GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

A proposed model for the flowering signaling pathway of sugarcane under photoperiodic control

C.P. Coelho[1], A.P. Costa Netto[2], J. Colasanti[3] and A. Chalfun-Junior[1]

[1]Departamento de Biologia, Setor de Fisiologia Vegetal,
Laboratório de Fisiologia Molecular de Plantas,
Universidade Federal de Lavras, Lavras, MG, Brasil
[2]Curso de Ciências Biológicas, Universidade Federal de Goiás, Campus Jataí,
Jataí, GO, Brasil
[3]Department of Molecular and Cellular Biology, University of Guelph, Guelph,
ON, Canada

Corresponding author: A. Chalfun-Junior
E-mail: chalfunjunior@dbi.ufla.br

Genet. Mol. Res. 12 (2): 1347-1359 (2013)
Received July 30, 2012
Accepted January 25, 2013
Published April 25, 2013
DOI http://dx.doi.org/10.4238/2013.April.25.6

ABSTRACT. Molecular analysis of floral induction in *Arabidopsis* has identified several flowering time genes related to 4 response networks defined by the autonomous, gibberellin, photoperiod, and vernalization pathways. Although grass flowering processes include ancestral functions shared by both mono- and dicots, they have developed their own mechanisms to transmit floral induction signals. Despite its high production capacity and its important role in biofuel production, almost no information is available about the flowering process in sugarcane. We searched the Sugarcane Expressed Sequence Tags database to look for elements of the flowering signaling pathway under photoperiodic control. Sequences showing significant similarity to flowering time genes of other species were clustered, annotated, and analyzed for conserved domains. Multiple alignments comparing the sequences found in the sugarcane database and those from other species

were performed and their phylogenetic relationship assessed using the MEGA 4.0 software. Electronic Northerns were run with Cluster and TreeView programs, allowing us to identify putative members of the photoperiod-controlled flowering pathway of sugarcane.

Key words: Floral induction; Photoperiodism; *Saccharum* spp; Flowering time genes; SUCEST; Biomass yield

INTRODUCTION

Transition from vegetative to reproductive growth is an important event in the development of higher plants. The change to a reproductive program is manifested in vegetative tissues and is regulated by both environmental and endogenous factors. The shoot apical meristem (SAM) is a population of undifferentiated cells that produce leaves and branches during vegetative growth. Under environmental and endogenous responses, SAM undergoes an identity change and produces floral primordia. Molecular analysis of floral induction has been extensively developed in *Arabidopsis* (Simpson and Dean, 2002). These studies have identified numerous flowering time genes that act in 4 response networks: autonomous, gibberellin (Baurle and Dean, 2006), photoperiod, and vernalization (Mouradov et al., 2002; Parcy, 2005).

The light-dependent flowering pathway can be controlled by 2 mechanisms; light quality and day length (Bernier and Periellux, 2005). These environmental factors lead to a cascade of responses that directly affect expression of the *Phytochrome B*, *Cryptochrome 2*, and *Phytochrome A* genes. A circadian clock-controlled mechanism integrates the inductive long-day (LD) photoperiod signals, which leads to the expression of *Gigantea* (*GI*), followed by the activation of *Constans* (*CO*) expression, and finally the induction of the *Flowering locus T* (*FT*) gene. The *FT* gene product acts as a leaf-synthesized florigen that migrates through the phloem to the SAM to cause flowering.

Although grass flowering processes utilize some functions also found in dicots, they also have developed their own mechanisms to produce floral induction signals (Colasanti and Coneva, 2009). Rice plants possess some genes that are absent in dicot plants, such as *GHD7* (*Grain number*, *Plant Height*, and *Heading Date7*) and *EHD1* (*Early heading date1*) (Greenup et al., 2009). These genes, which integrate information about short-day (SD)-induced flowering, act independently of the *GI-CO-FT* pathway.

Sugarcane, a monocot plant, is the main source of sugar production, representing almost two-thirds of the world production. A better understanding of the flowering process in sugarcane, an SD plant, is important because, among other factors, it is related to crop yield. The transition to reproductive growth leads to translocation of some of the sugar to the developing inflorescence, thus diverting the stored sugar away from stalks and, consequently, decreasing crop sugar yield.

Studies involving the flowering process in this crop plant will contribute to future insights into sugarcane genetics, physiological processes related to sucrose content/translocation, and the use of biotechnology approaches to increase sugar production. However, research related to the characterization of sugarcane flowering time genes is scarce.

The Sugarcane EST project (Figueiredo et al., 2001), SUCEST, encompasses a collection of 240,000 ESTs generated from 26 cDNA libraries constructed from different organs and/or tissues at different developmental stages, including apical meristem, flowers, leaf roll,

seeds, internode, stem bark, etiolated leaves, and lateral buds (Vettore et al., 2001).

In this study, we employed *in silico* analyses to search the SUCEST database for putative orthologs of flowering time genes that are involved in the photoperiod-controlled floral inductive signaling pathway in sugarcane.

MATERIAL AND METHODS

Database searches and alignments

Homologs of functionally characterized genes involved in the flowering signaling pathway under photoperiodic control (*GI→CO→FT*) were identified by data mining in the SUCEST database (http://compbio.dfci.harvard.edu/tgi/plant.html) using plant gene (BLASTn) and protein (tBLASTn) sequences as bait. Sequences with significant similarity (e-value $>10^{-4}$) were selected and submitted to clustering by the CAP3 program (Huang and Madan, 1999), forming the EST contigs and singlets.

