Genomics 92 (2008) 115–121

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Maturation of the nodule-specific transcript *MsHSF1c* in *Medicago sativa* may involve interallelic *trans*-splicing

Zhi-shui He [a,b], Hua-song Zou [a], Yan-zhang Wang [a], Jia-bi Zhu [a], Guan-qiao Yu [a,*]

[a] National Laboratory of Plant Molecular Genetics, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
[b] Graduate School of the Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 11 January 2008
Accepted 18 April 2008
Available online 12 June 2008

Keywords:
Allele
Heat-shock transcription factor
Medicago sativa
Root nodule
Single-base variations
Trans-splicing

ABSTRACT

In nonplant species, many heat-shock transcription factors (HSFs) undergo spatiotemporal-specific alternative splicing. However, little is known about the spatiotemporal-specific splicing of HSFs in plants. Previously, we reported that the alfalfa HSF gene *MsHSF1* undergoes multiple alternative splicing events in various tissues. Here, we identified another spliced transcript isoform, *MsHSF1c*, containing a 177-base tandem repeat, and showed that the low-abundance *MsHSF1c* is a nodule-specific transcript of *MsHSF1*. We also found that *MsHSF1* presents multiple alleles with single-base variations and the expression of *MsHSF1* alleles has allele-specific differences in alfalfa nodules. Because single-base variations at position 1006 change the AT of *MsHSF1b* to GT in *MsHSF1b-3*, creating a pair of donor/acceptor sites with the AG of *MsHSF1b*/*1b-1* at position 827–828 for pre-mRNA splicing, we suggest that *MsHSF1c* may be generated by *trans*-splicing between alleles *MsHSF1b-3* and *MsHSF1b* or *MsHSF1b-1*. These results provide new insight into the role of tissue-specific contribution in the transcription of plant HSF genes.

Plant heat-shock transcription factors (HSFs) have complex structures and specific functions. The known plant HSFs have been assigned to three classes (A, B, and C) according to their unique structural characteristics [1]. There are 21 HSFs in *Arabidopsis thaliana*, 23 in rice (*Oryza sativa*), 34 in soybean (*Glycine max*), at least 18 in tomato (*Lycopersicon esculentum*), and many other HSFs in maize (*Zea mays*) and other species [1–4]. Some of the genes encoding these HSFs undergo tissue-specific expression, such as the seed-specific *HSFA9* in sunflower (*Helianthus annuus*) and *Arabidopsis* [5,6], the root-specific *AtHSFA6b* [5], and the pollen-specific *AtHSFA5* (AtGenExpress microarray data: http://www.arabidopsis.org/info/expression/ATGenExpress.jsp). These genes have no known tissue-specific isoforms.

Many HSFs in nonplant species undergo spatiotemporal-specific alternative splicing. Stage-specific alternative splicing of *SmHSF* occurs in *Schistosoma mansoni* [7]. In vertebrates, HSF1 also undergoes tissue-dependent alternative splicing [8]. One zHSF1 isoform is tissue specific in zebrafish (*Danio rerio*) [9]. Previously, we reported that *MsHSF1* undergoes alternative splicing in various alfalfa tissues [10]. To date, spatiotemporal-specific splicing of plant HSFs has not been reported.

Trans-splicing is an important event for mRNA maturation in which separate pre-mRNAs contribute sequences to the mature mRNA. Spliced-leader (SL) *trans*-splicing was discovered in trypanosomes and in nematodes (reviewed in [11]). In SL *trans*-splicing, a short (~15–50 nt), capped 5′ noncoding mRNA as the leader sequence was joined to the mRNA that codes for protein [11,12]. A second type of *trans*-splicing is the "discontinuous group II intron" form of *trans*-splicing found in plant chloroplasts and mitochondria (reviewed in [12]). This type of *trans*-splicing was possibly generated through interactions between "intronic" RNA pieces [12]. Mature mRNAs generated by *trans*-splicing between separate nuclear-encoded pre-mRNAs from protein-coding genes were found extensively in mammals and in *Drosophila* [13–17]. Nuclear-encoded pre-mRNA *trans*-splicing in plants has been reported only by Kawasaki et al. [18].

In this study, we identified the alfalfa nodule-specific transcript *MsHSF1c*, which contains a 177-base tandem repeat. This kind of tandem repeat may be generated by *trans*-splicing between alleles *MsHSF1b-3* and *MsHSF1b* or *MsHSF1b-1*, because the single-base variations at position 1006 change the AT of *MsHSF1b* to GT in *MsHSF1b-3*, creating a pair of donor/acceptor sites with the AG of *MsHSF1b*/*1b-1* at position 827–828. Our results enable further understanding of the transcription of plant genes encoding HSFs.

Results

Molecular characterization of MsHSF1c

Earlier, we reported that the transcript *MsHSF1b* (GenBank Accession No. AY651249), composed of exon 1 and exon 4 of *MsHSF1*, encodes a class-A1 HSF protein [10]. In this work, we identified another cDNA clone, *MsHSF1c*, from alfalfa nodules (GenBank Accession No. DQ907238). Sequence analysis showed that

* Corresponding author. Fax: +86 21 54924015.
E-mail address: gqyu@sippe.ac.cn (G. Yu).

