

Tesis Doctoral

Desarrollo de marcadores funcionales y evaluación de la diversidad genética en *Eucalyptus globulus* con énfasis en genes potencialmente involucrados en características de calidad de la madera

Acuña, Cintia Vanesa

2011

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Acuña, Cintia Vanesa. (2011). Desarrollo de marcadores funcionales y evaluación de la diversidad genética en *Eucalyptus globulus* con énfasis en genes potencialmente involucrados en características de calidad de la madera. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Acuña, Cintia Vanesa. "Desarrollo de marcadores funcionales y evaluación de la diversidad genética en *Eucalyptus globulus* con énfasis en genes potencialmente involucrados en características de calidad de la madera". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2011.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales

**Desarrollo de marcadores funcionales y evaluación de la
diversidad genética en *Eucalyptus globulus* con énfasis
en genes potencialmente involucrados en características
de calidad de la madera**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires
en el área Ciencias Biológicas

Lic. Cintia Vanesa Acuña

Director de Tesis: Dr. Horacio Esteban Hopp

Directora Asistente: Dra. Susana Noemí Marcucci Poltri

Consejero de Estudios: Dr. Horacio Esteban Hopp

Lugar de trabajo: Instituto de Biotecnología e Instituto de Recursos Biológicos
CNIA, Instituto Nacional de Tecnología Agropecuaria, Castelar
Buenos Aires, 2011

Desarrollo de marcadores funcionales y evaluación de la diversidad genética en *Eucalyptus globulus* con énfasis en genes potencialmente involucrados en características de calidad de la madera

RESUMEN

Eucalyptus globulus es la especie forestal con mejor aptitud papelera y para la obtención de bioenergía a partir de celulosa, mayormente plantada en regiones templadas del mundo. Los proyectos genómicos en *Eucalyptus* han incrementado el número de secuencias disponibles en los bancos de datos públicos. Los marcadores funcionales públicos, si bien son frecuentes en diferentes cultivos, son aún escasos en especies forestales. De allí la importancia de la búsqueda y validación de regiones SSRs en genes de interés, para ser utilizados en futuros proyectos de mejoramiento asistido por marcadores (*marker assisted breeding*).

En este estudio, se identificaron secuencias no redundantes de ADN de *Eucalyptus* (genómicas y ESTs) depositadas en bancos de datos públicos para revelar secuencias microsatélites y predecir, *in silico*, su función putativa. Éstas fueron luego validadas en laboratorio para predecir su potencial uso en análisis de diversidad genética.

A partir de 12.690 ESTs de *Eucalyptus globulus* publicados en NCBI se identificaron 4.924 secuencias no redundantes. De éstas, 952 unigenes (19,3%) contenían 1.140 regiones SSR. Luego del análisis bioinformático de estos EST-SSRs, se diseñaron 979 oligonucleótidos novedosos y se predijo su función putativa, incluyendo categorías Gene Ontology (GO) según su proceso biológico, función molecular y componente celular.

Se identificaron así 29 SSRs en 24 genes candidatos (estructurales y reguladores) para calidad de madera. Los SSRs se encontraron en promotores, intrones y exones de genes candidatos (GC) de distintas rutas metabólicas (biosíntesis del fenilpropanoico, biosíntesis de celulosa, metabolismo de hemicelulosas, ruta metabólica del shikimato, metabolismo de la metionina y genes de tubulinas y el factor de transcripción LIM1).

Un total de 85 SSRs (56 EST-SSRs y 29 SSRs incluidos en GC) hallados en este trabajo fueron analizados para su validación en 8 genotipos de *E. globulus*, resultando positivos

un 65%. A partir de esta validación se obtuvieron 17 EST-SSRs y 12 GC-SSRs polimórficos. Con éstos se estimaron los valores de diversidad en una muestra de 60 individuos no emparentados de *E. globulus* representantes de seis razas que cubren el rango de distribución natural de la especie. Los valores de PIC, Ho y UHe variaron ampliamente entre aproximadamente 0,02 y 0,9, mientras que el número de alelos varió entre 2 y 16, con un promedio de 7,55.

Al realizar el test de Equilibrio de Hardy Weinberg (EHW) para los 29 *loci*, se encontró que 14 de ellos mostraron un significativo déficit de heterocigotas ($P < 0,01$). Este desvío del EHW podría explicarse por la presencia de subestructura poblacional y a la presencia de alelos nulos que sesgó las estimaciones de las frecuencias alélicas.

Se realizó el estudio de transferibilidad de 49 *loci* (37 EST-SSRs validados (polimórficos y monomórficos) y los 12 GC-SSRs polimórficos) a otras seis especies de *Eucalyptus* (*E. camaldulensis*, *E. dunnii*, *E. grandis*, *E. saligna*, *E. tereticornis* y *E. viminalis*). Se encontró que 33 marcadores amplificaron en las seis especies estudiadas y seis en al menos cinco, mostrando la alta transferibilidad de los mismos.

Finalmente, el estudio del polimorfismo y transferibilidad de los marcadores funcionales, permitió la selección de un conjunto de 13 (7 EST-SSRs y 6 GC-SSRs) altamente informativos y transferibles a otras seis especies de *Eucalyptus*.

El conjunto de marcadores desarrollados son altamente informativos tendrán un uso potencial en estudios de diversidad genética, taxonomía, mapeo de QTL y en facilitar, en un futuro, la selección asistida en el mejoramiento de *Eucalyptus*.

Palabras claves: *Eucalyptus globulus*, marcadores funcionales, SSR, diversidad genética, genes candidatos para calidad de madera, transferibilidad.

Functional markers development and genetic diversity analysis in *Eucalyptus globulus* with emphasis in wood quality candidate genes

ABSTRACT

Eucalyptus globulus is the most planted hardwood species for pulpwood in temperate regions. Genomic researches in *Eucalyptus* have increased the information available in DNA sequence's public databases. Functional genetic markers, while frequent in crop, are still scarce in forest species. Hence the detection and validation of SSRs in interesting genes to be used in future projects of marker-assisted breeding are needed.

Here, *Eucalyptus* DNA sequences (genomics and ESTs) from public databases were screened to identify non redundant sequences, to discover microsatellite sequences and to *in silico* predict putative gene functions. These were also validated in wet lab to predict their potential for functional genetic diversity analysis.

From 12,690 updated *E. globulus* EST database published in National Center for Biotechnology Information a total of 4,924 non-redundant sequences were identified. From these ones, 952 unigenes (19.3 %) contained 1,140 SSRs. A new set of 979 primers were designed for putative SSR-markers after bioinformatic analysis. The predicted functions of these EST-SSRs were adjudged, including biological process, molecular function and cellular component Gene Ontology (GO) categories.

Twenty four structural and regulatory candidate genes for wood quality carrying 29 SSR were indentified. Microsatellite sequences were located in promoters, introns and exons from candidate genes (CG) from: phenylpropanoid biosynthesis, cellulose biosynthetic process, hemicellulose metabolism, shikimate pathway, methionine metabolism, tubulin genes and the transcriptor factor LIM1.

Sixty five percent out of a total of 85 SSR (56 EST-SSRs and 29 SSR containg GC) detected in this work were validated for actual PCR amplification of tree DNA samples in eight genotypes of *E. globulus*. From this assessment a total of 17 polymorphic EST-SSRs and 12 polymorphic CG-SSRs markers were obtained. These ones were selected for further analyses, so as to accurately estimate genetic information content in a larger sample of 60 non related trees represented major geographical races of the species' natural distribution.

PIC, H_o and UH_e values varied over a wide range from around 0.02 to 0.9, whereas the allele number ranged from 2 to 16, with an average of 7.55.

Fourteen out of the 29 loci tested showed significant deviation from Hardy Weinberg Equilibrium (HWE) showing a significant deficit of heterozygotes ($P < 0.01$). This deviation from HWE could be explained by the presence of substructure population and the presence of null alleles that biased estimates of allele frequencies.

A set of 49 loci (37 validated EST-SSRs (polymorphic and monomorphic) and 12 polymorphic CG-SSRs) were also tested for cross-transferability to other six *Eucalyptus* species (*E. grandis*, *E. saligna*, *E. dunnii*, *E. viminalis*, *E. camaldulensis*, *E. tereticornis*). A total of 33 out of the 49 validated markers in *E. globulus* amplified in the six other species and six markers amplified in at least other five.

Finally, the analyses of polymorphism and transferability of functional markers, enabled the selection of a set of 13 (7 EST-SSRs and 6 GC-SSRs) highly informative and transferable to other six species of *Eucalyptus*.

The set of highly informative markers developed here will have potential use in studies of genetic diversity, taxonomy, gene mapping and will help the improvement of *Eucalyptus* through the assisted selection.

Keywords: *Eucalyptus globulus*, functional markers, SSR, genetic diversity, candidate genes for wood quality, transferability.

AGRADECIMIENTOS

Quiero agradecer a todas las personas que colaboraron de una u otra forma para la realización de esta tesis.

Al INTA por permitirme desarrollar este trabajo en los Institutos de Recursos Biológicos y Biotecnología de INTA y al CONICET por otorgarme la Beca de Postgrado.

A los Ing. Agr. Enrique Suárez e Ing. Agr. Marcelo Ferrer, directores del Instituto de Recursos Biológicos y Dra. Elisa Carrillo, directora del Instituto de Biotecnología del INTA-Castelar, por darme la oportunidad de trabajar en ambos Institutos y por poner a disposición los materiales e instalaciones de los laboratorios.

A Esteban por aceptar ser mi director, por la guía en los momentos complicados de los experimentos y por el constante apoyo.

A Susana por guiarme, enseñarme y por la constante dedicación. En lo personal, por todos los momentos compartidos durante estos 10 años de trabajar juntas. Por llevarme tantas veces a INTA y ¡aguantar los berrinches de Franco! Por aconsejarme como mamá y tenerme paciencia en mi aprendizaje.

A todo el personal del Instituto de Recursos Biológicos del INTA-Castelar por brindarme su ayuda y cariño permanente. Especialmente quiero agradecer a Susana T. y Noga por compartir mis primeros años forestales y brindarme su ayuda. A mis amigas Marce, Vir, Sabri, Renée, Mari y Ale por compartir tantas charlas, consejos y terapias de grupo!! A Carla por tantas cosas: por alentarme, compartir momentos tan importantes en nuestras vidas, por abrirme su casa y conocer a su hermosa familia, por luchar por un jardín mejor y por enseñarme a encarar las cosas con tanta energía! Gracias amiga!

A Pablo Pathauer por la colecta de las muestras y a Eduardo Cappa por brindarme su ayuda y aclarar mis dudas.

A todo el personal del Instituto de Biotecnología por recibirme tan cordialmente. No quisiera olvidarme de nadie (son tantos!), así que agradezco a TODOS por la ayuda en estos años, por preocuparse cuando tuve problemas familiares y por las sonrisas cruzadas en los pasillos.

A mis compañeros de laboratorio Mariano, Laura, Diego, Betiana y Ana por hacer que los ensayos sean más divertidos.

A mis compañeros “forestólogos” Pame y Martín por socorrerme en el lab cada vez que tenía que salir corriendo al jardín, por ayudarme en todo momento (¡especialmente con el GeneMapper!)

A Daniela y Marisa por las charlas en la oficina, por los consejos y por compartir anécdotas de nuestros hijos.

A los chicos del “locutorio”: Cori, Guille, Carlos, Diego, Lau, Carla, Lucila, Mariana, Luis, Flor, Ariel, Fede, Sebastián, Jere y Gabriel por las charlas, los mates y la ayuda con las dudas varias que van surgiendo...y por soportarme en estos últimos tiempos del doctorado!

A LauKa por las charlas, los consejos y por alentarme siempre.

A Vero Lía por el “Centro de asistencia al becario”, por tu predisposición para discutir sobre los análisis poblacionales y a Valeria Moreno por la discusión de los resultados del STRUCTURE.

A los chicos del secuenciador Apu, Pablo, Natalia y especialmente a Vero por correr siempre con mis muestras y preguntar siempre por mis resultados. Gracias!

A Paula F. por ayudarme con el Blast2go y a los chicos de bioinformática por la asistencia.

A todas mis compañeras con las que compartimos el aprendizaje de ser madres y tantos lindos recuerdos en el jardín de INTA. Gracias a Marie, Gabi y Adri por ayudarme con Franquito.

A mis amigos Leo, Ire, Negro, Vicky, Pato, Esteban, Juli, Martín, Ariel, Mariana, Caro, Mariano, Matías, por tantos momentos compartidos, por preguntarme siempre “¿cómo va la tesis?”.

A mis amigas del alma (por estricto orden alfabético): Adri, Gaby, Gise, Mari, Mere, Vale y Vani. Un GRACIAS enorme por todos estos años, por estar siempre juntas desde nuestro primer día en la UNLu! Las quiero mucho.

A mi amiga y hermana Yamila. Gracias por elegirme para ser la madrina de Binta. Creo que eso resume lo que significa para nosotras esta amistad de toda la vida. Estoy feliz por eso!

A mi familia por acompañarme, alentarme en todo momento y confiar en mí: a mamá, papá, tíos Marta y Moncho, a Isidro, primos Vivi, Gaby, Beto, a mi sobri Abril, a Naira y a Paito por llegar con una sonrisa a esta familia. A la abuela que me sigue acompañando siempre. A mi ahijada Vicky por enseñarme a luchar en la vida y a que se puede salir adelante a pesar de todo lo que nos toca. Quiero agradecerles especialmente a mis hermanos Vale y Ari, porque los amo y se que siempre puedo contar con ustedes! A Vale y Juan por estar siempre siempre siempre dispuestos a cuidar a sus sobris! Los quiero mucho!

A mi segunda familia por brindarme su amor y su ayuda. Especialmente a Mirta por la gran ayuda con los nietos, a Juampi y Myri por estar siempre y por mi nueva ahijadita Mora! Los quiero mucho.

Esta es la parte más difícil. ¿Cómo agradecerle por ser mi compañero de todos estos años, por elegirnos para formar una familia y tener dos hijos maravillosos, por ser mi pilar y ayudarme siempre, desde que éramos estudiantes para que yo pueda seguir adelante, sin bajar los brazos? Creo que con un “GRACIAS Hernán” enorme no alcanza. Sin vos no lo habría logrado. Te amo.

A mis hijos Franco y Bruno, por ser mi sol y lo mejor que me pasó en la vida. Por enseñarme a ser mamá y a saber cuáles son las prioridades en la vida. Por las sonrisas, los besos, por verlos crecer. Les doy las gracias por la paciencia durante este tiempo en que “mamá está muy ocupada escribiendo un libro gordo”. Los amo con todo mi corazón!

A Hernán

A Franco y Bruno

A mi familia y amigos

Parte de los resultados presentados en este trabajo de tesis fueron publicados

- **Acuña C**, Villalba P, Pathauer P, Hopp HE y Marcucci Poltri SN. Análisis de marcadores microsatélites en genes candidatos relacionados a calidad de madera en *Eucalyptus globulus* | Analysis of microsatellite markers in candidate genes related to wood quality in *Eucalyptus globulus*. (2010). Análisis de Semillas N° 14, Vol. 2, Tomo IV. pp 103-107.
- **Acuña C**, Fernandez P, Villalba P, García M, Hopp E, Marcucci Poltri S. Discovery, validation and *in silico* functional characterization of EST-SSR markers in *Eucalyptus globulus*. Tree Genetics and Genomes, aceptado para publicación con cambios menores
- **Acuña C**, Villalba P, García M, Pathauer P, Hopp E, Marcucci Poltri S. Characterization of novel microsatellite markers in candidate genes for wood properties for application in functional diversity assessment in *Eucalyptus globulus*. Enviado a Electronical Journal of Biotechnology.

ÍNDICE GENERAL

1. INTRODUCCIÓN	2
1.1 Forestación en Argentina	2
1.2 El género <i>Eucalyptus</i>	3
1.3 <i>Eucalyptus globulus</i> y su importancia	3
1.4 Programa de mejoramiento genético de <i>Eucalyptus globulus</i> en Argentina y el uso de marcadores moleculares	4
1.5 Perspectivas funcionales de los SSR	6
Figura 1.1: Función y efecto del cambio en el número de repeticiones en secuencias microsatélites.	7
1.6 Composición de la madera y genes candidatos para calidad	8
Figura 1.2: Ruta metabólica de síntesis de lignina.....	10
HIPÓTESIS y OBJETIVOS	13
2. MATERIALES Y MÉTODOS	16
2.1 Material vegetal.....	16
Tabla 2.1: Individuos de <i>Eucalyptus</i> utilizados en este estudio	17
2.2 Extracción de ADN genómico.....	18
2.3 Cuantificación de ADN	19
2.4 Análisis bioinformático para la identificación de regiones microsatélites en secuencias expresadas etiquetadas (ESTs)	19
2.4.1 Detección de unigenes.....	19
2.4.2 Determinación de secuencias microsatélites (SSR) en unigenes	19
Tabla 2.2. Posibles combinaciones de los di-y trinucleótidos.....	20
2.4.3 Determinación de posibles funciones de los unigenes que contienen microsatélites (EST-SSRs).....	21
Figura 2.1: Representación esquemática de las aplicaciones del programa Blast2GO.	22
2.5 Selección de genes candidatos e identificación de regiones microsatélites.....	22
2.6 Amplificación por PCR de las regiones microsatélites.....	23
2.7 Electroforesis de los fragmentos de amplificación.....	24
2.7.1 Visualización de ADN en geles de poliacrilamida	24
2.7.2 Métodos automáticos	25
2.7.3 Marcación directa.....	26
2.8 Purificación de fragmentos de ADN de geles de poliacrilamida.....	26
2.9 Análisis de microsatélites usando el programa GeneMapper® 4.0	27
Figura 2.3 Alelos del locus HMTe	28
2.10 Análisis estadístico de los datos obtenidos.....	28
2.10.1 Generación de la matriz básica de datos, matriz de distancia y análisis de agrupamiento	28
2.10.2 Estimadores del potencial de los marcadores para análisis de diversidad genética....	30
2.10.3 Equilibrio de HW, frecuencia de alelos nulos y estructura genética.....	31
3. RESULTADOS Y DISCUSIÓN	34
3.1 Capítulo I Búsqueda, caracterización funcional <i>in silico</i> y validación experimental de marcadores EST-SSRs en <i>Eucalyptus globulus</i>	34
3.1.1 Ensamblaje y definición de unigenes.....	34
3.1.2 Identificación de secuencias repetitivas microsatélites (SSR) en los unigenes.....	34
3.1.3 Caracterización de los motivos microsatélites	35
Figura 3.1: Frecuencia de los motivos repetitivos según el número de SSRs presentes en unigenes.	36
Figura 3.2: Frecuencia de los diferentes motivos repetitivos.....	38
Tabla 3.1: Posición de los distintos motivos repetitivos en ORFs o UTRs.....	39
3.1.4 Anotación funcional de los EST-SSRs de <i>Eucalyptus globulus</i>	40

Figura 3.3: Histograma del número de hits BLASTX encontrados en las distintas especies con secuencias de proteínas depositadas en GenBank	40
Figura 3.4: Asignación Gene Ontology (GO).....	42
3.1.5 Validación de los EST-SSRs en <i>E. globulus</i>	43
Tabla 3.2 Loci EST-SSRs validados en <i>Eucalyptus</i>	44
3.2 Capítulo II Caracterización de marcadores microsatélites en genes potencialmente involucrados en características de calidad de madera.....	49
3.2.1 Caracterización de regiones microsatélites.....	49
3.2.1.1 Microsatélites en secuencias ESTs	49
3.2.1.2 Microsatélites en secuencias genómicas de <i>Eucalyptus</i> depositadas en GenBank.....	51
3.2.2 Validación de los SSRs en genes candidatos para calidad de madera	52
3.2.3 Descripción de los SSRs polimórficos en GC para calidad de madera.....	54
Gen cinamoil-CoA reductasa (CCR)	54
Gen para la celulosa sintasa 3 (CesA3)	56
Gen del factor de transcripción LIM1.....	57
Gen de homocisteína S-metiltransferasa (HMT).....	58
Gen de shikimato kinasa (SK).....	60
Gen de xiloglucano endotransglicosilasa (XTH2)	61
Gen de glutatión S-transferasa (GST).....	62
Gen de glutamato descarbolixasa (GAD)	63
Gen de peroxidasa (PER).....	65
3.3 Capítulo III Evaluación de la capacidad para el análisis de variabilidad poblacional del conjunto seleccionado de marcadores funcionales y su transferibilidad a otras especies del género	68
3.3.1 Estudio de los patrones de distribución de los alelos de EST-SSRs y GC-SSRs de <i>E. globulus</i>	68
3.3.2 Estudio de los niveles de polimorfismo de EST-SSRs y GC-SSRs de <i>E. globulus</i>	72
3.3.3 Aplicación de los SSRs para el estudio de relación genética entre individuos de <i>E. globulus</i>	79
3.3.4 Transferibilidad de los marcadores funcionales a otras especies de <i>Eucalyptus</i>	82
Tabla 3.9 Transferibilidad de loci evaluados en <i>E. globulus</i> a otras seis especies del género.....	85
Figura 3.19 Patrón alélico de las siete especies analizadas.	87
Tabla 3.10 Evaluación de loci posibles de ser ubicados en los mapas de las tres poblaciones de cruzamientos controlados de <i>E. grandis</i>	88
3.3.5 Aplicación de los SSRs para la discriminación de los individuos entre las siete especies de <i>Eucalyptus</i> estudiadas	88
Figura 3.20 Histograma de las distancias DAS entre pares de individuos de las siete especies estudiadas	89
Figura 3.21: Dendrograma generado a partir de la matriz de distancia entre individuos de las siete especies analizadas.....	90
Figura 3.22: Test de Evanno resultante del análisis poblacional mediante STRUCTURE sin información <i>a priori</i>	91
Figura 3.24 Análisis de la muestra mediante STRUCTURE mostrando la asignación de todos los individuos a los k grupos obtenidos sin información <i>a priori</i> (A) y con información <i>a priori</i> (B y C)	94
Tabla 3.11 Matriz de diferenciación PhiPT entre las siete especies analizadas	96
Figura 3.25: Agrupamiento (UPGMA) generado a partir de la matriz de PhiPT entre especies.....	97
Figura 3.26: Gráfico representativo del análisis de la variancia molecular (AMOVA)	99
Figura 3.27: distribución de frecuencias de PhiPT.....	99
3.3.6 Selección de un grupo de marcadores funcionales más informativos.....	100
3.4 Recapitulación	101

4. CONCLUSIONES Y PERSPECTIVAS	104
Referencias bibliográficas	107
Tablas anexas	121

ABREVIATURAS Y SIGLAS

locus o loci: se utilizan en esta tesis en sentido *sensu lato*. Es decir, una posición genómica ocupada por un marcador molecular ubicada en una secuencia transcrita que puede o no ser codificante.

°C	grados centígrados
ADN	ácido desoxirribonucleico
ADNc	ADN copia
AN	número de acceso en el Genbank
ARN	ácido ribonucleico
ARNm	ARN mensajero
AFLP	polimorfismos de longitud de fragmentos amplificados, <i>Amplified Fragment Length Polymorphism</i>
ADNc	copia de ADN del ARNm
CTAB	Bromuro de cetiltrimetilamonio, <i>Cethyl Trimethyl Ammonium Bromide</i>
dNTP	desoxinucleótidos trifosfato
EC	electroforesis capilar
EDTA	Ácido etilendiaminotetraacético, <i>Ethylene-diaminetetraacetic acid</i>
EHW	Equilibrio de Hardy-Weinberg
EST	secuencias que se expresan, <i>Expressed Sequence Tags</i>
EST-SSR	secuencia EST que posee un microsatélite
EtBr	bromuro de etidio
EUA	Estados Unidos de América
g	gramo
h	hora
InDels	<i>insertions-deletions</i>
INTA	Instituto Nacional de Tecnología Agropecuaria
kpb	kilo pares de bases
DL	desequilibrio de ligamiento
M	molar
MCMC	métodos de Monte Carlo basados en cadenas de Markov
mg	miligramo
min	minuto
ml	mililitro
mM	milimolar
N°	número

ng	nanogramo
ORF	marco abierto de lectura, <i>Open Reading Frame</i>
pb	pares de bases
PCR	reacción en cadena de la polimerasa, <i>Polymerase Chain Reaction</i>
p/v	peso/volumen
PVP	polivinilpirrolidona
QTL	<i>loci</i> de características cuantitativas, <i>Quantitative Trait loci</i>
RAPD	ADN polimórfico amplificado al azar, <i>Random Amplified Polymorphic DNA</i>
RFLP	polimorfismos en la longitud de los fragmentos de restricción, <i>Restriction Fragment Length Polymorphism</i>
RFU	rango de fluorescencia
SNP	polimorfismo de nucleótido simple, <i>Single Nucleotide Polymorphisms</i> .
SSR	Secuencias Simples Repetidas o microsatélites, <i>Simple Sequence Repeats</i>
rpm	revoluciones por minuto
seg	segundo
TAE	<i>buffer</i> de electroforesis TRIS-acetato/EDTA
TBE	<i>buffer</i> de electroforesis TRIS-borato/EDTA
u	unidades
μl	microlitro
UPGMA	ligamiento promedio no ponderado, <i>Unweighted Pair Group Method with Arithmetic Mean</i>
UTR	Región no traducida/ por sus siglas en inglés: <i>Untranslated Region</i>
UV	luz ultravioleta

Introducción

1. INTRODUCCIÓN

1.1 Forestación en Argentina

Los árboles son organismos con ciclos de vida largos y debido a que los tiempos implicados para la expresión de la mayoría de los rasgos económicos importantes son bastante elevados, el proceso de cultivo y selección es muy lento, determinando que los programas de mejoramiento forestal sean una parte importante de la silvicultura intensiva. Si bien se han optimizado las técnicas de selección, aún existen algunas limitaciones tales como el tiempo que se necesita para alcanzar la madurez sexual y el tiempo requerido para evaluar el desarrollo del crecimiento. Debido a esto, la mayoría de los programas de mejoramiento forestal en el ámbito mundial, están orientados a mejorar simultáneamente más de una característica. Este escenario también se refleja en los actuales programas de mejoramiento nacional, donde el incremento del valor agregado de la madera destinada a la industria papelera, debe realizarse a partir de la selección de especies con propiedades papeleras específicas.

Argentina posee entre 18 y 20 millones de hectáreas de suelos con aptitud forestal, de las cuales 5 millones no compiten en el uso con otras actividades agropecuarias alternativas.

El 75% de las tierras de aptitud forestal se encuentra en tres regiones: la pampeana que presenta, sólo en la Provincia de Buenos Aires, una superficie apta de 6 millones de hectáreas, la mesopotámica, con 5 millones de hectáreas de aptitud y la Patagónica con 4 millones de hectáreas disponibles para este fin. Sólo el 20% de las superficies que no compiten con otras actividades agropecuarias se encuentran forestadas, correspondiendo a una superficie aproximada de 800 mil hectáreas.

En Argentina, la superficie actualmente forestada con eucaliptos se estima en unas 330.000 has, lo cual representa aproximadamente un 27 % del total plantado (1,2 millones de ha a 2004), de las cuales la región mesopotámica concentra más del 80%, seguida por la región pampeana y NOA en mucho menor escala. La tasa anual de reposición se estima en unas 50.000 ha, correspondiendo a *Eucalyptus* un 30% (15.000 ha).

El elevado número de especies del género *Eucalyptus*, con distintos requerimientos agroecológicos y calidad de madera, hacen de este género un recurso de gran importancia por su adaptación a los numerosos ecosistemas de Argentina.

1.2 El género *Eucalyptus*

El género *Eucalyptus* es originario de Australia, y comprende a más de 700 especies y subespecies que pertenecen a 13 linajes evolutivos principales (Brooker 2000). Los árboles pertenecientes a este género son diploides con un número de cromosomas haploide de 11 (Eldridge y col. 1993, Potts y Wiltshire, 1997). El tamaño del genoma fue calculado para varias especies de eucaliptos y sus híbridos por Grattapaglia y Bradshaw (1994). Se estimaron tamaños haploides de los genomas que van desde 370 hasta 700 millones de pares de bases (Mpb). Las especies del subgénero *Symphyomyrtus* presentaron un promedio de tamaño del genoma haploide de 650 Mpbs. Dentro de éste, las especies pertenecientes a la misma sección presentaron contenidos similares de ADN, con *E. globulus* y *E. dunnii* en el extremo inferior de la escala (530 Mpbs) y *E. saligna* en el extremo superior (710 Mpbs).

Al considerar el cultivo de eucalipto en el mundo, se observa que sobre un total de más de 700 especies que existen en la naturaleza, sólo algunas del subgénero *Symphyomyrtus* se destacan del resto por su importancia industrial: *E. grandis*, *E. globulus*, *E. gunnii*, *E. dunnii*, *E. saligna*, *E. viminalis*, *E. tereticornis*, *E. calmadulensis*, *E. nitens*, *E. cloeziana*, *E. robusta*, *E. urophylla*, *E. maculata*, *E. sideroxylon*, *E. botryoxylon* y algunos híbridos interespecíficos. Muchas de ellas son de interés por producir una gran cantidad de madera en un ciclo corto, capacidad de recuperación ante la acción negativa del fuego, la sequía y el ramoneo; colonización de suelos pobres deteriorados por la erosión o la agricultura irracional.

1.3 *Eucalyptus globulus* y su importancia

Entre los productos obtenidos a partir de las especies de *Eucalyptus* cultivadas se encuentran la madera aserrada, laminada, faqueada y aglomerada; pulpa para papel; postes; leña, carbón y taninos. Debe señalarse que es muy difícil encontrar una única especie apta para la producción de todos los productos (Carpineti y col. 1995).

Desde la óptica del mercado mundial de celulosa y papel sobre la base de eucaliptos, *Eucalyptus globulus*, *E. grandis* y el híbrido *E. grandis* x *E. urophylla*, concentran la principal oferta de materia prima. Dentro de ellas, *Eucalyptus globulus*, es el preferido por la industria papelera basada en este género debido a que su madera tiene una alta densidad, por lo tanto un alto rendimiento en el proceso industrial, y una buena blancura que permite lograr valores muy altos de blanco en el papel. El bajo contenido de lignina también es una característica importante para la utilización de los residuos forestales para la generación de biocombustibles (bioetanol) porque reduce los componentes inhibitorios y enriquece la concentración relativa de los hidratos de carbono fermentables.

Por tal motivo, esta especie originaria de Tasmania y sudeste de Victoria en Australia, ha sido introducida en varias regiones templadas del mundo (Argentina, Brasil, Francia, India, España y Portugal). En Argentina, la similitud climática y latitudinal registrada entre el área de origen y el Sudeste Bonaerense, sumada a las favorables condiciones edáficas que se presentan en la región, explican su excelente adaptación a las condiciones locales.

1.4 Programa de mejoramiento genético de *Eucalyptus globulus* en Argentina y el uso de marcadores moleculares

El programa de mejoramiento de *Eucalyptus globulus*, que se lleva a cabo desde 1993 en el Instituto Nacional de Tecnología Agropecuaria (INTA), contempló la introducción de una amplia base genética, establecida a campo por medio de ensayos concentrados en el sudeste de la provincia de Buenos Aires (López y col. 2001). Dicho programa tiene como objetivo principal aumentar la cantidad y calidad de la madera producida.

La Unidad Bosques Cultivados del Instituto de Recursos Biológicos (IRB) de INTA Castelar, en el marco del Proyecto Forestal de desarrollo para la Región Pampeana (SAGPyA BIRF) y el Convenio Internacional de Asistencia Técnica INTA-Sociedad Portuguesa de Papel (SOPORCEL), ha instalado desde el año 1995, una red de ensayos de orígenes y progenies de *E. globulus* en 4 localidades de la provincia de Buenos Aires. Los mismos provienen de ocho razas nativas australianas y cuatro razas locales (Argentina, Chile, Portugal y España). Su finalidad es la de establecer una población base, genéticamente amplia, para ser empleada en programas de mejoramiento.

El análisis de las características medidas en los árboles (diámetro del fuste y penetración de Pilodyn) a lo largo de distintas edades y diferentes sitios, permitió determinar el grado de adaptación de los orígenes y razas así como comparar sus afinidades, para luego integrar estos parámetros en un índice de selección individual (López y col. 2001). Actualmente se cuenta con parámetros genéticos individuales de densidad de madera, así como otros caracteres de interés forestal, como volumen y forma.

La caracterización y cuantificación de la variabilidad genética en especies de *Eucalyptus*, son actividades desarrolladas en el INTA Castelar desde 1998.

En *E. globulus* ya han sido evaluados con marcadores moleculares neutros aproximadamente 150 individuos, representativos de las razas geográficas descritas. Se emplearon 6 combinaciones de oligonucleótidos de AFLP, 10 *loci* de microsatélites (SSR), 3 microsatélites de cloroplasto (cpSSR) y el *locus* hipervariable JLA de cloroplasto para la cuantificación de la diversidad genética (Acuña y col. 2005; Torales y col. en preparación).

1.5 Marcadores moleculares neutros y funcionales

En la última década, se han utilizado diversos marcadores moleculares de ADN en estudios de diversidad genética, tales como: restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP) y microsatelites o simple sequence repeats (SSR) (Shepherd y Jones 2004). Los SSRs son ampliamente utilizados como marcadores moleculares por varias razones, como su relativa abundancia, distribución azarosa en el genoma, elevado polimorfismo (debido a su naturaleza multialélica) y herencia Mendeliana codominante (Morgante y Olivieri 1993; Powell y col. 1996). Otra ventaja de estos marcadores, es su relativamente simple detección utilizando amplificación por PCR, a partir de oligonucleótidos flanqueantes únicos. El método estándar de desarrollo de marcadores SSR involucra la creación de genotecas de pequeños insertos genómicos, la hibridización con oligonucleótidos de repeticiones en tándem, la secuenciación del clon candidato, haciendo el proceso costoso y laborioso (Gupta y Varshney 2000).

En *Eucalyptus*, se desarrollaron aproximadamente 450 SSRs de motivo (AG)_n y (AC)_n en *E. grandis* y *E. urophylla* (Brondani y col. 1998, 2002, 2006) y 8 SSRs en *S. sieberi* (Glaubitz y col. 2001). Cinco pares de oligonucleótidos se desarrollaron mediante técnicas de enriquecimiento de ISSR, conservadas entre distintas especies (*E. grandis*, *E. nitens*, *E. globulus*, *E. camaldulensis* y *E. urophylla*) (Van der Nest y col. 2000). En *E. globulus*, se desarrollaron oligonucleótidos para amplificar 20 regiones microsatélites (Steane y col. 2001), cuatro loci microsatélites se han caracterizado en *E. nitens* y seis en otras especies de *Eucalyptus* (Byrne y col. 1996).

El análisis de secuencias EST (*Expressed Sequence Tags*) no sólo es la herramienta más eficiente para el estudio y descubrimiento de genes, sino también es para el desarrollo de nuevos marcadores moleculares funcionales, como SSR y SNP.

El desarrollo de marcadores SSRs derivados de ESTs (EST-SSRs) tiene la ventaja, frente a los métodos tradicionales, de obtenerse a partir de bases de datos públicas de secuencias. Las regiones SSRs pueden buscarse en estas bases de datos, con el objetivo de reducir el tiempo y costo requeridos para su desarrollo. En los últimos años, debido al rápido incremento de información de secuencias, la búsqueda de SSRs en ESTs de bases públicas se ha tornado una interesante alternativa. Esta metodología se ha utilizado en muchas especies incluyendo girasol (Paniego y col. 2002), pasto llorón (*Eragrostis*) (Cervigni y col. 2008), caña de azúcar (Cordeiro y col. 2001), trigo (Eujayl y col. 2001; Leigh y col. 2003; Yu y col. 2004; Zhang y col. 2005), cebada (Thiel y col. 2003; Chabane y col. 2005), centeno (Hackauf and Wehling, 2002), melón (Fernandez-Silva y col. 2008), roble (Ueno y col. 2008), palmera (Singh y col. 2008) y té (Sharma y col. 2009).

En *Eucalyptus*, hay algunos consorcios públicos y privados para el desarrollo de ESTs (Proyecto FORESTs), aunque la disponibilidad de información liberada al público es aún restringida. Ceresini y col. (2005), crearon una base de datos de SSRs derivados de ESTs y desarrollaron una herramienta bioinformática, denominada Satellyptus, para la búsqueda y análisis de microsatélites en la base de ESTs de *Eucalyptus*.

A su vez, Rabello y col. (2005) realizaron la caracterización *in silico* en 19 genotecas de ESTs de *Eucalyptus* de FORESTs, conteniendo secuencias ADNc de cinco especies: *E. grandis*, *E. globulus*, *E. saligna*, *E. urophylla* y *E. camaldulensis*.

Hasta el momento, sólo un número relativamente acotado de secuencias han sido anotadas en bancos de datos públicos. Una búsqueda de ESTs de *E. globulus* disponibles en GenBank (www.ncbi.nlm.nih.gov/dbEST/, actualizado a Junio 2007) registró sólo 3.953 secuencias depositadas. A fines de 2007, Rasmussen-Poblete y col. (2008) publicaron 8.737 secuencias ESTs provenientes de genotecas de ADNc de *E. globulus* sujetos a bajas temperaturas.

Rengel y col. (2009) publicaron una colección de ESTs de *Eucalyptus* (llamada EUCAWOOD <http://www.polebio.scsv.ups-tlse.fr/eucalyptus/eucawood/>) provenientes de una genoteca normalizada de diferenciación de xilema y cuatro sustractivas: xilema vs hojas, xilema vs floema, madera joven vs madura y viceversa. Los autores utilizaron esta base de datos para la búsqueda de regiones repetitivas, encontrando 639 SSR.

Por otro lado, Yasodha y col. (2008) publicaron un conjunto de EST-SSRs provenientes de cinco especies de *Eucalyptus* de la base pública de ESTs (www.ncbi.nlm.nih.gov/dbEST/, actualizado a Agosto 2006). Recientemente, se publicaron 20 marcadores SSRs di- y trinucleotídicos y otros 20 tetra- y pentanucleotídicos, derivados de una colección de ESTs de *Eucalyptus* (Faria y col. 2010 a, b).

1.5 Perspectivas funcionales de los SSR

Aunque los SSRs son usualmente considerados como marcadores de ADN evolutivamente neutros, en los últimos años diversos estudios mostraron que podrían tener significancia funcional en varios fenómenos biológicos (Li y col. 2002) (Figura 1.1).

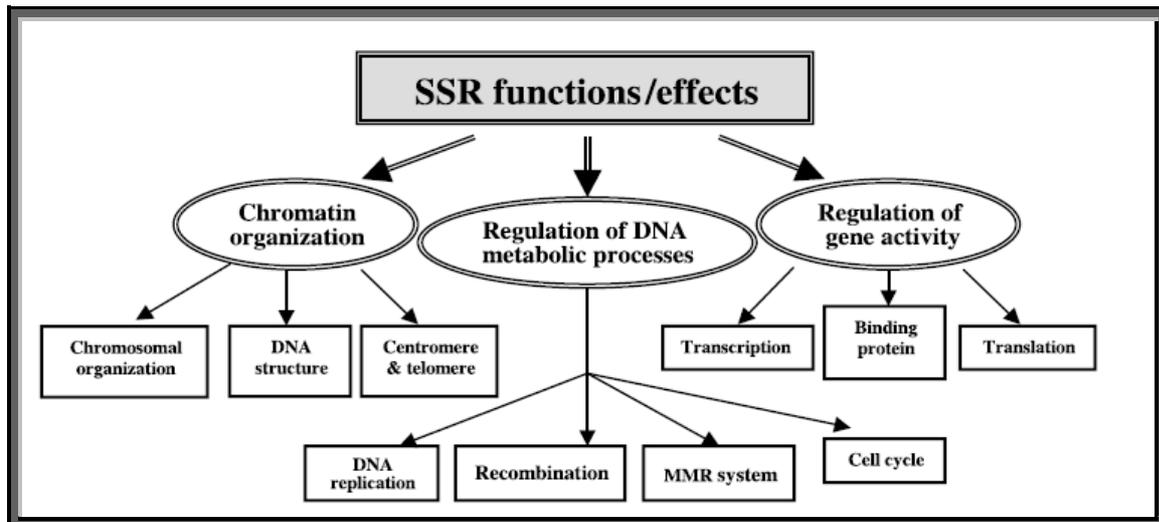


Figura 1.1: Función y efecto del cambio en el número de repeticiones en secuencias microsatélites. Extraído de (Li y col. 2002)

En los últimos años, una cantidad creciente de SSRs han sido caracterizados dentro de genes codificantes para proteínas y sus regiones no traducidas (UTRs). Éstos proveen información útil para el estudio de las posibles funciones que cumplen los SSRs. Existe evidencia que muestra que estos motivos no tienen una distribución al azar dentro de las regiones codificantes, UTRs e intrones dentro de los genes codificantes.

La expansión y contracción del tamaño de los SSRs (debido al cambio en el número de motivos repetitivos) en regiones codificantes para proteínas, podrían provocar la pérdida o ganancia de la función del gen, por medio de la mutación en el marco de lectura o la producción de ARNm tóxico. Las variaciones en los SSRs de las regiones 5'-UTRs podrían regular la expresión de los genes, afectando los procesos de transcripción y traducción a proteína. La expansión de SSR en regiones 3'-UTRs, podrían provocar cambios en la transcripción, produciendo una molécula de ARNm expandido que se podría acumular en el núcleo, alterando el empuje (o *splicing*) o exportación al citoplasma.

Los SSRs en intrones, también pueden afectar la transcripción de los genes, el empalme del ARNm o el transporte a citoplasma. Los SSRs trinucleotídicos localizados en UTRs o intrones también podrían inducir el silenciamiento de genes mediados por heterocromatina (Li y col. 2004).

Todos estos cambios podrían, eventualmente, provocar alteraciones en el fenotipo. Los SSRs dentro de genes están involucrados en los mismos procesos mutacionales que aquellos SSRs localizados en otras regiones genómicas, incluyendo mutaciones de punto y recombinación. Estos procesos de mutación generan cambios en el ADN que deben ser corregidos por el sistema de reparación del ADN. La mutación que no es corregida por este sistema, se convierte

en nuevo alelo del SSR y cambia el producto del gen, llevando a un eventual cambio en el fenotipo.

Así, los SSRs dentro de genes deberían estar sujetos a una mayor presión de selección que otras regiones genómicas por su importancia funcional. Estos SSRs podrían proveer la base molecular para una rápida adaptación a cambios medioambientales en organismos eucariotas y procariotas (Li y col. 2004).

En este contexto, la identificación y caracterización de regiones SSR dentro de genes, con el consiguiente desarrollo de marcadores funcionales serán de gran utilidad en el estudio de la variabilidad natural de genes candidatos, especialmente aquellos involucrados en calidad de madera, en las poblaciones de mejoramiento de *Eucalyptus*. Esto aportará conocimientos para la investigación de la base genética de la formación de la madera en aquellas características importantes para fines industriales y para la producción de bioenergía.

1.6 Composición de la madera y genes candidatos para calidad

Para fundamentar la selección de genes candidatos involucrados en calidad de madera, es necesario conocer cómo es el proceso de formación de ésta para definir los elementos biológicos que juegan un rol importante en dicho proceso.

Desde el punto de vista químico, la madera seca está constituida por celulosa (40-50%), hemicelulosa (25%) y lignina (25-35%), porcentajes que fluctúan entre especies y entre individuos de la misma especie.

La formación de la madera es un proceso complejo en el que las células del xilema y del *cambium* participan activamente en la síntesis de los metabolitos que forman parte de ella. En el proceso tienen lugar varias etapas sincronizadas: división celular, expansión y elongación de la célula, exudación de metabolitos y formación de pared secundaria y muerte celular.

En un corte transversal del tronco de un árbol se pueden observar al menos seis tipos distintos de madera: madera temprana, formada durante la primavera; madera tardía, formada durante el verano; madera juvenil, formada durante los primeros estados de desarrollo; madera adulta, formada durante el crecimiento después de alcanzar madurez sexual; madera de reacción y madera de compresión, ambas formadas en respuesta a la pérdida de la verticalidad del fuste. Cada una posee características propias y se manifiesta un proceso de crecimiento más lento durante la formación de madera tardía, que incluye una mayor acumulación de lignina (Plomion y col. 2001).

Las fluctuaciones en la formación de madera temprana y tardía redundan en las cualidades de productos para la industria papelera que incrementan la densidad de madera o cambian los

patrones de lignina. El contenido y calidad de ésta influye en el costo de su separación de la celulosa y hemicelulosa, en el proceso de producción de pasta para papel.

En las especies de *Eucalyptus*, la lignina representa entre el 20 y 30 % en peso de la madera, superando en algunos casos el 35% (Gion y col. 2001; Miranda y col. 2001) y su presencia influye negativamente en las propiedades de resistencia, brillo y envejecimiento del papel, además que resulta inhibitoria para los procesos de digestión enzimática de celulosa y posterior fermentación de la glucosa en la producción de etanol para biocombustible.

Debido a su mejor aptitud papelera, en los programas de mejoramiento forestal de *Eucalyptus globulus*, la selección de árboles con menor cantidad de lignina y/o lignina más fácilmente removible tendrá un impacto directo sobre el costo de producción de pasta, debido a la reducción de gastos en insumos químicos, energéticos y consumo de agua, que se utilizan para la extracción de la misma.

Por otro lado, el aumento de la densidad de la madera incide directamente en el aumento de la producción de pasta por unidad de volumen de procesamiento (productividad de los digestores). Desde una perspectiva molecular, el proceso de formación de la madera, involucra la expresión de un número de genes estructurales, coordinados por factores de transcripción, principalmente involucrados en la biosíntesis de polisacáridos (celulosa, hemicelulosa y pectinas), lignina y proteínas de la pared celular. Un gran número de ellos determina la composición de la pared celular y la forma que tienen las células. Los antecedentes que apoyan estas conclusiones se han obtenido por análisis bioquímicos clásicos, la aplicación de herramientas génicas (Le Provost y col. 2003), el estudio de perfiles de expresión proteica (Plomion y col. 2000) y el análisis de una gran colección de mutantes de *Arabidopsis thaliana* (Fagard y col. 2000).

Existen diversos estudios que analizan los genes expresados durante la formación de madera y, más particularmente, durante la xilogénesis (Egertsdotter y col. 2004; Foucart y col. 2006; Hertzberg y col. 2001; Israelsson y col. 2003; Paux y col. 2004; Paux y col. 2005; Yang y col. 2003; Yang y col. 2004).

Las rutas metabólicas que producen los componentes químicos de la madera, como la lignina o la celulosa, son las que cuentan con más investigación y resultan, por tanto, más conocidas, no obstante este conocimiento sigue siendo parcial. De hecho, los primeros datos genómicos disponibles, asociados a una característica de madera fueron justamente de los genes de la ruta metabólica de la lignina.

En *Eucalyptus*, se conocen varios genes estructurales y reguladores implicados en la biosíntesis de la lignina, incluyendo aquellos pertenecientes a la ruta del fenilpropanoico (PAL, C3H, C4H, COMT, CCoAOMT, y 4CL) (Gion y col. 2000; Thamarus y col. 2002), la ruta específica de los monolignoles (como CCR y CAD) (Gion y col. 2001; Poke y col. 2003), la ruta

metabólica del shikimato (CM, DAHP, SK), el metabolismo de la metionina (SAH, SAMS, HMT) (Kirst y col. 2004) (Figura 1.2). También fueron estudiados factores reguladores tales como el factor de transcripción MYB (Goicoechea y col. 2005) y LIM 1 (Kawaoka 2000; 2006), involucrados en la biosíntesis de éste polímero.

Todos estos genes parecen ser buenos candidatos para el desarrollo de nuevos marcadores moleculares funcionales que permitan realizar estudios de variabilidad genética.

