Molecular analysis of a sunflower gene encoding an homologous of the B subunit of a CAAT binding factor

Mariangela Salvini · Emanuela Sani ·
Marco Fambrini · Laura Pistelli · Chiara Pucciariello ·
Claudio Pugliesi

Abstract A genomic DNA fragment containing the complete *LEAFY COTYLEDON1-LIKE* (*HaL1L*) gene was retrieved by chromosome walking. Its sequence was confirmed and elongated by screening a sunflower genomic DNA BAC Library. *HaL1L*, whose cDNA had already been sequenced and characterized, encodes a NF-YB subunit of a CCAAT box-binding factor (NF-Y) involved in the early stages of zygotic and somatic embryogenesis in the *Helianthus* genus. In the *HaL1L* 5′-flanking region, elements specific to a putative TATA-box promoter and two "CG isles" were identified. An investigation of the methylation status of these CG rich DNA regions showed that differentially methylated cytosines were recognizable in the DNA of embryos on the fifth day after pollination in comparison to leaf DNA suggesting that during plant development epigenetic regulation of *HaL1L* transcription was achieved by methylating cytosine residues. We also searched the *HaL1L* nucleotide sequence for *cis*-regulatory elements able to interact with other transcription factors (TFs) involved in the *HaL1L* regulation. Of the elements identified, one of the most intriguing is WUSATA, the target sequence for the WUSCHEL (WUS) TF, which may be part of a complex regulation network controlling embryo development. In this article, we show that the WUSATA target site, located in the intron of *HaL1L*, is able to bind the TF WUS. Interestingly, we found auxin and abscisic acid responsive motifs in the *HaL1L* promoter region suggesting that this gene may additionally by under hormonal control. Finally, the presence of a cytoplasmic polyadenylation signal downstream to the coding region indicates that this gene may also be controlled at the translation level by a temporarily making the pre-synthesized *HaL1L* mRNA unavailable for protein synthesis.

M. Salvini (✉)
Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa,
Italy
e-mail: m.salvini@sns.it

M. Salvini · M. Fambrini · C. Pugliesi
Dipartimento di Biologia delle Piante Agrarie, Sezione di
Genetica, Università di Pisa, Via Matteotti 1B, 56124 Pisa, Italy

E. Sani
Faculty of Biomedical and Life Sciences, University of Glasgow,
Glasgow G12 8QQ, UK

L. Pistelli
Dipartimento di Biologia, Unità di Fisiologia Vegetale,
Università di Pisa, Via Mariscoglio 34, 56124 Pisa, Italy

C. Pucciariello
PlantLab, Institute of Life Sciences, Scuola Superiore
Sant'Anna, Piazza Martiri della Libertà 33, 56127 Pisa, Italy

Keywords CCAAT binding factor · Gene expression ·
Methylation · Promoter region · Transcription factors ·
EMSA · *Helianthus annuus* L

Introduction

Developmental processes are regulated in animals, like in plants, by complex gene expression programs in which dynamic regulation of chromatin folding represents the first level of control [1]. The first level of folding is the nucleosome, the functional unit of chromatin and it is formed when DNA interacts with a histone octamer [2]. Locally, the degree of chromatin compaction is controlled by chromatin-remodelling factors (e.g. CHD) which may

🖄 Springer

act by catalyzing the dissociation of core histones, facilitating nucleosome sliding or relocation of entire histone octamers. All these phenomena are intrinsically linked to the post-translational modifications that histone's N termini may potentially undergo [3]. Chromatin folding modification is of fundamental importance to allow access and proper recruitment of specific regulatory factors to their target DNA sequences, which may be dispersed anywhere with the respect to a gene [4]. In particular, promoters and enhancers of RNA polymerase II-transcribed genes are formed by a puzzle-like combination of short sequences recognized by sequence-specific regulators. Among such elements, the CCAAT box is known to be one of the most frequent, and it contributes significantly to the activity of the promoter. The protein complex that mainly interacts with this element is NF-Y, an ubiquitous heteromeric transcription factor (TF) made up of three subunits, NF-YA, NF-YB and NF-YC, all necessary for DNA binding. NF-YA and NF-YB, which share a conserved histone fold motif, join together independently of the presence of subunit C in a way that resembles the H2A-H2B histone dimer structure [5].

In plants, all three NF-Y subunit are encoded by multigenic families that complicates the task of assigning roles to single members. Studies carried out in *Arabidopsis thaliana* showed that in plants the HAP3 proteins, namely NF-YB subunits, include at least two distinct types: a LEAFY COTYLEDON1 (LEC1)-type and a non-LEC1-type which differ by 16 amino acid residues that serve as a signature of their B-domain [6–8]. The two classes differ also in expression pattern and function [8].

The LEC1-type comprises two separated peptides encoded by two different genes: *LEC1*, and a closely related *LEC1-LIKE* (*L1L*) (8). *LEC1* belongs to the class of *LEC* genes that as genetic analysis has shown play key roles in several aspects of embryo and seed development [9]. Early during embryogenesis, the *LEC* genes [*LEC1*, *LEC2* and *FUSCA3* (*FUS3*)] are required to maintain embryonic cell fate and to determine cotyledon identity [10]. These genes are also active later during embryogenesis when, together with *ABSCISIC ACID INSENSITIVE3* (*ABI3*), they are involved in the initiation and maintenance of the maturation phase [9]. Members of the *VP1/ABI3-LIKE* (*VAL*) family of B3 domain TFs function as global repressors of the LEC1/B3 TF network in germinating seedlings [11]. In addition, an upstream gene *PICKLE* (*PKL*) acts in concert with gibberellins (GAs) to repress embryonic traits during and after germination including the *LEC* gene expression [12–14]. *PKL* encodes an SWI/SNF-class chromatin-remodelling factor that belongs to the CHD3 group [12]. *pkl* mutants form swollen primary roots [15], also known as pickle roots, and they show ectopic expression of *LEC1*, *LEC2*, and *FUS3* in their primary

roots [13]. There is evidence that in germinating roots repression of *LEC1* is mediated by *PKL* activation of *Polycomb* group (*PcG*) genes that encode subunits of Polycomb repressive complexes responsible for trimethylating histone H3 at lysine 27 (H3K27me3) [16]. *L1L* is closely related to *LEC1*, but to date no works has definitely proved whether or not they have distinct or redundant roles. A different function during embryogenesis seems likely if we consider their specific expression patterns. *L1L*, which has been shown to be necessary during early embryogenesis, is expressed earlier than *LEC1* during embryo development [8, 9].

In plants, the zygote is not the only cell that has the ability to develop into an embryo. Zygotic embryogenesis is dependent on fertilization, but apomictic embryos produced in certain species provide clear evidence of in vivo embryo development without sex [17]. In epiphyllous species, even somatic cells of vegetative organs can spontaneously form in vivo ectopic embryos [18]. So far, little is known about the early inductive events underlying the initiation of embryogenesis and the transition of a somatic cell to an embryogenic mode of development. *LEC1* and *LEC1-LIKE* are fundamental for the re-acquisition of embryogenic competence [10]. However, other TFs, such as AGAMOUS LIKE15 (AGL15) and WUSCHEL (WUS), have also been shown to be required for early as well late steps of zygotic embryogenesis and their over-expression might be capable of triggering embryogenic programs in vegetative cells [19, 20]. The broad range of regulatory factors involved in the acquisition of embryogenic competence suggests that numerous pathways, both independent and inter-related, are functional in this process. Whether and how these genes work together to promote embryogenic cell formation is not yet known [21]. It is likely that epigenetic changes can decide the cell fate and that they constitute a layer of control superimposed to the activity of the plethora of TFs, involved in these processes [1, 22]. For example, several reports have suggested that DNA methylation may have an important role in cellular de-differentiation, changing cell fate, and in plant totipotency [23, 24]. In carrot the total DNA methylation level is altered during somatic embryo development [23]. In addition, treatment with drugs, such as 5-azacytidine or 2-amino 5-ethoxy-carbonyl-pyrimidine which induce hypomethylation, has been shown to suppress somatic embryogenesis [23, 25]. During carrot embryogenesis, it was hypothesized *CARROT-LEC1* (*C-LEC1*) expression may be sensitive to DNA methylation [26]. A *C-LEC1* promoter region showed a reduced level of DNA methylation during somatic embryogenesis followed by an increase during the transition from embryonic to vegetative growth [26].

The *L1L* cDNA from *Helianthus annuus* (*HaL1L*) has been isolated and characterized [27]. Expression studies in

the genus *Helianthus* showed that *HaL1L* encodes an homolog (56 % amino acid sequence identity) to the *LEC1-LIKE* of *A. thaliana* (*AtL1L*) and that the *HaL1L* gene is involved in early stages of both zygotic and somatic embryogenesis [27, 28].

In this paper we described the characteristics of a genomic DNA large insert containing the entire sequence of the *HaL1L* gene, isolated by chromosome walking and elongated by screening of genomic BAC library of *H. annuus*. Also, we provide the first evidence that *HaL1L* transcription could be under epigenetic regulation through methylation of cytosine residues and by potential interactions with hormones and TFs. In particular, we demonstrate that WUS can bind a specific target site in *HaL1L* intron.

Methods

Plant material and growth conditions

Helianthus annuus plants (inbred line HOR, provided by the Genetics Section of the Department of Crop Plant Biology, University of Pisa, Pisa, Italy) were grown in a growth chamber in pots containing a mixture of vermiculite, peat and soil at 23 ± 1 °C, under a 16 h daily photoperiod (200 µmol m^{-2} s^{-1}). Mercury vapour lamps were used for lighting (Osram HQI-TS 250 W/NDN, Wembley, England).

Isolation of the *HaL1L* gene and its flanking 5′- and 3′-region

Genomic DNA was extracted from young leaf blades of 30-day-old plants with the Nucleon DNA extraction and purification protocol for plant tissue (GE Healthcare Europe, UK), according to the manufacturer's instructions. Sequence information from the previously characterized *HaL1L* cDNA was used to isolate by PCR the intron/exon region of *HaL1L* [27]. *HaL1L*-specific primers CHI-TCDNA (forward, 5′-CTAGAGAGAGACAATTCC-3′) (located 39 nucleotides upstream from the ATG) and CHI-SR (reverse, 5′-GTAGATGGAGAGTGTCAG-3′) (located 136 nucleotides downstream from the translational stop codon) were used. The PCR conditions were: 95 °C for 2 min, 35 amplification cycles (30 s at 94 °C, 30 s at 53 °C, 1 min 30 s at 72 °C) and a final extension of 7 min at 72 °C.

