Genetics and Molecular Biology, 35, 3, 632-639 (2012)

www.sbg.org.br

Research Article

Functional characterization of sugarcane *mustang* domesticated transposases and comparative diversity in sugarcane, rice, maize and sorghum

Daniela Kajihara, Fabiana de Godoy, Thais Alves Hamaji, Silvia Regina Blanco, Marie-Anne Van Sluys and Magdalena Rossi

Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil.

Abstract

Transposable elements (TEs) account for a large portion of plant genomes, particularly in grasses, in which they correspond to 50%-80% of the genomic content. TEs have recently been shown to be a source of new genes and new regulatory networks. The most striking contribution of TEs is referred as "molecular domestication", by which the element coding sequence loses its movement capacity and acquires cellular function. Recently, domesticated transposases known as *mustang* and derived from the *Mutator* element have been described in sugarcane. In order to improve our understanding of the function of these proteins, we identified *mustang* genes from *Sorghum bicolor* and *Zea mays* and performed a phenetic analysis to assess the diversity and evolutionary history of this gene family. This analysis identified orthologous groups and showed that *mustang* genes are highly conserved in grass genomes. We also explored the transcriptional activity of sugarcane *mustang* genes in heterologous and homologous systems. These genes were found to be ubiquitously transcribed, with shoot apical meristem having the highest expression levels, and were downregulated by phytohormones. Together, these findings suggest the possible involvement of *mustang* proteins in the maintenance of hormonal homeostasis.

Key words: domesticated tranposases, *Saccharum* sp., *Sorghum bicolor,* transposable elements, *Zea mays*.

Received: December 2, 2011; Accepted: May 16, 2012.

Introduction

Transposable elements (TEs) have an important role in genome evolution because they are substrate for illegitimate recombination, generating chromosomal translocations, inversions and deletions (Sinzelle *et al.*, 2009). Several reports have shown that TEs can modify gene expression patterns by creating promoter and *cis* regulatory sequences, as well as alternative splicing and polyadenylation sites (Bonnivard and Higuet, 2008; Feschotte, 2008). The most direct impact that these mobile elements have on genomes is known as "molecular domestication" which occurs when a transposon loses its transpositional activity and becomes a host gene with an established function, thereby giving rise to a new gene (Sorek, 2007). This process has been described for several transposases in a wide variety of eukaryotic genomes from yeasts to humans (Feschotte, 2008).

Transposases, the most abundant proteins in nature (Aziz *et al.*, 2010), are responsible for the "cut and paste" mechanism of transposon mobility. During transposition, transposases recognize the flanking terminal inverted repeats (TIRs) of the element, excise the DNA fragment and insert it into a new target site. Transposases contain inherent DNA-binding activity. Most of the domesticated transposases reported so far have been characterized only at the sequence and structural levels and the few whose functional role has been investigated have been shown to be transcription factors. The first example reported for plants was the *Daysleeper* gene from *Arabidopsis thaliana* (Bundock and Hooykaas, 2005). This gene is derived from an *hAT* superfamily transposase and the encoded protein is a master transcription factor involved in the control of morphogenetic development. Other interesting examples include the *fhy3* and *far1* genes related to the *Mutator* superfamily transposase. FHY3 and FAR1 proteins are also transcription factors and have been implicated in maintaining homeostasis in the light response (Hudson *et al.*, 2003; Lin *et al.*, 2007).

Sugarcane is an economically important crop and, together with maize, rice and wheat, is one of the major agricultural commodities in terms of productivity (Devos, 2010). Sugarcane is commonly cultivated in tropical or subtropical regions and used mainly for sugar and biofuel production. The modern cultivars are inter-specific hybrids

Send correspondence to Magdalena Rossi. Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo (USP), Rua do Matão 277, 05508-090 São Paulo, SP, Brazil. E-mail: mmrossi@usp.br.

between two polyploid species, *Saccharum officinarum* (2n = 80) and *Saccharum spontaneum* (2n = 40-128); these hybrids have a complex genome with highly polyploid and aneuploidy (D'Hont and Glaszman, 2001). Allopolyploidization is one of the major factors in the evolution of the grass genome, leading to an increase in gene number, the activation of TEs and alterations in the epigenetic landscape, all of which can lead to new patterns of gene expression, creation and loss (Parisod *et al.*, 2010; Yaakov and Kashkush, 2011). Sugarcane therefore provides a useful example for studying the behavior of TEs in relation to the "genomic stress" caused by allopolyploidy.

The *Mutator* system is one of the most expressed transposons in sugarcane (Rossi *et al.*, 2001; Araújo *et al.*, 2005). An investigation of the diversity of *Mutator* transposase-related sequences in plants revealed the existence of four classes that emerged before the monocot-eudicot divergence. Further structural characterization and analysis of the genomic distribution revealed that Classes I and II corresponded to *bonafide* transposons while Classes III and IV were *mustang* domesticated transposases. A class-specific amplification in grasses with a burst of Class II elements has been identified based on an assessment of copy numbers in rice and sugarcane (Saccaro-Jr *et al.*, 2007).

The *mustang* domesticated transposases were first described in rice and *Arabidopsis* by Cowan *et al.* (2005) but no further functional characterization has been reported for this gene family. In the sugarcane polyploid genome, 6 and 26 copies of *mustang* genes were identified for Classes III and IV, respectively (Saccaro-Jr *et al.*, 2007). Recently, we sequenced and characterized three sugarcane *mustang* genes. For Class III, two haplotypes (one from each parental species) were selected, namely, SCMUG266BAC095 (*S. officinarum*) and SCMUG266BAC148 (*S. spontaneum*) while from Class IV the SCMUG148BAC249 gene was chosen. A comparative analysis of the genomic regions containing these *mustang* genes in rice, sorghum and sugarcane revealed perfect colinearity that supported their orthology and eliminated the possibility of horizontal transfer events (Marie-Anne Van Sluys, unpublished observation). To gain further insights about the function of this widely distributed and diverse transposon-derived gene family, in this work we screened for these genes in recently sequenced grass genomes, established their orthologous relationship, assessed their transcriptional activity in a heterologous system and produced an extended tissue and temporal transcription profile in sugarcane.