0888-7543/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2008.04.007

MsHSF1c contains a 177-nucleotide (59 amino acid residues) tandem repeat and encodes a protein of 561 amino acid residues. At both the nucleotide and the protein level, *MsHSF1c* shares 100% identity with *MsHSF1b*, except for the tandem repeat (Fig. 1). Thus, the domain structure of MsHSF1c is the same as that of MsHSF1b, and the two proteins have the same DNA-binding domain (DBD), oligomerization domain [carrying two adjacent hydrophobic heptad repeats (HR-A/B)], nuclear localization signal (NLS), AHA motif (a peptide with aromatic and hydrophobic amino acid residues in an acidic surrounding), and nuclear export signal (NES). The 59-residue tandem repeat is located between the NLS and the AHA (Fig. 1). Analysis of the tandem repeat revealed that it contains 13 serine residues (22%). Using NetPhos 2.0 software [19], we predicted that 6 residues are the probable phosphorylation sites among these 13 residues. However, there are only five likely phosphorylation sites throughout the whole nonrepeated region (Fig. 2).

MsHSF1c is a natural transcript of MsHSF1

It is possible that MsHSF1c is encoded by another HSF gene that contains the 177-nucleotide tandem repeat or that genomic DNA contains two tandem-linking alleles that generate *MsHSF1c* by alternative splicing. To rule out these possibilities, genomic PCR was performed with the primers TF1 and TR1 residing outside the tandem

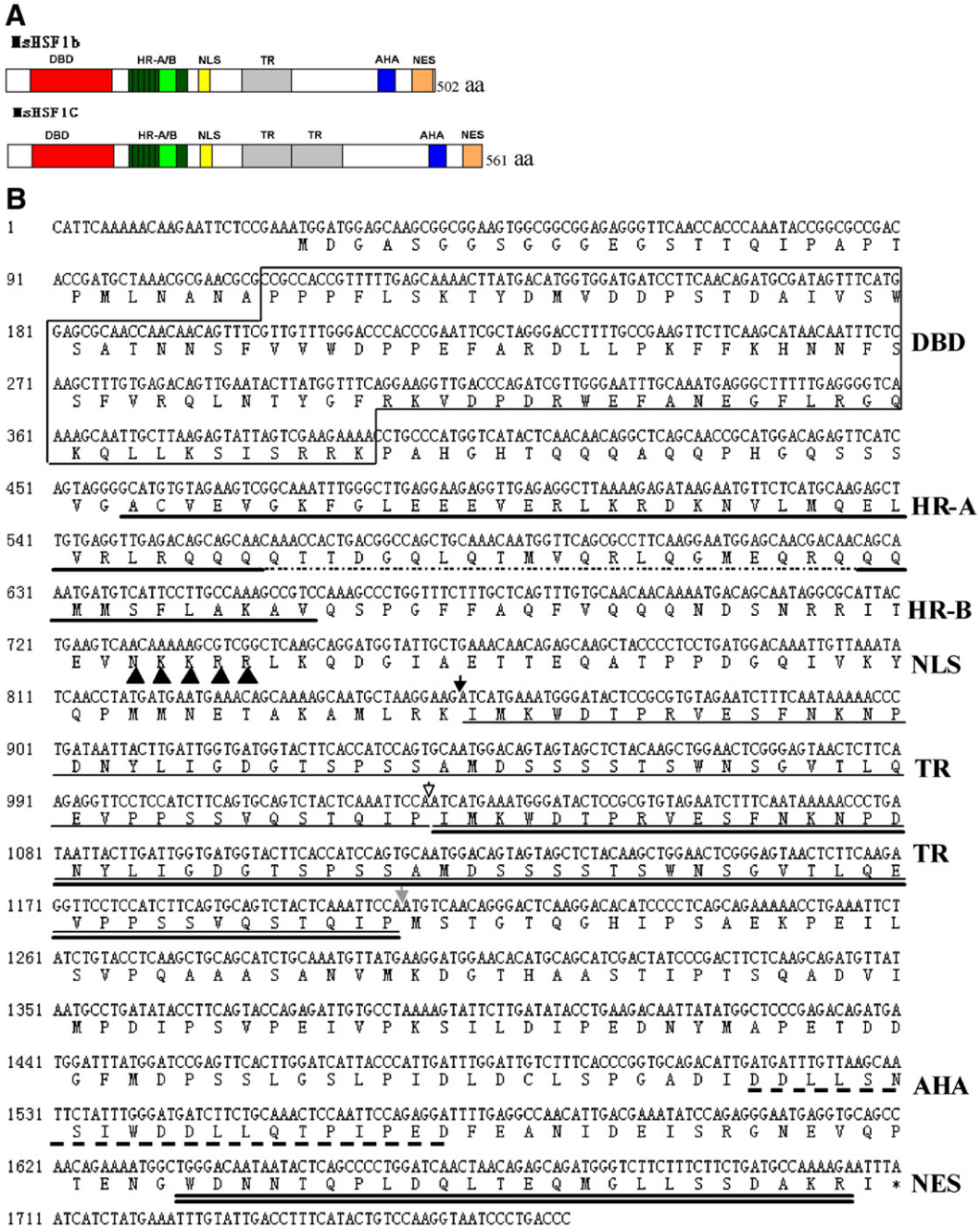

Fig. 1. cDNA and deduced protein sequences of *MsHSF1c*. (A) A representation of the domain structures of MsHSF1b and MsHSF1c. The domains, motifs, and tandem repeat (TR) regions are indicated. (B) The cDNA and deduced protein sequences of *MsHSF1c*. The DBD, HR-A/B, 21 amino acid residues inserted into HR-A and HR-B, NLS, NES, and AHA are boxed, bold single-underlined, dot-underlined, indicated by triangles, broken-underlined, and bold double-underlined, respectively. The region between the initiation and the junction point of TR is single-underlined. The region between the junction point and the termination of TR is unequal double-underlined. Black, white, and gray arrowheads indicate initiation, junction point, and termination positions of TR, respectively.