GenomeWalker DNA walking was used to find genomic DNA sequences adjacent to the 5′- and 3′-region of the *HaL1L* gene [29]. A preliminary analysis by DNA gel blot hybridization was conducted to identify the restriction enzymes suitable for 5′ and 3′ RAGE (rapid amplification of genomic end). A clone, *HaL1L-1* containing the *HaL1L* cDNA region located between the sequences corresponding to the CHI-TCDNA and CHI-SR primers, was used for

PCR amplification of two fragments containing the 5′-end (HaSON1) and 3′-end (HaSON3) of *HaL1L*. The HaSON1 fragment was amplified using the *HaL1L*-specific primers SON1F (forward, 5′-GAAACATATGGAACGTGGAG-3′) and SON1R (reverse, 5′-CAATGCACTCATTGTCTTC TG-3′) whereas the HaSON3 fragment was amplified using the *HaL1L*-specific primers SON3F (forward, 5′-GTGA GCGTGGGTCGATAAAGG-3′) and SON3R (reverse, 5′-AGTCTTTACACTGAGCATAC-3′). The resulting PCR products were electrophoretically purified and digoxigenin (DIG)-labelled with the DIG-DNA labelling Kit protocol (Roche Diagnostics, Mannheim, Germany) for use as probes. They were named *HaSON1*- and *HaSON3*-probe, respectively. Next, *H. annuus* genomic DNA was cut with restriction enzymes in single or double digestions using *Pvu*II, *Dra*I, *Eco*RV, *Stu*I, *Ssp*I and *Sty*I + *Pvu*II, according to the manufacturer's instructions (Fermentas Life Sciences-Thermo Fisher Scientific, Waltham, MA, USA). The resulting fragments were separated on a 1.2 % (w/v) agarose gel and blotted onto a positively charged nylon membrane (Roche). The membrane was then hybridized at 39 °C in DIG Easy Hyb Buffer (Roche) with the *HaSON1*- or *HaSON3*-probe according to the manufacturer's recommendations. The membrane was washed twice in 2× SSC, 0.1 % SDS at room temperature for 5 min, and washed twice in 0.1× SSC, 0.1 % (w/v) SDS for 10 min, at 65 °C.

For the 5′ RAGE, genomic DNA was digested with *Dra*I restriction enzyme and ligated to Genome Walker Adaptor, according to the manufacturer's instructions (Clontech-Takara Bio Inc., Otsu, Japan). The first PCR was conducted with the AP1 primer provided in the kit and the *HaL1L*-specific primer IN7 (reverse, 5′-ACGAAGACGCCTAG TGGTCAGCGATCAGCA-3′). The nested PCR was conducted with the nested adaptor primer (AP2) and the nested *HaL1L*-specific primer 1RAGECDS (reverse, 5′-GTGGA TGGGGAGCCTGTGGTAGCCATGGAA-3′). For the 3′ RAGE, genomic DNA was digested with *Ssp*I restriction enzyme and attached to Genome Walker Adaptor, according to the manufacturer's instructions (Clontech-Takara). The first PCR was conducted with the AP1 primer provided in the kit and the *HaL1L*-specific primer 3RAGE3 (forward, 5′-CTCTCTTGTTTTGTCCCCACGAATCTGT CA-3′). The nested PCR was conducted with the nested adaptor primers AP2 and 3RAGE2 (*HaL1L*-specific primer; forward, 5′-GGCCTGCTGGTTTCGAGCCGTATGC TCAGT-3′). The PCRs were conducted in conditions recommended by manufacturer (Clontech-Takara).

All PCR products were separated by electrophoresis on a 1.5 % TAE-agarose gel and visualized with ethidium bromide under UV light. Selected amplified products were purified using the Wizard® SV Gel and PCR Clean-UP System (Promega Corporation, Madison, WI, USA),

inserted into the pGEM-T Vector (Promega), and transformed in *Escherichia coli* JM109 competent cells (Promega). Plasmid DNA was prepared using Wizard® *Plus* Minipreps DNA Purification Kit (Promega). Several clones were automatically sequenced on both strands by MWG Eurofin Operon (Ebersberg, Germany). Sequence data were deposited in GenBank under the following accession number: AJ879074.

Helianthus annuus BAC-library

The library screened was a genomic DNA library of the BAC type (large inserts) representative of 8.3 genome equivalents of the *H. annuus* inbred line HA383. The library was built with *Hin*dIII DNA digested cloned in IndigoBac536 vector.

Probe labelling and filter hybridization for library screening

Eleven filters containing 202,752 clones of the DNA library were screened with a probe obtained by performing PCR on the clone HaK-3 with the primers SON1F and SON3R. The 650 bp long probe, containing the whole CDS of the *HaL1L* gene, was radioactively labelled with α^{-32}P using the Ladderman Labelling Kit (Takara Bio Inc., Otsu, Japan) according to the supplier instructions. Each of the 11 library filters were pre-hybridized for 30 min at 42°C in ULTRAhyb buffer (Ambion Inc.-Applied Biosystems, Carlsbad, CA, USA). The probe was then added directly to the pre-hybridization buffer and left hybridize was performed for 20 h at the same temperature. Two low stringency washes were performed in 2× SSC and SDS 0.1% for 10 min at room temperature, followed by two higher stringency washes in 1%, or 0.5% or 0.1% SSC and SDS 0.1%, for 15 min a 37°C. After the stringency washes, the hybridized filters were enveloped in plastic film and analysed by Typhoon Variable Mode Image (Perkin Elmer, Waltham, MA, USA). Among the BAC clones positive to hybridization, the clone 135I01 was chosen for automatic sequencing of the entire BAC by MWG Eurofin Operon (Ebersberg, Germany). The BAC sequence was analysed by the following software: GENESCAN availavle at http://www.mobyle. pasteur.fr/cgi-bin/portal.py?form=genescan; FASTA, BLAST and CLUSTALW available sites http://www.ebi.ac.uk/ and http://www.ncbi.nlm.nih.gov/; TSSP at the site htpp://www. sofberry.com/.

In silico analysis

The genomic sequences of the intron (864 bp), as well as upstream (1,598 bp) and downstream (350 bp) regions of the *HaL1L* gene were searched for putative TF binding sites using TRANSFAC (transcription factor binding sites database) [30] (http://www.gene-regulation.com/) and Mat-Inspector, which utilises the PLACE *database* for plants *cis-elements* (http://www.genomatix.de/cgi-bin/eldorado/ main.pl). The upstream region was searched for putative promoter sequence using the PromoterInspector *software* available at http://www.genomatix.de/cgi-bin/eldo rado/main.pl.

Sequence alignment and analysis

Database searches were carried out with the basic local alignment search tool (BLAST) program (BLAST 2.0a) at the National Center for Biotechnology Information [31]. A data set including HaL1L and other 17 amino acid sequences of HAP3 genes was constructed. The sequence alignment, manually adjusted, was obtained using ClustalW [32] (see electronic Supplementary Fig. 1 for all the sequences used in this analysis). Only the peptide regions corresponding to the four α helices were used for phylogenetic analyses, which was carried out using programs from the Phylogeny Inference Package (PHYLIP) group, PHYLIP, Version 3.6 [33]. A distance matrix was computed according to the Dayhoff point accepted mutation model using the Protdist program, which was then used as input for the Neighbour program. For the Neighbour program, the Unweighted Pair Group Method with Arithmetic mean method of clustering (UPGMA) was selected. The resulting tree is shown in Fig. 1. Boostrap values are reported at nodes. As support for the tree, a bootstrap analysis with 100 replicates was performed with the SeqBoot program and a strict consensus tree was obtained using the Consense program.

RNA isolation and semi-quantitative RT-PCR analysis of *Ha-L1L* expression

Total RNA was extracted from zygotic embryos (E) collected at 5 and 30 days after pollination (DAP) and mature leaves (15 cm long) of 60-day-old plants. Extraction was performed using the TriPure Isolation Reagent, according to the manufacturer's instructions (Roche). Four microgram of total RNA from each sample was used for the Superscript pre-amplification kit (Invitrogen S.R.L., Life Technologies, Milan, Italy) to produce the first strand cDNA, using conditions recommended by the manufacturer. PCRs were performed using *HaL1L* gene-specific primers. To normalize the amount of RNA in each sample, an amplification of the constitutively expressed sunflower *β-actin* transcript (*Haβ-actin*) was carried out. The number of PCR cycles was chosen in the exponential range of amplification. The primers used to amplify *HaL1L* were CHI-TCDNA and CHI-REL (5′-GAAGACAATGAGTG CATTG-3′). The primers used to amplify *Haβ-actin* were
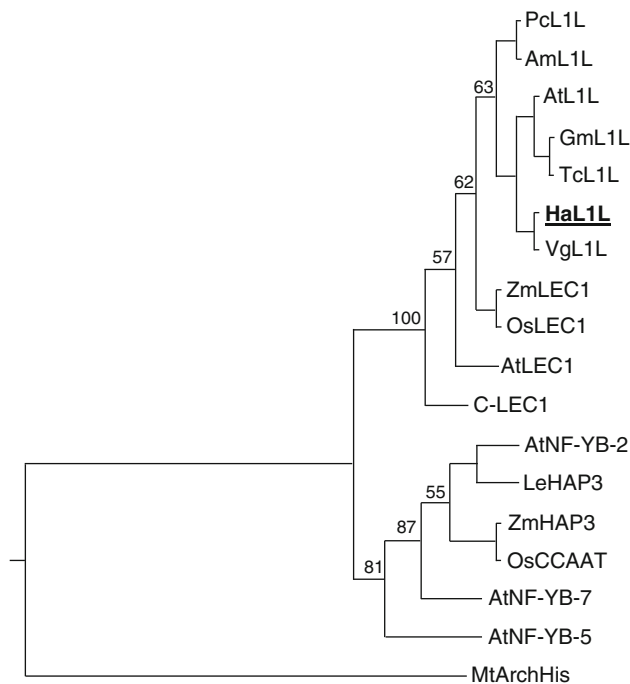
Fig. 1 Phylogenetic tree obtained with the unweighted pair group method with Arithmetic mean (UPGMA) method. *Helianthus annuus* HaL1L (GenBank accession no. AJ863116) was compared to non-LEC type HAP3, LEC-1 and LEC1-like (L1L) peptide sequences in monocot and dicot species. The GenBank accession numbers are: AAC63635.1 for *Arabidopsis thaliana* AtNF-YB-5, AAD22680 for *A. thaliana* AtNF-YB-7, AAK60334 for *A. thaliana* AtNF-YB-2, BT013228 for *Lycopersicum esculentum* LeHAP3, ×59714 for *Zea mays* ZmHAP3, OJ1280_A04.9 for *Oryza sativa* putative CCAAT-binding transcription factor (OsCCAAT), At1g21970 for *A. thaliana* AtLEC1, Q76FX3 for *Daucus carota* C-LEC1, Q6ZFM9 for *O. sativa* OsLEC1, AF4101762 for *Z. mais* ZmLEC1, AY058919 for *Vernonia galamensis* VgL1L, AF533650 for *Phaseolus coccineus* PcL1L, AY058917 for *Glycine max* GmL1L, AY058920 for *Argemone mexicana* AmL1L, A3KLH9 for *Theobroma cacao* TcL1L, and At5G47670 for *A. thaliana* AtL1L. The amino acid sequence of the *Methanothermobacter thermautotrophicus* archeal histone (MtArch-His) was used as an outgroup (GenBank accession no. M90086). At the nodes, bootstrap values are reported as support for the tree

ACT5 (5′-GATTCCGTTGCCCTGAGGTTC-3′) and ACT3 (5′-TCTCTGGAGGAGCAACCACC-3′). Primers were designed to yield 175 bp and 243 bp fragments for *HaL1L* and *Haβ-actin*, respectively. PCRs were performed in 20 μl of 10× buffer (Euroclone Spa-Life Division Sciences, Pero, Italy) containing 0.1 mM dNTPs, 0.75 μM of each primer for *HaL1L*, 0.15 μM of each primer for *Haβ-actin*, 1 mM MgCl$_2$, 1.25 U Euro*Taq* DNA polymerase (Euroclone) and 1 μl of single-stranded cDNA. Amplifications were carried out according to the following temperature profile: 94°C for 5 min for denaturation, followed by 30 amplification cycles (94°C for 30 s, 53°C for 25 s, 72°C for 25 s) and a final extension of 7 min at 72°C. The experiment was repeated with three independent RNA extraction.