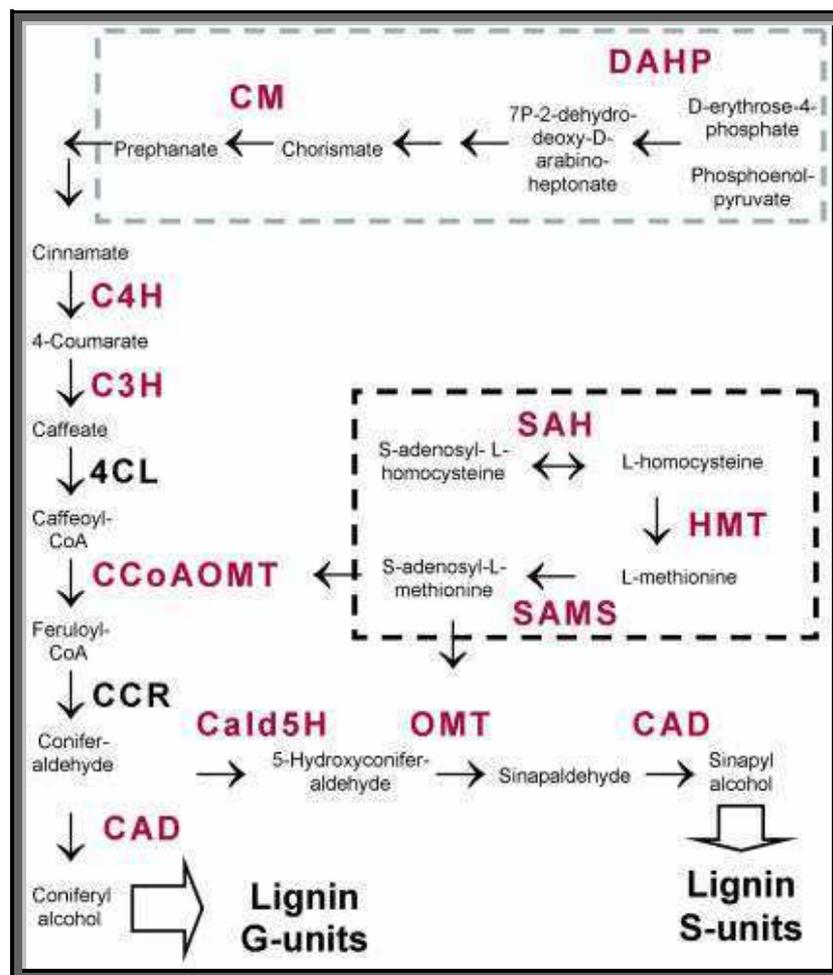


Figura 1.2: Ruta metabólica de síntesis de lignina. Incluye la ruta del fenilpropanoico y las rutas del shikimato (recuadro gris) y de la metionina (recuadro negro).
Fuente: Kirst. y col. (2004)

La exploración de la variabilidad natural de genes involucrados en rutas metabólicas como lignina y formación de la madera en *Eucalyptus globulus* no cuenta con precedentes en el país. La búsqueda de variantes alélicas de genes candidatos en poblaciones de mejoramiento de *Eucalyptus* es novedosa, tanto en la metodología, como en la inclusión de nuevos criterios de selección de árboles con destino a pasta de papel.

Debido a esto, el conocimiento de regiones genómicas involucradas en la densidad básica de madera y contenido/composición de lignina, permitirá a futuro una correcta selección asistida, generando mayor crecimiento del sector forestal y los segmentos productivos vinculados.

En el presente trabajo se propone desarrollar y caracterizar nuevos marcadores funcionales, haciendo énfasis en genes candidatos involucrados en características de madera, que permitan mediante estudios de asociación futuros, plantear nuevas estrategias para los actuales programas de mejoramiento de *E. globulus* que optimicen la calidad de madera. Principalmente se contemplará la variabilidad genética de marcadores funcionales relacionados a cantidad/composición de la lignina y densidad básica.

Hipótesis y objetivos

HIPÓTESIS

La identificación y caracterización de regiones SSRs dentro de genes, con el consiguiente desarrollo de marcadores funcionales serán de gran utilidad en el estudio de la variabilidad natural de genes candidatos, especialmente aquellos involucrados en calidad de madera, en las poblaciones de mejoramiento de *Eucalyptus*.

OBJETIVOS

OBJETIVO GENERAL

Descubrir, caracterizar y validar nuevas secuencias genómicas a través de análisis bioinformáticos para desarrollar marcadores moleculares funcionales que permitan medir variabilidad genética en poblaciones de *Eucalyptus globulus*, con énfasis en genes potencialmente involucrados en la cantidad/calidad de lignina, densidad básica de madera y otros caracteres de interés forestal.

OBJETIVOS ESPECÍFICOS

1. Explorar y analizar mediante herramientas bioinformáticas las colecciones de secuencias expresadas (ESTs) y secuencias de los principales genes involucrados en calidad de madera, depositadas en bancos de datos de dominio público de *Eucalyptus*, para la búsqueda de microsatélites no redundantes.
2. Diseñar y validar los marcadores funcionales encontrados.
3. Evaluar la variabilidad natural de dichos marcadores en los genotipos de *Eucalyptus globulus* seleccionados y su utilidad en estudios de diversidad genética poblacionales.
4. Analizar la posibilidad de transferibilidad de los marcadores funcionales desarrollados a otras especies del género *Eucalyptus*.

Materiales y Métodos

2. MATERIALES Y MÉTODOS

2.1 Material vegetal

El material vegetal utilizado en este estudio fue gentilmente cedido por la Unidad de Bosques cultivados del Instituto de Recursos Biológicos (IRB) de INTA Castelar. Se seleccionaron y analizaron 60 genotipos no emparentados (un individuo por familia), representantes de seis razas que cubren el rango de distribución natural de la especie (Dutkowski y Potts 1999) (Tabla 2.1a). Los mismos provienen de la población básica integrada por 276 familias, plantada entre los años 1995 y 1996 en cuatro sitios del sudeste de la provincia de Buenos Aires (Rodríguez Traverso 2000). Actualmente se cuenta con los valores del efecto genético aditivo de árbol individual (BLUP) (cuadrados medios mínimos familiares calculadas para cada individuo) de todas las variables de interés forestal, tales como crecimiento, forma, tiempo de transición a follaje adulto y grosor de la corteza y penetración de penetrómetro o pilodyn (López y col. 2001). Este último es una medida indirecta de la densidad básica de madera, debido a que están inversamente relacionadas (Greaves y col. 1996).

Para el análisis de validación de los marcadores moleculares desarrollados, se utilizó una muestra de ocho individuos, seleccionando uno de cada una de las razas (excepto Furneaux).

La muestra seleccionada está constituida por genotipos con valores genéticos individuales que cubren un amplio rango de densidad básica de madera. Se espera por un lado, que esto permita detectar variantes alélicas en los genes de las distintas rutas metabólicas involucradas en la formación de la madera y por otro, que la muestra sea representativa y permita una mejor estimación de la variabilidad de los genes funcionales detectados.

Para el análisis de transferibilidad de los marcadores moleculares diseñados se utilizaron entre siete y ocho genotipos de cada una de las otras seis especies de *Eucalyptus*: *E. grandis*, *E. saligna* (sección *Latoangulatae*); *E. dunnii*, *E. viminalis* (sección *Maidenaria*) y *E. camaldulensis*, *E. tereticornis* (sección *Exsertaria*) (Tabla 2.1 b). Para este estudio, el material vegetal fue gentilmente cedido por la Unidad de Bosques Cultivados del IRB de INTA Castelar y la Estación Experimental Agropecuaria Concordia (EEA INTA Concordia).

Tabla 2.1: Individuos de *Eucalyptus* utilizados en este estudio

Individuos de *E. globulus* (a) y otras seis especies de *Eucalyptus* (b) utilizados para el análisis de variabilidad y transferibilidad, respectivamente. Se detallan: razas de *E. globulus* (a) o especie (b) (primera columna), origen o procedencia (segunda columna), latitud (Lat., S), longitud (Long., E) y número de familias (un individuo por familia) en la última columna.

(a)

Raza	Origen/ Procedencia	Lat.	Long.	N
Eastern_Otways	Otways State Forest	38°36'	143°53'	5
Western_Otways	Otways National Park	38°47'	143°35'	12
NE_Tasmania	Seymour	41°40'	148°17'	3
S_Tasmania	Cradoc Hill	32°04'	138°29'	3
	Glendevie	43°14'	147°02'	2
SE_Tasmania	Moogara	42°47'	146°53'	4
Strzelecki	Jeeralang	38°20'	146°31'	17
Furneaux	Flinders Island	40°04'	148°00'	8
King_Island	King Island (TAS)	39°53'	143°59'	4
	Portugal OP			2
total				60

(b)

Especie	Origen/ Procedencia	Lat.	Long.	N
<i>E. grandis</i>	Woodum (QLD)	26°15'	152°44'	1
	Ring Tail S.F. (QLD)	26°19'	153°0'	1
	Kendall (NSW)	31°38'	152°42'	1
	South Africa			3
	Argentina, Concordia			2
<i>E. saligna</i>	Kangaroo Valley (NSW)	34°47'	150°34'	1
	Stewarts Brook (NSW)	32°00'	151°19'	3
	Watagan Mountains (NSW)	32°57'	151°14'	1
	origen desconocido			2
<i>E. dunnii</i>	Dead Horse Track (NSW)	36°10'	148°0'	5
	Oaky Creek (NSW)	31°1'	148°26'	1
	South Yabra (NSW)	28°36'	152°29'	1
	Argentina, Bs. As. Origen desconocido			1
<i>E. viminalis</i>	Bonang (VIC)	37°10'	148°53'	1
	Bonang River (VIC)	37°15'	148°45'	1
	Errinundra Road (VIC)	37°26'	148°52'	1
	Martins Creek (VIC)	37°28'	148°33'	1
	Mt. Baldhead (VIC)	37°26'	147°39'	2
	Noojee (VIC)	37°54'	145°58'	1
	Uriarra S.F. (NSW)	35°18'	148°55'	1
	Argentina, Bs. As. Origen desconocido			1
<i>E. camaldulensis</i>	Condamine (QLD)	26°56'	150°08'	1
	Gilgandra (NSW)	31°43'	148°39'	4
	Lake Albacutya (VIC)	35°45'	141°58'	2
	Argentina, Bs. As. Origen desconocido			1
<i>E. tereticornis</i>	N.of Taroom (QLD)	26°08'	149°57'	1
	Argentina, Oliveros. Origen desconocido			6
	Argentina, Bs. As. Origen desconocido			1
total				47

2.2 Extracción de ADN genómico

La extracción de ADN fue realizada utilizando un método derivado de Hoisington y col. (1994) con modificaciones menores para evitar la precipitación de aceites. El ADN se extrajo de hojas liofilizadas o frescas conservadas a $-70\text{ }^{\circ}\text{C}$. La liofilización se llevó a cabo a partir de tejido congelado en un liofilizador Labconco (EUA), durante 72 horas (a $-50\text{ }^{\circ}\text{C}$ y 1-5 micrones de mercurio de vacío). Las hojas de cada muestra luego fueron molidas por separado hasta convertirlas en un polvo muy fino. Se pesaron de 300 a 400 mg de este material en tubos de polipropileno de 15 ml al cual se agregaron 9 ml de solución de extracción CTAB (100 mM Tris-HCl, pH7,5; 10 mM EDTA, pH8; 700 mM NaCl; 140 mM 2-mecaptoetanol; 2% p/v bromuro de cetilmetilamonio (CTAB); 1% PVP) a 65°C . Se homogenizó cada muestra mediante inversiones suaves y se incubó durante 60-90 minutos a 65°C con movimiento continuo. Luego de dejarlas a temperatura ambiente durante 4-5 minutos, se agregaron 4,5 ml de cloroformo: octanol (24:1) seguido de varias inversiones suaves de los tubos (5 a 10 minutos). La separación de la fase acuosa fue realizada después de centrifugar los tubos durante 10 minutos en una centrífuga a 1300 xg. Luego, ésta fue trasvasada a un nuevo tubo y se repitió la extracción con cloroformo: octanol una segunda vez, transvasando la fase acuosa a un tubo de 50 ml. Finalmente, el ADN fue precipitado al agregar 2,5 volúmenes de etanol absoluto a 0°C y 800 μl de una solución de acetato de sodio 3M. El ADN precipitado fue rescatado utilizando un capilar de vidrio, siendo luego resuspendido en 1 ml de TE 1X (10 mM Tris-HCl, pH8; 1 mM EDTA, pH8). El ARN de la muestra fue eliminado mediante la incubación con 2 μl de ARNasa A (10 mg/ml) durante 30 minutos a temperatura ambiente.

Con el fin de obtener muestras con un elevado grado de pureza se agregaron a las mismas 400 μl de fenol-cloroformo-isoamílico (25:24:1) saturado en Tris-HCl (100mM) pH=8, se mezcló por inversión durante 10 minutos y luego se centrifugó a 1.300 xg durante el mismo tiempo. La fase acuosa fue recuperada y sometida a una segunda extracción orgánica que se llevó a cabo con 800 μl de cloroformo-octanol (24:1), repitiendo los pasos de mezcla y centrifugación. Finalmente, el ADN fue reprecipitado con el agregado de 1,2-1,5 ml de etanol frío y 80 μl de acetato de sodio (NaAc) (3M) hasta una concentración aproximada de 0,18 M. El ADN obtenido fue separado y resuspendido en 100 μl de TE 1X hasta concentraciones finales entre 100 y 500 ng/ μl .

2.3 Cuantificación de ADN

La concentración e integridad del ADN se determinó por medio de visualización con luz U.V., luego de la electroforesis en geles de agarosa (0,8% en TAE 1x -0,04 M Tris-acetato, 0,001M EDTA-) teñidos con bromuro de etidio. Todas las lecturas se refirieron a un patrón de concentración de ADN conocido y se ajustaron a 50 ng/μl en TE 1x. Las muestras de ADN obtenidas fueron almacenadas, hasta su posterior utilización, a una temperatura de 4 °C.

2.4 Análisis bioinformático para la identificación de regiones microsatélites en secuencias expresadas etiquetadas (ESTs)

Con el objetivo de detectar regiones microsatélites en la colección de ESTs de *Eucalyptus globulus* de dominio público, se analizaron 12690 secuencias publicadas en GenBank (actualizadas a Octubre de 2008), provenientes de 5 genotecas de:

- ADNc sustraído de *cambium* (Bossinger y Leitch 2000)
- ADNc de plantas con sometimiento a bajas temperaturas (Rasmussen-Poblete y col. 2008)
- ADNc sustraído de plantas con tratamientos de gravedad versus microgravedad (Ladouce y col. *no publicado*)
- ADNc de diferenciación de xilema (Rengel y col. 2009)
- ADNc de interacción *E. globulus*-ectomicorriza (*Eucalyptus globulus-Pisolithus tinctorius*) de 4 días (Voiblet y col. 2001)

2.4.1 Detección de unigenes

Se analizó la redundancia de secuencias de ESTs de la base de datos, mediante el programa EGassembler (Masoudi-Nejad y col. 2006, <http://egassembler.hgc.jp/>). Las secuencias que mostraron más del 85% de identidad, en tramos de 30 nucleótidos, fueron consideradas idénticas y se asignaron a un grupo específico ("contig"), el resto fueron consideradas secuencias únicas ("singletons").

2.4.2 Determinación de secuencias microsatélites (SSR) en unigenes

Se realizó la búsqueda de regiones microsatélites en las secuencias no redundantes obtenidas, mediante la utilización del servidor GDR SSR (*Genome Database for Rosaceae*; <http://www.rosaceae.org/bio/content?title=&url=/cgi->

[bin/gdr/gdr_ssr&style=height:1024px;width:950px;](#)). Se consideraron regiones microsatélites a aquellas secuencias con un mínimo de siete repeticiones para motivos dinucleotídicos, cuatro repeticiones para motivos trinucleotídicos y tres repeticiones para motivos tetra- y pentanucleotídicos.

En este análisis se consideraron las secuencias que contenían microsatélites perfectos, los cuales se caracterizan por tener un motivo de repetición único y sin interrupciones en la secuencia en tándem, y microsatélites compuestos, con más de un motivo de repetición en la misma secuencia.

Al momento de estimar las frecuencias fueron consideradas todas las posibles combinaciones de los cuatro nucleótidos. Los motivos de los microsatélites están representados por 4 dinucleótidos diferentes (AT)_n, (AG)_n, (AC)_n, (GC)_n. De la misma manera, existen 10 motivos trinucleotídicos, 33 tetranucleotídicos y 102 motivos pentanucleotídicos diferentes. Así, todos estos motivos pueden estar presentes de diversas formas a partir del mismo conjunto básico o su reverso complementario (Jurka y Pethiyagoda 1995; Katti y col. 2001). Por ejemplo, los diferentes motivos de los di- y trinucleótidos están representados en la Tabla 2.2

Tabla 2.2. Posibles combinaciones de los di- y trinucleótidos

Motivos dinucleótidos	Motivos trinucleótidos
AT/TA	AAT/ATA/TAA/ATT/TTA/TAT
AG/GA/CT/TC	AAG/AGA/GAA/CTT/TTC/TCT
AC/CA/TG/GT	AAC/ACA/CAA/GTT/TTG/TGT
GC/CG	ATG/TGA/GAT/CAT/ATC/TCA
	AGT/GTA/TAG/ACT/CTA/TCA
	AGG/GGA/GAG/CCT/CTC/TCC
	AGC/GCA/CAG/GCT/CTG/TGC
	ACC/CCA/CAC/GGT/GTG/TGG
	ACC/CCA/CAC/GGT/GTG/TGG
	GGC/GCG/CGG/GCC/CCG/CGC

Se diseñaron todos los oligonucleótidos para amplificar las regiones microsatélites encontradas mediante el software Primer3 (v. 0.4.0) (<http://frodo.wi.mit.edu/primer3/>), a través del servidor GDR SSR. Dicho programa excluye a aquellas regiones que se encontraron cercanas a los

extremos 5' o 3' y no dejan una longitud de secuencia flanqueante suficiente para el diseño de los mismos.

La predicción de marcos de lectura abiertos (ORF) se realizó también a través del programa GDR SSR. Este selecciona, a través de getORF EMBOSS (http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf?_pref_hide_optional=0), el ORF más largo como región codificante putativa.

2.4.3 Determinación de posibles funciones de los unigenes que contienen microsatélites (EST-SSRs)

Se utilizó el programa Blast2Go (Conesa y col. 2005; <http://www.blast2go.de/>), el cual describe los productos de los genes, en términos de: proceso biológico asociado (P), componente celular (C) y función molecular (F). Las anotaciones Gene Ontology (GO) se generaron en un proceso de 3 pasos:

- 1) SIMILITUD BLAST: se buscó similitud de secuencias con la base de datos (GenBank) no redundante (nr) de proteínas, utilizando un e-valor de corte de $E \leq 10^{-5}$.
- 2) MAPEO: se utilizaron los resultados del análisis de similitud BLASTX para obtener los términos *Gene Ontology* (GO) (<http://www.geneontology.org/>) mediante archivos de anotación provistos por el Consorcio GO que se incorporaron a la base de datos del servidor Blast2GO.
- 3) ANOTACIÓN: las secuencias se anotaron usando parámetros provistos por el usuario.

Los términos de InterPro se obtuvieron del servidor InterProScan (análisis funcional de proteínas) del EBI (European Bioinformatics Institute), los cuales pudieron ser convertidos y combinados con los términos GO.

Por último, el programa permitió realizar las anotaciones *Evidence Code* (EC) y de *Kyoto Encyclopedia of Genes and Genomes* (KEGG; <http://www.genome.jp/>), generadas a partir de GO.

El programa incluye varias herramientas visuales como: código de color de las secuencias, visualización gráfica de rutas metabólicas de KEGG, y gráficos estadísticos de términos GO (Figura 2.1)

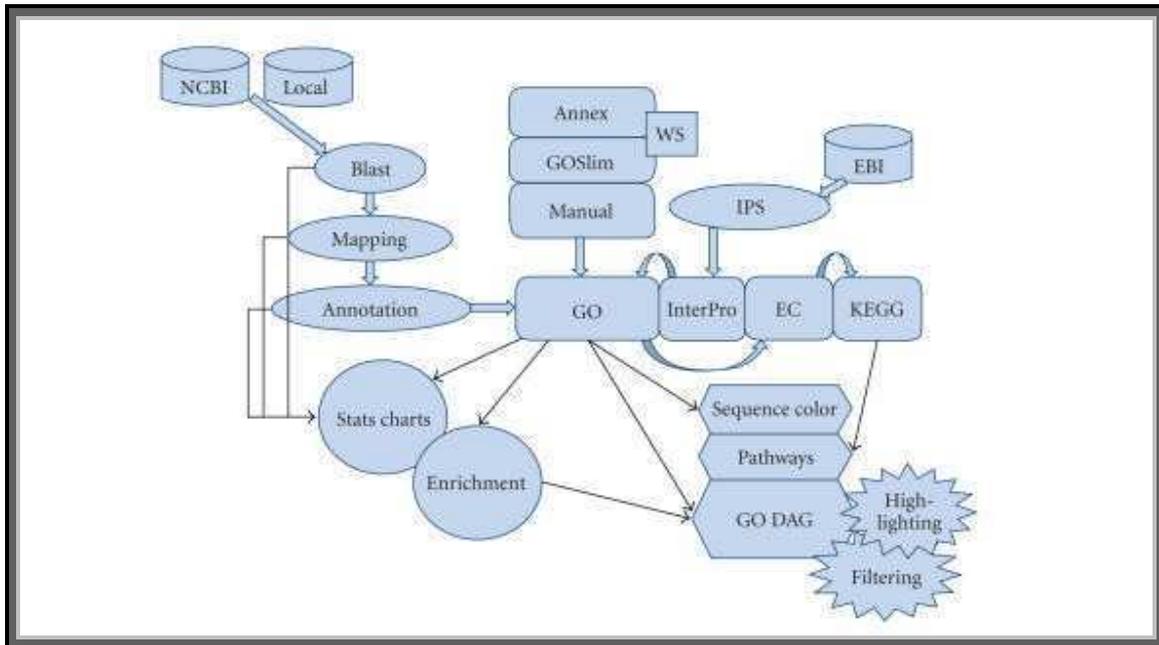


Figura 2.1: Representación esquemática de las aplicaciones del programa Blast2GO.

Fuente: Blast2GO (<http://www.blast2go.de/>)

2.5 Selección de genes candidatos e identificación de regiones microsatélites

La selección de genes candidatos para calidad de madera e identificación de SSR se realizó a partir de dos estrategias:

1. Se utilizó la asignación de posibles funciones de los unigenes que contienen microsatélites determinados a partir de los ESTs analizados del banco de secuencias (sección 2.4.1 de MyM), mediante programa Blast2GO.
2. Se estudiaron los genes mencionados en la bibliografía existente: Gion y col. (2000) y Thamarus y col. (2002) estudiaron genes pertenecientes a la ruta del fenilpropanoico (PAL, C3H, C4H, COMT, CCoAOMT, y 4CL); Kirst y col. (2004) analizaron también la expresión de estos genes y algunos pertenecientes a la ruta del shikimato (CM y DAHP) y del metabolismo de la metionina (SAH, HMT y SAMS). Los genes de la ruta específica de los monolignoles (como CCR y CAD) fueron estudiados por Poke y col. (2003), Gion y col. (2001) y Kirst y col. (2004). A su vez, algunos trabajos demuestran que existen factores reguladores tales como el factor de transcripción MYB (Goicoechea y col. 2005; Patzlaff y col. 2003) y LIM 1 (Kawaoka y col. 2000) que regulan la lignificación. Otros genes ampliamente estudiados son los involucrados en la síntesis de celulosa y hemicelulosa. Se han identificado familias multigénicas que codifican para las enzimas celulosa sintasa

(CesA) (Saxena y col 2005) y xiloglucano endotransglucosilasa-hidrolasa (XTH) las cuales cumplen una amplia gama de funciones, tanto en fenómenos de elongación y expansión celular, como en fenómenos de desarrollo que llevan a cambios importantes en la estructura de la pared, que debe ser degradada, sintetizada o reforzada (Kimura y Kondo 2002; Genovesi y col. 2008). Se realizó la búsqueda de secuencias de estos genes candidatos para calidad de madera en banco de genes de acceso público (<http://www.ncbi.nlm.nih.gov/>). Se analizó la analogía de la secuencia de cada gen candidato contra la base de datos del GenBank, en busca de regiones variables, mediante el algoritmo BIASTN y BLASTX (<http://www.ncbi.nlm.nih.gov/BLAST/>). Adicionalmente se utilizó una combinación de los programas *SSR Mining software* (localizado en GDR Server, http://www.rosaceae.org/bio/content?title=&url=/cgi-bin/gdr/gdr_ssr&style=height:1024px;width:950px;), Simple Sequence Repeat Identification Tool (SSRIT) software y la identificación manual por análisis visual de secuencias.

Luego de la elección de la región a ser amplificada, se diseñaron oligonucleótidos específicos mediante el programa Primer3 (<http://frodo.wi.mit.edu/primer3/>).

2.6 Amplificación por PCR de las regiones microsatélites

Los oligonucleótidos fueron sintetizados por un servicio externo (Alpha DNA, Canadá). Se realizaron reacciones de amplificación mediante reacción en cadena de la polimerasa (PCR).

Las condiciones utilizadas fueron las siguientes:

En 10 µl finales:

0,25 µM de cada cebador 0,2 mM dNTP 1,5 a 2 mM MgCl ₂ 1X buffer 1u Platinum Taq Polymerasa (Invitrogen, EUA) 20 ng de ADN

Para el análisis de transferibilidad de marcadores SSRs a otras especies de *Eucalyptus*, se utilizó una concentración final de MgCl₂ de 3 mM

El programa básico fue el siguiente: 5 min a 94° C, 35 ciclos de 1 min a 94° C, 1 min a temperatura de amplificación, 1 min 72° C y una extensión final de 10 min a 72° C.

Las temperaturas de amplificación de cada oligonucleótido se establecieron tomando como referencia las temperaturas proporcionadas por los programas de diseño utilizados, con la posterior puesta a punto de las diferentes condiciones de amplificación (Tabla 3.2 y Tabla 3.5).

2.7 Electroforesis de los fragmentos de amplificación

Para el análisis de validación de los marcadores moleculares encontrados *in silico*, se realizó la amplificación de cada fragmento en ocho individuos de *E. globulus*. A cada reacción de PCR, se les agregó 3 µl de buffer de carga con ambos colorantes (80% formamida, 10 mM EDTA (pH 8,0), 1 mg/ml azul de bromofenol, 1 mg/ml xileno cianol), antes de ser corridas electroforéticamente.

Las reacciones se calentaron a 95°C durante 3 minutos y luego se las colocó inmediatamente en hielo.

2.7.1 Visualización de ADN en geles de poliacrilamida

Los productos de reacción fueron analizados en geles desnaturalizantes de poliacrilamida al 6% (6% acrilamida / bisacrilamida 20:1, 7,5 M urea, 0,5 x TBE) de 0,4 mm de espesor. La corrida se realizó en buffer TBE 0,5x (cuba superior) y TBE 1x (cuba inferior) (TBE1x: 100 mM Tris / 100 mM ácido bórico / 2 mM EDTA), durante 3- 4 hs a 40 W- 50 W constantes en el sistema de Gibco- BRL Modelo S2 (EUA).

La tinción de geles de secuenciación con nitrato de plata se realizó con el sistema de Promega Biotech (EUA) "DNA staining kit", el cual consta de los siguientes pasos:

1. fijación del gel en ácido acético al 10% durante 20 min.
2. 3 lavados en agua bidestilada durante 2 min cada vez.
3. tinción con nitrato de plata al 1% y formaldehído 1,5 ml/litro de solución durante 30 min.
4. lavado de 3 seg en agua bidestilada fría.
5. revelado en solución de carbonato de sodio 30 g/l, formaldehído 1,5 ml/l y tiosulfato de sodio (10 mg/ml) 200 µl/l, hasta obtener coloración en las bandas.
6. detención de la reacción con ácido acético al 10% durante 2 min.
7. 3 lavados con agua bidestilada durante 2 min cada vez.
8. secado a temperatura ambiente.

Los patrones de bandas obtenidos en los vidrios fueron analizados directamente sobre un transiluminador de luz blanca. El tamaño de los alelos correspondiente a cada marcador se estableció en referencia al marcador de tamaño molecular de 25 pb (Invitrogen, EUA).

2.7.2 Métodos automáticos

Para realizar los estudios de polimorfismo de los marcadores moleculares en la población de 60 individuos de *E. globulus*, se automatizó el proceso de genotipificación mediante técnicas basadas en electroforesis capilar (EC) usando oligonucleótidos marcados con grupos fluorescentes. Los fragmentos se evaluaron por EC utilizando un analizador de ADN (ABI 3130xl, Applied Biosystems, EUA). Los oligonucleótidos sentido se marcaron con los siguientes fluoróforos:

Nombre del fluoróforo	Color emitido	Emisión de fluorescencia
6-FAM	azul	535
HEX	verde	560

El filtro utilizado en el equipo reserva el color rojo (ROX) para el estándar interno de peso molecular el cual acompaña la corrida electroforética. Para los fragmentos de 50 a 500 pb, el estándar de peso molecular fue GeneScan ROX500 (-250), tiene 15 fragmentos de: 35, 50, 75, 100, 139, 150, 160, 200, 300, 340, 350, 400, 450, 490 y 500 pb. El fragmento de 35 pb debe corregirse de manera manual en el análisis, ya que migra anormalmente y distorsiona los resultados de la curva estándar (Figura 2.2).

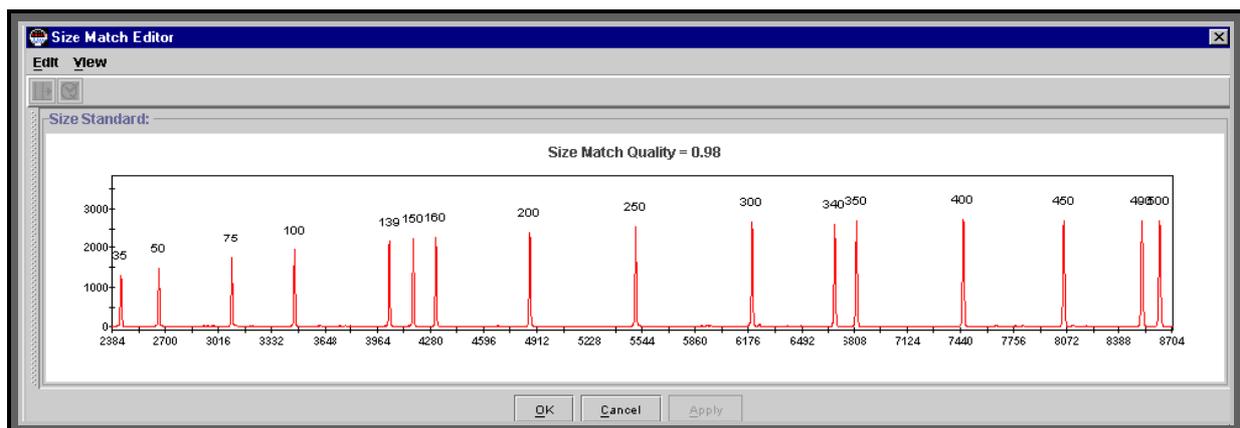


Figura 2.2 Fragmentos del estándar de tamaño molecular.

2.7.3 Marcación directa

Para la marcación directa los oligonucleótidos sentido se utilizaron los fluoróforos 6-FAM y HEX, como se mencionó previamente.

Para utilizar al máximo las posibilidades del sistema, ampliar la capacidad de procesamiento de muestras y hacerlo más económico, se tuvieron en cuenta dos aspectos importantes para definir la marcación de los oligonucleótidos y la preparación de las muestras para la electroforesis para una colección de microsatélites conocidos. Así, se utilizaron en cada panel dos marcadores compatibles, para aumentar la eficiencia de discriminación entre alelos en una sola corrida. Para ello, se usaron dos colores de fluoróforos diferentes (HEX, 6-FAM) para el diseño de los oligonucleótidos a ensayar y se combinaron los colores con los diferentes tamaños en pares de bases (pb) de los fragmentos amplificados, evitando la superposición de fragmentos del mismo tamaño y del mismo color.

En todos los casos, antes de preparar los paneles de co-electroforesis, los fragmentos se evaluaron en geles de agarosa observados mediante el escáner Typhoon trio (Amersham Biosciences, RU) incidiendo el láser a la longitud de onda correspondiente a cada fluoróforo y utilizando los filtros 520 BP40, 555 BP20 y 605 BP30 para los fluoróforos FAM, HEX y bromuro de etidio, respectivamente. Las imágenes se analizaron mediante el programa ImagenQuant (Amersham Biosciences, RU). Con las imágenes obtenidas se ajustaron las diluciones correspondientes, ya que es muy importante probar que la intensidad de la señal para los distintos fluorocromos se encuentre dentro del rango de fluorescencia adecuado (entre 200 y 4.000 RFU) y además de que no haya superposiciones de tamaños de fragmentos marcados con el mismo fluoróforo.

Para el protocolo de preparación de las muestras para su resolución por EC se realizó una optimización de la corrida con pocas muestras en una primera etapa y luego, en una segunda etapa, se realizó el ensayo con el total de las muestras a analizar.

Los fragmentos resueltos por EC fueron analizados de manera automática utilizando el programa GeneMapper® 4.0 (Applied Biosystem, EUA).

2.8 Purificación de fragmentos de ADN de geles de poliacrilamida

Una vez corridas las muestras de ADN mediante la técnica de electroforesis en geles de poliacrilamida e identificadas mediante su tinción con solución de nitrato de plata, se aislaron

las bandas de interés a partir del gel. Este procedimiento se realizó con los diferentes alelos del *locus* HMTi que presentaron mayores tamaños a los esperados. Cada fragmento de gel conteniendo un alelo fue resuspendido en 25 µl de TE 1X, desnaturalizando la muestra a 94°C durante 5 minutos para favorecer el proceso. Dicha solución fue utilizada como templado en la siguiente reacción de PCR, con el objetivo de obtener una mayor cantidad de producto de amplificación. Finalmente, los fragmentos se purificaron directamente a partir de la reacción de PCR utilizando el sistema comercial Qiaquick PCR Purification Kit (Qiagen, Alemania). Los mismos fueron secuenciados automáticamente utilizando un analizador de ADN (ABI 3130xl, Applied Biosystems, EUA).

2.9 Análisis de microsatélites usando el programa GeneMapper® 4.0 (Applied Biosystems, EUA)

El primer paso en el análisis de microsatélites usando el programa GeneMapper® 4.0 (Applied Biosystems, EUA) fue la creación de un proyecto dentro del cual se definió un grupo de paneles llamado “kit”, donde cada “panel” corresponde al grupo de marcadores que componen el conjunto de co-electroforesis definido previamente a la corrida por EC. Para cada uno de estos marcadores se completó una tabla conteniendo el nombre, la longitud del fragmento en pb, el color del fluoróforo y el tipo de motivo de repetición.

A continuación se seleccionaron los parámetros de análisis para el proyecto. Para esto, inicialmente se definió el método de análisis y el estándar de peso molecular. A partir de allí, y trabajando sobre un “panel” determinado, se corroboró que la asignación automática de tamaños a los picos del estándar de peso molecular sea correcta. En el caso de encontrar discrepancias, las asignaciones se corrigieron manualmente y se reanalizaron los datos antes de continuar el análisis. El programa brindó, a su vez, indicadores de calidad de las secuencias (SQ), visibles como etiquetas de colores asignadas a diferentes parámetros.

Luego, se identificaron los alelos correspondientes a cada marcador y se definió un “bin” (región del espectro definida para registrar la señal de un alelo) por alelo dentro del proyecto creado inicialmente. Cada bin entonces representó un alelo del marcador con dos atributos: el tamaño del fragmento en pb y el color del fluoróforo con el cual se marcó el oligonucleótido sentido correspondiente al marcador analizado (Figura 2.3).

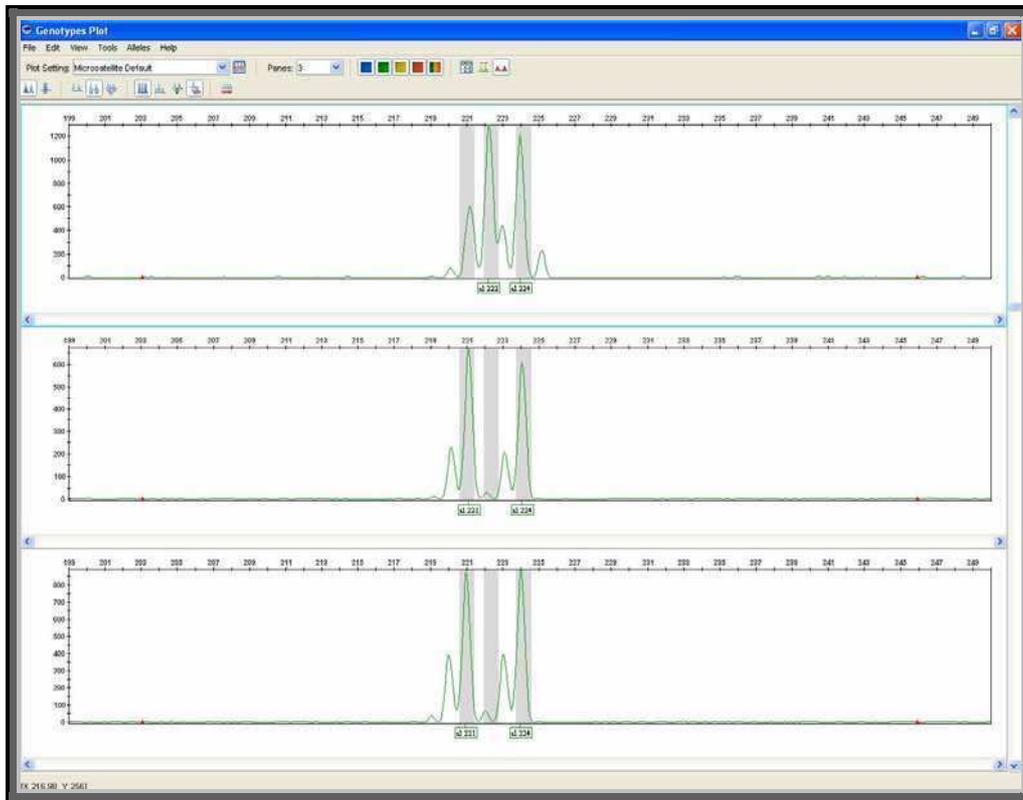


Figura 2.3 Alelos del *locus* HMTe indicados como picos para la muestra seleccionada y los rangos de tamaños (en pb) del marcador seleccionado. Las barras en gris representan los bins.

Para examinar los resultados y evaluar de manera conjunta los genotipos analizados en el proyecto, se utilizó la función *Genotype Plot* del programa, la cual permite visualizar los resultados de hasta ocho individuos simultáneamente.

Una vez considerados los criterios de calidad descritos para cada corrida y, en los casos necesarios, habiendo realizado las ediciones correspondientes, los genotipos asignados por el programa pudieron ser aceptados como confiables para continuar con el análisis de los mismos. Las tablas se exportaron como archivos de texto para su posterior utilización.

2.10 Análisis estadístico de los datos obtenidos

2.10.1 Generación de la matriz básica de datos, matriz de distancia y análisis de agrupamiento

Cada fragmento de amplificación (alelo) tiene una movilidad electroforética determinada. Ésta fue observada por medio de revelado con nitrato de plata o mediante genotipificación en

secuenciador automático. El tamaño de cada alelo fue medido en pares de bases (pb). Teniendo en cuenta esta codificación, se construyeron las matrices básicas para el conjunto de datos de SSRs utilizando el programa Excel (Microsoft, EUA). La matriz básica de datos (MBD), generada para el programa GenAIEx 6.2 (Peakall y Smouse 2006), representa a todos los individuos estudiados (OTUs: unidad taxonómica operacional) en las filas y las variables encontradas en las columnas (alelos). Los datos codominantes se presentan en dos columnas por *locus* como en la figura 2.4. Los alelos se codificaron como tamaño total en pares de bases (pb).

3	ind	pobl	pCCR		CesA3_A	
4	Ec2	camaldulensis	186	188	290	298
5	Ec3	camaldulensis	182	231	289	289
6	Ec4	camaldulensis	170	170	282	282
7	Ec5	camaldulensis	182	231	289	297
8	Ec6	camaldulensis	188	231	293	300
9	Ec7	camaldulensis	203	231	289	289
10	Ec8	camaldulensis	206	231	298	312
11	Ed1	dunnii	169	182	290	294
12	Ed2	dunnii	169	169	294	294
13	Ed3	dunnii	170	170	299	300
14	Ed4	dunnii	0	0	0	0
15	Ed5	dunnii	182	200	290	296
16	Ed6	dunnii	0	0	0	0
17	Ed7	dunnii	169	182	0	0
18	Ed8	dunnii	170	170	296	300

Figura 2.4: Ejemplo de datos codominantes de microsatélites, con alelos codificados como tamaño de fragmento para los marcadores pCCR y CesA3 A. En la columna 1 se identifican los individuos, en la columna 2 a las especies/poblaciones de las que provienen y en las restantes se indican los dos posibles alelos detectados para cada marcador.

Los casos de bandas dudosas o datos perdidos fueron codificados como 0 (cero).

La presencia de sólo un alelo de un microsatélite dado, fue considerada como un estado de homocigosis del alelo. Si bien la hipótesis de ausencia de alelos nulos no pudo ser confirmada debido a que no fueron realizadas pruebas de progenie, se realizaron estimaciones para estos valores y se muestran en la sección 3.3.1.1.

La Matriz de Distancia se construyó a partir de la Matriz Básica de Datos (con alelos re-codificados expresados en números del 01 al 99), mediante el programa *Populations* 1.2.30, utilizando un coeficiente de distancia DAS (shared allele distance) entre individuos (Chakraborty y Jin. 1993). Este índice DAS toma el mínimo valor de 0 cuando los individuos comparten todos los alelos y un valor máximo de 1, cuando los individuos no comparten alelos y se calcula mediante la siguiente fórmula: $D_{SAI} = 1 - P_{SAI}$

$$\text{siendo } P_{SAI} = \frac{\sum us}{2u}$$

donde s es la suma de todos los alelos en común en todos los *loci* u.

El Coeficiente de Distancia utilizado ofrece resultados similares al complemento del Coeficiente de Similitud de Jaccard (1908) más conocido corrientemente.

Se utilizó la técnica de SAHN (*Sequential-Agglomerative-Hierarchic-Nonoverlapping*), (Sneath y Sokal 1973), que agrupa las OTUs en forma jerárquica y secuencial, sobre la base de los niveles de distancia. El método de ligamiento utilizado fue UPGMA (Sokal y Michener 1958) que considera que el valor de distancia entre la OTU candidata a incorporarse y el grupo o núcleo es igual a una distancia promedio, resultante de los valores de distancia entre el candidato y cada uno de los integrantes del grupo o núcleo. El ligamiento promedio tiende a juntar grupos con poca varianza y producir grupos de varianza semejante. Para este análisis se utilizó el programa NTSys pc v2.0.

Los agrupamientos de individuos generados por el método de ligamiento descrito, originan un diagrama arborescente denominado dendrograma (representación gráfica de una clasificación jerárquica).

2.10.2 Estimadores del potencial de los marcadores para análisis de diversidad genética

El grado de polimorfismo medido para cada *locus* microsatélite evaluado fue cuantificado por medio del Índice de contenido polimórfico (PIC), según la fórmula de Botstein y col. (1980):

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

donde p_i es la frecuencia de cada alelo en el conjunto de alelos.

Otros estadísticos de cada marcador fueron analizados mediante el programa GenAlEx 6.2 (Peakall y Smouse 2006):

- número de alelos (N_a)
- número efectivo de alelos (N_e): $N_e = 1 / \sum_{i=1}^n p_i^2$
- Heterocigosidad observada (H_o)

- Heterocigosidad esperada insesgada (UH_e) (Nei 1987), corregida para el tamaño de cada muestreo (N) : $UH_e = \frac{2N}{2N-1} H_e$; donde $H_e = 1 - \sum_{i=1}^n p_i^2$
- Índice de Fijación (FI) (Wright 1965): $FI = 1 - \frac{H_o}{H_e}$
- Probabilidad de Identidad (PI): $PI = 2\left(\sum_{i=1}^n p_i^2\right)^2 - \sum_{i=1}^n p_i^4$

2.10.3 Equilibrio de Hardy-Weinberg, frecuencia de alelos nulos y estructura genética

Se realizaron las pruebas de equilibrio de Hardy-Weinberg (EHW) para cada *locus* mediante el método de cadenas de Markov, utilizando el programa GENEPOP 4.0.10 (Raymond y Rousset 1995). Las cadenas de Markov se basaron en una dememorización de 10000, 10000 batches y 10000 iteraciones por conjunto (*batch*).

La frecuencia de alelos nulos fue estimada mediante el programa INEST 1.0 (Chybicki y Burczyk 2009) que es adecuado para esta especie sujeta a un modelo de apareamiento mixto. Además, este programa estima el fracaso de la amplificación por PCR, a través del parámetro beta (β).

Se realizó el test de neutralidad de Ewens-Watterson de los *loci*, según el algoritmo de Manly y col. (1985), utilizando el programa Popgene versión 1.32 (<http://www.ualberta.ca/~fyeh/>).

Este test indica que puede haber presión selectiva sobre las variantes alélicas presentes en cada uno de los *loci* cuando los resultados son significativos. Se consideró significativo cuando el valor medio E-W está fuera del intervalo de confianza ($\alpha = 0,001$).

Para evaluar a los microsatélites seleccionados por su capacidad de diferenciación de grupos dentro de cada muestra (es decir, la población total de individuos de *E. globulus* y por otro lado el conjunto de especies seleccionadas para el estudio de transferibilidad), los individuos fueron asignados probabilísticamente a un número determinado de grupos (clusters) inferidos con un enfoque Bayesiano y los métodos de Monte Carlo basados en cadenas de Markov (MCMC: *Markov Chain Monte Carlo*). Se considera que el conjunto de individuos analizados representa una mezcla proveniente de un número desconocido de poblaciones (K), por ende se determinará un valor de K que maximice la probabilidad de los datos. De este modo se logra detectar el patrón de estructuración que muestran los materiales analizados. Para ello, se generó una cadena de Markov simple de 150000 ciclos. Las primeras 50000 iteraciones ("Burn-

in”) fueron descartadas, utilizando un modelo *admixture*, sin pre-adjudicación de estructura genética (para la población de 60 individuos de *E. globulus*) y con y sin pre-adjudicación de estructura genética (para el conjunto de siete especies de *Eucalyptus*), 10 corridas independientes para cada k, con k variables de 1 a 15. El valor más probable de k se eligió evaluando el estadístico Δk definido por Evanno et al. (2005) y la relación entre el logaritmo promedio de las probabilidades ($\ln(P)D$) en función de K. Este método fue implementado por el software STRUCTURE (Pritchard y col. 2000). El mismo se llevó a cabo utilizando los recursos de la Unidad de Servicios de Biología Computacional de la Universidad de Cornell (<http://cbsuapps.tc.cornell.edu/>). Según la documentación del programa Structure, el valor de K más apropiado es aquel más bajo capaz de capturar la mayor estructuración en el conjunto total de datos. Es decir, se deberá seleccionar aquel valor de K que presente la más alta probabilidad o mayor $\ln(P)D$ promedio, sin incorporar estructura ficticia. Los resultados fueron graficados mediante excel para proporcionar una visualización gráfica de la estructura de la población.

Se realizó además Análisis de la Varianza Molecular (AMOVA) mediante el programa GenAlEx 6.2 (Peakall y Smouse 2006) para estimar la distribución de la variabilidad genética entre y dentro de especies. Se calcularon los estadísticos PhiPT y se realizaron 9999 permutaciones para estimar su significancia.

Resultados y Discusión

Capítulo I

Búsqueda, caracterización funcional *in silico* y validación experimental de marcadores EST-SSRs en *Eucalyptus globulus*

3. RESULTADOS Y DISCUSIÓN

3.1 Capítulo I Búsqueda, caracterización funcional *in silico* y validación experimental de marcadores EST-SSRs en *Eucalyptus globulus*

3.1.1 Ensamblaje y definición de unigenes

Se evaluaron 12690 secuencias ESTs de *Eucalyptus globulus* depositadas en el banco de datos del GenBank (actualizado a Diciembre de 2008).

Luego del “clustering” y ensamblaje se obtuvieron 4924 secuencias únicas o unigenes, entre los que se encontraron 1495 *contigs* (Cg) y 3429 *singletons* (Sg). Estos resultados revelaron una redundancia de información de 2,6 veces en la base de datos y una relación Sg/Cg de 2,29. En un estudio similar, Rengel y col. (2009) realizaron el ensamblaje de 9.222 secuencias EST provenientes de 10.000 ADNc, del banco de datos de EucaWood (<http://www.polebio.scsv.upstlse.fr/eucalyptus/eucaWood/>). En el citado trabajo, se obtuvieron 3857 unigenes, que incluyeron 2.461 Cg y 1.396 Sg, resultando en una redundancia de 2,4 pero con una relación Sg/Cg inferior de 0,57.

3.1.2 Identificación de secuencias repetitivas microsatélites (SSR) en los unigenes

Utilizando estas secuencias únicas (unigenes), se realizó la búsqueda de regiones microsatélites mediante GDR SSR Server. Para determinar estas regiones, se consideraron aquellas secuencias con un mínimo de siete repeticiones para motivos dinucleotídicos, cuatro repeticiones para motivos trinucleotídicos y tres repeticiones para motivos tetra- y pentanucleotídicos.

De los 4.924 unigenes analizados, 952 (19,3%) contenían 1.140 regiones microsatélites. El 90% (1029) de los EST-SSRs encontrados presentaron regiones flanqueantes suficientes para el diseño de oligonucleótidos capaces de amplificar las regiones de interés (Tabla Anexa 1, en soporte CD).

