Cell Stress & Chaperones (2001) **6** (3), 177–189

Article no. csac. 2001.310

Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need?

Lutz Nover, Kapil Bharti, Pascal Döring, Shravan Kumar Mishra, Arnab Ganguli, and Klaus-Dieter Scharf

Biocenter of the Goethe University, Marie-Curie-Strasse 9, D-60439 Frankfurt/Main, Germany

Abstract Sequencing of the *Arabidopsis* genome revealed a unique complexity of the plant heat stress transcription factor (Hsf) family. By structural characteristics and phylogenetic comparison, the 21 representatives are assigned to 3 classes and 14 groups. Particularly striking is the finding of a new class of Hsfs (AtHsfC1) closely related to Hsf1 from rice and to Hsfs identified from frequently found expressed sequence tags of tomato, potato, barley, and soybean. Evidently, this new type of Hsf is well expressed in different plant tissues. Besides the DNA binding and oligomerization domains (HR-A/B region), we identified other functional modules of *Arabidopsis* Hsfs by sequence comparison with the well-characterized tomato Hsfs. These are putative motifs for nuclear import and export and transcriptional activation (AHA motifs). There is intriguing flexibility of size and sequence in certain parts of the otherwise strongly conserved N-terminal half of these Hsfs. We have speculated about possible exon-intron borders in this region in the ancient precursor gene of plant Hsfs, similar to the exon-intron structure of the present mammalian Hsf-encoding genes.

INTRODUCTION

Heat stress transcription factors (Hsfs) are the terminal components of a signal transduction chain mediating the activation of genes responsive to both heat stress and a large number of chemical stressors (Wu 1995; Nover et al 1996; Morimoto 1998; Scharf et al 1998b; Schöffl et al 1998; Nakai 1999). They recognize palindromic binding motifs, so-called heat stress elements (HSEs; 5′-AGAAnnTTCT-3′) conserved in promoters of heat stress–inducible genes of all eukaryotes (Bienz and Pelham 1987; Nover 1987, 1991). The initial cloning and characterization of the yeast Hsf (Sorger and Pelham 1988; Wiederrecht et al 1988) were rapidly followed by cloning of the corresponding genes from *Drosophila* (Clos et al 1990), mammals (Rabindran et al 1991; Sarge et al 1991; Schuetz et al 1991), and tomato (Scharf et al 1990). The analysis of the tomato Hsf system exposed 2 interesting peculiarities. First, there are at least 4 different Hsfs (Scharf et al 1990, 1993; Treuter et al 1993; Bharti et al 2000) belonging to 2 classes, ie, class A with Hsfs A1, A2, and A3 and class B with HsfB1. Second, 2 of the 4 Hsfs (HsfA2 and B1) are heat stress–inducible proteins themselves. Although, with few exceptions, multiple Hsfs or Hsf-related proteins were subsequently found in other organisms as well (Nover et al 1996; Morimoto 1998; Nakai 1999), the peculiarities of an extended HR-A/B region in the class A Hsfs (Fig 1) and the heat stress–dependent expression are unique features of the plant Hsf system. The sequencing of the *Arabidopsis thaliana* genome now allows a more detailed discussion on the plant Hsf system based on the combination of sequence comparison and results from the fairly advanced functional analysis of tomato Hsfs.

Correspondence to: Lutz Nover, Tel: 49 69–798–29284; Fax: 49 69–798–29286; E-mail: nover@cellbiology.uni-frankfurt.de.

Basic structure and classification of 21 *Arabidopsis* Hsfs

Based on the presence of the conserved DNA-binding domain (DBD) plus the adjacent HR-A/B region (Fig 1), we identified 21 open reading frames in the *Arabidopsis* genome encoding putative Hsfs. Other sequences annotated

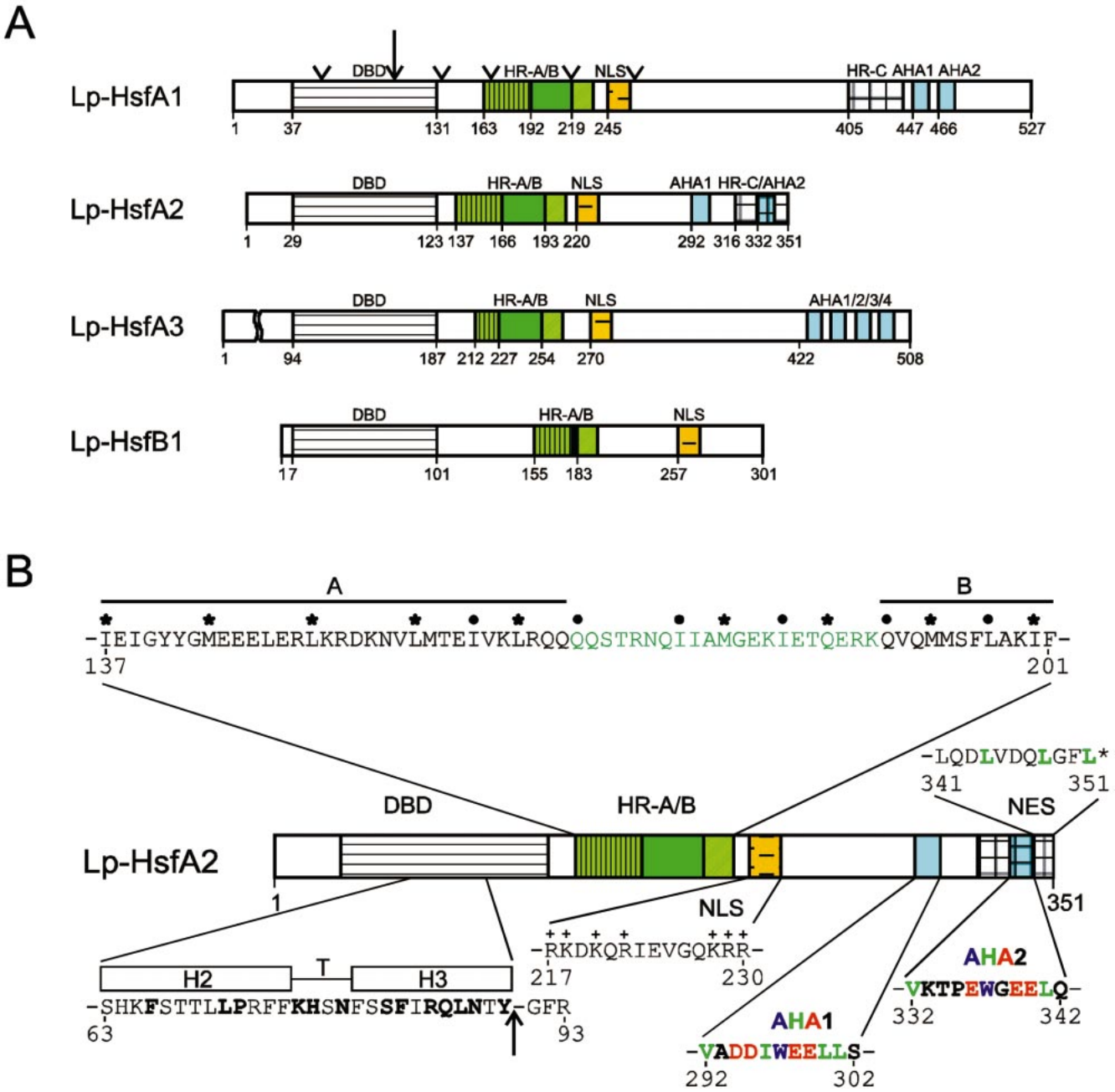


Fig 1. Basic structure of Hsfs. The block diagrams in part A represent tomato (*Lycopersicon peruvianum* [Lp]) Hsfs with their conserved functional domains. For abbreviations see part B. Arrowheads at the block diagram of Lp-HsfA1 indicate the positions of putative introns in the ancient *Hsf* gene. (B) Essential structural details are represented by the tomato HsfA2. (1) The central part of the DBD is the helix-turn-helix motif (H2-T-H3) with a considerable number of amino acid residues invariant among different organisms (boldfaced letters). The arrow indicates the position of the intron conserved in all plant Hsfs. (2) The oligomerization domain HR-A/B is characterized by the heptad pattern of hydrophobic residues (dots, asterisks). The insertion of additional 21 amino acid residues between parts A and B are marked in green. (3) The bipartite NLS represents a cluster of basic residues (K, R) recognized by the NLS receptor. (4) Central elements of the activator region are short motifs (AHA elements) rich in aromatic (W, Y, F), hydrophobic (L, I, V), and acidic amino acid residues (D, E). (5) A leucine-rich motif at the C-terminus functions as an NES.

as Hsf-like were not considered. Thus, the complexity of the plant Hsf system far exceeds that of any other organism whose genomic sequence is known. For comparison, there are a total of 4 Hsfs in vertebrates, only 1 Hsf in *Drosophila* and the nematode *Caenorhabditis elegans,* and 1 Hsf plus 3 Hsf-related proteins in yeast (Nover et al 1996; Nakai 1999).

