

## 2. ESTUDIO DEL SILENCIAMIENTO GÉNICO.

La expresión de Tnt1 en tabaco se activa por distintos estímulos bióticos y abióticos. La activación del promotor de Tnt1 por herida, elicitores microbianos y ataques por patógenos se mantiene en especies heterólogas como el tomate (*Lycopersicon esculentum*) o *Arabidopsis thaliana*. Sin embargo, en estas especies, a diferencia de lo que ocurre en su huésped natural, este retrotransposón también se transcribe ante estímulos abióticos como el ozono o la deshidratación (Moreau-Mhiri C *et al.*, 1996; Mhiri C *et al.*, 1997). Además, el promotor de Tnt1 es activo en partes aéreas de plantas de estas especies y, en particular, en tejidos florales (Moreau-Mhiri C *et al.*, 1996). Estas diferencias sugieren que existe una represión de la expresión de Tnt1 específica de tabaco (Grandbastien MA *et al.*, 2005).

Por otra parte, como ya se ha explicado en la introducción de esta memoria, los procesos de silenciamiento génico son los principales mecanismos por los que los genomas controlan la proliferación de secuencias extrañas, como los transposones. El número de copias de un elemento transponible en el genoma parece ser un determinante para que sea silenciado.

En el caso de Tnt1 se ha podido comprobar que, tras su introducción en el genoma de *Arabidopsis*, su amplificación desencadena su silenciamiento, no pudiendo superar un pequeño número de copias (Lucas H *et al.*, 1995). Una situación similar se ha observado para el retrotransposón Tto1, también de tabaco, cuando se introduce en *Arabidopsis*. Además, la actividad de los elementos Tto1 silenciados sólo se recupera en el contexto mutante *ddm1* (deficiente en metilación del DNA), en el que el silenciamiento transcripcional no es eficaz (Hirochika H *et al.*, 2000).

Pero, a pesar de la eficacia de los fenómenos de silenciamiento, Tnt1 puede expresarse en tabaco en determinados casos (ver introducción de este capítulo).

En este trabajo pretendemos conocer qué ocurre en estas situaciones, por qué puede llegar a expresarse Tnt1. Hemos formulado dos hipótesis que podrían explicar esta cuestión:

- (1) en determinadas condiciones de estrés el silenciamiento génico general se relaja;
- (2) las características de las LTRs de Tnt1 le permiten, en determinadas condiciones de estrés, superar el silenciamiento que opera sobre este retrotransposón, sin que el silenciamiento general se vea comprometido.

En este apartado se describen las estrategias seguidas para intentar confirmar alguna de estas dos hipótesis.

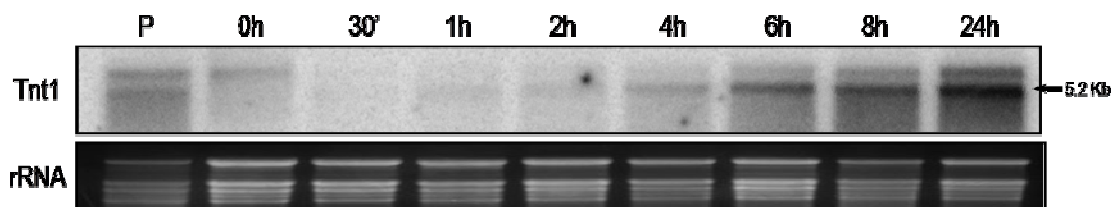
## 2.1. ESTUDIO DEL EFECTO DEL ESTRÉS SOBRE EL SILENCIAMIENTO.

El estudio del efecto del estrés sobre el silenciamiento lo hemos abordado utilizando sistemas de inducción de TGS sobre una región promotora, cuya activación está bien caracterizada.

### 2.1.1. Establecimiento de un sistema de inducción de la expresión de Tnt1.

En primer lugar, era necesario encontrar y establecer unas condiciones en las que Tnt1 se expresase, de forma reproducible, con tal de poder comparar las situaciones en las que este retrotransposón se transcribe (no está siendo silenciado) con la situación normal, en la que el promotor no es activo.

Trabajos anteriores habían demostrado que Tnt1 se expresaba en condiciones de estrés por herida e infección (ver introducción). Aprovechando estos datos diseñamos un sistema por el cual inducimos la expresión de Tnt1 controlando el tiempo necesario para poder detectar fácilmente los transcritos de este retrotransposón. Para ello seguimos dos estrategias, el aislamiento de protoplastos a partir de hojas de plantas de tabaco sombreadas y, por otro lado, herir estas mismas hojas, rayándolas en su reverso, e incubarlas en medio T<sub>0</sub> en presencia de la celulasa R-10 de Onozuka, que actúa como un elicitor fúngico. Se tomaron muestras a diferentes tiempos y se analizaron mediante hibridaciones de Northern.



**Figura 17.** Inducción de la expresión del retrotransposón Tnt1. Hibridación de Northern sobre RNA de protoplastos de tabaco (P) o de hojas de tabaco, rayadas e incubadas en medio MS más la celulasa R-10 durante distintos tiempos.

Los resultados muestran que con ambos sistemas hemos sido capaces de inducir la expresión de Tnt1 (figura 17). El resultado de la técnica de Northern sobre las hojas heridas e incubadas con la celulasa R-10 de Onozuka indica que a partir de las 4 horas de inducción se detecta el transcrito de Tnt1 (5,2 Kb) y su expresión continúa hasta las 24 horas.

Debido a la sencillez metodológica que presenta respecto a la obtención de protoplastos y a que requiere menos tiempo y se obtiene mayor cantidad de RNA a partir de la misma cantidad de muestra original, a partir de este momento las inducciones de la expresión del retrotransposón de Tnt1 se harán rayando el revés de las hojas de tabaco e incubándolas durante seis horas en medio MS en presencia de la celulasa R-10 de Onozuka.

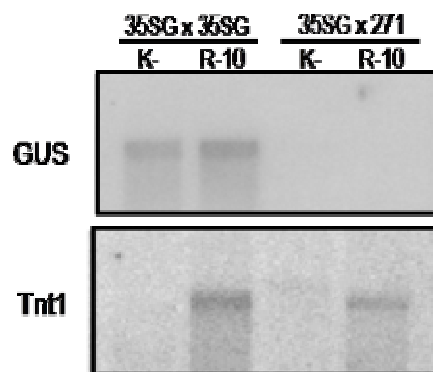
### **2.1.2. Estudio del locus silenciador 271.**

El grupo del Dr. Vaucheret ha caracterizado un locus silenciador, el locus 271, generado por la integración de múltiples copias del transgén 35Sp::Nitrito-reductasa, que tiene la particularidad de silenciar transcripcionalmente cualquier gen que contenga, al menos, 90 nt de la secuencia del promotor constitutivo 35S del CaMV (Thierry D and Vaucheret H, 1996).

Pensamos en utilizar esta línea para estudiar el efecto que tiene el estrés sobre el silenciamiento inducido por el locus 271. Para ello cruzamos plantas de *Nicotiana tabacum* var. *Xanthi* portadoras del gen marcador GUS bajo el control del promotor 35S (llamadas 35SG en este trabajo) con plantas portadoras del locus 271 (cedidas por el Dr. Hervé Vaucheret). Finalmente, se tomaron hojas de las plantas resultantes de este cruce y de plantas control (plantas 35SG autopolinizadas) y se estresaron de la forma indicada en el apartado anterior.

La figura 18 representa una hibridación de Northern con la sonda GUS. Observamos expresión del gen marcador GUS en las plantas 35SG autopolinizadas y, en cambio, no se expresa en las plantas cruzadas con las portadoras del locus silenciador 271. A pesar de que el tratamiento con la celulasa R-10 activa la transcripción de Tnt1 (como puede comprobarse en la hibridación de la misma membrana con la sonda Tnt1, ver figura 18), esta situación de estrés no afecta al silenciamiento inducido por el locus 271, puesto que en estas condiciones no se recupera la expresión de GUS en estas plantas.

Por lo tanto, vemos que existe un silenciamiento transcripcional sobre el promotor 35S que bloquea la transcripción de GUS y que este silenciamiento no desaparece en condiciones de estrés.



**Figura 18.** (A) Efecto del estrés sobre el silenciamiento del promotor 35S. Ver texto para leyenda.

Con el objeto de descartar un posible efecto transitorio del estrés sobre el silenciamiento, analizamos la expresión del gen GUS a distintos tiempos. No detectamos reactivación transitoria de la expresión de GUS con el tiempo. El silenciamiento del promotor 35S promovido por el locus 271 se mantiene a lo largo de todo el intervalo de tiempo analizado (hasta 24 horas).

Los datos obtenidos a partir de estos experimentos parecen indicar que el estrés no tiene ningún efecto sobre el silenciamiento ejercido por el locus 271, ya que no se recupera la expresión del gen GUS en protoplastos. Sin embargo, el locus 271 es un silenciador particularmente fuerte, manteniéndose incluso a lo largo de las generaciones cuando ya se ha segregado este locus (Park YD *et al.*, 1996) y puede que no sea el modelo ideal para estudiar la posible relajación transitoria de los mecanismos de silenciamiento.

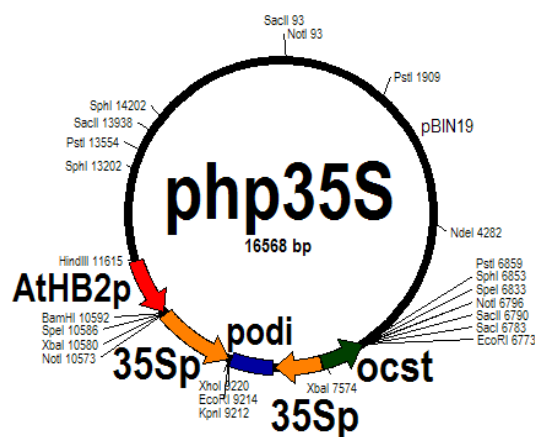
En el siguiente apartado se describe una estrategia alternativa que nos permitió estudiar el efecto del estrés sobre el silenciamiento del promotor 35S evitando la interferencia del potente locus silenciador 271.

### 2.1.3. Diseño de una estrategia de silenciamiento transitoria y local.

En los últimos años se han diseñado diferentes estrategias para silenciar secuencias concretas, desde el uso de cadenas simples de RNA antisentido (a la secuencia del RNA que se pretende eliminar) hasta los más eficaces hairpins (dobles cadenas con forma de horquilla) de RNA (Wesley SV *et al.*, 2001). En la mayoría de los casos todas estas técnicas estaban dirigidas a

conseguir la degradación de un RNA mensajero, es decir, a inducir un silenciamiento post-transcripcional. Sin embargo, se han utilizado también estrategias similares para el establecimiento del TGS en *Arabidopsis* (Mette MF *et al.*, 2000).

Siguiendo estas mismas estrategias, diseñamos y construimos un gen cuya expresión origina un transcrito que se pliega en forma de horquilla (hairpin), generando una región de doble cadena de RNA, cuya secuencia es la del promotor 35S. Esta molécula de dsRNA es reconocida por la maquinaria celular de silenciamiento, que la degrada generando siRNAs con la secuencia del promotor 35S. A esta construcción génica la hemos llamado hp35S (ver figura 19). Esta construcción sirvió para silenciar de forma transitoria y local el promotor 35S de plantas 35S::GUS, generadas a tal efecto. En concreto, se utilizó la línea 35SG 83.

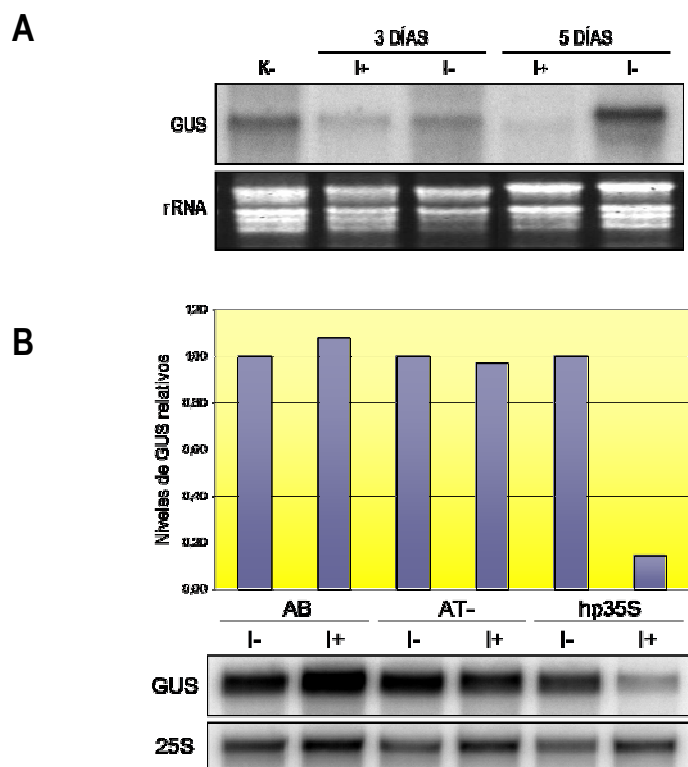


**Figura 19.** Esquema del plásmido binario conteniendo la construcción hp35S utilizado en los experimentos de agroinfiltración. El casete hp35S se ha construido a partir del plásmido pHANNIBAL (ver material y métodos). El plásmido binario utilizado como vector es el pBIN19 (ver material y métodos). AtHB2p: promotor del gen AtHB2 de *Arabidopsis*, de expresión constitutiva en tabaco (Martínez-García JF y Ludevid D, comunicación personal). 35Sp: promotor 35S del CaMV. podi: intrón pod. ocst: terminador del gen de la octopina sintasa.

Inyectamos una solución de *Agrobacterium tumefaciens* cepa LBA4404, conteniendo la construcción génica hp35S en hojas de plantas de tabaco transgénicas 35SG. La agroinfiltración se realizó en toda la superficie de la tercera o cuarta hoja verdadera, de plantas de unas 6 semanas de edad (con 4 ó 5 hojas verdaderas). En experimentos similares, en los que se inducía el PTGS, se ha podido comprobar que el tiempo óptimo requerido para estudiar la expresión y efecto del transgén es de 4 a 6 días tras la infiltración (Llave C *et al.*, 2002). Por ello,

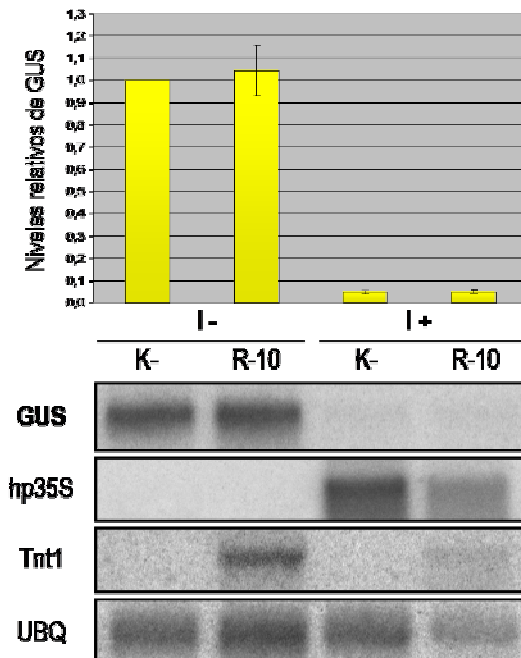
esperamos tres y cinco días tras la infiltración para recoger las muestras. Posteriormente analizamos la expresión del gen GUS mediante la técnica de Northern.

En la figura 20A se muestran los resultados obtenidos. A los tres días, la expresión de GUS disminuye en la hoja agroinfiltrada con *A.tumefaciens* conteniendo el hp35S, pero es a los cinco días cuando se observa una reducción notable del transcrito de GUS, aunque no llega a desaparecer totalmente. Este resultado es debido exclusivamente a la presencia del hp35S y no a otros efectos derivados de la infección o la infiltración, ya que la expresión de GUS no se ve alterada en hojas agroinfiltradas con la misma cepa de *A.tumefaciens* vacía (AT-, sin plásmido binario) o sólo infiltradas con la solución de infiltración (Agro buffer, AB, sin bacterias) (ver figura 20B).



**Figura 20.** Inducción del silenciamiento transcripcional del promotor 35S. Análisis por hibridación de Northern de la expresión del gen GUS, controlada por el promotor 35S, en hojas de tabaco transformadas con el transgén hp35S mediante agroinfiltración. (A) Comparación entre el TGS a los 3 días y a los 5 días tras de la agroinfiltración. (B) Comprobación de que el silenciamiento obtenido se debe exclusivamente a la presencia del transgén hp35S. La expresión del gen GUS se ha cuantificado respecto a la hibridación control (rRNA 25S) y se ha relativizado en función de los niveles obtenidos en las hojas no tratadas. K-: hoja control. I+: hoja agroinfiltrada. I-: hoja no agroinfiltrada. AB: hojas infiltradas sólo con Agro Buffer (tampón de agroinfiltración). AT-: hojas agroinfiltradas con *A. tumefaciens* vacío, sin vector binario. hp35S: hojas agroinfiltradas con *A. tumefaciens* conteniendo el plásmido php35S. 25S: sonda del RNA ribosomal 25S.

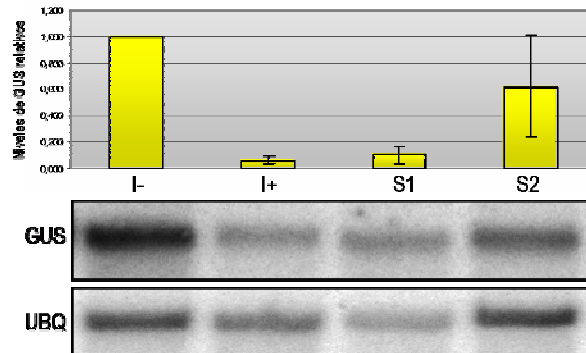
Puesto que nuestro objetivo era conocer el efecto que pueda ejercer el estrés sobre el silenciamiento génico, realizamos un experimento en el que después de agroinfiltrar y esperar 5 días, estresamos mitad de la hoja infiltrada. El análisis de la expresión del gen GUS (figura 21) indica que el estrés no afecta significativamente al silenciamiento dirigido sobre el promotor 35S.



**Figura 21.** Análisis del efecto del estrés sobre el silenciamiento del promotor 35S. Hibridaciones de Northern sobre RNAs extraídos de hojas agroinfiltradas con el plásmido php35S y estresadas por incubación con R-10. I-: hoja no agroinfiltrada. I+: hoja agroinfiltrada. K-: mitad de la hoja no estresada. R-10: mitad de la hoja sometida a estrés. El gráfico muestra el resultado obtenido tras cuatro repeticiones del experimento.

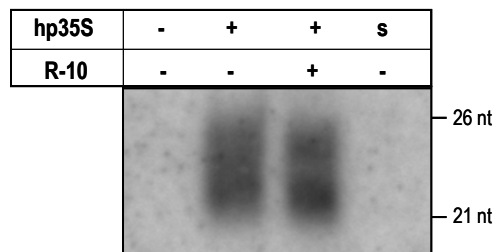
Por otra parte, el análisis de la expresión de GUS en las hojas no infiltradas (figura 22) parece indicar que el TGS podría transmitirse al resto de la planta. Diferentes trabajos han demostrado que el PTGS se transmite de forma sistémica y especulan sobre la naturaleza de la señal encargada de esta propagación (Palauqui JC *et al.*, 1997; Hamilton A *et al.*, 2002; Mlotshwa S *et al.*, 2002). Actualmente se postula que los siRNAs podrían viajar por el floema a otros tejidos donde la presencia del transcrito diana permitiría la amplificación de la señal de silenciamiento, mediante la acción de RNA polimerasas dependientes de RNA (Yoo BC *et al.*, 2004; Humber C *et al.*, 2003; Mlotshwa S *et al.*, 2002). Sin embargo, hasta el momento, no se ha podido detectar la transmisión del TGS. Este hecho se ha interpretado como una incapacidad de amplificar la señal de siRNA debido a que las dianas del TGS son secuencias no transcritas (Mlotshwa S *et al.*, 2002). También se ha propuesto que el PTGS y el TGS están mediados por distintas especies de siRNAs, de 21-23 nt y de 24-26 nt respectivamente, pudiendo residir en esta diferencia de

tamaños la capacidad de transmitirse que tiene el primer tipo de silenciamiento y la incapacidad para hacerlo que parece tener el segundo, aunque esto parece no ser así siempre (Hamilton A *et al.*, 2002; Yoo BC *et al.*, 2004; Himer C *et al.*, 2003; Timmons L, 2002).



**Figura 22.** Análisis de la transmisión del TGS inducido sobre el promotor 35S. I-: hoja no agroinfiltrada. I+: hoja agroinfiltrada con el plásmido php35S. S1: hoja sistémica inmediatamente más joven que la hoja agroinfiltrada. S2: siguiente hoja sistémica. UBQ: sonda de la ubiquitina. El gráfico muestra el resultado obtenido tras cuatro repeticiones del experimento.

Teniendo en cuenta todos estos trabajos anteriores, quisimos detectar los siRNAs que estarían dirigiendo el silenciamiento del transgén 35S::GUS en la hoja agroinfiltrada y en la hoja sistémica. En la figura 23 se muestra el resultado de la hibridación del Northern de sRNAs con la sonda del promotor 35S. Como puede observarse, se detectan sRNAs con homología de secuencia al promotor 35S en la hoja agroinfiltrada, pero no en la sistémica (s), sugiriendo que los siRNAs correspondientes a la secuencia del promotor 35S no se transmiten a la hoja sistémica. Sin embargo, podría ocurrir que la transmisión de bajos niveles de siRNAs, indetectables por la técnica de hibridación de Northern, pueda ser suficiente para desencadenar el TGS. Por otra parte, la señal que se transmite y promueve el silenciamiento en la hoja sistémica podría ser otro tipo de molécula distinta de los siRNAs.

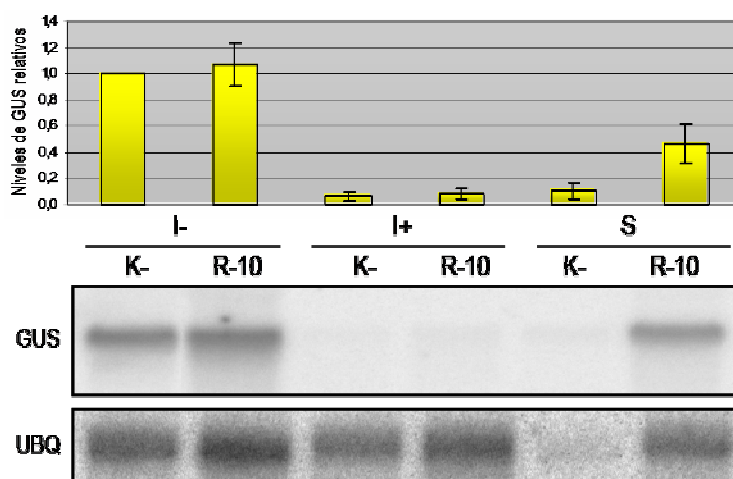


**Figura 23.** Análisis por Northern de los siRNAs con homología de secuencia al promotor 35S, originados a partir del transgén hp35S. R-10: hoja estresada. s: hoja sistémica.



El análisis por Northern (sobre mRNAs y siRNAs) de la hoja infiltrada muestra que tanto el nivel de expresión del transgén hp35S como la producción de siRNAs son elevados (ver figuras 21 y 23, respectivamente). Por lo tanto, podría ser que, aunque el estrés sea capaz de contrarrestar parcialmente el silenciamiento génico, este efecto no sea visible en las hojas infiltradas.

Dado que, en nuestro sistema, el TGS parecía poderse transmitir (aunque con menos intensidad), analizamos la expresión de GUS en la hoja siguiente a la infiltrada, es decir, la inmediatamente más joven, de la misma forma que se había hecho antes. La expresión del gen GUS en la hoja no agroinfiltrada es algo mayor que en la tratada, pero sigue siendo menor que la de una hoja control. Este resultado sugiere que el silenciamiento sistémico es menor. Cuando se somete a un estrés, la hoja sistémica recupera la expresión de GUS a niveles similares a los originales (figura 24). Estos resultados nos hacían pensar que el estrés podría recuperar la expresión de GUS en la hoja que presentaba un silenciamiento sistémico. La razón por la que el estrés afectara a los mecanismos de silenciamiento en la hoja sistémica y no en la hoja agroinfiltrada o “fuente” (de la señal que se transmite) podría ser que el silenciamiento que se estaría llevando a cabo “a distancia” sea más débil o de otra naturaleza.

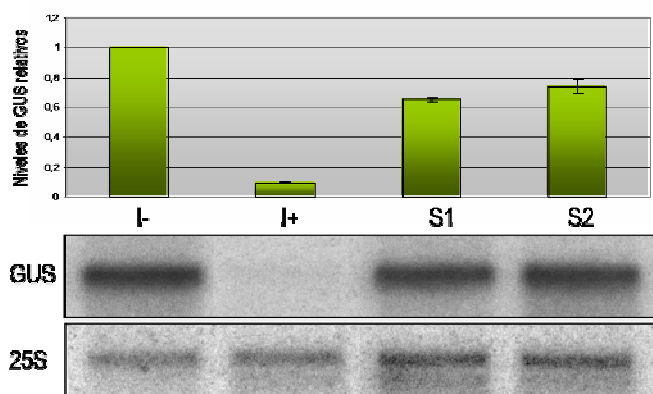


**Figura 24.** Análisis del efecto del estrés sobre el TGS sistémico del promotor 35S del CaMV. I-: hoja no agroinfiltrada. I+: hoja agroinfiltrada con el transgén hp35S. S: hoja sistémica. K-: mitad de la hoja no estresada. R-10: mitad de la hoja sometida a estrés. El gráfico muestra el resultado obtenido tras cuatro repeticiones del experimento.

A pesar de que distintas repeticiones de este experimento mostraban, en un primer momento, resultados similares (ver figura 24), repeticiones posteriores mostraron resultados diferentes. La

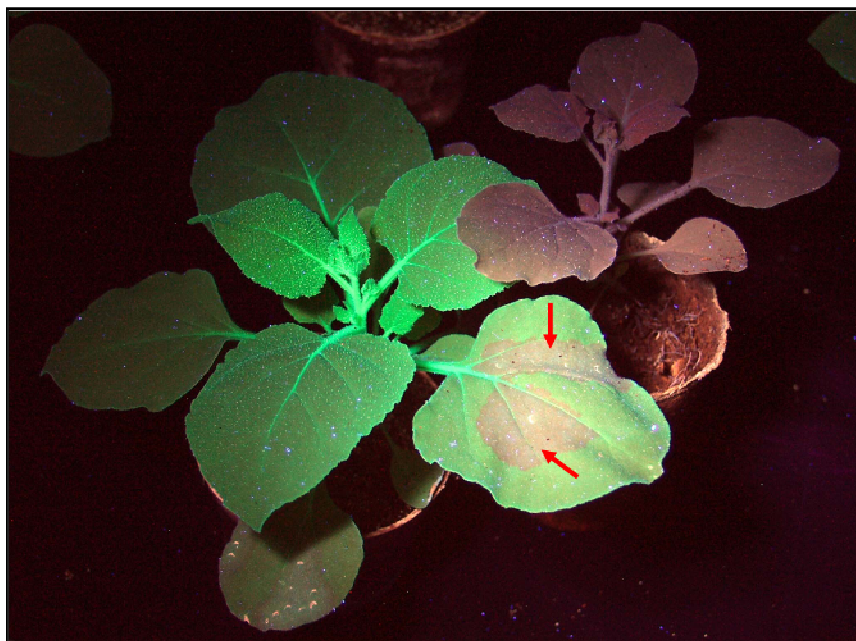
hoja agroinfiltrada siempre se comporta igual, se induce el silenciamiento transcripcional del promotor 35S, que resulta en la disminución de la cantidad del mRNA del gen GUS, que se mantiene a los mismos niveles en condiciones de estrés. En cambio, la expresión de GUS en la hoja sistémica variaba entre distintos experimentos. Las últimas réplicas de los experimentos, en las que todas las condiciones se controlaron cuidadosamente no nos han permitido confirmar de manera inequívoca la transmisión del TGS (ver figura 25).

Diferentes parámetros podían estar alterando la reproducibilidad del experimento. No podemos descartar que, al haber hecho los experimentos en distintas épocas del año, pequeñas variaciones en el desarrollo de las plantas puedan afectar al TGS.



**Figura 25.** Análisis de la transmisión del TGS inducido sobre el promotor 35S. I-: hoja no agroinfiltrada. I+: hoja agroinfiltrada con el plásmido php35S. S1: hoja sistémica inmediatamente más joven que la hoja agroinfiltrada. S2: siguiente hoja sistémica. El gráfico muestra el resultado obtenido tras tres repeticiones del experimento.

Por ello decidimos analizar, en paralelo, el TGS en otro sistema experimental distinto. Hemos utilizado plantas transgénicas 35S::GFP de la especie *Nicotiana benthamiana* (cedidas por la doctora Montserrat Martín, del LGMV, consorcio CSIC-IRTA). Infiltramos hojas de estas plantas con *A. tumefaciens* transformado con la construcción génica hp35S y hemos ido controlando la fluorescencia de los sectores agroinfiltrados de la hoja a lo largo de los días. A partir del quinto día tras la infiltración se puede observar claramente una pérdida significativa de fluorescencia únicamente en las áreas infiltradas (ver figura 26). Esta situación parece mantenerse indefinidamente, indicando que este tipo de silenciamiento no se transmite al resto de la hoja y de la planta.



**Figura 26.** Silenciamiento de la expresión de la GFP, mediante la inducción del TGS del promotor 35S por agroinfiltración de hojas de *N. benthamiana* 35S::GFP.

En conjunto, nuestros resultados muestran que el estrés no compromete de manera significativa al silenciamiento transcripcional. En las condiciones de estrés en las que se expresa Tnt1, el silenciamiento transcripcional general no parece relajarse.

## 2.2. ESTUDIO DEL SILENCIAMIENTO DE Tnt1.

Si, como parecen indicar los resultados anteriores, el estrés no afecta de manera significativa al silenciamiento, la capacidad de superar el silenciamiento y de expresarse, en determinadas condiciones, debe residir en el propio elemento Tnt1. Por ello decidimos estudiar cómo se inactiva este retrotransposón en su huésped natural, *Nicotiana tabacum*, y analizar cómo responden las regiones promotoras comprendidas en sus LTRs al silenciamiento.

Como he explicado anteriormente en esta memoria, pequeñas moléculas de RNA de 21-26 nt de longitud, a las que se les denomina pequeños RNAs de interferencia (siRNAs), parecen estar implicadas en los distintos procesos de silenciamiento génico, mediando y dirigiendo tanto la degradación o bloqueo de la traducción de los RNAs (en el silenciamiento post-transcripcional) como la modificación del estado de la cromatina y del DNA (en el silenciamiento transcripcional).

Estos siRNAs presentan homología, de distinto grado según el organismo y el proceso que medien, con la secuencia que se silencia (Hannon GJ, 2002; Voinnet O, 2002). Así pues, la presencia de estos siRNAs es indicadora de que se está produciendo un proceso de silenciamiento de sus secuencias homólogas. Sin embargo, procesos epigenéticos como la modificación de la cromatina por metilación o acetilación de histonas o la metilación del DNA pueden ser heredados y no siempre se ha encontrado una relación con siRNAs que pudieran estar dirigiendo estos procesos (Hirochika H *et al.*, 2000; Slotkin RK *et al.*, 2005).

En este trabajo, nosotros decidimos seguir diferentes aproximaciones para determinar si en las condiciones en las que Tnt1 no se expresa hay siRNAs que medien su silenciamiento y analizar si desaparecen, o no, cuando Tnt1 se transcribe. Buscamos la presencia de pequeños RNAs en las situaciones “control” (de no expresión de Tnt1) y de estrés, mediante la técnica de la hibridación de Northern. También decidimos realizar un Northern inverso, con el cual además de poder identificar la existencia de siRNAs correspondientes a Tnt1, podríamos determinar de qué región de este retrotransposón derivan o contra cuál están dirigidos. Asimismo, construimos una librería de los sRNAs presentes en hoja de tabaco en condiciones normales para tener una idea de qué tipo de pequeños RNAs se generan en esta planta, qué proporción presenta homología con secuencias repetidas y, como objetivo último, para identificar algún pequeño RNA que presente homología de secuencia con el retrotransposón Tnt1.

### **2.2.1. Análisis de los intermediarios del silenciamiento de Tnt1. I: construcción de una librería de sRNAs de *Nicotiana tabacum*.**

Cuando comenzamos el trabajo (primavera 2004) se había clonado un gran número de pequeños RNAs de distintos organismos como *Caenorhabditis elegans*, *Drosophila* o *Arabidopsis* y se disponía de información acerca del tipo de secuencias con las que presentaban homología estos sRNAs (Ambros V *et al.*, 2003; Aravin AA *et al.*, 2003; Llave C *et al.*, 2002). Los perfiles de este tipo de secuencias varían entre los diferentes organismos, pero en todos ellos se ha encontrado una proporción considerable de sRNAs presenta homología de secuencia con elementos transponibles. Se tenía muy poca información acerca de los pequeños RNAs de *Nicotiana tabacum*, pero también apuntaba a la existencia de siRNAs endógenos que presentan homología con transposones. Por ejemplo, un único trabajo publicado describe la existencia de

pequeños RNAs con homología a la secuencia de Tnt1, concretamente con las LTRs y con la región *gag*, basándose en hibridaciones de Northern (Hamilton A *et al.*, 2002).

Con el clonaje de la población de sRNAs de hoja de tabaco pretendemos conocer qué tipo de pequeños RNAs se producen en esta planta, qué proporción presenta homología con secuencias repetidas y, como objetivo último, identificar algún pequeño RNA que presente homología de secuencia con el retrotransposón Tnt1.

Para construir la librería de pequeños RNAs de hoja de tabaco se aisló y clonó toda la fracción de RNAs de un tamaño comprendido entre 21 y 26 nt de longitud aproximadamente. De todos los pequeños RNAs clonados disponemos, de momento, de 52 sRNAs secuenciados, del tamaño esperado. Se ha realizado una búsqueda por homología de secuencia sobre las bases de datos (de los sitios SOLGD, TAIR y NCBI) para conocer a qué secuencias corresponden los pequeños RNAs clonados (tabla 3 y apéndice 2), pero una de las limitaciones de este trabajo es que no hemos dispuesto de la secuencia completa del genoma de *Nicotiana tabacum*, ya que no es de acceso público.

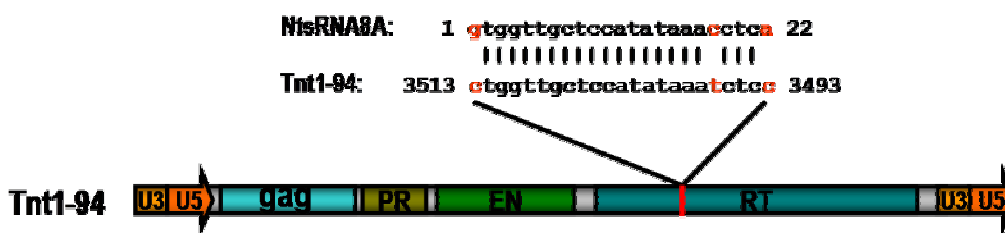
Tipo de secuencia	sRNAs
<b>Retrotransposones</b>	<b>8</b>
RETs con LTRs	4
SINEs	4
<b>tRNAs</b>	<b>3</b>
<b>rRNA</b>	<b>1</b>
<b>Secuencias génicas</b>	<b>8</b>
Conocidas	4
región codificante	2
Intrón	2
predichas/desconocidas	4

**Tabla 3.** Librería de sRNAs de *N. tabacum*. 20 de los 52 sRNAs clonados y secuenciados presentan homología con algún tipo de secuencia anotada en las bases de datos.

Ocho clones tienen homología con ESTs existentes en las bases de datos, de los cuales sólo cuatro presentan homología con algún gen anteriormente descrito. De estos últimos, tres se corresponden con factores de transcripción y en sólo dos de ellos la homología de secuencia con el sRNA tiene lugar dentro de la región codificante. Otro de los ESTs identificados presenta homología con una proteína putativa de respuesta a estrés hídrico de tomate. Los factores de

transcripción parecen ser dianas frecuentes del silenciamiento mediado por sRNAs, como lo demuestran la cantidad de siRNAs que presentan homología con estas secuencias clonados en otros trabajos. También es interesante la presencia de una secuencia con homología a una proteína inducida en estrés hídrico. Podría ser interesante analizar si este sRNA disminuye en situaciones de estrés, permitiendo la expresión del gen, lo que mostraría la existencia de una regulación de los genes de respuesta a estrés por silenciamiento.

Entre las secuencias identificadas, también han aparecido ocho que presentan similitud de secuencia con retrotransposones, la mitad son retrotransposones con LTRs y los otros cuatro están clasificados como SINEs de la familia TS. Uno de los ocho sRNAs que presentan homología de secuencia con retrotransposones, lo hace con el retrotransposón Tnt1, en concreto con la región que codifica para la retrotranscriptasa, una de las regiones más conservadas entre retrotransposones (figura 27).



**Figura 27.** Representación del sRNA 8A. Se muestra la homología existente con la secuencia del retrotransposón Tnt1 y la posición de la región homóloga dentro del elemento.

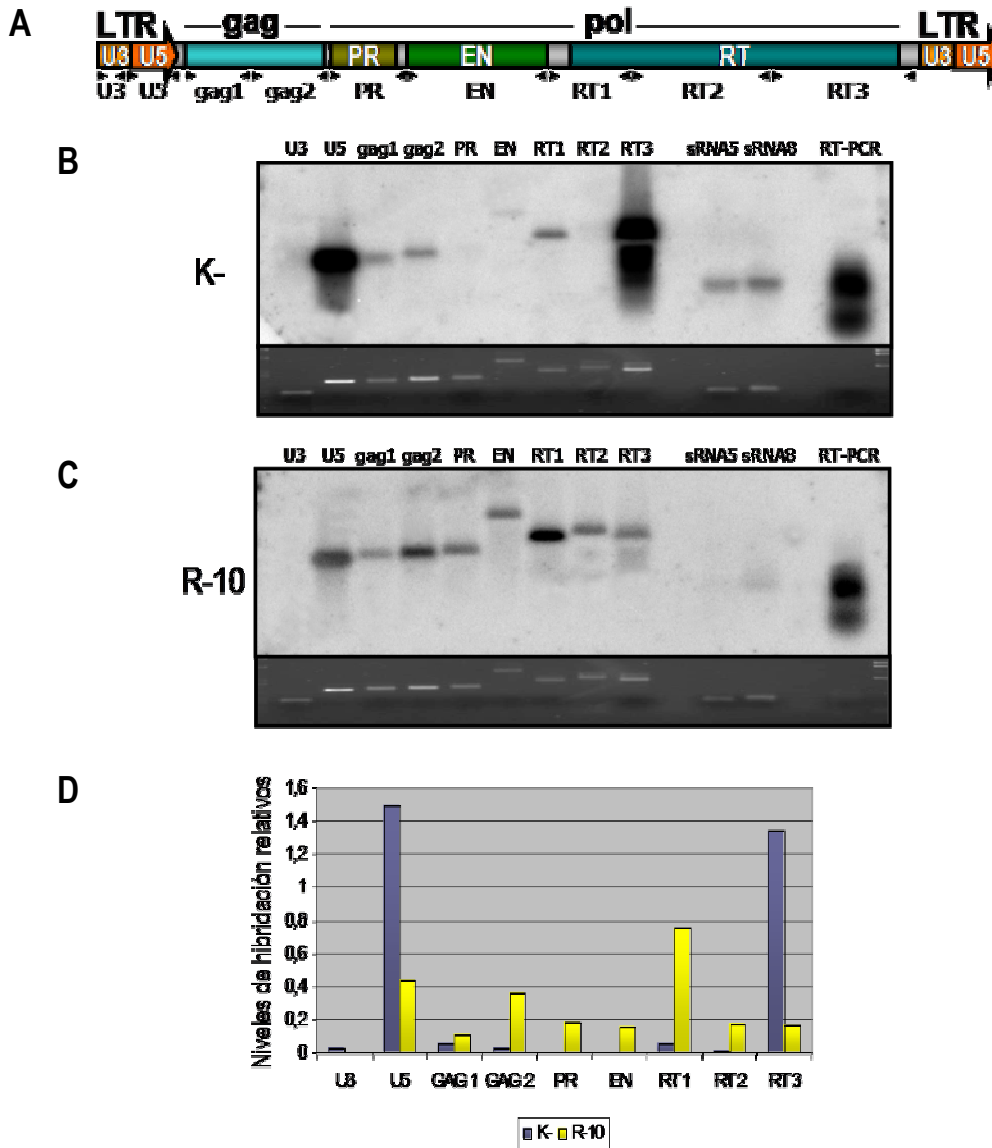
El hecho de que hayamos clonado varios pequeños RNAs que muestran homología de secuencia con elementos transponibles y, en concreto, con retrotransposones (ver tabla 3), está de acuerdo con la existencia de mecanismos activos de silenciamiento dirigidos contra transposones. Los elementos transponibles más abundantes en tabaco son los de clase I o retrotransposones, lo que puede explicar que sólo hayamos identificado sRNAs con homología a este tipo de transposones. Los SINEs de la familia TS son retrotransposones de pequeño tamaño derivados de tRNAs (Yoshioka Y *et al.*, 1993). Aunque estos elementos tienen características estructurales del tRNA-lys, podrían existir SINEs derivados de otros tRNAs que explicaran el hecho de que hayamos encontrado cuatro sRNAs con homología de secuencia a otros tRNAs. También puede ocurrir que algunas de las secuencias clonadas sean productos del metabolismo de los tRNAs, que comparten las características de los sRNAs (Llave C *et al.*, 2002),

en las que nos hemos basado para su clonación: un extremo 5'-fosfato y otro extremo 3'-hidroxilo libres. En distintos trabajos en los que se han creado librerías de sRNAs de distintos organismos también se han encontrado estos productos de la degradación de RNAs no codificantes (Sunkar R *et al.*, 2005; Aravin AA *et al.*, 2003).

### **2.2.2. Análisis de los intermediarios del silenciamiento de Tnt1. II: análisis de la acumulación de siRNAs.**

Mediante los Northern de los sRNAs, en los que comparamos la población de RNAs de bajo peso molecular (LMW RNAs) entre condiciones normales y de estrés, no hemos detectado siRNAs que presenten homología con Tnt1. Sin embargo, este resultado puede ser consecuencia de problemas metodológicos. En general, para detectar un determinado tipo de siRNAs se utilizan grandes cantidades de RNA y sondas de cadena simple complementarias al siRNA que se quiere detectar. El hecho de haber utilizado una sonda global (de toda la secuencia del retrotransposón) y el hecho de que en tabaco parece ser más difícil la detección de siRNAs que en *Arabidopsis* o en *N. benthamina* (Vaucheret, comunicación personal), podrían explicar este resultado negativo.

