

ORIGINAL RESEARCH

OPEN ACCESS
Full open access to this and
thousands of other papers at
<http://www.la-press.com>.

Identification of Drought-Responsive Universal Stress Proteins in Viridiplantae

Raphael D. Isokpehi, Shaneka S. Simmons, Hari H.P. Cohly, Stephen I.N. Ekunwe, Gregorio B. Begonia and Wellington K. Ayensu

Department of Biology, Jackson State University, Jackson MS 39217, USA.
Corresponding author email: raphael.isokpehi@jsums.edu

Abstract: Genes encoding proteins that contain the universal stress protein (USP) domain are known to provide bacteria, archaea, fungi, protozoa, and plants with the ability to respond to a plethora of environmental stresses. Specifically in plants, drought tolerance is a desirable phenotype. However, limited focused and organized functional genomic datasets exist on drought-responsive plant USP genes to facilitate their characterization. The overall objective of the investigation was to identify diverse plant universal stress proteins and Expressed Sequence Tags (ESTs) responsive to water-deficit stress. We hypothesize that cross-database mining of functional annotations in protein and gene transcript bioinformatics resources would help identify candidate drought-responsive universal stress proteins and transcripts from multiple plant species. Our bioinformatics approach retrieved, mined and integrated comprehensive functional annotation data on 511 protein and 1561 ESTs sequences from 161 viridiplantae taxa. A total of 32 drought-responsive ESTs from 7 plant genera *Glycine*, *Hordeum*, *Manihot*, *Medicago*, *Oryza*, *Pinus* and *Triticum* were identified. Two *Arabidopsis* USP genes At3g62550 and At3g53990 that encode ATP-binding motif were up-regulated in a drought microarray dataset. Further, a dataset of 80 simple sequence repeats (SSRs) linked to 20 singletons and 47 transcript assemblies was constructed. Integrating the datasets on SSRs and drought-responsive ESTs identified three drought-responsive ESTs from bread wheat (BE604157), soybean (BM887317) and maritime pine (BX682209). The SSR sequence types were CAG, ATA and AT respectively. The datasets from cross-database mining provide organized resources for the characterization of USP genes as useful targets for engineering plant varieties tolerant to unfavorable environmental conditions.

Keywords: drought, expressed sequence tags, microsatellite, plants, Pfam, salinity, simple sequence repeats, Uniprot, universal stress protein domain, viridiplantae

Bioinformatics and Biology Insights 2011:5 41–58

doi: [10.4137/BBI.S6061](https://doi.org/10.4137/BBI.S6061)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

Environmental stresses can negatively impact agricultural crop yield and quality.^{1,2} As an adaptive strategy, plant genomes encode genes that produce proteins that function in stress response and tolerance.^{3–5} Despite substantial research on response to abiotic and biotic stresses by plants, there are still knowledge gaps regarding the molecular mechanisms that regulate the diverse functions of environmental stress-associated plant genes and proteins.³ The increasing availability of genomic sequences of members of the viridiplantae (green algae and land plants) in combination with high-throughput bioinformatics tools and databases^{4,5} provide new opportunities for examining understudied gene families that could be central to stress response in plants.

Genes encoding proteins that contain the conserved 140–160 residues Universal Stress Protein (USP) domain (Pfam Accession: PF00582) are known to provide bacteria, archaea, fungi, protozoa, and plants with the ability to respond to a plethora of environmental stresses.^{6–9} Nutrient starvation, drought, high salinity, extreme temperatures and exposure to toxic chemicals are examples of conditions that induce expression of genes with the USP domain. Proteins containing domain PF00582 are often collectively referred to as universal stress proteins. In *Escherichia coli*, the USPs have been grouped into four classes according to their structural analysis and amino acid sequence—Class I: UspA, UspC, UspD; Class II: UspF and UspG; and Class III and Class IV: Two Usp domains of UspE.¹⁰ The UspA domain of MJ0577 (also called 1MJH) from *Methanocaldococcus jannaschii* crystallizes with a bound ATP while the UspA domain of *Haemophilus influenzae* lacks both ATP-binding activity and ATP-binding residues.^{11,12} Structural alignment has shown that the second and third conserved glycines in the polypeptide of the ATP-binding loop G-2X-G-9X-G-(S/T) in 1MJH are replaced by bulky amino acids glutamine and methionine in UspA.¹¹ The suggested ancestral function of the universal stress protein domain was nucleotide binding and signal transduction.¹⁶ Despite the knowledge of bacterial USP proteins, the functional diversity of the USPs in other organisms, including various plant species needs to be better defined.^{17,18}

Kerk et al¹³ examined the sequence and structure of 44 *Arabidopsis thaliana* proteins containing similarity

to the USP domain of bacteria and concluded that all *Arabidopsis* USPA domain-containing sequences have evolved from a 1MJH-like ancestor. Since the publication,¹³ there has been additional but limited studies aimed at understanding the function of universal stress proteins of *A. thaliana*.^{20–24} For example, AT5G54430 (AtPHOS32) and AT4G27320 (AtPHOS34) were shown to be phosphorylated in response to microbial elicitation of *Arabidopsis* cells.^{21,23} In addition, AtPHOS32 was proved to be a new substrate of the stress-regulated mitogen-activated protein kinases (MAPKs), AtMPK3 (AT3G45640) and AtMPK6 (AT2G43790). However, the precise functions of these two *Arabidopsis* USP as well as other members of the gene family are not yet established. In rice, another model plant species, OsUSP1, which is mediated by the gaseous plant hormone ethylene has been identified to potentially function in adaptation of deepwater rice plant to hypoxia.¹⁴ Additional plant USP genes have been characterized including legumes *Astragalus sinicus*¹⁵ and *Vicia faba*^{16,17} as well as in *Gossypium arboreum* (cotton).¹⁸ Recently, the USP genes of barley were identified, localized and their expression in anatomic and selected stress condition determined.¹⁹

Water-limiting condition (drought) is one of the key abiotic stresses that can adversely affect the growth, development and yield of crop and tree plants.²⁰ Drought induces biochemical and physiological responses in plants²¹ including reduced photosynthetic carbon and energy metabolism²² leading to oxidative stress. High salinity is also accompanied by drought.²⁰ Furthermore, wood production from forest trees can be hampered by drought.^{32,33} The ability to respond and tolerate drought stress is a desirable phenotype especially in plants that have to survive in environments with insufficient water. The molecular and cellular mechanisms for response and tolerance have been investigated using a range of powerful high-throughput genomic and proteomic techniques to dissect gene networks response to drought.²² Examples of drought-responsive USP genes have been reported in cotton¹⁸ and cowpea.²³ The identification of drought responsive USP genes from multiple plants species will present an array of research tools for genetic manipulation of plants for drought tolerance. Therefore, we sought to develop a bioinformatics screening strategy to identify drought-responsive USP



genes and transcripts from comprehensive protein and gene transcript databases.

There continues to be an increase in number and diversity of bioinformatics resources storing functional annotation of protein-coding sequences including those containing the USP domain.²⁴ The Pfam database of protein families represented by alignments and Hidden Markov Models contains at least 550 protein sequences from the viridiplantae (green algae and land plants) annotated to contain at least one USP domain.²⁵ These sequences have identifiers of the Universal Protein Resource (UniProt), which is the most comprehensive catalog for protein sequence and functional annotation data.²⁶ The UniProt entries have value-added cross-references to external databases that provide diverse annotation including structural, gene expression, literature and sequence diversity. In addition, there are specialized plant databases not yet linked to UniProt. For example, the Phytozome resource page (http://www.phytozome.net/Phytozome_resources.php) provides links to resources for general plant genomics; gene expression; gene indices and Expressed Sequence Tags (ESTs); *Arabidopsis*; grass and cereals; legumes; forest trees; other plant species and plant pathogen genomics. The overall objective of the investigation was to identify diverse plant universal stress proteins and Expressed Sequence Tags (ESTs) responsive to water-deficit stress. We hypothesize that cross-database mining of functional annotations in protein and gene transcript bioinformatics resources would help identify candidate drought-responsive universal stress proteins and transcripts from multiple plant species.

