Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome Wide Analysis and Characterization of Heat Shock Transcription Factors (Hsfs) in French Bean (Phaseolus Vulgaris L.)

B. Mallick
Orissa University of Agriculture and Technology
Meenu kumari (Meenu.Kumari@icar.gov.in)
Indian Institute of Horticultural Research https://orcid.org/0000-0003-0273-2873
S. K. Pradhan
Orissa University of Agriculture and Technology
Parmeswaran C.
Central Rice Research Institute
P. Naresh
Indian Institute of Horticultural Research
G. C. Acharya
Indian Institute of Horticultural Research

Research article

Keywords: PvHSF, heat stress, french bean, Phaseolus vulgaris, evolution

Posted Date: June 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36054/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background: Heat shock transcription factors (*HSF*s) play an important role as transcriptional regulatory proteins against heat stress by controlling the expression of heat responsive genes. French bean is a highly thermosensitive crop and therefore, its genome sequence information segregated, characterized here in terms of heat shock transcription factors alongwith its evolutionary significance.

Results: In this study, a total comprehensive set of 29 non-redundant full length *HSF* genes were identified and characterized from *Phaseolus vulgaris* L*. (PvHSF)* genome sequence*.* Detailed gene information such as chromosomal localization, domain position, motif organization and exon- intron identification were analyzed. All the 29 *PvHSF* genes were mapped on 8 out of 11 chromosomes indicating the gene duplication occurred in french bean genome. Motif analysis and exon- intron structure found to be conserved in each group which showed conserved structure of *PvHSFs* and heat induced response is highly influenced by the cytoplasmic proteins. Based on structural features and phylogenetic relationship, the *HSF* genes were grouped into three classes *i.e.* A to C and 14 groups. Presence of only one pair of paralog sequence suggests that it may be derived from the duplication event during evolution. Comparative genomics study indicated the influence of whole genome duplication and purifying selection on french bean genome during evolution. *In silico* expression analysis indicated active role of classs A and B family during abiotic stress condition and higher expression in floral organs.

Conclusion: A comprehensive analysis of Heat shock transcription factors resulted 29 non-redundant full length *PvHSF* genes, which were characterized for their occurrence in genome and functional importance through *In silico* approach.

Materials & Methods

Plants are the ancient organisms which have been confronted to various biotic and abiotic stress during the course of evolution through regulatory network of inherent adaptation measures. As sessile organisms, plants cannot avoid different biotic and abiotic stresses by changing their positions [1, 2]. Abiotic stress conditions such as transitory or constantly increasing temperature, drought, salinity, high light intensity, exposure to Ozone (O_3), chilling, freezing, exposure of the cell to toxins (ethanol, arsenic, trace metals, UV light) have considerable impact in the plant yield as it can reduce more than 50% productivity in many species. Heat stress is a major stress which causes the protein denaturation affecting the cytoskeleton arrangement of the cell and induces disruption in the metabolism, photosynthesis activity and cell membrane structure. Stressful environmental condition causes morpho-anatomical, physiochemical and biochemical changes at cellular level, which may affect the plant biomass production and economic yield in many areas worldwide. Therefore, major objective of agronomic research is to increase crop productivity under biotic and abiotic stresses.

In recent years, it have been proved that all living organisms possesses a common heat shock response mechanism which initiates the synthesis of “Heat shock proteins” known as “Stress induced proteins”

required for development of short term stress resistance [3, 4]. Among all 64 identified transcription factor groups in plants, Heat shock transcription factor family is ubiquitous and responsible for the expression of the heat shock proteins. Stress induced gene expression leads to the rapid accumulation of heat shock proteins which belong to 11 conserved multi-protein families. As the major HSPs are highly homologous among eukaryotes, the evolutionary conservation of the heat shock response is evidence which justifies that the production of HSPs is a fundamental and imperative process [5].

Characteristic feature of transcription factors is that they contain one or more DNA binding domain which binds to the promoter region for gene expression [6]. In general, *HSF* proteins have a core structure consisting a N- terminal DNA binding domain made up of a four-stranded antiparallel β-sheet and three helical bundle. It is characterized by a central helix-turn-helix (HTH) motif which recognizes and binds to the conserved cis acting elements known as “Heat Shock Element” (HSEs;5-AGAAnnTTCT-3). An adjacent domain commonly known as oligomerization domain with a heptad hydrophobic repeat (HR-A/B) required for the trimerization of *HSF* is connected to the DBD region by a linker group of amino acids. Arginine(R)& Lysine(K) rich NLS region and Leucine (L) rich NES regions are responsible for the dynamic distribution of *HSF* between the nucleus and cytoplasm [7, 8]. It has been seen that in most cases class A and C *HSF*s are having monopartite and bipartite NLSs (nuclear localization signals) adjacent to the oligomerization domain. AHA (Aromatic large hydrophobic amino acid residues) motif or the activation domain is the least defined part of *HSF* which consists of activators and repressor elements. It is a characteristic region of class A *HSF*s as class B and C lacks AHA motif. It is required for the transcription of HSPs, by binding to some basic transcription protein complexes. [9, 10].

In plants, ubiquity of heat shock proteins was first observed in *Glycine max* and Tobacco, using cell culture technique which revealed synthesis of certain proteins under heat condition. To our knowledge, *HSF*s have been investigated in many plats including *Lycopersicon peruviannum* (tomato), *Arabidopsis thaliana, Zea mays* (maize), *Glycine max* (soybean), *Medicago truncatula* (barrel clover), *Lotus japonicus*, *Triticum aestivum* (wheat), *Oryza sative* (rice), *Vigna radiata* (mung bean), *Populus trichocarpa* (black cottonwood). Here, well described 21 Arabidopsis *HSF* gene are used as seed sequence in order to identify and characterize the pool of *HSF* genes present in *Phaseolus vulgaris* (French bean) genome.

Material and methods:

Identification of *HSF*s:

Amino acid sequences corresponding to the 15 *AtHSFs* (*A. thaliana*) genes were retrieved from TAIR (The *Arabidopsis* Information Resource; http://www.*Arabidopsis*.org/. These protein sequences were used as query against database of Phytozome (http://www.phytozome.net)in standalone blast search with e values 0.001. Self-BLAST performed for sequences to remove redundancy. All the protein sequences thus obtained were confirmed with HMMER3.0 (http://hmmer.org/) on the basis of the HMM profile. Results from BLAST and HMMER hits were matched and parsed manually.

Characterization of *HSF* genes, chromosomal localization and protein domains:

To predict the physico-chemical properties, ProtParam tool (www.expacy.org/tools/protparam.html) was used. This tool is useful for computing the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY). The subcellular localization predicted by CELLO for eukaryotic sequences. PROSITE was used for documentation and information about the protein domains, functions, families as well as patterns and profiles (http://prosite.expacy.org/). MacDraw, a Raster based and vector graphic drawing application used for drawing chromosomal map. Gene structure display server (GSDS) was used for illustrating the gene structure.

Conserved domains, motif prediction and promoter region identification:

PSort and NetNES results used for predicting the NLS (Nuclear Localization Signal) and NES (Nuclear Export signal) regions, respectively. Using the amino acid sequences as query, the servers predicted possible regions in the sequence. DBD region and coiled coil regions were identified from the COIL server and HMMER server results. Using MEME suite server, the conserved motifs were predicted in 29 putative identified proteins. Minimum width and maximum width for motifs were decided 10 and 50 respectively in the advance option menu. The highly conserved motifs identified in species using MAST algorithms (http://meme.nbcr.net/).

Class A *HSF* genes in GTF/GFF3 format were analyzed by PLANTCARE database for finding the cis-acting elements. PLACE database was used for understanding the functionality of the promoter regions.

Multiple sequence alignment and phylogeny analysis:

Phylogeny analysis has always been an important factor for understanding the evolutionary relationship among species. Total 100 *HSF*s from *A. thaliana* (21), *G. max* (26), *V. radiate* (24) and *P. vulgaris* (29) were aligned using Clustal Omega program. The alignment file was used for constructing the phylogeny tree using Neighbor Joining algorithm with 1000 bootstrap replications and wrapped in MEGA 7.

Comparative Analysis:

Ortholog and paralog genes are two different types of homologous genes derived from the ancestral DNA by speciation and duplication events, respectively. Paralog sequences were identified by performing similarity search among the *PvHSF* proteins. OrthoVenn2 server was used for prediction of the ortholog genes present in other species. By performing BLAST program among the *HSF* proteins retrieved from different species, ortholog sequences were found in *Vigna radiata, Glycine max, Populous trichocarpa, Medicago truncatula* and *Arabidopsis thaliana*. CoGe database was used for analyzing and visualizing the synteny between *Phaseolus vulgaris* and *Arabidopsis thaliana*.