Se realizó un análisis de similitud de secuencia usando BLASTN ($E 10^{-5}$) entre el conjunto de marcadores EST-SSRs obtenidos en este trabajo y los publicados por otros autores. Se detectaron 14 SSR en 12 unigenes que presentaron similitud de secuencia con 88 secuencias

EST-SSRs en *Eucalyptus globulus* publicadas por Yasodha y col. (2008); 34 EST-SSRs con similitud con los publicados por Rengel y col. (2009) y sólo dos con los EST-SSRs publicados por Faria y col. (2010 a,b).

A partir de estos resultados se infiere que el presente trabajo *in silico* contribuye con una gran cantidad (979) de nuevos marcadores funcionales de tipo SSR potencialmente útiles.

La identificación de aproximadamente una secuencia SSR cada 5 unigenes analizados (952 de 4.924 unigenes) resultó mayor a los descritos por Yasodha y col. (2008) (96 de 4.005 ESTs de *E. globulus*), y Rengel y col. (2009) (512 de 3.857 unigenes de *Eucalyptus*). Además, esta proporción resultó mayor a las descritas para otras especies de dicotiledóneas (entre 2,65% y 16,82%).

Por otro lado, el porcentaje de unigenes conteniendo SSRs (19,3%) hallados en este estudio, se corresponde a aquellos descritos por Ceresini y col. (2005) y Rabello y col. (2005) (25,5% y 29% respectivamente). Estas concordancias se encontraron a pesar de que los otros autores incluyeron en el análisis genotecas de otras especies y otros parámetros para la definición de los microsatélites, así como la inclusión de las secuencias repetitivas de mononucleótidos.

La eficiencia en la identificación de secuencias SSR depende de diversos factores que incluyen, la herramienta informática utilizada en la búsqueda y el criterio utilizado en la búsqueda y el tamaño del conjunto de datos utilizados (Kumpatla y Mukhopadhyay 2005). Esto podría explicar las diferencias encontradas entre este trabajo y el de otros autores.

En diversos estudios se indica que los SSRs con mayor número de repeticiones poseen mayor probabilidad de mostrar polimorfismo dentro de poblaciones (Cho y col. 2000; Faria y col. 2010 a,b; La Rota y col. 2005; Temnykh y col. 2001). Si en este trabajo se hubiesen seleccionado sólo aquellas regiones microsatélites con mayores números de repeticiones (elevando el número de repeticiones a nueve para dinucleotídicos y a cinco para tri-, tetra- y pentanucleotídicos), sólo se hubiesen asignado 245 SSRs. Bajo estas condiciones, sólo dos EST-SSRs validados (denominados EST141 (AGA)₈ y EST191 (CT)₁₃), hubiesen sido seleccionados, perdiéndose un gran número de marcadores polimórficos (15 de los 17; ver Sección 3.1.5)

3.1.3 Caracterización de los motivos microsatélites

La presencia de SSRs dentro de genes es relativamente común, especialmente en las regiones no traducidas 5' y 3' (UTRs). Las repeticiones de tipo tri- y hexanucleotídicas se encuentran en mayor cantidad, tanto en regiones codificantes como no codificantes, pero otros tipos de

motivos repetitivos son menos frecuentes en regiones codificantes que en UTRs (Metzgar y col. 2000). En general, la baja frecuencia de SSRs encontrados en secuencias ESTs se relaciona con el bajo polimorfismo de regiones codificantes, en contraste con las no codificantes.

En este trabajo, los motivos trinucleotídicos y dinucleotídicos fueron los predominantes (58,6% y 22,8%, respectivamente), mientras los tetra- y pentanucleotídicos sólo representaron el 14,3% y 4,2% del total.

En diversos estudios de diferentes especies, se encontraron pocas repeticiones de dinucleótidos y tetranucleótidos en exones, pero una mayor cantidad de tri y hexanucleótidos (Li y col. 2004). La dominancia de secuencias repetitivas de tipo trinucleotídicas podría ser explicada sobre la base de la ventaja adaptativa que presentan éstas frente a las repeticiones no triméricas, posiblemente a causa del cambio en el marco de lectura que producen éstas (Metzgar y col. 2000).

Entre los 952 EST-SSRs, el 68,6% presentó un único SSR, mientras que otros unigenes presentaron entre dos y cinco SSRs en porcentajes de 26,1%, 4,2%, 0,7% y 0,4%, respectivamente. Para todas las categorías, el mayor porcentaje correspondió al tipo de motivo trinucleotídico (Figura 3.1).

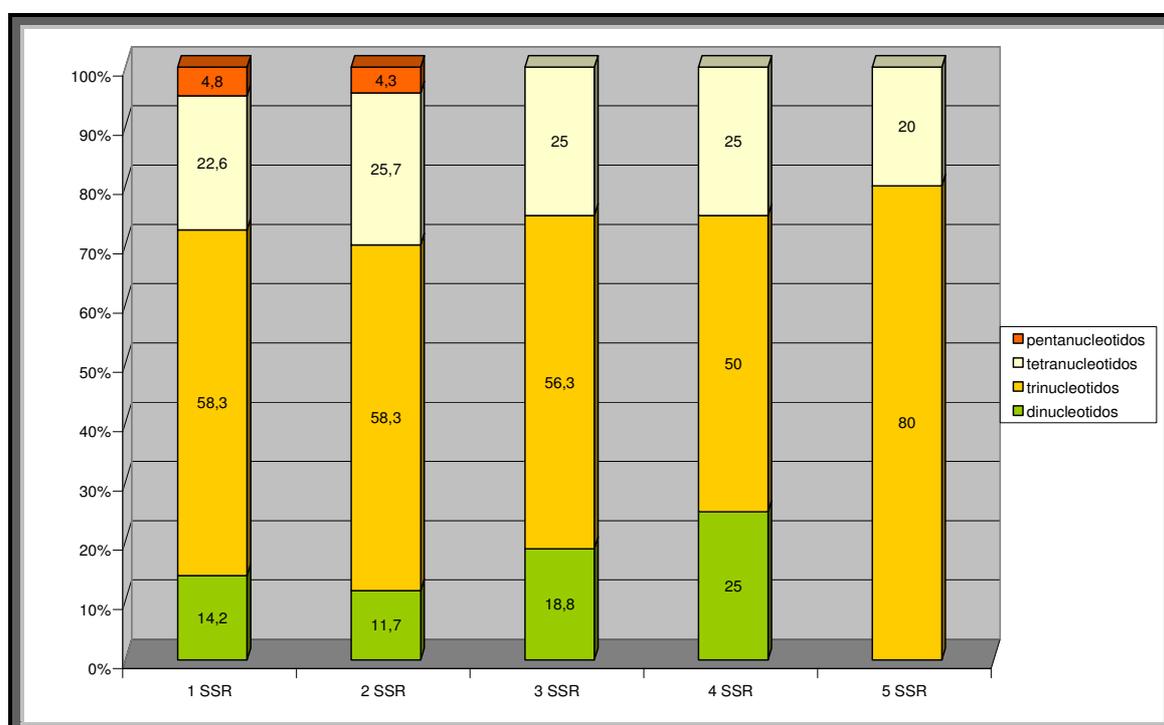


Figura 3.1: Frecuencia de los motivos repetitivos (di-, tri-, tetra- y pentanucleótidos) según el número de SSRs presentes en unigenes.

La frecuencia de los diferentes tipos de EST-SSRs se presenta en la figura 3.2

Entre las repeticiones dinucleotídicas, el motivo AG/CT fue el predominante (91,5%), mientras que los motivos AT, AC y CG fueron los menos representados (3,5%, 3,1% y 1,9% respectivamente). La frecuencia de motivos AG resultó mayor a los encontrados por Ceresini y col. (2005) (32%-13%) y Kumpatla y Mukhopadhyay (2005) (14,6%–54,5% del total de SSRs observados en 55 especies de dicotiledóneas). Un motivo dinucleotídico puede significar el agregado/pérdida y modificación de múltiples codones dependiendo del marco de lectura, y traducirse así en diferentes aminoácidos. Por ejemplo, el motivo AG/CT puede dar por resultado los codones GAG, AGA, UCU y CUC en el ARNm y traducirse en los aminoácidos Glu, Arg, Ser, Leu, respectivamente. Ala y Leu están presentes en alta frecuencia en proteínas, 8% y 10%, respectivamente (Lewin 1994).

Katti y col. (2000) describieron que entre las proteínas conocidas que contienen 10 o más aminoácidos repetidos, los más abundantes son la glutamina, alanina, glicina, ácido glutámico y serina. Éstas podrían ser unas de las razones adaptativas por las cuales el motivo GA/CT está muy representado en las colecciones de EST.

En este trabajo, entre los motivos repetitivos trinucleotídicos, la clase CCG/CGG fue la más representada (31,1%), y en menos porcentaje las clases AGG/CCT (23,2%), AAG/CTT (17,3%) y AGC/GCT (9,6%). Estos valores resultaron mayores a los descritos en Ceresini y col. (2005). De los 33 tipos diferentes de motivos tetranucleotídicos posibles, se encontraron 31. El motivo más abundante fue CCTT seguido por AGCG y AGGG (12,9% y 12,3%, respectivamente). Para las repeticiones pentanucleotídicas, los motivos CGAGC y AAGGG fueron los más abundantes (Figura 3.2).

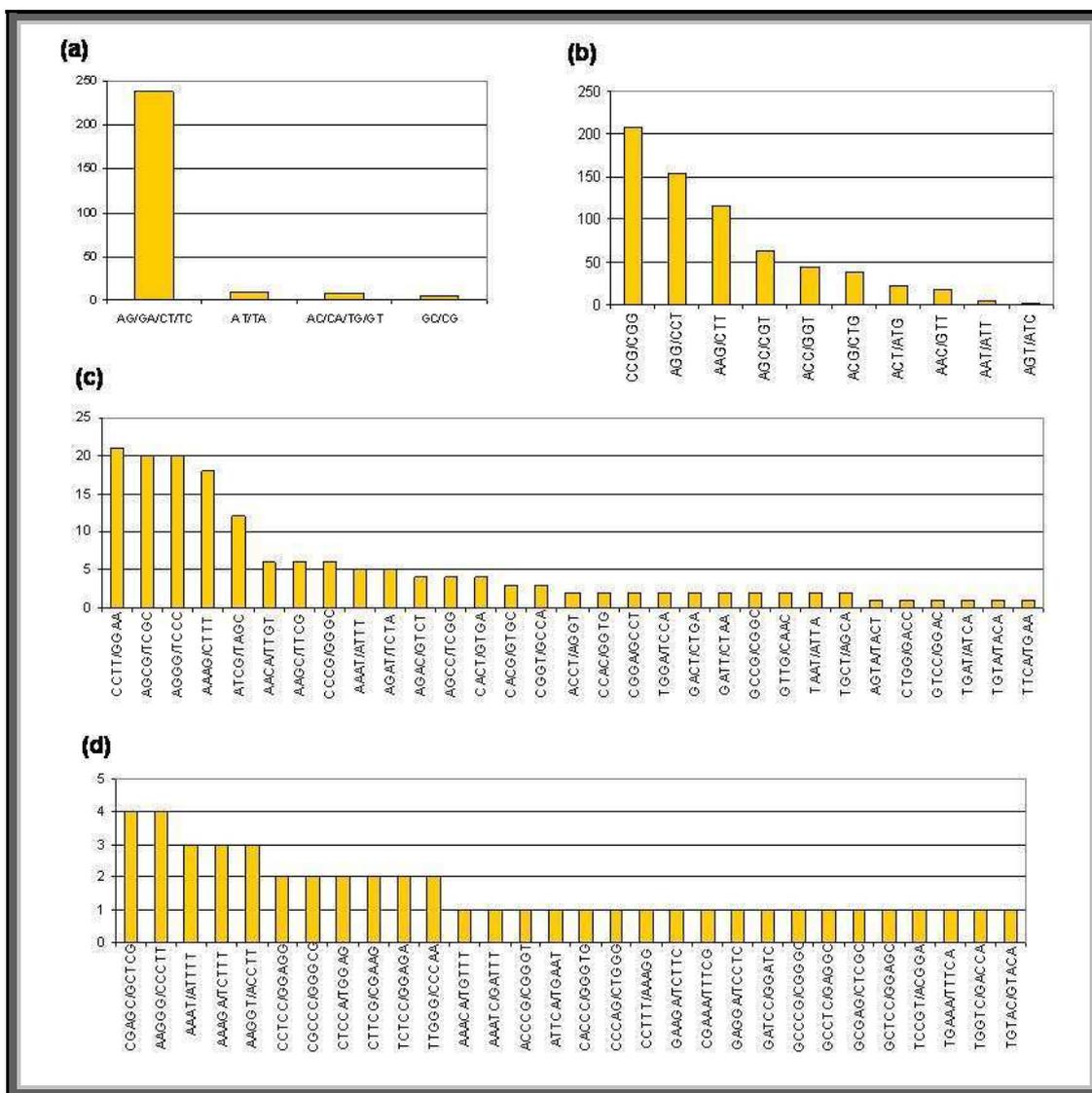


Figura 3.2: Frecuencia de los diferentes motivos repetitivos. Los histogramas muestran la distribución de las repeticiones de di- (a), tri- (b), tetra- (c) y penta- (d) nucleótidos.

Para una mayor caracterización de la posición de los microsatélites dentro de los genes, se analizó la distribución de estos motivos en regiones no codificantes (UTRs) y codificantes (ORFs), utilizando el programa GDR SSR. Sin embargo, como se trata de predicciones de la posición de los ORFs, distintos resultados podrían obtenerse si se utilizan diferentes programas para este análisis.

Los resultados mostraron que el 72,3% de los SSR se hallaban dentro de las regiones ORFs. Este inesperado resultado (mayor cantidad de SSRs dentro de los ORFs) podría deberse a una sobre-estimación de la longitud del ORF definido por el programa getORF EMBOSS. El mismo, selecciona el ORF más largo como posible región codificante e informa la ubicación del SSR en relación a éste.

La proporción relativa de los diferentes tipos de motivos repetitivos dentro de cada región puede observarse en la tabla 3.1

Tabla 3.1: Posición de los distintos motivos repetitivos en ORFs o UTRs

Tipo de repetición	ORFs	UTRs	Porcentaje en ORFs	Porcentaje en UTRs
Dinucleótido	98	60	62,0	38,0
Trinucleótido	549	120	82,1	17,9
tetranucleótido	148	114	56,5	43,5
pentanucleótido	30	21	58,8	41,2
total	825	315	72,3	27,7

En este tipo de análisis, se espera que los motivos repetitivos tri- y hexanucleotídicos sean los más representados en las regiones codificantes. La existencia de motivos SSRs particulares (especialmente los no-triméricos) podría tener implicancias en cómo las regiones codificantes son transcritas, debido a las posibles mutaciones en los marcos de lectura que podrían ocurrir cuando otras clases de SSR alteran el número de repeticiones (Metzgar y col. 2000).

En este trabajo se pudo observar que las repeticiones de trinucleótidos se encontraron en gran proporción en las regiones codificantes (82,1%). Sin embargo un número sustancial de motivos di- (62%), tetra- (56%) y pentanucleótidos (59%) también están presentes en esta región. Al analizar este elevado número repeticiones se encontró sin embargo, que el 26%, 96% y 100% de di-, tetra- y pentanucleótidos presentaron un número de repeticiones tal que, el largo total (en pb) del fragmento fue múltiplo de tres (por ejemplo: (TC)₉, (GA)₉, (CTCC)₃, (CGCCC)₃). Para el primer ejemplo, 9 motivos de 2 pb tienen un largo total de 18 bp, equivaliendo a 6 codones, manteniendo así el marco abierto de lectura.

3.1.4 Anotación funcional de los EST-SSRs de *Eucalyptus globulus*

Con el fin de asignar una función molecular probable a los marcadores derivados de EST-SSRs, las 952 secuencias correspondientes fueron analizadas usando el programa Blast2Go (Conesa y col. 2005, <http://www.blast2go.de/>), el cual les asigna una clasificación según *Gene Ontology* (GO).

De los EST-SSRs encontrados, 749 (78,7%) presentaron similitud significativa (13645 “hits”, $E \leq 10^{-5}$) con la base de proteínas no redundantes usando BLASTX (http://www.ncbi.nlm.nih.gov/blast/blast_databases.shtml).

Entre las diferentes especies de organismos con secuencias de proteínas depositadas en GenBank, los unigenes generados en este estudio presentaron el mayor número “hits” con *Arabidopsis thaliana* (19,8%), *Oryza sativa* (19,5%) y *Vitis vinifera* (15,5%). Dentro de las especies forestales, la mayor cantidad de “hits” encontrados fue con la especie *Populus trichocarpa*, una de las especie modelo dentro de este grupo (Figura 3.3).

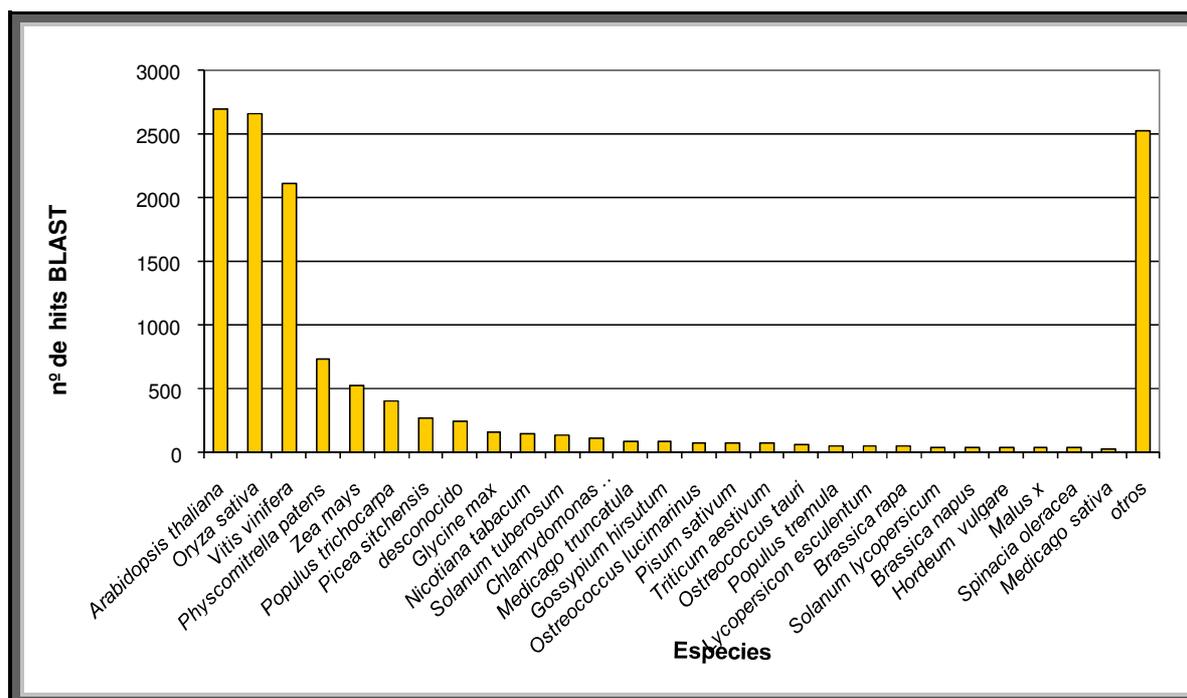


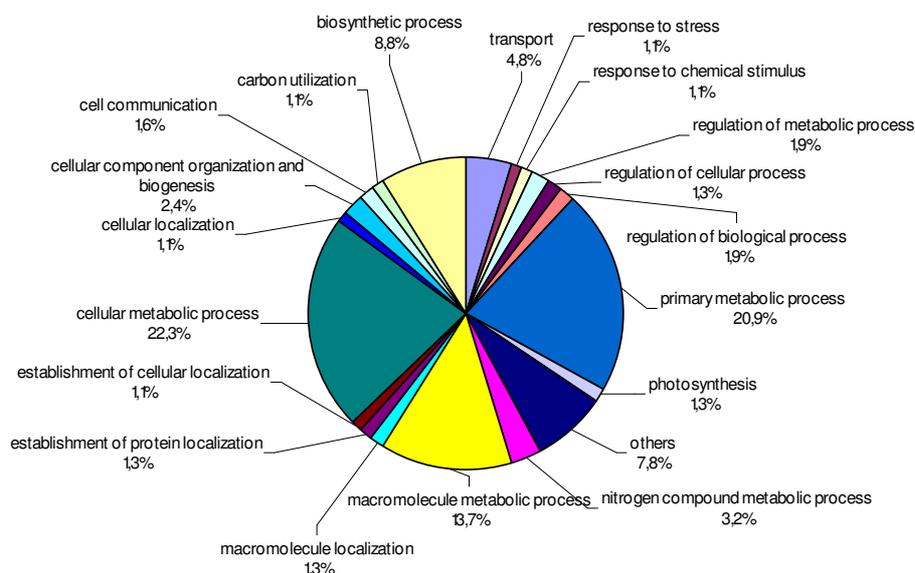
Figura 3.3: Histograma del número de hits BLASTX encontrados en las distintas especies con secuencias de proteínas depositadas en GenBank

A partir de éstos, se les asignaron 425 términos GO a 226 EST-SSRs (30,2%). De la totalidad de términos GO, 153 pertenecen a la categoría “Proceso Biológico”, 190 a “Función Molecular” y 82 a “Componente Celular”

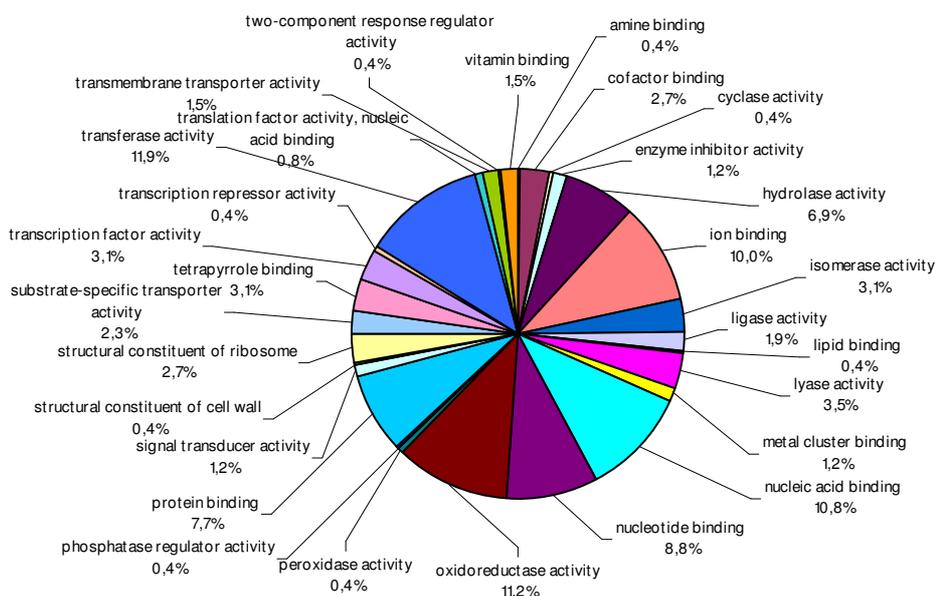
Los gráficos de la Figura 3.4 muestran la distribución de los términos GO a 3^o nivel, de acuerdo a la clasificación del Consorcio GO (Ashburner y col. 2000). La mayoría de los 153 términos GO en la categoría Proceso Biológico, pertenecen a la sub-categoría “Proceso metabólico celular” (22%) y “Proceso metabólico primario” (21%), entre otros menos representados. A su vez, la mayoría de los términos dentro de la clase Función Molecular, pertenecen a la sub-categoría “Actividad transferasa” (12%) y “Actividad oxidoreductasa” (11%). Estos resultados son similares a los descritos por Rengel y col. (2009) en paralelo a esta tesis, a pesar de que en este trabajo sólo se realizó la anotación funcional de los unigenes que poseyeron secuencias repetitivas SSR.

Los términos GO hasta el tercer nivel para cada una de las tres categorías, se encuentran descritos en la Tabla Anexa 2. Además, los términos GO asignados pertenecieron a 57 diferentes rutas metabólicas, en donde las más representadas fueron: fijación del carbono, ruta de las pentosas fosfato, metabolismo de la sacarosa y almidón y metionina (Tabla anexa 3). Estos resultados dan prueba de que los *loci* EST-SSRs de *Eucalyptus globulus* fueron funcional y metabólicamente representados.

(a) Proceso biológico



(b) Función molecular



(c) Componente celular

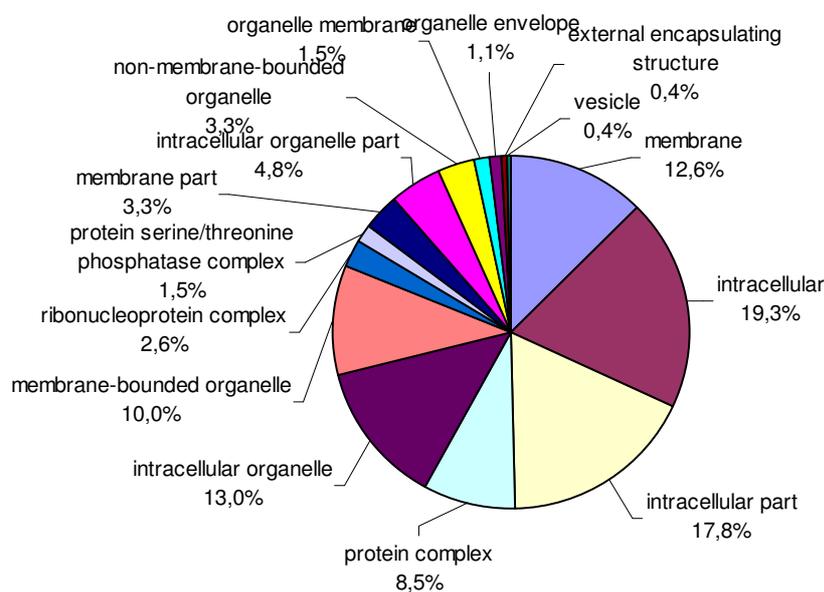


Figura 3.4: Asignación Gene Ontology (GO) (términos del 3º nivel) de 226 unigenes de *E. globulus*. El número de ESTs anotados por cada categoría es: (a) 153 para Proceso biológico, (b) 190 para Función molecular y (c) 82 para Componente celular.

3.1.5 Validación de los EST-SSRs en *E. globulus*

Con el objetivo de validar el panel de marcadores EST-SSRs diseñados, se tomaron al azar 56 combinaciones de oligonucleótidos que se utilizaron para amplificar fragmentos de ADN provenientes de ocho individuos procedentes de seis razas geográficas de *E. globulus*.

Como resultado, 37 *loci* mostraron un patrón de amplificación positivo y reproducible (Tabla 3.2), mientras que los 19 (34%) restantes no mostraron productos de amplificación. Esta falta de amplificación puede deberse a diferentes razones, que no fueron investigadas en este trabajo: las secuencias complementarias a los oligonucleótidos diseñados podrían localizarse a lo largo de un sitio de empalme (*splice*), la presencia de un intrón de gran tamaño en la secuencia genómica podría reducir significativamente la eficiencia de la reacción de amplificación, el empleo de secuencias derivadas de la base de datos de ESTs (que son secuencias que carecen de precisión por derivar de secuenciaciones simples sin curar) por lo que el diseño de los oligonucleótidos podría no resultar suficientemente complementario, entre otras razones. Otra razón que no puede descartarse es la falta de complementariedad total de los oligonucleótidos debido a la elevada diversidad nucleotídica encontrada en *Eucalyptus* (1 SNP cada 30 pb encontrada en *E. globulus* y 1 SNP cada 16 pb en *E. camaldulesis*) (Kulheim y col. 2009).

Entre los *loci* validados, 11 presentaron tamaños mayores (en pb) que los esperados, probablemente por la presencia de una secuencia intrónica dentro del fragmento genómico amplificado. A excepción del EST216, los 10 *loci* restantes presentaron tamaños de amplicones iguales o mayores a 500 pb y fueron considerados monomórficos (Tabla 3.2).

De los 37 oligonucleótidos validados, 17 reflejaron polimorfismos con un número de alelos entre dos y ocho, en la muestra de ocho individuos. A excepción de dos de éstos, todos los *loci* presentaron alguna función putativa asociada (respuesta a estrés, rol en señalización y defensa en plantas, factores de transcripción y otras funciones (Tabla 3.2)).

Al analizar la ubicación de los microsatélites que fueron validados, se encontró que 28 se ubicaban dentro de putativos ORFs, de los cuales el 43% resultaron polimórficos. Entre las nueve secuencias SSRs localizadas en extremos UTRs, cinco (56%) resultaron polimórficas.

La alta representación de SSRs validados dentro de regiones codificantes, podría revelar su potencial uso cuando se intente correlacionar marcadores en genes candidatos a rasgos fenotípicos o para llevar a cabo mejoramiento asistido por marcadores ("Marker assisted breeding o MAB").

Tabla 3.2 Loci EST-SSRs validados en *Eucalyptus*

Se incluye nombre del oligonucleótido, contig/singleton (C/S), tipo y número de repeticiones, posición dentro del ORF (S o N), secuencia de oligonucleótidos, temperatura de hibridización (°C), tamaño del amplicón (bp), polimorfismo (S/N), similitud BLASTX con secuencias de proteínas del GenBank.

nombre	C/S	motivo	ORF	secuencia de oligos (5' - 3')	temp. (C°)	tamaño amplicón (bp)	polim	función
EST 1	Contig 1146	(AT) ₁₀	N	F: TTGACAGTGATGGTCAGAAA R: GAACAAGGCATTTGGATTT	52	143	S	proteína inducida por auxinas [<i>Vigna radiata</i>]
EST 2	Contig 1229	(TGC) ₄	S	F: AGCCAGCTCTCTGTGTCTAC R: AGAAGAAGCAGTGGATGAAG	60-50	329	N	factor de transcripción [<i>Arabidopsis thaliana</i>]
EST 3	CT989441	(AAAAG) ₃	S	F: GAGGTCCTTCTTCTGATGTG R: CCATTCAATTCACAGTTTCT	60-50	240-250	S	proteína remorina [<i>Arabidopsis thaliana</i>]
EST 5	Contig 614	(GAG) ₄	S	F: CTTGGCCGAGGTCGAGAG R: ACCAAGATGAGGCTACACAC	60-50	333	S	proteína desconocida [<i>Arabidopsis thaliana</i>]
EST 9	Contig 299	(GCC) ₆	N	F: GTAAGCAATACACTTCTTCTTGC R: CAAGTGCTTCACCTCCTACA	52	>500	N	proteína rica en glicina [<i>Arabidopsis thaliana</i>]
EST 11	Contig 1185	(CT) ₈ / (GA) ₆	S	F: GTAGAACCAATCCAACAGGA R: TTATTACGCCCTCGTTTCTTC	60-50	240-250	S	Proteína tipo NtSar1 [<i>Solanum tuberosum</i>]
EST 12	Contig 1193	(GCC) ₅	S	F: GATGCTCCTGATGTACCAAT R: AAGTGCTAGACTCATACGCC	60-50	341	N	LOL1 (similar a LSD1) [<i>Arabidopsis thaliana</i>]
EST 13	Contig 1251	(CT) ₁₁	N	F: GTAATCAATTTCTGCGAGGA R: CAGTCTTGAGAACTGCTCC	54	375	N	MATH [<i>Medicago truncatula</i>]
EST 25	Contig 1361	(CAG) ₄	S	F: AGAGGATGGTGCTTCTTCTC R: TGACTGAGTGTTTGTGCATT	60-50	349	N	tipo homeodominio [<i>Medicago truncatula</i>]
EST 33	Contig 1255	(GGAT) ₃	S	F: ATAGTTTGCCTTGTGGAAGA R: AACAATTCATAAGCCGTTA	60-50	130-150	S	NADPH-ferrihemoproteína reductasa [<i>Vicia sativa</i>]
EST 36	Contig 1267	(TTC) ₄	S	F: TACGCCATTCTCTCTCACTT R: TGTTGTATTGCTGCTTTGAG	60-50	377	N	putativa AAA ATPasa [<i>Solanum tuberosum</i>]
EST 37	Contig 1255	(CATC) ₃	S	F: AACAATTCATAAGCCGTTA R: TTTATTAGTTGGCAAGGAGC	60-50	220-240	S	NADPH-citocromo P450 oxidoreductasa isoforma 1 [<i>Populus balsamifera</i> x <i>Populus deltoides</i>]

nombre	C/S	motivo	ORF	secuencia de oligos (5' - 3')	temp. (C°)	tamaño amplicón (bp)	polim	función
EST 46	Contig 620	(GCTA) ₄	N	F: AAGTTTCAACTGGACGAGG R: TGTTTCCATTGCGTAGTCA	60-50	210-220	S	Factor de transcripción union a DRE [<i>Gossypium hirsutum</i>]
EST 58	Contig 1373	(ACCAG) ₃	S	F: GGACAGTATACGGTCTTATT R: ATTTGGTTTGGATTCTTGTG	52	130-150	S	proteína de respuesta a fosfato [<i>Arabidopsis thaliana</i>]
EST 60	Contig 1357	(ATAG) ₃	S	F: ATTAAGTGACCGAACTCGAA R: TTCCACTTCCAGTTCCTATG	60-50	226	N	NADH dehidrogenasa subunidad 5 [<i>Eucalyptus globulus</i>]
EST 61	Contig 1313	(GAT) ₆	S	F: GGAGTTCAATCAAGGTCTCA R: ACAACATGCTTCACCTTCTT	54	>>500	N	PHAP2A [<i>Petunia x hybrida</i>]
EST 68	Contig 1361	(CGT) ₄	S	F: GCATCTTCGTTAGGACAGAC R: GTGGTTGCGGCGAAGGTA	60-50	>500	N	SYP52 [<i>Arabidopsis thaliana</i>]
EST 80	Contig 1364	(CT) ₇	S	F: GGGCAGGTACCTTAGAGTTT R: CATGGACTTCTCCAACCTTGT	60-50	>>500	N	prolil 4-hidroxilasa subunidad alfa [<i>Arabidopsis thaliana</i>]
EST 88	Contig 1356	(GAT) ₄	S	F: AGACCATGAACATCCAGAAC R: CATGCAGCTTACTACTCCC	60-50	225-250	S	proteína homeótica de respuesta BEL1 [<i>Solanum tuberosum</i>]
EST 96	Contig 1367	(TTG) ₄	N	F: GAGGCTCATCAAGAGAGATG R: ATCCAACCCTTAGCAAGA	52	253	N	cinamil alcohol dehidrogenasa [<i>Elaeis guineensis</i>]
EST 125	Contig 1397	(AGC) ₄	S	F: GGACAATGCTAAAGTGAAG R: GAGCGTTCGGTGTAAGATAC	60-50	200-210	S	proteína choque térmico [<i>Nelumbo nucifera</i>]
EST 134	Contig 252	(CA) ₈	S	F: TAGAGCAACGAAACCAAGTT R: AGGTTTCCCTATCTGAGGAG	52	300-340	S	Proteína tipo vtc2- [<i>Actinidia chinensis</i>]
EST 141	Contig 1404	(AGA) ₈	S	F: ATTCTGATGACGAGCTTGAT R: TTTAAGGGTGTCCAGAGAAA	54	310-315	S	proteína desconocida [<i>Arabidopsis thaliana</i>]
EST 155	Contig 1229	(CAG) ₄	S	F: AGAAGAAGCAGTGGATGAAG R: TTTGAAGGAAACAGCTAACC	60-50	398	N	factor de transcripción [<i>Arabidopsis thaliana</i>]
EST 156	Contig 1433	(TGG) ₄	N	F: GGTAAACAATAATCGCAGG R: AATTTACCTCGGACTCCTTC	60-50	116	N	sin similitud

nombre	C/S	motivo	ORF	secuencia de oligos (5´-3´)	temp. (C°)	tamaño amplicón (bp)	polim	función
EST 157	CT979846	(CT) ₇	N	F: AATTTACAGCATCCCATGAC R: ATAGAATGCTTCCACCACAC	60-50	290-325	N	proteína de control de ciclo celular [<i>Arabidopsis thaliana</i>]
EST 181	CB009717	(CGA) ₅	S	F: AACCTGTAACAACAACCACG R: CAGATCTCCACAGTACCCAC	60-50	500	N	proteína de respuesta rápida a deshidratación [<i>Pseudotsuga menziesii</i> var. <i>menziesii</i>]
EST 191	BF942594	(CT) ₁₃	S	F: AGTCGTGAAACACACACAGA R: ACGATCCCAAGATTCTAAT	54	205-225	S	serina treonina kinasa [<i>Cucumis sativus</i>]
EST 197	BF942556	(CTG) ₄	N	F: GCCTCAAATTTCAAATCAGT R: AGAGATTGGAGAGGAAGGAG	60-50	100-115	S	Proteína de unión a GTP [<i>Glycine max</i>]
EST 200	BF942514	(AG) ₈	S	F: GCTGAAAGATCAGATTGTCC R: AGGACCAAATGAATTGCTAA	54	330-345	S	Prot. de membrana de cloroplasto 37kDa [<i>Petunia x hybrida</i>]
EST 203	BF707491	(TCGTG) ₃	S	F: AAGCTCTTCGAGAACGAGTA R: ACAAGAGACCTTCCCTTTC	60-50	>500	N	NUDIX hidrolasa [<i>Medicago truncatula</i>]
EST 205	BF707487	(TTG) ₄	S	F: CACGGATGCTTTCTACTGTC R: AAACAAGCTTCATTAATCGC	60-50	500	N	Alanina transaminasa [<i>Arabidopsis thaliana</i>]
EST 208	BF707484	(CTC) ₄	S	F: TTACGTCTTCCCTTCTCGT R: TTGAATGGATAATCTGGAGG	60-50	>>500	N	Enzima conjugada a ubiquitina [<i>Arabidopsis thaliana</i>]
EST 210	BF707482	(CT) ₅ /(AG) ₇	S	F: TCCTCTCTCTTCCGTAGTT R: ACTCCATTGCTCTGATTCTG	60-50	>500	N	Prot. Dedos de zinc de respuesta a estrés [<i>Citrus</i> cv. Shiranuhi]
EST 214	AW600866	(TTGC) ₃	S	F: ATCCTCACTGGCATCTTCTA R: ATCCAGTTGTCATGGAAGTC	60-50	>500	N	Proteína intrínseca delta de tonoplasto [<i>Gossypium hirsutum</i>]
EST 216	AW600856	(GAA) ₇ / (GA) ₇	S	F: AAACACGAAGTGCAGAGC R: AATGCCAGATCATCAGATTC	54	309-405	S	Enzima conjugada a ubiquitina [<i>Populus x canadensis</i>]
EST 219	AW600851	(CT) ₈	N	F: AAGCACCACCATCTTCTA R: TAGCGTACTTGGTTAGGGAA	54	110-115	S	Proteína de estrés de maduración mediada por ABA [<i>Prunus persica</i>]

En resumen, en este capítulo se describió el desarrollo *in silico* de nuevos marcadores moleculares funcionales de tipo microsatélite (SSR) a partir de ESTs de *Eucalyptus globulus* publicados en GenBank y depurados para evitar redundancia de información. A partir de 12.690 ESTs públicos explorados, provenientes de distintas genotecas (de cambium, sometimiento a bajas temperaturas, microgravedad, diferenciación de xilema) se identificaron 4.924 secuencias no redundantes, de los cuales 952 (19% aproximadamente) contuvieron 1.140 regiones SSR. Se diseñaron oligonucleótidos para amplificar dichas regiones y se identificó la función putativa de 749 EST-SSRs utilizando BLASTX ($E \leq 10^{-5}$). A 226 de éstos se les asignaron bioinformáticamente (Blast2Go) categorías Gene Ontology (GO) según su función molecular, proceso biológico y componente celular. Para la validación de estos marcadores se amplificaron 56 *loci*, encontrándose amplificación positiva en 37 de ellos, con un total de 17 *loci* que demostraron ser polimórficos en los ocho individuos de *E. globulus* analizados.

En el conjunto seleccionado de EST-SSR de *Eucalyptus globulus* descrito se encontraron 22 SSRs en 21 genes candidatos (estructurales y reguladores) para calidad de madera, los cuales pertenecen a las rutas metabólicas de síntesis de celulosa, hemicelulosa, fenilpropanoico, metionina, shikimato, pentosa fosfato, así como tubulinas. Los mismos serán descritos en el capítulo II.

Resultados y Discusión

Capítulo II

**Caracterización de marcadores
microsatélites en genes potencialmente
involucrados en características de calidad
de madera**

3.2 Capítulo II Caracterización de marcadores microsatélites en genes potencialmente involucrados en características de calidad de madera

3.2.1 Caracterización de regiones microsatélites

3.2.1.1 Microsatélites en secuencias ESTs

A partir del análisis de la colección depurada de secuencias ESTs de *Eucalyptus globulus* depositadas en GenBank (desarrollado en el capítulo I), se encontraron regiones microsatélites en genes candidatos de las siguientes rutas metabólicas (se indican siglas por sus nombres en inglés):

Biosíntesis de celulosa: Celulosa sintasa 5 (CesA5), péptido de glicosilación reversible (RGP), pectina metilesterasa (PME) y celulosa sintasa-tipo glicosiltransferasa familia 2 (CSLD4)

Tubulinas: α -tubulina (AT1) y proteína similar a tubulina (TLP)

Biosíntesis de fenilpropanoico: cinamoil CoA reductasa (CCR), glutatión S-transferasa (GST), fenilalanina amonio-liasa (PAL), ácido cafeico o-metiltransferasa (COMT), peroxidasa (PER)

Ruta del shikimato: shikimato kinasa (SK)

Metabolismo de la metionina: homocisteína S-metiltransferasa (HMT), s-adenosil-l-homocisteína hidrolasa (SAHH), metionina sintasa (MS), s-adenosilmetionina sintasa (SAMS), glutamato decarboxilasa (GAD)

Metabolismo de hemicelulosas: xiloglucano endotransglicosilasa (XTH)

Biosíntesis de las pentosas fosfato: 24-esterol c-metiltransferasa (SMT1), fosfofructokinasa dependiente de pirofosfato subunidad beta (PPI-PFKb), transketolasa 1 (TK), y ribulosa-fosfato 3-epimerasa (RPE).

Los motivos repetitivos encontrados en estas secuencias ESTs pueden observarse en la siguiente tabla 3.3.

Tabla 3.3: Genes candidatos para calidad de madera, obtenidos mediante asignación de función de los EST-SSRs de *E. globulus*. Se indica la ruta metabólica, nombre del *locus*, número de acceso (GenBank o contig), enzima probable por similitud de secuencia, tipo de repetición (pb), motivo y número de repetición (Nº repet.).

Nombre del <i>locus</i>	Nº de acceso	Similitud de secuencia	pb	motivo	Nº repet.
Biosíntesis de celulosa					
RGP	Contig129	polipéptido de glicosilación reversible	2	ct	7
CesA5	ES593659	celulosa sintasa	3	tat	4
PME	ES594767	pectina metilesterasa	3	cct	6
CSLD4	ES596142	Celulosa sintasa-tipo glicosil transferasa familia 2	3	gct	4
Tubulinas					
AT1	Contig1034	alfa tubulina 1	3	ccg	5
TLP	ES588377	proteína tipo tubulina	3	aga	4
Ruta del fenilpropanoico					
PAL	ES591880	fenilalanina amonio-liasas	4	cgct	3
CCR-DRF	ES596432	Cinamoil CoA reductasa/ dihidroflavonol 4-reductasa	2	tc	7
COMT	Contig675	Ácido cafeico O-metiltransferasa	5	gagag	3
PER	ES588966	peroxidasa ATP19a	4	attt	3
Ruta del shikimato					
SK	BF942502	Shikimato kinasa	2	ct	6
Metabolismo de la metionina					
HMT	ES588789	selenocisteína metiltransferasa	3	ctc	5
SAHH	ES594621	S-adenosil-l-homocisteína hidrolasa	2	tc	7
MS	Contig938	metionina sintasa	4	gaga	3
SAMS	Contig423	S-adenosilmetionina sintasa	4	cttt	3
Metabolismo de hemicelulosas					
XTH	DQ100338	xiloglucano endotransglicosilasa	5	tttcg	3
Ruta de las pentosa fosfato					
SMT	Contig334	24-esterol c-metiltransferasa (SMT1)	2	ct	7
PPi-PFKb	Contig638	Fosfofructokinasa dependiente de pirofosfato subunidad β	3	cgc	4
TK	Contig453	transketolasa 1	3	gcc	4
RPE	Contig172	ribulosa-fosfato 3-epimerasa	2	ag	8
Desarrollo y respuesta a stress					
GST	CT98811	glutación S-transferasa	4	tatg	9
GAD_a	Contig557	glutamato decarboxilasa	4	caag	3
GAD_b	Contig557	glutamato decarboxilasa	4	aaga	3

Como era de esperar, se encontró una dominancia (34,8%) de motivos trinucleotídicos en estas secuencias derivadas de ESTs. Sin embargo, también se encontró una elevada proporción de motivos tetranucleotídicos (30,4%), seguidas por los di- (26%) y pentanucleotídicos (8,9%). Es importante destacar que, al igual que se mencionó en la sección 3.1.3, el 100% de tetra- y pentanucleótidos presentaron un número de repeticiones tal que, el largo total (en pb) del fragmento fue múltiplo de tres, manteniendo así el marco abierto de lectura.

3.2.1.2 Microsatélites en secuencias genómicas de *Eucalyptus* depositadas en GenBank

Para aumentar el número de genes candidatos a estudiar, se realizó además la búsqueda en GenBank de secuencias genómicas de *Eucalyptus* de función conocida así como secuencias descritas por otros autores (descritos en Sección 2.5 de Materiales y Métodos). La identificación y caracterización de regiones microsatélites fue realizada utilizando diferentes programas de búsqueda, así como la identificación manual por secuenciación directa de fragmentos de los genes candidatos amplificados.

Los motivos repetitivos encontrados en estas secuencias están descritos en la Tabla 3.4.

Tabla 3.4: Genes candidatos para calidad de madera, obtenidos mediante búsqueda en regiones genómicas de *Eucalyptus*. Se indica nombre del *locus*, número de acceso del GenBank, nombre de la proteína/enzima, ubicación en el gen, tipo de repetición, motivo y número de repetición.

Nombre del <i>locus</i>	nº de acceso	Similitud de secuencia	ubicación	bp	motivo	Nº repeticiones
CCR	AJ132750	Cinamoil CoA reductasa	promotor	2	ct	15
CesA3_A	EU165713	celulosa sintasa	promotor	2	ga	10
CesA3_B	EU165713	celulosa sintasa	promotor	2	ct	11
CesA3_C	EU165713	celulosa sintasa	promotor	3	aag	4
LIM 1	AB208710	factor de transcripcion LIM 1	promotor	2	ga	22
HMTi	FJ492059	homocisteína S-metiltransferasa	intrón	2	ct	

Como puede observarse, la mayoría de los motivos se encontraron en promotores y fueron de tipo dinucleotídico.

La presencia de microsatélites en promotores ya ha sido descrita por otros autores. En el género *Populus*, el 1,2% de los SSRs encontrados luego de secuenciación genómica (aproximadamente 1510), se encontraron en extremos 5'UTR (Arnaud y col. 2007).

Elementos repetitivos CT ya han sido descritos en *Arabidopsis thaliana* (Molina y Grotewold 2005). Zhou y col. (2007) lograron identificar microsatélites de motivo CT en promotores de

genes de miARNs los cuales son frecuentes y conservados en cuatro especies modelo analizadas (*Caenorhabditis elegans*, *Homo sapiens*, *Arabidopsis thaliana* y *Oryza sativa*), sugiriendo un importante rol en la transcripción de genes de microARNs.

3.2.2 Validación de los SSRs en genes candidatos para calidad de madera

Todos los SSRs que fueron identificados *in silico*, fueron validados mediante amplificación por PCR de muestras de ADN genómico de ocho individuos de *E. globulus* y algunos de ellos fueron confirmados por secuenciación automática (Tabla 3.5).

