Identifying *Eucalyptus* expressed sequence tags related to *Arabidopsis* flowering-time pathway genes

Marcelo Carnier Dornelas[1,2]* and Adriana Pinheiro Martinelli Rodriguez[1]

*[1]Universidade de São Paulo, Centro de Energia Nuclear na Agricultura, Laboratório de Biotecnologia Vegetal, Av. Centenário, 303, CEP 13400-970 Piracicaba, SP, Brasil; [2]Present address: Universidade Estadual de Campinas, Instituto de Biologia, Departamento de Fisiologia Vegetal, CP 6109, Cidade Universitária ¨Zeferino Vaz¨, CEP 13084-971 Campinas, SP, Brasil. *Corresponding author:*

Received: 21/05/2005, Accepted: 28/06/2005

Flowering initiation depends on the balanced expression of a complex network of genes that is regulated by both endogenous and environmental factors. The timing of the initiation of flowering is crucial for the reproductive success of plants; therefore, they have developed conserved molecular mechanisms to integrate both environmental and endogenous cues to regulate flowering time precisely. Extensive advances in plant biology are possible now that the complete genome sequences of flowering plants is available and plant genomes can be comprehensively compared. Thus, association studies are emerging as powerful tools for the functional identification of genes involved on the regulation of flowering pathways. In this paper we report the results of our search in the Eucalyptus Genome Sequencing Project Consortium (FORESTS) database for expressed sequence tags (ESTs) showing sequence homology with known elements of flowering-time pathways. We have searched the 33,080 sequence clusters in the FORESTS database and identified *Eucalyptus* sequences that codify putative conserved elements of the autonomous, vernalization-, photoperiod response- and gibberellic acid-controlled flowering-time pathways. Additionally, we have characterized *in silico* ten putative members of the *Eucalyptus* homologs to the *Arabidopsis* CONSTANS family of transcription factors.

Keywords: CONSTANS, data mining, EST, photoperiod, vernalization.

Identificação de etiquetas de seqüências expressas (ESTs) de *Eucalytpus* relacionadas a genes das vias controladoras do florescimento de *Arabidopsis*: As decisões envolvidas na iniciação do processo de florescimento dependem da expressão equilibrada de uma rede complexa de genes, que é regulada por fatores endógenos e ambientais. A regulação correta da transição para o florescimento é crucial para o sucesso reprodutivo das plantas; dessa forma elas desenvolveram mecanismos moleculares conservados para integrar tanto indícios ambientais como endógenos para regular precisamente o tempo do florescimento. Avanços recentes na biologia molecular vegetal, incluindo o seqüenciamento completo de genomas, tornaram possível a comparação detalhada de genomas vegetais. Assim, estudos comparativos de genômica funcional estão emergindo como ferramentas poderosas para a identificação de genes envolvidos no regulamento das vias metabólicas relacionadas ao florescimento. Neste artigo são apresentados resultados da busca no banco de dados do Projeto FORESTS (*Eucalyptus* Genome Sequencing Project Consortium) para a obtenção de etiquetas de seqüências expressas (ESTs) mostrando homologia com elementos-chave das vias reguladoras do florescimento. Os 33.080 ¨clusters¨ de ESTs de *Eucalyptus* spp. disponíveis no banco de dados do FORESTS foram analisados, e identificaram-se elementos conservados, possivelmente envolvidos nas vias controladoras do florescimento, quer sejam as autônomas, quer as controladas por vernalização, por ácido giberélico ou por fotoperíodo. Adicionalmente, foram caracterizados *in silico* dez possíveis membros da família CONSTANS de fatores de transcrição de *Arabidopsis*, em *Eucalyptus*.

Palavras-chave: CONSTANS, EST, fotoperíodo, mineração de dados, vernalização.

INTRODUCTION

Flowering is controlled by environmental conditions and developmental regulation (Reeves and Coupland, 2000; Samach and Coupland, 2000; Araki, 2001). The complexity of this regulation is created by an intricate network of signaling pathways (Reeves and Coupland, 2000; Samach and Coupland, 2000; Araki, 2001; Mouradov et al., 2002). Studies in the model plant *Arabidopsis* have led to the identification of components within individual signaling pathways that affect flowering, and to their positioning within molecular hierarchies. *Arabidopsis* is an excellent model system in which to approach this complexity, because it responds to many of the environmental conditions that control flowering in other species, and for which genetic tools are well developed (Levy and Dean, 1998; Simpson and Dean, 2002). Furthermore, distinct signaling pathways are known to converge on the activation of the same flowering-time genes (Mouradov et al., 2002; Simpson and Dean, 2002). This convergence of pathways on a common set of genes may enable the integration of different responses, so that the plant can produce a coordinated flowering process under conditions in which multiple environmental parameters are changing simultaneously (Levy and Dean, 1998; Simpson and Dean, 2002; Izawa et al., 2003). Also, genetic analysis of *Arabidopsis* varieties showing natural variation in flowering time has demonstrated how the activity of these pathways can be altered in nature and how balancing the effects of different environmental stimuli on flowering time is important in plants adapting to growth in different geographical locations (Koornneef et al., 1994; Samach and Coupland, 2000; Araki, 2001).

At present, the full complexity of the flowering network can only be approached in *Arabidopsis* and probably rice, where the necessary tools are available (Izawa et al., 2003), nevertheless extensive efforts are being made to describe related pathways in other plant species (Mouradov et al., 2002; Izawa et al., 2003; Dornelas and Rodriguez, 2001 and 2004; Dornelas et al., 2004;). Additionally, there is a need to understand how the full diversity in flowering responses is generated. For example, *Arabidopsis* responds to photoperiod, but all ecotypes are long-day plants that flower earlier under long than short days, whereas many other species show the reverse response and many are not responsive to photoperiod at all (Mouradov et al., 2002; Izawa et al., 2003). Also, all *Arabidopsis* ecotypes are annual plants, and understanding the perennial habit will require a different model species. Thus, to understand the diversity in flowering responses, there is a need to search for conserved key elements on the regulatory pathways involved in the control of flowering time in other plant species.

We are interested in characterizing genes involved in the early stages of floral development in woody angiosperm trees of the genus *Eucalyptu*s. With this aim, we have used the sequences of the key proteins of the different developmental pathways involved in the regulation of flowering-time available from *Arabidopsis* as baits to search the *Eucalyptus* Genome Sequencing Project Consortium (FORESTS) database for expressed sequence tags (ESTs) showing sequence homology with known elements of flowering-time pathways. We have identified *Eucalyptus* sequences that codify putative conserved elements of the vernalization, photoperiod response, autonomous and GA-controlled flowering-time pathways. Additionally, we have undertaken an extensive *in silico* characterization of the putative *Eucalyptus* homologues of the *CONSTANS* gene family, which, in *Arabidopsis*, mediate the cross-talk between the circadian clock and the genes controlling reproductive meristem identity. The *Arabidopsis CO-like* (*COL*) gene family is proposed to encode proteins with two zinc fingers loosely related to those of GATA transcription factors, termed the B-Box (Putterill et al., 1995) and contains a carboxy-terminal domain called CCT because it is present in CO, COL, and TIMING OF CAB 1 (TOC1) proteins (Strayer et al., 2000; Robson et al., 2001). We expect that our results will contribute to further studies describing how these pathways function in species that respond differently to environmental conditions than does *Arabidopsis*.