Expression Analysis:

For expression analysis of identified *PvHSFs*, GEO dataset (Accession: GSE 123381) were retrieved from NCBI. The EST sequence information were retrieved from Phytozome database and used for tissue specific expression analysis retrieved. Heat map visualization and further editing was carried out by using the “gplot” package of R studio.

Results

Identification and characterization of *PvHSFs*:

Two strategies were used to search for members of *HSF*s in French bean. Hidden Markov Model (HMM) and Standalone BLAST+ (version 2.2.23) software package. Sixty-five *HSF* loci were reported from Standalone BLAST against database Phytozome based on e value and identity percentage and score with significant alignments. Further with threshold e value and removal of incomplete, unanchored and redundant sequences, 29 *PvHSF* non-redundant candidate genes were identified which were similar with results of HMM search and characterized for biochemical properties and subcellular localization (**Table 1, S1. PV_sequence**). All the 29 *PvHSF* genes were mapped onto 11 *Phaseolus vulgaris* chromosomes (**Figure 1**). *PvHSF* genes were widely distributed in every chromosome except chromosome V, X and XI. Whereas, Chromosome I and II contains maximum number of genes *i.e.* 6 (20.68%) and 7 (24.13%), respectively. Five *HSFs* were detected on chromosome VII and four on chromosome III apart from single pair on Chromosome IV, VI and IX and chromosome VIII occupied only one *HSF* gene.

The key domain characterized by PROSITE results, which provided information about *HSF*_DOMAIN. The DBD domain located towards the N-terminal of sequence followed by oligomerization domain. PSort resulted absence of NLS region in class B including *PvHSFA8* and *PvHSFA3*. However, NES region is present in all the classes except few members. Through comparison analysis, conserved AHA motif was observed towards the C-terminal only in Class A *HSF*s. Using COILS and HMMER results, identified the coiled coil region, a characteristic of HR-A/B region (**Table 2**). The MEME suite server predicted conserved motifs of *PvHSFs* proteins. Thirty corresponding consensus motifs were predicted (**Fig. 2**). The members of class A contained the most conserved motifs with largest number 11 detected in *PvHSFA1A* and *PvHSFA1D*. Class B members possessed 3-7 motifs while 4 motifs aligned in member of class C family *i.e. PvHSFC1*. Motifs 1, 2 and 4 represent the *HSF* DBD domains found in all 29 *PvHSF*s. Motif 3 represents the HR-A/B domain from the class A and C *HSF*s. All class B *HSF* proteins exhibited the coiled-coil region of HR-A/B domain by motif 5. Motif 6 represents HR-A/B domain from class A *HSF*s. Motif 7 conserved for NLS domains in class A while motif 8 conserved for NLS of B class families. Furthermore, motifs 10, 11, 14, 16, 17, 18, 23, 26, 28 and 30 represented NES domains with (L) leucine rich. AHA motifs identified by 9, 15, 16, 20, 21, 22, 24 and 29 with W (tryptophan) rich domain.

Gene structure and phylogeny analysis:

In this study, the structure of *HSF* gene with one intron found to be mostly conserved throughout subfamilies among 72.41% of *PvHSF* members (**Fig. 3**). However, length and location of intron varies among the *HSF*s. Noticeably, two introns were observed among 24.13% of *HSF*s genes of which all

belongs to class A subfamilies (*PvHSFA1B, PvHSFA2A, PvHSFA4C, PvHSFA7A, PvHSFA8, PvHSFA9A*) except *PvHSFB2C* of class B. *PvHSFA2B* comprised of 3 introns with comparative smaller length.

The phylogenetic tree was constructed using MEGA 7 software with Neighbor-Joining method using multiple sequence alignment of 2 leguminosae members (*Glycine max; GmHSF*s and *Vigna radiata; VrHSFs*), 21 *AtHSF* and 29 *PvHSF* genes which categorized the 29 *PvHSFs* into 3 classes (A, B and C) and 14 groups (**Fig. 4**). Class A was the largest and contained 17 *PvHSF* which were again divided into 9 groups. Class B contained 11 members and Class C was the smallest consisting only one gene.

Regulatory cis acting elements

Sixty-four different promoter regions responsible for various expression identified among members (Table **S2**). The *PvHSFA5.2* contented maximum number of promoter regions while *PvHSFA4C* identified for least. CAAT-box and MYC promoter region is present in all the class A *HSF*s gene while MYB promoter region is present in all the members except *PvHSFA4* sub-group. ABRE (ABA responsive element) promoter region carried by all members except *PvHSFA8* and *A6* group. DRE region was only seen in *PvHSFA4A* and LTRE was observed in *PvHSFA5.2* and *PvHSFA9A*. P-box, W- box, box S, G-box and I-box were observed in members of class A. Light responsive element, 3-af1 binding site was only observed in *PvHSFA1B.*

Comparative genomics

For synteny analysis, 11 *HSF* genes from 4 groups (A1, A2, A6, B2) were compared with their orthologs sequences present in *Arabidopsis Thaliana*. The synteny analysis (**Fig. 6**) revealed that *PvHSFA2B, PvHSFA6A*, *PvHSFB2D* are in synteny with their orthologs though microsynteny can be observed in maximum sequences. However, perfect synteny was observed in *PvHSFB2D*. For *PvHSFA2A* and *A6B*, the synteny was observed in between 100kb-110kb and 85kb-185kb, respectively. Least synteny was observed in *PvHSFA1B/E*, in which the regions from *AtHSFA1E* (70kb-130kb) distributed over 110 kb region (20kb-190kb). In *PvHSFB2C*, the widely distributed genes present in *AtHSFB2B* were conserved in 85kb to 100kb. From the comparative analysis, we found 22 orthologous clusters between *HSF*s of six plant spp. (**Fig. 7**). Only one pair of paralog sequences were found in french bean *HSF* family.

In silico expression analysis

The expression data showed 19 *HSFs* out of 29 *PvHSF* genes were upregulated under drought stress and highest expression of *PvHSFB2A*, indicating its master role under abiotic stress (**Fig. 8a)**. The ubiquitous expression of *PvHSFB2A* throughout plant organs from ESTs dataset again supporting its key importance in gene regulatory network of french bean (**Fig. 8b**). Additionally, the relative expression of *PvHSFA6B* and *PvHSFB2D* were identified for having no or least expression in plant tissues.

Discussion

Identification and characterization of *HSF* genes in French bean:

The *HSF* genes are ubiquitous and their numbers varies in organisms. However, number of *HSF*s in plants outnumbers other organisms. Although from different studies in plants, oftenly it counts from 16 to 82 (16; strawberry and 82; wheat; **S3**). Cytoplasmic occurrence of *PvHSFs* are similar to other *HSF*s like *GmHSFs*, *LpHSFs* [9, 11]. The 473 Mb of 587 Mb french bean genome has been sequenced and assembled [12]. This crop is absolutely diploid (2n= 2x=11), much smaller than soybean genome (1115 Mb) and significantly higher than arabidopsis (125 Mb). Subsequently, the number of protein coding genes significantly varies among arabidopsis (25kb), soybean (46.43kb) and french bean (31.64 kb). In above mentioned species, number of *HSF*s lying very close from 21 to 29 (21; arabidopsis, 26; soybean and 29; french bean) likely to cereal crop rice 25 (genome size: 389 Mb). Howsoever, 29 *PvHSFs* reported in this study is adjacent to other previous reported leguminoseae members *i.e.* 24 *HSFs* in *Vigna radiata* [3], 20 *HSFs* in *Populous trichofera* and 16 *HSFs* in *Medicago truncatula* [13]. These results indicate that difference in copies of *HSF*s among these plant species is independent of genome size or protein coding genes and conserved in plant species of family with moderate evolutionary events like whole genome duplication (WGD) followed by gene loss and purifying selection [14, 15]. Considering the total chromosome number, it can be observed that french bean has moderate number of *HSF* members being a member of legume family.

Gene features such as exon-intron distribution, gene length and GC content have acute impacts during evolutionary events like WGD. Various studies also indicated that introns are responsible not only for gene expression but also for gene evolution [16]. Similar to this, only one intron gene feature also reported from mung bean[3] and most of the *HSF* members of rosaceae family [17]. However, number of introns ranged from zero to two in *Brassica oleracca* [18], and zero to five in *Brassica napus* [19]. Though, at least one member of class C in *Brassica napus* and *Brassica oleracca,* lacks intron. These results indicating that during expansion and divergence of leguminosae *HSF*s, intron losses escaped and consequently presence of at least one intron and absence of intron lacking are characteristics of leguminoseae *HSF* gene family.

Invariably, all the plant *HSF*s have 3 classes in *HSF* family. Observing the phylogeny tree, french bean has comparatively higher number of members in each class of family except class C. For instance, 17 *PvHSFs* with respect to 13 *VrHSFs* while *GmHSFs* consists 19 members in this class. In other legumes, like *Populus* and *Medicago* consisted 15 and 10 members in Class A, respectively [9, 11]. However, 11 members of class B *PvHSFs* is quite conserved among family which ranged from 6-12 in above discussed legume species. Complete conservation of class C *PvHSFs* with only one gene is observed likely to Arabidopsis and legumes. This report is presenting the considerable evidence of tightly conserved Class C and moderate conservation of Class B members, however evolutionary events most frequently observed among Class A members. The phylogeny tree and distance matrix indicated that *PvHSFs* gene family more closely related to the *VrHSFs* with respect to *GmHSFs*.

