

STRUCTURAL CHARACTERISATION OF *PINUS RADIATA* MADS-BOX DNA SEQUENCES ISOLATED BY PCR CLONING

DANIEL YUNQIU WANG*

Gene Expression Laboratory, Molecular and Cell Biology Section,
Bio-Resources Technology Division, Forestry and Forest Product Research Institute,
Tsukuba, 305 Japan

ROSIE E. BRADSHAW

Department of Microbiology and Genetics, Massey University,
Palmerston North, New Zealand

CHRISTIAN WALTER

New Zealand Forest Research Institute,
Private Bag 3020, Rotorua, New Zealand

MARIE B. CONNETT

Fletcher Challenge Forests, Te Teko, New Zealand

and DAVID W. FOUNTAIN

Department of Plant Biology and Biotechnology, Massey University,
Palmerston North, New Zealand

(received for publication 15 January 1997; revision 7 June 1997)

ABSTRACT

Flowering in all plant species analysed to date is regulated by highly conserved and developmentally regulated genes. Some of them were found to belong to the group of MADS-box genes. In the research described here, the construction of two opposing degenerate primers, targeted to highly conserved regions within the MADS box, allowed the amplification of a 78 bp segment from genomic DNA of *Pinus radiata* D. Don. By subcloning these PCR (Polymerase Chain Reaction) products into M13 and analysing the sequence of individual subclones, three different DNA sequences, each representing a conserved MADS-box region of three independent *P. radiata* genes were identified. The three MADS-box sequences shared nucleotide and amino acid sequence identity with homeotic genes of *Arabidopsis thaliana*:—two of them with the stamen-carpel-specific *AGAMOUS* gene and one with the stamen-petal-specific *AP3* gene. The sequences isolated here are currently being used to isolate and characterise full-length

* To whom correspondence should be addressed.

MADS-box gene sequences which could provide a useful tool for the manipulation of flowering in *P. radiata*.

Key words: MADS-box sequences; *AGAMOUS*-like; *AP3*-like; floral development; *Pinus radiata*.

INTRODUCTION

A stable transformation system for *Pinus radiata* has been developed to introduce novel traits into clonal material (Walter & Smith 1995; Walter *et al.* in prep.). The ultimate aims are to generate transgenic trees resistant to herbicides, insects, and pathogens, or to introduce other desirable traits. However, before commercial plantation is permitted the containment of transgenes inserted into genetically engineered forest trees may be necessary because of concern for ecological safety (Strauss *et al.* 1995). This ecological concern arises because of the movement of transgenes into the environment through the seasonal release of pollen and seeds, particularly through the release of the massive amount of pollen grains shed in early spring. In addition, nutrient allocation from vegetative growth to reproductive growth, such as pollen development, has been considered a waste of energy and resources (Ledig 1986). Moreover, pine pollen has also been reported as a potential source of allergens (Fountain & Cornford 1991). Because of these considerations, research directed to a greater understanding of male "floral" development in pines is of vital importance for modern plantation forestry. Both downregulation and acceleration of flowering are considered valuable for commercial tree plantations and breeding programmes. Manipulative treatments designed to influence reproductive activity can be attempted using molecular techniques. MADS-box genes (Schwarz-Sommer *et al.* 1990) encode a family of transcription factors, many of which control inflorescence and flower development in higher plants (Yanofsky 1995). They bind DNA *via* the MADS-box domain, a highly conserved region approximately 60 amino acids long (Yanofsky *et al.* 1990). The determination of organ identity in the four whorls of the angiosperm flower depends on the combining action of a set of genes (Bowman *et al.* 1991). To interpret the interaction of these organ identity genes, a model for the flower organ development process has been proposed where three different gene functions (A, B, and C) are combined to define four concentric zones in the flower meristem that confer separate identities for sepals (A), petals (AB), stamens (BC), and carpels (C) (Yanofsky 1995). The floral meristem identity gene *APETALA1* (*AP1*), from *Arabidopsis thaliana*, is involved in establishing the identity of the floral primordium as well as the A function (Mandel, Gustafson-Brown, Savidge, Yanofsky 1992). Other *A. thaliana* floral organ identity genes include *APETALA3* (*AP3*) a B-function gene (Jack *et al.* 1992), and *AGAMOUS* (*AG*) a C-function gene (Yanofsky 1995). It is now clear that these three *A. thaliana* floral-specific genes are all MADS-box genes (Mandel, Gustafson-Brown, Savidge, Yanofsky 1992; Jack *et al.* 1992; Yanofsky 1995). A growing number of MADS-box-containing sequences which affect flower development in plants have been cloned on the basis of similarity to sequences of mutationally defined MADS-box genes. These include *AGL-1, 2, 4, 5, 6* (*Arabidopsis thaliana*) (Ma *et al.* 1991), *BAG1* (*Brassica napus*) (Mandel, Bowman, Kempin, Ma, Meyerowitz, Yanofsky 1992), *FBP2* (*Petunia hybrida*) (Angenent *et al.* 1992), *DAL-1, 2, 3* (*Picea abies*) (Tandre *et al.* 1995), *OM1* (*Aranda deborah*) (Lu *et al.* 1993), *OsMADS1* (*Oryza sativa*) (Chung *et al.* 1994), *TAG1* (*Lycopersicon esculentum*) (Pnueli *et al.* 1994), and *ZAG1* and *ZAG2* (*Zea mays*) (Schmidt *et al.* 1993).

