IDENTIFICATION AND PARTIAL CHARACTERIZATION OF THE FLOWERING TIME GENE *ScID1* IN SUGARCANE

CARLA PRISCILA COELHO

2010

CARLA PRISCILA COELHO

IDENTIFICATION AND PARTIAL CHARACTERIZATION OF THE FLOWERING TIME GENE *ScID1* IN SUGARCANE

This dissertation is being submitted in a partial fulfillment of the requirements for the degree of Master in Plant Physiology of the Universidade Federal de Lavras.

Supervisor
Prof. Antonio Chalfun Junior

LAVRAS
MINAS GERAIS - BRASIL
2010

CARLA PRISCILA COELHO

**IDENTIFICATION AND PARTIAL CHARACTERIZATION OF
THE FLOWERING TIME GENE *ScID1* IN SUGARCANE**

This dissertation is being submitted in a partial
fulfillment of the requirements for the degree of
Master in Plant Physiology of the Universidade
Federal de Lavras.

APPROVED on March, 12th, 2010

Prof. Dr. Joseph Colasanti   University of Guelph

Prof. Dr. Antônio Paulino da Costa Netto   UFG

Prof. Antonio Chalfun Junior, PhD
UFLA
Supervisor

LAVRAS
MINAS GERAIS – BRASIL

"Sem a curiosidade que me move, que me inquieta, que me insere na busca, não aprendo nem ensino" Paulo Freire

"Se, a principio, a idéia não é absurda, então não há esperança para ela" Albert Einstein

Dedicated to

My beloved niece who is giving me the happiest days of my life.

To my father Beto, my mother Agnelina, my siblings Tamara and Pipe. You are the reason of my life.

To all my family for supporting me in all my decisions.

To my best friend Amanda just because.

SUMMARY

ABBREVIATION LIST

SAM   Shoot Apical Meristem
ID1   *INDETERMINATE1 gene*
IDD   *ID domain*
ScID1   Sugarcane *ID1*
SbID1   Sorghum *ID1*
OsID1   Rice *ID1*
SD   Short Day
LD   Long Day
PI   *PISTILATA*
AG   *AGAMOUS*
SEP   *SEPALLATA*
AP1   *APETALA 1*
AP2   *APETALA 2*
AP2   *APETALA3*
CAL   *CAULIFLOWER*
CO   *CONSTANS*
FLC   *FLOWERING LOCUS C*
FLF   *FLOWERING LOCUS F*
FRI   *FRIGIDA*
FT   *FLOWERING LOCUS T*
LFY   *LEAFY*
NLS   *nuclear localization sequence*
SOC1   *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS*
TOC1   *TIMING OF CAB EXPRESSION 1*
VERN1   *VERNALIZATION 1*
VERN2   *VERNALIZATION2*
VIN3   *VERNALIZATION INSENSITIVE 3*

GENERAL ABSTRACT

COELHO, Carla Priscila. **Identification and partial characterization of the flowering time gene *ID1* in sugarcane**. 2010. 68 p. Dissertation (Master´s Degree in Plant Physiology) – Universidade Federal de Lavras, Lavras.

Sugarcane is an important crop with C4 carbon metabolism, which makes this crop one of the most productive cultivated plants. Recently, it has gained more attention because of its potential use in biofuel productivity. Such fact has made Brazil one of the most promising countries in this field. The fact that sugarcane has one of the most complex genomes of cultivated plants is hampering the development of sugarcane genetics that could support breeding and crop improvement programs. Sugarcane belongs to *Saccharum* L genus, within Poaceae (Grass) family and *Andropogonae* tribe, which involves only polyploid species. After Genomic research, sugarcane genome has been studied and unravelled, but, still, more research needs to be developed due to sugarcane genome complexity. The aim of this research was to identify and analyse the expression of an orthologue of the maize *ID1* gene in sugarcane. Maize ID1 has an important flowering time regulator in this species. *In silico* analysis showed five contigs and four singlets that might be putative ID sequences in sugarcane. The isolation of partial sequence of *ScID1* suggests that it plays a role in flowering process. Expression analysis showed that this gene is only expressed in immature leaves of sugarcane and not expressed in the shoot apical meristem. The ScID1 protein may act as a transcription factor during flowering in sugarcane.

_____
[*]Guidance Committee: Antonio Chalfun Junior – UFLA (Orientador), Joseph Colasanti – University of Guelph

RESUMO GERAL

COELHO, Carla Priscila. **Identificação e caracterização parcial do gene do florescimento *ID1* em cana de açúcar.** 2010. 68 p. Dissertação (Mestrado em Fisiologia Vegetal) – Universidade Federal de Lavras, Lavras.*

A cana de açúcar é uma importante cultura que apresenta o metabolismo C4, sendo uma das mais plantas cultivadas mais produtivas. Recentemente, ela tem ganhado mais atenção devido ao seu potencial como fonte para a produção de biocombustíveis. Tal fato elevou o Brasil a um dos países mais promissores no setor. Contudo, a cana de açúcar, também, possui um genoma complexo, o que tem dificultado o desenvolvimento de pesquisas na área da genética, o que poderia dar suporte a programas de melhoramento da cultura. A cana de açúcar pertence ao gênero *Saccharum* L genus, que está dentro da família Poaceae (Gramineae) e da tribo *Andropogonae*, que envolve somente espécies poliplóides. Após o advento da Genômica, o genoma da cana de açúcar tem sido desvendado, contudo, a complexidade de seu genoma tem sido confirmada em muitos aspectos. O objetivo desse trabalho de dissertação foi identificar e analisar o perfil de expressão de um ortólogo do gene *ID1* de milho em cana de açúcar. ID1 de milho tem uma importante função de regulação do florescimento nessa espécie. Análises *in silico* demonstraram a presença de cinco contigs e quatro singlets, que são possíveis sequências ID em cana de açúcar. O isolamento da sequência parcial de *ScID1* sugere que esse gene possivelmente age no processo de florescimento em cana de açúcar. Análises de expressão sugerem que esse gene é expresso somente em folhas imaturas. A proteína ScID1 pode agir como um fator de transcrição, controlando o processo de florescimento nessa espécie.

_____

CHAPTER 1 - DEVELOPMENTAL PHYSIOLOGY OF SUGARCANE

1 INTRODUCTION

Transition from vegetative to reproductive growth is an important event in the development of higher plants. The change is observed in vegetative tissues that are 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 promotes identity change, producing floral primordia. Molecular analysis of floral induction has been extensively developed in *Arabidopsis* (Simpson & Dean, 2002). These studies have identified several flowering time genes related to four response networks: autonomous, gibberellin (Baurle & Dean, 2006), photoperiod and vernalization (Mouradov et al., 2002; Parcy, 2005).

Although grass flowering processes follows some ancestral dicot functions, they have developed their own mechanism to transmit floral induction signals (Colasanti & Coneva, 2009). In maize, the flowering transition is affected by *INDETERMINATE 1 gene* (*ID1*) (Colasanti et al., 1998). *ID1* is only expressed in immature maize leaves, indicating that its activity may be involved in signal transduction to the apical meristem (Kozaki et al., 2004).

Sugarcane is a monocot plant, which is the main source of sugar production, representing almost two thirds of world production (D´Hont, 2008). A better understanding of the flowering process in sugarcane is required for many reasons. The transition to the reproductive growth leads the translocation of part of the sugar to the developing inflorescence, which decreases the amount of sugar content in the stalks, and consequently the production yield. Studies involving the flowering process in this crop will contribute to future guides in sugarcane genetics, physiological processes related to sucrose content and the

use of biotechnology approaches to increase sugar production. The publications relating to *ID1* gene characterization and expression in sugarcane are scarce.

In this study it was aimed to identify and characterize the flowering time gene *ID1* in sugarcane.

2 BACKGROUND

2.1 Physiology of *Saccharum* ssp.

Sugarcane is a grass originating from Asia, cultivated in tropical and subtropical regions of the world. It is an important crop due to its ability to stock high concentrations of sugar in the internodes, to be efficient in the energy conversion process and because it is an important biofuel source (Tew & Cobill, 2008). Genus *Saccharum* comprises six species according to D'Hont et al. (1998): *S. spontaneum*, *S. officinarum*, *S. robustum*, *S. edule*, *S. barberi* and *S. sinense*.

Sugarcane is cultivated in a wide range of latitudes, between 35º N and 30º S, and altitudes between sea level and 1000 meters. Out of 121 countries produce sugarcane in a producing area of 20 millions of hectares (Food and Agricultural Organization – FAO, 2009), only 15 countries represent 86% of cultivated area and 87.1% of production.

Sugarcane has a C4 metabolism, presenting a high conversion efficiency of light energy into chemical energy (Lee & Bressan, 2006). Photosynthetic efficiency can vary among varieties and under different environmental conditions during the crop development. The energy conversion is affected by light (intensity and quantity), CO_2 concentration, water and nutrient availability

and temperature. Temperature is one of the most important factors for sugarcane production. Sugarcane is tolerant to high temperatures, growing in areas with an average temperature of 47 ºC, with water supply.

The estimation is that sugarcane production in Brazil for 2009/2010 will be around 612 million tons, harvested from an area of 7.5 millions of hectares (Companhia Nacional de Abastecimento – CONAB, 2009) and it is the highest producer in the world (FAO, 2009).

Commercial sugarcane is propagated by asexual reproduction and the germination is initiated from nodes in the stalks. During initial stages of germination, root primordia surrounding the nodes, produce a many roots. These roots are not connected to the development of the shoots, but are important to maintain stalk humidity. The primary shoot is initiated from internodes underground. Each node develops a bud and root primordia, which are essential to the new plant establishment. Root primordia will originate the roots, which are responsible for plant growth maintenance. After this event, shoots become independent of the original stalk (Bull, 2000). In the early stages, stem elongation is fast and the fiber quantity is high, but, at maturation time, this amount tends to decrease and sucrose quantity is increased. During this period, elongated internodes stop growing and carbon metabolites are stored as sucrose. Factors such as age, humidity and nitrogen status affect the sugarcane maturation and environmental factors such as water stress, temperature and nutrition affect sucrose accumulation in the stalks.

Sugarcane flowers are formed in an inflorescence-like opened panicle,, whose shape and size vary according to the variety. Inflorescence consists of a main axis and three more secondary branches. Associated to the branches are the spikelets arranged in pairs, containing individual flowers. Sugarcane flowers

consist of three stamens and a unique carpel with a typical stigma of wind-pollinated flowers (Moore & Nuss, 1987).

Flowering process in sugarcane is divided in four stages: changing of apical meristem to floral bud; floral bud to inflorescence; inflorescence development and inflorescence emission.

Flowering in field conditions is diverse, influenced by both endogenous and environmental factors. Floral initiation provokes the change in apical meristem from vegetative phase to floral development, ceasing stalk elongation, which affects crop yields (Moore & Maretzki, 1996). Floral development is initiated upon shortening of daylength, warm night temperatures and high diurnal temperatures, with high humidity. Flowers take from two to three months to become mature after initiation. Sugarcane is an out-crossing species, but auto-pollination can occur (Moore & Nuss, 1987). Pollen is small, pilous and dispersed by wind. It is rapidly dried after dehiscence, with a half-life of twelve minutes. As a result, sugarcane pollens are not easily dispersed in the field.

The switch from vegetative to reproductive growth interferes in the sucrose accumulation, as well. Flowering in commercial sugarcane presents some inconveniences when related to sugar production, as follows:

- Vegetative growth of stalks is ceased, leading to loss of sugar content;
- Once the life cycle is completed, flowering stalks become senescent;
- Flowered stalks can not be stocked in the field for too long.

Different varieties of sugarcane demonstrate specific rates of sugar content in the stalks. First, sugarcane varieties were mainly *Saccharum officinarum* clones, species with high sucrose content, domesticated from wild specie *S. robustum*. In the past century, sugar content has been enhanced with the crossing of *S. officinarum* and *S. spontaneum*, the latter is vigorous and

5

known for its disease resistance (Berding & Roach, 1987). The sugar content was increased after crossing these hybrids to *S. officinarum* aiming at the production of a modern sugarcane cultivar germoplasm.

2.2 Sugarcane genome

Sugarcane has a complex genome characterized by a high level of ploidy (Sreenivasan et al., 1987). Cultivated varieties have 2n = 100 to 130 chromosomes derived from *S. officinarum* (2n = 80 chromosomes) and *S. spontaneum* (2n = 40 to 128 chromosomes) crossing (Butterfield et al., 2001; D'Hont & Glaszmann, 2001). *Saccharum officinarum* has a basic chromosomal number of x=10, while *S. spontaneum,* x=8 (Grivet & Arruda, 2002). The difference in number of chromosomes between the two species enables the coexistence of distinct chromosomal organizations in a hybrid, and the possibilities of some chromosomes to be inherited from *S. spontaneum* or *S. officinarum,* or a result of ancestral species recombination (Grivet & Arruda, 2002). Despite the high level of ploidy, monoploid genome size of *S. officinarum* and *S. spontaneum* is estimated around 930 Mbp and 750 Mbp, respectively, having approximately the same size of the sorghum (760 Mbp) and around double the size of the rice genome (390 Mbp). Modern cultivars of sugarcane have around 120 chromosomes and a genome of 10,000 Mbp (D'Hont, 2005). As in other grasses, sugarcane has an arrangement of transposons, some of them located between coding genes (Jannoo et al., 2007). A large number of transposons are differently expressed in sugarcane tissues, affecting genomic stability of the crop (Araujo et al., 2005). Expressed transposons include homologues of maize mutator (Mu) and mudrA-like transposase. Sequence coding transposons found in sugarcane EST Project

6

(SUCEST) (Vettore et al., 2001) has been also found in *S. officinarum* and *S. spontaneum*, suggesting that these transposons were present in ancestral species of Andropogoneae tribe common to sorghum, maize and sugarcane (Rossi et al., 2004).

