The Molecular Evolution of the Small Heat-Shock Proteins in Plants

Elizabeth R. Waters

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721 Manuscript received April 11, 1995 Accepted for publication June 28, 1995

ABSTRACT

The small heat-shock proteins have undergone a tremendous diversification in plants; whereas only a single small heat-shock protein is found in fungi and many animals, over 20 different small heat-shock proteins are found in higher plants. The small heat-shock proteins in plants have diversified in both sequence and cellular localization and are encoded by at least five gene families. In this study, 44 small heat-shock protein DNA and amino acid sequences were examined, using both phylogenetic analysis and analysis of nucleotide substitution patterns to elucidate the evolutionary history of the small heatshock proteins. The phylogenetic relationships of the small heat-shock proteins, estimated using parsimony and distance methods, reveal that gene duplication, sequence divergence and gene conversion have all played a role in the evolution of the small heat-shock proteins. Analysis of nonsynonymous substitutions and conservative and radical replacement substitutions (in relation to hydrophobicity) indicates that the small heat-shock protein gene families are evolving at different rates. This suggests that the small heat-shock proteins may have diversified in function as well as in sequence and cellular localization.

"HE small heat-shock proteins are those proteins L produced in response to high temperature stress that are smaller than 30 kDa in size. Higher plants have at least 20 and some plant species may have as many as 40 different small heat-shock proteins (VIERLING 1991). In contrast, most other organisms have one or only a few small heat-shock proteins. Saccharomyces cerevisiae has one small heat-shock protein and Drosophila has four (ARRIGO and LANDRY 1994). The diversification of the plant small heat-shock proteins occurred after the split of the plant and animal lineages. This suggests that the tremendous diversification of small heat-shock proteins in plants may reflect adaptations to stresses unique to plants. The small heat-shock protein genes in plants comprise a large multigene family composed of at least five distinct gene families; all are nuclear encoded. The plant small heat-shock proteins have previously been divided into four classes based on sequence similarity and cellular localization (VIERLING 1991). One class of proteins localizes to the chloroplast (CP), one to the endoplasmic reticulum (ER), and two to the cytosol, classes I and II. Recently a fifth class of mitochondrial (MT)-localized proteins has been reported (LENNE and DOUCE 1994). The diversification of cellular localization of small heat-shock proteins is unique to plants; all of the nonplant small heat-shock proteins localize to the cytosol (ARRIGO and LANDRY 1994).

The plant small heat-shock proteins are related to the small heat-shock proteins in other organisms and to the vertebrate alpha-crystallin proteins (PLESOFSKY- VIG et al. 1992; JONG et al. 1993). All share a conserved heat-shock region in the carboxyl terminal domain. Comparisons of the amino acid sequences of the carboxyl terminal domain of some plant small heat-shock proteins and other small heat-shock proteins confirms that the plant proteins are related to but quite distinct from other small heat-shock proteins (PLESOFSKY-VIG et al. 1992; JONG et al. 1993). PLESOFSKY-VIG et al. (1992) concluded, based on branch lengths and tree topology, that the plant small heat-shock proteins have evolved more slowly than the animal small heat-shock proteins. They also concluded that the CP-localized protein originated from the chloroplast endosymbiotic event and is thus only distantly related to the other small heat-shock proteins (PLESOFSKY-VIG et al. 1992).

The *in vivo* function of the small heat-shock proteins is not known. Recent in vitro studies suggest that the small heat-shock proteins, like the large HSPs, may be molecular chaperones (JAKOB et al. 1993; MERCK et al. 1993; JAKOB and BUCHNER 1994; LEE et al. 1995). The biochemistry of the large heat-shock proteins (HSPs 70, 90 and 60) has been well studied (BECKMANN et al. 1990; GETHING and SAMBROOK 1992; BECKER and CRAIG 1994; CRAIG et al. 1994; SCHNEIDER et al. 1994). The evolution of HSP 70s has also been studied in some detail (BOORSTEIN et al. 1994; RENSING and MAIER 1994). These studies reveal that, in contrast to the small heatshock proteins, the genes coding for the HSP 70 proteins duplicated very early in the evolution of eukaryotes. The selective constraints on the large HSPs and the small HSPs are very different. Amino acid sequences of HSP 70 are highly conserved; there is almost 50% amino acid identity from Zea mays to Escherichia coli

Author e-mail: ewaters@ccit.arizona.edu

E. R. Waters

TABLE 1

Gene and protein accession numbers

Species	Protein	DNA accession number	Protein accession number
Chloroplast-localized proteins	······································	·····	
Arabidopsis thaliana	HSP 21	X54102	P31170
Glycine max	HSP 22	X07188	P09887
Petunia hybrida	HSP 21	X54103	P30222
Pisum sativium	HSP 21	X07187	P09886
Triticum aestivum	HSP 26A	X58280	000445
Triticum aestivum	HSP 26B	X67328	\$96581
Zea mays	HSP 26	L28712	520001
Mitochondrial-localized protein			
Chenopodium rubrum	HSP 23	X15333	
Endoplasmic reticulum-localized proteins		Ribbbb	
Arabidopsis thaliana	HSP 99	1111501	
Glycine max	HSP 22	X63198	P30936
Pisum sativum	HSP 22	M33898	1 30230
Class L cytocolically localized proteins	1101 22	M133030	
Arabidobsis thaliana	HSP 176	X16076	D19859
Arabidopsis thaliana	HSP 17.0	X10070 X17903	P10036
Arbidopsis thaliana	HSP 18 9	X17255 X17905	P10207
Chenopodium rubrum	HSP 18 3	X17233 X53870	F 10307 S90808
Daucus carota	HSP 18.0	X53870 X53859	520003 D97907
Daucus carota	HSP 17.8	X53851	1 47397 D97906
Chuine max	HSP 17.5	M11218	F27390 D04709
Clycine max	HSP 17.5	M11917	P04795 D04705
Chrine max	LISE 19 5	W11517 V07160	P04793
Helianthus annuus	HSP 17.6	X67100 X60701	P03478
Incohomicos con acculanteum	113F 17.0	X59701 VEC199	P30093
Lycopericoscon escutentum Medicago sativa	HSD 19 1	A30138 VE9710	P30221
Medicago sativa	HOF 10.1	X58710 X58711	P27879
Meanago sativa	H5P 16.2	X38/11 X60990	P27880
Oryza sativa	HSP 10.9	X00820	P2////
Discum antimum	HSP 17.4	D12035	P31673
Tisum sauvum	HSP 18.1	M33899	P19243
Truicum destivum	HSP 10.9A	X13431	P12810
Triticum aestivum	HSP 16.9B	X64618	\$21600
I riticum aestivum	HSP 16.9C	L14444	
Clea mays	HSP 17.2	X65725	
Class II cytocolically localized proteins	110D 15 C	X100440	Pagaga
Arabidopsis thaliana	HSP 17.6	X63443	P29830
Glycine max	HSP 17.9	X07159	P05477
Ipomea nil (Pharbatis nil)	HSP 18.8	M99430	QO1545
Ipomea nil (Pharbatis nil)	HSP 17.2	M99429	QO1544
Lilium longiflorum	HSP 18.2	BOUCHARD (1990)	
Lilium longiflorum	HSP 17.6	D21816	
Lilium longiflorum	HSP 16.5	D21818	
Pisum sativum	HSP 17.7	M33901	S12720
Triticum aestivum	HSP 17.3	X58279	S16525
Zea mays	HSP 17.5	X54076	P24631
Zea mays	HSP 17.8	X54075	P24632

(LINDQUIST and CRAIG 1988). The small heat-shock proteins evolve much more quickly; there is <40% amino acid identity between the small heat-shock protein in *S. cerevisiae* and the plant small heat-shock proteins. The different evolutionary histories of the large and small HSPs suggest that, even if both types of HSPs are molecular chaperones, the specific functions within the cell and the selective constraints on these groups of proteins are very different. Patterns of DNA sequence divergence can be very useful indicators of differences in selective constraint and possible functional divergence (HUGHES *et al.* 1990; HUGHES 1993a,b; KARLIN *et al.* 1992). In a study of the HSP 70 genes, HUGHES demonstrated that rates of nucleotide substitutions reflect the known functional differences among the HSP 70s (HUGHES 1993b). In this study of small heat-shock proteins, I examined both the complete DNA and amino acid sequences of 44 plant

Plant Small Heat-Shock Proteins

	10	20	30	40	50	60	70	80	90	100
T.aestivum 26a T.aestivum 26b Z.mays 26 P.sativum 21 G.max 21 A.thaliana 21 P.bybrida 21	MAAAN MAAAN MAAAP MAQSVSLSTIASP MASTLSFAASA MA.CKTLTCSASE	APFALVSRLSF APFAL.SRLSF FAIAGRLSPVA FILSQKPGS LLCSPLAPSF VLVSNGVVSATS	PAARL PIRAWF PAARL PFRAWF RL PVRA SVKSTPPCMA SVSSKSA SVSSKSA	RAARPAPLST RAARPAPVWT .WRPAHGFAS ASFPLRRQLPI FPFSVSI	GGRTRPLSVAS GRTRPLSVAS SS.GRARSLAVAS RLGLRNV FPRKIPSRI KCSVRKPASRLVA	AAQ ENRDN AAQ ENRDN AAQ ENRDN RAQ AGGDG G GDNKD RAQ DQREN OAT GDNKD	SVDVQ.V SVDVQ.V SVDVQ.V DNKDNSV NSVEVQH SIDVV TSVDVHV	SQAQNAGN.Q SQAQNAGN.Q SQNGGNRQ EVHRVNKDD. VSKGD .QQGQQKGN SNNNQGGNNQ	QGNAVQRRPRRA QGNAVQRRPRRA QGNAVQRRPRRA QGTAVERKPRRS QGTAVEKKPRRT QGSSVEKRPQQR QSAVE.RRPRRM	.GFDISP .GFDISP TALDISP .SIDISP .AMDISP LTMDVSP .ALDVSP
C rubrum 23	MA SMALRRLASE	NLVSGGTFR.		PLSVSRSFN	ΓN Α	OMG RVDHD	HELDDRS	NRAPISRRG.	DFPAS	FFSDVFD
L longiflorum 18.2									MGSKL	TREEYNT
L longiflorum 17.6									MGSKL	TREEYDT
L. longiflorum 16.5									MDSKF	EVDHSLI
Z.mavs 17.8									MDAVM	FGLET
Z.mays 17.5									MDGRM	FGLET
T.aestivum 17.3										FGLDA
P.sativum 17.7								• • • • • • • • • • • •	MDFRL	MDLDS
G.max 17.9							• • • • • • •		MDFRV	MGLES
I.nil 17.2									MDLRL	MGFDH
I.nil 18.8						• • • • • • • • • • • • • • • • • • • •	• • • • • • • •		MDLRN	FGLSNFG
A.thaliana 17.6II				· · · · · · · · · · · ·		••••	• • • • • • • •		MDLGR	F
D.carrota 18.0	•••••	• • • • • • • • • • •	•••••			••••	• • • • • • •	• • • • • • • • • • •	MS11P	SFFGS
D.carrota 17.8		• • • • • • • • • • •		••••••	• • • • • • • • • • • • •	••••	• • • • • • •	• • • • • • • • • • •	MS11P	SFFG.
M.sativus 18.1	• • • • • • • • • • • • • •	•••••		••••••		••••	• • • • • • • •	• • • • • • • • • • •		
M.sativus 18.2		• • • • • • • • • • • •		•••••		••••	• • • • • • • •		MSLIP	SFFG.
P.sativum 18.1	••••••	• • • • • • • • • • •	•••••		• • • • • • • • • • • • • •	•••	• • • • • • • •	• • • • • • • • • •	MOUT	SFFS.
G.max 17.5		• • • • • • • • • • • •	•••••	• • • • • • • • • • •	• • • • • • • • • • • • • •	••••	• • • • • • • •		MSD12	SIFG.
G.max 17.3		• • • • • • • • • • • • •		• • • • • • • • • • •		••••	•••••	• • • • • • • • • • •	MOLIP	N PPC
G.max 18.5	• • • • • • • • • • • • • •	• • • • • • • • • • • •	•••••	•••••	• • • • • • • • • • • • • • •	••••	• • • • • • • •		MOLTO	a TEC
G.Max 1/.6		• • • • • • • • • • • • •		•••••	• • • • • • • • • • • • • • •		• • • • • • • •	• • • • • • • • • • •	Met Th	D TEC
L.esculentum 17.8	•••••	•••••	••••••			••••			MSLIP	e TEC
A.Chailana 17.6		•••••		••••••		••••			MSLVD	S FFG
A. (nallana 1/.4		•••••			• • • • • • • • • • • • • •	••••			MGLTD	S IFC
A. (naliana 18.2		•••••					••••••		MGITD	S FFT
A antinum 17 9		••••					•••••		TTPRV	FGT
T section 16 9b		•••••							MSTV.	
T activum 16.90										
T aestivum 16.9a									MSIV.	
O sativa 16 9									MSLV.	
7 maye 17 2									MSLV.	
O cativa 17 4										
C. rubrum 18.3									MSLIP	NNWFNT.
P satism 22								MSLKPLNMLL	VPFLLLTLAADF	PLKAKGS
G may 22								MRLOOLNUF	FLLLCVA	. KANGS
A thaliana 22								MM KHLL	SIFFIGALLIGN	IKTSEGS
A. CHAITANA 22									221120100000	

FIGURE 1.—Amino acid alignment. Boxes mark conserved regions. #, highly conserved residue; *, completely conserved residue.

small heat-shock proteins. Using both distance- and parsimony-based phylogenetic methods, I constructed gene trees to determine the evolutionary relationships among and within the plant small heat-shock protein gene families. In addition I examined the rates of nucleotide substitutions among the plant small heat-shock proteins. I have found evidence of differences in selective constraint among the small heat-shock proteins suggesting that functional differences may also exist among the plant small heat-shock proteins.

MATERIALS AND METHODS

Sequence alignment: DNA and amino acid sequences of 44 small heat-shock proteins were obtained from the databases or the literature. Accession numbers or references are listed in Table 1. When amino acid sequences were not available, DNA sequences were translated using Translate in GCG (Genetics Computer Group 1991). The size of the HSPs (in kDa) were either taken from the literature or determined using the program PeptideSort in GCG. Amino acid sequences were aligned using PileUp in GCG. The alignment was further refined by hand in LineUp in GCG (Figure 1). The aligned protein sequences were imported into the program DNA Stacks (EERNISSE 1992). The unaligned coding regions of the DNA sequences were also imported. The DNA sequences were aligned by imposing the gaps in the amino acid alignment upon the DNA sequences (DNA alignment is available upon request from the author). Pairwise comparisons of overall sequence similarity were done using the program Gap in GCG.

