

Nueva estrategia para la construcción de una genoteca genómica de *Pinus pinaster* en cromosomas artificiales de bacterias

R. Bautista, D. P. Villalobos, S. Díaz-Moreno, F. R. Cantón, F. M. Cánovas y M. Gonzalo Claros*

Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias. Universidad de Málaga. 29071 – Málaga.

Resumen

Para poder conocer la estructura y organización de los genomas vegetales es necesario la clonación molecular de grandes fragmentos de secuencias genómicas para construir genotecas representativas. La construcción de genotecas genómicas en cromosomas artificiales de bacterias (BAC) es una de las herramientas más utilizadas con este fin y la estrategia elegida en este trabajo. Pero cuando se trabaja con especies que presentan genomas muy grandes, como el pino, realizar este tipo de genotecas es muy laborioso y costoso. En este estudio se describe un método para, a partir de cotiledones de *P. pinaster*, construir genotecas BAC en grupos de células, lo que disminuye drásticamente el coste, el espacio y el tiempo requerido. La genoteca BAC consta de 83 grupos de células con una media de 4000 clones por grupo y representa por ahora un 0,8 X del genoma de *P. pinaster*. También se demuestra que se puede realizar por PCR (reacción en cadena de la ADN-polimerasa) con rapidez la identificación del grupo que contiene un BAC de interés. Se ha identificado un grupo de células que contiene un clon BAC que porta una secuencia codificante similar a una quitinasa de clase I de *Picea abies*, y otro clon BAC en otro grupo que contiene secuencia homóloga a un ARNm desconocido de *P. taeda*. Esta estrategia permitirá rastrear y almacenar genotecas de organismos con grandes genomas y localizar en ella genes de interés con rapidez, describiéndose el método en detalle.

Palabras clave: Fd-GOGAT, genoma, genoteca genómica, BAC, *Pinus pinaster*.

Abstract

New strategy for *Pinus pinaster* genomic library construction in bacterial artificial chromosomes

To learn about the structure and organization of plant genomes, molecular cloning of large genomic fragments for representative gene libraries construction is required. The genome library construction on bacterial artificial chromosome (BAC) is one of the most frequently used tools for this purpose, and the chosen strategy in this work. But when working with species containing very large genomes, such as pine, construction of this kind of gene libraries is a laborious and expensive task. In this study we have developed a method that starts from *P. pinaster* cotyledons to construct a pooled BAC library that drastically reduces cost, space and time required to clone a genotype of interest. The current BAC library comprises 83 groups of cells with an average of 4000 clones per group and accounts for a 0.8 X genome content of *P. pinaster*. It is also shown that a BAC of interest can be quickly identified by PCR (polymerase chain reaction): a pool containing a BAC clone that carries a coding sequence similar to a Class I chitinase of *Picea abies* has been identified, as well as another BAC clone in a different pool containing a sequence with homology to an unknown *P. taeda* mRNA. This strategy will allow screening and storing gene libraries from large genome organisms and a quickly location of genes of interest, the method been described in detail.

Keywords: Fd-GOGAT, genome, genomic library, BAC, *Pinus pinaster*.

* Corresponding author: claros@uma.es

Received: 04-06-08. Accepted: 07-10-08.

1. Introducción

Las coníferas tienen una gran importancia ecológica y económica, dominan grandes ecosistemas, juegan un papel esencial en la fijación de carbono en nuestro planeta e incluyen especies muy utilizadas para la reforestación. También son una importante fuente de madera para la obtención de pulpa de papel y para la fabricación de muebles. Una de las coníferas más importantes en la península Ibérica es el pino resinero (*Pinus pinaster* Ait.), especie que se utilizará como modelo experimental en este trabajo. Se están realizando numerosos estudios de genómica funcional con pino como modelo, que van desde la producción de colecciones de EST (*expressed sequence tag*) de tejidos sometidos a distintos estímulos (Cantón *et al.*, 2004; Frigerio *et al.*, 2004; Lorenz *et al.*, 2006; Pavy *et al.*, 2005), hasta el análisis transcripcional utilizando micromatrices (*microarray*) (Joosen *et al.*, 2006; Ralph *et al.*, 2007; Ralph *et al.*, 2006; Yang y Loopstra, 2005; Zelena y Sorochnyn's'kyi, 2005), pasando por la proteómica (Costa *et al.*, 1999; Fernando, 2005; Lippert *et al.*, 2007). Pero se conoce poco de la genómica estructural del género *Pinus* y, en general, de las coníferas.

De la estructura del genoma de pino se sabe que se trata de un organismo diploide caracterizado por un genoma grande y complejo, lo que dificulta su análisis. El contenido de ADN nuclear es muy variable (Grotkopp *et al.*, 2004), aunque todas las especies poseen 12 cromosomas de morfología similar (Doudrick *et al.*, 1995), y va desde los 43,96 pg (*P. banksiana*) a los 75,36 pg (*P. gerardiana*) [1 pg = 960 kb, (Arumuganathan y Earle, 1991)]. El 75 % del genoma corresponde a secuencias altamente repetidas, mientras que hay un 25 % que corresponde a secuencias poco repetidas o en copia única. Los primeros análisis han revelado la existencia de un gran número de repeticiones, que incluyen retrotransposones (Heslop-Harrison *et al.*, 1997; Kamm *et al.*, 1996; Kossack y Kinlaw, 1999; Rocheta *et al.*, 2007; Stuart-Rogers y Flavell, 2001), repeticiones de secuencias simples (Schmidt *et al.*, 2000) y complejas familias de genes que podrían ser el resultado de duplicaciones (Kinlaw y Neale, 1997).

Para conocer la estructura del genoma de pino sería muy útil disponer de bancos de secuencias genómicas de gran tamaño. En los últimos años, la inmensa mayoría de las genotecas genómicas vegetales se han construido en cromosomas artificiales bacterianos (BAC: *bacterial artificial chromosome*) (Shizuya *et al.*, 1992). Con ellas se han generado contigs y se han aislado genes completos (Allouis *et al.*, 2003; Kim *et al.*, 1997;

Luo *et al.*, 2001; Song *et al.*, 2000; Vanhouten y MacKenzie, 1999; Zhang *et al.*, 1995), y también está siendo una herramienta básica para organizar los proyectos de secuenciación de genomas completos de maíz (<http://www.maize.org>), trigo (<http://www.wheatgenome.org>) y arroz (<http://rgp.dna.affrc.go.jp/IRGSP/>). Una de las principales características de los BAC es que sólo hay una copia por célula. Esta propiedad es fundamental para la estabilidad de los insertos grandes, sobre todo si derivan de secuencias repetidas que pueden sufrir procesos de recombinación. Al emplearse ADN genómico intacto para su construcción, los protocolos que se siguen para la generación de una genoteca en BAC son completamente diferentes a los que se aplican para las genotecas convencionales de insertos más pequeños, puesto que el ADN tiende a degradarse con facilidad y es muy fácil que se contamine con ADN cloroplastídico (Bautista *et al.*, 2007). En nuestro laboratorio se ha optimizado un método para preparar este ADN de alto peso molecular de pino y clonarlo en un vector de tipo BAC (Bautista *et al.*, 2007). Aunque los clones se conservaban aislados en placas de 96 o 384 pocillos, no es la tecnología ideal cuando se trata de organismos con genomas grandes y complejos, pues requiere un número extremadamente grande de clones, y resulta un problema cuando el espacio en el laboratorio es escaso. Por ejemplo, los genomas grandes como el de trigo (17 000 Mb) o el de pino (23 000 Mb) generarán un mínimo de 850 000 y 3 750 000 clones, respectivamente, para una genoteca 5X. Estas cifras son mucho mayores que los 7000 clones necesarios para una genoteca de similares características de *Arabidopsis thaliana*. Es más, almacenar las genotecas anteriores en placas de 384 pocillos implica generar entre 2 200 y 9 800 placas, que ocuparían el interior de 2 a 8 congeladores a -80 °C convencionales. Evidentemente, estos números hacen inviable en la práctica la construcción y mantenimiento de tales genotecas. En el trigo se ha conseguido reducir el tamaño de la genoteca al clonar sólo el cromosoma de interés (Janda *et al.*, 2004; Janda *et al.*, 2006), en el maíz se han clonado sólo las regiones de copia simple del genoma (Yuan *et al.*, 2003) o sólo los fragmentos sin metilar (maíz y trigo) porque contienen los genes que se están expresando (Lamoureux *et al.*, 2005; Rabinowicz, 2003). En nuestro laboratorio se ha comprobado que los métodos anteriores no son aplicables de forma sistemática a las coníferas (Bautista, 2008).