Chromosomal organization in grasses is highly conserved (Bennetzen & Freeling, 1993, 1997; Ming et al., 1998), possibly as a consequence of the short period of independent evolution that the species belonging to the same family diverged from a common ancestor. Comparative mapping demonstrated that sugarcane and sorghum chromosomes are highly syntenic, suggesting that DNA sequence organization in its chromosomes has been conserved (Dufour et al., 1997; Glaszmann et al., 1997; Guimarães et al., 1997). Analysis of genomic sequences of alcohol dehydrogenase in sugarcane and sorghum revealed a high co-linearity in the gene sequence, conserved gene structure and 95% of nucleotide similarity (Janoo et al., 2003). These findings support the comparative mapping results in sorghum studies, showing the syntenic relationship with sugarcane, which makes sorghum a model species for sugarcane studies (Dufour et al., 1997; Glaszmann et al., 1997; Guimarães et al., 1997; Ming et al., 1998). Consequently, access to the sorghum genome and to sugarcane transcriptome databases is an important source of information to further studies on sugarcane genetics. In addition to this, the close phylogenetic relationship between sorghum and maize enables a further source of information about sugarcane based on maize studies.

An important tool for genome research is the transcriptome databank produced by EST sequencing. A large collection of ESTs have been produced recently and are available in public databases (Carson & Botha, 2000; Arruda, 2001; Vettore et al., 2003; Casu et al., 2004). EST banks enable access to millions of expressed genes, identified by advanced technologies, such as

microarray. In addition to this, to use these tools it is desirable to know the complete genome structure of sugarcane, which so far, has not been unravelled.

2.3 Biofuel production from sugarcane

Sugarcane is used for both sugar and alcohol production. Many discussions have been raised about renewable energy since fossil oil is a polluting energy and is not extracted from a renewable source (Farrel et al., 2006; Jacobsson & Johnson, 2000; Dincer, 2000). Similarly, the price of petroleum barrels is increasing and the prediction is that it will become more expensive (Lee & Bressan, 2006).

When the oil crisis occurred in 1970, Brazil started the stimulation of ethanol production from sugarcane, through the PROALCOOL Program. Nine years later, the first car powered by ethanol fuel was released. After that, Brazil became practically economically independent on oil importation. Brazil is considered the most promising country in biofuel production from sugarcane, which is reinforced by its vast land, suitable climate, development of excellent varieties of sugarcane, and advanced biotechnological research on this crop.

Brazil is the largest ethanol producing country using sugarcane as the raw material. In 2009, more than a half of the harvested sugarcane was destined to alcohol production, reaching 55% of the total while for sugar production the average was around 45% (CONAB, 2009).

USA is the second largest ethanol producing country, using corn as the main raw material. Comparing sugarcane and corn, the latter yields around 3000 L/ha while sugarcane yields around 5000 to 7000 L/ha (Lee & Bressan, 2006). Comparing to the other raw materials used to produce bioethanol, sugarcane has the advantage of stocking carbon metabolites as sucrose, instead of starch. Corn and sugar beet, for example, require an additional stage to transform starch to

8

sucrose to, subsequently, ferment it into ethanol. This process increases the production costs from these species, while sugarcane ethanol costs 50% less than oil fuel. Similarly, USA spends 1 unit of oil fuel to generate 1.3 units of ethanol, while in Brazil, the same unit generates 8 or 9 sugarcane ethanol units.

More than 17 million litres of ethanol have been used as fuel each year, a fact that is bringing several environmental benefits such as the reduction of CO_2 emission. The increasing of the internal alcohol demand, due to petroleum environmental problems, has been supported by the national ethanol production from sugarcane.

The Brazilian Bioethanol Science and Technology (CTBE) inaugurated on January 22th, 2010, has the mission "*to contribute to the Brazilian leadership in the renewable energy sources and chemical industry raw material production sectors, mainly by improving the sugarcane bioethanol production chain, through state-of-the-art research, development and innovation*". The CTBE proposes to use and develop methodologies and models aimed at assessing the sustainability of the sugarcane bioethanol production chain.

Therefore, the growing interest in sugarcane bioethanol is stimulating huge investments in new-generation technologies to maintain Brazil as the world leader of bioethanol production. This fact is of crucial importance, since the energy provided is cheap, clean and renewable (Regalado, 2010).

2.4 Flowering process in sugarcane and *arabidopsis*

In *Arabidopsis* three many genes are involved in the control of floral transition. Genes that act at important regulatory points include *Flowering Locus T* (*FT*), *Suppressor of Overexpression of Constans 1* (*SOC1*) and *LEAFY* (*LFY*) Together they promote floral meristem identity gene activation, like *APETALA1*

(*AP1*), *CAULIFLOWER* (*CAL*) and *LFY*, and floral organ identity genes (Figure 1).
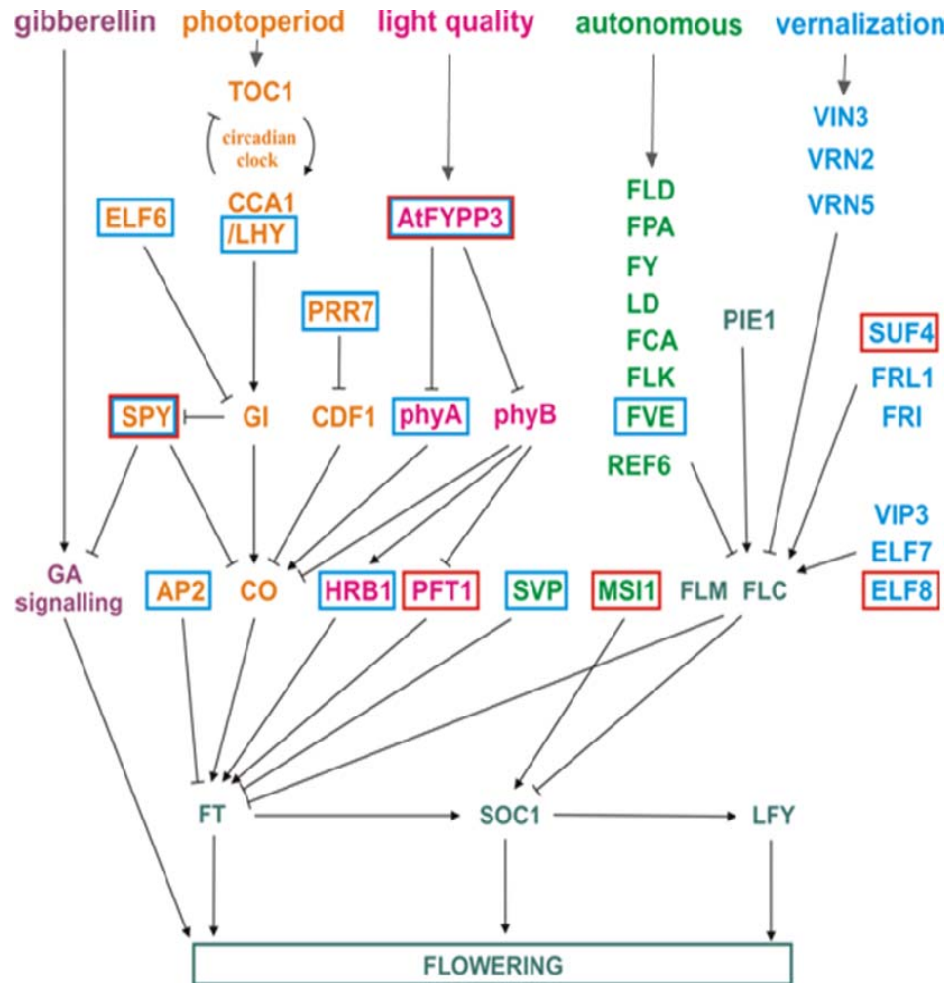
FIGURE 1 A simplified schematic view of the flowering pathways in *Arabidopsis* (Mouhu, 2009).

Flowering is the most important developmental switch in higher plants. The molecular mechanism of flowering has been widely studied in the model

plant *Arabidopsis thaliana* and several flowering time genes have been identified. So far, four pathways of flowering transition have been defined: autonomous, gibberellin (GA), photoperiod (involving light quality) and vernalization pathways.

Higher plants can be classified as long-day, short-day or neutral-day plants, according to their flowering response to specific photoperiods. Long and short-days plants require a specific photoperiod to starts the flowering process, while neutral-day plants flower after a particular number of leaves are formed. The latter plants are called autonomously flowering plants, as well.

Arabidopsis is a long-day plant, which means that flowering is accelerated by long-day exposures. This pathway response is composed of two components, circadian clock and light quality. The *CONSTANS* gene (CO) is a target of both signals, encoding a zinc-finger transcription factor. The autonomous pathway is composed by internal signals in which *FLF/FLC* seem to be the downstream target of this pathway, encoding a protein similar to MADS domain. The *LFY* gene plays a central role in the flowering cascade, regulating the final stages of flowering by activating the *AP1* and other homeotic genes in the shoot apical meristem (SAM) and leaf primordia (Wagner et al., 1999).

Flowering mechanism in higher plants is activated by a signal probably originated in the leaves in response to a critical photoperiod and/or other external factors. This signal seems to be a mobile regulator called "florigen", which is, then, translocated to the apical meristem, where the floral initiation process is triggered. The SAM is a group of indeterminate stem cells that initiate the lateral organ formation (Evans et al., 1997). The molecule "florigen" could not be isolated or identified, but recently FT has been widely believed to be a

florigen in Arabidopsis and possibly in rice (Bernier, 1988; Colasanti, 2005; Corbesier, 2007; Tamaki et al., 2007; Turck et al., 2008).

Flowering in sugarcane is a complex process that involves phytochrome, phytohormones, "florigen", environmental and endogenous factors (Castro, 1984). Apical meristem changes to inflorescence during short days, which means that sugarcane is a short day crop.

Despite several commercial varieties of seeds, they are only used for breeding programs, since the proportion of sugarcane seedlings with agronomic qualities related to parents are extremely low. Consequently, the sugarcane crop is reproduced vegetatively, resulting in clones with desirable features.

As the flowering in *Saccharum* is dependent on interaction of cultivars and environmental factors, some varieties can flower abundantly in their natural environment, but do not flower if they is in different regions (Bull & Glasziou, 1979). Several techniques have been developed to study the flowering process in sugarcane, such as photoperiod change to induce flowering, making the flowers available to cross when it is necessary (Bull & Glasziou, 1979).

Sugarcane is a short-day plant that needs to be exposed to inductive photoperiod and temperature, since leaf perception leads to substance accumulation required for the subsequent dark reactions. Then, the florigen is translocated to the apical meristem and accumulated in the SAM until a critical level of flowering substances is reached and consequently the floral bud differentiation is initiated.

Although there are many studies about the flowering time genes, so far, none of the characterized genes seemed to have a specific role in the signal transmission from the leaves to the SAM (Colasanti & Sundaresan, 2000). Corbesier (2007) discovered that FT protein movement can contribute to long-signaling for floral induction in *Arabidopsis*, since *FT* gene is expressed in the

12

leaves in response to the photoperiod, and then FT protein migrates to the apex meristem, activating flower identity genes.

More studies need to be done to establish whether this system operates in other plants and/or there is an alternative pathway to be discovered.

2.4 Flowering time gene *id1*

The flowering time gene *INDETERMINATE1* (*ID1),* previously characterized in maize, encodes nuclear-localized zinc finger proteins Cys2/His2, which perform a primary role in many biological processes (Colasanti et al., 2006; Agarwal et al., 2007; Englbrecht et al., 2004). Zinc finger proteins represent one of the largest classes of transcription factors in higher plants.

The *ID1* gene product binds to a specific DNA sequence, suggesting that ID1 acts in controlling the expression of other genes (Kozaki et al., 2004). The expression pattern of *ID1* in immature maize leaves exclusively suggests that it may act either in producing or transmitting the flowering signal from leaves to the SAM. Studies with *id1*maize mutants showed that plants were unable to flower or flowered much later than wild-type plants. These findings suggest that this gene may regulate the production of a floral stimulus from the leaves to the SAM (Colasanti et al., 1998).

In addition to *ID1*, at least, two other genes, *dlf1* and *vgt1*, identified in maize, play a role in controlling the transition to reproductive growth, since mutations in these genes postponed the floral transition (Colasanti & Muszynski, 2009). *ID1* belongs to a gene family composed of four zinc finger motifs called the ID-domain (IDD) (Figure 2). ID proteins and ID1-like protein have been isolated from rice and sorghum as well, suggesting that this protein has been conserved within monocot plants. There is a high similarity of *ID1* gene in

sorghum (*SbID1*) and rice (*OsID1*). Functional analysis of *OsID1,* also called as *Ehd2* and *RID1,* showed that this gene regulates the flowering process in rice, as well (Matsubara et al., 2008; Wu et al., 2008; Park et al., 2008).
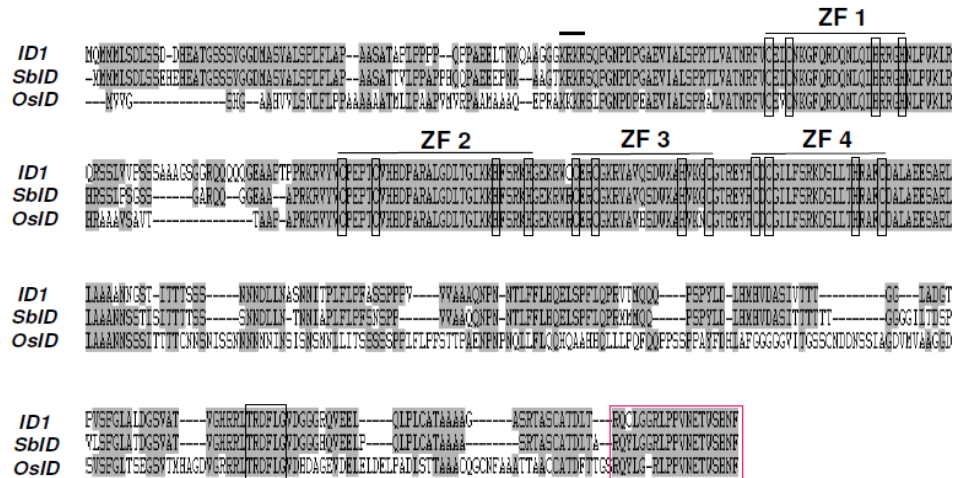
FIGURE 2 Peptide sequences of *ID1*, *Oryza sativa ID1* (*OsID)* and *Sorghum bicolor ID1* (*SbID)* showing the four zinc fingers of the ID-domain; the NLS region is in the dark bar (Colasanti et al., 2006).

