Theor Appl Genet (2002) 104:379–387

K.A. Thamarus · K. Groom · J. Murrell · M. Byrne
G.F. Moran

A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre, and floral traits

Received: 30 August 2000 / Accepted: 23 March 2001

Abstract A genetic linkage map containing potential candidate loci for wood, fibre and floral traits has been constructed for *Eucalyptus globulus* (Labill.) based on the segregation of 249 codominant loci in an outbred F_1 population of 148 individuals. The map contains 204 RFLP loci, including 31 cambium-specific expressed sequence tags (ESTs) and 14 known function genes, and 40 microsatellite and five isozyme loci. Independent male and female maps were constructed, and the 98 loci (39%) that segregated in both parents were used to combine the parental maps into an integrated map. The 249 loci mapped to 11 major linkage groups (n=11 in eucalypts) and a 12th small linkage group containing three loci that segregated in the male parent only. Total map distance is 1375 cM with an average interval of 6 cM. Forty one of the mapped loci identify known proteins (five isozymes) or sequences with known function (14 genes and 22 ESTs). The mapped genes include enzymes involved in lignin and cell-wall polysaccharide biosynthesis, and floral-development genes. This map will be used to locate quantitative trait loci for wood, fibre, and other traits in *Eucalyptus*.

Keywords Linkage map · Candidate gene · EST · Microsatellite · Lignin

Communicated by M.A. Saghai Maroof

K.A. Thamarus (✉) · K. Groom · J. Murrell · G.F. Moran
CSIRO Forestry and Forest Products and CRC
for Sustainable Production Forestry, PO Box E4008, Kingston,
ACT 2604, Australia
e-mail: Karen.Thamarus@ffp.csiro.au
Fax: +61-2-6281-8233

M. Byrne
CSIRO Forestry and Forest Products, PO Box E4008, Kingston,
ACT 2604, Australia

Present address:
M. Byrne, CALMScience, WA Herbarium, Locked Bag 104,
Bentley Delivery Centre, WA 6983, Australia

Introduction

Linkage maps are useful tools for breeders who wish to employ marker-aided selection (MAS) in breeding programs (Staub et al. 1996). In tree breeding, most traits of commercial interest are quantitative, and juvenile-mature trait correlations are poor. Trait evaluation may take several years; consequently, the potential benefits of early selection using molecular markers are very attractive. Numerous linkage maps for commercial tree species have been developed to locate quantitative trait loci (QTLs) and provide a basis for MAS (e.g. in *Pinus*, Groover et al. 1994; in *Populus*, Bradshaw and Stettler 1995; and in *Eucalyptus*, Grattapaglia et al. 1996; Byrne et al. 1997).

Genetic linkage maps contain various types of markers, including morphological, biochemical and DNA-based markers such as restriction fragment length polymorphisms (RFLPs), microsatellites or simple sequence repeats (SSRs), random amplified polymorphic DNA (RAPDs) and amplified fragment length polymorphisms (AFLPs). These DNA-based markers vary in the amount of DNA required, the costs of development and assay, the amount of genetic information revealed and transferability across taxa (Glaubitz and Moran 2000). RFLPs and microsatellites are codominant markers and more-informative than the generally dominant RAPD and AFLP markers when mapping outbred organisms (Maliepaard et al. 1997). In addition, RFLPs are readily shared between individuals and taxa, and maps constructed with RFLP markers may be integrated into generic maps for taxa. For example, a common set of RFLP markers was used to establish synteny between linkage maps from *Pinus radiata* and *Pinus taeda* (Devey et al. 1999).

In eucalypts, a genetic linkage map for *Eucalyptus nitens* has been constructed using RFLP and RAPD markers in an outbred F_2 pedigree (Byrne et al. 1995). Using the pseudo-testcross strategy in interspecific F_1 pedigrees, individual tree genetic-linkage maps have been constructed in *Eucalyptus grandis* and *Eucalyptus urophylla* based on RAPD markers (Grattapaglia and

Sederoff 1994; Verhaegen and Plomion 1996), and in *Eucalyptus globulus* and *Eucalyptus tereticornis* based on AFLP markers (Marques et al. 1998).

The candidate-gene approach to genetic mapping and QTL analysis utilises the expanding database of available plant genes. Known loci that could potentially give rise to the phenotype associated with an unknown QTL locus are candidate loci. Potential candidate loci may be identified empirically from known genes that map to the same locations as mapped QTLs, or *a priori* through knowledge of the regulatory or biochemical pathways associated with trait expression. Elements of these two strategies can be combined to enhance the potential to identify candidate genes (Sewell and Neale 2000). Known genes that may affect quantitative trait variation or ESTs isolated from tissues that express the trait are mapped and candidate gene-QTL associations are identified. For example, acetyl-CoA carboxylase, which catalyses the first step in fatty acid synthesis, was mapped in oat and has been associated with a major QTL for oil content (Kianian et al. 1999). This association is found in different oat inbred lines and across different environments. A QTL with a moderate effect on oil content is associated with the same gene in maize (Berke and Rocheford 1995). Methods for mapping known genes include sequencing mapped cDNAs and determining similarity to known genes (Neale et al. 1994; Jermstad et al. 1998), using isolated genes as probes for RFLP analysis (Tanksley et al. 1992; Li et al. 1999), homology cloning using primers developed by sequence alignment (Gentzbittel et al. 1998), and development and mapping of ESTs (Botella et al. 1997). Identification of a candidate gene-QTL association may circumvent molecular cloning of the QTL (Faris et al. 1999). In addition, having known genes on a linkage map will be useful for comparative mapping from *Eucalyptus* across to model plant species.

The purpose of this study is to map potential candidate genes and cambium-specific ESTs as part of a *E. globulus* QTL study in an effort to understand the molecular basis for quantitative trait variation in wood, fibre, and floral traits. In this paper, we report the construction of a genetic linkage map for *E. globulus* based on codominant markers, the linkage-group locations of lignin and cell-wall biosynthesis genes, floral development genes and ESTs isolated from wood-forming tissue, and a comparison of this map with other *Eucalyptus* genetic maps.

