

Nuevos retrotransposones de naranjo.  
Desarrollo de marcadores basados en  
retrotransposones para la evaluación de  
la biodiversidad de los cítricos.

Laura Rico Cabanas  
Barcelona, octubre de 2005



Departamento de Bioquímica y Biología Molecular  
Facultad de Farmacia

Memoria presentada por Laura Rico Cabanas, licenciada en Biología, inscrita en el programa de Doctorado del Departamento de Bioquímica y Biología Molecular de la Universidad de Barcelona en el bienio 2000-2002 para aspirar al grado de Doctor por la Universidad de Barcelona.

El presente trabajo ha sido realizado en el Laboratori de Genètica Molecular Vegetal del consorci CSIC-IRTA, bajo la dirección del Dr. José Antonio Martínez Izquierdo.

Becada por la Generalitat de Catalunya con la beca predoctoral para la formación del personal investigador (FI/FIAP) 2001FI 00143.

Director

Tutora

Autora

Dr. José Antonio Martínez Izquierdo

Dra. Montserrat Arró

Laura Rico Cabanas

# **INTRODUCCIÓN** **1**

<b>1.</b>	<b><i>LOS CÍTRICOS</i></b>	<b>1</b>
1.1	<i>ORIGEN DEL GÉNERO CITRUS</i>	1
1.2	<i>IMPORTANCIA ECONÓMICA Y DISTRIBUCIÓN GEOGRÁFICA DE LOS CÍTRICOS</i>	1
1.3	<i>CARACTERÍSTICAS MORFOLÓGICAS, REPRODUCTIVAS Y GENÉTICAS DEL GÉNERO</i>	1
1.3.1	Morfología de los cítricos	1
1.3.2	Propagación y reproducción de los cítricos	2
1.3.3	Genoma de los cítricos	3
1.4	<i>TAXONOMÍA DEL GÉNERO CITRUS</i>	4
<b>2.</b>	<b><i>COMPONENTES DE LOS GENOMAS</i></b>	<b>5</b>
<b>3.</b>	<b><i>LOS ELEMENTOS TRANSPONIBLES</i></b>	<b>6</b>
3.1	<i>CARACTERÍSTICAS ESTRUCTURALES Y CLASIFICACIÓN DE LOS ELEMENTOS TRANSPONIBLES</i>	7
3.1.1	Elementos de Clase I ( Retroelementos)	8
3.1.2	Elementos de Clase II (Transposones)	10
3.2	<i>IMPORTANCIA DE LOS ELEMENTOS TRANSPONIBLES EN LOS GENOMAS EUCARIOTAS</i>	12
3.2.1	Contribución de los retrotransposones al genoma de plantas	13
3.2.2	Distribución de los retrotransposones en el genoma de plantas	13
3.2.3	Variabilidad de los retrotransposones en el genoma de plantas	14
3.2.4	Impacto de los retrotransposones en la evolución del genoma de plantas	15
<b>4.</b>	<b><i>LOS RETROTRANSPOSONES-LTR</i></b>	<b>18</b>
4.1	<i>CARACTERÍSTICAS ESTRUCTURALES DE LOS RTN-LTR</i>	18
4.1.1	LTR	18
4.1.2	PBS ( <i>Primer Binding Site</i> ) y PTT ( <i>Poly Purine Track</i> )	20

4.1.3	DOMINIO INTERNO	20
o	El gen <i>gag</i>	21
o	El gen <i>pol</i>	21
o	El gen <i>env</i>	24
4.2	CICLO DE VIDA DE LOS RTN-LTR	25
4.3	RETROTRANSPOSONES-LTR DE PLANTAS CARACTERIZADOS	27
4.4	ACTIVIDAD DE LOS RTN-LTR	29
4.4.1	Regulación de la expresión y la transposición en RTN-LTR	29
o	Regulación a nivel transcripcional	29
o	Activación de la transcripción en condiciones de estrés biótico y abiótico	30
o	Secuencias reguladoras de la LTR del RTN	31
o	Metilación del DNA	33
o	Regulación a nivel post-transcripcional	33
4.4.2	Evidencias de la posible movilización de RTN-LTR	34
4.5	USO DE LOS RTN-LTR COMO HERRAMIENTAS MOLECULARES	35
<b>5</b>	<b>LOS MARCADORES</b>	<b>36</b>
5.1	MARCADORES MOLECULARES	37
5.2	LOS RTN-LTR COMO MARCADORES MOLECULARES	41
5.2.1	Marcadores moleculares SSAP	43
5.3	UTILIDAD DE LOS MARCADORES MOLECULARES	43
5.4	APLICACIÓN DE LOS MARCADORES MOLECULARES AL ESTUDIO DE LOS CÍTRICOS	45
5.4.1	Taxonomía del género <i>Citrus</i>	45
5.4.2	Origen de las variedades de cítricos comerciales	46

---

**OBJETIVOS** **49**

## **RESULTADOS 51**

### **A. CARACTERIZACIÓN DE RETROTRANSPOSONES DEL GENOMA DE CITRUS SINENSIS 51**

#### **A.1 OBTENCIÓN DE FRAGMENTOS DE RTN-LTR TIPO COPIA DEL GENOMA DE CITRUS SINENSIS 51**

- A.1.1 Estrategia de amplificación selectiva de RTN-LTR tipo *copia* 51
- A.1.2 Conservación de los dominios funcionales en los elementos CoSi 52
- A.1.3 Descripción de nuevas familias de RTN tipo *copia* en naranjo 53
- A.1.4 Estimación del número de copias de RTN tipo *copia* en el genoma de naranjo 55

#### **A.2 OBTENCIÓN DE FRAGMENTOS DE RTN-LTR TIPO GYPSY EN EL GENOMA DE CITRUS SINENSIS 57**

- A.2.1 Estrategia de amplificación selectiva de RTN-LTR tipo *gypsy* 57
- A.2.2 Conservación de los dominios funcionales en los elementos GySi 58
- A.2.3 Descripción de nuevas familias de RTN tipo *gypsy* en naranjo 60
- A.2.4 Estimación del número de copias de los RTN tipo *gypsy* en el genoma de naranjo 62

#### **A.3 CLONAJE, SECUENCIACIÓN Y CARACTERIZACIÓN DE UN NUEVO RTN TIPO COPIA DE NARANJO (CIRE1) 63**

- A.3.1 Estrategia de clonaje 63
  - Amplificación por PCR inversa de CIRE1 sobre DNA genómico 63
  - Amplificación por PCR directa de CIRE1 sobre DNA genómico 65
  - Localización del posible inicio de la LTR5' 68
  - Subclonaje, secuenciación y ensamblaje de las regiones amplificadas 72
- A.3.2 Análisis de las secuencias de CIRE1 clonadas del genoma de *C. sinensis* 73
  - Comparativa de secuencias 73
- A.3.3 Establecimiento de la secuencia completa de un elemento: CIRE1-1 77
- A.3.4 Integridad de los elementos CIRE1 en el genoma de *C. sinensis* 80
- A.3.5 Conservación de los dominios funcionales en CIRE1 81

A.3.6	Análisis filogenético de CIRE1 _____	84
A.3.7	Estimación del número de copias de CIRE1 _____	86
A.3.8	Presencia de CIRE1 en el genoma de los cítricos y otras especies vegetales _	88
○	Amplificación selectiva de CIRE1 por PCR en la familia Rutaceae _____	88
○	Amplificación selectiva de CIRE1 por PCR en otras especies vegetales _____	89
○	Análisis genómico por <i>Southern-blot</i> de CIRE1 en la familia Rutaceae _____	90
○	Análisis genómico por <i>Southern-blot</i> de CIRE1 en la familia Rutaceae y otras familias _____	91
<b>A.4 ESTUDIO DE LA FUNCIONALIDAD DEL PROMOTOR DE CIRE1 _____</b>		<b>92</b>
A.4.1	Análisis de la secuencia del promotor de la LTR de CIRE1 _____	92
A.4.2	Actividad del promotor de la LTR de CIRE1 _____	94
○	Experimentos de expresión transitoria en naranjo _____	94
○	Experimentos de expresión transitoria en células de epidermis de cebolla _____	95
○	Experimentos de expresión transitoria en suspensiones celulares de tabaco _____	97
<b>A.5 EXISTENCIA DE RTN TRANSCRIPCIONALMENTE ACTIVOS EN CITRUS SINENSIS _____</b>		<b>99</b>
A.5.1	Actividad transcripcional de RTN-LTR tipo <i> copia </i> en naranjo _____	99
○	Amplificación de transcritos de RTN-LTR tipo <i> copia </i> _____	99
○	Descripción de las familias de RTN-LTR tipo <i> copia </i> transcripcionalmente activas en naranjo _____	100
○	Comparación de la variabilidad de la población de secuencias de transcritos y de secuencias genómicas correspondientes a RTN-LTR tipo <i> copia </i> _____	102
○	Estudio de la expresión de RTN-LTR tipo <i> copia </i> en naranjo _____	104
A.5.2	Análisis de la transcripción de CIRE1 _____	106
○	Implicación de la herida mecánica y la adición exógena de las hormonas vegetales (JA y NAA) en la activación de la expresión de CIRE1 en <i>C. sinensis</i> _____	106

**B. ANÁLISIS DE LA BIODIVERSIDAD EN EL GÉNERO CITRUS UTILIZANDO RTN-LTR \_\_\_\_\_ 113**

**B.1 VARIABILIDAD DE LOS RTN-LTR TIPO COPIA EN EL GENOMA DE ESPECIES DEL GÉNERO CITRUS \_\_\_\_\_ 113**

B.1.1 Detección de fragmentos de RTN-LTR tipo *copia* en cítricos \_\_\_\_\_ 114

B.1.2 Análisis comparativo de las secuencias de RTN-LTR tipo *copia* en *Citrus* \_\_\_\_ 116

B.1.3 Estimación del número de copias de RTN-LTR tipo *copia* en cítricos \_\_\_\_\_ 123

**B.2 DESARROLLO DE MARCADORES MOLECULARES BASADOS EN LTR-RTN: TÉCNICA SSAP APLICADA A ESTUDIOS FILOGENÉTICOS \_\_\_\_\_ 125**

B.2.1 Aproximaciones experimentales utilizadas \_\_\_\_\_ 127

B.2.2 Análisis de los geles SSAP marcados con  $\gamma$ -p<sup>33</sup> obtenidos \_\_\_\_\_ 130

B.2.3 Tratamiento de datos SSAP \_\_\_\_\_ 135

B.2.4 Relaciones filogenéticas entre cítricos inferidas a partir de marcadores SSAP 136

**C. TÉCNICA SSAP PARA LA DETECCIÓN DE LA MOVILIDAD DE RTN-LTR \_\_\_\_\_ 139**

C.1 ESTUDIO DE LA ACTIVIDAD TRANSPOSICIONAL DE RTN COPIA EN NARANJO \_\_\_\_ 139

C.2 ANÁLISIS DE LAS ZONAS FLANQUEANTES A CIRE1 \_\_\_\_\_ 142

**DISCUSIÓN \_\_\_\_\_ 143**

**1. CARACTERIZACIÓN DE RTN-LTR EN CITRUS SINENSIS \_\_\_\_ 143**

1.1 Amplificación eficaz de poblaciones de RTN-LTR en *Citrus sinensis* usando cebadores degenerados \_\_\_\_\_ 143

1.2 Conservación a nivel de secuencia de los dominios funcionales en los fragmentos de retroelementos amplificados \_\_\_\_\_ 144

1.3 Descripción de las poblaciones de RTN-LTR presentes en *Citrus sinensis* \_\_\_\_\_ 145

1.4 Contribución de los RTN-LTR al genoma de *Citrus sinensis* \_\_\_\_\_ 146

<b>2. UN NUEVO RTN-LTR TIPO COPIA DEL GENOMA DE NARANJO: CIRE1</b>	<b>148</b>
2.1 <i>Análisis de los motivos estructurales y funcionales de CIRE1</i>	148
2.2 <i>Contribución de CIRE1 al genoma de Citrus sinensis</i>	150
2.3 <i>Presencia de CIRE1 en otras especies vegetales</i>	151
2.4 <i>Funcionalidad del Promotor de la LTR de CIRE1</i>	152
2.5 <i>Actividad transcripcional de RTN-LTR tipo copia en naranjo</i>	153
<b>3. ANÁLISIS DE LA BIODIVERSIDAD EN EL GÉNERO CITRUS UTILIZANDO RTN-LTR</b>	<b>156</b>
3.1 <i>Descripción de las poblaciones RTN-LTR tipo copia presentes en cítricos</i>	156
3.2 <i>Contribución de los RTN-LTR tipo copia al genoma de especies del género Citrus</i>	158
3.3 <i>Aplicación de la técnica SSAP basada en RTN-LTR para estudios filogenéticos dentro del género Citrus</i>	160
<b>4. APLICACIÓN DE LA TÉCNICA SSAP PARA LA DETECCIÓN DE LA MOVILIDAD DE RTN-LTR</b>	<b>162</b>
<b><u>CONCLUSIONES</u></b>	<b>165</b>
<b><u>MATERIAL Y MÉTODOS</u></b>	<b>167</b>
<b>1 MATERIAL BIOLÓGICO</b>	<b>167</b>
1.1 <i>Material vegetal</i>	167

1.2 Cepas bacterianas _____	167
<b>2 PLÁSMIDOS Y CONSTRUCCIONES _____</b>	<b>167</b>
<b>3 OBTENCIÓN DE DNA Y RNA _____</b>	<b>168</b>
3.1 Obtención de DNA a partir de tejidos vegetales _____	168
3.1.1 Extracción DNA genómico con DNAeasy Plant Mini Kit de Qiagen _____	168
3.1.2 Extracción DNA genómico mediante Solución MATAB _____	169
3.1.3 Extracción DNA genómico ultra-rápida _____	169
3.2 Obtención de DNA plasmídico _____	169
3.2.1 Minipreparación mediante lisis alcalina _____	169
3.2.2 Minipreparación con <i>NucleoSpin plasmid</i> de Macherey-Nagel _____	170
3.2.3 Midipreparación con <i>Qiagen Plasmid Midi kit</i> de Qiagen _____	170
3.3 Obtención de RNA total a partir de tejidos vegetales _____	170
3.3.1 Extracción RNA total con <i>RNAeasy Plant Mini Kit</i> de Qiagen _____	170
3.3.2 Extracción RNA total con Trizol <sup>®</sup> _____	171
3.3.3 Extracción RNA total mediante el tampón Z6 _____	171
3.3.4 Tratamiento del RNA con DNAsa _____	171
<b>4 MODIFICACIONES GENERALES DEL DNA _____</b>	<b>171</b>
4.1 Digestiones con enzimas de restricción y ligaciones _____	171
4.2 Defosforilación de extremos _____	172
<b>5 PURIFICACIÓN DE FRAGMENTOS DE DNA _____</b>	<b>172</b>
5.1 Extracción con Fenol:Cloroformo _____	172
5.2 Precipitación de ácidos nucleicos _____	173
5.3 Purificación de DNA a partir de geles de agarosa o de solución _____	173
<b>6 SUBCLONAJE DE FRAGMENTOS DE DNA EN PLÁSMIDOS _____</b>	<b>173</b>
6.1 Preparación de células competentes de <i>E.coli</i> DH5 $\alpha$ _____	173
6.2 Transformación de células competentes _____	173
<b>7 REACCIÓN EN CADENA DE LA POLIMERASA (PCR) _____</b>	<b>174</b>
7.1 Observaciones generales sobre la PCR _____	174
7.2 Condiciones y programas específicos de PCR utilizados en este trabajo _____	177
7.2.1 Amplificación específica de RTN-LTR tipo <i>Ty1-copia</i> _____	177
7.2.2 Amplificación específica de RTN-LTR de tipo <i>Ty3-gypsy</i> _____	177
7.2.3 PCR de colonias de <i>E.coli</i> transformadas _____	178

<b>8 PCR INVERSA (INVERSE PCR, i-PCR)</b>	<b>178</b>
8.1 Digestión simple de DNA genómico con enzimas de restricción	179
8.2 Ligación	179
8.3 Amplificación por i-PCR	179
8.3.1 Diseño de los cebadores para la amplificación	179
8.3.2 Reacción de PCR	180
<b>9 RT-PCR</b>	<b>180</b>
9.1 Retrotranscripción (RT)	180
9.2 Reacción de PCR sobre cDNA (RT-PCR)	181
<b>10 MARCAJE RADIATIVO DE DNA</b>	<b>181</b>
10.1 Marcaje por <i>Random priming</i>	181
10.2 Marcaje por PCR	181
<b>11 TRANSFERENCIA E HIBRIDACIÓN DE ÁCIDOS NUCLEICOS</b>	<b>182</b>
11.1 Hibridación Southern blot	182
11.1.1 Preparación, electroforesis y tratamiento del gel de DNA	182
11.1.2 Transferencia del gel de DNA	182
11.1.3 Prehibridación e Hibridación de la membrana	183
11.1.4 Lavados	184
11.1.5 Exposición	184
11.2 Hibridación Northern blot	184
11.2.1 Preparación, electroforesis y tratamiento del gel de RNA	184
11.2.2 Transferencia del gel de RNA	184
11.2.3 Prehibridación e Hibridación de la membrana	185
11.2.4 Lavados	185
11.2.5 Exposición	185
11.3 Hibridación Slot blot	186
11.4 Deshibridar un filtro hibridado con DNA/RNA	186
<b>12 TÉCNICA SSAP DE MARCADORES MOLECULARES</b>	<b>186</b>
12.1 Digestión simple del DNA genómico	187
12.2 Ligación de los adaptadores	187
12.2.1 Preparación de los adaptadores	187
12.2.2 Ligación de los adaptadores	188

12.3	Preamplificación	188
12.4a	SSAP-PCR no radiactiva	188
12.5a	Electroforesis en gel poliacrilamida 6%	189
12.5a.1	Limpieza de los vidrios	189
12.5a.2	Montaje de los vidrios	189
12.5a.3	Preparación del gel de poliacrilamida 6%	190
12.5a.4	Electroforesis en TBE 1x	190
12.5a.5	Fijación y tinción con plata	191
12.6a	Purificación de fragmentos de PCR a partir de geles de poliacrilamida teñidos con plata	191
12.4b	Marcaje radiactivo con $\gamma$ -p <sup>33</sup>	192
12.4b.1	Marcaje radiactivo del cebador específico de la LTR de CIRE1	192
12.4b.2	Marcaje radiactivo del marcador de peso molecular	193
12.5b	SSAP-PCR radiactiva	193
12.6b	Electroforesis en gel poliacrilamida 6%	194
12.6b.1	Limpieza de los vidrios	194
12.6b.2	Montaje de los vidrios	194
12.6b.3	Preparación del gel de poliacrilamida 6%	194
12.6b.4	Electroforesis en TBE 0,5x	195
12.6b.5	Tratamiento del gel SSAP radiactivo	195
<b>13</b>	<b>DETERMINACIÓN DE LA EXPRESIÓN DE GUS DE MUESTRAS TRANSFORMADAS</b>	<b>195</b>
13.1	Preparación de muestras vegetales para el bombardeo	195
13.1.1	Preparación de Epidermis de cebolla	195
13.1.2	Preparación de Suspensiones de tabaco W38	196
13.2	Transferencia de DNA mediante bombardeo	196
13.2.1	Preparación de las partículas de oro	196
13.2.2	Recubrimiento de las partículas de oro con DNA	197
13.2.3	Bombardeo	197
13.3	Determinación de la actividad $\beta$ -glucuronidasa	198
13.3.1	Preparación de extractos proteicos	198
13.3.2	Determinación de la concentración de proteínas del extracto	198
13.3.3	Ensayo de la actividad GUS por fluorimetría	199

13.3.4	Análisis histoquímica por tinción con X-Gluc	200
<b>14</b>	<b>GERMINACIÓN Y MULTIPLICACIÓN IN VITRO DE CITRUS</b>	<b>200</b>
14.1	Germinación a partir de semilla	200
14.1.1	Germinación a partir de semilla en invernadero	200
14.1.2	Germinación a partir de semilla <i>in vitro</i>	201
14.2	Multipliación <i>in vitro</i> de Citrus	202
<b>15</b>	<b>ANÁLISIS DE DATOS</b>	<b>204</b>
15.1	Análisis de secuencias	204
15.2	Análisis de datos SSAP	204
15.3	Estimación del números de copias por <i>Slot-blot</i>	205
<b>16</b>	<b>SECUENCIAS DEL EMBL</b>	<b>205</b>
	<b><u>BIBLIOGRAFÍA</u></b>	<b>207</b>

# INDICE DE FIGURAS

---

## INTRODUCCIÓN

<b>Figura 1.</b> (A) Hesperidio o fruto de los cítricos (B) Flor de Azahar _____	2
<b>Figura 2.</b> Esquema de la estructura del DNA y la clasificación del DNA según su número de copias _	5
<b>Figura 3.</b> Descubrimiento de los TE en plantas _____	6
<b>Figura 4.</b> Clasificación de los elementos transponibles (TE) según su modo de transposición _____	7
<b>Figura 5.</b> Clasificación de los retroelementos y virus relacionados _____	8
<b>Figura 6.</b> Organización estructural general de los diferentes tipos de TE en plantas _____	11
<b>Figura 7.</b> Organización de las LTR de un RTN _____	19
<b>Figura 8.</b> Organización estructural de un RTN-LTR _____	21
<b>Figura 9.</b> Ciclo de vida de los RTN-LTR _____	26
<b>Figura 10.</b> Representación gráfica de algunos de los principales tipos de marcadores moleculares ___	39
<b>Figura 11.</b> Esquema de las estrategias utilizadas en las técnicas de marcadores moleculares basados en RTN-LTR _____	42
<b>Figura 12.</b> Dendograma de 85 especies de <i>Citrus</i> y géneros relacionados basado en datos RFLP extraído de Federici et al. (1998) _____	46
<b>Figura 13.</b> Representación de las relaciones más probables entre los cítricos más comunes, según Moore (2001) _____	48

## RESULTADOS

<b>Figura 14.</b> Obtención de fragmentos de RTN <i> copia</i> del genoma de <i>C. sinensis</i> _____	51
<b>Figura 15.</b> Alineamiento múltiple por ClustalW de las secuencias aminoacídicas inferidas de los clones CoSi# y de zonas similares de RTN tipo <i> copia</i> ya descritos como son <i>Tnt1</i> de tabaco y <i> copia</i> de <i>D. melanogaster</i> . _____	52
<b>Figura 16.</b> Árbol filogenético basado en el alineamiento de secuencias aminoacídicas de zonas de la RT de tipo <i> copia</i> obtenidas del genoma de <i>C. sinensis</i> . _____	54
<b>Figura 17.</b> Experimento representativo de <i>Slot-blot</i> hibridado con la población completa de fragmentos amplificados en la PCR específica para la RT de RTN de tipo <i> copia</i> . _____	56
<b>Figura 18.</b> Obtención de fragmentos de RTN <i> gypsy</i> del genoma de <i>C. sinensis</i> _____	57
<b>Figura 19.</b> Alineamiento múltiple por ClustalW de las secuencias aminoacídicas inferidas de 3 clones GySi# y de zonas similares de RTN tipo <i> gypsy</i> ya descritos como son <i> Del1</i> de <i>Lilium</i> y <i> gypsy</i> de <i>D. melanogaster</i> _____	59
<b>Figura 20.</b> Árbol filogenético basado en el alineamiento de secuencias nucleotídicas de fragmentos de los dominios de a RT (A) y de la integrasa (B) de RTN de tipo <i> gypsy</i> clonados del genoma de <i>Citrus sinensis</i> _____	61
<b>Figura 21.</b> Resultado de la PCR inversa sobre DNA de <i>C. sinensis</i> flanqueando el fragmento CoSi14 _____	63
<b>Figura 22.</b> Secuencia H4. (A) Esquema de la estrategia de la amplificación y clonaje (B) Secuencia aminoacídica inferida y secuencia nucleotídica no codificante del clon H4 _____	64
<b>Figura 23.</b> Resultado de la electroforesis en gel 0,8% de agarosa de la amplificación por PCR directa sobre genoma de <i>C. sinensis</i> para obtener el extremo 5' del elemento _____	65
<b>Figura 24.</b> Secuencia MP20. (A) Esquema de la estrategia de clonaje. (B) Secuencia nucleotídica no codificante y secuencia aminoacídica inferida del clon MP20 _____	66

<b>Figura 25.</b> Alineamiento de las secuencias nucleotídicas de las dos posibles LTR del elemento clonado del genoma de <i>C. sinensis</i> _____	67
<b>Figura 26.</b> Secuencia nucleotídica no codificante del extremo 3' de H4. Diseño del cebador oligo1 ___	69
<b>Figura 27.</b> Secuencia nucleotídica no codificante del extremo 3' de H4. Diseño del cebador oligo12 _	69
<b>Figura 28.</b> Diseño del experimento para determinar si una secuencia pertenece o no a LTR _____	70
<b>Figura 29.</b> Secuencia nucleotídica no codificante del extremo 3' de H4. Diseño del cebador oligo21 _	71
<b>Figura 30.</b> Resultado de la electroforesis en gel 0,8% de agarosa de la amplificación por PCR directa sobre genoma de <i>C. sinensis</i> para obtener la secuencia del inicio de la LTR5' del elemento _____	71
<b>Figura 31.</b> Esquema resumen de los fragmentos amplificados sobre DNA genómico de <i>C. sinensis</i> ___	72
<b>Figura 32.</b> Alineamiento múltiple de secuencias nucleotídicas correspondientes a fragmentos de LTR aislados de <i>C. sinensis</i> _____	74
<b>Figura 33.</b> Secuencia de CIRE1-1 _____	78
<b>Figura 34.</b> Integridad de los elementos CIRE1 en <i>C. sinensis</i> _____	80
<b>Figura 35.</b> Motivos estructurales y funcionales de CIRE1-1, comparados con RTN-LTR tipo <i>copia</i> ___	82
<b>Figura 36.</b> Alineamiento múltiple de las secuencias aminoacídicas del dominio RT _____	83
<b>Figura 37.</b> Alineamiento múltiple de las secuencias aminoacídicas del motivo más conservado del dominio RH (TDAS) _____	84
<b>Figura 38.</b> Árbol filogenético basado en el alineamiento de las secuencias aminoacídicas de todo el dominio interno de CIRE1 y otros RTN _____	85
<b>Figura 39.</b> Representación gráfica del número promedio de copias del RTN CIRE1 calculado para diferentes regiones del elemento _____	87
<b>Figura 40.</b> Resultado de la amplificación selectiva de CIRE1 en el genoma de distintas especies del género <i>Citrus</i> y especies de géneros próximos _____	88
<b>Figura 41.</b> Resultado de la amplificación selectiva de CIRE1 en el genoma de distintas especies vegetales _____	89
<b>Figura 42.</b> Análisis genómico por <i>Southern-blot</i> de CIRE1 en especies familia Rutaceae _____	90
<b>Figura 43.</b> Análisis genómico por <i>Southern-blot</i> de CIRE1 en otras especies vegetales _____	91
<b>Figura 44.</b> Secuencia consenso del promotor LTR5' de CIRE1 indicando los posibles motivos reguladores de la transcripción _____	93
<b>Figura 45.</b> Localización de la actividad GUS en hojas de <i>C. sinensis</i> transformadas con pBSKLTR:: <i>gus</i> :: <i>nos</i> _____	95
<b>Figura 46.</b> Análisis de la expresión transitoria del promotor LTR5' de CIRE1 en células de epidermis de cebolla _____	96
<b>Figura 47.</b> Tinción histoquímica de suspensiones celulares de tabaco (W38) bombardeadas con las construcciones pBSKLTR:: <i>gus</i> :: <i>nos</i> , pBSK35S:: <i>gus</i> :: <i>nos</i> y pBSKgus:: <i>nos</i> _____	98
<b>Figura 48.</b> Actividad del promotor de la LTR5' de CIRE1 en comparación al promotor 35S en suspensiones celulares de tabaco, determinada por ensayo fluorimétrico de GUS _____	97
<b>Figura 49.</b> Obtención de transcritos de RTN-LTR <i>copia</i> en <i>C. sinensis</i> _____	100
<b>Figura 50.</b> Árbol fenético basado en el alineamiento de secuencias nucleotídicas de transcritos de RTN <i>copia</i> obtenidos por RT-PCR y fragmentos genómicos de RTN <i>copia</i> amplificados por PCR _____	101
<b>Figura 51.</b> Alineamiento múltiple por ClustalW de las secuencias amplificadas mediante RT-PCR de transcritos de RTN tipo <i>copia</i> presentes en <i>C. sinensis</i> _____	103
<b>Figura 52.</b> Análisis por RT-PCR de la acumulación de transcritos de RTN-LTR tipo <i>copia</i> en hoja de <i>C. sinensis</i> _____	105

<b>Figura 53.</b> Obtención de transcritos de CIRE1 en tejidos de <i>Citrus sinensis</i> sometidos a estrés por herida	108
<b>Figura 54.</b> Obtención de transcritos de CIRE1 en hojas de <i>C. sinensis</i> sometido a diferentes tratamientos	110
<b>Figura 55.</b> Árbol fenético basado en el alineamiento de transcritos de CIRE1 obtenidos por RT-PCR en muestras sometidas a diversos tratamientos y fragmentos genómicos de RTN-LTR tipo <i> copia</i> amplificados por PCR en <i>C. sinensis</i>	112
<b>Figura 56.</b> Frutos de especies cítricas comerciales y géneros próximos utilizadas en el estudio	113
<b>Figura 57.</b> Resultado de la amplificación de RTN-LTR tipo <i> copia</i> en el genoma de distintas especies del género <i>Citrus</i> y especies de géneros próximos	114
<b>Figura 58.</b> Alineamiento múltiple por ClustalW de las 66 secuencias aminoacídicas inferidas de los fragmentos de RT de RTN tipo <i> copia</i> , obtenidos de DNA genómico de distintas especies del género <i>Citrus</i> y especies de géneros próximos	115
<b>Figura 59.</b> Árboles filogenéticos tipo <i> Neighbour joining</i> de secuencias nucleotídicas de fragmentos de la RT de RTN tipo <i> copia</i> , aislados de DNA genómico de (A) <i>C.grandis</i> , (B) <i>C.limettioides</i> , (C) <i>Poncirus trifoliata</i> , (D) <i>C.medica</i> , (E) <i>C.clementina</i> y (F) <i>C.sinensis</i>	117
<b>Figura 60.</b> Análisis filogenético basado en el alineamiento de 66 secuencias nucleotídicas de fragmentos de la RT de RTN tipo <i> copia</i> , obtenidos de DNA genómico de 5 especies del género <i>Citrus</i> y 1 especie del género próximo <i>Poncirus</i>	119
<b>Figura 61.</b> Análisis filogenético basado en el alineamiento de 92 secuencias nucleotídicas de fragmentos de la RT de RTN tipo <i> copia</i>	122
<b>Figura 62.</b> Número de copias total de RTN tipo <i> copia</i> en el genoma de distintas especies del género <i>Citrus</i> y especies de géneros próximos	124
<b>Figura 63.</b> Esquema de la metodología de la técnica SSAP	126
<b>Figura 64.</b> Detección de bandas SSAP en geles de poliacrilamida 6% mediante tinción con plata	128
<b>Figura 65.</b> SSAP del RTN CIRE1 con el cebador selectivo Pcg, donde se favorece la separación correcta de los fragmentos generados de tamaño inferior a 1 kb y superior a 150 pb	130
<b>Figura 66.</b> SSAP del RTN CIRE1 con el cebador selectivo Pcg, donde se favorece la separación correcta de los fragmentos generados de menor tamaño: entre 330 pb y 50 pb	131
<b>Figura 67.</b> SSAP del RTN CIRE1 con el cebador selectivo Ttg, donde se favorece la separación correcta de los fragmentos generados de tamaño inferior a 1 kb y superior a 120 pb	132
<b>Figura 68.</b> SSAP del RTN CIRE1 con el cebador selectivo Ttg, donde se favorece la separación correcta de los fragmentos generados de menor tamaño: entre 330 pb y 50 pb	133
<b>Figura 69.</b> Análisis filogenético derivado de marcadores SSAP utilizando en la PCR selectiva un cebador específico de la LTR de CIRE1 en combinación con el cebador selectivo Pcg (A) o el cebador selectivo Ttg (B)	136
<b>Figura 70.</b> SSAP del RTN CIRE1 con el cebador selectivo Ttg en DNA genómico de <i>C. sinensis</i> sometido a diferentes tratamientos para ver actividad transposicional	141

## **MATERIAL Y MÉTODOS**

<b>Figura 71.</b> Multiplicación <i> in vitro</i> de plantas de <i>C. sinensis</i> ("Valencia late") a partir de semilla	203
--	-----

# INDICE DE TABLAS

---

## INTRODUCCIÓN

<b>Tabla I.</b> Cítricos designados como especies por Swingle (1943) _____	4
<b>Tabla II.</b> Contenido y distribución de elementos transponibles en diferentes genomas eucariotas secuenciados por completo (GENOME INTERNATIONAL SEQUENCING CONSORTIUM, 2001) _____	12
<b>Tabla III.</b> Mecanismos para minimizar los efectos adversos de la movilización de elementos transponibles en el genoma huésped, modificado de Kidwell y Lisch (1997) _____	17
<b>Tabla IV.</b> Principales características estructurales de algunos de los RTN-LTR de plantas aislados y caracterizados _____	28
<b>Tabla V.</b> Algunos de los RTN-LTR caracterizados que muestran actividad transcripcional _____	31
<b>Tabla VI.</b> Promotores funcionales de RTN-LTR de plantas _____	33
<b>Tabla VII.</b> Tabla comparativa de las principales técnicas de obtención de marcadores _____	40

## RESULTADOS

<b>Tabla VIII.</b> Número de copias total de RTN tipo <i>copia</i> en el genoma de <i>Citrus sinensis</i> _____	56
<b>Tabla IX.</b> Número de copias de RTN tipo <i>gypsy</i> en el genoma de <i>Citrus sinensis</i> _____	62
<b>Tabla X.</b> Comparación de las secuencias nucleotídicas correspondientes a fragmentos de LTR aislados de <i>C. sinensis</i> _____	75
<b>Tabla XI.</b> Comparación de las secuencias nucleotídicas correspondientes a regiones internas de RTN aisladas de <i>C. sinensis</i> _____	76
<b>Tabla XII.</b> Descripción de las secuencias aisladas del genoma de <i>C. sinensis</i> _____	77
<b>Tabla XIII.</b> Número de copias del RTN-LTR CIRE1 en el genoma de <i>Citrus sinensis</i> _____	86
<b>Tabla XIV.</b> Taxonomía de las especies utilizadas en el estudio _____	89
<b>Tabla XV.</b> Parámetros de variabilidad genética de la población de secuencias genómicas y transcritos de la región de la RT de RTN tipo <i>copia</i> calculadas con el programa DNAsp3.14 _____	104
<b>Tabla XVI.</b> Fracción de clones cuyas secuencias conceptuales de aminoácidos están interrumpidas por codones de parada o bien en las que son necesarios cambios de pauta de lectura _____	116
<b>Tabla XVII.</b> Diversidad nucleotídica ( $\pi$ ) entre las secuencias de RTN-LTR tipo <i>copia</i> obtenidas de los genomas de especies del género <i>Citrus</i> y especies de géneros próximos _____	118
<b>Tabla XVIII.</b> Variabilidad de las secuencias dentro de las distintas familias de RTN-LTR <i>copia</i> de los genomas de especies del género <i>Citrus</i> y especies de géneros próximos _____	120
<b>Tabla XIX.</b> Niveles de polimorfismo SSAP para el RTN-LTR CIRE1 en diversas especies de cítricos y especies de géneros próximos, utilizando dos cebadores selectivos diferentes _____	134

# ABREVIATURAS Y SÍMBOLOS

---

A	Adenina o adenosina
aa	Aminoácido
Adh	Alcohol deshidrogenasa
AFLP	<i>Amplified fragment length polymorphism</i>
APS	Amonium persulfate
ATP	Adenosín trifosfato
BSA	Albúmina sérica bovina
cDNA	DNA complementario
dsDNA	DNA de doble cadena
Ci	Curie, medida radioactiva
C-terminal	Extremo carboxilo
DMSO	Dimetilsulfóxido
DNA	Ácido desoxirribonucleico
DNAse	Desoxirribonucleasa
dNTP	Desoxirribonucleótidos
DO	Densidad Óptica
EDTA	Ácido etilendiamin-tetra-acético
EST	<i>Expressed sequences tag</i>
TE	Elementos transponibles
G	Guanina o guanosina
GTP	Guanosín trifosfato
GUS	$\beta$ -glucuronidasa
h	horas
INT	Integrasa
IRAP	<i>Inverse retrotransposon amplified polymorphism</i>
JA	Ácido jasmónico
Kb	Kilobases
LARD	<i>Large retrotransposon derivatives</i>
LB	Medi Luria-Bertani
LINE	<i>Long interspersed nuclear elements</i>
LTR	Repeticiones terminales largas
Mb	Megabases
MES	acid 2-morfolinoetansulfónico
Min	Minutos
MITE	<i>Miniature inverted-repeated transposable elements</i>
MOPS	3-N-morfolinopropanosulfonat
mRNA	RNA mensajero
NAA	Ácido naftalenacético
nd	nucleótidos
N-terminal	Extremo amonio
ON	Toda la noche
ORF	Pauta de lectura abierta
pb	Pares de bases
PBS	Sitio de unión del iniciador de la primera cadena de cDNA
PCR	Reacción en cadena de la polimerasa
PEG	Polietilenglicol

PNK	Polinucleótido quinasa
PPT	Zona rica en purinas iniciadora de la segunda cadena de cDNA
PR	proteasa
Pu	Purina
Py	Pirimidina
QTL	<i>Quantitative trait locus</i>
RAPD	<i>Random amplified polymorphic DNA</i>
RBIP	<i>Retrotransposon-based insertion polymorphism</i>
REMAP	<i>Retrotransposon-microsatellite amplified polymorphism</i>
RFLP	<i>Restriction fragment length polymorphism</i>
RI	Repeticiones invertidas
RIVP	<i>Retrotransposon internal variation polymorphism</i>
RNA	Ácido ribonucleico
RNAsa	ribonucleasa
RT	Transcriptasa inversa
RTN	Retrotransposón
s	Segundos
SA	Àcid salicílic
SCAR	<i>Sequence characterized amplified regions</i>
SDS	Dodecil sulfat de sodi
SINE	<i>Short interspersed nuclear elements</i>
SNP	<i>Single nucleotide polymorphism</i>
SSAP	<i>Sequence Specific Amplified Polymorphism</i>
SSR	Repetición de secuencia sencilla
T	Timidina
TA	Temperatura ambiente
TBE	Tris-borato-EDTA
TE	Tris-EDTA
TEMED	N,N,N',N'-tetrametiletildiamina
TIR	Repetición terminal invertida
TRIM	<i>Terminal-repeat retrotransposons in miniature</i>
tRNA	RNA de transferencia
TTP	Timidin trifosfato
VLP	Partículas similares a las virales

En las secuencias aminoacídicas:

<b>F</b> Phe, fenilalanina	<b>S</b> Ser, serina	<b>Y</b> Tyr, tirosina	<b>K</b> Lys, lisina	<b>W</b> Trp, triptófano
<b>L</b> Leu, leucina	<b>P</b> Pro, prolina	<b>H</b> His, histidina	<b>D</b> Asp, aspártico	<b>R</b> Arg, arginina
<b>I</b> Ile, isoleucina	<b>T</b> Thr, treonina	<b>Q</b> Gln, glutamina	<b>E</b> Glu, glutámico	<b>G</b> Gly, glicina
<b>M</b> Met, metionina	<b>A</b> Ala, alanina	<b>N</b> Asn, asparragina	<b>C</b> Cys, cisteína	<b>V</b> Val, valina

## **1. LOS CÍTRICOS**

### **1.1 ORIGEN DEL GÉNERO CITRUS**

Hay diversas hipótesis sobre el origen del género *Citrus*. La más aceptada lo sitúa hace miles de años en el sureste de Asia, con una posterior expansión por los otros continentes y siendo actualmente cultivado en todo el mundo. No se conocen exactamente las rutas de dispersión de los cítricos por los distintos continentes, aunque se sabe que están presentes en las escrituras antiguas de muchas culturas.

Los primeros cítricos fueron el resultado de mutaciones espontáneas de las plantas silvestres. El mejoramiento genético vegetal generó nuevas variedades e híbridos con tolerancia a condiciones de estrés biótico y abiótico, de alto rendimiento y con las características de calidad requeridas por el mercado. Este proceso de domesticación llevado a cabo por el hombre ha provocado que las principales especies modernas de cítricos no se encuentren en estado silvestre, pues son resultado de hibridaciones tanto naturales como seleccionadas por el hombre a través del tiempo.

### **1.2 IMPORTANCIA ECONÓMICA Y DISTRIBUCIÓN GEOGRÁFICA DE LOS CÍTRICOS**

Los cítricos ocupan el primer lugar en volumen de producción en fruticultura de acuerdo con las estadísticas sobre agricultura mundial (F.A.O., 2003).

Los principales países productores son Brasil, USA, China y España. España es el principal exportador de naranja para consumo directo, con una producción de 3.000.000 de toneladas y una superficie de cultivo de 285.000 hectáreas. Por regiones, el 68% de la superficie cultivada está en la Comunidad Valenciana, el 12% en Murcia y el 16% en Andalucía.

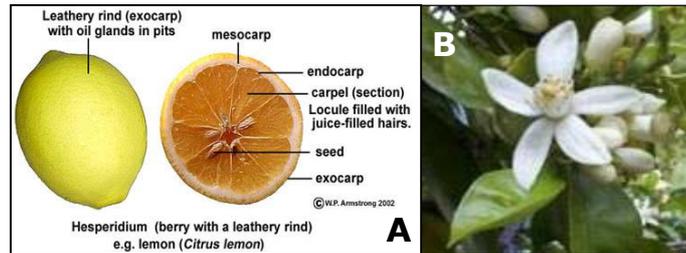
### **1.3 CARACTERÍSTICAS MORFOLÓGICAS, REPRODUCTIVAS Y GENÉTICAS DEL GÉNERO**

#### **1.3.1 Morfología de los cítricos**

Los cítricos son árboles, a veces espinosos, que se caracterizan por su fruto carnoso denominado hesperidio (**Figura 1A**). Las flores (Azahar), muy blancas, son apreciadas por su buen olor (**Figura 1B**).

## Introducción

**Figura 1. (A) Hesperidio o fruto de los cítricos.** El exocarpo es rico en glándulas que acumulan esencias aromáticas, y el mesocarpo es carnoso pero seco. Las dos envueltas constituyen la piel del fruto, mientras que el endocarpo forma unos velos que delimitan los diferentes gajos, al mismo tiempo que originan un número muy elevado de vesículas donde se acumula el zumo. **(B) Flor de azahar**



### 1.3.2 Propagación y reproducción de los cítricos

- **Propagación vegetativa**

La mayoría de las especies comerciales actuales se obtienen tras una cuidada selección de mutaciones espontáneas en las yemas que afectan a características de interés agronómico (sabor, tamaño, presencia semillas, etc.) y se propagan vegetativamente a partir de microinjertos sobre patrones definidos. De esta manera se obtienen árboles de forma uniforme y rápida.

- **Propagación por semilla**

Aunque, muchas especies son directamente infértiles (sin semillas) dado que es una característica agronómica deseable para su mejor comercialización, también se da la propagación por semilla.

Los cítricos tienen una característica muy peculiar, ya que en muchas especies se da la **embriogénesis somática** o **apomixis**, proceso por el cual células haploides o somáticas se desarrollan formando estructuras semejantes al embrión zigótico, y posteriormente originan una planta sin que se de fusión de gametos. Esta formación asexual de embriones ocurre comúnmente en la naturaleza en las variedades poliembriónicas. En estas especies, los embriones somáticos o apomícticos se desarrollan mejor que los zigóticos, y en realidad cuando se siembra una semilla se suele obtener una planta genéticamente idéntica a la planta madre permitiendo una propagación clonal. Este tipo de propagación clonal es interesante y explotada satisfactoriamente en agricultura ya que los patrones obtenidos por semilla son uniformes, pero en cambio dificulta mucho la obtención de poblaciones segregantes y por consiguiente los estudios genéticos, de ahí que tengan que elegir muy bien los parentales a utilizar en los cruzamientos.

Únicamente el pummelo (*C. grandis*), el cidro (*C. medica*) y algunos mandarinos, como el clementino (*C. clementina*) son monoembriónicos y se comportan como es habitual en otras especies.

### 1.3.3 Genoma de los cítricos

Prácticamente todas las especies de *Citrus* son diploides, con  $2n = 18$  cromosomas, y con un genoma relativamente pequeño, 382 Mpb o 0.40 pg (Arumuganathan et al., 1991).

El estudio citogenético del género está limitado por lo relativamente pequeños y similares que son los cromosomas. Sin embargo, cariotipos basados en el bandeo de Giemsa-C y tinción con fluorocromos como CMA y DAPI muestran muchos pares cromosómicos heteromórficos, quizás como consecuencia de la ampliamente extendida hibridación interespecífica durante la domesticación (Roose et al., 1998; Guerra et al., 1993). También se muestra que muchos cromosomas contienen largos bloques de heterocromatina terminal (Miranda et al., 1997).

En contraste con la elevada variabilidad en caracteres agronómicos entre los diferentes cultivares de una misma especie, la variabilidad genética detectada por RFLP, RAPD o ISSR (Herrero et al., 1996; Federici et al., 1998) es baja. Una posible explicación podría ser que la obtención de nuevos cultivares a partir de mutaciones espontáneas y posterior propagación vegetativa, no está afectando los loci detectados por los marcadores (Bretó et al., 2001).

El desarrollo de técnicas moleculares ha permitido realizar mapas de ligamiento del genoma de cítricos (Roose et al., 2000; Sankar et al., 2001; Ruiz et al., 2003) y disponer de marcadores asociados a caracteres de interés (Deng et al., 2001; Deng et al., 2003; Gmitter et al., 1996).

#### Proyecto de genómica funcional de *Citrus* (CFGP):

Además de las especies modelo como *Arabidopsis thaliana* o arroz, la genómica funcional (estudio de la función de los genes) se está aplicando también a plantas con un directo interés económico como son el maíz, o solanáceas como el tomate y la patata (*TIGR's Genome Projects*, <http://www.tigr.org/tdb/>). Como representante del género *Citrus* se ha escogido el "clementino de nules" (*C. clementina*) con el que se están realizando estudios de genómica funcional a gran escala dentro del **proyecto CFGP** ([http://citrusgenomics.ibmcp-ivia.upv.es/about\\_cfgp/](http://citrusgenomics.ibmcp-ivia.upv.es/about_cfgp/)) que generan colecciones de ESTs y librerías genómicas altamente informativas, que pueden servir como punto de partida a la eventual secuenciación del genoma entero. En la dirección del **Internacional Citrus Genome Consortium** (<http://int-citrusgenomics.org/>) se concentra además toda la información que se va generando acerca de la genómica de *Citrus*.

Con todo, la composición del genoma de los cítricos está en general todavía poco estudiada.

## Introducción

### 1.4 TAXONOMÍA DEL GÉNERO CITRUS

Los cítricos son dicotiledóneas que pertenecen a la familia *Rutaceae*, subfamilia *Aurantoideae*, la cual comprende seis géneros: *Citrus*, *Clymenia*, *Eremocitrus*, *Fortunella*, *Microcitrus* y *Poncirus*. De estos, solo tres tienen importancia comercial: *Poncirus* (naranja trifoliado), da frutos no comestibles y es empleado como patrón, *Fortunella* (kumquat) y *Citrus*, que dan frutos comestibles. De estos el género *Citrus* es el más importante económicamente.

La compleja biología reproductiva de *Citrus* dificulta mucho los análisis taxonómicos del género. Por una parte, el género contiene una variabilidad enorme de variedades, con hibridaciones cruzadas interespecie y entre géneros próximos, tanto naturales como provocadas por el hombre, que provocan descendencia fértil y un enorme rango de fenotipos diferentes. Además, la apomixis o reproducción asexual presente en el género previene el intercambio de genes y el aislamiento reproductivo, necesario según la definición clásica de especie (Según el concepto biológico de Mayr (1991) las especies son grupos de poblaciones naturales con cruzamientos entre sí que están aisladas reproductivamente de otros grupos). A esto, se le suma una larga historia de cultivo y gran dispersión.

En 1943, Swingle determinó una clasificación para el género *Citrus* donde se reconocían diez especies dentro del subgénero *Citrus* (**Tabla I**) mientras que en 1954 Tanaka definió 147 especies diferentes dentro del género. La mayor diferencia entre ambos sistemas de clasificación está en el tratamiento que recibe el grupo de las mandarinas; Swingle las sitúa todas dentro de *C. reticulata* exceptuando las especies salvajes de Japón (*C. tachibana*) y las especies de la India (*C. indica*), en cambio para Tanaka hay 36 especies diferentes de mandarinas. Estas grandes diferencias en el número de especies reconocidas reflejan la dificultad para realizar la clasificación taxonómica de un género, que hasta mediados de los años 70 estaba basada únicamente en datos morfológicos y geográficos. La aplicación de técnicas más modernas para el estudio del género se desarrolla en el **Apartado 5** de la Introducción.

