



Facultad de Ciencias Bioquímicas y Farmacéuticas - Universidad Nacional  
de Rosario

Tesis de Doctorado

**Estudios sobre la regulación de la expresión génica por  
microARNs en plantas mediante estrategias  
bioinformáticas**

**Presentada por: Uciel Pablo Chorostecki**

**Rosario, Argentina**

2016



# **Estudios sobre la regulación de la expresión génica por microARNs en plantas mediante estrategias bioinformáticas**

Uciel Pablo Chorostecki

Licenciado en Ciencias de la Computación

Universidad Nacional de Rosario

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias Biológicas, de la Universidad de Rosario y no ha sido presentada previamente para la obtención de otro título en esta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET), dependiente de la Facultad de Cs. Bioquímicas y Farmacéuticas, durante el período comprendido entre el ?? y el ??, bajo la dirección del Dr. Javier Palatnik.



I would like to dedicate this thesis to my loving parents ...



# Índice general

<b>Índice de figuras</b>	<b>ix</b>
<b>Índice de tablas</b>	<b>xi</b>
<b>1 Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco</b>	<b>1</b>
1.1 Introducción . . . . .	1
1.2 Resultados . . . . .	1
1.2.1 Predicción de genes regulados por microARNs. . . . .	1
1.2.2 comTAR: una herramienta para la predicción de genes blanco regulados por microARNs en plantas. . . . .	12
<b>2 My second chapter</b>	<b>15</b>
2.1 Short title . . . . .	15
<b>References</b>	<b>17</b>
<b>Appendix A Anexo</b>	<b>19</b>





# Índice de figuras

1.1	Estrategia . . . . .	3
-----	----------------------	---



# Índice de tablas

1.1	miARNs y sus genes blancos en plantas . . . . .	4
1.2	Detection of miRNA targets using different filters . . . . .	8
A.1	Especies y base de datos utilizadas para la búsquedas de genes blanco de miARNs conservados . . . . .	20
A.2	My caption . . . . .	21



## **Resumen**

Resumen: Deberá contener la siguiente información

1. Breve presentación del problema
2. Enfoque y planificación del problema
3. Datos significativos y hallazgos más importantes
4. Conclusiones



# Capítulo 1

## Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco

### 1.1 Introducción

### 1.2 Resultados

#### 1.2.1 Predicción de genes regulados por microARNs.

**Diseño de una estrategia para la identificación de genes blanco regulados por mi-croARNs basado en la conservación evolutiva del par microARN-gen blanco.**

Enfocamos nuestro análisis en 22 miARNs que están conservados en Angiospermas ([2], [4]). En general estos miARNs están codificados por pequeñas familias hasta 32 miembros. En los genomas completos de Arabidopsis, poplar y arroz es común encontrar variaciones en la secuencia de los miARNs pertenecientes a una misma familia, especialmente en el primer nucleótido y los nucleótidos 20 y 21 ([4]).

Sin embargo, observamos que la región entre la posición 2 y 19 está bastante conservada y pudimos encontrar una secuencia consenso presente en la mayoría de los miembros de cada familia de miARNs en esas tres especies. (tabla 1.1). Curiosamente, las bases variables fuera de esta región conservada son propensas a tener mismatches con genes blanco conocidos, lo que indica que podría existir una correlación entre la interacción miARN-gen blanco y la conservación de la secuencia del miARN.

Diseñamos una estrategia para identificar nuevos pares miARN-gen blanco principalmente basada en la conservación evolutiva de la secuencia del gen blanco (Figura 1.1). Las secuencias consenso de 18 nt de cada familia de miARN fueron usadas inicialmente para realizar la búsqueda de genes blanco en contigs de ESTs, de 41 especies de plantas, obtenidos de “Gene Index Project”<sup>1</sup> un proyecto mantenido y administrado por la universidad de Harvard que contiene un catálogo completo de genes en una amplia gama de organismos incluyendo plantas. Además se utilizaron ARNm completos para *A. thaliana*<sup>2</sup> y *Oryza Sativa*<sup>3</sup> (para ver la lista completa de especies, ver tabla A.1). Utilizando las secuencias consenso de 18nt y permitiendo 3 mismatches (errores), la búsqueda de genes blanco arrojó como resultado 38.597 genes distribuidos en las 43 especies (Figura 1.1, bin 1). Las interacciones G-U y los bulges fueron considerados como mismatches en esta primera búsqueda. Todos los genes blanco de *A. thaliana* conocidos hasta ese momento fueron identificados usando esta estrategia con la excepción de CSD2, un gen blanco del miR398 que contiene 4 mismatches (tabla A.1).

Teniendo en cuenta que la mayoría de los genes blanco arrojados presentan una escasa descripción del tipo genómica funcional, realizamos un BLASTx contra el proteoma de *A. thaliana*. El “locus ID” obtenido como “best hit” se utilizó como tag (etiqueta) para identificar al candidato en distintas especies (Figura 1.1). A pesar que esta estrategia no necesariamente identifica el gen ortólogo de Arabidopsis, sirve como propósito de clasificación de cada potencial gen blanco de miARN. Aunque la mayoría de los potenciales genes blanco pudieron ser fácilmente asignados con una etiqueta, algunos pocos casos, que incluye a los genes que representan ARNs no codificantes fueron perdidos en este paso.

Este enfoque permite la selección de los mejores candidatos basándose en la presencia de los genes blanco en un número distinto de especies. Utilizando 4 especies como el mínimo de especies requeridas (ya que tiene una buena especificidad), dio como resultado 3.781 genes que corresponden a 533 tags diferentes (Figura 1.1, bin 2).

La búsqueda también se puede hacer en combinación con filtros empíricos de interacción par miARN-gen blanco que tienen en cuenta la energía de interacción y la posición de los mismatches (ver Materiales y métodos). De los 38.597 candidatos iniciales, 9.375 pasan estos filtros (Figura 1.1, bin 4). Combinando filtros de energía y filtro de conservación evolutiva, la búsqueda arrojó como resultado 563 candidatos correspondientes a 146 tags (Figura 1.1, bin 5).

