Plant Omics Journal POJ

POJ 5(2):103-114 (2012) ISSN:1836-3644

Molecular cloning and characterization of two Trithorax-Group genes from *Phaseolus vulgaris* roots and symbiotic nodules

Juan Manuel Quiceno-Rico, José Alberto Camas-Reyes, Raúl Alvarez-Venegas*

CINVESTAV Unidad Irapuato, Departamento de Ingeniería Genética, Km. 9.6 Libramiento Norte, Carretera Irapuato-León C.P. 36821, Irapuato, Guanajuato, México

***Corresponding author: ralvarez@ira.cinvestav.mx**

Abstract

In eukaryotes, trithorax group proteins play critical roles in the regulation of transcription, cell proliferation, differentiation and development. In this work we report the molecular cloning and characterization of two cDNAs, *PvuTRX1h* and *PvuASH1h,* from the common bean *Phaseolus vulgaris*, both of which encode polypeptides homologues of trithorax group members described in animals and yeast. A full length clone of *PvuTRX1h* was isolated from total RNA prepared from roots and consisting of a 3270 bp ORF encoding 1089 amino acids, while the *PvuAsh1h* consists of a 1446 bp ORF encoding 481 amino acids. Characterization of the isolated sequences revealed that they contain all the canonical domains present in proteins from the TRX (trithorax) and ASH1 families. A comparison of the *PvuTRX1h* and *PvuASH1h* SET-domain sequences with homologous proteins from plants, animals and yeast, revealed that *PvuTRX1h* is phylogenetically related to the TRX family of histone lysine methyltransferases while *PvuASH1h* clusters with members of the ASH1 family. Quantitative RT-PCR (reverse transcription polymerase chain reaction) analyses of transcript abundance in roots and nodules, at different developmental stages, demonstrated that *PvuTRX1h* is particularly abundant at early stages of nodule development, whereas *PvuASH1h* functions at the stages of highest nitrogen-fixing activity of the nodules, suggesting that these genes could be involved in the formation of nitrogen-fixing nodules in *P. vulgaris*. This work reports the presence and characterization of Trithorax-group homolog genes in *P. vulgaris* and their expression patterns during nodule development.

Keywords: epigenetics; phylogenetics; RACE; SET-domain; trithorax.
Abbreviations: ASH1- ABSENT, SMALL, OR HOMEOTIC DISCS-1; ASHH1- ABSENT, SMALL, OR HOMEOTIC DISCS HOMOLOG-1; AWS- domain Associated With SET; ATX1- ARABIDOPSIS HOMOLOG OF TRITHORAX-1; DAST- Domain Associated with SET in Trithorax; dpi- days post-inoculation; KEGG- Kyoto Encyclopedia of Genes and Genomes; ORF- Open Reading Frame; PcG- Polycomb group; PHD- plant homeo domain; PWWP domain- for conserved Pro-Trp-Trp-Pro motif; RACE- rapid amplification of cDNA ends; SET- SET domain (for Su(var)3-9, E(z) and Trithorax); TILLING- Targeting Induced Local Lesions IN Genomes; TRX- Trithorax; trxG- trithorax group.

Introduction

It is now evident that cell development, cell differentiation, DNA repair, senescence, disease and cancer are under epigenetic control (Kouzarides, 2007; Surani et al., 2007). Epigenetic information is partly carried by reversible covalent modifications at the N-termini of histone proteins. In particular, the functional consequences of histone post-translational modifications (PTMs) can be either direct, causing structural changes to chromatin, or indirect, acting through the recruitment of effector proteins. Thus, epigenetic mechanisms play fundamental roles in shaping and maintaining cell identity and in patterning the body plan during development (Kiefer, 2007). In *Drosophila melanogaster*, the *Polycomb* group (PcG) proteins maintain a repressive state of homeotic gene (*HOX* gene) expression, while the *trithorax* group (trxG) proteins maintain *HOX* gene activity (Simons and Tamkun, 2002). The molecular analysis of PcG and trxG genes has revealed that their products act as large multimeric complexes at the level of chromatin structure, and that they are not required to initiate the regulation of *HOX* genes, but rather to maintain their expression state after the initial transcriptional regulators disappear from the embryo (Franke, et al., 1992; Dingwall et al., 1995; Papoulas et al., 1998; Shao et al., 1999; Schuettengruber et al., 2007). PcG and trxG genes are evolutionarily conserved in plants (Goodrich et al., 1997; Alvarez-Venegas and Avramova 2002; Springer et al., 2003; Thakur et al., 2003). In *Arabidopsis thaliana*, for example, several plant PcG genes have been identified in forward genetic screens for mutations affecting flowering time, flower and seed development and the vernalization response (Goodrich et al., 1997; Grossniklaus et al., 1998; Jackson et al., 2004; Köhler et al., 2003; Kim et al., 2005; reviewed in Alvarez-Venegas, 2010). On the other hand, the function of trxG genes in Arabidopsis and other plant model systems remains mostly unknown. For example, disruption of known plant Trithorax proteins, such as ARABIDOPSIS HOMOLOG OF TRITHORAX-1 (ATX1), a SET-domain containing protein with histone lysine methyltransferase activity (homolog to the Drosophila Trithorax gene), causes pleiotropic phenotypes, including homeotic stem growth, root and leaf defects, and is required to maintain normal expression levels of homeotic genes during flower

development (Alvarez-Venegas et al., 2003). ATX1 was shown to bind the *AGAMOUS* (*AG)* chromatin and to be required for the tri-methylation of lysine 4 at histone H3 (H3K4me3 deposition) at this locus (Saleh et al., 2007). In addition, ATX1 is directly involved in 'writing' the H3K4me3 marks on the *FLOWERING LOCUS C* (*FLC)* nucleosomes (Saleh et al., 2008a), and H3K4me3 deposition is accompanied by a decrease in H3K27me2 levels at the *FLC* locus (Pien et al., 2008). Furthermore, the expression of *ATX1* in roots is under developmental, in addition to tissue-specific, regulation, and the promoter of *ATX1* has been shown to be active in vascular tissues of six-days-old and two-week-old roots, while in latter developmental stages *ATX1* is expressed in dividing root tip cells (Saleh et al., 2008b). A second Drosophila SET-domain containing gene, *absent, small, or homeotic discs 1 (Ash1)*, has also been classified as a trxG gene. Accordingly, sequence analysis in Arabidopsis has also identified gene homologs to the Drosophila *Ash1* gene. For example, *ASHH2* (*ABSENT, SMALL, OR HOMEOTIC DISCS HOMOLOG-2*, also known as *EFS*) was originally isolated as a novel early-flowering mutant, *early flowering in short days* (*efs*), involved in controlling an inhibitor of flowering (Soppe et al., 1999). It has been shown that loss of function of *ASHH2* results in reduced dimethylation of lysine 36 of histone H3 (H3K36), particularly in chromatin associated with the *FLC* gene. *ashh2* mutants display reduced *FLC* expression and flower early, establishing ASHH2-mediated H3K36 methylation as a novel epigenetic memory code required for *FLC* expression in preventing early flowering (Zhao et al., 2005). On the other hand, *efs (ashh2)* mutations suppress *FLC* expression in *FRIGIDA (FRI)*-containing or autonomous pathway mutant backgrounds. Lesions in *EFS* also reduce the level of histone H3K4me3 in *FLC* chromatin (Kim et al., 2005). These results indicate that ASHH2 is a multifunctional enzyme with H3K4 and H3K36 methylation activity. Taking into account that a knock-out mutation of *ASHH2* has a pleiotropic effect (including reduced root growth; Soppe et al., 1999), *ashh2* mutants also exhibited increased shoot branching and repression of *SPS* (*SUPERSHOOT*) transcript, suggesting that ASHH2 also plays an important role in regulating the expression of genes controlling shoot branching in Arabidopsis (Dong et al., 2008). Recently, Cazzonelli, (2009) reported that ASHH2 is also involved in regulating carotenoid biosynthesis by modifying the histone methylation status of chromatin surrounding the *CAROTENOID ISOMERASE (CRTISO)* gene. In contrast to the early-flowering phenotype of the *ashh2* mutants, *ashh1* mutants show a late-flowering phenotype associated with up-regulation of the *FLC* gene, suggesting that ASHH1 essentially contributed to maintaining repression of genome transcription (Xu et al., 2008). Furthermore, ASHH1 has been shown to localize to cortex cells of the Arabidopsis root (Xu et al., 2008), although no specific histone methyl-transferase activity has as yet been reported for ASHH1. As part of our research addresses the study of trxG genes involved in the development of nitrogen-fixing nodules in legumes, we have isolated and cloned the full-length cDNAs of *PvuTRX1h* and *PvuASH1h* from *Phaseolus vulgaris*, two genes encoding TRX and ASH1 homolog genes, respectively. The isolated sequences and their protein structures were compared to homologous genes present in other plant genomes, together with homologous genes characterized previously in animals and yeast. Phylogenetic analysis clusters these two genes with their respective TRX and ASH1 families of histone lysine methyltransferases. As an approach to gain further insights into the function of these genes, the expression pattern of *PvTRX1h* and *PvASH1h* was analyzed by quantitative RT-PCR (Q-RT-PCR) during several stages of root and nodule development. Q-PCR analyses from nitrogen-fixing nodules showed that *PvuTRX1h* is particularly abundant at early stages of nodule development, suggesting that this gene could be involved in the formation of nitrogen-fixing nodules in *P. vulgaris*.