Materials and methods

Plant material

Mapping analyses were conducted on an outbred F_1 population of 148 full-sib progeny, the result of an inter-population cross of King Island (female parent) and Taranna (male parent) individuals. This cross is part of the North Forest Products/CSIRO Division of Forestry hybrid trials (Volker 1995) that were planted in 1990. Subsets of the 148 progeny were planted at seven sites throughout Australia spanning the natural and planted range of *E. globulus*: with four sites in Tasmania and one site each in Victoria, New South Wales and Western Australia.

Sources of RFLP probes

The genomic DNA (gDNA) and cDNA probes used in this study were from a *E. nitens* genomic library and a *E. globulus* cDNA library (Byrne et al. 1994). Genomic and cDNA loci are identified by 'g' and 'c', respectively, followed by an identifying number. The ESTs were isolated from a *E. globulus* cambium-specific subtracted cDNA library (Bossinger and Leitch 2000), and are identified by 'e', followed by the last three numerals of the GenBank accession number (Table 1). The floral genes were isolated from *E. globulus* and are eucalypt homologues of *Arabidopsis* floral genes: two *AGAMOUS* homologues (AGE1, AGE2, Harcourt et al. 1995), an *APETALA 1* (EAP1, Kyozuka et al. 1997), and a *LEAFY* homolog (ELF1, Southerton et al. 1998). The cell-wall genes were isolated from *E. globulus* and are homologous to cellulose synthase (ECS1), cellulase (ECA1), and xylan synthase (EXS1 and EXS2) (S. Southerton, personal communication). S-adenosyl homocysteine hydrolase (MsaS2) was isolated from *Medicago sativa* and is thought to have an associated role in lignification (Abrahams et al. 1995).

Five more RFLP probes, with sequence homology to genes of the monolignol biosynthetic pathway (Whetten et al. 1998), were developed from *E. globulus* genomic DNA for this study and showed segregation in the pedigree: phenylalanine ammonia-lyase (PAL), caffeoyl Coenzyme A O-methyltransferase (CCoAOMT), 4-coumarate Coenzyme A ligase (4CL), caffeate O-methyltransferase (COMT) and cinnamoyl Coenzyme A reductase (CCR). Cinnamoyl alcohol dehydrogenase (CAD) was also isolated from *E. globulus* DNA but no variation was detected in the parents when DNA was digested with 12 different enzymes.

Eucalypt lignin gene-probe development

Lignin gene-specific primers were designed from sequences available in public databases accessed through the NCBI web server (http://www.ncbi.nlm.nih.gob/BLAST/). For each lignin gene-primer pair, except 4CL, the base sequence for primer design was from a eucalypt species (Table 2). Genomic DNA and cDNA sequences were used in each sequence alignment to distinguish intron splice sites so that primers were not placed near these regions. Conserved regions of aligned DNA or protein sequences were identified using MACAW (Schuler et al. 1991). PCR primers were designed using PRIMER (Version 0.5, Lincoln et al. 1991). Primer pairs were selected to span at least one intron and were designed with the following selection criteria: length 18–22 nucleotides, 60°C melting temperature, minimum self-complementarity and a 3′ GC clamp. Polymerase chain reaction (PCR) assays were carried out in a final volume of 25 µl in 10 mM Tris HCl (pH 8.4), 50 mM KCl, 0.2 mM of each of the four dNTPs, 1.5 mM $MgCl_2$, 0.16 µM of both forward and reverse primers, 0.2 mg/ml of bovine serum albumin, 20–30 ng of genomic DNA, and 1 unit of *Taq* polymerase (GibcoBRL). The concentration of the forward and reverse primers for 4CL amplification was increased to 0.2 µM. PCR was carried out for 30 cycles of 95°C for 15 s, 55°C for 30 s, and 72°C for 1 min. PCR-generated fragments of *E. globulus* DNA were size-checked on 1.5% agarose gel, purified with a QIAquick PCR Purification kit (Qiagen Pty Ltd, Victoria, Australia), and cloned into plasmids using the pGEM-T Easy Vector System (Promega Corp, Madison, Wis., USA). The inserts were sequenced using the ABI Prism® Big DyeTerminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems), and a BLAST 2.0 search (Altschul et al. 1997) was conducted on either the nucleotide or the predicted amino-acid sequence to confirm fragment-sequence homology.

RFLP procedures

Genomic DNA (4 µg) was extracted from leaf tissue, digested with *Eco*RI, *Eco*RV, *Dra*I, *Bgl*II, *Hin*dIII *Bcl*I, or *Hpa*II restriction enzymes, and hybridised to filters as described in Byrne et al. (1993), except that filters were capillary blotted in 0.4 M NaOH

Table 1 Candidate gene and cambium-specific EST loci by linkage group, with sequence homology and GenBank accession number or reference