---

<sup>1</sup><http://compbio.dfci.harvard.edu/tgi/>

<sup>2</sup><http://arabidopsis.org>

<sup>3</sup><http://rice.plantbiology.msu.edu>



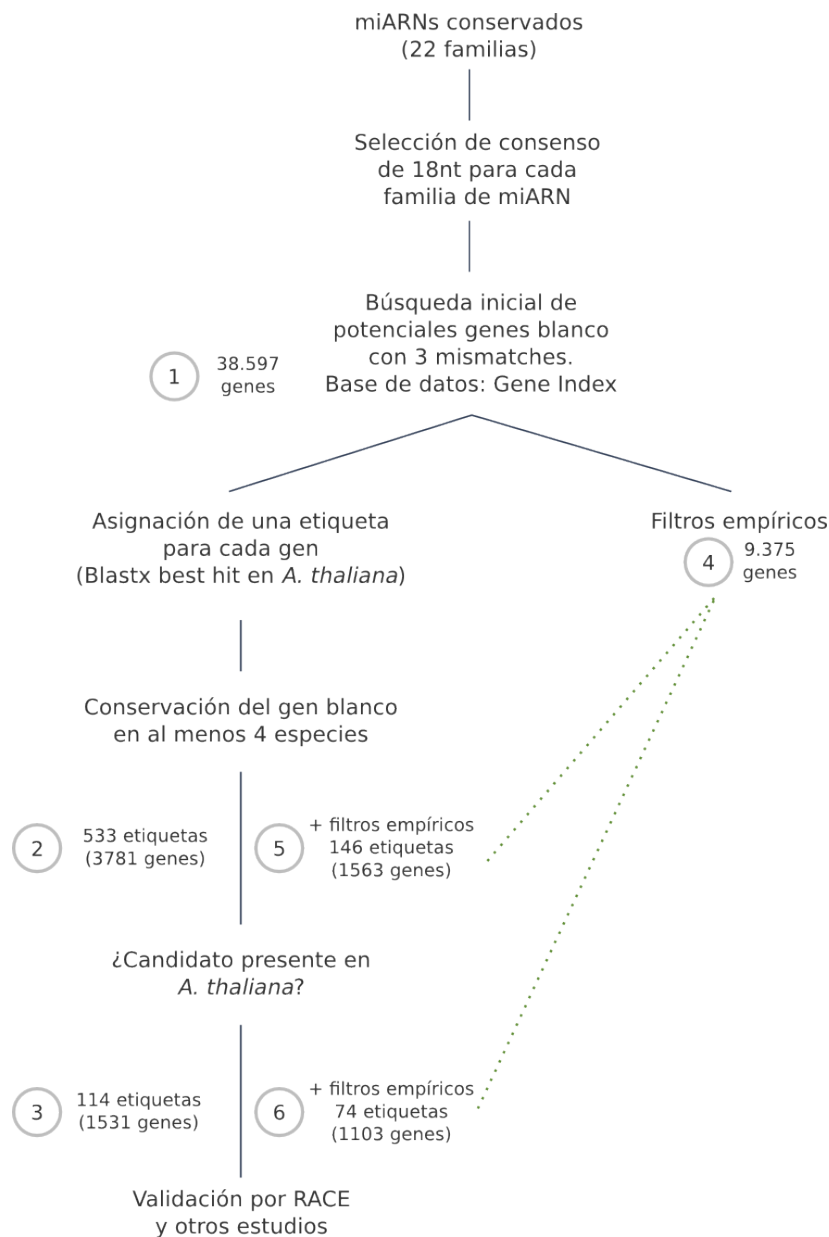


Fig. 1.1 Esquema de la estrategia para la identificación de nuevos genes blanco. El número de genes blanco está identificado en cada paso. Luego de aplicar el análisis de conservación, todos los genes que tienen el mismo hit en Arabidopsis, fueron considerados como un solo gen blanco. El lado derecho muestra la búsqueda hecha con filtros empíricos: bin 5 y 6 incluyen genes blanco seleccionados con ambos filtros, empíricos y de conservación. Mientras que el bin 2 y 3 muestra los potenciales genes blanco seleccionados sólo con el filtro de conservación.

Table 1.1 miARNs y sus genes blancos en plantas

miARN	Consenso (18 nt)	Targets conocidos <sup>(a,b)</sup>
miR156	GACAGAAGAGAGTGAGCA	factores de transcripción SPL
miR159	TTGGATTGAAGGGAGCTC	factores de transcripción MYB, <b>NOZZLE (NZL)</b>
miR160	GCCTGGCTCCCTGTATGC	factores de transcripción ARF
miR162	CGATAAACCTCTGCATCC	DCL1
miR164	GGAGAAGCAGGGCACGTG	factores de transcripción NAC
miR166	CGGACCAGGCTTCAATTC	factores de transcripción HDZip
miR167	GAAGCTGCCAGCATGATC	factores de transcripción ARF, <b>IAA-ALANINE RESISTANT 3 (IAR3)</b>
miR168	CGCTTGGTGACGTCGGG	AGO1
miR169	AGCCAAGGATGACTTGCC	factores de transcripción CCAAT-HAP2
miR171	TTGAGCCGTGCCAATATC	factores de transcripción GRAS
miR172	GAATCTTGATGATGCTGC	factores de transcripción AP2
miR319	TGGACTGAAGGGAGCTCC	factores de transcripción TCP
miR390	AGCTCAGGAGGGATAGCG	TAS RNA
miR393	CCAAAGGGATCGCATTGA	TIR1 proteins, F-BOX proteins
miR394	TGGCATTCTGTCCACCTC	proteínas F-BOX
miR395	TGAAGTGTGTGGGGGAAC	ATP-sulfurilasas, transportadores de sulfato
miR396	TCCACAGCTTTCTTGAAC	factores de transcripción GRF, <b>MMG4.7, FLUORESCENT IN BLUE LIGHT (FLU)</b>
miR397	CATTGAGTGCAGCGTTGA	Laccases
miR398	GTGTTCTCAGGTCACCCC	Cu/Zn SODs, CytC oxidase protein subunit, Chaperona de cobre (CCS)
miR399	GCCAAGGAGATTGCCC	Enzima E2 de conjugación de ubiquitina
miR408	TGCACTGCCTCTTCCCTG	Blue copper proteins, Laccases, <b>P-TYPE ATPase (PAA2), PAC1 (Proteasome component)</b>
miR827	TAGATGACCATCAGCAAA	SPX proteins

a Los genes blancos fueron agrupados según sus funciones.

b Nuevos genes blancos validados experimentalmente en este estudio están indicados en negrita.

### Parámetros empíricos y de conservación evolutiva pueden actuar de manera sinérgica para identificar genes blanco regulados por miARNs.

Potenciales genes blanco de miARNs fueron clasificados de acuerdo al mínimo número de especie en donde fueron detectados (Figura 2A-E). Como control para cada microARN generamos 10 secuencias “scramble” (al azar), dividiendo las secuencias originales de a di-nucleótidos y luego generando nuevas secuencias al azar conservando la composición de los di-nucleótidos. Estas secuencias al azar fueron utilizadas para realizar búsqueda de genes blanco del mismo modo que lo hicimos para las secuencias originales. La relación señal/ruido fue calculada como el cociente entre el número de genes blanco para los miARNs y el número promedio obtenido de las secuencias al azar. El radio fue de 1,2 para todos los miARNs juntos sin requerir conservación y esa relación incrementa con el número de especie en donde los genes blanco fueron detectados (Figura 1.2 A, recuadro). Los datos para todos los miARNs y sus potenciales genes blancos conservados en al menos 4 especies están incluidos en la tabla 1.2.