Tabla 3.5 Validación de 29 SSR en GC para calidad de madera. Se indica nombre del *locus*, número de acceso del GenBank, nombre de la enzima, secuencia de oligonucleótidos (F y R), temperatura de hibridación, tamaño esperado del fragmento de amplificación y resultados de la validación (monomórfico, polimórfico y productos de amplificación inespecíficos).

nombre del locus	homología de secuencia	oligonucleótidos (FyR)	T ^o hibrid.	tamaño esp.	validación
Biosíntesis de celulosa					
RGP	polipéptido de glicosilación reversible	F:TAAATAGCTGTCGTTGGGGG R:TAGGAAGGACAGGAACGTGG	60-50	269	inesp.
CesA5	celulosa sintasa 5	F: TTCACAGTTGGCTCCTGTTG R:ACTGTTAAATGCCGGACTION	60-50	292	mono
CesA3_A	celulosa sintasa 3	F:TTGGAAGTAATGGAGCATT R:AGTTTGTTCTTCCCTCCCTC	54	292	poli
CesA3_B	celulosa sintasa 3	F:TCCATTTCTTATGATCCGTC R:CCAAGTGGAGTATCCTCGTA	54	340	poli
CesA3_C	celulosa sintasa 3	F:GGGATTCATGCTAGACACAT R:TGCTTTATCTCCACCACTT	54	374	no
PME	pectina metilesterasa	F:GCTCAAACCAGAAAACCGAG R:GAGTGCCTTGGCAGGTAGAG	60-50	159	mono
CSLD4	Tipo celulosa sintasa	F:GATCTTGTACGCGATGTTG R:CCACTACAAGAGCCCTCAGC	60-50	133	no
tubulinas					
AT1	alfa tubulina 1	F:CAACTCCTCCGCCTTCCT R:TCTTGTCACTTGGCATCTGG	60-50	216	no
TLP	proteína tipo tubulina	F:TCCCTGTCTTTGCTTGAACC F:AGAGCCGACGAAATTAGCAA	60-50	183	no
Ruta del fenilpropanoico					
PAL	fenilalanina amonio-liasa	F:GGGACTTGACAGTTTCGGTC R:GCCTGAGAAGAGGAGTGGAA	60-50	147	mono
pCCR	Cinamoil CoA reductasa (promotor)	F: ATGGAAAATAAGGGC R: ACTTTTGGACGCTTTTGGAC	42	222	poli
CCR _e	Cinamoil CoA reductasa (exón)	F.AGTACAAGTTTCCGGTCCGGG R:TGGCAAAGACTGATACGCAG	60-50	140	no

nombre del locus	homología de secuencia	oligonucleótidos (FyR)	Tº hibrid.	tamaño esp.	validación
COMT	ácido cafeico o-metiltransferasa	F:TCTCGCGGAAGAGAGAGAAG	60-50	184	mono
		R:GCCATGATCTCGAGGAGGT			
PER	peroxidasa ATP19a	F:CGTTTCGTGTGCTGACATCT	57	266	poli
		R:ATGATACGTGTGGCCAATCA			
ruta del shikimato					
SK	shikimato kinasa	F:TGGCATATGTTGTGGTTAAA	54	336	poli
		R:GATGGGAAGAGAGACAGACA			
Metabolismo de la metionina					
HMTi	homocisteína S-metiltransferasa (intron)	F:TAGGGAGGGTCTTCCCCTTA	57	307	poli
		R:GATCTCCTCGGTGATGGTA			
HMTe	homocisteína S-metiltransferasa (exón)	F:TTCTTCCGCGTACAAAATCC	60-50	229	poli
		R:GGGTGAAGTGATGAGGCACT			
SAHH	s-adenosil-l-homocisteína hidrolasa	F:CCGGTCCGGATTCCC	60-50	102	inesp.
		R:CTGCTGCTCTGCTTCTTCT			
MS	metionina sintasa	F:TCCTCTCTATCCCTCCCC	60-50	182	inesp.
		R:CGCTGCTCTTTTATCCCAG			
SAMS	s-adenosilmetionina sintasa	F:TCGCACAAAGTATGAGTGCC	60-50	196	inesp.
		R:GTGGAAGTTCACAGGGGAGA			
metabolismo de hemicelulosas					
XTH	xiloglucano endotransglucosilasa	F:ATCAAGACATCACGAGCAA	58	132	poli
		R:CTCGCGATTCTAGAGAAGAA			
Ruta de las pentosa fosfato					
SMT	24-esterol c-metiltransferasa (SMT1)	F:GATTCGCGGATCTTTGACT	60-50	172	mono
		R:ACGTCGCTCTGTCCATCTT			
PPi-PFKb	Fosfofructokinasa dependiente de pirofosfato subunidad β	F:CTTCGTGTACAACGGCGAT	60-50	280	no
		R:CAACTTCTGGCTTGGTCCAT			
TK	transketolasa 1	F:GTAACTGACCAACGGCGAT	60-50	259	inesp.
		R:GGATGGTGTGACCGACTTC			
RPE	ribulosa-fosfato 3-epimerasa	F:CGCACCCAACCTTCTCTCTC	60-50	211	inesp.
		R:CTTCACCACCGTCTGGATCT			
Desarrollo y respuesta a stress					
GST	glutación S-transferasa	F:TTTGATCTCCTGCTTTCTGT	54	118	poli
		R:CGAAGAACGTGCTATTAGGT			
GAD_a	glutamato decarboxilasa	F:TAAAGCGAAAGCAAAACCGT	60-50	250	poli
		R:AATTGCGGACGTATCTGGAG			
GAD_b	glutamato decarboxilasa	F:AAAGCCAACCTTCCCCTGTCT	60-50	189	poli
		R:AGAATTGCGGACGTATCTGG			
factor de transcripción					
LIM1	factor de transcripción LIM	F:GCTTCCCTTCTTATCCTCCA	58	178	poli
		R:CTGTGAGCTTGTCCACCAGA			

A partir de este análisis, 12 regiones microsatélites en 9 genes candidatos involucrados en calidad de madera, resultaron polimórficas en *E. globulus*: cinamoil CoA reductasa (CCR) de *E. gunnii* (GenBank: AJ132750), celulosa sintasa (CesA3) de *E. grandis* (GenBank: EU165713) (dos regiones microsatélites denominadas CesA3 A y B), el factor de transcripción LIM1 (Eglim1) de *Eucalyptus globulus* (GenBank: AB208710), dos secuencias ESTs de *Eucalyptus* que presentan homología con los genes de homocisteína S-metiltransferasa (HMT) (GenBank: CB967988 (Kirst y col. 2004) y ES588789), y los ESTs con similitud de secuencia con las enzimas peroxidasa (PER) (GenBank: ES588966), glutamato decarboxilasa (dos regiones microsatélites, denominadas GAD_A y GAD_B) (contig 557, GenBank: ES595076, ES590002, ES590669), shikimato quinasa (SK) (GenBank: BF942502), xiloglucano endotransglicosilasa (XTH) (GenBank: CT989247) y glutatión S-transferasa (GST) (GenBank: CT988111).

3.2.3 Descripción de los SSRs polimórficos en GC para calidad de madera

A continuación se describen los genes cuyos motivos repetitivos SSRs resultaron polimórficos en el estudio de validación.

En todos los casos, la amplificación de los SSRs dieron patrones claros y reproducibles. Si bien en algunos casos los SSRs se encontraron en GC pertenecientes a familias multigénicas, no se observaron patrones de bandas que sugirieran que la amplificación se dio en diferentes lugares del genoma.

Gen cinamoil-CoA reductasa (CCR)

Una secuencia microsatélite de motivo CT fue encontrada en el promotor del gen CCR de *Eucalyptus gunnii* (GenBank: AJ132750) entre las posiciones -164 y -135, no descrita previamente. Esta región es cercana a la descrita por Lacombe y col. (2000), entre las posiciones -119 y -77, la cual es necesaria y suficiente para la expresión en tejido vascular en tallo.

Cinamoil-CoA reductasa (CCR) es una enzima que cataliza la conversión del éster de cinamoil CoA a su correspondiente cinamoaldehído, el primer paso específico en la síntesis de monómeros de lignina, siendo un buen gen candidato en el control de la biosíntesis de dicho polímero.

Se han descrito más de una copia del *locus* CCR en el genoma de diferentes especies vegetales. Por ejemplo, se han encontrado 10 genes en *A. thaliana* (Jones y col. 2001) y dos en *Zea mays* (Pichon y col. 1998). Sin embargo, diferentes estudios realizados en el género

Eucalyptus mostraron que dicho gen posee una copia única en *E. globulus*, *E. tenuiramis* (Poke y col. 2003) y *E. gunnii* (Lacombe y col. 1997).

En un estudio realizado por Chabannes y col. (2001) se encontró que líneas de tabaco transgénicas con expresión significativamente reducida del gen CCR, no sólo mostraron una disminución del contenido de lignina, sino un aumento de xilosa y glucosa asociada a pared celular. Así, éste y otro estudio realizado por Hu y col. (1999) demostraron que la disminución de la síntesis de lignina promueve, a su vez, la síntesis y acumulación de celulosa. Esto sugiere la importancia de la búsqueda de variables alélicas dentro de este gen, que podrían estar involucradas en cambios en los dos mayores componentes de la pared celular.

Diversos estudios mostraron polimorfismo dentro de este gen en eucaliptos. En un estudio de Gion y col. (2001) se ha descrito que el *locus* CCR co-localiza con QTL para contenido y relación de monómeros siringil–guayacil de lignina en otras especies de *Eucalyptus* (*E. grandis* x *E. urophylla*), explicando entre el 13,5% y 24,7% de la variación en estos caracteres, respectivamente.

Poke y col. (2003) identificaron 21 y 23 polimorfismos de simple nucleótido (SNP) en exones e intrones del gen CCR en *E. globulus*, respectivamente. Estas mutaciones incluyeron diferencias en aminoácidos altamente conservados dentro de la secuencia de la proteína. No obstante, Poke y col. (2004), mostraron que ciertos cambios en aminoácidos altamente conservados dentro de la enzima, no se correlacionaron con cambios en el contenido y composición de la lignina o densidad de la madera, en un análisis de 68 individuos F1.

Por otro lado, Thumma y col. (2005) encontraron polimorfismos en el gen CCR asociados con la variación en el ángulo de microfibrilla (MFA) en un estudio de mapeo por desequilibrio de ligamiento (DL), usando marcadores SNP en *E. nitens*.

En este estudio, se identificó una región microsatélite de motivo CT dentro del promotor del gen (Figura 3.5), el cual podría estar involucrado en la regulación de la transcripción. Como se mencionó, esta región está próxima a la posición descrita por Lacombe y col. (2000) que direcciona la expresión génica localizada en xilema.

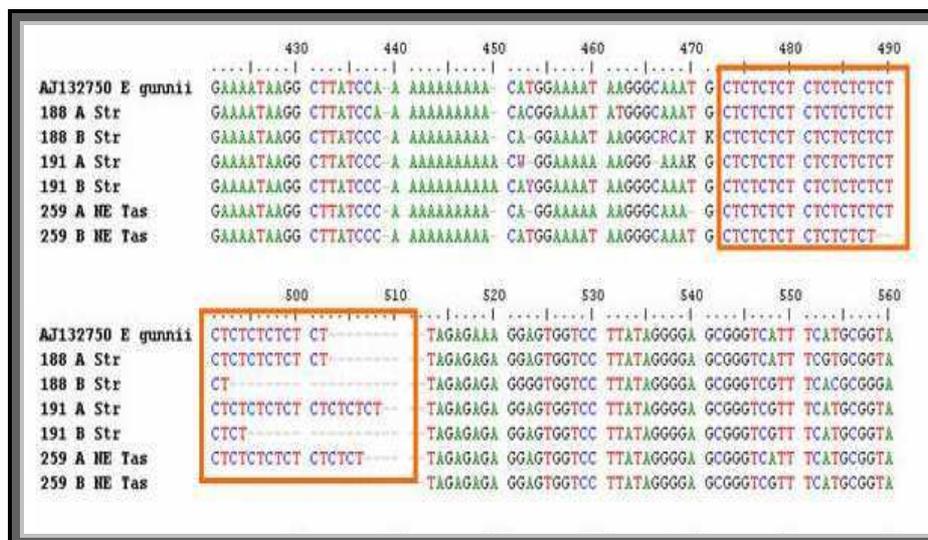


Figura 3.5: Alineamiento de 7 secuencias mostrando la variabilidad del motivo microsatélite CT encontrado entre la región promotora del gen CCR de *E. gunnii* (GenBank: AJ132750) y fragmentos secuenciados de 6 individuos de *E. globulus*. Se indican los motivos repetitivos (recuadro rojo)

Gen para la celulosa sintasa 3 (CesA3)

La síntesis coordinada del polímero de glucosa es dirigida por un complejo de celulosa sintasa (CelS) unido específicamente a membrana plasmática. Se postula que el complejo CelS está integrado por aproximadamente 36 subunidades de celulosa sintasa (CESA). Las plantas superiores contienen familias de genes de estas subunidades catalíticas, los cuales codifican para los componentes de un complejo enzimático embebido en la membrana celular. Un estudio realizado por Ranik y Myburg (2006) en seis ADNc completos de CesA en *E. grandis* (EgCesA1 a 6) demostró que existen dos grupos de genes CesA, asociados a deposición de pared celular primaria (EgCesA4 y 5) y secundaria (EgCesA1 a 3), mientras que EgCesA6 se expresó débilmente en todos los tejidos analizados.

En este estudio se identificaron dos regiones microsatélites, ambas presentes en los promotores altamente conservados de los genes CesA1 y CesA3 de *E. grandis* (GenBank: EU737100 y EU165713, respectivamente) (Creaux y col. 2008). Uno de estos SSR, con motivo de repetición GA, fue encontrado entre las posiciones -804 y -787, mientras el otro de motivo CT fue localizado entre las posiciones -248 y -226. Ambas regiones genómicas fueron denominadas CesA3a y CesA3b, respectivamente (Figura 3.6).

Sin embargo, y como se dijo previamente, no se observaron patrones de bandas que sugirieran que la amplificación haya ocurrido en las distintas copias (promotores de CesA1 y CesA3),

encontrándose como máximo dos fragmentos amplificados por individuo y por marcador. Por otro lado, cuando se analizó la segregación de ambos marcadores (CesA3_A y CesA3_B) en un cruzamiento de *E. grandis*, no se observó distorsión en las mismas, sugiriendo que cada uno de ellos estarían amplificando sólo en un lugar del genoma, también en esta especie (Martín García, comunicación personal).

Si bien se conoce el número de genes de CesA en *E. grandis*, hasta el momento no existen estudios de esta familia multigénica en *E. globulus*, resultando sumamente interesante para su investigación en etapas posteriores a este estudio.

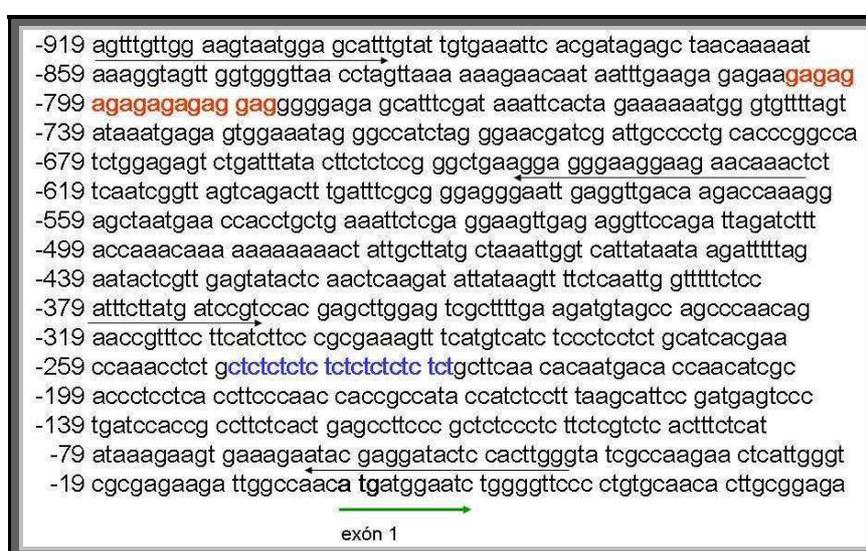


Figura 3.6: Motivos microsatélites encontrados en la región promotora del gen CesA3 de *E. grandis* (GenBank: EU165713). Se indican los motivos denominados CesA3a (en rojo) y CesA3b (en azul) y el codón de iniciación de la transcripción (en negrita).

Gen del factor de transcripción LIM1

Los motivos ricos en AC (Pal-box) constituyen un importante elemento de activación en cis para la expresión de genes involucrados en la biosíntesis del fenilpropanoico. Kawaoka y col. (2000) postularon que el factor de transcripción LIM puede unirse a los elementos Pal-box (secuencias consenso de unión de factores de transcripción en cis) en el promotor de genes de la biosíntesis de monolignoles y la represión del gen LIM conducía a la inhibición de la biosíntesis de lignina en plantas de tabaco transgénicas. Luego, estos autores informaron una

reducción del contenido de lignina en plantas de *E. camaldulensis* debido a la supresión de la expresión de este factor de transcripción (Kawaoka y col. 2006).

Además, Dejardin y col. (2004) desarrollaron 4 genotecas de ADNc, las cuales fueron utilizadas para colecciones de ESTs, que reflejan la expresión de genes en diferentes tejidos involucrados en la formación de la madera en álamo. La distribución de ESTs en diferentes tejidos mostró una elevada expresión de una proteína homóloga a LIM en madera en tensión, que se forma en la parte superior de los tallos doblados, que está enriquecido en celulosa debido a la formación de una capa gelatinosa suplementaria.

En este estudio se identificó una región microsatélite con motivo repetitivo GA entre las posiciones -43 y -1 del gen que codifica para el factor de transcripción LIM1 de *E. globulus* (Figura 3.7) (GenBank: AB208710)

```

-106 ggcttcctt tttatcctc cattctctc tctctctc ctacactca cagacacaat
-46 cacagagaga gagagagaga gagagagaga gagagagaga gagagaaagg cattgcagg
 15 aacaaccag aagtgcattg cctgtgagaa gacagtctat ctggaggaca agctcacagc
 75 tgacaataga atctaccaca aggcctgctt cagatgccac cattgcaaaag ggacttcaa
135 ggtatgcat gataaaaact gttagatctc aaggatttct cagtaattaa caaatgatca
195 gatgtgagtt tgatatattc cccaattag aggtccaaga gagttgcatt tctgtagatc

```

Figura 3.7: Motivo microsatélite AG encontrado en la región promotora del gen del factor de transcripción LIM1 de *E. globulus* (GenBank: AB208710). Se indican: el motivo de repetición (en rojo) y el exón 1 (línea azul).

Gen de homocisteína S-metiltransferasa (HMT)

La homocisteína S-metiltransferasa (HMT), también llamada metionina sintasa (METH) es una enzima implicada en la ruta metabólica de la metionina y cataliza la biosíntesis de S-adenosilmetionina, un sustrato de la ruta del fenilpropanico.

En *Arabidopsis*, la actividad enzimática de HMT es codificada por tres genes (HMT1, HMT2 y HMT3) (Ranocha y col. 2000, 2001), mientras que en tomate se encontró un gen con un único o bajo número de copias (Zeh y col. 2002). En un estudio realizado en brócoli, Lyi y col. (2007) describieron dos ADNc que codifican para proteínas con actividad HMT. Mientras que el producto de uno de esos genes exhibió sólo actividad HMT, el otro mostró una función adicional

(actividad selenocisteína S-metiltransferasa (SMT)) en aquellas especies que acumulan sulfuro/selenio. Ambos son genes de copia única dentro del genoma de brócoli.

Kirst y col. (2004) encontraron que los niveles de expresión de aquellos ARNm de *Eucalyptus grandis* transcritos a partir de genes que codifican para esta enzima (GenBank: ESTs CD669142, CD669275 and CB967988), se correlacionaron negativamente con el crecimiento.

En este estudio se diseñaron oligonucleótidos para amplificar una región del EST CB967988, encontrándose fragmentos de tamaños variables y mayores al esperado. Luego del aislamiento, purificación y secuenciación automática de estos fragmentos, se identificó una región microsatélite de motivo CT. El alineamiento múltiple de estos fragmentos junto a la secuencia del EST GenBank CB967988 de *E. grandis* similar a *Arabidopsis thaliana* homocisteina S-metiltransferasa (HMT), descrito por Kirst y col. (2004) permitió comprobar que el SSR se encontraba dentro de una región intrónica (Figura 3.8). La secuencia del microsatélite fue publicada en GenBank con número de acceso FJ492059 (Acuña et al., enviado a EJB) y el locus fue nombrado HMTi en este estudio.

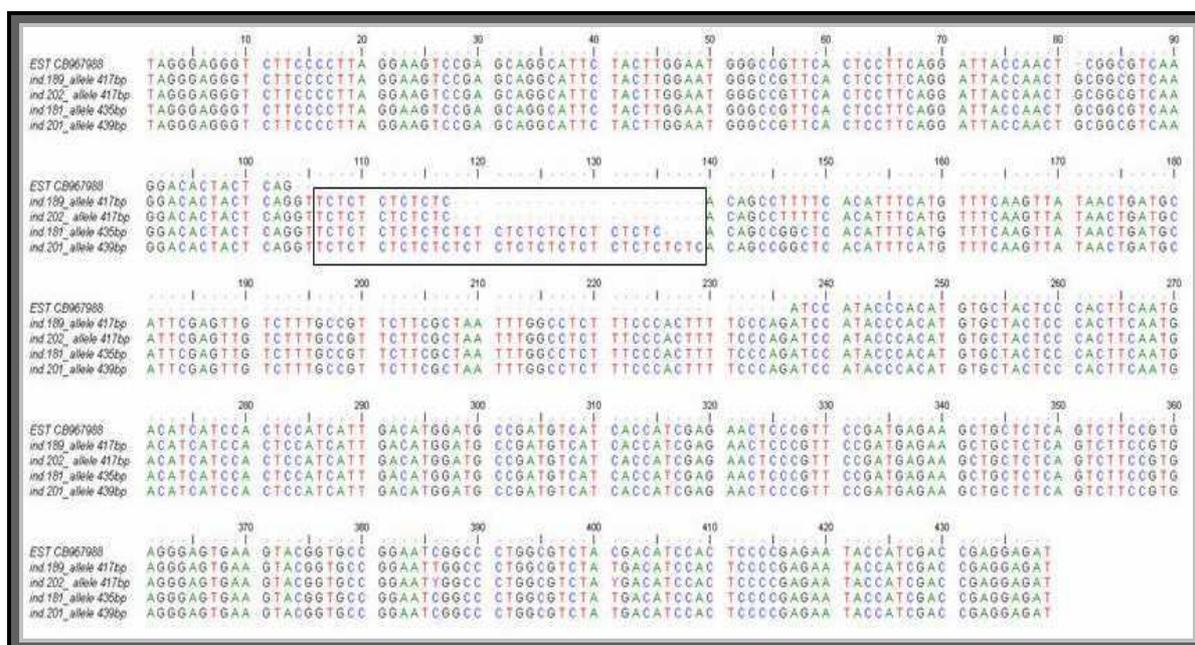


Figura 3.8: Alineamiento múltiple entre el EST CB967988 de *E. grandis* similar a HMT y 4 secuencias de ADN genómico de *E. globulus*. Se muestra la variabilidad del motivo microsatélite CT encontrado dentro del intrón del gen HMT putativo, denominado HMTi (recuadro negro).

El desarrollo y análisis de marcadores moleculares SSRs en intrones resulta interesante ya que la variación de longitud de estos motivos repetitivos puede afectar potencialmente la transcripción del gen, el procesamiento (“splicing”) del ARNm, la exportación al citoplasma e inducir al silenciamiento génico.

Además del motivo SSR antes descrito, se encontró otro motivo trinucleotídico (CTC) en el EST ES588789 con similitud a la enzima selenocisteína metiltransferasa. Dicho *locus* fue nombrado HMTe en este estudio.

Gen de shikimato kinasa (SK)

La ruta del shikimato une el metabolismo de carbohidratos con la biosíntesis de compuestos aromáticos. En una secuencia de siete pasos metabólicos, el fosfoenolpiruvato y la eritrosa 4-fosfato son convertidos en chorismato, el precursor de aminoácidos y diversos metabolitos secundarios aromáticos (Herrmann y Weaver, 1999). La vía del shikimato está involucrada en la biosíntesis de L- fenilalanina que se incorpora como sustrato de la ruta metabólica del fenilpropanoico (Fig. 1.2).

En esta ruta, la enzima shikimato kinasa (SK) es la implicada en la biosíntesis de la fenilalanina, tirosina y triptofano. La misma pertenece a una familia de transferasas, específicamente aquellas que transfieren grupos conteniendo fósforo (fosfotransferasas) con un grupo alcohol como aceptor.

En este estudio se identificó un EST (GenBank: BF942502) de *Eucalyptus globulus* similar a una enzima de la familia de SK de *A. thaliana* que contiene una región microsatélite de motivo CT (Figura 3.9).

```

CGGCACGAGCTTCTTTGATAGGTAAGTCTGATTTTATTGGTTGAGAGAGCT
GAAGGAATTTCAATACAGTGTCTGCGGAGTGTATGGATTGCTATGTGTAA
CTTGTGGTTTGTGCTCCAGTGTATGCATTGGTCGAGGAAGCTGCTGGCG
GTGAAGCTTCTGCAAACTTCTTAGGGAGACAGATGAGCAAGGATTTGCT
GAGTCTGAGGTGATATATTAGGGTGTGTTGCTTTTGTGCTTTTGCACATTA
GGCCTAGACTAGATGGCGGGAGTTCAATGGACTGTTATCTGGTGAAGTAA
GTCTTAGCGGCATCAGGTTGAATGAAAAGTGTGGGAACTTAGTTCTC
GTGGCATAATGTTGTTGTTAAAAGTGTGTTTGTGTTTGTGTTTGTGTTT
CAATTTTCTTGTAAACTCCTGTTTCATCAGTAGATAAATTTATATTGAG
GTTCACTAGGCATCTAATCAGGGGACTTCTGTTTTAGGCAGAAGTTTTAA
AGCAGTTATCATCCATGGGTCGCTTGGTTGTTTGTGCTGGAGATGGTGCA
GTTTCAGAGTTCAACTAATCTGTATGGAAATAATCTCTCTCTCTCGCTCTGT
CCTTCTCCATCTCCATCTCCATCTCCATCTCCGTCCTCCATCTCTGTTTCTG
TCTGTCTGTCTGTCTGTCTCTCTTCCATCTTAAGCTGCGCCTAAAAATTT
GCAGGGGATTCTTANACATGGGA
  
```

Figura 3.9: Motivo microsatélite CT encontrado en el EST (GenBank: BF942502) de *Eucalyptus globulus* similar a la enzima SK de *A. thaliana*. Se indican: el motivo de repetición (en rojo) y la ubicación de los oligonucleótidos diseñados (flechas).

Gen de xiloglucano endotransglicosilasa (XTH2)

El xiloglucano es el tipo de hemicelulosa más abundante de la pared celular de las plantas dicotiledóneas. Estas moléculas se adhieren a la superficie de las microfibrillas de celulosa actuando como elementos de enlace. Químicamente, consisten en cadenas de (1, 4)- β -D-glucanos sustituidas con residuos de xilosil, galactosil y fucosil.

La enzima xiloglucano endotransglicosilasa (XTH) es miembro de una familia de genes involucrada en la expansión celular. Las mismas poseen actividad xiloglucano endotransglicosilasa (actividad XET) y xiloglucano endohidrolasa (actividad XEH) (Rose y col. 2002). Así, XTH corta y liga moléculas de xiloglucanos, incluyéndolos en la pared celular (Genovesi y col. 2008). Debido a esto, el tamaño de las moléculas de xiloglucanos puede ser alterada luego de su deposición influyendo así en la dureza, flexibilidad y porosidad en la pared celular (Hrmova y col. 2007).

En un estudio realizado en *Arabidopsis thaliana* ecotipo Columbia, se encontraron 33 genes XTH dispersos en los cinco cromosomas del genoma (Rose y col. 2002).

Hasta la actualidad, fueron depositadas en GenBank cuatro secuencias de ARNm de correspondientes a genes que codifican para XTH1, XTH2, XTH3 y XTH4 en *E. globulus* (GenBank: DQ100337, DQ100338, DQ100339 y DQ100340; Goulao, sin publicar). En este estudio, luego de analizar la secuencia de ARNm que codifica para la enzima xiloglucano endotransglicosilasa 2 (XTH2) de *E. globulus* (GenBank: DQ100338), se encontró un motivo SSR pentanucleotídico (TTTCG).

Al realizar un BLASTn de esta secuencia con las otras correspondientes a genes que codifican para XTH1, XTH3 y XTH4 no se encontró similitud de secuencia en la región que contenía el SSR. Esto nos podría estar indicando que dicho motivo repetitivo se encuentra en sólo uno de los genes de esta familia multigénica.

```

1 actccgccgg caccgtgacc gcttactact tgfcttcgca agggccgact cacgacgaga
61 ttgactcga gttcctgggg aacctgagcg gcgacccta catcctccac accaacgtct
121 tcaactcaagg gaagggcaac aggggagcagc agttctacct gtggfittgac cccaccagga
181 atttccacac atactccgtc atctggaagc cccagcatal catcttcttg gtagacaaca
241 ttcctattag agttttcaag aatggagagt caattggcgt gccctcccc aagagccagc
301 ccatgaagat atactcgagc ctctggaatg ccatgattg ggccacgaga ggcggactga
361 tcaagacgga ctggtcgaaa gcgccctca cggcatacta caggaagttc caggccactg
421 cctgcacctg gtccacgggc tctctctct cggagatcgg acggcccgtc tctactctg
481 gatccacatg gaaaatcaat gagctcgatg cctatggccg gaggcggctc aggtgggtcc
541 agaagtact catgatctac gactactgtc cggacggcaa gaggtccct caaggcatcc
601 cagccgaatg caagcgtcg cgattctaga gaagaaaaag agggaaatca acgtgttcga
661 atggaaggtc gctcggcagt atagtagtaa tgcattttt cgtttcgttt cgcattgtca
721 agtcatgact tgctcgtgat tctctgattc caggtgggaa ataaaaaaaa aatggccttt
781 gtacaatgtt tctgttgaa agttgtggg gtgggggtgct gaattatgct ctgtaaagcc
841 acggtgtga cataaagata tgttttagc agttgttct acctgtgta ataaaacttt
901 gcattctct tcaaaaaaaaa aaaaaaaaaa aaaaaa

```

Figura 3.10: Motivo microsatélite TTTTCG encontrado en la secuencia ARNm para la enzima xiloglucano endotransglicosilasa (XTH2) de *E. globulus* (GenBank: DQ100338). Se indican: el motivo de repetición (en rojo) y la ubicación de los oligonucleótidos diseñados (flechas).

Gen de glutatión S-transferasa (GST)

Las enzimas glutatión S-transferasas de plantas (GSTs) comprenden una superfamilia heterogénea de proteínas multifuncionales agrupadas en cuatro clases (phi, tau, zeta y theta), de las cuales tau es la mayormente representada (Edwards y col. 2000). Las GST se conjugan con glutatión (GSH) y con una gran variedad de compuestos que son lipofílicos y poseen un centro electrofílico.

En maíz (*Zea mays* L. Pioneer híbrido 3906) se purificó la enzima glutatión S-transferasa (GST) que mostró actividad con ácido p-cumárico y otros fenilpropanoides insaturados (Dean y col. 1995) Además del ácido p-cumárico, la enzima mostró actividad con los ácidos o-cumárico, m-cumárico, trans-cinámico, ferúlico y alcohol coniferil. Sin embargo, existe evidencia de que esta actividad es mediada por la enzima peroxidasa y no GST (Dean y col. 1997).

Hasta el momento, los estudios de la función de esta enzima en especies cultivadas, mostraron que posee roles diversos como tolerancia a estrés oxidativo, detoxificación de compuestos xenobióticos y metales pesados, transporte de productos secundarios tóxicos y señalización celular durante la respuesta a estrés (Kampranis y col. 2000; Kilili y col. 2004; Loyall y col. 2000; Mueller y col. 2000; Thom y col. 2002).

La anotación del genoma de *Populus trichocarpa*, permitió la identificación de 81 genes GST, de los cuales nueve fueron considerados pseudogenes. La distribución de los genes GST entre los cromosomas resultó desigual: en los cromosomas 7, 9, 17 y 18 no se encontraron genes GST, mientras que en los cromosomas 1, 8, 10, 11, 14 y 19 la densidad fue relativamente alta, disponiéndose en “clusters”.

Todos estos genes fueron divididos en ocho clases distintas, las cuales se diferenciaron en sus tasas de evolución, estructura del gen, expresión de las respuestas a factores de estrés abiótico y en sus propiedades enzimáticas (Lan y col. 2009).

En este estudio se identificó una región microsatélite con motivo tetranucleotídico TATG en la secuencia EST CT988111 de *E. globulus* del GenBank (Figura 3.11), homóloga a la enzima glutatión S-transferasa clase tau de *Populus trichocarpa* (GenBank GQ377235).

```

GCGGCCGCCCGGGCAGGTACGAGACTTGCGGGGGTT
TCAGCATCGAGGCCGAGTGCCCCAAGCTCGTTGCGTG
GGCCAAGAGGTGCTTGCTCCGAGAGAGCGTGTCCAAG
GCCCTTCCTGACCCTGAGAAAAGTTCACGGGTTCGTGCT
GGAGCTCAAGAAAAAATTTGGGGTCGAGTGGGCTGTA
GTTTTCAGGTTGGAATAGAATGGGGCTCATTTTGATCT
CCTGCTTTCTGTAAGGGATATATATATATATATATAT
GTATGTATGTATGTATGTATGTATGTATGTATGTAT
GTATGTATGTATGTATGTATGTATGTATGTATGTAT
TTGTGATTAAGAACCTAATAGCACGTTCTTCGCCTAATA
ATGTGTAATTATAAAGTTGCGGAACACTGCTCCACGTC
ATGTCTGTAATTCTTCAAAAAAAAAAAAAAAAAAAAAA
AAAA

```

Figura 3.11: Motivo microsatélite TATG en la secuencia EST CT988111. Se indican: el motivo de repetición (en rojo) y la ubicación de los oligonucleótidos diseñados (flechas).

Gen de glutamato descarboxilasa (GAD)

Glutamato descarboxilasa (GAD) es una enzima que cataliza la conversión de L-glutamato a ácido gamma-aminobutírico (GABA), un aminoácido no proteico presente en células procariontas y eucariotas.

A GABA se le atribuyen diversas funciones, tales como: acumulación y transporte de nitrógeno, regulación del pH citosólico, protección contra estrés oxidativo, osmoregulación, defensa contra insectos, crecimiento del tubo polínico, entre otras.

En plantas y animales, GABA es principalmente metabolizado en una vía corta compuesta por tres enzimas: una enzima citosólica denominada glutamato descarboxilasa (GAD) y dos enzimas mitocondriales denominadas GABA transaminasa (GABA-T) y semialdehído succínico deshidrogenada (SSADH).

El calcio (Ca^{2+}) y la calmodulina (CaM) tienen una función crítica en la señalización en plantas y se encuentran involucrados en la regulación del crecimiento y en respuesta a cambios ambientales. En plantas, varias enzimas son reguladas por Ca^{2+} y su molécula sensora calmodulina, entre las que se encuentra GAD.

La actividad de GAD ha sido caracterizada *in vitro* en extracto de proteínas de variadas especies y tejidos, la cual es regulada por pH y la unión de Ca^{2+} /calmodulina a su dominio C-terminal (Gut y col. 2009). Baum y col. (1996), al utilizar plantas que expresaban una enzima GAD silvestre o una GAD mutante sin el dominio de unión a CaM, encontraron evidencias *in vivo* de la importancia de esta unión para la regulación de GAD y su importancia en el desarrollo de la planta.

Los primeros genes de GAD identificados fueron en plantas de petunia (Baum y col.1993), tomate (Gallego y col.1995), tabaco (Yu y Oh, 1998) y *A. thaliana* (Turano y Fang, 1998; Zik y col. 1998).

Los análisis *in silico* del genoma de *A. thaliana* revelaron la presencia de cinco genes codificantes de GAD, de los cuales GAD1 se expresa específicamente en tejido de raíz (Bouche y col. 2004).

En este estudio se identificaron dos regiones microsatélite con motivo tetranucleotídico (GAD_A: CAAG Y GAD_B: AAGA) en la secuencia del contig 557 (formado por EST ES595076, ES590002 y ES590669 de *E. globulus* del GenBank), homólogo a GAD de *Ricinus communis* (E value 6×10^{-124}) y a GAD 4 de *Arabidopsis thaliana* (E value 8×10^{-121}) (Figura 3.12)

```

GGGTATTATTAGACAAGTTTCGGGTCGGGATTCCCGGGATCTGGATTAAAGCGAAAGC
AAAACCGTTCCCTTACGCGCACAGACAAGCAAGCAAGCTCTCTCCCTTATAAAGCCAAC
TTCCCTGTCTCCCTCCGCGAGCGTGCCCGTTTCTCTCGTTTCGGTTTTCTTTGTGT
TCGTCGGGTGAGAGAGGAAGGAAGAAAGAGGAGGAAGGATGGTGATCACCACGACC
GTGTCGGAATCCGATGAGAATTTCACTCCACTTTCGCCTCCAGATACGTCCGCAATTCT
CTTCCCAAGTTCAAGATGCCGGAGCAGTCGATACCAGAAAGAAGCAGCTTACCAGATAATA
AACGATGAACTGATGTTGGATGGGAATCCGAGGCTGAAGTTGGCTTCTTCGTGACCACG
TGGATGGAGCCAGAGTGCAGATAAGCTGATGATGGCTGCGATCAACAAGAATTATGTCGAT
ATGGATGAATACCGTGTCCACACAGAGCTCCAGAACCAGGTCGTTAAACATGATAGCCAC
CTGTTCAATGCTCCCATGGGGATAAAGAGAACGCTGTTGGTGTGGGACGGTGGGGTCA
TCTGAGGCTATTATGTTGGCAGGTTGGCAATGAAGAGGAAGTGGCAGAAAAGAAGGCAG
GCTGAAGGAAAACCCATGATAAGCCGAATATTGTTACTGGAGCCAATGTCAGGTTTGC
TGGGAGAAGTTCGCTAGGTACTTTGAGGTCGAGCTGAAGGAAGTGAAGTGAAGTGAAGG
TATTATGTGATGGACCCGGTGAAGCTGTTGAGATGGTCGACGAGAATACCATATGTTG
GCAGCGATACTCGGCTCGACCTTAACTGGAGAATTCGAGGATGGAAGCTCCCTAATGAT
CTCCTACTAAGAAGAACAGGGAACCGGATGGGACACTCCATTACGTGATGCTGCGA
GCGGAGGGTCAATTTGCCGCTGTACCGCATCGAATGGGATTCGTCACCCCTGGTAGAGC
ATACGTAGCGTCACATATGCTGTCATCGTTTCGATGTCCTGAGACAGAAGTCCGAGACTC
TCTCATATACTGATGATCACCT

```

Figura 3.12: Motivos microsatélites encontrados en contig 557 de *E. globulus*. Se indican los motivos denominados GAD_A (en rojo) y GAD_B (en azul).

Gen de peroxidasa (PER)

La peroxidasa es una enzima que cataliza la oxidación de un amplio número de sustratos orgánicos e inorgánicos, como fenoles (guayacol, pirogalol) y aminas aromáticas (o-fenilendiamina) por medio de peróxidos (H_2O_2). El rol de PER en plantas no está completamente identificado, aunque hay evidencias de su asociación a la biosíntesis de pared celular, resistencia a enfermedades y reparación de tejidos dañados (Préstano y Manzano 1993). En la biosíntesis de lignina está implicada en la oxidación de los monolignoles, permitiendo la polimerización de los mismos.

Estructuralmente, la enzima peroxidasa pertenece a una superfamilia que contiene tres grupos principales (I, II, y III). En términos de catálisis, los tres grupos participan en un ciclo que involucra la oxidación por peróxido de hidrógeno seguido de la reducción de diversos compuestos fenólicos (incluyendo monómeros de lignina, obteniéndose las especies radicales libres correspondientes). Sin embargo, sólo la clase I (ascorbato peroxidadas) y clase III se encuentran en las plantas (Blee y col. 2003).

Las peroxidadas de clase III pertenecen a una familia multigénica, cuya evolución parece estar relacionada con el aumento en la complejidad de la arquitectura de la pared celular de las plantas (Passardi y col. 2004).

Se han identificado isoenzimas peroxidadas involucradas en el proceso de lignificación en diversas plantas como tabaco (Lagrimini y col. 1987), álamo (Christensen y col. 1998, 2001;

Sasaki y col. 2004), tomate (Blee y col. 2003), arabidopsis (Demont-Caulet y col. 2010), entre otras.

En este estudio, se identificó una región microsatélite con motivo ATTT en el EST ES588966 de *E. globulus*, similar a la enzima peroxidasa 27 (PER27) (E value 1 E^{-31}) de *Arabidopsis thaliana*.

```

GGGAACTATAGAAACAGTTTCGGTCCGGATTCCCGGGATAATCATAGTTCTTCTTGGCAC
CAAAGAATGGCTTGTCCAAAGCTCCTTGTCTTGTTTGTGCTTCAATTGATGTTATTT
TTTCTTGTCCCTGAGGTTGCTCAAGGTCACGGTCTCAAAGTAGGGTTTTACCAGAAGGC
TGCCCAAACGCAGAGGCCATCGTGAAGGAAGTCATCGGCCAAGTCTTGTCCATCGATCCG
AGTCTCTCCGGCCCTATTTGAGGCTGCACTTTCACGATTGCTTCGTTAGGGGCTGTGAA
GGTTCGGTGCTACTGGAATCCCGACCTATCAATCCGAGAAGTACGCGATTGCGAAGCTT
GGCCTTCGAGGTTACCAAGAGATTGATAGAGTCAAACAGCGTTGGAGAAGGCATGCCCT
GGCATCGTTTCGTGTGCTGACATCTTGGCCCTTGTCACTAGAGAAATTGTCGTGCGGGT
CGCATATTCATAAATCACCTCCTCAATATTTTATCGCCAATTTGATTGTGGAAAGCAAAT
TTTGGCCACACGAATCATAGATGATAAGTACCCACAATTTATTTTGAATGAGA
AAAACATTTTATATAAATATTGGGTTTTTTTTTTTCTTGTGCTGTCGCTTTAGACCGGA
GAACTTTTCTATGATTGGCCACACGTATCATAGAGGATAAGTACCCACAATTTATTTTA
TTTTTTTGAATGAGAAAAACATTTTATATAGATATTGGAGTTTTTTTTTCTTGTGTTGT
CGCTTTAGAC

```

Figura 3.13: Motivos microsatélites encontrados en secuencia ES588966 de *E. globulus*. Se indica el motivo ATTT (en rojo) y la ubicación de los oligonucleótidos diseñados (flechas).

En resumen, en este capítulo se describió la identificación y caracterización de 29 marcadores microsatélites dentro de 24 genes candidatos potencialmente involucrados en calidad de madera en *Eucalyptus*. Los mismos fueron amplificados en una muestra de 8 individuos de *Eucalyptus globulus* de los cuales 17 fueron validados y 12 SSR resultaron polimórficos.

Estos *loci* polimórficos se analizaron en 60 individuos, para estudiar su nivel de polimorfismo y se evaluó su transferibilidad a otras seis especies de *Eucalyptus*. Estos resultados se desarrollan en el capítulo III.

El hallazgo de marcadores moleculares tipo SSR en genes candidatos para calidad de madera no cuenta con precedentes y resulta sumamente interesante para futuros estudios de identificación de genotipos, mapeo genético, co-segregación y mejoramiento asistido por marcadores.

Resultados y Discusión

Capítulo III

**Estudio de la capacidad de los marcadores
funcionales seleccionados para análisis de
variabilidad poblacional y su transferibilidad
a otras especies del género**

3.3 Capítulo III Evaluación de la capacidad para el análisis de variabilidad poblacional del conjunto seleccionado de marcadores funcionales y su transferibilidad a otras especies del género

3.3.1 Estudio de los patrones de distribución de los alelos de EST-SSRs y GC-SSRs de *E. globulus*

Entre los *loci* validados, se seleccionaron 29 polimórficos (17 EST-SSRs y 12 GC-SSR), cuya amplificación, mediante oligonucleótidos específicos, produjeron un patrón de bandas claras y reproducibles (Tablas 3.2 y 3.5). Para estimar el potencial de estos marcadores, se evaluó una muestra de 60 individuos no emparentados de *E. globulus*, provenientes de distintas razas geográficas. Si bien esta colección es representativa de la distribución natural de la especie en Australia, estos individuos son parte de un ensayo de campo y por lo tanto no constituyen una población natural. Sin embargo, el cálculo de estimadores que evalúan aspectos de genética de poblaciones, es de suma utilidad para valorar el potencial de los marcadores y la evaluación del comportamiento de los mismos.

En la Tabla Anexa 4 se muestran los tamaños de los alelos (en pb) de cada *locus* para cada individuo. Estos 29 marcadores fueron capaces de detectar un total de 219 alelos diferentes en la población, en un amplio rango de tamaños (83 pb a 465 bp para los *loci* GST y HMTi, respectivamente), mostrando un alto nivel de polimorfismo. El número de alelos por *locus* varió entre 2 y 16, con un promedio de 7,55 (d.e. 4,59). El número máximo (16 alelos) se encontró al amplificar los SSRs del EST 157 y del *locus* LIM (éste último posee el mayor número de repeticiones evaluadas *in silico* (AG)₂₂), ambos fuera de los ORFs. Por otro lado, el menor número de alelos (2) se encontró en los *loci* EST5, EST33 y EST88 (Tabla 3.7), los cuales se encontraron dentro de los ORFs.

Las frecuencias alélicas para cada *locus* se muestran en la Figura 3.14

Figura 3.14 Frecuencias alélicas en 12 GC-SSR y 17 EST-SSRs

El histograma muestra las frecuencias alélicas de los *loci* en 60 individuos de *E. globulus*. Se indica, en línea verde, la frecuencia 0,5.

Como puede observarse en el gráfico, la mitad de los *loci* (14) (pCCR, HMTe, XTH2, GAD_A, GAD_B, EST5, EST11, EST33, EST58, EST88, EST125, EST141, EST197 y EST200) presentó un alelo más frecuente (con una frecuencia mayor a 0,5). En nueve de estos *loci*, el alelo más frecuente fue el de menor tamaño (menor número de repeticiones) (Figura 3.14), sugiriendo que éstos son los más conservados en estos genes. Otra posible explicación del predominio del alelo de secuencia más corta es que en los individuos heterocigotas que presentan un alelo de secuencia corta y otro de secuencia más larga, sólo el alelo corto se detectó (Wattier y col. 1998). Esto pudo haber ocurrido en este estudio ya que la mayoría de los alelos cortos más frecuentes aparecen en estado homocigota, especialmente en los *loci* pCCR, GAD_A, EST125 y EST197 en donde se encontraron frecuencias de alelos nulos estimadas mayores a 0,10 (ver más adelante en Tabla 3.7).

Del mismo modo, en cada *locus* también fueron detectados otros alelos con muy baja frecuencia.

En *Eucalyptus*, existe una alta probabilidad de mantener alelos en frecuencias bajas o alelos raros (<5%) dentro de una población y las mutaciones que producen alelos raros se mantienen en las poblaciones a bajas frecuencias por un gran número de generaciones (Grattapaglia y Kirst 2008).

En este estudio, se encontraron alelos únicos (presentes en sólo un individuo, en estado homocigota o heterocigota) en todos los *loci* excepto XTH, EST3, EST11, EST88 y EST197, como puede observarse en la tabla 3.6.

Tabla 3.6 Alelos únicos por *locus* en la población estudiada. Se indica el *locus* (primera columna), proporción (nº de alelos únicos/ nº alelos totales de ese *locus*) (segunda columna) y el tamaño del alelo en pb (tercera columna). La mayoría de éstos están en estado heterocigota, excepto los marcados con asterisco.

<i>locus</i>	frecuencia	alelo (pb)	<i>locus</i>	frecuencia	alelo (pb)
pCCR	0.364	194*	GAD_A	0.33	252
		198*			257
		210	GAD_B	0.50	175
		214			186
CesA3_A	0.385	285	PER	0.13	262
		293	EST1	0.17	151*
		299	EST5	0.50	339
		302	EST33	0.50	136
		307	EST46	0.25	213
CesA3_B	0.231	360	EST58	0.33	124
		361			146
		367	EST 125	0.25	220
LIM1	0.250	164	EST134	0.11	299
		179	EST 141	0.33	315
		184	EST157	0.13	302
		194			306
HMTi	0.231	415	EST191	0.30	204
		442			226
		465			228
HMTe	0.333	222*	EST200	0.33	337
SK	0.400	332*	EST216	0.20	409
		352			412
		353	EST219	0.36	102
		369			114
		371			122
		386			124
GST	0.083	83			

Para los *loci* CesA3_b, HMTi, GST, PER, EST46, EST125, EST141, EST191, EST200 y EST216, los alelos menos frecuentes (únicos) fueron los de menor o mayor tamaño del conjunto encontrado (menor o mayor número de repeticiones del motivo) (Figura 3.14), mostrando que en estos *loci* las mutaciones que produjeron alelos de tamaños extremos, fueron las menos frecuentes.

Los *loci* GAD_B, EST5 y EST33 presentaron la mayor proporción de alelos únicos. En estos casos, el alto porcentaje se debe a que los 3 *loci* tienen muy pocos alelos (dos los EST5 y EST33 y cuatro GAD_B). Por otro lado, *locus* GST presentó el menor porcentaje (8,3%) de alelos únicos.