MATERIAL AND METHODS

Searching Eucalyptus *ESTs homologs to* Arabidopsis *flowering-time genes:* The overall goal of this study was to retrieve from the FORESTS data set, *Eucalyptus* spp homologs to all genes described to be involved in the control of flowering time, according to the processes showed in figure 1. In order to achieve this, data mining in the FORESTS database was carried out using published plant gene sequences as baits, as well as keyword searches in the FORESTS home page (https://forests.esalq.usp.br). Plant gene sequences used as baits were retrieved from public gene databases (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) using their corresponding accession numbers or by the use of keyword-oriented searches (Mouradov et al., 2002; Izawa et al., 2003). Protein (deduced amino acid) sequences from the retrieved bait sequences were compared to *Eucalyptus* clustered EST

sequences using tBLASTn or a combination of different Blast algorithms (Altshul et al., 1997), with the BLOSUM62 scoring matrix, with a threshold of $e<10^{-10}$ for positive hits. The identity (in terms of donor cDNA library) and number of sequence read composition of each individual candidate cluster were checked to access their potential expression pattern.

For the results presented in table 1, we have obtained e-values using the BLASTp algorithm (Altshul et al., 1997) as described above and calculated the identity and the similarity, at the amino acid level, relative to the corresponding *Arabidopsis* putative homolog, within the extension of the successful sequence alignment produced by the pair-wise comparison.

In silico characterization of the Eucalyptus homologs belonging to the CONSTANS *gene family:* The *Arabidopsis CONSTANS* (*CO*) gene family codifies putative transcription factors defined by two conserved domains (Putterill et al., 1995; Griffiths et al., 2001). The first is a zinc finger region near the amino terminus that resembles B-boxes, which regulate protein-protein interactions in several animal transcription factors (Putterill et al., 1995). The second is a region of 43 amino acids near the carboxy terminus termed the CCT *(C*O, *C*O-like, *TOC*1) domain (Robson et al., 2001). We have identified *Eucalyptus* homologs to the *Arabidopsis CO* gene family, using the *Arabidopsis* sequences as baits and the BLAST algorithms (Altshul et al., 1997) as described above. Only comparisons that produced an e-value lower than e^{-50} were considered highly significant. When the obtained e-values were between e^{-50} and e^{-5}, a re-clusterization of all reads identified was performed using the CAP3 algorithm from the BioEdit Software (Hall, 1999). The novel cluster consensus sequences obtained were re-submitted to BLAST and frequently better e-values were obtained. To identify the presence of conserved domains in the deduced protein sequence, we analyzed them using the CDD algorithm (Marchler-Bauer et al., 2005).

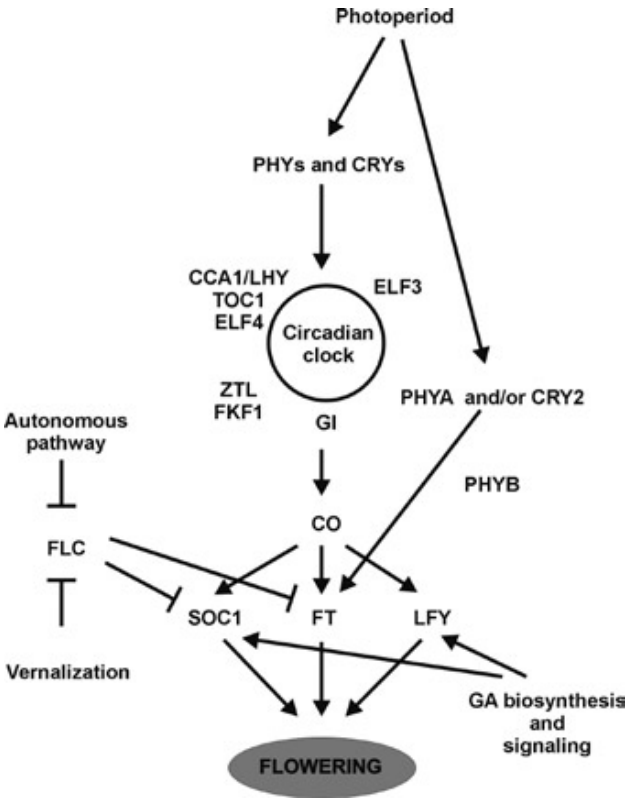

Figure 1. Overview of the relationships among the elements involved in the flowering-time pathways in the model plant *Arabidopsis thaliana* (after Mouradov et al., 2002; Izawa et al., 2003). The data underlying the model and the corresponding homologs in *Eucalyptus* are presented in table 1 and in the text. For abbreviations and gene names see table 1.

Comparative and phylogenetic analysis of CONSTANS *gene family homologs:* To examine the relationships between the *Eucalyptus CO*-like genes and their putative *Arabidopsis* homologs in more detail, their nucleotide and predicted peptide sequences were used to determine genetic distances and to construct phylogenetic trees. Because the middle regions of the genes were the most diverged, they could not be aligned with confidence. Therefore, neighbor-joining (NJ) and maximum parsimony (MP) trees were constructed using B-box (and CCT domain sequences when available) following the alignments obtained using the CLUSTALX software (Thompson et al., 1994). The alignments were eventually corrected by hand. Phylogenetic trees were obtained using parsimony and/or genetic distance calculations. Neighbour-joining (Saitou and Nei, 1987) and Bootstrap (with 1000 replicates) trees were built using the MEGA software (http://www.megasoftware.net).

RESULTS

Eucalyptus *ESTs homologs to* Arabidopsis *flowering-time genes:* Based on the systematic search in the FORESTS database using *Arabidopsis* sequences as baits, we have identified 536 *Eucalyptus* spp. EST sequences, organized in 21 clusters, representing putative *Eucalyptus* homologs to flowering-time genes. Some of these genes are required for the daylength response, and some encode regulatory proteins

specifically involved in the control of flowering, while others encode components of light signal transduction pathways or are involved in circadian clock function. A representation of the relationships among these processes is shown in figure 1 and the putative homologs of the key players in *Eucalyptus* are presented in table 1. The role of each of these elements in the flowering-time pathways and their implication for the understanding of the *Eucalyptus* flowering processes are presented in the Discussion section.

Comparative and phylogenetic analysis of Eucalyptus CONSTANS-*like gene family:* We have found 187 *Eucalyptus* EST sequences showing significant (e-value lower than e^{-5}) similarity to the *Arabidopsis COL* genes, by means of a combination of the tBLASTn and keyword search in the FORESTS database. When submitted to the CAP3 algorithm, these sequences were organized initially in 37 clusters. Their deduced amino acid sequences were submitted to BLASTp and the number of clusters showing significant (e-value lower than e^{-5}) similarity to the *Arabidopsis COL* genes was reduced to 20.

We thus restricted our analysis to *Eucalyptus* sequences showing the conserved B-box and CCT domains, according to the definition of the COL family provided by Griffiths et al., (2003). The number of *Eucalyptus* EST clusters encoding proteins significantly similar to those from the *Arabidopsis* COL family was then reduced to ten. We thus assumed that each of these clusters corresponds to a member of the *Eucalyptus COL* gene family (table 2).

As most of the CCT domain sequences are not available for the EgCOL proteins, we produced alignments of the predicted peptides of the conserved B-box region for all *Arabidopsis* AtCO and AtCOL proteins and their putative *Eucalyptus* homologs (figure 2A). The alignment showed that the *Eucalyptus* EgCO and EgCOL2-10 had consensus CO-like amino acid residues at the amino-terminus. Variation within the B-box domain suggested that the CO-like genes could be further subdivided. For example, distinctive amino acid residues in the B-boxes grouped *Arabidopsis* COL3 to COL5 peptides with EgCOL2 and EgCOL3 (figure 2A).