Considering the importance of class A *HSF* subfamilies with conserved features which were observed to play important role during heat stress and therefore here *PvHSFs* members from Class A only analyzed through PLANTCARE database for understanding the *cis-acting* elements (**Fig. 5**). CAAT-box and MYC

were found to be dominant elements present in all members of class A along with an unknown promoter region. MYB, STRE, LTRE, ABRE, DRE, MYC were some abiotic stress cis acting elements, found in class A *PvHSFs*. The above mentioned regions were reported to be important for heat shock response in plants. For instance, STRE was the first discovered cis acting element to be responsive to stress and provides binding site for *HSFA1A* in *Arabidopsis Thaliana* and deletion of STRE (Stress Responsive Elements) highly affects the promoter activity under stress condition. [20, 21]. In mung bean, all the genes were reported for having multiple DREs (Drought Responsive Elements) and ABREs. *PvHSFA4A* contained DRE1 responsible for dehydration response. DRE combining with DRE binding proteins (DREB 2A) involved in transcriptional regulation of stress in plants such as drought, salt and low temperature responses [22, 23]. LTR was observed in *PvHSFA5.2* and *PvHSFA9A* which illustrates the low temperature response provided under stress condition [24]. Absence of LTRE in maximum members suggests lower response to low temperature in comparison to other plants. *HSF*a1a was reported before to be the master regulator for thermotolerance in tomato [25]. In *Arabidopsis,* it was reported that *HSF*A1a/b are responsible for the induction of class B genes under heat stress condition. [26, 27]. ABA responsive elements are also responsive to ABA and GA responses, *via* combination with ABRE binding proteins. In *PvHSFA1A* many ABA responsive elements like MYC and MYB were observed. Besides this, W box and box S, observed among several members responsible for wound responsive elements. Therefore, cis-element analysis of identified *PvHSF* members presents evidence of these genes for differential gene regulation during abiotic stress mostly thermotolerance and involvement in other regulatory functions like light responsive, wound responsive etc.

Synteny analysis of *HSF*s was performed between *Phaseolus* and *Arabidopsis* for the four different *HSF*s groups namely A1, A2, A6, and B2. These four *HSF*s groups showed highly conserved induced expression during heat stress in multiple crops such as rice, wheat, chickpea etc [1, 6, 14, 27]. Further, *HSF*A1a and *HSF*A2 was characterized as master regulator of basal and acquired thermotolerance, respectively in *Arabidopsis*, rice, tomato and other members of class A (A2 and A6) along with B2 functions as synergistic co activator. Therefore, considering their importance during heat stress response [9], a detailed analysis was conducted only for the members of these four *HSF*s groups (*HSF*A1, A2, A6 and B2) in our analysis. The synteny analysis (**Fig.7**) revealed that *Arabidopsis HSF*s and its orthologs in *Phaseolus* (*PvHSF*A2B, *PvHSF*A6A, *PvHSF*B2D) were in synteny and showed high degree of sequence conservation.

Specifically, perfect synteny was observed between Arabidopsis B2D and *PvHSF*B2D. However, *PvHSF*A2A and A6B showed synteny for the distance of 100kb-110kb and 85kb-185kb, respectively between *Arabidopsis* and *Phaseolus*. Besides, least synteny was observed for the *HSF*s namely *PvHSF*A1B/E, in which the regions from At*HSF*A1E (70kb-130kb) was distributed over 110 kb region (20kb-190kb). Further, *PvHSF*B2C, the widely distributed genes present in At*HSF*B2B were conserved in 85kb to 100kb. These findings in *Phaseolus* suggested that, during the course of evolution *PvHSF*A2B, *PvHSF*A6A and *PvHSF*B2D had conserved their gene structure in comparison with *Arabidopsis* indicating that the diversification might have took place due to speciation or duplication event. However, *PvHSF*B2C showed highly conserved synteny with 15 kn region. Additionaly, maximum members of *Phaseolus HSF*s

showed microsynteny which can be correlated to the whole-genome duplication, gene loss or mutation [18, 20, 28].

Expressions of *PvHSF* genes under abiotic stress (drought) condition and tissue specific expression were illustrated with the objective of validation of putatively identified genes. The digital expression data (RPKM) from GEO datasets were retrieved and expressed in heatmap (**Fig. 8a**). This analysis shows that both class A and B family is directly involved in signaling pathways during abiotic stress. Also it may be inferred from this dataset that B4 group (a, b, c and d) of class B and A1 group (b and d), A3, A5.1 and A8 of class A alongwith *PvHSFC*1were upregulated during drought stress with wide variation. Likely to this class A and B family members played key role during heat stress driven experiments for instance, *LpHSFA1* was reported to be the master regulator for thermotolerance in tomato [25]. Also two soybean *HSF*s from class A2 (*GmHSF*12, *GmHSF* 28) and 3 from class B (*GmHSF* 34, *GmHSF* 35 and *GmHSF* 47) were upregulated during stress condition [11]. Therefore, class A and B family members activation in signaling pathway during abiotic stress is established for *PvHSF* genes also. Ubiquitous expression of *PvHSFB2A* in all plant organs and highest even downregulated expression during abiotic stress stipulated its prime importance among all identified *PvHSFs* (**Fig. 8b**). However *PvHSF*B2D, B4C and D, *PvHSF*A6A and B had no expression in 1 to 4 organelles which is likely to *PbHSFA6C* of Chinese white pear where nil expression observed during its four developmental stages [17]. In french bean, *PvHSF*C1 has showed considerable expression in many samples and slight upreguation during drought stress showed its functional importance, but information is still incomplete.

Conclusion:

In this study, a comprehensive set of 29 full length Heat shock factor genes (*PvHSFs*) were identified and characterized in *Phaseolus vulgaris* with additional gene information such as chromosomal localization and exon intron identification. All the genes were mapped on 8 out of 11 chromosomes indicating the gene duplication occurred in french bean genome. Motif analysis and exon intron structure found to be sustained in each group which concludes that structure of *HSF* family is conserved in french bean. It can be inferred that, heat shock response is highly influenced by the cytoplasmic proteins in french bean. According to the structural characteristics of the proteins and phylogeny analysis, the *HSF* genes were grouped into three classes (A, B and C) and 14 groups. Presence of only one pair of paralog sequence suggests that it may be derived from the duplication event during evolution. From the synteny analysis, it could be referred that whole genome duplication and purifying selection have highly influenced the french bean genome. Expression analysis validated class A and B *HSF*s involvement during abiotic stress and key importance of *PvHSFB2A* due to their comparative higher occurrence under abiotic stress and ubiquitous presence in floral organs. This comprehensive analysis establishes better understanding of *PvHSF* genes which will facilitate further functional characterization in french bean.

Declarations

Contributors: The research topic was conceived and manuscript preparation including analysis was carried out by B. Mallick and Meenu Kumari, result interpretation was contributed by SK Pradhan and Parmeswaram C., editing and software scripts were supported by P Naresh and G C Acharya. All authors approved this manuscript.

Compliance with ethical standards: This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest: The authors declare that they have no conflict of interest.

Funding Source: Not applicable.

Acknowledgement: Authors acknowledge Director, ICAR- IIHR and CHES, Bhubaneswar for providing basic facilities.

References:

1. Duan S, Liu B, Zhang Y, Li G, Guo X. Genome-wide identification and abiotic stress-responsive pattern of heat shock transcription factor family in Triticum. 2019;:1–20.
2. Zhu JK. Abiotic Stress Signaling and Responses in Plants. Cell. 2016;167:313–24. doi:10.1016/j.cell.2016.08.029.
3. Li S, Wang R, Jin H, Ding Y, Cai C. Molecular characterization and expression profile analysis of heat shock transcription factors in mungbean. Front Genet. 2019;10 JAN:1–13.
4. Al-Whaibi MH. Plant heat-shock proteins: A mini review. J King Saud Univ - Sci. 2011;23:139–50. doi:10.1016/j.jksus.2010.06.022.
5. Vierling E. THE ROLES OF HEAT SHOCK PROTEINS IN PLANTS. 1991.
6. Scharf KD, Berberich T, Ebersberger I, Nover L. The plant heat stress transcription factor (*HSF*) family: Structure, function and evolution. Biochim Biophys Acta - Gene Regul Mech. 2012;1819:104–19. doi:10.1016/j.bbagrm.2011.10.002.
7. Kotak S, Larkindale J, Lee U, von Koskull-Döring P, Vierling E, Scharf KD. Complexity of the heat stress response in plants. Curr Opin Plant Biol. 2007;10:310–6.
8. Dirk G, Kutay U. Transport between the cell nucleus and the cytoplasm. 1999.
9. Nover L, Bharti K, Döring P, Mishra SK, Ganguli A, Scharf K-D. Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones. 2001;6:177.
10. Soares-Cavalcanti NM, Belarmino LC, Kido EA, Pandolfi V, Marcelino-Guimarães FC, Rodrigues FA, et al. Overall picture of expressed heat shock factors in Glycine max, lotus japonicusand medicago truncatula. Genet Mol Biol. 2012;35 SUPPL.1:247–59.
11. Chung E, Kim KM, Lee JH. Genome-Wide Analysis and Molecular Characterization of Heat Shock Transcription Factor Family in Glycine max. J Genet Genomics. 2013;40:127–35.

doi:10.1016/j.jgg.2012.12.002.

12. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46:707–13. doi:10.1038/ng.3008.
13. Wang F, Dong Q, Jiang H, Rt-pcr HÁ. Genome-wide analysis of the heat shock transcription factors in Populus trichocarpa and Medicago truncatula. 2012;:1877–86.
14. Lin Y, Cheng Y, Jin J, Jin X, Jiang H, Yan H, et al. Genome duplication and gene loss affect the evolution of heat shock transcription factor genes in legumes. PLoS One. 2014;9.
15. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, et al. Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. Science (80-). 2014;345:950–3.
16. Rose AB. Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol. 2008;326:277–90.
17. Qiao X, Li M, Li L, Yin H, Wu J, Zhang S. Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (Pyrus bretschneideri) and five other Rosaceae species. 2015;:1–16.
18. Lohani N, Golicz AA, Singh MB, Bhalla PL. Genome-wide analysis of the *HSF* gene family in Brassica oleracea and a comparative analysis of the *HSF* gene family in B. oleracea, B. rapa and B. napus. Funct Integr Genomics. 2019;19:515–31.
19. Zhu X, Huang C, Zhang L, Liu H, Yu J, Hu Z, et al. Systematic Analysis of *HSF* Family Genes in the Brassica napus Genome Reveals Novel Responses to Heat, Drought and High CO2 Stresses. Front Plant Sci. 2017;8 July:1–15.
20. Guo M, Lu JP, Zhai YF, Chai WG, Gong ZH, Lu MH. Genome-wide analysis, expression profile of heat shock factor gene family (Ca*HSF*s) and characterisation of Ca*HSF*A2 in pepper (Capsicum annuum L.). BMC Plant Biol. 2015;15:1–20. doi:10.1186/s12870-015-0512-7.
21. Haralampidis K, Milioni D, Rigas S, Hatzopoulos P. Combinatorial interaction of cis elements specifies the expression of the Arabidopsis AtHsp90-1 gene. Plant Physiol. 2002;129:1138–49.
22. Ahuja I, de Vos RCH, Bones AM, Hall RD. Plant molecular stress responses face climate change. Trends Plant Sci. 2010;15:664–74. doi:10.1016/j.tplants.2010.08.002.
23. Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional Regulatory Network of Plant Heat Stress Response. Trends Plant Sci. 2017;22:53–65. doi:10.1016/j.tplants.2016.08.015.
24. Maestrini P, Cavallini A, Rizzo M, Giordani T, Bernardi R, Durante M, et al. Isolation and expression analysis of low temperature-induced genes in white poplar (Populus alba). J Plant Physiol. 2009;166:1544–56. doi:10.1016/j.jplph.2009.03.014.
25. Mishra SK, Tripp J, Winkelhaus S, Tschiersch B, Theres K, Nover L, et al. In the complex family of heat stress transcription factors, *HSF*A1 has a unique role as master regulator of thermotolerance in tomato. Genes Dev. 2002;16:1555–67.

26. Lohmann C, Eggers-Schumacher G, Wunderlich M, Schöffl F. Two different heat shock transcription factors regulate immediate early expression of stress genes in Arabidopsis. Mol Genet Genomics. 2004;271:11–21.
27. von Koskull-Döring P, Scharf KD, Nover L. The diversity of plant heat stress transcription factors. Trends Plant Sci. 2007;12:452–7.
28. Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, et al. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet. 2013;45:487–94.

Tables

Table 1: PvHsf gene information and physico-chemical characters

Sr. No.	Name	Gene ID	Subcellular localization	ORF (AA)	MW (KDA)	pI	CDs	GRAVY
1	PvHSFA1A	Phvul.002G138500.1	PP	490	54.23	5.02	3506	-0.624
2	PvHSFA1B/E	Phvul.002G040500.1	OM,EC	465	51.72	5.26	2753	-0.589
3	PvHSFA1D	Phvul.003G283300.1	PP,EC	474	52.77	4.90	3325	-0.682
4	PvHSFA2A	Phvul.001G037000.1	PP,CP,EC	373	42.46	4.91	1873	-0.613
5	PvHSFA2B	Phvul.009G078300.1	CP,OM	373	41.68	4.90	1326	-0.523
6	PvHSFA3	Phvul.007G223400.1	OM,CP,EC	452	51.27	4.83	2674	-0.647
7	PvHSFA4A	Phvul.002G242000.1	OM	403	45.80	5.14	1340	-0.687
8	PvHSFA4C	Phvul.006G171700.1	PP,CP	386	43.87	5.01	1300	-0.698
9	PvHSFA5.1	Phvul.001G266300.1	EC	441	49.82	5.61	1400	-0.585
10	PvHSFA5.2	Phvul.006G034200.1	EC	487	54.53	5.44	4479	-0.726
11	PvHSFA6A	Phvul.007G278200.1	CP	334	38.66	4.82	1721	-0.755
12	PvHSFA6B	Phvul.007G061800.1	CP	353	40.99	5.29	1571	-0.657
13	PvHSFA7A	Phvul.001G154700.1	CP	366	42.24	5.54	1791	-0.878
14	PvHSFA7B	Phvul.007G174800.2	CP	357	41.49	5.33	1417	-0.723
15	PvHSFA8	Phvul.002G322700.1	CP	360	41.20	4.90	2136	-0.623
16	PvHSFA9A	Phvul.003G135200.1	CP	404	46.42	5.83	1711	-0.711
17	PvHSFA9B	Phvul.003G135400.1	CP	404	46.30	5.96	1692	-0.742
18	PvHSFB1A	Phvul.003G244000.1	EC	278	30.72	5.92	2580	-0.697
19	PvHSFB1B	Phvul.002G019100.1	EC,OM	284	31.52	8.44	1255	-0.729
20	PvHSFB2A	Phvul.004G119600.1	PP,EC	340	37.15	5.62	1128	-0.570
21	PvHSFB2B	Phvul.002G155300.1	OM,PP	356	38.91	4.93	1147	-0.499
22	PvHSFB2C	Phvul.007G067800.1	OM	304	34.04	7.02	998	-0.580
23	PvHSFB2D	Phvul.002G228400.1	CP	206	23.90	7.61	1718	-0.701
24	PvHSFB3	Phvul.001G131000.1	OM	232	26.58	9.08	1901	-0.857
25	PvHSFB4A	Phvul.009G065800.1	EC	364	40.83	8.44	1191	-0.602
26	PvHSFB4B	Phvul.001G022900.1	EC	339	38.84	7.31	1569	-0.615
27	PvHSFB4C	Phvul.008G202500.1	8	PP,CP	269	31.20	7.22	1149
28	PvHSFB4D	Phvul.006G015700.1	6	PP,OM,CP	269	31.48	6.50	1305
29	PvHSFC1	Phvul.004G163300.1	4	EC	317	35.40	5.95	1362

(PP- Periplasmic, CP- cytoplasmic, OM- outer membrane, EC- Extra cellular, MW- Molecular weight, GRAVY- Grand Average of hydropathy)