En este artículo se describe la posibilidad de elaborar una genoteca de *P. pinaster* en BAC utilizando una estrategia que ya se ha aplicado con éxito en otras espe-

cies (Isidore *et al.*, 2005; Ma *et al.*, 2000; Salimath y Bhattacharyya, 1999). La idea central consiste en mantener la genoteca en forma de grupos de células, de tal forma que se pueda disminuir el coste, el tiempo y el espacio requerido para su construcción y almacenamiento. Se ha comprobado que cada grupo puede contener más clones de lo que se había descrito, y que las mezclas de ligación se pueden congelar sin una pérdida significativa de eficacia. También se demuestra que se pueden localizar con rapidez clones de interés en los grupos de células por PCR y se demuestra la utilidad de incluir un clon de control en los grupos de células.

2. Materiales y Métodos

2.1. Preparación del vector de clonación pIndigoBAC536

La calidad del vector BAC es clave porque está directamente relacionada con la eficacia de la transformación. Se utilizó el vector pCUGIBAC1 (proporcionado por el profesor David Frisch de la Clemson University Genomics Institute, EE.UU) que contiene el vector pIndigoBAC536 clonado a su vez en el vector pGEM-4z (Luo *et al.*, 2001; Vanhouten y MacKenzie, 1999) para facilitar la preparación del vector. El pIndigoBac536 se aísla de la siguiente forma: se digieren 10 µg de ADN del vector con 50 U de *Bam*HI (Roche) a 37 °C durante 16 h, en un volumen final de 200 µl. Se purifica el ADN del vector con fenol:cloroformo (1:1). Se precipita el ADN del vector añadiendo 1/10 volumen de acetato sódico a 3 M, pH 5,2 más 2 volúmenes de etanol absoluto, se incuba 30 minutos a -80 °C y se centrifuga a 13 000 rpm a 4 °C durante 20 minutos. A continuación, se lava el precipitado con etanol al 70 %, se seca al aire y se resuspende en 50 µl de agua desionizada estéril. Después, se desfosforilan 10 µg de ADN del vector con 0,4 U de CIP (Roche) por µg de ADN y se incuba 30 minutos a 37 °C, parando la reacción con EDTA a 0,05 M pH 8. El producto final se extrae con fenol:cloroformo (1:1) y se precipita con etanol absoluto como antes. Luego se incuban los 10 µg de ADN del vector con 400 U de la ADN-ligasa de T4 (NEB) en un volumen final de 40 µl durante 2 h a temperatura ambiente para eliminar las moléculas sin desfosforilar. Los distintos tipos (pIndigoBAC536 y pGEM4-z) y formas (lineal y circular) de vectores se separan en un gel de agarosa al 0,8 % en tampón TAE 1X (tris-acetato a 40 mM, Na₂EDTA a 1 mM). Se extrae del gel de agarosa el ADN lineal del

pIndigoBAC536 (7,5 kb) mediante columnas de purificación Qiaquick (Qiagen), siguiendo las instrucciones del fabricante para fragmentos de gran tamaño. Finalmente, se ajusta la concentración final del vector a 10 ng/µl, se reparte en alícuotas de 50 µl, y se almacenan en glicerol al 50 % a -20 °C (para un uso frecuente) y a -80 °C (para un uso a más largo plazo).

2.2. Construcción del clon BAC de control

Se ha construido un clon BAC que sirve de control positivo para el rastreo de la genoteca (Figura 1) con un fragmento genómico de *P. pinaster* que contiene la secuencia de un pseudogén de la glutamato-sintasa dependiente de ferredoxina (Fd-GOGAT) previamente clonado en el laboratorio (Bautista *et al.*, 2007) al que se unió un fragmento del gen delator *GUS* de *E. coli*. De esta forma, se puede identificar por PCR el control dentro de un grupo de clones BAC utilizando los cebadores GOGAT1/GOGAT2, pero más específicamente, los GUS1/T7 descritos en la Tabla 1.

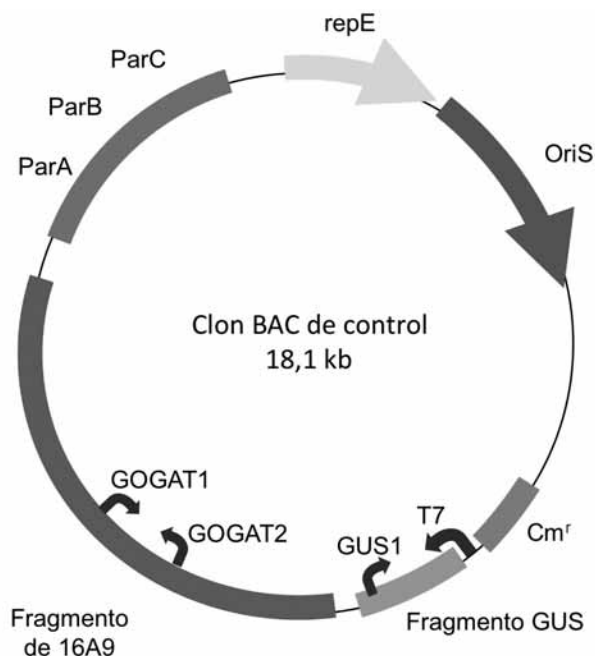


Figura 1. Estructura del clon BAC de control de 18,1 kb. Contiene un fragmento genómico de 9 kb procedente del clon BAC 16A9 clonado en el sitio *Bam*HI del vector pIndigoBAC536 y un fragmento de 1,7 kb del gen *GUS* en el sitio *Eco*RI. Se indican con flechas azules la posición de los cebadores y su nombre utilizados para rastrear este ADN por PCR.

Tabla 1. Características de los cebadores utilizados en este estudio

<i>Cebador</i>	<i>Características</i>	<i>T^a hibridación</i>	<i>Tamaño amplificado</i>	<i>Secuencia</i>
GS1a prom1	Secuencia promotora	56 °C	428 pb	5'-TCAAATGATCACCTCGTCA-3'
GS1a prom2	de GS1a	57 °C		5'-GTAACACGCATTCCCATT-3'
Gs1b prom1	Secuencia promotora	57 °C	333 pb	5'-ACTGTCGGATGTGCATGGTA-3'
Gs1b prom2	de GS1b	55 °C		5'-GGCATGCTATGCTTTGGAAT-3'
GOT3'-A	Fragmento exón-	52 °C	456 pb	5'-CTGAAAGAATGGGAAAAATA-3'
GOT3'-B	zona 3' UTR de GOGAT	60 °C		5'-TGACTAGAGACCCAAGAGCA-3'
T7	Fragmento Gus-	48 °C	1700 pb	5'-TAATACGACTCACTATAGGG-3'
GUS1	pIndigoBAC536	45 °C		5'-TGGATCGCGAAAACGTGT-3'
GOGAT1	Fragmento clon	52 °C	200 pb	5'-TGATATCACCTCCTCCACATC-3'
GOGAT2	BAC 16A9	58 °C		5'-TGCTCCAGTGCCCCCATCATG-3'
ELP 5'	EST homóloga a	60 °C	440 pb	5'-GGAATGCAAGGGGGTTTATTGC-3'
ELP 3'	ELP	55 °C		5'-CAAGTTTCAGTGCTTCACC-3'
LHT 5'	EST homóloga a	57 °C	460 pb	5'-TGGTGCAGTTCTGGGATTGA-3'
LHT 3'	LHT	60 °C		5'-CACAGGATTGGAATCTCGCAT-3'
GLP-3' seq	EST homóloga a	51 °C	800 pb	5'-AGCAGCAACCAATTCTCA-3'
FL GLP-3'	GLP	55 °C		5'-TTGCTTCAAGTTCCAAAGCC-3'

