Australian Journal of Crop Science| Pages 28-37 | DOI: 10.21475/ajcs.21.15.09.sp-3

Identification of small open reading frames (sORFs) associated with heat tolerance in nitrogen-fixing root nodules of *Phaseolus vulgaris* wild-type and cv BAT93

Alejandra Zayas-del Moral[1], Damián Martínez-Reyes[2], Carmen Quinto[1], Federico Sanchez† and Claudia Díaz-Camino[1]*

[1]Departamento de Biología Molecular de Plantas Instituto de Biotecnología, Universidad Nacional Autónoma de México Av. Universidad 2001, CP 62210, Cuernavaca, Morelos, México
[2]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México AP 565A Cuernavaca, Morelos, México
† In loving memory.

***Corresponding author: claudia@ibt.unam.mx**

Abstract

Common bean is an important legume crop and a major source of protein for low-income groups around the world. Legumes have the ability to engage symbiotic interactions with nitrogen-fixing soil bacteria. In this study, next-generation sequencing technology was used to perform transcriptome analyses of a yet unexplored group of peptides encoded by small open reading frames (sORFs; < 150 codons) in nitrogen-fixing symbiotic nodules of two heat-tolerant genotypes of common bean (*Phaseolus vulgaris* L): the cultivar BAT93 and a wild genotype (named *P. vulgaris* 7) from the south of Mexico. After heat stress, total RNA was isolated and used for transcriptome analysis. Sixty differentially expressed sORFs were identified between control and heat stress treatments. The expression profiles of these sORFs suggest that, regardless the evolutionary closeness between *P. vulgaris* BAT93 and *P. vulgaris* 7, each genotype has independently adapted their molecular signaling pathways to survive heat stress. The dataset developed may provide a useful resource for future genetic and genomic studies in these species.

Keywords: Heat stress, small open reading-frames, common bean, legume-rhizobia symbiosis, biological nitrogen-fixation, next-generation sequencing, transcriptome analysis.
Abbreviations: sORFs - short open reading frames, SPs - small proteins.

Introduction

The world´s human population is expected to reach 9.1 billion in 2050 (Food and Agriculture Organization of the United Nations, http://www.fao.org/). Over-population is associated with increasing global consumption of resources, food security and climate change. Recent climate models estimate that the global surface temperature is likely to rise by 4.8 °C in the worst-case scenario (IPCC, 2014). In semi-arid and tropical regions, which are among the most populated and under-developed, the increase in surface temperatures will severely affect crop production (IPCC, 2014). Legumes include important grain, pasture, and agroforestry species, and are second to cereal crops in agricultural importance based on area harvested and total production (http://www.fao.org/). Pulses (dry seeded legumes) are protein rich and affordable foods, and an important component in human sustenance, especially in the dietary pattern of low-income people in developing countries. In addition to their nutritional value, most legumes have symbiotic associations with nitrogen-fixing bacteria mainly belonging to the Rhizobiaceae family (rhizobia) (Dénarié et al., 1996). This remarkable biological interaction culminates with the formation of specialized root organs, the symbiotic nodules, where biological fixation of atmospheric nitrogen takes place. Nodulated legumes produce substantial amounts of organic nitrogen fertilizer and play a key role in sustainable agriculture in tropical and temperate climates (Peoples et al., 1995; Tate, 1995). Numerous studies have established that high temperatures (30°C to 40°C depending on species) have negative impacts on Rhizobium soil survival, root bacterial attraction and infection, and also nodule development (Lebrazi & Fikri Benbrahim, 2014; Abd-Alla et al., 2014). Knowledge of nodule functioning after heat stress, such as those experienced by legume plants in the field during a day is limited.

Small proteins (SPs) have emerged as an important class of signaling molecules involved in nodulation (Batut et al., 2011), and also in growth, development, and in response to stress (Hanada et al., 2012; Marmiroli & Maestri, 2014). SPs are encoded by short open reading frames (sORFs) and distinguished from other ORFs by their sizes (30-150 codons in length). Although many sORFs play important roles as regulators of diverse biological processes, this gene group usually escapes gene annotation because they are particularly difficult to predict by computational biology due to their small size. Thus, sORFs have been studied in only a few plant species and their biological importance is little understood. Here a comparative analysis was made of the expression of sORFs of the root-nodule transcriptome of two *P. vulgaris* genotypes under control and stress treatments (sudden and prolonged heat exposure). Computational strategies were deployed to identify sORFs that were up-regulated in active nitrogen-fixing nodules under heat stress. This information may be relevant in selecting new bean genotypes able to harbour active nitrogen-fixing nodules resilient to heat stress.

Results

Phenotypic and molecular responses to heat stress in P. vulgaris heat-stress resistant genotypes

To evaluate the ability of *P. vulgaris* cv. BAT93 and *P. vulgaris* 7 to adapt to sudden heat stress (without any priming, known as basal thermo-tolerance), the plants of each genotype were subjected to heat stress for 6 continuous hours. After stress, the aerial plant parts of *P. vulgaris* BAT93 and *P. vulgaris* 7 were photographed (Figure 1a and 1b). *P. vulgaris* BAT93 wilted (Fig 1a), and only a third of the plants subjected to heat were able to recover after one week in benign conditions. Recovered plants of BAT93 just produced one pod, in general with 1 seed (Table 1). In contrast *P. vulgaris* 7 showed no differences in foliar turgor nor seed production between control and stress treatments (Table 1). There were no differences in nodule size and number of nodules per root plant (Figure 1e), but the results show a significant and similar increase in chaperone transcript accumulation (Figure 1c), an indication that nodules of both genotypes responded to heat stress. Interestingly, the rate of nitrogen-fixation in nodules elicited by *R. tropici* CIAT899 in *P. vulgaris* BAT93 was severely reduced in heat-shock treated plants compared to control, but this effect was not observed in nodules of the wild *P. vulgaris* 7 (Figure 1 d). The data indicate that the basal thermo-tolerance of *P. vulgaris* 7 is higher than *P. vulgaris* BAT93, and that the biological nitrogen fixation process is not altered in *P. vulgaris* 7 by heat stress.