Locus[a]	Linkage group	Marker type	Sequence homology[b]	GenBank accession[c] or reference
e345c	1	EST	New[d]	AW191345
e301	1	EST	Hydroxymethyltransferase	AW191301
e331	1	EST	New	AW191331
e377	1	EST	GH1 protein; auxin inducible (IAA16; Ath)	AW191377
e345d	1	EST	New	AW191345
e343	1	EST	Transmembrane protein	AW191343
e371	2	EST	Serine carboxypeptidase II-3 precursor	AW191371
e378	2	EST	New	AW191378
e340	2	EST	New	AW191340
e360a	2	EST	Xyloglucan endotransferase (XET)	AW191360
e364	2	EST	Beta-galactosidase (Ath)	AW191364
CCoAOMT	2	Gene	Caffeoyl Coenzyme A O-methyltransferase	This paper, Table 2
ECA1	2	Gene	Cellulase homolog	S. Southerton, unpublished
4CL	3	Gene	4-coumarate:Coenzyme A ligase	This paper, Table 2
e370b	3	EST	S-adenosyl-L-methionine synthetase (SAMS)	AW191370
e329	3	EST	(1–4)-beta-mannan endohydrolase	AW191329
AGE1	3	Gene	Agamous homolog 1	Harcourt et al. 1995
ELF1	3	Gene	Leafy homolog	Southerton et al. 1998
EAP1	4	Gene	Squamousa homolog	Kyozuka et al. 1997
e330b	4	EST	New	AW191330
PAL	4	Gene	phenylalanine ammonia-lyase	This paper, Table 2
e344b	4	EST	P-glycoprotein	AW191344
e314	5	EST	Calmodulin, calcium dependent protein kinase	AW191314
e344a	5	EST	P-glycoprotein	AW191344
EXS2	5	Gene	Xylan synthase homolog 2	S. Southerton, unpublished
e376	6	EST	Subtilisin like protease (SBT1)	AW191376
EXS1	6	Gene	Xylan synthase homolog 1	S. Southerton, unpublished
e319	6	EST	Cytochrome c oxidase precursor	AW191319
e346	6	EST	1,3-glucanase (Glc1)	AW191346
e348	6	EST	GTP-binding protein	AW191348
e310	7	EST	Laccase	AW191310
e330a	7	EST	New	AW191330
COMT	7	Gene	Caffeate O-methyltransferase	This paper, Table 2
e370a	7	EST	S-adenosyl-L-methionine synthetase (SAMS)	AW191370
e365	7	EST	Methionine synthase	AW191365
e345a	8	EST	New	AW191345
ECS1	8	Gene	Cellulose synthase homolog	S. Southerton, unpublished
e353	8	EST	S. cerevisiae EMP70 protein precursor (Homo)	AW191353
e351	8	EST	Calcineurin B-like protein	AW191351
AGE2	9	Gene	Agamous homolog 2	Harcourt et al. 1995
e345b	9	EST	New	AW191345
CCR	10	Gene	Cinnamoyl Coenzyme A reductase	This paper, Table 2
e360b	10	EST	Xyloglucan endotransferase (XET)	AW191360
e358	10	EST	Elongation factor-1 alpha (Tef S1) gene	AW191358
MsaS2	11	Gene	S-adenosyl-homocysteine hydrolase	Abrahams et al. 1995

[a] EST loci are designated by lower case "e" followed by the last three digits of the GenBank accession number
[b] EST loci sequence homology provided by G. Bossinger. Other locus sequence homology identified by referenced source
[c] EST loci GenBank accession number provided by G. Bossinger
[d] New sequences show no significant homology to database sequences

overnight rather than vacuum-blotted. RFLP probes were prepared by the PCR amplification of inserts and purified using QIAquick spin columns. RFLP procedures followed Byrne et al. (1994) with the exception of the heterologous S-adenosyl homocysteine hydrolase probe from *M. sativa*, which was hybridised at 55°C. The parents of the pedigree were screened to identify polymorphic loci, and the progeny were screened with 55 cDNA, 88 gDNA, 24 EST and 14 gene probes.

Microsatellite procedures

The progeny were assayed at eight microsatellite loci from *E. nitens* (Byrne et al. 1996, http://www.ffp.csiro.au/tigr/molecular/eucmsps.html), three loci from *Eucalyptus sieberi* (Glaubitz et al. 1999, 2001), 26 loci of *E. globulus* (http://www.ffp.csiro.au/tigr/molecular/eucmsps.html) and three loci from *E. grandis* (Brondani et al. 1998). Microsatellite loci are designated by a species identifier (*En*, *Es*, *Eg* and *Embra*), followed by a unique number. Seven microsatellite loci were detected by polyacrylamide gel-electrophoresis and 33 were detected by ABI fragment length analysis.

Isozyme procedures

The mapping parents were assessed for variation in 14 enzyme systems following the procedures in Byrne et al. (1995). Segrega-

Table 2 Eucalypt lignin gene-primer sequences and source-database sequences used in multiple alignment and primer design

Lignin gene[a]	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	Species, sequence source, and GenBank Accession number
PAL	CACATTTTGGATGGAAGCG	TCGTACTGGACTCGAACGC	*Eucalyptus globulus*, mRNA, AF167487 *Populus kitakamiensis*, gene 1, D30657 *Populus kitakamiensis*, gene 2, D43802
CCoAOMT	GAGGAGAGCCAGACCCAAG	CAAGAATAGGCAAAGCAGGG	*Eucalyptus globulus*, mRNA, AF046122 *Eucalyptus* gunnii, mRNA, Y12228 *Populus trichocarpa*, gene 1, AJ223621 *Populus trichocarpa*, gene 2, AJ223620 *Petroselinum crispum*, gene, Z54183
CCR	CTCGAGCGAGGCTACACC	CCAAGACAAGGACGTGGG	*Eucalyptus gunnii*, mRNA, X79566 *Eucalyptus gunnii*, gene, X97433 *Populus trichocarpa*, gene, AJ224986
COMT	GACAAAATCCTCATGGAAAGC	AGTGCCACCTCCGACATC	*Eucalyptus gunnii*, mRNA, X74814 *Populus tremuloides*, gene 1, U13171 *Populus kitakamiensis*, gene 1, D49710
4CL[b]	GGIATGACIGARGCIGGICC	AAIGCIACIGGIACYTCICC	*Populus balsamifera×P. deltoides*, mRNA 1, AF008184 *Populus balsamifera×P. deltoides*, mRNA 2, AF008183 *Petroselinum crispum*, mRNA 2, X13325 *Petroselinum crispum*, mRNA 1, X13324 *Populus tremuloides*, mRNA 1, AF041049 *Populus tremuloides*, mRNA 2, AF041050 *Solanum tuberosum*, gene, M62755 *Pinus taeda*, gene 2, U39404 *Pinus taeda*, gene 1, U39405 *Oryza sativa*, gene, X52623

[a] See text for enzyme name
[b] Degenerate primers: I inosine, R purine, Y pyrimidine

tion of the progeny was assessed for the five enzyme systems that were polymorphic: malate dehydrogenase (MDH, EC 1.1.1.37), uridine diphosphogluconoic pyrophosphatase (UGP, EC 2.7.7.9), glucose phosphate isomerase (GPI, EC 5.3.1.9), phosphogluconate dehydrogenase (PGD, EC 1.1.1.44), and aspartate aminotransferase (AAT, EC 2.6.1.1).

