Title	Molecular characterization and in silico expression analysis of a chalcone synthase gene family in Sorghum bicolor
Author(s)	Lo, C; Coolbaugh, RC; Nicholson, RL
Citation	Physiological And Molecular Plant Pathology, 2002, v. 61 n. 3, p. 179-188
Issued Date	2002
URL	http://hdl.handle.net/10722/48505
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Molecular Characterization and *in silico* Expression Analysis of a Chalcone Synthase Gene Family in *Sorghum bicolor*

Clive Lo[1, 2], Ronald C. Coolbaugh[1], and Ralph L. Nicholson[1]*

[1]*Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA and* [2]*Department of Botany, The University of Hong Kong, Pokfulam Road, Hong Kong*

*To whom all correspondence should be addressed. E-mail: nicholson@btny.purdue.edu.

Key words: chalcone synthase; *Sorghum bicolor*; *in silico* gene expression analysis

ABSTRACT

Recent use of *Sorghum bicolor* as a target for grass genomics has presented new resources for gene discovery in novel metabolic pathways in Poaceae. Sorghum synthesizes a unique class of flavonoid phytoalexins, the 3-deoxyanthocyanidins, in response to fungal infection. The biosynthetic pathways for 3-deoxyflavonoids are largely uncharacterized but are known to involve transcriptional activation of chalcone synthase (CHS). CHS, or naringenin CHS, catalyzes the formation of naringenin, the precursor for different flavonoids. We have isolated seven sorghum CHS genes, CHS1 to 7, from a genomic library on high-density filters. CHS1 to 7 are highly conserved and closely related to the maize C2 and Whp genes. Several of them are also linked in the genome. These findings suggest that they are the result of recent gene duplication events. Expression of the individual CHS genes was studied *in silico* by examination of expressed sequence tag (EST) data available in the public domain. Our analyses suggested that CHS1-7 were not differentially expressed in the various growth and developmental conditions represented by the cDNA libraries used to generate the EST data. However, we identified a CHS-like gene, CHS8, with significantly higher EST abundance in the pathogen-induced library. CHS8 shows only 81-82% identity to CHS1 to 7 and forms a distinct subgroup in our phylogenetic analysis. In addition, the active site region contains substitutions that distinguish CHS8 from naringenin CHS. We propose that CHS8 has evolved new enzymatic functions that are involved in the synthesis of defense-related flavonoids, such as the 3-deoxyanthocyanidins, during fungal infection.

INTRODUCTION

Sorghum (*Sorghum bicolor* L. Moench) is the fifth-ranking grain crop in the world and is well known for its adaptability to adverse growing conditions. Because of its small genome size (760 Mb) and its relatedness to maize, sorghum is considered a logical complement to rice as a genome model for cereals [12]. Genomics resources for sorghum, such as genomic libraries on high-density filters [45], integrated physical and genetic maps [12], and large collections of expressed sequence tag (EST) databases, are being developed by different research institutes. Recently, stable transformation of sorghum using *Agrobacterium* has been demonstrated to be feasible [47]. All these resources are extremely useful for rapid discovery of genes that influence specific traits in sorghum. We are interested in sorghum because it synthesizes a unique class of flavonoid phytoalexins, the 3-deoxyanthocyanidins, as an essential component of defense mechanisms against pathogen infection [28, 39]. Flavonoids are plant secondary metabolites that are also important in pigmentation [20], UV protection [36], induction of pollen tube germination and growth [30], and inhibition of auxin transport [22]. The 3-deoxyanthocyanidins are structurally similar to the anthcyanidin aglycones that accumulate in response to light in mesocotyls of sorghum juvenile seedlings. In sorghum, accumulation of both 3-deoxyanthocyanidins and anthocyanidins is preceded by accumulation of transcripts encoding chalcone synthase (CHS) (EC 2.3.1.74), a key enzyme in flavonoid biosynthesis [19, 27].

CHS, or naringenin CHS, is a plant specific polyketide synthase that catalyzes the condensation of three units of malonyl-CoA with *p*-coumaroyl-CoA to form naringenin chalcone. This reaction is generally regarded as the committed step leading to the synthesis of different flavonoid compounds [11]. Several CHS-like enzymes, such as stilbene (STS), acridone synthases, and pyrone synthases, are believed to have evolved independently from naringenin CHS in related plant species [37]. Recently, five CHS-related proteins with different substrate preferences have been identified in the most primitive vascular plant *Psilotum nudum* [46]. The activities of CHSs are largely regulated at the level of transcription. For example, some CHS genes contain light-responsive elements in their promoter regions [17, 38]. In addition, expression of CHS genes is controlled by developmental stages of the plant, as well as biological and environmental stresses, such as infection and wounding [10, 43]. In higher plants, CHS is encoded by a family of genes. Individual CHS genes in legumes often exhibit differential expression patterns that could be correlated to the types of flavonoid compounds being synthesized. In sorghum, CHS mRNA accumulation in mesocotyls is induced by light, just as the synthesis of anthocyanin pigments is light mediated [27]. In contrast, synthesis of 3-deoxyanthocyanidins is independent of light, and CHS mRNA accumulation was detected in mesocotyl tissues following fungal inoculation, regardless of whether the plants were maintained in the light or the dark [44]. These results suggested the presence of a CHS gene family in sorghum, the members of which are regulated by different stimuli and have specific roles in plant defense.

Studies of flavonoid metabolism in monocots have been focused primarily on anthocyanin biosynthesis in maize [40}. In contrast, the biosynthesis of 3-

deoxyflavonoids is not well understood. Accumulation of phlobaphenes (polymerized forms of 3-deoxyflavonoids) is controlled by P1 and Y genes in maize and sorghum, respectively [7]. However, structural genes involved in the 3-deoxyflavonoid biosynthesis have not been identified. We take advantage of the new genomic resources to dissect the unknown 3-deoxyanthocyanindin pathways in sorghum. In this paper, we report the isolation of a family of seven naringenin CHS genes, CHS1 to 7, from *S. bicolor*. The expression of the individual CHS genes was studied by analysis of the sorghum EST databases released by the Comparative Grass Genome Center (CGGC, University of Georgia). Examination of the EST databases also led to the discovery of a CHS-related gene, CHS8, which is preferentially expressed in infected seedlings. The significance of CHS8 is discussed in the context of CHS evolution and phytoalexin biosynthesis in sorghum.

MATERIALS AND METHODS

Genomic cloning of sorghum CHS genes

The high-density filters containing a genomic BAC library of *S. bicolor* (cultivar BTx623) [45] were obtained from the BAC Workshop at Texas A&M University (College Station, TX). The library was screened with a ^{32}P-labeled CHS probe by hybridization. The probe was a 620 bp PCR fragment derived from a sorghum CHS gene (U51569) [9]. Labeling reaction was performed with (α-^{32}P)-dCTP (Amersham Pharmacia, Newark, NJ) using a Deca-Prime labeling kit (Ambion, Austin, TX). The filters were pre-hybridized for 15 min at 65 $^{\circ}$C in 0.5 M sodium phosphate buffer (pH 7.0) containing 7% (w/v) SDS, and was hybridized for 18 h in fresh buffer with the

addition of the denatured probe. The hybridized filters were washed three times in 0.5 X SSC (Na citrate, 7.5 mM ; NaCl, 75 mM), 0.5% SDS at 65 °C, and were exposed to X-ray films with intensifying screens at -70 °C.