Materials and Methods

Phenogram and comparative sequence analysis

The *mustang* gene phenogram was constructed using sequences of Class III and Class IV *Mutator*-like transposases from sugarcane (three genomic haplotypes recently obtained by our group), *Sorghum bicolor* (Phytozome database), *Zea mays* (MaizeGDB, database), *Oryza sativa* (TIGR database) and *A. thaliana* (TAIR database). Sequence *loci* and accession numbers are detailed in Table S1 (Supplementary Material). Sugarcane cDNAs belonging to Class I (TE165) and Class II (TE109), as well as the *Arabidopsis mudrA*-like locus At2g07100, were used as external groups. Full-length nucleotide coding regions were aligned with ClustalW (Higgins *et al.*, 1994). A distance matrix based on the neighbor-joining algorithm (Saitou and Nei, 1987) was generated with MEGA 3.1 (Kumar *et al.*, 2004). A regular bootstrap test was applied with 1,000 repeats. Gene identity analyses were done by global alignment using the EMBOSS Pairwise Alignment Algorithm.

Promoter-reporter gene fusions

The putative promoter regions of sugarcane *mustang* genes were amplified with two pairs of primers (Table S2. Supplementary Material). ClassIII-promF and ClassIII-promR amplified fragments of 1,921 bp and 1,685 bp from SCMUG266BAC148 and SCMUG266BAC095 loci, respectively. ClassIV-promF and ClassIV-promR amplified a 2,000 bp fragment from the SCMUG148BAC249 gene. PCR reactions were done using BAC DNA from the corresponding genome (BAC095-F04, BAC249-C12 and BAC148-J07), according to Almeida *et al.* (2011). The amplicons were cloned into the entry vector pDONR221 (Invitrogen) after which the putative promoter regions were recombined into the PHGWFS7 binary vector (Karimi *et al.*, 2002) using GATEWAY technology (Invitrogen). This vector contains two reporter genes, *gfp* and *gus,* downstream from the putative promoter. The constructs were tested in transient expression assays involving particle bombardment of *Allium cepa* epidermis, according to Rech and Aragão (1998), using a helium version of the PDS-1000 device (Sanford *et al.*, 1991) developed by EMBRAPA (Brasília, Brazil). Three independent experiments were done in triplicate. The pCambia 1391Z.35S vector, which harbors the *gus* gene under control of the 35S promoter of cauliflower mosaic virus, and the PHGWFS7 empty vector were used as positive and negative controls, respectively. After 24 h, GUS activity was detected by incubation overnight in buffer containing X-gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid) at 37 °C (McCabe *et al.*, 1988).

Promoter analysis in a heterologous system

Promoter functionality was assayed in a stable transgenic system of tobacco BY2 cell cultures (Nagata *et al.*, 1992). Cell suspensions were grown in modified Murashige-Skoog (MS) medium (Murashige and Skoog, 1962) containing 3% (w/v) sucrose and 0.2 mg of 2,4-dichlorophenyoxyacetic acid (2,4-D)/mL, pH 5.7 and subcultured into fresh media every 6-7 days. The *Agrobacterium tumefaciens*-mediated transformation was done according to An (1985). After co-culture, BY2 cell suspensions were

transferred to fresh selective medium (40 μg of hygromycin/mL). Isolated calli were maintained on agar medium or cultured in liquid medium to obtain transgenic cell suspensions. The presence of the transgene fusion was confirmed by PCR using promoter-specific primers (Almeida *et al.*, 2011). GUS activity was detected as described above. GFP fluorescence was detected using a Zeiss LSM 410 confocal microscope equipped with a 488 nm excitation filter and an LP515 emission filter. To evaluate the promoter responsiveness to growth regulators, 6-7-day-old transgenic cell lines were transferred to fresh media containing indole 3-butyric acid (IBA, 10 μM), isopentenyl *adenine* (IP, 10 μM) or abscisic acid (ABA, 20 μM) and incubated for 24 h. The hormone concentrations were chosen based on previous reports (Bueno *et al.*, 1998; Motyka *et al.*, 2003; Campanoni and Nick, 2005). Three biological replicates were used for each hormone treatment. Samples were frozen in liquid nitrogen and stored at -70 °C until RNA extraction. Reporter gene mRNA was quantified by qPCR as described below.

Samples for developmental expression pattern

Sugarcane (*Saccharum* sp.) variety R-570 plants were cultivated in a greenhouse at room temperature and with natural illumination. Culm fragments containing internodes were set in plastic bags with soil and vermiculite (1:1). Shoot meristem, leaf and root samples were collected at three stages of development (15 and 30 days after bud emergence and from adult plants). The material was immediately frozen in liquid nitrogen and stored at -70 °C until RNA extraction.

Quantitative RT-PCR (qPCR)

Total RNA of BY2 or sugarcane samples was extracted using the Trizol reagent (Invitrogen), according to the manufacturer's instructions. The RNA concentration of the samples was measured with a NanoDrop ND-100 spectrophotometer (NanoDrop Technologies). The intactness of the RNA samples was assessed by agarose gel electrophoresis. DNA contamination was removed by treating with DNAse I (Invitrogen). The absence of DNA was confirmed by PCR amplification (Almeida *et al.*, 2011) using actin intron-flanking and *mustang* promoter primers for BY2 and sugarcane RNA, respectively. For cDNA synthesis, Superscript III reverse transcriptase (Invitrogen) was used. For the experiments with BY2 cell cultures, specific primers for the *gus* reporter gene that amplified a 131 bp fragment were designed. Ribosomal L25 and 1α-elongation factor housekeeping genes were used to normalize the mRNA expression levels. The expression of sugarcane *mustang* genes was profiled with class-specific primers. The amplicon sizes were 170 bp and 171 bp for Classes III and IV, respectively. Ubiquitin was used as a housekeeping control. All of the primer sequences are shown in Table S2.