Por este motivo se decidió analizar la presencia de siRNAs de Tnt1 por Northern inverso. La secuencia de Tnt1 se dividió en nueve regiones que se amplificaron independientemente por PCR (figura 28A). Posteriormente, los distintos fragmentos se migraron electroforéticamente en un gel de agarosa y se transfirieron a una membrana de nylon para hibridarlas con la fracción de pequeños RNAs procedente de hoja de tabaco, marcados radiactivamente. En la figura 28 se muestran los resultados de las hibridaciones con sRNAs procedentes de hoja control y con sRNAs de hoja estresada. En ambas situaciones hemos detectado pequeños RNAs que hibridan con Tnt1. En las condiciones normales, en las que Tnt1 no se expresa, hemos detectado hibridación con la región U5 de las LTRs y con la parte final de la región que codifica para la retrotranscriptasa (RT56). Pero también aparecen sRNAs que hibridan, en menor medida, con la región *gag* y el comienzo de la región de la RT (RT12). En condiciones de estrés hemos detectado hibridación con casi toda la secuencia del retrotransposón Tnt1 (no hay señal para la región U3 de la LTR), aunque esta hibridación no es tan fuerte como la detectada en el caso anterior para las regiones U5 y RT56.



**Figura 28.** Análisis de los sRNAs con homología de secuencia a Tnt1. (A) Representación de las 9 regiones en las que se ha dividido la secuencia del retrotransposón Tnt1 y situación de los oligos utilizados para su amplificación. (B) Northern reverse sobre la secuencia de Tnt1 en el que se ha utilizado como sonda radiactiva la fracción de sRNAs aislada a partir de una hoja de tabaco no estresada (K-). (C) Northern reverse sobre la secuencia de Tnt1 en el que se ha utilizado como sonda radiactiva la fracción de sRNAs aislada a partir de una hoja de tabaco estresada (R-10). (D) Comparación de los niveles de hibridación de los sRNAs entre ambas situaciones.

Podemos encontrar dos explicaciones a que hayamos detectado hibridación a lo largo de toda la secuencia en las condiciones en que Tnt1 se expresa. Puede tratarse de siRNAs implicados en un proceso de PTGS sobre el mRNA de Tnt1 cuando éste se transcribe, o bien podrían ser productos de la degradación de este mRNA. En los experimentos de inducción de la expresión de Tnt1, hemos visto que este RNA se degrada con facilidad, detectándose un rastro de



hibridación a lo largo de todo el carril. En cualquier caso, de tratarse de un fenómeno de PTGS, éste no sería muy efectivo puesto que el mRNA de Tnt1 se acumula en cantidades considerables.

Por otro lado, estos resultados parecen indicar que en situaciones control, en las que Tnt1 no se expresa, existen siRNAs endógenos dirigidos contra su secuencia, los cuales podrían estar dirigiendo su inactivación. El hecho de que haya algunas especies de siRNAs más abundantes que otras puede ser debido a la especificidad del proceso de silenciamiento. La región de la RT es una de las más conservadas entre estos tipos de secuencias, de forma que la presencia de sRNAs dirigidos contra esta región podría ayudar a reprimir muchas copias e, incluso, otras familias de retrotransposones (que han podido divergir en otras partes de su secuencia) al mismo tiempo. Los sRNAs que hibridan con la región U5 de la LTR hacen pensar que realmente existe un silenciamiento transcripcional sobre Tnt1, ya que estarían afectando a la región promotora. No hemos detectado sRNAs que hibriden con la región U3 de la LTR, donde se localizan las principales secuencias promotoras de la transcripción de Tnt1. Sin embargo, ésta es la región más variable entre las distintas copias de Tnt1 y posiblemente un mecanismo de silenciamiento dirigido contra esta secuencia no conseguiría bloquear todos los elementos Tnt1.

El hecho de que los niveles de hibridación con la región U5 en la situación de estrés sean mayores que en condiciones de control, parece indicar que existen siRNAs dirigidos contra esta secuencia y que éstos disminuyen cuando Tnt1 se expresa. Puede ser que estos siRNAs estén promoviendo el silenciamiento de la U5 y que en condiciones de estrés desaparezcan, permitiendo la transcripción del retrotransposón.

Todos los datos obtenidos a partir de los experimentos anteriores sugieren que existe un mecanismo de silenciamiento génico del retrotransposón Tnt1 mediado por pequeños RNAs. Los datos obtenidos del Northern reverse sobre la secuencia de Tnt1 así como la aparición, entre los clones secuenciados, de una secuencia de un sRNA con homología de secuencia con este retrotransposón, parecen indicar que existen moléculas con características de pequeños RNAs derivadas o dirigidas contra la secuencia de este retrotransposón, en particular con la región U5 de la LTR. El haber localizado las regiones contra las que puedan estar dirigidos los sRNAs de Tnt1, nos puede permitir, en el futuro, realizar de nuevo los experimentos de Northern con sondas de cadena simple correspondientes a estas regiones.

### **2.2.3. Análisis de los intermediarios del silenciamiento de Tnt1. III: modificaciones de la cromatina.**

El silenciamiento génico transcripcional se manifiesta en cambios conformacionales en la cromatina. La región silenciada adquiere las características de heterocromatina, de forma que se inactivan las regiones promotoras y se bloquea la transcripción del gen que controlan. La heterocromatina se caracteriza por presentar un patrón de metilación de DNA e histonas determinado, así como otras modificaciones en las histonas (como la desacetilación) que hacen que se asocien a estas regiones del DNA una serie de proteínas específicas, impidiendo el acceso a la maquinaria de transcripción (Harikrishnan KN *et al.*, 2005; Mutskov V and Felsenfeld G, 2004; Volpe TA *et al.*, 2002; Mette MF *et al.*, 2000).

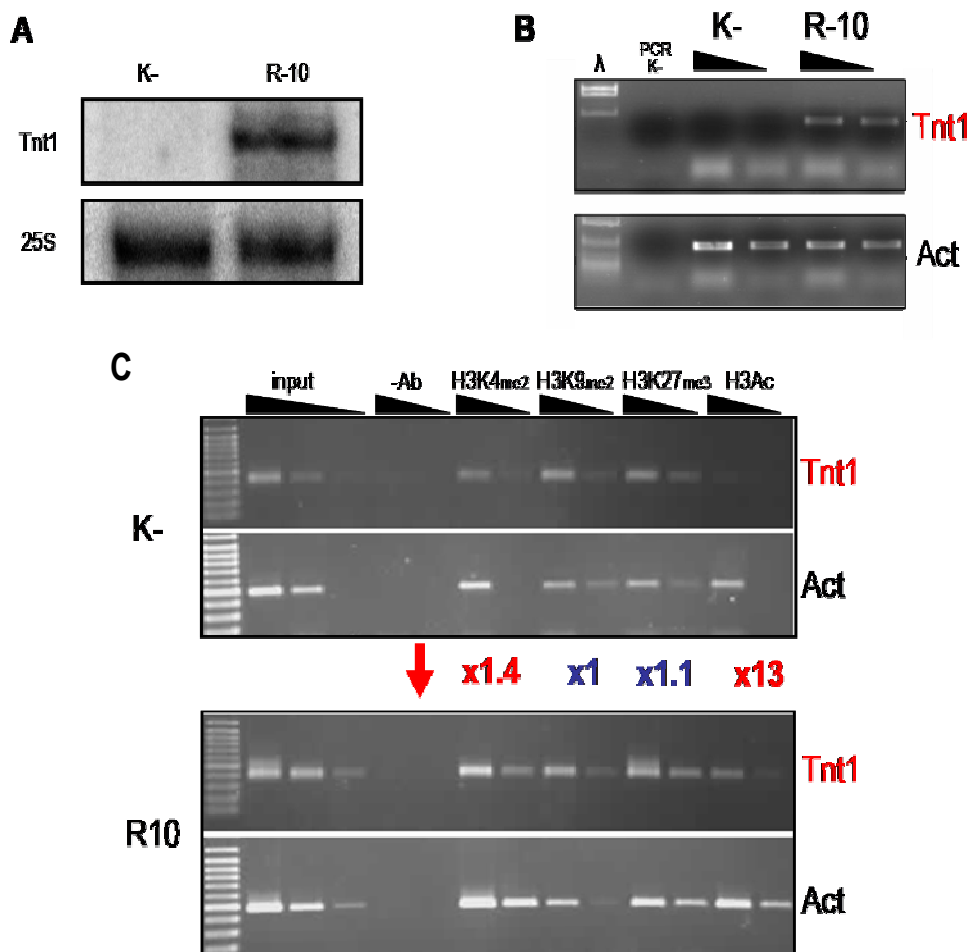
Se considera que existe un “código de las histonas”, definido por las modificaciones post-traduccionales que presentan estas proteínas (Cosgrove MS and Wolberger C, 2005), que permiten a la maquinaria nuclear diferenciar entre eucromatina (o cromatina activa) y heterocromatina. En *Arabidopsis* se ha visto que, mientras que en la eucromatina las H3 aparecen principalmente dimetiladas en su lisina número 4 (H3K4me<sub>2</sub>), en la heterocromatina las H3 suelen estar dimetiladas en sus lisinas 9 (H3K9me<sub>2</sub>). Pero además, la mono- y dimetilación de las lisinas 27 (H3K27me y H3K27me<sub>2</sub>) también se asocian a la heterocromatina silenciada.

En este trabajo hemos comparado el estado de la cromatina de las secuencias endógenas de Tnt1 en una situación de estrés, en la que se expresa, respecto a las condiciones normales, en las que no se expresa. Estos experimentos se han realizado en colaboración con la Dr. M. Lluïsa Espinás (IBMB).

La inducción de Tnt1 se ha realizado como se explica en el apartado 2.1.1. En la figura 29 se muestra el resultado de esta inducción, comprobado mediante Northern, sobre RNA total de hoja de tabaco utilizando una sonda específica de Tnt1, y RT-PCR, utilizando los oligos Avi y 13 (específicos de Tnt1). Las mismas muestras fueron tratadas convenientemente para llevar a cabo la técnica de ChIP (inmuno-precipitación de la cromatina) utilizando distintos anticuerpos que reconocen diferentes modificaciones de las histonas (anti-H3K4me<sub>2</sub>, anti-H3K9me<sub>2</sub>, anti-H3K27me<sub>3</sub> y anti-H3Ac). Posteriormente se amplificó un fragmento de Tnt1, utilizando los oligos Avi y 13 (ver material y métodos), a partir de cada una de las fracciones precipitadas. Como

control se amplificó el DNA de la actina, que es activo tanto en condiciones control como de estrés.

La figura 29C muestra el resultado de estas PCRs. En una hoja control se observa una mayor amplificación de secuencias de Tnt1 a partir de las fracciones inmunoprecipitadas con características de heterocromatina, es decir con anticuerpos anti-H3K9me2 y anti-H3K27me3. Por el contrario, cuando analizamos hoja estresada vemos que aumenta considerablemente la amplificación de secuencias de Tnt1 a partir de la fracción inmunoprecipitada con el anticuerpo contra la H3 acetilada y también aumenta en la fracción H3K4me2, mientras que no varía en la correspondiente a la H3K9me2.



**Figura 29.** Análisis del estado de la cromatina asociada a la expresión el retrotransposón Tnt1. La hibridación de Northern (A) y la RT-PCR (B) muestran la inducción de la expresión de Tnt1. (C) Experimento de ChIP. Electroforesis del producto de las PCRs realizadas sobre las distintas fracciones inmunoprecipitadas con los diferentes anticuerpos indicados. En las distintas PCRs se han utilizados cantidades crecientes de DNA molde (1 y 5  $\mu$ l). K-: hoja control, no estresada. R-10: hoja estresada. input: DNA total. -Ab: control de la inmunoprecipitación, no se ha utilizado anticuerpo.

Sin embargo, la amplificación de secuencias de Tnt1 asociadas a H3K27me3 también aumenta, aunque muy levemente. Recientemente se ha descrito que, en *Arabidopsis*, la trimetilación de la lisina 27 en las H3 heterocromáticas (H3K27me3) se correlaciona con una reactivación transcripcional (Mathieu O *et al.*, 2005). Quizás, a diferencia de lo que ocurre en animales, esto es general para las plantas.

Estos resultados parecen indicar que se han producido modificaciones en la cromatina asociada a Tnt1 paralelamente a su inducción, pasando de tener características de heterocromatina a cromatina transcripcionalmente activa. Esto sugiere que en situaciones control la transcripción de Tnt1 está reprimida a nivel cromatínico, mientras que esta inactivación desaparece en condiciones de estrés.

Puesto que existe un gran número de copias de Tnt1 y que no todas se inducen por igual en situaciones de estrés, hemos querido conocer qué secuencias están sufriendo estos cambios en la cromatina y cuáles se transcriben. Se han clonado y secuenciado los productos de la RT-PCR de la figura 29B y de la PCR de la figura 29C. Todas las secuencias expresadas, obtenidas por RT-PCR, pertenecen a la familia Tnt1A. Sin embargo, el análisis de la PCR a partir de las distintas fracciones inmunoprecipitadas indica que la modificación a cromatina activa se da sobre las tres familias de Tnt1. Este resultado induce a pensar que el retrotransposón Tnt1 se expresa como resultado de dos procesos, en cierto modo, independientes. En primer lugar, desaparece el silenciamiento transcripcional sobre las distintas copias de Tnt1 debido a cambios en la conformación en la cromatina, probablemente como respuesta al estrés. Y, en segundo lugar, se induce la expresión de las distintas familias de Tnt1 dependiendo del tipo de estrés que esté actuando. Una vez la cromatina se encuentra en un estado transcripcionalmente activo, los diferentes factores de transcripción podrían acceder a los promotores de las distintas copias del retrotransposón e inducir la expresión de una determinada.

### **2.3. ESTUDIO DEL EFECTO DE LA SECUENCIA DE LAS LTRs DE Tnt1 SOBRE EL SILENCIAMIENTO.**

Nuestros resultados indican que en situaciones de estrés disminuyen los siRNAs correspondientes a las LTRs de Tnt1 y que, paralelamente, la cromatina asociada a estas

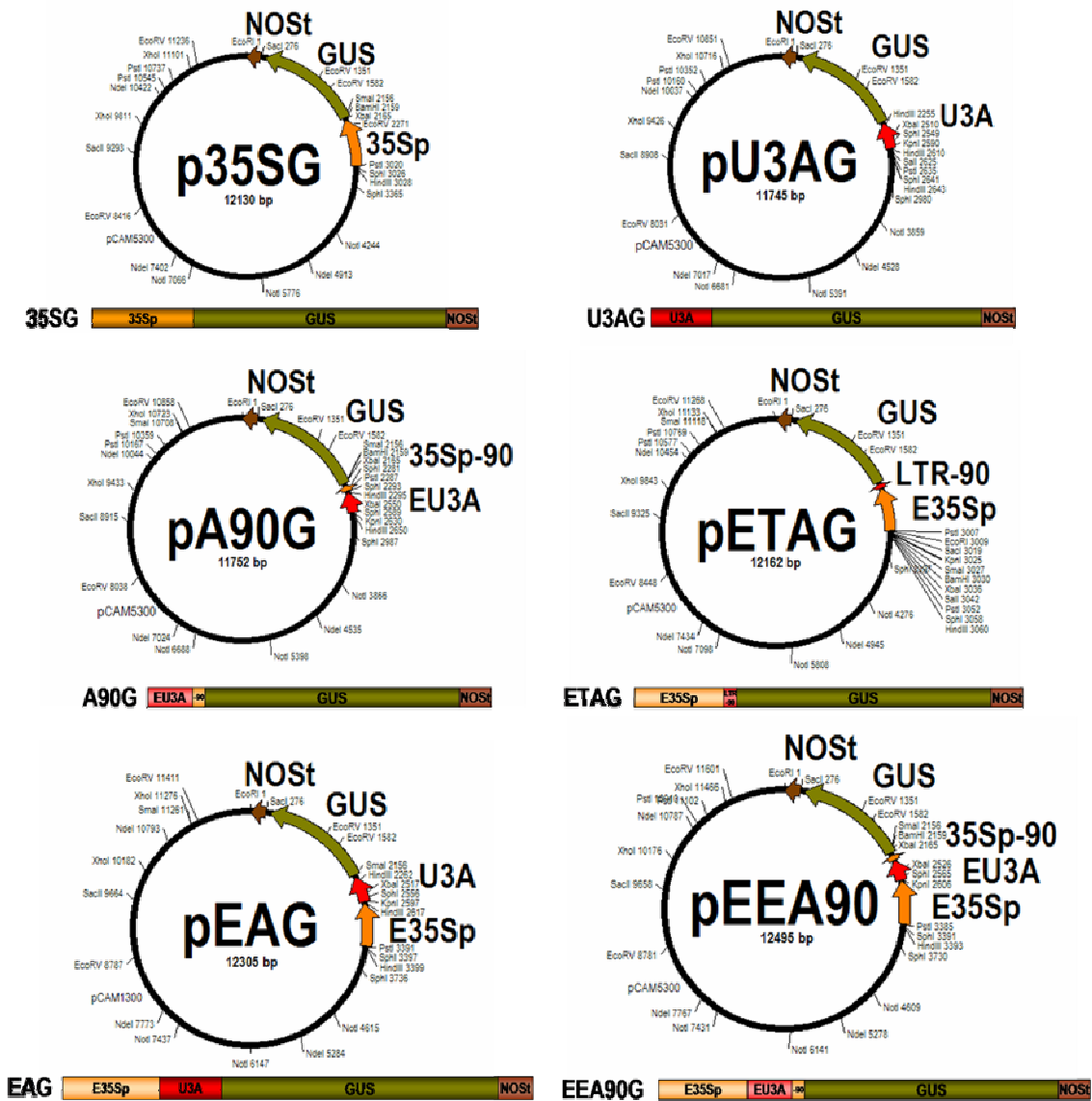
regiones se modifica para permitir la transcripción. Dado que no hemos visto cambios en el silenciamiento de otros loci reprimidos en condiciones de estrés (análisis del silenciamiento transcripcional del promotor 35S), creemos que, en estas situaciones, se produce una activación o des-silenciamiento específico de los elementos Tnt1.

Puesto que el TGS actúa sobre los promotores, inactivándolos, decidimos analizar si las regiones promotoras de Tnt1 son particularmente resistentes al silenciamiento en situaciones de estrés. Para ello decidimos construir promotores quiméricos que contuvieran secuencia de la región U3 de la LTR, donde se localizan los principales elementos activadores de la transcripción de Tnt1 (Casacuberta JM and Grandbastien MA, 1993; Vernhettes S *et al.*, 1997) y del promotor 35S del CaMV y que controlaran la expresión del gen marcador GUS. Posteriormente analizamos cómo se expresa del gen GUS cuando silenciamos las secuencias del promotor 35S de forma específica (como se ha explicado en los apartados anteriores). Pretendemos determinar si la presencia de la secuencia de la LTR de Tnt1 afecta al silenciamiento transcripcional ejercido sobre el promotor 35S y si es suficiente para que en condiciones de estrés se recupere la expresión del gen GUS. Con este objetivo generamos plantas de tabaco transgénicas para cada una de las construcciones génicas que se detallan a continuación.

### **2.3.1. Construcciones génicas.**

En la figura 30 se esquematizan las construcciones génicas que creamos para llevar a cabo el análisis del efecto de las LTR de Tnt1 sobre el silenciamiento de este retrotransposón. Hemos utilizado la secuencia de la LTR de Tnt1A ya que sabemos que es la familia que se expresa preferentemente en estrés generado por herida e infección. Hemos dividido la secuencia de la región U3 de la LTR en los fragmentos EU3A (*enhancer* de la U3A) y LTR-90 (constituida por los 90 nucleótidos que preceden al lugar de inicio de la transcripción). En la región EU3A se localizan los principales elementos activadores de la transcripción de Tnt1. La región LTR-90 es la que contiene la caja TATA, de inicio de la transcripción. De la misma manera hemos dividido el promotor 35S en dos regiones, a las que hemos denominado *enhancer* (E35Sp) y 35Sp-90, para complementar las anteriores y crear promotores quiméricos funcionales. La región -90 del promotor 35S es la que contiene la caja TATA.

Todas las construcciones génicas se clonaron en vectores binarios, pCAM5300 o pCAM1300 (ver material y métodos), para poder transformarlas en plantas de tabaco.

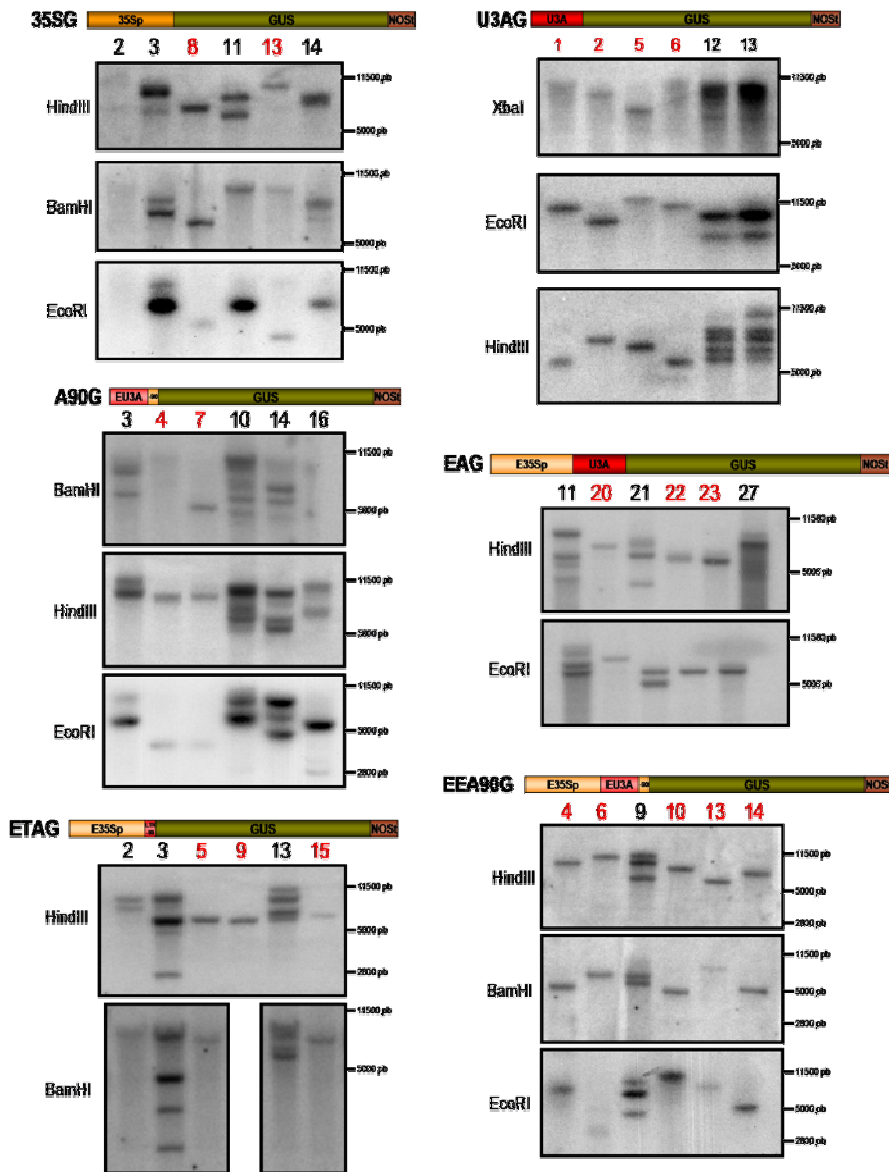


**Figura 30.** Representación esquemática de las construcciones génicas y los plásmidos binarios utilizados en el análisis del efecto de la secuencia de las LTRs de Tnt1 sobre el silenciamiento. Ver texto para explicación.

### 2.3.2. Generación de las plantas transgénicas.

Las diferentes plantas transgénicas se han generado siguiendo el método de transformación que utiliza la bacteria *Agrobacterium tumefaciens* como vector del transgén que queremos introducir en la planta. De esta manera, primero transformamos la cepa LBA4404 de *A. tumefaciens* con el vector binario conteniendo el transgén de interés. Posteriormente utilizamos estas bacterias para infectar hojas de *N. tabacum* var. Xanthi, transfiriendo el transgén a algunas células e induciendo

la formación de callos a partir de las mismas. A partir de estos callos indujimos la formación de plántulas individuales y las aislamos mediante selección por antibiótico.



**Figura 31.** Hibridación de Southern para analizar el número de copias del transgén en las diferentes plantas transformadas. Para cada construcción se muestra el análisis de 6 plantas independientes. Las plantas que poseen una sola copia del transgén se indican en rojo. Se indica el enzima de restricción utilizado en cada caso para digerir el DNA genómico.

A partir de estas plántulas regeneramos al menos 25 plantas totalmente desarrolladas para cada construcción génica y analizamos su dotación génica. Buscamos plantas que sólo hubieran incorporado una copia del transgén para evitar efectos de silenciamiento que interfirieran en nuestros experimentos. De las distintas plantas regeneradas para cada transgén buscamos al

menos 3 líneas monocopia independientes. En la figura 31 se muestran los resultados del análisis por hibridación de Southern de algunas de las líneas estudiadas para cada construcción. Una vez identificadas las plantas que sólo contenían una copia del transgén, las crecimos para autofecundarlas y conseguir una descendencia homocigota para la construcción génica introducida. Se obtuvo descendencia de al menos 2 líneas monocopia independientes para cada una de las construcciones a analizar. Determinamos si una planta era heterocigoto u homocigota germinando las semillas resultantes de su autofecundación en medio selectivo. Cuando obtuvimos plantas homocigotas monocopia las crecimos para poder analizar la expresión del transgén y llevar a cabo los experimentos de silenciamiento y estrés.

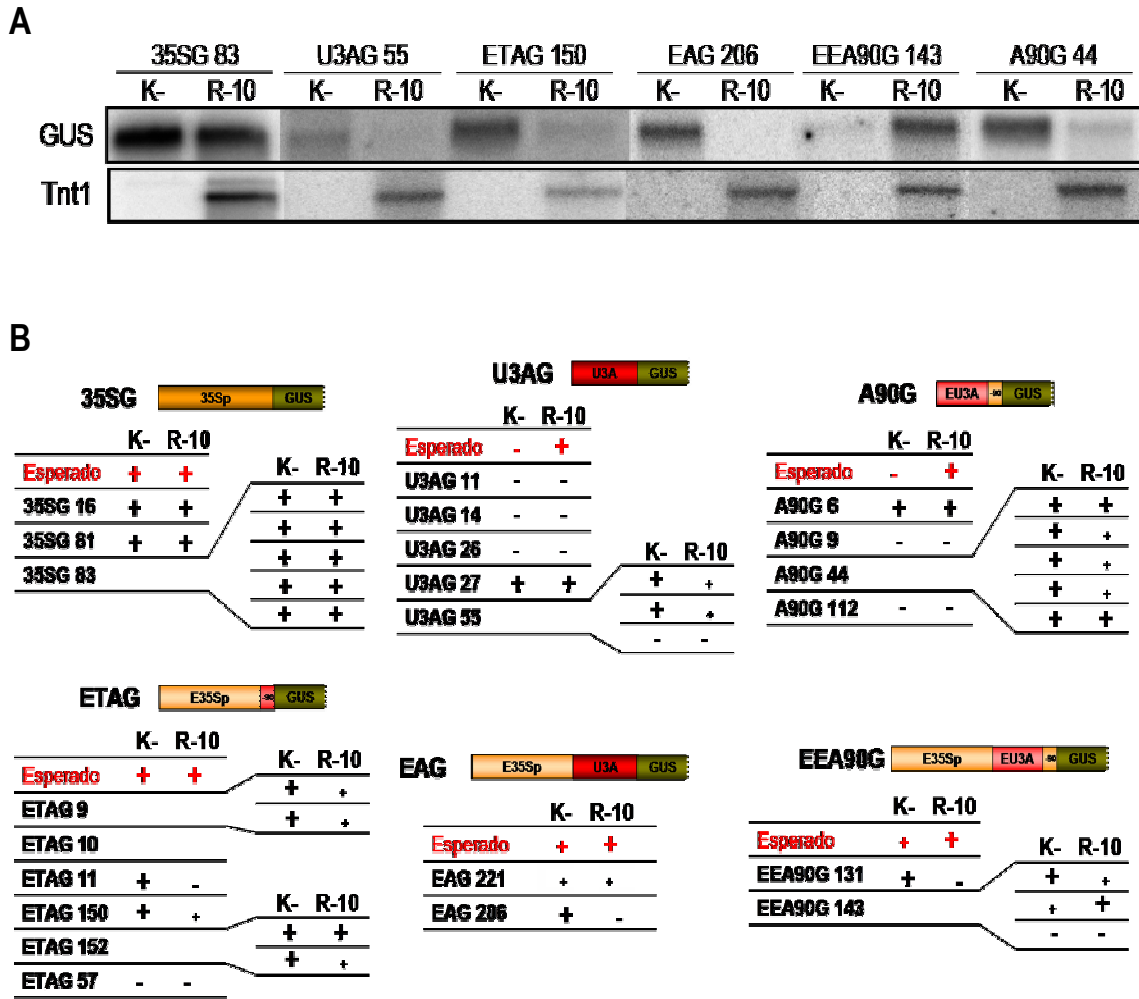
### **2.3.3. Análisis de la expresión de las construcciones quiméricas.**

Todos los estudios de la expresión de las diferentes construcciones génicas se han realizado mediante la técnica de hibridación de Northern. En primer lugar, analizamos la expresión del gen GUS en hoja de varias plantas de distintas líneas de cada transgén, comparando la situación control (una hoja normal) con la situación de estrés (hoja rayada e incubada con R-10 durante 6 horas, como se explica anteriormente en esta memoria). Algunos de estos resultados se muestran en la figura 32A y en la figura 32B se representa un resumen de los resultados obtenidos para todas las construcciones. El patrón de expresión de la mayoría de las construcciones génicas transformadas no es el que esperábamos según lo que se conoce acerca de las LTRs de Tnt1 y del promotor 35S. Se han analizado muchas plantas y se han hecho varias repeticiones para varias de ellas. En algunos casos se han obtenido patrones de expresión diferentes entre distintas repeticiones del experimento.

Todas las líneas de plantas transgénicas 35SG expresan el transgén GUS según lo esperado. El promotor 35S del CaMV promueve la transcripción constitutiva de estas secuencias y detectamos el RNA de GUS en todos los casos analizados.

En cambio, las plantas U3AG deberían expresar el transgén GUS solamente en condiciones de estrés, cuando se expresa Tnt1, ya que poseen la región U3 de los elementos Tnt1A colocada como promotor. Sin embargo, en ninguna de las plantas analizadas se activa la expresión del gen GUS en condiciones de estrés. Además, en varias de ellas, contrariamente a lo esperado, detectamos transcrito de GUS en situación control (ver figura 32B).





**Figura 32.** Representación del resultado del análisis por hibridación de Northern de la expresión del transgén GUS en hoja de algunas de las plantas transgénicas generadas. (A) Hibridación de la planta más representativa de los resultados obtenidos para cada una de las construcciones génicas analizadas. K-: hoja control. R-10: hoja estresada. Se muestra la inducción de Tnt1 como control de la activación de la expresión de este retrotransposón. (B) Compilación de los resultados obtenidos para cada una de las plantas analizadas. En rojo se indican los resultados esperados para cada una de las construcciones transformadas. Se han analizado plantas de líneas independientes y, en algunos casos, plantas independientes de la misma línea. +: expresión del gen GUS. -: no expresión. Aquellos casos en los que la expresión del gen GUS es diferente entre las dos situaciones analizadas, control (K-) y estrés (R-10), se han utilizado signos + de distinto tamaño, en concordancia con los niveles de expresión.

En las plantas transformadas con la construcción A90G esperábamos observar un patrón similar de expresión al que teóricamente debían mostrar las plantas U3AG. La construcción A90G contiene los elementos responsables de la inducción de la expresión de Tnt1 en condiciones de estrés en su región U3 y tan sólo la región flanqueante a la caja TATA del promotor 35S, que por sí sola no es capaz de promover la transcripción (Fang RX *et al.*, 1989). Entre todas las plantas

analizadas no hemos encontrado ninguna que se exprese de la misma forma que lo hace el retrotransposón Tnt1. Además, la mayoría expresan GUS en situación control. Puede ocurrir que la combinación de la región 35Sp-90 con las secuencias activadoras contenidas en el fragmento EU3A, sea suficiente para promover la transcripción del transgén en hoja.

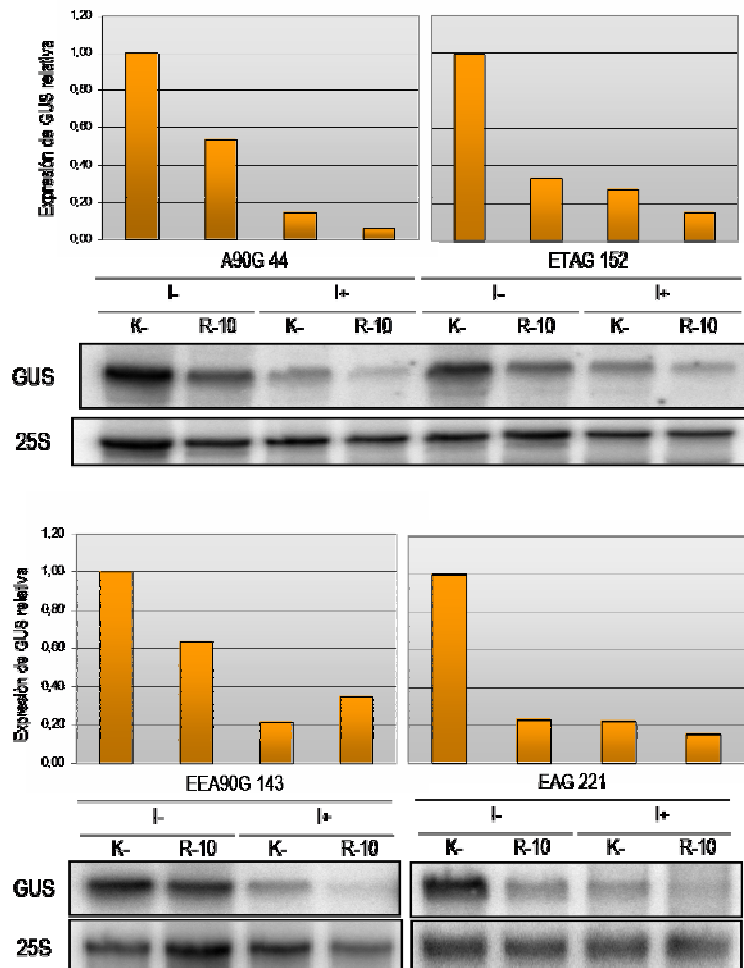
La construcción ETAG es la complementaria a la anterior, ya que posee los elementos activadores del promotor 35S, además de la caja TATA de Tnt1 (región LTR-90). Por tanto, debería expresarse constitutivamente. En algunas de las plantas transformadas con este transgén hemos detectado transcrito de GUS en las dos situaciones analizadas, pero, a diferencia de lo que esperábamos, los niveles de este RNA en condiciones de estrés son mucho menores a los del control.

La plantas portadoras del transgén EAG, poseen los elementos activadores del promotor 35S, además de los de la región U3A, junto con la caja TATA de ésta última, dirigiendo la transcripción del gen GUS. Esperábamos, así, una expresión constitutiva del transgén. Pero, en ninguna planta analizada, portadora de esta construcción, hemos detectado transcrito de GUS en condiciones de estrés.

Las plantas transgénicas que contienen la construcción génica EEA90G también deberían expresar el gen marcador GUS constitutivamente. Sin embargo esto no ocurre así y, sorprendentemente, algunas de estas plantas constituyen el único caso en el que parece que hay respuesta de estas secuencias al estrés.

Esta variabilidad y discordancia de los resultados obtenidos es difícil de explicar y supone un problema para llevar a cabo los análisis sobre el silenciamiento que pretendíamos realizar a continuación. La naturaleza del transgén que estamos utilizando influye, probablemente, sobre su expresión. En este análisis estamos introduciendo parte de una secuencia (la LTR de Tnt1) que está repetida en cientos de copias dentro del genoma de *Nicotiana tabacum*, siendo claramente susceptible a ser diana del silenciamiento génico. Esto podría explicar el hecho de que no detectemos transcrito de GUS, controlado por la región U3 de las LTRs, en condiciones de estrés, en varias de las líneas transgénicas generadas. Es posible que estas construcciones génicas estén siendo silenciadas y, además, parecería que este silenciamiento es más fuerte en las condiciones en las que se expresan las copias endógenas de Tnt1. Sin embargo, al analizar de qué manera está siendo controlado epigenéticamente Tnt1, no hemos detectado la existencia de siRNAs dirigidos contra la región U3. El silenciamiento de estos transgenes debe explicarse de otra manera, siendo necesarios nuevos análisis para determinar la causa de su inactivación.

No obstante, decidimos llevar a cabo los experimentos de silenciamiento sobre algunas de las líneas de plantas transgénicas generadas (aquellas que expresaban el transgén de forma lo más semejante a lo esperado). Infiltramos una hoja de cada una de las distintas plantas transgénicas con *Agrobacterium* conteniendo la construcción génica hp35S. A los cinco días de la agroinfiltración, congelamos la mitad de la hoja y estresamos la otra mitad (como se describe en el apartado 2.1.1). Posteriormente comparamos la expresión del gen GUS entre hojas no infiltradas e infiltradas y entre mitades no estresadas y estresadas. Los resultados de estos experimentos se representan en la figura 33. En aquellos casos en los que GUS se expresaba en hoja vemos que, cuando se introduce el transgén hp35S, disminuyen drásticamente los niveles del transcrito. Parece ser que el silenciamiento del promotor 35S es efectivo, como se ha descrito antes en este capítulo, y, además, que éste no se recupera en ningún caso.



**Figura 33.** Análisis del efecto de la secuencia de las LTRs de Tnt1 sobre el TGS del promotor 35S. Se representa el análisis por hibridación de Northern de la línea más representativa de cada una de las construcciones génicas que contienen promotores quiméricos LTR/35S. I-: hoja no agroinfiltrada. I+: hoja agroinfiltrada con el transgén hp35S. K-: mitad no estresada de la hoja. R-10: mitad estresada de la hoja.

Los resultados de estos análisis indican que las regiones de la LTR que hemos utilizado en estas construcciones no son suficientes para superar el silenciamiento transcripcional dirigido contra la secuencia del promotor 35S y que, por el contrario, son dianas de silenciamiento, en particular bajo condiciones de estrés.

En conclusión, es posible que, debido al elevado número de copias de las LTRs existente en el genoma de tabaco, las secuencias colocadas en los transgenes estén siendo silenciadas. Suponemos que otras secuencias de la LTR, que no hemos utilizado, o de la región interna del transposón pueden ser las responsables de que pueda darse la expresión de Tnt1 en condiciones de estrés. Es necesario realizar otros análisis, utilizando la secuencia de la LTR entera o bien la región U5, para poder confirmar esta hipótesis. Del mismo modo, estudios de este tipo en sistemas heterólogos, en los que no exista un fondo de este tipo de secuencias en el genoma, podrían ayudar a conocer si las LTRs de Tnt1 ejercen alguna influencia sobre el silenciamiento transcripcional de otros promotores (como el caso aquí analizado) y si es así, saber a que nivel afectan al silenciamiento y qué regiones son las implicadas en este proceso.

# DISCUSIÓN GENERAL.

*“De la misma manera, un discípulo nunca puede imitar los pasos de su guía. Porque cada uno tiene una forma de ver la vida, de convivir con las dificultades y con las conquistas. Enseñar es mostrar qué es posible. Aprender es volver posible a sí mismo”.*

**Paulo Coelho.** *El Peregrino de Compostela (Diario de un mago).*

## **DISCUSIÓN GENERAL.**

Los elementos transponibles forman parte de los genomas desde tiempos evolutivos ancestrales. Ésta es la conclusión resultante del análisis de una gran variedad de genomas, pertenecientes a todos los grupos filogenéticos en los que se agrupan a los seres vivos. En todos ellos se ha encontrado algún tipo de secuencia móvil, si bien todos los tipos de transposones no están presentes en todos los organismos y su abundancia es, asimismo, desigual.

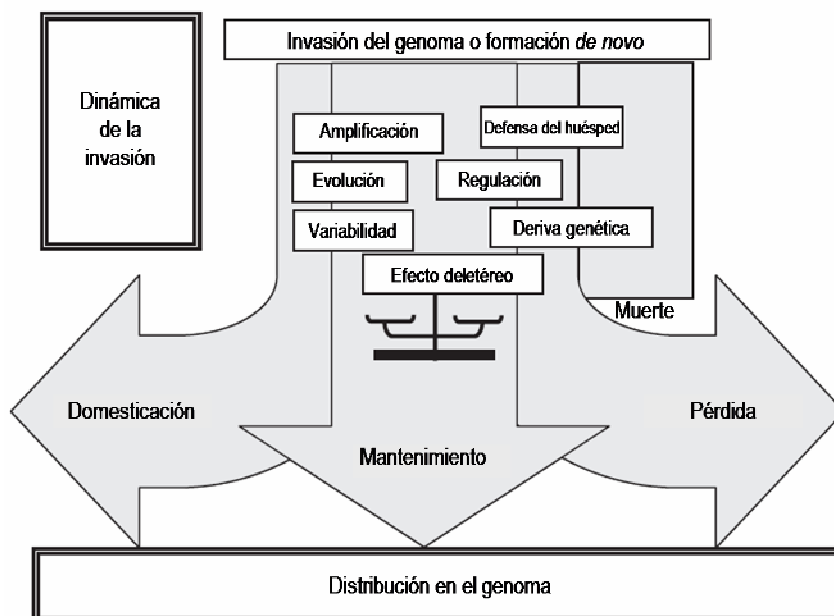
Los elementos transponibles son un conjunto de secuencias cuya característica común es su capacidad de movilizarse dentro de los genomas. Sin embargo, existe una gran variabilidad estructural y de modo de transposición entre los ETs. Estas diferencias han permitido, probablemente, que estos elementos persistan en todos los genomas, habiendo encontrado la forma de adaptarse o mantenerse en los distintos organismos.

La transposición es un mecanismo altamente mutágeno, que puede afectar gravemente a la expresión de los genes. Por ello, es un proceso que debe estar fuertemente controlado por los genomas. En el caso en el que los transposones consiguieran evadir los mecanismos de control y pudieran movilizarse libremente dentro del genoma, podrían alterar y afectar gravemente la expresión génica y estructura del genoma y su efecto sería tan deletéreo que el organismo sería inviable. En esta situación, el transposón y sus nuevas copias desaparecerían evolutivamente, al no transmitirse a una la descendencia, que sería inexistente. En el caso contrario, en el que los mecanismos de control de la transposición fueran extraordinariamente efectivos, los elementos inactivados también se perderían a lo largo de la evolución. De forma que, la persistencia o mantenimiento de un determinado tipo de elemento transponible en un genoma concreto es el resultado del equilibrio establecido, básicamente, entre su efectividad a la hora de movilizarse y el control ejercido por el genoma hospedador para evitar la transposición y conseguir la inactivación y eliminación del transposón.