Analysis of cytosine methylation status

To verify if cytosine methylation has a role in *HaL1L* expression, a study was performed comparing the methylation pattern of DNA extracted from mature leaves with DNA from zygotic embryos at 5 and 30 DAP. The sequence of the intergenic DNA fragment spanning from the start codon of the *HaL1L* coding region back to the stop codon of the upstream gene was examined with "CpG-finder" *software* available at the site http://www.softberry.com to look for CpG isles. Two isles, Is1 (−730/−531) and Is2 (−153/−56), containing cytosine residues competent to be differentially methylated, were identified.

Sample preparation and sodium bisulfite treatment

To establish the Is1 and Is2 methylation patterns, the CpGGenome^{TM} FastDNA (Chemicon, Temecula, CA, USA) kit was used, which employs sodium bisulfite to chemically change not-methylated cytosine into uracile. For each sample, 5 μg genomic DNA were digested, at 37°C for 16 h, by 50 U of *Ssp*I, producing a 2,024 bp long fragment containing both isles. Subsequently, the samples were treated with Proteinase K (600 U\ml), for 1 h at 37°C and purified by "Wizard SV gel and PCR clean-up system" (Promega). The enzymatic fragmentation may be not sufficient to assure the complete DNA denaturation required for the bisulfite action on not-methylated cytosines. The protocol was greatly improved adding a chromatography step on hydroxyapatite (130-0520 DNA Grade Bio-Gel HTP Hydroxyapatite, Bio-Rad, Italy) to enrich the samples in single strand DNA (ssDNA) *vs* double strand DNA (dsDNA). Chromatography on hydroxyapatite was performed as follow: before starting, 0.2 g of resin was suspended in 0.5 ml of 0.01 M phosphate buffer pH 6.8 (PB), boiled for 5 min and allowed to decant at 55°C. Three μg of *Ssp*I digested DNA, solubilized in 100 μl H$_2$O, was boiled for 5 min and immediately put on ice. Next, 2 μl of 0.5 M PB were added up to a final concentration of 0.01 M. The DNA sample was heated at 55°C, added to the pre-equilibrated resin, and the DNA was allowed to bind for 10 min at 55°C. At the end of the incubation, the mix was loaded into a Pasteur pipette used as chromatography column. The ssDNA was eluted twice with 200 μl of pre-heated (55°C) 0.16 M PB, then the residual dsDNA was recovered with 0.25 M and 0.5 M PB. The ssDNA was briefly dialyzed against ice water on Millipore dialysis filters (0.025 μm).

One microgram of DNA, prepared as above, was added to 550 μl of Modification Reagent, containing sodium bisulfite. The mix was neutralized to pH 7 with NaOH, and incubated at 55°C for 6–18 h, according to instructions of the CpGGenome^{TM} FastDNA kit (Chemicon).

Finally, the DNA sample was subjected to two consecutive PCRs to amplify the Is1 and Is2. In the first PCR, the following primers were used: BLU (forward, 5′-GTTGATAYATGAAATTTTAT-3′) and NERO (reverse, 5′-TCTTTAAAACCTTRAATAATA-3′) for Is1; GIALLO (forward, 5′-AAATTATGATTAAATYTGTT-3′) and INDACO (reverse, 5′-TATTATATATATACACACAC-3′) for Is2. For both isles, amplifications were carried out according to the following temperature profile: 94°C for 2 min for denaturation, 35 amplification cycles (94°C for 30 s, 42°C for 30 s, 72°C for 1 min 30 s) and a final extension of 7 min at 72°C.

For the nested PCRs, the following primers were used: ROSSO (forward, 5′-TYYAGATTTATTGAGTTTTT-3′) and VIOLA (reverse, 5′-ACATTATTTCCTTRATTTATA-3′) for Is1; ARANCIO (forward, 5′-GTGGTTAAATCAAATYTGTT-3′) and PIPPO (reverse, 5′-CCCATAAAATTAAAAAAACATTAT-3′) for Is2. For both isles, amplifications were carried out according to the following temperature profile: 94°C for 2 min for denaturation, 35 amplification cycles (94°C for 30 s, 42°C for 30 s, 72°C for 30 s) and a final extension of 7 min at 72°C.

Amplification products of predicted size were purified with "Wizard SV gel and PCR clean-up system" (Promega) and cloned with "TOPO-TA Cloning Kit" (Invitrogen). The sequences of selected clones, automatically obtained by MWG, were aligned to search sites occupied by C residues in untreated DNA, which are replaced by T in bisulphyte treated DNA. The following softwares were used: BiQAnalyzer at the site http://biq-analyzer.bioinf.mpi-sb.mpg.de; CyMATE at the site http://www.gmi.oeaw.ac.at/CyMATE; CpGfinder at the site http://www.softberry.com.

The results are shown as the mean of three independent experiments (treatments) each with 10 replicates (sequences). The data were processed using analysis of variance procedures, and the means were compared by Tukey's test ($P = 0.05$). Homogeneity of variances was evaluated using Bartlett's test ($P = 0.05$). Statistical analyses on percentage data were performed after arcsine transformation.

RNA extraction and HaWUSCHEL (HaWUS) cDNA isolation

Total RNA was extracted from H. annuus vegetative shoot with the TriPure Isolation Reagent, according to the manufacturer's instructions (Roche). Two microgram of total RNA was used with the Superscript pre-amplification kit (Invitrogen), to produce first strand cDNA in conditions recommended by manufacturer. A cDNA fragment was obtained with the degenerate primers WUSF (forward, 5′-GAWCTBTACTACAWCAAWG-3′) and WUSR (reverse, 5′-MATMGGRAATAAKGGAAGWGT-3′), placed in con-

served regions of the WUS gene determined by comparing the sequence of the WUS gene from Antirrhinum majus ROSULATA (ROA) mRNA (EMBL AY162209); Solanum lycopersicum mRNA (EMBL AJ538329); Petunia hybrida mRNA (EMBL AF481951); Populus trichocarpa mRNA (EMBL AM234747); A. thaliana mRNA (EMBL AJ012310).

The WUSF-WUSR cDNA fragment sequencing allowed us to choose primers for 5′- and 3′-RACE (rapid amplification of cDNA ends), according to the manufacturer's instructions (Invitrogen). 3′-RACE was performed with the specific primer WUS3F (forward, 5′-GTGGCTATGGATCTGTTGCT-3′) with the following PCR conditions: 1 min at 94°C, 30 amplification cycles (30 s at 94°C, 45 s at 60°C, 30 s at 72°C), then 72°C for 7 min. For the 5′-RACE the HaWUS-specific primers WUS15R (reverse, 5′-GGTGGTGAGATGGAATCTGCAT-3′) and WUS25R (reverse, 5′-TTTCTTCTGGCGTTCACGAGCT-3′) were used in the first and second PCR, respectively. The first and second PCR conditions were: 94°C for 1 min, 30 cycles (30 s at 94°C, 30 s at 62°C, 30 s at 72°C), 72°C for 7 min. The PCR amplifications yielded products of 566 and 367 bp for the 3′ and 5′ ends, respectively, which were cloned into the pDrive Cloning Vector (Qiagen, Hilden, Germany) and sequenced. A full-length HaWUS cDNA was obtained with the following primers: WUS-CDSF (forward, 5′-GGGTTTTGTTTCAACCCGTT-3′), and WUS-CDSR (reverse, 5′-GCTAATCAGACATCACCCAT-3′). The PCR conditions were: 94°C for 1 min, 30 amplification cycles (30 s at 94°C, 30 s at 58°C, 30 s at 72°C), and 72°C for 7 min. A product of 1,022 bp was cloned into the pDrive Cloning Vector (Qiagen, Hilden, Germany) and several clones containing the entire sequence were sequenced on both strands.

All PCRs were performed with DreamTaq™ DNA Polymerase (Fermentas).

The HaWUS cDNA was deposed in nucleotide database at EMBL with the accession number HE616565.

Linear template construction for recombinant protein synthesis

A construct was produced which consists in a linear DNA template for in vitro synthesis of a C-terminal 6His-tagged protein. The RTS Wheat Germ LinTempGenSet (5 PRIME, Hamburg, Germany) was employed. Two consecutive PCRs were performed according to the supplier's instructions. For the first PCR two primers were designed that contained gene-specific portions complementary to a region of HaWUS nucleotide sequence as well as portions indicated by the supplier. The sequences of the primers used (gene-specific nucleotides are underlined) were: Sense 5′-CTTTAAGAAGGAGATATACCATGGAAACTCAAC

AACAACAAGC-3′ and AntiSense 5′-TGATGATGAGAA CCCCCCCCCGTGAGATGGAATCTGCAT-3′. These primers allowed the amplification of a 122 aa long region of *HaWUS* (Supplementary Fig. 2) that PROSO II *software* had identified as being soluble in the wheat germ *cell free* protein system. PROSO II is available at http://mips. helmholtz-muenchen.de/prosoII/prosoII.seam. This particular protein segment was chosen due to the insoluble character of the full length recombinant protein (293 aa). In the second PCR, two DNA fragments were placed in the reaction mixture. These fragments contained regulative elements for the correct transcription and translation of the final construct and they had extremities which overlapped with the ends of the first PCR product. The amplification happened via the flanking primers corresponding to the ends of the added DNA fragments. The nucleotide sequences of both DNA fragments and flanking primers are not available in the RTS Wheat Germ LinTempGenSet manual. The final product was a linear construct containing the first PCR product lengthened with the above mentioned DNA fragments. After sequencing to confirm the correct open reading frame, the construct was ready for use as DNA templates in RTS^{TM} Wheat Germ Continuous Exchange Cell-Free (CECF) in vitro protein expression reactions (5 PRIME). All PCRs were performed using the Expand^{TM} High Fidelity PCR System (Roche).

Recombinant protein synthesis

Recombinant protein synthesis was carried out in a eukaryotic cell-free protein synthesis system based on a wheat germ lysate, the RTS^{TM} 100 CECF Wheat Germ (5 PRIME). The reaction was conducted in a container divided into two compartments by a membrane, according to the supplier's protocol. The first compartment was filled with 1 ml of Feeding solution and the second compartment was filled with 50 μl of Reaction solution plus 2 μg of linear template prepared as above. The container was shaken at 900 rpm at 24°C for 24 h in an Eppendorf Thermomixer Comfort 5355 (Eppendorf AG, Hamburg, Germany). Then the reaction solution containing the 6His-tagged WUS recombinant protein (tagged at the carboxylic end) was carefully recovered and stored frozen until purification.

Purification of the 6His-tagged WUS recombinant protein

The 6His-tagged WUS recombinant protein was purified under native conditions to maintain its DNA binding ability. For this purpose 100 μl of binding buffer were added to a tube with 25 μl of 50% PerfectPro Agarose slurry (PerfectPro Ni–NTA Agarose; 5 PRIME) and 50 μl

of synthesized protein. The resulting solution was gently mixed on a rotary shaker (200 rpm) at 4°C for 60 min, then it was centrifuged for 10 s at $1,000 \times g$ to pellet the resin. 10 μl of the supernatant, that can contain any proteins not tied to the resin, were transferred to a fresh tube for SDS-PAGE analysis. The pellet, which contained the recombinant protein bounded to the resin, was washed twice with 100 μl of wash buffer centrifuging for 10 s at $1,000 \times g$. The recombinant protein was eluted 5 times with 15–20 μl of elution buffer centrifuging for 10 s at $1,000 \times g$ between each elution step and the supernatants were carefully transferred into fresh tubes. All the recovered elutes were analysed by SDS-PAGE to confirm the expected WUS recombinant protein band at a MW of 15 KD.