To further examine the relationship between the putative EgCOL homologs and their *Arabidopsis* counterparts, the sequence alignment shown in figure 2A was used to determine genetic distances and to construct phylogenetic trees. Therefore, neighbor-joining (figure 2B) and maximum parsimony trees were constructed, giving similar results (data not shown). This consistently grouped the proteins into three main clades (figure 2B). Group III genes comprised *Arabidopsis* and *Eucalyptus* proteins with two zinc finger domains, the second of which was diverged from the CO-type B-box. Group II genes comprised *Arabidopsis* and *Eucalyptus* proteins with a single B-box. Group I comprised the majority of *CO*-like genes and included the *Eucalyptus* EGCO, EgCOL2 and EgCOL3 proteins. There were subdivisions within Group I, but these had low bootstrap values (figure 2B). EgCOL7 had the most diverged B-Box domain and the phylogenetic tree placed it, and the related *Arabidopsis* proteins ATCOL16 and AtCOL6-7, within Group II. Other two *Eucalyptus* proteins, EgCOL5 and EgCOL6 were consistently divergent and were also placed in Group II, at the base of the tree, but in a different clade to which no *Arabidopsis* homologs could be assigned.

DISCUSSION

The apical meristem in most *Eucalyptus* species generally remains vegetative throughout their life cycle. Lateral meristems, arising in the leaf axils, may give rise to a leafy shoot or to an inflorescence (Dornelas et al., 2004), in response to inductive environmental conditions such as day length and temperature, as well concentrations of endogenous substances such as gibberellic acid (GA), if the tree is sufficiently mature (Moncur and Hasan, 1994; Southerton et al., 1998). Molecular and genetic approaches have identified genes required for the flowering-time control and their relationships were established in the model plant *Arabidopsis* (figure1). We have identified putative homologs of the key players in *Eucalyptus* (table 1). Additionally, we have characterized *in silico*, the putative *Eucalyptus* homologs to the *COL* gene family. In the following sections we discuss the role of each of these elements in the flowering-time pathways and their implication for the understanding of the *Eucalyptus* flowering process.

The photoperiod response pathway and the role of the *CONSTANS* gene family

One of the most important environmental factors controlling flowering time is the duration of the daily light period, or photoperiod. *Arabidopsis* is a facultative long-day plant, which flowers earlier under long days but eventually flowers under short days (Corbesier et al., 1996). The products of the genes *CONSTANS (CO)*, *CRYPTOCHROME2/FHA (CRY2)*, *GIGANTEA (GI)*, and *FT* are part of this long day–promoting pathway (Koornneef et al., 1991; Mouradov et al., 2002). *CO* is the latest acting of the known genes that is specific to this

A

```
                                  B-box 1                                                B-box 2
AtCOL2      CDTCRS-AACTVYCEADSAYLCTTCDARVHAANRVASRHERVRVCQSCESAPAA-----FLCKADAASLCTACDAEIHS---------ANPLARRHQRVPI
AtCOL1      CDTCRS-AACTVYCRADSAYLCSSCDAQVHAANRLASRHERVRVCQSCERAPAA-----FFCKADAASLCTTCDSEIHS---------ANPLARRHQRVPI
AtCO        CDTCRS-NACTVYCHADSAYLCMSCDAQVHSANRVASRHKRVRVCESCERAPAA-----FLCEADDASLCTACDSEVHS---------ANPLARRHQRVPI
EgCO        CDTCRS-AACTVYCRADLAYLCSSCDARIHAANRVASRHERVWVCEACERAPAA-----FLCKADAASLCTACDADIHS---------ANPLARRHHRVPI
AtCOL4      CDSCKS-ATAALYCRPDAAFLCLSCDSKVHAANKLASRHARVHMCEVCEQAPAH-----VTCKADAAALCVTCDRDIHS---------ANPLARRHERVPV
AtCOL3      CDSCKS-TAATLFCRADAAFLCGDCDGKIHTANKLASRHERVWLCEVCEQAPAH-----VTCKADAAALCVTCDRDIHS---------ANPLSRRHERVPI
AtCOL5      CDACKS-VTAAVFCRVDSAFLCIACDTRIHSFT----RHERVWVCEVCEQAPAA-----VTCKADAAALCVSCDADIHS---------ANPLASRHERVPV
EgCOL3      CDSCKT-AAAAVFCRADAAFLCLGCDARVHGAAKLSARHERVWVCEVCEQAPAA-----VTCKADAAALCVTCDADIHS---------ANPLARRHERVPV
EgCOL2      CDSCKS-ATAALFCRADSAFLCPSCDAKVHAANKLASRHARVVVCEVCEQSPAH-----VTCRADAASLCISCDRDIHS---------ANPVAQKHDRVPV
AtCO9       CDFCGE-QRSMVYCRSDAACLCLSCDRSVHSANALSKRHSRTLVCERCNAQPAT-----VRCVEERVSLCQNCDWSGHDNSNNNNSSSSTSPQQHKRQTI
AtCOL10     CDFCGE-QRSMVYCRSDAACLCLSCDRNVHSANALSKRHSRTLVCERCNAQPAS-----VRCSDERVSLCQNCDWSGHDGKN-------STTTSHHKRQTI
EgCOL9      CDFCGE-QRSTVYCRSDAACLCLSCDRNVHSANALSRRHSRTLLCERCNSQPAL-----VRCTEERVSLCQNCDWIGHG---------ASTTASGHKRQTI
EgCOL10     CEFCGE-QRSMVYCRSDAACLCLSCDRNVHSANALSRRHSRTLLCERCNSQPAS-----VICAEERVSLCQNCDWIGHG---------ASTSASTHERXTL
AtCOL11     CDFCGT-EKALIYCKSDSAKLCLNCDVNVHSANPLSQRHTRSLLCEKCSLQPTA-----VHCMNENVSLCQGCQNTASN----------CTGLGHRLQSL
AtCOL12     CDHCAT-SQALIYCKSDLAKLCLNCDVHVHSANPLSHRHIRSLICEKCFSQPAA-----IRCLDEKVSYCQGCHNHESN----------CSELGHRVQSL
EgCOL8      CEFCRA-LRPVIYCQADAAQLCLSCDAKVHSANALSSRHLRSLLCDSCQYRPAF-----VHCLDHKLYICRNCDRRIHN----------GSCQHRTQAI
AtCOL13     CDYCDS-SVALVYCKADSAKLCLACDKQVHVANQLFAKHFRSLLCDSCNESPSS-----LFCETERSVLCQNCDWQHHT----------ASSSLHSRRPF
EgCOL4      CDYCGG-TTALLYCRADTARLCFACDREVHATNPLFAKHNRSLLCDACDSASAS-----FFCEAERLLLCQNCDWDRHDGGG-----GGGGAAAVHNRRPV
AtCOL14     CEFCGE-RTAVLFCRADTAKLCLPCDQHVHSANLLSRKHVRSQICDNCSKEPVS-----VRCFTDNLVLCQECDWDVHG---------SCSSSATHERSAV
AtCOL15     CDFCGE-RTAVLFCRADTAKLCLPCDQQVHTANLLSRKHVRSQICDNCGNEPVS-----VRCFTDNLILCQECDWDVHG---------SCSVSDAHVRSAV
AtCOL6      CDSCVK-RRARWYCAADDAFLCHACDGSVHSANPLARRHERVRLKSASAGKYRH-----ASPPHQ-ATWHQGFTRKARTPRG----GKKSHTMVFHDLVPE
AtCOL16     CDSCVK-RRARWYCAADDAFLCQSCDSLVHSANPLARRHERVRLKTASPAVVKHSNHSSASPPHEVATWHHGFTRKARTPRGS---GKKNNSSIFHDLVPD
AtCOL7      CDACMKRSRASWYCPADDAFLCQSCDASIHSANHLAKRHERVRLQS---SSPTE-----TA-DKTTSVWYEGFRRKARTPRS-----KSCAFEKLLQIESN
AtCOL8      CELCLN-KHAVWYCASDDAFLCHVCDESVHSANHVATKHERVCLRTNEISNDVR-----GG-TTLTSVWHSGFRRKARTPRS-----R-YEKKPQQKIDDE
EgCOL7      CDSCLR-RRARWYCTADDAFLCQGCDSSVHSANQLARRHERVRPWDGLVVVFQDR---LWRRCQVAVVSPTPICRQFSSSG------VARRVHPQGPDAPD
EgCOL6      CELCNN--LAATYCGSDGANLCWDCDSSVHGANFLVAKHSRNLLCRSCQGPTPW----NASGRRLPPTV-SVCEKCVDCR--------GCDFRGGSARSSC
EgCOL5      CELCGS--RACLYCQADDAYLCRQCDRWVHEANFLAQRHIRCLLCTSCQRLTQR----YVLGPSLVVVLPSMVESCADG---------EDSSDAQDKSSE
            *: *       :*  * * **  **  :*       :* *
```