Similarly to *FT* gene, *ID1* is expressed in leaves but no expression is noticed in the SAM, indicating that *ID1* gene could be related to the floral stimulus in maize and/or other grasses. *ID1* downstream target genes were screened aiming at verifying whether any of these genes are related to ortholog flowering time genes previously characterized in other species. So far, studies developed in maize, showed that there are homologs of *CO/FT* genes in maize. Expression patterns of the gene *ZCN8* in this species suggests that it may play a role as a FT ortholog in maize (Danilevskya et al., 2008).

3 REFERENCES

AGARWAL, P.; ARORA, R.; RAY, S.; SINGH, A. K.; SINGH, V. P.; TAKATSUJI, H.; KAPOOR, S.; TYAGI, A. K. Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. **Plant Molecular Biology**, Dordrecht, v. 65, p. 467-485, July 2007.

ARAUJO, P. G.; ROSSI, M.; JESUS, E. M.; SACCARO JUNIOR, N. L.; KAJIHARA, D.; MASSA, R.; FELIX, J. M.; DRUMMOND, R. D.; FALCO, M. C.; CHABREGAS, S. M.; ULIAN, E. C.; MENOSSI, M.; SLUYS, M-A. V. Transcriptionally active transposable elements in recent hybrid sugarcane. **The Plant Journal**, London, v. 44, p. 707-717, Dec. 2005.

ARRUDA, P. Sugarcane transcriptome: a landmark in plant genomics in the tropics. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, p. 1-2, Jan. 2001.

BAURLE, I.; DEAN, C. The timing of developmental transitions in plants. **Cell**, Cambridge, v. 125, p. 655-664, May 2006.

BENNETZEN, J. L.; FREELING, M. Grasses as a single genetic system: genome composition, collinearity and compatibility. **Trends in Genetics**, Netherlands, v. 9, p. 259-261, Aug. 1993.

BENNETZEN, J. L.; FREELING, M. The unified grass genome: synergy in synteny. **Genome Research**, Cold Spring Harbor, v. 7, p. 301-306, Apr. 1997.

BERDING, N.; ROACH, B. T. Germplasm collection, maintenance, and use. In: HEINZ, D. J. (Ed.). **Sugarcane improvement through breeding**. Amsterdam: Elsevier, 1987. chap. 4, p. 143-210.

BERNIER, G. The control of floral evocation and morphogenesis. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 39, p. 175-219, June 1988.

BULL, T. A.; GLASZIOU, K. T. Sugarcane. In: LOVETT, J. V.; LAZENBY, A. (Ed.). **Australian field crops.** Melbourne: Angus and Robertson, 1979. p. 95-113.

BULL, T. The sugarcane plant. In: HOGARTH, M.; ALLSOPP, P. (Ed.). **Manual of cane growing**. Australia: Bureau of Sugar Experimental Stations, Indooroopilly, 2000. chap. 4, p. 71-83.

BUTTERFIELD, M. K.; D'HONT, A.; BERDING, N. The sugarcane genome: a synthesis of current understanding, and lessons for breeding and biotechnology. **Proceedings of South African Sugar Technologists Association**, Mount Edgecombe, v. 75, p. 1-5, July/Aug. 2001.

CARSON, D. L.; BOTHA, F. C. Preliminary analysis of expressed sequence tags for sugarcane. **Crop Science,** Madison, v. 40, p. 1769-1779, Nov./Dec. 2000.

CASTRO, P. R. C. Controle da florescência da cana-de-açúcar. **Álcool & Açúcar**, São Paulo, v. 4, n. 17, p. 44-49, 1984.

CASU, R. E.; DIMMOCK, C. M.; CHAPMAN, S. C.; GROF, C. P.; MCINTYRE, C. L.; BONNETT, G. D.; MANNERS, J. M. Identification of differentially expressed transcripts from maturing stem of sugarcane by *in silico* analysis of stem expressed sequence tags and gene expression profiling. **Plant Molecular Biology**, Dordrecht, v. 54, p. 503-517, Mar. 2004.

COLASANTI, J.; CONEVA, V. Mechanisms of floral induction in grasses: something borrowed something new. **Plant Physiology**, Rockville, v. 149, p. 56-62, Jan. 2009.

COLASANTI, J. Decoding the floral stimulus: what's next? **Flowering Newsletter**, Liege, v. 40, p. 24-26, Sept. 2005.

COLASANTI, J.; MUSZYNSKI, M. G. The maize floral transition. In: HAKE, S. C.; BENNETZEN, J. L. (Ed.). **Handbook of maize**: its biology. New York: Springer Science, 2009. v. 1, p. 41-55.

COLASANTI, J.; TREMBLAY, R.; WONG, A. Y.; CONEVA, V.; KOZAKI, A.; MABLE, B. K. The maize *Indeterminate1* flowering time regulator defines a highly conserved zinc finger protein family in higher plants. **BMC Genomics**, London, v. 7, p. 158-175, June 2006.

COLASANTI, J.; SUNDARESAN, V. Florigen enters the molecular age: long-distance signals that cause plants to flower. **Trends in Biochemical Sciences**, Netherlands, v. 25, p. 236-240, May 2000.

COLASANTI, J.; YUAN, Z.; SUNDARESAN, V. The *indeterminate* gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. **Cell**, Cambridge, v. 93, p. 593-603, May 1998.

COMPANHIA NACIONAL DE ABASTECIMENTO. **1º Levantamento de cana de açúcar.** 2009. Available in: <http://www.conab.gov.br/conabweb/>. Accessed on: mar. 21, 2010.

CORBESIER, L. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. **Science**, Washington, v. 316, p. 1030-1033, Sept. 2007.

DANILEVSKAYA, O. N.; MENG, X.; HOU, Z.; ANANIEV, E. V.; SIMMONS, C. R. A genomic and expression compendium of the expanded PEBP gene family from maize. **Plant Physiology**, Rockville, v. 146, p. 250-264, Jan. 2008.

D'HONT, A.; GLASZMANN, J. C. Sugarcane genome analysis with molecular markers, a first decade of research. **Proceedings of International Society of Sugarcane Technologists**, Australia, v. 24, p. 556-559, 2001.

D'HONT, A.; ISON, D.; ALIX, K.; ROUX, C.; GLASZMANN, J. C. Determination of basic chromosome numbers in the genus Saccharum by physical mapping of ribosomal RNA genes. **Genome**, Ottawa, v. 41, p. 221-225, 1998.

D'HONT, A. Sugarcane: a major source of sweetness, alcohol and bio-energy. In: MOORE, P. H.; MING, R. (Ed.). **Tropical crop plants genomics.** New York: Springer-Verlag, 2008. chap. 21, p. 483-513.

D'HONT, A. Unravelling the genome structure of polyploids using fish and gish; examples of sugarcane and banana. **Cytogenetic and Genome Research**, Basel, v. 109, n. 1/3, p. 27-33, 2005.

DINCER, I. Renewable energy and sustainable development: a crucial review. **Renewable and Sustainable Energy Reviews,** Golden, v. 4, p. 157-175, June 2000.

DUFOUR, P.; DEU, M.; GRIVET, L.; D'HONT, A.; PAULET, F.; BOUET, A.; LANAUD, C.; GLASZMANN, J. C.; HAMON, P. Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. **Theoretical and Applied Genetics**, Berlim, v. 94, p. 409-418, Mar. 1997.

ENGLBRECHT, C. C.; SCHOOF, H.; BOHM, S. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. **BMC Genomics**, London, v. 5, p. 39-56, July 2004.

EVANS, M. M. S.; BARTON, M. K. Genetics of angiosperm shoot apical meristem development. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 48, p. 673-701, June 1997.

FARREL, A. E.; PLEVIN, R. J.; TURNER, B. T.; JONES, A. D.; O'HARE, M.; KAMMEN, D. M. Ethanol can contribute to energy and environmental goals. **Science**, Washington, v. 311, p. 506-508, Jan. 2006.

FOOD AND AGRICULTURAL ORGANIZATION. **Major food and agricultural commodities and producers.** 2009. Available in: <http://www.fao.org/es/ess/top/commodity.html?lang=en&item=156&year=2005>. Accessed on: dec. 16, 2009.

GLASZMANN, J. C.; DUFOUR, P.; GRIVET, L.; D'HONT, A.; DEU, M.; PAULET, F.; HAMON, P. Comparative genome analysis between several tropical grasses. **Euphytica**, Netherlands, v. 96, p. 13-21, July 1997.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current Opinion in Plant Biology**, London, v. 5, p. 122-127, Apr. 2002.

GUIMARÃES, C. T.; SILLS, G. R.; SOBRAL, B. W. S. Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 94, p. 14261-14266, Dec. 1997.

JACOBSSON, S.; JOHNSON, A. The diffusion of renewable energy technology: an analytical framework and key issues for research. **Energy Policy**, Oxford, v. 28, p. 625-640, July 2000.

JANNOO, N.; GRIVET, L.; CHANTRET, N.; GARSMEUR, O.; GLASZMANN, J. P.; ARRUDA, P.; D'HONT, A. Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. **The Plant Journal**, England, v. 50, p. 574-585, Apr. 2007.

JANOO, N.; GRIVET, L.; D'HONT, A.; ARRUDA, P. Genomic sequencing in sugarcane: first insight into the physical organization of the genome and mycrosynteny with other grasses. In: PLANT AND ANIMALS GENOMES CONFERENCE, 11., 2003, San Diego. **Abstracts…** Available in: <http://www.intlpag.org/pag/11/abstracts/W26_W182_XI.html>. Accessed on: dec. 13, 2009.

KOZAKI, A.; HAKE, S.; COLASANTI, J. The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. **Nucleic Acids Research**, Oxford, v. 32, p. 1710-1720, Mar. 2004.

LEE, T. S. G.; BRESSAN, E. A. The potencial of ethanol production from sugarcane in Brazil. **Sugar Tech**, India, v. 8, p. 195-198, Dec. 2006.

MATSUBARA, K.; YAMANOUSHI, U.; WANG, Z-X.; MINOBE, Y.; IZAWA, T.; YANO, M. *Ehd2*, a rice ortholog of the maize *Indeterminate1* gene, promotes flowering by up-regulating *Ehd1*. **Plant Physiology**, Rockville, v. 148, p. 1425-1435, Nov. 2008.

MING, R.; LIU, S. C.; LIN, Y. R.; SILVA, J.; WILSON, W.; BRAGA, D.; DEYNZE, V. A.; WENSLAFF, T. F.; WU, F. F.; MOORE, P. H.; BURNQUIST, W.; SORRELLS, M. E.; IRVINE, J. E.; PATERSON, A. H. Alignment of the *Sorghum* and *Saccharum* chromosomes: comparative genome organization and evolution of a polysomic polyploid genus and its diploid cousin. **Genetics**, Baltimore, v. 150, p. 1663-1682, Dec. 1998.

MOORE, P. H.; MARETZKI, A. Sugarcane. In: ZAMSKI, E.; SCHAFEER, A. A. (Ed.). **Photoassimilate distribution in plants and crops:** source–sink relationships. New York: Marcel Dekker, 1996. p. 643-669.

MOORE, P. H.; NUSS, K. J. Flowering and flower synchronization. In: HEINZ, D. J. (Ed.). **Sugarcane improvement through breeding.** New York: Elsevier, 1987. p. 273-311.

MOUHU, K.; HYTONEN, T.; FOLTA, K.; RANTANEN, M.; PAULIN, L.; AUVINEN, P.; ELOMAA, P. Identification of flowering genes in strawberry, a perennial SD plant. **BMC Plant Biology**, London, v. 9, p. 122-137, Sept. 2009.

MOURADOV, A.; CREMER, F.; COUPLAND, G. Control of flowering time: interacting pathways as a basis for diversity. **Plant Cell**, Rockville, v. 14, p. 111-130, May 2002.

PARCY, F. Flowering: a time for integration. **International Journal of Developmental Biology**, Bilbao, v. 49, n. 5/6, p. 585-593, 2005.

PARK, S. J.; KIM, S. L.; LEE, S.; JE, B. I.; PIAO, H. L.; PARK, S. H.; KIM, C. M.; RYU, C. H.; PARK, C. H.; XUAN, Y-H.; COLASANTI, J.; AN, G.; HAN, C. D. Rice *Indeterminate1* (*OsID1*) is necessary for the expression of *Ehd1* (*Early heading date 1*) regardless of photoperiod. **Plant Journal**, Oxford, v. 56, p. 1018-1029, Oct. 2008.