Similar to many other proteins regulating gene activity,

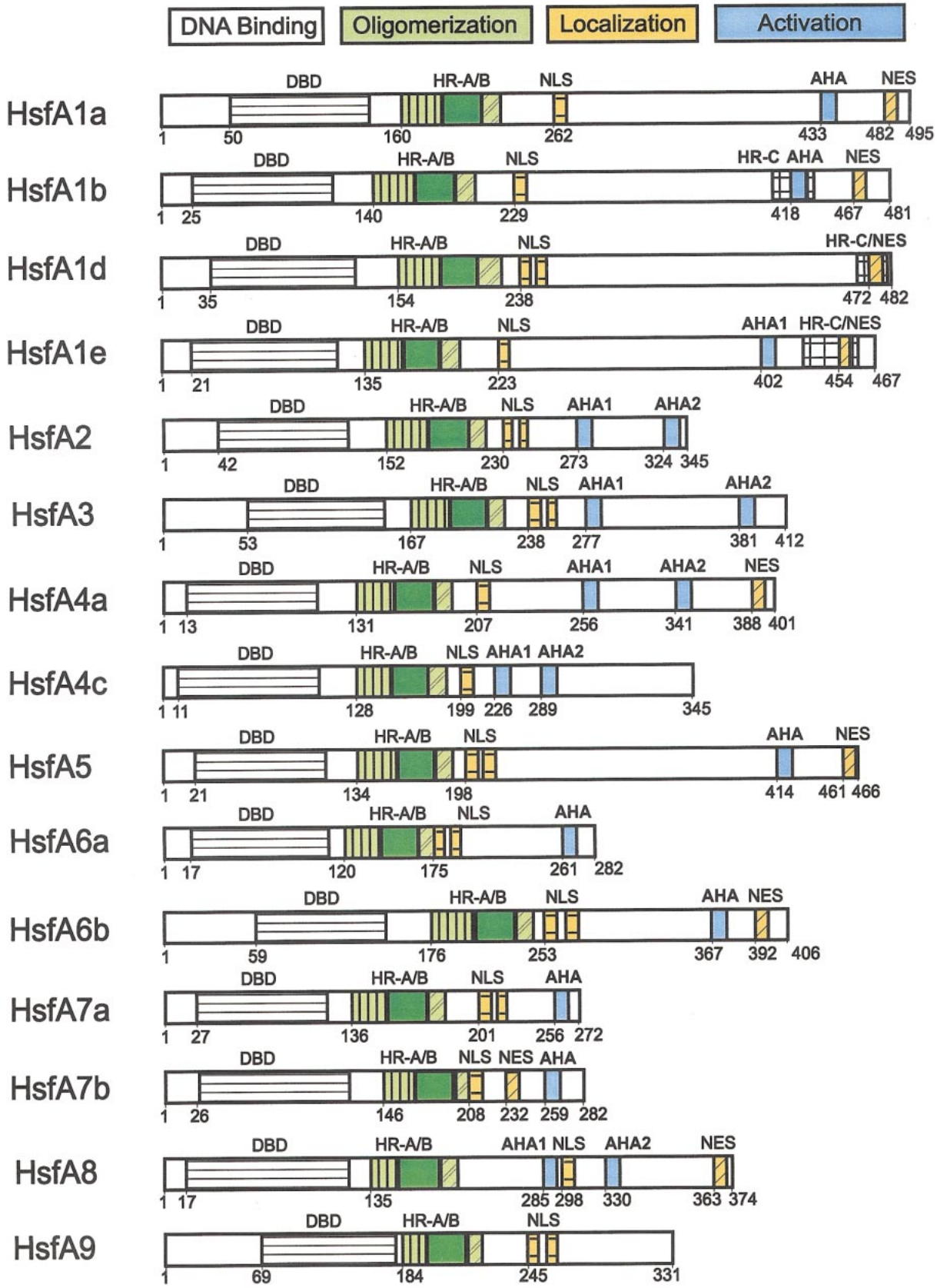


Fig 2. Survey of *Arabidopsis* Hsfs. For explanations see legend to Figure 1.

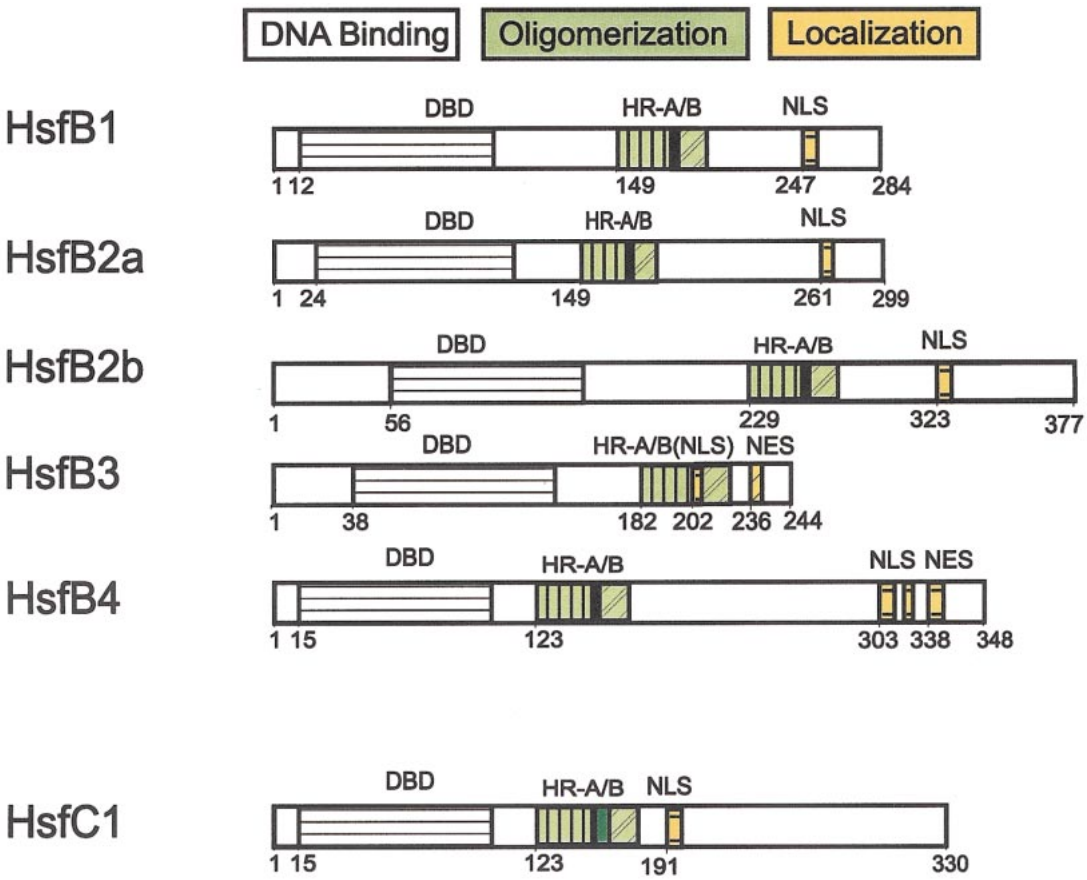


Fig 2. Continued.

Hsfs have a modular structure. Despite considerable variability in size and sequence, their basic structure is conserved among eukaryotes. This is exemplified by tomato HsfA2, with sequence details given for modules identified by mutation and functional analysis (Fig 1B). This knowledge helped to identify corresponding functional motifs for the *Arabidopsis* Hsfs compiled in Figure 2 and Table 1.

Close to the N-terminus, the highly structured DBD is the most conserved part of Hsfs. It consists of a 3-helical bundle (H1, H2, H3) and a 4-stranded antiparallel β-sheet (β1, β2, β3, β4). The hydrophobic core of this domain ensures the precise positioning of the central helix-turn-helix motif (H2-T-H3) required for specific recognition of the palindromic HSEs (Damberger et al 1994; Harrison et al 1994; Vuister et al 1994; Schultheiss et al 1996). The only crystal structure of a Hsf-DNA complex was reported for the DBD of the *Kluyveromyces lactis* Hsf (Littlefield and Nelson 1999). Interestingly, binding of 2 monomers of the DBD to the HSE motif involves protein-protein contacts mediated by the 10 amino acid residues of the loop (wing) between β3 and β4 strands, which is lacking in plant Hsfs. It will be interesting to elaborate the differences in the arrangement of DNA-bound Hsf subunits between plants and other organisms.

Similar to all known *Hsf* coding genes of other organisms, the DBD of plant Hsfs is encoded in 2 parts separated by the only intron, which is inserted immediately upstream of the coding part for the H2-T-H3 DNA binding motif (Fig 1). The position of the intron is identical in all cases, but the size is highly variable (Table 1). In the plant *Hsf* genes, the exon-intron borders are defined by the codons for the invariant Tyr residue (codons UAU or UAC) at the end of the HTH motif and the Gly residue (codons GGG, GGA, or GGT) in the following turn to β3 (intron sequence is indicated by the small case letters): 5′ TAT(C)-ag.....ag-GG(T, A, G). The arrow in Figure 1 indicates the position of the intron. Interestingly, the mammalian *Hsf* genes contain many additional introns (Zhang et al 1998; Manuel et al 1999). Based on structural similarities between mammalian and plant Hsfs, we marked the hypothetical positions of these introns by arrowheads at the block diagram of LpHsfA1 (Fig 1A). We will discuss arguments supporting a similar exon-intron organization in the ancient precursor gene encoding plant Hsfs.

Table 1 Survey of plant Hsfs

Number	Name (former name)	Chromosomal localization	Intron (nucl.)	ORF (aa)	MW (kDa)	pI	Protein accession number	Reference
				Class A				
	Group A1							
1	AtHsfA1a (Hsf1)	IV	147	495	55.7	4.9	CAB10555	Hübel and Schöffl 1994
2	AtHsfA1b (Hsf3)	V	606	481	53.6	4.6	CAA74397	Prändl et al 1998
3	AtHsfA1d	I	1419	482	54.2	4.7	AAF81328	DS
4	AtHsfA1e	III	968	468	52.0	4.6	AAF26960	DS
5	LpHsfA1 (Hsf8)	VIII	3477[a]	505	55.7	5.0	CAA47869	Scharf et al 1990, 1993
	Group A2							
6	AtHsfA2	II	324	345	39.1	5.0	AAC31222	Lin et al 1999
7	GmHsfA2 (Hsf21)		cDNA	PC			S59537	Czarnecka-Verner et al 1995
8	LpHsfA2 (Hsf30)	VIII	488	351	40.2	4.6	CAA47870	Scharf et al 1990, 1993
9	PsHsfA2 (HsfA)		cDNA	334	38.1	4.6	CAA09300	Aranda et al 1999
10	OsHsfA2	X	1669	358	40.8	4.7	AC027658[b]	
	Group A3							
11	AtHsfA3	V	532	412	46.4	5.2	CAB82937	DS
12	LpHsfA3		727	508	55.8	4.6	AAF74563	Bharti et al 2000
	Group A4							
13	AtHsfA4a (Hsf21)	IV	77	401	46.2	5.2	CAA16745	DS
14	AtHsfA4c	V	93	345	39.6	5.5	BAB09213	DS
15	MsHsfA4		cDNA	402	46.2	5.7	AAF37579	DS
16	NtHsfA4 (Hsf2)		cDNA	408	46.4	5.0	BAA83711	Shoji et al 2000
17	ZmHsfA4		cDNA	308	35.4	6.5	CAA58117	Gagliardi et al 1995
	Group A5							
18	AtHsfA5 (HsfA4b)	IV	328	466	52.4	6.2	CAB10177	DS
	Group A6							
19	AtHsfA6a	V	80	282	33.2	5.4	BAB11313	DS
20	AtHsfA6b	III	694	406	46.7	4.7	BAB01258	Sato et al 2000b
	Group A7							
21	AtHsfA7a	III	507	272	31.8	6.0	CAB41311	DS
22	AtHsfA7b	III	496	282	32.6	5.0	CAB86436	DS
	Groups A8 and A9							
23	AtHsfA8	I	345	374	42.6	4.6	AAF16564	DS
24	AtHsfA9	V	79	331	38.2	5.4	BAA97129	Sato et al 2000a
				Class B				
	Group 1							
25	AtHsfB1 (Hsf4)	IV	193	284	31.3	6.3	CAB16764	Prändl et al 1998, Nover et al 1996
26	GmHsfB1 (Hsf34)		cDNA	282	31.2	9.2	S59538	Czarnecka-Verner et al 1995
27	LpHsfB1 (Hsf24)	II	cDNA	301	33.2	5.6	CAA39034	Scharf et al 1990, 1993
28	NtHsfB1 (Hsf1)		cDNA	292	32.1	6.0	BAA83710	Shoji et al 2000
	Group B2							
29	AtHsfB2a (Hsf6)	V	83	299	34.1	5.8	CAB63802	Sato et al 2000a
30	AtHsfB2b (Hsf7)	IV	89	377	39.7	4.7	CAB39937	DS
	Group B3							
31	AtHsfB3	II	108	244	28.3	5.1	AAB84350	DS
	Group B4							
32	AtHsfB4	I	233	348	39.6	7.6	AAG34256	DS
33	GmHsfB4a (Hsf5)		cDNA	370	42.1	7.7	S59539	Czarnecka-Verner et al 1995
				Class C				
34	AtHsfC1	III	79	330	37.7	5.7	BAB02003	Kaneko et al 2000
35	OsHsfC1	I	84	339	36.9	6.2	BAB19067	DS