Among the ESTs and cDNA resources listed in Phytozome, we observed that the TIGR Plant Transcript Assemblies database (Plantta)²⁷ had a wide collection of 254 plant species (as of July 2007). The ESTs and full-length cDNA are being used for discovery of genes in plant species as well as evidence of gene expression in conditions as well as anatomic parts. The identification of ESTs encoding universal stress proteins could facilitate further studies on selection of markers for comparative mapping, plant breeding and forward genetics.^{28,29}

The Plantta resource contains simple sequence repeats (SSR) or microsatellite annotation for some transcripts. Microsatellites are 1–6 bp tandemly repeated

DNA sequences that occupy a significant fraction of the nuclear genome of all eukaryotes.³⁰ Microsatellites in protein-coding genes can inactivate or activate genes or truncate protein.³¹ In plants, microsatellites derived from EST sequences (EST-SSRs) have been proposed to be better candidates for gene tagging and are preferred over genomic-SSR markers for plant improvement programs owing to their higher inter-specific transferability rate.³² Thus, we investigated the presence of SSR on transcript assemblies and singleton sequences in Plantta. Furthermore, since our primary interest was on drought-responsive genes, we sought to identify USP-annotated Plantta ESTs that contain text relevant to drought in their dbEST³³ entries. The keyword search provided an indication of the experimental condition for generating the cDNA libraries. Finally, we determined the overlap of EST dataset containing SSR entries with the EST dataset annotated with drought or water stress.

The bioinformatics strategy described can be adapted for analyzing a set of viridiplantae protein sequences defined by a Pfam protein domain. Furthermore, plant transcripts from other abiotic and biotic stress conditions can be mined and analyzed. In summary, we identified diverse plant universal stress proteins and transcripts responsive to drought including those that contain microsatellite markers that may regulate their function.

Methods

Construction of dataset of viridiplantae universal stress proteins

Viridiplantae proteins annotated in Pfam database²⁵ with Pfam domain PF00582 were downloaded and computationally processed with a suite of UNIX and PERL scripts to retrieve their respective UniProt Identifiers. Subsequently, for non-obsolete or deleted UniProt entries, the protein domain architecture, organism source of sequence, protein sequence length and protein molecular weight were extracted from XML-formatted UniProt entries (UniProt release 2010_10—Oct 5, 2010). These selected annotations are typically available for UniProt entries. Overview of the USP dataset construction is illustrated in Figure 1. Analysis of the protein domain architecture annotation provided a prediction of the number of USP domains as well as additional types of protein domain(s) present.

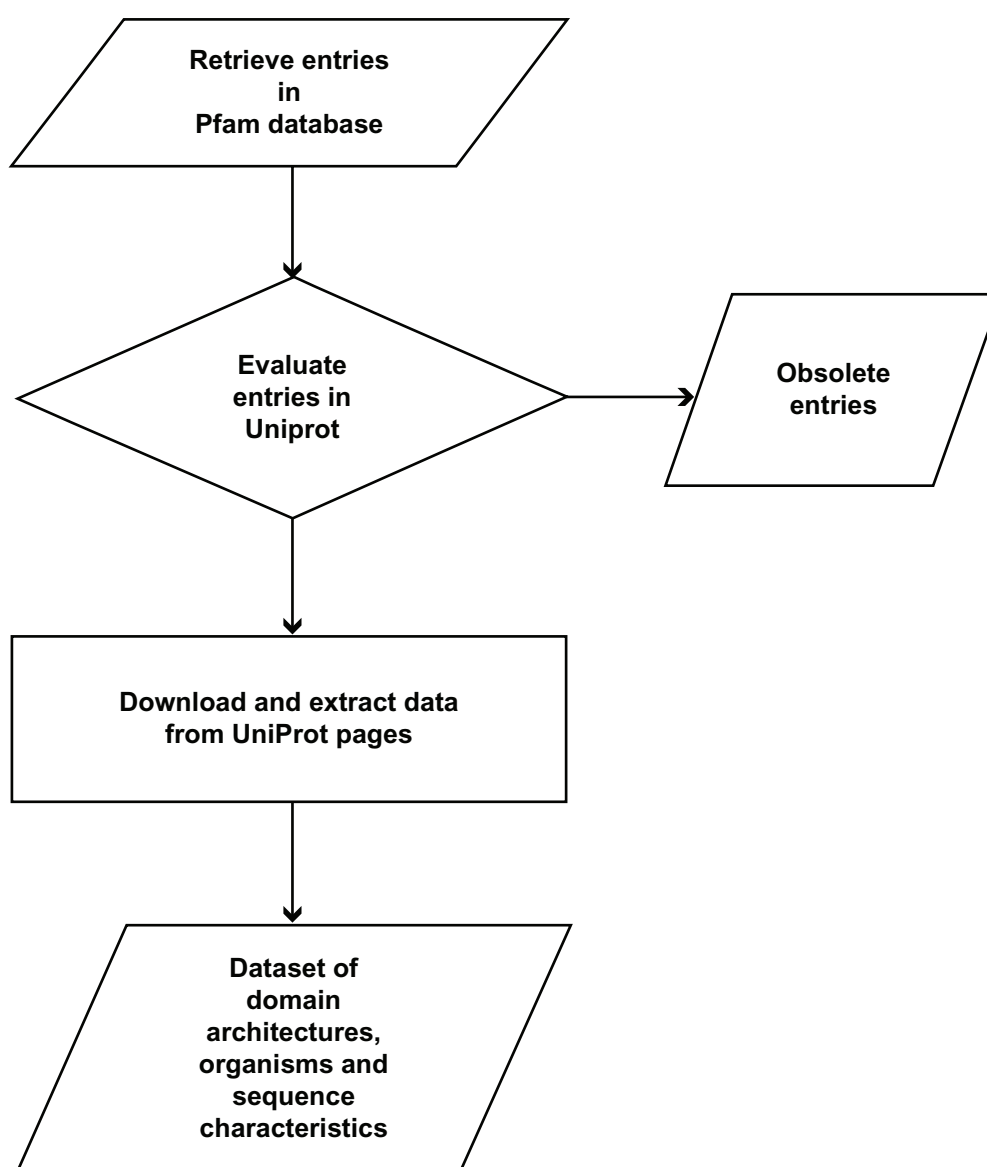


Figure 1. Flowchart for constructing dataset of viridiplantae universal stress proteins.

Orthologous viridiplantae drought-responsive genes encoding universal stress proteins

A UniProt entry for a protein sequence contains value-added cross-references to other databases (<http://www.uniprot.org/docs/dbxref>). The cross-referenced databases for each viridiplantae USP entry was computationally extracted from the XML formatted files. A non-redundant list of the databases was assembled and used to construct a presence-absence matrix consisting of rows of UniProt protein identifiers and columns of selected databases. A zero (0) was used to encode absence of cross-referencing to a database and

one (1) for presence of cross-reference to a database. This matrix was then searched for USP entries with cross-reference to the Gene Expression Atlas (a subset of ArrayExpress)⁴⁵ and Ortholog Matrix Project (OMA) Browser.³⁴ The matrix was visualized using a Linux version of matrix2png.³⁵ The Gene Expression Atlas (GXA) stores microarray and other gene expression data and was selected because it had annotation for “Experimental Factors”, which included a subsection on “Environmental Stresses” such as drought. Furthermore, the OMA Browser allows for exploration of orthologous relations between protein sequences for 1000 species (Release of May 2010).



A combination of the data from GXA and OMA allowed us to identify orthologous plant proteins in which a member has been demonstrated to be responsive to drought. Additional homologous sequences for the identified drought up-regulated USPs were retrieved from PLAZA—a resource for plant comparative genomics³⁶ and their multiple sequence alignment generated using ClustalW2 at <http://www.ebi.ac.uk/clustalw/>.

Viridiplantae universal stress protein transcripts derived from drought conditions

The TIGR Plant Transcript Assemblies (Plantta; <http://plantta.jcvi.org/>)²⁷ consists of a collection of transcripts (assembled ESTs and singletons) for at least 215 plant species. The content of webpage for each USP transcripts in the Plantta resource was also parsed to identify those with microsatellite (SSR) annotation. We sought to identify universal stress protein ESTs from cDNA library source derived from drought stress. The first step involved retrieving from Plantta, transcripts annotated with the text “universal stress protein”. In the second step, all the ESTs identifiers in dbEST³³ associated with the Plantta transcripts were retrieved and the entries in GenBank downloaded and searched for text “drought”. Another search strategy, the dbEST entries were searched for text “water” and then the retrieved subset searched with text “stress”. The assumption was that the presence of “drought” or combination of “water” and “stress” was indicative of a cDNA library derived from drought stress conditions. This mining of text in the dbEST entries was done to help identify universal stress protein ESTs as research tools for understanding stress response in a large number of plant species of agricultural, economic, ecological or industrial importance but without complete genome sequences.