The *Saccharum officinarum* EST-contigs and EST-singlets obtained were manually annotated, and data validation was performed by local tBLASTx and tBLASTn searches of the retrieved sequences against the GenBank database. Selected sequences were then used as bait in another search against the SUCEST database, aiming at uncovering additional reads, as well as to remount incomplete clusters. This process was repeated until no more new significant reads were found. Open reading frames (ORFs) of validated sequences were obtained through the ORFinder tool, from NCBI (http://www.ncbi.nlm.nih.gov), and their protein sequences were generated through the translation tool found in the ExPASY (http://www.expasy.ch) protein database. The protein sequence alignments were performed by the ClustalW program (Thompson et al., 1994), using default parameters.

Phylogenetic analysis

The putative orthology of the deduced amino acid sequences of sugarcane transcripts, compared to homologs from other species, was assessed by phylogenetic trees formed by the MEGA software, version 4.0 (Tamura et al., 2007), with the neighbor-joining comparison model (Saitou and Nei, 1987), *p*-distance method and pairwise suppression. Bootstrap values from 1000 replicates were used to assess the robustness of the trees.

In silico gene expression analysis

In silico qualitative gene expression profiling was performed using virtual Northern blot analyses. For each EST-contig and EST-singlet, frequencies of reads that form each EST-contig and EST-singlet in the libraries in which they were expressed were calculated. This procedure required that the data have been previously normalized to give a more accurate idea of the degree of expression of the sequences in each treatment and plant organ when all libraries were considered in this study.

Normalization consisted of multiplying each read by the quotient between the number of reads from the library where it was expressed and the sum of reads of all libraries where expression was found. The results were plotted in a matrix and gene expression patterns among

ESTs and libraries were obtained by hierarchical clustering, performed by the Cluster v.3 program (Eisen et al., 1999). Graphic outputs were generated by the TreeView v.1.6 software (Eisen et al., 1999) and presented in a color scale from black to red, where red indicated higher expression levels. Undetectable expression was noted in gray.

RESULTS

The main components of the flowering pathway under photoperiodic control were compiled, and their sequences were used to search sugarcane EST-contigs. Taken together, the results of the phylogenetic analyses, electronic Northern and BLASTp searches allowed for the identification of candidates for several genes that may be involved in the sugarcane flowering pathway under photoperiodic control.

Analyses at the SUCEST database revealed 16 reads related to the *GI* gene, clustered into 7 contigs and 2 singlets. As shown in the phylogenetic tree (Figure 1A) and in Table 1, GiC2 and GiC5 peptide candidates showed high similarity with maize *GI* (*ZmGI*), with amino acid identity ranging from 94 to 97%. It was possible to divide the phylogenetic tree into 2 subgroups, with one group corresponding to neutral day plants, which included GiC2 and another group corresponding to LD and SD plants, including the GiC5. The electronic Northern showed that these contigs are expressed in 6 (GiC2) and 12 (GiC5) different libraries, in no tissue-specific manner, including tissues where the *GI* typically acts, such as apical meristem surrounding immature and mature leaves and in the inflorescence. BLASTp analyses revealed high identity of GiC2 (94%) and GiC5 (97%) with a *ZmGI* ortholog.

Twenty-three reads related to *CO* were found and clustered into 1 contig and 8 singlets. The motif analyses showed that all sequences contained conserved domains for the family of *CO*-like genes, and they could be categorized into the 3 subgroups of *CO* family genes (Griffiths et al., 2003; Wenkel et al., 2006). The results for *CO* gene similarity (Figure 1B) indicated that sugarcane possesses some candidates for each subtype of the *CO*-like superfamily. This result was confirmed by BLASTp analyses, since the candidates found could be related to *COL1*, *COL5*, *COL6*, and *COL10*. Additionally, contig CoS1 was considered a candidate for the *CO* gene because it showed a high level of similarity with the maize HD1 protein (81% identity). Phylogenetic analyses showed that this singlet was grouped with *CO* orthologs of related species, such as rice and maize, and the electronic Northern suggested that its expression was specific to inflorescence tissue.

Candidates for putative *EHD1* and *GHD7* orthologs were clustered into 5 contigs and 4 singlets derived from 20 reads, and 4 contigs and 2 singlets derived from 11 reads, respectively. It was possible to identify some candidates for *EHD1* and *GHD7*, which are monocot-specific genes, in the SUCEST database. According to phylogenetic analyses and the electronic Northern, Ehd1C4 is closely related to the rice *EHD1* gene (*OsEHD1*) and is detected in leaf libraries, such as mature leaf tissues (Figure 1C). After comparing conserved domains, it was possible to identify the CCT domain of the *GHD7* gene. The candidate Ghd7C4 (Figure 1D) showed very high abundance in the mature leaf tissue library. Through BLASTp analyses, Ghd7C4 was found to correlate with the barley *COL7* gene, with an identity of 67%.

It was possible to detect 20 reads for the *FT* gene in searches, clustered into 3 contigs and 5 singlets, all of them containing the PEBP conserved domain and with high similarity to the maize *ZCN* superfamily (Danilevskaya et al., 2008). BLASTp results revealed 2 can-

didates for *Terminal Flower1* (*TFL1*)-like genes in the SUCEST database: FtC3 and FtS2. These candidates showed high similarity to a rice putative *TLF1* (83 and 79%, respectively), and an electronic Northern showed that these sequences are expressed in the leaves (FtS2) and lateral bud (FtC3) tissues. Results of the phylogenetic analyses suggested that the FtS1 and FtS5 EST-contigs were the most related to *ZCN26* (Figure 1E), and the electronic Northern showed expression in root-shoot transition libraries. Phylogenetic results suggested that FtC2 and FtC1 are candidates of the *FT*-like I group orthologs, *ZCN14*-*ZCN15* and *ZCN19*-*ZCN25*, respectively. FtS2, FtC3, FtS1, and FtS5 could be categorized as candidates for the *FT*-like II group, related to *ZCN8* and *ZCN26*. A candidate for a *Mother* of *FT* (*MFT*)-like subfamily gene, *ZCN11*, was also found, i.e., the FtS4 EST-contig.

