $\mu\text{l}$  de buffer de ligación (HML), 1  $\mu\text{l}$  de inhibidor de ARNasa y 1  $\mu\text{l}$  de ligasa 2 T4 ARN *Deletion mutant*. Cuatro microlitros de esta solución de reacción fue agregada a la primera solución conteniendo las muestras de pequeños ARNs y adaptador 3', obteniendo un volumen total de 10  $\mu\text{l}$ . El tubo de reacción fue incubado 1 h a 28 °C. Luego, la reacción fue inactivada agregando 1  $\mu\text{l}$  de Solución Stop (STP) al tubo de reacción e incubándolo 15 min a 28 °C. Las muestras se conservaron en hielo hasta la siguiente reacción.

### **Ligación de los adaptadores 5'**

A cada mezcla de reacción del paso Ligación de los adaptadores 3' se les agregó 1  $\mu\text{l}$  de adaptador 5' y se incubó durante 2 min a 70 °C. Inmediatamente, las muestras se colocaron en hielo y se adicionó 1  $\mu\text{l}$  de ATP a 10 mM y 1  $\mu\text{l}$  de ligasa T4 ARN a cada muestra. La reacción final (14  $\mu\text{l}$ ) fue incubada 1 h a 28 °C y luego se mantuvo en hielo hasta el paso siguiente.

### **Reacción de transcripción reversa**

La reacción de transcripción reversa se llevó a cabo adicionando 1  $\mu\text{l}$  de *primer* ARN RT (RTP) a 6  $\mu\text{l}$  de las muestras de pequeños ARNs ligados a los adaptadores 5' y 3'. La reacción (7  $\mu\text{l}$  totales) fue incubada 2 min a 70 °C e inmediatamente colocada en hielo. En un tubo aparte se preparó una mezcla de reacción conteniendo 2  $\mu\text{l}$  de buffer *First Strand 5X*, 0,5  $\mu\text{l}$  de dNTPs a 12,5 mM, 1  $\mu\text{l}$  de DTT a 100 mM, 1  $\mu\text{l}$  de inhibidor de ARNasa y 1  $\mu\text{l}$  de enzima de transcriptasa reversa *SuperScript II* (Invitrogen). Esta mezcla de reacción fue agregada al tubo de reacción, la solución final de 12,5  $\mu\text{l}$  fue incubada a 50 °C durante 1 h, y mantenida en hielo nuevamente.

### **Amplificación por PCR**

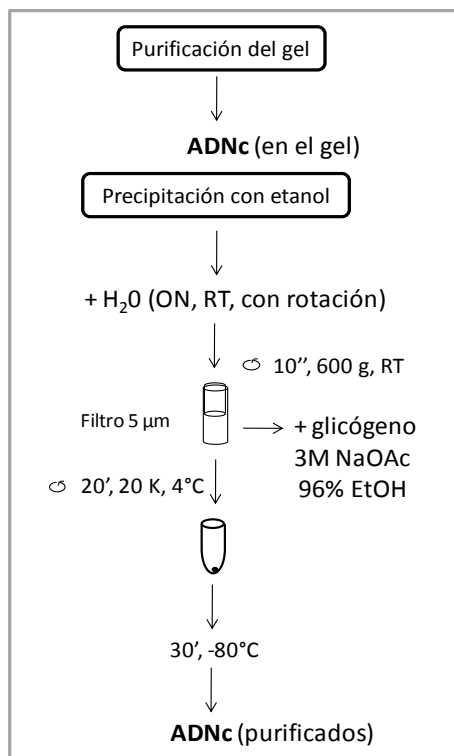
En esta etapa se utilizaron diferentes adaptadores indexados para cada condición que permiten luego *pool*ear las bibliotecas y realizar la secuenciación en una misma línea abaratando así el costo de secuenciación. Para cada muestra se preparó una mezcla de reacción siguiente (PCR *master mix*) conteniendo 8,5  $\mu\text{l}$  de agua miliQ, 25  $\mu\text{l}$  de *Mix* de PCR (PML), 2  $\mu\text{l}$  de *Primer* PCR (RP1), 2  $\mu\text{l}$  de *Primer* Índice PCR (RPIX, en el cual X corresponde al número del índice usado). Los *TruSeq Index* que fueron utilizados son: RPI1: ATCACG (CTL #1), RPI2: CGATGT (SC15 #1), RPI3: TTAGCC (55N1 #1), RPI4: TGACCA (CTL #2), RPI5: ACAGTG (SC15 #2), RPI6: GCCAAT (55N1 #2). Los 37,5  $\mu\text{l}$  de PCR *master mix* fueron agregados a los 12,5  $\mu\text{l}$  de la reacción de transcripción reversa obteniendo un volumen final de 50  $\mu\text{l}$ . Los tubos de reacción fueron colocados en el termociclador para realizar la amplificación por PCR utilizando



el programa siguiente: 30 segundos a 98 °C; luego de 11 ciclos de [10 seg. a 98°C, 30 seg. a 60 °C, 15 seg. a 72 °C]; 10 min a 70 °C; por un tiempo total de 28 min.

### Purificación del producto de PCR

Los ADNc obtenidos luego de la reacción de amplificación por PCR fueron separados en un un gel de poliacrilamida 6 % (p/v). El gel de 8 ml se preparó mezclando 1,2 ml de acrilamida 40 %-bisacrilamida 2 %, con 1,6 ml de TBE 5X, completando con 5,1 ml agua miliQ estéril. Se agregaron 56 µl de APS y 4 µl de TEMED, se mezcló por inversión, se colocó en el sistema de arando de geles Mini-PROTEAN 3 (Bio-Rad) y se incubó durante 30 min a temperatura ambiente. Los pocillos del gel fueron limpiados con TBE 1X estéril. El volumen total de muestra a sembrar en el gel fue de 60 µl (50 µl de ADNc + 10 µl del buffer de carga de ADN a 6X). Cada muestra fue sembrada en dos pocillos. Cuatro µl de la solución de *Custom Ladder* y 2 µl de la solución de *High Resolution Ladder* fueron cargados en el gel. La electroforesis se realizó a 145 V durante 40 min. El gel se tiñó en una solución de bromuro de etidio a 0,5 µg/ml durante 5 min. Las bandas de los productos de amplificación de la biblioteca (correspondiendo al ADNc, los adaptadores y los oligonucleótidos) con un tamaño de ~140 y 160 nts marcado por el *Custom Ladder* fueron visualizadas en transiluminador bajo luz UV. Con ayuda de un bisturí se cortaron las bandas de interés y se colocaron en tubos de 0,5 ml libres de ARNasa y estériles que fueron previamente perforados con una aguja. Los tubos de 0,5 ml conteniendo las porciones de gel fueron insertados en unos tubos de 2 ml libres de ARNasa y estériles. Ambos fueron centrifugados a 13000 rpm a temperatura ambiente para obtener los fragmentos de gel en el tubo de 2 ml para proceder a su purificación (Figura 77).



**Figura 77: Purificación y precipitación del ADN.** Los productos de la PCR fueron sembrados en un gel de poliacrilamida a 6 %, para seleccionar los ADNc de interés. Los fragmentos de gel recolectados fueron filtrados y el ADN precipitado con etanol.

El ADN conteniendo los fragmentos de gel fue eluido por incubación con 300 µl de agua miliQ estéril a temperatura ambiente toda la noche con agitación. El contenido del tubo fue transferido en un filtro de 5 µm que fue centrifugado 10 seg a 600 g.

Después de descartar el filtro, se adicionaron 30  $\mu$ l de NaOAc a 3 M, 975  $\mu$ l de etanol a 96 % (p/v) frío y 2  $\mu$ l de glicogeno como agente coprecipitante. Inmediatamente la solución fue centrifugada a 13000 rpm a 4 °C durante 20 min. El sobrenadante fue descartado y el precipitado fue lavado con 500  $\mu$ l de etanol a 70 % (p/v). Se centrifugó a 13000 rpm durante 2 min a temperatura ambiente, se descartó el sobrenadante y el precipitado se dejó secar en un baño seco a 37 °C durante ~10 min. El ADN se resuspendió en 10  $\mu$ l de Tris-HCl a 10 mM, pH 8,5 y se conservó a -20 °C hasta su utilización.

## 9.4 Validación y secuenciación de las bibliotecas

### Control de calidad y normalización

La concentración e integridad de las bibliotecas fueron analizadas mediante electroforesis capilar utilizando un Agilent Technologies 2100 Bioanalyzer. Un  $\mu$ l de una dilución 1/2 del producto de PCR fue separado en un *chip* DNA-1000 siguiendo las recomendaciones del proveedor. Teniendo en cuenta los resultados de concentración provistos por el Bioanalyzer, cada biblioteca fue normalizada a una concentración final de 24,9 nM ( $\approx$ 2,33 ng/ $\mu$ l), correspondiendo al valor de concentración de la muestra menos concentrada. Las muestras con concentraciones más altas, fueron diluidas con agua miliQ estéril. Finalmente, 3,6  $\mu$ l de cada muestra normalizada (6 muestras, ~20  $\mu$ l) fueron *pool*eadas en un mismo tubo para su secuenciación.

### Secuenciación

Las bibliotecas *pool*eadas fueron secuenciadas en un secuenciador HiSeq2000 utilizando el servicio HTS de la University of California Riverside, USA. Para ello, se realizaron corridas que arrojan lecturas simples (*single reads*) de 51 pares de bases. Los datos fueron descargados de la plataforma de secuenciación (<http://illumina.ucr.edu/ht/>).

## 10. Análisis bioinformáticos

### 10.1 Bases de datos de secuencias genómicas

Se utilizó la base de datos genómica de *P. vulgaris* v2.1, la cual corresponde al genoma del cultivar andino Race Peru G19833 (Schmutz et al., 2014) y se encuentra disponible públicamente en *Phytozome* v12. (<https://phytozome.jgi.doe.gov/pz/portal.html>). El genoma mesoamericano de la variedad BAT93 fue secuenciado recientemente (Vlasova et al., 2016) y se encuentra libremente disponible en la SRA (*Sequence Read Archive*) de la NCBI (*National Center for Biotechnology Information*), bajo los números de acceso: PRJNA221782 (*BioProject ID*); SRS488731 (*genomic 454, SOLiD y HiSeq reads*); SRS488023, SRS488025, SRS489191-255 (*GBS HiSeq reads*); y SRS498664, SRS498673-76, SRS498904-933 (*RNA-Seq HiSeq reads*).

## 10.2 Análisis de datos de secuenciación masiva

### Procesamiento de la data cruda con Galaxy

La conversión de los datos de secuenciación del formato *Fastq* a *Fastq Sanger* se realizó a través de la interface Galaxy versión 2.2.6.2 (<https://usegalaxy.org/>, (Blankenberg et al., 2010b; Goecks et al., 2010; Blankenberg et al., 2010a; Blankenberg and Hillman-Jackson, 2014) usando la herramienta *NGS: QC Manipulation > FastQ Groomer*. Se procedió al análisis de la calidad de las lecturas con *NGS: QC Manipulation > FastQC*. Para convertir los archivos de *fastq* a *fasta*, o viceversa, se usó la herramienta *Converts Formats > FASTQ to FASTA converter*, o *NGS: QC Manipulation > combine FASTA and QUAL into FASTQ* (Blankenberg et al., 2010a), respectivamente.

### Filtrado de la data de secuenciación con Workbench

Las etapas de filtrado se realizaron con el programa UEA Workbench, The small RNA Workbench Versión 3.2 (Stocks et al., 2012; Beckers et al., 2017; Mohorianu et al., 2017), siguiendo las recomendaciones previstas por el manual del programa (Lott, 2012, The UEA sRNA Toolkit). The UEA small RNA Workbench es un programa que puede descargarse libremente para el procesamiento y análisis de data de secuenciación masiva de pequeños ARNs (<http://srna-workbench.cmp.uea.ac.uk/>). El procesamiento de los datos y su subsecuente análisis fueron realizados en una computadora Inter Pentium I7 con 4 núcleos y 32G RAM. Las herramientas del programa fueron ejecutadas bajo Windows 10, con el asesoramiento del Dr. Mancini Ulises, Profesional Asistente de CONICET en bioinformática y biología de sistemas.

Los archivos *fastq* obtenidos en Galaxy fueron filtrados por tamaño y calidad usando la herramienta '*Tools > Helper Tools > Adapter Removal*', entrando la secuencia 3' del adaptador de Illumina (TGGAATTCTCGGGTGCCAAGGAAGTCCAGTCAC) remanente en la data de secuenciación. Se aplicó un filtro de tamaño de 18 a 26 nucleótidos, así como también filtros de calidad eliminando la baja complejidad y las secuencias inválidas. Luego se eliminaron las secuencias correspondiendo a los ARN ribosomales y ARN de transferencia (ARNr-t) usando la herramienta *Filter Tools > Helper Tools > Filter* con el parámetro *Discard T-R RNA* seleccionado. Esta herramienta utiliza las secuencias *\$INSTALL\_PATH/data/t\_and\_r\_RNAs.fa* disponibles en el paquete del programa proviniendo de RFAM (Schmutz et al., 2014; Griffiths-Jones et al., 2003; Gardner et al., 2009) versión 10, the Genomic tRNA Database (Chan and Lowe, 2009) y EMBL (*European Molecular Biology Laboratory*, Kulikova et al., 2004) versión 95. Los archivos de salida obtenidos fueron las secuencias redundantes y no-redundantes (seleccionando la opción '*Make Output Non-Redundant*'), ambos en formato *fasta*, sumados a los datos de filtrado en formato de texto.

### Mapeo contra los genomas de *P. vulgaris* con Galaxy

Los alineamientos de las lecturas a los genomas de *P. vulgaris* disponibles se realizaron mediante la herramienta *NGS: Mapping > Bowtie2 - map reads against reference genome de Galaxy* (Langmead et al., 2009; Langmead and Salzberg, 2012; Langmead, 2013) entrando los datos de secuencia en formato *fasta*, (*pvulgaris\_v10.fa*

y *P.vulgaris\_218.fa*) para el genoma mesoamericano y el genoma andino, respectivamente. Se obtuvieron archivos de las secuencias alienadas y non-alineadas, archivos *.bam* y *.bai*. Estos últimos archivos sirvieron de material de partida (*input*) para realizar un análisis estadístico que permite comparar los alineamientos entre cada genoma: NGS: *Picard > Collect Alignment Summary Metrics*, *output*: archivo excel).

### **Análisis de la data de secuenciación con Workbench**

En cada etapa de análisis se usó la herramienta de mapeo contra el genoma seleccionando el parámetro '*Filter reads that do not match the genome*', entrando el archivo del genoma de referencia. A partir de los archivos redundantes filtrados por tamaño, calidad y limpiados de adaptadores y de los ARN ribosomales y de transferencia, se procedió a la identificación de miARN usando la herramienta en *Workbench Tools > Analysis Tools > miRCat* (Moxon et al., 2008), con los genomas de referencia de *P. vulgaris* andino y mesoamericano en corridas independientes. Se aplicaron los parámetros default de plantas (*sRNA Sequence Flank Extension=100*, *Min Free Energy Threshold=-25.0*, *sRNA Sequence Length=20-22*, *Max Genome Hits=16*, *Max Unpaired Bases=50%*, *Max Consecutive Gaps=3*, *Min GC=20%*, *Min sRNA Locus Size=1*, *Min sRNA Orientation=80%*, *Min Hairpin Length=60*, *Min sRNA Abundance=1*, *Loci Separation Distance=200*, *Min Paired Bases=17*, *Maximun Overlap=80*, *P\_Val=0.05*, *Thread Count=7*), con la opción de bucles complejos seleccionada. La versión de miRBase al momento del análisis fue la versión 21. Se exportaron los archivos de miARNs identificados en formato *fasta* y los resultados en *.csv*.

Para el análisis de los phasiARNs, se usó la herramienta *Tools > Analysis Tools > TA-SI Prediction*. El programa implementa un algoritmo descrito en Chen et al., (2007) para calcular la probabilidad de secuencias de 21 nts en fase basándose en los datos genómicos de referencia. Se usaron los mismos archivos redundantes filtrados usados como *input* para el análisis de miARNs, aplicando los parámetros default sin modificación (*Min abundance=2*, *Pvalue=0.0001*, *Phasing register=21*). Se obtuvieron 4 archivos de datos por cada condición: los phasiARNs en texto con su secuencia y su abundancia, dos archivos *fasta* (tasiARN y sus *loci* respectivos), y una tabla de resultados *.csv*.

Las herramientas del programa UEA Workbench usadas para identificar miARNs y tasiARNs no permiten discriminar el origen de las secuencias identificadas, y por lo tanto no pueden relacionar la abundancia de una secuencia a un *locus* cuando aquellos son múltiples. Para identificar el precursor más probable de un miARN *multi-loci*, se analizaron en detalle las estructuras secundarias de tallo-rulo de los precursores putativos usando la herramienta *Tools > Visualisation > RNA Folding / Annotation* de UEA Workbench, usando las secuencias del miARN maduro y del miARN complementario (miARN\*) correspondiente. En el caso de un phasiARN *multi-loci* se averiguó que la secuencia esté en fase con las secuencias de 21 nts identificadas que pertenece a un mismo locus.

## Visualización de la data de secuenciación

Para visualizar las lecturas alineadas al genoma se usó el programa IGV (*Integrative Genomics Viewer*) (Thorvaldsdóttir et al., 2013; Robinson et al., 2011), el cual puede descargarse libremente (<http://software.broadinstitute.org/software/igv/>) y posee una interface intuitiva que acepta una gran variedad de fuente de datos, como datos de secuenciación masiva, genomas y anotaciones. Adicionalmente se usó la herramienta *Tools > Visualisation > VisSR* de UEA Workbench que tiene el mismo propósito.

Para realizar los diagramas de Venn se identificaron los miARNs y phasiARNs identificados según el genoma Mesoamericano o Andino utilizado. Las listas con las secuencias de estos sARNs fueron cargadas en la herramienta web *Calculate and draw custom Venn diagrams*, disponible en el portal de la Universidad de Gent (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). El programa genera el diagrama mostrando la superposición de elementos, así como también, tablas mostrando los elementos compartidos y únicos entre categorías. Para realizar los *heat maps* se usó el programa Morpheus del Broad Institute (<https://software.broadinstitute.org/morpheus/>).