Taxonomía de Swingle	Nombre común
<i>C. aurantifolia</i> Christm.	Lima
<i>C. aurantium</i> L.	Naranja amarga
<i>C. indica</i> Tan.	especie india , no cultivada
<i>C. maxima</i> Cerril; o <i>C. grandis</i> Osberck	Pummelo
<i>C. limon</i> (L.) Burm. f.	Limón
<i>C. medica</i> L.	Cidro
<i>C. paradisi</i> Macf.	Pomelo
<i>C. reticulata</i> Blanco	Mandarina, tangerina, clementina, satsuma
<i>C. sinensis</i> Osbeck	Naranja dulce
<i>C. tachibana</i> (Mak.) Tan.	Especie japonesa, no cultivada

**Tabla I. Cítricos designados como especies por Swingle (1943).** Tabla modificada de Moore, 2001

## 2. COMPONENTES DE LOS GENOMAS

Con los proyectos de secuenciación de genomas y EST a gran escala, se está viendo que la cantidad de genes en los diferentes organismos no es proporcional al tamaño del genoma (lo que se conoce como la paradoja del valor C). En gran parte esto es debido al DNA repetitivo, el cual puede presentar grandes diferencias tanto a nivel de secuencia como en abundancia aun entre genomas de especies muy próximas, por lo que es importante para estudios evolutivos, genéticos, taxonómicos y aplicados.

El DNA repetitivo constituye gran parte del genoma en células eucariotas, alrededor del 35% de media, y es especialmente abundante en el genoma de las plantas. En *Arabidopsis thaliana* un 37% del genoma son secuencias repetidas (EUROPEAN UNION ARABIDOPSIS GENOME SEQUENCING CONSORTIUM, 1999), en maíz el 60-80% del genoma es DNA repetitivo (Heslop-Harrison et al., 2000) y en trigo las secuencias repetidas llegan incluso a ocupar el 83% del genoma (Wicker et al., 2001). Existe una buena base de datos de las secuencias altamente repetidas del genoma de plantas, especialmente del género *Triticeae*, disponible en la dirección <http://w3lamc.umbr.cas.cz/PlantSat/> (Macas et al., 2002).

El DNA repetitivo puede clasificarse a grandes rasgos como moderadamente repetitivo (normalmente repeticiones dispersas) y altamente repetitivo (repeticiones en tándem) (**Figura 2**).

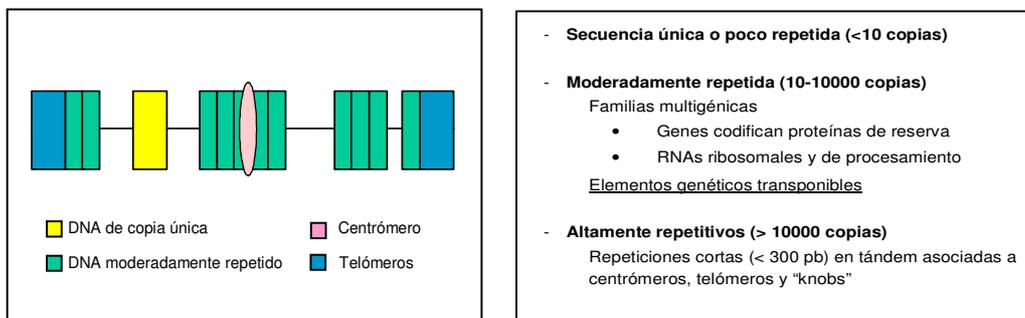


Figura 2. Esquema de la estructura del DNA y la clasificación del DNA según su número de copias.

Las secuencias altamente repetidas en tándem tienen normalmente localizaciones cromosómicas características: subteloméricas o centroméricas, e incluyen los DNA satélites (longitudes monoméricas variables siendo frecuentes 140-180 o 300-360 bp), los microsátélites (repeticiones de motivos cortos de 1-5 pb) y los minisátélites (unidades de repetición algo más complejas y largas, hasta 40 pb). Las secuencias teloméricas consisten en repeticiones de una secuencia de 7 pb conservada (CCCTAAA)

## Introducción

---

que es añadida al final de los cromosomas por acción de la telomerasa, manteniendo así la integridad de los cromosomas.

Pero, la mayor parte del DNA repetitivo lo constituyen los elementos transponibles.

### 3. LOS ELEMENTOS TRANSPONIBLES

Los elementos transponibles (TE, *Transposable Elements*) son fragmentos de DNA que tienen la capacidad de insertarse en nuevas localizaciones cromosómicas, pudiendo realizar frecuentemente copias de ellos mismos en el proceso. Con los avances en la secuenciación de DNA a gran escala se está viendo que no son un componente raro del genoma, sino que constituyen una fracción muy importante del material genético de muchos eucariotas. Ocupan el 45% del genoma humano (GENOME INTERNATIONAL SEQUENCING CONSORTIUM, 2001) y hasta un 50-80% de genomas de gramíneas (Meyers et al., 2001).

En 1983 Barbara McClintock recibió el premio Nóbel por su descubrimiento de fragmentos de DNA capaces de moverse por los cromosomas del maíz, los transposones. Su descubrimiento databa del 1950 y estaba basado únicamente en estudios citogenéticos y observaciones de los fenotipos de las plantas y las panochas de maíz en diferentes cruces (**Figura 3**). Posteriormente los estudios moleculares de John McDonald se centraron más en la importancia que han tenido los TE a lo largo de la evolución.

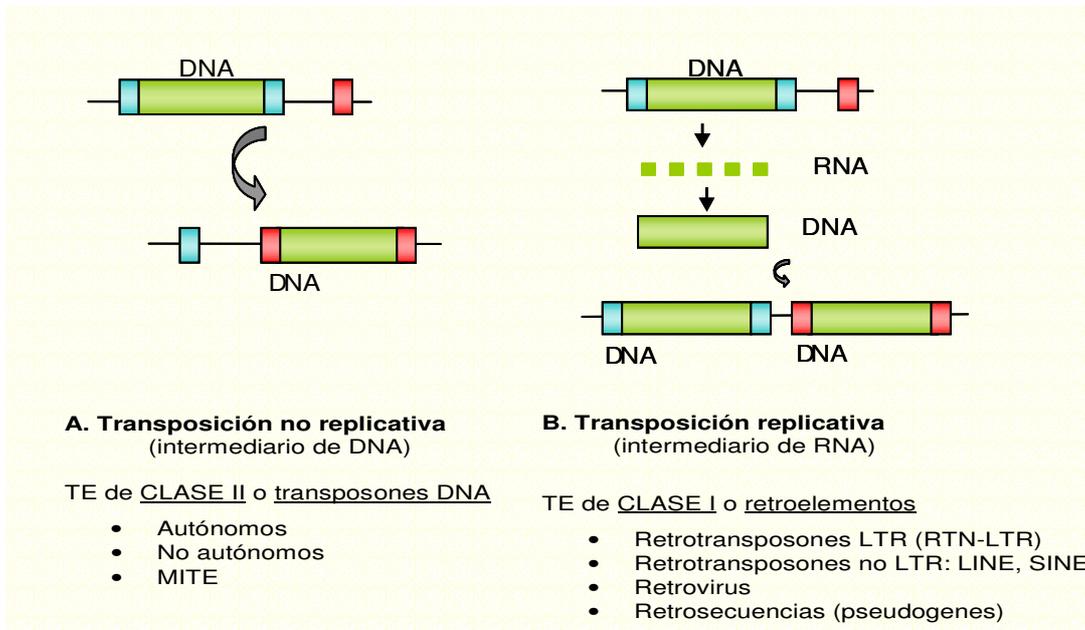


**Figura 3. Descubrimiento de los TE en plantas.** B. McClintock determinó que los efectos que observaba en la pigmentación de las panochas de maíz eran causados por transposones (*Ac/Ds*) capaces de insertarse en distintas regiones del genoma y que eran causantes de mutaciones, ya que afectaban a la función del gen con el que se asociaban, en este caso genes de la vía biosintética de los antocianos.

Años después, los TE han sido encontrados en el genoma de todos los organismos estudiados, como por ejemplo las secuencias *Alu* en humanos, los elementos *Ty1-5* de *Saccharomyces cerevisiae* o los elementos *Copia* y *Gypsy* de *Drosophila*, entre muchos más.

### 3.1 CARACTERÍSTICAS ESTRUCTURALES Y CLASIFICACIÓN DE LOS ELEMENTOS TRANSPONIBLES

Los TE se dividen en dos clases según si su intermediario en la transposición es RNA (**Clase I o retroelementos**) o DNA (**Clase II o transposones de DNA**). Mientras que en los elementos de clase II es el propio DNA el que se moviliza, en los elementos de clase I, el DNA se transcribe a RNA, éste se retrotranscribe y es la copia de DNA la que se integra en la nueva localización. Este modo replicativo de transposición (integración de nuevas copias hijas vía un intermediario de RNA) de los retroelementos puede provocar un incremento rápido del número de copias de los elementos en el genoma huésped. Sin embargo, los transposones de DNA se mueven directamente vía un intermediario de DNA, mediante un mecanismo de "corte y pega" de manera que no aumentan significativamente su número de copias en el genoma huésped (**Figura 4**).



**Figura 4. Clasificación de los elementos transponibles (TE) según su modo de transposición:** (A) Transposición vía directamente un intermediario de DNA (modo no replicativo o conservativo) en TE de clase II o transposones de DNA. (B) Transposición vía la transcripción inversa de un intermediario de RNA (modo replicativo) en TE de clase I o retroelementos. En ambos tipos se duplica la diana de inserción como resultado de su movilización.

El sistema de elementos autónomos y no-autónomos fue descrito primeramente en elementos de Clase II (Fedoroff et al., 1983), donde la movilización del elemento no-autónomo depende de la transposasa (enzima que lleva a cabo la transposición) de su pareja autónoma. Este sistema de transactivación es también aplicable a los elementos

## Introducción

---

de Clase I donde más recientemente se están describiendo también retroelementos de carácter no autónomo (Kalendar et al., 2004; Witte et al., 2001).

### 3.1.1 Elementos de Clase I ( Retroelementos)

Dentro de esta clasificación se incluyen los retrovirus, las retrosecuencias y los retrotransposones (**Figura 5**), aunque nos vamos a centrar exclusivamente en estos últimos.

CLASE. Retroelementos  
ORDEN. Retrales  
SUBORDEN. Retroposineae → **Retrotransposones no-LTR**  
SUBORDEN. Retronineae (intrones mitocondriales grupo II)  
ORDEN. Retrovirales  
SUBORDEN. Orthoretrovirineae [RNA, ciclo con fase de integración]  
SUBORDEN. Pararetrovirineae [DNA, sin función integrasa]  
FAMILIA. Hepadnaviridae (infecta animales; p.e. HBV)  
FAMILIA. Caulimoviridae (infecta plantas, p.e. CaMV)  
SUBORDEN. Retrotransposineae → **LTR- Retrotransposones**  
FAMILIA. Pseudoviridae (Ty1/copia)  
FAMILIA. Metaviridae (Ty3/gypsy)

**Figura 5. Clasificación de los retroelementos y virus relacionados, según Hansen C y Heslop-Harrison JS (2004).** Siendo, HBV: Hepatitis B Virus; CaMV: Cauliflower Mosaic Virus

Los retrotransposones (RTN) son la clase más abundante y extendida de TE en eucariotas. Pueden dividirse en dos grandes grupos según si presentan o no repeticiones directas terminales largas (**LTR**, *Long Terminal Repeat*) flanqueando la región interna codificante.

Los elementos autónomos tienen como mínimo dos genes: ***gag*** y ***pol*** que se sintetizan en forma de poliproteína, que posteriormente se procesa en diferentes péptidos funcionales mediante una proteasa codificada por el propio elemento. El gen ***gag*** codifica para proteínas estructurales de la cápside. El gen ***pol*** codifica para actividades enzimáticas necesarias para el ciclo de vida autónomo del RTN, como son: la proteasa (**PR**), integrasa (**INT**), transcriptasa inversa (**RT**) y RNasa H (**RH**). Los elementos no-autónomos han perdido parte o toda la región codificante.

- **Retrotransposones con LTR (RTN-LTR)**

Los **RTN-LTR** presentan repeticiones largas terminales (LTR) de tamaño variable (desde 100 pb hasta varias kb) flanqueando el dominio interno. Las LTR contienen normalmente secuencias reguladoras tanto promotoras como terminadoras de la transcripción del elemento.

Este grupo comprende los RTN de tipo **Ty3-gypsy** y los RTN de tipo **Ty1-copia**. Los RTN-LTR son elementos estructuralmente similares a los retrovirus y tienen también un ciclo de vida similar, aunque no comparten su capacidad infectiva. Los miembros de cada tipo de RTN-LTR presentan similitudes tan elevadas entre miembros de especies vegetales diferentes como con elementos de la misma especie, apoyando la divergencia tan antigua que hay entre los dos tipos de RTN-LTR (Xiong y Eickbush, 1990).

Estos dos grupos de RTN-LTR, *gypsy* y *copia*, se diferencian entre sí por la conservación a nivel de secuencia de sus dominios internos y por el orden relativo de sus dominios enzimáticos dentro del gen *pol*. Mientras que en los RTN tipo *copia* el dominio de la INT está situado a 5' de la RT, inmediatamente después de la región que codifica para la PR, en los de tipo *gypsy* el dominio de la INT está situado a 3' de la RT, justo al final del gen *pol*.

Recientemente se están describiendo nuevos tipos de retroelementos con LTR de carácter no autónomo como son los **LARD** (*Large Retrotransposon Derivate*) o los **TRIM** (*Terminal Repeat Retrotransposon In Miniature*). En los LARD, la región codificante está substituida por un largo dominio conservado no codificante. La secuencia de este dominio interno es suficientemente divergente como para no alinearse correctamente con las secuencias codificantes de RTN-LTR típicos, pero a su vez está bien conservada y presenta polimorfismos entre distintos cereales (Kalendar et al., 2004). Los TRIM son retroelementos análogos a los transposones de DNA no autónomos que presentan repeticiones directas terminales cortas (TDR, *Terminal Direct Repeat*), de entre 100-250 pb similares a LTR, pero carecen de dominio interno (Witte et al., 2001).

- **Retrotransposones sin LTR (RTN no-LTR)**

Los RTN no-LTR no presentan LTR y se transcriben a partir de un promotor interno. Dentro de este grupo encontramos los **LINE** (*Long Interspersed Nuclear Elements*), los **SINE** (*Short Interspersed Nuclear Elements*) y los retro-pseudogenes. La replicación de estos elementos se da en el mismo sitio de integración, cuando la endonucleasa corta el DNA huésped para generar extremos 3' hidroxil que sirven como cebadores para el inicio de la retrotranscripción.

## **Introducción**

---

Así como los LINE son elementos autónomos que contienen un dominio interno que codifica para los genes *gag* y *pol* (codifica para la endonucleasa y la transcriptasa inversa), los SINE no tienen ninguna capacidad codificante. Estos últimos derivan de transcritos de la polimerasa III (como los tRNA o los 7SL RNA) y utilizan las funciones de los LINE para transponerse, actuando como elementos no autónomos.

Un ejemplo de LINE en plantas es BLIN que está presente en 40-50 copias en el genoma de *Hordeum* (Vershinin et al., 2002).

### **3.1.2 Elementos de Clase II (Transposones)**

Los transposones de DNA presentan repeticiones terminales invertidas (TIR, *Terminal Inverted Repeated*) que son reconocidas por la transposasa, enzima que lleva a cabo la movilización de "corte y pega" de estos elementos. Los miembros no autónomos de una familia derivan normalmente de deleciones internas de elementos autónomos de la misma familia que provocan la pérdida de las secuencias codificantes. Estos elementos defectivos pueden ser transactivados por transposasas codificadas por elementos relacionados.

Ejemplos de familias de transposones de DNA son los elementos P de *Drosophila*. En plantas entre otros, *Ac/Ds* donde *Ac* (*Activator*) es el elemento autónomo capaz de transactivar a *Ds* (*Dissociation*) o *pSpm/dSpm* donde *pSpm* (*Supresor-mutator*) transactiva el elemento defectivo *dSpm* (*defective Spm*).

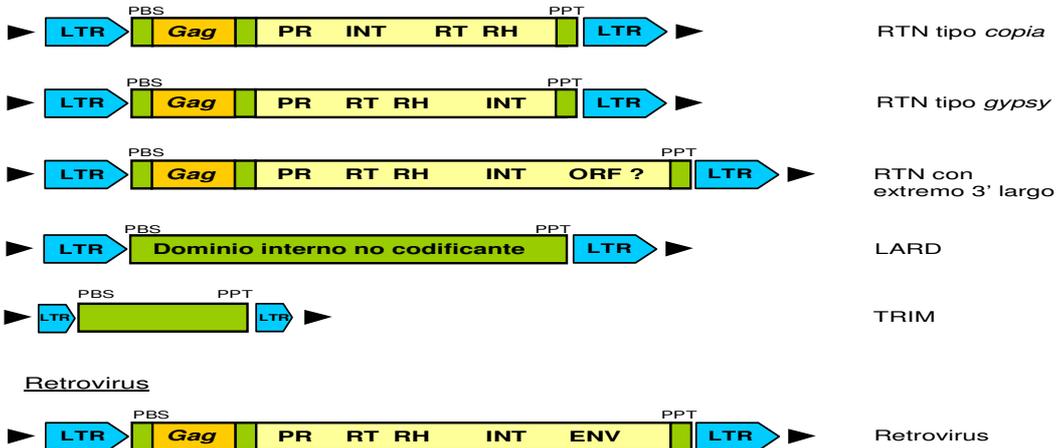
Los **MITE** (*Miniature Inverted-repeat Transposable Element*) son estructuralmente como TE de clase II defectivos, pues presentan TIR y no tienen capacidad codificante. Sin embargo, parecen tener un tipo de movilización replicativa propia de TE de clase I, ya que se encuentran en un número muy elevado de copias y cada subfamilia de MITE está muy conservada a nivel de tamaño y secuencia, lo que sugiere una amplificación a partir de un número limitado de progenitores. El mecanismo de amplificación de los MITE es hasta el momento desconocido, aunque se ha propuesto que actuarían como un tipo particular de elementos de clase II no-autónomos que necesitan de transposasas codificadas por elementos autónomos relacionados para poder movilizarse (Feschotte et al., 2002).

**Retrotransposones o Elementos Transponibles de Clase I**

Retrotransposones no-LTR



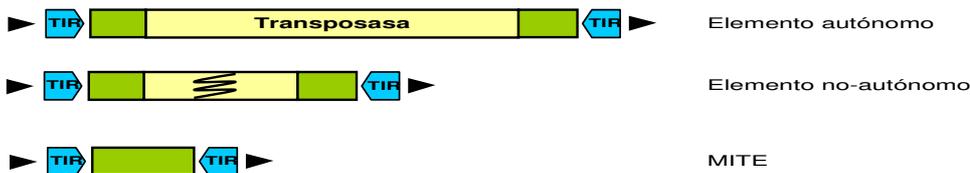
Retrotransposones LTR



Retrovirus



**Transposones o Elementos Transponibles de Clase II**



**Figura 6. Organización estructural general de los diferentes tipos de TE en plantas.** Los triángulos negros representan las duplicaciones de las dianas de inserción (DI) o TSD (*Target Site Duplication*), resultado directo de la actividad transposicional de los TE. Dentro de los **RTN no-LTR**, los **LINE** poseen algunos genes que codifican proteínas necesarias para su multiplicación: gen *gag* que codifica para proteínas de la cápside, y las proteínas endonucleasa (EN) y transcriptasa inversa (RT), mientras que los **SINE** no presentan región codificante. Ambos tipos de elementos acaban con una repetición simple, generalmente una cola poli(A). Los **RTN-LTR** presentan repeticiones largas terminales (LTR) flanqueando el dominio interno que codifica para las proteínas de la cápside (gen *gag*), proteasa (PR), integrasa (INT), transcriptasa inversa (RT) y RNAsaH (RH). Ambos tipos de RTN-LTR, **gypsy** y **copia**, se diferencian entre sí por el orden relativo de sus dominios enzimáticos dentro del dominio interno. También presentan PBS (*Primer Binding Site*) y PPT (*Poly Purine Track*), secuencias donde se une el cebador para la síntesis de la cadena cDNA (-) y cDNA (+), respectivamente. Los **retrovirus** presentan la misma organización estructural que los RTN-LTR tipo **gypsy** pero contienen en el extremo 3', adyacente a la INT, el gen *env* que codifica para proteínas de la envuelta. Los **LARD** presentan LTR, PBS y PPT pero la región interna está substituida por un largo dominio conservado no codificante. Los **TRIM** son retroelementos no autónomos con LTR cortas y sin dominio interno. Los **transposones de DNA** presentan repeticiones invertidas terminales (TIR). Los miembros **autónomos** codifican para la transposasa, mientras que los **no autónomos** presentan las secuencias codificantes truncadas. Los **MITE** presentan TIR y no tienen capacidad codificante. El esquema no está a escala, los RTN-LTR miden desde unos pocos kb hasta 15 kb [incluso 22kb el elemento tipo **gypsy** *Ogre* de *Pisum sativum* (Neumann et al., 2003)]. Los LINE son algo más pequeños, desde 1kb hasta 8 kb, mientras que los SINE normalmente tienen entre 100 y 300 pb de tamaño. Los MITE tienen menos de 600 pb.

## Introducción

### 3.2 IMPORTANCIA DE LOS ELEMENTOS TRANSPONIBLES EN LOS GENOMAS EUCARIOTAS

Tanto en animales como en plantas, los retroelementos son la clase mayoritaria de DNA repetitivo disperso en su genoma, aunque difieren en el tipo de retroelemento mayoritario y su ubicación relativa a los genes (**Tabla II**).

	<i>H. sapiens</i>		<i>D. melanogaster</i>		<i>C. elegans</i>		<i>A. thaliana</i>	
	% bases	Nºfamilias	% bases	Nºfamilias	% bases	Nºfamilias	% bases	Nºfamilias
LINE/SINE	33,4	6	0,7	20	0,4	10	0,5	10
LTR	8,1	100	1,5	50	0,0	4	4,8	70
DNA	2,8	60	0,7	20	5,3	80	5,1	80
total	44,4	170	3,1	90	6,5	94	10,5	160

**Tabla II. Contenido y distribución de elementos transponibles en diferentes genomas eucariotas secuenciados por completo** (GENOME INTERNATIONAL SEQUENCING CONSORTIUM, 2001). En la tabla, LINE/SINE hace referencia al grupo de retrotransposones no LTR (RTN no LTR); LTR, hace referencia al grupo de retrotransposones con LTR(RTN-LTR); DNA, hace referencia al grupo de transposones de DNA; total, hace referencia a la suma de las diferentes contribuciones.

Así por ejemplo, del total de elementos transponibles del genoma humano, el 10% (1 millón de copias) corresponde a una única familia: las secuencias Alu que es el único SINE activo en el genoma humano (GENOME INTERNATIONAL SEQUENCING CONSORTIUM, 2001). En cambio en *Arabidopsis thaliana*, los transposones de DNA son los TE que ocupan un mayor porcentaje del genoma con más de 80 familias descritas entre las que encontramos 80 miembros de MULE (*Mutator like*), 7 miembros de *Ac-like* o 15 miembros de MITE (Quang Hien Le et al., 2000). Además, los RTN-LTR son los que están mayoritariamente representados dentro de los retroelementos y los SINE están raramente presentes en su genoma (EUROPEAN UNION ARABIDOPSIS GENOME SEQUENCING CONSORTIUM, 1999). En *Drosophila melanogaster* se han identificado hasta 90 familias de elementos transponibles y además contiene en su genoma retrovirus activos (Alberola et al., 1996; Song et al., 1994). En *Caenorhabditis elegans* en cambio, no se han identificado prácticamente retroelementos y la mayoría de elementos transponibles parecen ser de DNA.

Este estudio de contenido y distribución de elementos repetitivos se ha visto enriquecido con la información que puede extraerse de los genomas que ya están totalmente secuenciados ([ncbi.nlm.nih.gov/genbank/genomes](http://ncbi.nlm.nih.gov/genbank/genomes)).

Con todo, los RTN son la clase más extendida de TE en eucariotas y el genoma de las plantas contiene una cantidad extraordinariamente abundante y diversa de RTN

comparándolo con el resto de genomas eucariotas. A continuación, se desarrolla con mayor detalle la contribución de los RTN al genoma de plantas.

### **3.2.1 Contribución de los retrotransposones al genoma de plantas**

Los RTN tienen el potencial para alcanzar un número de copias muy elevado dado el tipo de transposición replicativa que presentan.

Los RTN-LTR tipo  *copia* se encuentran en un número de copias muy alto en numerosas plantas y comprenden fracciones importantes de sus genomas (Pearce y Harrison, 1996; San Miguel et al., 1996). Por ejemplo, el RTN *BARE-1* está presente en más de 50.000 copias en el genoma de cebada (Suoniemi et al., 1996) y *Opie1* en más de 30.000 copias en el genoma de maíz (SanMiguel et al., 1996).

También encontramos ejemplos de RTN-LTR tipo *gypsy* altamente repetitivos, como *Cinful-1* con 20.000 copias en el genoma de maíz (SanMiguel y Bennetzen, 1998) y *Del1* con 13.000 copias en *Lilium henryi* (representa el 0,4% del genoma) o más de 40.000 copias en *Lilium longiflorum* (1% del genoma) (Leeton y Smyth, 1993).

En lo que respecta a los RTN no-LTR, por ejemplo el elemento LINE *Del2* representa el 4% del genoma de *Lilium speciosum* con 250.000 copias (Leeton y Smyth, 1993) y entre los elementos SINE, destacar las 50.000 copias de la familia TS en tabaco o las 500 copias de S1<sub>BN</sub> en *Brassida napus* (Deragon et al., 1994).

### **3.2.2 Distribución de los retrotransposones en el genoma de plantas**

Diversos estudios han demostrado la ubicuidad de estos elementos. Así, podemos encontrar ejemplos de RTN-LTR tipo  *copia* (Voytas et al., 1992; Flavell et al., 1992; Suoniemi et al., 1998), RTN-LTR tipo *gypsy* (Suoniemi et al., 1998), LINEs (Schmidt, 1999; Noma et al., 1999), SINE (Deragon et al., 1994) así como MITEs (Bureau y Wessler, 1992, Feschotte et al., 2002) caracterizados en multitud de especies, tanto en gimnospermas como en angiospermas, abarcando todo el reino vegetal.

El modo replicativo de transposición que presentan los RTN hace que hayan invadido con éxito prácticamente todas las partes del genoma de las plantas, llegando a ocupar en ocasiones hasta el 80% del DNA nuclear (SanMiguel et al., 1998).

En plantas, cada grupo de RTN tiene familias que se concentran o que son deficientes en regiones específicas de los cromosomas (Kumar y Bennetzen, 1999). Así, aproximaciones mediante Hibridaciones *in situ* (FISH) de cromosomas metafásicos o en

## **Introducción**

---

profase nuclear, revelan que los RTN-LTR tipo *copia* se transponen preferentemente de manera dispersa por la eucromatina concretamente en regiones intergénicas, mientras que en *Arabidopsis thaliana* se concentran en regiones paracentroméricas (EUROPEAN UNION ARABIDOPSIS GENOME SEQUENCING CONSORTIUM, 1999). Los elementos tipo *gypsy* en cereales se sitúan preferentemente en la región centromérica y en un tipo concreto de disposición unos dentro de otros en forma anidada o *nested*, tal y como se observa al analizar la región de la *Adh1* en maíz (SanMiguel et al., 1996). Hay por ejemplo, elementos tipo *copia* que son abundantes preferentemente en regiones heterocromáticas terminales de los cromosomas de *Allium cepa* [cebolla, familia Liliaceae] pero en cambio no están presentes en la misma región cromosómica de *Secale cereale* [centeno, familia Graminaeae] (Kumar y Bennetzen, 1999). Vemos por tanto, que la localización de los RTN dependerá tanto de la especie vegetal como del elemento analizado.

No hay evidencias claras sobre las preferencias de inserción de los RTN en plantas. En algunos casos, se localizan en regiones cromosómicas que se están transcribiendo activamente, particularmente en o cerca de promotores (White et al., 1994; Hirochika et al., 1996), y en otros casos se localizan en regiones repetitivas heterocromáticas (Heslop-Harrison et al., 1997). Aún así, en general suelen encontrarse ampliamente dispersos por los cromosomas, preferentemente en zonas intergénicas y dentro de otras secuencias repetitivas. Esto lleva a pensar que los RTN han evolucionado hacia una movilización dentro de zonas silenciosas del genoma, para evitar aumentar la tasa de mutación génica y no llegar a niveles letales. Así, los RTN pueden proliferar como secuencias repetidas dispersas sin resultar deletéreos, contribuyendo activamente al aumento del tamaño del genoma huésped. Hay estudios que confirman la localización de inserciones recientes del elemento *Tnt1* (RTN-LTR tipo *copia*) en regiones génicas de tabaco mientras que las inserciones más antiguas están concentradas en zonas repetitivas incluyendo TE. Este resultado sugiere que las nuevas inserciones de *Tnt1* se dan cerca de genes pero con el tiempo los RTN-LTR van siendo excluidos de las regiones génicas (Quang Hien LE, ICPMB 2003).

### **3.2.3 Variabilidad de los retrotransposones en el genoma de plantas**

Muchos estudios sobre la variabilidad de los RTN-LTR se basan en la amplificación por PCR utilizando cebadores degenerados complementarios a secuencias altamente conservadas de la región de la RT (Voytas et al., 1992; Flavell et al., 1992; Suoniemi et al., 1998). Mediante esta aproximación, se ha conseguido detectar específicamente elementos tipo *copia* en el genoma de multitud de especies vegetales.

El análisis de las secuencias obtenidas muestra que estos elementos además de ser ubicuos y presentarse en un número muy elevado en el genoma de las plantas, son extremadamente heterogéneos, mucho más que en *Drosophila* o levadura (Flavell et al., 1992). Estudios en *Vicia spp.* muestran que el número de copias de RTN-LTR tipo *copia* está correlacionado con la heterogeneidad de las secuencias (Pearce et al., 1996). Pero, esta variabilidad en los RTN-LTR no se explica únicamente por el hecho de que se encuentren en un número de copias tan alto, ya que hay familias de RTN-LTR tipo *copia* en *Arabidopsis* como la *Ta* que aun siendo extremadamente pequeña es altamente heterogénea (Flavell et al., 1992).

La presencia de familias de RTN-LTR tipo *copia* tan heterogéneas, sugiere que estos elementos existían ya en pasos muy tempranos de la evolución y que han divergido en subgrupos altamente heterogéneos antes de que aparecieran los órdenes de plantas modernas (Gribbon et al., 1999; White et al., 1994).

El propio ciclo de vida del RTN-LTR, similar al de los retrovirus, genera tasas de mutación muy elevadas dada la baja fidelidad y la incapacidad para corregir errores de las enzimas implicadas en el proceso (RNA polimerasa II y RT) (Vershinin et al., 1999).

### **3.2.4 Impacto de los retrotransposones en la evolución del genoma de plantas**

- **Los TE como parásitos genómicos**

La naturaleza parásita intrínseca de los TE viene de su habilidad para incrementar su número de copias y/o mantener su supervivencia en nuevas especies a través de largos periodos evolutivos. El ciclo en el linaje filogenético de un TE pasa por las fases de replicación dinámica, inactivación y degradación. La transposición es un mecanismo con errores (las enzimas implicadas en el proceso carecen de mecanismos para su corrección) que produce elementos no autónomos los cuales a menudo reprimen la tasa de transposición de los elementos activos. En estos casos la transferencia horizontal a nuevos huéspedes es una posible vía para los TE activos para iniciar un nuevo ciclo.

Hasta el momento únicamente hay evidencias claras de transferencia horizontal en elementos de Clase II. Los ejemplos más concluyentes son para los TE como los elementos P o *mariner* en insectos (Robertson, 1993). En lo que se refiere a retroelementos, hay ejemplos en *Drosophila melanogaster* donde el elemento *gypsyDm* que codifica para una proteína ENV funcional (proteína de la cubierta) es capaz de movilizarse (Song et al., 1994) e incluso en cocultivos permite la transferencia celular

## **Introducción**

---

de RTN que carecen del gen *env* (Syomin et al., 2002). En plantas, sin embargo las evidencias para los retroelementos son ambiguas (Voytas et al., 1992; Hirochika et al., 1993; Kidwell et al., 1992).

- **Los TE como fuente de variación genética**

El tipo de variación genética introducida por los TE va desde cambios en el tamaño y reorganización de todo el genoma (actúan como moduladores del genoma) hasta simples cambios nucleotídicos, lo cual puede producir bien cambios en caracteres fenotípicos o bien cambios silenciosos sólo detectables a nivel de DNA.

En general, las diferencias en el contenido de RTN es una de las razones de la elevada variabilidad en el tamaño de los genomas de plantas. Se dice por tanto que actúan como **moduladores del genoma**. Así por ejemplo, estudios sobre la contribución de los RTN al genoma de los cereales sugieren que la diferencia en el tamaño del genoma de maíz (2.400 Mb) y sorgo (1.200 Mb) está mayoritariamente provocada por la acumulación de RTN en el maíz por inserciones que tuvieron lugar hace entre 2 y 6 millones de años esto es, después de que se diera la divergencia entre estas especies (SanMiguel et al., 1998).

Además de la variación genética que generan, la transposición de TE en zonas génicas puede modificar las secuencias promotoras, terminadoras o truncar las secuencias codificantes de los genes donde se insertan, actuando como **agentes mutagénicos**. Así por ejemplo, la inserción de los transposones Ac/Ds en el locus *waxy* de maíz provoca pérdida de pigmentación del grano tal y como determinó B. McClintock, o la inserción del elemento *Mutator* en la caja TATA del gen *Adh1* provoca cambios en la especificidad de tejido de la transcripción en maíz (Kidwell y Lisch, 1997). Por otro lado, la inserción de RTN-LTR en regiones intergénicas puede modificar también la expresión de los genes adyacentes dada la presencia de secuencias reguladoras en las LTR.

Muchos de los RTN-LTR de plantas se han identificado por estar insertados en alelos mutados. Así, *Tnt1* (RTN-LTR tipo *copia*) se aisló a partir de una mutación en el gen de la nitrato reductasa en cultivos celulares de tabaco (Grandbastien et al., 1989) o *Magellan* (RTN-LTR tipo *gypsy*) provoca el fenotipo mutante *waxy* al quedar insertado en el locus *Waxy-M* (Purugganan y Wessler, 1994).

- **Co-Evolución del TE y en el genoma huésped**

La transposición parece haber tenido un importante papel en la evolución de la estructura y regulación de los genes de plantas pues se ha visto que los RTN-LTR y los MITE tienen importancia como modificadores de la expresión de genes de la planta (Wessler et al., 1995). Muchos genes de plantas presentan en sus regiones flanqueantes reminiscencias de RTN-LTR tipo *copia* que les aportarían secuencias reguladoras. En algunos casos, estas secuencias similares a RTN-LTR se encuentran en la misma posición en diferentes miembros de una misma familia génica, lo cual indica que la inserción del RTN es anterior al proceso de duplicación génica (White et al., 1994).

Por otro lado, una transposición no regulada es perjudicial para el organismo huésped. Para permitir la co-existencia del TE en el genoma huésped se han desarrollado una serie de mecanismos que reducen los efectos adversos que la movilización de estos elementos pueden provocar en la supervivencia de la célula huésped (**Tabla III**).

Mecanismo	Ejemplo
Inserción orientada hacia regiones no codificantes	Inserción preferencial en regiones reguladoras, como los MITE en la región 5' no codificante de los genes (Wessler et al., 1995)  Inserción en elementos preexistentes (en <i>nested</i> ), muy típico en maíz (SanMiguel et al., 1996)
Transposición específica de tejido	Represión de la transposición de los elementos P de <i>Drosophila</i> en las células somáticas
Regulación del número de copias	Metilación  Autorregulación, como en <i>Ac</i> y <i>Spm</i> que cuando están en un número de copias elevado reducen su transposición (Kidwell y Lisch, 1997).

**Tabla III. Mecanismos para minimizar los efectos adversos de la movilización de elementos transponibles en el genoma huésped**, modificado de Kidwell y Lisch, 1997.

En el **Apartado 4.4** se explican más ampliamente estos mecanismos desarrollados por los genomas eucariotas para disminuir la actividad de los TE y así controlar su capacidad mutagénica.

Un claro ejemplo de que los TE pueden llegar a realizar funciones vitales para la célula huésped lo encontramos en *Drosophila* donde retroelementos como HET-A y TART tienen la función de reemplazar los finales dañados de los cromosomas, mientras que en otros insectos la integridad de los cromosomas la mantienen normalmente las telomerasas (Pardue et al., 1996).

### 4. LOS RETROTRANSPOSONES-LTR

En este trabajo nos centramos en los retrotransposones con LTR (RTN-LTR) de plantas (Revisiones: Casacuberta et al., 2003; Feschotte et al., 2002; Kumar y Bennetzen, 1999).

#### 4.1 CARACTERÍSTICAS ESTRUCTURALES DE LOS RTN-LTR

Los RTN-LTR poseen en sus extremos dos repeticiones largas terminales (LTR, *Long Terminal Repeats*) que flanquean el dominio interno el cual codifica las proteínas necesarias para completar el ciclo de vida del retrotransposón, tal y como se describirá más adelante.

A continuación del final de la LTR5' se encuentra el PBS (*Primer Binding Site*) y la LTR3' está precedida por una región rica en purinas (PPT, *Poly Purine Track*), que son los sitios de unión al cebador que utilizan los RTN para la síntesis de la cadena de cDNA (-) y cDNA(+) respectivamente, durante la retrotranscripción. La LTR5' proporciona el iniciador funcional de la transcripción del elemento mientras que la LTR3', el terminador funcional. Las LTR son además importantes para la integración del elemento ya que la integrasa reconoce sus extremos.

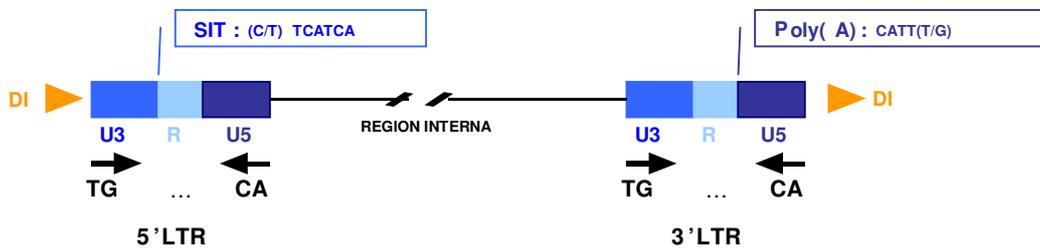
##### 4.1.1 LTR

Las LTR están situadas en idéntica orientación en ambos extremos del RTN-LTR, y pueden tener tamaños comprendidos entre 100 pb (*Tos17*, Hirochika et al., 1992) y 5,2 Kb (*Ogre*, Newmann et al., 2003).

Una de las características intrínsecas de las LTR es que ambas son idénticas en secuencia en el momento de la inserción del elemento. Con el tiempo las LTR van acumulando mutaciones y el número de sustituciones entre ambas pueden servir para datar el momento de la inserción (SanMiguel et al, 1998).

Ambas LTR presentan repeticiones cortas invertidas (IR, *Inverted Repeat*) a los extremos. El consenso descrito para los RTN-LTR es de 5' -"TG..CA" -3', pero el número de nucleótidos repetidos varía de unos RTN o otros (Neuvéglise et al., 2002).

Las LTR están flanqueadas por una duplicación de la diana de inserción (DI) o TSD (*Target Site Duplication*) del genoma huésped. Cuando se inserta un RTN, un segmento corto del DNA celular (normalmente 4 - 5 pb) se duplica en el sitio de inserción como consecuencia del propio proceso.



**Figura 7: Organización de las LTR de un RTN-LTR.** Las LTR son repeticiones directas que flanquean la región interna del elemento, y se dividen típicamente en tres dominios: U3, R y U5. La transcripción del elemento comienza en el sitio de inicio de la transcripción (SIT) de la LTR5', que delimita la región U3 y R. El final de la transcripción se produce en la LTR3', donde está el sitio de poliadenilación funcional que delimita las regiones R y U5. En la figura se indican las secuencias consenso en plantas para el SIT (Gómez, 2002) y el sitio de poliadenilación. Se señalan también las repeticiones cortas invertidas terminales "TG..CA" y la duplicación de la diana de inserción (DI) a ambos extremos del elemento.

Las LTR pueden estar divididas en tres dominios funcionales contiguos:

- **U3:** Comprende desde el lugar de inicio de la LTR hasta el lugar de inicio de la transcripción (SIT).

Contiene una secuencia promotora para la RNA polimerasa II, presente en ambas LTR pero sólo actúa como promotor fuerte en la LTR5'. La secuencia de la LTR3' podría actuar como promotor débil para los genes adyacentes al lugar de inserción (White et al., 1994).

Se ha visto que deleciones en la región U3 del RTN-LTR tipo *copia Tnt1* de tabaco provocan la inhibición de la transcripción del elemento ya sea, por pérdida de la actividad transcripcional, defectos en la terminación de la transcripción o disminución en la estabilidad del mRNA (Casacuberta et al., 1995).

En esta región U3, no sólo está presente la caja TATA, sino también las cajas responsables de la regulación de la transcripción en diferentes tejidos o en diferentes condiciones ambientales, como se ha visto en *Tnt1* de tabaco (Vernhettes et al., 1997; Beguiristain et al., 2001), *BARE-1* de cebada (Suoneimi et al., 1996) y *Tto1* de arroz (Takeda et al., 1998).

La variabilidad de secuencia en las regiones reguladoras del dominio U3 sirve para definir las diferentes subfamilias de *Tnt1* (Vernhettes et al., 1998).

## **Introducción**

---

- **R:** Es la zona comprendida entre el inicio y el final de la transcripción. Este dominio puede presentar agrupaciones repetidas en tándem.

Aunque ambas LTR presentan las secuencias, el iniciador funcional de la transcripción está en la LTR5' y el terminador funcional en la LTR3'. La región R se encuentra a ambos extremos del transcrito. Permite a la cadena (-) corta de cDNA saltar de la LTR5' a la LTR3' durante la transcripción inversa (ver **Figura 9**).

- **U5:** Comprende desde el sitio de poliadenilación (que marca el final de la transcripción) y el final de la LTR.

### **4.1.2 PBS (Primer Binding Site) y PTT (Poly Purine Track)**

A continuación del final de la LTR5' se encuentra el **PBS** (*Primer Binding Site*) que es el sitio de unión al cebador que utilizan los RTN para el inicio de la síntesis de la primera cadena de cDNA (-).

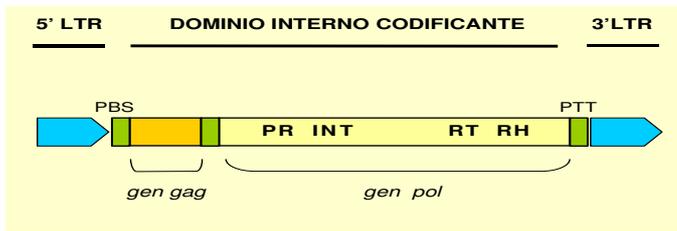
Es una secuencia complementaria a una parte de un tRNA, el más común en plantas es complementario al tRNA metionina, de secuencia TGGTATCAGAGC, aunque se han descrito casos de RTN que contienen PBS complementarios a otros tRNA, como el tRNA de Glu en *Bagy2* de cebada (Vicient et al., 2001) y *Cyclops2* de guisante (Chavanne et al., 1998), o el tRNA de Arg en *Grande* de maíz (Vicient, 1995) o *RIRE2* de arroz (Ohtsubo et al., 1999).

La LTR3' está precedida por una zona rica en purinas o **PPT** (*Poly Purine Track*) de 9-17 pb, que actúa como lugar de inicio de la síntesis de la segunda cadena de cDNA (+) durante la retrotranscripción.

### **4.1.3 DOMINIO INTERNO**

Los RTN-LTR se movilizan a una nueva localización cromosómica con un paso intermedio de retrotranscripción.

El dominio interno codifica para todas las actividades enzimáticas necesarias para el ciclo de vida autónomo del RTN. Se sintetizan en forma de poliproteína, que posteriormente se procesa en diferentes péptidos funcionales mediante una proteasa codificada por el propio elemento. Tal y como se ha comentado anteriormente, el gen *gag* codifica para proteínas estructurales de la cápside. El gen *pol*, codifica para actividades enzimáticas como la proteasa (**PR**), integrasa (**INT**), transcriptasa inversa (**RT**) y RNasa H (**RH**). En algunos RTN hay excepcionalmente un dominio a 3' extra, similar al gen *env* de retrovirus.



**Figura 8. Organización estructural de un RTN-LTR.** Las LTR flanquean el dominio interno que contiene el gen *gag* (qué codifica proteínas estructurales de la cápside), el gen *pol* (qué codifica actividades enzimáticas como la proteasa (PR), integrasa (INT), transcriptasa inversa (RT) y RNasa H (RH)). El elemento representado es de tipo *copia* pues el dominio de la INT está situado a 5' de la RT.

## El gen *gag*

El gen *gag* codifica proteínas estructurales que sirven para formar la partícula similar a la viral (VLP, *Virus Like Particle*) como se ha visto en RTN de levadura (Palmer et al., 1997), *Drosophila* (Bachmann et al., 2004) y cebada (Jääskeläinen et al., 1999).

Esta región está poco conservada a nivel de secuencia en los distintos RTN, pero en ella se ha identificado uno o varios dominios ricos en cisteínas (CX<sub>2</sub>CX<sub>3</sub>GHX<sub>4</sub>C) que se presentan altamente conservados en diferentes RTN y retrovirus. Se cree que este motivo es necesario para la unión del RNA genómico a la VLP (Covey, 1986), así como para colocar el tRNA iniciador en posición correcta para el inicio de la retrotranscripción (Vicent, 1995; Kossack y Kinlaw, 1999).

## El gen *pol*

El gen *pol*, codifica diferentes actividades enzimáticas y está situado siempre a continuación del gen *gag*, pudiendo estar codificado por la misma pauta de lectura (ORF, *Open Reading Frame*) o por una ORF propia.

Como el ensamblaje de la VLP requiere un exceso de proteínas estructurales respecto a las catalíticas, es necesario que haya algún mecanismo de regulación de la expresión de ambos tipos de proteínas. En algunos virus, las regiones *gag* y *pol* están separadas por un codón de parada y la traducción del mRNA da lugar a proteínas GAG, pero ocasionalmente el codón de parada es substituido por una tRNA para glutamina y se produce una proteína de fusión GAG-POL. Un mecanismo muy común en los RTN es el uso de cambios de pauta de lectura mediante la formación de *stem loop* o por "codones raros" en el RNA. Otros RTN presentan un control post-transcripcional de la proporción adecuada de GAG frente a POL ya que, aunque se sintetiza una poliproteína GAG-POL posteriormente la proteína madura POL tiene menor estabilidad proteica o bien se degrada preferencialmente. Otros RTN tienen directamente 2 ORF.

## Introducción

---

- **Proteasa (PR)**

La proteasa codificada por el propio elemento provoca el autoprocesamiento de la poliproteína dando lugar a los péptidos funcionales.

Las proteasas de RTN y retrovirus son de tipo aspártico y contienen un motivo altamente conservado que actúa como centro activo: D(T/S)G. La forma activa de la proteína es un dímero.

La especificidad de corte de las PR ha sido estudiada en algunos retrovirus, y no refleja la existencia de una secuencia específica de corte.

- **Retrotranscriptasa (RT)**

La RT convierte las moléculas de RNA en DNA de doble cadena. Para ello consta de actividad DNA-polimerasa dependiente de RNA [copia el RNA genómico a cDNA (-)], actividad DNA-polimerasa dependiente de DNA [sintetiza la segunda cadena de cDNA (+)] y actividad Ribonucleasa H [elimina el molde de RNA del híbrido DNA-RNA].

Los retrovirus mutan en una tasa  $10^4$ - $10^6$  veces mayor que los genes celulares. Aunque no se conoce el dato exactamente para RTN se supone que la tasa de mutación es similar, debido a que las enzimas implicadas en el proceso de replicación del RTN (RNA polimerasa II y RT) carecen de mecanismos de corrección de errores. La RT presenta una tasa de sustitución 1,4 veces inferior a la INT.

El dominio RT es el más conservado a nivel de aminoácidos dentro de los RTN (Doolittle et al., 1989). Existen una serie de secuencias altamente conservadas como el motivo YVDD del centro activo de la RT, así como las secuencias TAFLHG y SLYGLKQA típicas de RT de RTN-LTR tipo *copia* (Flavell et al, 1992) o el motivo SKCEF en RTN-LTR tipo *gypsy* con el residuo lisina (K) siempre presente en las polimerasas RNA-dependientes (Suoniemi et al., 1998).