Distribution of sORFs in P. vulgaris BAT93 and G19833 genotypes and in other model legumes

Some 64,692 and 31,638 ORFs from *P. vulgaris* genotypes Mesoamerican BAT93 and Andean G19833, respectively (Vlasova et al., 2016, Schmutz et al., 2014), 88,647 ORFs from *Glycine max* (Schmutz et al., 2010), 10,979 ORFs from *Lotus japonicus* (Sato et al., 2008) and 62,319 ORFs from *Medicago truncatula* (Young et al., 2011) were collected from Phytozome version 11 (www.phytozome.net; Goodstein et al., 2012) and from miyakogusa.jp version 3.0 for *L. japonicus* (http://www.kazusa.or.jp/lotus). The ratio of sORFs (30 to 150 amino acids length) *versus* the total number of ORFs reported for each genome version was calculated (Figure 2 and Table 2). Although the annotations of total ORFs have changed in recent years in all the genomes of listed legumes (Guillén et al., 2013), the highest frequency of sORFs was found in the best-studied genomes of leguminous plants, i.e. *M. truncatula* and *L. japonicus* (0.2 and 0.3, respectively), while the proportion of sORFs/ORFs annotated in the *P. vulgaris* genomes fluctuates between 0.11 to 0.14, a slight difference that may be due to annotation systems used in these genomes.

Differential expression analysis of sORFs under heat-shock conditions in P. vulgaris

Out of 235 differentially expressed ORFs in *P. vulgaris* BAT93 under heat-stress (data not shown), 16 (6.8%) were sORFs. Most differentially expressed sORFs could not be assigned to any gene ontology (GO) category (Table 3 and 4), so these were analyzed by the MEME Suite (Figure S1), and also by BLASTP, which was found to be the most informative algorithm. In stressed root nodules of *P. vulgaris* BAT93, a histone and a thymidine kinase domain are present in three down-regulated sORFs (Figure 3 and Table 3). Five sORFs in *P. vulgaris* BAT93 nodules under heat-stress were up-regulated, andwe identified known protein domains in two of them (Figure 3a and Table 3): a domain found in SL33 plant splicing factors (PHASIBEAM10F006374), and a cytocrome-c oxidase domain (PHASIBEAM10F012744). In *P. vulgaris* 7 significant expression changes were detected in 1,064 ORFs, and 44 (4.1%) of them were identified as sORFs. A GO associated function could be annotated in 26 sORFs (Table 4), but in this case, neither the use of MEME (Figure S1), nor the BLASTP algorithm gave additional information over the putative biological function of some other sORFs of this group (Table 4). In heat-stressed nodules of *P. vulgaris* 7, 13 sORFs were down-regulated (Figure 3 b and Table 4). Most protein domains of these sORFs are yet unknown, or belong to proteins with no described biological function. However, some protein domains found in stress-related proteins were identified (Phvul.008G112900.1, Phvul.008G189400.1, Phvul.009G027600.1.) in a growth factor (Phvul.003G233400.1) (Yang et al., 2001), cytochrome b5 (Phvul.006G115900.1) and in proteins responsive to gibberellic acid (Phvul.008G235300.1), respectively. Up-regulated sORFs in these root nodules included calmodulin-like domains present in proteins involved in the signaling of calcium (Phvul.001G155400.1, Phvul.001G260700.1, Phvul.003G115800.1, Phvul.007G111200.1, Phvul.007G278900.1), and in phyto-hormone responsive proteins, such as ethylene, auxin or gibberellin (Phvul.007G193400.1, Phvul.007G219700.1, Phvul.009G015900.1, Phvul.010G019700.1).

Discussion

Small proteins encoded by small open reading frames (sORFs, 30 to 150 codons) have been shown to be relevant in legume-rhizobia interactions as well as in plant growth and development, and in response to stress (Batut et al., 2011; Hanada et al., 2012; Marmiroli & Maestri, 2014). sORFs identification can be predicted by bioinformatics approaches, such as web-based tools [sORFfinder (Hanada et al., 2010), HAltORF (Vanderperre et al., 2012), or uPEPperoni (Skarshewski et al., 2014)] by homology with other related-species, or by sequence analysis and clustering. Several molecular techniques are used to confirm sORFs gene expression, among these next-generation-sequencing technologies are reliable, e.g. RNA-seq. In this work, RNA-seq technology was used to gather relevant data on changes in gene expression of small proteins encoded by sORFs in root nodules of two *P. vulgaris* genotypes elicited by *R. tropici* CIAT899, a bacterium resistant to heat (Martínez-Romero et al. 1991), under prolonged heat stress conditions.

P. vulgaris BAT93, a representative cultivar of the Mesoamerican common bean gene pool, was bred for high productivity in tropical conditions at the Centro Internacional de Agricultura Tropical (CIAT), Colombia (Voysest, 1983, 2000). This breeding line has been well-studied, and its genome has been recently sequenced (Vlasova et al., 2016). Taken in consideration all these advantages, *P. vulgaris* BAT93 was chosen as the reference genotype to compare with *P. vulgaris* 7, which is a wild-type genotype collected from the south of México. Plant responses to heat stress in both common bean genotypes were confirmed by the strong induction of heat-shock proteins (HSps) (Figure 1 c) (Wang et al., 2004, Aparicio et al., 2005, Larkindale et al., 2005; Kim et al., 2011). Interestingly, although the induction of HSPs was similar in both common bean genotypes (Figure 1c), deleterious phenotypic effects at the whole plant level were observed only in *P. vulgaris* BAT93 (Figure 1a compared to 1b and Table 1).

Table 1. Phenotypic responses to heat stress in *Phaseolus vulgaris* cv. BAT93 and in *P. vulgaris* 7 genotypes.