## 10.3 Análisis de secuencias nucleotídicas y proteicas

### Alineamiento y comparación de secuencias

Para buscar regiones de similitud a nivel de nucleótidos o de amino ácidos con las bases de datos públicamente disponibles se usaron las herramientas *online* BLAST (*Basic Local Alignment Search Tool*) de NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) y Phytozome (*Tools > BLAST*, versión 12) (<https://phytozome.jgi.doe.gov/pz/portal.html>). Teniendo en cuenta las diferencias de anotación y terminología entre los genomas Andino y Mesoamericano (genomas organizados en cromosomas o *scaffold* respectivamente), para la comparación de los datos genómicos obtenidos entre ambos genomas se usó una programa de BLAST local descargado desde NCBI (Nucleotide-Nucleotide BLAST 2.4.0+). Este programa se corrió en Windows usando una línea de comando (`C:\Program Files\NCBI\blast-2.4.0+bin>blastn -query seq.txt -db pvulgaris.v10.fa -out outseq`).

Se usó el programa Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) para el alineamiento de secuencias de ADN, ARN y proteínas. Estos mismos alineamientos fueron tratados en BOXSHADE ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)) para modificar su formato.

Para el estudio de sintenia de uso la plataforma PLANT GENOME DUPLICATION DATABASE, (<http://chibba.agtec.uga.edu/duplication/index/locus>).

### Herramientas para clonados y biología molecular

Para analizar mapas de vectores y diversos análisis de secuencias se usó el programa Vector NTI (Invitrogen, [www.lifetechnologies.com](http://www.lifetechnologies.com)). Adicionalmente se usó Serial Cloner (versión 2.6.1) (<http://www.serialbasics.com>), que es un programa libre de acceso que

proporciona numerosas herramientas de biología molecular como análisis de secuencias de ADN, proteínas, realizar PCR, clonados, recombinaciones y restricción *in silico*.

Para obtener la secuencia complementaria y reversa de secuencias nucleotídicas se usó Reverse Complement ([http://www.bioinformatics.org/sms/rev\\_comp.html](http://www.bioinformatics.org/sms/rev_comp.html)).

### **Herramientas específicas para ARN**

Se usó miRBase (<http://www.mirbase.org/>) como referencia para base de datos de miARNs (Griffiths-Jones, 2005). La búsqueda se realizó por secuencia o identificador de miARN (ID).

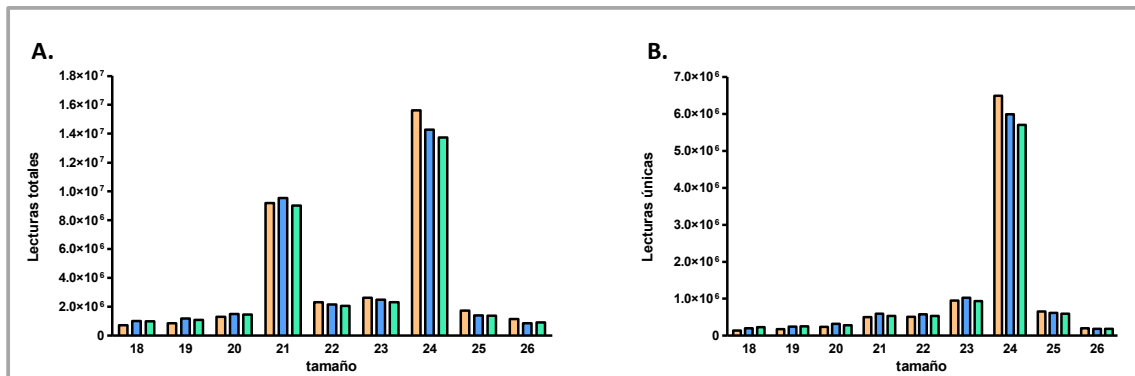
El plegamiento de ácidos nucleicos y la predicción de hibridación se analizaron a través de mfold (<http://unafold.rna.albany.edu/?q=mfold/rna-folding-form>) usando la herramienta RNA Folding Form, (Mathews et al., 1999; AM., 1999; Barciszewski and Clark, 1999; Rouillard et al., 2003).

Para la predicción de target de miARNs en plantas se usó el programa psRNATarget ([http://plantgrn.noble.org/v1\\_psRNATarget/](http://plantgrn.noble.org/v1_psRNATarget/)): A Plant Small RNA Target Analysis Server (versión 2011) (Dai and Zhao, 2011), o la nueva versión del mismo (actualizada en 2017, <http://plantgrn.noble.org/psRNATarget/>) (Dai et al., 2011; Dai et al., unpublished).

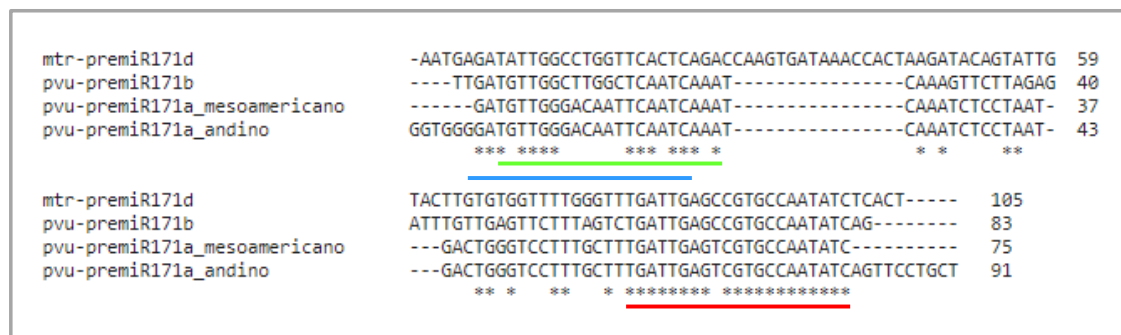
*Anexos*

# Anexos

## 1. Figuras Suplementarias



**Figura S1: Distribución de las lecturas según el tamaño.** Las secuencias totales (A) y únicas (B) (redundantes, y no redundantes respectivamente) libres de adaptadores y filtradas por tamaño y calidad fueron clasificadas según el tamaño (en pares de bases) según cada condición. Las dos replicas biológicas (#1 y #2) fueron graficadas juntas, valores expresadas en número de secuencias.



**Figura S2: Precursores de los miR171a y -b de *P. vulgaris*. y -d de *M. truncatula*.** Los alineamientos de secuencias múltiples se realizaron con Clustal Omega. Los asteriscos representan nucleótidos conservados. En rojo están indicados los miR171 maduros, en verde el mi171a\* y azul el mi171b\* presentes en la data de secuenciación. Los tamaños de los precursores (en nts) están indicados a la derecha. mtr = *M. truncatula*, pvu = *P. vulgaris*.



## 2. Tablas Suplementarias

**Tabla S1: Cantidad de lecturas filtradas de bibliotecas individuales y totales.** Cantidad de secuencias totales (redundantes) y únicas (no redundantes) para cada biblioteca. Los valores en cantidad de lectura y porcentaje corresponden a los valores remanentes después de aplicar cada filtro.

Biblioteca	CTL #1		SC15 #1		55N1 #1		CTL #2		SC15 #2		55N1 #2	
	totales	únicas	totales	únicas	totales	únicas	totales	únicas	totales	únicas	totales	únicas
Lecturas totales	28821434	6961601	32521580	7258370	28569207	6552488	28418340	6697383	29084409	6130675	25053599	6340575
Filtradas por adaptadores	28255351	6309527	31822048	6588184	27946906	5948847	27996576	6086765	28710136	5502044	24697171	5761143
Filtradas por adaptadores (%)	98.0	90.6	97.8	90.8	97.8	90.8	98.5	90.9	98.7	89.7	98.6	90.9
Filtradas por tamaño (18-26)	17295863	4946797	20116552	5445514	17336928	4693967	18169623	4929130	14326001	4303799	15563994	4550323
Filtradas por tamaño (18-26) (%)	60.0	71.1	61.9	75.0	60.7	71.6	63.9	73.6	49.3	70.2	62.1	71.8
Filtradas por calidad	17289792	4941968	20109304	5440363	17331093	4689541	18163139	4924460	14321028	4300042	15558791	4546473
Filtradas por calidad (%)	60.0	71.0	61.8	75.0	60.7	71.6	63.9	73.5	49.2	70.1	62.1	71.7
Filtradas por ARNt y ARNr	14793348	4903371	18257213	5403083	14935595	4651104	16290462	4886391	12389942	4265304	14318672	4513560
Filtradas por ARNt y ARNr (%)	51.3	70.4	56.1	74.4	52.3	71.0	57.3	73.0	42.6	69.6	57.2	71.2

**Tabla S2: miARNs conocidos identificados con el genoma mesoamericano.** Lista de los miARNs de miRBase identificados con miRCat en las dos replicas biológicas para cada condición. Las secuencias están listadas por orden alfabético.

<b>miRBase - Mesoamericano - replicas biológicas</b>		
<b>CTL #1 #2</b>	<b>SC15 #1 #2</b>	<b>55N1 #1 #2</b>
AAGCTCAGGAGGGATAGCACC	AAGCTCAGGAGGGATAGCACC	AAGCTCAGGAGGGATAGCACC
AAGCTCAGGAGGGATAGCGCC	AAGCTCAGGAGGGATAGCGCC	AAGCTCAGGAGGGATAGCGCC
AGAATCTTGATGATGCTGCAT	AGAATCTTGATGATGCTGCAT	AGAATCTTGATGATGCTGCAT
ATGCACTGCCCTCTCCCTGGC	ATGCACTGCCCTCTCCCTGGC	ATGCACTGCCCTCTCCCTGGC
CAGCCAAGAATGACTTGCCGG	CAGCCAAGAATGACTTGCCGG	CAGCCAAGAATGACTTGCCGG
CTTGACTGAAGGGAGCTCCC	CTTGACTGAAGGGAGCTCCC	CTTGACTGAAGGGAGCTCCC
TCATTGAGTGCAGCGTTGATG	TCATTGAGTGCAGCGTTGATG	TCATTGAGTGCAGCGTTGATG
TCATTTTGCCTGCAATGATCTG	TCATTTTGCCTGCAATGATCTG	TCATTTTGCCTGCAATGATCTG
TCGATAAACCTCTGCATCCAG	TCCAAAGGGATCGCATTGATCT	TCCAAAGGGATCGCATTGATCT
TCGCTTGGTGCAGGTCCGGAA	TCGATAAACCTCTGCATCCAG	TCGATAAACCTCTGCATCCAG
TCGGACCAGGCTTCATTCCCC	TCGCTTGGTGCAGGTCCGGAA	TCGCTTGGTGCAGGTCCGGAA
TCTCGGACCAGGCTTCATTCC	TCGGACCAGGCTTCATTCCCC	TCGGACCAGGCTTCATTCCCC
TGAAGCTGCCAGCATGATCT	TCTCGGACCAGGCTTCATTCC	TCTCGGACCAGGCTTCATTCC
TGAAGCTGCCAGCATGATCTG	TGAAGCTGCCAGCATGATCT	TGAAGCTGCCAGCATGATCT
TGAAGCTGCCAGCATGATCTT	TGAAGCTGCCAGCATGATCTG	TGAAGCTGCCAGCATGATCTG
TGACAGAAGAGAGTGAGCAC	TGACAGAAGAGAGTGAGCAC	TGACAGAAGAGAGTGAGCAC
TGATTGAGCCGTGCCAATATC	TGATTGAGCCGTGCCAATATC	TGATTGAGCCGTGCCAATATC
TGATTGAGTCGTGCCAATATC	TGATTGAGTCGTGCCAATATC	TGATTGAGTCGTGCCAATATC
TGCCAAAGGAGAGTTGCCCTG	TGCCAAAGGAGAGTTGCCCTG	TGCCAAAGGAGAGTTGCCCTG
TGCCAAAGGAGATTTGCCCTG	TGCCAAAGGAGATTTGCCCTG	TGCCAAAGGAGATTTGCCCTG
TGCCCTGGCTCCCTGTATGCCA	TGCCCTGGCTCCCTGTATGCCA	TGCCCTGGCTCCCTGTATGCCA
TGGACTGAAGGGAGCTCCTTC	TGGACTGAAGGGAGCTCCTTC	TGGACTGAAGGGAGCTCCTTC
TGGACTGAAGGGGAGCTCCTTC	TGGACTGAAGGGGAGCTCCTTC	TGGACTGAAGGGGAGCTCCTTC
TGGAGAAGCAGGGCACGTGCA	TGGAGAAGCAGGGCACGTGCA	TGGAGAAGCAGGGCACGTGCA
TGTGTTCTCAGGTGCCCCCTG	TGGAGAAGCAGGGCACGTGCA	TGGAGAAGCAGGGCACGTGCA
TTAGATTCACGCACAAACTTG	TGTGTTCTCAGGTGCCCCCTG	TGTGTTCTCAGGTGCCCCCTG
TTATAGTCTGACATCTGGAAT	TTAGATTCACGCACAAACTTG	TTAGATTCACGCACAAACTTG
TTTCAATTTGAAAATAGGCATT	TTATAGTCTGACATCTGGAAT	TTATAGTCTGACATCTGGAAT
TTCCACAGCTTTCTTGAAGCTG	TTTCAATTTGAAAATAGGCATTG	TTTCAATTTGAAAATAGGCATTG
TTCCACAGCTTTCTTGAAGCTT	TTCCACAGCTTTCTTGAAGCTG	TTCCACAGCTTTCTTGAAGCTG
TTGACAGAAGAGAGAGAGCAC	TTCCACAGCTTTCTTGAAGCTT	TTCCACAGCTTTCTTGAAGCTT
TTGACAGAAGATAGAGAGCAC	TTGACAGAAGAGAGAGAGCAC	TTGACAGAAGAGAGAGAGCAC
TTGAGCCGCGTCAATATCTCA	TTGACAGAAGATAGAGAGCAC	TTGACAGAAGATAGAGAGCAC
TTGGACTGAAGGGAGCTCCC	TTGAGCCGCGTCAATATCTCA	TTGACAGAAGATAGAGAGCAC
TTGGCATTCTGTCCACCTCC	TTGGACTGAAGGGAGCTCCC	TTGAGCCGTGCCAATATCACA
TTTGGACTGAAGGGAGCTCCT	TTGGCATTCTGTCCACCTCC	TTGGACTGAAGGGAGCTCCC
TTTGGATTGAAGGGAGCTCTA	TTTGGACTGAAGGGAGCTCCT	TTGGCATTCTGTCCACCTCC
	TTTGGATTGAAGGGAGCTCTA	TTTGGACTGAAGGGAGCTCCT
		TTTGGATTGAAGGGAGCTCTA

**Tabla S3: Nuevos miARNs identificados con el genoma mesoamericano.** Lista de los nuevos miARNs presentando por lo menos una secuencia del miARN\* identificados con miRCat en las dos replicas biológicas para cada condición. Las secuencias están listadas por orden alfabético.

<b>Nuevos - Mesoamericano - replicas biológicas</b>		
<b>CTL #1 #2</b>	<b>SC15 #1 #2</b>	<b>55N1 #1 #2</b>
CATCTGATGAATTGTATCACT	AACCAGGCTCTGATACCATGA	AACCAGGCTCTGATACCATGA
CGGAAGTTGTTTGGGGGTTTC	AGGGCAAGAAGAAGAAGAAGC	CGAGCCGAATCAATATTACTC
CTAATCAAGGAAATCACAGTAG	CATCTGATGAATTGTATCACT	CGGAAGTTGTTTGGGGGTTTC
GGAATGGGCTGATTGGGAAGC	CGAGCCGAATCAATATTACTC	CTAATCAAGGAAATCACAGTAG
GTTGGGTCGATCGGTCCGCCT	CTAATCAAGGAAATCACAGTAG	GAGGAGTGGACTTTTGAGGCCT
TGCTGGAAGTTTTTGAAGGAAT	GAGGAGTGGACTTTTGAGGCCT	GGAATGGGCTGATTGGGAAGC
TGGAGGATTAGGTAAAACAAC	GGAATGGGCTGATTGGGAAGC	TAAGTGAATATTCTTAAAGCCT
	TAAGTGAATATTCTTAAAGCCT	TGCTGGAAGTTTTTGAAGGAAT
	TGCTGGAAGTTTTTGAAGGAAT	TGGAGGATTAGGTAAAACAAC
	TGGAGGATTAGGTAAAACAAC	TTCTTTCAAACAGGCCCTGAG
	TTGCTGTGATGATGAGATCT	

**Tabla S4: miARNs identificados con el genoma mesoamericano.** Valores normalizadas (en RPM) de los miARNs identificados de nuestra data de secuenciación presentes en miRBase. Las secuencias en **negrita** corresponden a los miARNs que fueron graficados en el Capítulo I y los miR390 del Capítulo II. Las secuencias están listadas por orden alfabético de la identificación de miRBase.