- **RNAse H (RH)**

La actividad RH se sitúa en la zona C-terminal de la misma proteína que la RT y es capaz de degradar el RNA de un híbrido RNA-DNA. *In vitro* posee actividad endo y exonucleasa 3' y 5'. Durante el proceso de retrotranscripción la RH elimina el molde de RNA (para permitir la síntesis de la segunda cadena de DNA) y los iniciadores de RNA una vez acabado el proceso.

Se han definido para la RH dos motivos altamente conservados en los distintos RTN, el motivo1 de secuencia RTKHI(D/E) y el motivo2 de secuencia AD(I/M/F/L)LTK (Pearce et al., 1999). También está muy conservado el motivo TDAS situado en la mitad N-terminal de la proteína.

- **Integrasa (INT) o endonucleasa**

La integrasa es necesaria para la inserción del cDNA del RTN en el genoma.

La proteína presenta en la zona N-terminal un motivo de dedo de zinc cuya secuencia  $HX_{3-7}HX_{20-32}CX_2C$  es de las más conservadas en los diferentes RTN. Parece ser que este motivo interviene en el reconocimiento del DNA de la LTR previo a la integración, aunque no es imprescindible para que se de el proceso. En la zona central de la proteína se encuentra el dominio catalítico cuya secuencia  $DD_{35}E$  está perfectamente conservada en INT retrovirales y de RTN y es esencial para su actividad enzimática (Suoniemi et al., 1998). La zona C-terminal de la proteína es una región muy básica, que une inespecíficamente DNA, pero no parece estar demasiado conservada en los distintos RTN o retrovirus.

En algunos casos, por ejemplo el RTN Ty1 de *Saccharomyces cerevisiae*, se ha visto que la INT tiene dos dominios diferentes, uno requerido para la integración y otro requerido para la localización nuclear [presenta una señal de localización nuclear bipartita (NLS)]. Este mecanismo de direccionamiento nuclear del complejo de preintegración mediado por la INT puede ser una estrategia general utilizada por los RTN y retrovirus para infectar células que no están en división (Gómez, 2002).

### El gen *env*

La diferencia entre retrovirus y retrotransposones se basa en la presencia del gen *env* que codifica para una proteína transmembrana necesaria para la expansión extracelular.

El gen *env* está situado en el extremo 3' del dominio interno contiguo al inicio de la LTR3', en una pauta de lectura diferente (ORF3) que codifica las glucoproteínas de la cubierta que reconocen proteínas receptoras de la superficie de la célula huésped, permitiendo su fusión con la membrana plasmática y posterior entrada a la célula, facilitando la movilidad celular y la infectividad.

Algunos RTN de invertebrados, entre ellos *gypsy* de *Drosophila*, contienen dominios *env*. La secuencia de *gypsyDm* (de *Drosophila melanogaster*) contiene una ORF3 que codifica para una proteína ENV totalmente funcional y además estas secuencias pueden ser transferidas horizontalmente. Se han encontrado elementos idénticos a *gypsyDm* en poblaciones naturales de *Drosophila suboscuro* coexistiendo con elementos *gypsyDs* (de *Drosophila suboscuro*) indicando que *Drosophila suboscuro* ha sido invadida por elementos *gypsyDm* infecciosos. Este es el primer retrovirus de insectos descrito. En cambio las ORF3 presentes en *gypsyDs* y *gypsyDv* (de *Drosophila virilis*) carecen de algunos de los motivos esenciales para la función de esta proteína, por lo cual se piensa que pueden ser formas degeneradas de estos retrovirus (Alberola et al., 1996; Song et al., 1994).

No se ha documentado la presencia de retrovirus en plantas pero sí de pararetrovirus (virus que encapsidan DNA y cuya multiplicación se da vía retrotranscripción, aunque no se integran en el genoma huésped) como el virus del mosaico de la coliflor (CaMV).

Pero en algunos miembros de RTN del tipo *gypsy* es frecuente encontrar una región más o menos larga entre la integrasa y la LTR3' que recuerda al gen *env*, como en *Tat4-1* y *Athila1-1* de *Arabidopsis* (Wright y Voytas, 1998), *Calypso* de haba (Peterson-Burch et al., 2000), *Cyclpos2* de guisante (Chavanne et al., 1998), *Bagy2* de cebada o *Rigy2* de arroz (Vicent et al., 2001). Además, se ha visto que *Bagy2* es capaz de transcribirse y que el dominio *env* presenta *splicing* en el RNA genómico de una manera similar a la vista para retrovirus, además la proteína ENV predicha contiene los dominios típicos de proteínas retrovirales (Vicent et al., 2001).

También encontramos ejemplos dentro del grupo *copia* como SIRE1, elemento aislado de haba, que contiene una segunda ORF en el extremo 3' que codifica una proteína de 70 kDa similar a ENV (Laten et al., 1998). Este elemento está considerado como un proretrovirus.

### 4.2 CICLO DE VIDA DE LOS RTN-LTR

En la **Figura 9** está representado el ciclo de vida de un RTN-LTR autónomo.

La **transcripción** de los RTN-LTR se realiza mediante una RNA polimerasa II celular. Se inicia en el extremo 5' del dominio R de la LTR5' y termina en el extremo 3' del dominio R de la LTR3', formándose una única molécula de mRNA (**Figura 9, Paso 1**).

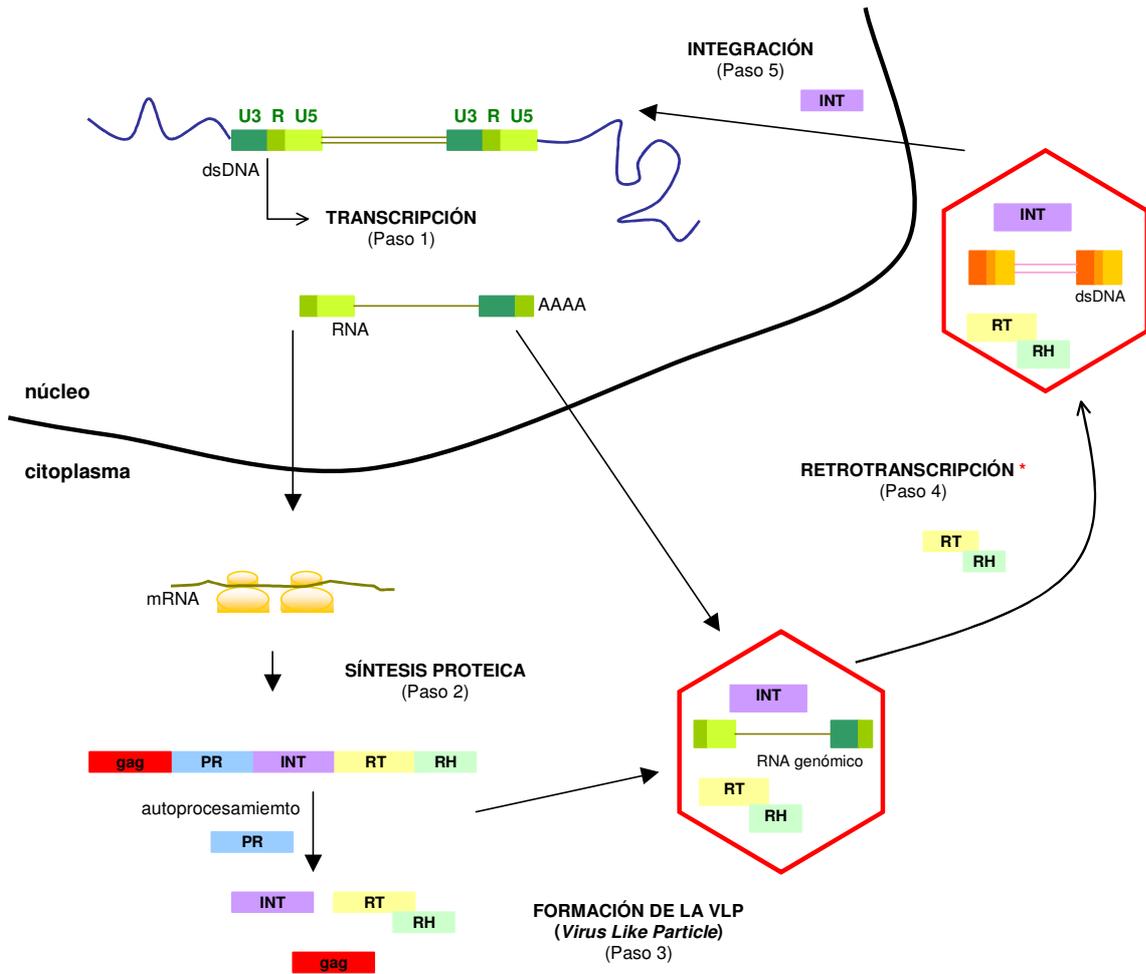
Algunos de los RNA son traducidos y se generan las proteínas necesarias para la replicación. Es la proteasa codificada por el propio elemento la que provoca el autoprocésamiento de la poliproteína dando lugar a los péptidos funcionales (**Figura 9, Paso 2**). Otros RNA actúan como RNA genómico haciendo de molde para la retrotranscripción.

A continuación se da el ensamblaje de la VLP (*Virus Like Particle*). Parte de las proteínas generadas se unen sobre el RNA genómico formando las partículas semejantes a virus (VLP) donde tiene lugar la replicación (**Figura 9, Paso 3**).

La **retrotranscripción (Paso 4)** se inicia con la síntesis de una cadena de cDNA (-) complementaria al RNA genómico. La secuencia del PBS del RNA es complementario a un RNA celular, normalmente el extremo 3' de tRNA, que la RT utiliza como cebador para sintetizar una cadena (-) corta complementaria a los dominios R-U5 de la LTR5' (**4a**). La polimerasa de la RT deja de copiar dada la presencia de una señal de parada en el dominio R de la cadena (-), la RH degrada la región R y U5 del RNA, y a continuación la cadena corta de DNA copiada es transferida al final 3' de la molécula de RNA (por complementariedad de los dominios R) donde se continúa la síntesis de la cadena DNA (-) hasta llegar a la señal de parada del dominio R (**4b**). Conforme se avanza en la síntesis de DNA, la RH va degradando el RNA del híbrido RNA-DNA resultante. La síntesis de la segunda cadena de DNA se inicia al unirse el cebador a la PTT y se sintetizan las regiones U3-R-U5 y PBS (**4c**). Se produce un nuevo salto al extremo 5' (donde hay complementariedad de secuencias del PBS) y se continúa con la síntesis de las dos cadenas de DNA (**4d**).

Cuando se completa la síntesis del dsDNA, el elemento nuevo posee ambas LTR y puede integrarse en una nueva localización cromosómica gracias a la acción de la INT que reconoce las secuencias de las LTR (**Figura 9, Paso 5**). Durante la inserción del RTN, la INT produce un corte asimétrico en el DNA huésped dando lugar a extremos cohesivos que luego son reparados a ambos lados del elemento provocando la duplicación de la diana de inserción (DI) del genoma huésped.

## Introducción



### \* RETROTRANSCRIPCIÓN (Paso 4)

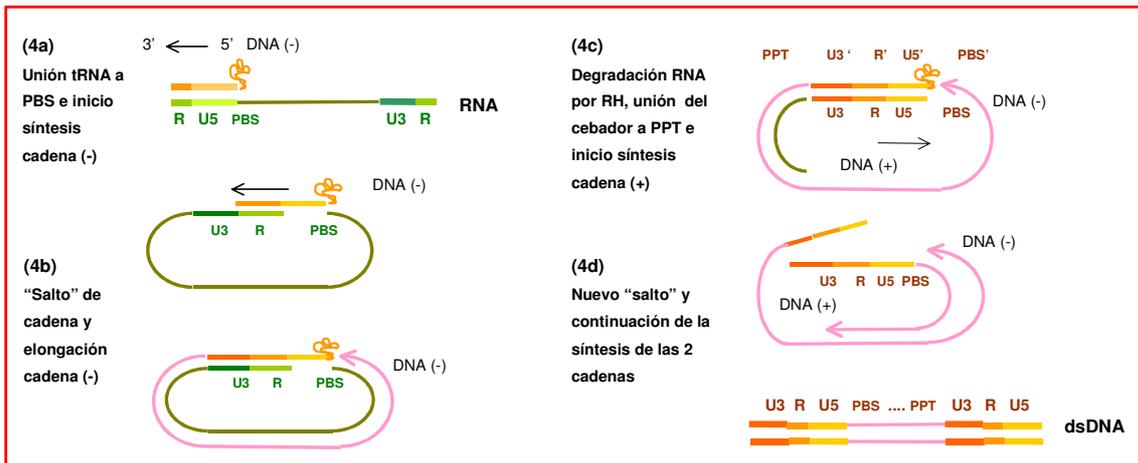


Figura 9. Ciclo de vida de los RTN-LTR. Los diferentes pasos del ciclo están desarrollados en el texto

Basándose en los conocimientos actuales, el mecanismo de transcripción, replicación e integración de los retroelementos es igual para RTN-LTR tipo *copia*, tipo *gypsy* y retrovirus, pero los retrovirus tienen un paso extracelular al presentar cubierta (codificada por gen *env*) (Kumar y Bennetzen, 1999).

### **4.3 RETROTRANSPOSONES-LTR DE PLANTAS CARACTERIZADOS**

Los dos tipos de RTN-LTR, *gypsy* y *copia*, se diferencian entre sí por el orden relativo de sus dominios enzimáticos dentro del gen *pol* y por la conservación a nivel de secuencia de sus dominios internos, en referencia a su similitud con los prototipos *GypsyDm* (Marlon et al., 1986) y *CopiaDM* (Mount y Rubin, 1985) descritos en *Drosophila melanogaster*.

Los RTN-LTR tipo *copia* fueron los primeros identificados en plantas (Voytas et al., 1992) y son el grupo mejor estudiado. La ubicuidad de los RTN-LTR tipo *gypsy* en plantas está también claramente demostrada (Suoniemi et al., 1998).

En la siguiente tabla se describen las características principales de algunos de los RTN-LTR aislados y caracterizados del genoma de plantas (**Tabla IV**).

Elemento	Tipo de RTN	Especie en que se aisló	Tamaño (pb)	Longitud LTR (IR) (pb)	Nº copias en la especie huésped	tRNA	Otras características estructurales	Referencia	Nº Acc. BBDD
BARE1	copia	<i>Hordeum vulgare</i> (cebada)	12000	1.829 [6]	5.000	Met	4 % divergencia LTR. Diana inserción: GAAC	Manninen y Shulman, 1993	Z17327
Panzee	copia	<i>Cajanus cajan</i>	4.947	372-383 [4]	n.d	Met	3,3 % divergencia LTR. Diana inserción de 5pb	Lall et al., 2002	AJ000893
Tnt1	copia	<i>Nicotiana tabacum</i> (tabaco)	5.300	610	> 100	Met	Diana inserción de 5pb	Grandbastien et al., 1989	X13777
Tto1	copia	<i>Nicotiana tabacum</i> (tabaco)	5.500	574 [2]	> 30	Met	0 % divergencia LTR. Diana inserción de 5pb	Hirochika, 1993	D83003
Bs1	copia	<i>Zea mays</i> (maíz)	3.200	302	1-5			Jim y Bennetzen, 1989	M25397
Opie1	copia	<i>Zea mays</i> (maíz)	8.700	1.256	>30.000			SanMiguel et al., 1996	U68408
Tos17	copia	<i>Oryza sativa</i> (arroz)	4.100	138	1-5			Hirochika, 1992, 1996	D85876
Athlia	gypsy	<i>Arabidopsis thaliana</i>	10.500	1539 [5]	30	Asp		Pelissier et al, 1995	X81801
Cintul1	gypsy	<i>Zea mays</i> (maíz)	8.600	586	20.000			SanMiguel y Bennetzen, 1998	
Cyclops2	gypsy	<i>Pisum</i> (guisante)	12.500	1.504	1.000	Asp		Chavanne et al., 1998	AJ000640
Del1	gypsy	<i>Lilium henry</i>	9.300	2.400	13.000	Met		Smyth et al., 1989	X13886
Grande1	gypsy	<i>Zea</i> (maíz)	13.500	600	>1300	Arg		Martinez-Izquierdo et al., 1997	
Magellan	Gypsy	<i>Zea mays</i> (maíz)	5.700	341 [2]	4-8	Met		Purugganan et al., 1994	
Fire2	gypsy	<i>Oryza sativa</i> (arroz)	11.000	441 [3]	5.000	Arg	Posee ORF extra a 3'	Wright et al., 1998	
Bagy2	gypsy	<i>Hordeum vulgare</i> (cebada)	10.000	1520-1537	10.000	Glu		Ohtsubo et al., 1999	AF254799-
PDR1-1	copia	<i>Pisum sativum</i> (guisante)	3925	156	200		0,6% divergencia LTR	Shirasu et al., 2000	AF279072
								Vicient et al., 2001	
								Lee et al., 1990	X66399

### 4.4 ACTIVIDAD DE LOS RTN-LTR

Tal y como se ha comentado anteriormente, los RTN-LTR tienen un mecanismo de transposición replicativo que los convierte en elementos potencialmente muy invasivos. Por este motivo, el control de la distribución y número de copias en que se encuentran los RTN-LTR es crucial para minimizar los posibles efectos deletéreos en el genoma huésped. Los genomas eucariotas parecen haber desarrollado mecanismos para reducir la actividad de los elementos móviles y así controlar su capacidad mutagénica. Los mecanismos de silenciamiento son probablemente los más generales y efectivos. Por ello, la mayor parte de los RTN-LTR de plantas están silenciados en la mayoría de tejidos vegetales durante el desarrollo.

#### 4.4.1 Regulación de la expresión y la transposición en RTN-LTR

Como hemos visto, los RTN-LTR no pueden transponerse sin la presencia de RNA que actúe como molde para la retrotranscripción así como la producción de mRNA para la síntesis proteica, por tanto uno de los pasos fundamentales para regular la actividad de un elemento será controlando el inicio de la transcripción.

##### **A) Regulación a nivel transcripcional**

El silenciamiento transcripcional (TGS, *Transcriptional Gene Silencing*) es un mecanismo de inactivación del promotor de las secuencias repetidas que directamente evita su transcripción. Hay varios factores que inducen el TGS pero parece ser que el más importante es la presencia de múltiples copias de la secuencia en concreto, así como la presencia de siRNA (*short interfering RNA*) correspondientes a secuencias de RTN (Casacuberta et al., 2003).

Las bases de datos EST (*Expressed Sequence Tag*) pueden aportar una idea de la representación general que tienen los transcritos de RTN en los diferentes genomas. Por ejemplo, para cereales el 0,1% de los clones en bases de datos EST corresponden a transcritos de RTN, llegando a ser del 0,3% en cebada (Vicient et al., 2001).

Aunque muchos de los RTN-LTR aislados no se transcriben en condiciones normales, hay en la actualidad diversos ejemplos de RTN-LTR que muestran patrones de transcripción en determinados tejidos o momentos de desarrollo. Así, se han detectado transcritos específicos de *Tnt1* en raíz de planta adulta de tabaco a niveles bajos pero no en hoja (Pouteau et al., 1991), *BARE1* en hoja y callos de cebada (Suoniemi et al., 1996), *Cinful* en hojas de maíz (SanMiguel y Bennetzen, 1998) u *Opie1* en raíz de maíz (SanMiguel et al., 1996), entre otros.

## **Introducción**

---

También, se ha observado que los elementos *Tto1* de tabaco (Hirochika, 1993) y *Tos17* de arroz (Hirochika et al., 1996) tienen además un número de copias incrementado en cultivos celulares donde el nivel de RNA del elemento es también elevado, confirmando que existe una correlación entre la transcripción y la transposición en estos elementos.

En cambio, es posible observar un número elevado de transcritos de *BARE1* en hoja de cebada mientras que la transposición del elemento no ha podido ser demostrada en estas condiciones (Kumar y Bennetzen, 1999).

Asimismo, aunque la transcripción de *Tnt1A* no es detectable en cultivos de suspensiones celulares, sí ha podido detectarse un incremento en el número de copias del elemento en esta situación (Hirochika, 1993).

Por tanto, aunque la transcripción sea un requisito necesario para la transposición, y en algunos casos como *Tto1* y *Tos17* la movilización sí está regulada a nivel transcripcional, vemos que para la movilización de los RTN-LTR son necesarios otros pasos que han de estar igualmente regulados.

- **Activación de la transcripción en condiciones de estrés biótico y abiótico**

Muchos de los RTN-LTR de plantas mejor caracterizados se transcriben activamente en respuesta a diferentes factores de estrés tanto biótico como abiótico. Muchos se expresan durante el proceso de aislamiento de protoplastos o en cultivos celulares *in vitro* (Hirochika, 1993; Pouteau et al., 1991), indicando que quizás la transposición contribuye significativamente a la variación somaclonal que aparece en plantas regeneradas a partir de cultivos celulares (Grandbastien, 1998).

Se conoce que *Tnt1A* de tabaco activa su expresión en respuesta a herida, en hoja tratada con la criptogeina (elicitor de hongos) y por la adición de metiljasmonato, mientras que *Tnt1B* se transcribe sólo en cultivos celulares y *Tnt1C* activa su expresión en hojas tratadas con ácido salicílico y la hormona 2,4-D (Beguiristain et al., 2001). También se observa activación de la transcripción de *Tnt1* en respuesta a agentes microbianos (Pouteau et al., 1994). De la misma manera, *Tto1* de tabaco es inactivo en hoja pero se induce en respuesta a herida y en cultivos celulares (Hirochika, 1997; Takeda et al., 1999) mientras que *Tto5* se activa en respuesta a ácido salicílico (Takeda et al., 1998). Ambos elementos (*Tto1* y *Tto5*) se activan también cuando el tabaco es infectado con el virus del mosaico del tabaco (TMV) (Hirochika, 1995). El elemento *Tos17* se induce en cultivos celulares de arroz (Hirochika, 1997).

<b>Elemento</b>	<b>Actividad transcripcional en...</b>	<b>Referencia</b>
<i>Tnt1</i> (tabaco)	Raíz planta adulta, respuesta a herida, en cultivos celulares, protoplastos, respuesta a patógenos microbianos	Pouteau et al., 1991 Beguiristain et al., 2001 Pouteau et al., 1994
<i>BARE-1</i> (cebada)	Hoja y callos,	Suoniemi et al., 1996
<i>Cinful</i> (maíz)	Hojas	SanMiguel y Bennetzen, 1998
<i>Opie1</i> (maíz)	Raíz	SanMiguel et al., 1996
<i>Tto1</i> (tabaco)	Cultivos celulares, respuesta a herida, protoplastos, infección por virus	Hirochika, 1993 Takeda et al., 1999 Hirochika, 1995
<i>Tos17</i> (arroz)	Protoplastos, Cultivos celulares	Hirochika et al., 1996
<i>OARE-1</i> (trigo)	Respuesta a herida, luz UV, ácido salicílico, ácido jasmónico, infección por hongos	Kimura et al., 2001

**Tabla V.** Tabla donde se muestran algunos de los RTN-LTR caracterizados que muestran actividad transcripcional

- **Secuencias reguladoras de la LTR del RTN**

Los RTN-LTR poseen en las LTR secuencias reguladoras en *cis* que controlan sus patrones de expresión en el genoma huésped.

La estructura de los elementos reguladores de la transcripción en secuencias repetidas en tándem (que son inestables durante la retrotranscripción) favorecen la aparición de copias mutadas o defectivas que compiten con las copias activas por los factores necesarios para la expresión, de esta manera los RTN-LTR siguen una estrategia similar a la de los virus para mantener su población en niveles que no afecten a la viabilidad del genoma huésped (Casacuberta et al., 1995).

La estructura y función de los promotores de los distintos RTN-LTR mejor caracterizados se ha estudiado mediante transformación de plantas con la construcción LTR-GUS, donde el gen informador se coloca bajo el control del promotor de la LTR del elemento concreto. Por ejemplo, ensayos de expresión tanto estable como transitoria de la proteína fusión LTR-GUS, conteniendo fragmentos de distintas longitudes de la LTR, han permitido definir para *Tnt1A* (Pouteau et al., 1991, 1994) y *Tto1* (Takeda et al., 1998, 1999) que la mayor parte de las secuencias que controlan la expresión del RTN-LTR están situadas en el dominio U3 de la LTR5'. Además, las diferentes subfamilias de *Tnt1* (*Tnt1A*, *Tnt1B* y *Tnt1C*) que se expresan en tabaco bajo diferentes condiciones de estrés tal y como habíamos visto en el apartado anterior, difieren en sus dominios U3 los cuales probablemente contienen las secuencias necesarias para el control de su expresión (Beguiristain et al., 2001). Parece ser que la elevada variabilidad de las secuencias reguladoras de *Tnt1* ha permitido al RTN evolucionar y adquirir diferentes

## Introducción

---

nuevos mecanismos de regulación permitiendo coexistir y persistir en el genoma de *Nicotiana* (Verhenettes et al., 1998). También, se ha descrito el mismo caso para las LTR de *BARE* (Vicient et al., 2005).

Los promotores de *Tnt1* y *Tto1* que se activan inducidos por diferentes condiciones de estrés, muestran homología de secuencia con promotores de genes de defensa en plantas y por tanto, se unen a los mismos factores de transcripción inducidos en respuesta a defensa (Mhiri et al., 1997; Grandbastien et al., 1997; Takeda et al., 1999). También el promotor de *BARE1* comparte homologías con componentes de otros promotores de genes de defensa (Suoniemi et al., 1996).

Este tipo de estudios demuestran también que construcciones LTR-GUS con promotores de *Tnt1* o *Tto1* (ambos RTN-LTR de tabaco) son activos en sistemas heterólogos como *Arabidopsis*, tomate o arroz, y que los patrones de expresión del RTN-LTR que se observan en las plantas transgénicas son similares a los vistos en tabaco, su huésped natural. Además, las activaciones por estrés de *Tnt1* y *Tto1* se mantienen en estos sistemas heterólogos. Por tanto, los factores involucrados en la retrotranscripción están conservados en las plantas superiores. Pero, mientras que en el sistema homólogo (tabaco) la activación del promotor de *Tnt1* se da únicamente en raíz, en plantas transgénicas de *Arabidopsis* y tomate hay expresión también en flor (Mhiri et al., 1997). Estos resultados sugieren que en la regulación de la expresión de *Tnt1* inducida por estrés, además de las secuencias conservadas que actúan en *cis* en la LTR, participan también reguladores específicos del huésped (Kumar y Bennetzen, 1999).

Diferentes fracciones de la región U3 de la LTR de *Tnt1* pueden estimular la actividad de un promotor mínimo heterólogo (CaMS-35S mínimo) adyacente en protoplastos de tabaco (Casacuberta y Grandbastien, 1993).

En la siguiente tabla se muestran algunos de los RTN-LTR caracterizados en los que se ha estudiado la funcionalidad del promotor de la LTR mediante experimentos con el gen informador *gus* (**Tabla VI**).

Fusión a GUS	Condiciones activación	Sistema	Referencia
Partes de LTR <i>Tnt1</i>	Protoplastos	Homólogo: tabaco	Casacuberta y Grandbastien, 1993
LTR de <i>Tnt1</i>	Hoja herida (mecánica), frío, ácido salicílico y especies reactivas oxígeno, CuCl <sub>2</sub>	Heterólogo: tomate y arabidopsis	Mhiri et al., 1997
U3 de LTR de <i>Tnt1</i>	Hoja tratada con salicilato de sodio, CuCl <sub>2</sub> , Cryptogeina (elicitors microbianos)	Homólogo: tabaco	Verhettes et al., 1997
LTR de <i>Tnt1A</i> , <i>Tnt1B</i> , <i>Tnt1C</i>	Cryptogeina, MJA, A. salicílico, Auxinas	Homólogo: tabaco	Beguiristain et al., 2001
LTR de <i>Tnt1</i>	Cryptogeina, Capsiceina (elicitors microbianos)	Homólogo: tabaco	Pouteau et al., 1994
LTR y LTR* de <i>Tto1</i> (LTR* delección de 8 pb)	Hoja herida y MJA	Homólogo: tabaco	Takeda et al., 1998
LTR de <i>Panzee</i>	Especificidad tejido: activación alta en raíz, baja en órganos florales y muy baja en hoja	Heterólogo: tabaco	Lall et al., 2002

**Tabla VI. Promotores funcionales de RTN-LTR de plantas.** La tabla incluye aquellos RTN-LTR de plantas cuya funcionalidad del promotor de la LTR haya sido probada en la bibliografía mediante experimentos de transformación de plantas con la construcción LTR-GUS.

- **Metilación del DNA**

La metilación del DNA se ha asociado a la región heterocromática e inactividad genética en muchos eucariotas superiores, incluyendo las plantas. El análisis del DNA repetitivo (mayoritariamente RTN) presente en el genoma de plantas muestra citosinas metiladas en las secuencias 5'-CG-3' y 5'-CNG-3'. Esta metilación se asocia tanto a inactividad transcripcional del elemento como a una elevada tasa de mutaciones puntuales por transición de C a T, que hace que los elementos degeneren en formas no funcionales más rápidamente (SanMiguel et al., 1998).

Se ha demostrado que en el caso de *Tto1* la metilación del DNA es responsable de su silenciamiento en Arabidopsis (Hirochika et al., 2000). En este trabajo se observa que el número de copias de *Tto1* se incrementa en líneas transgénicas de Arabidopsis en cultivos celulares durante el proceso de transformación, y que seguidamente el elemento se silencia. Este silenciamiento está correlacionado con una reducción de los niveles de transcritos de *Tto1* y con la metilación de las copias silenciadas. En cambio, cuando la línea silenciada está en un fondo mutante deficiente en metilación (*ddm1*) entonces *Tto1* pasa a estar hipometilado y vuelve a activarse su transcripción.

### **B) Regulación a nivel post-transcripcional**

Los RTN pueden también inactivar otros RTN pre-existentes insertándose dentro de ellos, en una estructura anidada (*nested*) muy frecuente en las regiones intergenéticas del genoma de maíz (SanMiguel et al., 1996).

## **Introducción**

---

Se ha visto que la recombinación homóloga entre LTR, para generar solo-LTR puede ser otro mecanismo de control para reducir el número de copias de determinados RTN-LTR (Vicient et al., 1999). También en levaduras, se ha visto que la recombinación entre LTR evita una invasión excesiva de elementos *Ty* permitiendo el mantenimiento en el genoma huésped de una población pequeña de elementos todavía activos (Jordan and McDonald, 1999).

El silenciamiento post-transcripcional (PTGS, *Post- Transcriptional Gene Silencing*) es un mecanismo general de defensa antiviral en plantas y consiste en una degradación de RNA específica de secuencia (Casacuberta et al., 2003).

### **4.4.2 Evidencias de la posible movilización de RTN-LTR**

Aunque el número de RTN-LTR descritos en plantas es muy elevado (Kumar y Bennetzen, 1999) se han obtenido evidencias de una movilización reciente en muy pocos de ellos.

Además de la activación de la transcripción para varios RTN-LTR descrita anteriormente, también se han detectado proteínas de la GAG e INT de *BARE1* así como partículas similares a virus (VLP) de este RTN-LTR en cebada (Jääskeläinen et al., 1999) y en otros cereales (Vicient et al., 2001) sugiriendo que algunas familias de RTN no sólo son activos transcripcionalmente sino que también se traducen e integran.

El RTN-LTR *BARE1* es altamente polimórfico entre diferentes cultivares de *H. vulgare*, indicando actividad transposicional del elemento en estas especies en un pasado reciente (Kalendar et al., 1999). Además, también presenta polimorfismos de inserción en otras especies de la familia Triticeae lo cual sugiere que la actividad transposicional del elemento ha persistido durante millones de años en múltiples especies (Gribbon et al., 1999). Por otra parte, se ha observado que cambios microclimáticos marcados pueden modificar el número de copias de *BARE1* en cebada (Kalendar et al., 2000).

Se han obtenido polimorfismos de inserción entre especies próximas (Pearce et al., 2000) y entre variedades y poblaciones (Kalendar et al., 2000; Vicient et al., 1999) para distintos elementos. Hay RTN-LTR transposicionalmente activos en un pasado reciente en *Zea spp.* (Purugganan et al., 1994; White et al., 1994), en *Pisum spp.* (Ellis et al., 1998; Pearce et al., 2000) o *Avena spp.* (Yu y Wise, 2000). También se ha demostrado la presencia de nuevas inserciones en el caso de *Tnt1* (Melayah et al., 2001), *Tto1* (Hirochika, 1993) y *Tos17* (Hirochika et al., 1997) en respuesta a determinadas condiciones de estrés.

## 4.5 USO DE LOS RTN-LTR COMO HERRAMIENTAS MOLECULARES

### Taguado de genes (*Gene Tagging*)

Entendemos como genética reversa los métodos para asignar funciones biológicas a genes mediante la generación y cribage de gran número de mutantes, de los que se seleccionan fenotipos de interés y seguidamente se busca el gen mutado que los produce. Existen varias formas de generar mutantes, ya sea por mutagénesis química o física (radiaciones) o mediante la inserción de transposones (*transposon tagging*) o T-DNA.

En el *transposon tagging* tradicionalmente se utilizan elementos transponibles de clase II como *Ac/Ds*, *En/Spm* o *Mutator*, que en determinadas situaciones pueden escindir y revertir la mutación (Greco et al., 2001).

Los RTN de plantas presentan una serie de características que los convierten en herramientas muy útiles para el taguado de genes, por encima de otros elementos de inserción tradicionales: (i) la mutación que generan es estable pues presentan un modo de transposición replicativo; (ii) no presentan dianas fijas de inserción generando movilizaciones al azar, facilitando así la generación de colecciones de mutagénesis saturadas; (iii) su transposición puede estar regulada por factores bióticos o abióticos; (iv) son ubicuos en plantas y (v) algunos pueden llegar a ser altamente mutagénicos al transponer en regiones génicas.

En arroz, el RTN-LTR *Tos17* está silenciado en condiciones normales pero se activa en condiciones de cultivo celular. En plantas regeneradas a partir del cultivo se pueden encontrar entre 5-30 copias nuevas del elemento (Hirochika et al., 1997). *Tos17* transpone de forma preferente en regiones ricas en genes posibilitando su utilización como herramienta mutagénica (Miyao et al., 2003). *Tos17* se vuelve inactivo en las plantas regeneradas y las mutaciones producidas por sus inserciones se heredan de forma estable en las siguientes generaciones (Hirochika et al., 2004).

De la misma manera, el RTN-LTR de tabaco *Tto1* se ha demostrado que es activo en cultivos tisulares también de *Arabidopsis* dirigido por su propio promotor, confirmando que es un sistema eficiente para la mutagénesis en esta especie (Okamoto y Hirochika, 2000).

### Marcadores moleculares

Los RTN-LTR son ubicuos, se presentan en un número de copias elevado y en poblaciones extremadamente heterogéneas, presentan una distribución dispersa por los cromosomas y muestran polimorfismos de inserción tanto intra como inter-especie en plantas. Estas características intrínsecas de los RTN-LTR los convierte en buenos marcadores moleculares tal y como se desarrolla ampliamente en el siguiente capítulo.

### 5 LOS MARCADORES

Los marcadores tienen gran importancia en estudios evolutivos y de variabilidad.

Un marcador es un carácter de determinación sencilla empleado en los programas de mejora cuando el carácter de interés que se quiere seleccionar posee una determinación complicada.

Un marcador ideal ha de ser:

- Polimórfico, variable entre los diferentes individuos analizados.
- Reproducible experimentalmente.
- Codominante, que permita diferenciar entre individuos homocigotos y heterocigotos.
- Discriminante, permitiendo detectar diferencias entre individuos estrechamente relacionados.
- Ampliamente distribuido por el genoma.
- No debe estar influenciado por el ambiente.
- Ha de comportarse como un gen neutro, sin afectar al desarrollo de la planta.
- Su detección y análisis tienen que ser simples y de bajo coste.

Hay diferentes tipos de marcadores:

#### A. Marcadores de caracteres morfológicos

Los primeros marcadores utilizados fueron los que afectaban a la morfología de las plantas. Pero, los rasgos fenotípicos además de ser escasos presentan una serie de limitaciones ya que tienen niveles de polimorfismo bajos y también suelen presentar dominancia, pleiotropía (un gen o alelo determina diferentes fenotipos), epistasia (un carácter determinado por más de un locus) y además, algunos son deletéreos.

#### B. Marcadores bioquímicos

Los **Isozimas** fueron los primeros marcadores bioquímicos que se emplearon en mejora genética, aunque actualmente están más en desuso. Son enzimas que catalizan una misma reacción bioquímica pero presentan algunas diferencias en la cadena de aminoácidos que pueden detectarse mediante tinción con sustratos específicos. Son marcadores generalmente codominantes y la metodología de detección es rápida y sencilla. Como desventaja, pueden estar influenciados por cambios ambientales y durante el desarrollo.

### C. Marcadores moleculares

Un marcador molecular es esencialmente una secuencia nucleotídica correspondiente a una localización física particular en el genoma. Esta secuencia ha de ser suficientemente polimórfica para que su patrón de herencia sea fácilmente seguido. Los marcadores moleculares se encuentran en cualquier especie, son fenotípicamente neutros, pueden detectarse en estadios tempranos del desarrollo y en cualquier tejido, y no presentan interacciones epistáticas.

#### 5.1 MARCADORES MOLECULARES

En la actualidad hay un gran número de tipos de marcadores moleculares con unas características específicas que los hacen adecuados para unas determinadas aplicaciones.

- Los **RFLP** (*Restriction Fragment Length Polimorphism*) fueron la primera técnica desarrollada basada en marcadores de DNA hacia el 1970. En ella se digiere el DNA genómico y tras *Southern-blotting* se hibrida con un fragmento de DNA conocido (Helentjaris et al., 1985). Suelen segregar como marcadores codominantes, son altamente reproducibles y su número es "ilimitado". Su principal inconveniente es su elevado coste económico y de tiempo, además son poco polimórficos (**Figura 10A**).
- Los **RAPD** (*Randomly Amplified Polimorphic DNA*) es una técnica rápida basada en la PCR que no requiere conocimiento previo de la secuencia. Utiliza un cebador de 10 nucleótidos de secuencia aleatoria y la metodología de detección es rápida y sencilla (Williams et al., 1990). Sus principales inconvenientes es que presentan dominancia, tienen una baja información polimórfica, poca correlación con la información derivada de otros marcadores y baja reproductibilidad, dada las bajas temperaturas de anillamiento utilizadas en las reacciones de PCR (**Figura 10C**).
- Los **SCAR** (*Sequence Characterized Amplified Regions*) surgieron para solucionar algunos de los problemas que presentan los RAPD. Consiste en clonar y secuenciar los fragmentos polimórficos de los RAPD para diseñar nuevos cebadores más largos y específicos. Así, se eliminan problemas de reproductibilidad y se consigue en ocasiones transformar el polimorfismo en codominante.
- En los años 90s se desarrollaron en plantas métodos para detectar variabilidad basados en **microsatélites** o **SSR** (*Simple Sequence Repeats*). Se trata de secuencias de 2-6 pb repetidas en tándem, en un número que va de dos hasta cientos de veces, y que se presentan distribuidos aleatoriamente en el genoma de los eucariotas. Las repeticiones del tipo (TA/AT)<sub>n</sub> o (GA/CT)<sub>n</sub> son las más abundantes en plantas. Estos marcadores tienen un nivel de polimorfismo muy elevado que se produce por el cambio

## Introducción

---

en el número de repeticiones. La mayoría se comportan como codominantes. Son muy útiles en la construcción de mapas genéticos saturados y altamente informativos. La metodología de detección se basa en la amplificación de fragmentos de DNA utilizando cebadores diseñados a ambos lados del microsatélite y su posterior resolución en un gel de acrilamida. El método es relativamente simple, reproducible y puede ser automatizado (Tautz et al., 1989) (**Figura 10D**).

- Los **ISSR** (*Inter Simple Sequence Repeat*) se basan igualmente en secuencias repetidas en tándem, pero en este caso se amplifica la zona comprendida entre dos microsatélites utilizando cebadores homólogos a los microsatélites a los que se le añaden dos bases aleatoriamente (Gupta et al., 1994). Son marcadores menos polimórficos que los SSR pero con un gran potencial para estudios de poblaciones naturales (Wolfe AD y Liston A., 1998).

- A mediados de los años 90, se introdujo el método de **AFLP** (*Amplified Fragment Length Polymorphism*), que combina el RFLP y la PCR ya que, aunque el método deriva de la PCR el polimorfismo que se detecta deriva de las variaciones que se dan en los sitios de restricción (Vos et al., 1995). La metodología de detección consiste en una digestión del DNA genómico con dos enzimas de restricción, después se realiza una ligación con unos adaptadores que se unen a los fragmentos de restricción, se amplifican estos fragmentos con cebadores homólogos a la secuencia de los adaptadores y se resuelve en un gel de poliácridamida. Modificando los nucleótidos de selección se amplifican grupos de fragmentos diferentes por lo que el empleo de la técnica AFLP permite la identificación de un número "ilimitado" de marcadores (Mueller UG y Wolfenbarger L., 1999). Los AFLP generan muchas bandas por reacción, además no requieren conocimiento previo de la secuencia de la especie y son muy abundantes. Como desventajas, destacar que son marcadores dominantes y que fragmentos generados con el mismo tamaño no tienen porque ser homólogos (**Figura 10B**).

- Los **SNP** (*Single Nucleotide Polymorphism*) detectan cambios de una sola base en la secuencia de DNA que se dan con una frecuencia alélica de como mínimo el 1% (Marth et al., 1999; Martin et al., 2002). Es el tipo más común para diferenciar secuencias entre alelos. En el genoma humano hay cerca de un millón de SNP, que son altamente polimórficos y pueden ser usados como marcadores genéticos simples. Existen diferentes métodos para generar SNPs como la secuenciación directa de un fragmento aleatorio de DNA, de una banda AFLP, RAPD o EST (*Expressed Sequence Tag*) o de genes conocidos (Rafalski et al., 2002).



## Introducción

Marcadores	Técnica					Requerimientos			
	Basada en PCR	Nº marcadores	Polimorfismo	Co-Dominancia	Replicabilidad	Cantidad / Calidad de DNA	Conocimiento secuencia DNA	Detección radiactiva	Sistema de gel
<b>Isozimas</b>	No	<90	Bajo	Sí	Alta	-	No	No	No
<b>RFLP</b>	No	Ilimitados	Medio	Sí	Alta	Alta/Alta	No	Sí/No	Agarosa
<b>RAPD</b>	Sí	Ilimitados	Medio	No	Variable	Baja/Baja	No	No	Agarosa
<b>SCAR</b>	Sí	Ilimitados	Medio	Sí /No	Alta	Baja/Alta	Sí	Sí/No	Agarosa
<b>SSR</b>	Sí	Ilimitados	Muy alto	Sí	Alta	Baja/Media	Sí	No	Acrilamida-Agarosa
<b>ISSR</b>	Sí	Ilimitados	Medio	No	Alta	Baja/Media	Sí/No	No	Acrilamida-Agarosa
<b>AFLP</b>	Sí	Ilimitados	Alto	No	Alta	Baja/Alta	No	Sí/No	Acrilamida
<b>SNP</b>	Sí	Ilimitados	Muy alto	Sí	Alta	Baja/Alta	Sí	No	Secuenciar
<b>EST</b>	Sí	Ilimitados	Medio	Sí	Alta	Baja/Alta	Sí	Sí/No	Acrilamida-Agarosa

**Tabla VII. Tabla comparativa de las principales técnicas de obtención de marcadores comentadas.** En la tabla consideramos los diferentes criterios de forma general pero, hay que tener en cuenta que estas técnicas son muy dependientes de la especie que se está analizando.

Una técnica de marcadores moleculares ideal ha de explotar los cambios físicos que se dan en el genoma (mutaciones puntuales/inserciones-delecciones/reordenamientos) para visualizar la diversidad genética. El loci analizado por el método ha de estar disperso por el genoma en una elevada frecuencia, permitiendo generar mapas de recombinación saturados y bien distribuidos. El método ha de ser universal para que sea fácilmente aplicado en distintas especies. El patrón de bandas generado tiene que ser reproducible. Los RTN-LTR cumplen todos estos requerimientos y por ello están siendo desarrollados como marcadores moleculares (Kumar, A. y Hirochika, H., 2001; Schulman, Flavell y Ellis, 2004).

### 5.2 LOS RTN-LTR COMO MARCADORES MOLECULARES

Los RTN-LTR se encuentran de forma ubicua y abundante en el genoma de las plantas. Se ha demostrado que se distribuyen de manera dispersa y aleatoria por todo el genoma, aunque hay zonas de mayor inserción tal y como se ha comentado en apartados anteriores. Además, el tipo de cambio que generan (inserciones) es permanente, generando marcadores de localización temporal.

Su utilización como marcadores moleculares es altamente informativa ya que cada elemento tiene su propia historia transposicional. Así, aquellos elementos activos en un pasado reciente presentarán poblaciones altamente heterogéneas y polimórficas dentro de una especie, siendo muy útiles en estudios de diversidad genética intraespecífica o varietal (Ellis et al., 1998; Waugh et al. 1997). En cambio, aquellos elementos activos hace millones de años serán más útiles en la determinación de relaciones interespecíficas o entre géneros (SanMiguel et al., 1998).

Existen una serie de técnicas que aprovechan las características intrínsecas de los RTN-LTR para desarrollarlos como marcadores moleculares altamente informativos:

- Los **SSAP** (*Sequence Specific Amplified Polymorphism*) son muy similares a los AFLP, pero como cebadores se utiliza uno complementario a la secuencia del adaptador del enzima de restricción y otro cebador homólogo a una secuencia específica, en este caso del RTN. Normalmente se generan menos fragmentos que con los AFLP pero más polimórficos (Waugh et al., 1997). Esta aproximación se describe más detalladamente a continuación.
- Los **IRAP** (*Inter Retrotransposon Amplified Polymorphism*) son marcadores que se basan en la amplificación de la zona que se encuentra entre dos RTN situados próximos y en orientaciones opuestas en el genoma. Se utilizan cebadores específicos complementarios a una secuencia del RTN. Genera gran cantidad de bandas polimórficas, aunque la mayoría son dominantes (Kalendar et al., 1999).
- Los **REMAP** (*Retrotransposon-Microsatellite Amplified Polymorphism*) detectan polimorfismos entre la zona amplificada comprendida entre un RTN y un microsatélite (SSR) (Kalendar et al., 1999). Existen estudios que confirman que las secuencias de RTN están íntimamente asociadas con microsatélites (Ramsay et al., 1999).
- Los **RBIP** (*Retrotransposon-Based Insertion Polymorphism*) utilizan como cebadores una secuencia del RTN y una secuencia del DNA flanqueante a este elemento. Es un método codominante que detecta polimorfismos de inserción (presencia/ausencia del RTN) (Flavell et al., 1998).
- Los **RVIP** (*Retrotransposon Internal Variation Polymorphism*) utilizan como cebadores secuencias internas a los RTN, y sirven para revelar variaciones en la estructura interna de las diferentes copias de un elemento o familia (Vershinin et al., 1999).