Results

Cloning of the Phaseolus vulgaris trithorax homologs

The SET domain [for Su(var)3-9, E(z) and Trithorax], a ~130 amino-acids conserved sequence motif with histone lysine methyl-transferase activity, may be found in proteins belonging to families of both repressors (the Polycomb group, PcG) and activators (the Trithorax group, trxG) (Dorn et al., 1993; Kennison, 1995; Huang et al., 1998; Jenuwein et al., 1998; Sedkov et al., 1999). However, plant gene homologs to the Trithorax superfamily have barely been characterized, especially in Arabidopsis, while nothing is known in legumes. Thus, our goal was to characterize putative gene homologs to the Drosophila Trithorax (TRX) and Ash1 genes (both trx-G members) in the common bean.
In order to attain cDNA clones of the TRX and Ash1 gene homologs in *P. vulgaris*, first, we aligned the *Arabidopsis thaliana* cDNA *ATX1* sequence (At2g31650) with two predicted soybean (*Glycine max*) Trithorax sequences (Glyma08g29010.1 and Glyma18g51890.1) that we identified from the Phytozome database (http://www.phytozome.net/). Next, we looked for conserved regions between all three genes and designed a set of consensus primers to amplify an internal fragment from *P. vulgaris* (Fig S1). With total RNA extracted from three-days old BAT 477 wild type cultivar roots and used as a template in an RT-PCR reaction, a 1300 bp fragment was amplified. After cloning and sequencing this fragment, primers were designed to perform 5'- and 3'-RACE experiments. With the results from these experiments we were able to amplify the whole cDNA sequence, a 3270 bp fragment, using RT-PCR. Subsequently, after sequencing this cDNA, we did a sequence comparison of the entire coding region from the *P. vulgaris* gene, with the Arabidopsis *ATX1* and the two soybean Trithorax-gene homolog sequences. Thus, we determined that the *P. vulgaris* gene (named here *PvuTRX1h*) has an 81% sequence identity at the DNA level with the Arabidopsis *ATX1* (69% at the protein level) and a 93% sequence identity with the Glyma08g29010.1 and Glyma18g51890.1 soybean sequences (both, at the DNA and protein levels) (Fig 1a). It is important to note that the *PvuTRX1h* cDNA (3270 bp) has practically the same length as the Arabidopsis *ATX1* cDNA (3365 bp), as well as the same length for the two soybean sequences used as reference (Glyma08g29010.1 and Glyma18g51890.1; 3267 bp each). This suggested that besides the presence of the evolutionarily conserved SET domain, these proteins should have the same domain architecture. Indeed, we found that the *PvuTRX1h* gene conserves the arrangement and all the structural domains present in the other plant sequences (that is, the PWWP-, DAST- and SET- domains; Fig 1c and S1). Thus, the phylogenetic relationship between these trithorax-like proteins is supported by their architectural content (S1) as well as by the relatedness of their SET domains. Therefore, *Arabidopsis thaliana* and *Phaseolus vulgaris* (as well as *Glycine max*) have inherited the TRX-gene from the common ancestor before the separation of fabids and malvids

(a)

MAAAFQEGGGGDDDATIDVHTTRGTPIRYLPLDHLYSATSPCRVTASGSSNVMSKKVKARKLTAHHHFNNHDHK
KATSSSSSSSQPPSSKPPLLFAYSRRRKRHSPSTAPFYDSLCKTEGEVNADENENEKRLLKKRKIGSTELERLGVDL
NTAIGDVDGPRLRECRNQFGNSGAAGNFKCGSLENLPKGSPDSRPVKKWIGLSFDNADPEVFIGLRCKVYWPMDL
KAYTGYVRSYDKETKIHHVKYDDGDEENLILSNENIRFHVSRDEVKHLKLNFAKVRDNNVSDYDVEEMLAMAA
SLDDCQDFEPGDIIWAKLTGHAMWPAVVLDESLARNCKGLKMILGGRSVPVQFFGTHDFARVRVQQVKSFLSGL
LTDLHSKCKKHSFIEGLEEAKRYLSEQKLPLEMIELQKRCTADNCNNVSGEDGGCTDSGDECLNGERTLMALENV
ETFPYVVGDLQILSLGKIVKDSAFRDGKFIWPEGYTAVRKFTSVTDPEVFAPYKMEVLRDPESKVRPLFRVTVDGG
EQFNGNTPSACWSEVFKKIKKMEKNASEGTLAEGGVEKGYESGSDMFGFSNPKVLKLIKGLSKSKISSKNSFCKL
GSQRHNNLPLGYRQVHINWFDLDKCNVCHMDEEYENNLFLQCDKCRMMVHARCYGELEPVNGVLWLCNLCRS
GAPPPPCCLCPLIGGAMKPTTDGRWAHLACAMWIPETCLADVKRMEPIDGLSRISKDRWKLLCSICGVSYGACIQ
CSNSSCRVAYHPLCARAAGLCVELENEDRLYLLSVDDDEDQCIRLLSFCKKHRQPSNEHSVADDRIVRVAGLCSD
YEPPPNPSGCARSEPYDYFGRRGRKEPEALAAASLKRLFVENQPYIVGGYCQHGALNDPEPSGRGVCSKFFCSQQ
RLRTSLIDTSNNILSISEKYKYMRETFRKRL**AFGKSRIHGFGIFAKHPHKGGDMVIEYTGELVRPPIADRREHFI**
YNSLVGAGTYMFRIDDERVIDATRAGSIAHLINHSCAPNCYSRVISVNGDEHIIIFAKRDIKQWEELTYDYRFF
SIDERLSCYCGFPKCRGIVNDTEAEERAATLYAPRRELIDWRGE-

(b)

MDPFTEELPQYIHINQNEFFMRRQKKQKEEDIAICECKYDANDTDSACGDSCLNVLTSTECTPGYCPCDILCKNQK
FQKCEYAKT**KLFKTEGRGWGLLAGEDLKAGQFVIEYCGEVISWKEAKRRSQAYENQGLKDAFIICLNASESI**
DATRKGSLARFINHSCRPNCETRKWNVLGEIRVGIFAKHDVPIGTELAYDYNFEWFGGAKVRCLCGALKCSG
FLGAKSRGFQEDTYLWEDDDDRYSVEKIPVYDSAEDEPVSNVNGRTESPLDVMLKDEQLSESTGFNVQSLDSVQ
MKGLDVKKIKTEVTDEDMHLYNHDTEQTLSQKNAMISRIRSNAAGRNYHIGPRSMSTKRSRAYNGGRFKNLVEK
KIDAKFAAGLLASKEAQEEILNCEKRKDDATSTLDSLYDEIRPAIEEHERDSQDSVSTTVAEKWIQVCCLKLKAEF
DLYSSIVKNVACTAQRAPGQAKPTEVDNENEIKLLTG-

(c)

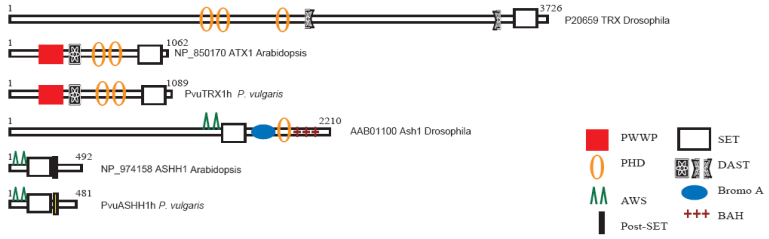

Fig 1. Protein sequences and domain architecture of PvuTRX1h and PvuASH1h. (a) Predicted amino acid sequence from the *PvuTRX1h* cDNA sequence (GenBank: JF262910). The SET-domain is shown in red. (b) Predicted amino acid sequence from the *PvuASH1h* cDNA sequence (GenBank: JF262911). (c) Protein domain comparison of PvuTRX1h and PvuASHH1h with homologous proteins from Drosophila (GenBank P20659 and AAB01100) and Arabidopsis (GenBank NP_850170 and NP_974158). SET- SET domain; DAST- Domain Associated with SET in Trithorax; PWWP- for conserved Pro-Trp-Trp-Pro motif; AWS- Associated With SET; PHD- plant homeo domain; BAH- Bromo Adjacent Homology domain; Bromo_A- Post-SET- post SET domain.