miRBase	Secuencia	Tamaño	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2
<b>mir1514a</b>	<b>TTCATTTTAAAAATAGGCATTG</b>	22	371.2	342.5	269.7	660.9	589.9	496.7
mir1515	TCATTTTGCCTGCAATGATCTG	22	15.5	13.9	10.4	15.2	18.0	17.1
mir156	TGACAGAAGAGAGTGAGCAC	20	137.7	218.6	159.8	233.9	258.7	223.9
mir156	TGACAGAAGAGAGAGAGCAC	21	0.3	0.2	0.9	0.6	1.1	0.4
mir156mir157	TTGACAGAAGATAGAGAGCAC	21	632.2	602.3	694.0	601.0	942.7	686.1
mir159	TTTGGATTGAAGGGAGCTCTA	21	3470.3	2728.0	2520.0	2969.5	4374.4	2423.1
mir160	TGCCTGGCTCCCTGAATGCCA	21	0.0	0.1	0.0	0.1	0.2	0.2
mir160	TGCCTGGCTCCCTGTATGCCA	21	1.1	1.2	1.7	3.4	3.8	3.4
mir162	TCGATAAACCTCTGCATCCAG	21	82.3	135.4	128.6	123.7	117.8	144.4
mir164	TGGAGAAGCAGGGCACGTGC	20	0.1	0.0	0.0	0.0	0.0	0.0
mir164	TGGAGAAGCAGGGCACGTGCA	21	2.8	2.8	2.9	2.1	4.9	2.2
mir164	TGGAGAAGCAGGGCACGTGCG	21	0.0	0.0	0.0	0.0	0.1	0.0
mir166	TCGGACCAGGCTTCATTCCCC	21	22707.0	32628.0	26946.1	33832.1	28523.3	46775.4
mir166	TCTCGGACCAGGCTTCATTCC	21	75936.4	118393.2	102475.5	130740.3	81324.8	115327.2
<b>mir167c-d iso</b>	<b>TGAAGCTGCCAGCATGATCT</b>	20	5.1	11.0	7.8	8.8	12.8	12.2
<b>mir167c-d</b>	<b>TGAAGCTGCCAGCATGATCTG</b>	21	76.5	102.1	52.1	111.0	121.6	97.9
mir167	TGAAGCTGCCAGCATGATCTT	21	12.4	0.0	8.2	12.5	17.3	0.0
mir168	TCGCTTGGTGCCAGGTCGGGAA	21	675.0	627.6	786.0	674.9	638.7	617.6
mir169	AGCCAAGGATGACTTGCCGG	20	0.1	0.0	0.0	0.0	0.0	0.0
mir169	CAGCCAAGAATGACTTGCCGG	21	3.2	1.5	2.2	3.4	3.8	2.5
mir169	TAGCCAAGGATGACTTGCCCT	20	0.0	0.1	0.1	0.0	0.0	0.0
mir170mir171	TGATTGAGCCGTGTCAATATC	21	0.1	0.0	0.1	0.0	0.2	0.0
mir171b	TGATTGAGCCGTGCCAATATC	21	0.3	0.5	0.6	0.5	0.4	0.6
mir171k	TTGAGCCGCGCCAATATCACT	21	0.0	13.1	0.0	0.0	0.0	0.0
mir171m-t	TTGAGCCGCGTCAATATCTCA	21	2.0	2.7	0.0	0.7	1.6	1.1
mir171f-iso	TTGAGCCGTGCCAATATCAC	20	0.0	0.1	0.0	0.0	0.0	0.0
mir171q	TTGAGCCGTGCCAATATCACA	21	0.0	1.5	1.1	0.0	0.0	0.6
<b>mir171a</b>	<b>TGATTGAGTCGTGCCAATATC</b>	21	20.8	40.8	33.3	13.8	40.0	27.0
mir172	AGAATCTTGATGATGCTGCA	20	0.0	0.0	0.1	0.0	0.2	0.0
mir172	AGAATCTTGATGATGCTGCAT	21	1.1	1.7	1.1	0.7	0.9	0.7
mir2118mir2218	TTGCCGATCCACCCATTCCCTA	22	0.0	0.0	0.0	0.0	36.2	0.0
mir319	CTTGACTGAAGGGAGCTCCC	21	856.9	1048.0	1216.7	1381.2	1395.4	1250.5
mir319	TGGACTGAAGGGAGCTCCTTC	21	57.4	149.6	192.8	258.9	167.9	157.2
mir319	TGGACTGAAGGGAGCTCCTTC	22	31.8	68.1	65.2	95.7	61.3	58.9
mir319	TTGGACTGAAGGGAGCTCCC	20	411.6	756.2	1054.5	1010.2	871.4	823.5
mir319	TTTGGACTGAAGGGAGCTCCT	21	15.5	26.6	21.2	31.5	39.1	25.2
<b>mir390a</b>	<b>AAGCTCAGGAGGGATAGCGCC</b>	21	20.4	19.9	22.6	20.5	23.5	19.8
<b>mir390b</b>	<b>AAGCTCAGGAGGGATAGCACC</b>	21	16.4	102.3	50.4	32.2	42.0	42.8
mir393	TCCAAAGGGATCGCATTGATC	21	0.0	18.6	18.1	0.0	0.0	0.0
mir393	TCCAAAGGGATCGCATTGATCT	22	0.0	60.5	49.1	55.6	54.8	53.2
mir393	TTCCAAAGGGATCGCATTGATC	22	49.6	0.0	0.0	0.0	0.0	0.0
mir394mir384	TTGGCATTCTGTCCACCTCC	20	1.1	2.5	3.3	4.7	7.4	3.8
mir395	CTGAAGTGTTTGGGGAACTC	21	0.0	0.0	0.2	0.0	0.0	0.0
mir395	TGAAGTGTTTGGGGAACTC	20	0.0	0.0	0.2	0.1	0.0	0.0
mir395	TGAAGTGTTTGGGGAACTTT	21	0.0	0.0	0.0	0.1	0.0	0.0
mir396	TTCCACAGCTTCTTGAAGCTG	21	1543.5	1736.0	2252.8	1929.0	1778.2	1859.5
mir396	TTCCACAGCTTCTTGAAGCTT	21	584.4	737.6	446.0	746.8	662.1	920.7
mir397	TCATTGAGTGCAGCGTTGATG	21	2.0	1.0	1.7	2.2	1.9	2.5
mir398	TGTGTTCTCAGGTCGCCCTG	21	11.2	14.0	15.7	15.7	0.0	28.4
mir399	TGCCAAAGGAGAGACTGCCCTG	21	0.0	0.0	0.1	0.0	0.0	0.3
mir399	TGCCAAAGGAGAGTTGCCCTG	21	6.2	5.6	7.4	7.2	7.7	6.6
mir399	TGCCAAAGGAGATTGCCCCAG	21	0.0	0.1	0.0	0.1	0.0	0.0
mir399	TGCCAAAGGAGATTGCCCTG	21	0.3	0.1	0.2	0.1	0.2	0.2
mir403	TTAGATTCACGCACAAACTTG	21	110.5	163.2	117.8	173.4	155.6	145.3
mir408	ATGCACTGCCTCTTCCCTGGC	21	7.6	4.8	6.9	8.6	6.4	12.5
<b>mir5374-5p</b>	<b>TTATAGTCTGACATCTGGAAT</b>	21	233.6	411.9	396.2	334.0	452.1	425.0

**Tabla S5: Nuevos miARNs identificados con el genoma mesoamericano.** Valores normalizadas (en RPM) de los nuevos miARNs presentando por lo menos una secuencia del miARN\* identificados de nuestra data de secuenciación. Las secuencias en negrita corresponden a dos miARNs del genoma mesoamericano que fueron graficados en el Capítulo 1 (n° 1 y 30), y las dos isoformas del miR5924 (n° 4 y 10) que fue estudiado en el Capítulo III. Las secuencias están listadas por orden alfabético.

	Secuencia	Tamaño	CTL #1	SC15 #1	55N1 #1	CTL #2	SC15 #2	55N1 #2
1	<b>AACCAGGCTCTGATACCATGA</b>	21	956.2	1324.6	1213.7	960.0	655.6	1074.6
2	ACCATGTGATGAATTCGGATT	21	1.0	0.9	1.8	0.0	133.1	4.1
3	AGCCAAGGGTGATTTGCCGGC	21	0.0	0.0	0.0	11.7	9.2	10.3
4	<b>AGGGCAAGAAGAAGAAGC</b>	21	1.6	2.6	0.0	3.0	3.5	2.7
5	AGTTCCTCTGAGCACTTCACA	21	0.0	0.0	0.0	0.2	0.0	0.1
6	<b>CAAGGACTATGTGATCTTGGTC</b>	22	1.6	2.7	1.4	2.0	2.1	1.1
7	CAATGAGGGCATGTTGTAGGC	21	886.0	959.9	876.3	1057.9	987.3	941.1
8	CACCGGTCGGGTCGTCCCTTC	21	0.0	0.0	0.0	45.2	0.0	37.5
9	CAGCCAAGGGTGATTTGCCGGC	22	9.3	12.6	8.8	0.0	0.0	0.0
10	<b>CAGGGCAAGAAGAAGAAGC</b>	22	0.0	0.0	0.4	0.0	0.0	0.0
11	CATCTGATGAATTTGTATCACT	21	1.8	4.5	2.1	5.8	2.9	3.4
12	CATGATTAATTTTCGGATTACA	22	0.3	0.3	0.2	0.4	0.5	0.0
13	CATGGTGGTTAAGATACTCA	20	0.0	0.0	0.0	0.1	0.0	0.0
14	CCACCCACGAATATCTAGATC	21	0.0	0.0	0.0	0.0	0.0	2.5
15	CGACTCCCGGTAGGACCTTTT	21	0.1	0.2	0.0	0.0	0.0	0.0
16	CGAGCCGAATCAATATTACTC	21	0.0	256.5	92.5	210.5	184.0	190.5
17	CGGAAGTTGTTTGGGGGTTTC	20	8.2	7.8	4.8	6.9	0.0	8.9
18	CGGTCGGCAGCGGAGGTCGGC	21	0.0	0.0	0.1	0.0	0.0	0.0
19	CTAATCAAGGAAATCACAGTAG	22	81.4	72.7	56.0	65.0	56.6	49.8
20	CTAGTCTAGTTTGTGGGTCT	21	0.2	0.4	0.2	0.4	0.6	0.5
21	CTCAACTCGAGCCTGAGGACT	21	0.0	0.0	0.0	0.0	0.2	0.0
22	GAGGAGTGGACTTTTGAGGCCT	22	23.6	19.3	15.1	14.4	17.8	21.1
23	GATAGCCACGTGGACTGTTCT	21	0.0	0.0	0.0	0.0	0.1	0.0
24	GGAATGGGCTGATTGGGAAGC	21	65792.0	83458.8	65620.8	68654.2	64507.3	59316.6
25	GGGATATTGCAGGATATGACT	21	0.0	0.0	0.0	0.1	0.0	0.0
26	GGGTTACCTGAGCTCAGAGCCT	22	0.1	0.0	0.0	0.0	0.0	0.0
27	GGTTGTCTAAATTGGAGGGTA	21	0.5	0.0	0.0	0.0	0.0	0.0
28	GTACGAGTATCGGTGTCCGAC	21	0.2	0.0	1.0	0.0	0.0	0.0
29	GTAGGTCGATTCGGTGTCTTA	21	0.0	0.0	0.0	0.0	0.1	0.0
30	<b>GTTGGGTCGATCGGTCCGCCT</b>	21	115.0	157.4	171.3	147.1	196.3	178.7
31	TAAGTGAATATTTCTTAAAGCCT	22	981.9	2097.3	1397.5	2175.7	1796.5	1505.3
32	TAATACTTTCCTTTTGGGCTT	21	0.0	0.0	0.5	0.9	0.0	0.0
33	TAATCTGCATCCTGAGGTTT	20	13.9	12.4	13.1	6.8	2.9	4.9
34	TACGGGTCGCTCTCACCTAGG	21	0.3	0.2	0.2	0.2	0.0	0.0
35	TCAATTAAGAAAGCTGTAACC	21	0.0	0.0	0.0	0.0	0.3	0.0
36	TCACAGGGTTGTCTGACAGACT	22	0.0	27.7	0.0	0.0	0.0	0.0
37	TCATAGTTACACTGATAGAG	20	0.0	0.4	0.0	0.0	0.0	0.0
38	TCATGTGATGAAATGGATA	20	0.0	0.0	0.0	0.0	49.4	0.7
39	TGAAGTGTTTGGGGGAACCTCT	21	0.2	0.3	0.7	0.4	0.9	0.3
40	TGATTTACAAGTGGCGCTATT	21	0.1	0.0	0.0	0.0	0.0	0.0
41	TGCTGGAAGTTTTTGAAGGAAT	22	1.1	0.9	0.7	1.7	0.4	1.3
42	TGCTGTAGAATCGATTTTCCTA	22	0.0	0.1	0.0	0.0	0.0	0.0
43	TGGAGGATTAGGTAAAACAAC	21	294.7	327.5	312.5	374.9	346.9	245.9
44	TTCGGCGGCGCTTTTCTGCCG	21	0.0	0.0	0.3	0.1	0.0	0.0
45	TTCTCTGTAAGCTAAAATTAG	21	0.0	0.0	0.0	0.3	0.0	0.0
46	TTCTTTCAAACAGGCCCTGAG	21	0.0	0.0	11.6	0.3	0.0	0.0
47	TTGAGCCCGCTCAATATCTC	20	0.0	0.0	1.2	0.0	0.0	0.0
48	TTGCTGTGATGATGAGATCT	20	0.0	0.9	0.0	0.0	1.0	0.9

**Tabla S6: miARNs identificados con el genoma mesoamericano.** Valores normalizadas del promedio de las dos replicas biológicas (en RPM) y *fold change* del logaritmo en base 2 de los miARNs identificados de nuestra data de secuenciación presentes en miRBase. Los valores nulos de un cociente están representados por el infinito (-inf o inf cuando el dividendo o el divisor es igual a 0 respectivamente). Las secuencias están listadas por orden alfabético de la identificación de miRBase.

miRBase	Secuencia	CTL	SC15	55N1	CTL vs SC15	CTL vs 55N1	55N1 vs SC15
mir1514a	TTCATTTTGAAAATAGGCATTG	516.1	466.2	383.2	-0.1	-0.4	-0.3
mir1515	TCATTTTGCGTGCAATGATCTG	15.3	15.9	13.7	0.1	-0.2	-0.2
mir156	TGACAGAAGAGAGTGAGCAC	185.8	238.6	191.8	0.4	0.0	-0.3
mir156	TTGACAGAAGAGAGAGAGCAC	0.4	0.6	0.6	0.5	0.5	0.0
mir156mir157	TTGACAGAAGATAGAGAGCAC	616.6	772.5	690.0	0.3	0.2	-0.2
mir159	TTTGGATTGAAGGGAGCTCTA	3219.9	3551.2	2471.5	0.1	-0.4	-0.5
mir160	TGCCTGGCTCCCTGAATGCCA	0.0	0.1	0.1	2.1	1.8	-0.4
mir160	TGCCTGGCTCCCTGTATGCCA	2.3	2.5	2.6	0.1	0.2	0.0
mir162	TCGATAAACCTCTGCATCCAG	103.0	126.6	136.5	0.3	0.4	0.1
mir164	TGGAGAAGCAGGGCACGTGC	0.1	0.0	0.0	-inf	-inf	inf
mir164	TGGAGAAGCAGGGCACGTGCA	2.5	3.9	2.6	0.6	0.0	-0.6
mir164	TGGAGAAGCAGGGCACGTGCG	0.0	0.0	0.0	inf	inf	-inf
mir166	TCGGACCAGGCTTCATTCC	28269.5	30575.6	36860.8	0.1	0.4	0.3
mir166	TCTCGGACCAGGCTTCATTCC	103338.4	99859.0	108901.4	0.0	0.1	0.1
mir167c-d iso	TGAAGCTGCCAGCATGATCT	7.0	11.9	10.0	0.8	0.5	-0.3
mir167c-d	TGAAGCTGCCAGCATGATCTG	93.7	111.8	75.0	0.3	-0.3	-0.6
mir167	TGAAGCTGCCAGCATGATCTT	12.5	8.6	4.1	-0.5	-1.6	-1.1
mir168	TCGCTTGGTGCAGGTCGGGAA	675.0	633.2	701.8	-0.1	0.1	0.1
mir169	AGCCAAGGATGACTTGCCGG	0.1	0.0	0.0	-inf	-inf	inf
mir169	CAGCCAAGAATGACTTGCCGG	3.3	2.7	2.4	-0.3	-0.5	-0.2
mir169	TAGCCAAGGATGACTTGCCCT	0.0	0.0	0.0	inf	inf	0.3
mir170mir171	TGATTGAGCCGTGTCAATATC	0.0	0.1	0.1	1.3	1.0	-0.3
mir171	TGATTGAGCCGTGCCAATATC	0.4	0.4	0.6	0.2	0.7	0.5
mir171	TTGAGCCCGCCCAATATCACT	0.0	6.6	0.0	inf	inf	-inf
mir171	TTGAGCCCGGTCAATATCTCA	1.4	2.2	0.6	0.7	-1.3	-2.0
mir171	TTGAGCCGTGCCAATATCAC	0.0	0.0	0.0	inf	inf	-inf
mir171	TTGAGCCGTGCCAATATCACA	0.0	0.7	0.9	inf	inf	0.3
mir171a	TGATTGAGTCGTGCCAATATC	17.3	40.4	30.2	1.2	0.8	-0.4
mir172	AGAATCTTGATGATGCTGCA	0.0	0.1	0.1	inf	inf	-0.3
mir172	AGAATCTTGATGATGCTGCAT	0.9	1.3	0.9	0.5	0.0	-0.5
mir2118mir2218	TTGCCGATTCCACCCATTCCTA	0.0	18.1	0.0	inf	inf	-inf
mir319	CTTGACTGAAGGGAGCTCC	1119.1	1221.7	1233.6	0.1	0.1	0.0
mir319	TGGACTGAAGGGAGCTCCTTC	158.2	158.8	175.0	0.0	0.1	0.1
mir319	TGGACTGAAGGGAGCTCCTTC	63.7	64.7	62.0	0.0	0.0	-0.1
mir319	TTGGACTGAAGGGAGCTCC	710.9	813.8	939.0	0.2	0.4	0.2
mir319	TTTGGACTGAAGGGAGCTCCT	23.5	32.8	23.2	0.5	0.0	-0.5
mir390a	AAGCTCAGGAGGGATAGCGCC	20.5	21.7	21.2	0.1	0.1	0.0
mir390b	AAGCTCAGGAGGGATAGCACC	24.3	72.1	46.6	1.6	0.9	-0.6
mir393	TCCAAAGGATCGCATTGATC	0.0	9.3	9.0	inf	inf	0.0
mir393	TCCAAAGGATCGCATTGATCT	27.8	57.7	51.2	1.1	0.9	-0.2
mir393	TTCCAAAGGATCGCATTGATC	24.8	0.0	0.0	-inf	-inf	inf
mir394mir384	TTGGCATTCTGTCCACCTCC	2.9	4.9	3.6	0.8	0.3	-0.5
mir395	CTGAAGTGTGGGGGAACCTC	0.0	0.0	0.1	inf	inf	inf
mir395	TGAAGTGTGGGGGAACCTC	0.1	0.0	0.1	-inf	0.7	inf
mir395	TGAAGTGTGGGGGAACCTTT	0.0	0.0	0.0	-inf	-inf	inf
mir396	TTCCACAGCTTTCTTGAACCTG	1736.2	1757.1	2056.1	0.0	0.2	0.2
mir396	TTCCACAGCTTTCTTGAACCTT	665.6	699.9	683.3	0.1	0.0	0.0
mir397	TCATTGAGTGCAGCGTTGATG	2.1	1.4	2.1	-0.5	0.0	0.6
mir398	TGTGTTCTCAGGTCGCCCTG	13.5	7.0	22.0	-0.9	0.7	1.7
mir399	TGCCAAAGGAGAGCTGCCCTG	0.0	0.0	0.2	inf	inf	inf
mir399	TGCCAAAGGAGAGTTGCCCTG	6.7	6.7	7.0	0.0	0.1	0.1
mir399	TGCCAAAGGAGATTGCCCCAG	0.0	0.0	0.0	-0.2	-inf	-inf
mir399	TGCCAAAGGAGATTGCCCTG	0.2	0.1	0.2	-0.4	0.1	0.5
mir403	TTAGATTACGCACAAACTTG	141.9	159.4	131.6	0.2	-0.1	-0.3
mir408	ATGCACTGCCTCTCCCTGGC	8.1	5.6	9.7	-0.5	0.3	0.8
mir5374-5p	TTATAGTCTGACATCTGGAAT	283.8	432.0	410.6	0.6	0.5	-0.1

**Tabla S7: Nuevos miARNs identificados con el genoma mesoamericano.** Valores normalizadas del promedio de las dos replicas biológicas (en RPM) y *fold change* del logaritmo en base 2 de los nuevos miARNs presentando por lo menos una secuencia del miARN\* identificados de nuestra data de secuenciación. Los valores nulos de un cociente están representados por el infinito (-inf o inf cuando el dividendo o el divisor es igual a 0 respectivamente). Los números 4 y 10 corresponden a las dos isoformas del miR5924 y el número 6 a la del mir9624. Las secuencias están listadas por orden alfabético.