## Introducción

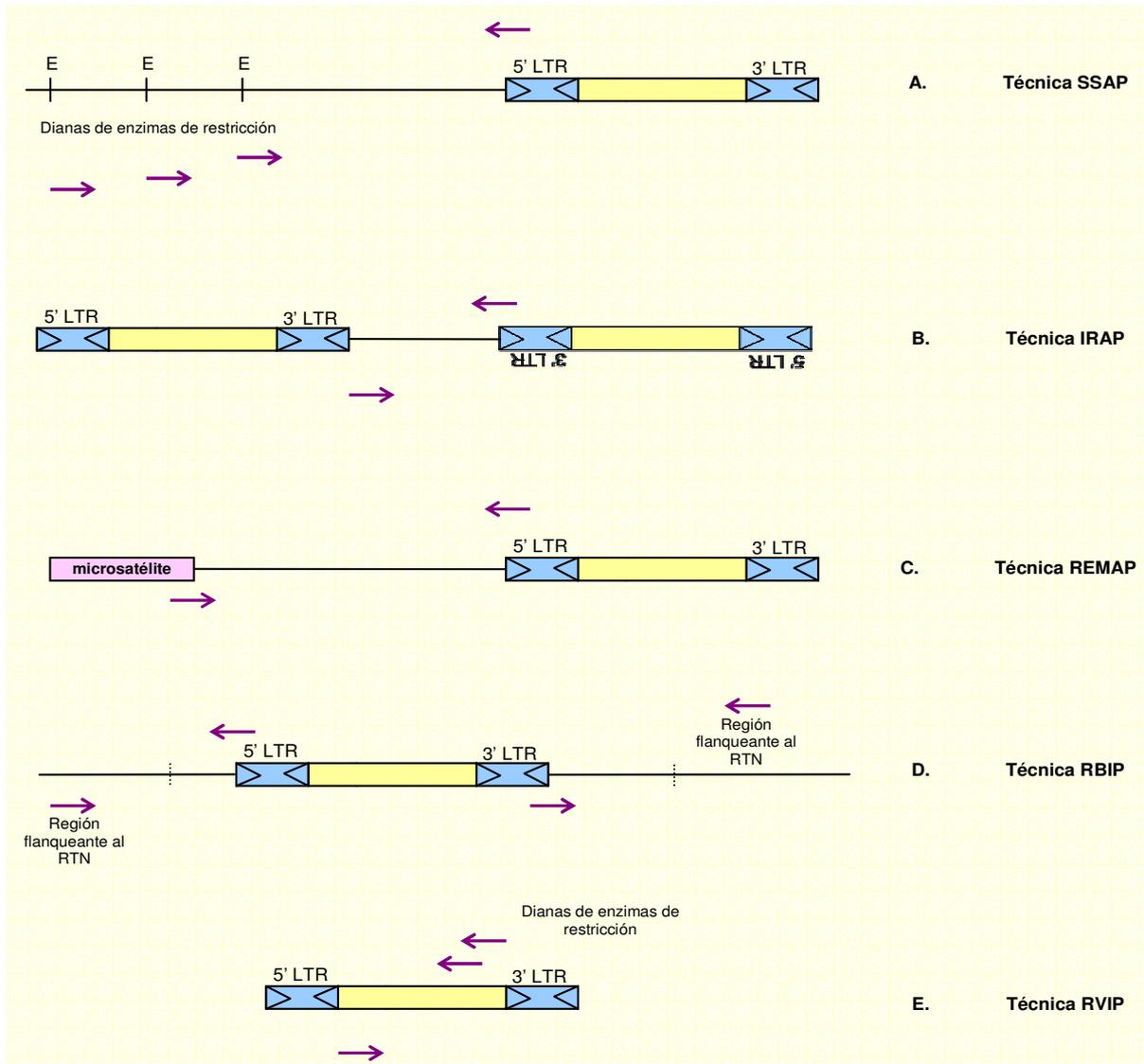


Figura 11. Esquema de las estrategias utilizadas en las técnicas de marcadores moleculares basados en RTN-LTR.

- (A) SSAP. *Sequence Specific Amplification Polymorphism.*
- (B) IRAP. *Inverse Retrotransposon Amplified Polymorphism.*
- (C) REMAP. *Retrotransposon-Microsatellite Amplified Polymorphism.*
- (D) RBIP. *Retrotransposon-Based Insertion Polymorphism.*
- (E) RVIP. *Retrotransposon Internal Variation Polymorphisms*

### 5.2.1 **Marcadores moleculares SSAP**

Tal y como se ha comentado anteriormente, los SSAP (*Sequence Specific Amplified Polymorphism*) son marcadores muy similares a los AFLP tanto en metodología experimental como por el tipo de fragmentos amplificados que generan. Cada banda amplificada en un gel SSAP representa los fragmentos obtenidos por PCR utilizando un cebador diseñado a partir de una secuencia del RTN-LTR y otro cebador complementario a un sitio de restricción (Waugh et al., 1997). Generalmente se obtienen menos fragmentos que con los AFLP, pero más polimórficos. Típicamente se utiliza como secuencia específica del RTN-LTR el final conservado de la LTR (Waugh et al., 1997). En este sentido, se han desarrollado técnicas que permiten, sin conocimiento previo de secuencia de RTN-LTR, el aislamiento rápido de secuencias pertenecientes a la LTR de elementos tipo *copia* (Pearce et al., 1999), lo cual posibilita la aplicación de la técnica SSAP a cualquier especie vegetal aunque no se tenga conocimiento de ningún RTN-LTR aislado. Asimismo, también pueden utilizarse para generar cebadores SSAP otras secuencias del RTN-LTR conservadas en los elementos de una misma familia como son el PBS o PPT (Pearce et al., 2000).

Los marcadores SSAP son dominantes (al igual que los AFLP) y permiten detectar variación de secuencia debida al sitio de inserción del RTN-LTR (presencia/ausencia del elemento), variación de la secuencia genómica que flanquea el elemento y variación de la secuencia del propio elemento (Ellis et al., 1998). Al ser un sistema de marcadores múltiple es necesario aislar y secuenciar cada banda por separado para detectar en cada experimento la causa del polimorfismo SSAP (Waugh et al., 1997).

La metodología de los SSAP se desarrolla ampliamente en el Apartado 12 de Material y Métodos y en el Capítulo B2 de Resultados de este trabajo.

## 5.3 **UTILIDAD DE LOS MARCADORES MOLECULARES**

Los marcadores moleculares presentan múltiples aplicaciones en la **mejora genética de plantas**. Permiten **identificar** el origen y/o naturaleza de los diferentes individuos de una población. Por ejemplo permiten distinguir entre individuos zigóticos o nucelares en plantas que presentan apomixis como los cítricos (Torres et al., 1982), así como identificar híbridos o bien distinguir entre cultivares comerciales. También pueden ser utilizados para estudios de relaciones filogenéticas inter e intraespecíficas. Otra utilidad la encontramos en la caracterización y evaluación de colecciones de germoplasma, así como la detección de variación somaclonal que aparece a veces después de la regeneración a partir de tejidos tisulares.

También permiten **localizar genes** que controlan caracteres de interés agronómico de resistencia a enfermedades o caracteres cuantitativos (QTL, *Quantitative Trait Locus*) que pueden ser usados en mejora genética por **selección asistida por marcadores** y

## **Introducción**

---

en clonación basado en el **mapeo fino de un gen**. Los mapas genéticos son esenciales en la localización de genes y QTLs. La selección asistida por marcadores en la mejora genética permite simplificar mucho los procesos de retrocruzamiento reduciendo el número de generaciones necesarias para la obtención del genotipo deseado; esto es especialmente ventajoso en especies con largos periodos de generación como son, por ejemplo, los cítricos.

Marcadores moleculares basados en RTN-LTR del tipo SSAP han sido aplicados para la construcción de mapas de ligamiento en cebada (Manninen et al., 2000; Waugh et al., 1997), trigo (Queen et al., 2004) y guisante (Ellis et al., 1998). Asimismo, también han sido utilizados para mapear genes de interés agronómico, como son un gen de resistencia al hongo *Pyrenophora teres Drechs. f. teres Smedegen* que causa la lesión *net blotch* en cebada mediante marcadores tipo IRAP y REMAP con los RTN-LTR *BARE1* y *Sukkula* (Manninen et al., 2000).

Los marcadores moleculares también se aplican en **estudios filogenéticos**. Los marcadores moleculares basados en RTN-LTR permiten el estudio de la diversidad genética y las relaciones filogenéticas entre especies. Por ejemplo, los miembros activos de una familia de RTN producen nuevas inserciones en el genoma creando polimorfismos. Mediante estas técnicas, las nuevas inserciones pueden ser detectadas y utilizadas para establecer secuencias temporales de los distintos eventos de inserción, ayudando así a determinar filogenias.

De esta manera, los marcadores moleculares basados en RTN-LTR han sido aplicados a estudios filogenéticos y de biodiversidad en diferentes especies vegetales como son *Hordeum* (cebada) (Kumar et al., 1999; Waugh et al., 1997; Gribbon et al., 1999), *Pisum* (guisante) (Ellis et al., 1998; Pearce et al., 2000), *Zea mays* (maíz) (García-Martínez y Martínez-Izquierdo, 2003), *Triticum* (trigo) (Queen et al., 2004) o *Lycopersicum* (tomate) (Tam et al., 2005).

En algunas de estas especies vegetales los marcadores SSAP parecen estimar mejor las relaciones filogenéticas que los marcadores AFLP; es el caso de cebada (Waugh et al., 1997) y tomate (Tam et al., 2005) entre otros. Asimismo, se ha visto que aproximaciones multi-RTN generan informaciones mucho más detalladas y profundas permitiendo incluso detectar fenómenos de domesticación como en el caso del género *Pisum* (Ellis et al., 1998).

A su vez, métodos como los SSAP y RBIP permiten detectar actividad de RTN en forma de inserciones recientes y por tanto son una herramienta muy útil para evaluar la variabilidad genómica inducida por RTN en plantas que se encuentran bajo condiciones de estrés biótico (como infecciones por patógenos) o abiótico (como estrés ambiental o en cultivos celulares).

## **5.4 APLICACIÓN DE LOS MARCADORES MOLECULARES AL ESTUDIO DE LOS CÍTRICOS**

### **5.4.1 Taxonomía del género *Citrus***

Tal y como se ha comentado anteriormente, la compleja biología reproductiva de los cítricos dificulta enormemente los análisis taxonómicos del género.

Hasta mediados de los años 70s, la clasificación taxonómica del género se basaba únicamente en datos morfológicos y geográficos, y los sistemas más aceptados reconocían desde 10 especies (Swingle, 1943) hasta 147 especies (Tanaka, 1954) dentro del género *Citrus*.

A partir de los años 70s, se comenzaron a aplicar técnicas más modernas para el estudio del género. Los estudios de Barrett y Rhodes en 1976 y Scora en 1975, basados en características morfológicas y bioquímicas, consideran que son sólo tres las especies básicas del género [*C. maxima* (o *C. grandis*), *C. medica* y *C. reticulata*], el resto de especies estarían dentro del subgénero *Citrus* que podría haber derivado por hibridaciones dentro de estas tres especies (hibridaciones intraespecíficas), o bien entre ellas (hibridaciones interespecíficas) o con especies de géneros muy próximos (hibridaciones intergénicas) (Moore et al., 2001). Estudios más recientes basados en marcadores bioquímicos (Herrero et al., 1996) y moleculares (Nicolosi et al., 2000) refuerzan esta hipótesis.

Los marcadores de DNA son ampliamente utilizados para estudios taxonómicos. En *Citrus*, se han utilizado marcadores RAPD para la identificación de cultivares y el mapeo físico (Cai et al., 1994). También se han estudiado las relaciones filogenéticas entre *Citrus* y géneros relacionados combinando marcadores RAPD y RFLP (Federici et al., 1998; Liou et al., 1996). Asimismo se ha determinado mediante análisis de RFLP organulares las relaciones filogenéticas dentro del género *Citrus* (Green et al., 1986) y entre especies y cultivares de *Citrus*, *Poncirus* y *Fortunella* (Yamamoto et al., 1993).

Estudios más recientes con RAPDs y SCARs (Nicolosi et al., 2000), ISSRs (Sankar et al., 2001) y IRAPs (Bretó et al., 2001) están siendo aplicados para aclarar las relaciones filogenéticas dentro del género *Citrus*, aunque los resultados no siempre coinciden.

En la siguiente Figura se muestra por ejemplo un árbol filogenético del género basado en marcadores moleculares del tipo RFLP según Federici y colaboradores (1998).

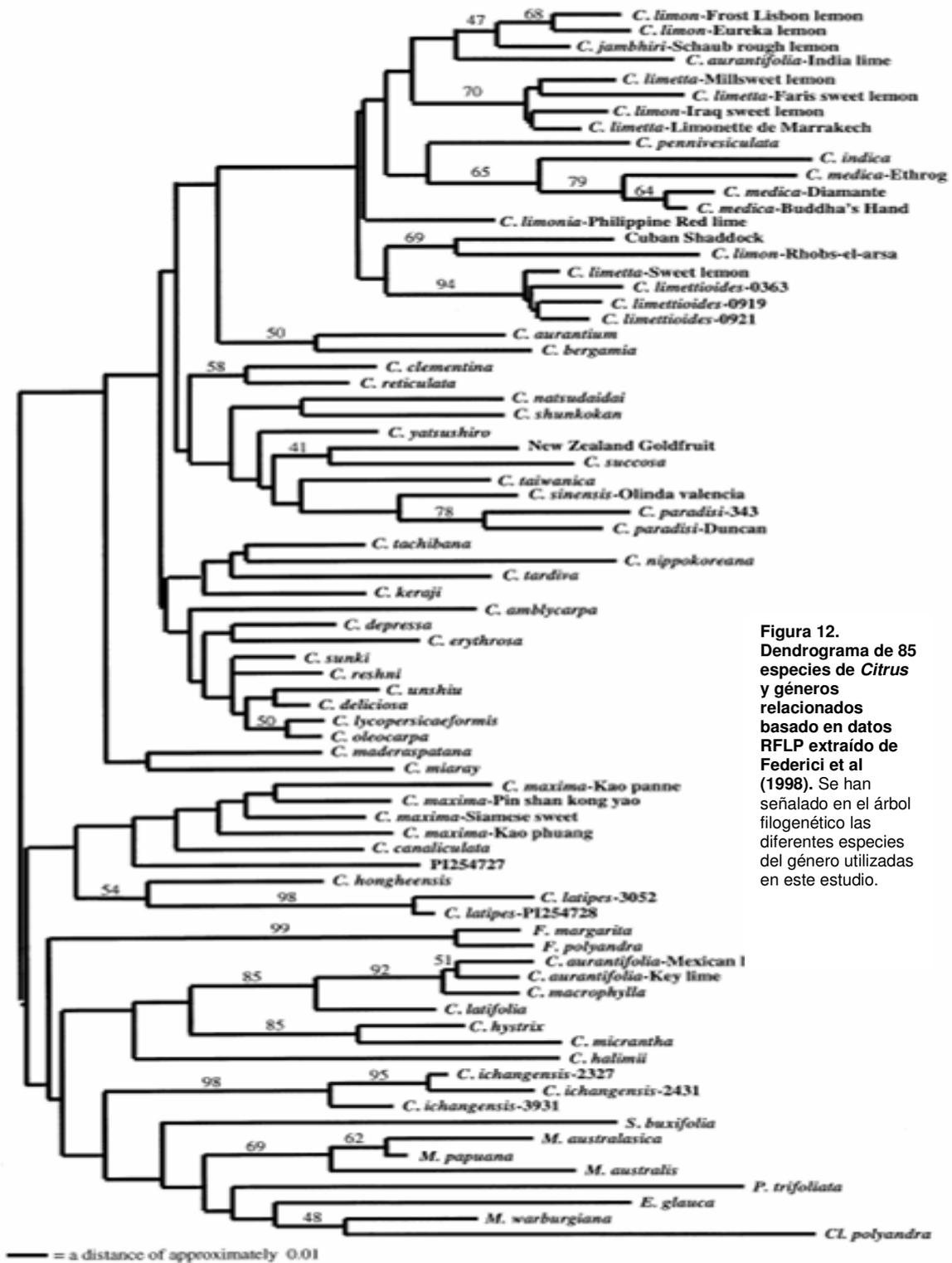


Figura 12. Dendrograma de 85 especies de *Citrus* y géneros relacionados basado en datos RFLP extraído de Federici et al (1998). Se han señalado en el árbol filogenético las diferentes especies del género utilizadas en este estudio.

### 5.4.2 Origen de las variedades de cítricos comerciales

Estrechamente ligado a la taxonomía del grupo, el origen de los diferentes cítricos comerciales también suscita diferentes hipótesis. Tal y como se ha comentado en el apartado anterior, estos parecen haberse originado a partir de diversas hibridaciones entre las tres especies verdaderas de Barret y Rhodes (1976).

Tanto *C. grandis* (pummelo), *C. medica* (cidro) como *C. reticulata* (mandarino), presentan patrones de hibridación únicos al realizar análisis *Southern* del DNA cloroplástico (Nicolosi et al., 2000) lo cual apoya su estatus de especie. Estudios filogenéticos con marcadores SSR confirman también al pummelo, cidro y mandarino como las especies básicas de los cítricos cultivados (Fang et al., 2003).

A continuación se resumen algunos de las hipótesis más aceptadas sobre el origen de algunos de los cítricos comerciales más conocidos:

- **LIMA** Barret y Rhodes (1976) sugieren que probablemente se originaron por un cruce trihíbrido intergénico entre *C. medica*, *C. grandis* y una especie de *Microcitrus*. Análisis posteriores con marcadores de DNA del tipo RAPD o SCAR muestran que las limas (*C. aurantifolia*, *C. limettioides*) presentan marcadores comunes a *C. medica* y a *Microcitrus* (Nicolosi et al., 2000) y otros estudios con RFLP también apoyan el parentesco de *C. medica* y las limas (Federici et al., 1998).
- **LIMONERO** En el caso del limonero (*C. limon*), determinan que es un complejo híbrido similar a las limas pero con una mayor proporción de genes de *C. medica*. Estudios moleculares determinan que los limoneros posiblemente se originaron a partir de *C. medica* y *C. aurantium* (Nicolosi et al., 2000).
- **NARANJO** Estos mismos autores, sugieren que el naranjo amargo (*C. aurantium*) presenta predominantemente el genotipo de *C. reticulata* con pequeñas introgresiones de genes de *C. grandis*. Este dato ha sido igualmente confirmado con análisis de marcadores de tipo RAPD y SCAR (Nicolosi et al., 2000). El naranjo dulce (*C. sinensis*) presenta predominantemente el genotipo de *C. reticulata* con pequeñas introgresiones de genes de *C. grandis*, de la misma manera que pasaba en *C. aurantium* (Pedrosa et al., 2000).
- **POMELO** El caso de *C. paradisi* (pomelo) está bien documentado. Tiene un origen híbrido, probablemente por hibridación natural, entre *C. grandis* y *C. sinensis*, y una posterior selección hacia mutantes sin semillas y con el color rojo del fruto incrementado. Los datos históricos y morfológicos están confirmados nuevamente por datos moleculares mediante aproximaciones RAPD y SCAR (Nicolosi et al., 2000).

## Introducción

- **MANDARINO** El grupo de las mandarinas es el fenotípicamente más heterogéneo del género. Estudios moleculares (análisis del tipo RAPD y SCAR) no permiten identificar el origen del mismo pero confirman su carácter de especie (Nicolosi et al., 2000; Fang et al., 1998).

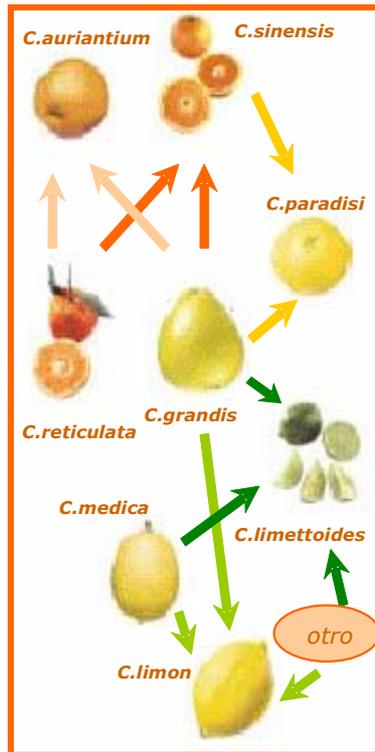


Figura 13. Representación de las relaciones más probables entre los cítricos más comunes, según Moore, 2001.

De todas maneras, nuevos datos son necesarios para clarificar las relaciones entre las diferentes especies del género *Citrus*.

Este trabajo tiene como objetivo principal estudiar y caracterizar retrotransposones con LTR (RTN-LTR) del genoma del naranjo (*Citrus sinensis*) como especie representativa del género *Citrus*. De esta manera se quieren aplicar los conocimientos adquiridos al estudio de los cítricos para entender la elevada variabilidad existente entre ellos, así como desarrollar marcadores moleculares basados en RTN-LTR para evaluar la biodiversidad del género.

En detalle se persiguen los siguientes objetivos:

1. Clonar y secuenciar RTN-LTR de ambos tipos *copia* y *gypsy* del genoma de *Citrus sinensis* (naranjo dulce, variedad "Valencia late") y realizar estudios filogenéticos para ver el espectro de familias de RTN-LTR presentes así como determinar la contribución de éstos al genoma.
2. Evaluar la contribución de los RTN-LTR tipo *copia* al genoma de diferentes especies del género *Citrus*, mediante la descripción de las poblaciones presentes y el número de copias en que se encuentran en los diferentes genomas.
3. Caracterizar algún RTN-LTR tipo *copia* o *gypsy* considerado relevante incluyendo (i) Análisis bioinformático de la secuencia (ii) Determinación del número de copias en que se encuentra (iii) Búsqueda de RTN-LTR activos transcripcionalmente en el genoma de *C. sinensis* (iv) En un elemento activo, estudio de la funcionalidad del promotor de la LTR.
4. Utilizar el RTN-LTR caracterizado en *Citrus sinensis* para poner a punto la técnica SSAP en este sistema y desarrollar marcadores SSAP basados en el RTN.
5. Aplicar la técnica SSAP al estudio filogenético dentro del género *Citrus* y géneros próximos de la familia Rutaceae. Asimismo, aplicar la técnica SSAP para la detección de polimorfismos de inserción de cara a revelar transposiciones recientes de RTN-LTR en *C. sinensis*.

## **Objetivos**

---

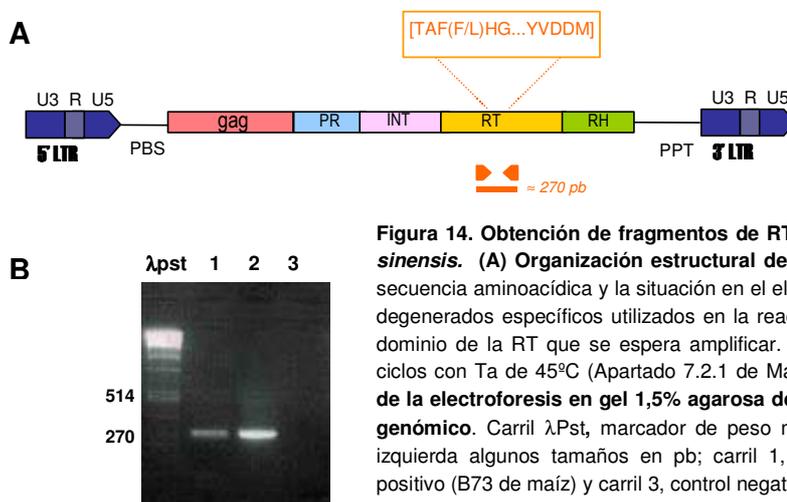
## A. CARACTERIZACIÓN DE RETROTRANSPOSONES DEL GENOMA DE *CITRUS SINENSIS*

### A.1 OBTENCIÓN DE FRAGMENTOS DE RTN-LTR TIPO COPIA DEL GENOMA DE *CITRUS SINENSIS*

#### A.1.1 Estrategia de amplificación selectiva de RTN-LTR tipo copia

La transcriptasa inversa (RT) es entre las proteínas codificadas por retrovirus y retrotransposones (RTN) la que está más conservada (Doolittle et al., 1989). Esta elevada conservación es debida a la presencia en la RT de dominios con actividad enzimática altamente conservados, así como regiones más variables (Xiong and Eickbush, 1990). La mayoría de los elementos tipo *Ty1/copia* identificados en plantas se han aislado gracias a una aproximación experimental mediante PCR, utilizando como cebadores oligonucleótidos degenerados complementarios a dominios conservados de la RT (como son el dominio TAF(L)HG o el dominio YVDDM) (Flavell et al., 1992a; Voytas et al., 1992). En este trabajo, hemos utilizado la misma estrategia para detectar la presencia de retroelementos del grupo *copia* en el genoma de *Citrus sinensis*. Esta técnica permite la identificación preferente de elementos del grupo *copia*, frente a los de otras familias de RTN (Voytas et al., 1992) (**Figura 14A**).

A partir de DNA genómico extraído de hoja joven de *C.sinensis* (con un rendimiento de extracción de 3,75 µg DNA / 100 mg tejido), se realizó la PCR tal y como se describe en Material y Métodos según Voytas et al. (1992). Se visualizó en gel de agarosa 1,5 % el producto amplificado de tamaño esperado (aprox. 270 pb) y se clonó en el vector pGEMT (**Figura 14B**). Se obtuvo la secuencia nucleotídica de 18 clones individuales escogidos al azar en dos reacciones de PCR independientes, mediante el secuenciador automático ABI PRISM 377, Applied Biosystems. La nomenclatura utilizada para denominar los clones conteniendo fragmentos correspondientes a RTN fue CoSi# (retrotransposón *Copia* de *C.Sinensis*).



**Figura 14. Obtención de fragmentos de RTN *copia* en el genoma de *C. sinensis*.** (A) Organización estructural de RTN tipo *copia*. Se indica la secuencia aminoacídica y la situación en el elemento de los oligonucleótidos degenerados específicos utilizados en la reacción de PCR y el tamaño del dominio de la RT que se espera amplificar. La PCR se ha realizado a 30 ciclos con  $T_a$  de 45°C (Apartado 7.2.1 de Material y Métodos). (B) Imagen de la electroforesis en gel 1,5% agarosa de la amplificación sobre DNA genómico. Carril  $\lambda$ Pst, marcador de peso molecular  $\lambda$ Pst, indicando a la izquierda algunos tamaños en pb; carril 1, *C. sinensis*; carril 2, Control positivo (B73 de maíz) y carril 3, control negativo (sin DNA).

## Resultados

### A.1.2 Conservación de los dominios funcionales en los elementos CoSi

Los fragmentos obtenidos por PCR correspondientes a hipotéticas regiones de RT de RTN-LTR tipo *copia*, se clonaron y secuenciaron. Las secuencias resultantes, una vez comprobadas como pertenecientes a RT, se alinearon utilizando el programa ClustalW (Thompson et al., 1994). El análisis de las secuencias revela múltiples variaciones entre ellas, correspondiendo de hecho a 18 elementos diferentes. Para obtener más información de estos datos, la secuencia nucleotídica de los diferentes clones se tradujo a secuencia aminoacídica y estas secuencias deducidas conceptualmente se compararon entre si y con secuencias de otros RTN tipo *copia* ya descritos y presentes en las bases de datos, como son *copiaDM* de *Drosophila melanogaster* (Nº Acc.: P04146) y *Tnt1* de *Nicotiana tabacum* (Nº Acc.: P10978). Este análisis permite conocer qué regiones están más o menos conservadas y, confirma la presencia en todas ellas de una secuencia consenso (SLYGLKQA) característica de secuencias de la RT de RTN de tipo *copia* (Figura 15). De los clones secuenciados, 7 están interrumpidos por codones de parada en la pauta de lectura que corresponde a una RT de RTN-LTR tipo *copia*, pero ninguno presenta inserciones o deleciones, y todos tienen motivos funcionales conservados. Al realizar comparaciones de secuencias con las bases de datos (GenBank) mediante BLASTN (compara la secuencia nucleotídica con bases de datos nucleotídicas) y BLASTX (compara la secuencia nucleotídica traducida a todas las pautas de lectura con bases de datos proteicas) obtenemos en todos los casos similitudes significativas con fragmentos de poliproteínas de RTN tipo *copia* de diferentes especies vegetales incluidas en ellas.

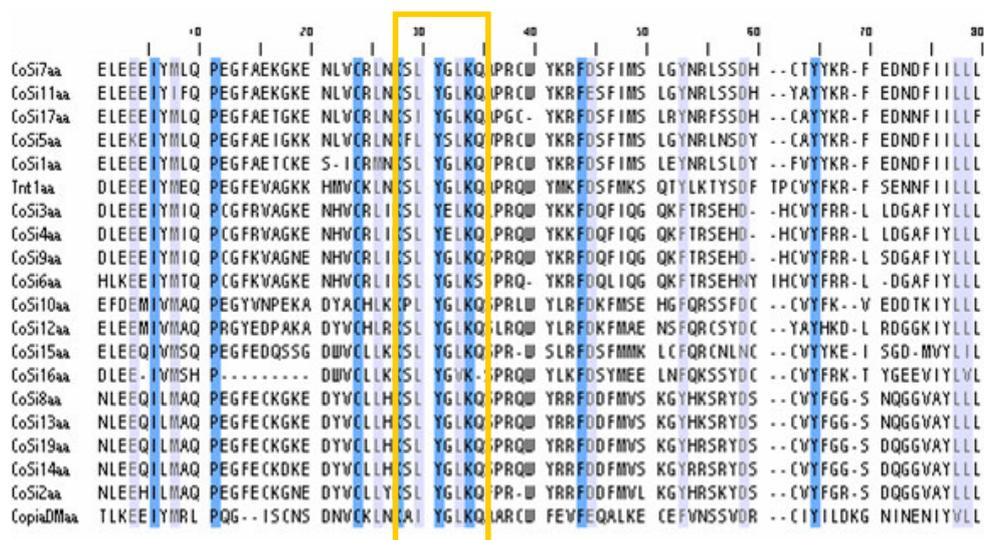


Figura 15. Alineamiento múltiple por ClustalW de las secuencias aminoacídicas inferidas de los clones CoSi# y de zonas similares de RTN tipo *copia* ya descritos como son *Tnt1* de tabaco y *copia* de *D. melanogaster*. El orden relativo de las secuencias está directamente relacionado con el nivel de similitud entre ellas. Los puntos indican ausencia de aminoácido en esa posición. El sombreado oscuro indica identidad de secuencia del 100% y el claro, elevada similitud. Se ha recuadrado el consenso SLYGLKQA típico de secuencias de RT de RTNs de tipo *copia*. CoSi#: clones de *C. sinensis*. Copia (Nº Acc.: P04146) de *Drosophila melanogaster*. Tnt1 (Nº Acc.: P10978) de *Nicotiana tabacum*. No se muestran los dominios conservados típicos de RTN-LTR *copia* (TAFLHG y YVDDM) que flanquean las secuencias porque corresponden a los cebadores degenerados utilizados en la amplificación por PCR y su inclusión no revela las variaciones de secuencia existentes en los productos amplificados sino las secuencias impuestas por los cebadores.

**A.1.3 Descripción de nuevas familias de RTN tipo copia en naranjo**

Para tener una visión del espectro de similitudes entre los fragmentos aislados, se han realizado análisis filogenéticos de las secuencias tanto nucleotídicas como aminoacídicas mediante árboles tipo *Neighbour joining* (NJ) que nos han permitido agruparlas según la divergencia de secuencia entre ellas y entre otros RTN de tipo *copia* ya descritos. En la **Figura 16** se muestra el árbol NJ elaborado a partir del alineamiento de la **Figura 15**. La mayoría de las secuencias caen en diferentes subgrupos, reflejando cierta heterogeneidad entre los fragmentos de secuencias de RT presentes en el genoma de *C. sinensis*, aunque no hay ningún grupo muy divergente respecto a los demás.

Todas las secuencias de *C. sinensis* analizadas en este estudio, forman parte de un mismo supergrupo según la definición de Feng & Doolittle (1987) citada en Gribbon et al. (1999) ya que contiene miembros separados 100 unidades de distancia, lo que corresponde a 60% de identidad nucleotídica o 40% de identidad aminoacídica. La identidad nucleotídica más baja encontrada en comparaciones dos a dos entre secuencias CoSi# es del 60%.

Un análisis más profundo permitió clasificar 14 de los 18 elementos CoSi# obtenidos dentro de 3 familias claras, denominadas CIRE1-CIRE3 (acrónimo de Citrus Retrotransposon). Los elementos son asignados a una misma familia si comparten una identidad nucleotídica  $\geq 85\%$  entre ellos. La identidad nucleotídica media entre los 5 clones aislados de CIRE1 es del 93,5%, para los 5 clones de CIRE2 es del 85,4% y para los 4 clones de CIRE3 del 88,9%. Hemos considerado, que los 4 elementos restantes son miembros de 4 familias diferentes (CIRE4-CIRE7) con un solo representante entre las secuencias obtenidas en este estudio. Se han señalado las diferentes familias en la **Figura 16**, donde vemos que los tres grupos mayoritarios de secuencias (CIRE1-CIRE3) están apoyados por valores de *bootstrap* muy altos (99-100).

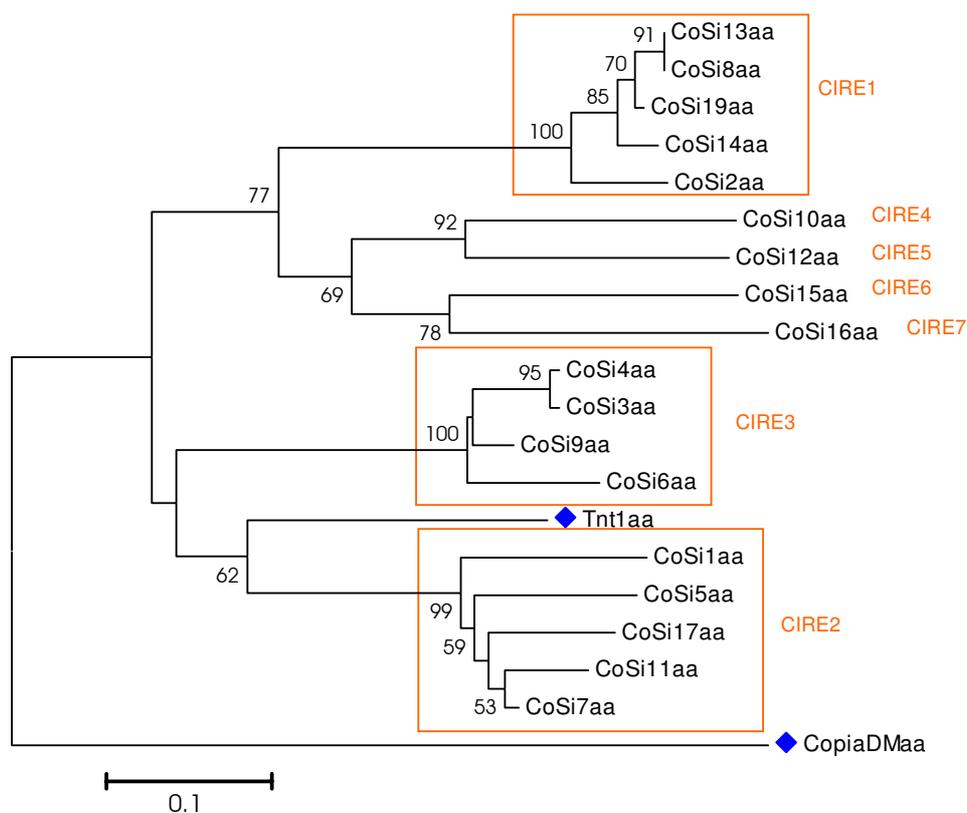
Se observa también que *Tnt1* de tabaco se encuentra bastante próximo a uno de los subgrupos de RTN-LTR CoSi originados, en concreto comparte una identidad media del 69% con los miembros de CIRE2 (**Figura 16**). Se observa que *CopiaDM* se ubica claramente como *outgroup* en el análisis, siendo evidentemente la secuencia más divergente.

Al realizar búsquedas en las bases de datos (GeneBank), se han encontrado secuencias de la RT de RTN-LTR tipo *copia* de *C. sinensis* (NºAcc.: AY654418, AY654419, AY654420, AY654422 y AY654423). Al compararlas mediante BLASTN con los RTN-LTR descritos en este trabajo, se encuentra una secuencia, AY654422 con el 90% de identidad nucleotídica a elementos de CIRE2, indicando que esta familia de RTN-LTR *copia* corresponde probablemente a la misma familia de elementos que AY654422, descrito anteriormente en naranjo. Por el contrario, el resto de familias CIRE no presentan similitudes superiores al 50% con RTN-LTR de *C. sinensis* encontrados en las Bases de datos, pero muestran una elevada similitud con secuencias de RT de RTN de especies no

## Resultados

relacionadas como son AJ000893 de *panzee*, RTN de *Cajanus cajan* (85% de identidad nucleotídica con miembros de CIRE2), AY624347 de *Populus ciliata* (70% con CIRE3), X13777 de *Tnt1*, RTN de *Nicotiana tabacum* (69% con CIRE2), AF232981 de *Amaranthus quitensis* (69% con CIRE1), AF398199 de *Zea mays* (68% con CIRE3) y D12827 de *Tto1*, RTN de *Nicotiana tabacum* (63% con CIRE2), entre otras.

La diversidad nucleotídica ( $\pi$ : diferencias nucleotídicas entre pares de secuencias por posición analizada) en la región de la RT es de 0,34.



**Figura 16. Árbol filogenético basado en el alineamiento de secuencias aminoacídicas de zonas de la RT de RTN-LTR tipo copia obtenidas del genoma de *C. sinensis*.** Árbol filogenético tipo *Neighbour joining* (Saitou y Nei, 1987) generado con el programa MEGA2 (Kumar et al., 2001). Las opciones utilizadas para generar el árbol se describen en Material y Métodos. La divergencia entre las secuencias en unidades de distancia genética viene definida por la longitud de las ramas horizontales. Esta indicado en cada bifurcación los valores de bootstrap  $\geq 50\%$  para 1000 réplicas. Se han recuadrado las diferentes familias en qué quedan divididas las secuencias, el nombre de cada familia está indicado a la derecha de la figura. Secuencias de *Drosophila melanogaster* (copiaDM, N<sup>o</sup> Acc.: P04146) y de *Nicotiana tabacum* (Tnt1, N<sup>o</sup> Acc.: P10978) se han incluido en la comparación (♦).

### **A.1.4 Estimación del número de copias de RTN tipo copia en el genoma de naranjo**

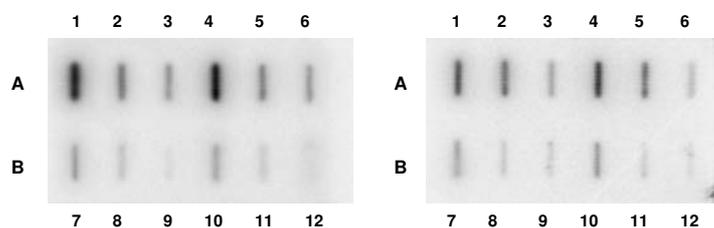
Se ha estimado el número de copias total de RTN tipo *copia* por genoma haploide de *C. sinensis* mediante reconstrucción genómica a partir del análisis por hibridación *Slot-blot*, tal y como se describe en la literatura (Aledo et al., 1995; Pearce et al., 1996). La sonda utilizada es el conjunto de fragmentos de DNA producto directo de la PCR anteriormente descrita para aislar específicamente este tipo de RTN. Este tipo de sonda permite reconocer toda la variabilidad de secuencias amplificadas entre los dos cebadores de la PCR y como se pretende determinar el número de copias total, es preferible en este caso a sondas clonadas que solo representan la secuencia de un elemento, ya que así la estimación es más representativa sobretodo si el RTN-LTR es abundante y heterogéneo (García-Martínez y Martínez-Izquierdo, 2003).

Dicha sonda se marca radiactivamente y se hibrida con diluciones seriadas de cantidades conocidas de DNA genómico, así como diluciones en serie de una cantidad conocida de la misma sonda que se utilizará en la hibridación, que se usa como patrón (**Figura 17**). Como el tamaño del genoma de *C. sinensis* es un valor conocido, la comparación entre la señal de hibridación sobre el DNA genómico y la que se obtiene sobre las diferentes diluciones del fragmento de DNA utilizado como sonda, permite calcular su número de copias presente en cada rectángulo del filtro (*slot*), y por extrapolación en el genoma (ver Material y Métodos).

#### Fiabilidad experimental de la técnica utilizada:

Para cada membrana, se cargan por duplicado tres cantidades diferentes de DNA genómico y tres cantidades de la sonda patrón. La comparación 1 a 1 entre las señales de hibridación genera 36 valores de la estimación del número de copias que hay en cada rectángulo del filtro (*slot*), y por tanto en el genoma. A partir de estos valores se calcula el promedio  $\pm$  error estándar de número de copias por membrana individual. Únicamente se tienen en cuenta aquellos experimentos individuales donde el porcentaje de error experimental es  $\leq 10\%$  [siendo  $E = (SEM/Prom)*100$ ; E: % error; SEM: error estándar de la media; Prom: promedio]. Como mínimo se realizan dos experimentos de transferencia a membrana e hibridación independientes. El número de repeticiones de cada experimento individual dependerá de la reproducibilidad del resultado entre membranas experimentalmente fiables ( $E \leq 10\%$ ).

## Resultados



**Figura 17. Experimento representativo de *Slot-blot* hibridado con la población completa de fragmentos amplificados en la PCR específica para la RT de RTN de tipo *copia*.** La fila A corresponde a diluciones seriadas de la sonda a cantidades: 2,5 ng (slot 1 y 4), 0,5 ng (slot 2 y 5) y 0,1 ng (slot 3 y 6). La fila B corresponde a las diluciones de DNA genómico de naranjo a cantidades: 500 ng (slot 7 y 10), 100 ng (slot 8 y 11) y 50 ng (slot 9 y 12). El filtro se hibrida con la sonda marcada radiactivamente, como se describe en Material y Métodos. Para cada DNA genómico diferente se realizan como mínimo dos réplicas del *slot-blot* y dos experimentos de hibridación independientes.

Se determina que, bajo las condiciones concretas en que se han realizado los experimentos, hay  $9.928 \pm 972,9$  copias de fragmentos de 270 pb de la RT de RTN-LTR de tipo *copia* en el genoma de *C.sinensis* (**Tabla VIII**).

Nº copias - Exp1	Nº copias - Exp2	NÚMERO COPIAS TOTAL <i>copia</i>
9.255 ± 899	10.601 ± 1.809	<b>9.928 ± 972,9</b>

**Tabla VIII. Número de copias total de RTN tipo *copia* en el genoma de *Citrus sinensis*.** Tabla que recoge los valores estimados en las distintas réplicas del experimento de *Slot-blot* y el promedio  $\pm$  error estándar calculado.

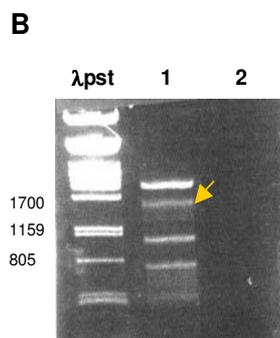
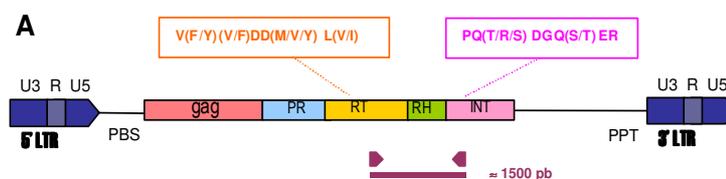
Si suponemos que todos estos fragmentos forman parte de elementos completos de RTN-LTR y que los elementos tipo *copia* descritos tienen como promedio unas 5 kb de tamaño, entonces las aproximadamente 10.000 copias de RTN *copia* constituyen aproximadamente el 13,12 % del genoma haploide de *Citrus sinensis*.

## A.2 OBTENCIÓN DE FRAGMENTOS RTN-LTR TIPO GYPSY EN EL GENOMA DE CITRUS SINENSIS

### A.2.1 Estrategia de amplificación selectiva de RTN-LTR tipo gypsy

Los dos tipos de RTN-LTR, *copia* y *gypsy*, se diferencian en la conservación a nivel de secuencia de sus dominios codificantes y el orden relativo que ocupa la transcriptasa inversa (RT) en el dominio proteico interno en relación a la integrasa (INT), estando después de la INT en *copia* y precediendo a la INT en *gypsy*. Esta característica distintiva ha sido aprovechada para diseñar una estrategia que permite detectar y amplificar por PCR específicamente elementos del grupo Ty3/*gypsy*, excluyendo a los de otras familias de RTN (Suoniemi et al, 1998). La técnica consiste en utilizar como cebadores oligonucleótidos degenerados que corresponden a dos cajas conservadas de los dominios de la RT y la integrasa, en este orden concreto (**Figura 18A**).

La amplificación a partir de DNA genómico con los cebadores degenerados específicos de RTN *gypsy* resulta en un patrón múltiple de bandas tal y como se muestra en la **Figura 18B**. De los fragmentos generados, se purificó y clonó en el vector pGEMT una banda de un tamaño aproximado de 1,3-1,7 kb que corresponde al esperado para la zona comprendida entre la RT y la integrasa de RTN-LTR tipo *gypsy*. Se obtuvo la secuencia nucleotídica de 9 clones individuales escogidos al azar, mediante el secuenciador automático ABI PRISM 377, Applied Biosystems. La nomenclatura utilizada para denominar las los clones conteniendo distintos fragmentos de RTN-LTR fue GySi# (retrotransposon Gypsy de C.Sinensis).



**Figura 18. Obtención de fragmentos de RTN *gypsy* del genoma de *C. sinensis*.** (A) Organización estructural de RTN tipo *gypsy*. Se indica la secuencia aminoacídica y la situación en el elemento de los oligonucleótidos degenerados específicos utilizados en la reacción de PCR y el tamaño de la zona comprendida entre la RT e integrasa que se espera amplificar. La PCR se ha realizado a 6 ciclos con Ta de 45°C y 30 ciclos con Ta de 56°C (Apartado 7.2.2 de Material y Métodos). (B) Imagen de la electroforesis en gel 0,8% agarosa de la amplificación por PCR sobre DNA genómico. Carril λPst, marcador de peso molecular λPst, indicando algunos tamaños en pb a la izquierda; carril 1, *C. sinensis*; carril 2, control negativo (sin DNA). La flecha indica la banda de tamaño esperado, cuyo DNA ha sido purificado y clonado como perteneciente a RTN-LTR tipo *gypsy*.

### A.2.2 Conservación de los dominios funcionales en los elementos GySi

Las secuencias obtenidas se han alineado entre sí en dos partes, por un lado las secuencias correspondientes a la RT, y por otro las de la zona de la integrasa. Posteriormente, las secuencias se han traducido a las seis pautas de lectura. Todos los clones estudiados que muestran similitud con la RT o la integrasa de RTN, presentan unas pautas de lectura truncadas cuando se interrumpe la traducción conceptual que da lugar a una proteína de RTN, por ello han sido tratadas para poder hacer los alineamientos de secuencias aminoacídicas; únicamente uno de los clones, GySi9, presenta una hipotética pauta de lectura continua y no interrumpida por ningún codón de parada. Se observa una clara conservación a nivel de secuencia entre los elementos aislados, preferentemente en dominios funcionales, ya que se confirma la presencia en todos ellos de unos motivos consenso de las secuencias de RTN-LTR de tipo *gypsy*, conservados (Chavanne et al., 1998, Suoniemi et al., 1998).

En la **Figura 19** se muestra el alineamiento por ClustalW de las secuencias aminoacídicas inferidas de 3 de los clones GySi# secuenciados por completo (desde la RT hasta la integrasa): GySi9 (1589 pb), GySi43 (1493 pb) y GySi84 (1602 pb), y de zonas similares de RTN tipo *gypsy* ya descritos como son *Del1* de *Lilium henryi* (NºAcc.: X13886) y *GypsyDM* de *D.melanogaster* (NºAcc.: P10401). Comprobamos que los motivos funcionales de la RT, la RNAsa H y la integrasa están conservados en los elementos *gypsy* aislados del genoma de *C. sinensis*. Tal como queda señalado en la **Figura 19**, el centro activo de la RT (motivo DD) está altamente conservado, y también encontramos, aproximadamente a 40 residuos cadena abajo, el motivo SKCEF perfectamente conservado, conteniendo el residuo lisina (K) siempre presente en las polimerasas RNA-dependientes (Suoniemi et al., 1998). Por otra parte, está presente en los clones aislados de *C. sinensis*, el motivo TDAS que define el centro activo de la RNAsa H, seguido por un residuo de glutamato (E), esencial para la actividad catalítica de la RNAsa H y también el motivo N<sub>3</sub>DXL perfectamente conservado. Para la integrasa, es visible en la región N-terminal el dominio de unión al DNA (HX<sub>6</sub>HX<sub>29</sub>CX<sub>2</sub>C), que forma una estructura de dedo de zinc altamente conservada. El centro activo enzimático se inicia con el motivo altamente conservado GLLQPLPI (Suoniemi et al., 1998), motivo muy similar al presente en los clones aislados de *C. sinensis*. Por último, el motivo DX<sub>60</sub>DX<sub>35</sub>E que está muy conservado en integrasas retrovirales y de RTN y que es esencial para su actividad enzimática, está claramente presente en los clones GySi#.

El orden que los dominios RT, RNAsa H e integrasa ocupan en la secuencia determina que los fragmentos clonados son de RTN-LTR de tipo *gypsy*. La confirmación de la identidad del conjunto de clones obtenidos vino al realizar comparaciones con las bases de datos (GenBank) mediante BLASTX, dando en todos los casos similitudes muy altas (70 %) con fragmentos de poliproteínas de RTN de tipo *gypsy*.