Linkage analysis

All loci were scored independently by at least two people, and the segregation of loci was tested for goodness of fit to expected Mendelian segregation ratios using the chi-square test. The type of segregation was determined for each locus, and loci were separated into male and female segregation datasets as described in Byrne et al. (1995). Linkage analysis was conducted on male and female datasets independently using Joinmap 2.0, which is able to integrate the segregation data of various types and recombination estimates from various sources into a single map (Stam 1993; Stam and van Ooijen 1995). Male and female datasets were then combined and analysed to produce the integrated map. Parental mapping analyses were used to identify genotyping errors, and check linkage groups and marker order in the integrated map. Linkage groups were assigned with a minimum LOD threshold of 4.0 (female dataset) or 4.5 (male and combined datasets). Loci that were completely linked were identified and the less informative locus (or loci) was removed from the combined dataset prior to determining the marker order within groups. These loci were later included in the combined map at the position of the mapped loci to which they were linked.

Within linkage groups, marker order was determined with the following JOINMAP 2.0 mapping module parameters: a minimum LOD threshold of 0.5, a maximum recombination threshold of 0.45 and a chi-square jump threshold of 3. These parameter settings are within ranges recommended by the JOINMAP 2.0 manual. JOINMAP uses a least-squares approach to sequentially build a map, evaluating markers from pairwise data (Stam 1993). The minimum LOD threshold, in combination with the maximum recombination threshold, restricts the inclusion of markers into calculations for marker-order "goodness of fit." Minimum LOD values of 0.01 are relaxed whereas values of 2.0 are "solid." These LOD thresholds are not directly comparable to overall marker-order LOD values from multilocus map programs such as Mapmaker (Lander et al. 1987). The chi-square jump threshold limits the allowable increase in marker-order "goodness of fit" after the addition of a particular marker. A low jump threshold (e.g. 3) identifies "troublesome" markers with greater stringency, whereas a higher chi-square jump threshold (e.g. 5) allows a marker that causes a larger jump in the marker-order "goodness of fit" to be automatically included in the map. Kosambi's mapping function (Kosambi 1944) was used to determine the distance between markers. Marker order was checked by comparing the combined map with the male and female maps. Marker order was also evaluated by determining if recombination rates among common and male or female segregation markers were roughly additive. If marker order was consistent among maps and the combined map was constructed without relaxing the above parameters, then the final order was accepted. Loci that caused chi-square jumps greater than 3 were checked for genotyping errors and corrected or removed from analysis. The "fixed order" option was used to help reduce chi-square jumps in linkage groups 2, 4, 6, 8, 9 and 11. Loci segregating in both parents were used to establish a "fixed order." The order within linkage groups was accepted if the mean chi-square for all loci was less than 1.

Table 3 Number of loci detected by type of marker and segregation

Marker type / Segregation type	RFLP				Microsatellite (39)[a]	Isozyme (5)[b]	Total (%)
	cDNA (55)[c]	gDNA (88)[c]	EST (24)[a]	Gene (14)[a]			
Male – 2 alleles	31	34	11	6	9	2	93 (38)
Female – 2 alleles	17	19	7	4	7	2	56 (22)
Both – 2 alleles	1		1				2 (1)
Both – 3 or 4 alleles	16	41	12	4	24	1	98 (39)
Total loci	65	94	31	14	40	5	249

[a] Number of primer pairs used [b] Number of enzyme systems used [c] Number of probes used

Results

Inheritance of markers

All loci placed on the *E. globulus* map are codominant. One hundred and eighty one RFLP probes (55 cDNAs, 88 gDNAs, 24 ESTs, 14 genes), 39 microsatellite PCR primer pairs and five isozymes were screened, and 249 loci were analysed (Table 3). Only two loci (c238 and c069c) departed from expected Mendelian segregation ratios at a chi-square probability value of 0.01. These two loci were not excluded from the analysis because for 249 loci at α=0.01, two loci would be expected to be distorted by chance. The number of loci detected by each marker type and each segregation type are summarized in Table 3. For each marker type, departure from the distribution of all loci into the various segregation types (totals column) was tested using a chi square test for independence with α=0.05. There were significant differences between the distribution of all loci into segregation types and the distributions of microsatellite (p<0.0001) and cDNA (p <0.01) loci. Microsatellite primers detected the greatest number of fully informative loci (60%), and almost half of the cDNA loci (48%) segregated in the male parent only. The distributions of gDNA, gene and EST loci into segregation types were not significantly different from the overall distribution. The distribution of isozyme loci was not tested due to the small sample size. In all, 39% of the loci detected three or more alleles and these loci provided the framework for combining the two parental maps into the *E. globulus* map.

Linkage analysis

Two hundred and forty nine loci were mapped into 11 major linkage groups and a 12th small linkage group, which contained three loci segregating in the male parent only (Fig. 1). Lengths of the 11 major linkage groups ranged from 90 cM to 167 cM, with an average size of 124 cM (n=11 in eucalypts; Eldridge et al. 1993). The number of loci per major linkage group varied from 35 (group1) to 15 (group 11). Six gaps on the *E. globulus* map were 20 cM or more in length. The largest distance between markers was 27 cM on linkage group 9. Twenty two loci had the same map position as other mapped loci. Most share the same position with only one other marker, although in a few cases more than two loci mapped to the same position (e.g. at 167 cM on linkage group 1). Total map distance was 1375 cM with an average interval of 6 cM based on the 227 loci that mapped to unique positions.