	GENOTYPE			
FEATURES	*P. vulgaris* BAT93		*P. vulgaris* 7	
	CTRL	HS	CTRL	HS
Survival after heat stress (6h 37ºC, 3 plants per replicate, 3 technical replicates)	NA	1/3	NA	3/3
Average Pods per plant	3	0.33	>3	>3
Average Seeds per pod	5	1	4	4
Average seed weight (g)	0.186 g	0.168 g	0.046 g	0.058 g
Average nodule number per plant	84.48		44.71	
Average nodule dry weight per plant (g)	0.02		0.0055	

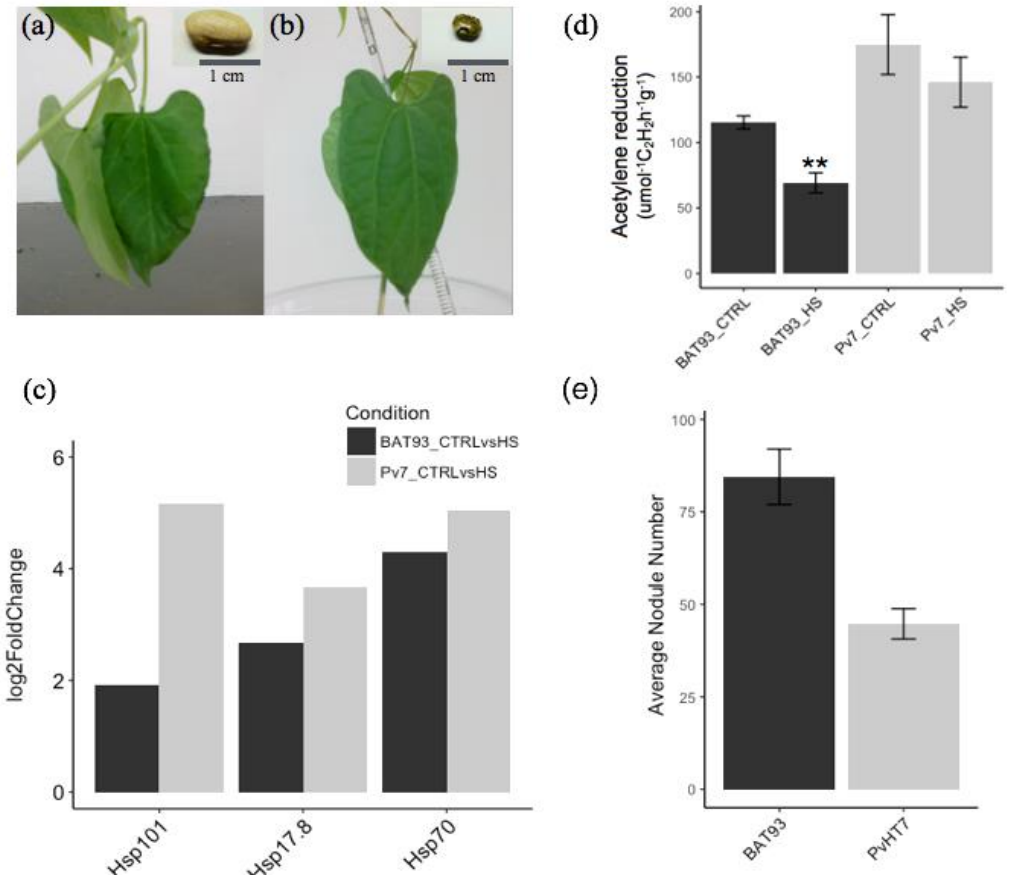

Fig 1. Heat-stress response in P. vulgaris cv. BAT93 and P. vulgaris 7 genotypes. (a) Foliar turgor changes observed in P. *vulgaris* BAT93 (a) and in P. *vulgaris* 7 plants after the heat-shock treatment (37ºC/6 h). Insets in (a) and (b) show seeds of the corresponding bean genotypes. Bar size, 1 cm. (c) Expression ratio of HSP101, HSP17.8 and HSP70 chaperones in root nodules of P. *vulgaris* BAT93 (in black) and P. *vulgaris* 7 (in grey), either in control conditions (CTRL) or subjected to heat stress (HS). The fold change in expression was obtained by DESeq of each heat-stress molecular marker from root nodules of control plants versus its expression in root nodules of heat-stressed plants. Values in both graphs represent the Log2 fold change of three biological replicates. (d) Effects of the thermal shock on the nitrogenase activity of P. *vulgaris* BAT93 or P. *vulgaris* 7 root nodules, either in control conditions or after the heat-shock treatment. ** P < 0.01, n=15. (e) Average nodule number of each genotype at 20 dpi. n=15

Table 2. Comparison of total number of open reading frames (ORFs) and small open reading frames (sORFs) in *Phaseolus vulgaris* cv. BAT93 and G19833, *Glycine max*, *Lotus japonicus* and *Medicago truncatula*.

	Phaseolus vulgaris BAT93	*Phaseolus vulgaris* G19833	*Glycine max*	*Lotus japonicus*	*Medicago truncatula*
ORFs	64692	31638	88647	10979	62319
sORFs	7414	4560	14979	2195	18688
sORFs/ORFs ratio	0.11	0.14	0.16	0.19	0.29

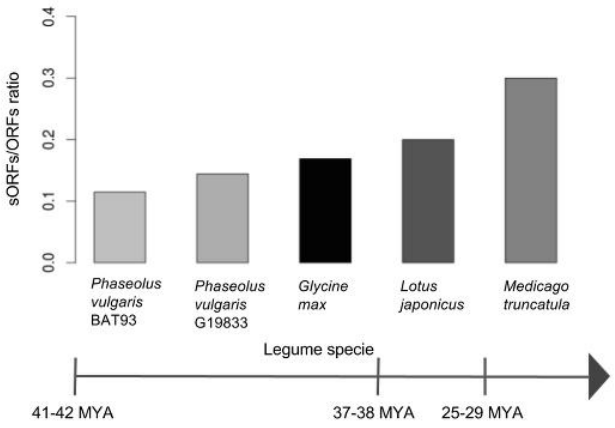

Figure 2

Fig 2. Proportion of sORFs detected in legume plant genomes. *P. vulgaris* G19833, *G. max*, *L. japonicus* and *M. truncatula* protein sizes in Phytozome version 1.11, and *P. vulgaris* BAT93 in *The Novo Genome Assembly and Annotation Team*. Arrow represents timeline evolution of these plant legumes based on archaeological and molecular data (Choi et al., 2004). Intersections indicate the time of divergence between clades. MYA, *million years ago*.

Table 3. List of differentially expressed sORFs in *P. vugaris* cv. BAT93 after heat-stress.