from the rosids-plant lineage, and no additional domain was acquired during the evolution of these SET domain-containing proteins. On the other hand, we used the Arabidopsis *ASHH1* sequence to search the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (DBGET Search - P.vulgaris_est) and identified a highly homologous (81% homology) expressed sequence tag (EST sequence; entry number 13627). However, this *P. vulgaris* EST sequence contained only part of the 5'-UTR and 5'-gene sequences. Then, in order to identify the 3'-end sequence and to clone the full length cDNA, this EST (*epvu13627*) was aligned with two homologous soybean sequences, Glyma20g30870.1 and Glyma10g36720.1 (www.phytozome.net; Fig S2). We identified regions of homology and designed a set of primers to amplify the full length sequence from *P. vulgaris*. The cDNA (consisting of 1446 bp and named here *PvuASH1h*) has a 62% sequence identity at the amino acid level with the Arabidopsis ASHH1 and a 90% sequence identity at the amino acid level with both Glyma20g30870.1 and Glyma10g36720.1 soybean sequences (Fig 1b). In addition, the *PvuASH1h* homolog also conserves the arrangement and all the domains present in homologous sequences (that is, the AWS-, SET-, and Post-SET- domains; Fig 1c). Both cDNA sequences, *PvuTRX1h* and *PvuASH1h*, have been deposited in the GenBank database (accession numbers JF262910 and JF262911, respectively).

Protein structure prediction of the Phaseolus and Arabidopsis trithorax homologs

To gain insight into the biological function of the *P. vulgaris* PvuTRX1h and PvuASH1h proteins, we determined their predicted protein structures by using the "iterative threading assembly refinement" (I-TASSER) program (Roy et al., 2010). We used the SET-domain structure of the human SET7/9 (Kwon et al., 2003; Xiao et al., 2003) protein, as the known 3D model, to train the I-TASSER program, in order to search for possible topological similarities and to build a reasonable model of the SET-domain structures of the PvuTRX1h and PvuASH1h sequences. The I-TASSER predicted structures are presented in Figure 2. As shown, in all cases, the SET-domain forms the classical β-fold configuration, with a series of curved β-strands forming several small sheets, packed together with the N-SET and C-SET domains or regions. In all the predicted structures, the SET domain is folded into several small β-sheets surrounding

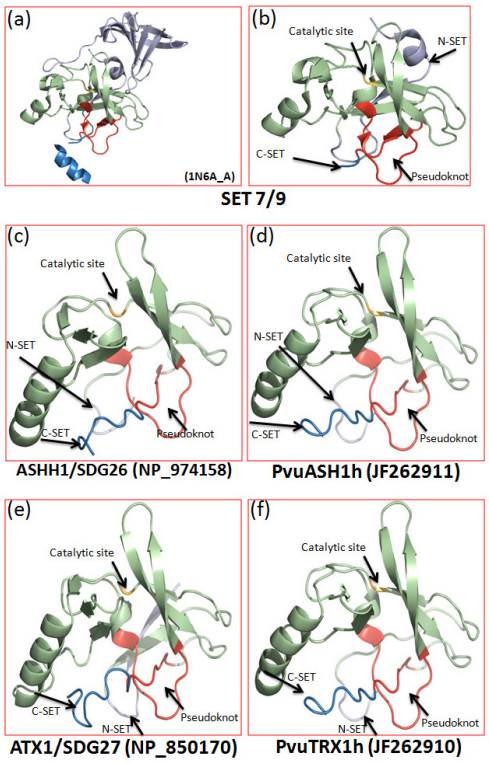

Fig 2. SET-domain structure prediction. Protein structure prediction of the: (a) human SET7/9 full-length protein; (b) human SET7/9 SET-domain only; (c) the Arabidopsis ASHH1 SET-domain; (e) the Arabidopsis ATX1 SET-domains, and the *P. vulgaris* (d) PvuASH1h and (f) PvuTRX1h SET-domains. In all cases, the SET-domain forms the classical β-fold configuration, with a series of curved β-strands forming several small sheets, surrounding a knot-like structure or 'pseudoknot' that brings together the two most-conserved sequence motifs of the SET domain (RFINHXCXPN and ELx[F/Y/W/]DY) to form the active site in a location immediately next to the pocket where the methyl donor binds. Arrow heads indicate the conserved regions present in all the sequences: N-SET, catalytic site, Pseudoknot, and C-SET.

a knot-like structure or 'pseudoknot' that brings together the two most conserved sequence motifs of the SET domain (RFINHXCXPN and ELx[F/Y/W/]DY) to form the active site in a location immediately next to the pocket where the methyl donor binds (Fig 3; Dillon et al., 2005). Thus, the predicted PvuTRX1h and PvuASH1h proteins (as well as the Arabidopsis proteins) have similar sequence motifs surrounding the SET domain, and also share a higher level of similarity in the SET domain (Fig 3), making them putative histone lysine methyltransferases. That is, they have the aromatic ring amino acid at the fourth position in the conserved motif ELx(F/Y/W/)DY, which is important for the specificity of the SET domain peptide in establishing mono-, di-, or trimethylation of a single Lys residue (Zhang et al., 2003; Cheng et al., 2005; Collins et al., 2005; Thorstensen et al., 2006), as well as the invariant residues in binding to AdoMet and the target lysine, the catalytic site, the structural pseudoknot, an intra-molecular interacting salt bridge, and a F/Y switch (Fig 3).

Phylogenetic analysis of the PvuTRX1h and PvuASH1 proteins

To evaluate whether *PvuTRX1h* and *PvuASH1h* are true homologs to the Drosophila Trx and Ash1 genes, a phylogenetic tree was constructed using proteins from several species (Fig 4). Only SET-domain sequences with the highest scores were used to generate Maximum Likelihood phylogenetic trees. Thus, the criterion for distributing the proteins to one of the four different families (Su(var) 3-9, E(z), TRX, and ASH1) was the relatedness of their SET-peptides. The phylogenetic analysis revealed that both proteins were grouped within the clades of their homologs; that is, the *P. vulgaris* homologous proteins grouped with homologs from legumes, Arabidopsis and other plant proteins (Fig 4). On the other hand, the clustering pattern is supported by the architecture of the proteins (Fig 1c; Alvarez-Venegas and Avramova, 2002). This confirmed that PvuTRX1h and PvuASH1h are true trxG homologs in *P. vulgaris* and, as we have previously stated, for an unknown gene, a function similar to an already established gene might be expected when the protein architecture of the unknown gene is the same as the protein architecture of the gene with known function. In cases when only one conserved element is considered (e.g. the SET domain), cladistic analysis of its sequence similarity should place it in the same sister-group within the same clade of the family as the gene with known function (Alvarez-Venegas and Avramova, 2002), as is shown here. However, the clustering of PvuTRX1h and PvuASH1h into specific legume subclades, separated from Arabidopsis, indicated a potential divergent methylation activity from the activity reported in Arabidopsis for their homologs, and this is something that deserves further research.

Analysis of -cis elements at the promoter region of PvuTRX1h

In addition, as has been proposed, a comparison of orthologous genes and their upstream regions can be screened for common regulatory signals (*cis*-acting elements) in order to make functional predictions for an unknown gene (Bork et al., 1998). In particular, the structural properties of several nodule-specific promoters have been studied by carrying out deletion analyses of the promoter regions, leading to the identification of several regulatory elements (e.g. enhancers, organ or cell-specific elements, and strong positive elements), in addition to core promoter motifs (Stougaard et al., 1990; Nakawaga et al., 2003; Rodriguez-Llorente et al., 2003). Thus, in order to identify regulatory elements and to predict the function of *PvuTRX1h* (as a true trxG gene homolog), we compared the *ATX1* promoter region (~1500 bp) with the *PvuTRX1h* promoter region (a 445 bp region upstream of the start codon), for transcription factor (TF) binding motifs or *cis*-acting elements (Methods; Fig 5). We found additional putative TF binding sites upstream of *ATX1* not present in *PvuTRX1h*. However, the *PvuTRX1h* promoter region so far identified is shorter than the *ATX1* region. All motifs from the *PvuTRX1h* promoter region are also present in the *ATX1* promoter region, indicating that similar regulation of the two genes is also possible. Noteworthy is the fact that both promoters have consensus sequence motifs of organ-specific elements (OSE) characteristic of the promoters activated in infected cells of root nodules (the OSE2ROOTNODULE