In this paper, we report three *Pinus radiata* DNA sequences amplified by PCR which share nucleotide and amino acid sequence identity with the MADS-box family.

MATERIALS AND METHODS

The needle fascicle tissue used for DNA extraction was collected from *P. radiata* Clone 880-606 growing in the central North Island of New Zealand. The template genomic DNA was extracted from the needle fascicle tissue using the CTAB method modified from that of Murray & Thompson (1980).

MADS-box DNA sequences from *P. radiata* were amplified by polymerase chain reaction (PCR) with degenerate primers

forward primer: 5'-CGGAATTCMGICARGTIACITT-3',

reverse primer: 5'-GCTCTAGATCIGCRTRCAIARIAC-3',

designed according to Tikka *et al.* (1993); M = C or A; I = Inosine; R = A or G. The aim was to amplify the conserved region between the RQVT and VLCDAE amino acid motifs found in the MADS-box domain of the *Arabidopsis thaliana* *AGAMOUS* gene. For convenience in cloning, restriction sites *Eco* RI and *Xba* I (underlined) were included at the ends of the primers. PCR was carried out in a 100- μ l reaction mixture containing 1 μ l genomic DNA, 0.2 μ M each primer, 1.25 mM each dNTPs, 1 \times Taq PCR reaction buffer, and 2 units of Taq DNA polymerase (Boehringer Mannheim GmbH). The PCR reaction mixture was preheated at 94°C for 5 min initially and then subjected to 20 cycles of 94°C for 45 s, 37°C for 45 s, and 72°C for 60 s, and another 20 cycles of 94°C for 45 s, 45°C for 45 s, and 72°C for 60 s. After the 40 cycles were complete the reactions were incubated at 72°C for 5 min and then stored at 4°C or -20°C. A 3% agarose minigel was used to check 10 μ l of the amplified DNA. PCR products were cloned in M13 mp18 and sequenced (Sanger *et al.* 1977). *Arabidopsis thaliana* DNA was included as a control in PCR.

Alignments of sequences isolated were carried out using PILEUP with a gap weight of 5.0 and a gap length weight of 0.3 (Devereux 1989).

To further characterise the relationship of the pine DNA sequences among themselves and their relationship with other MADS-box genes, a dendrogram tree based on the PILEUP program of the GCG package was plotted to show the clustering relationships of these sequences.

PILEUP creates a multiple sequence alignment using a simplification of the progressive alignment method of Feng & Doolittle (1987). A dendrogram tree is a representation of clustering relationships, showing the order of the pairwise alignments of the selected DNA sequences. This clustering method is sensitive to the order in which sequences are aligned. A clustering algorithm determines this order from the pairwise similarities calculated before the final alignments are done. According to PILEUP, the distance along the horizontal axis is proportional to the difference between clusters and sequences, the vertical axis has no significance at all (Devereux 1989).