One hundred and thirty five RFLP probes used to detect loci on the *E. globulus* map were also used on the *E. nitens* map (Byrne et al. 1995). Of these probes, 122 (90%) detect loci that appear syntenic to *E. nitens* loci, based on linkage group, marker order and distance between markers. Fifty nine RFLP probes have also been mapped in *Eucalyptus marginata* (Moran, unpublished data), and 51 (86%) detect apparent syntenic loci in *E. globulus*. Linkage groups were assigned the same numbers as in the *E. nitens* map (see Byrne et al. 1995) based on these shared markers. The similar grouping of loci, among species and between the male and female *E. globulus* maps, lends a high degree of confidence to the assignment of loci to the linkage groups.

The individual female and male maps that were merged to produce the integrated map were 1262 and 1353 cM in total length respectively. More loci segregated in the male parent than in the female parent (193 vs 156). However, there was no significant difference in the male and female rate of recombination among consecutive pairs of fully informative loci (paired *t*-test, α=0.05). The difference in map-length estimates for the parents was primarily due to addition of loci segregating in the male parent only at the top of linkage group 1 (11 cM), the bottom of linkage group 5 (25 cM), the top of linkage group 9 (48 cM), and all of linkage group 12 (12 cM). This 12th linkage group should merge with another linkage group when more linked loci are identified, probably linkage group 6 based on the map of *E. nitens.*

There were forty one loci corresponding to known proteins (five isozymes) and sequences of known or hypothesised function (14 genes and 22 cambium-specific ESTs) on the *E. globulus* map (Table 1). Some of the EST loci are also associated with lignin or cell-wall synthesis, e.g. laccase (locus e310) and xyloglucan endotransferase (loci e360a and e360b). There was little evidence that genes encoding enzymes in the same or re-

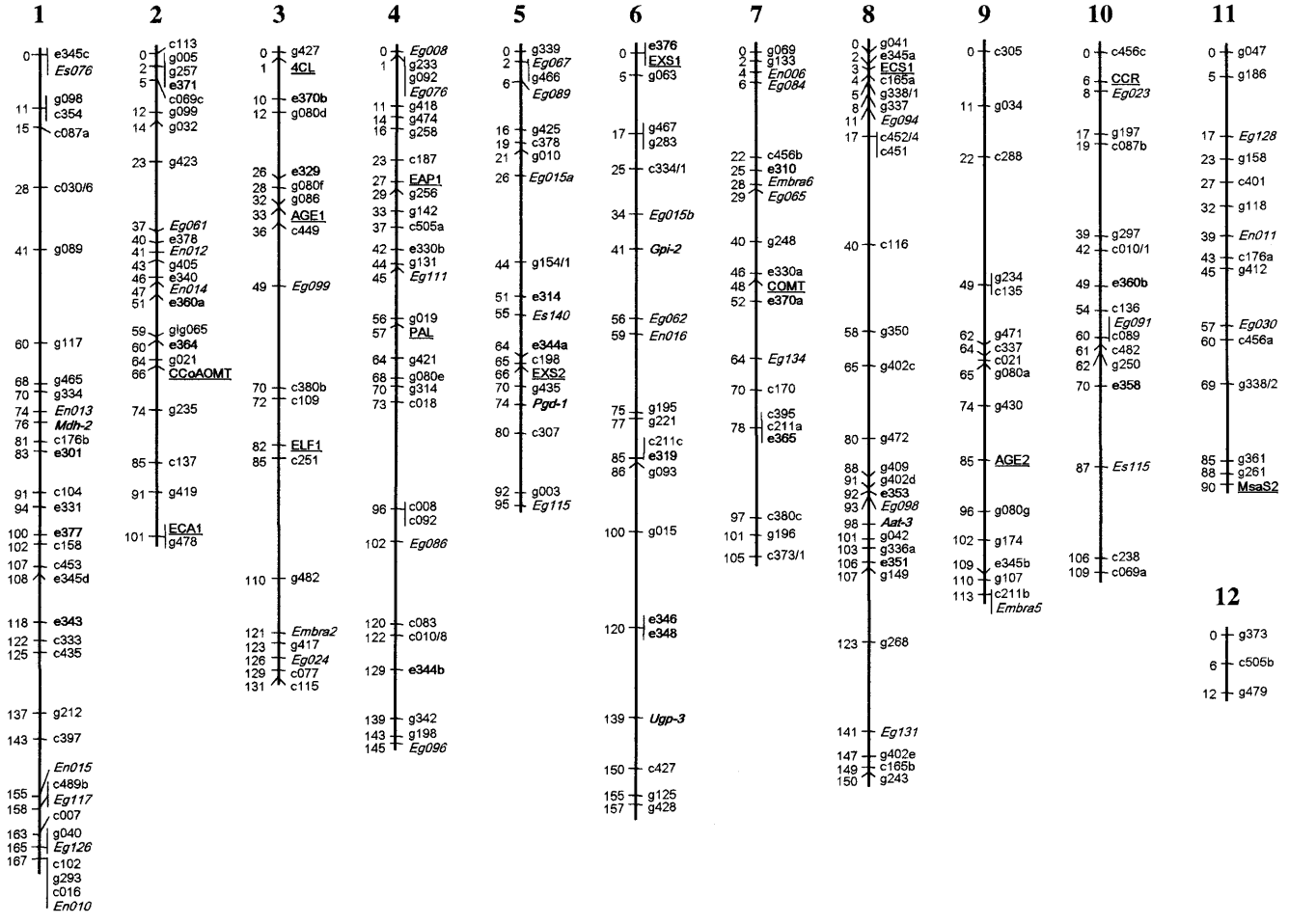

Fig. 1 Genetic linkage map of *E. globulus*. Marker identification is given to the right of the bars; cumulative distance (in Kosambi cM) of markers along groups is indicated to the left of the bars. Microsatellite markers are in *italics* and designated by the *prefix* "Eg", "En", "Es" or "Embra" followed by an *identifying number*. Isozyme markers are in ***bold italics***. Candidate gene markers are *underlined*. EST, gDNA and cDNA markers are designated by the *prefix* "e", "g" and "c" respectively, followed by the *identifying number*. Multiple loci from the same hybridising probe are indicated by *lower case letters* following the marker identification. Loci that correspond to sequences with known or hypothesised function are in *bold*

