

Universitat Autònoma de Barcelona

CONTROL DE LA TRANSPOSICIÓN E IMPACTO DE LOS ELEMENTOS TRANSPONIBLES EN GENOMAS VEGETALES: ANÁLISIS DEL RETROTRANSPOSÓN Tnt1 DE TABACO Y DEL MITE *Emigrant* DE ARABIDOPSIS.



TESIS DOCTORAL

NÉSTOR SANTIAGO GONZALVO

Barcelona, Febrero de 2006



Consortio CSIC – IRTA
Consejo Superior de Investigaciones Científicas
Laboratorio de Genética Molecular Vegetal
Instituto de Biología Molecular de Barcelona
Departamento de Genética Molecular

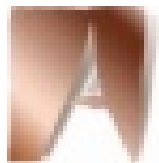


Universidad Autónoma de Barcelona
Departamento de Genética y Microbiología
Unidad de Genética

CONTROL DE LA TRANSPOSICIÓN E IMPACTO DE LOS ELEMENTOS TRANSPONIBLES EN GENOMAS VEGETALES: ANÁLISIS DEL RETROTRANSPOSÓN Tnt1 DE TABACO Y DEL MITE *Emigrant* DE ARABIDOPSIS.

NÉSTOR SANTIAGO GONZALVO

Barcelona, Febrero de 2006



Universitat Autònoma de Barcelona

Facultat de Ciències

Dptment. de Genètica i Microbiologia

Unitat de Genètica

**CONTROL DE LA TRANSPOSICIÓN E
IMPACTO DE LOS ELEMENTOS
TRANSPONIBLES EN GENOMAS
VEGETALES: ANÁLISIS DEL
RETROTRANSPOSÓN Tnt1 DE TABACO
Y DEL MITE *Emigrant* DE ARABIDOPSIS.**

Memoria presentada por NÉSTOR SANTIAGO GONZALVO, inscrita en el programa de Genética de tercer ciclo de la Universidad Autónoma de Barcelona en el bienio 2000-2002, para optar al grado de Doctor en Genética por la Universitat Autònoma de Barcelona.

Este trabajo se ha realizado en el Instituto de Biología Molecular de Barcelona del Consejo Superior de Investigaciones Científicas, bajo la dirección del Dr. JOSEP M^a CASACUBERTA y la tutoría del Dr. ORIOL CABRÉ.

Con el consentimiento de,

Director de tesis

Tutor

Autor

Dr. Josep M^a Casacuberta Suñer

Dr. Oriol Cabré Fabré

Néstor Santiago Gonzalvo

A mis padres.

“Helios brillaba muy lejos y alto, y desde el pálido cielo no derramaba calor alguno, sino sólo el destello de su plateado cabello”.

Mary Renault. *Alexias de Atenas.*

AGRADECIMIENTOS.

Este trabajo ha podido ser realizado gracias a la ayuda y al apoyo que he recibido por parte de un gran número de personas, a las que debo este agradecimiento.

En primer lugar quiero agradecer a Pep, Dr. Josep M. Casacuberta, el apoyo y la confianza que ha depositado en mí. La continua ayuda que me ha brindado y lo mucho que he aprendido con y gracias a él. El haberme infundado cierto espíritu crítico, tan necesario en la Ciencia. También, la paciencia y buen humor que siempre le acompañan. ¡Muchas gracias, Pep!

Al Dr. Oriol Cabré por admitirme bajo su tutela en la Universidad y mostrar interés por mi trabajo.

A mis compañeros de laboratorio. Enric, gracias por aguantarme, incluso en los momentos más tensos, por tu ayuda y tus consejos. Ahora quiero desearte mucha suerte y paciencia para la escritura de tu manuscrito, que estoy seguro será una gran tesis.

Céline, ¡qué más te voy a contar que no sepas ya! Muchísimas gracias por todo, por tu simpatía, amabilidad, ayuda, consejos, por los cursos de francés... Por tantos momentos de convivencia, dentro y fuera del laboratorio. Eres una gran compañera de piso. Sobre todo, muchas gracias por la gran amistad que me has ofrecido desde el primer momento y que se ha ido consolidando poco a poco.

A mis primeros compañeros, en el laboratorio *Verde*, Blanca, Miriam y Torben. Los que me enseñaron a manejarme en el laboratorio y me hicieron muy agradable el inicio de esta etapa.

A los demás compañeros temporales que han pasado por el laboratorio *Gris*. A Alicia, Mima, Maryaline, Vivienne por su ayuda con los experimentos. A Marcela, Cédric y Nico por amenizar el laboratorio.

A Maribel y sus chicas, Isabel, Helga y Fátima, con las que compartimos, todavía no sé cómo, laboratorio. Y también a Jonatan.

A la Dra. M. Lluïsa Espinás, por su ayuda con los experimentos de ChIP.

A todo el departamento de Genética Molecular del IBMB, los de ahora y los de antes.

A las Wendys, Núria, Irma y Elisabet. Ellas sí que me han tenido que aguantar. Me han ayudado en todo lo posible y me han alegrado en todo momento. Buenos momentos hemos pasado y buenos momentos pasaremos. Os espero allá donde vaya. Guardo mi carné del club.

A Jaume, Jordi y los demás *lilas*.

A Carles y los *naranjas*. A las *amarillas*, Laura y Eli. Con ellas a Mina y, ahora también, a Bea. A los *rosas*, Sonia, Bia, Jordi, Silvia, Lidia y Jorge. ¡Ánimo Sonia! Aquí incluyo a Marta, que aunque ahora esté entre ratones, también ha formado una parte importante del departamento.

A Maite, Marc y los *rojos* de todos los tiempos. A los *verdes* Cristian, Ignacio, Carlos. A Cristina, Alicia, Sami y demás *azules*. A Fathy, Valeria y demás *negros*. A Desi y otros *fruitis*.

A mis compañeros de la sexta planta, Maida, Cristina C., Cristina S., Soraya, Sergi, Paloma, Mariano, Nahuel, Juanjo, Ana, Cristina V. y todos los demás. Uno se engancha al café si lo toma junto a vosotros, sois unos grandes amigos.

A Pilar y Eva, por su continua disponibilidad y gran ayuda con las plantas. También quiero dar las gracias a Alejandro, Leire y César.

A Mercé, Mónica, Mireia, Ángel y demás personal del resto de servicios del Instituto.

Todos ellos, de una forma u otra me han ayudado en mi trabajo, bien materialmente o bien anímicamente, que es casi más importante. Va a ser difícil olvidaros a cada uno de vosotros. Gracias.

A la gente de Madrid, del CIB, que me hicieron sentir como en casa. César, por su ayuda con los sRNAs. Gracias por ser tan buen maestro y amigo. Elisa y Belén, gracias por vuestra naturalidad y ayuda. Parece que os conozca de toda la vida.

To my friends from Leicester. Teo, Alessandra, Celia, Estelle and the others. I spend nice times with you. I hope to see you someday again. Thank you very much for your help and kindness.

A mis amigos, que me han hecho disfrutar de muy buenos momentos.

A María, Jesús, Cristina, Pablo y los demás oscenses. Tantos años y tantas experiencias. Nunca se olvida. A Jana, Gemma, Cris, mis compañeras de la Universidad. A Érika, Mónica, Ángeles, Sandra, Carmen y Jesús. A Nico. A Javi, Zaira y el resto de la *penya*. A mis excompañeros de piso, en especial a Joan, Luis y Herbert.

A mi familia, que siempre ha estado a mi lado y me ha ayudado en todo momento. ¡Gracias papá, mamá y tata! A mis tíos, que me acogieron en su casa como un hijo más. A mis abuelas y abuelo, que siempre han confiado en mí.

ABREVIATURAS.

2,4-D	2,4-ácido Diclorofenoxiacético
°C	Celsius degrees (grados centígrados)
A	Amperio(s)
Ab	Antibody (anticuerpo)
APS	Amonium Persulfate (persulfato de amonio)
aRNA	RNA aberrante
Au	Aurum (oro)
BAC	Bacterial Artificial Chromosome (cromosoma artificial bacteriano)
BAP	6-BezylAminoPurine
BrEt	Bromuro de Etidio
BSA	Bovine Serum Albumin (albúmina sérica bovina)
CaMV	Cauliflower Mosaic Virus (virus del mosaico de la coliflor)
cfu	colony forming unit (unidad formadora de colonias)
ChIP	Chromatin Immuno-Precipitation (inmuno-precipitación de la cromatina)
cm	centímetro(s)
cr	cromosoma
Cryp	Cryptogein (criptogeína)
DMSO	Dimetilsulfóxido
DNA	DeoxyriboNucleic Acid (ácido desoxirribonucleico)
DO	Densidad Óptica
dNTP	deoxyNucleotide Tri-Phosphate (desoxinucleótido tri-fosfato)
dsRNA	double stranded RNA (RNA de cadena doble)
EDTA	Ácido Etileno Diamino Tetracético
EST	Expressed Sequence Tag (secuencia expresada, transcrita)
ET	Elemento Transponible
EtOH	Etanol
F	Faradio(s)
g	gramo(s)
GUS	β-Glucoronidasa
h	hora(s)
H ₂ O	agua
HAc	ácido acético
HPLC	High Performance Liquid Chromatography (Cromatografía líquida de alta resolución)
KAc	Kalium Acetate (acetato de potasio)
NaClO	Hipoclorito de sodio (lejía)
Kb	Kilobase(s)
Kg	Kilogramo(s)
L	Litro
LB	Luria-Bertani
LMW	Low Molecular Weight (bajo peso molecular)
Mb	Megabase
MeJa	Metil Jasmonato
MES	2-(N-Morpholino)ethane-sulfonic acid
μF	microFaradio
μg	microgramo(s)
μl	microlitro(s)
μM	microMolar
miRNA	micro RNA
mJ	miliJulios
ml	mililitro
mM	miliMolar
min	minuto(s)
MITE	Miniature Inverted-repeat Transposable Element
MOPS	3-(N-Morpholino)propanesulfonic acid
MW	Molecular Weight (peso molecular)
NAA	1-NaphthaleneAcetic Acid (ácido naftalenacético)

NaAc	N atrium A cetate (acetato de sodio)
ng	n anogramo(s)
nm	n anómetro(s)
NOS	N opalin S intasa
nt	n ucleótido(s)
Ω	O hmio(s)
oligo	o ligonucleótido
ORF	O pen R eading F rame (pauta abierta de lectura)
pb	p ares de b ases
PCR	P olimerase C hain R eaction (reacción en cadena de la polimerasa)
pmol	p icomol(es)
PMSF	P henyl m ethylsulfonyl fluoride
PNK	P oly N ucleotide K inase (polinucleótido quinasa)
protA	P roteína A
psi	p ounds per s quare inch (libras por pulgada cuadrada).
PTGS	P ost T ranscriptional G ene S ilencing (silenciamiento post-transcripcional)
pTi	T umor inducing p lasmid (plásmido de virulencia)
rDNA	r ibosomal D N.A (ADN ribosomal)
RdRP	R N.A-dependent R N.A P olymerase (RNA polimerasa dependiente de RNA)
RESites	R elated E mpy S ites (sitios vacíos asociados)
RET	R etro E lemento T ransponible (retrotransposón)
Rif	R ifampicina
RISC	R N.A I nduced S ilencing C omplex (Complejo del silenciamiento inducido por RNA)
RITS	R N.A-induced I nitiation of T ranscriptional G ene S ilencing (Inicio del silenciamiento génico transcripcional inducido por RNA)
RNA	R ibo N ucleic A cid (ácido ribonucleico)
rpm	r evoluciones p or m inuto
RT	R etro T ranscriptasa
s	s egundo(s)
Sal	S alicylic acid (ácido salicílico)
SINE	S hort I nterspersed repetitive E lement
Sm	S treptomycin sulphate (estreptomicina)
SDS	S odium dodecyl sulfate
sp.	e specie
sRNA	s mall R N.A (pequeño RNA)
siRNA	s hort interfering R N.A (pequeño RNA de interferencia)
SSC	S odium s alt citrate
ssDNA	s ingle s tranded D N.A (DNA de cadena sencilla)
ssRNA	s ingle s tranded R N.A (RNA de cadena sencilla)
T ^a	T emperatura
TAE	T ris- a cetate E DTA
Taq	T hermus a quaticus
Taq pol	p olimerasa de T hermus a quaticus
TBE	T ris- b orate E DTA
TE	T ris E DTA
TFB	T ransformation B uffer (solución de transformación)
TGS	T ranscriptional G ene S ilencing (silenciamiento génico transcripcional)
TIR	T erminal I nverted R epeat (repetición terminal invertida)
T _m	m elting T emperature (temperatura de hibridación)
TSD	T arget S ite D uplication (duplicación del sitio de inserción)
u	u nidad(es)
U.V.	U ltra V ioleta
v	v olumen
var.	v arietad
V	V oltios
w	w eight (peso)
WT	W ild T ype (silvestre)
w/v	w eight/ v olume (peso/volumen)
YEB	Y east E xtract B roth

ÍNDICE.

INTRODUCCIÓN.	1
1. ¿QUÉ SON LOS GENOMAS?	3
2. LOS ELEMENTOS TRANSPONIBLES.	10
2.1. CLASIFICACIÓN DE LOS ELEMENTOS TRANSPONIBLES.	10
2.1.1. Elementos transponibles de clase I. Retrotransposones.	11
2.1.2. Elementos transponibles de clase II. Transposones de DNA.	13
2.2. ABUNDANCIA DE LOS TRANSPOSONES EN LOS GENOMAS.	14
2.3. DISTRIBUCIÓN DE LOS ELEMENTOS TRANSPONIBLES.	15
3. EFECTO DE LOS ELEMENTOS TRANSPONIBLES SOBRE LA EVOLUCIÓN DE GENES Y GENOMAS.	17
3.1. FUNCIONES CELULARES DESARROLLADAS POR TRANSPOSONES.	18
3.2. FUNCIONES ADAPTATIVAS Y GENES DERIVADOS DE ELEMENTOS TRANSPONIBLES.	20
3.3. EFECTOS DE LOS TRANSPOSONES SOBRE LA EXPRESIÓN Y REGULACIÓN DE OTROS GENES.	21
3.4. EFECTO DE LOS TRANSPOSONES SOBRE LA EVOLUCIÓN DE LAS SECUENCIAS CODIFICANTES.	24
4. MECANISMOS DE CONTROL DE LOS TRANSPOSONES: SILENCIAMIENTO GÉNICO.	26
OBJETIVOS.	31
RESULTADOS Y DISCUSIÓN.	35
CAPÍTULO 1. ANÁLISIS DE LA FAMILIA <i>Emigrant</i> DE MITES DE <i>Arabidopsis thaliana</i>.	37
INTRODUCCIÓN.	37
1. DISEÑO DE UN PROGRAMA INFORMÁTICO PARA LA BÚSQUEDA DE MITES.	38
2. IDENTIFICACIÓN Y CARACTERIZACIÓN DE NUEVOS MITES <i>Emigrant</i> EN EL GENOMA DE ARABIDOPSIS.	39
3. ANÁLISIS FILOGENÉTICO. Dentro del genoma de <i>Arabidopsis thaliana</i> existen diferentes subfamilias de elementos <i>Emigrant</i> .	41
4. ANÁLISIS EVOLUTIVO.	43
4.1. Análisis de la variabilidad de secuencia y tamaño entre los distintos grupos de <i>Emigrant</i> .	43
4.2. Distribución.	44
4.3. Sitios de inserción.	46
4.4. Localización de los elementos <i>Emigrant</i> respecto a los genes.	48
4.4. Evidencias de actividad en un pasado reciente.	53

CAPÍTULO 2. ESTUDIO DEL CONTROL DE LA EXPRESIÓN DEL RETROTRANSPOSÓN Tnt1 DE TABACO.	59
INTRODUCCIÓN.	59
1. ESTUDIO DE LA TERMINACIÓN DE LA TRANSCRIPCIÓN.	61
1.1. ANÁLISIS DE LOS SITIOS DE POLIADENILACIÓN DE Tnt1.	62
1.2. DIFERENCIAS EN LA POLIADENILACIÓN ENTRE LAS DIFERENTES FAMILIAS DE Tnt1.	64
1.3. EFECTO DEL ESTRÉS SOBRE LA POLIADENILACIÓN DE Tnt1.	67
1.4. ESTUDIOS DE LA INTERACCIÓN ENTRE EL INICIO Y LA TERMINACIÓN DE LA TRANSCRIPCIÓN.	70
2. ESTUDIO DEL SILENCIAMIENTO GÉNICO.	73
2.1. ESTUDIO DEL EFECTO DEL ESTRÉS SOBRE EL SILENCIAMIENTO.	74
2.1.1. Establecimiento de un sistema de inducción de la expresión de Tnt1.	74
2.1.2. Estudio del locus silenciador 271.	75
2.1.3. Diseño de una estrategia de silenciamiento transitoria y local.	76
2.2. ESTUDIO DEL SILENCIAMIENTO DE Tnt1.	83
2.2.1. Análisis de los intermediarios del silenciamiento de Tnt1. I: construcción de una librería de sRNAs de <i>Nicotiana tabacum</i> .	84
2.2.2. Análisis de los intermediarios del silenciamiento de Tnt1. II: análisis de la acumulación de siRNAs.	87
2.2.3. Análisis de los intermediarios del silenciamiento de Tnt1. III: modificaciones de la cromatina.	90
2.3. ESTUDIO DEL EFECTO DE LA SECUENCIA DE LAS LTRs DE Tnt1 SOBRE EL SILENCIAMIENTO.	92
2.3.1. Construcciones génicas.	93
2.3.2. Generación de las plantas transgénicas.	94
2.3.3. Análisis de la expresión de las construcciones quiméricas.	96
DISCUSIÓN GENERAL.	101
1. EL SILENCIAMIENTO GÉNICO COMO CONTROL PRINCIPAL DE LA TRANSPOSICIÓN.	105
2. IMPACTO DE LOS ELEMENTOS TRANSPONIBLES SOBRE GENES Y GENOMAS.	107
CONCLUSIONES.	111
MATERIAL Y MÉTODOS.	115
1. MATERIALES.	117
1.1. MATERIAL VEGETAL.	117
1.1.1. ESTERILIZACIÓN Y GERMINACIÓN DE SEMILLAS.	117
1.2. MATERIAL BACTERIANO.	117
1.3. PLÁSMIDOS.	118
1.3.1. VECTORES UTILIZADOS PARA REALIZAR LAS CONSTRUCCIONES GÉNICAS.	118
1.3.2. VECTORES BINARIOS.	118
1.3.3. VECTORES UTILIZADOS PARA LA CLONACIÓN DE PRODUCTOS DE PCR.	118

1.4. MEDIOS.	119
1.5. ANTIBIÓTICOS Y REACTIVOS.	119
1.6. OLIGONUCLEÓTIDOS.	119
2. MÉTODOS GENERALES.	121
2.1. EXTRACCIÓN DE ÁCIDOS NUCLEICOS.	121
2.1.1. EXTRACCIÓN DE DNA PLASMÍDICO.	121
2.1.1.2. Extracción de DNA plasmídico de <i>A.tumefaciens</i> .	121
2.1.2. EXTRACCIÓN DE DNA GENÓMICO.	122
2.1.2.1. Extracción rápida de DNA genómico.	123
2.1.3. EXTRACCIÓN DE RNA.	123
2.1.3.1. Tratamiento con DNasa.	124
2.2. REACCIÓN EN CADENA DE LA POLIMERASA (PCR).	125
2.2.1. Mezcla de la reacción de PCR.	125
2.2.2. Condiciones de la reacción.	125
2.3. REACCIÓN DE RT-PCR.	125
2.3.1. SÍNTESIS DE LA MOLÉCULA DE cDNA.	125
2.4. PURIFICACIÓN DE LOS PRODUCTOS DE PCR.	126
2.5. CLONACIÓN DE PRODUCTOS DE PCR.	126
2.6. ELECTROFORESIS DE ÁCIDOS NUCLEICOS EN GEL DE AGAROSA.	126
2.7. EXTRACCIÓN DE FRAGMENTOS DE DNA DESDE GEL DE AGAROSA.	126
2.8. REACCIONES DE MODIFICACIÓN DEL DNA.	126
2.9. LIGACIÓN DE FRAGMENTOS DE DNA.	127
2.10. TRANSFORMACIÓN BACTERIANA.	127
2.10.1. TRANSFORMACIÓN DE <i>E.coli</i> POR CHOQUE TÉRMICO.	127
2.10.1.1. Preparación de células competentes de <i>E.coli</i> para transformación por choque térmico.	127
2.10.1.2. Transformación de <i>E.coli</i> por choque térmico.	128
2.10.2. TRANSFORMACIÓN DE <i>E.coli</i> POR ELECTROPORACIÓN.	128
2.10.2.1. Preparación de células competentes de <i>E.coli</i> para transformación por electroporación.	128
2.10.2.2. Transformación de <i>E.coli</i> por electroporación.	129
2.10.3. TRANSFORMACIÓN DE <i>A.tumefaciens</i> POR CHOQUE TÉRMICO.	129
2.10.3.1. Preparación de células competentes de <i>A.tumefaciens</i> para transformación por choque térmico.	129
2.10.3.2. Transformación de <i>A.tumefaciens</i> por choque térmico.	129
2.11. CUANTIFICACIÓN DE ÁCIDOS NUCLEICOS.	130
2.12. SECUENCIACIÓN DE ÁCIDOS NUCLEICOS.	130
2.13. MARCAJE RADIACTIVO DE ÁCIDOS NUCLEICOS.	130
2.13.1. MARCAJE DE DNA.	130
2.13.2. MARCAJE DE sRNAs.	130
2.14. PURIFICACIÓN O AISLAMIENTO DE ÁCIDOS NUCLEICOS MARCADOS RADIACTIVAMENTE.	130
2.14.1. Purificación de sondas de DNA marcadas radiactivamente.	130
2.14.2. Purificación de sRNAs marcados radiactivamente.	130
3. TÉCNICA DE SOUTHERN.	131
3.1. DIGESTIÓN DEL DNA GENÓMICO.	131
3.2. ELECTROFORESIS EN GEL DE AGAROSA.	131
3.3. TRANSFERENCIA SOUTHERN.	131
3.4. HIBRIDACIÓN SOUTHERN.	132
3.5. EXPOSICIÓN Y REVELADO.	132
3.6. DESHIBRIDACIÓN DE LAS MEMBRANAS.	132

4. TÉCNICA DE NORTHERN.	133
4.1. ELECTROFORESIS DE RNA EN GEL DESNATURALIZANTE.	133
4.2. TRANSFERENCIA NORTHERN.	133
4.3. HIBRIDACIÓN NORTHERN.	133
4.4. EXPOSICIÓN Y REVELADO.	133
5. ANÁLISIS DE LOS sRNAs.	134
5.1. AISLAMIENTO DE LMW-RNAs.	134
5.2. ELECTROFORESIS EN GEL DESNATURALIZANTE DE POLIACRILAMIDA.	134
5.3. PURIFICACIÓN DE LOS sRNAs POR ELUCIÓN DESDE GEL DE POLIACRILAMIDA.	134
5.4. TÉCNICA DE NORTHERN.	135
5.4.1. ELECTROFORESIS EN GEL DESNATURALIZANTE DE POLIACRILAMIDA.	135
5.4.2. TRANSFERENCIA.	135
5.4.3. HIBRIDACIÓN.	135
5.4.4. EXPOSICIÓN Y REVELADO.	135
5.5. TÉCNICA DE NORTHERN REVERSO.	135
5.6. CONSTRUCCIÓN DE UNA LIBRERÍA DE sRNAs DE TABACO.	136
6. MÉTODOS PARA EL TRATAMIENTO DEL MATERIAL VEGETAL.	137
6.1. INDUCCIÓN DE LA EXPRESIÓN DEL RETROTRANSPOSÓN Tnt1 DE TABACO.	137
6.2. TRANSFORMACIÓN DE <i>Nicotiana tabacum</i> .	137
6.2.1. TRANSFORMACIÓN ESTABLE DE TABACO.	137
6.2.1.1. Preparación del material bacteriano.	137
6.2.1.2. Preparación del material vegetal.	137
6.2.1.3. Co-cultivo.	138
6.2.1.4. Inducción de callos transgénicos.	138
6.2.1.5. Regeneración de plantas a partir de los callos.	139
6.2.2. TRANSFORMACIÓN TRANSITORIA DE TABACO POR AGROINFILTRACIÓN.	139
6.2.3. TRANSFORMACIÓN TRANSITORIA DE HOJAS DE TABACO POR BIOLÍSTICA.	140
6.2.3.1. Preparación de la reserva de partículas de oro.	140
6.2.3.2. Recubrimiento de las partículas de oro con DNA, por precipitación.	140
6.2.3.3. Proceso de transformación de las hojas de tabaco: bombardeo.	141
7. ENSAYO DE INMUNOPRECIPITACIÓN DE LA CROMATINA (ChIP).	141
7.1. UNIÓN COVALENTE DE PROTEÍNAS Y DNA.	142
7.2. INMUNOPRECIPITACIÓN DE LOS COMPLEJOS PROTEÍNA-DNA.	142
8. MÉTODOS BIOINFORMÁTICOS.	143
8.1. HERRAMIENTAS INFORMÁTICAS.	144
8.2. BÚSQUEDA E IDENTIFICACIÓN DE ELEMENTOS <i>Emigrant</i> .	144
8.3. ANÁLISIS FILOGENÉTICO.	144
BIBLIOGRAFÍA.	147

APÉNDICE 1 - Tabla completa de los MITEs *Emigrant* identificados en *Arabidopsis*.

APÉNDICE 2 - Tabla de la librería de los sRNAs clonados de *Nicotiana tabacum*.

APÉNDICE 3 - Tabla de las plantas transgénicas generadas.