ID	Size	Associated protein	MOTIF MEME	BLASTP (Superfamilies)	Sequences producing significant alignments	Associated processes
PHASIBEAM10B038118 (T1)	103	N/A	N/A	No Putative Conserved Domains	Spidroin-1-like [Glycine max]	
PHASIBEAM10B045106 (T1)	103	N/A	N/A	H4 superfamily	Histone H3.2 [Cajanus cajan]	
PHASIBEAM10F001830 (T1)	103	N/A	N/A	No Putative Conserved Domain	Glutamine dumper 5-like [Cicer arietinum] transmembrane protein [Medicago truncatula]	
PHASIBEAM10F002901 (T1)	116	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
PHASIBEAM10F003368 (T1)	123	N/A	N/A	AAI_LTSS superfamily	Lipid transfer protein DIR1 [Vigna angularis, Medicago truncatula]	
PHASIBEAM10F004274 (T1)	140	N/A	N/A	No Putative Conserved Domain	Lysine-rich arabinogalactan protein 19-like [Vigna angularis] transmembrane protein, putative [Medicago truncatula]	
PHASIBEAM10F006225 (T1)	145	N/A	N/A	HMG-box superfamily	High mobility group B protein 7-like [Glycine soja] PREDICTED [Vigna radiata]/ HMGB-UBF_HMG-box, class II and III members of the HMG-box superfamily of DNA-binding proteins	
PHASIBEAM10F006374 (T1)	81	N/A	N/A	RRM_SF superfamily	Serine/arginine-rich SC35-like splicing factor SCL33 isoform X1 [Vigna radiata]	Hormonal control (Cruz et al., 2014, Suzuki et al., 2016)
PHASIBEAM10F007017 (T1)	137	Histone H2B.6	N/A	H2B superfamily	probable histone H2B.3 [Vigna radiata]	DNA package (Iliakis et al., 2008; Kim et al. 2015; Kantidze et al., 2016)

PHASIBEAM10F008866 (T3)	130	Thymidine kinase a	N/A	TK superfamily	Thymidine kinase-like [Vigna radiata]	Nucleotide synthesis (Wang & Liu, 2006; Garton et al., 2007); nucleotide salvage pathway (Moffat et al. 2002)
PHASIBEAM10F011464 (T1)	130	Histone H3.2	N/A	H4 superfamily	Histone H3.2 [Cajanus cajan] histone H3 [Triticum aestivus] core histone H2A/H2B/H3/H4	DNA package (Iliakis et al., 2008; Kim et al. 2015, Kantidze et al., 2016)
PHASIBEAM10F012744 (T1)(T2)	75(T1) 67(T2)	N/A	Motif B	COX7a_Cyt_c_Oxidase_VIIa superfamily	Cytochrome-c oxidases, electron carriers [Theobroma cacao]	Stress response (Gong et al. 1998, Huang et al. 2016))
PHASIBEAM10F019557 (T1)(T2)	88(T1) 93(T2)	N/A	Motif C/A	SANT_Superfamily/ Myb_DNA-Binding	PREDICTED: protein RADIALIS-like 3 [Vigna radiata] MYB transcription factor MYB142 [Glycine max]	
PHASIBEAM10F022486 (T1)	74	N/A	N/A	No Putative Conserved Domain	Transmembrane protein, putative [Medicago truncatula]	
PHASIBEAM10F025436 (T1)	72	N/A	N/A	No Putative Conserved Domain	Hypotetical protein LR48_Vigan08g167400 [Vigna angularis]	
PHASIBEAM10F026060 (T1)	144	N/A	N/A	Alpha-crystallin-HSPs_p23-like superfamily/IbpA	PREDICTED: 15.7 kDa heat shock protein, peroxisomal [Vigna angularis][Vigna radiata]	(Vierling et al. 1997)

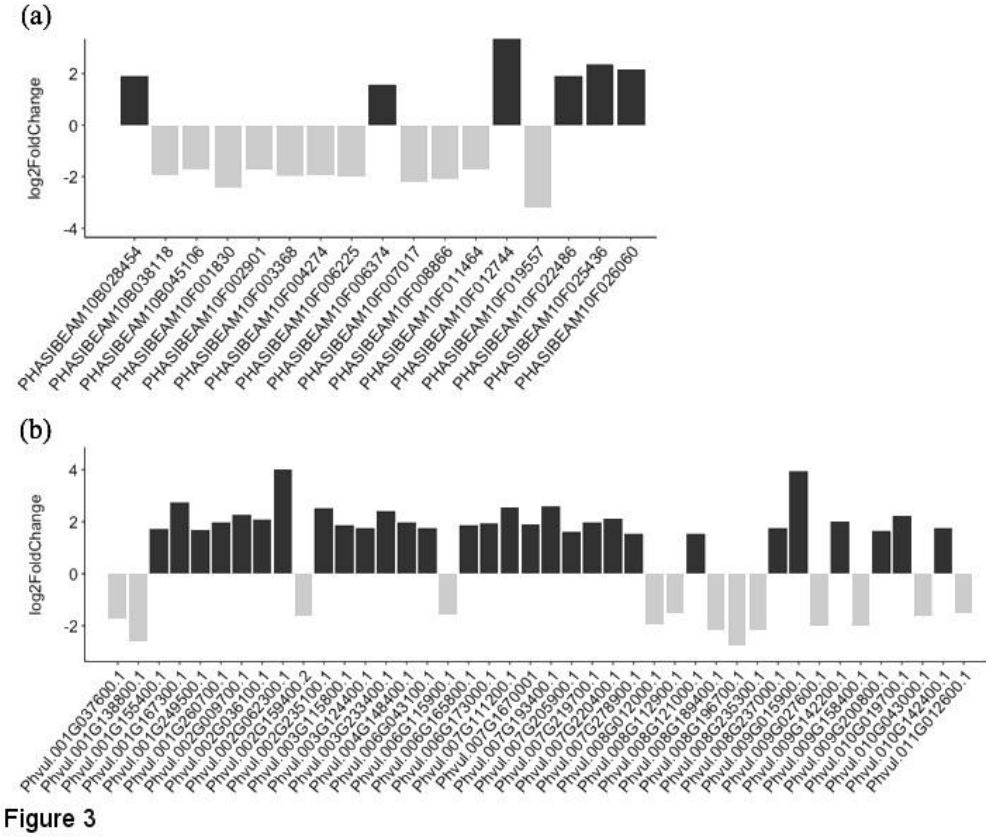

Figure 3

Fig 3. sORFs differentially expressed in (a) *P. vulgaris* BAT93 and (b) *P. vulgaris* 7. Gene expression values between control or heat-stressed 20 dpi nodules are expressed as the Log2 of the fold change. Accession numbers are indicated on the X axis.

