

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Agronomy & Horticulture -- Faculty Publications

Agronomy and Horticulture Department

2004

IMPLEMENTATION OF A MITOCHONDRIAL MUTATOR

Sally Mackenzie

Lincoln, NE, sam795@psu.edu

Ricardo Vilela Abdelnoor

Lincoln, NE

Follow this and additional works at: <https://digitalcommons.unl.edu/agronomyfacpub>



Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Botany Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

Mackenzie, Sally and Abdelnoor, Ricardo Vilela, "IMPLEMENTATION OF A MITOCHONDRIAL MUTATOR" (2004). *Agronomy & Horticulture -- Faculty Publications*. 1327.

<https://digitalcommons.unl.edu/agronomyfacpub/1327>

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



US 20040210962A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2004/0210962 A1**
Mackenzie et al. (43) **Pub. Date: Oct. 21, 2004**

(54) **IMPLEMENTATION OF A MITOCHONDRIAL
MUTATOR**

Related U.S. Application Data

(75) Inventors: **Sally Mackenzie**, Lincoln, NE (US);
Ricardo Vilela Abdelnoor, Lincoln, NE
(US)

(60) Provisional application No. 60/456,318, filed on Mar.
20, 2003.

Publication Classification

Correspondence Address:
JONDLE & ASSOCIATES P.C.
9085 EAST MINERAL CIRCLE
SUITE 200
CENTENNIAL, CO 80112 (US)

(51) **Int. Cl.⁷** **A01H 1/00**; C12N 15/82;
C12Q 1/68; C07H 21/04;
C12N 5/04

(52) **U.S. Cl.** **800/278**; 536/23.6; 435/6;
435/69.1; 435/419; 435/468;
530/370

(73) Assignee: **The Board of Regents of the University
of Nebraska**

(57) **ABSTRACT**

(21) Appl. No.: **10/806,038**

Plant MSH1 polynucleotides and polypeptides are
described. Also described are methods for the use and
modulation of such MSH1 polynucleotides and polypep-
tides.

(22) Filed: **Mar. 22, 2004**

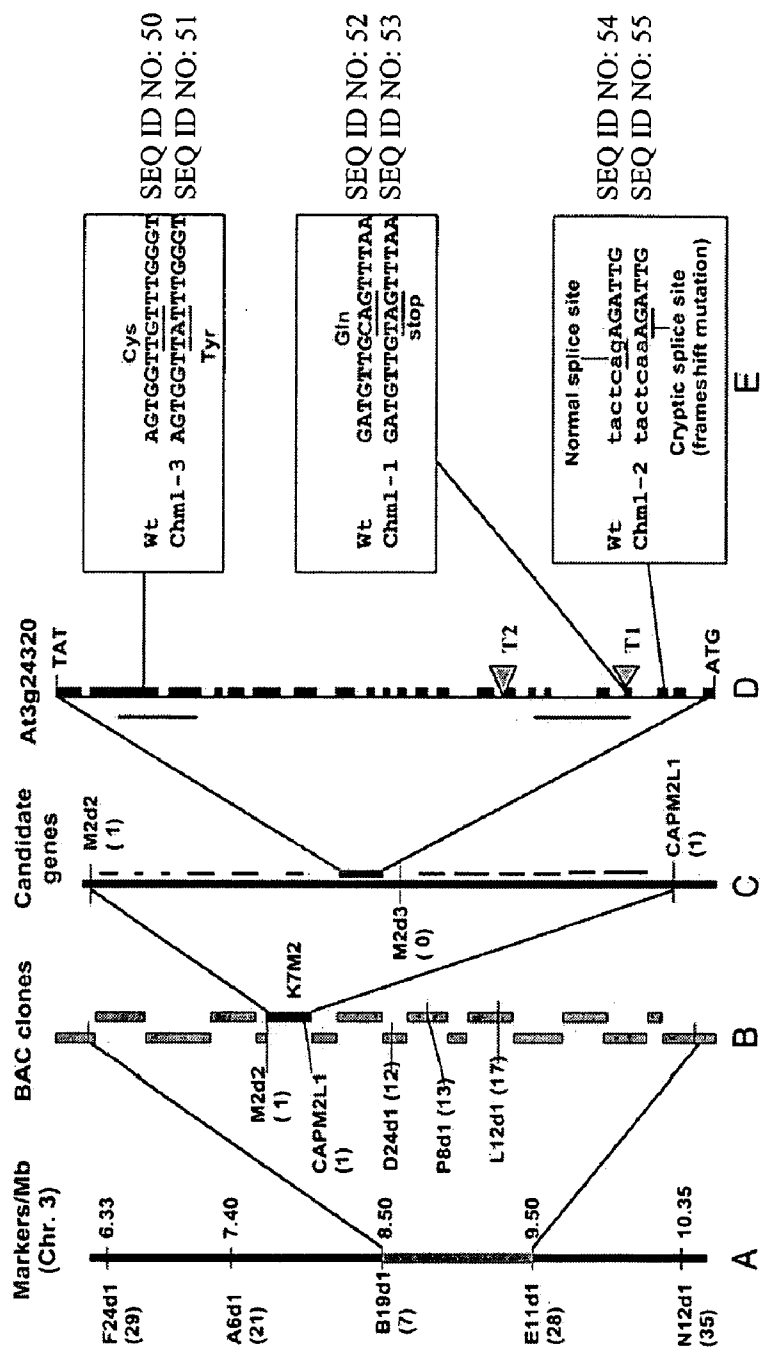


Figure 1

	1	50
Arabidopsis	(1) --MHWTATRNAVVSFPKWR---FFFRSSYRTYSSSLKPSSPITLNNRRYSEG	
Common Bean	(1) --MYRAVTRNVAMFLPRCSRSLSHSHSLPPFTISSLPSRFTRNGRVKRV	
Soybean	(1) --MYRVATRNVAFFPRCCSFAHVTPLPIETSFAPSRFTRNGCVKRV	
Rice	(1) MATORTLASSLVAATPRWLP-----VAADSFLRRRHRPRCSPTPALIFNR	
Tomato	(1) --MYWVTAKNVAVSVERWR-----GLSLFLRPPLRRRFLSFSPHTLC	
Consensus	(1) MYRV TRNVVVS PRWR F SSF F S PSR L ING V N	
	51	100
Arabidopsis	(46) ISCLRDKSLKRIITASKKVKTSDDVITDKDISHLVWWKERLOTCKKPSI	
Common Bean	(49) SYMDNNVSRGSSRTTKKPKVFNMMVDDKDLPHISWWKERLOMCKKFSI	
Soybean	(49) PSYTDKKVS-RGSSRAKKPKIFNNVDDKDLPHILWWKERLOMCKKFSI	
Rice	(46) RSWSKPRKVSRSISIVSRKMNKQGDLCNFGMLPHILWWKERMERCKKPSI	
Tomato	(41) REQTRCVKERKFFATTAKKIKOKSIPPEKDYVNMWWKERMEFLRKPSI	
Consensus	(51) SYIR K R S SKKLK P VLDDKDLPHILWWKERLQ CRKPSI	
	101	150
Arabidopsis	(96) VQLTERLMYTNLGLDPSLRNGSLKDGSLNWEMLQFKSRFPREVLLCRVG	
Common Bean	(99) VQLTORLESNGLGIDSKLNGSVKEGTLNWEMLQFKSRFPREVLLCRVG	
Soybean	(98) VQLTERLESNGLGINSNLKNGSLKEGTLNWEMLQFKSRFPREVLLCRVG	
Rice	(96) MQLTORLVYSNLGLDPTLRNGSLKDGSLNTEMLQFKSRFPREVLLCRVG	
Tomato	(91) ALLAKRLTYCNLGLVDPSLRNGSLKEGTLNSEMLQFKSRFPREVLLCRVG	
Consensus	(101) VQLI RL YSNLLGLDPSLRNGSLKEGTLNWEMLQFKSRFPREVLLCRVG	
	151	200
Arabidopsis	(146) EFYEAGIGDACLVEYAGLNPFGLRSDSIPKAGCPVNLROTLDLTLRN	
Common Bean	(149) EFYEAWGIDACMLVEYAGLNPCGGLOSDSVPRAGCPVNLROTLDLTLRN	
Soybean	(148) EFYEAWGIDACLVEYVGLNPIGGLRSDSIPRASCVPVNLROTLDLTLRN	
Rice	(146) DFYEAWGFDACLVEHAGLNPFGLRSDSIPKAGCPVNLROTLDLTLRN	
Tomato	(141) DFYEAGIGDACLVEYAGLNPFGLHSDSIPKAGCPVNLROTLDLTLRN	
Consensus	(151) EFYEAGIGDACLVEYAGLNPFGLRSDSIPKAGCPVNLROTLDLTLRN	
	201	250
Arabidopsis	(196) GYSVCIVEEVQGPTPARSRKGRFISGHAHPGSPYVYGLVGDHDLDFPDP	
Common Bean	(199) GYSVCILEEVQGPTQARSRKRRFISGHAHPGNPYVYGLAAMDHDLNFPDP	
Soybean	(198) GYSVCIVEEAQGPSQARSRKRRFISGHAHPGNPYVYGLATVDHDLNFPDP	
Rice	(196) GYSVCIVEETQGPTQARARKGRFISGHAHPGSPYVYGLAAMDHDLNFPDP	
Tomato	(191) GYSVCIVEEVQGPTQARARKSRFISGHAHPGSPYVYGLVXDDQDLDFPDP	
Consensus	(201) GYSVCIVEEVQGPTQARSRK RFISGHAHPGSPYVYGLA VDHDLDFPDP	
	251	300
Arabidopsis	(246) MPVVGISRARGYCMISTFETMKAYSLEDGLTEEAVTKLRTCRCHHFL	
Common Bean	(249) MPVVGISHSARGYCTNMVLETMTKTYSEDCLTEEAVTKLRTCRYHHLFL	
Soybean	(248) MPVVGISHSARGYCTNMVLETMTKTYSEDCLTEEAVTKLRTCRYHHLFL	
Rice	(246) MPVVGISRARGYCLTSVLETMTKTYSAEEGLTEEAVTKLRTCRYHHLFL	
Tomato	(241) MPVVGISRARGYCTISVYETMTKTYSEDGLTEEAVTKLRTCRCHHFL	
Consensus	(251) MPVVGISRARGYCIISVLETMTKTYSEDGLTEEAVTKLRTCRYHHLFL	
	301	350
Arabidopsis	(296) HASLRHNASGTCRWGEFEGGGLLWGECSRNFEWFDGDTLSELLSRVKDV	
Common Bean	(299) HTSLTQDSCGTSKRWGEFEGGGLLWGECSRHFEWFDGSPSLDLVKVKEL	
Soybean	(298) HTSLRRNSCGTCNWGEFEGGGLLWGECSRHFDWFDGNPVSLLAKVKEL	
Rice	(296) HSSLRNSSGTSRWGEFEGGGLLWGECSCKSFEWFDGNPIDELLCKVKEL	
Tomato	(291) HNSLKNSSGTSRWGEFEGGGLLWGECSNARQQEWLDGNPIDELLFKVKEL	
Consensus	(301) HTSLRNSSGTSRWGEFEGGGLLWGECSRNFEWFDGNPISELL KVKEL	
	351	400

Figure 3

Arabidopsis	(346)	YGLDDEVSFERNVNPSSKNRERPLHLGTATQIGALPTEGIPCLLKVLLPST
Common Bean	(349)	YGLDDEVTFRNITVSSSRHRARPLTLGTSTQIGAIHTEGIPSLKVLSPS
Soybean	(348)	YSLDDEVTFRNITVSSGHRAARPLTLGTSTQIGALPTEGIPSLKVLSPSN
Rice	(346)	YGLEEKTVFRNVSVSLEGRPOPLMLGTATQIGVLPTEGIPSLKVLPPN
Tomato	(341)	YGLNDDIPFRNVTVVSENRRPLHLGTATQIGALPTEGIPCLLKVLLPPH
Consensus	(351)	YGLDDEVTFRNVTVSS RPRPLHLGTATQIGAIPTGIPSLKVLPP
		401 450
Arabidopsis	(396)	CSGLPSLYVRDLLLNPPAYDIALKIQETCKLMSTVTCSEIPEFTCVSSAKL
Common Bean	(399)	CNGLPVLYIRNLLLNPPSYEIASKIQETCKLMSSITCSEIPEFTCVSSAKL
Soybean	(398)	CNGLPVLYIRELLLNPPSYEIASKIQAETCKLMSSVTCSSEIPEFTCVSSAKL
Rice	(396)	FGGLPSLYTRDLLLNPPSPDVASSVOEACRLMGSTCSEIPEFTCTPAAKL
Tomato	(391)	CSGLPVLYTRDLLLNPPAYETSSDIOEACRLMMSVTCSIPDFTCTSSAKL
Consensus	(401)	C GLPVLYIRDLLLNPPSYEIASKIQETCKLMSSVTCSSEIPEFTCVSSAKL
		451 500
Arabidopsis	(446)	VKLEQREANYIEFCRIKNVLDDVLHMRHAEIVELKLLMDPTWVATGL
Common Bean	(449)	VKLEWREYNHMEFCRIKNVLDEILHMYKSELNEILKNLMDPTWATGL
Soybean	(448)	VKLEWREYNHMEFCRIKNVLDEILQMYSTSELNEILKHLMDPTWVATGL
Rice	(446)	VKLESKEVNHMEFCRIKNVLDEVLFMGSNAELSAIINKLMDPAATVTGF
Tomato	(441)	VKLELREANHVFEFCIKSMVEEILQLYRNSLRATMELLMDPTWVATGL
Consensus	(451)	VKLE REYNHIEFCRIKNVLDEIL MYR SEL EILK LIDPTWVATGL
		501 550
Arabidopsis	(496)	KIDFDTLVNECHWASDTIGEMISLDENESHQNVSKCDNVPNEFFYDMESS
Common Bean	(499)	DIDFETLVSGCEVASSKISEPISLDGCGN-DQKINSLSIIPYEFFEDTESK
Soybean	(498)	EIDFETLVAGCEIASSKIGETVSLDDEN-DQKINSFSFTPHIEFFEDMESK
Rice	(496)	KVEADILVNECSFISORIAEVISLGGES-DQAITSEYTPKEFFNCGMESS
Tomato	(491)	KVDFTLVNECGKISCRISEISLVHGEN-DQKISSYPIIPNDFFEDMELL
Consensus	(501)	KIDFDTLVNEC AS KISEIISLDGEN DQKISS IP EFFEDMES
		551 600
Arabidopsis	(546)	WGRVKGTHTEEEITQVEKSAEALSLAIAEDFHPTISRTKATTASTGGPK
Common Bean	(548)	WKGRTKRVHIDEVETAVQKAAEVLHIAVTEDFVPVVSRIKATIAPIGGPK
Soybean	(547)	WKGRTKRIHDDVETAVEKAAEALHIAVTEDFVPVVSRIKATIAPIGGPK
Rice	(545)	WKGRTKRVHAEFEESNVDAIAEALSTAVIEDFLPIISRVKSVMSNGSSK
Tomato	(540)	WKGRTKRIHLEEAIAEVEKAADALSLAIAEDFLPIISRIKATMAPICCTK
Consensus	(551)	WKGRTKRIHIEE FT VEKAAEALSIAVTEDFLPIISRIKATMAPLGGPK
		601 650
Arabidopsis	(596)	GEISYAREHESVWFKGKRFPTSLWAGTAGEDQIKQLKPAALDSKGKKVGEE
Common Bean	(598)	GEISYAREHEAVWFRGKRFPTSLWSGSPGEEQIKQLRHALDSKGKRVGEE
Soybean	(597)	GEISYAREQEAVWFKGKRFPTNLWAGSPGEEQIKQLRHALDSKGKRVGEE
Rice	(595)	GEISYAKEHESVWFKGKRFPTPNWANTPGELQIKQLKPAALDSKGKRVGEE
Tomato	(590)	GEILYAREHGAVWFKGKRFVPTVWAGTAGEEQIKQLRPAALDSKGKKVGEE
Consensus	(601)	GEISYAREHEAVWFKGKRFPTSLWAGTPGEEQIKQLRPAALDSKGKKVGEE
		651 700
Arabidopsis	(646)	WFTTPKVEIALVRYHEASENAKARVLELLREL SVKLOTKINVLVFASMLL
Common Bean	(648)	WFTTPKVEAALTRYHEANAKATERVLELLRELATELHYSINTLVFSSSTLL
Soybean	(647)	WFTTPKVEAALTRYHEANAKAKERVLELLRGLAELOYSINTLVFSSSMLL
Rice	(645)	WFTTIKVENALTRYHEACDNAKRKVLELLRGLSSELQDKINVLVFCSTML
Tomato	(640)	WFTTMRVEDALTRYHEASARAKSRVLELLRGLSSELLSKINILVFASVLN
Consensus	(651)	WFTTPKVE ALTRYHEA AKAK RVLELLRGLSSELQ KINILVFASMLL

Figure 3 (cont'd)

		701		750
Arabidopsis	(696)	VISKALFSAHCEGRRRRKWFPTLVGFSLDEGAKPLDGASRMKLTGSPYW		
Common Bean	(698)	VITKALFAHASEGRRRRWVFPTLAESNGFEDVKSSDKIHGMKIVGLAPYW		
Soybean	(697)	VIKALFAHASEGRRRRWVFPTLVESHGFEDVKSLDKIHGMKISGLLPYW		
Rice	(695)	VITKALFGHVSEGRRRGWVLPITSP--LCKDNVTEFISSEMELSGTFPYW		
Tomato	(690)	VIKSLFSAHVSEGRRRNWLEPTTITQFNKCDTEALNGIDGMKILGSPYW		
Consensus	(701)	VITKALFAHASEGRRRRWVFPTL	ED KSLD T GMKISGLSPYW	
		751		800
Arabidopsis	(746)	FDVSSGTAVHNTVDMQSLFLLTGPNGGGKSSILRSTICAAALLGTCGLMVP		
Common Bean	(748)	FHTAEG-IVRNDVDMQSLFLLTGPNGGGKSSILRSTICAAALLGTCGLMVP		
Soybean	(747)	FHTAEG-VVRNDVDMQSLFLLTGPNGGGKSSILRSTICAAALLGTCGLMVP		
Rice	(743)	LDTNQGNAILNDVHMHSIFLLTGPNGGGKSSILRSTICAAALLGTCGLMVP		
Tomato	(740)	FDAARGTGVDTVDMQSLFLLTGPNGGGKSSILRSTICAAALLGTCGLMVP		
Consensus	(751)	FDIA G AV NDVDMQSLFLLTGPNGGGKSSILRSTICAAALLGTCGLMVP		
		801		850
Arabidopsis	(796)	AESACIPHFDSIMLHMKSYSVPDCKSSFOVEMSEIRSIIVSOATSRSLVL		
Common Bean	(797)	AESAVIPYFDSITLHMKSYSVPADCKSSFOVEMSEIRSIIGGTTKRSLVL		
Soybean	(796)	AESALIPYFDSITLHMKSYSVPADCKSSFOVEMSEIRSIIGGTTNRSLVL		
Rice	(793)	AASAVIPHFDSIMLHMKAYSVPADCKSSFOVEMSEIRSLVCRATARSRLVL		
Tomato	(790)	AESAVIPHFDSIMLHMKSYSVPDCKSSFOVEMSEIRSLITCATSRSLVL		
Consensus	(801)	AESAVIPHFDSIMLHMKSYSVPADCKSSFOVEMSEIRSIIGATSRSLVL		
		851		900
Arabidopsis	(846)	IDEICRGTTETAKGTCIAGSVVESLDTSGCLGIVSTHLHGIFSLPLTAKNI		
Common Bean	(847)	VDEICRGTTETAKGTCIAGSIETLERIGCLGVVSTHLHGIFTLPLNKKST		
Soybean	(846)	VDEICRGTTETAKGTCIAGSIETLDGIGCLGIVSTHLHGIFTLPLNKKNT		
Rice	(843)	IDEICRGTTETAKGTCIAGSIETRLDNVGCIGITSTHLHGIFDLPLSTHNT		
Tomato	(840)	IDEICRGTTETAKGTCIAGSVIETLDEIGCLGIVSTHLHGIFDLPLKTKKT		
Consensus	(851)	IDEICRGTTETAKGTCIAGSIETLDIGCLGIVSTHLHGIFTLPLIKNT		
		901		950
Arabidopsis	(896)	TYKAMGAENVEGQTKPTWKLTIDGVCRESLAFETAKREGVPESVIQRAEAL		
Common Bean	(897)	VHKAMGTTICIDGOILPTWKLTIDGVCKESLAFETAIREGTPEPIIRRAECL		
Soybean	(896)	VHKAMGTTSIDGOIMPTWKLTIDGVCKESLAFETAKREGTPEHIVVRAEYL		
Rice	(893)	DFKAMGTEIIDRCIOPTWKLMGTCRESLAFQTARKEGMPDLIRRAEEL		
Tomato	(890)	VYKAMGAEYVDGQOPTWKLTIDGTCCKESLAFETAOREGTPEIIRQAEEL		
Consensus	(901)	VHKAMGTEIDGQIIPTWKLTIDGVCKESLAFETAKREGIPEIIRRAEL		
		951		1000
Arabidopsis	(946)	YLSVYAK-----DASAEVMKPDQIITSSNN-----DQIQKPV		
Common Bean	(947)	YKSVYA-----EENFPNEEFSTCINNINNNTTSLY-----SKGFLSGA		
Soybean	(946)	YQLVYAKEMLFAENFPNEEFSTCINNINNNGTHLH-----SKRFLSGA		
Rice	(943)	YLAMSTN-----SKHTSSAVHHEISTANSTVNSIVEKPNYLRNGLELOS		
Tomato	(940)	YNSAYGNQIPRKIDQIRPLRSDIDLNSTDNSDOLNGTRQIALDSSTKLM		
Consensus	(951)	Y SVYA	EK S INI NL TTSL	A
		1001		1050
Arabidopsis	(979)	SSERSLEKDLAKATVVKICGKKMTEP-----EATECLSIGARELP		
Common Bean	(986)	NQMEGFROEVERAITVICODYIMERKNKKIALELPEIKCLLIGKREQPPP		
Soybean	(990)	NQMEVLRREVERAVTVICDHIKDLKCKKIALELPEIKCLLIGTRELPPP		
Rice	(987)	GSFGLLRKELESVVTICKKKGLDLYNKRSISELIEVVCVAVGAREQPPP		
Tomato	(990)	HRMGTSSKKLEDAICLICCKLTLEYKMKNPSEMPMNCVLTAAAREQAP		
Consensus	(1001)	M ILRKELERAITVIC KKIIL KK	EL EI CLLIGAREQPPP	
		1051		1100

Figure 3 (cont'd)

Arabidopsis	(1020)	STVGSSCVYVMRRPDKRLYTGQTDDEGRVRAHRAKEGTQGSSEFLYLMVQ
Common Bean	(1036)	SVVGSSSVYVMFTPDKKLYVGETDDLEGRVRRHRLKEGMDEASEFLYFLVP
Soybean	(1040)	SVVGSSSVYVMFRPDKKLYVGETDDLEGRVRRHRLKEGMHDEASEFLYFLVP
Rice	(1037)	STVGRSSHYVMIRRDSEKLYTGQTDDELVGRISAHRSKEGMDATILYILVP
Tomato	(1040)	STVGASSVYVMLRPDKKLYVGQTDDEGRVRAHRLKEGMENASEFLYFLVS
Consensus	(1051)	STVGSSSVYVM RPDKKLYVGQTDDEGRVRAHRLKEGM DASFLYFLVP
		1101 1150
Arabidopsis	(1070)	GKSMACQLETLLINQLHEOGVSLANLADGKHRNFGTSSSLSTSDVVSIL-
Common Bean	(1086)	GKSMACQFESLLINQLSSQGFOLSNMADGKHRNFGTSMYYA-----
Soybean	(1090)	GKSMACQFESLLINQLSGQGFOLSNMADGKHRNFGTSMYYT-----
Rice	(1087)	GKSMACQLETLLINQLPLKGEKLINKADGKHRNFGISLMPGEAIAA----
Tomato	(1090)	GKSMACQLETLLINQLPNHGFOLSNMADGKHRNFG-----
Consensus	(1101)	GKSIACQLETLLINQL QGFQLSNIADGKHRNFGTS L

Arabidopsis	SEQ ID NO: 3
Common Bean	SEQ ID NO: 47
Soybean	SEQ ID NO: 31
Rice	SEQ ID NO: 22
Tomato	SEQ ID NO: 40
Consensus	SEQ ID NO: 65

Figure 3 (cont'd)

IMPLEMENTATION OF A MITOCHONDRIAL MUTATOR

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. §119 from U.S. Application Ser. No. 60/456,318, filed Mar. 20, 2003, which is incorporated herein in its entirety by reference.

GOVERNMENT LICENSE RIGHTS

[0002] The U.S. Government has a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of the contracts awarded by the National Science Foundation and the Department of Energy.

TECHNICAL FIELD

[0003] This invention relates to using molecular and evolutionary techniques to identify polynucleotide and polypeptide sequences corresponding to commercially relevant traits in domesticated plants.

BACKGROUND OF THE INVENTION

[0004] The plant mitochondrial genome is retained in a multipartite structure that arises by a process of repeat-mediated homologous recombination. Low frequency ectopic recombination also occurs, often producing sequence chimeras, aberrant open reading frames, and novel subgenomic DNA molecules. This genomic plasticity may distinguish the plant mitochondrion from mammalian and fungal types. In plants, relative copy number of recombination-derived subgenomic DNA molecules within mitochondria is controlled by nuclear genes, and a genomic shifting process can result in their differential copy number suppression to near-undetectable levels. We have cloned a nuclear gene that regulates mitochondrial substoichiometric shifting in *Arabidopsis*. The CHM gene was shown to encode a protein related to the MutS protein of *E. coli* that is involved in mismatch repair and DNA recombination. We postulate that the process of substoichiometric shifting in plants may be a consequence of ectopic recombination suppression or replication stalling at ectopic recombination sites to effect molecule-specific copy number modulation.

[0005] Argument for the mitochondrion as a central regulator of cellular functions has become increasingly persuasive in the past several years, as information expands detailing cell metabolic functions (Golden & Melov, (2001) *Mech. Aging Dev.* 122, 1577-1589; Naviaux (2000) *Eur. J. Ped.* 159, 5219-5226), programmed cell death (Ravagnan, et al. (2002) *Cell. Physiol.* 192,131-137), and intracellular signaling (Epstein et al. (2001) *Molec. Biol. Cell.* 12,297-308). The disclosures of Golden & Melov, Naviaux, and all other patents and publications referred to herein, are incorporated herein in their entirety by reference. In higher plants, mitochondrial functions and behavior have clearly been influenced by the plant cell's unique context. Co-evolution of mitochondria and chloroplasts has permitted economy of function via protein dual-targeting (Small, et al. (1998) *Plant Molec. Biol.* 38, 265-277, Peeters & Small (2001) *Biochim. Biophys. Acta* 1541, 54-63), genome capacity and coding have been altered (Knoop & Brennicke (2002) *Crit. Rev.*

Plant Sci. 21,111-126), and the mitochondrial genomes of plants have acquired structural and maintenance features distinct from their animal counterparts.

[0006] The plant mitochondrial genome appears to be organized as a collection of small circular and large, circularly-permuted linear molecules (Oldenburg & Bendich (2001) *Molec. Biol.* 310, 549-562; Backert, et al. (1997) *Trend Plant Sci.* 2, 477-483), not unlike what has been postulated for yeast (Maleszka, et al. (1991) *EMBO J.* 10, 3923-3929; Lecrenier & Foury (2000) *Gene* 246,37-48). DNA replication may be conducted by a rolling circle mechanism, and experimental difficulties identifying replication origins have led to the suggestion of recombination-mediated replication initiation (Backert & Borner (2000) *Curr. Genet.* 37, 304-314). In fact, a distinct feature of plant mitochondrial genome organization is the prominent role of recombination.

[0007] High frequency inter- and intra-molecular recombination is detected within the higher plant mitochondrial genome at large repeated sequences that can be readily identified by physical mapping (Fauron, et al. (1995) *Trends Genet.* 11, 228-235). Their presence in direct orientation permits the subdivision of the genome into a collection of molecules, each containing only a portion of the genetic information. More intriguing, however, is the common observation in plants of intragenic ectopic recombination events that can occur at sites containing as few as seven nucleotides of homology (Andre, et al. (1992) *Trends Genet.* 8, 128-132). Ectopic recombination results in expressed gene chimeras that cause cytoplasmic male sterility, plant variegation and other aberrant phenotypes (Mackenzie & McIntosh (1999) *Plant Cell* 11, 571-585; Sakamoto, et al. (1996) *Plant Cell* 8, 1377-1390).

[0008] A phenomenon rendering the plant mitochondrial genome unusually variable in structure is termed substoichiometric shifting. First reported in maize (Small, et al. (1987) *EMBO J.* 6, 865-869) as the stable presence of subgenomic mitochondrial DNA molecules within the genome at near-undetectable levels, the process appears to be highly dynamic. Mitochondrial genomic shifting involves rapid and dramatic changes in relative copy number of portions of the mitochondrial genome over one generation's time (Janska, et al. (1998) *Plant Cell* 10,1163-1180). These substoichiometric forms have been estimated at levels as low as one copy per every 100-200 cells (Arrieta-Montiel, et al. (2001) *Genetics* 158, 851-864). Generally the rapid shifting process involves only a single subgenomic DNA molecule, often containing recombination-derived chimeric sequences, and the process is apparently reversible (Janska, et al., *ibid.*, Kanazawa, et al. (1994) *Genetics* 138, 865-870). Genomic shifting can alter plant phenotype because the process activates or silences mitochondrial sequences located on the shifted molecule. Observed phenotypic changes have included plant tissue culture properties (Kanazawa, et al., *ibid.*), leaf variegation and distortion (Sakamoto, et al., *ibid.*), and spontaneous reversion to fertility in cytoplasmic male sterile crop plants (Janska, et al., *ibid.*, Smith, et al. (1991) *Theor. Appl. Genet.* 81,793-798). It has been postulated that substoichiometric shifting may have evolved to permit the species to create and retain mitochondrial genetic variation in a silenced but retrievable form (Small, et al. (1989) *Cell* 58, 69-76).

[0009] Mitochondrial substoichiometric shifting has been shown in at least two cases to be under nuclear gene control, involving the Fr gene in *Phaseolus vulgaris* (Mackenzie & Chase (1990) Plant Cell 2, 905-912) and the CHM gene in *Arabidopsis* (Martinez-Zapater, et al. (1992) Plant Cell 4, 889-899; Redei (1973) Mut. Res. 18,149-162). Mutation of the nuclear CHM gene results in a green-white leaf variegation that, in subsequent generations, displays maternal inheritance (Redei, *ibid.*). The appearance of the variegation phenotype is accompanied by a specific rearrangement (Martinez-Zapater, et al., *ibid.*) that includes amplification of a mitochondrial DNA molecule encoding a chimeric sequence (Sakamoto, et al., *ibid.*). Genetic analysis suggests that the wildtype form of CHM actively suppresses copy number of the subgenomic molecule carrying the chimeric sequence. Loss of proper function of the CHM gene, characterized by two available EMS-derived mutant alleles *chm1-1*, *chm1-2* (Redei, *ibid.*) and a tissue culture-derived mutant allele *chm1-3* (Martinez-Zapater, et al., *ibid.*), results in rapid and specific copy number amplification of the subgenomic molecule, producing the consequent leaf variegation. It is not clear whether the copy number amplification or suppression of a single subgenomic molecule occurs by differential replication or a recombination mechanism.

SUMMARY OF THE INVENTION

[0010] The present invention provides an isolated nucleic acid molecule selected from the group consisting of: a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and SEQ ID NO:45; a nucleic acid molecule comprising at least a portion of any of these nucleic acid molecules; a complement of any of these nucleic acid molecules; and a nucleic acid molecule comprising an allelic variant of a nucleic acid molecule comprising any of these nucleic acid sequences.

[0011] In some embodiments, the nucleic acid molecule is a plant nucleic acid molecule, a nucleic acid molecule selected from the group consisting of *Arabidopsis*, *Oryza*, *Glycine*, *Hordeum*, *Zea*, *Medicago*, *Allium*, *Citrus*, *Solanum*, *Sorghum*, *Saccharum*, *Nicotiana*, *Lycopersicon*, *Triticum*, *Zinnia*, and *Phaseolus* nucleic acid molecules, a nucleic acid molecule selected from the group consisting of: a nucleic acid molecule comprising a nucleic acid sequence that encodes a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47, and SEQ ID NO:65; and a nucleic acid molecule comprising an allelic variant of a nucleic acid molecule encoding a protein having any of said amino acid sequences.

[0012] The present invention also provides an isolated MSH1 protein. In some embodiment, the protein is encoded by a plant MSH1 nucleic acid molecule that hybridizes to the complement of a nucleic acid molecule having a nucleic acid

sequence SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, or SEQ ID NO:45 under stringent hybridization conditions. In some embodiments, the protein is SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47 or SEQ ID NO:65, or a protein comprising at least a portion of an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47 and SEQ ID NO:65.

[0013] The present invention also provides a method to identify a compound capable of inhibiting MSH1 activity of a plant, said method comprising: contacting an isolated plant MSH1 nucleic acid molecule selected from the group consisting of SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and SEQ ID NO:45 with a putative inhibitory compound which, in the absence of said compound, said plant MSH1 nucleic acid molecule has the activity of suppressing ectopic recombination; and determining if said putative inhibitory compound inhibits said activity. In some embodiments, the putative inhibitory compound is a RNA molecule suspected of having RNAi activity. The invention also provides compounds identified by the method

[0014] Further provided is a method for identification of plant mutants arising from mitochondrial ectopic recombination comprising providing a plant, suppressing expression of an MSH1-homologous gene in the plant, and detecting an aberrant phenotype,

[0015] whereby a plant mutant is identified. In some embodiments, the suppression is effected by a compound identified by the above-described method. In some embodiments, the aberrant phenotype is cytoplasmic male sterility. The invention also provides plant mutants identified by the method of claim 12.

BRIEF DESCRIPTION OF THE FIGURES

[0016] FIG. 1. Positional cloning of the CHM candidate locus. The use of molecular markers permitted the establishment of a genetic map (A) and identification of the intervening overlapping bacterial artificial chromosome clones for physical mapping (B) All physical mapping information was derived from the *Arabidopsis* Genome Initiative (50). High resolution mapping with three markers permitted delimitation of the locus to a 80-kb interval contained within a single bacterial artificial chromosome

clone (C) A gene candidate was identified within the interval based on predicted mitochondrial targeting features. The candidate CHM locus contains 22 exons (D) with two MutS-like conserved intervals denoted by red lines. Analysis of two EMS-derived mutants, *chm1-1* and *chm1-2*, and one tissue culture-derived mutant *chm1-3*, as well as two TDNA insertion mutations (T1 and T2), provided definitive evidence of CHM identity (E). The numbers in parentheses in (A) correspond to the number of recombinants identified between the marker and the gene.

[0017] **FIG. 2.** Alignment of AtMSH1 with MutS and MutS homologs. The amino acid sequence alignment was performed using the ClustalW software and includes the MutS sequence from *E. coli*, MSH1 from *Saccharomyces cerevisiae*, and AtMSH6 and CHM (AtMSH1) from *Arabidopsis*. (A) Alignment of the region of the DNA-binding domain that encompasses the conserved motif for mismatch recognition and DNA binding. (B) Alignment of a portion of the ATPase domain. The characteristic motifs for this domain are indicated by red lines. M1—Walker motif; M2—ST motif; M3—DE motif (Walker B motif); M4—TH motif (Obmolova, et al. (2000) Nature 407, 703-710; Lamers, et al. (2000) Nature, 407, 711-717). The asterisks (*) indicate residues that are identical and the arrow indicates the site of amino acid substitution in mutant *chm1-3*.

[0018] **FIG. 3.** Alignment of MSH proteins.

DETAILED DESCRIPTION OF THE INVENTION

[0019] The present invention provides a plant nuclear gene and corresponding gene product, in *Arabidopsis thaliana* that influences mitochondrial genome organization. The gene is designated AtMSH1, and it is believed to suppress ectopic (illegitimate) recombination of the mitochondrial genome. The present invention provides for isolated MSH1 proteins, isolated MSH1 nucleic acid molecules, antibodies directed against MSH1 proteins and other inhibitors of MSH1 activity. As used herein, the terms isolated MSH1 proteins and isolated MSH1 nucleic acid molecules refers to MSH1 proteins and esterase nucleic acid molecules derived from plants and, as such, can be obtained from their natural source or can be produced using, for example, recombinant nucleic acid technology or chemical synthesis. The term “plant” refers to an individual living plant or population of same, a species, subspecies, variety, cultivar or strain. In some preferred embodiments, the domesticated organism is a plant selected from the group consisting of maize, wheat, rice, sorghum, tomato or potato, or any other domesticated plant of commercial interest. A “plant” is any plant at any stage of development, including a seed plant. Also included in the present invention is the use of these proteins, nucleic acid molecules, antibodies and inhibitors to generate transgenic plants, and mutant plants, as well as in other applications, such as those disclosed below.

[0020] The present invention is the result of studies investigating the unusual plant phenomenon of mitochondrial substoichiometric shifting and the role of the nuclear gene CHM. This gene, located on chromosome III, was shown to encode a protein that is targeted to mitochondria and that has homology to a yeast mitochondrial MutS protein. A summary of this investigation is provided in the EXAMPLES section.