Table 2 Characteristics domains of PvHsfs

No	HSF	NLS	NES	AHA motif
1	PvHSFA1A	(224) NKKRRLKR	nd	(432) PSIWDEILQT
2	PvHSFA1B/E	(216) KKRRLHR	nd	(400) DEFWELFFRP
3	PvHSFA1D	(231) NKKRRLK	nd	(416) PNFWDDIVRT
4	PvHSFA2A	(241) NRKRR	nd	(313) DTILEDFFIK
5	PvHSFA2B	(240) RKRR	nd	(313) DAIWEDLLNQD
6	PvHSFA3	Nd	(375) L-L	(410) SDIWDIGSGSL
7	PvHSFA4A	(204) RKRRLPR	(248) L-Q	(339) DVFWEQFLTE
8	PvHSFA4C	(206) RKRRLPR	nd	(336) DIFWERFLTE
9	PvHSFA5.1	(204) NKKRRLPP	nd	(389) DVFWEQFLTE
10	PvHSFA5.2	(211) NKKRRLP	nd	(435) DVFWEQFLTE
11	PvHSFA6A	(229) NKRRR	(268) L	(294) EVLLEELLNEG
12	PvHSFA6B	(240) KKRR	(287) LDLGLNL	(319) EVFWQDLIKED
13	PvHSFA7A	(243) KKRRRP	(291) L-M	(322) EGFWEELFSE
14	PvHSFA7B	(234) KKRR	(332) I	(314) EEFWEELLISE
15	PvHSFA8	Nd	(173) LQ	(300) DGSWEQLFLG
16	PvHSFA9A	(263) KRPR	(198) I--L	nd
17	PvHSFA9B	(263) KRPR	(198) I--L	nd
18	PvHSFB1A	Nd	nd	nd
19	PvHSFB1B	Nd	nd	nd
20	PvHSFB2A	Nd	(199) L	nd
21	PvHSFB2B	Nd	nd	nd
22	PvHSFB2C	Nd	(184) L-KELE---SL	nd
23	PvHSFB2D	Nd	(270) FEPLNL	nd
24	PvHSFB3	Nd	(199) LFGVRL-V	nd
25	PvHSFB4A	Nd	(349) LDKDDLGL-L	nd
26	PvHSFB4B	Nd	nd	nd
27	PvHSFB4C	Nd	(175) LL	nd
28	PvHSFB4D	Nd	nd	nd
29	PvHSFC1	(197) KKRR	(203) I	nd