The amount of amplified mRNA was calculated using the Ct (cycle threshold) values based on a gene-specific standard curve constructed with serial cDNA dilutions (1/3, 1/9, 1/27 and 1/81). qPCR reactions were done with 1 μL of the 1/3 cDNA dilution (~15 ng), 200-300 nM of each primer and 12.5 μL of SYBR Green® Master Mix (Applied Biosystems) in a final volume of 25 μL. The cycling conditions were 10 min at 95 °C, 40 cycles of 20 s at 95 °C, 20 s at the primer-specific annealing temperature and 60 s at 72 °C; the reactions were done in a 7500 PCR Real Time thermocycler (Applied Biosystems). The results were expressed as the mean of three biological replicates normalized with respect to the expression of the housekeeping gene and were analyzed statistically using InfoStat software (InfoStat Group, FCA, Universidad Nacional de Córdoba, Argentina) and the non-parametric Kruskal-Wallis test.

Results

mustang genes in grasses

To gain understanding about the evolutionary history and diversity of the *mustang* gene family in grasses, an *in silico* search of the recently sequenced genomes of *S. bicolor* and *Z. mays* was undertaken. By using sugarcane and rice sequences as queries (Table S1) six and nine copies of *mustang* genes were retrieved from the sorghum and maize genomes, respectively. Full-length protein sequences of *Arabidopsis*, rice, sorghum, maize and the newly obtained sugarcane *mustang* genomic copies (SCMUG266 BAC095, SCMUG266 BAC148, and SCMUG148 BAC249) were used in a phenetic analysis. The tree topology agreed with that previously reported by our group (Saccaro-Jr *et al.*, 2007). A single domestication event prior to the monocot-eudicot divergence was confirmed and Classes III and IV appeared as sister groups. Two Class III and three Class IV clades were also identified. Grass orthologs grouped together within each clade mostly in agreement with the species' phylogenetic relationships. For every clade, the *S. bicolor* orthologous gene was identified. In contrast, multiple *Z. mays* sequences were identified for Class III clades while the ortholog for Class IV clade 2 was missing. As expected, SCMUG266 BAC095 and SCMUG266 BAC148 Class III sugarcane genes, from *S. officinarum* and *S. spontaneum*, respectively, that were identified with the same cDNA clones (Saccaro-Jr *et al.*, 2007), clustered together (Figure 1).

After identifying the orthologous groups a detailed comparative analysis of the coding and non-coding regions was done for *mustang* genes belonging to Class III clade 1 and Class IV clade 3. The 2,000 bp regions upstream of the translation initiation codon, or up to the neighbor coding sequence, were considered as the putative promoter regulatory region. Regions with > 90% identity were identified in the sugarcane, sorghum and maize coding sequences,

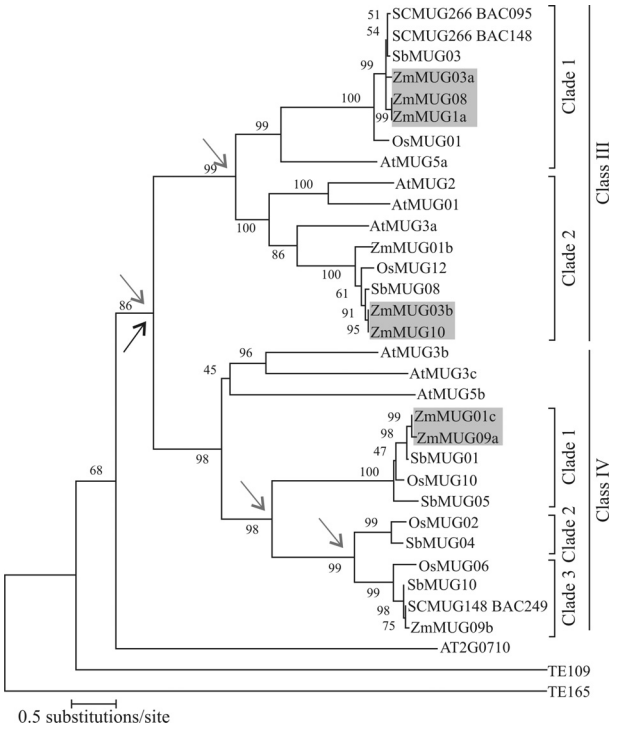

Figure 1 - Phenogram of *mustang* genes in grasses. AtMUGXX, OsMUGXX, SCMUGXX, SbMUGXX and ZmMUGXX are *mustang* genes of *Arabidopsis*, rice, sugarcane, sorghum and maize, respectively. All the rice and *Arabidopsis* sequences used were described by Saccaro-Jr *et al.* (2007). The sorghum and maize sequences are available at http://www.phytozome.net/sorghum and http://www.maizegdb.org. The shaded maize sequences probably arose through polyploidization. Brackets indicate orthologous groups. The black arrow indicates the possible domestication event. Grey arrows indicate duplication events. Numbers on the branches indicate bootstrap values greater than 50%.

whereas the identity between these species and rice decreased to ~70% and ~80% for Classes III and IV, respectively. For the regulatory regions, the identities ranged between 60% and 40%, except for both Class III sugarcane promoters which showed 81% identity (Table 1).