All the three buffers used for recombinant protein purification consisted of 50 mM phosphate buffer and 300 mM NaCl, pH 8, but they contained increasing amount of imidazole at concentration of 10 mM, 20 mM, 250 mM in binding, washing and elution buffer, respectively.

Electrophoretic mobility shift assay (EMSA)

EMSA was performed with a DIG-labelled synthetic double strand oligonucleotide 39 bp long, as a probe, according to DIG Gel Shift Kit, 2nd generation (Roche) protocol. The probe was prepared as follows: a 39 mer single strand sense oligonucleotide (named EMSA6) and its complementary antisense, corresponding to the *HaL1L* intron region containing the WUSATA motif, were purchased from the MWG Eurofin Operon. The sequence of EMSA6 sense was: 5′-TATGTATATATGATGTCATT AATATGATGTTATATATTC-3′, where the WUS target sequences is underlined. Equal volumes of two solutions containing the sense and antisense oligonucleotides, respectively, both concentrated 12 μM in TEN-buffer, were mixed, incubated 10 min at 95°C then cooled slowly to 25°C to obtain a double strand oligonucleotide that was diluted with more TEN-buffer to a final concentration of 3 μM.

Subsequently, the double strand oligo (3.85 pmol) was DIG-labelled according to the manufacturer's protocol (Roche) and was then ready for use as a probe. For the gel shift reaction three 25 μl samples were made following the kit instructions: Sample1 contained purified recombinant protein (75 nmol) and DIG-labelled probe (15.5 fmol); Sample2 was the same as Sample1 but contained unlabelled double strand oligo in excess (125-fold) for specific competition; Sample3 was the same as Sample1 but without recombinant protein. Given that the probe sequence is A/T rich, Poly [d(A-T)] was also added to all three samples for non-specific competition. The samples were carefully mixed, incubated for 15 min at 20°C then placed on ice.

The reaction mixtures were loaded onto a 5% polyacrylamide pre-casted gel in 0.5× TBE buffer (5% polyacrylamide TBE Criterion XT precast gel, Bio-Rad, Berkeley, CA, USA). The gel was pre-run (10 min) and run at 8 V/cm until the front reached 2/3 of the way to the bottom. We recommend the use of loading buffer without bromophenol blue; to follow the run loading buffer with bromophenol blue can be loaded in a empty lane of the gel. After the run, the gel content was transferred to a positively charged membrane (Roche) by electroblotting performed for 30 min at 400 mA. Finally, chemiluminescent detection was conducted as described in the DIG Gel Shift Kit procedure to make visible the protein/probe complex.

Results

Organization of DNA region containing *HaL1L*

A genomic DNA fragment containing the entire *HaL1L* gene was analysed. A first portion consisting in the full-length intron/exon region of *HaL1L* (comprehensive of two exons and one intron) 1,684 bp in length was amplified by

PCR with the CHITCDNA/CHISR primers (Fig. 2; Supplementary Fig. 3). The intron (864 bp) had recognizable canonical GT/AG borders. The position of intron was deduced by comparing *HaL1L* cDNA [27] with the genomic DNA sequence. Next, genomic DNA walking was used to isolate the entire *HaL1L* gene and its flanking regions (Fig. 3a–c). Amplified products containing the 5′-(2,523 bp) and 3′-end (542 bp), respectively, were cloned and sequenced (Fig. 3b, c). The nucleotide sequences were confirmed and lengthened by screening a sunflower genomic DNA BAC Library. An intergenic region of 1,598 bp spans from the *HaL1L* translation start codon back to the translation stop codon of the upstream gene, which encodes for a putative TF similar to the *A. thaliana* gene At5G47660.1 (Fig. 2a; Supplementary Fig. 3). Downstream to *HaL1L*, an intergenic region of 350 bp precedes a gene encoding a RNA binding/tRNA methyl transferase similar to the gene At5G47680.1 of *A. thaliana*.

The promoter region of *HaL1L* was located by TSSP and CpGfinder *software* basing the search on the consensus sequences that in plants characterize regions flanking the coding sequence. The *HaL1L* promoter is of the 'TATA-type' for the sequence TTATA found at −29 bp before the
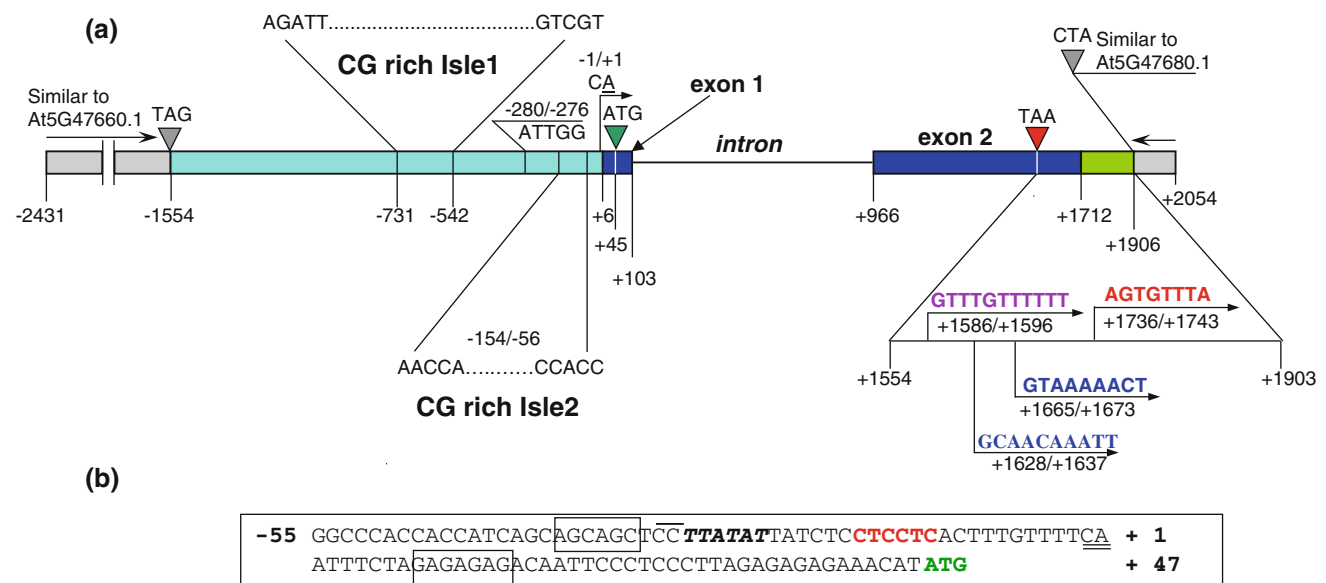
Fig. 2 Schematic representation of the *LEAFY COTYLEDON1-LIKE* gene of *Helianthus annuus* (*HaL1L*). The GenBank accession number is AJ879074. **a** The *light blue rectangle* indicates the promoter region of *HaL1L*. The position of two GC rich isles (1 and 2) are reported. CA indicate the putative *HaL1L* transcription start site (TSS) as well as the start of nucleotide sequence numeration (−1/+1). ATG and TAA represent the *HaL1L* translation start and stop codon, respectively. The *blue rectangles* represent the two exons of *HaL1L* joined by an intron depicted as a line. The *green rectangle* represents the 3′-region of *HaL1L*. The citoplasmatic polyadenilation signal is shown in *purple*. The nuclear polyadenilation signals are shown in *blue*. The signal of transcription end is shown in *red*. The upstream gene similar

to At5G47660.1 of *A. thaliana* (TAG: translation stop codon) and the downstream gene similar to At5G47680.1 of *A. thaliana* (in reverse orientation; CTA: translation stop codon) are shown in gray. **b** The nucleotide sequence of the *HaL1L* proximal promoter. The TATA Box is written in *italic* and *bold* characters. The CC dimer, upstream of the TATA box and specific for plants, is upper-lined. The enhancers of transcription are boxed. CA, the putative *HaL1L* transcription start site (TSS) and start nucleotide sequence numeration (−1/+1), is double underlined. The Y-Patch, specific to proximal plant promoters, is written in *bold* and *red* characters. The ATG is written in *bold* and *green* characters
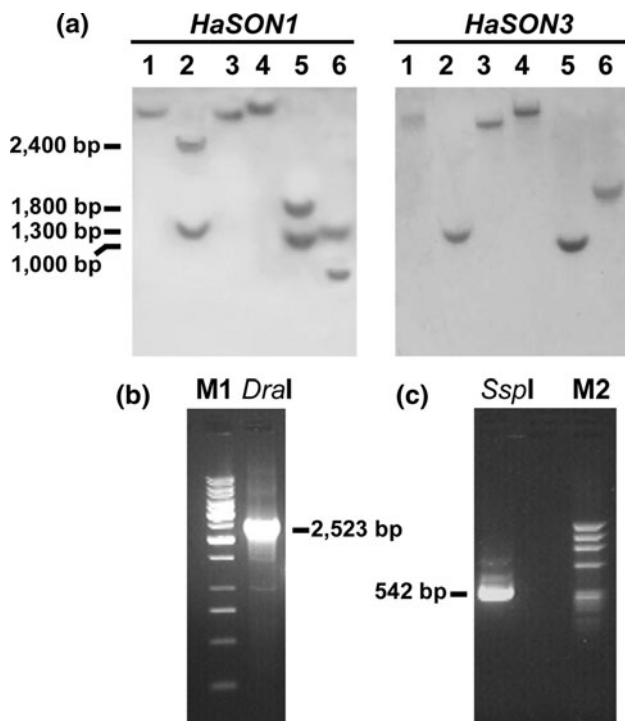
Fig. 3 Genome walking on the *LEAFY COTYLEDON1-LIKE* gene of *Helianthus annuus* (*HaL1L*). GenBank accession number AJ879074. **a** DNA gel blot hybridization. Genomic DNA was cut with restriction enzymes in single or double digestions with *Pvu*II (*lane 1*), *Dra*I (*lane 2*), *Eco*RV (*lane 3*), *Stu*I (*lane 4*), *Ssp*I (*lane 5*) and *Sty*I + *Pvu*II (*lane 6*) and separated on a 1.2% (w/v) agarose gel. Hybridization was performed with the probes *HaSON1* or *HaSON3*. **b, c** Nested PCRs of 5′ (**b**) and 3′ (**c**) RAGE. The PCR products were resolved on ethidium bromide-stained 1.5% TAE-agarose gel and visualized under UV light. M1 and M2 (Marker DNA ladders) stand for the: 1 Kb marker and PhiX174 Hae III Digested DNA, respectively (New England, BioLabs, Inc)

transcription start site (TSS) (Fig. 2a, b; Supplementary Fig. 3). The TSS was identified in the CA dimer, where the adenine residue represent the first transcribed nucleotide [the numeration of DNA sequence starts at this dimer (−1/+1)]. The YR-rule (where TSS is R underlined, Y stands for C or T and R stands for A or G) is respected as many as 77% of the *Arabidopsis* promoters. Also, a Initiator motif (Inr) is visible surrounding the TSS: YYANYY, where TSS is underlined, and Y = C\T; N = any base.