[0021] MSH1 proteins and nucleic acid molecules of the present invention have utility because they represent novel targets for modulation which would effect mitochondrial ectopic recombination. The products and processes of the present invention are advantageous because they enable the express and inhibition of processes that involve MSH1. While not being bound by theory, it is believed these newly discovered proteins have contributed adaptive advantage by a strategy that may be unique to the Plant Kingdom.

[0022] A. MSH1 Polypeptides

[0023] One embodiment of the present invention is an isolated plant MSH1 polypeptide. As used herein, an MSH1 polypeptide, in one embodiment, is a polypeptide that is related to (i.e., bears structural similarity to) the *A. thaliana* polypeptide of about 1118 amino acids and having the sequence depicted in **FIG. 3** (SEQ ID NO: 3). The original identification of such a polypeptide is detailed in the Examples.

[0024] A preferred MSH1 polypeptide is encoded by a polynucleotide that hybridizes under stringent hybridization conditions to a gene encoding an MSH1 polypeptide (i.e., an *A. thaliana* gene). It is to be noted that the term “a” or “an” entity refers to one or more of that entity; for example, a gene refers to one or more genes or at least one gene. As such, the terms “a” (or “an”), “one or more” and “at least one” can be used interchangeably herein. It is also to be noted that the terms “comprising,” “including,” and “having” can be used interchangeably.

[0025] As used herein, stringent hybridization conditions refer to standard hybridization conditions under which polynucleotides, including oligonucleotides, are used to identify molecules having similar nucleic acid sequences. Such standard conditions are disclosed, for example, in Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Labs Press, 1989. Examples of such conditions are provided in the Examples section of the present application.

[0026] As used herein, an *A. thaliana* AtMSH1 gene includes all nucleic acid sequences related to a natural *A. thaliana* AtMSH1 gene such as regulatory regions that control production of the *A. thaliana* AtMSH1 polypeptide encoded by that gene (such as, but not limited to, transcription, translation or post-translation control regions) as well as the coding region itself. In one embodiment, an *A. thaliana* AtMSH1 gene includes the nucleic acid sequence SEQ ID NO:1. Nucleic acid sequence SEQ ID NO:X represents the deduced sequence of a cDNA (complementary DNA) polynucleotide, the production of which is disclosed in the Examples. It should be noted that since nucleic acid sequencing technology is not entirely error-free, SEQ ID NO:1 (as well as other sequences presented herein), at best, represents an apparent nucleic acid sequence of the polynucleotide encoding an *A. thaliana* AtMSH1 polypeptide of the present invention.

[0027] In another embodiment, an *A. thaliana* AtMSH1 gene can be an allelic variant that includes a similar but not identical sequence to SEQ ID NO:1. During higher plant evolution, natural allelic variation for the MSH1 locus likely revealed the adaptive advantage that arises from sporadic copy number modulation of mitochondrial genomic variants. Some of these variants, when amplified, condition male

sterility that could facilitate advantageous outcrossing activity in natural populations (Arrieta-Montiel, et al., *ibid.*). An allelic variant of an *A. thaliana* AtMSH1 gene including SEQ ID NO: 1 is a locus (or loci) in the genome whose activity is concerned with the same biochemical or developmental processes, and/or a gene that occurs at essentially the same locus as the gene including SEQ ID NO:1, but which, due to natural variations caused by, for example, mutation or recombination, has a similar but not identical sequence. Because genomes can undergo rearrangement, the physical arrangement of alleles is not always the same. Allelic variants typically encode polypeptides having similar activity to that of the polypeptide encoded by the gene to which they are being compared. Allelic variants can also comprise alterations in the 5' or 3' untranslated regions of the gene (e.g., in regulatory control regions). Allelic variants are well known to those skilled in the art and would be expected to be found within a given cultivar or strain since the genome is diploid and/or among a population comprising two or more cultivars or strains.

[0028] According to the present invention, an isolated, or biologically pure, polypeptide, is a polypeptide that has been removed from its natural milieu. As such, "isolated" and "biologically pure" do not necessarily reflect the extent to which the polypeptide has been purified. An isolated MSH1 polypeptide of the present invention can be obtained from its natural source, can be produced using recombinant DNA technology or can be produced by chemical synthesis. An MSH1 polypeptide of the present invention may be identified by its ability to perform the function of natural MSH1 in a functional assay. By "natural MSH1 polypeptide," it is meant the full length MSH1 polypeptide of *A. thaliana*. The phrase "capable of performing the function of a natural MSH1 in a functional assay" means that the polypeptide has at least about 10% of the activity of the natural polypeptide in the functional assay. In other embodiments, the MSH1 polypeptide has at least about 20% of the activity of the natural polypeptide in the functional assay. In other embodiments, the MSH1 polypeptide has at least about 30% of the activity of the natural polypeptide in the functional assay. In other embodiments, the MSH1 polypeptide has at least about 40% of the activity of the natural polypeptide in the functional assay. In other embodiments, the MSH1 polypeptide has at least about 50% of the activity of the natural polypeptide in the functional assay. In other embodiments, the polypeptide has at least about 60% of the activity of the natural polypeptide in the functional assay. In other embodiments, the polypeptide has at least about 70% of the activity of the natural polypeptide in the functional assay. In other embodiments, the polypeptide has at least about 80% of the activity of the natural polypeptide in the functional assay. In still other embodiments, the polypeptide has at least about 90% of the activity of the natural polypeptide in the functional assay. Examples of functional assays are detailed elsewhere in this specification.

[0029] As used herein, an isolated plant MSH1 polypeptide can be a full-length polypeptide or any homologue of such a polypeptide. Examples of MSH1 homologues include MSH1 polypeptides in which amino acids have been deleted (e.g., a truncated version of the polypeptide, such as a peptide), inserted, inverted, substituted and/or derivatized (e.g., by glycosylation, phosphorylation, acetylation, myristylation, prenylation, palmitoylation, amidation and/or

addition of glycerophosphatidyl inositol) such that the homologue has natural MSH1 activity.

[0030] In one embodiment, when the homologue is administered to an animal as an immunogen, using techniques known to those skilled in the art, the animal will produce a humoral and/or cellular immune response against at least one epitope of a natural MSH1 polypeptide. MSH1 homologues can also be selected by their ability to perform the function of MSH1 in a functional assay.

[0031] Plant MSH1 polypeptide homologues can be the result of natural allelic variation or natural mutation. MSH1 polypeptide homologues of the present invention can also be produced using techniques known in the art including, but not limited to, direct modifications to the polypeptide or modifications to the gene encoding the polypeptide using, for example, classic or recombinant DNA techniques to effect random or targeted mutagenesis.

[0032] In accordance with the present invention, a mimetope refers to any compound that is able to mimic the ability of an isolated plant MSH1 polypeptide of the present invention to perform the function of an MSH1 polypeptide of the present invention in a functional assay. Examples of mimetopes include, but are not limited to, anti-idiotypic antibodies or fragments thereof, that include at least one binding site that mimics one or more epitopes of an isolated polypeptide of the present invention; non-polypeptideaceous immunogenic portions of an isolated polypeptide (e.g., carbohydrate structures); and synthetic or natural organic molecules, including nucleic acids, that have a structure similar to at least one epitope of an isolated polypeptide of the present invention. Such mimetopes can be designed using computer-generated structures of polypeptides of the present invention. Mimetopes can also be obtained by generating random samples of molecules, such as oligonucleotides, peptides or other organic molecules, and screening such samples by affinity chromatography techniques using the corresponding binding partner.

[0033] The minimal size of an MSH1 polypeptide homologue of the present invention is a size sufficient to be encoded by a polynucleotide capable of forming a stable hybrid with the complementary sequence of a polynucleotide encoding the corresponding natural polypeptide. As such, the size of the polynucleotide encoding such a polypeptide homologue is dependent on nucleic acid composition and percent homology between the polynucleotide and complementary sequence as well as upon hybridization conditions *per se* (e.g., temperature, salt concentration, and formamide concentration). It should also be noted that the extent of homology required to form a stable hybrid can vary depending on whether the homologous sequences are interspersed throughout the polynucleotides or are clustered (i.e., localized) in distinct regions on the polynucleotides. The minimal size of such polynucleotides is typically at least about 12 to about 15 nucleotides in length if the polynucleotides are GC-rich and at least about 15 to about 17 bases in length if they are AT-rich. Preferably, the polynucleotide is at least 12 bases in length.

[0034] As such, the minimal size of a polynucleotide used to encode an MSH1 polypeptide homologue of the present invention is from about 12 to about 18 nucleotides in length. There is no limit, other than a practical limit, on the maximal size of such a polynucleotide in that the polynucleotide can

include a portion of a gene, an entire gene, or multiple genes, or portions thereof. Similarly, the minimal size of an MSH1 polypeptide homologue of the present invention is from about 4 to about 6 amino acids in length, with preferred sizes depending on whether a full-length, fusion, multivalent, or functional portions of such polypeptides are desired. Preferably, the polypeptide is at least 30 bases in length.

[0035] Any plant MSH1 polypeptide is a suitable polypeptide of the present invention. Suitable plants from which to isolate MSH1 polypeptides (including isolation of the natural polypeptide or production of the polypeptide by recombinant or synthetic techniques) include maize, wheat, barley, rye, millet, chickpea, lentil, flax, olive, fig almond, pistachio, walnut, beet, parsnip, citrus fruits, including, but not limited to, orange, lemon, lime, grapefruit, tangerine, minneola, and tangelo, sweet potato, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, pepper, celery, squash, pumpkin, hemp, zucchini, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tomato, sorghum, sugarcane, sugarbeet, sunflower, rapeseed, clover, tobacco, carrot, cotton, alfalfa, rice, potato, eggplant, cucumber, *Arabidopsis*, and woody plants such as coniferous and deciduous trees, with soybean, tomato, potato, rice, wheat, and barley being preferred.

[0036] A preferred plant MSH1 polypeptide of the present invention is a compound that when expressed or modulated in a plant, is capable of suppressing ectopic recombination of the mitochondrial genome.

[0037] One embodiment of the present invention is a fusion polypeptide that includes an MSH1 polypeptide-containing domain attached to a fusion segment. Inclusion of a fusion segment as part of a MSH1 polypeptide of the present invention can enhance the polypeptide's stability during production, storage and/or use. Depending on the segment's characteristics, a fusion segment can also act as an immunopotentiator to enhance the immune response mounted by an animal immunized with an MSH1 polypeptide containing such a fusion segment. Furthermore, a fusion segment can function as a tool to simplify purification of an MSH1 polypeptide, such as to enable purification of the resultant fusion polypeptide using affinity chromatography. A suitable fusion segment can be a domain of any size that has the desired function (e.g., imparts increased stability, imparts increased immunogenicity to a polypeptide, and/or simplifies purification of a polypeptide). It is within the scope of the present invention to use one or more fusion segments. Fusion segments can be joined to amino and/or carboxyl termini of the MSH1-containing domain of the polypeptide. Linkages between fusion segments and MSH1-containing domains of fusion polypeptides can be susceptible to cleavage in order to enable straightforward recovery of the MSH1-containing domains of such polypeptides. Fusion polypeptides are preferably produced by culturing a recombinant cell transformed with a fusion polynucleotide that encodes a polypeptide including the fusion segment attached to either the carboxyl and/or amino terminal end of a MSH1-containing domain.

[0038] Exemplary fusion segments for use in the present invention include a glutathione binding domain; a metal binding domain, such as a poly-histidine segment capable of

binding to a divalent metal ion; an immunoglobulin binding domain, such as Polypeptide A, Polypeptide G, T cell, B cell, Fc receptor or complement polypeptide antibody-binding domains; a sugar binding domain such as a maltose binding domain from a maltose binding polypeptide; and/or a "tag" domain (e.g., at least a portion of β -galactosidase, a strep tag peptide, other domains that can be purified using compounds that bind to the domain, such as monoclonal antibodies). Other fusion segments suitable for use in the invention include metal binding domains, such as a poly-histidine segment; a maltose binding domain; a strep tag peptide.

[0039] Preferred plant MSH1 polypeptides of the present invention are *Arabidopsis* MSH1 polypeptides, soybean MSH1 polypeptides, tomato MSH1 polypeptides, rice MSH1 polypeptides, and common bean MSH1 polypeptides. Other preferred plant MSH polypeptides include corn MSH1 polypeptides, wheat MSH1 polypeptides, sugar cane MSH1 polypeptides, medicago MSH1 polypeptides, onion MSH1 polypeptides, orange MSH1 polypeptides, zinnia MSH1 polypeptides, tobacco MSH1 polypeptides, and barley MSH1 polypeptides.

[0040] One preferred *A. thaliana* AtMSH1 polypeptide of the present invention is a polypeptide encoded by an *A. thaliana* polynucleotide that hybridizes under stringent hybridization conditions with complements of polynucleotides represented by SEQ ID NO:1. Such an AtMSH1 polypeptide is encoded by a polynucleotide that hybridizes under stringent hybridization conditions with a polynucleotide having nucleic acid sequence SEQ ID NO:1.

[0041] Inspection of AtMSH1 genomic nucleic acid sequences indicates that the genes comprise several regions, including an ATP-binding domain, comprised of four well conserved motifs designated M1-M4 (Obmolova, et al., *ibid.*; **FIG. 2B**), and a DNA binding domain (aa 129-206) containing the aromatic doublet (FY) motif.

[0042] Translation of SEQ ID NO:1 suggests that the *A. thaliana* AtMSH1 polynucleotide includes an open reading frame. The reading frame encodes an *A. thaliana* AtMSH1 polypeptide of about 1118 amino acids, the deduced amino acid sequence of which is represented herein as SEQ ID NO:3, assuming an open reading frame having an initiation (start) codon spanning from about nucleotide 124 through about nucleotide 126 of SEQ ID NO:1 and a termination (stop) codon spanning from about nucleotide 3478 through about nucleotide 3480 of SEQ ID NO:1.

[0043] Similarly, translation of SEQ ID NO:20 suggests that the *Oryza sativa* MSH1 polynucleotide includes an open reading frame. The reading frame encodes an *Oryza sativa* MSH polypeptide of about 1132 amino acids, the deduced amino acid sequence of which is represented herein as SEQ ID NO:22, assuming an open reading frame having an initiation (start) codon spanning from about nucleotide 1 through about nucleotide 3 of SEQ ID NO:22 and a termination (stop) codon spanning from about nucleotide 3394 through about nucleotide 3396 of SEQ ID NO:20.

[0044] Similarly, translation of SEQ ID NO:29 suggests that the *Glycine max* MSH1 polynucleotide includes an open reading frame. The reading frame encodes an *Glycine max* MSH polypeptide of about 1130 amino acids, the deduced amino acid sequence of which is represented herein as SEQ ID NO:31, assuming an open reading frame having an

initiation (start) codon spanning from about nucleotide 1 through about nucleotide 3 of SEQ ID NO:29 and a termination (stop) codon spanning from about nucleotide 3391 through about nucleotide 3393 of SEQ ID NO:20.

[0045] Similarly, translation of SEQ ID NO:38 suggests that the *Lycopersicon esculentum* MSH1 polynucleotide includes an open reading frame. The reading frame encodes an *Lycopersicon esculentum* MSH polypeptide of about 1124 amino acids, the deduced amino acid sequence of which is represented herein as SEQ ID NO:40, assuming an open reading frame having an initiation (start) codon spanning from about nucleotide 1 through about nucleotide 3 of SEQ ID NO:38 and a termination (stop) codon spanning from about nucleotide 3369 through about nucleotide 3371 of SEQ ID NO:20.

[0046] Similarly, translation of SEQ ID NO:45 suggests that the *Phaseolus vulgaris* MSH1 polynucleotide includes an open reading frame. The reading frame encodes an *Phaseolus vulgaris* MSH polypeptide of about 1126 amino acids, the deduced amino acid sequence of which is represented herein as SEQ ID NO:47, assuming an open reading frame having an initiation (start) codon spanning from about nucleotide 1 through about nucleotide 3 of SEQ ID NO:45 and a termination (stop) codon spanning from about nucleotide 3379 through about nucleotide 3381 of SEQ ID NO:20.

[0047] Additional EST sequences having at least 60% sequence identity to a portion of SEQ ID NO:1 or a complement of SEQ ID NO:1 have been found. These include MSH1 polynucleotides from corn (SEQ ID NO:11), potato (SEQ ID NO:18), wheat (SEQ ID NO:41), sugar cane (SEQ ID NO:32 and SEQ ID NO:34), medicago (SEQ ID NO:13), onion (SEQ ID NO:14), orange (SEQ ID NO:16), zinnia (SEQ ID NO:43), tobacco (SEQ ID NO:36), and barley (SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10). Polypeptides encoded by the foregoing nucleic acid molecules can be deduced using methods well known in the art. In general, the polynucleotide or its complement is aligned with the *Arabidopsis* AtMSH1 polynucleotide, a reading frame is determined, and the resulting polypeptide sequence is translated. Polypeptides encoded by the foregoing nucleic acid molecules or their complements include corn (SEQ ID NO:12), potato (SEQ ID NO:19), wheat (SEQ ID NO:42), sugar cane (SEQ ID NO:33 and SEQ ID NO:35), onion (SEQ ID NO:15), orange (SEQ ID NO:17), zinnia (SEQ ID NO:44), and barley (SEQ ID NO:7, SEQ ID NO:9), and consensus (SEQ ID NO:65).

[0048] Comparison of the various *A. thaliana*, soybean, corn, tomato, potato, rice, wheat, common bean, sugar cane, medicago, onion, orange, zinnia, tobacco, and barley MSH1 nucleic acid sequences and amino acid sequences described herein indicates that these species of plants possess similar MSH1 genes and polypeptides. The nucleotide sequences of the coding region of MSH1 from the various plants have >60% sequence identity when compared to each other, which makes clear that they are homologous.

[0049] Finding this degree of identity between soybean, corn, tomato, potato, rice, wheat, common bean, sugar cane, medicago, onion, orange, zinnia, tobacco, and barley MSH1 nucleic acid sequences and amino acid sequences supports the ability to obtain any plant MSH1 polypeptide and polynucleotide given the polypeptide and nucleic acid sequences disclosed herein.

[0050] These plant MSH1 polypeptides, and the polynucleotides that encode them, represent novel compounds with utility in ectopic recombination of the mitochondrial genome.

[0051] Preferred plant MSH1 polypeptides of the present invention include polypeptides comprising amino acid sequences that are at least about 30%, preferably at least about 50%, more preferably at least about 75% and even more preferably at least about 90% identical to one or more of the amino acid sequences disclosed herein for *A. thaliana* AtMSH1 polypeptides of the present invention. More preferred plant MSH1 polypeptides of the present invention include: polypeptides encoded by at least a portion of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:29, SEQ ID NO:38 and/or SEQ ID NO:45 and, as such, have amino acid sequences that include at least a portion of SEQ ID NO:3, SEQ ID NO:22, SEQ ID NO:31, SEQ ID NO:40 and/or SEQ ID NO:47; polypeptides encoded by at least a portion of SEQ ID NO:1, SEQ ID NO:20, SEQ ID NO:29, SEQ ID NO:38 and/or SEQ ID NO:45 and, as such, have amino acid sequences that include at least a portion of SEQ ID NO:3, SEQ ID NO:22, SEQ ID NO:31, SEQ ID NO:40 and/or SEQ ID NO:47. Also preferred are polypeptides that have amino acid sequences that include at least a portion of SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:42, and/or SEQ ID NO:44; and polypeptides encoded by at least a portion of SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:41, and/or SEQ ID NO:43, or a complement of any of the foregoing SEQ ID NO:s. As used herein, "at least a portion" of a polynucleotide or polypeptide means a portion having the minimal size characteristics of such sequences, as described above, or any larger fragment of the full length molecule, up to and including the full length molecule. For example, a portion of a polynucleotide may be 12 nucleotides, 13 nucleotides, 14 nucleotides, 15 nucleotides, and so on, going up to the full length polynucleotide. Similarly, a portion of a polypeptide may be 4 amino acids, 5 amino acids, 6 amino acids, 7 amino acids, and so on, going up to the full length polypeptide. The length of the portion to be used will depend on the particular application. As discussed above, a portion of a polynucleotide useful as hybridization probe may be as short as 12 nucleotides. A portion of a polypeptide useful as an epitope may be as short as 4 amino acids. A portion of a polypeptide that performs the function of the full-length polypeptide would generally be longer than 4 amino acids.

[0052] Particularly preferred plant MSH1 polypeptides of the present invention are polypeptides that include SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47 and/or SEQ ID NO:65 (including, but not limited to the encoded polypeptides, full-length polypeptides, processed polypeptides, fusion polypeptides and multivalent polypeptides thereof) as well as polypeptides that are truncated homologues of polypeptides that include at least portions of the aforemen-

tioned SEQ ID NOs. Examples of methods to produce such polypeptides are disclosed herein, including in the Examples section.

[0053] Plant MSH1 polypeptides may have DNA binding and ATPase activities. Identification of the chm1-3 mutation as a cysteine-tyrosine substitution within the predicted ATP binding domain does suggest the importance of this region to protein function. Substitution of the bulkier tyrosine would likely create distortion in the region, affecting ATP binding or hydrolysis.

[0054] Mismatch repair components appear to be involved in not only the binding and excision of nucleotide mismatches during the replication process, but also suppression of ectopic recombination (Harfe & Jinks-Robertson (2000) *Annu. Rev. Genet.* 34, 359-399; Chen & Jinks-Robertson (1999) *Genetics* 151,1299-1313). Investigation of the mitochondrial substoichiometric shifting phenomenon suggests two alternative models for the influence of MSH1. It is conceivable that the MSH1 gene has shared or relinquished its mismatch repair function, such that its primary role in the plant mitochondrial genome is to regulate non-homologous recombination. Disruption of MSH1 could, thus, result in the enhancement of intra-molecular ectopic recombination activity detected as apparent amplification of novel mitochondrial DNA forms. A possible weakness in this model arises in reports that several plant systems with mitochondrial DNA molecules susceptible to shifting appear to be derived from a DNA exchange that involved at least one molecular form no longer present in high copy number. Some also appeared to contain unique sequences. Therefore, the shifted molecules were thought to replicate autonomously (Andre, et al., *ibid*; Kanazawa, et al., *ibid*; Janska & Mackenzie (1993) *Genetics* 135, 869-879).

[0055] If mitochondrial DNA molecules that undergo shifting are, in fact, replicated autonomously, an alternative model for molecule-specific substoichiometric shifting might apply. The *Arabidopsis* MSH1 product likely participates as a component of the DNA replication apparatus. Mitochondrial DNA molecules subject to copy number shifting may have originated by earlier ectopic recombination events during the evolution of the lineage. In this case, the resulting chimeric sites might serve to trigger a process of site-specific replication stalling by the MSH1 protein during vegetative growth.

[0056] Both models assume that the replicative form of the mitochondrial genome within meristematic (undifferentiated) tissues differs from that of vegetative (somatic). Hence, stoichiometric shifting events in vegetative tissues do not condition irreversible loss of the suppressed genetic information. Presumably, the complete mitochondrial genetic complement is retained within the transmitting (meristematic) tissues (Arrieta-Montiel, et al., Janska & Mackenzie, *ibid*).

[0057] B. MSH1 Polynucleotides

[0058] One embodiment of the present invention is an isolated plant polynucleotide that hybridizes under stringent hybridization conditions with an *A. thaliana* AtMSH1 gene. The identifying characteristics of such genes are heretofore described. A polynucleotide of the present invention can include an isolated natural plant MSH1 gene or a homologue thereof, the latter of which is described in more detail below.

A polynucleotide of the present invention can include one or more regulatory regions, full-length or partial coding regions, or combinations thereof. The minimal size of a polynucleotide of the present invention is the minimal size that can form a stable hybrid with one of the aforementioned genes under stringent hybridization conditions. Suitable and preferred plants are disclosed above.

[0059] In accordance with the present invention, an isolated polynucleotide is a polynucleotide that has been removed from its natural milieu (i.e., that has been subject to human manipulation). As such, "isolated" does not reflect the extent to which the polynucleotide has been purified. An isolated polynucleotide can include DNA, RNA, or derivatives of either DNA or RNA.

[0060] An isolated plant MSH1 polynucleotide of the present invention can be obtained from its natural source either as an entire (i.e., complete) gene or a portion thereof capable of forming a stable hybrid with that gene. An isolated plant MSH1 polynucleotide can also be produced using recombinant DNA technology (e.g., polymerase chain reaction (PCR) amplification, cloning) or chemical synthesis. Isolated plant MSH1 polynucleotides include natural polynucleotides and homologues thereof, including, but not limited to, natural allelic variants and modified polynucleotides in which nucleotides have been inserted, deleted, substituted, and/or inverted in such a manner that such modifications do not substantially interfere with the polynucleotide's ability to encode an MSH1 polypeptide of the present invention or to form stable hybrids under stringent conditions with natural gene isolates.

[0061] A plant MSH1 polynucleotide homologue can be produced using a number of methods known to those skilled in the art (see, for example, Sambrook et al., *ibid*). For example, polynucleotides can be modified using a variety of techniques including, but not limited to, classic mutagenesis techniques and recombinant DNA techniques, such as site-directed mutagenesis, chemical treatment of a polynucleotide to induce mutations, restriction enzyme cleavage of a nucleic acid fragment, ligation of nucleic acid fragments, polymerase chain reaction (PCR) amplification and/or mutagenesis of selected regions of a nucleic acid sequence, synthesis of oligonucleotide mixtures and ligation of mixture groups to "build" a mixture of polynucleotides and combinations thereof. Polynucleotide homologues can be selected from a mixture of modified nucleic acids by screening for the function of the polypeptide encoded by the nucleic acid (e.g., ability to elicit an immune response against at least one epitope of an MSH1 polypeptide, ability to suppress ectopic recombination in a transgenic plant containing an MSH1 gene and/or by hybridization with an *A. thaliana* AtMSH1 gene).

[0062] An isolated polynucleotide of the present invention can include a nucleic acid sequence that encodes at least one plant MSH1 polypeptide of the present invention, examples of such polypeptides being disclosed herein. Although the phrase "polynucleotide" primarily refers to the physical polynucleotide and the phrase "nucleic acid sequence" primarily refers to the sequence of nucleotides on the polynucleotide, the two phrases can be used interchangeably, especially with respect to a polynucleotide, or a nucleic acid sequence, being capable of encoding an MSH1 polypeptide. As heretofore disclosed, plant MSH1 polypeptides of the

present invention include, but are not limited to, polypeptides having full-length plant MSH1 coding regions, polypeptides having partial plant MSH1 coding regions, fusion polypeptides, multivalent protective polypeptides and combinations thereof.

[0063] At least certain polynucleotides of the present invention encode polypeptides that selectively bind to immune serum derived from an animal that has been immunized with an MSH1 polypeptide from which the polynucleotide was isolated.

[0064] A preferred polynucleotide of the present invention, when suppressed in a suitable plant, is capable of generating economically useful mutant plants. As will be disclosed in more detail below, such a polynucleotide can be, or encode, an antisense RNA, a molecule capable of triple helix formation, a ribozyme, or other nucleic acid-based compound.

[0065] One embodiment of the present invention is a plant MSH1 polynucleotide that hybridizes under stringent hybridization conditions to an MSH1 polynucleotide of the present invention, or to a homologue of such an MSH1 polynucleotide, or to the complement of such a polynucleotide. A polynucleotide complement of any nucleic acid sequence of the present invention refers to the nucleic acid sequence of the polynucleotide that is complementary to (i.e., can form a complete double helix with) the strand for which the sequence is cited. It is to be noted that a double-stranded nucleic acid molecule of the present invention for which a nucleic acid sequence has been determined for one strand, that is represented by a SEQ ID NO, also comprises a complementary strand having a sequence that is a complement of that SEQ ID NO. As such, polynucleotides of the present invention, which can be either double-stranded or single-stranded, include those polynucleotides that form stable hybrids under stringent hybridization conditions with either a given SEQ ID NO denoted herein and/or with the complement of that SEQ ID NO, which may or may not be denoted herein. Methods to deduce a complementary sequence are known to those skilled in the art. Preferred is an MSH1 polynucleotide that includes a nucleic acid sequence having at least about 60 percent, at least about 65 percent, preferably at least about 70 percent, more preferably at least about 75 percent, more preferably at least about 80 percent, more preferably at least about 85 percent, more preferably at least about 90 percent and even more preferably at least about 95 percent homology with the corresponding region(s) of the nucleic acid sequence encoding at least a portion of an MSH1 polypeptide. Particularly preferred is an MSH1 polynucleotide capable of encoding at least a portion of an MSH1 polypeptide that naturally is present in plants.

[0066] Particularly preferred MSH1 polynucleotides of the present invention hybridize under stringent hybridization conditions with at least one of the following polynucleotides: SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and/or SEQ ID NO:45, or to a homologue or complement of such polynucleotide.

[0067] A preferred polynucleotide of the present invention includes at least a portion of nucleic acid sequence SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and/or SEQ ID NO:45 that is capable of hybridizing (i.e., that hybridizes under stringent hybridization conditions) to an *A. thaliana* AtMSH1 gene of the present invention, as well as a polynucleotide that is an allelic variant of any of those polynucleotides. Such preferred polynucleotides can include nucleotides in addition to those included in the SEQ ID NOs, such as, but not limited to, a full-length gene, a full-length coding region, a polynucleotide encoding a fusion polypeptide, and/or a polynucleotide encoding a multivalent protective compound.

[0068] The present invention also includes polynucleotides encoding a polypeptide including at least a portion of SEQ ID NO:3, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:7, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:9, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:12, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:15, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:17, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:19, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:22, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:24, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:26, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:31, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:33, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:35, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:40, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:42, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:42, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:44, polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:47, and/or polynucleotides encoding a polypeptide having at least a portion of SEQ ID NO:65, including polynucleotides that have been modified to accommodate codon usage properties of the cells in which such polynucleotides are to be expressed.

[0069] Knowing the nucleic acid sequences of certain plant MSH1 polynucleotides of the present invention allows one skilled in the art to, for example, (a) make copies of those polynucleotides, (b) obtain polynucleotides including at least a portion of such polynucleotides (e.g., polynucleotides including full-length genes, full-length coding regions, regulatory control sequences, truncated coding regions), and (c) obtain MSH1 polynucleotides for other plants. Such polynucleotides can be obtained in a variety of ways including screening appropriate expression libraries with antibodies of the present invention; traditional cloning techniques using oligonucleotide probes of the present invention to screen appropriate libraries or DNA; and PCR amplification of appropriate libraries or DNA using oligo-

nucleotide primers of the present invention. Preferred libraries to screen or from which to amplify polynucleotides include libraries such as genomic DNA libraries, BAC libraries, YAC libraries, cDNA libraries prepared from isolated plant tissues, including, but not limited to, stems, reproductive structures/tissues, leaves, roots, and tillers; and libraries constructed from pooled cDNAs from any or all of the tissues listed above. In the case of rice, BAC libraries, available from Clemson University, are preferred. Similarly, preferred DNA sources to screen or from which to amplify polynucleotides include plant genomic DNA. Techniques to clone and amplify genes are disclosed, for example, in Sambrook et al., *ibid.* and in Galun & Breiman, *TRANS-GENIC PLANTS*, Imperial College Press, 1997.

[0070] The present invention also includes polynucleotides that are oligonucleotides capable of hybridizing, under stringent hybridization conditions, with complementary regions of other, preferably longer, polynucleotides of the present invention such as those comprising plant MSH1 genes or other plant MSH1 polynucleotides. Oligonucleotides of the present invention can be RNA, DNA, or derivatives of either. The minimal size of such oligonucleotides is the size required to form a stable hybrid between a given oligonucleotide and the complementary sequence on another polynucleotide of the present invention. Minimal size characteristics are disclosed herein. The size of the oligonucleotide must also be sufficient for the use of the oligonucleotide in accordance with the present invention. Oligonucleotides of the present invention can be used in a variety of applications including, but not limited to, as probes to identify additional polynucleotides, as primers to amplify or extend polynucleotides, as targets for expression analysis, as candidates for targeted mutagenesis and/or recovery, or in agricultural applications to alter MSH1 polypeptide production or activity. Such agricultural applications include the use of such oligonucleotides in, for example, antisense-, triplex formation-, ribozyme- and/or RNA drug-based technologies. The present invention, therefore, includes such oligonucleotides and methods in a plant by use of one or more of such technologies.

[0071] The predicted features of the candidate CHM-encoded protein denoted MSH1 suggest that the gene encodes the mitochondrial MSH1 counterpart in higher plants. MSH1 encodes a mitochondrial mismatch repair protein in yeast, though its counterpart in animals has not yet been identified. The CHM candidate sequence showed strongest homology with the *Arabidopsis* nuclear MSH6 sequence (**FIG. 2**), consistent with suggestions that nuclear mismatch repair components likely derived from a progenitor to MSH1 (Culligan, et al. (2000) *Nucl. Acids Res.* 28, 463-471).

[0072] Although the predicted CHM candidate protein displayed several features suggesting its involvement in mismatch repair, lines containing mutations in the locus showed no evidence of mitochondrial point mutation accumulation. The primary effect within the mitochondrion appeared to be the reproducible substoichiometric shifting phenomenon. This assumption is based on the observation of identical mitochondrial DNA restriction fragments arising upon substoichiometric shifting in all *chm* mutants when tested repeatedly (Sakamoto, et al., *ibid.*, Martinez-Zapater, et al., *ibid.*, this report). Moreover, no evidence of progressive decline in plant growth features has been observed over

time. The *chm1-1* and *chm1-2* mutants, reported in the 1970's (Redei, *ibid.*), appear identical to one another in phenotype and mitochondrial DNA configuration. Although detailed sequence analysis would be required to estimate the incidence of mismatch accumulation in the *chm* mutants, one would anticipate a random pattern of mitochondrial DNA polymorphism and progressive phenotypic decline in *chm* mutants were the mismatch accumulation rate enhanced.

[0073] Mutation of the MSH1 locus in yeast results in rapid accumulation of mitochondrial genomic rearrangements leading to disruption of mitochondrial function. Interestingly, a reproducible pattern of DNA restriction fragment polymorphism was reported in some of the *petit* mutants arising in yeast MSH1 mutant strains (Reenan & Kolodner). This observation may be indication that *msh1*-associated mitochondrial genomic rearrangements are similar in plants and fungi. Alignment between the yeast MSH1 protein and the *Arabidopsis* CHM (MSH1) candidate shows only 17% amino acid identity overall, with ca. 28% identity within the predicted functional domains for ATP and DNA binding, but with well conserved motifs (**FIG. 2**). The yeast MSH1 protein has been shown to have both DNA mismatch binding and ATPase activity (Chi & Kolodner (1994) *J Biol. Chem.* 269,29984-29992; Chi & Kolodner. (1994) *J. Biol. Chem.* 269, 29993-29997).

[0074] C. Recombinant Molecules

[0075] The present invention also includes a recombinant vector, which includes at least one plant MSH1 polynucleotide of the present invention, inserted into any vector capable of delivering the polynucleotide into a host cell. Such a vector contains heterologous nucleic acid sequences, that is nucleic acid sequences that are not naturally found adjacent to polynucleotides of the present invention and that preferably are derived from a species other than the species from which the polynucleotide(s) are derived. As used herein, a derived polynucleotide is one that is identical or similar in sequence to a polynucleotide or portion of a polynucleotide, but can contain modifications, such as modified bases, backbone modifications, nucleotide changes, and the like. The vector can be either RNA or DNA, either prokaryotic or eukaryotic, and typically is a virus or a plasmid. Recombinant vectors can be used in the cloning, sequencing, and/or otherwise manipulating of plant MSH1 polynucleotides of the present invention. One type of recombinant vector, referred to herein as a recombinant molecule and described in more detail below, can be used in the expression of polynucleotides of the present invention. Preferred recombinant vectors are capable of replicating in the transformed cell.

[0076] Suitable and preferred polynucleotides to include in recombinant vectors of the present invention are as disclosed herein for suitable and preferred plant MSH1 polynucleotides *per se*. Particularly preferred polynucleotides to include in recombinant vectors, and particularly in recombinant molecules, of the present invention include SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and/or SEQ ID NO:45.

[0077] Isolated plant MSH1 polypeptides of the present invention can be produced in a variety of ways, including production and recovery of natural polypeptides, production and recovery of recombinant polypeptides, and chemical synthesis of the polypeptides. In one embodiment, an isolated polypeptide of the present invention is produced by culturing a cell capable of expressing the polypeptide under conditions effective to produce the polypeptide, and recovering the polypeptide. A preferred cell to culture is a recombinant cell that is capable of expressing the polypeptide, the recombinant cell being produced by transforming a host cell with one or more polynucleotides of the present invention. Transformation of a polynucleotide into a cell can be accomplished by any method by which a polynucleotide can be inserted into the cell. Transformation techniques include, but are not limited to, transfection, electroporation, microinjection, lipofection, adsorption, and protoplast fusion. A recombinant cell may remain unicellular or may grow into a tissue, organ or a multicellular organism. Transformed polynucleotides of the present invention can remain extrachromosomal or can integrate into one or more sites within a chromosome of the transformed (i.e., recombinant) cell in such a manner that their ability to be expressed is retained. Suitable and preferred polynucleotides with which to transform a cell are as disclosed herein for suitable and preferred plant MSH1 polynucleotides per se. Particularly preferred polynucleotides to include in recombinant cells of the present invention include SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and/or SEQ ID NO:45.

[0078] Suitable host cells to transform include any cell that can be transformed with a polynucleotide of the present invention. Host cells can be either untransformed cells or cells that are already transformed with at least one polynucleotide. Host cells of the present invention either can be endogenously (i.e., naturally) capable of producing plant MSH1 polypeptides of the present invention or can be capable of producing such polypeptides after being transformed with at least one polynucleotide of the present invention. Host cells of the present invention can be any cell capable of producing at least one polypeptide of the present invention, and include bacterial, fungal (including yeast and rice blast, *Magnaporthe grisea*), parasite (including nematodes, especially of the genera *Xiphinema*, *Helicotylenchus*, and *Tylenchlohyinchus*), insect, other animal and plant cells.