Los genes GS1a, GS1b y Fd-GOGAT, y el clon 16A9 mencionados son de *P. pinaster*. Las EST para las que se han diseñado los cebadores son también de *P. pinaster*. ELP es una quitinasa de clase I, LHT es un transportador de lisina y GLP es una proteína tipo germinina

2.3. Preparación de ADN de alto peso molecular a partir de cotiledones de *P. pinaster*

Los reactivos específicos para este método son: el tampón de homogeneización HB a 10X (trizma base a 0,1 M, KCl a 0,8 M, Na₂EDTA a 0,1 M, espermidina a 40 mM y espermina a 10 mM; la espermidina y la espermina se deben añadir después de esterilizar en autoclave; el pH se ajusta a pH 9,5 con NaOH) que se guarda a 4 °C, el tampón de lavado (HB 1X, sacarosa a 0,5 M, Triton X-100 al 0,5 %, PVP-40 000 (polivinilpirrolidona) al 2%, ácido ascórbico al 0,1 % y 2-mercaptoetanol al 0,15 %) que se prepara justo antes de usar y se esteriliza por filtración, la solución SCE (sorbitol a 1 M, citrato de sodio a 100 mM y Na₂EDTA a 60 mM; el pH se ajusta a 7 con KOH), y el tampón de lisis EPS (Na₂EDTA a 0,5 M, lau-

ril-sarcosina de sodio a 1 % y proteinasa K a 0,1 mg/ ml, que se añade justo antes de usar).

El procedimiento realizado para preparar los núcleos intactos comienza embebiendo las semillas de *P. pinaster* en agua desionizada durante 12 horas en condiciones de aireación continua. Luego se germinan en vermiculita (Europerlite S.A) a una temperatura constante de 22 °C con 16 h de luz durante 9 a 12 días. Las plántulas obtenidas se pueden guardar a -80 °C si no se van a usar inmediatamente. Se toman 20 g de los cotiledones almacenados a -80 °C y se muelen con nitrógeno líquido en un mortero. El tejido disgregado se transfiere a un vaso de precipitado de 500 ml frío que contiene 200 ml de HB 1X (tampón de lavado). Esto debe realizarse con cuidado porque el tampón con sacarosa en ausencia de cationes divalentes cambia las condiciones osmóticas y hace que los núcleos sean extremadamente frágiles. Se reco-

mienda emplear un pincel grueso n.º 12 para ayudar a romper los grumos que se forman. El tampón tiene además un pH elevado para inhibir la acción de las endonucleasas internas. A continuación, se homogeneiza con un imán en un agitador magnético dando 6 pulsos de 20 segundos cada 2 minutos. La solución se filtra a través de 2 capas de gasa y 2 capas de Miracloth (Calbiochem). Se transfiere el filtrado a tubos de centrifugación de 250 ml fríos y se centrifugan en un rotor angular fijo a 1200 x g durante 20 minutos a 4 °C. Se decanta el sobrenadante y se resuspende el sedimento (que contiene los núcleos) en 2 ml de tampón de lavado con ayuda de un pincel fino (n.º 2). Se añaden 28 ml más de tampón de lavado. Se filtra la solución a través de una capa de Miracloth para eliminar los restos de tejido y se transfiere el filtrado a tubos de centrifuga de fondo redondo fríos, centrifugándose a 1000 x g durante 15 minutos a 4 °C en un rotor angular flotante. A continuación se decanta el sobrenadante y se resuspende el sedimento que contiene los núcleos en 2 ml de tampón de lavado con ayuda de un pincel fino n.º 2. Se añaden otros 28 ml más de tampón de lavado y se centrifuga a 1000 x g durante 15 minutos a 4 °C. Lávese dos o tres veces más en función del estado de pureza de la muestra. En el último lavado, se resuspende la muestra en 500 µl de SCE frío con ayuda del pincel fino n.º 2. Finalmente, se verifica la integridad de los núcleos tiñendo 100 µl de la suspensión nuclear con 20 µl de una disolución de DAPI (4'-6-dimidino-2-fenilindol) a 20 µg/ml en tampón SCE. E incubando la mezcla de 1 a 2 minutos a 4 °C en oscuridad. Se emplea 10 µl de suspensión nuclear teñida para su observación al microscopio de epifluorescencia.

El procedimiento para embeber los núcleos en agarosa de bajo punto de fusión comienza fundiendo el stock de agarosa de bajo punto de fusión (Seaplaque, FMC Bioproducts, EE.UU) al 1,6 % en SCE y manteniéndolo a 45 °C. Hay que atemperar la muestra de núcleos 5 minutos a 45 °C y luego añadir 1 volumen de agarosa de bajo punto de fusión a 1 volumen de muestra de núcleos, manteniendo la mezcla a 45 °C. La mezcla se transfiere a los moldes utilizando una pipeta con una punta con el extremo cortado. Hay que dejar que se solidifiquen los bloques a 4 °C unos 30 minutos y luego se guardan en TE 1X a 4 °C hasta su uso.

La preparación del ADNapm se basa en la lisis de los núcleos contenidos en los bloques de agarosa. Para ello se incuban los bloques con 10 volúmenes de EPS por bloque suplementado con PVP-40 000 (polivinilpirrolidona) al 2% y ácido ascórbico al 0,1 %. Después se añade proteinasa K a una concentración final de 0,1

mg/ml y se incuba a 50 °C sin agitación al menos 36 horas, refrescando el tampón de incubación y la enzima cada 12 horas. Los bloques se lavan con 10 volúmenes por bloque de una solución de EDTA a 0,5 M a pH 8 y se incuban 1 hora a 50 °C. A continuación se lavan los bloques con 10 volúmenes por bloque de la solución de EDTA a 0,5 M a pH 8 frío y se incuban 1 hora en hielo. Hay que repetir esta operación hasta que el tampón esté completamente transparente. Finalmente, se almacenan los bloques en EDTA a 0,5 M, pH 8, a 4 °C hasta que llegue el momento de usarlos. Este ADNapm se puede mantener intacto en EDTA a 0,5 M hasta 5 años (Bautista *et al.*, 2007). Alternativamente, los bloques se pueden almacenar en tubos de polipropileno de 50 ml con etanol al 70% que se han mantenido a temperatura ambiente al menos 4 horas y luego se han guardado a -20 °C. El ADN en etanol no sufre degradación en los 6 primeros meses de almacenamiento. Cuando se vaya a utilizar este ADN, es necesario eliminar los restos de etanol que hayan podido difundir dentro del bloque; para ello se incuban a 4 °C en TE 1X durante varias horas y se cambia el tampón varias veces a lo largo de la incubación.

En caso de que el bloque de agarosa adquiriera una tonalidad marrón, se recomienda no utilizar porque probablemente no funcionen los tratamientos del ADN necesarios para construir la genoteca.