APÉNDICE 4 - Análisis estadísticos.

NOTA.- Los apéndices 1, 2 y 3 del manuscrito impreso sólo muestran una imagen representativa de las tablas completas, contenidas en el CD que se adjunta con esta copia.

CONTENIDO DEL CD.

TESIS.pdf

Documento PDF que contiene toda la información presentada en la copia impresa de este manuscrito.

APÉNDICES. Carpeta que contiene las tablas completas, presentadas en los apéndices 1, 2 y 3.

Apéndice1 - Tabla completa MITEs *Emigrant*.xls

Apéndice2 - Tabla librería sRNAs tabaco.xls

Apéndice3 - Tabla plantas transgénicas.xls

PUBLICACIONES. Carpeta que contiene dos archivos en formato PDF correspondientes a los dos artículos científicos publicados y derivados del trabajo que se presenta en esta tesis doctoral.

Plant LTR-retrotransposons and MITEs.pdf

Genome-wide Analysis of the *Emigrant* Family of MITEs of *Arabidopsis thaliana*.pdf

INTRODUCCIÓN.

“Los descubrimientos científicos no son una transferencia unidireccional de información de la inequívoca naturaleza a unas mentes siempre abiertas. Es una interacción recíproca entre una variopinta y confusa naturaleza y unas mentes lo suficientemente receptivas (como no lo son muchas) para extraer un esquema razonable de entre todo ese ruido. (...) La ciencia es una interacción equilibrada entre la mente y la naturaleza”.

Stephen Jay Gould. *Dientes de gallina y dedos de caballo.*

INTRODUCCIÓN.

1. ¿QUÉ SON LOS GENOMAS?

Etimológicamente, el término genoma hace referencia al conjunto de los genes de un individuo o de una especie, contenido en un juego haploide de cromosomas (RAE). Sin embargo, actualmente entendemos el genoma como un concepto más amplio, como todo el DNA contenido en un juego haploide de cromosomas, aunque no esté formando parte de un gen. El genoma se encarga de almacenar y transmitir la información genética.

Hace diez años se publicó la primera secuencia completa de un genoma. Se trataba del genoma de la bacteria *Haemophilus influenzae* Rd (Fleischmann RD *et al.*, 1995) consistente en una única molécula de DNA circular de 1,83 Mb de tamaño. Desde entonces se ha llevado a cabo un gran número de proyectos de secuenciación de genomas a gran escala, abarcando todos los reinos y principales órdenes de los seres vivos (ver tabla 1) y de virus.

Organismo	Completado	En proceso de ensamblaje	En proceso de Secuenciación
Virus	1495		
Arquea	24		25
Bacterias	257		521
Eucariotas	20	82	171
Protistas	4	9	35
Hongos	9	41	16
Plantas	3	2	28
Animales	4	29	79
Mamíferos	2	15	24
Peces		2	2
Aves		1	
Anfibios			1
Insectos	1	8	35
Nemátodos	1	2	5
Platelmintos			2
Otros (moluscos, cnidarios, poríferos,...)		2	10

Tabla 1. Los diferentes proyectos de secuenciación de genomas abarcan todos los grupos taxonómicos de los seres vivos y de virus. Según NCBI (Noviembre 2005).

Introducción.

La finalización de la secuenciación de genomas tan lejanos filogenéticamente como el de la bacteria *Escherichia coli*, del nematodo *Caenorhabditis elegans*, de la planta *Arabidopsis thaliana*, de la mosca *Drosophila melanogaster* y del ser humano supone una auténtica revolución en los estudios de la genética evolutiva. El análisis y la comparación de la enorme cantidad de información arrojada por estos proyectos están aportando datos muy interesantes acerca de la composición y evolución de los genomas. Gracias a este tipo de estudios se están conociendo qué tipo de regiones y secuencias contenidas en el genoma se han conservado a lo largo de la historia evolutiva y, por tanto, son esenciales para la organización de un ser vivo y cuáles se han ido adquiriendo a lo largo de la evolución y definen las características mínimas de cada grupo de organismos. Como ejemplo, en la figura 1 se representa el resultado de la comparación de las proteínas humanas predichas a partir de la secuencia del genoma y que refleja la especialización a nivel proteico que ha tenido lugar hasta alcanzar el nivel de complejidad del ser humano.

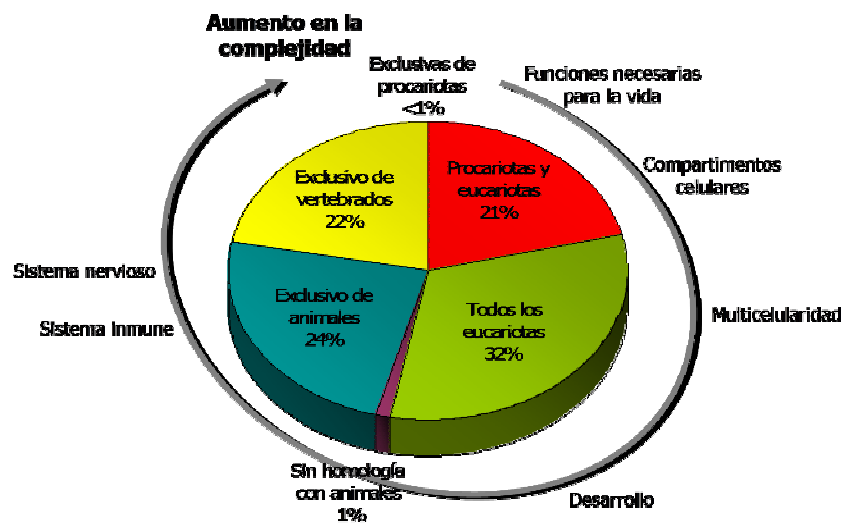


Figura 1. Los genes que codifican nuevas funciones se han ido añadiendo a lo largo de la evolución. Distribución de los homólogos de las proteínas humanas predichas. Modificado de Lander ES *et al.*, 2001.

Dos de los datos más sorprendentes derivados del estudio de los genomas hacen referencia al tamaño de los genomas y al número de genes. Contrariamente a lo que podría esperarse, ambos parámetros no están en función directa el uno respecto al otro, ni siguen una relación directa con la complejidad de los organismos. Aquellos genomas que contienen un mayor número de genes no siempre son los de mayor tamaño y algunos genomas de especies semejantes y que contienen un número similar de genes presentan grandes diferencias de

Introducción.

tamaño. Los ejemplos más sorprendentes pueden ser que el tamaño del genoma de la ameba *Amoeba dubia* (~650 Gb), ser unicelular extraordinariamente simple, sea 200 veces mayor que el del ser humano (~3 Gb), un organismo multicelular de elevada complejidad (Lander ES *et al.*, 2001; Petrov DA, 2001), o que el genoma de la avena (~11,3 Gb) sea 25 veces mayor que el del arroz (~0,46 Gb), dos especies de plantas monocotiledóneas (Gale MD and Devos KM, 1998).

Aunque existe una tendencia general a que el número de genes aumente a medida que se sube en la escala evolutiva, esta regla no es universal. A lo largo de la evolución, y mayoritariamente en organismos superiores, se han producido grandes saltos en la complejidad de los mismos que no aparecen reflejados en una diferencia en el número de genes. Un ejemplo significativo es el caso del ratón y del ser humano, ambos animales poseen un número similar de genes (Waterston RH *et al.*, 2002). Pero existen muchos otros ejemplos como que el número estimado de genes de *C. elegans* sea mayor que el de *D. melanogaster* (tabla 2).

Especie	Tamaño genoma (Mb)	Número genes (aprox.)
<i>Mycoplasma genitalum</i>	0,58	500
<i>Haemophilus influenzae</i>	1,83	1700
<i>Bacillus subtilis</i>	4,2	4100
<i>Escherichia coli</i>	4,6	4300
<i>Schizosaccharomyces pombe</i>	13,8	4900
<i>Saccharomyces cerevisiae</i>	13,5	5600
<i>Caenorhabditis elegans</i>	97	19000
<i>Arabidopsis thaliana</i>	125	25500
<i>Oryza sativa</i>	465	30000
<i>Drosophila melanogaster</i>	180	13600
<i>Mus musculus</i>	2500	22500
<i>Homo sapiens</i>	3000	24500

Tabla 2. Tamaño del genoma y número de genes de algunos organismos.

Todos estos datos están haciendo cambiar la visión clásica de que las secuencias génicas codificantes son las únicas responsables de la evolución de los organismos. La comparación entre la secuencia de genomas cercanos como el del ratón y la rata o el del ser humano y el del chimpancé están aportando todavía más datos acerca de las fracciones del genoma y los mecanismos responsables del aumento de la complejidad observada a nivel fisiológico, pero no a nivel del número de genes. Además, el primer análisis del borrador de la secuencia del genoma del chimpancé (*Pan troglodites*) muestra que existe solamente un 1,23% de divergencia nucleotídica entre este primate y el ser humano, que difícilmente explica las diferencias

Introducción.

fisiológicas existentes entre ambos organismos (reflejadas principalmente a nivel conductual) (Chimpanzee Sequencing and Analysis Consortium, 2005). Estos datos apoyan aún más la hipótesis de que además de las secuencias codificantes, en los genomas han evolucionado toda una serie de mecanismos de regulación de la expresión génica, que han permitido el desarrollo de organismos más *sofisticados*.

Las secuencias del DNA que codifican proteínas, es decir, las que llegan a ser traducidas, comprenden, generalmente, tan sólo una pequeña parte de los genomas. Por ejemplo, se ha estimado que este tipo de secuencias constituyen tan sólo el 1,5% de todo el genoma humano (Lander ES *et al.*, 2001). Cuando se habla de genes, la fracción que éstos ocupan en los genomas es algo superior, pues la mayoría de los genes son entidades mayores que, además de la secuencia codificante (exones), están constituidos por intrones y regiones reguladoras, que controlan y dirigen su correcta expresión. La mayor parte de los genes que codifican proteínas son genes de copia única, aunque existen algunas familias génicas representadas por varias copias en un genoma, como los genes que codifican histonas o las globinas humanas. Siguiendo con el ejemplo del genoma humano, la suma de todas las secuencias génicas, incluyendo los intrones y aquellas regiones del DNA que codifican RNAs estructurales (tRNAs y rRNAs), sólo llegan a explicar un 25% del mismo (Lander ES *et al.*, 2001). Este porcentaje varía según el organismo que tengamos en consideración, estando en función del tamaño del genoma. *Schizosaccharomyces pombe*, la levadura de fisión, es el organismo eucariota con menor número de genes descrito hasta el momento. Las secuencias que codifican proteínas ocupan cerca del 60,2% de las 13,8 Mb que comprenden su genoma. Y el contenido de DNA de secuencia única de este genoma supera el 90% de su totalidad (Wood V *et al.*, 2002). *Arabidopsis* es una planta que también tiene un genoma de pequeño tamaño, en comparación a otras plantas y organismos de complejidad semejante. En su caso, las regiones génicas comprenden alrededor del 50% de las 125 Mb de su genoma (AGI, 2000).

Normalmente se habla de DNA extragénico para referirse a aquellas fracciones del genoma que no están constituidas o forman parte de genes. En 1968, Britten y Kohne describieron la existencia de grandes fracciones de DNA genómico que se organizaba de forma diferente a los genes de copia única, en eucariotas superiores (Britten RJ and Kohne DE, 1968). Se trata de secuencias que aparecen repetidas en el genoma, en un número variable de copias o repeticiones, a las que se denominó DNA repetitivo. Bajo el término de DNA repetitivo se agrupan diferentes tipos de secuencias, que aparecen detalladas en la tabla 3. El contenido de

DNA repetitivo en los genomas es muy variable. Mientras que en organismos procariontes es apenas inexistente, este tipo de DNA puede llegar a suponer más del 90% del genoma en algunas especies vegetales (Leeton PR and Smyth DR, 1993) o de anfibios (Marracci S *et al.* 1996). De hecho, es el contenido en DNA repetitivo y no el número de genes, el que explica la mayor parte de la variación de tamaño entre los genomas eucariotas. En concreto y, como se explica más adelante en esta introducción, los elementos transponibles son los principales componentes que determinan el tamaño de los genomas (Kidwell MG, 2002).

Tabla 3. Tipos de DNA repetitivo.

-
1. **DNA altamente repetido.**
 - 1.1. **DNA satélite:** repeticiones de dominios de longitud variable (5-100 pb) en miles o millones de copias, ocupando fragmentos de hasta 100 Mb de longitud (ej. DNA satélite α centromérico humano).
 2. **DNA medianamente repetido.**
 - 2.1. **DNA microsátélite:** repeticiones en tándem de grupos de hasta 4 pb en miles de copias, situadas principalmente en regiones centroméricas.
 - 2.2. **DNA minisátélite:** repeticiones en tándem de fragmentos de unos 25 pb de media en miles de copias, situadas principalmente en regiones teloméricas.
 - 2.3. **DNA repetitivo disperso o intercalado:** principalmente constituido por elementos transponibles o secuencias derivadas, pero también comprende las familias génicas presentes en alto número de copias en el genoma (como rDNA o histonas).
-

Excluyendo el DNA repetitivo génico (rDNA y familias génicas presentes en alto número de copias en el genoma), algunos autores hablan de *DNA basura* para referirse al DNA repetitivo, argumentando que no aporta ningún beneficio ni funcionalidad al genoma y que, a pesar de su abundancia, puede ser dispensable para el genoma. Pero la teoría del *DNA basura* plantea otras cuestiones. ¿Por qué se ha mantenido este tipo de DNA en los genomas? Y, ¿por qué lo ha hecho de forma diferente en los distintos genomas? A medida que se ha avanzado en los estudios de los genomas, estas respuestas han comenzado a poder ser aclaradas. Al mismo tiempo la teoría del *DNA basura* ha ido perdiendo cada vez más fuerza.

Actualmente se conoce que parte de este DNA repetitivo sí tiene una funcionalidad en los genomas. La hipótesis del "DNA esquelético" (Cavalier-Smith T and Beaton MJ, 1999) propone que una de las funciones principales de este tipo de secuencias es definir y participar en la estructura de los cromosomas, asegurando la correcta transmisión del material genético a las células hijas. El DNA repetitivo está formando parte de los orígenes de replicación, centrómeros y telómeros en la mayoría de los genomas, pero además constituyen señales que dirigen la compactación del genoma con proteínas y RNA para formar determinadas estructuras de la

Introducción.

cromatina y señales para el anclaje del DNA a estructuras celulares o nucleares concretas (Shapiro JA, 2005). Los centrómeros están compuestos principalmente por dos tipos de DNA repetitivo, los elementos transponibles (de los que hablaré más adelante) y el DNA satélite (como el DNA satélite α de humanos o la repeticiones CentC de maíz), que se dispone formando de 100 a 5000 repeticiones en tándem, algunas de las cuales constituyen sitios de unión para la proteína de unión específica a centrómero, CENP-B (Wong LH and Choo KHA, 2004; Dawe RK, 2003). La presencia de este alto número de repeticiones hace que haya menos orígenes de replicación en esta región, asegurando la replicación tardía de los centrómeros. Los telómeros están formados por DNA medianamente repetitivo, esencialmente por secuencias minisatélites y transposones (ver más adelante). Además, las regiones pericentroméricas y subteloméricas son muy ricas en DNA altamente repetitivo, ayudando a definir y delimitar las estructuras anteriores.

Los elementos transponibles son un tipo muy especial y el más abundante de DNA repetitivo. Debido a su gran diversidad, ubicuidad y capacidad de movilización, los ETs han sido elementos que han participado activamente en la evolución de los genomas, moldeándolos y aportándoles gran plasticidad. En este trabajo hemos querido estudiar el papel que han ejercido este tipo de secuencias sobre la evolución de los genomas, por lo que más adelante hablaré más detalladamente acerca de estos elementos.

El importante papel, a veces esencial, que juegan algunos tipos de DNA repetitivo sobre los genomas explica por qué se mantienen estas secuencias a lo largo de la evolución. Es indiscutible que la selección natural ha sido el factor último que ha determinado la persistencia de este DNA extragénico (de igual forma que lo hace sobre las secuencias génicas) en función de su posible aprovechamiento para llevar a cabo funciones celulares o, simplemente, tolerando la acumulación de aquel DNA que no supone una desventaja evolutiva para el genoma.

Más difícil de explicar es por qué esta acumulación de DNA extragénico y expansión del genoma ha sido mayor en unos organismos que en otros. Se han descrito caracteres fenotípicos correlacionados con el tamaño del genoma, entre los que se encuentran la duración de la mitosis y meiosis, la tasa de metabolismo basal de algunos mamíferos y aves, el tiempo mínimo de generación de algunas plantas, el tamaño de la semilla o el tiempo de desarrollo embrionario de algunas salamandras (Petrov DA, 2001). Además, parece que para que la célula funcione correctamente es necesaria una relación cito-nuclear relativamente constante, es decir, el volumen celular y el volumen nuclear (tamaño del genoma) deben estar directamente relacionados (Bowen NJ and Jordan IK, 2002). Existen distintos mecanismos responsables de la

Introducción.

variación del tamaño de los genomas (figura 2). Aunque la poliploidía y las grandes duplicaciones cromosómicas explican aquellos casos más extremos, parece que la movilización de los elementos transponibles y los sucesos de pequeñas inserciones y deleciones (indels) son las mayores fuentes de diversidad en tamaño de los genomas (Petrov DA, 2001). Por qué la selección natural es menos restrictiva en unas especies que en otras o en determinados momentos (donde la deriva genética jugaría un papel fundamental) no está totalmente resuelto. En situaciones desfavorables o de adaptación, la poliploidía o las grandes duplicaciones cromosómicas pueden aportar copias de genes existentes con las que poder experimentar, manteniendo también la función original. Como se explicará más adelante la transposición es una importante fuente de variabilidad que puede suponer una ventaja adaptativa en ciertas condiciones. Todos estos procesos han de abordarse a nivel poblacional para poder entender la ventaja que puede suponer la expansión de los genomas.

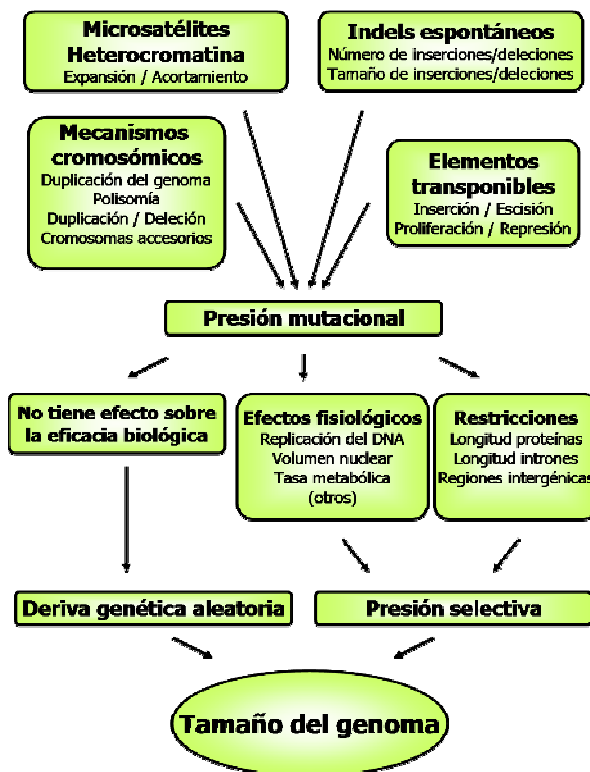


Figura 2. Procesos involucrados en la variación del tamaño del genoma.
Adaptado de Petrov DA, 2001.

Con todo lo explicado en este apartado se puede afirmar que la idea y visión que se tenía de los genomas hasta hace, prácticamente, una década está cambiando. Los genomas son entidades heterogéneas, diversas y dinámicas, cuya evolución ha llevado a un aumento de la complejidad de los organismos. Pero esta complejidad se ha adquirido gracias a la reorganización de los

genomas a diferentes niveles, desde la creación de nuevos genes o dominios, la reordenación de los mismos o la amplificación de familias génicas, hasta el desarrollo de nuevos mecanismos de regulación de la expresión génica: regiones promotoras, activadoras y terminadoras, splicing alternativo, creación de dominios de la cromatina y modificaciones de la misma, silenciamiento génico y miRNAs, etc. Las secuencias repetidas, y en concreto los transposones, han tenido un papel esencial en la evolución de los genomas al aportarles la plasticidad necesaria para poder alcanzar estos niveles de complejidad.

2. LOS ELEMENTOS TRANSPONIBLES.

Los elementos transponibles (ETs) son entidades genéticas capaces de movilizarse (transponer) dentro de los genomas, cambiando de localización y, a veces, aumentando su número de copias. Los efectos de la transposición se habían descrito, indirectamente, mucho antes de conocerse la fuente de los cambios fenotípicos observados, hablándose del fenómeno de variegación (citado en Lönnig WE and Saedler H, 2002). Fue Barbara McClintock quien, a mediados del siglo XX, propuso la existencia de unos “elementos controladores” causantes de la inestabilidad cromosómica y la variegación que observaba en el maíz. Postuló que estos elementos se movilizaban entre múltiples genes, activando o reprimiendo la expresión de los mismos (McClintock B, 1948). Pero fue el genetista Alexander Brink quien acuñó el término de *elementos transponibles*, al considerarlo “menos interpretativo” (citado en Wessler SR, 2001).

Sin embargo, debieron pasar 20 años para que se caracterizara molecularmente el primer elemento transponible, una secuencia de inserción (IS) bacteriana de clase II (Jordan E *et al.*, 1967). Desde entonces, se ha caracterizado un enorme número de transposones en distintos organismos.

2.1. CLASIFICACIÓN DE LOS ELEMENTOS TRANSPONIBLES.

La amplia diversidad de ETs existente en el conjunto de los organismos hace difícil establecer un sistema universal de clasificación de estos elementos. David Finnegan, en 1989, definió una primera clasificación de los transposones en función de su mecanismo de transposición (Finnegan DJ, 1989). Pero a medida que se han ido describiendo nuevas secuencias móviles

esta clasificación ha debido ser modificada. Hull, en 2001, estableció una nomenclatura taxonómica de tipo viral para definir los distintos grupos de retrotransposones (Hull R, 2001). Curcio y Derbyshire propusieron en 2003 una nueva forma de agrupar los elementos transponibles según el modo en el que transponen, definido por sus enzimas efectores o transposasas (Curcio MJ and Derbyshire KM, 2003). Sin embargo, estas clasificaciones resultan complicadas y presentan ciertas limitaciones a la hora de agrupar a todos los ETs (Capy P, 2005). Por ello, en este trabajo se ha utilizado la clasificación propuesta por Capy en 1997, y revisada en 2005, y que es una síntesis de la nomenclatura utilizada por los diferentes autores a partir de la clasificación original de Finnegan (Capy P, 2005).

De esta forma, los elementos transponibles se dividen en dos grandes grupos de acuerdo con su modo de transposición: los elementos de clase I o retrotransposones y los elementos de clase II o transposones de DNA. En la figura 3 se representan las principales clases de ETs eucariotas.

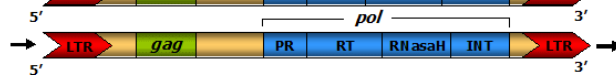
Elementos transponibles de clase I o Retrotransposones

Retrotransposones con LTRs

Tipo Ty1-*copia*



Tipo Ty3-*gypsy*



Retrotransposones sin LTRs

LINEs



SINEs



Elementos transponibles de clase II o Transposones de DNA

Elementos autónomos



Elementos no autónomos



MITEs



Helitrones

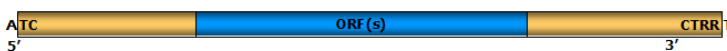


Figura 3. Los diferentes tipos de transposones eucariotas. Los tamaños son sólo indicativos, no están a escala. Las flechas indican la orientación de las repeticiones terminales. Ver abreviaturas en texto.

2.1.1. Elementos transponibles de clase I. Retrotransposones.

Los retrotransposones se movilizan dentro del genoma a través de un intermediario de RNA, que es inversamente transcrito a DNA antes de su integración en otra posición del genoma. Este mecanismo de movilización recibe el nombre de transposición de tipo replicativa, ya que el

Introducción.

número de copias de estos elementos aumenta con cada suceso de transposición. La estructura de estos ETs es muy similar a la de los retrovirus, los cuales codifican adicionalmente para proteínas de envoltura (ENV) que les proporcionan capacidad invasora, por lo que se postula un origen retroviral de los retrotransposones.

Los retrotransposones pueden ser subdivididos, a su vez, en diferentes grupos, en base a su estructura y mecanismo de transposición: los retrotransposones con LTRs y los retrotransposones sin LTRs (Kumar A and Bennetzen JL, 1999).

Los retrotransposones con LTRs presentan largas repeticiones terminales (LTRs) de longitud variable (desde 100 pb hasta varias Kb) orientadas ambas en el mismo sentido, de forma directa, flanqueando la región interna codificante. La transcripción de estos retrotransposones comienza en la LTR 5' y termina en la LTR 3'. Generalmente, las LTRs contienen las secuencias promotoras así como reguladoras de la terminación de la transcripción. Los elementos autónomos contienen, al menos, dos genes, denominados *gag* y *pol*, que se transcriben y traducen en forma de poliproteína, que es fragmentada en distintos péptidos funcionales por la acción de una proteasa codificada por el mismo elemento. El gen *gag* codifica una proteína estructural de tipo cápsida (CP), importante para el empaquetamiento del RNA del retrotransposón. El gen *pol* codifica los enzimas necesarios para completar el ciclo vital del retrotransposón, que aportan actividad proteasa (PR), transcriptasa inversa (RT), RNasa H e integrasa (INT). El orden de estos dominios proteicos dentro de la secuencia de DNA define dos grupos de LTR retrotransposones, tipo Ty1-*copia* y tipo Ty3-*gypsy* (figura 3).