A lo largo de la evolución, este equilibrio ha dado lugar a distintas situaciones. Como se explica en la introducción de este trabajo, los transposones han participado activamente en la evolución de los genomas. Se conocen varios genes que derivan de antiguos elementos transponibles, otros transposones llevan a cabo funciones celulares esenciales para el mantenimiento y transmisión del material genético, mientras que otros, en cambio, se han mantenido en el genoma como secuencias móviles que parecen responder, únicamente, a su propia supervivencia sin intervenir, al menos directamente, en las funciones celulares.

## Discusión general.

En la figura 1 se esquematiza el posible destino de los ETs en los genomas, como resultado de la combinación de diversos factores.



**Figura 1.** Una vez en el genoma, el destino de un determinado transposón está determinado por el balance establecido entre diferentes factores, resultando en la domesticación, mantenimiento o pérdida del transposón, si sus efectos deletéreos no han llevado a la muerte del organismo hospedador. Adaptado de Hua-Van A *et al.*, 2005.

El trabajo presentado en esta memoria se ha centrado en el estudio del control de la transposición y de la evolución de los elementos transponibles en el genoma, que pueden ser indicativo del impacto de los transposones sobre los genes y genomas. Como modelo, se han analizado dos transposones, el retrotransposón Tnt1 de tabaco y el MITE *Emigrant* de *Arabidopsis*, que son representativos de los grupos de ETs más abundantes en los genomas vegetales y que presentan estrategias evolutivas distintas. Mientras que Tnt1 limitaría su impacto expresándose únicamente en condiciones de estrés, *Emigrant* lo haría al haber reducido su tamaño, como consecuencia de la separación las secuencias portadoras de la actividad de la secuencia que se moviliza.

## 1. EL SILENCIAMIENTO GÉNICO COMO CONTROL PRINCIPAL DE LA TRANSPOSICIÓN.

A pesar de que la transposición puede estar regulada a distintos niveles, por ejemplo, a nivel de la terminación de la transcripción y poliadenilación (que también hemos analizado en este trabajo), el silenciamiento génico es el principal mecanismo de control de la proliferación de los elementos transponibles.

Uno de los principales determinantes para que una secuencia determinada sea silenciada es el que se encuentre en un elevado número de copias dentro del genoma. Y éste es el caso de los ETs, como el retrotransposón Tnt1 y el MITE *Emigrant*, que están presente en cientos de copias en el genoma de tabaco y de *Arabidopsis*, respectivamente.

En este trabajo hemos querido conocer cómo está controlada la expresión de Tnt1 y a qué niveles actúa el silenciamiento génico sobre este tipo de secuencias. Los resultados presentados en esta memoria sugieren que, efectivamente, Tnt1 está silenciado en el genoma de tabaco. En concreto, la asociación de las copias de este retrotransposón a cromatina inactiva y la existencia de sRNAs que hibridan con la LTR de Tnt1 parecen indicar que este transposón está sometido a un silenciamiento de tipo transcripcional en hojas de plantas sanas. Por el contrario, este control de la expresión de Tnt1 se relaja cuando se somete la planta a un estrés biótico o abiótico, ya que es capaz de expresarse en estas condiciones (Casacuberta JM and Grandbastien MA, 1993; Pouteau S *et al.*, 1991; Pouteau S *et al.*, 1994; Mhiri C *et al.*, 1997).

A pesar de que se ha descrito que en condiciones de estrés el silenciamiento post-transcripcional de algunos virus de plantas deja de ser efectivo, permitiendo su amplificación (Taliensky M *et al.*, 2004), los resultados que hemos obtenido muestran que el silenciamiento génico transcripcional no se ve comprometido.

Esto sugiere que la expresión de Tnt1 en condiciones de estrés resulta de la relajación o desaparición específica del silenciamiento de este tipo de secuencias. Nuestros resultados indican que la especificidad del proceso podría residir en la secuencia de Tnt1, posiblemente en la región promotora contenida en las LTRs. Los resultados presentados en esta memoria, sugieren que la transcripción de Tnt1 está asociada a modificaciones de la cromatina. Podemos relacionar la disminución de los sRNAs que hibridan con la secuencia de la región U5 con las

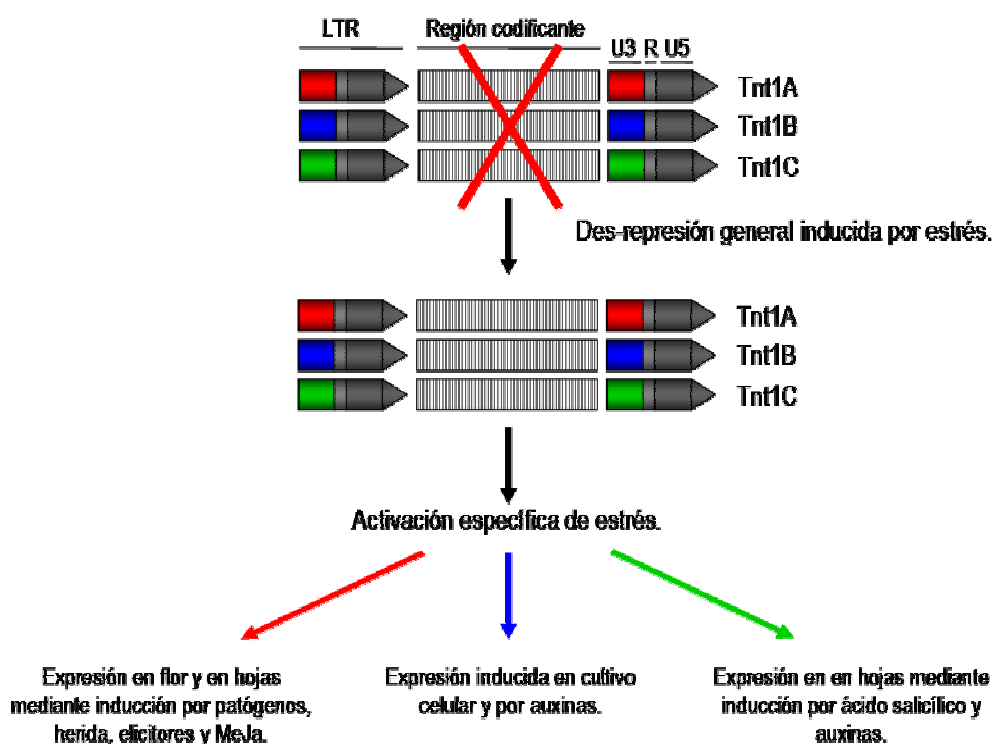


## Discusión general.

modificaciones que hacen que la cromatina asociada a las LTRs de Tnt1 adquiera características de cromatina activa y, finalmente, con la transcripción del retrotransposón.

Las modificaciones de la cromatina asociadas al estrés con R-10, parecen tener lugar sobre secuencias pertenecientes a las distintas familias de Tnt1, aunque en estas condiciones sólo se expresan las copias de Tnt1A, que son las que tienen un promotor que responde a este tipo de inducción.

Estos resultados sugieren que la expresión del retrotransposón Tnt1 estaría controlada a dos niveles. En condiciones de estrés, en primer lugar, desaparecería de forma general la represión transcripcional de todas las copias de Tnt1 y, entonces, se expresarían las copias asociadas al tipo de estrés que esté sufriendo la planta, dependiendo de la región U3 (distinta entre las diferentes familias de Tnt1) que contengan en su secuencia.



**Figura 2.** Modelo para la activación y expresión de los distintos elementos Tnt1 de tabaco.

Los transposones del tipo MITE parecen haber desarrollado otro mecanismo para minimizar su efecto sobre los genomas. La gran mayoría de ellos son secuencias pequeñas sin capacidad codificante. Se ha propuesto que las actividades encargadas de su transposición se encuentran codificadas en uno o pocos elementos relacionados. De hecho, en *Arabidopsis* se ha identificado

una única secuencia, *Lem1* (Feschotte C and Mouches C, 2000), que podría ser la responsable de la reciente movilización de la familia *Emigrant* de MITEs, descrita en esta memoria. De hecho, en nuestro laboratorio se ha demostrado que el producto codificado por este elemento, relacionado con la transposasa de la familia Tc1/*mariner*, es capaz de unirse a los TIRs de los MITEs *Emigrant* (Loot C and Casacuberta JM, no publicado). Al tratarse de una única copia, este posible elemento autónomo no estaría siendo silenciado, lo que permitiría su expresión y la movilización de los elementos *Emigrant*. Sin embargo, estos MITEs, presentes en un elevado número de copias en el genoma de *Arabidopsis* seguramente son diana de los mecanismos de silenciamiento y se probablemente encuentran metilados. La metilación de estos elementos puede estar favoreciendo, a su vez, la mutación rápida de su secuencia, que terminaría por no ser reconocida por la maquinaria encargada de su movilización. Esto podría explicar el hecho de que, como indica la forma en estrella de los árboles filogenéticos de las distintas familias de *Emigrant* generados, solamente una o pocas copias hayan actuado como secuencias fundadoras en los distintos sucesos de amplificación.

## **2. IMPACTO DE LOS ELEMENTOS TRANSPONIBLES SOBRE GENES Y GENOMAS.**

Para conocer cómo se desarrolla la interacción real que tiene lugar entre genomas y transposones, es necesario estudiar la dinámica de ciertos ETs en un genoma concreto a lo largo de la evolución. En este trabajo se ha llevado a cabo un análisis de este tipo sobre la familia *Emigrant* de MITEs en el genoma de *Arabidopsis*.

El estudio de la asociación de las distintas copias de los MITEs *Emigrant*, de diferente edad, con las secuencias génicas, nos proporciona una visión evolutiva de la dinámica que han seguido estas secuencias y el impacto que han tenido dentro del genoma.

El resultado del trabajo presentado en esta memoria muestra que existe una asociación de los MITEs *Emigrant* con los genes. Así como las copias más jóvenes en el genoma se localizan principalmente lejos de las secuencias génicas, las copias más antiguas aparecen, en mayor frecuencia, cerca o dentro de genes. Parece que, a lo largo de la evolución, se han mantenido los elementos que han transpuesto en secuencias génicas y se han ido eliminando las otras

## Discusión general.

copias. Esta acumulación, con el tiempo, de los MITEs *Emigrant* en secuencias génicas sugiere que los elementos insertados cerca de los genes se han seleccionado positivamente a lo largo de la evolución de *Arabidopsis*.

Pero esta variación en la distribución de las copias antiguas de estos transposones respecto a las jóvenes puede deberse a que los elementos situados en regiones pobres en genes están siendo eliminados más eficazmente. Uno de los mecanismos por los que puede aclararse el número de ETs del genoma es la recombinación entre distintas copias de un mismo transposón. Este proceso está mucho menos tolerado en regiones ricas en genes, ya que puede resultar en deleciones o recombinaciones desiguales que afectarían negativamente a los genes. Ésta es una de las hipótesis propuestas para explicar la acumulación de secuencias *Alu* en regiones ricas en GC (Medstrand P *et al.*, 2002).

En cualquier caso, los elementos *Emigrant* asociados a secuencias génicas se han mantenido a lo largo de la evolución y deben haber influido seguramente en la evolución de los genes en *Arabidopsis*.

Entre todas las secuencias analizadas, hemos encontrado algunos *Emigrant* que se localizan en la región 5' de los genes, otros en la 3' no traducida e incluso algunos dentro de la secuencia transcrita. De esta forma, pueden estar alterando la regulación de los genes al modificar sus regiones promotoras o terminadoras o el propio producto génico. Puede ser que este tipo de elementos estén suponiendo algún beneficio a los genes junto a los que se insertan, como la aportación de nuevos elementos reguladores o activadores. También puede ocurrir que la transposición de copias situadas junto a genes ayude a distribuir ciertos elementos reguladores o activadores por el genoma.

Pero en el caso concreto de *Arabidopsis*, cuyo genoma es muy compacto y las regiones intergénicas son muy cortas, los MITEs podrían estar desarrollando un papel importante al servir de aisladores génicos. De hecho, se ha observado una asociación entre MITEs, en maíz y sorgo, y regiones de anclaje a la matriz nuclear (MARs) (Tikhonov AP *et al.*, 2000). Asimismo, experimentos con mutantes de metilación en *Arabidopsis* sugieren que los transposones y en particular los MITEs podrían ser puntos de anclaje de lazos de DNA eucromáticos (Franz P *et al.*, 2003). La presencia de MITEs asociados a los genes podría ayudar a crear dominios génicos independientes, en los que la expresión de un determinado gen no interfiera con la de otros genes adyacentes, relativamente cercanos.

## Discusión general.

En el caso del retrotransposón Tnt1 no se tiene conocimiento del impacto que su transposición ha tenido sobre los genes. Sin embargo, se ha descrito una preferencia de inserción por zonas ricas en genes (Melayah D *et al.*, 2001), por lo que es muy probable que muchas copias de Tnt1 se encuentren asociadas a secuencias génicas. Por ello, es posible que su activación por estrés afecte a los genes vecinos, ya sea por la transcripción de los mismos en sentido directo o inverso a partir de estos retrotransposones o bien porque las modificaciones de la heterocromatina que hemos descrito asociadas a estas secuencias pudieran propagarse a las regiones vecinas, permitiendo la activación de los genes adyacentes. En un futuro, sería interesante analizar los genes localizados junto a las diferentes copias de Tnt1 para conocer de qué manera se encuentran afectados por este transposón y, en particular, ver si se expresan asociados a la inducción del retrotransposón.

## **CONCLUSIONES.**

## CONCLUSIONES.

Los resultados obtenidos del trabajo que se ha presentado en esta memoria permiten extraer las siguientes conclusiones:

1. El programa TRANSPO es una buena herramienta para detectar, dentro de un genoma, copias de una determinada familia de MITEs, incluso aquellas más divergentes.
2. En el genoma de *Arabidopsis* existen, al menos, 151 elementos pertenecientes a la familia *Emigrant* de MITEs.
3. Los distintos elementos *Emigrant* se pueden clasificar en diferentes grupos y subfamilias. Estas subfamilias presentan distinta variabilidad, indicando que tienen una edad evolutiva diferente. Por tanto, han tenido lugar varios eventos de amplificación de MITEs pertenecientes a la familia *Emigrant* a lo largo de la historia evolutiva de *Arabidopsis*. Las distintas subfamilias se han originado a partir de la amplificación de uno o pocos elementos fundadores.
4. Algunas inserciones de elementos *Emigrant* no están fijadas en las poblaciones naturales de *Arabidopsis*, indicando que ha existido una actividad reciente de esta familia de MITEs en el genoma de *Arabidopsis*.
5. Los elementos *Emigrant* más antiguos se encuentran localizados cerca o dentro de secuencias génicas de *Arabidopsis* en mayor frecuencia que las copias más recientes, que se localizan preferentemente apartados de los genes. Esto indica que los elementos *Emigrant* insertados junto a secuencias génicas se han mantenido preferentemente a lo largo de la evolución de *Arabidopsis*.
6. En las hojas de tabaco heridas e incubadas con la celulasa R-10 de Onozuka se expresan mayoritariamente elementos pertenecientes a la subfamilia Tnt1A y los transcritos generados terminan y se poliadenilan en un único sitio

## Conclusiones.

7. La eficiencia de la terminación de la transcripción y la incorporación de la cola poliA de los mRNAs de Tnt1 se ven afectadas por el tipo de molécula que se utiliza para inducir la expresión de este retrotransposón.
8. Existe una influencia de la región U3 promotora del retrotransposón Tnt1 sobre la terminación y poliadenilación de los transcritos de este elemento.
9. El transgén hp35S promueve el silenciamiento transcripcional del promotor 35S en plantas transgénicas de *Nicotiana tabacum* 35S::GUS y en plantas transgénicas de *Nicotiana benthamiana* 35S::GFP en experimentos de transformación transitoria.
10. El silenciamiento génico transcripcional no se ve afectado por el estrés. En condiciones de estrés no se recupera la expresión del transgén 35S::GUS que ha sido silenciado transcripcionalmente por el locus 271. Tampoco se recupera esta expresión tras haber sido silenciado por el transgén hp35S.
11. En hoja sana de tabaco existen sRNAs que hibridan fuertemente con las regiones U5 de la LTR y la región terminal del gen de la RT de Tnt1. Los niveles de estos sRNAs disminuyen en hoja estresada, cuando se expresa el retrotransposón Tnt1.
12. En condiciones de estrés inducido por R-10, cuando Tnt1 se expresa, la cromatina asociada a las regiones promotoras de Tnt1 adquiere características de cromatina activa. Mientras que las modificaciones de la cromatina se dan sobre elementos pertenecientes a las distintas familias de Tnt1, solamente se expresan las copias pertenecientes a la familia Tnt1A.
13. Transgenes que contienen la secuencia completa o parcial de la región U3 de la LTR en su región promotora son silenciados en plantas de tabaco, particularmente, en las condiciones en las que se expresan las copias endógenas del retrotransposón Tnt1.

# **MATERIAL Y MÉTODOS.**



# MATERIAL Y MÉTODOS.

## 1. MATERIALES.

### 1.1. MATERIAL VEGETAL.

Las plantas de *Arabidopsis thaliana* germinaron y crecieron en maceta en el invernadero en las siguientes condiciones: 18-22°C de temperatura; 45% de humedad; 16h luz / 8h oscuridad. Para el análisis por PCR se extrajo DNA genómico de plantas de 3 semanas de edad. En este trabajo se han utilizado los siguientes ecotipos de *Arabidopsis thaliana* sp.: Col-0 (Columbia), La-0 (Landsberg), RLD1 (Netherlands), WS (Wassilewskija), Cnt-1 (Canterbury), Co-4 (Coimbra), Dijon G, Est-0 (Estland), Ge-0 (Geneve), Kas-1 (Kashmir), Ms-0 (Moscow), Nd-1 (Niederzenz), Tsu-0 (Tsu), No-0 (Nossen).

Las plantas de *Nicotiana tabacum* var. *Xanthi* y de *Nicotiana benthamiana* han germinado y crecido en maceta en el invernadero en las siguientes condiciones: 22-26°C de temperatura; 60% de humedad; 14-16h luz / 8-10h oscuridad, dependiendo de la estación del año. Las plantas de *N. tabacum* var. *Xanthi* germinadas y crecidas *in vitro* lo hicieron en las siguientes condiciones: 24-26°C de temperatura; 16h luz / 8h oscuridad; no hay control de humedad ya que el material se encuentra confinado en placa, tubo o jarra de cultivo, donde se mantiene una elevada humedad.

#### 1.1.1. ESTERILIZACIÓN Y GERMINACIÓN DE SEMILLAS.

Las semillas germinadas en placa *in vitro* necesitan una esterilización previa, para evitar contaminación por hongos. Las semillas germinadas en tierra en el invernadero no necesitan esta esterilización previa. Se ha seguido el método de esterilización en líquido, por el cual se sumergen las semillas en una solución al 10% (v/v) de NaClO y 0,1% (v/v) Tween-20® durante 5-10 min y, posteriormente, se aclaran 3-4 veces con H<sub>2</sub>O estéril.

Las semillas germinadas *in vitro* se han sembrado en placas de medio MS/gelrita (ver apartado 1.4.). Las semillas transgénicas se han sembrado en placas de medio MS/gelrita más el antibiótico correspondiente para su selección (ver apartado 1.5. para conocer la concentración final necesaria para la selección en planta). Para conseguir una óptima germinación de las semillas de *Arabidopsis*, una vez sembradas en placa o en tierra, se les ha sometido a un periodo de vernalización (48-72 h a 4°C en oscuridad) suficiente para romper su estadio de dormancia. También se ha sometido a un periodo de vernalización a las semillas de tabaco sembradas en placa, para conseguir una germinación sincrónica de las mismas. Las semillas de tabaco sembradas en tierra no necesitan un periodo de vernalización para que consigan germinar.

### 1.2. MATERIAL BACTERIANO.

***Escherichia coli* K-12 DH5 $\alpha$** : *supE44*,  $\Delta$ *lacU169* ( $\Phi$ 80 *lacZ*  $\Delta$ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*. (Hanahan D, 1983). Esta cepa se ha utilizado para la amplificación del DNA plasmídico.

***Agrobacterium tumefaciens* LBA4404:** Ach5 pTiAch5 (agrocinopina, tipo octopina); Sm/Sp(R) en el plásmido de virulencia. Se ha eliminado todo el T-DNA del pTiAch5, creando el plásmido de virulencia pAL4404 (Hoekema A *et al.*, 1983). Se ha utilizado esta cepa para la transformación de *N.tabacum*.

### 1.3. PLÁSMIDOS.

#### 1.3.1. VECTORES UTILIZADOS PARA REALIZAR LAS CONSTRUCCIONES GÉNICAS.

**pBluescript® II SK (+/-) y KS (+/-).** Stratagene. Fagémidos derivados del plásmido pUC19 (número de acceso de GenBank: L09137). Estos vectores se han utilizado para la clonación de fragmentos de DNA.

**pBI221.** Clontech. Plásmido derivado del pUC19, al que se le ha introducido el promotor 35S de CaMV, la región codificante del gen GUS y el terminador NOS. Número de acceso de GenBank: AF502128.

**pBSX1.** Plásmido derivado del pBluescript® II KS+, al que se le ha introducido en la diana *Xho*I la secuencia completa del retrotransposón Tnt 1-94 (número de acceso de GenBank: X13777). (Grandbastien MA *et al.*, 1989).

**pHANNIBAL.** Vector utilizado para la generación de RNA de doble cadena. Número de acceso de GenBank: AJ311872. (Wesley SV *et al.*, 2001).

**pUC19.** Plásmido utilizado como vector de clonaje original para la mayoría de las construcciones génicas generadas en este trabajo. Número de acceso de GenBank: X02514. (Yanisch-Perron C *et al.*, 1985).

#### 1.3.2. VECTORES BINARIOS.

**pCAMBIA 1300.** CAMBIA™. Vector binario utilizado en la transformación de plantas mediada por *Agrobacterium*. Número de acceso de GenBank: AF234296.

**pCAMBIA 5300.** Plásmido derivado del pCAMBIA 1300, al que se le ha introducido el intrón de la catalasa de ricino dentro del gen *hptII* (resistencia a higromicina). La presencia del intrón de la catalasa dentro del gen *hptII* confiere a la planta transgénica mayor resistencia a la higromicina y permite una mayor eficiencia en la eliminación del *Agrobacterium* durante la regeneración de las plantas transgénicas. Este vector binario se ha utilizado en la transformación de plantas mediada por *Agrobacterium*..

**pBIN 19.** Vector binario utilizado en la transformación de plantas mediada por *Agrobacterium*. Número de acceso de GenBank: U12540. (Bevan M, 1984).

#### 1.3.3. VECTORES UTILIZADOS PARA LA CLONACIÓN DE PRODUCTOS DE PCR.

**pTZ57R/T** (Fermentas); **pGEM®-T Easy** (Promega); **pCR®II** y **pCR®2.1** (Invitrogen).

## 1.4. MEDIOS.

**LB** (Luria-Bertani) (Sambrook J and Russell DW, 2001): 1% (w/v) bacto-triptona; 0,5% (w/v) extracto de levadura; 1% (w/v) NaCl. (1,5% (w/v) agar bacteriológico). Ajustar a pH 7,5 con NaOH. Autoclavar a 115°C durante 20 min.

**MS.** El medio MS utilizado en este trabajo es una complementación del medio de Murashige-Skoog (Murashige T and Skoog F, 1962): 0,44% (w/v) MS+Vitaminas (*Duchefa*); 1% (w/v) Sacarosa; 0,05% (w/v) MES. (0,2% (w/v) gelrita ó 0,8% (w/v) agar). Ajustar a pH 5,8 con KOH. Autoclavar.

**SOB** (Sambrook J and Russell DW, 2001): 2% (w/v) bacto-triptona; 0,5% (w/v) extracto de levadura; 10 mM NaCl; 2,5 mM KCl; 10 mM MgSO<sub>4</sub>; 10 mM MgCl<sub>2</sub>. Ajustar a pH 7,0 con NaOH. Autoclavar.

**T<sub>0</sub>.** El medio T<sub>0</sub> preparado en este trabajo es una variante simplificada del descrito por Bourgin para el cultivo de protoplastos de tabaco (Bourgin *et al.*, 1979): 0,44% (w/v) MS+Vitaminas; 0,07% (w/v) MES; 2% (w/v) sacarosa; 8% (w/v) manitol; 8 · 10<sup>-4</sup>% (v/v) Tween-80®; 1 µg/ml NAA; 0,2 µg/ml BAP. (Las hormonas NAA y BAP se disuelven en DMSO).

**YEB** (Yeast Extract Broth) (Van Larebeke *et al.*, 1977): 0,5% (w/v) extracto de buey; 0,1% (w/v) extracto de levadura; 0,5% (w/v) peptona bacteriológica; 0,5% (w/v) sacarosa; 0,048% (w/v) MgSO<sub>4</sub>. (1,5% (w/v) agar bacteriológico). Ajustar a pH 7,2 con NaOH / HCl. Autoclavar a 115°C durante 20 min.

## 1.5. ANTIBIÓTICOS Y REACTIVOS.

Antibiótico o reactivo	Abreviación	Concentración final (µg/ml)			Solvente
		<i>E.coli</i>	<i>A.tumefaciens</i>	planta	
Carbenicilina	Carb	100	100	-	H <sub>2</sub> O
Kanamicina	Kan	25	50	50	H <sub>2</sub> O
Rifampicina	Rif	-	100	-	DMSO
Estreptomicina	Sm	20	300	-	H <sub>2</sub> O
Higromicina B	Hyg	-	-	25	-
Ticarcilina / ácido clavulánico	timentin	-	100	100	H <sub>2</sub> O
Isopropyl-β-D-thiogalactopyranoside	IPTG	23,8	-	-	H <sub>2</sub> O
5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside	X-Gal	40	-	-	N-N'-dimetil formamida

## 1.6. OLIGONUCLEÓTIDOS.

Oligo	Secuencia	T <sub>m</sub>
<b>35S</b>	5' CGTAAAGACTGGCGAAC 3'	52°C
<b>2x35S</b>	5' CTTGCAAGACCCTTCC 3'	54°C
<b>GUS</b>	5' GTCGAGTTTTTTGATTTACAGGG 3'	56°C
<b>GUS3'</b>	5' CCTCGCAAGGCATATTGC 3'	56°C
<b>M13-20 (DIRECT)</b>	5' GTAAAACGACGGCCAGT 3'	52°C
<b>M13 REVERSE</b>	5' CAGGAAACAGCTATGAC 3'	50°C
<b>Tocs</b>	5' TATCATGCGATCATAGGCGTC 3'	55°C

Material y métodos.

<b>Oligo</b>	<b>Secuencia</b>	<b>T<sub>m</sub></b>
ACTINA F	5' CGGAAAGAGGTTTATTCATTAC 3'	64°C
ACTINA R	5' GGTGCTGAGAGAAGCTAAGATAG 3'	68°C
AVI	5' TATGCTGACCAAGGTGGTAC 3'	60°C
13	5' TTATACCTTGTCTGTGAAACC 3'	58°C
TRU	5' GGATTCGAGAAATGGTAGATG 3'	60°C
TRT	5' GATGTGAGATATCATTGGATTG 3'	64°C
OligoA1	5' GATGGCACCCTGCGTTTGCTGGC 3'	50°C
OligoA2	5' GATGGTGCCACTCTTGACG 3'	50°C
Oligo(dT) <sup>1</sup>	5' CCTAGGCCACTGTGGCCTTTTTTTTTTTTTT 3'	62°C
U3 1	5' TGATGATGTCCATCTC 3'	46°C
U3 2	5' TGTGAAGAAATGAGAG 3'	44°C
U5 1	5' ACACCAACAAAGAGAG 3'	46°C
U5 2	5' GATACCAGTTGTTGGG 3'	48°C
GAG 1	5' GTAGTACTATACTTGG 3'	44°C
GAG 2	5' AGCTGTGTGATTAGTC 3'	46°C
GAG 3	5' ACACAGCTTGCCAACC 3'	50°C
GAG 4	5' ATTTTGCACCATGGCG 3'	48°C
PR 1	5' AATGATAATGTTGTCCTC 3'	48°C
PR 2	5' GAAATCTCATCTTGTGC 3'	46°C
EN 1	5' ATTTCTGTAGATTTATGG 3'	46°C
EN 2	5' CTTACCTTTTCTGAC 3'	46°C
RT 1	5' GAAGAATGGTATAATTCC 3'	48°C
RT 2	5' GCTAGCTGCTAAGCTC 3'	50°C
RT 3	5' CCTAGATCTTGAAGTG 3'	46°C
RT 4	5' GCGTGAGCAATATCAG 3'	48°C
RT 5	5' CAGTTGGTGTGTCAG 3'	48°C
RT 6	5' TCTCATTCATCCAGAG 3'	46°C
e1 F	5' CTGGAATGAAGGAAACTTC 3'	58°C
e1 R	5' CAAGTCCTTGCTTTGACTATC 3'	60°C
e3 F	5' GGAGATAATTTCAATAGTTTG 3'	54°C
e3 R	5' GGCGATAATAAAGCTAATTAG 3'	56°C
e13 F	5' CGGTCTTAATTTCTGTGAAAC 3'	58°C
e13 R	5' CCTTTTATATGCTAGTATATG 3'	54°C
e43 F	5' GCTTAATTGATAAAGACTATA 3'	52°C
e43 R	5' CACCATTTTAATTATTATTC 3'	54°C
e87 F	5' GGGATCAAATTGATCATTGTA 3'	56°C
e87 R	5' GCCATAATATCTTTGCTTCC 3'	58°C
e97 F	5' CGTGACTTGATACTCTTAAC 3'	56°C
e97 R	5' CCATTAATCAAATCATTAAC 3'	54°C
e107 F	5' CGACGTAATAATAAGATTGG 3'	49°C
e107 R	5' ACTCTATAGAATAATAATCTC 3'	46°C
e116 F	5' CGTCATTTGACAATAATCAC 3'	50°C
e116 R	5' GTCTTTTAAGTATTGAATCTTG 3'	48°C
e126 F	5' CAGCGATAACATTTGATCTTC 3'	58°C
e126 R	5' GCATGTATCTTAAACCATTG 3'	54°C
e128 F	5' GGAGCAAAACATATAAATC 3'	50°C
e128 R	5' GCGGATCTAATTAATTGA 3'	48°C
e130 F	5' GCGTTTGCTTCTAGGTTTAAAC 3'	60°C
e130 R	5' GCCATTAGATCGTTAGAATAC 3'	58°C
e158 F	5' CCATATTCACAATTTTAC 3'	46°C
e158 R	5' GCTTAAATAAATAGAAAGAG 3'	50°C

<sup>1</sup>El Oligo(dT) utilizado contiene una extensión en 5' (Casacuberta JM *et al.*, 1995).

## 2. MÉTODOS GENERALES.

Los métodos que se describen en este apartado hacen referencia a técnicas básicas generales, ampliamente utilizadas en biología molecular. En aquellos casos en los que se han introducido modificaciones respecto al protocolo original, se detalla el protocolo modificado que se ha seguido.

### 2.1. EXTRACCIÓN DE ÁCIDOS NUCLEICOS.

#### 2.1.1. EXTRACCIÓN DE DNA PLASMÍDICO.

La **minipreparación** de DNA plasmídico de *E.coli* se ha realizado mediante el método de la lisis alcalina, basado en el descrito por Birnboim y Dolly (Birnboim HC and Dolly J, 1979) y modificado, posteriormente como se describe en *Molecular Cloning: A Laboratory Manual* (Sambrook J and Russell DW, 2001).

Para obtener mayores cantidades de DNA plasmídico se han realizado **midipreparaciones** utilizando el kit comercial de QIAGEN® (Plasmid Midi Kit), siguiendo las indicaciones del fabricante.

#### 2.1.1.2. Extracción de DNA plasmídico de *A.tumefaciens*.

Debido a que *Agrobacterium* posee pared celular (es una bacteria gram negativa), el protocolo de extracción de DNA plasmídico, arriba descrito, ha sido modificado para facilitar la lisis celular y eliminar residuos de la pared que puedan actuar como inhibidores de distintas reacciones enzimáticas.

Para la extracción de DNA plasmídico de *Agrobacterium* se ha seguido el siguiente protocolo:

1. Centrifugar en un vial de microcentrífuga 1,5 ml de cultivo bacteriano 5 minutos a 10000 rpm. Repetir 3 veces hasta haber centrifugado un total de 4,5 ml de cultivo. Descartar el sobrenadante.
2. Lavar el sedimento bacteriano añadiendo 1 ml de solución STE<sup>1</sup> 1x. Centrifugar 5 min a 10000 rpm. Descartar el sobrenadante.
3. Resuspender el sedimento bacteriano en 100 µl de solución I de lisis alcalina<sup>2</sup> con lisozima<sup>3</sup>. Trabajar con cuidado con tal de no lisar las células.
4. Añadir 200 µl de solución II de lisis alcalina<sup>4</sup> e invertir suavemente con tal de no fragmentar el DNA genómico. Reposar a temperatura ambiente 5 minutos.
5. Añadir 150 µl de solución III de lisis alcalina<sup>5</sup>. Neutraliza solución II y precipita el DNA genómico. Invertir suavemente. Poner en hielo 10 minutos.
6. Centrifugar 10 minutos a máxima potencia. Precipitación de DNA genómico.
7. Inmediatamente pasar el sobrenadante a otro vial de microcentrífuga, con tal de que no se resuspenda el precipitado (DNA genómico, que se descarta).
8. Añadir fenol-cloroformo 1:1 (450 µl aprox.). Mezclar agitando en vórtex.
9. Centrifugar 5 minutos a máxima potencia.
10. Pasar el sobrenadante a otro vial de microcentrífuga. Descartar la fase inferior (fenol).
11. Repetir pasos 8-10.
12. Añadir 1 ml de EtOH 100%. Mezclar agitando en vórtex.

## Material y métodos.

13. Centrifugar 20 minutos.
14. Eliminar sobrenadante y conservar el precipitado, DNA plasmídico.
15. Lavar el precipitado, añadiendo 250 ml de EtOH 70% (lavado) y centrifugando 5 minutos.
16. Eliminar el sobrenadante y dejar secar el precipitado.
17. Resuspender en 30  $\mu$ l de H<sub>2</sub>O + RNasa A<sup>6</sup>. Utilizar 15  $\mu$ l por cada reacción de digestión.

<sup>1</sup>STE 1x: 100 mM NaCl; 10 mM Tris-HCl pH 8,0; 1 mM EDTA pH 8,0.

<sup>2</sup>Solución I de lisis alcalina: 50 mM EDTA pH 8,0; 25 mM Tris-HCl pH 8,0.

<sup>3</sup>Lisozima: 4 mg/ml.

<sup>4</sup>Solución II de lisis alcalina: 0,2 M NaOH; 1% (w/v) SDS.

<sup>5</sup>Solución III de lisis alcalina: 3 M KAc; 0,2 M HAc.

<sup>6</sup>RNasa A: 20  $\mu$ g/ml.

### 2.1.2. EXTRACCIÓN DE DNA GENÓMICO.

La extracción de DNA genómico, tanto de arabis como de tabaco, se ha realizado siguiendo el protocolo descrito a continuación. Se trata de una modificación del método descrito por Dellaporta y colaboradores (Dellaporta SL *et al.*, 1983), que permite hacer extracciones de DNA rápidas sin que sea necesario purificar el DNA en un gradiente de CsCl.

1. Triturar 1-2 g de tejido en mortero con N<sub>2</sub> líquido. Pasar el material pulverizado y congelado a un tubo de centrifuga de 15 ml (Corex<sup>®</sup>).
2. Añadir 7,5 ml de tampón de extracción<sup>1</sup> (mezclado con 2- $\beta$ -mercaptoetanol) y 0,5 ml de SDS 20%. Mezclar por inversión.
3. Incubar 10 min a 65°C.
4. Añadir 2,5 ml de KAc 5M y mezclar por inversión. Incubar 20 min en hielo.
5. Centrifugar 20 min a 10000 rpm.
6. Pasar el sobrenadante a un tubo de centrifuga de 30 ml (Corex<sup>®</sup>), a través de una gasa de nylon Miracloth<sup>™</sup> (Calbiochem, La Jolla, CA).
7. Añadir 5 ml de isopropanol. Mezclar bien e incubar 30 min a -20°C.
8. Centrifugar 20 min a 10000 rpm.
9. Eliminar el sobrenadante y dejar secar el precipitado al aire.
10. Resuspender el precipitado en 0,7 ml de H<sub>2</sub>O y transferir la solución a un vial de microcentrifuga de 1,5 ml.
11. Dar un pulso de centrifuga para eliminar restos insolubles y pasar el sobrenadante a un nuevo vial de 1,5 ml.
12. Añadir 75  $\mu$ l de NaAc 3M pH 5,2 y 0,5 ml de isopropanol. Mezclar suavemente. Observar la aparición del precipitado en forma de hilos.
13. Centrifugar 30 s a 10000 rpm.
14. Lavar el precipitado con etanol 70% y dejar secar un poco.
15. Resuspender el precipitado en 100  $\mu$ l de H<sub>2</sub>O.
16. Añadir RNasa A<sup>2</sup> (libre de DNasas).
17. Digerir 20 min a 37°C.
18. Añadir 100  $\mu$ l de fenol/cloroformo/alcohol isoamílico (25:24:1, V/V/V).
19. Centrifugar 5 minutos a 10000 rpm.
20. Pasar sobrenadante a un nuevo vial de microcentrifuga y descartar la fase inferior.
21. Añadir 0,1 V de NaAc 3 M pH 5,2 y 2,5 V de EtOH 100%. Mezclar suavemente.
22. Centrifugar 2 min a 10000 rpm.
23. Descartar sobrenadante y lavar el precipitado con EtOH 70%.
24. Resuspender el precipitado en 100  $\mu$ l de H<sub>2</sub>O estéril.

25. Guardar a 4°C.

<sup>1</sup>**Tampón de extracción de DNA:** 0,1 mM Tris-HCl pH 8,0; 50 mM EDTA pH 8,0; 0,5 mM NaCl; 10 mM 2-β-mercaptoetanol (añadir en el momento de utilización).

<sup>2</sup>**RNasa A:** 100 µg/ml.

### 2.1.2.1. Extracción rápida de DNA genómico.

Para análisis por PCR (polimorfismos y genotipado), en los que la integridad del DNA genómico no es un factor crítico, se ha seguido un protocolo adaptado del método de extracción rápida de DNA genómico de planta descrito por Edwards *et al.* (Edwards K *et al.*, 1991). Todos los pasos de este protocolo se pueden realizar a temperatura ambiente.

1. Recoger la muestra (muy pequeña cantidad), colocarla dentro de un vial de microcentrífuga.
2. Añadir 400 µl de tampón de extracción<sup>3</sup> y triturar la muestra en el mismo vial, con ayuda de un émbolo estéril durante 15 s y mezclar bien mediante agitación en vórtex durante 5 s.
3. Centrifugar a 14000 rpm durante 1 min.
4. Transferir 300 µl del sobrenadante a un nuevo vial. Añadir 1 V de isopropanol e incubar durante 2 min.
5. Centrifugar a 14000 rpm durante 5 min.
6. Descartar el sobrenadante y dejar secar el precipitado.
7. Resuspender el precipitado en 100 µl de H<sub>2</sub>O. (2,5 µl son suficientes para el análisis por PCR estándar).

<sup>3</sup>**Tampón de extracción rápida de DNA:** 200 mM Tris-HCl pH 7,5; 250 mM NaCl; 25 mM EDTA pH 8,0; 0,5% (w/v) SDS.

### 2.1.3. EXTRACCIÓN DE RNA.

La extracción de RNA total de planta se ha realizado utilizando el reactivo TRIZOL® (Invitrogen™), siguiendo las indicaciones del fabricante. Todo el material se ha tratado antes con 10% (w/v) SDS, para desnaturalizar cualquier enzima residual.

Alternativamente, se ha utilizado un método de extracción de RNA de plantas modificado a partir del descrito por Logeman y colaboradores (Logemann J *et al.*, 1987). A continuación se detalla el protocolo seguido:

1. Triturar el tejido vegetal en mortero con N<sub>2</sub> líquido y pasar el material triturado a un vial de microcentrífuga de 1,5 ml.
2. Añadir 400 µl de tampón Z6<sup>1</sup> (ya mezclado con 2-β-mercaptoetanol), mezclar por agitación en vórtex exhaustivamente y congelar en N<sub>2</sub> líquido.
3. Dejar descongelar la muestra en hielo (mientras tanto se puede ir triturando el resto de muestras).
4. Añadir 400 µl de fenol/cloroformo/alcohol isoamílico (25:24:1, V/V/V).
5. Mezclar vigorosamente agitando en vórtex.
6. Centrifugar 5 minutos a 13000 rpm.
7. Pasar sobrenadante a un nuevo vial de microcentrífuga de 1,5 ml y descartar la fase inferior.
8. Repetir pasos 4-7.
9. Añadir 0,1 V de 1M HAc y mezclar vigorosamente.

## Material y métodos.

10. Añadir 0,7 V de etanol absoluto (EtOH 100%) y mezclar vigorosamente.
11. Dejar precipitar 30 minutos a  $-20^{\circ}\text{C}$  o bastante tiempo a  $4^{\circ}\text{C}$ .
12. Centrifugar 20 minutos a 13000 rpm.
13. Descartar sobrenadante y guardar el precipitado.
14. Añadir 400  $\mu\text{l}$  de EtOH 70%.
15. Centrifugar 5 minutos a 13000 rpm.
16. Descartar sobrenadante y guardar el precipitado.
17. Añadir 400  $\mu\text{l}$  de NaAc 3 M pH 5,5.
18. Mezclar vigorosamente, para disolver todos los polisacáridos posibles.
19. Centrifugar 10 minutos a 13000 rpm.
20. Descartar sobrenadante y guardar el precipitado.
21. Añadir 400  $\mu\text{l}$  de EtOH 70%.
22. Centrifugar 5 minutos a 13000 rpm.
23. Descartar sobrenadante y guardar el precipitado.
24. Añadir 400  $\mu\text{l}$  de EtOH 70%.
25. Centrifugar 5 minutos a 13000 rpm.
26. Descartar sobrenadante, eliminando muy bien todo el EtOH y dejar secar el precipitado (el menor tiempo posible, para evitar degradación de RNA y facilitar su resuspensión).
27. Resuspender el precipitado en 50-100  $\mu\text{l}$  de  $\text{H}_2\text{O}$  estéril (dependiendo de la cantidad de material de partida), agitando vigorosamente con el vórtex.
28. Calentar a  $65^{\circ}\text{C}$  durante 5 min.
29. Poner en hielo y dejarlo 1 min.
30. Agitar en vórtex.
31. Centrifugar 5 minutos a 13000 rpm.
32. Pasar el sobrenadante a otro vial de microcentrífuga (descartar el precipitado que haya podido quedar).