lated biochemical pathways mapped together. The five gene loci of the lignin pathway mapped to different linkage groups and similarly the two xylan synthase loci mapped to linkage groups 5 and 6. Two flowering genes, AGE1 and ELF1, mapped to linkage group 3 but were essentially unlinked (49 cM apart). The five isozyme loci mapped to linkage groups 1, 5, 6 (two loci, 98 cM apart) and 8. The distribution of microsatellite loci was reasonably even among linkage groups with the number on each group corresponding approximately to the length of the groups, with the exception of linkage group 9 which has only one microsatellite locus.

Discussion

Construction of an integrated genetic map in an outbred pedigree relies on the number and distribution of fully informative codominant markers (Byrne et al. 1995). All markers on the *E. globulus* map are codominant, and the fully informative loci are well-spaced across linkage groups. Thirty six percent of the RFLP loci mapped in *E. globulus* detect three or more alleles. This is similar to the numbers reported for the *E. nitens* mapping pedigree, where 33% (67/210) of the RFLP loci detect three or more alleles (Byrne et al. 1995). The percentage of microsatellite loci which are fully informative in the *E. globulus* intraspecific mapping pedigree (60%) is less than that found in the *E. grandis* and *E. urophylla* interspecific mapping pedigree (80%; Brondani et al. 1998). This difference is probably due to an increased number of alleles detectable at a particular locus in the interspecific cross as compared to the intraspecific cross.

The total map distance for *E. globulus* reported in this paper (1375 cM) is comparable to the 1462 cM reported for the *E. nitens* integrated map (Byrne et al. 1995). Both maps were constructed using similar methods and markers (RFLPs). However, the *E. nitens* map has 80 more mapped loci than the *E. globulus* map, primarily from the addition of RAPD loci. RAPD loci extend the *E. nitens* map by a total of 57 cM on three linkage

Table 4 Comparison of observed map lengths for linkage maps developed in *Eucalyptus*

Species	Number of markers	Number of linkage groups[a]	Observed map length (cM)[b]	Reference
E. grandis	240 RAPD	14	1551 (female)	Grattapaglia and Sederoff 1994
E. urophylla	251 RAPD	11	1101 (male)	As above
E. nitens	210 RFLP 125 RAPD 4 Isozyme	12	1462 (integrated)	Byrne et al. 1995
E. grandis	236 RAPD	11	1415 (male)	Verhaegen and Plomion 1996
E. urophylla	269 RAPD	11	1331 (female)	As above
E. tereticormis	268 AFLP	14	919 (female)	Marques et al. 1998
E. globulus	200 AFLP	16	967 (male)	As above
E. globulus	365 AFLP	11	1309 (male)	Myburg et al. 2000
E. globulus	204 RFLP 40 SSR 5 Isozyme	12	1375 (integrated)	This paper

[a] Haploid chromosome number=11 in *Eucalyptus* [b] Kosambi's mapping function

groups (7, 10 and 12). Other reported map distances for eucalypts are summarized in Table 4. For *E. globulus*, the 1375-cM distance reported in this study is comparable to the 1309-cM reported by Myburg et al. (2000). These two *E. globulus* map distances are greater than the 967-cM reported by Marques et al. (1998). The difference among *E. globulus* map distances may be due to the number of loci mapped to unique positions (120 framework, Marques et al. 1998; 168 framework, Myburg et al. 2000; 227, this study) and the coalescing of mapped loci into a minimum number of linkage groups, rather than any difference due to marker type. Based on a genome estimate of 530 Mbp/1 C for *E. globulus* (Grattapaglia and Bradshaw 1994), and the map distance obtained in this study (1375 cM), one centimorgan is approximately equivalent to 385 kbp.

Comparative map studies rely on a set of common, transferable markers that segregate in the species or pedigrees of interest. The *Mdh-2* locus on group 1 and the *Pgd-1* locus on group 5 (Fig. 1) mapped to the same positions in *E. nitens* (Byrne et al. 1995). No linkage between isozyme loci has been detected in the pedigrees from either *E. globulus* or *E. nitens*. The gDNA and cDNA markers used in this map were also used to generate the *E. nitens* map, and a high degree of synteny was observed between the two species. Other published eucalypt linkage maps have used different markers (RAPDs and AFLPs, Table 4) and direct comparisons between these maps over many loci and all linkage groups are not possible. Microsatellites, known genes, and identified ESTs present an opportunity to make initial comparisons among linkage maps.

Three EMBRA microsatellite loci (Brondani et al. 1998) are mapped in this study. Although each microsatellite locus is on a different linkage group (Fig. 1 and Brondani et al. 1998), the positions of two of the three loci within linkage groups are similar among the maps. Embra5 maps to the end of linkage group 5 in *E. grandis* and *E. urophylla*, and to the end of linkage group 9 in *E. globulus*. Embra6 maps to the centre of linkage group 1 in *E. grandis* and *E. urophylla*, and to the top third of linkage group 7 in *E. globulus*. The third locus, Embra2, that was not mapped in *E. urophylla*, maps to the centre third of linkage group 11 in *E. grandis* and to the bottom fifth of linkage group 3 in *E. globulus*.