Los retrotransposones sin LTRs no están flanqueados por LTRs, en cambio presentan en su extremo 3' una secuencia repetida simple, que generalmente es una cola poli(A). Se transcriben a partir de un promotor interno. En función a su capacidad codificante estos elementos se dividen en LINEs y SINEs. Los LINEs codifican para tres proteínas con funciones similares a las que aparecen en los retrotransposones con LTRs: una proteína del tipo *gag*, una endonucleasa (EN) y una transcriptasa inversa (RT). Los SINEs, o retroposones, no tienen capacidad codificante ni similitud con ninguna región codificante, siendo, por tanto, elementos no autónomos (no son capaces de transponer por sí solos). Los SINEs derivan de transcritos de la RNA polimerasa III (como los tRNAs o los RNAs 7SL) y el extremo 3' de algunos de ellos comparte homología con la región 3' de algunos LINEs presentes en el mismo genoma, indicando que los SINEs podrían utilizar la maquinaria de transposición de los LINEs (Kajikawa M and Okada N, 2002).

2.1.2. Elementos transponibles de clase II. Transposones de DNA.

La mayoría de los transposones de DNA siguen un mecanismo de transposición no replicativo, de escisión e inserción. Estos ETs codifican para una proteína con función integrasa, denominada transposasa, encargada de su movilización a través de un producto intermediario, que es el propio transposón de DNA. La transposasa reconoce sus repeticiones terminales invertidas (TIRs) y corta la secuencia del elemento móvil por ambos extremos para integrarlo, posteriormente, en otra localización del genoma. En función de la homología existente tanto entre su secuencia interna y codificante, como la aparecida entre los TIRs o entre las TSDs (duplicaciones del sitio de inserción), se han definido siete superfamilias de transposones de DNA eucariotas (Robertson HM, 2002; Feschotte C *et al.*, 2002b). Así mismo, por semejanza estructural de los elementos, se han podido establecer relaciones entre transposones de DNA eucariotas y procariotas. En concreto se ha visto que los elementos de la superfamilia *Tc/mariner* eucariotas son los que más similitud presentan con las IS bacterianas (Chandler M and Mahillon J, 2002).

Como en el caso de los retrotransposones, los diferentes grupos de elementos móviles de clase II contienen miembros autónomos y miembros no autónomos, que comparten las secuencias en *cis* necesarias para su movilización. La transposición de los elementos no autónomos se produce en *trans*, gracias a la maquinaria aportada por los elementos autónomos.

Un tipo especial de elementos transponibles de DNA no autónomos son los MITEs (Miniature Inverted-repeat Transposable Elements). Los MITEs son transposones de pequeño tamaño (que varía entre 100 pb y 700 pb) con TIRs de 10 a 15 nt de longitud, que flanquean una secuencia interna no codificante rica en AT. La presencia de los TIRs y de repeticiones subterminales (SIRs) confieren a estos elementos la capacidad de formar estructuras secundarias (Bureau TE and Wessler SR, 1992). Una característica especial de estos ETs es que parecen tener un mecanismo replicativo de transposición, debido al elevado número de copias de ciertas familias, existentes en diferentes genomas (Feschotte C *et al.*, 2002b). Pero la identificación en el genoma de *Arabidopsis* de un elemento potencialmente autónomo, de la familia *Tc1/mariner*, relacionado con la familia *Emigrant* de MITEs, sugiere que estos elementos podrían ser movilizados por ETs de clase II activos presentes en el mismo genoma (Feschotte C and Mouches C, 2000). Sin embargo hasta la publicación en 2003 de tres trabajos, en un mismo número de la revista *Nature*, que describían la movilización de una familia de MITEs de arroz, denominada *mPing*, no

se había observado transposición de estos pequeños transposones (Kikuchi K *et al.*, 2003; Nakazaki T *et al.*, 2003; Jiang N *et al.*, 2003). En estos trabajos se postula que la movilización de estos MITEs podría deberse a la actividad de un elemento autónomo emparentado con los mismos, denominado *Pong*, que codifica para una transposasa de la familia *PIF/Harbinger*.

Recientemente se ha identificado un nuevo grupo de transposones de clase II en genomas eucariotas, que presentan características especiales y similares a las de los transposones IS (secuencias de inserción) bacterianos. Estos elementos móviles, clasificados bajo el nombre de *Helitron* en plantas, no presentan TIRs ni TSDs y transponen mediante replicación en círculo rodante (Kapitonov VV and Jurka J, 2001; Feschotte C and Wessler SR, 2002).

2.2. ABUNDANCIA DE LOS TRANSPOSONES EN LOS GENOMAS.

Los transposones son elementos ubicuos, se han encontrado secuencias de este tipo en todos los organismos analizados, con la excepción de la cepa secuenciada de la bacteria *Bacillus subtilis* (Kunst F *et al.*, 1997), y de los protozoos parasitarios *Leishmania* y *Plasmodium falciparum* (Gardner MJ *et al.*, 2002). En concreto, los transposones de DNA aparecen a lo largo de toda la escala evolutiva de los seres vivos (Hua-Van A *et al.*, 2005), sugiriendo un origen ancestral de este tipo de elementos. Los ETs han formado parte de los genomas de los distintos organismos durante largo tiempo y son el tipo de DNA más abundante en muchos de ellos, llegando a constituir cerca del 99% de algunos genomas del género *Lilium* (Leeton PR and Smyth DR, 1993). En la tabla 4 se muestran algunos ejemplos de la fracción de los genomas ocupada por los elementos transponibles.

Especie	Fracción de ETs En el genoma	Referencias
<i>Saccharomyces cerevisiae</i>	3,1%	Kim JM <i>et al.</i> , 1998
<i>Caenorhabditis elegans</i>	6,5%	<i>C.elegans</i> Sequencing Consortium. 1998
<i>Drosophila melanogaster</i>	3,1%	Kaminker JS <i>et al.</i> , 2002
<i>Anopheles gambiae</i>	16%	Holt RA <i>et al.</i> , 2002
<i>Mus musculus</i>	37,9%	Waterston RH <i>et al.</i> , 2002
<i>Homo sapiens</i>	45%	Lander ES <i>et al.</i> , 2001
<i>Arabidopsis thaliana</i>	14%	Arabidopsis Genome Initiative, 2000
<i>Zea mays</i>	hasta 80%	Meyers BC <i>et al.</i> , 2001
<i>Lilium sp.</i>	99%	Leeton PR and Smyth DR, 1993

Tabla 4. Algunos ejemplos de la fracción de los genomas ocupada por los ETs.

Sin embargo, la abundancia de los diferentes tipos de transposones en los distintos organismos es muy variable. Los organismos procariotas sólo contienen transposones de DNA, mientras que los eucariotas contienen ambos tipos de ETs, si bien los más abundantes en organismos superiores son los elementos transponibles de clase I o retrotransposones (Hua-Van A *et al.*, 2005). Las levaduras y la mayoría de los protozoos parasitarios secuenciados no poseen elementos transponibles de clase II. Los transposones más abundantes en plantas son los retrotransposones con LTRs y los MITEs (Feschotte C *et al.*, 2002a). Estos últimos pueden suponer más del 70% de todos los transposones de arroz o el 50% de los de *Arabidopsis* (Le QH *et al.*, 2000). Los MITEs también son los elementos transponibles más abundantes en *Caenorhabditis elegans* (Hua-Van A *et al.*, 2005). Los retrotransposones son los ETs predominantes en insectos, como *Anopheles gambiae* y *Drosophila* (Holt RA *et al.*, 2002; Kaminker JS *et al.*, 2002), y en mamíferos, como el ratón o el ser humano (Waterston RH *et al.*, 2002; Lander ES *et al.*, 2001). Los retrotransposones sin LTRs constituyen la mayor fracción de los elementos transponibles del genoma humano. En concreto, la familia L1 de LINEs ocupa el 21% del total de la secuencia genómica humana y la familia Alu de SINEs el 10% (Lander ES *et al.*, 2001).

2.3. DISTRIBUCIÓN DE LOS ELEMENTOS TRANSPONIBLES.

La distribución de los elementos transponibles dentro de los genomas es el resultado de la preferencia o especificidad de inserción de estas secuencias móviles y de la eficiencia de eliminación de las mismas (aclarado) por parte del genoma. Por otra parte, la especificidad de inserción y el efecto de la evolución son distintos para cada tipo de elemento y para cada genoma. Existen pocos estudios en los que se haya podido diferenciar claramente el papel jugado por estas dos fuerzas en la distribución de un determinado transposón en su genoma huésped. La mayoría de los trabajos en los que se analiza la localización de los ETs en los genomas presentan la disposición de estos elementos en un momento concreto, en el que se realiza el análisis, por lo que aportan una idea global de la especificidad de inserción de las secuencias móviles y del papel de la evolución en dicha distribución.

Sin embargo, el análisis de familias jóvenes de ETs en los genomas o de la movilización de algunos de estos elementos ha permitido conocer la especificidad de inserción de determinados

Introducción.

transposones. Éstos pueden mostrar preferencia por ciertas secuencias de DNA, por elementos funcionales del DNA o por la estructura de la cromatina.

Los elementos Ty de levaduras constituyen un buen ejemplo de transposones que presentan una clara especificidad de inserción. Prácticamente ninguno de estos ETs parece dirigirse hacia una secuencia de DNA concreta, pero sí lo hacen hacia determinados elementos funcionales del DNA. Los transposones Ty3 se insertan específicamente en una posición que precede por unos pocos nucleótidos al inicio de transcripción de genes transcritos por la RNA polimerasa III, por la interacción de su integrasa con los factores de transcripción TFIIIB y TFIIIC (Chalker DL and Sandmeyer SB, 1992; Yieh L *et al.*, 2000; Yieh L *et al.*, 2002). Esto es importante porque los genes transcritos por la Pol III tienen un promotor interno e insertarse en 5' de estos genes equivale a insertarse en zonas sin genes. Los elementos Ty1 se insertan a unos 700 pb a 5' de genes transcritos también por la Pol III (Devine SE and Boeke JD, 1996; Bachman N *et al.*, 2005). En cambio, el retrotransposón Ty5 se inserta cerca de las porciones teloméricas inactivas y en el locus silenciado de determinación sexual (mating type) del genoma de *Saccharomyces cerevisiae*. Esta preferencia de inserción se debe a que su integrasa interacciona específicamente con las proteínas Sir4, que conforman la cromatina inactiva (Zou S and Voytas DF, 1997; Zhu Y *et al.*, 2003).

Los elementos transponibles de plantas también presentan cierta especificidad de inserción, aunque ésta es mucho menos estricta que la observada en levaduras. Por ejemplo, los transposones de clase II parecen insertarse (y retenerse) en regiones genéticamente activas, no metiladas (presuntamente eucromáticas) (Bennetzen JL, 2000). Los transposones de DNA de maíz, como los elementos *Ac/Ds*, *Mu1* o *Spm*, transponen preferentemente en regiones transcripcionalmente activas (Lönning WE and Saedler H, 2002; Bennetzen JL, 2000). Los elementos que presentan pocas copias en un genoma pueden insertarse en regiones transcripcionalmente activas para facilitar su transcripción y transposición (Bennetzen JL, 2000). Estudios realizados sobre la movilización del retrotransposón Tnt1 de tabaco o Tos17 de arroz, indican que los retrotransposones de plantas también parecen insertarse en regiones transcripcionalmente activas (Grandbastien MA *et al.*, 2005; Miyao A *et al.*, 2003).

No obstante, estos elementos terminan acumulándose en regiones heterocromáticas. En general, parece existir una clara asociación de los retrotransposones con la cromatina inactiva y metilada (presuntamente DNA heterocromático), como la que ocupa los centrómeros, las regiones pericentroméricas, los telómeros o los "nudos" (knobs) de heterocromatina situados en los brazos de los cromosomas de muchas plantas (Bennetzen JL, 2000). Éste es el caso de los

genomas de arroz o *Arabidopsis*, donde los retrotransposones ocupan, principalmente, las regiones centroméricas (*Arabidopsis* Genome Initiative, 2000; Goff SA *et al.*, 2002).

En muchos casos los transposones tienden a insertarse junto a o dentro de la secuencia de otros ETs, en regiones ricas en este tipo de elementos (SanMiguel P *et al.*, 1996; Hua-Van A *et al.*, 2005). Se piensa que los retroelementos transponen en regiones heterocromáticas o pobres en genes para minimizar los efectos derivados de su transposición de tipo replicativa.

Esta dinámica de distribución de los retrotransposones también se ha visto en animales. En concreto, se han llevado a cabo estudios del efecto de la evolución sobre la distribución de estos elementos en el genoma humano. Los resultados obtenidos de algunos de estos análisis sugieren que los retrotransposones con LTRs se insertan en zonas transcripcionalmente activas, pero que con el tiempo las copias localizadas en estas regiones se van perdiendo y aumenta la densidad de retroelementos en áreas del genoma ricas en AT o pobres en genes (Medstrand P *et al.*, 2002; Medstrand P *et al.*, 2005).

El gran tamaño de los retrotransposones, junto con la presencia en sus LTRs de fuertes elementos promotores, pueden alterar gravemente la función de los genes junto a o en los que se insertan, explicando, así, su aclaramiento de las regiones eucromáticas.

Sin embargo, no se han realizado estudios de este tipo sobre transposones de menor tamaño, cuyo efecto sobre los genes podría no ser tan drástico. En este trabajo hemos querido analizar el efecto de la evolución sobre la distribución de la familia *Emigrant* de MITEs de *Arabidopsis*.

3. EFECTO DE LOS ELEMENTOS TRANSPONIBLES SOBRE LA EVOLUCIÓN DE GENES Y GENOMAS.

La autonomía de movilización que poseen los transposones, así como su origen ancestral y larga convivencia en los distintos genomas, han hecho que estos elementos hayan tenido un papel muy importante sobre la evolución de genes y genomas. Papel que ha sido muy discutido desde el momento de su descubrimiento y del que todavía hoy en día queda mucho por conocer.

Barbara McClintock defendía la idea de que los transposones juegan un rol activo en la evolución de los genomas. Hablaba de los transposones como *elementos controladores*, capaces de

modificar los patrones de expresión de los genes en el desarrollo, al insertarse cerca de cualquier locus, así como de remodelar los genomas como respuesta a estrés, favoreciendo la aparición de nuevas variantes con una eficacia biológica mayor, las cuales podrían adaptarse a o superar estas condiciones desfavorables con mayor facilidad (McClintock B, 1987; McClintock B, 1984). Frente a esta visión funcional de los ETs, otros autores desarrollaron la hipótesis del *DNA egoísta*, en la que se considera a los elementos transponibles como secuencias parásitas que se han ido acumulando a lo largo de la evolución, gracias a su habilidad de replicarse de forma autónoma, sin que aporten ningún beneficio al genoma en el que hospedan (Doolittle WF and Sapienza C, 1980; Orgel LE and Crick FH, 1980). Sin embargo, en los últimos años se han acumulado evidencias de que aunque no todos los ETs ejercen funciones beneficiosas para los genomas en los que se hospedan, ya que la transposición es un fenómeno fundamentalmente mutágeno, muchos de ellos son esenciales para el correcto funcionamiento celular. Por otra parte, su papel en la evolución genes y genomas está fuera de toda duda. A continuación se detallan algunos ejemplos de funciones celulares desarrolladas por transposones, así como sus principales efectos mutágenos.

3.1. FUNCIONES CELULARES DESARROLLADAS POR TRANSPONES.

La inserción de los transposones fuera de secuencias génicas ha aportado funciones de carácter estructural a los genomas, contribuyendo de esta forma a su evolución. Los elementos transponibles pueden estar participando en la organización de la cromatina en niveles superiores al estar formando parte de los centrómeros, telómeros y de las MARs (regiones de anclaje a la matriz nuclear).

Los centrómeros de organismos como las plantas, *Drosophila* o humanos están constituidos principalmente por grandes fragmentos de DNA microsatélite interrumpidos por elementos transponibles. Existen transposones específicos de centrómero, descritos en cereales, cuya evolución, así como la de otras secuencias también derivadas de elementos transponibles, ha tenido un papel esencial en la definición de los centrómeros (Wong LH and Choo KHA, 2004) (figura 4). Por ejemplo, se ha visto que retroelementos específicos de centrómero de maíz (CRM) interaccionan con la proteína CENH3 (histona H3 centromérica), cuya variante humana se denomina CENP-A, encargada de iniciar la formación del cinetocoro (Zhong CX *et al.*, 2002; Wong LH and Choo KHA, 2004; Collins KA *et al.*, 2004).

Introducción.

En *Drosophila melanogaster*, los telómeros consisten en repeticiones en tándem de dos retrotransposones sin LTRs, HeT-A y TART, cuya actividad transposicional (dirigida específicamente a las regiones teloméricas) compensa la ausencia del enzima telomerasa en este género (Pardue ML and Debaryshe PG, 2003; Abad JP *et al.*, 2004).

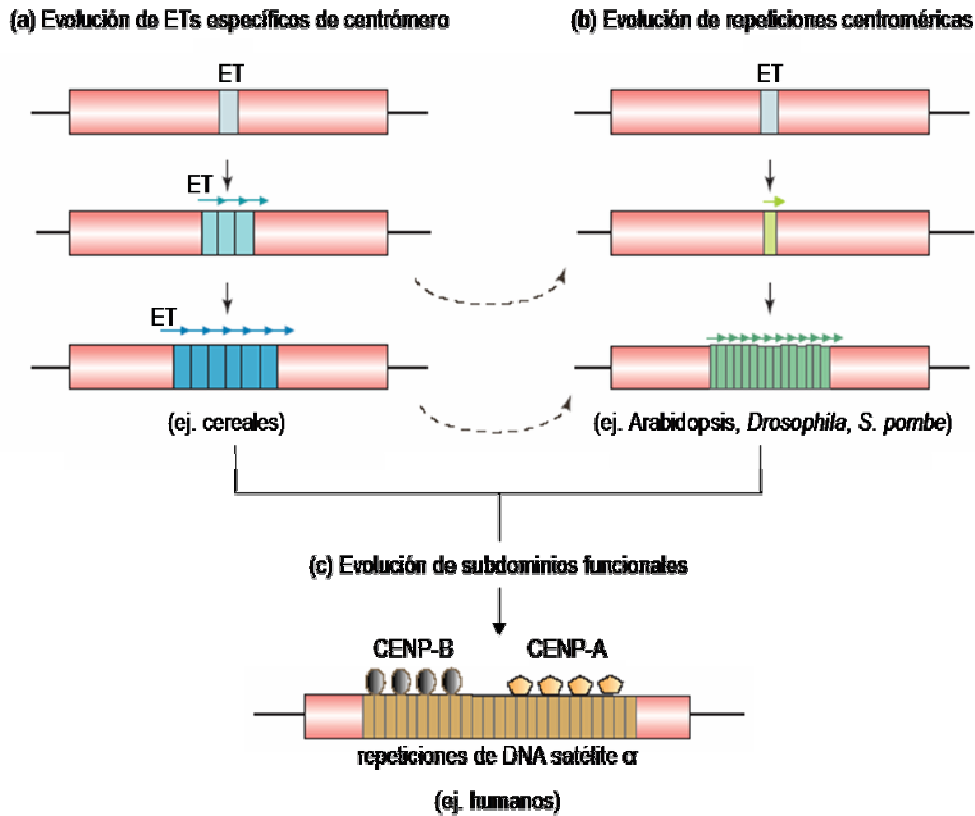


Figura 4. Los elementos transponibles (ETs) y las transposasas intervienen en la evolución de los centrómeros a través de tres posibles vías convergentes. Adaptado de Wong LH and Choo KHA, 2004.

En la literatura aparecen varios ejemplos de la relación entre MARs y transposones tanto de clase I, retrotransposones *Ty* de levaduras (Wyrick JJ *et al.*, 2001), *gypsy* de *Drosophila* (Byrd K and Corces VG, 2003; Nabirochkin S *et al.*, 1998), secuencias Alu y LINEs humanos (Jordan IK *et al.*, 2003), como de clase II, MITEs de maíz y sorgo capaces de unirse a proteínas de la matriz nuclear (Avramova Z *et al.*, 1998; Tikhonov AP *et al.*, 2000). Estos datos apuntan a que estos elementos pueden actuar como aisladores génicos, interviniendo en procesos de regulación génica, al crear dominios independientes de la cromatina que pueden estar separando secuencias activadoras y promotoras de genes adyacentes (Nabirochkin S *et al.*, 1998; Holmes-Davis R and Comai L, 1998).

3.2. FUNCIONES ADAPTATIVAS Y GENES DERIVADOS DE ELEMENTOS TRANSPONIBLES.

La presencia y actividad de los transposones también ha sido aprovechada por los organismos en los que hospedan, desarrollando funciones adaptativas, derivadas de antiguas secuencias móviles. Se ha descrito que hay, al menos, 47 genes humanos que derivan de elementos transponibles (Lander ES *et al.*, 2001).

Uno de estos genes es la proteína centromérica B humana (CENP-B), componente del cinetocoro. Esta proteína presenta gran similitud con las transposasas de la familia Tc1/*mariner* (Kipling D and Warburton PE, 1997). Podría tratarse del producto de un antiguo transposón con especificidad por las regiones centroméricas, como los que se describen en el apartado anterior, que, simplemente, ha retenido la capacidad de unirse a estas secuencias, siendo incapaz de movilizarlas (figura 4).

El sistema inmune específico de vertebrados también constituye un ejemplo representativo de la adquisición de funciones transposicionales en un momento dado de la evolución de los animales. Los genes *RAG1* y *RAG2*, involucrados en el sistema de recombinación V(D)J de las inmunoglobulinas que tiene lugar en los linfocitos B, parecen derivar de transposones. La actividad coordinada de ambos genes puede catalizar la transposición *in vitro*, al crear conjuntamente un enzima similar a las transposasas de clase II (Agrawal A *et al.*, 1998; Hiom K *et al.*, 1998).

La telomerasa es una retrotranscriptasa que se encarga de mantener los extremos de los cromosomas lineares (telómeros), mediante la replicación de fragmentos de DNA contenidos en estas regiones. Este enzima esencial para la supervivencia celular tiene una estructura semejante a la de las retrotranscriptasas de los ETs de clase I sin LTRs, lo que hace pensar que la telomerasa eucariota constituye otro ejemplo de un gen celular derivado de un elemento transponible (Arkhipova IR *et al.*, 2003; Eickbush TH, 1997; Nakamura TM and Cech TR, 1998).

Existen otros ejemplos de neofuncionalización, o reclutamiento de ETs para ejercer funciones celulares, como el gen *sincitina 2* de primates o la familia génica *Mart* de mamíferos. El gen de la sincitina codifica para un polipéptido involucrado en la morfogénesis de la placenta, con

características de proteínas de la envoltura retroviral (Mi S *et al.*, 2000; Blaise S *et al.*, 2003). Los genes *Mart* están relacionados con el gen *gag* de retrotransposones con LTRs del tipo Sushi, presente en peces y anfibios, pero no en mamíferos. Estos genes han perdido su capacidad de retrotransposición, pero se expresan durante el desarrollo embrionario y en individuo adulto en ratón y parecen estar involucrados en el control de la proliferación celular y apoptosis (Brandt J *et al.*, 2005).

3.3. EFECTOS DE LOS TRANSPONES SOBRE LA EXPRESIÓN Y REGULACIÓN DE OTROS GENES.

La transposición es un proceso altamente mutágeno. Puede alterar la función de otros genes de forma negativa, modificando a las regiones reguladoras de éstos o el producto que codifican (ver figura 5). Los transposones afectan a la expresión de los genes al insertarse dentro o cerca de los mismos. Los elementos promotores contenidos en la secuencia de los elementos transponibles pueden promover la transcripción de los genes adyacentes, creando nuevos patrones de expresión para estos últimos. Además, la transcripción de los elementos transponibles puede continuar, en ciertos casos, más allá de la secuencia del propio transposón, incorporando secuencias de genes adyacentes.

Cuando los ETs se insertan dentro de las regiones transcritas pueden variar la eficiencia de la transcripción del gen, modificar la estructura del mRNA y alterar la estructura de las proteínas al inducir un splicing alternativo, promoviendo el inicio de la transcripción desde un intrón, una terminación prematura del transcrito o rompiendo el gen y creando varios subproductos génicos o productos no funcionales (Han JS and Boeke JD, 2005; Shapiro JA, 2005).

A nivel cromosómico la transposición puede causar roturas cromosómicas. Además la recombinación entre diferentes copias de un transposón puede promover reorganizaciones cromosómicas a gran escala, como deleciones, inversiones o duplicaciones. Se ha visto que los ETs están implicados en el origen de algunas inversiones cromosómicas naturales en distintos organismos. En *Drosophila*, por ejemplo, la inversión 2j ha sido originada por la recombinación ectópica entre dos copias de un mismo ET del tipo *Foldback* (Cáceres M *et al.*, 1999).

Introducción.

Como ejemplo de efectos deletéreos de la transposición, se estima que cerca del 0,5% de los desórdenes genéticos humanos, entre los cuales aparecen a neurofibromatosis de tipo 1, las hemofilias de tipo A y B o la hipercolesterolemia familiar, son el resultado directo de inserciones de secuencias Alu o de la recombinación no homóloga entre diferentes copias de estos SINES (Hedges DJ and Batzer A, 2005). La expresión ectópica de algunos *neogenes* derivados de ETs también es causa de enfermedades humanas. Este es el caso de la anteriormente mencionada sincitina, que parece estar involucrada en el desarrollo de la esclerosis múltiple, cuando se expresa en el cerebro (Perron H *et al.*, 2005). No obstante, no llegamos a ver los efectos más drásticos de la transposición a nivel individual, debido a que la nueva reordenación letal se pierde con el individuo.