[0079] Suitable host viruses to transform include any virus that can be transformed with a polynucleotide of the present invention, including, but not limited to, rice stripe virus, and echinocloa hoja blanca virus.

[0080] In a preferred embodiment, non-pathogenic symbiotic bacteria, which are able to live and replicate within plant tissues, so-called endophytes, or non-pathogenic symbiotic bacteria, which are capable of colonizing the phyllosphere or the rhizosphere, so-called epiphytes, are used. Such bacteria include bacteria of the genera *Agrobacterium*, *Alcaligenes*, *Azospirillum*, *Azotobacter*, *Bacillus*, *Clavibacter*, *Enterobacter*, *Erwinia*, *Flavobacter*, *Klebsiella*, *Pseudomonas*, *Rhizobium*, *Serratia*, *Streptomyces* and *Xan-*

thomonas. Symbiotic fungi, such as *Trichoderma* and *Gliocladium* are also possible hosts for expression of the inventive nucleotide sequences for the same purpose.

[0081] A recombinant cell is preferably produced by transforming a host cell with one or more recombinant molecules, each comprising one or more polynucleotides of the present invention operatively linked to an expression vector containing one or more transcription control sequences. The phrase "operatively linked" refers to insertion of a polynucleotide into an expression vector in a manner such that the molecule is able to be expressed in the correct reading frame when transformed into a host cell. As used herein, an expression vector is a DNA or RNA vector that is capable of transforming a host cell and of effecting expression of a specified polynucleotide. Preferably, the expression vector is also capable of replicating within the host cell. Expression vectors can be either prokaryotic or eukaryotic, and are typically viruses or plasmids. Expression vectors of the present invention include any vectors that function (i.e., direct gene expression) in recombinant cells of the present invention, including in bacterial, fungal, parasite, insect, other animal, and plant cells. Preferred expression vectors of the present invention can direct gene expression in bacterial, yeast, fungal, insect and mammalian cells and more preferably in the cell types heretofore disclosed.

[0082] Recombinant molecules of the present invention may also (a) contain secretory signals (i.e., signal segment nucleic acid sequences) to enable an expressed MSH1 polypeptide of the present invention to be secreted from the cell that produces the polypeptide and/or (b) contain fusion sequences which lead to the expression of polynucleotides of the present invention as fusion polypeptides. Examples of suitable signal segments and fusion segments encoded by fusion segment nucleic acids are disclosed herein. Eukaryotic recombinant molecules may include intervening and/or untranslated sequences surrounding and/or within the nucleic acid sequences of polynucleotides of the present invention. Suitable signal segments include natural signal segments or any heterologous signal segment capable of directing the secretion of a polypeptide of the present invention. Preferred signal and fusion sequences employed to enhance organ and organelle specific expression include, but are not limited to, arcelin-5, see Goossens, A. et. al. The arcelin-5 Gene of *Phaseolus vulgaris* directs high seed-specific expression in transgenic *Phaseolus acutifolius* and *Arabidopsis* plants. Plant Physiology (1999) 120:1095-1104, phaseolin, see Sengupta-Gopalan, C. et. al. Developmentally regulated expression of the bean beta-phaseolin gene in tobacco seeds. PNAS (1985) 82:3320-3324, hydroxyproline-rich glycoprotein, serpin, see Yan, X. et. al. Gene fusions of signal sequences with a modified beta-glucuronidase gene results in retention of the beta-glucuronidase protein in the secretory pathway/plasma membrane. Plant Physiology (1997) 115:915-924, N-acetyl glucosaminyl transferase 1, see Essi, D. et. al. The N-terminal 77 amino acids from tobacco N-acetylglucosaminyltransferase I are sufficient to retain reporter protein in the Golgi apparatus of *Nicotiana benthamiana* cells. Febs Letters (1999) 453(1-2):169-73, albumin, see Vandekerckhove, J. et. al. Enkephalins produced in transgenic plants using modified 2S seed storage proteins. BioTechnology 7:929-932 (1989) and PR1, see Pen, J. et. al. Efficient production of active industrial enzymes in plants. Industrial Crops and Prod. (1993) 1:241-250, and other sequences as described in the Examples.

[0083] Polynucleotides of the present invention can be operatively linked to expression vectors containing regulatory sequences such as transcription control sequences, translation control sequences, origins of replication, and other regulatory sequences that are compatible with the recombinant cell and that control the expression of polynucleotides of the present invention. In particular, recombinant molecules of the present invention include transcription control sequences. Transcription control sequences are sequences which control the initiation, elongation, and termination of transcription. Included are those transcription control sequences which are sufficient to render promoter-dependent gene expression controllable for cell-type specific, tissue-specific or inducible by external signals or agents; such elements may be located in the 5' or 3' regions of the native gene. Particularly important transcription control sequences are those which control transcription initiation, such as promoter, enhancer, operator and repressor sequences. Suitable transcription control sequences include any transcription control sequence that can function in at least one of the recombinant cells of the present invention. A variety of such transcription control sequences are known to those skilled in the art. Preferred transcription control sequences include those which function in bacterial, yeast, fungal, insect and mammalian cells, such as, but not limited to, *tac*, *lac*, *trp*, *trc*, *oxy-pro*, *omp/lpp*, *rrnB*, bacteriophage lambda (λ) (such as λp_L and λp_R and fusions that include such promoters), bacteriophage T7, T7lac, bacteriophage T3, bacteriophage SP6, bacteriophage SP01, metallothionein, α -mating factor, *Pichia* alcohol oxidase, alphavirus subgenomic promoters (such as *Sindbis* virus subgenomic promoters), antibiotic resistance gene, baculovirus, *Heliothis zea* insect virus, vaccinia virus, herpesvirus, poxvirus, adenovirus, cytomegalovirus (such as intermediate early promoters, simian virus 40, retrovirus, actin, retroviral long terminal repeat, Rous sarcoma virus, heat shock, phosphate and nitrate transcription control sequences as well as other sequences capable of controlling gene expression in prokaryotic or eukaryotic cells.

[0084] Particularly preferred transcription control sequences are plant transcription control sequences. The choice of transcription control sequence will vary depending on the temporal and spatial requirements for expression, and also depending on the target species. Thus, expression of the nucleotide sequences of this invention in any plant organ (leaves, roots, seedlings, immature or mature reproductive structures, etc.) or at any stage of plant development is preferred. Although many transcription control sequences from dicotyledons have been shown to be operational in monocotyledons and vice versa, ideally dicotyledonous transcription control sequences are selected for expression in dicotyledons, and monocotyledonous promoters for expression in monocotyledons. However, there is no restriction to the provenance of selected transcription control sequences; it is sufficient that they are operational in driving the expression of the nucleotide sequences in the desired cell.

[0085] Preferred transcription control sequences that are expressed constitutively include but are not limited to promoters from genes encoding actin or ubiquitin and the CaMV 35S and 19S promoters. The nucleotide sequences of this invention can also be expressed under the regulation of promoters that are chemically regulated. This enables the MSH1 polypeptide to be synthesized only when the crop plants are treated with the inducing chemicals.

[0086] A preferred category of promoters is that which is induced by the physiological state of the plant (i.e. wound inducible, water-stress inducible, salt-stress inducible, disease inducible, and the like). Numerous promoters have been described which are expressed at wound sites and also at the sites of phytopathogen infection. Ideally, such a promoter should only be active locally at the sites of infection, and in this way the MSH1 polypeptides only accumulate in cells in which the accumulation is desired. Preferred promoters of this kind include those described by Stanford et al. *Mol. Gen. Genet.* 215: 200-208 (1989), Xu et al. *Plant Molec. Biol.* 22: 573-588 (1993), Logemann et al. *Plant Cell* 1: 151-158 (1989), Rohrmeier & Lehle, *Plant Molec. Biol.* 22: 783-792 (1993), Firek et al. *Plant Molec. Biol.* 22: 129-142 (1993), and Warner et al. *Plant J.* 3: 191-201 (1993).

[0087] Preferred tissue-specific expression patterns include but are not limited to green tissue specific, root specific, stem specific, and flower specific. Promoters suitable for expression in green tissue include many which regulate genes involved in photosynthesis and many of these have been cloned from both monocotyledons and dicotyledons. A preferred promoter is the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula, *Plant Molec. Biol.* 12: 579-589 (1989)). A preferred promoter for root specific expression is that described by de Framond (*FEBS* 290: 103-106 (1991); EP 0 452 269 to Ciba-Geigy). A preferred stem specific promoter is that described in U.S. Pat. No. 5,625,136 (to Ciba-Geigy) and which drives expression of the maize *trpA* gene.

[0088] A recombinant molecule of the present invention is a molecule that can include at least one of any polynucleotide heretofore described operatively linked to at least one of any transcription control sequence capable of effectively regulating expression of the polynucleotide(s) in the cell to be transformed, examples of which are disclosed herein.

[0089] A recombinant cell of the present invention includes any cell transformed with at least one of any polynucleotide of the present invention. Suitable and preferred polynucleotides as well as suitable and preferred recombinant molecules with which to transfer cells are disclosed herein.

[0090] Recombinant cells of the present invention can also be co-transformed with one or more recombinant molecules including plant MSH1 polynucleotides encoding one or more polypeptides of the present invention and one or more other polypeptides useful when expressed in plants.

[0091] It may be appreciated by one skilled in the art that use of recombinant DNA technologies can improve expression of transformed polynucleotides by manipulating, for example, the number of copies of the polynucleotides within a host cell, the efficiency with which those polynucleotides are transcribed, the efficiency with which the resultant transcripts are translated, and the efficiency of post-translational modifications. Recombinant techniques useful for increasing the expression of polynucleotides of the present invention include, but are not limited to, operatively linking polynucleotides to high-copy number plasmids, integration of the polynucleotides into one or more host cell chromosomes, addition of vector stability sequences to plasmids, substitutions or modifications of transcription control signals (e.g., promoters, operators, enhancers), substitutions or

modifications of translational control signals (e.g., ribosome binding sites, Shine-Dalgarno sequences), modification of polynucleotides of the present invention to correspond to the codon usage of the host cell, deletion of sequences that destabilize transcripts, and use of control signals that temporally separate recombinant cell growth from recombinant enzyme production during fermentation. The activity of an expressed recombinant polypeptide of the present invention may be improved by fragmenting, modifying, or derivatizing polynucleotides encoding such a polypeptide.

[0092] Recombinant cells of the present invention can be used to produce one or more polypeptides of the present invention by culturing such cells under conditions effective to produce such a polypeptide, and recovering the polypeptide. Effective conditions to produce a polypeptide include, but are not limited to, appropriate media, bioreactor, temperature, pH and oxygen conditions that permit polypeptide production. An appropriate, or effective, medium refers to any medium in which a cell of the present invention, when cultured, is capable of producing an MSH1 polypeptide of the present invention. Such a medium is typically an aqueous medium comprising assimilable carbon, nitrogen and phosphate sources, as well as appropriate salts, minerals, metals and other nutrients, such as vitamins. The medium may comprise complex nutrients or may be a defined minimal medium. Cells of the present invention can be cultured in conventional fermentation bioreactors, which include, but are not limited to, batch, fed-batch, cell recycle, and continuous fermentors. Culturing can also be conducted in shake flasks, test tubes, microtiter dishes, and petri plates. Culturing is carried out at a temperature, pH and oxygen content appropriate for the recombinant cell. Such culturing conditions are well within the expertise of one of ordinary skill in the art.

[0093] Depending on the vector and host system used for production, resultant polypeptides of the present invention may either remain within the recombinant cell; be secreted into the fermentation medium; be secreted into a space between two cellular membranes, such as the periplasmic space in *E. coli*; or be retained on the outer surface of a cell or viral membrane.

[0094] The phrase "recovering the polypeptide" refers simply to collecting the whole fermentation medium containing the polypeptide and need not imply additional steps of separation or purification. Polypeptides of the present invention can be purified using a variety of standard polypeptide purification techniques, such as, but not limited to, affinity chromatography, ion exchange chromatography, filtration, electrophoresis, hydrophobic interaction chromatography, gel filtration chromatography, reverse phase chromatography, concanavalin A chromatography, chromatofocusing and differential solubilization. Polypeptides of the present invention are preferably retrieved in "substantially pure" form. As used herein, "substantially pure" refers to a purity that allows for the effective use of the polypeptide as a diagnostic or test compound, and means, with increasing preference, at least 50%, 60%, 70%, 80%, 90%, 95%, or 98% homogeneous.

[0095] D. Transfected Plant Cells and Transgenic Plants

[0096] With regard to MSH1, particularly preferred recombinant cells are plant cells. By "plant cell" is meant any self-propagating cell bounded by a semi-permeable

membrane and containing a plastid. Such a cell also requires a cell wall if further propagation is desired. Plant cell, as used herein includes, without limitation, algae, cyanobacteria, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

[0097] The particular arrangement of the MSH1 sequence in the transformation vector will be selected according to the type of expression of the sequence that is desired. In some embodiments, expressing MSH1 polypeptides is desirable, while in others, a reduction of activity is desirable. The former embodiment is discussed first.

[0098] In one embodiment, at least one of the MSH1 polypeptides or an allele thereof, of the invention is expressed in a higher organism, e.g., a plant. A nucleotide sequence of the present invention is inserted into an expression cassette, which is then preferably stably integrated in the genome of said plant. In another preferred embodiment, the nucleotide sequence is included in a non-pathogenic self-replicating virus. Plants transformed in accordance with the present invention may be monocots or dicots and include, but are not limited to, maize, wheat, barley, rye, millet, chickpea, lentil, flax, olive, fig almond, pistachio, walnut, beet, parsnip, citrus fruits, including, but not limited to, orange, lemon, lime, grapefruit, tangerine, minneola, and tangelo, sweet potato, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, pepper, celery, squash, pumpkin, hemp, zucchini, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tomato, sorghum, sugarcane, sugarbeet, sunflower, rapeseed, clover, tobacco, carrot, cotton, alfalfa, rice, potato, eggplant, cucumber, *Arabidopsis*, and woody plants such as coniferous and deciduous trees.

[0099] Once a desired nucleotide sequence has been transformed into a particular plant species, it may be propagated in that species or moved into other varieties of the same species, particularly including commercial varieties, using traditional breeding techniques.

[0100] Accordingly, the present invention provides a method for producing a transfected plant cell or transgenic plant comprising the steps of a) transfecting a plant cell to contain a heterologous DNA segment encoding a protein and derived from an MSH1 polynucleotide not native to said cell (the polynucleotide indeed could be native but the expression pattern could be developmentally altered, still leading to the preferred effect); wherein said polynucleotide is operably linked to a promoter that can be used effectively for expression of transgenic proteins; b) optionally growing and maintaining said cell under conditions whereby a transgenic plant is regenerated therefrom; c) optionally growing said transgenic plant under conditions whereby said DNA is expressed, whereby the total amount of MSH1 polypeptide in said plant is altered. In a preferred embodiment, the method further comprises the step of obtaining and growing additional generations of descendants of said transgenic plant which comprise said heterologous DNA segment wherein said heterologous DNA segment is expressed. As used herein, "heterologous DNA", or, in some cases, "transgene" refers to foreign genes or polynucleotides, or additional, or modified versions of native or endogenous genes

or polynucleotides (perhaps driven by different promoters) in order to alter the traits of a plant in a specific manner.

[0101] The invention also provides plant cells which comprise heterologous DNA encoding an MSH1 polypeptide. In a preferred embodiment, the transgenic plant cell is a propagation material of a transgenic plant. The present invention also provides a transfected host cell comprising a host cell transfected with a construct comprising a promoter, enhancer or intron polynucleotide from an MSH1 polynucleotide, and a polynucleotide encoding a reporter protein.

[0102] The present invention also provides a method of preparing a transgenic plant comprising: a) producing a transfected plant cell having a transgene encoding an MSH1 polypeptide whereby MSH1 expression in said plant cell is altered; and b) growing a transgenic plant from the transfected plant cell wherein the MSH1 transgene is expressed in the transgenic plant. The expression of the transgene includes an increase or decrease in MSH1 expression. In some embodiments, the expression of the transgene produces an RNA that may interfere with a native MSH1 gene such that the expression of the native gene is either eliminated or reduced, resulting in a useful outcome.

[0103] The invention also provides a transgenic plant containing heterologous DNA which encodes an MSH1 polypeptide that is expressed in plant tissue, including expression in a vector introduced into the plant.

[0104] The present invention also provides an isolated polynucleotide which includes a transcription control element operably linked to a polynucleotide that encodes the MSH1 gene in plant tissue. In preferred embodiment, the transcription control element is the promoter native to an MSH1 gene.

[0105] In some embodiments, a nucleotide sequence of this invention is expressed in transgenic plants, thus causing the biosynthesis of the corresponding MSH1 polypeptide in the transgenic plants. In this way, transgenic plants with characteristics related to MSH1 expression are generated. For their expression in transgenic plants, the nucleotide sequences of the invention may require modification and optimization. Although preferred gene sequences may be adequately expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray et al. Nucl. Acids Res. 17:477-498 (1989)). All changes required to be made within the nucleotide sequences such as those described above are made using well known techniques of site directed mutagenesis, PCR, and synthetic gene construction using the methods described in the published patent applications EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol), and WO 93/07278 (to Ciba-Geigy).

[0106] For efficient initiation of translation, sequences adjacent to the initiating methionine may require modification. For example, they can be modified by the inclusion of sequences known to be effective in plants. Joshi has suggested an appropriate consensus for plants (NAR 15: 6643-6653 (1987)) and Clontech suggests a further consensus translation initiator (1993/1994 catalog, page 210). These consensus are suitable for use with the nucleotide sequences of this invention. The sequences are incorporated

into constructions comprising the nucleotide sequences, up to and including the ATG (while leaving the second amino acid unmodified), or alternatively up to and including the GTC subsequent to the ATG (with the possibility of modifying the second amino acid of the transgene).

[0107] Expression of the nucleotide sequences in transgenic plants is driven by transcription control elements shown to be functional in plants. Transformation of plants with a polynucleotide under the control of these regulatory elements provides for controlled expression in the transformed plant. Such transcription control elements have been described above. In addition to the selection of a suitable initiator of transcription, constructions for expression of MSH1 polypeptide in plants require an appropriate transcription terminator to be attached downstream of the heterologous nucleotide sequence. Several such terminators are available and known in the art (e.g. tml from CaMV, E9 from rbcS). Any available terminator known to function in plants can be used in the context of this invention.

[0108] Numerous other sequences can be incorporated into expression cassettes described in this invention. These include sequences which have been shown to enhance expression such as intron sequences (e.g. from Adhl and bronze1) and viral leader sequences (e.g. from TMV, MCMV and AMV).

[0109] It may be preferable to target expression of the nucleotide sequences of the present invention to different cellular localizations in the plant. In some cases, localization in the cytosol may be desirable, whereas in other cases, localization in some subcellular organelle may be preferred. Subcellular localization of heterologous DNA encoded polypeptides is undertaken using techniques well known in the art. Typically, the DNA encoding the target peptide from a known organelle-targeted gene product is manipulated and fused upstream of the nucleotide sequence. Many such target sequences are known for the chloroplast and their functioning in heterologous constructions has been shown. The expression of the nucleotide sequences of the present invention is also targeted to the endoplasmic reticulum or to the vacuoles of the host cells. Techniques to achieve this are well-known in the art.

[0110] Vectors suitable for plant transformation are described elsewhere in this specification. For *Agrobacterium*-mediated transformation, binary vectors or vectors carrying at least one T-DNA border sequence are suitable, whereas for direct gene transfer any vector is suitable and linear DNA containing only the construction of interest may be preferred. In the case of direct gene transfer, transformation with a single DNA species or co-transformation can be used (Schocher et al. Biotechnology 4: 1093-1096 (1986)). For both direct gene transfer and *Agrobacterium*-mediated transfer, transformation is usually (but not necessarily) undertaken with a selectable marker which may provide resistance to an antibiotic (kanamycin, hygromycin or methotrexate) or a herbicide (basta). The choice of selectable marker is not, however, critical to the invention.

[0111] In another preferred embodiment, a nucleotide sequence of the present invention is directly transformed into the plastid genome. A major advantage of plastid transformation is that plastids are capable of expressing multiple open reading frames under control of a single promoter. Plastid transformation technology is extensively

described in U.S. Pat. Nos. 5,451,513, 5,545,817, and 5,545,818, in PCT application no. WO 95/16783, and in McBride et al. (1994) Proc. Natl. Acad. Sci. USA 91, 7301-7305. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the gene of interest into a suitable target tissue, e.g., using biolistics or protoplast transformation (e.g., calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate homologous recombination with the plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab, Z., Hajdukiewicz, P., and Maliga, P. (1990) Proc. Natl. Acad. Sci. USA 87, 8526-8530; Staub, J. M., and Maliga, P. (1992) Plant Cell 4, 39-45). This resulted in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub, J. M., and Maliga, P. (1993) EMBO J. 12, 601-606). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-polypeptide antibiotic resistance genes with a dominant selectable marker, the bacterial *aadA* gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3'-adenyltransferase (Svab, Z., and Maliga, P. (1993) Proc. Natl. Acad. Sci. USA 90, 913-917). Previously, this marker had been used successfully for high-frequency transformation of the plastid genome of the green alga *Chlamydomonas reinhardtii* (Goldschmidt-Clermont, M. (1991) Nucl. Acids Res. 19: 4083-4089). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention. Typically, approximately 15-20 cell division cycles following transformation are required to reach a homoplasmic state. Plastid expression, in which genes are inserted by homologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclear-expressed genes to permit expression levels that can readily exceed 10% of the total soluble plant polypeptide. In a preferred embodiment, a nucleotide sequence of the present invention is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants homoplasmic for plastid genomes containing a nucleotide sequence of the present invention are obtained, and are preferentially capable of high expression of the nucleotide sequence.

[0112] In some embodiments, a reduction or suppression of MSH1 polypeptide activity is desired. In some embodiments, a reduction of MSH1 polypeptide activity may be obtained by introducing into plants an antisense construct based on an MSH1 cDNA or gene sequence. For antisense suppression, an MSH1 cDNA or gene is arranged in reverse orientation relative to the promoter sequence in the transformation vector. The introduced sequence need not be a full length MSH1 cDNA or gene, and need not be exactly homologous to the native MSH1 cDNA or gene found in the plant type to be transformed. Generally, however, where the introduced sequence is of shorter length, a higher degree of homology to the native MSH1 sequence will be needed for effective antisense suppression. The introduced antisense

sequence in the vector generally will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous MSH1 gene in the plant cell. Although the exact mechanism by which antisense RNA molecules interfere with gene expression has not been elucidated, it is believed that antisense RNA molecules bind to the endogenous mRNA molecules and thereby inhibit translation of the endogenous mRNA. The production and use of anti-sense constructs are disclosed, for instance, in U.S. Pat. No. 5,773,692 (using constructs encoding anti-sense RNA for chlorophyll a/b binding protein to reduce plant chlorophyll content), and U.S. Pat. No. 5,741,684 (regulating the fertility of pollen in various plants through the use of anti-sense RNA to genes involved in pollen development or function).

[0113] Suppression of endogenous MSH1 gene expression can also be achieved using ribozymes. Ribozymes are synthetic RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Pat. No. 4,987,071 to Cech and U.S. Pat. No. 5,543,508 to Haselhoff. Inclusion of ribozyme sequences within antisense RNAs may be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that bind to the antisense RNA are cleaved, leading to an enhanced antisense inhibition of endogenous gene expression.

[0114] Constructs in which an MSH1 cDNA or gene (or variants thereof) are over-expressed may also be used to obtain co-suppression of the endogenous MSH1 gene in the manner described in U.S. Pat. No. 5,231,021 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire MSH1 cDNA or gene be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous MSH1 gene. However, as with antisense suppression, the suppressive efficiency will be enhanced as (1) the introduced sequence is lengthened and (2) the sequence similarity between the introduced sequence and the endogenous MSH1 gene is increased.

[0115] Constructs expressing an untranslatable form of an MSH1 mRNA may also be used to suppress the expression of endogenous MSH1 activity. Methods for producing such constructs are described in U.S. Pat. No. 5,583,021 to Dougherty et al. such constructs may be prepared by introducing a premature stop codon into an MSH1 ORF.

[0116] Polynucleotides of the present invention may also be used to specifically suppress gene expression by methods such as RNA interference (RNAi), which may also include cosuppression and quelling. This and other techniques of gene suppression are well known in the art. A review of this technique is found in Science 288:1370-1372, 2000. Traditional methods of gene suppression, employing antisense RNA or DNA, operate by binding to the reverse sequence of a gene of interest such that binding interferes with subsequent cellular processes and thereby blocks synthesis of the corresponding protein. RNAi also operates on a post-transcriptional level and is sequence specific, but suppresses gene expression far more efficiently

[0117] Studies have demonstrated that one or more ribonucleases specifically bind to and cleave double-stranded RNA into short fragments. The ribonuclease(s) remains associated with these fragments, which in turn specifically bind to complementary mRNA, i.e. specifically bind to the transcribed mRNA strand for the gene of interest. The mRNA for the gene is also degraded by the ribonuclease(s) into short fragments, thereby obviating translation and expression of the gene. Additionally, an RNA polymerase may act to facilitate the synthesis of numerous copies of the short fragments, which exponentially increases the efficiency of the system. A unique feature of this gene suppression pathway is that silencing is not limited to the cells where it is initiated. The gene-silencing effects may be disseminated to other parts of an organism and even transmitted through the germ line to several generations.

[0118] Specifically, polynucleotides of the present invention are useful for generating gene constructs for silencing specific genes. Polynucleotides of the present invention may be used to generate genetic constructs that encode a single self-complementary RNA sequence specific for one or more genes of interest. Genetic constructs and/or gene-specific self-complementary RNA sequences may be delivered by any conventional method known in the art. Within genetic constructs, sense and antisense sequences flank an intron sequence arranged in proper splicing orientation making use of donor and acceptor splicing sites. Alternative methods may employ spacer sequences of various lengths rather than discrete intron sequences to create an operable and efficient construct. During post-transcriptional processing of the gene construct product, intron sequences are spliced-out, allowing sense and antisense sequences, as well as splice junction sequences, to bind forming double-stranded RNA. Select ribonucleases bind to and cleave the double-stranded RNA, thereby initiating the cascade of events leading to degradation of specific mRNA gene sequences, and silencing specific genes. Alternatively, rather than using a gene construct to express the self-complementary RNA sequences, the gene-specific double-stranded RNA segments are delivered to one or more targeted areas to be internalized into the cell cytoplasm to exert a gene silencing effect.

[0119] Using this cellular pathway of gene suppression, gene function may be studied and high-throughput screening of sequences may be employed to discover sequences affecting gene expression. Additionally, genetically modified plants may be generated.

[0120] Finally, dominant negative mutant forms of the disclosed sequences may be used to block endogenous MSH1 activity. Such mutants require the production of mutated forms of the MSH1 protein that interact with the same molecules as MSH1 but do not have MSH1 activity.

[0121] E. MSH1 Antibodies

[0122] The present invention also includes isolated antibodies capable of selectively binding to an MSH1 polypeptide of the present invention or to a mimetope thereof. Such antibodies are also referred to herein as anti-MSH1 antibodies. Particularly preferred antibodies of this embodiment include anti-*A. thaliana* MSH1 antibodies.

[0123] Isolated antibodies are antibodies that have been removed from their natural milieu. The term "isolated" does not refer to the state of purity of such antibodies. As such,

isolated antibodies can include anti-sera containing such antibodies, or antibodies that have been purified to varying degrees.

[0124] As used herein, the term "selectively binds to" refers to the ability of antibodies of the present invention to preferentially bind to specified polypeptides and mimetopes thereof of the present invention. Binding can be measured using a variety of methods known to those skilled in the art including immunoblot assays, immunoprecipitation assays, radioimmunoassays, enzyme immunoassays (e.g., ELISA), immunofluorescent antibody assays and immunoelectron microscopy; see, for example, Sambrook et al., *ibid.*, and Harlow & Lane, 1990, *ibid.*

[0125] Antibodies of the present invention can be either polyclonal or monoclonal antibodies. Antibodies of the present invention include functional equivalents such as antibody fragments and genetically-engineered antibodies, including single chain antibodies, that are capable of selectively binding to at least one of the epitopes of the polypeptide or mimetope used to obtain the antibodies. Antibodies of the present invention also include chimeric antibodies that can bind to more than one epitope. Preferred antibodies are raised in response to polypeptides, or mimetopes thereof, that are encoded, at least in part, by a polynucleotide of the present invention.

[0126] A preferred method to produce antibodies of the present invention includes (a) administering to an animal an effective amount of a polypeptide or mimetope thereof of the present invention to produce the antibodies and (b) recovering the antibodies. In another method, antibodies of the present invention are produced recombinantly using techniques as heretofore disclosed to produce MSH1 polypeptides of the present invention. Antibodies raised against defined polypeptides or mimetopes can be advantageous because such antibodies are not substantially contaminated with antibodies against other substances that might otherwise cause interference in a diagnostic assay.

[0127] Antibodies of the present invention have a variety of potential uses that are within the scope of the present invention. For example, such antibodies can be used (a) as reagents in assays to detect expression of MSH1 by plant, (b) as tools to screen expression libraries and/or to recover desired polypeptides of the present invention from a mixture of polypeptides and other contaminants and/or (c) to modulate the function of an MSH1 polypeptide (e.g., increase or decrease the level or activity of an MSH1 polypeptide). Antibodies of the present invention can be used to target cytotoxic, therapeutic or imaging agents to subjects in order to deliver therapeutic agents or localize imaging agents to RA-affected organs or tissues. Targeting can be accomplished by conjugating (i.e., stably joining) such antibodies to the therapeutic or imaging agents using techniques known to those skilled in the art.

[0128] F. Methods for Effecting Mitochondrial Ectopic Recombination and Identification of Mutants Arising from Mitochondrial Ectopic Recombination

[0129] In one embodiment, the invention provides a method to identify a compound capable of inhibiting MSH1 activity (e.g., effecting ectopic recombination) of a plant, said method comprising contacting an isolated plant MSH1 nucleic acid molecule with a putative inhibitory compound

which, in the absence of said compound, said plant MSH1 nucleic acid molecule has the activity of suppressing ectopic recombination; and determining if said putative inhibitory compound inhibits said activity. The present invention also comprises a method for effecting mitochondrial ectopic recombination comprising providing a plant, and suppressing expression of an MSH1-homologous gene in the plant. A preferred inhibitory compound is an RNA molecule having RNAi activity.

[0130] The invention further provides a method for identification of mutants arising from mitochondrial ectopic recombination comprising providing a plant, and suppressing expression of an MSH1-homologous gene in the plant, and detecting an aberrant phenotype, whereby a mutant is identified. A preferred aberrant phenotype includes cytoplasmic male sterility. Cytoplasmic male sterility is a plant trait that facilitates a cost-effective strategy for the production of proprietary hybrids. Hybrid seed is valued for producing higher yields and more uniform crop stands. Hybrids are important in a large number of horticultural and agronomic crops including corn, sorghum, rice, wheat, tomato, rape, sunflower, carrot, onion, sugar beet, to name few. Cytoplasmic male sterility (CMS) mutations arise as the consequence of ectopic recombination events that produce novel expressed DNA sequences within the mitochondrial genome. This is well documented in the scientific literature. The present invention also includes mutants identified the method of the invention.

EXAMPLES

Example 1

Identification of the AtMSH1 Gene

[0131] A. Gene mapping, cloning, and sequence analysis. A map-based cloning strategy for the isolation of the CHM locus involved the design of PCR-based co-dominant markers, using the Cereon *Arabidopsis* polymorphism collection (Jander, et al., *ibid.*) to distinguish between the Col-0 and Landsburg *erecta* ecotypes used in the F₂ mapping populations. The markers were designed in a 5-Mb region of Chromosome III based on information from the classical mapping experiments of CHM (Martinez-Zapater, et al., *ibid.*; Redei, *ibid.*). The primer sequences for markers are available upon request. The F₂ mapping population was derived from a cross between the chm1-1 mutant line and Landsburg *erecta* ecotype (pollen donor). A segregating sub-population of 172 variegated plants was analyzed. Genomic DNA purification was conducted according to Li and Chory, *ibid.* DNA gel blot analysis was conducted using the protocol of Sambrook et al., *ibid.* High resolution mapping of the CHM locus on *Arabidopsis* Chromosome III delimited the gene to an 80-kb interval as shown in FIG. 1.

[0132] DNA sequencing of the candidate locus in chm1-1, chm1-2 and chm1-3 mutants (Kanazawa, et al., *ibid.*) was conducted in a Beckman/Coulter CEQ2000XL 8-capillary DNA sequencer. Two independent PCR samples for each mutant were sequenced. The 5' RACE analysis was done with the GeneRacer® Kit (Invitrogen, Carlsbad, Calif.). Mutants chm1-1 and chm1-2 were obtained from the *Arabidopsis* Biological Resource Center, and mutant chm1-3 was provided by a colleague. Sequence analysis of the interval revealed a gene candidate with similarity in

sequence features to the MutS gene of *E. coli* (FIG. 2). MutS is a component of the *E. coli* mismatch repair and DNA recombination apparatus (Marti, et al., *ibid.*). The gene, comprised of 22 exons, was predicted to encode a 43-amino acid mitochondrial targeting presequence with mitochondrial targeting values of 0.916 (MitoProt), 0.943 (Predator) and 0.856 (TargetP). RNA gel blots showed that the transcript derived from this gene was 3.5 kb in size and the encoded protein 1118 amino acids in length, predicting a 124-kDa polypeptide.

[0133] The two sequence-indexed T-DNA insertion mutants were identified on the SiGnAL (Salk Institute Genomic Analysis Laboratory) website (Accessions SALK041951 (SEQ ID NO:5) and SALK046763 (SEQ ID NO:4)), and seed for the mutants obtained from the *Arabidopsis* Biological Resource Center (ABRC). The T-DNA insertion positions were confirmed by DNA sequencing of the insertion junctions. The first insertion was located within the fourth exon and the second within the eighth intron. Analysis of the T-DNA mutants (T3 generation) revealed mild green-white leaf variegation, growing more intense in the following selfed generation. Variegated plants having a green-white variegation phenotype carried a mitochondrial genome rearrangement similar to that observed in the mutants chm1-1 and chm1-2. A population of 60 T4 plants segregating for one of the T-DNA (SALK041951) mutations (16 wildtype, 31 hemizygous, 13 homozygous for the T-DNA) showed co-segregation of the T-DNA with the mitochondrial shifting phenotype. Of the 13 progeny homozygous for the T-DNA insertion, eight were variegated and the remaining five showed no obvious variegation phenotype. Incomplete penetrance of the variegation phenotype is characteristic of chm1-1 and chm1-2 mutants (Redei, *ibid.*).

[0134] DNA gel blot hybridization analysis of mitochondrial genome configuration using the mitochondrial atp9-rp116 junction sequence associated with substoichiometric shifting (Sakamoto, et al., *ibid.*) as probe. Total genomic DNA was digested with BamHI, subjected to gel electrophoresis, blotted and probed. Lane Wt designates wildtype ecotype Columbia-0, lane C1 designates mutant chm1-1, and T1 and T2 designate two sister lines containing the T-DNA1 insertion mutation. DNA band pattern changes previously associated with substoichiometric shifting were noted (Martinez-Zapater, et al., *ibid.*).

[0135] Cosegregation analysis of mitochondrial substoichiometric shifting with the T-DNA1 insertion mutation. A three-primer PCR-based assay to detect substoichiometric shifting (Sakamoto, et al., *ibid.*) was used to assay wildtype Col-0 (Wt), mutant chm1-1 (C1) and individual plants segregating for presence of the T-DNA insertion within the candidate CHM locus.

[0136] All progeny homozygous for the T-DNA insertion mutation showed the mitochondrial shifting phenotype. None of the segregants hemizygous for, or lacking, the T-DNA mutation showed evidence of variegation. The hemizygous plants showed no mitochondrial shifting. Similar co-segregation results were obtained for the second TDNA (SALK046763) mutation as well.

[0137] To test further the possibility that the identified MutS-homologous sequence was CHM, we sequenced the chm1-1 and chm1-2 alleles of the gene. The chm1-1 line had

a single nucleotide (C-T) substitution that gave rise to a premature stop codon within the fourth exon (**FIG. 1E**). The chm1-2 mutant had a single nucleotide (G-A) substitution at the intron-exon junction of Exon 2 (**FIG. 1E**). This substitution resulted in two-nucleotide slippage of the intron splice site, producing a frameshift and premature termination of translation five amino acids beyond the mutation site. Therefore, in both chm1-1 and chm1-2 mutant lines, the CHM candidate locus is predicted to give rise to highly truncated, inactive peptides.

[0138] Sequence analysis of the chm1-3 allele, derived from a tissue culture line by Martinez-Zapater et al. (Martinez-Zapater, et al., *ibid.*), revealed an amino acid substitution (Cys-Tyr) within the ATP binding domain (**FIG. 1E**). The mutant phenotype in this case may be due to the substitution of a bulkier amino acid within a site essential for protein function.