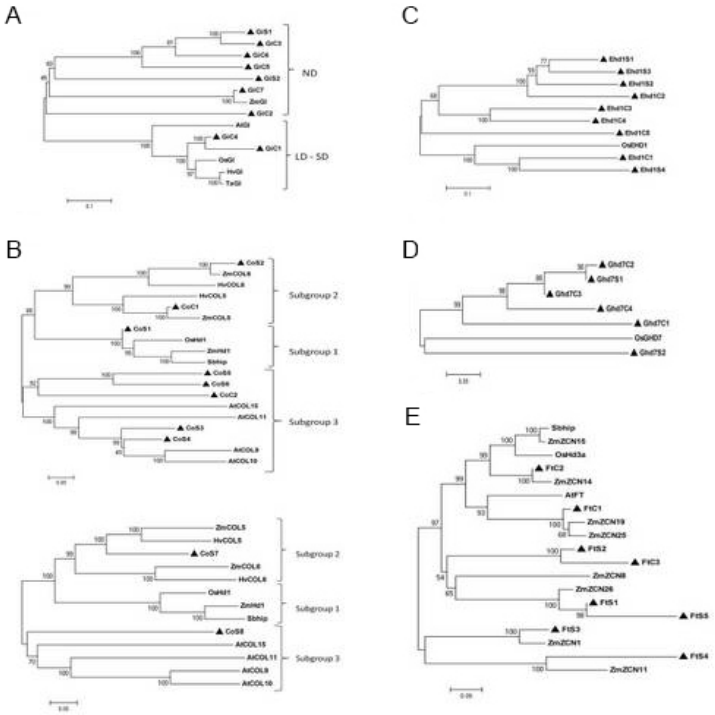

Figure 1. Phylogenetic analyses involving the sugarcane putative genes of the flowering pathway under photoperiodic control (triangles) and homolog sequences obtained from the NCBI database related to **A.** *GI*; **B.** *CO*; **C.** *EHD1*; **D.** *GDH7*; **E.** *FT.* Neighbor-joining trees were built for sugarcane-deduced amino acids and protein sequences from other species aligned with ClustalW2. Bootstrap values from 1000 replications were used to assess the robustness of the trees. Genetic distances are shown at the given scales. The protein sequences from other species and their respective accession numbers are as follows: **A.** *Zea mays* [ZmGI (ABZ81992.1)], *Arabidopsis thaliana* [AtGI (AAF00092.1)], *Oryza sativa* [OsGI (BAD68052.1)], *Hordeum vulgare* [HvGI (ACM49849.1)], *Triticum aestivum* [TaGI (AAL08497.2)]. **B.** *Arabidopsis thaliana* [AtCOL9 (NP_187422.1), AtCOL10 (NP_199636.1), AtCOL11 (NP_193260.2), AtCOL15 (NP_174126.1)], *Zea mays* [ZmCOL5 (NP_001147679.1), ZmCOL6 (NP _001148229.1), ZmHd1 (ABW82153.1)], *Hordeum vulgare* [HvCOL5 (AAL99264.1), HvCOL6 (AAL99267.1)], *Oryza sativa* [OsHd1 (BAI59731.1)], *Sorghum bicolor* [Sbhip (XP_002436860.1)]. **C.** *Oryza sativa* [OsEHD1 (ABF95340.1)]. **D.** *Oryza sativa* [OsGHD7 (ACA14489.1)]. **E.** *Zea mays* [ZmZCN1 (ABW96224.1), ZmZCN8 (ABX11010.1), ZmZCN11 (NP_001106264.1), ZmZCN14 (NP_001106251.1), ZmZCN15 (ABW96237.1), ZmZCN19 (NP_001106256.1), ZmZCN25 (NP_001106257.1), ZmZCN26 (ABW96244.1)], *Oryza sativa* [OsHd3a (BAF15064.1)], *Arabidopsis thaliana* [AtFT (AAF03936.1)], *Sorghum bicolor* [Sbhip (XP_002436509.1)].

Table 1. Comparison of the sugarcane ESTs related to the flowering pathway under photoperiodic control found in the SUCEST database and their best hit of BLASTp analysis at NCBI database.