RESULTS AND DISCUSSION

The goal of this research was to detect DNA fragments homologous to MADS-box genes and analyse them on a sequence basis. Fourteen DNA sequences were amplified from

P. radiata using MADS-box primers and were aligned with the conserved MADS-box region of the *AGAMOUS* (*AG*) gene from *A. thaliana* (Fig. 1).

The sequences showed varying degrees of identity to the *AG* MADS-box, ranging from 28.9% to 84.4%. Apart from Pm7 and Pm12, the DNA sequences amplified in this study had a high degree of identity (62.2%–84.4%) to the *AG* MADS-box and we concluded that they were amplified from pine genomic DNA. The pine MADS-box sequences were not identical to each other, and they did not have 100% identity to the *AG* MADS-box sequence. This

```

Pm12  ***G**G**G  **G**G**TTC  *TT***AGG*  ATT*CAT**A  TGTG*TTG*C  **TCTA*GC*  ***CT*****  **C**C**
Pm7   ***G**G**G  **G**G**TTC  *TT***AGG*  ATT*CAT**A  TGTG*TTG*C  **TCTA*GC*  ***C*****  **C**C**
Pm14  *A**G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm5   ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm10  ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  *CA**C**
Pm6   ***GT***G  *TG**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm1   ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm13  ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm4   *A**G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm11  ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm8   ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm2   ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm9   ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
Pm3   ***G**G**G  **G**C**CG*  *G**GC**G*T  *G**G**T  **A**G**AC  *G**G**T**  **C**C*****  **C**C**
AG    ACGTCAAGTC  ACCTTTTTCGA  AACGTAGAAA  TGGTTTGCTC  AAGAAAGCTT  ACGAGCTCTC  TGTCTCTGT  GATGCTGA

```

FIG. 1—Fourteen (Pm1–Pm14) DNA sequences isolated from the genomic DNA of *Pinus radiata* by PCR, using primers based on the conserved MADS-box region, were aligned with the conserved MADS-box DNA sequence of *AG* (bottom sequence). Nucleotides different from the *AG* sequence are indicated (bold letters). Positions indicated by * are the same nucleotides as in the *AG* sequence.

eliminates the possibility of these amplified DNA sequences being contaminated by *A. thaliana* DNA during PCR, which is a concern since *A. thaliana* DNA was consistently used as a positive control during the amplification of pine genomic DNA by PCR. When the amplified DNA sequences were compared with one another, it was found that they could be divided into four groups, represented by Pm1 (Pm1, 5, 6, 10, 14), Pm2 (Pm2, 3, 8, 9, 11), Pm4 (Pm4, 13), and Pm7 (Pm7, 12). Within each of these groups, minor sequence differences were detected (e.g., one nucleotide difference between Pm3 and Pm9) but we cannot exclude the possibility that these minor differences were due to misincorporation during PCR amplification.

A comparison of the deduced amino acid sequence of these four representative pine sequences and fourteen major MADS-box genes (Fig. 2) showed that Pm4 and *DAL-2* (the *Picea abies* MADS-box gene described by Tandre *et al.* 1995) shared identical amino acid sequences with *AG*, *AGL-1*, and *AGL-5* of *A. thaliana* (hereafter referred to as the “*AG* group”) in the region analysed. Pm2, *DAL-1*, and *DAL-3* shared an identical amino acid sequence with *AGL-6*. Pm2 was different from the *AG* group at only one amino acid by substitution of serine (S) instead of cysteine (C). It is of interest that *AGL-2* and *AGL-4* also differed from the *AG* group at the same amino acid position, having an alanine (A) instead of a cysteine (C). Due to differences in hypothesised function of these groups of MADS-box