Cabe destacar, que la clasificación como alelo raro es aún preliminar debido al número de individuos utilizado para este análisis (60) y además que éstos no provienen de una selección al azar (ver MyM, sección 2.1).

3.3.2 Estudio de los niveles de polimorfismo de EST-SSRs y GC-SSRs de *E. globulus*

El nivel de polimorfismo de los 29 *loci* SSR analizados puede observarse en la tabla 3.7

Tabla 3.7 Indicadores del potencial de los 29 *loci* SSRs en términos de la diversidad genética detectada en la muestra de 60 individuos. Se indica motivo de repetición, N: tamaño de la muestra, Na: número de alelos, PIC: índice de contenido polimórfico, PI: probabilidad de identidad, H_o: Heterocigosidad observada, UH_e: Heterocigosidad esperada inesgada, F: índice de fijación, P-val EHW: p valor para la prueba de equilibrio de Hardy-Weinberg y Frec Nul: frecuencia de alelos nulos.

<i>locus</i>	motivo	N	N _a	PIC	PI	H _o	UH _e	F	P-val EHW	frec nul (beta)
pCCR	2	46	11	0,616	0,151	0,304	0,640	0,519	0,0000	0,187 (0,002)
CesA3_a	2	54	13	0,749	0,076	0,519	0,781	0,329	0,0000	0,095 (0,002)
CesA3_b	2	53	13	0,808	0,050	0,698	0,837	0,158	0,0000	0,053 (0,002)
LIM1	2	55	16	0,860	0,028	0,782	0,881	0,104	0,0699	0,041 (0,005)
HMTi	2	55	13	0,791	0,057	0,727	0,821	0,106	0,0345	0,043 (0,002)
HMTe	3	55	3	0,159	0,702	0,145	0,169	0,134	0,0155	0,074 (0,002)
SK	2	55	15	0,758	0,073	0,582	0,791	0,258	0,0001	0,085 (0,002)
XTH2	5	48	3	0,302	0,471	0,354	0,354	-0,012	1,0000	0,054(0,0001)
GST	4	57	12	0,849	0,033	0,825	0,871	0,045	0,5072	0,034 (0,001)
GAD_A	4	59	6	0,462	0,288	0,390	0,514	0,234	0,0203	0,097 (0,001)
GAD_B	4	59	4	0,255	0,541	0,237	0,289	0,173	0,3795	0,073 (0,001)
PER	4	49	8	0,738	0,084	0,490	0,776	0,362	0,0003	0,155 (0,001)
EST1	2	48	6	0,645	0,143	0,313	0,702	0,550	0,0000	0,243 (0,001)
EST3	5	52	5	0,621	0,158	0,212	0,678	0,685	0,0000	0,280 (0,001)
EST5	3	57	2	0,017	0,966	0,018	0,018	-0,009	-	0,078 (0,001)
EST11	2	54	5	0,620	0,156	0,500	0,667	0,243	0,0446	0,115 (0,001)
EST33	4	57	2	0,017	0,966	0,018	0,018	-0,009	-	0,081 (0,001)
EST46	4	60	4	0,562	0,205	0,417	0,636	0,340	0,0001	0,145 (0,001)
EST58	5	52	6	0,546	0,212	0,519	0,598	0,123	0,6156	0,064 (0,001)
EST88	3	57	2	0,034	0,933	0,035	0,035	-0,018	1,0000	0,078 (0,001)
EST 125	3	52	4	0,361	0,397	0,231	0,413	0,435	0,0001	0,157 (0,001)
EST134	2	51	9	0,731	0,084	0,471	0,763	0,377	0,0000	0,182 (0,001)
EST 141	3	51	3	0,279	0,502	0,255	0,334	0,230	0,1355	0,097 (0,001)
EST157	2	53	16	0,873	0,025	0,642	0,892	0,274	0,0000	0,128 (0,001)
EST191	2	45	10	0,738	0,085	0,622	0,780	0,193	0,0060	0,119 (0,001)
EST197	3	58	4	0,511	0,247	0,328	0,596	0,446	0,0000	0,163 (0,001)
EST200	2	41	3	0,166	0,688	0,195	0,180	-0,095	1,0000	0,062 (0,001)
EST216	3/2	42	10	0,717	0,09	0,667	0,747	0,097	0,0054	0,080 (0,001)
EST219	2	57	11	0,707	0,103	0,912	0,752	-0,223	0,1553	0,015 (0,001)
media			7,55	0,534		0,304	0,640	0,519		
d,e,			4,59	0,272		0,519	0,781	0,329		

El índice de contenido polimórfico (PIC) varió en un amplio rango (0,017 a 0,873), con un valor promedio de 0,534 (d.e. 0,272). Sin embargo, 21 *loci* SSRs presentaron valores de PIC mayores a 0,5 que correspondieron a aquellos que poseían más de 4 alelos. Los valores de PIC de los marcadores con motivo de repetición dinucleotídicos fueron mayores (0,699) a los de motivo tri- (0,227), tetra- (0,481) y pentanucleotídicos (0,490). Resultados similares se encontraron en dos estudios realizados por Faria y col. (2010a,b) en donde los SSRs con motivos dinucleotídicos presentaron los mayores valores de PIC (0,724 para di-, 0,658 para tri- y un promedio de 0,468 para tetra-, penta- y hexanucleotídicos).

Al analizar los genes candidatos para calidad de madera, se observó que los *loci* con valores de PIC menores a 0,50 fueron GAD_A, GAD_B, HMTe y XTH, cuyos SSRs se encontraban dentro de los exones, lo cual explicaría su menor variabilidad. Por otro lado, los *loci* con PIC mayores a 0,75 correspondieron a LIM1 y CesA3_A (SSRs en el promotor), HMTi (SSR en el intrón) y GST (SSR en el exón). Excepto para GST, el elevado valor de PIC de estos *loci* tendría relación con la mayor variabilidad que presentan regiones que no se expresan (en zonas de promotores e intrones), es decir que estarían sujetas a menor presión de selección.

Otro estimador muy útil para evaluar el potencial de un marcador, es su capacidad de discriminación de individuos. El mismo estima la Probabilidad de Identidad (PI), y calcula la probabilidad de que dos individuos no relacionados presenten el mismo genotipo para ese *locus*. La PI varió entre 0,025 y 0,702, excepto para los EST5, EST33 y EST88, que mostraron valores cercanos a 1, y por lo tanto con muy bajo poder de discriminación (la genotipificación de la mayoría de los individuos para cada uno de los *loci* resultó igual).

La probabilidad de identidad combinada (calculada como el producto de las PI individuales) resultó igual a $7,19 \times 10^{-23}$, similares a los descriptos para Faria y col. (2010b) en donde se utilizaron 21 marcadores EST-SSRs. A su vez, 18 de los *loci* evaluados presentaron un PI inferior a 0,15 indicando que pueden utilizarse para la identificación de individuos, tomando conjuntamente a todos o en grupos parciales de menor número de marcadores.

Al analizar los valores de H_o y U_{H_e} en todos los *loci* se encontró que éstos variaron en un amplio rango (0,018 a 0,912 y 0,018 a 0,881, respectivamente). El valor promedio de ambas fue 0,304 y 0,640, similares a los encontrados en otros trabajos de estudios de variabilidad genética con marcadores EST-SSRs en *Eucalyptus* (Faria y col. 2010 a,b). Sin embargo, 13 y 22 *loci* SSRs presentaron valores de H_o y U_{H_e} mayores a 0,5, respectivamente. Nuevamente, los SSRs con motivos de dinucleótidos fueron los más polimórficos ($H_o=0,567$ y $U_{H_e}=0,731$), mientras que los de motivos trinucleotídicos resultaron los menos polimórficos ($H_o=0,169$ y

$U_{H_e}=0,261$). Esta clara distinción en el contenido de información genética entre los marcadores SSR di- y trinucleotídicos es un fenómeno también observado en otras especies de plantas (Vigouroux y col. 2002).

Para los 29 *loci* evaluados en 60 individuos, la heterocigosidad esperada insesgada (U_{H_e}) bajo el supuesto de Equilibrio de Hardy-Weinberg (EHW) resultó mayor a la heterocigosidad observada (H_o), excepto para los *loci* XTH2, EST5, EST33, EST88, EST200 y EST219. Consecuentemente con esto, los índices de fijación (F) variaron entre -0,223 y 0,685, mostrando una clara evidencia de un exceso de individuos heterocigotas en los seis *loci* antes mencionados y un déficit de individuos heterocigotas en los 23 *loci* SSRs restantes, respecto de los valores estimados para el EHW.

Al realizar el test de EHW para los 29 *loci*, se encontró que 14 de ellos mostraron un significativo déficit de heterocigotas $P < 0,01$ (Tabla 3.7).

La desviación del EHW puede deberse a una variedad de causas. Como se encontró en este estudio, algunos *loci* mostraron un exceso de heterocigotos ($F < 0$), lo cual podría atribuirse eventualmente a la presencia de una selección sobredominante.

Por el contrario, el déficit de heterocigotas encontrado en la mayoría de los *loci* podría deberse a:

1. altos niveles de consanguinidad en la población
2. el *locus* está bajo selección,
3. presencia de alelos nulos en esa población,
4. la presencia de una cierta subestructuración reproductiva de la población (efecto «Wahlund») (Nei, 1987).

Se analizaron por separado cada una de las hipótesis y se detallan a continuación los resultados de dicho análisis:

1. Debido a que, como se dijo previamente, en este estudio se analizaron individuos no emparentados, provenientes de diferentes familias la hipótesis 1 antes mencionada difícilmente puede explicar el desvío del EHW encontrado. Entonces, este desvío podría deberse a que los *loci* están bajo presión de selección, a la presencia de alelos nulos o por la presencia de subestructura genética causada por el efecto de muestreo.

2. Para analizar la hipótesis 2, se realizó el test de neutralidad Ewens-Watterson. Se encontró que los 29 *loci* no se desviaron significativamente de la neutralidad esperada ($\alpha = 0,001$). Esto

se evidenció por el hecho de que el valor de F observado fue similar a la media del F-estadístico simulado (mediante 1000 permutaciones) y se encontró dentro del intervalo de confianza L95-U95 (Tabla 3.8).

Tabla 3.8 Test de Neutralidad de Ewens-Watterson. Se indica: *Locus*: nombre del *locus*, *n*: número de *thali* empleados en el test, *k*: número de alelos, *F obs*: F-estadística observada, *F prom*: valor promedio de la F simulada, *ES*: error estándar de la F simulada, *L95*: intervalo de confianza del 95% inferior, *U95*: intervalo de confianza del 95% superior.

<i>Locus</i>	<i>n</i>	<i>k</i>	<i>F obs</i>	<i>F prom</i>	<i>ES</i>	<i>L95</i>	<i>U95</i>
pCCR	92	11	0,3674	0,2361	0,0061	0,1392	0,4336
CesA3a	108	13	0,2267	0,2109	0,005	0,1217	0,3944
CesA3b	106	13	0,1712	0,2087	0,0048	0,1221	0,3957
LIM1	110	16	0,1274	0,1669	0,0028	0,1008	0,3063
HMTi	110	13	0,1868	0,2104	0,005	0,1212	0,3957
HMTe	110	3	0,8321	0,6701	0,0335	0,3706	0,9641
SK	110	15	0,2162	0,1804	0,0035	0,1086	0,3299
XTH2	96	3	0,65	0,6682	0,0319	0,3661	0,959
GST	114	12	0,1367	0,2362	0,0076	0,1291	0,4632
GADa	118	6	0,4908	0,4364	0,022	0,2346	0,781
GADb	118	4	0,7132	0,5736	0,0316	0,3128	0,9178
PER	98	8	0,2318	0,3282	0,0125	0,1847	0,6258
EST1	96	6	0,3049	0,4249	0,0202	0,2292	0,753
EST3	104	5	0,329	0,4939	0,0263	0,2679	0,8713
EST5	114	2	0,9826	0,8069	0,0282	0,5038	0,9826
EST11	108	5	0,3393	0,4914	0,0248	0,2719	0,8278
EST33	114	2	0,9826	0,8187	0,0279	0,5014	0,9826
EST46	120	4	0,369	0,5748	0,033	0,3068	0,9192
EST58	104	6	0,4079	0,4272	0,0212	0,2406	0,7722
EST88	114	2	0,9655	0,817	0,0268	0,5038	0,9826
EST125	104	4	0,5913	0,5726	0,0295	0,3134	0,9072
EST134	102	9	0,2445	0,3007	0,0107	0,1672	0,5656
EST141	102	3	0,6692	0,6811	0,0335	0,3754	0,9614
EST157	106	16	0,1162	0,1676	0,0028	0,1025	0,3106
EST191	90	10	0,2291	0,2614	0,0072	0,1496	0,4822
EST197	116	4	0,4092	0,5856	0,0304	0,3294	0,9163
EST200	82	3	0,8218	0,6576	0,0324	0,3623	0,9521
EST216	84	10	0,2616	0,2607	0,008	0,1514	0,4915
EST219	114	11	0,2542	0,2567	0,008	0,1457	0,4974

3. Para analizar la hipótesis 3 como probable causa del desvío del EHW se estudió la posible existencia de alelos nulos en los 29 *loci* en la población. Los alelos nulos pueden ocurrir cuando existen mutaciones en las regiones flanqueantes de los microsatélites, evitando el anillamiento de los oligonucleótidos por falta de complementariedad suficiente.

Ante la presencia de alelos nulos en un *locus* particular, la heterocigosidad observada sería en gran medida subestimada, afectando especialmente las estimaciones de parámetros poblacionales que se basan en la proporción de heterocigotas, tales como el coeficiente de consanguinidad F de Wright. Los alelos nulos pueden detectarse experimentalmente de varias maneras:

- podrían inferirse de los resultados de cruzamientos controlados (Pastorelli y col. 2003), sin embargo este método no es eficiente para evaluar la frecuencia de alelos nulos en una población debido a que éstos están en general, distribuidos aleatoriamente entre los genotipos.
- se podrían rediseñar los oligonucleótidos en otras secuencias genómicas adyacentes, sin embargo esto requiere de un mayor costo y esfuerzo y podría ser insuficiente si las nuevas regiones en donde se diseñen no son lo suficientemente conservadas.
- se podría estimar su frecuencia utilizando la información disponible de las proporciones genotípicas.

En este estudio, se optó por la estimación de dichas frecuencias mediante el programa INEST1.0 (Chybicki y Burczyk 2009) el cual discrimina alelos nulos de aquellos alelos no amplificados por fracaso de la técnica de PCR, a través del parámetro beta (β).

La frecuencia de alelos nulos estimada varió entre 0,017 y 0,281. Los menores valores correspondieron a los *loci* EST5 y EST33, los únicos con 2 alelos en esta población. Por otro lado, los valores mayores correspondieron al EST3, el cual también presentó el mayor Índice de Fijación ($F=0,685$).

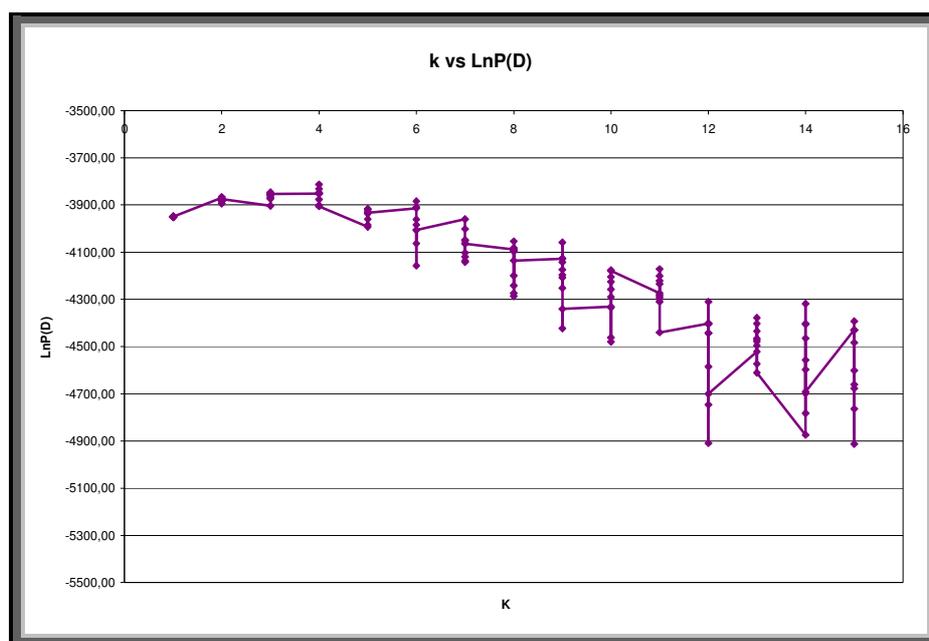
Además, la probabilidad β de la falta de amplificación al azar estimada para cada *locus* varió entre el 0,1% y 0,5%, lo cual indicó que las frecuencias encontradas no son consistentes con el fracaso al azar de la amplificación, sino con la presencia de verdaderos alelos nulos.

La aparición de alelos nulos en *loci* microsatélites se han estudiado en numerosos organismos, tales como humanos (Callen y col. 1993), mariposa (*Polyommatus bellargus*) (Harper y col. 2003), incluyendo especies forestales como pino (Vogl y col. 2002), picea (Nascimento de Souza y col. 2005), haya europea (*Fagus sylvatica* L.) (Chybicki y col. 2008) y eucaliptos (Glaubitz y col. 2001, Faria y col. 2010 a,b).

Los resultados encontrados en este estudio coinciden con aquellos publicados por otros autores, en donde se encontró también que la mayoría de los *loci* que se apartaron del EHW, presentaron frecuencias de alelos nulos entre 0,10 y 0,30 (Faria y col. 2010 a,b; Harper y col. 2003).

4. Para analizar la hipótesis 4 se realizó la asignación probabilística de los individuos a un número determinado de grupos (clusters) utilizando un enfoque Bayesiano implementado por el software STRUCTURE (Pritchard y col. 2000). El análisis se realizó sin información *a priori* sobre las razas de *E. globulus*. Al graficar el estadístico ΔK , basado en la tasa de cambio de la probabilidad (Evanno y col. 2005), variando K de uno a 15, se encontró un número K de poblaciones ideales igual a cuatro (K=4) (Figura 3.15).

A)



B)

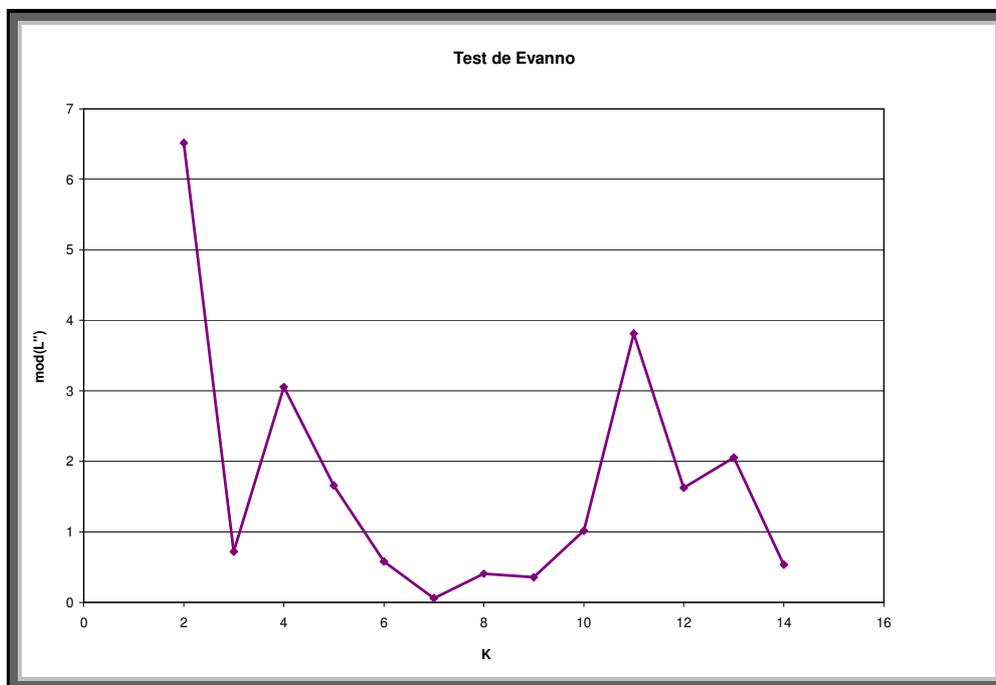


Figura 3.15 Análisis de los resultados del programa STRUCTURE para encontrar las k poblaciones ideales que forman la muestra de *E. globulus* analizada.

A) logaritmo de la probabilidad *a posteriori* (L) como función del número de poblaciones ideales evaluadas ($k=1$ a 15). Los distintos puntos representan las corridas independientes para cada valor k.

B) Δk calculado de acuerdo a Evanno y col. (2005), en función de los números de poblaciones ideales evaluadas ($k=1$ a 15).

El valor modal de la distribución (Figura 3.15-B) es considerado como el número real de poblaciones k. En este caso se observaron dos valores modales a $k=4$ y $k=10$. Sin embargo, al analizar el gráfico de la Figura 3.15-A se observó que para $k=4$, los valores del logaritmo de la probabilidad ($\ln(P)D$), tienden a estabilizarse (meseta) para comenzar a descender con valores más dispersos a mayores valores de k. Por lo tanto, se tomó $k=4$ como número real de grupos o "clusters". Los gráficos correspondientes a la corrida que mostró la mayor probabilidad *a posteriori* de los datos para $k=4$ se observan en la Figura 3.16.

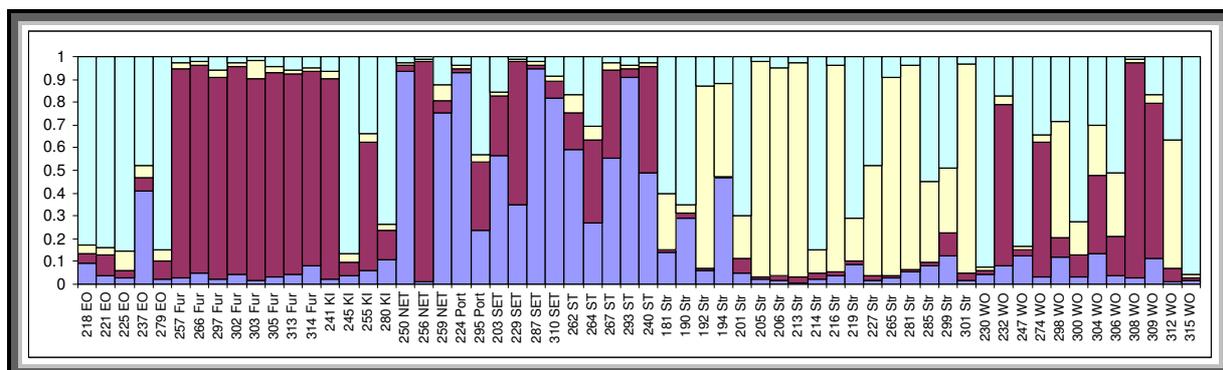


Figura 3.16: Resultado de la inferencia de estructura poblacional utilizando el programa STRUCTURE. Cada individuo está representado por una línea vertical particionada en k segmentos de colores cuyo largo es proporcional al coeficiente individual de pertenencia a cada uno de los k grupos inferidos. Se indica cada individuo con raza geográfica de pertenencia (EO: Eastern Otways, Fur: Furneaux, KI: King Island, NET: NE de Tasmania, Port: raza local de Portugal, SET: SE de Tasmania, ST: S de Tasmania, Str: Strzelecki, WO: Western Otways).

Los grupos formados en este estudio también fueron encontrados por nuestro grupo de trabajo al analizar una población de 194 individuos de *E. globulus* utilizando 2300 marcadores DArT (datos no mostrados).

Por lo tanto, el déficit de heterocigotas encontrado en la mayoría de los *loci* podría deberse a la presencia de la subestructuración (efecto «Wahlund») (Nei, 1987) y a la existencia de alelos nulos.

Independientemente de cuál sea explicación la correcta, el déficit de individuos heterocigotas se ha observado en otros estudios utilizando SSRs en *Eucalyptus* (Marcucci Poltri y col. 2003); (Zelener y col. 2005)) y otras especies forestales (*Pseudotsuga menziesii* (Slavov y col. 2004), *Picea abies* (Nascimento de Sousa y col. 2005), *Picea asperata* (Wang y col. 2005).

3.3.3 Aplicación de los SSRs para el estudio de relación genética entre individuos de *E. globulus*

Como se mencionó previamente, los marcadores funcionales estudiados mostraron un gran potencial para discriminar individuos. Esto fue evidenciado también mediante el cálculo de distancias genéticas entre pares de individuos de la muestra evaluada y su respectivo análisis de agrupamiento.

Se construyó la Matriz de Distancia a partir de la Matriz Básica de Datos (incluyendo todos los *loci* analizados) mediante el programa Populations 1.2.30, utilizando un coeficiente de distancia

DAS (shared allele distance) (Chakraborty y col. 1993) que permitió medir el grado de diferencia entre todos los pares de OTUs (individuos) en relación a las variables estudiadas (alelos).

Los valores de distancia entre individuos variaron entre 0,267 (individuos 267 y 293, ambos procedentes del Sur de Tasmania) y 0,744 (individuos 237 y 264, procedentes de Eastern Otways y Sur de Tasmania, respectivamente), con un promedio de 0,52 (d.e 0,07). En el histograma de los valores obtenidos en la matriz de distancia se puede observar en la Figura 3,17, en donde se detectó que el 66,5% de los valores de distancia fueron mayores a 0,5.

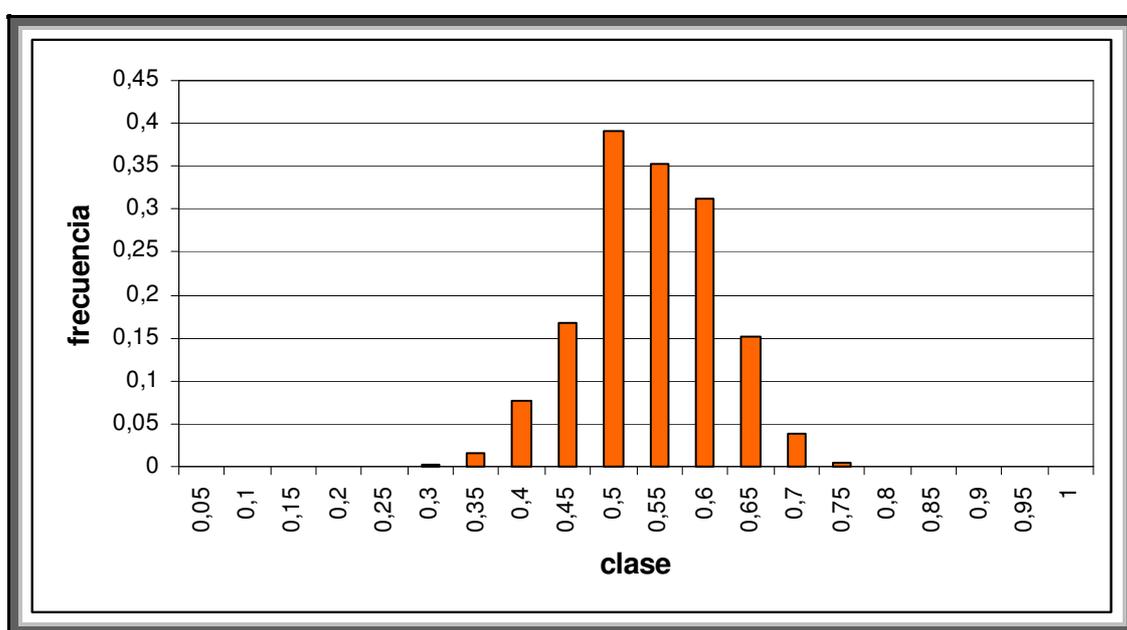


Figura 3.17 Histograma de las distancias entre pares de individuos de *E. globulus* (DAS)

Los agrupamientos de individuos, generados por el método de ligamiento descrito anteriormente (sección 2.10.3 de MyM), originaron el siguiente dendrograma (Figura 3.18).

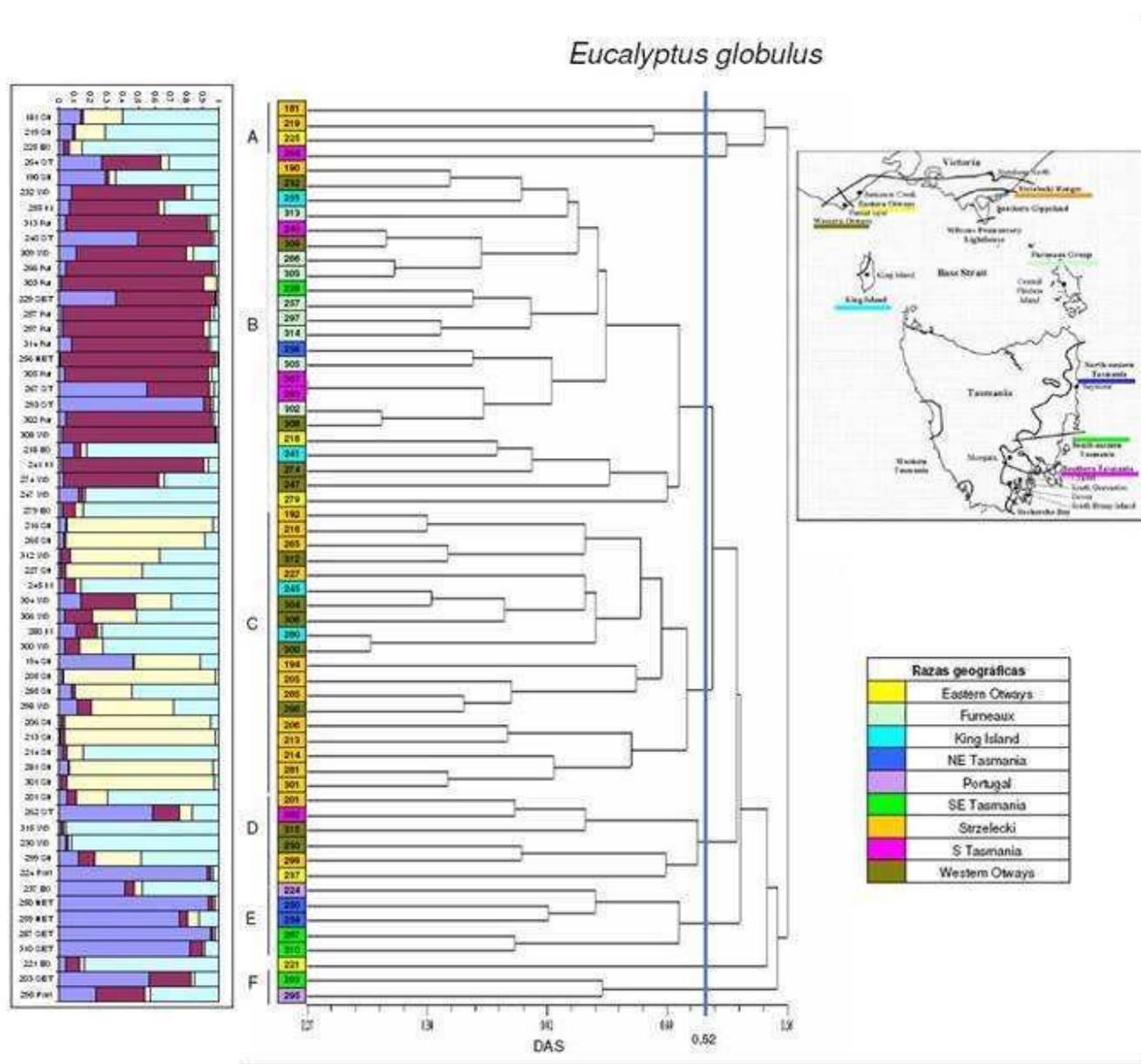


Figura 3.18: Dendrograma generado a partir de la matriz de distancia entre individuos y mapa de Australia indicando las razas analizadas. Se indica la distancia promedio (línea azul) y los grupos formados (A-F) con la clasificación de los individuos según su raza geográfica. A la izquierda se muestra la composición genética de cada individuo según la Figura 3.16.

Los análisis del dendrograma y de la composición genética de cada individuo estimada mediante STRUCTURE, mostraron una coincidencia en los principales grupos formados.

De los 60 individuos analizados mediante STRUCTURE, 33 presentaron una buena asignación (constitución genética mayor al 80%) a uno de los cuatro grupos formados. Al analizar estos 33 individuos, se observó, en términos generales, que el grupo representado en color verde estuvo formado por individuos del Suroeste de Victoria (Eastern Otways y Western Otways), mientras que el grupo representado en color amarillo, estuvo formado por individuos de la raza geográfica Strzelecki (que se corresponden al grupo C del dendrograma). Estas tres tienen un origen geográfico continental cercano y ya se ha demostrado en otros trabajos que poseen afinidad genética, utilizando tanto parámetros genéticos individuales de densidad de madera (López y col. 2001) como marcadores moleculares (Jones y col. 2006, Steane y col. 2006).

El grupo representado en azul estuvo formado por individuos de la Isla de Tasmania (NE, SE y S), junto con los dos individuos proveniente de la llamada raza local de Portugal, presentes en los grupos E y F del dendrograma. La afinidad genética de las razas de la Isla de Tasmania ya ha sido demostrada por otros autores (Jones y col. 2006, Jordan y col. 1994, López y col. 2001). Además el agrupamiento obtenido entre razas locales de Portugal con nativas de Tasmania se correspondió con el encontrado en estudios anteriores (Acuña y col. 2005; Torales y col. *en preparación*) realizado con marcadores neutros AFLP y SSR.

El grupo representado en rojo (el más numeroso) estuvo conformado por individuos de diversas razas geográficas australianas, entre los que se encuentra un sub-grupo con individuos procedentes de Furneaux, presentes en el grupo B del dendrograma.

Los cuatro grupos formados en este estudio también fueron encontrados por nuestro grupo de trabajo al analizar una población de 194 individuos de *E. globulus* utilizando 2300 marcadores DArT (datos no mostrados).

Es decir, los resultados obtenidos mostraron que el conjunto de marcadores funcionales analizados detectaron la misma estructura poblacional que otros marcadores y mediante diferentes análisis.

3.3.4 Transferibilidad de los marcadores funcionales a otras especies de *Eucalyptus*

En virtud de la conservación de las secuencias de ADN en las regiones codificantes del genoma, se espera que una proporción significativa de los pares de oligonucleótidos de EST-SSRs y GC diseñados, funcionen en especies relacionadas.

Con el objetivo de estudiar el potencial uso de estos marcadores diseñados en *E. globulus* en otras especies del género, se amplificaron 36 EST-SSRs validados (polimórficos y monomórficos, excepto el EST219 cuyo patrón no fue claro para su evaluación) y los 12 GC-SSRs en muestras de individuos de *E. camaldulensis*, *E. dunnii*, *E. grandis*, *E. saligna*, *E. tereticornis* y *E. viminalis* (ver sección 2.1 de MyM).

De estos 48 *loci* analizados, 33 fueron transferibles a las otras seis especies estudiadas. A su vez, seis marcadores amplificaron en cinco especies y uno de los marcadores transfirió sólo a tres especies. El resto (ocho *loci*) no transfirió a ninguna especie, y correspondieron a *loci* monomórficos y/o con tamaños mayores a 500 pb en *E. globulus*, los cuales no resultaron informativos para los análisis realizados en este estudio (Tabla 3.9).

En la Tabla Anexa 5 se muestran los tamaños de los alelos (en pb) de cada *locus* para cada individuo. Los alelos encontrados en las otras especies mostraron el mismo rango de tamaños que en las muestras de *E. globulus*, excepto para el *locus* HMTi, donde los alelos presentaron tamaños entre 160 bp y 180 bp, menores a los observados en *E. globulus* (420 bp a 460 bp). Esta diferencia en el largo de los fragmentos amplificados podría deberse a una delección/inserción dentro del intrón que contiene al SSR, aunque esto no ha sido probado en este estudio.

En general, los porcentajes de transferibilidad entre especies resultaron elevados: 77,1% para *E. viminalis* y *E. camaldulensis*, 79,2% para *E. saligna*, 81,3% para *E. tereticornis*, y 83,3% para *E. dunnii* y *E. grandis*.

Al analizar el potencial de transferibilidad de los marcadores polimórficos, se encontró amplificación positiva de los mismos en las otras seis especies, excepto el *locus* SK que no amplificó en *E. camaldulensis*, el EST197 que no amplificó en *E. viminalis* y el EST11 que amplificó sólo en *E. viminalis*, *E. dunnii* y *E. grandis* (Tabla 3.9). A su vez, la mayoría de estos *loci* polimórficos en *E. globulus* resultaron también polimórficos en las otras especies (a pesar del pequeño tamaño de la muestra de cada una de ellas), mostrando su potencial uso en estudios de diversidad genética.

Estos altos porcentajes de transferibilidad se encontraron también en otros trabajos dentro del género *Eucalyptus*. Byrne y col. (1996) publicó un conjunto de cuatro *loci* SSR (desarrollados a partir de *E. nitens*) con conservación completa entre las especies *E. globulus*, *E. grandis* y *E. camaldulensis* dentro del subgénero *Symphyomyrtus* (secciones *Maidenaria*, *Transversaria* y *Exertaria*, respectivamente). Van der Nest y col. (2000) también informó los resultados preliminares de la amplificación interespecífica de cinco *loci* SSR en *Eucalyptus*. Marques y col.

(2002) describieron un conjunto de 180 marcadores SSR cuya transferibilidad fue del 78% entre especies del subgénero *Symphyomyrtus* (*Eucalyptus grandis*, *Eucalyptus urophylla*, *Eucalyptus tereticornis* y *Eucalyptus globulus*), los cuales fueron utilizados para estudiar la homología entre los grupos de ligamiento en estas especies. En ese estudio, aunque se identificaron algunos rearrreglos, se encontró una gran colinearidad entre genomas de especies del subgénero *Symphyomyrtus*. En otro estudio realizado por Steane y col. (2001) se encontró una elevada transferibilidad de los marcadores SSRs diseñados en *E. globulus* a las otras tres especies más estrechamente relacionados (*E. maidenii*, *E. bicostata* y *E. pseudoglobulus*). En ese estudio se encontró que la mayoría de los oligonucleótidos amplificó también en muestras de *E. nitens* (serie *Globulares*), pero el éxito de transferibilidad disminuyó a medida que aumentaba la distancia filogenética con *E. globulus*. La similitud morfológica y la compatibilidad genética en los cruzamientos sexuales entre eucaliptos del mismo subgénero podrían explicar el elevado nivel de conservación en las secuencias de ADN y la alta transferibilidad entre especies relacionadas (Myburg y col. 2007).

Tabla 3.9 Transferibilidad de *loci* evaluados en *E. globulus* a otras seis especies del género. Se indica: nombre del locus, relación nº alelos/nº individuos amplificados para cada *locus* en cada especie. La última columna indica el número de especies en las que transfiere cada *locus* y en la última fila, el porcentaje de *loci* transferibles en cada especie.

<i>locus</i>	<i>E. globulus</i>	<i>E. camaldulensis</i>	<i>E. dunnii</i>	<i>E. grandis</i>	<i>E. saligna</i>	<i>E. tereticornis</i>	<i>E. viminalis</i>	Nº <i>sspp</i>	
POLIMÓRFICOS	pCCR	5/7	7/7	4/6	6/7	6/6	7/7	11/8	6
	Ces3_A	5/7	8/7	5/5	2/7	3/6	7/8	11/7	6
	CesA3_B	6/8	8/7	6/5	6/7	9/6	10/8	8/8	6
	LIM	5/8	8/7	5/5	5/8	8/6	9/8	8/8	6
	HMTi	7/8	9/7	4/5	1/8	4/4	8/6	10/7	6
	HMTe	2/8	1/7	3/6	2/6	1/2	3/8	3/8	6
	SK	7/8	0	7/7	5/7	2/5	4/3	6/8	5
	XTH	3/8	2/7	3/7	1/8	1/6	1/8	1/7	6
	GST	6/8	7/7	6/8	7/8	5/6	8/8	8/8	6
	GAD_A	2/8	4/7	5/6	4/8	6/6	6/7	2/7	6
	GAD_B	2/8	4/7	3/6	3/8	5/6	6/8	4/8	6
	PER	6/8	5/7	3/5	3/8	4/6	7/8	3/6	6
	EST1	4/8	7/6	3/4	3/7	3/4	5/7	1/4	6
	EST3	5/7	2/5	3/5	4/4	5/4	6/6	1/8	6
	EST5	2/8	1/7	3/6	3/6	2/5	1/6	2/7	6
	EST11	3/8	0	4/6	2/6	0	0	3/5	3
	EST33	2/8	4/7	2/8	3/8	2/6	4/8	5/8	6
	EST37	2/8	5/7	1/7	2/8	1/6	3/8	2/8	6
	EST46	3/8	2/7	1/8	2/8	2/6	1/6	1/8	6
	EST58	3/7	3/7	4/7	1/7	2/4	2/8	3/8	6
	EST88	2/8	2/6	2/6	1/7	2/5	2/7	1/7	6
	EST 125	3/8	3/7	3/7	3/6	3/6	6/7	2/3	6
	EST134	4/8	7/7	5/8	5/8	5/6	7/8	3/7	6
	EST 141	2/6	3/6	4/6	2/6	2/5	3/6	3/5	6
EST191	3/7	4/5	5/8	5/8	3/3	6/8	8/8	6	
EST197	4/7	5/7	4/7	3/8	5/6	5/7	0	5	
EST200	2/8	5/4	6/5	4/5	3/3	7/5	4/5	6	
EST216	7/7	6/4	8/5	5/4	8/6	4/5	4/6	6	
MONOMÓRFICOS	EST 2	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6
	EST 9	monom.	0	0	0	0	0	0	0
	EST 12	monom.	0	0	0	0	0	0	0
	EST 13	monom.	0	0	0	0	0	0	0
	EST 25	monom.	0	0	0	0	0	0	0
	EST 36	monom.	0	0	0	0	0	0	0
	EST 60	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6
	EST 61	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6

	<i>locus</i>	<i>E. globulus</i>	<i>E. camaldulensis</i>	<i>E. dunnii</i>	<i>E. grandis</i>	<i>E. saligna</i>	<i>E. tereticornis</i>	<i>E. viminalis</i>	Nº <i>sspp</i>
	EST 68	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6
	EST 80	monom.	monom.	monom.	monom.	0	monom.	monom.	5
	EST 96	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6
	EST 155	monom.	0	0	0	0	0	0	0
	EST 156	monom.	0	monom.	monom.	monom.	monom.	monom.	5
	EST 157	monom.	monom.	monom.	monom.	monom.	monom.	0	5
	EST 181	monom.	0	0	0	0	0	0	0
	EST 203	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6
	EST 205	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6
	EST 208	monom.	0	0	0	0	0	0	0
	EST 210	monom.	monom.	monom.	monom.	monom.	monom.	0	5
	EST 214	monom.	monom.	monom.	monom.	monom.	monom.	monom.	6
	% transferibilidad		77,1	83,3	83,3	79,2	81,3	77,1	

Esta elevada tasa de transferencia resulta interesante ya que aporta al conocimiento genómico de especies en las que no se dispone de una cantidad significativa de marcadores funcionales.

Así, este es el primer trabajo en donde se describen microsatélites disponibles para *E. viminalis*, mientras que para *E. tereticornis* sólo fue reportado un trabajo en donde se validan tres EST-SSRs en una muestra de 12 individuos (Yasodha y col. 2008).

Para realizar un estudio preliminar de la variabilidad de estos marcadores en las otras seis especies estudiadas, se calculó número de alelos (N_a), número efectivo de alelos (N_e) y número de alelos privados (N_p) para cada una de ellas. La media de dichos resultados se graficaron en el histograma de la figura 3.19.

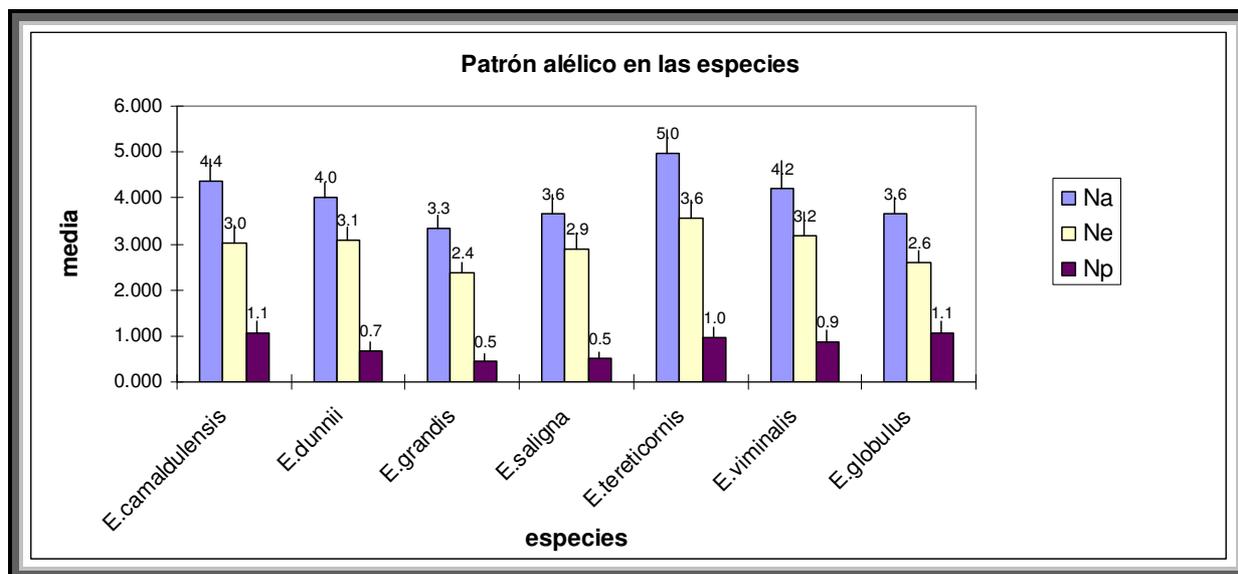


Figura 3.19 Patrón alélico de las siete especies analizadas. Se indica: N_a : número de alelos, N_e : número efectivo de alelos y N_p : número de alelos privados de cada especie.

Si bien este estudio comparativo se realizó con pocos individuos (ocho), se pudo observar un elevado número de alelos y una buena distribución de los mismos (N_a y N_e) en todas las especies. *E. tereticornis* fue la que mostró mayor variabilidad y además, en todas las especies se amplificaron alelos privados (es decir, alelos que aparecen en dicha especie pero no en las otras). Estos alelos especie-específicos podrían ser utilizados en estudios de discriminación entre individuos de dichas especies. Sin embargo, debe considerarse que el número de muestras es pequeño y esta afirmación debe tomarse con cautela. Así, se propone realizar este análisis en una etapa posterior a este trabajo, contemplando la selección, colecta y análisis de un mayor número de individuos por especie.

En este estudio además, se evaluó la posibilidad de incluir estos marcadores en los mapas de ligamiento de *E. grandis*, que se están construyendo como parte de otra tesis doctoral del grupo de trabajo, en donde se utilizan tres cruzamientos de árboles contrastates para densidad de mandera (C182-5 x c32-20; G x 36, K x B). Los resultados encontrados se indican en la tabla 3.10.

Tabla 3.10 Evaluación de *loci* posibles de ser ubicados en los mapas de las tres poblaciones de cruzamientos controlados de *E. grandis*.

Se indican los nombres de los parentales involucrados en cada cruzamiento (C182-5 x c32-20; G x 36, K x B)

cruzamiento	c182-5 x c32-20	G x 36	K x B	cruzamiento	c182-5 x c32-20	G x 36	K x B
pCCR			X	EST5	X	X	
CesA3_A		X		EST11			X
CesA3_B	X	X		EST33	X		
LIM	X	X	X	EST37	X		
HMTi				EST46		X	X
HMTe				EST58			
SK	X		X	EST88			
XTH2				EST 125	X		
GST	X	X	X	EST134	X	X	X
GAD_A	X	X	X	EST 141			X
GAD_B	X	X	X	EST191	X	X	X
PER		X		EST197	X		
EST1			X	EST200			X
EST3			X	EST216			

De estos resultados se puede concluir que la mayoría de los *loci* (22 de los 28) pueden ser ubicados en los mapas de al menos una de las poblaciones segregantes de *E. grandis*. Esto resulta particularmente interesante para los genes candidatos para calidad de madera, debido a que sólo algunos de ellos han sido localizados en los mapas hasta el momento (genes homólogos a celulosa sintasa en eucaliptos (Thamarus y col. 2002; Thuma y col. 2010), CCR (Moran y col. 2002, Thuma y col. 2010).