Category	Contig/aa	BLASTp	e-value	Identity	Positives
GI	GiC1/514	DAA06172.1| TPA_inf: gigantea 1A [*Zea mays*]. 1162 aa	0.0	424/469 (90%)	437/469 (93%)
	GiC2/332	DAA06172.1| TPA_inf: gigantea 1A [*Zea mays*]. 1162 aa	0.0	311/330 (94%)	315/330 (95%)
	GiC3/193	BAD68053.1| putative gigantea [*Oryza sativa* Japonica Group]. 261 aa	$4e^{-92}$	143/186 (77%)	156/186 (84%)
	GiC4/289	CAB56058.1| gigantea homologue [*Oryza sativa*]. 975 aa	$9e^{-175}$	252/284 (89%)	261/284 (92%)
	GiC5/271	DAA06172.1| TPA_inf: gigantea 1A [*Zea mays*]. 1162 aa	$6e^{-180}$	261/269 (97%)	266/269 (99%)
	GiC6/184	DAA06172.1| TPA_inf: gigantea 1A [*Zea mays*]. 1162 aa	$2e^{-88}$	140/161 (87%)	151/161 (94%)
	GiC7/317	DAA06172.1| TPA_inf: gigantea 1A [*Zea mays*]. 1162 aa	0.0	297/309 (96%)	304/309 (98%)
	GiS1/117	AF469490_1 gigantea-like protein [*Triticum aestivum*].190 aa	$1e^{-42}$	70/83 (84%)	75/83 (90%)
	GiS2/87	ABR26000.1| gigantea protein [*Oryza sativa* Indica Group]. 169 aa	$3e^{-27}$	56/87 (64%)	64/87 (74%)
CO	CoC1/127	AAM74069.1|AF490474_1 CONSTANS-like protein [*Hordeum vulgare* subsp *vulgare*]. 323 aa	$1e^{-56}$	93/119 (78%)	102/119 (86%)
	CoC2/206	NP_195607.2| B-box-type zinc finger-containing protein [*Arabidopsis thaliana*]. 183 aa	$3e^{-73}$	116/207 (56%)	145/207 (70%)
	CoS1/75	ABW82153.1| Hd1 [*Zea mays*]. 398 aa	$2e^{-35}$	63/78 (81%)	68/78 (87%)
	CoS2/279	NP_001148229.1| CONSTANS-like protein CO6 [*Zea mays*]. 364 aa	$2e^{-142}$	242/267 (91%)	245/267 (92%)
	CoS3/116	NP_199636.1| zinc finger protein CONSTANS-LIKE 10 [*Arabidopsis thaliana*]. 373 aa	$3e^{-54}$	83/114 (73%)	97/114 (85%)
	CoS4/91	NP_199636.1| zinc finger protein CONSTANS-LIKE 10 [*Arabidopsis thaliana*]. 373 aa	$7e^{-38}$	65/90 (72%)	72/90 (80%)
	CoS5/163	ACC85612.1| zinc finger protein [*Phyllostachys edulis*]. 256 aa	$5e^{-65}$	110/158 (70%)	118/158 (75%)
	CoS6/159	ACF35275.1| B-box zinc finger protein [*Bambusa oldhamii*]. 256 aa	$9e^{-64}$	106/145 (73%)	113/145 (78%)
	CoS7/152	NP_001147679.1| CONSTANS-like protein CO5 [*Zea mays*]	$5e^{-48}$	91/135 (67%)	100/135 (74%)
	CoS8/136	AAZ86536.1| truncated COL1 [*Lolium perenne*]. 168 aa	$1e^{-66}$	96/135 (71%)	108/135 (80%)
FT	FtC1/109	ABW96241.1| ZCN19 [*Zea mays*]. 175 aa	$2e^{-73}$	104/109 (95%)	104/109 (95%)
	FtC2/115	ABW96236.1| ZCN14 [*Zea mays*]. 236 aa	$6e^{-79}$	113/115 (98%)	114/115 (99%)
	FtC3/85	BAD73176.1| putative terminal flower1 [*Oryza sativa* Japonica Group]. 180 aa	$2e^{-35}$	57/69 (83%)	61/69 (88%)
	FtS1/175	ABW96244.1| ZCN26 [*Zea mays*]. 187 aa	$3e^{-119}$	163/173 (94%)	169/173 (98%)
	FtS2/179	BAD73176.1| putative terminal flower1 [*Oryza sativa* Japonica Group]. 180 aa	$3e^{-95}$	136/272 (79%)	148/172 (86%)
	FtS3/173	ABW96224.1| ZCN1 [*Zea mays*]. 173 aa	$3e^{-113}$	158/173(91%)	163/173 (94%)
	FtS4/128	ABW96234.1| ZCN11 [*Zea mays*]. 180 aa	$7e^{-48}$	81/93 (87%)	86/93 (92%)
	FtS5/78	ABX11026.1| ZCN26 [*Zea mays*]. 187 aa	$3e^{-24}$	59/65 (91%)	62/65 (95%)
EHD1	Ehd1C1/215	ADX60172.1| ARR-B transcription factor [*Zea mays*]. 631 aa	2^{e-128}	201/209 (96%)	206/209 (99%)
	Ehd1C2/130	ADX60157.1| ARR-B transcription factor [*Zea mays*]. 631 aa	$9e^{-41}$	70/111 (63%)	84/111 (76%)
	Ehd1C3/252	ABF95340.1| two-component response regulator-like PRR73 [*Oryza sativa* Japonica Group]. 473 aa	$1e^{-118}$	189/237 (80%)	201/237 (85%)
	Ehd1C4/170	NP_001151536.1| two-component response regulator-like PRR95 [*Zea mays*]. 630 aa	$2e^{-111}$	161/165(95%)	165/169 (98%)
	Ehd1C5/548	ADO51647.1| PRR59 [*Zea mays*]. 695 aa	0.0	454/547 (83%)	481/547 (88%)
	Ehd1S1/125	ADX60172.1| ARR-B transcription factor [*Zea mays*]. 631 aa	$4e^{-59}$	92/109 (84%)	98/109 (90%)
	Ehd1S2/146	NP_001104864.1| response regulator 10 [*Zea mays*]. 686 aa	$2e^{-92}$	130/139 (94%)	136/139 (98%)
	Ehd1S3/153	ADX60172.1| ARR-B transcription factor [*Zea mays*]. 631 aa	$5e^{-89}$	132/146 (90%)	139/146 (95%)
	Ehd1S4/184	ADX60172.1| ARR-B transcription factor [*Zea mays*]. 631 aa	$2e^{-38}$	85/144 (59%)	102/144 (71%)
GHD7	Ghd7C1/152	AF490474_1 CONSTANS-like protein [*Hordeum vulgare* subsp *vulgare*]. 323 aa	$5e^{-48}$	117/153 (76%)	120/153 (78%)
	Ghd7C2/75	NP_001148229.1| CONSTANS-like protein CO6 [*Zea mays*]. 364 aa	$3e^{-49}$	67/71 (94%)	67/71 (94%)
	Ghd7C3/73	AAL99266.1| CONSTANS-like protein CO6 [*Hordeum vulgare* subsp *vulgare*]. 357 aa	$5e^{-25}$	63/75 (84%)	67/75 (89%)
	Ghd7C4/105	AAL99269.1| CONSTANS-like protein CO7 [*Hordeum vulgare* subsp *vulgare*]. 244 aa	$1e^{-25}$	69/103 (67%)	71/103 (69%)
	Ghd7S1/89	NP_001148229.1| CONSTANS-like protein CO6 [*Zea mays*]. 364 aa	$1e^{-46}$	77/81(95%)	77/81 (95%)
	Ghd7S2/127	AAL99270.1| CONSTANS-like protein CO8 [*Hordeum vulgare* subsp *vulgare*]. 247 aa	$2e^{-16}$	44/81 (54%)	51/81 (63%)