<sup>1</sup>**Tampón de extracción Z6:** 8 M Hidrocloruro de Guanidinio; 20 mM MES<sup>2</sup> pH 7,0; 20 mM EDTA pH 8,0. Justo antes de utilizar añadir 1% (V/V) 2- $\beta$ -mercaptoetanol.

<sup>2</sup>**MES 0,5 M, pH 7,0:** disolver primero los 4,88 g de MES en 40 ml de  $\text{H}_2\text{O}$ ; ajustar pH con KOH 5 M; añadir  $\text{H}_2\text{O}$  hasta enrasar los 50 ml; guardar a  $-20^{\circ}\text{C}$ .

### 2.1.3.1. Tratamiento con DNasa.

Una vez hemos extraído el RNA lo sometemos a un tratamiento con DNasa, con tal de eliminar el DNA y tener un RNA más puro.

1. Mezclar en un vial de microcentrífuga de 1,5 ml: 50  $\mu\text{l}$  de RNA de la extracción; 5  $\mu\text{l}$  de 10x tampón de DNasa<sup>1</sup>; 1  $\mu\text{l}$  RNasina<sup>2</sup>; 1  $\mu\text{l}$  DNasa<sup>3</sup>.
2. Incubar 30 minutos a  $37^{\circ}\text{C}$ .
3. Añadir 1 volumen de fenol/cloroformo/alcohol isoamílico (25:24:1, V/V/V).
4. Mezclar vigorosamente agitando en vórtex.
5. Centrifugar 5 minutos a 13000 rpm.
6. Pasar sobrenadante a un nuevo vial de microcentrífuga de 1,5 ml y descartar la fase inferior.
7. Añadir 0,1 V de NaAc 3M pH 5,5 y 2,5 V de EtOH 100%.
8. Dejar precipitar 30 minutos a  $-20^{\circ}\text{C}$  o bastante tiempo a  $4^{\circ}\text{C}$
9. Centrifugar 20 minutos a 13000 rpm.
10. Descartar sobrenadante y guardar el precipitado.
11. Añadir 200  $\mu\text{l}$  de EtOH 70%.
12. Centrifugar 5 minutos a 13000 rpm.



13. Descartar sobrenadante y guardar el precipitado
14. Resuspender el precipitado en 50  $\mu$ l de H<sub>2</sub>O estéril.
15. Guardar el RNA a -80°C.

<sup>1</sup>**Tampón DNasa 10x**: 0,4 M Tris-HCl pH 7,5; 60 mM MgCl<sub>2</sub>.

<sup>2</sup>**RNasina**: RNAGuard™ RNase Inhibitor Porcine (24,8 U/ $\mu$ l), Amersham Pharmacia Biotech.

<sup>3</sup>**DNasa**: DNaseI RNase free (10 U/ $\mu$ l), Boehringer Mannheim.

## 2.2. REACCIÓN EN CADENA DE LA POLIMERASA (PCR).

### 2.2.1. Mezcla de la reacción de PCR.

Para las diferentes PCRs de este trabajo se ha partido de 10 ng o 100 ng de DNA molde, dependiendo de si se trata de DNA plasmídico o de DNA genómico, respectivamente. Para llevar a cabo la amplificación del fragmento deseado se añade a este DNA la mezcla de la reacción de la PCR<sup>1</sup>.

<sup>1</sup>**Mezcla de la reacción de la PCR**: 1x tampón de PCR<sup>2</sup>; 200  $\mu$ M dNTPs; 1,5 mM MgCl<sub>2</sub>; 10 pmol oligonucleótido 5'; 10 pmol oligonucleótido 3'; 2,5 u *Taq* polimerasa.

<sup>2</sup>**Tampón de PCR 10x**: 500 mM KCl; 200 mM Tris-HCl pH 8,6; 1% Tritón X-100.

### 2.2.2. Condiciones de la reacción.

Las condiciones a las que se han realizados las PCRs han sido:

**a. Pre-desnaturalización**: 94°C durante 3 min.

**b. Amplificación** (25-35 ciclos):

**b.i. Desnaturalización**: 94°C durante 30 s.

**b.ii. Hibridación**: 48-65°C durante 30 s. (Esta temperatura depende de cada par de oligos, se ha utilizado una temperatura 2°C inferior a la T<sub>m</sub> de los oligos).

**b.iii. Extensión**: 72°C durante el tiempo necesario para completar la extensión del fragmento deseado. La *Taq* pol extiende a razón de 1 Kb/min.

**c. Extensión final**: 72°C durante 5 min.

## 2.3. REACCIÓN DE RT-PCR.

La reacción de RT-PCR permite la amplificación de las moléculas de RNA. Esta reacción consta de dos pasos: síntesis de la molécula de cDNA (DNA copia) y de una reacción de PCR convencional, que utiliza como molde este cDNA.

### 2.3.1. SÍNTESIS DE LA MOLÉCULA DE cDNA.

1. Mezclar, en un volumen final de 11  $\mu$ l, 5  $\mu$ g de RNA; 12,4 U RNasina; 100 pmol Oligo(dT)<sup>1</sup>.
2. Incubar la muestra 5 minutos a 65°C. Enfriar en hielo.
3. Llevar la mezcla a 20  $\mu$ l de volumen final conteniendo 10 mM DTT; 1x tampón de M-MLV RT; 500  $\mu$ M dNTPs; 24,8 U RNasina; 200 U M-MLV RT (GIBCO BRL).
4. Incubar 1 hora a 37°C.
5. Añadir 80  $\mu$ l H<sub>2</sub>O estéril y guardar a -20°C.

<sup>1</sup>El Oligo(dT) utilizado contiene una extensión en 5', ver tabla de oligos (Casacuberta JM *et al.*, 1995).

## 2.4. PURIFICACIÓN DE LOS PRODUCTOS DE PCR.

La purificación de los productos de la PCR se ha realizado utilizando el **QIAquick® PCR Purification Kit** (QIAGEN), siguiendo las indicaciones especificadas por el fabricante.

## 2.5. CLONACIÓN DE PRODUCTOS DE PCR.

Para la clonación de los productos de PCR, a lo largo de este trabajo, se han utilizado cuatro kits comerciales diferentes, dependiendo de la disponibilidad de los mismos en el laboratorio. Se han seguido las indicaciones especificadas en cada uno de ellos para clonar los productos obtenidos de la reacción de PCR en los correspondientes plásmidos. Los kits utilizados han sido: **pGEM®-T Easy Vector System I** (Promega), **InsT/Aclone™ PCR Product Cloning Kit** (Fermentas), **TA Cloning® Kit** (Invitrogen) y **TOPO TA Cloning® Kit** (Invitrogen).

## 2.6. ELECTROFORESIS DE ÁCIDOS NUCLEICOS EN GEL DE AGAROSA.

La migración electroforética del DNA se ha llevado a cabo, generalmente, en geles de agarosa al 1% (w/v); 1x TAE<sup>1</sup> (ó 0,5x TBE<sup>2</sup>); 0,5 µg/ml BrEt. El porcentaje de agarosa se ha variado en aquellos casos que así lo requerían, aumentándolo o disminuyéndolo con el fin de poder discriminar entre fragmentos pequeños de DNA o entre fragmentos de gran tamaño de DNA, respectivamente. Para cargar el DNA en el gel se le añade tampón colorante de carga en gel<sup>3</sup> (ver soluciones y tampones generales). La migración se realiza, normalmente, a un voltaje constante de 100 V en 1x TAE (ó 0,5x TBE).

La electroforesis del RNA para su visualización después de una extracción se realiza de la misma forma que la arriba descrita para el DNA. Simplemente se limpia con 10% SDS<sup>4</sup> todo el material y se calientan las muestras a 65°C durante 5 min.

La electroforesis del RNA en gel desnaturizante se describe en el apartado 4.1.

<sup>1</sup>TAE (Tris Acetato pH 8,1) 50x: 2 M Tris base; 1 M HAc; 0,1M EDTA pH 8,0.

<sup>2</sup>TBE (Tris Borato pH 8,3) 10x: 0,89 M Tris base; 0,89 M Ácido Bórico; 0,1 M EDTA pH 8,0.

<sup>3</sup>Tampón colorante de carga en gel 10x: 50% (v/v) glicerol; 10 mM EDTA pH 8,0; 0,25% (w/v) azul de bromofenol; 0,25% (w/v) xylene cianol FF.

<sup>4</sup>SDS 10%: 100 g de SDS en 1 L de H<sub>2</sub>O.

## 2.7. EXTRACCIÓN DE FRAGMENTOS DE DNA DESDE GEL DE AGAROSA.

Para la extracción de fragmentos de DNA desde gel de agarosa, tras la migración de los mismos, se ha utilizado el kit **QIAquick® Gel Extraction Kit** (QIAGEN), según las indicaciones del fabricante.

## 2.8. REACCIONES DE MODIFICACIÓN DEL DNA.

Las reacciones de modificación del DNA realizadas en este trabajo comprenden la digestión del DNA con enzimas de restricción, fosforilación y de-fosforilación de extremos, reacciones de

rellenado de extremos protuberantes y reacciones de limado de los mismos para crear extremos romos. Los métodos para todas estas reacciones aparecen descritos en los manuales de biología molecular *Molecular Cloning: A Laboratory manual* y *Current Protocols in Molecular Biology* (Sambrook J and Russell DW, 2001; Ausubel FM *et al.*, 2001).

## 2.9. LIGACIÓN DE FRAGMENTOS DE DNA.

La eficiencia de la reacción de ligación viene determinada, básicamente, por las cantidades de vector e inserto presentes en ella. El cálculo de estas cantidades, cuya proporción se indica a continuación, así como la propia reacción de ligación se han realizado según se indica en *Molecular Cloning: A Laboratory Manual* (Sambrook J and Russell DW, 2001). Este protocolo se ha variado según indicaciones del fabricante cuando se han utilizado ligasas de características especiales.

Fórmula para calcular la cantidad de inserto en la mezcla de reacción:

$$\text{ng inserto} = \text{ng vector} \times \text{MW inserto (Kb)} / \text{MW vector (Kb)}$$

Esta fórmula determina una relación inserto : vector de 1:1. Generalmente, para las reacciones de ligación se utiliza una relación 3:1 (inserto vector). Pero para algunas reacciones de ligación de extremos romos se ha utilizado la proporción 5:1 (inserto : vector).

## 2.10. TRANSFORMACIÓN BACTERIANA.

### 2.10.1. TRANSFORMACIÓN DE *E.coli* POR CHOQUE TÉRMICO.

#### 2.10.1.1. Preparación de células competentes de *E.coli* para transformación por choque térmico.

El protocolo seguido para la preparación de células competentes de *E.coli* para transformación por choque térmico es una variación del descrito por Hanahan (Hanahan D, 1983), que permite obtener una eficiencia teórica de transformación de, aproximadamente,  $10^6$ - $10^7$  cfu/ $\mu$ g DNA.

1. Inocular un precultivo de 5 ml de LB a partir de una colonia aislada de *E.coli* DH5 $\alpha$ , crecida en placa de medio LB, e incubar a 37°C en agitación durante 14-16 h (toda la noche).
2. Inocular 2,5 ml del precultivo crecido en 250 ml de LB. Incubar a 37°C en agitación hasta alcanzar una DO<sub>600nm</sub> de 0,5 (2-3 h).
3. Enfriar el cultivo en hielo durante 15 min. Trabajar siempre en frío.
4. Centrifugar a 4500 rpm durante 10 min a 4°C.
5. Decantar el sobrenadante y resuspender el precipitado de células en 50 ml de solución TFB1<sup>1</sup> fría (con cuidado de no lisar las células).
6. Incubar 10 min en hielo.
7. Centrifugar a 4500 rpm durante 10 min a 4°C.
8. Decantar el sobrenadante y resuspender el precipitado de células en 5 ml de solución TFB2<sup>2</sup> fría (con cuidado de no lisar las células).
9. Hacer alícuotas de 100  $\mu$ l y congelar inmediatamente en N<sub>2</sub> líquido. Almacenar a -80°C.

## Material y métodos.

<sup>1</sup>**TFB1:** 30mM KAc; 50 mM MnCl<sub>2</sub>·4H<sub>2</sub>O; 10 mM CaCl<sub>2</sub>·2H<sub>2</sub>O; 100 mM RbCl; 15% (v/v) glicerol. Esterilizar por filtración y guardar a 4°C.

<sup>2</sup>**TFB2:** 10 mM MOPS; 75 mM CaCl<sub>2</sub>·2H<sub>2</sub>O; 10 mM RbCl; 15% (v/v) glicerol. Esterilizar por filtración y guardar a 4°C.

### 2.10.1.2. Transformación de *E.coli* por choque térmico.

1. Añadir 50-100 ng de DNA a una alícuota de células competentes, previamente descongelada en hielo. Mezclar e incubar en hielo durante 20 min.
2. Aplicar un choque térmico de 42°C durante 2 min.
3. Incubar en hielo durante 2 min.
4. Añadir 300 µl de medio SOB e incubar durante 45 min a 37°C en agitación.
5. Centrifugar a 6000 rpm durante 2 min. Resuspender el sedimento bacteriano en 200 µl del sobrenadante.
6. Sembrar en placas de LB/agar (con el antibiótico correspondiente para la selección de transformantes) e incubar a 37°C, en posición invertida, durante toda la noche (12-16 h).

### 2.10.2. TRANSFORMACIÓN DE *E.coli* POR ELECTROPORACIÓN.

La transformación de células bacterianas por electroporación se ha utilizado en aquellos casos que presentaban una mayor dificultad: transformación de plásmidos de gran tamaño (es el caso de los plásmidos binarios) o de reacciones de muy baja eficiencia de ligación.

#### 2.10.2.1. Preparación de células competentes de *E.coli* para transformación por electroporación.

El protocolo seguido para la preparación de células competentes de *E.coli* para transformación por electroporación es una variación del descrito por Seidman y colaboradores (Seidman CG *et al.*, 1997), que permite obtener una eficiencia teórica de transformación de, aproximadamente, 10<sup>9</sup> cfu/µg DNA.

1. Inocular un precultivo de 5 ml de LB a partir de una colonia aislada de *E.coli* DH5α, crecida en placa de medio LB, e incubar a 37°C en agitación durante 14-16 h (toda la noche).
2. Inocular los 5 ml del precultivo crecido en 1 L de LB. Incubar a 37°C en agitación hasta alcanzar una DO<sub>600nm</sub> de 0,6-0,9.
3. Enfriar el cultivo en hielo durante 15 min. Trabajar siempre en frío.
4. Centrifugar a 4000 rpm durante 5 min a 4°C.
5. Decantar el sobrenadante y resuspender suavemente el sedimento bacteriano en 100 ml de H<sub>2</sub>O estéril fría.
6. Centrifugar a 4000 rpm durante 5 min a 4°C.
7. Decantar el sobrenadante y resuspender suavemente el sedimento bacteriano en 20 ml de solución fría de glicerol al 10% (v/v).
8. Centrifugar a 4000 rpm durante 5 min a 4°C.
9. Decantar el sobrenadante y resuspender suavemente el sedimento bacteriano en 2-3 ml de solución fría de glicerol al 10% (v/v).
10. Hacer alícuotas de 40 µl y congelarlas inmediatamente en N<sub>2</sub> líquido. Almacenarlas a -80°C.

### **2.10.2.2. Transformación de *E.coli* por electroporación.**

1. Añadir 20-40 ng de DNA, previamente dializado (mediante membranas de nitrocelulosa de 0,025  $\mu\text{m}$  de poro), a una alícuota de células electrocompetentes, descongeladas en hielo. Incubar durante 1 min en hielo y transferir a una cubeta de electroporación (de 1 mm de luz), también en hielo.
2. Aplicar un pulso de 1,7 kV de potencia. Condiciones de la electroporación: 25  $\mu\text{F}$  de capacidad eléctrica; 200  $\Omega$  de resistencia.
3. Añadir inmediatamente 300  $\mu\text{l}$  de medio SOB e incubar en hielo durante 2 min.
4. Incubar 45 min a 37°C en agitación.
5. Centrifugar a 6000 rpm durante 2 min. Resuspender el sedimento bacteriano en 200  $\mu\text{l}$  del sobrenadante.
6. Sembrar en placas de LB/agar (con el antibiótico correspondiente para la selección de transformantes) e incubar a 37°C, en posición invertida, durante toda la noche (12-16 h).

### **2.10.3. TRANSFORMACIÓN DE *A.tumefaciens* POR CHOQUE TÉRMICO.**

#### **2.10.3.1. Preparación de células competentes de *A.tumefaciens* para transformación por choque térmico**

La preparación de bacterias competentes de *Agrobacterium tumefaciens* cepa LBA4404 se realiza de acuerdo con el protocolo descrito por An y colaboradores (An G *et al.*, 1988).

1. Hacer un precultivo, en medio YEB (100  $\mu\text{g/ml}$  Rif; 300  $\mu\text{g/ml}$  Sm), de *A.tumefaciens* cepa LBA4404, a partir de una colonia aislada crecida en placa, y dejar crecer a 28°C durante 48-72 h.
2. Inocular 1 ml del precultivo en 100 ml de YEB (100  $\mu\text{g/ml}$  Rif; 300  $\mu\text{g/ml}$  Sm) y dejar crecer a 28°C hasta que la  $\text{DO}_{595\text{nm}}$  llegue a 0,6.
3. Centrifugar a 3000 g 10 min a 4°C.
4. Resuspender el sedimento celular en 2 ml de  $\text{CaCl}_2$  20 mM estéril y frío.
5. Hacer alícuotas de 200  $\mu\text{l}$  y congelar en  $\text{N}_2$  líquido. Se guardan a -80°C un máximo de 3 meses.

#### **2.10.3.2. Transformación de *A.tumefaciens* por choque térmico.**

1. Añadir 1  $\mu\text{g}$  de DNA a una alícuota de células competentes descongelada en hielo.
2. Incubar 5 min en hielo.
3. Incubar 5 min a -80°C.
4. Incubar 5 min a 37°C.
5. Añadir 1 ml de medio YEB e incubar 4 h en agitación a 28°C.
6. Centrifugar a 6000 rpm durante 2 min. Resuspender el sedimento bacteriano en 200  $\mu\text{l}$  del sobrenadante.
7. Sembrar en placas de YEB/agar (con los antibióticos correspondientes para la selección de transformantes) e incubar a 28°C, en posición invertida, durante 48-72h.

## **2.11. CUANTIFICACIÓN DE ÁCIDOS NUCLEICOS.**

La cuantificación de ácidos nucleicos se ha realizado por espectrofotometría utilizando el NanoDrop® ND-1000 Spectrophotometer y el programa informático ND-1000 V3.1.0 (NanoDrop Technologies Inc.).

## **2.12. SECUENCIACIÓN DE ÁCIDOS NUCLEICOS.**

Las reacciones de secuenciación necesarias a lo largo de este trabajo han sido realizadas en el Servicio de Secuenciación del IBMB-CSIC.

## **2.13. MARCAJE RADIATIVO DE ÁCIDOS NUCLEICOS.**

### **2.13.1. MARCAJE DE DNA.**

Las sondas radiactivas de DNA utilizadas para la mayoría de las hibridaciones descritas en este trabajo se han sintetizado mediante la técnica de random priming. Se ha utilizado el kit comercial Random Primed DNA Labeling Kit (Roche) y se han seguido las indicaciones del fabricante para completar la reacción de marcaje.

### **2.13.2. MARCAJE DE sRNAs.**

El marcaje de los sRNAs se ha llevado a cabo mediante la incorporación de un grupo fosfato radiactivo en su extremo 5' mediante una reacción catalizada por la T4 polinucleótido quinasa (PNK). Se ha utilizado una T4 PNK comercial (Fermentas) según las indicaciones del fabricante.

## **2.14. PURIFICACIÓN O AISLAMIENTO DE ÁCIDOS NUCLEICOS MARCADOS RADIATIVAMENTE.**

### **2.14.1. Purificación de sondas de DNA marcadas radiactivamente.**

La purificación de las sondas de DNA marcadas radiactivamente se ha realizado utilizando las columnas **ProbeQuant™ G-50 Micro Columns** (Amersham Biosciences), siguiendo las indicaciones especificadas por el fabricante.

### **2.14.2. Purificación de sRNAs marcados radiactivamente.**

La purificación de los sRNAs marcados radiactivamente se ha realizado siguiendo el protocolo que se detalla a continuación:

1. Precipitar los sRNAs marcados radiactivamente en  $\text{NH}_4\text{OAc}$  0,5 M (final); EtOH 70% (final); glicógeno 0,1  $\mu\text{g}/\mu\text{l}$  (final). Mezclar bien e incubar 2 h a  $-20^\circ\text{C}$ .
2. Centrifugar a 14000 rpm durante 30 min a  $4^\circ\text{C}$ .
3. Eliminar el sobrenadante.
4. Volver a precipitar añadiendo al precipitado 1 ml EtOH 100%; 200  $\mu\text{l}$   $\text{NH}_4\text{OAc}$  5 M y 60  $\mu\text{l}$  de  $\text{H}_2\text{O}$  estéril.
5. Mezclar bien e incubar 30 min a  $-20^\circ\text{C}$ .

## Material y métodos.

6. Centrifugar a 14000 rpm durante 30 min a 4°C.
7. Eliminar el sobrenadante.
8. Lavar el precipitado con 500 µl EtOH 70%.
9. Centrifugar a 14000 rpm durante 10 min a 4°C.
10. Eliminar el sobrenadante y dejar secar el precipitado.
11. Resuspender en 40 µl de H<sub>2</sub>O estéril. Guardar en hielo.

### 3. TÉCNICA DE SOUTHERN.

Gracias al método de detección de secuencias específicas de DNA descrita por Southern (Southern EM, 1975), se puede conocer el número de copias de un transgén que se han integrado en el genoma de un determinado individuo. Esta técnica se ha utilizado en este trabajo para seleccionar las plantas monocopia para el transgén transformado.

#### 3.1. DIGESTIÓN DEL DNA GENÓMICO.

En cada reacción de digestión se han digerido 10 µg de DNA genómico en 200 µl de volumen final, utilizando 100 u. de enzima de restricción y 2.5 mM de espermidina, durante 3-4 h.

#### 3.2. ELECTROFORESIS EN GEL DE AGAROSA.

Los DNAs digeridos se han desnaturalizado a 65°C durante 5 min e incubado, posteriormente, en hielo durante 2 min. La migración de los mismos se ha realizado a 100 V durante 3 h en un gel de agarosa al 0,8% (w/v); 0,5x TBE<sup>1</sup>; 0,5 µg/ml BrEt.

<sup>1</sup>TBE (Tris Borato pH 8,3) 10x: 0,89 M Tris base; 0,89 M Ácido Bórico; 0,1 M EDTA pH 8,0.

#### 3.3. TRANSFERENCIA SOUTHERN.

Para que pueda transferirse correctamente el DNA, y haya una óptima hibridación posterior, éste debe ser desnaturalizado previamente. El gel de agarosa donde han migrado los DNAs es sometido a los siguientes tratamientos: 0,25 N HCl durante 15 min (para facilitar la transferencia de fragmentos de DNA de gran tamaño); solución desnaturalizante<sup>1</sup> durante 20 min; solución neutralizante<sup>2</sup> durante 20 min; 10x SSC<sup>3</sup> durante 20 min (para equilibrar el gel en el tampón de transferencia).

La transferencia del DNA desde el gel de agarosa hasta la membrana de nylon (Hybond™-N, Amersham-Pharmacia-Biotech) se realiza mediante el método de transferencia capilar en líquido (Sambrook J and Russell DW, 2001), durante toda la noche (12-16 h) en 10x SSC.

Una vez transferido, se fija el DNA a la membrana mediante irradiación con luz U.V. La luz U.V. consigue que se entrecrucen las moléculas del ácido nucleico con las fibras de nylon de la membrana y, de esta forma, queden fijadas en la misma. La fijación por UV se ha realizado utilizando el UV Stratalinker® 2400 de Stratagene, seleccionando la opción de *auto crosslink* (irradiación de una energía total de 120 mJ). En su defecto, o complementariamente, se puede

## Material y métodos.

fijar los ácidos nucleicos a la membrana mediante el tueste de la misma a 80°C durante 2-3 horas.

<sup>1</sup>**Solución desnaturalizante:** 0,5 M NaOH; 1,5 M NaCl.

<sup>2</sup>**Solución neutralizante:** 0,5 M Tris; 1,5 M NaCl. Ajustar a pH 8,0 con HCl 25% (v/v).

<sup>3</sup>**SSC 20x:** 3 M NaCl; 300 mM citrato de sodio pH 7,0.

### 3.4. HIBRIDACIÓN SOUTHERN.

Se ha pre-hibridado la membrana de DNA incubándola en 20 ml de solución de hibridación de Southern<sup>1</sup> a 42°C, durante 1-2 h en agitación. Posteriormente se ha hibridado la membrana con la sonda marcada (apartado 2.13.), previamente desnaturalizada (5 min a 95°C), en 15 ml de solución de hibridación de Southern y 10% (w/v) dextrán sulfato, a 42°C durante toda la noche (14-18 h) en agitación.

Tras la hibridación se ha lavado la membrana, para eliminar la sonda unida inespecíficamente al DNA, en las siguientes condiciones: 2x SSC, 0,1% (w/v) SDS a 65°C durante 20 min; 1x SSC, 0,1% (w/v) SDS a 65°C durante 20 min; 0,5x SSC, 0,1% (w/v) SDS a 65°C durante 20 min; 0,2x SSC, 0,1% (w/v) SDS a 65°C durante 20 min.

<sup>1</sup>**Solución de hibridación de Southern:** 40% (v/v) formamida; 0,5% (w/v) SDS; 5x SSPE<sup>2</sup>; 5x reactivo Denhardt's<sup>3</sup>; 0,5 µg/µl DNA de esperma de salmón (desnaturalizado 10 min a 98°C).

<sup>2</sup>**SSPE 20x:** 3,6 M NaCl; 0,2 M NaH<sub>2</sub>PO<sub>4</sub>; 0,02 M EDTA.

<sup>3</sup>**Reactivo Denhardt's 100x:** 2% (w/v) Ficoll 400; 2% (w/v) PVP (polivinilpirrolidona); 2% (w/v) BSA (fracción V). Esterilizar por filtración y almacenar a -20°C.

### 3.5. EXPOSICIÓN Y REVELADO.

Una vez lavada la membrana, se introduce en una bolsa de plástico, de manera que no se seque completamente. La membrana se deja exponiendo dentro de un casete de exposición (Hypercassette™, Amersham Biosciences) frente a una pantalla de fósforo (Storage Phosphor Screen, Amersham Biosciences), cuya superficie se ve alterada (activada) por la radiación que recibe desde la membrana, pudiendo ser leída (revelada) posteriormente en el escáner **Molecular Imager® FX** (BIO-RAD) utilizando el paquete informático **Quantity One® 4.1.1.**, proporcionado por la misma casa comercial. La pantalla de fósforo se desactiva otra vez al exponerla ante la fuente de luz emitida por el **Screen Eraser – K** (BIO-RAD).

### 3.6. DESHIBRIDACIÓN DE LAS MEMBRANAS.

Las membranas que han sido hibridadas con una sonda pueden ser rehibridadas con otra sonda diferente, una vez deshibridada la sonda anterior. No es conveniente hibridar una membrana más de 3 ó 4 veces, ya que cada vez que se hibrida y se deshibrida se alteran las características de los ácidos nucleicos fijados en la misma y se deterioran.

La deshibridación de las membranas, indistintamente si contienen DNA o RNA, se realiza lavando las membranas hibridadas con H<sub>2</sub>O hirviendo en agitación durante 10-15 min. Repetir el lavado si todavía hay sonda unida a la membrana.



## 4. TÉCNICA DE NORTHERN.

### 4.1. ELECTROFORESIS DE RNA EN GEL DESNATURALIZANTE.

La muestra ha de ser desnaturalizada antes de correr la electroforesis. Para ello se mezcla 1 volumen de RNA (10 µg) con 2 volúmenes de tampón de carga de Northern<sup>1</sup>, se incuba a 65°C durante 15 minutos y se deja enfriar en hielo durante 2 minutos antes de cargar en el gel. La electroforesis se realiza en gel desnaturalizante<sup>2</sup> en MEN<sup>3</sup> 1x, a un voltaje fijo de 100 V durante 2-3 horas.

<sup>1</sup>**Tampón de carga de Northern:** 1,5x MEN pH 7,0; 60% (v/v) formamida desionizada; 6,85% (v/v) formaldehído; 1.5x tampón colorante de carga en gel; 0,75 µg/ml BrEt.

<sup>2</sup>**Gel desnaturalizante:** 1% agarosa, 1x MEN pH 7,0; 7,4% (v/v) formaldehído.

<sup>3</sup>**MEN 10x:** 200 mM MOPS pH 7,0; 50 mM NaAc; 10 mM EDTA pH 8,0. Ajustar a pH 7,0 y autoclavar. Conservar a 4°C.

### 4.2. TRANSFERENCIA NORTHERN.

Se sumerge el gel en NaOH 0,01 N durante 20 minutos en agitación, para optimizar la transferencia de los fragmentos grandes de RNA y eliminar el máximo de formaldehído posible del gel (la presencia de formaldehído dificulta la transferencia del RNA a la membrana de nylon). Seguidamente se sumerge el gel en 10x SSC durante 20 minutos en agitación para equilibrarlo en el tampón de transferencia.

La transferencia del RNA desde el gel de agarosa hasta la membrana de nylon y su posterior fijación a esta última se realiza como se describe en el apartado 3.3., de la misma forma que para el DNA.

### 4.3. HIBRIDACIÓN NORTHERN.

La membrana que contiene el RNA se prehibrida en 20 ml de solución de hibridación<sup>1</sup> a 65°C durante 1 hora. Posteriormente se hibrida con la sonda marcada (ver apartado 2.13.), previamente desnaturalizada (5 min a 95°C), en 15 ml de solución de hibridación a 65°C durante toda la noche (14-18 h). Transcurrido este tiempo, se lava la membrana 3 veces con 50 ml de solución de lavado<sup>2</sup> durante 20 minutos a 65°C.

<sup>1</sup>**Solución de hibridación** (Church GM and Gilbert W, 1984): 250 mM tampón fosfato<sup>3</sup> pH 7,2; 7% (w/v) SDS; 1 mM EDTA pH 8,0.

<sup>2</sup>**Solución de lavado:** 20 mM tampón fosfato<sup>3</sup> pH 7,2; 1% (w/v) SDS; 1 mM EDTA pH 8,0.

<sup>3</sup>**Tampón (sodio) fosfato 1M pH 7,2:** 68,4% (v/v) 1 M Na<sub>2</sub>HPO<sub>4</sub>; 31,6% (v/v) 1 M NaH<sub>2</sub>PO<sub>4</sub>.

### 4.4. EXPOSICIÓN Y REVELADO.

La exposición y revelado de las membranas se ha realizado como se indica en el apartado 3.5.

## 5. ANÁLISIS DE LOS sRNAs.

### 5.1. AISLAMIENTO DE LMW-RNAs.

Para la obtención de la fracción aislada de los RNAs de bajo peso molecular (LMW-RNAs) se ha utilizado el QIAGEN® RNA/DNA Midi Kit y se ha seguido el protocolo proporcionado por el fabricante.

### 5.2. ELECTROFORESIS EN GEL DESNATURALIZANTE DE POLIACRILAMIDA.

La migración electroforética de los LMW-RNAs y sRNAs se ha realizado en gel desnaturalizante de poliacrilamida<sup>1</sup>, a un amperaje constante de 12 mA (o voltaje de 180V) en 0,5x TBE<sup>2</sup>. Se realiza una migración previa de 30-45 min, en las mismas condiciones, para equilibrar el gel.

La muestra (LMW-RNAs) se mezcla con 1 volumen de formamida, se desnaturaliza durante 5 min a 95°C, se incuba 2 min en hielo, se mezcla con tampón colorante de carga en gel<sup>3</sup> y se carga en el gel.

Como marcadores de tamaño se han utilizado combinaciones de oligos de DNA de 20 a 26 nt, añadiendo, al menos, 500 pmoles, de forma que puedan ser fácilmente detectados mediante tinción con BrEt. Se debe tener en cuenta que la migración de los oligos de DNA es ligeramente más rápida que la del RNA.

Una vez finalizada la electroforesis, se sumerge el gel en una solución 0,5x TBE; 0,5 µg/ml BrEt, durante 15 min, con tal de poder visualizar los ácidos nucleicos migrados. Posteriormente, se puede desteñir el gel sumergiéndolo en TBE 0,5x durante otros 15-20 min en agitación.

**<sup>1</sup>Gel desnaturalizante de poliacrilamida:** 12% poliacrilamida; 7 M urea; 0,5x TBE.

Preparación del gel (40 ml):

1. Mezclar 12 ml de 40% poliacrilamida (acrilamida:bisacrilamida, 29:1), 16,8 g de urea y 2 ml de 10x TBE. Llevar a 37°C para disolver bien la urea.
2. Enrasar a 40 ml con H<sub>2</sub>O estéril y mezclar bien, con ayuda del vórtex.
3. Añadir 320 µl de APS al 10% y mezclar bien.
4. Añadir 15 µl de TEMED y mezclar bien.
5. Verter de inmediato en el molde, evitando la formación de burbujas y dejar polimerizar unos 30 min.

**<sup>2</sup>TBE (Tris Borato pH 8,3) 10x:** 0,89 M Tris base; 0,89 M Ácido Bórico; 0,1 M EDTA pH 8,0.

**<sup>3</sup>Tampón colorante de carga en gel 10x:** 50% (v/v) glicerol; 10 mM EDTA pH 8,0; 0,25% (w/v) azul de bromofenol; 0,25% (w/v) xylene cianol FF.

### 5.3. PURIFICACIÓN DE LOS sRNAs POR ELUCIÓN DESDE GEL DE POLIACRILAMIDA.

La purificación de los sRNAs migrados en un gel de poliacrilamida se ha realizado mediante su elución desde el gel con NaCl 0,3 M, como se indica a continuación:

11. Cortar la región del gel correspondiente al tamaño de los sRNAs que se quieren obtener y trocearla en pedazos muy pequeños.

## Material y métodos.

12. Sumergir completamente estos pedazos en NaCl 0.3 M e incubar a 4°C en agitación durante toda la noche.
13. Recuperar la fase acuosa, pasándola a un nuevo vial y descartar los trozos de acrilamida.
14. Precipitar los sRNAs en NH<sub>4</sub>OAc 0,5 M (final); EtOH 70% (final); glicógeno 0,1 µg/µl (final). Mezclar bien e incubar 2 h a -20°C.
15. Centrifugar a 14000 rpm durante 30 min a 4°C.
16. Eliminar el sobrenadante y lavar el precipitado con EtOH 70%.
17. Centrifugar a 14000 rpm durante 15 min a 4°C.
18. Eliminar el sobrenadante y dejar secar el precipitado.
19. Resuspender en 20 µl de H<sub>2</sub>O estéril. Guardar en hielo.

### 5.4. TÉCNICA DE NORTHERN.

Debido al pequeño tamaño de los sRNAs se ha modificado la técnica original de Northern (descrita en el apartado 4) para optimizar la resolución e hibridación de estos pequeños RNAs.

#### 5.4.1. ELECTROFORESIS EN GEL DESNATURALIZANTE DE POLIACRILAMIDA.

Se han corrido 20-25 µg de LMW-RNA de cada muestra en un gel desnaturizante de poliacrilamida al 12%, como se describe en el apartado 5.2.

#### 5.4.2. TRANSFERENCIA.

La transferencia de los ácidos nucleicos desde el gel de poliacrilamida a la membrana de nylon se ha realizado utilizando el Trans-Blot SD Semi-Dry Transfer Cell (BIO-RAD), aplicando una intensidad de 400 mA durante 1 h. Posteriormente se ha fijado el RNA a la membrana mediante irradiación con luz U.V., como se indica en el apartado 3.3.

#### 5.4.3. HIBRIDACIÓN.

Se ha realizado una prehibridación de la membrana (conteniendo los RNAs) con 12 ml de la solución de hibridación comercial PerfectHyb™ Plus Hybridization Buffer (Sigma) a 37°C en agitación durante 1-2 h. Posteriormente se ha hibridado la membrana con la sonda marcada (ver apartado 2.12 y 2.13) y desnaturizada (5 min a 100°C) en 10 ml de PerfectHyb™ Plus Hybridization Buffer (Sigma) a 37°C en agitación durante 16-18 h. Finalmente, se ha lavado la membrana con una solución 2x SSC; 0,2% SDS, 3 veces a 50°C durante 15 min en agitación.

#### 5.4.4. EXPOSICIÓN Y REVELADO.

La exposición de la membrana hibridada y su posterior revelado se han llevado a cabo como se describe en el apartado 3.5.

### 5.5. TÉCNICA DE NORTHERN REVERSO.

La técnica de Northern reverso es una variación de la de Northern convencional. La peculiaridad de este método es que en la membrana se fija el DNA y la sonda utilizada es de RNA. En el caso concreto de los Northern reversos realizados en el trabajo que se explica en esta memoria, se han utilizado como sonda moléculas de sRNAs de tabaco.



## 6. MÉTODOS PARA EL TRATAMIENTO DEL MATERIAL VEGETAL.

### 6.1. INDUCCIÓN DE LA EXPRESIÓN DEL RETROTRANSPOSÓN Tnt1 DE TABACO.

La inducción de la expresión del retrotransposón Tnt1 de tabaco se ha realizado en hojas jóvenes (tercera hoja, generalmente) de plantas de *Nicotiana* de 5-6 semanas de edad (presentan 5-6 hojas verdaderas) crecidas en invernadero. Se hiere el reverso de la hoja, previamente esterilizada en solución al 5% (v/v) NaClO y 0.1% (v/v) Tween-20® durante 10 min, con ayuda de un bisturí y se incuba durante 6 horas en oscuridad en medio T<sub>0</sub> (ver apartado 1.4.) en presencia de 0,1% (w/v) celulasa R-10 de Onozuka (Yakult Honsha CO., LTD.).

### 6.2. TRANSFORMACIÓN DE *Nicotiana tabacum*.

#### 6.2.1. TRANSFORMACIÓN ESTABLE DE TABACO.

La transformación estable persigue conseguir la generación de plantas individuales que contengan, integrada en el genoma de cada una de sus células, la construcción génica (transgén) foránea que se quiere estudiar.

El método seguido en este trabajo para la transformación estable de plantas de *Nicotiana tabacum* var. *Xanthi* mediante infección por *Agrobacterium tumefaciens* (cepa LBA4404) se detalla a continuación.

##### 6.2.1.1. Preparación del material bacteriano.

1. Crecer un precultivo de *A.tumefaciens* LBA4404 transformado con el plásmido binario (conteniendo la construcción génica de interés) en 3 ml de medio YEB<sup>1</sup> (100 µg/ml Rif; 300 µg/ml Sm; 50 µg/ml Kan) durante 48-72 h a 28°C en agitación. No añadir la resistencia del plásmido binario (Kan) en el cultivo que se utilizará como control de transformación.
2. Inocular 1 ml del precultivo crecido en 100 ml de YEB (100 µg/ml Rif; 300 µg/ml Sm; 50 µg/ml Kan) y dejar crecer durante toda la noche (unas 21 h) a 28°C en agitación.
3. Dejar crecer hasta que la DO<sub>600nm</sub> alcance un valor comprendido entre 0,6 y 0,8.
4. Centrifugar a 4500 rpm durante 15 min.
5. Eliminar el sobrenadante y resuspender inmediatamente el sedimento bacteriano en 50 ml de medio MS líquido.
6. Incubar 20 min a 28°C en agitación para que las bacterias se recuperen y activen los genes de virulencia.

<sup>1</sup>Medio YEB: ver apartado 1.4.

##### 6.2.1.2. Preparación del material vegetal.

Se han crecido plantas de *Nicotiana tabacum* variedad *Xanthi* en condiciones *in vitro* durante unas 8 semanas. Para la transformación se utilizan hojas jóvenes pero bien desarrolladas: la tercera o cuarta a partir del ápice.

## Material y métodos.

1. Cortar la hoja y colocarla con el reverso hacia arriba en una placa de Petri con 20 ml de medio MS<sup>2</sup>.
2. Con ayuda de un bisturí, eliminar los bordes (en primer lugar) y los nervios (después) de la hoja.
3. Hacer cuadrados de 0,5 cm de lado con los 4 lados heridos. (De una hoja se obtienen 15-20 cuadrados; para cada construcción son necesarios 50 cuadrados procedentes de 2-3 hojas).
4. Traspasar los cuadrados de hoja a una placa con 7 ml de medio MS (con el reverso de la hoja hacia arriba).

<sup>2</sup>Medio MS: ver apartado 1.4.

### 6.2.1.3. Co-cultivo.

1. Añadir 3 ml del cultivo de *A.tumefaciens* a los 7 ml de MS con los trozos de hoja.
2. Co-incubar los trozos de hoja con el *Agrobacterium* durante 5 min, agitando periódicamente.
3. Secar los trozos de hoja con papel de filtro estéril.
4. Transferir los trozos de hoja a placas Corning® conteniendo medio MS<sup>3</sup>, colocándolas con el anverso hacia arriba.
5. Poner 10 trozos por placa. En total habrá 5 placas (50 trozos de hoja) por construcción, más una placa (10 discos) para el control de transformación.
6. Incubar en la oscuridad durante 48 h a 26°C.

<sup>3</sup>Medio MS': medio MS/gelrita (ver apartado 1.4.) complementado con hormonas; 1 µg/ml BAP; 0,1 µg/ml NAA.

### 6.2.1.4. Inducción de callos transgénicos.

Tras la incubación en oscuridad:

1. Observar los trozos de hoja bajo la lupa para comprobar y evaluar el crecimiento de *Agrobacterium*. Se ha de observar un crecimiento de la bacteria alrededor de la hoja, sin que sea excesivo. Si la contaminación es excesiva se deben limpiar las hojas con medio líquido MS con 200 µg/ml timentin (ver apartado 1.5.).
2. Si la contaminación bacteriana no es excesiva, se pasan los trozos de hoja a placas Corning® conteniendo medio selectivo MSI<sup>4</sup> (con hormonas para la inducción de callo, antibiótico de selección y timentin (ver apartado 1.5.), para eliminar el *Agrobacterium*), secándolas, previamente, en papel de filtro estéril.
3. Incubar a 26°C con luz durante unas 2-3 semanas. (En este periodo las hojas se curvan y endurecen. Poco a poco comenzarán a aparecer callos en los bordes).
5. Cuando los callos alcancen un tamaño de unos milímetros, se cortan los trozos de hojas con tal de aislar los diferentes callos y se colocan en nuevas placas con medio MSI.
6. Incubar a 26°C con luz.

<sup>4</sup>Medio MSI: Medio MS' (ver apartado 6.2.1.3.) complementado con antibióticos; 100 µg/ml timentin; 25 µg/ml higromicina B.