Este estudio sugiere que los nuevos marcadores funcionales desarrollados pueden ser utilizados para amplificar estos *loci* en otras especies de *Eucalyptus*. Este conjunto de marcadores funcionales está así disponible para estudios de genética de poblaciones y filogenia entre especies del sugénero *Symphyomyrtus* y complementarán estudios realizados en este aspecto (Steane y col. 2011).

3.3.5 Aplicación de los SSRs para la discriminación de los individuos entre las siete especies de *Eucalyptus* estudiadas

Para analizar el potencial de los marcadores desarrollados para discriminar individuos de las siete especies estudiadas, se realizaron tres análisis: el primero mediante índices de distancia, tal como se mencionó previamente, el segundo, mediante el enfoque bayesiano del programa STRUCTURE y el tercero mediante análisis de la variancia molecular (AMOVA).

En la primera evaluación, se construyó la Matriz de Distancia a partir de la Matriz Básica de Datos (incluyendo todas las especies de *Eucalyptus*), utilizando un coeficiente de distancia DAS, de la misma manera que se describe en el ítem 3.2.2.

La distribución de los valores obtenidos en la matriz de distancia se pueden observar en el histograma de la figura 3.20

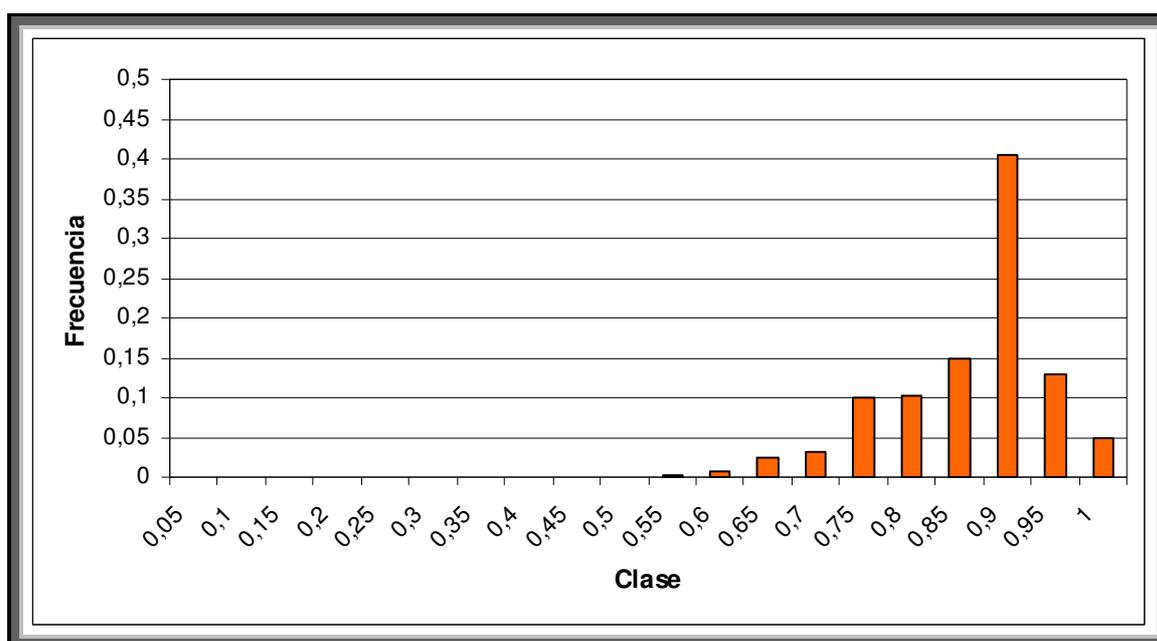


Figura 3.20 Histograma de las distancias DAS entre pares de individuos de las siete especies estudiadas

Los valores de distancia entre individuos variaron entre 0,5 (entre dos individuos de la especie *E. camaldulensis*) y 1 (para dos pares de individuos (*E. dunnii*-*E. grandis* y *E. viminalis*-*E. globulus*), con un promedio de 0,84 (d.e 0,08). Estos valores de distancia genética resultaron mayores a los encontrados en un trabajo de Faria y col. (2010b) (DAS entre 0,214 y 0,972, con un promedio de 0,694) en donde se analizaron un total de 64 individuos de las especies *E. grandis*, *E. globulus*, *E. urophylla* y *E. camaldulensis*, mediante 20 marcadores EST-SSRs.

Como puede observarse en la matriz de distancia triangular (Tabla Anexa 6), los valores de distancia dentro de cada especie fueron, en términos generales, inferiores a los valores entre especies, como era de esperar. Así, la mayoría de los valores de distancias DAS entre individuos de la misma especie resultó menor a 0,8, mientras que la mayoría de los valores

mayores a 0,8 se encontraron entre individuos de diferentes especies, excepto entre individuos de *E. camaldulensis* y *E. tereticornis*, como se grafica en la Figura 3.21.

Con el objeto de visualizar fácilmente las distancias genéticas entre los individuos de las siete especies estudiadas, se realizó la construcción del dendrograma a partir de la matriz de distancia (Figura 3.21).

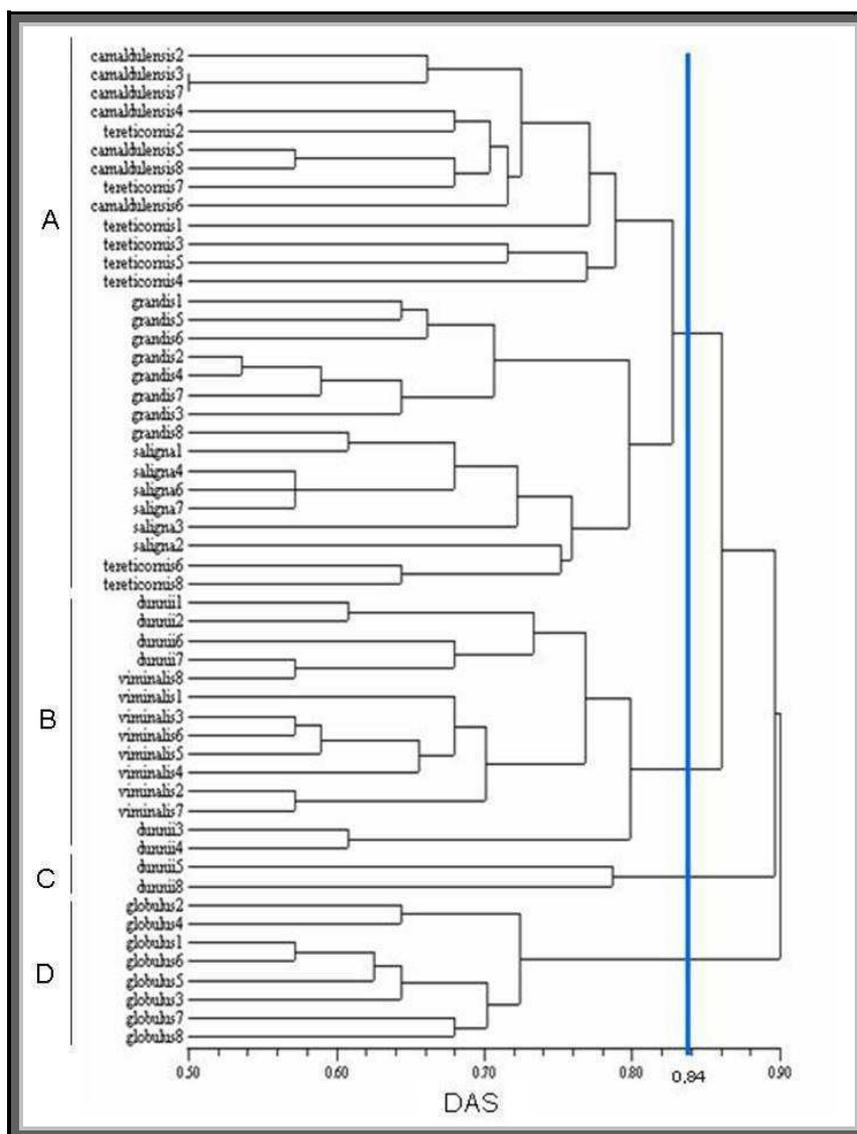


Figura 3.21: Dendrograma generado a partir de la matriz de distancia entre individuos de las siete especies analizadas en este estudio. Se indica la distancia promedio (línea azul).

En términos generales, se pudo observar que los individuos de una misma especie se agruparon entre sí, excepto algunos individuos de *E. tereticornis* y *E. dunnii*.

Al tomar como referencia la distancia DAS promedio (0,84), se obtuvieron cuatro grupos:

Grupo A: formado por dos subgrupos con individuos de *E. camaldulensis*-*E. tereticornis* (sección *Exsertaria*) y *E. grandis*-*E. saligna* (sección *Latoangulatae*), respectivamente.

Grupo B: formado por todos los individuos de *E. viminalis* y seis de *E. dunnii* (sección *Maidenaria*)

Grupo C: formado sólo por los dos individuos de *E. dunnii* que se separaron del grupo B

Grupo D: formado por los ocho individuos de *E. globulus* (sección *Maidenaria*).

En la segunda evaluación se realizó el análisis mediante el programa STRUCTURE. En primera instancia, se evaluó sin asignar información *a priori* sobre la pertenencia de los individuos a cada una de las especies estudiadas. Al graficar el estadístico ΔK , basado en la tasa de cambio de la probabilidad (Evanno y col. 2005), variando K de uno a diez, se encontró un número K grupos igual a cinco (K=5) (Figura 3.22).

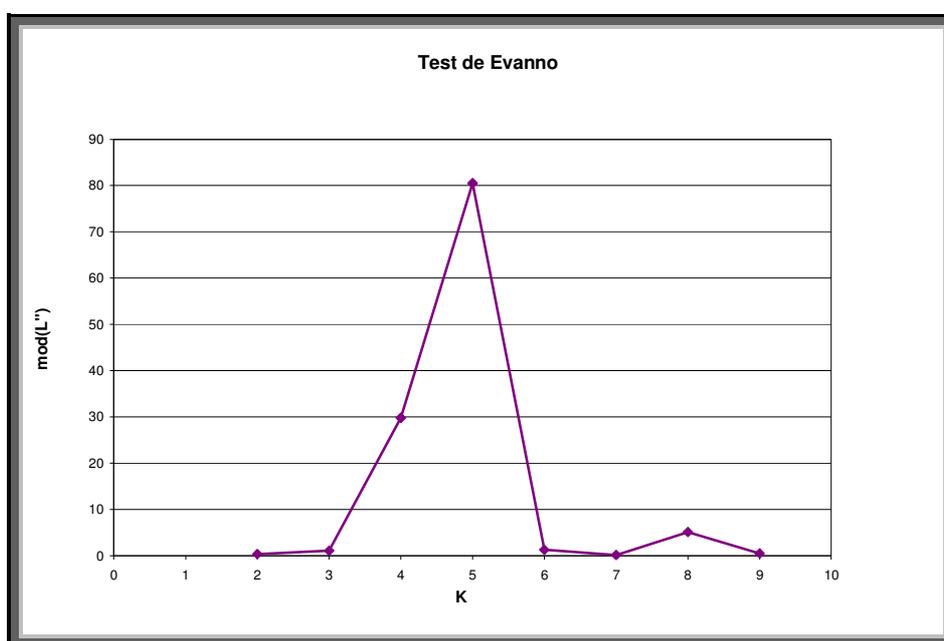


Figura 3.22: Test de Evanno resultante del análisis poblacional mediante STRUCTURE sin información *a priori*

La matriz Q utilizada para graficar los resultados obtenidos para $K=5$ se describe en la Tabla Anexa 7. El gráfico correspondiente se muestra más adelante en la Figura 3.24-A.

En términos generales, los grupos teóricos (caracterizados por color) obtenidos están conformados con los individuos de las siguientes especies:

Anaranjado: formado por individuos de las especies *E. camaldulensis* y *E. tereticornis*

Violeta: individuos de *E. grandis* y *E. saligna*

Verde: individuos de *E. dunnii*

Amarillo: individuos de *E. viminalis*

Azul: individuos de *E. globulus*

La mayoría de los individuos tuvo una buena asignación (constitución genética mayor al 85%) a uno de los cinco grupos formados, excepto en los individuos *E. tereticornis* 3 (Et3 en el gráfico) y *E. saligna* 2 (Es2). Según el histograma de probabilidades, el individuo Et3 tiene una importante contribución genética del grupo representado por el color violeta, mientras que el individuo Es2 tuvo una mayor contribución del genoma caracterizado en el grupo anaranjado. Esto último se corresponde con el dendrograma de la Figura 3.19 en donde el individuo Es2 se agrupó con dos individuos de *E. tereticornis* (Et6 y Et8).

Así, los únicos grupos que no pudieron ser discriminados con este análisis fueron los formados por individuos de *E. grandis* y *E. saligna* (en violeta) y de *E. camaldulensis* y *E. tereticornis* (grupo representado en anaranjado). Los individuos de este último grupo tampoco pudieron ser discriminados por especie en el dendrograma.

Por otro lado, se realizó el mismo estudio considerando información *a priori* de pertenencia de los individuos a las siete especies. Sin embargo, el Test de Evanno realizado mostró un número K de poblaciones igual a cuatro ($K=4$) más probable y también un $k=7$, lo cual era el resultado esperado (Figura 3.23).

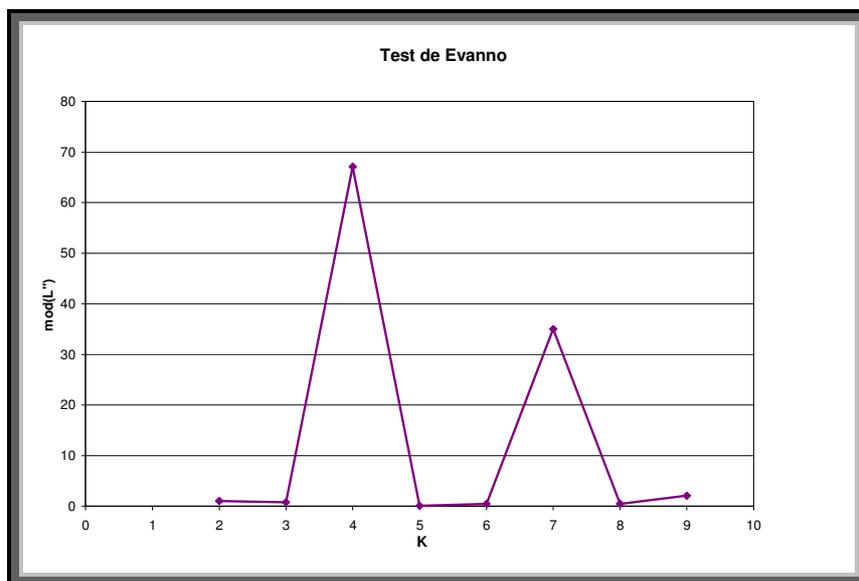


Figura 3.23 Test de Evanno resultante del análisis poblacional mediante STRUCTURE con información *a priori*

Con estos resultados, y teniendo en cuenta la información previa del número de grupos, se decidió realizar los análisis utilizando los k que mostraron las mayores probabilidades *a posteriori* de los datos ($k=4$ y $k=7$) (Figura 3.24_B y C). Las matrices Q utilizadas ($K=4$ y $K=7$) se muestran en la Tabla anexa 7.

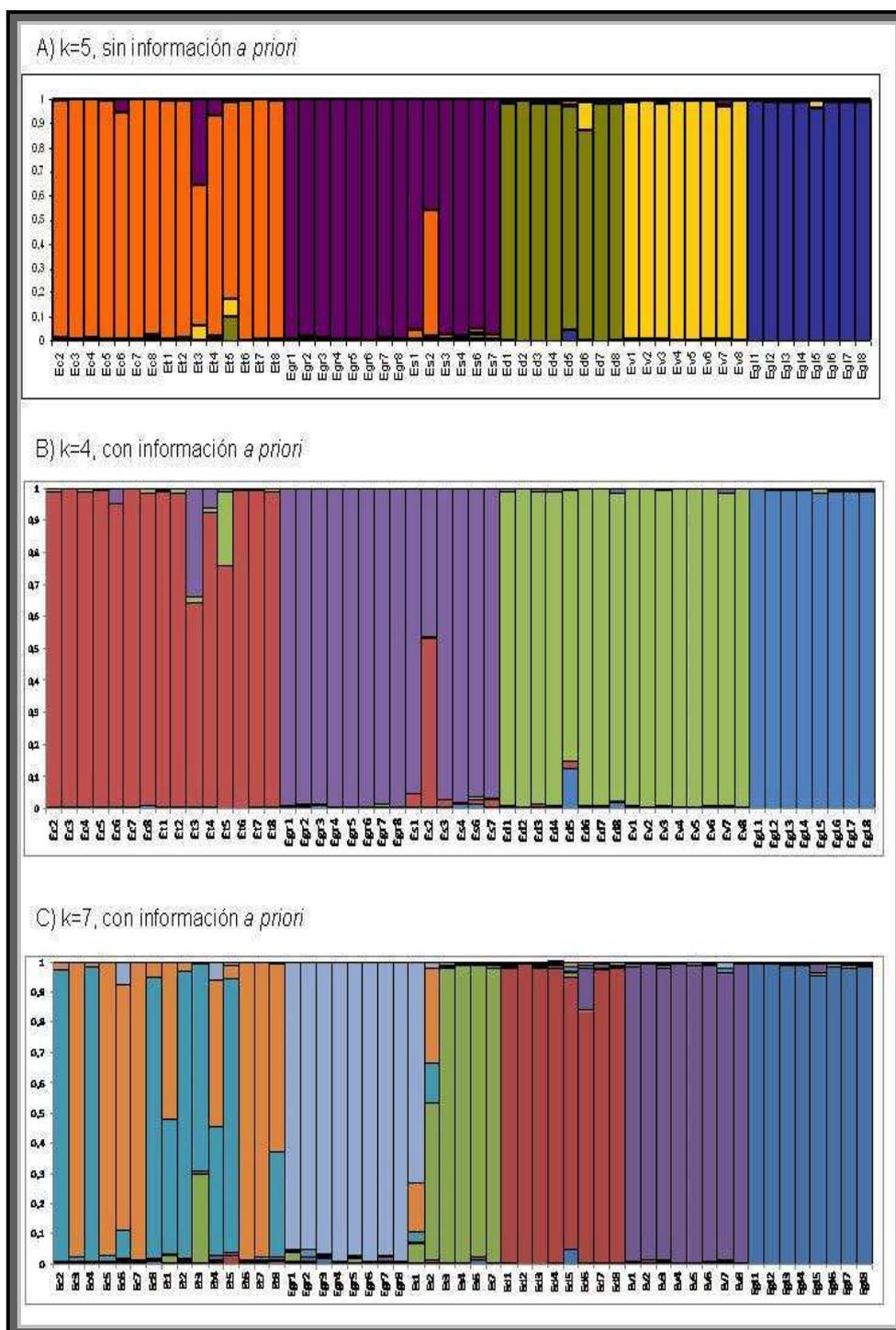


Figura 3.24 Análisis de la muestra mediante STRUCTURE mostrando la asignación de todos los individuos a los k grupos obtenidos sin información *a priori* (A) y con información *a priori* (B y C)

Al incluir información *a priori* de la pertenencia de los individuos a las especies para $k=4$, tampoco pudieron ser discriminados los individuos de *E. camaldulensis* y *E. tereticornis* (grupo representado en rojo), *E. grandis* y *E. saligna* (en violeta) y *E. dunnii* y *E. viminalis* (en verde) (Figura 3.24-B).

Excepto los individuos Et3, Et5 y Es2 todos los individuos mostraron una constitución genética mayoritaria (>80%) correspondientes a una de los cuatro grupos ideales.

Por otro lado, la asignación de los individuos a los siete grupos ($k=7$), mostró que los mismos presentaron una constitución genética mayor al 80% de cada grupo, para las especies *E. dunnii*, *E. grandis*, *E. viminalis*, *E. globulus* y la mayoría de los individuos de *E. saligna*. Para esta última especie, el individuo Es1 se agrupó (con una probabilidad *a posterior* de 0,73) al grupo de los individuos de *E. grandis* (color celeste). Esto podría deberse a que este individuo en particular, podría ser un híbrido de *E. saligna* por otra especie, de la cual no se tiene registro. Según los resultados de este análisis, esta especie podría ser *E. grandis*. Por otra parte, el individuo, Es2 parecería tener contribución genética del grupo formado por *E. tereticornis* y *E. camaldulensis*.

Sin embargo, los individuos de *E. camaldulensis* y *E. tereticornis* no pudieron ser discriminados, validando los resultados obtenidos en el análisis de agrupamiento (Figura 3.21) y la baja diferenciación PhiPT (0,069) que se encontró entre estas especies (ver más adelante, Tabla 3.11 y Figura 3.25).

Estos resultados muestran el potencial que poseen estos marcadores para la discriminación de individuos de las especies estudiadas, a excepción de *E. camaldulensis* y *E. tereticornis* que no pudieron diferenciarse, aún asignando como dato *a priori* la división en siete grupos. A su vez, sin tener información *a priori* de las especies, estos marcadores lograron discriminar grupos entre las sección *Exsertaria* (agrupando *E. camaldulensis* y *E. tereticornis*), *Latoangulatae* (agrupando *E. grandis* y *E. saligna*) y también dentro de la sección *Maidenaria* (diferenciando *E. dunnii*, *E. viminalis* y *E. globulus*).

En la tercera evaluación, se calcularon los coeficientes de diferenciación entre especies mediante AMOVA. La matriz de coeficientes de diferenciación PhiPT obtenida mediante GenAlex 6.2 (Tabla 3.11) mostró que los valores variaron entre 0,069 (*E. camaldulensis* y *E.*

tereticornis) y 0,414 (*E. globulus* y *E. viminalis*), todos altamente significativos ($P \leq 0,005$) luego de 9999 permutaciones.

Utilizando estos valores, se realizó un análisis de agrupamiento mediante el programa NTSyS.

El mismo mostró nuevamente, que las especies más relacionadas entre sí son *E. camaldulensis* y *E. tereticornis*, mientras que *E. globulus* fue la que más se diferenció del resto (Figura 3.25 y Tabla 3.11).

Tabla 3.11 Matriz de diferenciación PhiPT entre las siete especies analizadas

	<i>E. camaldulensis</i>	<i>E. dunnii</i>	<i>E. grandis</i>	<i>E. saligna</i>	<i>E. tereticornis</i>	<i>E. viminalis</i>	<i>E. globulus</i>
<i>E. camaldulensis</i>	0						
<i>E. dunnii</i>	0,282	0					
<i>E. grandis</i>	0,310	0,237	0				
<i>E. saligna</i>	0,258	0,250	0,185	0			
<i>E. tereticornis</i>	0,069	0,187	0,194	0,162	0		
<i>E. viminalis</i>	0,341	0,194	0,344	0,316	0,248	0	
<i>E. globulus</i>	0,383	0,344	0,376	0,350	0,343	0,414	0

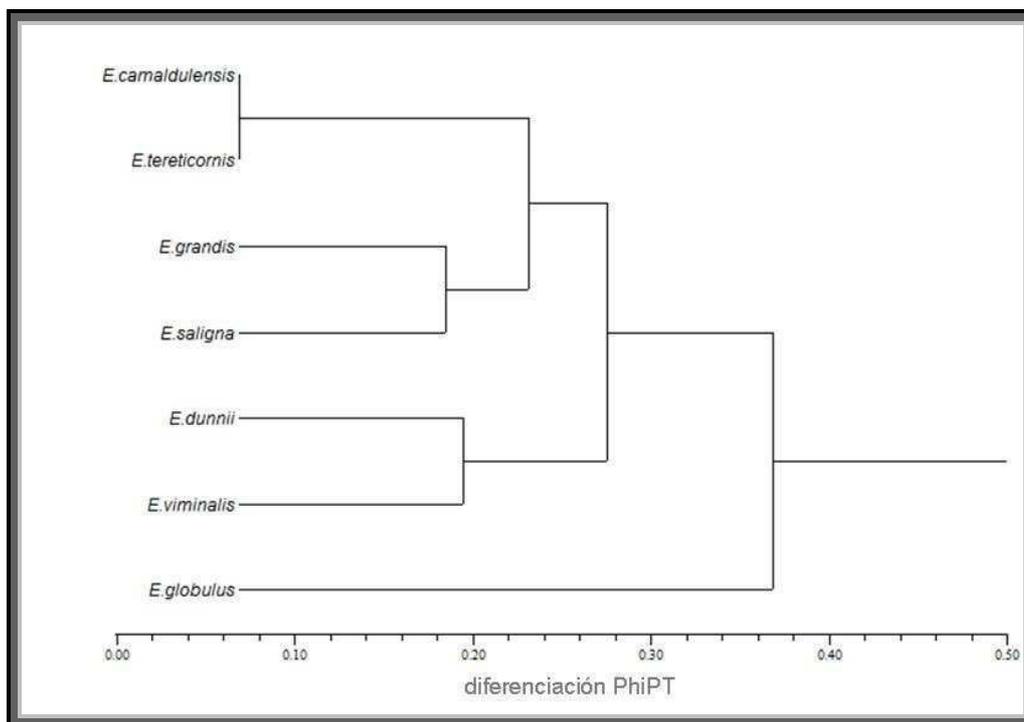


Figura 3.25: Agrupamiento (UPGMA) generado a partir de la matriz de PhiPT entre especies

Los dendrogramas obtenidos (Figura 3.21 y 3.25) fueron consistentes con las relaciones filogenéticas conocidas y podrían ser explicados por la fuerte asociación comúnmente encontrada en *Eucalyptus* entre la similitud genética y la proximidad geográfica de especies y poblaciones (Butcher y col. 2002; Steane y col.2006; Payn y col. 2008).

Así, *E. camaldulensis* y *E. tereticornis* que pertenecen a la Sección *Exsertaria*, se agruparon juntas con un valor de Phi de 0,069. La afinidad de estas dos especies fue también encontrada por Steane y col. (2011), analizando mediante tecnología DArT (*Diversity Arrays Technology*, Jaccourd y col. 2001), en un estudio que involucró 94 taxa con 8354. A su vez, estas dos especies estrechamente relacionadas poseen propiedades de madera y características de crecimiento similares. De hecho, en regiones australianas en donde coexisten ambas especies, se encuentran híbridos interespecíficos naturales (Eldridge y col. 1994).

Por otro lado se agruparon las especies *Eucalyptus grandis* y *E. saligna* las cuales pertenecen a la misma sección (*Latoangulatae*). En un estudio taxonómico llevado a cabo por Burgess y Bell (1983) se analizaron 24 poblaciones tomadas a lo largo del rango de distribución natural de ambas especies. Los autores encontraron que existían poblaciones claramente asignadas a

una de las dos especies, mientras que otras poblaciones, con caracteres fenotípicos intermedios, fueron difíciles de clasificar.

A su vez, las especies de las secciones *Exsertaria* y *Latoangulatae* presentaron menor distancia genética entre sí, lo que concuerda con estudios filogenéticos realizados con secuencias ITS por Steane y col. 2002

Por último, las especies *E. dunni*, *E. viminalis* y *E. globulus* pertenecen a la sección Maidenaria. En este estudio se encontró menor distancia entre las especies *E. dunni* y *E. viminalis* que se agruparon juntas en el dendrograma, mientras *E. globulus* se separó del resto de las especies analizadas. La diferenciación de *E. globulus* podría deberse a la mayor distancia geográfica que existe entre las procedencias de las muestras analizadas (sur de Victoria e Isla de Tasmania) con respecto a *E. dunnii* (New South Wales) y *E. viminalis* (Victoria) y a la falta de flujo génico entre ellas. Esta diferenciación se observó también en el estudio de Faria y col. (2010 b).

Estos resultados demuestran que, a pesar del muestreo pequeño, los marcadores desarrollados mostraron afinidades entre especies similares a las encontradas en otros estudios filogenéticos utilizando marcadores moleculares neutros y caracteres fenotípicos.

A partir del AMOVA entre especies, no obstante el tamaño pequeño de las muestras analizadas por especie, se pudo observar que el 28% de la variabilidad se encuentra entre ellas, mientras que el 72% restante, dentro de cada especie (Figura 3.26). En la figura 3.27, se muestra el valor de PhiPT significativo ($P < 0,000$, barra roja), ya que dicho valor fue superior a todos los valores de PhiPT encontrados en una distribución nula (barras de color negro) generadas por 9999 permutaciones aleatorias.

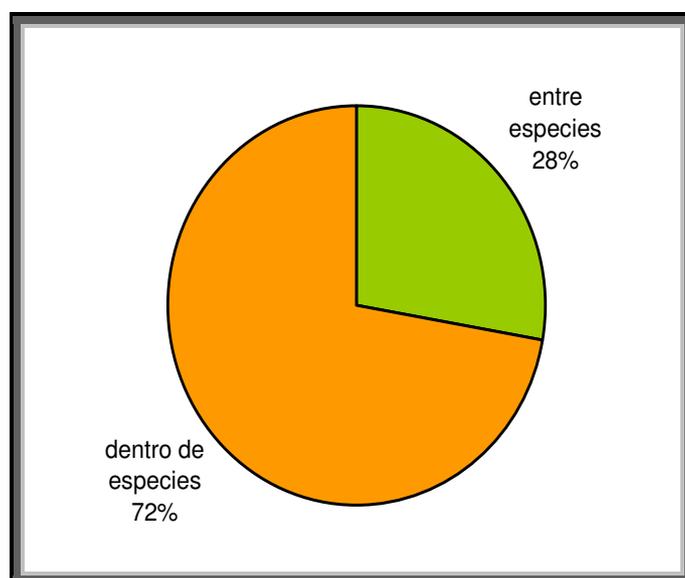


Figura 3.26: Gráfico representativo del análisis de la variancia molecular (AMOVA)

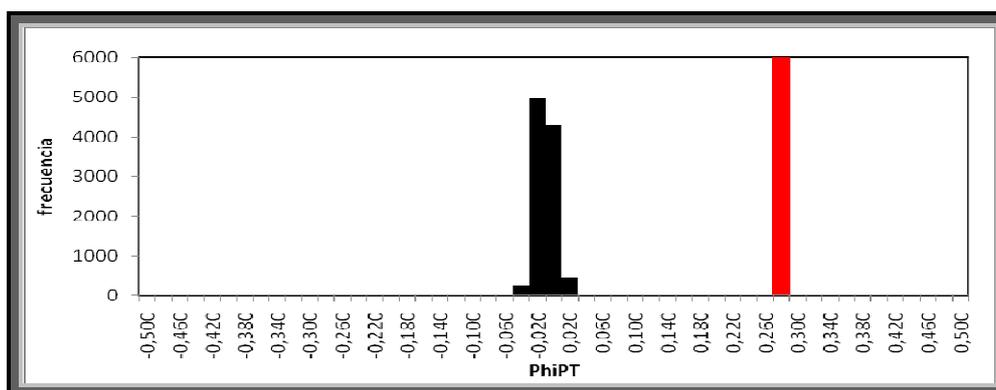


Figura 3.27: distribución de frecuencias de PhiPT permutadas (9999) (barras negras) respecto del PhiPT observado (barra roja)

En resumen, los resultados de estos tres tipos de análisis mostraron que los marcadores funcionales analizados permitieron discriminar al conjunto de individuos dentro y entre las especies, encontrándose una distancia genética (DAS) elevada entre los mismos (Figura 3.20) y corroborados mediante el AMOVA.

3.3.6 Selección de un grupo de marcadores funcionales más informativos

En general, los SSR derivados de genotecas genómicas muestran un mayor nivel de polimorfismo que aquellas provenientes de bancos de datos de ESTs públicos (Cho y col. 2000; Kim y col. 2008). Los resultados de este estudio confirman esto, aunque fueron ligeramente inferiores o similares a aquellos parámetros estimados por Brondani y col. (1998), Kirst y col. (2005), Marcucci Poltri y col. (2003), Steane y col. (2006) y Zelener y col. (2005). Esto puede explicarse debido a que los SSR funcionales se comportaron como neutros según el Test de Neutralidad de Ewens-Watterson realizado. A su vez, ésta es la mejor razón de porqué son interesantes y tienen un uso potencial como marcadores funcionales en el mejoramiento y en estudios de mapeo comparativo y diversidad genética.

Dado que algunos de los *loci* analizados resultaron ser moderada a altamente informativos se seleccionó un subconjunto de marcadores con un valor de PIC mayor a 0,55, frecuencias de alelos nulos inferiores a 0,15 y transferibles a las otras especies de *Eucalyptus* analizadas (Sección 3.3.4).

Dichos marcadores son: los GC para calidad de madera CesA3_a, CesA3_b, LIM1, HMTi, SK, GST, y los EST-SSRs 11, 46, 58, 157, 191, 216 y 219. Los EST-SSRs selectos tienen similitud de secuencia con enzimas con funciones similares, tales como: proteína tipo NtSar1 (crecimiento celular), factor de unión a DRE (respuesta a estrés), proteína de respuesta a fosfato, proteína de control del ciclo celular, proteína serina treonina kinasa (relacionadas a señalización y defensa en plantas), enzima conjugada a ubiquitina y proteína de estrés de maduración mediada por absícico (respuesta a estrés).

La Probabilidad de Identidad (PI) para este conjunto de marcadores seleccionado resultó muy bajo ($1,97 \times 10^{-15}$), resultando así complementarios a los publicados por Faria y col. (2010 a,b). Por lo tanto, este conjunto de marcadores funcionales podrá utilizarse en análisis de diversidad genética, identificación de individuos ("fingerprinting") y estudios poblacionales comparativos entre especies.

Estos 13 marcadores de elevada calidad representan más del 15% de los marcadores funcionales evaluados (29 GC y 56 EST-SSRs).

Con estos resultados, se podría esperar que aproximadamente 150 EST-SSRs de los nuevos 979 desarrollados *in silico* sean potencialmente informativos.

3.4 Recapitulación

Considerando los resultados obtenidos en este estudio, se podría afirmar que las bases de datos públicas que contienen tanto ESTs como secuencias genómicas, son de suma utilidad para la búsqueda y desarrollo de marcadores funcionales. En este trabajo se presentaron 979 nuevos EST-SSRs que, junto con el análisis de las secuencias de ADN de distintos genes, permitieron el desarrollo y validación de 17 EST-SSRs y 12 marcadores SSRs dentro de genes candidatos para calidad de madera. Estos microsatélites fueron evaluados de acuerdo a su polimorfismo y transferibilidad a otras seis especies de *Eucalyptus* representes en las secciones filogenéticas principales del subgénero *Symphomyrtus*. Se analizó la utilidad de estos marcadores para la discriminación tanto de individuos como de especies.

Debido a que fueron desarrollados a partir de regiones conservadas, es probable que la transferibilidad y el polimorfismo de los microsatélites se extienda a otras 300 o más especies en el subgénero.

Este conjunto de marcadores ofrece una herramienta para estudios de genética de poblaciones y para aplicaciones al mejoramiento genético ya que miden directamente la variabilidad funcional de genes candidatos para calidad de madera en el género *Eucalyptus*.

Hasta el momento, los importantes avances en la genómica de *Eucalyptus* a nivel mundial justifican el esfuerzo en la investigación por el enorme impacto que tendrá el mejoramiento asistido por marcadores para los caracteres de calidad de madera. Este progreso se vio facilitado por la aparición de las tecnologías moleculares y la mejora en la precisión de las mediciones de madera y fibra, necesarios para la fenotipificación.

Por otro lado, el germoplasma de *E. globulus* es altamente variable y representa un recurso clave para estudios de mapeo genético y de asociación. El número y la calidad de QTLs se ha incrementado en la última década. Para la construcción de mapas genéticos se están utilizando, además de marcadores anónimos, la estrategia de genes candidatos, más precisa y confiable. Esto es posible gracias a la identificación de numerosos genes candidatos potenciales mediante cartografía física y descubrimiento y asignación de posible función a ESTs. Hasta el momento, algunos genes, principalmente implicados en la biosíntesis de la lignina y la pared celular, se han validado y se presentan como buenos candidatos potenciales. CAD y CCR, así como otros cuatro genes implicados en la ruta de fenilpropanoides fueron encontrados en cuatro regiones

distintas del mapa de *E. urophylla* construido mediante RAPD (Gion y col. 2000). Dentro de estos genes, se detectaron 53 SNPs en la secuencia de CCR y 8 SNP en CAD en *E. globulus*, pero hasta ahora los estudios de asociación no han mostrado ninguna evidencia de que estas sustituciones de un solo aminoácido en el CCR afecte el contenido de lignina (Poke y col. 2003).

Este estudio hace una importante contribución a los marcadores SNPs hallados dentro de genes candidatos (Külheim y col. 2009, Novaes y col., 2008) ya que no se han descrito hasta el momento marcadores SSRs en los mismos, de fácil medición y alta variabilidad.

Además de los genes estructurales se hizo enfoque en genes reguladores, tales como factores de transcripción, que puedan tener un alto impacto en el carácter. Los primeros datos sobre genes candidatos en eucaliptos hasta ahora indican que los genes de efectos mayores existen, lo que sugiere que el estudio de estos genes es una estrategia realista para la aplicación del MAS.

Conclusiones y perspectivas

4. CONCLUSIONES Y PERSPECTIVAS

- Se analizaron bioinformáticamente las colecciones de secuencias expresadas (ESTs) de bancos de datos de dominio público de *Eucalyptus globulus*. El ensamblado de las 12.690 secuencias permitió contar con 4.925 secuencias únicas (unigenes), de las cuales se pudo obtener información no redundante.
- Las herramientas bioinformáticas utilizadas permitieron la búsqueda de secuencias repetitivas tipo microsatélites en la colección no redundante de ESTs. El hallazgo de 979 nuevos EST-SSRs hace una importante contribución a los marcadores publicados por otros autores dentro del género *Eucalyptus* (Ceresini y col. 2005; Faria y col. 2010 a,b; Rabello y col. 2005; Rengel y col. 2009, Yasodha y col. 2008).
- Las herramientas bioinformáticas, junto con la abundante información disponible de secuencias de proteínas, permitió una importante asignación de función (78%) de los unigenes que contienen microsatélites. Esto permitió el hallazgo de 22 genes candidatos para calidad de madera, conteniendo 23 SSRs.
- Por otro lado, las secuencias genómicas depositadas en GenBank, junto con los datos de bibliografía existente, permitieron la selección y búsqueda de secuencias de genes candidatos para calidad de madera, especialmente aquellos involucrados en la formación de la madera y sus componentes principales (lignina, celulosa y hemicelulosa). En estas secuencias se reconocieron seis SSRs en cuatro genes, tanto en regiones de promotores como en un intrón.
- La utilización de una pequeña muestra de 8 individuos de *E. globulus* en el ensayo de validación, permitió detectar un elevado porcentaje de marcadores polimórficos (17 EST-SSRs y 12 SSR de GC a partir de 57 EST-SSRs y de los 29 GC-SSRs).
- Aunque se espera que los marcadores funcionales sean menos variables que los neutros, el estudio del polimorfismo de los marcadores ensayados mostró un elevado potencial de los mismos en estudios de diversidad genética.
- El análisis de los valores de PIC, H_o y U_h mostró que los marcadores SSRs en promotores o intrones resultaron más variables que los encontrados en regiones expresadas debido a presión de selección diferencial entre ellas.

- Los marcadores utilizados fueron capaces de detectar la presencia de subestructura poblacional, una de las principales causas del desvío del Equilibrio de Hardy Weinberg (EHW) encontrado en 14 de los 29 *loci*.
- La identificación y verificación de alelos raros (en 13 EST-SSRs y 11 SSR en GC, en la población de 60 individuos de *E. globulus*) podría resultar de importancia en el mejoramiento genético y conservación de la variabilidad en las poblaciones.
- El conjunto de marcadores funcionales analizados detectó la misma estructura poblacional en *E. globulus* que otros marcadores (2300 DArT) y mediante diferentes análisis (distancia entre individuos y composición genética de cada uno estimada mediante STRUCTURE).
- La elevada transferibilidad de los EST-SSRs y SSR de GC a las especies de *E. camaldulensis*, *E. dunnii*, *E. grandis*, *E. saligna*, *E. tereticornis* y *E. viminalis*, mostró su potencial para el estudio de variabilidad genética y sintenia entre especies.
- El estudio de transferibilidad permitió describir y validar los primeros microsatélites disponibles en *E. viminalis*, los cuales serán utilizados en estudios de variabilidad genética en el programa de mejoramiento de esta especie en INTA.
- La utilización de una pequeña muestra de 8 individuos de cada una de las siete especies de *Eucalyptus* permitió el agrupamiento en forma consistente con las relaciones filogenéticas conocidas y la proximidad geográfica de éstas.
- Por extrapolación de nuestros resultados de validación se espera que, de los 979 nuevos EST-SSRs, más de 290 (30%) sean polimórficos y más de 240 (25%) sean transferibles a las especies *E. grandis*, *E. saligna*, *E. dunnii*, *E. viminalis*, *E. camaldulensis*, *E. tereticornis*.
- El estudio del polimorfismo y transferibilidad de todos los marcadores funcionales, permitió la selección de un conjunto de 13 (7 EST-SSRs y 6 GC-SSRs) altamente informativos, con bajas frecuencias estimadas de alelos nulos, altos PIC, bajos PI y transferibles a otras seis especies de *Eucalyptus*. Los mismos están siendo transferidos al Programa de Mejoramiento Genético de *Eucalyptus*.
- Los resultados derivados de los estudios de diversidad genética servirán también para sustentar acciones de conservación, uso de los recursos genéticos y de identificación de genotipos, aportando descriptores moleculares y marcadores funcionales en los procesos de obtención de nuevos materiales según UPOV.

El hallazgo de motivos SSRs en genes candidatos en *Eucalyptus* no cuenta con precedentes, con lo cual se hace una importante contribución al estudio de variabilidad natural dentro de los mismos, complementando otros trabajos de búsqueda de SNPs en genes candidatos. Estos marcadores podrían contribuir a la verificación de sintenia y colinearidad entre diferentes mapas genéticos de *E. globulus*, validación de genes y posición de QTLs en diferentes especies de las secciones *Maidenaria*, *Exsertaria* y *Latoangulatae*, a las que pertenecen la mayor parte de las especies comerciales.

Lic. Cintia V. Acuña
Tesisista

Dr. H. Esteban Hopp
Director

Dra. Susana N. Marcucci Poltri
Director Asistente

Referencias bibliográficas

Referencias bibliográficas

Acuña C, Hopp E, Pathauer P, Villalba P, Marcucci Poltri SN (2011) Characterization of novel microsatellite markers in candidate genes for wood quality for application in functional diversity assessment in *Eucalyptus globulus*. Enviado a *Electronical Journal of Biotechnology*

Acuña C, Torales S, Zelener N, López G, Pathauer P, Marcucci Poltri SN (2005) Uso de AFLP para determinar los posibles orígenes de razas locales de *Eucalyptus globulus*. *Revista IDIA XXI* 5: 149-152

Arnaud D, Dejardin A, Leple JC, Lesage-Descauses MC, Pilate G (2007) Genome-wide analysis of LIM gene family in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*. *DNA Res* 14: 103-116

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25: 25-29

Baum G, Chen Y, Arazi T, Takatsuji H, Fromm H. (1993) A plant glutamate decarboxylase containing a calmodulin binding domain. Cloning, sequence, and functional analysis. *Journal of Biological Chemistry* 268(26):19610–19617

Baum G, Lev-Yadun S, Fridmann Y, Arazi T, Katsnelson H, Zik M, Fromm H. (1996) Calmodulin binding to glutamate decarboxylase is required for regulation of glutamate and GABA metabolism and normal development in plants. *The EMBO Journal* 15:2988–2996

Blee K.A., Choi J.W., O'Connell A.P., Schuch W., Lewis N.G., Bolwell G.P (2003) A lignin-specific peroxidase in tobacco whose antisense suppression leads to vascular tissue modification. *Phytochemistry* 64 (1):163-176

Bossinger G, Leitch M (2000) Isolation of cambium-specific genes from *Eucalyptus globulus* Labill. In: Savidge R BJ, Napier R (ed) *Cell and molecular biology of wood formation*. BIOS Scientific, Oxford, pp 203–207

Botstein D, White R L, Skolnick M and Davis RW (1980). Construction of genetic linkage map in man using restriction length polymorphisms. *Am J Hum Genet* 32: 314–331

Bouche N., A. Fait, M. Zik and H. Fromm (2004) The root-specific glutamate decarboxylase (GAD1) is essential for sustaining GABA levels in *Arabidopsis*. *Plant Mol. Biol.* 55: 315–325

Brooker MIH (2000) A new classification of the genus *Eucalyptus* L'Her. (*Myrtaceae*) *Aust Syst Bot* 13:79–148

Brondani RP, Brondani C, Grattapaglia D (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Mol Genet Genomics* 267: 338-347

- Brondani RP, Williams ER, Brondani C, Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. *BMC plant biology* 6: 20
- Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *TAG Theoretical and Applied Genetics* 97: 816-827
- Burguess IP, Bell J (1983) Comparative morphology and allozyme frequencies of *Eucalyptus grandis* Hill ex Maiden and *E. saligna* Sm. *Aust. For. Res* 13:133-149
- Butcher PA, Otero A, McDonald MW, Moran GF. (2002) Nuclear RFLP variation in *Eucalyptus camaldulensis* Dehnh from northern Australia. *Heredity*. 88:402–412.
- Byrne M, Marquez-Garcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite *loci* in the genus *Eucalyptus*. *Australian Journal of Botany* 44:331–34
- Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR. (1993) Incidence and origin of 'null' alleles in the (AC)_n microsatellite markers. *Am J Hum Genet*. 52:922–927
- Carpineti L, Dalla Tea F, Glade JE, Marcó MA (eds) (1995) Manual para productores de eucaliptos de la Mesopotamia Argentina. Grupo Forestal, EEA. INTA Concordia
- Ceresini PC, Silva CLSP, Missio RF, Souza EC, Fischer CN, Guilherme IR, Gregorio I, Silva EHTd, Cicarelli RMB, Silva MTAd, Garcia JF, Avelar GA, Porto Neto LR, Marçon AR, Bacci Junior M, Marini DC (2005) Satellyptus: analysis and database of microsatellites from ESTs of *Eucalyptus*. *Genetics and Molecular Biology* 28: 589-600
- Cervigni GD, Paniego N, Diaz M, Selva JP, Zappacosta D, Zanazzi D, Landerreche I, Martelotto L, Felitti S, Pessino S, Spangenberg G, Echenique V (2008) Expressed sequence tag analysis and development of gene associated markers in a near-isogenic plant system of *Eragrostis curvula*. *Plant molecular biology* 67: 1-10
- Chabane K, Ablett G, Cordeiro G, Valkoun J, Henry R (2005) EST versus Genomic Derived Microsatellite Markers for Genotyping Wild and Cultivated Barley. *Genetic Resources and Crop Evolution* 52: 903-909
- Chabannes M, Barakate A, Lapierre C, Marita JM, Ralph J, Pean M, Danoun S, Halpin C, Grima-Pettenati J, Boudet AM (2001) Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. *Plant J*. 28, 257–270
- Chakraborty R, Jin L (1993) A unified approach to study hyper-variable polymorphisms: Statistical considerations of determining relatedness and population distances. In: *DNA fingerprinting: State of the Science* (eds PenaDJ, ChakrabortyR., EplensJT, JeffreysAJ), pp. 153–175.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, et al (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 100:713-722.

Christensen JH, Bauw G, Welinder KG, van Montagu M and Boerjan W (1998) Purification and characterization of peroxidases correlated with lignification in poplar xylem. *Plant Physiology* 118:125–13

Christensen JH, Overney S, Rohde A, Diaz WA, Bauw G, Simon P, van Montagu M and Boerjan W (2001) The syringaldazine-oxidizing peroxidase PXP 3-4 from poplar xylem: cDNA isolation, characterization and expression. *Plant Molecular Biology* 47: 581–593.

Chybicki and Jaroslaw Burczyk (2009) Simultaneous Estimation of Null Alleles and Inbreeding Coefficients. *J Hered* 100 (1):106-113.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)* 21: 3674-3676

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant science (Shannon, Ireland)* 160: 1115-1123

Creux NM, Ranik M, Berger DK, Myburg AA (2008) Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. *New Phytol* 179(3):722-737.

Dean JV, Devarenne TP (1997) Peroxidase-mediated conjugation of glutathione to unsaturated phenylpropanoids. Evidence against glutathione S-transferase involvement. *Physiologia plantarum* 99: 271-278.