Two cytosine residues are present contiguous to the 5′-end of the TATA box and to the *Y-Patch* motif (CTCCTC) at −20 bp from TSS (Fig. 2a, b; Supplementary Fig. 3). Other typical recognizable elements are the CCAAT sequence, localized in reverse orientation at −280 bp from TSS, and two CG rich isles spreading from −730 to −531 bp (Is1) and from −153 to −56 bp (Is2) from the TSS. A transcription termination signal as well as two nuclear polyadenylation signals involved in the mRNA maturation process, were identified at the 3′-region of the *HaL1L* gene by a silico analysis (Fig. 2; Supplementary Fig. 3).

Putative target sites of regulatory proteins are scattered throughout the *HaL1L* sequence

Putative target sites of both positive and negative regulatory proteins were found by the MatInspector *software* [40] in the genomic sequences of the intron, as well as upstream and downstream regions of the *HaL1L* gene. Sites with matrix similarity >0.9 are shown in Fig. 4 (Supplementary Figs. 4, 5, 6 for more exhaustive lists). In particular, six WUSATA motifs were found, which represent the putative target sites of WUS TF.

To verify the functionality of the putative DNA target sites, we decided to test whether or not the WUSATA motifs were able to bind the WUS TF protein. The information necessary to build DNA constructs, which we needed in order to synthesize the WUS protein in vitro (using the wheat *cell free* system for protein synthesis), came from *HaWUS* cDNA sequence. The full length, 293 aa long HaWUS protein was found insoluble. Therefore, a soluble fragment (122 aa) of the protein that coincided with the amino terminus was used in gel retardation tests (Supplementary Fig. 2). The WUS target site, out of the six ones identified in *HaL1L* genomic sequence, which we tested was located in the intron. We decided to test this specific WUSATA motif for the complete identity of nucleotide sequence with the WUSATA motif of *A. thaliana AGAMOUS* (*AG*) gene and for its same location in the intron. We found that the HaWUS fragment was able to bind and slow down the probe, as the protein/probe complex was electrophoretically delayed in comparison to the free probe (Fig. 5, lanes 1, 3). That the binding was specific was proved by the ability of the unlabelled probe present in excess to remove the labelled probe from the complex with the protein fragment (Fig. 5, lane 2).

DNA methylation as control of *HaL1L* transcription

To check if cytosine methylation is involved in the control of *HaL1L* transcription, we analysed of the methylation status of two CG rich isles Is1 and Is2.

In the 200 bp long Is1 the overall methylated cytosine percentage was found to be 38.33% of the total residues in leaves, 33.33% in 30 DAP embryos and 19.05% in 5 DAP embryos (Fig. 6a). Based on nucleotides contiguous to the cytosine, different classes of methylated sites may be identified. The percentage of methylated cytosines in the class of CHG sites (where H is not G.) was 16.88% for 5 DAP embryos, 70.91% for leaves and 59.09% for 30 DAP embryos (Fig. 6a, b). As for the GCN class (where N stands for any base), the percentage of methylated cytosines was: 60.32% for 5 DAP embryos, 75.93% for 30 DAP embryos and 86.67% for leaves. Unlike the previous two
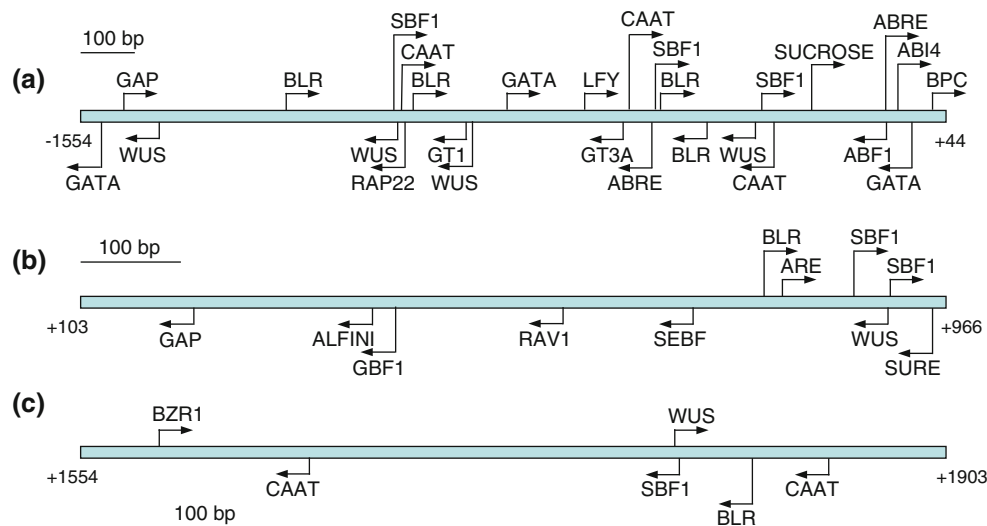
Fig. 4 5′-upstream (a), intron (b) and 3′-downstream (c) regions of the *LEAFY COTYLEDON1-LIKE* gene of *Helianthus annuus* (*HaL1L*), GenBank accession number AJ879074. Analyses were performed by using MatInspector *software* that utilises the PLACE *database* for plants *cis-element*s [34]. Some of putative transcription factor-binding sites and *cis*-acting elements discussed in the text are shown. A complete lists of putative transcription factor-binding sites and *cis*-acting elements is reported in Supplementary Fig. 4, 5, 6. ABF1 and ABRE: ABA response elements; ABI4: Coupling element 1 binding factor; ALFINI: Salt/drought responsive elements; ARE and SEBF: Auxin response elements; BLR: *Arabidopsis* BELLRINGER homeobox protein; BPC: GAGA elements; BZR1: Brassinosteroid response elements; CAAT: CAAT binding factor; GAP: GAP-Box (light response elements); GATA: plant I-box site; GBF1: Plant G-box/C-box bZIP proteins; GT1, GT3A and SBF1: GT-box elements; LFY: LEAFY binding site; RAP22: Light responsive element motif; RAV1: 5′-part of bipartite RAV1; SUCROSE: Sucrose box; SURE: Sugar response elements; WUS: *Arabidopsis* WUSCHEL homeobox protein. The position of the motifs are in scale to the *HAL1L* regions. The *arrows* indicate the strand on which the motifs are located

classes, the CHH class showed very similar values in all the experimental materials: 2.97% in leaf, 3.03% in 30 DAP embryos, 3.25% in 5 DAP embryos (Fig. 6b). Finally, by examining cytosines one by one in order of their distribution in the Is1 at least two conclusions could be drawn: (i) eight cytosine residues located in positions 19, 20, 36, 76, 109, 154, 160 and 172 accounting for 19% of total cytosines in the Is1 and belonging to all classes of methylated sites, were found never methylated in 5 DAP embryos only (Fig. 7a-c and Supplementary Fig. 7); (ii) the percentage of methylated cytosine at site 97 (CHH class) were higher in 5 DAP embryos (71.43%) than in leaves (20%), while an intermediate value was detected in 30 DAP embryos (33.33%). Differently from Is1, no differences in methylation status were observed in the DNA of the Is2 region in our sample.

Discussion

In the model plant *A. thaliana* the LEC1-type HAP3 subunits include two peptides, LEC1 and L1L, that are encoded by genes with a common origin and that are both required for embryogenesis [8, 27, 34, 35]. L1L can functionally replace LEC1 when expressed ectopically in *lec1* mutants [7, 8], though it is not yet known if the two

subunits have distinct or redundant roles. RNA interference mediated suppression of *L1L* expression induces defects in embryo development that differ from those found in *lec1* mutants, suggesting that LEC1 and L1L are not completely redundant and that they have different roles in embryogenesis [8, 34]. Unlike the reported embryo-lethal phenotype of L1L-RNAi lines of Kwong et al. [8], other authors have reported that a loss-of-function T-DNA insertion line, l1l-1, does not present apparent phenotypical alterations during seed development [41]. This behaviour was interpreted as due to *LEC1* redundancy in comparison to *L1L*. Vice versa, the evident defects in *lec1* mutant embryo development might derive from the reduction of both LEC1 and L1L NF-YB subunits. Indeed, in LEC1-deficient embryos L1L is also reduced, probably as the result of the FUS3 regulation by LEC1 [41, 42].

HaL1L, has been identified as a LEC1-type HAP3 subunit of a NF-Y TF by a phylogenetic analysis of HAP3 peptide sequences from monocot and dicot species (Fig. 1). In the consensus tree, LEC1 class peptides and non-LEC1 class peptides were grouped in well separated clades. Similarly, the subclass L1L was clearly distinguishable and HaL1L was more closely related to L1L than to LEC1 (this work and [27]). On the other hand, it is still unclear if an orthologous gene to *AtLEC1* actually exists in the *H. annuus* genome. A search in EST databases returned only
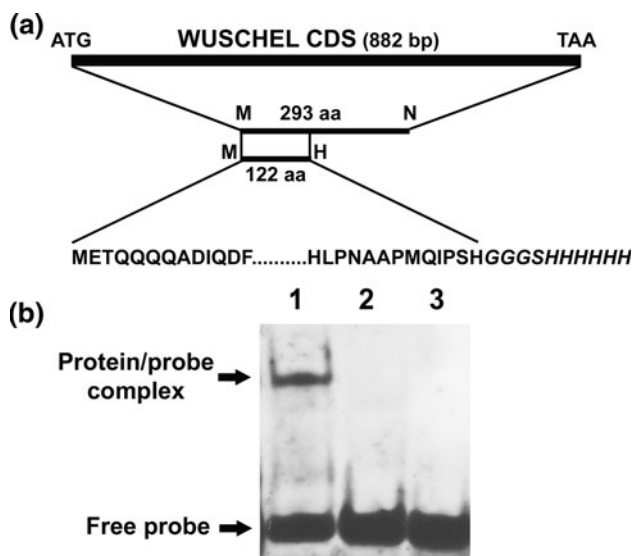
Fig. 5 HaWUS binds to the *HaL1L* intron. **a** Illustration of the construct containing a WUS protein fragment [122 amino acids (aa)] 6His-tagged at the C-terminal end. The construct was used in EMSA. The Coding Sequence (CDS) *WUSCHEL* (*WUS*) (682 bp) is depicted as a black rectangle, ATG and TAA are the translation start and stop codon, respectively. M (metionine) and N (asparagine) indicate the amino acids that delimitate the full-length WUS protein (293 aa), respectively. H (histidine) indicates the last amino acid of the WUS protein fragment. See also Supplementary Fig. 2. **b** Immunodetection of EMSA performed with recombinant WUS (RW). *Lane 1*: RW/ probe complex delayed in comparison to the free probe. The probe is a DIG-labelled 39 ds oligonucleotide containing the WUS binding motif like in the *HaL1L* intron. *Lane 2*: The probe is removed from the complex with RW by the unlabelled 39 ds oligonucleotide. *Lane 3*: Free probe

HaL1L and other less related sequences; in addition, a preliminary screening of a DNA genomic BAC library did not reveal the presence of a second *HAP3-LEC1* type gene (data not shown).

Organization of the DNA region containing *HaL1L*

The analysis of a DNA fragment containing the *HaL1L* nucleotide sequence led to the identification of several interesting features.

The *HaL1L* DNA region, comprehensive of the upstream and downstream genes, was found to be homologous to the *A. thaliana* chromosome V region carrying *AtL1L* (At5G47670). A 1,598 bp (424 bp in *A. thaliana*) flanking sequence separated *HaL1L* from the 5′ upstream gene encoding a putative TF, similar to the At5G47660.1 gene of *A. thaliana*. Downstream from *HaL1L*, an intergenic region of 350 bp (350 bp in *A. thaliana*) precedes a gene similar to the *A. thaliana* At5G47680.1 gene that encodes an RNA binding/tRNA methyl transferase.