Promoter-reporter gene fusions and transient expression

The putative promoters of the three sugarcane haplotypes were functionally characterized in order to gain insight into the function of *mustang* genes and the conservation of several slightly different copies during angiosperm evolution. The regulatory sequences (Table 1) were amplified and cloned into the PHGWFS7 binary vector upstream of the *gfp-gus* fusion (Figure 2A). The three sequences studied in the transient expression assay were found to be active promoters. Both copies of the Class III *mustang* genes were active, which suggested functional redundancy (Figure 2B).

mustang promoter activity in a heterologous system

BY2 cell lines were stably transformed via *Agrobacterium* using the previously described constructs (Figure 2A). After growth in selective medium and confirmation of the presence of the transgene by PCR five transgenic lines were selected for each promoter-reporter gene fusion and the presence of both reporter proteins was analyzed. GFP and GUS were identified in all of the transgenic lines by confocal microscopy and a histochemical assay, respectively (Figure 3). Reporter gene expression was clearly higher in the lines that expressed the *gfp-gus* fusion under control of the *mustang* Class IV promoter.

After demonstrating that the regulatory regions of the *mustang* genes are indeed active promoters, and motivated by the presence of hormone-responsive DNA elements (Daniela Kajihara, unpublished observation), we investigated the effect of growth regulators on the transcription level of these genes. Transgenic cell lines were treated with IBA (indol-butyric acid) and IP (isopentenyladenine), the most abundant naturally occurring auxin and cytokinin, respectively, and the stress-related hormone ABA (abscisic acid). The levels of *gus* mRNA were quantified by qPCR

Table 1 - Percentage of identity between orthologous groups of *mustang* genes from grasses.

Class III	SCMUG266BAC148	SbMUG03	ZmMUG08	OsMUG01
SCMUG266BAC095 1685/1785[a]	81/98	44/97	42/91	45/72
SCMUG266BAC148 1921/1785[a]	-	45/97	44/91	46/73
SbMUG03 2000/1785[a]	-	-	43/90	42/72
ZmMUG08 2000/1725[a]	-	-	-	42/69
OsMUG01 2000/2070[a]	-	-	-	-

Class IV	SbMUG10	ZmMUG09b	OsMUG06
SCMUG148BAC249 2000/2237[a]	45/95	46/93	40/81
SbMUG10 2000/2163[a]	-	59/96	43/84
ZmMUG09b 2000/2233[a]	-	-	42/84
OsMUG06 1444/2169[a]	-	-	-

SCMUGXX: sugarcane sequences. OsMUGXX: rice sequences. SbMUGXX: *Sorghum bicolor* sequences. ZmMUGXX: *Zea mays* sequences. [a]Promoter size and coding region size in bp.

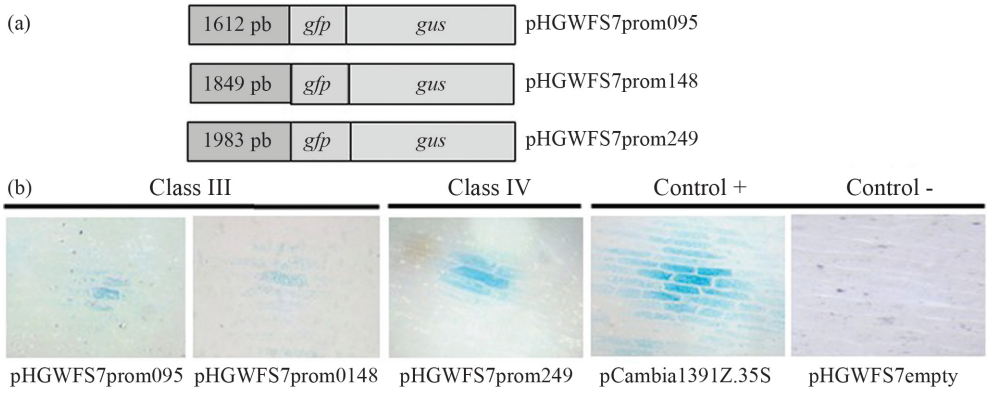

Figure 2 - Promoter-reporter gene fusions. (A) Promoter-*gfp-gus* fusion constructs for SCMUG266 BAC095, SCMUG266 BAC148 and SCMUG148 BAC249 genes. (B) Histochemical analysis of GUS activity in onion epidermis bombarded with pHGWFS7prom095, pHGWFS7prom148 and pHGWFS7prom249 plasmids. The vectors pCambia1391Z.35S and pHGWFS7empty were used as positive and negative controls, respectively. The presence of GUS activity is indicated by the blue staining.

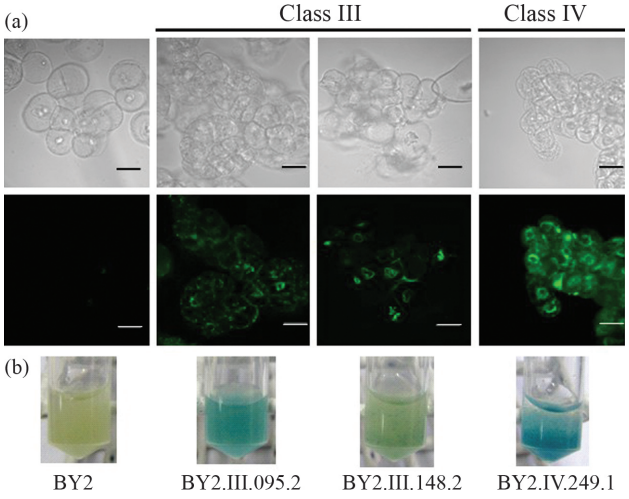

Figure 3 - Transcriptional activity of *mustang* gene promoters in BY2 transgenic cell lines. (A) Light and fluorescent (GFP) microscopy images of transgenic lines carrying pHGWFS7prom095 (BY2.095.2), pHGWFS7prom148 (BY2.148.2) and pHGWFS7prom249 (BY2.249.1) and the corresponding wild type negative control (BY2). White and black scale bars: 25 μm. (B) Histochemical detection of GUS activity in the same BY2 cell lines.

and expressed relative to that of untreated cells (Figure 4). Interestingly, the promoters of SCMUG266 BAC148 and SCMUG148 BAC249 (Classes III and IV, respectively) showed a similar reduction of 30%-40% in their expression in response to treatment with auxin. The SCMUG266 BAC095 gene was only slightly affected by cytokinin and abscisic acid (~10% decrease in relative expression).

mustang expression profile in sugarcane

The transcription profile of sugarcane *mustang* genes *in planta* was assessed by qPCR in leaf, shoot apical meristem and root during plant development from bud emergence until adult age (Figure 5). In this case, class-specific primers were used and all copies within each class were amplified. Both classes were expressed in all three tissues and stages tested. In general, the differences observed showed low statistical significance. The Class III *mustang* genes showed higher expression in shoot meristem at all stages. The expression in leaves and meristems was maximal in 30-day-old plants while roots had the highest expression in adult plants. For Class IV, the expression was similar in all tissues but slightly higher in shoot meristem. Class IV genes were generally more expressed than Class III genes.