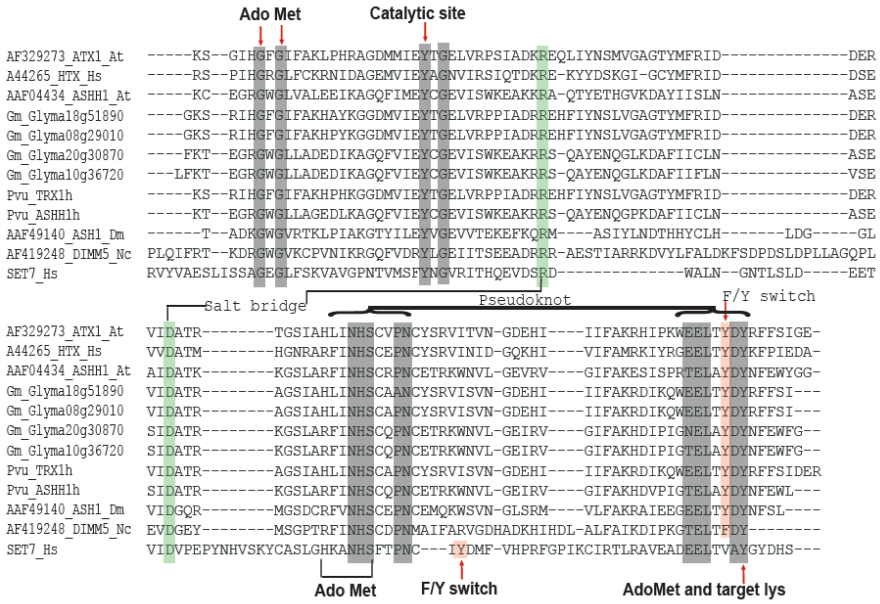

Fig 3. Protein sequence alignment of the SET domains of several representative histone lysine methyltransferases (HKMT). The predicted PvuTRX1h and PvuASH1h proteins (as well as the Arabidopsis proteins) have similar sequence motifs surrounding the SET domain, making them putative histone lysine methyltransferases. This means that they have the aromatic ring amino acid at position 4 in the conserved motif ELx(F/Y/W/)DY, which is important for the specificity of the SET domain peptide in establishing mono-, di-, or trimethylation of a single Lys residue. The involvement of invariant residues in binding to AdoMet and the target lysine, the catalytic site, the structural pseudoknot, an intra-molecular interacting salt bridge, and an F/Y switch, are indicated. The black text on a gray background denotes conserved residues. Species abbreviations: *At*, *Arabidopsis thaliana*; *Dm*, *Drosophila melanogaster*; *Hs*, *Homo sapiens*; *Nc*, *Neurospora crassa*; *Pvu, Phaseolus vulgaris; Gm, Glycine max* (modified from Dillon et al, 2005).

type) (Fehlberg et al., 2005), as well as the OSE nodulin consensus sequences (NODCON2GM) required for efficient nodule-specific expression (Stougaard et al., 1990; Fig 5). This might suggest that in *P. vulgaris* the *PvuTRX1h* gene is also involved in nodule and root development, additionally to all other unknown functions that this gene might have, whereas its ortholog in Arabidopsis (*ATX1*) is involved in root and lateral root development (Alvarez-Venegas et al., 2003; Alvarez-Venegas et al., 2006; Saleh et al., 2008b). Considering that there is no nodule formation in Arabidopsis and also that the *ATX1* promoter has ROOTMOTIFTAPOX1 *cis*-elements (Elmayan and Tepfer, 1995; Wally et al., 2008), a response element not present in the *PvuTRX1h* promoter, we hypothesize that *PvuTRX1h* is involved in nitrogen-fixing nodule development. Furthermore, as has been proposed, the acquisition of novel unrecognized motifs or changes in the modular combinations of TF binding boxes in the promoters (such as the organ-specific elements in the *ATX1* and *PvuTRX1h* promoters) may guide the specific expression pattern of the orthologs and allow the possible divergence of their functions (Wendel, 2000; Hughes, 2002; Duarte et al., 2006), something that deserves further analysis in the future.

It would be interesting to generate and analyze, in the future, mutagenized lines for both genes, either in the promoter or in the coding regions. This could be done, for example, by using a non-transgenic technique called Targeting Induced Local Lesions IN Genomes (TILLING); as well as by EcoTILLING, a rapid method of discovery and analysis of single-nucleotide polymorphism (SNP) in natural populations. These techniques are independent of genome size, ploidy level and reproductive system of plants that can be applied to economically important crops such as *P. vulgaris* (Rashid et al., 2011).

Expression patterns of PvuTRX1h and PvuASH1h

In order to gain further insights into the function of the *P. vulgaris PvuTRX1h* and *PvuASH1h*, semi-quantitative RT-PCR was used to determine the expression pattern of these two genes in roots, nodules, leaves and flowers (Fig 6). Using specific primers, RT-PCR analysis revealed to some extent different expression patterns. However, both genes were expressed ubiquitously in all tested tissues of the two *P. vulgaris* lines (DOR364 and BAT477 wild type lines). Lowest expression levels were obtained with mature leaves at 12 dpi (days post-inoculation with *Rhizobium etli*), while the patterns of gene expression in mature flowers were very similar for both lines. In contrast, the root expression levels for *PvuTRX1h* appeared higher right after inoculation with *R. etli* (day 0 post-inoculation, D0), in particular at the elongation zone (EZ), for both DOR364 and BAT477 wild type lines. On the other hand, the *PvuASH1h* gene showed a tendency to increase its expression at 6dpi in both lines. Against this background, the stronger expression of *PvuTRX1h*, as compared to *PvuASH1h*, in the early stages of nodule development (day 0 to day 3 post-inoculation, D0 to D3), suggested that *PvuTRX1h* might be more involved in nodule formation than *PvuASH1h* is.

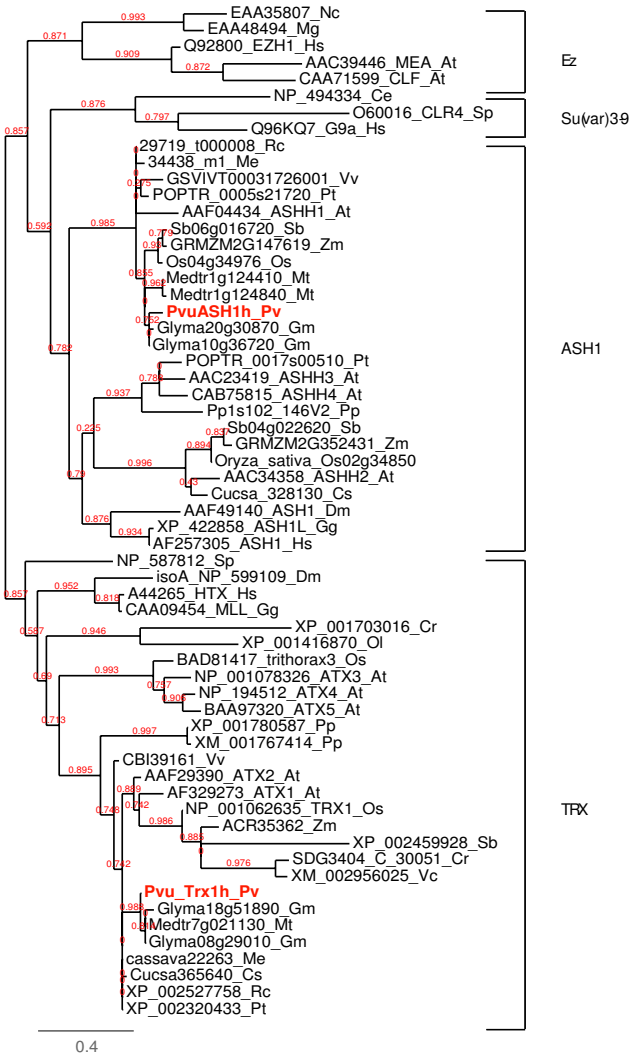

Fig 4. Maximum Likelihood phylogeny of SET-domain containing proteins. The SET domain of the Arabidopsis ATX1 and ASHH1 proteins was used to perform database searches using the BLAST program. SET-domain sequences with the highest scores were used to generate the Maximum Likelihood phylogenetic tree. The *Phaseolus vulgaris* PvuTRX1h and PvuASH1h sequences are in red. Branch length is proportional to the number of substitutions per site. Numbers in red correspond to the branch support values. GenBank accession numbers are followed by species abbreviations: *A. thaliana (At), N. crassa (Nc), M. grisea (Mg), C. elegans (Ce), S. pombe (Sp), H. sapiens (Hs), R. communis (Rc), M. esculenta (Me), V. vinifera (Vv), P. trichocarpa (Pt), S. bicolor (Sb), Z.mays (Zm), O. sativa (Os), M. truncatula (Mt), G. max (Gm), P. patens (Pp), C. sativus (Cs), D. melanogaster (Dm), G. gallus (Gg), C. reihardtii (Cr), O. lucimarinus (Ol), V. carteri (Vc), M. truncatula (Mt), Pv (P. vulgaris).*