Five lignin genes mapped in this study were also mapped by Gion et al. (2000) on a *E. urophylla*× *E. grandis* linkage map which incorporated the single-strand conformational polymorphism gene markers with previously mapped RAPD markers (Verhaegen and Plomion 1996). Comparison between the two *Eucalyptus* maps based on these genes is limited because only one gene per linkage group is in common. However, based on the relative position of the lignin gene within a linkage group the following tentative comparisons may be made: linkage groups 7 (COMT1 and COMT2), 10 (CCoAOMT) and 11 (4CL) of the hybrid eucalypt map may correspond to linkage groups 7 (COMT), 2 (CCoAOMT) and 3 (4CL) of this *E. globulus* map. Tentative comparisons based on PAL and CCR are not possible since these loci map to the same linkage group in the hybrid map (group 6) and on different linkage groups in the *E. globulus* map (group 4 and group 10). CCR maps to the end of each linkage group, and PAL maps to the centre of each linkage group. PAL gene families are known in trees such as *Populus* (Subramaniam et al. 1993) and pines (Butland et al. 1998), and in other plants such as sunflower (three loci, Gentzbittel et al. 1999), wheat (four loci, Li et al. 1999) and maize (two loci, Davis et al. 1999). It may be possible that a PAL gene family exists in the *Eucalyptus* genome and that different PAL loci were assayed in the *E. urophylla*×*E grandis* map and the *E. globulus* map reported here. Southern-blot analysis indicated at least two PAL genes in the *E. globulus* genome (data not shown); however, only one locus was polymorphic in the mapping pedigree. Alternatively, there may be chromosomal rearrangements between these species.

In other tree genera, known genes have been located on genetic linkage maps of Douglas-fir (Jermstad et al. 1998) and a *Populus* hybrid (Bradshaw et al. 1994). Un-

fortunately, these genes cannot be used to make comparisons between maps of these genera since none of the genes identified on the Douglas-fir map are similar to those mapped in this paper, and there is only one gene in common between the *Populus* and *Eucalyptus* maps, phenylalanine ammonia-lyase (PAL). The recent increase in sequencing of previously mapped cDNAs and the mapping of ESTs will provide more loci for a comparison of maps in different species and genera. Databases of identified ESTs in hybrid aspen (http://www.biochem.kth.se/PopulusDB/) and pine cDNAs (http://www.cbc.med.umn.edu/ResearchProjects/Pine/IFG.loblolly/index.html) that have been mapped and sequenced are established on the World Wide Web. In particular, the loblolly pine-EST project aims to combine linkage mapping, QTL mapping and sequence information to identify genes that determine wood properties (Neale et al. 1994).

In this paper, we report a linkage map for *E. globulus* containing lignin, cell-wall and floral genes and cambium-specific ESTs that will be used for the QTL mapping of wood, fibre, and floral traits. The map is based on the segregation of codominant loci that are transferable to other eucalypt species and provides a basis for investigating genome organization and evolution in the genus. In the future, mapping of more known genes will allow a comparison of genome organization in eucalypts to that of other plant genera and, in particular, identify synteny with model species such as *Arabidopsis*.

Acknowledgements We thank Charlie Bell, Penny Butcher, Mike Devey and Jeff Glaubitz for helpful discussions on data analysis. We also thank our colleagues for providing known gene probes; Gerd Bossinger (cambium-specific ESTs), Liz Dennis (floral genes), Simon Southerton (cell wall genes) and John Watson (MsaS2). Penny Butcher and Simon Southerton provided helpful reviews of the manuscript. Finally, we thank the Co-operative Research Centre for Sustainable Production Forestry for funding this research project.

References

Abrahams S, Hayes CM, Watson JM (1995) Expression patterns of three genes in the stem of lucerne (*Medicago sativa*). Plant Mol Biol 27:513–528

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Berke TG, Rocheford TR (1995) Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. Crop Sci 35:1542–1549

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

Botella MA, Coleman MJ, Hughes DE, Nishimura MT, Jones JDG, Somerville SC (1997) Map positions of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. Plant J 12:1197–1211

Bradshaw HD, Stettler RF (1995) Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. Genetics 139:963–973

Bradshaw HD, Villar M. Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet 89:167–178

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theor Appl Genet 97:816–827

Butland SL, Chow M, Ellis BE (1998) A diverse family of phenylalanine ammonia-lyase genes expressed in pine trees and cell cultures. Plant Mol Biol 37:15–24

Byrne M, Moran GF, Tibbits WN (1993) Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. J Hered 84:218–220

Bryne M, Moran GF, Murrell JC, Tibbits WN (1994) Detection and inheritance of RFLPs in *Eucalyptus nitens*. Theor Appl Genet 89:397–402

Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theor Appl Genet 91:869–875

Byrne M, Marques-Garcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus.* Aust J Bot 44:331–341

Byrne M, Murrell JC, Owen JV, Kriedmann P, Williams ER, Moran GF (1997) Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens.* Theor Appl Genet 94:674–681

Davis GL, McMullen MD, Baysdorfer C, Musket T, Grant D, Staebell M, Xu G, Polacco M, Koster L, Melia-Hancock S, Houchins K, Chao S, Coe EH (1999) A maize-map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics 152:1137–1172

Devey MF, Sewell MM, Uren TL, Neale DB (1999) Comparative mapping in loblolly and radiata pine using RFLP and microsatellite markers. Theor Appl Genet 99:656–662

Eldridge K, Davidson J, Harwood C, van Wyk G (1993) Eucalypt domestication and breeding. Oxford University Press, Oxford

Faris JD, Li WL, Liu DJ, Chen PD, Gill BS (1999) Candidate gene analysis of quantitative disease resistance in wheat. Theor Appl Genet 98:219–225

Gentzbittel L, Mourzeyar S, Badaoui S, Mestries E, Vear F, Tourvieille De Labrouhe D, Nicolas P (1998) Cloning molecular markers for disease resistance in sunflower, *Helianthus annuus* L. Theor Appl Genet 96:519–525