Phylogenetic analysis: Phylogenetic analysis of the aligned DNA and amino acid sequences were conducted using parsimony in PAUP (SwoFFORD 1993) version 3.1.1 and distance (DNAdist, Protdist and NeighborJoining) in PHYLIP (FELSEN-STEIN 1993) version 3.5c. PHYLIP is available by anonymous FTP at "evolution.genetics.washington.edu."

The parsimony analyses were conducted as follows: heuristic searches with 100 random addition replicates, with MUL-PARS and TBR branch swapping (steepest descent was not invoked), were conducted to find the most parsimonious trees. All trees were found in the first or second replicate, no additional trees were found in the next 98 replicates. The strict consensus of the most parsimonious trees was constructed. Support for branches was evaluated by bootstrap analysis: 100 Bootstrap replicates with the same conditions as above were conducted.

The tree presented in this paper is arbitrarily rooted with the sequences for the chloroplast proteins. At the present time it is also not possible to unequivocally choose a root for the small heat-shock proteins. Additions of other eukaryotic small heat-shock proteins (from yeast and humans) to the data matrix make alignment more difficult and, in addition, do not resolve the relationships among the plant small heatshock protein gene families.

The analysis of the DNA sequences were first conducted with the complete sequences and then with the transit peptides and the third positions removed. Transit sequences were