Dean JV, Devarenne TP, Lee IS, Orlofsky LE (1995) Properties of a Maize Glutathione S-Transferase That Conjugates Coumaric Acid and Other Phenylpropanoids. *Plant Physiol* 108: 985-994

Dejardin A, Leple JC, Lesage-Descauses MC, Costa G, Pilate G (2004) Expressed sequence tags from poplar wood tissues--a comparative analysis from multiple libraries. *Plant biology (Stuttgart, Germany)* 6: 55-64

Demont-Caulet N, Lapierre C, Jouanin L, Baumberger S, Méchin V (2010) *Arabidopsis* peroxidase-catalyzed copolymerization of coniferyl and sinapyl alcohols: Kinetics of an endwise process. *Phytochemistry* 71(14-15):1673-1683

Dutkowski GW, Potts BM (1999) Geographic Patterns of Genetic Variation in *Eucalyptus globulus* ssp. *globulus* and a Revised Racial Classification. *Aust J Bot* 47:237-263

Edwards R, Dixon DP, Walbot V (2000) Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends Plant Sci* 5 (5):193-198.

- Egertsdotter U, van Zyl LM, MacKay J, Peter G, Kirst M, Clark C, Whetten R, Sederoff R (2004) Gene expression during formation of earlywood and latewood in loblolly pine: expression profiles of 350 genes. *Plant biology (Stuttgart, Germany)* 6: 654-663
- Eldridge K, Davidson J, Hardwood C, Van Wyk G (1994) *Eucalypt domestication and breeding*. Oxford University Press, Oxford
- Eldridge K, Davidson J, Harwood C, vanWyk G (1993) *Eucalypt domestication and breeding*. Clarendon, Oxford, UK
- Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. *Euphytica* 119: 39-43
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14:2611–2620.
- Fagard M, Höfte H, Vernhettes S (2000) Cell wall mutants. *Plant Physiology and Biochemistry* 38: 15-25
- Faria DA, Mamani EMC, Pappas MR, Pappas GJ and Grattapaglia D (2010a) A selected set of EST-derived microsatellites, polymorphic and transferable across 6 species of *Eucalyptus*. *Journal of Heredity* 101: 512-520
- Faria, D.A., Mamani, E.M.C., Pappas, G.J. and Grattapaglia, D (2010b) Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. *Tree Genetics and Genomes* DOI: 10.1007/s11295-010-0315-9
- Fernandez-Silva I, Eduardo I, Blanca J, Esteras C, Picó B, Nuez F, Arús P, Garcia-Mas J, Monforte A (2008) Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.). *TAG Theoretical and Applied Genetics* 118: 139-150
- Foucart C, Paux E, Ladouce N, San-Clemente H, Grima-Pettenati J, Sivadon P (2006) Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*. *The New phytologist* 170: 739-752
- Gallego P.P. y col. (1995) A role for glutamate decarboxylase during tomato ripening: the characteristics of a cDNA encoding a putative glutamate decarboxylase with a calmodulin-binding site. *Plant Mol. Biol.* 27:1143–1151
- Genovesi V, Fornalé S, Fry SC, Ruel K, Ferrer P, Encina A, Sonbol FM, Bosch J, Puigdoménech P, Rigau J, Caparrós-Ruiz D (2008) ZmXTH1, a new xyloglucan endotransglucosylase/ hydrolase in maize, affects cell wall structure and composition in *Arabidopsis thaliana*. *Journal of Experimental Botany* 59: 875–889
- Gion J-M, Rech P, Grima-Pettenati J, Verhaegen D, Plomion C (2000) Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. *Molecular Breeding* 6: 441-449

Gion JM, Boudet C, Grima-Pettenati J, Pichavant FH, Plomion C, Bailleres H, Verhaegen D (2001) A candidate genes approach identifies CCR, PAL and C4H as *loci* for syringyl/guaiacyl ratio in an interespecific hybrid between *E. urophylla* and *E. grandis*. In: Conference Pol (ed) Developing the eucalypt of the future, Valdivia, Chile.

Glaubitz JC, Emebiri LC, Moran GF (2001) Dinucleotide microsatellites from *Eucalyptus sieberi*: inheritance, diversity, and improved scoring of single-base differences. *Genome* 44: 1041-1045

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J (2005) EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. *Plant J* 43: 553-567

Grattapaglia D, Kirst M (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. *The New phytologist* 179: 911-929

Grattapaglia D, Bradshaw HD (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Can J For Res* 24:1074–1078

Greaves BL, Borralho NMG, Raymond CA, Farrington A (1996) Use of a Pilodyn for the indirect selection of basic density in *Eucalyptus nitens*. *Can. J. For. Res.* 26: 1643-1650

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163-185

Gut H, Dominici P, Pilati S, Astegno A, Petoukhov MV, Svergun DI, Grütter MG, Capitani G (2009) A Common Structural Basis for pH- and Calmodulin-mediated Regulation in Plant Glutamate Decarboxylase. *Journal of Molecular Biology* 392(2): 334-351.

Hackauf B, Wehling P (2002) Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breeding* 121: 17-25

Harper GL, Maclean N, Goulson D (2003) Microsatellite markers to assess the influence of population size, isolation and demographic change on the genetic structure of the UK butterfly *Polyommatus bellargus*. *Molecular Ecology*, 12(12), 3349-3357

Herrmann KM, Weaver LM (1999) The Shikimate Pathway. *Annu Rev Plant Physiol Plant Mol Biol*, 50: 473-503.

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G (2001) A transcriptional roadmap to wood formation. *Proceedings of the National Academy of Sciences of the United States of America* 98: 14732-14737

Hoisington D, Khairallah M, Gonzalez-de-Leon D (1994) Laboratory protocols: CIMMYT Applied Molecular Genetics Laboratory. CIMMYT, Mexico, D.F.

Hrmova M, Farkas V, Lahnstein J, Fincher GB (2007) A Barley xyloglucan xyloglucosyl transferase covalently links xyloglucan, cellulosic substrates, and (1,3;1,4)-beta-D-glucans. *The Journal of biological chemistry* 282: 12951-12962

Hu WJ et al. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat. Biotechnol* 17: 808–812

Israelsson M, Eriksson ME, Hertzberg M, Aspeborg H, Nilsson P, Moritz T (2003) Changes in gene expression in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth. *Plant molecular biology* 52: 893-903

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bull Soc Vaudoise Sci Nat*, 44: 223–270.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucl. Acids Res.* 29, e25.

Jones L, Ennos AR and Turner SR (2001) Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of *Arabidopsis*. *Plant J.* 26: 205–216.

Jones ME, Shepherd M, Henry RJ, Delves A (2006) Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Conserv. Genet.* 7, 691–703.

Jordan, GJ and Borralho, NMG and Tilyard, P and Potts, BM (1994) Identification of races in *Eucalyptus globulus* ssp *globulus* based on growth traits in Tasmania and geographic distribution. *Silvae Genetica*, 43 (5-6). pp. 292-298. ISSN 0037-5349

Jurka J, Pethiyagoda C (1995) Simple repetitive DNA sequences from primates: compilation and analysis. *Journal of molecular evolution* 40: 120-126

Kampranis SC, Damianova R, Atallah M, Toby G, Kondi G, Tschlis PN, Makris AM (2000) A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. *The Journal of biological chemistry* 275: 29207-29216

Katti MV, Sami-Subbu R, Ranjekar PK, Gupta VS (2000) Amino acid repeat patterns in protein sequences: their diversity and structural-functional implications. *Protein Sci* 9: 1203-1209

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular biology and evolution* 18: 1161-1167

Kawaoka A, Kaothien P, Yoshida K, Endo S, Yamada K, Ebinuma H (2000) Functional analysis of tobacco LIM protein Ntlm1 involved in lignin biosynthesis. *The Plant Journal* 22: 289-301

Kawaoka A, Nanto K, Ishii K, Ebinuma H (2006) Reduction of lignin content by suppression of expression of the LIM domain transcription factor in *Eucalyptus camaldulensis*. *Silvae Genetica* 55: 269-277

- Killili KG, Atanassova N, Vardanyan A, Clatot N, Al-Sabarna K, Kanellopoulos PN, Makris AM, Kampranis SC (2004) Differential roles of tau class glutathione S-transferases in oxidative stress. *The Journal of biological chemistry* 279: 24540-24551
- Kim K, Ratcliffe S, Wade French B, Liu L, Sappington T (2008) Utility of EST-derived SSRs as population genetics markers in a beetle. *J Hered* 99(2):112–124.
- Kimura S, Kondo T (2002) Recent progress in cellulose biosynthesis. *Journal of plant research* 115: 297-302
- Kirst M, Cordeiro CM, Rezende GD and Grattapaglia D (2005) Power of Microsatellite Markers for Fingerprinting and Parentage Analysis in *Eucalyptus grandis* Breeding Populations. *Journal of Heredity* 96(2):161–166. doi:10.1093/jhered/esi023
- Kirst M, Myburg AA, De Leon JP, Kirst ME, Scott J, Sederoff R (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol* 135: 2368-2378
- Külheim C, Yeoh S H, Maintz J, Foley WJ and Moran GF (2009) Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* 10, 452
- Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48: 985-998
- La Rota M, Kantety R, Yu JK, Sorrells M (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6 (1):23.
- Lacombe E, Hawkins S, Van Doorselaere J, Piquemal J, Goffner D, Poeydomenge O et al (1997) Cinnamoyl coA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *Plant J.* 11: 429–441.
- Lacombe E, Van Doorselaere J, Boerjan W, Boudet AM, Grima-Pettenati J (2000) Characterization of cis-elements required for vascular expression of the cinnamoyl CoA reductase gene and for protein-DNA complex formation. *Plant J* 23: 663-676
- Lagrimini LM, Burkhart W, Moyer M and Rothstein S (1987) Molecular cloning of complementary DNA encoding the lignin-forming peroxidase from tobacco: molecular analysis and tissue-specific expression. *Proceedings of the National Academy of Sciences of the United States of America* 84:7542–7546.
- Lan T, Yang ZL, Yang X, Liu YJ, Wang XR, Zeng QY (2009) Extensive functional diversification of the *Populus* glutathione S-transferase supergene family. *The Plant Cell Online*, 21(12), 3749. *Am Soc Plant Biol.*
- Le Provost G, Paiva J, Pot D, Brach J, Plomion C (2003) Seasonal variation in transcript accumulation in wood-forming tissues of maritime pine (*Pinus pinaster* Ait.) with emphasis on a cell wall glycine-rich protein. *Planta* 217: 820-830

Leigh F, Lea V, Law J, Wolters P, Powell W, Donini P (2003) Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica* 133: 359-366

Lewin B (1994) *Genes V*, New York

Li Y, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. vol 11, pp 2453-2465

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution* 21: 991-1007

López GA, Potts BM, Dutkowski GW, Rodriguez Traverso JM (2001) Quantitative Genetics of *Eucalyptus globulus*: Affinities of Land Race and Native Stand Localities. *Silvae Genetica* 50: 244-252

Loyall L, Uchida K, Braun S, Furuya M, Frohnmeyer H (2000) Glutathione and a UV light-induced glutathione S-transferase are involved in signaling to chalcone synthase in cell cultures. *The Plant cell* 12: 1939-1950

Lyi SM, Zhou X, Kochian LV and Li L (2007) Biochemical and molecular characterization of the homocysteine S-methyltransferase from broccoli (*Brassica oleracea* var. *italica*). *Phytochemistry*, 68, 1112–1119.

Manly BFJ (1985) *The statistics of natural selection*. Chapman and Hall, London.

Marcucci Poltri SN, Zelener N, Rodriguez Traverso J, Gelid P, Hopp HE (2003) Selection of a seed orchard of *Eucalyptus dunnii* based on genetic diversity criteria calculated using molecular markers *Tree Physiology* 23: 625-632

Marques M, Brondani V, Grattapaglia D, Sederoff R (2002) Conservation and synteny of SSR *loci* and QTLs for vegetative propagation in four *Eucalyptus* species. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 105(2-3), 474-478.

Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S (2006) EGAssembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. *Nucleic acids research* 34: W459-462

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome research* 10: 72-80

Miranda I, Almeida MH, Pereira H (2001) Provenance and site variation of wood density in *Eucalyptus globulus* Labill. at harvest age and its relation to a non-destructive early assessment. *Forest Ecology and Management* 149: 235-240

Molina C, Grotewold E (2005) Genome wide analysis of *Arabidopsis* core promoters. *BMC genomics* 6: 25

- Moran GF, Thamarus KA, Raymond CA, Qiu D, Uren T, Southerton SG (2002) Genomics of *Eucalyptus* wood traits, 59: 645-650.
- Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J* 3: 175-182
- Mueller LA, Goodman CD, Silady RA, Walbot V (2000) AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiol* 123: 1561-1570
- Myburg AA, Potts B, Marques CM, Kirst M, Gion JM, Grattapaglia D, Grima-Pettenati J (2007) *Eucalyptus*. In: Kole C, ed. *Genome mapping and molecular breeding in plants*, Vol. 7: Forest trees. New York, NY, USA: Springer, 115–160.
- Nascimento de Sousa S, Finkeldey R, Gailing O (2005) Experimental verification of microsatellite null alleles in Norway spruce (*Picea abies* [L.] Karst.): Implications for population genetic studies. *Plant Molecular Biology Reporter* 23: 113-119
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.
- Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR and Kirst M (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9, 312
- Paniego N, Echaide M, Munoz M, Fernandez L, Torales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suarez EY, Hopp HE (2002) Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). *Genome* 45: 34-43
- Passardi F, Penel C, Dunand C (2004) Performing the paradoxical: how plant peroxidases modify the cell wall. *Trends in Plant Science* 9(11):534-540.
- Pastorelli R, Smulders MJM, Van't Westende WPC, Vorman B, Giannini R, Vettori C, Vendramin GG (2003) Characterization of microsatellite markers in *Fagus sylvatica* L and *Fagus orientalis* Lipsky. *Mol Ecol Notes* 3:76-78
- Patzlaff A, McInnis S y col. (2003) Characterisation of a pine MYB that regulates lignification. *Plant J* 36(6): 743-54.
- Paux E, Tamasloukht M, Ladouce N, Sivadon P, Grima-Pettenati J (2004) Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant molecular biology* 55: 263-280
- Paux E, Carocha V, Marques C, Mendes de Sousa A, Borralho N, Sivadon P, Grima-Pettenati J (2005) Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *The New phytologist* 167: 89-100
- Payn KG, Dvorak WS, Janse BJH, Myburg AA (2008) Microsatellite diversity and genetic structure of the commercially important tropical tree species *Eucalyptus urophylla*, endemic to seven islands in eastern Indonesia. *Tree Genet Genomes*. 4:519–530.

Peakall R, Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. vol 6, pp 288-295

Pichon M, Courbou I, Beckert M, Boudet AM, Grima-Pettenati J (1998) Cloning and characterization of two maize cDNAs encoding Cinnamoyl-CoA Reductase (CCR) and differential expression of the corresponding genes. *Plant Mol Biol* 38(4): 671-676.

Plomion C, Pionneau C, Brach J, Costa P, Bailleres H (2000) Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). *Plant physiology* 123: 959-969

Plomion C, Leprovost G, Stokes A (2001) Wood formation in trees. *Plant Physiol* 127: 1513-1523

Poke FS, Vaillancourt RE, Elliott RC, Reid JB (2003) Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase 2 (CAD2). *Molecular Breeding* 12: 107-118

Poke F, Raymond CA, Reid J, Vaillancourt RE (2004) The effect of a single amino acid substitution in a lignin biosynthesis enzyme on wood properties in *Eucalyptus globulus*. Borralho N PJ, Marques C, Coutinho J, Madeira M, Tomé M (eds) (ed) IUFRO silviculture and improvement of eucalypts, Raiz Instituto, Aveiro, pp pp 388–394

Potts BM, Wiltshire RJE (1997) *Eucalypt* genetics and genecology. In: Williams J, Woinarski J (eds) *Eucalypt Ecology: Individuals to Ecosystems*. Cambridge University Press, Cambridge, UK

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225-238

Préstamo G, Manzano P (1993) Peroxidases of selected fruits and vegetables and the possible use of ascorbic acid as an antioxidant. *HortScience*, 28, 48–50.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.

Rabello E, Souza ANd, Saito D, Tsai SM (2005) In silico characterization of microsatellites in *Eucalyptus* spp.: abundance, length variation and transposon associations. *Genetics and Molecular Biology* 28: 582-588

Ranik M, Myburg AA (2006) Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. *Tree physiology* 26: 545-556

Ranocha P, Bourgis F, Ziemak MJ, Rhodes D, Gage DA and Hanson AD (2000) Characterization and functional expression of cDNAs encoding methionine-sensitive and -insensitive homocysteine S-methyltransferases from *Arabidopsis*. *J. Biol. Chem.* 275, 15962–15968.

Ranocha P, McNeil SD, Ziemak MJ, Li C, Tarczynski MC and Hanson AD (2001) The S-methylmethionine cycle in angiosperms: ubiquity, antiquity and activity. *Plant J.* 25, 575–584

Rasmussen-Poblete S, Valdés J, Gamboa MC, Valenzuela PDT, Krauskopf E (2008) Generation and analysis of an *Eucalyptus globulus* cDNA library constructed from seedlings subjected to low temperature conditions. *Electronic Journal of Biotechnology Pontificia Universidad Católica de Valparaíso, Chile.* <http://www.ejbiotechnology.info/content/vol11/issue2/full/14/>

Raymond M, Rousset F (1995) GENEPOP Version 1.2: Population genetics software for exact tests and ecumenicism. *J Heredity.* 86:248–249.

Rengel D, San Clemente H, Servant F, Ladouce N, Paux E, Wincker P, Couloux A, Sivadon P, Grima-Pettenati J (2009) A new genomic resource dedicated to wood formation in *Eucalyptus*. *BMC plant biology* 9: 36

Rodríguez Traverso J (2000) Conclusiones del cierre del convenio de mejoramiento genético INTASOPORCEL (I Parte) In: INTA (ed) I Seminario Internacional del *Eucalyptus globulus* en la Argentina, Argentina, pp 11-24

Rose JKC, Braam J, Fry SC and Nishitani K (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature, *Plant Cell Physiol.* 43:1421–1435

Sasaki S, Nishida T, Tsutsumi Y, Kondo R (2004) Lignin dehydrogenative polymerization mechanism: a poplar cell wall peroxidase directly oxidizes polymer lignin and produces in vitro dehydrogenative polymer rich in [beta]-O-4 linkage. *FEBS Lett* 562(1-3):197-201.

Saxena IM and Brown RM (2005) Cellulose biosynthesis: current views and evolving concepts. *Ann Bot (Lond)* 96(1): 9-21.

Sharma RK, Bhardwaj P, Negi R, Mohapatra T, Ahuja PS (2009) Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC plant biology* 9: 53

Shepherd M, Jones M (2004) Molecular markers in tree improvement: characterisation and use in *Eucalyptus*. In: Wenzel HLG (ed) *Molecular marker systems in plant breeding and crop improvement.* Springer, Berlin, pp 399-409

Singh R, Zaki N, Ting N-C, Rosli R, Tan S-G, Low E-T, Ithnin M, Cheah S-C (2008) Exploiting an oil palm EST database for the development of gene-derived SSR markers and their exploitation for assessment of genetic diversity. *Biologia* 63: 227-235

Slavov GT, Howe GT, Yakovlev I, Edwards KJ, Krutovskii KV, Tuskan GA, Carlson JE, Strauss SH, Adams WT (2004) Highly variable SSR markers in Douglas-fir: Mendelian inheritance and map locations. *TAG Theoretical and Applied Genetics* 108: 873-880

Sneath PHA, Sokal RR (1973) *Numerical Taxonomy.* In: Freeman (ed), San Francisco

- Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38: 1409-1438
- Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM (2001) Development and characterisation of microsatellite *loci* in *Eucalyptus globulus* (Myrtaceae). *Silvae Genetica* 50: 89-91
- Steane DA, Nicolle Dean, McKinnon GE, Vaillancourt RE, Potts BM (2002) Higher-level relationships among the eucalypts are resolved by ITS-sequence data. *Australian Systematic Botany* 15, 49–62.
- Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM (2006) A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. *Tree Genetics & Genomes* (2006) 2: 30–38.
- Steane DA, Nicolle D, Sansaloni C, Petroli CD, Carling J, Kilian A, Myburg AA, Grattapaglia D, Vaillancourt RE (2011) Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using highthroughput, genome-wide genotyping. *Mol. Phylogenet. Evol.* doi:10.1016/j.ympev.2011.02.003
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11:1441-1452.
- Thamarus KA, Groom K, Murrell J, Byrne M, Moran GF (2002) A genetic linkage map for *Eucalyptus globulus* with candidate *loci* for wood, fibre, and floral traits. *TAG. Theoretical and applied genetics* 104: 379-387
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and applied genetics* 106: 411-422
- Thom R, Cummins I, Dixon DP, Edwards R, Cole DJ, Laphorn AJ (2002) Structure of a tau class glutathione S-transferase from wheat active in herbicide detoxification. *Biochemistry* 41: 7008-7020
- Thumma BR, Nolan MF, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171: 1257-1265
- Thumma BR, Southerton SG, Bell JC, Owen JV, Henery ML, Moran GF (2010) Quantitative trait *locus* (QTL) analysis of wood quality traits in *Eucalyptus nitens*. *Tree Genet & Genomes* 6:305–317.
- Torales S, Acuña C, Zelener N, Marcucci Poltri S Genetic diversity and population structure in australian races and landraces of *Eucalyptus globulus* using molecular markers. En preparación.

- Turano FJ and Fang TK (1998) Characterization of two glutamate decarboxylase cDNA clones from *Arabidopsis*. *Plant Physiol* 117:1411–1421.
- Ueno S, Taguchi Y, Tsumura Y (2008) Microsatellite markers derived from *Quercus mongolica* var. *crispula* (*Fagaceae*) inner bark expressed sequence tags. *Genes & genetic systems* 83: 179-187
- Van der Nest MA, Steenkamp ET, Wingfield BD, Wingfield MJ (2000) Development of simple sequence repeat (SSR) markers in *Eucalyptus* from amplified inter-simple sequence repeats (ISSR). *Plant Breeding* 119 (5):433-436
- Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WF, Smith JSC, Doebley J. 2002. Rate and pattern of mutation at microsatellite *loci* in maize. *Mol Biol Evol.* 19:1251–1260.
- Vogl C, Karhu A, Moran G, Savolainen O (2002) High resolution analysis of mating systems: inbreeding in natural populations of *Pinus radiata*. *Journal of Evolutionary Biology* 15(3): 433-439
- Voiblet C, Duplessis S, Encelot N, Martin F (2001) Identification of symbiosis-regulated genes in *Eucalyptus globulus-Pisolithus tinctorius* ectomycorrhiza by differential hybridization of arrayed cDNAs. *The Plant Journal* 25: 181-191
- Wang Y, Luo J, Xue X, Korpelainen H, Li C (2005) Diversity of microsatellite markers in the populations of *Picea asperata* originating from the mountains of China. *Plant Science* 168: 707-714
- Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite *loci*: experimental evidence at the dinucleotide *locus* Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, 7, 1569–1573.
- Wright S (1965) The interpretation of population structure by Fstatistics with special regard to systems of mating. *Evolution* 19: 395-420.
- Yang J, Park S, Kamdem DP, Keathley DE, Retzel E, Paule C, Kapur V, Han KH (2003) Novel gene expression profiles define the metabolic and physiological processes characteristic of wood and its extractive formation in a hardwood tree species, *Robinia pseudoacacia*. *Plant molecular biology* 52: 935-956
- Yang S-H, van Zyl L, No E-G, Loopstra CA (2004) Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). *Plant Science* 166: 1185-1195
- Yasodha R, Sumathi R, Chezian P, Kavitha S, Ghosh M (2008) *Eucalyptus* microsatellites mined in silico: survey and evaluation. *Journal of genetics* 87: 21-25
- Yu SJ and Oh S-H (1998) Cloning and characterization of a tobacco cDNA encoding calcium/calmodulin-dependent glutamate decarboxylase. *Mol. Cell* 8 pp. 125–129.

Yu JK, Dake TM, Singh S, Benscher D, Li W, Gill B, Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome* 47: 805-818

Zeh Michaela, Georg Leggewie, Rainer Hoefgen and Holger Hesse (2002) Cloning and characterization of a cDNA encoding a cobalamin-independent methionine synthase from potato (*Solanum tuberosum* L.) *Plant Molecular Biology* 48: 255–265, 2002.

Zelener N, Marcucci Poltri SN, Bartoloni N, Lopez CR, Hopp HE (2005) Selection strategy for a seedling seed orchard design based on trait selection index and genomic analysis by molecular markers: a case study for *Eucalyptus dunnii*. *Tree physiology* 25: 1457-1467

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. *Theoretical and applied genetics* 111: 677-687

Zhou X, Ruan J, Wang G, Zhang W (2007) Characterization and identification of microRNA core promoters in four model species. *PLoS computational biology* 3: e37

Zik M. y col. (1998) Two isoforms of glutamate decarboxylase in *Arabidopsis* are regulated by calcium/calmodulin and differ in organ distribution. *Plant Mol. Biol.* 37:967–975.

Tablas anexas

TABLA ANEXA 2

Información detallada de la distribución de secuencias hasta el 3º nivel gráfico, términos GO (en inglés), número de secuencias y “paterns” para las categorías de: Proceso biológico, Función Molecular y Componente Celular. Los términos se indican en idioma inglés

Proceso Biológico				
Nivel	Término GO	Nº secuencias	Score	Paterns
1	biological_process	153	44,12	
2	biological adhesion	2	1,632	biological_process
2	biological regulation	11	1,742	biological_process
2	cellular process	96	20,98	biological_process
2	developmental process	1	0,254	biological_process
2	establishment of localization	19	6,189	biological_process, localization
2	localization	19	3,713	biological_process
2	metabolic process	127	57,21	biological_process
2	multicellular organismal process	1	0,124	biological_process
2	multi-organism process	1	0,36	biological_process
2	reproduction	1	0,215	biological_process
2	reproductive process	1	0,36	biological_process, reproduction
2	response to stimulus	7	3,984	biological_process
3	biosynthetic process	33	12,16	metabolic_process
3	carbon utilization	4	2,8	metabolic_process
3	catabolic process	3	0,171	metabolic_process
3	cell adhesion	2	2,719	biological adhesion, cellular process
3	cell communication	6	2,424	cellular process
3	cell development	1	0,215	cell differentiation, cellular process
3	cellular component organization and biogenesis	9	2,404	cellular process
3	cellular developmental process	1	0,653	developmental process, cellular process
3	cellular homeostasis	3	1,799	homeostatic process, cellular process
3	cellular localization	4	0,863	cellular component organization and biogenesis, localization
3	cellular metabolic process	83	21,20	metabolic process, cellular process

Continuación Tabla Anexa 2

Nivel	Término GO	Nº secuencias	Score	Paterns
3	death	1	0,215	developmental process
3	defense response	2	2,0	response to stimulus
3	establishment of cellular localization	4	1,44	establishment of localization, cellular localization
3	establishment of protein localization	5	2,280	establishment of localization, protein localization
3	generation of precursor metabolites and energy	2	0,72	cellular metabolic process
3	macromolecule localization	5	0,820	localization
3	macromolecule metabolic process	51	10,53	metabolic process
3	megagametogenesis	1	0,96	embryo sac development, developmental process
3	multicellular organismal development	1	0,207	multicellular organismal process, developmental process
3	nitrogen compound metabolic process	12	4,763	metabolic process
3	photosynthesis	5	5,0	cellular metabolic process
3	pollination	1	0,6	multi-organism process, reproductive process
3	primary metabolic process	78	17,53	metabolic process
3	regulation of biological process	7	1,655	biological process, biological regulation
3	regulation of biological quality	3	0,647	biological regulation
3	regulation of cellular process	5	1,799	cellular process, regulation of biological process
3	regulation of metabolic process	7	2,04	regulation of biological process, metabolic process
3	regulation of molecular function	1	0,6	biological regulation
3	response to abiotic stimulus	1	0,6	response to stimulus
3	response to biotic stimulus	1	1,0	response to stimulus
3	response to chemical stimulus	4	1,680	response to stimulus
3	response to endogenous stimulus	3	1,08	response to stimulus
3	response to stress	4	2,32	response to stimulus
3	secondary metabolic process	1	0,215	metabolic process
3	secretion	1	0,36	establishment of localization

Función Molecular				
Nivel	Término GO	Nº secuencias	Score	Paterns
1	molecular_function	190	75,38	
2	antioxidant activity	1	0,6	molecular_function
2	binding	103	52,97	molecular_function
2	catalytic activity	116	59,34	molecular_function
2	enzyme regulator activity	4	1,075	molecular_function
2	molecular transducer activity	3	1,08	molecular_function
2	structural molecule activity	9	5,8	molecular_function
2	transcription regulator activity	10	6,0	molecular_function
2	translation regulator activity	2	0,72	molecular_function
2	transporter activity	10	4,300	molecular_function
3	peroxidase activity	1	1,0	antioxidant activity, oxidoreductase activity, acting on peroxide as acceptor
3	amine binding	1	0,6	binding
3	cofactor binding	7	4,199	binding
3	ion binding	26	7,343	binding
3	lipid binding	1	0,6	binding
3	metal cluster binding	3	1,799	binding
3	nucleic acid binding	28	20,72	binding
3	nucleotide binding	23	6,592	binding
3	protein binding	20	16,64	binding
3	tetrapyrrole binding	8	4,8	binding
3	vitamin binding	4	2,4	binding
3	cyclase activity	1	0,6	catalytic activity
3	hydrolase activity	18	5,574	catalytic activity
3	isomerase activity	8	2,159	catalytic activity
3	ligase activity	5	0,964	catalytic activity
3	lyase activity	9	2,832	catalytic activity
3	oxidoreductase activity	29	21,13	catalytic activity
3	transferase activity	31	9,573	catalytic activity

Continuación Tabla Anexa 2

Nivel	Término GO	Nº secuencias	Score	Patterns
3	transcription factor activity	8	8,0	DNA binding, transcription regulator activity
3	enzyme inhibitor activity	3	1,432	enzyme regulator activity
3	phosphatase regulator activity	1	0,36	enzyme regulator activity
3	signal transducer activity	3	1,799	molecular transducer activity
3	two-component response regulator activity	1	1,0	signal transducer activity, transcription regulator activity
3	structural constituent of cell wall	1	1,0	structural molecule activity
3	structural constituent of ribosome	7	7,0	structural molecule activity
3	transcription repressor activity	1	1,0	transcription regulator activity
3	translation factor activity, nucleic acid binding	2	1,2	translation regulator activity, nucleic acid binding
3	substrate-specific transporter activity	6	2,069	transporter activity
3	transmembrane transporter activity	4	0,608	transporter activity
Componente Celular				
1	cellular_component	82	15,30	
2	cell part	81	30,20	cell, cellular_component
2	cell	81	18,12	cellular_component
2	macromolecular complex	30	9,524	cellular_component
2	organelle	35	6,120	cellular_component
2	extracellular region	3	3,0	cellular_component
2	organelle part	13	2,289	cellular_component, organelle
2	envelope	3	0,388	cellular_component
3	membrane	34	26,56	cell part
3	intracellular	52	21,01	cell part
3	intracellular part	48	15,02	intracellular, cell part
3	protein complex	23	11,45	macromolecular complex
3	intracellular organelle	35	9,786	organelle, intracellular part
3	membrane-bounded organelle	27	7,257	organelle

Nivel	Término GO	Nº secuencias	Score	Patterns
3	ribonucleoprotein complex	7	4,416	macromolecular complex, intracellular part
3	protein serine/threonine phosphatase complex	4	3,6	protein complex, cell part
3	membrane part	9	3,412	membrane, cell part
3	intracellular organelle part	13	3,383	organelle part, intracellular organelle, intracellular part
3	non-membrane-bounded organelle	9	2,943	organelle
3	organelle membrane	4	1,126	membrane-bounded organelle, organelle part, membrane
3	organelle envelope	3	0,647	membrane-bounded organelle, envelope, intracellular organelle part
3	external encapsulating structure	1	0,36	cell part
3	vesicle	1	0,077	organelle

TABLA ANEXA 3

Información de las posibles enzimas que contienen secuencias SSR. La predicción se basa en homología de secuencia (BLASTX) y asignación GO mediante programa Blast2GO. Se incluye información de rutas metabólicas asociadas, mapas KEGG, los nombres de las enzimas e información de las secuencias (contig/singletons) que las representan. Las enzimas y rutas metabólicas se indican en idioma inglés

Ruta metabólica	Mapa Kegg	Enzimas	secuencias
Carbon fixation	map00710.gif	ec:2.7.2.3 - phosphoglycerate kinase	Contig352
		ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:2.2.1.1 - transketolase	Contig453
		ec:5.1.3.1 - ribulose-phosphate 3-epimerase	Contig172, gi 162322247 gb ES591526.1 ES591526
		ec:4.1.1.39 - ribulose-bisphosphate carboxylase	gi 162325275 gb ES590523.1 ES590523, gi 162322349 gb ES593784.1 ES593784, Contig915
Pentose phosphate pathway	map00030.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:2.7.1.11 - 6-phosphofructokinase	Contig638
		ec:2.2.1.1 - transketolase	Contig453
		ec:5.1.3.1 - ribulose-phosphate 3-epimerase	Contig172, gi 162322247 gb ES591526.1 ES591526
Propanoate metabolism	map00640.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:1.3.99.3 - acyl-CoA dehydrogenase	gi 162324385 gb ES594197.1 ES594197, gi 162324247 gb ES594185.1 ES594185, Contig945
Starch and sucrose metabolism	map00500.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:3.1.1.11 - pectinesterase	gi 162328606 gb ES592242.1 ES592242
		ec:4.1.1 - pyruvate decarboxylase	Contig557
		ec:2.4.1.21 - starch synthase	Contig595
Methionine metabolism	map00271.gif	ec:3.3.1.1 - adenosylhomocysteinase	gi 162323228 gb ES594621.1 ES594621

Ruta metabólica	Mapa Kegg	Enzimas	secuencias
Methionine metabolism	map00271.gif	ec:2.1.1.14 - 5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	Contig938
		ec:2.1.1.10 - homocysteine S-methyltransferase	gi 162325711 gb ES588789.1 ES588789
		ec:4.1.1 - pyruvate decarboxylase	Contig557
		ec:2.5.1.6 - methionine adenosyltransferase	Contig423
Nitrogen metabolism	map00910.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:6.3.1.2 - glutamate--ammonia ligase	Contig361
		ec:4.1.1 - pyruvate decarboxylase	Contig557
		ec:4.2.1.1 - carbonate dehydratase	Contig1098
beta-Alanine metabolism	map00410.gif	ec:1.3.99.3 - acyl-CoA dehydrogenase	gi 162324385 gb ES594197.1 ES594197, gi 162324247 gb ES594185.1 ES594185, Contig945
		ec:4.1.1.15 - glutamate decarboxylase	Contig557
Sulfur metabolism	map00920.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:2.3.1.30 - serine O-acetyltransferase	gi 162325605 gb ES589775.1 ES589775, gi 162324132 gb ES595414.1 ES595414
Glycolysis / Gluconeogenesis	map00010.gif	ec:2.7.2.3 - phosphoglycerate kinase	Contig352
		ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:2.7.1.11 - 6-phosphofructokinase	Contig638
Selenoamino acid metabolism	map00450.gif	ec:3.3.1.1 - adenosylhomocysteinase	gi 162323228 gb ES594621.1 ES594621
		ec:4.1.1 - pyruvate decarboxylase	Contig557
		ec:2.5.1.6 - methionine adenosyltransferase	Contig423
Valine, leucine and isoleucine degradation	map00280.gif	ec:1.3.99.3 - acyl-CoA dehydrogenase	gi 162324385 gb ES594197.1 ES594197, gi 162324247 gb ES594185.1 ES594185, Contig945
PPAR signaling pathway	map03320.gif	ec:1.3.99.3 - acyl-CoA dehydrogenase	gi 162324385 gb ES594197.1 ES594197, gi 162324247 gb ES594185.1 ES594185, Contig945

Ruta metabólica	Mapa Kegg	Enzimas	secuencias
Insulin signaling pathway	hsa04910.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
		ec:2.7.1.11 - 6-phosphofructokinase	Contig638
		ec:4.1.1 - pyruvate decarboxylase	Contig557
Glyoxylate and dicarboxylate metabolism	map00630.gif	ec:4.1.1.39 - ribulose-bisphosphate carboxylase	gi 162325275 gb ES590523.1 ES590523, gi 162322349 gb ES593784.1 ES593784, Contig915
Histidine metabolism	map00340.gif	ec:2.4.2.17 - ATP phosphoribosyltransferase	Contig1012, gi 162321034 gb ES595731.1 ES595731
		ec:6.1.1.21 - histidine--tRNA ligase	gi 162326821 gb ES589253.1 ES589253
Cysteine metabolism	map00272.gif	ec:4.1.1 - pyruvate decarboxylase	Contig557
		ec:2.3.1.30 - serine O-acetyltransferase	gi 162325605 gb ES589775.1 ES589775, gi 162324132 gb ES595414.1 ES595414
Reductive carboxylate cycle (CO2 fixation)	map00720.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:4.1.1 - pyruvate decarboxylase	Contig557
Pyruvate metabolism	map00620.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:4.1.1 - pyruvate decarboxylase	Contig557
Butanoate metabolism	map00650.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:4.1.1.15 - glutamate decarboxylase	Contig557
Pentose and glucuronate interconversions	map00040.gif	ec:3.1.1.11 - pectinesterase	gi 162328606 gb ES592242.1 ES592242
		ec:5.1.3.1 - ribulose-phosphate 3-epimerase	Contig172, gi 162322247 gb ES591526.1 ES591526
Fatty acid metabolism	map00071.gif	ec:1.3.99.3 - acyl-CoA dehydrogenase	gi 162324385 gb ES594197.1 ES594197, gi 162324247 gb ES594185.1 ES594185, Contig945
Biosynthesis of ansamycins	map01051.gif	ec:2.1.1 - nicotinamide N-methyltransferase	gi 162326362 gb ES592322.1 ES592322, Contig334
		ec:2.2.1.1 - transketolase	Contig453
Ubiquitin mediated proteolysis	map04120.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
		ec:5.2.1.8 - peptidylprolyl isomerase	Contig162

Ruta metabólica	Mapa Kegg	Enzimas	secuencias
Oxidative phosphorylation	map00190.gif	ec:3.6.1.1 - inorganic diphosphatase ec:3.6.1.10 endopolyphosphatase	gi 162327917 gb ES589901.1 ES589901, gi 162327901 gb ES595863.1 ES595863
Alanine and aspartate metabolism	map00252.gif	ec:4.1.1.15 - glutamate decarboxylase	Contig557
		ec:4.1.1 - pyruvate decarboxylase	Contig557
Glutamate metabolism	map00251.gif	ec:6.3.1.2 - glutamate--ammonia ligase	Contig361
		ec:4.1.1.15 - glutamate decarboxylase	Contig557
p53 signaling pathway	map04115.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
		ec:1.17.4.1 - ribonucleoside-diphosphate reductase	gi 162328099 gb ES593737.1 ES593737 gi 162326362 gb ES592322.1 ES592322, Contig334
Nicotinate and nicotinamide metabolism	map00760.gif	ec:2.1.1 - nicotinamide N-methyltransferase	Contig334
Galactose metabolism	map00052.gif	ec:2.7.1.11 - 6-phosphofructokinase	Contig638
		ec:5.1.3.2 - UDP-glucose 4-epimerase	Contig446
Inositol phosphate metabolism	map00562.gif	ec:5.5.1.4 - inositol-3-phosphate synthase	Contig114, gi 162325138 gb ES595314.1 ES595314
Neurodegenerative Diseases	map01510.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
		ec:1.15.1.1 - superoxide dismutase	Contig980
Taurine and hypotaurine metabolism	map00430.gif	ec:4.1.1.15 - glutamate decarboxylase	Contig557
		ec:4.1.1 - pyruvate decarboxylase	Contig557
Cyanoamino acid metabolism	map00460.gif	ec:3.5.5.4 - cyanoalanine nitrilase	Contig1008
		ec:4.2.1.65 - 3-cyanoalanine hydratase	Contig1008
Streptomycin biosynthesis	map00521.gif	ec:5.5.1.4 - inositol-3-phosphate synthase	Contig114, gi 162325138 gb ES595314.1 ES595314
Jak-STAT signaling pathway	map04630.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Wnt signaling pathway	map04310.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Calcium signaling pathway	map04020.gif	ec:5.2.1.8 - peptidylprolyl isomerase	Contig162
T cell receptor signaling pathway	map04660.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Amyotrophic lateral sclerosis (ALS)	map05030.gif	ec:1.15.1.1 - superoxide dismutase	Contig980
Porphyrin and chlorophyll metabolism	map00860.gif	ec:4.2.1.24 - porphobilinogen synthase	Contig371
Two-component system - General	map02020.gif	ec:6.3.1.2 - glutamate--ammonia ligase	Contig361
Chronic myeloid leukemia	map05220.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283

Ruta metabólica	Mapa Kegg	Enzimas	secuencias
Cell cycle	map04110.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Type I diabetes mellitus	map04940.gif	ec:4.1.1.15 - glutamate decarboxylase	Contig557
ErbB signaling pathway	map04012.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Drug metabolism - other enzymes	map00983.gif	ec:2.1.1.67 - thiopurine S-methyltransferase	gi 162322157 gb ES590288.1 ES590288
Citrate cycle (TCA cycle)	map00020.gif	ec:4.1.1 - pyruvate decarboxylase	Contig557
Fructose and mannose metabolism	map00051.gif	ec:2.7.1.11 - 6-phosphofructokinase	Contig638
Asthma	map05310.gif	ec:1.11.1.7 - peroxidase	gi 162327290 gb ES588966.1 ES588966
Bladder cancer	map05219.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Parkinson's disease	map05020.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Melanoma	map05218.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Dentatorubropallidoluysian atrophy (DRPLA)	map05050.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Prostate cancer	map05215.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Glioma	map05214.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Pyrimidine metabolism	map00240.gif	ec:1.17.4.1 - ribonucleoside-diphosphate reductase	gi 162328099 gb ES593737.1 ES593737
Phenylpropanoid biosynthesis	map00940.gif	ec:1.11.1.7 - peroxidase	gi 162327290 gb ES588966.1 ES588966
Lysine biosynthesis	map00300.gif	ec:5.1.1.7 - diaminopimelate epimerase	gi 162328079 gb ES591393.1 ES591393
TGF-beta signaling pathway	map04350.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Aminoacyl-tRNA biosynthesis	map00970.gif	ec:6.1.1.21 - histidine--tRNA ligase	gi 162326821 gb ES589253.1 ES589253
Methane metabolism	map00680.gif	ec:1.11.1.7 - peroxidase	gi 162327290 gb ES588966.1 ES588966
Phenylalanine metabolism	map00360.gif	ec:1.11.1.7 - peroxidase	gi 162327290 gb ES588966.1 ES588966
Huntington's disease	map05040.gif	ec:6.3.2.19 - ubiquitin--protein ligase	gi 162320648 gb ES596283.1 ES596283
Nucleotide sugars metabolism	map00520.gif	ec:5.1.3.2 - UDP-glucose 4-epimerase	Contig446
Purine metabolism	map00230.gif	ec:1.17.4.1 - ribonucleoside-diphosphate reductase	gi 162328099 gb ES593737.1 ES593737
Peptidoglycan biosynthesis	map00550.gif	ec:6.3.1.2 - glutamate--ammonia ligase	Contig361
Glycine, serine and threonine metabolism	map00260.gif	ec:4.1.1 - pyruvate decarboxylase	Contig557

TABLA ANEXA 4

Matriz de datos de los alelos encontrados en los 60 individuos de *E. globulus* analizados mediante 27 marcadores SSR funcionales (EST-SSRs y GC-SSRs). Se indica el nombre de los individuos en la primera columna. Los datos codominantes se presentan en dos columnas por *locus*. Los alelos se codificaron como tamaño total en pares de bases (pb).