Heat stress transcription factors (Hsfs) from *Arabidopsis thaliana* (At), *Glycine max* (Gm), *Lycopersicon peruvianum* (Lp), *Medicago sativa* (Ms), *Nicotiana tabacum* (Nt), *Pisum sativum* (Ps), *Oryza sativa* (Os), and *Zea mays* (Zm) are listed with their new names based on their structure and phylogenetic relationship (see Figs 3 and 4). The existence of many more Hsfs also in tomato (*L esculentum*, totally more than 16 Hsfs) and soybean (*G max*, totally more than 15 Hsfs) can be deduced from sequence searches in the expressed sequence tag (EST) libraries. A survey of relevant information is compiled in Figure 5. The molecular weight (MW) and isoelectric points (pI) of Hsfs were calculated on the basis of the amino acid (aa) sequences of the open reading frames (ORF) using Clone Manager 5 software. The chromosomal location and size of the intron are indicated whenever possible. For identification, the accession numbers and references are given. DS, direct submission to the database as part of *Arabidopsis* or rice genomic sequencing projects; DNA, complementary DNA; PC, partial clone. Because of an additional exon added by the computer program to the N-terminal part of At-HsfA5, the ORF in the database has extended to 834 aa residues. We did not include this exon but rather considered the methionine residue preceding the DNA binding domain by 15 aa residues as putative translational start site. This gives an ORF of 466 aa residues. A similar situation holds true for At HsfA6a. In this case, an additional exon with a stop codon was introduced in the C-terminal domain, creating a truncated version of the Hsf with 251 aa residues. Our ORF with 282 aa residues simply neglects this putative exon 3 and includes the end of the normal exon 2, which includes also a classic AHA motif (see Table 3).

[a] The size of the intron was derived from the *L esculentum* HsfA1 clone (LEHSF8, accession no. X67599).

[b] The rice HsfA2 was not yet assigned. It is coded on BAC AC027658 derived from chromosome X in the region defined by nucl. 106 800 to 103 000.

Table 2 Comparison of the linker/HR-A/B regions of Hsfs

Hsf	DBD	Linker/HR-A	HR-A core	Insert	HR-B
AtHsfA1a	**ISRRK**	28aa	SCVEVGKFG**L**EEEVEQ**L**KRDKNV**L**MQELVK**L**	rqqqqttdnklqvlvkhlqvmeqrqqq	**IMSFLAKAV**
AtHsfA1b	**IVRRK**	19aa	ACVEVGKFG**I**EEEVER**L**KRDKNV**L**MQELVR**L**	rqqqqatenqlqnvgqkvqvmeqrqqq	**MMSFLAKAV**
AtHsfA1d	**ITRRK**	23aa	ACVEVGKFG**L**EEEVER**L**KRDKNV**L**MQELVR**L**	rqqqqstdnqlqtmvqrlqgmenrqq	**QLMSFLAKAV**
AtHsfA1e	**IVRRK**	18aa	ACVEVGKFG**L**EEEVER**L**QRDKNV**L**MQELVR**L**	rqqqqvtehhlqnvgqkvhvmeqrqqq	**MMSFLAKAV**
LpHsfA1	**ISRRK**	30aa	ACVEVGKFG**L**EEEVER**L**KRDKNV**L**MQELVR**L**	rqqqqstdnqlqgmvqrlqgmelrqqq	**MMSFLAKAV**
AtHsfA2	**IKRRR**	14aa	SCVEVGQYG**F**DGEVER**L**KRDHGV**L**VAEVVR**L**	rqqqhssksqvaameqrllvtekrqqq	**MMTFLAKAL**
PsHsfA2	**IKRRR**	14aa	ACIELGEFG**L**EGEIER**L**RRDRSV**L**VAEIVK**L**	rqqqnnsrdqisamearllitekkhqq	**MMAFLARAL**
LpHsfA2	**IKRRR**	12aa	ACIEIGYYG**M**EEELER**L**KRDKNV**L**MTEIVK**L**	rqqqqstrnqiiamgekietqerkqvq	**MMSFLAKIF**
OsHsfA2	**IKRRK**	14aa	SC**L**EVGEFG**F**EEEIDR**L**KRDKNI**L**ITEVVK**L**	rqeqqatkdhvkamedrlraaeqkqvq	**MMGFLARAM**
AtHsfA3	**IHRRR**	12aa	SQSQGSPTE**V**GGEIEK**L**RKERRA**L**MEEMVE**L**	qqqsrgtarhvdtvnqrlkaaeqrqkq	**LLSFLAKLF**
LpHsfA3	**IQRRR**	8aa	GSSAEAGKGTMDEIEK**L**RNEKSL**M**MQEVVE**L**	qqqqhgtvqlmesvneklqaaeqrqkq	**MVSFLAKVL**
AtHsfA4a	**IHRRK**	15aa	PLTDSERVR**M**NNQIER**L**TKEKEG**L**LEELHK**Q**	deerevfemqvkelkerlqhmekrqkt	**MVSFVSQVL**
AtHsfA4c	**IHRRK**	14aa	PLTESERRS**M**EDQIER**L**KNEKEG**L**LAELQN**Q**	eqerkefelqvttlkdrlqhmeqhqks	**IVAYVSQVL**
NtHsfA4	**IHRRK**	15aa	PLTESERQG**Y**KEDIQK**L**KHENES**L**HLDLQR**H**	qqdrqglelqmqvftervqhvehrqkt	**MLSALARML**
MsHsfA4	**IHRRK**	14aa	SLTESERQS**M**IDEIEK**L**KQDREQ**L**LVETKR**Y**	qhdwerheiqmhcskdqleklehkqqk	**MLSSVSEAL**
ZmHsfA4	**IHRRK**	13aa	PLPDTERRD**Y**EEEIER**L**KCDNAA**L**TSELEK**N**	Aqkklvtekrmqdledklifledrqkn	**LMAYVRDIV**
AtHsfA5	**IHRRK**	10aa	SSTDQERAV**L**QEQMDK**L**SREKAA**I**EAKLLK**F**	kqqkvvakhqfeemtehvddmenrqkk	**LLNFLETAI**
AtHsfA6a	**IKRRK**	----	TSSQTQTQS**L**EGEIHE**L**RRDRMA**L**EVELVR**L**	rrkqesvktylhlmeeklkvtevkqem	**MMNFLLKKI**
AtHsfA6b	**IRRRK**	21aa	FCIEVGRYG**L**DGEMDS**L**RRDKQV**L**MMELVR**L**	rqqqqstkmyltlieeklkkteskqkq	**MMSFLARAM**
AtHsfA7a	**IKRRN**	11aa	ACNE----------**L**RREKQV**L**MMEIVS**L**	rqqqqttksyikameqriegterkqrq	**MMSFLARAM**
AtHsfA7b	**IKRRT**	17aa	HDPGVELPQ-------**L**REERHV**L**MMEIST**L**	rqeeqrargyvqameqringaekkqrh	**MMSFLRRAV**
AtHsfA8	**IRRKN**	14aa	TTYAQEKSG**L**WKEVDI**L**KGDKQV**L**AQELIK**V**	rqyqevtdtkmlhledrvqgmeesqqe	**MLSFLVMVM**
AtHsfA9	**IKRRS**	13aa	TTTETEVES-------**L**KEEQSP**M**RLEMLK**L**	kqqqeesqhqmvtvqekihgvdteqqh	**MFSFFAKLA**
AtHsfC1	**IARRK**	5aa	YGGDLEDGE**I**VREIER**L**KEEQRE**L**EAEIQR**M**	nrriea----------tekrepeq	**MMAFLYKVV**
OsHsfC1	**RRKKR**	40aa	EDVLAKEAALFEEVQR**L**RHEQTA**I**GEELAR**M**	sqrlqa----------terrpdq	**LMSFLAKLA**
AtHsfB1	**IRRRK**	41aa	GSVENMVAD**L**SGENEK**L**KRENNN**L**SSELAA**A**	kkqr----------de	**LVTFLTGHL**
NtHsfB1	**IRRRK**	44aa	TPGKSQFAD**L**SDENEK**L**KKDNQM**L**SSELAQ**A**	kkqc----------de	**LVAFLNQYV**
LpHsfB1	**IRRRK**	45aa	PGKLSQFTD**L**SDENEK**L**KKDNQM**L**SSELVQ**A**	kkqc----------ne	**LVAFLSQYV**
GmHsfB1	**IKRRK**	48aa	ETNTTPSHQ**L**SSENEK**L**KKDNET**L**SCELAR**A**	rkqc----------de	**LVAFLRDRLM**
AtHsfB2a	**IQRRK**	51aa	TGNGGLSVE**L**LEENEK**L**RSQNIQ**L**NRELTQ**M**	ksic----------dn	**IYSLMSNYV**
AtHsfB2b	**IQRRK**	69aa	TTSCTTAPE**L**VEENER**L**RKDNER**L**RKEMTK**L**	kgly----------an	**IYTLMANFT**
AtHsfB3	**IRRRK**	41aa	TSSSFVYTA**L**LDENKC**L**KNENEL**L**SCELGKTkkkc----------kq		**LMELVERYR**
AtHsfB4	**IHRRK**	61aa	IDTAAQVTA**L**SADNER**L**RRSNTV**L**MSELAH**M**	kkly----------nd	**IIYFVQNHV**
GmHsfB4a	**IHRRK**	69aa	SNNYNTVTA**L**SEDNER**L**RRSNNM**L**MSELAH**M**	kkly----------nd	**IIYFVQNHV**

As indicated on top of the table, sequence details of the linker/HR-A/B regions of heat stress transcription factors (Hsfs) are given in the 1-letter code for amino acid (aa) residues. The arrowheads mark the positions of the introns in the putative *Hsf* precursor gene. The linker size is indicated by the number of aa residues between the cluster of basic aa residues at the C-terminal end of the DNA binding domain (DBD) (boldfaced letters on the left). The HR-A region is separated into 2 parts by the position of the putative intron, (1) the N-terminal optional part, which is modified or lacking in some Hsfs, and (2) the HR-A core formed by the 2 internal heptad repeats. Only this core part was included into the phylogenetic tree (Fig 3).