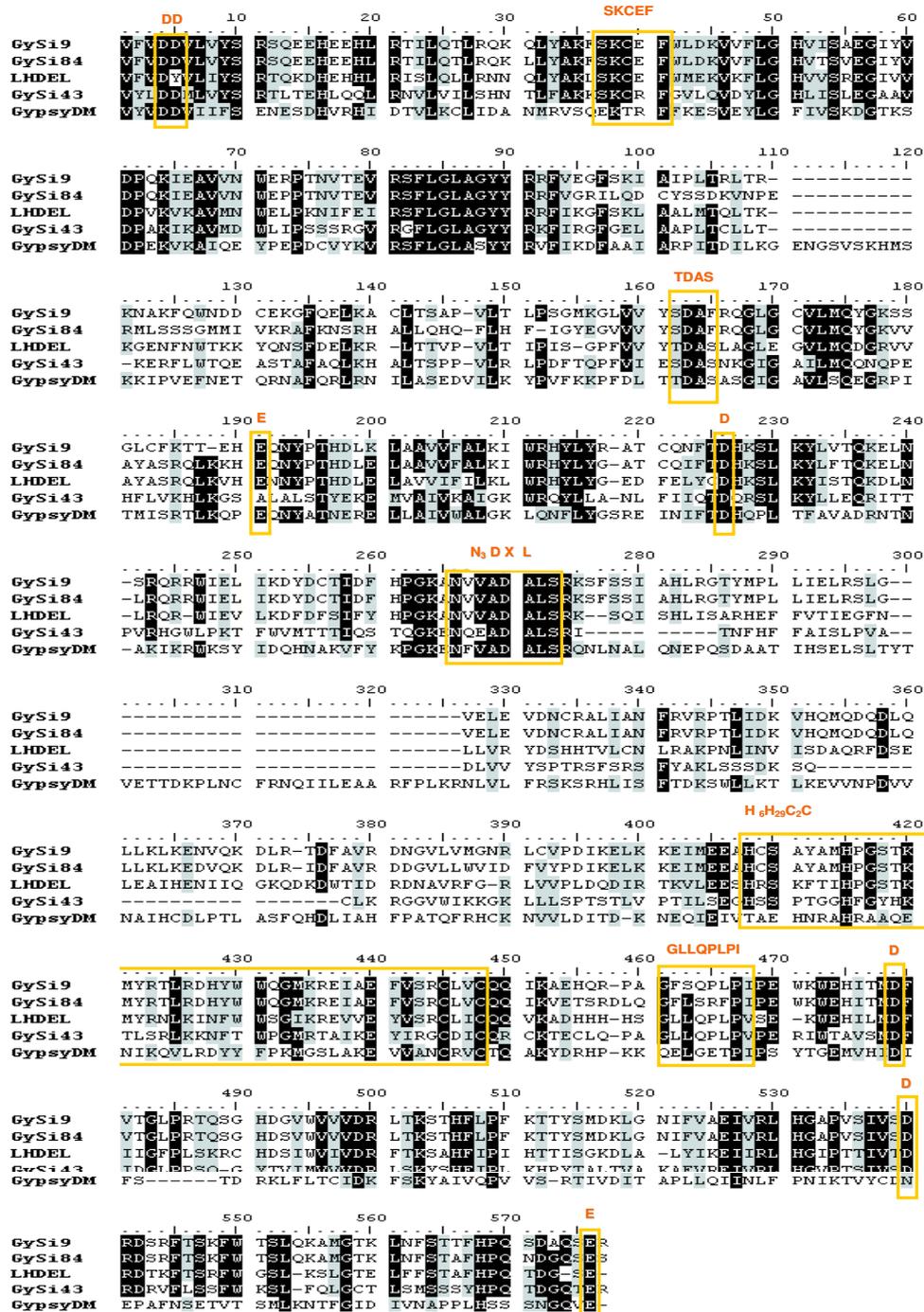


Figura 19. Alineamiento múltiple por ClustalW de las secuencias aminoácidas inferidas de 3 clones GySi# y de zonas similares de RTN tipo *gypsy* ya descritos como son *Del1* de *Lilium henryi* y *GypsyDm* de *D. melanogaster*. El orden relativo de las secuencias está directamente relacionado con el nivel de similitud entre ellas. Los guiones indican ausencia de aminoácido y son añadidos para optimizar el alineamiento. El sombreado oscuro indica identidad de secuencia 95% y el claro, elevada similitud. Se han recuadrado diferentes motivos de RT e integrasa típicos de secuencias de RTNs de tipo *gypsy*. GySi#: clones de *C. sinensis*. *GypsyDm* de *Drosophila melanogaster* (Nº Acc.: P10401). *LHdel1* de *Lilium henryi* (NºAcc.: PX13886).

## Resultados

---

### A.2.3 Descripción de nuevas familias de RTN tipo *gypsy* en naranjo

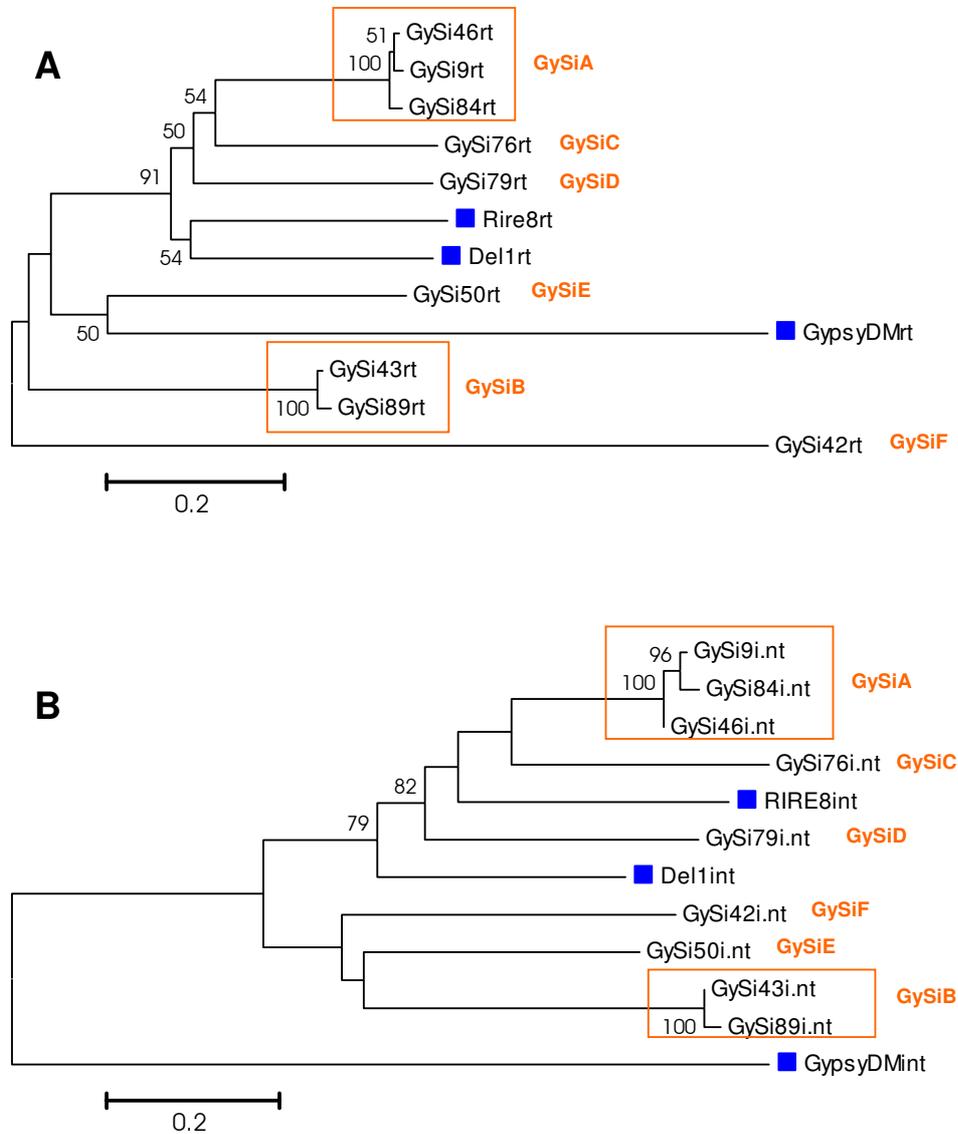
Se han realizado análisis filogenéticos mediante árboles tipo NJ a partir del alineamiento de las secuencias nucleotídicas de la zona de la RT y de la zona de la integrasa por separado (resultados no mostrados), comparándolos con zonas similares de otros RTN-LTR de tipo *gypsy* ya descritos como son *Del1* de *Lilium henryi* (NºAcc.: X13886), *GypsyDM* de *D.melanogaster* (NºAcc.: X13886) y *Rire8* de arroz (NºAcc.: AB014740) (**Figura 20**). En ambos casos se obtienen árboles con distribuciones de secuencias muy parecidas que reflejan una clasificación de los 9 elementos GySi# clonados dentro de 2 familias claras (GySiA y GySiB), con 3 y 2 elementos muy próximos y el resto de secuencias en ramas más divergentes. Los elementos son asignados a una misma familia si comparten una identidad nucleotídica  $\geq 85\%$  entre ellos. La similitud media es del 97% tanto entre los clones de GySiA como para los clones de GySiB. Ambos grupos están apoyados por valores de *bootstrap* muy altos (100). La misma consideración hecha para los elementos tipo *copia* de naranjo es válida para los *gypsy* donde, los 4 elementos con porcentaje de identidad inferior al 85% pasan a ser considerados como pertenecientes a 4 familias diferentes (GySiC-GySiF) con un solo representante entre las secuencias obtenidas en este estudio.

Al analizar más detalladamente las representaciones gráficas de la **Figura 20**, se observa que el grado de divergencia de secuencia entre la región de la RT y la de integrasa es similar, excepto en un caso para el dominio de la RT: secuencia GySi42rt, qué se muestra muy alejado al resto de secuencias GySi# (**Figura 20A**), este dato se explica al analizar su secuencia ya que vemos que está muy truncada, con una elevada presencia de codones de parada y cambios continuos de la hipotética pauta de lectura. Por esta razón, no consideramos este clon en los estudios posteriores.

En el resto de secuencias, el dominio de la RT parece estar más conservado entre los RTN-LTR tipo *gypsy* de distintas especies que el dominio de la integrasa ya que, en el árbol NJ basado en el alineamiento de secuencias de la zona de la integrasa (**Figura 20B**) las secuencias son más divergentes y se observa que *GypsyDM* se ubica claramente como *outgroup*.

La diversidad nucleotídica ( $\pi$ ) es algo menor en la región de la RT (0,38) que en integrasa (0,40). Además, la longitud de las ramas en el árbol del alineamiento de la zona RT es menor, lo que nos está indicando una menor distancia genética. Para la región de la RT, los miembros de GySiA presentan una identidad nucleotídica del 64% con una región similar del RTN *Del1* de *Lilium* (NºAcc.: X13886), del 65% con *RIRE8* de arroz (NºAcc.: AB014740), del 44% con los miembros de GySiB y del 3% con *GypsyDM* de *D.melanogaster* (NºAcc.: P10401). A su vez, los miembros de GySiB presentan una identidad nucleotídica del 44% con *Del1*, del 52% con *RIRE8*, y del 4% con *GypsyDM*. Para la región de la integrasa, los miembros de GySiA presentan una identidad

nucleotídica del 50% con una región similar de *Del1*, del 57% con *RIRE8*, del 40% con los miembros de GySiB y del 3% con *GypsyDM*. A su vez, los miembros de GySiB presentan una identidad nucleotídica del 35% con *Del1*, del 40% con *RIRE8* y del 6% con *GypsyDM*.



**Figura 20. Árbol filogenético basado en el alineamiento de secuencias nucleotídicas de fragmentos de los dominios de la RT (A) y de la integrasa (B) de RTN de tipo *gypsy* clonados del genoma de *C. sinensis*. Árbol filogenético tipo *Neighbour joining* (NJ) generado con el programa MEGA2. La divergencia entre las secuencias en unidades de distancia genética viene definida por la longitud de las ramas horizontales. Esta indicado en cada bifurcación los valores de bootstrap  $\geq 50\%$ . Se han recuadrado las diferentes subfamilias en qué quedan divididas las secuencias, el nombre de cada subfamilia está indicado a la derecha de la figura. Secuencias de *Drosophila melanogaster* (*GypsyDM*, N<sup>o</sup> Acc.: P10401), de *Liium henryi* (*Del1*, N<sup>o</sup> Acc.: X13886) y de *Oryza sativa* (*RIRE8*, N<sup>o</sup> Acc.: AB014740) se han incluido en la comparación (■).**

## Resultados

---

### A.2.4 Estimación del número de copias de los RTN-LTR tipo *gypsy* en el genoma de naranjo

Se ha estimado el número de copias total de RTN-LTR tipo *gypsy* por genoma haploide en *C. sinensis* mediante reconstrucción genómica a partir del análisis por hibridación *Slot-blot*, tal y como se detalla en el apartado A.1.4 para RTN tipo *copia*. En este caso, la sonda utilizada es directamente el producto de PCR utilizando los cebadores degenerados específicos para aislar RTN tipo *gypsy*.

Se determina que, bajo las condiciones concretas en que se han realizado los experimentos, hay  $4.902,7 \pm 610,6$  copias de fragmentos de 1500 pb de la RT-integrasa de RTN de tipo *gypsy* en el genoma de *C. sinensis* (Tabla IX). Si, como en el caso de *copia*, suponemos que todos estos fragmentos forman parte de elementos completos de RTN (los elementos *gypsy* descritos tienen como promedio unas 8 kb de tamaño), entonces las aproximadamente 5.000 copias de RTN *gypsy* constituyen alrededor del 10,49 % del genoma de *Citrus sinensis*.

Nº copias - Exp1	Nº copias - Exp2	Nº copias - Exp3	NÚMERO COPIAS TOTAL <i>gypsy</i>
$6116 \pm 467,4$	$4416 \pm 557,4$	$4176 \pm 553,9$	<b><math>4902,7 \pm 610,6</math></b>

Tabla IX. Número de copias de los RTN-LTR tipo *gypsy* en el genoma de *Citrus sinensis*. Tabla que recoge los valores estimados en las distintas réplicas del experimento de *Slot-blot* y el promedio  $\pm$  error estándar calculado.

El número total de RTN-LTR por genoma haploide se calcula como la suma del número de elementos *copia* y *gypsy*. Así, en este estudio se estima mediante análisis por hibridación *Slot-blot* que los RTN-LTR constituyen aproximadamente el 23% del genoma del naranjo, siendo los *copia* más abundantes que los *gypsy* en esta especie. Se confirma que los RTN-LTR son elementos altamente repetitivos del genoma de *C. sinensis*.

### **A.3 CLONAJE, SECUENCIACIÓN Y CARACTERIZACIÓN DE UN NUEVO RTN TIPO COPIA DE NARANJO (CIRE1)**

#### **A.3.1 Estrategia de clonaje**

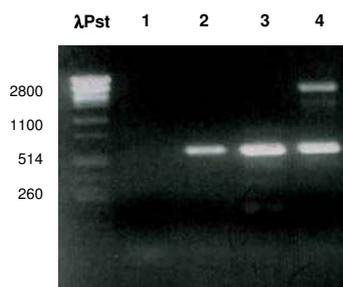
Una vez demostrada la presencia de RTN-LTR tipo *copia* en el genoma de *Citrus sinensis* mediante la amplificación por PCR de parte de la región de la RT, el siguiente paso dentro de la investigación fue clonar un elemento entero para proceder a su caracterización. El clonaje del elemento *copia* entero se ha realizado sobre DNA genómico de *Citrus sinensis*, variedad "Valencia late".

##### **A.3.1.1 Amplificación por PCR inversa de CIRE1 sobre DNA genómico**

La técnica de la PCR inversa (*inverse-PCR*, i-PCR) permite amplificar fragmentos de DNA adyacentes a una secuencia conocida en el genoma. Así, a partir de las secuencias de las que disponemos del dominio de la RT (de los fragmentos generados por PCR) se ha aplicado esta técnica para secuenciar totalmente aquellos RTN considerados más relevantes e informativos para estudios posteriores.

El clon del dominio de la RT escogido como representante de los RTN tipo *copia* en el genoma de *C. sinensis*, y punto de partida para el clonaje del RTN entero, fue el CoSi14 por presentar una única pauta de lectura abierta en la secuencia conocida, correspondiente al dominio de la RT. Este elemento pertenece a una de las familias mayoritarias de RTN *copia* descritas en naranjo (CIRE1).

A partir de DNA genómico digerido con EcoRI o HindIII y ligado entre si (ver Material y Métodos) se han amplificado por i-PCR fragmentos de DNA adyacentes a la secuencia conocida del clon, utilizando para la reacción cebadores internos encarados hacia fuera y complementarios a las zonas más externas de la secuencia conocida de CoSi14. La secuencia de estos cebadores (denominados oligo5 y oligo6) y las condiciones de amplificación se describen en Material y Métodos. El resultado de la i-PCR se muestra en la **Figura 21**.

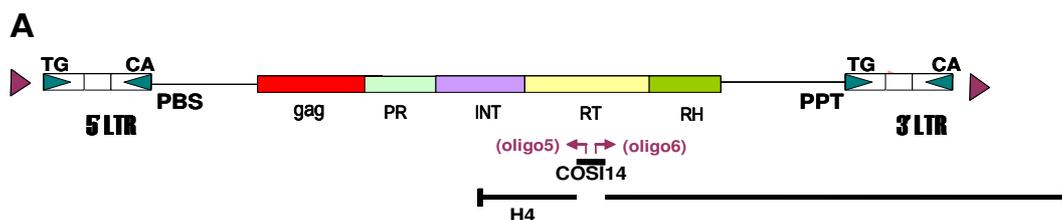


**Figura 21. Resultado de la PCR inversa sobre DNA de *C.sinensis* flanqueando al fragmento de CoSi14.** Imagen del gel al 1% de agarosa tras la electroforesis de los fragmentos de DNA amplificados. Carril  $\lambda$ Pst, marcador de peso molecular  $\lambda$ Pst, indicando algunos tamaños en pb; carril 1, control negativo (sin DNA); carriles 2 y 3, DNA genómico digerido con EcoRI; carril 4, DNA genómico digerido con HindIII. La PCR se ha realizado a 30 ciclos con Ta de 61°C (Apartado 8 de Material y Métodos).

## Resultados

Aquellos productos de amplificación con un tamaño superior a 1 kb fueron purificados, clonados y secuenciados. Cada nueva secuencia obtenida se enfrentaba contra las bases de datos (Genebank) mediante Blast/Fasta para confirmar si era de RTN.

El fragmento de mayor tamaño amplificado aplicando la i-PCR sobre DNA genómico total corresponde al clon H4 de 2873 pb. La comparación con las bases de datos confirma la identidad de la secuencia amplificada como perteneciente a un RTN de tipo *copia*. H4 presenta una única pauta de lectura continua sin interrupciones por codones de parada cuando se compara con secuencias de RTN. El alineamiento de la secuencia aminoacídica deducida conceptualmente del elemento con regiones similares de RTN tipo *copia* como *Tnt1* (NºAcc.: P10978) y *BARE1* (NºAcc.: Z17327) con los que muestra similitud (50%), ha servido para definir una región RT-RNasa H que incluye motivos consenso muy conservados para estas regiones. Seguidamente al motivo 2 de la RNasa H, el elemento muestra una zona no codificante donde, siempre teniendo como referente la secuencia nucleotídica de *Tnt1*, se han detectado posibles inicios de la LTR3' ("TG..") precedidos de una zona rica en purinas (PPT), estructura típica y presente en todos los RTN-LTR. También detectamos potenciales cajas de la región del promotor (como CAAT-box o TATA-box, entre otras) que nos confirman que se trata realmente de la secuencia de una LTR (**Figura 22**).



**B**

ORF1 (1-2183) 728 aa

```
KLP RSFWAEAVNPASYLVTRSPSSAMASKHQKNCGMENQPPNLRVLGALLTLTSSRQAEARALKEVFGYDPGVKGY
KIWCKGSRKCI VSRDVFVHESVLLKESAHDAGLQDNPAANNRSGSETSKVNVELLTDKSEKEAASDDERATAESE
HEVSELPQADLQNYQLARDVRREVRAPVRYGYADLIAYALLCADEVITIEEPANFSEAMESVHCDKWLEAMQDEMESL
QRNQWTWTLIPNPGNKRLINCKWIFKRNEGIPDVEPPKYKARLVARGFTQREGVDFNEIFSPVVKHSSIRILLAMVALL
DLELEQMDVK TAFLHGNLEEQILMAQPEGFECKDEYVCLLHK SLYGLKQSPRQWYRRFDDFMVSKGYRRSRYSV
YFGGSDQGGVA YLLLYVDDMLIASKYKSEIERLKNLLKAEFEMKDLGNAKRILGMDIIRDRSAGTLFLSQGKYIKKVL
ERFEMQDCKPVQTP LGPQFKLSAATTSEDESQMNFFYAQAVGSLMYAMVCTRSDIAYAVSVVSRYLSCPGKVHWNVA
KWIIMRYLKGSSSTCGLLYGKTRSDKIEVMGFVDSDFAGDLDRRSTSRYMFLVNSCLISWKSLLQSGGALSSTAEFIA
TTEAVKEAMWLRGLLNLWLNQKIVQVFCDNQNAIHLVKNQMYHE RTKHIDVKLFIRDEVGKGTVVVSKIHTSVN PA
DALTKSLP TAKFEFCVNL MGI MPKSN *
```

Región No codificante (2187-2872pb)

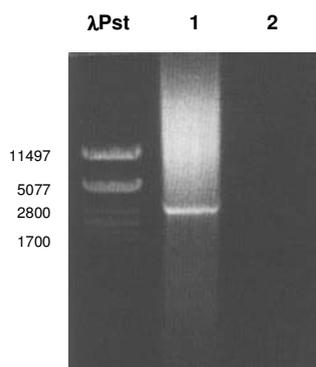
```
TAAATTCAAGAAGATGAATTGTAGTTGAATGAAGGTTTTTATAGGTGCTGATTATTTGAAGGGATCAACTGAATCAACTGCAAGGTGGAGATTGTG
AAGATATGTGTATATGGACTTAAATGAAGAACTGGAAAAGTAGAGAGCTGTTGTGCAGATTGACTTCAGTTGATGGGGTCTATTTTGATATCT
TCCGCTGACGTGGCAGTAATCCGGTTATAATCTGTTATAACGAAGCCACAGCACATGATATAAAAACAGAGGCTGACGAATGGTGCCTCATACTG
TGCTAATAICGICTTCATTTGTTTCATCATTTTGCATAGAATCTGTTTCATTGTGATCTGTGCATAATCTTCTGTAATACACCTGTGATGGGTT
AGTTCTTGATCAATAATAATGAGTAAATTTTCTTCTAAGITTTGCTGGTCTGTGCTTTTGTGTTTGGCTCGGGTTAGCTTCTTTGGTCTGGGTT
TTTCTTGATTCGATTGGTGCTTTTGGTGTCTACTACCAAGATTGCTGAGTCTTGCTGTGGTTAGACTTAAATCTTAACACAAAGAAATG
TGCCAGATGAAGACTTTCCAGATGTGCTTAATAGTGCTTACACCACATAATCTTCCGCTCAGAGATGGTGTACTGAGAAAAGTGGGCTGAGA
GACAACAATAACTGGATTGCCGAAGAAGCTT
```

**Figura 22. Secuencia H4. (A) Esquema de la estrategia de amplificación y clonaje.** A partir de una secuencia parcial de la RT de 266 pb (CoSi14) se han diseñado dos cebadores internos encarados hacia fuera (oligo5 y oligo6) con los que se ha conseguido amplificar mediante i-PCR un fragmento de 2873 pb, correspondiente al clon H4. **(B) Secuencia aminoacídica inferida y secuencia nucleotídica no codificante del clon H4.** Se resaltan las secuencias consenso propias de la RT (TAFLHG/SLYGLKQA/YVDDM) y los motivos 1 y 2 conservados de la Rnasa H (RTKHID/ADIFLTK). El asterisco (\*) indica el codón de parada (TAA) que da paso a la región no codificante. Se subrayan también potenciales zonas ricas en purinas (PPT), posibles inicios de la 3'LTR, e hipotéticas cajas del promotor presentes en la región 3' secuenciada. En color verde se señala la secuencia del oligo9, utilizado en el siguiente apartado de los resultados.

### A.3.1.2 Amplificación por PCR directa de CIRE1 sobre DNA genómico

Aprovechando una de las características intrínsecas de las LTR (son secuencias idénticas entre si en el momento de la inserción del elemento) hemos diseñado un oligonucleótido complementario a una zona de la hipotética LTR3' ya secuenciada. Este cebador se ha utilizado para amplificar por PCR directa el resto del elemento, no incluido en el clon H4.

En una primera aproximación, el cebador directo (oligo9) se ha diseñado a partir de una zona cadena abajo de cualquiera de los posibles inicios "TG" de la LTR3' (próxima a la potencial caja TATA), descartando que la secuencia del cebador esté situada en la secuencia no codificante cadena arriba de la LTR3', que no tendría secuencia complementaria en la LTR5' y por tanto no amplificaría ningún fragmento. Como cebador inverso, se utiliza un oligo diseñado a partir de una zona de la RT (oligo5). La secuencia de estos cebadores y las condiciones de amplificación se describen en Material y Métodos. Como puede verse en la **Figura 23**, se consiguió amplificar una banda mayoritaria de aproximadamente 3 kb, qué fue purificada, clonada y secuenciada con el nombre de MP20.

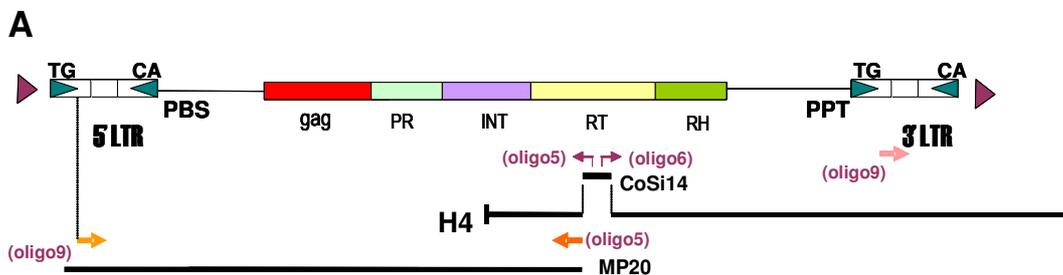


**Figura 23. Resultado de la electroforesis en gel 0,8% de agarosa de la amplificación por PCR directa de CIRE1 sobre genoma de *C.sinensis* para obtener el extremo 5' del elemento.** En la reacción de PCR se ha utilizado la combinación de cebadores: oligo9 y oligo5. La secuencia de los cebadores y las condiciones de amplificación se describen en Material y Métodos. La PCR se ha realizado a 25 ciclos con Ta de 54°C. Carril λPst, marcador de peso molecular λPst, indicando algunos tamaños en pb. Carril 1, DNA *C. sinensis*. Carril 2, control negativo (sin DNA).

El clon MP20 tiene 3285 pb y solapa en una región de 1044 pb con H4. Comparando las secuencias nucleotídicas de la región solapada, se observan 63 cambios nucleotídicos entre los dos fragmentos clonados, lo que supone un 94% de identidad. La secuencia nucleotídica de MP20 se tradujo a secuencia aminoacídica observándose la presencia,

## Resultados

por este orden de un motivo consenso (CX<sub>2</sub>CX<sub>3</sub>GHX<sub>4</sub>C) característico de secuencias de *gag*, una secuencia D(T/S)G típica de proteasas, un dominio de unión al DNA (HX<sub>3-7</sub>HX<sub>20-32</sub>CX<sub>2</sub>C) propio de la integrasa y como era de esperar, el típico motivo TAFLH de la RT próximo a la secuencia complementaria al cebador utilizado en la reacción de PCR (**Figura 24**). Se confirma claramente que la secuencia de MP20 se corresponde con la del extremo 5' de un elemento tipo *copia* (dado el posición relativa que ocupan la integrasa y la RT) y qué juntamente con la secuencia de H4 podría ser suficiente para contener un elemento entero [dado que el tamaño estándar de los RTN *copia* caracterizados es de 5-6 kb (Kumar y Bennetzen, 1999)]. Cabe destacar que la secuencia conceptual de aminoácidos de MP20 se obtuvo sin saltos en la pauta de lectura aunque su secuencia aminoacídica está interrumpida por 6 codones de parada situados mayoritariamente en la región del elemento que codifica la integrasa.



Región No codificante (1-432 pb)

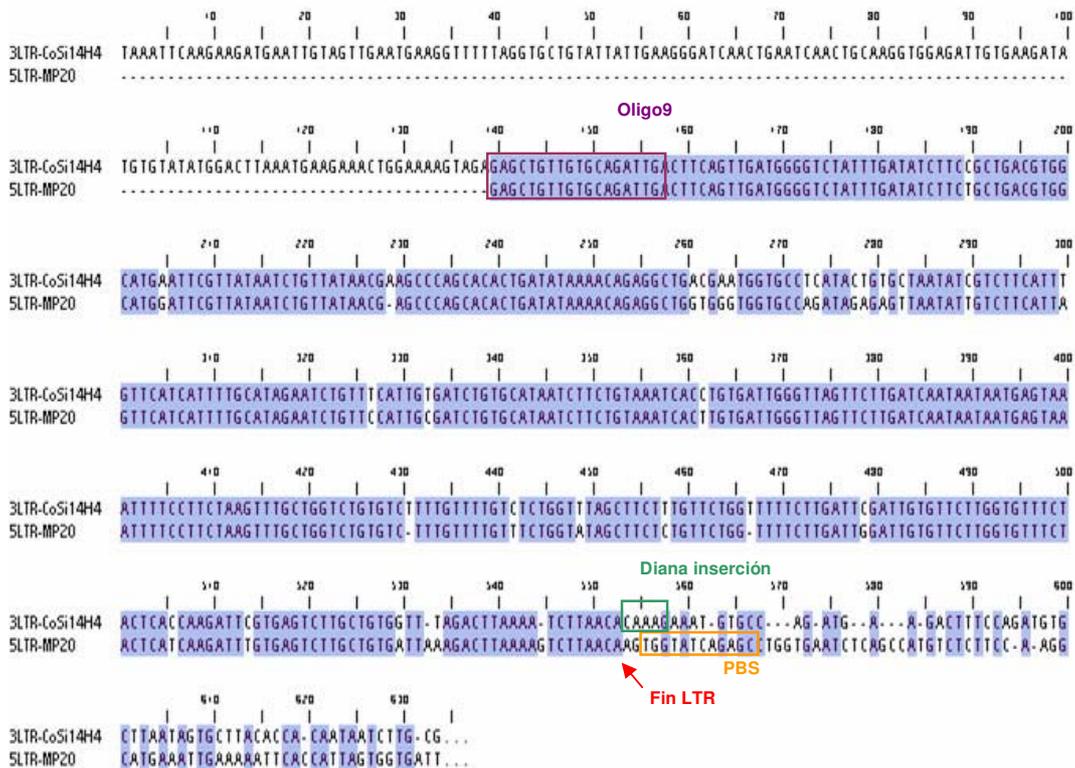
```
GAGCTGTTGTGCAGATTGACTTCAGTTGATGGGGTCTATTTGATATCTTCTGCTGACGTGGCATGGATTCTGTTATAACGAGCCC
AGCACACTGATATAAAACAGAGGCTGGTGGGTGGTCCAGATAGAGAGTAAATATTGTCTTCATTAGTTCATCATTTTTGCATAGAATCTGTCC
ATTGCGATCTGTCATAATCTTCTGTAATCACTTGTGATTGGGTTAGTCTTGTGCAATAATAATGAGTAAATTTCTTCTAAGTTTGTCTGG
TCGTGCTCTTTGTTTTGTTCTGGTATAGCTTCTCTGTTCTGGTTTCTTGTGATTGGATTGTGTTCTTGGTGTCTACTCATCAAGATTGTGA
GTCTTGCTGTGATTAAGACTTAAAGTCTTAACAAGTGGTATCAGAGCCTGGTGAATCTCAGCCATG
```

ORF1 (433-2285) 942 aa

```
MSLPRHEIEKFTISGDFSLWKLKMRALLVHQARICPEEEDPEASTGVSVIDDKKRIQNRHSTLILSLGDSILREISE
EKTALGIWNKVVETLCMKKSLAHLRFLKRLYTFMSREWVTIQDHIDTFNKIILDLEGVENVKICDEDKAFFLLSSLPK
SYEGFVDTMLYGRITLTLLEDVKASLSKEI*KNELETSNEGELIARSEKKKDKNKNQKGGHGNQETDDKDKKRRK
CFYCRNEGHYIRDCFEKSKESQEKSGDAVAASDDGSDGYQSADLLVASNSNTKQWVIDSGCSFHLCPKTLFYKYE
AVDGGRVLMGNMNCNIVGIGSVKIKMFDGDIRSLHEVRHAPRLKRNLI SLGMLDNLGYFFKSKSGGLEIRKGTIIVM
KGVKENGLYVLQGSVSVQEGISAVYEEDRTKLWHLRLGHIINIKGLLELSKQGLLGGDRIQQLFCENCFIFKSHRSK
FHKGEHMSKQVLDYAHTDLWGPAQVPSLGGRYFMSLIDDYSRVVWYIILKTKDQALEKFNV*KSLVENQSGFKLKLCL
RTDNGLEFYSEFEYEQKHGKIKMHTVRFPTQONGLAERMNRTLVDKTRCMLIIQSFPFSFVVEAVNTTSYLINRSP
SSTIGFKTPEELWNGKPANYQNLRVFGCPAYLHIKQGLDARALK*VFGYPDGVKGYKIWCKEKGCIVSRDVPFHE
SALLKESAEYDTGLQHNSAVNRQSGTSKVKVELLTDKGSEKETTSDDERATAESEHHDVSELPQADLQSYQLARGVRV
REVRAPVR*GYADLMAYALLCADEMTIEEPSNFSEAMESVHCDKWLEAMQDKMESLQRNKTWTLIPNPGNKRLISCKW
IFKRKE*IPDVEPPKYKARLVARGFTQREGVDFNEIFSPVVKHSSITILLAMVALLDLELEQMDVKTAFLHGNLEEYI
LMVQPEGFCECKD
```

**Figura 24. Secuencia MP20. (A) Esquema de la estrategia de clonaje.** A partir de la secuencia H4 correspondiente a la parte de LTR3' se diseña un cebador (oligo9) que hibridará también con la región complementaria de la LTR5' y va a permitir amplificar por PCR directa, junto al cebador inverso, complementario a la zona de la RT (oligo5) la región 5' del elemento, obteniéndose el clon MP20. **(B) Secuencia nucleotídica no codificante y secuencia aminoacídica inferida del clon MP20.** En verde se señala la secuencia del oligo9. Se marca la posible caja TATA del promotor presente en 5' en la región secuenciada. El codón **ATG** que codifica para la metionina (**M**) da paso a la región codificante. Se resaltan también en azul las secuencias consenso propias de la *gag* (CX<sub>2</sub>CX<sub>3</sub>GHX<sub>4</sub>C), la proteasa (D[T/S]G), la integrasa (HX<sub>3-7</sub>HX<sub>20-32</sub>CX<sub>2</sub>C) y la RT (TAFLHG). Los asteriscos (\*) indican codones de parada en la secuencia

Las secuencias de que disponemos nos permiten hacer ya una comparación de las posibles LTR del elemento. En la **Figura 25** se muestra el alineamiento entre las dos secuencias que contienen las hipotéticas LTR amplificadas esto es, la de la LTR5' incluida en el extremo 5' no codificante del clon MP20, desde el inicio de la secuencia (oligo9) hasta el ATG iniciador que define el comienzo de la región codificante, y la de la LTR3' incluida en el extremo 3' del clon H4 (zona no codificante a partir del codón de parada hasta el final de la secuencia). Vemos que la alta similitud se mantiene desde el oligo9 hasta los nucleótidos CA, con algunos cambios puntuales. La homología se pierde bruscamente en un punto que coincide con una secuencia CA típica de las repeticiones invertidas terminales de las LTR.



**Figura 25. Alineamiento de las secuencias nucleotídicas de las dos posibles LTR del elemento clonado del genoma de *C. sinensis*.** La flecha señala el punto donde se pierde bruscamente la alta similitud entre las secuencias, que coincide con una secuencia CA típica de las repeticiones invertidas terminales de las LTR. Se encuentra inmediatamente después del final de la LTR3' la posible diana de inserción del elemento en el genoma del naranjo, así como la secuencia correspondiente al PBS presente a 2 pb cadena abajo del final de la LTR5'. Se ha recuadrado también la secuencia complementaria al cebador (oligo9) diseñado a partir de la secuencia de H4 y que ha sido utilizado para amplificar la secuencia MP20 mediante PCR directa sobre DNA genómico.

## Resultados

---

Si este es el final de la LTR de 5' a continuación debería encontrarse la secuencia consenso del PBS (*primer binding site*) que es el sitio de unión al cebador que utilizan los RTN para el inicio de la retrotranscripción, el más común es complementario al tRNA metionina, cuya secuencia es exactamente la misma que encontramos en nuestro elemento (TGGTATCAGAGC). La secuencia cadena abajo de la LTR3' corresponde ya a DNA genómico del cítrico, cuyos 5 primeros nucleótidos constituyen presumiblemente la diana de inserción del elemento. Cuando se inserta un RTN en el genoma huésped, un segmento corto de este DNA (mayoritariamente 5 pb para los RTN-LTR) se duplica en el sitio de inserción. Así que probablemente en nuestro caso la diana de inserción sea la secuencia CAAAG.

Tenemos por tanto identificadas en las secuencias disponibles: la LTR3' y a continuación la posible diana de inserción, así como gran parte de la LTR5' y la secuencia del PBS. Faltará únicamente el inicio de la LTR5' y, precediéndolo, la duplicación de la diana de inserción al inicio del elemento, para poder afirmar que el RTN-LTR está completo.

### A.3.1.3 Localización del posible inicio de la LTR5'

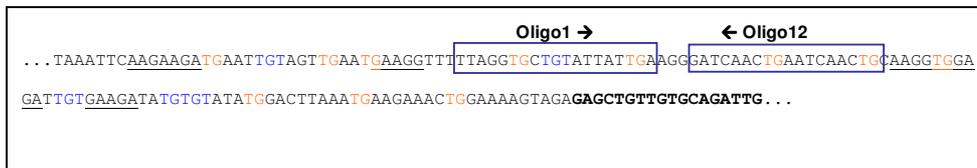
A continuación en la **figura 26**, se muestra la secuencia nucleotídica no codificante de una zona de unos 140 pb del extremo 3' de H4, desde el codón de parada de la RNasa H hasta la secuencia utilizada para diseñar el oligo9. Sabemos que la secuencia complementaria al oligo9 pertenece claramente a la LTR5' ya que nos ha permitido amplificar por PCR directa un fragmento de la región 5' del elemento; por lo tanto, el potencial inicio "TG" de la LTR3' tiene que estar situado necesariamente dentro de estos aproximadamente 140 pb.

Otra de las características intrínsecas de las LTR es que presentan repeticiones invertidas terminales así que, como la secuencia final de las LTR definidas para el elemento de CIRE1 es "..TCTTAACA", se espera que las LTR empiecen con la secuencia "TGTTAAGA...", si la repetición terminal es perfecta y de 8 nucleótidos, ya que, dependiendo del RTN la longitud de la repetición invertida es variable y no siempre perfecta, habiendo elementos donde únicamente se repite el binomio "TG..CA". Además, el potencial inicio de LTR3' ha de estar precedido también por una zona rica en purinas (PPT). Basándose en todo esto, se consideraron todos los posibles inicios "TG" presentes en estos 140 pb, que han sido tenidos en cuenta para diseñar cebadores con los que mediante PCR directa sobre DNA genómico, intentar amplificar la parte de la LTR5' que le falta al elemento (**Figura 26**).



**Figura 26. Secuencia nucleotídica no codificante del extremo 3' de H4.** Se muestran los 140 pb desde el codón de parada (TAA) hasta la secuencia utilizada para diseñar el oligo9 (en negrita). En azul están marcadas las secuencias "TGT.." y en naranja las secuencias "TG.." presentes en la secuencia. También, se han subrayado zonas ricas en purinas y se ha recuadrado la secuencia complementaria al cebador diseñado.

En una primera aproximación, se utilizaron como cebadores el oligo1 (directo) y el oligo19 (inverso, diseñado a partir de una zona de la integrasa). La secuencia de estos cebadores y las condiciones de amplificación se describen en Material y Métodos. No se consigue amplificar ningún fragmento con estas condiciones. La zona utilizada para diseñar el oligo1 pertenece probablemente a la secuencia que precede la LTR3' que al no estar repetida en la región 5' del elemento no amplifica ninguna banda en la PCR. Para confirmar este punto, se diseñó un nuevo cebador (oligo12) complementario a la secuencia situada 3 pb cadena abajo de oligo1 y encarado en sentido contrario a este (**Figura 27**).

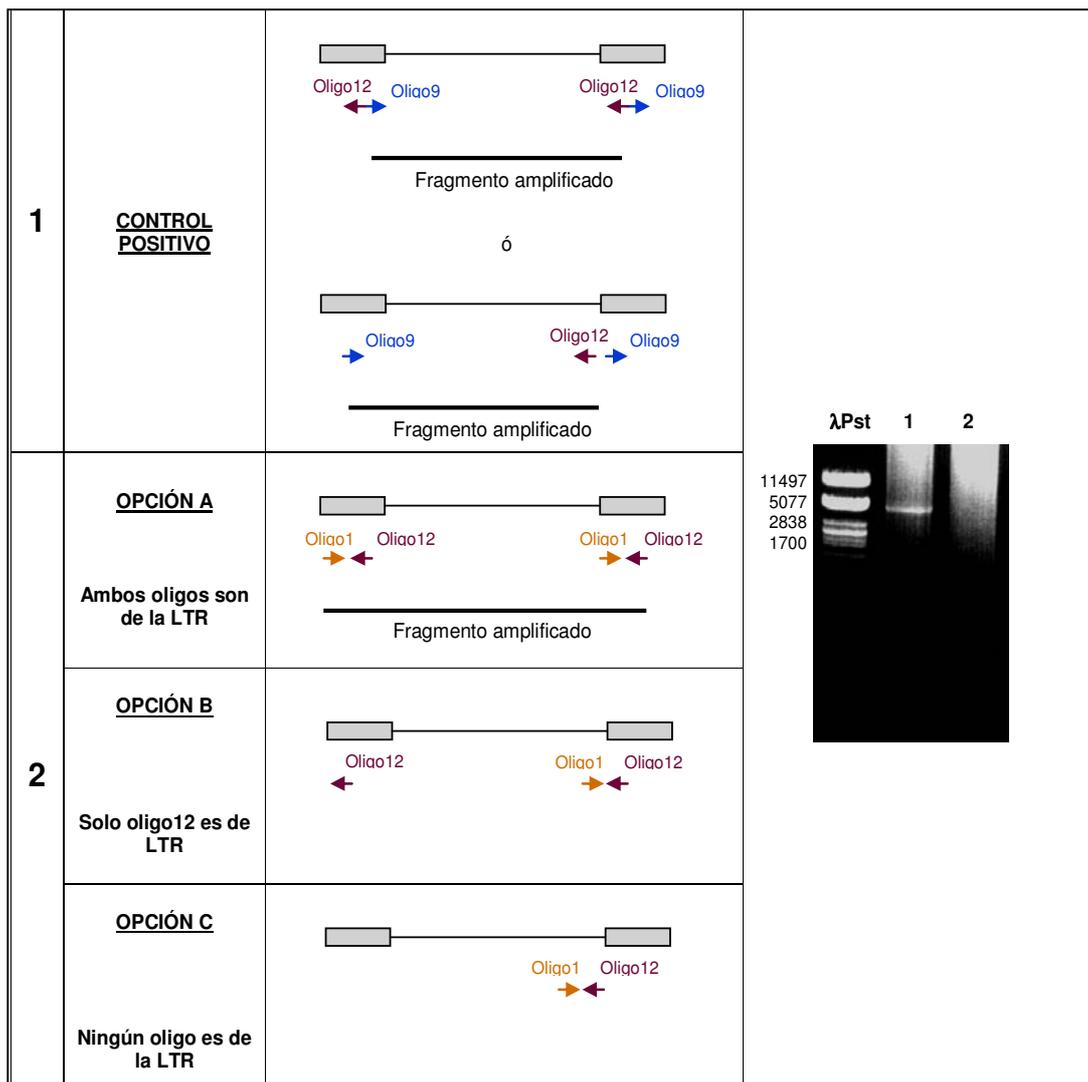


**Figura 27. Secuencia nucleotídica no codificante del extremo 3' de H4.** Se muestra la misma Figura26 con la adición de los cebadores diseñados y la orientación de los mismos.

El planteamiento del experimento es el siguiente: si la secuencia complementaria a oligo1 pertenece a la LTR entonces la de oligo12 necesariamente también, en este caso realizando una PCR sobre genómico con estos dos cebadores se amplificará, un fragmento del tamaño del elemento menos una LTR (**Figura 28, opción A**); en cambio si la secuencia complementaria a oligo1 queda fuera de la LTR entonces no conseguiremos amplificar ningún fragmento, independientemente de que oligo12 sea o no de la LTR (**Figura 28, opción B o C**). Como control positivo de la reacción de PCR utilizamos los cebadores oligo9 (complementario a una zona de la LTR) y oligo12 que generarán un fragmento del tamaño del elemento menos una LTR en la amplificación independientemente de que la secuencia del oligo12 sea o no de la LTR (**Figura 28, control**). Como hipótesis inicial del experimento consideramos que no hay ningún otro elemento o parte de uno (que sea reconocido por los mismos cebadores) entre las 2 LTR. Las condiciones de amplificación se describen en Material y Métodos. Conseguimos

## Resultados

amplificar un fragmento de tamaño esperado al realizar la PCR sobre DNA genómico utilizando la combinación de cebadores: oligo9 y oligo12 (**Figura 28, carril 1**) en cambio, no se obtiene ningún tipo de amplificación utilizando la combinación de cebadores: oligo1 y oligo12 (**Figura 28, carril 2**) lo cual nos está indicando que la secuencia complementaria al oligo1 correspondiente al elemento no pertenece a la LTR del mismo.



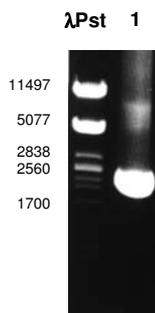
**Figura 28. Diseño del experimento para determinar si una secuencia de la región 3' de CIRE1 pertenece o no a LTR.** A la izquierda de la figura se muestra el esquema del planteamiento de las reacciones de PCR y los posibles resultados. A la derecha se muestra la foto del gel 0,8% de agarosa de la electroforesis del resultado de la reacción de PCR. Carril  $\lambda$ Pst, marcador de peso molecular  $\lambda$ Pst, se indican algunos tamaños en pb. Carril 1, amplificación con oligo9+oligo12 (control positivo); carril 2, amplificación con oligo1+oligo12.

Se diseñó un nuevo cebador (oligo21) complementario a una zona un poco más interna de la LTR3', cuya secuencia era una repetición invertida imperfecta (7 de 8 nucleótidos) del final de la LTR3', tal y como se muestra en la **Figura 29**. Además, para clonar el inicio de la LTR5' del mismo elemento, se añadió a la secuencia del cebador, la secuencia de la diana de inserción cadena arriba del potencial inicio "TGT..", tal y como debería encontrarse duplicada en el genoma flanqueando al RTN-LTR CIRE1. De esta manera también se consigue aumentar la especificidad de la amplificación. Nuevamente utilizamos como cebador inverso el oligo19 (complementario al zona de la integrasa). La secuencia de estos cebadores y las condiciones de amplificación se describen en Material y Métodos.



**Figura 29. Secuencia nucleotídica no codificante del extremo 3' de H4.** A la misma Figura 26 se le ha añadido los últimos 15 nucleótidos de la secuencia de la LTR3'. No se ha escrito la secuencia de la LTR3' entera para simplificar. En negrita se han resaltado los 8 últimos nucleótidos de la LTR3', los cuales están repetidos e invertidos (repetición imperfecta de 7 sobre 8) en el inicio de la secuencia complementaria al cebador oligo21 diseñado (remarcado con un cuadro en la figura). A la secuencia del oligo21 se le ha añadido a 5' la secuencia CAAAG correspondiente a la diana teórica de inserción del elemento.

En este caso sí se consigue amplificar una banda del tamaño esperado, aplicando la PCR directa sobre DNA genómico total con los cebadores oligo21 y oligo19 (**Figura 30**), confirmando la localización más probable del inicio de la LTR. Este fragmento amplificado de aproximadamente 2300 pb fue purificado, clonado y secuenciado con el nombre de insert3.25.