### 6.2.1.5. Regeneración de plantas a partir de los callos.

Después de una o dos semanas aparecen espontáneamente los primeros brotes a partir de los callos. Cuando los brotes están bien desarrollados (tallo con un par o tres de hojas) se cortan por el tallo, asegurando de coger un brote individual y se traspasan a tubos de cultivo con medio MS<sup>1</sup>/<sub>2</sub><sup>5</sup>. Es importante no traspasar ningún trozo de callo. En estas condiciones los brotes echan raíces en una o dos semanas y siguen creciendo desarrollándose una planta completa. Estas plantas con raíces bien desarrolladas pueden ser transplantadas a jarras de cultivo *in vitro* para que sigan creciendo. Y, posteriormente, cuando tienen raíces bien desarrolladas pueden ser transplantadas a maceta en el invernadero, pasando por condiciones de alta humedad (protegidas por cúpulas translúcidas) durante 1 semana, para aclimatarlas a las condiciones del invernadero.

<sup>5</sup>**Medio MS<sup>1</sup>/<sub>2</sub>**: Modificación del medio MS (ver apartado 6.2.1.2.) en la que se han disminuido las sales y vitaminas a la mitad y se han aumentado tres veces los hidratos de carbono. 0,22% (w/v) MS+Vitaminas (*Duchefa*); 3% (w/v) Sacarosa; 0,05% (w/v) MES; 0,8% (w/v) agar. Ajustar a pH 5,8 con KOH. Autoclavar. Complementar con antibióticos; 100 µg/ml timentin; 25 µg/ml higromicina B.

### 6.2.2. TRANSFORMACIÓN TRANSITORIA DE TABACO POR AGROINFILTRACIÓN.

La técnica de la agroinfiltración ha sido utilizada en este trabajo para conseguir la transformación, de manera transitoria, de ciertas partes de la planta. Este método de transformación consigue la incorporación del transgén en el genoma de las células del área de la planta que ha sido agroinfiltrada, a diferencia de la transformación estable donde todas las células del organismo contienen el transgén.

El protocolo seguido en la técnica de agroinfiltración ha sido el siguiente:

1. Aislar colonias transformadas de *A.tumefaciens* en placas conteniendo el medio YEB selectivo (YEB + antibióticos correspondientes).
2. Inocular un precultivo de 5 ml de medio YEB selectivo (100 µg/ml Rif; 300 µg/ml Sm; 50 µg/ml Kan; para la cepa y plásmidos utilizados en este trabajo, ver apartados 1.2 y 1.3) a partir de una colonia transformante aislada. Crecer durante 24-48 h a 28°C en agitación.
3. Inocular un agro-cultivo<sup>1</sup> de 30 ml a partir de 3 ml del precultivo. Crecer durante 18-20 h a 28°C en agitación.
4. Centrifugar los 30 ml de cultivo a 4500 rpm durante 15 min.
5. Resuspender el sedimento celular en agro-buffer<sup>2</sup> hasta una DO<sub>600nm</sub> de 0,5.
6. Incubar los cultivos diluidos a temperatura ambiente durante toda la noche (14-16 h).
7. Proceder a infiltrar las plantas. En este trabajo se han utilizado jeringas de polipropileno con las que se inyecta el medio de *Agrobacterium* presionando suavemente sobre el reverso de las hojas de la planta.

Posteriormente se han dejado crecer las plantas agroinfiltradas en el invernadero y se ha recogido la muestra a los 5 días, a partir de la infiltración.

<sup>1</sup>**Agro-cultivo**: 10 mM MES pH 5,7<sup>3</sup>; 20 µM Acetosyringone<sup>4</sup>; 100 µg/ml Rif; 300 µg/ml Sm; 50 µg/ml Kan.

<sup>2</sup>**Agro-buffer**: 10 mM MgCl<sub>2</sub>; 10 mM MES pH 5,7<sup>3</sup>; 150 µM Acetosyringone<sup>4</sup>.

<sup>3</sup>**1 M MES pH 5,7**: 19,5 g de MES en 100 ml H<sub>2</sub>O. Llevar a 37°C para disolverlo bien. Ajustar a pH 5,7 con 10 N NaOH. Esterilizar por filtración. Guardar a -20°C.

<sup>4</sup>**100 mM acetosyringone** (3,5-dimethoxy-4-hydroxyacetophenon): 0,2 g de acetosyringone en 10 ml DMSO. Calentar para disolverla.

### 6.2.3. TRANSFORMACIÓN TRANSITORIA DE HOJAS DE TABACO POR BIOLÍSTICA.

El análisis de la terminación de la transcripción y poliadenilación de las distintas construcciones génicas conteniendo las regiones U3 de las LTRs de Tnt1 se ha realizado *in vivo* transformando hojas de tabaco mediante biolística.

#### 6.2.3.1. Preparación de la reserva de partículas de oro.

1. Pesar con precisión 60 mg de partículas de oro ( $\varnothing = 1 \mu\text{m}$ , BioRad).
2. Lavar las partículas con EtOH. Añadir 1 ml de EtOH HPLC<sup>1</sup>. Mezclar agitando en vórtex durante 10 minutos aproximadamente. Centrifugar 1 minuto a 10000 rpm.
3. Lavado con H<sub>2</sub>O destilada estéril (x4 veces). Añadir 1 ml de H<sub>2</sub>O destilada estéril. Mezclar agitando en vórtex durante 1 minuto. Dejar reposar las partículas 5-10 minutos. Centrifugar 1 minuto a 10000 rpm.
4. Resuspender las partículas de oro. Resuspender el precipitado en 1 ml de glicerol 50% estéril. Guardar a 4°C. Siempre guardar la siguiente relación: 60 mg de partículas de Au / 1 ml de EtOH, H<sub>2</sub>O destilada estéril o glicerol 50%. Es importante mantener siempre la misma relación para que el número de partículas de Au por  $\mu\text{l}$  sea constante. Si no fuera así existirían diferencias en la eficiencia de transformación.

<sup>1</sup>El etanol ha de ser de calidad y completamente deshidratado, para evitar la agregación de las partículas de oro. Utilizar EtOH HPLC de botella nueva (Scharlau®). Alicuotarlo en tubos Falcon™ de 15 ml (preferiblemente) o botellas pequeñas con buen cierre. Guardar las alícuotas en un desecador.

#### 6.2.3.2. Recubrimiento de las partículas de oro con DNA, por precipitación.

1. Presonificar las partículas de oro durante 3 minutos en el baño a 4°C (cámara fría).
2. Alicuotar 15  $\mu\text{l}$  de partículas en cada vial de microcentrifuga mezclando y agitando en vórtex entre paso y paso. (1 vial de 1.5 ml por hoja o muestra a bombardear).
3. Mezclar agitando en vórtex 10 segundos.
4. Añadir los 20  $\mu\text{l}$  de la mezcla de DNAs<sup>1</sup> a los 15  $\mu\text{l}$  de partículas de Au. Mezclar agitando en vórtex durante 10 segundos.
5. Añadir 50  $\mu\text{l}$  de CaCl<sub>2</sub> 2,5 M. Mezclar agitando en vórtex durante 10 segundos.
6. Añadir 20  $\mu\text{l}$  de espermidina<sup>2</sup> 0,1 M. Mezclar agitando en vórtex durante 30 segundos. Dejar reposar 15 minutos en hielo.
7. Dar un toque de centrifuga y eliminar el sobrenadante. El toque de centrifuga ha de ser muy leve para evitar que se formen aglomerados de partículas de oro.
8. Añadir 500  $\mu\text{l}$  de EtOH HPLC. Mezclar agitando en vórtex durante 30 segundos. Dejar reposar 15 minutos en hielo.
9. Dar un toque de centrifuga y eliminar el sobrenadante.
10. Añadir 200  $\mu\text{l}$  de EtOH HPLC. Mezclar agitando en vórtex durante 30 segundos. Dejar reposar 15 minutos en hielo.
11. Dar un toque de centrifuga y eliminar el sobrenadante.
12. Añadir 100  $\mu\text{l}$  de EtOH HPLC. Mezclar agitando en vórtex durante 30 segundos. Dejar reposar 15 minutos en hielo.
13. Dar un toque de centrifuga y eliminar el sobrenadante.
14. Añadir 25  $\mu\text{l}$  de EtOH HPLC.
15. Guardar en hielo hasta el momento de bombardear.



## Material y métodos.

**<sup>1</sup>Mezcla de DNAs:** añadir 3-4 ng del DNA problema en 20  $\mu$ l de volumen final de H<sub>2</sub>O. En aquellos casos que se crea conveniente se pueden añadir plásmidos control (1 ng), que contengan genes marcadores como la luciferasa, para comprobar la eficiencia de la transformación.

**<sup>2</sup>Espermidina 100 mM.** Calentar el bote de espermidina *free-base* a 65°C durante 5 minutos. Hacer alícuotas de 14  $\mu$ l. Guardar a -20°C. Justo antes de utilizarla, añadir 986  $\mu$ l de H<sub>2</sub>O destilada estéril.

### 6.2.3.3. Proceso de transformación de las hojas de tabaco: bombardeo.

Para la transformación transitoria de hojas de tabaco se ha utilizado una pistola de partículas PDS-1000/He (BioRad®).

1. Coger hojas jóvenes (primeras o segundas) de plantas de *N. tabacum* var. *Xanthi* crecidas en el fitotrón (en oscuridad) o en sombra y sumergir el ápice en H<sub>2</sub>O.
2. Esterilizar el material: los filtros y rejillas de macrocarrier con EtOH absoluto y los discos de ruptura (1100 psi) con isopropanol.
3. Preparar placas de Petri pequeñas con 3 ml de medio T<sub>0</sub>.
4. Abrir bomba de Helio hasta un poco más de 1100 psi y abrir bomba de vacío.
5. Colocar la muestra sobre los filtros. Para ello, antes hay que sonicar cada vial de microcentrífuga conteniendo las partículas de oro recubiertas del DNA, sumergiéndolos en el baño de ultrasonidos durante 15 segundos (x3 veces) y mezclar agitando con vórtex durante 10 segundos. Inmediatamente pipetear 11  $\mu$ l de la mezcla y verter en un filtro, seguidamente coger otros 11  $\mu$ l y poner en otro filtro (la misma muestra en dos discos porque bombardearemos haz y envés de la hoja). Apurar todo el volumen, en el caso de que quedara muestra en el vial.
6. Repetir paso 5 para cada muestra (mezcla de DNAs).
7. Dejar secar completamente.
8. Montar el macrocarrier y cargar el disco de ruptura.
9. Colocar la hoja a bombardear sobre una placa de Petri grande y esta última sobre el soporte en el penúltimo nivel.
10. Cerrar la cámara de vacío.
11. Hacer el vacío hasta la marca y mantenerlo.
12. Disparar. Mantener hasta que la presión suba a 1100 psi; entonces se oye el disparo.
13. Disparar también sobre la otra cara de la hoja.
14. Cortar un disco de la hoja, donde ésta haya recibido el impacto. Colocar el disco sobre la placa de Petri pequeña con el medio.
15. Dejar incubar en oscuridad a 26°C toda la noche.
16. Congelar los discos en caso de que no se vaya a trabajar con ellos inmediatamente.

## 7. ENSAYO DE INMUNOPRECIPITACIÓN DE LA CROMATINA (ChIP).

El método de inmunoprecipitación de la cromatina (ChIP) permite conocer las asociaciones existentes *in vivo* entre determinadas proteínas (como las histonas modificadas) y determinadas secuencias del DNA, a las que se unen (Kuo MH and Allis CD, 1999; Bowler C *et al.*, 2004). El ensayo de ChIP se divide en dos etapas, el entrecruzamiento de las moléculas de proteínas y DNA en presencia de formaldehído y la inmunoprecipitación de los complejos proteína-DNA con anticuerpos específicos a partir de extractos sonicados para fragmentar la cromatina. Finalmente, los fragmentos de cromatina inmunoprecipitados pueden ser analizados por PCR, utilizando oligonucleótidos específicos.

Mediante esta técnica, en este trabajo, se ha querido conocer qué tipo de modificaciones de las histonas están asociadas a las secuencias del retrotransposón Tnt1.

## 7.1. UNIÓN COVALENTE DE PROTEÍNAS Y DNA.

El entrecruzamiento o establecimiento de uniones covalentes entre las proteínas y la molécula de DNA se consigue gracias a la presencia del formaldehído

1. Sumergir la muestra (hojas) en 50 ml de tampón de cross-link<sup>1</sup>.
2. Infiltrar al vacío durante 10 min.
3. Añadir 5 ml de glicina 1M y mezclar.
4. Infiltrar al vacío 5 min.
5. Eliminar el líquido.
6. Lavar las muestras con H<sub>2</sub>O estéril (x2).
7. Secar y congelar en N<sub>2</sub> líquido.
8. Pulverizar las muestras en presencia de N<sub>2</sub> líquido.
9. Pasar el polvo congelado a un vial de 50 ml.
10. Añadir 10 ml de tampón A<sup>2</sup> y mezclar suavemente.
11. Filtrar, a través de Miracloth<sup>®</sup> en embudo, sobre otro Falcon de 50 ml.
12. Centrifugar 15 min a 4°C a 2800 rpm (precipitan núcleos más cloroplastos).
13. Eliminar sobrenadante con cuidado.
14. Resuspender el precipitado en 0,5-1 ml de tampón de lisis<sup>3</sup>.
15. Guardar a 4°C.

<sup>1</sup>**Tampón de cross-link:** 13,6% (w/v) sacarosa; 1 mM Tris-HCl pH 8; 0,05% (v/v) Tritón X-100<sup>®</sup>; 1% (v/v) formaldehído (añadir el formaldehído en el momento de utilizar).

<sup>2</sup>**Tampón A:** 8,5% (w/v) sacarosa; 10 mM KCl; 10 mM MgCl<sub>2</sub>; 10 mM Hepes pH 7,8; 5 mM EDTA pH 8; 0,5% (v/v) Tritón X-100<sup>®</sup>; 200 μM DTT (añadir DTT en el momento de utilizar).

<sup>3</sup>**Tampón de lisis:** 15 mM Tris-HC pH 7,4; 60 mM KCl; 2 mM EDTA pH 8; 1mM DTT (añadir DTT en el momento de utilizar).

## 7.2. INMUNOPRECIPITACIÓN DE LOS COMPLEJOS PROTEÍNA-DNA.

Una vez entrecruzada la cromatina, se sonicar los núcleos lisados para obtener fragmentos de DNA de unos 500-1000 pb. Estos fragmentos y los complejos de proteínas unidos a éstos son inmunoprecipitados con anticuerpos que reconocen las proteínas de interés.

1. Sonicar 350 μl de cada muestra. En un sonicador Branson<sup>®</sup> someter la muestra a tres ciclos de 10 minutos de sonicación al 30% de amplitud.
2. Centrifugar las muestras sonicadas a 14000 rpm durante 15 minutos, para eliminar los residuos celulares.
3. Recuperar el sobrenadante, que contienen los fragmentos de cromatina solubles, colocando 100 μl del extracto por vial de microcentrífuga de 1,5 ml.
4. Añadir 900 μl de tampón de inmunoprecipitación<sup>1</sup> en cada vial.
5. Hacer un prelavado añadiendo 30 μl de ProtA-sefarosa, 2 μg de ssDNA (para bloquear la ProtA-sefarosa) y, opcionalmente, 1 μg de anticuerpo pre-inmune. Tanto la ProtA-sefarosa como el anticuerpo pre-inmune unirán aquellas otras proteínas que hayan podido quedar presentes en la muestra, que puedan interferir en el proceso posterior al poder establecer uniones inespecíficas.

## Material y métodos.

6. Incubar durante 1 h a 4°C en rotación.
7. Centrifugar a 3000 rpm durante 2 minutos.
8. Pasar el sobrenadante a un nuevo vial de microcentrífuga y añadir el anticuerpo específico para la proteína que se quiera precipitar.
9. Incubar durante 12-16 h (toda una noche) a 4°C en rotación.
10. Añadir 60 µl de ProtA-sefarosa y 2 µg de ssDNA.
11. Incubar durante 2,5 h a 4°C en rotación.
12. Centrifugar a 3000 rpm durante 2 minutos y eliminar la fase superior.
13. Lavar la fase inferior (más densa, conteniendo los complejos de la proteína de interés unida al DNA, a su anticuerpo y a la protA-sefarosa) con tampón TSE 1<sup>2</sup> durante 10 minutos a 4°C en rotación.
14. Centrifugar a 3000 rpm durante 2 minutos y eliminar la fase superior.
15. Lavar la fase inferior con tampón TSE 2<sup>3</sup> durante 10 minutos a 4°C en rotación.
16. Centrifugar a 3000 rpm durante 2 minutos y eliminar la fase superior.
17. Lavar la fase inferior con tampón TSE 3<sup>4</sup> durante 10 minutos a 4°C en rotación.
18. Centrifugar a 3000 rpm durante 2 minutos y eliminar la fase superior.
19. Lavar (x3) la fase inferior con TE<sup>5</sup> (se puede hacer a temperatura ambiente).
20. Eluir (x3), agitando en el vórtex durante 2 minutos, con 100 µl de tampón de elución<sup>6</sup>.
21. Centrifugar a 4000 rpm durante 2 minutos y pasar la fase superior (conteniendo los fragmentos de DNA que se han liberado de la unión covalente a las proteínas) a un nuevo vial de microcentrífuga.
22. Eliminar las posibles proteínas solubles mediante la utilización de fenol/cloroformo o pasando la muestra a través de columnas de purificación de productos de PCR (ver apartados 2.1. y 2.4.).

Una vez se dispone de una solución limpia conteniendo sólo el DNA, que estaba unido a las proteínas inmunoprecipitadas, se puede realizar el análisis por PCR de estos fragmentos.

<sup>1</sup>**Tampón de inmunoprecipitación:** 1% (v/v) Tritón; 2 mM EDTA pH 8,0; 150 mM NaCl; 20 mM Tris-HCl pH 8,0; 100 µM PMSF; 10 µg/ml aprotinina; 21 µM leupeptina.

<sup>2</sup>**TSE 1:** 2 mM EDTA pH 8,0; 0,1% (w/v) SDS; 1% (v/v) Tritón X-100®; 150 mM NaCl; 20 mM Tris-HCl pH 8,0.

<sup>3</sup>**TSE 2:** 2 mM EDTA pH 8,0; 0,1% (w/v) SDS; 1% (v/v) Tritón; 500 mM NaCl; 20 mM Tris-HCl pH 8,0.

<sup>4</sup>**TSE 3:** 0,25 M LiCl; 1% (v/v) NP40; 1% (w/v) deoxycholate sódico; 1 mM EDTA pH 8,0; Tris-HCl pH 8,0.

<sup>5</sup>**TE, pH 8,0:** 100 mM Tris-HCl pH 8,0; 10 mM EDTA pH 8,0. Autoclavar durante 20 minutos a 15 psi y guardar a temperatura ambiente.

<sup>6</sup>**Tampón de elución:** 1% (w/v) SDS; 0,1 M NaHCO<sub>3</sub>. Preparar en el momento de utilización.

## 8. MÉTODOS BIOINFORMÁTICOS.

El análisis informático presentado en esta memoria ha requerido la utilización de toda una serie de programas, paquetes de programas y bases de datos. Algunos de estos programas han sido desarrollados específicamente para realizar el trabajo descrito en esta memoria en colaboración con el grupo dirigido por el profesor Xavier Messeguer del Departament de Llenguatges i Sistemes Informàtics (LSI) de la Universitat Politècnica de Catalunya (UPC). Estas aplicaciones informáticas se describen a continuación y están disponibles en el sitio web del LSI (<http://algggen.lsi.upc.es>).

## 8.1. HERRAMIENTAS INFORMÁTICAS.

El programa **TRANSP**O (Herráiz C, 2001; Santiago N *et al.*, 2002) ha sido diseñado para encontrar cualquier par de secuencias que se encuentren en sentido inverso separadas una cierta distancia, sobre cualquier secuencia genómica, génica o BAC que se le proporcione. Se trata de una implementación del algoritmo desarrollado por Myers (1998), que encuentra todos los sitios en los que una secuencia en cuestión (la secuencia del TIR de los elementos *Emigrant*, en este caso) coincide aproximadamente con otra secuencia dada (la secuencia de los cinco cromosomas de *arabidopsis*).

**SIMMAT** (**SIM**milarity **MAT**rix) es un programa diseñado para generar una matriz de similitud entre secuencias de ADN, de forma independiente respecto al resto de las aplicaciones utilizadas en el alineamiento de las secuencias (Goñi JR, 2002; Santiago N *et al.* 2002).

**SPAT** (**SPA**wning **T**ree clustering method) es un programa diseñado para agrupar secuencias en conjuntos, eliminando las más divergentes de cada grupo determinado, para poder aplicar análisis filogenéticos convencionales sobre aquellas que se asemejan entre sí (Goñi JR, 2002; Santiago N *et al.*, 2002).

**CLUPH** (**CLU**stering **grA**PH) dibuja un mapa de la localización de las secuencias de ADN clasificadas en grupo, según la posición que ocupan en un gen, cromosoma o genoma (Goñi JR, 2002; Santiago N *et al.*, 2002).

## 8.2. BÚSQUEDA E IDENTIFICACIÓN DE ELEMENTOS *Emigrant*.

Se ha utilizado el programa **TRANSP**O para buscar, en toda la secuencia disponible del genoma de *arabidopsis* ([www.arabidopsis.org](http://www.arabidopsis.org)), secuencias repetidas invertidas que presentan al menos un 75% de homología con los 20 primeros nucleótidos del TIR consenso de *Emigrant* (Casacuberta E *et al.*, 1998) y están separadas entre sí una distancia comprendida entre 200 y 700 nt. Se han eliminado los elementos coincidentes generados a partir de secuencias repetidas invertidas subterminales. Se ha calculado la matriz de similitud de las secuencias encontradas mediante la aplicación **SIMMAT**. Estas secuencias han sido agrupadas utilizando el programa **SPAT**. El programa **CLUPH** ha generado una distribución gráfica de los diferentes **MITE**s *Emigrant*, divididos en las familias previamente definidas, en los 5 cromosomas de *arabidopsis*. Con el fin de obtener información de los **ORF**s situados cerca de los elementos *Emigrant*, así como de la posible existencia de **RES**ites, se han utilizado como muestra los 30 nt que flanquean a cada uno de ellos, tanto en 5' como en 3', en búsquedas de similitud de secuencia utilizando la aplicación **BLAST** (versión 2.0; Altschul SF *et al.*, 1990: <http://www.ncbi.nlm.nih.gov/blast/>) sobre el **GenBank** (Nacional Center for Biotechnology Information, **NCBI**; <http://www.ncbi.nlm.nih.gov>).

## 8.3. ANÁLISIS FILOGENÉTICO.

Aquellas secuencias agrupadas dentro de un mismo grupo por el programa **SPAT** fueron alineadas con el programa de alineamiento múltiple **CLUSTAL W**, tomando los parámetros que aparecen por defecto (versión 1.5; Thompson JD *et al.*, 1994), con pequeños refinamientos. La visualización y optimización de los alineamientos nucleotídicos se realizó mediante el programa **GENEDOC** (versión 2.6.002; Nicholas KB and Nicholas HB Jr., 1997). La matriz de distancia basada en el algoritmo de **Jukes-Cantor** (Jukes TH and Cantor CR, 1969) se generó utilizando el

## Material y métodos.

programa DNADIST del paquete de programas de Felsenstein PHYLIP (Felsenstein J, 1989). Esta matriz se utilizó para generar los árboles filogenéticos por el método de neighbor-joining (Saitou N and Nei M, 1987). Los análisis de fiabilidad de las relaciones (bootstrap) establecidas por los programas anteriores se llevaron a cabo mediante los programas Seqboot y Consense, también pertenecientes al paquete PHYLIP (Felsenstein J, 1989). La variabilidad existente entre las diferentes secuencias, medida como la diversidad nucleotídica ( $\pi$ ) de Nei (Nei M, 1987), y su desviación estándar se calcularon a través del programa DnaSP (Rozas J and Rozas R, 1999).

# **BIBLIOGRAFÍA.**

## BIBLIOGRAFÍA.

**Abad J. P., B. De Pablos, K. Osoegawa, P. J. De Jong, A. Martin-Gallardo, and A. Villasante.** Genomic analysis of *Drosophila melanogaster* telomeres: full-length copies of HeT-A and TART elements at telomeres. *Mol Biol Evol* 21 (9):1613-1619, 2004.

**Agrawal A., Q. M. Eastman, and D. G. Schatz.** Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394 (6695):744-751, 1998.

**Altschul S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** Basic local alignment search tool. *J Mol Biol* 215 (3):403-410, 1990.

**Ambros V., R. C. Lee, A. Lavanway, P. T. Williams, and D. Jewell.** MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol* 13 (10):807-818, 2003.

**An G., P. R. Ebert, A. Mitra, and S. B. Ha.** *Binary Vectors*. In: *Plant Molecular Biology Manual*. Gelvin, S. B., and Schilperoort, R. A. (eds.) pp. A3: 1-19, Kluwer Academic Publishers, Dordrecht, Netherlands, 1988.

**Arabidopsis Genome Initiative (AGI).** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408 (6814):796-815, 2000.

**Aravin A. A., M. Lagos-Quintana, A. Yalcin, M. Zavolan, D. Marks, B. Snyder, T. Gaasterland, J. Meyer, and T. Tuschl.** The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 5 (2):337-350, 2003.

**Ares M. Jr and N. J. Proudfoot.** The spanish connection: transcription and mRNA processing get even closer. *Cell* 120 (2):163-166, 2005.

**Arkipova I. R., K. I. Pyatkov, M. Meselson, and M. B. Evgen'ev.** Retroelements containing introns in diverse invertebrate taxa. *Nat Genet* 33 (2):123-124, 2003.

**Ausubel F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G Seidman, J. A. Smith, and K. Struhl.** *Current Protocols in Molecular Biology*. John Willey & sons eds., 2001.

**Avramova Z., A. Tikhonov, M. Chen, and J. L. Bennetzen.** Matrix attachment regions and structural colinearity in the genomes of two grass species. *Nucleic Acids Res* 26 (3):761-767, 1998.

**Bachman N., M. E. Gelbart, T. Tsukiyama, and J. D. Boeke.** TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of lsw2p to *S. cerevisiae* tDNAs. *Genes Dev* 19 (8):955-964, 2005.

**Beguiristain T., M. A. Grandbastien, P. Puigdomenech, and J. M. Casacuberta.** Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiol* 127 (1):212-221, 2001.

**Bennetzen J. L.** Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* 42 (1):251-269, 2000.

**Bennetzen J. L.** Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr Opin Genet Dev* 15 (6):621-627, 2005.

- Bevan M.** Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Res* 12 (22):8711-8721, 1984.
- Birnboim H. C. and J. Dolly.** A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513-1523, 1979.
- Blaise S., N. de Parseval, L. Benit, and T. Heidmann.** Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci U S A* 100 (22):13013-13018, 2003.
- Bourgin J. P., Y. Chupeau, and C. Missioner.** Plant regeneration from mesophyll protoplasts of several *Nicotiana* species. *Physiol.Plant* 45:288-292, 1979.
- Bowen N. J. and I. K. Jordan.** Transposable elements and the evolution of eukaryotic complexity. *Curr Issues Mol Biol* 4 (3):65-76, 2002.
- Bowler C., G. Benvenuto, P. Laflamme, D. Molino, A. V. Probst, M. Tariq, and J. Paszkowski.** Chromatin techniques for plant cells. *The Plant Journal* 39:776-789, 2004.
- Brandt J., S. Schrauth, A. M. Veith, A. Froschauer, T. Haneke, C. Schultheis, M. Gessler, C. Leimeister, and J. N. Volf.** Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. *Gene* 345 (1):101-111, 2005.
- Britten R. J. and D. E. Kohne.** Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 161 (841):529-540, 1968.
- Brookfield J. F.** The ecology of the genome - mobile DNA elements and their hosts. *Nat Rev Genet* 6 (2):128-136, 2005.
- Bureau T. E. and S. R. Wessler .** Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4 (10):1283-1294, 1992.
- Byrd K. and V. G. Corces.** Visualization of chromatin domains created by the gypsy insulator of *Drosophila*. *J Cell Biol* 162 (4):565-574, 2003.
- C.elegans Sequencing Consortium.** Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282 (5396):2012-2018, 1998.
- Cáceres M., J. M. Ranz, A. Barbadilla, M. Long, and A. Ruiz.** Generation of a widespread *Drosophila* inversion by a transposable element. *Science* 285 (5426):415-418, 1999.
- Calvo O. and J. L. Manley.** Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol Cell* 7 (5):1013-1023, 2001.
- Capy P.** Classification and nomenclature of retrotransposable elements. *Cytogenet Genome Res* 110 (1-4):457-461, 2005.
- Casacuberta E., J. M. Casacuberta, P. Puigdomenech, and A. Monfort.** Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the Emigrant family of elements. *Plant J* 16 (1):79-85, 1998.
- Casacuberta E., P. Puigdomenech, and A. Monfort.** Distribution of microsatellites in relation to coding sequences within the *Arabidopsis thaliana* genome. *Plant Science* 157 (1):97-104, 2000.



- Casacuberta J. M. and M. A. Grandbastien.** Characterisation of LTR sequences involved in the protoplast specific expression of the tobacco Tnt1 retrotransposon. *Nucleic Acids Res* 21 (9):2087-2093, 1993.
- Casacuberta J. M., S. Vernhettes, and M. A. Grandbastien.** Sequence variability within the tobacco retrotransposon Tnt1 population. *EMBO J* 14 (11):2670-2678, 1995.
- Casacuberta J. M., S. Vernhettes, C. Audeon, and M. A. Grandbastien.** Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. *Genetica* 100 (1-3):109-117, 1997.
- Casacuberta J. M. and N. Santiago.** Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. *Gene* 311:1-11, 2003.
- Cavalier-Smith T. and M. J. Beaton.** The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica* 106 (1-2):3-13, 1999.
- Chalker D. L. and S. B. Sandmeyer.** Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev* 6 (1):117-128, 1992.
- Chandler M. and J. Mahillon.** Insertion Sequences Revisited. In: *Mobile DNA II*. N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (eds.). ASM Press, p. 305-366. Washington DC, 2002.
- Chimpanzee Sequencing and Analysis Consortium.** Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437 (7055):69-87, 2005.
- Chory J., C. Peto, R. Feinbaum, L. Pratt, and F. Ausubel.** *Arabidopsis thaliana* mutant that develops as a light-grown plant in the absence of light. *Cell* 58 (5):991-999, 1989.
- Church G. M. and W. Gilbert.** Genomic sequencing. *Proc Natl Acad Sci U S A* 81 (7):1991-1995, 1984.
- Collins K. A., S. Furuyama, and S. Biggins.** Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. *Curr Biol* 14 (21):1968-1972, 2004.
- Cosgrove M. S. and C. Wolberger.** How does the histone code work? *Biochem Cell Biol* 83 (4):468-476, 2005.
- Courtial B., F. Feuerbach, S. Eberhard, L. Rohmer, H. Chiapello, C. Camilleri, and H. Lucas.** Tnt1 transposition events are induced by in vitro transformation of *Arabidopsis thaliana*, and transposed copies integrate into genes. *Mol Genet Genomics* 265 (1):32-42, 2001.
- Curcio M. J. and K. M. Derbyshire.** The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* 4 (11):865-877, 2003.
- Damert A., J. Lower, and R. Lower.** Leptin receptor isoform 219.1: an example of protein evolution by LINE-1-mediated human-specific retrotransposition of a coding SVA element. *Mol Biol Evol* 21 (4):647-651, 2004.
- Dawe R. K.** RNA interference, transposons, and the centromere. *Plant Cell* 15 (2):297-301, 2003.
- Dellaporta S. L., J. Wood, and J. B. Hicks.** A plant DNA miniprep: version II. *Plant Mol Bio Rep* 1:19-21, 1983.
- Depicker A., S. Stachel, P. Dhaese, P. Zambryski, and H. M. Goodman.** Nopaline synthase: transcript mapping and DNA sequence. *J Mol Appl Genet* 1 (6):561-573, 1982.

**Devine S. E. and J. D. Boeke.** Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev* 10 (5):620-633, 1996.

**Djikeng A., H. Shi, C. Tschudi, and E. Ullu.** RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA* 7 (11):1522-1530, 2001.

**Doolittle W. F. and C. Sapienza.** Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284 (5757):601-603, 1980.

**Druker R., T. J. Bruxner, N. J. Lehrbach, and E. Whitelaw .** Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Res* 32 (19):5800-5808, 2005.

**Edwards K., C. Johnstone, and C. Thompson.** A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* 19 (6):1349, 1991.

**Eickbush T. H.** Telomerase and retrotransposons: which came first? *Science* 277 (5328):911-912, 1997.

**Fang R. X., F. Nagy, S. Sivasubramaniam, and N. H. Chua.** Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *Plant Cell* 1 (1):141-150, 1989.

**Felsenstein J.** PHYLIP-phylogenetic inference package (version 3.56). *Cladistics* 5:164-166, 1989.

**Feschotte C. and C. Mouches.** Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. *Mol Biol Evol* 17 (5):730-737, 2000.

**Feschotte C. and S. R. Wessler.** Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. *Proc Natl Acad Sci U S A* 98 (16):8923-8924, 2001.

**Feschotte C., X. Zhang, and S. R. Wessler.** Miniature Inverted-repeat Transposable Elements and their relationship to established DNA transposons. In: *Mobile DNA II*. N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (eds.). ASM Press, p. 1147-1158. Washington, D.C., 2002.

**Feschotte C., N. Jiang, and S. R. Wessler.** Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3 ( 5):329-341, 2002.

**Feschotte C., M. T. Osterlund, R. Peeler, and S. R. Wessler.** DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITEs. *Nucleic Acids Res* 33 (7):2153-2165, 2005.

**Finnegan D. J.** Eukaryotic transposable elements and genome evolution. *Trends Genet* 5 (4):103-107, 1989.

**Fleischmann R. D., M. D. Adams , O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, and J. M. Merrick.** Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269 (5223):496-512, 1995.

**Franz P., W. Soppe, and I. Schubert.** Heterochromatin in interphase nuclei of *Arabidopsis thaliana*. *Chromosome Res* 11 (3):227-240, 2005.

**Gale M. D. and K. M. Devos.** Plant comparative genetics after 10 years. *Science* 282 (5389):656-659, 1998.

M. J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419 (6906):498-511, 2002.

Goñi J. R. Desenvolupament d'eines informàtiques per a la classificació i anàlisi de seqüències d'ADN. Facultat d'Informàtica de Barcelona. Universitat Politècnica de Barcelona. 2002.

Goff S. A., D. Ricke, T. H. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma, D. Hadley, D. Hutchison, C. Martin, F. Katagiri, B. M. Lange, T. Moughamer, Y. Xia, P. Budworth, J. Zhong, T. Miguel, U. Paszkowski, S. Zhang, M. Colbert, W. L. Sun, L. Chen, B. Cooper, S. Park, T. C. Wood, L. Mao, P. Quail, R. Wing, R. Dean, Y. Yu, A. Zharkikh, R. Shen, S. Sahasrabudhe, A. Thomas, R. Cannings, A. Gutin, D. Pruss, J. Reid, S. Tavtigian, J. Mitchell, G. Eldredge, T. Scholl, R. M. Miller, S. Bhatnagar, N. Adey, T. Rubano, N. Tusneem, R. Robinson, J. Feldhaus, T. Macalima, A. Oliphant, and S. Briggs. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296 (5565):92-100, 2002.

Grandbastien M. A., C. Audeon, E. Bonnard, J. M. Casacuberta, B. Chalhoub, A. P. Costa, Q. H. Le, D. Melayah, M. Petit, C. Poncet, S. M. Tam, M. A. Van Sluys, and C. Mhiri. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet Genome Res* 110 (1-4):229-241, 2005.

Grandbastien M. A., A. Spielmann, and M. Caboche. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 337 (6205):376-380, 1989.

Grandbastien M. A., H. Lucas, J. B. Morel, C. Mhiri, S. Vernhettes, and J. M. Casacuberta. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. *Genetica* 100 (1-3):241-252, 1997.

Grandbastien M. A. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci* 3 (3):181-187, 1998.

Hamilton A., O. Voinnet, L. Chappell, and D. Baulcombe. Two classes of short interfering RNA in RNA silencing. *EMBO J* 21 (17):4671-4679, 2002.

Hammond S. M. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett* 579 (26):5822-5829, 2005.

Han J. S. and J. D. Boeke. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays* 27 (8):775-784, 2005.

Hanahan D. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166 (4):557-580, 1983.

Hannon G. J. RNA interference. *Nature* 418 (6894):244-251, 2002.

Harikrishnan K. N., M. Z. Chow, E. K. Baker, S. Pal, S. Bassal, D. Brasacchio, L. Wang, J. M. Craig, P. L. Jones, S. Sif, and A. El-Osta. Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat Genet* 37 (3):254-264, 2005.

Hedges D. J. and M. A. Batzer. From the margins of the genome: mobile elements shape primate evolution. *Bioessays* 27 (8):785-794, 2005.

**Herr A. J., M. B. Jensen, T. Dalmay, and D. C. Baulcombe.** RNA polymerase IV directs silencing of endogenous DNA. *Science* 308 (5718):118-120, 2005.

**Herráiz C.** Aplicación web para la búsqueda, análisis y agrupación de transposones. Facultat d'Informàtica de Barcelona. Universitat Politècnica de Barcelona. 2001.

**Himber C., P. Dunoyer, G. Moissiard, C. Ritzenthaler, and O. Voinnet.** Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J* 22 (17):4523-4533, 2003.

**Hiom K., M. Melek, and M. Gellert.** DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94 (4):463-470, 1998.

**Hirochika H.** Activation of tobacco retrotransposons during tissue culture. *EMBO J* 12 (6):2521-2528, 1993.

**Hirochika H., H. Okamoto, and T. Kakutani.** Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation. *Plant Cell* 12 (3):357-369, 2000.

**Hoekema A., P. J. Hooykaas, and R. A. Schilperoort.** Transfer of the octopine T-DNA segment to plant cells mediated by different types of *Agrobacterium* tumor- or root-inducing plasmids: generality of virulence systems. *J Bacteriol* 158 (1):383-385, 1984.

**Holmes-Davis R. and L. Comai.** Nuclear attachment regions and plant gene expression. *Trends Plant Sci* 3 (3):91-97, 1998.

**Holt R. A., G. M. Subramanian, A. Halpern, G. G. Sutton, R. Charlab, D. R. Nusskern, P. Wincker, A. G. Clark, J. M. Ribeiro, R. Wides, S. L. Salzberg, B. Loftus, M. Yandell, W. H. Majoros, D. B. Rusch, Z. Lai, C. L. Kraft, J. F. Abril, V. Anthouard, P. Arensburger, P. W. Atkinson, H. Baden, Berardinis de, V. D. Baldwin, V. Benes, J. Biedler, C. Blass, R. Bolanos, D. Boscus, M. Barnstead, S. Cai, A. Center, K. Chaturverdi, G. K. Christophides, M. A. Chrystal, M. Clamp, A. Cravchik, V. Curwen, A. Dana, A. Delcher, I. Dew, C. A. Evans, M. Flanigan, A. Grundschober-Freimoser, L. Friedli, Z. Gu, P. Guan, R. Guigo, M. E. Hillenmeyer, S. L. Hladun, J. R. Hogan, Y. S. Hong, J. Hoover, O. Jaillon, Z. Ke, C. Kodira, E. Kokoza, A. Koutsos, I. Letunic, A. Levitsky, Y. Liang, J. J. Lin, N. F. Lobo, J. R. Lopez, J. A. Malek, T. C. McIntosh, S. Meister, J. Miller, C. Mobarry, E. Mongin, S. D. Murphy, D. A. O'Brochta, C. Pfannkoch, R. Qi, M. A. Regier, K. Remington, H. Shao, M. V. Sharakhova, C. D. Sitter, J. Shetty, T. J. Smith, R. Strong, J. Sun, D. Thomasova, L. Q. Ton, P. Topalis, Z. Tu, M. F. Unger, B. Walenz, A. Wang, J. Wang, M. Wang, X. Wang, K. J. Woodford, J. R. Wortman, M. Wu, A. Yao, E. M. Zdobnov, H. Zhang, Q. Zhao, S. Zhao, S. C. Zhu, I. Zhimulev, M. Coluzzi, Torre A. della, C. W. Roth, C. Louis, F. Kalush, R. J. Mural, E. W. Myers, M. D. Adams, H. O. Smith, S. Broder, M. J. Gardner, C. M. Fraser, E. Birney, P. Bork, P. T. Brey, J. C. Venter, J. Weissenbach, F. C. Kafatos, F. H. Collins, and S. L. Hoffman.** The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298 (5591):129-149, 2002.

**Hua-Van A., A. Le Rouzic, C. Maisonhaute, and P. Capy.** Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogenet Genome Res* 110 (1-4):426-440, 2005.

**Hull R.** Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. International Committee on Taxonomy of Viruses. *Arch Virol* 146 (11):2255-2261, 2005.

**Jiang N., Z. Bao, X. Zhang, H. Hirochika, S. R. Eddy, S. R. McCouch, and S. R. Wessler.** An active DNA transposon family in rice. *Nature* 421 (6919):163-167, 2003.

**Jordan E., H. Saedler, and P. Starlinger.** Strong-polar mutations in the transferase gene of the galactose operon in *E.coli*. *Mol Gen Genet* 100 (3):296-306, 1967.

**Jordan I. K., I. B. Rogozin, G. V. Glazko, and E. V. Koonin.** Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet* 19 (2):68-72, 2003.

**Jukes T. H. and C. R. Cantor.** Evolution of protein molecules. In: *Mammalian protein metabolism*, edited by H. N. Munro. Academic Press, p. 21-132. New York, 1969.

**Kajikawa M. and N. Okada.** LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 111 (3):433-444, 2002.

**Kaminker J. S., C. M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, D. A. Wheeler, S. E. Lewis, G. M. Rubin, M. Ashburner, and S. E. Celniker.** The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol* 3 (12): research0084.1-0084.20, 2005.

**Kanno T., B. Huettel, M. F. Mette, W. Aufsatz, E. Jaligot, L. Daxinger, D. P. Kreil, M. Matzke, and A. J. Matzke.** Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* 37 (7):761-765, 2005.

**Kapitonov V. V. and J. Jurka.** Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica* 107 (1-3):27-37, 1999.

**Kapitonov V. V. and J. Jurka.** Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A* 98 (15):8714-8719, 2001.

**Kashkush K., M. Feldman, and A. A. Levy.** Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat Genet* 33 (1):102-106, 2003.

**Katayama S., Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C. C. Yap, M. Suzuki, J. Kawai, H. Suzuki, P. Carninci, Y. Hayashizaki, C. Wells, M. Frith, T. Ravasi, K. C. Pang, J. Hallinan, J. Mattick, D. A. Hume, L. Lipovich, S. Batalov, P. G. Engstrom, Y. Mizuno, M. A. Faghihi, A. Sandelin, A. M. Chalk, S. Mottagui-Tabar, Z. Liang, B. Lenhard, and C. Wahlestedt.** Antisense transcription in the mammalian transcriptome. *Science* 309 (5740):1564-1566, 2005.