Quantitative RT-PCR analysis from roots and nodules

To confirm the semi-quantitative PCR results, and with the aim of determining the possible function of *PvuTRX1h* and *PvuASH1h*, real-time RT-PCR (qRT-PCR) was used to quantitatively determine the expression pattern of these two genes during root and nodule development (root ramification zone and nodule; Fig 7). The expression of a housekeeping gene, actin, was used as an internal control. We found that the expression of *PvuTRX1h* appeared to be at a higher level at day three (D3) post-inoculation (ramification zone), for both DOR364 and BAT477 wild type lines (Fig 7). Following this peak in expression, the *PvuTRX1h* transcript levels (relative quantity) tended to remain constant from D3 to D12 post-inoculation at the ramification zone, and from D12 to D21 in the nodule. In contrast, expression of the *PvuASH1h* gene tended to be almost entirely absent from D0 and gradually increased until it reached its peak in expression at D21 post-inoculation, in both wild type lines. These observations suggest different functional involvement for the two genes. While both genes were ubiquitously expressed, *PvuTRX1h* displayed a more specialized pattern, with the highest expression level apparently occurring during the early stages of nitrogen-fixing nodule development, whereas *PvuASH1h* transcript levels increased alongside the stages of most nitrogen-fixing activity of the nodules. Together, these data indicate that transcriptional down-regulation of *PvuTRX1h*, and up-regulation of *PvuASH1h*, could be directly involved in nodule development.

Discussion

Plant genes homologous to the Trithorax superfamily have not been characterized in *Phaseolus vulgaris*, and our first goal was to establish the structure of two putative trithorax-Group homologs, the *PvuTRX1h* and *PvuASH1h* genes. RACE and PCR analysis were carried out to clone and sequence the full length cDNAs of both genes. We found that both predicted proteins have almost the same number of amino acids as their Arabidopsis homologs (Fig 1). After further analyzing the molecular structure of *PvuTRX1h* and *PvuASH1h* we found that, in addition to the SET domain, both genes have all the same architectural elements as the ATX1 and ASHH1 proteins; that is, PvuTRX1h has the PWWP-, DAST-, PHD- and SET- domains in the same order as in ATX1, whereas the PvuASH1h protein has the characteristic AWS- (associated with SET) and SET-domains (Fig 1). Furthermore, we have determined that the SET-domain of both Phaseolus proteins forms the classical β-fold configuration, surrounding a knot-like structure or 'pseudoknot' (Fig 2). In addition, the predicted PvuTRX1h and PvuASH1h proteins (as well as the Arabidopsis proteins) have similar sequence motifs surrounding the SET domain, making them putative histone lysine methyltransferases (Fig 3). Phylogenetic analysis also revealed that both proteins were grouped within the clades of their homologs (Fig 4). However, the clustering of PvuTRX1h and PvuASH1h into specific legume subclades suggests a potential divergent methylation activity in legumes from the activity reported in Arabidopsis for their homologs. For example, the presence of transcription factor binding motifs, or *cis*-acting elements, at the promoter region of *PvuTRX1h* suggested that this gene could be involved in nitrogen-fixing nodule development in *P. vulgaris*. *PvuTRX1h* therefore has a higher density of

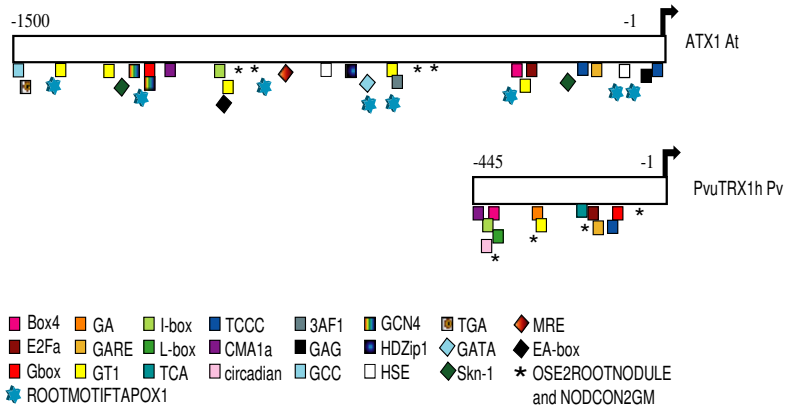

Fig 5. Transcription factor binding motifs recognized at the promoter regions of *ATX1* and *PvuTRX1*. Numbers indicate nucleotide coordinates up-stream from start codons. Transcription start codon is represented by an arrow. Names and colored symbols correspond to the DNA *cis*-acting elements present at the promoter regions.

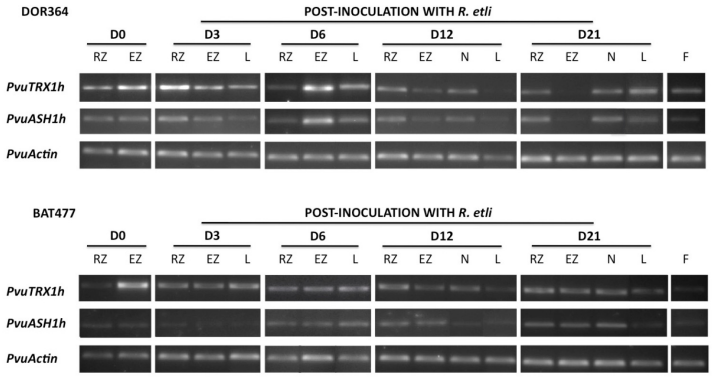

Fig 6. RT-PCR analysis. Expression patterns of the *P.vulgaris PvuTRX1h* and *PvuASH1h* genes, as determined by semi-quantitative RT-PCR in roots, nodules, leaves and flowers of two different common bean lines at 0, 3, 6, 12 and 21 days post-inoculation with *R. etli*. (a) Expression patterns in BAT477 WT line; (b) Expression patterns in DOR364 WT. (RZ = root ramification zone; EZ = elongation zone; L = leaves; F = mature flowers). *PvuActin* was used as control.

organ-specific elements (OSE) characteristic of the promoters activated in infected cells of root nodules (the OSE2ROOTNODULE type) (Fehlberg et al, 2005), as well as the OSE nodulin consensus sequences (NODCON2GM) required for efficient nodule-specific expression (Stougaard et al, 1990) (Fig 5). Thus, as a first step to distinguishing the function of the two *P. vulgaris* genes, we examined their expression patterns. RT-PCR analysis using specific primers to discriminate between the two genes revealed that both genes are ubiquitously expressed (Fig 6) and that different expression patterns are observed during nodule development after inoculation with *R. etli* (Fig 7). *PvuTRX1h* could therefore function as an activator of homeotic genes (involved in root and nodule development), similar to Trithorax in animal systems, whereas PvuASH1h could be involved in the final stages of nitrogen-fixing activity of the nodules. However, it is important to note that the finding of plant genes that are members of the trithorax family does not imply that they necessarily function within a complex of synergistically acting factors similar to the trx-G complex of Drosophila, and the presence of a SET domain belonging to the Trithorax superfamily (Jenuwein et al., 1998) does not automatically imply that the genes would have activating function. The SET domain itself may be found in both activators and repressors and is regarded as a dual-function motif, capable of assembling both types of complex.