B

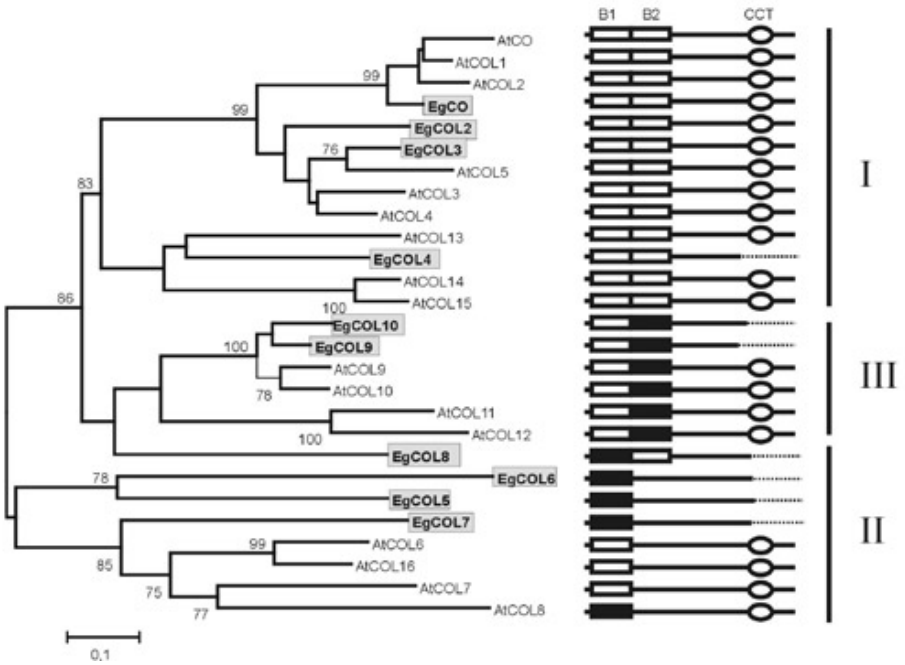

Figure 2. Characterization of the putative *CONSTANS* gene family in *Eucalyptus*. **A**. Alignment of predicted peptides of *Eucalytus CO*-like putative homologs and related genes from *Arabidopsis*. The region of the proteins aligned corresponds to the conserved B-box domains of the CO-like family (Robson et al., 2001; Griffiths et al., 2003). The asterisks and double mark signs under the alignment identify the fully conserved and partially conserved residues, respectively. **B**. Phylogenetic analysis of *CO*-like genes. A Neighbour-Joining tree was built based on the B-box domain alignment shown in A. The *Eucalyptus* names of the deduced protein sequences are given in shaded boxes. Genetic distances are shown at the given scale. Bootstrap values from 1,000 replicates were used to assess the robustness of the trees. Only bootstrap values above 75% are shown. The domain structures of each protein is also shown to the right side of their names. B1 and B2 are *CO*-like B-boxes (white rectangles) or derived zinc finger domains (solid rectangles). CCT is the conserved CCT carboxy-terminus domain (Robson et al., 2001). The dotted lines represent incomplete sequences. The *Eucalyptus* EST clusters corresponding to the respective gene names are listed in table 2. *Arabidopsis* MIPS codes are as follows: AtCO (At5g15840); AtCOL1 (At5g15850); AtCOL2 (At3g02380); AtCOL3 (At2g24790); AtCOL4 (At5g24930); AtCOL5 (At5g57660); AtCOL6 (At1g68520); AtCOL7 (At1g73870); AtCOL8 (At1g49130); AtCOL9 (At3g07650); AtCOL10 (AB023039); AtCOL11 (At4g15250); AtCOL12 (At3g21880); AtCOL13 (At2g47890); AtCOL14 (At2g33500); AtCOL15 (At1g28050); AtCOL16 (At1g25440).

pathway; all the other genes also act in other pathways or have more general effects (figure 1). *FT* acts downstream of *CO* and also in other pathways (Kardailsky et al., 1999; Kobayashi et al., 1999; Onouchi et al., 2000), whereas *GI* and *CRY2* act upstream of *CO* (Guo et al., 1998; Suarez-Lopez et al., 2001).

The *CO* gene was proposed to encode a protein with two zinc fingers loosely related to those of GATA transcription factors (Putterill et al., 1995) and contains a carboxy-terminal domain called CCT because it is present in CO, CO-like (COL), and TIMING OF CAB 1 (TOC1) proteins (Strayer et al., 2000; Robson et al., 2001). Comparison of 16 *Arabidopsis COL* genes that encode proteins sharing the B-box and CCT domains (Robson et al., 2001) demonstrated that the B-box is a type of zinc finger identified in animal proteins and believed to mediate protein–protein interactions (Borden, 1998). In an analysis of the *Arabidopsis CO*-like genes sequences by Lagercrantz and Axelsson (2000), they concluded that this gene family evolves rapidly, particularly

Table 1. *Eucalyptus* ESTs that share homology to flowering-time genes of *Arabidopsis*.

Category	*Arabidopsis*[a]	MIPS code	*Eucalyptus* [b]	e-value[c]	ID/SIM[d]	ext[e]
Photoreceptor	PHYA	At1g09570	EGUTSL1041B03	e-170	87/93	70
	PHYB	At2g18790	EGBGLV3212C11	e-102	82/84	88
	PHYC	At5g35840	EGEPCL2212D06	e-102	84/85	76
	CRY1	At4g08920	EGEQBK1001E10	0	89/92	89
	CRY2	At1g04400	EGEQST2200A05	e-136	75/81	76
Circadian clock	CCA1	At2g46830	EGCCRT6011H06	2e-50	62/74	68
	LHY	At1g01060	EGCCLV2228F02	e-71	78/83	69
			EGCCRT6011H06	6e-61	85/89	79
	GI	At1g22770	EGEQST2001F03	0	89/94	98
	TOC1	At5g61380	EGACRT3322D08	e-102	87/92	79
	ELF3	At2g25930	EGSBST2083B09	3e-33	74/81	89
	ZTL	At5g57360	EGEQSL1007C08	2e-64	86/91	87
	LKP2	At2g18910	EGJMST6139E10	e-114	86/89	88
Circadian clock	FKF1	At1g68050	EGEQBK1002B03	0	89/94	98
mediator	CO	At5g15840	EGCBSL7207A07	3e-84	61/70	73
Floral pathway	FT	At1g65480	EGUTFB1296G08	3e-78	68/74	84
integrator	SOC1	At2g45660	EGAGLV2211H06	6e-69	69/72	78
	LFY	At5g61850	not found (see text)			
Vernalization pathway	FLC	At4g18280	EGEQFB1200B04	2e-45	69/75	76
Chromatin-related	EMF2	At5g51230	EGCEST6029D06	e-105	89/94	81
	FIE	At3g20740	EGJERT6044F11	e-73	86/91	76
	LHP	At5g17690	EGRFRT3020D04	2e-57	68/71	72