[0139] B. The CHM candidate has features of a mismatch repair component. The MutS-homologous gene identified as a candidate for CHM displayed several features characteristic of a mismatch repair component. These features included an ATP-binding domain (aa 761-946) comprised of four well conserved motifs designated M1-M4 (Obmolova, et al., *ibid.*; **FIG. 2B**). In addition to ATPase function, this domain appears to be involved in dimerization of the protein (Obmolova, et al.; Lamers, et al.), although this has not yet been demonstrated for mitochondrial MutS homologs. A DNA binding domain (aa 129-206) was also identified (**FIGS. 1, 2**) to contain the aromatic doublet (FY) motif that is characteristic of this domain in MutS and MutS-like proteins (**FIG. 2A**). This doublet was shown to be essential for mismatch recognition and specific DNA binding activity (33, 34). We were unable to detect three other conserved domains characteristic of MutS. A connector domain, involved in inter-domain interactions, a core domain and a clamp domain, involved in nonspecific double-strand DNA binding, did not appear to be well conserved. The CHM candidate protein likely localizes to mitochondria. To confirm that the MutS-like protein localized to the mitochondrion, we conducted RACE-PCR and discovered a transcript start site at 578 residues upstream to the site predicted in the Munich Information Center for Protein Sequences (MIPS) database (Schoof, et al.) and in GenBank (Accession AP000382). No start site was observed by RACE analysis at the point predicted by the MIPS database, and three clustered transcription start sites were detected at the upstream site. The confirmed start site added 102 amino acids to the predicted protein product and permitted the identification of a mitochondrial targeting presequence that was omitted from the previous database entries. The sequence was annotated

based on cDNA sequence analysis and is available as GenBank Accession AY191303.

Example 2

Plant Transformation and Biolistic Delivery

[0140] The amino acid sequence of AtMSH1 was analyzed with MitoProt (Claros & Vincens (1996) *Eur. J. Biochem.* 241, 779-786), and the first 213 nucleotides of the gene were PCR amplified with the primers

MSHtranspFor
5' GGCCATGGTGTGAATTGCATAGTCGTCG3' (SEQ ID NO:48)
and

MSHtranspRev
5' GGCCATGGAAA CATCACTTGACGTCTTC3'. (SEQ ID NO:49)

[0141] PCR products were ligated to the Pgm®-T Easy Vector System (Promega) and digested with NcoI to release the insert. Insert fragments were ligated to the pCambia 1302 vector at the NcoI site that resides at the start of gfp. This vector utilizes the CaMV 35S promoter. Bombardment experiments used 4-week-old leaves of *Arabidopsis* (Col-0) with tungsten particles and the Biolistic PDS-1000/He system (Bio-Rad). Particles were bombarded into *Arabidopsis* leaves using 900-psi rupture discs under a vacuum of 900-psi (1 psi=6.9 kPa). After the bombardment, *Arabidopsis* leaves were allowed to recover for 18-22 h on Murashige and Skoog media plates at 22° C. in 16 h daylight. Localization of GFP expression was conducted by confocal laser scanning microscopy with Bio-Rad 1024 MRC-ES using 488 nm excitation and two-channel measurement of emission, 522 nm (green/GFP) and 680 nm (red/chlorophyll). Mitochondria were identified by their characteristic movement and rapid inter-conversions from small round to highly elongated, shapes. Plastids located in the cells emit red autofluorescence. Positive controls for mitochondrial (F1-ATPase gamma subunit provided by Dr. D. Stern) and chloroplast (Rubisco Pea/SSU/TPSS, provided by Dr. L. Alison) targeting were included with each experiment.

Example 3

Identification of Homologs

[0142] Homologs were identified by BLAST search using the tblastn program against the est_others database. The MSH1 protein sequence was used as the Query sequence. The search was done using the BLOSUM62 matrix, word size of 3 and low complexity filter.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 65

<210> SEQ ID NO 1

<211> LENGTH: 3673

<212> TYPE: DNA

<213> ORGANISM: *Arabidopsis thaliana*

<400> SEQUENCE: 1

-continued

agaggactgt gagattgtga attgcatagt cgtcgtcttc tggcgggaaa agaagcccta	60
gaaaagggt gaaaggtagaa aactctactt cttcttcttc ttcttcttca gagtgtgaga	120
gagatgcatt ggattgctac cagaaacgcc gtcgtttcat tcccaaatg gcggttcttc	180
ttccgctcct catatcgac ttactcttcc ctcaaaccct cctcccaat tctacttaat	240
agaaggtagt ctgaggggat atcttgtctc agagatggaa agtctttgaa aagaatcaca	300
acggcttcta agaaagtga gacgtcaagt gatgttctca ctgacaaaga tctctctcat	360
ttggtttggt ggaaggagag attgcagaca tgtaagaaac catctactct tcagcttatt	420
gaaaggctta tgtacaccaa tttacttggt ttggacccta gcttgaggaa tggaagtta	480
aaagatggaa acctcaactg ggagatgttg cagtttaagt caaggtttcc acgcgaagtt	540
ttgctctgca gagtaggaga attttatgag gctattggaa tagatgcttg tatacttggt	600
gaatatgctg gtctcaatcc ttttggtggt cttcgatcag atagtattcc aaaggctggc	660
tgccaatta tgaatcttcg acagactttg gatgacctga caccgaatgg ttattcagtg	720
tgtattgtgg aggaagtta ggggccaaaca ccagcacgct cccgtaaagg tcgatttatt	780
tcaggcatg cacatccagg aagtccttat gtatatgggc ttgtcgggtg tgaccatgat	840
cttgactttc ctgactctat gcctgttggt gggatatctc gttcagcaag ggggtattgt	900
atgatatacta ttttcgagac tatgaaagca ttttcgctag atgatggtct aacagaagaa	960
gccttagtta ccaagctccg cactcgtcgc tgtcatcatc ttttcttaca tgcactggtg	1020
aggcacaatg catcagggac gtgccgctgg ggagagtttg ggggaagggg tctactctgg	1080
ggagaatgca gtagcaggaa ttttgaatgg tttgaaggag atactctttc cgagctctta	1140
tcaagggtca aagatgttta tggctctgat gatgaagttt cctttagaaa tgtcaatgta	1200
ccttcaaaaa atcggccacg tccgttgcat cttggaacgg ctacacaaat tggtgccetta	1260
cctactgaag gaataccttg tttgttgaa gtgttacttc catctactg cagtggctg	1320
ccttctttgt atgttaggga tcttcttctg aacctctctg cttacgatat tgctctgaaa	1380
attcaagaaa cgtgcaagct catgagcaca gtaacatggt caattccaga gtttacctgc	1440
gtctcttctg ctaagcttgt gaagcttctt gagcaacggg aagccaacta cattgagttc	1500
tgtcgaataa aaaatgtgct tgatgatgta ttacatatgc atagacatgc tgagcttggtg	1560
gaaatcctga aattattgat ggatcctacc tgggtggcta ctggtttgaa aattgacttt	1620
gacacttttg tcaacgaatg tcattggcg tctgatacaa ttggtgaaat gatctcttta	1680
gatgagaatg aaagtcatca gaatgtaagt aaatgtgaca atgtcccgaa cgaattcttt	1740
tatgatatgg agtcttcatg gcgaggtcgc gtttaaggaa ttcatataga ggaagaaatc	1800
actcaagtag aaaaatcagc tgaggcttta tctttagcag tagctgagga ttttaccct	1860
attatatcaa gaattaaggc caccactgct tcaacttggtg gccgaaagg cgaatcgca	1920
tatgcaagag agcatgagtc tgtttggttc aaggggaaac ggtttacgcc atctatctgg	1980
gctggtactg caggggaaga ccaaataaaa cagctgaaac ctgccttaga ctcgaaagga	2040
aaaaagggtg gagaagaatg gtttacgacc ccaaagggtg aaattgcttt agtcagatac	2100
catgaagcta gtgagaatgc aaaagctcgg gtgttggaac tgttgcgca gttatccgtt	2160
aaattgcaaa caaaaataaa tgttcttctg tttgcatcta tgcttctggt catttcaaaa	2220
gcattatttt cccatgcttg tgaagggaga aggcgaaagt gggtttttcc aacgcttgct	2280

-continued

```

ggattcagtt tagatgaggg cgcaaaacca ttagatggtg ccagtcgaat gaagctgaca 2340
ggcctgtcac cttattggtt tgatgtatct tctggaaccg ctgttcacaa taccgttgac 2400
atgcaatcac tgtttcttct aactggacct aacggtggtg gtaaatcgag tttgctcaga 2460
tcaatatgcg cagctgctct acttggaatt tccggtttaa tggttccagc tgaatcagct 2520
tgtattcctc actttgattc catcatgctt cacatgaaat catatgacag cctgtagac 2580
ggaaaaagtt ctttccaggt agaaatgtcg gaaatacgat ctattgtaag ccaggctact 2640
tcgagaagcc tagtgcttat agatgagata tgccgagggg cagagacagc aaaaggcacc 2700
tgtatcgctg gtagtggtg agagagtctt gacacaagtg gttgtttggg tattgtatct 2760
actcatctcc atggaatctt cagtttacct cttacagcga aaaacatcac atataaagca 2820
atgggagccg aaaatgtcga agggcaaac aagccaactt ggaaattgac agatggagtc 2880
tgcagagaga gtcttgctt tgaaacagct aagagggag gtgttccga gtcagttatc 2940
caaagagctg aagctcttta cctctcggtc tatgcaaaag acgcatcagc tgaagttgtc 3000
aaaccgacc aaatcataac ttcattccaa aatgaccagc agatccaaaa accagtcagc 3060
tctgagagaa gtttgagaa ggacttagca aaagctatcg tcaaaatctg tgggaaaaag 3120
atgattgagc ctgaagcaat agaattgtct tcaattggtg ctctgtgagct tccacctcca 3180
tctacagttg gttcttcagc cgtgtatgtg atgcggagac ccgataagag attgtacatt 3240
ggacagaccg atgatcttga aggacgaata cgtgcgcac gagcaaagga aggactgcaa 3300
gggtcaagtt ttctatacct tatgggtcaa ggtaagagca tggcttgtca gttagagact 3360
ctattgatta atcaactcca tgaacaaggc tactctctgg ctaacctagc cgatggaaag 3420
caccgtaatt tcggaacgtc ctcaagcttg agtacatcag acgtagtcag catcttatag 3480
tttgaacat tagctgtggt ttagttagat catctctatg tgcaattgaa caagtcagtt 3540
tgctagaact agagtagatt actaagaac catgccgtt ttcattttga gatattgcaa 3600
aacggcatgc agttcgggta agtcggatgc cgcaattacc aattttgggt cagtctgtgt 3660
aattgtcgtt tca 3673

```

```

<210> SEQ ID NO 2
<211> LENGTH: 3673
<212> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (124)..(3480)

```

```

<400> SEQUENCE: 2

```

```

agaggactgt gagattgtga attgcatagt cgtcgtcttc tggcgggaaa agaagcccta 60
gaaaaagggt gaaaggtgaa aactctactt cttctcttc ttctcttca gagtgtgaga 120

gag atg cat tgg att gct acc aga aac gcc gtc gtt tca ttc cca aaa 168
Met His Trp Ile Ala Thr Arg Asn Ala Val Val Ser Phe Pro Lys
1 5 10 15

tgg cgg ttc ttc ttc cgc tcc tca tat cgc act tac tct tcc ctg aaa 216
Trp Arg Phe Phe Phe Arg Ser Ser Tyr Arg Thr Tyr Ser Ser Leu Lys
20 25 30

ccc tcc tcc cca att cta ctt aat aga agg tac tct gag ggg ata tct 264
Pro Ser Ser Pro Ile Leu Leu Asn Arg Arg Tyr Ser Glu Gly Ile Ser
35 40 45

```

-continued

tgt ctc aga gat gga aag tct ttg aaa aga atc aca acg gct tct aag	312
Cys Leu Arg Asp Gly Lys Ser Leu Lys Arg Ile Thr Thr Ala Ser Lys	
50 55 60	
aaa gtg aag acg tca agt gat gtt ctc act gac aaa gat ctc tct cat	360
Lys Val Lys Thr Ser Ser Asp Val Leu Thr Asp Lys Asp Leu Ser His	
65 70 75	
ttg gtt tgg tgg aag gag aga ttg cag aca tgt aag aaa cca tct act	408
Leu Val Trp Trp Lys Glu Arg Leu Gln Thr Cys Lys Lys Pro Ser Thr	
80 85 90 95	
ctt cag ctt att gaa agg ctt atg tac acc aat tta ctt ggt ttg gac	456
Leu Gln Leu Ile Glu Arg Leu Met Tyr Thr Asn Leu Leu Gly Leu Asp	
100 105 110	
cct agc ttg agg aat gga agt tta aaa gat gga aac ctc aac tgg gag	504
Pro Ser Leu Arg Asn Gly Ser Leu Lys Asp Gly Asn Leu Asn Trp Glu	
115 120 125	
atg ttg cag ttt aag tca agg ttt cca cgc gaa gtt ttg ctc tgc aga	552
Met Leu Gln Phe Lys Ser Arg Phe Pro Arg Glu Val Leu Leu Cys Arg	
130 135 140	
gta gga gaa ttt tat gag gct att gga ata gat gct tgt ata ctt gtt	600
Val Gly Glu Phe Tyr Glu Ala Ile Gly Ile Asp Ala Cys Ile Leu Val	
145 150 155	
gaa tat gct ggt ctc aat cct ttt ggt ggt ctt cga tca gat agt att	648
Glu Tyr Ala Gly Leu Asn Pro Phe Gly Gly Leu Arg Ser Asp Ser Ile	
160 165 170 175	
cca aag gct ggc tgc cca att atg aat ctt cga cag act ttg gat gac	696
Pro Lys Ala Gly Cys Pro Ile Met Asn Leu Arg Gln Thr Leu Asp Asp	
180 185 190	
ctg aca cgc aat ggt tat tca gtg tgt att gtg gag gaa gtt cag ggg	744
Leu Thr Arg Asn Gly Tyr Ser Val Cys Ile Val Glu Glu Val Gln Gly	
195 200 205	
cca aca cca gca cgc tcc cgt aaa ggt cga ttt att tca ggg cat gca	792
Pro Thr Pro Ala Arg Ser Arg Lys Gly Arg Phe Ile Ser Gly His Ala	
210 215 220	
cat cca gga agt cct tat gta tat ggg ctt gtc ggt gtt gac cat gat	840
His Pro Gly Ser Pro Tyr Val Tyr Gly Leu Val Gly Val Asp His Asp	
225 230 235	
ctt gac ttt cct gat cct atg cct gtt gtt ggg ata tct cgt tca gca	888
Leu Asp Phe Pro Asp Pro Met Pro Val Val Gly Ile Ser Arg Ser Ala	
240 245 250 255	
agg ggg tat tgt atg ata tct att ttc gag act atg aaa gca tat tcg	936
Arg Gly Tyr Cys Met Ile Ser Ile Phe Glu Thr Met Lys Ala Tyr Ser	
260 265 270	
cta gat gat ggt cta aca gaa gaa gcc tta gtt acc aag ctc cgc act	984
Leu Asp Asp Gly Leu Thr Glu Glu Ala Leu Val Thr Lys Leu Arg Thr	
275 280 285	
cgt cgc tgt cat cat ctt ttc tta cat gca tcg ttg agg cac aat gca	1032
Arg Arg Cys His His Leu Phe Leu His Ala Ser Leu Arg His Asn Ala	
290 295 300	
tca ggg acg tgc cgc tgg gga gag ttt ggg gaa ggg ggt cta ctc tgg	1080
Ser Gly Thr Cys Arg Trp Gly Glu Phe Gly Glu Gly Gly Leu Leu Trp	
305 310 315	
gga gaa tgc agt agc agg aat ttt gaa tgg ttt gaa gga gat act ctt	1128
Gly Glu Cys Ser Ser Arg Asn Phe Glu Trp Phe Glu Gly Asp Thr Leu	
320 325 330 335	
tcc gag ctc tta tca agg gtc aaa gat gtt tat ggt ctt gat gat gaa	1176
Ser Glu Leu Leu Ser Arg Val Lys Asp Val Tyr Gly Leu Asp Asp Glu	
340 345 350	

-continued

gtt tcc ttt aga aat gtc aat gta cct tca aaa aat cgg cca cgt ccg	1224
Val Ser Phe Arg Asn Val Asn Val Pro Ser Lys Asn Arg Pro Arg Pro	
355 360 365	
ttg cat ctt gga acg gct aca caa att ggt gcc tta cct act gaa gga	1272
Leu His Leu Gly Thr Ala Thr Gln Ile Gly Ala Leu Pro Thr Glu Gly	
370 375 380	
ata cct tgt ttg ttg aag gtg tta ctt cca tct acg tgc agt ggt ctg	1320
Ile Pro Cys Leu Leu Lys Val Leu Leu Pro Ser Thr Cys Ser Gly Leu	
385 390 395	
cct tct ttg tat gtt agg gat ctt ctt ctg aac cct cct gct tac gat	1368
Pro Ser Leu Tyr Val Arg Asp Leu Leu Leu Asn Pro Pro Ala Tyr Asp	
400 405 410 415	
att gct ctg aaa att caa gaa acg tgc aag ctc atg agc aca gta aca	1416
Ile Ala Leu Lys Ile Gln Glu Thr Cys Lys Leu Met Ser Thr Val Thr	
420 425 430	
tgt tca att cca gag ttt acc tgc tgc tct tct gct aag ctt gtg aag	1464
Cys Ser Ile Pro Glu Phe Thr Cys Val Ser Ser Ala Lys Leu Val Lys	
435 440 445	
ctt ctt gag caa cgg gaa gcc aac tac att gag ttc tgt cga ata aaa	1512
Leu Leu Glu Gln Arg Glu Ala Asn Tyr Ile Glu Phe Cys Arg Ile Lys	
450 455 460	
aat gtg ctt gat gat gta tta cat atg cat aga cat gct gag ctt gtg	1560
Asn Val Leu Asp Asp Val Leu His Met His Arg His Ala Glu Leu Val	
465 470 475	
gaa atc ctg aaa tta ttg atg gat cct acc tgg gtg gct act ggt ttg	1608
Glu Ile Leu Lys Leu Leu Met Asp Pro Thr Trp Val Ala Thr Gly Leu	
480 485 490 495	
aaa att gac ttt gac act ttt gtc aac gaa tgt cat tgg gcg tct gat	1656
Lys Ile Asp Phe Asp Thr Phe Val Asn Glu Cys His Trp Ala Ser Asp	
500 505 510	
aca att ggt gaa atg atc tct tta gat gag aat gaa agt cat cag aat	1704
Thr Ile Gly Glu Met Ile Ser Leu Asp Glu Asn Glu Ser His Gln Asn	
515 520 525	
gta agt aaa tgt gac aat gtc ccg aac gaa ttc ttt tat gat atg gag	1752
Val Ser Lys Cys Asp Asn Val Pro Asn Glu Phe Phe Tyr Asp Met Glu	
530 535 540	
tct tca tgg cga ggt cgc gtt aag gga att cat ata gag gaa gaa atc	1800
Ser Ser Trp Arg Gly Arg Val Lys Gly Ile His Ile Glu Glu Glu Ile	
545 550 555	
act caa gta gaa aaa tca gct gag gct tta tct tta gca gta gct gag	1848
Thr Gln Val Glu Lys Ser Ala Glu Ala Leu Ser Leu Ala Val Ala Glu	
560 565 570 575	
gat ttt cac cct att ata tca aga att aag gcc acc act gct tca ctt	1896
Asp Phe His Pro Ile Ile Ser Arg Ile Lys Ala Thr Thr Ala Ser Leu	
580 585 590	
ggg ggc ccg aaa ggc gaa atc gca tat gca aga gag cat gag tct gtt	1944
Gly Gly Pro Lys Gly Glu Ile Ala Tyr Ala Arg Glu His Glu Ser Val	
595 600 605	
tgg ttc aag ggg aaa cgg ttt acg cca tct atc tgg gct ggt act gca	1992
Trp Phe Lys Gly Lys Arg Phe Thr Pro Ser Ile Trp Ala Gly Thr Ala	
610 615 620	
ggg gaa gac caa ata aaa cag ctg aaa cct gcc tta gac tcg aaa gga	2040
Gly Glu Asp Gln Ile Lys Gln Leu Lys Pro Ala Leu Asp Ser Lys Gly	
625 630 635	
aaa aag gtt gga gaa gaa tgg ttt acg acc cca aag gtg gaa att gct	2088
Lys Lys Val Gly Glu Glu Trp Phe Thr Thr Pro Lys Val Glu Ile Ala	
640 645 650 655	

-continued

tta gtc aga tac cat gaa gct agt gag aat gca aaa gct cgg gtg ttg	2136
Leu Val Arg Tyr His Glu Ala Ser Glu Asn Ala Lys Ala Arg Val Leu	
660 665 670	
gaa ctg ttg cgc gag tta tcc gtt aaa ttg caa aca aaa ata aat gtt	2184
Glu Leu Leu Arg Glu Leu Ser Val Lys Leu Gln Thr Lys Ile Asn Val	
675 680 685	
ctt gtc ttt gca tct atg ctt ctg gtc att tca aaa gca tta ttt tcc	2232
Leu Val Phe Ala Ser Met Leu Leu Val Ile Ser Lys Ala Leu Phe Ser	
690 695 700	
cat gct tgt gaa ggg aga agg cga aag tgg gtt ttt cca acg ctt gtc	2280
His Ala Cys Glu Gly Arg Arg Arg Lys Trp Val Phe Pro Thr Leu Val	
705 710 715	
gga ttc agt tta gat gag ggc gca aaa cca tta gat ggt gcc agt cga	2328
Gly Phe Ser Leu Asp Glu Gly Ala Lys Pro Leu Asp Gly Ala Ser Arg	
720 725 730 735	
atg aag ctg aca ggc ctg tca cct tat tgg ttt gat gta tct tct gga	2376
Met Lys Leu Thr Gly Leu Ser Pro Tyr Trp Phe Asp Val Ser Ser Gly	
740 745 750	
acc gct gtt cac aat acc gtt gac atg caa tca ctg ttt ctt cta act	2424
Thr Ala Val His Asn Thr Val Asp Met Gln Ser Leu Phe Leu Leu Thr	
755 760 765	
gga cct aac ggt ggt ggt aaa tcg agt ttg ctc aga tca ata tgc gca	2472
Gly Pro Asn Gly Gly Gly Lys Ser Ser Leu Leu Arg Ser Ile Cys Ala	
770 775 780	
gct gct cta ctt gga att tcc ggt tta atg gtt cca gct gaa tca gct	2520
Ala Ala Leu Leu Gly Ile Ser Gly Leu Met Val Pro Ala Glu Ser Ala	
785 790 795	
tgt att cct cac ttt gat tcc atc atg ctt cac atg aaa tca tat gac	2568
Cys Ile Pro His Phe Asp Ser Ile Met Leu His Met Lys Ser Tyr Asp	
800 805 810 815	
agc cct gta gac gga aaa agt tct ttc cag gta gaa atg tcg gaa ata	2616
Ser Pro Val Asp Gly Lys Ser Ser Phe Gln Val Glu Met Ser Glu Ile	
820 825 830	
cga tct att gta agc cag gct act tcg aga agc cta gtg ctt ata gat	2664
Arg Ser Ile Val Ser Gln Ala Thr Ser Arg Ser Leu Val Leu Ile Asp	
835 840 845	
gag ata tgc cga ggg aca gag aca gca aaa ggc acc tgt atc gct ggt	2712
Glu Ile Cys Arg Gly Thr Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly	
850 855 860	
agt gtg gta gag agt ctt gac aca agt ggt tgt ttg ggt att gta tct	2760
Ser Val Val Glu Ser Leu Asp Thr Ser Ser Gly Cys Leu Gly Ile Val Ser	
865 870 875	
act cat ctc cat gga atc ttc agt tta cct ctt aca gcg aaa aac atc	2808
Thr His Leu His Gly Ile Phe Ser Leu Pro Leu Thr Ala Lys Asn Ile	
880 885 890 895	
aca tat aaa gca atg gga gcc gaa aat gtc gaa ggg caa acc aag cca	2856
Thr Tyr Lys Ala Met Gly Ala Glu Asn Val Glu Gly Gln Thr Lys Pro	
900 905 910	
act tgg aaa ttg aca gat gga gtc tgc aga gag agt ctt gcg ttt gaa	2904
Thr Trp Lys Leu Thr Asp Gly Val Cys Arg Glu Ser Leu Ala Phe Glu	
915 920 925	
aca gct aag agg gaa ggt gtt ccc gag tca gtt atc caa aga gct gaa	2952
Thr Ala Lys Arg Glu Gly Val Pro Glu Ser Val Ile Gln Arg Ala Glu	
930 935 940	
gct ctt tac ctc tcg gtc tat gca aaa gac gca tca gct gaa gtt gtc	3000
Ala Leu Tyr Leu Ser Val Tyr Ala Lys Asp Ala Ser Ala Glu Val Val	
945 950 955	

-continued

```

aaa ccc gac caa atc ata act tca tcc aac aat gac cag cag atc caa    3048
Lys Pro Asp Gln Ile Ile Thr Ser Ser Asn Asn Asp Gln Gln Ile Gln
960                      965                      970                      975

aaa cca gtc agc tct gag aga agt ttg gag aag gac tta gca aaa gct    3096
Lys Pro Val Ser Ser Glu Arg Ser Leu Glu Lys Asp Leu Ala Lys Ala
                      980                      985                      990

atc gtc aaa atc tgt ggg aaa aag atg att gag cct gaa gca ata gaa    3144
Ile Val Lys Ile Cys Gly Lys Lys Met Ile Glu Pro Glu Ala Ile Glu
                      995                      1000                      1005

tgt ctt tca att ggt gct cgt gag ctt cca cct cca tct aca gtt    3189
Cys Leu Ser Ile Gly Ala Arg Glu Leu Pro Pro Pro Ser Thr Val
                      1010                      1015                      1020

ggt tct tca tgc gtg tat gtg atg cgg aga ccc gat aag aga ttg    3234
Gly Ser Ser Cys Val Tyr Val Met Arg Arg Pro Asp Lys Arg Leu
                      1025                      1030                      1035

tac att gga cag acc gat gat ctt gaa gga cga ata cgt gcg cat    3279
Tyr Ile Gly Gln Thr Asp Asp Leu Glu Gly Arg Ile Arg Ala His
                      1040                      1045                      1050

cga gca aag gaa gga ctg caa ggg tca agt ttt cta tac ctt atg    3324
Arg Ala Lys Glu Gly Leu Gln Gly Ser Ser Phe Leu Tyr Leu Met
                      1055                      1060                      1065

gtt caa ggt aag agc atg gct tgt cag tta gag act cta ttg att    3369
Val Gln Gly Lys Ser Met Ala Cys Gln Leu Glu Thr Leu Leu Ile
                      1070                      1075                      1080

aat caa ctc cat gaa caa ggc tac tct ctg gct aac cta gcc gat    3414
Asn Gln Leu His Glu Gln Gly Tyr Ser Leu Ala Asn Leu Ala Asp
                      1085                      1090                      1095

gga aag cac cgt aat ttc gga acg tcc tca agc ttg agt aca tca    3459
Gly Lys His Arg Asn Phe Gly Thr Ser Ser Ser Leu Ser Thr Ser
                      1100                      1105                      1110

gac gta gtc agc atc tta tag tttgaaacat tagctgtgtt tgtagttgat    3510
Asp Val Val Ser Ile Leu
                      1115

catctctatg tgcaattgaa caagtcagtt tgctagaact agagtagatt actaagaaac    3570

catgccgttt ttcattttga gatattgcaa aacggcatgc agttcgggta agtcggatgc    3630

cgcaattacc aattttgggt cagtctgtgt aattgtcgtt tca    3673

```

<210> SEQ ID NO 3

<211> LENGTH: 1118

<212> TYPE: PRT

<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 3

```

Met His Trp Ile Ala Thr Arg Asn Ala Val Val Ser Phe Pro Lys Trp
1                      5                      10                      15

```

```

Arg Phe Phe Phe Arg Ser Ser Tyr Arg Thr Tyr Ser Ser Leu Lys Pro
20                      25                      30

```

```

Ser Ser Pro Ile Leu Leu Asn Arg Arg Tyr Ser Glu Gly Ile Ser Cys
35                      40                      45

```

```

Leu Arg Asp Gly Lys Ser Leu Lys Arg Ile Thr Thr Ala Ser Lys Lys
50                      55                      60

```

```

Val Lys Thr Ser Ser Asp Val Leu Thr Asp Lys Asp Leu Ser His Leu
65                      70                      75                      80

```

```

Val Trp Trp Lys Glu Arg Leu Gln Thr Cys Lys Lys Pro Ser Thr Leu
85                      90                      95

```

-continued

Gln	Leu	Ile	Glu	Arg	Leu	Met	Tyr	Thr	Asn	Leu	Leu	Gly	Leu	Asp	Pro	100	105	110
Ser	Leu	Arg	Asn	Gly	Ser	Leu	Lys	Asp	Gly	Asn	Leu	Asn	Trp	Glu	Met	115	120	125
Leu	Gln	Phe	Lys	Ser	Arg	Phe	Pro	Arg	Glu	Val	Leu	Leu	Cys	Arg	Val	130	135	140
Gly	Glu	Phe	Tyr	Glu	Ala	Ile	Gly	Ile	Asp	Ala	Cys	Ile	Leu	Val	Glu	145	150	155
Tyr	Ala	Gly	Leu	Asn	Pro	Phe	Gly	Gly	Leu	Arg	Ser	Asp	Ser	Ile	Pro	165	170	175
Lys	Ala	Gly	Cys	Pro	Ile	Met	Asn	Leu	Arg	Gln	Thr	Leu	Asp	Asp	Leu	180	185	190
Thr	Arg	Asn	Gly	Tyr	Ser	Val	Cys	Ile	Val	Glu	Glu	Val	Gln	Gly	Pro	195	200	205
Thr	Pro	Ala	Arg	Ser	Arg	Lys	Gly	Arg	Phe	Ile	Ser	Gly	His	Ala	His	210	215	220
Pro	Gly	Ser	Pro	Tyr	Val	Tyr	Gly	Leu	Val	Gly	Val	Asp	His	Asp	Leu	225	230	235
Asp	Phe	Pro	Asp	Pro	Met	Pro	Val	Val	Gly	Ile	Ser	Arg	Ser	Ala	Arg	245	250	255
Gly	Tyr	Cys	Met	Ile	Ser	Ile	Phe	Glu	Thr	Met	Lys	Ala	Tyr	Ser	Leu	260	265	270
Asp	Asp	Gly	Leu	Thr	Glu	Glu	Ala	Leu	Val	Thr	Lys	Leu	Arg	Thr	Arg	275	280	285
Arg	Cys	His	His	Leu	Phe	Leu	His	Ala	Ser	Leu	Arg	His	Asn	Ala	Ser	290	295	300
Gly	Thr	Cys	Arg	Trp	Gly	Glu	Phe	Gly	Glu	Gly	Gly	Leu	Leu	Trp	Gly	305	310	315
Glu	Cys	Ser	Ser	Arg	Asn	Phe	Glu	Trp	Phe	Glu	Gly	Asp	Thr	Leu	Ser	325	330	335
Glu	Leu	Leu	Ser	Arg	Val	Lys	Asp	Val	Tyr	Gly	Leu	Asp	Asp	Glu	Val	340	345	350
Ser	Phe	Arg	Asn	Val	Asn	Val	Pro	Ser	Lys	Asn	Arg	Pro	Arg	Pro	Leu	355	360	365
His	Leu	Gly	Thr	Ala	Thr	Gln	Ile	Gly	Ala	Leu	Pro	Thr	Glu	Gly	Ile	370	375	380
Pro	Cys	Leu	Leu	Lys	Val	Leu	Leu	Pro	Ser	Thr	Cys	Ser	Gly	Leu	Pro	385	390	395
Ser	Leu	Tyr	Val	Arg	Asp	Leu	Leu	Leu	Asn	Pro	Pro	Ala	Tyr	Asp	Ile	405	410	415
Ala	Leu	Lys	Ile	Gln	Glu	Thr	Cys	Lys	Leu	Met	Ser	Thr	Val	Thr	Cys	420	425	430
Ser	Ile	Pro	Glu	Phe	Thr	Cys	Val	Ser	Ser	Ala	Lys	Leu	Val	Lys	Leu	435	440	445
Leu	Glu	Gln	Arg	Glu	Ala	Asn	Tyr	Ile	Glu	Phe	Cys	Arg	Ile	Lys	Asn	450	455	460
Val	Leu	Asp	Asp	Val	Leu	His	Met	His	Arg	His	Ala	Glu	Leu	Val	Glu	465	470	475
Ile	Leu	Lys	Leu	Leu	Met	Asp	Pro	Thr	Trp	Val	Ala	Thr	Gly	Leu	Lys	485	490	495

-continued

Ile	Asp	Phe	Asp	Thr	Phe	Val	Asn	Glu	Cys	His	Trp	Ala	Ser	Asp	Thr
			500					505					510		
Ile	Gly	Glu	Met	Ile	Ser	Leu	Asp	Glu	Asn	Glu	Ser	His	Gln	Asn	Val
		515					520					525			
Ser	Lys	Cys	Asp	Asn	Val	Pro	Asn	Glu	Phe	Phe	Tyr	Asp	Met	Glu	Ser
	530				535						540				
Ser	Trp	Arg	Gly	Arg	Val	Lys	Gly	Ile	His	Ile	Glu	Glu	Glu	Ile	Thr
545					550				555					560	
Gln	Val	Glu	Lys	Ser	Ala	Glu	Ala	Leu	Ser	Leu	Ala	Val	Ala	Glu	Asp
			565					570						575	
Phe	His	Pro	Ile	Ile	Ser	Arg	Ile	Lys	Ala	Thr	Thr	Ala	Ser	Leu	Gly
		580						585					590		
Gly	Pro	Lys	Gly	Glu	Ile	Ala	Tyr	Ala	Arg	Glu	His	Glu	Ser	Val	Trp
		595					600					605			
Phe	Lys	Gly	Lys	Arg	Phe	Thr	Pro	Ser	Ile	Trp	Ala	Gly	Thr	Ala	Gly
	610					615					620				
Glu	Asp	Gln	Ile	Lys	Gln	Leu	Lys	Pro	Ala	Leu	Asp	Ser	Lys	Gly	Lys
625					630					635				640	
Lys	Val	Gly	Glu	Glu	Trp	Phe	Thr	Thr	Pro	Lys	Val	Glu	Ile	Ala	Leu
			645						650					655	
Val	Arg	Tyr	His	Glu	Ala	Ser	Glu	Asn	Ala	Lys	Ala	Arg	Val	Leu	Glu
		660						665					670		
Leu	Leu	Arg	Glu	Leu	Ser	Val	Lys	Leu	Gln	Thr	Lys	Ile	Asn	Val	Leu
		675					680					685			
Val	Phe	Ala	Ser	Met	Leu	Leu	Val	Ile	Ser	Lys	Ala	Leu	Phe	Ser	His
	690					695					700				
Ala	Cys	Glu	Gly	Arg	Arg	Arg	Lys	Trp	Val	Phe	Pro	Thr	Leu	Val	Gly
705					710					715				720	
Phe	Ser	Leu	Asp	Glu	Gly	Ala	Lys	Pro	Leu	Asp	Gly	Ala	Ser	Arg	Met
			725						730					735	
Lys	Leu	Thr	Gly	Leu	Ser	Pro	Tyr	Trp	Phe	Asp	Val	Ser	Ser	Gly	Thr
		740						745						750	
Ala	Val	His	Asn	Thr	Val	Asp	Met	Gln	Ser	Leu	Phe	Leu	Leu	Thr	Gly
		755					760					765			
Pro	Asn	Gly	Gly	Gly	Lys	Ser	Ser	Leu	Leu	Arg	Ser	Ile	Cys	Ala	Ala
	770					775					780				
Ala	Leu	Leu	Gly	Ile	Ser	Gly	Leu	Met	Val	Pro	Ala	Glu	Ser	Ala	Cys
785					790					795				800	
Ile	Pro	His	Phe	Asp	Ser	Ile	Met	Leu	His	Met	Lys	Ser	Tyr	Asp	Ser
			805					810						815	
Pro	Val	Asp	Gly	Lys	Ser	Ser	Phe	Gln	Val	Glu	Met	Ser	Glu	Ile	Arg
		820						825					830		
Ser	Ile	Val	Ser	Gln	Ala	Thr	Ser	Arg	Ser	Leu	Val	Leu	Ile	Asp	Glu
		835					840					845			
Ile	Cys	Arg	Gly	Thr	Glu	Thr	Ala	Lys	Gly	Thr	Cys	Ile	Ala	Gly	Ser
	850					855					860				
Val	Val	Glu	Ser	Leu	Asp	Thr	Ser	Gly	Cys	Leu	Gly	Ile	Val	Ser	Thr
865					870					875				880	
His	Leu	His	Gly	Ile	Phe	Ser	Leu	Pro	Leu	Thr	Ala	Lys	Asn	Ile	Thr
			885						890					895	
Tyr	Lys	Ala	Met	Gly	Ala	Glu	Asn	Val	Glu	Gly	Gln	Thr	Lys	Pro	Thr