Luego estudiamos la selección de candidatos teniendo en cuenta los filtros empíricos. Para esto aplicamos una versión modificada de los filtros descritos anteriormente y requiriendo (i) una energía mínima de hibridación (MFE) de al menos 72% del apareamiento perfecto de cada secuencia consenso y (ii) que sólo un mismatch pudiera estar presente entre la posición 1 y la 11 de la secuencia consenso (2-12 del miARN). De la búsqueda inicial 9.375 genes pasaron estos filtros conteniendo el 97% de los genes validados anteriormente de Arabidopsis. (Figura 1.1, bin 4).

Al aplicar solamente este filtro empírico, dio como resultado una relación señal/ruido de 1,7, al agrupar todos los miARNs juntos (Figura 1.2 A). Observamos que aplicar si-

multáneamente los filtros empíricos y de conservación aumentaron significativamente la relación señal/ruido para todos los miARNs juntos (Figura 1.2 A recuadro) y también de cada miARN individualmente (Figura 1.2 B-E, recuadros y tabla 1.2). En varios casos, esta relación llega hasta 10 cuando se requiere de que el gen blanco este presente en más de 5 especies y que pase los filtros empíricos (Figure 1.2 A–D). Este efecto sinérgico indica que el filtro de conservación evolutiva y los parámetros empíricos pueden estar seleccionando aspectos diferentes de la interacción miARN-gen blanco.

Observamos que el número de genes blanco candidato y la relación señal/ruido es variable entre los distintos miARNs. El miR396 tiene la mayor cantidad de potenciales genes blanco, 92 de ellos presentes en al menos 4 especies y 26 de ellos pasan además los filtros empíricos (Tabla 1.2 y Figura 1.2 B). El miR408 y el miR398 también tienen un número alto de potenciales genes blanco y buenas relaciones de señal/ruido (Figura 1.2 C-D).

En contraste, ciertos miARNs como el miR162, miR168 y miR399 tienen un solo potencial gen blanco conservado en al menos 4 especies de acuerdo con nuestra búsqueda (Tabla 1.2 y Figura 1.2 E). Al menos en el caso del miR162 y del miR168 este resultado podría estar reflejando su rol específico en la regulación por retroalimentación de la biogénesis del miARN, ya que controlan los niveles de expresión DCL1 y AGO1 respectivamente ([15],[16]).

Como control adicional para nuestra estrategia hicimos la búsqueda de genes blanco del miR158 y miR173, que son miARNs presentes solamente en *A. thaliana* y especies bien cercanas (17). Como era esperado estos miARNs no generaron más candidatos que sus versiones al azar (Tabla 1.2 y Figura 1.2 F).

Luego chequeamos si los pares miARN-gen blanco altamente conservados tenían una interacción más fuerte que los que están presentes en pocas especies. Para esto calculamos la energía mínima de hibridación para cada interacción detectada en nuestro trabajo. Observamos que los pares miARN-gen blanco presentes en muchas especies tienden a tener energía de interacción mayores que los que están presentes en menos especies (Figure 1.3 A). De todos modos, la correlación no fue notoria y algunas interacciones miARN-gen blanco tuvieron una baja energía de hibridación (Figure 1.3 A). Estos resultados muestran que una alta conservación podría no ser necesariamente equivalente a una fuerte interacción, la misma podría proporcionar una explicación para los efectos sinérgicos causados por los filtros de evolución y empíricos sobre la relación señal/ruido.

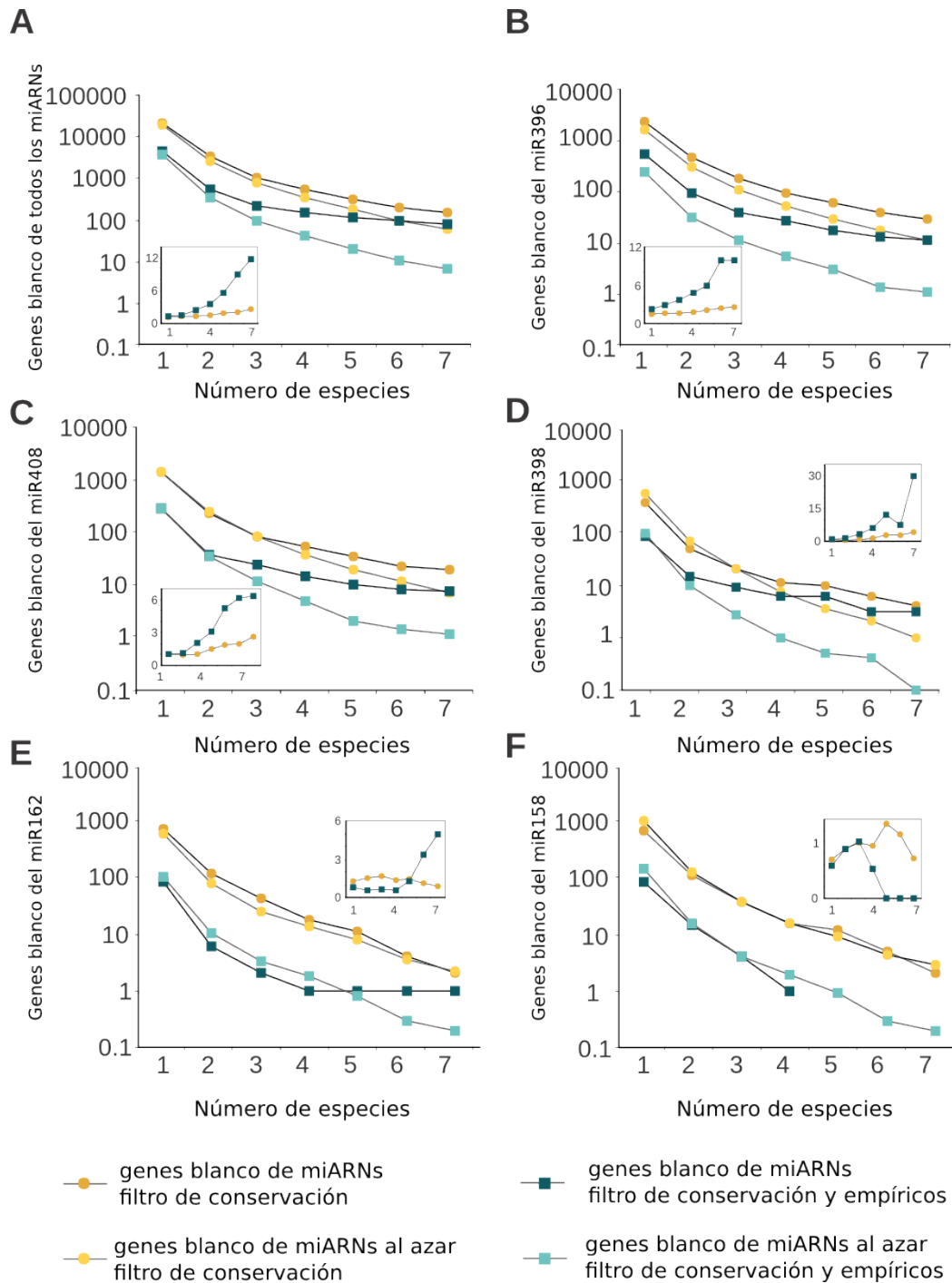


Fig. 1.2 Conservación de potenciales genes blanco en distintas especies. Todos los miARNs (A), miR396 (B), miR408 (C), miR398 (D), miR162 (E), miR158 (F). Puntos naranja representan los genes blanco de miARNs usando filtro evolutivo. Puntos amarillos representan los genes blanco de las secuencias al azar usando filtro evolutivo. El cuadrado azul muestra los genes blanco de miARNs luego de aplicar filtros empíricos y evolutivos, mientras que el cuadrado celeste representa los genes blanco de las secuencias al azar en las mismas condiciones. Los recuadros muestran la relación señal/ruido.