E. R. Waters

<pre>7. destivum 26aFeU DIPSIMETRIGNEDM DELP DAVGEPTRESPA BAR. BENNEDLENDERVINGE DMOLESEVRINGEDALIDGENKELBAGEG 7. aestivum 21FeU DIPSIMETRIGNEDMIN DELP DAVGEP. TRESPA BAR. BENNEDLENDERVINGE DMOLARDEVRINGEDMINGEDERKBAGEG 7. aestivum 21FeU DIPSIMETRIGNEDMIN DELP DAVGEP. TRESPA BART. SUPPEIR KEDERVINGE DMOLARDEVRINGEDTU INGENKELBAGEG 7. aestivum 21FeU DIPSIMETRIGNEDMIN DELP DAVGEP. BINIGGE I RIVNEI KEDERKINGE DMOLARDEVRINGEDTU INGENKESFEN 7. abstivum 21FeU DIPSIMETRIGNEDMIN DEVE FORMTFIG.BNIGGE I RIVNEI KEDERKINGE DMOLARDEVRINGEDKUN INGENKESFEN 7. abstivum 21FEU DIPSIMETRIGNEDMIN DEVE FORMTFIG.BNIGGE I RIVNEI KEDERKINGE DMOLARDEVRINGEDKUN INGERKESFEN 7. abstivum 21FEU DIPSIMETRIGNEDMIN DEVE FORMTFIG.BNIGGE I RIVNEI KEDERKINGE DMOLARDEVRINGEDKUN INGERKESFEN 7. abstivum 12FEU DARSPHRITHKOMUTT DRELF EDTIMTFIG.BNIEGGE I RAPHOL KEDERKINGE DMOLARDEVRING SUPPLI VIGORKESFEN 7. abstivum 12FEU DARSPHRITHKOMUTT DRELF EDTIMTFIG.BNIEGGE I RAPHOL KEDERKINGE DMOLARDEVRING SUPPLI VIGORKES 1. longificum 14 1. longificum 15 1. LAAPK KLIVVELVASVERD</pre>			110	120	130	140	150	160	170	180	190	200
T. aest ivum 26b FGLV DPMSPRRTMRQQLDTM DRLF DDAVGPF. TARSPAR RAKTP. REPOID HEDEREVKREP MEGUSEEEVVAVUEDALVIRGEHKEE. LAEGG S. ast ivum 21 FGLV DPMSPRRMRQQLDTM DRLF DDAVGPCRRSPAR TAUDY. RLFPDI VEDEKVRKER DMSQCSEEVVAVUEDALVIKGHKEE (AEGGS G. max 21 FGLD DPMSPRRMRQQLDTM DRLF DDAVTFG.NIGGGE IRXPPDI KDEEHEINREP MEGUSKEEVVAVUEDALVIKGKKESDERG A. thal ana 21 . FGLD DPMSPRRMRQQUDTM DRVF EDTTFPS.NIGGGE IRXPPDI KDEEHEINREP MEGUSKEEVVAUEDALVIKGKKESDERG A. thal ana 21 . DFVS DFJSPRTMRQQMIDTM DRVF EDTTFPS.NIGGGE IRXPPDI KDEEHEINREP MEGUSKEEVVAUEDALVIKGKKESDERG C. rubrum 23 P. FAAT R. SVQQLMLANDQLM NPFF. COTTFPSGSNRGTGE IRXPPDI KDEEHEINREP MEGUSKEEVVAUEDALVIKGKKESDERG C. rubrum 23 P. FAAT R. SVQQLMLANDQLM NPFF. COTTFPSGSNRGTGE IRXPPDI KDEEHEINREP MEGUSKEEVVAUEDALVIKIGKKESDERG C. rubrum 24 P. FAAT R. SVQQLMLANDQLM NPFF	T.aestivum 26a	FGLV	DPMSPMRTMROM	LDTM DRI	F DDAVGFP.	.TRRSPAA	RAR RRMPWD1	MEDEKEVKMRF	DMPGLSREEV	RVMVEDDALV	TRGEHKKE	AGEGO
Z. mays 26 SPEGLV DPMSPHRTMROULDTM DRLF DDAVGTPMCTRSPAT TCOV. RLPMPE INDERLEMRE DMSGLARDEVKVAVEDDTLVIRGEHKEE LABOOS P. saft Vun FGLL DPMSPHRSMROULDTM DRLF EDATTEG. NNIGGE IRVPMEI NDERLEMRE DMSGLAREDVKVSVEDDULVIRGEHKEEENG G. max 21 FGLL DPMSPHRSMROULDTM DRVF EDTMTFG. NNIGGE IRVPMEI NDERLEMRE DMSGLAREDVKVSVEDDULVIRGEHKEEENG P. hybrida 21 FGLL DPMSPHRTMROUNDT DRLF EDTMTFGSRNROTE IRVPMEI KDEELKNRP DMSGLAREDVKVSVEDDULVIRGEHKE L. longiflorum 18.2 LLAAFH KLTVRLEVASVERD	T.aestivum 26b	FGLV	DPMSPMRTMROM	LDTM DRI	F DDAVGFP	. TARSPAR	RAKTP . RMPWDI	MEDEKEVKMRF	DMPGL SREEV	RVMVEDDALV	TRGEHKKE.	AGEGO
P. aat 1 EGL DEWSPREEMENDLITM DETF EDATTFEG. NITGGGE I RAPPD I KDEFEITRMEP MEGUSKEVVKVSUEDUULVISCHER, E EARD A. thai ana 21 EFGL DEWSPREEMENDLITM DEWF EDTMTPRG. NITGGGE I RAPPD I KDEFEITRMEP MEGUSKEVVKVSUEDUULVISCHERKE Phybrida 21 EGL DEWSPREEMENDLITM DEWF EDTMTPRG. NITGGGE I RAPPD I KDEFEITRMEP MEGUSKEVVKVSUEDUULVISCHERKE C. rubrum 23 EGL DEWSPREEMENDLITM DEWF EDTMTPRG. NITGGGE I RAPPD I KDEFEITRMEP MEGUSKEVVKVSUEDUULVISCHERKE C. rubrum 71 ELL DEWSPREEMENDLITM DEWF EDTMTPRG. NITGGGE I RAPPD I KDEFEITRMEP MEGUSKEVVKVSUEDUULVISCHERKE EEE L. longiflorum 18. 2 L. LAPPH KUTVRLEVASVEKD	Z.mavs 26	SPFGLV	DPMSPMRTMROM	LOTM DRI	FDDAVGFPM	GTRRSPAT	TGDV RL PWD1	VEDEKEVKMRT	DMPGLARDEV	KVMVEDDTLV	TRGEHKKEE	GARGOS
G.max 21	P.sativum 21	FGLL	DPWSPMRSMROM	LOTM DR	FEDAITIPG	RNIGGGE	TRVPWE1	KDEEHETRMRF	DMPGVSKEDV	KVSVEDDVLV	TKSDHR	FFNG
A. thaliana 21	G.max 21	FGIL	DPWSPMRSMROI	LOTM DRV	F EDTMTFPG	RNTGGGE	T RAPWD1	KDEEHETRMRF	DMPGLAKEDV	KVSVEDDMLV	TROCHESE	OFHG
P. hybrida 21	A.thaliana 21	DPWS	DPLSPMRTMROM	LDTM DR	IF EDTMPVSG	RNRGGSG	V. SETRAPWDI	KEEEHEIKMRF	DMPGLSKEDV	KISVEDNVLV	IKGEOKKE	
C.rubrum 23 P. FRAT R. SVGQLMNLMDQLM ENFF MAASR GSGRAMRROWEV REDEELSLKV DMPGLAKEDVKVSVEDNTLI IKSEAEKE L longiflorum 16.2 L AAFH KLTVRLEVASVFKD. ATPADI KNLPANLYFI MMFWRGEIKVSVEDDBL/VISGERKKR. EEL L Ongiflorum 16.5 L AAFH KLTVRLEVASVFKD. ANTRADI KNLPANLYFI MMFWRGEIKVSVEDDBL/VISGERKKR. EEL Z mays 17.5 L AAFH KLTVRLEVASVFKD. ANTRADI KNLPANLYFI MMFWRGEIKVSVEDDBL/VISGERKRE. EEL Z mays 17.5 MMAALGHLLDVPDGDA GAGD DXX.GGGARTYTVP DAARMAATTDI KDMFGAYAFVV DMFGLGTGIKVQVEDEEVLVISGERKRE. EEL Z mays 17.5 MMAALGHLDLVPDGDA GAGD DXX.GGGARTYTVP DAARMAATTDI VMFGAYAFVV DMFGLGTGIKVQVEDEEVLVISGERKRE. EEL Z mays 17.5 MMAALGHLDLVPDGDA GAGD DXX.GGGARTYTVP DAARMAATTDIVKELFGAYAFVV DMFGLGTGIKVQVEDEEVLVISGERKRE. EEL Z mays 17.5 MMAALGHLDLVPDGDA GAGD DXX.GGGARTYTVP DAARMAATTDIVKELFGAYAFVV DMFGLGTGIKVQVEDEEVLVISGERKRE. EEL G max 17.9 LFMTLHHIDLTDD. T TKKN. DFPEDITATYP DAARMAATTDIVKELFGAYAFVV DMFGLGKGIKVQVEDEEVLVISGERKRE. EKD A.thaliana 17.6II LFF. HHIMDYAGD. D XSSK SSPERTFML DAAMAATTDIVKEYPNSYVFI DMFGLGSGIKVQVEDDUVVSGERTRE. EKD D.carrota 18.0 SNFDFSLDVMDFK DFPL VISSA.SEFGK ETAAFANTHIDW KEYPNSVYVFI DMFGLGSGIKVQVEDDUVVSGERTRE. KED D.carrota 18.0 SNFDFFSLDVMDFK DFPT NISALSASSFPC SNFDFFSLDVMDFK DFPF SISSASSFC SNSFDFFRLVKVVETEDDAVLGISGERKKE. KED G max 17.5 G RRSNFDFFSLDVMDFK DFPF PISLSA	P.hybrida 21	FGLL	DPMSPMRTMROM	MDTM DRI	FEDTMTFPG	SRNRGTGE	IRAPWDI	KDDENEIKMRF	DMPGLSKEEV	KVSVEDDVLV	IKGEHKKE.	
L longiflorum 18.2 LL AAFH KLTVRLEVASVPKD	C. rubrum 23	P. FRAT	R. SVGOLMNLM	DOLM EN	F	MAASR	GSGRAMRRGWDV	REDEFALELKV	DMPGLAKEDV	KVSVEDNULT	IKSEAFKE	
L longiflorum 17.6 LLAAFH KLTVVLEVASVPKD	L longiflorum 18 2	LLAAFH	KLTVRLEVASVE	800 DI			ΔΤΡΑΠΙ	KNLPDAVLVET	AMDDUDTOFT	KVEVEDDODI	MITCOEPER	
L longiflorum 16.5 AK LNQL TEFL. ANRAGPLRAPFV DARAMATAATDI KOMPGALVFII DMPGUSEEIKIDVEGUNU JUGUKAL, EEEE Z.mays 17.8 P LANALQHLLDYPODA GAGG DNKTOSGGANTRIYK DARAMATPADV KELGANAFVV DMPGLGTDIRVQVDERVLJVUSGERRR EEE P. activum 17.3 P MAALQHLLDYPODA GAGG DNK. GGGGTMTIYK DARAMATPADV KELGANAFVV DMPGLGTDIRVQVDERVLJVUSGERRR EEE P. activum 17.3 P MAALQHLLDIPODA GAGG DNK. GGGGTMTIYK DARAMATPADV KELGANAFVV DMPGLGTDIRVQVDERVLJVUSGERRR EEE G.max 17.9 P LFHTLQHMDMSED.G AGDI K GGGGTMTIYK DARAMATPADV KELPGANAFVV DMPGLGSDIKVQVEDRVLJUSGERRR EEE G.max 17.9 P LFHTLQHMDMSED.G AGDI K THNAPTWSVK DAKAMATPADV KELPGANAFVV DMPGLGSDIKVQVEDRVLJUSGERRR EEE G.max 17.9 P LFHTLQHMDMSED.G AGDI K THNAPTWSVK DAKAMATPADV KENPSVFEI DMPGLKSODIKVQVEDRVLJUSGERR EEE G.max 17.9 P LFHTLQHMDMSED.G AGDI K THNAPTWSVK DAKAMATPADV KEYPNSVVFI DMPGLKSODIKVQVEDRVLJUSGERRK EEE G.max 17.6 P LISTLQMLIPFDHN FK TRNNSFVYR DAKAMATPADV KEYPNSVVFI DMPGLKSODIKVQVEDRVLJVSGERRK KETA J. J. J. S. ALGU K J.	L longiflorum 17 6	LI. AAFH	KLTVRLEVASVP	кр Кр		•••••	ΔΠΡΑΠΙ	KNLDDAVLVET	DMPGVPTGET	RVEVEDDGAL	TICEPREE	EEE
Instruction Introduction	L longiflorum 16 5	AK LNOL	TEFI.	KD	ΔNDNO	DIDADEVD		KDMDCAVVETT	DMPOVEGEET	KVEVEDUSALI VIDVEEONMIN	TCCEPTOE	EEE
<pre>2. mays 17.5 2. mays 17.5 2. mays 17.5 2. mays 17.5 3. mp MMAALQHLLDYPGGDA GAGG DKAL GGGCPTHTYAD DARAMAYTPADVKELPGAYAFVV DMPGLGTGDIXVQVEDERVLJISGERRE EKE P. sativum 17.3 3. mp MMAALQHLLDYPGGDA GAGG DKAL GGGCPTHTYAD DARAMAATPADVKELPGAYAFVV DMPGLGTGDIXVQVEDERVLJISGERRE EKE 5. max 17.9 4. thaliana 17.611 5. mp LFHTLGHMMDMSED.G AGG NKTHNAPTMSYVF DAKAMAATPADVKELPGAYAFVV DMPGLGTGDIXVQVEDERVLJISGERRE EKE 5. max 17.9 5. mp LFHTLGHMMDMSED.G AGG NKTHNAPTMSYVF DAKAMAATPADVKEYPNSVYFID DMPGLKSGDIXVQVEDDNULJCGER.KK DEE 5. ml 11 18.8 5. mp LFHTLGHMDMSED.G AGG NKTHNAPTMSYVF DAKAMAATPADVKEYPNSVYFID DMPGLKSGDIXVQVEDDNULVVSGERRE EKE 5. max 17.611 5. mp LISTIQDMLDFADDHD RAGR APPEQPIRAYK DAKAMAATPADVKEYPNSVYFID DMPGLKSGDIXVQVEDDNULVVSGERRE KENE 5. carrota 18.0 5. carrota 18.0 5. carrota 18.0 5. carrota 18.0 5. carrota 18.0 5. carrota 18.0 5. carrota 18.1 5. ms NVDPFFSLDVMDFK DFF INSALSASFFGK ETAAFANTHIDW KETPGAHVFKA DLPGLKKEEVKVEVEEGKVLDISGERNEKEE 5. max 18.1 5. max 18.1 5. max 17.5 5. m</pre>	7 mays 17 8	D D D D D	IMANIOULIDVE	DODA GAO		CAMDIVID		KDHP GATVFII	DMPGVESEEI	DUOVEDEDUIN	TICCEPEDE	LELL
<pre>1. assivum 17.3 </pre>	7 mays 17 5	P	LMVALOHLLDVP	DODA GAC	C DKA CCC	COTOTVIA		KELDGAVAEW	DMPGLGTGDI	KVQVEDERVI I	TECEPERE	TDT
 Asativum 17.7 PERTLAHINDITOLT FERNLNAPTENTYVE DARAMATPADV KEHENSYVENV DMPGVKSGDIKVQVEDONULLIGGER.RREEE G. max 17.9 PERTLAHINDITOLT TERNLNAPTENTYVE DARAMATPADV KEHENSYVENV DMPGVKSGDIKVQVEDONULLIGGER.RREEE I. nil 17.2 PLFHTLGHMIDMSED.G AGDN KTHNAPTENTYVE DARAMATPADV KEYENSYVET I DMPGUKSGDIKVQVEDONULLIGGER.RREEE I. nil 17.2 PLFHTLGHMIDMSED.G AGDN KTHNAPTENTYVE DARAMATPADV KEYENSYVET I DMPGUKSGDIKVQVEDONULUSGERERREEE I. nil 17.2 PLFHTLGHMIDMSED.G AGDN KTHNAPTENTYVE DARAMATPADV KEYENSYVET I DMPGUKSGDIKVQVEDONULUSGERERREAEE I. nil 17.2 PLFHTLGHMIDMSED.G AGDN KTRNEPKVYRE DARAMATPADV KEYENSYVET I DMPGUKSGDIKVQVEDONULUSGERERREKEE I. nil 18.8 LEF.Q LLSTIQDMLDFADDHD RAGR APPEQPIRAYVP DARAMATPADV KEYENSYVET ID MPGUKSGDIKVQUEDONULUSGERENREKEE O. carrota 18.0 SR R SNVLPFSLDIWDFFK DFFF TNSALSASSFPQ ENSAFVSTRUDM KETPGANVFKA DLPGLKKEEVKVEVEEGKULQISGERNVEKEP Sativus 18.1 DFFSLDWDPFK DFFF TNSALSASSFPQ ENSAFVSTRUDM KETPEANVFKA DLPGLKKEEVKVELEDDORVLGISGERNVEKEP S. SNVFDFFSLDWDPFK DFFF FNSLSAENSAFVSTRUDM KETPEANVFKA DLPGLKKEEVKVELEDDORVLGISGERNVEKEP G. max 17.3 GR R SNVFDFFSLDWDPFK DFFF PSLSAENSAFVSTRUDM KETPEANVFKA DLPGLKKEEVKVQIEDORVLGISGERNVEKEP G. max 17.6 GR R SNVFDFFSLDWDFFK DFFF PSLSAENSAFVSTRUDM KETPEANVFKA DLPGLKKEEVKVELEDORVLGISGERNVEKEP G. max 17.6 GR R SNVFDFFSLDWDFFK PFF PSLSAENSAFVSTRUDM KETPEANVFKA DLPGLKKEEVKVELEDORVLGISGERNVEKEP G. max 17.6 GR R SNVFDFFSLDWDFFK FFF FFF.F. P.SLSAANCENCHYPEANVFNA DLPGLKKEEVKVELEDORVLGISGERNVEKEP A. thaliana 17.6 GR R SNVFDFFSLDWDFFF FFF.F. P.SLSAANCENCHYPEANTIDM KETPEANVFKA DLPGLK	T aestivum 17 3	P	MMAALOHLLDTP	DODA GAC	PFK 0	COTRAVUE		KELPCAVAEW	DMPGIGEGDT	KVQVEDERVD' KVOVEDEBVI V	TSCEPPER	TVE
C. max 17.9The LFHTLQHMADDMSED.G AGDN KTHNAFMSTVR DAKAMATPADV KEYPNSTVET DMPGLKSGDIKVQVEDONLLLIGGEN KRDEEI. nii 17.2	P catinum 17 7	r	LENTLUUTMDLT	משתית ממ				KEUDNOVJEMU	DMPCVKSCDT	KVQVEDEKVD	TECEP VD	EAE
J. nil 17.2 J. nil 12.2 J. nil 12.2 J. nil 17.2 J. nil 17.2 J. nil 17.2 J. nil 17.2 J. nil 12.2 J. nil 12.2 GR R SNVEDPFSLOWMOPK DFPF INSALSASFP Q ENSAFVSTRUDW KETPEAHVFKA DLPGLKKEEVKVELEDDRVLQISGERNVE KEP J. sativus 18.1 J DPFSLOWMOPK DFPF SNSPSG ENSAFVSTRUDW KETPEAHVFKA DLPGLKKEEVKVELEDDRVLQISGERNVE KEP J. max 17.5 GR R SNVEDPFSLOWMOPK DFPF SNSPSG ENSAFVSTRUDW KETPEAHVFKA DLPGLKKEEVKVUELDDRVLQISGERNVE KEP J. max 17.6 GR R SNVEDPFSLOWMOPK DFPF PSISLSA ENSAFVSTRUDW KETPEAHVFKA DLPGLKKEEVKVUELDDRVLQISGERNVE KEP J. max 17.6 GR R SNVEDPFSLOWDPFK DFPF PSISLSA ENSAFVSTRUDW KETPEAHVFKA DLPGLKKEEVKVUELDDKULQISGERNVE KED J. max 17.6 </td <td>G may 17 9</td> <td>r</td> <td>LEHTLOHMONS</td> <td>FD G AG</td> <td>NK THN</td> <td>APTWEVUP</td> <td>DAKAMAATPADV</td> <td>KEVPNGVVFFT</td> <td>DMPGLKSCDT</td> <td>KVOVEDENVLI</td> <td>TCGEP KP</td> <td></td>	G may 17 9	r	LEHTLOHMONS	FD G AG	NK THN	APTWEVUP	DAKAMAATPADV	KEVPNGVVFFT	DMPGLKSCDT	KVOVEDENVLI	TCGEP KP	