-continued

900					905					910					
Trp	Lys	Leu	Thr	Asp	Gly	Val	Cys	Arg	Glu	Ser	Leu	Ala	Phe	Glu	Thr
915					920					925					
Ala	Lys	Arg	Glu	Gly	Val	Pro	Glu	Ser	Val	Ile	Gln	Arg	Ala	Glu	Ala
930					935					940					
Leu	Tyr	Leu	Ser	Val	Tyr	Ala	Lys	Asp	Ala	Ser	Ala	Glu	Val	Val	Lys
945					950					955					
Pro	Asp	Gln	Ile	Ile	Thr	Ser	Ser	Asn	Asn	Asp	Gln	Gln	Ile	Gln	Lys
965					970					975					
Pro	Val	Ser	Ser	Glu	Arg	Ser	Leu	Glu	Lys	Asp	Leu	Ala	Lys	Ala	Ile
980					985					990					
Val	Lys	Ile	Cys	Gly	Lys	Lys	Met	Ile	Glu	Pro	Glu	Ala	Ile	Glu	Cys
995					1000					1005					
Leu	Ser	Ile	Gly	Ala	Arg	Glu	Leu	Pro	Pro	Pro	Ser	Thr	Val	Gly	
1010					1015					1020					
Ser	Ser	Cys	Val	Tyr	Val	Met	Arg	Arg	Pro	Asp	Lys	Arg	Leu	Tyr	
1025					1030					1035					
Ile	Gly	Gln	Thr	Asp	Asp	Leu	Glu	Gly	Arg	Ile	Arg	Ala	His	Arg	
1040					1045					1050					
Ala	Lys	Glu	Gly	Leu	Gln	Gly	Ser	Ser	Phe	Leu	Tyr	Leu	Met	Val	
1055					1060					1065					
Gln	Gly	Lys	Ser	Met	Ala	Cys	Gln	Leu	Glu	Thr	Leu	Leu	Ile	Asn	
1070					1075					1080					
Gln	Leu	His	Glu	Gln	Gly	Tyr	Ser	Leu	Ala	Asn	Leu	Ala	Asp	Gly	
1085					1090					1095					
Lys	His	Arg	Asn	Phe	Gly	Thr	Ser	Ser	Ser	Leu	Ser	Thr	Ser	Asp	
1100					1105					1110					
Val	Val	Ser	Ile	Leu											
1115															

<210> SEQ ID NO 4

<211> LENGTH: 80393

<212> TYPE: DNA

<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 4

```

aagcttgcat ctctcaagaa ctctgacaat cagtcaaagg gattttctct gagccttata 60
agcaaaacttg aggtaagatc atgtttctcc aaactcaaca ctttcttctc gattttacaa 120
ttccatcata aacaaaggtc ttctctgttc cttctccagg tcattgttgt cagtcttatt 180
taatctctca aatcgaacat ttaggatatg taataacatg attaagccat tgtcctagac 240
caaatacagt ttgcaacatt gtagatgtaa gtgcttttgg ttttgctcag cgtaacgatt 300
tgtctcttgc tctactagtc aggaactagc gaactccttg gatttagaga cccggcaaaa 360
cttatcaggc ttcattgatg ctgttgagaa aatactcttg cagcaaaccg gtgaagaact 420
caagtccaat gaatcctccc aaaagtgagt accaagaacc acctcaagag tttgtgcagt 480
ttctatctcc ttattgtttt tgtcttgggt tggtatctgc aactcttctg tgtaattact 540
ttggaaattg gaattgtatc aatactcttg cttatgtcct caagtttctt ttatatatca 600
atatacctaa acacattatt atcttactct ccgcatata taatcagaga actaatgaaa 660
taatctccaa ataactctca aatacttctc cattctgggt aatggaggaa aacaccaacc 720

```

-continued

atatataact tacggagaag taatctccaa ataactctcca aatacttctc cattcttggtt	780
aatggaggaa aacaccaacc atatataact tagtgattac ttctccgtaa ggtgacaaca	840
aaactcagag gtctctttaa atggaagaag ctaaagttgt tctttgagtg ttttaactgt	900
tacttttagtc aatttaagga agtcaatatg gcttagtaaa tcaattaaga aaaccattta	960
aaatctacca agtttagtct aaaagactgt tgtctacact tagctaaacc aaaaaacaga	1020
accagaacca gcaccaaata aacctaataa aatttccatt ttaaatataa ccacgttcag	1080
attgttcttg tcttcaactg agattcttgt gcagtggaa cttcttgctt tgtagaaagt	1140
ttttcaatcg ttgtcttcaa gacaggtttc aatctttctc ttacaagagt gacctcgcgc	1200
agtttctgat tcaaccggtt caagatcgca ggtaaaactc tttgatgatt ttctttcact	1260
atttcaagca tatctctacc atataaattt tctgcacatc tgtagaactt cctccaaga	1320
tctttcaggt cctctctccc gttaataact tcctctgcat tcttcgctgc acttgtaga	1380
aatcccacca acatatctag ctcgaaatc tgatcttccc agttgtacat atctttttcg	1440
atatcggtta ggttcttctg tgcttgata tcataactgt tcacaacca agtggtgttt	1500
agtactgtag acgagaccgg tgtctgttct tccttgggaa ttagcttata gttaggtgtc	1560
gctcgttcag tctttctcct tttcttcaga ggtctctcag gttggtctga ttcttttga	1620
ggcaacattc ctctcttccc ctttccctcc aacttgatct tatccgactc tgtgtcttg	1680
gggttatcat caactttaac cctagttcca gcctcatctt ctccatgacc gtaggcaatg	1740
tctacgaacc gttgataaag attcttgtat ttatcgatca aacgttgaag agacttaacg	1800
aattggctgt aatcgattct atgagactta aagcttttac acaagcaagt gtaaaactctc	1860
aattccgacc gagttagtct ctccctgata gccttcaata actctgtcaa cttctgtttc	1920
gcttctaaac ggtccttcat ctttctctcc acatcgctct taggtttata aactatccgt	1980
atagcaggta tcatcttttg ttcatataaga agaatagggt tctcttccgg cttctcatca	2040
tgatcaatct ttcgattaag gttattgtct ttaagcctcc taagttctct agtgagttcg	2100
ggatttctga gtagacgtac cgaagatcgt aaccttgtct gtaccgatga agactgtgtc	2160
ttgacaccgt tcgtcttcat caactcatca ggtcagcca tatcttctga atctcgatag	2220
cgtgtgagcg aaacaaacga cttagggtag agtatatata aggtgttttc gttttccttt	2280
ttttttgagg agattaacta tctagatttg ctgaaagtac gcagattact tgttttctaa	2340
ccaatggtaa taaatcaatt tatttaaatt agttttcttt caactgatta aaaacaatta	2400
aacgcaaatt aaattttgag aatcagtatt tgtcttatat ataaatttat acatgaaaaa	2460
tataaaaat ttacaacaaa aatcatcatc ttatgaagag ttcatgaatc tcatcaaatc	2520
tctctgtttt ctcttgcaac aaaactcacg gtgatgtcgt ctatacaaat cttgttgctt	2580
tttatattag aatttgtag attctagtag gctggaaggg atcttgaagg atttggtgtt	2640
agagatttat gtcattctca ttttaggagg tggagaagtc aaaagagaat caaccatcta	2700
gcagctcaat tggaaagtat tgaagtgatt gatgatgaca tgttttatag tttgtttct	2760
attatatgaa gcttggtccc catctgaaat tttctatatt tctcttagtt ctttctctg	2820
gctaacatac ctaagatact gtatgatag catccaacct caaagttata tttgttgaa	2880
tcttaagggt tctgtcataa cctatcttaa aactgcaggc aggtgaaggt acactcttca	2940
gtgctagaca caagatcttc agtccagtg gttgtcgcca aagtgaaga tataattatt	3000

-continued

tctaacaact ttgagaaaag attttgtgat gagttcgaaa acaattagtt gageccttccc	3060
tttcagtagc ttaagaggct gaatttttct ttcgtagcta actcatgtca ttgagttctc	3120
tgagcaagcc aattgactgc atgtttttgtg atcctatggc aggcacacgt tcgaatacta	3180
cgacaaagat gaaacaattg tgactcatat agctggaggt attgatgcat ttttaaaggt	3240
ctctgatggt tggccactgc tgaatacccc attaaatgga agaagctaaa gttagtgttt	3300
taactattac tatatagtca atttaggaaa gtcaatattg cttagttaat caattaagaa	3360
aaccatttaa aatctacaaa ctcaagtcta aaagattgtt gtctacagtt agctaaacca	3420
aaaaacagaa ccaacaccaa aaacaaatta acctaatagaa aattacattc taaattaaac	3480
caggttcaga ttgtttgtgt cttcacttga gattcttgtg cagtggagg tcatttggt	3540
tatgtgtagg tttgggattt ggctcgagtg atctctacat gtggtgagtt tgtaagaaaa	3600
tgaagttcgg gaatcgaaaa aaacgcttca taaaccaggt aaagctacaa gagatggatg	3660
ttacaactaa ataacacttg cacttgaaaa ctaaacaaag tttttatata aagagtgcc	3720
caagtggagt ggtggtatata acagagtaaa aatgagacaa tgtaaacaaag agaaggccag	3780
tcaagctaga cgagtggatc tgtcaaaggc ctcaatgact tgataacat atctatgctt	3840
acacaaatga atactcggg tgacgaggac ggaggtgaga ctgtgcggaa gatcaatacc	3900
ggtgcgttta aaaccccaaa tgagacgctg aactacatag tccccacac gatctttcac	3960
caaattctca agtaactat caaactcaga ggcaatcaac gtaagacacc tttcatcaag	4020
catatcaaga aacttctgca cagggtaatt agagaaggag cctgatgaaa gggacacaaa	4080
gtgaccagag aacttgttta ccctttctta tcaaaggcag ctatcaagca atcattacca	4140
aacatggaga gcaaaagtga accagcatga tgtataagca gtaatagggt cttgttacca	4200
atctcattac gggagagtgc aacaaaagtt gcaccaactt gaatcaggaa cttgaacaaa	4260
agttgcacca ttctcatatt gacttgattt aaatattaac aaaaaattat atctatgtta	4320
ttgtagagaa gaaaaagaa aaacgtatat aattaaagat taggaaactc aagtcctttg	4380
aagaaaaata agaaataaaa aaacccaagc cttaaagtaa ccttgataaa tcgtcatacc	4440
ttttctcaac gatactattc tataatacat atttaactag ataagtaac cgcacctgt	4500
gcagaaaaaa aaattattaa ataaatttaa cattggatta aatatagaag ttattctagt	4560
ggccaaaatc tataatgact ttagggttta tatagatata tgtgatatcc attatatgat	4620
caataatata ttggcaaata tagaacagaa catttatatt ttattaatat gtttagtata	4680
taaaatgatg attcttttga tattataact tattctttat aagataaata atattgctaa	4740
ataattcgct taacgcatta gtaatactat ttttatgtat ttttgtaaag gtgaaaacga	4800
ttttatgtaa tatatgtttg tgaaaaattt gtagtattgt ggatgaatat gatattggt	4860
ttctatttat aggaaaaatg gtttttacct atgtaggaaa ttgtcttttt tccttgcaat	4920
tgtttttagaa atttgttaca agatattaat gcgggaaaaa tgtaaaatttt tgtttcggaa	4980
tatctatagt aaaatatgat tttgttccca tattttctaa tattttactt atgtattatt	5040
agttaatgat atacgcagtt attctaattg gcggcagatt tgctatatat acttgattag	5100
gggtagtttt cttttgtaga cacatcaatc tgtatgaaat ttacaggatg aaaaacactc	5160
ggccacgaag agccaaaatt taggtttgcg ttttgctgtg tgaagagaaa atatgagtca	5220
agtatggggt tgaaattcag taatttgaga agagttcatt gactaagtat tgacgtgat	5280

-continued

atggtagaat atacaataaa ccatatactt acataactaa tcataaaagg aattgatagt	5340
cttatactac atgcactaag tcactaacca acctggctta gcattttctaa acacgaattt	5400
acaaactggt aagatctcgt ggcgtctgac aacatccaat aatgacaatg gcgatatgaa	5460
gaaacaaaag cagagggcta gaccaagcat gaagtctttt aacgttattt ttgtttaaga	5520
gactgaagaa gaaagaagga ttctatggag attggagaga gtggttgagc tcgggcaaga	5580
gaataaacca atgtaccact ttagatacaa tgttttgagc tttctgaaga gaaaaagaa	5640
gagaaaaacg agactaaaca attattagtt tttcatagtt atatttaact aaaaactgat	5700
tgttttcggg gtatgtgtat aggtacccaa tccattggct ttcgagattg aaattaatta	5760
agaaactcat aattccatcac atacatacac aatatatgga cgaaatttga gtaaaattca	5820
ctaagcttct gtccataaaa ccttaagtga aaaagggaga tcaatcccaa gaatctattg	5880
aaaatatgat aatcttaagt aaaaagcaa actcaagaat catagcttta tggatagaac	5940
aaattataac gctcttgaa gaaaacaaga atcatccggg tcctattctc taccaaaaca	6000
aagagtgcca ccacacagca aacttataca cttcgttttc atcgagaatg ttcattgtata	6060
tgaggaaatc ttgagattta gggcaacgctc ttgggaatct cctacttata ataaattgaa	6120
atcaagaaaa aatataactc atgcaatgat cagattccag attcaagatc atataataaa	6180
tcaaaccccta tatcttcctg aaaacaccac aagacatata agaacaatca ctacgagagc	6240
agtgaattta gtagaatca atacaataca taagggttta tcaattgctt gagtaacaat	6300
gaacccttaa acaataatgt cttagagataa aagatatgta aatcgaaaag aaataaaccc	6360
taagagaatg atcgaaacta attgatatcg gaatcatcaa aatcatcctc gaacttggtg	6420
aagaaatcgc tgtcttcggg cttcatttga aagggtgaaa ttcctccaac ttcaaactgc	6480
atatccatgg gcccgaaatc gtcttcacatc atctccatct gcattaggcat cgagtatcct	6540
ggatactcca tctcaccag aaagtcacgg tacaccatcg gcctcatcgg gtcttggtcc	6600
caatatctat ccacgttttc gcttcctttg cgtgaaaatc aaacaaaacc taaggataaa	6660
gagaaagaga ggtgttctac tttttgaggg ttaggacacc cttaagtaat atccgggaca	6720
ttataaaggc ccattagtag aggagattgg tccattatcac agtcaagata aattaagaga	6780
tttctttctt tatgtgtcat caacaattag tttttaagat ttttttctaa aacaaagatt	6840
ttgtattaga ttatatccaa catatgtagt atctactaaa ataagccttt ctaataatta	6900
taatgacctat tataaagaaa aaagtatatt tatttttatt tcttattttt ggtattgatt	6960
aaaatataca tatctaactg ctatataaga aacatatgtg ttttacaatt atcaaaactg	7020
attacttcca acagaaggaa acatacaaat ttgttaacat gtattagctg aagctcttca	7080
ccagtaaaat cattgcgaag atatatgatt caaaattggg aaaaatcttc cctagtagag	7140
aacaactcaa tgacaatgtg aggacaaaag cagagaagaa taaacactaa aagagaacc	7200
ttttgtatca tagtatttac aacatttaac gatcataatt ttgagcattt tcaggattga	7260
tcaaaatcta tatatatata tatgataaga gacttcattt tgaataatca aataagatac	7320
aaacaacctt aatccgtata atagtcatta atttattagc ttgtggtgga ggtggttttt	7380
ctgccacctt taggattctc acacgggaca acatgcaagc atgcatattg tcagaaacat	7440
actcaagaaa gccaaagaaa tgaccggcct cgagaatgac cagtatgcaa ggacgctata	7500
ctatccctaag tcctaactcc taatccatag aaaaatactc tgaacacgtg gtacgtgact	7560

-continued

tgtcttctta aacgatactt ggctagaaac ttgcagtttc gtaatatgta gccgtccaaa	7620
cttttacttc aaagtcaatc tcggttaaga aaacatgttc tgtccaccaa ggtaataaag	7680
atcaagagta ctgtatcgac tataatttca tcaaatatat cagcacttga cagggtcacc	7740
tacttatgac ggagggtgc gatattggtt acaaacacat cctcgctgac gatcattcgg	7800
ttataacctt tcaataaatt tcaattttct ctaatccttt caaatagatg acaaacacgt	7860
cttttggtga ataaatcttt cattcatacc aaaaataaat aaagtgcata tgtgcaaatg	7920
attgcatata aggaaaaata ataaacgtac ctatgtatct taagtgcacac tttctaggat	7980
ttctcagcca taatgattcc accgttcttc atggtgctgt gagaaagagg ttcatgttag	8040
attctatcca ttaggtaga aatttttgtt ttctcaaatc atcgatatgg tttatttagt	8100
tcctcatgta agaaaatact ttgtttatga aagttttctc tgtggattga tttttcacac	8160
agctccgttt cttcggttcc ctttggaatg tacatgtaag aagtatgatg tattccattt	8220
tcaatatggt ttatttaaca taaaatctag attttcgacc aataaccttt tcaagtaaaa	8280
gaccagaaca agtcaattta gtcatttttc gaacatctaa aacttaaaact tatgatgatg	8340
ttacgactca atctttttta aagttccctt attacatcac tagagcaata gtttttgtgt	8400
aacttttagt ctaaaaagga gcgagtctaa taagtgcata aacaaggat atactaacag	8460
aatttgggaa aaatagttaa attaaaaaaa aaggtgcaca agaacacaaa catagataga	8520
tcactcaaaa caaaacaaaa caaaacaaaa cccaaaaaaa aatccaatca agatttgtat	8580
ttcaaaagca aaatcccaat aatacttaac tggatataaa tttcaaaaca gcttgaaaat	8640
caaacggctg gagggtagg cgcacgaggt actggagatc caggcactcc aatggaatca	8700
tcacatttat catggtagat ataagcaaaa ccaccatgac ctgctaaatc cattcctcgc	8760
atttcatgct gtcctgagat cctaagcaaa ttgagctttt tgaggatgaa gaagagtgtt	8820
cccattgtgg cactaaccca tccaacaatc acaattatct gaaccagttg tgctcccaat	8880
agcttccttc ctccgcccac aaatagcccg tagtgccttc ctgggctcgc gccgtaaac	8940
tcatttatat acttctcttt tgcaaacagt cctacaaata tcaaacccca agcaccacac	9000
cctccgtgta gttgtgcggc ctcaagtgga tcgtcatatt ttaagagctc tgcgagcttg	9060
ttgcatccga taaggacgag ggaagccacg aagccacata cgatecgtgc ccatggatca	9120
accacagagc aacctgccgt tatggccgca aacctccga gtaacctggt gcaaacgtca	9180
gttacgttcc agtggcctga taggagacgt tttccgaaga gtgtggttag agccgctgtg	9240
catcccagaa gtgtagtgtt aaccgcggtg cggcctattc cactccattg gccatagtgt	9300
gaaccagaat ttagaggat gatgatcttg gtgaaggaa cagggttgaa cccgtacca	9360
caaaccaga gaaggaaggt ccctaagacg acgagtgagg cagagtggcc tcgcagagca	9420
atagcatgac ccccatcagg aaaccgacca atccgagggc cttcaataag ggcaccccat	9480
aatcctgcga taccaccaac catgtgaaca acacctgacc cagcaaagtc gatggctcca	9540
gtgcaaaaca aacggctctc tgaacggaag ggactagccc atccatccgg agaccagaac	9600
cagttagaga caactgggta aacaaacccg gtcaaaaaag aagagtatat caaatacgcc	9660
acgaacttag tcctctcggc aatggagcgg ctggtgatcc cagcggctgc gattgcaaac	9720
gcccattggt agaggaagaa ggagtaatcc gaggtgagag tcggaaggtt ttgaagacca	9780
aagttgtgtc ttccaatgaa tccatcggag gattcaccaa aggcacaaatg ataaccacaa	9840

-continued

aggtagtaga agagtccctcc ggctgcagca tcaaggacat tagtgagcat gatgttcac	9900
gtgttcttag ctctaaccga gccagcacia agcatagcga agccgagctg catcgcaaa	9960
acaagatagg cagagaagag gaggtacgtg ttgtctatag cgtaggctgc atcggtaa	10020
ttgtgttaa cggaacccaa ctggccgcaa atgtagtcag ccgctgccgt ggcatttgg	10080
ccgagtaggg ctgagagatc agccgcagag caagtaatag ctccctgacat gttttaga	10140
cttgagagag gaaccacacg aataagaaag attgtctctt gtgcagctat aaaaagaaa	10200
agaggaggag tttctgattc tggttgacc aatcaaattc gtagttaagt ggagagatat	10260
tcattattaa attattagag atggaatcta acaaatctta taaactactt gcaatatac	10320
agaatcttat acggttttgg aaatttagtg attggcatgt actttttatt tggatatcaa	10380
cgcaatagtt tcttattata aaaaaagaga gagagggagt ttaaattccc ttatcacat	10440
cttcttttag caatctaagc aagctgatgg aagcaaattt aacccaaaat atgaatgac	10500
acattataat attgcaaaaa gatattaaaa ataaaaata tgatctattc aattgtagca	10560
tatgtaagca aaacatattt gaccaaggaa tctaagaaga tccccgtaga gagttggctg	10620
ccgttttcga ctccactgt cacaccacct cacacgaag ggaagctcat tcgtgctttt	10680
actttcgggg aaaaaacgtt tacatgtacc tgagtatctt ctaatttcta ttttaata	10740
attatataaa caaaattagt acttctaggg tttctagcaa atcaatttgt ggcaccttg	10800
tagttgtaga tcagcaactc tttttttggg aatcaggtca aaattcagca ttcttaata	10860
gaaaatctta aatataatga acttgtaaag tattaatata attgctgacg aaactgc	10920
tgataacatg gtttagcggg aaccacgttt tattagtga aaatttcact tgcgatct	10980
aaattcaaca aatcaatggc gtttaatgtg attctatata ttagcaacca caatgg	11040
ataagaatga gtatagttaa tagataaaat ttagcaata agtaatgttc caactatg	11100
tatatgtaaa cttaatcatt ataagacata cttgtaaatt cactagtga tatatac	11160
cgaaggagaa taaataagtt tactagatta caaattagct gatggaaaac gaagaagt	11220
acgatttaga atacattaat attgggattt gacactcagg aaaatagaca aaaccaa	11280
ttttttttaa tctaattgtc tcgacgaacg ataaatattt taattgcaac gggagct	11340
aaaatcatac agagtaacct gacatcattg attaccacat aacacaaatt aattatt	11400
taatacgttt aatttactga ttaattatag tgcataattt attttagttt aggtca	11460
acaattaaact ataagaacta aaaatttact aaatacgaaa agccacattg gga	11520
tgttttatga aattatgcaa cggcgcatg cggcgtctct agctatcctc tatgcat	11580
agactataaa atatttgttc gtgaaattac tttttatgtt ctataaatag gattgag	11640
ttaataacca agaacatcat aaaaaaact ctatagataa tcataataat tgcatta	11700
tatttggcta tttgccaaaa caatttaact caatactggg tgatttgcaa tgctag	11760
ttttaagaa ccatcttaga ttctaagtag taaacaaata tataaatcaa gtggtagt	11820
tttagcgatt cccattctag gaagcgttgc cttgtcggtc tagaacaaga aaagatt	11880
tacaaaacat tgaattatct acatcaattg ttatttacct agtcgaaata gtgttat	11940
ttttttcttt attattagaa aatttgggtg atacactagt acttgacgac tataatc	12000
ctaagaaatt agttgggtgg taaaatgtcg gattaccgat agataattaa aataaa	12060
aactctaagt ctaaaactata cgggcaagaa tatttaactg gttcgtaatt atttta	12120

-continued

tttttcttac	tacgacccgt	taaaggtaat	caaaattctt	gtacaatgtt	ctgtctttgt	12180
ttgggtgaca	tatgaaatat	acgtttgaga	atattttaat	acagtttgac	gactaatatg	12240
cacaaaaaga	tgatgtgcaa	ggtataagct	actggaattg	catcgtaatt	ccatgattct	12300
acaggtctcg	aaagtgtata	caaacaagga	atatatgtta	taaaaatttt	tgttttagcag	12360
aatttgaaat	aaataaaaaa	tactattgaa	atattaggtg	ctaaaattta	tcaaccacaa	12420
aatatgggtc	atactccttt	gtttaaagtt	atcgaggacg	caagaatttg	atttatattt	12480
taattttagt	ttgtacgatt	aatagcaaca	attgttaggc	taataaaacc	cattaggcctt	12540
agttcaatag	ttctaacaa	gtgaccagtt	acatagatgt	aatgacatgt	catatttctt	12600
ctagtattat	ttattaacca	acatcggtca	agtaaaagat	aatatctacg	tgtacctaa	12660
cgtggctagt	taactaatg	gaaaaaatcg	ctttacgtgt	acataaggga	atgactaagt	12720
tctgcttttg	gtacgaaatg	gtacacaaat	gatttttcag	aatttatatt	attctaagaa	12780
caacttttgt	gtactccaat	ttcctaattg	cgtcggcaga	acaaaaacct	aagcaaagtt	12840
gtatacagta	tatgcataac	tctccttttt	ggatcaccaa	aagttgtata	tataaatctt	12900
taagaaatag	cataaacatt	agccgaggca	tgtaaaatta	aaattcaaga	aaaaaaaaaa	12960
aaacataaac	taaaacactt	ataaacacaa	aaagacgata	ggtcacttaa	aacaaaccaa	13020
aacatccaaa	tatttatatt	tcaaaaccaa	aacccaaaca	atacttaaac	ggtaataaat	13080
tacaaaaaag	ttgaaaatta	aacgcgagga	ggagtagctg	atcgaggga	aggagatcca	13140
ggatccactc	tatgagactc	atcatcatta	tcatggtaga	tataagcaaa	gccaccgtga	13200
cgtgtcatat	ccatcccttg	catttcattg	tgtcccgaga	tcctaagcag	attgagcctt	13260
ttgaggatga	agaagagtgt	tcccattgtg	gcactaacc	atcctacaat	cacaagtatt	13320
tgaaccaatt	gtgtcccaa	cagcttcctt	cctccgccca	taaatagtcc	atatggcctt	13380
cccggttg	cgccataaac	ctcgtttaga	tacttctctt	tggcaaaca	tcctacgaat	13440
atcaacccc	acgcgccaca	ccctccatgt	agttgggctg	cctcgagtgg	atcatcatat	13500
tgtacaagct	ccgcgagctt	gttgcatccg	ataaggacga	cagaagccat	gaagccgcac	13560
acaatcgctg	cccatggctc	tacgacggag	caacctgcgg	ttatggccgc	aaaccaccg	13620
agtaacccgt	tgcaaacgtc	cgttacgttc	cagtggcctg	ataggagacg	tttaccaaa	13680
agtgtggtta	gagctgctgt	gcatcctgag	agtgtggtgt	taaccgctgt	acggccgatt	13740
ccgtccatt	ggccgtagtt	ggaaccagaa	ttatacgga	cgagtatctt	agtgaaggaa	13800
ccggggttga	aaccatacca	tccaacat	aggaggaagg	ttcctaagac	tactagcgag	13860
gcagagtggc	cgcgacagc	aatagcgcga	ccacctttct	cgaaccgacc	acgacgagga	13920
ccttcaataa	gagcacccca	taaacctgct	atgccaccaa	ccatgtgaac	aacaccggag	13980
ccagcaaagt	caatggctcc	ggtgctaaac	aaacgatcat	ccgtgaacg	aaagggactg	14040
gcccattccat	ccggggacca	aaaccagtga	gagacaaccg	ggtaaacaaa	tccggttaag	14100
aaagaagagt	atatcaagta	agccacgaac	tgagtcctct	ctgcgatcga	accacttgtg	14160
attccagcgg	ccgcgattgc	gaacgccc	tggtagagga	agaaagagta	atcagctgtg	14220
ggagtcggaa	agtctctaag	agcaaagttg	tgtcttccaa	tgaacccttc	ggaggatcct	14280
caaagggcaa	aggcgtaacc	aaagagatag	tagaagagtc	ctccggctgc	agcgtcaagg	14340
acattggtaa	gcatgatgtt	catcgtatct	ttggctctaa	cagaaccagc	acaaagcata	14400

-continued

gcgaagccga gctgcatggc gaagacaagg taggcagaga agaggaggta ggtgtgtct	14460
atggcggaagg ctgcatcggg gaacttgttg ttaacgggta ctaattggcc gcaaatgtag	14520
tcggccgccc cgtggcggtt ggggcccaagt aggggtggcg gatcggccgc agagcatgtt	14580
attgtccttg acatgtttga gagagctgag agagagaaag agagatacgg agaagaaagg	14640
ttgtctcttg tgtgtctctt aaaaggaaga aggaggagtt taggggtgac tttagattcc	14700
ggctagacca attaagtctg tatttaagag agtttatcaa ttatttaaga agttatgtga	14760
cgtaacaaca ttattgaatt agagatgtaa tctgccaaat cttcttaata acttgcaa	14820
tacgatccta atcctatacg gtttttgggtg attaagtgat tgccatgtaa tgtgtgaata	14880
aaaaaaagg aaattaaagg cttgtataaa aaaattacaa aaataaattt gtgaataaaa	14940
ttgatactag tatttaaagt aataaatcaa agaataattg tcctagctca accgtatatg	15000
ttcaagcaaa ccgatcagaa cttactacat agatgtcgat tttagatcag cagttttatg	15060
aaattaagtc aagattcagc actctttata cgacaatctt atgataatct tgcaaaattt	15120
taataaagg ttcataccaaa tattaattaa gtttgctgaa ggttgaaacc aaaaatgata	15180
acatgggtta gcgggaacca cgttttattt gctgaaaatt gcactttaca cccttaactc	15240
ttagatatgt ccaaaattac tcagatcatc aacagaagca ttaaaaaaa ttatcttgg	15300
ttgtataat aatactacat gcagtataa agtattcaat taggaaaaca aaaaatgaa	15360
gacacacgac atatagtctt aagcaagaga tatatacaat aattcgatcc aattgacatt	15420
gaaacctatt aacagttcat gaataattca cgtttatacc aaaacaaaa cattcatgtt	15480
taagatatat atatatatat atatatatat atattttaag atatttaag attgaaaaa	15540
gacatttatc tctcagtttg tgccaagcca tgaaaattca agtcaagtat tacttttagg	15600
aacatgatga tctcttttga tcaaaaatct attatgatgt atttgaataa atatactta	15660
aagttgttac aagtgctgat gtttcaccaa taggcttttt attgtcacta actagtctta	15720
aaaactgaaa gctttctata tatgtatgtg tgtgtgtgtc tgaccttaat catgttttaa	15780
gtcaacctaa cttaatcatt atgggagaca cttgtgactt gttatctata tatttactga	15840
tgagttgtaa atgagagttg acactcagaa aaatagacaa agcaaactgt tttctaaaat	15900
aatgaacggt aaattatgca catatttttag tagacggtag ctttctaaat cgtacaaaat	15960
aatgtcaaga attgacgtga catcattgat taccacataa caaatattga ttcttttagta	16020
atacgtagta actattgatt aattgtagta catatttttt aatttaggtc aactgaaaat	16080
tataaatgta aaattgttta ctaaaacga aaagccatag taagtgtatt acattatatt	16140
ggtttatgaa ataaaactaa agcaggcttg ttatactgtg taactaggta agcaccacgg	16200
ctcgacgcaa tagcacgtta gggagtctat agctcgctc tatgcatatc aaaattcaca	16260
atttttgttc gtgaaattct ttttttatcc tctgatgtta gaaaccaaag gcatcataag	16320
aaaaattgag ataatacaaga cttttgcatt aacttttgtg ccaaaaaaca attacgcaa	16380
gagtggtaga ctttgcattg cttcatgatg actttaaga accctgctgg attctaagta	16440
gtaccattca aatctattgt ttccagctag ttctatatat tttaaagaat gagttattga	16500
ctctattagt ctctaattat tatacttctt agattttttt tttttttttg gttcaaagtt	16560
caacgagtag acattcacag caagattatt ttaaatgccg aaagtatttc tattaatatt	16620
cttgcaacta ccactttaaa tgtaaatcaa attattgtac agtataattg aaagtttcgc	16680

-continued

atttggtgtc	tttgtttagc	tgacatatga	aagatacatt	taacaagaaa	aaaatgcgtg	16740
atgcaaatac	agtttgacta	atatgcagaa	aatatgatgt	gcaaggaaaa	gctattggac	16800
ctgatcgcat	ttaatatcgt	gataacatga	tactacagtt	atccaaactg	aatctaacga	16860
atgaataaaa	ggatttgtat	tgggttttagg	aacatagaat	gttattgaat	tcttcatcta	16920
ggcaaaatta	agttctataa	attaatcgcc	taccaaata	gggccactca	ttttttgttt	16980
aggaagtctt	tgcaaagatt	ttaaggcaac	gaatggtttag	gcaataaag	cccatttagg	17040
ctcgaccctg	aaacgtttac	aagtggccac	ttagatttac	tgacatcacc	tatatcaagg	17100
ttattattat	tcatcagcct	aagtattttg	gttggttaaag	tctaaaagga	aatatccacg	17160
tatacctaaa	cgcagctagc	taactagttag	agaaaaaatc	gttcaagtta	catataagat	17220
tactttgaat	aactttatct	ttgagatata	atttgtttag	ggaatttatg	cttcgattct	17280
tgtttcgaaa	cgatgtacca	tactatcatc	ataactaggt	ccagaattct	agacggcgcc	17340
cttgtaaact	acaagaggaa	tgatcaagca	agtaaagagg	taatgctgaa	tccaagacta	17400
aattagaaag	taagagattt	cattagcttt	aatgagttca	aagggaattt	cacctagtgg	17460
aagtcggatt	cataactatt	tgttccttaa	aagcggattc	atgtttaaca	tattttattg	17520
ttaaagtaag	aataattaat	atgggtcaaa	ccgtgttctt	gtttttcttt	tccgtagcac	17580
tatgccacta	tctttattat	acaaaagctt	cggaaattgg	atgggacacg	tttagccaaa	17640
tccaaatgat	agatatcgat	tccttgttac	ggaaaaaaa	aaaaaaaaag	atatcgattc	17700
ctttgtttta	gttttaattt	tttttgcgt	ctacttcac	actattacaa	gatattaata	17760
ttttcctatc	cggtaaaact	tactttatta	acatacaaat	ttcgattttt	tccgataatt	17820
gaattccaaa	cataaaataa	aataaacttt	ttggtcaaac	acttaatgat	atgagcttcc	17880
aaaaatatta	atacacattt	aagatctaaa	taatttttga	aagaaagcta	aataaatatg	17940
aacaaacaag	ggaaattaac	tagaaatttt	ttttgtttta	aaaaagcgga	tcactgggtt	18000
ttatttttat	tttcttttaa	aatttatgtt	taattttact	ttctaataat	gaactagtgt	18060
atgatgatga	tccggaagac	tatccaacat	gcagttaagt	gttaaaaaaa	attttaatta	18120
agtttacgag	aggctttaat	aaaagagcta	agtacatact	ggaactatta	aaagcaacca	18180
aattgaacag	aacactttct	attaatctac	aagaaactca	ttaggataat	ccaagttgga	18240
ataatctact	gtaatatctt	aatctcatca	agaagatatt	ttcactacaa	tatgcaagtg	18300
acaaaaaaaa	tatttgatct	attggagtga	gttgtgttgt	actaccattg	acataagcat	18360
tcttccacat	actttcttaa	attctactac	atcactcttt	tatgataaaa	ccaaatgagc	18420
ttcttacatc	acacacaatt	cataccaaaa	ttatatatgt	taaaacccaa	aatctcaa	18480
aaatgaaacc	acaaactaaa	ctcaaactcg	acctaagttg	tcggtcaagt	ttttctcttt	18540
cattatccgg	cagataagat	attttaaagg	tcaagttaac	agaagggaat	gaccatgttc	18600
tggaaatagt	gcttccttgg	ggcacaagca	gcgccaccaa	aacttccttg	tccttcaaaa	18660
tccagattcc	ataaacgcgc	ccaaatagcg	ttatcttcat	tcacattagc	ctcaggtatt	18720
gttaccggaa	ccaccgggta	ataccggtta	ctatttgggt	cacaaccaga	agttgttgtt	18780
gaatgtgtga	tctgatgagg	tacatataat	gctccacttt	cgccactact	gctactgctg	18840
aagtggttat	caccattgtt	gttgtogtca	tctagtaacg	acatgatctt	tttcatgtcg	18900
atttgattga	attgaagcaa	ctgttgttct	tgctgcaact	ccatttgtct	ttgctgctga	18960