**Kidwell M. G.** Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115 (1):49-63, 2002.

**Kikuchi K., K. Terauchi, M. Wada, and H. Y. Hirano.** The plant MITE mPing is mobilized in anther culture. *Nature* 421 (6919):167-170, 2003.

**Kim J. M., S. Vanguri, J. D. Boeke, A. Gabriel, and D. F. Voytas.** Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8 (5):464-478, 1998.

**Klessig D. F. and J. Malamy.** The salicylic acid signal in plants. *Plant Mol Biol* 26 (5):1439-1458, 1994.

**Kuhlmann M., B. E. Borisova, M. Kaller, P. Larsson, D. Stach, J. Na, L. Eichinger, F. Lyko, V. Ambros, F. Soderbom, C. Hammann, and W. Nellen.** Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res* 33 (19):6405-6417, 2005.

**Kumar A. and J. L. Bennetzen.** Plant retrotransposons. *Annu Rev Genet* 33:479-532, 1999.

Kunst F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, and A. Danchin. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390 (6657):249-256, 1997.

Kuo M. H., and C. D. Allis. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* 19 (3):425-433, 1999.

Lander E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, and Y. J. Chen. Initial sequencing and analysis of the human genome. *Nature* 409 (6822):860-921, 2001.

Le Q. H., S. Wright, Z. Yu, and T. Bureau. Transposon diversity in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 97 (13):7376-7381, 2000.

Leeton P. R. and D. R. Smyth. An abundant LINE-like element amplified in the genome of *Lilium speciosum*. *Mol Gen Genet* 237 (1-2):97-104, 1993.

Lippman Z., B. May, C. Yordan, T. Singer, and R. Martienssen. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol* 1 (3):420-428, 2003.

- Lippman Z. and R. Martienssen.** The role of RNA interference in heterochromatic silencing. *Nature* 431 (7006):364-370, 2004.
- Llave C., K. D. Kasschau, M. A. Rector, and J. C. Carrington.** Endogenous and silencing-associated small RNAs in plants. *Plant Cell* 14 (7):1605-1619, 2002.
- Logemann J., J. Schell, and L. Willmitzer.** Improved method for the isolation of RNA from plant tissues. *Anal Biochem* 163 (1):16-20, 1987.
- Loke J. C., E. A. Stahlberg, D. G. Strenski, B. J. Haas, P. C. Wood, and Q. Q. Li.** Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures. *Plant Physiol* 138 (3):1457-1468, 2005.
- Lonnig W. E. and H. Saedler.** Chromosome rearrangements and transposable elements. *Annu Rev Genet* 36:389-410, 2005.
- Magnotta S. M. and J. P. Gogarten.** Multi site polyadenylation and transcriptional response to stress of a vacuolar type H<sup>+</sup>-ATPase subunit A gene in *Arabidopsis thaliana*. *BMC Plant Biol* 2:3, 2002.
- Marracci S., R. Batistoni, G. Pesole, L. Citti, and I. Nardi.** Gypsy/Ty3-like elements in the genome of the terrestrial Salamander hydromantes (Amphibia, Urodela). *J Mol Evol* 43 (6):584-593, 1996.
- Mathieu O., A. V. Probst, and J. Paszkowski.** Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. *EMBO J* 24 (15):2783-2791, 2005.
- McClintock B.** Mutable loci in maize. In: *Canergie Institution of Washington Year Book*, p. 155-169, 1948.
- McClintock B.** The significance of responses of the genome to challenge. *Science* 226 (4676):792-801, 1984.
- McClintock B.** *The Discovery and Characterization of Transposable Elements: The Collected Papers of Barbara McClintock*. Garland Publishing Co., New York, 1987.
- Medstrand P., L. N. van de Lagemaat, C. A. Dunn, J. R. Landry, D. Svenback, and D. L. Mager.** Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res* 110 (1-4):342-352, 2005.
- Medstrand P., L. N. van de Lagemaat, and D. L. Mager.** Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* 12 (10):1483-1495, 2002.
- Melayah D., E. Bonnivard, B. Chalhoub, C. Audeon, and M. A. Grandbastien.** The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. *Plant J* 28 (2):159-168, 2001.
- Mette M. F., W. Aufsatz, J. van der Winden, M. A. Matzke, and A. J. Matzke.** Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* 19 (19):5194-5201, 2000.
- Meyers B. C., S. V. Tingey, and M. Morgante.** Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 11 (10):1660-1676, 2001.
- Mhiri C., J. B. Morel, S. Vernhettes, J. M. Casacuberta, H. Lucas, and M. A. Grandbastien.** The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Mol Biol* 33 (2):257-266, 1997.

- Mi S., X. Lee, X. Li, G. M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X. Y. Tang, P. Edouard, S. Howes, J. C. Jr Keith, and J. M. McCoy.** Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403 (6771):785-789, 2000.
- Miller W. J., J. F. McDonald, and W. Pinsker.** Molecular domestication of mobile elements. *Genetica* 100 (1-3):261-270, 1997.
- Miyao A., K. Tanaka, K. Murata, H. Sawaki, S. Takeda, K. Abe, Y. Shinozuka, K. Onosato, and H. Hirochika.** Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* 15 (8):1771-1780, 2003.
- Mlotshwa S., O. Voinnet, M. F. Mette, M. Matzke, H. Vaucheret, S. W. Ding, G. Pruss, and V. B. Vance.** RNA silencing and the mobile silencing signal. *Plant Cell* 14 Suppl:S289-S301, 2005.
- Moran J. V., R. J. DeBerardinis, and H. H. Jr Kazazian.** Exon shuffling by L1 retrotransposition. *Science* 283 (5407):1530-1534, 1999.
- Moreau-Mhiri C., J. B. Morel, C. Audeon, M. Ferault, M. A. Grandbastien, and H. Lucas.** Regulation of the tobacco Tnt1 retrotransposon in heterologous species following pathogen-related stress. *Plant J* 9:409-419, 1996.
- Muotri A. R., V. T. Chu, M. C. Marchetto, W. Deng, J. V. Moran, and F. H. Gage.** Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435 (7044):903-910, 2005.
- Murashige T. and F. Skoog.** A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol.Plant* 15:473-497, 1962.
- Mutskov V. and G. Felsenfeld.** Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J* 23 (1):138-149, 2004.
- Nabirochkin S., M. Ossokina, and T. Heidmann.** A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. *J Biol Chem* 273 (4):2473-2479, 1998.
- Nakamura T. M. and T. R. Cech.** Reversing time: origin of telomerase. *Cell* 92 (5):587-590, 1998.
- Nakazaki T., Y. Okumoto, A. Horibata, S. Yamahira, M. Teraishi, H. Nishida, H. Inoue, and T. Tanisaka.** Mobilization of a transposon in the rice genome. *Nature* 421 (6919):170-172, 2003.
- Nei M.** *Molecular Evolutionary Genetics*. Columbia University Press, New York, 1987.
- Nekrutenko A. and W. H. Li.** Transposable elements are found in a large number of human protein-coding genes. *Trends Genet* 17 (11):619-621, 2001.
- Nolan T., L. Braccini, G. Azzalin, A. De Toni, G. Macino, and C. Cogoni.** The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in *Neurospora crassa*. *Nucleic Acids Res* 33 (5):1564-1573, 2005.
- Noma K., T. Sugiyama, H. Cam, A. Verdel, M. Zofall, S. Jia, D. Moazed, and S. I. Grewal.** RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet* 36 (11):1174-1180, 2004.



- Okamoto H. and H. Hirochika.** Silencing of transposable elements in plants. *Trends Plant Sci* 6 (11):527-534, 2001.
- Onodera Y., J. R. Haag, T. Ream, P. C. Nunes, O. Pontes, and C. S. Pikaard.** Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120 (5):613-622, 2005.
- Orgel L. E. and F. H. Crick.** Selfish DNA: the ultimate parasite. *Nature* 284 (5757):604-607, 1980.
- Palauqui J. C., T. Elmayan, J. M. Pollien, and H. Vaucheret.** Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J* 16 (15):4738-4745, 1997.
- Pardue M. L. and P. G. DeBaryshe.** Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu Rev Genet* 37:485-511, 2005.
- Park Y. D., I. Papp, E. A. Moscone, V. A. Iglesias, H. Vaucheret, A. J. Matzke, and M. A. Matzke.** Gene silencing mediated by promoter homology occurs at the level of transcription and results in meiotically heritable alterations in methylation and gene activity. *Plant J* 9 (2):183-194, 1996.
- Peaston A. E., A. V. Evsikov, J. H. Graber, W. N. de Vries, A. E. Holbrook, D. Solter, and B. B. Knowles.** Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev* 7 (4):597-606, 2004.
- Perepelitsa-Belancio V. and P. Deininger.** RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet* 35 (4):363-366, 2003.
- Perron H., F. Lazarini, K. Ruprecht, C. Pechoux-Longin, D. Seilhean, V. Sazdovitch, A. Creange, N. Battail-Poirot, G. Sibai, L. Santoro, M. Jolivet, J. L. Darlix, P. Rieckmann, T. Arzberger, J. J. Hauw, and H. Lassmann.** Human endogenous retrovirus (HERV)-W ENV and GAG proteins: physiological expression in human brain and pathophysiological modulation in multiple sclerosis lesions. *J Neurovirol* 11 (1):23-33, 2005.
- Petersen G. and O. Seberg.** Phylogenetic evidence for excision of Stowaway miniature inverted-repeat transposable elements in triticeae (Poaceae). *Mol Biol Evol* 17 (11):1589-1596, 2000.
- Petrov D. A.** Evolution of genome size: new approaches to an old problem. *Trends Genet* 17 (1):23-28, 2001.
- Phillips C., S. Jung, and S. I. Gunderson.** Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA. *EMBO J* 20 (22):6443-6452, 2001.
- Plasterk R. H.** The origin of footprints of the Tc1 transposon of *Caenorhabditis elegans*. *EMBO J* 10 (7):1919-1925, 1991.
- Pouteau S., E. Huttner, M. A. Grandbastien, and M. Caboche.** Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. *EMBO J* 10 (7):1911-1918, 1991.
- Pouteau S., M. A. Grandbastien, and M. Boccardo.** Microbial elicitors of plant defence responses activate transcription of a retrotransposon. *Plant J* 5:535-542, 1994.
- Proudfoot N.** New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* 16 (3):272-278, 2004.

- Puig M., M. Cáceres, and A. Ruiz.** Silencing of a gene adjacent to the breakpoint of a widespread *Drosophila* inversion by a transposon-induced antisense RNA. *Proc Natl Acad Sci U S A* 101 (24):9013-9018, 2004.
- Ricci P., P. Bonnet, J. C. Huet, M. Sallantin, F. Beauvais-Cante, M. Bruneteau, V. Billard, G. Michel, and J. C. Pernollet.** Structure and activity of proteins from pathogenic fungi *Phytophthora* eliciting necrosis and acquired resistance in tobacco. *Eur J Biochem* 183 (3):555-563, 1989.
- Robertson H. M.** Evolution of DNA transposons in eucaryotes. In: *Mobile DNA II*. N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (eds.). ASM Press, p. 1093-1110. Washington DC, 2002.
- Rothnie H. M.** Plant mRNA 3'-end formation. *Plant Mol Biol* 32 (1-2):43-61, 1996.
- Rozas J. and R. Rozas.** DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15 (2):174-175, 1999.
- Rusterucci C., J. L. Montillet, J. P. Agnel, C. Battesti, B. Alonso, A. Knoll, J. J. Bessoule, P. Etienne, L. Suty, J. P. Blein, and C. Triantaphylides.** Involvement of lipoxygenase-dependent production of fatty acid hydroperoxides in the development of the hypersensitive cell death induced by cryptogein on tobacco leaves. *J Biol Chem* 274 (51):36446-36455, 1999.
- Saito M., J. Yonemaru, G. Ishikawa, and T. Nakamura.** A candidate autonomous version of the wheat MITE Hikkoshi is present in the rice genome. *Mol Genet Genomics* 273 (5):404-414, 2005.
- Saitou N. and M. Nei.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4 (4):406-425, 1987.
- Sambrook J. and D. W. Russell.** *Molecular Cloning: A Laboratory Manual*. 3rd. ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York, 2001.
- SanMiguel P., A. Tikhonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P. S. Springer, K. J. Edwards, M. Lee, Z. Avramova, and J. L. Bennetzen.** Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274 (5288):765-768, 1996.
- Seidman, C. G., K. Struhl, J. Sheen, and T. Jessen.** Introduction of plasmid DNA into cells, p. 1.8.-1.8.10. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), *Current protocols in molecular biology*, vol. 1. Wiley Interscience, New York, N.Y, 1997.
- Scott L., D. LaFoe, and C. F. Weil.** Adjacent sequences influence DNA repair accompanying transposon excision in maize. *Genetics* 142 (1):237-246, 1996.
- Shapiro J. A.** A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* 345 (1):91-100, 2005.
- Shapiro J. A.** Retrotransposons and regulatory suites. *Bioessays* 27 (2):122-125, 2005.
- Sijen T. and R. H. Plasterk.** Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* 426 (6964):310-314, 2003.
- Slotkin R. K., M. Freeling, and D. Lisch.** Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat Genet* 37 (6):641-644, 2005.

**Southern E. M.** Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98 (3):503-517, 1975.

**Sunkar R., T. Girke, and J. K. Zhu.** Identification and characterization of endogenous small interfering RNAs from rice. *Nucleic Acids Res* 33 (14):4443-4454, 2005.

**Takeda S., K. Sugimoto, H. Otsuki, and H. Hirochika.** A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J* 18 (4):383-393, 1999.

**Takeda S., K. Sugimoto, T. Kakutani, and H. Hirochika.** Linear DNA intermediates of the Tto1 retrotransposon in Gag particles accumulated in stressed tobacco and *Arabidopsis thaliana*. *Plant* 28 (3):307-317, 2001.

**Taliansky M., S. H. Kim, M. A. Mayo, N. O. Kalinina, G. Fraser, K. D. McGeachy, and H. Barker.** Escape of a plant virus from amplicon-mediated RNA silencing is associated with biotic or abiotic stress. *Plant J* 39 (2):194-205, 2004.

**Thierry D. and H. Vaucheret.** Sequence homology requirements for transcriptional silencing of 35S transgenes and post-transcriptional silencing of nitrite reductase (trans)genes by the tobacco 271 locus. *Plant Mol Biol* 32 (6):1075-1083, 1996.

**Thompson J. D., D. G. Higgins, and T. J. Gibson.** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22 (22):4673-4680, 1994.

**Tikhonov A. P., J. L. Bennetzen, and Z. V. Avramova.** Structural domains and matrix attachment regions along colinear chromosomal segments of maize and sorghum. *Plant Cell* 12 (2):249-264, 2000.

**Timmons L.** The long and short of siRNAs. *Mol Cell* 10 (3):435-437, 2002.

**Tu Z.** Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 98 (4):1699-1704, 2001.

**Van Larebeke N., C. H. Genetello, J. P. Hernasteens, A. De Picker, I. Zaenen, E. Messens, M. Van Montagu, and J. Schell.** Transfer of Ti plasmids between *Agrobacterium* strains by mobilization with the conjugative plasmid RP4. *Mol.gen Genet.* 152:1119-1124, 1977.

**Vance V. and H. Vaucheret.** RNA silencing in plants--defense and counterdefense. *Science* 292 (5525):2277-2280, 2001.

**Vastenhouw N. L. and R. H. Plasterk.** RNAi protects the *Caenorhabditis elegans* germline against transposition. *Trends Genet* 20 (7):314-319, 2004.

**Vaucheret H. and M. Fagard.** Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends Genet* 17 (1):29-35, 2001.

**Verdel A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi, S. I. Grewal, and D. Moazed.** RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303 (5658):672-676, 2004.

**Vernhettes S., M. A. Grandbastien, and J. M. Casacuberta.** In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon Tnt1. *Plant Mol Biol* 35 (5):673-679, 1997.

**Vernhettes S., M. A. Grandbastien, and J. M. Casacuberta.** The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Mol Biol Evol* 15 (7):827-836, 1998.

**Vigdal T. J., C. D. Kaufman, Z. Izsvak, D. F. Voytas, and Z. Ivics.** Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. *J Mol Biol* 323 (3):441-452, 2002.

**Voinnet O.** RNA silencing: small RNAs as ubiquitous regulators of gene expression. *Curr Opin Plant Biol* 5 (5):444-451, 2002.

**Volpe T., V. Schramke, G. L. Hamilton, S. A. White, G. Teng, R. A. Martienssen, and R. C. Allshire.** RNA interference is required for normal centromere function in fission yeast. *Chromosome Res* 11 (2):137-146, 2005.

**Volpe T. A., C. Kidner, I. M. Hall, G. Teng, S. I. Grewal, and R. A. Martienssen.** Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297 (5588):1833-1837, 2002.

**Wahle E. and U. Rugegger.** 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* 23 (3):277-295, 1999.

**Waterston R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyas, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. LeVine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapozhnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, and E. S. Lander.** Initial sequencing and comparative analysis of the mouse genome. *Nature* 420 (6915):520-562, 2002.

**Wesley S. V., C. A. Helliwell, N. A. Smith, M. B. Wang, D. T. Rouse, Q. Liu, P. S. Gooding, S. P. Singh, D. Abbott, P. A. Stoutjesdijk, S. P. Robinson, A. P. Gleave, A. G. Green, and P. M. Waterhouse.** Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J* 27 (6):581-590, 2001.

**Wessler S. R., T. E. Bureau, and S. E. White.** LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr Opin Genet Dev* 5 (6):814-821, 1995.

**Wessler S. R.** Turned on by stress. Plant retrotransposons. *Curr Biol* 6 (8):959-961, 1996.

**Wessler S. R.** McClintock at 100 - Reason to Celebrate. *Science* 294 (1):62-63, 2002.

**Wong L. H. and K. H. Choo.** Evolutionary dynamics of transposable elements at the centromere. *Trends Genet* 20 (12):611-616, 2004.

**Wood V., R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schafer, S. Muller-Auer, C. Gabel, M. Fuchs, A. Dusterhoff, C. Fritz, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dreano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sanchez, F. del Rey, J. Benito, A. Dominguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, and P. Nurse.** The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415 (6874):871-880, 2002.

**Wu X. and S. M. Burgess.** Integration target site selection for retroviruses and transposable elements. *Cell Mol Life Sci* 61 (19-20):2588-2596, 2004.

**Wyrick J. J., J. G. Aparicio, T. Chen, J. D. Barnett, E. G. Jennings, R. A. Young, S. P. Bell, and O. M. Aparicio.** Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294 (5550):2357-2360, 2001.

**Yang G., J. Dong, M. B. Chandrasekharan, and T. C. Hall.** Kiddo, a new transposable element family closely associated with rice genes. *Mol Genet Genomics* 266 (3):417-424, 2001.

**Yanisch-Perron C., J. Vieira, and J. Messing.** Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33 (1):103-119, 1985.

**Yieh L., G. Kassavetis, E. P. Geiduschek, and S. B. Sandmeyer.** The Bf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the gypsy-like element, Ty3. *J Biol Chem* 275 (38):29800-29807, 2000.

**Yieh L., H. Hatzis, G. Kassavetis, and S. B. Sandmeyer.** Mutational analysis of the transcription factor IIIB-DNA target of Ty3 retroelement integration. *J Biol Chem* 277 (29):25920-25928, 2002.

**Yonaha M. and N. J. Proudfoot.** Transcriptional termination and coupled polyadenylation in vitro. *EMBO J* 19 (14):3770-3777, 2000.

**Yoo B. C., F. Kragler, E. Varkonyi-Gasic, V. Haywood, S. Archer-Evans, Y. M. Lee, T. J. Lough, and W. J. Lucas.** A systemic small RNA signaling system in plants. *Plant Cell* 16 (8):1979-2000, 2004.

**Yoshioka Y., S. Matsumoto, S. Kojima, K. Ohshima, N. Okada, and Y. Machida.** Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc Natl Acad Sci U S A* 90 (14):6562-6566, 1993.

**Yu J., S. Hu, J. Wang, G. K. Wong, S. Li, B. Liu, Y. Deng, L. Dai, Y. Zhou, X. Zhang, M. Cao, J. Liu, J. Sun, J. Tang, Y. Chen, X. Huang, W. Lin, C. Ye, W. Tong, L. Cong, J. Geng, Y. Han, L. Li, W. Li, G. Hu, J. Li, Z. Liu, Q. Qi, T. Li, X. Wang, H. Lu, T. Wu, M. Zhu, P. Ni, H. Han, W. Dong, X. Ren, X. Feng, P. Cui, X. Li, H. Wang, X. Xu, W. Zhai, Z. Xu, J. Zhang, S. He, J. Xu, K. Zhang, X. Zheng, J. Dong, W. Zeng, L. Tao, J. Ye, J. Tan, X. Chen, J. He, D. Liu, W. Tian, C. Tian, H. Xia, Q. Bao, G. Li, H. Gao, T. Cao, W. Zhao, P. Li, W. Chen, Y. Zhang, J. Hu, S. Liu, J. Yang, G. Zhang, Y. Xiong, Z. Li, L. Mao, C. Zhou, Z. Zhu, R. Chen, B. Hao, W. Zheng, S. Chen, W. Guo, M. Tao, L. Zhu, L. Yuan, and H. Yang.** A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296 (5565):79-92, 2002.

**Zarudnaya M. I., I. M. Kolomiets, A. L. Potyahaylo, and D. M. Hovorun.** Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res* 31 (5):1375-1386, 2003.

**Zhang X., C. Feschotte, Q. Zhang, N. Jiang, W. B. Eggleston, and S. R. Wessler.** P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc Natl Acad Sci U S A* 98 (22):12572-12577, 2001.

**Zhong C. X., J. B. Marshall, C. Topp, R. Mroczek, A. Kato, K. Nagaki, J. A. Birchler, J. Jiang, and R. K. Dawe.** Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 14 (11):2825-2836, 2002.

**Zhu Y., J. Dai, P. G. Fuerst, and D. F. Voytas.** Controlling integration specificity of a yeast retrotransposon. *Proc Natl Acad Sci U S A* 100 (10):5891-5895, 2003.

**Zilberman D., X. Cao, and S. E. Jacobsen.** ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299 (5607):716-719, 2003.

**Zou S. and D. F. Voytas.** Silent chromatin determines target preference of the *Saccharomyces* retrotransposon Ty5. *Proc Natl Acad Sci U S A* 94 (14):7412-7416, 1997.

# APÉNDICE 1.

## APÉNDICE 1.

### Tabla completa de los MITEs *Emigrant* identificados en Arabidopsis.

Elemento y grupo	Número de acceso	Localización en el BAC	Cr.	Contexto génico			
				A 5' del MITE		A 3' del MITE	
				gen	dist. (pb)	gen	dist. (pb)
e1r A2	AC006273	27301-27809	I	T1461.0	16386	T1460.7, adenoprotein P1460 protein	(5') 3666
e2 E3	AC006278	51446-50962	I	F14963.97, Strong similarity to g8421.650 E2NSP precursor from <i>Daucus carota</i>	1729	F14963.95	(3') 228
e3 O	AC026975	82546-82791	I	T0022.97	(3') 641	T0022.95, similar to high affinity ribulose biphosphate ADP4Z g8421.650	(3') 372
e4 E3	AC025806	92385-92926	I	F1001426, similar to EST g84.382433	1889	F1001427, similar to membrana-associated wall-associated protein	(5') 902
e5r E3	AC069252	78871-79331	I	F262.19, hypothetical protein	1867	F262.20, similar to MADS-box protein NDR1 7 g8421.650	(3') 428
e6r C1	AC006661	32147-31490	I	F1268.9, similar to LUCIF1	(3') 75	F1268.8, very similar to retrotransposon Tair19 reverse transcriptase	1492

**Tabla A1.**

En esta tabla se enumeran todos los MITEs de la familia *Emigrant* identificados en el trabajo presentado en esta memoria. Se especifica el grupo al que pertenece cada elemento, el número de acceso del BAC donde se localiza y su posición dentro del mismo, el cromosoma en el que se encuentra y el contexto génico que rodea al transposón. Se indica la secuencia génica más próxima al MITE a 5' y a 3' y la distancia que la separa del elemento *Emigrant*. En aquellos casos en los que el transposón se localiza a menos de 500 pb de la secuencia codificante (donde consideramos que se localizan las regiones reguladoras no traducidas), se indica, entre paréntesis, si se encuentra a 3' o a 5' del gen. El CD que se adjunta con este manuscrito contiene la tabla completa.



## **APÉNDICE 2.**

## APÉNDICE 2.

### Tabla de la librería de los sRNAs clonados de *Nicotiana tabacum*.

sRNA	Secuencia (5'→3')	Tamaño (nt)	Homología
1	CTTCAGATCCAGCTCCACACGAC	24	gn15042110058pansaricus NL_001_00447111 gp2518797/Arabidopsis1517511.1NICO517511 gsp-pul gene, Ty3-gypsy retrotransposon-like
2	GGTGGGGCCCTCCGAGTCCAT	24	No hits found
3	AGGGGTGTTGATGGAGCCAG	21	No hits found
5	ATTTCCACCCATAAACCCAGTTGAG	24	similar to gypsy type retrotransposon protein orf3 CB:BN484461 GI:5902448 ( <i>Oryza sativa</i> )
6	AAATCCGACACXCGAATCGACGTCC	24	No hits found
7	AGCTCGACGACTTGAATAAGAG	23	No hits found
8A	GTCGTTGCTCCATATAAACCTCA	23	Tm1 <i>N. tabacum</i> and Tm57 RT gene de <i>Oryza sativa</i> / TM31 <i>Ipomoea batatas</i>

**Tabla A2.**

En esta tabla se detallan todas las secuencias de los sRNAs clonados a partir de un extracto de RNA total de hoja no estresada de *Nicotiana tabacum*. Se indica el número otorgado a cada sRNA, su secuencia, tamaño (en pares de bases) y las secuencias con las que presenta homología, obtenidas mediante una búsqueda en las bases de datos. El CD que se adjunta con este manuscrito contiene la tabla completa.

## **APÉNDICE 3.**

### APÉNDICE 3.

#### Tabla de las plantas transgénicas generadas.

PLANTAS TRANSGÉNICAS: T <sub>1</sub>						
DNA						
Planta	Genómico PCR	Genómico	Homocigota?	Semilla	T <sub>2</sub> resistente?	
ETAG 5						
50			68,18%			30/44
51						48/48
52			70,00%			35/50
53						48/49
54						51/51
55						28/28
56			73%			35/48
57						49/49
58			74%			37/50
59						48/48

**Tabla A3.**

En esta tabla se enumeran todas las plantas transgénicas generadas en el trabajo presentado en esta memoria. Se indica si se trata de las líneas parentales (T<sub>0</sub>) o de sus descendientes (T<sub>1</sub> y T<sub>2</sub>). Para cada construcción se generaron distintas plantas transgénicas, que fundaron distintas líneas. Se especifica el número de la planta (o línea), si se dispone en el laboratorio de DNA genómico, si se trata de una planta homocigota para el transgén, si se dispone de semilla almacenada y la proporción de descendientes que han heredado el transgén (y son resistentes al antibiótico de selección). El DNA genómico de PCR se obtuvo mediante una extracción rápida de DNA (ver material y métodos). Los cuadros sombreados indican bien la disposición de DNA y de semillas o bien que la planta es homocigota. El CD que se adjunta con este manuscrito contiene la tabla completa.

## **APÉNDICE 4.**

## APÉNDICE 4.

# ANÁLISIS ESTADÍSTICOS.

Se han realizado diferentes pruebas estadísticas sobre los resultados obtenidos de la distribución de los elementos *Emigrant* de MITEs dentro del genoma de *Arabidopsis thaliana*. Para cada uno de los grupos *Emigrant* se han llevado a cabo pruebas de Chi-cuadrado independientes.

El primer análisis (tabla A4.1) compara la distribución observada de los elementos *Emi* respecto a los genes con una localización al azar de los mismos en el genoma de *Arabidopsis*. Para realizar este análisis se han establecido, arbitrariamente, cuatro intervalos en los que pueden estar situados los MITEs: 0 (los MITEs están insertados dentro de la secuencia transcrita, tanto exones como intrones), <500 nt (posibles regiones reguladoras), 500-1000 nt (probablemente fuera de una región génica) y >1000 nt ("lejos" de genes).

Las frecuencias esperadas se han calculado en función de las características del genoma de *A.thaliana* (ver capítulos de introducción y discusión).

Los resultados de la prueba de Chi-cuadrado, con un 99% de confianza, muestran que *EmiA2* es la única subfamilia que presenta una distribución respecto a los genes significativamente no aleatoria. Un valor de  $\chi^2$  de 19.76 (cuando el valor de  $\chi^2$  para 3 grados de libertad al 99% de confianza es 11.34, tabla A4.3) indica una clara tendencia a su inserción lejos de los genes. Los valores de Chi-cuadrado (al 99% de confianza) para el resto de las subfamilias entran dentro de lo esperado en una distribución al azar.

Se realizó un segundo análisis (tabla A4.2) para comparar la distribución de los elementos más antiguos (*Emi0*) respecto a la distribución mostrada por los más modernos (*EmiA2*). Esta prueba permite detectar cualquier diferencia existente en la distribución de los MITEs *Emi* a lo largo del tiempo, asumiendo el patrón que muestran los *EmiA2* como el primitivo, el correspondiente a la especificidad de inserción de estos elementos. Por lo tanto, las frecuencias esperadas se han calculado a partir de la localización de las copias *EmiA2* respecto a los genes.

Un valor de Chi-cuadrado igual a 30.25 muestra una distribución de los elementos arcaicos significativamente diferente a la de los más recientes.

**Tabla A4.1. Distribución de *Emi* respecto a los genes.**

Grupo	4 intervalos <sup>1</sup>	2 intervalos <sup>2</sup>
	$\chi^2$	$\chi^2$
<i>Emi0</i>	8,85	1,07
<i>EmiA1</i>	7,99	4,44
<i>EmiA2</i>	19,76	12,7
<i>EmiA3</i>	3,05	0,89
<i>EmiB1</i>	2,68	1,78
<i>EmiB2</i>	9	0,25
<i>EmiB3</i>	10,45	0,07
<i>EmiC1</i>	6,42	1,04
<i>EmiC2</i>	4,61	2,67
<i>EmiC3</i>	3,5	2,67

<sup>1</sup> 0, <500, 500-1000, >1000 nt; son 3 grados de libertad

<sup>2</sup> <500, >500 nt; es 1 grado de libertad

**Tabla A4.2. Distribución de *Emi0* respecto a *EmiA2*.**

Grupo	4 intervalos <sup>1</sup>	2 intervalos <sup>2</sup>
	$\chi^2$	$\chi^2$
0	30,25	27,28

<sup>1</sup> 0, <500, 500-1000, >1000 nt; son 3 grados de libertad

<sup>2</sup> <500, >500 nt; es 1 grado de libertad

**Tabla A4.3. Valores críticos de Chi-cuadrado.**

g.l.	Probabilidad	
	0,05	0,01
1	3,84	6,63
2	5,99	9,21
3	7,81	11,34

g.l. significa grados de libertad

*“Pero, de pronto, hacia las dos o las tres de la mañana, un viento suave se abrió paso por el río y por la ventana y el tejado del molino se llenaron de repente de una lluvia compacta y amarilla. Eran las hojas muertas de los chopos, que caían, la lenta y mansa lluvia del otoño que de nuevo regresaba a las montañas para cubrir los campos de oro viejo y los caminos y los pueblos de una dulce y brutal melancolía”.*

**Julio Llamazares.** *La lluvia amarilla.*



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

Néstor Santiago,\* Cristina Herráiz,† J. Ramón Goñi,† Xavier Messeguer,† and Josep M. Casacuberta\*

\*Department of Genètica Molecular, IBMB-CSIC, Barcelona; and †Department of Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya

Miniature inverted-repeat transposable elements (MITEs) are structurally similar to defective class II elements, but their high copy number and the size and sequence conservation of most MITE families suggest that they can be amplified by a replicative mechanism. Here we present a genome-wide analysis of the *Emigrant* family of MITEs from *Arabidopsis thaliana*. In order to be able to detect divergent ancient copies, and low copy number subfamilies with a different internal sequence we have developed a computer program to look for *Emigrant* elements based solely on the terminal inverted-repeat sequence. We have detected 151 *Emigrant* elements of different subfamilies. Our results show that different bursts of amplification, probably of few active, or master, elements, have occurred at different times during *Arabidopsis* evolution. The analysis of the insertion sites of the *Emigrant* elements shows that recently inserted *Emigrant* elements tend to be located far from open reading frames, whereas more ancient *Emigrant* subfamilies are preferentially found associated to genes.

## Introduction

Transposable elements (TEs) are important components of eukaryote genomes, accounting for a high fraction of them. Most TEs can be grouped in two major classes according to their mode of transposition: class I elements transpose by a replicative mechanism invoking an RNA intermediate, whereas class II elements usually transpose by a nonreplicative “cut-and-paste” mechanism. Whereas most genomes contain elements of both classes, class I elements are in general much more abundant than class II elements, probably as a consequence of their replicative mode of transposition. This is the case for the human genome, where the most abundant TEs are the L1 long interspersed nuclear element (LINE) and the Alu family of short interspersed nuclear elements (SINEs) (International Human Genome Sequencing Consortium 2001), both belonging to the class I group of TEs. Plant genomes also contain retrotransposons in a high copy number, although in this case retroviral-like long terminal repeat–retrotransposons are the most abundant (Kumar and Bennetzen 1999). In addition, plant genomes contain miniature inverted-repeat transposable elements (MITEs) in a very high copy number (Wessler, Bureau, and White 1995; Le, Wright, and Bureau 2000; Turcotte, Srinivasan, and Bureau 2001).

MITEs are a particular class of TE first described in plants but later found to be present in other eukaryote genomes (Bureau and Wessler 1992; Oosumi, Garlick, and Belknap 1996; Tu 1997). They are structurally similar to defective class II elements, but their high copy number and the size and sequence conservation of most MITE families suggest that they can be amplified by a replicative mechanism. It has been recently proposed

that MITEs could be a particular type of defective class II element (Feschotte, Jiang, and Wessler 2002), some of them related to the pogo subclass of Mariner transposons or to bacterial insertion sequence elements (Feschotte and Mouchès 2000a; Le, Wright, and Bureau 2000; Zhang et al. 2001). Nevertheless, whereas it has been proposed that some MITE families could still be active in plants (Casacuberta et al. 1998; Zhang, Arbuckle, and Wessler 2000; Zhang et al. 2001), the characterization of a mobile MITE copy allowing the analysis of its transposition mechanisms is still lacking. In this context, the analysis of the evolution of MITE families of elements within their host genomes is probably the best approach to analyze the lifestyle of these elements and the impact of their mobility on host genomes.

Here we present a genome-wide analysis of the *Emigrant* family of MITEs in *Arabidopsis thaliana*. In order to be able to detect elements with a divergent internal sequence, representing either ancient *Emigrant* elements, or previously undescribed low copy number *Emigrant* subfamilies, we have developed a computer program to detect putative MITEs in a genomic sequence based solely on their terminal inverted-repeat (TIR) sequences. This approach has allowed us to perform, for the first time, an evolutionary analysis of a family of MITEs within a particular genome. Our results show that different *Emigrant* subfamilies of elements have probably been generated by the amplification of a small number of founder elements. Our results also show that, although *Emigrant* elements target very rich AT regions for insertion, elements closely linked to genes are more frequently maintained during evolution.

## Materials and Methods

### TRANSPO and SPAT Programs

TRANSPO implements the fast bit-vector algorithm (Myers 1998) that finds all locations at which a query (the *Emigrant* TIR sequence in this case) approximately matches a sequence (the sequence of five *Arabidopsis* chromosomes). The expected time is linear on the length of the sequence. Although this algorithm is

Key words: *Arabidopsis*, evolution, MITE, *Emigrant*, master element.

Address for correspondence and reprints: Josep M. Casacuberta, Department of Genètica Molecular, IBMB-CSIC, Jordi Girona 18, 08034 Barcelona, Spain. E-mail: jesgmp@cid.csic.es.

*Mol. Biol. Evol.* 19(12):2285–2293. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

based on dynamic programming with quadratic cost on the lengths of the sequence and the query, the linearity can be accomplished if the length of the query is shorter than the length of the computer word.

The clustering program SPAT groups the sequences into a hierarchical classification, i.e., a nested sequence of partition (Gordon 1999). Given a similarity measure between each pair of sequences, the complete weighted graph is designed where the nodes are the sequences and the weight of each edge is the similarity measure between the pair of sequences connected by this edge. Then the maximum spanning binary tree (MST) is found. This tree “spans” the graph connecting all the nodes in such away that the sum of the weights of the edges is maximized. The algorithm proceeds by removing the edge with minimum weight and dividing the tree into two disjoint subtrees (Zahn 1971; Delattre and Hansen 1980). A cluster is determined by the value of the removed edge that creates it and the value of the edge that divides it. A cluster of sequences is cohesive when a consecutive extraction of a significant number of edges does not change the composition of the cluster.

TRANSP0 and SPAT are available at [www.lsi.upc.es/~algen](http://www.lsi.upc.es/~algen).

#### Emigrant Element Mining

The TRANSP0 program was used to look at the entire available *Arabidopsis* genome sequence ([www.arabidopsis.org](http://www.arabidopsis.org)) for inverted repeated sequences 75% identical to the first 20 nt of the previously defined *Emigrant* TIR (CAGTAAACCTCTATAAATT) located within a range of 200–700 nt. Overlapping elements generated from subterminal inverted-repeat sequences were eliminated. A pairwise similarity matrix was calculated and sequences were grouped using the SPAT program, and a graphical distribution of the different elements in *Arabidopsis* chromosomes was obtained using the program CLUPH.

To obtain information about the open reading frame (ORF) located close to the *Emigrant* elements, the 30 nt flanking each *Emigrant* element upstream and downstream were used as probes in sequence similarity searches (BLAST 2.0; Altschul et al. 1990; <http://www.ncbi.nlm.nih.gov/blast/>). A table containing the BAC accession number and the nucleotide position of each *Emigrant* element, as well as the name and the distance of the elements to the closest upstream and downstream ORF can be obtained as additional information. Sequence similarity searches (BLAST) with the sequences flanking *Emigrant* elements were also used to look for related empty sites (RESites).

#### Phylogenetic Analysis

Sequences belonging to a particular group obtained with SPAT were aligned using the multiple-alignment program CLUSTAL W using the default parameters (version 1.5; Thompson, Higgins, and Gibson 1994), with some minor refinements. DNADIST in Felsenstein’s package PHYLIP (Felsenstein 1989) was used to generate a distance matrix based on the Jukes-Cantor al-

gorithm (Jukes and Cantor 1969). This was used to generate neighbor-joining trees (Saitou and Nei 1987). Bootstrap analyses were performed using the programs Seqboot and Consense from PHYLIP (Felsenstein 1989). Sequence variability, as measured by Nei’s measure of nucleotide diversity,  $\pi$  (Nei 1987), and its standard deviation were calculated using the program DnaSP (Rozas and Rozas 1999).

#### Results

Different Groups of the *Emigrant* Element can be found in the Genome of *A. thaliana*

We have designed a new computer program named TRANSP0 that looks at a given sequence for the presence of a particular inverted-repeat motif within a given range of distances. The independence of the search from a conserved internal sequence allows the localization of previously undetected low copy number subfamilies of a particular MITE that share the TIR sequences but differ in their internal sequence, as well as ancient MITE copies that have lost most of the sequence homology by insertion, deletions, or point mutations, within the sequence between the TIRs.

We have searched the complete available *Arabidopsis* sequence, which covers 115.4 Mbp of the 125-Mbp genome and does not include telomeres, centromeres or the rDNA repeated regions, for *Emigrant* elements by looking for *Emigrant* TIRs (tolerating up to 25% divergence) separated by more than 200 nt and less than 700 nt. We have localized 151 sequences that could represent *Emigrant* elements. All these sequences present *Emigrant*-like TIRs, are very AT-rich, and do not have coding capacity, and most of them are flanked by the dinucleotide TA. Although all these characteristics are reminiscent of MITE-related sequences, most of these sequences are not annotated as *Emigrant* elements, MITEs, or possible TEs in the databases.

The high variability of the internal sequence does not allow the correct alignment of the 151 sequences and their analysis by phylogenetic methods. We have developed the program SPAT that proceeds by elimination of the most divergent sequence of a given group in order to tentatively group the sequences and be able to apply conventional phylogenetic methods. SPAT gave a tentative classification of most of the 151 sequences into three main groups, that we have named *EmiA* (41 sequences), *EmiB* (26 sequences), and *EmiC* (37 sequences), based on pairwise identity comparisons. Forty-seven sequences were too divergent to be included in any of the defined groups and have been named as *Emi0* elements.