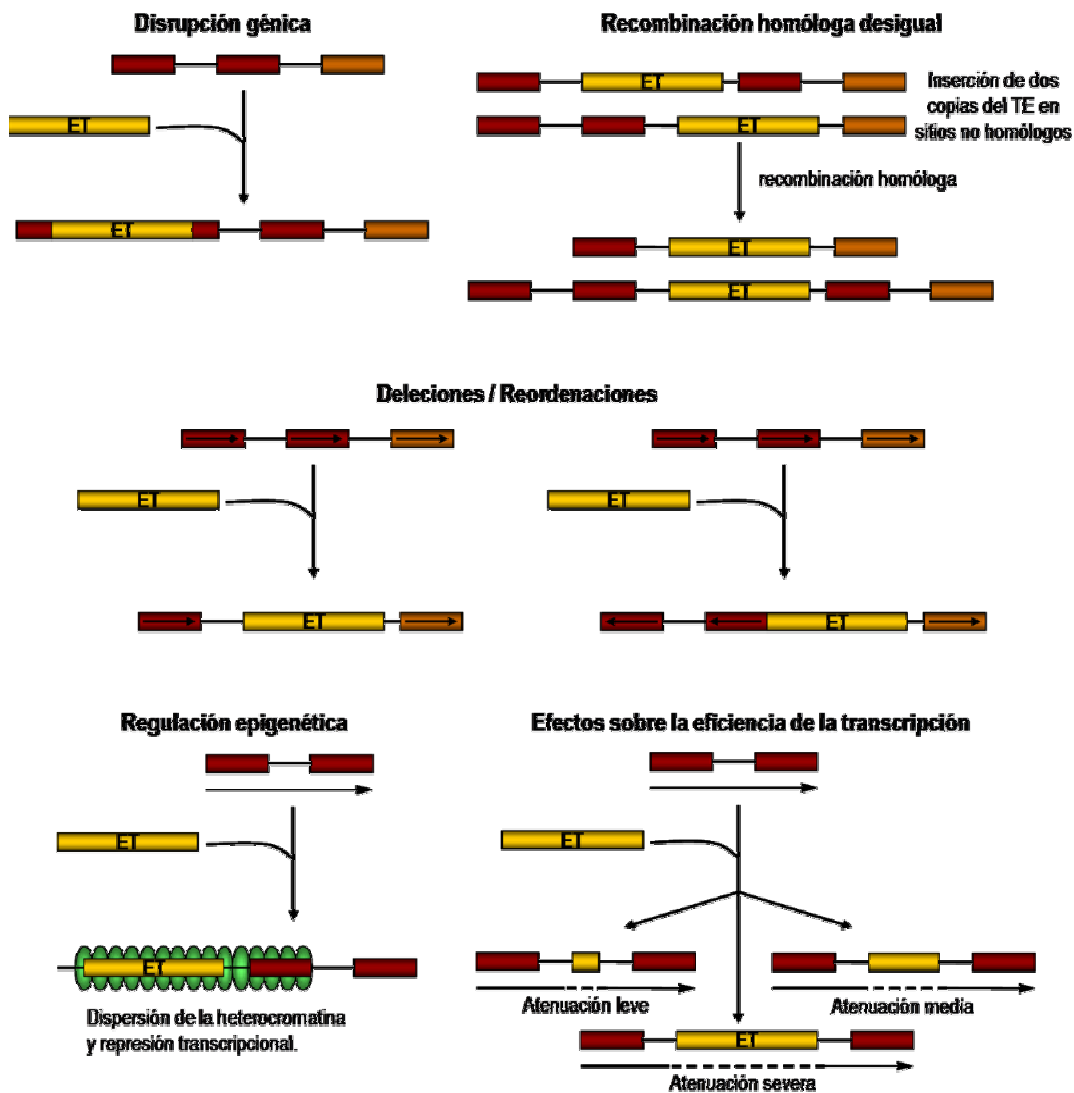


Figura 5. Posibles mecanismos por los que los ETs pueden alterar la expresión de otros genes y comprometer su función. Adaptado de Han JS and Boeke JD, 2005.

Introducción.

Sin embargo, en algunos casos, la transposición ha servido para establecer nuevos mecanismos de control de la expresión génica, que han sido reclutados a lo largo de la evolución por los distintos genomas.

Peaston y colaboradores han descrito que los retrotransposones pueden ayudar a regular de forma simultánea un conjunto de funciones que se expresan a lo largo del desarrollo de los oocitos y embriones preimplantacionales de ratón (Peaston AE *et al.*, 2004). En este trabajo han detectado que un 13% de todos los transcritos procesados de oocitos maduros de ratón contienen secuencias de retrotransposones, en su mayoría pertenecientes a la familia MT de retrotransposones no autónomos. Alrededor de un cuarto de estos transcritos son mRNAs quiméricos, correspondientes a genes de ratón, previamente identificados, que contienen secuencias alternativas derivadas de retrotransposones en su región 5'. Estas secuencias actúan como promotores alternativos, regulando a lo largo del desarrollo la expresión de la amplia variedad de genes a los que aparecen fusionados.

En un trabajo publicado este año, Muotri y colaboradores muestran que el retrotransposón L1 humano es capaz de transponer *in vitro* en células precursoras neuronales de ratón, afectando a la expresión de varios genes neuronales (Muotri AR *et al.*, 2005). Los autores postulan que este proceso podría ocurrir de forma natural en el tejido nervioso y que la actividad de los elementos transponibles podría causar algunos de los cambios genéticos que determinan la unicidad de cada individuo, al definir su circuito neuronal.

Un caso similar a los anteriores es el de los retrotransposones Wis 2-1A, cuya expresión en nuevas plantas anfiploides de trigo (generadas a partir de hibridaciones interespecíficas) resulta en la transcripción de secuencias de genes adyacentes en el mismo transcrito del transposón (Kashkush K *et al.*, 2003). De esta manera, los genes correspondientes pueden estar activando o silenciando. Este último caso se puede dar cuando el gen se encuentra en orientación inversa al elemento transponible y se transcribe la cadena complementaria.

En *Drosophila*, se ha observado que la expresión de un gen situado junto al punto de rotura proximal de la inversión 2j se ve drásticamente reducida como consecuencia de la transcripción de la cadena complementaria de este gen a partir de un transposón que se ha insertado al otro lado del punto de rotura (Puig M *et al.*, 2004).

Introducción.

Recientemente, se ha visto que la transcripción de la cadena complementaria del DNA es un proceso general en mamíferos y se ha propuesto que puede ser un mecanismo de control de la expresión de muchos genes (Katayama S *et al.*, 2005).

Se han descrito muchos otros ejemplos en los que los ETs pueden estar afectando a la expresión de los genes. En un trabajo en el que se analizaron unas 2000 regiones promotoras de genes humanos se vio que casi el 25% de las mismas contenían secuencias derivadas de transposones, demostrando el potencial de los ETs para afectar a la expresión de los genes humanos al aportarles nuevos elementos reguladores en *cis* (Jordan IK *et al.*, 2003). Así mismo, en otro estudio que abarcaba casi 14000 genes humanos reveló que un 4% de los mismos contenían inserciones de ETs dentro de su secuencia transcrita y que la gran mayoría de estas inserciones (cerca del 90%) se daba en intrones preexistentes, pudiendo resultar en la creación de nuevos exones (Nekrutenko A and Li WH, 2001). En este mismo análisis se vio que en más de 100 casos los elementos transponibles eran responsables de la divergencia entre genes ortólogos de diferentes especies (humanos, ratón y rata). En los dos estudios anteriores los elementos transponibles identificados más abundantes pertenecían a la familia *Alu* de SINEs.

Finalmente, la regulación epigenética de los ETs (que se explica en el siguiente apartado) también puede afectar a los genes adyacentes al transposón o a aquellos en los que se inserta una secuencia móvil. Éste es el caso del gen *Cabp* de ratón, en el que la inserción del retrotransposón IAP genera transcritos a partir de la LTR de este elemento, que están regulados por metilación (Druker R *et al.*, 2004).

3.4. EFECTO DE LOS TRANSPONES SOBRE LA EVOLUCIÓN DE LAS SECUENCIAS CODIFICANTES.

En algunos casos, el proceso de transposición ha servido para crear nuevas variantes génicas, nuevas proteínas o nuevos dominios proteicos (ver figura 6). Se han descrito varios ejemplos en los que distintos elementos transponibles son capaces de movilizar fragmentos de otros genes, dentro del genoma, como resultado de su transposición o promover reorganizaciones cromosómicas, como inversiones o la recombinación homóloga desigual entre dos copias semejantes de una misma familia de ETs (Hedges DJ and Batzer MA, 2005).

Introducción.

Los retrotransposones L1 humanos son capaces de movilizar secuencias situadas a 3' del elemento como resultado de un fallo en la terminación de la transcripción, que continúa más allá de la secuencia del elemento. Se ha demostrado la transducción de exones mediante este mecanismo *in vivo* (Moran JV *et al.*, 1999), así como la creación de nuevas isoformas proteicas derivadas de la actividad de los LINEs L1, promoviendo, así, la evolución proteica *en trans*. Éste es el caso de la aparición de una nueva isoforma del receptor de la leptina humano, resultante de la retrotransposición de un elemento SVA no autónomo mediada por la maquinaria de un elemento L1 (Damert A *et al.*, 2004).

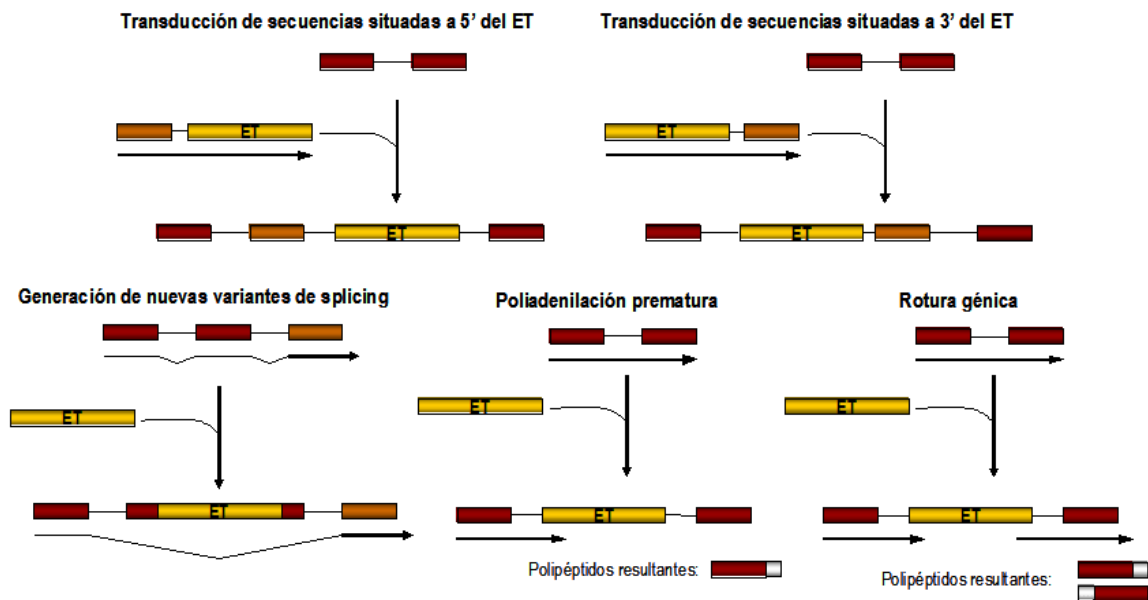


Figura 6. Posibles mecanismos de creación de nuevos alelos funcionales. Adaptado de Han JS and Boeke JD, 2005.

En plantas se ha observado la adquisición y fusión de fragmentos génicos por distintos tipos de transposones, tanto ETs de clase I (retrotransposón con LTRs *Bs1* de maíz), como de clase II (Pack-MULEs de arroz y los *Helitrones* de maíz). Además muchos de estos putativos genes quiméricos se expresan, aunque todavía no se ha demostrado la creación de un gen en plantas mediada por la actividad de un elemento transponible (Bennetzen JL, 2005).

Estos ejemplos, en humanos y plantas, sugieren que los transposones podrían redistribuir genes y fragmentos génicos dentro de los genomas y, de esta manera, ayudar a crear ordenaciones

génicas inéditas que sirvan de base para la evolución de nuevos genes y funciones génicas. Esta situación podría explicar la hipótesis de Barbara McClintock en la que define a los elementos transponibles como *máquinas* activas capaces de remodelar el genoma como respuesta al estrés (comentada anteriormente).

En este apartado se describen múltiples ejemplos de los efectos de la transposición sobre genes y genomas. En conjunto, todos estos datos ofrecen una clara visión del importante papel que han desarrollado los elementos transponibles en aportar plasticidad a los genomas y servir de fuente de variación de los mismos a lo largo de la evolución.

Pero la presencia de los ETs en los genomas no debe entenderse únicamente como el resultado de que puedan suponer un beneficio para los genomas, si no más bien a la tolerancia que estos últimos desarrollan frente a la naturaleza autónoma y móvil de los transposones. No podemos obviar el aprovechamiento de las características y funciones de los transposons por parte de los genomas. Algunos autores hablan del genoma como un *ecosistema* definido por la interacción de los elementos transponibles y los organismos en los que residen. E incluso utilizan el término *mutualismo* para explicar la presencia de las secuencias móviles y de las funciones derivadas de las mismas en los genomas (Brookfield JF, 2005). Otros autores hablan de *domesticación molecular* de los transposones por parte de los organismos, para terminar desarrollando funciones celulares (algunas de las cuales se citan en este apartado), que en ciertos casos son esenciales para el organismo (Miller WJ *et al.*, 1997; Hua-Van A *et al.*, 2005).

4. MECANISMOS DE CONTROL DE LOS TRANSPOSONES: SILENCIAMIENTO GÉNICO.

Los distintos organismos han desarrollado diversas estrategias para controlar la transposición y así evitar la proliferación de los elementos transponibles y minimizar así sus efectos deletéreos o capacidad mutagénica. La diversidad de los mecanismos de regulación está en relación a la diversidad de los procesos de transposición y de las familias de transposones. El control de la transposición puede darse en diferentes etapas del proceso, a nivel transcripcional, de la síntesis de las proteínas necesarias para la movilización del elemento o durante la escisión o inserción del ET (Hua-Van A *et al.*, 2005).

Introducción.

Como ya se ha indicado antes, el principal control de los retrotransposones de levadura, los elementos Ty, reside en su especificidad de inserción, que los dirige a regiones desprovistas de genes (Zhu Y *et al.*, 2003). Por el contrario, los ETs de genomas de plantas se insertan en cualquier tipo de secuencia y el principal mecanismo de control de su actividad es el control de su transcripción, transcribiéndose únicamente bajo ciertas situaciones de estrés (Grandbastien MA, 1998; Takeda *et al.*, 1999). Sin embargo, existen mecanismos de control que son probablemente generales de los genomas eucariotas y que regulan la actividad y proliferación de la mayoría de transposones y secuencias extragénicas. A estos procesos se les conoce bajo el nombre de silenciamiento génico y, probablemente, constituyen el principal mecanismo de control de las secuencias repetidas y exógenas en el genoma.

El silenciamiento génico ejerce su función a diferentes niveles. El silenciamiento transcripcional o TGS (Transcriptional Gene Silencing) actúa principalmente sobre las regiones promotoras de los genes, bloqueando la transcripción. Este proceso se desencadena, entre otros factores, por la presencia en el genoma de múltiples copias de una misma secuencia, por lo que podría ser el principal mecanismo de control de los elementos transponibles (Vaucheret H and Fagard M, 2001; Kuhlmann M *et al.*, 2005). El silenciamiento post-transcripcional o PTGS (Post Transcriptional Gene Silencing) opera a nivel del RNA, una vez transcrito, degradándolo o bloqueando su traducción. El PTGS tiene lugar, generalmente, cuando hay una acumulación anormalmente alta de ciertas especies de RNAs en el citoplasma. En plantas parece constituir el principal mecanismo de control de la multiplicación de virus de RNA (Vance V and Vaucheret H, 2001). Otros procesos epigenéticos, relacionados con el TGS y que afectan a la conformación de la cromatina, como la metilación y desacetilación de las histonas o la metilación del DNA, también juegan un papel importante en el control de la proliferación de las secuencias repetidas dentro de los genomas (Lippman Z and Martienssen R, 2004; Lippman Z *et al.*, 2003; Volpe T *et al.*, 2002; Volpe T *et al.*, 2003).

El efecto final del silenciamiento génico de los ETs es el bloqueo de la transposición, pero la forma de conseguirlo varía entre distintos transposones y organismos (Okamoto and Hirochika, 2001; Nolan T *et al.*, 2005). Los transposones Tc de *Caenorhabditis elegans* y el retrotransposón *Tad* de *Neurospora crassa* están regulados post-transcripcionalmente (Vastenhouw NL and Plasterk RH, 2004; Nolan T *et al.*, 2005). En plantas, en cambio, la mayoría de los elementos transponibles están silenciados a nivel transcripcional (Casacuberta JM and Santiago N, 2003). Varios trabajos publicados en los últimos años sugieren que los ETs de plantas están

Introducción.

controlados por TGS. Por ejemplo, se ha observado la reactivación de distintos transposones en mutantes de proteínas involucradas en el TGS de *Arabidopsis* (Lippman Z *et al.*, 2003). También se ha visto que, el retrotransposón Tto1 de tabaco se silencia en *Arabidopsis* cuando aumenta su número de copias en el genoma de este organismo heterólogo, por efecto de su transposición, y que este silenciamiento es debido a la hipermetilación de la secuencia del ET, como demuestra su reactivación en un contexto deficiente en metilación (Hirochika H *et al.*, 2000). Del mismo modo, el retrotransposón Tnt1 de tabaco también se silencia cuando se introduce en el genoma de *Arabidopsis* e incrementa su número de copias (Lucas H *et al.*, 1995).

Sin embargo, el TGS y el PTGS, lejos de ser procesos independientes, presentan muchos elementos comunes. Ambos mecanismos parecen estar mediados por pequeños RNAs (sRNAs) (Lippman Z and Martienssen R, 2004), que son moléculas de RNA de 21 a 26 nt de longitud, que proporcionan la especificidad al proceso y se acomplejan con las proteínas efectoras del silenciamiento génico. Se ha demostrado que, en plantas y levadura, la formación de heterocromatina está dirigida por pequeños RNAs de interferencia (siRNAs) asociados a un complejo proteico, al que se le ha denominado RITS, en el que participa la proteína Argonaute (Ago) y que presenta gran similitud con el complejo RISC, que interviene en el PTGS (figura 7) (Noma K *et al.*, 2004; Zilberman D *et al.* 2003; Lippman Z *et al.*, 2003; Volpe T *et al.*, 2002).

Los siRNAs son los productos del procesamiento de moléculas de RNA de doble cadena (dsRNA), llevado a cabo por un enzima con actividad helicasa y RNasIII, denominado Dicer. Se han propuesto diversos mecanismos por los que pueden generarse dsRNAs derivados de ETs, como la transcripción de elementos adyacentes en orientación inversa, la formación de estructuras secundarias dentro del propio transcrito del transposón, como consecuencia de las repeticiones terminales o internas, o la acción de enzimas con actividad RNA polimerasa dependiente de RNA (RdRP), que utilizan como molde RNAs de cadena sencilla (ssRNA) o RNAs aberrantes (aRNAs) derivados de la propia secuencia del transposón (Okamoto H and Hirochika H, 2001; Vastenhouw NL and Plasterk RH, 2004).

Los siRNAs generados pueden guiar bien la degradación o bien una represión traduccional de RNAs homólogos, a través de su asociación con el complejo proteico RISC (complejo del silenciamiento inducido por RNA), cuyo componente central son proteínas de la familia Argonaute (Ago) (Hammond SM, 2005). Pero los siRNAs también se pueden asociar a otro complejo proteico, denominado complejo RITS (complejo de inicio del silenciamiento génico

Introducción.

transcripcional inducido por RNA), cuyo componente central también son proteínas de la familia Argonaute y que se asocia, a su vez, a enzimas con actividad modificadora de la cromatina (como metilasas del DNA o de histonas) y de mantenimiento de la heterocromatina (Verdel A *et al.*, 2004; Noma K *et al.* 2004).

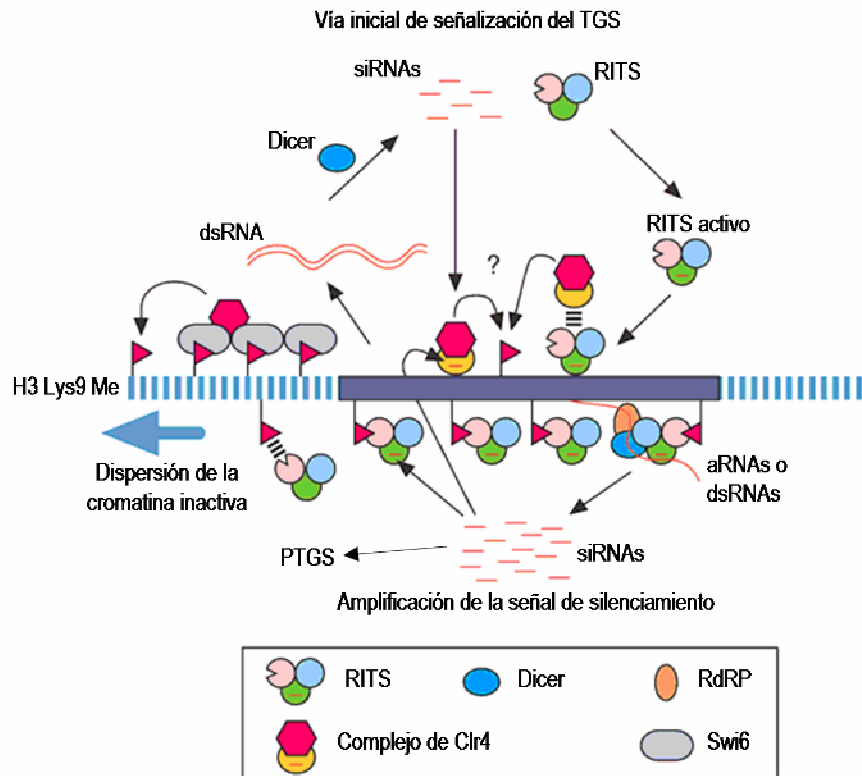


Figura 7. Modelo de la participación de los siRNAs y RITS en el silenciamiento génico transcripcional (TGS) y su relación con el silenciamiento génico post-transcripcional (PTGS), de las secuencias centroméricas de *S. pombe*. Este modelo podría aplicarse también al silenciamiento de los elementos transponibles. La proteína Clr4 es un enzima modificador de la cromatina, que se encarga de metilar las histonas. La proteína Swi6 es una proteína de definición o mantenimiento de la heterocromatina. Adaptado de Noma K *et al.*, 2004.

El DNA metilado puede estar dirigiendo y/o manteniendo, a su vez, el silenciamiento de secuencias homólogas. Varios trabajos publicados este año muestran que en plantas existe una RNA polimerasa (Pol IV) que es capaz de transcribir loci metilados, originando un RNA que podría servir para la producción de siRNAs, asegurando el mantenimiento del silenciamiento de estas regiones e induciendo, así, el TGS de otros loci homólogos (Kanno T *et al.*, 2005; Herr AJ *et al.*, 2005; Onodera Y *et al.*, 2005).

Introducción.

Se han identificado sRNAs con homología a distintos elementos transponibles en diferentes organismos, como *C. elegans* (Sijen T and Plasterk RH, 2003), *Drosophila* (Aravin AA *et al.*, 2003), *Trypanosoma brucei* (Djikeng A *et al.*, 2001), tabaco (Hamilton A *et al.*, 2001), arroz (Sunkar R. *et al.*, 2005) o *Arabidopsis* (Llave C. *et al.*, 2002), indicando la existencia de un mecanismo activo y general de silenciamiento que controla la expresión y proliferación de estas secuencias repetidas autónomas en los genomas.

De una forma indirecta, el silenciamiento de un transposón puede llevar a su pérdida evolutiva. La fuerte metilación de las secuencias transponibles silenciadas puede incrementar su tasa de mutación, llevando a su completa inactivación. Por otra parte, la probabilidad de que un transposón se pierda definitivamente en la evolución está directamente relacionada con el tiempo durante el que este elemento se mantenga inactivo.

El control de la transposición es fundamental para la supervivencia de los distintos organismos. Pero este control no significa la inactivación definitiva de ciertos elementos transponibles. En la mayoría de los genomas, existen transposones capaces de movilizarse. Los retrotransposones de plantas, por ejemplo, pueden activarse en condiciones de estrés (Grandbastien MA, 1998; Takeda *et al.*, 1999). Esto sugiere que la relajación de la represión de la transposición y/o las estrategias adoptadas por ciertos elementos transponibles para poder expresarse han permitido la amplificación de este tipo de secuencias dentro de los genomas. Pero el mantenimiento y coexistencia de los transposones en los genomas a lo largo de la evolución es el balance resultante de la capacidad de amplificación del transposón (mutagenicidad) y el control ejercido sobre este proceso.

OBJETIVOS.

OBJETIVOS.

Con el trabajo que aquí se presenta hemos querido analizar como se controlan los elementos transponibles y el impacto que tienen sobre la evolución de los genes y los genomas. Para ello hemos tomado dos elementos transponibles de plantas como modelo y nos hemos establecido los siguientes objetivos:

1. Identificar, en el genoma de Arabidopsis, todos los elementos de tipo MITE pertenecientes a la familia *Emigrant* existentes.
2. Analizar la evolución de los MITEs *Emigrant* en el genoma de Arabidopsis.
3. Analizar el impacto de los transposones *Emigrant* sobre los genes de Arabidopsis.
4. Analizar la terminación de la transcripción y la poliadenilación de los mRNAs del retrotransposón Tnt1 de tabaco.
5. Analizar si la expresión del retrotransposón Tnt1 de tabaco puede estar controlada por procesos epigenéticos.
6. Determinar el efecto del estrés sobre el silenciamiento génico transcripcional.
7. Analizar la correlación de la expresión del retrotransposón Tnt1 y cambios en el silenciamiento génico de estos elementos.