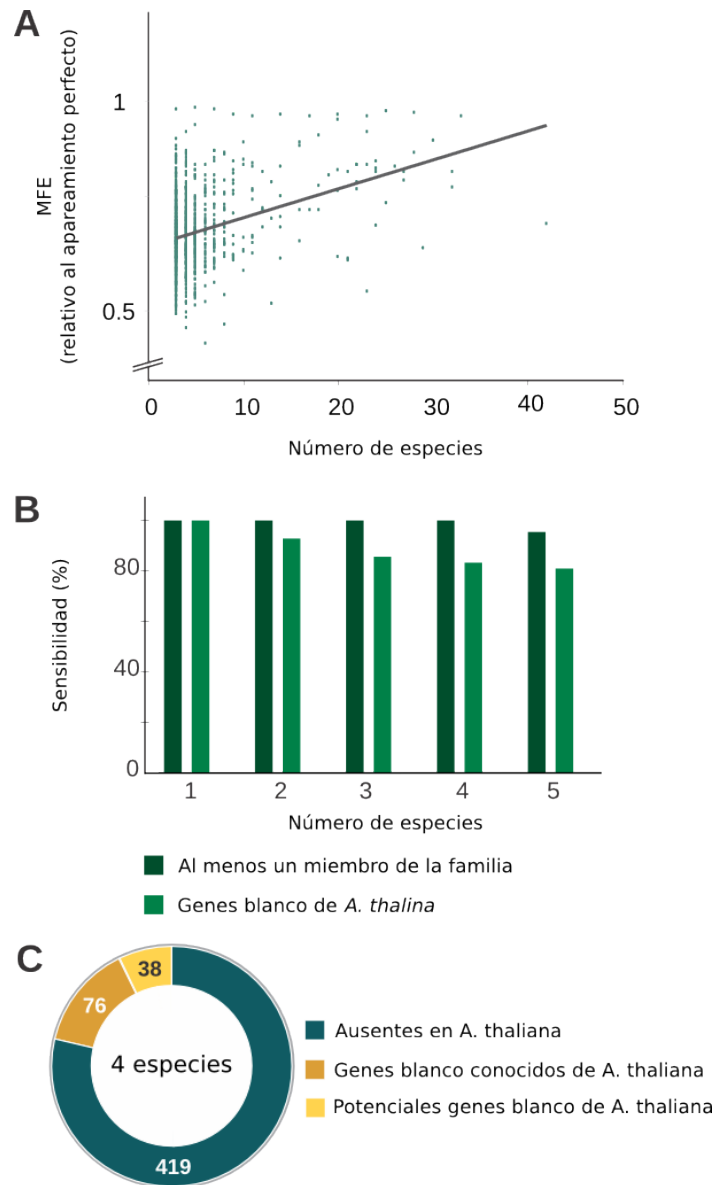


Fig. 1.3 Selección de genes blanco por conservación evolutiva de la secuencia. **(A)** Relación entre MFE y el número de especies en donde cada gen blanco fue detectado. **(B)** Sensibilidad de la estrategia, analizado de dos modos distinto. Verde clarito: evaluando la presencia de genes validados en *Arabidopsis* y en verde oscuro teniendo en cuenta la presencia de por lo menos un gen blanco de cada familia regulada por miARNs. **(C)** Clasificación de los potenciales genes blanco presentes en al menos 4 especies.

Table 1.2 Detection of miRNA targets using different filters

	Sin filtros			Filtros empíricos			Conservación 4 especies			Todos los filtros		
	miARN	scramble	ratio	miARN	scramble	ratio	miARN	scramble	ratio	miARN	scramble	ratio
miR156	3915	3994.4	± 149.9	1.0	890	704.7	± 45.2	1.3	34	39.7	± 3.1	0.9
miR159	1663	1283.7	± 47.8	1.3	472	254.9	± 21.9	1.9	20	10.1	± 1.1	2.0
miR160	793	695.6	± 30.5	1.1	277	157.5	± 28.8	1.8	5	4.4	± 0.9	1.1
miR162	1191	930.2	± 139.5	1.3	108	164.7	± 24.1	0.7	18	13.5	± 3.5	1.3
miR164	2486	1480.2	± 60.4	1.7	678	333.2	± 32.2	2.0	39	12.4	± 1.9	3.1
miR166	879	815.5	± 45.0	1.1	231	129	± 14.5	1.8	16	10.6	± 1.4	1.5
miR167	1777	1364.2	± 146.6	1.3	478	214.8	± 27.5	2.2	22	20.2	± 3.6	1.1
miR168	962	797.5	± 48.5	1.2	209	185	± 14.2	1.1	6	4.4	± 0.8	1.4
miR169	1540	1047.2	± 69.7	1.5	464	181.4	± 15.6	2.6	26	11.1	± 2.1	2.3
miR171	884	723.4	± 32.1	1.2	202	113.8	± 13.4	1.8	7	6.6	± 1.4	1.1
miR172	3007	1693.7	± 124.7	1.8	540	288.1	± 40.3	1.9	34	17.7	± 1.7	1.9
miR319	1363	1274.2	± 113.6	1.1	324	249.2	± 22.3	1.3	18	15	± 2.8	1.2
miR390	873	814.4	± 64.3	1.1	335	173	± 22.5	1.9	8	4.7	± 1.2	1.7
miR393	986	844.6	± 58.7	1.2	276	124.6	± 11.1	2.2	14	7.1	± 1.2	2.0
miR394	1569	1531.4	± 57.5	1.0	188	237.1	± 25.0	0.8	26	21.4	± 2.2	1.2
miR395	1472	1226.7	± 66.7	1.2	426	217.6	± 16.5	2.0	11	8.8	± 1.3	1.3
miR396	4641	2979.3	± 246.6	1.6	1246	390.5	± 38.8	3.2	92	51.4	± 5.9	1.8
miR397	1426	1050.9	± 27.9	1.4	368	236.5	± 23.5	1.6	26	9.7	± 0.8	2.7
miR398	935	834	± 34.5	1.1	376	144	± 18.1	2.6	11	7.5	± 1.6	1.5
miR399	1192	1137.6	± 72.0	1.0	275	207.8	± 24.9	1.3	5	13.6	± 1.7	0.4
miR408	2782	2502.9	± 103.6	1.1	695	468.7	± 50.8	1.5	51	35.1	± 3.0	1.5
miR827	2261	2000.1	± 119.8	1.1	317	297.1	± 45.0	1.1	44	23.4	± 3.9	1.9
Total	38597	31021.7	± 1859.8	1.2	9375	5473.2	± 576.3	1.7	533	348.4	± 47.0	1.5
Control												
miR158	1364	1462.8	± 69.1	0.9	170	208.7	± 15.8	0.8	15	16	± 1.7	0.9
miR173	1386	1232.1	± 101.7	1.1	243	215.6	± 23.4	1.1	11	12	± 2.4	0.9

a Sin filtros, búsqueda inicial utilizando los miARN consenso de 18nt y 3 mismatches.