Several typical elements allow the identification of the *HaL1L* core promoter that is of the 'TATA-type'. The
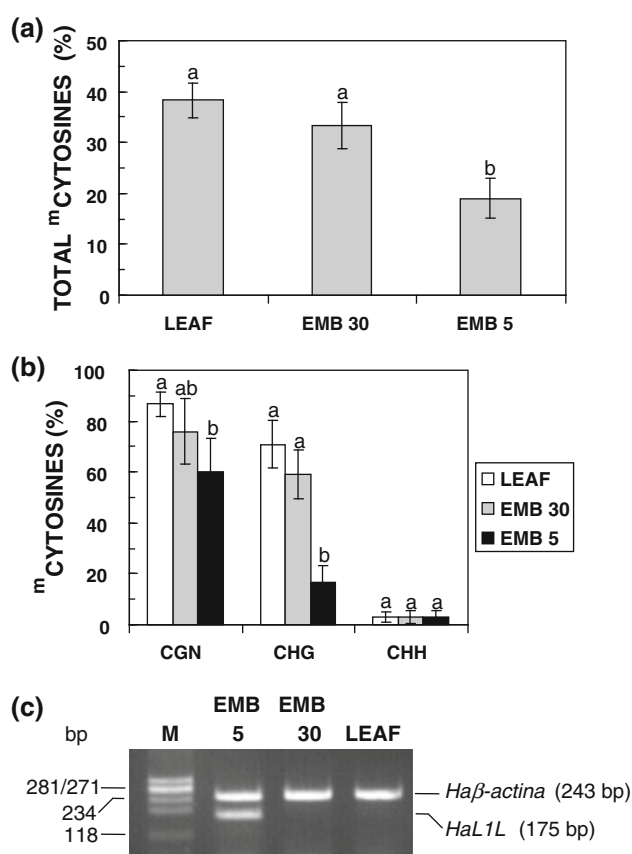
Fig. 6 Analysis of *LEAFY COTYLEDON1-LIKE* (*HaL1L*) transcription and DNA methylation in *HaL1L* 5′-upstream Isle1 (Is1) during the transition from vegetative to embryonic growth. **a** Methylation profiles for *HaL1L* 5′-upstream Is1. DNA methylation was analysed by bisulfite sequencing in leaves and in zygotic embryos at 5-, and 30-DAP. The vertical axis represents the percentage of methylated cytosines (᷾cytosines) compared to the total number of cytosine residues. **b** The extent of methylation in different sequence contexts, CG, CNG (N = A, T, G, or C), and CHH (H = A, T, or C) sites, in leaves and zygotic embryos (5- and 30-DAP) expressed in ᷾cytosine percentage. The results are presented as the mean (±sd) of three independent experiments (treatments) each with 10 replicates (sequences). Values with the *same letter* are not significantly different at the *P* = 0.05 level according to Tukey's test. **c** Semi-quantitative RT-PCR analysis for the steady state *HaL1L* transcripts in sunflower embryos (EMB) at different days (5 and 30) after pollination (DAP) (EMB 5 and EMB 30) and in mature leaves (LEAF). The sunflower *β-actin* gene (*Haβ-actin*) was used as internal control. The RT-PCR products were resolved on ethidium bromide-stained 2.0% (w/v) TAE-agarose gels and visualized under UV light. M: Marker DNA ladders, PhiX174 Hae III Digested DNA (New England, BioLabs, Inc)

transcription start site was identified in a CA dimer, which respects the YR-rule as do 77% of *Arabidopsis* promoters [38]. A pyrimidine-rich Initiator motif (Inr) was also recognizable surround the TSS. Inr elements are known in some mammalian promoters [36], and they are also functional in plants [37]. A survey of *Arabidopsis* TSSs revealed that a limited number of promoters (less than 10%) have the Inr motif around the TSS [38].
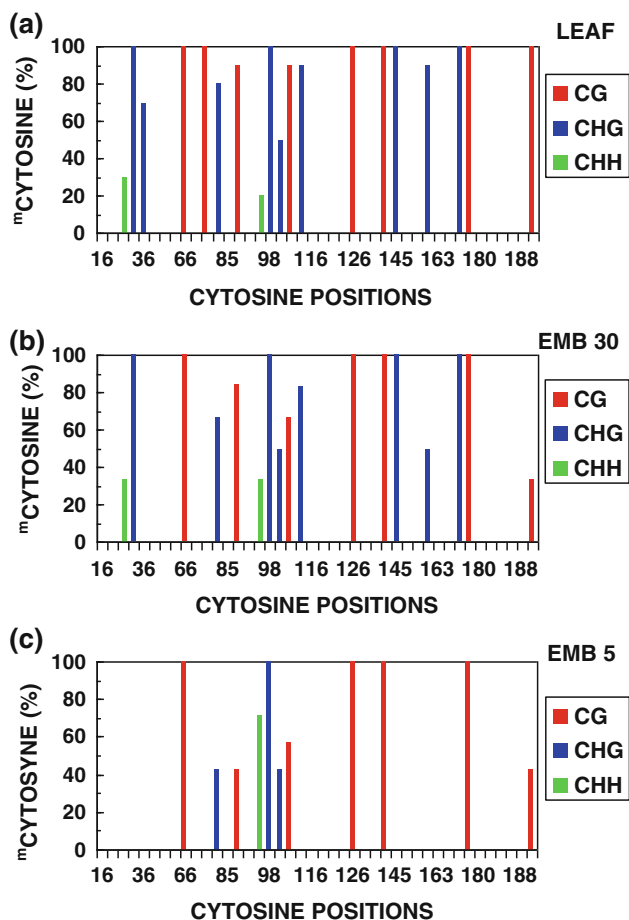
Fig. 7 Cytosine methylation in *LEAFY COTYLEDON1-LIKE* (*HaL1L*) 5′-upstream Isle1 (Is1) during development. DNA methylation was analysed by bisulfite sequencing in leaves (**a**) and zygotic embryos at 30- (**b**) and 5-DAP (**c**). The *horizontal axes* represents all cytosines present in Is1 of the *HaL1L* 5′-upstream region. The *vertical axes* indicates the percentage ^mcytosine

As for the coding region, analysis showed that *HaL1L* is composed of two exons and an intron bordered by canonical GT/AG dimers. Interestingly, we found that the intron contains target motif for DNA-binding proteins suggesting that it may play a role in regulating *HaL1L* transcription. This will be discussed further in the next section.

The cytoplasmic polyadenylation signal which was identified downstream to the coding region suggested a negative *HaL1L* post-transcription regulation. It is known that poly(A) tails lengthened in the cytoplasm are able to form complexes with regulative proteins that segregate and temporarily inactivate mRNAs in the cytoplasm [43, 44]. Elongation of the mRNA poly(A) tail is done by cytoplasmatic poly(A)polymerases (PAPs), that during embryo development act under hormonal control [45]. The existence of a mRNA translation control is supported by in situ hybridization analysis which shows that *HaL1L* transcripts accumulate in maternal tissues of sunflower embryos [27].

DNA methylation controls *HaL1L* transcription

In sunflower, high levels of *HaL1L* mRNA were detected in the developing embryo, in the embryo suspensor, endosperm, integument, and integumentary tapetum cells [27]. Localized *HaL1L* expression and IAA accumulation in leaf epidermis domains represent early events of somatic embryogenesis displayed by an epiphyllous clone (EMB-2) of the interspecific hybrid *H. annuus* × *H. tuberosus* [28].

In our experiments, semi-quantitative RT-PCR showed that 5 DAP embryos contains high amounts of *HaL1L* transcript, while in 30 DAP embryos *HaL1L* transcript was undetectable (Fig. 6c). Also *HaL1L* transcript accumulation was very low or absent in leaves (Fig. 6c), stems and roots [27]. In view of the inverse relationship between DNA methylation and transcription, we believe that *HaL1L* is hyper-methylated in its regulative regions in organs with low *HaL1L* transcript accumulation. The comparison between embryos (5 and 30 DAP) and leaves suggests that an epigenetic mechanism involving cytosine methylation, is active during the sunflower embryogenesis. Indeed, overall methylated cytosine percentages in Is1, located in the *HaL1L* promoter, were significantly higher in leaves and 30 DAP embryos (38.33 and 33.33%, respectively) than in 5 DAP embryos (19.05%) (Fig. 6a). Observing the pattern of methylated cytosines by class of sites may be particularly significant. Major differences, which seems rather organ specific, were evident in methylation of CHG and GCN classes, while no difference were observed for the CHH class. The percentage of methylated cytosines in CHG sites was much lower in 5 DAP embryos (16.88%) than in 30 DAP embryos (59.09%), and the difference was even greater when compared to leaves (70.91%). Regarding the GCN class, 5 DAP embryos showed lower percentage of methylated cytosines (60.32%) only in comparison to leaves (86.67%). Indeed 30 DAP embryos (75.93%) were not really different from 5 DAP embryos and leaves (Fig. 6b). Within classes, differences in methylation of single cytosines, in selected sites, might exert a biological role. This may be the case of the eight cytosine residues (19% of total cytosines) in Is1 that were never methylated in 5 DAP embryos (Fig. 7c and Supplementary Fig. 7) and of the site 97 (CHH class) where the percentage of methylation was higher in 5 DAP embryos (71.43%) than in leaves (20%), with an intermediate value being detected in 30 DAP embryos (33.33%) (Fig. 7a–c). Notably, in plants, de novo methylation is often observed at CNG and CHH sites [46].

Our data suggest that DNA methylation could regulate the expression of *HaL1L*, in agreement with the recent proofs obtained in carrot [26]. Carrot *C-LEC1* expression is regulated by DNA methylation of a 5′-upstream region (Region 1), which shows a lower number of methylated

cytosines during somatic embryogenesis than after the transition from embryonic to vegetative growth. Moreover, in vitro hypermethylation of the Region 1 induced a dramatic reduction in *C-LEC1* expression, indicating a negative correlation between DNA methylation and *C-LEC1* expression. The same authors also speculate that the *C-LEC1* 5′-upstream region includes other important controlling elements, such as target sites for not yet identified TFs [26].

HaL1L transcription regulation may be exerted by regulative DNA-binding proteins

In plants, like in other higher eukaryotes, transcription regulation is mediated by a large number of TFs that control the expression of hundreds of target genes in numerous, sometimes intertwined, signal transduction cascades. TFs recognize specific binding sites, known as regulatory motifs, which are functional elements able to determine the timing and location of transcriptional activity. These elements are primarily located in the non-coding sequences upstream of a gene, although functional elements in introns and 3′ regions have been described as well. The discovery of regulatory motifs and their organization is an important first step to improve our understanding of gene expression and regulation [39]. However, not all motifs present are functional elements or simultaneously active, since the cooperation between TFs is context dependent, therefore only stringent tests can determine the actual working significance of each specific target site.

Bioinformatic analysis carried out on the nucleotide sequences of the entire *HaL1L* gene, highlighted the complex regulation to which this gene is subjected. The putative target sites found in the *HaL1L* sequence, can be grouped in several classes on the basis of the relative binding protein functions. In the next section we discuss the significance of the auxin response element (*ARE*), ABA responsive elements (*ABRE*), GATA, GAGA-box and WUSATA motif, that respectively suggest that hormones (ABA and auxin), light signals, *Polycomb* and homeotic (*WUS*) genes are involved in the *HaL1L* transcription regulation system.