RESULTADOS Y DISCUSIÓN.

“Mis evidencias convencen al ignorante, y las evidencias del sabio me convencen a mí. Pero aquel cuyo razonamiento se halla a mitad de camino entre la sabiduría y la ignorancia, ni puedo convencerle yo a él, ni puede convencerme él a mí”.

Khalil Gibran. *Máximas espirituales.*

CAPÍTULO 1. ANÁLISIS DE LA FAMILIA *Emigrant* DE MITEs DE *Arabidopsis thaliana*.

INTRODUCCIÓN.

El trabajo presentado en este capítulo persigue aportar nuevos datos acerca del impacto de los elementos transponibles sobre la evolución de los genomas. En concreto, aquí se estudia la historia evolutiva de los MITEs de la familia *Emigrant* en el genoma de la planta modelo *Arabidopsis thaliana*. Con este objetivo hemos querido identificar y caracterizar todas las secuencias contenidas en el genoma de *Arabidopsis* pertenecientes a la familia *Emigrant* de MITEs.

A diferencia de lo que ocurría cuando fueron descritos los primeros elementos *Emigrant* (Casacuberta E *et al.*, 1998), en el momento en el que se comenzó el presente estudio la mayor parte del genoma de *Arabidopsis thaliana* ya estaba secuenciada y disponible para su análisis. Entonces se pensó en diseñar una estrategia para poder encontrar todos los *Emigrant* contenidos en el mismo. En colaboración con el grupo de Xavier Messeguer, del Departament de Llenguatges i Sistemes Informàtics (LSI) de la Universitat Politècnica de Catalunya (UPC), se creó un programa informático que nos permitiera realizar esta tarea.

Debido a que los MITEs se caracterizan por no presentar secuencia codificante alguna, éstos son elementos muy variables en secuencia. Por lo tanto, una búsqueda de nuevos elementos basada en la similitud de su región interna con los *Emigrant* previamente descritos, como se había estado haciendo hasta el momento con este tipo de transposones, obviaría aquellas secuencias pertenecientes a distintas subfamilias así como aquellas que, perteneciendo a la misma subfamilia, pudieran haber divergido a lo largo del tiempo. La originalidad de la estrategia que hemos seguido, descrita a continuación, es el resultado de nuestra pretensión de identificar el máximo de secuencias, sino todas, pertenecientes a la familia *Emigrant* de MITEs contenidas en el genoma de *Arabidopsis*.

1. DISEÑO DE UN PROGRAMA INFORMÁTICO PARA LA BÚSQUEDA DE MITEs.

Los MITEs son transposones de pequeño tamaño que, de igual forma que otros ETs de clase II, presentan repeticiones terminales invertidas (TIRs) flanqueando su secuencia interna. Los MITEs se han agrupado en distintas familias en función de la secuencia de sus TIRs.

En este proyecto se pensó en crear un programa informático que fuera capaz de identificar transposones del tipo MITE. Para ello nos pusimos en contacto con el grupo del Profesor Xavier Messeguer y definimos las bases del programa. Queríamos diseñar una herramienta informática que fundamentara la búsqueda en las secuencias de los TIRs, teniendo en cuenta también otras características de los MITEs, como su pequeño tamaño o la presencia de repeticiones de la diana de inserción.

Cumpliendo los requisitos definidos se concibió el programa TRANSPO (Herráiz C, 2002), que busca, en una secuencia dada, dos secuencias repetidas e invertidas separadas entre sí un cierto rango de distancias. El programa fue diseñado de tal forma que pueda ser aplicado para buscar cualquier secuencia que presente las características de los MITEs, ya que es el usuario quien define los distintos parámetros requeridos (secuencia de los TIRs, horquilla de distancias entre las que buscar y la secuencia genómica sobre la que realizar el análisis).

Con el objetivo de comprobar la existencia de las duplicaciones de la diana de inserción (resultado del proceso de transposición) definimos que el programa TRANSPO también proporcionara, junto con el MITE identificado, las secuencias que lo flanquean por cada extremo. El hecho de no utilizar una región interna como objeto de la búsqueda permite la detección de nuevos elementos, no descritos previamente, que comparten similitud en sus TIRs pero no en el resto de su secuencia. De esta forma se pueden encontrar subfamilias de MITEs compuestas por muy pocas secuencias, así como copias antiguas de MITEs que han perdido gran parte de homología de secuencia debido a sucesos de inserción, delección o mutaciones puntuales en su región interna. Para facilitar todavía más la identificación de copias antiguas de MITEs pertenecientes a una familia concreta, propusimos al grupo de informática añadir otra variable a la búsqueda. Teniendo en cuenta que los TIRs, de la misma forma que el resto de la secuencia de los MITEs, pueden ir acumulando mutaciones a lo largo de las generaciones, pensamos en permitir cierta variabilidad dentro de la secuencia de los TIRs definida para la búsqueda. El grado de diferencias respecto a la secuencia original también puede ser determinado por el usuario.

2. IDENTIFICACIÓN Y CARACTERIZACIÓN DE NUEVOS MITES *Emigrant* EN EL GENOMA DE ARABIDOPSIS.

Una vez definidos todos los parámetros que considerábamos necesarios para la identificación de transposones del tipo MITE, se utilizó el programa TRANSPO para identificar nuevos elementos de la familia *Emigrant* de MITES sobre la secuencia completa disponible de Arabidopsis (que comprende 115.4 Mb de los 125 Mb que tiene su genoma, al no incluir telómeros, centrómeros o regiones repetidas de rDNA). La secuencia utilizada en la búsqueda ha sido el consenso de los TIRs de *Emigrant* (CAGTAAAACCTCTATAAATT, Casacuberta E *et al.*, 1998), tolerando hasta un 25% de divergencia. Hemos definido una horquilla de distancias entre dos repeticiones invertidas que comprende desde los 200 nt a los 700 nt, de acuerdo con el rango de tamaños de los distintos MITES descritos previamente (revisado en Feschotte C *et al.*, 2002).

El programa TRANSPO genera un archivo con el resultado del análisis realizado. Este archivo es un listado de todas las secuencias identificadas como MITES de la familia *Emigrant* encontrados en la secuencia del genoma de Arabidopsis. Para cada secuencia se especifica el cromosoma en el que se encuentra y su posición dentro del mismo, los 30 nt que flanquean el transposón por ambos lados, la secuencia y localización en el cromosoma de los TIRs identificados y la posición, longitud y secuencia de la región interna (ver figura 1).

```
e109>
4 - 4574531
LEFT_TARGET: ATTCTTGCCATACGTTTTCTTCCATATA
TIR: CAGTAAAACt*TCTATAAATT start-end: 4574531-4574550
BODY: length: 492 start-end: 4574551-4575041
RIT: AATTTATAc*AGGTTTTACTG start-end: 4575042-4575061
RIGHT_TARGET: TACTCTAACTCATCTTAATTC AATTTAACCC

AATACTCGTTAAATTAATAACTTCTATAAATTAATAAATTCCTGTTCTAAGTTGGGAACAGTGTA AAAAGTGACACAAATCGATAAATT
GATAAGATAATAATTTTTTGAAAATCTTATATAAATTTATGGTCCCATCAATATCATAAATTAATAATTGTATGAATGTATCAATTATATATA
TACATTATATGTA AAAAGTTTCTTTGAAATATAATTCTAATATTTGTTGCTTAAATTTGTATTTATTCCTATTTGAAATTTAGTTTTAGTATTT
TACTGTATCAAAAACCTTTTGGTGTGTTTTCTAAACATGATATTTTATTACTTGTAATTGATTCTTGAAAGATATTTCTATAATGAAAAAA
ATCCTATAGAAATTTTTGAAAAGTTACTAAATTAGGAAAATCTCTCTATAAATTAATAATTAATTAACCTTATCGATAAATTA AACCTCTCAAA
ATTTTTGCAGTCCCAACATTTTT
```

Figura 1. Ejemplo de la información proporcionada por el programa TRANSPO correspondiente a una secuencia identificada como un MITE *Emigrant*. En la primera línea se indica el número de la secuencia y en la segunda el cromosoma en el que se encuentra (el 4, en este caso) y la posición donde comienza el MITE. En las siguientes filas aparecen anotadas las secuencias flanqueantes, secuencia y posición de la terminaciones repetidas invertidas (TIR y RIT) y la longitud, posición y secuencia (en último lugar) de la región interna del elemento transponible. La duplicación de la diana de inserción TA se indica en rojo. Las diferencias en la secuencia de los TIRs respecto a la secuencia consenso aparecen indicadas con una letra minúscula y un asterisco.

De esta forma, han sido identificados 151 posibles elementos *Emigrant*. Algunos de estos MITEs encontrados coinciden, como cabía esperar, con los ya descritos anteriormente (Casacuberta E *et al.*, 1998). Todos los elementos detectados son ricos en A+T, no poseen capacidad codificante y la mayoría están flanqueados, en ambos extremos, por el dinucleótido TA. Se han descartado aquellas secuencias solapadas, generadas a partir de las repeticiones subterminales. Aunque las características anteriores son típicas de secuencias relacionadas con MITEs, la mayoría de las encontradas en este trabajo no aparecían anotadas como miembros de la familia *Emigrant*, ni como MITEs u otros posibles ETs en las bases de datos.

La disponibilidad de los 60 nt de las regiones flanqueantes a cada elemento *Emigrant* nos ha permitido realizar una búsqueda de RESites (Related Empty Sites) o sitios vacíos asociados (Le QH *et al.*, 2000; Petersen G and Seberg O, 2000; Tu Z, 2001) que, en genomas como el de *Arabidopsis* (en el que las regiones duplicadas comprenden un 60% del mismo, AGI, 2000), representan sucesos de duplicación de regiones genómicas anteriores a la inserción de elementos transponibles en alguna de ellas. La existencia de estos RESites se ha utilizado previamente como indicador de transposición al analizar putativos ETs dentro de genomas concretos (Le QH *et al.*, 2000; Tu Z, 2001). Mediante búsquedas por homología de secuencia sobre la base de datos del NCBI (GenBank), a partir de las regiones flanqueantes a los *Emigrant* identificados, hemos encontrado 14 RESites bien conservados. Estos RESites corresponden a los elementos e7r, e19, e20, e25r, e26, e44r, e46r, e57, e67, e83r, e85r, e89r, e139r y e143r. En la figura 2 se muestran algunos ejemplos.

El resultado anterior, junto con la presencia de la duplicación TA acompañando a gran parte de las secuencias identificadas, sugiere que todos los elementos encontrados en este análisis son verdaderos transposones pertenecientes a la familia *Emigrant* de MITEs.

Pero la estrategia que hemos seguido también tiene sus limitaciones. El programa TRANSPO no es capaz de encontrar aquellos MITEs que, debido a deleciones u otras mutaciones, carezcan de uno de los dos TIRs. De hecho, conocemos que existen elementos que han perdido uno de sus TIRs con posterioridad a su inserción (Casacuberta E *et al.*, 1998). La única manera de llegar a identificar estos otros miembros sería mediante una búsqueda basada en la homología de secuencia con el consenso generado a partir de las regiones internas de los MITEs obtenidos gracias al programa TRANSPO. De todas formas, no se consideró necesario realizar este análisis, de una manera sistemática, para el objeto de nuestro estudio, debido a que el número

de elementos *Emigrant* encontrados es significativamente grande y, además, se encuentra dentro del rango predicho en trabajos anteriores (Casacuberta E *et al.*, 1998) si tenemos en cuenta que en nuestro caso no hemos utilizado las secuencias centroméricas y teloméricas del genoma de *Arabidopsis*.

EmiA

```

flank e46r: 2   actacaaaaccetttcgggttggtttcataa e46r ta aattttattcttctaaatctccatata 58
              ||||| ||||| ||| ||||| ||||| |||
ATF26B15 : 31489 actataaaacccttcagggttggtttcataa aattttattcttctaaatcaccatttaga 31493
    
```

EmiB

```

flank e139r: 1   tcccttggcgaagagtgatgcaatttatata e139r ta tctttacatatttcttgetatatttt 58
              ||||| ||||| ||||| ||||| |||
AC005824 : 87109 tcccttggcgaagagtgatgcaattttttta tctttacatggttcttctgetatatttt 87052
    
```

EmiC

```

flank e25r: 1   tatccattgatcgtacataaagtgcacata e25r ta taccctaatttcataaagtctaagacaatg 58
              ||||| ||||| ||||| ||||| |||
AC051631: 95782 tatccattgatcgtacataaagtgcacata taccctaatttcataaagtttaagacaatg 95725
    
```

EmiD

```

flank e89: 1   gatattaatatattttgaaagttttattt e89 ta tagtaaccaatcgtgtggttaagtggaga 58
              ||||| ||||| ||||| ||||| |||
AP001309.1 : 64854 gatattaatatattttgaaagttttattt tagtaaccaatcgtgttagttaagttagaga 64797
    
```

Figura 2. RESites. Se muestra un ejemplo de RESites de elementos pertenecientes a cada uno de los grupos de MITEs *Emigrant*.

3. ANÁLISIS FILOGENÉTICO. Dentro del genoma de *Arabidopsis thaliana* existen diferentes subfamilias de elementos *Emigrant*.

Un primer intento de alinear y agrupar las 151 secuencias obtenidas del análisis anterior mostró la elevada variabilidad de la región interna de todos estos MITEs, siendo imposible llevar a cabo este propósito utilizando los métodos y herramientas habituales y disponibles. Debido a esto, y también en colaboración con el grupo del LSI, se desarrolló el programa SPAT (Goñi R, 2002), que compara secuencias dos a dos y elimina aquellas más divergentes de cada grupo. SPAT clasificó la mayoría de los 151 elementos *Emigrant* en tres grupos, llamados *EmiA* (compuesto por 41 secuencias), *EmiB* (26 secuencias) y *EmiC* (37 secuencias), cada uno de los cuales son

suficientemente homogéneos para ser sometidos a un análisis filogenético convencional. 47 secuencias son tan divergentes que SPAT no ha podido incluirlas en ninguno de los grupos definidos y los hemos agrupado bajo el nombre de *Emi0*. Todos los miembros de la familia *Emigrant* previamente descritos (Casacuberta E *et al.*, 1998) pertenecen al grupo *EmiA*. Se ha utilizado el programa CLUSTALW para alinear las secuencias pertenecientes a cada uno de los anteriores grupos y los alineamientos obtenidos han servido para generar árboles filogenéticos por neighbor-joining. La figura 3 muestra los árboles resultantes. Dentro de cada uno de los árboles se pueden establecer diferentes grupos monofiléticos definidos por valores de bootstrap elevados (que representan la robustez de los nodos establecidos), de forma que la mayoría de las secuencias se clasifican en tres subfamilias diferentes dentro de cada grupo *Emigrant* (A1, A2 y A3; B1, B2 y B3; C1, C2 y C3).

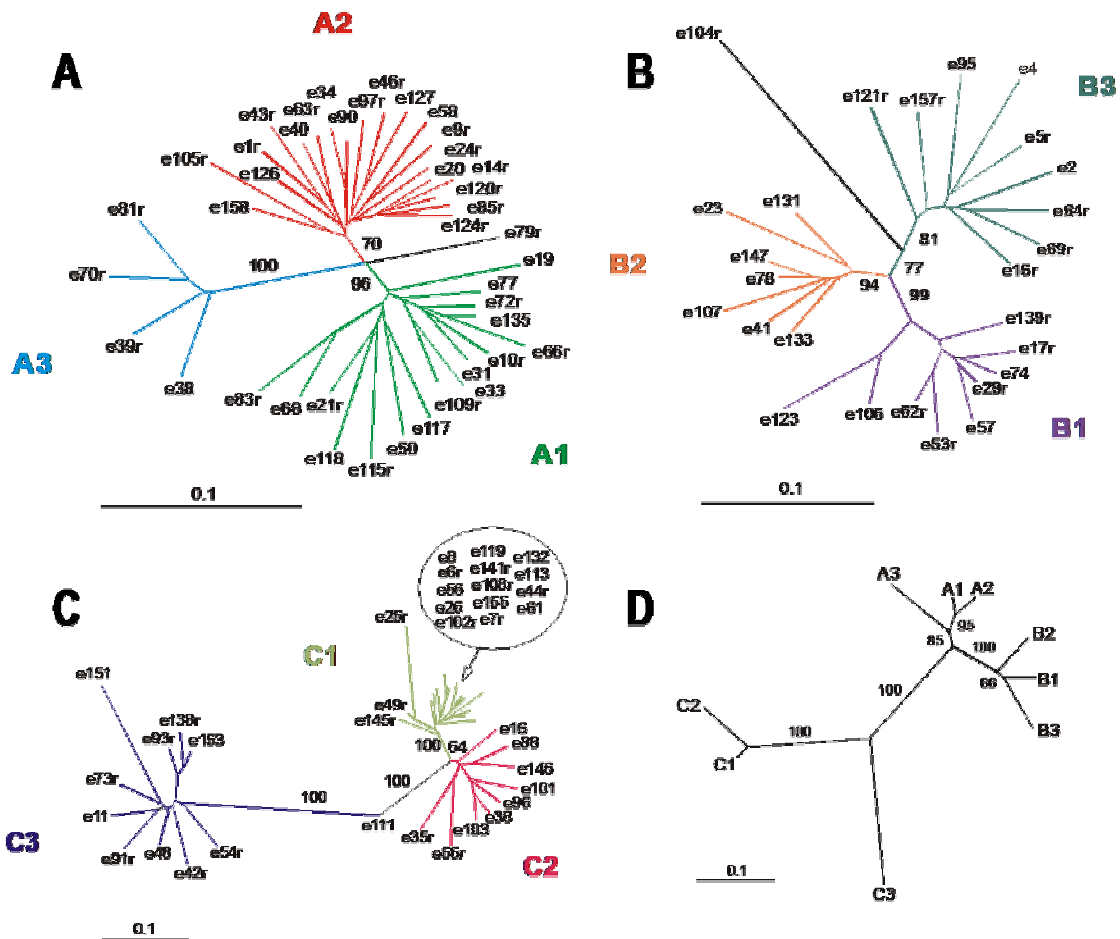


Figura 3. Análisis filogenético de los elementos *Emigrant*. Árboles obtenidos por el método de neighbor-joining correspondientes a los grupos *EmiA* (A), *EmiB* (B) y *EmiC* (C). (D) Árbol neighbor-joining obtenido por comparación de los consensos de las distintas familias *Emigrant* de MITes.

Las subfamilias han sido tratadas de la misma forma que los grupos originales y, a partir de los alineamientos generados, se ha deducido una secuencia consenso para cada una de ellas. Con

tal de conocer las relaciones filogenéticas existentes entre las subfamilias de *Emigrant* se han comparado los 9 consensos (figura 3). Las tres subfamilias del grupo *EmiA*, así como las tres subfamilias *EmiB* y las tres *EmiC* parecen estar relacionadas filogenéticamente entre ellas, tal y como muestran los valores de *bootstrap* de los nodos que agrupan estas subfamilias de tres en tres entre ellas.

4. ANÁLISIS EVOLUTIVO.

4.1. Análisis de la variabilidad de secuencia y tamaño entre los distintos grupos de *Emigrant*.

Una vez conocido que las secuencias identificadas pueden agruparse en subfamilias y conocidas también las relaciones existentes entre éstas, se pasó a estudiar cuán homogéneas eran estas subfamilias para intentar tener nociones de la antigüedad de las mismas. Tanto la diversidad nucleotídica (π) como la variabilidad de tamaño dentro de cada subfamilia se calcularon a partir de los alineamientos anteriormente obtenidos. La variabilidad de tamaño se ha calculado respecto al tamaño medio de cada subfamilia. Debido a que los elementos agrupados en *Emi0* no presentan similitudes significativas entre sí, no han sido sometidos a estos análisis.

Tabla 1. Variabilidades de secuencia y de tamaño

	#	Diversidad nucleotídica (π)	Desviación estándar de π	Variabilidad de tamaño
<i>EmiA1</i>	16	0.108	0.0069	5.6%
<i>EmiA2</i>	20	0.095	0.0040	2.7%
<i>EmiA3</i>	4	0.081	0.0149	5.2%
<i>EmiB1</i>	9	0.102	0.0142	3.9%
<i>EmiB2</i>	7	0.077	0.0097	3.8%
<i>EmiB3</i>	9	0.115	0.0092	8.2%
<i>EmiC1</i>	17	0.068	0.0059	7.1%
<i>EmiC2</i>	9	0.112	0.0087	4.2%
<i>EmiC3</i>	10	0.131	0.0167	5.9%
<i>Emi0</i>	47	n.d.	n.d.	n.d.

NOTA.- En la segunda columna aparece el número de elementos que componen cada subfamilia. No se determinan (n.d.) los datos para las secuencias *Emi0* porque su cálculo no tiene sentido, al no ser un grupo homogéneo.

Los resultados obtenidos muestran que las distintas subfamilias presentan diferentes grados de variabilidad (tabla 1). Mientras que algunas son muy homogéneas tanto en secuencia como en

tamaño, véase *EmiA2*, otras, como *EmiB3*, son mucho más variables. Cada uno de los tres grupos originales (*EmiA*, *EmiB* y *EmiC*) contiene subfamilias de distinta variabilidad. Como ejemplo, las subfamilias *EmiA2* y *EmiB2* son muy homogéneas tanto en secuencia como en tamaño, *EmiC1* es muy homogénea en secuencia pero es relativamente variable en tamaño y, en cambio, subfamilias como *EmiC3* y *EmiB3* son bastante heterogéneas.

La existencia de diferentes grupos de elementos *Emigrant* en el genoma de *Arabidopsis* y el hecho de que cada una de ellas se subdivide en subfamilias de distinta variabilidad, indica que la transposición de estos elementos se ha producido mayoritariamente siguiendo sucesos de amplificación de unas pocas secuencias en determinados momentos de la evolución. Tras la inserción en sus nuevas localizaciones dentro del genoma, las diferentes copias han ido divergiendo y perdiendo su homología. La topología de los árboles filogenéticos obtenidos en forma de estrella, con ramas muy cortas dentro de cada grupo, apoya la hipótesis de que se han producido distintas “explosiones” de amplificación a partir de pocas secuencias. Pero además, parece ser que cualquier secuencia es susceptible a ser amplificada en un momento dado, ya que encontramos subfamilias de elementos *Emigrant* más homogéneas y otras más heterogéneas dentro de los diferentes grupos definidos.

4.2. Distribución.

De igual forma que ocurre en otros organismos (ver introducción de esta memoria) se ha observado que ciertos transposones se insertan preferentemente en regiones concretas del genoma de *Arabidopsis* (Kapitonov VV and Jurka J, 1999; AGI, 2000). Éste es el caso de los centrómeros, donde se concentran los ETs de clase I, o de las regiones pericentroméricas y otros dominios heterocromáticos, donde se agrupan los ETs de clase II (AGI, 2000). En este trabajo hemos querido conocer si los MITEs de la familia *Emigrant* presentan también una cierta predilección por los dominios pericentroméricos heterocromáticos, como se ha descrito anteriormente (AGI, 2000), o bien no posee ninguna preferencia manifiesta.

Los datos proporcionados por el programa TRANSPO incluyen, como se ha explicado antes, la posición cromosómica que ocupan los elementos identificados. Hemos analizado estos resultados y los hemos utilizado también para obtener una representación gráfica de la distribución de los 151 elementos *Emigrant* sobre los cinco cromosomas que conforman el

genoma de *Arabidopsis*. Para realizar esta tarea hemos diseñado, también en colaboración con el grupo del LSI, el programa CLUPH (Goñi R, 2002). En la figura 4 se representan los resultados generados por el programa CLUPH.

Todos los cromosomas contienen secuencias *Emigrant*, aunque con pequeñas diferencias en su densidad, variando de los 1.8 *Emi*/Mb que aparecen en el cromosoma 4 a los 1.0 *Emi*/Mb en el cromosoma 5. Aunque la concentración de los distintos grupos de *Emigrant* también varía ligeramente entre los 5 cromosomas, todos éstos contienen representantes de cada grupo. Las diferencias de concentración de los elementos *EmiA* llegan a ser hasta seis veces mayores entre los distintos cromosomas (comparación entre cromosomas 4 y 5); mientras que en el caso de los elementos *EmiB*, el cromosoma más denso en estas secuencias (cr. 1) no llega a duplicar la concentración del menos abundante en las mismas (cr. 5).