individuo	pCCR		CesA3_a		CesA3_b		LIM1		HMTi		HMTe		SK		XTH2		GST		GAD_A		GAD_B		PER		EST1		EST3		EST5	
181	182	208	294	296	346	346	161	169	435	435	221	221	345	349	129	134	104	112	255	255	185	189	266	267	0	0	245	246	333	333
190	194	194	290	290	356	358	161	161	435	440	221	221	351	351	129	129	100	120	248	248	185	185	264	268	0	0	243	243	333	333
192	182	204	290	296	0	0	169	175	417	417	221	224	351	367	129	129	104	116	256	256	185	189	267	270	161	161	246	246	333	333
194	182	182	290	290	346	359	161	169	419	435	221	221	351	355	129	129	100	112	256	256	189	189	266	267	168	168	243	246	333	333
201	204	204	290	294	342	356	169	173	439	439	221	221	351	351	129	129	112	112	256	256	185	185	265	267	0	0	243	243	333	333
203	196	196	288	306	348	348	159	169	417	425	0	0	331	349	129	129	93	93	0	0	185	189	266	269	0	0	245	245	333	333
205	182	182	294	298	346	367	169	169	425	435	221	224	367	367	0	0	93	107	248	256	185	185	0	0	0	0	246	246	333	333
206	182	182	298	298	346	346	159	159	415	418	0	0	351	367	134	134	93	104	248	256	185	185	268	268	132	132	246	246	333	333
213	182	182	298	298	346	346	169	169	417	419	221	221	351	386	0	0	95	107	256	256	185	185	0	0	130	130	245	245	333	333
214	182	182	294	302	342	350	167	169	417	417	221	221	345	367	129	129	100	112	248	248	185	185	268	268	130	130	243	243	333	333
216	183	0	290	298	346	356	169	179	417	417	221	224	351	351	0	0	93	116	248	256	185	189	269	269	140	140	246	246	333	333
218	188	206	294	306	348	356	165	175	417	425	221	224	347	351	129	134	100	112	248	248	185	185	265	267	140	140	252	252	333	333
219	182	182	290	296	348	365	169	169	417	435	221	224	351	359	129	134	100	100	248	256	185	185	268	269	130	130	251	252	333	333
221	208	208	290	290	358	358	167	171	425	425	221	221	352	369	129	134	93	107	248	248	185	185	267	267	130	130	245	245	333	333
224	216	216	290	304	346	358	163	175	425	431	221	221	331	355	129	129	93	116	248	248	189	189	267	268	0	0	246	246	333	333
225	188	188	290	290	346	354	169	186	435	437	221	224	351	351	129	134	112	112	248	248	185	185	267	268	151	151	246	246	333	333
227	182	182	290	290	356	361	169	194	417	417	221	221	349	367	129	134	104	112	248	255	185	185	269	269	140	140	246	246	333	333
229	182	182	290	306	348	356	0	0	427	427	221	221	331	331	129	129	93	116	248	248	185	185	268	268	132	132	246	246	333	333
230	188	204	294	294	350	360	159	169	417	425	222	222	347	359	129	129	112	117	248	248	185	185	267	267	132	140	243	243	333	333
232	182	0	290	294	345	348	159	161	435	435	221	221	331	351	129	129	95	112	248	248	185	185	268	268	132	132	251	252	333	333
237	206	0	296	296	356	358	159	175	0	0	0	0	351	351	0	0	113	116	248	249	185	185	265	269	0	0	243	243	333	333
240	182	182	290	290	348	348	159	175	417	425	221	221	331	351	129	129	93	100	248	248	185	189	0	0	132	132	246	246	333	333
241	182	188	289	294	346	354	159	177	417	417	221	221	331	351	129	134	97	112	248	248	185	185	267	268	130	130	245	245	333	333
245	0	0	290	290	0	0	0	0	0	0	221	224	0	0	0	0	0	0	248	249	185	185	0	0	0	0	245	246	333	333
247	208	210	294	306	348	356	159	161	417	435	221	221	349	355	129	129	83	112	248	248	185	185	265	265	140	140	0	0	333	333
250	182	182	288	288	356	356	161	175	435	440	221	221	331	331	129	129	100	116	248	256	175	185	266	268	140	140	0	0	333	333
255	182	196	290	290	356	358	159	171	442	465	221	221	331	351	122	129	107	112	248	256	185	185	268	268	130	132	243	243	0	0

individuo	pCCR		CesA3_a		CesA3_b		LIM1		HMTi		HMTe		SK		XTH2		GST		GAD_A		GAD_B		PER		EST1		EST3		EST5	
256	182	182	296	296	354	354	173	173	417	439	221	221	331	331	0	0	97	97	248	248	185	185	269	269	132	132	245	246	333	333
257	182	182	290	290	350	356	171	173	435	439	221	221	331	351	0	0	97	116	248	248	185	185	268	268	130	132	245	245	333	333
259	188	206	296	296	350	350	169	178	417	425	221	221	331	355	129	129	100	116	248	248	0	0	268	268	140	140	246	246	333	333
262	182	182	294	296	346	346	173	173	418	431	221	221	331	355	129	129	100	107	248	256	185	185	265	268	132	132	243	243	333	333
264	182	182	294	294	346	348	159	159	425	435	221	221	332	332	129	134	112	112	248	248	185	189	266	268	130	140	251	252	333	333
265	182	204	290	290	342	350	159	169	435	437	221	221	351	351	129	129	112	116	248	256	185	185	265	269	140	140	246	246	333	333
266	182	182	290	296	348	356	159	159	425	440	221	221	351	351	129	129	104	116	248	248	185	185	268	269	130	132	246	246	333	333
267	0	0	0	0	0	0	0	0	0	0	221	221	331	331	129	129	93	107	248	248	185	189	0	0	132	132	246	246	333	333
274	204	208	294	306	348	356	164	184	417	435	221	221	351	351	129	134	97	112	248	248	185	185	267	268	132	132	245	246	333	333
279	198	198	294	294	348	348	165	165	417	425	221	221	351	371	129	134	97	112	248	256	185	189	270	270	132	132	246	246	333	333
280	206	216	290	290	356	358	159	161	417	417	221	221	331	349	129	134	97	112	256	256	185	185	267	267	130	132	246	246	333	339
281	182	182	298	298	0	0	0	0	419	437	221	221	351	367	129	134	100	107	248	252	185	185	0	0	140	140	243	243	333	333
285	196	204	285	294	350	356	159	169	417	435	221	221	359	367	129	134	112	116	248	256	185	189	268	268	0	0	252	252	333	333
287	182	182	288	290	356	359	159	169	418	435	221	221	355	355	129	129	93	116	248	248	185	189	264	266	140	140	243	243	333	333
293	206	206	290	299	348	356	159	163	435	435	221	221	331	355	0	0	93	93	248	248	185	189	266	268	0	0	0	0	333	333
295	196	214	304	306	350	350	163	171	417	425	221	221	351	351	129	129	97	116	256	256	185	189	268	269	132	132	0	0	333	333
297	0	0	289	293	0	0	171	173	419	425	221	221	0	0	129	129	97	112	248	249	185	185	0	0	130	132	0	0	0	0
298	0	0	0	0	0	0	0	0	425	425	221	221	0	0	0	0	97	107	248	249	185	185	0	0	130	168	0	0	333	333
299	0	0	0	0	0	0	159	165	417	435	0	0	0	0	0	0	0	0	248	249	185	185	0	0	0	0	245	246	0	0
300	182	204	290	290	356	358	161	173	0	0	221	221	0	0	129	129	112	113	256	256	185	185	0	0	130	132	246	246	333	333
301	182	182	290	298	345	346	169	169	435	437	221	221	367	367	129	129	107	112	248	255	185	185	262	269	161	161	243	246	333	333
302	182	0	290	290	356	358	163	167	417	0	221	221	331	331	129	134	104	104	248	248	185	185	266	266	132	132	245	245	333	333
303	182	0	294	296	348	348	175	177	419	437	221	221	351	351	122	129	97	97	248	248	185	185	268	268	130	132	246	246	333	333
304	182	182	0	0	356	365	159	169	417	439	221	221	351	353	129	129	97	100	248	248	185	185	267	267	130	140	245	246	333	333
305	182	182	294	296	348	348	167	177	417	437	221	221	331	331	129	129	97	97	248	255	185	185	266	266	130	132	246	246	333	333
306	0	0	290	290	348	356	161	161	435	439	0	0	351	359	129	134	104	107	248	255	185	185	268	268	130	132	246	246	333	333
308	182	182	0	0	348	356	163	173	417	417	221	221	331	331	0	0	97	104	248	256	185	185	268	268	132	132	243	243	333	333
309	0	0	0	0	354	356	165	175	435	435	221	221	351	351	0	0	0	0	248	248	185	185	265	268	132	132	246	246	333	333
310	182	182	290	290	348	356	163	175	417	427	221	221	331	349	129	129	112	116	248	248	189	189	264	267	0	0	0	0	333	333
312	182	182	289	290	346	346	159	171	437	439	221	221	351	361	134	134	107	113	248	249	185	185	0	0	130	132	246	246	333	333
313	182	182	294	294	348	356	159	171	417	435	221	224	331	331	129	129	100	117	248	248	185	185	268	268	132	132	243	243	333	333
314	0	0	289	290	356	358	171	175	417	435	221	221	331	355	129	129	97	112	248	248	185	185	267	267	132	132	0	0	333	333
315	0	0	296	307	356	356	161	186	417	439	221	221	351	361	129	129	112	120	256	257	185	186	267	267	130	132	243	243	333	333

individuo	EST11		EST33		EST46		EST58		EST88		EST125		EST134		EST141		EST157		EST191		EST197		EST200		EST216		EST219	
181	243	245	144	144	211	212	129	140	233	233	204	204	318	318	305	308	323	341	216	222	95	95	333	333	404	404	0	0
190	0	0	144	144	211	212	135	135	233	249	202	202	318	318	305	305	0	0	0	0	95	95	333	333	404	407	108	110
192	243	245	144	144	211	212	0	0	233	249	202	202	303	316	305	305	310	317	212	214	95	96	333	335	0	0	108	110
194	245	245	144	144	212	212	0	0	233	233	202	202	316	316	305	305	294	294	212	214	96	96	333	333	407	409	104	108
201	245	249	144	144	211	212	129	135	233	233	202	202	316	316	305	305	0	0	222	222	95	95	333	333	397	404	108	108
203	243	249	144	144	211	211	135	135	233	233	202	207	322	322	305	308	0	0	204	220	95	104	333	333	397	404	106	110
205	243	243	0	0	211	211	129	140	233	233	202	202	314	318	305	305	310	323	214	214	96	96	333	333	396	404	104	108
206	241	245	144	144	211	211	140	140	233	233	0	0	320	320	305	305	337	337	212	214	95	95	333	333	402	412	108	110
213	245	245	144	144	211	211	0	0	233	233	202	202	320	320	305	305	310	317	0	0	95	95	333	333	404	404	110	128
214	243	245	144	144	211	211	135	146	233	233	202	207	0	0	305	305	323	341	218	220	95	96	333	333	402	411	110	114
216	243	243	144	144	210	211	135	140	0	0	202	202	0	0	305	305	310	310	214	214	95	96	0	0	404	407	0	0
218	0	0	144	144	210	211	119	135	233	233	202	202	303	316	305	305	294	294	0	0	96	96	333	333	393	404	108	110
219	249	249	144	144	211	211	135	135	233	233	207	207	318	318	305	308	335	337	212	212	95	95	333	333	400	404	104	108
221	243	243	144	144	211	211	129	135	233	233	202	207	303	316	305	305	325	325	0	0	95	104	333	335	404	407	104	108
224	245	245	144	144	211	212	0	0	0	0	202	207	322	322	305	305	317	318	214	220	96	96	333	335	0	0	106	108
225	243	247	144	144	210	211	135	140	233	233	207	207	316	318	305	308	317	339	0	0	96	96	333	337	397	400	104	108
227	245	245	144	144	210	212	119	129	233	233	202	202	297	297	305	305	306	323	0	0	95	95	333	333	404	404	102	110
229	247	247	144	144	211	211	135	135	233	233	202	207	322	322	0	0	294	320	220	220	104	105	333	333	0	0	106	108
230	245	245	144	144	212	213	135	135	233	233	202	202	322	322	305	305	317	317	0	0	95	95	333	335	0	0	0	0
232	241	245	144	144	210	210	135	135	233	233	0	0	0	0	305	305	320	323	220	220	95	95	333	333	404	407	106	110
237	245	245	144	144	210	210	0	0	233	233	204	204	316	316	305	305	310	310	216	216	95	95	333	335	0	0	106	124
240	245	245	144	144	210	210	129	135	233	233	202	207	316	318	308	308	0	0	0	0	96	96	333	333	400	404	110	116
241	241	245	144	144	210	210	119	135	233	233	0	0	316	316	305	305	294	294	214	220	96	105	333	333	404	404	106	110
245	245	249	136	144	210	211	0	0	233	233	220	202	0	0	305	308	294	313	212	212	105	105	0	0	396	404	106	108
247	247	249	144	144	212	212	129	135	233	233	202	207	316	316	305	305	323	327	212	220	96	105	333	333	393	393	106	108
250	245	245	144	144	211	212	135	135	233	233	207	207	316	322	305	308	294	323	0	0	96	96	333	335	397	404	106	108
255	243	245	144	144	210	210	129	135	233	233	0	0	316	318	305	305	323	323	220	220	95	104	0	0	397	404	108	110
256	0	0	144	144	211	211	135	135	233	233	202	202	0	0	0	0	325	325	220	220	95	95	0	0	0	0	108	110
257	247	247	144	144	211	211	135	135	233	233	202	207	0	0	0	0	325	325	214	224	96	96	0	0	0	0	110	128
259	0	0	144	144	211	212	135	135	233	233	202	202	316	316	305	308	294	294	214	214	95	95	333	335	0	0	108	110
262	245	245	144	144	211	212	129	135	233	233	0	0	318	318	0	0	320	323	0	0	95	95	333	333	397	397	106	108
264	245	247	144	144	211	211	140	140	233	233	207	207	322	322	0	0	313	325	214	220	96	96	333	333	397	404	110	116
265	243	243	144	144	212	212	0	0	233	233	202	202	316	316	305	305	0	0	212	214	96	96	333	333	0	0	110	122
266	245	247	144	144	210	210	129	135	233	233	202	202	316	322	305	305	294	294	0	0	95	95	333	333	0	0	110	112

individuo	EST11		EST33		EST46		EST58		EST88		EST125		EST134		EST141		EST157		EST191		EST197		EST200		EST216		EST219		
267	245	247	0	0	210	211	135	135	233	233	202	202	316	316	308	308	310	317	0	0	95	95	0	0	404	404	104	106	
274	245	247	144	144	211	211	135	135	233	233	0	0	0	0	0	0	317	317	214	220	96	96	0	0	404	404	110	112	
279	0	0	144	144	212	212	119	135	233	233	202	202	0	0	305	305	302	325	212	212	95	96	333	333	400	400	106	110	
280	249	249	144	144	210	211	135	135	233	233	207	207	316	316	305	0	317	323	212	220	0	0	333	333	396	404	108	110	
281	245	245	144	144	211	211	135	140	233	233	202	202	316	320	305	305	323	327	214	216	95	96	0	0	400	400	108	108	
285	245	245	144	144	211	211	129	140	233	233	202	202	303	303	305	0	0	0	0	0	0	96	96	333	333	0	0	104	108
287	245	245	144	144	211	211	129	135	233	233	204	207	314	318	308	308	318	325	0	0	95	96	333	333	400	404	104	108	
293	0	0	144	144	211	211	135	135	233	233	202	202	316	316	305	305	318	327	216	220	95	95	0	0	0	0	104	108	
295	245	245	144	144	210	211	119	129	0	0	202	207	0	0	0	0	0	0	0	0	96	104	333	333	397	407	106	110	
297	247	247	144	144	210	211	135	135	233	233	202	202	316	322	305	305	310	325	220	220	96	96	333	333	400	402	106	112	
298	243	245	144	144	212	212	129	140	233	233	202	202	314	320	305	308	294	294	212	220	96	96	333	333	396	404	108	108	
299	245	245	144	144	211	212	129	129	233	233	202	202	316	322	305	308	335	337	214	216	95	95	0	0	0	0	110	110	
300	241	247	144	144	212	212	135	135	233	233	202	202	297	316	305	305	317	339	212	220	95	104	333	333	396	404	108	108	
301	245	245	144	144	211	212	135	135	233	233	202	202	316	316	305	305	317	323	220	220	95	96	0	0	0	0	110	110	
302	245	247	0	0	210	210	135	135	233	233	202	202	312	316	305	305	327	329	220	220	95	95	333	333	404	404	106	108	
303	245	247	144	144	210	211	129	129	233	233	202	202	314	316	305	305	310	310	214	214	95	104	333	333	0	0	110	110	
304	245	245	144	144	210	211	135	135	233	233	202	202	314	316	305	305	294	310	212	214	95	96	333	333	404	404	106	108	
305	247	249	144	144	211	211	129	129	233	233	202	207	314	316	305	305	313	323	220	222	95	95	0	0	0	0	108	110	
306	243	245	144	144	210	211	119	135	233	233	202	202	312	316	305	308	323	327	212	214	95	96	333	333	404	407	108	110	
308	245	247	144	144	210	210	129	135	233	233	202	202	316	316	305	305	323	325	222	224	95	95	0	0	404	404	108	108	
309	243	245	144	144	210	210	0	0	233	233	202	202	312	316	308	308	317	323	220	220	0	0	0	0	397	400	110	110	
310	245	245	144	144	211	211	129	135	233	233	0	0	312	314	305	315	310	310	218	218	95	95	0	0	404	404	108	108	
312	243	243	144	144	211	212	135	135	233	233	202	202	316	318	305	305	327	329	212	214	96	96	0	0	402	411	108	110	
313	245	249	144	144	210	210	135	140	233	233	202	202	322	322	305	308	325	325	220	226	95	104	0	0	0	0	110	112	
314	245	247	144	144	210	210	135	140	233	233	0	0	299	316	305	305	294	294	220	220	96	96	0	0	0	0	108	110	
315	245	249	144	144	211	212	124	135	233	233	202	202	312	312	305	308	317	323	214	228	95	95	0	0	402	402	106	108	

TABLA ANEXA 5

Matriz de datos de los alelos encontrados en los 47 individuos de las siete especies estudiadas: *E. camaldulensis*, *E. dunnii*, *E. grandis*, *E. saligna*, *E. tereticornis*, *E. viminialis* y *E. globulus*, analizados mediante 27 marcadores SSR funcionales (EST-SSRs y GC-SSRs).

Se indica el nombre de los individuos y especie en la primera y segunda columna, respectivamente. Los datos codominantes se presentan en dos columnas por *locus*. Los alelos se codificaron como tamaño total en pares de bases (pb).

individuo	especie	pCCR		CesA3_A		CesA3_B		LIM		HMTi		HMTe		SK		XTH2		GST		GAD_A		GAD_B		PER	
Ec2	<i>E. camaldulensis</i>	186	188	290	298	346	352	165	182	433	437	221	221	0	0	129	130	86	86	248	255	185	191	265	265
Ec3	<i>E. camaldulensis</i>	182	231	289	289	344	350	157	175	0	0	221	221	0	0	129	129	114	114	248	255	185	191	265	265
Ec4	<i>E. camaldulensis</i>	170	170	282	282	334	350	182	194	444	444	221	221	0	0	129	129	122	137	249	249	186	190	259	267
Ec5	<i>E. camaldulensis</i>	182	231	289	297	343	344	175	196	421	421	221	221	0	0	129	129	114	114	248	248	185	185	265	265
Ec6	<i>E. camaldulensis</i>	188	231	293	300	340	344	173	173	411	457	221	221	0	0	129	129	96	96	248	249	185	190	259	264
Ec7	<i>E. camaldulensis</i>	203	231	289	289	336	344	175	175	421	431	221	221	0	0	129	129	114	114	248	254	185	191	265	266
Ec8	<i>E. camaldulensis</i>	206	231	298	312	334	346	163	165	450	452	221	221	0	0	129	129	97	99	248	248	185	185	265	267
Ed1	<i>E. dunnii</i>	169	182	290	294	350	361	163	163	437	437	0	0	335	371	134	134	90	90	249	257	190	190	0	0
Ed2	<i>E. dunnii</i>	169	169	294	294	346	350	0	0	0	0	221	224	333	380	134	134	97	97	261	264	190	190	0	0
Ed3	<i>E. dunnii</i>	170	170	299	300	354	365	173	173	0	0	221	221	380	380	129	134	105	107	249	264	186	190	265	270
Ed4	<i>E. dunnii</i>	0	0	0	0	0	0	169	179	431	431	221	222	333	363	134	134	90	90	257	264	186	190	264	265
Ed5	<i>E. dunnii</i>	182	200	290	296	350	371	159	159	431	435	221	224	363	371	134	134	107	107	256	261	189	190	264	270
Ed6	<i>E. dunnii</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	103	116	0	0	0	0	0	0
Ed7	<i>E. dunnii</i>	169	182	0	0	0	0	0	0	431	433	224	224	361	361	134	139	103	116	0	0	0	0	264	264
Ed8	<i>E. dunnii</i>	170	170	296	300	346	371	159	159	435	437	222	224	367	371	134	134	107	107	249	261	190	190	265	270
Eg1	<i>E. grandis</i>	169	231	289	289	340	340	161	165	411	411	0	0	359	367	129	129	90	103	251	258	187	188	265	265
Eg2	<i>E. grandis</i>	169	231	289	289	340	348	161	161	411	411	221	221	367	367	129	129	95	124	249	251	186	188	265	265
Eg3	<i>E. grandis</i>	169	200	289	289	348	348	169	169	411	411	221	221	353	367	129	129	99	99	249	249	186	186	263	265
Eg4	<i>E. grandis</i>	0	0	290	290	339	340	169	174	411	411	221	221	0	0	129	129	107	111	249	251	186	188	263	264
Eg5	<i>E. grandis</i>	179	198	0	0	0	0	169	169	411	411	221	222	353	353	129	129	103	103	245	249	186	186	265	265
Eg6	<i>E. grandis</i>	169	196	290	290	340	354	169	172	411	411	221	221	353	367	129	129	107	107	249	251	186	188	265	265
Eg7	<i>E. grandis</i>	169	169	289	290	346	356	165	172	411	411	221	221	341	369	129	129	103	111	245	251	186	188	264	265
Eg8	<i>E. grandis</i>	196	198	289	289	340	346	172	174	411	411	0	0	353	353	129	129	103	103	251	251	188	188	265	265

individuo	especie	pCCR		CesA3_A		CesA3_B		LIM		HMTi		HMTe		SK		XTH2		GST		GAD_A		GAD_B		PER	
Es1	<i>E. saligna</i>	196	231	289	293	336	354	161	170	411	411	0	0	367	367	129	129	103	103	248	249	185	186	265	265
Es2	<i>E. saligna</i>	231	231	289	293	359	361	167	173	431	431	221	221	0	0	129	129	103	103	249	256	185	186	264	265
Es3	<i>E. saligna</i>	168	168	290	293	356	361	175	177	437	437	221	221	353	353	129	129	92	92	248	258	187	188	263	266
Es4	<i>E. saligna</i>	168	168	289	293	365	369	170	173	0	0	0	0	353	353	129	129	107	113	256	256	185	189	266	266
Es6	<i>E. saligna</i>	170	170	293	293	367	369	169	173	0	0	0	0	353	353	129	129	99	99	256	260	185	189	264	264
Es7	<i>E. saligna</i>	182	192	289	290	365	373	175	180	448	448	0	0	353	353	129	129	103	103	250	250	187	187	265	265
Et1	<i>E. tereticornis</i>	182	188	289	289	346	361	173	175	413	413	208	208	0	0	129	129	91	91	0	0	185	186	261	264
Et2	<i>E. tereticornis</i>	182	184	302	302	344	346	167	186	423	429	221	221	0	0	129	129	128	140	249	249	190	190	265	267
Et3	<i>E. tereticornis</i>	168	168	294	302	343	363	167	177	427	439	221	222	0	0	129	129	121	121	250	260	187	189	265	266
Et4	<i>E. tereticornis</i>	191	231	289	293	346	348	153	175	429	445	221	222	325	325	129	129	86	105	249	252	186	189	263	266
Et5	<i>E. tereticornis</i>	169	231	294	294	344	357	167	205	0	0	221	221	363	371	129	129	91	107	249	249	186	186	265	266
Et6	<i>E. tereticornis</i>	231	231	293	293	344	359	155	184	0	0	221	221	0	0	129	129	75	75	249	255	190	191	259	265
Et7	<i>E. tereticornis</i>	231	231	296	296	336	338	175	175	421	421	221	221	0	0	129	129	75	105	248	252	185	189	264	266
Et8	<i>E. tereticornis</i>	0	0	300	305	336	346	155	167	431	431	221	221	329	329	129	129	91	120	248	249	185	186	264	264
Ev1	<i>E. viminalis</i>	200	211	294	294	344	350	177	177	427	435	208	224	340	347	129	129	116	116	252	252	185	185	264	265
Ev2	<i>E. viminalis</i>	204	210	312	314	356	359	164	170	419	437	221	224	341	353	129	129	99	99	252	257	185	190	264	264
Ev3	<i>E. viminalis</i>	202	214	296	298	343	354	155	161	429	437	224	224	353	359	129	129	103	116	252	252	185	189	263	264
Ev4	<i>E. viminalis</i>	167	206	0	0	350	359	155	170	398	398	224	224	347	359	129	129	128	128	252	252	185	185	263	265
Ev5	<i>E. viminalis</i>	169	211	274	277	363	363	170	192	427	439	224	224	340	353	129	129	116	120	252	252	185	185	264	265
Ev6	<i>E. viminalis</i>	169	202	300	310	356	358	164	164	441	454	224	224	353	364	129	129	111	117	252	252	185	189	0	0
Ev7	<i>E. viminalis</i>	169	184	290	293	356	358	175	182	419	445	224	224	347	353	129	129	99	103	252	257	185	186	263	265
Ev8	<i>E. viminalis</i>	198	204	294	294	358	359	170	175	0	0	224	224	347	347	0	0	116	118	0	0	185	185	0	0
EgL1	<i>E. globulus</i>	182	182	290	290	346	359	161	169	419	435	221	221	351	355	129	129	99	111	256	256	189	189	266	267
EgL2	<i>E. globulus</i>	182	182	290	296	348	365	169	169	417	435	221	224	351	359	129	134	99	99	248	256	185	185	268	269
EgL3	<i>E. globulus</i>	208	210	294	306	348	356	159	161	417	435	221	221	349	355	129	129	83	111	248	248	185	185	265	265
EgL4	<i>E. globulus</i>	182	182	294	294	346	348	159	159	425	435	221	221	332	332	129	134	111	111	248	248	185	189	266	268
EgL5	<i>E. globulus</i>	182	182	288	290	356	359	159	169	418	435	221	221	355	355	129	129	92	116	248	248	185	189	264	266
EgL6	<i>E. globulus</i>	182	182	0	0	356	365	159	169	417	439	221	221	351	353	129	129	96	99	248	248	185	185	267	267
EgL7	<i>E. globulus</i>	182	0	294	296	348	348	175	177	419	437	221	221	351	351	122	129	96	96	248	248	185	185	268	268
EgL8	<i>E. globulus</i>	206	216	290	290	356	358	159	161	417	417	221	221	331	349	129	134	96	111	256	256	185	185	267	267

Continuación

individuo	especie	EST1		EST3		EST5		EST11		EST33		EST37		EST46		EST58		EST88		EST 125		EST134		EST 141	
Ec2	<i>E. camaldulensis</i>	141	149	243	243	311	311	0	0	142	144	233	236	214	214	137	139	230	230	210	210	308	336	0	0
Ec3	<i>E. camaldulensis</i>	134	138	243	243	311	311	0	0	144	144	220	236	214	214	137	137	228	230	210	210	308	313	295	300
Ec4	<i>E. camaldulensis</i>	0	0	243	248	311	311	0	0	142	145	236	237	214	214	137	137	230	230	210	220	308	325	293	293
Ec5	<i>E. camaldulensis</i>	134	134	0	0	311	311	0	0	134	144	236	236	214	214	137	137	230	230	203	220	325	330	293	300
Ec6	<i>E. camaldulensis</i>	150	150	243	243	311	311	0	0	144	145	236	236	214	214	137	142	228	230	210	210	313	315	293	300
Ec7	<i>E. camaldulensis</i>	132	144	243	243	311	311	0	0	144	144	219	236	214	214	137	137	230	230	210	210	313	325	300	300
Ec8	<i>E. camaldulensis</i>	134	134	0	0	311	311	0	0	144	144	236	236	211	214	137	137	0	0	210	210	311	336	295	295
Ed1	<i>E. dunnii</i>	0	0	243	243	311	311	293	293	144	144	236	236	217	217	142	142	230	230	207	210	323	323	293	293
Ed2	<i>E. dunnii</i>	134	144	243	243	311	311	293	298	144	144	236	236	217	217	141	142	230	230	203	203	323	323	293	298
Ed3	<i>E. dunnii</i>	0	0	247	250	311	311	293	302	136	144	236	236	217	217	142	148	230	230	203	203	323	323	293	302
Ed4	<i>E. dunnii</i>	0	0	0	0	313	319	293	302	136	144	236	236	217	217	141	142	230	230	203	203	323	323	293	302
Ed5	<i>E. dunnii</i>	133	134	0	0	311	311	298	307	144	144	236	236	217	217	131	141	228	230	203	210	322	330	298	308
Ed6	<i>E. dunnii</i>	133	134	247	250	311	311	298	298	136	144	236	236	217	217	142	148	230	230	203	210	313	323	298	298
Ed7	<i>E. dunnii</i>	144	144	247	250	0	0	0	0	144	144	236	236	217	217	142	142	0	0	0	0	323	323	0	0
Ed8	<i>E. dunnii</i>	0	0	0	0	0	0	0	0	144	144	0	0	217	217	0	0	0	0	203	203	322	345	0	0
Eg1	<i>E. grandis</i>	0	0	247	248	306	306	0	0	136	144	236	236	210	210	142	142	230	230	203	207	325	325	0	0
Eg2	<i>E. grandis</i>	134	142	0	0	308	311	293	293	144	147	236	239	210	210	142	142	230	230	205	207	311	325	293	293
Eg3	<i>E. grandis</i>	141	141	246	246	308	311	293	293	144	144	236	236	210	214	142	142	230	230	203	207	318	323	293	293
Eg4	<i>E. grandis</i>	141	141	0	0	311	311	293	293	144	144	236	236	210	210	142	142	230	230	0	0	315	315	293	293
Eg5	<i>E. grandis</i>	134	134	247	247	306	306	293	293	136	144	236	236	210	214	142	142	230	230	0	0	325	325	293	293
Eg6	<i>E. grandis</i>	134	141	245	247	306	306	289	293	136	144	236	236	214	214	142	142	230	230	207	207	315	325	289	293
Eg7	<i>E. grandis</i>	134	141	0	0	0	0	293	293	144	144	236	236	210	210	142	142	230	230	207	207	323	323	293	293
Eg8	<i>E. grandis</i>	134	134	0	0	0	0	0	0	144	144	236	236	210	210	0	0	0	0	203	203	311	323	0	0
Es1	<i>E. saligna</i>	141	141	0	0	0	0	0	0	144	144	236	236	210	314	126	126	230	230	207	210	325	330	0	0
Es2	<i>E. saligna</i>	0	0	237	248	308	311	0	0	136	144	236	236	210	210	0	0	230	230	207	207	322	345	300	300
Es3	<i>E. saligna</i>	134	134	244	246	308	308	0	0	144	144	236	236	210	210	142	142	228	230	200	207	325	340	300	300
Es4	<i>E. saligna</i>	0	0	246	246	308	308	0	0	144	144	236	236	210	210	0	0	230	230	200	207	322	340	295	300
Es6	<i>E. saligna</i>	134	138	0	0	308	308	0	0	144	144	236	236	210	210	142	142	0	0	200	207	325	330	300	300
Es7	<i>E. saligna</i>	134	138	240	246	308	311	0	0	144	144	236	236	210	210	126	142	230	230	200	207	325	330	300	300

individuo	especie	EST1		EST3		EST5		EST11		EST33		EST37		EST46		EST58		EST88		EST 125		EST134		EST 141	
Et1	<i>E. tereticornis</i>	136	138	244	246	311	311	0	0	144	144	219	236	214	214	142	142	230	230	203	230	313	345	0	0
Et2	<i>E. tereticornis</i>	134	143	235	250	311	311	0	0	136	144	236	236	214	214	137	142	230	230	210	210	323	323	295	300
Et3	<i>E. tereticornis</i>	148	148	244	248	311	311	0	0	144	147	236	239	0	0	142	142	230	230	200	210	308	325	293	293
Et4	<i>E. tereticornis</i>	134	134	243	250	311	311	0	0	144	145	236	236	214	214	137	142	230	230	203	207	313	330	293	293
Et5	<i>E. tereticornis</i>	138	138	0	0	311	311	0	0	136	147	236	239	214	214	142	142	230	230	210	216	313	323	293	293
Et6	<i>E. tereticornis</i>	134	134	244	248	0	0	0	0	144	144	236	236	214	214	137	137	228	228	210	216	313	325	300	300
Et7	<i>E. tereticornis</i>	0	0	0	0	311	311	0	0	145	145	236	236	214	214	137	142	230	230	210	216	313	313	295	295
Et8	<i>E. tereticornis</i>	134	134	244	248	0	0	0	0	144	144	236	236	0	0	142	142	0	0	0	0	311	323	0	0
Ev1	<i>E. viminalis</i>	0	0	243	243	311	311	0	0	142	144	236	239	217	217	142	142	230	230	0	0	0	0	0	0
Ev2	<i>E. viminalis</i>	0	0	243	243	311	311	298	302	143	144	236	236	217	217	142	148	230	230	0	0	323	323	298	302
Ev3	<i>E. viminalis</i>	134	134	243	243	311	311	0	0	142	144	236	236	217	217	141	142	230	230	0	0	323	323	0	0
Ev4	<i>E. viminalis</i>	134	134	243	243	311	311	298	302	144	144	239	239	217	217	142	142	230	230	205	205	323	325	298	302
Ev5	<i>E. viminalis</i>	134	134	243	243	311	311	302	302	136	144	236	236	217	217	141	142	230	230	205	205	311	323	302	302
Ev6	<i>E. viminalis</i>	134	134	243	243	311	311	295	295	136	144	236	236	217	217	142	148	230	230	203	203	323	323	295	295
Ev7	<i>E. viminalis</i>	0	0	243	243	306	311	302	302	144	144	236	236	217	217	142	148	230	230	0	0	323	323	302	302
Ev8	<i>E. viminalis</i>	0	0	243	243	0	0	0	0	145	145	236	236	217	217	148	148	0	0	0	0	323	323	0	0
EgL1	<i>E. globulus</i>	168	168	243	246	333	333	245	245	144	144	0	0	212	212	0	0	230	230	203	203	316	316	305	305
EgL2	<i>E. globulus</i>	132	132	251	252	333	333	249	249	144	144	0	0	211	211	137	137	230	230	207	207	318	318	305	308
EgL3	<i>E. globulus</i>	141	141	0	0	333	333	247	249	144	144	0	0	212	212	131	137	230	230	203	207	316	316	305	305
EgL4	<i>E. globulus</i>	132	141	251	252	333	333	245	247	144	144	0	0	211	211	142	142	230	230	207	207	322	322	0	0
EgL5	<i>E. globulus</i>	141	141	243	243	333	333	245	245	144	144	0	0	211	211	131	137	230	230	205	207	313	318	308	308
EgL6	<i>E. globulus</i>	132	141	245	246	333	333	245	245	144	144	0	0	210	211	137	137	230	230	203	203	313	316	305	305
EgL7	<i>E. globulus</i>	132	133	246	246	333	333	245	247	144	144	0	0	210	211	131	131	230	230	203	203	313	316	305	305
EgL8	<i>E. globulus</i>	132	133	246	246	333	339	249	249	144	144	0	0	210	211	137	137	230	230	207	207	316	316	305	0

individuo	especie	EST191		EST197		EST200		EST216	
Ec2	<i>E. camaldulensis</i>	219	219	85	85	0	0	395	406
Ec3	<i>E. camaldulensis</i>	205	205	96	96	341	353	396	402
Ec4	<i>E. camaldulensis</i>	208	211	85	96	0	0	400	404
Ec5	<i>E. camaldulensis</i>	0	0	85	96	0	0	0	0
Ec6	<i>E. camaldulensis</i>	211	211	85	95	334	334	0	0
Ec7	<i>E. camaldulensis</i>	0	0	96	104	347	347	396	402
Ec8	<i>E. camaldulensis</i>	205	208	85	88	336	336	0	0
Ed1	<i>E. dunnii</i>	211	219	96	100	332	350	0	0
Ed2	<i>E. dunnii</i>	205	205	96	96	0	0	0	0
Ed3	<i>E. dunnii</i>	205	205	96	96	334	346	396	402
Ed4	<i>E. dunnii</i>	206	211	96	96	346	352	399	402
Ed5	<i>E. dunnii</i>	207	211	88	96	336	350	400	408
Ed6	<i>E. dunnii</i>	211	219	82	88	0	0	0	0
Ed7	<i>E. dunnii</i>	211	211	0	0	0	0	396	398
Ed8	<i>E. dunnii</i>	205	205	96	100	334	334	404	407
Eg1	<i>E. grandis</i>	198	211	96	97	332	332	0	0
Eg2	<i>E. grandis</i>	207	215	96	98	0	0	0	0
Eg3	<i>E. grandis</i>	207	207	96	96	334	334	404	407
Eg4	<i>E. grandis</i>	198	214	96	96	0	0	0	0
Eg5	<i>E. grandis</i>	198	214	96	97	332	340	406	408
Eg6	<i>E. grandis</i>	207	207	96	97	334	334	0	0
Eg7	<i>E. grandis</i>	211	211	96	96	332	344	402	402
Eg8	<i>E. grandis</i>	207	211	96	98	0	0	404	404
Es1	<i>E. saligna</i>	0	0	96	96	334	334	395	401
Es2	<i>E. saligna</i>	205	205	85	85	332	332	393	412
Es3	<i>E. saligna</i>	208	208	97	97	332	340	404	404
Es4	<i>E. saligna</i>	0	0	92	97	0	0	404	408
Es6	<i>E. saligna</i>	211	211	95	97	0	0	396	408
Es7	<i>E. saligna</i>	0	0	92	97	0	0	399	404
Et1	<i>E. tereticornis</i>	215	215	0	0	336	336	399	399
Et2	<i>E. tereticornis</i>	208	208	92	96	0	0	399	399
Et3	<i>E. tereticornis</i>	208	208	96	98	332	336	0	0

individuo	especie	EST191		EST197		EST200		EST216	
Et4	<i>E. tereticornis</i>	207	211	93	98	334	340	0	0
Et5	<i>E. tereticornis</i>	209	221	92	96	338	354	405	407
Et6	<i>E. tereticornis</i>	207	207	93	96	0	0	0	0
Et7	<i>E. tereticornis</i>	207	207	92	92	334	350	399	410
Et8	<i>E. tereticornis</i>	207	207	85	85	0	0	399	399
Ev1	<i>E. viminalis</i>	209	225	0	0	338	344	394	394
Ev2	<i>E. viminalis</i>	206	211	0	0	332	334	403	405
Ev3	<i>E. viminalis</i>	225	225	0	0	332	332	405	406
Ev4	<i>E. viminalis</i>	211	211	0	0	334	338	406	406
Ev5	<i>E. viminalis</i>	205	221	0	0	0	0	405	405
Ev6	<i>E. viminalis</i>	206	215	0	0	0	0	0	0
Ev7	<i>E. viminalis</i>	219	219	0	0	334	334	403	405
Ev8	<i>E. viminalis</i>	211	215	0	0	0	0	0	0
EgL1	<i>E. globulus</i>	211	214	96	96	332	332	407	409
EgL2	<i>E. globulus</i>	211	211	95	95	332	332	400	404
EgL3	<i>E. globulus</i>	211	219	96	105	332	332	393	392
EgL4	<i>E. globulus</i>	214	219	96	96	332	332	397	404
EgL5	<i>E. globulus</i>	0	0	95	96	332	332	400	404
EgL6	<i>E. globulus</i>	211	214	95	96	332	332	404	404
EgL7	<i>E. globulus</i>	214	214	95	104	332	332	0	0
EgL8	<i>E. globulus</i>	211	219	0	0	332	332	396	404

TABLA ANEXA 7

Matriz de pertenencia de cada individuo con respecto a cada número de poblaciones (k) tomado como ideal

a) Análisis realizado con 60 individuos de *E. globulus* (k=4)

Individuo	Pertenencia k=4			
218 EO	0,092	0,045	0,034	0,829
221 EO	0,039	0,091	0,031	0,838
225 EO	0,027	0,033	0,083	0,857
237 EO	0,411	0,059	0,054	0,476
279 EO	0,022	0,081	0,046	0,851
257 Fur	0,026	0,92	0,027	0,027
266 Fur	0,046	0,916	0,019	0,019
297 Fur	0,02	0,887	0,032	0,061
302 Fur	0,042	0,913	0,016	0,03
303 Fur	0,018	0,887	0,077	0,018
305 Fur	0,032	0,899	0,025	0,044
313 Fur	0,044	0,883	0,012	0,061
314 Fur	0,083	0,853	0,015	0,048
241 KI	0,019	0,883	0,033	0,065
245 KI	0,037	0,061	0,035	0,866
255 KI	0,061	0,565	0,035	0,339
280 KI	0,106	0,131	0,029	0,734
250 NET	0,934	0,029	0,011	0,027
256 NET	0,009	0,972	0,01	0,009
259 NET	0,754	0,052	0,071	0,123
224 Port	0,928	0,019	0,014	0,039
295 Port	0,234	0,306	0,031	0,43
203 SET	0,564	0,262	0,018	0,156
229 SET	0,35	0,631	0,006	0,013
287 SET	0,945	0,015	0,019	0,021
310 SET	0,816	0,079	0,019	0,086
262 ST	0,589	0,166	0,077	0,169
264 ST	0,269	0,367	0,055	0,309
267 ST	0,553	0,389	0,031	0,027
293 ST	0,907	0,041	0,014	0,037

Individuo	Pertenencia k=4			
240 ST	0,49	0,465	0,017	0,028
181 Str	0,139	0,014	0,243	0,604
190 Str	0,29	0,022	0,039	0,648
192 Str	0,059	0,011	0,801	0,129
194 Str	0,466	0,009	0,407	0,117
201 Str	0,051	0,061	0,191	0,697
205 Str	0,021	0,009	0,947	0,023
206 Str	0,016	0,02	0,914	0,05
213 Str	0,008	0,025	0,941	0,026
214 Str	0,024	0,027	0,101	0,848
216 Str	0,039	0,013	0,913	0,035
219 Str	0,085	0,018	0,185	0,712
227 Str	0,017	0,023	0,484	0,476
265 Str	0,029	0,01	0,871	0,09
281 Str	0,056	0,011	0,893	0,04
285 Str	0,08	0,016	0,356	0,548
299 Str	0,121	0,103	0,287	0,488
301 Str	0,017	0,032	0,917	0,033
230 WO	0,043	0,016	0,017	0,924
232 WO	0,079	0,71	0,039	0,172
247 WO	0,124	0,028	0,015	0,833
274 WO	0,03	0,595	0,031	0,344
298 WO	0,116	0,086	0,513	0,285
300 WO	0,031	0,097	0,144	0,728
304 WO	0,134	0,343	0,224	0,299
306 WO	0,036	0,175	0,276	0,514
308 WO	0,026	0,949	0,012	0,014
309 WO	0,111	0,685	0,039	0,165
312 WO	0,011	0,059	0,562	0,368
315 WO	0,016	0,012	0,013	0,959

b) Análisis realizado con 53 individuos de siete especies de *Eucalyptus*, sin datos *a priori* (k=5) y con datos *a priori* (k=4 y k=7)

Individuo	Perteneencia k=5				
Ec2	0,006	0,004	0,008	0,978	0,004
Ec3	0,002	0,003	0,002	0,992	0,002
Ec4	0,003	0,012	0,002	0,981	0,003
Ec5	0,003	0,002	0,002	0,99	0,003
Ec6	0,006	0,004	0,003	0,934	0,054
Ec7	0,004	0,003	0,002	0,99	0,002
Ec8	0,011	0,008	0,008	0,971	0,002
Et1	0,003	0,004	0,003	0,985	0,005
Et2	0,001	0,01	0,008	0,978	0,002
Et3	0,002	0,003	0,062	0,58	0,353
Et4	0,003	0,005	0,015	0,914	0,064
Et5	0,001	0,1	0,074	0,814	0,01
Et6	0,001	0,002	0,002	0,992	0,003
Et7	0,002	0,002	0,004	0,99	0,002
Et8	0,001	0,006	0,007	0,982	0,004
Ed1	0,005	0,982	0,004	0,003	0,006
Ed2	0,001	0,994	0,002	0,001	0,001
Ed3	0,001	0,985	0,004	0,004	0,006
Ed4	0,002	0,986	0,005	0,004	0,004
Ed5	0,045	0,929	0,003	0,017	0,006
Ed6	0,003	0,869	0,118	0,007	0,003
Ed7	0,002	0,978	0,015	0,003	0,003
Ed8	0,004	0,984	0,002	0,003	0,007

Individuo	Perteneencia k=5				
Ev1	0,003	0,006	0,983	0,007	0,002
Ev2	0,003	0,01	0,982	0,003	0,002
Ev3	0,003	0,007	0,977	0,006	0,006
Ev4	0,001	0,002	0,993	0,002	0,002
Ev5	0,001	0,003	0,991	0,004	0,001
Ev6	0,003	0,005	0,986	0,004	0,002
Ev7	0,003	0,006	0,962	0,008	0,021
Ev8	0,002	0,003	0,991	0,003	0,002
Egr1	0,002	0,004	0,002	0,002	0,99
Egr2	0,002	0,002	0,012	0,008	0,977
Egr3	0,008	0,004	0,002	0,003	0,983
Egr4	0,003	0,002	0,002	0,002	0,991
Egr5	0,001	0,002	0,002	0,002	0,993
Egr6	0,001	0,002	0,001	0,002	0,994
Egr7	0,003	0,007	0,007	0,003	0,98
Egr8	0,002	0,002	0,002	0,002	0,992
Es1	0,003	0,002	0,003	0,039	0,953
Es2	0,007	0,014	0,002	0,522	0,455
Es3	0,007	0,002	0,003	0,017	0,972
Es4	0,01	0,003	0,003	0,006	0,978
Es6	0,013	0,02	0,005	0,016	0,946
Es7	0,004	0,004	0,002	0,02	0,97
Egl1	0,994	0,002	0,001	0,001	0,001
Egl2	0,993	0,002	0,003	0,001	0,001
Egl3	0,992	0,002	0,003	0,002	0,002
Egl4	0,993	0,003	0,001	0,001	0,002
Egl5	0,965	0,003	0,028	0,002	0,002
Egl6	0,99	0,001	0,002	0,002	0,006
Egl7	0,989	0,002	0,004	0,003	0,002
Egl8	0,99	0,002	0,003	0,003	0,002

b) K=4

Individuo	Pertenenca k=4			
Ec2	0,007	0,982	0,008	0,004
Ec3	0,002	0,994	0,002	0,002
Ec4	0,003	0,989	0,004	0,003
Ec5	0,003	0,992	0,003	0,003
Ec6	0,006	0,944	0,003	0,047
Ec7	0,003	0,993	0,002	0,002
Ec8	0,01	0,976	0,012	0,002
Et1	0,002	0,989	0,003	0,006
Et2	0,002	0,981	0,015	0,002
Et3	0,003	0,637	0,022	0,338
Et4	0,003	0,922	0,012	0,062
Et5	0,001	0,754	0,235	0,009
Et6	0,001	0,994	0,002	0,003
Et7	0,002	0,993	0,003	0,002
Et8	0,002	0,988	0,007	0,004
Ed1	0,004	0,005	0,982	0,009
Ed2	0,001	0,002	0,995	0,002
Ed3	0,002	0,011	0,974	0,014
Ed4	0,002	0,006	0,983	0,009
Ed5	0,126	0,024	0,844	0,006
Ed6	0,003	0,006	0,987	0,003
Ed7	0,003	0,005	0,989	0,004
Ed8	0,017	0,006	0,963	0,014
Ev1	0,002	0,008	0,988	0,002
Ev2	0,004	0,003	0,99	0,002
Ev3	0,004	0,006	0,985	0,005
Ev4	0,002	0,003	0,993	0,002

Individuo	Pertenenca k=4			
Ev5	0,001	0,004	0,993	0,002
Ev6	0,003	0,005	0,99	0,002
Ev7	0,002	0,007	0,977	0,014
Ev8	0,002	0,005	0,99	0,002
Eg1	0,002	0,002	0,004	0,992
Eg2	0,002	0,008	0,004	0,987
Eg3	0,008	0,002	0,004	0,986
Eg4	0,003	0,002	0,002	0,993
Eg5	0,001	0,002	0,002	0,994
Eg6	0,001	0,002	0,001	0,995
Eg7	0,003	0,003	0,009	0,985
Eg8	0,002	0,002	0,003	0,994
Es1	0,003	0,041	0,003	0,954
Es2	0,007	0,521	0,006	0,466
Es3	0,007	0,021	0,002	0,969
Es4	0,012	0,006	0,003	0,979
Es6	0,013	0,015	0,01	0,962
Es7	0,004	0,025	0,002	0,968
EgL1	0,996	0,001	0,002	0,001
EgL2	0,995	0,001	0,003	0,001
EgL3	0,994	0,002	0,002	0,002
EgL4	0,995	0,002	0,002	0,002
EgL5	0,981	0,003	0,013	0,003
EgL6	0,992	0,002	0,002	0,005
EgL7	0,991	0,003	0,003	0,002
EgL8	0,992	0,003	0,003	0,002

b) K=7

Individuos	Pertenenca k=7						
Ec2	0,003	0,002	0,002	0,004	0,964	0,023	0,002
Ec3	0,002	0,004	0,002	0,002	0,012	0,976	0,001
Ec4	0,002	0,006	0,002	0,002	0,973	0,012	0,002
Ec5	0,003	0,003	0,003	0,003	0,014	0,971	0,003
Ec6	0,005	0,006	0,004	0,003	0,096	0,81	0,075
Ec7	0,004	0,003	0,002	0,001	0,003	0,986	0,001
Ec8	0,008	0,004	0,002	0,005	0,93	0,049	0,002
Et1	0,002	0,003	0,026	0,002	0,447	0,516	0,003
Et2	0,002	0,008	0,003	0,005	0,953	0,029	0,002
Et3	0,002	0,002	0,295	0,008	0,687	0,003	0,003
Et4	0,003	0,005	0,008	0,011	0,427	0,483	0,064
Et5	0,001	0,026	0,002	0,013	0,901	0,047	0,009
Et6	0,001	0,002	0,003	0,002	0,006	0,984	0,002
Et7	0,002	0,003	0,003	0,007	0,01	0,974	0,001
Et8	0,001	0,008	0,004	0,008	0,352	0,621	0,006
Ed1	0,005	0,974	0,005	0,004	0,005	0,002	0,006
Ed2	0,001	0,991	0,001	0,003	0,002	0,001	0,002
Ed3	0,001	0,976	0,003	0,005	0,003	0,005	0,006
Ed4	0,002	0,976	0,002	0,006	0,007	0,003	0,006
Ed5	0,048	0,9	0,017	0,003	0,016	0,014	0,002
Ed6	0,003	0,836	0,003	0,139	0,009	0,006	0,004
Ed7	0,002	0,972	0,003	0,014	0,003	0,002	0,003
Ed8	0,004	0,975	0,002	0,002	0,003	0,002	0,012
Eg1	0,002	0,005	0,034	0,002	0,004	0,003	0,95
Eg2	0,002	0,002	0,004	0,014	0,024	0,003	0,952
Eg3	0,02	0,002	0,006	0,002	0,003	0,002	0,965
Eg4	0,002	0,002	0,002	0,002	0,002	0,002	0,988
Eg5	0,001	0,002	0,016	0,002	0,003	0,002	0,973
Eg6	0,001	0,001	0,002	0,001	0,002	0,002	0,991
Eg7	0,004	0,004	0,003	0,012	0,004	0,003	0,97
Eg8	0,001	0,002	0,002	0,002	0,002	0,001	0,99
Es1	0,003	0,002	0,062	0,006	0,036	0,157	0,735
Es2	0,005	0,009	0,517	0,003	0,132	0,315	0,019

Individuos	Pertenenca k=7						
Es3	0,004	0,001	0,974	0,003	0,003	0,004	0,011
Es4	0,002	0,002	0,986	0,003	0,002	0,002	0,003
Es6	0,015	0,006	0,966	0,003	0,003	0,003	0,004
Es7	0,003	0,003	0,973	0,002	0,005	0,008	0,007
Ev1	0,002	0,005	0,002	0,974	0,011	0,004	0,002
Ev2	0,003	0,01	0,003	0,975	0,006	0,002	0,002
Ev3	0,003	0,006	0,004	0,966	0,011	0,003	0,006
Ev4	0,001	0,002	0,002	0,99	0,003	0,002	0,002
Ev5	0,001	0,003	0,002	0,982	0,008	0,002	0,001
Ev6	0,002	0,005	0,002	0,979	0,006	0,003	0,002
Ev7	0,002	0,005	0,007	0,951	0,012	0,003	0,021
Ev8	0,002	0,002	0,002	0,985	0,003	0,003	0,003
EgL1	0,992	0,002	0,001	0,001	0,002	0,001	0,001
EgL2	0,991	0,002	0,001	0,003	0,001	0,001	0,001
EgL3	0,988	0,001	0,001	0,003	0,003	0,002	0,002
EgL4	0,988	0,003	0,002	0,001	0,002	0,001	0,002
EgL5	0,953	0,003	0,007	0,031	0,003	0,002	0,001
EgL6	0,984	0,001	0,006	0,002	0,002	0,001	0,004
EgL7	0,978	0,002	0,009	0,004	0,002	0,004	0,001
EgL8	0,985	0,002	0,003	0,003	0,004	0,002	0,001