The analysis of embryonic hormonal control emphasizes the role of auxin, abscisic acid (ABA) and GAs [9]. The presence of the ARE motif (TGTCTC) in the intron region of *HaL1L* suggest that Auxin Response Factors (ARFs) might regulate the expression of this gene (Fig. 4; Supplementary Fig. 5). The key role of auxin in plant embryogenesis has been well described [47]. The establishment of auxin synthesis and polar auxin transport are key steps in embryo development. During early zygotic embryogenesis auxin accumulates dynamically at specific positions that correlate with developmental decisions

regarding transport regulation and response machinery [48]. Auxin can induce ectopic *FUS3* expression in epidermal cells of cotyledons leading to a reduced venation pattern [49]. Also, ectopic expression of *LEC2* causes rapid activation of auxin-responsive genes [50]. Nonetheless, only minor auxin-related phenotypes have been observed during early embryogenesis in the *lec* mutants, such as the occasionally detected abnormal development of the suspensor [34, 35]. Conversely, ectopic expression of *LEC1* and *LEC2* in vegetative cells can trigger embryo-like structure initiation [34, 35]. It has been also proposed that the role of *LEC1* in the promotion of embryonic cell identity and division requires auxin [51]. Interestingly, although the ectopic expression of *HaL1L* seems to be not directly mediated by auxin levels per se, localized *HaL1L* expression and IAA accumulation in leaf epidermis domains represent early events of somatic embryogenesis displayed by epiphyllous leaves of the interspecific (*Helianthus annuus* × *H. tuberosus*) EMB-2 clone [28].

Molecular analysis of ABA-responsive gene promoters has led to the identification of several motifs capable of conferring ABA responsiveness to a minimal promoter. One such motif is ABRE, which belong to the so-called G-box family. ABRE contains an ACGT core, a sequence known to be recognized by plant bZIP proteins [52–55]. A single copy of ABRE is not sufficient for ABA-mediated induction of transcription. The multiple ABREs and the so-called coupling element (CE) detected in the 5′-region of *HaL1L* (Fig. 4; Supplementary Fig. 4) could be sufficient to establish a minimal ABA-responsive complex (ABRC), and thereby confer ABA responsiveness to *HaL1L* promoter [56]. As the embryo enters in the maturation phase, ABA content increases, and the resulting high ABA/GA ratio promotes maturation. It also induces dormancy and inhibits cell-cycle progression, embryo growth and germination. ABA is required for the induction of some seed storage proteins (SSPs) in seedlings ectopically expressing *FUS3* [42, 57] or *ABI3* [58]. Also, it has been proved that the activity of *ABI3* and *FUS3* can be regulated at the post-translational level by ABA [49, 59]. Additionally, it has been suggested that *LECs* might confer competence to somatic cells to undergo embryogenesis by repressing GA levels and, perhaps, by enhancing ABA levels, creating a cellular environment similar to that of maturation phase of zygotic embryos [10].

Sucrose box motifs were found in the intron and the 5′-region of *HaL1L* (Fig. 4; Supplementary Fig. 4 and 5). Sugars can function as hormone-like signalling molecules that adjust metabolism, growth, and development of plants. Sugar signalling operates at all phases of the plant life cycle and dominates plant metabolism. Most sugar signalling effects appear to be mediated through transcriptional control. Changes in the sugar concentration cause

induction or repression of gene transcription [60]. A large but specific set of genes are positively regulated by sugars. The majority of genes identified, that are induced by elevated sugar levels, encode products that help set capacity for carbon storage, utilization, and import. Other important classes of genes regulated by sugar, include genes for storage proteins, defence genes, and genes involved in secondary product pathways [60]. It has been suggested that in addition to auxin, the role of *LEC1* needs sucrose to activate cell division and promote an embryonic cell fate [51]. Indeed, the phenotype penetrance of *lec1-d^{tnp}*, a gain-of-function mutant of *AtLEC1,* is almost entirely dependent on the presence of a carbon source in the growth medium, with sucrose found to be the most effective [51]. It has been suggest that in *lec1-d^{tnp}*, sucrose may act as a signal to induce starch production, and that sucrose action may be increased by *LEC1* ectopic expression.

Several motifs, which we found scattered throughout the *HaL1L* sequence, lead back to TFs involved in regulative mechanisms responsive to light (Fig. 4; Supplementary Figs. 4, 5, 6). NF-Y was recently discovered to be involved in the control of photoperiod regulated flowering time [61]. In particular, the presence of the GATA motif suggests that *HaL1L* may interact with GATA proteins. The *Arabidopsis* zinc-finger-type transcriptional activator CONSTANS (CO) is a CCT domain protein (CCTp) which belongs to the family of GATA proteins. In *Arabidopsis*, CO is the key regulator of photoperiod-induced flowering time [62]. A study of functional and structural relationship between CCTps and NF-Y complexes showed that *AtLEC1* interacts both with CO and the related protein CO-LIKE 15 (COL15). *AtL1L* however, interacts only with CO [62]. Within the CCT domain, a region, that shares a conserved motif with NF-YA subunits and that is involved in DNA–protein or protein–protein complexes has been described. It was hypothesised that CCTps can regulate gene expression by replacing subunits of the NF-Y complex. Since a CO/NF-YB/NF-YC complex does actually exist, so it is thought that CO might exerts its regulatory effect on *FLOWERING LOCUS T* (*FT*) expression by competing with NF-YA subunits in the NF-Y heterotrimer assembling [62]. The amount of CO mRNA is regulated by the circadian clock and by light so that it accumulates at the end of the day under long days and during the night under short days [63, 64]. These data suggest that in plants: (i) the NF-Y complex is able to induce gene activation by interacting with CCT domain proteins that play important roles in environmental responses [62]; (ii) the NF-YB and NF-YC subunits can physically interact with full-length CO, and are genetically required for CO-mediated floral promotion [64].

Both biologically active wavelengths of light and plant hormones act through specific signal transduction

mechanisms to coordinate many aspects of development in higher plants [65, 66]. It is known that the same signal transduction mechanism can be used by two different signals, for example blue light and ABA. *AtLEC1* was identified as a fundamental member both of the transduction system and the signalling mechanisms inhibiting the ABA-mediated delay in seed germination. On the contrary, *AtL1L* seems to have a role only in the ABA response, confirming the hypothesis that its function does not completely overlap with that of *AtLEC1*.

Another motif present in the 5′-region of *HaL1L* that deserves mention, is the GAGA-box which binds Pc proteins (Fig. 4; Supplementary Fig. 4). In metazoan NF-Y results joined with active/inducible genes linked with active histone marks, such as H3K9ac and H3k4me3, but also with inactive genes associated with the repressive histone marks H4K20me3 and H3K27me3. NF-Y appears to act both as a repressor and an activator depending on its interacting cofactors [67]. It is known that the deposition of the negative histone mark H3K27me3 is attributed to Pc and recent experiments with *Caenorhabditis elegans* had genetically linked NF-Y function to Pc through direct interaction with ESC/E(Z) components [68]. In plants, mutations which affect genes for members of two distinct Pc group complexes, PRC1 and PRC2 [69], cause reorganization of heterochromatin domains and influence several developmental processes as well the initiation of ectopic embryos and meristems [70].

One of the most intriguing motifs, present in both the 5′- and 3′-flanking regions as well in the *HaL1L* intron, is WUSATA (Fig. 4; Supplementary Figs. 4, 5, 6). WUS-ATA represents the target sequence for WUS [71], a homeodomain TF with a crucial role in promoting meristem growth [72, 73]. WUS has a decisive role in the embryogenic program, in which it induces the transition from vegetative to embryogenic status and/or maintain the embryonic stem cells identity. *WUS* overexpression appears to repress the *LEC1* gene [19]. The WUSATA motif was first identified in the intron of *AG,* a gene that helps establish floral organ identity [73], and on which *WUS* exerts transcriptional activation. However, in floral meristem, the control of stem cell maintenance depends on the *WUS* repression which requires the direct involvement of *AG* and the recruitment of Polycomb proteins [74].

The *HaL1L* proximal promoter and the *AG* intron both contain the putative binding motif for BELLRINGER (BLR) (Fig. 4), an homeodomain protein that was shown to repress *AG* activity [75]. *WUS*, together with the gene of the floral meristem identity *LEAFY* (*LFY*), as well as *BLR* and *AG*, are hypothesized to be part of a negative feedback loop which acts to suppress floral meristem formation. The presence of motifs binding *WUS*, *LFY* and *BLR* in *HaL1L* intron and 3′- and 5′-flanking regions (Fig. 4; Supplementary Figs. 4, 5, 6)

may indicate that also *HaL1L* is part of the above negative feedback loop.

WUS-binding sites are also found in the promoter sequence of the *Medicago truncatula SOMATIC EMBRYO RELATED FACTOR1* (*MtSERF1*) gene encoding a TF that belongs to the APETALA2 (AP2)/ERF superfamily. *MtSERF1* is expressed in zygotic embryos, somatic embryos and in embryogenic callus. *MtSERF1* is also induced by ethylene. Based on these characteristics, *MtSERF1* appears to be an essential connection between stress and development [76, 77].

Though gene sequence dissection is a powerful tool for the identification of *cis*-regulatory sites, the simple annotation of a transcription factor binding site is not *per se* sufficient to obtain satisfactory functional information. We have shown that a WUSATA target site was able to bind WUS transcription factor (Fig. 5). The WUS target site, out of the six ones identified in *HaL1L* genomic sequence, which we tested showed nucleotide identity with the WUSATA motif in *A. thaliana AG* gene and the same intron location.

Our results provide evidence which supports the hypothesis that *HaL1L* transcription is under epigenetic regulation via cytosine methylation and via putative relationships with a plethora of transcription factors and hormones. We believe that our results might suggest future research projects which will help clarify the complex regulative mechanisms involved in controlling development processes.