b Filtros empíricos, energía de al menos 72% del apareamiento perfecto y 1 mismatch en la posición 2-12 del par miARN-gen blanco.

c Conservación del ID tag en al menos cuatro especies.

d Todos los filtros, combinación de los filtros empíricos y de conservación en al menos cuatro especies.

e miARN, genes blanco para cada miARN específico.

f scramble, promedio de los genes blanco de 10 versiones al azar de cada miARN ± error estándar.

### **Identificación de nuevos genes blanco en *A. thaliana* por conservación de la secuencia del gen blanco.**

Para encontrar nuevos genes blanco nos enfocamos en los genes potenciales que fueron seleccionados de nuestra estrategia utilizando solamente conservación evolutiva, debido a que los parámetros empíricos ya fueron utilizados extensamente en trabajos anteriores. [e.g.,([1],[6],[14])]. En primer lugar, analizamos la detección de genes blanco validados previamente en *A. thaliana* [basado en ([5])] usando nuestra estrategia y encontramos que el 84% de ellos estaban presentes en al menos 4 especies (Figura 1.3 B). Consideramos esto como un buen resultado ya puede ser que no todos los genes blanco de *Arabidopsis* estén conservados evolutivamente.

Generalmente los miARNs en plantas regulan genes que codifican para proteínas de la misma familia, es por esto que evaluamos si por lo menos un miembro de cada familia era detectado en nuestro enfoque. Encontramos genes blanco pertenecientes a casi todas las familias de genes codificantes para proteínas presentes en cuatro especies (Figura 1.3 B), con la excepción de TAS3, que es regulado por el miR390, al ser un ARN no codificante no es detectado por Blastx.

Para encontrar nuevos genes blanco regulados por miARNs, nos enfocamos únicamente en los potenciales genes blanco conservados en 4 especies, donde una de ellas es *A. thaliana* (Figura 1, bin 1). Genes blanco de miARNs que no están presentes en *A. thaliana* podrían incluir genes que perdieron su regulación durante la evolución o genes que hayan adquirido control por un miARN conservado más reciente en otras especies. La conservación en cuatro especies fue elegida como un filtro evolutivo porque provee buena sensibilidad para genes blanco conocidos.

Identificamos 114 potenciales genes que satisfacen este criterio. Donde 76 de ellos son genes validados anteriormente o genes muy relacionados (Figura 1.3 C). Curiosamente encontramos 38 genes que no tienen relación con genes blanco conocidos de miARNs y decidimos estudiar este grupo con mayor detalle. Nos enfocamos primero en los genes que estaban presentes en un gran número de especies para tener mejor especificidad (Figura 1.2) e intentamos validarlos utilizando 5' RACE PCR modificada ([8],[7]).

Un potencial gen blanco del MiR408 era At5g21930 que codifica para P-TYPE ATPase OF ARABIDOPSIS 2 (PAA2) y estaba presente en 22 especies distintas incluido monocotiledóneas y dicotiledóneas. MiR408 es inusual debido a que tiene un 5'-A, sin embargo >30% de las secuencias maduras del miR408 corresponden a una variante corrida 1 nt que empieza con 5'-U ([9]) (Figura 1.4 A). La validación experimental reveló fragmentos de ARNm compatible con este último sitio de corte (Figura 1.4 A). PAA2 es necesaria para

el transporte de iones de cobre a plastocianina ([11]), y su regulación por el miR408 está relacionada con el rol de este miARN en la homeostasis de cobre ([17]).

Otro potencial candidato del miR408 era At3g22110 que codifica para PROTEASOME ALPHA SUBUNIT C1 (PAC1) y estaba presente en 20 especies. Por medio de 5' RACE PCR demostramos que este gen es gen blanco del miR408 (Figura 1.4 A). Curiosamente la interacción del par miARN-gen blanco tiene 3 mismatches en la región 5', y se hubiera perdido como potencial gen blanco si se aplicaban solamente los filtros empíricos.

Luego estudiamos los genes blanco del miR396, donde los genes SVP y SUI1 estaban presentes en 29 y 19 especies respectivamente. Pero en ambos casos fallamos al obtener producto de la PCR utilizando 5' RACE PCR modificada. La falta de regulación de este gen por el miR396 podría estar relacionado a la débil energía de hibridación del par miARN-gen blanco, aunque no podemos descartar que el miR396 esté controlando su traducción.

Otros dos potenciales genes blanco del miR396 eran At5g43060 y At3g14110 que codifican para la proteasa MMG4.7 y FLUORESCENT IN BLUE LIGHT (FLU), respectivamente. Y en ambos casos pudimos detectar el corte (Figura 1.4 C y D).

En contraste con el miR408 y miR396, donde tienen varios potenciales genes blanco, obtuvimos un solo potencial gen blanco para el miR159, un factor de transcripción MYB que regula desarrollo del estambre y polen. ([10]) El otro potencial gen blanco era At4g27330, conocido como NOZZLE/SPOROXYLESS. Este factor de transcripción, que participa en desarrollo del estambre y óvulo [13, 18], fue también validado por 5' RACE PCR (Figura 1.4 E). Es interesante notar que al menos las funciones de NOZZLE y PAA2 pueden estar directamente relacionadas con el rol de genes blanco, ya descritos anteriormente, del miR159 y miR408 respectivamente.

PAA2, FLU y NOZZLE fueron detectados en mono y dicotiledóneas mientras que PAC1 y MMG4.7 fueron detectadas solamente en dicotiledóneas (Figura 1.4 A-E). Las posiciones del sitio de unión del miARN-gen blanco están altamente conservadas y muchas de las posiciones variables corresponden a mismatches con el miARN o variaciones del tipo G-C/G-U. Además este método no requiere que el sitio del gen blanco esté conservada, sino más bien que haya una interacción predicha con el miARN en distintas especies. De esta manera el sitio de NOZZLE, donde la secuencia cambia en diferentes especies (Figura 1.4 E), pudo ser detectado por este enfoque.