The oligomerization domain (HR-A/B region) is connected to the characteristic cluster of basic amino acid residues at the C-terminus of the DBD by a flexible linker of 9 to 39 amino acid residues for class A, 50 to 78 amino acid residues for class B, and 14 to 49 amino acid residues for class C Hsfs (Fig 2, Table 2). From experiments with the yeast and mammalian Hsfs (Flick et al 1994; Liu and Thiele 1999), it is apparent that this linker or at least part of it is important for the oligomerization behavior. The heptad pattern of hydrophobic amino acid residues in the HR-A/B region (Fig 1B) suggest a coiled-coil structure similar to that reported also for leucine-zipper–type protein interaction domains (Peteranderl et al 1992, 1999).

In plants, there are 3 classes in the Hsf protein family (classes A, B, and C), which are discriminated by peculiarities of their flexible linkers and HR-A/B regions (Figs 1 and 2 and Table 2). The HR-A/B region of class B Hsfs are similar to all nonplant Hsfs, whereas all class A and class C Hsfs have an extended HR-A/B region due to an insertion of 21 (class A) and 7 (class C) amino acid residues between the A and B parts (see sequence details in Table 2). It is remarkable that representatives of the third class (class C) were unnoticed so far. From analysis of expressed sequence tag (EST) libraries, they are evidently well expressed in tomato, soybean, potato, barley, and *Arabidopsis,* at least on the RNA level. This HsfC1 type is clearly separated from all others by sequence details of the DBDs and by the characteristics of the HR-A/B region. The significance of these extended oligomerization domains in class A and class C Hsfs for the coiled-coil structure and oligomerization behavior is not yet clear. However, there is evidence that the tomato HsfB1 exists as a dimer, whereas HsfA1 and HsfA2 are trimers.

The flexibility of size and sequence in this part of the Hsfs between the C-terminus of the DBD and the B part of the HR-A/B region is remarkable (Table 2). Because of the lack of sufficient experimental data, we can only speculate that the individual properties of Hsfs, ie, their olig-

Table 3 Functional motifs of plant Hsfs

Number	Hsf	NLS	NES	AHA motifs
	Class A Hsfs			
1	AtHsfA1a	(262) NKKRRLR	(482) LIEELGLL	AHA (433) FE**FL**EE**YM**PE
2	AtHsfA1b	(229) NKKRR	(467) LTEQMGLL	AHA (418) DP**FW**EQ**FF**SV
3	AtHsfA1d	(238) NRR-5aa-KKRRFKR	(472) LLSPETLDL	nd
4	AtHsfA1e	(223) NKKRR	(454) LTEQMGLL	AHA (402) DS**FW**EQ**FI**GE
5	LpHsfA1	(222) NKR-5aa-KKRRIK	(486) MEHLTEQM	AHA1 (428) I**D**W**Q**S**GLL**D**E
				AHA2 (446) D**P**F**W**E**K**F**L**Q**S**
6	AtHsfA2	(230) KEKK-8aa-RKRR	(334) MVDQMGFL	AHA1 (273) E**M**L**F**AAA**I**D**D**
				AHA2 (324) L**D**W**D**S**Q**D**L**H**D**
7	LpHsfA2	(217) RKDKQR-4aa-QKRR	(344) LVDQLDFL	AHA1 (294) D**D**I**W**EE**LL**S**E**
				AHA2 (335) P**E**W**G**EE**L**Q**D**L**
8	PsHsfA2	(204) KRKRR	(321) VDLQNL	AHA1 (277) L**S**D**W**EE**LL**N**Q**
				AHA2 (296) E**V**L**I**G**D**F**S**Q**I**
9	OsHsfA2	(229) KRK-7aa-KKRRR	(345) LAQQLGYL	AHA1 (265) P**YLF**D**S**G**V**L**N**
				AHA2 (314) D**D**F**W**AE**LL**V**E**
10	AtHsfA3	(238) KGKEK-7aa-KARKK	nd	AHA1 (277) D**D**W**E**R**LL**M**Y**D**
				AHA2 (381) D**V**C**W**E**Q**F**A**A**G**
11	LpHsfA3	(272) RQMK-11aa-RKFVKH	nd	AHA1 (429) EE**LW**G**M**G**F**E**A**
				AHA2 (447) P**E**L**W**D**S**L**S**S**Y**
				AHA3 (467) S**D**L**W**D**I**D**P**L**Q**
				AHA4 (483) V**D**K**W**P**A**D**G**S**P**
12	AtHsfA4a	(207) RKRRFPR	(388) ITEQLGHL	AHA1 (256) I**A**I**W**EN**LV**S**D**
				AHA2 (341) D**G**F**W**Q**Q**F**F**S**E**
13	AtHsfA4c	(199) RRKRR	nd	AHA1 (226) L**T**F**W**EN**LV**S**E**
				AHA2 (289) D**D**F**W**EQC**L**T**E**
14	MsHsfA4	(202) RKRR	nd	AHA1 (338) D**V**F**W**EQ**FL**T**E**
				AHA2 (377) G**R**F**W**W**N**M**R**K**S**
15	NtHsfA4	(205) RKRR	(395) LAEQLGHL	AHA1 (257) L**T**F**W**EN**VL**Q**D**
				AHA2 (344) D**I**F**W**EQ**FL**T**E**
16	ZmHsf4	(196) HGKKRR	(298) LVREIRSL	nd
17	AtHsfA5	(198) KNFGKK-10aa-KKRR	(461) IEQLTL*	AHA (414) DV**FW**EQ**FL**T**E**
18	AtHsfA6a	(175) KKIKK-7aa-RKRN	nd	AHA (261) EG**IW**KG**FV**L**S**
19	AtHsfA6b	(253) KEKRK-7aa-KKRQR	(392) LIQQLGYL	AHA (367) EG**FW**ED**LL**N**E**
20	AtHsfA7a	(201) KKIK-8aa-KRKR	nd	AHA (256) D**G**F**W**EE**LL**S**D**
21	AtHsfA7b	(208) QKRDR (?)	(232) LSELEALAL	AHA (259) D**G**F**W**EE**LL**M**N**
22	AtHsfA8	(298) RKKTKK	(363) LTEQMELL	AHA1 (285) D**G**A**W**EK**LL**L**L**
				AHA2 (330) K**S**Y**M**L**K**L**I**S**E**
23	AtHsfA9	(245) KKRKMK-11aa-KKLK	nd	nd
	Class B Hsfs			
24	AtHsfB1	(247) RKKRDR	nd	nd
25	LpHsfB1	(255) KEKKKR	nd	nd
26	NtHsfB1	(246) KENKKKR	nd	nd
27	GmHsfB1	(240) KRNNHKRDR	nd	nd
28	AtHsfB2a	(261) KRTK (?)	nd	nd
29	AtHsfB2b	(323) KRARR	nd	nd
30	AtHsfB3	(202) KTKKCK	(236) LKLFGVKL	nd
31	AtHsfB4	(303) RKTK-9aa-KKR	(338) LALNLM	nd
32	GmHsfB4a	(318) KTK-6aa-QSKKR	(359) LLGLNLM	nd
	Class C Hsfs			
33	AtHsfC1	(191) KKKRR	nd	nd
34	OsHsfC1	(229) RKRRRQH	nd	nd

Numbers in brackets indicate position of the putative nuclear localization signal (NLS), nuclear export signal (NES) and activator (AHA) motifs in the C-terminal domain. Putative NLS can be monopartite (eg, nos. 1, 2, 4) or bipartite (eg, nos. 3, 5). In the latter case, 2 basic clusters are separated by a number of amino acid residues as indicated. For the AHA motifs, aromatic and large hydrophobic residues are set in boldface type. nd, no motifs detectable by sequence homology. Function of the underlined motifs of tomato heat stress transcription factors (Hsfs) were tested experimentally (see Lyck et al 1997 for NLS; Treuter et al 1993, Döring et al 2000, and Bharti et al 2000 for AHA motifs; and Heerklotz et al 2001 for the NES of HsfA2).

omerization, their specific role and regulatory behavior in the Hsf network, and their interaction with other proteins, are dependent on sequence information in this region. In some cases the linker between DBD and the first repeat of the HR-A part is very short and/or the first repeat is lacking (see examples in Table 2, Lp-HsfA3, At-Hsfs A6a, A7a, A7b, and A9). In this context, the peculiarities of the HsfC1 in the linker/HR-A/B region are particularly striking.

In most cases, the nuclear localization signals (NLSs) of class A and class C Hsfs are monopartite or bipartite clusters of basic amino acid residues adjacent to the HR-A/B region. The corresponding motifs are marked with NLS in Figure 2 (for details of sequences and positions

Table 4 Expression of *Arabidopsis* Hsfs

	EST libraries[a]						Northern/RT-PCR[b]	
Hsf	**Mix**	**RL**	**Seedling**	**Root**	**DS**	**GS**	**Leaves**	**Cell culture**
HsfA1a							+	++
HsfA1b				+	+		++	++
HsfA1d	+						+(hs)	+
HsfA1e	+						+(hs)	+
HsfA2							+++(hs)	+++(hs)
HsfA3							++(hs)	+(hs)
HsfA4a	+						++	++
HsfA4c	++						+(hs)	++
HsfA5						+	+(hs)	+
HsfA6a							—	—
HsfA6b							+(hs)	—
HsfA7a			+				+++(hs)	++(hs)
HsfA7b							++(hs)	++(hs)
HsfA8							(+)	(+)
HsfA9		+					—	—
HsfB1	++			++			++(hs)	++
HsfB2a							(+)	++(hs)
HsfB2b	++		+	+	+		+(hs)	+(hs)
HsfB3							(+)	—
HsfB4							(+)	+
HsfC1	++			++			++	+

[a] Expressed sequence tag (EST) libraries from the dbEST database of NCBI were screened with each heat stress transcription factor (Hsf). +, 1 EST; ++, 2 or more ESTs found for a given Hsf; Mix, data from the Ohio State clone set and from a mixed library containing ESTs derived from RNA from etiolated seedlings, roots, leaves from vegetative plants, stems, flowers, and siliques. RL, rosette leaves; DS, developing seeds; GS, green silique. By sequence comparison, the following ESTs were assigned: HsfA1b, AV548883, BE523447; HsfA1d, N38285; HsfA1e, N96842; HsfA4a, AW004500; HsfA4c, AI995151, H37587, H76687; HsfA5, AV558506, F15453; HsfA7a, AI992565, AV544578, R65204, AA042693; HsfA9, AI997827; HsfB1, AV538768, AA712283, R90161, AV551082, AI993273, T75808, T44458; HsfB2b, AI996379, AA605426, R90511, BE526155, W43651; HsfC1, H36030, AV552195, T21116, AV543793.