2.4. Digestión parcial del ADNapm

Durante la extracción del ADNapm siempre existe una fragmentación inespecífica por cizalla o por acción de endonucleasas inespecíficas. Estos fragmentos inhibirán la posterior ligación, por lo que conviene retirarlos antes de realizar otras manipulaciones. Para ello, se migra el ADN de los bloques durante 12 horas a 4 V/cm con un pulso inicial de 2,5 segundos y uno final de 5,5 segundos en un gel de agarosa al 0,7% en tampón TBE 0,33X. A continuación, el ADN intacto contenido en el bloque se digiere parcialmente y luego se lavan en TE 1X a 4 °C durante 1 h; cada 30 minutos se cambia el tampón. Después, se coloca medio bloque de agarosa por tubo y se añade 1 ml de tampón de incubación (tampón de la enzima de restricción más espermidina a 4 mM, DTT a 1 mM, Lys-HCl a 2,5 mM y EGTA a 60 mM). Deben utilizarse EGTA y Lys-HCl, que son inhibidores de nucleasas, para evitar la degradación del ADN durante la digestión (Bautista *et al.*, 2007; Liu y Wu, 1999; Peterson *et al.*, 2000). Tras incuban 1 h a 4 °C, se retira el tampón de incubación y se añaden 170 µl de tampón de

digestión (tampón de digestión de la enzima más 0,75 U de *Bam*HI, espermidina a 4 mM, DTT a 1 mM, BSA a 0,5 mg/ml, Lys-HCl a 2,5 mM y EGTA a 60 mM) a cada medio bloque. Se incuba 1 h a 4 °C y luego 3 minutos a 37 °C, y se para la reacción con 20 µl de EDTA a 0,5 M pH 8. Ahora se depositan los medios bloques en un gel de agarosa al 0,7 %, en tampón TBE 0,33X, con la ayuda de una espátula y se migran en FIGE a 4 V/cm, con un pulso inicial de 2,5 segundos y uno final de 5,5 segundos, durante 20 h a 11 °C. Al acabar la electroforesis, se escinde la parte del gel utilizada como control de digestión, se tiñe con bromuro de etidio a 0,5 µg/ml y se fotografía esta parte del gel junto con una regla como referencia para ubicar la región del gel que contiene el ADNpm de entre 100 y 240 kb. Dicha región se escinde cuidadosamente y se introduce en una bolsa de diálisis junto con 300 µl de TBE 0,33X, asegurándose de que no queden burbujas cuando se cierran los extremos de la bolsa con las pinzas de diálisis. Se electroeluye durante 2 horas y 30 minutos utilizando los mismos parámetros con los que se realizó la migración. Después se invierte el sentido de la migración durante 1 minuto para despejar el ADN de las paredes. Se retiran los fragmentos de agarosa y se recoge el tampón que contiene los fragmentos de ADN con una punta con el extremo cortado. Finalmente, se dializa y concentra el ADN sobre un filtro de 0,0025 µm VSWP (Millipore) que flota en una placa de Petri con TE 0,5X y PEG al 10 % durante 90 minutos a temperatura ambiente. La concentración del ADN se cuantifica por la absorbancia a 260 nm

2.5. Construcción de la genoteca en grupos de clones

Se incuba una proporción de inserto:vector de 10:1 (habitualmente 40 ng de vector pIndigoBAC536 desfosforilado y 400 ng de ADNpm) con 400 U de ADN-ligasa de T4 en un volumen final de 150 µl durante una noche a 16 °C. A continuación, esta mezcla de ligación se concentra y dializa como se ha descrito para el ADN digerido en el apartado anterior, lo que reduce el volumen de ligación a una media de 60 µl. Se descongelan en hielo las células Electromax DH10B (Invitrogen) y se añaden de 2 a 4 µl de la mezcla de ligación desalada en un volumen de 20 µl de suspensión de células en una cubeta de electroporación de 0,1 cm (el resto de la ligación se congela en alícuotas de 10 µl). Se electropora a 1800 V, 10 µF de capacitancia y un máximo de resistencia de 600 Ω en el Micropulser Electroporator de Bio-

Rad. Las células se resuspenden en 1000 µl de SOC y se incuban 1 hora a 37 °C con agitación lenta. Se diluye el cultivo para sembrar entre 3000 y 5000 clones distintos (Figura 2) en placas de 254 mm x 254 mm (Nunc, Dinamarca) con LB (triptona al 1%, extracto de levadura al 0,5%, NaCl al 1% y agar al 1,5 %, pH 7,5) que contiene X-gal a 40 µg/ml, IPTG a 100 µg/ml y cloranfenicol a 12,5 µg/ml. Además, se deposita en un extremo de la placa una colonia del clon BAC de control. Se resuspenden todos los clones que crecieron en la placa en 3 ml de tampón de almacenamiento (LB suplementado con KCl a 10 mM, MgSO₄/MgCl₂ a 20 mM, glucosa a 40 mM y glicerol al 25 %) y se incuban las células resuspendidas 1 hora a 37 °C con agitación (220 rpm). El volumen de células obtenido se reparte en 3 alícuotas distintas, de las que una se utiliza para propagar los clones y extraer el ADN directamente, y las otras dos se almacenan a -80°C (Figura 2).

2.6. Rastreo de la genoteca por PCR

Los grupos de la genoteca se pueden rastrear por PCR utilizando como molde el ADN de los BAC de cada grupo y un par de cebadores que cubran una región del gen de interés (Tabla 1). El rastreo se comienza inoculando 100 µl de cada grupo de células en 4 ml de LB con cloranfenicol a 12,5 µg/ml y después se incuban al menos 6 h a 37 °C en agitación (220 rpm). Después se extrae el ADN de los grupos utilizando columnas de purificación (Omega Bio-Tek). También se puede utilizar el protocolo descrito (Villalobos *et al.*, 2004) cuando son pocos grupos. Se ordenan todas las extracciones en placas de 96 pocillos (Figura 2) y se diluyen 1/10 con TE 1X. Cada reacción de PCR se prepara con 5 µl del ADN de cada grupo diluido 1/10, 0,025 U de BioTaq (BioLine), tampón de PCR 1X (BioLine), los cuatro dNTP a 200 mM, cebadores a 0,4 mM y agua destilada desionizada hasta completar los 50 µl finales. La amplificación se realiza sobre un termociclador (Biometra) a la temperatura de hibridación y tiempo de extensión final que aparecen en la Tabla 1.

3. Resultados y discusión

3.1. Construcción de la genoteca de pino

A diferencia de lo que ocurre cuando se aplican otros métodos estándar (Peterson *et al.*, 2000; Zhang *et al.*,

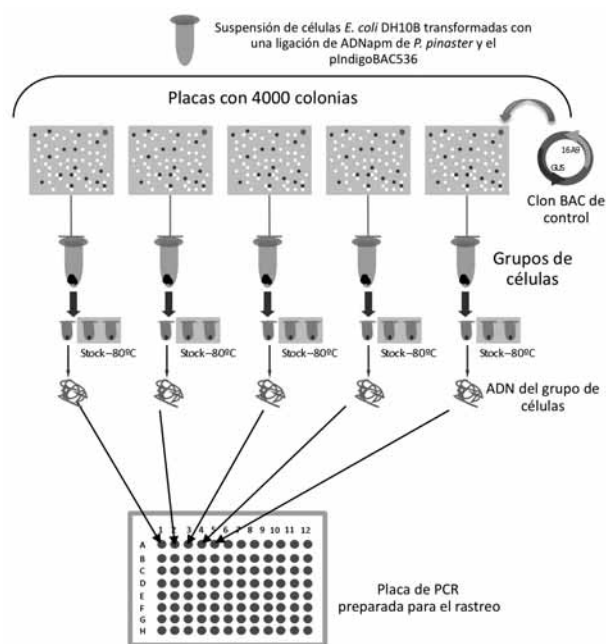


Figura 2. Esquema de la preparación de la genoteca por grupos a partir de células de *E. coli* DH10B transformadas con ADN de *P. pinaster* clonado en el vector pIndigoBAC536. Véanse los detalles en el texto.