Conclusions

The specific function of the *P. vulgaris* trithorax homologs has not yet been established. However, our results demonstrated that the two genes displayed different expression patterns in differentiated (roots and nodule) tissues. These observations suggest a different functional involvement for the two genes. While both genes were ubiquitously expressed, *PvuTRX1h* displayed a more specialized pattern with the highest expression level apparently occurring during early the stages of nitrogen-fixing nodule development, whereas *PvuASH1h* transcript levels increase concomitantly with the stages of most nitrogen-fixing activities of the nodules. In the future, it would be interesting in the future to generate and analyze mutagenized lines for both genes, as well as to determine the

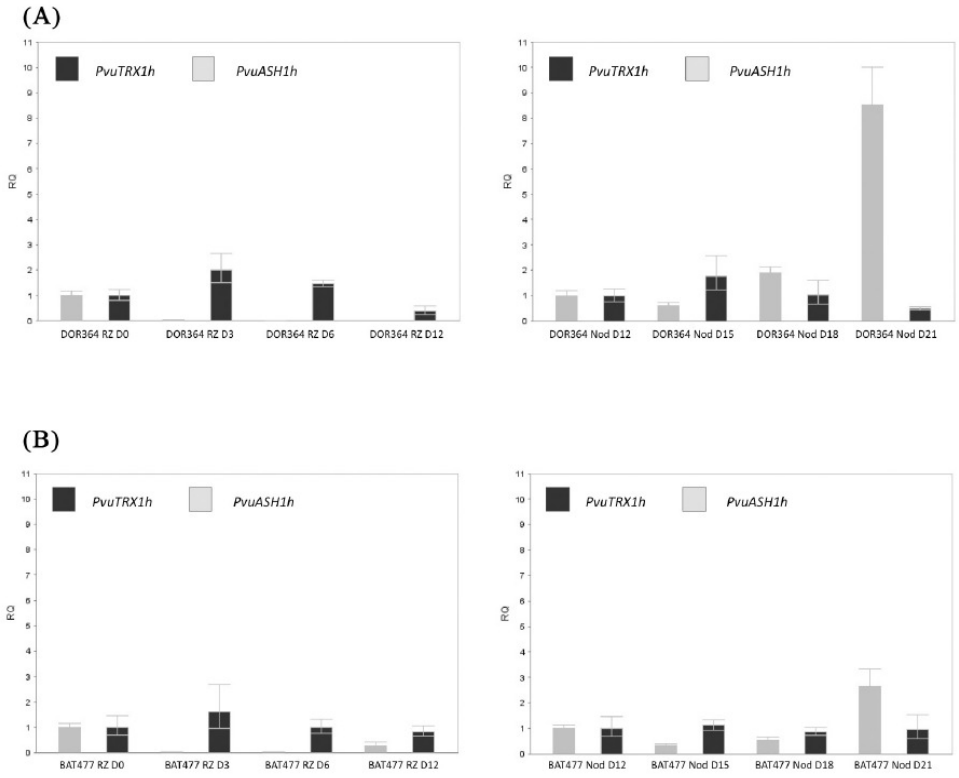

Fig 7. Expression patterns of the *P.vulgaris PvuTRX1h* and *PvuASH1h* genes as determined by real-time RT-PCR in roots and nodules of two different common bean lines at 0, 3, 6, 12 and 21 days post-inoculation with *Rhizobium etli*. Total RNA was reverse-transcribed and *PvuTRX1h* and *PvuASH1h* cDNAs were investigated by relative quantification analysis. The expression levels of the target genes were calculated using comparative Ct method. The lines inside the boxes denote the medians. (a) Expression patterns in *P.vulgaris* DOR364 wild type line; and (b) expression patterns in *P.vulgaris* BAT477 wild type line (RZ = root ramification zone; Nod = nodules). To normalize the quantity of transcripts in each sample, *PvuTRX1h* and *PvuASH1h* mRNA levels were normalized to the amount of actin cDNA.

type of histone methyl-transferase activity of these genes and their target genes.

Materials and methods

Plant materials

Wild type *Phaseolus vulgaris* cultivars BAT 477 and DOR 364 (Davis et al., 1988) were acquired from the germplasm bank at the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, INIFAP-Celaya, México.

Seed germination

P. vulgaris seeds from wild type (BAT 477 and DOR 364) lines were surface sterilized with 2% sodium hypochlorite for five minutes, rinsed five times with sterile H_20, placed in sterile plates containing wet towel paper, covered with tin foil and placed for three days at 28^oC in a Percyval growth chamber before Rhizobium inoculation.

Nodule formation

Once germinated, the seedlings were transferred to pots containing vermiculite and placed in a greenhouse. At the same time, *Rhizobium etli* strain CE3 was grown at 30^oC for 48 to 72 h on PY media (5 g/l of peptone, 3 g/l of yeast extract, 0.7 g/l of calcium chloride) containing 20 mg/ml nalidixic acid and 100 mg/ml streptomycin, until the bacterial culture reaches an OD_{600} of 0.5 to 0.6 (Cárdenas et al, 2006). Then, the roots from the different plants were inoculated with 1 ml of the Rhizobium suspension and watered every other day with a B&D solution without nitrogen (Broughton and Dilworth, 1971). Roots, and roots plus nodules, were collected at 0, 3, 6, 12 and 21 days post-inoculation, while nodules were collected at 12, 15, 18 and 21 days post-inoculation. The no-root hair zone (or root elongation zone) and the developing and mature root hair zones (root ramification zone) were separated (according to Bhuvaneswari et al., 1981), frozen in liquid nitrogen and stored at -80^oC.

Protein sequence analysis and structure prediction

cDNA sequences for *ATX1* and *ASHH1* were initially recognized from Arabidopsis and used as queries in various sources. Database searches were performed using the BLAST programs (Altschul et al., 1990 and 1997) on the KEGG (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/) and the Phytozome (http://www.phytozome.net/) databases. Soybean (*Glycine max*) sequences homologous to the Arabidopsis *ATX1*

sequence (Phytozome accession numbers Glyma08g29010.1 and Glyma18g51890.1) and to the ASHH1 (Glyma20g30870.1 and Glyma10g36720.1) were identified and analyzed. Each of these sequences was translated and examined for the presence of the different domains by searching the Conserved Domain Database (CDD) available from NCBI (Marchler-Bauer et al., 2005) as well as the Simple Modular Architecture Research Tool (SMART) database (Schultz et al., 1998; Letunic et al., 2004). Multiple sequence alignments were performed with the CLUSTALX (version 2.0; Thompson et al., 1997).

On the other hand, we used the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) for protein structure prediction of the Arabidopsis (ATX1 and ASHH1) and *P. vulgaris* sequences (PvuTRX1h and PvuASH1h) (Zhang, 2008; Roy et al., 2010). All models were visualized using PyMOL (version 1.3).

RACE analysis and cloning of the PvuTRX1h gene

To generate full-length cDNA we made use of the 5'- and 3'- Rapid Amplification of cDNA Ends method (SMART RACE cDNA kit, Clontech Laboratories USA). Total RNA was extracted from three-days old BAT 477 wild type roots using the Trizol reagent (Invitrogen Life Technologies USA). In order to clone the gene homolog to the TRX gene (*PvuTRX1h*), we used the *ATX1* sequence to generate a multiple sequence alignment with the two *G. max* sequences obtained from the BLAST searches (genes *Glyma08g29010.1* and *Glyma18g51890.1*). With this alignment, we designed a set of primers with homology to both *G. max* genes (forward 5'-GATGGTCCATGCTAGATG-3'; reverse 5'-CTTCGTG-GTGCATATAGGGTA-3') and used them to amplify a 1300 bp cDNA fragment from *P. vulgaris* nodules. After sequencing this fragment, an antisense primer for 5'-RACE was designed (5'-GATGCAGCTGCAAGTGCTTC-3') and a fragment of ~2500 bp was obtained. At the same time, with the sequence from the same 1300 bp fragment that we initially obtained, a 3'-RACE primer was designed (5'-GCGAAGAGGGACATTAAACAGTA-3'), and a 604 bp amplification product was obtained. After sequencing the 5'- and 3'- RACE products, we designed a set of primers to amplify the whole gene sequence (forward 5'-ATGGCAGCGGCGTTCCAG-3'; reverse 5'-TCATCCTC-CTCTCCAATCTATTAATTCCC-3'), a 3270 bp cDNA. This PCR product was cloned into the Topo 2.1 (Invitrogen) and sequenced with an ABI 3730xl machine. The sequence has been deposited at GenBank (BankIt1430436: *PvuTRX1h* JF262910).

Cloning of the PvuASH1h gene homolog

We used the Arabidopsis *ASHH1* sequence to search the *Phaseolus vulgaris* (common bean) (EST) KEGG Genes Database (http://www.genome.jp/dbget-bin/www_bget?epvu: 13627) and identified a highly homologous (81% homology) EST sequence (entry number *13627*), that contained only the 5'-UTR and 5'- gene sequence. Then, in order to identify the 3'-end sequence and to clone the full length cDNA, this EST was aligned with two homologous soybean sequences, *Glyma20g30870.1* and *Glyma10g36720.1* (Phytozome: www.phytozome.net), and a set of primers was designed. The forward primer (5'-ATGGATCCTTTTACTGAAGAGTT-GCC-3') was complementary to the *P. vulgaris* EST, while the reverse primer (5'-TCAACCTGTCAGGAGCTTAATTT-CG-3') was complementary to the two soybean sequences. The PCR reaction amplified a 1446 bp fragment that was cloned into the Topo 2.1 (Invitrogen) and sequenced (ABI 3730xl machine). The sequence has been deposited at GenBank (BankIt1430436: *PvuASH1h* JF262911).