[b] Northern analyses of leaf RNA were published for HsfsA1a (Hsf1), A1b (Hsf3), and B1 (Hsf4) by Hübel and Schöffl (1994) and Prändl et al (1998). Reverse transcriptase–polymerase chain reaction (RT-PCR) data for messenger RNA of other Hsfs in cell cultures and leaves are from A. Ganguli (unpublished data). (+), +, ++, increasing intensity of signals; (hs), messenger RNA level markedly increased in heat-stressed cells.

see Table 3). Because of the investigations on the nuclear import signals of tomato Hsfs A1 and A2 (Lyck et al 1997), it is clear that only these clusters, and not the clusters of conserved basic amino acid residues at the end of the DBD (Table 2), contribute to the nuclear import of Hsfs.

A cluster of arginine and lysine residues close to the C-terminus of tomato HsfB1 is responsible for its permanent nuclear localization (Scharf et al 1998a; Heerklotz et al 2001). Similar motifs are also found in other representatives of this group and in groups B2 and B4. An exception is At-HsfB3, which is the smallest of all Hsfs identified so far. In this case, the only cluster of basic amino acid residues, which might function as a NLS, is present in the linker region between HR-A and HR-B (Fig 2 and Table 3).

The nucleocytoplasmic distribution of proteins can be markedly influenced by nuclear export. Because of a leucine-rich export signal in the HR-C region (nuclear export signal [NES], see Fig 1 and Table 3 for sequence details), the tomato HsfA2 is mostly found in the cytoplasm unless complexed in hetero-oligomers with HsfA1 (Scharf et al 1998a; Heerklotz et al 2001). This phenomenon of changing nucleocytoplasmic distribution due to the balance of NLS and NES, which may be altered by protein modification or by interaction with other proteins, is crucial for many signaling pathways involving transcription factors (for references see Heerklotz et al 2001; Görlich and Kutay 1999). Inspection of the C-terminal parts of *Arabidopsis* Hsfs was shown in several cases to be similar to leucine-rich sequences, which might function as NESs (Fig 2 and Table 3).

The C-terminal activation domains (CTAD) of the Hsfs are the least conserved in sequence and size. For all class A Hsfs, the CTADs are acidic and enriched in proline, serine, threonine, glutamic acid, and aspartic acid residues. The function of Hsfs as transcription-activating proteins is evidently connected with short peptide motifs (AHA motifs, see Nover and Scharf 1997; Döring et al 2000), which are characterized by aromatic (W, F, Y), large hydrophobic (L, I, V), and acidic (E, D) amino acid residues (see examples given in Fig 1 and Table 3). With few exceptions, such AHA motifs are found in the center of the CTADs of most *Arabidopsis* class A Hsfs. Similar AHA motifs are also in the center of the activation domains of human, *Drosophila*, and yeast Hsfs and of many other transcription factors of mammals, eg, VP16, RelA, Sp1, Fos, Jun, steroid receptors, and the yeast, eg, Gal4 and

Gcn4 (see references in Nover and Scharf 1997; Döring et al 2000). Most likely, they represent the essential sites of contact with subunits of the basal transcription complex as shown by pull-down experiments (Yuan and Gurley 2000; Döring, in preparation) .

It is remarkable that the C-terminal domains (CTDs) of class B Hsfs are completely different. For the group of the most conserved Hsfs in plants, ie, HsfB1, the CTD is positively charged, and AHA motifs are lacking. We have evidence that HsfB1 plays a special role in gene activation as a synergistic partner of HsfA1 (Bharti et al, unpublished data). The CTDs of Hsfs B2, B3, B4, and C1 are neutral, with clusters of basic amino acid residues interspersed. AHA motifs are not detectable. It will be interesting to test the function of these Hsfs in reporter assays alone and in combination with other Hsfs to find out which of them are synergistic coactivators (Bharti et al, in preparation) or repressors (Czarnecka-Verner et al 2000).

Four tomato Hsfs with their structural and functional identities

Although the picture is far from complete, the experimental data obtained with the tomato Hsfs indicate that the multiplicity may be connected with distinct functions in the Hsf network. The question of how many Hsfs we need cannot be answered at present. However, the multiplicity of regulatory effects in the Hsf system of tomato with only 4 Hsfs cloned and experimentally studied so far is surprising. It gives an idea of the real complexity of the Hsf network with 21 representatives in *Arabidopsis.* In fact, the overall complexity of Hsfs in tomato and other plants is comparable. Searching EST libraries, we found expression data for 15 new tomato Hsfs with representatives in practically all groups and classes defined for *Arabidopsis* (see additional information compiled in Fig 5 available in the online version only).

In the following, we will briefly summarize the relevant experimental data, indicating well-defined and nonredundant roles of the 4 tomato Hsfs in the Hsf network.

First, the constitutively expressed HsfA1 (527 aa residues) is the largest of the 4 tomato Hsfs studied so far. There is evidence that HsfA1 plays a central role for the heat stress response in general and for the expression and function of the other Hsfs. The CTAD harbors 2 AHA motifs (Table 3) that are essential for the activator function (Döring et al 2000).

Second, synthesis of HsfA2 (351 amino acid residues) is strictly heat stress dependent. It accumulates to fairly high levels in tomato cell cultures and different tomato tissues, especially in periods of repeated heat stress and recovery. Similar to HsfA1, HsfA2 is a strong transcription activator with 2 AHA motifs in its CTAD (Fig 1B, Table 3). In the course of a heat stress regimen, HsfA2 exists in 3 different forms characterized by their intracellular distribution and modes of protein interaction: (1) Nuclear form: Because of the strong C-terminal NES (see details in Fig 1B), significant nuclear retention of HsfA2 and activator function are found only in the presence of HsfA1, ie, in form of the HsfA1/A2 heterooligomer (Scharf et al 1998; Heerklotz et al 2001). (2) Cytoplasmic insoluble form: The ongoing accumulation of HsfA2 during long-term heat stress and its stability coincides with its interaction with Hsp17 class II (Scharf et al, unpublished data). During heat stress, both proteins are reversibly incorporated into the cytoplasmic chaperone complexes built of the heat stress granules (HSGs). (3) Cytoplasmic soluble form: After dissociation of the HSG complexes during the recovery, HsfA2 and small Hsps are found in soluble oligomers in the cytoplasm. Release of HsfA2 from the HSG complexes needs the activity of the Hsp90 chaperone machinery, ie, it is inhibited by geldanamycin (Scharf et al 1998a, unpublished results).

Third, HsfA3 (508 amino acid residues) is the least studied of the 4 tomato Hsfs. It is found constitutively expressed in cell cultures but is barely detectable in tomato leaves. It may represent a developmentally regulated Hsf with expression only in rapidly dividing cells. Four AHA motifs in the CTAD with a central Trp residue contribute to its activator function (Bharti et al 2000).

Fourth, HsfB1 (301 amino acid residues) is the only class B Hsf so far studied in tomato. Its very low level found in unstressed cell cultures or tissues is transiently increased several-fold after heat stress. HsfB1 is relatively short-lived and always found in the nucleus. In contrast to the class A Hsfs, HsfB1 probably has no activator function itself. This can be explained by the differences in the CTDs (see above). However, we have preliminary evidence that coexpression of low levels of HsfA1 with HsfB1 can result in strong synergistic effects in reporter gene activation (Bharti et al, unpublished data).

Expression of *Arabidopsis* Hsfs

Expression data for *Arabidopsis* Hsfs are fragmentary and stem from different sources (see compilation in Table 4). First, a number of ESTs are found in the database. They were derived from RNAs isolated from different tissues of *Arabidopsis* (see details given in footnote b to Table 4). Second, Schöffl's group published data from Northern analyses for Hsfs A1a, A1b, and B1 using control and heat stress leaves (Hübel and Schöffl 1994; Prändl et al 1998). Third, we did some preliminary studies using RNA from control and heat stress cell cultures and leaves (A. Ganguli, unpublished data). Among the 21 Hsfs, 15 are represented in Table 4, whereas Hsfs A3, A6a, A6b,