REGALADO, A. Race for cellulosic fuels spurs Brazilian research program. **Science**, Washington, v. 327, p. 928-929, Feb. 2010.

ROSSI, M.; ARAUJO, P. G.; JESUS, E. M.; VARANI, A. M.; SLUYS, M. A. V. Comparative analysis of *Mutator*-like transposases in sugarcane. **Molecular Genetics and Genomics**, New York, v. 27, p. 194-203, Sept. 2004.

SIMPSON, G. G.; DEAN, C. *Arabidopsis*, the Rosetta stone of flowering time? **Science**, Washington, v. 296, n. 5566, p. 285-289, 2002.

SREENIVASAN, T. V.; AHLOOWALIA, B. S.; HEINZ, D. J. Cytogenetics. In: HEINZ, D. J. (ED.). **Sugarcane improvement through breeding**. New York: Elsevier, 1987. p. 211-253.

TAMAKI, S.; MATSUO, S.; WONG, H. L.; YOKOI, S.; SHIMAMOTO, K. Hd3a protein is a mobile flowering signal in rice. **Science**, Washington, v. 316, p. 1033-1036, May 2007.

TEW, T. L.; COBILL, R. M. Genetic improvement of sugarcane (Saccharum ssp.) as an energy crop. In: VERMERRIS, W. (Ed.). **Genetic improvement of bioenergy crops**. New York: Springer, 2008. chap. 9, p. 273-294.

TURCK, F.; FORNARA, F.; COUPLAND, G. Regulation and identity of florigen: Flowering Locus T moves center stage. **Annual Review of Plant Biology**, Palo Alto, v. 59, p. 573-594. 2008.

VETTORE, A. L.; SILVA, F. R. da; KEMPER, E. L.; ARRUDA, P. The libraries that made Sucest. **Genetics and Molecular Biology,** Ribeirão Preto, v. 24, n. 1/4, p. 1-7, 2001.

VETTORE, A. L.; SILVA, F. R.; KEMPER, E. L.; SOUZA, G. M.; SILVA, A. M.; FERRO, M. I. T.; HENRIQUE-SILVA, F.; GIGLIOTI, E. A.; LEMOS, M. V. F.; COUTINHO, L. L.; NOBREGA, M. P.; CARRER, H.; FRANÇA, S. C.; BACCI JUNIOR, M.; GOLDMAN, M. H. S.; GOMES, S. L.; NUNES, L. R.; CAMARGO, L. E. A.; SIQUEIRA, W. J.; SLUYS, M-A. V.; THIEMANN, O. H.; KURAMAE, E. E.; SANTELLI, R. V.; MARINO, C. L.; TARGON, M. L. P. N.; FERRO, J. A.; SILVEIRA, H. C. S.; MARINI, D. C.; LEMOS, E. G. M.; MONTEIRO-VITORELLO, C. B.; TAMBOR, J. H. M.; CARRARO, D. M.; ROBERTO, P. G.; MARTINS, V. G.; GOLDMAN, G. H.; OLIVEIRA, R. C.; TRUFFI, D.; COLOMBO, C. A.; ROSSI, M.; ARAUJO, P. G.; SCULACCIO, S. A.; ANGELLA, A.; LIMA, M. M. A.; ROSA JUNIOR, V. E.; SIVIERO, F.; COSCRATO, V. E.; MACHADO, M. A.; GRIVET, L.; MAURO, S. M. Z. D.; NOBREGA, F. G.; MENCK, C. F. M.; BRAGA, M. D. V.; TELLES, G. P.; CARA, F. A. A.; PEDROSA, G.; MEIDANIS, J.; ARRUDA, P. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, Cold Spring Harbor, v. 13, p. 2725-2735, Nov. 2003.

WAGNER, D.; SABLOWSKI, R. W. M.; MEYEROWITS, E. M. Transcriptional activation of *Apetala1* by *Leafy*. **Science**, Washington, v. 285, p. 582-584, July 1999.

WU, C.; YOU, C.; LI, C.; LONG, T.; CHEN, G.; BYRNE, M. E.; ZHANG, Q. *RID1*, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 105, p. 12915-12920, Sept. 2008.

CHAPTER 2 – *IN SILICO* ANALYSIS OF *ID1* GENE IN SUGARCANE BY SUCEST DATABASE

1 ABSTRACT

ID1 gene is a flowering time gene characterized in maize, that belongs to the family of the ID domain. The aim of this work was to identify the *ID* genes in sugarcane by the sugarcane EST databank (SUCEST). First of all, a search for reads was done in the Gene Project plataform and the *reads* were clustered in EST-contigs and, then, analysed as putative sugarcane ID sequences. After a manual annotation, the sequences were screened for the presence of ID conserved domain, surrounding the NLS region and zinc finger domains. Five contigs and four singlets were selected to be used in this work, considered putative sugarcane ID sequences. The alignment was performed comparing the putative sugarcane *ID* sequences, maize *ID1* and rice and sorghum *ID1-like*. A pairwise comparison was done to verify the percentage of similarity between pairs of sequences. The phylogenetic analysis showed that Singlet 1 is the closest related to the maize, rice and sorghum *ID1*, but it does not present the region surrounding the NLS, because most of the sequences found in the SUCEST were incomplete. From these results it can be concluded that all the sequences might be putative ID sequences in sugarcane, but the entire gene sequences need to be identified.

2 INTRODUCTION

Large scale ESTs have been used to rapidly access sequence information. ESTs from more than 1000 species comprise the EST databank (dbEST) in the National Center for Biotechnology Information (NCBI). Plants represent 18% of all ESTs in dbEST, with half of the sequences being from rice, wheat, maize, sugarcane, sorghum and barley. This amount of grass transcript information, especially from sugarcane, sorghum and maize, provides a good opportunity to compare expression profiles between the species. Characteristics associated with metabolism, development, and biotic and/or abiotic stress responses can be accessed by comparative analysis of gene expression between related species (Vettore et al., 2001). 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 (apical meristem, flowers, leaf roll, seeds, internode, stem bark, etiolated leaves, lateral bud, stem bark, etc.) (Vettore et al., 2001). These ESTs were organized in contigs, named as sugarcane assembled sequences (SAS), which allowed the finding of more than 30,000 sugarcane genes (Vettore et al., 2003). Transcriptome comparison of sugarcane with other monocots and dicots showed that genes shared between these plants have been highly conserved, despite the long period of independent evolution (Vincentz et al., 2004). Therefore, approximately 71% of sugarcane proteins are similar to *Arabidopsis* proteins, and 82% are similar to rice. Approximately, 11% of SASs may correspond to monocots specific genes, while 18% of SASs may represent sugarcane proteins (Vincentz et al., 2004). SUCEST database analysis revealed genes coding a wide range of functions associated to important agronomic features (Vettore et al., 2003). SASs were grouped in 18 categories related to

24

their biological roles in many cellular processes or routes. Almost 50% of all SAS were associated in 5 categories: cellular dynamics, stress response, protein metabolism, bioenergetics and cellular communication/signal transduction. One category "unable to classify" comprising 17% of all SAS categorized correspond to genes with unknown roles. Proteins were also categorized according to the presence and/or number of conserved domains. SUCEST allowed a preliminary analysis of tissue-specific expression profiles. In general, 60% of transcription factors found in *Arabidopsis* or rice are present in SUCEST as well (Vettore et al., 2003).

More than 3,500 genes involved in different aspects of signal transduction, transcription, development, cellular cycle, stress response and pathogen interaction were identified in SUCEST (Souza et al., 2001). Expression profile by microarray of these signalling compounds in cultivated plants showed an expression pattern of different signalling proteins in flowers, roots, leaves, lateral buds and internodes (Papini-Terzi et al., 2005). To identify the flowering time gene *ID1* in sugarcane, an *in silico* analysis was performed comparing the sequences found at SUCEST database with already known *ID1* sequences from maize, rice and sorghum.

3 MATERIAL AND METHODS

3.1 Search for EST-contigs in Genbank

A search for *ID1* sequences in related species was performed by keyword at NCBI database and obtained sequences were annotated. Additionally, a search using BLAST (*Basic Local Alignment Search Tool*) in the same database was performed using maize *ID1* sequence as query. The sequences, presenting significant alignment (e-value $< 10^{-5}$), were annotated and used to compare with sequences deposited in the SUCEST database.

3.2 Comparison of sequences found at Genbank with sequences deposited at SUCEST database

The sequences found at NCBI were confronted using *blastn* at the *The Gene Index Project* (Quackenbush et al., 2001) database, where sugarcane EST sequences are deposited. Sequences were grouped in EST-contigs and annotated using *CLC Main WorkBench 5.1* software (CLC bio®). After assembling the sequences, they were analyzed to select those containing the ID domain, involving the nuclear localization sequence (NLS) and the four zinc fingers (Colasanti et al., 2006). Only sequences involving the conserved domain were selected for subsequent analysis.

3.3 Phylogenetic analysis of possible ID1 sequences with related species

EST-contigs alignments were performed using CLC Main Workbench 5.6.1 following the default parameters, using translated sequences. A phylogenetic tree was constructed using *CLC Main Workbench 5.6.1* software with the default parameters, *Neighbor-Joining* comparison model, distance *p*

26

method and *pairwise* supression. The phylogenetic distance of clusters was measured by *bootstraps*.

4 RESULTS AND DISCUSSION

Sequences presenting the ID-domain (*IDD* sequences) were identified through the SUCEST database. 20 reads were found, using *ID1* gene as query. The clusterization process generated 9 EST-contigs. The protein sequence of each putative sugarcane *IDD* gene found was analyzed to find the conserved domain and motifs.

Although some of the selected sequences did not present all of the four zinc fingers, sequences presenting at least one of the zinc fingers were kept. A saturation process was performed aiming at saturating the search and possibly get the entire gene, but, still, most of the sequences were incomplete. This can be related to the fact that, although the SUCEST database comprises a large amount of deposited ESTs, cDNA clones used to create the libraries were only partially sequenced (Vettore et al., 2003).

The homology among sugarcane contigs with deposited *IDD* sequences were analyzed using *tblastn* tool at the NCBI database (Table 1).

TABLE 1 NCBI *tblastn* result with identified sugarcane EST-contigs

	Sequence	Size	Identity	e-value
Contig 1 **312 aa**	*Zea mays* INDETERMINATE-related PROTEIN 10	583	96%	2e-133
Contig 2 **277 aa**	*Zea mays* clone 506559 INDETERMINATE-related protein 1	543	88%	6e-95
Contig 3 **153 aa**	*Lolium multiflorum* putative zinc finger protein ID1	407	75%	1e-33
Contig 4 **217 aa**	*Zea mays* INDETERMINATE-related protein 10	583	52%	1e-51
Contig 5 **202 aa**	*Arabidopsis thaliana* AtIDD11	513	82%	1e-51
Singlet 1 **222 aa**	*Zea mays* clone 506559 INDETERMINATE-related protein 1	543	81%	6e-72
Singlet 2 **148 aa**	*Arabidopsis thaliana* AtIDD2 (Arabidopsis thaliana Indeterminate(ID)	452	55%	1e-24
Singlet 3 **148 aa**	*Zea mays* INDETERMINATE-related protein 10	583	66%	1e-38
Singlet 4 **167 aa**	*Zea mays* INDETERMINATE-related protein 10	583	65%	5e-54

As shown in Table 1, sugarcane EST-contigs were related with maize and *IDD* genes from *Arabidopsis* and maize.

Colasanti et al. (2006) showed the alignment among *IDD* genes of maize, sorghum and rice, where the ID domain is highly conserved. This result supports the findings that sugarcane might have putative *IDD* genes, since all of these species presents a close phylogenetic relationship (Bouchenak-Khelladi et al., 2008).

It is to be classified yet which of each EST-contig correlates with a specific member of the IDD family. With this propose, all the genes need to be identified.

The *ID1* gene can be differentiated from the *IDD* genes because the presence of a spacer between ZF1 and ZF2 of *ID1* that is absent in *IDD* genes (Figure 1).