Three positive clones, E16, C19, and G18, were identified and the corresponding stock cultures were ordered from the BAC workshop. DNA was prepared from the BAC clones following the recommended protocol available at the web site of the BAC workshop (http://hbz.tamu.edu/bac.html). To isolate CHS genes from the BAC clones, DNA samples were digested with either *Hin*d III or *Xba* I and cloned into pGEM-3Z plasmid vectors (Promega, Madison, WI). CHS plasmid clones were screened by colony hybridization using the CHS probe. DNA sequencing was performed at the DNA Sequencing Facility at Iowa State University (Ames, IA). Seven independent CHS genomic clones, CHS1-7, with complete coding sequence were obtained (Table I). CHS1 to 3 were isolated from BAC clone E16 as *Hin*d III fragments. CHS4 to 5 and CHS 6 to 7 were isolated from BAC clones C19 and G18, respectively, as *Xba* I fragments. Computer analyses of DNA and protein sequences were performed using Lasergene software for Windows (DNASTAR, Inc., Madison, WI).

In silico Analysis of Sorghum EST Databases

S. bicolor (cultivar BTx623) EST sequences generated from various non-normalized cDNA libraries were deposited in dbEST, GenBank by the CGGC. Information on the cDNA libraries (Table II) was provided by L. Pratt (University of Georgia, Athens, GA). Virtual transcripts of CHS 1-7, consisting of sequences

downstream of the TATA box through the polyadenylation signal (AATAA) without the intronic regions, were used to perform BLASTN searches [2] against the sorghum EST databases. Sequences with an Expect value [2] smaller than 1.0×10^{-10} were extracted from dbEST (Table III). The ESTs were assembled with the CHS virtual transcripts using the SeqMan II program in Lasergene software with a minimum match size of 200 bp and a minimum sequence identity of 98%. Under these conditions, the CHS virtual transcripts formed individual contiguous sequences with their matching ESTs. The CHS gene identity for each EST was then assigned. The *S. bicolor* cDNA clones were sequenced from both 5'-and 3'-ends at CGGC and the corresponding EST sequences are in separate files in dbEST. The records of the EST files were examined electronically to identify sequences derived from the same EST clone. Thus, a clone sequenced from 5' and 3' ends was considered only once in our analyses. Differential expression of a particular gene was inferred statistically by pairwise comparison of its EST abundance in the cDNA libraries using Equation 2 in Audic and Claverie (1997) [5]. A computer program is also available on their website for statistical analysis of digital transcription profiles data (http://igs-server.cnrs-mrs.fr/~audic/significance.html)

RESULTS

Isolation and Features of Sorghum CHS Genes

A CHS PCR-fragment was used to screen a genomic library of *S. bicolor* (cv BTx623) in bacterial artificial chromosome (BAC) vectors on high density filters [45]. The probe was derived from a CHS gene in sorghum [9]. Seven CHS genes, CHS1-7, with complete coding sequences were isolated from three BAC clones, E16, C19, and

G18 (Table I). Since BTx623 is a sorghum inbred line [45], the different CHS genomic clones represent members of a gene family instead of allelic sequences. Presence of multiple CHS genes in the same BAC clones (Table I) suggests that these genes occur in clusters in the genome. Nucleotide sequences for CHS1 to 7 were deposited in GenBank under the accession numbers AF152548 to AF152554. BLASTN [2] searches against the GenBank nucleotide database revealed a high degree of homology between the sorghum CHS genes and a maize CHS (C2) gene (GenBank accession no. X60205). The coding regions of CHS1 to 7 were deduced by sequence comparison with C2. Alignment of the nucleotide sequences of the sorghum CHS genes and C2 revealed the presence of two conserved regions of homology (87-90% identity) that corresponded to the C2 coding sequence (Fig. 1). CHS1 to 7 genes share at least 91% sequence identity in their coding regions. CHS1 to 6 encode predicted polypeptides of 401 amino acids. There is a deletion of the gly-396 residue in the gene product of CHS7. Amino acid sequence identity among CHS1 to 7 is at least 97.5%. The maize C2 protein is at least 87% identical to the gene products of CHS1 to 7 (Table I).

A TATA box-like sequence upstream of the start codon (ATG) and a potential polyadenylation signal (AATAA) downstream of the stop codon (TGA) were identified in each of the sorghum CHS genes (Fig. 1). The coding regions of CHS1 to 7 are all interrupted at a TG(T/C) codon (Cys-64) by single introns (Fig. 1). An intron at this position is also present in the maize C2 gene and CHS genes from other plants such as barley (X58339), *Arabidopsis* (AF144533), and soybean (L07647). The sizes of the introns range from150 bp in CHS1 to over 1000 bp in CHS5 and 6 (Table I). There is

considerable sequence identity, not only in the coding regions, but also in the 5' and 3'-untranslated regions in CHS1 to 7. The major differences in these regions are additions or deletions instead of significant stretches of sequence dissimilarity (data not shown).

In silico Expression Analysis of the CHS Gene Family in Sorghum

S. bicolor EST data recently deposited in GenBank (dbEST) by the CGGC were used to perform *in silico* expression analysis of the individual sorghum CHS genes. The origins of the major *S. bicolor* cDNA libraries represented in the EST databases are shown in Table II. As of October 25, 2001, each of the cDNA libraries in dbEST was represented by 5500 to 7000 EST clones. Sorghum CHS EST sequences (Table III) were identified by BLASTN searches against dbEST. Many of the CHS ESTs contain enough sequence information to confirm of the predicted intron positions in the individual CHS genes (data not shown). The availability of CHS1-7 genomic sequences allowed us to accurately assign gene identities to most of the CHS ESTs in the different cDNA libraries (Table III). Careful examination of the cDNA library data set revealed that some of the ESTs differed from CHS1 to 7. Eight ESTs from the PI1 library and one EST from the LG1 library were less than 90% identical to the CHS1 to 7 sequences (data not shown). We assembled these EST sequences using a minimum match size of 50 bp and a minimum identity of 98%. A consensus sequence was obtained and termed CHS8 hereafter.

The majority of the CGGC sorghum EST data have been assembled into tentative consensus sequences (TCs) by the Institute for Genome Research (TIGR). The sorghum

TCs are annotated and can be accessed through the *S. bicolor* Gene Index (SbGI, http://www.tigr.org/tdb/sbgi/). A total of six TCs were retrieved from SbGI when a text search with "chalcone synthase" was performed. At TIGR, TCs are assembled based on two or more ESTs that overlap for at least 40 bases with at least 95% sequence identity (http://www.tigr.org/tdb/tgi_info.html). However, these conditions are not likely to be stringent enough to accurately assign the CHS ESTs to individual members of the gene family. For example, TC14662 was annotated as both CHS2 and CHS6 while TC14665 was annotated as both CHS3 and CHS5 in SbGI (data not shown). The SbGI TC numbers were examined in the all CHS EST sequences (Table III). TC14664 and TC14665 are chimeric contigs containing ESTs from two different CHS genes. On the other hand, CHS2 ESTs were represented by two different TC numbers in the SbGI.