-continued

aattgttgcc	tcttcaagat	tcggttcttt	gtcttctccg	cactgttagt	tggagacttc	19020
gtcttcttct	tgaaatgggt	tcgccaatag	ttcttgatth	cgttgtctgt	tcttcccggt	19080
aaactacgtg	caatcgtaga	ccacctacaa	atacaaatca	tcagacaaaa	aaccttagaa	19140
taaggaaaaa	gagatgagtt	acttattagg	gtaaagctac	aattgtttca	tatttttagtt	19200
tcgaatgtaa	tgaaaaaaag	ttaagagaag	aaaattctat	ttagcatact	agaggatgtg	19260
aataaacaat	caaagcttag	gttaattacc	tattgcccc	cttagcatgt	aactcaagga	19320
taatggtttc	ttcatgagga	gtgatttgct	ctctcttgag	gtctggctct	aggtagttaa	19380
cccatcttaa	cctgcagctt	ttccatttct	tcttcaaccc	cgcgagcctc	gcaacagagt	19440
tccatcgccc	ttcacctgta	agctgcacat	aatcgatcaa	aaggcggtct	tcttcagcag	19500
tccaaggccc	ttttctccaa	ccttcttcta	ccattcccca	tcctcctccc	atccctcccc	19560
acaaactcat	tatttcttct	tagttttttt	ttcttttagtg	aaagctaaat	aagagaagta	19620
gtgtcttctc	cttctctaac	agtcccacca	atctcttttt	gtctctctgt	cttggtgaaac	19680
ctaagtgttc	ctctgttttg	tctttgtgtt	gtgttttggt	ctgtttttcc	tcttccccct	19740
tcattattta	tatatatact	cactctagct	ccttggctcc	accacttttt	ttgtttgttt	19800
gttttctata	gagatataga	ctagtgcaca	gacaatcttt	ttacattttt	agagttgtgt	19860
acctttttaga	atcataaata	tcacttaacc	actaactaaa	agaaaattcc	gtgcaatata	19920
catacctttt	tccccttaaa	ctacaaaaat	agttctacaa	gctcttaatt	agcaataatt	19980
agaatttaat	gtaaacacaa	aagaattacg	agaacttggg	tgctctcggt	accctttttc	20040
actggctcgtt	ttcataagtc	tttcgtatgt	atcatttcct	ctgtccattc	aactcttgat	20100
ctccaaaggc	atcatcatac	acaaattact	aaattacccc	tcttgttggg	atatttcttt	20160
tgagttttta	tcatggatga	aaacatgcat	tgatacatta	aaccaaaact	tttcttcttt	20220
tctccaacta	cattttgagt	ccttagttgc	taaccaaatc	tgaagacaa	aagtttacia	20280
gtttaatttt	tttaaaaaa	attctaagaa	tctaagctta	atagtctctg	tatgggaatt	20340
acaagtatat	aaattaataa	aacccgagaa	attagaaaag	ttaacacatt	gtaaactagt	20400
gtcatggtag	gttgaaattt	tttgatttaa	ttccaaggag	acgtaacaat	ggtgagttgt	20460
ggctacttca	tcacaatcag	tttcaaaaac	ctatcaatat	atatccttat	agaaaaatat	20520
tcttcataac	ctataacaaa	attttcatct	tcaaatgaaa	cagtgaaggg	tccaatgaaa	20580
attttccctt	gaagcactat	ctgatcgtct	ccctataggc	tataagcaaa	tacttcttac	20640
gcatatgggt	gaatgaatca	gaagaatttt	acgagctatc	gatccacat	caaaatttct	20700
gaaaagggtct	tcccaagag	tatctttcct	acataacctc	gacagaccaa	gataaccgtt	20760
cacttactac	acaagtttta	aactcattag	ttttagggtt	tttgttttta	atatgtttaa	20820
ccaaaaaaa	gaaaaataaa	ttaaatacaa	cagaaattat	atttgtatat	atatatatat	20880
atatatatat	atataaacga	attcagccta	aatcaaaaac	taaacagatc	caaagcataa	20940
tttatttgat	ttcagtaga	attctattga	aaagttaag	aaactaaatc	aaattttaa	21000
gttataatta	ttaagccacc	caagcttcat	cattctcata	tacacgaaac	gaggatatta	21060
cagggtattt	cattttgatc	cattatttta	tgcaagatta	aagaaataat	gttaaggcat	21120
taggtaacat	ggtttaacaa	ggcatcaaat	ccaatagaat	aatcctttta	atctaaattt	21180
ttatcttttt	gataggggta	agctaaagtt	atttacaact	tgtgttttgt	tgccacgtat	21240

-continued

atgatctaga ggtttaatgc agggacatat attatgatta ttataattat ttaacttcct	21300
taatgaacat tttcttagaa agcataatta aaagtaaaag aatatcatgg tgctggatta	21360
ttttaaggtt atacacatta ttgaaccact aaggtttctc caagtgtcaa ggatatttat	21420
acatgcgac ttacactaat ctttgttgtt gtcgtataag attcgttgga ttccacataa	21480
ggagttatta caataaacta tagaataaaa ccaataata tgaataagtt gatgggtgtt	21540
gtacgtcacc atgggacagg tgggtcttata attctctcct cttagatatt tgtaatgtct	21600
cttctcattt ctttctagat attgtgatca atccaaaaa atgaatgtct agcgtttcca	21660
tctttttcca tagtagaaat caaataaaaa aatactgcaa aatactaata taacaagaaa	21720
tatctacgtg aaagatgagt tattatatgt cattaatagc gttttatgaa aatcacaaac	21780
tatgtattat gtctctggtt tcatgttggg tacgtatttc agatgaacct aaaacaaaa	21840
ctatcggttg gtttttgttt ctaaaataca tataacattt tcagtaacat ccaataaata	21900
ggatactgaa tgtaagttt acgtgtatag atattgcaac ttatttatca aaatatcaat	21960
tgagcggttg aaatttcttt tgtaagttcc atgacaaaa taacgtcaac tccaaatttt	22020
tgatgcgaca aaacatagta gttagttaat tagcactaca aacattagac taattttcta	22080
gaactagggt cggatttagg gtatttcacg ccaagcttga ccgatcgact gaccgatcca	22140
cttgctagtc aaaggaata gattttcttt ttaatataga ttacctatga atgaatcttt	22200
taaatgccaa aataaggtac atcatcaaat cataattact taaataatgc gtaaaactgat	22260
cgtatgggtt ttgaatctct caagctgtta cggtggaact attggggcac ttcgaagatt	22320
aggacaattt ctaagaagtg aaacgaaata cgttaagacc taatctctag aatatagttc	22380
ccgtgtcttt ccaaagacat ttgatgccta taaaatataa atcacatgac tgatcggaag	22440
agtcgaaaa cacaataga actatggtga aaattcaaat ggttcaacag atttttat	22500
tctaagtatt ttaatcagat ttgtttcaca aagcaattct caaatttggt tacctcaacg	22560
gctcagagac tacagacaca aaggcgtgac tataagacgt cagcaatatc attcactgat	22620
tatacataaa aaataatttt taattgaatg attagacact aatcactttt ttaattccac	22680
atgattaatt atttaattga atcatctaata aaagttttgt cgttgagggt catttgaact	22740
ccgtgtttcg tacgggtttt catattattt ctgaccttg tactgcaaag acatgtttag	22800
cagctaagca tgtgattttg ctagatatgc ttttacgtta agaattattt atttttttt	22860
aaagtatgtc ttttcaggag aattattatt tttttttttg ttttttttca catcatatcc	22920
ctctcaaatt atttgtagtt tatgtgtgat ttgttttggt atatataatg tgtgaatttg	22980
tttttttaga aacactagag actccgtatt gaaatgttta ttttaattat ttgatagaac	23040
tcttttttat aaacacatgt taaaaagacg gtcttgctct atatcagaaa ataaattatc	23100
aaaacttact cctaaactat atccgctaga atttgttgtt ggagaaatat cgaaaatata	23160
aattcacatt tgggataaac aaaacataaa tttttaaaat gtcataattt aaaatttagt	23220
ttaatgtttt tgataaaacg gtcaaaattc gttgatttgg ctacttgagg taaatcatta	23280
aatttttgtt aacaagttta atttttatta aaaaataaga aagaatattg aaccaaatac	23340
gaattgatga tcttttagac taatacaaca aactttaaac cattgaagca cgatagcatt	23400
ttcgtatgct agatataata ctaggtagta agagatgtga aagaaaaaat gattgcaaag	23460
aacggataac tttgactaat ttcttactag atacgacaat tttatcaaat atttcgaatc	23520

-continued

gtacaagtaa	taaaataaac	aaaaattcaa	atccaacggt	tctgcgtgag	atgcaacttg	23580
gcatgtactt	tgttgatcaa	atcatcgtaa	tcttttgtgt	aacttatata	acattttcgt	23640
catttttttt	ttaagtttcc	cccctttttg	gtgactttct	ttgtcgatta	aaaataaagt	23700
ttgcacgata	ctttggaatt	agctttttaa	ttaattataa	taatgttggt	cctatcttaa	23760
gttgaaagtg	cagggatgtg	tacttgccaa	taatcttcta	gaaccagagt	gaatcattgt	23820
tattctcaca	gctttgatta	caaacatttc	cccccaaaaa	taaaataata	caaaaaccat	23880
atatcaattg	caaactctatg	ggttcatata	tggttttggg	aaattctttt	atctttgggt	23940
gtgaccata	tggagtatga	ccttggctgt	gaccttattt	gtggattcta	aatctaggac	24000
ttggttatag	tcgatagact	caatatggat	tgtaatcca	taaataggat	ttatgcaaaa	24060
aatgtatatg	gggaattgta	ttagtaaaaa	aaaaatcatt	ttgcggcata	tcttaccaaa	24120
gttcgaattc	cggttattgc	gttatcatgc	ataatagaga	ttgaattatg	cgacttgtcc	24180
attccaataa	tgtcaaaaac	aattgacaat	tttttataac	actactgaaa	ataggcacat	24240
ttaaactttt	attatataaa	attagaaata	ggcttactta	gagggttttc	cataaacttt	24300
tccaaaaact	atttgtttct	gcacacactc	tgaagaagac	tcactctctc	ccatctaaag	24360
atttcagact	catcacagt	ctccttactc	aagctattat	ctagaacatt	tagaaagagc	24420
agaattcaag	aaccatata	gtttcacggt	cgccggcttc	catagcaagt	gacatcataa	24480
caagatgttt	cttggccttt	caaacttcta	tatttattac	atcaggccat	tctccacatc	24540
cttttggttt	tagcattatg	atccaaccac	accatgacga	tttcatgtgc	aagtttcctg	24600
aaatcttctt	ataggaagag	tttttttgcc	attataattc	acatgttttt	cttghtaattt	24660
tttatctctc	tcttcttttt	attctaagtt	tcatagttaa	atattttgta	atgtcttact	24720
tattcaatag	aagtagaata	atattttattc	acatacaact	ttcctaaact	tgatagaatc	24780
ctatctaggt	aaacaaaatt	ggtcaagaac	ttaaaccatc	tatttctttt	tattatgcgt	24840
cggataagac	gtttgatttt	ttttgaagta	tatattcctc	acatgtcttt	ttctgaaact	24900
tttttcgtca	ctgctcttct	tttttatata	actacaatgt	gttattgtgt	cgaagaataa	24960
attcaaaaca	gacaacatta	tattatttct	acgtttttga	ttaaaacata	atcactttcc	25020
ctaaatcatc	catcgaaaga	atgatttcta	tcaagaaatt	acaagatggg	agagttaatc	25080
tcactttcag	aataggatatt	tagtataaat	ccatcataga	tatagataac	cttatctctt	25140
atacaataaa	taactctttt	gaccaaatta	attcatatgc	tgcaatcttc	aactctttat	25200
gtttaatttc	gtatatacac	ctcaaacctc	agagtatata	aattggaata	tcagaaagga	25260
tgcaaaatta	tcgtatttta	ttatatatca	gaaaatatat	acttattgta	tactttttta	25320
aaagaaatcg	catcacaaat	tgaatcaata	tttgtgttct	cctaattttt	ggaagaaaag	25380
aaaccaactt	ttaaccagggt	cccggttggt	cagtggattt	gagtagtatt	ggctaccgag	25440
gtttgagtaa	tttaacatag	ttttccgaac	ctcaggatta	attttttgag	tctaggaagc	25500
ctttaaaatc	acttgagtta	tcgattaaaa	aagaaagatg	agttattcaa	gataggatta	25560
cgttatatat	gtagataaaa	acataatata	ttatttacia	aacaacataa	acaataatgt	25620
tatocaaaga	accaactttt	tcatttaaaa	attaggtaac	atagtaaaat	tagacattcc	25680
atattattat	tatatagtaa	aggtacagtc	attagactat	tcttaggata	aatttcactt	25740
tacatcaccg	gcctagcgtt	atgtagatgt	cgggtgttct	aatacatgat	ctgagtcttt	25800

-continued

ttagtccaaa gcaatcgtca tgataaagtt gttcggacat gaaggaacta gaaaatttcta	25860
ctaacttttg acaaaattca gttttaaata gttgataata taaccatag ttataatcag	25920
ttaatcttct atttcctggt ctatgtcgcc tagaatgata tacgaattat aaatcttatt	25980
tgatagaaaa gaaaatgtta taagagacaa ggcacgcaag aatgattgaa gatggtagat	26040
aatttaggaa tccagaatac atgataaatt attaccttac ggcattcgaa aaagtgttca	26100
aattatcaat tcttattctt ttttttctta taacttgtga ttttgtgtgt tttagtggat	26160
tttctaatac ttaggtaaaa agtgtgttca tttagttaa caaaattgta aatatacgaat	26220
ttgatacaaa atgaaattaa tgagttaaaa aataaagggt tgacaaaaaa aaaagagttt	26280
aaaaataaat aaaaaaactt agagaataca aatgagagtt ttaattcttg tcatgttttg	26340
ttttgatgca agaaaaagta tccgaggatg ctaaatagct tagaatgtca tgaatacgtg	26400
atgaaatcat taccacagct tgtatcatat gaaagtattt agggaaaatt gtcaattcag	26460
aatttcttac aatgagtcaa tagtaggaaa gttttgtaac aattttataa aatttgggtt	26520
aacatttgta aggtaaaaaa aactacattg aaggaccact acatattcga ctatatttta	26580
ttttgtttta ataatacagag gtggtgttgg aaacaacgta aggttggtt aagcaaatgt	26640
gaagcatcca aatcgtaaat taattaatga cagcaaatca agtttggttg ttaaacacta	26700
aagaccaatc cacctctagc tatctatttt actttgacga aaccatttta tgagttgtga	26760
aataattact ctttgattac tttataaaac cttaacctaa caagtgaac tcttttaatt	26820
acttttgctt agctcaagtc tactggttga aagaaaagat acattgtctt ttctctattc	26880
caagttcgag tgtagtgaat ggattgaaaa ctctagaaga tatatgagtc atctaaaagt	26940
taagaaagtg agagatattg agtagaagt caaacctta ggttaaggtc gaggggtgtg	27000
gggtattttc atcaacaatt tgcgcacgaa tgccaaaaca aagcttatca attttttttg	27060
gcatagtata tctaattagt gttgttaggc aaatacttac atcattcttt taaaatattc	27120
tttcttcttt ttgttataaa gaaataaata actatttttt ttcttccttt ccacaaaatt	27180
tattttttat gacctctca tttttaaaga taacttgata gattaggaac tcacatttaa	27240
actagtttca aaatctgtaa aagatctatc ttactattt atattacaaa ggcttatggt	27300
aatttccaca ctttataaat gggattttgt gcaaaagaaa cccttcttag tcgcctattt	27360
ttgcaacttt accctctatc tccgaaaaat gcataataac tctatgaagt taaaaataat	27420
aaacaatata ttggctattg tcaagttttt aaaatttaat aagtaatata tcaaatcgag	27480
actctcgaga cgaggagttc aactctttat aattcctttt taaacccttt agaatatcat	27540
atatttcttc tagagtgtta ctttctttct acaaaattat tatctaactt ttactctgta	27600
atattacatt tgaaaagttg cacattcatg tatgcatcat ttcctaattt tagaaaagat	27660
atatttccta gttaatgata tacggatcaa ttttccaaac aattatttgt gatattctaa	27720
agaaacttta atttcagttg tccatgaaac atgttgtaat atagagagca ttcctaggta	27780
acaacatcaa cttgtaatat atatactct gaacaatgtt acaagcatct aatgagcttc	27840
atagttcatc gtgcatgctg tgttgtagag gaatattaac agttggaact gtattgccat	27900
caaaactgtt tccgagcggc ataacatcat ctctacacc atttgctgct tttccagcta	27960
attgagcttt aatagcagca gtagccattc gttgcagttt cagaattccg atgogttttg	28020
aaacaacgta aacaacagca aagacaaaaga gagaaaatcc aactatgagg attatcctgt	28080

-continued

tgatagtgcc	acagaaaaca	agagttagtt	agtattatta	ttcacctata	gaatggtttg	28140
gtaaaacaga	gtctattctc	ttttcaaata	cctgtcaatt	acatcttgac	gctgcattgt	28200
tgaaagtaga	ttacgggtcc	tcgacaacag	agatctatga	cctttgtatt	cactctcggc	28260
ttttttaagt	actccagtag	attcatctac	aatgtttaac	atgtataata	ataataatat	28320
gtcagtttta	tgtcttaatc	tgttttctact	aaggacttaa	ttatctggca	tattatgaat	28380
gacataccaa	aagccacgag	agtgtttgta	cttcgttcaa	cctcctgtaa	cccaagtcaa	28440
gagagtattt	tgatcacaca	ttgtttataa	attaagaagc	cgaatatgca	aacaagtgtg	28500
aatgcacaat	catttcataa	ctagtttttt	tctgacctga	accatcagct	gccgtgaacg	28560
cctgagactt	tcagttatgc	tttcagcatc	ggatgtcacg	ccagcatttg	ctctataaac	28620
aaccacaaat	gtagatttca	ctggaaacta	acactgaact	cttttttttag	tgcatgatgc	28680
atttaccaat	aaacaaaagc	ataatacaac	ttaaactatgt	ccaagaggag	tcccaattga	28740
attgctggat	tttgtaggat	aaatatagca	gagtttccaa	gatgtacata	ctgtcgttta	28800
cgtctgagaa	ctgtagactc	tgttccacca	ccgaggagaa	gctctctcta	caatcattga	28860
aagcactcct	tacacaaaaa	tttcaaaagg	agctggctgg	taaaaacagt	ggcagagaga	28920
gagaatgctt	aagccatgag	actttacctc	ttcttgagct	gctttcctca	tgttatcctt	28980
agcttgcaaa	ttagcacttc	tcaagtttac	cctcaactg	accacaaaca	cagtgcacaa	29040
actcatatta	gtttcaccca	tttatcttct	cctgaactca	tactctcact	ctcttggtgt	29100
accagaattc	taagtaactg	agatcaaaact	caatttcaag	gaaatacttt	aaaacactca	29160
tatatcattg	aaaaggatat	ataacattcc	tctcgttctt	acaaactact	acaacgatga	29220
ctcaagtacc	cttatgcaa	aagggttaac	gcacactatc	acatcatttg	tcctagagat	29280
tgaattactc	aactgcctaa	agacaacctt	aattgtacaa	atgcaagggt	cctatatctt	29340
acttgtagaa	tgttgtagat	tccacaagca	gacagtaaac	caaatttcca	agtacctgtg	29400
atattgattc	ttccacgttt	caaggagtga	ctgagtggat	tgaacctgat	catcagaagg	29460
aagctgaggt	gcgagcagat	ccaaattgaa	ttgaagggaa	ttgagaagag	agagaccatc	29520
ctgagccaat	ccgttgagcc	tctgtaacga	aaacttctct	tctcctccat	ctcccctcct	29580
tgatttaccg	tactgttgaa	tcgagataat	atgtccaatc	gtcttctcgt	acgcttcttc	29640
ccattctctc	ttcgtcttct	ccacttcaac	aacaacctca	tccatgtcgt	aactgcaaat	29700
ttcataaatt	ataaaactgc	agagtttagt	aatcaaaaga	gatataaaca	agctacggaa	29760
tctcccaatg	caatctgcta	atagctctct	aaacatccaa	cacgtataac	catctaattg	29820
ttctgtatg	aattatgatc	gaatcactac	aggaacggta	tcagaagact	cgattgattt	29880
gatgggatat	aagggtcaca	aacttacatt	gggtgatcga	aagagatgaa	gagcggagaa	29940
aatgagaatg	aaattgcgaa	actaggggat	aagggtgaag	aagatctggg	aaccagagag	30000
tgttgagatg	caatattgat	ctgtagaaga	cacctcaaat	tttgatttct	ggggaagaag	30060
agaaataact	aaaaacgacg	tagtttttaga	taaacctatg	ctattctaaa	cccggttatc	30120
ttgcaattgc	tatcattgtc	tgggtccaatg	tccggtttat	cagttattta	gttatatggt	30180
catattctta	ttggttaaaa	tttctagcat	tacttattga	tgaataaaga	agtttactgc	30240
tgagttttgt	tcaaagtaca	cgtaaatcgg	atttatgaaa	cgacaattac	acagactgac	30300
ccaaaattgg	taattgcggc	atccgactta	cccgaactgc	atgccgtttt	gcaaaatctc	30360

-continued

aaaatgaaaa	acggcatggt	ttcttagtaa	tctactctag	ttctagcaaa	ctgacttggt	30420
caattgcaca	tagagatgat	caactacaaa	cacagctaata	gtttcaaact	ataagatgct	30480
gactacgtct	gatgtactca	agcttgagga	cgttccgaaa	ttacggtgct	ttccatcggc	30540
taggttagcc	agagagtagc	cttgttcatg	gagttgatta	atcaatagag	tctctaactg	30600
acaagccatg	ctcttacctt	gaaccataag	gtatagaaaa	cttgaccctt	gcagtccttc	30660
ctttgctcga	tgcgcacgta	ttcgtccttc	aagatcatcg	gtctgcagtt	tcgtaacaga	30720
aatcctcaga	atcaataact	ttcgttttct	gtctattcta	agagaagttt	tataatatta	30780
atatggttaag	tatgtacctg	tccaatgtac	aatctcttat	cgggtctccg	catcacatac	30840
acgcatgaag	aaccaactgt	agatggaggt	ggaagctcac	gagcaccaat	tgaagacat	30900
tctattgctt	caggctcaat	catctttttc	ccacagattt	tgacgatagc	tttgctaaag	30960
tccttctcca	aacttctctc	agagctgact	ggtttttgga	tctgctggtc	attggttgat	31020
gaagttatga	tttggtcggg	tttgacaact	tcagctgatg	cgtcttttgc	atagaccgag	31080
aggtaaagag	cttcagctct	ttggataact	gactcgggaa	caccttcctt	cttagctgtt	31140
tcaaacgcaa	gactctctct	gcagactcca	tctgtcaatt	tccaagtggg	cttggtttgc	31200
ccttcgacat	tttcggctcc	cattgcttta	tatgtgatgt	ttttcgtgtg	aagaggtaaa	31260
ctgaagattc	catggagatg	agtagatata	atacccaaac	aaccacttgt	gtcaagactc	31320
tctaccacac	taccagcgat	acagggtgct	tttgctgtct	ctgtccctcg	gcatactctc	31380
tctataagca	ctaggcttct	cgaagtagcc	tggcttacia	tagatcgtat	ttccgacatt	31440
tctacctgat	ggcacagaac	cagttgagag	gatatcaaac	attacaatta	aagaaaaaaa	31500
aacagtttgg	tgaattctag	atttgatgct	ttatatatat	acctggaaag	aacttttttc	31560
gtctacaggg	ctgtcatatg	atttcattgt	aagcatgatg	gaatcaaagt	gaggaatata	31620
agctgattca	gttggaacca	ttaaaccgga	aattccaagt	agagcagctg	cgcataattg	31680
tctgagcaaa	ctcgattttc	caccaccgtt	agggtccagt	agaagaaaca	gtgattgcat	31740
gtcaacggta	ttgtgaacag	cgggttcaga	agatacatca	aaccaataag	gtgacaggcc	31800
tgctagcttc	attcgaactg	caccatctaa	tggttttgct	ccctatatta	aaatataatc	31860
aaataggtgc	atgagatata	taagaataag	acagcccagc	atgttttggtg	atttcttaga	31920
aattgcacgc	atctgtctac	aaactcaacg	actagcaagc	acagttacct	catctaaact	31980
gaatccgaca	agcgttgtaa	aaacccactt	tcgccttctc	ccttcactgt	agttgtgcac	32040
agtaactaga	ttaaaactac	agaaatttct	taatgagaat	tgattaagcg	caaacttacc	32100
aagcatggga	aaataatgct	tttgaaatga	ccagaagcat	agatgcaaag	acaagaacat	32160
ttatttttgt	ttgcaattta	acggataact	cgcgcaacag	ttccaacacc	cgagcttttg	32220
cattctcact	agcttcatgg	tatctgcaac	aaaagcataa	catgatatac	caatacatac	32280
ccagtattat	tgaataaatt	tcttttggtg	ggttcaataa	aacgaggaaa	caacctgact	32340
aaagcaattt	ccaccttttg	ggtcgttaac	cattcttctc	caaccttttt	tcctttcgag	32400
tctaaggcag	gtttcagctg	ttttatttgg	tcttcccctg	cagtagcagc	ccagatagat	32460
ggcgtaaaac	gtttcccctt	gaaccaaaca	gactcatgct	ctcttgcata	tgcgatttctg	32520
cctttcgggc	caccaagtga	agcagtggtg	gccttaattc	ttgatataat	agggtgaaaa	32580
tcctcagcta	cctacacatg	gatgttttga	ctcttataaa	tacgacaaga	tcaccagcac	32640

-continued

cattatattt	aagcaaaaaa	gtactatatg	agcatctgtt	atcgttcaaa	gaaataagct	32700
gcaattgttg	aatacttact	gctaagata	aagcctcagc	tgatttttct	acttgagtga	32760
tttcttcctc	tatatgaatt	cccttaacgc	gacctcgcca	tgaagactcc	atatcataaa	32820
agaattcggt	cgggacattg	tcacatttac	ttacattctg	atgactttca	ttctcatcta	32880
aagagatcat	ttaccaatt	gtatcagacg	cccaatgaca	ttcgttgacc	tggaagaaa	32940
ccgaaaacta	cgattttatac	tcatactcat	aagctagtgt	gattataaaa	aatggcattc	33000
aaattcaacg	cttatttttct	gaactcgtga	agggatagga	acttatcatc	aaggaaaata	33060
gataagaaaa	caaagcaggg	agacatacac	tcagagtcac	gtttaaagtc	ataaacttac	33120
aaaaagtgtc	aagtcaattt	tcaaaccagt	agccaccag	gtaggatcca	tcaataattt	33180
caggatttcc	acaagctcag	catgtctatg	catatgtaat	acatcatcaa	gcacattttt	33240
tattcgacag	aactcaatgt	agttggcttc	ccgttgctca	agaagcttca	caagctgcaa	33300
aaagatgcaa	ataatcacac	tgtaagagat	gagagttcca	taagtgaac	gacaaaagtt	33360
agcatctttc	agaagaattc	tatgtaggaa	gtgacttaaa	ccttagcaga	agagacgcag	33420
gtaaactctg	gaattgaaca	tgttactgtg	ctcatgagct	tgacgcttcc	tgcaataaac	33480
aatagtttga	cagggacaat	gagaagcctt	aatcataaga	aaccctaaaa	ggaataaact	33540
aaggaaactc	tcttcaaate	accttgaatt	ttcagagcaa	tatcgtaagc	aggaggggtc	33600
agaagaagat	ccctaacata	cctgaaagga	taaaggcatt	ggaagaagta	gaggagaaaa	33660
acaataagaa	cgtaataaag	atccttcacc	caaacattac	ttacaaagaa	ggcagaccac	33720
tgacgtaga	tggaagtaac	accttcaaca	aacaaggtat	tccttcagta	ggaagggcac	33780
ctgcagaacc	acataatata	tattatcatg	ctttgtcatt	cataaactcc	taaacctcct	33840
agtgaagcac	tattcagcat	gaagaactat	gccaaccaat	ttgtgtagcc	gttccaagat	33900
gcaacggacg	tgccgatttt	tttgaaggta	cattgacatt	tctaaaggaa	acttcatcat	33960
caagaccata	aacatctttg	acctgagggg	gaacaataga	aagtagttgt	ttttcacaaa	34020
actgaaatgt	cacaattgca	tatttagtca	gcagccaaag	tccgtcgttc	tgacccttga	34080
taagagctcg	gaaagagtat	ctccttcaaa	ccattcaaaa	ttcctgctac	tgacattctcc	34140
ccagagtaga	cccccttccc	caaacctctc	ccagcggcac	gtccctgcat	aataaaaaat	34200
taacaagaat	tgtatcagaa	aagaccacaa	agctcaaaag	tcaattaagt	tgacacaaaa	34260
catgaatcgt	atttctcttg	taactgtctc	cagactctct	aaagtatttg	ctagatcata	34320
tgctaccacc	tgagaatact	tatgttcatt	tgaggataaa	atggcaccgc	tctacattct	34380
attattggca	ctatcttcaa	cagttcagag	atgctaagac	aatgtagaat	tgaagatagt	34440
attcaaaacc	tgatgcattg	tgccctcaacg	atgcatgtaa	gaaaagatga	tgacagcgac	34500
gagtgcggag	cttggttaact	aaggcttctt	ctgttagacc	atcatctagc	gaatatgctt	34560
tcatagtctc	gaaaatagat	atcatacaat	accccttgcc	tgaacgagat	atccctgtac	34620
ttacagaaaa	aatcaagtca	tgtaataact	atatataaat	ttttaatagg	ttgtagcgca	34680
aaataatata	tactgacta	acaaaatacc	ataagaaata	cttaccaaca	acaggcatag	34740
gatcaggaaa	gtcaagatca	tggtcaacac	cgacaagccc	atatacataa	ggacttctctg	34800
gatgtgcatg	cctatataca	gtacatgtca	gacagacgaa	tgcaagaaaa	aaaaagtctc	34860
agtgatattt	cctagctaata	aagataatga	atattgcagc	ttgtttaatc	agaaaaattcc	34920

-continued

aggttctatc atattacata ccctgaaata aatcgacctt tacgggagcg tgetggtgtt	34980
ggccctgaa cttcctccac aatacactac agtgaagaaa aaaacatgca gtaattttta	35040
tttaacatgc ttgagggaca tggtaaatgg taaggaatta tttttctcaa ctcaccactg	35100
aataaccatt gcgtgtcagc tcatccaaag tctgtcgaag attctgtaga aagaaaacaa	35160
gaatttgtgg caaagaaaa attaatgtga gacaacgaaa gcctgaaggg aagagtaatg	35220
atgattaatt gcagcagaac atatgcttgc tatccaacaa aaagcaccac aacttgattc	35280
cgtgtacgca atggccacta aagtagctgt ttatgtaact tgcccaagtt ctgectcaat	35340
gattacaact ctgagagact aaatagaatc aaaaaagatt ataactttac tacaagcaaa	35400
gattaagaag ctataatggc tgtgtatatt tatttgcccc cttaagaaga tgtggctctc	35460
ctctgtttta acagtaattt tctcaacaca ataccagaga cgaattagtc tattttatca	35520
taggaacaaa ttataaactt gaatcacata aatgctcata atgtgagtaa gcttgaaatt	35580
caaaaaggat tgtgtacat aattgggcag ccagcctttg gaatactatc tgatcgaaga	35640
ccacaaaaag gattgagacc agcatattca acaagtatac aagcatctat tccaatagcc	35700
tcataaaatt ctctacctg gagaacaaa tttcagaaca gtaacagtct tcataagata	35760
aagacttcaa atgacattat tctgctaatt tttgttctga agtaacaaag atcacaaaac	35820
tcaatggaat tagaaaaaca acaaacatac tctgcagagc aaaacttcgc gtggaaacct	35880
tgacttaaac tgcaacatct cccagttgag gtttccatct tttaaactga aatgaagaaa	35940
attccaaata ttgcatacat aaaggccaca aagttaccgt aacaaactgg actaaagtag	36000
tggactagga ataaaactaa tggattgttt ccaactaaag tacttagagc cttcacaact	36060
aaagtagctt aagtcatact aagatacact agctactttt catttgataa ccaagctaaa	36120
aatgattaaa ggctgacatc tagaaggtag gatgaaatct tctaatagtc tcagaagaga	36180
actgaccttc cattcctcaa gctaggggcc aaaccaagta aattgggtga cataagcctt	36240
tcaataagct gaagagtaga tggtttctta catgtctgca atctctgagt agacacagaa	36300
tataaaaaat cagtatataa cgaaacaata tcaagacttg aaatttctta accaaaaaac	36360
aaaaaagct ttacctcctt ccaccaaacc aaatgagaga gatcttctgc agtgagaaca	36420
tcacttgacg tcttcacttt cttagaagcc gttgtgattc ttttcaaaga ctttccatct	36480
ctgagacaag atatccctc agagtacct ctattactgt caacacataa caaaatacaa	36540
aaaaagtttc ataacaagat tcttatcaag aactagacca aacacaaaaa tcaaagaaat	36600
tgagttcaat attcaaaagt acaggaattc ataaaataca aaaaatctga acttgctaac	36660
atattagaga agattatata aggagatgag aggattacag tagaattggg gagggagggt	36720
tgagggaaga gtaagtgcga tatgaggagc ggaagaagaa ccgccatttt gggaatgaaa	36780
cgacggcggt tctggtagca atccaatgca tctctctcac actctgaaga agaagaagaa	36840
gaagaagtag agttttcacc tttcaccctt tttctagggc ttcttttccc gccagaagac	36900
gacgactatg caattcacia tctcacagtc ctctaattga aaacatgttt tctcttctct	36960
tcttgccttc ttagtttaaa ggtaattagg attttagcct tttaaaaaga gtaatttctc	37020
atttgcctcc tataatgttt caaattacat tcttgttcat taatttacca aaattaatca	37080
tcgggtagtt ttgtatcttt ttggaatgac taacaaagaa cccatctcta tgaacacatc	37140
tctaattctc cagtttttgt tgtaaagatt gtaatttctc tgcttttcaa agaggatcaa	37200

-continued

aagttttgta	attagaaatc	ttaaaatgac	tatcagttgt	ttctgaattc	tatacaaaact	37260
taataactcta	atgagtatta	ataataagaa	aacaaaacca	ctaaatgtct	agactcagtg	37320
cagtggtgta	gtgtgtgtgt	ttgtttggca	caacaaatac	gaaacatcta	gaattacata	37380
agaatttggt	gtgagtcatt	ggatttatca	gataggcatg	cacaaagtgt	ttcaccaatg	37440
gctattttgc	atgttggtgc	tgcttcttta	actgaactaa	gtacaaatca	caaatatcca	37500
atcacatttc	atcttctccc	ttaacaaaga	ctccataaag	aaaagagaag	aaaagactca	37560
taaacagaga	aaaaaatggc	aggagagaga	tcaaagttaa	caacaaacca	cttttataat	37620
catcagatta	ttctttgtta	ctttcttata	atctcacaag	tttccattgc	ctcaagcaac	37680
acgagtaatg	taggagttaa	ctggggaata	atggcgagtc	accagcttcc	accagaaaag	37740
gttgtgaaga	tgctaattga	caatagtctc	actaagctga	aactatttga	agccgaccaa	37800
aacatcttag	acgctcta	tggtctcagc	attgaagtca	tgataggaat	accaaaccgg	37860
tttcttaaag	aaatggctca	agatacatct	gtagcagctt	catgggttga	agaaaacggt	37920
actgcttatt	cttacaacgg	tgagtgcaac	atcaagtaca	tagctgttgg	aaacgagcct	37980
ttccttcaga	catataatgg	aacttacggt	gagttcacat	taccagctct	tatcaacatc	38040
caacgagcac	tagaggaagc	tgatctgaaa	aatgtgaaag	tactgttcc	tttcaacgca	38100
gacatctatt	tttccccgtg	agcgaaccct	gttccatcag	ctggagactt	tagacctgag	38160
ctaagggatg	caacgattga	gataatcaat	ttcttgattt	cacatgattc	gcctttcacc	38220
gttaacatat	acccttttct	tagtctctat	ggaaatgctt	actttccttt	ggattttgcc	38280
ttctttgatg	ggactaacia	gtctttgaga	gatggaaatt	tggtctacac	caatgtgttt	38340
gatgcgaatc	tcgacacttt	gatttggtgc	atggagagat	atagcttctt	ggggatgaag	38400
atcattgtag	gagaggtcgg	gtggcctacg	gatggagaca	agaacgctaa	tgtaaagagt	38460
gcaaagagat	tcaatcaggg	aatggtaaag	catgctatgt	ctggaaatgg	aacgcctgcg	38520
aggaaaggag	tgattatgga	tgtttacctt	ttcagccttg	ttgatgagga	cgccaagagt	38580
atagcaccgg	ggacttttga	gaggcactgg	gggatccttg	agtttgatgg	gagaccgaaa	38640
tatgagcttg	acttgctcag	ttaaaggaaat	gacaagcctt	tggttcctgt	ggaagatgtg	38700
aagtatctgc	ccaaaacttg	gtgtattctt	gaccctaatt	catataacct	cgatgacttg	38760
cctgataata	tcgactatgc	ttgtagtgtg	tctgattgca	cagcactcgg	gtatggatcc	38820
tcttgtaacc	atctcactgc	tacaggaaat	gtttcatatg	cctttaatat	gtattatcag	38880
atgcacgac	agaaaacatg	ggactgcgat	ttcttggggt	tggtttgatg	cacagatgaa	38940
gatccatctg	atgaactttg	cgagttccct	gtaatgattg	atacaggaga	ttcaacgagg	39000
ttgcagcctg	gatcttcaag	agtgttgacc	agagtcgcag	ccgcggtttt	agttatgttg	39060
gtccttccca	tcttgtagta	agtacctgtt	gtttgttatg	tcttgtaaat	ctctcaagtt	39120
ttacttcctg	caaaagaatc	tttcaaaacta	aaggttggtc	ccttcgatta	gcatagcctt	39180
ggaaaataca	caacttcatt	aactagaaga	aattgaatca	cagaattaca	tattactaga	39240
gtgatctggt	ccaaagctgg	aagttaagta	atataaacac	ctagagtctc	agagtgggat	39300
taacagagag	aatgatattt	tcaatcagaa	gaagctggta	ttcatgtcag	atttcttttg	39360
gtgataaagt	atcttctcaa	atatgtgttt	cagcttctcg	tgccgaacca	tctogtcgag	39420
gattctatct	ttggaaacaa	cttcgttgtt	aatcggcctg	tccttctcat	tggtagaaga	39480