a. Abbreviations: CCA1: CIRCADIAN CLOCK ASSOCIATED 1; CK2: casein kinase2; CO: CONSTANS; CRY: CRYPTOCHROME; ELF3: EARLY FLOWERING3; EMF2: EMBRYONIC FLOWERING2; FIE: FERTILIZATION INDEPENDENT ENDOSPERM; FKF1: FLAVIN-BINDING, KELCH-REPEATS, F-BOX1; FLC: FLOWERING LOCUS C; FT: FLOWERING LOCUS T; LFY: LEAFY; LHP1: LIKE HETEROCHROMATIN PROTEIN1; LHY: LATE ELONGATED HYPOCOTYL; GI: GIGANTEA; PHY: PHYTOCHROME; SOC1: SUPPRESSOR OF OVEREXPRESSION OF CO1; TOC1: TIMING OF CAB EXPRESSION1.
b. When more than one putative homolog is present in *Eucalyptus*, all possible clusters are presented.
c. Using the BLASTp algorithm (Altschul et al., 1997).
d. ID= identity; SIM= similarity; both based on the amino acid sequence, relative to the putative *Arabidopsis* homologs.
e. ext= extension of the successful alignment including eventual insertion/deletion events.

in the middle regions. Their analysis focused on B-box sequences only, and they included the *Arabidopsis STO (SALT TOLERANC*E) gene. *ST*O-like genes have B-boxes but no CCT domain. Additionally, their analysis included in the *COL* family, the related *ZIM* gene, which contains an additional ZIM motif. This short motif is found in a variety of plant transcription factors that contain GATA domains and its conserved amino acids form the pattern TIFF/YXG (Lagercrantz and Axelsson, 2000; Griffiths et al., 2003).

We have identified ten members of the *COL* gene family in *Eucalyptus* (table 2; figure 2). Each member was represented by a single EST sequence (i.e. a singleton, e.g. *EgCOL10*) or by the re-clusterization of more than one cluster, represented by more than 40 sequences (e.g. *EgCOL2*). Among the sequences that composed the clusters, we found many from cDNA libraries derived from tissues of *E. camaldulensis*, *E. urophylla* and *E. saligna*, additionally to those derived from *E. grandis*. We could identify COL members coding for all three conserved groups of COL proteins (figure 2B). These three groups were identified previously and are thought to have evolved prior to the divergence of monocots and dicots

Table 2. *Eucalyptus* putative homologs to the *CONSTANS*-like genes of *Arabidopsis*.

Name	Cluster [a]	e-value[b]	ID/SIM[c]	ext[d]
EgCO	EGCBSL7207A07	3e-84	61/70	73
EgCOL2	EQSL1007C08			
	EQFB1202G12	5e-85	53/62	100
EgCOL3	EQST7201G10			
	UTBK1010C01	6e-79	44/54	100
EgCOL4	EGBFFB1044A07	5e-34	55/71	32
EgCOL5	EGACST6260A06	3e-18	52/64	68
EgCOL6	EGABSL1083B09	1e-24	68/76	33
EgCOL7	EGJFLV3236G03			
	EGQHSL1104H01	2e-32	30/40	78
EgCOL8	EGJEFB1029F07	3e-32	38/54	44
EgCOL9	EGBFLV2250A04	1e-107	52/61	100
EgCOL10	EGCBRT6017H05	3e-54	73/84	36

a. When more than one putative homolog is present in *Eucalyptus*, all possible clusters are presented. In this case, the corresponding sequences were re-aligned with CAP3 and the novel cluster consensus sequence was used in the comparison.
b. Using the BLASTp algorithm (Altschul et al., 1997).
c. ID= identity; SIM= similarity; both given in percentage and based on the amino acid sequence, relative to the closest *Arabidopsis* CO or CO-like protein.
d. ext= extension percentage of the successful alignment including eventual insertion/deletion events.

(Griffiths et al., 2003). COL proteins belonging to different groups are expected to perform distinct biological roles, although only the *CO* role in controlling flowering time has been studied in detail (Onouchi et al., 2000; Suarez-Lopez et al., 2001; Griffith et al., 2003).

When constitutively expressed in *Arabidopsis*, *CO* induces early flowering and loss of photoperiod sensitivity (Onouchi et al., 2000). Accordingly, mutations in the *FT*, *SUPPRESSOR OF OVEREXPRESSION OF CO1 (SOC1*, also called *AGAMOUS-LIKE 20* [*AGL2*0]) genes partially suppress the early flowering phenotype of *35S*::*CO* (Onouchi et al., 2000).

The *CO* transcript level shows a diurnal peak in *Arabidopsis* plants grown in long days. This peak is narrower in short days, ending 4 h earlier (Suarez-Lopez et al., 2001). Plants entrained in long days show a circadian rhythm in *CO* transcript level when transferred to continuous light (LL), indicating that this rhythm is controlled by the circadian clock (Suarez-Lopez et al., 2001). *FT* is an early target of *CO* (Samach et al., 2000), and its transcript level also follows a circadian rhythm (Suarez-Lopez et al., 2001). This peak is absent in the *co* mutant. In addition, the circadian clock–related mutants, *LATE ELONGATED HYPOCOTYL (LHY)*, *GI*, and *EARLY FLOWERING3 (ELF3)*, show an altered rhythm in *CO* transcript level that correlates with their flowering phenotype. Overall, *CO* appears to mediate between the circadian clock and the flowering-time gene *FT* (Suarez-Lopez et al., 2001). Candidates to photoreceptors that might be involved in the post-transcriptional activation of CO protein under long days are CRY2 and PHYTOCHROME A (PHYA).

The circadian clock elements

The molecular mechanism that generates the circadian rhythms is often described as being composed of three interrelated parts. The first one is an input pathway that synchronize the clock mechanism to daily cycles of light and dark; the second one is a central oscillator that generates the 24-h time-keeping mechanism; and the third one is a group of output pathways that regulate particular processes (Roenneberg and Merrow, 2000; McClung, 2001). The control of flowering via *CO* and *FT* represents one such output pathway (Suarez-Lopez et al., 2001). Another output pathway, represented by the *CHLOROPHYLL A/B BINDING PROTEIN2 (CAB*2) gene, has enabled detailed analysis of clock regulation. The *Arabidopsis timing of CAB (toc)* mutants flowered early in short days (Kreps and Simon,

1997; Somers et al., 1998). The TOC1 protein shows two characteristic domains (Strayer et al., 2000). The first of these, at the N terminus, is similar to the receiver domain of response regulators from two-component signal transduction systems. The second motif, at the C terminus, is the CCT domain, shared with the CO family of transcriptional regulators. The consensus cluster sequence encoding the putative *Eucalyptus* TOC1 homolog was formed by an alignment of sequences derived from cDNA libraries made with *E. grandis*, *E. saligna* and *E. urophylla* tissues, indicating that *TOC* homologs may be conserved among *Eucalyptus* species.

Other likely components of the oscillator are the *CIRCADIAN CLOCK ASSOCIATED1 (CCA1*) and *LATE ELONGATED HYPOCOTYL (LHY*) genes, which are also involved in the photoperiodic induction of flowering. These genes encode highly similar single Myb domain DNA binding proteins and are each regulated by the circadian clock, with a peak in their expression soon after dawn (Schaffer et al., 1998; Wang and Tobin, 1998). The overexpression of *LHY* or *CCA1 in Arabidopsis* disrupts the circadian rhythm and causes late flowering under long-day conditions (Schaffer et al., 1998; Wang and Tobin, 1998; Mizoguchi et al., 2002). Recent reports, nevertheless, suggest that *LHY* overexpression may not disrupt rhythms in expression of the circadian clock–regulated gene *EARLY FLOWERING 3 (ELF*3), which would be an unexpected exception to the effect of *LHY* (Hicks et al., 2001). *ELF3* encodes a nuclear protein proposed to act as a transcriptional activator (Hicks et al., 2001; Liu et al., 2001).