FIGURE 1 Alignment of deduced ID-domain amino acid sequences of maize, rice and *Arabidopsis* and PCP1, an *IDD* gene from potato (Colasanti et al., 2006). (continue on the next page)

According to this, none of the sequences found could be classified as a possible *ScID1* gene, because they do not present the spacer, but they are candidates for sugarcane *IDD* genes.
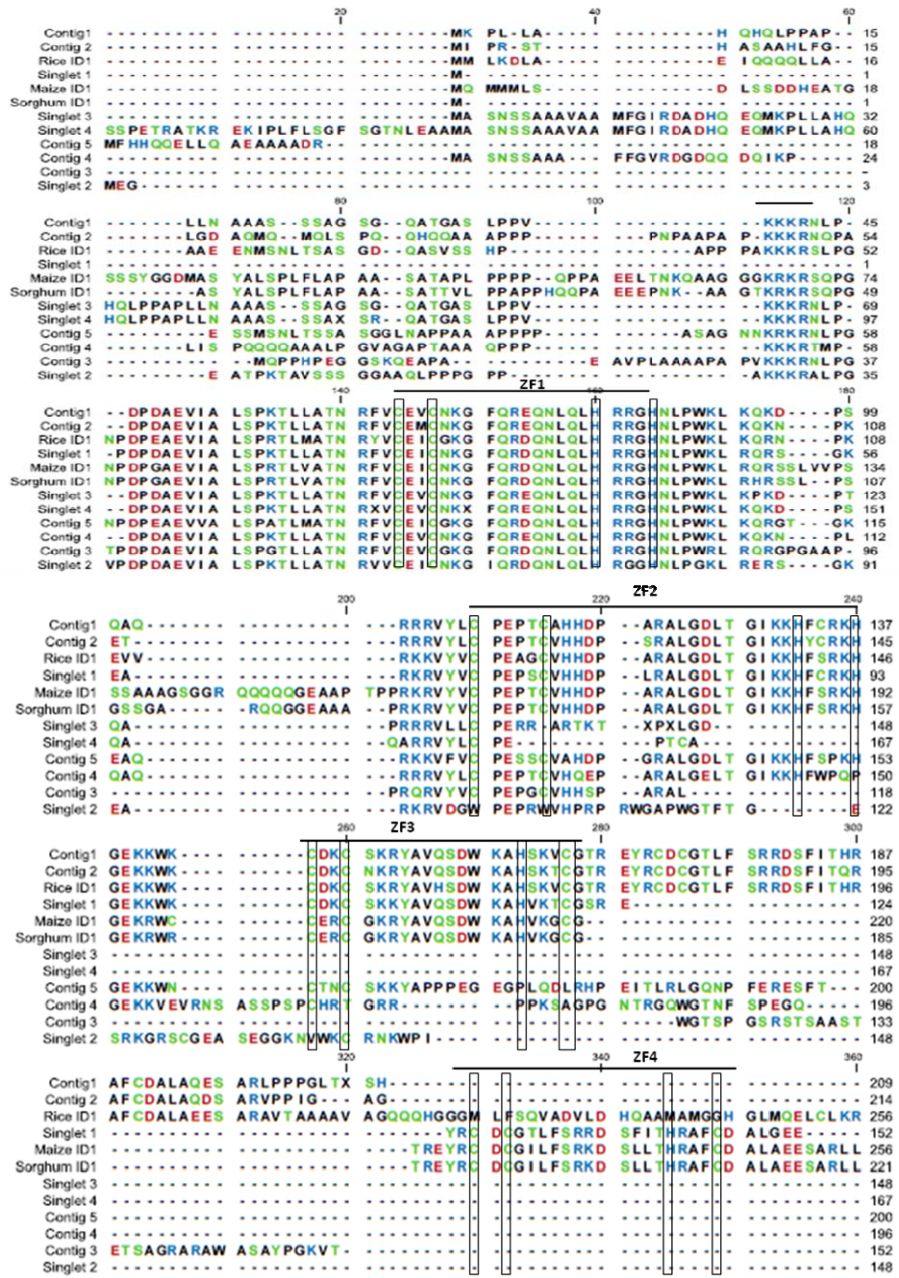
FIGURE 2 Alignment of sugarcane EST-contigs with maize, rice and sorghum *ID1* sequences found in NCBI Genbank. Amino acids shared by at

34

least two sequences are shown in the same color; the putative NLS is shown by a dark bar. Each zinc finger is indicated as ZF1, ZF2, ZF3 and ZF4, and the boxed areas show the C and H residues.

All of the sugarcane EST-contigs presented, at least two of the zinc fingers, and, except for Singlet 1, showed the putative nuclear localization signal, which marks the beginning of the ID-domain. This evidence suggests that they comprise genes from the *IDD* family, even though just partial sequences were found in the SUCEST.

Two of the zinc fingers, ZF2 and ZF3, have the hallmarks of the TFIIIA zinc finger class of proteins, which binds to a specific DNA sequence, whereas the other two zinc fingers, ZF1 and ZF4, have atypical structures, possibly working as mediating protein interactions (Kozaki et al., 2004).

The NLS region consists of charged lysines (K) and arginine (R) residues in the N-terminal border (Raikhel, 1992). So far, in maize and *Arabidopsis* sixteen *IDD* genes have been identified whereas in rice fourteen *IDD* genes, with at least 61% of similarity with *ID1* gene (Colasanti et al., 2006).

Although there is a good indication of the presence of an *ID-like* gene in sugarcane, none of the sequences found presented the characteristics of the *ID1* gene.

Phylogenetic tree comparison of maize, sorghum and rice *ID1* with sugarcane ID-like sequences (Figure 3) show that the *ID1* from the three species are closely related to each other, which it was previously published by Colasanti et al. (2006). Comparing the sequence of Singlet 1 with the *ID1* gene, it was noticed that although it is classified as the closest sequence, it is a partial sequence that is missing the NLS, but involves all the four zinc fingers.
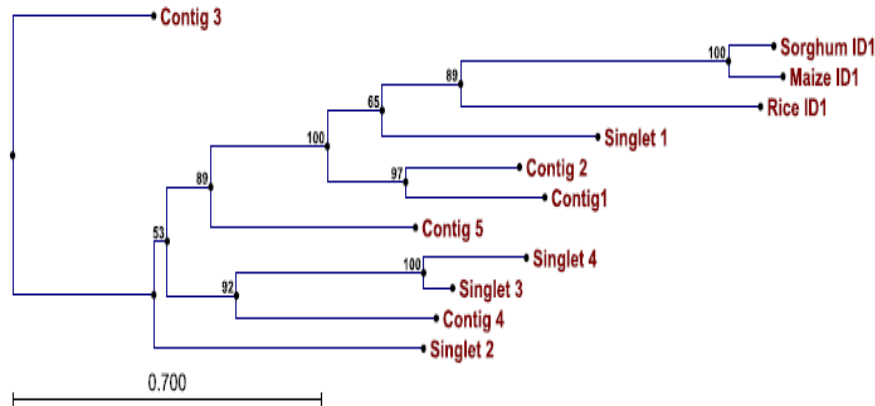
FIGURE 3 Phylogenetic comparison of possible *ScID* with *ID1*, *SbID* and *OsID* (Acession number for *ID1*, *SbID* and *OsID* are 162461280, 242039639 and 55773665, respectively).

The Contig 1 and 2 are the largest sequences found, involving all the IDD motifs. The high similarity of genes in different species, mainly those involving conserved domains, indicates that the function of the sequences might be conserved, and these genes are called ortholog genes. Orthology is the relationship of two characters where the sequence divergence follows the speciation (Fitch, 1970).

According to this, the IDD sequences found in the SUCEST might have functions related to the ZmIDD, AtIDD and OsIDD. Tanimoto et al. (2008) characterized one of the AtIDD genes, the *AtIDD15* gene, proposing that this gene regulates the transcription of other genes required for the starch metabolism, playing a role at amyloplast sedimentation.

5 CONCLUSIONS

After an *in silico* analysis at the SUCEST database it was possible to identify nine *IDD* gene candidates. These findings allowed the identification of the motifs and conserved domains of the possible *IDD* sequences found in the SUCEST, indicating that they could be members of the sugarcane IDD family.

6 REFERENCES

BOUCHENAK-KHELLADI, Y.; SALAMIN, N.; SAVOLAINEN, V.; FOREST, F.; BANK, M.; CHASE, M. W.; HODKINSON, T. R. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. **Molecular Phylogenetics and Evolution**, Orlando, v. 47, n. 2, p. 488-505, 2008.

COLASANTI, J.; TREMBLAY, R.; WONG, A. Y.; CONEVA, V.; KOZAKI, A.; MABLE, B. K. The maize *Indeterminate1* flowering time regulator defines a highly conserved zinc finger protein family in higher plants. **BMC Genomics**, London, v. 7, p. 158-175, June 2006.

FIGUEIREDO, R. C.; BRITO, M. S.; FIGUEIREDO, L. H. M.; QUIAPIN, A. C.; VITORELLI, P. M.; SILVA, L. R.; SANTOS, R. V.; MOLFETTA, J. B.; GOLDMAN, G. H.; GOLDMAN, M. H. S. Dissecting the sugarcane expressed sequence tag (SUCEST) database: unravelling flower-specific genes. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1/4, p. 77-84, 2001.

FITCH, W. M. Distinguishing homologous from analogous proteins. **Systematic Zoology**, Washington, v. 19, p. 99-113, June 1970.

KOZAKI, A.; HAKE, S.; COLASANTI, J. The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. **Nucleic Acids Research**, Oxford, v. 32, p. 1710-1720, Mar. 2004.

PAPINI-TERZI, F. S.; ROCHA, F. R.; VENCIO, R. Z.; OLIVEIRA, K. C.; FELIX, J. M.; VINCENTINI, R.; ROCHA, C. S.; SIMÕES, A. C.; ULIAN, E. C.; MAURO, S. M. di; SILVA, A. M.; PEREIRA, C. A.; MENOSSI, M.; SOUZA, G. M. Transcription profiling of signal transduction-related genes in sugarcane tissues. **DNA Research**, Oxford, v. 12, n. 1, p. 27-38, 2005.

QUACKENBUSH, J.; CHO, J.; LEE, D.; LIANG, F.; HOLT, I.; KARAMYCHEVA, S.; PARVIZI, B.; PERTEA, G.; SULTANA, R.; WHITE, J. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. **Nucleic Acids Research**, Oxford, v. 29, n. 1, p. 159-164, 2001.

RAIKHEL, N. Nuclear targeting in plants. **Plant Physiology**, Rockville, v. 100, p. 1627-1632, Dec. 1992.

SOUZA, G. M.; SIMÕES, A. C. Q.; OLIVEIRA, K. C.; GARAY, H. M.; FIORINI, L. C.; GOMES, F. S.; NISHIYAMA-JUNIOR, M. Y.; SILVA, A. M. The sugarcane signal transduction (Sucast) catalogue: prospecting signal transduction in sugarcane. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1/4, p. 25-34, 2001.

TANIMOTO, M.; TREMBLAY, R.; COLASANTI, J. Altered gravitropic response, amyloplast sedimentation and circumnutation in the *Arabidopsis shoot gravitropism 5* mutant are associated with reduced starch levels. **Plant Molecular Biology**, Dordrecht, v. 67, p. 57-69, Feb. 2008.

VETTORE, A. L.; SILVA, F. R. da; KEMPER, E. L.; ARRUDA, P. The libraries that made Sucest. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1/4, p. 1-7, 2001.

VETTORE, A. L.; SILVA, F. R.; KEMPER, E. L.; SOUZA, G. M.; SILVA, A. M.; FERRO, M. I. T.; HENRIQUE-SILVA, F.; GIGLIOTI, E. A.; LEMOS, M. V. F.; COUTINHO, L. L.; NOBREGA, M. P.; CARRER, H.; FRANÇA, S. C.; BACCI JUNIOR, M.; GOLDMAN, M. H. S.; GOMES, S. L.; NUNES, L. R.; CAMARGO, L. E. A.; SIQUEIRA, W. J.; SLUYS, M-A. V.; THIEMANN, O. H.; KURAMAE, E. E.; SANTELLI, R. V.; MARINO, C. L.; TARGON, M. L. P. N.; FERRO, J. A.; SILVEIRA, H. C. S.; MARINI, D. C.; LEMOS, E. G. M.; MONTEIRO-VITORELLO, C. B.; TAMBOR, J. H. M.; CARRARO, D. M.; ROBERTO, P. G.; MARTINS, V. G.; GOLDMAN, G. H.; OLIVEIRA, R. C.; TRUFFI, D.; COLOMBO, C. A.; ROSSI, M.; ARAUJO, P. G.; SCULACCIO, S. A.; ANGELLA, A.; LIMA, M. M. A.; ROSA JUNIOR, V. E.; SIVIERO, F.; COSCRATO, V. E.; MACHADO, M. A.; GRIVET, L.; MAURO, S. M. Z. D.; NOBREGA, F. G.; MENCK, C. F. M.; BRAGA, M. D. V.; TELLES, G. P.; CARA, F. A. A.; PEDROSA, G.; MEIDANIS, J.; ARRUDA, P. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, Cold Spring Harbor, v. 13, p. 2725-2735, Nov. 2003.

VINCENTZ, M.; CARA, F. A. A.; OKURA, V. K.; SILVA, F. R.; PEDROSA, G. L.; HEMERLY, A. S.; CAPELLA, A. N.; MARINS, M.; FERREIRA, P. C.; FRANÇA, S. C.; GRIVET, L.; VETTORE, A. L.; KEMPER, E. L.; BURNQUIST, W. L.; TARGON, M. L. P.; SIQUEIRA, W. J.; KURAMAE, E. E.; MARINO, C. L.; CAMARGO, L. E. A.; CARRER, H.; COUTINHO, L. L.; FURLAN, L. R.; LEMOS, M. V. F.; NUNES, L. R.; GOMES, S. L.; ZANTELLI, R. V.; GOLDMAN, M. H.; BACCI JUNIOR, M.; GIGLIOTI, E. A.; THIEMANN, O. H.; SILVA, F. H.; SLUYS, M-A. V.; NOBREGA, F. G.; ARRUDA, P.; MENCK, C. F. M. Evaluation of monocot and eudicot divergence using the sugarcane transcriptome. **Plant Physiology**, Rockville, v. 134, p. 951-959, Mar. 2004.

CHAPTER 3 – ISOLATION AND PARTIAL CHARACTERIZATION OF AN *ID1* HOMOLOGOUS GENE FROM SUGARCANE GENOMIC DNA

1 ABSTRACT

The flowering process in monocots follows most of the complex network characterized in eudicots. Nevertheless, some mechanisms of the regulation process that is explained by the well known flowering molecular model ABCE are divergent between the two classes. Monocots present paleas and lemmas instead of sepals and petals. It is suggested that specific genes in monocots might be involved in the flowering process control. *ID1* is a flowering time gene characterized in maize, expressed only in immature leaves, which indicates that this gene plays a role as a long-distance signal to flowering. The aim of the work was to isolate, clone, sequence and compare the tissue-specific expression pattern of *ScID1* comparing to maize *ID1*. The results of the amplification reaction with sugarcane genomic DNA allowed the isolation of part of the possible *ScID1* and expression analysis from cDNA demonstrated that the gene is expressed in immature leaves and the results from western blot analysis indicate that the ID1 protein is expressed in both immature and mature leaves of sugarcane. Further studies such as functional analyses need to be developed to confirm the involvement of this gene in the flowering process of sugarcane.