-continued

gaagaccagt	tcggatatgc	gatgtttctc	actttgctta	cagtacttgt	tccactcaga	39540
tgtgtctttg	accatgagat	ggtaaatgaa	cacagcacgc	ttttgaccaa	ttctgaaagc	39600
tcgacttatg	gcttggtctt	caacagaagg	gttccaaaca	acatcaagaa	tcacaactct	39660
tgaagcgcca	acaagactga	taccctcaga	acatgctttt	gttgatgcca	gtagaacttt	39720
agatccactg	tctggtttgt	tgaagtgtgc	gatcatatgc	tgctgtctc	tttgttcaac	39780
tttaccatgc	atcaagagaa	tttgttcccc	ttcagtccag	tcacattctg	caatgagctg	39840
ctccatgac	agcttcagag	tgtcaatata	ttggctatac	accaacactt	tctctttcac	39900
ggttccactg	atgcggatga	agtcgataag	gaactttggt	ttcacacctt	cctcgatatt	39960
aagtctaagc	ctcttgagag	ttcctaattg	tgctggccca	ataaccaagt	cttccttctt	40020
ggtcggattg	cagcacaaat	acagagatgg	gtgtactgaa	acagctgaaa	gcttggtctc	40080
aaattcaaat	gtgttttgag	aagtgtcaat	tctgtcaaga	attttcttct	gttgaaaagg	40140
cgggttcaac	acgacaacac	agtctcttag	acctggaagg	ctttcttgaa	ggatggttcc	40200
ttcatggaca	tgacacaaat	gagcaatcat	agccttaaga	tcacacattc	tgttttctct	40260
attcaccttt	ccgtgttctc	cttcttggtt	acacttgcta	agttcatgta	tcctcgaaga	40320
aatcgtatct	ttatcagctg	gtcttgccag	gcataagacg	ttagaaagtt	ccttgaaatt	40380
attctggaac	agtgctccct	aaagaaagat	acgcttttct	gttctgactt	cagtaagcac	40440
tttccaaatg	agactactct	ggttttctcg	ggtgtggccc	tcgtcaagga	ccagcaaccc	40500
cggcaattct	accaacatcc	tcctaaacac	ttgcattcct	tcagtgttct	tatttgccgc	40560
gagcttctca	tacagagggt	agctaattcc	aagaatgctt	ttctgcttcc	accacgaaac	40620
tagcttaacc	atacggtatg	aattatgatg	cctatttccc	tctaaacgtg	agacagcttc	40680
agcatcctca	taccctgata	actgcagact	attcatgtta	taaaatggaa	tgtttacgtt	40740
ccatttcctg	acttcgtctt	cccaagtacg	cattagagtt	gcaggagcta	tgaccatagg	40800
atggctatct	ggaaaccgtt	ttaagtatga	ctgaagaaaa	acgacgggtc	aacgtgtctt	40860
accagttcca	gccttatgct	aaataatgca	tccgccactt	cccttaaccc	ctacgctgtt	40920
caactcatta	atttttgttg	tcccagctaa	attcttccaa	ataaattcaa	acccttcttg	40980
ttggtgtggg	tacaaagtat	ctttgatgcc	agggacgtat	tgccagacag	ttccttctat	41040
attatccaaa	ggggcgacaa	aactgctggg	atctgaagca	tcaaaactca	gcctgttagg	41100
taaaggttca	ccctttctat	cactacattt	cttattgtca	ttaacactag	gacgatactt	41160
gtcctgcaca	atgcaggaaa	atgaacatgt	taagaactca	gataagaaaa	tcaacaaata	41220
gtacaactca	atctaggatt	tcagagcgct	ataaaaatgt	taccatggct	ggtgagatgt	41280
ctttgatctc	gacagccaca	tacgcacagt	ggacgcattt	caaaccgatt	tcacatccca	41340
ggacaaagtc	gtgtgtccct	ttgctacata	gcataatctc	attctaatac	atcagagagc	41400
aagtaattta	atttcatggt	gaatcaatga	tacaaaaaaa	aaaggtggag	atacgaacca	41460
aaaagcacgc	aacaagattg	tcaggcatag	cctgggcaat	atagaagaat	tagtagtctt	41520
actttatcag	gggtagatga	gtgcattcct	tccaatgtta	gagcaacgtt	catatcttcc	41580
cacaagctgt	ctaattcttt	ctcctcttct	gttttctcga	taagcacagg	ctcctcgag	41640
ccaaatctca	agtttaaaag	aggtgaatct	tctgtcgagt	agtttattgt	ctcttctcca	41700
tcacagagtt	tttcgccacc	ataaaaactc	tcaccattca	agtggttctt	ctcctccacc	41760

-continued

ctatggaac	tcctcogttc	ccttggtttc	ccatgttcat	ttaccttctc	acttgagctc	41820
tcctaacta	caggatcttc	tctagaatca	acctctgcta	tttgtccca	ggagaaaata	41880
tcttccttga	aaacgtcttt	gctctccaac	atagatttag	ctagcagggt	aattacatcg	41940
aagtatatgt	tcctacgaaa	agtcctcgac	ttcttatgat	ggtatacttt	ttcacttggg	42000
ttcttttcgc	atgtagcatc	gtctctggta	ccacctctt	catcctcaga	acaaacgaaa	42060
tcactatcag	agctctctac	atcagaagaa	tctgaacttt	ccccacgta	atcagatgag	42120
tcagaatcat	tagcatcttc	atccatatca	ctctcaccag	aatctctatc	ttcgccacta	42180
acttcctctc	ttgaatcagt	ccctaattcc	tctaaagggt	cttcctcatc	atcagaacta	42240
gaacttagac	ttacaacctc	atcagacaca	taagtcttct	cctcacccct	caaattagca	42300
tcatccagaa	gaatatcaca	aacctcagaa	ccagcattga	catcctcgac	atgctgattc	42360
tctcccgca	tagttccaag	gaaaacaaca	tcatcatcat	catccggaga	taaagggttc	42420
tctttaccac	tcacttttagc	atcatcctca	tcaaaatcac	aaaccctagg	agaaatcacc	42480
gaagcagaac	caacattgtc	atcatcctct	acatgatcat	tctctctctg	aacagttcca	42540
acgaaaacaa	catcatcatc	atcatctata	gggttaaaat	tagaagcttt	ctcttcacat	42600
cccaaattac	gatcatcagc	atcaaaatca	caaaccctat	caccaaatac	aaacgatttc	42660
gattgtaaat	tcacagaagt	agaaccaaca	ttctcatcat	ctcgttacc	ttcaggatat	42720
tcagttctaa	caaatacaac	atcgctcgta	tcaccccttc	tccgcccgtc	tttctctctc	42780
ctcggaag	gtgaacaagc	atcacgcag	ttaactctcc	tcttctcagt	cctcgaattt	42840
acacaacca	gagactgac	ttcttctct	ccagaaatcc	cctttgattt	attgagaatc	42900
gagttcaagt	aagattcagt	cctcgaacga	gttcttctag	caacacaact	agtcataatc	42960
atggctgaag	agagaaatcg	aaaaccctag	aatagtggat	ttgacgaaga	aggaagaaag	43020
ctcatgtcct	ctctcctctg	ttgtctctgc	aaattttttt	ttttttgcct	ttggcggtta	43080
cgattcaact	ttctcgaaac	attcaataaa	aaaggttttt	ttttgtttgt	ttctttatta	43140
gggtaaattg	taatatattga	aaagatgagg	ggctaagtga	aaattaatag	ataagtttgc	43200
ttcatattcta	ttggtgggaa	ttgattgaca	tgagagtttc	aaagttaaag	gtggatgtta	43260
ctggttctgt	cagtagatag	aggattttaa	taggggtcaa	aatattagt	gaaaaatgaa	43320
aagcttttag	tatttcaa	atattttaca	aattgggttac	taattttttt	tgttgtaatg	43380
gtgctggtaa	ggacaaaaat	tctattattt	gcttaggatt	ctctccttta	tgttttgatg	43440
tgagtatgtg	actaatattt	ctgcaaacga	tttagtgttg	ggttattaga	gaaagatatc	43500
ttctggagaa	cgaaaatctg	gaggtgagat	cattgctgat	ttgaattgta	aattttttat	43560
gaaaattttg	tggtaatcaa	atctaattgc	gaaaattcat	agacagagtc	actcacagag	43620
atatgcaatt	tatacaactt	ttacgattat	atattttttg	cggtttaatac	tatcaaaaca	43680
tgtattataa	cttaaataga	agattattat	tatttggttg	tttcaaaaat	aaaaagagtt	43740
tgcggaataa	ttagtagaac	atggttagagt	taataataat	atgatagaga	atgagttaaa	43800
tagaatgtta	actttgacac	ttctaatttt	ataattacag	taataaatta	tacataccag	43860
aggctatata	taaaaatact	aaaaaaataa	attatatcac	tttaatactt	atttaaaatg	43920
tgtacaacatc	taaaaaatg	gtctggtagt	atattacgaa	ttaaatttta	ttagttatat	43980
aaaaaaaaat	atgatgaaga	ggctcaattc	ctcaaagcaa	tttatgtgag	agttcaatta	44040

-continued

ttattagatt tctaaaatta tatttttgag attggatttc taaagttcta tttttgagtt	44100
tgaactttca acgagttata aaattggaaa tacaaaatct tatggaatcg aattatctaa	44160
ccgggttttag gttccaccga tcagactaat cggatcagct caagttctaa aatatgtaaa	44220
agctcgtaac gagaactttc aaatttgggg caaaaaaaca tgaactttct aatgtcatt	44280
taaatctcaa agtttggttt gactttataa taaaatggtt aagttttgtt gaccgagtca	44340
ttttagggcat gtcgttaaat ttttgtaaat taaaagatga ccgagttaac ttttcataat	44400
gtttgggttt agaacgatac cttatatgta tttagaaaga tacctcataa gctttctgat	44460
accctacggt gtgatggttt ttctctttaa tgacatcata tatgtatata agatttattt	44520
atttcccaaa cgaaaaatta actctgtcgt ctttttatta acgaaaattt atcgatatgc	44580
ctataatgat tcaatcaacg aaatttacta ttttattata aagtc aaacc aaaatttaag	44640
attttattaa cattttgaaa attcatgttt ttatgtccca aatttgaaag ttgcaaaaa	44700
aattttgaca tttttccga tctaaattta ccatttaaac aataaaaagt aaaaggtcaa	44760
aatgtaacat tttagaaaca caaaggacaa aatgataaat atcatgccaa cgcgttgatc	44820
ggtttttagc tcttttcttc cgatttcgtt ttttctttct tcatctctaa gattaagcac	44880
accttggtcat ccattttcgt ttgaatctca acgttttttg cttgatttta gcttctgaga	44940
gagtcagat ccaaggaatt ggaggaagaa taatgtcggc aaggcatggg caatcgtcgt	45000
atcgagatcg gagtgacgaa ttcttcaaaa ttgtggagac tctaaggaga tcgattgctc	45060
cggctccggc ggcaataat gtgccgatg gtaataatcg gaacgatggg gcgagaagag	45120
aggatcttat caacaaatcc gagttcaaca agagagcttc tcatattggt ttagctatta	45180
atcaaacgtc gcagaagcta tcgaagcttg cgaacgtaa gttgttttag tgtttagctc	45240
atttgatttg gtgagatctt gtgaatctgg attagttaac ctgtaatggc tttagcttag	45300
gaattagaat ggtgcttaga tcgagaaccg atttatttag tgggtgtaat ttctcaagtg	45360
cttatcttag tactcgagaa tgtgtaattg gtatggaaga tgctgtattt tgtgcataat	45420
tgtaacttt gtcgtttttg ttgcaaatca gttgcaaaga ggacatcagt gtttgatgat	45480
cctactcagg agatacaaga gctgacgta gtcacaaagc aagagatctc tgctctaaat	45540
tctgctctgg tcgaccttca attgttccgc agttctcaga atgatgaagg gaataactct	45600
agagacaggg acaaaagtac tcaactcagca actgttgtt acgatctaaa gtatcgtttg	45660
atggatacta ccaaagagtt taaggatgtt cttaccatga gaaccgaggt tttttctttg	45720
ttccgtcgat ttcttttcga taaagcttga gtatgcagtc gtgactgagc tatgatttac	45780
ttgtctggtg ttgcagaata tgaaggtcca tgaaagtaga aggcaactct tctcttcaaa	45840
tgcttcaaaa gaatcaacaa acccattcgt tcgccagcgt cctttggctg ctaaggctgc	45900
tgctagttaa tctgttcctc ttccatgggc aaatgggtct tcttcatctt catctcagtt	45960
agtccatgg taaagtacta ttaaccgaat aaccacagtt gcaaatacag catacttctt	46020
gaaatctggc taagaatttg gtctaattgg agtccagaaa gttgtcatgg gtttactaga	46080
atgttttctg taccatgtca gtcgttgctg atttaagatc tggcttgtca agaactcaag	46140
acgtttcata atctctgctt ttacaggaag ccgggagagg gagaatcttc accattgttg	46200
cagcagagtc aacaacaaca gcaacagcaa cagcaacaaa tggccccatt gcaagacaca	46260
tatatgcagg gtcgagcaga agctctacac accgttgaat caacaatcca tgagctaagc	46320

-continued

agtatcttca cacaactagc aaccatggtt tctcaacaag gggaaattgc aatcagggttc	46380
gtcttcctcc tgaggaaaa catcaatatg tgttataatg tccaatgatt catagaacat	46440
ttgtggaag gaagtataat ataatgttgc tggtttatac aggatcgatc aaaacatgga	46500
agatacatta gcaaattgtg aaggcgcaca gagccaactg gccaggatc tcaacagtat	46560
atcatcaaac cgatggctaa tgatgaagat tttcttcgta ctcatcgcat ttctcatgat	46620
ttctctcttc ttcgtggcat aaacattgag actaaaaatt gattttaaag agtttcaaca	46680
tctagacatg gttcttttgt cctatcgat cttattgaca tgtaaagact taagagaaaa	46740
aagggatctt atgctctggt cagattaatg gagagatata taatgtcact ttcatgtaaa	46800
attgctttca agttttccaa acaaatTTTT gtgattctgc ttctcttttc attaccactg	46860
ccaaccatta ggaattgttt cttattccat ttgtacttca tttattatgt aaagacaaaa	46920
aaaaaattt cattaccgtg tttatatatt tatttgtttg taagttaatc aatctacgac	46980
ttgaaaaatt gataaatata tgggatattg tcaaaagctg caacacaata atcaccaatt	47040
catgatctaa gcaatatctc tgttgtttg ttcatacgtt tagctcttca aattcaggac	47100
ttaatggcat aaagagcgcg tccacttcgg tttcatctac ttcgtctaag ctcgctggtt	47160
tccatttcgg attctgcaac aaaaaacaca ctcatgaata agcttatagt aggcactaca	47220
gagacgaaag agtgtgaaat gcagtgagta ggaagataaa aatgacagag acattcttga	47280
ttttataaac tatgttcac tttgacgtg gatatatata tactcagctt catgtttttg	47340
acccaactca agaaattaga aagcacacaa aaaccaattg acagaatgaa ctaaattcatt	47400
aaacattgag tggcaatgaa caaaaacctg gtccttgta gtcaagactg ctctaaccac	47460
ttcagtgaaa tcaactgcga atgcagatct tagagcaatg cggtaactcag ttatcatcac	47520
accgttaagc ttcaaatg aaagaatagt aagcattacg aaggaaaaac tcgtaattta	47580
tcggatacaa agcgtttgaa ctcaactgtt ccatggcgtt gttggtttta cccttggcac	47640
aagcaacttt ggagaagtac ttgtgtgtta ggtaaagaga aaaagggtgcc cctttttcaa	47700
tgcttggac tgcttcatta gcccattctg caactgtacg caaacattgt gtccaaattt	47760
ctcagacaca agacaacgat ttcaatttca ttaatgtaaa cagcaactaa cagtagtctt	47820
tcttagaaat caagattggt aacatacaag cacaaggggt ttggtatttc aatataaaag	47880
gcagaaaaga agtttctttg tcgaaacctt aagagactaa tgcatacat accggaagcc	47940
tcaactgctct gctggaactt tttcagctct tctattgttt ccttaacaga cttacttaca	48000
ctaaaagctg attctatctg aggcaaaagc atttgagggt gggactcagt ttcaggatca	48060
ctgctgtaat ttgataaagt tgcttgatg tgtgtgtgg gatccttgg actgtattac	48120
caacacaaat tctttcttag agactaaaca aagtaatact aaaagctcct atatttgaaa	48180
tatacacatg aaatatatcg aacaaagtat tatacgaggg cctacaggtc agctgacaaa	48240
atggcttctc tgagtgaacc tagctttcca gatgtacat aatgtgtccc aagaccaca	48300
aacagtgcac ctgaaggtgt ggaaatcctt ctcccagtc tacctaggta agcacctgca	48360
gaatttgaag gttgaaaagt tgctggtgga taatgatact acagcaaatt gttatgcaaa	48420
ttgttgagg tctcagctat atgaaacata ccaacagatc cttcgccagg gctatgagct	48480
gcaatatatg aaaatccaac atctggaaac aagccgattc cattttcttg cattgcgagg	48540
accgtcctct ggaaatggag aacaaaacag tgacagaggt aagaaatgca tgataaatgc	48600

-continued

catccacaca	acagtacata	acaaggggtt	actatgtgca	cctctgttat	cactcggtag	48660
cgtccatgtc	cagaaagacc	aagaccaa	cccattgtta	tgccatccat	taaagagata	48720
tatggttttc	tgtatccagc	aatcttgcat	atcaggctat	actcagctgt	aaatacctaa	48780
aaggaagcaa	agttatgaaa	ttaggacatc	ataaacaaga	aacatacatt	ttgaagggtta	48840
atcaaatgaa	ctaaagcaat	tgatggtgaa	taaaaattgt	aagacactga	aggtgactac	48900
ttcaaagaag	ggcagatcca	agcaaaaaa	gagaaagcat	ggagtaaatac	tctagaatcc	48960
tagactaaag	agtggtacca	cacagattat	tttgatcaac	agaaataaag	aatagaccac	49020
ctaacatcat	caaaactagaa	gtttcctact	acttcaagta	cattcagtag	aagtagtatt	49080
tctctatttc	tgtttcagta	atcatcatat	cctgacttat	gggggaaaca	attgcacaac	49140
cagacctcta	tcatacttga	caattgattt	ttggatgtaa	tctgcttcac	atcaccgcct	49200
gcagataaac	taagaccatg	gttataaaaa	tttctatcca	gtgtgaattt	gataaagaaa	49260
actcctttga	tgcaagaaag	ggaaacccaa	aaaacctcac	atagaacaaa	acacagagta	49320
caaaaatacc	aaagagtgtc	aaaatgacat	tttgtatgtt	taaagaaaga	gataagcctt	49380
gatacagcca	ataacacaca	attcatataa	aaaccttttg	cacaagagac	gtatttttgt	49440
ccattaggat	ctctgcaaca	actcctttta	tatccattcc	tgaccattc	ataaatggtc	49500
actctaaac	cttcccattg	cattaaccaa	gaccaaagt	tttctataag	ataagataag	49560
atcagatgca	acatactttc	aacaaattta	tcctacaatc	ttagaaacca	gaatccacta	49620
tataagcaca	ctcaagaaaa	gtttctttta	cctgcacaaa	agggccgaga	tgtgcttcct	49680
tcgacgacaa	cacatttgac	tccaggatca	tattcccatt	catcaagcaa	gcttttgtag	49740
ttcaaatcca	tttctgtaaa	aaaaaatct	tcaataattg	aacttggtt	gtgaaaaagc	49800
ttaattcttt	atgaaaacg	acctagattc	atggcggtga	gagctttagg	tcgatcaaga	49860
gtaatgagag	caacgccatt	tggtataaca	tttcctttca	cgaactcgtc	gcttccactg	49920
gccattactg	agaatttccg	acgatgagat	acagagagct	ttgagaagg	aataagaaaa	49980
gatggagagg	ttctagagaa	gaagatgcga	gcagaagcca	gagataaact	cttgatcatc	50040
actgtgattc	ctaattccac	gtttgtcgga	gaaaccaa	atctgtgact	ttgttttctt	50100
ccaaccacca	aaagtactct	gtgtcagttt	aagggtccac	taatctggag	atatgtcggt	50160
cggctcaata	caaaaagccc	aatagtgacc	aataacagag	ttttgtttta	gggtaaattg	50220
caggccacat	ccgttgacca	tgttttattt	ccatggattg	tcaaagtcaa	atgtggtttt	50280
caccatttgt	cttttcgcct	gcaaaataac	gcttctgccc	ttgcgcttct	ctcttagcaa	50340
ctgttgaaat	gtacggttta	gtttagcagg	tttagttacg	tacactgttc	cgttcagtt	50400
gaagattttt	tgttcacgtt	gacgcgattc	ggttccgtta	caagtgagat	tattccgaag	50460
tgacatcaac	atttaatgcc	cccaaaaagt	ttagatgtga	ctattgtcac	aattaatgca	50520
aaatcatcat	tacatcaa	aatattaaaa	aaattattaa	taataataaa	aaaaaaattt	50580
gttgagataa	agaagatgaa	cccaggaaca	aagaattatt	gcgaacgtac	attgctgctg	50640
aagaagtgtt	gatcttgttt	ttgtaccaag	tttgagcat	tgctcgtgta	agtaatgaag	50700
ttcttttaag	aagttttttt	tatgtcatat	taggatcttg	attgtgttcg	aactccattt	50760
ttttatcaga	atatttgcaa	aatctatatt	tgtttgttta	aaatagtcaa	atgataattg	50820
gtatttaggt	ttataaatgt	gtagtatcga	cattccattt	gttttttata	tgtatagtat	50880

-continued

ttttcagttg	catttaaatgt	gtgtctcctg	aatttatgat	gcagggtgggt	tacggatatatt	50940
cccgtattta	ccatgggtcgg	tgagtggaaa	tgcaaagaag	gagaatggaa	atttgaacct	51000
gaggaaggca	tcttttggtcg	ttgtgtacgc	gttcaggaaa	ccatgacgta	cacggacttt	51060
gtccgtacac	tccgagaggc	tttcagtttg	aagtccaccg	agatcaatcc	cataattagc	51120
tactggatgc	cgggtgaaat	gtcagtgtcg	atagatacta	aacgccctcc	tgtatacatt	51180
gacagtcaga	tgggtctcga	gacgtttttt	ttggttcgtg	gtgggtatct	ttctctcaac	51240
ttgttcgtct	ctttcaataa	tacagtgaag	gcgttcggga	atactgtctc	taggtctcat	51300
ggtgaagttt	ctgaaagtga	ccgtgtgttt	gacagtgccta	caaagtgtga	tgaaaataat	51360
gaggatgcag	aggagcaagc	cgatgttgaa	gacgatacaa	cagaagagga	tgaattggag	51420
gagggcaacg	aagatgatga	tgggtgtgat	gattatcttg	gtgtggatgg	atctattggg	51480
ggaagtacc	tgtatcctac	cacagacggt	agtgaaggta	ttggcgaaga	atacgattac	51540
aacaagagga	acgatgtaat	agtagaagaa	tatgtacaag	gaaatcttga	ggaagacgtc	51600
agtctcactc	aacagactgc	tcaacatcca	gggaaattca	gtcgtttatt	ggaacgtgt	51660
acgtacacgg	gtgttagtgt	acctggcggg	gatacagggg	aattcagtcg	tctcttgga	51720
acgtgtacgt	acactggtct	tagtgtacct	ggcgaagaaa	caatctctcc	attgggttac	51780
ggtacacgct	gcgaaggcca	aagggattcc	aatcttattt	gtatagacga	agtcaccggg	51840
gatgaagtga	tcgaagggtat	tggagcttcc	tcaactggga	cacaaaatga	taatggggta	51900
gttaccgcag	ctgaggccta	ccaagtgtat	aatgaatttc	ctaagttaca	tgggtgcgag	51960
ttggcgaatc	tcggtaacga	tgtctctcct	gtgttcgatg	acttactcaa	ttttcgagct	52020
gatgagacac	aggttgaaat	cagcaccaca	ggcgaactt	tgtttgttgg	aagtgttttt	52080
aagaatagaa	aagtcttgca	acaaacaatg	tctcttcaag	cgataaagca	atgcttttgc	52140
ttcaagcaac	ctaagtcagt	tcctaaaaca	ttgaagatgg	tgtgcgttga	tgagacttgt	52200
caatggcaat	tgactgctcg	cgctgtgaag	gattcagaaa	gtttcaaaat	cacttcgtat	52260
gctacgacac	atacgtgtaa	catcgactct	cggaagaact	acaataaaca	tgctaattat	52320
aagctcctcg	gagaagtgtg	gaggagcaga	tacagttcta	cgcaggggtga	gccgcgagct	52380
gtcgacttac	cacagctgct	cctaaatgat	ctgaatgtgc	gtattttgta	ttctaccgct	52440
tggagagcaa	aagaggttgc	agtggaaaat	gtacgggggg	atgagatagc	aaattacagg	52500
tttttgccga	cctatctgta	tcttctccaa	ttagctaata	cgggtacaat	aacacaccta	52560
cactatacac	cagaagtatga	tggtaagcag	cgcttcaagt	atgtctttgt	ctctcttggc	52620
gcttctatca	aaggtctgat	atatatgagg	aaggtagtgt	tggtagatgg	aacgcagcta	52680
gtcggacctt	acaaaggatg	tctccttatt	gcattgtgcc	aagatgggaa	cttccaaata	52740
ttcccaatag	cttttggtgt	tgttgatggt	gagaccgatg	cttcttgggc	atgggttttt	52800
gaaaagttag	ctgagattgt	tccagacagt	gacgatttaa	tgattgtctc	ggacagacat	52860
tcatctatat	acaaaggcct	aagtgttgtt	tacccgagag	cgcacatgag	agcatgtgct	52920
gttcaccttg	agcgaacacct	ctccacctat	tatggtaagt	ttggtgtgtc	tgctttattt	52980
ttcagcgctg	ccaaagctta	tagggtcagg	gatttcgaga	aataatttga	actgttaagg	53040
gaaaagagtg	ctaaatgcgc	aaaataccta	gaggacatag	gatttgagca	ttggacaaga	53100
gctcactgta	gaggagaacg	ctacaatatc	atgtctagca	acaactctga	gtccatgaat	53160

-continued

catgtgttaa	caaaagtaaa	aacttatccg	attgtttaca	tgatcgagtt	cattcgggat	53220
gtcctaatagc	gatggttcgc	atcgaggagg	aagaaagttg	ctaggtgcaa	atcttctgtg	53280
acacctgagg	ttgatgagag	atttctacaa	gagttgcctg	cgtcaggtaa	atacgctgtg	53340
aagatgtctg	gaccgtggag	ctatcaagtg	acaagtaaat	ccggggagca	ttttcatgtc	53400
gttctggacc	agtgtacgtg	tacgtgtctt	aggtacacca	agttatgaat	cccatgtgaa	53460
cacgctctag	cagctgcaat	cgagcatgga	attgatccga	agtcttgtgt	cggatgggtg	53520
tatggccttc	aaacgttttc	tgattctttc	caagaaccaa	ttctacctat	cgcggatccg	53580
aaagatgttg	tcatcccaca	acatatagtg	gatttgatac	tcattcctcc	atacagtaga	53640
cggccaccag	gaagaccact	gtccaaaagg	atcccatcta	gaggtgaaaa	ccgggtatgt	53700
actaacatcg	ttcattcatc	tacttgatgt	ttgagtatta	ccatgcggag	atttgtaact	53760
tcctgtcttt	gtcacttcag	aaaagcaggt	cataacaaga	agacgtgcaa	aaatccaatc	53820
tagttccggt	gacaagtttc	tgagccggcc	aatatagtg	acgcttctgt	gttttctcgg	53880
tgtagcgaag	tttatatata	atccacctgt	aactatgtac	gttgtgtacg	tacgtttaaa	53940
ctctacgttc	gtataactat	gaacgttttt	gtctgtgtac	gtgttaaagt	tcatgtaata	54000
agggtagaca	gaggtggttg	ctcaatattt	tgagtaattt	gaaataagtt	ttttgaaaat	54060
ctaacttatt	ccccctctta	aatgaacgga	gttgagagaca	tggttgggta	tattgtggtt	54120
gtgttggtat	ttaatggaaa	ggacgtaact	gttgctgttt	attgtcaa	caacgttatg	54180
attaccctat	atgacgggta	ccaggctatt	acaaaaaatt	tgaatcaagt	gtagttgtag	54240
ccgcattaat	aactgtagcc	gttcagatt	tcctctat	aagtctctc	ccttccttaa	54300
ttttatccac	aatcttcatt	ctaaacccat	ttcttgtgcc	ggcagttata	tcattcacaa	54360
aaaaaaattt	agtctttcga	atcgcttggt	cttgctatcc	gtgagatgga	ggggtttcat	54420
acggtctcag	aattgtgtcc	tctgatcact	gggtggcgaa	tttatgtctt	tgctgtgagg	54480
gtcttcaaga	aggttatatc	gcccacgctc	ttcgaactcg	gccttatcct	ggcagattac	54540
gaggttat	ccgttaaatc	attatgtgtt	tttctcgctc	attcgcttaa	cttatatatt	54600
aaactctcta	ctatcatggt	ctgtatgtaa	agaataaccg	tattgaggcc	accgttgatc	54660
gtcgccttgc	tcogttttac	gaagaccgat	tcgttgaaaa	tgagtggag	acgatcacct	54720
ctttcttggt	ctgtaagtca	actgacttag	tcagagcaac	aaagcatgaa	tacggcatac	54780
tgtttatgga	tcagatcgct	gtcgttcatt	ccccaccgag	atcaccacca	gttcagatt	54840
tcgatttcac	tcctctcgac	tacattttgg	ataaatctgc	ttacaagaac	gttttagtcg	54900
gtgaggagga	ataactctga	cgcttcattt	attacatttt	tgaaaatggt	tgtttattat	54960
ctttgtttct	ccatcatcta	gatgtcattg	gagcgttggt	tgatgttggc	gcattgacta	55020
cagactatta	cggtctcaaa	ctgagtttca	agttaaaaga	tagatagtaa	gtgttctttt	55080
catttgtttt	atttacaatg	tcactctaaa	cgctaacagt	ccttatttct	aacgtagcaa	55140
tgaagctctg	gagtgcaag	cccgcaatca	acatgcggag	tacttgatg	gttactttta	55200
gagtttgggt	aaagaaaatt	ttgttgtcgc	actcagcttt	tgccgtctaa	ccgaactgtc	55260
taaccggaag	cttgagagcc	atggtgctat	ttcaaagtg	gttgccaatc	cggacagagc	55320
agaagttg	gatatctcga	tggtgggttt	ttagctagag	ggaatgaacc	attattctcc	55380
gttttaatat	aatataagac	tacagttcaa	ttaagttttc	tttccaggt	gttgtgtgtg	55440

-continued

tttcataag aaccaaacta catttcagtc gttgttggtt cttgtttccg tttaaatttt	55500
aatctaattt attgttttat gtacgtgcct tgttttattg ttcaaactct ctcgcaatgt	55560
acgtatagtt tctggtgtac gctatcgtgt acgtatatta gttttggtgc cttcaacatc	55620
aaaaactggt agatttatac ttatgttgtg tacgcaattt cccaatgtac gataaacgaa	55680
gtttaatata ctatatagtg tacgtacacg aagacatcta gtaatactta aagtggtaat	55740
tcattcaaaa gccagaaaca taaaaaagga caggcagttt atataaaggg aaattcataa	55800
gataagataa ataaactgga acgattcaaa aaacataatt ttttaaacga agcatacaca	55860
tgtgttgcaa aatatgtgtg catccgtcat accaaaagac tcgagagtag agccgagtag	55920
agtagcttag accttagcag ctttgcttct agtagaatat agccatctcc ctggcgtagc	55980
agtagaaaaa acaaaccctg cagccatgat ggaacgtgca ttaccctacg caactcggcg	56040
ggtggacaaa cctccaagac ctttccagc ggcgagcatt ttcagtggtga agctggctga	56100
tgtgatacat caccatctct gcaaccgcgt ggccagcttc aaggggtgggt gtcatttccc	56160
agaacttttc gctgataggg aggaaatcca agtgattacc ggtacacgta aggaacaaac	56220
cgggtgcgaa gcaagcaaag ggatcagata aagcgacggt tgtcatgact gaaaggccct	56280
gagctactgt cacgaagttg ctgcgttgcc gaacaccttc aaggtaaacc gccgggatgg	56340
atccacgttc tacaagcacc tcgaagtatg gccagaaggg agaccccggt tggagtaacg	56400
acggccacgt catgaattca tcgattgcaa ggggttggtg gtcagaactg tagactaagt	56460
tccttagcgt ctctatggga gtctggctga cggggaggtg gttgacaagt atgcttatgg	56520
ggtttgcaa ttggattcct aggtacaact cggttggagg ggttagcgga ggtaggaaag	56580
gcattgtttt tcaaacgagg aggtctagac ggatttgtgt attcaactaa gtgtgtttgt	56640
catgagtttt aaagaggaga ataaccatca atgtaagtgg gtgggtgacg tggatgtata	56700
tttagtgcat catattagta acttttaatt taaacaaatt cctgtgtacc tctcataatc	56760
gcgccgtgta cgtctacaaa atactcatta cgttctctga cagaagggtg agttgacgtc	56820
tgtttgtatg gcatataagt ggaacaatat ggcttcatgt atagtgtat attaggggtc	56880
aacgcattta ctaacaatct tttcaataa acattaactt acttatttta tctgccaaat	56940
caattgagta cattgctcgg taaccgtgta cgtctaatag agatattaat gagagtgtga	57000
tggaggtcac ttctagatg tatttgacag cgagtaatta atttgagcg aaatcttcta	57060
tatatgggtt atacgacgtc aaacaaattt ccttttgtct tcatttacia atatcgctgc	57120
tcactatgta atctttatag ttggaattaa actatgccac gtgtacatta tcatatatcg	57180
tgtacgtaga tattttatta ttttattatt ttaatgatgt ctcatattag agtactttga	57240
gctagtatgt ttcaatcaat tacaaagtgg taatttaatt aaaaacaaac ttccattact	57300
gataatcaac gtaaaaagct cacatattgt tgcattgaaa acttataaat atgacaggat	57360
cgtggcgaat gtgagctcta caggatctgg ctaatccgtt tagcacacaa ccagatgtaa	57420
cattggttgc aaagattatt gaagagtctc gatctaattg aacacacctc cgcgcattca	57480
ggagtgccta cgtcaacaca ttccgggaac gaaaaactgt tagcgtatgt gtattttaaa	57540
gtattacat atttctttat atcttttagc acctcctcac aaatgtcacg tgcgtcctcc	57600
gattccaaag cataatggtt gcttcogaag agccgaagggt agacaccacc catgtgagca	57660
ttaccagcac atatgtaaa aattgcgcat gcaagggttg cagcggcggt ccttggggcg	57720

-continued

atcgctcta agttaggat agctccctgg atgtcagact catgtgtaac gagacgcagt	57780
ccttcattggt aaatagccgt ggggttacca gcttgtaagc accgttgaaa gaaggatcta	57840
tagcgatctt cggagttgat gtcatttgga tcgtggcctg ccgcgtagaa gtcacgcgga	57900
tcgtcgaca tgctgaaat gtttgcatth ttgaggacat ccggacggtg gacaatgtct	57960
cttcacagag gaccggattt taacataggt ccgaggtacc accaacattt gtcagccatt	58020
ttcttggtta tcttcgcaag caaatcgtca ggaatatttg ggtttgcat atttaggagt	58080
aagggttttc gagaaaatga aatttgaaca cttaaataag catcattgaa gatatgggtg	58140
ggtaagttat ggttgatttt attgcaaagg tattaagtga ttatgtgtat tcatattgtc	58200
aaatcaaagt aatagtattc catatataat ttgttatcgt tgttatgagc aacctctttt	58260
taataacagc ttaaaactag acgtgtacgt ttactgacg gtcttagtgt acgtccacat	58320
ttacatttct acatttactc aacaacaggt gtacgttgta gtgtatgttt tagtgaacgt	58380
ccacatttac atttctacat ttgcccaaca aacagtgtac gttgtagtgt acgtttaagt	58440
gtacgtccac atttacattt ctacatttgc ccaacaaaca gtgtacgttg tagtgtacgt	58500
tttagtgtac gtccatattt acatttctac atttactcaa cagacagtgt acgctgtagt	58560
gtactattag tgtacgtcca ttcataaata tcaccattta tgagacaaac caaagacctc	58620
atatgtttgc atgtgttatt ttttagtgta cgttagagtt gatattctcat gctagtgaac	58680
gtccatatct agttttccga gacaaagaaa aaacctctaa gtattatttg gtatgtgcac	58740
gtgtacggag ttgtggacgc ttagatttta atatccaaat ttacatttac tgcagtgtct	58800
aaatatcata tgtgaatttg gcgaaaaaat attcaacttg agaacataa cacaccttgc	58860
aaattttcta agcaataata taatttcaac ataacataa acaatatagt agaaggctta	58920
tcataatttg aaacataaca tagcggataa cataaacaaa catataaagt agaattggaat	58980
aactatagca tttagtaaac acgcttgcca cagaccaga ggtaacagcg gttgcaaacg	59040
ttttgaaaag ctctgatac catgtaacaa tataaggcgc aaggaggcat actaatcca	59100
tggttggtag gataagagaa cgtaggacca tatgtattgc tgtatggagg gtcaaacttc	59160
ttttttcct cgatgaactc atcaccocaa actcgagtgg caaccgagtc caatggataa	59220
tggttgccgg tgaagagctg tagaacaag ccgcccataa aatcataccc agcacatatg	59280
aatacaatgg cgcagcaag tgttgcatth gctcgtactg gagcatgacg ctgtaagagc	59340
ctgatggctc cattgatgtt tcgttcacgc gttagaacac gaataccttc gtaatacacg	59400
gccgtgggat tattagctgc aaaacacctt aagaaaaatg ttcgatgtcg gccttcatca	59460
gcgatccgg agggattgtg gcccctccag cttggaagtc atcgggcgag tcacacaagg	59520
tgagaatatt ggcattcttg aggcacatcc gacggtagac aatgtctctt ccacgagtgc	59580
cggctcgtac caatggacca aggtcggacc agcattggtc agccagggtc ttagctatct	59640
tcgagagtaa atcgtcggga agagtccggt taaccatggt tgaaatgaga tagtcagagg	59700
tcgtcagtaa tttcttctact ttaataacca ctactgtgaa tttggttggg gcaattacgg	59760
ttgaaatgaa tgtcatactt gtgttggtta gcgttggtg tcaaatcaat ttatcacttg	59820
agttatgtta taaagttaat atgcagtgtg cgagacatat agatatgtac gtacacattt	59880
acattatgta attataagtg aacgatgtgt acgacaatat gatagaggac aataaattaa	59940
ctctcgacaa tctccaaatt aaatgtagtt gcaatcatat atataagtca ggtttgttca	60000