	Secuencia	CTL	SC15	55N1	CTL vs SC15	CTL vs 55N1	55N1 vs SC15
1	AACCAGGCTCTGATACCATGA	958.1	990.1	1144.1	0.0	0.3	0.2
2	ACCATGTGATGAATTCGGATT	0.5	67.0	3.0	7.0	2.5	-4.5
3	AGCCAAGGGTGATTTGCCGGC	5.9	4.6	5.1	-0.3	-0.2	0.2
4	AGGGCAAGAAGAAGAAGC	2.3	3.0	1.3	0.4	-0.8	-1.2
5	AGTTCCTCTGAGCACTTCACA	0.1	0.0	0.0	-inf	-1.8	inf
6	CAAGGACTATGTGATCTTGGTC	1.8	2.4	1.3	0.4	-0.5	-0.9
7	CAATGAGGGCATGTTGTAGGC	971.9	973.6	908.7	0.0	-0.1	-0.1
8	CACCGGTCGGGTCGTCCCTTC	22.6	0.0	18.8	-inf	-0.3	inf
9	CAGCCAAGGGTGATTTGCCGGC	4.7	6.3	4.4	0.4	-0.1	-0.5
10	CAGGGCAAGAAGAAGAAGC	0.0	0.0	0.2	inf	inf	inf
11	CATCTGATGAATTTGTACTACT	3.8	3.7	2.7	0.0	-0.5	-0.4
12	CATGATTAATTTTCGGATTACA	0.4	0.4	0.1	0.0	-1.9	-1.9
13	CATGGTGGTTAAGATACTCA	0.0	0.0	0.0	-inf	-inf	inf
14	CCACCCACGAATATCTAGATC	0.0	0.0	1.3	inf	inf	inf
15	CGACTCCCGGTAGGACCTTTT	0.1	0.1	0.0	0.7	-inf	-inf
16	CGAGCCGAATCAATATTACTC	105.2	220.3	141.5	1.1	0.4	-0.6
17	CGGAAGTTGTTTGGGGGTTT	7.6	3.9	6.8	-0.9	-0.1	0.8
18	CGGTCGGCAGCGGAGGTCGGC	0.0	0.0	0.1	inf	inf	inf
19	CTAATCAAGGAAATCACAGTAG	73.2	64.6	52.9	-0.2	-0.5	-0.3
20	CTAGTCTAGTTTGTGGGTCT	0.3	0.5	0.3	0.7	0.1	-0.5
21	CTCAACTCGAGCCTGAGGACT	0.0	0.1	0.0	inf	inf	-inf
22	GAGGAGTGGACTTTTGAGGCCT	19.0	18.5	18.1	0.0	-0.1	0.0
23	GATAGCCACGTGGACTGTTCT	0.0	0.0	0.0	inf	inf	-inf
24	GGAATGGGCTGATTGGGAAGC	67223.1	73983.1	62468.7	0.1	-0.1	-0.2
25	GGGATATTGCAGGATATGACT	0.0	0.0	0.0	-inf	-inf	inf
26	GGGTACCTGAGCTCAGAGCCT	0.0	0.0	0.0	-inf	-inf	inf
27	GGTTGTCTAAATTGGAGGGTA	0.2	0.0	0.0	-inf	-inf	inf
28	GTACGAGTATCGGTGTCCGAC	0.1	0.0	0.5	-inf	2.3	inf
29	GTAGGTCGATTCGGTGTCCCTA	0.0	0.0	0.0	inf	inf	-inf
30	GTTGGGTCGATCGGTCCGCCT	131.1	176.9	175.0	0.4	0.4	0.0
31	TAACTGAATATCTTAAAGCCT	1578.8	1946.9	1451.4	0.3	-0.1	-0.4
32	TAATACTTTCCTTTTGGGCTT	0.4	0.0	0.3	-inf	-0.7	inf
33	TAATCTGCATCCTGAGGTTT	10.3	7.7	9.0	-0.4	-0.2	0.2
34	TACGGGTCGCTCTCACCTAGG	0.3	0.1	0.1	-1.7	-1.4	0.3
35	TCAATTAAGAAAGCTGTAACC	0.0	0.2	0.0	inf	inf	-inf
36	TCACAGGGTTGCTGACAGACT	0.0	13.8	0.0	inf	inf	-inf
37	TCATAGTTACACTGATAGAG	0.0	0.2	0.0	inf	inf	-inf
38	TCATGTGATGAAATTGGATA	0.0	24.7	0.3	inf	inf	-6.1
39	TGAAGTGTGGGGGAACTCT	0.3	0.6	0.5	1.0	0.7	-0.3
40	TGATTTACAAGTGGGCTATT	0.0	0.0	0.0	-inf	-inf	inf
41	TGCTGGAAGTTTTTGAAGGAAT	1.4	0.7	1.0	-1.1	-0.5	0.6
42	TGCTGTAGAAATCGATTTTCCTA	0.0	0.0	0.0	inf	inf	-inf
43	TGGAGGATTAGGTAAAACAAC	334.8	337.2	279.2	0.0	-0.3	-0.3
44	TTCGGCGGCGCTTTTCGCGG	0.1	0.0	0.1	-inf	1.1	inf
45	TTCTCTGTAAGCTAAAATTAG	0.2	0.0	0.0	-inf	-inf	inf
46	TTCTTTCAAACAGGCCCTGAG	0.2	0.0	5.8	-inf	5.2	inf
47	TTGAGCCCGCTCAATATCTC	0.0	0.0	0.6	inf	inf	inf
48	TTGCTGTGATGATGAGATCT	0.0	1.0	0.5	inf	inf	-1.1

**Tabla S8: phasiARNs identificados con el genoma mesoamericano.** Lista de los phasiARNs en fase identificados con *Phasi Prediction* en las dos replicas biológicas para cada condición. Las secuencias están listadas por orden alfabético.

<b>phasiARNs en fase - Mesoamericano - replicas biológicas</b>		
<b>CTL #1 #2 (25)</b>	<b>SC15 #1 #2 (40)</b>	<b>55N1 #1 #2 (36)</b>
AACCAAACCAGTCAGATCTGC	AAAAGATGAGAAGAATGGGAA	AAAAGATGAGAAGAATGGGAA
AGATCTGACTGGTTTGGTTCC	AACCAAACCAGTCAGATCTGC	AGAGTGAAGAACACGTTTCCT
AGGCTCCACAGTCACTTCGC	AACGAGTTGTGACAAGAATTC	ATGTTGTTTCGTAGGTAATATC
AGTTTTGGCTGAGATTTGAGC	ACAACGAGTTGTGACAAGAAT	ATTCCGAAGGATGTCTTGAAT
ATTCCGAAGGATGTCTTGAAT	AGAGTGAAGAACACGTTTCCT	CAAATGGAATCTGAAACTATT
CAAATGGAATCTGAAACTATT	AGATCTGACTGGTTTGGTTCC	CACGTGTCCAAGTCTGTAAC
CAACATCTCCAAAGCATCCTT	AGTTTCGGTAGAGGATAGCCT	CAGATGGAATCTGAAACCAAG
CAATTGCAGGCTATTGTTGGC	ATTCTTGTCACTCGTTGT	CAGATTCATCTGAAACTACC
CAGATGGAATCTGAAACCAAG	CAAATGGAATCTGAAACTATT	CAGATTCATCTGAAACTATG
CAGATTCATCTGAAACTACC	CAACATCTCCAAAGCATCCTT	CAGATTCATCTGAAACTATT
CAGATTCATCTGAAACTATG	CAATTGCAGGCTATTGTTGGC	CAGATTCATCTTCTGAAACT
CAGATTCATCTTCTGAAACT	CACGTGTCCAAGTCTGTAAC	CCAAGGTTGAATGTGGAGATC
CCAAGGTTGAATGTGGAGATC	CAGATGGAATCTGAAACCAAG	CGGGCCCCACGGTGGGCGCCA
CCAGAAAAACAATCGGTTGTT	CAGATTCATCTGAAACTAAG	CTTGAGCGTCGGAGTGCAATC
CTATTCGAAGGATGTCTTGA	CAGATTCATCTGAAACTACC	GAAGTACTGTGGGAGCCTGC
GAAGTACTGTGGGAGCCTGC	CAGATTCATCTGAAACTATG	GATTGTAGCCTGGAAGATGGG
GAATAGAGAGTCAATAATTGT	CAGATTCATCTGAAACTATT	GCAGGCTCAAGAAAAACTCCC
TAGTTTCAGATGGAATCTGCA	CAGATTCATCTTCTGAAACT	GGTTGGGCAAGAACATACTCG
TAGTTTCAGATGGAATCTGCT	CATAGTTTCAGATGGAATCTG	TAGTTTCAGATGGAATCTGCA
TCAAGACATCCTTCGGAATAG	CGACAAAACAATCGGTTGTTT	TAGTTTCAGATGGAATCTGCT
TCCCTTTGGATGTCGTCCTGC	CTACCAAGAGAACCTTCTTCT	TATGGGGCTCCAGGAAGTAGA
TCGCAGAAACAATCGGTTGTT	CTATCGGGCCGCTATCGGACC	TCAAGACATCCTTCGGAATAG
TCTCCACATTC AACCTTGGCA	CTTGCTGAGTTGGTTAAAGAA	TCATAGTACACATGTTGGACG
TCTTCTTGGCCAGCAACTTGC	GCAGGCTCAAGAAAAACTCCC	TCCAACATGTGTACTATGACA
TTTCAGAAGATGGAATCTGCT	GCCCAAGTCAGTGACCGAACC	TCCCTTTGGATGTCGTCCTGC
	GGTTCGGTCACTGACTTGGGC	TCTAACGATTATCGACCCAGG
	TACTTCCTGGAGCCCCATAGC	TCTCCACATTC AACCTTGGCA
	TAGTTTCAGATGGAATCTGCA	TCTTCTTGGCCAGCAACTTGC
	TAGTTTCAGATGGAATCTGCT	TGGACACGTGGCTCGCATCCA
	TATGGGGCTCCAGGAAGTAGA	TTAGGAAAAAAATCTTCTT
	TCAAACGGATGAGTATTCACC	TTCTGAATCCGTAGGTAAATT
	TCATAGTACACATGTTGGACG	TTGCACTCCGACGCTCAAGTC
	TCCAACATGTGTACTATGACA	TTGCTTCAAAAATTGTTTTTG
	TCTAACGATTATCGACCCAGG	TTTACCTACGGATTCAGAATC
	TCTCCACATTC AACCTTGGCA	TTTCAGAAGATGGAATCTGCT
	TCTTCTTGGCCAGCAACTTGC	TTTTTGGATGATGTTTGGAA
	TGAATACTCATCCGTTTGACC	
	TTGCTTCAAAAATTGTTTTTG	
	TTTCAGAAGATGGAATCTGCT	
	TTTTTGGATGATGTTTGGAA	



**Tabla S9: phasiARNs identificados con el genoma mesoamericano.** Valores normalizadas del promedio de las dos replicas biológicas (en RPM) y *fold change* del logaritmo en base 2 de los phasiARNs identificados en fase en nuestra data de secuenciación, considerando únicamente los que son consistentes en las dos replicas biológicas. Los valores nulos de un cociente están representados por el infinito (-inf o inf cuando el dividendo o el divisor es igual a 0 respectivamente).

phasiARN	Secuencia	CTL	SC15	55N1	CTL vs SC15	CTL vs 55N1	55N1 vs SC15
phasi#1	CAGATTCATCTTCTGAAACT	0.2	0.2	0.3	0.5	1.1	0.6
phasi#2	CAGATGGAATCTGAAACCAAG	0.3	0.3	0.1	0.0	-1.3	-1.3
phasi#3	CAAATGGAATCTGAAACTATT	40.0	45.2	40.1	0.2	0.0	-0.2
phasi#4	CAGATTCATCTGAAACTACC	0.4	0.4	0.5	0.2	0.6	0.4
phasi#5	CAGATTCATCTGAAACTATG	0.6	0.7	0.6	0.2	-0.1	-0.3
phasi#6	TTTCAGAAGATGGAATCTGCT	38.9	39.0	37.0	0.0	-0.1	-0.1
phasi#7	TAGTTTCAGATGGAATCTGCT	33.8	31.6	30.1	-0.1	-0.2	-0.1
phasi#8	TAGTTTCAGATGGAATCTGCA	1.2	1.6	1.7	0.3	0.5	0.1
phasi#9	TCTCCACATTC AACCTTGCA	0.9	0.7	0.5	-0.3	-0.7	-0.4
phasi#10	TCTTCTTGCCAGCAACTTGC	3.8	3.6	5.4	-0.1	0.5	0.6
phasi#11	CACGTGTCCAAGTCTGTA ACT	0.0	0.3	0.2	inf	inf	-0.7
phasi#12	TCTAACGATTATCGACCCAGG	0.0	0.2	0.2	inf	inf	-0.2
phasi#13	GAATAGAGAGTCAATAATTGT	2.5	0.0	0.0	-inf	-inf	inf
phasi#14	ACAACGAGTTGTGACAAGAAT	0.0	0.3	0.0	inf	inf	-inf
phasi#15	CTATCGGGCCGCTATCGGACC	0.0	0.3	0.0	inf	inf	-inf
phasi#16	GGTTGGGCAAGAACATACTCG	0.0	0.0	0.1	inf	inf	inf
phasi#17	TTAGGAAAAAAAAATTCTTCTT	0.0	0.0	0.1	inf	inf	inf
phasi#18	TGGACACGTGGCTCGCATCCA	0.0	0.0	0.2	inf	inf	inf

**Tabla S10: Loci de phasiARNs identificados con el genoma mesoamericano.** Loci comunes de secuencias en fase de phasiARNs entre las dos replicas biológicas y para estos loci, los que fueron compartidos entre cada condición.

Loci de phasiARNs	SC15 #1 + #2 = 13	55N1 #1 + #2 = 13
<b>CTL #1 + #2 = 11</b>	PHASED_Chromo:scaffold00968	PHASED_Chromo:scaffold00012
PHASED_Chromo:scaffold00214	PHASED_Chromo:scaffold00037	PHASED_Chromo:scaffold00968
PHASED_Chromo:contig118694	PHASED_Chromo:scaffold00416	PHASED_Chromo:scaffold00029
PHASED_Chromo:scaffold00968	PHASED_Chromo:scaffold00334	PHASED_Chromo:scaffold00021
PHASED_Chromo:scaffold01370	PHASED_Chromo:scaffold00537	PHASED_Chromo:scaffold00264
PHASED_Chromo:scaffold00080	PHASED_Chromo:scaffold00019	PHASED_Chromo:contig125036
PHASED_Chromo:scaffold00346	PHASED_Chromo:scaffold00259	PHASED_Chromo:scaffold00674
PHASED_Chromo:scaffold00103	PHASED_Chromo:scaffold00012	PHASED_Chromo:contig88356
PHASED_Chromo:scaffold00587	PHASED_Chromo:scaffold00070	PHASED_Chromo:scaffold00448
PHASED_Chromo:scaffold00537	PHASED_Chromo:scaffold00029	PHASED_Chromo:scaffold00048
PHASED_Chromo:scaffold00019	PHASED_Chromo:scaffold00310	PHASED_Chromo:scaffold01490
PHASED_Chromo:scaffold00910	PHASED_Chromo:contig125036	PHASED_Chromo:scaffold01180
<b>SC15 + 55N1 = 3</b>	PHASED_Chromo:scaffold00010	PHASED_Chromo:scaffold00019
PHASED_Chromo:scaffold00029	<b>CTL+SC15 = 1</b>	<b>CTL+SC15+55N1 = 2</b>
PHASED_Chromo:scaffold00012	PHASED_Chromo:scaffold00537	PHASED_Chromo:scaffold00019
PHASED_Chromo:contig125036	<b>CTL+55N1 = 0</b>	PHASED_Chromo:scaffold00968

**Tabla S11: Datos de secuenciación de las 3 isoformas del miR5924 de bibliotecas de sARNs.** Las isoformas de 20, 21 y 22 nts fueron secuenciadas en bibliotecas de *P. vulgaris* de diferentes tejidos (Formey et al., 2015). La isoforma de 21 nts fue también secuenciada en las bibliotecas de pelos radicales de Formey et al., 2016 (valores en negrita). Los valores controles (CTL) corresponde a pelos radicales sin inocular y NF con tratamiento por factor Nod a 6 hpi. Los valores (*E*) y (*UPE*) indicados corresponden a la probabilidad máxima (*Maximum expectation*) y la accesibilidad del mensajero al miARN (*Target accessibility*) obtenidos en psRNATarget.

tamaño	20 nts	21 nts	22 nts
secuencia	GGGCAAGAAGAAGAAGAAC	AGGGCAAGAAGAAGAAGAAC	CAGGGCAAGAAGAAGAAGAAC
nombre	miRNov248	miRNov250	miRNov246
miRdeep score	19.8	8.8 / <b>15.4</b>	84.3
miRanda	25	25	24
miARN*	si	si	si
bibliotecas (expresión sin normalizar)	flores (23) plántulas (3) raíces (21) nódulos (7) hojas (7)	flores (48) plántulas (8) raíces (2) nódulos (1) hojas (2) <b>pelos radicales CTL (8) NF (6)</b>	flores (55) plántula (0) raíces (5) nódulos (13) hojas (6)
blanco putativo anotación	<i>Phvul.006G188700.1</i> disease resistance protein (TIR class), putative	<i>Phvul.006G188700.1</i> disease resistance protein (TIR class), putative	<i>Phvul.006G188700.1</i> disease resistance protein (TIR class), putative
psRNA target ( <i>E</i> / <i>UPE</i> )	2 / 3.327	<b>1 / 4.944</b>	2 / 4.929
inhibición	clivaje	clivaje	clivaje
degradoma ( <i>Pvalue</i> )	no	no	no
blanco putativo anotación	<i>Phvul.007G259200.1</i> SNF2 domain-containing protein / helicase domain-containing protein	<i>Phvul.005G066000.2</i> D-mannose binding lectin protein with Apple-like carbohydrate-binding domain	<i>Phvul.003G125700.1</i> jasmonate-zim-domain protein 12
psRNA target ( <i>E</i> / <i>UPE</i> )	2 / 0.951	no	no
inhibición	clivaje	clivaje	clivaje
degradoma ( <i>Pvalue</i> )	no	<b>0.1397405</b>	0.0363209993406077

### 3. Secuencias de miARN, miARN\* y precursores

Las secuencias en formato *fasta* son listadas según el orden de aparición en el manuscrito.