RT-PCR reactions from Phaseolus vulgaris plant tissues

Total RNA was isolated from *P. vulgaris* (wild type DOR364 and BAT477 lines) leaves, roots, nodules and flowers by using the RNAeasy Plant Isolation kit from Qiagen (Germantown, MD, USA) according to the manufacturer's instructions. Samples of total RNA (1 µg) were reverse transcribed into cDNA in 50 µl reactions with 50 U/µl of reverse transcriptase and oligo(dT) primer using the MMLV-RT from Fermentas (Burlington, Ontario CA). An aliquot of cDNA synthesis (2 µl) was used in RT-PCR reactions.

We determined gene expression in leaves, flowers, and roots. For root tissue, we determined the expression patterns in three different zones: the no-root hair zone (or root elongation zone), the developing and mature root hair zones (root ramification zone), and the nodules (Bhuvaneswari et al., 1981),

The RT-PCR conditions were as follows: 95 ^{o}C for 5 min followed by 25 cycles of 95 ^{o}C for 1 min, 60 ^{o}C for 30 s, and 72 ^{o}C for 1 min. Primers are shown 5' to 3' as follows: PvuTRX1hF: GATGGTCCATGCTAGATG; PvuTRX1hR: GATGCAGCTGCAAGTGCTTC; PvuASH1hF: CTGTAC-TTCGCTGGCACTG; PvuASH1hR: CTCCACAGTATTCG-ATGAC; and actin as control (KEGG accession number 3899) PvuActinF: CGCATCTCTTCTAGGCTACAGGACC; PvuActinR: CCAGGCGTTCAACTATACCAGGG. Samples were analyzed in 1% agarose gels.

Quantitative RT-PCR reactions from roots and nodules

Total RNA was extracted using the PureLink kit according to the manufacturer's instructions (Invitrogen) and treated with DNase I. Roots (from the root ramification zone) were collected at 0, 3, 6, 12 and 21 days post-inoculation, and nodules were collected at 12, 15, 18 and 21 days post-inoculation with *R. etli*. All experiments were performed using three biological replicates and two technical replicates. Two micrograms of total RNA from each sample were used for the first-strand cDNA synthesis in a 20 µl reverse transcription reaction. Reverse transcription was performed according to the established protocol (MMLV-RT, Fermentas). The first-strand cDNA was used for SYBR Green monitored qRT-PCR (Applied Biosystems, FastSYBR®Green Master Mix) in an Applied Biosystems StepOnePlus™ Real-Time PCR System. The qRT-PCR analysis was performed with Step One software (Applied Biosystems). The primer pairs were used for qRT-PCR analysis of *PvuTRX1h* (5'- CGGCTTTCTTGTTACTG-TGGC-3' and 5'- CGAAACAATGGGAAGAATCAG-3'); *PvuASH1h* (5'- CTGTACTTCGCTGGCACTG-3' and 5'-CCTCCTTCTGCTTCTTCTGC-3'). Amplification of the *PvuActin* (KEGG accession #epvu3899) gene was used for reaction normalization with the primers 5'- CGCATCTC-TTCTAGGCTACAGGACC-3' and 5'- CCAGGCGTTCA-ACTATACCAGGG-3'). Amplification was performed under the following conditions: 95 ^{o}C for 10 min, followed by 40 cycles of 95°C/15sec-62°C/1min; and 72 ^{o}C, 40 s, and a melting curve analysis of 95°C/15sec-60°C/1min (+0.3°C)-95°C/15sec. Gene expression levels were calibrated to the average value of all tissues to obtain a calibrated delta-Ct value for each gene.

Phylogenetic analysis of the Trithorax homologs from P. vulgaris

In order to determine the phylogeny of the *P. vulgaris PvuTRX1h* and *PvuASH1h* sequences, the Arabidopsis ATX1 and ASHH1 SET-domains were used to perform database searches using BLASTP and TBLASTN at the NCBI network service (http://www.ncbi.nlm.nih.gov/blast) and at the Phytozome databank (http://www.phytozome.net). Only SET-domain sequences with the highest scores were used to generate the Maximum Likelihood phylogenetic trees. The phylogenetic analyses were performed by using the Phylogeny Pipeline at http://phylogeny.lirmm.fr/ (Dereeper et al., 2008); that is, multiple sequence alignment of proteins was performed with the MUSCLE software (Edgar, 2004), curation of the alignment was performed with the GBLOCKS program (Castresana, 2000), Maximum Likelihood trees with approximate Likelihood Ratio Test for branches (PhyML+aLRT) was performed with the PhyML3.0 program (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006), and trees were drawn with TREEDYN software (Chevenet et al., 2006). The accession numbers for the different proteins corresponded to those reported at the NCBI and Phytozome databanks.

Identification of cis-elements at the promoter region of PvuTRX1h

Information available from the Ohio State University site (http:// arabidopsis.med.ohio-state.edu/AtcisDB) was used to obtain the *ATX1* promoter region.

For the *PvuTRX1h* promoter region, we used its cDNA sequence to perform a BLASTN search against the *P. vulgaris* draft genome sequence developed at the National Laboratory of Genomics for Biodiversity (Langebio, México; Common Bean Genome Sequencing Project). We found a contig (Contig51613.1) containing 445nt upstream of the start codon of *PvuTRX1h* and corresponding to its partial promoter region.

For analyses of transcription factors (TF) binding motifs, once the *PvuTRX1h* and *ATX1* promoter regions were obtained, we used the PLACE database (a Database of Plant Cis-acting Regulatory DNA Elements, http://www.dna.affrc. go.jp/PLACE/signalscan.html; Higo et al, 1999 and Prestridge, 1991) and the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; Lescot et al, 2002), to identify *cis*-regulatory sequences. No contigs containing the promoter region for the *PvuASHH1h* gene have so far been found in the Common Bean Genome Sequencing Project.

Acknowledgements

We would like to thank Dr. Jorge Acosta-Gallegos at INIFAP-Celaya, México, for kindly providing seeds from the *P. vulgaris* cultivars BAT 477 and DOR 364. We are most grateful to Dr. Luis Cárdenas Torres at the Instituto de Biotecnología (UNAM, México) for providing us with the *Rhizobium etli* strain CE3. We also thank Dr. Alfredo Herrera-Estrella, at the National Laboratory of Genomics for Biodiversity (Langebio, México), for allowing us to access the *P. vulgaris* draft genome sequence. This work was supported by the "Consejo Nacional de Ciencia y Tecnología", grant CB-2011/167693 to RAV.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

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

Alvarez-Venegas R (2010) Regulation by Polycomb and Trithorax Group Proteins in Arabidopsis. In: Last R (ed) The Arabidopsis Book. Rockville, MD, USA

Alvarez-Venegas R, Avramova Z (2002) SET-domain proteins of the Su(var), E(z) and Trithorax families. Gene 285: 25-37

Alvarez-Venegas R, Pien S, Sadder M, Witmer X, Grossniklaus U, Avramova Z (2003) ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. Curr Biol 13: 627-637

Alvarez-Venegas R, Sadder M, Hlavacka A, Baluska F, Xia Y, Lu G, Firsov A, Sarath G, Moriyama H, Dubrovsky JG, Avramova Z (2006) The Arabidopsis homolog of trithorax, ATX1, binds phosphatidylinositol 5-phosphate, and the two regulate a common set of target genes. P Natl Acad Sci USA 103(15): 6049-6054

Anisimova M, Gascuel O (2006) Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. Systems Biol 55(4): 539-52

Broughton WJ, Dilworth MJ (1971) Control of Leghaemoglobin Synthesis in Snake Beans. Biochem J 125: 1075-1080

Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y (1998) Predicting function: from genes to genomes and back. J Mol Biol 283: 707-725

Bhuvaneswari TV, Bhagwat AA, Bauer WD (1981) Transient susceptibility of root cells in four common legumes to nodulation by Rhizobia. Plant Physiol 68: 1144-1149

Cárdenas L, Alemán E, Nava N, Santana O, Sánchez F, Quinto C (2006) Early responses to Nod factors and mycorrhizal colonization in a non-nodulating Phaseolus vulgaris mutant. Planta 223(4): 746-754

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4): 540-52

Cazzonelli CI, Cuttriss AJ, Cossetto CB, Pye W, Crisp P, Whelan J, Finnegan EJ, Turnbull C, Pogsona BJ (2009) Regulation of carotenoid composition and shoot branching in Arabidopsis by a chromatin modifying histone methyltransferase, SDG8. Plant Cell 21(1): 39-53

Cheng X, Collins RE, Zhang X (2005) Structural and sequence motifs of protein (histone) methylation enzymes. Annu Rev Bioph Biom 34: 267-294