Inil 18.8 LPP.0 LLSTIQUMLDFADDHD RAGR APPEQPIRATYR DAKAMAATPADV KEVPNSYVFIA DMFOKAAELNQVEDDNVLVVSGERGREKEKD A. thaliana 17.611 PILSLEDMLEVFEDHN NEKTRNNFSKVINK DAKAMAATPADV KEVPNSYVFIA DMFOKAAELNQVEDDNVLVVSGERGREKEKD D. carrota 18.0 SRNSTVLNFSLDINDFFO DYPL ITSSGTSSEFGK ETAAFANTHIDW KETPAHVFKA DLPGLKKEEVKVEEGKVLQISGERNKEKEE M. sativus 18.1 DFSLDVNDPFK DYFF NNSLSASSFPQ ENSAFVSTRIDW KETPEAHVFKA DLPGLKKEEVKVEIEDDRVLQISGERSVEKED M. sativus 18.2 GR RSNVFDFFSLDVNDPFK DFFF NNSLSASSFPQ ENSAFVSTRIDW KETPEAHVFKA DLPGLKKEEVKVEIEDDRVLQISGERSVEKED S. max 17.5 GR RSNVFDFFSLDVNDPFK DFFF SINSSPSA.SFPR ENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEIEDDRVLQISGERSVEKED G. max 17.3 GR RSNVFDFFSLDWNDPFK DFFF PNSLSA.SFPR ENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEIEDDRVLQISGERNVEKED G. max 17.3 GR RSNVFDFFSLDWNDPFK DFFF PSSLSAENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEIEDDRVLQISGERNVEKED G. max 17.6 GP RSNVFDFFSLDWNDPFK DFFF PSSLSAENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEDDRVLQISGERNVEKED S. max 17.6 GR RNVFDFFSLDWNDPFK DFFF PSSLSAENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEDGDRVLQISGERNVEKED S. max 17.6 GR RSNVFDFFSLDWDPFF EFFF PSSLSAENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEDGDRVLQISGERNVEKED S. max 17.6 GR RSNVFDFFSLDWPFF EFFF PSSLSAENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEDGORVLQISGERNVEKED J. max 12 GR R	T nil 17.2	p	LE HHIMDYA		NS S	A PSRTEMI.	DAKAMAATPADU	KEVPNSVVETT	DMPGLKSGDT	KVOVDCDNVL	SISCER KR	FAFF
A. thaliana 17.611	T nil 18 8	LEP O	LISTIONLOFA			ODTDAVUD	DAKAMAATPADU	KEVDNEVVETA	DMPGUKAAFT	KVOVEDDAVI I	NICCEPTER	EALL
D. carrota 18.0SRRSRVLNPFSLDIWDPFQ DYPL ITSSTSSEFG K ETAAFANTHIDW KETPQAHVFKA DLPGLKKEEVKVEVEEGKVLQISGERNKE KEED. carrota 17.8GRRSNVLDPFSLDVWDPFK DYPL VTSSA.SEFG K ETAAFANTHIDW KETPQAHVFKA DLPGLKKEEVKVEVEEGKVLQISGERNKE KEEM. sativus 18.1	A thaliana 17 GTT	DC1Q	TISTLEDMLEVP	FDUN NEW	TPN	NPCRVVMR		TENDNAVAEUN	DMPGTKGDFT	KVOVENDNVLA	NSGERIER	NKEN
D. carrota 17.8GR N SWUDNFSLDWUPFK DFPL TISSISSEFGK ETAAFANITIDW RETPANTYKA DLPGLKKEEVKUVEEGAVLQISGERKEKEEM. sativus 18.1 DFFSLDWUPFK DFPF TNSALSASSFPQ ENSAFVSTRIDW KETPEANFYKA DLPGLKKEEVKUVEEGDRVLQISGERKEKEDP. sativus 18.1GR N SNVFDFFSLDWUPFK DFPF TNSALSASSFPQ ENSAFVSTRIDW KETPEANFYKA DLPGLKKEEVKUVEEDDRVLQISGERSVEP. sativus 18.1GR N SNVFDFFSLDWUPFK DFPF TNSALSASSFP R ENSAFVSTRUDW KETPEANFYKA DLPGLKKEEVKUVEUEDDRVLQISGERSVEP. sativus 18.1GR R SNVFDFFSLDWUPFK DFPF SDLSS.G. max 17.5GR R SNVFDFFSLDWUPFK DFPF PSLSAS. max 17.5GR R SVVFDFSLDWUPFK DFPF PSLSA.S. max 17.6GR R SNVFDFSLDWUPFK DFPF PNISSASFPESR.S. SMFDFSLDWUPFK DFPF FINSG.ENSAFVSTRVDW KETPEANFYKA DLPGLKKEEVKVEUEDDRVLQISGERNVEJ. esculentum 17.8DR R SSVFDFFSLDWUPFK DFPF PNISSASFPESR.S. SMFDFSLDWUPFK DFPF ELGP PSNSCESSAFNTRIDW KETPEANFYKA DLPGLKKEEVKVEVEDDRVLQISGERNVEJ. esculentum 17.6GR R TNVFDFSLDVUPFF ELGP PSNSCA. thaliana 17.6GR R TNVFDFFSLDUVUPFE GFLT P.SGLNAP.AK DVAAFTNAKVDW RETPEANFYKA DLPGLKKEEVKVEVEDGNILQISGERSEA. thaliana 17.6GR R SNVFDFFSLDLWUPFE GFLT P.SGLNAP.AK DVAAFTNAKVDW RETPEANFYKA DLPGLKKEEVKVEVEDGNILQISGERSEP. sativum 17.9GR R SNVFDFFSLDLWUPFE GFLT P.SGLNAP.AK DVAAFTNAKVDW RETPEANFYKA DLPGLKKEEVKVEVEDGNILQISGERSEP. sativum 16.90R SNVFDFFSLDLWUPFE GFT FSSLANASTA CDVAAFTNAKVDW KETPEANFYKA DLPGLKKEEVKVEVEDGNULVISGERKEP. sativum 16.90R SNVFDFFSLDLWDPFD TF R.SIVPAISGGSS ETAAFANARVDW KETPEANFYKA DLPGVKKEEVKVEVEDGNULVISGERKEP. sativum 16.90R.R SNVFDFFSLDLWDPFD TF R.SIVPAISGGS	D garrets 19 0	CDD	CARLINDER DTW	DDEO DVI		REC V		VEDDONINEVA	DI DOLVEDO	NUTUETOVUL (TOGENQUE	INICEIN
A. CaliblaC. CaliblaCaliblaC. Calibla <thcalibla< th=""></thcalibla<>	D carrota 17.9	CP P	CINVENET SEDIW		L 11556155	EFGK	ETAAFAN TRIDW	KETPQARVFKA	DIPCINEEV	NVEVEEGRVLY	DISCERNKE	· · · KEE
M. sativus 18.1	M aptime 19 1	GR R			E DNCALCAC	CED O	EIAAFVNIEIDW	KEIFQANVFKA	DI DCI VKEEV	NUEVEEGRVLY	LISGERNAE	NEL
A. SalivasGR.C.SNVFDFFSLDVNDFFDiffNNSDFS.SFFRENSAFVSTRVDWEFTEANVFAADiffDiffNVFDFSLDVNDFKDiffSNSFS.SFFRENSAFVSTRVDWEFTEANVFAADiffDiffNVFDFSLDVNDFKDiffSNSFS.SFFRENSAFVSTRVDWKETEANVFAADiffSNSFDASNS	M cativas 10.1	CP P	CARVEDDECI DVW		F INSALSAS	SFFQ	ENGAFVSIRIDW	KEIPEARVPKA	DEFGERREEV.	KVEIEDDRVLY	TECEREVE	KED
G. max 17.5GR RGR RGR RGR RGR RGR RGR RSVFDPFSLDUWDPFKFF PTSLSAENSAFVNTVDW KETPEAHVFKA DIPGLKKEEVKVQIEDDRVLQISGERNLEKEDG. max 18.5GR RSVFDPFSLDUWDPFKDFF PTSLSASFPEFSR.ENSAFVNTVDW KETPEAHVFKA DIPGLKKEEVKVQIEDDRVLQISGERNVEKEDG. max 17.6GP RSNVFDPFSLDWDPFKDFF PTSLSASFPEFSR.ENSAFVNTVDW KETPEAHVFKA DIPGLKKEEVKVQIEDDRVLQISGERNVEKEDJ. max 17.6GR RSNFDPFSIDVFDPFRELGF PSTNSG.ENSAFVNTVDW KETPEAHVFKA DIPGLKKEEVKVEVEDRVLQISGERNVEKEDJ. esculentum 17.8DR RSSSMFDPFSIDVFDPFRELGF PSTNSG.ESSAFANTRIDW KETPEAHVFKA DLPGLKKEEVKVEVEDRVLQISGERNVEKEDA. thaliana 17.6GR RTNVFDPFSLDVHPFFGFLT PGLTNAP.AKDVAAFTNAKUDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERSENEEA. thaliana 17.6GR RSNIFDPFSLDVHPFFGFLT PGLTNAP.AKDVAAFTNAKUDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERSENEEH. annuus 17.6SK RSNIFDPFSLDWDPFQ GIISTEPARETAAFNAHIDW KETPEAHVFKA DLPGKKKEEVKVEVEDGNULVGISGERSENEEP. aestivum 16.9b.RSNIFDPFSLDWDPFD TF R.SIVPAISGGS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNULVSGERSKEKEDD. sativa 16.9.RSNVFDPFSUDWDPFD TF R.SIVPAISGGS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNULVSGERSKEKEDD. sativa 17.2.R.SNVFDPFSUDWDPFD TF R.SIVPAISGGS ETAAFANARUW KETPEAHVFKA DLPGVKKEEVKVEVEDGNULVSGERSKEKEDD. sativa 17.4.R.SNVFDPFSUDWDPFD TF R.SIVPA	A.Sativus 10.2	GRR	CINVEDPESLOVW		F NNSALSA.	OFPR	ENSAFVSTRVDW	KETPEANVFKA	DLPGMKKEEV.	KVEIEDDRVL(DISCERSVE	KED
GR R.L.SNVFDPFSLDVWDPFK DFHF PISLSAENSAFVSTRVDW KETPEAHVFKA DIPGLKKEEVKVQIEDDRVLQISGERNVE KEDG. max 17.3GR RSNVFDPFSLDVWDPFK DFPF PSSLSAENSAFVSTRVDW KETPEAHVFKA DIPGLKKEEVKVQIEDDRVLQISGERNVE KEDG. max 17.6GP RSNVFDPFSLDWWDPFK DFPF PSSLSAENSAFVSTRVDW KETPEAHVFKA DIPGLKKEEVKVQIEDDRVLQISGERNVE KEDL. esculentum 17.8DR R.SSSMFDPFSLDVFDPFR ELGF PSTNSGENSAFVSTRVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERNVE KEDA. thaliana 17.6GR RTNVFDPFSLDVFDPFE GFLT P. SGLANAP.AMVAAFTNAKVDW RETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERSNE NEEA. thaliana 17.4GR RSNVFDPFSQDUWDPFE GFLT P. SGLANAP.AK DVAAFTNAKVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERSNE NEEA. thaliana 17.4GR RSNVFDPFSQDUWDPFE GFLT P. SGLANAP.AK DVAAFTNAKVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERSNE NEEA. thaliana 17.6SK RSNVFDPFSQDUWDPFE GFLT P. SGLANAP.AK DVAAFTNARVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNULQISGERSNE NEEA. thaliana 17.6GR RSNVFDPFSQDUWDPFE GFLT P. SGLANAP.A R DVAAFTNARVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNULQISGERSNE NEEA. thaliana 17.6GR RSNVFDPFSQDUWDPFQ GIISTEPASTAAFANAHIDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNULQISGERSNE NEEB. sativum 17.9GR RSNVFDPFSQDUWDPFQ MFQL ARSATCTTNETAAFANARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNULVSGERSNE KEDJ. saestivum 16.9cTSNVFDPFSLDWDPFD TF R. SIVPAISGGS ETAAFANARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNULVVSGERTKE KED<	C mar 17 E	GRR	CINVED DE CLOVW		F SNSSF5A.	5FFR	ENPAPVSIKVDW	KEIPEARVERA	DLPGLKKEEV.	KVEVEDDRVLG	JISGERSVE	KED
GR RSVFDPFSDDVWDPFK DFFF DFFF DFFF DFFF STUSAENSAFVSTRVDW KETPEANVFKA DIFGLKKEEVKVQLEDGVVLQISGERNVE KED3. max 17.6GP RSNVFDPFSDDVWDPFK DFFP PTLSSASSFPEFSRENSAFVSTRVDW KETPEANVFKA DIFGLKKEEVKVUEDGDVLQISGERNVE KEDL. esculentum 17.8DR RSSSMFDPFSDDVVDPFK ELGF PSTNSGESSAFANTRIDW KETPEANVFKA DIFGLKKEEVKVEVEDGNILQISGERNVE KEDA. thaliana 17.6GR RTNVFDPFSDDVVDPFF GFLT P.SGLANAP.AM DVAAFTNAKVDW RETPEANVFKA DLPGLKKEEVKVEVEDGNILQISGERNVE KEDA. thaliana 17.4GR RTNVFDPFSDDVDPFF GFLT P.SGLANAP.AK DVAAFTNAKVDW RETPEANVFKA DLPGLKKEEVKVEVEDGNILQISGERSKE NEEA. thaliana 18.2GR RSNVFDPFSDDVDPFF GFT PSSALANASTAR DVAAFTNAKVDW KETPEANVFKA DLPGLKKEEVKVEVEDGNULQISGERSKE NEEP. sativum 17.9GR RSNVFDPFSDDVWDFF GFT PSSALANASTAR DVAAFTNARVDW KETPEANVFKA DLPGLKKEEVKVEVEDGNULQISGERSKE KEDP. sativum 16.9b.R RSNVFDPFSDDVWDFF OG II STEPAETAAFANARVDW KETPEANVFKA DLPGVKKEEVKVEVEDGNULVSGERSKE KEDP. aestivum 16.9c	$G_{\text{max}} = 17.3$	GRR	COVED DECLOVA		F PISLSA	• • • • • • • • •	ENSAFVNTRVDW	KETPEAHVFEA	DIPGLAREEV.	KVQIEDDRVLQ	21SGERNLE	· · · KED
GR RINVEDPESDUWUPFE DFFPFFFILSASFEEFSK.ENSAFVNEVA KEPEANVFA DIFELKKEEVKVUEDDRVLQISGERNVE KEDJ. max 17.6GP RSNVEDPFSLDWUPFK DFFELGFPSSVSAENSAFVNEVDW KETPEANVFKA DIFGLKKEEVKVEVEDDRVLQISGERNVE KEDJ. esculentum 17.8DR RSSSMFDPFSLDVFDPFE GFLT P. SGLANAP.ABNSAFVNEVW KETPEANVFKA DIFGLKKEEVKVEVEDGNILQISGERNVE KEDA. thaliana 17.6GR RTNVFDPFSLDVWDPFE GFLT PGLENAP.AKDVAAFTNAKVDW RETPEANVFKA DIFGLKKEEVKVEVEDGNILQISGERSKE NEEA. thaliana 18.2GR RSNVFDPFSLDVWDPFE GFLT PGLENAP.AKDVAAFTNAKVDW RETPEANVFKA DIFGLKKEEVKVEVEDGNILQISGERSKE NEEA. thaliana 17.6SK RSNVFDPFSLDVWDPFE GFLT PGLENAP.AKDVAAFTNARVDW KETPEANVFKA DLPGLKKEEVKVEVEDGNILQISGERSKE NEEA. thaliana 17.6GR RSNVFDPFSLDVWDPFE GFLT PGLENAP.AKDVAAFTNARVDW KETPEANVFKA DLPGLKKEEVKVEVEDGNILQISGERSKE NEEA. thaliana 17.6GR RSNVFDPFSLDWDPFQ GIISTEPARETAAFANARVDW KETPEANVFKA DLPGVKKEEVKVEVEDGNILVISGERSKE NEEP. sativum 16.9R RSNVFDPFSLDWDPFD T.F R.SIVPAISGGS ETAAFANARVDW KETPEANVFKA DLPGVKKEEVKVEVEDGNILVVSGERTKE KEDJ. sativum 16.9aR RSNVFDPFSLDWDPFD T.F R.SIVPAISGGS ETAAFANARIDW KETPEANVFKA DLPGVKKEEVKVEVEDGNILVVSGERTKE KEDJ. sativum 16.9aR RSNVFDPFSLDLWDPFD SV.F R.SVVPATSDN. DTAAFANARIDW KETPEANVFKA DLPGVKKEEVKVEVEDGNILVVSGERTKE KEDJ. sativum 17.2R RSNVFDPFSLDLWDPFD GFF G SGSGSL.FPRANS DAAFASARIDW KETPEANVFKA DLPGVKKEEVKVEVEDGNULVSGERTKE <td>G. max 17.3</td> <td>GRR</td> <td>SSVFDPFSLDVW</td> <td></td> <td>F PSSLSA</td> <td>DDDDD</td> <td>ENSAFVSTRVDW</td> <td>KETPEAHVFKA</td> <td>DIPGLKKEEV.</td> <td>KLEIQUGRVLG</td> <td>21 SGERNVE</td> <td>· · · KED</td>	G. max 17.3	GRR	SSVFDPFSLDVW		F PSSLSA	DDDDD	ENSAFVSTRVDW	KETPEAHVFKA	DIPGLKKEEV.	KLEIQUGRVLG	21 SGERNVE	· · · KED