#### >pvu-miRNov1

AACCAGGCTCTGATACCATGA

#### >pvu-miRNov1\*

TCGTGGTATCAGGTCCTGCTT

#### >pvu-premiRNov1

TCGTGGTATCAGGTCCTGCTTCACCAAATGATCTTGAGTTCAACTCCTACACAAGCACTTGGTTAACC  
AGGCTCTGATACCATGA

#### >pvu-miRNov30

GTTGGGTCGATCGGTCCGCCT

#### >pvu-miRNov30\*

CGGGTCCGATGTCTGACGTCC

#### >pvu-premiRNov30

Anexos

GTTGGACCTTGGGTTGGGTCCGATCGGTCCGCCTCTGGTGTGCACCGGTCCGGTCCGATGTCTGACG  
TCCGATGTCTGAGGTCCGAT

**>pvu-miR167c-d (21 nts)**  
TGAAGCTGCCAGCATGATCTG

**>pvu-miR167c-d\***  
GATCATGTGGCTGCTTCACC

**>pvu-premiR167c-d**  
CAGCAGTTGAAGCTGCCAGCATGATCTGAGTTTACCTTCTATAATGGTAAAAACAGATCATGTGGCTG  
CTTCACCTGTTG

**>pvu-miR167c-d (20 nts)**  
TGAAGCTGCCAGCATGATCT

**>pvu-premiR167c-d (20 nts)**  
AGGAAAAAGTGAAGCTGCCAGCATGATCTAGCTTTGGTTAGTGTGAGCCACTGAGGTGCTAACCCCTTA  
CTAGGTCATGCTGTGCTAGCCTCACTCCTTCCT

**>pvu-miR5374**  
TTATAGTCTGACATCTGGAAT

**>pvu-miR5374\***  
TTCGAATGTCAGATTATAAAA

**>pvu-premiR5374**  
TTGAAGTCATCGAAATTTTATAGTCTGACATCTGGAATTTAAATTTCTCAACAAGGGCATCAGATCCAA  
AACCTGCATCTTCATCGCTTTGAAACAGAAAGGGCACTCCAGATCCACGTGCAGAGAGAAGCTCTTGT  
TGGGAAATTTGAATTTCAATGTCAGATTATAAAAACATTGGTGATTTCAA

**>pvu-miR1514a**  
TTCATTTTAAAAATAGGCATTG

**>pvu-miR1514a\***  
TGCCTATTTTAAGATGAAAA

**>pvu-premiR1514a (chr3)**  
TCCTGTTTTCATTTTAAAAATAGGCATTGCATTACATTTTCGATGCGTCTTCCAATTCGAATGCCTATTTT  
AAGATGAAAAACAAGGA

**>pvu-miR1514a\***  
TGCCTATTTCAAGATGAAAA

**>pvu-premiR1514a (chr7)**  
GTTTTCATTTTAAAAATAGGCATTGCATAATTGCATTGCATTTTCGATGCATCTTCCAATTCGAATGCCTA  
TTTCAAGATGAAAAAC

**>pvu-miR171a**  
TGATTGAGTCGTGCCAATATC

**>pvu-miR171a\***  
TGTTGGGACAATTCAATCAAA

**>pvu-premiR171a (mesoamericano)**  
GATGTTGGGACAATTCAATCAAAATCAAATCTCCTAATGACTGGGTCTTTGCTTTGATTGAGTCGTGCC  
AATATC

**>pvu-premiR171a (andino)**  
GGTGGGATGTTGGGACAATTCAATCAAAATCAAATCTCCTAATGACTGGGTCTTTGCTTTGATTGAG  
TCGTGCCAATATCAGTTCTGCT

**>pvu-miR171b**  
TGATTGAGCCGTGCCAATATC

**>pvu-miR171b\***  
TGATGTTGGCTTGGCTCAATC

**>pvu-premiR171b**  
TTGATGTTGGCTTGGCTCAATCAAAATCAAAGTTCTTAGAGATTTGTTGAGTTCTTTAGTCTGATTGAGC  
CGTGCCAATATCAG

**>mtr-miR171a**  
TGATTGAGTCGTGCCAATATC

**>mtr-premiR171a**  
TGAATTCCTCCGCTTTTGGATGTTGGCTTGTCTCAATCAAAATCAAAGTTCTTGAAATTTGAGTTCTTT  
AGTCTGATTGAGTCGTGCCAATATCATATTAAGCGATAAAAAGTC

**>mtr-miR171d**  
TGATTGAGCCGTGCCAATATC

**>mtr-premiR171d**

Anexos

AATGAGATATTGGCCTGGTTCACTCAGACCAAGTGATAAACCACTAAGATACAGTATTGTACTTGTGTG  
GTTTTGGGTTTGATTGAGCCGTGCCAATATCTCACT

**>mtr-miR171h**  
CGAGCCGAATCAATATCACTC

**>lja-miR171c**  
TGAGCCGAATCAATATCACTC

**>gma-miR171x**  
TGATTGAGCCGTGCCAATATC

**>gma-premiR171x**  
GATGTTGGACGGTTCAATCAAATCAAATCTCCTAATGGCTGGGTCTTTGGTATGATTGAGCCGTGCC  
AATATC

**>pvu-miR390a**  
AAGCTCAGGAGGGATAGCGCC

**>pvu-miR390a\***  
CGCTATCCATCCTGAGTTTC

**>pvu-premiR390a (chr10)**  
GTGTAGAAGAATCTGTAAAGCTCAGGAGGGATAGCGCCATGGATGGTCATTGCTCATCTCTCTTCTCT  
TTCTCCCTTTGCCTCTTTGATCTTCTCTTGCCTATCCATCCTGAGTTTCATGGCTTCTTCTACAC

**>pvu-premiR390a (chr6 sin miARN\*)**  
AAGCTCAGGAGGGATAGCGCCATGATTCTCACATCATAGTGTCTTTGGCGCTATCTATCCTGAGTTT

**>pvu-miR390b**  
AAGCTCAGGAGGGATAGCACC

**>pvu-miR390b\***  
CGCTATCTATCTTGAGCTTC

**>pvu-premiR390b**  
GGGAGAATCTGTAAAGCTCAGGAGGGATAGCACCTTGATACTGATGATGCTATACATGTTTCATAGCAG  
TATATACTTGGCGCTATCTATCTTGAGCTTCATGGCTTCTTCT

**>mtr-premiR390a-b**  
AAGCTCAGGAGGGATAGCGCC

**>mtr-premiR390a**  
CTGTAAAGCTCAGGAGGGATAGCGCCATAGAATGTCTTCTTTTTTTGGTTCCATTTTATTGATTGATCT  
TCTCTTGCCTATCCATCCTGAGTTTCATGG

**>mtr-premiR390b**  
GAGTATAGGAGGGTCGGTAAAGCTCAGGAGGGATAGCGCCATTGATAAATGTGTGTACGTGGTATTT  
GGCGCTATCCATCCTGAGTTTCACCGTTCTTCTTACTAGCT

**>pvu-miR5924**  
AGGGCAAGAAGAAGAAGAAGC

**>pvu-miR5924\***  
TCTGTTTCTTCTGTCTCTGT

**>pvu-premiR5924**  
GGATTTTGTCTGTTTCTTCTGTCTCTGTGGCTGTGCTTTTTCTGAGTAAATTAATAAAGCAGCAGCA  
GCAGGGCAAGAAGAAGAAGAAGCAAATTC

## 4. Oligonucleótidos

Tabla S12: Oligonucleótidos (listados por orden alfabético).

Identificación	Secuencia (5' - 3')	Utilización
EF1 $\alpha$ F	AACTACCACCGGCACTTGAT	Real Time
EF1 $\alpha$ R	AGCACCCAGGCATACTTGAAT	Real Time
M13 F	GTAAAACGACGGCCAG	Vector pENTRY
M13 R	CAGGAAACAGCTATGAC	Vector pENTRY
miR171a F	CGATGTGATTGAGACGTGCC	Stem loop PCR
miR171a RT	GTCGTATCCAGTGCCTGTCTGTGGAGTCGGCAATTGCACTGGATACGACGAT ATT	Stem loop RT
miR390b F	TCGCTAAGCTCAGGAGGGAT	Stem loop PCR
miR390b RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTGCACTGGATACGACGGTGCT	Stem loop RT
miR5924 F	GGTCCAGGGCAAGAAGAAGA	Stem loop PCR
miR5924 RT	GTGCGTGTCTGTGGAGTCGGCAATTGCACTGGATACGACGCTTCT	Stem loop RT
NodC F	TGCTTGACACAACCAGCACTG	PCR
NodC R	GACAGCCAGTCGCTATTGCTCA	PCR
NF-YA1 F	TCATTTGGATCTTCTGTGCACAC	Real Time
NF-YA1 R	GCTTCTTCATCTGGTCTCATAAAGG	Real Time
NF-YC1 F	GCAGGGCTATATCGGTCTTTTC	Real Time
NF-YC1 R	GAGTAACATTTTGGATTCAATG	Real Time
pK7WG2D 35S F	GATGACGCACAATCCCCTATCC	Vector OE
pK7WG2D Kan R	CGTAAAACGGCTTGTCCCG	Vector OE
PremiR390b F	CACCGGCTTGGGTCATGGAAGATGGTT	Clonado + Real Time
PremiR390b R	CGCCATTTAGAAGACATGCCTCACAAT	Clonado + Real Time
PremiR5924 F	CACCGGGTGGGGAGTGTTTGTGTT	Clonado + Real Time
PremiR5924 R	GGCAGAGATGTGTGCTCTTTTCAT	Clonado + Real Time
Reerse universal	GTGCAGGGTCCGAGGT	Stem loop PCR
TAS3 F	CCGTTGCCACCCAACACTCAT	Real Time
TAS3 R	CCCCATTGAGGGAGGTTCTAA	Real Time
phasiARFs F	TCGCTTTCTTGACCTTGATA	Stem loop PCR
phasiARFs RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTGCACTGGATACGACAAGGTC	Stem loop RT
U6 snRNA F	GGAACGATACAGAGAAGATTAGCA	Stem loop PCR
U6 snRNA RT	GTGCAGGGTCCGAGGTTTGGACCATTTCTCGAT	Stem loop RT

# *Bibliografía*

# *Bibliografía*

- Aguilar, O.M., Riva, O., and Peltzer, E.** (2004). Analysis of *Rhizobium etli* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. *Proc. Natl. Acad. Sci.* **101**: 13548–13553.
- Allen, E. and Howell, M.D.** (2010). miARNs in the biogenesis of trans-acting siRNAs in higher plants. *Semin. Cell Dev. Biol.* **8**: 798–804.
- Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C.** (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**: 207–221.
- Amarger, N.** (2001). Rhizobia in the Field. *Advances* **73**: 109–168.
- Ambros, V. et al.** (2003). A uniform system for microRNA annotation. *RNA* **9**: 277–279.
- Ameres, S.L. and Zamore, P.D.** (2013). Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14**: 475–488.
- Amor, B. Ben, Shaw, S.L., Oldroyd, G.E.D., Maillet, F., Penmetsa, R.V., Cook, D., Long, S.R., Dénarié, J., and Gough, C.** (2003). The NFP locus of *Medicago truncatula* controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *Plant J.* **34**: 495–506.
- Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D.G., and de Carvalho-Niebel, F.** (2007). AP2-ERF Transcription Factors Mediate Nod Factor Dependent Mt ENOD11 Activation in Root Hairs via a Novel cis-Regulatory Motif. *Plant Cell* **19**: 2866–2885.
- Aravin, A.A., Hannon, G.J., and Brennecke, J.** (2007). The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. *Science* **318**: 761–764.
- Arenas-Huertero, C., Pérez, B., Rabanal, F., Blanco-Melo, D., De La Rosa, C., Estrada-Navarrete, G., Sanchez, F., Covarrubias, A.A., and Reyes, J.L.** (2009). Conserved and novel miRNAs in the legume *Phaseolus vulgaris* in response to stress. *Plant Mol. Biol.* **70**: 385–401.
- Ariel, F. et al.** (2012). Two Direct Targets of Cytokinin Signaling Regulate Symbiotic Nodulation in *Medicago truncatula*. *Plant Cell* **24**: 3838–3852.
- Arikit, S., Xia, R., Kakrana, A., Huang, K., Zhai, J., Yan, Z., Valdés-López, O., Prince, S., Musket, T.A., Nguyen, H.T., Stacey, G., and Meyers, B.C.** (2014). An Atlas of Soybean Small RNAs Identifies Phased siRNAs from Hundreds of Coding Genes. *Plant Cell Online* **26**: 4584–4601.
- Arrighi, J.-F.** (2006). The *Medicago truncatula* Lysine Motif-Receptor-Like Kinase Gene Family Includes NFP and New Nodule-Expressed Genes. *Plant Physiol.* **142**: 265–279.
- Aslam, S.N. et al.** (2008). Bacterial Polysaccharides Suppress Induced Innate Immunity by Calcium Chelation. *Curr. Biol.* **18**: 1078–1083.
- Aukerman, M.J. and Sakai, H.** (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**: 2730–41.
- Axtell, M., Westholm, J., and Lai, E.** (2011). Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* **12**: 221.
- Axtell, M.J.** (2013). Classification and Comparison of Small RNAs from Plants. *Annu. Rev. Plant Biol.* **64**: 137–159.
- Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P.** (2006). A Two-Hit Trigger for siRNA Biogenesis in Plants. *Cell* **127**: 565–577.
- Axtell, M.J. and Meyers, B.C.** (2018). Revisiting criteria for plant miRNA annotation in the era of big data. *Plant Cell: tpc.00851.2017*.
- Axtell, M.J., Snyder, J.A., and Bartel, D.P.** (2007). Common Functions for Diverse Small RNAs of Land Plants. *Plant Cell* **19**: 1750–1769.
- Banba, M., Gutjahr, C., Miyao, A., Hirochika, H., Paszkowski, U., Kouchi, H., and Imaizumi-Anraku, H.** (2008). Divergence of evolutionary ways among common sym genes: CASTOR and CCaMK show functional conservation between two symbiosis systems and constitute the root of a common signaling pathway. *Plant Cell Physiol.* **49**: 1659–1671.
- Bapaume, L. and Reinhardt, D.** (2012). How membranes shape plant symbioses: signaling and transport in nodulation and arbuscular mycorrhiza. *Front. Plant Sci.* **3**: 1–29.
- Barciszewski, J. and Clark, B.F.C.** (1999). RNA biochemistry and biotechnology (Kluwer). pp.

- 347–352.
- Bartel, D.P.** (2009). MicroRNA Target Recognition and Regulatory Functions. *Cell* **136**: 215–233.
- Battaglia, M., Ripodas, C., Clua, J., Baudin, M., Aguilar, O.M., Niebel, A., Zanetti, M.E., and Blanco, F.A.** (2014). A Nuclear Factor Y Interacting Protein of the GRAS Family Is Required for Nodule Organogenesis, Infection Thread Progression, and Lateral Root Growth. *Plant Physiol.* **164**: 1430–1442.
- Baudin, M., Laloum, T., Lepage, A., Ripodas, C., Ariel, F., Frances, L., Crespi, M., Gamas, P., Blanco, F.A., Zanetti, M.E., de Carvalho-Niebel, F., and Niebel, A.** (2015). A phylogenetically conserved group of NF-Y transcription factors interact to control nodulation in legumes. *Plant Physiol.* **33**: pp.01144.2015.
- Becerra-Velásquez, V.L. and Gepts, P.** (1994). RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. *Genome* **37**: 256–263.
- Beckers, M., Mohorianu, I., Stocks, M., Applegate, C., Dalmay, T., and Moulton, V.** (2017). Comprehensive processing of high-throughput small RNA sequencing data including quality checking, normalization, and differential expression analysis using the UEA sRNA Workbench. *RNA* **23**: 823–835.
- Beebe, S., Gonzalez, A.V., and Rengifo, J.** (1999). Improving Human Nutrition Through Agriculture: The Role of International Agricultural Research Research on Trace Minerals in Common Bean. **1996**: 217–226.
- Beebe, S.E., Rao, I.M., Blair, M.W., and Acosta-Gallegos, J.A.** (2013). Phenotyping common beans for adaptation to drought. *Front. Physiol.* **4**: 1–20.
- Bella, J., Hindle, K.L., McEwan, P.A., and Lovell, S.C.** (2008). The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **65**: 2307–2333.
- Bhawna, Bonthala, V.S., and Gajula, M.P.** (2016). PvTFDB: a *Phaseolus vulgaris* transcription factors database for expediting functional genomics in legumes. *Database (Oxford)*. **64**: 217–226.
- Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P.S., Logozzo, G., Stougaard, J., McClean, P., Attene, G., and Papa, R.** (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci.* **109**: 788–796.
- Bitocchi, E., Rau, D., Bellucci, E., Rodriguez, M., Murgia, M.L., Gioia, T., Santo, D., Nanni, L., Attene, G., and Papa, R.** (2017). Beans (*Phaseolus* spp.) as a Model for Understanding Crop Evolution. *Front. Plant Sci.* **8**: 1–21.
- Blair, M.W., Díaz, L.M., Buendía, H.F., and Duque, M.C.** (2009). Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **119**: 955–972.
- Blanco, F.A., Peltzer Meschini, E., Zanetti, M.E., and Aguilar, O.M.** (2009). A Small GTPase of the Rab Family Is Required for Root Hair Formation and Preinfection Stages of the Common Bean-Rhizobium Symbiotic Association. *Plant Cell Online* **21**: 2797–2810.
- Blankenberg, D. et al.** (2010a). Galaxy: A Web-Based Genome Analysis Tool for Experimentalists. In *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc.: Hoboken, NJ, USA), p. 19.10.1-19.10.21.
- Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., and Team, G.** (2010b). Manipulation of FASTQ data with galaxy. *Bioinformatics* **26**: 1783–1785.
- Blankenberg, D. and Hillman-Jackson, J.** (2014). Analysis of Next-Generation Sequencing Data Using Galaxy. In (Humana Press, New York, NY), pp. 21–43.
- Boisson-Dernier, A., Andriankaja, A., Chabaud, M., Niebel, A., Journet, E.-P., Barker, D.G., and de Carvalho-Niebel, F.** (2005). *MtENOD11* Gene Activation During Rhizobial Infection and Mycorrhizal Arbuscule Development Requires a Common AT-Rich-Containing Regulatory Sequence. *Mol. Plant-Microbe Interact.* **18**: 1269–1276.
- Bologna, N.G., Schapire, A.L., and Palatnik, J.F.** (2013). Processing of plant microRNA precursors. *Brief. Funct. Genomics* **12**: 37–45.
- Bologna, N.G. and Voinnet, O.** (2014). The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in *Arabidopsis*. *Annu. Rev. Plant Biol.* **65**: 473–503.
- Bond, J. and Gresshoff, P.** (1993). Soybean transformation to study molecular physiology. *Plant responses to Environ. Gresshoff*: 25–44.
- Bonnet, E., Wuyts, J., Rouzé, P., and Van de Peer, Y.** (2004). Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random