Results

Construction of dataset of viridiplantae universal stress proteins

A total of 511 viridiplantae proteins annotated with universal stress protein domain (PF00582) from 43 unique taxa (NCBI Taxonomy IDs) were downloaded from UniProt on October 24, 2010 (Table 1). The protein count per taxa ranged from 1 to 88. The protein

counts for Liliopsida (monocotyledons), dicotyledons, and other viridiplantae including green algae were 235, 203 and 73 respectively. Furthermore, land plants with at least 50 USP records in UniProt from the Pfam dataset were *Oryza sativa subsp. japonica*, *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa subsp. indica* and *Zea mays*. The green algae genera represented in the dataset were *Chlamydomonas*, *Ostreococcus* and *Micromonas*. The sequence length ranged from 29 (A7Y7Q4) to 1223 (A8HRL3) with 251 unique lengths observed (Supplementary File 1 and Fig. 2). Finally, 39 sequences were annotated as fragments.

A total of 17 Pfam protein domains arranged in 17 architectures were associated with the dataset (Table 2 and Fig. 3). Ten of the 17 protein domains occurred only in one protein, most of which are uncharacterized as with sequences from *Oryza sativa subsp. indica*, *Oryza sativa subsp. japonica*, *Vitis vinifera* and *Zea mays*. Two sequences in this subset had names that indicated possible function: flagellar associated protein from *Chlamydomonas reinhardtii* and Anti-bacterial protein from *Solanum tuberosum* (potato). As expected the universal stress protein family (PF00582) domain was present in all the proteins analyzed. The protein kinase domain (PF00069), U-box domain (PF04564) and protein tyrosine kinase (PF07714) were found in at least 20 proteins (Table 2). The combination of domains for the USP and the transmembrane sodium/hydrogen exchanger family (PF00999) was observed in 5 proteins: B9S492 (*Ricinus communis*), A5BEW1 (*Vitis vinifera*), B9I6U4 (*Populus trichocarpa*), B9INS2 (*Populus trichocarpa*) and A9T441 (*Physcomitrella patens*). A total of 387 protein sequences had only the USP domain. In a subset of 12 sequences having tandem USP domains, 9 sequences were from green algae (Table 3).

Orthologous viridiplantae drought-responsive genes encoding universal stress proteins

The UniProtKB database cross-references for each viridiplantae USP entry stored in the XML format were extracted to determine the availability of each database annotation across the dataset of entries. Table 4 shows databases that were used to annotate at least 100 USPs. The complete list of 45

**Table 1.** Dataset of viridiplantae universal stress proteins entries in UniProt.

Scientific name	Common name	NCBI Taxonomy ID	Number of UniProt entries	USP domain only
<i>Oryza sativa</i> subsp. <i>japonica</i>	Rice	39947	88	60
<i>Arabidopsis thaliana</i>	Mouse-ear cress	3702	78	53
<i>Populus trichocarpa</i>	Western balsam poplar	3694	59	45
<i>Oryza sativa</i> subsp. <i>indica</i>	Rice	39946	52	32
<i>Zea mays</i>	Maize	4577	52	45
<i>Ricinus communis</i>	Castor bean	3988	43	32
<i>Picea sitchensis</i>	Sitka spruce	3332	21	21
<i>Physcomitrella patens</i>	Moss	3218	18	15
<i>Vitis vinifera</i>	Grape	29760	18	11
<i>Micromonas pusilla</i> CCMP1545		564608	10	10
<i>Medicago truncatula</i>	Barrel medic	3880	8	8
<i>Micromonas</i> sp. RCC299		296587	8	8
<i>Brassica campestris</i>	Field mustard	3711	7	7
<i>Chlamydomonas reinhardtii</i>		3055	6	4
<i>Ostreococcus lucimarinus</i> (strain CCE9901)		436017	5	5
<i>Ostreococcus tauri</i>		70448	4	4
<i>Vicia faba</i>	Broad bean	3906	4	4
<i>Brachypodium distachyon</i>	Purple false brome	15368	3	1
<i>Gossypium arboreum</i>	Tree cotton	29729	2	2
<i>Oryza sativa</i>	Rice	4530	2	0
<i>Arachis hypogaea</i>	Peanut	3818	1	1
<i>Astragalus sinicus</i>	Chinese milk vetch	47065	1	1
<i>Brachypodium sylvaticum</i>	False brome	29664	1	0
<i>Brassica oleracea</i> var. <i>alboglabra</i>	Chinese kale	3714	1	1
<i>Capsicum chinense</i>	Scotch bonnet	80379	1	1
<i>Cicer arietinum</i>	Chickpea	3827	1	1
<i>Gossypium barbadense</i>	Sea-island cotton	3634	1	1
<i>Hordeum bulbosum</i>	Bulbous barley	4516	1	1
<i>Hordeum vulgare</i>	Barley	4513	1	1
<i>Hordeum vulgare</i> var. <i>distichum</i>	Two-rowed barley	112509	1	1
<i>Marchantia polymorpha</i>	Liverwort	3197	1	0
<i>Mirabilis jalapa</i>	Garden four-o'clock	3538	1	1
<i>Pisum sativum</i>	Garden pea	3888	1	1
<i>Populus trichocarpa</i> × <i>Populus deltoids</i>	Black cottonwood × Eastern cottonwood	3695	1	1
<i>Potamogeton distinctus</i>	Roundleaf pondweed	62344	1	1
<i>Prunus dulcis</i>	Almond	3755	1	1
<i>Solanum lycopersicum</i>	Tomato	4081	1	1
<i>Solanum tuberosum</i>	Potato	4113	1	1
<i>Sonneratia alba</i>	Mangrove Apple	122812	1	1
<i>Sonneratia apetala</i>	Mangrove	122813	1	1
<i>Sonneratia caseolaris</i>	Mangrove Crabapple	122814	1	1
<i>Sonneratia ovata</i>	Mangrove	122816	1	1
<i>Triticum aestivum</i>	Wheat	4565	1	1

cross-references is available in Supplementary File 1. The Gene Ontology, InterPro, NCBI Taxonomy, and Pfam were found in all the 511 UniProt entries. In order to construct a matrix, 40 of the cross-references were selected with references present in all entries removed as well as RefSeq, which had an identical

number of entries with Entrez Gene database. The matrix is available in the Supplementary File 1.

Twelve USP sequences were annotated with both the ArrayExpress and Ortholog Matrix Project (OMA) Browser (Fig. 4). Three *Arabidopsis* USP genes (Q93W91 [At3g62550], Q9LPF5 [At1g44760]

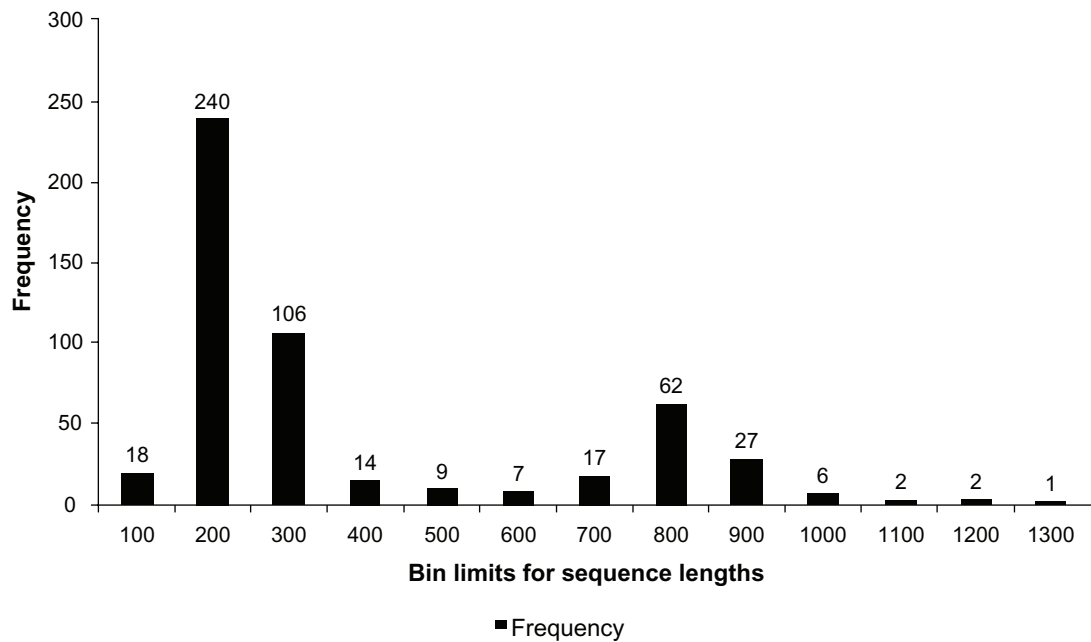


Figure 2. Distribution of sequence length of 511 viridiplantae universal stress proteins.

and Q9M328 [AT3g53990]) were up regulated in a drought microarray experiment stored in ArrayExpress and were annotated in the OMA Browser. Box plots of the three genes obtained from ArrayExpress as well as multiple sequence alignment of orthologs are

Table 2. Distribution of protein families in viridiplantae universal stress proteins.