nd: not detected by sequence similarity search

Figures

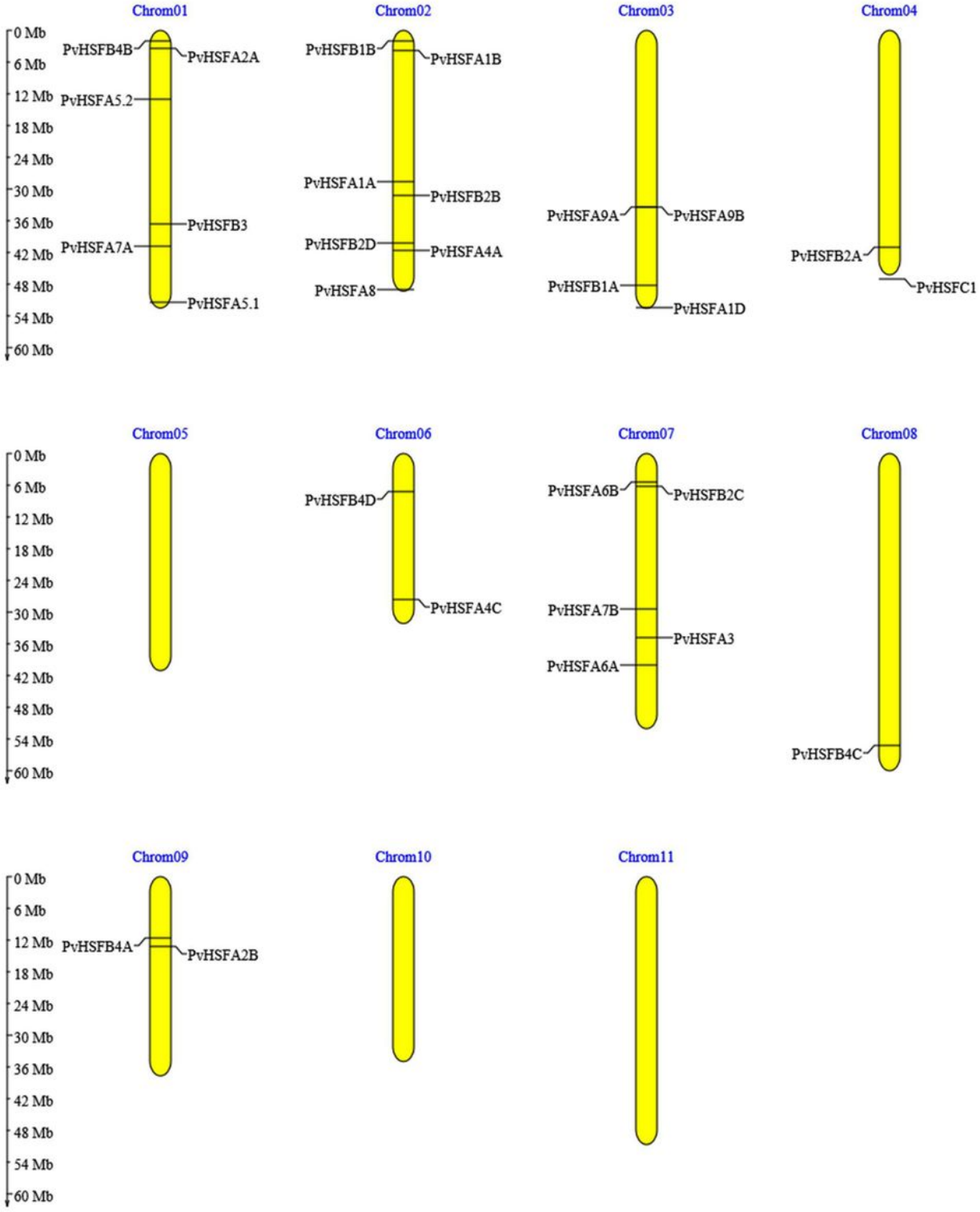

Figure 1

Chromosomal distribution of PvHSFs

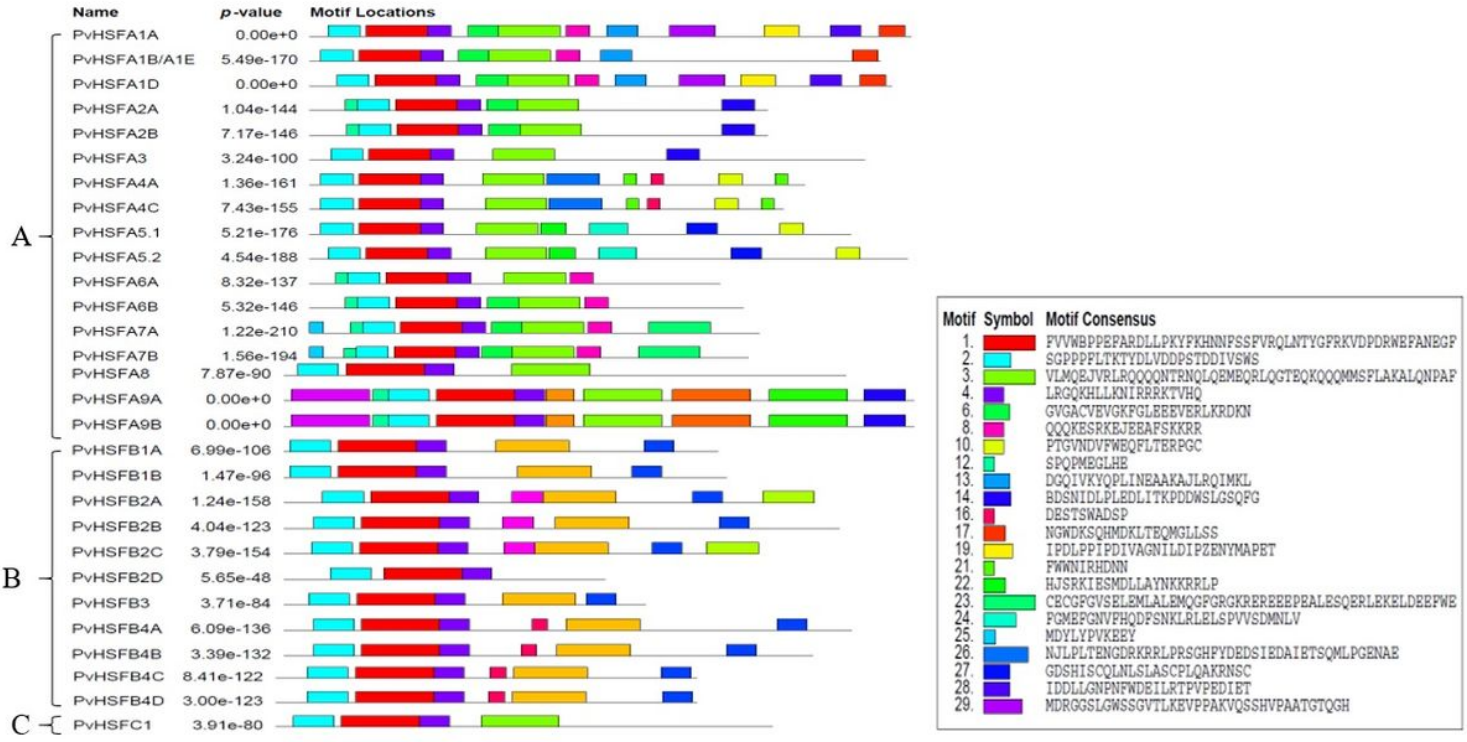

Figure 2

Distribution of motifs in PvHSF family members

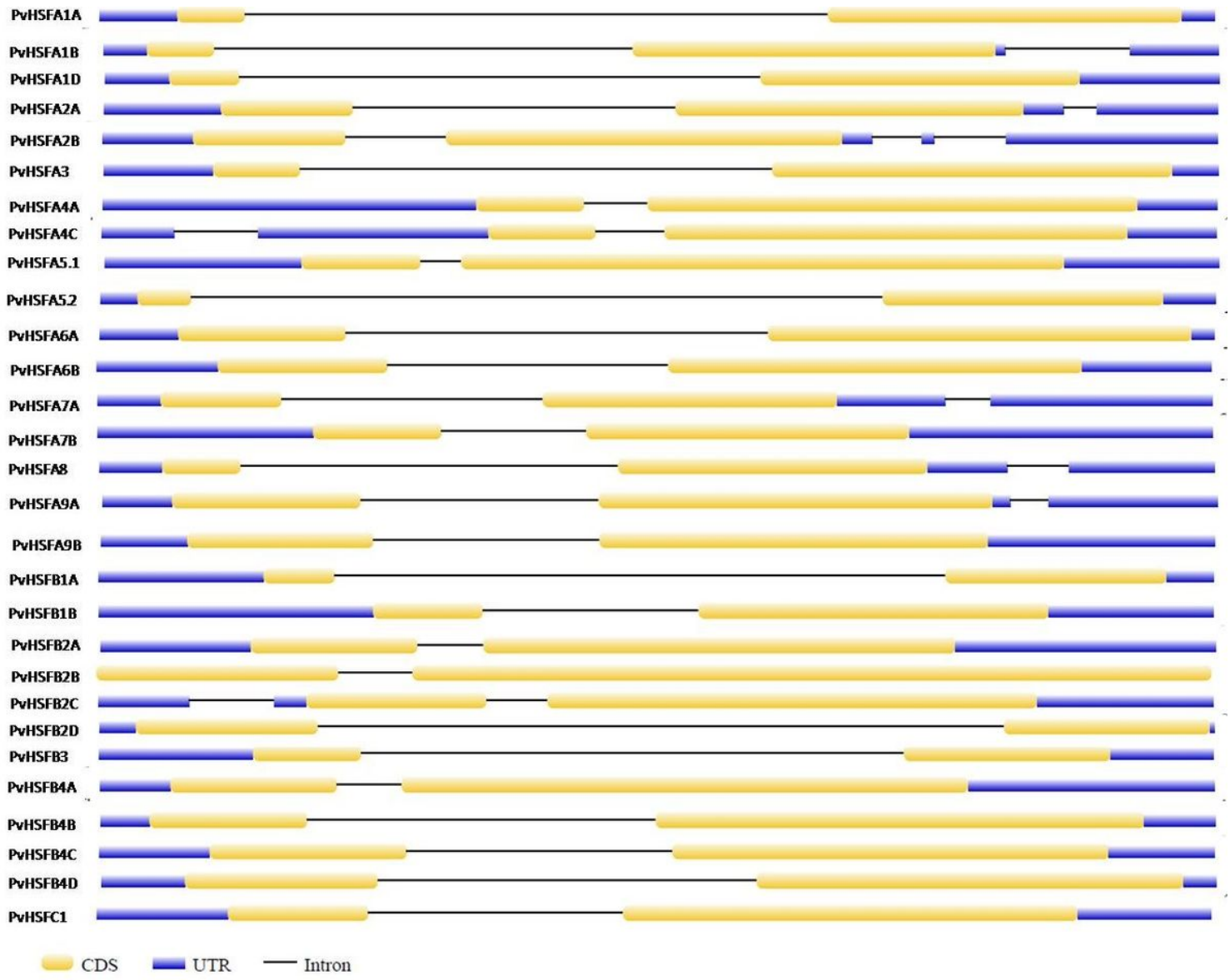

Figure 3

Gene structure of PvHSFs from GSDS server

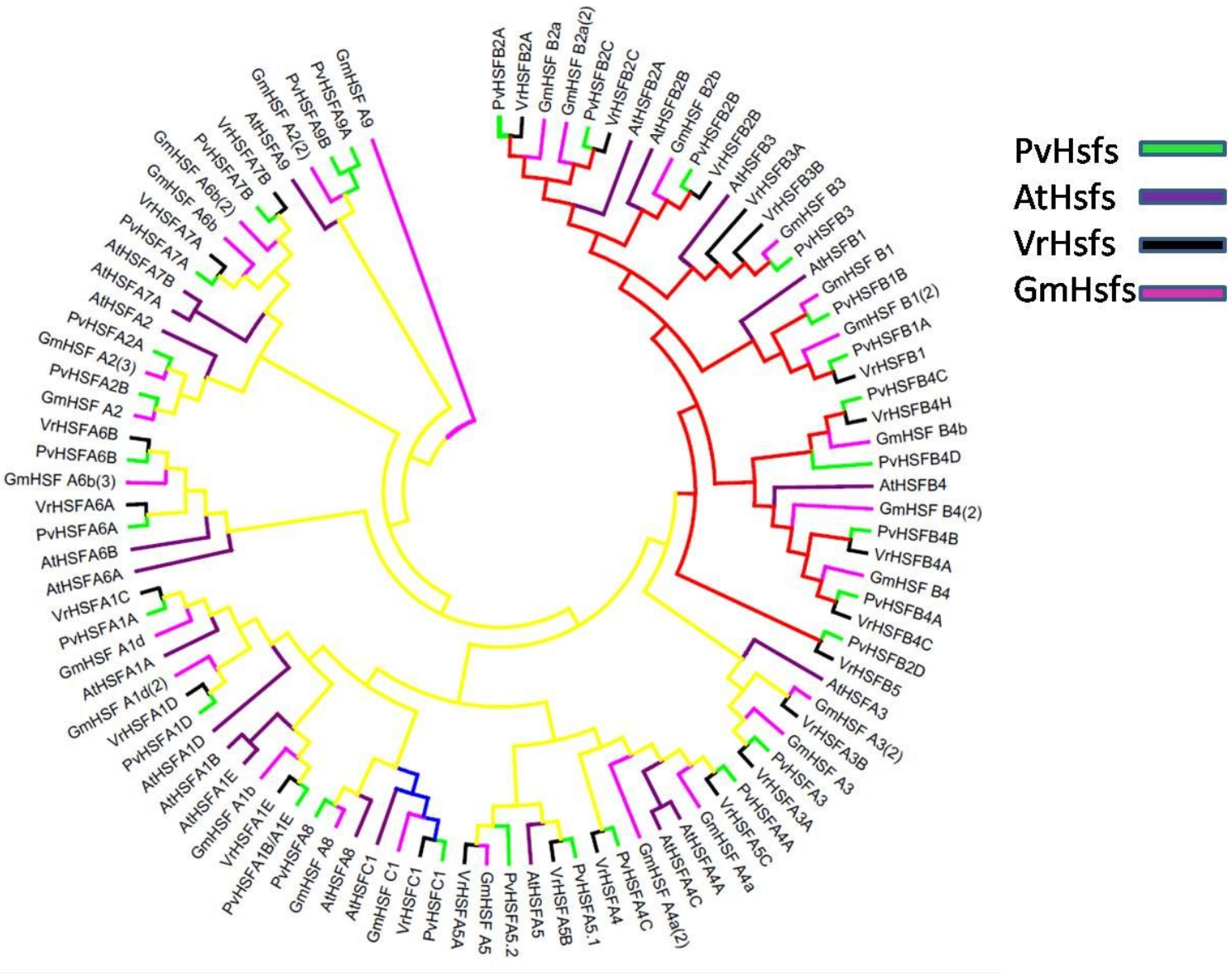

Figure 4

Phylogeny analysis.of HSF members from Phaseolus vulgaris L., Arabidopsis thaliana, Vigna radiata, and Glycine max L.. Different class of families denoted as: Class A: Yellow, Class B: Red, Class C: Blue