Chevenet F, Brun C, Banuls AL, Jacq B, Chisten R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7: 439

Collins RE, Tachibana M, Tamaru H, Smith KM, Jia D, Zhang X, Selker EU, Shinkai Y, Cheng X (2005) In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem 280: 5563–5570

Davis J, Guiller K, Kipe-Nolt J, Awah M (1988) Non-nodulating mutants in common bean. Crop Sci 28: 859–860

Dereeper A, GuignonV, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie, JM, Gascuel, O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36: W465-9

Dillon SC, Zhang X, Trievel RC, Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol 6: 227

Dingwall AK, Beek SJ, McCallum CM, Tamkun JW, Kalpana GV, Goff SP, Scott MP (1995) The Drosophila snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex. Mol Biol Cell 6(7): 777-791

Dong G, Ma DP, Li J (2008) The histone methyltransferase SDG8 regulates shoot branching in Arabidopsis. Biochem Bioph Res Co 373: 659–664

Dorn R, Krauss V, Reuter G, Saumweber H (1993) The enhancer of position-effect variegation of Drosophila, E(var)3-93D, codes for a chromatin protein containing a conserved domain common to several transcriptional regulators. P Natl Acad Sci USA 90(23):11376-11380

Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW (2006) Expression pattern shift following duplication indicative of subfractionalization and neofractionalization in regulatory genes in Arabidopsis. Mol Biol Evol 23: 469–478

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5): 1792-1797

Elmayan T, Tepfer M (1995) Evaluation in tobacco of the organ specificity and strength of the rolD promoter, domain A of the 35S promoter and the 35S2 promoter. Transgenic Res 4(6): 388-396

Fehlberg V, Vieweg MF, Dohmann EM, Hohnjec N, Puhler A, Perlick AM, Kuster H (2005) The promoter of the leghaemoglobin gene VfLb29: functional analysis and identification of modules necessary for its activation in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots. J Exp Bot 56: 799-806

Franke A, DeCamillis M, Zink D, Cheng N, Brock HW, Paro R (1992) Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. EMBO J 11(8): 2941-2950

Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, et al (1997) A polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature 386: 44–51

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 280: 446–450

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5): 696-704

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27(1): 297-300

Huang N, vom Baur E, Garnier JM, Lerouge T, Vonesch JL, Lutz Y, Chambon P, Losson R (1998) Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. EMBO J 17: 3398–3412

Hughes AL (2002) Adaptive evolution after gene duplication. Trends Genet 18: 433–434

Jackson JP, Johnson L, Jasencakova Z, Zhang X, Perez-Burgos L, Singh PB, Cheng X, Schubert I, Jenuwein T, Jacobsen SE (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 112(6): 308-15

Jenuwein T, Laible G, Dorn R, Reuter G (1998) SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell Mol Life Sci 54(1): 80-93

Kennison JA (1995) The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annu Rev Genet 29:289-303

Kiefer JC (2007) Epigenetics in development. Dev Dynam 236: 1144–1156

Kim SY, He Y, Jacob Y, Noh YS, Michaels M, Amasino R (2005) Establishment of the vernalization-responsive, winter-annual habit in Arabidopsis requires a putative histone H3 Methyl Transferase. Plant Cell 17: 3301–3310

Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W (2003) Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO J 22(18): 4804-4814

Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693–705

Kwon T, Chang JH, Kwak E, Lee CW, Joachimiak A, Kim JC, Lee JW, Cho J (2003) Chromatin and transcription mechanism of histone lysine methyl transfer revealed by the structure of SET7/9–AdoMet. EMBO J 22: 292-303

Lescot M, Dhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouz P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30(1): 325-327

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P (2004) SMART 4.0: towards genomic data integration. Nucleic Acids Res 32: D142-4

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH (2005) CDD: a Conserved Domain Database for protein classification. Nucleic Acids Res 33: D192-6

Nakagawa T, Takane K, Sugimoto T, Izui K, Kouchi H, Hata S (2003) Regulatory regions and nuclear factors involved in nodule enhanced expression of a soybean phosphoenolpyruvate carboxylase gene: implications for molecular evolution. Mol Genet Genomics 269: 163–172

Papoulas O, Beek SJ, Moseley SL, McCallum CM, Sarte M, Shearn A, Tamkun JW (1998) The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. Development 125(20): 3955-66

Pien S, Fleury D, Mylne J, Crevillen P, Inze D, Avramova Z, Dean C, Grossniklaus U (2008) Unraveling trithorax functions in plants: ATX1 dynamically regulates the activation of FLC via histone H3 lysine 4 trimethylation. Plant Cell 20: 580-588

Prestridge DS (1991) SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. CABIOS 7: 203-206

Rashid M, He G, Guanxiao Y, Khurram Z (2011) Relevance of tilling in plant genomics. Aust J Crop Sci 5(4): 411-420

Rodriguez-Llorente ID, Perez-Hormaeche J, Dary M, Caviedes MA, Trinh TH, Kondorosi A, Ratet P, Palomares AJ (2003) MsPG3 polygalacturonase promoter elements necessary for expression during Sinorhizobium meliloti–Medicago truncatula interaction. Plant Soil 257: 19–26

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725-738

Saleh A, Al-Abdallat A, Ndamukong I, Alvarez-Venegas R, Avramova Z (2007) The Arabidopsis homologs of trithorax (ATX1) and enhancer of zeste (CLF) establish 'bivalent chromatin marks' at the silent AGAMOUS locus. Nucleic Acids Res 35(18): 6290-6296

Saleh A, Alvarez-Venegas R, Avramova Z (2008a) Dynamic and stable histone H3 methylation patterns at the Arabidopsis FLC and AP1 loci. Gene 423: 43-47

Saleh A, Alvarez-Venegas R, Yilmaz M, Oahn-Le, Hou G, Sadder M, Al-Abdallat A, Xia Y, Lu G, Ladunga I, Avramova Z (2008b) The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions. Plant Cell 20(3): 568-579

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G (2007) Genome regulation by Polycomb and Trithorax proteins. Cell 128: 735–745

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. P Natl Acad Sci USA 95(11): 5857-5864

Sedkov Y, Benes JJ, Berger JR, Riker KM, Tillib S, Jones RS, Mazo A (1999) Molecular genetic analysis of the Drosophila trithorax-related gene which encodes a novel SET domain protein. Mech Dev 82(1-2):171-179

Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell 98(1): 37-46

Simons JA, Tamkun JW (2002) Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. Curr Opin Genet Dev 12: 210–218

Soppe WJ, Bentsink L, Koornneef M (1999) The early-flowering mutant efs is involved in the autonomous promotion pathway of Arabidopsis thaliana. Development 126: 4763-4770

Springer NM, Napoli CA, Selinger DA, Pandey R, Cone KC, Chandler VL, Kaeppler HF, Kaeppler SM (2003) Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots. Plant Physiol 132: 907-925

Stougaard J, Jorgensen JE, Christensen T, Kühle A, Marcker KA (1990) Interdependence and nodule specifity of cis-acting regulatory elements in the soybean leghaemoglobin lbc3 and N23 gene promotoer. Mol Genet Genomics 220: 353–360

Surani MA, Hayashi K, Hajkova P (2007) Genetic and Epigenetic Regulators of Pluripotency. Cell 128: 747–762

Thakur JK, Malik MR, Bhatt V, Reddy MK, Sopory SK, Tyagi AK, Khurana JP (2003) A POLYCOMB group gene of rice (Oryza sativa L. subspecies indica), OsiEZ1, codes for a nuclear localized protein expressed preferentially in young seedlings and during reproductive development. Gene 314: 1–13

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882

Thorstensen T, Fischer A, Sandvik SV, Johnsen SS, Grini PE, Peuter G, Aalen RB (2006) The Arabidopsis SUVR4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. Nucleic Acids Res 34: 5461–5470

Wally O, Jayaraj J, Punja ZK (2008) Comparative expression of beta-glucuronidase with five different promoters in transgenic carrot (*Daucus carota* L.) root and leaf tissues. Plant Cell Rep 27(2): 279-87

Wendel J (2000) Genome evolution in polyploids. Plant Mol Biol 42: 225–249

Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, Howell S, Taylor IA, Blackburn GM, Gamblin SJ (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature 421(6923): 652-656

Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou J-P, Steinmetz A, Shen W-H (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. Mol Cell Biol 28(4): 1348–1360

Zhang X, Yang Z, Khan SI, Horton JR, Tamaru H, Selker EU, Cheng X (2003) Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell 12: 177–185

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40

Zhao Z, Yu Y, Meyer D, Wu C, Shen WH (2005) Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. Nat Cell Biol 7: 1156–1160