Pfam ID*	Pfam name	Count
PF00582	Universal stress protein family	511
PF00069	Protein kinase domain	87
PF04564	U-box domain	34
PF07714	Protein tyrosine kinase	21
PF00999	Sodium/hydrogen exchanger family	5
PF03107	C1 domain	2
PF07649	C1-like domain	2
PF00651	BTB/POZ domain	1
PF01370	NAD dependent epimerase/dehydratase family	1
PF02637	GatB domain	1
PF03061	Thioesterase superfamily	1
PF04147	Nop14-like family	1
PF04185	Phosphoesterase family	1
PF05139	Erythromycin esterase	1
PF05699	hAT family dimerisation domain	1
PF08879	WRC	1
PF08880	QLQ	1

Note: *Description of protein domains available at <http://pfam.sanger.ac.uk/>.

presented in Figure 5. The OMA Browser provides multiple sequence alignment for groups of orthologs for each protein sequence (Fig. 5). Orthologous sequences were from *Oryza sativa*, *Sorghum bicolor*, *Populus trichocarpa* and *Vitis vinifera*.

Visual inspection of the alignments showed that the G-2X-G-9X-G (S/T) motif for small phosphoryl/ribosyl-binding residues of Adenosine Triphosphate (ATP)⁴⁹ was present in Q9M328 and Q93W91 but absent in Q9LPF5. Additional homologous sequences for the drought-responsive proteins provided by PLAZA³⁶ and ClustalW2 generated sequence alignments can be found in the Supplementary File 2. The multiple sequence alignment for 16 homologous sequences including drought responsive ATP-binding motif containing At3g62550 is presented in Figure 6. The conserved Aspartate (D) residue in position 12 of At3g62550 is known to be involved in adenine binding in ATP-binding USPs.^{15,50}

Viridiplantae universal stress protein gene transcripts derived from drought conditions

A total of 1561 ESTs clustered into 360 singletons and 185 Transcript Assemblies from 137 unique viridiplantae members (82 genera) and annotated with text “universal stress protein” were obtained from the TIGR Plant Transcript Assemblies


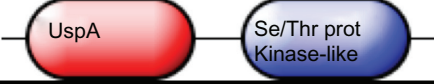
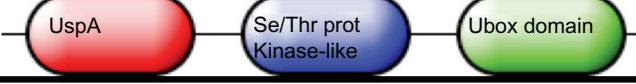
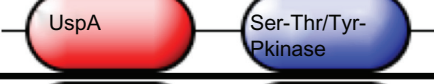
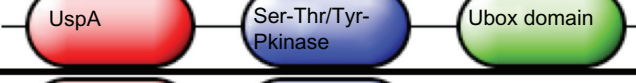


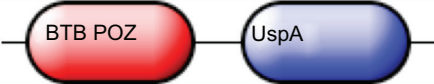
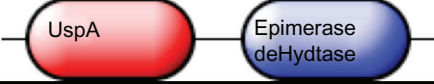
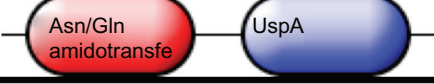
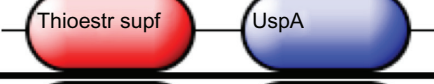

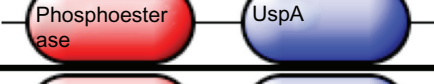



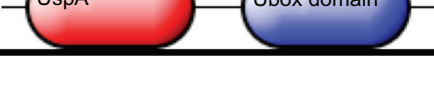
Domain Architecture	Pfam ID	Example	Count
	PF00582	A0MFH3_ARATH	387
	PF00582 PF00069	A2WRR9_ORYSI	59
	PF00582 PF00069 PF04564	A1YKF2_BRASY	27
	PF00582 PF07714	A9S9S5_PHYPA	14
	PF00582 PF07714 PF04564	A2X833_ORYSI	6
	PF00999 PF00582	A5BEW1_VITVI	5
	PF00582 PF03107 PF07649	Q84R08_ARATH	2
	PF00651 PF00582	A3CDI2_ORYSJ	1
	PF00582 PF01370	A3CFI4_ORYSJ	1
	PF02637 PF00582	A5B9V0_VITVI	1
	PF03061 PF00582	A5AGB5_VITVI	1
	PF04147 PF00582	A5AVR8_VITVI	1
	PF04185 PF00582	Q6WBY1_SOLTU	1
	PF00582 PF05139	A8JBW7_CHLRE	1
	PF00582 PF05699	A2WJV1_ORYSI	1
	PF08879 PF08880 PF00582	B4F168_MAIZE	1
	PF00582 PF04564	PUB54_ARATH	1

Figure 3. Protein domain architectures, examples and counts in dataset of plant universal stress proteins. Architecture images obtained from InterPro (www.ebi.ac.uk/interpro), an integrated database of predictive protein “signatures” for protein annotation and classification. The examples are UniProt identifiers with abbreviations for the plant taxa as follows—ORYSI: *Oryza sativa* subsp. *indica* (Rice); BRASY: *Brachypodium sylvaticum* (False brome); ARATH: *Arabidopsis thaliana* (Mouse-ear cress); ORYSJ: *Oryza sativa* subsp. *japonica* (Rice); VITVI: *Vitis vinifera* (Grape); CHLRE: *Chlamydomonas reinhardtii*; SOLTU: *Solanum tuberosum* (Potato); PHYPA: *Physcomitrella patens* subsp. *patens*; MAIZE: *Zea mays* (Maize).

**Table 3.** Viridiplantae universal stress proteins with tandem USP domains.

UniProt Identifier	Organism	Protein Length	Domain Coordinates
A8IXV1	<i>Brassica campestris</i>	220	13-56 79-198
B0YQX1	<i>Gossypium arboreum</i>	169	4-64 78-169
C1N7W4	<i>Micromonas pusilla</i> CCMP1545	343	90-160 187-322
C1MYP7	<i>Micromonas pusilla</i> CCMP1545	581	84-241 518-569
C1N599	<i>Micromonas pusilla</i> CCMP1545	396	48-102 155-249
C1E4R1	<i>Micromonas</i> sp"RCC299	567	295-355 431-567
C1FHK1	<i>Micromonas</i> sp"RCC299	267	12-93 120-256
A2ZLH5	<i>Oryza sativa</i> subsp"indica	320	21-155 167-312
A4RVM8	<i>Ostreococcus lucimarinus</i> (strain CCE9901)	274	67-164 195-250
A4RZS6	<i>Ostreococcus lucimarinus</i> (strain CCE9901)	401	26-215 237-378
Q015R5	<i>Ostreococcus tauri</i>	401	173-212 234-376
Q01BC4	<i>Ostreococcus tauri</i>	215	23-83 85-191

(Supplementary File 1). *Triticum aestivum* (bread wheat), *Oryza sativa* Japonica Group and *Glycine max* (soybean) had at least 100 ESTs annotated as encoding universal stress proteins. The 82 plant genera represented in the universal stress protein gene transcript dataset were clustered according to number of species or species combination (Table 5).

A dataset of 80 simple sequence repeats (SSRs) linked to 20 singletons and 47 transcript assemblies was constructed (Supplementary File 1). A total of 31 types of SSRs (3 uninucleotide; 7 dinucleotides; 16 trinucleotides; 1 tetranucleotide; 3 pentanucleotides; and 1 hexa-nucleotides) were retrieved (Table 6). The transcript count associated with each SSRs was also determined to identify potential unique EST-SSR markers. For example, the dinucleotide TA was unique for singleton DY959747 from *Lactuca sativa* (lettuce). The suggested primers for the identified EST-SSRs are available from the Planta website at <http://planta.jcvi.org/>.