Table 4. List of differentially expressed sORFs in *P. vugaris* 7 after heat-stress.

ID	Size	GO	MOTIF MEME	BLASTP (Superfamilies)	Sequences producing significant alignments	Associated process
Phvul.010G019700.1	112	Uncharacterised protein family SERF	N/A	4F5	Gibberellin regulated protein [Cynara cardunculus var. Scolymus]	Hormonal control
Phvul.010G142400.1	114	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.010G043000.1	97	Domain of unknown function (DUF581)	N/A	zf-FLZ superfamily	uncharacterized genes	
Phvul.003G124400.1	74	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.003G233400.1	75	phytosulfokine 4 precursor	N/A	PSK superfamily	phytosulfokines-like [Glycine max]	Growth
Phvul.003G115800.1	121	Ca2+-binding protein 1	Motif F/A	Efh superfamily (EF-hand7)	hypersensitivye reaction associated Ca2+-binding protein [Phaseolus vulgaris] calmodulin-like [Vigna angularis]	Calcium signaling (Liu et al., 2003; Al-Quaraan et al., 2010)
Phvul.009G200800.1	141	N/A	N/A	G_glu_transpept superfamily	transmembrane protein, putative [Medicago truncatula]	
Phvul.009G158400.1	58	N/A	N/A	No Putative Conserved Domain	aldo/keto reductase [Desulfitobacterium metallireducens]	
Phvul.009G015900.1	101	SAUR-like auxin-responsive protein family	N/A	Auxin_inducible superfamily	auxin-induced protein ARG7 [Cajanus cajan] Predicted: auxin-induced protein 15A [Vigna angularis]	Hormonal control
Phvul.009G142200.1	115	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.009G027600.1	150	Heavy metal transport/detoxification superfamily protein	N/A	HMA_superfamily	Predicted: heavy metal-associated isoprenylated plant protein 22 [Vigna angularis]	Stress response
Phvul.011G012600.1	86	Domain of unknown function, DUF642	N/A	PLN03089/hypotetical	uncharacterized genes	
Phvul.008G196700.1	44	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.008G121000.1	101	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.008G235300.1	97	Gibberellin-regulated family protein	N/A	GASA superfamily	gibberellic acid - stimulated protein 1 [Glycine soja]	Hormonal control
Phvul.008G112900.1	101	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	N/A	AAI_LTSS superfamily	predicted: putative ipid-transfer protein DIR1 [Vigna angularis]	Stress response
Phvul.008G189400.1	134	Heavy metal transport/detoxification superfamily protein	N/A	HMA_superfamily	Predicted: copper transport protein ATX1-like [Glycine max]	Stress response
Phvul.008G012000.1	137	Calcium-binding EF-hand family protein	N/A	EFh superfamily/EF-hand7	calcium-binding EF-hand protein [Medicago truncatula]	
Phvul.008G237000.1	144	HSP20-like chaperones superfamily protein	N/A	alpha-crystallin-HSPs_p23-like superfamily	Predicted: 15.7 kDa heat shock protein, peroxisomal [Vigna angularis]	
Phvul.004G148400.1	71	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	

Phvul.007G220400.1	54	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.007G205900.1	62	Low temperature and salt responsive protein family	N/A	No Putative Conserved Domain	Predicted: hydrophobic protein RCI2B [Vigna adiata]/Stress-induced hydrophobic peptide [Theobroma cacao]	
Phvul.007G111200.1	118	calmodulin-like 11	Motif D/F	EFh superfamily/EF-hand7	Predicted: calmodulin-like protein 11 [Vigna radiata]	Calcium signaling (Liu et al., 2003; Al-Quaraan et al., 2010)
Phvul.007G193400.1	147	Integrase-type DNA-binding superfamily protein	N/A	AP2 superfamily	Ethylene-responsive transcription factor ERF098 [Glycine soja]	Hormonal control
Phvul.007G167000.1	131	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.007G278900.1	150	calmodulin-like 11	Motif D/F/G	EFh superfamily	Predicted: calmodulin-like protein 8 [Vigna radiata var. radiata]	Calcium signaling (Liu et al., 2003; Al-Quaraan et al., 2010)
Phvul.007G219700.1	96	SAUR-like auxin-responsive protein family	N/A	Auxin_inducible superfamily	Predicted: auxin-induced protein X15-like [Glycine max]	
Phvul.001G037600.1	119	Domain of unknown function (DUF3511)	N/A	DUF3511 superfamily	uncharacterized genes	
Phvul.001G138800.1	67	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.001G167300.1	127	RmlC-like cupins superfamily protein	N/A	Cupin_3/Cupin_like superfamily	RmlC-like cupins superfamily protein	
Phvul.001G155400.1	148	calmodulin-like 11	Motif D/F/G	EFh superfamily	Predicted: calmodulin-3-like [Vigna radiata var. radiata]	Calcium signaling (Liu et al., 2003; Al-Quaraan et al., 2010)
Phvul.001G249500.1	67	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.001G260700.1	84	N/A	N/A	No Putative Conserved Domain	uncharacterized genes/ F-box domain, cyclin-like protein [Cynara cardunculus var. scolymus]	Cell proliferation
Phvul.006G173000.1	93	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.006G165800.1	127	jasmonate-zim-domain protein 8	N/A	tify_superfamily / CCT_2 superfamily	Predicted: protein TIFY 5A-like [Vigna radiata var. radiata]	
Phvul.006G043100.1	94	N/A	N/A	No Putative Conserved Domain	uncharacterized genes	
Phvul.006G115900.1	143	cytochrome B5 isoform E	N/A	Cyt-b5 superfamily	cytochrome b5-like [Vigna angularis]	
Phvul.002G062300.1	56	N/A	N/A	DUF4534 superfamily	uncharacterized genes	
Phvul.002G235100.1	73	N/A	N/A	DUF761 superfamily	uncharacterized genes	
Phvul.002G036100.1	113	cytochrome c-2	N/A	Cytochrom_C superfamily	Cytochrome c [Cajanus cajan][Medicago truncatula]	