Our EST abundance analysis indicated that the LG1, DG1, and PI1 databases contained approximately equal numbers of total CHS ESTs (Table IV). There was only one CHS EST in each of the WS1, IP1, EM1 databases. No CHS EST was detected in the OV1&2 databases. Since the sorghum EST data sets are derived from non-normalized libraries (L. Pratt, personal communication), the number of ESTs for a given gene also reflects the abundance of its mRNA in the population used to prepare a particular cDNA library [32]. Audic and Claverie established rigorous statistical tests for identifying differentially expressed genes by pair-wise comparison of EST abundance [5]. In the different sorghum cDNA libraries, EST counts for CHS1 to 7 were between 0 to 4 (Table IV). At a confidence level of 95%, these values were not significantly different from 0 and were considered basal levels of expression [5, 29]. In contrast, the number of

ESTs (8) for the newly identified CHS8 gene in the PI1 library (Table IV) were significantly different from the values observed in other libraries at a confidence level of 95%. Thus, our data suggested that CHS8 is differentially expressed in 2-week-old seedlings inoculated with the anthracnose pathogen *Colletotrichum graminicola* (PI1 library, Table II), compared to the uninoculated tissues represented by other EST libraries analyzed in this study.

Features of the Sorghum CHS8 Gene

All CHS8 ESTs were assembled into the same consensus sequence, TC20269, in SbGI (Table III). We present the reverse complementary sequence of TC20269 identified with an open reading frame of 1194 bp in Fig. 2. A putative polyadenylation signal, AATAA, is located downstream of the stop codon. The TGT codon, which is interrupted in CHS1 to 7, is found in the same position in CHS8 (Cys-64). The translated regions of CHS8 shows 77-79% and 81-82% sequence identity to the coding sequences of CHS1 to 7 at nucleotide and amino acid levels, respectively. The sequence of TC20269 was confirmed by sequencing of a near full-length CHS8 cDNA clone, PI1_67_F08, obtained from the CGGC. The cDNA sequence (GenBank Accession Number AY069951) is identical to TC20269 except that 21 and 34 bases are missing from the 5'- and 3'-untranslated regions of this clone, respectively.

Comparison and Phylogenetic Analysis of CHS and CHS-related proteins

The deduced protein sequences of sorghum CHS1-7 and CHS8 were compared to those of selected CHS and CHS-related proteins from different plant species (Fig. 3).

Alignment of these enzymes revealed a number of strictly conserved amino acids including Cys-168 (sequence numbering based on sorghum CHS1 protein), His-306, Asn-340, Pro-380 (data not shown). These amino acid residues are located in the four loops that largely define the active site of CHS and related enzymes [42]. Cys168 was identified as the active site nucleophile in CHS and stilbene synthase [26] and forms an ionic pair with His-306 [42]. The G^{379}PGPG loop contains a proline residue that provides rigidity and the unique configuration of this loop[41]. The putative active site cysteine residues in CHS1 to 7 are located in the strictly conserved environment, MMYQQGCFAGGTVLR (Fig. 3), as in other naringenin CHSs in angiosperms [26]. In CHS8, there are two notable differences in the amino acid residues in this putative active site environment: the Gln to His and the Thr to Met substitutions in positions -3 and +5 relative to the Cys residue, respectively (Fig. 3). Other CHS-like proteins, including barley HvCHS2, *Gerbera* pyrone synthase, and stilbene synthases, contain different substitution events in their active site regions.

The evolutionary relationships among these CHS and CHS-related proteins are presented in Figure 4. Proteins from the same species or the same plant family generally clustered together in the phylogenetic tree. Among the monocot proteins, barley HvCHS2 formed a distinct monophyletic group. Sorghum CHS1-7 proteins are closely related to the two CHS proteins from maize. CHS8, however, lies in a separate subgroup away from CHS1 to 7 and the maize proteins.

DISCUSSION

In this study, we identified a family of eight CHS genes in sorghum: CHS1 to 7 (putative naringenin CHS genes) by genomic cloning and CHS8 (CHS-like gene) by analysis of the public EST databases. In higher plants, CHS appears to exist frequently as a family of genes. *Petunia hybrida* [24] and several leguminous plants such as soybean [1] and bean [35] contain 8 to 10 CHS genes in their genomes. However, only single copies of CHS genes were found in parsley [18] and *Arabidopsis* [14]. Members of the grass family appear to have a small number of CHS genes. For example, there are two CHS genes in maize, C2 and Whp, and they are located on different chromosomes [15]. Sorghum and maize may have diverged from a common ancestor 16.5 million years ago [16]. Extensive conservation of gene content and gene order has been observed in their genomes [21]. The maize genome, which is approximately 2,500 Mb in size [4], contains more gene duplications and repetitive sequences than the 760-Mb sorghum genome [34]. Apparently, CHS genes represent the small portion of genes that are more abundant in sorghum. The clustering of the highly conserved CHS1 to 7 genes suggests that they have been generated by recent gene duplications long after sorghum and maize diverged from their common ancestor.

A common feature of gene families is that individual members are differentially regulated in response to developmental and environmental signals. Traditionally, detection of the expression of specific members of a gene family has been performed using gene-specific oligonucleotides in northern analyses or RNase protection assays [23, 24]. However, design of such probes depends largely on sequence divergence in the gene

family. This approach has been shown to be laborious and ineffective for the CHS1 to 7 genes in sorghum (data not shown). Instead, we took advantage of the large collections of *S. bicolor* EST data released recently by the CGGC to examine the expression of the individual CHS genes in sorghum. The gene identities of the CHS ESTs can be accurately determined by direct comparison to the DNA sequences of individual sorghum CHS genomic clones (Table III).

We did not observe differential expression of specific CHS genes in certain growth stages or conditions that are represented by most of the EST libraries in this study. For example, none of the sorghum CHS genes were differentially expressed in 10-14-day-old light grown seedlings (LG1 library, Table IV). Light-responsive elements, such as Box I and Box II, have been identified in the promoter regions of CHS genes from a number of plants including *Arabidopsis* [17] and parsley [38]. However, light-induced expression of flavonoid genes was also regulated temporally in *Arabidopsis*. Thus, CHS, chalcone isomerase, and dihydroflavonol reductase were only transiently expressed in 3-day-old seedlings but not in 7-day old seedlings [25]. Similarly, we detected transient accumulation of CHS mRNA in mesocotyls of 4-day-old etiolated seedlings of sorghum upon illumination [27], although it was not clear which specific CHS genes were involved. Flavonoid biosynthesis is one of the major metabolic activities during flower development. However, CHS genes did not appear to be up-regulated in sorghum immature panicles, as revealed from the EST abundance data (IP1, Table IV). It is possible that transcriptional regulation of flavonoid biosynthesis occurred downstream of CHS during flower development in sorghum.

Our *in silico* expression analysis revealed the presence of a CHS-related gene, CHS8, in sorghum. CHS8 is only 81-82% identical to CHS1 to 7 and appeared to be more distantly related in the phylogenetic analysis (Fig. 4). These findings suggested that CHS8 have been duplicated from the ancestral form of CHS1 to 7 and diverged in protein coding sequence. Gene duplication followed by sequence divergence has been used as a model to explain how novel biochemical functions have arisen from CHSs. For example, STSs and pyrone synthases are believed to have evolved recently from CHS genes in related plants [37]. STS catalyzes the biosynthesis of the backbone of stilbene phytoalexins using *p*-coumaroyl CoA and three units of malonyl CoA, the identical substrates for CHS. Pyrone synthase is a novel class of CHS-like enzymes recently identified in *Gerbera hybrida* [13]. The enzyme uses acetyl-CoA to perform two condensation reactions with malonyl-CoA, instead of three as in CHS, to form the pyrone backbone of gerberin and parasorboside. In barley, the expressed HvCHS2 protein showed an unusual substrate preference for feruloyl-CoA and caffeoyl-CoA, resulting in the formation of homoeriodictyol and eriodictyol, respectively, in *in vitro* enzyme assays [8]. HvCHS2 was strongly induced upon inoculation with the powdery mildew pathogen *Blumeria graminis* [8]. However, phytoalexin compounds have not been identified in infected barley plants.