Gr RSNVFDPFSLDWWDPF DFN PIRV FISSAENSNFVN RVN KLUDARVLADIPELKEVKVVEDEDKVQISGERKVEKEDL. esculentum 17.8DR.SSSMFDPFSLDVFDPFR ELGF PSTNSGESSAFANTRIDW KETPERHVFKV DLPGLKKEEVKVVEDEDRVLQISGERKVEKEDA. thaliana 17.6GR RTNVFDPFSLDVFDPFR EJELF P.SGLANAP.AM DVAAFTNAKVDW RETPERHVFKA DLPGLKKEEVKVVEDGNILQISGERSNENEEA. thaliana 17.4GR RTNVFDPFSLDVWDPFE GFLT P.SGLANAP.AK DVAAFTNAKVDW RETPERHVFKA DLPGLKKEEVKVVEDGNILQISGERSKENEEA. thaliana 18.2GR RSNVFDPFSDDWDPFE GFT PSSALANASTA R DVAAFTNAKVDW RETPERHVFKA DLPGLKKEEVKVVEDGNULQISGERSKENEEH. annuus 17.6SK RSNIFDPFSLDTWDPFQ GII.STEPASTEPAETAAFANARVDW KETPERHVFKA DLPGUKKEEVKVEVEDGNULVISGERSKENEEP. sativum 17.9GR RTNAFDPFSLDLWDPFQ NFQL ARSATGTTNETAAFANARVDW KETPERHVFKA DLPGUKKEEVKVEVEDGNULVISGERSKEKEDT. aestivum 16.9b.R RSNVFDPFADLWADPFD TF R.SIVPAISGGSS ETAAFANARVDW KETPERHVFKA DLPGUKKEEVKVEVEDGNULVVSGERTKEKEDD. sativa 16.92J. aestivum 16.9a.R RSNVFDPFSDLDWDPFD TF R.SIVPAISGGSS ETAAFANARVDW KETPERHVFKA DLPGUKKEEVKVEVEDGNULVUSGERTKED. sativa 17.4.R RSNVFDPFSDLDWDPFD SV.F R.SVVPATSDN.DTAAFANARIDW KETPERHVFKA DLPGUKKEEVKVEVEDGNULVISGQRSRED. sativa 17.4.R RSNVFDPFSDLDWDPFD GFF G.SGSGSL.FPRANS DAAAFAGARIDW KETPERHVFKA DLPGUKKEEVKVEVEDGNULVISGQRSRED. sativa 17.4.R RSNVFDPFSLDEUMPFP	G. max 17.6	GRR	NNVFDPFSLDVW		F PNTLSSAS	FPEFSR	ENSAFVSTRVDW	KETPEAHVENA	DIPGLEREEV.	KVQIEDDKVL(21 SGERNVE	KED
D. escurent. DK R.S. SSMEDPFSLDVFDFFE ELGF FSIDV. ESSMEDPFSLDVFDFLELGF FSIDV. ESSMEDPFSLDVEDEDKUDISGERKNE NEE A. thaliana 17.6 GR R TNVFDPFSLDVWDPFE GFLT P.SGLANAP.AK DVAAFTNAKVDW RETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERSNE NEE A. thaliana 18.2 GR R SNVFDPFSQDLWDPFE GFTT P.SGLANAP.AK DVAAFTNAKVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNILQISGERSKE NEE H. annuus 17.6 GR R SNVFDPFSQDLWDPFE GFTT P.SGLANAP.AK DVAAFTNAKVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNULQISGERSKE NEE H. annuus 17.6 GR R TNVFDPFSLDTWDPFQ GII T GLTNAP.AK DVAAFTNAKVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNULQISGERSKE NEE H. annuus 17.6 GR R SNIFDPFSDLDWDPFQ GII T SGLANAP.A K DVAAFTNAKVDW KETPEAHVFKA DLPGLKKEEVKVEVEDGNULQISGERSKE NEE H. annuus 17.6 GR R TNAFDPFSLDWDPFQ GII T SGLANAP.A K DVAAFTNAKVDW KETPEAHVFKA DLPGUKKEEVKVEVEDGNULQISGERSKE NEE P. aaesivum 16.9b .R R SNVFDPFADLWADPFD TF R.SIVPAISGGSS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNULVVSGERTKE KED J. sativa 16.9c	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	GPR	CONFORMED		PISSVSA.	• • • • • • • • •	ENSAFVNIKVDW	KEIQEAHVLKA	DI PGLKKEEV.	KVQIEDDKVLQ	TCOERNVE	KED
A. thaliana 17.4 GR R INVFDPFSLDVVDPFE GFLT PGLTNAP.AK DVAAFTNARVDW RETPEAHVFKA DVPGLKKEEVKVEVEDGNLQISGERSSE NEE A. thaliana 17.4 GR R TNVFDPFSLDVVDPFE GFLT PGLTNAP.AK DVAAFTNARVDW RETPEAHVFKA DVPGLKKEEVKVEVEDGNLQISGERSSE NEE A. thaliana 18.2 GR R SNVFDPFSLDTWDPFG GIT PGLTNAP.A K DVAAFTNARVDW RETPEAHVFKA DLPGLKKEEVKVEVEDGNLQISGERSKE NEE H. annuus 17.6 SK R SNIFDPFSLDTWDPFG GIISTEPA R ETAAIVNARIDW KETPEAHVFKA DLPGLKKEEVKVEVEDGRVLQISGERSKE NEE P. sativum 17.9 GR R TNAFDPFSLDLWADPFQ NFQL ARSATGTTN ETAAFANAHIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGRVLVISGERSKE KED T. aestivum 16.9b .R R SNVFDPFADLWADPFD TF R. SIVPAISGGSS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGRVLVVSGERSKE KED T. aestivum 16.9c	http://www.icencum.icencum.icencum.icencum.icencum.icencum.icencum.icencum.icencum.icencum.icencum.icencum.icen	DRRS	SSMFDPF SIDVF		F PSINSG	·····	LOSAFANTKIDW	DEMDENNIARY	DUPGLAREEV.	KVEVEEDRVLQ	DISGERNVE	· · · KED
A. thaliana 18.2 GRR INVEDPESDDVWDPFE GFLT P.SALANASTA R DVAAFINARVDW RETPEAHVFKA DLPGLKKEEVKVEVEDGNULDISGERSKE NEE A. thaliana 18.2 GRR SNVFDPFSQDLWDPFE GFT PSSALANASTA R DVAAFINARVDW RETPEAHVFKA DLPGLKKEEVKVEVEDGNULDISGERSKE NEE H. annuus 17.6 SK R SNVFDPFSQDLWDPFE GFT PSSALANASTA R DVAAFINARVDW RETPEAHVFKA DLPGLKKEEVKVEVEDGNULDISGERSKE NEE P. sativum 17.9 GR R TNAFDPFSLDLWDPFQ GII. STAFANARVDW RETPEAHVFKA DLPGVKKEEVKVEVEDGNULVSGERSKE NEE T. aestivum 16.9b .R R SNVFDPFADLWADFFD T F R.SIVPAISGGSS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNULVSGERSKE KED T. aestivum 16.9c	A.LNallana 17.0	GRR			T P.SGLANA	P.AM	DVAAFINAKVDW	REIPEANVFRA	DUPGLEREEV.	KVEVEDGNIL(TCOPDOGE	NEE
A. Challana 10.2 GR R SNVFDPFSQDLWDFFE GFF FSSALANASIA KDVAAFINARVDW KETPEAHVFAA DLPGGKKEEVKVEVEDGRVLQISGERCRE QEE P. anuus 17.6 SK R SNVFDPFSLDLWDFFQ SI ISTEPA R ETAAIVNARIDW KETPEAHVFAA DLPGGKKEEVKVEVEDGRVLQISGERCRE QEE P. sativum 16.9b .R R SNVFDPFADLWADPFD T F R. SIVPAISGGSS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVVSGERSKE KED T. aestivum 16.9b .R R TNVFDPFADLWADPFD T F R. SIVPAISGGSS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVVSGERSKE KED D. sativa 16.9c	A thaliana 17.4	GRR	INVEDPESIDVW.		T PGLINA	P.A	DVAAF INALVDW	KEIPLANVENA	DV PGLKKEEV	KVEVEDGNILG	TROPROVE	NEE
A. anildus 17.9SR K SNIFDPFSDD/WDPFQ NFQ SI1SIFA KEIAAIVNARIDW KEIPEANVEKA DLPGUKKEEVKVEVEDGNVLVISGERKTE KEDP. sativum 16.9bR R SNVFDPFSDLWDPFQ NFQL ARSATGTTN FTAAFANARHDW KETPEANVFKA DLPGVKKEEVKVEVEDGNVLVVSGERSRET. aestivum 16.9c	A. Chailana 18.2	GRR	SNVFDPFSQULW.	DPFE GFF	T PSSALANA	DIAR	DVAAF TNAR VDW	VETPEANVERA	DIPGLKKEEV	NVEVEDANVLÇ	TCCEPCPE	NEE
P. salivum 16.9b RR SNVFDPFADLWDPFQ NFQL ARSNIGIN	D antinuus 17.0	OD D	SNIFDPF SLDIW.			EFA	EIAAIVNARIDW	VERDENINERN	DLPGMAREEV.	VEVEDGRVLL VUTT TODDVILL	TOOPDYNE	· · · QEE
1. aestivum 16.9D 1. KK SNVFDFFADLWADFFD 1F K.SIVFAISGGSS EIAAFANARVDW KEIPEAHVFKA DLPGVKKEEVKVEVEDGNVLVVSGERTKEKED 7. aestivum 16.9c DTFRSIVFAISGGSS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVVSGERTKEKED 7. aestivum 16.9c DTFRSIVFAISGGSS ETAAFANARVDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVVSGERTKEKED 0. sativa 16.9	P.Sacivum 17.9	GRA	INAF DFF SLDLW.		E ARBAIGII	20000	EIAAF ANARIDW	VEDDEN IN/PKN	DI POWKKEEV	VELEEDRVLF	NCOEDODE	KED
1. aestivum 16.9a .RRTNVFDPFADLWADFFD TFR.SIVPALSGGS ETAAFANARADW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVISGERTKEKED 2. sativa 16.9 .RRSNVFDPFSDLWADFFD TFR.SIVPALTSDN. DTAAFANARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVISGERTKEKED 2. sativa 16.9 .RRSNVFDPFSDLWDPFD TM.FRSIVPALTSDN. DTAAFANARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVISGERTKEKED 2. sativa 17.2 .RRSNVFDPFSMDLWDPFD TM.FRSIVPSATSTN.SETAAFASARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVISGERTKEKED 2. sativa 17.4 .RRSNVFDPFSMDLWDPFD GFPF G.SGSGSL.FPRANS DAAAFAGARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLVISGERTKEKED 2. sativa 17.4 .RRSNVFDPFSLDLWDPFD GFPF G.SGSGSL.FPRANS DAAAFAGARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLQISGERTKEQEE 2. sativa 17.4 .RRSNVFDPFSLDLWDPFD GFPF GLPSTLSTVPRSETAA ETAAFANARIDW KETPEAHVFKA DLPGVKKEEVKVEVEDGNVLQISGERTKEQEE 2. sativa 22 LL PFID SPNTLL.SDLWSDRFP DFF VLEQIPYGVEKHEFSI TLSHARVDW KETPEGHVIMU DVPGLKKEUKVEVEENRVLRVSGERTKEEEK 3. max 22 LL PFMD PPITLL.ADLWSDRFP DFF VLEHIPFGVDKDEASM AMSPARVDW KETPEGHVIML DVPGLKKEUKVEVEENRVLRVSGERTKEEEK 4. thaliana 22 LS SALE TTPGSLLSDLWLDRFP DFFK ILERIPLGLERDT.SV ALSPARVDW KETAEGHEIML DIPGLKKDEVKIEVEENRVLRVSGERTKEEEK	T.aestivum 16.90	. R R	SNVFDPFADLWA		F K.SIVPAL		EIAAFANARVDW	VETDEAUVEVA	DLPGVKKEEV	NVEVEDGINVLV	VSGERSKE	KED
1. acstivul 10.94	T. aestivum 16.90	 			R .SIVPAL.		E TAAF AWAR V DW	VEDENINERA	DLPGVKKEEV.	VEVEDGIVULV	VEGERIKE	··· KED
2. mays 17.2 SNVFDPFSMDLWDPFD SV.F.K.SVPATSDN.S DIAMFANARIDW KETPEANVFKA DLPGVKKEEVKVEVEDGNVLVISGQRSREKED 2. mays 17.2 SNVFDPFSMDLWDPFD TM.F.RSIVPSATSTN.S ETAAFASARIDW KETPEANVFKA DLPGVKKEEVKVEVEDGNVLQISGQRSREKED 2. sativa 17.4 SNVFDPFSMDLWDPFD GFPF GSGSGSL.FPRANS DAAAFAGARIDW KETPEANVFKA DLPGVKKEEVKVEVEDGNVLQISGQRSRE KED 2. subrum 18.3 GR R SNVFDPFSDLEWDPF FGLP STUTYPRSETAA ETAAFANARIDW KETPEANVFKA DLPGVKKEEVKVEVEDGNVLQISGQRSRE KED 9. sativum 22 LL PFID SPNTLL.SDLWSDFFP DPFR VLEQIPYGVEKHEPSI TLSHA RVDW KETPEGHVIMV DVPGLKKEEVKVEVEDRNVLRVSGERKKE EDK 3. max 22 LL PFID SPNTLL.ADLWSDRFP DPFR VLEHIPFGVDKDEASM AMSPA RVDW KETPEGHVIMV DVPGLKKEEVKVEVEENRVLRVSGERKKE EEK 4. thaliana 22 LS SALE TTPGSLLSDLWLDRFP DPFK ILERIPLGLERDT.SV ALSPA RVDW KETPEGHVIML DIPGLKKDEVKIEVEENRVLRVSGERKKE EEK	Castiva 16 0	.K.K	INVFDFFADLWA		F R. SIVPAL		EIAAFANAEMDW DRAAEANAD TDU	VETTERUVETA	DIPOVKKEEV	NVEVEDGINVLV	TECOREVE	NED
2.mays 17.2 RK SNVFDPFSMDLWDFPD IM.F KSIVFSAISTN.S EIAAFASARIDW KETPEAHVFKADDPGGKKEEVKVEVEDGRVLVISGGKSKEQEE 0.sativa 17.4 RK SNVFDPFSDLWDFP G.S.GSGSL.FPRANS DAAAFAGARIDW KETPEAHVFKADDPGKKEEVKVEVEDGRVLVISGGKRKEQEE 2.rubrum 18.3 GR R SNIFDPFSDLEIMDPF GFDSTLSTVPRSETAA ETAAFANARIDW KETPEAHVFKADDPGKKEEVKVEVEDGRVLRISGQRAREKEE P. sativum 22 LL PFID SPNTLL.SDLWSDRFP DPFR VLEQIPYGVEKHEPSI TLSHARVDW KETPEGHVIMU DVPGLKKEDKVEVEDGRVLRVSGERKKEEDK 3.max 22 LL PFMD PPITLL.ADLWSDRFP DPFR VLEHIPFGVDKDEASM AMSPARVDW KETPEGHVIML DVPGLKREEIKVEVEENRVLRVSGERKKEEEK A.thaliana 22 LS SALE TTPGSLLSDLWLDRFP DPFK ILERIPLGLERDT.SV ALSPARVDW KETAEGHEIML DIPGLKKDEVKIEVEENRVLRVSGERKKEEEK	J.Saliva 10.9	.KK	SIVEDPE SEDEW.	DPFD SV.	F R.SVVPA.	ISDN.	DIAAFANARIDW	KEIFEORVERA	DIPOVKKEEV		TEGORERE	··· KED
7. Sativa 17.4 INFEDERAL DEFECT. SUBJECT FOR	C. Mays 17.2	.KK	SNVFDPF SMDLW	DPFD IM.	F RELVESA.		DAAAFASARIDW	KETPEAHVFKA	DUPGVKKEEV	CVEVEDGINVLV	TECEPTER	KED
	0. saliva 17.4	.K.K	ONTEDPESIDEN			L.FPRAND	DAAAFAGARIDW	VENDENDURA	DV PGLAREEV		TROOPADE	···QLL
 Sacivum 22 LL PFID SPNTLL.SDLWSDRFP DPFK VLEQIPYGVEKHEPSI TLSHARVDW KETPEGHVIMV DVPGLKKDDIKIEVEENRVLRVSGERKKEEDK G. max 22 LL PFMD PPITLL.ADLWSDRFP DPFR VLEHIPFGVDKDEASM AMSPARVDW KETPEGHVIML DVPGLKKDEVKEVEVEENRVLRVSGERKKE EEK A. thaliana 22 LS SALE TTPGSLLSDLWLDRFP DPFK ILERIPLGLERDT.SV ALSPARVDW KETAEGHEIML DIPGLKKDEVKIEVEENRVLRVSGERKKE EEK 		GK K	DIVITOPTSLUED	ODED DEE	rSTLST	VPROLIAA	EIAAFAWARIDW	KEIPEANVFKA	DUPOL VNCCT	VEVEDONVLP	LAGORARE	NEE
5. max 22 LL FFMD PFTTLL.ADLWSDRFP DFFR VLEHIFFGUKDEASM AMSFARVDW KETPEGHVINL DVPGLKREEIKVEVEENRVLRVSGERKKEEEK A.thaliana 22 LS SALE TTPGSLLSDLWLDRFP DFFK ILERIPLGLERDT.SV ALSPARVDW KETAEGHEIML DIPGLKKDEVKIEVEENRVLRVSGERKREEEK	P. Sacivum 22	LL PFID	SPNTLL.SDLWS	UKFP DPF	K VLEQIPYG	VERHEPSI	TLSHAKVDW	KETPEGHVIMV	DV PGLKKDDI	LEVEENKVLF	WAGERKKE	EDK
A. CHAILANA ZZ LS BALE TTYGSLLSULWLUKFY UPFK ILEKIPLGLEKUT. SV ALSFA KVUW KETAEGHEIRL UI FGLKKDEVKIEVEENGULKVSGEKKKEJ EEK	j.max 22	LL PFMD	PPITEL ADLWS	UKFP DPF	K VLEHIPFG	VDKDEASM	AMSPAKVDW	KETPEGHVIML	DV PGLKKEELI	VEVEENKVLF	VEGERAKE	EEK
	A.LNAIIANA 22	LS SALE	TTPGSLLSDLWL	UKFP DPF	K IDERIPLG	LERDI'.SV	ALSPAKVDW	AETAEGHEIML	TIAGPEKEDEAI	TEARENGAPA	VEGERARE	EEK