Phvul.002G159400.1	99	SPIRAL1-like2	Motif E	No Putative Conserved Domain	Predicted: protein SPIRAL-like 5 [Vigna angularis]
Phvul.002G159400.2	99	SPIRAL1-like2	Motif E	No Putative Conserved Domain	Predicted: protein SPIRAL-like 5 [Vigna angularis]

The observed differences were accompanied by lower nitrogen-fixation levels (Figure 1d), supporting the hypothesis that reduced metabolic activity caused by heat stress reduces nitrogen fixation rates. Interestingly, compared to *P. vulgaris* BAT93, the average nodule number per root in plants of *P. vulgaris 7* was considerably lower (Figure 1e), although the level of nitrogen-fixation was higher (Figure 1d). This finding suggest that, compared to *P. vulgaris* BAT93, *P. vulgaris 7* root nodules are not only more resistant to heat stress but more efficient in fixing nitrogen.

To reveal the presence and quantity of any RNA in a biological sample by RNA-seq, statistical estimation of data is required. Three statistical methods were used to validate changes in gene expression, and only sORFs with a significant differential expression were considered. Compared to unstressed symbiotic nodules (control), 15 sORFs were differentially expressed in *P. vulgaris* BAT93 root nodules, whereas 44 sORFs were identified in *P. vulgaris* 7. Contrary to *P. vulgaris* 7, in *P. vulgaris* BAT93 most sORFs were down-regulated, with only a few being up-regulated (Figure 3 and Table 3). RNA-seq data on heat stressed nodules from both common bean genotypes suggest the involvement of phytohormones and antioxidant systems in the signaling for thermo-tolerance acquisition (Suzuki et al., 2016). However, the most remarkable difference at the molecular level observed among heat-stressed nodules of these genotypes was the notable abundance of sORFs transcripts related to calcium signaling in *P. vulgaris* 7 (Table 4). This finding suggests that, regardless of the evolutionary closeness of the domesticated *P. vulgaris* BAT93 and the wild *P. vulgaris* 7, each genotype has independently adapted their molecular responses to preserve the biological nitrogen fixation process under heat stress (Figure 3, and Tables 3 and 4). This ability becomes highly relevant in nitrogen deprived soils, such as those of tropical and temperate regions. In this sense, *P. vulgaris* 7 as well as other *P. vulgaris* wild relatives of Mexico are important reservoirs of genetic variation that could be sourced for crop improvement.

To our knowledge, this is the first report of a set of sORFs being associated with heat stress. Taking into consideration the highest resistance to heat stress shown by *P. vulgaris* 7

(Figure 1), induced sORFs under heat should be subject of further functional genomics studies. Although these studies are necessary to prove the biological function of each of these sORFs, the described procedure opens new possibilities to detect potentially relevant genes involved in heat stress response.

Materials and Methods

Plant growth and heat-stress treatments

Dry, mature seeds of *Phaseolus vulgaris* cv. BAT93 and a *Phaseolus vulgaris* wild heat-tolerant genotype (named *P. vulgaris* 7) were surface sterilized as previously described (Estrada-Navarrete et al., 2007). Sterilized seeds were transferred to sterile trays containing wet paper towels. Trays were covered with foil and incubated at 28°C for 2 days (Estrada-Navarrete et al., 2007). Two-day-old common bean sprouts were inoculated with *Rhizobium tropici* CIAT899 (Martínez-Romero et al., 1991) and grown at 28ºC/18ºC day/night temperature, 65% relative humidity, 180-300 µ-1mol photon m-2s-1 and 14 h photoperiod 20 days in a growth chamber. Common bean plants were ered every third day with N-free sterile B&D nutrient solution (Broughton & Dilworth 1971). After this period, plants were subjected to a sudden heat-stress (37ºC), sustained for 6 h. Twenty days post-inoculation (dpi), root-nodules from 5 plants of each genotype were harvested, frozen in liquid nitrogen, and stored at -80 ºC.

Nitrogenase activity was evaluated from 20 dpi inoculated roots (following methods of Ramírez et al., 1999; Verdoy et al., 2004) under control and heat stress conditions in both bean genotypes. Nodulated roots were incubated in acetylene gas for 1 h and ethylene production was determined by gas chromatography (Varian model 3300). Specific activity was expressed as µmol $^{-1}C_2H_2h^{-1}g^{-1}$ nodule dry weight.

Bacterial strain and culture

The *Rhizobium tropici* CIAT899 strain was selected as it has known resistance to heat (37ºC; Martínez-Romero et al., 1991). Two-day-old bean sprouts were inoculated with *R. tropici* CIAT899 according to Ramírez et al. (2005) with some minor modifications. Briefly, *R. tropici* CIAT899 was grown in peptone yeast liquid medium [0.5% bactopeptone (w/v), 0.3% yeast extract (w/v), 7 mM $CaCl_2 \cdot 2H_2O$] supplemented with 20 g/mL nalidixic acid at 30 ºC to a cell density of 5 to 8×10^8 mL^{-1}. 1 mL was applied to the root.

RNA extraction, cDNA libraries preparation and sequencing using Illumina Hiseq2000

Twenty dpi symbiotic nodules were isolated, frozen in liquid nitrogen and ground to a fine powder with a mortar and pestle. The sample was immediately processed for total RNA isolation using the extraction kit ZR Plant RNA MiniPrep (Zymo Research, USA) according to manufacturer's instructions. Total RNA in each sample was more than 5 µg. RNA integrity was confirmed using a 2100 Bioanalyzer (Agilent Technologies, Inc.) with a minimum RNA integrity number (RIN) value of 7.0. cDNA library templates from 3 biological replicates of each genotype, and from both control and heat-stress conditions (24 cDNA libraries in total), were prepared using a Truseq™ RNA Sample Prep Kit (Illumina) according to the manufacturer's recommendations at the University Unit for Massive Sequencing (UUSM) from the Universidad Nacional Autónoma de México (UNAM). These libraries were sent to Macrogen Inc. (Korea; www.macrogen.com) for sequencing by Illumina Hiseq2000 (http://www.illumina.com).