STSs and pyrone synthase contain amino acid substitutions in the predicted active site environment that distinguish them from naringenin CHSs (Fig. 3). Interestingly, the Gln-His substitution characteristic of the grape and peanut STSs was also found in

sorghum CHS8. On the other hand, the Thr-Met substitution is unique to CHS8 among the proteins examined in this study. STSs have not been identified in any monocot species, but sugarcane has been reported to synthesize the stilbene phytoalexin piceatannol [6]. It is impossible to conclude what enzymatic activities have evolved in sorghum CHS8 without detailed biochemical characterization. However, because of the differential expression of CHS8 in pathogen-induced sorghum seedlings, we reason that it is involved in the biosynthesis of defense-related metabolites, such as the 3-deoxyanthocyanidins. The PI1 cDNA library, in which significant counts of CHS8 ESTs were observed, was prepared from 2-week-old sorghum seedlings 48 h after inoculation with *C. graminicola* (Table I). We have previously demonstrated the accumulation of specific 3-deoxyanthocyanidin phytoalexins, such as luteolinidin and apigeninidin, in sorghum tissues of different stages inoculated with the same fungus [28, 39]. Importantly, the sorghum phytoalexins represent a significant component for the expression of resistance to *C. graminicola* [28, 39]. Despite their structural similarities to the anthocyanidins, synthesis of 3-deoxyanthocyanidins in sorghum occurs through some distinct but yet unidentified pathways [19]. The correlation between CHS8 expression, as estimated by EST abundance, and the presence of 3-deoxyanthocyanidins suggests that CHS8 is involved in their biosynthesis. We speculate that CHS8 encodes a CHS-related enzyme with a different substrate requirement that catalyzes the formation of a "chalcone" precursor unique to 3-deoxyanthocyanidin biosynthesis. Thus, CHS8 may represent the key to the dissection of this novel flavonoid pathway. Finally, it is worth emphasizing that the biological relevance of *in silico* expression data is far from being conclusive and needs to be confirmed by further experimental work. Nevertheless, this

type of analysis represents a quick and efficient way of identifying potential genes of interest under specific biological conditions.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (grant no. MCB-96603439 to RLN). This is journal article number 16688 of the Purdue University Agricultural Experimental Station. The authors thank Dr. P. B. Goldsbrough for valuable advice during the investigation. We also thank the Pratt laboratory (CGGC), funded by NSF grant DBI-9872649, for providing the EST clones and data.

REFERENCES

1. **Akada S, Dube SK.** 1995. Organization of soybean chalcone synthase clusters and characterization of a new member of the family. *Plant Molecular Biology* **29:** 189-199.
2. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller M, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* 25:3389-3402.
3. **Arias JA, Dixon RA, Lamb CJ.** 1993. Dissection of the functional architecture of a plant defense gene promoter using a homologous in vitro transcription initiation system. *Plant Cell* **5:** 485-496
4. **Arumunganathan K, Earle ED**. 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9**: 208-219
5. **Audic S, Claverie J-M.** 1997. The significance of digital gene expression profiles. *Genome Research* **7:** 986-995

6. Brinker AM, Seigler DS. 1991. Isolation and identification of piceatannol as a phytoalexin from sugarcane. *Phytochemistry* **30:** 3229-3232

7. Chopra S, Brendel V, Zhang J, Axtell JD, Peterson T. 1999. Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. *Proceedings of the National Academy of Science, USA* **96:**15330-5

8. Christensen AB, Gregersen PL, Schroder J, Collinge DB. 1998. A chalcone synthase with an unusual substrate preference is expressed in barley leaves in response to UV light and pathogen attack. *Plant Molecular Biology* **37**: 849-857

9. Cui Y, Magill J, Frederiksen R, Magill C. 1996. Chalcone synthase and phenylalanine ammonia-lyase mRNA levels following exposure of sorghum seedlings to three fungal pathogens. *Physiological and Molecular Plant Pathology* **49**:187-199.

10. Dangl JL, Hahlbrock K, Schell J. 1989. Regulation and structure of chalcone synthase. In IK Vasil, J Schell, eds, *Cell Culture and Somatic Cell Genetics of Plants*, Vol 6. Academic Press, New York, pp 155-173.

11. Dixon RA, Paiva NL. 1995. Stress-induced phenylpropanoid metabolism. *Plant Cell* **7:**1085-1097.

12. Draye X, Lin Y-R, Qian X-Y, Bowers JE, Burow GB, Morrell PL, Peterson DG, Presting GG, Ren S-X, Wing RA, Paterson AH. 2001. Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the sorghum genome as a foundation. *Plant Physiology* **125**: 1325-1341.

13. Eckermann S, Schröder G, Schmidt J, Strack D, Edrada RA, Helariutta Y, Elomaa P, Kotilainen M, Kilpeläinen I, Proksch P, Teeri TH, Schröder J. 1998. New pathway to polyketides in plants. *Nature* **396:** 387-390.

14. Feinbaum RL, Ausubel FM. 1988. Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. *Molecular Cell Biology* **8**: 1985-1992 (1988).

15. Franken P, Niesbach-Klösgen U, Weydemann U, Maréchal-Drouard L, Saedler H, Wienand H. 1991. The duplicated chalcone synthase genes *C2* and *Whp* (white pollen) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by the anthocyanin intensifying gene *in*. *European Molecular Biology Organization* **10**:2605-2612.

16. Gaut BS, Le Thierry I' Enneguin M, Peek AS, Saukins MC. 2000. Maize as a model for the evolution of plant nuclear genomes. *Proceedings of the National Academy of Science, USA* **97:** 7008-7015.

17. Hartman U, Valentine WJ, Christie JM, Hays J, Jenkins GI, Wessjaar B. 1998. Identification of UV/blue light-responsive elements in the *Arabidopsis thaliana* chalcone synthase promoter using a homologous protoplast transient expression system. *Plant Molecular Biology* **36**: 741-754.

18. Herrmann A, Schulz W, Hahlbrock K. 1988. Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. *Molecular and General Genetics* **212:** 93-98.

19. Hipskind JD, Goldsbrough PB, Urmeev F, Nicholson RL (1996) Synthesis of 3-deoxyanthocyanidin phytoalexins in sorghum does not occur via the same pathway as 3-hydroxylated anthocyanidins and phlobaphenes. *Maydica* **41**: 155-166.

20. Holton TA, Cornish EC. 1995. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* **7**:1071-1083.

21. Hulbert SH, Richter TE, Axtell JD, Bennetzen JL. 1990. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proceedings of the National Academy of Science, USA* **87:** 4251-4255.

22. Jacobs M, Rubery PH. 1988. Naturally occurring auxin transport regulators. *Science* **241**: 239-253.

23. Junghans H, Dalkin K, Dixon RA. 1993. Stress responses in alfalfa (*Medicago sativa* L.). 15. Characterization and expression patterns of members of a subset of the chalcone synthase multigene family. *Plant Molecular Biology* **22**: 239-253.