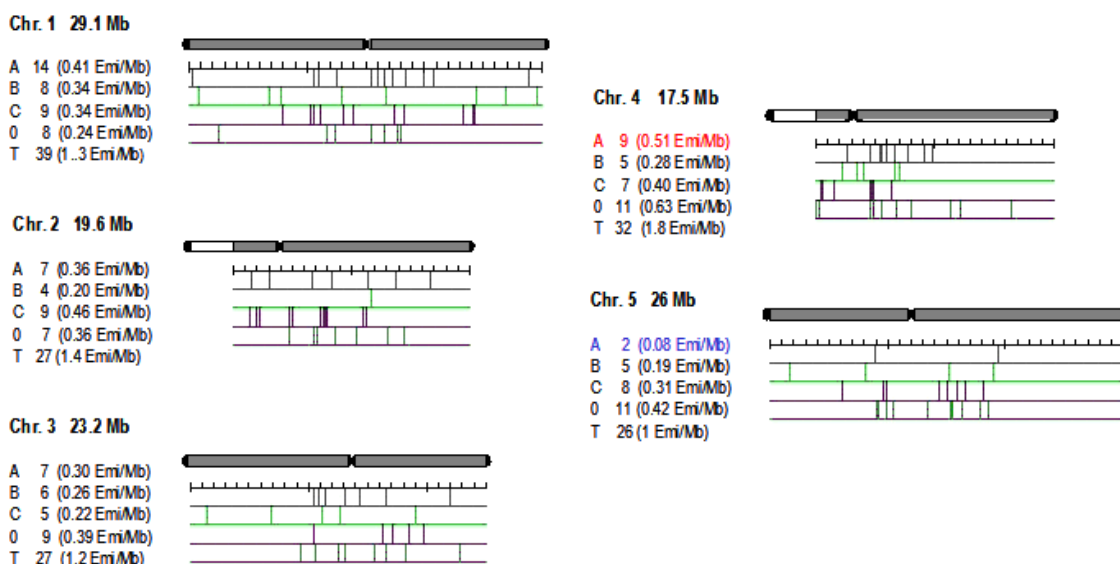


Figura 4. Representación de los 151 elementos *Emigrant* en los cinco cromosomas de *A.thaliana*. Se indica el número de MITEs *Emigrant*, divididos en grupos, por Mb.

Además, los MITEs de la familia *Emigrant* se distribuyen a lo largo de todo el cromosoma, indistintamente del cromosoma o familia de elementos que tengamos en consideración. Parece que los transposones *Emigrant* no muestran preferencia por insertarse en regiones heterocromáticas, ya que los elementos que, suponemos, se han movilizad más recientemente (aquellos pertenecientes a la subfamilia *EmiA2*), aparecen localizados en todos los cromosomas de forma más o menos homogénea, incluso sobre las regiones heterocromáticas de cuya secuencia se dispone, siendo el dominio pericentromérico del cromosoma 4 una de ellas.

Pero el genoma de *Arabidopsis* no estaba totalmente secuenciado en el momento en el que se realizó este trabajo, faltaba la secuencia de la mayor parte de las regiones centroméricas, teloméricas y de las repeticiones del rDNA, suponiendo una limitación para nuestro estudio. No podíamos identificar los MITEs *Emigrant* situados en estas regiones y por lo tanto no podemos saber si en estas regiones existe gran concentración de estos elementos, aunque estimas sobre el número de copias de estos transposones realizadas con anterioridad no parecen indicarlo (Casacuberta E *et al.*, 1998).

Para intentar solventar este impedimento pensamos en realizar hibridaciones fluorescentes *in situ* (FISH) sobre núcleos metafásicos de *Arabidopsis*. Con la técnica FISH se pueden localizar físicamente y directamente sobre los cromosomas distintas secuencias al mismo tiempo, utilizando fluoróforos diferentes para distintas sondas. De esta forma se diseñaron experimentos para hibridar los distintos MITEs junto con sondas específicas de regiones heterocromáticas o de telómeros. Estas hibridaciones presentan gran dificultad debido a las características de las secuencias de los MITEs ya que son elementos de pequeño tamaño y su secuencia interna es altamente divergente. Probablemente ésta es la causa por la que no funcionaron las hibridaciones realizadas con la secuencia de diferentes MITEs.

4.3. Sitios de inserción.

La integración de los transposones dentro de la secuencia de DNA es una etapa crítica en el proceso de transposición, siendo uno de los pasos más controlados de este mecanismo. La mayoría de los ETs no se integran en el DNA al azar, sino que presentan cierto grado de especificidad sobre la secuencia en la que se integran (Vigdal TJ *et al.*, 2002; Wu X and Burgess SM, 2004). Muchos ETs se insertan preferentemente en regiones ricas en otras secuencias repetidas, ricas en A+T (pobres en genes), de forma que minimizan su efecto deletéreo. En nuestro trabajo también hemos querido conocer si los MITEs de la familia *Emigrant* se integran en un determinado tipo de secuencias.

Las características de los sitios en los que se insertan esta familia de MITEs se han estudiado mediante el análisis de los 60 nt que flanquean a cada uno de los 151 elementos *Emigrant*. Los transposones *Emigrant* parecen dirigirse a regiones muy ricas en A+T, de igual forma que lo hacen otros MITEs o MLEs (mariner-like elements) (Le QH *et al.*, 2000). Las secuencias

flanqueantes a los elementos *Emigrant* están constituidas en un 74.3% por A+T, cuando el contenido promedio de A+T en el genoma de *Arabidopsis* varía entre el 64.5% y el 66.6%, llegando a alcanzar un 67.7% en regiones no codificantes (AGI, 2000). Además, aunque el microsatélite más frecuente en el genoma de *Arabidopsis thaliana* consiste en repeticiones TA (Casacuberta E *et al.*, 2000), no parece que los *Emigrant* se inserten en microsatélites, puesto que sólo el 3% de los 151 elementos aquí descritos aparecen en secuencias con más de cuatro repeticiones del dinucleótido TA.

Hemos buscado si existe alguna especificidad de secuencia de inserción para los elementos *Emigrant*. Para ello hemos comparado las secuencias flanqueantes de los miembros de la subfamilia *EmiA2*, puesto que son los que suponemos han transpuesto más recientemente y, por tanto, cuya posible secuencia de inserción habrá variado menos. En la figura 5 se muestra el consenso de las secuencias flanqueantes a estos elementos.

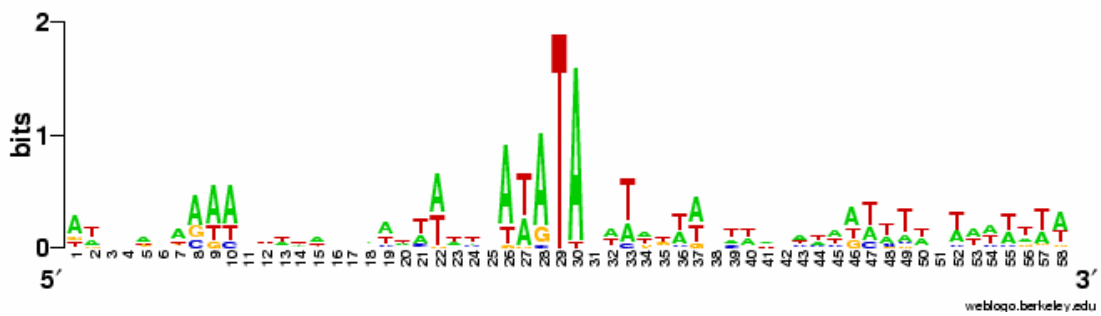


Figura 5. Análisis de las regiones flanqueantes a los MITEs *Emigrant* de la subfamilia A2. Se han comparado las secuencias que preceden y suceden a todos los elementos *EmiA2* y se representa la proporción de las bases que ocupan cada posición, obtenidas mediante el programa Seqlogo. El eje de ordenadas representa la robustez de la información. El máximo para una secuencia de DNA corresponde a 2 bits. El elemento *Emigrant* se localiza entre las bases 29 y 30, que en la gran mayoría de casos corresponden con el dinucleótido TA.

Los MITEs *Emigrant* presentan una clara preferencia por insertarse dentro del dinucleótido TA, el cual duplican cuando se integran en esta posición. Esta predilección también ha sido observada en más familias de MITEs diferentes, mientras que otras se insertan en otras secuencias como TAA o TTAA (Feschotte C *et al.*, 2002). Sin embargo, esta predilección no es estricta, sino que más bien, parece que es la presencia de una alternancia de TAs en la región del DNA en la que se integran los elementos *Emigrant* la que define el lugar de inserción de estos transposones.

Se ha propuesto que es la estructura del DNA y no tanto la secuencia del mismo es la que puede estar determinando la especificidad de inserción de los elementos transponibles. Determinadas

conformaciones de la molécula del DNA pueden facilitar el acceso de la maquinaria de transposición, que integrará la nueva copia del ET (Vigdal TJ *et al.*, 2002).

La presencia de abundantes timinas y adeninas en la secuencia pueden estar creando estructuras secundarias en la molécula del DNA, que, además, al poseer menos enlaces de hidrógeno entre las bases, es más maleable. Quizás es esta característica la que sirve como diana para la inserción de las nuevas copias de *Emigrant*.

4.4. Localización de los elementos *Emigrant* respecto a los genes.

A pesar de que los MITEs muestran una preferencia de inserción por regiones ricas en A+T, frecuentemente se han encontrado en secuencias génicas (Wessler SR *et al.*, 1995; Yang G *et al.*, 2001). Pero esta asociación de los MITEs con regiones ricas en genes puede ser debida a que la identificación de nuevos ETs ha derivado, durante mucho tiempo, del estudio de los genes y sus secuencias reguladoras y codificantes. Actualmente, gracias a la disponibilidad de la secuencia de los genomas de algunos organismos como *Arabidopsis* o arroz (AGI, 2000; Goff SA *et al.*, 2002; Yu J *et al.*, 2002), se pueden diseñar estrategias independientes al estudio de ESTs para identificar todas las copias de elementos transponibles, MITEs en concreto (como la descrita en este trabajo), presentes en los distintos genomas.

Por otra parte, la localización de los ETs en el genoma es el resultado de su especificidad de inserción junto con la presión selectiva ejercida sobre los mismos. Partiendo de esta premisa, decidimos llevar a cabo un estudio evolutivo de la asociación de los MITEs *Emigrant* con los genes y el papel que ha jugado la selección sobre la distribución de estos elementos.

En el trabajo en el que se describió el primer elemento *Emigrant* (Casacuberta E *et al.*, 1998) se observó que los MITEs de esta familia no se encontraban asociados a genes. Pero este trabajo sólo contaba con 14 secuencias que, como hemos visto antes, en su mayoría pertenecen a las subfamilias *EmiA1* y *EmiA2*, dos de las más recientes. Ahora disponemos de un mayor número de secuencias que, además, muestran diferente antigüedad dentro del genoma, lo que constituye una mejor base para realizar este tipo de análisis.

Para conocer el contexto génico en el que se encuentran los distintos elementos identificados se han contrastado todas las secuencias individualmente en la base de datos del NCBI (GenBank) utilizando la aplicación BLAST. Se han analizado las regiones flanqueantes a cada uno de los

elementos *Emigrant* y se ha calculado la distancia que separa al MITE del ATG o codón STOP del gen más cercano. En la mayoría de los casos se trata de genes putativos o predichos. En el apéndice 1 se adjunta una tabla donde están anotados todos los MITEs *Emigrant* encontrados, con el número de acceso del BAC en el que se encuentran, así como la posición que ocupan dentro del mismo, secuencias que flanquean a cada elemento y la distancia a la que se encuentra el ORF más cercano.

El resultado de este análisis muestra que un 10% de los MITEs *Emigrant* se encuentran localizados dentro de genes, tanto en intrones (7%) como en exones (3%), el 25.5% están situados a menos de 500 nt de un ORF y el 23% a una distancia comprendida entre 500 y 1000 pb. Un 29% de los elementos distan más de 1 Kb del ORF más cercano y el 13% restante ocupan regiones repetidas del genoma.

Interesantemente, la situación de los MITEs *Emigrant* respecto a los genes varía de forma sustancial entre las diferentes subfamilias analizadas (ver tabla 2). Así como más de la mitad (55%) de los elementos *EmiB3*, o el 42% de los *Emi0*, se encuentran localizados dentro o cerca de un ORF predicho (a menos de 500 pb de distancia), tan sólo el 15% de los MITEs *EmiA2*, la subfamilia más homogénea, presentan el mismo comportamiento.

Tabla 2. Distancia al ORF más cercano (nt)

	0 (dentro)	<500	500-1000	>1000	Repetitiva	n.d.
<i>EmiA1</i> (16)	12.5% (2)	19% (3)	19% (3)	44% (7)	6% (1)	
<i>EmiA2</i> (20)		15% (3)	20% (4)	50% (10)	15% (3)	
<i>EmiA3</i> (4)		25% (1)	50% (2)		25% (1)	
<i>EmiB1</i> (9)	12.5% (1)	12.5% (1)	12.5% (1)	38% (3)	25% (2)	(1)
<i>EmiB2</i> (7)		43% (3)	43% (3)		14% (1)	
<i>EmiB3</i> (9)		55.5% (5)	33% (3)	11% (1)		
<i>EmiC1</i> (17)	20% (3)	20% (3)	7% (1)	40% (6)	13% (2)	(2)
<i>EmiC2</i> (9)	11% (1)	22% (2)	44% (4)	22% (2)		
<i>EmiC3</i> (10)	10% (1)	20% (2)	30% (3)	30% (3)	10% (1)	
<i>Emi0</i> (47)	14% (6)	28% (12)	21% (9)	18.5% (8)	18.5% (8)	(4)
Total (151)	10% (15)	25.5% (37)	23% (33)	27.5% (40)	13% (19)	(7)

NOTA.- Entre paréntesis aparece el número de elementos *Emigrant* que aparecen en cada localización.

Repetitiva hace referencia a regiones descritas como repetidas en el genoma de *A.thaliana*.

No se determina (n.d.) la distancia al ORF más cercano de 7 *Emigrant* porque no había información de la región genómica en las bases de datos. Estas secuencias no se han tomado en cuenta al calcular los porcentajes. Los elementos e79 (grupo *EmiA* y que está a <500 nt de un ORF), e104 (*EmiB* y que está a <500 nt de un ORF) y e111 (*EmiC* y dentro de un ORF) no se han incluido en los datos de ninguna subfamilia, pero sí se han tenido en cuenta a la hora de calcular los datos globales.

Estos resultados nos indican que la selección ha tenido un efecto (quizás indirecto) sobre la posición de los elementos *Emigrant* a lo largo de la evolución. Mientras que las copias más jóvenes se sitúan “lejos” de los genes, las más antiguas aparecen asociadas a genes en mayor número, sugiriendo que los elementos *Emigrant* se insertan preferentemente distanciados de las

regiones génicas, pero que aquellos asociados a secuencias génicas se mantienen, con mayor frecuencia, a lo largo de la evolución. En la figura 6 se representa gráficamente la localización de los elementos *EmiA2* (los más recientes evolutivamente) y de los *Emi0* (los más viejos). La variación de la distribución de estos MITEs a lo largo del tiempo puede ser debida a que los elementos asociados a los genes han sido seleccionados positivamente, quizás porque estén aportándole al gen nuevas características beneficiosas. Pero también puede estar ocurriendo que, puesto que la eliminación de las copias alejadas de la secuencia codificante supone un riesgo menor que la de las copias situadas junto a los genes, se estén perdiendo los elementos situados en regiones pobres en genes como consecuencia, por ejemplo, de procesos de recombinación entre copias ectópicas.

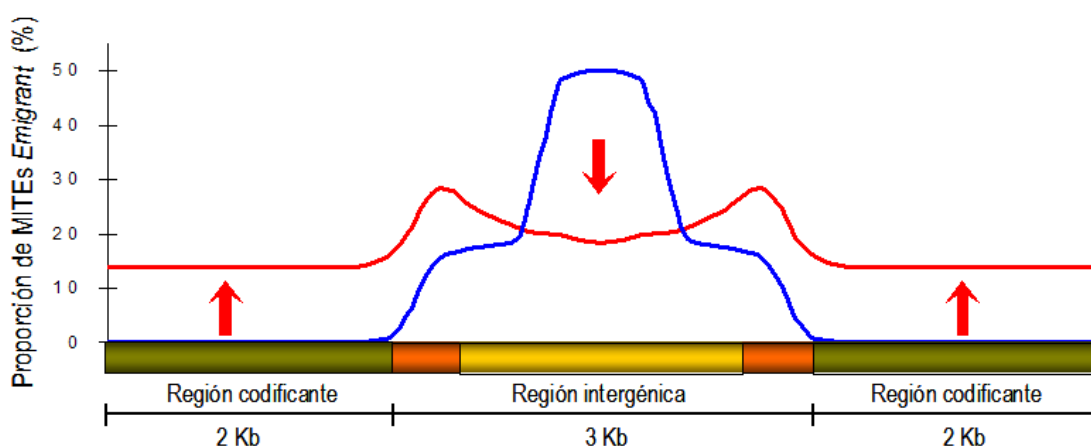


Figura 6. Los elementos *Emigrant* asociados a genes parecen haberse mantenido a lo largo de la evolución de *Arabidopsis*. La línea azul representa el porcentaje de MITEs *EmiA2*, las copias más recientes de los elementos *Emigrant*, situados en una determinada posición respecto a las regiones codificantes de *Arabidopsis*. La línea roja representa la localización de los elementos más antiguos, *Emi0*. En el eje de coordenadas se representa la distribución media de los genes en *Arabidopsis* (AGI, 2000). Las cajas naranjas corresponden a los 500 nt que flanquean a las regiones codificantes y que supuestamente coinciden con regiones reguladoras. Las flechas rojas representan el cambio de la distribución de los MITEs *Emigrant* en el tiempo.

De todos los elementos *Emigrant* insertados a menos de 500 nt de la secuencia codificante más cercana, el 46.5% lo hace en su región 3', el 27.5% en su región 5' y el 26% se localizan dentro del mismo ORF. Estos MITEs pueden ejercer un efecto sobre la actividad promotora, de procesamiento diferencial, terminación de la transcripción o estabilidad del RNA, así como sobre la capacidad codificante del gen. El hecho de que haya casi el doble de elementos en 3' del ORF, que en 5' o dentro del mismo, puede estar indicando que el efecto de la transposición sobre los genes en esta región es menos grave, pudiendo ser mantenida en el genoma al no ser tan deletérea como las otras dos situaciones. La inserción a 5' de un ORF puede alterar el patrón de

expresión del gen o cambiar la pauta de lectura del mismo, como podría ocurrir con la inserción dentro de la secuencia codificante, teniendo un efecto mucho más dramático. La presencia de un MITE a 3' del ORF podría afectar a la terminación de la transcripción, aportando en algunos casos nuevas señales de poliadenilación (seguidamente se muestra un ejemplo).

Se han analizado con más detalle todas estas inserciones cerca o dentro de ORFs, aunque solamente uno de los genes implicados, DET1, había sido estudiado más a fondo y se dispone de información sobre su expresión (Chory J *et al.*, 1989). El resto son ORFs putativos, obtenidos mediante programas de predicción de genes. En la figura 7 se muestran dos ejemplos de estas asociaciones entre MITEs *Emigrant* y genes.

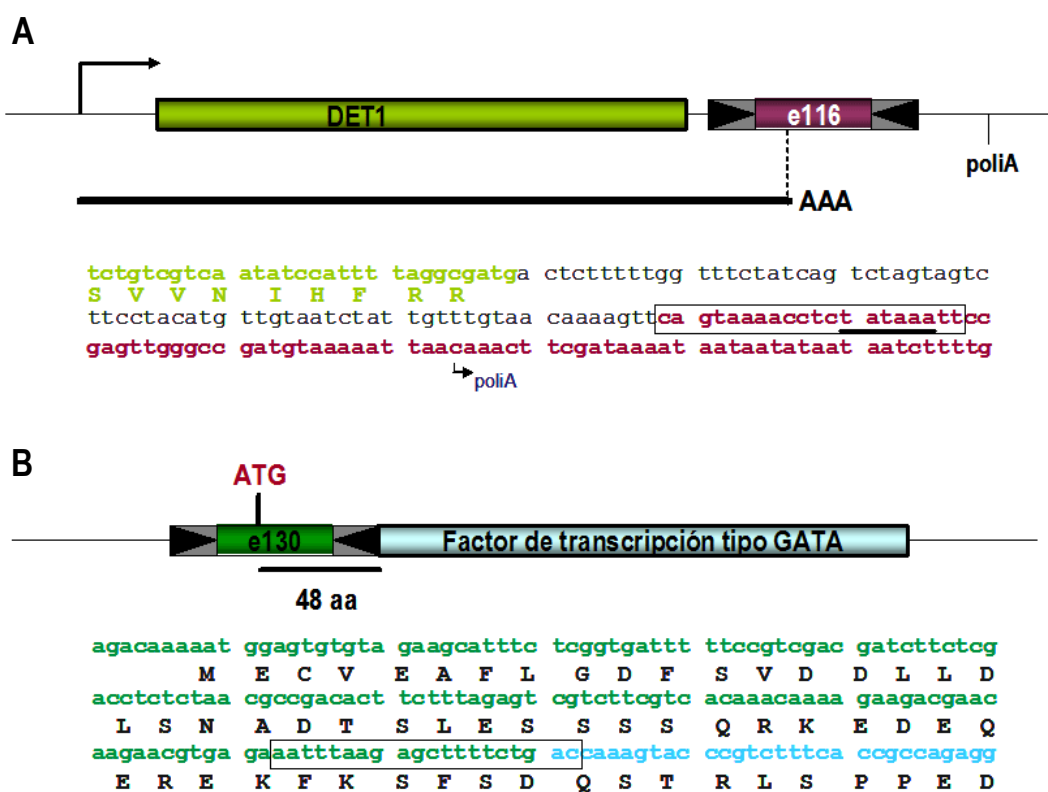


Figura 7. Elementos *Emigrant* insertados dentro de ORFs. Representación esquemática de las inserciones de *Emi116* (A) y *Emi130* (B) dentro de ORFs. Los TIRs aparecen recuadrados en las secuencias. La señal de poliadenilación aparece subrayada y el sitio poliA también se indica.

Como ejemplo de un elemento *Emigrant* situado a 3' de un ORF, en la figura 7A se representa un MITE *Emi0* (e116) localizado en la región 3' transcrita del gen DET1. Analizando la secuencia genómica y del cDNA disponibles de DET1 hemos visto que la transcripción de este gen termina dentro del MITE, probablemente respondiendo a las señales de poliadenilación contenidas en la

secuencia del e116. Hemos buscado polimorfismos para este transcrito entre diferentes ecotipos de *Arabidopsis* pero no hemos encontrado ninguna diferencia. Este MITE es un elemento *Emigrant* antiguo que parece estar fijado ya en las diferentes poblaciones de esta especie. Por otra parte, en un análisis similar sobre distintas especies del género *Brassica*, no pudimos amplificar por PCR la región 3' correspondiente de DET1. Al tratarse de géneros diferentes, aunque cercanos evolutivamente, las secuencias de la región no codificante de este gen pueden haber divergido de forma que los oligonucleótidos diseñados (a partir de la secuencia de *Arabidopsis*) no hibridan en *Brassica*.

En la figura 7B se muestra un ejemplo de otro elemento *Emi0* (e130) situado, en este caso, dentro de un gen putativo, que codificaría para un factor de transcripción del tipo GATA. La inserción del MITE podría estar aportando un nuevo ATG (codón de inicio de la traducción) y 48 nuevos aminoácidos en la región N-terminal de la proteína. Sin embargo, serían necesarios análisis funcionales para determinar el impacto de la inserción del elemento *Emigrant* sobre a expresión de este gen.

Otro ejemplo interesante lo constituyen cinco elementos *Emigrant* que aparecen situados entre dos ORFs, a menos de 500 nt de cada uno de ellos, pudiendo afectar a la expresión de ambos genes. La elevada densidad génica del genoma de *Arabidopsis*, derivada del pequeño tamaño de este último, resulta en la existencia de regiones intergénicas extremadamente cortas (AGI Project, 2000). Se ha propuesto que los MITEs podrían actuar como MARs (Matrix Attachment Regions) (Avramova Z *et al.*, 1998; Tikhonov AP *et al.*, 2000), ayudando a evitar una posible interferencia transcripcional entre dos genes próximos. Las MARs son regiones del DNA de unión a la matriz nuclear, que sirven, al mismo tiempo, de anclaje de la cadena del DNA a la matriz proteica y de regiones espaciadoras de genes, definiendo y delimitando unidades estructurales y funcionales de la cromatina. Las MARs delimitan unidades génicas transcripcionalmente independientes, aislando diferentes genes o grupos génicos entre ellos y reduciendo tanto efectos de posición como de silenciamiento génico (Avramova Z *et al.*, 1998). No parece probable que los MITEs puedan llegar a determinar, por sí solos, la estructura cromosómica, pero sí podrían ser seleccionados para aportar nuevas funciones locales como la regulación génica o de MAR, una vez presentes en una nueva localización genómica (Tikhonov AP *et al.*, 2000). Esta situación puede ser de particular interés dentro del denso genoma de *Arabidopsis*.

4.4. Evidencias de actividad en un pasado reciente.

Todo el trabajo presentado hasta el momento se ha realizado sobre el ecotipo Columbia de *Arabidopsis thaliana*, ya que es el que se ha utilizado en el proyecto de secuenciación del genoma de *Arabidopsis* y, por tanto, del que se dispone de su secuencia. Los ecotipos de *Arabidopsis* hacen referencia a poblaciones naturales de esta especie vegetal crecidas en distintas zonas geográficas, presentando diferencias fisiológicas, de forma y desarrollo, así como diferencias genéticas. El estudio de los polimorfismos existentes entre los distintos ecotipos de *Arabidopsis* nos puede dar idea de la edad evolutiva de ciertas regiones del genoma y reordenaciones génicas ocurridas en los mismos.

Teniendo en cuenta todo esto, hemos pretendido relacionar los polimorfismos de los elementos *Emigrant* anteriormente descritos (Casacuberta E *et al.*, 1998), que indicaban que estos ETs se habían movilizado en un pasado reciente, con los datos obtenidos del análisis evolutivo, explicados en el apartado 4.1 (diferencias de tamaño y variabilidad nucleotídica entre las diferentes subfamilias).