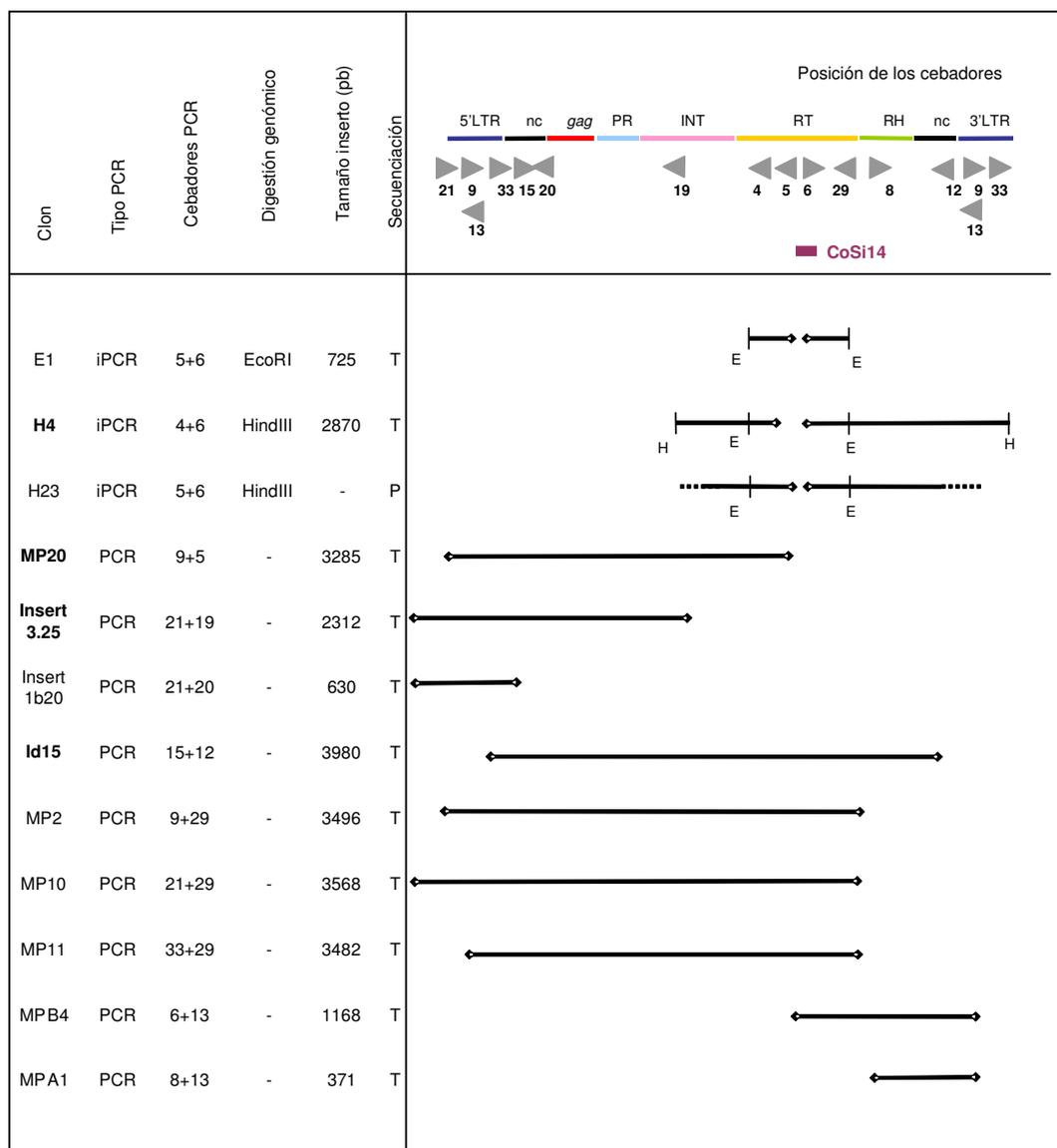


**Figura 30. Resultado de la electroforesis en gel 0,8% de agarosa de la amplificación por PCR directa sobre genoma de *C.sinensis* para obtener la secuencia del inicio de la LTR5' del RTN-LTR.** En la reacción de PCR se ha utilizado la combinación de cebadores: oligo21 y oligo19. La secuencia de los cebadores y las condiciones de amplificación se describen en Material y Métodos. La PCR se ha realizado a 25 ciclos con Ta de 52°C. Carril λPst, marcador de peso molecular λPst, indicando algunos tamaños en pb. Carril 1, DNA *C. sinensis*.

## Resultados

### A.3.1.4 Subclonaje, secuenciación y ensamblaje de las regiones amplificadas

Además de los insertos anteriormente descritos, se han conseguido clonar otros fragmentos mediante amplificación por PCR con distintas combinaciones de cebadores diseñados a partir de las secuencias nucleotídicas conocidas, según queda esquematizado en la **Figura 31**.



**Figura 31. Esquema resumen de los fragmentos amplificados sobre DNA genómico de *C.sinensis*.** Sobre el dibujo que muestra la organización estructural de los RTN tipo  *copia* se marca la posición de los distintos cebadores diseñados y utilizados en las amplificaciones. Debajo se presentan los distintos fragmentos generados situados bajo el esquema del elemento en relación a la zona que representan. A la izquierda, se resume la información referente a cada fragmento secuenciado, esto es: nombre (clon), si ha sido generado por PCR o i-PCR (Tipo PCR), la combinación de cebadores utilizada para generarlo (cebadores PCR), enzima de restricción en el caso de i-PCR (digestión genómico), el tamaño en pb (tamaño inserto) y si se conoce toda su secuencia (secuenciación total [T] o parcial [P]).

El inserto secuenciado con el nombre Insert1b20, representa una zona del elemento que va desde el teórico inicio de la 5'LTR hasta la región de la *gag*, en total 630 pb. El inserto secuenciado con el nombre id15 tiene 3980 pb e incluye toda la región interna del elemento. Los insertos MP2, MP10 y MP11 de 3496, 3568 y 3482 pb, respectivamente, representan zonas del elemento desde diferentes puntos de la LTR5' hasta la región que codifica por la RNAsa H. Los insertos MPA1 y MPB4, representan las regiones más a 3' del elemento, MPB4 de 1168 pb va desde una zona que codifica para la RT hasta una región de la LTR3', mientras que MPA1, más corto (371 pb) va desde una zona que codifica para la RNAsa H hasta la misma región de la LTR3'.

Los fragmentos generados todos por PCR directa sobre DNA genómico, una vez purificados fueron clonados en el vector pGEMT y posteriormente secuenciados al azar. Dado el elevado tamaño de los insertos generados, la secuenciación se realizó mediante *primer walking* esto es, mediante el diseño de diferentes cebadores internos a la secuencia conocida que son utilizados en la siguiente reacción para ir avanzando en la secuenciación. Esto genera una serie de secuencias más cortas que han de ensamblarse entre si para generar la secuencia entera original.

### **A.3.2 Análisis de las secuencias de CIRE1 clonadas del genoma de *Citrus sinensis***

#### **A.3.2.1 Comparativa de secuencias**

- **Zona de las LTR**

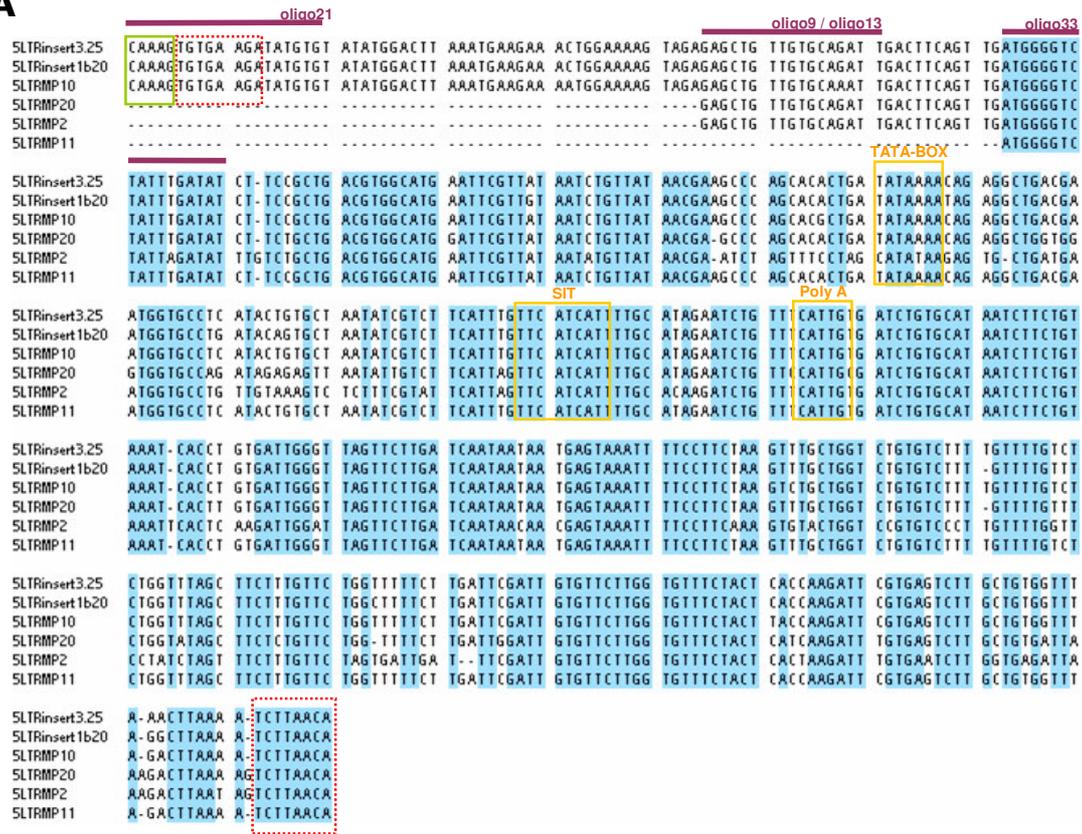
A continuación se muestra el alineamiento múltiple de todas las secuencias nucleotídicas correspondientes a fragmentos de LTR clonadas de *C. sinensis* (**Figura 32**). Se han secuenciado 3 clones que contienen regiones (parciales o totales) de LTR3' y 7 clones correspondientes a regiones (totales o parciales) de LTR5'.

El tamaño de las LTR completas es de alrededor de 460 pb. Este tamaño es similar al visto en otros RTN tipo *copia* descritos como son *Panzee* con una LTR5' de 372 pb y LTR3' de 383 pb (Lall.Maneesha et al., 2002), *Tnt1* cuyas LTR tienen 610 pb (Grandbastien et al., 1989) o *Tto1* con LTR de 576 pb (Hirochika, 1993), aunque el tamaño de las LTR no es un carácter distintivo de los RTN-LTR tipo *copia* (o *gypsy*) pues *BARE1* de cebada tiene unas LTR de 1829 pb (Manninen y Shulman, 1993).

A continuación del inicio asignado a las LTR de CIRE1 se observan las repeticiones invertidas cortas típicas de los extremos de las LTR. En este caso la repetición es de 3 pb (TGT...ACA), aunque si no tenemos en cuenta el cuarto nucleótido, la repetición invertida llega a ser de 8 pb (TGTTGAAGA ... TCTTAACA), siguiendo los estándares descritos para otros RTN (Neuvéglise et al., 2002).

## Resultados

**A**



**B**



**Figura 32. Alineamiento múltiple por ClustaIW de secuencias nucleotídicas correspondientes a fragmentos de LTR aislados de *C. sinensis*. (A) Fragmentos correspondientes a LTR5' (B) Fragmentos correspondientes a LTR3'. El sombreado indica identidad de secuencia del 100%. La ausencia de nucleótido en una determinada posición se indica con un guión. Se ha señalado con una línea horizontal la situación de las secuencias complementarias a cebadores presentes en esta región. Se ha recuadrado en verde la teórica diana de inserción (CAAAG) presente en la secuencia de la LTR3' de H4 y añadida a la secuencia del cebador en el diseño del oligo21. Se ha recuadrado en rojo las repeticiones cortas invertidas de los extremos de las LTR. Se ha recuadrado en naranja secuencias típicas de promotores presentes en la secuencia (caja TATA, SIT y sitio de poliadenilación). El origen de cada secuencia está descrito en la Figura 31.**

El contenido nucleotídico de las LTR es alto en A+T en un 63,3%, contenido similar al que encontramos en RTN como *Tnt1* (62,5% A+T) o *Del1* (64,2% A+T) y algo superior al de *BARE1* (51,8% A+T) (Manninen et al., 1993).

Si comparamos las secuencias de las LTR, tomadas una a una, presentan una identidad de secuencia nucleotídica máxima del 39% con las LTR de *Tnt1* (NºAcc.: X13777) o del 41% con las LTR de *Panzee* (NºAcc.: AJ000893), con lo que podemos afirmar que no son similitudes significativas (superior al 50%). Identidades similares presentan otros RTN tipo *copia* entre ellos al comparar sus LTR, por ejemplo las secuencias de las LTR de *BARE1* son idénticas en un 37% a *Tnt1* y en un 43% a *CopiaDM* (Manninem et al., 1993).



	MP20	Insert325	Insert1b20	MP10	MP11	MP2
<b>H4</b>	8.7 36/417 91.3	1.9 9/466 98.1	3.2 15/466 96.8	2.8 13/466 97.2	2.1 8/389 98.9	19 79/417 81
MP20		7 29/417 93	7.9 33/417 92.1	7.7 32/417 92.3	7.2 28/417 92.8	17.3 72/417 82.7
Insert325			1.3 6/466 98.7	1.3 6/466 98.7	0.3 1/389 99.7	18.3 76/417 81.7
Insert328			1.9 9/466 98.1	0.6 3/466 99.4	0.5 2/389 99.5	18.8 78/417 81.2
Insert1b20				2.2 10/466 97.8	1.5 6/389 98.5	18.8 78/417 81.2
MP10					1.3 3/389 98.7	18.3 76/417 81.7
MP11						19.3 75/389 80.7

**Tabla X. Comparación de las secuencias nucleotídicas correspondientes a fragmentos de LTR clonados de *C. sinensis*.** En verde, se indica el porcentaje de diferencias entre cada par de secuencias comparada. En negro se indica el número de nucleótidos diferentes respecto al número total de nucleótidos coincidentes. En morado, se indica el porcentaje de identidad para cada par de secuencias. Se ha resaltado en amarillo el análisis comparativo entre el fragmento del clon H4 que corresponde a LTR3' y el resto de fragmentos clonados correspondientes todos ellos a LTR5'. El origen de cada secuencia está descrito en la figura 31.

En la **Tabla X** se muestra el análisis comparativo de las LTR de los diferentes clones secuenciados en el estudio. El porcentaje de identidad entre cada par de secuencias comparada va del 99,7% al 80,7%. Al comparar la LTR3' (correspondiente al clon H4) con las LTR5' vemos que, sin tener en cuenta el clon MP2 (señalado con una flecha en la **Tabla X**), el porcentaje de identidad entre las LTR comparadas es superior al 90%, siendo en promedio del 96,5%. Este elevado nivel de identidad encontrado entre la

## Resultados

secuencia correspondiente a LTR3' y las secuencias correspondientes a LTR5', hace compatible el hecho de que se trate de fragmentos clonados del mismo RTN-LTR, pues se han encontrado porcentajes de identidad similares al comparar las dos LTR de RTN tipo *copia* como es *BARE1* (95,8%) (Suoniemi et al., 1996) o las dos LTR de RTN tipo *gypsy* como son *Cyclops* (95,9%) (Chavanne et al., 1998) o *Grande1-4* (95%) (Martínez-Izquierdo et al., 1997).

Al realizar la comparación de todos los clones secuenciados al azar que se corresponden con LTR5' entre ellos, vemos que en su mayoría difieren como máximo en un 2,2%. Para RTN como *BARE1* se ha visto que la identidad entre LTR de clones diferentes es del 92-94% (Manninen et al., 1993). Nuevamente, exceptuamos la LTR5' de MP2, que es la más diferente con un grado de similitud media del 81.5% con el resto.

Los diferentes clones contienen secuencias típicas de los promotores eucariotas. El análisis de la secuencia del promotor de la LTR se desarrolla en el **apartado A.4.1**.

- **Dominio Interno**

A continuación se muestra el análisis comparativo de las secuencias nucleotídicas de los diferentes clones aislados en el estudio correspondientes a regiones internas del RTN (**Tabla XI**). Vemos que el porcentaje de identidad entre las secuencias del dominio interno es superior al visto para secuencias de LTR, ya que va del 100% al 91%.

	MPA1	MPB4	MP20	Insert325	MP-id15	MP10	MP11	MP2
<b>H4</b>	100%	98%	94%	-	92%	98%	99%	91%
<b>MPA1</b>		99%	-	-	91%	-	-	-
<b>MPB4</b>			-	-	94%	-	-	-
<b>MP20</b>				95%	93%	94%	94%	93%
<b>Insert325</b>					93%	99%	99%	93%
<b>id15</b>						92%	92%	99%
<b>MP10</b>							98%	92%
<b>MP11</b>								92%

**Tabla XI. Comparación de las secuencias nucleotídicas correspondientes a regiones internas de RTN aisladas de *C. sinensis*.** Se indica el porcentaje de identidad entre cada par de secuencias comparada, calculado por el número de coincidencias en la comparación 1 a 1 respecto a la longitud en pb de la zona que solapan. El guión indica que las secuencias comparadas no se solapan. El origen de cada secuencia está descrito en la figura 31.

Mirando en detalle cada secuencia, vemos que **MP-id15** únicamente difiere en un 1% con la secuencia nucleotídica que comparte con **MP2**. Las secuencias de los clones **MP10**, **MP11** y **insert3.25** se solapan a lo largo de una región del dominio interno de 1856 pb donde la similitud de secuencia nucleotídica entre ellas es del 99%. Por último, la secuencia del clon **MP20** solapa en una región de aproximadamente 1000 pb con H4, id15, MP2, MP10 y MP11 con las que comparte una similitud de secuencia media del 94%, y la secuencia del clon **H4** presenta a su vez una elevada similitud de secuencia con MPA1, MP11, MP10 y MPB4 (identidad media del 98%) y es algo más diferente a las secuencias MP2 y MP-id15 (identidad media del 91,5%).

El nivel de acumulación de mutaciones en las secuencias aisladas, así como otras características como el contenido nucleotídico o los motivos conservados que presentan, se analiza en detalle a continuación. Destaca el hecho que es en la región correspondiente a la integrasa donde se concentran la mayoría de las mutaciones observadas y descritas en la **Tabla XII**. De la misma manera, observamos que las secuencias de MP2 y id15, que son las que presentaban menor similitud con el resto, son también las que presentan más mutaciones.

	Mutaciones que provocan cambios en la pauta de lectura correspondiente a RTN			Contenido Nucleotídico	Motivos funcionales
	Codones parada	Inserciones	Deleciones		
<b>Id15</b>	21	1 de 7pb	5 de 1pb, 1 de 14 pb, 1 de 6 pb	39,9% CG 59,9% AT	Gag, PR, INT, RT, RH
<b>MP2</b>	20	1 de 7pb	3 de 1pb, 1 de 14 pb, 1 de 6 pb	39,3% CG 60,7% AT	Gag, PR, INT, RT, RH
<b>MP10</b>	0	1 de 1pb	1 de 1pb	41,4% CG 58,6% AT	Gag, PR, INT, RT, RH
<b>MP11</b>	1	1 de 1pb	3 de 1pb	41,2% CG 58,3% AT	Gag, PR, INT, RT, RH
<b>Insert3.25</b>	0	0	0	40,6% CG 59,4% AT	Gag, PR, INT
<b>MP20</b>	6	0	0	40,1% CG 59,9% AT	Gag, PR, INT, RT
<b>H4</b>	0	0	0	41,7% CG 58,3% AT	RT, RH

Tabla XII. Descripción de las secuencias aisladas del genoma de *C. sinensis*

### **A.3.3 Establecimiento de la secuencia completa de un elemento: CIRE1-1**

Se han obtenido diferentes clones que presentan elevadas similitudes de secuencia entre ellos y entre regiones de RTN. Esta elevada identidad indica que todos los fragmentos obtenidos son porciones del mismo RTN-LTR tipo *copia* CIRE1.

## Resultados

Para analizar extensamente la secuencia de CIRE1, se han considerado dos clones que contienen largos segmentos de CIRE1 y que por si solos incluyen toda la longitud del RTN-LTR. Así, para determinar la secuencia completa del elemento CIRE1-1 se ha considerado el clon insert3.25, que define el extremo 5' del elemento (desde la LTR5' hasta una zona de la INT), y el clon H4, que define el extremo 3' (desde una zona de la INT. hasta la LTR3'). El elemento CIRE1-1 está flanqueado por una misma diana teórica duplicada pues, el cebador utilizado para amplificar el clon insert3.25 (oligo21) presenta en su secuencia la diana teórica de inserción que está igualmente adyacente a la LTR3' de H4. Ambos fragmentos comparten 5 pb que solapan sus secuencias en una región correspondiente a la INT y se encuentran intactos en el genoma (con una única pauta de lectura continua y sin interrupciones por codones de parada en la región codificante). La suma de ambas secuencias abarca toda la longitud del RTN-LTR y representa al elemento CIRE1-1 secuenciado en su totalidad (**Figura 33**).

### *Citrus sinensis* copia-like LTR retrotransposon, CIRE1-1

```
organism="Citrus sinensis" (sweet orange)
mol_type="genomic DNA"
tissue_type="leaf"
variety="Valencia late"
```

#### Sequence

```
tgtgaagata tgtgtatatg gacttaaatg aagaaactgg aaaagtagag agctgttgtg 60
cagattgact tcagttgatg gggctctatt gatatcttcc gctgacgtgg catgaattcg 120
ttataatctg ttataacgaa gccacgcaca ctgatataaa acagaggctg acgaatggtg 180
cctcatactg tgctaataatc gtcttcattt gttcatcatt ttgcatagaa tctgtttcat 240
tgtgatctgt gcataatctt ctgtaaatca cctgtgattg ggttagttct tgatcaataa 300
taatgagtaa attttccttc taagtttgcg ggtctgtgtc tttgttttg tctctggttt 360
agcttctttg ttctggtttt tcttgattcg attgtgttct tgggttttct actcaccaag 420
attcgtgagt cttgctgtgg tttaaactta aaatcttaac aagtggatc agagcctggt 480
gaatttcagc catgtctctt ccaaggcatg aaatcgaaaa attcaccatt ggtggtgatt 540
tctcgccttg gaagctgaaa atgaaagctc tccttgtgca tcaagggctt gaatctgccc 600
tggacgaaga agactagag gcctctacag gttctgggat tgatgacaaa aggagacaaa 660
tacagaatcg agctcatagc actctgattt tgagtcttgg tgaactcaatc ttgagagaga 720
tctctgaaga gaagacagcc ttaggaaatt ggaacaaggt tgaaacctctg tgtatgaaga 780
aatccctggc tcatagattg ttcttgaaga aaaggttgta cacatttcca atgagagaag 840
gggttacaat acaagatcac attgacactt ttaacaagat cattcttgat cttgaagggg 900
tggaaaaatg aaaaatctgt gatgaagaca aggcattctt cctgctgagc tctctacca 960
agtcttatga aggttttga gatacaatgc tctatgggag gaccactttg actcttgaag 1020
atgtcaaacg atccctcagc tcaaaggaga ttcagaagaa ctgtgagctt gaacaagta 1080
atggtgaggg gctgatggca agaacagaaa agaagaaga tcagaagaac aaaaatcaag 1140
gaaaaaggca tggaaagaac caagaaactg cagacaagaa gaaaaagaaa agaaagtgtc 1200
tctactgcag aaaggagggg cttacataaa gggattgctt tgagaagaag aagaagaaa 1260
gtcaagaaaa gtcaggagat gcagcagtag cctctgatga tggttctgat ggatatcaa 1320
gtgctgattt acttgtggcc tcaaacagca atacaaaagg tcagtgggtc atagactctg 1380
gggtctcttt tcatttatgt cctgaaaaaa cctattttta caaatatgaa gctggtgatg 1440
gtggttagagt cttaatggga aacaataatg tgttaatat tgttggattt tggttctgta 1500
aaagatcaag atgtttgatg gaactattaa gatctctaca tgaggtgagg catgctccta 1560
ggctaagag aaatttaatc tctttaggca tgctagacag cttagggtac ttctttaaat 1620
caagaattgg tggtttagag gtaagaaaag ggactgagat agtaatgaag ggggttaacg 1680
agaatggcct atatgtattg caaggctcat cagtgcctgt tcaagaagga gtttctgctg 1740
tttctgagga agatagaacc aagtatggc accttaggct tggatcatatg agtatcaag 1800
ggctgcaaga gcttagtaaa caagggtcac ttggtgggga tagaatccag caactggagt 1860
tttgtgaaaa ctgtatttt ggcaaatcac acaggtccaa gtttaacaaa ggggagcata 1920
tgtctaaca agttttggac tatgcacata ttgaccattg gggacctgct caggtccctt 1980
ccttgagtgg tggcagatat ttcatgtcac ttatgatga ttactcaagg aaagtatgga 2040
tctacatact caagatcaaa gatcaagctt tggaaaaatt caaggtctgg aatcatttgg 2100
tagaaaaatc gctgacttc aagctgaaat gcttgaggac tgataatggc ttggaatttt 2160
gtagcaaaatg atttgaagaa tattgcaaaa aacatggcat aaagaggcac aaaacagtga 2220
ggttcaccct ccagcagaat ggcttggctg agaggatgaa caggaccttg gtggataaaa 2280
ccagatgtat cgtgataaat tcaaagcttc caagaagttt ttgggcagaa gctgtcaata 2340
cagcaagtta tttagttaac agaagcccat cttctgcaat tggcttcaaa acaccagaag 2400
aattgtggaa tggaaaacca gcaaaactacc aaaaatctcag agtttttggg tgcctgctt 2460
acctcaacat caatcaaggc aagctggaag caagggcctt gaaaggagt cttgttggat 2520
atcctgatgg ggttaagggg tacaaaatct ggttaagga tcaaggaaaa tgcatagtca 2580
gcagggatgt tgtgtttcat gaatctgtat tgctgaaaga aagtgcagaa catgatgctg 2640
gactgcaaga caatccagca gctaacaac gatctggatc tgaacttcc aaggttaatg 2700
tggagctatt gactgataaa agcagtgaaa aagaagcagc ttctgatgat gagagggcca 2760
```

cagctgaaag	tgaggaacat	gaggtatcag	aattgccaca	agctgactta	cagaactatc	2820
agttagctag	agatagagtt	agaaggggaag	tcagagcacc	agtgaggat	ggctatgctg	2880
acttaattgc	ctatgcacta	ttgtgtgctg	atgaagtgc	aattgaagag	cctgctaact	2940
tctctgaagc	catggaagt	gtgcaactgtg	acaagtggct	tgaggcaatg	caagatgaaa	3000
tggagtcact	tcaaaagaaat	cagacctgga	cacttatacc	aaatccagga	aacaagaggt	3060
tgatcaactg	taagtggatt	ttcaagagga	atgaaggaat	tccagatgtt	gaacctcca	3120
aatataaagc	aaggctagtg	gctaggggtt	ttcacacagag	ggaaggtgta	gatttcaatg	3180
aaattttttc	tcctgtagtg	aagcacagtt	caataagaat	cttattagcc	atggttagctt	3240
tacttgactt	ggagctggag	cagatggatg	taaaaacagc	cttccttcat	ggaatcttg	3300
aagaacaaat	attgatggca	cagcctgaag	gatttgagtg	taaggataag	gaggattatg	3360
tgtgtctgct	gcataagtcc	ttgtatggac	tcaagcaatc	ccctcgacag	tggtacagaa	3420
ggtttgatga	ttttatggtc	tcaaagggct	atcgcagaag	cagatagac	agctgtgctt	3480
acttttggtg	atcagatcaa	gggtggagttg	cttatttctg	tctgtatgtt	gacgatatgt	3540
taattgcaag	caaatacaag	tcagagatag	agaggttaaa	gaatctgta	aaggctgagt	3600
ttgaaatgaa	ttttatggc	aatgcaaaaa	gaatccttgg	gatggacatt	atcagagata	3660
gatctgctgg	cactttattc	ttgagtcaag	ggaatatat	caagaaagt	ctggagaggt	3720
ttgaaatgca	agactgtaaa	cctgttcaga	ctccacttgg	tccacaattc	aagttgtcag	3780
cagctacaac	atcagaagat	gagctcaga	tgaatgaatt	cccatatgca	caggctgtgg	3840
ggagcctgat	gtatgctatg	gtatgcacca	ggtcagacat	agcttatgca	gtaatgttg	3900
tcagcagata	tctcagttgt	ccaggcaagg	ttcactggaa	tgctgtcaa	tgataaatga	3960
ggttatctaaa	gggttcttct	acttgtggtt	tgctgtatgg	gaagactaag	agtgacaaaa	4020
ttgaagtcac	gggttttgta	gactctgact	ttgccggaga	tcttgacaga	agaaagtcta	4080
catcaagata	tatgtttgtg	ttaaataagt	gcctgatcag	ctggaagtca	tctcttcagt	4140
ctggtggtgc	tttatcctca	actgaagctg	agttcattgc	gactacagag	gcagtaagg	4200
aggccatgtg	gctgaggggc	ttgcttaatg	agctatggtt	aaatcagaaa	atgttccagg	4260
ttttttgtga	taaccaaaa	gccatacact	tggtgaagaa	tcaaatgtat	catgaaagaa	4320
ccaagcacat	tgatgtgaag	ctccagttta	tcagagatga	agttggaaag	ggtacagtgg	4380
ttgtgtccaa	gatccatacc	agtgatgaatc	cagctgatgc	tctaacaaaa	tacttccca	4440
ctgcaaaaat	tgagttctgt	gtgaatctaa	tgggcatcat	gccccaaatc	aactaaattc	4500
aagaagatga	attgtagttg	aatgaaggtt	tttaggtgct	gtattattga	agggatcaac	4560
tgaatcaact	gcaaggtgga	gattgtgaag	atatgtgtat	atggacttaa	atgaagaaac	4620
tggaaaagta	gagagctggt	gtgcagattg	acttcagttg	atggggtcta	tttgatatct	4680
tcgctgacg	ttgcatgaat	ctgttataat	ctgttataac	gaagcccagc	acactgatat	4740
aaaacagagg	ctgacgaatg	gtgcctcata	ctgtgctaata	atcgtcttca	tttgttcatc	4800
attttgcata	gaatctgttt	cattgtgatc	tgtgcataat	cttctgtaaa	tcacctgtga	4860
ttgggttagt	tcttgatcaa	taataatgag	taaattttcc	ttctaagttt	gctggctgtg	4920
gtctttgtgt	ttgtctctgg	tttagcttct	ttgttctggt	tttcttgat	tcgattgtgt	4980
ccttgggtgt	tctactcacc	aagattctgtg	agtcttgctg	tggtttagac	ttaaaatctt	5040
aaca	5044					

LTRs 1..461 ; 4584..5044

"putative minus strand primer":PBS 464..475

"protease motif" 1374..1382

"integrase motif" 1770..1874

"reverse transcriptase motif" 3276..3299

"reverse transcriptase motif" 3519..3549

"RNase H motif" 4037..4052

"putative poliprotein" 492..4493

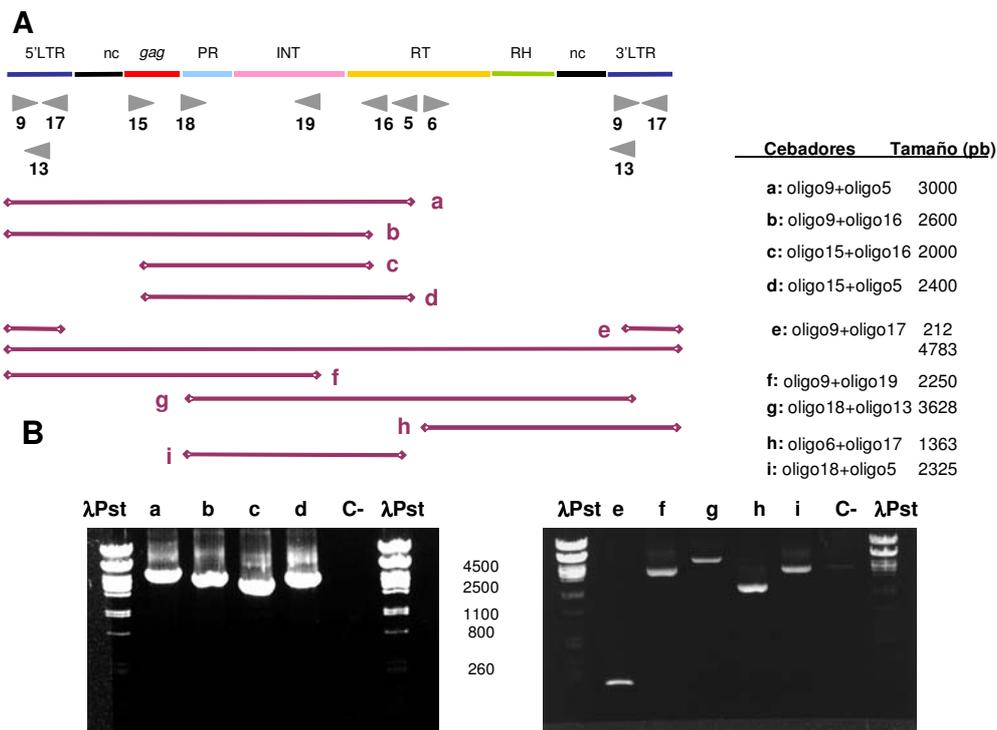
Translation="MSLPRHEIEKFTIGGDFSLWKLKMKALLVHQGLESALDEEDLEASTGSGIDDKRRQIQNRAHSTLILSL  
GDSILREISEEKTALGIWNKVETLCMKKSLAHLRFLKRLRYTFMREGVTIQDHIDTFNKIILDLEGVENVKICDEDKAFFL  
LSSLPKSYEGFVDTMLYGRFTLTLEDVKASLSSKEIQKNCLETSSNGEGLMARTEKKKDQKNKNQKGGHGNQETADKKKK  
RKCFYCRKEGHIYRDCFEKKKKESQEKSGDAAVASDDGSDGYSADLLVASNSNTKGQWVIDSGCSFHLCPKTLFYKYEAV  
DGGRVLMGNNNVNCNIVGIWFCRKRSLMELLRLSLHEVRHAPRLKRNLI SLGMLDSLGYFFKSRIGGLEVRKGTETVMKGVNE  
NGLYVLQSSVVPVQEGVSAVSEEDRTKLWHLRLGHMSIKGLQELSKQGLLGGDRIQQLFCENCIFGKSHRSKFNKGEHMSK  
QVLDYAHIDHWGPAQVPSLSGGRYFMSLIDDSYRKYWIYILKIKDQALEKFKVWVSLVENQSDFKLCLRTDNGLEFC SKVF  
EEYCQKHGIKRHKTVRFTPQONGLAERMNRITLVDKTRCMLINSKLPFSWAEAVNTASYLVNRSFSAIGFKTPEELWNGKP  
ANYQNLRVFGCPAYLHINQKLEARALKGVFVGPDPGVKGYKIWCKDQKCIVSRDVFVHESVLLKESAEHDAQLQDNPAAN  
KRSGETSKVNVVELLTDKSSKEAASDDERATAESEEHEVSELPQADLQNYQLARDVRREVVRAPVRYGYADLIAYALLCAD  
EVTIEEPANFSEAMESVHCDKWLEAMQDEMESLQRNQTTWTLIPNPNKRLINCKWIFKRNREGIPDVEPPKYKARLVARGFTQ  
REGVDFNEIFSPVVKHSSIRILLAMVALLDLELEQMDVKTAFLHGNLEEQILMAQPEGFECCKDEDYVCLLHKSlyGLKQSP  
RQWYRRFDDFMVSKGYRRSRYSVYFVGGSDQGGVAYLLLYVDDMLIASKYKSEIERLKNLLKAEFEMKDLGNAKRILGMDI  
IRDRSAGTLFLSQGKYIKKVLERFEMQDCKPVQTPVLPQPKLSAATSEDESQMNFPYAQAVGSLMYAMVCTRSDIAYAVS  
VVSRYLSCPQKGVHNAVKWIMRYLKGSSCTGLLYGKTKSDKIEVMGFVDSDFAGDLDRRSTSRYMVFLNSCLISWSSLSQS  
GGALSSTEAEFIATTEAVKEAMWLRGLLNLWLNQIVQVCDNQNAIHLVKNQMYHERTKHIDVKLQFIRDEVGKGTVVVS  
KIHTSVNPADALTKSLPTAKFEFCVNLGMIPKSN"

Figura 33. Secuencia de CIRE1-1

## Resultados

### A.3.4 Integridad de los elementos CIRE1 en el genoma de *C. sinensis*

Para comprobar si la mayoría de las copias de CIRE1 estaban íntegras en el genoma de *C. sinensis* se ideó un experimento de PCR donde se utilizaban diferentes combinaciones de cebadores diseñados a partir de la secuencia de CIRE1-1 de la que se dispone. Se diseñó de manera que las diferentes parejas de oligos amplificaran diferentes partes del elemento y que en conjunto lo cubrieran por completo. De esta manera se quiere ver si la mayoría de las copias de CIRE1 presentan deleciones o inserciones importantes en su secuencia. En la **figura 34** se resumen las combinaciones de cebadores utilizadas y el resultado obtenido para cada una. Efectivamente, en todos los casos se observa una banda única, clara y del tamaño esperado en la amplificación, indicando que estos elementos están en su mayoría enteros en el genoma. En la combinación de cebadores: oligo9 + oligo17, al ser ambos complementarios a secuencias de la LTR y estar encarados uno al otro, se esperan dos fragmentos (uno de 212 pb en cada una de las dos LTR) y otro de 4783 pb (que incluye parte de las dos LTR y la región incluida entre ellas) No obstante, sólo se observa el fragmento más corto de los dos posibles que podrían amplificarse (**Figura 34, carril e**), probablemente debido a una mayor eficiencia en la amplificación por PCR de fragmentos cortos frente a los largos.



**Figura 34. Integridad de los elementos CIRE1 en *C. sinensis*.** (A) Esquema del elemento CIRE1. Se ha indicado la posición de los diferentes cebadores utilizados en la PCR y el tamaño de banda que se espera amplificar. La secuencia de los cebadores y las condiciones de la amplificación se describen en Material y Métodos. La PCR se ha realizado a 25 ciclos con Ta de 52-54°C. (B) Electroforesis en gel de agarosa al 1% del resultado de la amplificación por PCR de las distintas regiones del elemento descritas en (A). Carril λPst, marcador de peso molecular λPst, se indican algunos tamaños en pb.

### **A.3.5 Conservación de los dominios funcionales en CIRE1**

El elemento completo CIRE1-1 secuenciado tiene un tamaño de aproximadamente 5 kb y contiene dos LTR de unos 460 pb con un contenido de A+T medio de 63,3%. El contenido nucleotídico del dominio interno es alto en A+T en una proporción del 59% y presenta una única ORF de 3999 nucleótidos. Comparaciones entre las ORF de *Tnt1* de *Nicotiana tabacum* (NºAcc.: P10978) y CIRE1-1 revelan que ambos comparten las regiones estructurales y funcionales comunes a los RTN-LTR tipo  *copia*. Se observa un elevado grado de similitud a nivel aminoacídico entre las posibles proteínas codificadas por las ORF de CIRE1-1 y *Tnt1*: similitudes del 36, 38, 48 y 53% para los dominios de *gag*, PR, IN y RT, respectivamente. Estos valores tienen en cuenta la gran similitud de diversos motivos altamente conservados en RTN y retrovirus.

Se ha realizado el alineamiento de la secuencia aminoacídica deducida para CIRE1-1 con las secuencias aminoacídicas de otros RTN-LTR tipo  *copia* ya descritos y totalmente caracterizados cuya secuencia entera está en las bases de datos. Así, las secuencias del dominio interno de *BARE1* de cebada (NºAcc.: Z17327), *RIRE1* de arroz (BAB21127), *Ta1-3* de Arabidopsis (S05465), *Tlc1* de tomate (AAK29467), *Tnt1* de tabaco (P10978), *Tto1* de tabaco (BAA11674.1) y *CopiaDM* de *Drosophila melanogaster* (P04146) se han sumado a la de CIRE1 para establecer las diferentes regiones del elemento y el grado de conservación de los motivos de referencia de RTN tipo  *copia* comparados con los otros elementos.

En la **Figura 35** se resumen los motivos estructurales y funcionales de CIRE1-1. Se observa la presencia de aminoácidos altamente conservados correspondientes a las regiones más conservadas de cada dominio proteico en todas las secuencias aminoacídicas de RTN alineadas, tal y como se describe en detalle a continuación:

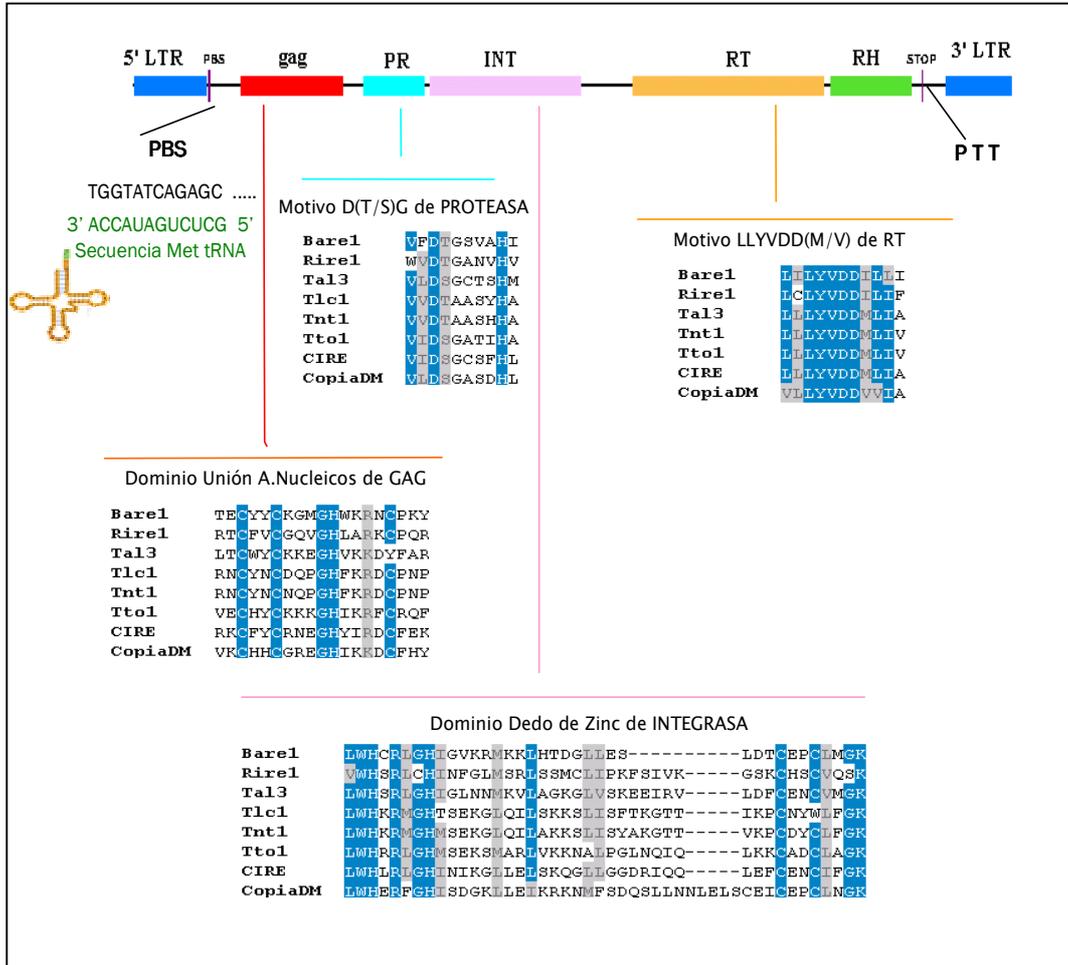
- Región GAG

Esta región no presenta similitudes significativas con ningún RTN aunque, podemos determinar claramente en el dominio GAG de CIRE1-1 la posición de un motivo de unión a ácidos nucleicos perfecto: CX<sub>2</sub>CX<sub>3</sub>GHX<sub>4</sub>C, presencia que sí comparte con el resto de retroelementos, como puede verse en la **Figura 35**.

- Región Proteasa

En la **Figura 35** se muestra el alineamiento del dominio PR alrededor del motivo más conservado D(T/S)G (centro activo de la proteasa), confirmando el alto grado de conservación de este motivo en CIRE1-1. El resto de secuencia del dominio proteasa, fuera del centro activo, es mucho más diversa.

## Resultados



**Figura 35. Motivos estructurales y funcionales de CIRE1-1, comparados con los de RTN-LTR tipo copia.** El dominio interno de CIRE1-1 de aproximadamente 4123 pb está flanqueado por dos LTR unos 460 pb aproximadamente. En el esquema, las cajas situadas dentro del dominio interno del elemento representan los dominios aminoacídicos que codifican: GAG; PR proteasa; INT integrasa; RT transcriptasa inversa; RH RNasa H. La secuencia del hipotético PBS (*primer binding site*) sitio de unión del cebador para la síntesis de la primera cadena de DNA, es complementaria al extremo 3' de la secuencia del tRNA para metionina. Se indica la situación del PPT (*polypurine tract*). Se muestra el alineamiento múltiple por ClustalW de las secuencias aminoacídicas deducidas para varios dominios de CIRE1 con las regiones correspondientes de otros RTN presentes en las bases de datos. Están sombreados aquellos residuos más conservados en las distintas secuencias. Se indican algunos de los motivos más altamente conservados de cada dominio: región de unión a ácidos nucleicos (CX<sub>2</sub>CX<sub>3</sub>GHX<sub>4</sub>C) del dominio de GAG, la altamente conservada secuencia D(T/S)G de la proteasa, el motivo de dedo de zing (HX<sub>3-7</sub>HX<sub>20-32</sub>CX<sub>2</sub>C) de la integrasa o el casi invariante motivo LLYVDD de la RT.

- Región Integrasa

Como se ha comentado anteriormente, la región del elemento que codifica la integrasa es la que presenta más mutaciones en los diferentes clones aislados y secuenciados de CIRE1. Aun así, aquellos motivos considerados como los más conservados en los diferentes RTN también están presentes en CIRE1-1. Así, presenta un motivo de dedo de zinc cuya secuencia  $HX_{3-7}HX_{20-32}CX_2C$  está perfectamente conservada tal y como vemos en la **Figura 35**. Otro de los motivos más conservados del dominio INT es el  $PQX_2GVXER$ , éste no es un dominio que presenten todos los RTN (Vicent, 1995) y en el caso de CIRE1-1, de los 6 residuos más conservados presenta 5. El motivo  $DD_{35}E$  que está perfectamente conservado en integrasas retrovirales y de RTN y es esencial para su actividad enzimática (Suoniemi et al., 1998), también se observa en CIRE1-1 aunque no perfectamente conservado.

- Región RT

Se conoce que la región que codifica la RT es el dominio menos variable de los RTN (Doolittle et al., 1989).

El dominio RT suele contener 7 cajas que corresponden a las regiones más conservadas del dominio, separadas por secuencias de tamaño variable y menor conservación. En la **Figura 36** se muestra el alineamiento de la secuencia aminoacídica de las 7 cajas de CIRE1-1 con regiones equivalentes de otros RTN como son *Tnt1* de tabaco (como representante del grupo *copia*) y *Del1* de *Lilium henry* (como representante del grupo *gypsy*). Este tipo de estructuración está bastante conservado entre los RTN, tanto de tipo *copia* como *gypsy*, pero con un carácter distintivo entre los dos tipos de RTN, con menor divergencia entre dos secuencias de dos RTN del mismo tipo que entre dos RTN de los dos tipos (**Figura 36**). Además, los RTN de tipo *copia* presentan la caja 6 más corta y la secuencia entre las cajas 6 y 7 más larga y conservada (Vicent, 1995). Como CIRE1-1 cumple, efectivamente, con esas características, podemos confirmar de nuevo que este elemento pertenece al grupo *copia*.



**Figura 36. Alineamiento múltiple de las secuencias aminoacídicas del dominio RT.** Las secuencias de *Del1* (1510387A) RTN tipo *gypsy* de *Lilium* y de *Tnt1* (P10978) RTN tipo *copia* de tabaco, se han unido a la de CIRE1-1 en la comparación. La definición de las cajas del dominio de la RT está basada en Xiong y Eickbush (1990) según Vicent, 1995.

## Resultados

Al realizar comparaciones mediante BLASTX con las bases de datos (GeneBank) de la secuencia aminoacídica de CIRE1-1 correspondiente a toda la región de la RT conseguimos identidades del 56% con *Ta1-3* de *Arabidopsis*, 53% con *Tnt1* y 51% con *Tto1*, ambos RTN-LTR de tabaco. El porcentaje de identidad aumenta hasta un 88% con *Ta1-3* si únicamente tenemos en cuenta la región más conservada correspondiente a la caja 5.

En la **Figura 35** se muestra el alineamiento de la región de la RT correspondiente a parte de la caja 5 donde se sitúa el motivo DD. Este motivo se corresponde con el centro activo y está presente en todas las RT e incluso en algunas polimerasas (Velasco et al., 1992). Se observa que la secuencia de CIRE1-1 para esta región es exactamente igual a la que presentan *Ta1-3*, *Tnt1* y *Tto1*.