- sequences. *Bioinformatics* **20**: 2911–2917.
- Boualem, A., Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combiér, J.P., Niebel, A., Crespi, M., and Frugier, F.** (2008). MicroRNA166 controls root and nodule development in *Medicago truncatula*. *Plant J.* **54**: 876–887.
- Bouis, H.E.** (2003). Micronutrient fortification of plants through plant breeding: can it improve nutrition in man at low cost? *Proc. Nutr. Soc.* **62**: 403–411.
- Bozsoki, Z., Cheng, J., Feng, F., Gysel, K., Vinther, M., Andersen, K.R., Oldroyd, G., Blaise, M., Radutoiu, S., and Stougaard, J.** (2017). Receptor-mediated chitin perception in legume roots is functionally separable from Nod factor perception. *Proc. Natl. Acad. Sci. U. S. A.* **114**: 8118–8127.
- Brockwell, J., Bottomley, P.J., and Thies, J.E.** (1995). Manipulation of rhizobia microflora for improving legume productivity and soil fertility: A critical assessment. *Plant Soil* **174**: 143–180.
- Broghammer, A. et al.** (2012). Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc. Natl. Acad. Sci.* **109**: 13859–13864.
- Broughton, W.J., Hernández, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J.** (2003). Beans (*Phaseolus* spp.) – model food legumes. *Plant Soil* **252**: 55–128.
- Buermans, H.P.J. and den Dunnen, J.T.** (2014). Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**: 1932–1941.
- Burle, M.L., Fonseca, J.R., Kami, J.A., and Gepts, P.** (2010). Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. *Theor. Appl. Genet.* **121**: 801–813.
- Bustos-Sanmamed, P., Bazin, J., Hartmann, C., Crespi, M., and Lelandais-Brière, C.** (2013). Small RNA pathways and diversity in model legumes: lessons from genomics. *Front. Plant Sci.* **4**: 1–19.
- Cao, Y., Halane, M.K., Gassmann, W., and Stacey, G.** (2017). The Role of Plant Innate Immunity in the Legume-Rhizobium Symbiosis. *Annu. Rev. Plant Biol.* **68**: 535–561.
- Castro-Guerrero, N.A., Isidra-Arellano, M.C., Mendoza-Cozatl, D.G., and Valdés-López, O.** (2016). Common Bean: A Legume Model on the Rise for Unraveling Responses and Adaptations to Iron, Zinc, and Phosphate Deficiencies. *Front. Plant Sci.* **7**: 1–7.
- Catoira, R., Galera, C., de Billy, F., Penmetsa, R. V, Journet, E.P., Maillet, F., Rosenberg, C., Cook, D., Gough, C., and Dénarié, J.** (2000). Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *Plant Cell* **12**: 1647–66.
- Cerri, M.R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G.E.D., Barker, D.G., Fournier, J., and de Carvalho-Niebel, F.** (2012). *Medicago truncatula* ERN Transcription Factors: Regulatory Interplay with NSP1/NSP2 GRAS Factors and Expression Dynamics throughout Rhizobial Infection. *Plant Physiol.* **160**: 2155–2172.
- Chan, P.P. and Lowe, T.M.** (2009). GtRNadb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**: 93–97.
- Chapman, E.J. and Carrington, J.C.** (2007). Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**: 884–896.
- Charpentier, M., Sun, J., Martins, T.V., Radhakrishnan, G. V, Findlay, K., Soumpourou, E., Thouin, J., Véry, A., Sanders, D., Morris, R.J., and Oldroyd, G.E.D.** (2016). Symbiotic Calcium Oscillations. *Science* **352**: 1102–1105.
- Charron, D., Pingret, J.-L., Chabaud, M., Journet, E.-P., and Barker, D.G.** (2004). Pharmacological Evidence That Multiple Phospholipid Signaling Pathways Link Rhizobium Nodulation Factor Perception in *Medicago truncatula* Root Hairs to Intracellular Responses, Including Ca<sup>2+</sup> Spiking and Specific ENOD Gene Expression. *Plant Physiol.* **136**: 3582–3593.
- Chen, C. et al.** (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* **33**: 1–9.
- Chen, C., Fan, C., Gao, M., and Zhu, H.** (2009). Antiquity and function of CASTOR and POLLUX, the twin ion channel-encoding genes key to the evolution of root symbioses in plants. *Plant Physiol.* **149**: 306–17.
- Chen, H.-M., Chen, L.-T., Patel, K., Li, Y.-H., Baulcombe, D.C., and Wu, S.-H.** (2010). 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl. Acad. Sci.* **107**: 15269–15274.
- Chen, H.-M., Li, Y.-H., and Wu, S.-H.** (2007). Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in *Arabidopsis*.

- Proc. Natl. Acad. Sci. **104**: 3318–3323.
- Chen, Y., Sun, S.-K., Tang, Z., Liu, G., Moore, K.L., Maathuis, F.J.M., Miller, A.J., McGrath, S.P., and Zhao, F.-J.** (2017). The Nodulin 26-like intrinsic membrane protein OsNIP3;2 is involved in arsenite uptake by lateral roots in rice. *J. Exp. Bot.* **68**: 3007–3016.
- Cheng, H.P. and Walker, G.C.** (1998). Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by *Rhizobium meliloti*. *J. Bacteriol.* **180**: 5183–91.
- Chi, X., Yang, Q., Chen, X., Wang, J., Pan, L., Chen, M., Yang, Z., He, Y., Liang, X., and Yu, S.** (2011). Identification and Characterization of microRNAs from Peanut (*Arachis hypogaea* L.) by High-Throughput Sequencing. *PLoS One* **6**: e27530.
- Chuck, G., Cigan, A.M., Saetern, K., and Hake, S.** (2007). The heterochronic maize mutant *Corngrass1* results from overexpression of a tandem microRNA. *Nat. Genet.* **39**: 544–549.
- Clúa, J., Roda, C., Zanetti, M.E., and Blanco, F.A.** (2018). Compatibility between legumes and rhizobia for the establishment of a successful nitrogen-fixing symbiosis. *Genes (Basel)*. **9**.
- Combiér, J., Frugier, F., Billy, F. De, Boualem, A., El-yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P., Crespi, M., and Niebel, A.** (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Plant Physiol.* **141**: 3084–3088.
- Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C.** (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **17**: 997–1003.
- D'haene, B., Mestdagh, P., Hellemans, J., and Vandesompele, J.** (2012). miRNA Expression Profiling: From Reference Genes to Global Mean Normalization. In *Methods in molecular biology (Clifton, N.J.)*, pp. 261–272.
- 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., Frugier, F., and Goormachtig, S.** (2011). Transcriptional and post-transcriptional regulation of a NAC1 transcription factor in *Medicago truncatula* roots. *New Phytol.* **191**: 647–661.
- Dai, X. and Zhao, P.X.** (2011). PsRNATarget: A plant small RNA target analysis server. *Nucleic Acids Res.* **39**: 155–159.
- Dai, X., Zhuang, Z., and Zhao, P.X.** (2011). Computational analysis of miRNA targets in plants: Current status and challenges. *Brief. Bioinform.* **12**: 115–121.
- Dai, X., Zhuang, Z., and Zhao, P.X.** psRNATarget: A Plant Small RNA Target Analysis Server (2017 update). unpublished.
- Dalla Via, V., Narduzzi, C., Aguilar, O.M., Zanetti, M.E., and Blanco, F.A.** (2015). Changes in the Common Bean Transcriptome in Response to Secreted and Surface Signal Molecules of *Rhizobium etli*. *Plant Physiol.* **169**: 1356–1370.
- Dalla Via, V., Traubenik, S., Rivero, C., Aguilar, O.M., Zanetti, M.E., and Blanco, F.A.** (2017). The monomeric GTPase RabA2 is required for progression and maintenance of membrane integrity of infection threads during root nodule symbiosis. *Plant Mol. Biol.* **93**: 549–562.
- Dangl, J.L.** (2013). Pivoting the Plant Immune System. *Science* **341**: 746–751.
- Deakin, W.J. and Broughton, W.J.** (2009). Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nat. Rev. Microbiol.* **7**: 312–20.
- Deleris, A., Halter, T., and Navarro, L.** (2016). DNA Methylation and Demethylation in Plant Immunity. *Annu. Rev. Phytopathol.* **54**: 579–603.
- Dénarié, J. and Cullimore, J.** (1993). Lipo-Oligosaccharide Nodulation Factors: A Minireview New Class of Signaling Molecules Mediating Recognition and Morphogenesis. *Plant Physiol.* **74**: 951–954.
- Dénarié, J., Debelle, F., and Promé, J.-C.** (1996). *Rhizobium* Lipo-Chitooligosaccharide Nodulation Factors: Signaling Molecules Mediating Recognition and Morphogenesis. *Annu. Rev. Biochem.* **65**: 503–535.
- Desaki, Y., Miyata, K., Suzuki, M., Shibuya, N., and Kaku, H.** (2017). Plant immunity and symbiosis signaling mediated by LysM receptors. *Innate Immun.* **24**: 92–100.
- Devers, E.A., Branscheid, A., May, P., and Krajinski, F.** (2011). Stars and Symbiosis: MicroRNA- and MicroRNA\*-Mediated Transcript Cleavage Involved in Arbuscular Mycorrhizal Symbiosis. *Plant Physiol.* **156**: 1990–2010.
- Dezulian, T., Palatnik, J., Huson, D., and Weigel, D.** (2005). Conservation and divergence of

- microRNA families in plants. *Genome Biol.* **6**: P13.
- Diaz-Pendon, J.A., Li, F., Li, W.-X., and Ding, S.-W.** (2007). Suppression of antiviral silencing by cucumber mosaic virus 2b protein in *Arabidopsis* is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* **19**: 2053–63.
- Ding, Y.F. and Zhu, C.** (2009). The role of microRNAs in copper and cadmium homeostasis. *Biochem. Biophys. Res. Commun.* **386**: 6–10.
- Downie, J.A.** (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol. Rev.* **34**: 150–170.
- Doyle, J.J.** (2011). Phylogenetic Perspectives on the Origins of Nodulation. *Mol. Plant-Microbe Interact.* **24**: 1289–1295.
- Dubrovsky, J.G., Gambetta, G.A., Hernández-Barrera, A., Shishkova, S., and González, I.** (2006). Lateral root initiation in *Arabidopsis*: Developmental window, spatial patterning, density and predictability. *Ann. Bot.* **97**: 903–915.
- Dupont, L., Alloing, G., Pierre, O., El, S., Hopkins, J., Hrouart, D., and Frendo, P.** (2012). The Legume Root Nodule: From Symbiotic Nitrogen Fixation to Senescence. *Senescence.*
- Elbashir, S.M., Elbashir, S.M., Lendeckel, W., Lendeckel, W., Tuschl, T., and Tuschl, T.** (2001). RNA interference is mediated 1- and 22-nucleotide RNAs. *Genes Dev.* **15**: 188–200.
- Elmore, J.M., Lin, Z.D., and Coaker, G.** (2012). A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Curr. Opin. Plant Biol.* **14**: 365–371.
- Estrada-Navarrete, G. et al.** (2006). *Agrobacterium rhizogenes* transformation of the *Phaseolus* spp.: a tool for functional genomics. *Mol Plant Microbe Interact* **19**: 1385–1393.
- Estrada-Navarrete, G., Alvarado-Affantranger, X., Olivares, J.E., Guillén, G., Díaz-Camino, C., Campos, F., Quinto, C., Gresshoff, P.M., and Sanchez, F.** (2007). Fast, efficient and reproducible genetic transformation of *Phaseolus* spp. by *Agrobacterium rhizogenes*. *Nat. Protoc.* **2**: 1819–1824.
- Etemadi, M., Gutjahr, C., Couzigou, J.-M., Zouine, M., Laressergues, D., Timmers, A., Audran, C., Bouzayen, M., Bécard, G., and Combier, J.-P.** (2014). Auxin Perception Is Required for Arbuscule Development in Arbuscular Mycorrhizal Symbiosis. *Plant Physiol.* **166**: 281–292.
- Fahlgren, N. and Carrington, J.C.** (2010). miRNA Target Prediction in Plants. In Humana Press, pp. 51–57.
- Fang, X. and Qi, Y.** (2016). RNAi in Plants: An Argonaute-Centered View. *Plant Cell* **28**: 272–285.
- Fang, Y. and Spector, D.L.** (2007). Identification of Nuclear Dicing Bodies Containing Proteins for MicroRNA Biogenesis in Living *Arabidopsis* Plants. *Curr. Biol.* **17**: 818–823.
- Fei, Q., Li, P., Teng, C., and Meyers, B.C.** (2015). Secondary siRNAs from *Medicago* NB-LRRs modulated via miRNA-target interactions and their abundances. *Plant J.* **83**: 451–465.
- Fei, Q., Xia, R., and Meyers, B.C.** (2013). Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. *Plant Cell* **25**: 2400–2415.
- Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M.H., Lin, Y.H., Reid, D.E., and Gresshoff, P.M.** (2010). Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* **52**: 61–76.
- Ferguson, B.J. and Mathesius, U.** (2014). Phytohormone Regulation of Legume-Rhizobia Interactions. *J. Chem. Ecol.* **40**: 770–790.
- Foo, E., McAdam, E.L., Weller, J.L., and Reid, J.B.** (2016). Interactions between ethylene, gibberellins, and brassinosteroids in the development of rhizobial and mycorrhizal symbioses of pea. *J. Exp. Bot.* **67**: 2413–2424.
- Formey, D., Iñiguez, L.P., Peláez, P., Li, Y.F., Sunkar, R., Sánchez, F., Reyes, J.L., and Hernández, G.** (2015). Genome-wide identification of the *Phaseolus vulgaris* sRNAome using small RNA and degradome sequencing. *BMC Genomics* **16**: 423.
- Formey, D., Martín-Rodríguez, J.Á., Leija, A., Santana, O., Quinto, C., Cárdenas, L., and Hernández, G.** (2016). Regulation of small RNAs and corresponding targets in nod factor-induced *phaseolus vulgaris* root hair cells. *Int. J. Mol. Sci.* **17**: 1–15.
- Froussart, E., Bonneau, J., Franche, C., and Bogusz, D.** (2016). Recent advances in actinorhizal symbiosis signaling. *Plant Mol. Biol.* **90**: 613–622.
- Gage, D.J.** (2004). Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiol. Mol. Biol. Rev.* **68**: 280–300.

- Ganesan, K. and Xu, B.** (2017). Polyphenol-rich dry common beans (*Phaseolus vulgaris* L.) and their health benefits. *Int. J. Mol. Sci.* **18**: 2331.
- Gao, M., D'Haese, W., De Rycke, R., Wolucka, B., and Holsters, M.** (2001). Knockout of an azorhizobial dTDP-L-rhamnose synthase affects lipopolysaccharide and extracellular polysaccharide production and disables symbiosis with *Sesbania rostrata*. *Mol. Plant. Microbe. Interact.* **14**: 857–66.
- Gardner, P.P., Daub, J., Tate, J., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., Eddy, S.R., and Bateman, A.** (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res.* **37**: 136–1140.
- Genre, A., Chabaud, M., Balzergue, C., Puech-Pagès, V., Novero, M., Rey, T., Fournier, J., Rochange, S., Bécard, G., Bonfante, P., and Barker, D.G.** (2013). Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca<sup>2+</sup> spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. *New Phytol.* **198**: 190–202.
- Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Peret, B., Laplaze, L., Franche, C., Parniske, M., and Bogusz, D.** (2008). SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiacteria. *Proc. Natl. Acad. Sci.* **105**: 4928–4932.
- Gobbato, E.** (2015). Recent developments in arbuscular mycorrhizal signaling. *Curr. Opin. Plant Biol.* **26**: 1–7.
- Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy Team, T.** (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**: 86.
- Goettel, W., Liu, Z., Xia, J., Zhang, W., Zhao, P.X., and An, Y.Q.** (2014). Systems and evolutionary characterization of microRNAs and their underlying regulatory networks in soybean cotyledons. *PLoS One* **9**.
- González, V. et al.** (2003). The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biol.* **4**: R36.
- Gonzalez, V., Santamaria, R.I., Bustos, P., Hernandez-Gonzalez, I., Medrano-Soto, A., Moreno-Hagelsieb, G., Janga, S.C., Ramirez, M.A., Jimenez-Jacinto, V., Collado-Vides, J., and Davila, G.** (2006). The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. *Proc. Natl. Acad. Sci.* **103**: 3834–3839.
- Gough, C. and Cullimore, J.** (2011). Lipo-chitooligosaccharide Signaling in Endosymbiotic Plant-Microbe Interactions. *Mol. Plant-Microbe Interact.* **24**: 867–878.
- Graham, P.H.** (1981). Some problems of nodulation and symbiotic nitrogen fixation in *Phaseolus vulgaris* L.: a review. *F. Crop. Res.* **4**: 93–112.
- Graham, P.H. and Vance, C.P.** (2003). Update on Legume Utilization Legumes: Importance and Constraints to Greater Use. *Plant Physiol.* **131**: 872–877.
- Grange, L. and Hungria, M.** (2004). Genetic diversity of indigenous common bean (*Phaseolus vulgaris*) rhizobia in two Brazilian ecosystems. *Soil Biol. Biochem.* **36**: 1389–1398.
- Griffiths-Jones, S.** (2005). Annotating Non-Coding RNAs with Rfam. In *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc.: Hoboken, NJ, USA), p. Unit 12.5.
- Griffiths-Jones, S.** (2010). MiRBase: MicroRNA sequences and annotation. *Curr. Protoc. Bioinforma.* **34**: 1291–12910.
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R.** (2003). Rfam: An RNA family database. *Nucleic Acids Res.* **31**: 439–441.
- Groth, M., Takeda, N., Perry, J., Uchida, H., Dräxl, S., Brachmann, A., Sato, S., Tabata, S., Kawaguchi, M., Wang, T.L., and Parniske, M.** (2010). NENA, a *Lotus japonicus* homolog of Sec13, is required for rhizodermal infection by arbuscular mycorrhiza fungi and rhizobia but dispensable for cortical endosymbiotic development. *Plant Cell* **22**: 2509–26.
- Gutjahr, C.** (2014). Phytohormone signaling in arbuscular mycorrhiza development. *Curr. Opin. Plant Biol.* **20**: 26–34.
- Gutjahr, C., Banba, M., Croset, V., An, K., Miyao, A., An, G., Hirochika, H., Imaizumi-Anraku, H., and Paszkowski, U.** (2008). Arbuscular Mycorrhiza-Specific Signaling in Rice Transcends the Common Symbiosis Signaling Pathway. *PLANT CELL ONLINE* **20**: 2989–3005.
- Gutjahr, C. and Parniske, M.** (2013). Cell and Developmental Biology of Arbuscular Mycorrhiza Symbiosis. *Annu. Rev. Cell Dev. Biol.* **29**: 593–617.
- Haag, A.F., Arnold, M.F.F., Myka, K.K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert,**