<u>Pm7</u>	* * * * L F H C K E F H L C D C H S M P * * * * *
AP3	* * * * Y S * * * M * * F * * * H * * T * F * * * R
AP1	* * * * * S * * * A * * * * * H * I * * * * * *
<u>Pm1</u>	* * * * * S * * * M * * * * * Q * * * * * * * *
<u>Pm2</u>	* * * * * S * * * * * * * * * * * * * * * * * *
DAL-1	* * * * * S * * * * * * * * * * * * * * * * * *
DAL-3	* * * * * S * * * * * * * * * * * * * * * * * *
AGL-6	* * * * * S * * * * * * * * * * * * * * * * * *
<u>Pm4</u>	* *
DAL-2	* *
AGL-1	* *
AGL-5	* *
AGL-2	* * * * * A * * * * * * * * * * * * * * * * * *
AGL-4	* * * * * A * * * * * * * * * * * * * * * * * *
FBP2	* * * * * A * * * * * * * * * * * * * * * * * *
OsMADS1	* * * * * A * * * * * * * * * * * * * * * L * * * *
OM1	* * * * * A * * * K R * * * * * * * * * * * * *
AG	<u>R O V T F A K R R N G L L K K A Y E L S V L C D A E</u>

FIG. 2—An alignment of the deduced amino acid sequences of four PCR clones (**Pm1**, **Pm2**, **Pm4**, and **Pm7**) of *Pinus radiata* with deduced amino acid sequences of conserved MADS-box-like regions from various plant species (primer sequences underlined). The deduced amino acid sequences of four DNA sequences isolated from *P. radiata* were determined using the TRANSLATE programme. *AG*, *AGL1-6*, *AP1*, and *AP3* are from *Arabidopsis thaliana*, *OM1* is from *Aranda Deborah*, *FBP2* is from *Petunia hybrida*, *DAL1*, *DAL2*, and *DAL3* are from *Picea abies*, *OsMADS1* is from *Oryza sativa*. Amino acids different from the *AG* sequence are shown in bold letters. Asterix indicate identical amino acids with regard to the *AG* sequence.

genes and the importance of cysteine residues in determining protein folding (Lehninger 1975), it is suggested that changes at this position may have functional significance. Pm1 was different from the *AG* group in three amino acids. The positions at which Pm1 differed from the *AG* group were similar to those at which the *A. thaliana* *AP3* and *AP1* (genes controlling the earlier stages of floral development) differed from the *AG* group (genes controlling the later stages of floral development). Pm7 showed an almost completely different DNA and amino acid sequence from the rest of the MADS-box genes, and we have consequently removed it from further analysis.

To further characterise the relationships of the three groups of pine MADS-box-like DNA sequences (Pm1, 2, 4) to each other and other MADS-box sequences, a dendrogram tree based on the PILEUP program was plotted to show the clustering relationships of these sequences (Fig. 3).

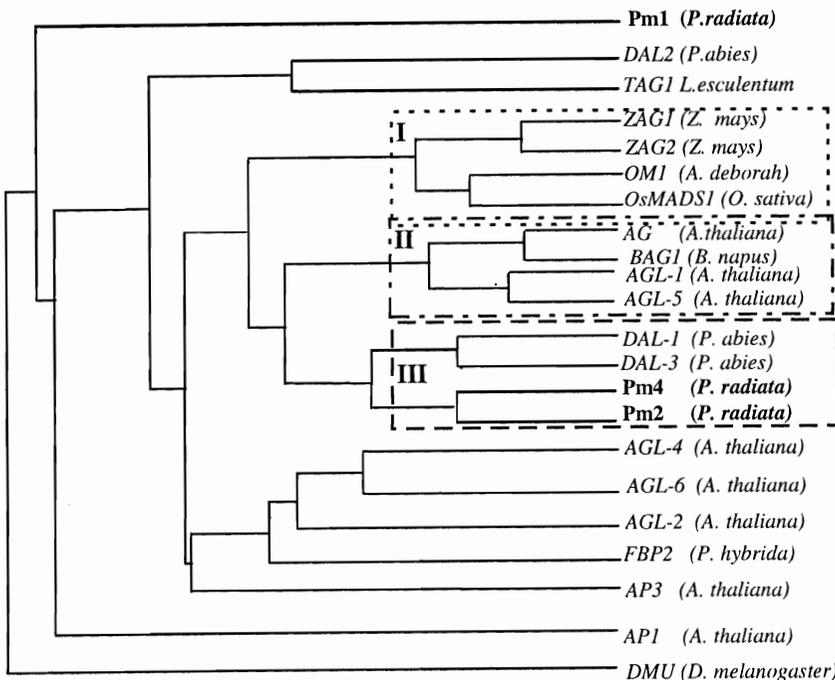


FIG. 3—A dendrogram based on the pairwise sequence alignment, showing relatedness of three pine DNA sequences excluding primer regions (**Pm1**, **Pm2**, and **Pm4**) with MADS-box DNA sequences from other species. This dendrogram used the PILEUP program. Box I includes four genes from monocotyledonous species, box II includes the *AG* gene and three genes closely related to *AG* from dicotyledonous species, and box III includes two pine DNA sequences from this study and two genes from another conifer species.