### **Identificación de nuevos genes blanco permitiendo interacciones G-U.**

Los genes blanco identificados utilizando la estrategia descrita anteriormente, tienen varios mismatches y bulges con sus miARNs, lo que puede ayudar a explicar por que se perdieron en trabajos anteriores. También notamos que muchas de estas nuevas interacciones miARN-gen



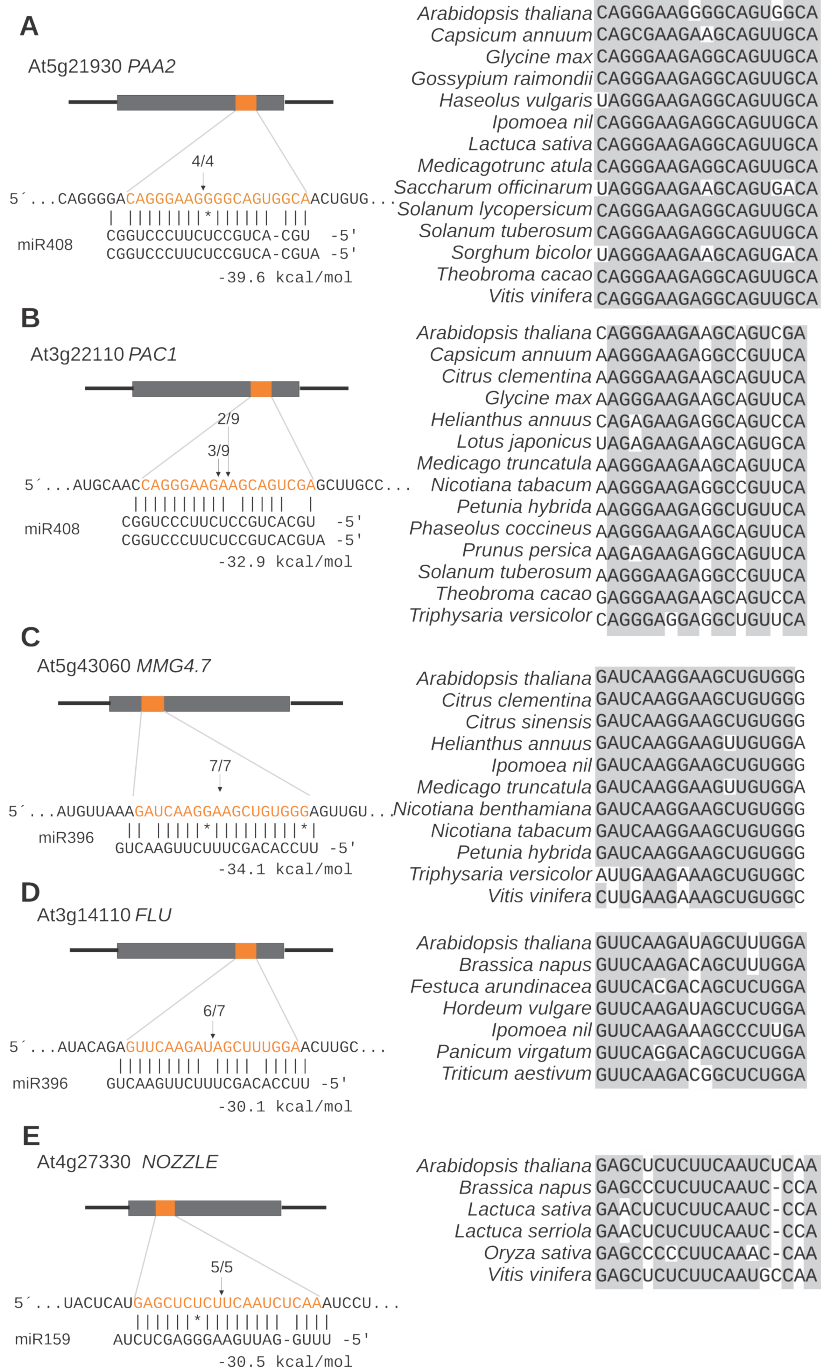


Fig. 1.4 Nuevos genes blanco validados en *A. thaliana*. El alineamiento entre el miARN y los nuevos genes blanco identificados se muestran en la izquierda. La conservación evolutiva de la secuencia del sitio reconocido por el miARN en las especies seleccionadas se muestra a la derecha. La figura muestra las interacciones del miR408 con PAA2 (A), miR408 con PAC1 (B), miR396 con MMG4.7 (C), miR396 con FLU (D), miR159 con NOZZLE (E). Las flechas marcan el sitio de corte determinado por 5'RACE-PCR y los números indican la frecuencia de clonado de cada fragmento.

blanco contenían posiciones que variaban alternadamente entre G-C y G-U en distintas especies. Como consideramos G-U como mismatch en nuestra búsqueda inicial, decidimos realizar nuevamente la búsqueda con los miARNs consenso de 18nt pero permitiendo ahora 4 mismatches, donde al menos uno de ellos tiene que ser del tipo G-U. Esta búsqueda permitiría interacciones miARN-gen blanco con sólo 14 bases apareadas perfectamente.

Para compensar el uso de estos parámetros relajados en términos de mismatches, queremos que el gen blanco aparezca en al menos 10 especies distintas para aumentar la especificidad (Figura 1.5 A). Encontramos 125 potenciales genes blanco en *A. thaliana* teniendo en cuenta este criterio (Figura 1.5 A) y 34 de ellos no aparecían en las búsquedas anteriores. El gen blanco CSD2 regulado por el miR398, que no apareció anteriormente, fue detectado con estos parámetros.

Luego encontramos que el miR167 que regula factores de respuesta a auxina (ARFs), también regulaba potencialmente a un gen denominado IAA-ALANINE RESISTANT 3 (IAR3) (Figura 1.5B y C), que está involucrado en el control de niveles libre de auxina ([3, 12]).

IAR3 en *Arabidopsis* tiene 3 mismatches con respecto al miR167, pero en la posición 12 de la interacción miARN-gen blanco, hay una interacción G-U en varias especies (Figura 5B y C). La técnica de 5' RACE PCR confirmó que el gen realmente era gen blanco del miR167 (Figura 5C).