FIGURE 1.—Continued

removed because they are under very different selective pressures than the rest of the proteins and evolve very quickly. The third codon positions were removed after it was determined that, in most pairwise comparisons, synonymous substitutions were saturated *i.e.*, greater than two substitutions per site. The topology of the trees generated using complete sequences and without the transit sequences and third positions were almost identical. Removal of the transit sequences and third positions decreased resolution for some closely related sequences but significantly increased the overall consistency index. The tree presented in this paper was constructed from data matrices in which the transit sequences and third positions were removed. There were 311 informative sites in the DNA data matrix. A 5:1 transitions:transversions weighting was used because this ratio was found to be the empirical values for these substitutions among the plant small heatshock protein data.

Amino acid distances were generated with Protdist in PHY-LIP using the categories option. The distance matrices were then used to construct trees with the neighbor joining (NJ) method. One hundred bootstrap replicates were generated using Seqboot and the consensus trees generated in Consense.

Rate analysis: Estimates of synonymous (Ks) and nonsynonymous (Ka) substitutions were generated by the program Li93 (LI 1993). Positions that included gaps were removed from the analysis. Estimates of the number of conservative and radical amino acid replacement substitutions per site were generated by the program SCR-PC (HUGHES *et al.* 1990). Sta-

tistical significance of pairwise comparisons were estimated with T tests.

RESULTS

Sequence conservation and divergence among small heat-shock proteins: The small heat-shock proteins are more conserved, across protein families, in the carboxyl-terminal (C-terminal) domain than in the aminoterminal (N-terminal) domain. In the N-terminal domain (amino-acids 1-152) there are family specific conserved regions (Figure 1). The chloroplast (CP)-, mitochondrial (MT)- and endoplasmic reticulum (ER)-localized proteins all have transit sequences that are specific for each organelle (Figure 1). The CPlocalized proteins also have a Met-rich region (amino acids 103-124) in the N-terminal domain (Figure 1 and VIERLING 1991). The class I cytosolic proteins have a consensus region in the N-terminal region (amino acids 107-120). The class II cytosolic proteins also have a small conserved region (amino acids 143-154) not present in the other protein classes at the very end of the N-terminal region.

The alignment of the small heat-shock proteins clearly shows the higher conservation in the C-terminal

Plant Small Heat-Shock Proteins

		210	220	230	240	25	0 260	270	280
т.	aestivum 26a	GEGGDGWWKERSVSS	YDMRLAL . PDE	CDKSQVRAEL	KNGVLLVSV	PKR		. ETERKVIDVQ	/Q
T.	aestivum 26b	GEGGDGWWKERSLSS	YDMRLAL.PDE	CDKSQVRAEL	KNGVLLVSV	KPR		.ETERKVIDVQ	/Q
z.	mays 26	GGDGDGWWKQRSVSS	YDMRLAL.PDE	CDKSKVRAEL	KNGVLLVTV	PKT		.EVERKVIDVQ	/Q
P.	sativum 21	GEDCWSRKSYSC	YDTRLKL.PDN	CEKEKVKAEL	KDGVLYITI	PKT		.KIERTVIDVQ	IQ
G.	max 21	GDDSWSSRTYSS	YDTRLKL.PDN	CEKDKVKAELI	KNGVLYITI	PKT		.KVERKVIDVQ	/Q
A.	thaliana 21	DSDDSWSGRSVSS	YGTRLQL . PDN	CEKDKIKAEL	KNGVLFITI	PKT	• • • • • • • • • • • • • • • •	.KVERKVIDVQ	IQ
Ρ.	hybrida 21	.SGKDDSWGRN.YSS	YDTRLSL.PDN	VDKDKVKAELI	KNGVLLISI	PKT		.KVEKKVTDVE	I
C.	rubrum 23	TEEEEQRRR	YSSRIELTPNL	YKIDGIKAEM	KNGVLKVTV	PKI		.KEEEKKDVFQ	VMVD
L.	longiflorum 18.2	KYQIMERWTGR	RMRKFER . PKN	RDTKAVSAVW	KNGVLAVTV	GKLLA	WEVAGLFFNIERLP	VPLPTKTKSIE	<i>V</i> KIEVKIA
L.	longiflorum 17.6	KYQMMERWTGK	RMRKFEL.PEN	ADTKAVSAVW	KNGVLAVTV	RKLPA	WEVAGISFNIERLP	VPLPTKTKSIE	/KIA
L.	longiflorum 16.5	RYLEMQRRMGK	MMRKFKL.LEN	ANSGAISAVC	KNGVLTVTV	EKLPS		.QEPKAIE	IKIA
Ζ.	mays 17.8	.DDAKYLRMERRMGK	FMRKFVL.PDN	ADVDKVAAVC	RDGVLTVTV	EKLPP		. PEPKKPKTIE	/KVA
Ζ.	mays 17.5	DAKYLRMERRMGK	FMRKFVL.PDN	ADMDKISAVC	RDGVLTVTV	EKLPP		.PEPKKPKTIE	/KVA
т.	aestivum 17.3	DAKYLRMERRMGK	LMRKFVL.PEN	ADMEKISP.C	RDGVLTVTV	DKLPP		. PEPKKPKTIQ	/QVA
₽.	sativum 17.7	KEGVKYLKMERRIGK	LMRKFVL.PEN	ANIEAISAIS	QDGVLTVTV	NKLPP		. PEPKKPKTIQ	/KVA
G.	max 17.9	KEGAKYLRMERRVGK	LMRKFVL.PEN	ANTDAISAVC	QDGVLSVTV	QKLPP		.PEPKKPRTIQ	<i>T</i> KVA
Ι.	nil 172	KEGAKYVRMERRVGK	LMRKFVL.PEN	ANKEKITAVC	QDGVLTVTV	ENVPP		.PEPKKPRTIE	/KIG
I.	nil 18. 8	KDGVKYLRMERRVGK	FMRKFVL . PEN	ANVEAINAVY	QDGVLQVTV	EKLPP		. PEPKKPKTVE	/KVA
A.	thaliana 17.6II	.EGVKYVRMERRMGK	FMRKFQL.PEN	ADLDKISAVC	HDGVLKVTV	QKLPP		. PEPKKPKTIQ	/QVA
D.	carrota	.KNDKWHPLEVSSGK	FLRRFRL.PEN	ANVDEVKAGM	ENGVLTVTV	PKVE.	• • • • • • • • • • • • • • • •	MKKPEVKSIH	[SG
D.	carrota 17.8	.KNDKWHRVERSSGK	FLRRFRL.PEN	AKVDEVKAAM	ANGVVTVTV	PKVE.		. IKKPEVKAID	ISG
Μ.	sativus 18.1	.KNDQWHRVERSSGK	FMRRFRL.PEN	akmd <u>o</u> vkaam	ENGVLTVTV	PKEE.		. IKKPEVKSIE:	ISS
М.	sativus 18.2	.KNDQWHRLERSSGK	FMRRFRL.PEN	akmdqvkaam!	ENGVLTVTV	PKEE.	• • • • • • • • • • • • • • •	VKKPEVKTID	[SG
Ρ.	sativum 18.1	.KNDQWHRVERSSGK	FMRRFRL.PEN	akmdqvkaam	ENGVLTVTV	PKEE.	• • • • • • • • • • • • • • • • •	. IKKAEVKSIE	ISG
G.	max 17.5	.KNDTWHRVERSSGN	FMRRFRL.PEN	AKVEQVKASM	ENGVLTVTV	PKEE.	• • • • • • • • • • • • • • • •	.VKKPDVKAIE	[SG
G.	max 17.3	.KNDTWHRVERSSGK	LVRRFRL.PENA	AKVDQVKASM	ENGVLTVTV	PKEE.	• • • • • • • • • • • • • • •	IKKPDVKAID	[SG
G.	max 18.5	.KNDTWHRVERSSGK	FMRRFRL.PENA	AKVEQVKASM	ENGVLTVTV	PKEE.	•••••	VKKPDVKAIE	[SG
G.	max 17.6	.KNDTWHRVDRSSGK	FMRRFRL.PENA	AKVEQVKACM	ENGVLTVTI	PKEE.	•••••	.VKKSDVKPIE	LSG
L.	esculentum 17.8	.KNDKWHRMERSSGK	FMRRFRL.PEN	AKMDQVKASM	ENGVLTVTV	PKEE.	• • • • • • • • • • • • • • • •	.VKKPEVKSIE	[SG
Α.	thaliana 17.6	.KNDKWHRVERSSGK	FTRRFRL . PEN	AKMEEIKASM	ENGVLSVTV	PKVP.	•••••	EKKPEVKSID	[SG
Α.	thaliana 17.4	.KSDTWHRVERSSGK	FMRRFRL . PENA	AKVEEVKASM	ENGVLSVTV	PKVQ.	•••••	ESKPEVKSID.	[SG
A.	thallana 18.2	. KNDKWHRVERASGK	FMRRFRL PENA	AKMEEVKAIM	ENGVLTVVV	PKAP.	•••••	EKKPQVKSID.	LSGAN
н.	annuus 17.6	.KDDIWHRVERSSGK	FIRRFRL.PENA	AKMDEVKAMM	ENGVLTVVV	PKEE.	•••••	EKKPMVKAID.	LSG
Р. —	sativum 17.9	.KNDTWHRVERSOGS	FLERFEL PENA	AKVDQVKAAMI	ENGVLTVTV	PREE.	• • • • • • • • • • • • • • • •	VKKPEAKPIQ.	L'IG
Τ.	aestivum 16.9b	.KNDKWHRVERSSGK	FVRRFRL.PEDA	AKVEEVKAGLI	ENGVLTVTV	PKAE.	•••••	VKKPEVKAIE.	LSG
T.	aestivum 16.90	.KNDKWHRVERSSGK	FVRRFRL PEDA	AKVEEVKAGLI	ENGVETVTV	PKAE.	• • • • • • • • • • • • • • • • •	VKKPEVKAIE.	LSG
<i>T</i> .	aestivum 16.9a	. KNDKWHRVERSSGK	FVRRFRL.LEDA	AKVEEVKAGLI	ENGVETVTV	PKAE.	•••••	VKKPEVKAIQ.	LSG
0.	sativa 16.9	.KNDKWHRVERSSGQ	FMRRFRL.PENA	AKVDQVKAGLI	ENGVLTVTV	PKAE	• • • • • • • • • • • • • • • • •	VKKPEVKALE.	LSG
Z.	mays 1/.2	. KUDKWHRVERSSGQ	FIRRFRL PDDA	AKVDQVKAGLI	ENGVLTVTV	PKAE.	• • • • • • • • • • • • • • •	EKKPEVKAIE	SG
<i>U</i> .	sativa 1/.4	. KTDKWHRVERSSGK	FERRERL. PED	TKPEQIKASMI	ENGVLTVTV	FKEE.	• • • • • • • • • • • • • • • • •	PKKPDVKSIQ	TG
Ľ.	ruprum 18.3	.KNDTWHKVERSSGQ	FMRKFRL PENA	AK V DQVKAGMI	ENGVLTVTV	FKNE.	DETECTORIS	APKPQVKAIN	· · · · · · · · · · · · · · · · · · ·
۲.	Salivum 22	.KGDHWHRVERSYGK	FWRQFKL . PQN	ULDSVKAKMI	ENGVETETE	RESH	DELEGREMVSIVEE	JUAPSKI VNDEI	A
. ب ا	max 22	. KGDAWARVERSYGK	FWRQFRL.PQN	ULUSVKAKLI	ENGVETETE	DKLSP	GALAGERVVSLAGE		KQEL
n -	LIIAIIANA 22	. KGDQWHRVERSYGK	FWRQFRL, PDN	UMESVIALL		TUPPL	EVAVOLKANITAVE	UQTAKI SSSES	SKEL
		#	# #	Ŧ	~ " # ###				

FIGURE 1. — Continued

domain (amino acids 152-282) (Figure 1). This domain contains four completely conserved and 15 highly conserved amino acids. The plant small heat-shock proteins share a consensus region (amino acids 166-193) (Figure 1 and VIERLING 1991) not present in other eukaryotic small heat-shock proteins. All plant small heatshock proteins also share a eukaryotic HS region (amino acids 214-250). The proline . . . glycine, valine, leucine amino-acid motif (amino acids 224, 239, 240, 241) in the HS domain is highly conserved among all eukaryotic small heat-shock proteins. This motif is highly conserved in the plant small heat-shock proteins. In the class II Lilium longiflorum HSP 16.5 and in Triticum aestivum HSP 16.9b the proline has been replaced by a leucine. The leucine at position 241 has been replaced by a valine in Daucus carota HSP 17.8.

Phylogenetic relationships of the small heat-shock proteins: To determine paralogous and orthologous relationships among the small heat-shock proteins, aligned amino acid and DNA sequences were analyzed using both distance (NJ)- and parsimony-based phylogenetic programs. Results from all of the analyses support the conclusion that the five major gene families form monophyletic groups and are most likely the result of gene duplications that occurred before the diversification of the angiosperms (Figures 2 and 3). The NJ tree generated from DNA distance matrices and the parsimony trees generated from amino acid data matrices are not shown but are highly congruent with the trees presented. In the NJ and parsimony trees the branches for individual gene families are highly supported by bootstrap analysis (Figures 2 and 3). It is not possible to deduce from this analysis the order of gene duplication events that gave rise to the five families, although the presence of both monocot and dicot sequences within each family indicates that the duplications occurred before the divergence of these two groups.