DISCUSSION

Previous studies in *Arabidopsis* and rice have shown that, although some pathways are conserved between mono- and dicot plants, each group has developed specific mechanisms to control the flowering process.

Almost nothing is known about the flowering pathway in sugarcane, although this process is extremely important and related to crop yield. A search for 5 flowering time genes in sugarcane, *GI*, *CO*, *EHD1*, *GHD7*, and *FT*, which are both mono- and dicot-specific genes, was performed in the SUCEST database.

GI

GI is a large protein that is nuclear localized (Huq et al., 2000) and regulates flowering through the integration of circadian rhythm periods, acting upstream of the *CO* gene (Samach and Coupland, 2000). Analysis of this gene was performed since sugarcane floral induction is controlled by photoperiod and the circadian clock. The circadian clock is a pacemaker that controls rhythmic processes that occur within a period of 24 h (Hayama and Coupland, 2003). Mutation of *GI* impairs circadian rhythms and delays the flowering process, suggesting that its functions is to couple the circadian clock to the day-length perception inducing downstream genes, including those specifically related to the floral transition, such as *CO* and *FT* (Mizoguchi et al., 2005). *GI* is highly conserved in higher plants, including monocot species such as rice (Hayama et al., 2002). A functional hierarchy of *GI*-*CO*-*FT* acting together to connect the circadian oscillator to the flowering pathway has been established (Mizoguchi et al., 2005). The 2 candidates found in sugarcane contain a conserved domain present in *GI*, although the sequences found encode incomplete ORFs. The contig GiC2 was grouped with GI proteins of day-neutral plants, and the GiC5 contig was grouped with *GI* of LD and SD plants. This information supports the idea that each candidate may act differently, depending on the environmental conditions in which sugarcane is subjected. Results from Higuchi et al. (2011) suggest that in *Pharbitis nil*, an SD plant with an absolute requirement for SD photoperiods to induce flowering, a *GI* ortholog functions as a suppressor of flowering through the repression on an *FT* ortholog. Although sugarcane is an SD plant (Araldi et al., 2010), other variables interact with photoperiod signals to determine floral induction, such as low temperatures. Taken together with the electronic Northern (Figure 2A), it is possible to predict that the *GI* ortholog (GiC5) is important in the sugarcane flowering network, although its specific function (such as inducing or repressing downstream *FT*-like genes) needs to be confirmed by functional analyses.

CO

The *CO* gene, downstream of *GI*, is a key regulatory protein that integrates signals from the circadian clock to control flowering (Putterill, 2001; Valverde, 2011). *CO* expression exhibits a circadian rhythm under continuous light, in which *CO* has a diurnal expression pattern with a peak in the night, regulated by the circadian clock (Hayama and Coupland, 2003). Rice, which is an SD plant, has a *CO* ortholog (the *Hd1* gene) whose expression is repressed under LD and induced under SD floral inductive photoperiods (Izawa et al., 2003). Since sug-

arcane is also an SD plant, a mechanism of flowering control via *CO* orthologous genes may be shared between these 2 species. All the *CO*-like candidates found contained an amino terminus B-box superfamily conserved domain, which regulates protein-protein interactions (Torok and Etkin, 2001), and/or the carboxyl terminus CCT domain. They could be grouped into specific classes of the *CO*-like gene family. Cereals possess specific classes that are absent in *Arabidopsis*; i.e., the group I class genes that contain a single-B-box domain and the group IV class, which lacks the B-box domain and has only the CCT domain (Griffiths et al., 2003). In group IV, candidate genes were found in sugarcane. Complete ORFs of possible *CO* orthologs in sugarcane could be found, such as CoS1, which has the B-box superfamily-conserved domain. CoC1 and CoC2 transcripts were present in different leaf libraries (Figure 2B). BLASTp analyses identified barley and *Arabidopsis CO*-like genes; phylogenetic analyses suggested that they are related to subgroups I and III, respectively. The BLASTp and the phylogenetic results showed a closer relationship of CoC1 to monocot *CO* orthologs, suggesting that this contig may be the candidate of *CO* in sugarcane. Functional characterization needs to be performed to verify under which conditions (SD or LD) this gene is induced and/or repressed to determine whether or not this mechanism of control is shared between sugarcane and rice.