24. Koes RE, Spelt CE, Mol JNM. 1989. The chalcone synthase multigene family of *Petunia hybrida* (V30): differential, light-regulated expression during flower development and UV induction. *Plant Molecular Biology* **12**: 213-225.

25. Kubasek WL, Ausubel FM, Shirley BW. 1998. A light-independent developmental mechanisms potentiates flavonid gene expression in *Arabidopsis* seedlings. *Plant Molecular Biology* **37**: 217-223.

26. Lanz T., Tropf S, Marner F-J, Schröder J, Schröder G. 1991. The role of cysteines in polyketide synthases. Site directed mutagenesis of resveratrol and chalcone synthases, two enzymes in different plant-specific pathways. *Journal of Biological Chemistry* 266: 9971-9976.

27. Lo SCC, Nicholson RL. 1998. Reduction of light-induced anthocyanin accumulation in inoculated sorghum mesocotyls. Implication for a compensatory role in the defense response. *Plant Physiology* **116:** 979-989.

28. Lo SC, de Verdier K, Nicholson RL. 1999. Accumulation of 3-deoxyanthocyanidin phytoalexins and resistance to *Colletotrichum sublineolum* in sorghum. *Physiological and Molecular Plant Pathology* **55:** 263-273.

29. Mekhedov S, de Ilárduya OM, Ohlrogge J. 2000. Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. *Plant Physiology* **122:** 389-401.

30. Mo Y, Nagel C, Taylor LP. 2001. Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proceedings of the National Academy of Science, USA* **89:** 7213-7217.

31. Nicholson RL, Hammerschmidt R. 1992. Phenolic compounds and their role in disease resistance. *Annual Review of Phytopathology* **30:** 369-389.

32. Ohlrogge J, Benning C . 2000. Unraveling plant metabolism by EST analysis. *Current Opinions in Plant Biology* **3**: 224-228.

33. Paterson AH, Lin Y-R, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S-C, Stansel JW, Irvine JE. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269:** 1714-1717.

34. Pereira MG, Lee M, Bramel-Cox P, Woodman W, Doebley J, Whitkus R. 1994. Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome* **37**: 236-243.

35. Ryder TB, Headrick SA, Bell JN, Liang X, Clouse SD, Lamb CJ. 1987. Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Molecular and General Genetics* **210:** 219-233.

36. Schmelzer E, Jahanen W, Hahlbrock K. 1988. In situ localization of light induced chalcone synthase mRNA, chalcone synthase, and flavonoid end products in epidermal cells of parsley leaves. *Proceedings of the National Academy of Sciences, USA.* **85**: 2989-2993.

37. Schroder J. 2000. The family of chalocone synthase-related proteins: functional diversity and evolution. In JT Romero, eds, Evolution of Metabolic Pathways. *Elsevier Science Ltd.*, pp 55-89.

38. Schulze-Lefert P, Becker-André, Schulz W, Hahlbrock K, Dangl JL. 1989. Functional architecture of the light-responsive chalcone synthase promoter from parsley. *Plant Cell* **1:** 707-714.

39. Snyder BA, Nicholson RL (1990) Synthesis of phytoalexins in sorghum as a site specific response to fungal ingress. *Science* **248:** 1637-1639.

40. Stafford HA. 1998. Teosinte to maize-some aspects of missing biochemical and physiological data concerning regulation of flavonoid pathways. *Phytochemistry* **49**: 285-293.

41. Suh D-Y, Fukuma K, Kagami J, Yamazaki Y, Shibuya M, Ebizuka Y, Sankawa U. 2000a. Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. *Journal of Biochemistry* **350:** 229-235.

42. Suh D-Y, Kagami J, Fukuma K, Sankawa U. 2000b. Evidence for catalytic cysteine-histidine dyad in chalcone synthase. *Biochemistry et biophysica acta. Reviews* **275:** 725-730.

43. van der Meer IM, Stuitje AR, Mol JNM. 1993. Regulation of general phenylpropanoid and flavonoid gene expression. CRC Press, Boca Ratan, pp 125-155.

44. Weiergang I, Hipskind JD, and Nicholson RL. 1996. Synthesis of 3-deoxyanthocyanidin phytoalexins in sorghum occurs independent of light. *Physiological and Molecular Plant Pathology* **49**: 377-388.

45. Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA. 1994. Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Research* **22**: 4922-4931.

46. Yamazaki Y, Suh D-Y, Sitthithaworn W, Ishiguro K, Kobayashi Y, Shibuya M, Ebizuka Y, Sankawa U. 2001. Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, Psilotum nudum. *Planta* **214**: 75-84.

47. Zhao Z-Y, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, Pierce D. 2000. *Agrobacterium*-mediated transformation in sorghum. *Plant Molecular Biol*ogy **44:** 789-798.

FIGURE LEGENDS

FIG. 1. Diagrammatic representation of sorghum CHS genes. The coding sequence (cds) of the maize C2 gene was aligned to the genomic sequences of CHS1-7 to predict the coding regions in the sorghum genes. The predicted coding regions are all interrupted by a single intron in the cysteine codon (TGT/C) in the same positions. The CHS probe used to isolate the genes from the BAC library aligned to the second exon. Putative TATA box and polyadenylation signals (AATAA) were identified as shown.

FIG. 2. cDNA and predicted amino acid sequences of sorghum CHS8. Nucleotide sequence was determined by assembling of highly conserved CHS ESTs in the PI1 and DG1 cDNA libraries. This sequence is complementary to the tentative consensus sequence TC20269 at SbGI. The TGC codon (dotted line), which is interrupted by introns in CHS1 to 7, is located in the same position in CHS8. The putative polyadenylation signal is underlined.

FIG. 3. Alignment of the active site regions in CHS and CHS-like proteins from different plant species. The active site nucleophile cys residue is located in a highly conserved environment (underlined) in these proteins. The active site environment is strictly conserved in sorghum CHS1 to 7, as in all the known naringenin CHS. Substitution events (bold) were identified in CHS8 and other CHS-like enzymes. Accession no. for protein sequences: sorghum CHS1 to 8, this study; maize C2, P24824; maize Whp, P24825; Petunia CHSA, P08894; Soybean CHS1, SYSYC3; Tomato CHS, S12224; Alfalfa CHS1, P30073; *Gerbera* CHS1, P48390; Barley CHS1, P26018; Pea CHS1, S33610; Barley HvCHS2, Q96562; *Gerbera* pyrone synthase (PYS), CAA86219; Peanut STS1, AAA96434; Grape STS1, AAL09046.

FIG. 4. Evolutionary relationships between CHS and CHS-like proteins from different plant species. The phylogenetic tree was constructed in Lasergene software, based on the multiple alignment by Jotun Hein Method with PAM250 as the residue weight table. Units on the scale underneath the tree represent 10X number of substitution events. Protein sequences examined were the same as those in Fig. 3.

Table I. Features of sorghum CHS genes isolated from the BAC library

Gene	BAC[a]	Identity (%)		Intron (bp)
		C2 cds[b]	C2 aa[c]	
CHS1	E16	87.0	90.3	205
CHS2	E16	88.7	89.8	165
CHS3	E16	89.4	90.0	263
CHS4	C19	88.0	90.0	647
CHS5	C19	89.8	89.3	1967
CHS6	G18	90.0	88.5	1570
CHS7	G18	89.8	87.3	150

[a]BAC clone from which the CHS gene was isolated.

[b]Coding sequence of the maize *C2* gene (X60205).