2 INTRODUCTION

Floral organs are generated by pluripotent cells derived from the shoot apical meristem (SAM). SAM is the target of floral inductive signals originated in vegetative tissues, mostly from the leaves. The floral transition is affected by both environmental and endogenous pathways, which activate the floral inductive signal, called florigen. So far, not much is known about this mobile signal.

The specification of the sepals in whorl 1 is conferred by class A homeotic genes *APETALA1* (*AP1)* and *APETALA2* (*AP2)*. The combination of the class A with the class B homeotic genes *PISTILLATA* (*PI)* and *APETALA3* (*AP3)* confer the petal identity in whorl 2. Class B combined with class C homeotic gene *AGAMOUS (AG)* specifies stamen identity in whorl 3. And class C in combination with class E homeotic gene *SEPALLATA (SEP)* confer carpel identity in whorl 4 (for review, see Krizek & Fletcher, 2005).

Monocot flowers present stamen and carpels, but differ from eudicot because they do not present the outer whorls (sepals and petals). Grass flowers have paleas (pa), lemmas (le) and lodicules (lo). Studies with maize and rice suggest that class B genes play the same role as in eudicot plants (Whipple et al., 2004). Loss of function of *AP3-like* gene in maize results in replacement of lodicules by paleas or lemmas, and replacement of stamens by carpels. *Arabidopsis thaliana* class B mutants present the same transformations, indicating that monocots paleas and lemmas are homologous to sepals, and lodicules to petals (for review, see Krizek & Fletcher, 2005).

Transition to flowering in monocots follow some pathways of the complex network of genes in dicots, but recent studies show that grasses have

developed unique flowering regulation mechanisms (Colasanti & Coneva, 2009).

The *ID1* gene determines the switch to flowering in maize, and this gene has been identified in related species such as sorghum and rice. Although sugarcane has been extensively studied (Guimarães et al., 1997; Figueiredo et al., 2001; Papini-Terzi, 2009; Casu et al., 2005; Calsa Júnior & Junqueira, 2007; Moore, 2005), no studies relating the flowering process to specific genes of this complex network has been reported. The flowering process in sugarcane is very important, as it is in any higher plants, but specifically in this crop, the transition to flowering is followed by the isoporization process. This process is defined as a dehydration of the stalk tissues, followed by a decrease of the cane mill, resulting in a higher fibre quantity (Caputo et al., 2007). The isoporization leads to the decrease in the yields of sugar in the stalks, due to the translocation of the carbon assimilates for seed production. Because of the importance of the flowering time gene *ID1* in maize, it is of main importance develop an initial characterization of this gene expression in sugarcane, since maize and sugarcane are closely related species. The aim of this work was to isolate, clone and verify the expression pattern of the *ID1* gene and ID1 protein in sugarcane, compared to maize *ID1*.

3 MATERIAL AND METHODS

3.1 Plant material

Experiments were carried out at the Colasanti Lab at the Department of Molecular and Cellular Biology of the University of Guelph, Canada. RB72 454 sugarcane variety donated by Professor Antônio Paulino da Costa Netto, collected at the Experimental field of the Universidade do Estado de Minas Gerais, at Passos, Brazil, was used. Plants were kept in a growth chamber at the Phytotron Facility of the University of Guelph under the following conditions: 80% of humidity, 1000 umol.m^2.s^{-1} and 28 ºC.

3.1 Extraction and purification of sugarcane genomic DNA

Leaves from sugarcane, variety RB72 454 were harvested and used for genomic DNA extraction. Tissues were frozen in liquid nitrogen and ground in microcentrifuge tubes. 750 µl extraction buffer and 50 µl 10 % SDS were added and mixed vigorously. Tubes, containing the extracts, were incubated at 65 °C for 10 minutes, centrifuged at maximum speed (14.000 xg) for 5 min. Supernatant was transferred to 1.5 mL microtubes that contained 250 µl of 5 M potassium acetate, mixed by inversion and incubated on ice for 20 min. After that, they were microcentrifuged once again at max speed for 5 minutes and supernatant transferred to a 1.5 mL tube containing 500 µl isopropanol, mixed and incubated at -20°C for 20 min. Then, these solutions were microfuged at max speed for 15 minutes, supernatant removed using aspirator and pellets washed in 300 µl of 70% ethanol. Supernatant was removed using aspirator and pellets dried and ressuspended in 50-100 µl water. After that, templates were

treated with RNAse I, *E. coli* (Fermentas Life Sciences) and left 60 minutes in incubator at 37º. Sugarcane genomic DNA was quantified using Nanodrop 8000 (Thermo Scientific).

3.2 Amplification of possible *ID1* sequences by PCR with degenerated primers

Six pairs of degenerated primers were designed based on maize *ID1* motif (Colasanti et al., 2006) and used to amplify possible *ID1* sequences in sugarcane genomic DNA.

Denaturating, annealing and extension temperatures and mastermix concentrations for PCR were optimized for each pair of the degenerated primers. PCR program was set up according to the melting temperature of each pair of primers. Programs were based on 40 cycles at 95 ºC for 30 seconds, annealing temperature in a gradient of 58 ºC to 60 ºC for 30 seconds and 72ºC for 1.5 minutes. The products were visualized in 1% electrophoresis gel, stained with ethidium bromide. Amplified region with c4scF/c4scR pair of primers was extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen) and quantified using Nanodrop 8000 (Thermo Scientific).

TABLE 2 Degenerated primers designed using the universal codes for specifying a wobble: R=A/G, Y=C/T, M=A/C, K=G/T, S=C/G, W=A/T, B=C/G/T, D=A/G/T, H=A/C/T, V=A/C/G, and N=A/C/G/T

Primer pairs	Sequence 5'-3'	Estimated amplicon*
c1scF	ATGATGATGGYTNWSNGAYYT	
c1scR	RAARTTRTGNSWCCANGTYTC	2940 bp
c2scF	ATGCARCARCARCCNWSNCC	
c2scR	CCANGTYTCRTTNACNGGNGG	294 bp
c3scF	WSNCCNTTYCTNCARCCNMG	
c3scR	YTGNAGYTCYTCNACYTGNC	211 bp
c4scF	GGNGGNGAYATGGCNWSNTA	
c4scR	YTGRTCNCKYTGRAANCCYTT	349 bp
c5scF	GAYTGGAARGCNCAYGTNAA	
c5scR	YTCYTGRTGNAGRAARAANAG	1799 bp
c6scF	AARAGRAARAGRWSNCARCCG	
c6scR	RAARTGYTTYTTDATNCCNGT	576 bp

* The amplicon size is estimated based on maize genomic gene, since the genome of sugarcane has not been sequenced yet. Sizes may vary due to presence of intronic regions.
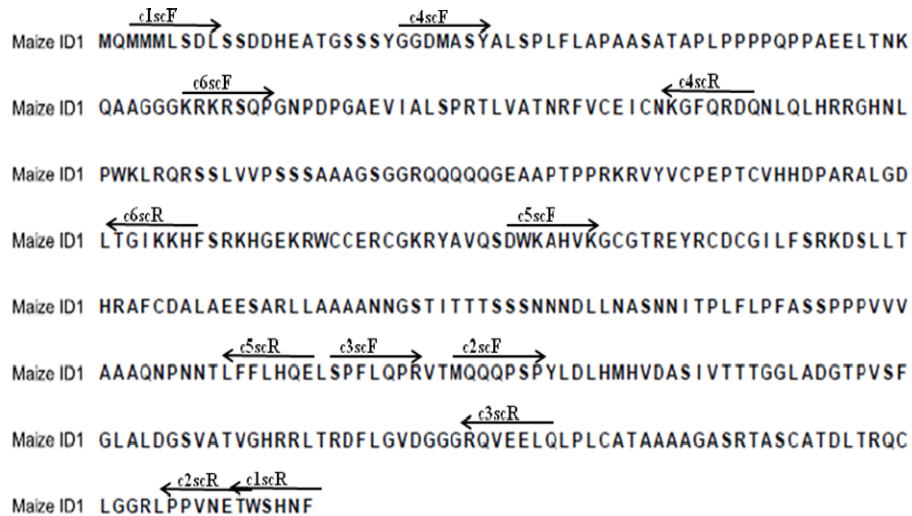
FIGURE 2 Translated sequence from maize *ID1* showing the regions from where the degenerated primers were designed.

3.3 Cloning and transformation of amplification products

Before the cloning, blunt-ends were added to the amplicons with T4 DNA polymerase (Fermentas), according to manufacturer's instructions. To 1uL of DNA was added 2uL of 10X of O buffer, dNTP mix and T4 DNA polymerase. The mastermix was mixed thoroughly, spun briefly and incubated for 5 minutes at room temperature. Reaction was stopped by heating at 75 ºC for 10 min.

Blunted-end DNA was ligated to the vector pCR-Blunt II-TOPO, using Zero Blunt Cloning kit (Invitrogen®) following manufacturer instructions.
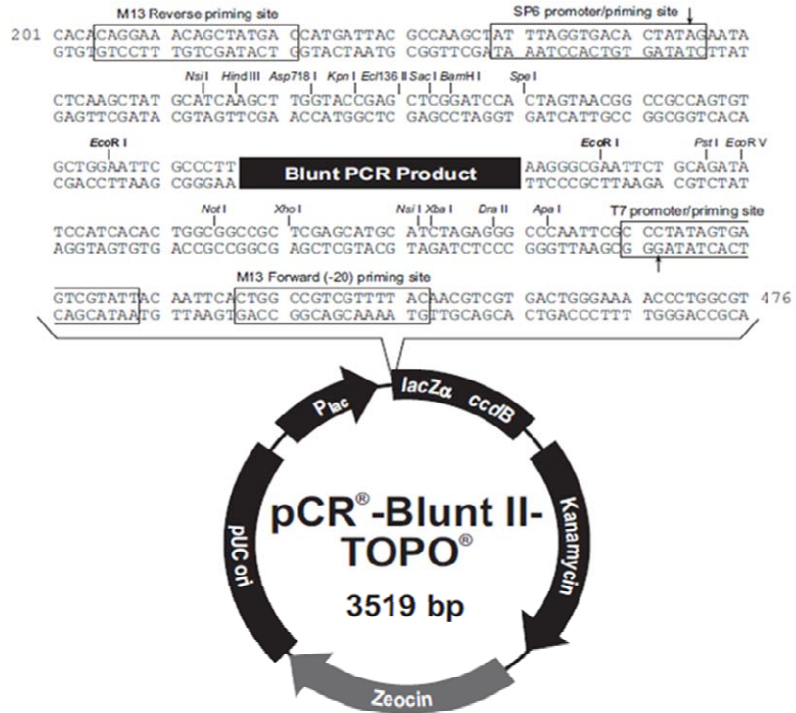
FIGURE 3 Schematic representation of pCR-Blunt II – TOPO vector construction indicating cloning site delimited by restriction sites (Invitrogen®).

3.4 Chemical transformation of *E. coli* competent cells

Recombinant vectors were inserted in *E. coli*, DH5α strain chemically competent (Invitrogen®) in Kan50 supplemented medium. 2 µL of ligation products were mixed to 40 µL *E. coli*, incubated on ice for 30 minutes, followed by heat shock of 42ºC for 45 seconds, and incubated again on ice for 2 minutes. 250 µL of SOC medium (Invitrogen®) were added to the recombinant cells,

incubated in a shaker (300 rpm) at 37 ºC for 45 minutes. After this, cells were plated on solid LB medium enriched with 50 µL/mL of kanamycine (Kan50), and left for 16 hours at 37 ºC.

3.5 PCR analysis of transformed colonies

Transformed colonies were submitted to a PCR, using an inoculum of each colony resistant to the selective medium added to a mastermix containing primers c4scF/c4scR and the GoTaq (Promega®). Amplification was performed submitting the cells to 42 cycles in a denaturing period of 30 seconds at 95 ºC, annealing for 30 seconds at 58 ºC and extension at 72 ºC for 1.5 minutes. PCR products were submitted to an electrophoresis gel in 1% agarose and stained with ethidium bromide.

3.6 Extraction of plasmid DNA with possible *ID1* sequences

Plasmid DNA was extracted from transformed colonies using Plasmid Miniprep Kit (Qiagen). One colony containing the gene was added to 10 mL of LB medium supplemented with Kan50 and kept in orbital shaker at 200 rpm, 37 ºC for 16 hours. After that, cultures were centrifuged to pellet the cells, and used for plasmid DNA extraction.

Miniprep protocols are based on alkaline lysis by the addition of buffers provided by the manufacturer, following the plasmid DNA ligation to the anionic column. DNA was diluted in DEPC water and stored at -80º.

3.7 Digestion with restriction enzymes

Plasmid DNA (at a concentration of 417.6 ng/µL, and A260/280 of 1.88) was digested with *Eco*RI based on pCR® vector construction that includes

the inserted site. 2 uL of 10X O buffer and 1uL of *Eco*RI was added to 1uL of plasmid DNA, mixed thoroughly and left at 37 ºC for 1 hour. Digestion was analysed by gel electrophoresis in agarose 1%, stained with ethidium bromide.

3.8 Sequencing of circular DNA

Clones selected were sequenced using ABI PRISM 3100 genetic analyser (Applied Biosystems) at the Genomic Facility of the Science complex, University of Guelph. M13 reverse and M13 forward primers were used to generate sequences.