EHD1

A putative rice *EHD1* B-type response regulator (RR) domain was detected in all sugarcane *EHD1* candidates found at the SUCEST database. In rice, this gene acts as a floral inducer under SD conditions by controlling the expression of the *FT* gene, and independently of the *Hd1* gene, induces expression of *FT*-like genes after SD treatment in *Hd1*-deficient strains. Rice *Hd1* is expressed only under SD conditions, but *EHD1* is expressed in both conditions, independent of *Hd1* (Doi et al., 2004). As in rice, a candidate sugarcane ortholog of *EHD1* may perform this function. A complete ORF of Ehd1C4 was found to possess the RR-conserved domain. Additionally, this contig was detected in leaf libraries (Figure 2C), indicating that sugarcane may possess a 2-component flower signaling pathway, such as in rice plants (Doi et al., 2004; Endo-Higashi and Isawa, 2011).

GHD7

The *GHD7* gene is responsible for inactivation of *EHD1* under LD conditions in rice. Similar to *EHD1*, the *GHD7* gene is a monocot-specific gene, so far found only in rice, whose expression is related to crop grain number (Xue et al., 2008). The GHD7 protein contains a CCT motif, which mediates protein-protein interaction and nuclear localization, as found in CO proteins. *GHD7* candidates found in sugarcane EST database are very similar to rice *GHD7* and *CO*-like genes because they share the same conserved domain. Due to this feature, BLASTp analyses found sequences related to *CO* genes (Table 1). However, when the analyses for *CO* genes were performed, these genes were not found, suggesting that they are *GHD7* candidates. *GHD7* contributes to the adaptation of rice cultivars to cold-climate regions (Xue et al., 2008), and it has been recently revealed that *GHD7* transcription is mediated through phytochrome signaling and is gated in a photoperiod-dependent manner (Itoh et al., 2010). Electronic Northern results showed high abundance of potential *GHD7* orthologs in sugarcane leaf tissue libraries, such as Ghd7C4, as in rice, where *GHD7* is strongly expressed in

the blades of fully expanded leaves (Xue et al., 2008). This suggests that this candidate may have an important function in the flowering process (Figure 2D), although expression in leaves alone does not mean that this gene is involved in the floral induction. There is a correlation between the *EHD1* levels and *GHD7* induction under non-inductive LD conditions, where *GHD7* represses transcription of *EHD1*, thus subsequently affecting expression of *Heading date 3a* (*Hd3a*, a rice florigen) (Itoh et al., 2010). The existence of this pathway may be predicted in sugarcane, since candidate orthologs were found in SUCEST.

FT

The *FT* gene, which encodes a phosphatidylethanolamine-binding protein (PEBP)-related protein, is highly conserved in plants (Kobayashi et al., 1999). The FT protein has been found to move to the SAM, via the phloem, where it acts as a floral stimulus by activating the FD transcription factor (Corbesier et al, 2007). *FT*-related genes have been found in monocot plants such as wheat (19 genes), maize (30 genes, according to Chardon and Damerval, 2005; and 25 genes according to Danilevskaya et al., 2008), and rice (19 genes). Within the *ZCN* superfamily there are 3 major subfamilies: *FT*-like, *MFT*-like, and *TFL1*-like (Danilevskaya et al., 2008). Candidates for all members of the *ZCN* superfamily could be found, including members of the *FT*-like I group, comprising monocot floral activators such as *Hd3a* (Danilevskaya et al., 2008). Maize possesses *TFL1*-like genes, named *ZCN1* to *ZCN6* genes, which when expressed ectopically modify the flowering time and inflorescence architecture in transgenic maize plants and maintain the indeterminacy of vegetative meristems (Danilevskaya et al., 2010). *TLF1*, which is a meristem identity gene, is an antagonist of the *FT* gene, despite the similarity in their sequences (Kobayashi et al., 1999; Tahery et al., 2009). The candidate FtS2 is a possible sugarcane *FT* ortholog; it contains the complete PEBP superfamily domain, and transcripts have been found in mature leaf libraries (Figure 2E), as would be expected of a potential florigen-encoding gene. Despite a BLASTp analysis suggesting that FtS2 is highly similar to the TFL1 protein, which is an antagonist of the FT protein, FtS2 is expressed in mature leaf tissues, indicating that it is more likely an *FT* ortholog rather than *TFL1*, which is expressed in inflorescence meristem tissues (Tahery et al., 2009).

ZCN8

Recent findings show that *ZCN8* is an *FT* ortholog gene in maize and teosinte (Lazakis et al., 2011; Meng et al., 2011). Exposure of teosinte plants to SD photoperiods that cause flowering is correlated with a large increase in *ZCN8* expression. A more moderate increase in *ZCN8* expression has been observed in maize, a neutral-day plant, under SD conditions. Other evidence has shown that *ZCN8* is involved in photoperiod sensitivity, acting as a florigen, since ectopic expression in *Arabidopsis* rescues the *FT* mutant phenotype (Lazakis et al., 2011). Moreover, *ZCN8*-silencing experiments showed late flowering of transgenic maize plants (Meng et al., 2011). No clear ortholog of *ZCN8*/*FT* was found in BLASTp searches, and the expression profile showed no transcript in leaf tissues (Figure 2E). However, through phylogenetic analyses, the FtS1 and FtS5 EST-contigs were the most related to *ZCN8*. Together, this information suggests that *ZCN*-like superfamily gene function is conserved between the species, and that some of them may act as floral activators in sugarcane.