Discussion

The grasses exhibited a diverse spectrum of genome size and complexity (Gaut *et al.*, 2000). At one end of the spectrum, diploid rice (*O. sativa*) has 12 chromosomes (2n = 24) and a haploid genome size of 389 Mb (Matsumoto *et al.*, 2005) while at the other extreme, sugarcane cultivars, which have developed through several generations of crosses since the early interspecific hybrids, are highly polyploid and aneuploidy, with 100-130 chromosomes and a genome size (2C) of around 10,000 Mb (D'Hont, 2005). Diploid maize (*Z. mays*) is of intermediate size, with a haploid genome containing 3,283 Mb and 10 chromosomes (2n = 20) (Gaut *et al.*, 2000). This wide diversity, together with the recent availability of genomic sequences for sugarcane, maize and sorghum, led us to study the distribution of *mustang* genes in grass genomes.

Nine genes were identified in *Z. mays* and six orthologs in *S. bicolor*. The phenetic analysis of *mustang* genes corroborated the existence of two and three clades within Classes III and IV, respectively, as reported by Saccaro-Jr *et al.* (2007). Within each clade, the tree topology and sequence identities reflected the phylogenetic relationships among the studied species (Roulin *et al.*, 2009). As expected, sorghum *mustang* genes followed the same distribution pattern as rice orthologs since both are diploid species and show no recent duplication event (Adams and Wendel, 2005). In contrast, as with sugarcane in which polyploidy has produced several haplotypes per clade

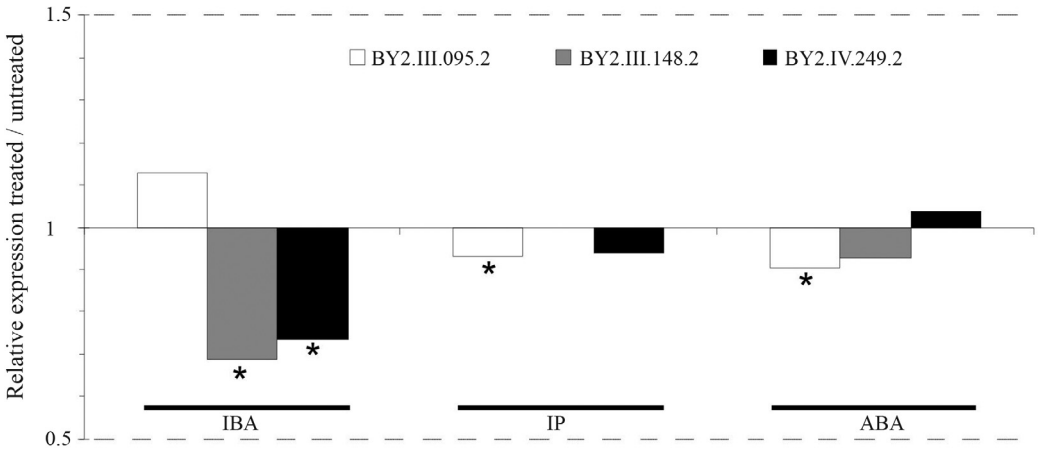

Figure 4 - Transcriptional activity of *mustang* gene promoters in response to auxin, cytokinin and abscisic acid. BY2 transgenic cell lines were treated for 24 h with auxin (IBA), cytokinin (IP) and abscisic acid (ABA). The experiment was done in triplicate. The asterisks indicate significance differences compared to the untreated control ($p < 0.05$; Kruskal-Wallis test).

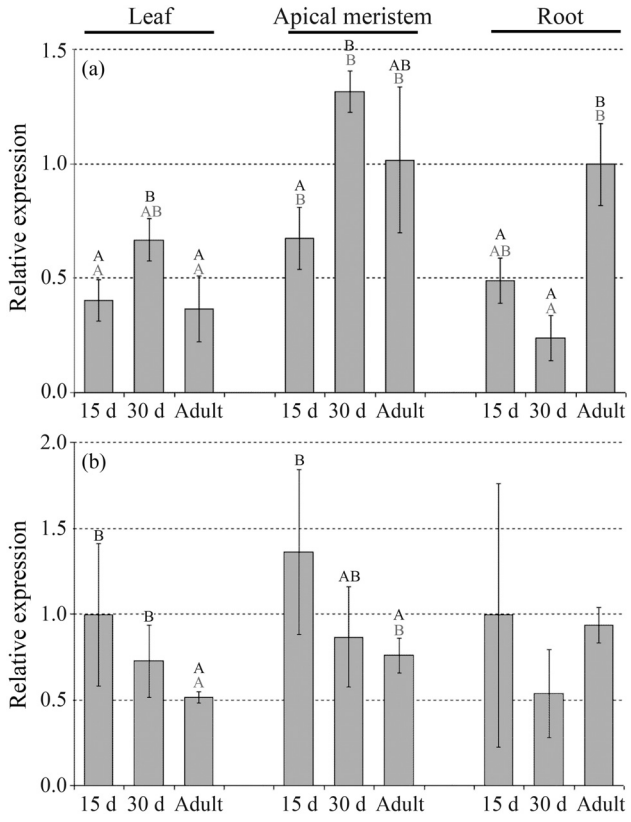

Figure 5 - Expression of Class III and IV *mustang* genes during the development of sugarcane plants. Expression of Class III (A) and Class IV (B) *mustang* genes in leaf, apical meristem and root of 15-day-old (15 d), 30-day-old (30 d) and adult (Adult) plants. The columns are the mean ± SE of three independent experiments. Different letters indicate significant differences ($p < 0.05$; Kruskal-Wallis test). Black letters indicate comparisons between different ages of the same tissue. Gray letters indicate comparisons between different tissues of the same age.