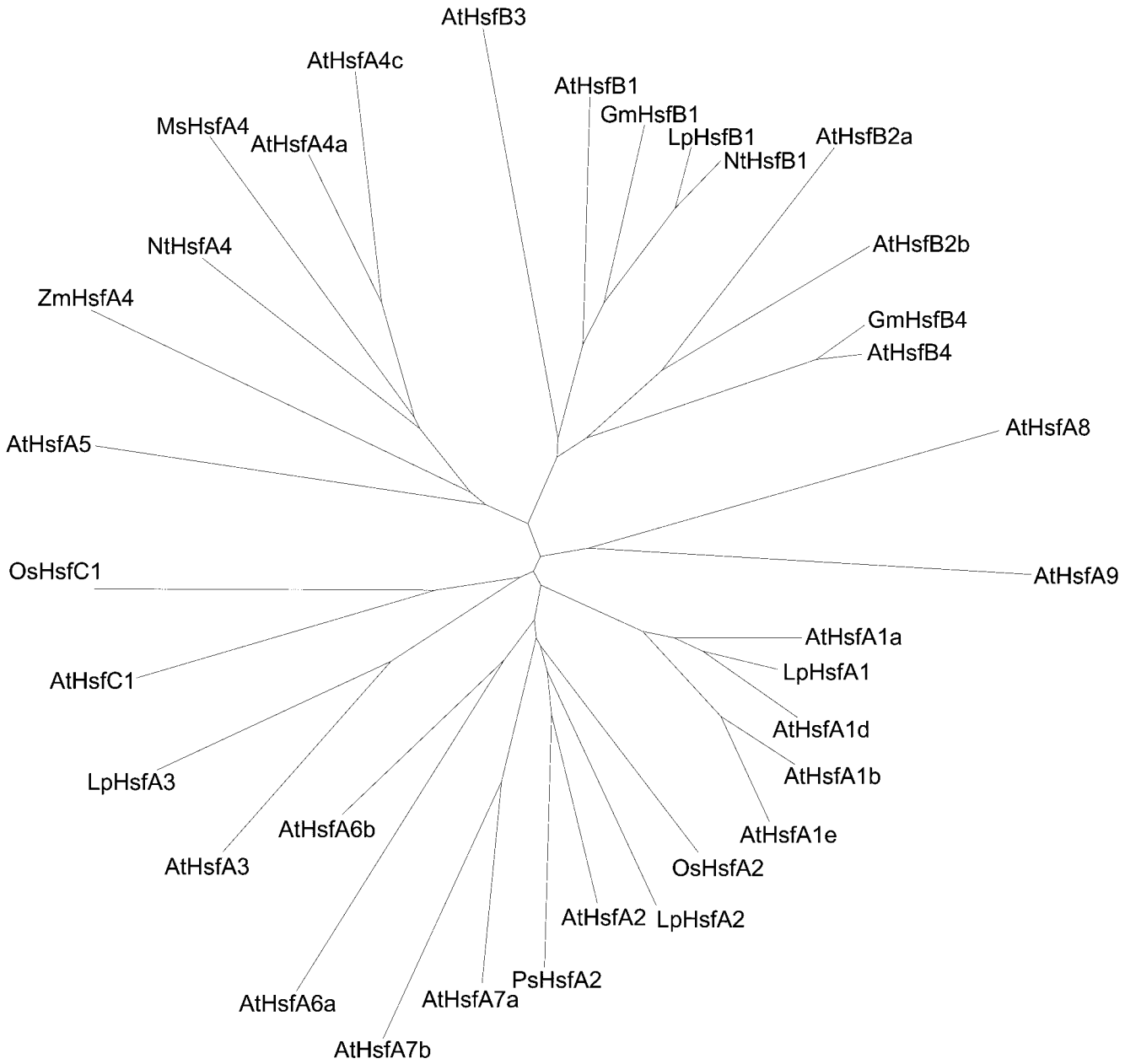
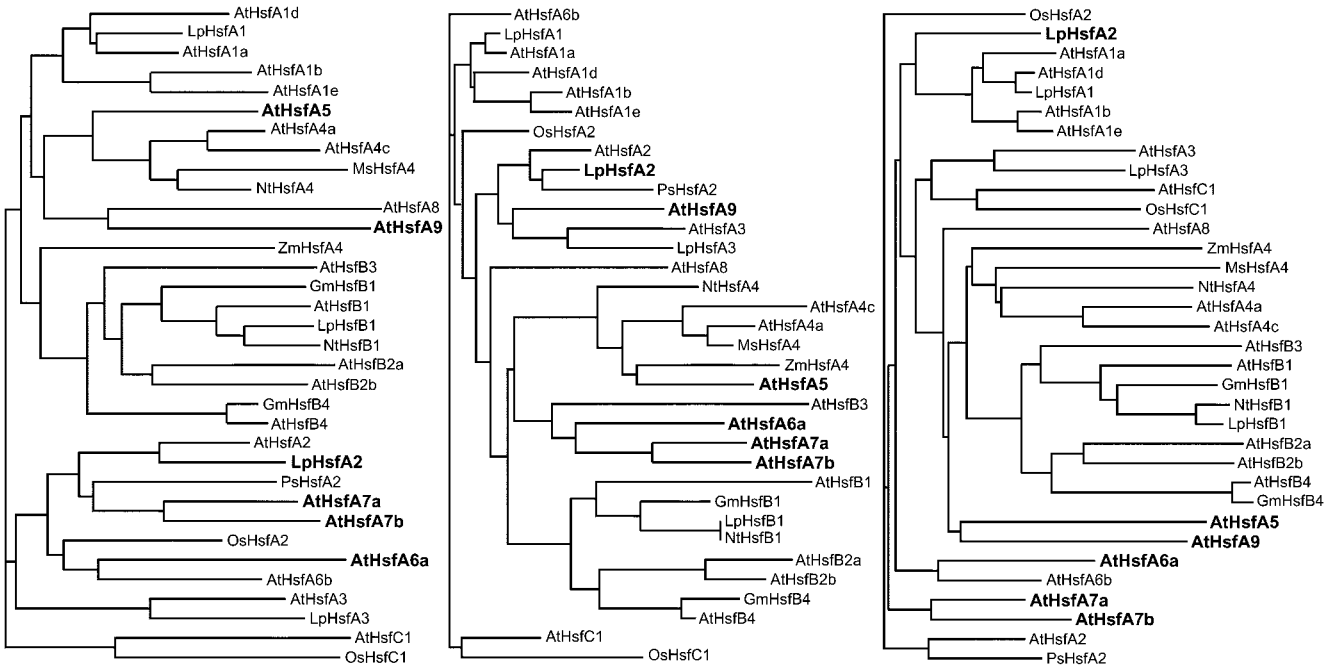


Fig 3. Phylogenetic relationship of Hsfs based on amino acid sequence comparison of the DBDs and HR-A/B regions (Clustal analysis). The consensus tree for all Hsfs was elaborated using the Clustalx 1–8_msw and Tree view software. For names, accession numbers, and identification of the Hsfs, see Table 1 and explanations given to this table. An extended version of this tree is available in the online version (Fig 5). It includes sequence information of many additional Hsfs from tomato, potato, barley, and soybean as derived from EST libraries.

A7b, B2a, and B3 could not be detected in any of the tissues or conditions analyzed so far. However, a severe drawback of the data from EST libraries is the lack of samples from heat stress tissues. This is best illustrated for HsfA2. Comparable to the situation with the tomato HsfA2, its messenger RNA was not detectable in control cells but was detected very strongly in heat stress cells. From this expression pattern, it is not surprising that no EST was found in the libraries created exclusively from RNA isolated from control tissues. Closer inspection of the putative promoter/5′-UTR regions indicate an intriguing pattern of HSE modules for all Hsf-encoding genes of *Arabidopsis* (see Fig 6 in the online version only). It was shown earlier (Treuter et al 1993; Czarnecka-Verner et al 2000) that, depending on the position of the HSE, upstream or downstream of the TATA box, Hsfs may activate or repress the transcription of the adjacent gene. It is tempting to speculate that the complex HSE patterns



Hsfs	DBD(N-terminal part)	DBD(C-terminal part)	HR-A/B
LpHsfA2	A2/A7	A2/A3/A9	A1
AtHsfA5	A4/A5	A4/A5	A5/A9
AtHsfA6a	A6	B3/A6a/A7	A6
AtHsfA7a/7b	A2/A7	B3/A6a/A7	A7
AtHsfA9	A8/A9	A2/A3/A9	A5/A9

Fig 4. Clustal analysis of the phylogenetic relationship based on the comparison of sequences of different parts of the N-terminal half of plant Hsfs. Three different conserved parts were used to create the phylogenetic trees: (1) the N-terminal part of the DBD until the position of the intron, (2) the C-terminal part of the DBD, and (3) the HR-A/B region, including 2 heptad repeats of HR-A, the insert, and HR-B (see borders defined in Table 2). Most Hsfs are found in identical groups irrespective of the sequence parts used for the analysis. However, a few Hsfs marked by boldfaced letters change their positions (see also the summary given at the bottom of the figure). We assume that, similar to the situation in the present mammalian *Hsf* genes, these 3 parts were separated by introns in the ancient plant *Hsf* precursor gene, ie, exon shuffling could have generated mosaic Hsfs. For AtHsfs A7a and A7b, the following situation is envisaged. The N-terminal part of the DBD and the HR-A/B region are derived from the putative *A2/A7* precursor gene, whereas the C-terminal part of the DBD stems are from the *B3/A7/A6a* precursor gene.

in the promoter/5′-UTR regions indicate a network of regulatory interactions between Hsfs and their encoding genes.

Phylogenetic analysis of plant Hsfs

Analyses of the amino acid sequences in the conserved N-terminal half of the Hsfs revealed 2 remarkable features (Table 2): (1) Two highly variable parts, ie, the flexible linker between DBD and HR-A and the insert connecting the HR-A and HR-B parts, are interspersed between the very conserved DBD and the 2 parts of the oligomerization domain. (2) The positions of most of the Hsfs in the phylogenetic tree (Fig 3) are fixed irrespective of the sequence parts used for the generation, ie, the N-terminal or C-terminal parts of the DBD, the whole DBD, the HR-A/B region, or the DBD plus HR-A/B region as used for the generation of the tree presented in Figure 3. However, there are a few exceptions to this rule, ie, Hsfs changing their position because they are evidently mosaic proteins formed from different phylogenetic parts (see examples and explanations given in Fig 4).

A possible explanation for this behavior can be found by comparing the exon-intron organization of mammalian and plant *Hsf* genes. As mentioned already, there is only one intron in the coding part of the DBD in all plant Hsfs. However, besides this conserved intron in the DBD, there are 11 additional introns in the mouse *hsf1* and *hsf*2 genes separating structural and functional modules (Zhang et al 1998; Manuel et al 1999). This exon-intron organization is particularly striking for the region encoding the DBD/HR-A/B part of these Hsfs. It is tempting to speculate that the position of introns in an ancient precursor gene of plant Hsfs is similar to the present mammalian genes. At least insertion of these hypothetical borders in the block diagram for the Lp-HsfA1 (Fig 1A) in-

dicates that separate exons may have generated the C-terminal part of the DBD, the flexible linker, the HR-A part plus insertion, and the HR-B part plus adjacent NLS region. This hypothetical exon-intron structure helps to understand the variability of sequence and length of the linker and of the insertion between HR-A and HR-B, making 6 amino acid residues for all class B Hsfs, 6 + 7 amino acid residues for HsfC1, and 6 + 21 amino acid residues for all class A Hsfs (Table 2). We further hypothesize that a type of exon shuffling and elimination of most of the introns in early times of the evolution of plant *hsf* genes contributed to the fixation of the present state and the generation of the mosaic type of Hsfs exemplified in Figure 4 for Lp-HsfA2 and At-Hsfs A5, A6a, A7a/A7b, and A9.

ACKNOWLEDGMENTS

Experimental work in our group was supported by grants from the Deutsche Forschungsgemeinschaft and by the Fonds der Chemischen Industrie.

REFERENCES

Aranda MA, Escaler M, Thomas CL, Maule AJ. 1999. A heat stress transcription factor in pea is differentially controlled by heat and virus replication. *Plant J* 20: 153–161.

Bharti K, Schmidt E, Lyck R, Bublak D, Scharf KD. 2000. Isolation and characterization of HsfA3, a new heat stress transcription factor of *Lycopersicon peruvianum. Plant J* 22: 355–365.

Bienz M, Pelham HRB. 1987. Mechanisms of heat-shock gene activation in higher eukaryotes. *Adv Genet* 24: 31–72.

Clos J, Westwood JT, Becker PB, Wilson S, Lambert U, Wu C. 1990. Molecular cloning and expression of a heaxameric *Drosophila* heat stress factor subject to negative regulation. *Cell* 63: 1085–1097.

Czarnecka-Verner E, Yuan CX, Fox PC, Gurley WB. 1995. Isolation and characterization of six heat stress transcription factor cDNA clones from soybean. *Plant Mol Biol* 29: 37–51.

Czarnecka-Verner E, Yuan CX, Scharf KD, Englich G, Gurley WB. 2000. Plants contain a novel multi-member class of heat stress factors without transcriptional activator potential. *Plant Mol Biol* 43: 459–471.

Damberger FF, Pelton JG, Harrison CJ, Nelson HCM, Wemmer DE. 1994. Solution structure of the DNA-binding domain of the heat stress transcription factor determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Protein Sci* 3: 1806–1821.

Döring P, Treuter E, Kistner C, Lyck R, Chen A, Nover L. 2000. Role of AHA motifs for the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. *Plant Cell* 12: 265–278.