The class I cytosolic gene family contains paralogous genes. The phylogenetic relationships among the class I sequences are not always congruent with organismal relationships. The dicot sequences *H. annuus* HSP 17.6, *C. rubrum* HSP 18.3 and the *P. sativum* HSP 17.9 are consistently more closely related to the monocot (*T. aestivum*, *Z. mays* and *O. sativa*) sequences than to the other dicot sequences (Figures 2 and 3). This indicates that there have been duplications within the class I family.

There is evidence of gene conversion within the class I gene family: With the exception of the P. sativum HSP 17.9 and 18.1, and the O. sativa HSP 17.4 and 16.9 sequences, class I sequences from a single species are 790



FIGURE 2.—Parsimony tree based on DNA sequences. Strict consensus of the six most parsimonious trees. Tree length, 1619; consistency index, 0.456. Branch lengths are proportional to changes found along the branches. The tree is rooted with the sequences for the CP-localized proteins. The number of times out of the 100 bootstrap replicates that a branch was present is noted above the branch; values below 50 are not noted.

each other's closest relatives (Figures 2 and 3). This pattern suggests that gene conversion is homogenizing some of the class I sequences. Separate parsimony analysis of the DNA sequences coding for the N-terminal and C-terminal domains have the same topology (data not shown), suggesting that if gene conversion is occurring it is not localized to one part of the genes.

Duplication and divergence of class II sequences: The class II genes from *L. longiflorum, Z. mays* and *I. nil* are developmentally and differentially expressed (BOUCHARD 1990; KRISHNA *et al.* 1992; KOBAYASHI *et al.* 1994). However, nothing is known about the function of these proteins. I examined the rates of nucleotide substitution and amino acid replacements for evidence of functional divergence among the class II proteins.

Sequences from *L. longiflorum* were isolated from meiotic cDNA libraries generated from microgametophyte tissue (BOUCHARD 1990; KOBAYASHI *et al.* 1994). *L. longiflorum* HSP 18.2 is induced by both meiosis and heat (BOUCHARD 1990); *L. longiflorum* HSP 17.6 and 16.5 are expressed during meiosis and it is not known if they are also expressed during heat shock (KOBAYASHI 1994). All three *L. longiflorum* proteins are clearly class II small heat-shock proteins although 18.2 and 17.6 have lost part (six amino acids) of the class II consensus region. Pairwise comparisons of the class II *L. longiflorum* sequences show an interesting pattern of sequence divergence, in that the DNA sequences are more similar than the corresponding amino acid sequences (Table 2). This pattern of similarity was not found in any of the other pairwise comparisons of the other plant small heat-shock proteins. On closer inspection the DNA alignments revealed that many of the third codon positions were conserved among these sequences while first and second codon positions were not. There are no significant differences in percentage G + C content or codon usage among the Lilium genes.

To explore this pattern of sequence divergence in more detail synonymous and nonsynonymous substitutions among the *L. longiflorum* genes were examined. Comparisons were made with complete sequences (Table 3). In addition class II sequences from *I. nil* and *Z. mays* were examined. The *I. nil* HSP 18.8 gene is induced by both heat-shock and the photoperiod changes that induce flowering, whereas 17.2 is induced by heat shock alone (KRISHNA *et al.* 1992). *Z. mays* HSP 17.5 is induced by heat shock and during pollen development (meiosis); while *Z. mays* HSP 17.8 is induced only by heat shock (ATKINSON *et al.* 1993).

When protein sequences are constrained by function, synonymous substitutions (Ks) are expected to be significantly higher than the nonsynonymous substitutions (Ka). In most, but not all, of the pairwise comparisons of the class II gene sequences the number of synonymous substitutions were higher than the number of nonsynonymous substitutions. The Ks between both *L. longiflorum* HSP 18.2 and 16.5, and *L. longiflorum* HSP 17.6 and 16.5 is not significantly greater than Ka (Table 3A).

The pattern of nonsynonymous substitutions was examined using the program of HUGHES *et al.* (1990), which distinguishes between conservative and radical amino acid replacements. Proteins under strong selection to maintain function are expected to have more conservative (within the same amino acid chemical group) than radical replacements (across chemical groups). In comparisons of the class II sequences, I used the category of hydrophobicity, since hydrophobicity is conserved in the C-terminal domain among all the eukaryotic small heat-shock proteins (NOVER 1990). It is hypothesized (NOVER 1990) that the conserved hydropathy profiles of these proteins reflect strong selective constraints related to the ability of the small heatshock proteins to form oligomers.

Comparisons of the *L. longiflorum* HSP 18.2 and 17.6 genes reveal that although Ks is higher than Ka, conservative replacements are not significantly more frequent than radical replacements (Table 3). Between *L. longiflorum* HSP 18.2 and 16.5, Ks is not significantly greater than Ka. However, conservative replacements are significantly more frequent than radical replace-



FIGURE 3.—NJ tree based on amino acid sequences. The number of times out of the 100 bootstrap replicates that a branch was present is noted above the branch; values below 50 are not noted.

ments (Table 3). The *I. nil* sequences that are differential expressed do not have significantly more conservative than radical replacement substitutions (Table 3).

Small heat-shock proteins do not evolve at equal rates: Relative rate tests (Wu and LI 1985) were conducted within the gene families *i.e.*, CP, II, I and ER to see if there are any differences in evolutionary rates within gene families. No evidence of differences in evolutionary rates within families (data not shown) were found.

It was then determined if rates of substitution were variable among gene families. To examine this, rates of evolution were compared among species pairs from which sequences of at least three gene families are available. Taxa were examined in pairs to control for organismal divergence time. For example, the divergence time should be the same for all of the genes in *Z. mays* and *T. aestivum*. If all of the small heat-shock protein genes are evolving at the same rate, then the number of substitutions per site between each family of orthologous genes (*e.g.*, between the CP and ER genes) of *Z. mays* and *T. aestivum* should be the same.

TABLE 2

Pairwise comparisons of Lilium longiflorum sequences

Comparison	DNA identity (%)	Amino acid identity (%)
18.2 vs 17.6	89.4	85.8
18.2 vs. 16.5	63.0	46.3
17.6 vs. 16.5	73.6	53.3

Percentage identity was estimated with GAP in GCG.

Rates of nonsynonymous substitution: I examined the total number of Ka of the complete gene sequences and of the portion of the genes coding for the N and C terminal domains (data not shown). The class II and ER genes are evolving more quickly than the CP and class I genes (Table 4). The genes for the ER proteins had a consistently higher Ka than the CP and class I genes. The class II genes also had a higher Ka than the CP and class I genes, but this difference in rate was not statistically significant in the Z. mays vs. T. aestivum comparison (Table 4). Compared to the other gene families the class II genes had significantly higher Ka values in the portion of the genes coding for the Nterminal domain (data not shown). The gene families had more similar Ka values in the portion of the genes coding for the C-terminal domain (data not shown).

Rates of conservative and radical amino acid replacements: The nonsynonymous substitutions were examined in more detail and designated as conservative and radical according to hydrophobicity. The CP, class II and ER proteins had significantly higher conservative than radical replacements (Table 5). This pattern is expected under strong selection if hydrophobicity is important for function. None of the class I gene comparisons had significantly higher conservative than radical replacements. However the class I genes had significantly more conservative substitutions than radical replacement substitutions in the portion of the gene coding for the C-terminal domain (data not shown).

DISCUSSION

Increasing complexity of gene families reflects the inceasing complexity of organisms and functional di-

E. R. Waters

TABLE	3
-------	---

Pairwise comparisons of class II sequences

	Ks	Ka
A. Pairwise estir (Ka) substi	nates of synonymous (Ks) and nonsyno tutions per site for the Class II sequence	nymous es
L. longiflorum 18.2 vs. 17.6	0.233 ± 0.096	$0.104 \pm 0.027^{***}$
L. longiflorum 18.2 vs. 16.5	0.512 ± 0.157	0.417 ± 0.068
L. longiflorum 17.6 vs. 16.5	0.287 ± 0.090	0.284 ± 0.051
I. nil 17.2 vs. 18.8	0.650 ± 0.212	$0.149 \pm 0.034^{***}$
Z. mays 17.8 vs. 17.5	0.182 ± 0.080	$0.040 \pm 0.017^{***}$
	Con	Rad
B. Pairwise (Rad) sub	estimates of conservative (Con) and rad ostitutions among the Class II sequences	ical s
L. longiflorum 18.2 vs. 17.6	0.100 ± 0.021	0.060 ± 0.022
L. longiflorum 18.2 vs. 16.5	0.383 ± 0.038	$0.261 \pm 0.046^{***}$
L. longiflorum 17.6 vs. 16.5	0.275 ± 0.038	0.226 ± 0.043
I. nil 17.2 vs. 18.8	0.203 ± 0.027	0.167 ± 0.035
Z. mays 17.8 vs. 17.5	0.063 ± 0.017	$0.024 \pm 0.014^{***}$

Values are means \pm SE. * indicates that Ks is significantly greater than Ka at the 0.05 probability level, ** 0.01 level and *** 0.001 level.

versification of gene products (OHTA 1991). Gene duplication has long been recognized as an important process in genome evolution. Once a gene duplicates, the new copy can accumulate substitutions and eventually diverge enough that a new function becomes possible. Gene duplication and divergence has been examined theoretically (NAGALAKI 1984; WALSH 1987, 1995; OHTA 1988a-c, 1991). This study has shown that gene duplication, sequence divergence and gene conversion have all played a role in the evolution of the small heat-shock protein genes in plants. The small heat-shock protein genes have evolved from the single gene found in most animals and fungi into a large super gene family in angiosperms. The diversification of small heat-shock proteins in plants may reflect molecu-

lar adaptations to stressful conditions unique to plants as well as evolution of functions not related specifically to high temperature stress. Analysis of patterns of substitutions reveals that the selective constraints on the small heat-shock protein gene families are not identical. Differences in selective constraint frequently reflect functional differences. This suggests that functional divergence has occurred among the small heatshock proteins in plants.

Evolutionary relationships among small heat-shock protein gene families: The order of the gene duplications that gave rise to the five small heat-shock protein gene families is not known and cannot be deduced from the phylogenetic analysis of the available sequences. More data on small heat-shock proteins in

Species comparison	СР		Class II	Class	I	ER
() = 0.000	A. No	onsynonymou	s substitutions, Ka, pe	r site		
T. aestivum vs. Z. mays	0.063 ± 0.000	.015	0.086 ± 0.025	0.080 ± 0.000	0.017	
G. max vs. P. sativum	0.083 ± 0.000	.017	0.110 ± 0.019	0.063 ± 0	0.015	0.104 ± 0.017
A. thaliana vs. P. sativum	0.143 ± 0.0143	.022	0.219 ± 0.029	0.136 ± 0.000	0.022	0.209 ± 0.025
A. thaliana vs. G. max	0.121 ± 0.000	.012	0.255 ± 0.031	0.106 ± 0	0.027	0.231 ± 0.028
Species comparison	CP vs. II	CP vs. I	CP vs. ER	II vs. I	II vs. ER	I vs.ER
T. aestivum vs. Z. mays	NS	NS		NS		
G. max vs. P. sativum	NS	NS	NS	**	NS	***
A. thaliana vs. P. sativum	***	NS	***	***	NS	***
A. thaliana vs. G. max	***	NS	***	***	NS	***

 TABLE 4

 Comparisons of nonsynonymous substitutions among gene families

Values are means \pm SE. * indicaes that the Ka for the two gene classes are different at the 0.05 probability level, ** 0.01 level and *** 0.001 level; NS indicates that the Ka for the two genes are not statistically different.

Comparisons of conservative and radiear substitutions among gene and the									
<u> </u>		СР		Class II		Class I		ER	
comparisons	Con	Rad	Con	Rad	Con	Rad	Con	Rad	
T. aestivum vs.	0.105	0.048	0.149	0.033	0.115	0.089			
Z. mays	± 0.012	$\pm 0.019^{**}$	± 0.0245	$\pm 0.016^{***}$	± 0.022	± 0.027			
G. max vs.	0.123	0.018	0.148	0.052	0.088	0.072	0.108	0.084	
P. sativum	± 0.022	$\pm 0.012^{***}$	± 0.024	$\pm 0.021^{***}$	± 0.019	± 0.025	± 0.019	± 0.024	
A. thaliana vs.	0.145	0.070	0.254	0.094	0.195	0.170	0.210	0.132	
P. sativum	± 0.024	$\pm 0.025^{***}$	± 0.029	$\pm 0.028^{***}$	± 0.026	± 0.035	± 0.025	$\pm 0.030^{***}$	
A. thaliana vs.	0.131	0.073	0.273	0.131	0.211	0.172	0.230	0.122	
G. max	± 0.022	$\pm 0.025^{**}$	± 0.030	$\pm 0.031^{***}$	± 0.027	± 0.036	± 0.026	$\pm 0.028^{***}$	

TABLE 5

Comparisons of conservative and radical substitutions among gene families

Values are means \pm SE. Con, conservative; Rad, radical. * indicates that Con is greater than Rad at the 0.05 probability level, ** 0.01 level, and *** 0.001 level.

early plants will be needed to determine the order of gene duplications. This work is in progress.

PLESOFSKY-VIG et al. (1992) hypothesized that the CPlocalized protein may have been transferred to the plant nucleus from a photosynthetic endosymbiont and therefore the CP protein family is only distantly related to the other plant small heat-shock protein families. The sequence conservation among the small heat-shock proteins argues against the hypothesis of an endosymbiotic origin of the CP protein. All of the plant small heatshock proteins share a plant consensus region in the Cterminal domain, in addition to the heat-shock region that is shared with other eukaryotic small heat-shock proteins. The plant consensus region is not conserved in other eukaryotic small heat-shock proteins (VIERLING 1991; PLESOFSKY-VIG et al. 1992; JONG et al. 1993). If the CP proteins were bacterial in origin, they would not share this region with the other plant small heat-shock proteins. It is then more likely that early in the plant lineage a single small heat-shock protein gene existed that had the plant consensus region. Multiple duplications of this gene gave rise to the many small heat-shock protein gene families early in the evolution of plants (*i.e.*, at least before the rise of the angiosperms).