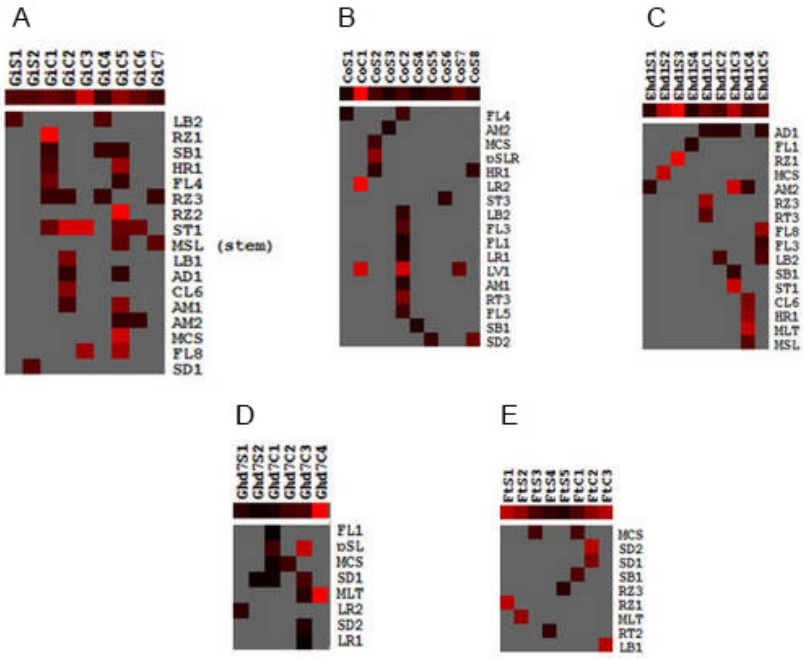

Figure 2. *In silico* expression profile of putative elements of the flowering pathway under photoperiodic control: **A.** *Gigantea*; **B.** *Constans*; **C.** *Early heading date1*; **D.** *Heading date7*; **E.** *Flowering locus T*. The normalized numbers of reads for the transcripts in each library are represented in a scale from black to red. The contigs (C) and singlets (S) are represented as columns and the sugarcane libraries as lines. Sugarcane libraries are as follows: FL4 (developed inflorescence and rachis), AM2 (apical meristem and tissues surrounding of immature plants), MCS (stem), pSRL (leaf roll including apex after floral induction), HR1 (seedling inoculated with *Herbaspirilum rubrisubal*), LR2 [(leaf roll from field-grown adult plants (small insert)], ST3 (fourth apical stalk internodes of adult plants), LB2 (lateral buds from adult plants), FL3 (base of developing inflorescence), FL1 (inflorescence at the beginning of development), LR1 [leaf roll from field-grown adult plants (large insert)], LV1 (etiolated leaves from *in vitro*-grown seedlings), AM1 (apical meristem and tissues surrounding of mature plants), RT3 (root apex from adult plants), FL5 (developed inflorescence), SB1 (stalk bark from adult plants), SD2 (developing seeds), AD1 (seedlings inoculated with *Gluconacetobacter diazot*), FL1 [inflorescence at beginning of development (1 cm long)], RZ1 [shoot-root transition zone from young plants (large insert)], RZ3 (shoot-root transition zone from adult plants), FL8 [developing inflorescence and rachis (10 cm long)], ST1 (first apical stalk internodes of adult plants), CL6 [pool of sugarcane calli submitted to low temperature (4°C)], MLT (mature leaf tissue), MSL (sugarcane mature stem library), SD1 [developing seeds (large insert library)], RT2 [root tips (0.3 cm long) from adult plants], LB1 (lateral buds from field-grown adult plants), pSL (leaves after floral induction), RZ2 [shoot-root transition zone from young plants (small insert)].

A model for the photoperiodic mechanism of the flowering pathway (Figure 3) is proposed based on a comparative transcriptome and sequence conservation analysis. This preliminary study of the flowering time genes in sugarcane provides basic information for in-depth studies relating to the flowering process in this important crop. Further analyses of the genes identified will also provide a better understanding of the photoperiodic control of this crucial metabolic process. Moreover, functional characterization will help to unravel the molecular basis of the flowering process in different varieties of sugarcane with distinct florigenic potential.

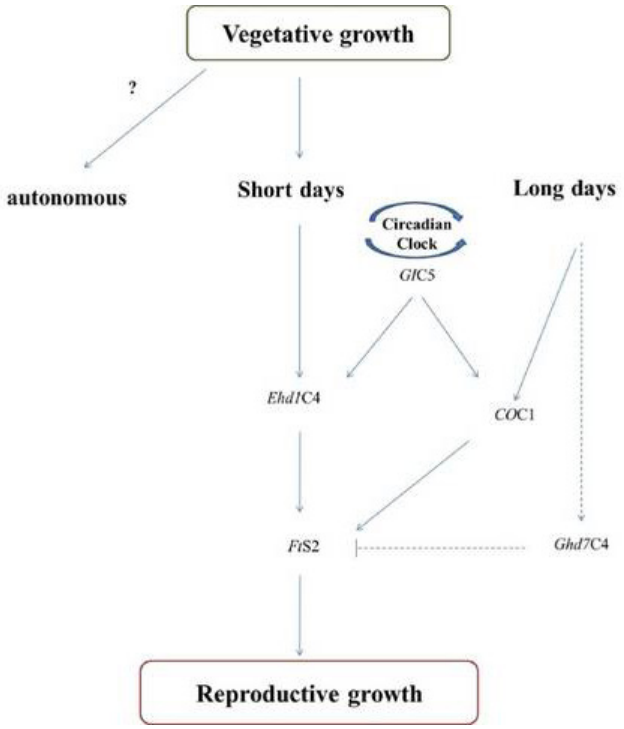

Figure 3. A hypothetical model of the gene network controlling floral induction under photoperiodic control in sugarcane. Photoperiodism and circadian clock controls the pathway by regulation of *GI*, *EHD1*, and *CO* ortholog genes. Two distinct mechanisms are assumed to be involved in the regulation of the *FT* ortholog gene, one under long-day and short-day conditions, which induces *FtS2* expression and another mechanism through the component *GHD7* ortholog, which suppresses *FtS2* expression in non-inducing conditions. Each element of the pathway corresponds to the EST-contig selected through the *in silico* characterization of the putative flowering time genes.