(Saccaro-Jr *et al.* 2007), maize also displayed multiple *loci*. This result reflects the origin of the maize genome. Most modern plant species have evidence of polyploidization followed by massive silencing and gene loss; this also seems to be the case for maize (Adams and Wendel, 2005). The *Z. mays* diploid progenitors diverged 20.5 million years ago (MYA) and, after the divergence of sorghum 16.5-11.4 MYA, a segmental allotetraploidization event occurred followed by a "rediploidization" (Gaut *et al.*, 2000). As a result, 72% of single copy genes in rice or sorghum are duplicated in the maize genome.

Genes whose products are transcription factors or that participate in signal transduction are generally retained after genome duplication, resulting in greater plasticity and better organism adaptation (Gaut *et al.*, 2000; Adams and Wendel, 2005). This phenomenon can explain the multiple Class III *mustang* genes identified in maize given that their protein products harbor the transposase DNA-binding domain (Rossi *et al.*, 2004) and all domesticated transposases that have been functionally characterized so far are transcription factors. These findings indicate that polyploidization involves more than the simple fusion of progenitor genomes, but rather a wide repertoire of molecular and physiological adjustments. These events cover genomic rearrangements, gene loss and changes in gene expression. In addition, polyploidization protects the genome from transpositional mutations (Gaut *et al.* 2000). These various actions indicate that polyploidization is an important force in plant evolution and the differential gene content among species reflects the evolutionary history of the genomes.

The results of the phenetic and comparative sequence analyses raise several questions. Why are there two classes of *mustang* genes in plants? Why have both classes undergone duplications and diverged into different clades? Is there any functional specificity underlining this apparent redundancy? To address these complex issues, we examined the activity of sugarcane *mustang* genes at the transcriptional level. Transient expression assays with promoter-reporter gene fusions showed that Class III (SCMUG266BAC095 and SCMUG266BAC148) and Class IV (SCMUG148BAC249) genes harbor active promoters. Stably transformed cell lines displayed constitutive

reporter gene expression regardless of the upstream regulatory region. Moreover, all three promoters were negatively modulated by phytohormones and no class-specific response was observed. While SCMUG266BAC148 (Class III) and SCMUG148BAC249 (Class IV) were repressed by auxin, SCMUG266BAC095 (Class III) showed reduced mRNA levels upon treatment with cytokinin and abscisic acid. Auxin caused the strongest repression. These results, together with the high level of conservation along angiosperms, suggest that the functional diversification seen here contributed to the positive selection of the different *mustang* genes and helped to retain them in the sugarcane genome (Adams and Wendel, 2005). The responsiveness of the SCMUG266BAC095 promoter to cytokinin and abscisic acid would appear to be contradictory even though both hormones participate in the control of the cell cycle checkpoints G2-M and G1-S, respectively (Swiatek *et al.*, 2002; Wolters and Jürgens, 2009).

Comparison of the *mustang* transcriptional behavior seen in the heterologous system with the expression pattern seen in sugarcane leads to several interesting conclusions. The three genes studied were ubiquitously expressed, with little differences in the intensity of expression; the Class IV regulatory region showed the highest levels of transcription. The tissue in which they were most abundantly transcribed was shoot apical meristem. This finding agrees with the expression pattern of the MUG1 rice *mustang* gene (Class III, clade 1) for which shoot apical meristem displayed the highest level of expression, together with the flowers (Kwon *et al.*, 2009).

The analysis of hormone-mediated regulation is still a complicated task because of the extensive functional redundancy among gene family members, as well as feedback regulation and crosstalk within and among different hormonal pathways. Cytokinin promotes the proliferation of stem cell daughters by inhibiting their differentiation, whereas auxin initiates and maintains the population of organ founder cells (Wolters and Jürgens, 2009). These opposite stimuli must act in a spatially and temporally precise manner. Consequently, an efficient regulatory mechanism that allows the system to circumscribe the hormonal response is necessary. A competent solution for this has already been reported (Hudson *et al.*, 2003; Lin *et al.*, 2007). The *Arabidopsis* mutants *fhy3* and *far1* display a phenotype of reduced inhibition of hypocotyl elongation that is specific for far-red light and the phytochrome A (phyA) signaling pathway. Functional analyses demonstrated that FHY3 and FAR1, which are also related to *Mutator* transposases, are transcription factors whose expression is negatively regulated by phyA signaling. These findings led to the proposal that these proteins modulate the homeostasis of phyA signaling homeostasis in higher plants in response to light (Hudson *et al.*, 2003; Lin *et al.*, 2007). Overall, the functional data reported here suggest that a similar mechanism could be valid for *mustang* genes. In this regard, we propose that these genes may downregulate signaling pathways in order to accurately control temporal and spatial responses to hormones.

In conclusion, we consider that this study contributes to the understanding of the evolutionary history of *mustang* genes in grass genomes and provides new insights about the biological processes in which these domesticated transposases participate, particularly the response to hormones. Our results also provide functional evidence to support the proposition that transposable elements can serve as a source of new transcription factors that allow populations to adapt and species to evolve (Biémont and Vieira, 2006). Additional functional studies are required to test this hypothesis and provide more information about the functionality of *mustang* genes, thereby reinforcing the role of TEs in the evolution of the plant genome.