Evolutionary relationships within small heat-shock protein gene families: The relationships of the genes for CP- and ER-localized proteins are congruent with organismal relationships and therefore these two gene families are most likely composed of orthologous genes. The phylogenetic relationships among the class I sequences is however more complex.

The phylogenetic relationships of the class I sequences suggests that gene conversion is occurring among some but not all of the class I genes. When gene conversion is frequent, all paralogous genes involved in the gene conversion event will be each others closest relatives in a phylogenetic analysis (SANDERSON and DOYLE 1992). When gene conversion does not occur at all or very infrequently, each group of paralogous genes will reflect organismal relationships (SANDERSON and DOYLE 1992). The class I sequences from *D. carota, M. sativum, G. max* and *A. thaliana* are all most closely related to other con-specific class I genes, *i.e., A. thaliana* HSP 17.6, 17.4 and 18.2. This pattern suggests that either there are new duplications in each species, or, more likely, that gene conversion is maintaining sequence similarity among class I genes. The relationships of the class I sequences could also be explained by numerous independent duplications within each lineage. However if gene duplications were this frequent, one would expect to see many more small heat-shock proteins than have been observed.

The sequence divergence among the class I genes within species suggests that while gene conversion occurs it is not frequent. In a study of globin genes, FITCH *et al.* (1991) were able to detect which portion of the gene was undergoing gene conversion by constructing trees using different regions of the globin genes. Trees constructed separately from small heat-shock protein gene sequences for the N and C terminal domain had the same topology as the trees based on the entire gene sequence. This indicates that gene conversion is not limited to either the N or C terminal domains. A similar pattern to that seen with the small heat-shock protein genes was reported with the genes for the small subunit of ribulose bisphosphate carboxylase (MEAGER *et al.* 1989).

Comparisons of some of the class II genes suggest that functional divergence is occurring within the class II family. It has been previously established that some of the class II genes are developmentally expressed. However it is not known if the differences in expression reflect differences in function. The class II genes in both Z. mays (ATKINSON et al. 1993) and L. longiflorum (BOUCHARD 1990) are expressed during heat-shock and flower development. The I. nil HSP 17.2 gene is induced during heat shock and is also induced by changes in photoperiod (KRISHNA et al. 1992). The I. nil HSP 18.8 gene is heat-inducible but is not induced by changes in photoperiod (KRISHNA et al. 1992). In comparisons of both Z. mays and I. nil class II sequences synonymous substitutions are significantly greater than nonsynonymous substitutions. However, the patterns of amino-acid replacement substitution (conservative vs. radical) between the I. nil small heatshock protein genes indicates that there may be functional divergence among the I. nil small heat-shock proteins.

Rapid divergence after gene duplication has been reported for other genes (LI and GOJOBORI 1983; LI 1985; GOODMAN et al. 1987). In these cases there was enough phylogenetic information to place the timing of gene duplications on a phylogenetic tree and to asses rates of nonsynonymous substitutions before and after the duplication events. These studies show that while the rate of nonsynonymous substitution may be high immediately after duplication, this rate does eventually slow down. The difficulty with interpreting the L. longiflorum data is that we do not have sequences from other closely related organisms and so it is not possible to date the duplications. There may be no selective constraint on the L. longiflorum HSP 16.5 kDa protein at all; it may be drifting with neutral substitutions. Another possibility is that after the duplication event this gene had a burst of nonsynonymous substitutions but is now under selection to maintain a new function. The ratio of Ks to Ka is both a function of selective constraints and the time since duplication. The equality in rates of synonymous and nonsynonymous substitutions may reflect the fact that the synonymous substitutions, which accumulate as function of time, are now reaching the level of the nonsynonymous substitutions. If these genes were sampled sometime in the future, Ks would be higher than Ka.

It is unlikely that the *L. longiflorum* HSP genes are pseudogenes. They are expressed and they do not have any misplaced start or stop codons. They have the conserved class II consensus region, in addition to the conserved plant heat-shock domain and the eukaryotic heat-shock domain. If they were pseudogenes, they would accumulate amino acid replacements at the same rate across the entire sequence and these conserved regions would not be maintained. Most likely the *L. longiflorum* genes are recently duplicated genes that are in the process of diverging in both sequence and function from an ancestral gene. More complete sampling within Lilium and related taxa will be needed before this can be determined with greater confidence.

Selective constraints among the small heat-shock protein gene families: The differences in evolutionary rate among the small heat-shock protein gene families found in this study suggest that these gene families have diverged in function. Equality of rates of nonsynonymous substitutions indicate that proteins are under similar selective constraints. The CP proteins have significantly fewer nonsynonymous substitutions than the class II sequences. The ER and class II sequences have significantly more nonsynonymous substitutions than the class I sequences. There are also differences in the ratio of conservative to radical amino acid replacement substitutions among the gene families. If the ratio of conservative to radical replacements reflects functional constraints, then the class I sequences are functionally distinct from the other classes. Recent in vitro studies indicate that some small heat-shock proteins can act as molecular chaperones (JAKOB et al. 1993; MERCK et al. 1993; JAKOB and BUCHNER 1994; LEE et al. 1995). If the small heat-shock proteins are molecular chaperones, the differences in selective constraint revealed by this study suggest that the individual small heat-shock protein families may have very different substrate specificities. It is also possible that some small heat-shock protein families may have evolved entirely new functions.

The evolution of the small heat-shock proteins in plants from a single gene to a very large multigene family composed of at least five gene families is an important example of gene family diversification. The application of molecular evolutionary analysis to DNA and amino acid sequences of unknown function can help to establish paralogous groupings and, most importantly, can identify possible instances of functional divergence. The assumption underlying this analysis is that sequence divergence reflects functional divergence. Where functional differences have already been established for other proteins (KARLIN et al. 1992), this has proved to be true. Our ability to obtain DNA and amino acid sequences has far outstripped our ability to conduct detailed in vitro and in vivo studies of protein function. The use of sequence analysis can help in the formulation of hypotheses concerning function that can then be tested in the laboratory.

This work was supported by a postdoctoral fellowship from the Research Training Group in the Analysis of Biological Diversification at the University of Arizona. I thank G. BHARATHAN, N. MORAN, R. SHEEHY and E. VIERLING for many helpful discussions. A. COLWELL, G. BHARATHAN, L. MCDADE, N. MORAN and E. VIERLING provided useful comments on previous versions of this manuscript.

LITERATURE CITED

- ARRIGO, A.-P., and J. LANDRY, 1994 Expression and function of the low-molecular-weight heat-shock proteins, pp. 335–373 in *The Biology of Heat Shock Proteins and Molecular Chaperones*, edited by R. I. MORIMOTO. Cold Spring Harbor Laboratory Press, Plainview, NY.
- ATKINSON, B. G., M. RAIZADA, R. A. BOUCHARD, J. R. FRAPPIER and D. WALDEN, 1993 The independent stage-specific expression of the 18-kDa heat shock protein genes during microsporogenesis in Zea mays L. Dev. Genet. 14: 15-26.
- BECKER, J., and E. A. CRAIG, 1994 Heat-shock proteins as molecular chaperones. Eur. J. Biochem **219**: 11–23.
- BECKMANN, R. P., L. A. MIZAEN and W. J. WELCH, 1990 Interaction of Hsp 70 with newly synthesized proteins: implications for protein folding and assembly. Science 248: 850–854.
- BOORSTEIN, W. R., T. ZIEGELHOFFER and E. A. CRAIG, 1994 Molecular evolution of the HSP70 multigene family. J. Mol. Evol. 38: 1–17.
- BOUCHARD, R. A., 1990 Characterization of expressed meiotic prophase repeat transcript clones of *Lilium*: meiosis-specific expres-

sion, relatedness, and affinities to small heat shock protein genes. Genome **33:** 68–79.

- CRAIG, E. A., J. S. WEISSMAN and A. L. HORWICH, 1994 Heat shock proteins and molecular chaperones: mediator of protein conformation and turnover in the cell. Cell 78: 365–372.
- EERNISSE, D. J., 1992 DNA translator and aligner—hypercard utilities to aid phylogenetic analysis of molecules. CABIOS 8: 177–184.

FELSENSTEIN, J., 1993 PHYLIP (Phylogeny Inference Package), Seattle.

- FITCH, D. H. A., W. J. BAILEY, D. A. TAGLE, M. GOODMAN, L. SIEU et al., 1991 Duplication of the Yn-globin gene is mediated by repetitive L1 LINE sequences in an early ancestor of simian primates. Proc. Natl. Acad. Sci. USA 88: 7396-7400.
- GENETICS COMPUTER GROUP, 1991 GCG Package. Genetics Computer Group, Madison, WI.
- GETHING, M.-J., and J. SAMBROOK, 1992 Protein folding in the cell. Nature 355: 33-45.
- GOODMAN, M., J. CZELUSNIAK, B. F. KOOP, D. A. TAGLE and J. L. SLIGHTOM, 1987 Globins: a case study in molecular phylogeny. Cold Spring Harbor Symp. Quant. Biol. 52: 875–890.
- HUGHES, A. L., 1993a Contrasting evolutionary rates in the duplicate chaperonin genes of *Mycobacterium tuberculosis* and *M. leprae*. Mol. Biol. Evol. 10: 1343–1359.
- HUGHES, A. L., 1993b Nonlinear relationships among evolutionary rates identify regions of functional divergence in heat-shock protein 70 genes. Mol. Biol. Evol. 10: 243–255.
- HUGHES, A., T. OTA and M. NEI, 1990 Positive darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. 7: 515-525.
- JAKOB, U., and J. BUCHNER, 1994 Assisting spontaneity: the role of HSP90 and small HSPs as molecular chaperones. Trends Biochem. Sci. 19: 205-211.
- JAKOB, U., M. GAESTEL, K. ENGEL and J. BUCHNER, 1993 Small heat shock proteins as molecular chaperones. J. Biol. Chem. 268: 1517-1520.
- JONG, W. W. D., J. A. M. LEUSNISSEN and C. E. M. VOORTER, 1993 Evolution of the alpha-crystallin/small heat-shock protein family. Mol. Biol. Evol. 10: 103-126.
- KARLIN, S., V. BRENDEL and P. BUCHER, 1992 Significant similarity and dissimilarity in homologous proteins. Mol. Biol. Evol. 9: 152-167.
- KNACK, G., Z. LIU and K. KLOPPSTECH, 1992 Low molecular mass heat-shock proteins of a light-resistant photoautotrophic cell culture. Eur. J. Cell Biol. 59: 166–175.
- KOBAYASHI, T., E. KOBAYASHI, S. SATO, Y. HOTTA, N. MIYAJIMA et al., 1994 Characterization of cDNAs induced in meiotic prophase in Lily micorosporocytes. DNA Research 1: 15–26.
- KRISHNA, P., R. F. FELSHEIM, J. C. LARKIN and A. DAS, 1992 Structure and light-induced expression of a small heat-shock protein gene of *Pharbitis nil*. Plant Physiol. 100: 1772–1779.
- LEE, G. J., N. POKALA and E. VIERLING, 1995 Structure and molecular chaperone activity of cytosolic small heat shock proteins from pea. J. Biol. Chem. 270: 10432–10438.
- LENNE, C., and R. DOUCE, 1994 A low molecular mass heat-shock proteins is localized to higher plant mitochondria. Plant Physiol. 105: 1255-1261.

- LI, W.-H., 1985 Accelerated evolution following gene duplication and its implication for the neutralist-selectionist controversy, pp. 333-352 in *Population Genetics and Molecular Evolution*, edited by T. OHTA and K. AOKI. Japan Scientific Society, Tokyo.
- LI, W.-H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36: 96–99.
- LI, W.-H., and T. GOJOBORI, 1983 Rapid evolution of goat and sheep globin genes following gene duplication. Mol. Biol. Evol. 1: 94-108.
- LINDQUIST, S., and E. A. CRAIG, 1988 The heat-shock proteins. Annu. Rev. Genet. 22: 631-677.
- MEAGHER, R. B., S. BERRY-LOWE and K. RICE, 1989 Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. Genetics 123: 845–863.
- MERCK, K. B., P. J. T. A. GROENEN, C. E. M. VOORTER, W. A. D. HAARD-HOEKMAN, J. HOROWITZ *et al.*, 1993 Structural and functional similarities of bovine alpha-crystallin and mouse small heat-shock protein. A family of chaperones. J. Cell Biol. 114: 255-261.
- NAGALAKI, T., 1984 Evolution of multigene families under intrachromosomal gene conversion. Proc. Natl. Acad. Sci. USA 81: 3796-3800.
- NOVER, L., 1991 Heat Shock Response. CRC Press, Boca Raton, FL.
- OHTA, T., 1988a Evolution by gene duplication and compensatory advantageous mutations. Genetics 120: 841-847.
- OHTA, T., 1988b Multigene and supergene families, pp. 41–65 in Oxford Survey in Evolutionary Biology, edited by P. H. HARVEY and L. PARTRIDGE. Oxford University Press, New York.
- OHTA, T., 1988c Time for acquiring a new gene by duplication. Proc. Natl. Acad. Sci. USA 85: 3509-3512.
- OHTA, T., 1991 Multigene families and the evolution of complexity. J. Mol. Evol. 33: 34–41.
- PLESOFSKY-VIG, N., J. VIG and R. BRAMBL, 1992 Phylogeny of the alphacrystallin-related heat shock proteins. J. Mol. Evol. 35: 537–545.
- RENSING, S. A., and U.-G. MAIER, 1994 Phylogenetic analysis of the stress-70 protein family. J. Mol. Evol. 39: 80–86.
- SANDERSON, M. J., and J. J. DOYLE, 1992 Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst. Biol. 41: 4–17.
- SCHNEIDER, H.-C., J. BERTHOLD, M. F. BAUER, K. DIETMEIER, B. GUIARD et al., 1994 Mitochondrial Hsp70/MIM44 complex facilitates protein import. Nature 371: 768-774.
- SWOFFORD, D. L., 1993 PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1. Computer program distributed by the Illinois History Survey, Champaign, IL.
- VIERLING, E., 1991 The heat shock response in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42: 579-620.
- WALSH, J. B., 1987 Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion. Genetics 117: 543-557.
- WALSH, J. B., 1995 How often do duplicated genes evolve new functions. Genetics 139: 421–428.
- WU, C.-I., and W.-H. LI, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82: 1741–1745.

Communicating editor: W.-H. LI