Flick KE, Gonzalez L, Harrison CJ, Nelson HCM. 1994. Yeast heat stress transcription factor contains a flexible linker between the DNA-binding and trimerization domains—implications for DNA binding by trimeric proteins. *J Biol Chem* 269: 12475–12481.

Gagliardi D, Breton C, Chaboud A, Vergne P, Dumas C. 1995. Expression of heat stress factor and heat stress protein 70 genes during maize pollen development. *Plant Mol. Biol* 29: 841–856.

Görlich D, Kutay U. 1999. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* 15: 607–660.

Harrison CJ, Bohm AA, Nelson HCM. 1994. Crystal structure of the DNA binding domain of the heat stress transcription factor. *Science* 263: 224–227.

Heerklotz D, Döring P, Bonzelius F, Winkelhaus S, Nover L. 2001. The balance of nuclear import and export determines the intracellular distribution of tomato heat stress transcription factor HsfA2. *Mol Cell Biol* 21: 1759–1768.

Hübel A, Schöffl F. 1994. *Arabidopsis* heat stress factor: isolation and characterization of the gene and the recombinant protein. *Plant Mol Biol* 26: 353–362.

Kaneko T, Katoh T, Sato S, Nakamura A, Asamizu E, Tabata S. 2000. Structural analysis of *Arabidopsis thaliana* chromosome 3, II: sequence features of the 4,251,695 bp regions covered by 90 P1, TAC and BAC clones. *DNA Res* 7: 217–221.

Lin X, Kaul S, Rounsley SD, et al. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana. Nature* 402: 761–768.

Littlefield O, Nelson HCM. 1999. A new use for the 'wing' of the ''winged'' helix-turn-helix motif in the HSF-DNA cocrystal. *Nat Struct Biol* 6: 464–470.

Liu PCC, Thiele DJ. 1999. Modulation of human heat stress factor trimerization by the linker domain. *J Biol Chem* 274: 17219–17225.

Lyck R, Harmening U, Höhfeld I, Treuter E, Scharf KD, Nover L. 1997. Intracellular distribution and identification of the nuclear localization signals of two plant heat-stress transcription factors. *Planta* 202: 117–125.

Manuel M, Sage J, Mattei MG, Morange M, Mezger V. 1999. Genomic structure and chromosomal localization of the mouse Hsf2 gene and promoter sequences. *Gene* 232: 115–124.

Morimoto RI. 1998. Regulation of the heat stress transcriptional response: cross talk between family of heat stress factors, molecular chaperones, and negative regulators. *Genes Dev* 12: 3788–3796.

Nakai A. 1999. New aspects in the vertebrate heat stress factor system: HsfA3 and HsfA4. *Cell Stress Chaperones* 4: 86–93.

Nover L. 1987. Expression of heat stress genes in homologous and heterologous systems. *Enzyme Microb Technol* 9: 130–144.

Nover L. 1991. *Heat Shock Response.* CRC Press, Boca Raton, FL.

Nover L, Scharf KD. 1997. Heat stress proteins and transcription factors. *Mol Cell Life Sci* 53: 80–103.

Nover L, Scharf KD, Gagliardi D, Vergne P, Czarnecka-Verner E, Gurley WB. 1996. The Hsf world: classification and properties of plant heat stress transcription factors. *Cell Stress Chaperones* 1: 215–223.

Peteranderl R, Nelson HCM. 1992. Trimerization of the heat stress transcription fact factor by a triple-stranded alpha-helical coiled-coil. *Biochemistry* 31: 12272–12276.

Peteranderl R, Rabenstein M, Shin Y, Liu CW, Wemmer DE, King DS, Nelson HCM. 1999. Biochemical and biophysical characterization of the trimerization domain from the heat stress transcription factor. *Biochemistry* 38: 3559–3569.

Prändl R, Hinderhofer K, Eggers-Schumacher G, Schöffl F. 1998. Hsf3, a new heat stress factor from *Arabidopsis thaliana,* derepresses the heat stress response and confers thermotolerance when overexpressed in transgenic plants. *Mol Gen Genet* 258: 269–278.

Rabindran SK, Giorgi G, Clos J, Wu C. 1991. Molecular cloning and expression of a human heat stress factor. *Proc Natl Acad Sci U S A* 88: 6906–6910.

Sarge KD, Zimarino V, Holm K, Wu C, Morimoto RI. 1991. Cloning and characterization of two mouse heat stress factors with dis-

tinct inducible and constitutive DNA binding ability. *Genes Dev* 5: 1902–1911.

Sato S, Nakamura Y, Kaneko T, Katoh T, Asamizu E, Kotani H, Tabata S. 2000a. Structural analysis of *Arabidopsis thaliana* chromosome 5, X: sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones. *DNA Res* 7: 31–63.

Sato S, Nakamura Y, Kaneko T, Katoh T, Asamizu E, Tabata S. 2000b. Structural analysis of *Arabidopsis thaliana* chromosome 3, I: sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones. *DNA Res* 7: 131–135.

Scharf KD, Heider H, Höhfeld I, Lyck R, Schmidt E, Nover L. 1998a. The tomato Hsf system: HsfA2 needs interaction with HsfA1 for efficient nuclear import and may be localized in cytoplasmic heat stress granules. *Mol Cell Biol* 18: 2240–2251.

Scharf KD, Höhfeld I, Nover L. 1998b. Heat stress response and heat stress transcription factors. *J Biosci* 23: 313–329.

Scharf KD, Rose S, Thierfelder J, Nover L. 1993. Two cDNAs for tomato heat stress transcription factors. *Plant Physiol* 102: 1355–1356.

Scharf KD, Rose S, Zott W, Schöffl F, Nover L. 1990. Three tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF. *EMBO J* 9: 4495–4501.

Schöffl F, Prändl R, Reindl A. 1998. Regulation of the heat stress response. *Plant Physiol* 117: 1135–1141.

Schuetz TJ, Gallo GJ, Sheldon L, Tempst P, Kingston RE. 1991. Isolation of a cDNA for HSF2: evidence for two heat stress factor genes in humans. *Proc Natl Acad Sci U S A* 88: 6911–6915.

Schultheiss J, Kunert O, Gase U, Scharf KD, Nover L, Rüterjans H. 1996. Solution structure of the DNA-binding domain of the tomato heat stress transcription factor HSF24. *Eur J Biochem* 236: 911–921.

Shoji T, Kato K, Sekine M, Yoshida K, Shinmyo A. 2000. Two types of heat stress factors in cultured tobacco cells. *Plant Cell Rep* 19: 414–420.

Sorger PK, Pelham HRB. 1988. Yeast heat stress factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell* 54: 855–864.

Treuter E, Nover L, Ohme K, Scharf KD. 1993. Promoter specificity and deletion analysis of three heat stress transcription factors of tomato. *Mol Gen Genet* 240: 113–125.

Vuister GW, Kim SJ, Orosz A, Marquardt J, Wu C, Bax A. 1994. Solution structure of the DNA-binding domain of *Drosophila* heat stress transcription factor. *Nat Struct Biol* 1: 605–614.

Wiederrecht G, Seto D, Parker CS. 1988. Isolation of the gene encoding the *S. cerevisiae* heat stress transcription factor. *Cell* 54: 841–853.

Wu C. 1995. Heat stress transcription factors. *Annu Rev Cell Biol* 11: 441–469.

Yuan CX, Gurley WB. 2000. Potential targets for HSF1 within the preinitiation complex. *Cell Stress Chaperones* 5: 229–242.

Zhang Y, Koushik S, Dai RJ, Mivechi NF. 1998. Structural organization and promoter analysis of murine heat stress transcription factor-1 gene. *J Biol Chem* 273: 32514–32521.