1995), los núcleos preparados según se describe en este artículo no se oxidan sino que permanecen blancos (resultados no mostrados), esto es, aptos para cualquier modificación enzimática posterior. El ADN de estos núcleos se digirió parcialmente con *Bam*HI en presencia de inhibidores de endonucleasas, se ligó al vector pIndigoBAC536, y se transformó *E. coli* DH10B con 4 µl de la ligación, bien recién preparada o bien recién descongelada (véase el apartado 3.2). La congelación de la ligación permitió construir la genoteca en distintas jornadas a partir de la misma mezcla de ligación. Los mejores resultados se obtuvieron cuando se realizaron transformaciones con 20 µl de suspensión celular competente de *E. coli* DH10B y 4 µl de ligación sin congelar (Tabla 2).

La siembra de las células transformadas en placas de 245 mm x 245 mm generó de 3000 a 5000 clones (media: 4000 clones), que se recogieron en un único tubo para formar un grupo (Figura 2). Los grupos se mantuvieron congelados en presencia de glicerol (25%) y glucosa (40 mM) para minimizar la pérdida de clones durante el almacenamiento a -80 °C y la posterior propagación en medio líquido, como se sugirió en (Isidore *et al.*, 2005; Ma *et al.*, 2000). Además, cada grupo siempre incluyó una colonia del clon BAC de control añadi-

da con posterioridad. La genoteca consta de 83 grupos de células (21/04/2008) con 332 000 clones, lo que equivale a una representación 0,8X al asumirse un tamaño medio de inserto de 107 kb, que es lo publicado para este método (Bautista *et al.*, 2007).

3.2. Efecto de la congelación sobre la mezcla de ligación

Debido a que una genoteca 3X de pino en las condiciones descritas necesitaría 2 100 000 clones, y teniendo en cuenta que el número de clones que se generan en una transformación tiene que ser manejable, se hace imprescindible almacenar las mezclas de ligación, si bien los mejores resultados se obtuvieron con las mezclas de ligación recién preparadas (Tabla 2). Ma y colaboradores (Ma *et al.*, 2000) describen una pérdida máxima del 10 % de la eficacia de transformación cuando se someten las mezclas de ligación a un ciclo de congelación con nitrógeno líquido y posterior almacenamiento a -80 °C. Estos resultados se han confirmado con ADN de pino (Tabla 2) y, además, se ha observado que la eficacia de transformación disminuye drásticamente con un segundo ciclo de congelación (dato no mostrado). Por tanto, se puede realizar la ligación de la genoteca en un procedimiento puntual y luego guardar la ligación a -80 °C para ir transformando las células competentes y completar la genoteca en poco tiempo.

3.3. Propagación de los grupos de la genoteca

Puesto que el análisis de los grupos de la genoteca implica su amplificación, había que determinar si, dentro del mismo grupo de células, los clones pueden competir entre sí durante la propagación, es decir, si unos clones se pierden mientras que otros se hacen demasiado abundantes. Los dos principales factores que afectan

Tabla 2. Datos sobre eficacia de la ligación

Volumen de células	25 µl	20 µl
Volumen de ligación	2,5 µl	4 µl
Relación células/ligación	10:1	5:1
Fresca	1,47 x 10 ⁵ ufc/µg	
1ª Congelación	1,31 x 10 ⁵ ufc/µg	2,29 x 10 ⁵ ufc/µg

a la competición son el número de clones por grupo y el tiempo de propagación. Respecto a lo segundo, se sabe que la representación de clones en cultivos de grupos de células propagados durante una noche (menos de 16 h) es independiente del tamaño del inserto (Suzuki *et al.*, 2002), lo que indica que la pérdida de clones, de producirse, sería aleatoria. Además, la posibilidad de rastrear la presencia del clon de control (véase el apartado 3.4) revelaría un posible sesgo de la población clonal.

3.4. Optimización del procedimiento de detección por PCR utilizando el clon BAC de control

Dadas las características del proceso, se necesita tener la seguridad de que es posible identificar, por PCR, un clon concreto dentro del conjunto de ADN que se obtiene al extraer un grupo de clones BAC mediante lisis alcalina. Para ello ha sido necesario construir un clon BAC de control (Figura 1) que se propagara con los clones del grupo para permitir su rastreo por PCR, ya que los grupos utilizados en los trabajos en los que se basa este método nunca contenían unos 3000 clones de media, mientras que en nuestro caso la media era de 4000, por lo que nuestros posibles positivos estarían incluidos en una población celular mayor de lo descrito en la bibliografía. Con los cebadores GUS1 y T7 (Tabla 1) se puede amplificar una secuencia quimérica que sólo está presente el clon BAC de control que también contiene una secuencia de pino para simular la estructura de los clones BAC de la genoteca. A su vez, el ADN de pino que contiene el BAC de control podría utilizarse también para rastrear, pero dado que se trata de la misma especie, se verá que son frecuentes las amplificaciones inespecíficas de igual forma que amplificar sólo el gen *GUS* provocaría también la amplificación de dicho gen a partir del ADN genómico de *E. coli* que contamina las preparaciones plasmídicas), por lo que la mejor opción será localizar dicho clon con los cebadores GUS1 y T7.

Con la pareja de cebadores GUS1/T7 se ha podido identificar el clon BAC de control en un grupo de células que contiene más de 5000 clones distintos, del que se pusieron a propagar 100 µl en un volumen de LB de 4 ml (Figura 3). Los análisis de PCR mostraron que GUS1 y T7 amplificaban la misma banda de ADN cuando se utilizaba como molde el ADN del clon BAC de control, así como el grupo de la genoteca que lo con-

tiene (grupo+), mientras que no aparece ni cuando la genoteca no contiene el clon BAC de control ni cuando no hay ADN molde (Figura 3). También se comprobó que, utilizando diluciones 1/10 y 1/50 de la preparación del ADN de los grupos, se obtenían resultados repetitivos y específicos con la dilución 1/10, en la que la banda específica era claramente más intensa que otras amplificaciones inespecíficas que aparecieron, comportamiento que no era tan obvio con la dilución 1/50. Por tanto, se puede concluir que se puede detectar con claridad un clon BAC en un grupo utilizando una dilución 1/10. La presencia de amplificaciones inespecíficas se podría deber a que la temperatura de hibridación de la pareja GUS1/T7 es de tan solo 50 °C, a pesar de estar por encima de las temperaturas de fusión (45°C y 48°C, respectivamente, Tabla 1). De hecho, con temperaturas iguales inferiores a 50-55°C se producían hibridaciones inespecíficas con los grupos de la genoteca, probablemente debidas al ADN contaminante de *E. coli* o a las secuencias repetidas o pseudogenes del pino (resultados no mostrados). Por tanto, está claro que, si se consigue identificar el clon de control en un grupo, se podrá identificar cualquier otro clon que exista en él.