All the previously described *Emigrant* elements (Casacuberta et al. 1998) belong to the *EmiA* group. We have previously demonstrated that *EmiA* elements were mobile in the recent past because some of them were found to be polymorphic among *Arabidopsis* ecotypes (Casacuberta et al. 1998). In order to obtain data on the possible mobility of the other groups of elements, we searched the *Arabidopsis* genome for RESites, representing genome duplication events occurring prior to the

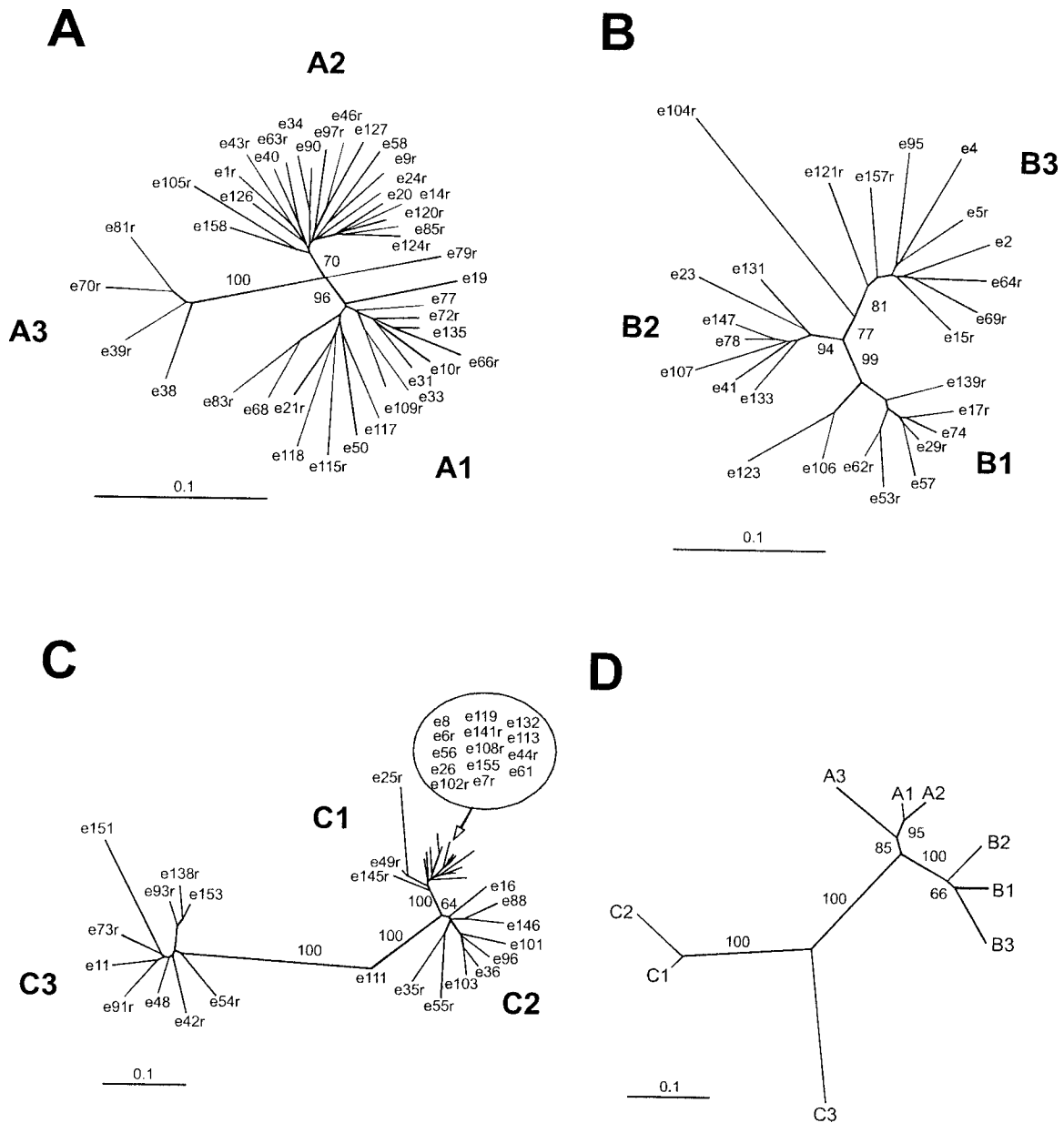


FIG. 1.—Phylogenetic analysis of the *Emigrant* elements. Neighbor-joining trees obtained comparing the *EmiA* (A), *EmiB* (B), and the *EmiC* (C) elements. The subfamilies defined within each *Emigrant* group are shown as A#, B#, or C#. (D) Neighbor-joining tree obtained comparing the consensus sequences of the different *Emigrant* subfamilies. Bootstrap values above 60% supporting major clusters are shown. Distances are proportional to evolutionary divergence expressed in substitutions per hundred sites.

transposition of these newly described elements. The presence of RESites within a genome has been successfully used as an indication of mobility when analyzing possible TEs within a single genome (Le, Wright, and Bureau 2000; Tu 2001). We found more than 20 well-conserved RESites corresponding to the different groups of *Emigrant* elements, although we found more RESites corresponding to the *EmiA*, *EmiB*, and *EmiC* classes than to the *Emi0* class (not shown). These data, and the presence in each case of a TA duplication accompanying the insertion of the element, strongly suggest that the different elements described here are indeed mobile elements related to the *Emigrant* family of MITEs.

#### Analysis of the Sequence and Size Variability of the Different *Emigrant* Subfamilies

The relatively high sequence identity within each of the *EmiA*, *EmiB*, and *EmiC* groups of elements has allowed us to perform conventional phylogenetic analysis. Sequences belonging to each group were aligned using CLUSTAL W, and the alignments were used to obtain neighbor-joining trees. Figure 1 presents the trees obtained. Different monophyletic groups supported by high bootstrap values can be defined within each tree. Within each *Emigrant* group most of the sequences can be subdivided into three different subfamilies (A1, A2,

**Table 1**  
**Sequence, Size variability and Position with Respect to the Closest Predicted Genes of the Different *Emigrant* Subfamilies**

	#	Nucleotide diversity ( $\pi$ )	Standard deviation of $\pi$	Size variability (%)	Position with respect to the closest predicted gene						
					0 (within)	<500	500–1000	>1000	rep.	n.d.	
<i>EmiA1</i> ..	16	0.108	0.0069	5.6	12.5% (2)	19% (3)	19% (3)	44% (7)	6% (1)		
<i>EmiA2</i> ..	20	0.095	0.0040	2.7		15% (3)	20% (4)	50% (10)	15% (3)		
<i>EmiA3</i> ..	4	0.081	0.0149	5.2		25% (1)	50% (2)		25% (1)		
<i>EmiB1</i> ...	9	0.102	0.0142	3.9	12.5% (1)	12.5% (1)	12.5% (1)	38% (3)	25% (2)	1	
<i>EmiB2</i> ...	7	0.077	0.0097	3.8		43% (3)	43% (3)		14% (1)		
<i>EmiB3</i> ...	9	0.115	0.0092	8.2		55.5% (5)	33% (3)	11% (1)			
<i>EmiC1</i> ...	17	0.068	0.0059	7.1	20% (3)	20% (3)	7% (1)	40% (6)	13% (2)	2	
<i>EmiC2</i> ...	9	0.112	0.0087	4.2	11% (1)	22% (2)	44% (4)	22% (2)			
<i>EmiC3</i> ...	10	0.131	0.0167	5.9	10% (1)	20% (2)	30% (3)	30% (3)	10% (1)		
<i>Emi0</i> ...	47	n.d.	n.d.	n.d.	14% (6)	28% (12)	21% (9)	18.5% (8)	18.5% (8)	4	
Total ...	151	n.d.	n.d.	n.d.	10% (15)	25.5% (37)	23% (33)	27.5% (40)	13% (19)	7	

NOTE.—Sequence variability is shown as the  $\pi$  measure (Nei 1987) of nucleotide diversity. Size variability has been calculated as the standard deviation from the mean size. Nucleotide and size variability was not determined (n.d.) for *Emi0* because these elements do not constitute a homogeneous group and cannot be aligned. For the distances to the predicted genes we show the percentage of sequences in each category, as well as the number of sequences (in parentheses). We did not determine (n.d.) the distance to the closest ORF for seven *Emigrant* elements because we have not obtained the annotation of the corresponding genomic region. The e79 element (belonging to the *EmiA* group and lying <500 nt of an ORF), the e104 element (belonging to the *EmiB* group and lying at <500 nt of an ORF), and the e111 element (belonging to the *EmiC* group and lying within an ORF) have not been included in any of the *Emigrant* subfamilies but have been included in the global analysis of elements.

and A3; B1, B2, and B3; C1, C2, and C3). By performing new alignments with the sequences belonging to each subfamily, we have deduced a consensus sequence for each of them and compared these consensus sequences in order to obtain information about the phylogenetic relationships among the different *Emigrant* subfamilies. A neighbor-joining tree, obtained comparing the consensus sequences of each subfamily, is also shown in figure 1. The three *EmiA* subfamilies, the three *EmiB* subfamilies, and the three *EmiC* subfamilies seem phylogenetically related because the three different groups cluster together with high bootstrap values.

The alignments of the sequences belonging to each subfamily were also used to calculate the nucleotide diversity,  $\pi$  (Nei 1987), and the size variability for each *Emigrant* subfamily. These results show that each *Emigrant* subfamily displays a different degree of variability (table 1). Whereas some subfamilies such as *EmiA2* are highly homogeneous both in size and sequence, other subfamilies like the *EmiB3* subfamily are highly variable. Each *Emigrant* group contains subfamilies of different variability. Within the *EmiA* group the A2 subfamily is more homogeneous than the A1 subfamily; within the *EmiB* group the most homogeneous is the B2 subfamily, and the most variable is B3, the B1 subfamily displaying an intermediate degree of variability. Within the *EmiC* group the C1 subfamily is the most homogeneous in sequence although relatively variable in size, and C3 seems to be the most variable group.

#### Analysis of *Emigrant* Insertion Sites

Figure 2 shows the distribution of *Emigrant* elements in *Arabidopsis* chromosomes. The five chromosomes of *Arabidopsis* contain *Emigrant* elements al-

though the density of insertions varies slightly among them. Chromosome 4 has the highest concentration of *Emigrant* elements (1.8 *Emigrant* (*Emi*)/Mbp), whereas chromosome 5 the lowest (1 *Emi*/Mbp), and chromosomes 1, 2, and 3 intermediate *Emigrant* concentrations. Although the concentration of the different *Emigrant* groups varies slightly between the five chromosomes, they all contain representatives of each *Emigrant* group. The concentration of *EmiA* elements in chromosome 4 is six times higher than in chromosome 5, and twice the concentration of *EmiC* elements is present in chromosomes 2 and 4 compared with chromosome 3, whereas there are more *Emi0* elements (3 $\times$ ) in chromosome 4 than in chromosome 1.

The analysis of 60 nt around the insertion site of the 151 *Emigrant* insertions shows that, in addition to insertion within TA sequences, *Emigrant* elements, like other MITEs and mariner-like elements (MLEs) (Le, Wright, and Bureau 2000), target regions of very high AT content. The average AT content of the *Arabidopsis* chromosomes ranges from 64.5% to 66.6%, rising to 67.6% in noncoding regions (The *Arabidopsis* Genome Initiative 2000), whereas sequences flanking *Emigrant* elements have 74.3% AT. Although the most frequent microsatellite in *Arabidopsis* consists of TA repetitions (Casacuberta, Puigdomènech, and Monfort 2000), *Emigrant* does not seem to target microsatellites for integration because whereas 25% of the *Emigrant* elements analyzed here lie in a TATA sequence, only 3% are found in a sequence containing a repetition of more than 4 TAs.

#### Position of *Emigrant* Elements Relative to ORFs

Although MITEs seem to target AT-rich regions, they have often been found close to transcribed sequen-

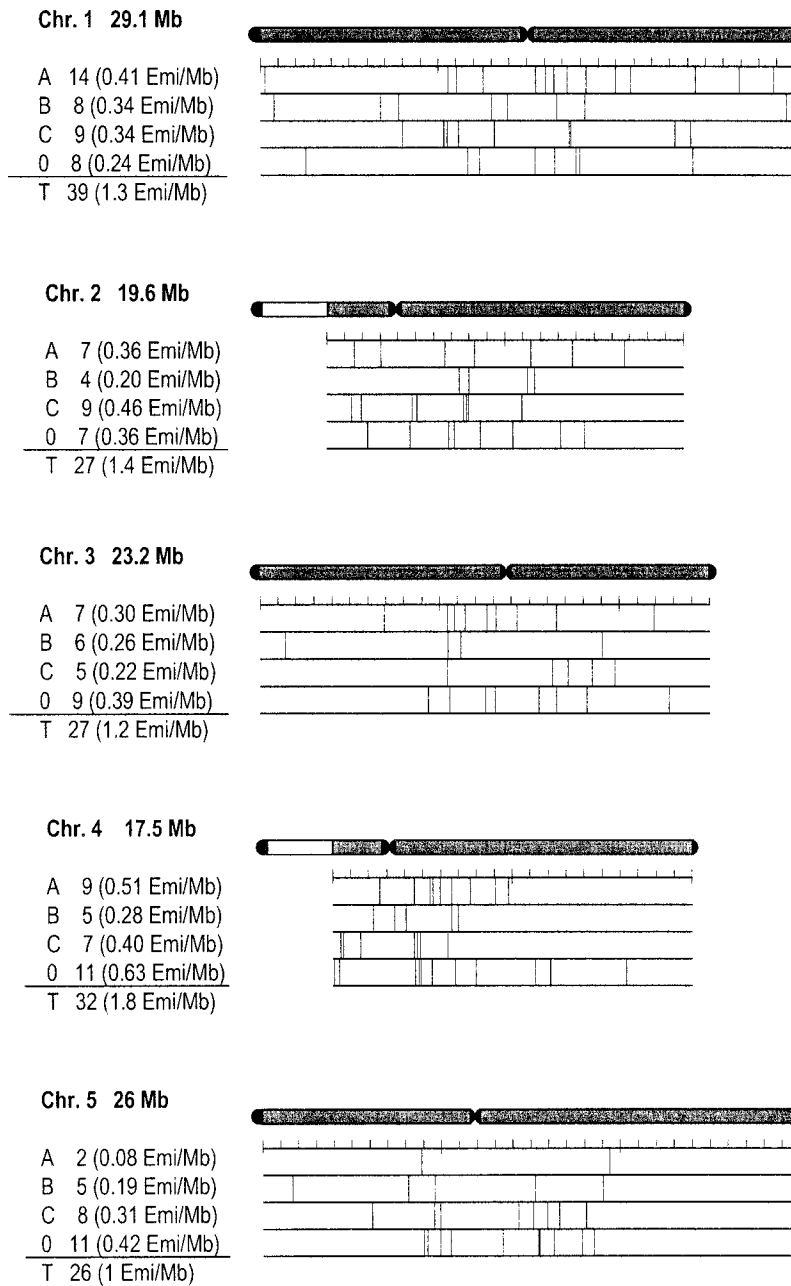


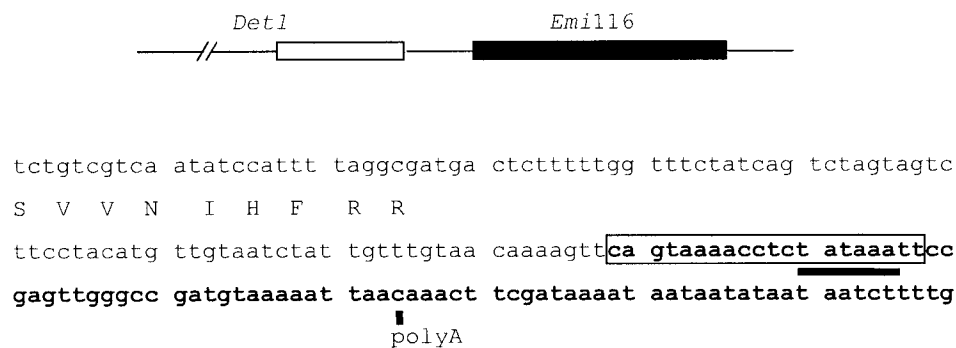
FIG. 2.—Distribution of *Emigrant* elements of the different groups in the five chromosomes of *Arabidopsis*. The number of *Emigrant* elements per million base pairs is indicated for each subfamily.

es (Wessler, Bureau, and White 1995; Yang et al. 2001). We analyzed the regions flanking the 151 *Emigrant* insertions and calculated the distance from the ATG or STOP codon of the closest predicted gene. Ten percent of the elements lie within a predicted gene (7% in introns and 3% in exons), 24% lie at less than 500 nt from an ORF, and 23% lie at more than 500 nt and less than 1,000 nt from an ORF. Twenty-nine percent are located at more than 1,000 nt from any ORF, and 13% are inserted within a repetitive region. Nevertheless, the position of the *Emigrant* elements with respect to the predicted genes greatly varies among the different subfamilies analyzed (see table 1). While 55.5% of the *EmiB3* elements are found at less than 500 nt from an ORF, and

42% of *EmiO* are located within or close to a predicted gene, the vast majority of the *EmiA2* elements (85%) are located at more than 500 nt from any ORF.

Among the 53 *Emigrant* elements located at less than 500 nt from the closest ORF, 46.5% are located downstream, 27.5% are located upstream, and 26% are located within a predicted gene. These elements can affect promoter activity, splicing, transcriptional termination, or RNA stability, as well as the coding capacity of the ORF. We have thus analyzed these insertions in some detail, and figure 3 shows examples of such close-gene insertions. Figure 3A shows an *EmiO* element, found within the transcribed downstream region of the *Det1* gene, as an example of an element lying downstream of

## A



## B

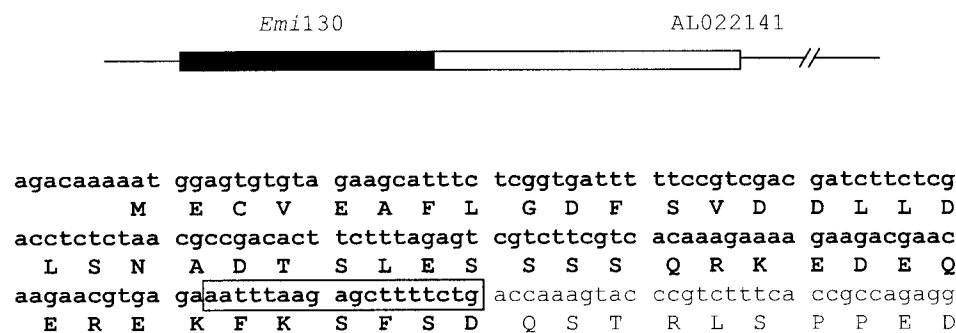


FIG. 3.—*Emigrant* elements inserted within ORFs. Schematic representation of the insertions of *Emi116* (A) and *Emi130* (B) within ORFs. Open boxes represent coding sequences, and filled boxes represent *Emigrant* elements. The name of the gene or the accession number of the ORF in which the *Emigrant* elements are inserted is indicated. The nucleotide and deduced amino acid sequences are shown under the scheme. Sequences corresponding to the *Emigrant* element are in bold, and the TIR sequence is boxed. The poly-A site is also indicated, and the polyadenylation signal is underlined.

an ORF. The availability of the genomic and the cDNA sequence for the *Det1* gene has allowed us to determine that the transcription of the *Det1* gene stops within the *Emigrant* element, probably using polyadenylation sequences provided by *Emi116*. Figure 3B shows an example of an *Emi0* element located within a predicted gene coding for a GATA-like transcription factor. The insertion of the element has provided a new putative ATG and 48 new amino acids within the C-terminal region of the protein. We also found five *Emigrant* elements lying at less than 500 nt from two different ORFs. The insertion of those elements could potentially affect the expression of both the upstream and the downstream genes. Alternatively, the insertion of an *Emigrant* element in these extremely short intergenic regions could help to avoid transcriptional interference between both genes. Related to this, it is interesting to note that it has been proposed that some MITEs could act as matrix attachment regions isolating their neighboring genes (Tikhonov, Bennetzen, and Avramova 2000). This possible effect of MITE insertion could be particularly useful in *Arabidopsis*, which has a very compact genome, and genes are sometimes found extremely close to one another.

## Discussion

### A New Approach to Study MITE Evolution

Most MITE families described to date are characterized by a high degree of sequence and size conservation (Bureau and Wessler 1992; Casacuberta et al. 1998; Feschotte and Mouchès 2000b; Oosumi, Garlick, and Belknap 1996; Tu 1997; Yang et al. 2001). However, after insertion MITEs are subjected to random mutation, and the sequence and size homogeneity of a particular MITE family will decrease with time. Ancient divergent copies, and low copy number families of MITEs, are difficult to detect by sequence similarity-based search methods and have probably been missed in searches performed to date. In order to get access to *Emigrant* divergent elements and analyze the evolution of this particular family of MITEs we developed a computer program, TRANSPO, based solely on the presence of relatively conserved TIR sequences. A similar approach has recently been used to describe new MITE families in *Anopheles gambiae*. Tu (2001) developed a computer program, named FINDMITE, that looks in a given sequence for unknown TIRs flanked by short direct repeats. The sequences of both the TIR and the

direct repeat are not fixed, which makes FINDMITE suitable for looking for new MITEs but not for finding old elements that have very imperfect TIRs or have lost the short direct duplications. TRANSPO can be used for looking for the presence of a particular TIR sequence with a low stringency and can detect elements that have lost the short direct duplication flanking them, which makes it more suitable for looking for ancient elements with a known TIR. The program TRANSPO has allowed us to detect all the sequences present within the genome of *Arabidopsis* that contain TIRs 75% identical to the previously defined *Emigrant* TIR (Casacuberta et al. 1998), separated by more than 200 nt and less than 700 nt. The presence of target site duplications and RESites for the different *Emigrant* groups described here has allowed us to confirm the mobile nature of these sequences.

#### Different Amplification Bursts of *Emigrant* Elements have Occurred During *Arabidopsis* Evolution

The 47 *Emi0* sequences are too divergent to be included in any of the nine *Emigrant* subfamilies defined here. The high divergence of these elements suggests that they represent old *Emigrant* insertions that have accumulated a high number of mutations. The phylogenetic analysis of the other *Emigrant* elements shows that they belong to different subfamilies with different degrees of variability. Whereas the 20 *EmiA2* elements are highly homogeneous, the *EmiB3* subfamily is highly variable. This suggests that different amplification bursts have occurred at different times during the evolution of *Arabidopsis*, giving rise to these different subfamilies—the more variable a subfamily is, the more ancient the amplification burst that has generated it should be. The start-type topology of the *Emigrant* subfamilies in the different trees suggests that each subfamily has been generated from the amplification of a single *Emigrant* element. This could be explained by the presence of only one active or master element capable of amplification at a particular moment, as predicted by the master gene model developed for SINES (Deininger and Batzer 1995), or simply by assuming that the amplification of MITEs is an extremely rare event occurring stochastically on any *Emigrant* element that would thus act as a founder element for a new subfamily. In any of both scenarios the result will be that only very few *Emigrant* elements had been amplified during the evolution of *Arabidopsis* and that the insertion dynamics of *Emigrant* elements has been very similar to that of most SINES, in spite of the important differences in the transposition mechanisms.

This evolutionary dynamics shown here for the *Emigrant* element is probably shared by other elements, as the presence of highly conserved subfamilies within a single host genome has been described for other MITEs (Tu 2001; Yang et al. 2001).

#### Elements Close to Genes have been Preferentially Maintained During *Arabidopsis* Evolution

Although MITEs seem to target very highly AT-rich regions, they have often been found close to tran-

scribed sequences (Wessler, Bureau, and White 1995; Yang et al. 2001; Feschotte, Jiang, and Wessler 2002). Nevertheless, it is not known if this preferential location is the result of their insertion specificity. On the other hand, a recent survey failed to detect transposon insertions in *A. thaliana* coding regions, suggesting a purifying selection against deleterious mutation (Le, Wright, and Bureau 2000). The presence of mobile elements at particular locations within a genome is the result of their transpositional activity and of the selection of the best fit genomes. Thus, elements transposing randomly within a genome can be found at particular locations as the result of a positive selection of their insertion within those sites or the negative selection of insertion in other locations. The effect of target site specificity should be more easily detected for recently inserted elements, whereas the effect of selection will be more apparent for ancient insertions. The comparison of the distribution of ancient versus recent elements should reveal the effect of selection and thus the impact of transposon insertions. So, we compared the relative distribution of the different subfamilies of *Emigrant* elements, which represent amplification bursts occurring at different times of the evolution of *Arabidopsis*, with respect to predicted genes in order to determine their insertion specificity as well as the effect of selection and the impact of *Emigrant* insertions.

*EmiA2* is the most homogeneous subfamily described here, both in sequence and size, and probably represents the most recent burst of amplification of *Emigrant* elements. Eighty-five percent of the 20 *EmiA2* elements lie at more than 500 nt from the closest ORF (see table 1). The genome of *Arabidopsis* is extremely compact, and the intergenic regions are very short. The mean size of *Arabidopsis* genes is 2 kbp and there is one gene every 5 kbp, which implies that the mean distance between two genes is only 3 kbp (The Arabidopsis Genome Initiative 2000). Thus genic regions occupy 40% of the genome space, and the regions closely linked to the genes, that most probably contain gene regulatory regions (arbitrarily taken here as 500 nt) occupy 20% of the genome space, which means that 60% of the genome is occupied by genes and their potentially regulatory regions. The regions not linked to genes occupy only 40% of the genome space (20% the region arbitrarily defined here as between 500 and 1,000 nt, and 20% the region arbitrarily defined here as >1,000 nt). The distribution of *EmiA2* elements is thus far from random, with *Emigrant* elements inserting preferentially far from ORFs. This difference is statistically highly significant, with a chi-square value of 19.76, whereas the chi-square value with three degrees of freedom and 99% probability is 11.34. The strict specificity of *Emigrant* and other MITEs for the TA dinucleotide as insertion site, as well as the preference for very highly AT-rich regions (74.3% AT in the case of *Emigrant*) probably helps these elements to avoid genes even in extremely compact genomes such as that of *Arabidopsis*.

This preference for regions far from genes is less pronounced for other *Emigrant* subfamilies. More than 50% of the *EmiB3* elements, and 43% of the *Emi0* group

of elements, are less than 500 nt from the closest ORF. Interestingly, the *Emi0* group contains the most divergent *Emigrant* elements, and the *EmiB3* subfamily is one of the most variable subfamilies, suggesting that both the *Emi0* group and the *EmiB3* subfamily represent the most ancient insertions and have been subjected to selection for a relatively long period of time. In particular, the differences of distribution of *Emi0* and *EmiA2* elements are statistically significant, with a chi-square value of 19.76 (three degrees of freedom; 99% probability  $\chi^2 = 11.34$ ). The other *Emigrant* subfamilies show different distribution patterns with respect to ORFs, but although the low number of elements makes it difficult to draw conclusions in some cases, the more variable a subfamily, the more closely it is associated to genes.

A possible explanation for these results would be a domestication of the *Emigrant* transposase that would have learned to avoid genes during evolution, inserting *Emigrant* elements farther and farther from genes. Alternatively, these results suggest that whereas *Emigrant* elements preferentially insert far from ORFs, the elements closely linked to genes are more frequently maintained during evolution. This is reminiscent of what has been shown for the Alu family of SINEs in the human genome. Alu's tend to insert in AT-rich regions, and recently transposed Alu subfamilies are found in gene-poor regions, whereas ancient Alu subfamilies are found preferentially in GC-rich regions closely associated to genes (International Human Genome Sequencing Consortium 2001). Although other possible explanations have been pointed out (Brookfield 2001; Batzer and Deininger 2002), it has been proposed that a positive selection in favor of the minority of Alu's in GC-rich DNA could explain the difference in distribution between old and new Alu subfamilies (International Human Genome Sequencing Consortium 2001). *Emigrant* and other MITEs resemble SINEs in their short size and their high copy number, and here we have shown that their amplification dynamics is also very similar. Thus, although the redistribution of *Emigrant* elements could also be the result of a preferential loss of those elements located far from genes, it is tempting to hypothesize that, as has been proposed for Alu elements within the human genome, there has been a positive selection for *Emigrant* elements lying within or close to genes during *Arabidopsis* evolution.

#### A Role for *Emigrant* Elements in the Evolution of *Arabidopsis* Genes

Over the last 10 years a growing body of evidence has pointed toward a modular nature for the regulation of gene expression. Promoters, and probably terminators, are constituted by a complex array of regulatory elements. Most of these elements are found in many different promoters or terminators, although each promoter-terminator contains a particular combination of them. With the completion of genome sequencing projects it has become more and more clear that coding regions of eukaryote genes are also often composed of domains or modules that have been reshuffled during

evolution. There are many different mechanisms that can account for the amplification and distribution of particularly successful coding or regulatory modules, but short replicative elements such as SINEs and MITEs would be particularly suitable candidates for such a function. SINEs are frequently found within or close to genes in *Arabidopsis* (Lenoir et al. 2001) and other organisms (Makalowski 1995), and it has been recently found that some of them can play an important biological role as coding or transcriptional regulatory regions (Shimamura et al. 1998; Ferrigno et al. 2001; Goodyer, Zheng, and Henty 2001; Landry, Medstrand, and Mager 2001; Ackerman et al. 2002). Moreover, it has been proposed that B2 SINEs may have the potential to distribute a functional pol II promoter throughout the genome (Ferrigno et al. 2001). Here we show that a high number of *Emigrant* elements within potential promoters, terminators, introns, and coding sequences which may affect gene coding capacity or regulation have been conserved during evolution. Although molecular experiments to determine unambiguously the impact of these insertions have yet to be performed, our results suggest that the insertion of *Emigrant* elements has played an important role in the evolution of *Arabidopsis* genes. MITEs, as has been proposed for SINEs, could have been recruited by genomes in an evolutionary mechanism to generate novel coding or regulatory sequences. The fact that MITEs can probably be excised (Petersen and Seberg 2000; Yang et al. 2001) makes them even more suitable for such a function.

#### Acknowledgments

We would like to thank E. Casacuberta, M. L. Espinás, J. Martínez-García, and P. Puigdomènech for critical reading of the manuscript. This work was supported by a grant from the Ministerio de Ciencia y Tecnología to J.M.C. (grant BIO2000-0953).

#### LITERATURE CITED

- ACKERMAN, H., I. UDALOVA, J. HULL, and D. KWIATKOWSKI. 2002. Evolution of a polymorphic regulatory element in interferon-gamma through transposition and mutation. *Mol. Biol. Evol.* **19**:884–890.
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- BATZER, M. A., and P. L. DEININGER. 2002. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**:370–379.
- BROOKFIELD, J. F. 2001. Selection on Alu sequences? *Curr. Biol.* **11**:R900–R901.
- BUREAU, T. E., and S. R. WESSLER. 1992. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* **4**:1283–1294.
- CASACUBERTA, E., J. M. CASACUBERTA, P. PUIGDOMÈNECH, and A. MONFORT. 1998. Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. *Plant J.* **16**:79–85.
- CASACUBERTA, E., P. PUIGDOMÈNECH, and A. MONFORT. 2000. Distribution of microsatellites in relation to coding sequences within the *Arabidopsis thaliana* genome. *Plant Sci.* **157**:97–104.



- DEININGER, P. L., and M. A. BATZER. 1995. SINE master genes and population biology. Pp. 43–60 in R. J. Maraia, ed. *The impact of short interspersed elements (SINEs) on the host genome*. RG Landes Company, Austin, Tex.
- DELATTRE, M., and P. HANSEN. 1980. Bicriterion cluster analysis. *IEEE Trans. Pattern Anal. Mach. Intelligence* **4**:277–291.
- FELSENSTEIN, J. 1989. PHYLIP—phylogeny inference package (version 3.56). *Cladistics* **5**:164–166.
- FERRIGNO, O., T. VIROLLE, Z. DJABARI, J. P. ORTONNE, R. J. WHITE, and D. ABERDAM. 2001. Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nat. Genet.* **28**:77–81.
- FESCHOTTE, C., N. JIANG, and S. R. WESSLER. 2002. Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* **3**:329–341.
- FESCHOTTE, C., and C. MOUCHES. 2000a. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. *Mol. Biol. Evol.* **17**:730–737.
- . 2000b. Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: characterization of the Mimo family. *Gene* **250**:109–116.
- GOODYER, C. G., H. ZHENG, and G. N. HENDY. 2001. Alu elements in human growth hormone receptor gene 5′ untranslated region exons. *J. Mol. Endocrinol.* **27**:357–366.
- GORDON, A. D. 1999. *Classification*. Chapman & Hall/CRC, New York.
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–922.
- JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 in H. N. Munro, ed. *Mammalian protein metabolism*. Academic Press, New York.
- KUMAR, A., and J. BENNETZEN. 1999. Plant retrotransposons. *Annu. Rev. Genet.* **33**:479–532.
- LANDRY, J. R., P. MEDSTRAND, and D. L. MAGER. 2001. Repetitive elements in the 5′ untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. *Genomics* **76**:110–116.
- LE, Q. H., S. WRIGHT, and T. BUREAU. 2000. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **97**:7376–7381.
- LENOIR, A., L. LAVIE, J. L. PRIETO, C. GOUBELY, J. C. COTE, T. PELISSIER, and J. M. DERAGON. 2001. The evolutionary origin and genomic organization of SINEs in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **18**:2315–2322.
- MAKALOWSKI, W. 1995. SINEs as a genomic scrap yard: an essay on genomic evolution. Pp. 81–104 in R. J. Maraia, ed. *The impact of short interspersed elements (SINEs) on the host genome*. RG Landes Company, Austin, Tex.
- MYERS, G. 1998. A fast bit-vector algorithm for approximate string matching based on dynamic programming. *Proc. Ninth Combinatorial Pattern Matching Conference*. Springer-Verlag LNCS Series **1448**:1–13.
- NEI, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- OOSUMI, T., B. GARLICK, and W. R. BELKNAP. 1996. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J. Mol. Evol.* **43**:11–18.
- PETERSEN, G., and O. SEBERG. 2000. Phylogenetic evidence for excision of Stowaway Miniature Inverted-Repeat Transposable Elements in Triticeae (Poaceae). *Mol. Biol. Evol.* **17**:1589–1596.
- ROZAS, J., and R. ROZAS. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**:174–175.
- SAITOU, N., and N. NEI. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SHIMAMURA, M., M. NIKAIIDO, K. OHSHIMA, and N. OKADA. 1998. A SINE that acquired a role in signal transduction during evolution. *Mol. Biol. Evol.* **15**:923–925.
- THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, population-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- THE ARABIDOPSIS GENOME INITIATIVE. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**:796–815.
- TIKHONOV, A. P., J. L. BENNETZEN, and Z. V. AVRAMOVA. 2000. Structural domains and matrix attachment regions along colinear chromosomal segments of maize and sorghum. *Plant Cell* **12**:249–264.
- TU, Z. 1997. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **94**:7475–7480.
- . 2001. Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **98**:1699–1704.
- TURCOTTE, K., S. SRINIVASAN, and T. BUREAU. 2001. Survey of transposable elements from rice genomic sequences. *Plant J.* **25**:169–179.
- WESSLER, S., T. BUREAU, and S. E. WHITE. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**:814–821.
- YANG, G., J. DONG, M. B. CHANDRASEKHARAN, and T. C. HALL. 2001. *Kiddo*, a new transposable element family closely associated with rice genes. *Mol. Genet. Genomics* **266**:417–424.
- ZAHN, C. T. 1971. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans. Comput.* **C-20**:68–86.
- ZHANG, Q., J. ARBUCKLE, and S. R. WESSLER. 2000. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions in maize. *Proc. Natl. Acad. Sci. USA* **97**:1160–1165.
- ZHANG, X., C. FESCHOTTE, Q. ZHANG, N. JIANG, W. R. EGGLESTON, and S. R. WESSLER. 2001. P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc. Natl. Acad. Sci. USA* **98**:12572–12577.

PIERRE CAPY, reviewing editor

Accepted August 22, 2002



Review

# Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes

Josep M. Casacuberta\*, Néstor Santiago

*Department of Molecular Genetics, IBMB-CSIC, Jordi Girona 18, 08034 Barcelona, Spain*

Received 30 December 2002; received in revised form 19 February 2003; accepted 25 March 2003  
Received by A.J. van Wijnen

## Abstract

Transposons are genetic elements that can move, and sometimes spread, within genomes, and that constitute an important fraction of eukaryote genomes. Two types of transposons, long terminal repeat (LTR)-retrotransposons and miniature inverted-repeat transposable elements (MITEs), are highly represented in plant genomes, and can account for as much as 50–80% of the total DNA content. In the last few years it has been shown that, in spite of their mutagenic capacity, both LTR-retrotransposons and MITEs can be found associated to genes, suggesting that their activity has influenced the evolution of plant genes. In this review we will summarise recent data on the control of the activity and the impact of both LTR-retrotransposons and MITEs on the evolution of plant genes and genomes.

© 2003 Elsevier Science B.V. All rights reserved.

*Keywords:* Transposable element; Silencing; Stress; Transcription

## 1. Structural characteristics of retrotransposons and miniature inverted-repeat transposable elements (MITEs)

Transposable elements (TEs) are usually classified in two different groups according to their mode of transposition: class I elements transpose through an RNA intermediate, while class II elements transpose directly via a DNA intermediate. The replicative mode of transposition of retrotransposons can rapidly increase their copy number, which can be extremely high in eukaryote genomes. On the contrary, class II TEs are usually present in a low copy number, probably as a consequence of their ‘cut and paste’ mechanism of transposition. MITEs constitute a particular type of TEs with characteristics of both class I and class II elements. While their structural characteristics are similar to defective class II elements, their high copy number and the existence of subfamilies showing high sequence and size

conservation suggest that they can be amplified from a very limited number of progenitors (Feschotte et al., 2002a).

Retrotransposons are the most widespread class of eukaryotic TE. They can be divided into two principal groups, the long terminal repeat (LTR) and the non-LTR retrotransposons. LTR retrotransposons are further subdivided into the Ty1-*copia* and the Ty3-*gypsy* groups, while non-LTR retrotransposons are subdivided into long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). LTR retrotransposons have long terminal repeats (LTRs) of variable length (from 100 bp to several Kb) that flank the internal coding region. Both Ty1-*copia* and Ty3-*gypsy* groups encode a number of proteins in two major genes, *gag* and *pol* that are synthesised as a polyprotein, which is cleaved into functional peptides by an element-encoded protease. *Gag* encodes structural proteins important for the packaging of retrotransposon RNA while the *pol* gene encodes the enzymatic activities needed for the retrotransposon life cycle. The order in which these enzymatic activities are encoded within the *pol* gene differs between Ty1-*copia* and Ty3-*gypsy* elements. While integrase precedes the reverse transcriptase, and RNaseH, and is located just downstream of the protease coding capacity in Ty1-*copia* elements, it is located at the end of the *pol* gene in Ty3-*gypsy* elements (Fig. 1). Transcription of

*Abbreviations:* LTR, long terminal repeat; MITE, miniature inverted-repeat transposable element; TIR, terminal inverted repeat; VLP, virus like particle; PTGS, post-transcriptional gene silencing; TGS, transcriptional gene silencing; ORF, open reading frame; MAR, matrix attachment region.

\* Corresponding author. Tel.: +34-93-400-6142; fax: +34-93-204-5904.

*E-mail address:* [jcsgmp@cid.csic.es](mailto:jcsgmp@cid.csic.es) (J.M. Casacuberta).

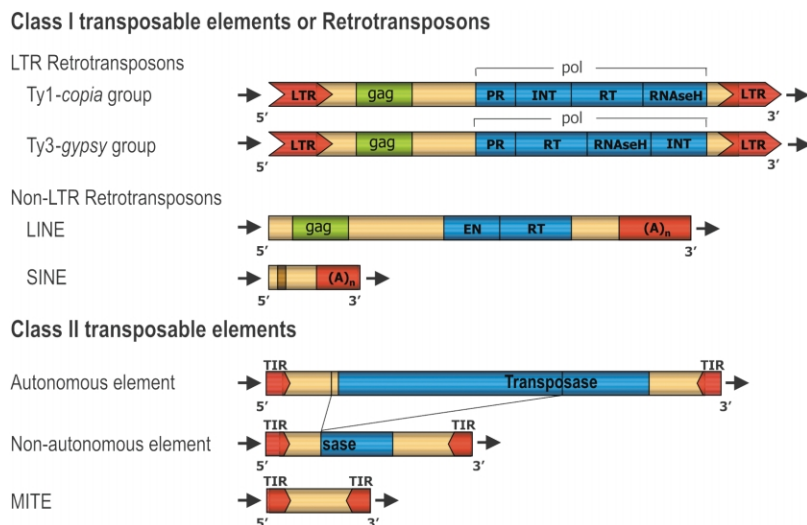


Fig. 1. Structure of the different types of plant transposable elements.

LTR retrotransposons starts at the 5' LTR and ends at the 3' LTR, and LTRs usually contain regulatory sequences for both promoting and terminating transcription of the element.

Non-LTR retrotransposons lack LTRs and are transcribed from an internal promoter. LINES, like LTR-retrotransposons, have *gag* and *pol* genes encoding structural and enzymatic activities, and it has been proposed that LINES could be the precursors of LTR-retrotransposons (Xiong and Eickbush, 1990). On the contrary the small retrotransposons called SINEs are very different from the rest, and they do not have any coding capacity. SINEs derive from polymerase III transcripts (like tRNAs or 7SL RNAs), and use LINE-specified functions to transpose (Kajikawa and Okada, 2002). Transcription is the first step of retrotransposition, as the synthesised RNA is used as template for reverse transcription to generate a new copy of the element prior to integration. In the case of retroelements with coding capacity (e.g. LTR retrotransposons and LINES), this RNA is also used as mRNA for the synthesis of the encoded proteins.

Most class II elements transpose by a 'cut and paste' mechanism mediated by a transposase that recognise their short terminal inverted repeated sequences (TIRs). The presence of certain conserved motives within transposases, as well as sequence and length similarities in the TIRs and in the target site duplications generated upon insertion, allow to classify eukaryotic class II transposons in 7 different superfamilies (Robertson, 2002; Feschotte et al., 2002b). Internal deletions within the coding sequences of transposons can generate defective elements that are no longer able to transpose autonomously, but can be transactivated by active transposases expressed by related elements. MITE structure resembles to that of defective class II transposons in the absence of the coding capacity and the presence of TIRs. However, the high copy number and the sequence and size conservation of each MITE subfamily suggest that

MITEs can be highly amplified from a limited number of progenitors, which is a characteristic of class I elements. For these reason MITEs remained long time unclassified. Recently, however, a direct link between a MITE family and a potential autonomous element was found suggesting that MITEs are a particular type of defective class II transposons. The example of transposase-encoding element related to a MITE family was found in Arabidopsis, where an element closely related to the *Emigrant* family of MITEs was found to encode a *pogo*-like transposase (Feschotte and Mouches, 2000). Since then, transposase-encoding elements related to most MITE families have been found in plants and other organisms (Feschotte and Wessler, 2002; Feschotte et al., 2002b; Le et al., 2000; Turcotte et al., 2001; Turcotte and Bureau, 2002; Yu et al., 2000; Zhang et al., 2001), and it has been proposed that MITEs are a particular type of defective class II elements mobilised by transposases encoded by their related autonomous elements (Feschotte et al., 2002a,b). Nevertheless, the mechanism by which these elements are amplified remains unknown (Fig. 2).