### **Identificación de genes blanco específicos de *Solanaceae*.**

#### **1.2.2 comTAR: una herramienta para la predicción de genes blanco regulados por microARNs en plantas.**

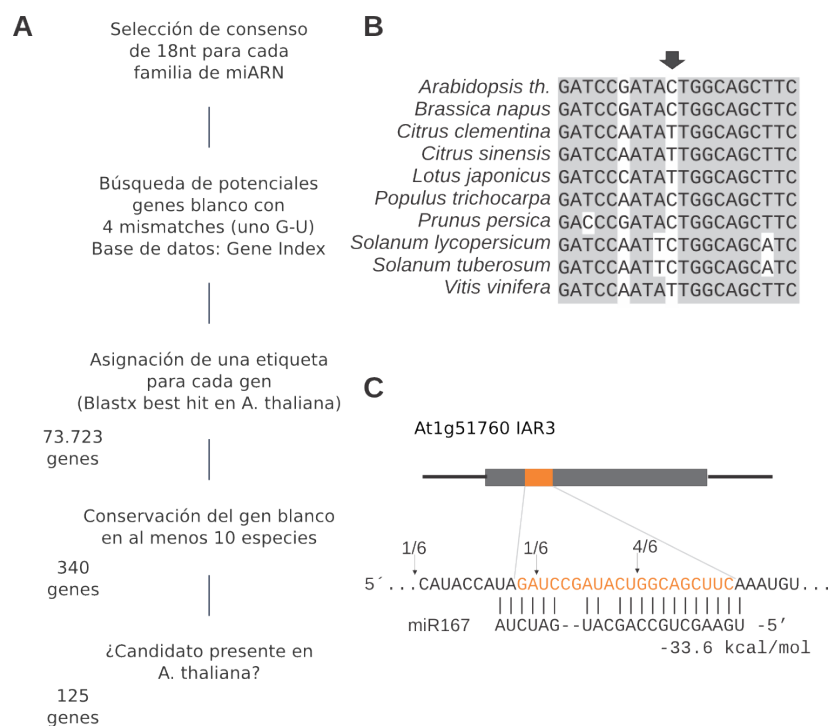


Fig. 1.5 Identificación de un nuevo gen blanco de miARN, relajando los parámetros de interacción pero incrementando el parámetro de conservación evolutiva. **(A)** Esquema de la estrategia modificada para identificar genes blanco de miARNs. **(B)** Conservación del sitio blanco reconocido por el miARN en distintas especies. La flecha indica una variación de G-C o G-U con el miARN dependiendo de la especie. **(C)** Alineamiento en *Arabidopsis thaliana* del gen blanco IAR3 con el miR167. La flecha indica la posición del corte indicada por 5'RACE-PCR y el número indica la frecuencia de clonado de cada fragmento.

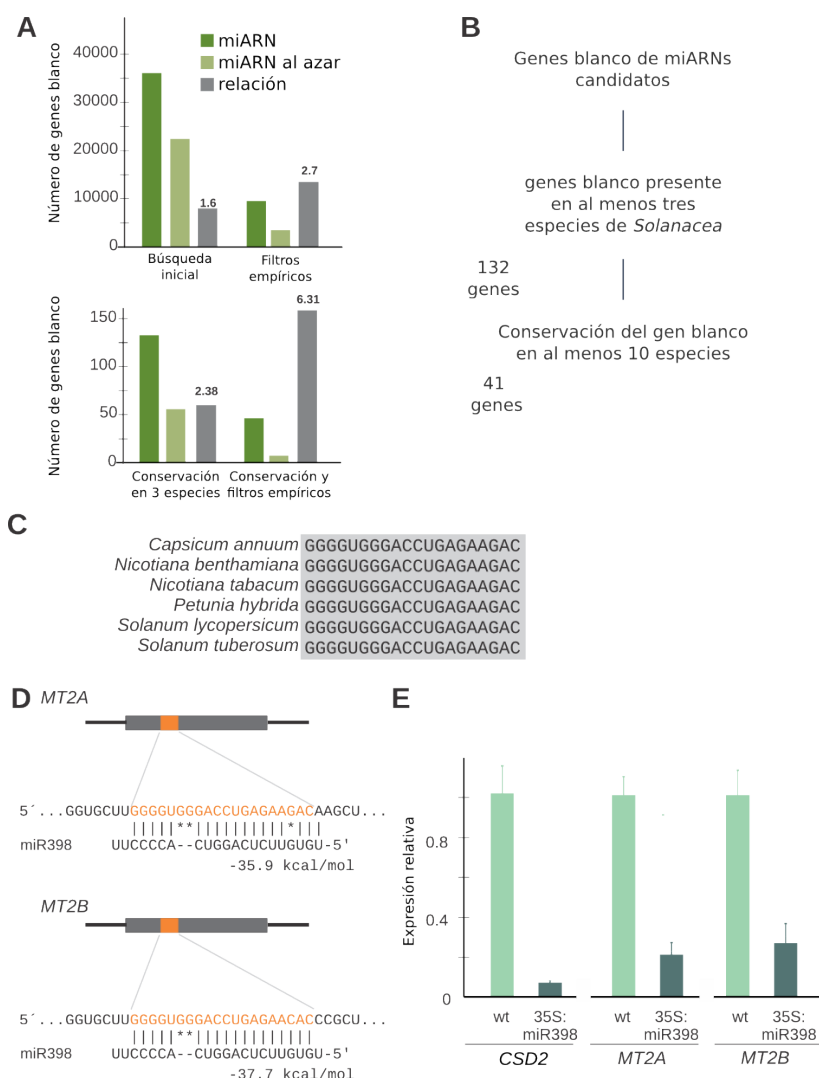


Fig. 1.6 Identificación de genes blanco de miARN, específicos de *Solanaceae*. (A) Predicción de genes blanco de miARN en cinco especies de *Solanaceae*. El número de genes blanco de todos los miARNs conservados juntos se muestra luego de aplicar distintos filtros. También se muestran los genes blanco obtenidos a partir de las secuencias al azar. (B) Esquema que muestra la estrategia para identificar genes blanco específicos de *Solanaceae*. (C) Conservación del sitio reconocido por el miR398 con MT2A específico de *Solanaceae*. (D) Esquema que muestra el sitio de unión entre el miR398 y MT2A y MT2B. (E) Niveles de transcritos de CSD2, MT2A y MT2B en plantas salvajes y plantas transgénicas de tabaco (cv Petit havana) que sobreexpresan el miR398.

## **Capítulo 2**

### **My second chapter**

#### **2.1 Enfoque bioinformático para el estudio de la evolución y biogénesis de microARNs en plantas.**



# References

- [1] Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, 121(2):207 – 221.
- [2] Cuperus, J. T., Fahlgren, N., and Carrington, J. C. (2011). Evolution and functional diversification of miRNA genes. *The Plant cell*, 23(2):431–442.
- [3] Davies, R. T., Goetz, D. H., Lasswell, J., Anderson, M. N., and Bartel, B. (1999). IAR3 Encodes an Auxin Conjugate Hydrolase from Arabidopsis. 11(March):365–376.
- [4] Debernardi, J. M., Rodriguez, R. E., Mecchia, M. A., and Palatnik, J. F. (2012). Functional Specialization of the Plant miR396 Regulatory Network through Distinct MicroRNA–Target Interactions. *PLoS Genet*, 8(1):e1002419.
- [5] Fahlgren, N., Jogdeo, S., Kasschau, K. D., Sullivan, C. M., Chapman, E. J., Laubinger, S., Smith, L. M., Dasenko, M., Givan, S. a., Weigel, D., and Carrington, J. C. (2010). MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. *The Plant cell*, 22(4):1074–89.
- [6] Jones-Rhoades, M. W. and Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell*, 14(6):787 – 799.
- [7] 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. *Developmental Cell*, 4(2):205–217.
- [8] 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. 297(September):2053–2056.
- [9] Maunoury, N. (2011). AGO1 and AGO2 Act Redundantly in miR408-Mediated Plantacyanin Regulation. 6(12).
- [10] Millar, A. a. and Gubler, F. (2005). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *The Plant cell*, 17(3):705–21.
- [11] Niyogi, K. K., Pilon, M., Shikanai, T., Abdel-ghany, S. E., and Mu, P. (2005). Two P-Type ATPases Are Required for Copper Delivery in Arabidopsis thaliana Chloroplasts. 17(April):1233–1251.