According to the current molecular model (figure 1), LHY and CCA1 feed back to repress their own expression, and as their protein levels fall, the expression of *TOC1* rises. TOC1, in turn, promotes the expression of *LHY* and *CCA1*, so initiating another cycle (Alabadi et al., 2001). All central elements of the model have been recently described to occur in rice (Izawa et al., 2003) and have shown to have putative homologs in *Eucalyptus* (table 1), suggesting the conservation of the molecular mechanism regulating the circadian clock in angiosperms

Synchronizing the molecular circadian clock

Circadian rhythms must be synchronized with the daily rhythm of light and dark and of temperature (Roennenberg and Merrow, 2000). The phase of circadian rhythms can therefore be reset by environmental conditions such as light and temperature that fluctuate during day/night cycles. The photoreceptors PHYA, PHYB, CRY1, and CRY2 all influence clock entrainment under specific light conditions (McClung, 2001; Yanovsky et al., 2000). Mutations in an *Arabidopsis* PHYB-interacting protein, ZEITLUPE (ZTL), also have dramatic effects on clock function and flowering time (Somers et al., 2000). Constitutive overexpression of *ZTL* gave a late-flowering phenotype in *Arabidopsis* plants growing in long-days (Nelson et al., 2000). The ZTL protein contains a PAS domain, an F-box, and six repeated kelch motifs (Adams et al., 2000). The PAS domain mediates protein–protein interactions and has been found in a group of blue light photoreceptors (Briggs and Huala, 1999). Accordingly, protein–protein interactions have been described between ZTL and PHYB and between ZTL and CRY1 (Jarillo et al., 2001). *ZTL* is a member of a family of three genes in *Arabidopsis* that also includes *FKF1* and *LKP2* (Somers et al., 2000). Putative homologs for all three genes have been found in *Eucalyptus* (table 1.). The *fkf1* mutant shares the late-flowering phenotype of *ztl*. The late-flowering phenotype of *fkf1* can be rescued by vernalization and gibberellin (GA) application, suggesting that *fkf1* may be associated with the autonomous flowering pathway (Nelson et al., 2000).

Mutant alleles of the *Arabidopsis* gene *GIGANTEA (GI)* also cause alterations in period length of gene expression rhythms. The GI protein contains six putative transmembrane domains (Fowler et al., 1999; Park et al., 1999; Huq et al., 2000), suggesting a membrane protein, but GUS:GI and GFP:GI fusion proteins were targeted to the nucleus (Park et al., 1999). Three *Eucalyptus* EST clusters span the *GI* gene, but the one containing the most extensive 5′ region is EGBMST2013E01 (table 1).

The vernalization response pathway

Exposure to low temperatures for several weeks will often accelerate flowering in *Arabidopsis* (Michaels and Amasino, 2000). The genetic control of vernalization was related to two different loci, *FLOWERING LOCUS C* (*FLC*) and *FRIGIDA* (*FRI*; Burn et al., 1993; Lee and Amasino, 1993; Clarke and Dean, 1994). *FLC* encodes a MADS box transcription factor (Sheldon et al., 1999; Michaels and Amasino, 2000) that is a repressor of flowering, and high-level expression of *FLC* correlates with the vernalization requirement of winter annual *Arabidopsis* varieties (Michaels and Amasino, 2000; Sheldon et al., 1999; 2000). The other locus at which dominant alleles confer a vernalization requirement is *FRI*. The product of this gene somehow increases *FLC* mRNA

abundance (Sheldon et al., 1999; Michaels and Amasino, 2000). The biochemical function of FRI protein is unknown, but it contains coiled-coil domains that may be involved in protein–protein interactions (Johanson et al., 2000). FRI homologs were not clearly identified in rice (Izawa et al., 2003). Further, no orthologues of the recently identified *VERNALIZATION1* (*VRN1*) and *VRN2* (Gendall et al., 2001; Levy et al., 2002) have been found in rice (Izawa et al., 2003), suggesting that vernalization-related genes have been lost from the rice genome during evolution, probably due to the adaptation to tropical regions. Similarly, no *Eucalyptus* ESTs were found that showed significant similarities to *FRI, VRN1* or *VRN2*.

Cloning of *VRN2* showed that its protein sequence is related to a *Drosophila* Polycomb group protein, Su(z)12, that regulates gene expression by modifying chromatin structure (Birve et al., 2001; Gendall et al., 2001). This suggests that there could be some molecular mechanism relating chromatin structure and flowering time control.

The relation between chromatin structure and flowering control

Arabidopsis mutants have been described that flower without forming any adult leaves, advancing directly from embryonic development to flowering (Sung et al., 1992; Yang et al., 1995; Kinoshita et al., 2001). The class of mutants named *embryonic flower* (*emf1*; Sung et al., 1992 and *emf2* Yang et al., 1995) form a reduced inflorescence and abnormal flowers that lack petals without first forming any rosette (Sung et al., 1992; Yang et al., 1995). The predicted EMF2 protein contains a zinc finger, an N-terminal basic domain and a C-terminal acidic domain (Yoshida et al., 2001). This protein belongs to the same family as *VRN2* and also showed similarity to the *Drosophila* Polycomb group protein Su(z)12. The EMF2 protein may then act as part of a protein complex that during embryo and seedling development represses the expression of genes that promote reproductive development. The importance of Polycomb group genes in repressing reproductive development was also emphasized by the demonstration that another *Arabidopsis* Polycomb group gene, *FERTILIZATION INDEPENDENT ENDOSPERM (FIE*), is related to the EXTRA SEX COMBS protein of *Drosophila* (Kinoshita et al., 2001). Loss-of-function *fie* alleles were originally described because they allow partial endosperm formation prior to fertilization. However, the effect of the mutation in the seedling or adult was not described, because maternal *FIE* alleles are essential for embryo development (Ohad et al., 1999). By expressing *FIE* from a defective *FIE* promoter that allows expression during seed development but not in the germinated seedling, it was demonstrated that *FIE* is required for the repression of flowering in the seedling (Kinoshita et al., 2001). Plants homozygous for *fie* and expressing *FIE* from such a defective promoter initiate reproductive development as seedlings and resemble *emf* mutants. Putative homologs for both EMF2 and FIE were found among *Eucalyptus* ESTs in the FORESTS database (table 1).

The GA pathway

The growth regulator gibberellic acid (GA) promotes flowering in *Arabidopsis* (Wilson et al., 1992; Putterill, et al., 1995; Blazquez et al., 1998). On the other hand, exogenous application of paclobutrazol reduced the concentration of endogenous GA in apical tissues of different *Eucalyptus* species and enhanced the reproductive activity of grafted trees (Moncur and Hasan, 1994), suggesting that in *Eucalyptus*, high concentrations of GA inhibits the flowering process. One of the ways in which GAs promote flowering is by increasing the transcriptional activity of the floral meristem identity gene *LEAFY* (*LFY*). Overexpression of *LFY* also restores flowering of *ga1-3* in short days (Blazquez et al., 1998). No *Eucalyptus* EST in the FOREST database showed significant similarity to LFY. Nevertheless, LFY orthologs have been identified in *E. grandis* (Dornelas et al., 2004) and in *E. globulus* (Southerton et al., 1998), indicating that the GA-regulated flowering pathway mediated by LFY may be conserved in *Eucalyptus*.