- P., and Ferguson, G.P.** (2012). Molecular insights into bacteroid development during *Rhizobium*-legume symbiosis. *FEMS Microbiol. Rev.* **37**: 364–383.
- Haage, K. and Parniske, M.** (2013). Molecular Mechanisms Governing Arbuscular Mycorrhiza Development and Function. In *Molecular Microbial Ecology of the Rhizosphere* (John Wiley & Sons, Inc.: Hoboken, NJ, USA), pp. 457–465.
- Hacquard, S., Spaepen, S., Garrido-Oter, R., and Schulze-Lefert, P.** (2017). Interplay Between Innate Immunity and the Plant Microbiota. *Annu. Rev. Phytopathol.* **55**: 565–589.
- Hadri, A., and Bisseling, T.** (1998). Responses of the Plant to Nod Factors. In *The Rhizobiaceae* Springer N. H. Spaink, A. Kondorosi, and P.J. Hooykaas, eds.
- Hardarson, G.** (1993). Methods for enhancing symbiotic nitrogen fixation. *Plant Soil* **152**: 1–17.
- Harrison, M.J.** (2012). Cellular programs for arbuscular mycorrhizal symbiosis. *Curr. Opin. Plant Biol.* **15**: 691–698.
- Harrison, M.J.** (2005). Signaling in the arbuscular mycorrhizal symbiosis. *Annu. Rev. Microbiol.* **59**: 19–42.
- Hassan, S. and Mathesius, U.** (2012). The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. *J. Exp. Bot.* **63**: 3429–3444.
- Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., Ordoukhanian, P.** (2014). Library construction for next-generation sequencing: Overviews and Challenges. *Biotechniques* **56**: 61–68.
- Den Herder, G. and Parniske, M.** (2009). The unbearable naivety of legumes in symbiosis. *Curr. Opin. Plant Biol.* **12**: 491–499.
- Hirsch, S., Kim, J., Munoz, A., Heckmann, A.B., Downie, J.A., and Oldroyd, G.E.D.** (2009). GRAS Proteins Form a DNA Binding Complex to Induce Gene Expression during Nodulation Signaling in *Medicago truncatula*. *Plant Cell Online* **21**: 545–557.
- Hobecker, K.V., Reynoso, M.A., Bustos-Sanmamed, P., Wen, J., Mysore, K.S., Crespi, M., Blanco, F.A., and Zanetti, M.E.** (2017). The MicroRNA390/TAS3 Pathway Mediates Symbiotic Nodulation and Lateral Root Growth. *Plant Physiol.* **174**: 2469–2486.
- Högslund, N. et al.** (2009). Dissection of symbiosis and organ development by integrated transcriptome analysis of *Lotus japonicus* mutant and wild-type plants. *PLoS One* **4**: 6556.
- Holt, D.B., Gupta, V., Meyer, D., Abel, N.B., Andersen, S.U., Stougaard, J., and Markmann, K.** (2015). micro RNA 172 (miR172) signals epidermal infection and is expressed in cells primed for bacterial invasion in *Lotus japonicus* roots and nodules. *New Phytol.* **208**: 241–256.
- Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D.** (2007). The *Drosophila* RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. *Curr. Biol.* **17**: 1265–1272.
- Huijser, P. and Schmid, M.** (2011). The control of developmental phase transitions in plants. *Development* **138**: 4117–29.
- Hutvagner, G., McLachlan, J., Bálint, É., Tuschl, T., and Zamore, P.D.** (2001). A cellular function for the RNA interference enzyme Dicer in small temporal RNA maturation. *Science* **93**: 834–838.
- Hutvagner, G. and Simard, M.J.** (2008). Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9**: 22–32.
- Imaizumi-Anraku, H. et al.** (2005). Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature* **433**: 527–531.
- Jagadeeswaran, G., Zheng, Y., Li, Y.F., Shukla, L.I., Matts, J., Hoyt, P., MacMil, S.L., Wiley, G.B., Roe, B.A., Zhang, W., and Sunkar, R.** (2009). Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume-specific microRNA families. *New Phytol.* **184**: 85–98.
- Jez, J.M., Lee, S.G., and Sherp, A.M.** (2016). Sequencing and Improved Computational Power for Data Analysis Are Key for Marker-Assisted Se-. *Science* **353**: 1241–1244.
- Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., and Wang, E.** (2016). DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. *Nat. Commun.* **7**: 1–14.
- Jones-Rhoades, M.W. and Bartel, D.P.** (2004). Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* **14**: 787–799.
- Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B.** (2006). MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**: 19–53.
- Jones, J.D.G. and Dangl, J.L.** (2006). The plant immune system. *Nature* **444**: 323–329.

- Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E., and Graham, C. (2009). NIH Public Access. October 5: 619–633.
- Jones, K.M. and Walker, G.C. (2008). Responses of the model legume *Medicago truncatula* to the rhizobial exopolysaccharide succinoglycan. *Plant Signal. Behav.* **3**: 888–90.
- Journet, E.-P., El-Gachtouli, N., Vernoud, V., de Billy, F., Pichon, M., Dedieu, A., Arnould, C., Morandi, D., Barker, D.G., and Gianinazzi-Pearson, V. (2001). *Medicago truncatula* *ENOD11*: A Novel RPRP-Encoding Early Nodulin Gene Expressed During Mycorrhization in Arbuscule-Containing Cells. *Mol. Plant-Microbe Interact.* **14**: 737–748.
- Kaló, P. et al. (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* **308**: 1786–1789.
- Kami, J., Velasquez, V.B., Debouck, D.G., and Gepts, P. (1995). Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci.* **92**: 1101–1104.
- Kanamori, N. et al. (2006). A nucleoporin is required for induction of Ca<sup>2+</sup> spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proc Natl Acad Sci U S A* **103**: 359–364.
- Karas, B., Murray, J., Gorzelak, M., Smith, A., Sato, S., Tabata, S., and Szczyglowski, K. (2005). Invasion of *Lotus japonicus* root hairless 1 by *Mesorhizobium loti* involves the nodulation factor-dependent induction of root hairs. *Plant Physiol.* **137**: 1331–1344.
- Kaschuk, G., Hungria, M., Andrade, D.S., and Campo, R.J. (2006). Genetic diversity of rhizobia associated with common bean (*Phaseolus vulgaris* L.) grown under no-tillage and conventional systems in Southern Brazil. *Appl. Soil Ecol.* **32**: 210–220.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **5**: 0479–0493.
- Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Dev. Cell* **4**: 205–217.
- Kawaharada, Y. et al. (2015). Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523**: 308–312.
- Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**: 645–9.
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**: 1278–1284.
- Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**: 126–139.
- Kistner, C. (2005). Seven *Lotus japonicus* Genes Required for Transcriptional Reprogramming of the Root during Fungal and Bacterial Symbiosis. *Plant Cell Online* **17**: 2217–2229.
- Kouchi, H. et al. (2004). Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. *DNA Res.* **11**: 263–74.
- Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa, T., Umehara, Y., Suganuma, N., and Kawaguchi, M. (2010). How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. *Plant Cell Physiol.* **51**: 1381–1397.
- Kramer, M.F. (2011). Stem-loop RT-qPCR for miRNAs. *Curr. Protoc. Mol. Biol.* **7**.
- Kretschmar, T., Kohlen, W., Sasse, J., Borghi, L., Schlegel, M., Bachelier, J.B., Reinhardt, D., Bours, R., Bouwmeester, H.J., and Martinoia, E. (2012). A petunia ABC protein controls strigolactone-dependent symbiotic signalling and branching. *Nature* **483**: 341–344.
- Kulcheski, F.R. et al. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* **12**: 307.
- Kulikova, T. et al. (2004). The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* **32**: 27–30.
- Kurihara, Y. and Watanabe, Y. (2004). From The Cover: *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci.* **101**: 12753–12758.
- Langmead, B. (2013). Bowtie2. *Nat. Methods* **9**: 357–359.
- Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: 25.
- Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs

- with probably regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858–862.
- Lauressergues, D., Delaux, P.M., Formey, D., Lelandais-Brière, C., Fort, S., Cottaz, S., Bécard, G., Niebel, A., Roux, C., and Combiér, J.P.** (2012). The microRNA miR171h modulates arbuscular mycorrhizal colonization of *Medicago truncatula* by targeting NSP2. *Plant J.* **72**: 512–522.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V.N.** (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415–419.
- Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N.** (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**: 4051–4060.
- Lelandais-Brière, C., Moreau, J., Hartmann, C., and Crespi, M.** (2016). Noncoding RNAs, Emerging Regulators in Root Endosymbioses. *Mol. Plant-Microbe Interact.* **29**: 170–180.
- Lelandais-Brière, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J., and Crespi, M.** (2009). Genome-Wide *Medicago truncatula* Small RNA Analysis Revealed Novel MicroRNAs and Isoforms Differentially Regulated in Roots and Nodules. *Plant Cell Online* **21**: 2780–2796.
- Li, D., Wang, L., Liu, X., Cui, D., Chen, T., Zhang, H., Jiang, C., Xu, C., Li, P., Li, S., Zhao, L., and Chen, H.** (2013). Deep Sequencing of Maize Small RNAs Reveals a Diverse Set of MicroRNA in Dry and Imbibed Seeds. *PLoS One* **8**: 1–14.
- Li, H., Deng, Y., Wu, T., Subramanian, S., and Yu, O.** (2010). Misexpression of miR482, miR1512, and miR1515 Increases Soybean Nodulation. *Plant Physiol.* **153**: 1759–1770.
- Li, W.X., Oono, Y., Zhu, J., He, X.J., Wu, J.-M., Iida, K., Lu, X.Y., Cui, X., Jin, H., and Zhu, J.K.** (2008). The Arabidopsis NFYA5 Transcription Factor Is Regulated Transcriptionally and Posttranscriptionally to Promote Drought Resistance. *Plant Cell Online* **20**: 2238–2251.
- Li, X., Lei, M., Yan, Z., Wang, Q., Chen, A., Sun, J., Luo, D., and Wang, Y.** (2014). The REL3 - mediated *TAS3* ta-siRNA pathway integrates auxin and ethylene signaling to regulate nodulation in *Lotus japonicus*. *New Phytol.* **201**: 531–544.
- Libault, M., Farmer, A., Brechenmacher, L., May, G.D., and Stacey, G.** (2010). Soybean root hairs: a valuable system to investigate plant biology at the cellular level. *Plant Signal. Behav.* **5**: 419–21.
- Limpens, E., Franken, C., Smit, P., Willemsse, J., Bisseling, T., and Geurts, R.** (2003). LysM Domain Receptor Kinases Regulating Rhizobial Nod Factor-Induced Infection. *Science* **302**: 630–633.
- Llaca, V. and Gepts, P.** (1996). Pulsed-field gel electrophoresis analysis of the phaseolin locus region in *Phaseolus vulgaris*. *Genome* **39**: 722–9.
- Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C.** (2002). Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science* **297**: 2053–2056.
- Lohar, D.P., Sharopova, N., Endre, G., Peñuela, S., Samac, D., Town, C., Silverstein, K.T., and VandenBosch, K.** (2006). Transcript analysis of early nodulation events in *Medicago truncatula*. *Plant Physiol.* **140**: 221–234.
- Lopez-Gomez, M., Sandal, N., Stougaard, J., and Boller, T.** (2012). Interplay of flg22-induced defence responses and nodulation in *Lotus japonicus*. *J. Exp. Bot.* **63**: 393–401.
- Lott, M., Mapleson, M., Mohorianu, I., Moulton, V., Moxon, S., and Schwach, F.** (2012). The UEA sRNA Toolkit: A User Guide for the Perl Implementation. SPECTRE.
- Lu, C., Meyers, B.C., and Green, P.J.** (2007). Construction of small RNA cDNA libraries for deep sequencing. *Methods* **43**: 110–117.
- De Luis, A., Markmann, K., Cognat, V., Holt, D.B., Charpentier, M., Parniske, M., Stougaard, J., and Voinnet, O.** (2012). Two MicroRNAs Linked to Nodule Infection and Nitrogen-Fixing Ability in the Legume *Lotus japonicus*. *Plant Physiol.* **160**: 2137–2154.
- Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., and Stougaard, J.** (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**: 637–640.
- Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K., and Stougaard, J.** (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nat. Commun.* **1**: 1038.
- Maillet, F. et al.** (2011). Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469**: 58–63.

- Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P.** (2004). MicroRNA control of PHABULOSA in leaf development: Importance of pairing to the microRNA 5' region. *EMBO J.* **23**: 3356–3364.
- Mallory, A.C. and Vaucheret, H.** (2006). Functions of microRNAs and related small RNAs in plants. *Nat. Genet.* **38**: S31–S36.
- Mao, G., Turner, M., Yu, O., and Subramanian, S.** (2013). miR393 and miR164 influence indeterminate but not determinate nodule development. *Plant Signal. Behav.* **8**: 26753.
- Marin, E., Jouannet, V., Herz, A., Lokerse, A.S., Weijers, D., Vaucheret, H., Nussaume, L., Crespi, M.D., and Maizel, A.** (2010). miR390, *Arabidopsis* TAS3 tasiRNAs, and Their *AUXIN RESPONSE FACTOR* Targets Define an Autoregulatory Network Quantitatively Regulating Lateral Root Growth. *Plant Cell* **22**: 1104–1117.
- Markmann, K., Giczey, G., and Parniske, M.** (2008). Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biol.* **6**: 0497–0506.
- Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.R., Schultze, M., Ratet, P., and Oldroyd, G.E.D.** (2007). *Medicago truncatula* NIN Is Essential for Rhizobial-Independent Nodule Organogenesis Induced by Autoactive Calcium/Calmodulin-Dependent Protein Kinase. *Plant Physiol.* **144**: 324–335.
- Martin, F.M., Uroz, S., and Barker, D.G.** (2017). Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**: 4501.
- Martínez-Romero, E.** (2003). Diversity of *Rhizobium-Phaseolus vulgaris* symbiosis: overview and perspectives. *Plant Soil* **252**: 11–23.
- Masson-Boivin, C., Giraud, E., Perret, X., and Batut, J.** (2009). Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol.* **17**: 458–466.
- Mathesius, U.** (2008). Goldacre paper: Auxin: At the root of nodule development? *Funct. Plant Biol.* **35**: 651–668.
- Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H.** (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**: 911–940.
- Matzke, M.A., Kanno, T., and Matzke, A.J.M.** (2015). RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu. Rev. Plant Biol.* **66**: 243–267.
- Mazziotta, L., Reynoso, M.A., Aguilar, O.M., Blanco, F.A., and Zanetti, M.E.** (2013). Transcriptional and functional variation of NF-YC1 in genetically diverse accessions of *Phaseolus vulgaris* during the symbiotic association with *Rhizobium etli*. *Plant Biol.* **15**: 808–818.
- Megraw, M., Cumbie, J.S., Ivanchenko, M.G., and Filichkin, S.A.** (2016). Small Genetic Circuits and MicroRNAs: Big Players in Polymerase II Transcriptional Control in Plants. *Plant Cell* **28**: 286–303.
- Meschini, E.P., Blanco, F.A., Zanetti, M.E., Beker, M.P., Küster, H., Pühler, A., and Aguilar, O.M.** (2008). Host Genes Involved in Nodulation Preference in Common Bean (*Phaseolus vulgaris*)–*Rhizobium etli* Symbiosis Revealed by Suppressive Subtractive Hybridization. *Mol. Plant-Microbe Interact.* **21**: 459–468.
- Meyers, B.C. et al.** (2008). Criteria for Annotation of Plant MicroRNAs. *Plant Cell Online* **20**: 3186–3190.
- Mohorianu, I., Bretman, A., Smith, D.T., Fowler, E.K., Dalmay, T., and Chapman, T.** (2017). Comparison of alternative approaches for analysing multi-level RNA-seq data. *PLoS One* **12**: 1–24.
- Moissiard, G., Parizotto, E.A., Humber, C., and Voinnet, O.** (2007). Transitivity in *Arabidopsis* can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. *RNA* **13**: 1268–78.
- Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C.** (2008). Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation. *Cell* **133**: 128–141.
- Moran, Y., Agron, M., Praher, D., and Technau, U.** (2017). The evolutionary origin of plant and animal microRNAs. *Nat. Ecol. Evol.* **1**: 1–8.
- Morozova, O. and Marra, M.A.** (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**: 255–264.
- Mostasso, L., Mostasso, F., Dias, B., Vargas, M.A.T., Hungria, M.** (2002). Selection of bean