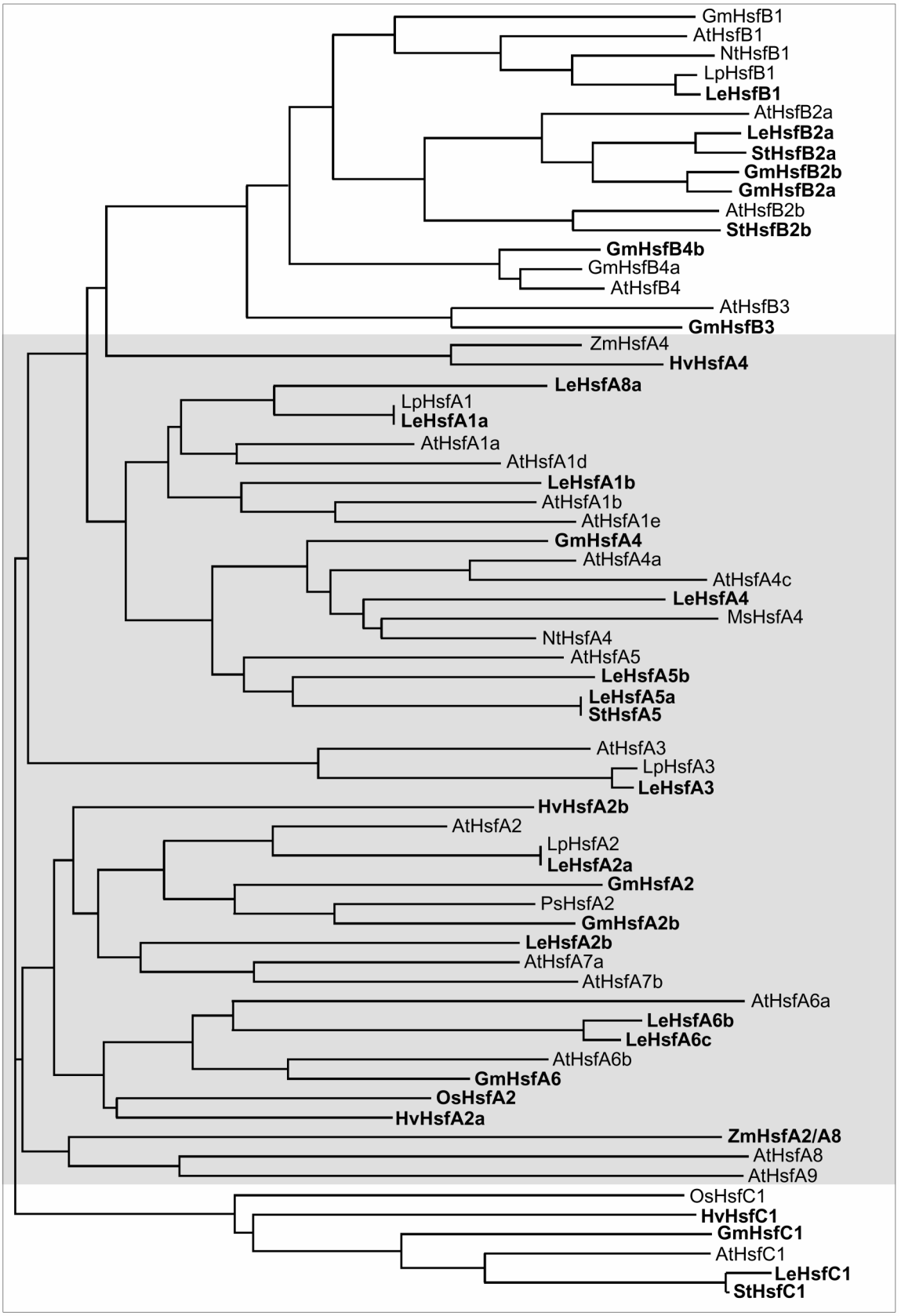


Fig 5. Extended version of phylogenetic relationships of plant Hsfs. To emphasize the complexity of the plant Hsf family in general, an extended version of the phylogenetic tree was created by Clustal analysis based on the N-terminal parts of the DBD of those Hsfs contained in Fig 3 plus a considerable number of additional partial clones derived from the database, mostly from EST libraries of tomato (*Lycopersicon esculentum*), potato (*Solanum tuberosum*), soybean (*Glycine max*), and barley (*Hordeum vulgare*). For other abbreviations see Fig 3 and Table 1. New entries in the figure are marked by boldfaced letters. Information was derived from the following ESTs and accession numbers: *L esculentum* HsfA1a (AW933448, AW399336, AW223123), HsfA1b (BE354387), HsfA2a (AW034874), HsfA2b (AW930998), HsfA3 (BE433610, AI895834, AW034135, AW035854, AW035844, AW030642, AW033013; all ESTs represent incompletely spliced messenger RNAs), HsfA4 (AW038959, AW933529), HsfA5a (AW217982, AW041695, AW030725, BF096782), HsfA5b (AW034402), HsfA6b (AW036683, AW932142, AW222011, BE 434585, BE433803, AI895294, AI489721, BG132247), HsfA6c (BG351853), HsfA8a (AW738023, mosaic Hsf, whose annotation to the HsfA8 group depends on the C-terminal part of the DBD and the HR-A/B region), HsfA8b (AW931892, not included, because only C-terminal part with HR-A/B region available), HsfB1 (BF097217, BG134658, AI895934), HsfB2a (AW931781, AW220758, AW931176), HsfC1 (AW738534, AW979619, BE451302, AW649243); *G max* HsfA2a (Hsf21 fragment, Z46952), HsfA2b (AW164509), HsfA4 (BE611683, AW756148, BG405291, BE330669), HsfA6 (BG041837), HsfB2a (BE346810, BF067962), HsfB2b (AW703969), HsfB3 (BE019974), HsfB4b (BF597135), HsfC1 (BE347442, AW596493, BG352891); *S tuberosum* HsfA5 (BF459947), HsfB2a (BE473183), HsfB2b (BF052865), HsfC1 (BF459641); *H vulgare* HsfA2a (BE603513), HsfA4 (BE216310), HsfA2b (BF264338), HsfC1 (BF616419); *Zea mays*: ZmHsfA2/A8 (Hsfa fragment, S61458).

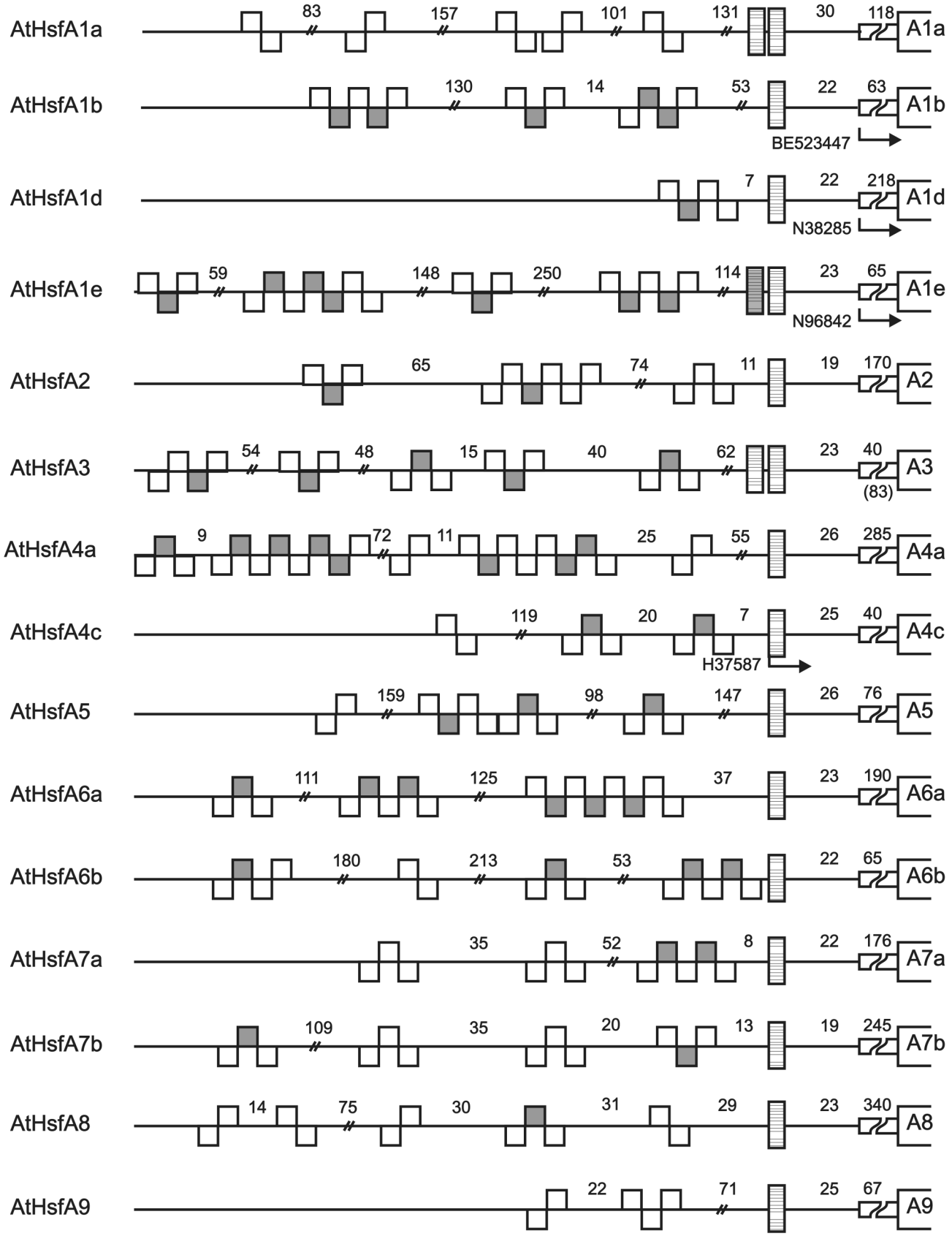


Fig 6. Promoter architecture of *Arabidopsis Hsf* genes with respect to the positioning of HSE clusters. For orientation and positioning of the putative TATA box, we used the well-defined start site of the open reading frames, data for genomic clones with experimentally defined transcription start sites, and the starting points of the indicated ESTs. Symbols are explained at the bottom of the figure. Numbers on the left refer to the corresponding numbers in Table 1.

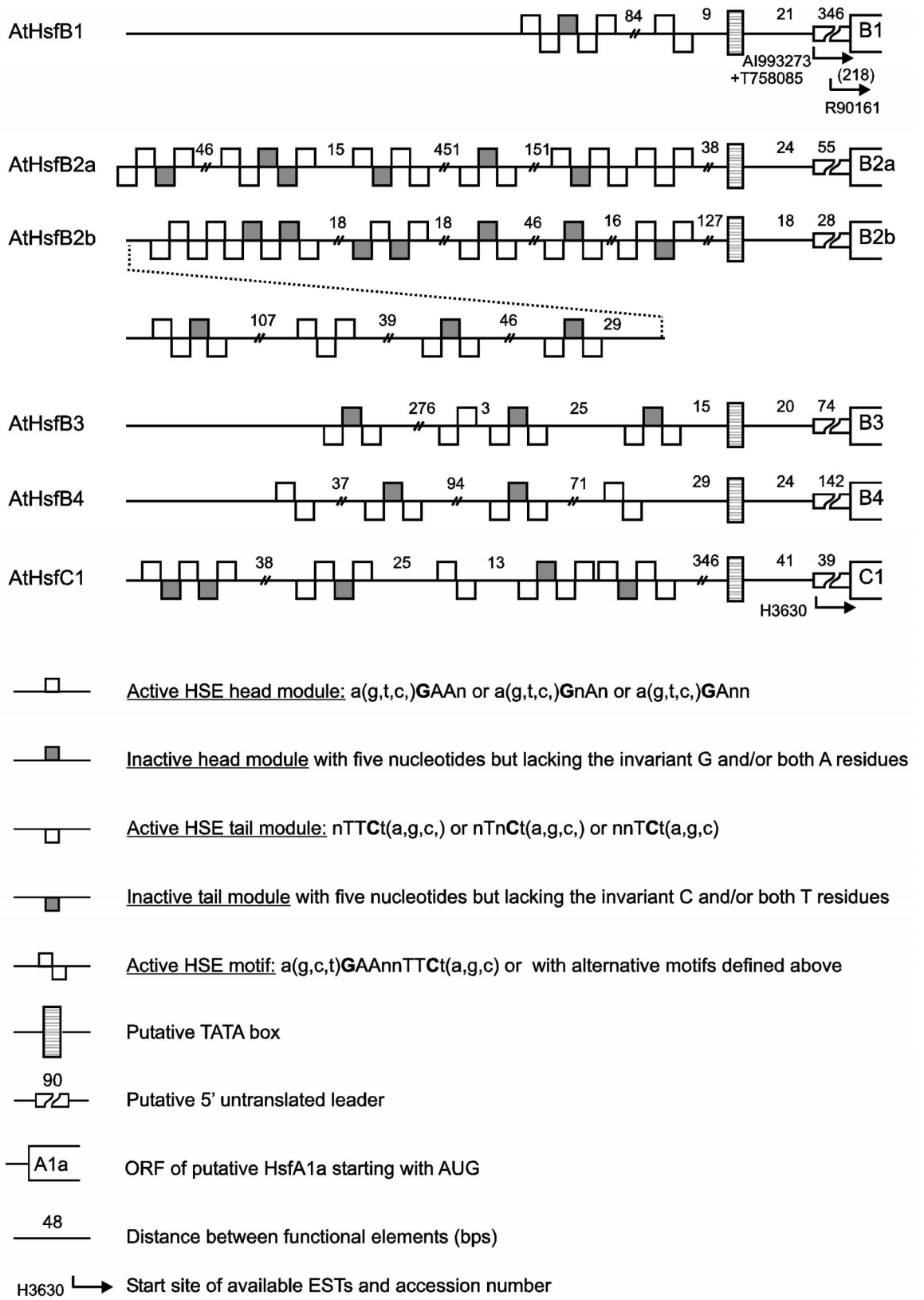


Fig 6. Continued.