A close relationship between Pm2 and Pm4 is suggested, and the sequences are shown to form a *Pinus* cluster, which is closely related to the *Picea* cluster of *DAL-1* and *DAL-3*, forming a conifer group (box III). The immediate neighbouring group is an angiosperm group consisting of four MADS-box genes (*AG* and three *AG*-like genes) from two dicotyledon species *A. thaliana* and *Brassica napus*, of the family Brassicaceae (box II). The

next neighbouring group to these two closely related groups is a monocotyledon group consisting of four MADS-box genes from three monocotyledon species of two families, Poaceae and Orchidaceae (box I). Many of the genes in boxes I, II, and III have been implicated in late floral development.

Comparing the DNA sequence and deduced amino acid sequence data, we hypothesise that three groups of DNA sequences isolated from *P. radiata*, represented by Pm1, Pm2, and Pm4, may be part of *Pinus* homologues to genes that could have a role in controlling floral development in angiosperms. Pm2 and Pm4 pine DNA sequences are more closely related to MADS-box genes controlling late floral development, such as *AG* from *A. thaliana* and *AGAMOUS*-like genes from a number of other plant species. Their close relationship to *DAL-1,2,3* genes from *Picea abies* not only suggests a common coniferous origin but also indicates that these *Pinus radiata* DNA sequences may derive from MADS-box genes which have a similar function to *DAL-1,2,3* genes. Products of *DAL-1,2,3* genes from *Picea abies* have been found expressed during the development of both male and female reproductive organs, with *DAL-2* being expressed only in developing male and female cones (Tandre *et al.* 1995). Based on this comparison, it is hypothesised that Pm2 and Pm4 DNA sequences arise from pine genes which are functional as well as structural homologues of the angiosperm class C genes. Pm1 is more closely related to homeotic genes controlling earlier steps of floral development, such as class B genes controlling petal and stamen development in angiosperms. We have used these DNA sequences as probes in order to identify a number of MADS-box genes from male cone tissue of *Pinus radiata*. These are possibly genes involved in the developmental processes leading to male cone formation and are therefore candidates for a molecular approach to slow down cone development (Walden *et al.* in prep.).

ACKNOWLEDGMENTS

This research was supported by grants from the Foundation for Research Science and Technology (FRST) and the Massey University Graduate Research Fund. We gratefully acknowledge the generous gift of *agamous* cDNA from Dr E.M.Meyerowitz, and valuable comments on the PCR study given by Dr D.W.R.White.

REFERENCES

- ANGENENT, G.C.; BUSSCHER, M.; FRANKEN, J.; MOL, J.N.M.; VAN TUNEN, A.J. 1992: Differential expression of two MADS-box genes in wild type and mutant *Petunia* flowers. *Plant Cell* 4: 983–993.
- BOWMAN, J.L.; SMYTH, D.R.; MEYEROWITZ, E.M. 1991: Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112: 1–20.
- CHUNG, Y-Y.; KIM, S-R.; FINKEL, D.; YANOFSKY, M.F.; AN, G. 1994: Early flowering and reduced apical dominance result from ectopic expression of a rice MADS box gene. *Plant Molecular Biology* 26: 657–665.
- DEVEREUX, J. 1989: “The GCG Sequence Analysis Software Package”, version 6.0. University Research Park, Madison, Wisconsin, USA.
- FENG, D-F.; DOOLITTLE, R.F. 1987: Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* 35: 351–360.
- FOUNTAIN, D.W.; CORNFORD, C.A. 1991: Aerobiology and allergenicity of *Pinus radiata* pollen in New Zealand. *Grana* 30: 71–75.