[c]Amino acid sequence of *C2* protein (CAA42764)

Table II. *S. bicolor* (BTx623) cDNA libraries represented in dbEST, GenBank. EST sequences were released by the University of Georgia CGGC.

Library	Description	Number of EST [a]
Dark-grown (DG1)	5-d-old seedlings grown in dark growth room	6865
Light-grown (LG1)	10 to 14-d-old seedlings grown in greenhouse	6748
Pathogen Induced (PI1)	14-d-old seedlings, 48 h after inoculation with *Colletotrichum graminicola*	6135
Water-stressed (WS1)	5-week-old plants 7-8 days after withholding water	6265
Immature panicles (IP1)	developing pre-anthesis panicles	5623
Ovary (OV1&2)	mixtures of ovaries of varying immature stages from 8-week-old plants	5882
Embryo (EM1)	embryos germinated for 24 h	5926

[a]Data as of October 24, 2001. Most EST clones were represented by 2 sequence files (5'- and 3'-ends) in dbEST and were only considered once.

Table III. CHS ESTs identified in the sorghum EST databases. GenBank accession no., EST ID name, library, CHS gene identity, and TC number for each sequence are shown.

Accession No.	EST ID[a]	Library	Gene[b]	TC No.[c]
BE359668	DG1_56_G02.b2_A002	DG1	CHS1	14644
BE359724	DG1_56_G02.g2_A002	DG1	CHS1	14644
BE362244	DG1_85_H08.b1_A002	DG1	CHS3	14645
BE362275	DG1_85_H08.g1_A002	DG1	CHS3	14645
BE362252	DG1_85_F07.b1_A002	DG1	CHS4	14643
BE362290	DG1_85_F07.g1_A002	DG1	CHS4	14643
BE357552	DG1_21_G10.b1_A002	DG1	CHS5	14645
BE357672	DG1_21_G10.g1_A002	DG1	CHS5	14645
BE357763	DG1_22_G09.b1_A002	DG1	CHS5	14645
BE357828	DG1_22_G09.g1_A002	DG1	CHS5	14645
BE358776	DG1_32_D08.b1_A002	DG1	CHS5	14645
BE358855	DG1_32_D08.g1_A002	DG1	CHS5	14645
BE361141	DG1_70_C01.b1_A002	DG1	CHS5	14645
BE361236	DG1_70_C01.g1_A002	DG1	CHS5	14645
BE356020	DG1_121_B02.b1_A002	DG1	CHS6	14645
BE356024	DG1_121_B02.g1_A002	DG1	CHS6	14642
BE355662	DG1_116_C08.b1_A002	DG1	CHS8[d]	20269
BE355743	DG1_116_C08.g1_A002	DG1	CHS8	20269
AW287293	LG1_269_B01.b1_A002	LG1	CHS2	14644
AW284288	LG1_269_B01.g1_A002	LG1	CHS2	14642
AW283840	LG1_260_D05.g1_A002	LG1	CHS3	14642
AW286037	LG1_260_E08.b1_A002	LG1	CHS3	14645

1	AW672357	LG1_359_D12.b1_A002	LG1	CHS3	14645
2	AW287061	LG1_265_A05.b2_A002	LG1	CHS4	14643
3	AW283518	LG1_265_A05.g1_A002	LG1	CHS4	14643
4	AW287080	LG1_265_C05.b2_A002	LG1	CHS4	14643
5	AW283504	LG1_265_C05.g1_A002	LG1	CHS4	14643
6	AW671975	LG1_353_E05.b1_A002	LG1	CHS4	14643
7	AW565912	LG1_353_E05.g1_A002	LG1	CHS4	14643
8	AW565339	LG1_342_D02.g1_A002	LG1	CHS5	14645
9	AW564589	LG1_296_G07.b1_A002	LG1	CHS6	14645
10	AW672334	LG1_359_A01.b1_A002	LG1	CHS6	14645
11	AW671616	LG1_348_A08.b1_A002	LG1	U.D. [e]	14641
12	AW672448	LG1_359_D12.g1_A002	LG1	U.D.	14642
13	BE596741	PI1_59_A12.g1_A002	PI1	CHS1	14644
14	BE596891	PI1_59_H01.b1_A002	PI1	CHS1	14644
15	BE594398	PI1_33_A03.b1_A002	PI1	CHS5	14645
16	BE366441	PI1_33_A03.g1_A002	PI1	CHS5	14645
17	BE363966	PI1_10_H06.g1_A002	PI1	CHS8	20269
18	BE364784	PI1_16_D11.b1_A002	PI1	CHS8	20269
19	BE367323	PI1_44_A08.b1_A002	PI1	CHS8	20269
20	BE598061	PI1_65_C02.b1_A002	PI1	CHS8	20269
21	BE597782	PI1_65_C02.g1_A002	PI1	CHS8	20269
22	BE598245	PI1_67_F08.b1_A002	PI1	CHS8	20269
23	BE597962	PI1_67_F08.g1_A002	PI1	CHS8	20269
24	BE597503	PI1_70_D03.b1_A002	PI1	CHS8	20269
25	BE597283	PI1_70_D03.g1_A002	PI1	CHS8	20269
26	BE600157	PI1_80_A04.b1_A002	PI1	CHS8	20269
27	BE600826	PI1_90_F05.b1_A002	PI1	CHS8	20269
28	BE601240	PI1_90_F05.g1_A002	PI1	CHS8	20269

1	BE595826	PI1_54_F02.b1_A002	PI1	U.D.	14643
2	AW678390	WS1_15_D04.g1_A002	WS1	CHS6	14645
3	BI075945	IP1_23_C09_b1_A002	IP1	CHS5	N/A
4	BI075646	IP1_23_C09_g1_A002	IP1	CHS5	N/A
5	BF176830[f]	EM1_3_B02.b1_A002	EM1	CHS4	14643
6	BF177058[f]	EM1_3_B02.g1_A002	EM1	CHS7	14641

7

8 [a]"b1/b2" and "g1/g2" are extensions of EST ID names, representing the 5' and 3' sequences of cDNA
9 clones, respectively. [b]CHS gene identity was determined by direct sequence comparison to *CHS1-7*. [c]TC
10 No., tentative consensus sequence number in SbGI. Most CGGC EST sequences have been assembled by
11 TGIR into TC sequences which are presented in SbGI (www.tigr.org/tdb/sbgi). [d]*CHS8* was identified as a
12 new CHS gene by assembling of EST sequences with less than 90% identity to *CHS1-7*. [e]U.D.,
13 undetermined. [f]BF176830 and BF177058 are 5'- and 3'-end sequences from the same EST clone
14 (EM1_3_B02). It is unclear why they aligned to different CHS genes. This EST clone may represent
15 another new CHS gene.

16
17

Table IV. EST abundance of CHS genes in the *S. bicolor* cDNA libraries.

	CHS EST Counts[a]						
Gene	DG1	LG1	PI1	WS1	IP1	OV1&2	EM[b]
All CHS	11	9	12	1	1	0	1
CHS1	0	1	2	0	0	0	0
CHS2	1	0	0	0	0	0	0
CHS3	3	1	0	0	0	0	0
CHS4	3	1	0	0	0	0	(1)
CHS5	1	4	1	0	1	0	0
CHS6	1	1	0	1	0	0	0
CHS7	0	0	0	0	0	0	(1)
CHS8	0	1	8[c]	0	0	0	0
U.D.[d]	2	0	1	0	0	0	0
EF1A[e]	12	16	12	11	32	31	71

[a]EST clones with 5'- and 3'- end sequences in dbEST were only counted once. [b]The 5'- and 3'-end sequences of the EST clone from this library aligned to different CHS genes (Table III). [c]For a confidence level of 95%, the observed count was significantly different from counts for *CHS8* in other libraries (Audic and Claverie, 1997). [d]U.D., undetermined. [e]EF1A, translation elongation factor 1A (TC34091, SbGI), highly induced after seed germination, was used as a positive reference for *in silico* expression analysis.