Acknowledgments

D.K. and T.A.H. were supported by student scholarships from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) and F.G. was so by a scholarship from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil). This work was partially supported by grants from FAPESP and CNPq and was done in compliance with current laws governing genetic experimentation in Brazil.

References

Adams KL and Wendel JF (2005) Novel patterns of gene expression in polyploid plants. Trends Genet 21:539-543.

Almeida J, Quadrana L, Asís R, Setta N, de Godoy F, Bermúdez L, Otaiza SN, Corrêa da Silva JV, Fernie AR, Carrari F and Rossi M (2011) Genetic dissection of vitamin E biosynthesis in tomato. J Exp Bot 62:3781-3798.

An G (1985) High efficiency transformation of cultured tobacco cells. Plant Physiol 79:568-570.

Araújo PG, Rossi M, de Jesus EM, Saccaro Jr NL, Kajihara D, Massa R, Felix JM, Drummond RD, Falco MC, Chabregas SM, *et al.* (2005) Transcriptionally active transposable elements in recent hybrid sugarcane. Plant J 44:707-717.

Aziz RK, Breitbart M and Edwards RA (2010) Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Res 38:4207-4217.

Biémont C and Vieira C (2006) Genetics: Junk DNA as an evolutionary force. Nature 443:521-524.

Bonnivard E and Higuet D (2008) Fluidity of eukaryotic genomes. C R Biol 332:234-240.

Bundock P and Hooykaas P (2005) An *Arabidopsis* hAT-like transposase is essential for plant development. Nature 436:282-284.

Bueno P, Piqueras A, Kurepa J, Savouré A, Verbruggen N, Van Montagu M and Inzé D (1998) Expression of antioxidant enzymes in response to abscisic acid and high osmoticum in tobacco BY-2 cell cultures. Plant Sci 138:27-34.

Campanoni P and Nick P (2005) Auxin-dependent cell division and cell elongation. 1-naphthaleneacetic acid and 2,4-dichlorophenoxyacetic acid activate different pathways. Plant Physiol 137:939-948.

Cowan RK, Hoen DR, Schoen DJ and Bureau TE (2005) MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. Mol Biol Evol 22:2084-2089.

D'Hont A and Glaszman JC (2001) Sugarcane genome analysis with molecular markers, a first decade of research. Proc Int Soc Sugarcane Technol 24:556-559.

D'Hont A (2005) Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenet Genome Res 109:27-33.

Devos K (2010) Grass genome organization and evolution. Curr Opin Plant Biol 13:139-145.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev 9:397-405.

Gaut BS, Le Thierry d'Ennequin M, Peek AS and Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci USA 97:7008-7015.

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

Hudson ME, Lisch DR and Quail PH (2003) The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J 34:453-471.

Karimi M, Inze D and Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci 7:193-195.

Kumar S, Tamura K and Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150-163.

Kwon SJ, Park KC, Son JH, Bureau T, Park CH and Kim NS (2009) Sequence diversity of a domesticated transposase gene, MUG1, in *Oryza* species. Mol Cells 27:459-465.

Lin R, Ding L, Casola C, Ripoll DR, Feschotte C and Wang H (2007) Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. Science 318:1302-1305.

Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio BA, Baba T, *et al.* (2005) The map-based sequence of the rice genome. Nature 436:793-800.

McCabe DE, Swan WF and Matineli BJ (1988) Stable transformation of soybean (*Glycine max*) by particle bombarment. Biotechnology 6:923-926.

Motyka V, Vankova R, Capkova V, Petrasek J, Kaminek M and Schmülling T (2003) Cytokinin-induced upregulation of cytokinin oxidase activity in tobacco includes chances in enzyme glycosylation and secretion. Physiol Plant 117:11-21.

Murashige T and Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15:473-497.

Nagata T, Nemoto Y and Hasezawa S (1992) Tobacco BY2 cell-line as the Hela-cell in the cell biology of higher-plants. Int Rev Cytol 132:1-30.

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B and Grandbastien MA (2010) Impact of transposable elements in organization and functioning of allopolyploid genomes. New Phytol 186:37-45.

Rech EL and Aragão FJL (1998) Biolobalística. In: Brasileiro ACM and Carneiro VTC (eds) Manual de Transformação de Genética de Plantas. EMBRAPA, Brasília, pp 99-107.

Rossi M, Araujo PG and Van Sluys MA (2001) Survey of transposable elements in sugarcane expressed sequence tags (ESTs). Genet Mol Biol 24:147-154.

Rossi M, Araújo PG, de Jesus EM, Varani AM and Van Sluys MA (2004) Comparative analysis of sugarcane *Mutator*-like transposases. Mol Genet Genomics 272:194-203.

Roulin A, Piegu B, Fortune PM, Sabot F, D'Hont A, Manicacci D and Panaud O (2009) Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae. BMC Evol Biol 9:e58.

Saccaro-Jr NL, Van Sluys MA, Varani AM and Rossi M (2007) MudrA-like sequences from rice and sugarcane cluster as two bona fide transposon clades and two domesticated transposases. Gene 392:117-125.

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

Sanford JC, DeVit MJ, Russell JA, Smith FD, Harpending PR, Roy MK and Johnston SA (1991) An improved, helium-driven biolistic device. Technique 3:3-16.

Sinzelle L, Izsvák Z and Ivics Z (2009) Molecular domestication of transposable elements: From detrimental parasites to useful host genes. Cell Mol Life Sci 66:1073-1093.

Sorek R (2007) The birth of new exons: Mechanisms and evolutionary consequences. RNA 10:1603-1608.

Swiatek A, Lenjou M, Van Bockstaele D, Inzé D and Van Onckelen H (2002) Differential effect of jasmonic acid and abscisic acid on cell cycle progression in tobacco BY-2 cells. Plant Physiol 128:201-211.