The bioinformatics strategy retrieved 32 drought-responsive ESTs from 7 plant genera *Glycine*, *Hordeum*, *Manihot*, *Medicago*, *Oryza*, *Pinus* and *Triticum* (Table 7). Furthermore, the strategy revealed differentially expressed ESTs. In domesticated barley, two ESTs BM369974 and BQ761388 were expressed in the root while CD662497 was expressed in the lower leaf epidermis. In rice, two ESTs CK665047 and CA764828 were expressed in drought stressed leaf and drought stress panicle respectively. Integrating the datasets on SSRs and drought-responsive ESTs identified three drought-responsive ESTs

from *Triticum aestivum* (BE604157), *Glycine max* (BM887317) and *Pinus pinaster* (BX682209) (Table 8). The SSR sequence types were (CAG)₄, (ATA)₄ and (AT)₅ respectively.

Discussion

Plants are continuously exposed to abiotic and biotic stresses that require adaptation for survival. The availability of genomic sequences from a variety of viridiplantae has facilitated the dissection of the molecular, cellular and developmental responses to environmental stresses including drought.³⁷ Our investigation demonstrates the benefits of integrating data on universal stress proteins from comprehensive protein and transcript databases. The value-added and prioritized datasets produced presents new opportunities to better investigate the function of universal stress proteins from diverse plants. According to the focus of the investigation, the protein and gene transcript datasets are discussed in the context of response to drought and salt stress.

Construction of dataset of viridiplantae universal stress proteins

We have retrieved, mined and integrated comprehensive functional annotation data on 511 universal stress protein and 1561 ESTs sequences from the viridiplantae. A total of 161 plants with unique NCBI Taxonomy Identifier were associated with the sequences. Thus, we have provided a catalog of protein and gene transcripts from model and non-model plant species those of importance in agriculture,

**Table 4.** Selected UniProt cross-reference resources linked to plant universal stress proteins.

Database	USP UniProt entry count	Database	Web server
GO	511	Gene Ontology	http://www.geneontology.org/
InterPro	511	Integrated resource of protein families, domains and functional sites	http://www.ebi.ac.uk/interpro/
NCBI Taxonomy	511	NCBI Taxonomy Database	http://www.ncbi.nlm.nih.gov/taxonomy
Pfam	511	Pfam protein domain database	http://pfam.sanger.ac.uk/
EMBL	506	EMBL nucleotide sequence database	http://www.ebi.ac.uk/embl/
ProteinModelPortal	407	Protein Model Portal of the PSI-Nature Structural Biology Knowledgebase	http://www.proteinmodelportal.org/
Gene3D	379	Gene3D Structural and Functional Annotation of Protein Families	http://gene3d.biochem.ucl.ac.uk/Gene3D/
PubMed	366	PubMed	http://www.pubmed.gov
DOI	361	Digital Object Identifier	http://www.doi.org/
GeneID	264	Database of genes from NCBI RefSeq genomes	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene
RefSeq	264	NCBI Reference Sequences	http://www.ncbi.nlm.nih.gov/RefSeq/
EnsemblPlants	236	EnsemblPlants	http://plants.ensembl.org/
KEGG	234	KEGG: Kyoto Encyclopedia of Genes and Genomes	http://www.genome.jp/kegg/
PRINTS	227	Protein Motif fingerprint database; a protein domain database	http://umber.sbs.man.ac.uk/dbbrowser/PRINTS/
UniGene	163	UniGene gene-oriented nucleotide sequence clusters	http://www.ncbi.nlm.nih.gov/sites/entrez?db=UniGene
SMR	128	SWISS-MODEL Repository—a database of annotated 3D protein structure models	http://swissmodel.expasy.org/repository/
HOGENOM	116	The HOGENOM Database of Homologous Genes from Fully Sequenced Organisms	http://pbil.univ-lyon1.fr/databases/hogenom.php
PROSITE	110	PROSITE; a protein domain and family database	http://www.expasy.org/prosite/
SUPFAM	110	Superfamily database of structural and functional annotation	http://supfam.org
ProtClustDB	108	Entrez Protein Clusters	http://www.ncbi.nlm.nih.gov/sites/entrez?db=proteinclusters

ecology, industry and alternative energy. A catalog limited to *Arabidopsis* universal stress proteins has been published.¹³ The cross-database references available in our investigation present other researchers with a “one-stop-shopping” for sequences information on viridiplantae universal stress proteins.

The bioinformatics strategy extracted functional annotation data from comprehensive public domain protein and gene transcript databases. The Pfam

protein family database³⁶ served as the source of protein sequences for which their functional annotation data in the UniProt protein resource²⁶ were extracted and integrated with other specialized databases including those storing data on gene expression³⁸ and protein sequence evolution.³⁴ We also extracted functional annotation data from the Plantta EST resource, since ESTs are a source of genomic information especially for plants without complete genome

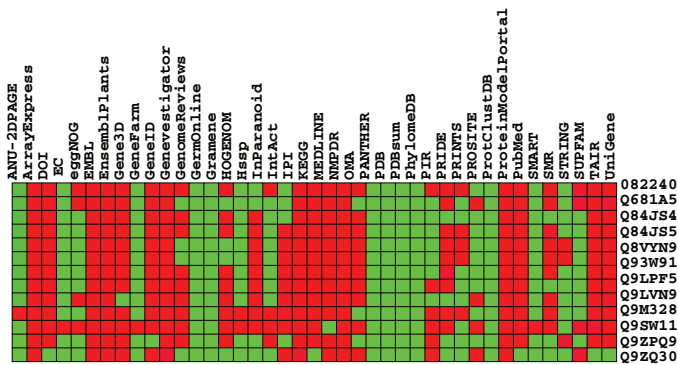


Figure 4. Visualization of matrix of availability of annotation with 40 external database references for selected plant universal stress proteins in UniProt. Description of column headings is documented in Supplementary File 1. **Notes:** Red, presence of database annotation; Green, absence of database annotation.

sequencing projects. The bioinformatics approach presented could be useful for other researchers interested in other protein families.

The particular function of a protein depends on its combination of domains. In general, the presence of the USP domain may provide the ability for the function of the other domain to be expressed under stress conditions. The USP domain appears as a single domain in small USP proteins (~14–15 kDa), as two domains arranged in tandem in larger USP proteins (~30 kDa), or as one or two USP domains together with other functional domains.^{9,13} Our analysis extracted and organized the domain combinations present in the 511 plant USPs thereby providing function-categorized subsets of the dataset. The categories can be investigated for shared function and regulation. Protein phosphorylation by kinases is a known pathway utilized by plants to response to osmotic stress.^{52,53}

Five proteins had annotation for the sodium/hydrogen exchanger family domain (PF00999), a domain for transport of sodium ions either out of cell or organelles in exchange for hydrogen ions to prevent toxic accumulation of sodium ions.^{54,55} The *Arabidopsis* gene encoding Na⁺/H⁺ exchanger termed salt overly sensitive (SOS1) is an important determinant of salt tolerance.³⁹ The list of uncharacterized proteins with both USP and Na⁺/H⁺ Exchanger included protein A9T441 from the moss *Physcomitrella patens*, the oldest clade of land plants⁴⁰ and that is highly tolerant against hyper salinity and severe water limitations.⁴¹ The 18 *P. patens* USPs in the dataset warrants further

investigation to understand the evolution of USPs from small land plants to higher plants after 450 years. The recognition of *P. patens* has a versatile tool for plant functional genomics could accelerate additional research of benefit to higher plants of importance in agriculture (eg, grapevine), industry (eg, castor plant) and cellulosic biofuels (eg, poplar).

Nine of the 12 protein sequences with tandem USP domains were from green algae. There are currently a limited number of reports on functional characterization of proteins with tandem USP domains.^{10,42,43} In *Escherichia coli*, mutants of UspE that contain tandem USP domains were unable to form cell-cell interactions and cell aggregates in stationary phase. In *Mycobacterium tuberculosis*, which has 8 of its 10 USPs having tandem domains, Rv2623 has growth-regulating capability linked to ATP-binding.⁴² A recent investigation observed higher degree of sequence identity between tandem domains in prokaryotes compared to eukaryotes.⁴⁴ The dataset analyzed did not including tandem USP domains. A starting point for characterization of tandem USP domain of plants could be to determine the sequence conservation between the domains.