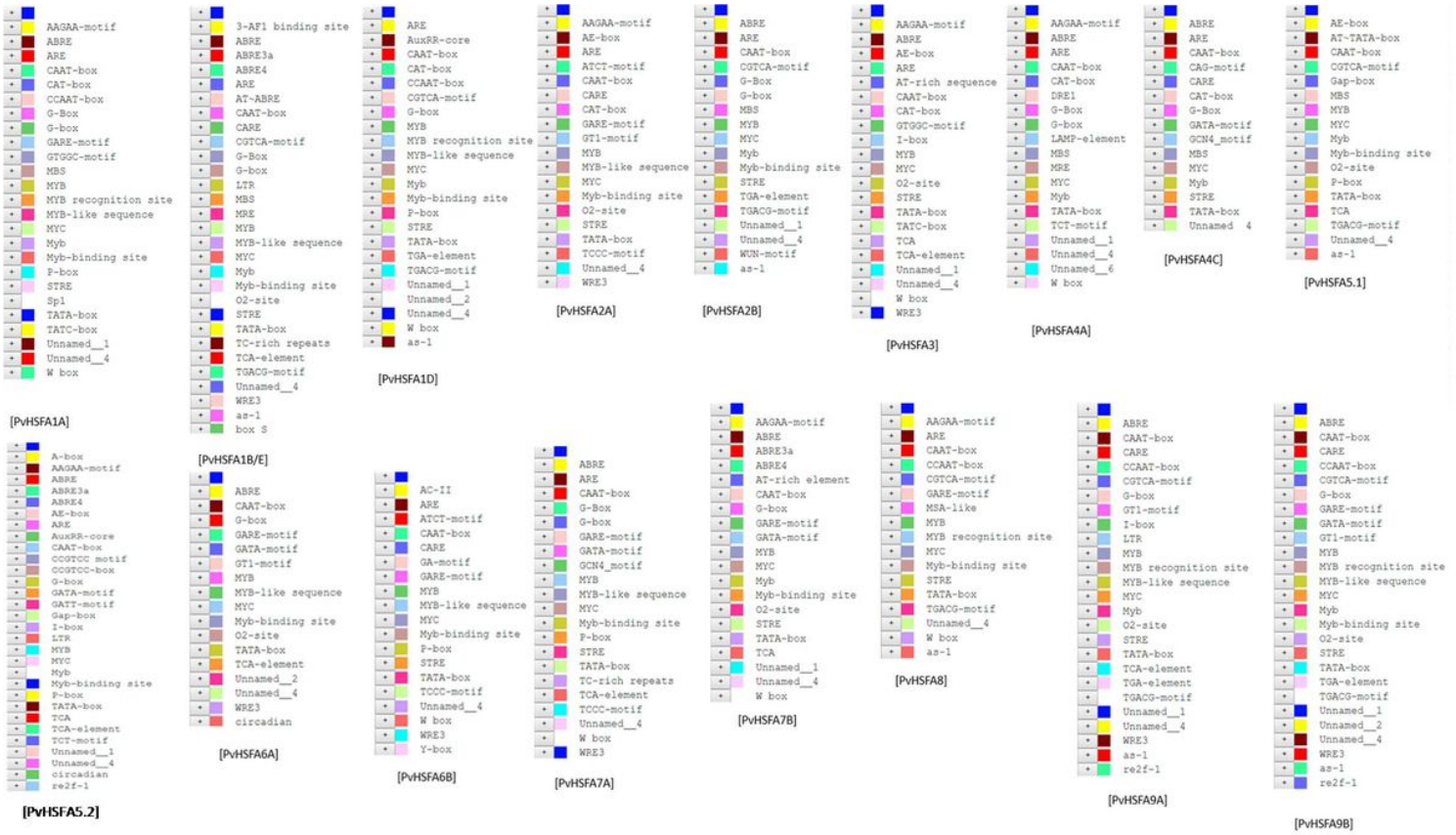

Figure 5

Cis acting elements of candidate PvHSFs

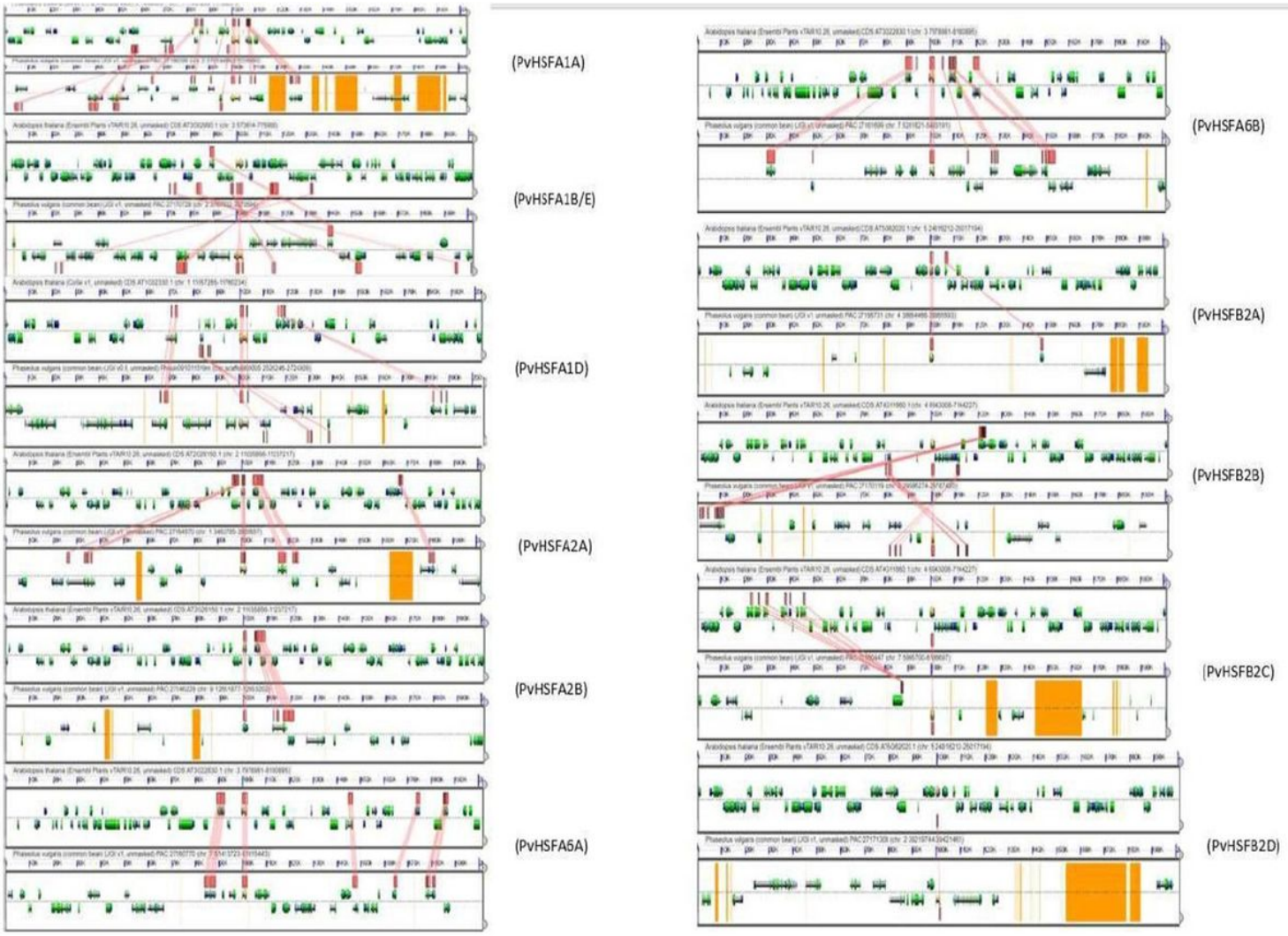

Figure 6

Synteny analysis between PvHSFs of A1, A2, A6 and B2 subclasses and AtHSFs

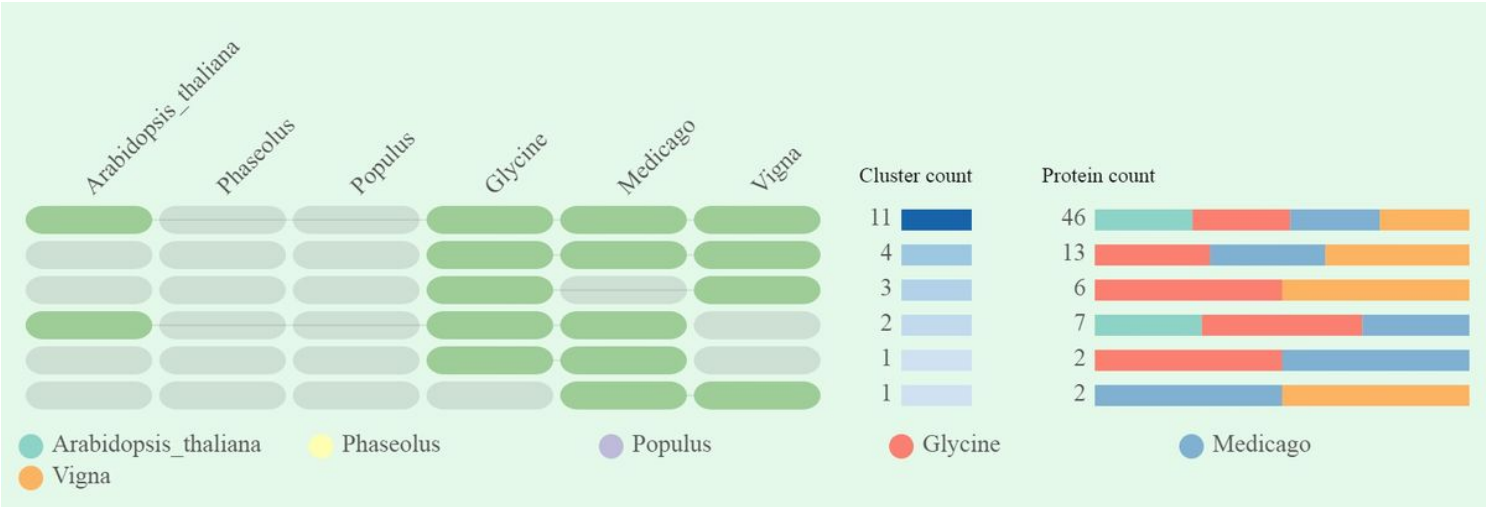

Figure 7

Orthologs identified between members of PvHSFs, AtHSFs,VrHSFs, PtHSFs, GmHSFs and MtHSFs