References

1. Birnbaum KD, Sánchez Alvarado A (2008) Slicing across kingdoms: regeneration in plants and animals. Cell 132:697–710
2. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125:315–326
3. Cosgrove MS, Boeke JD, Wolberger C (2004) Regulated nucleosome mobility and the hystone code. Nature Struct Mol Biol 11:1037–1043
4. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14:283–291
5. Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239:15–27
6. Gusmaroli G, Tonelli C, Mantovani R (2002) Regulation of novel members of the *Arabidopsis thaliana* of the CCAAT-binding nuclear factor Y subunits. Gene 283:41–48
7. Lee H, Fischer RL, Goldberg RB, Harada JJ (2003) Arabidopsis *LEAFY COTYLEDON1* represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc Natl Acad Sci USA 100:2152–2156
8. Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003) *LEAFY COTYLEDON1-LIKE* defines a class of regulators essential for embryo development. Plant Cell 15:5–18
9. Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. Plant J 54:608–620
10. Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. Trends Plant Sci 13:624–630
11. Suzuki M, Wang HH-Y, McCarty DR (2007) Repression of the *LEAFY COTYLEDON 1/B3* regulatory network in plant embryo development by *VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3* genes. Plant Physiol 143:902–911
12. Ogas J, Kaufmann S, Henderson J, Somerville C (1999) PICKLE is a CHD3 chromatin-remodelling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. Proc Natl Acad Sci USA 96:13839–13844
13. Dean Rider S Jr, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J (2003) Coordinate repression of regulators of embryonic identity by *PICKLE* during germination in *Arabidopsis*. Plant J 35:33–43
14. Henderson JT, Li H-C, Dean Rider S, Mordhorst AP, Romero-Severson J, Cheng J-C, Robey J, Sung R, de Vries SC, Ogas J (2004) *PICKLE* acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. Plant Physiol 134:995–1005
15. Ogas J, Cheng JC, Sung ZR, Somerville C (1997) Cellular differentiation regulated by gibberellin in the *Arabidopsis thaliana pickle* mutant. Science 277:91–94
16. Aichinger E, Villar CBR, Farrona S, Reyes JC, Henning L, Köhler C (2009) CHAD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. PLOS Genet 5(8):e1000605. doi:10.1371/journal.pgen.1000605
17. Asker SE, Jerling L (1992) Apomixis in plants. CRC Press, London
18. Dickinson TA (1978) Epiphylly in angiosperms. Bot Rev 44:181–232
19. Harding EW, Tang WN, Nichols KW, Fernandez DE, Perry SE (2003) Expression and maintenance of embryogenic potential is enhanced through constitutive expression of *AGAMOUS-Like 15*. Plant Physiol 133:653–663
20. Zuo J, Niu Q-W, Frugis G, Chua N-H (2002) The *WUSCHEL* gene promotes vegetative-to-embryonic transition in *Arabidopsis*. Plant J 30:349–359
21. Verdeil J-L, Alemanno L, Niemenak N, Tranbarger TJ (2007) Pluripotent versus totipotent plant stem cells: dependence versus autonomy? Trends Plant Sci 12:245–252
22. Costa S, Shaw P (2007) Open minded cells: how cells can change fate. Trends Cell Biol 17:101–106
23. Lo Schiavo F, Pitto L, Giuliano G, Torti G, Nuti-Ronchi V, Marazziti D, Vergara R, Orselli S, Terzi M (1989) DNA methylation of embryogenic cell cultures and its variation as caused by mutation, differentiation, hormones and hypomethylating drugs. Theor Appl Genet 77:325–331
24. Grafi G (2004) How cells dedifferentiate: a lesson from plants. Dev Biol 268:1–6
25. Yamamoto N, Kobayashi H, Togashi T, Mori Y, Kikuchi K, Kuriyama K, Tokuji Y (2005) Formation of embryogenic cell clumps from carrot epidermal cells is suppressed by 5-azacytidine, a DNA methylation inhibitor. J Plant Physiol 162:47–54
26. Shibukava T, Yazawa K, Kikuchi A, Kamada H (2009) Possible involvement of DNA methylation on expression regulation of carrot *LEC1* gene in its 5′-upstream region. Gene 437:22–31

27. Fambrini M, Durante C, Cionini G, Geri C, Giorgetti L, Michelotti V, Salvini M, Pugliesi C (2006) Characterization of *LEAFY COTYLEDON1-LIKE* gene in *Helianthus annuus* and its relationship with zygotic and somatic embryogenesis. Dev Genes Evol 216:253–264

28. Chiappetta A, Fambrini M, Petrarulo M, Rapparini F, Michelotti V, Bruno L, Greco M, Baraldi R, Salvini M, Pugliesi C, Bitonti MB (2009) Ectopic expression of *LEAFY COTYLEDON1-LIKE* gene and localized auxin accumulation mark embryogenic competence in epiphyllous plants of *Helianthus annuus × H. tuberosus*. Ann Bot 103:735–747

29. Siebert PD, Chenchik A, Kellog DE, Lukyanov KA, Lukyanov SA (1995) An improved method for walking in uncloned genomic DNA. Nucl Acids Res 23:1087–1088

30. Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Ohnhäuser R, Prüß M, Schacherer F, Thiele S, Urbach S (2001) The TRANSFAC system on gene expression regulation. Nucl Acids Res 29:281–283

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

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

33. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

34. Lotan T, Ohto M, Yee KM, West MAL, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) *Arabidopsis LEAFY COTYLEDON1* is sufficient to induce embryo development in vegetative cells. Cell 93:1195–1205

35. Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) *LEAFY COTYLEDON2* encodes a B3 domain transcription factor that induces embryo development. Proc Natl Acad Sci USA 98:11806–11811

36. Carey M, Smale ST (2001) Concepts and strategies: I. promoter and the general transcription machinery. In: *Transcriptional regulation in eukaryotes*, Cold Spring Harbor Laboratory Press, New York, pp 99–115

37. Nakamura M, Tsunoda T, Obokata J (2002) Photosynthesis nuclear genes generally lack TATA-boxes: a tobacco photosystem I gene responds to light through an initiator. Plant J 29:1–10

38. Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics 8:67. doi:10.1186/1471-2164-8-67

39. Vandepoele K, Casneuf T, Van de Peer Y (2006) Identification of novel regulatory modules in dicotyledonous plants using expression data and comparative genomics. Genome Biol 7:R103. doi:10.1186/gb-2006-7-11-r103

40. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21:2933–2942

41. Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T (2009) Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J 58:843–856

42. Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T (2005) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of *FUSCA3* and *ABSCISIC ACID INSENSITIVE3*. Plant Cell Physiol 46:399–406

43. Arnold DR, Françon P, Zhang J, Martin K, Clarke HJ (2008) Stem-loop binding protein expressed in growing oocytes is required for accumulation of mRNAs encoding histones H3 and H4 and for early embryonic development in the mouse. Dev Biol 313:347–358

44. Reverdatto SV, Dutko JA, Chekanova JA, Hamilton DA, Belostotsky DA (2004) mRNA deadenylation by PARN is essential for embryogenesis in higher plants. RNA 10:1200–1214

45. Rothnie HM (1996) Plant mRNA 3'-end formation. Plant Mol Biol 32:43–61

46. Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci USA 99:16491–16498

47. Jenik PD, Barton MK (2005) Surge and destroy: the role of auxin in plant embryogenesis. Development 132:3577–3585

48. Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, Hamann T, Offringa R, Jürgens G (2003) Efflux-dependent auxin gradients establish the apical-basal axis of *Arabidopsis*. Nature 426: 147–153

49. Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P (2004) The transcription factor *FUSCA3* controls developmental timing in *Arabidopsis* through the hormones gibberellin and abscisic acid. Dev Cell 7:373–385

50. Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proc Natl Acad Sci USA 103:3468–3473

51. Casson SA, Lindsey K (2006) The *turnip* mutant of *Arabidopsis* reveals that *LEAFY COTYLEDON1* expression mediates the effects of auxin and sugars to promote embryonic cell identity. Plant Physiol 142:526–541

52. Marcotte WR Jr, Russell SH, Quatrano RS (1989) Abscisic acid-responsive sequences from the *em* gene of wheat. Plant Cell 1:969–976

53. Shen QJ, Zhang P, Ho TH (1996) Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. Plant Cell 8:1107–1119

54. Seo M, Koshiba T (2002) Complex regulation of ABA biosynthesis in plants. Trends Plant Sci 7:41–48

55. Gómez Porras JL, Riaño-Pachón DM, Dreyer I, Mayer JE, Mueller-Roeber B (2007) Genome-wide analysis of ABA-responsive elements ABRE and CE3 reveals divergent patterns in Arabidopsis and rice. BMC Genomics 8:260. doi:10.1186/1471-2164-8-260

56. Narusaka Y, Nakashima K, Shinwari ZK, Sakuma Y, Furihata T, Abe H, Narusaka M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Interaction between two *cis*-acting elements, ABRE and DRE, in ABA dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. Plant J 34:137–148

57. Kagaya Y, Okuda R, Ban A, Toyoshima R, Tsutsumida K, Usui H, Yamamoto A, Hattori T (2005) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in *Arabidopsis*. Plant Cell Physiol 46:300–311

58. Parcy F, Valon C, Kohara A, Misera S, Giraudat J (1997) The *ABSCISIC ACID INSENSITIVE3*, *FUSCA3* and *LEAFY COTYLEDON1* loci act in concert to control multiple aspects of *Arabidopsis* seed development. Plant Cell 9:1265–1277

59. Zhang X, Garreton V, Chua N-H (2005) The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. Genes Dev 19:1532–1543

60. Koch KE (1996) Carbohydrate-modulated gene expression in plants. Annu Rev Plant Physiol Plant Mol Biol 47:509–540

61. Wenkel S, Turck F, Singer K, Gissot L, Le Gourriere J, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell 18:2971–2984

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

63. Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, Marion CM, Hempel FD, Ratcliffe OJ (2008) The nuclear factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*. Planta 228:709–723

64. Kumimoto RW, Zhang Y, Siefers N, Holt BF III (2010) NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*. Plant J 63:379–391

65. Warpeha KM, Upadhyay S, Yeh J, Adamiak J, Hawkins SI, Lapik YR, Anderson MB, Kaufman LS (2007) The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in *Arabidopsis*. Plant Physiol 143:1590–1600

66. Borevitz JO, Maloof JN, Lutes J, Dabi T, Redfern JL, Trainer GT, Werner JD, Asami T, Berry CC, Weigel D, Chory J (2002) Quantitative trait loci controlling light and hormone response in two accessions of *Arabidopsis thaliana*. Genetics 160:683–696

67. Ceribelli M, Dolfini D, Merico D, Gatta R, Viganò AM, Pavesi G, Mantovani R (2008) The histone-like NF-Y is a bifunctional transcription factor. Mol Cell Biol 28:2047–2058

68. Deng H, Sun Y, Zhang Y, Luo X, Hou W, Yan L, Chen Y, Tian E, Han J, Zhang H (2007) Transcription factor NFY globally represses the expression of the *C. elegans* Hox gene *Abdominal-B* homolog *egl-5*. Dev Biol 308:583–592

69. Costa S, Shaw P (2006) Chromatin organization and cell fate switch respond to positional information in *Arabidopsis*. Nature 439:493–496

70. Xu L, Shen WH (2009) Polycomb silencing of *KNOX* genes confines shoot stem cell niches in *Arabidopsis*. Curr Biol 19:82–87

71. Mayer KFX, Schoof H, Haecker A, Lenhard M, Jürgens G, Laux T (1998) Role of *WUSCHEL* in regulating stem cell fate in *Arabidopsis* shoot meristem. Cell 95:805–815

72. Laux T, Mayer KF, Berger J, Jürgens G (1996) The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. Development 122:87–96

73. Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, Simon R, Weigel D (2001) A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. Cell 105:793–803

74. Liu X, Kim YJ, Müller R, Yamul RE, Liu C, Pan Y, Ca X, Goodrich J, Chen X (2011) *AGAMOUS* terminates floral stem cell maintenance in *Arabidopsis* by directly repressing *WUSCHEL* through recruitment of Polycomb group proteins. Plant Cell 23:3654–3670

75. Bao X, Franks RG, Levin JZ, Liu Z (2004) Repression of *AGAMOUS* by *BELLRINGER* in floral and inflorescence meristems. Plant Cell 16:1478–1489

76. Mantiri FR, Kurdyukov S, Lohar DP, Sharopova N, Saeed NA, Wang XD, Vanden Bosch KA, Rose RJ (2008) The transcription factor MtSERF1 of the ERF subfamily identified by transcriptional profiling is required for somatic embryogenesis induced by auxin plus cytokinin in *Medicago truncatula*. Plant Physiol 146:1622–1636

77. Chen S-K, Kurdyukov S, Kereszt A, Wang X-D, Gresshoff PM, Rose RJ (2009) The association of homeobox gene expression with stem cell formation and morphogenesis in cultured *Medicago truncatula*. Planta 230:827–840