ACKNOWLEDGMENTS

Research supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), including fellowships and the "Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG", for the financial support.

REFERENCES

Baurle I and Dean C (2006). The timing of developmental transitions in plants. *Cell* 125: 655-664.

Bernier G and Perilleux C (2005). A physiological overview of the genetics of flowering time control. *Plant Biotechnol. J.* 3: 3-16.

Chardon F and Damerval C (2005). Phylogenomic analysis of the PEBP gene family in cereals. *J. Mol. Evol.* 61: 579-590.

Colasanti J and Coneva V (2009). Mechanisms of floral induction in grasses: something borrowed, something new. *Plant Physiol.* 149: 56-62.

Corbesier L, Vincent C, Jang S, Fornara F, et al. (2007). FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* 316: 1030-1033.

Danilevskaya ON, Meng X, Hou Z, Ananiev EV, et al. (2008). A genomic and expression compendium of the expanded PEBP gene family from maize. *Plant Physiol.* 146: 250-264.

Danilevskaya ON, Meng X and Ananiev EV (2010). Concerted modification of flowering time and inflorescence architecture by ectopic expression of *TFL1*-like genes in maize. *Plant Physiol.* 153: 238-251.

Doi K, Izawa T, Fuse T, Yamanouchi U, et al. (2004). *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. *Genes Dev.* 18: 926-936.

Eisen MB, Spellman PT, Brown PO and Botstein D (1999). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95: 14863-14868.

Endo-Higashi N and Isawa T (2011). Flowering time genes *Heading date 1* and *Early heading date 1* together control panicle development in rice. *Plant Cell Physiol.* 52: 1083-1094.

Figueiredo RC, Brito MS, Figueiredo LHM, Quiapin AC, et al. (2001). Dissecting the sugarcane expressed sequence tag (SUCEST) database: unraveling flower-specific genes. *Genet. Mol. Biol.* 24: 77-84.

Greenup A, Peacock WJ, Dennis ES and Trevaskis B (2009). The molecular biology of seasonal flowering-responses in *Arabidopsis* and the cereals. *Ann. Bot.* 103: 1165-1172.

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

Hayama R and Coupland G (2003). Shedding light on the circadian clock and the photoperiodic control of flowering. *Curr. Opin. Plant Biol.* 6: 13-19.

Hayama R, Izawa T and Shimamoto K (2002). Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. *Plant Cell Physiol.* 43: 494-504.

Higuchi Y, Sage-Ono K, Sasaki R, Ohtsuki N, et al. (2011). Constitutive expression of the *GIGANTEA* ortholog affects circadian rhythms and suppresses one-shot induction of flowering in *Pharbitis nil*, a typical short-day plant. *Plant Cell Physiol.* 52: 638-650.

Huang X and Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868-877.

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

Itoh H, Nonoue Y, Yano M and Izawa T (2010). A pair of floral regulators sets critical day length for Hd3a florigen expression in rice. *Nat. Genet.* 42: 635-638.

Izawa T, Takahashi Y and 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.

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

Lazakis CM, Coneva V and Colasanti J (2011). ZCN8 encodes a potential orthologue of *Arabidopsis FT* florigen that integrates both endogenous and photoperiod flowering signals in maize. *J. Exp. Bot.* 62: 4833-4842.

Meng X, Muszynski MG and Danilevskaya ON (2011). The *FT*-like *ZCN8* gene functions as a floral activator and is involved in photoperiod sensitivity in maize. *Plant Cell* 23: 942-960.

Mizoguchi T, Wright L, Fujiwara S, Cremer F, et al. (2005). Distinct roles of *GIGANTEA* in promoting flowering and regulating circadian rhythms in *Arabidopsis*. *Plant Cell* 17: 2255-2270.

Mouradov A, Cremer F and Coupland G (2002). Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14 (Suppl): S111-S130.

Parcy F (2005). Flowering: a time for integration. *Int. J. Dev. Biol.* 49: 585-593.

Putterill J (2001). Flowering in time: genes controlling photoperiodic flowering in *Arabidopsis*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356: 1761-1767.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

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

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

Tahery H, Abdullah MP, Norlia B, Kafilzadeh F, et al. (2009). Terminal flower 1 (*TFL1*) homolog genes in monocots. *Eur. J. Sci. Res.* 38: 26-37.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

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

Torok M and Etkin LD (2001). Two B or not two B? Overview of the rapidly expanding B-box family of proteins. *Differentiation* 67: 63-71.

Valverde F (2011). CONSTANS and the evolutionary origin of photoperiodic timing of flowering. *J. Exp. Bot.* 62: 2453-2463.

Vettore AL, Silva FR da, Kemper EL and Arruda P (2001). The libraries that made SUCEST. *Genet. Mol. Biol.* 24: 1-7.
Wenkel S, Turck F, Singer K, Gissot L, et al. (2006). CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* 18: 2971-2984.
Xue W, Xing Y, Weng X, Zhao Y, et al. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40: 761-767.