- (*Phaseolus vulgaris* L.) rhizobial strains for the Brazilian Cerrados. *Field Crops Research*. **73**: 121–132.
- Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G.** (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**: 720–728.
- Moxon, S., Schwach, F., Dalmay, T., MacLean, D., Studholme, D.J., and Moulton, V.** (2008). A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* **24**: 2252–2253.
- Mückstein, U., Tafer, H., Hackermüller, J., Bernhart, S.H., Stadler, P.F., and Hofacker, I.L.** (2006). Thermodynamics of RNA-RNA binding. *Bioinformatics* **22**: 1177–1182.
- Nova-Franco, B., Iñiguez, L.P., Valdés-López, O., Alvarado-Affantranger, X., Leija, A., Fuentes, S.I., Ramírez, M., Paul, S., Reyes, J.L., Girard, L., and Hernández, G.** (2015). The Micro-RNA172c-APETALA2-1 Node as a Key Regulator of the Common Bean-*Rhizobium etli* Nitrogen Fixation Symbiosis. *Plant Physiol.* **168**: 273–291.
- 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., Hernandez, G., and Vance, C.P.** (2014). An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* **15**: 1–16.
- Oldroyd, G.E.D.** (2013). Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**: 252–263.
- Oldroyd, G.E.D. and Downie, J.A.** (2008). Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annu. Rev. Plant Biol.* **59**: 519–546.
- Oldroyd, G.E.D., Murray, J.D., Poole, P.S., and Downie, J.A.** (2011). The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annu. Rev. Genet.* **45**: 119–144.
- Op den Camp, R., Streng, A., De Mita, S., Cao, Q., Polone, E., Liu, W., Ammiraju, J.S.S., Kudrna, D., Wing, R., Untergasser, A., Bisseling, T., and Geurts, R.** (2011). LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. *Science* **331**: 909–12.
- Pandey, S.P., Shahi, P., Gase, K., and Baldwin, I.T.** (2008). Herbivory-induced changes in the small-RNA transcriptome and phytohormone signaling in *Nicotiana attenuata*. *Proc. Natl. Acad. Sci. U. S. A.* **105**: 4559–64.
- Pant, B.D., Musialak-Lange, M., Nuc, P., May, P., Buhtz, A., Kehr, J., Walther, D., and Scheible, W.-R.** (2009). Identification of nutrient-responsive Arabidopsis and rapeseed microRNAs by comprehensive real-time polymerase chain reaction profiling and small RNA sequencing. *Plant Physiol.* **150**: 1541–55.
- Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S.** (2005). Nuclear processing and export of microRNAs in Arabidopsis. *Proc. Natl. Acad. Sci.* **102**: 3691–3696.
- Park, W., Li, J., Song, R., Messing, J., and Chen, X.** (2002). CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis thaliana. *Curr Biol.* **12**: 1484–1495.
- Parniske, M.** (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat. Rev. Microbiol.* **6**: 763–775.
- Patriarca, E.J., Tatè, R., and Iaccarino, M.** (2002). Key Role of Bacterial NH<sub>4</sub><sup>+</sup> Metabolism in Rhizobium-Plant Symbiosis. *Microbiol. Mol. Biol. Rev.* **66**: 203–222.
- Peltier, H.J. and Latham, G.J.** (2008). Normalization of microRNA expression levels in quantitative RT-PCR assays: Identification of suitable reference RNA targets in normal and cancerous human solid tissues. *Rna* **14**: 844–852.
- Perotto, S., Brewin, N.J., and Kannenberg, E.L.** (1994). Cytological Evidence for a Host-Defense Response That Reduces Cell and Tissue Invasion in Pea Nodules By Lipopolysaccharide-Defective Mutants of *Rhizobium-Leguminosarum* Strain-3841. *Mol. Plant-Microbe Interact.* **7**: 99–112.
- Prüfer, K., Stenzel, U., Dannemann, M., Green, R.E., Lachmann, M., and Kelso, J.** (2008). PatMaN: Rapid alignment of short sequences to large databases. *Bioinformatics* **24**: 1530–1531.
- Raabe, C.A., Tang, T.-H., Brosius, J., and Rozhdestvensky, T.S.** (2014). Biases in small RNA deep sequencing data. *Nucleic Acids Res.* **42**: 1414–1426.
- Radutoiu, S., Madsen, L.H., Madsen, E.B., Jurkiewicz, A., Fukai, E., Quistgaard, E.M.H., Albrektsen, A.S., James, E.K., Thirup, S., and Stougaard, J.** (2007). LysM domains mediate lipochitin-oligosaccharide recognition and Nfr genes extend the symbiotic host range. *EMBO J.* **26**: 3923–3935.

- Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P.** (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* **20**: 3407–3425.
- Rajewsky, N.** (2006). MicroRNA target predictions in animals. *Nat. Genet.* **38**: 8–13.
- Ramachandran, V. and Chen, X.** (2008). Degradation of microRNAs by a Family of Exoribonucleases in *Arabidopsis*. *Science* **321**: 1490–1492.
- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., and Guarino, L.** (2010). A Gap analysis methodology for collecting crop gene pools: A case study with Phaseolus beans. *PLoS One* **5**: 13497.
- Redmond, J.W., Batley, M., Djordjevic, M.A., Innes, R.W., Kuempel, P.L., and Rolfe, B.G.** (1986). Flavones induce expression of nodulation genes in *Rhizobium*. *Nature* **323**: 632–635.
- Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P.** (2002). MicroRNAs in plants. *Genes Dev.* **16**: 1616–1626.
- Reynoso, M.A., Blanco, F.A., Bailey-Serres, J., Crespi, M., and Zanetti, M.E.** (2013). Selective recruitment of mRNAs and miRNAs to polyribosomes in response to rhizobia infection in *Medicago truncatula*. *Plant J.* **73**: 289–301.
- Richard, M.M.S. and Takken, F.L.W.** (2017). Plant Autoimmunity: When Good Things Go Bad. *Curr. Biol.* **27**: 361–363.
- Riely, B.K., Lougnon, G., Ané, J.M., and Cook, D.R.** (2007). The symbiotic ion channel homolog DMI1 is localized in the nuclear membrane of *Medicago truncatula* roots. *Plant J.* **49**: 208–216.
- Rípodas, C., Castaingts, M., Clúa, J., Blanco, F., and Zanetti, M.E.** (2015). Annotation, phylogeny and expression analysis of the nuclear factor Y gene families in common bean (*Phaseolus vulgaris*). *Front. Plant Sci.* **5**: 761.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P.** (2011). Integrative genomics viewer. *Nat. Biotechnol.* **29**: 24–26.
- Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G., and Papa, R.** (2009). Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol. Appl.* **2**: 504–522.
- Rouillard, J.-M., Zuker, M., and Gulari, E.** (2003). OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.* **31**: 3057–62.
- Saito, K. et al.** (2007). NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. *Plant Cell* **19**: 610–24.
- Sambrook, J. and Russel, D.W.** (2001). *Molecular Cloning: a Laboratory Manual*. Cold Spring (New York).
- Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A.** (2002). DICER-LIKE1: Blind men and elephants in *Arabidopsis* development. *Trends Plant Sci.* **7**: 487–491.
- Schauser, L., Roussis, A., Stiller, J., and Stougaard, J.** (1999). A plant regulator controlling development of symbiotic root nodules. *Nature* **402**: 191–195.
- Scheidle, H., Groß, A., and Niehaus, K.** (2004). The Lipid A substructure of the *Sinorhizobium meliloti* lipopolysaccharides is sufficient to suppress the oxidative burst in host plants. *New Phytol.* **165**: 559–566.
- Schmutz, J. et al.** (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**: 707–713.
- Schmutz, J. et al.** (2010). Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178–183.
- Schwab, R., Palatnik, J.F., Rießer, M., Schommer, C., Schmid, M., and Weigel, D.** (2005). Specific effects of microRNAs on the plant transcriptome. *Dev. Cell* **8**: 517–527.
- Simms, E.L.** (2002). Partner Choice in Nitrogen-Fixation Mutualisms of Legumes and Rhizobia. *Integr. Comp. Biol.* **42**: 369–380.
- Simms, E.L. and Taylor, D.L.** (2002). Partner Choice in Nitrogen-Fixation Mutualisms of Legumes and Rhizobia. *Integr. Comp. Biol.* **42**: 369–380.
- Singh, S., Katzer, K., Lambert, J., Cerri, M., and Parniske, M.** (2014). CYCLOPS, A DNA-Binding Transcriptional Activator, Orchestrates Symbiotic Root Nodule Development. *Cell Host Microbe* **15**: 139–152.
- Singh, S. and Parniske, M.** (2012). Activation of calcium- and calmodulin-dependent protein kinase (CCaMK), the central regulator of plant root endosymbiosis. *Curr. Opin. Plant Biol.* **15**: 444–453.
- Smit, P.,** -(2005). NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced

- transcription. *Science* **308**: 1789–91.
- Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., and Bisseling, T.** (2007). Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. *Plant Physiol.* **145**: 183–191.
- Song, Q.X., Liu, Y.F., Hu, X.Y., Zhang, W.K., Ma, B., Chen, S.Y., and Zhang, J.S.** (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol.* **11**: 5.
- de Sousa Cardoso, T.C., Portilho, L.G., de Oliveira, C.L., Mckeown, P.C., Maluf, W.R., Gomes, L.A.A., Teixeira, T.A., do Amaral, L.R., Spillane, C., and de Souza Gomes, M.** (2016). Genome-wide identification and in silico characterisation of microRNAs, their targets and processing pathway genes in *Phaseolus vulgaris* L. *Plant Biol.* **18**: 206–219.
- Soyano, T. and Hayashi, M.** (2014). Transcriptional networks leading to symbiotic nodule organogenesis. *Curr. Opin. Plant Biol.* **20**: 146–154.
- Soyano, T., Hirakawa, H., Sato, S., Hayashi, M., and Kawaguchi, M.** (2014). NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proc. Natl. Acad. Sci.* **111**: 14607–14612.
- Soyano, T., Kouchi, H., Hirota, A., and Hayashi, M.** (2013). NODULE INCEPTION Directly Targets NF-Y Subunit Genes to Regulate Essential Processes of Root Nodule Development in *Lotus japonicus*. *PLoS Genet.* **9**.
- Stocks, M.B., Moxon, S., Mapleson, D., Woolfenden, H.C., Mohorianu, I., Folkes, L., Schwach, F., Dalmay, T., and Moulton, V.** (2012). The UEA sRNA workbench: A suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* **28**: 2059–2061.
- Strack, D., Fester, T., Hause, B., Schliemann, W., and Walter, M.H.** (2003). Arbuscular mycorrhiza: biological, chemical, and molecular aspects. *J. Chem. Ecol.* **29**: 1955–79.
- Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K., and Parniske, M.** (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* **417**: 959–962.
- Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K., and Yu, O.** (2008). Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics* **9**: 1–14.
- Szarzynska, B., Sobkowiak, L., Pant, B.D., Balazadeh, S., Scheible, W.R., Mueller-Roeber, B., Jarmolowski, A., and Szweykowska-Kulinska, Z.** (2009). Gene structures and processing of *Arabidopsis thaliana* HYL1-dependent pri-miRNAs. *Nucleic Acids Res.* **37**: 3083–3093.
- Szittyá, G., Moxon, S., Santos, D.M., Jing, R., Fevereiro, M.P.S., Moulton, V., and Dalmay, T.** (2008). High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. *BMC Genomics* **9**: 1–9.
- Takken, F.L.W. and Goverse, A.** (2012). How to build a pathogen detector: Structural basis of NB-LRR function. *Curr. Opin. Plant Biol.* **15**: 375–384.
- Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D.** (2003). A biochemical framework for RNA silencing in plants A biochemical framework for RNA silencing in plants. *Genes & Dev.* **2**: 49–63.
- Taylor, R.G., Walker, D.C., and McInnes, R.R.** (1993). *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.* **21**: 1677–8.
- Taylor, R.S., Tarver, J.E., Hiscock, S.J., and Donoghue, P.C.J.** (2014). Evolutionary history of plant microRNAs. *Trends Plant Sci.* **19**: 175–182.
- Thies, J.E., Singleton, P.W., and Bohlool, B.B.** (1991a). Influence of the size of indigenous rhizobial populations on establishment and symbiotic performance of introduced rhizobia on field-grown legumes. *Appl. Environ. Microbiol.* **57**: 19–28.
- Thies, J.E., Singleton, P.W., and Bohlool, B.B.** (1991b). Modeling symbiotic performance of introduced rhizobia in the field by use of indices of indigenous population size and nitrogen status of the soil. *Appl. Environ. Microbiol.* **57**: 29–37.
- Thorneley, R.N., Abell, C., Ashby, G.A., Drummond, M.H., Eady, R.R., Huff, S., Macdonald, C.J., and Shneier, A.** (1992). Posttranslational modification of *Klebsiella pneumoniae* flavodoxin by covalent attachment of coenzyme A, shown by <sup>31</sup>P NMR and electrospray mass spectrometry, prevents electron transfer from the nifJ protein to nitrogenase. A possible new regulatory mechanism for biological nitrogen fixation. *Biochemistry* **31**: 1216–24.
- Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P.** (2013). Integrative Genomics Viewer

- (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**: 178–192.
- Triplett, E.W. and Sadowsky, M.** (1992). Nodulation of Legumes. *Annu. Rev. Microbiol.* **46**: 399–428.
- Turner, M., Adhikari, S., and Subramanian, S.** (2013a). Optimizing stem-loop qPCR assays through multiplexed cDNA synthesis of U6 and miRNAs. *Plant Signal. Behav.* **8**: 1–4.
- Turner, M., Nizampatnam, N.R., Baron, M., Coppin, S., Damodaran, S., Adhikari, S., Arunachalam, S.P., Yu, O., and Subramanian, S.** (2013b). Ectopic Expression of miR160 Results in Auxin Hypersensitivity, Cytokinin Hyposensitivity, and Inhibition of Symbiotic Nodule Development in Soybean. *Plant Physiol.* **162**: 2042–2055.
- Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E.F., and Hellens, R.P.** (2007). Protocol: A highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* **3**: 1–12.
- Vernié, T., Kim, J., Frances, L., Ding, Y., Sun, J., Guan, D., Niebel, A., Gifford, M.L., de Carvalho-Niebel, F., and Oldroyd, G.E.D.** (2015). The NIN Transcription Factor Coordinates Diverse Nodulation Programs in Different Tissues of the *Medicago truncatula* Root. *Plant Cell* **27**: 3410–3424.
- Vlasova, A. et al.** (2016). 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**: 1–18.
- Voinnet, O.** (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell* **136**: 669–687.
- Wang, Y., Wang, L., Zou, Y., Chen, L., Cai, Z., Zhang, S., Zhao, F., Tian, Y., Jiang, Q., Ferguson, B.J., Gresshoff, P.M., and Li, X.** (2014). Soybean miR172c Targets the Repressive AP2 Transcription Factor NNC1 to Activate *ENOD40* Expression and Regulate Nodule Initiation. *Plant Cell Online* **26**: 4782–4801.
- Wong, C.E., Zhao, Y.T., Wang, X.J., Croft, L., Wang, Z.H., Haerizadeh, F., Mattick, J.S., Singh, M.B., Carroll, B.J., and Bhalla, P.L.** (2011). MicroRNAs in the shoot apical meristem of soybean. *J. Exp. Bot.* **62**: 2495–2506.
- Wong, K.H., Jin, Y., Moqtaderi, Z., Wong, K.H., Jin, Y., and Moqtaderi, Z.** (2013). Multiplex Illumina Sequencing Using DNA Barcoding. In *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc.: Hoboken, NJ, USA), p. 7.11.1–7.11.11.
- Xia, R., Meyers, B.C., Liu, Z., Beers, E.P., Ye, S., and Liu, Z.** (2013). MicroRNA Superfamilies Descended from miR390 and Their Roles in Secondary Small Interfering RNA Biogenesis in Eudicots. *Plant Cell* **25**: 1555–1572.
- Xia, R., Xu, J., and Meyers, B.C.** (2017). The Emergence, Evolution, and Diversification of the miR390-TAS3-ARF Pathway in Land Plants. *Plant Cell* **7**: 1232–1247.
- Xie, Z.** (2005). Expression of Arabidopsis MIRNA Genes. *Plant Physiol.* **138**: 2145–2154.
- Xie, Z., Kasschau, K.D., and Carrington, J.C.** (2003). Negative Feedback Regulation of Dicer-Like1 in Arabidopsis by microRNA-Guided mRNA Degradation. *Curr. Biol.* **13**: 784–789.
- Yan, J., Cai, X., Luo, J., Sato, S., Jiang, Q., Yang, J., Cao, X., Hu, X., Tabata, S., Gresshoff, P.M., and Luo, D.** (2010). The REDUCED LEAFLET Genes Encode Key Components of the trans-Acting Small Interfering RNA Pathway and Regulate Compound Leaf and Flower Development in *Lotus japonicus*. *Plant Physiol.* **152**: 797–807.
- Yan, Z. et al.** (2016). Identification and functional characterization of soybean root hair microRNAs expressed in response to *Bradyrhizobium japonicum* infection. *Plant Biotechnol. J.* **14**: 332–341.
- Yan, Z., Hossain, M.S., Arikiti, S., Valdés-López, O., Zhai, J., Wang, J., Libault, M., Ji, T., Qiu, L., Meyers, B.C., and Stacey, G.** (2015). Identification of microRNAs and their mRNA targets during soybean nodule development: functional analysis of the role of miR393j-3p in soybean nodulation. *New Phytol.* **207**: 748–59.
- Yang, S., Tang, F., Gao, M., Krishnan, H.B., and Zhu, H.** (2010). R gene-controlled host specificity in the legume-rhizobia symbiosis. *Proc. Natl. Acad. Sci.* **107**: 18735–18740.
- Yano, K. et al.** (2008). CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc. Natl. Acad. Sci.* **105**: 20540–20545.
- Yoro, E., Suzaki, T., Toyokura, K., Miyazawa, H., Fukaki, H., and Kawaguchi, M.** (2014). A Positive Regulator of Nodule Organogenesis, NODULE INCEPTION, Acts as a Negative Regulator of Rhizobial Infection in *Lotus japonicus*. *Plant Physiol.* **165**: 747–758.
- Yoshikawa, M.** (2013). Biogenesis of trans-acting siRNAs, endogenous secondary siRNAs in plants. *Genes Genet. Syst.* **88**: 77–84.
- Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W.,**

- Lagrange, T., Walker, J.C., and Chen, X.** (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc. Natl. Acad. Sci.* **105**: 10073–10078.
- Zanetti, M.E., Blanco, F.A., Beker, M.P., Battaglia, M., and Aguilar, O.M.** (2010). A C Subunit of the Plant Nuclear Factor NF-Y Required for Rhizobial Infection and Nodule Development Affects Partner Selection in the Common Bean– *Rhizobium etli* Symbiosis. *Plant Cell* **22**: 4142–4157.
- Zanetti, M.E., Rípodas, C., and Niebel, A.** (2017). Plant NF-Y transcription factors: Key players in plant-microbe interactions, root development and adaptation to stress. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860**: 645–654.
- Zhai, J. et al.** (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* **25**: 2540–2553.
- Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., and Anderson, T.A.** (2006). Conservation and divergence of plant microRNA genes. *Plant J.* **46**: 243–259.
- Zhang, Y.** (2005). miRU: An automated plant miRNA target prediction server. *Nucleic Acids Res.* **33**: 701–704.
- Zhou, Z.S., Huang, S.Q., and Yang, Z.M.** (2008). Bioinformatic identification and expression analysis of new microRNAs from *Medicago truncatula*. *Biochem. Biophys. Res. Commun.* **374**: 538–542.
- Zipfel, C. and Oldroyd, G.E.D.** (2017). Plant signalling in symbiosis and immunity. *Nature* **543**: 328–336.
- Zuker, M.** (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406–15.