- Región RH

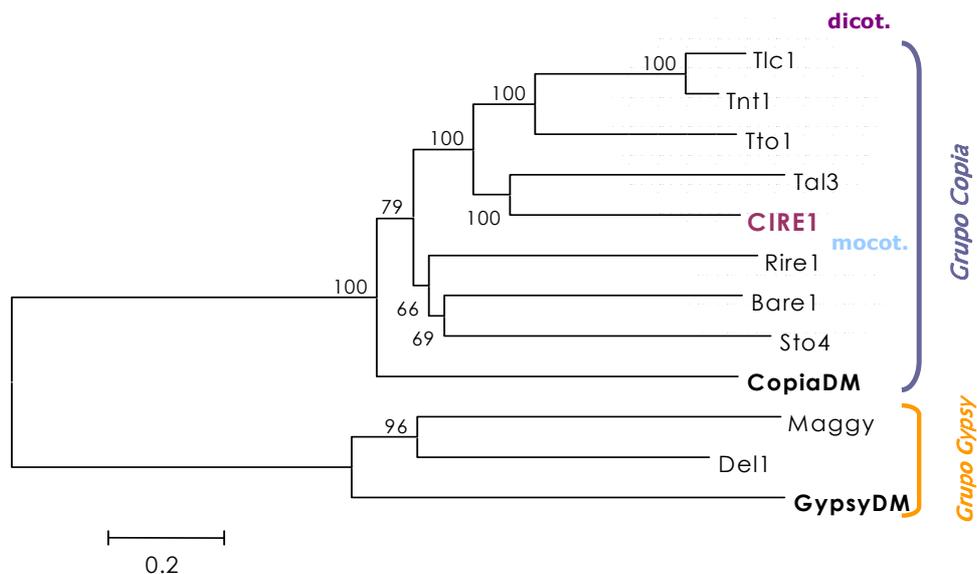
Se han definido para la región de la RNasa H dos motivos altamente conservados en los diferentes RTN, motivo1: RTKHI(D/E) y motivo 2: AD(I/M/F/L)LTK (Pearce et al., 1999). CIRE1-1 presenta el motivo 1 perfectamente conservado y mantiene 5 residuos conservados de los 6 del motivo 2. En lo que respecta al motivo TDAS, observamos que CIRE1-1 no lo tiene totalmente conservado (VDSD) pero curiosamente presenta exactamente la misma secuencia que *copiaDM* de *Drosophila melanogaster* tal y como se muestra en la **Figura 37**.

Gypsy	LTTDASASGI
Del1	VYTDASLAGL
Tnt1	GYTDADMAGD
Copia	GYVDS DWAGS
CIRE1	GFVDSDFAGD

**Figura 37. Alineamiento múltiple de las secuencias aminoacídicas del motivo más conservado del dominio RH (TDAS).** Las secuencias de *gypsyDM* (NºAcc.: P10401), *Del1* (Nº Acc.: 1510387A) RTN del grupo *gypsy*, y de *copiaDM* (P04146) y *Tnt1* (P10978) RTN del grupo *copia*, se han unido a la de CIRE1-1 en la comparación.

### A.3.6 Análisis filogenético de CIRE1

Como ya se ha comentado anteriormente, CIRE1 presenta una ordenación de los dominios GAG-PR-INT-RT-RH propia de los RTN de tipo *copia* y también presenta una conservación de motivos propia de elementos del grupo *copia* como se ha visto para la RT. Aún así, para confirmar la ubicación de CIRE1 dentro de los RTN-LTR de tipo *copia* hemos realizado un análisis filogenético basándonos en el alineamiento de las secuencias aminoacídicas correspondientes a todo el dominio interno de diferentes RTN. El resultado se muestra en la **Figura 38**.



**Figura 38. Árbol filogenético basado en el alineamiento de las secuencias aminoacídicas de todo el dominio interno de CIRE1 y otros RTN.** Árbol filogenético tipo *Neighbour joining* (NJ) generado con el programa MEGA2. La divergencia entre las secuencias en unidades de distancia genética viene definida por la longitud de las ramas horizontales. Están indicados en cada bifurcación los valores de bootstrap  $\geq 50\%$ . Se indica a la derecha de la figura las dos ramas de RTN tipo *copia* o *gypsy* en que quedan agrupadas las secuencias. Secuencias incluidas en la comparación: RTN del grupo copia: *Tnt1* secuencia de tabaco (P10978), *Tal3* secuencia de *Arabidopsis thaliana* (S05465), *Tto1* secuencia de tabaco (BAA11674.1), *Tlc1* secuencia de tomate (AAK29467), *Sto4* secuencia de maíz (T17429), *Rire1* secuencia de arroz (BAB21127), *BARE1* secuencia de cebada (Z17327) y *CopiaDM* secuencia de *Drosophyla melanogaster* (P04146). RTN del grupo Gypsy: *Maggy* secuencia de *Magnaporthe grisea* (AAA33420), *Del1* secuencia de *Lilium henryi* (1510387A), y *GypsyDM* secuencia de *Drosophyla melanogaster* (P10401).

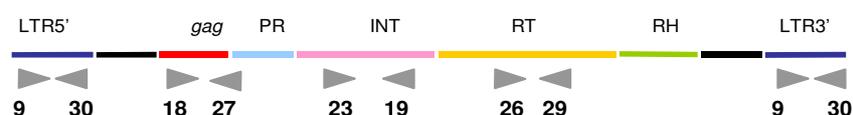
El árbol filogenético resultante muestra dos grandes ramas distribuyéndose las secuencias en dos grandes grupos correspondientes a RTN-LTR tipo *copia* y RTN-LTR tipo *gypsy*, división apoyada por valores de bootstrap muy elevados. Los elementos *copiaDM* y *gypsyDM* aislados de *Drosophila* aparecen, entre los RTN-LTR de plantas, como *outgroup* en sus correspondientes grupos. CIRE1 se ubica claramente dentro del grupo *copia*, confirmando todos los resultados anteriores.

Destacar la aparición en el árbol filogenético, dentro del grupo *copia*, de dos subagrupaciones. En la primera, se incluyen todos los RTN de dicotiledóneas analizados, como *Tnt1* y *Tto* de tabaco, *Tlc* de tomate, *Ta1-3* de *Arabidopsis* y CIRE1 de naranjo (**dicot, Figura 38**); mientras que el segundo bloque agrupa todos los RTN de monocotiledóneas analizados, como *RIRE1* de arroz, *BARE1* de cebada y *Sto4* de maíz (**mocot, Figura 38**).

## Resultados

### A.3.7 Estimación del número de copias de CIRE1

Se ha estimado el número de copias de CIRE1 por genoma haploide en *C. sinensis* mediante reconstrucción genómica a partir del análisis por hibridación *Slot-blot*, tal y como se ha descrito anteriormente (**Apartado A.1.4 y A.2.4 de Resultados**). La sonda utilizada se obtiene por PCR (no es sonda clonada) pero en este caso se utilizan cebadores altamente específicos de la familia de la que queremos conocer el número de copias, de esta manera sólo se evaluarán los elementos de la familia de CIRE1. A partir de la secuencia conocida de CIRE1-1, se diseñaron diferentes oligonucleótidos complementarios a distintas regiones del elemento, que se utilizan como cebadores para generar las diferentes sondas del experimento. Se utiliza la combinación oligo9+oligo30 para generar un fragmento de 389 pb de la LTR, la combinación oligo18+oligo27 para generar un fragmento de 297 pb del gen *gag*, la combinación oligo23+oligo19 para generar un fragmento de 470 pb de la INT y finalmente la combinación oligo26+oligo29 para generar un fragmento de 403 pb de la RT. La secuencia de estos oligonucleótidos y las condiciones de la amplificación se indican en Material y Métodos.



	Nº copias - Exp1	Nº copias - Exp2	Nº copias - Exp3	NÚMERO COPIAS TOTAL CIRE1
LTR	5740,1 ± 563,2	4686,2 ± 343,6	3494,7 ± 205,2	<b>4640,3 ± 437,45</b>
GAG	2551,4 ± 235,1	2087,5 ± 114,7	1962,2 ± 78,6	<b>2201,7 ± 121,75</b>
INT	3255,1 ± 307,7	2382,1 ± 89,8	1257,3 ± 93,7	<b>2298,2 ± 190,02</b>
RT	3334,1 ± 304,4	2612,1 ± 172,0	2077,2 ± 180,2	<b>2674,4 ± 245,63</b>

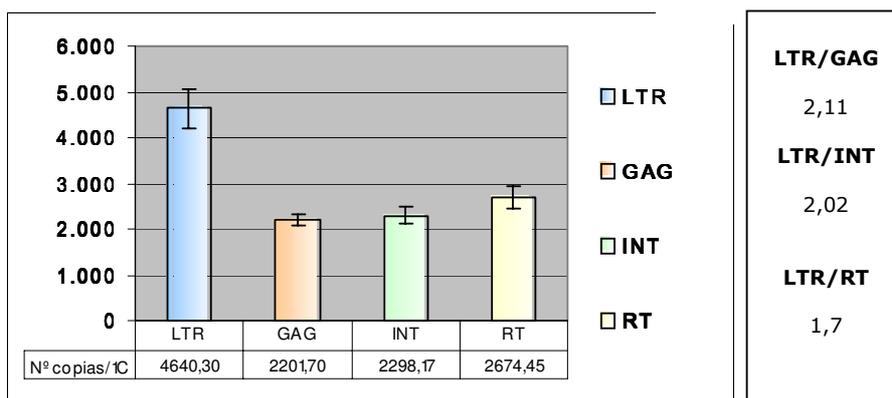
**Tabla XIII. Número de copias del RTN-LTR CIRE1 en el genoma de *Citrus sinensis*.** Tabla que recoge los valores estimados en las distintas réplicas del experimento de *Slot-blot* y el promedio ± error estándar calculado.

Bajo las condiciones concretas de hibridación y lavado de los filtros (ver Material y Métodos) en que se han realizado los experimentos, se ha estimado que hay unas 2200 copias de CIRE1 en el genoma de *C. sinensis* (**Tabla XIII**). El gen *gag* en su conjunto es el más divergente dentro de los dominios codificantes de los RTN. Normalmente, el número estimado de copias de un RTN usando una sonda del gen *gag*, representa el menor valor y el más ajustado al valor real frente a sondas de otros dominios del RTN,

tal y como se observa en los resultados obtenidos (2.202 copias, Tabla XIII). En cambio, el número de copias estimado con una sonda del dominio de la RT, mucho más conservada entre los distintos RTN, resulta en un número de copias mayor (2674 copias, Tabla XIII), sobreestimando el número de copias real de CIRE1. En cualquier caso, se observa un número de copias equivalente para las diferentes regiones del elemento CIRE1, indicando que estos se encuentran en su mayoría íntegros en el genoma, tal y como se había apuntado en el **Apartado A.3.4.**

Los elementos CIRE1 constituyen por tanto el 22% de los aproximadamente 10.000 elementos tipo *copia* de *C.sinensis* estimados en el Apartado A.1.4 de Resultados. Teniendo en cuenta el tamaño de CIRE1 (5 kb), las 2.200 copias por genoma haploide representan el 2,9% del genoma en esta especie.

Los RTN-LTR completos presentan dos LTR por cada dominio interno como consecuencia de los mecanismos de replicación y transposición. En este caso, el número de copias de las LTR en relación al dominio interno ha de ser similar a 2, tal y como se ha descrito para el RTN *Grande1* en *Zea* (García-Martínez y Martínez-Izquierdo, 2003). La aparición de solo-LTR se produce como consecuencia de recombinaciones homólogas intracromosómicas que provocan la pérdida del dominio interno y de una de las LTR del RTN. En estos casos, la relación LTR:dominio interno pasa a ser superior a 2, tal y como se ha descrito para *BARE1* de *Hordeum*, siendo el ejemplo más espectacular de exceso en solo-LTR (Vicient et al., 1999). Tal y como queda representado en la **Figura 39**, en el caso de CIRE1 la relación LTR:dominio interno es muy próxima a 2 en todos los casos, descartando una presencia significativamente importante de solo-LTR en el genoma de naranjo.



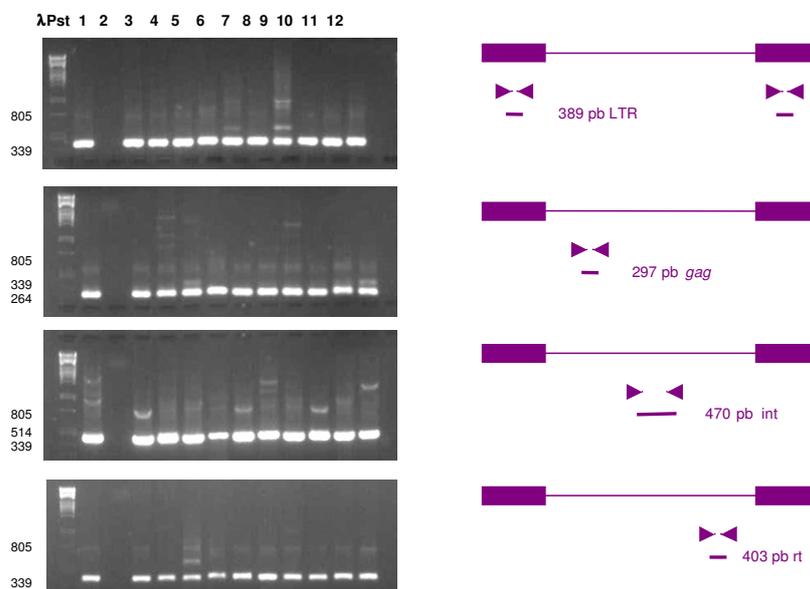
**Figura 39. Representación gráfica del número promedio de copias del RTN-LTR CIRE1 calculado para diferentes regiones del elemento.** Se indica el promedio de número de copias por genoma haploide (nº copias/1C) estimado en el experimento de *Slot-blot* utilizando como sonda múltiples DNA originados por PCR correspondientes a los dominios de la LTR, de la GAG, la INT y la RT. A la derecha, se indica el número de copias de CIRE1 relativo estimado por la LTR en relación a las diferentes regiones del dominio interno.

## Resultados

### A.3.8 Presencia de CIRE1 en el genoma de cítricos y otras especies vegetales

#### A.3.8.1 Amplificación selectiva de CIRE1 por PCR en la familia Rutaceae

Se decidió buscar en qué especies de cítricos está presente CIRE1 mediante el diseño de parejas de cebadores de diferentes partes del elemento que permiten su amplificación selectiva por PCR. Se utilizaron las mismas combinaciones de cebadores que las usadas para determinar el número de copias de CIRE1 en *C. sinensis*, tal y como se detalla en el apartado anterior. Por tanto, se espera amplificar una región de 389 pb de la LTR del RTN, una región de 297 pb del gen *gag*, una región de 470 pb de la integrasa y finalmente una región de 403 pb correspondiente a la RT. CIRE1 se buscó en 10 especies diferentes del género *Citrus* y 2 especies de géneros próximos de la familia Rutaceae. Las especies utilizadas se describen ampliamente en el **apartado B** de esta tesis.



**Figura 40. Resultado de la amplificación selectiva de CIRE1 en el genoma de distintas especies del género *Citrus* y especies de géneros próximos.** Imagen de la electroforesis en gel 2 % de agarosa de la amplificación sobre DNA genómico de parte de la LTR, gen *gag*, *int* y *rt* respectivamente. Carril λPst, marcador de peso molecular λPst, indicando algunos tamaños en pb. Carril 1, control positivo (DNA de *C. sinensis*); carril 2, control negativo de la reacción de PCR (sin DNA), carril 3, *C. clementina*; carril 4, *C. grandis*; carril 5, *C. ichangensis*; carril 6, *C. limettioides*; carril 7, *C. limon*; carril 8, *C. macroptera*; carril 9, *C. medica*; carril 10, *C. unshiu*; carril 11, *Fortunella*; carril 12, *Poncirus*. A la derecha de la figura se ha representado esquemáticamente el RTN con la localización de los diferentes cebadores utilizados y el tamaño de banda que se espera amplificar en cada región. La PCR se ha realizado a 30 ciclos con  $T_a$  de 52°C. Las secuencias de los cebadores se describen en Apartado 7 de Material y Métodos.

Tal y como se muestra en la **Figura 40**, el RTN CIRE1 se ha amplificado en todas las especies investigadas y su secuencia está suficientemente conservada en las especies de la familia Rutaceae analizadas como para poder ser amplificada en todos los casos utilizando cebadores diseñados a partir de la secuencia del elemento aislado de *C. sinensis*

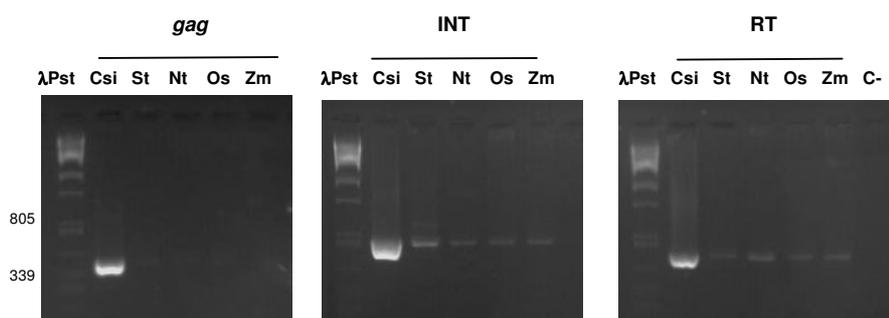
**A.3.8.2 Amplificación selectiva de CIRE1 por PCR en especies vegetales fuera de la familia Rutaceae**

Se quiso averiguar también si como en Rutaceae, CIRE1 está presente también en otras especies vegetales más lejanas con la idea de seguir la historia evolutiva del RTN. Para ello se ha utilizado DNA genómico de hoja de tabaco (*Nicotiana tabacum*), patata (*Solanum tuberosum*), melón (*Cucumis melo*), arabidopsis (*Arabidopsis thaliana*), arroz (*Oryza sativa*) y maíz (*Zea mays*). En la **Tabla XIV** se muestra un breve resumen taxonómico de las especies utilizadas.

MONOCOTILEDONEAS		
Familia	Género	Especie
Gramineae	Oryza	<i>Oryza sativa</i> (arroz)
	Zea	<i>Zea mays mays</i> (maíz)
DICOTILEDONEAS		
Familia	Género	Especie
Brassicaceae	Arabidopsis	<i>Arabidopsis thaliana</i>
Cucurbitaceae	Cucumis	<i>Cucumis melo</i> (melón)
Solanaceae	Solanum	<i>Solanum tuberosum</i> (patata)
	Nicotiana	<i>Nicotiana tabacum</i> (tabaco)
Rutaceae	Citrus	<i>Citrus sinensis</i> (naranja)

Tabla XIV. Taxonomía de las especies utilizadas en el estudio.

Nuevamente se realiza la amplificación selectiva de CIRE1 por PCR, el resultado de la cual se muestra en la **Figura 41**. El resultado de la amplificación por PCR, en las regiones de la INT y la RT de CIRE1 en especies fuera de la familia Rutaceae, es una banda poco intensa de tamaño algo superior a la obtenida en el naranjo (**Figura 41**). En ninguna de las especies testadas conseguimos amplificar fragmentos de la región del gen *gag* de CIRE1.



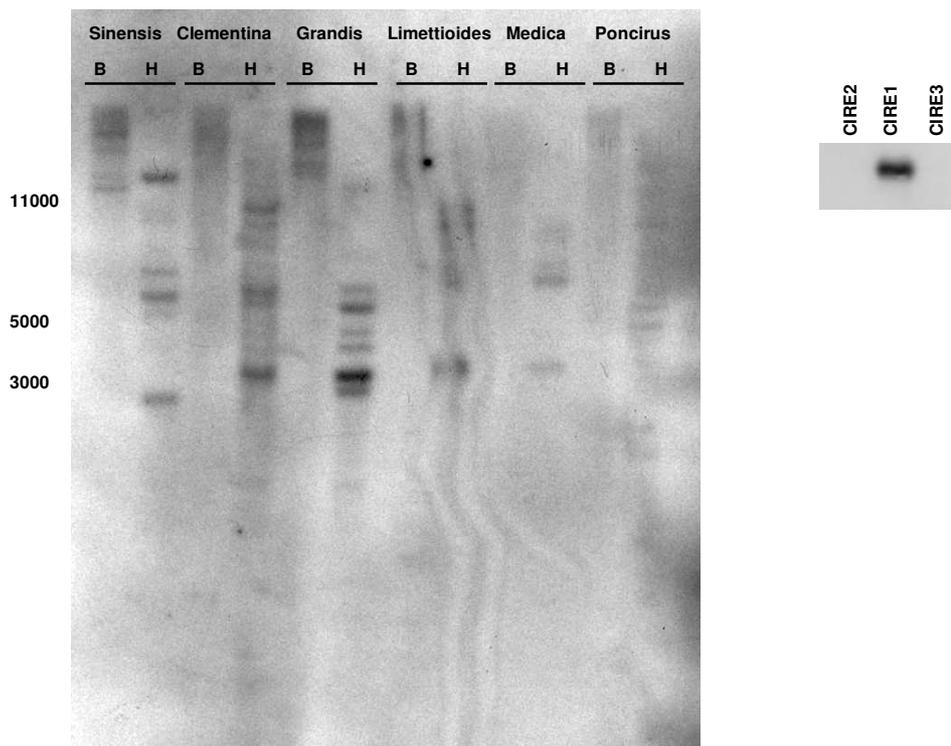
**Figura 41. Resultado de la amplificación selectiva de CIRE1 en el genoma de distintas especies vegetales.** Imagen del gel de agarosa al 2% tras la electroforesis del resultado de la amplificación sobre DNA genómico de parte de los genes *gag*, *int* y *rt*, respectivamente. Carril Pst, marcador de peso molecular Pst, indicando algunos tamaños en pb. Csi, *C. sinensis* (naranja); St, *Solanum tuberosum* (patata); Nt, *Nicotiana tabacum* (tabaco); Os, *Oryza sativa* (arroz); Zm, *Zea mays* (maíz); carril C-, control negativo de la reacción de PCR (sin DNA). La localización de los cebadores y el tamaño de banda que se espera amplificar son los mismos que en la **Figura 40**.

## Resultados

### A.3.8.3 Análisis genómico por Southern-blot de CIRE1 en la familia Rutaceae

Los DNAs genómicos de seis de las especies de la familia Rutaceae que contienen CIRE1, se han digerido con los enzimas de restricción BamHI y HindIII y se han analizado por *Southern blot* utilizando como sonda un fragmento interno de CIRE1 (correspondiente a una región de la RT). Paralelamente, se ha hibridado una membrana donde se han transferido los diferentes fragmentos de las secuencias consenso para esta región de las principales familias de RTN copia vistas en *C.sinensis*: CIRE1, CIRE2 y CIRE3 (Apartado A.1.3 de resultados). Este mismo fragmento de CIRE1 es el que se utiliza después como sonda para hibridar ambas membranas, y así probar su especificidad.

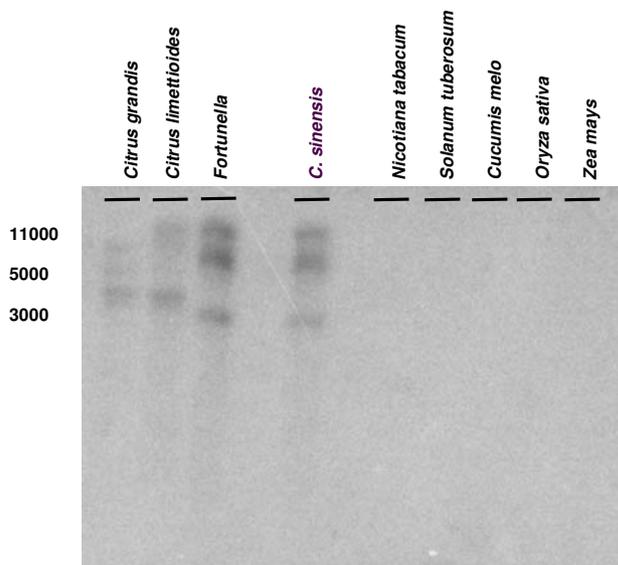
Se observa en la **Figura 42** que la sonda y las condiciones de hibridación utilizadas para el *Southern* son altamente específicas para esta familia, pues la sonda únicamente se une al fragmento correspondiente a CIRE1. También observamos un patrón de hibridación múltiple en las diferentes especies y digestiones analizadas. Este patrón es compatible con un alto número de copias de los elementos de la familia CIRE1 por el genoma de las seis especies analizadas, especialmente en *C. grandis*, *C. clementina* y *C. sinensis*.



**Figura 42. Análisis genómico por Southern-blot de CIRE1 en diferentes especies de la familia Rutaceae.** Se han digerido 10 µg de DNA genómico de *C. sinensis*, *C. clementina*, *C. grandis*, *C. limettioides*, *C. medica* y *Poncirus trifoliata* con BamHI (B) o HindIII (H) y tras transferirlos a una membrana, se ha hibridado con un fragmento de la RT de CIRE1. A la derecha se muestra una membrana con los diferentes clones de las secuencias consenso para fragmentos de RT de CIRE1, CIRE2 y CIRE3, que ha sido hibridada con la misma sonda CIRE1 para probar su especificidad. A la izquierda se indican algunos tamaños de banda en pb.

**A.3.8.4 Análisis genómico por *Southern-blot* de CIRE1 en especies de Rutaceae y otras familias**

Los DNAs genómicos de las diferentes especies descritas en el apartado A.3.8.2, se han digerido con el enzima de restricción HindIII y se han analizado por *Southern blot* utilizando como sonda el mismo fragmento interno de CIRE1 descrito en el apartado anterior. Se deduce por el análisis por *Southern-blot* de la **Figura 43** que la familia de CIRE1 tiene numerosas copias en el genoma de las especies del género *Citrus* y *Fortunella*, ya que aparecen varias bandas de relativamente alta intensidad. En cambio, no se observa ninguna banda en el genoma del resto de especies. Estos datos indican la ausencia del RTN CIRE1 en los genomas de las especies ajenas a la familia Rutaceae analizadas.



**Figura 43. Análisis genómico por *Southern-blot* de CIRE1 en diferentes especies vegetales.** 10 µg de DNA genómico de *Citrus grandis*, *Citrus limettioides*, *Fortunella*, *Citrus sinensis*, *Nicotiana tabacum*, *Solanum tuberosum*, *Cucumis melo*, *Oryza sativa* y *Zea mays* se han digerido con HindIII y tras transferirlos a una membrana, se han hibridado con un fragmento de la RT de CIRE1-1. La sonda específica de CIRE1 es la misma que en la **Figura 42**. A la izquierda se indican algunos tamaños de banda en pb.

### **A.4 ESTUDIO DE LA FUNCIONALIDAD DEL PROMOTOR DE LA LTR DE CIRE1**

El RTN-LTR CIRE1 secuenciado completamente presenta en su dominio interno una ORF sin interrupciones por codones de parada, y la traducción conceptual a aminoácidos da lugar a una secuencia con similitud significativa a la de proteínas esenciales del ciclo de vida de un RTN como son GAG, proteasa, integrasa, transcriptasa inversa o RNAsaH, sin necesidad de realizar ningún cambio en la pauta de lectura. Además, la mayoría de esas proteínas presentan los motivos funcionales descritos en otros RTN altamente conservados. Estos datos hacen factible que las hipotéticas proteínas de CIRE1 sean funcionales, pero para ello sería necesario primero que el elemento fuera activo transcripcionalmente, para lo cual sería imprescindible que el promotor del RTN, situado en la LTR 5', fuera funcional. Para determinar si el promotor de CIRE1 era funcional, se decidió analizar la LTR5' y su uso como promotor heterólogo del gen informador *gus*.

#### **A.4.1 Análisis de la secuencia del promotor de la LTR de CIRE1**

Las LTR de RTN funcionales contienen los motivos de los promotores necesarios para iniciar la síntesis de RNA, así como señales de terminación de la transcripción. Pueden contener también elementos reguladores que controlan la expresión en diferentes tejidos y bajo diferentes condiciones ambientales.

Las secuencias obtenidas del genoma de *C. sinensis* que corresponden a regiones de LTR de CIRE1, que han sido analizadas en el **apartado A.3.2.1**, presentan identidades que van del 100% al 97,8%, y en todos los casos contienen secuencias típicas de los promotores eucariotas (**Figura 32**).

De las 14 hipotéticas cajas de la LTR de CIRE1 con elevada similitud al consenso de las cajas TATA de los promotores de genes de plantas, se ha considerado como la más probable para ser la caja TATA funcional real, la que se indica en la **Figura 44**. Esta caja TATA está situada a unos 50 nucleótidos del hipotético sitio de inicio de la transcripción (SIT), de igual secuencia (C/T)TCATCA que el consenso en plantas (Joshi, 1987). Por otra parte, la hipotética señal de polinización está ubicada 11 nucleótidos cadena abajo del SIT y el posible sitio de poliadenilación, de secuencia igual al consenso CATT(T/G), situado 8 nucleótidos en 3' a la señal (**Figura 44**). La zona comprendida entre el inicio de la LTR y el hipotético inicio de la transcripción (SIT) corresponde al dominio U3 de la LTR de 215 pb de largo. El dominio R de la LTR en CIRE1 es muy corto, de 25 pb, y va desde el SIT hasta el sitio de poliadenilación. Desde este punto hasta el final de la LTR corresponde al dominio U5 de la LTR, en CIRE1 de 225 pb.

### LTR5' de CIRE1

```

+ TGTGAAGATATGTGTATATGGACTTAAATGAGAAACTGGAAAAGTAGAGAGCTGTTGTGCAGATTGACTTCA GTTGA
- ACACCTTCATACACATATACCTGAATTTACTTCTTTGACCTTTTCATCTCTCGACAACACGCTTAAC TGAA GTCAACT

+ TGGGGTCTATTTGATATCTCCGCTGACGTGGCATGAATTCGTTATAACTGTTATAACGAAGCCCAGCACAC
- ACCCCAGATAAAC TATAGAAGGCGACTGCACCGTACTTAAGCAATATTAGACAATATTGCTTCGGGTCGTTGG

+ TGGATATAAAACAGAGGC TGACCAATGGTGCCTCATACTGTGCTAATATCGTCTTCATTTGTTTCATCATTTTGC
- ACTATATTTTGTCTCCGACTGCTTACCACGGAGTATGACACGATTATAGCAGAAGTAAACAAGTAGTAAAACG

+ ATAGAATCTGTTTCATTGTGATCTGTGCATAATCTTCTGTAACACCTGTGATTGGGTTAGTTCTTGATCAA
- TATCTTAGACAAAGTAACACTAGACACGTATTAGAAGACATTTAGTGGACACTAACCAATCAAGAAGTAGTT

+ TAATAATGAGTAAATTTCTTCTAAGTTTGCTGGTCTGTGCTTTTGTGTTTGTCTCTGGTTTAGCTTCTTTG
- ATTATTACTCATTTAAAGGAAGATTCAAACGACCAGACACAGAAAACAAAACAGAGACCAAATCGAAGAAAC

+ TTCTGGTTTTTCTTGATTCGATTGTGTTCTTGGTGTCTTACTACCAAGATTCGTGAGTCTTGCTGTGGTTT
- AAGACCAAAAAGAACTAAGCTAACACAAGAACCACAAAGATGAGTGGTTCTAAGCACTCAGAACGACACCAAA

+ AAACCTAAAATCTTAACA
- TTTGAATTTTGAATTGT
  
```

<p><b>ABRE:</b> motivo de respuesta a ácido abscisico</p> <p><b>AuxRR-core:</b> motivo de respuesta a auxinas</p> <p><b>TGA-box:</b> motivo de respuesta a auxinas</p> <p><b>CGTCA-motif:</b> motivo de respuesta a MeJA</p> <p><b>TGACG-motif:</b> motivo de respuesta a MeJA</p> <p><b>HSE:</b> motivo de respuesta a calor</p> <p><b>MBS:</b> motivo de respuesta a sequedad</p> <p><b>As1</b> motivo de expresión específica en raíz</p> <p><b>CAAT-box</b></p> <p><b>TATA-box</b></p>	<p><input type="checkbox"/> <b>Hipotética caja TATA funcional</b></p> <p>CIRE1            aCACTGATATAaAa</p> <p>Consenso        TCACT - ATATATAG</p> <p><input type="checkbox"/> <b>Hipotético sitio de inicio de la transcripción (SIT)</b></p> <p>CIRE1            T TCATCA</p> <p>Consenso        (C/T)TCATCA</p> <p><input type="checkbox"/> <b>Hipotética señal de poliadenilación</b></p> <p>CIRE1            cATAgAA</p> <p>Consenso        AATA-AA</p> <p><input type="checkbox"/> <b>Hipotético sitio de poliadenilación</b></p> <p>CIRE1            CATTG</p> <p>Consenso        CATT(T/G)</p>
--	--

**Figura 44. Secuencia de la LTR5' de CIRE1 indicando los posibles motivos reguladores de la transcripción.** Se indican enmarcados los posibles caja TATA, señal de poliadenilación y lugar de inicio de la transcripción y poliadenilación. Determinado por lo anterior, se indican también las posibles regiones en que se divide la LTR: U3, R y U5. En la parte inferior de la figura, se indica a la izquierda, las diferentes señales y cajas del promotor que se han hallado utilizando las siguientes bases de datos: *PlantCARE: a database of plant cis-acting regulatory elements* y *PLACE: plant cis-acting regulatory DNA elements database*. A la derecha, se comparan las cajas y sitios propuestos con los equivalentes consensos descritos para los genes de plantas (Vicent, 1995). Todos estos motivos y cajas se indican en distintos colores en la secuencia

## Resultados

---

Las LTR de CIRE1 presentan también 10 secuencias CAAT (típicas de promotores de animales) y diferentes motivos descritos como de respuesta a: ácido abscísico (ABA), ácido salicílico, luz, auxinas, jasmonato de metilo (MeJA), estrés térmico o sequedad, herida y cajas asociadas a la expresión específica en raíz. La mayoría de estas señales se sitúan en el dominio U3 de la LTR en CIRE1, que contiene el promotor del elemento. Todos los motivos descritos están conservados en los diferentes clones de CIRE1 y cumplen los estándares descritos para promotores de genes de plantas en cuanto a su secuencia y distancia entre ellos (**Figura 44**).

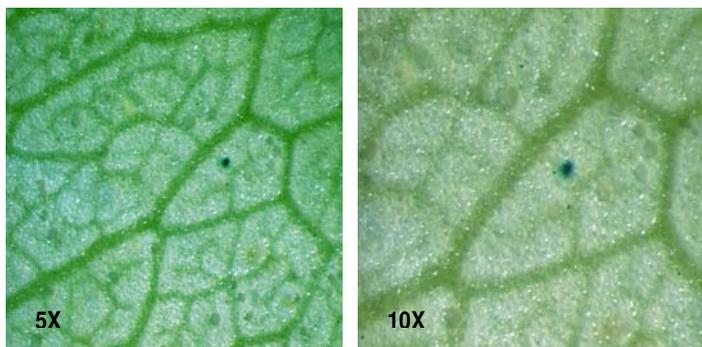
Tras confirmar la presencia de las cajas de los promotores eucariotas y los hipotéticos elementos reguladores en la secuencia de las LTR de CIRE1, se decidió continuar con el estudio de la funcionalidad del promotor mediante experimentos de expresión transitoria bombardeando tejido vegetal con una construcción que contiene la secuencia entera de la LTR5' de CIRE1 (mostrada en la Figura 44) fusionado al gen informador *gus* (Casacuberta et al., 1993; Suoniemi et al., 1996; Hirochika et al., 1996; Vernhettes et al., 1997; Mhiri et al., 1997; Pouteau et al., 1994; Pauls et al., 1994). Para realizar la construcción de DNA utilizada en la transformación, el hipotético promotor de CIRE1 (la secuencia entera de la LTR5') se unió con el gen informador *gus* y con el terminador NOS (ver Material y Métodos) obteniéndose la construcción pBSKLTR::*gus*::*nos*. También se utilizó una construcción en la que el promotor constitutivo 35S dirige la expresión del gen informador *gus* con el fin de utilizarla como control positivo del experimento. Como control negativo del experimento se utilizó el plásmido donde se había clonado la secuencia de la LTR, que sólo contiene el gen *gus* y el terminador NOS sin ninguna secuencia promotora.

### **A.4.2 Actividad del promotor de la LTR5' de CIRE1**

#### **A.4.2.1 Experimentos de expresión transitoria en naranja**

Con el fin de saber si el promotor de la LTR5' de CIRE1 era capaz de dirigir la actividad GUS se seleccionaron hojas de plantas de poca edad cultivadas *in vitro*. Se escogieron estas mejor que hojas de plántulas de naranja crecidas en invernadero ya que al ser el primero un tejido más blando permitirá mejor la transformación. Se bombardearon las hojas mediante microproyectiles que llevaban adheridos el DNA de la construcción pBSKLTR::*gus*::*nos*. Igualmente se bombardearon las construcciones control [promotor del 35S (Control positivo) y plásmido sin promotor (Control negativo)]. El tejido bombardeado se mantuvo durante 24 horas en oscuridad a 26°C y seguidamente se determinó la actividad GUS por el método histoquímico.

Los resultados obtenidos tras la tinción histoquímica muestran que el promotor de la LTR5' de CIRE1 es capaz de dirigir la actividad GUS en células de hoja de *Citrus sinensis*, aunque el número de células transformadas resultó ser muy bajo (**Figura 45**).



**Figura 45.**  
Localización de la actividad GUS en hojas de *C.sinensis* transformadas con pBSKLTR:gus::nos

#### **A.4.2.2 Experimentos de expresión transitoria en células de epidermis de cebolla**

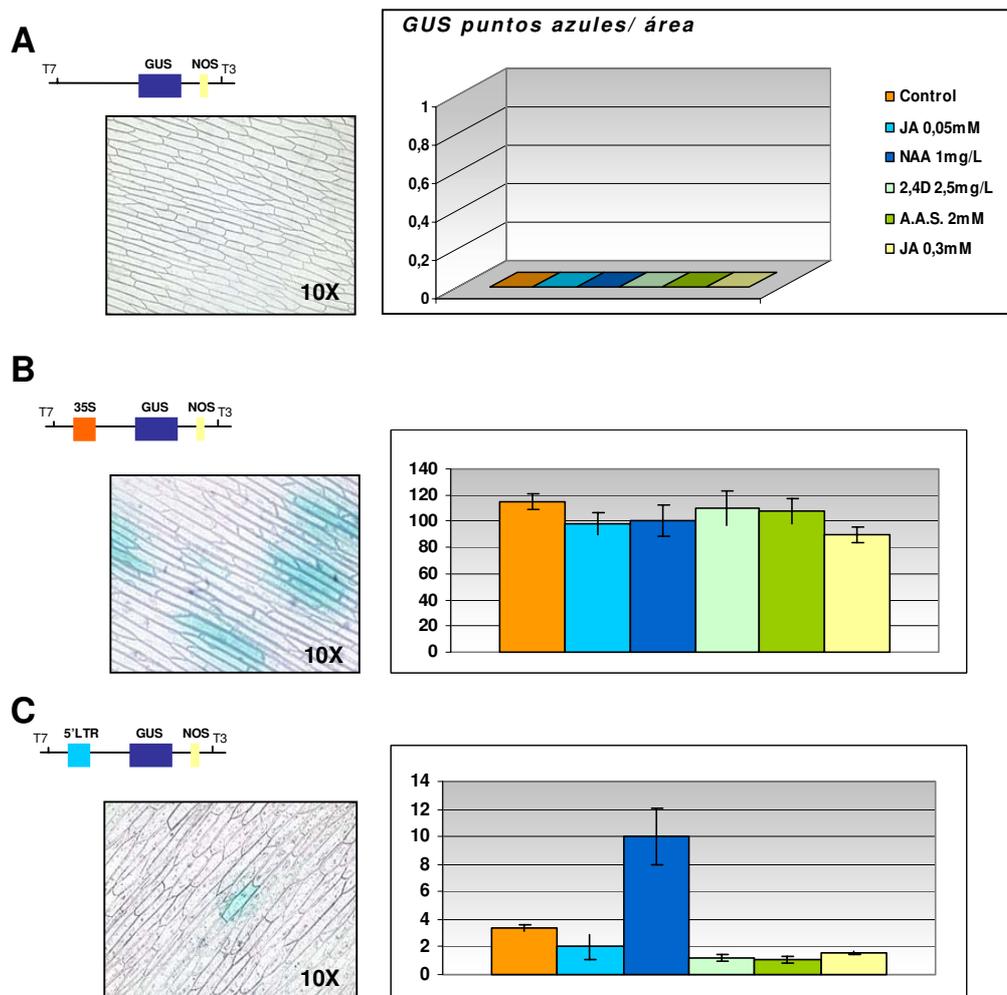
Se realizaron también experimentos de transformación de células de epidermis de cebolla mediante bombardeo con las mismas construcciones del apartado anterior. Estas células son grandes, transparentes y generalmente están en monocapa. El sistema de transformación es fácil y la expresión transitoria y su detección son rápidas, por lo que se trata de un ensayo muy conveniente para mirar actividad de promotores. Tras el bombardeo se determinó la actividad GUS por el método histoquímico. Los resultados obtenidos se muestran en la **Figura 46** (medio MS control). En este caso sí vemos actividad GUS en las células transformadas con las construcciones pBSK35S::gus::nos (**B**) y pBSKLTR::gus::nos (**C**), siendo la actividad de esta última un orden de magnitud menor que la construcción con el promotor 35S.

Para averiguar si el promotor de la LTR5' es inducible se plantearon nuevos experimentos sometiendo el tejido transformado a diferentes tratamientos. Así, se bombardearon células de cebolla con las mismas construcciones del experimento anterior, y después se incubaron 24 horas en medio sólido MS con suplemento exógeno de 50  $\mu$ M o 0,3mM de ácido jasmónico (JA) (Takeda et al., 1998), 1 mg/L de ácido naftalenacético (NAA) (Hirochika et al. 1993; Casacuberta et al., 1995), 2,5 mg/L de 2,4 D o 2mM de ácido salicílico. Posteriormente se determinó la actividad GUS mediante tinción histoquímica.

Los resultados presentados en la **Figura 46C** muestran que la LTR5' de CIRE1 es capaz de dirigir la expresión de GUS en todas las condiciones probadas en células de la epidermis de cebolla a un nivel similar o inferior al control, excepto en el tratamiento

## Resultados

con la auxina NAA, donde se observa un incremento de la actividad GUS, probablemente debido a una inducción del promotor de CIRE1. En el caso del promotor 35S se observa en las células bombardeadas una actividad GUS un orden de magnitud mayor que con el de la LTR, siendo además no inducible (**Figura 46B**). No se observa actividad GUS en células bombardeadas con la construcción sin promotor (**Figura 46A**).

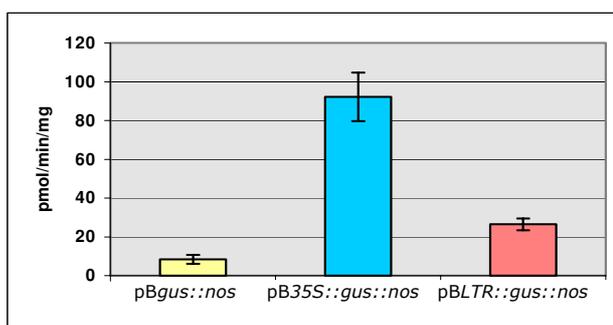


**Figura 46. Análisis de la expresión transitoria del promotor de la LTR5' de CIRE1 en células de epidermis de cebolla.** Detección histoquímica de células de la epidermis de cebolla bombardeadas con las construcciones (A) pBSKGus::nos (control negativo): plásmido que contiene el gen informador *gus* sin promotor, (B) pBSK35S::gus::nos (control positivo): plásmido que contiene el gen informador *gus* y el promotor constitutivo 35S y (C) pBSKLTR::gus::nos: plásmido que contiene el gen informador *gus* y la LTR5' de CIRE1. Previo a la tinción con X-Gluc., el tejido bombardeado se incubó 24 horas en medio MS suplementado con las hormonas vegetales indicadas en la leyenda de la gráfica superior. Las barras representan las medias del número de puntos azules detectados por segmento de área equivalente (2 réplicas por construcción, en un mínimo de 2 experimentos independientes), y las barras de error son el error estándar de la media.

#### A.4.2.3 Experimentos de expresión transitoria en suspensiones celulares de tabaco

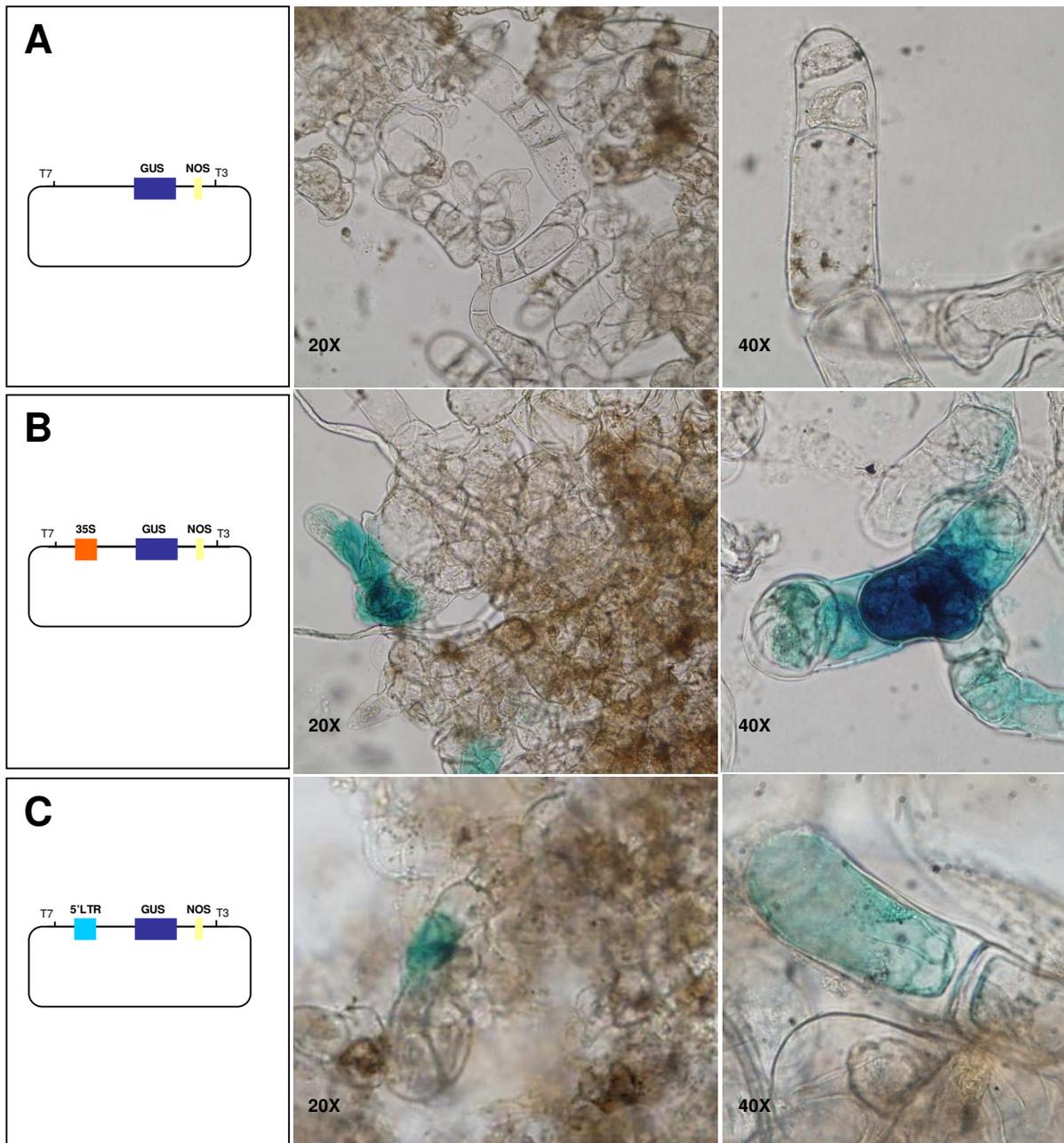
Una vez demostrado que el promotor de la LTR5' de CIRE1 era activo en células de epidermis de cebolla, se decidió continuar con el análisis de la actividad del promotor bombardeando suspensiones celulares de tabaco (dicotiledónea como el naranjo) con las construcciones de DNA del apartado anterior; utilizando un método más cuantitativo que el histoquímico para determinar la actividad GUS: el método fluorimétrico. Este método, frente al histoquímico que es más cualitativo, se basa en la cuantificación de la actividad GUS por fluorimetría del producto 4-MU de la reacción catalizada por GUS sobre el sustrato MUG. La enzima GUS proviene del extracto de proteína total del tejido bombardeado después de 24 horas de la transformación. La cuantificación de la actividad del promotor mediante el ensayo histoquímico se hace por el conteo de puntos azules observados, que corresponden a eventos individuales de transformación. Es un ensayo rápido y sencillo pero difícil de cuantificar pues la unidad de cuantificación (presencia / ausencia de punto azul) no discrimina entre el tamaño del punto o si el color de éste es fuerte o débil. No reflejando entonces si la expresión del gen informador ha sido más o menos alta, y por tanto si el nivel de actividad del promotor ha sido más o menos alto. Una clara consecuencia de este hecho puede ser la sobrevaloración de la actividad de promotores débiles y / o la infravaloración de la actividad de promotores fuertes.