3.9 Alignment and phylogenetic analysis of sugarcane ID1 sequences with orthologs of related species.

Sequences from sugarcane were aligned with *ID1* sequence from sorghum, maize and rice using *ClustalW* 1.8 software, according to default parameters. Phylogenetic trees were constructed using *CLC Main Workbench 5.1* software using *Neighbour-Joining* comparison model (Saitou & Nei, 1987), *p* distance method and *pairwise* supression. Phylogenetic distance of clusters was measured by *bootstraps*.

3.10 Specific primer design based on sugarcane *ID1* sequence

After the sequencing reaction of c4scF/c4scR amplified region, a c4F specific primer was designed based on part of sugarcane *ID1* sequence. And, a specific reverse primer, c4R, was designed based on the conserved domain of maize *ID1* gene.

TABLE 2 Specific primers designed after sequencing from part of the putative
sugarcane *ID1*.

Primer pairs	Sequence 5'-3'	Amplified fragment
c4F	CCTCTCTTCCTCGCACCGGC	~ 700 bp
c4R	AAGCCTCGCGCTCTCCTCTG	


```
                        c4F
Maize ID1  MQMMMLSDLSSDDHEATGSSSYGGDMASYALSPLFLAPAASATAPLPPPPQPPAEELTNK

Maize ID1  QAAGGGKRKRSQPGNPDPGAEVIALSPRTLVATNRFVCEICNKGFQRDQNLQLHRRGHNL

Maize ID1  PWKLRQRSSLVVPSSSAAAGSGGRQQQQQGEAAPTPPRKRVYVCPEPTCVHHDPARALGD

Maize ID1  LTGIKKHFSRKHGEKRWCCERCGKRYAVQSDWKAHVKGCGTREYRCDCGILFSRKDSLLT
                c4R
Maize ID1  HRAFCDALAEESARLLAAAANNGSTITTTSSSNNNDLLNASNNITPLFLPFASSPPPVVV

Maize ID1  AAAQNPNNTLFFLHQELSPFLQPRVTMQQQPSPYLDLHMHVDASIVTTTGGLADGTPVSF

Maize ID1  GLALDGSVATVGHRRLTRDFLGVDGGGRQVEELQLPLCATAAAAGASRTASCATDLTRQC

Maize ID1  LGGRLPPVNETWSHNF
```

FIGURE 4 Translated sequence from maize *ID1* showing the regions from
where the c4F/c4R primers were designed.

3.11 RNA extraction of immature sugarcane leaves

Total RNA of immature and mature sugarcane leaves, and shoot apical
meristem were extracted using TRIZOL Reagent (Invitrogen[®]) as described by

Colasanti et al. (1998). Tissue were collected, weighed and frozen in liquid nitrogen. Subsequently, it was ground with mortar and pestle, an amount of 1mL of TRIZOL was added to each 100mg of tissue (1mL/100mg) and mixed vigorously in a vortex. Extracts were incubated for 5 minutes at room temperature and centrifuged at 4 ºC for 10 minutes at 12.000 xg. The supernatant was transferred to a new tube containing chloroform (0.2 mL/mL of Trizol) and mixed for 15 seconds. The solution was centrifuged for 20 minutes at 4 ºC at 12.000 xg, and the aqueous layer was transferred to a new tube. Isopropanol (0.5 mL/mL of Trizol) was added to the solution, mixed by inversion and incubated for 60 minutes at -20ºC. The samples were centrifuged for 10 minutes at 4ºC (12.000 xg) and the supernatant discarded. The pellets was washed with 1 mL of cold ethanol 75% and the tubes were submitted to a new centrifugation for 5 minutes at 4ºC (12.000 g). The supernatant was removed and the RNA ressuspended in 30 µL of DEPC water. The sample concentration was quantified using NanoDrop 8000 (Thermo Scientific), and then stocked at -80ºC.

3.12 cDNA synthesis and RT-PCR amplification

Sugarcane cDNA was synthesized using qScript cDNA Supermix kit (Quanta Biosciences) according to manufacture instructions. 2 µL of qScript cDNA Supermix was added to 1 µL of RNA template and DEPC water to a total volume of 10 µL. Samples were mixed and centrifuged briefly. The mixture was incubated for 5 minutes at 25ºC, 30 minutes at 42ºC, 5 minutes at 85ºC and hold at 4ºC. After incubation, cDNA was diluted 5x to perform subsequent amplification reactions.

3.13 *ID1* gene expression profile analysis by semi-quantitative RT-PCR

Immature leaves and SAM cDNAs were used to verify the expression pattern of sugarcane putative *ID1* in different tissues. c4F and c4scR primers were used to amplify *ID1* gene. Semi quantitative RT/PCR program was set up to 37 cycles at 95ºC for 30 seconds, 58ºC for 30 seconds and 72ºC for 1.5 minutes. Expression was analysed by staining with ethidium bromide eluted in 1% agarose gel electrophoresis.

3.14 Western blot of ScID1 protein using maize ID1 protein antibody

Nuclei isolation to protein extraction was performed using an adapted protocol of Jay Hollick (Colasanti et al., 2006). Western blots were performed according to the Steinmuller & Apel (1986) method with modifications. Plant tissues (1.0 to 5.0 grams), were collected in liquid nitrogen, ground to a fine powder in liquid nitrogen and 5 mL of cold isolation buffer (IB) was added: (IB: 20 mM Tris-HCl, pH 7.8, 250 mM of sucrose, 5 mM $MgCl_2$, 5 mM KCl, 40% (v/v) glycerol, 0.25% (v/v) TritonX-100, 0.1% (v/v) β-mercaptoethanol). This suspension was filtered in a cheesecloth and then, filtered again in a nylon membrane, centrifuged at 6.000 xg for 10 minutes, the pellet was washed twice with IB and ressuspended in 50 μL of IB. 50 μL of Laemmli buffer (62.5 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 25% (v/v) glicerol, 0.01% (w/v) of bromophenol blue) was added to the suspension. Duplicated samples of proteins (50 to 100 μg) were submitted to polyacrilamide gels 10% (w/v) and transferred to nitrocellulose membranes by iBlot® Dry Blotting System (Invitrogen®), or stained with Comassie blue to confirm protein concentration and evaluate protein integrity. For blotting, nitrocellulose membranes were blocked in 5% (w/v) milk in PBS (20 mM Tris-buffered, 137 mM NaCl, pH 7.7) with 0.2% (v/v) of Tween-20. Purified antibody, primary anti-ID1 was used in a dilution of

1:5000 in milk solution of PBS-Tween/5% and incubated for 1 hour in room temperature. Blots were washed with PBS-Tween 0.2%, 6 times for 10 min each and then incubated with secondary antibody conjugated with peroxidase (Amersham) in a dilution of 1:10.0000, and then washed as described above. Ligated antibodies were detected with a Super Signal West Pico Chemiluminescent substrate (Pierce, Woburn, MA) and exposed to XAR X-ray film (Kodak).

4 RESULTS AND DISCUSSION

The sugarcane genome is still to be sequenced, and, because of this, degenerated primers were designed based on maize *ID1* gene. From primer pairs c4scF/c4scR it was possible to isolate a sequence of 400 bp by gel extraction (concentration of 25 ng/μL), subsequently the template was cloned and chemically transformed in *E. coli* competent cells. Plasmid DNA was extracted, and to confirm the ligation into the vector, a digestion was performed with *Eco*RI. From the digestion it was possible to verify a band in the expected size of the insert (400 bp) and another one in the size of the vector (Figure 5).
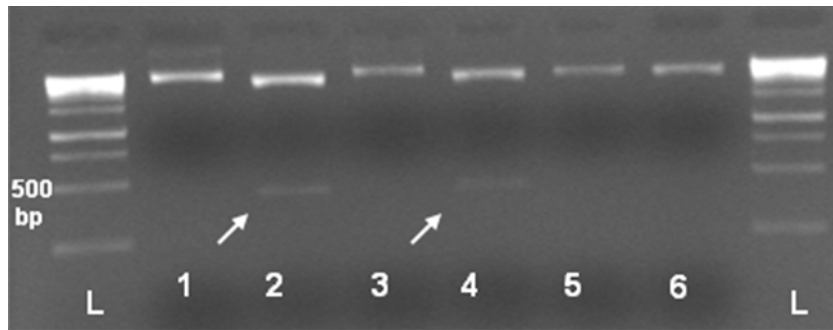

FIGURE 5 Digestion of plasmid DNA with *Eco*RI. 1 – Clone 1A uncut, 2 – Digested Clone 1A, 3 – Clone 1B uncut, 4 – Digested Clone 1B, 5 – Clone 1E uncut, 6 – Digested Clone 1E. Digested clones 1A and 1B presented the expected fragment of 400 bp. L: Ladder of 1 Kb.

After that, the fragment was sequenced and the translated sequence was analysed to verify if it contained the conserved domain. It seems that the genomic fragment presents a small intron surrounding the zinc finger 1, if compared to the maize *ID1* translated sequence, which presents an intronic region of 89 bp in the same position (Figure 6). Although supported data suggests that this region comprises an intron like *ID1*, sugarcane cDNAs need to be sequenced to confirm the right position and characterize the sugarcane intronic region.

FIGURE 6 Aligment of translated sequence of sugarcane c4scF/c4scR amplicon to the respective region of maize *ID1*. In the black box is highlighted the possible intronic region of *ScID1* compared to the maize *ID1* intronic sequence.

The translated sequence was aligned to the translated sequence of *ID1* and *SbID1* gene (Figure 7). The sequenced region includes the zinc finger 1 (z1) involving the *Nuclear Localization Sequence* (NLS) KRKR, which marks the beginning of the ID domain (Colasanti et al., 2006).

These findings suggest that the sequence found may be an orthologous gene of the *ID1* in sugarcane because it shows one of the *ID1* conserved

57

domains and presents a possible intronic region similarly to the *ID1* gene. Although functional analyses were not performed to verify whether or not *ScID1* plays a role as a flowering time gene, an ortholog gene can be identified by phylogenetic reconstructions based on conserved protein domains (Nardmann & Werr, 2007). Based on this, the possible conserved function needs to be confirmed by gene expression patterns or mutant analyses. Consequently, it will be possible to confirm whether the putative *ScID1* act as a flowering time gene in sugarcane.

An alignment was performed between the coding region of sequenced c4scF/c4scR *ScID1* with the respective regions of *ID1* and *SbID1* (Figure 7).
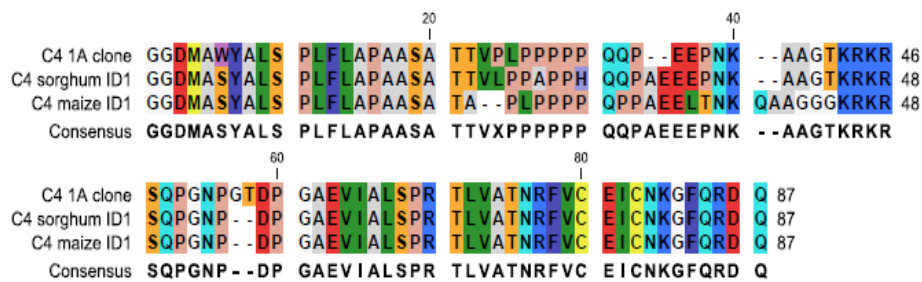

FIGURE 7 Alignment of the sugarcane c4scF/c4scR amplicon (C4 1A clone) to the respective coding regions of maize *ID1* and *SbID1*.

It can be noticed that there is a higher similarity between *SbID1* than putative *ScID1* and *ID1*. These findings are supported by the fact that sorghum and sugarcane present a closer phylogenetic relationship than sugarcane and maize.

There is a connection between the transition to flowering and the assimilate partitioning in several species, following an increase of soluble carbohydrates in the apical bud (Corbesier et al., 1998; Bernier, 1993). So, the increased diversion of assimilates to the SAM might be related to the floral induction, and therefore to the floral stimulus, as well (Sachs & Hackett, 1983).

58

An increase of the sucrose content in the shoot apex of *Sinapis alba* (a long-day plant) is correlated to the stimulation of flowering when exposed to a long-day (Bodson & Outlaw, 1985). Wong & Colasanti (2007) found that *ID1* expression is not related to the sink-source status and, since it is unperturbed by external stimuli, it is developmentally regulated. All these facts support the hypothesis that, in maize, the *ID1* gene acts through a novel autonomous floral induction pathway (Wong & Colasanti, 2007; Coneva et al., 2007).

A few downstream genes are differentially expressed in normal and *id1* mutant plants, and a significant part of them is related to photosynthesis and carbohydrate metabolism. These findings substantiate the hypothesis of a possible connection between floral inductions and assimilate partitioning (Coneva et al., 2007).

Similarly, in sugarcane there is a process called isoporization, in which the major part of the assimilates stocked in the stalks is translocated to the reproductive organs (Salata et al., 1982). This process is responsible for the reduction of the production of sugar and ethanol.

Still, functional analysis needs to be performed with downstream targets of *ID1* to indicate whether these genes play a role in the flowering process (Coneva et al., 2007). Additionally, functional characterization of the *ScID1* will allow the identification of downstream regutated genes and a better understanding of the relation assimilates partitioning/flowering induction.

For the preliminary analysis of semi-quantitative RT-PCR, RNA obtained from different tissues of sugarcane was used to synthesize cDNAs. Due to technical problems of extracting high quantity of RNA from mature sugarcane leaves, only SAM and immature leaves from sugarcane were used to synthesize cDNA (Table 3 and Figure 8).