Orthologous viridiplantae drought-responsive genes encoding universal stress proteins

Cross-referencing of specialized databases to a protein sequence entry in UniProtKB provides additional functional annotation that can help accelerate selection of plant USPs for characterization. The UniProtKB provides links to at least 126 specialized resources including plant bioinformatics databases such as The Arabidopsis Resource (TAIR),⁴⁵ Gramene,⁴⁶ and EnsemblPlants.⁴⁷ We have integrated available database cross-references to provide a visual view of databases across the viridiplantae USPs analyzed. The utility of such view was demonstrated on a subset of proteins that were annotated with ArrayExpress⁴⁵ and Ortholog Matrix Project (OMA) Browser.³⁴ This view enabled us to easily identify Q9SW11 (U-box domain-containing protein 35; At4g25160, PUB35) as an enzyme based on the presence of the Enzyme Commission (EC) number (Fig. 4: Column 4, Row 10). The U-box domain for regulated protein ubiquitination and degrada-



Figure 5. Gene expression and protein sequence alignment of *Arabidopsis thaliana* USPs up-regulated in response to drought. Detail gene expression and protein sequence alignment can be obtained by using the following weblinks respectively by replacing the <accession> with the UniProt protein identifier.

<http://www.ebi.ac.uk/gxa/experiment/E-MEXP-1863/<accession>>.
<http://omabrowser.org/cgi-bin/gateway.pl?f=DisplayGroup&p1=<accession>>.

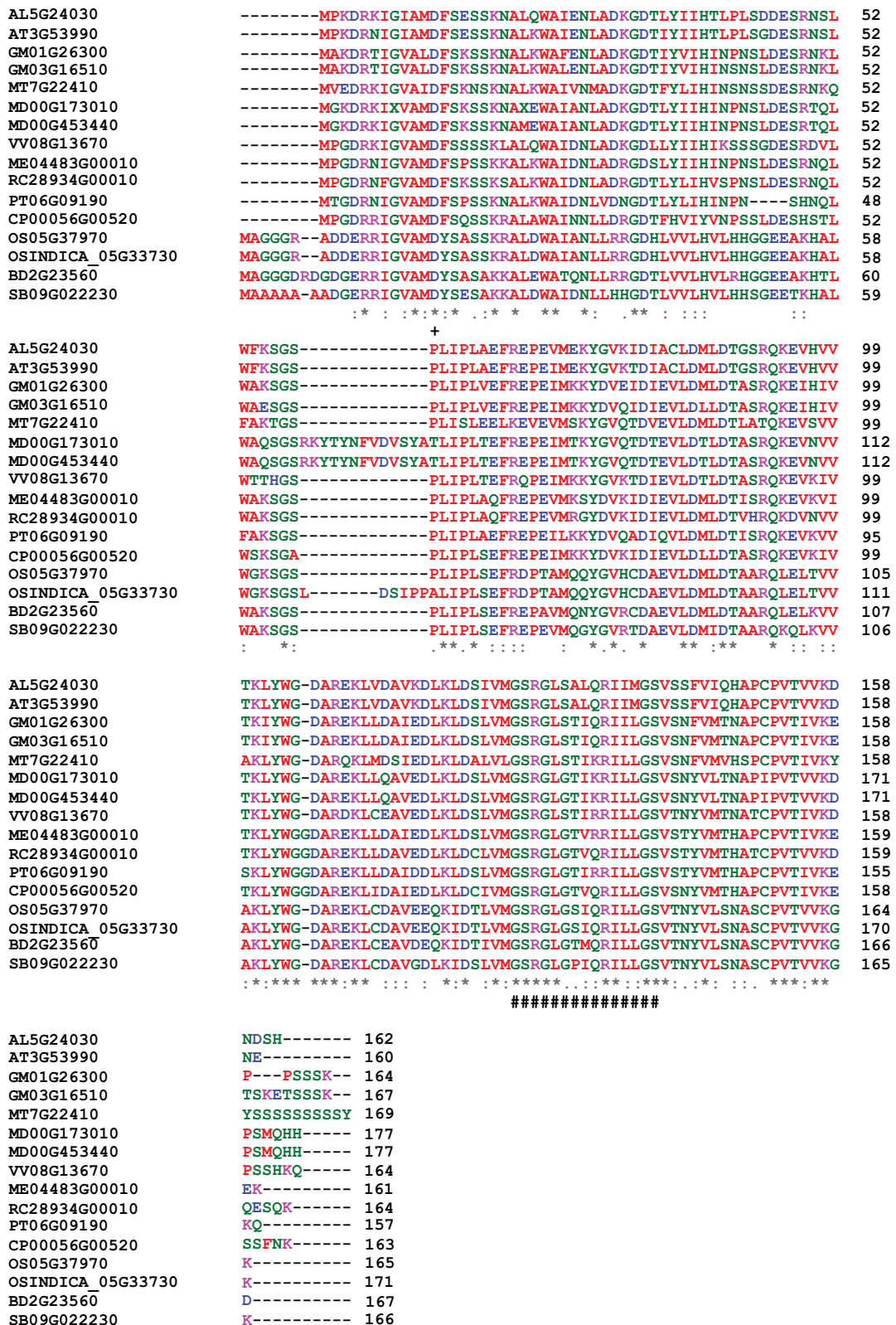


Figure 6. Multiple sequence alignment of drought-responsive Arabidopsis thaliana universal stress protein At3g53990 and homologs. The conserved Aspartate (D) residue in position 12 of At3g62550 (marked with +) is known to be involved in adenine binding in ATP-binding USPs.^{12,44} The region for small phosphoryl/ribosyl-binding residues of ATP is indicated with a series of #. The first two letters of the sequence name correspond to the plant: AL, Arabidopsis lyrata; AT, Arabidopsis thaliana; BD, Brachypodium distachyon; CP, Carica papaya; GM, Glycine max; MD, Malus domestica; ME, Manihot esculenta; MT, Medicago truncatula; OS, Oryza sativa ssp. Japonica; OSAINDICA, Oryza sativa ssp. Indica; PT, Populus trichocarpa; RC, Ricinus communis; SB, Sorghum bicolor; VV, Vitis vinifera.

**Table 5.** Plant genera represented in universal stress protein gene transcripts dataset.

Genus	Number of species
<i>Populus</i>	9
<i>Helianthus</i>	6
<i>Citrus</i>	5
<i>Oryza Picea</i>	4
<i>Fragaria Gossypium Lactuca Medicago Nicotiana Prunus Saccharum Sorghum Triticum</i>	3
<i>Agrostis Apium Arachis Centaurea Euphorbia Hordeum Petunia Phaseolus Pinus</i>	2
<i>Pseudotsuga Rosa Solanum Taraxacum Vitis</i>	
<i>Aegilops Allium Ananas Antirrhinum Avena Avicennia Brachypodium Brassica Capsicum Catharanthus Ceratopteris Chlamydomonas Cichorium Coffea Curcuma Cyamopsis Cycas Eragrostis Eucalyptus Festuca Ginkgo Glycine Hedyotis Ipomoea Juglans Lolium Lotus Malus Manihot Marchantia Mesembryanthemum Mesostigma Mimulus Panax Panicum Pennisetum Phalaenopsis Physcomitrella Pisum Rhododendron Ricinus Salvia Secale Selaginella Sesamum Syntrichia Thellungiella Theobroma Vaccinium Welwitschia Zamia Zantedeschia Zingiber Zinnia</i>	1

Table 6. Simple Sequence Repeats (SSR) linked to universal stress protein gene transcripts.