Puesto que el amplificado con GUS1 y T7 sólo podía aparecer cuando el grupo de células contenga el BAC de control, había que confirmar los resultados amplifican-

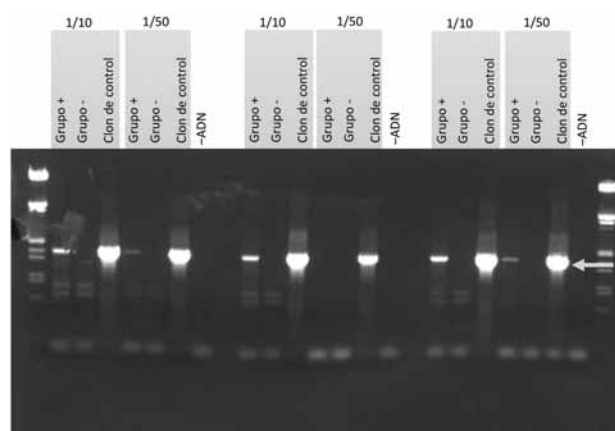


Figura 3. Detección del clon de control por PCR con los cebadores GUS1 y T7 que amplifican un fragmento de 1,7 kb. La flecha señala la amplificación del tamaño esperado. 'Grupo-': preparación de ADN del grupo de células que no contiene el clon de control. 'Grupo+': preparación de ADN del lote de células que contiene el clon de control. 'Clon de control': preparación de ADN que sólo contiene el BAC de control, '-ADN': preparación que no contiene ADN. 1/10 y 1/50 indican las diluciones realizadas sobre el ADN preparado del grupo de células.

do un ADN de pino. Para ello se utilizó la pareja de cebadores GOGAT1 y GOGAT2 (Tabla 1) que amplifican un fragmento genómico de 200 pb de *P. pinaster* que está contenido en el clon BAC de control. El producto de PCR con la pareja GOGAT1/GOGAT2 era visible (Figura 4) con el ADN del clon de control y con el ADN genómico de pino, pero no aparecía con el ADN genómico de *E. coli* ni en ausencia de ADN. En los grupos, la banda aparecía claramente cuando estaba el clon de control (grupo+), pero también se observaba, con menor intensidad, en las muestras sin el clon BAC de control (grupo-). Esto podría hacer pensar varias cosas: 1) que se están amplificando secuencias homólogas dentro del grupo de células, ya que ese fragmento deriva de un posible pseudogén de Fd-GOGAT (Bautista *et al.*, 2007); 2) que la PCR en estas condiciones conlleva algo de inespecificidad; y 3) que la secuencia del pseudogén de Fd-GOGAT es muy abundante en el genoma de pino. No se puede tampoco descartar que se amplifiquen secuencias homólogas procedentes del ADN genómico de *E. coli* contaminante porque aunque en la Figura 4 no se observa dicha amplificación, sí que aparece en la Figura 5. Puesto que la pareja T7/GUS1 identifica, de forma inequívoca, los grupos de ADN a los que se añadió el clon BAC de control, no encontrándose ningún tipo de amplificación en los ADN restantes (Figura 3), se puede

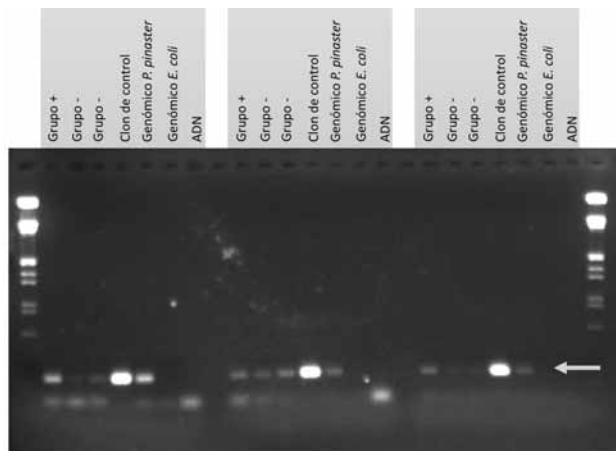


Figura 4. Detección por PCR del fragmento de 200 pb de la glutamato-sintasa dependiente de ferredoxina en tres grupos distintos de la genoteca con los cebadores GOGAT1 y GOGAT2. La interpretación de cada pocillo y la flecha es la misma que para la figura 3. Además, 'Genómico *P. pinaster*' indica que es un ADN genómico obtenido por métodos convencionales a partir de cotiliones de *P. pinaster*, y 'Genómico *E. coli*' indica que es un ADN genómico obtenido de un cultivo de dicha bacteria.

considerar que el método es suficientemente sensible para rastrear el ADN de una colonia presente en un grupo de la genoteca, aunque seguramente aparecerán bandas de amplificación inespecíficas que deberían ser menos intensas que la de amplificación específica.

3.5. Rastreo de nuevos genes en la genoteca

A pesar de que el 75 % de las secuencias del genoma de las coníferas son repetitivas (Guevara *et al.*, 2005), se puede comenzar a realizar rastreos con genotecas incompletas para determinar la validez del método. Para ello se diseñaron parejas de cebadores de los genes ELP, LHT y GLP aislados a partir de un banco de secuencias de EST de *P. pinaster* (Cantón *et al.*, 2004; Frigerio *et al.*, 2004) y a partir de las secuencias promotoras de los genes *GS1a* y *GS1b* de *P. pinaster*. Todas las parejas de cebadores de la Tabla 1 se comprobaron sobre ADN genómico total de *Pinus pinaster* (resultado no mostrado) y se verificó que el fragmento amplificado era del tamaño esperado o de mayor tamaño, posiblemente por amplificación de una zona exón-intrón. De todas las parejas de cebadores diseñadas, se desecharon aquellas que tenían una temperatura de hibridación inferior a 50 °C, para evitar las amplificaciones inespecíficas (como las comentadas en el apartado anterior). Durante el rastreo, todas las parejas empleadas (salvo ELP-3'/ELP-5' y Got3'-A/Got3'-B, Tabla 1) dieron un resultado negativo, sin duda porque la genoteca no cubre representativamente el 100 % del genoma. Con la pareja de cebadores ELP-3'/ELP-5', que dio un resultado positivo, se logró amplificar una banda de ~400 pb en el grupo 7 (Figura 5, izquierda), cuyo tamaño coincide con el que se obtiene utilizando el ADN genómico de *P. pinaster*. Por su parte, la pareja de cebadores Got3'-A/Got3'-B para la Fd-GOGAT, aunque amplificaba en la mayoría de los grupos un producto de 560 pb, mayor de lo esperado, producía una banda específica de 450 pb en el grupo n.º 71, que tampoco aparece en el control con ADN genómico de *P. pinaster*.

Para comprobar que las bandas amplificadas correspondían al ADN esperado, se secuenció el fragmento amplificado por los cebadores ELP-3'/ELP-5' y se encontró que la secuencia era un 89% homóloga a la quitinasa de clase I de *Picea abies* (AY450922), lo que indicó que se había obtenido de un clon BAC que contenía parte de la secuencia codificante de un gen homólogo a una quitinasa de clase I de *P. abies*. En cambio, la secuenciación del fragmento amplificado de 450