## 2. Retrotransposon and MITE copy number and plant genome size

Both LTR and non-LTR retrotransposons are widespread in plant genomes where they can reach very high copy numbers. For example, the Ty1-*copia* elements, BARE-1 from barley and Opie-1 and Huck2 from maize are present in 20,000–200,000 copies (Meyers et al., 2001; SanMiguel et al., 1996; Vicient et al., 1999), the Ty3-*gypsy* Cinfu-1 is present in 20,000 copies in the genome of maize, the LINE Del2 is present at 250,000 copies in Lilium (Leeton and Smyth, 1993), and the SINE TS is present at 50,000 copies in tobacco (Yoshioka et al., 1993). Retrotransposons have been found to be present at a high copy number in

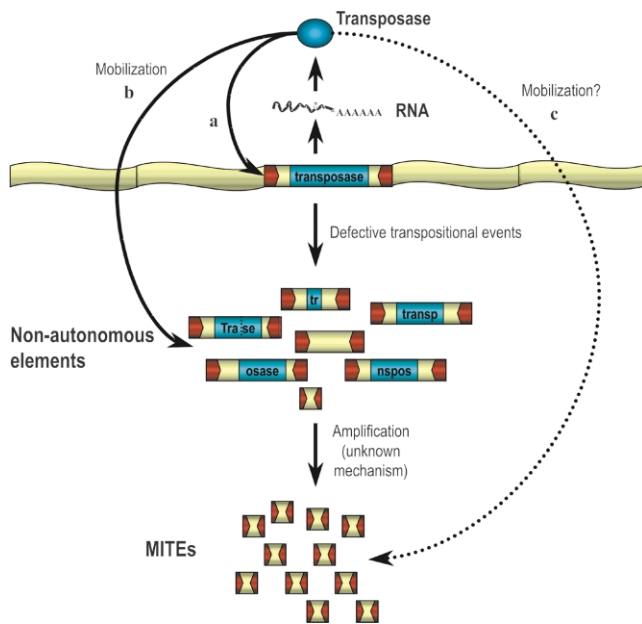


Fig. 2. Model for the origin of MITEs. DNA transposons code for a transposase that directs their mobilization (a). Incomplete transpositional events can generate defective copies, no longer able to autonomously transpose. These elements can be mobilised in trans by transposases coded by their related autonomous elements (b). Some short non-autonomous copies could be amplified by an already unknown mechanism, and generate a family of MITEs. Mobilization of such elements is supposed to be catalysed by the transposase encoded by the autonomous transposon (c) or by related elements.

heterochromatic regions including centromeres (Arabidopsis Genome Initiative, 2000; Feng et al., 2002), but they are also found interspersed with genes. Indeed, gene-flanking regions have been frequently found to contain sequences related to LTR retrotransposons (White et al., 1994). Retrotransposons can insert within pre-existing retrotransposons giving rise to nested structures, as it has been shown analysing the *adh1* region in maize (SanMiguel et al., 1996). The comparison of this region between maize and sorghum showed that the later was devoid of retrotransposons, suggesting that they were inserted after the divergence of these two species. This hypothesis was further confirmed by the analysis of the variability between the two LTRs of these elements that allowed to date the insertions at 2 to 6 million years, thus coinciding with the time of species divergence between sorghum and maize (SanMiguel et al., 1998). This study and the high copy number of retrotransposons in maize compared to sorghum suggest that the difference in genome size between both species is largely due to an important retrotransposon accumulation in maize after the divergence of both species. This is also the case of two *Oryza* species, *O. sativa* and *O. australiensis*, in which the variation in number of RIRE1 copies alone can explain one-third of their differences in genome size (Vicent and Schulman, 2002). In general, differences in retrotransposon content is probably one of the reasons of the high variability

of genome size in plants. While in small genomes like *Arabidopsis thaliana* retrotransposons represent only the 4–8% of the genome, in large genomes like maize they can account for more than 50–80% of their DNA content (Kumar and Bennetzen, 1999). The copy number of retrotransposons increases with their activity due to their replicative mechanism of transposition and has probably played a major role in plant genome expansion (Bennetzen and Kellogg, 1997), but the presence of repetitive sequences within the genome can also favour recombination events between them reducing genome size (Devos et al., 2002). In the case of LTR-retrotransposons, it has been shown that recombination between LTRs, to generate solo LTRs, can be an important mechanism to reduce the number of copies of particular retrotransposons (Vicent et al., 1999). Therefore, the copy number of a retrotransposon in a given genome will be the result of its retrotransposition activity but also of the ability of the genome to eliminate the newly inserted copies, and it will be different for each retrotransposon family and each host genome.

MITEs are also abundant in plant genomes, but differently to what it was found for retrotransposons, the analysis of the rice genome has shown that they are preferentially located within euchromatic regions (Feng et al., 2002; Zhang et al., 2000), where they can also form nested structures (Jiang and Wessler, 2001). Most MITE families are present in hundreds of copies of different subfamilies in plant genomes (reviewed in Feschotte et al., 2002b), but in some cases, like the maize *mPIF* element, which is present in more than 6000 copies (Zhang et al., 2001), their copy number can be particularly high. The recent analysis of the sequence of rice chromosome 4 has shown that MITEs constitute almost 50% of the total number of repetitive DNA elements (Feng et al., 2002). On the other hand, although MITEs are very short elements compared to retrotransposons, they can also account for an important fraction of plant genomes. This is the case of the *Stowaway* superfamily of elements that account for 2% of the rice genome (Mao et al., 2000). This high invasivity differentiates MITEs from other class II-related mobile elements that are maintained at a very low copy number within genomes. Whether amplification of MITEs is part of their particular mechanism of transposition or it is an independent phenomenon remains to be elucidated. In any case, the already mentioned relationship of MITEs and class II transposons suggest that they are mobilised by enzymes closely related to class II transposases (Feschotte et al., 2002a). Transposases catalyse both the insertion and the excision of their associated transposons, and thus MITEs should be expected to excise with a certain frequency and, as it is the case of class II transposons, leave excision footprints upon excision. Indeed, the only active MITE described to date, the rice *mPing* element, excises imprecisely and leaves different types of excision footprints (Nakazaki et al., 2003; Kikuchi et al., 2003). Nevertheless, although possible excision footprints have been found for

other MITEs (Yang et al., 2001; Petersen and Seberg, 2000; Casacuberta, unpublished), most MITE insertions seem relatively stable, as particular MITEs are frequently found at the same genomic position in different related species (Wessler, 1998). Abortive transpositions in which excision of an element is not followed by its reinsertion are relatively frequent for some class II transposons like the Ac/Ds elements (Gorbunova and Levy, 2000 and references therein). This mechanism could also account for a loss of MITE elements, and reverse the increase of copy number that seems to accompany their particular mode of transposition.

### 3. Retrotransposons and MITEs as mutagens

The movement of transposons, and in particular that of LTR-retrotransposons and MITEs, can generate a great variety of mutations in plant genomes. It was the characterisation of an insertional mutant in the maize *Adh* gene that allowed the first description of a retrotransposon in plants, the *Bs1* element (Johns et al., 1985). In a similar way, Tnt1, the first active retrotransposon described in plants, was also isolated after its insertion within the tobacco *Nitrate reductase* gene (Grandbastien et al., 1989). Since then, many examples of mutant phenotypes generated by retrotransposon insertions within coding sequences have been characterised (see for example Vignols et al., 1995; Takano et al., 2001). But the insertion of retrotransposons in non-coding sequences can also generate mutations. Their insertion within introns can result in tissue specific alternative splicing leading to the production of fully active or truncated proteins in different tissues (Marillonnet and Wessler, 1997; LePrince et al., 2001; Varagona et al., 1992), and the insertion of LTR-retrotransposons in non-coding regions close to genes can also modify their transcription or transcriptional termination due to the presence in their LTR of transcriptional promoters, regulators, and terminators. This ability of retrotransposons to generate mutations has been recently used as a tool to generate mutant collections in rice (Hirochika, 2001). On the other hand, although MITEs are miniature elements, their insertion can also generate mutations. Indeed, *Tourist*, the first MITE family described, was initially identified as an insertion within the maize *waxy* gene leading to a mutant phenotype (Bureau and Wessler, 1992), and the *mPing* element, the first active MITE described has been shown to be responsible for a *slender glume* mutant allele in rice (Nakazaki et al., 2003).

This capacity to generate mutations of both retrotransposons and MITEs seem somehow contradictory with the high copy number those elements can reach within fully viable plant genomes. The first possible explanation to this dilemma is the high prevalence of polyploidy in plants that can buffer the mutagenic activity of TEs. Indeed, the proportion of angiosperms that have experienced one or more episodes of chromosome doubling in their evolutionary

history might be more than 70% (Wendel, 2000). In addition, most MITE and retrotransposon copies present in plant genomes are probably defective elements that are no longer able to transpose and generate mutations. This is particularly clear for LTR-retrotransposons. Although the number of LTR-retrotransposons described in plants is very high (Kumar and Bennetzen, 1999) and continues to increase with the completion of genome projects, evidences for recent activity has only been obtained for a handful of them. Insertion polymorphisms between closely related species (Pearce et al., 2000; Kumar and Bennetzen, 1999), or among different varieties and populations (Vicent et al., 1999; Kalendar et al., 2000), have been obtained for a number of elements. In some cases, virus like particles (VLPs), proteins or a low level of transcription, has been detected, suggesting a low level of transpositional activity for some elements (Jaaskelainen et al., 1999; Vicent et al., 2001a, 2001b). Nevertheless, a high level of expression associated to a copy number increase within a genome has only been shown for a very few number of elements. The tobacco retrotransposon Tnt1 can be activated generating new insertions (Melayah et al., 2001), the tobacco Tto1 retrotransposon actively transposes in cell culture (Hirochika, 1993), and the rice Tos17 element can also increase its copy number in tissue culture conditions (Hirochika, 1997). Although retrotransposition seems to have been important in the evolutionary history of many plant genomes, very few plant retrotransposons have maintained their transpositional capacity during evolution.

A similar situation is found for MITEs. Although MITEs are abundant in plant genomes and the number of different MITE families described has greatly increased in the last few years (see Feschotte et al., 2002b for a review), and some of these families show insertion polymorphisms among individuals or populations of the same species (see for example Casa et al., 2000; Casacuberta et al., 1998), only one active MITE family, the rice *mPing* element, has been characterised to date (Jiang et al., 2003; Kikuchi et al., 2003; Nakazaki et al., 2003).

### 4. Control of TEs by silencing mechanisms

The existence of a high number of sequences within plant genomes that can be considered as remnants of mobile elements reveals the existence of efficient transposon inactivating mechanisms. Indeed, eukaryote genomes seem to have developed mechanisms to reduce the activity of mobile elements and control their mutagenic activity. Among them, silencing mechanisms are probably the most general and effective. Silencing was first described in transgenic plants, but related phenomena have now been described in a broad range of normal organisms. Post-transcriptional gene silencing (PTGS) is a sequence-specific RNA degradation that probably constitutes a general antiviral defence mechanism in plants, while the promoter

inactivation mechanism named transcriptional gene silencing (TGS), could be a mechanism primarily directed to abolish transcription of mobile elements (Vance and Vaucheret, 2002; Vaucheret and Fagard, 2001). Different factors influence the induction of TGS, but the presence of multiple copies of the target sequence seems to be a major factor leading to gene silencing. In the least few years examples of TE inactivation by high copy number-induced silencing have been reported. For example, the activity of the *Drosophila*I element (a LINE retrotransposon), is repressed by the introduction of multiple copies of a transgene expressing a small internal region of this element (Jensen et al., 1999), and the tobacco Tto1 retrotransposon becomes silent in Arabidopsis after several rounds of retrotransposition leading to a copy number increase (Hirochika et al., 2000). On the other hand, the presence of short interfering RNA (siRNA), a mediator of silencing, corresponding to retrotransposon sequences (Hamilton et al., 2002; Llave et al., 2002), also confirms that these elements are indeed targeted by genome silencing mechanisms.

Silenced promoters are hypermethylated and have an increased resistance to DNase I, suggesting that they form secondary DNA structures that attract methylation and heterochromatin components (Vaucheret and Fagard, 2001). Consistent with this, it has been shown that mutations affecting different chromatin remodelling factors reactivate silent mobile elements (Miura et al., 2001; Singer et al., 2001; Wright and Voytas, 2002), as do mutants of paramutation, a phenomenon closely related to gene silencing mechanisms (Lisch et al., 2002). Moreover, it has been shown that the inactivation of the tobacco Tto1 retrotransposon in Arabidopsis, which is associated to its hypermethylation, can be reversed in a methylation deficient context (Hirochika et al., 2000).

The strong hypermethylation of silenced elements can accelerate their mutation rate rendering them definitively inactive. On the other hand, the chances for an evolutionary loss of a TE increase with the time it is maintained inactive. In yeast it has been shown that recombination between LTRs allows a high turn over of Ty retrotransposons, and the maintenance of small populations of active elements (Jordan and McDonald, 1999). In plants, phylogenetic analysis of the retrotransposon Reverse Transcriptase (RT) gene showed evidences of purifying selection in species with low copy numbers of Ty1-copia elements, suggesting also a high turn-over of low copy number populations of retrotransposons (Navarro-Quezada and Schoen, 2002). Nevertheless, plants also contain very high copy number retrotransposon families that are in most cases inactive, and that have not been eliminated, probably because plant genomes can support huge variations in their genome content without important consequences.

The high copy number that MITEs and retrotransposons can attain in plant genomes, as well as the existence of a few active MITE and retrotransposon families, suggests that some elements can escape to the genomic control

mechanisms. TGS is directed against repeated sequences, and its effect is the inactivation of the promoters contained within these sequences. Non-autonomous defective TEs are a particular type of transposons in which the mobilised sequence is different from the one that codes for the enzyme needed for transposition. In the case of MITEs, the transposase-encoding element is probably present in one or very few copies, while the transposing and proliferating unit, the MITE itself, reaches very high copy number. Under this situation, silencing mechanisms will be directed towards the MITE, which cannot be inactivated by TGS because it is not transcribed, while the very low copy number transposase-encoding element will not be a target of silencing. As the transposase-encoding element is supposed to share the terminal sequences with its deletion derivative, the MITE, a MITE-directed TGS could also affect the expression of the transposase. Nevertheless, it has been suggested that MITEs could be transactivated by transposases encoded by related elements other than the source element, with very limited sequence similarity with the MITE itself (Feschotte et al., 2002b).

Although MITEs cannot be inactivated by TGS, silencing-associated processes, such as methylation, could influence its ability to transpose. Methylation has been shown to modify the capacity of the TIRs of different class II transposons to bind the transposase, and thus influence their transposon competence (Benito and Walbot, 1997; Ros and Kunze, 2001). Holomethylated Ac/Ds elements cannot bind Ac transposase and are unable to transpose, while hemimethylated elements bind the transposase with high affinity and transpose actively (Ros and Kunze, 2001). It has been suggested that the preference of Ac transposase for hemimethylated DNA could explain the link between Ac/Ds transposition and DNA replication (Wang et al., 1996). It is tempting to hypothesise that a similar replication-dependent transposition could help MITEs to attain the extremely high copy number they present in eukaryote genomes.

In any case, the amplification process leading to a new MITE family, which has been suggested to represent the last step of the life cycle of a transposon preceding its immediate death (Feschotte et al., 2002b) could also be interpreted as part of a strategy allowing their spread and maintenance within a genome.

Very short defective elements related to LTR-retrotransposons, known as TRIMs have also been described (Witte et al., 2001), although a direct link to an autonomous element has not been found yet. Nevertheless, in this case, and differently to what happens with MITEs, these short elements are supposed to transpose through an RNA intermediate, and can thus be inactivated. In addition, not only the short defective elements but also retrotransposons themselves can reach very high copy numbers. Active retrotransposons have thus to escape silencing by other mechanisms. Individual elements located in particular locations within chromosomes could perhaps be less sensible to inactivation by silencing allowing them to amplify. Related to this, it has been shown that the repeated induction of a

promoter located nearby induces the progressive demethylation and de-silencing of a particular copy of the mouse IAP retrotransposon (Barbot et al., 2002). Nevertheless, the effect of particular chromosomal locations on silencing cannot explain the simultaneous transcription of multiple copies of a retrotransposon family. This is the case, for example, of the tobacco retrotransposon Tnt1 that is expressed as a population of related but different sequences originated by the concomitant transcription of many different elements (Casacuberta et al., 1995). Another possibility is that their high sequence variability could help LTR-retrotransposons to escape silencing. Interestingly, it has been shown that the sequence variability of the tobacco Tnt1 retrotransposon is not homogeneously distributed along the sequence, the promoter, which is the target of the TGS mechanisms, being the most variable region (Vernhettes et al., 1998) (Fig. 3). Nevertheless, it seems difficult that the sequence variability displayed by retrotransposons could be sufficient to escape the extremely efficient TGS mechanisms, which are able to detect and inactivate repeated sequences as short as 90 nt (Vaucheret et al., 1998). A particularity of the very few plant retrotransposons that have maintained their ability to transpose is that they are active only under stress situations (Grandbastien, 1998). Different reasons can be invoked to explain this association of transposon mobility with stress (see Section 5), but it would well be that the genomic silencing mechanisms are somehow relaxed under these situations, allowing TEs to temporally escape to the genomic control. Interestingly, it has recently been shown that a cold stress can lead to a severe demethylation and activation of a retrotransposon related sequence in maize (Steward et al., 2002).

## 5. Stress activation of plant LTR-retrotransposons and MITEs

Transcription is the first step of the retrotransposition process and seems to be a major controlling step for plant

retrotransposons. Transcription, and subsequent transposition, is only detectable under certain conditions that in all cases can be considered as stress conditions. In the case of Tnt1, the three different subfamilies described, Tnt1A, Tnt1B and Tnt1C (Vernhettes et al., 1998), are all transcribed under stress situations associated to plant defence reactions (Casacuberta et al., 1997; Beguiristain et al., 2001) (Fig. 3). Tnt1A is transcribed in roots, and strongly induced in leaves treated with the fungal elicitor cryptogein or methyl jasmonate (Pouteau et al., 1994; Vernhettes et al., 1997; Beguiristain et al., 2001) while Tnt1C transcription can be induced in leaves treated with salicylic acid or 2,4-D, and Tnt1B is transcribed in cell cultures (Beguiristain et al., 2001). In a similar way, Tto1 expression is induced by wounding and cell culture associated stresses (Hirochika, 1997; Takeda et al., 1999), and Tos17 activity is also strongly induced in cell culture (Hirochika, 1997). Environmental stresses can also activate retrotransposition. It has been shown that sharp microclimate changes can modify the copy number of the BARE-1 retrotransposon in wild barley (Kalendar et al., 2000).

Transcriptional regulation of both Tnt1 and Tto1 has been studied in some detail and shown to be strictly controlled. The promoter of Tnt1A contains two different boxes, located within the U3 region of the LTR, that have been shown to be important for the element's transcription and that show sequence similarities with plant defence promoters (Vernhettes et al., 1997). One of these boxes specifically interacts in vivo with proteins that are induced in defence-related stresses (Vernhettes et al., 1997). Tnt1B and Tnt1C are also expressed in tobacco under different stress situations and differ from Tnt1A in their U3 sequence that probably contains the sequences needed to control their expression (Beguiristain et al., 2001). In the case of Tto1, a 13-bp motif has been identified as a cis-regulatory sequence associated to the induction of Tto1 expression in defence-related stresses (Takeda et al., 1999). Interestingly, this motif specifically binds different MYB transcription factors, one of which, that has been named LBM1, is identical to

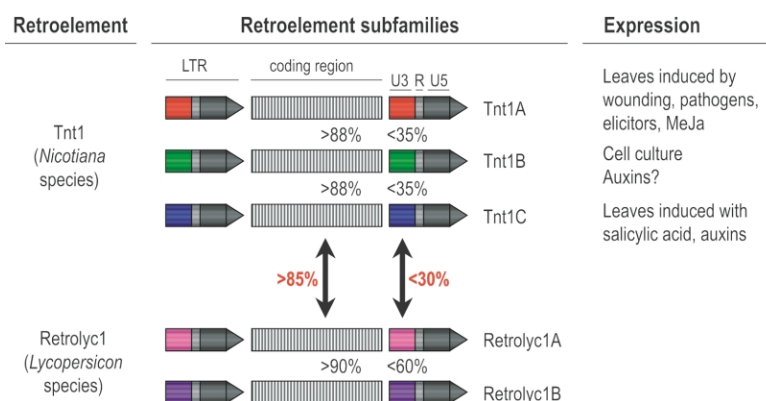


Fig. 3. Evolution of stress regulated promoters in Tnt1-related retrotransposons. Percentages of identity between U3 regions (coloured boxes) and coding regions (based on RT domain) between different subfamilies are indicated in black and between Tnt1 and Retroylc1 elements in red. Stress conditions under which the three different Tnt1 subfamilies are expressed are also indicated.

a previously described MYB-1 factor induced by virus infection (Sugimoto et al., 2000). The overexpression of another of these MYB transcription factors, NtMYB2, activates transcription of both Tto1 and the PAL defence-related gene in tobacco (Sugimoto et al., 2000). Moreover, extended homologies are found between promoters of Tto1 and an asparagus defence gene, AoPR1 (Takeda et al., 1999). All these results suggest that both Tnt1 and Tto1 are activated in defence-associated stresses because their promoters are similar to that of plant defence genes and bind the same defence induced transcription factors.

It has been proposed that retrotransposons could have captured plant defence promoters from normal genes or inversely, that they could have provided their inducible promoters to some plant defence genes (Grandbastien et al., 1997; Takeda et al., 1999). The distribution of particular successful promoters throughout the genome is a suggestive hypothesis to explain the co-ordinate regulation of groups of genes (e.g. plant defence genes). Nevertheless, evolutionary analysis of Tnt1 promoters seem to indicate that they have not evolved outside the rest of the retrotransposon sequence (Vernhettes et al., 1998), which suggests that the similarities among retrotransposon plant defence gene promoters could be the result of a convergent evolution. Retrotransposons are structurally and functionally very similar to retroviruses and it has been proposed that, as retroviruses, they could display a high sequence plasticity allowing them to rapidly evolve parts of their sequence, and acquire stress associated promoters (Casacuberta et al., 1997). In agreement with that possibility, it has been recently shown that the high variability of Tnt1 U3 region has allowed to this family of elements to evolve three different stress inducible promoters in tobacco (Beguiristain et al., 2001), and that the Tnt1-related element Retrolyc1, has evolved different promoters in tomato (Araujo et al., 2001) (Fig. 3). The driving force for the selection of stress promoters could be that stress is a rare event and thus stress induced TEs will transpose few enough to not compromise host genome viability. On the other

hand, the variability that the movement of TEs generates could also help to rapidly evolve the genome when facing a situation to which it is not well adapted, as it was initially proposed by McClintock (McClintock, 1984).

MITE transposition also seems to be induced by stress. The first active MITE described, the rice *mPing* element, increases its copy number in cell cultures (Jiang et al., 2003), and excises and reinserts at new locations in anther-derived calli (Kikuchi et al., 2003). Although the transposase responsible for *mPing* mobilisation has not been described yet, these data suggest that it is probably induced by stress.

## 6. Impact of retrotransposons and MITEs on the evolution of genes and genomes

Besides the mutagenic effect of TE insertion, transposition could generate variability useful for evolution. Transposition of LTR-retrotransposons and MITEs seems to have been a major player in plant gene evolution, as both types of elements have been frequently found associated to genes in maize (Wessler et al., 1995).

Although most MITEs insert within a TA or a TAA sequence, and, in general they seem to target very high AT-rich regions for integration, they have often been found close to transcribed sequences (Wessler et al., 1995; Yang et al., 2001; Feschotte et al., 2002a). Nevertheless, the *Arabidopsis Emigrant* element was found to be inserted relatively far from genes, and a survey of the published genomic sequence failed to detect transposon insertions in *Arabidopsis thaliana* coding regions (Le et al., 2000). This apparent contradiction has recently been solved for the *Emigrant* element. A phylogenetic analysis has shown that while young *Emigrant* sequences are located far from genes, the ancient *Emigrant* insertions are often associated to genes, suggesting that *Emigrant* elements preferentially insert far from open reading frames (ORFs), but the elements closely linked to genes are more frequently maintained during

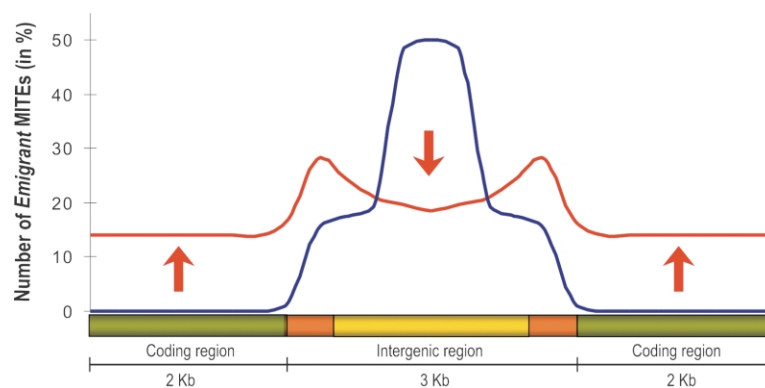


Fig. 4. Maintenance of *Emigrant* elements close to genes during *Arabidopsis* evolution. Blue line represents the percentage of young *Emigrant* elements found at a particular location with respect to *Arabidopsis* coding regions. Red line represents the percentage of old *Emigrant* elements found at a particular location with respect to *Arabidopsis* coding regions. Green boxes represent coding regions; orange boxes represent 0.5 Kb regions close to coding regions supposed to contain regulatory sequences; yellow box represents intergenic regions. Arrows show the shift of *Emigrant* distribution with time.



evolution (Santiago et al., 2002) (see Fig. 4). This is reminiscent of what has been shown for the Alu family of SINEs in the human genome. Alus tend to insert in AT rich regions, and recently transposed Alu subfamilies are found in poor-gene regions, while ancient Alu subfamilies are found preferentially in GC-rich regions closely associated to genes (International Human Genome Sequencing Consortium, 2001). This shift of MITE distribution with time could suggest that the elements close to genes have been positively selected during Arabidopsis evolution. Alternatively, it could be the result of a more frequent elimination of MITEs located far from genes. Indeed, in the case of human Alus it has been proposed that recombination plays a major role on removing them, re-creating insert-free alleles, and that this process is less frequent in gene-rich regions due to the adverse effects that deletions and unequal recombinations could have in these regions (Medstrand et al., 2002). Whatever the mechanism could be this concentration with time in gene-rich regions could be a particularity of short elements like Alus and MITEs, as it has been shown that LTR retrotransposon insertions present a totally opposite dynamics in the human genome (Medstrand et al., 2002).

The insertion of MITEs within genes can modify the promoter and terminator sequences, as well as the translational start and coding sequences (Wessler et al., 1995; Yang et al., 2001; Santiago et al., 2002; El Amrani et al., 2002). LTR-retrotransposon sequences are also frequently found associated to genes (Wessler et al., 1995), suggesting that the modification of the regulation of the expression of target genes due to the presence within the LTRs of promoter and terminator sequences has been an important mechanism in the evolution of plant genes. On the other hand, the insertion of retrotransposons in intergenic regions can also modify the expression of adjacent genes. It has been recently shown that interspecific hybridisations can reactivate the transcription of the wheat WIS2 retrotransposon, which can drive the readout synthesis of new transcripts from adjacent sequences including sense and antisense strands of genes located nearby, resulting on the silencing or the activation of these genes (Kashkush et al., 2003).

A particular case of rapidly evolving gene loci is that of plant resistance genes. Genes conferring race-specific resistance are often clustered in the genome forming large tandem repeats of highly polymorphic genes. It has been shown that the rice *Xa21* gene family contains a high number of TEs (including LTR-retrotransposons and MITEs) inserted within the different genes, and it has been proposed that besides the unequal exchange between the different copies, the high variability needed to evolve new resistance specificities is generated by the TEs insertions (Richter and Ronald, 2000).

TE insertion outside genes can also contribute to genome evolution. It has been shown that MITEs often coincide with sequences showing matrix attachment region (MAR) activity, and it has been proposed that some MITEs could

act as MARs isolating their neighbouring genes (Tikhonov et al., 2000). On the other hand, plant centromere and pericentromeric regions often contain retrotransposons that could be important for the functionality of these regions (Pelissier et al., 1996; Fukui et al., 2001; Cheng et al., 2002; Jiang et al., 2002). A centromere specific LTR-retrotransposon has been described in rice (Cheng et al., 2002) and other grasses (Miller et al., 1998) suggesting that these elements can play a role in plant chromosome organisation. Indeed, it has been recently shown that the maize centromere-specific retroelement can interact with the kinetochore protein CENH3 (Zhong et al., 2002). Interestingly, it has been shown that the interspecific cross between two different mammalian species to generate a hybrid, activates the transposition of a centromere specific retrotransposon, and it has been suggested that this could facilitate rapid karyotypic evolution (O'Neill et al., 1998).

## 7. Concluding remarks: McClintock revisited

Insertion of TEs can modify the expression or the coding capacity of genes and thus transposition can be an extremely deleterious event. For this reason, since the work of Barbara McClintock in the forties, the role of transposable elements (TEs) has been the object of an intense debate. McClintock's idea of mobile elements being active genome remodelling machines in response to stress (McClintock, 1984), was countered by those considering TEs as merely selfish or parasite elements (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). During the last 20 years examples of TEs fulfilling important roles in genomes have been described (Pardue et al., 1996) but it has appeared clearly that it is not necessary to suppose a benefit for the host to explain the presence within genomes of most TEs. Still, many evidences have shown that a high number of transposons are indeed activated by stress and that their mobility has reshaped eukaryote genomes in many ways.

## Acknowledgements

We would like to thank Carlos M. Vicient and M. Lluïsa Espinas for critical reading of the manuscript. We are also grateful to the referees for their helpful comments and suggestions. Work in our lab is funded by a grant from Spanish government (MCyT BIO2000-0953) and by the 'Centre de Referència en Biotecnologia' (Generalitat de Catalunya).

## References

- Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Araujo, P.G., Casacuberta, J.M., Costa, A.P., Hashimoto, R.Y.,

- Grandbastien, M.A., Van Sluys, M.A., 2001. Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus. *Mol. Genet. Genomics* 266, 35–41.
- Barbot, W., Dupressoir, A., Lazar, V., Heidmann, T., 2002. Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Res.* 30, 2365–2373.
- Beguiristain, T., Grandbastien, M.A., Puigdomenech, P., Casacuberta, J.M., 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiol.* 127, 212–221.
- Benito, M.I., Walbot, V., 1997. Characterization of the maize Mutator transposable element MURA transposase as a DNA-binding protein. *Mol. Cell Biol.* 17, 5165–5175.
- Bennetzen, J.L., Kellogg, E.A., 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9, 1509–1514.
- Bureau, T.E., Wessler, S.R., 1992. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4, 1283–1294.
- Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., Wessler, S.R., 2000. Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. *Proc. Natl. Acad. Sci. USA* 97, 10083–10089.
- Casacuberta, E., Casacuberta, J.M., Puigdomenech, P., Monfort, A., 1998. Presence of miniature inverted-repeat transposable elements (MITES) in the genome of *Arabidopsis thaliana*: characterisation of the Emigrant family of elements. *Plant J.* 16, 79–85.
- Casacuberta, J.M., Vernhettes, S., Audeon, C., Grandbastien, M.A., 1997. Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. *Genetica* 100, 109–117.
- Casacuberta, J.M., Vernhettes, S., Grandbastien, M.A., 1995. Sequence variability within the tobacco retrotransposon Tnt1 population. *EMBO J.* 14, 2670–2678.
- Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R., Jiang, J., 2002. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14, 1691–1704.
- Devos, K.M., Brown, J.K., Bennetzen, J.L., 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12, 1075–1079.
- Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603.
- El Amrani, A., Marie, L., Ainouche, A., Nicolas, J., Couee, I., 2002. Genome-wide distribution and potential regulatory functions of AtATE, a novel family of miniature inverted-repeat transposable elements in *Arabidopsis thaliana*. *Mol. Genet. Genomics* 267, 459–471.
- Feng, Q., Zhang, Y., Hao, P., et al., 2002. Sequence and analysis of rice chromosome 4. *Nature* 420, 316–320.
- Feschotte, C., Mouches, C., 2000. Evidence that a family of miniature inverted-repeat transposable elements (MITES) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. *Mol. Biol. Evol.* 17, 730–737.
- Feschotte, C., Wessler, S.R., 2002. Mariner-like transposases are widespread and diverse in flowering plants. *Proc. Natl. Acad. Sci. USA* 99, 280–285.
- Feschotte, C., Jiang, N., Wessler, S.R., 2002a. Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3, 329–341.
- Feschotte, C., Zhang, X., Wessler, S.R., 2002b. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), *Miniature Inverted-repeat Transposable Elements (MITES) and their relationship with established DNA transposons*, Mobile DNA II, American Society of Microbiology Press, Washington, DC, pp. 1147–1158.
- Fukui, K.N., Suzuki, G., Lagudah, E.S., Rahman, S., Appels, R., Yamamoto, M., Mukai, Y., 2001. Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat. *Plant Cell Physiol.* 42, 189–196.
- Gorbunova, V., Levy, A.A., 2000. Analysis of extrachromosomal Ac/Ds transposable elements. *Genetics* 155, 349–359.
- Grandbastien, M.A., 1998. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* 3, 181–187.
- Grandbastien, M.A., Lucas, H., Morel, J.B., Mhiri, C., Vernhettes, S., Casacuberta, J.M., 1997. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. *Genetica* 100, 241–252.
- Grandbastien, M.A., Spielmann, A., Caboche, M., 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 337, 376–380.
- Hamilton, A., Voinnet, O., Chappell, L., Baulcombe, D., 2002. Two classes of short interfering RNA in RNA silencing. *EMBO J.* 21, 4671–4679.
- Hirochika, H., 1993. Activation of tobacco retrotransposons during tissue culture. *EMBO J.* 12, 2521–2528.
- Hirochika, H., 1997. Retrotransposons of rice: their regulation and use for genome analysis. *Plant Mol. Biol.* 35, 231–240.
- Hirochika, H., 2001. Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* 4, 118–122.
- Hirochika, H., Okamoto, H., Kakutani, T., 2000. Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* 12, 357–369.
- International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Jaaskelainen, M., Mykkanen, A.H., Arna, T., Vicient, C.M., Suoniemi, A., Kalendar, R., Savilahi, H., Schulman, A.H., 1999. Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. *Plant J.* 20, 413–422.
- Jensen, S., Gassama, M.P., Heidmann, T., 1999. Taming of transposable elements by homology-dependent gene silencing. *Nat. Genet.* 21, 209–212.
- Jiang, N., Bao, Z., Temnykh, S., Cheng, Z., Jiang, J., Wing, R.A., McCouch, S.R., Wessler, S.R., 2002. Dasheng: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* 161, 1293–1305.
- Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., Wessler, S.R., 2003. An active DNA transposon family in rice. *Nature* 421, 163–167.
- Jiang, N., Wessler, S.R., 2001. Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell* 13, 2553–2564.
- Johns, M.A., Mottinger, J., Freeling, M.A., 1985. A low copy number, copia-like transposon in maize. *EMBO J.* 4, 1093–1102.
- Jordan, I.K., McDonald, J.F., 1999. Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. *Genetics* 151, 1341–1351.
- Kajikawa, M., Okada, N., 2002. LINES mobilize SINES in the eel through a shared 3' sequence. *Cell* 111, 433–444.
- Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., Schulman, A.H., 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci. USA* 97, 6603–6607.
- Kashkush, K., Feldman, M., Levy, A.A., 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nat. Genet.* 33, 102–106.
- Kikuchi, K., Terauchi, K., Wada, M., Hirano, H.Y., 2003. The plant MITE mPing is mobilized in anther culture. *Nature* 421, 167–170.
- Kumar, A., Bennetzen, J.L., 1999. Plant retrotransposons. *Annu. Rev. Genet.* 33, 479–532.
- Le, Q.H., Wright, S., Yu, Z., Bureau, T., 2000. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 97, 7376–7381.
- Leeton, P.R., Smyth, D.R., 1993. An abundant LINE-like element amplified in the genome of *Lilium speciosum*. *Mol. Gen. Genet.* 237, 97–104.
- Leprince, A.S., Grandbastien, M.A., Meyer, C., 2001. Retrotransposons of the Tnt1B family are mobile in *Nicotiana plumbaginifolia* and can induce alternative splicing of the host gene upon insertion. *Plant Mol. Biol.* 47, 533–541.
- Lisch, D., Carey, C.C., Dorweiler, J.E., Chandler, V.L., 2002. A mutation

- that prevents paramutation in maize also reverses Mutator transposon methylation and silencing. *Proc. Natl. Acad. Sci. USA* 99, 6130–6135.
- Llave, C., Kasschau, K.D., Rector, M.A., Carrington, J.C., 2002. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* 14, 1605–1619.
- Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, R.A., Wing, R.A., 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Res.* 10, 982–990.
- Marillonnet, S., Wessler, S.R., 1997. Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing. *Plant Cell* 9, 967–978.
- McClintock, B., 1984. The significance of responses of the genome to challenge. *Science* 226, 792–801.
- Medstrand, P., Van De Lagemat, L.N., Mager, D.L., 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* 12, 1483–1495.
- Melayah, D., Bonnivard, E., Chalhoub, B., Audeon, C., Grandbastien, M.A., 2001. The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors. *Plant J.* 28, 159–168.
- Meyers, B.C., Tingey, S.V., Morgante, M., 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* 11, 1660–1676.
- Miller, J.T., Dong, F., Jackson, S.A., Song, J., Jiang, J., 1998. Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* 150, 1615–1623.
- Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., Kakutani, T., 2001. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* 411, 212–214.
- Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H., Tanisaka, T., 2003. Mobilization of a transposon in the rice genome. *Nature* 421, 170–172.
- Navarro-Quezada, A., Schoen, D.J., 2002. Sequence evolution and copy number of Ty1-copia retrotransposons in diverse plant genomes. *Proc. Natl. Acad. Sci. USA* 99, 268–273.
- O'Neill, R.J., O'Neill, M.J., Graves, J.A., 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393, 68–72.
- Orgel, L.E., Crick, F.H., 1980. Selfish DNA: the ultimate parasite. *Nature* 284, 604–607.
- Pardue, M.L., Danilevskaia, O.N., Lowenhaupt, K., Slot, F., Traverse, K.L., 1996. *Drosophila* telomeres: new views on chromosome evolution. *Trends Genet.* 12, 48–52.
- Pearce, S.R., Knox, M., Ellis, T.H., Flavell, A.J., Kumar, A., 2000. Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Mol. Gen. Genet.* 263, 898–907.
- Pelissier, T., Tutois, S., Tourmente, S., Deragon, J.M., Picard, G., 1996. DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. *Genetica* 97, 141–151.
- Petersen, G., Seberg, O., 2000. Phylogenetic evidence for excision of Stowaway miniature inverted-repeat transposable elements in triticeae (Poaceae). *Mol. Biol. Evol.* 17, 1589–1596.
- Pouteau, S., Grandbastien, M.A., Boccara, M., 1994. Microbial elicitors of plant defence responses activate transcription of a retrotransposon. *Plant J.* 5, 535–542.
- Richter, T.E., Ronald, P.C., 2000. The evolution of disease resistance genes. *Plant Mol. Biol.* 42, 195–204.
- Robertson, H.M., 2002. In: Craig, N., Craigie, R., Gellert, M., Lambowitz, A. (Eds.), *Molecular evolution of DNA transposons*, Mobile DNA II, American Society of Microbiology Press, Washington, DC.
- Ros, F., Kunze, R., 2001. Regulation of activator/dissociation transposition by replication and DNA methylation. *Genetics* 157, 1723–1733.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L., 1998. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20, 43–45.
- SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., Bennetzen, J.L., 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274, 765–768.
- Santiago, N., Herráiz, C., Goñi, J.R., Messeguer, X., Casacuberta, J.M., 2002. Genome-wide analysis of the *Emigrant* family of MITEs of *Arabidopsis thaliana*. *Mol. Biol. Evol.* 19, 2285–2293.
- Singer, T., Yordan, C., Martienssen, R.A., 2001. Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes Dev.* 15, 591–602.
- Steward, N., Ito, M., Yamaguchi, Y., Koizumi, N., Sano, H., 2002. Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *J. Biol. Chem.* 277, 37741–37746.
- Sugimoto, K., Takeda, S., Hirochika, H., 2000. MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. *Plant Cell* 12, 2511–2528.
- Takano, M., Kanegae, H., Shinomura, T., Miyao, A., Hirochika, H., Furuya, M., 2001. Isolation and characterization of rice phytochrome A mutants. *Plant Cell* 13, 521–534.
- Takeda, S., Sugimoto, K., Otsuki, H., Hirochika, H., 1999. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J.* 18, 383–393.
- Tikhonov, A.P., Bennetzen, J.L., Avramova, Z.V., 2000. Structural domains and matrix attachment regions along colinear chromosomal segments of maize and sorghum. *Plant Cell* 12, 249–264.
- Turcotte, K., Bureau, T., 2002. Phylogenetic analysis reveals stowaway-like elements may represent a fourth family of the IS630-Tc1-mariner superfamily. *Genome* 45, 82–90.
- Turcotte, K., Srinivasan, S., Bureau, T., 2001. Survey of transposable elements from rice genomic sequences. *Plant J* 25, 169–179.
- Vance, V., Vaucheret, H., 2002. RNA silencing in plants - defense and counterdefense. *Science* 292, 2277–2280.
- Varagona, M.J., Purugganan, M., Wessler, S.R., 1992. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* 4, 811–820.
- Vaucheret, H., Fagard, M., 2001. Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends Genet.* 17, 29–35.
- Vaucheret, H., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Mourrain, P., Palauqui, J.C., Vernhettes, S., 1998. Transgene-induced gene silencing in plants. *Plant J.* 16, 651–659.
- Vernhettes, S., Grandbastien, M.A., Casacuberta, J.M., 1998. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Mol. Biol. Evol.* 15, 827–836.
- Vernhettes, S., Grandbastien, M.A., Casacuberta, J.M., 1997. In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon Tnt1. *Plant Mol. Biol.* 35, 673–679.
- Vicient, C.M., Jaaskelainen, M.J., Kalendar, R., Schulman, A.H., 2001a. Active retrotransposons are a common feature of grass genomes. *Plant Physiol.* 125, 1283–1292.
- Vicient, C.M., Kalendar, R., Schulman, A.H., 2001b. Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Res.* 11, 2041–2049.
- Vicient, C.M., Schulman, A.H., 2002. Copia-like retrotransposons in the rice genome: few and assorted. *Genome Lett.* 1, 35–47.
- Vicient, C.M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E., Schulman, A.H., 1999. Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell* 11, 1769–1784.
- Vignols, F., Rigau, J., Torres, M.A., Capellades, M., Puigdomenech, P., 1995. The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. *Plant Cell* 7, 407–416.

- Wang, L., Heinlein, M., Kunze, R., 1996. Methylation pattern of Activator transposase binding sites in maize endosperm. *Plant Cell* 8, 747–758.
- Wendel, J.F., 2000. Genome evolution in polyploids. *Plant. Mol. Biol.* 42, 225–249.
- Wessler, S.R., 1998. Transposable elements associated with normal plant genes. *Physiol. Plant.* 103, 581–586.
- Wessler, S.R., Bureau, T.E., White, S.E., 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* 5, 814–821.
- White, S.E., Habera, L.F., Wessler, S.R., 1994. Retrotransposons in the flanking regions of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. *Proc. Natl. Acad. Sci. USA* 91, 11792–11796.
- Witte, C.P., Le, Q.H., Bureau, T., Kumar, A., 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc. Natl. Acad. Sci. USA* 98, 13778–13783.
- Wright, D.A., Voytas, D.F., 2002. Athila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses. *Genome Res.* 12, 122–131.
- Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9, 3353–3362.
- Yang, G., Dong, J., Chandrasekharan, M.B., Hall, T.C., 2001. Kiddo, a new transposable element family closely associated with rice genes. *Mol. Genet. Genomics* 266, 417–424.
- Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., Machida, Y., 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. *Proc. Natl. Acad. Sci. USA* 90, 6562–6566.
- Yu, Z., Wright, S.I., Bureau, T.E., 2000. Mutator-like elements in Arabidopsis thaliana. Structure, diversity and evolution. *Genetics* 156, 2019–2031.
- Zhang, Q., Arbuckle, J., Wessler, S.R., 2000. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. *Proc. Natl. Acad. Sci. USA* 97, 1160–1165.
- Zhang, X., Feschotte, C., Zhang, Q., Jiang, N., Eggleston, W.B., Wessler, S.R., 2001. P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc. Natl. Acad. Sci. USA* 98, 12572–12577.
- Zhong, C.X., Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J., Dawe, R.K., 2002. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 14, 2825–2836.