SSR	Universal stress protein gene transcripts	Nucleotide count	Transcript count
A	CN463769 TA10796_2711 TA10796_2711 TA33493_4530 TA36088_4113 TA36088_4113 TA70577_4565	1	7
T	CI395583 EC939973 TA70577_4565 TA72057_3847	1	4
G	TA70577_4565	1	1
AG	DW142996 TA1316_69721 TA35794_29760 TA4456_3988	2	4
AT	TA1967_153471 TA2761_80863 TA3030_71647	2	3
GA	TA14075_2711 TA3367_309804	2	2
TG	TA49270_4530 TA49271_4530	2	2
CT	TA16418_3330	2	1
GT	TA33493_4530	2	1
TA	DY959747	2	1
CAG	AJ610677 AL825196 CJ661413 CJ668094 TA52848_4565 TA53203_4565 TA53303_4565 TA53312_4565 TA53408_4565	3	9
CGC	CA181583 CI776988 DT694744 TA26627_4558 TA32279_4513 TA33491_4530 TA36096_4547 TA3991_132711	3	8
CCG	TA2412_4568 TA41380_4513 TA41381_4513 TA41598_4513 TA55332_4565 TA55439_4565	3	6
AAG	DY942109 DY953240 TA10188_4232 TA1962_3197	3	4
GAA	DT694744 DY923603 EE657448 TA25103_4558	3	4
ATA	BQ473543 TA48543_3847 TA48544_3847	3	3
GGT	CI602544 TA49270_4530 TA49271_4530	3	3
CGT	TA46082_3847 TA48499_4547	3	2
GCA	CD220323 TA25103_4558	3	2
GGC	TA2176_4120 TA51553_4530	3	2
AGG	TA75441_4530	3	1
ATT	TA36088_4113	3	1
CGA	DT694744	3	1
CGG	TA1497_94328	3	1
GAG	TA15268_29730	3	1
TGT	DR575687	3	1
TTGT	TA3984_73275	4	1
AAAAT	TA762_4615	5	1
CACCC	TA32279_4513	5	1
TTTAA	TA3585_36596	5	1
GCGGCT	TA41381_4513	6	1

**Table 7.** Drought-annotated plant Expressed Sequence Tags (ESTs)

EST	Plant	Source of EST library
BM886962	<i>Glycine max</i> (soybean)	*
BM887317	<i>Glycine max</i> (soybean)	*
CD662497	<i>Hordeum vulgare</i> subsp. <i>vulgare</i> (domesticated barley)	Lower leaf epidermis
BM369974	<i>Hordeum vulgare</i> subsp. <i>vulgare</i> (domesticated barley)	Root
BQ761388	<i>Hordeum vulgare</i> subsp. <i>vulgare</i> (domesticated barley)	Root
DV442544	<i>Manihot esculenta</i> (cassava)	**
DV442765	<i>Manihot esculenta</i> (cassava)	**
DV443464	<i>Manihot esculenta</i> (cassava)	**
DV444643	<i>Manihot esculenta</i> (cassava)	**
DV446035	<i>Manihot esculenta</i> (cassava)	**
DV446427	<i>Manihot esculenta</i> (cassava)	**
DV447334	<i>Manihot esculenta</i> (cassava)	**
DV454753	<i>Manihot esculenta</i> (cassava)	***
DV455089	<i>Manihot esculenta</i> (cassava)	***
DV455235	<i>Manihot esculenta</i> (cassava)	***
DV455909	<i>Manihot esculenta</i> (cassava)	***
DV456031	<i>Manihot esculenta</i> (cassava)	***
DV456176	<i>Manihot esculenta</i> (cassava)	***
DV456576	<i>Manihot esculenta</i> (cassava)	***
DV456911	<i>Manihot esculenta</i> (cassava)	***
DV457684	<i>Manihot esculenta</i> (cassava)	***
BE248764	<i>Medicago truncatula</i> (barrel medic)	Plantlets
BF631735	<i>Medicago truncatula</i> (barrel medic)	Plantlets
BF634145	<i>Medicago truncatula</i> (barrel medic)	Plantlets
BF634785	<i>Medicago truncatula</i> (barrel medic)	Plantlets
CK665047	<i>Oryza sativa</i> Indica Group	Leaf
CA764828	<i>Oryza sativa</i> Indica Group	Panicles
BX680935	<i>Pinus pinaster</i>	Root
BX682209	<i>Pinus pinaster</i>	Root
BE604157	<i>Triticum aestivum</i> (bread wheat)	Leaf
BE428779	<i>Triticum turgidum</i> subsp. <i>durum</i> (durum wheat)	Root
BE429106	<i>Triticum turgidum</i> subsp. <i>durum</i> (durum wheat)	Root

Notes: *Leaf, drought stressed, 1 month old plants, greenhouse grown; **Mature leaf and petiole, young leaf and apical meristem, root, tuber and tuber peel, young leaf and apical meristem midnight; ***Young leaf and apical meristem, mature leaf and petiole, root, tuber and tuber peel from water stressed plants.

tion is a modified RING-finger domain involved in protein that lacks metal-binding ability.⁴⁸ Comparative structural and functional assays could reveal the interactions of the USP domain and the enzyme domains present in Q9SW11. Orthologous drought-responsive universal stress proteins could be candidates to engineer desired phenotypes in plants. Our analyses identified three *Arabidopsis* proteins (Fig. 5) and their orthologs in *Oryza sativa*, *Sorghum bicolor*, *Populus trichocarpa* and *Vitis vinifera*. Q9M328 and Q93W91 and their homologs

could be regulated by ATP based on the presence of ATP-binding motif (Fig. 6).

Viridiplantae universal stress protein gene transcripts derived from drought conditions

Expressed Sequence Tags generated from stress-challenged plant tissues have been used as high quality transcripts to discover genes, identify candidate stress-responsive genes/transcripts and identify functional markers such as genic microsatellites and



Table 8. Drought-responsive Expressed Sequence Tags (ESTs) with microsatellites.

EST	Plant	Tissue	Plantta TA	Plantta SSR ID	SSR	Number of repeats	Transcript length (bp)	Start	End
BE604157	<i>Triticum aestivum</i> (bread wheat)	Leaf	TA53312_4565	233072	CAG	4	810	205	216
BM887317	<i>Glycine max</i> (soybean)	Leaf, drought stressed, 1 month old plants, greenhouse grown	TA48544_3847	815126	ATA	4	889	544	555
BX682209	<i>Pinus pinaster</i> (maritime pine)	Root	TA3030_71647	751586	AT	5	414	199	208

single nucleotide polymorphisms.^{49–51} The effects of SSR type as well as number of repeats on gene regulation, transcription and protein function are poorly understood in plants when compared to human or animal systems.⁵¹ In this article we report automatic extraction of information on simple sequence repeats (SSRs) associated with 1561 ESTs in the Plantta resource.²⁷ Our analysis identified candidate USP gene transcripts in multiple plants (Supplementary File 1 and Table 5); organized the SSRs into types (Table 6), drought-annotated USP ESTs (Table 7) and USP EST-SSRs from drought-stress tissues (Table 8). The majority (49 of 80) of the USP EST-SSRs was the trinucleotide type, which has been reported to be the most abundant in rice, wheat and barley^{52,53} as well as peanut⁵⁴ and citrus.⁵⁵ All together, our analyses provide a comprehensive collection of USP ESTs including those responsive to drought. We have clustered the plant genera based on the number of species to facilitate investigating the EST-SSR and EST-Single Nucleotide Polymorphisms (SNPs) in USP genes for comparative mapping, transferability, genetic diversity and plant improvement.

Conclusions

The molecular mechanisms by which genes encoding the universal stress protein domain are able to confer in plants the ability to respond and adapt to environmental changes are not well defined. We have computationally retrieved, mined and integrated functional annotations on protein and gene transcripts that encode the universal stress protein domain. The datasets from cross-database mining provide organized resources for the characterization of USP genes as useful targets for engineering plant varieties tolerant to unfavorable environmental conditions.

Acknowledgments

Mississippi NSF-EPSCoRAward (EPS-0903787); Research Centers in Minority Institutions (RCMI)—Center for Environmental Health at Jackson State University (NIH-NCRR G12RR013459); Pittsburgh Supercomputing Center's National Resource for Biomedical Supercomputing (T36 GM008789); US Department of Homeland Security Science and Technology Directorate (2007-ST-104-000007; 2009-ST-062-000014; 2009-ST-104-000021). SSS was a Louis Stokes Mississippi Alliance for Minority Participation (LSMAMP)



Fellow in 2005 and is currently a PhD Candidate in the Environmental Science PhD Program at Jackson State University. Disclaimer: The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.