La diferencia de intensidad en el color azul debida a GUS por fluorimetría se detectó en algunas células transformadas al observarlas al microscopio. Tal y como se muestra en la **Figura 47** la coloración conseguida en las células transformadas con la construcción pBSK35S::gus::nos (**B**) es mayor frente la de las células transformadas con la construcción pBSKLTR::gus::nos (**C**), y como era de esperar no se observa coloración azul en las células bombardeadas con el control negativo (**A**). Pero la diferencia más clara en la actividad de los promotores analizados se observa tras el estudio fluorimétrico, qué muestra que en las condiciones en que éste se ha realizado, la LTR5' de CIRE1 es funcional ya que se observa una clara actividad de GUS dirigida por el promotor de la LTR5' cuando se compara con el control sin promotor, aunque esta actividad da 5 veces inferior a la dirigida por el promotor 35S (**Figura 48**).



**Figura 48.** Actividad del promotor de la LTR5' de CIRE1 en comparación al promotor 35S en suspensiones celulares de tabaco, determinada por ensayo fluorimétrico de GUS. Suspensiones celulares de tabaco W38 se bombardearon con las construcciones indicadas en abscisas y tras 24h. se extrajo la proteína total y se determinó GUS fluorimétricamente. Los valores dados son la media y el error estándar calculados para 4 réplicas por construcción en cada experimento, en al menos 2 experimentos independientes.

## Resultados



**Figura 47. Tinción histoquímica de suspensiones celulares de tabaco (W38) transformadas con las construcciones (A) pBSK*gus*::*nos* (control negativo): plásmido que contiene el gen informador *gus* sin promotor, (B) pBSK35S::*gus*::*nos* (control positivo): plásmido que contiene el gen informador *gus* y el promotor constitutivo 35S y (C) pBSKLTR::*gus*::*nos*, plásmido que contiene el gen informador *gus* y la LTR5' de CIRE1.**

## **A.5 EXISTENCIA DE RTN-LTR TRANSCRIPCIONALMENTE ACTIVOS EN CITRUS SINENSIS**

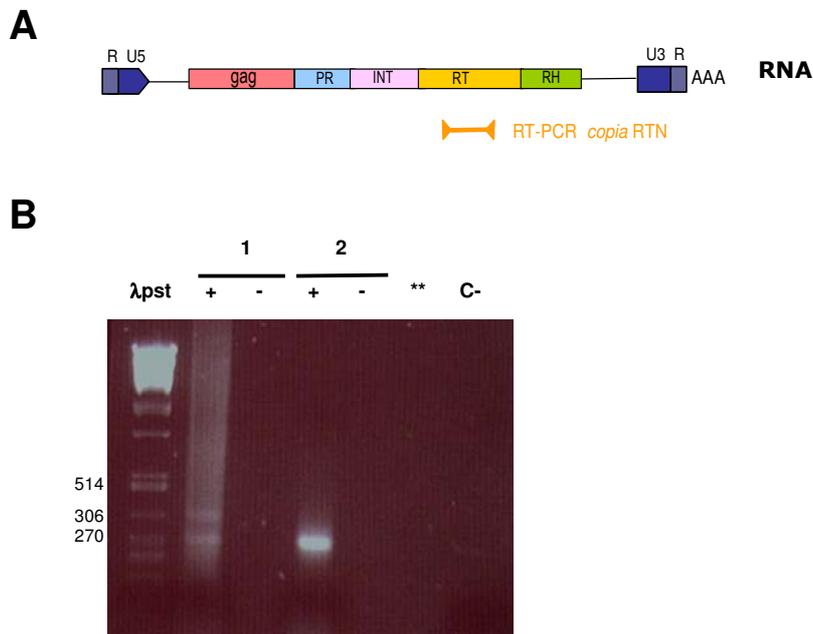
### **A.5.1 Actividad transcripcional de RTN-LTR tipo *copia* en naranjo**

#### **A.5.1.1 Amplificación de transcritos de RTN-LTR tipo *copia***

La amplificación por PCR de DNA genómico utilizando cebadores degenerados de regiones altamente conservadas del dominio RT de RTN-LTR tipo *copia*, resulta ser un buen método para clonar, secuenciar y analizar estos RTN en plantas. Del análisis de las secuencias de la RT (u otros dominios codificantes), se deduce que muchos de los RTN identificados previamente no son, previsiblemente, funcionales al menos en sus dominios proteicos. Para conocer si el conjunto de RTN-LTR tipo *copia* de *C. sinensis* son transcripcionalmente activos (tengan dominios proteicos funcionales o no) se decidió realizar experimentos de RT-PCR sobre RNA total de tejidos vegetales, como se ha descrito en Hirochika (1993). En la amplificación por PCR se utilizan los mismos cebadores degenerados específicos de RTN-LTR tipo *copia* descritos anteriormente en el apartado A.1.1., usando como molde en este caso el conjunto de cDNA (provenientes de los transcritos copiados a DNA mediante la transcripción inversa) en lugar del DNA genómico. De esta manera, habrá amplificación en la PCR si existen transcritos correspondientes a RTN-LTR de tipo *copia*, lo que indicará que estos elementos son transcripcionalmente activos. El uso de la RT-PCR en una primera aproximación, en lugar de experimentos de *northern*, se justifica por el bajo nivel de transcripción de la mayoría de los RTN, cuyos transcritos se detectan mejor por RT-PCR al ser una técnica más sensible.

La **Figura 49** muestra los resultados de la RT-PCR obtenidos con RNA total procedente de hoja y raíz de plantas jóvenes crecidas en invernadero. Se observa que en plantas crecidas en condiciones normales hay expresión clara de RTN-LTR tipo *copia* en raíz donde se obtiene una banda única del tamaño esperado para la región de la RT que se está amplificando [**Figura 49, 2(+)**]. En cambio en hoja, se detecta amplificación pero resulta ser comparativamente más inespecífica e ineficiente que en raíz [**Figura 49, 1(+)**]. Para estar totalmente seguros que la amplificación proviene del cDNA cada muestra de RNA total se trata por duplicado, sometiendo una réplica a la reacción de transcripción inversa y la otra no. Así cualquier amplificación que se de en la réplica no sometida a la transcripción inversa será debida a contaminación de DNA genómico y ambas muestras se descartarán. Como se observa en la **Figura 49** en los controles no se observa amplificación [**Figura 49, 1(-) y 2 (-)**] dando validez a la reacción.

## Resultados

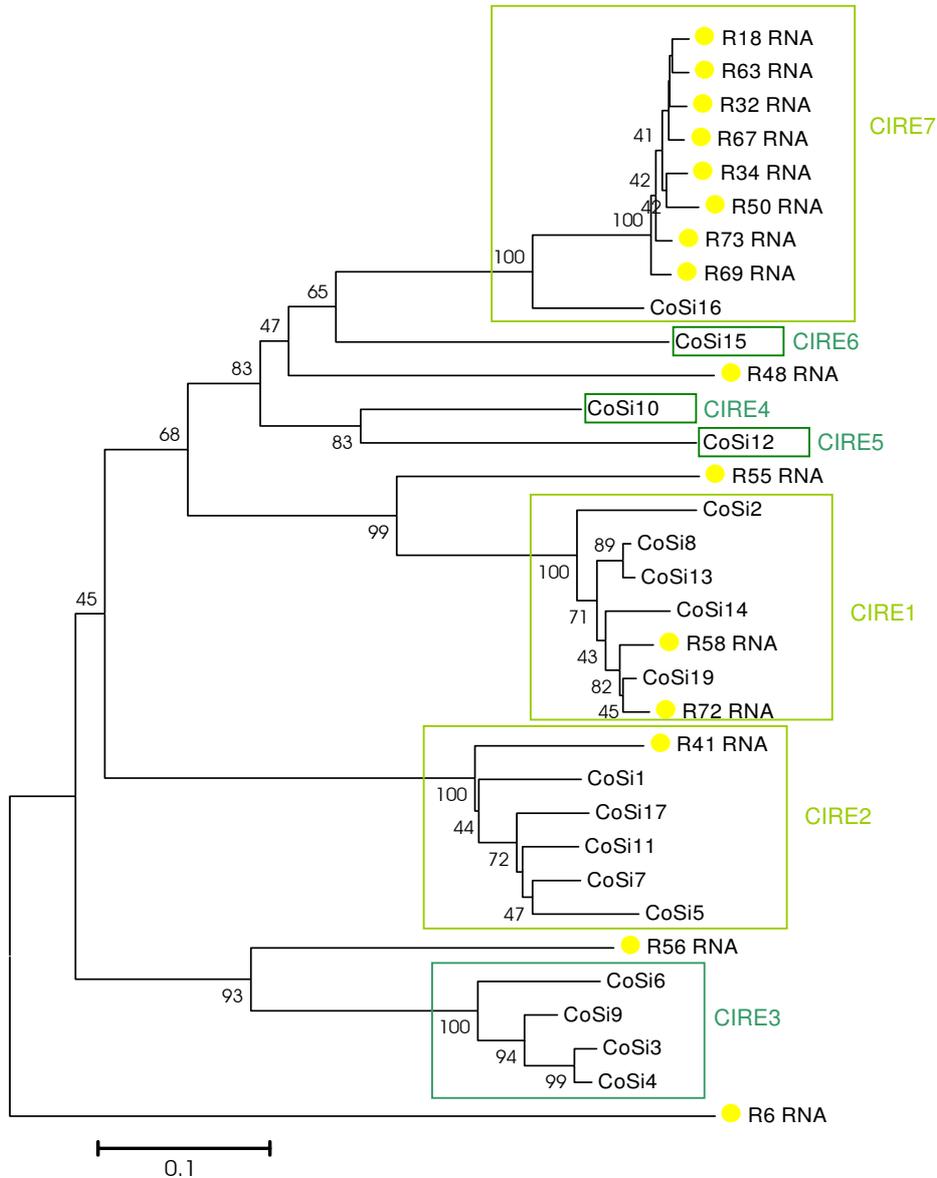


**Figura 49. Obtención de transcritos de RTN-LTR *copia* en *C.sinensis*. (A) Esquema que muestra la región del elemento tipo *copia* amplificada por RT-PCR. Los cebadores utilizados son los oligonucleótidos degenerados específicos para RTN-LTR tipo *copia*. La secuencia de los cebadores y las condiciones de amplificación se describen en el Apartado 7.2.1 de Material y Métodos. La PCR se ha realizado a 30 ciclos con Ta de 45°C. (B) Imagen del gel al 1,5% de agarosa tras la electroforesis de los productos amplificados por RT-PCR. Carril  $\lambda$ Pst, marcador de peso molecular  $\lambda$ Pst, indicando algunos tamaños en pb; (1) Hoja congelada inmediatamente después de ser cortada, (2) Raíz congelada inmediatamente después de ser cortada, (+) cDNA (RNA sometido a reacción de RT), (-) RNA (RNA sin paso previo de transcripción inversa). (\*\*) Control negativo de la mezcla de reacción de transcripción inversa, (C-) Control negativo de la mezcla de reacción de PCR.**

Los productos de tamaño esperado (270 pb) de dos reacciones de PCR independientes, fueron clonados en el vector pGEMT y se determinó la secuencia de 15 clones escogidos al azar para la muestra amplificada a partir RNA total de raíz. En el caso de la RT-PCR sobre RNA de hoja, las secuencias clonadas resultaron ser inespecíficas.

### **A.5.1.2 Descripción de las familias de RTN-LTR tipo *copia* transcripcionalmente activas en naranjo**

A partir del alineamiento múltiple por ClustalW de las 15 secuencias nucleotídicas procedentes de los transcritos aislados de RNA de raíz junto las 18 secuencias nucleotídicas procedentes de DNA genómico descritas en el apartado A.2.1, se obtiene el árbol por el método de *Neighbor joining* que se muestra en la **Figura 50**.



**Figura 50.** Árbol fenético basado en el alineamiento de secuencias nucleotídicas de transcritos de RTN-LTR *copia* obtenidos por RT-PCR y fragmentos genómicos de RTN-LTR *copia* amplificados por PCR en *C. sinensis*. Árbol tipo *Neighbour joining* (NJ) generado con el programa MEGA2. La divergencia entre las secuencias en unidades de distancia genética viene definida por la longitud de las ramas horizontales. Esta indicado en cada bifurcación los valores de bootstrap  $\geq 40\%$ . Se han recuadrado las diferentes familias en qué quedan divididas las secuencias, el nombre de cada familia está indicado a la derecha de la figura. Las familias CIRE marcadas de color verde claro tienen representantes transcripcionalmente activos y las marcadas de color verde oscuro no. CoSi#: clones de la amplificación de DNA genómico de *C. sinensis*. (•) R# RNA: secuencias obtenidas de la reacción de RT-PCR sobre RNA total de raíz. Tanto en la PCR sobre genómico como en la RT-PCR se han utilizado los mismos cebadores degenerados específicos de RTN-LTR tipo *copia* descritos en el capítulo 7.2.1 de Material y Métodos.

## Resultados

---

En el árbol se observa que 8 de las secuencias correspondientes a transcritos: R18, R32, R34, R50, R63, R67, R69 y R73 son muy similares entre sí (identidad de secuencia del 97%) y a su vez comparten una identidad nucleotídica media del 86% con CoSi16. Las secuencias de los clones de los transcritos R58 y R72 (identidad de secuencia del 96%) son como media un 93% idénticas a las secuencias de CoSi2, CoSi19, CoSi14, CoSi8 y CoSi13, llegando a ser en concreto la secuencia nucleotídica de R72 en un 97% idéntica a la de CoSi19. El transcrito R41 comparte una similitud de secuencia media del 85% con CoSi1, CoSi17, CoSi11, CoSi5 y CoSi7. El resto de secuencias obtenidas en la RT-PCR están distribuidas por el árbol sin pertenecer a ningún grupo claro definido previamente (los elementos son asignados a un mismo grupo si comparten una identidad nucleotídica  $\geq 85\%$  entre ellos). Así, la secuencia de R55 es en un 74% idéntica a CoSi2, la de R56 lo es en un 70% a CoSi6 y R48 en un 70% a CoSi16.

Del análisis de los clones secuenciados al azar en este estudio se concluye que, de la población de elementos expresados en raíz el 53% (R18, R32, R34, R50, R63, R67, R69 y R73) corresponden a productos de transcripción correspondientes a elementos de la familia CIRE7, el 13% (R58 y R72) corresponden a la familia de secuencias CIRE1 y el 7% (R41) corresponde a la familia de secuencias CIRE2.

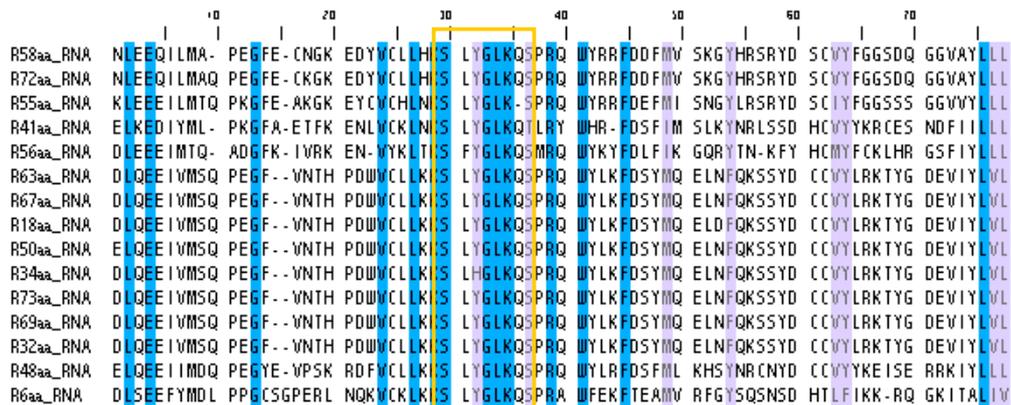
Por tanto, no todos los elementos CIRE que hemos clonado y secuenciado de naranjo en este estudio aplicando el método de amplificación por PCR sobre DNA genómico (descritos en el **apartado A.1**) se agrupan con secuencias de elementos transcripcionalmente activos en raíz, ni todas las secuencias de transcritos obtenidas se corresponden con familias de RTN-LTR tipo *copía* definidas por ahora en naranjo, tal y como queda representado en la **Figura 50**.

Se realiza una búsqueda en las colecciones de EST (*Expressed Sequence Tags*) para localizar posibles secuencias derivadas de RTN en el genoma de naranjo. El análisis de las 45.192 entradas de EST de *C. sinensis* no muestra ninguna secuencia similar a RTN.

### **A.5.1.3 Comparación de la variabilidad de la población de secuencias de transcritos y de secuencias genómicas correspondientes a RTN-LTR tipo *copía***

El análisis de las secuencias nucleotídicas de los transcritos revela que se están transcribiendo diferentes elementos tipo *copía*, de hecho todos los clones escogidos al azar y secuenciados resultan ser diferentes entre sí, aunque hay una familia (CIRE7) que está mayoritariamente representada.

Las secuencias aminoacídicas deducidas conceptualmente confirman, al igual que ocurría en las secuencias genómicas CoSi#, la presencia en todas ellas de dominios conservados característicos de secuencias de RT de RTN-LTR tipo *copia* (**Figura 51**). Para uno de los clones (R50) ha sido necesario arreglar la secuencia para realizar el alineamiento, ya que presenta la inserción de 1 pb que hace necesario realizar un salto en la pauta de lectura correspondiente a RT de RTN.



**Figura 51. Alineamiento múltiple por ClustalW de las secuencias amplificadas mediante RT-PCR de transcritos de RTN-LTR tipo *copia* presentes en *C. sinensis*.** El orden relativo de las secuencias está directamente relacionado con el nivel de similitud entre ellas. Los guiones indican ausencia de aminoácido en esa posición. El sombreado oscuro indica identidad de secuencia del 100% y el más claro, elevada similitud. Se ha recuadrado el consenso SLYGLKQA típico de secuencias de RT de RTN-LTR tipo *copia*. No se muestran los dominios conservados correspondientes a los cebadores que flanquean las secuencias (ver pie **Figura 15**).

Para la población de secuencias nucleotídicas de la RT de DNA genómico (clones CoSi#) el número de posiciones polimórficas (S) es de 164 de los 274 nucleótidos (N) que tienen las secuencias. En los sitios polimórficos, se observan un total de 36 mutaciones, de las cuales 20 son sinónimas y 16 son no-sinónimas esto es, que dan lugar a cambios aminoacídicos. La mayoría de los cambios en los sitios polimórficos se deben a una sola sustitución por sitio. El 39% de las secuencias obtenidas presentan mutaciones puntuales que llevan a la aparición de codones de parada, pero ninguna presenta cambios por inserción/delección que provoquen un salto en la pauta de lectura. En el caso de las secuencias de transcritos de la misma región (secuencias R#\_RNA), el número de sitios polimórficos (S) es de 191 de los 274 (N) que tienen las secuencias. Estos sitios polimórficos son además en su mayoría hipervariables ya que cada secuencia puede presentar sustituciones nucleotídicas diferentes, por tanto el cambio se debe a varias sustituciones por sitio. El número total de mutaciones es de 57, de las cuales 13 son sinónimas y 44 provocan cambios aminoacídicos. El 80% de los transcritos presentan codones de parada en su secuencia y 1 clon de los 15

## Resultados

secuenciados presenta una inserción de 1 pb que hace necesario realizar un salto en la pauta de lectura correspondiente a RT de RTN.

En la **Tabla XV** se muestran algunos parámetros de variabilidad genética calculados para la población de DNA genómico y la población de transcritos de la misma región de la RT de RTN-LTR tipo *copia*. Estos parámetros se han obtenido utilizando el programa DNAsp3.14 (Rozas y Rozas, 1999). Se observa que la diversidad nucleotídica ( $\pi$ ) para ambas poblaciones (DNA genómico y transcritos) es muy similar.

	<b>N</b>	<b>m</b>	<b>S</b>	<b>S/m</b>	<b>K</b>	<b><math>\pi</math></b>	<b><math>\pi</math> (JC)</b>
<b>DNA</b>	274	231	164	0,710	78,02	0,337	0,475
<b>RNA</b>	274	263	191	0,726	80,13	0,304	0,443

N: número total de nucleótidos

m: número total de posiciones analizadas (excluyendo los que tienen gaps)

S: número de sitios polimórficos

S/m: número de sitios polimórficos por posiciones analizadas

K: número medio de diferencias nucleotídicas entre parejas de secuencias

$\pi$ : Diversidad nucleotídica (siendo  $\pi = K/m$ )

$\pi$  (JC): diversidad nucleotídica (con la corrección de Jukes y Cantor, 1969)

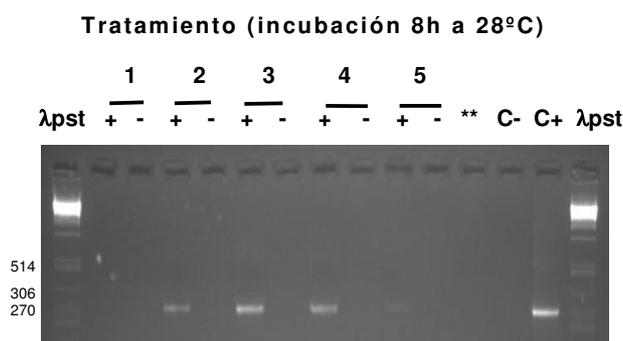
**Tabla XV. Parámetros de variabilidad genética de la población de secuencias genómicas y transcritos de la región de la RT de RTN-LTR tipo *copia* calculadas con el programa DNAsp3.14 (Rozas y Rozas, 1999)**

### A.5.1.4 Estudio de la expresión de RTN-LTR tipo *copia* en naranjo

Una vez confirmada la presencia en naranjo de retroelementos transcripcionalmente activos en condiciones normales pertenecientes al grupo de RTN-LTR tipo *copia*, el siguiente estudio fue identificar si hay elementos cuya expresión se activa durante determinadas condiciones y ver si esta activación se da de la misma manera en hoja que en raíz. Para ello se ha aislado RNA procedente de discos de hoja de naranjo de la variedad "Valencia late" sometidos a diversos tratamientos (herida mecánica o estrés abiótico). Además del daño mecánico propiamente dicho (Bowles et al., 1990), suplementar de forma exógena con jasmonato de metilo (JA) se ha visto que activa la expresión inducida por herida de varios RTN-LTR de plantas (Takeda et al., 1998). Por ello se probó el posible efecto de esta hormona sobre la activación transcripcional de elementos de naranjo suplementando el medio de incubación con 50 $\mu$ M JA. Asimismo, se conoce que los transposones de Clase II de maíz Ac/Ds y Spm se activan durante cultivos celulares (Peschke & Phillips, 1991), de la misma manera que algunos RTN-LTR de tabaco se activan en protoplastos (Pouteau et al., 1991) y en cultivos celulares (Hirochika et al., 1993; Casacuberta et al., 1995). Una de las hipótesis es que la

activación es debida a la adición de hormonas en el medio de cultivo durante el proceso, por ello se quiso probar el efecto de hormonas vegetales sobre la activación transcripcional de RTN-LTR en naranjo suplementando el medio de incubación con la hormona ácido naftalenacético (NAA) a 1mg/L.

En detalle, se cortaron las hojas en pequeños trozos (~ 1 cm<sup>2</sup>), se agruparon en diferentes subgrupos, y posteriormente se incubaron en tampón MES durante 8h, con o sin hormona en el medio, o con o sin cortes en las hojas, según el tratamiento (Ver Figura 52). En la reacción de RT-PCR se utilizaron nuevamente los cebadores degenerados específicos de RTN-LTR tipo *copia* que amplifican una región de la RT de aproximadamente 270 pb. Vemos que en condiciones control no se observa una expresión clara en RNA extraído de hoja [Figura 52, 1(+)], corroborando el bajo nivel de producto amplificado obtenido anteriormente [Figura 49, 1(+)]. En cambio aparece una banda clara en todas aquellas muestras de RNA extraído de hoja sometida a estrés o tratamiento hormonal [Figura 52, 2-4(+)], salvo en la tratada con JA [Figura 52, 5(+)] donde la banda es más débil. Nuevamente en los controles negativos no aparecen productos de amplificación [Figura 52, carriles (-)] excluyendo falsos positivos y validando la reacción.



**Figura 52. Análisis por RT-PCR de la acumulación de transcritos de RTN-LTR tipo *copia* en hoja de *C. sinensis* sometida a diferentes tipos de tratamientos.** Imagen del gel de agarosa al 1,2% tras la electroforesis de los productos amplificados en la RT-PCR utilizando como cebadores los oligonucleótidos degenerados específicos para RTN-LTR tipo *copia* descritos en el Apartado 7.2.1 de Material y Métodos. La PCR se ha realizado a 30 ciclos con Ta de 45°C. Carril λPst, marcador de peso molecular λPst, se indican algunos tamaños en pb. **Carriles (+):** cDNA (RNA sometido a reacción de transcripción inversa), **Carriles (-):** RNA (RNA sin paso previo de transcripción inversa). **1. Control:** Hoja congelada inmediatamente después de ser cortada. **2. Corte:** Hoja cortada e incubada en tampón 0,05% MES-KOH pH 5,7. **3. Herida:** Hoja cortada y con heridas, incubada en tampón 0,05% MES-KOH pH 5,7. **4. NAA:** Hoja cortada e incubada en tampón 0,05% MES-KOH pH 5,7 + 1mg/L NAA. **5. JA:** Hoja cortada e incubada en tampón 0,05% MES-KOH pH 5,7 + 50 μM JA. **(\*\*)** Control negativo de la mezcla de la reacción de transcripción inversa. **(C-)** Control negativo de la mezcla de la reacción de PCR. **(C+)** Control positivo sobre DNA genómico.

## Resultados

---

### A.5.2 Análisis de la transcripción de CIRE1

Los resultados del apartado anterior confirman la existencia de transcritos de elementos del tipo  *copia* en *C. sinensis*. Por los análisis de secuencias sabemos que hay elementos de la familia CIRE1 que se están expresando en raíz en plantas sin tratar. Se pretende saber si esta familia de elementos  *copia* es capaz de expresarse en hoja en respuesta a hormonas vegetales o en condiciones de estrés concretas como la herida mecánica.

#### A.5.2.1 Implicación de la herida mecánica y la adición exógena de hormonas vegetales (JA y NAA) en la activación de la expresión de CIRE1 en *C. sinensis*

Las señales que controlan la expresión de los RTN-LTR están localizadas fundamentalmente en la región promotora de las LTR, aunque solo se utiliza como promotor la de la LTR5'. Se han identificado en la región U3 de la LTR5' ciertos elementos en *cis* que parecen estar involucrados en la inducción de la transcripción en RTN-LTR en respuesta a estrés biótico o abiótico. Existe dentro de esta región una elevada variabilidad entre las diferentes familias de RTN-LTR de plantas que se ha relacionado con la capacidad de los promotores de estos elementos a responder a diferentes factores asociados al estrés (Verhenettes et al., 1998; Beguiristain et al., 2001; Vicient et al., 2005).

Entre otros, se han identificado en la LTR5' de CIRE1 hipotéticos elementos reguladores en *cis* análogos a elementos de respuesta a herida y elementos de respuesta a hormonas como jasmonato de metilo (motivos CGTCA y TGACG) y auxinas (AuxRR-core y TGA box) (Ver **Figura 44, apartado A.4.1**). Tal y como se ha desarrollado ampliamente en el **apartado A.4**, se ha comprobado mediante experimentos de expresión transitoria que esos elementos podrían jugar un papel en la modulación de la transcripción del elemento, ya que el promotor que los contiene responde con un incremento en la expresión por la acción de dichas hormonas.

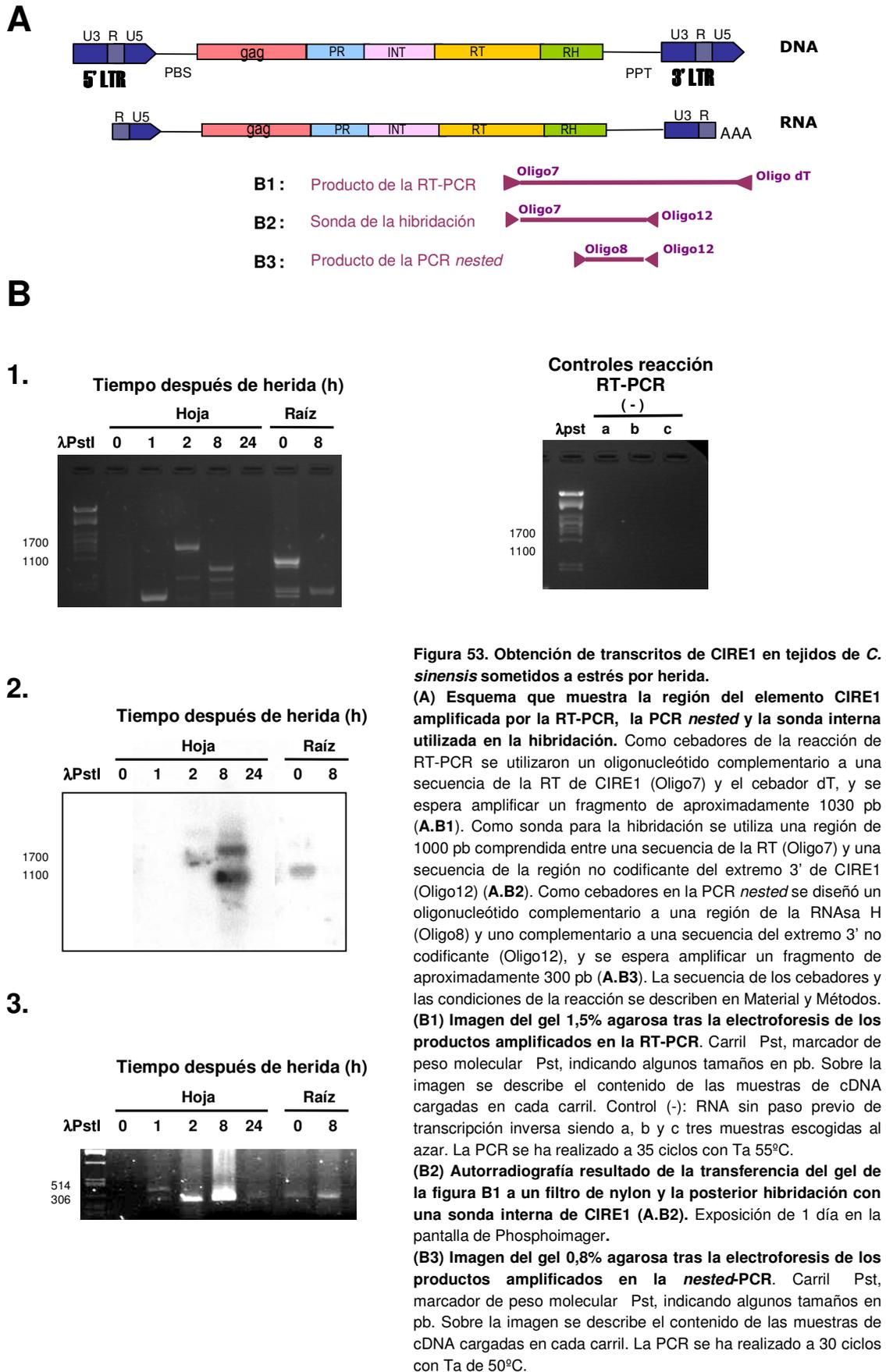
En una primera aproximación, se ha probado mediante análisis por RT-PCR si la expresión de CIRE1 en hoja y raíz puede verse inducida por herida mecánica. Para ello, se cortaron las hojas en pequeños trozos (~ 1 cm<sup>2</sup>), se agruparon en diferentes subgrupos, se les realizaron pequeños cortes y posteriormente se incubaron en tampón MES durante un mínimo de 1h y un máximo de 24h, recogiendo y congelando muestras a diferentes tiempos de incubación. Un grupo de hojas fue congelado antes de la incubación para ser utilizado como control. Para raíz, las heridas se realizaron mediante agujas. Se aisló el RNA total y, de la misma manera descrita anteriormente, se amplificó por RT-PCR una región que corresponde a los últimos 1030 nucleótidos de la molécula

de RNA, que contiene 700 nucleótidos de la parte 3' de la región que codifica para la RT y RH, 100 nucleótidos de la región 3' no traducida situada entre el final de la ORF y el principio de la LTR3', y 230 nucleótidos correspondientes a las regiones U3 y R de la LTR3' (ver **Figura 53A**). Para ello, se utilizaron como cebadores de la reacción de PCR un oligonucleótido complementario a una secuencia de la región de la RT de CIRE1 (Oligo7) y el cebador dT, que se va a unir a la cola poliA del RNAm. En CIRE1, al igual que el resto de RTN-LTR, la secuencia correspondiente a la señal de poliadenilación está localizada en la frontera entre la región R y U5 de la LTR3' (Casacuberta et al., 1995).

Uno de los cebadores utilizados en la reacción de RT-PCR es oligodT que permite seleccionar de manera preferente RNA mensajeros (con cola poliA). No obstante, por si hubiera alguna traza de DNA genómico en las muestras que mostrara similitud con la secuencia del cebador, se ha realizado también la reacción de PCR en paralelo sobre muestras de RNA no sometidas a la reacción de transcripción inversa. Estos controles de la reacción de RT-PCR no muestran bandas de amplificación, validando así la reacción [**Figura 53B1, a, b y c (-)**]. Tal y como se muestra en la **Figura 53B1**, no se observan bandas de amplificación por RT-PCR en hoja sin tratar (tiempo 0) pero sí en raíz, confirmando la expresión diferencial de tejido en condiciones normales ya vista en apartados anteriores. Además de en raíz, también se obtiene una banda de aproximadamente el tamaño esperado ( $\approx 1030$  pb) en hoja 8h después de la herida, aunque aparecen también bandas de menor tamaño. Lo mismo pasa en la muestra de raíz de 8h después de la herida y la muestra de hoja de 1h después de la herida. Estas bandas de menor tamaño podrían deberse a amplificaciones inespecíficas. Lo mismo se podría decir para hoja recogida 2h después de la herida, pero en este caso el tamaño de banda es superior.

Para confirmar que el resultado de la amplificación es específico se transfirió el gel de la electroforesis de la reacción de RT-PCR mostrado en la **Figura 53B1** a una membrana de nylon, y seguidamente se hibridó con una sonda específica de CIRE1 de 1000 nucleótidos, que incluye una región comprendida entre el extremo 3' de la región que codifica para la RT y RH y la región no codificante de 3' (situada entre el final de la ORF y el principio de la LTR3'), tal y como se muestra en la **Figura 53A(B2)**. El resultado de la hibridación se muestra en la **Figura 53B2**. La autorradiografía muestra que en las muestras sometidas a herida mecánica, la hibridación es positiva sobre las bandas amplificadas de alrededor de 1030 nucleótidos en la muestra de hoja tras 8 horas de la herida y en raíz sin herida (tiempo 0). También se observa una banda de tamaño superior al esperado en la muestra de hoja de 2 y 8 h después de la herida.

## Resultados

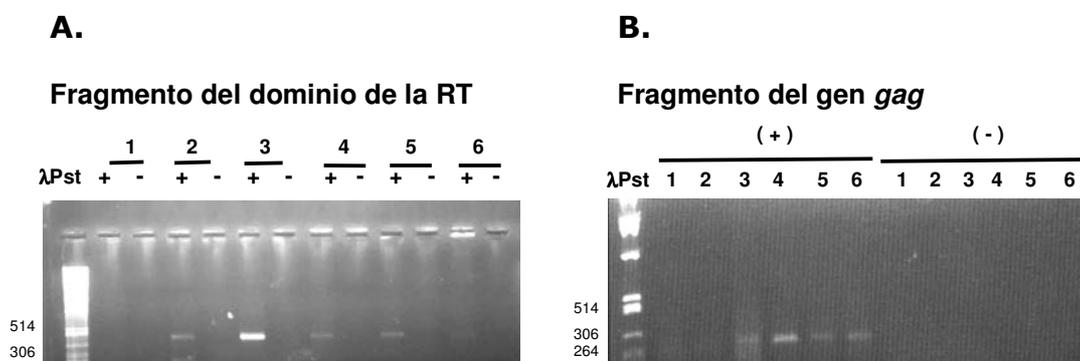


Una nueva confirmación de la especificidad de la RT-PCR la obtenemos al realizar una PCR *nested* o PCR interna sobre el producto amplificado en la RT-PCR utilizando unos cebadores diseñados a partir de las secuencias de CIRE1. El fragmento amplificado de aproximadamente 300 nucleótidos corresponde a una región del elemento comprendida entre una región de la RH y una secuencia del extremo 3' no codificante, por tanto y tal y como se muestra en la **Figura 53A(B3)** se amplifica una zona interna a la región amplificada con los cebadores utilizados en la RT-PCR. El resultado se muestra en la **Figura 53B3**. Se observa que no hay amplificación del fragmento interno esperado en la muestra de hoja no sometida a herida mecánica (tiempo 0). Se observa una banda débil pero del tamaño esperado en muestras de hojas tras 1h y 24h de haberlas sometido a herida mecánica, y en el resto de muestras y tiempos se observa una banda clara del tamaño esperado, confirmando que entre los fragmentos amplificados anteriormente en la RT-PCR hay transcritos específicos de CIRE1. Estos resultados reflejan que en el tratamiento por herida existen bandas en la primera amplificación por RT-PCR (**Figura 53B1**) que no corresponden al tamaño esperado, y solo existe amplificación suficiente y del tamaño esperado para ser detectada específicamente por una sonda de CIRE1 en hoja a las 8h de la herida y en raíz sin tratar. En cambio, cuando se realiza una PCR *nested* sobre los primeros productos de la RT-PCR, se observa también amplificación de la banda de tamaño esperado en ambos tejidos y todos los tiempos de tratamiento, exceptuando hoja no herida (tiempo 0, hoja congelada inmediatamente tras ser cortada).

Para evaluar más finamente el patrón de expresión de CIRE1 en respuesta a herida mecánica o bien a la aplicación exógena de hormonas vegetales, y así confirmar los resultados de la RT-PCR, se ha analizado por *northern blot* el RNA total extraído de diferentes muestras de tejido de naranjo sometidas a tratamiento por hormonas o a estrés. Para determinar el efecto de las hormonas sobre la posible activación transcripcional de CIRE1, se cortaron hojas en pequeños trozos (~ 1 cm<sup>2</sup>), se agruparon en diferentes subgrupos, se infiltraron durante 1h con tampón MES sólo (muestra control) o suplementado con 50 µM JA o 1mg/L NAA, posteriormente se incubaron 3h más a 28°C, y seguidamente se recogieron y congelaron las muestras hasta su posterior análisis. En el caso de la herida mecánica, plantas jóvenes crecidas en invernadero fueron sometidas *in situ* a pequeños cortes en las hojas y raíz, y posteriormente se recogió muestra de ambos tejidos a las 4h y 8h después de la herida. Asimismo, se ha buscado la presencia de RNA de CIRE1 en hoja y raíz de naranjo (*C. sinensis*, variedad "Valencia late") cultivado bien en invernadero o bien *in vitro*. En ninguno de los casos se han conseguido niveles de transcritos de CIRE1 suficientes para ser detectados por *northern-blot* (resultados no mostrados).

## Resultados

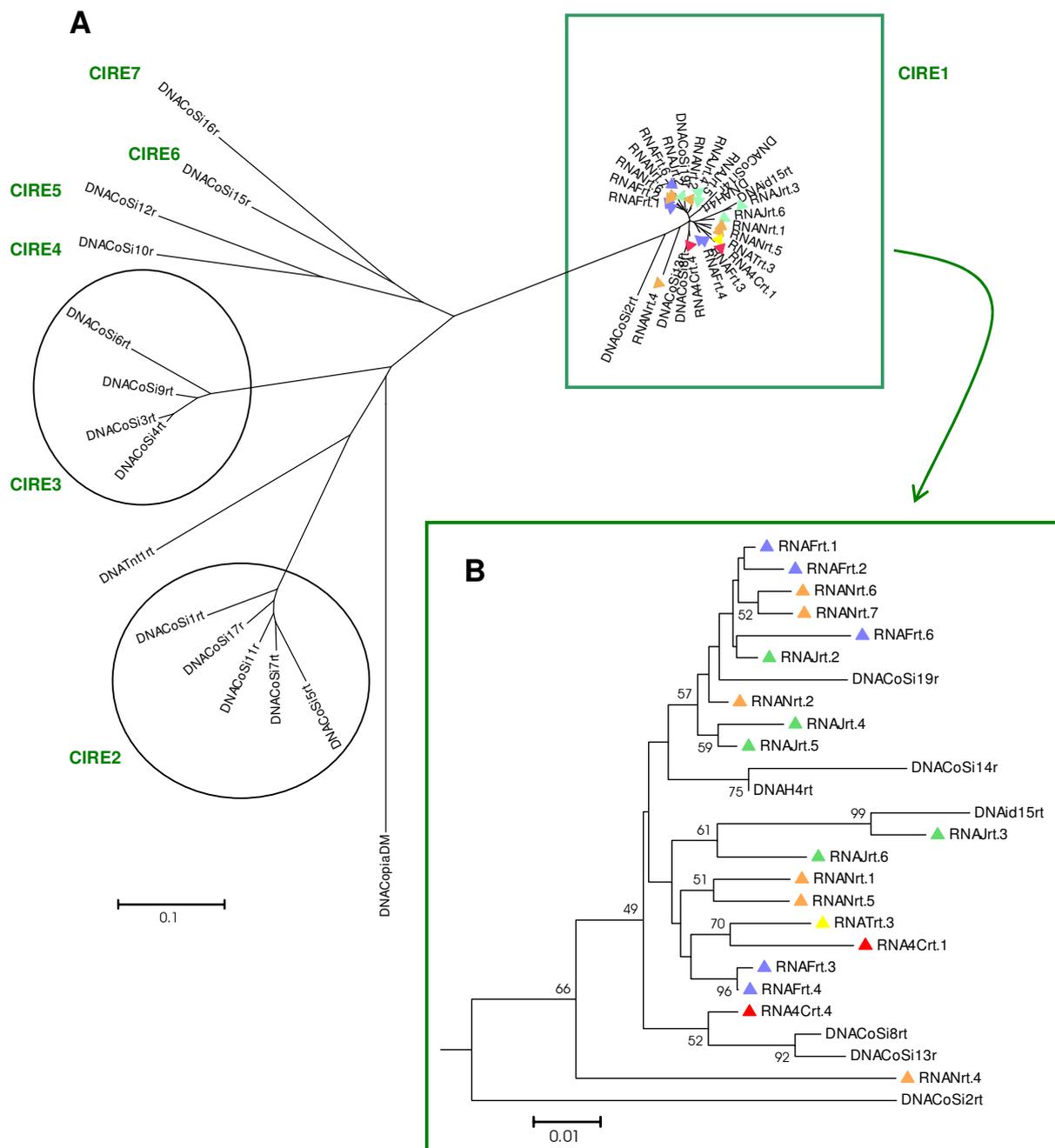
Finalmente y para confirmar la integridad del RNA transcrito de CIRE1 (es decir, si más de un dominio del RTN-LTR está presente en los transcritos, concluyéndose entonces que se podría tratar del RNA genómico completo), se sometieron discos de hojas de naranjo a diversos tratamientos (ver pie de figura 54) y a continuación se realizó RT-PCR a partir del RNA total utilizando diversos cebadores diseñados a partir de las secuencias disponibles de CIRE1. Nuevamente se utilizaron en la amplificación las combinaciones específicas de oligos para cada región del elemento descritas en el **apartado A.3.7**. Los fragmentos amplificados corresponden a diversas regiones del elemento como son el gen *gag* y el dominio de la RT. En la **Figura 54** se muestra el resultado de la amplificación. Se obtienen bandas del tamaño esperado para el fragmento del dominio de la RT en todas las muestras sometidas a estrés [**Figura 54A, 2-6(+)**] y no en la muestra control [**Figura 54A, 1(+)**], mientras que la amplificación del fragmento del gen *gag* es positiva en las mismas muestras tratadas [**Figura 54B, 2-6(+)**] aunque en el caso de la muestra de hoja cortada y simplemente incubada en tampón [**Figura 54B, 2(+)**] se observa una banda muy débil. No hay amplificación de la región del gen *gag* sobre el cDNA de la muestra control [**Figura 54B, 1(+)**]. En ninguno de los dos casos, se observa amplificación en las muestras no sometidas al paso de retrotranscripción [**Figura 54, 1-6 (-)**] descartando falsos positivos debidos a contaminaciones con DNA genómico.



**Figura 54. Obtención de transcritos de CIRE1 en hojas de *C. sinensis* sometidas a diferentes tratamientos.** Imagen del gel al 1,2% de agarosa tras la electroforesis de los productos amplificados en la RT-PCR utilizando como cebadores los oligonucleótidos específicos de CIRE1 para amplificar una región de 403 nucleótidos de la RT (Oligo26+Oligo29) y una región de 297 nucleótidos del gen *gag* (Oligo18+Oligo27). La secuencia de estos cebadores y las condiciones de la amplificación se describen en Material y Métodos. La PCR se ha realizado a 30 ciclos con Ta de 52°C. Carril λPst, marcador de peso molecular λPst, se indican algunos tamaños en pb. **Carril (+):** cDNA (RNA sometido a reacción de transcripción inversa), **Carril (-):** RNA (RNA sin paso previo de RT). **1. Control:** Hoja congelada inmediatamente después de ser cortada. **2. Corte:** Hoja cortada e incubada en tampón 0,05% MES-KOH pH 5,7, 8h a 28°C y luz. **3. Herida:** Hoja cortada y con heridas, incubada en tampón 0,05% MES-KOH pH 5,7, 8h a 28°C y luz. **4. NAA:** Hoja cortada e incubada en tampón 0,05% MES-KOH pH 5,7 + 1mg/L NAA, 8h a 28°C y luz. **5. JA:** Hoja cortada e incubada en tampón 0,05% MES-KOH pH 5,7 + 50 μM JA, 8h a 28°C y luz. **6. Frío:** Hoja cortada e incubada en tampón 0,05% MES-KOH pH 5,7, 8h a 4°C y luz.

Para confirmar la autenticidad como CIRE1 de los fragmentos de la región del dominio de la RT generados en la RT-PCR anterior, se clonaron en el vector pGEMT y se determinó la secuencia de 19 clones escogidos al azar. Las secuencias nucleotídicas obtenidas corresponden a 1 clon de la muestra de hoja sometida al corte (RNATrt#), 5 clones de la muestra de hoja herida (RNAFrt#), 5 clones de la muestra de hoja suplementada con JA (RNAJrt#), 6 clones de la muestra de hoja suplementada con NAA (RNANrt#) y 2 clones de la muestra de hoja sometida a frío (RNA4Crt#). A partir del alineamiento múltiple por ClustalW de estas secuencias junto con las secuencias nucleotídicas procedentes de DNA genómico descritas en el apartado A.1 (DNACoSi#rt) se obtiene el árbol tipo *Neighbour joining* que se muestra en la **Figura 55A**. Hay que tener en cuenta que en este caso, las secuencias de RNArt# y DNACoSi#rt aunque corresponden todas a una misma región de la RT, no han sido obtenidas utilizando la misma combinación de cebadores. El análisis realizado nos permite confirmar claramente que los transcritos amplificados pertenecen a la familia CIRE1. Por otro lado, si consideramos únicamente la rama del árbol que incluye los transcritos (**Figura 55B**) podemos afirmar que no hay una expresión diferencial de determinados elementos bajo los diferentes tratamientos ya que las diferentes secuencias quedan repartidas por todo el árbol. No obstante, se aprecia una mayor similitud entre pares de secuencias de transcritos procedentes de hojas con un tratamiento determinado (**Figura 55B**).

## Resultados



**Figura 55. Árbol fenético basado en el alineamiento de cDNA de transcritos de CIRE1 obtenidos por RT-PCR en muestras sometidas a diversos tratamientos y fragmentos genómicos de RTN-LTR *copia* amplificados por PCR en *C. sinensis* de la región del dominio de la RT. (A) Árbol fenético radial tipo NJ generado con el programa MEGA2. Se indica el nombre de las diferentes familias de RTN-LTR tipo *copia* presentes en *C. sinensis*. DNACoSi#: clones de la amplificación de DNA genómico de *C. sinensis*. (Δ) RNA#: transcritos obtenidos en la reacción de RT-PCR. Las secuencias nucleotídicas de la región correspondiente de la RT de los fragmentos id15 y H4 (apartado A.3.1.4), de *Tnt1* (X13777) y *CopiaDM* (P04146) se han unido también a la comparación. (B) Árbol tipo NJ obtenido de la ampliación de la rama del árbol (A). Están indicados en cada bifurcación los valores de bootstrap  $\geq 40\%$ . (Δ) RNATrt#: transcritos obtenidos en RT-PCR sobre muestra de hoja cortada. (Δ) RNAFr#: transcritos obtenidos en RT-PCR sobre muestra de hoja herida. (Δ) RNAJrt#: transcritos obtenidos en RT-PCR sobre muestra de hoja suplementada con JA. (Δ) RNANr#: transcritos obtenidos en RT-PCR sobre muestra de hoja suplementada con NAA. (Δ) RNA4Crt#: transcritos obtenidos en RT-PCR sobre muestra de hoja en frío.**