Strategy for large-scale discovery of putative sORFs in P. vulgaris BAT93 and G19833 genotypes

sORFs (30 to 150 aa in length) of *P. vulgaris* BAT93 were gathered from CoGe and *The Novo Genome Assembly and Annotation Team* (CoGe database [https://genomevolution.org/CoGe/] and [http://denovo.cnag.cat/genomes/bean/], genome ID 20365) while the sORFs from *P. vulgaris* G19833 were collected from Phytozome (Phytozome version 11 database [www.phytozome.net]; *P. vulgaris* v1.0), respectively. In both cases, all sORFs with no initial methionine were discarded to avoid truncated transcripts.

Gene expression and motif-based analysis of P. vulgaris sORFs

In order to estimate transcript abundance for each experimental condition tested, raw sequence data from Illumina Hiseq2000 were analyzed using FASTQC software (www.bioinformatics.babraham.ac.uk/projects/). The short sequence reads obtained (of around 100 bp in length) were aligned to the reference genome; *P. vulgaris* BAT93 (CoGe database [https://genomevolution.org/CoGe/], genome ID 20365) or *P. vulgaris* G19833 (*P. vulgaris v1.0;* Phytozome version 11 database [www.phytozome.net]) to uncover their identity. The SMALT software (http://www.sanger.ac.uk/science/tools/smalt-0) was used to this purpose. Finally, differential expression was estimated with DESeq, an R/Bioconductor package performing a pairwise differential expression analysis (Anders & Huber, 2010, Bioconductor V3.3, R V 3.3, http://bioconductor.org/packages/2.11/bioc/, Robinson 2010). Only *P. vulgaris* sORFs validated by this method with a 2-fold change and a *P*-value < 0.05 between control and the heat stress condition of each common bean genotype were considered for study. Selected *P. vulgaris* sORFs were classified according the GO annotation (http://www.agbase.msstate.edu/cgi-bin/tools/GOanna.cgi, McCarthy et al., 2006), and further analyzed by the MEME Suite (http://meme-suite.org/tools/meme, Bailey et al., 2009), and by the BLASTP algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi, Altschul et al., 1990).

Acknowledgments

We acknowledge Alfonso Leija (CCG, UNAM) for technical support with the acetylene reduction analysis. We are very grateful to Jorge Acosta, INIFAP - CE Bajío and Alfonso Delgado, Instituto de Biología, UNAM for supplying *Phaseolus vulgaris* seeds. We are thankful to the University Unit for Massive Sequencing (UUSM) from the National University of Mexico (UNAM) for Illumina sample sequencing and data analyisis.
This work was funded by FAO/IAEA CRP 23029 Research Contract 16601, PAPIIT IN206415 and IN206815 from Universidad Nacional Autónoma de México and CONACYT No. 177744 and 177207. Eunice Alejandra Zayas-Del Moral is a doctoral student from Programa de Doctorado en Ciencias Biomédicas (UNAM) and received fellowship No. 317533 from CONACYT.

References

Abd-Alla MH, Issa AA, Ohyama T (2014) Impact of harsh environmental conditions on nodule formation and dinitrogen fixation of legumes. Advances in biology and ecology of nitrogen fixation. ISBN *978-953-51-1216-7*, P.978-953.

Al-Quraan NA, Locy RD, Singh NK (2010) Expression of calmodulin genes in wild type and calmodulin mutants of *Arabidopsis thaliana* under heat stress. Plant Physiol Biochem. 48 (8): 697-702.

Altschul, SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215: 403-410.

Anders S, and Huber W (2010) Differential expression analysis for sequence count data. Genome Biol. 11: R106.

Aparicio F, Thomas CL, Lederer C, Niu Y, Wang D, and Maule AJ (2005) Virus induction of Heat Shock Protein 70 reflects a general response to protein accumulation in the plant cytosol. Plant Physiol. 138: 529-536.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37 (Web Server Issue): W202-W208.

Batut J, Mergaert P, and Masson-Boivin C (2011) Peptide signalling in the rhizobium-legume symbiosis. Curr Opin Microbiol. 14(2): 181-187.

Broughton WJ, and Dilworth MJ (1971) Control of leghaemoglobin synthesis in snake beans. Biochem J. 125: 1075-1080.

Denarié J, Debellé F, and Promé JC. (1996) Rhizobium lipochitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Annu Rev Biochem. 65: 503-35.

Estrada-Navarrete G, Alvarado-Affantranger X, Olivares JE, Guillén G, Díaz-Camino C, Campos F, Quinto C, Gresshoff PM, and Sanchez F (2007) Fast, efficient and reproducible genetic transformation of *Phaseolus spp*. by *Agrobacterium rhizogenes*. Nature Prot. 2: 1819-1824.

Garton S, Knight H, Warren GJ, Knight MR, and Thorlby GJ (2007) Crinkled leaves 8-a mutation in the large subunit of ribonucleotide reductase–leads to defects in leaf development and chloroplast division in Arabidopsis thaliana. Plant Journal 50: 118–127.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40: D1178-D1186.

Guillén, G., Díaz-Camino, C., Loyola-Torres, C.A., Aparicio-Fabre, R., Hernández-López, A., Díaz-Sánchez, M., and Sanchez, F. (2013) Detailed analysis of putative genes encoding small proteins in legume genomes. Front. Plant Sci. 4: 208.

Hanada, K., Akiyama, K., Sakurai, T., Toyoda, T., Shinozaki, K., and Shiu, S.H. (2010) sORF finder: a program package to identify small open reading frames with high coding potential. Bioinformatics. 26(3): 399-400.

Hanada, K., Higuchi-Takeuchi, M., Okamoto, M., Yoshizumi, T., Shimizu, M., Nakaminami, K., Ohashi, C., Ilda, K., Tanaka, M., Horii, Y., et al. (2012) Small open reading frames associated with morphogenesis are hidden in plant genomes. Proc Acad Nat Sci. 110(6): 2395-2400.

Huang YC, Niu CY, Yang CR, Jinn TL (2016) The heat-stress factor HSFA6b connects ABA signaling and ABA-mediated heat responses. Plant Physiology Preview. 172(2):1182-1199

Iliakis G, Wu W, and Wang M (2008) DNA double strand break repair inhibition as a cause of heat radiosensitization: re-evaluation considering backup pathways of NHEJ. Int J Hyperthermia. 24 (1): 17-29.