Figure 1

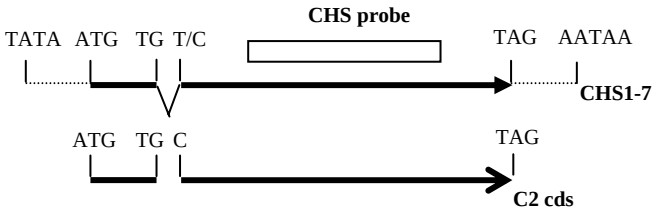


Figure 2

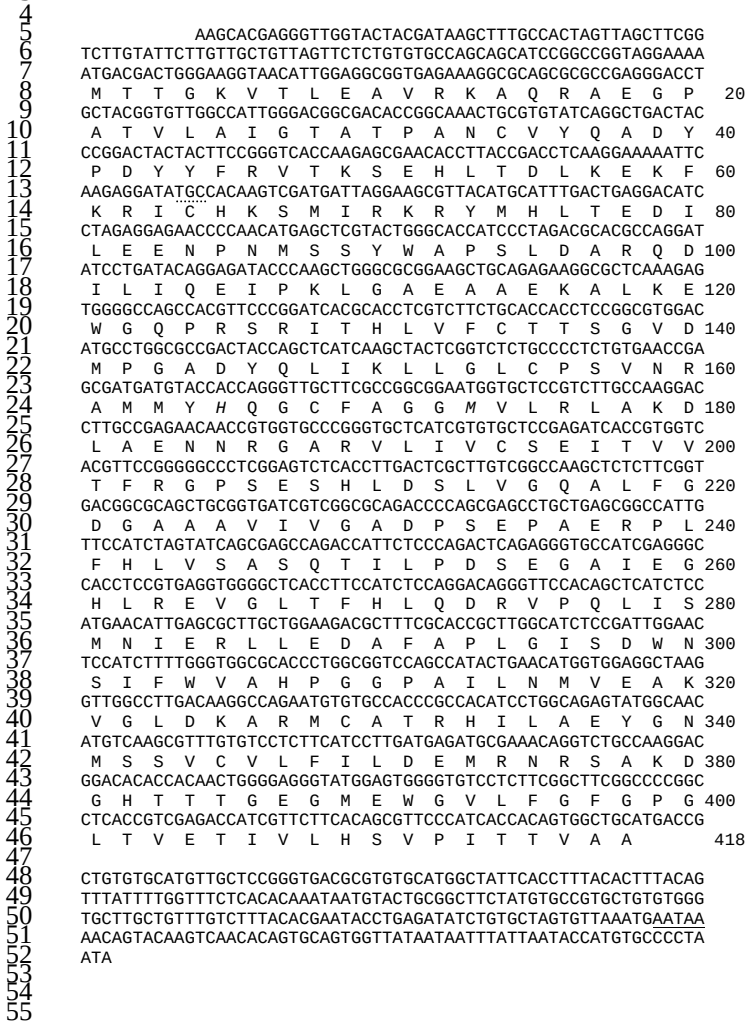

AAGCACGAGGGTTGGTACTACGATAAGCTTTGCCACTAGTTAGCTTCGG
TCTTGTATTCTTGTTGCTGTTAGTTCTCTGTGTGCCAGCAGCATCCGGCCGGTAGGAAAA
ATGACGACTGGGAAGGTAACATTGGAGGCGGTGAGAAAGGCGCAGCGCGCCGAGGGACCT
M T T G K V T L E A V R K A Q R A E G P 20
GCTACGGTGTTGGCCATTGGGACGGCGACACCGGCAAACTGCGTGTATCAGGCTGACTAC
A T V L A I G T A T P A N C V Y Q A D Y 40
CCGGACTACTACTTCCGGGTCACCAAGAGCGAACACCTTACCGACCTCAAGGAAAAATTC
P D Y Y F R V T K S E H L T D L K E K F 60
AAGAGGATATGCCACAAGTCGATGATTAGGAAGCGTTACATGCATTTGACTGAGGACATC
K R I C H K S M I R K R Y M H L T E D I 80
CTAGAGGAGAACCCCAACATGAGCTCGTACTGGGCACCATCCCTAGACGCACGCCAGGAT
L E E N P N M S S Y W A P S L D A R Q D 100
ATCCTGATACAGGAGATACCCAAGCTGGGCGCGGAAGCTGCAGAGAAGGCGCTCAAAGAG
I L I Q E I P K L G A E A A E K A L K E 120
TGGGGCCAGCCACGTTCCCGGATCACGCACCTCGTCTTCTGCACCACCTCCGGCGTGGAC
W G Q P R S R I T H L V F C T T S G V D 140
ATGCCTGGCGCCGACTACCAGCTCATCAAGCTACTCGGTCTCTGCCCCTCTGTGAACCGA
M P G A D Y Q L I K L L G L C P S V N R 160
GCGATGATGTACCACCAGGGTTGCTTCGCCGGCGGAATGGTGCTCCGTCTTGCCAAGGAC
A M M Y H Q G C F A G G M V L R L A K D 180
CTTGCCGAGAACAACCGTGGTGCCCGGGTGCTCATCGTGTGCTCCGAGATCACCGTGGTC
L A E N N R G A R V L I V C S E I T V V 200
ACGTTCCGGGGGCCCTCGGAGTCTCACCTTGACTCGCTTGTCGGCCAAGCTCTCTTCGGT
T F R G P S E S H L D S L V G Q A L F G 220
GACGGCGCAGCTGCGGTGATCGTCGGCGCAGACCCCAGCGAGCCTGCTGAGCGGCCATTG
D G A A A V I V G A D P S E P A E R P L 240
TTCCATCTAGTATCAGCGAGCCAGACCATTCTCCCAGACTCAGAGGGTGCCATCGAGGGC
F H L V S A S Q T I L P D S E G A I E G 260
CACCTCCGTGAGGTGGGGCTCACCTTCCATCTCCAGGACAGGGTTCCACAGCTCATCTCC
H L R E V G L T F H L Q D R V P Q L I S 280
ATGAACATTGAGCGCTTGCTGGAAGACGCTTTCGCACCGCTTGGCATCTCCGATTGGAAC
M N I E R L L E D A F A P L G I S D W N 300
TCCATCTTTTGGGTGGCGCACCCTGGCGGTCCAGCCATACTGAACATGGTGGAGGCTAAG
S I F W V A H P G G P A I L N M V E A K 320
GTTGGCCTTGACAAGGCCAGAATGTGTGCCACCCGCCACATCCTGGCAGAGTATGGCAAC
V G L D K A R M C A T R H I L A E Y G N 340
ATGTCAAGCGTTTGTGTCCTCTTCATCCTTGATGAGATGCGAAACAGGTCTGCCAAGGAC
M S S V C V L F I L D E M R N R S A K D 380
GGACACACCACAACTGGGGAGGGTATGGAGTGGGGTGTCCTCTTCGGCTTCGGCCCCGGC
G H T T T G E G M E W G V L F G F G P G 400
CTCACCGTCGAGACCATCGTTCTTCACAGCGTTCCCATCACCACAGTGGCTGCATGACCG
L T V E T I V L H S V P I T T V A A 418
CTGTGTGCATGTTGCTCCGGGTGACGCGTGTGCATGGCTATTCACCTTTACACTTTACAG
TTTATTTTGGTTTCTCACACAAATAATGTACTGCGGCTTCTATGTGCCGTGCTGTGTGGG
TGCTTGCTGTTTGTCTTTACACGAATACCTGAGATATCTGTGCTAGTGTTAAATGAATAA
AACAGTACAAGTCAACACAGTGCAGTGGTTATAATAATTTATTAATACCATGTGCCCCTA
ATA