Gentzbittel L, Mestries E, Mouzeyar S, Mazeyrat F, Badaoui S, Vear F, Tourvieille de Labrouhe D, Nicolas P (1999) A composite map of expressed sequences and phenotypic traits of the sunflower (*Helianthus annuus* L.) genome. Theor Appl Genet 99:218–234

Gion J-M, Rech P, Grima-Pettenati J, Verhaegen D, Plomion C (2000) Mapping candidate genes in Eucalyptus with empahsis on lignification genes. Molecular Breed 6:441–449

Glaubitz JC, Strk J, Moran GF (1999) Genetic impacts of different silvicultural practices in native eucalypt forests. In: Matyas CS (ed) Forest genetics and sustainability. Kluwer, Dordrecht, The Netherlands, pp 183–196

Glaubitz JC, Moran GF (2000) Genetic tools. In: Young A, Boyle T, Boshier D (eds) Forest conservation genetics: principles and practice. CSIRO Publishing, Melbourne, pp 39–59

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

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

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121–1137

Grattapaglia D, Bertolucci FLG, Penchel R, Sederoff RR (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. Genetics 144:1205–1214

Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchel-Olds T, Sherman B, Vujcic S, Williams C, Neale D (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics 138: 1293–1300

Harcourt R, Kyozuka J, Zhu X, Southerton S, Llewellyn D, Dennis E, Peacock J (1995) Genetic engineering for sterility in temperate plantation eucalypts. In: Potts BM, Borralho NMG, Reid JB, Cromer RN, Tibbits WN, Raymond CA (eds) Eucalypt plantations: improving fibre yield and quality. Proc. CRC-IUFRO Conf., Hobart, 19–24 Feb, CRC for Temperate Hardwood Forestry. Hobart, Tasmania, pp, 403–405

Jermstad KD, Bassoni DL, Kinlaw CS, Neale DB (1998) Partial DNA sequencing of Douglas-fir cDNAs used for RFLP mapping. Theor Appl Genet 97:771–776

Kianian SF, Egli MA, Phillips RL, Rines HW, Somers DA, Gengenbach BG, Webster FH, Livinston SM, Groh S, O'Donoughue LS, Sorrells ME, Wesenberg DM, Stuthman DD, Fulcher RG (1999) Association of a major groat oil content QTL and an acetyl-CoA carboxylase gene in oat. Theor Appl Genet 98:884–894

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Kyozuka J, Harcourt R, Peacock WJ, Dennis ES (1997) *Eucalyptus* has functional equivalents of the *Arabidopsis AP1* gene. Plant Mol Biol 35:573–584

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Li WL, Faris JD, Chitoor JM, Leach JE, Hulbert SH, Liu DJ, Chen PD, Gill BS (1999) Genomic mapping of defense response genes in wheat. Theor Appl Genet 98:226–233

Lincoln SE, Daly MJ, Lander ES (1991) Primer software, available from E. Lander, Whitehead Institute, Cambridge, Massachusetts

Maliepaard C, Jansen J, van Ooigen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Genet Res 70:237–250

Marques CM, Araújo JA, Ferreira JG, Whetten R, O'Malley DM, Liu B-H, Sederoff R (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. Theor Appl Genet 96:727–737

Myburg AA, Griffin AR, Sederoff RR, Whetten R (2000) Genetic analysis of interspecific backcross families of a hybrid of *Eucalyptus grandis* and *Eucalyptus globulus*. In: Dungey HS, Dieters MJ, Nikles DG (eds) Hybrid breeding and genetics of forest trees. Proc QFRI/CRC-SPF Symposium, 9–14 April 2000, Noosa, Queensland, Australia, pp 462–466

Neale DB, Kinlaw CS, Sewell MM (1994) Genetic mapping and DNA sequencing of the loblolly pine genome. For Genet 1:197–206

Schuler GD, Altschul SF, Lipman DJ (1991) A workbench for multiple alignment construction and analysis. Proteins Struct Funct Genet 9:180–190

Sewell MM, Neale DB (2000) Mapping quantitative traits in forest trees. In: Jain SM, Minocha SC (eds) Molecular biology of woody plants, vol. 1. Kluwer Academic Publishers, The Netherlands, pp 407–423

Southerton SG, Strauss SH, Olive MR, Harcourt RL, Decroocq V, Zhu X, Llewellyn DJ, Peacock WJ, Dennis ES (1998) *Eucalyptus* has a functional equivalent of the *Arabidopsis* floral meristem identifing the gene *LEAFY*. Plant Mol Biol 37: 897–910

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Joinmap. Plant J 3:739–744

Stam P, van Ooijen JW (1995) Joinmap version 2.0: software for the calculation of genetic linkage maps. CPRO-DLO. Wageningen

Staub JE, Serquen FC, Gupta M (1996) Genetic markers, map construction, and their application in plant breeding. Hort Sci 31:729–740

Subramaniam R, Reinold S, Molitor EK, Douglas DJ (1993) Structure, inheritance, and expression of hybrid poplar (*Populus tricocarpa*×*Populus deltoides*) phenylalanine ammonia-lyase genes. Plant Physiol 102:71–83

Tanksley SD, Ganal MW, Prince JP, de Vincente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder S, Wing RA, Wu W, Young ND (1992) High-density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Verhaegen D, Plomion C (1996) Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. Genome 39:1051–1061

Volker PW (1995) Evaluation of *Eucalyptus nitens*×*globulus* for commercial forestry. In: Potts BM, Borralho NMG, Reid JB, Cromer RN, Tibbits WN and Raymond CA (eds) Eucalypt plantations: improving fibre yield and quality. Proc CRC-IUFRO Conf, Hobart, 19–24 Feb, CRC for Temperate Hardwood Forestry, Hobart, Tasmania, pp 222–225

Whetten RW, MacKay JJ, Sederoff RR (1998) Recent advances in understanding lignin biosynthesis. Annu Rev Plant Physiol Plant Mol Biol 49:585–609