TABLE 3 RNA concentration of sugarcane immature and mature leaves and maize immature leaves

RNA samples	ng/µL	A260/280
Sugarcane immature leaves	596.6	1.76
Sugarcane SAM	475.9	1.87
Sugarcane mature leaves	47.45	1.73
Maize immature leaves	285.6	1.8

cDNAs were submitted to RT-PCR to determine if the sugarcane *ID1* expression profile follows the same pattern as maize. The expected size of the PCR products, based on maize *ID1* sequence, was around 700 bp. It has been suggested that *ID1* gene function is related to a long-distance response, because, although its function involves the determinacy of the transition to reproductive stage in maize, it is expressed only in immature leaves (Colasanti et al., 1998).
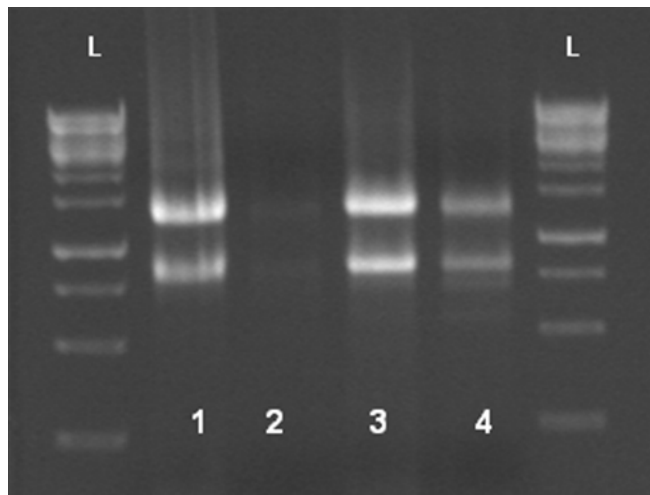
FIGURE 8 RNA profile from sugarcane and maize tissues. 1 –sugarcane immature leaves, 2 –sugarcane mature leaves, 3 – sugarcane SAM, 4 –maize immature leaves. L:Ladder of 1Kb.

C4 primers (forward and reverse) were used to detect expression levels in immature sugarcane leaves and shoot apical meristem. Similarly to maize, the results for sugarcane showed *ID1* expression in immature leaves (Figure 9), suggesting that the pattern might be conserved between the species. The expression profile shows that immature sugarcane leaves present higher expression compared to immature maize leaves, used as a positive control. No expression was observed in the SAM or in the negative control (H_2O).
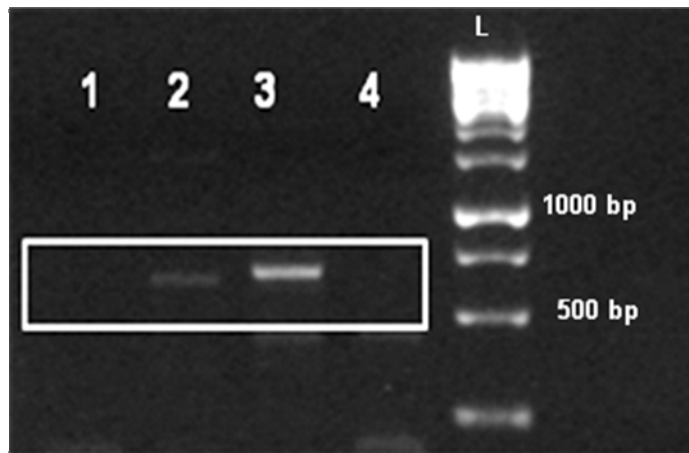

FIGURE 9 Expression profile of *ID1* gene between SAM and immature leaves of sugarcane compared with maize immature leaves. 1- H_2O (negative control), 2 –maize immature leaves (positive control), 3 – sugarcane immature leaves, 4 – sugarcane SAM. L: Ladder of 1 Kb.

These results are still preliminary, the reaction parameters need to be optimized and, additionally, quantitative PCR need to be performed to confirm the expression profile.

To investigate whether the ScID1 protein presents similar expression patterns of ID1, an anti-ID1 antibody specific for the C-terminal 20 amino acids of ID1 was used. Related grass species, sorghum and rice, present deduced C-terminal peptide sequences of SbID1 and OsID1 proteins nearly identical to ID1 and are expressed exclusively in immature leaves (Colasanti et al., 2006).

First of all, a staining with Comassie Blue was developed to verify the protein quantity and quality (Figure 10). After that, a western blotting was performed to verify if the anti-ID1 antibody cross-reacted with putative ScID1.
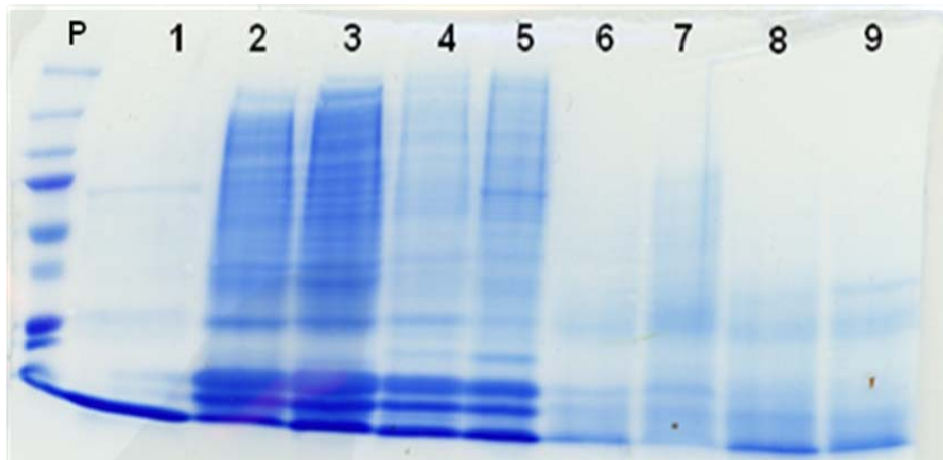

FIGURE 10 Protein profile developed by comassie staining. 1 – in vitro *ID1* protein, 2 – sugarcane immature leaves, 3 – sugarcane SAM, 4 – sugarcane mature leaves, 5 – maize mature leaves, 6 – supernatant from sugarcane immature leaves, 7 – supernatant from sugarcane SAM, 8 – supernatant from sugarcane mature leaves, 9 – supernatant from maize mature leaves. P: Protein standard of 250 kD (BIO-RAD).

Interestingly, sugarcane ID1 proteins hybridized with anti-ID1 antibody, suggesting that the protein shares most of the amino acids localized in the C-terminal region.

Differently from ID1, SbID1 and OsID1, ScID1 protein was detected in both immature and mature leaves, and is absent in the region of the shoot apex (Figure 11).

FIGURE 11 Western blotting showing the hybridization of ID1- antibody with immature and mature leaves of sugarcane. Film was exposed for 5 minutes with primary antibody at 1:5000 and secondary antibody at 1:10000. 1 – *in vitro* ID1 (positive control), 2 – sugarcane immature leaves, 3 – sugarcane SAM, 4 – sugarcane mature leaves, 5 – maize mature leaves (negative control). Blue lane: molecular weight standard of 50 kDa.

Nevertheless, observing the blotting it can be noticed that the levels of ScID1 protein in immature leaves is higher than sugarcane mature leaves, supporting the results found with the expression profile of the RT-PCR, where no expression was verified in the sugarcane SAM and lower expression was detected in maize immature leaves when compared to sugarcane immature leaves. This result suggests that, although the protein can be found in mature leaves, in immature leaves its function might be more intense. The absence of the protein in SAM suggests that the ScID1 may have a function similar to maize ID1. Furthermore, the restriction of ID1 expression to developing leaves suggests that this putative regulatory protein controls a leaf derived floral inductive signals (Colasanti et al., 2006).

A sequencing reaction of the ScID1 peptide sequence needs to be done to confirm the identity of the protein and to analyse the percentage of identity of the amino acids.

5 CONCLUSIONS

Isolation of partial sequence of *ScID1* and high homology with *ID1* translated sequence suggests that sugarcane may have an *ID1* ortholog. Functional analysis is needed to prove its function in flowering.

6 REFERENCES

BERNIER, G.; HAVELANGE, A.; HOUSSA, C.; PETITJEAN, A.; LEJEUNE, P. Physiological signals that induce flowering. **Plant Cell**, Rockville, v. 5, p. 1147-1155, Oct. 1993.

BODSON, M.; OUTLAW, W. H. Elevation in the sucrose content of the shoot apical meristem of *Sinapis alba* at floral evocation. **Plant Physiology**, Rockville, v. 79, n. 2, p. 420-424, Oct. 1985.

CALSA JÚNIOR, T.; JUNQUEIRA, A. Serial analysis of gene expression in sugarcane (*Saccharum ssp.)* leaves revealed alternative C4 metabolism and putative antisense transcripts. **Plant Molecular Biology**, Netherlands, v. 63, n. 10, p. 745-762, Apr. 2007.

CAPUTO, M. M.; SILVA, M. A.; BEAUCLAIR, E. G. F.; GAVA, G. J. C. Acúmulo de sacarose, produtividade e florescimento de cana de açúcar sob reguladores vegetais. **Interciencia**, Venezuela, v. 32, p. 834-839, dic. 2007.

CASU, R. E.; MANNERS, J. M.; BONNETT, G. D.; JACKSON, P. A.; MCINTYRE, C. L.; DUNNE, R.; CHAPMAN, S. C.; RAE, A. L.; GROF, C. P. L. Genomics approaches for the identification of genes determining important traits in sugarcane. **Field Crops Research**, Netherlands, v. 92, p. 137-147, June 2005.

COLASANTI, J.; CONEVA, V. Mechanisms of floral induction in grasses: something borrowed something new. **Plant Physiology**, Rockville, v. 149, p. 56-62, Jan. 2009.

COLASANTI, J.; TREMBLAY, R.; WONG, A. Y.; CONEVA, V.; KOZAKI, A.; MABLE, B. K. The maize *Indeterminate1* flowering time regulator defines a highly conserved zinc finger protein family in higher plants. **BMC Genomics**, London, v. 7, p. 158-175, June 2006.

COLASANTI, J.; YUAN, Z.; SUNDARESAN, V. The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. **Cell**, Cambridge, v. 93, p. 593-603, May 1998.

CONEVA, V.; ZHU, T.; COLASANTI, J. Expression differences between normal and *indeterminate1* maize suggest downstream targets of *ID1*, a floral transition regulator in maize. **Journal of Experimental Botany**, Oxford, v. 58, p. 3679-3693, Oct. 2007.

CORBESIER, L.; LEJEUNE, P.; BERNIER, G. The role of carbohydrates in the induction of flowering in *Arabidopsis* thaliana: comparison between the wild type and a starchless mutant. **Planta**, Germany, v. 206, p. 131-137, July 1998.

FIGUEIREDO, R. C.; BRITO, M. S.; FIGUEIREDO, L. H. M.; QUIAPIN, A. C.; VITORELLI, P. M.; SILVA, L. R.; SANTOS, R. V.; MOLFETTA, J. B.; GOLDMAN, G. H.; GOLDMAN, M. H. S. Dissecting the sugarcane expressed sequence tag (SUCEST) database: unravelling flower-specific genes. **Genetics and Molecular Biology**, Ribeirão Preto, v. 24, n. 1/4, p. 77-84, 2001.

GUIMARÃES, C. T.; SILLS, G. R.; SOBRAL, B. W. S. Comparative mapping of Andropogenae: Saccharum L. (sugarcane) and its relation to sorghum and maize. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 94, p. 14261-14266, Dec. 1997.

KRIZEK, B. A.; FLETCHER, J. C. Molecular mechanisms of flower development: an armchair guide. **Nature**, England, v. 6, p. 688-698, Sept. 2005.

MOORE, P. H. Integration of sucrose accumulation processes across hierarchical scales: towards developing an understanding of the gene-to-crop continuum. **Field Crops Research**, Amsterdam, v. 92, p. 119-135, June 2005.

NARDMANN, J.; WERR, W. The evolution of plant regulatory networks: what *Arabidopsis* cannot say for itself. **Current Opinion in Plant Biology**, London, v. 10, p. 653-659, Aug. 2007.

PAPINI-TERZI, F. S. Sugarcane genes associated with sucrose content. **BMC Genomics**, London, v. 10, p. 120-141, Mar. 2009.

SACHS, R. M.; HACKETT, W. P. Source-sink relationships and flowering. In: MEUDT, W. J. (Ed.). **Strategies of plant reproduction, Beltsville Symposium on Agricultural Research**. Totowa, New Jersey: Allenheld-Osmun, 1983. p. 263-272.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution,** Lawrence, v. 4, n. 4, p. 406-425, July 1987.

SALATA, J. C.; FERREIRA, L. J.; CASAGRANDE, A. A. Interferência de florescimento nas qualidades agroindustriais de algumas variedades comerciais de cana de açúcar (*Saccharum ssp.*). **Brasil Açucareiro**, São Paulo, v. 93, n. 1, p. 45-55, 1982.

STEINMULLER, K.; APEL, K. A Simple and efficient procedure for isolating plant chromatin which is suitable for studies of DNAse i-sensitive domains and hypersensitive sites. **Plant Molecular Biology**, Dordrecht, v. 7, p. 87-94, May 1986.

WHIPPLE, C. J.; CICERI, P.; PADILHA, C. M.; AMBROSE, B. A.; BANDONG, S. L.; SCHMIDT, R. J. Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. **Development**, Cambridge, v. 131, p. 6083-6091, Oct. 2004.

WONG, A. Y. M.; COLASANTI, J. Maize floral regulator protein *Indeterminate1* is localized to developing leaves and is not altered by light or the sink/source transition. **Journal of Experimental Botany**, Oxford, v. 58, p. 403-414, Sept. 2007.