Figure 3

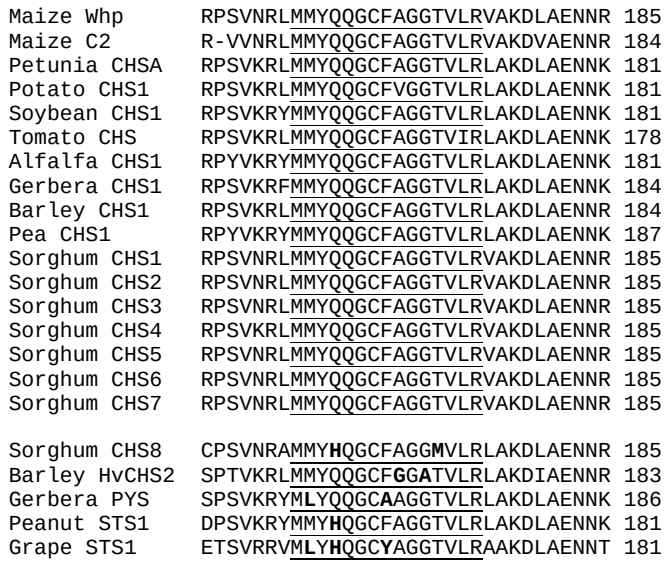
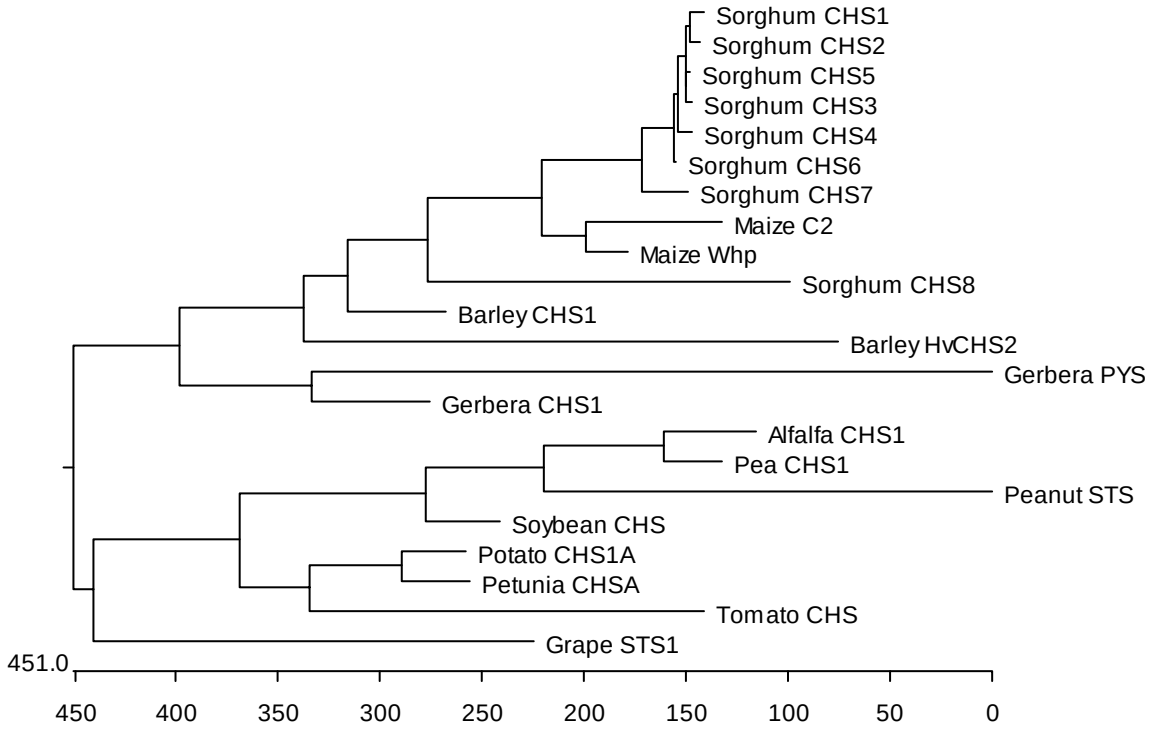

Maize Whp RPSVNRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Maize C2 R-VVNRLMMYQQGCFAGGTVLRVAKDVAENNR 184
Petunia CHSA RPSVKRLMMYQQGCFAGGTVLRLAKDLAENNK 181
Potato CHS1 RPSVKRLMMYQQGCFVGGTVLRLAKDLAENNK 181
Soybean CHS1 RPSVKRYMMYQQGCFAGGTVLRLAKDLAENNK 181
Tomato CHS RPSVKRLMMYQQGCFAGGTVIRLAKDLAENNK 178
Alfalfa CHS1 RPYVKRYMMYQQGCFAGGTVLRLAKDLAENNK 181
Gerbera CHS1 RPSVKRFMMYQQGCFAGGTVLRLAKDLAENNK 184
Barley CHS1 RPSVKRLMMYQQGCFAGGTVLRLAKDLAENNR 184
Pea CHS1 RPYVKRYMMYQQGCFAGGTVLRLAKDLAENNK 187
Sorghum CHS1 RPSVNRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Sorghum CHS2 RPSVNRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Sorghum CHS3 RPSVNRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Sorghum CHS4 RPSVKRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Sorghum CHS5 RPSVNRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Sorghum CHS6 RPSVNRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Sorghum CHS7 RPSVNRLMMYQQGCFAGGTVLRVAKDLAENNR 185
Sorghum CHS8 CPSVNRAMMYHQGCFAGGMVLRLAKDLAENNR 185
Barley HvCHS2 SPTVKRLMMYQQGCFGGATVLRLAKDIAENNR 183
Gerbera PYS SPSVKRYMLYQQGCAAGGTVLRLAKDLAENNK 186
Peanut STS1 DPSVKRYMMYHQGCFAGGTVLRLAKDLAENNK 181
Grape STS1 ETSVRRVMLYHQGCYAGGTVLRAAKDLAENNT 181

Figure 4


Sorghum CHS1
Sorghum CHS2
Sorghum CHS5
Sorghum CHS3
Sorghum CHS4
Sorghum CHS6
Sorghum CHS7
Maize C2
Maize Whp
Sorghum CHS8
Barley CHS1
Barley HvCHS2
Gerbera PYS
Gerbera CHS1
Alfalfa CHS1
Pea CHS1
Peanut STS
Soybean CHS
Potato CHS1A
Petunia CHSA
Tomato CHS
Grape STS1
451.0
450
400
350
300
250
200
150
100
50
0