Integration of flowering pathways

Separate genetic pathways regulate flowering in response to different environmental signals, but these pathways eventually converge to regulate the expression of the same downstream genes. For example, all of the flowering-time pathways ultimately lead to the transcriptional activation of the same set of floral identity genes that act within the floral primordia (Pineiro and Coupland, 1998). *LFY* is the earliest of the known floral identity genes to be expressed, and directly activates later genes (Wagner et al., 1999). Plants carrying fusions of the *LFY* promoter to the *GUS* marker gene were used to demonstrate that *LFY* expression responds both to the long-day flowering pathway and to GA. Thus the GA and long-day pathways converge on the *LFY* promoter, rather than both pathways activating an earlier acting gene that in turn increases the expression

of *LFY* (Blazquez and Weigel, 2000). Convergence of the long-day and autonomous pathways has also been studied. These pathways are clearly separate upstream *FLC* and *CO* genes (figure 1). Nevertheless, flowering-time genes that act downstream of both *CO* and *FLC* have been identified. The *SOC1* (or *AGL20*) gene that encodes a MADS box protein is both activated by CO and repressed by FLC, suggesting that the pathways represented by these genes converge on *SOC1* (Borner et al., 2000; Lee et al., 2000; Samach et al., 2000; Michaels and Amasino, 2001). *SOC1* is also regulated by GA, and therefore is a common target of all three flowering pathways (Borner et al., 2000). Expression of *FT*, another flowering-time gene regulated by more than one pathway, is reduced by mutations that impair the function of the long-day and autonomous pathways and activated by overexpression of *CO*, indicating that *FT* acts downstream of both of these pathways (Samach et al., 2000). The relationship between *FT* and floral meristem identity gene expression has been studied at the genetic and molecular levels (Nilsson et al., 1998). *FT* is a member of a small gene family in *Arabidopsis* that also contains *TERMINAL FLOWER 1 (TFL1)*. The exact biochemical function of this family of proteins in plants is not known, nevertheless both *FT* and *TFL1* putative homologs were found in *Eucalyptus* (table1). The large conservation of all key regulators of the flowering-time pathways between *Arabidopsis* and *Eucalyptus* suggests that most of the molecular mechanism regulating the flowering process may be evolutionarily conserved among Myrtaceae and Brassicaceae, despite their different flowering strategies.

Conclusions and Perspectives

Genetic analysis in *Arabidopsis* has enabled the isolation of genes that control flowering time and the identification of interacting pathways that promote flowering in response to different environmental conditions. However, the function of key pathway components is likely to be altered in other herbaceous as well as woody plants to change their target genes or the activity of the protein complexes in which they act. These changes may explain, for example, why GA promotes flowering in *Arabidopsis* but represses flowering in *Eucalyptus* and some short-day plants (Moncur and Hasan, 1994; Izawa et al., 2003). Thus, there are flowering strategies for which *Arabidopsis* may be a less-effective model. Accordingly, although *Arabidopsis* genes can be used to drive very early flowering of trees and to overcome developmental delays of flowering (Weigel and Nilsson, 1995; Peña et al., 2001), studies of an annual plant such as *Arabidopsis* will probably not reveal the molecular mechanisms underlying many of the unique features of perennial plants (Battey, 2000). Thus, we hope that the identification of putative homologs of the key genes involved in the control of the transition to flowering in *Eucalyptus* presented in this article, as well as future studies on their expression patterns and on their manipulated expression in transgenic plants, will enable comparisons of the molecular mechanisms that regulate the floral transition in higher plants.

Acknowledgments: The authors acknowledge FAPESP, CAPES and CNPq for financial support.

REFERENCES

Adams J, Kelso R, Cooley L (2000) The kelch repeat superfamily of proteins: Propellers of cell function. Trends Cell Biol. 10:17–24.

Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Mas P, Kay SA (2001) Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. Science 293:880–883.

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

Araki T (2001) Transition from vegetative to reproductive phase. Curr. Opin. Plant Biol. 4:63–68.

Bagnall DJ, King RW, Whitelam GC, Boylan MT, Wagner D, Quail PH (1995) Flowering responses to altered expression of phytochrome in mutants and transgenic lines of *Arabidopsis thaliana* (L) Heynh. Plant Physiol. 108: 1495–1503.

Battey NH (2000) Aspects of seasonality. J. Exp. Bot. 51: 1769– 1780.

Birve A, Sengupta AK, Beuchle D, Larsson J, Kennison JA, Rasmuson-Lestander A, Muller J (2001) Su(z)12 a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants. Development 128:3371–3379.

Blazquez MA, Weigel D (2000) Integration of floral inductive signals in *Arabidopsis.* Nature 404:889–892.

Blazquez MA, Green R, Nilsson O, Sussman MR, Weigel D (1998) Gibberellins promote flowering of *Arabidopsis* by activating the LEAFY promoter. Plant Cell 10:791–800.

Borden KLB (1998) RING fingers and B-boxes: Zinc-binding protein-protein interaction domains. Biochem. Cell Biol. 76:351–358.

Borner R, Kampmann G, Chandler J, Gleissner R, Wisman E, Apel K, Melzer S (2000) A MADS domain gene involved in the transition to flowering in *Arabidopsis.* Plant J. 24:591–599.

Briggs WR, Huala E (1999) Blue-light photoreceptors in higher plants. Annu. Rev. Cell Dev. Biol. 15:33–62.

Burn JE, Bagnall DJ, Metzger JD, Dennis ES, Peacock WJ (1993) DNA methylation vernalization and the initiation of flowering. Proc. Natl. Acad. Sci. USA 90:287–291.

Clarke JH, Dean C (1994) Mapping FRI a locus controlling flowering time and vernalization response in *Arabidopsis thaliana.* Mol. Gen. Genet. 242:81–89.

Corbesier L, Gadisseur I, Silvestre G, Jacqmard A, Bernier G (1996) Design in *Arabidopsis thaliana* of a synchronous system of floral induction by one long day. Plant J. 9:947–952.

Dornelas MC, Rodriguez APM (2001) A genomic approach to elucidating grass flower development. Gen. Mol. Biol. 24:69-76.

Dornelas MC, Rodriguez APM (2004) Identification of differentially expressed genes during reproductive development in sugarcane (*Saccharum* sp) by the analysis of expressed sequence tags. Flowering Newsletter 37:40-45.

Dornelas MC, Amaral WAN, Rodriguez APM (2004) *EgLFY*, the *Eucalyptus grandis* homolog of the *Arabidopsi*s gene *LFY* is expressed in reproductive and vegetative tissues. Braz. J. Plant Physiol. 16:105-114.

Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Coupland G, Putterill J (1999) GIGANTEA: A circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. EMBO J. 18:4679–4688.

Gendall AR, Levy YY, Wilson A, Dean C (2001) The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in *Arabidopsis.* Cell 107:525-535.

Griffiths S, Dunford RD, Coupland G, Laurie DA (2003) The evolution of the *CONSTANS*-like gene families in barley, rice and *Arabidopsis*. Plant Physiol. 131:1855-1867.

Guo HW, Duong H, Ma N, Lin CT (1999) The *Arabidopsis* blue light receptor cryptochrome 2 is a nuclear protein regulated by a blue light-dependent post-transcriptional mechanism. Plant J. 19:279–287.

Guo HW, Yang WY, Mockler TC, Lin CT (1998) Regulations of flowering time by *Arabidopsis* photoreceptors. Science 279:1360–1363.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41:95-98.

Hicks KA, Albertson TM, Wagner DR (2001) EARLY FLOWERING3 encodes a novel protein that regulates circadian clock function and flowering in *Arabidopsis.* Plant Cell 13:1281– 1292.