Disclosures

This manuscript has been read and approved by all authors. This paper is unique and not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

- Porter JR, Semenov MA. Crop responses to climatic variation. Philosophical transactions of the Royal Society of London Series B. *Biological Sciences*. 2005;360:2021–35.
- Valliyodan B, Nguyen HT. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Current Opinion in Plant Biology*. 2006;9:189–95.
- Hirayama T, Shinozaki K. Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant Journal*. 2010;61:1041–52.
- Mochida K, Shinozaki K. Genomics and bioinformatics resources for crop improvement. *Plant and Cell Physiology*. 2010;51:497–523.
- Tatusova T, Smith-White B, Ostell J. A collection of plant-specific genomic data and resources at NCBI. *Methods in Molecular Biology*. 2007;406:61–87.
- Kvint K, Nachin L, Diez A, Nystrom T. The bacterial universal stress protein: function and regulation. *Current Opinion in Microbiology*. 2003;6:140–5.
- Nystrom T, Neidhardt FC. Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in *Escherichia coli*. *Molecular Microbiology*. 1992;6:3187–98.
- Persson O, Valadi A, Nystrom T, Farewell A. Metabolic control of the *Escherichia coli* universal stress protein response through fructose-6-phosphate. *Molecular Microbiology*. 2007;65:968–78.
- Siegele DA. Universal stress proteins in *Escherichia coli*. *Journal of Bacteriology*. 2005;187:6253–4.
- Nachin L, Nannmark U, Nystrom T. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *Journal of Bacteriology*. 2005;187:6265–72.
- Sousa MC, McKay DB. Structure of the universal stress protein of *Haemophilus influenzae*. *Structure*. 2001;9:1135–41.
- Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, et al. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proceedings of the National Academy of Sciences of the United States*. 1998;95:15189–93.
- Kerk D, Bulgrien J, Smith DW, Gribskov M. *Arabidopsis* proteins containing similarity to the universal stress protein domain of bacteria. *Plant Physiology*. 2003;131:1209–19.
- Sauter M, Rzewuski G, Marwedel T, Lorbiecke R. The novel ethylene-regulated gene OsUsp1 from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. *Journal of Experimental Botany*. 2002;53:2325–31.
- Chou MX, Wei XY, Chen DS, Zhou JC. A novel nodule-enhanced gene encoding a putative universal stress protein from *Astragalus sinicus*. *Journal of Plant Physiology*. 2007;164:764–72.
- Becker JD, Moreira LM, Kapp D, et al. The nodulin vfENOD18 is an ATP-binding protein in infected cells of *Vicia faba* L. nodules. *Plant Molecular Biology*. 2001;47:749–59.
- Hohnjec N, Kuster H, Albus U, et al. The broad bean nodulin VfENOD18 is a member of a novel family of plant proteins with homologies to the bacterial MJ0577 superfamily. *Molecular and General Genetics*. 2000;264:241–50.
- Maqbool A, Zahur M, Husnain T, Riazuddin S. GUSP1 and GUSP2, two drought-responsive genes in *Gossypium arboreum* have homology to universal stress proteins. *Plant Molecular Biology Reporter*. 2009;27:6.
- Li WT, Wei YM, Wang JR, et al. Identification, localization, and characterization of putative USP genes in barley. *Theoretical and Applied Genetics*. 2010;121:907–17.
- Yokota A, Kawasaki S, Iwano M, et al. Citrulline and DRIP-1 protein (ArgE homologue) in drought tolerance of wild watermelon. *Annals of Botany*. 2002;89 Spec No:825–32.
- Shinozaki K, Yamaguchi-Shinozaki K. Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany*. 2007;58:221–7.
- Bray EA. Classification of genes differentially expressed during water-deficit stress in *Arabidopsis thaliana*: an analysis using microarray and differential expression data. *Annals of Botany*. 2002;89 Spec No:803–11.
- Coetzer N, Gazendam I, Oelofse D, Berger DK. SSHscreen and SSHdb, generic software for microarray based gene discovery: application to the stress response in cowpea. *Plant Methods*. 2010;6:10.
- Brazas MD, Yamada JT, Ouellette BF. Providing web servers and training in Bioinformatics: 2010 update on the Bioinformatics Links Directory. *Nucleic Acids Research*. 2010;38 Suppl:W3–6.
- Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Research*. 2010;38:D211–22.
- UniProt Consortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research*. 2010;38:D142–8.
- Childs KL, Hamilton JP, Zhu W, et al. The TIGR Plant Transcript Assemblies database. *Nucleic Acids Research*. 2007;35:D846–51.
- Lazo GR, Chao S, Hummel DD, et al. Development of an expressed sequence tag (EST) resource for wheat (*Triticum aestivum* L.): EST generation, unigene analysis, probe selection and bioinformatics for a 16,000-locus bin-delineated map. *Genetics*. 2004;168:585–93.
- Tang S, Okashah RA, Cordonnier-Pratt MM, et al. EST and EST-SSR marker resources for Iris. *BMC Plant Biology*. 2009;9:72.
- Cuadrado A, Cardoso M, Jouve N. Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. *Cytogenetic and Genome Research*. 2008;120:210–9.
- Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution*. 2004;21:991–1007.
- Aishwarya V, Sharma PC. UgMicroSatdb: database for mining microsatellites from unigenes. *Nucleic Acids Research*. 2008;36:D53–6.
- Boguski MS, Lowe TM, Tolstoshev CM. dbEST—database for “expressed sequence tags”. *Nature Genetics*. 1993;4:332–3.
- Schneider A, Dessimoz C, Gonnet GH. OMA Browser—exploring orthologous relations across 352 complete genomes. *Bioinformatics*. 2007;23:2180–2.
- Pavlidis P, Noble WS. Matrix2png: a utility for visualizing matrix data. *Bioinformatics*. 2003;19:295–6.
- Proost S, van BM, Sterck L, et al. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell*. 2009;21:3718–31.
- Tuberosa R, Salvi S. Genomics-based approaches to improve drought tolerance of crops. *Trends in Plant Science*. 2006;11:405–12.
- Parkinson H, Kapushesky M, Kolesnikov N, et al. ArrayExpress update— from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Research*. 2009;37:D868–72.
- Qiu QS, Barkla BJ, Vera-Estrella R, Zhu JK, Schumaker KS. Na⁺/H⁺ exchange activity in the plasma membrane of *Arabidopsis*. *Plant Physiology*. 2003;132:1041–52.
- Reski R, Frank W. Moss (*Physcomitrella patens*) functional genomics—Gene discovery and tool development, with implications for crop plants and human health. *Briefings in Functional Genomics and Proteomics*. 2005;4:48–57.



41. Frank W, Ratnadewi D, Reski R. *Physcomitrella patens* is highly tolerant against drought, salt and osmotic stress. *Planta*. 2005;220:384–94.
42. Drumm JE, Mi K, Bilder P, et al. *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. *PLoS Pathogens*. 2009;5:e1000460.
43. Hingley-Wilson SM, Lougheed KE, Ferguson K, Leiva S, Williams HD. Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable in vitro. *Tuberculosis (Edinb)*. 2010;90:236–44.
44. Reshef D, Itzhaki Z, Schueler-Furman O. Increased sequence conservation of domain repeats in prokaryotic proteins. *Trends in Genetics*. 2010;26:383–7.
45. Lamesch P, Dreher K, Swarbreck D, et al. Using the *Arabidopsis* information resource (TAIR) to find information about *Arabidopsis* genes. *Current Protocols in Bioinformatics*. 2010;Chapter 1:Unit 1.
46. Jaiswal P. Gramene database: a hub for comparative plant genomics. *Methods in Molecular Biology*. 2011;678:247–75.
47. Kersey PJ, Lawson D, Birney E, et al. Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic Acids Research*. 2010;38:D563–9.
48. Aravind L, Koonin EV. The U box is a modified RING finger—a common domain in ubiquitination. *Current Biology*. 2000;10:R132–4.
49. Guo B, Chen X, Dang P, et al. Peanut gene expression profiling in developing seeds at different reproduction stages during *Aspergillus parasiticus* infection. *BMC Developmental Biology*. 2008;8:12.
50. Raju NL, Gnanesh BN, Lekha P, et al. The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). *BMC Plant Biol*. 2010;10:45.
51. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*. 2005;23:48–55.
52. Gupta PK, Rustgi S, Sharma S, et al. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics*. 2003;270:315–23.
53. La RM, Kantety RV, Yu JK, Sorrells ME. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics*. 2005;6:23.
54. Liang X, Chen X, Hong Y, et al. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biology*. 2009;9:35.
55. Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr. Mining and characterizing microsatellites from citrus ESTs. *Theoretical and Applied Genetics*. 2006;112:1248–57.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>