IPCC (2014) Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field CB, Barros VR, Dokken DJ, Mach KJ, Mastrandrea MD, Bilir TE, Chatterjee M, Ebi KL, Estrada YO, Genova RC, Girma B, Kissel ES, Levy AN, MacCracken S, Mastrandrea PR, and White LL (Eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 1132.

Kantidze OL, Velichko AK, Luzhin AV, and Razin SV (2016) Heat stress-induced DNA damage. Acta Naturae. 8(2):75-78.

Kim DH, Xu ZY, Na YJ, Yoo YJ, Lee J, Sohn EJ, and Hwang I (2011) Small heat shock protein HSP17.8 functions as an AKR2A cofactor in the targeting of chloroplast outer membrane proteins in Arabidopsis. Plant Physiol. 157(1): 132-146.

Kim JM, Sasaki T, Ueda M, Sako K, and Seki M (2015) Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. Front Plant Sci. 6: 114.

Larkindale J, Hall JD, Knight MR, and Vierling E. (2005) Heat stress phenotypes of Arabidopsis mutants implicate multiple signaling pathways in the acquisition of thermotolerance. Plant Physiol. 138: 882-897.

Lebrazi S & Fikri Benbrahim K (2014) Environmental stress conditions affecting the N2 fixing *Rhizobium*-legume symbiosis and adaptation mechanisms. African Journal of Microbiology Research. 8(53):4053-4061

Liu HT, Li B, Shang ZL, Li XZ, Mu RL, Sun DY, and Zhou RG (2003) Calmodulin Is Involved in Heat Shock Signal Transduction in Wheat. Plant Physiology, 132:1186–1195.

Marmiroli N, and Maestri E (2014) Plant peptides in defense and signaling. Peptides 56: 30-44.

Martínez-Romero E, Segovia L, Martins MF, Franco AA, Graham P, and Pardo MA (1991) *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L. beans and *Leucaena sp.* trees. Int J Syst Bacteriol. 41(3): 417-426.

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP et al. (2006) AgBase: a functional genomics resource for agriculture. BMC Genomics. 7: 229.

Moffatt BA, Stevens YY, Allen MS, Snider JD, Pereira LA, Todorova MI, Summers PS, Weretilnyk EA, Martin-McCaffrey L, Wagner C (2002) Adenosine kinase deficiency is associated with developmental abnormalities and reduced transmethylation. Plant Physiology 128: 812–821

Peoples MB, Herridge DF, and Ladha JK. (1995) Biological nitrogen fixation: an efficient source of nitrogen for sustainable agricultural production. Plant Soil. 174: 3-28.

Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, and Lara M (2005) Sequencing and analysis of common bean ESTs: Building a foundation for functional genomics. Plant Physiol. 137: 1211-1227.

Ramírez M, Valderrama B, Arredondo-Peter R, Soberón M, Mora J, and Hernández G (1999) Rhizobium etli genetically engineered for the heterologous expression of Vitreoscilla sp. Hemoglobin: effects on free-living and symbiosis. MPMI. 12(11):1008-1015

Robinson MD, McCarthy DJ, and Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. 26(1): 139-140.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K et al. (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res. 15: 227-239.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, David L, Hyten L, Song Q, Jay J et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature. 463: 178-183.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, et al. (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genetics. 46:707-713

Skarshewski A, Stanton-Cook M, Huber T, Al Mansoori S, Smith R, and Beatson SA (2014) uPEPperoni: an online tool for upstream open reading frame location and analysis of transcript conservation. BMC Bioinformatics. p. 15-36

Suzuki N, Bassil E, Hamilton JS, Inupakutika MA, Zandalinas SI, Tripathy D, Luo Y, Dion E, Fukui G, Kumazaki A, Nakano R, Rivero RM, Verebeck GF, Azad RK, Blumwald E, Mittler R (2016) ABA is required for plant acclimation to a combination of salt and heat stress. PLoS One. e0147625.

Tate RL (1995) Soil microbiology (symbiotic nitrogen fixation). New York, N.Y. John Wiley & Sons, Inc. p. 307-333.

Vanderperre B, Lucier JF, Roucou X (2012) HAltORF: A database of predicted out-of-frame alternative open reading frames in human. Database (Oxford). Bas025.

Verdoy D, Lucas MM, Manrique E, Covarrubias AA, De Felipe MR, Pueyo JJ (2004) Differential organ-specific response to salt stress and water deficit in nodulated bean (*Phaseolus vulgaris*). Plant, Cell & Environment. 27(6):757-767

Vierling E. (1997) The small heat shock proteins in plants are members of an ancient family of heat induced proteins. Acta Physiol Plant. 19:539-547.

Vlasova, A., Capella-Gutiérrez, S., Rendón-Anaya, M., Hernández-Oñate, M., Minoche AE, Erb J, Câmara F, Prieto-Barja P, Corvelo A, Sanseverino W, et al. (2016) Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. Genome Biol. 17: 32.

Voysest O (1983) Variedades de frijol en América Latina y su origen. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. p. 87 (20823).

Voysest O (2000) Mejoramiento genético del frijol (*Phaseolus vulgaris* L.): Legado de variedades de América Latina 1930-1999. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. p. 195 (Publicación CIAT no. 321) (SB 327.V69).

Wang W, Vinocur B, Shoseyov O, and Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci. 9: 244-252.

Wang C, Liu Z (2006) Arabidopsis ribonucleotide reductases are critical for cell cycle progression, DNA damage repair, and plant development. Plant Cell. 18: 350–365.

Wang F, Dong Q, Jiang H, Zhu S, Chen B, Xiang Y (2012) Genome-wide analysis of the heat shock transcription factors in Populus trichocarpa and Medicago truncatula. Mol Biol Rep. 39: 1877-1886.

Yang H, Matsubayashi Y, Nakamura K, Sakagami Y (2001) Diversity of Arabidopsis Genes Encoding Precursors for Phytosulfokine, a Peptide Growth Factor1. Plant Physiology. 127(3):842-851.

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, et al. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature. 480: 520-524.

Supplementary information

Fig S1. Motifs identified by MEME in the differentially expressed sORFs in cv. BAT93 (a) and *Phaseolus vulgaris* 7 (b) heat stress transcriptomes.