Se han realizado varias PCRs sobre catorce ecotipos de *Arabidopsis thaliana* para amplificar distintos elementos *Emigrant* y conocer si existen polimorfismos debidos a la presencia o ausencia de estos MITEs entre los diferentes genomas. Se escogieron cinco elementos pertenecientes al grupo *EmiA2*, el más homogéneo, y cinco secuencias *Emi0*, supuestamente más antiguas. Ninguna de las cinco inserciones de elementos *Emi0* examinadas muestra un polimorfismo de inserción entre los catorce ecotipos analizados (figura 8). En cambio, dos de las cinco inserciones de MITEs *EmiA2* (las correspondientes a los elementos e126 y e158) sí muestran polimorfismo entre los ecotipos aquí estudiados.

Los resultados estarían en concordancia con los obtenidos del análisis filogenético, indicando que aquellos elementos más divergentes (*Emi0*) llevan más tiempo en el genoma de *Arabidopsis*, habiéndose fijado en el mismo antes de la separación de los distintos ecotipos.

En la figura 8 se puede observar la aparición de dos bandas en la PCR del elemento e158 en el ecotipo Co-4. La banda superior corresponde con el tamaño esperado si el MITE está presente en la posición genómica estudiada y la banda inferior, de menor intensidad, corresponde a la

ausencia del *Emigrant*. Este resultado indica que existen polimorfismos de esta secuencia dentro de un mismo ecotipo. Entonces, se decidió crecer y analizar diferentes plantas del ecotipo Co-4.

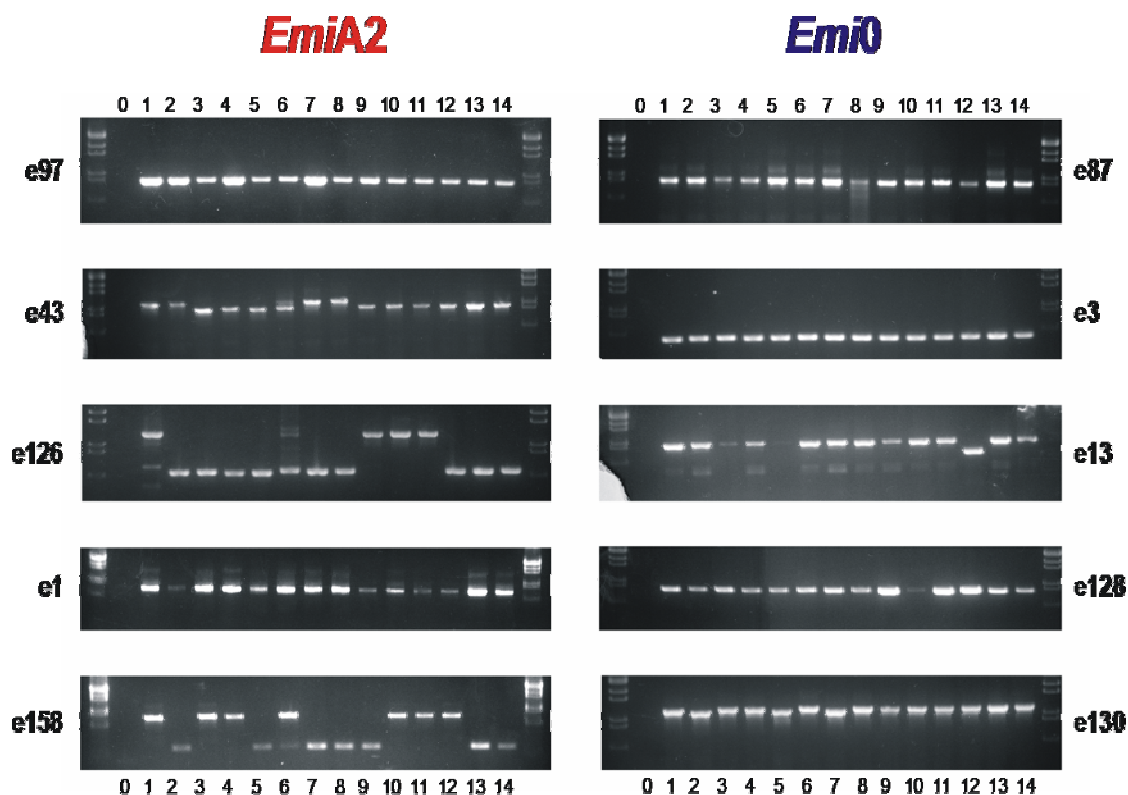


Figura 8. Análisis del polimorfismo de elementos *EmiA2* y *Emi0* entre distintos ecotipos de *Arabidopsis thaliana*. 0.Control de PCR (no DNA); 1.Columbia; 2.Landsberg; 3.RLD; 4.WS; 5.Cnt-1; 6.Co-4; 7.Dijon d; 8.Est-0; 9.Ge-0; 10.Kas-0; 11.Ms-0; 12.Nd-1; 13.Tsu-0; 14.No-0.

La PCR realizada sobre el DNA genómico de plantas Co-4 muestra claramente la existencia de polimorfismos debidos a la presencia o ausencia del elemento e158 entre distintos individuos de este ecotipo y, más aún, muestra la existencia de alelos diferentes dentro del genoma de un mismo individuo (figura 9). Para estudiar estas PCRs más detalladamente, se clonaron y secuenciaron estas bandas.

El análisis de las secuencias obtenidas a partir de las PCRs realizadas sobre los distintos ecotipos nos ha aportado datos que parecen indicar que ha existido, o existe todavía, una movilización de estas secuencias en el genoma de *Arabidopsis*. Y que esta movilización parece afectar a los elementos pertenecientes al grupo *EmiA2*, los más homogéneos en secuencia y tamaño. De esta forma parece corroborarse que los diferentes grupos de elementos *Emigrant*

han sufrido diferentes amplificaciones en el tiempo y que aquellos más homogéneos son aquellos que poseen una historia evolutiva más reciente.

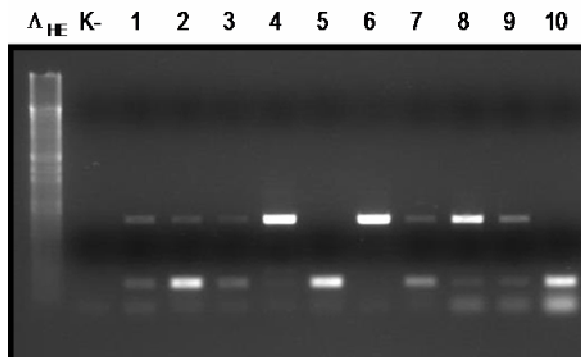


Figura 9. Análisis por PCR del polimorfismo del elemento e158 entre distintos individuos del ecotipo Co-4 de *Arabidopsis*. K-: control de la PCR.

Hasta hace poco tiempo se tenía poca información acerca del mecanismo de amplificación de los MITEs, aunque en función a la homología que presentaban en la secuencia de sus TIRs y de la secuencia de las TSDs (duplicaciones de los sitios de inserción) con familias de transposones de DNA se pensaba que los MITEs podían ser movilizados por transposasas codificadas por elementos transponibles autónomos de clase II presentes en el mismo genoma. Hace cinco años se publicó el primer trabajo que apuntaba en esta dirección, donde se identificaba un elemento transponible, relacionado con la familia *Emigrant* de MITEs, que contiene un ORF con homología a transposasas de la superfamilia *Tc1/mariner* (Feschotte C and Mouchès C, 2000). Trabajos posteriores han identificado diferentes ETs autónomos relacionados con familias de MITEs en genomas como el de maíz, arroz o trigo (Zhang X *et al.*, 2001; Feschotte C *et al.*, 2005; Saito M *et al.*, 2005) y tres trabajos coetáneos han descrito la movilización de la familia *mPing* de MITEs en arroz y han propuesto que son movilizados por transposasas codificadas en elementos autónomos de la misma familia o relacionados (Kikuchi K *et al.*, 2003; Nakazaki T *et al.*, 2003; Jiang N *et al.*, 2003). Todos estos resultados juntos muestran que los MITEs pueden ser escindidos de su posición original mediante mecanismos activos de transposición, cuyo primer paso es la unión de la transposasa a la secuencia de los TIRs y de las repeticiones subterminales de estos ETs (Feschotte C *et al.*, 2005 y datos obtenidos en nuestro laboratorio no publicados todavía). Además, como ocurre con otros transposones de clase II, la escisión de estos elementos suele dejar huellas o *footprints* de escisión en la secuencia del DNA (Kikuchi K

et al., 2003; Nakazaki T *et al.*, 2003; Jiang N *et al.*, 2003), debido a que la transposasa puede efectuar cortes que no son limpios y/o a la subsiguiente reparación de la cadena de DNA (Plasterk RH, 1991; Scott L *et al.*, 1996).

Las secuencias analizadas procedentes de la PCR del elemento e158 sobre distintos individuos del ecotipo Co-4 muestran que las bandas inferiores corresponden a deleciones del MITE respecto a su posición original (banda superior), ya que aparecen *footprints* o huellas de escisión, características de la transposición (ver figura 10). Este resultado nos está indicando que el polimorfismo de la presencia del *Emigrant* e158 observado entre distintas plantas del ecotipo Co-4 de Arabidopsis parece no deberse a una nueva inserción del transposón en la localización del genoma analizada, sino a un proceso de eliminación del elemento de esta posición. En cualquier caso se trata de una copia que todavía no había sido fijada en la población, como indica la variación existente tanto entre alelos de un mismo individuo, o entre individuos de un mismo ecotipo, como entre individuos de ecotipos diferentes de Arabidopsis.



Figura 10. Representación del alineamiento de las secuencias amplificadas en la PCR del elemento e158 sobre un individuo heterocigoto del ecotipo Co-4 de Arabidopsis. Sobre el alineamiento se esquematiza la estructura del MITE *Emigrant* e158. Las cajas rojas muestran las diferencias nucleotídicas en la región flanqueante al elemento, que definen el *footprint* o huella dejado tras la escisión del transposón.

También hemos podido observar que algunas de las secuencias amplificadas en determinados ecotipos varían ligeramente en tamaño respecto a la longitud esperada y observada en los otros. Éste es el caso del elemento e43. Como ejemplo, la banda obtenida de la PCR realizada sobre el ecotipo RLD es algo más pequeña que la obtenida en Columbia. Para conocer a qué es debida esta diferencia se clonaron y secuenciaron ambas bandas. Los resultados de la secuenciación mostraron que se ha producido una deleción dentro del elemento e43 del ecotipo RLD, en concreto se trata de una deleción de 202 nt en la región 3' del elemento, entre los cuales encontramos una repetición subterminal del TIR en 3'. Como se ha comentado antes, la desaparición de las repeticiones subterminales puede resultar en que la transposasa no se

pueda unir eficientemente a la secuencia del MITE y se vea imposibilitada la movilización del elemento. Aunque no se ha secuenciado, suponemos que el elemento e13 del ecotipo Nd-1 también presenta una delección en su secuencia, puesto que la diferencia de tamaño respecto al resto de las bandas amplificadas no se corresponde con la ausencia de todo el MITE.

En conjunto, el trabajo presentado en este capítulo ofrece una visión del dinamismo de las secuencias móviles dentro del genoma, del constante equilibrio entre la transposición y la eliminación o inactivación de las mismas. En concreto, la familia *Emigrant* de MITEs parece haber llevado a cabo diferentes *explosiones* de amplificación a lo largo de la evolución de *Arabidopsis*. El último de estos sucesos de transposición parece ser muy reciente. Además, la aparición de *footprints* de escisión parece indicar que hay una eliminación de los elementos *Emigrant* del genoma de *Arabidopsis*, lo que puede ser un mecanismo de defensa ante posibles nuevos eventos de transposición. La delección de ciertas secuencias de los MITEs, como sus TIRs o repeticiones subterminales, también ayuda a inactivar estas copias e impedir así su movilización.

CAPÍTULO 2. ESTUDIO DEL CONTROL DE LA EXPRESIÓN DEL RETROTRANSPÓSÓN Tnt1 DE TABACO.

INTRODUCCIÓN.

En este capítulo se describen las aproximaciones seguidas para llegar a determinar cómo está controlada la expresión del retrotransposón Tnt1 de tabaco y estudiar los mecanismos de silenciamiento génico que operan sobre este tipo de secuencias. Con este trabajo pretendemos conocer algo más acerca de las herramientas de las que disponen los genomas para controlar la proliferación de elementos transponibles y otras secuencias exógenas, como pueden ser transgenes o ácidos nucleicos virales.

Tnt1 fue el primer retrotransposón activo descrito en plantas, identificado como resultado de su inserción dentro del gen de la *Nitrato reductasa* de tabaco (Grandbastien MA *et al.*, 1989). Actualmente, este transposón está bien caracterizado y constituye uno de los pocos ejemplos de elementos transponibles vegetales transcripcionalmente activos (Melayah *et al.*, 2001). Pero su transcripción, y posterior transposición, sólo ha sido detectada en determinadas condiciones de estrés (Casacuberta JM and Grandbastien MA, 1993; Pouteau S *et al.*, 1991; Pouteau S *et al.*, 1994; Mhiri C *et al.*, 1997), como ocurre, generalmente, con los transposones de clase I activos en plantas descritos hasta el momento (Hirochika H, 1993; Wessler SR, 1996; Grandbastien MA, 1998; Kumar A and Bennetzen JL, 1999; Takeda S *et al.*, 2001).

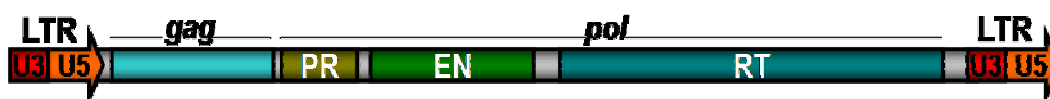


Figura 11. Representación esquemática del retrotransposón Tnt1. Las LTRs flanquean el elemento en ambos extremos y se dividen en dos grandes regiones: U3 y U5. La región codificante comprende los genes *gag* y *pol*. El gen *pol* codifica las actividades enzimáticas proteasa (PR), integrasa o endonucleasa (EN) y retrotranscriptasa (RT).

Tnt1 es un retrotransposón del tipo Ty1-copia (esquema en figura 11) que está presente en cientos de copias en el genoma de *Nicotiana tabacum* y otras especies del género *Nicotiana* (Grandbastien MA *et al.*, 1989; Vernhettes S *et al.*, 1997). La actividad promotora de estos elementos reside, esencialmente, en sus repeticiones directas terminales o LTRs (Vernhettes S

et al., 1997; Kummar A and Bennetzen JL, 1999), cuya estructura y función se ha estudiado mediante la transformación de plantas con fusiones transcripcionales del tipo LTR-GUS. Gracias a experimentos de expresión estable y transitoria de estas construcciones génicas, que contenían fragmentos de la LTR de diferente longitud, se ha podido determinar que los principales elementos que regulan la expresión de Tnt1 *in cis* se localizan en la región U3 de la LTR 5' (Casacuberta JM and Grandbastien MA, 1993). Estos elementos están contenidos, básicamente, en una corta secuencia palindrómica y en una serie de repeticiones en tándem de 31 pb, denominadas caja BI y cajas BII, respectivamente, y que presentan similitud de secuencia con otros elementos contenidos en promotores de genes de defensa de plantas (Vernhettes S *et al.*, 1997; Casacuberta JM and Grandbastien MA, 1993). En concreto, la caja BII interacciona *in vivo* y de forma específica con proteínas inducidas por estrés asociado a defensa (Vernhettes S *et al.*, 1997).

El análisis de la expresión de los elementos de Tnt1 de tabaco mostró que las copias de este retrotransposón existentes en el genoma presentan cierta variabilidad de secuencia, de manera que se pueden definir tres subfamilias (Tnt1A, Tnt1B y Tnt1C) en función de las regiones U3 reguladoras que presentan (Vernhettes S *et al.*, 1998). Las tres subfamilias se expresan en situaciones de estrés asociado a defensa (Casacuberta JM *et al.*, 1997; Beguiristain T *et al.*, 2001) y, aunque conservan homología de secuencia en el resto de la LTR y región codificante, la diferencia en sus regiones promotoras U3 es suficiente para que presenten patrones de expresión distintos (Beguiristain T *et al.*, 2001). Tnt1A se transcribe en raíz y se induce fuertemente en hojas heridas o tratadas con metil jasmonato o con el elicitor fúngico criptogeína (Pouteau S *et al.*, 1994; Moreau-Mhiri C *et al.*, 1996; Vernhettes S *et al.*, 1997; Beguiristain T *et al.*, 2001). Tnt1B se expresa casi exclusivamente en células en cultivo y Tnt1C lo hace, principalmente, en hoja inducida por ácido salicílico o por la auxina 2,4-D (Beguiristain T *et al.*, 2001).

Gracias a sus características, Tnt1 es un buen modelo para estudiar el control de la transcripción de los retrotransposones. La retrotransposición en plantas parece estar controlada principalmente a nivel transcripcional (Grandbastien MA, 1998; Casacuberta JM and Santiago N, 2003). Hasta ahora, los estudios del control de la transcripción de Tnt1 se han centrado en el análisis del inicio de la transcripción, caracterizando las secuencias promotoras y activadoras contenidas en las LTRs (Casacuberta JM and Grandbastien MA, 1993; Vernhettes S *et al.*, 1997). Sin embargo, el control epigenético, que afecta a la conformación de la cromatina, o el control de

la terminación de la transcripción también constituyen otros niveles a los que puede estar regulada la transcripción. En los últimos años estos procesos han sido estudiados en mayor grado, conociéndose cada vez más su funcionamiento y mostrando la gran importancia que tienen en la regulación de la expresión génica. El trabajo que se presenta en esta memoria aborda estos dos procesos y discute el papel que puedan tener en el control de la expresión de los retrotransposones.

1. ESTUDIO DE LA TERMINACIÓN DE LA TRANSCRIPCIÓN.

La terminación de la transcripción supone un punto crítico para la correcta expresión de los genes (Magnotta SM and Gogarten JP, 2002; Phillips C *et al.*, 2001; Wahle E and Rügsegger U, 1999). En las células eucariotas los RNAs transcritos sufren una serie de modificaciones post-transcripcionales (splicing, o eliminación de intrones, y la adición de la caperuza en el extremo 5' de la molécula y de una cola poliA en su extremo 3') esenciales para la estabilidad, transporte al citoplasma y correcta traducción de la molécula de RNA. Estas modificaciones parecen estar relacionadas entre sí y pueden tener lugar simultáneamente (Ares M and Proudfoot NJ, 2005; Minvielle-Sebastia L and Keller W, 1999). Normalmente la poliadenilación se produce tras la eliminación del extremo 3' del transcrito por una nucleasa que corta a unos veinte nucleótidos a 3' de la señal de poliadenilación contenida en la región 3' no codificante (3' UTR) de la molécula de RNA. En animales esta señal de poliadenilación está bastante conservada, correspondiendo a la secuencia AAUAAA en la mayoría de los casos (aunque también se han descrito otras), y parece ser suficiente para dirigir la terminación y poliadenilación del RNA (Wahle E and Rügsegger U, 1999).

En plantas este proceso está menos estudiado, pero también parece estar controlado por secuencias o regiones, menos definidas, presentes en la 3' UTR y que se localizan anterior y posteriormente al sitio de corte y poliadenilación (Rothnie HM, 1996; Loke JC *et al.*, 2005). A estos elementos se les ha denominado FUE (Far-Upstream Element), NUE (Near-Upstream Element), sitio de corte (CS), elemento de corte (CE) y elementos en 3'. El elemento FUE se localiza de la posición -25 a -160 del sitio de corte, es una secuencia rica en uracilos y su longitud varía entre 60 y 100 nt, comprendiendo combinaciones de motivos de 6 a 18 nt. El NUE corresponde con la señal de poliadenilación presente en animales, pero la secuencia AAUAAA no está tan conservada, aunque siempre es rico en adenosinas, y se localiza entre las

posiciones -13 y -30 respecto al CS. El sitio de corte suele ser un dinucleótido cuya secuencia consenso es YA (donde Y representa a cualquiera de las dos pirimidinas, C ó U). El elemento de corte o CE es una pequeña región situada a ambos lados de CS y está formado por un conjunto de pequeños elementos en cis, ricos en U. Los elementos situados a 3' del sitio de corte son más variables en plantas que en animales, pero se ha descrito que su delección afecta a los niveles de expresión de algunos genes vegetales (Loke JC *et al.*, 2005).

En animales, se ha visto que todos estos elementos forman estructuras secundarias en la región 3' no traducida de la molécula del RNA, necesarias para la correcta terminación de la transcripción y poliadenilación. Estos dos procesos parecen estar íntimamente relacionados y regulados de forma conjunta (Yonaha M and Proudfoot N, 2000; Proudfoot N, 2004).

Existen genes con múltiples de sitios de poliadenilación, cuya transcripción resulta en la producción de moléculas de RNA de diferente tamaño, características y eficiencia de traducción. Aunque codifiquen para la misma proteína difieren en su región 3' no traducida, donde pueden localizarse algunos elementos reguladores. La presencia o ausencia de estas secuencias puede suponer un mecanismo por el cual estos genes están regulados post-transcripcionalmente. En plantas se supone que este mecanismo puede actuar de forma similar.

La terminación de la transcripción podría constituir una etapa de regulación de la transposición. Además de influir en la estabilidad y traducción del mRNA, en el caso de los retrotransposones, una terminación prematura podría dificultar el uso del transcrito como molde en el proceso de retrotranscripción y bloquear, así, la transposición.

Se ha visto, por ejemplo, que la existencia de diversas señales de poliadenilación en la secuencia de un retrotransposón del tipo LINE-1 humano parece constituir un mecanismo de control que limita la actividad transposicional del elemento (Perepelitsa-Belancio V and Deininger P, 2003).

1.1. ANÁLISIS DE LOS SITIOS DE POLIADENILACIÓN DE Tnt1.

En el estudio del control de la terminación de la transcripción de Tnt1, analizamos, primero, el lugar en el que se incorpora la cola poliA en los RNAs resultantes de la expresión de este retrotransposón. Para ello indujimos la expresión de Tnt1 mediante incubación de hojas de tabaco heridas con la celulasa R-10 de Onozuka (ver apartado 2.1.1, donde se explica más detalladamente). Posteriormente extrajimos el RNA total de estas hojas y amplificamos mediante

Resultados y discusión. Capítulo 2.

RT-PCR los transcritos de Tnt1 generados (figura 12). La reacción de la PCR se realizó utilizando los oligonucleótidos OligodT y Avi, que hibrida con la región terminal de la RT de Tnt1 (ver material y métodos). Posteriormente se clonaron y secuenciaron los productos de la PCR.

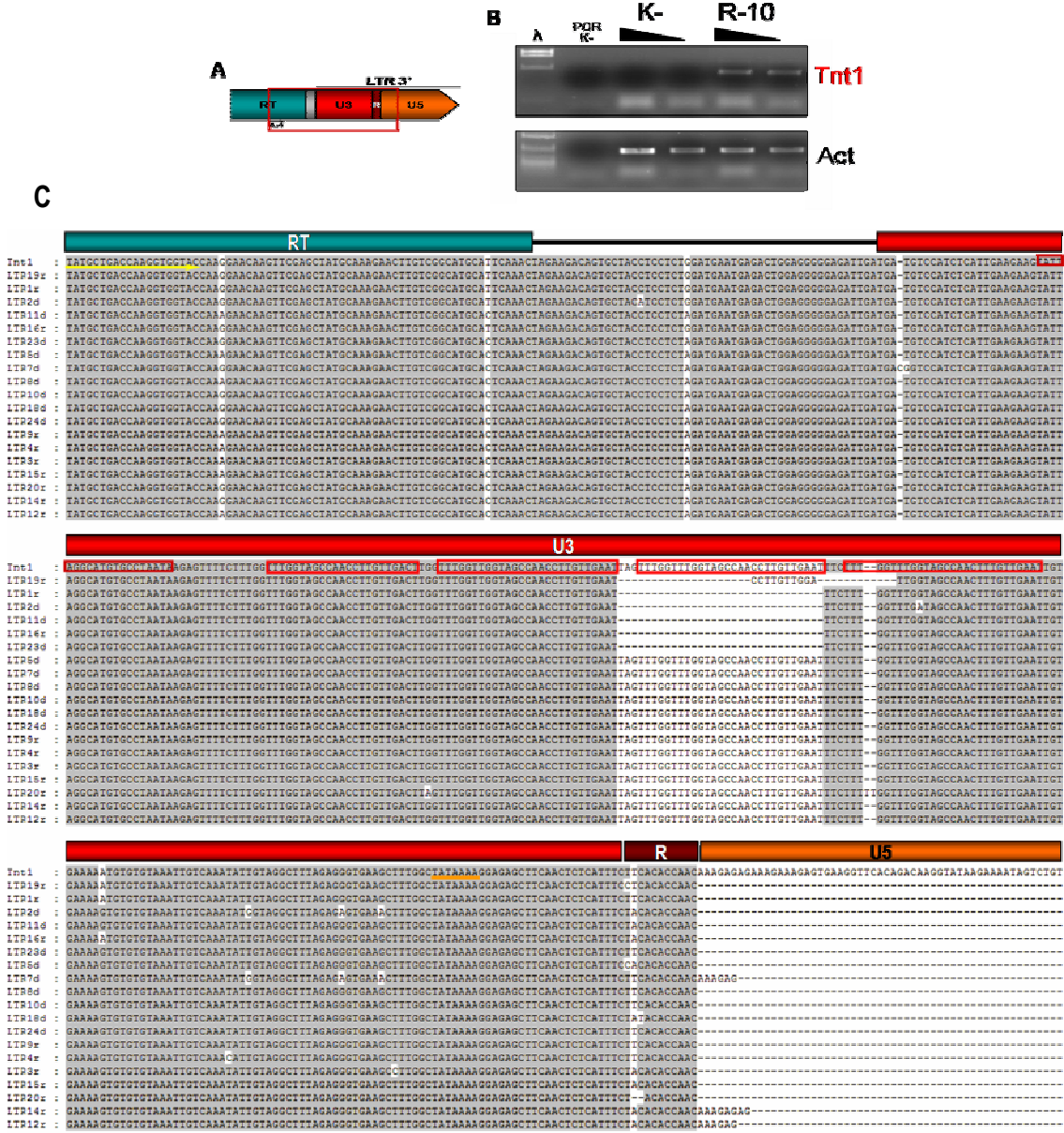


Figura 12. Análisis del sitio de poliadenilación de Tnt1. (A) Esquema representativo de la región 3' terminal de la secuencia del retrotransposón Tnt1, que incluye el extremo 3' de la RT y la LTR 3'. La LTR se divide en tres regiones: U3, R y U5. Se indica el lugar de hibridación del oligo Avi y la región representada en el alineamiento. (B) Electroforesis de los productos de la RT-PCR sobre Tnt1 y actina (control). PCR K-: control de la PCR. K-: hoja control. R-10: hoja estresada. (C) Representación del alineamiento de los productos de la PCR clonados y secuenciados. Tnt1: secuencia consenso de Tnt1A (Casacuberta JM *et al.*, 1995). Sobre los alineamientos se representa las regiones en las que se divide la secuencia de Tnt1. La secuencia del oligo Avi aparece subrayada en amarillo. Las cajas BI y BII aparecen recuadradas en granate y rojo, respectivamente. La señal de poliadenilación principal aparece subrayada en naranja.