Yaakov B and Kashkush K (2011) Methylation, transcription, and rearrangements of transposable elements in synthetic allopolyploids. Int J Plant Genomics 2011:569826.

Wolters H and Jürgens G (2009) Survival of the flexible: Hormonal growth control and adaptation in plant development. Nat Rev Genet 10:305-317.

Internet Resources

Phytozome database, http://www.phytozome.net/ (accessed 11/2010).

MaizeGDB database, http://wwww.maizegdb.org, (accessed 11/2010).

TIGR database, http://plantta.jcvi.org/, (accessed 11/2010).

TAIR database, http://www.arabidopsis.org/, (accessed 11/2010).

EMBOSS Pairwise Alignment Algorithm, http://www.ebi.ac.uk/Tools/emboss/align (on line accessed program).

InfoStat software, http://www.infostat.com (free-download software, after requirement).

Supplementary Material

The following online material is available for this article:

Table S1 - Sequence information on *Mustang* genes identified in plant genomes

Table S2 - Primer used for promoter-*gfp-gus* fusion constructs and real time PCR experiments.

This material is available as part of the online article from http://www.scielo.br/gmb.

Associate Editor: Marcia Pinheiro Margis

Supplementary Table S1. Sequence information.

Species/Database [(1)]	**Name (class)** [(2)]	***Loci*/accession number (position)** [(3)]
Saccharum sp. (sugarcane) GenBank	SCMUG266BAC095 (Class III)	GU080321
	SCMUG266BAC148 (Class III)	GU080318
	SCMUG148BAC249 (Class IV)	GU080320
	TE165	DQ115055
	TE109	DQ115045
Sorghum bicolor Phytozome	SbMUG01 (Class IV)	Sb01g027410
	SbMUG03 (Class III)	Sb03g002010
	SbMUG04 (Class IV)	Sb04g006220
	SbMUG05 (Class IV)	Sb05g027680
	SbMUG08 (Class III)	Sb08g020220
	SbMUG10 (Class IV)	Sb10g024700
Oryza sativa TIGR	OsMUG01 (Class III)	LOC_Os01g41210
	OsMUG02 (Class IV)	LOC_Os02g09900
	OsMUG06 (Class IV)	LOC_Os06g42640
	OsMUG10 (Class IV)	LOC_Os10g01550
	OsMUG12 (Class III)	LOC_Os12g40530
Arabidopsis thaliana TAIR	AtMUG5a (Class III)	AT5G16505.1
	AtMUG5b (Class IV)	AT5G34853.1
	AtMUG2 (Class III)	AT2G30640.1
	AtMUG01 (Class III)	AT1G06740.1
	AtMUG3a (Class III)	AT3G04605.1
	AtMUG3b (Class IV)	AT3G05850.1
	AtMUG3c (Class IV)	AT3G06940.1
	AT2G07100	AT2G07100
Zea mays MaizeGDB	ZmMUG01a (Class III)	AC177909 (246725-245351)
	ZmMUG03a (Class III)	AC190982 (202769-204551)
	ZmMUG08 (Class III)	AC201818 (110632-112416)
	ZmMUG01b (Class III)	AC204517 (72081-74315)
	ZmMUG10 (Class III)	AC194150 (47510-49765)
	ZmMUG03b (Class III)	AC221000 (28198-30453)
	ZmMUG01c (Class IV)	AC202094 (94687-100808)
	ZmMUG09a (Class IV)	AC215910 (131798-132870)
	ZmMUG09b (Class IV)	AC209850 (21913-24150)

Mustang genes identified in sugarcane (Saccaro-Jr et al., 2007), S. bicolor (this work), O. sativa (Cowan et al., 2005), A. thaliana (Cowan et al., 2005) and Z. mays (this work) genomes.

(1) Genomic sequence database: http://www.ncbi.nlm.nih.gov/ for sugarcane, http://www.phytozome.net/ for *S. bicolor*, http://plantta.jcvi.org/ for *O. sativa*, http://www.arabidopsis.org/ for *A. thaliana*, and http://wwww.maizegdb.org/ for *Z. mays*.

(2) Name used in the phenetic analysis shown in Figure 1 and the corresponding class between brackets.

(3) *Loci* identification or accession number (nucleotide position) in the corresponding database.

Supplementary Table S2. Primer used for promoter-*gfp-gus* fusion constructs and real time PCR experiments.

Primer name	**Amplified *loci***	**Sequence 5´-3´ (1)**
ClassIII-promF	SCMUG266 BAC148	CCTTGTGAGAAACCTTCAAG
ClassIII-promR	SCMUG266 BAC095	CGACTCCGCTCCACTCTC
ClassIV-promF	SCMUG148BAC249	CACCTGTACATGATCCAAGGAGT
ClassIV-promR		GGTCTGGCTCGGGGATAAG
actinF	actin	TGGCATCATACCTTTTACAA
actinR		TCCGGGCATCTGAACCCTCT
gusF	*gus* reporter gene	GGACAAGGCACTAGCGGGACTT
gusR		CCGACGCGAAGCGGGTAGATAT
L25F	ribosomal L25	CCCCTCACCACAGAGTCTGC
L25R		AAGGGTGTTGTTGTCCTCAATCTT
EFF	1α-elongation factor	TGAGATGCACCACGAAGCTC
EFR		CCAACATTGTCACCAGGAAGTG
ClassIIIF	SCMUG266 BAC148	CCGGACAGGACCCACTGG
ClassIIIR	SCMUG266 BAC095	CGACCACACTGAACGATCCG
ClassIVF	SCMUG148BAC249	TTAGCTGTTCTTCAAATAGGTGGAG
ClassIVR		GAATGCGAACTGGTCAGGC
UbiF	ubiquitin	CGTCCGCAGTCCCCAAT
UbiR		TGAGAGGATCGCGAGGATTC

(1) The primers were designed for the present paper.