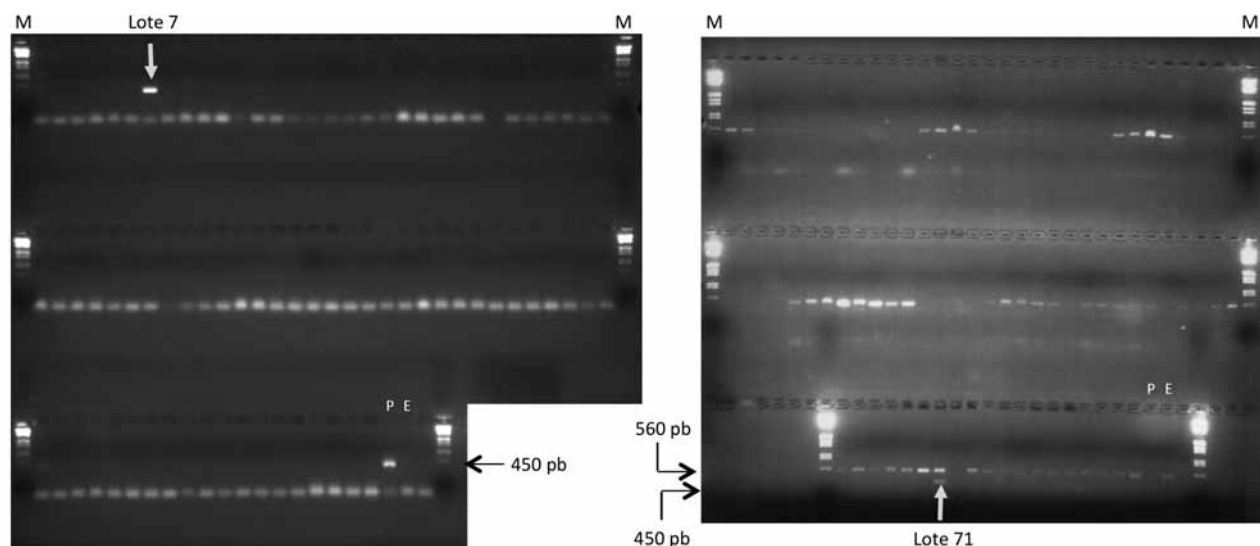


Figura 5. Rastreo por PCR de los 83 grupos de la genoteca de *P. pinaster* utilizando las parejas de cebadores ELP-3'/ELP-5' (izquierda), y Got3'A/Got3'B (derecha). Los amplificados se migraron en un gel de agarosa al 1,5%. Se señala con una flecha las amplificaciones específicas. 'P' es ADN genómico de *P. pinaster*. 'E' es ADN genómico de *E. coli*. 'M' es el marcador de pesos moleculares.

pb utilizando los cebadores Got3'A/Got3'B de Fd-GOGAT no mostró homología con la Fd-GOGAT sino con un ARNm desconocido de una genoteca de raíz de *P. taeda* (BQ654855) a una identidad del 72 %. El aislamiento de clones de Fd-GOGAT en el genoma de pino que no corresponden realmente a este gen es un hecho comprobado (Bautista *et al.*, 2007) que es coherente con la dispersión de la señal de hibridación de este gen por el genoma (Bautista, 2008).

Una vez identificado el grupo en el cual se encuentra un clon BAC de interés, sólo quedaría aislarlo entre los clones que forman el grupo. Para esta fase, es necesario realizar una dilución del cultivo del grupo, extender dicha dilución para aislar los clones y ordenarlos de forma individual. En esta etapa es muy recomendable poder disponer de un sistema robotizado que permita seleccionar las colonias, ordenarlas e imprimirlas sobre filtros de nylon para realizar un rastreo por hibridación (Isidore *et al.*, 2005; Salimath y Bhattacharyya, 1999).

3.6. Conclusión

El método expuesto en este trabajo permite construir una genoteca genómica de *Pinus pinaster* de una forma abordable. Gracias a la formación de grupos de clones se podrá almacenar en un espacio razonable a pesar del

gran número de clones necesario (entre 2 y 7 millones de clones, según la cobertura). Además, se ha desarrollado un sistema rápido de detección por PCR, que se ha validado al identificar un clon BAC de control dentro de un grupo de células al localizar el clon BAC de control introducido, así como sendos BAC con secuencias de otros genes. Esto, a su vez, confirma la validez del procedimiento y la genoteca obtenida. Sin embargo, no hay que descartar cierta frecuencia de aparición de amplificadas como el de los cebadores Got3'A/Got3'B que no producen la secuencia esperada. Posiblemente eso se deba a la naturaleza repetitiva del genoma de pino y a que unas secuencias son más abundantes que otras (Bautista, 2008).

Aunque hubiera sido factible completar la genoteca 3X necesaria para una representación significativa de todos los genes, no se ha llevado a cabo debido a que esta genoteca no podría usarse en un proyecto de secuenciación porque los cotiledones utilizados como material de partida corresponden a genotipos distintos. Puesto que recientemente se ha aprobado un proyecto piloto para la secuenciación del genoma de *P. pinaster* donde se contempla la generación de una genoteca completa a partir de tejido haploide de callos del megagametofito, se ha considerado prioritario dar a conocer el método a la comunidad científica que esperar a la construcción de la nueva genoteca completa para describirlo.

Agradecimientos

Queremos agradecer a Remedios Crespillo su inestimable asistencia técnica y a los Servicios Centralizados de Apoyo a la Investigación de la Universidad de Málaga por poner a nuestra disposición sus instalaciones. Este trabajo ha sido financiado por los proyectos BIO2006-06216 y AGL2006-07360 del Plan Nacional de I+D+I y el Proyecto de Excelencia AGR663 de la Junta de Andalucía.

Bibliografía

- ALLOUIS S., MOORE G., BELLEC A., SHARP R., FAIVRE RAMPANT P., MORTIMER K., PATEYRON S., FOOTE T.N., GRIFFITHS S., CABOCHE M., CHALHOUB B., 2003. Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm 'Chinese Spring'. *Cereal Research Communications* 31, 331-338.
- ARUMUGANATHAN K., EARLE E.D., 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* 9, 208-218.
- BAUTISTA R., VILLALOBOS D.P., DÍAZ-MORENO S.M., CANTÓN F.R., CÁNOVAS F.M., CLAROS M.G., 2007. Toward a *Pinus pinaster* bacterial artificial chromosome library. *Annals Forest Science* 64, 855-864.
- BAUTISTA R., 2008. Estrategias para el estudio del genoma de *P. pinaster*. Universidad de Málaga
- CANTÓN F.R., LE PROVOST G., GARCIA V., BARRÉ A., FRIGÉRIO J.M., PAIVA J., FEVEREIRO P., ÁVILA C., MOURET J.F., DE DARUVAR A., CÁNOVAS F.M., PLOMION C., 2004. Transcriptome analysis of wood formation in maritime pine. In: RITTER E., ESPINEL S., BARREDO Y., Eds., *Sustainable Forestry, Woods Products and Biotechnology*. Vitoria, Spain, pp. 333-348.
- COSTA P., PIONNEAU C., BAUW G., DUBOS C., BAHRMANN N., KREMER A., FRIGERIO J.M., PLOMION C., 1999. Separation and characterization of needle and xylem maritime pine proteins. *Electrophoresis* 20, 1098-1108.
- DOUDRICK R.L., HESLOP-HARRISON J.S., NELSON C.D., SCHMIDT T., NANCE W.L., SCHWARZACHER T., 1995. Karyotype of slash pine (*Pinus elliottii* var *elliottii*) using patterns of fluorescence in situ hybridization and fluorochrome banding. *J. Hered* 86, 289-296.
- FERNANDO D.D., 2005. Characterization of pollen tube development in *Pinus strobus* (Eastern white pine) through proteomic analysis of differentially expressed proteins. *Proteomics* 5, 4917-4926.
- FRIGERIO J.M., DUBOS C., CHAUMEIL P., SALIN F., GARCIA V., BARRÉ A., PLOMION C., 2004. Using transcriptome analysis to identify osmotic stress candidate genes in maritime pine (*Pinus pinaster* Ait.). In: RITTER E., ESPINEL S., BARREDO Y., Eds., *Sustainable Forestry, Wood products & Biotechnology*. Vitoria, Spain, pp. 348-362.
- GROTKOPP E., REJMANEK M., SANDERSON M.J., ROST T.L., 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates, supertree analysis. *Evolution* 58, 1705-1729.
- GUEVARA M.A., SOTO A., COLLADA C., PLOMION C., SAVOLAINEN O., NEALE D.B., GONZÁLEZ-MARTÍNEZ S.C., CERVERA M.T., 2005. Genomics applied to the study of adaptation in pine species. *Invest Agrar: Sist Recur For* 14, 292-306.
- HESLOP-HARRISON J.S., BRANDES A., TAKETA S., SCHMIDT T., VERSHININ A.V., ALKHIMOVA E.G., KAMM A., DOUDRICK R.L., SCHWARZACHER T., KATSIOTIS A., KUBIS S., KUMAR A., PEARCE S.R., FLAVELL A.J., HARRISON G.E., 1997. The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* 100, 197-204.
- ISIDORE E., SCHERRER B., BELLEC A., BUDIN K., FAIVRE-RAMPANT P., WAUGH R., KELLER B., CABOCHE M., FEUILLET C., CHALHOUB B., 2005. Direct targeting and rapid isolation of BAC clones spanning a defined chromosome region. *Funct Integr Genomics* 5, 97-103.
- JANDA J., BARTOS J., SAFAR J., KUBALAKOVA M., VALÁRIK M., CIHALIKOVA J., SIMKOVA H., CABOCHE M., SOURDILLE P., BERNARD M., CHALHOUB B., DOLEZEL J., 2004. Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat. *Theor Appl Genet* 109, 1337-1345.
- JANDA J., SAFAR J., KUBALAKOVA M., BARTOS J., KOVAROVA P., SUCHANKOVA P., PATEYRON S., CIHALIKOVA J., SOURDILLE P., SIMKOVA H., FAIVRE-RAMPANT P., HRIBOVA E., BERNARD M., LUKASZEWSKI A., DOLEZEL J., CHALHOUB B., 2006. Advanced resources for plant genomics: a BAC library specific for the short arm of wheat chromosome 1B. *Plant J* 47, 977-986.
- JOSEN R.V., LAMMERS M., BALK P.A., BRONNUM P., KONINGS M.C., PERKS M., STATTIN E., VAN WORTDRAGEN M.F., VAN DER GEEST A.L., 2006. Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. *Tree Physiol* 26, 1297-1313.
- KAMM A., DOUDRICK R.L., HESLOP-HARRISON J.S., SCHMIDT T., 1996. The genomic and physical organization of Ty1-copia-like sequences as a component of large genomes in *Pinus elliottii* var. *elliottii* and other gymnosperms. *Proc Natl Acad Sci U S A* 93, 2708-2713.
- KIM C.S., LEE C.H., SHIN J.S., CHUNG Y.S., HYUNG N.I., 1997. A simple and rapid method for isolation of high qua-