- [12] Rampey, R. A., Leclere, S., Kowalczyk, M., Ljung, K., Bartel, B., Biology, C., and Texas, R. A. R. (2004). A Family of Auxin-Conjugate Hydrolases That Contributes to Free Indole-3-Acetic Acid Levels during Arabidopsis Germination 1. 135(June):978–988.
- [13] Schiefthaler, Balasubramanian, Sieber, Chevalier, Wisman, and Schneitz (1999). Molecular analysis of NOZZLE , a gene involved in pattern formation and early sporogenesis during sex organ development in Arabidopsis thaliana. *Proc. Natl Acad. Sci.*, 96(September):11664–11669.
- [14] Schwab, R., Palatnik, J. F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of micornas on the plant transcriptome. *Developmental Cell*, 8(4):517 – 527.
- [15] Vazquez, F., Gascioli, V., Cr  t  , P., and Vaucheret, H. (2004). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Current biology : CB*, 14(4):346–51.
- [16] Xie, Z., Kasschau, K. D., and Carrington, J. C. (2003). Negative Feedback Regulation of Dicer-Like1 in Arabidopsis by microRNA-Guided mRNA Degradation. *Current Biology*, 13(9):784–789.
- [17] Yamasaki, H., Abdel-Ghany, S. E., Cohu, C. M., Kobayashi, Y., Shikanai, T., and Pilon, M. (2007). Regulation of copper homeostasis by micro-RNA in Arabidopsis. *The Journal of biological chemistry*, 282(22):16369–78.
- [18] Yang, W.-c., Ye, D., Xu, J., and Sundaresan, V. (1999). The SPOROCTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. pages 2108–2117.



# Appendix A

## Anexo

### **Predicción de genes regulados por microARNs.**

#### **For 64bit OS**

```
edit ~/.bashrc file and add following lines
PATH=/usr/local/texlive/2011/bin/x86_64-linux:$PATH;
export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH;
export MANPATH
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

Table A.1 Especies y base de datos utilizadas para la búsquedas de genes blanco de miARNs conservados

Specie	Database
Allium_cepa	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Aquilegia	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Arabidopsis_thaliana	<a href="http://arabidopsis.org/">http://arabidopsis.org/</a>
Beta_vulgaris	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Brassica napus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Capsicum_annuum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Citrus_clementina	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Citrus_sinensis	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Coffea_canephora	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Euphorbia_esula	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Festuca_arundinacea	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Glycine_max	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Gossypium	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Gossypium_raidmondii	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Haseolus_vulgaris	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Helianthus_annuus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Hordeum_vulgare	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Ipomoea_nil	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Lactuca_sativa	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Lactuca_serriola	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Lotus_japonicus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Malus_x_domestica	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Medicago_truncatula	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Mesembryanthemum_crystallinum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Nicotiana_benthiana	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Nicotiana_tabacum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Oryza_sativa	<a href="http://www.jcvi.org/">http://www.jcvi.org/</a>
Panicum_virgatum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Petunia_hybrida	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Phaseolus_coccineus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Populus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Prunus_persica	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Saccharum_officinarum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Secale_cereale	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Solanum_lycopersicum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Solanum_tuberosum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Sorghum_bicolor	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Theobroma_cacao	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
triphysaria	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Triphysaria_versicolor	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Triticum_aestivum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Vitis_vinifera	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Zea_mays	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>

Table A.2 My caption

microRNA	Target	ID
miR156/miR157	SPL	At1g27370
miR156/miR157	SPL	At1g53160
miR156/miR157	SPL	At2g33810
miR156/miR157	SPL	At3g15270
miR156/miR157	SPL	At5g43270
miR156/miR157	SPL	At1g69170
miR156/miR157	SPL	At2g42200
miR156/miR157	SPL	At3g57920
miR156/miR157	SPL	At5g50670
miR159/miR319	TCP	At1g30210
miR159/miR319	TCP	At1g53230
miR159/miR319	TCP	At2g31070
miR159/miR319	MYB	At3g11440
miR159/miR319	TCP	At3g15030
miR159/miR319	TCP	At4g18390
miR159/miR319	MYB	At5g06100
miR159/miR319	MYB	At2g26950
miR159/miR319	MYB	At2g32460
miR159/miR319	MYB	At5g55020
miR160	ARF	At1g77850
miR160	ARF	At2g28350
miR160	ARF	At4g30080
miR162	DCL	At1g01040
miR164	NAC	At1g56010
miR164	NAC	At3g15170
miR164	NAC	At5g07680
miR164	NAC	At5g53950
miR164	NAC	At5g61430
miR164	NAC	At3g12977
miR164	NAC	At5g39610
miR166/miR165	HD-ZIPIII	At1g30490
miR166/miR165	HD-ZIPIII	At1g52150
miR166/miR165	HD-ZIPIII	At2g34710
miR166/miR165	HD-ZIPIII	At5g60690
miR166/miR165	HD-ZIPIII	At4g32880
miR167	ARF	At1g30330
miR167	ARF	At5g37020
miR168	AGO	At1g48410
miR169	HAP2	At1g17590
miR169	HAP2	At1g54160
miR169	HAP2	At1g72830
miR169	HAP2	At3g05690
miR169	HAP2	At3g20910
miR169	HAP2	At5g06510
miR170/miR171	SCL	At2g45160
miR170/miR171	SCL	At3g60630
miR170/miR171	SCL	At4g00150
miR172	AP2	At2g28550
miR172	AP2	At4g36920
miR172	AP2	At5g60120
miR172	AP2	At5g67180
miR172	AP2	At2g39250
miR172	AP2	At3g54990
miR390/miR391	TAS3	At3g17185
miR390/miR391	TAS3	At5g49615
miR390/miR391	TAS3	At5g57735
miR393	TIR1/AFB	At1g12820
miR393	bHLH	At3g23690
miR393	TIR1/AFB	At3g26810
miR393	TIR1/AFB	At3g62980
miR393	TIR1/AFB	At4g03190
miR394	F-Box	At1g27340
miR395	APS	At3g22890
miR395	AST	At5g10180
miR395	APS	At5g43780
miR395	APS	At4g14680
miR396	GRF	At2g22840
miR396	GRF	At2g36400
miR396	GRF	At2g45480
miR396	GRF	At4g24150
miR396	GRF	At4g37740
miR396	GRF	At5g53660
miR396	GRF	At3g52910
miR397	LAC	At2g29130
miR397	LAC	At2g38080
miR397	LAC	At5g60020
miR398	CSD	At1g08830
miR398	CSD	At2g28190
miR398	CytC oxidase	At3g15640
miR399	E2-UBC	At2g33770
miR399	E2-UBC	At2g33770
miR408	LAC	At2g30210
mir408	PLC	At2g02850
miR827	SPX	At1g02860