- JACK, T.; BROCKMAN, L.; MEYEROWITZ, E.M. 1992: The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* 68: 683–688.
- LEDIG, F.T. 1986: Conservation strategies for forest gene resources. *Forest Ecology & Management* 14: 77–90.
- LEHNINGER, A.L. 1975: "Biochemistry". Worth Publishers Inc., NY.
- LU, Z-X.; WU, M.; LOH, C-S.; YEONG, C-Y.; GOH, C-J. 1993: Nucleotide sequence of a flower-specific MADS box cDNA clone from orchid. *Plant Molecular Biology* 23: 901–904.
- MA, H.; YANOFSKY, M.F.; MEYEROWITZ, E.M. 1991: *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Development* 5: 484–495.
- MANDEL, M.A.; GUSTAFSON-BROWN, C.; SAVIDGE, B.; YANOFSKY, M.F. 1992: Molecular characterisation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* 360: 273–277.
- MANDEL, M.A.; BOWMAN, J.L.; KEMPIN, S.A.; MA, H.; MEYEROWITZ, E.M.; YANOFSKY, M.F. 1992: Manipulation of flower structure in transgenic tobacco. *Cell* 71: 133–143.
- MURRAY, G.; THOMPSON, W.F. 1980: Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Research* 8(19): 4321–4325.
- PNUELI, L.; HAREVEN, D.; ROUNSLEY, S.D.; YANOFSKY, M.F.; LIFSCHITZ, E. 1994: Isolation of the tomato *AGAMOUS* gene *TAG1* and analysis of its homeotic role in transgenic plants. *Plant Cell* 6: 163–173.
- SANGER, F.; NICKLEN, S.; COULSON, A.R. 1977: DNA sequencing with chain terminating inhibitors. *Proceedings National Academy Science USA* 74: 5463–5467.
- SCHMIDT, R.J.; VEIT, B.; MANDEL, M.A.; MENA, M.; HAKE, S.; YANOFSKY, M.F. 1993: Identification and molecular characterisation of *ZAG1*, the maize homologue of the *Arabidopsis* floral homeotic gene *AGAMOUS*. *Plant Cell* 5: 729–737.
- SCHWARZ-SOMMER, Z.; HUIJSER, P.; NACKEN, W.; SAEDLER, H.; SOMMER, H. 1990: Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250: 931–936.
- STRAUSS, S.H.; ROTTMANN, W.H.; BRUNNER, A.M.; SHEPPARD, L.A. 1995: Genetic engineering of reproductive sterility in forest trees. *Molecular Breeding* 1: 5–26.
- TANDRE, K.; SUNDÅS, A.; ALBERT, V.A.; ENGSTRÖM, P. 1995: Conifer homologues to genes that control floral development in angiosperms. *Plant Molecular Biology* 27: 69–78.
- TIKKA L.; KARJALAINEN, A.; SOPANEN, T. 1993: Floral MADS-box genes in birch (*Betula pendula*). In: Proceedings of the 1993 IUFRO Symposium Section S2.01-05: Biology and Control of Reproductive Processes in Forest Trees.
- WALDEN, A.R.; WANG, D.Y.; WALTER, C.; GARDNER, R.C.: A large family of TM3 orthologs in *Pinus radiata* includes two members with deletions of the conserved K domain (in prep.)
- WALTER, C.; SMITH, D.R. 1995: Transformed *Pinus radiata* now growing in greenhouse at the New Zealand Forest Research Institute (NZ FRI). *Dendrome: Forest Tree Genome Research Updates* 2(2).
- WALTER, C.; GRACE, L.J.; WAGNER, A.; WHITE, D.W.R.; WALDEN, A.R.; DONALDSON, S.S.; HINTON, H.; GARDNER, R.C.; SMITH, D.R. Stable transformation and regeneration of transgenic plants of *Pinus radiata* D Don (in prep.)
- YANOFSKY, M.F. 1995: Floral meristem to floral organs: genes controlling early events in *Arabidopsis* flower development. *Annual Review Plant Physiology Plant Molecular Biology* 46: 167–188.
- YANOFSKY, M.F.; MA, H.; BOWMAN, J.L.; DREWS, G.; FELDMANN, K.; MEYEROWITZ, E.M. 1990: The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature* 346: 35–39.