En total se secuenciaron 19 fragmentos de DNA. Todas las secuencias corresponden a transcritos de elementos pertenecientes a la subfamilia Tnt1A. Seis de estas secuencias presentaban la delección parcial o total de la tercera caja BII, respondiendo a la variabilidad existente entre las distintas copias de Tnt1 (Casacuberta JM *et al.*, 1995).

Dieciséis de estas secuencias terminan o se poliadenilan en la misma posición, que coincide con el sitio típico de poliadenilación descrito anteriormente para los mRNAs de Tnt1, expresados en protoplastos de hoja de tabaco (Casacuberta JM *et al.*, 1995) y tras una inducción con criptogeína (Beguiristain T and Casacuberta JM, no publicado). Las otras tres secuencias obtenidas en este análisis incorporan la cola poliA seis u ocho nucleótidos después.

Estos resultados indican que en las condiciones que hemos utilizado para inducir la expresión del retrotransposón Tnt1, se expresan mayoritariamente elementos pertenecientes a la subfamilia Tnt1A y que los transcritos generados terminan y se poliadenilan en un único sitio, indicando que el terminador de la transcripción está funcionando de forma eficiente.

1.2. DIFERENCIAS EN LA POLIADENILACIÓN ENTRE LAS DIFERENTES FAMILIAS DE Tnt1.

Los determinantes principales de terminación de la transcripción suelen situarse a 5' del sitio de poliadenilación. En el caso de Tnt1, estarían localizados en las regiones U3 y R (secuencia puente comprendida entre el sitio de inicio de la transcripción y el de poliadenilación y que separa las regiones U3 y U5) de la LTR. Puesto que la secuencia de la región U3 varía entre las distintas subfamilias de Tnt1, para empezar a analizar las señales que dirigen la poliadenilación del transcrito de Tnt1, hemos colocado la secuencia U3 de la LTR de cada una de las tres subfamilias controlando la terminación del gen marcador GUS. Tras la región U3 hemos añadido la secuencia del terminador NOS, que permite asegurar que los transcritos que no reconozcan las señales de Tnt1 terminen correctamente. El promotor constitutivo 35S del CaMV dirige la expresión del este gen. Como control se ha utilizado el plásmido pBI221, en el que el terminador NOS se sitúa directamente a continuación de la región codificante del gen GUS. En la figura 13 se esquematizan las construcciones génicas utilizadas en este experimento.

La transcripción de estas construcciones se ha analizado *in vivo* mediante su bombardeo sobre discos de hoja de tabaco, incubados seguidamente en medio MS durante 6 horas para que se

distintas especies de transcritos, en función del lugar en el que incorporan la cola poliA, para cada una de las construcciones utilizadas.

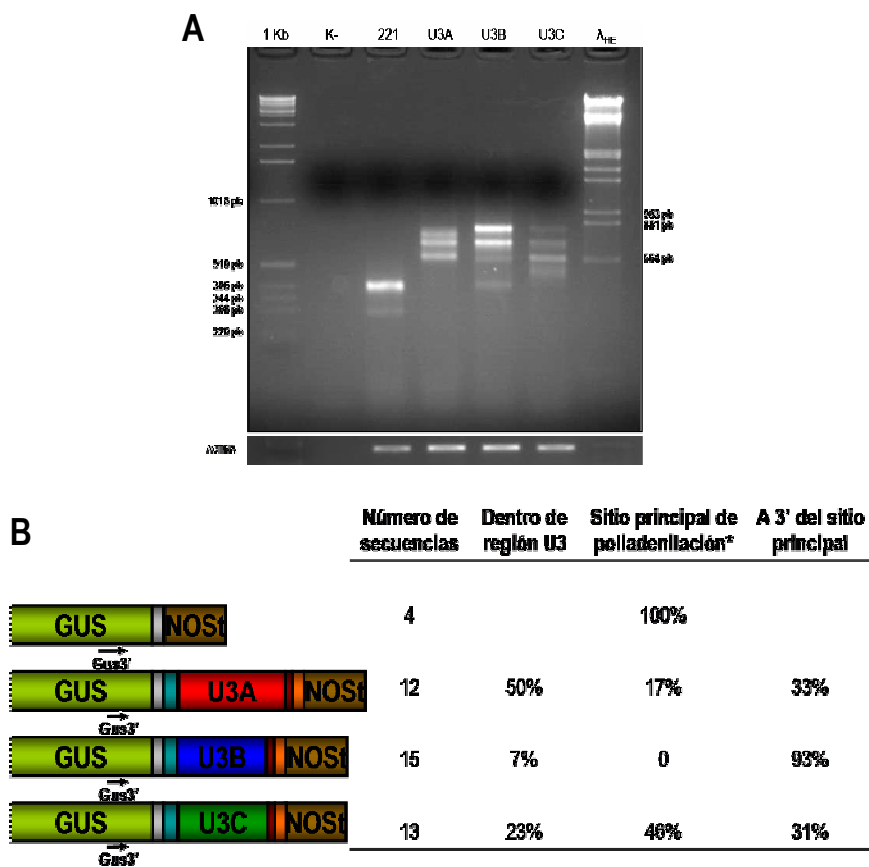


Figura 14. (A) Electroforesis del producto de las PCRs realizadas sobre el mRNA transcrito a partir de las construcciones conteniendo las diferentes regiones U3 de la LTR como terminador. (B) Tabla con las proporciones de las diferentes especies de transcritos generados, definidas en función del sitio de poliadenilación, para cada una de las construcciones analizadas. *: El sitio principal de poliadenilación de la primera construcción (pBI221) es el del terminador NOS.

Todos los amplicones derivados de la expresión del plásmido pBI221 terminan en un único sitio dentro del terminador NOS (Depicker A *et al.*, 1982), mostrando que en las condiciones ensayadas este terminador funciona correctamente.

De los 12 fragmentos, correspondientes a transcritos de la construcción GU3A, secuenciados sólo 2 han incorporado la cola poliA en el sitio principal de poliadenilación. Consideramos como sitio principal de poliadenilación el descrito anteriormente para los transcritos de Tnt1 expresados en protoplastos de hoja de tabaco o en hojas tratadas con criptogeína o R-10 (ver apartado anterior). Seis secuencias terminan antes de esta posición, en distintos lugares dentro de la

región U3A, y las otras 4 secuencias han continuado su transcripción más allá del sitio principal de poliadenilación.

Ninguno de los 15 transcritos expresados a partir de la construcción GU3B, amplificados y clonados, parece responder a la señal principal de poliadenilación. Uno de ellos termina mucho antes de encontrar esta señal, dentro de la región U3B, y los 14 restantes prolongan su transcripción en 3' del sitio principal de poliadenilación hasta responder, las cuatro secuencias más largas, a la señal de poliadenilación contenida en el terminador NOS.

Por otra parte, sólo la mitad de los transcritos resultantes de la expresión de la construcción génica GU3C (seis de 13 secuenciados) se poliadenilan en el sitio principal. Tres de las secuencias lo hacen dentro de la región U3C y las otras cuatro acaban en diferentes posiciones dentro de la secuencia del terminador NOS, a 3' del sitio principal de poliadenilación.

Los resultados obtenidos de este análisis sugieren que el terminador del gen de la Nopalín sintasa es un terminador fuerte en las condiciones ensayadas, mientras que las señales de poliadenilación contenidas en las distintas regiones U3 son menos eficientes. Además las diferentes U3 presentan distinta eficiencia a la hora de promover la incorporación de la cola poliA en la posición correspondiente. En el caso de la región U3A los resultados muestran que la terminación de los transcritos del transgén GU3A es menos eficiente que la de los elementos Tnt1A expresados bajo inducción con R-10.

1.3. EFECTO DEL ESTRÉS SOBRE LA POLIADENILACIÓN DE Tnt1.

Mediante el sistema utilizado en el este análisis se obtiene una expresión constitutiva del transgén, mientras que el retrotransposón Tnt1 sólo se transcribe en condiciones de estrés. Trabajos anteriores sobre la expresión de Tnt1 muestran que, además de la inducción preferencial de una subfamilia respecto a las otras, dependiendo del tipo de estrés que se esté aplicando, puede variar la eficiencia de la terminación y de la poliadenilación de los transcritos. El análisis de las secuencias de Tnt1 expresadas muestra que la cola poliA no se incorpora siempre en el lugar indicado por la señal de poliadenilación "principal", si no que aparecen otros sitios de poliadenilación "atípicos". El tipo de estrés parece influir incluso en la terminación de los transcritos de una misma subfamilia, la subfamilia Tnt1A (Beguiristain T and Casacuberta JM, no publicado). En este trabajo hemos querido comenzar a estudiar este fenómeno. Para ello, decidimos analizar el efecto del estrés sobre la terminación de los transcritos originados a partir

de la expresión de la construcción que contiene la región U3 de los elementos Tnt1A como terminador (GU3A, ver apartado anterior).

Bombardeamos discos de hoja de plantas de tabaco silvestres (WT) con la construcción GU3A y se infiltraron (por vacío) e incubaron durante 6 horas en medio MS conteniendo distintos elicitores. Los elicitores utilizados para este experimento son la criptogeína, elicitorina producida por el hongo *Phytophthora cryptogea* y que origina una respuesta de estrés ante infección (Ricci P *et al.*, 1989), MeJa, metil-jasmonato, molécula intermediaria en las vías de señalización de la respuesta de defensa en plantas, componente de la cascada de señalización desencadenada por la criptogeína (Rusterucci C *et al.*, 1999), 2,4-D, 2,4-ácido diclorofenoxiacético, auxina responsable de la inducción transitoria de Tnt1 en cultivo celular de tabaco (Beguiristain T *et al.*, 2001) y el ácido salicílico, también molécula intermediaria en vías de señalización de la respuesta de defensa ante herida e infección (Klessig DF and Malamy J, 1994). Posteriormente se extrajo el RNA, sobre el que se realizó la RT-PCR utilizando los oligos GUS3' y OligodT (ver material y métodos).

Los resultados de la RT-PCR se representan en la figura 15A y muestran la existencia de distintas bandas, resultantes de una poliadenilación diferencial del RNA. En todos los tipos de estrés aplicados encontramos un patrón de bandas similar, aunque la abundancia relativa de las distintas bandas varía dependiendo del elicitor utilizado.

Se han clonado y secuenciado los productos de las PCRs realizadas sobre el material tratado con criptogeína y ácido salicílico (figura 15B y 15C). De los transcritos obtenidos a partir de la expresión de la construcción GU3A en presencia de criptogeína, casi la mitad de ellos incorporan la cola poliA en el sitio principal de poliadenilación, que se ha definido antes. Tres de las nueve secuencias terminan antes de este sitio y dos lo hacen posteriormente. En cambio, la mayor parte de los transcritos de GU3A en presencia de ácido salicílico continúan su transcripción más allá del sitio principal de poliadenilación. Sólo dos de los 17 amplicones secuenciados incorporan la cola poliA en el sitio principal.

El perfil de las muestras tratadas con MeJa se asemeja al de las tratadas con criptogeína y el de las tratadas con 2,4-D al de las tratadas con ácido salicílico, mientras que en ausencia de elicitores (infiltración con MS solo) se observa una abundancia similar de todas las especies de mRNAs generados.

Los transcritos resultantes de la infiltración con criptogeína y MeJa parecen responder mejor a las secuencias terminadoras presentes en la región U3. Por el contrario, cuando se utiliza ácido salicílico o 2,4-D la transcripción parece continuar y acabar más allá de la región R.

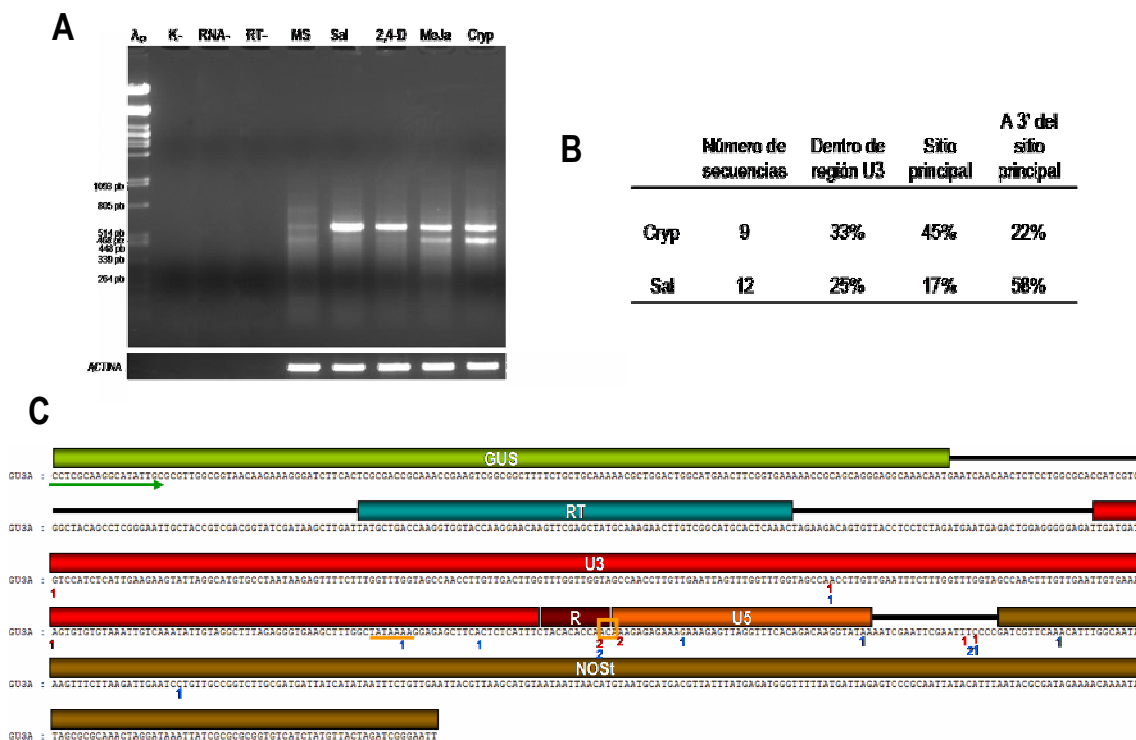


Figura 15. (A) Electroforesis del producto de las RT-PCRs sobre los transcritos de la construcción GU3A. PCR realizada con los oligos GUS3' y OligodT. λp : DNA del fago λ digerido con *Pst*I. K-: control de la PCR. RNA-: control de la retrotranscripción, reacción sin RNA. RT-: control de la retrotranscripción, reacción sin RT. INF-: hoja no infiltrada. Sal: ácido salicílico. Cryp: criptogeína. (B) Tabla con las proporciones de las diferentes especies de transcritos generados, definidas en función del sitio de poliadenilación, para los tratamientos con criptogeína y ácido salicílico. (C) Representación de la secuencia de la región terminal de la construcción GU3A. Se indica la posición donde se ha incorporado la cola poliA y el número de transcritos, para criptogeína (rojo) y ácido salicílico (azul). La flecha verde indica la posición del oligo GUS3'. La señal de poliadenilación y el sitio principal se indican en naranja, subrayada y encuadrado, respectivamente.

La presencia de distintos sitios de poliadenilación y eficiencias de terminación diferentes dependiendo del elicitador utilizado, sugiere que podrían existir elementos reguladores de la terminación y poliadenilación en la región U3 de las LTRs que responden de manera diferente ante ciertos factores. Es importante resaltar que hemos utilizado la región U3 de los elementos Tnt1 pertenecientes a la familia A, que constituyen la principal población de secuencias transcritas tras la inducción con criptogeína o MeJa (ver en introducción) y su terminador podría estar particularmente adaptado a estas condiciones.

De todas formas, encontramos una gran proporción de secuencias que no terminan en la región U3, por lo que no parece que la región U3 sea un terminador muy eficiente en este sistema. Esto podría ser debido a la presencia de elementos responsables para la terminación fuera de la región U3. De hecho, en sistemas animales existen regiones situadas a 3' del sitio de poliadenilación que contribuyen a que se dé este proceso (Zarudnaya MI *et al.*, 2003). En plantas se ha descrito la existencia de elementos similares en 3', que podrían estar ejerciendo la misma función (Loke JC *et al.*, 2005). Por ello, podría ser que en la región U5 se localizaran algunos elementos importantes para la terminación. Entonces, sería interesante analizar su efecto utilizando construcciones similares a las que hemos empleado en nuestro estudio, pero conteniendo la región U5 de la LTR.

1.4. ESTUDIO DE LA INTERACCIÓN ENTRE EL INICIO Y LA TERMINACIÓN DE LA TRANSCRIPCIÓN.

Sin embargo, en los últimos años se ha demostrado que, en muchos casos, factores implicados en el inicio de la transcripción también afectan a la terminación, poniendo de relieve una relación promotor-terminador (Calvo O and Manley JL, 2001).

En todos los ensayos presentados en los apartados anteriores hemos utilizado el promotor 35S, que es distinto del promotor que dirige la transcripción de Tnt1 *in vivo* (que se encuentra en la región U3 de la LTR).

Para comprobar si la presencia de la región U3 dirigiendo la expresión del gen afecta a la eficiencia de la terminación controlada por esta misma secuencia, decidimos comparar la transcripción del gen marcador GUS flanqueado en ambos extremos por la U3 de Tnt1A con la del transgén utilizado anteriormente, en el que la U3 sólo se encontraba a 3' de GUS (ver esquema de los plásmidos utilizados en la figura 16A).

Los diferentes plásmidos se bombardearon sobre discos de hoja distintos, que se infiltraron con MS o MS + criptogeína, para inducir la expresión de los transgenes que contienen la U3 a 5'. Posteriormente se analizó la terminación de los transcritos generados de la misma forma que se ha descrito en los apartados anteriores, amplificando la región 3' de los RNAs, utilizando los oligos GUS3' y OlidodT (ver material y métodos).

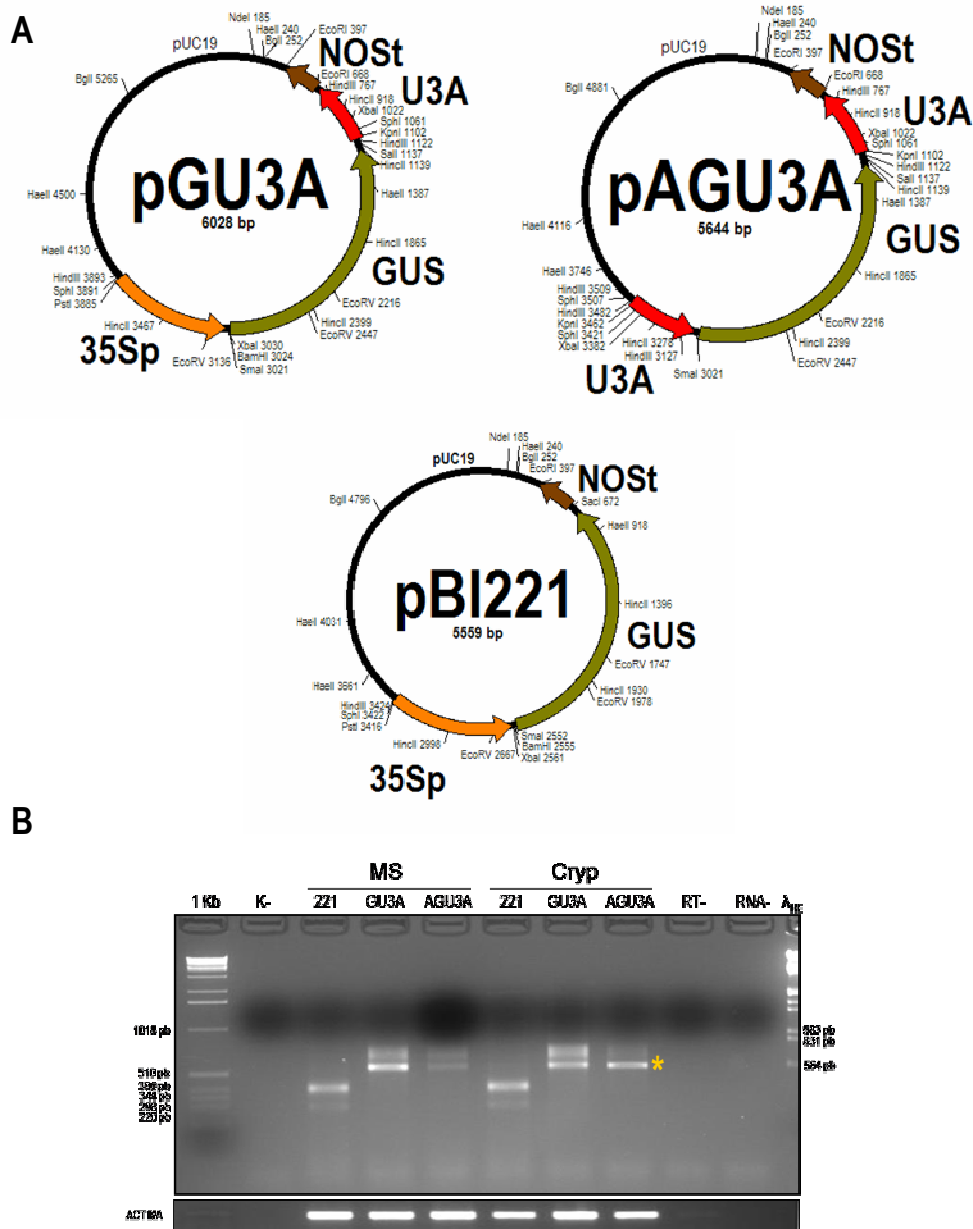


Figura 16. (A) Plásmidos utilizados en el bombardeo para el estudio de la influencia del inicio de la transcripción sobre su terminación. Todos los plásmidos derivan del pUC19. 35Sp: promotor 35S del CaMV. GUS: región codificante del gen de la β-glucoronidasa. NOST: terminador del gen de la nopalina sintasa. U3A: región U3 de la LTR de los elementos Tnt1A. (B) Electroforesis del producto de las PCRs realizadas con los oligos GUS3' y OligodT. PCR realizada con los oligos GUS3' y OligodT. 1Kb: marcador de tamaño. λ_{HE}: DNA del fago λ digerido con *HindIII* y *EcoRI*. K-: control de la PCR. RNA-: control de la retrotranscripción, reacción sin RNA. RT-: control de la retrotranscripción, reacción sin RT. MS: discos infiltrados con medio MS. Cryp: discos infiltrados con criptogeína. El asterisco naranja indica el fragmento que ha sido clonado y secuenciado.

En la figura 16B se puede observar una diferencia entre el patrón de bandas resultante de la expresión de las distintas construcciones génicas. La banda que aparece en los discos

bombardeados con el plásmido control pBI221 se corresponde en tamaño con la terminación del transcrito respondiendo a la señal contenida en la secuencia terminadora NOS. A pesar de no estar presente el inductor, las hojas bombardeadas con AGU3A e infiltradas con medio MS también presentan expresión del transgén, aunque a niveles más bajos. Este hecho puede ser debido a que el bombardeo y posterior infiltración de la hoja estén desencadenando una cierta respuesta a estrés por herida, que activaría la U3A promotora.

La expresión de la construcción GU3A genera transcritos con distinta terminación, como se ha mostrado en los experimentos anteriores. En cambio, aunque la construcción AGU3A también genera cierta variedad de transcritos, aparece una especie claramente mayoritaria en presencia del inductor de la expresión, criptogéina. Este fragmento se ha clonado y secuenciado y corresponde a un transcrito que termina en el tercer sitio de poliadenilación que se representa en la figura 15C. Se trata de un mRNA que termina antes de la señal de poliadenilación principal, descrita antes. De momento sólo hemos obtenido un clon de este producto de PCR; sería necesario analizar un mayor número de clones para confirmar que ésta es la secuencia de la especie mayoritaria.

Los resultados presentados en este apartado indican que la presencia de la región U3A promoviendo la transcripción del transgén GUS::U3A lleva a un aumento de la proporción de transcritos que terminan dentro de la región U3A terminadora., pudiendo existir una influencia de la región promotora o del proceso de inicio de la transcripción sobre la terminación y poliadenilación.

El hecho de que se generen preferentemente ciertas especies de mRNAs (definidas en función a su lugar de terminación) cuando se induce específicamente la expresión del retrotransposón en determinadas condiciones, sugiere que existe una influencia de la región promotora sobre la terminadora. Esta relación puede constituir un mecanismo de control de la transposición, que asegura que sólo transpongan los distintos elementos en las situaciones de estrés que promueven específica y fuertemente su expresión.