- lity genomic DNA from fruit trees and conifers using PVP. Nucl Acids Res 25, 1085-1086.
- KINLAW C.S., NEALE D.B., 1997. Complex gene families in pine genomes. Trends Plant Sci 2, 356-359.
- KOSSACK D.S., KINLAW C.S., 1999. IFG, a gypsy-like retrotransposon in *Pinus* (*Pinaceae*), has an extensive history in pines. Plant Mol Biol 39, 417-426.
- LAMOUREUX D., PETERSON D.G., LI W., FELLERS J.P., GILL B.S., 2005. The efficacy of Cot-based gene enrichment in wheat (*Triticum aestivum* L.). Genome 48, 1120-1126.
- LIPPERT D., CHOWRIRA S., RALPH S.G., ZHUANG J., AESCHLIMAN D., RITLAND C., RITLAND K., BOHLMANN J., 2007. Conifer defense against insects: proteome analysis of Sitka spruce (*Picea sitchensis*) bark induced by mechanical wounding or feeding by white pine weevils (*Pissodes strobi*). Proteomics 7, 248-270.
- LIU D., WU R., 1999. Protection of Megabase-Sized Chromosomal DNA from Breakage by DNase Activity in Plant Nuclei. BioTechniques 26, 258-261.
- LORENZ W.W., SUN F., LIANG C., KOLYCHEV D., WANG H., ZHAO X., CORDONNIER-PRATT M.M., PRATT L.H., DEAN J.F., 2006. Water stress-responsive genes in loblolly pine (*Pinus taeda*) roots identified by analyses of expressed sequence tag libraries. Tree Physiol 26, 1-16.
- LUO M., WANG Y.-H., FRISCH D., JOOBEUR T., WING R.A., DEAN R.A., 2001. Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (*Fom-2*). Genome 44, 154-162.
- MA Z., SONG W., SHARP P.J., LIU C., 2000. No-gridded library: a new approach for BAC (bacterial artificial chromosome) exploitation in hexaploid wheat (*Triticum aestivum*). Nucleic Acid Res 28, 106.
- PAVY N., LAROCHE J., BOUSQUET J., MACKAY J., 2005. Large-scale statistical analysis of secondary xylem ESTs in pine. Plant Mol Biol 57, 203-224.
- PETERSON D.G., TOMKINS J.P., FRITSCH D.A., WING R.A., PATERSON A.H., 2000. Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. J. Agric Genomics 5, <http://www.ncgr.org/jag/index.html>.
- RABINOWICZ P.D., 2003. Constructing gene-enriched plant genomic libraries using methylation filtration technology. Methods Mol Biol 236, 21-36.
- RALPH S.G., YUEH H., FRIEDMANN M., AESCHLIMAN D., ZEZNICK J.A., NELSON C.C., BUTTERFIELD Y.S., KIRKPATRICK R., LIU J., JONES S.J., MARRA M.A., DOUGLAS C.J., RITLAND K., BOHLMANN J., 2006. Conifer defence against insects: microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. Plant Cell Environ 29, 1545-1570.
- RALPH S.G., JANCSEK S., BOHLMANN J., 2007. Dirigent proteins in conifer defense II: Extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea* spp.). Phytochemistry 68, 1975-1991.
- ROCHETA M., CORDEIRO J., OLIVEIRA M., MIGUEL C., 2007. PpRT1: the first complete gypsy-like retrotransposon isolated in *Pinus pinaster*. Planta 225, 551-562.
- SALIMATH S.S., BHATTACHARYA M.K., 1999. Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the Rps1-k disease resistance gene. Theor Appl Genet 98, 712-720.
- SCHMIDT A., DOUDRICK R.L., HESLOP-HARRISON J.S., SCHMIDT T., 2000. The contribution of short repeats of low sequence complexity to large conifer genome. Theor Appl Genet 101, 7-14.
- SHIZUYA H., BIRREN B., KIM U.J., MANCINO V., SLEPAK T., TACHIIRI Y., SIMON M., 1992. Cloning and stable maintenance of 300-kilobase pair fragments of human DNA in *Escherichia coli* using a F-factor-based vector. Proc Natl Acad Sci USA 89, 8794-8797.
- SONG J., DONG F., JIANG J., 2000. Construction of a bacterial artificial chromosome (BAC) library for potato molecular cytogenetics research. Genome 43, 199-204.
- STUART-ROGERS C., FLAVELL A.J., 2001. The evolution of Ty1-copia group retrotransposons in gymnosperms. Mol Biol Evol 18, 155-163.
- SUZUKI G., SOOK DO G., MUKAI Y., 2002. Efficient Storage and Screening System for Onion BAC Clones. Breeding Science 52, 157-159.
- VANHOUTEN W., MACKENZIE S., 1999. Construction and characterization of a common bean bacterial artificial chromosome library. Plant Mol Biol 40, 977-983.
- VILLALOBOS D.P., BAUTISTA R., CÁNOVAS F.M., CLAROS M.G., 2004. Isolation of Bacterial Artificial Chromosome DNA by Means of Improved Alkaline Lysis and Double Potassium Acetate Precipitation. Plant Molecular Biology Reporter 22, 419-425.
- YANG S.H., LOOPSTRA C.A., 2005. Seasonal variation in gene expression for loblolly pines (*Pinus taeda*) from different geographical regions. Tree Physiol 25, 1063-1073.
- YUAN Y.N., SANMIGUEL P.J., BENNETZEN J.L., 2003. High-Cot sequence analysis of the maize genome. Plant Journal 34, 249-255.
- ZELENA L.B., SOROCHYNS'KYI B.V., 2005. [Characteristics of gene expression in morphologically abnormal needles of Scots pine]. Ukr Biokhim Zh 77, 78-83.
- ZHANG H.-B., ZHAO X., DING X., PATERSON A.H., WING R.A., 1995. Preparation of megabase-size DNA from plant nuclei. Plant J 7, 175-184.