Huq E, Tepperman JM, Quail PH (2000) GIGANTEA is a nuclear protein involved in phytochrome signaling in *Arabidopsis.* Proc. Natl. Acad. Sci. USA 97:9789–9794.

Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis.* Curr. Opin. Plant Biol. 6:113-120.

Jarillo JA, Capel J, Tang R-H, Yang H-Q, Alonso JM, Ecker JR, Cashmore AR (2001) An *Arabidopsis* circadian clock component interacts with both CRY1 and phyB. Nature 410:487–490.

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of FRIGIDA a major determinant of natural variation in *Arabidopsis* flowering time. Science 290:344–347.

Johnson E, Bradley M, Harberd NP, Whitelam GC (1994) Photoresponses of light-grown phyA mutants of *Arabidopsis*: Phytochrome A is required for the perception of daylength extensions. Plant Physiol. 105:141–149.

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D (1999) Activation tagging of the floral inducer FT. Science 286: 1962–1965.

Kinoshita T, Harada JJ, Goldberg RB, Fischer RL (2001) Polycomb repression of flowering during early plant development. Proc. Natl. Acad. Sci. USA 98:14156–14161.

Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960–1962.

Koornneef M, Blankestijn-de-Vries H, Hanhart C, Soppe W, Peeters T (1994) The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg erecta wild-type. Plant J. 6: 911–919.

Koornneef M, Hanhart CJ, Van Der Veen JH (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana.* Mol. Gen. Genet. 229:57–66.

Kreps JA, Simon AE (1997) Environmental and genetic effects on circadian clock-regulated gene expression in *Arabidopsis.* Plant Cell 9:297–304.

Lagercrantz U, Axelsson T (2000) Rapid evolution of the family of *CONSTANS* like genes in plants. Mol. Biol. Evol. 17: 1499–1507.

Lee H, Suh S-S, Park E, Cho E, Ahn JH, Kim S-G, Lee JS, Kwon YM, Lee I (2000) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis.* Genes Dev. 14:2366–2376.

Lee IAB, Amasino R (1993) Analysis of naturally occurring late flowering in *Arabidopsis thaliana.* Mol. Gen. Genet. 237:171–176.

Lee Y, Lloyd AM, Roux SJ (1999) Antisense expression of the CK2 alpha-subunit gene in *Arabidopsis* Effects on light-regulated gene expression and plant growth Plant Physiol. 119:989–1000.

Levy YY, Dean C (1998) Control of flowering time Curr Opin Plant Biol 1:49-54

Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C (2002) Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. Science 297:243-246.

Liu XL, Covington MF, Fankhauser C, Chory J, Wanger DR (2001) ELF3 encodes a circadian clock-regulated nuclear protein that functions in an *Arabidopsis* PHYB signal transduction pathway. Plant Cell 13:1293–1304.

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH (2005) CDD: a Conserved Domain Database for protein classification. Nucl. Acids Res. 33:192-196.

McClung CR (2001) Circadian rhythms in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 52:139–162.

Michaels SD, Amasino RM (2000) Memories of winter: Vernalization and the competence to flower. Plant Cell Environ. 23:1145–1153.

Michaels SD, Amasino RM (2001) Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of

FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. Plant Cell 13:935–941.

Mizoguchi T, Wheatley K, Hanzawa Y, Wright L, Mizoguchi M, Song H-R, Carré IA, Coupland G (2002) LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in *Arabidopsis.* Dev. Cell 2:629-41.

Moncur MW, Hasan O (1994) Floral induction in *Eucalyptus nitens.* Tree Physiol. 14:1303-1312.

Mouradov A, Cremer F, Coupland G (2002) Control of flowering time: interacting pathways as a basis for diversity. Plant Cell Supplement: S11-S130.

Nelson DC, Lasswell J, Rogg LE, Cohen MA, Bartel B (2000) FKF1 a clock-controlled gene that regulates the transition to flowering in *Arabidopsis.* Cell 101:331–340.

Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL (1999) Mutations in FIE a WD polycomb group gene allow endosperm development without fertilization. Plant Cell 11:407–415.

Onouchi H, Igeno MI, Perilleux C, Graves K, Coupland G (2000) Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among *Arabidopsis* flowering-time genes. Plant Cell 12:885–900.

Park DH, Somers DE, Kim YS, Choy YH, Lim HK, Soh MS, Kim HJ, Kay SA, Nam HG (1999) Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis GIGANTEA* gene. Science 285:1579–1582.

Peña L, Martin-Trillo M, Juarez J, Pina JA, Navarro L, Martinez-Zapater JM (2001) Constitutive expression of *Arabidopsis LEAFY* or *APETALA1* genes in citrus reduces their generation time. Nature Biotechnol. 19:263–267.

Pineiro M, Coupland G (1998) The control of flowering time and floral identity in *Arabidopsis.* Plant Physiol. 117:1–8

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80:847–857.

Reeves PH, Coupland G (2000) Response of plant development to environment: Control of flowering by daylength and temperature. Curr. Opin. Plant Biol. 3:37–42.

Robson F, Costa MMR, Hepworth S, Vizir I, Pineiro M, Reeves PH, Putterill J, Coupland G (2001) Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants. Plant J. 28:619–631.

Roenneberg T, Merrow M (2000) Circadian clocks: Omnes viae Romam ducunt. Curr Biol 10:R742–R745.

Saitou N, Nei M (1987) The neighbour joining method: a new method for reconstructing phylogenetic trees. Molec. Biol. Evol. 4:406-425.

Samach A, Coupland G (2000) Time measurement and the control of flowering in plants. Bioessays 22 38–47.

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G (2000) Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis.* Science 288:1613–1616.

Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carre IA, Coupland G (1998) The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. Cell 93:1219–1229.

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES (1999) The FLF MADS box gene: A repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. Plant Cell 11:445–458.

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: The central role of FLOWERING LOCUS C (FLC). Proc. Natl. Acad. Sci. USA 97:3753–3758.

Simpson GG, Dean C (2002) *Arabidopsis* the Rosetta stone of flowering time? Science 296:285-289.

Somers DE, Schultz TF, Milnamow M, Kay SA (2000) ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis.* Cell 101 319–329.

Somers DE, Webb AAR, Pearson M, Kay SA (1998b) The short-period mutant *toc1–1* alters circadian clock regulation of multiple outputs throughout development in *Arabidopsis thaliana.* Development 125:485–494.

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 identity gene LEAFY. Plant Mol. Biol. 37: 897–910.

Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P, Panda S, Kreps JA, Kay SA (2000) Cloning of the *Arabidopsis* clock gene TOC1 an autoregulatory response regulator homolog. Science 289:768–771.

Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis.* Nature 410:1116–1120.

Sung ZR, Belachew A, Shunong B, Bertrand-Garcia R (1992) *EMF* an *Arabidopsis* gene required for vegetative shoot development. Science 258:1645–1647.

Thompson JD, Higgins DG, Gibson, TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22:4673-4680.

Wagner D, Sablowski RWM, Meyerowitz EM (1999) Transcriptional activation of APETALA1 by LEAFY. Science 285:582–584.

Wang Z-Y, Tobin EM (1998) Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. Cell 93:1207–1217.

Weigel D, Nilsson O (1995) A developmental switch sufficient for flower initiation in diverse plants. Nature 377: 495–500.

Wilson RN, Heckman JW, Somerville CR (1992) Gibberelin is required for flowering in *Arabidopsis thaliana* under short days. Plant Physiol. 100:403–408.

Yang C-H, Chen L-J, Sung ZR (1995) Genetic regulation of shoot development in *Arabidopsis*: Role of the EMF genes. Dev. Biol. 169:421–435.

Yanovsky MJ, Mazzella MA, Casal JJ (2000) A quadruple photoreceptor mutant still keeps track of time. Curr. Biol. 10:1013–1015.