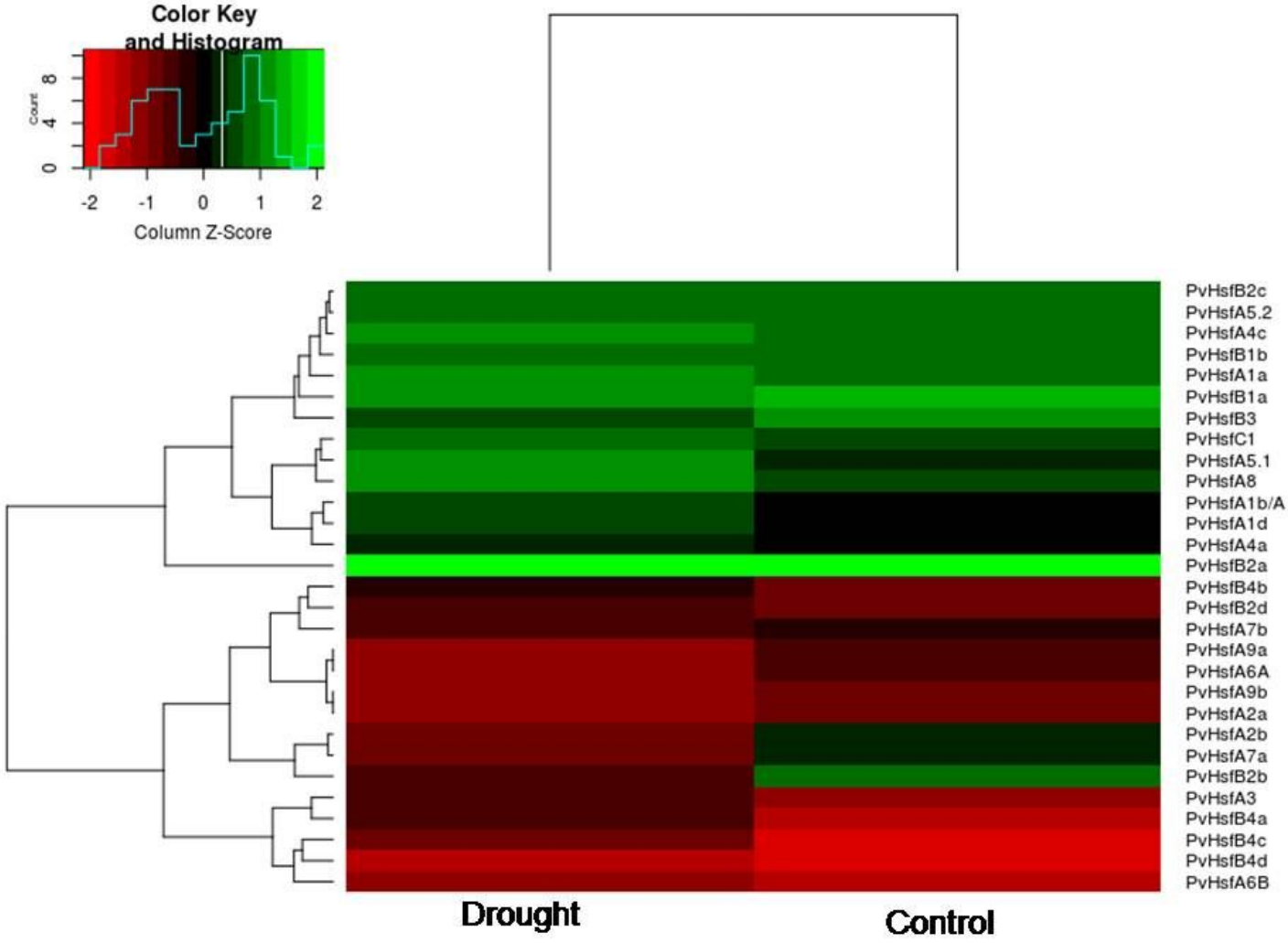

Figure 8

In silico expression analysis of putatively identified PvHSFs under abiotic stress (drought)

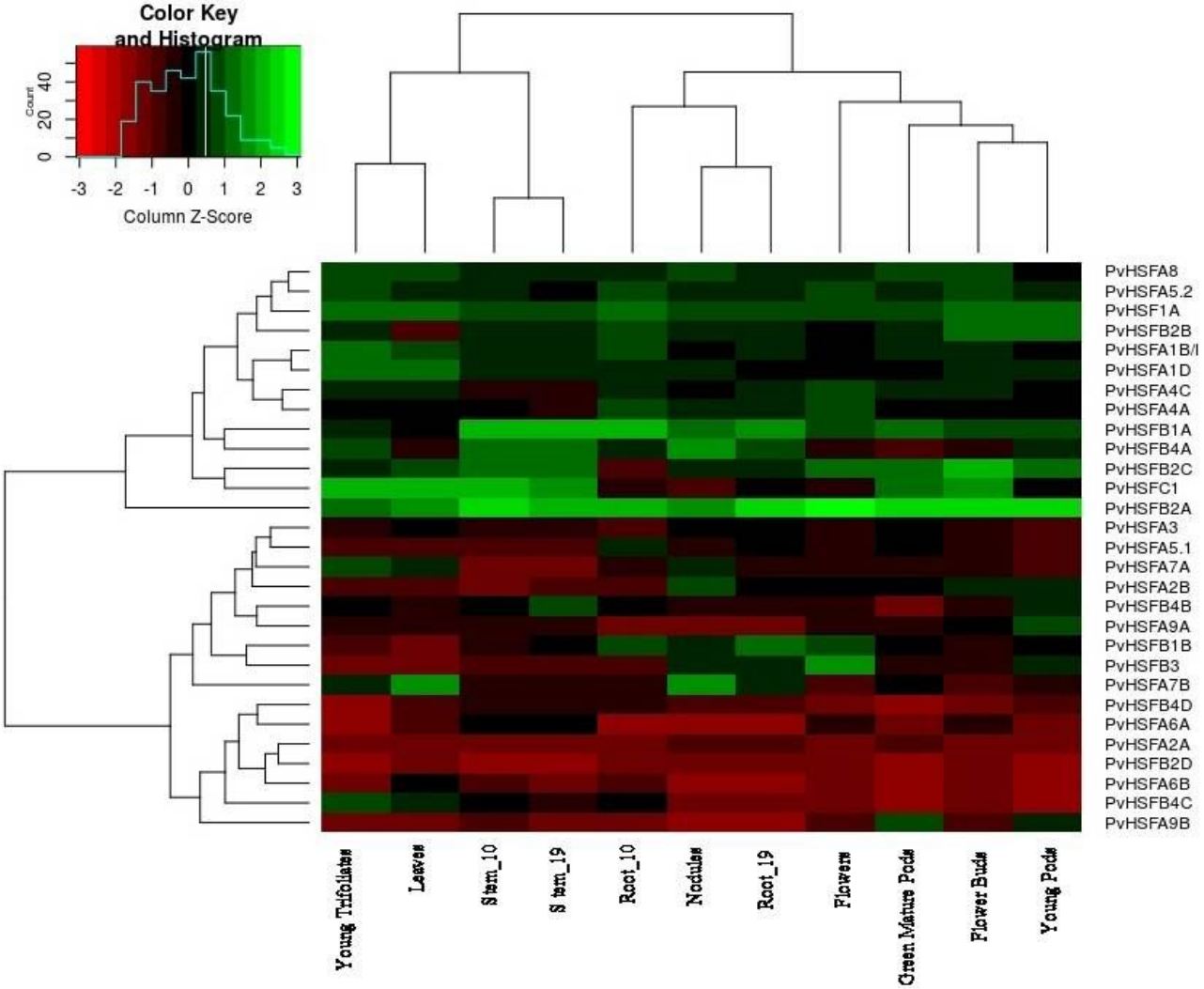

Figure 9

In silico expression analysis of putatively identified PvHSFs in different tissues of plant parts

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- S3.Hsfsinformationinotherplantspp..docx
- S2.Ciselement.xlsx
- S1.PVsequence.txt