-continued

taataattgg ttgcaaagaa ccagaatatt gttcttggac gattaaacca agtgcacagt	60060
accggcctcc cgcttgccgt gctcgtagat gtcaacagag tatttctgac ggtagaagac	60120
catatctgct tctcgaatgg ttgtcagatc cgaaaaaggg tgtccaaacg caagcaactc	60180
cagggaacttc atcgtgtatg gcccgcgaatc tcttgtagtt gggttttgag caacggtagg	60240
acaacgcaca tactcaaaag gttcaactga gtacggagaa atcaatacat cctggcacat	60300
cgctctgact aagtacggca tcatttcgca aataggagtc attctcgcct tcacagccga	60360
ctcccacgtg tgcgatatta aagcgtcata cacagttatt agccgttcat taaggttgat	60420
tcccaaaaca acccaatggt cactcttcca gttcatcgga gcatacacia catcgacatc	60480
cttcacccac tttatattgg gttcacggtt tatgtgaact ccattagcaa tgcgggatag	60540
taattttacc catttgaaac cgcttcttatt aatacacttc ttgaagtcgc ccacccgctt	60600
ggtaagaaga tgcgtcagca tcatgtccaa tacaatacat ctgtttgcta tcatctgttc	60660
cccggtcttg tgccacatca acattgccat ctttctaaa tgctgaaatg ttcaaaaaat	60720
atatcagtg acgtacactt ccacgaaaaa tttaatcatc tatgtacaat ttataacaag	60780
gaatgtacgt atataaatgc aacgtaccgt actttcgacc cattcaccag gtgttagaag	60840
gttcgcaaaa aagtcgttgg gcacttttcg aaacccgaaa agagtgtgaa tcctgtaaaa	60900
tccgaaaaa tatttgataa caaccttaaa aatgatgatt gagatattaa ataatccatt	60960
ttactataat gtacttacac agagggtcct ttaatgatat tgctgaattc ctctactttt	61020
gaagggtcaa ccacgtccag ggggttgat gagacatttc cagccctgat aacaacctct	61080
atctgttgag ctggtacatc gggcaccaat actgtttttg ccttaaaacc atgaggatgg	61140
gcacatccct gttcatagac aagttgttct ttcactttct ttggggacga ttttttttct	61200
tcttttgctt cgacagcagc agcattgtct gattcatcct gttacaaat atgattgtca	61260
acaacagttt catagctttt tctatgttta aaataggtac atacattatt tttaggaata	61320
tcggtcttgt ctgcttcacg tgaccctggg ttaggatcct cagtttctac agacccctcg	61380
acaatacacg gctcttccat gactacctgt gcaagttaac gtacacaagt atatataatg	61440
tacacatata atcctatgta tgctattccg aaaataagaa ttgaacaca caaattgtta	61500
cctccaaatc agtggctgca gctaaatcac aaacggcctt ggtatccaac tcccatatga	61560
cctgcataat ataaatttta tacatttgaa cagtaaatac tacattttat ttgttgcaag	61620
agttagctta cattgttaac tacttgaaat ggtgacgggg gatcactgtg ggtcttctat	61680
ggtatttcaa cttccatcat tgtcactaat ttggattctc ctgcccctc aacctgtaga	61740
atttagcatc ggaaagaaac tttttagtaa tgtatgaaac aatatatatg tacacagaac	61800
atagtgacaa aaatgtacgt acacttgtgg caggtacgtg agctgtgtct ccatcagaag	61860
gctcataagt ttcttctgac tcgctgccat tagcttcagt cgatgcactg gcatctcttg	61920
cttcttcctg aaacaaaaat aatttgcaaa gtcttatggg cgtacacata tggaaaaaga	61980
acaagtgctt accatctcta tgtcagcatc agctatttca tcggcccctg tatgggtacc	62040
aactaagcca ggttcagcaa catcctttgg tcattatcc tgaacatag gctaccatag	62100
agttaagcgt tagataatg aagaaattaa tgtaatagtt atgcatatag tcatggaaat	62160
ctgtatataa cggttaccag aagatctacc tctgaaactc ttgcagacac attagatgga	62220
acttcttcca tctcgggtgc aacagtagct tottccacag gcctctcttc ctgtgttaca	62280

-continued

aagacagtcc	cagacattga	acctaagcag	acagtcttac	caacaacaat	atttcctcta	62340
tcctgaaaca	tagtttaact	aacttcttaa	gcgaatatgt	acgtacacca	tttaaacaga	62400
ttgttaagag	tattagtaga	cgacatacca	gatgggattg	tgtggatcca	gcataaggaa	62460
ctattaaaaa	tccagtgctc	aatatctctt	cttcagcggc	agaagaattc	aatccatcat	62520
ttgcagtaaa	atcactttca	ttattagcgc	catcctgtat	cgaaccagac	agtacagtta	62580
tacatatgga	ttataaaata	tttagcaatg	tacgtacatg	acttacatgt	tcatcaatga	62640
atccttgcat	atcttcccta	ttggaatcag	gttcctcaat	agtggtcaaca	gaaaagaaa	62700
catttttctt	cggctttgaa	ggttttcttc	ttccgcagtg	gtatgtactc	cagggacatc	62760
caaatcttcc	tctttggaat	tacagccaac	agaaagtcca	ggagtgtcca	catcatgaga	62820
aacacgagtc	aaccttgaag	actttctcgt	tccagtgctt	cctttttag	acatccttct	62880
cgatggcttc	ttcgcagtaa	ccggtgcctt	atccatagaa	ataggtgggt	ccgcgcaatc	62940
atgggaaggt	atggtttgtt	ctggcctttt	gccttttggg	tcaagggcag	cacgtatctg	63000
gtcacggaca	aagtctttca	tatcttcatg	cattggatac	atttggctct	ccagctcctt	63060
caaagacctc	ctcatttcaa	caataaacct	ctcaggcaca	ggctctccaa	aacaatcctc	63120
agcagtctct	tctttatcac	agtccatttc	agatttctgt	tttttcataa	caggttcttt	63180
agggctcggg	cccggtgcat	ttcgtttctt	cttaacatct	acaatagacc	agttgtctgt	63240
attaattggt	tttcttctag	gtgcaatacc	aaaaaaattt	gtaaaaaaa	cattttatact	63300
tatacttaca	tgtctcttct	agttgtggtg	ggcgcgactg	ctttggtact	ctagcatagc	63360
ctccaaccca	ttcatcctcc	tgccactggg	gcccacttat	caacagtgcc	tcaatgtaat	63420
caacccttgg	atcatcaact	tcatcatccc	atgaaagaga	tggaggacaa	acattatcac	63480
ctggtttcac	aatatagttg	acctcaacct	acacaaataa	gtatgttaac	catagataat	63540
gtacatttac	ttattatgta	cgtacataca	caatttaaca	cacctcatca	gcagcctcac	63600
actccaaaat	gcgagatgtc	cggattgtct	gcagagaggg	aagacgatgg	atagacctct	63660
cagcaaatgt	cctgttaggt	acatcatcag	gtcctagctt	agcaattgat	gggattgttt	63720
cgaaagctaa	aagctggaga	gctaattggg	aaccatggag	tgcataagac	cctacaagta	63780
taccoccttat	tagtttctga	gcatcgtaag	gagtttgga	atttgctatc	ctccccaaag	63840
tacgggtaaa	ggacaccctt	ccccatgggt	atttgcagaa	aaactccagg	ttcttagtag	63900
tctcgacagt	cttcgaagta	ggcctgttgg	ggttactgtg	ggccgctacc	acaccgtcaa	63960
ccaaaataat	aagagatagt	gccagttgct	tccacccttc	catgctttcc	tcctcctgca	64020
gccagctcac	caaatcagca	attgttgga	ccgtgtcccc	aaacctctca	tgaacagag	64080
tcttccaaac	actctcacac	tcaggtttca	cactgataac	atcatcaacg	tctttcttct	64140
tgggatactt	cccacattcc	agcccagtta	agatactaaa	ttctcttaat	ctaaatctga	64200
ttggatgacc	gccgaataca	atccacatct	cattcacctt	ctttgtcact	aactgacggc	64260
atattagtcc	aaggacaagc	ttagcattga	aagatgcctt	gtttttgggt	attttaaaaa	64320
gtttcccaaa	tggagaatcg	agaagaaact	gcactctctg	tttccccttt	agtacttttg	64380
ctatgtgaga	aatgtactcc	ggttttagagt	acgcgttcat	cttcgtctcg	gaagggtaac	64440
ggtctagagc	aaagagtctt	ggtggttaact	cttcaactgt	aagatcttca	gtcggagcaa	64500
cacttggaact	acctgcaatt	aaaagtcata	tactgttgag	caaacagaag	tagtgtacgt	64560

-continued

acacagacat tttctagatt ataactcaaa atcatatttt tggcggaaac agagattttcc	64620
aattctactt ctcaattttt catcctaaag atatgctaaa ttccgacaaa ttctaactact	64680
ttaatcaaac aagcttttaca attctgatac acccaaagac aacaaacggg aaattagaga	64740
ttctaaccga gttcctccgt tgcctttgtc gcagagcggt ttctccgcgg ggaactcctt	64800
aatacacgct tctccttccc tggaccatgc tgcttcctcg ggacttcgtc gttaacttta	64860
ttcccagcca tagctaaaga tgctatttcg ccggagggtt aaatttcgcc ggaaaaaaca	64920
gtaactttct cttggcttcg ccgggaaaaa cagtaacttt ctcttggttt cgcgggaaa	64980
aacagtaact ttctcttggc ttgcgcggga aaaacagtaa ctttctcttg gcttcgccgg	65040
gaaaaacagt aagtttcact tcatcggtatg tgggagttca gtggcacaaat tcaactaccc	65100
aacaattcga agttatgtcg aattttgaag caatggatac tgaatttcgt ttcgctttct	65160
gattctctct caatctctcc gttattctct ttttcgtttt ctcggtttct tcgtttcagt	65220
tttttttttt ttgtctttt tcaaaactga accgcatcaa ctaaacgaa accaactgct	65280
gtagggtttg gtttacgcg tcaacttaac cgatacgaac tgctgcacgg ttgggtttgt	65340
gccattatta agttttccca aggtaatttg gtccgaaatt gaatgaaaag acaaacgtg	65400
aaaaccagtt ctcaacttga caaacggtga gagcaaatct ttttctgtcc acttttgtga	65460
aaatttcttt ttgtttaatt gatgattgaa aagaaaacgg cacaaccaa acctcaatct	65520
ttgttttttc ctcaatgtga gtattattat tagcatcaca tcttctatgg aaagatacca	65580
aagaatgaaa aaaaatatat atacaaaatc atataagtgt atcatatata gatcgtagtt	65640
caatagagtt aaaacatttc ctcaaatctt tacagttaaa aagaagaatt caaaaaaaaa	65700
aaaacatgta atttcttgca tctttaacca taaccttgac cgggtctttt aatattgccg	65760
gtcgcctttc cctgtgtggt ctgaccgtca gtgctcgacc aggacgggta caaatcatat	65820
tcactaattg ggtaactata taaatcttga gtctcaagtc ccactttcct aaacttggtc	65880
atgccttcac tgtcttggct cgagtcataa tctgtgcttc ctccagatga gccatacaca	65940
ttgtgtgtgc ccggtgtgat tccttggttt agatctgatg gtgatatgtt tccttctagc	66000
acacgtgcaa cctgctcggt atttttacac caatcactca ttcgattagc attagtaaca	66060
ttttaaatg actactattt agcaaattga aaaaagggtg agctaagaaa acatgtttgt	66120
cccattgcta gggctgatca aataagcaaa atataaggct aggcatacat accaaaccaa	66180
ctgagatcga acacggtatt aaccgaaacc aattcggagc ggcaaccaa acgacccaaa	66240
tataaatcaa tctttgtaag ttatatgttt ccgaagccaa aaactaaact gaactgacaa	66300
atgaaaaatat cttggttcta ggtttctaac tccgtatata gtttgaaaac cggaaagcgt	66360
aaatgccttt agagaaaaga aaagaatgat taatttacct gatccatgcg aggtctacgt	66420
ggagctgtgg accgaacaca agctgcggca caagcaacca tgcgagccat ctcttctttg	66480
tcatactcat tattcagttt ttatcaacc acaacctcaa agtttcctaa ttcagatact	66540
tggttaagca aaggtcgtgc ctggtccaaa ttattcacac atagggtttt cttcattcaa	66600
aaatcacatc gtaaacatta tgtgaagggt agattatttg catgtaatta tattacttac	66660
ccaatcaacc aagctgttat ctgcatggac attgtttaca tcaatagggc gacgtccagt	66720
tattagctcc agaagtacaa cgccaaatga gaaaacgtca gacttttccg tgagttttcc	66780
gcttgaagca tattccggag ccaaatacct aaaatccaaa agtttcataa tatgtaaaga	66840

-continued

ttcacataaa	cttttaacaa	aactgcataa	atgtagatat	ttggtagctt	caagattatc	66900
tttaccctaaa	agttcccatc	acacgtgtag	atacatgagt	atttgtatca	gaagcaatct	66960
tggaagacc	aaaatcagca	acctaagaaa	tagaagagag	acacatgata	caaagatact	67020
tcatgacaat	atcgtaacaa	ataacaaata	aaaggaagaa	taaatcatac	ctttgcttca	67080
aatttgaaat	ctatcaatat	gtttgacgcc	ttgatatcac	ggtgaataat	tttaggattg	67140
cctgcatata	tataaatgcc	tagcgagtaa	acaacatga	cctagctata	tattttattta	67200
ttttgagaaa	caaatatgga	attataccat	gacctatatt	tattacacat	gatcttatag	67260
acttacaatt	ttcatgaaga	taagacaatc	ctttggcaga	accaacagca	atcttcaatc	67320
ttgaactcca	ttccatcgta	ggccgtccct	ttcctgtcac	aaattcgatt	cccaagaatt	67380
tactaatga	aaacattgaa	ggatgcaaga	aacaataaga	gagtttattt	tttatttttt	67440
ctaaccatgg	aggtgaaact	cgagagtgtt	gttgggaaca	aactcataga	caagcaatct	67500
ttgagcatcg	gcgatgcaat	aaccgacaag	agcaaccaaa	tgtctatgat	gtactcggct	67560
aatgatccca	acctctgctt	gaaactctct	ttctccttga	gaactccctt	ctttcaattg	67620
tttcacagca	acttctttcc	cattacgtaa	catacctttg	aacacgtatc	cgaaccgcc	67680
ttgtcctaac	aaattggcct	cagagaatcc	attggtggct	ctagatagct	cctcgtaatt	67740
gaaagtgcct	tgatatatgc	ctaagcctaa	tgcaagccct	ggtgatgggtg	gaggaagaac	67800
tgattgatcc	gagtagtttg	agtcgtatgc	tccgtgctg	ccgtgctca	tgaaatgcgg	67860
tgagcgtggt	ggtgcagagg	agggagactt	cgggtggtgg	actgacatca	ccacgtgatc	67920
tgatcgtcgt	gatgcatttt	gttgttgctg	accaccgtaa	gggactccat	ctgagaaacg	67980
catgaacata	atatctataa	aaacggtctc	tctatggatc	catatacaag	tttaaaaatg	68040
tagaaaaaac	atgtaataga	attatcataa	aatgcatatt	catgattaat	ttccccaaat	68100
atatgaaaat	gacgtatatg	taaaagaaaa	tgtagataat	ctagttagca	caactatatt	68160
acgttctaaa	tgattatata	tagccgagta	gaagggaata	tatagttcaa	gaaaatgtag	68220
ataagtcata	cttcaagatt	caattttatt	taaatctttg	ttttacctat	gggagcaggg	68280
ggtgcttcat	tgtctcttcg	tcgtttcttc	ttacaaagga	aaaatatcaa	agccaatgcc	68340
acaagaagca	caaatcctcc	tcctatggcg	atcccgacca	tagctccctt	tgataattct	68400
ttggaagatt	gtgccggtgg	agaagagcct	ggagtggtag	atggagtact	aggagatctc	68460
ccgcagaggg	gactcgtcgg	aggtggaggg	ctcagactgc	ctacccgagg	aggaggagtg	68520
ggtgttgagg	gaggtggcgg	agatcttgcc	ggagtgaag	gagtggcggg	agagggcggg	68580
ggaggagaag	aaggcctcaa	gggagaagac	ggtggagggg	aaggcgtcaa	gggaggtgaa	68640
ggaatgcttg	gagatggagg	cggaggtgaa	gggtggtggg	gagacggagg	cagaggagat	68700
ggtgtcaaa	gaggtgaagg	agtgatggca	ggtggaggag	aaggagtcaa	gggaggtgaa	68760
ggagtgtggg	gagaaggcgt	caaggagggt	gaaggagtgg	tgggagatgg	aggcagagga	68820
gaaggggtca	agggaggtgg	gggagaaggg	gtagagggag	gtataggagg	aacggtggta	68880
ggtggagggtg	gaggaggaag	ggcgggtgga	agggcagtag	gcggaggtgg	aggggggaag	68940
gcgggaggaa	gggcagtagg	tggaggtgga	ggagtaacgg	gcagaggctg	aggcggagga	69000
ggaattggga	gaggctgtgg	aggaggtgac	ggagtaccac	cgggagggcg	cgccgatgac	69060
atatcttccc	gctagtaaaa	aaaaataaag	gacgggataa	agaagaaaaa	agaatgttaa	69120

-continued

aagcgggaaa	agaggtctga	tctctcttgc	caccgagact	agacttgagt	tttgattact	69180
tcaatctttt	taaaaaacct	ttggtacacc	gatatagagg	agaatatacg	taatcttatt	69240
tttgtatgag	ttatatagtt	tgtgttcttc	ttgcatcaaa	agcgaaagaa	agaaatgaga	69300
gcacttacta	cattgacatt	tctccgatgt	ttattattaa	ctatctaaca	tcaagttaca	69360
ccaaaataat	ccattcaaac	tgaaacgtac	ggacaataaa	taagtatttt	gttaatcata	69420
ttcaatcaag	gttcttaata	acagctttta	caaaaaaaaa	aaggttcgtg	ataacaggcc	69480
aattatatca	atatggttct	ttcagatatc	cagataatgt	ttttgctctt	cagataatat	69540
atgtttatgt	aaataattat	tattgtatca	gggtcaatgt	cgtatcattc	tatatatctt	69600
gtttaaaaca	tcccacaaga	cgcttaatct	aagtagaaat	aaaacagaa	gtcaacgact	69660
gattgcaatg	atactttgtc	aagtttggtt	tggtcaaata	ttttgttttt	gtcaaattga	69720
tttaataaaa	ataccagtga	tatatgggcc	gaacaaccaa	ttttacttaa	tgttggtaat	69780
agtattatct	ttatttttgt	taacttagga	aatcttaggt	cgaacctctg	actatgtaat	69840
aatatatagg	gcttataaaa	catcggacaa	acaacgaacg	ttcaaacgaa	gaacttaata	69900
cacgacgtca	ttcaagaatg	gatcaagctt	ttgttgatct	tcaagtgtgt	atgtttacaa	69960
gaaggtcaaa	ataatttggt	agtggacatt	ggacaagcat	gccaaaatgg	ccaaaacaaa	70020
gtccactatc	aattgtctac	aacactataa	cctttttcat	tatgtatttg	tctttttctt	70080
ggaatgtttt	attcaacaac	aacaaaatta	tcgtaagaaa	taaaaaat	gacagtgtga	70140
tataaagaaa	cttataacat	attactttca	atcctatggt	taaaaaaaa	acactaaaaa	70200
tgaaaatcat	tgtgctaaga	atataataac	acattcacat	tacacaaaca	tatatgaaaa	70260
attctaatag	accactttcg	ttatatagtt	aatacgatat	tatagacaag	aagaaactcg	70320
atcaaagcac	gcaaagctct	tatataagcc	ttaaaactca	aggaggcgcc	tcatgacgcc	70380
aagcagctcc	aaatgtgatg	tcattctgtg	aaaatgcccg	atcgcgtcct	cgatgatctc	70440
cactgtcgat	ttccccttaa	tcttctcctg	catgaagtat	gcgacggaga	caggcactac	70500
gacgtcggtt	cctggttgta	tgacatggca	aggcactgac	acttgtccca	aaatctctct	70560
ctcgtcgcta	ccaaacacta	tcttagctaa	ggcaagagct	gtctcgggct	tcatcttctt	70620
aagggaactc	tcgaaccttt	ggaccgagag	agagtctctc	gagtcgacca	cgaatgagga	70680
gaaatcaacc	gcccaagctt	cgtagttgga	gccaatgctg	gtaatgatcg	tgtcaatgtc	70740
ttttgactca	aaccctcctt	tgtaatcctc	actgtttata	tacctatttg	catttcaatc	70800
tagtaacggt	actaatttga	tcatctaaat	atcattttct	ctttgcgcaa	aaaaaaaaaa	70860
tcattttctt	ttatattata	agtacgttgc	attgatatat	atagtcaaac	attatcaatt	70920
gttgcattaa	attatggccg	actttctcac	ctataactat	aataaatatc	taaagtaatc	70980
agtcaaagta	tgatacttca	aacagttaac	tccatttcga	taatttctat	atataggaga	71040
atatagtatg	ttagtgggcc	acgtgatata	taaagatggg	ttaagtttta	gtacttctaa	71100
ctttaaaaat	gacaataaat	atattgattc	atcatatgat	aataaaatca	taacacattt	71160
tttctaattt	tcgatgtatc	tgtatgaaat	gaatgtctta	ttgacatatt	gatagcatat	71220
gtagcatggt	tacatatatt	catctacgta	tatatgcgta	gactgcagag	aagtaagaga	71280
agatgtgatc	atatgtcata	tccatccaac	aagctagaag	ctgatcatgt	ccacatgttc	71340
aacacttggc	aacagttaca	cttcattaca	tatagatatc	aaccacataa	aaactattaa	71400

-continued

atattggcta	atatagtata	ttattgtaaa	tcattatattt	tattttgtct	ttttcctcaa	71460
acacatatc	acgaaggaat	cttgtagcat	cggtcagact	gtataacctt	ataattttga	71520
ttgcatgtgc	ttagattttc	ctttcgaacg	taaattataa	attctaaatc	atcggctctc	71580
atcttattta	cagatatatt	atcatatttt	ataaatagat	tcaatatata	ggcctaaatg	71640
atgttataca	gaacagaaat	agattataca	tatgctttga	tttttacttc	caattatgat	71700
taagggtaa	tggtatacta	agagtgaagc	gtaccttgga	gaagcagcaa	taaggagaa	71760
atgtgtaaac	aagtcggggc	ttttaataga	agcagcacia	ccgatcacc	ccgacatgga	71820
atggcctacg	aacacgacag	ggccaaactt	caactcctcc	ataagagcaa	tgagatcgtc	71880
agagaagacg	tctaaggagt	tatactttga	agggtcatag	agagtttgat	ctttaatggc	71940
tccagagaaa	agccagtcac	agacaaaaac	tttgaaggat	tgggacaaga	ccggtattat	72000
tttatcccaa	accgactggg	cgcctccgaa	accgtgtgcc	aagaccatcg	atctctctcc	72060
tgaaccgatg	attttggcgt	tcatagcgga	tgcaaggccg	gatattctct	gattaaccac	72120
catgcttaag	tacaagagtt	ttgttagatg	ggacaacttt	gataagttgt	gtgcgtatat	72180
ataaaggaga	agagagatgg	attaggttta	tggttcttta	attgacattc	aaccaccgat	72240
tttattaggc	tcgcattttg	gtacatccaa	acatttatgg	ccttttcaac	aatggctgta	72300
ttgtatttca	ttgtccatct	acatagcatc	ttagaggata	aaacaatatc	catcataaat	72360
aagaatatac	tactccatgt	gaaaacgaaa	ctctactttt	tggtagtatt	aaaaaggtaa	72420
taattatttg	gactcgatat	gaagtatata	tactagtttt	gtttgtatgt	gtaaagttta	72480
gagtatcaaa	gtgtctttct	tttttttttg	tgtgtatcaa	agtgtctttt	tttctatcaa	72540
acaatgcatt	tggtgtaaaa	acaatatattt	tttcttttaa	agctaattta	ttattatttt	72600
cgaaaaaata	aatgcattaa	ctaactgtac	atttagacgt	caacatgatc	tcaaattctc	72660
aatacattaa	cataaccata	aatcttaatt	aagttatagt	caatgaagaa	taaagcacac	72720
ggcaaatctc	aataaattga	ttaacatcaa	tccttgctaa	tcataggagt	catctacgaa	72780
acgtataaaa	tgaccttttg	tcttacaata	ataacaacgc	aaattaatat	aattctatta	72840
tatatacatc	attttcgttt	ctctccctta	gactttcgca	tatatagaca	catcgaaaa	72900
atataaaata	caagcaacta	attgtaatca	agtttctttc	tttaataaca	aatcttcaga	72960
agaatggata	tgaagaaaaa	cgatatattt	tgtaatataa	tgttttcaat	ttatatagta	73020
taataaaaca	catatatatt	tatacaccaa	tgactcatta	gctatatatt	atatttcaaa	73080
agcagtagct	gtgcgtggag	tcaatatctc	gccaccaacc	ggggaaactg	gtatttaaaa	73140
gtgtataaaa	acggagataa	aaaaaaaaag	tgtacttctt	tgacttctct	tgaacaaaa	73200
ttttctagct	agctgcaaat	acttggttta	taaactggaa	tccatgctct	ttaaatttct	73260
gtgatgaaga	ggaaaagcct	aaaaacggaa	taaaaaaaca	gtacacgtga	gatgacttgg	73320
aaaaatctta	aagactcatt	gaaaggactg	aattaagtgt	gaaaatattg	tgttttatat	73380
gtataaacca	tatatatata	acttatataa	caaagtaatg	cgtctgtgta	acttttcata	73440
agaaaagata	cccaacaat	aaaaacaatg	aacaatcgtc	gaaaaccta	tctagggata	73500
gccaataaaa	aagcctctaa	aatatcgaaa	ccagaacgtg	taagggtccc	aataacttcg	73560
cttaaaactc	aaaagtactc	aaatttatgt	gaggtttaaa	ttggttcaaa	ggtaaaacca	73620
gagatatctg	ggcgagaaaa	aaaaagacaa	gaagcacaca	accgcccatg	aggaatctga	73680

-continued

aggaccagac	cccaaaagtt	ttgttttgtc	cataccttgt	cattttcttc	ttccaaat	73740
ggaccaatta	ttattttaca	ttaaactcag	aaaagggttg	aaattttcct	ttatcttcac	73800
agttaatgat	ttagagtaaa	tgtgtaaagc	attcgaatat	aattactggt	ttcacaaaaa	73860
gaagaagaat	aatattttag	gtgttgtttt	tgcattgtatg	gatacaaaag	ctagcctcta	73920
tgcctaattg	tgttagtata	gagtgatcga	tactatagaa	gctgtacttt	ttcccgtaga	73980
taactaaaga	cgaggggcaa	agcgattctc	cacaagaaga	gtaattccat	catatgtata	74040
gcataacttt	tttgactttt	tgacctaaag	aaaaaaagca	tacttatttg	actaaataaa	74100
ttcgggtctg	tataaaggct	gaaatactgg	agaaaaagct	tagttttgtc	gtgtagtgtat	74160
caatgtatat	tgacaaaaga	gaagagactg	gaaaaaagtc	tacgaccaac	ttttccttcc	74220
ctttggcctt	aacttcactc	acacatcatt	ttttaatat	acccaaat	gtccattttc	74280
aaaccattaa	gacgcgatat	tgtgcatttt	gagattgtta	gcttgtagag	aattaaactt	74340
cgaatacctt	agttacaatt	ttatttcata	caaggtgtca	cagatccaca	ccataccgtt	74400
accttgttga	tatacaagtt	gaagttgtcc	ttagatatta	tgagcttgtt	tcacaaatca	74460
taaggctctaa	aaaaagattg	cgatatgagt	ttgttacaat	cattagaata	aaagaaaaag	74520
agagtaaaag	aaaataagat	aataaatgag	gaaaaatgaga	gttggagagt	ggagagacta	74580
gagaatacaa	attgcctttt	gcctttaaaa	gccggttcct	actctctttc	tccgaaataa	74640
atgtcgttag	cattattatc	gatcgtatct	atctatacac	acatgcgtac	caccacgtgt	74700
tttgtgctat	aatcatccta	taattcttat	taccaatctc	tagatttgtt	gttacttgtt	74760
ttataaaact	ttttgttttt	cggtattcatt	tggcacaacc	tatccaaaaa	aaaggtttaa	74820
atgagtatgt	tttaaaaaga	aactaatgtt	attgccaatc	tgagtactcc	gaagagatca	74880
agctaaaatt	gtaaggacca	aataacaaac	aatctttgca	aaccttttta	gctcaaaagg	74940
aaactactac	taaacacatt	ataagattga	acatggtgac	ataaatacaa	actactaatt	75000
acaactagca	gtttaaatat	gattcgtata	tcattgatac	taacaagatt	agaaaaaaa	75060
aacactatat	ataatgattg	attaacaagt	ttcttgtatc	ttagaaaaaa	gaggaaaaaa	75120
catattgatt	gttttttagt	ttatacaact	ttgtacacat	aattaacctg	caaagagagt	75180
atttcgacta	tgaaggaat	atgtaaatca	tgaaggaag	ctatttatga	aagatacaat	75240
aatataaacg	tgagttgttc	tcaaaaggca	ttagcctctc	atttcccatc	gtatttaa	75300
gggagtattg	caaggaaaac	atatatgttg	gtgtaaatga	ctttcttggg	tatgcagaag	75360
aatgaaatta	aacttcgaaa	atctccagac	ttacgggtta	gcttaagtgt	gatggcaa	75420
ccagccaagt	cgttgtgaaa	tctccctgaa	cattttataa	aacaaaataa	ctgtttat	75480
ttcacatatc	ttgaagctta	gtttgaaatt	tttgcattgt	attttggttt	actatgcaac	75540
aataaaacgg	gtttaatacg	tataatttgg	tgaactaac	tgttttttcc	ttacatggtc	75600
cactctcaac	tgttttaaag	agatgaaaag	ttatcatcat	tgattacgaa	aatgatattc	75660
aactctctct	atacgcccca	catccacaca	aatgccttaa	tgaagaggtt	atcgcgtttg	75720
cgattagaac	aaacccaaac	gttaccttca	atcataat	gcgatcttgt	tttggttagc	75780
taatagagtt	atagatcaac	tacgtcctag	tccaaaatct	gtggttataa	ctgtagttaa	75840
taatagtaaa	aataaaagct	tacagtaacc	ctcttttcat	cattgtaaga	ttcataacaa	75900
tgctctta	atgaatatac	aaatttcatt	ttcaagagat	atcgtaaagc	gaaaacatat	75960

-continued

tggtcaagaa	ctagacttcg	actggagaca	atggaggaac	atctactagc	ctttcaattt	76020
cttccaactg	gtcatatgga	gcaatctgca	gaataaccaa	acccaagact	cataaaattt	76080
cacatgtttt	ggtaatatct	ttgtttatct	agttttgatt	gtgatttcta	agttgtattt	76140
atttatacct	ggctaaccac	gtcggggcag	gttatcgaga	cagcgtagtt	tcccattggt	76200
cgtatgtcct	caggttcgat	atcttccgct	acatcgccat	atagaacttt	ttgctctcca	76260
gtccattcat	cctgtttaac	aaagatccaa	acattgacac	tttgaattct	tgttttaagt	76320
tttaaaagaa	gctttaagaa	gaaacagatt	ctcaggttga	atatctttca	taccacactt	76380
tggtgcagatc	tatcattttct	tctgacggtt	gcagggtgca	gtaagaactc	ttcgtctgag	76440
tttggtacct	tcactcta	tgccttgaga	tacttgctgt	atgtcacggc	cgttgatact	76500
gtttttaaca	taaaatgcac	agaaaagaaa	gagtccttta	acttaattca	tgcattgata	76560
ctaaaagaga	aaggaagggt	ttttgacatg	gttggttggt	taccttgctg	gcgtatcttg	76620
gcgcattggt	gcactacaca	tacaccaaga	tcctggaacg	ttctggcaac	gtcacttaga	76680
ggatccgaca	ctacttcagg	agttccgcta	tccccgaag	cagataacta	acaatcaaaa	76740
aggaatttaa	gaaaaactgt	ttagaacaat	gagtcagta	atattactgt	tcaatcacct	76800
gatataacaa	aagaattaga	tgaacattcc	aaggacatca	aaatagcttt	tcattcctaa	76860
agtcaaaagc	taaatattca	actgtgttgt	ttcatagaac	tcttacatca	tatataactt	76920
atatacatac	cgttggtcta	atggggaggt	caaagaggtg	aggtatgccg	aattgcttga	76980
ccacctgcag	ccaaagatgt	attgtcagaa	gttatctaaa	tgtgtttaac	gtacatgtcg	77040
ctgttgaaat	attcattttg	gtttcttgaa	gtaaaactaa	gagacagtat	aaagtcagct	77100
acctcagaac	ctgaaccttt	cccaaaagg	taataacggt	tcccatcagc	gtcaaagtgg	77160
cacatattct	ccacaacagc	aacgcaagc	acctactccg	taaaaccaa	caggaaataa	77220
atctcggaag	ttcatatctg	ttgttgtag	agttgagaag	aaaaacaagt	tcacaagctc	77280
catataaaag	aaaaaggaga	gtttatgatc	aaagtctgga	acagggaat	aaccttaagt	77340
tttgagaaca	tccttacacc	ttttgcaaca	tcaataaacg	ccaacttttg	aggggtggtg	77400
acaattaccg	ctgctgtcaa	tgggcgaacc	tgataccaag	aacattatag	tcatagcatc	77460
taacagaaga	atttaataat	attcatcata	tttcagttgc	tagcaaagtt	attcagttgt	77520
tacatgcaag	ggaaaaagca	ttggcacata	tatatgttga	ttactttcat	tcctacata	77580
tgcgaaagaa	ctaaaagtaa	agtagaaatg	gttggtataat	attcaaaacc	gaacctggca	77640
taaggtcagt	tgtatatcac	cagttccagg	aggcatgtcg	ataacaagat	agtcacgctc	77700
tccctgcgat	caaatccctc	aaaaagattt	ataaatgttt	ctggaaaaaa	catgaagtca	77760
atgattttat	ggaaatgaac	aaaccattca	gttggtgtaa	ggagttggtt	tataacacca	77820
gacaccatag	gacctctcat	aatggcacgc	cctgttcctg	caaatccaaa	tgagactagc	77880
ttgacgccca	tgtattctgt	tggaatgatg	gtcttcttct	ccgggttcta	ttgaacatat	77940
acagaactcg	aaaaacatta	gtcctttgtg	tggtactgct	atcacaaggt	gagtcagaga	78000
atattaccat	ttccaatata	cggctctcag	gattgaccat	ggttggtaga	cttgaccat	78060
agacatcagc	atcaaagata	ccaactctag	caccataacc	agctaagtga	taagcaagat	78120
ttacagctac	tgttgatttc	ccaacaccac	cctgaaatgt	tcacttaaca	acagcgtgtc	78180
atgaactaaa	agggacgcta	gaaagaaaga	tacagacaga	ttactttcta	tgtcacaaag	78240

-continued

aggaaatacc agctactgtt ccgtttctat cttttccaaa aaaacttgaa cgagacaata	78300
aacaaagaac acaggacatc ttattgatat cacatcacta agactacttt ggacagaagag	78360
tatgtagcat tcgtattttac acagcagagc acatcccata ttattaactt aaatttgaga	78420
agaaacttca gttcaaaaag aaacatcaag agctgtgtaa aaagaaattc attggaagat	78480
atatgtacct tgcaactaga aacagcagatg atgttcgaaa ttcttgataa tccaaaggga	78540
agctgccctg caaaaatggg cttggctggt tgtgctgaca ttgtcacatt taccttcttc	78600
acccatggaa gggctgcaac tacctcattt gccttggtct caaactacaa aagagtcgta	78660
atccatccat caaagataca ctaaaagaac tgattcattg cttatataac caatgatata	78720
gctccaaaat cctcctaatt ggatgaagtt ctatcagtag taatatgata aaagaatagt	78780
gaatggtaat gtcattttacc atgtctttga ctggacatgc ggggtgtgtc agctccaaac	78840
ggaacgaaac ctacaacaag agaccaaaca tgtttcagca gccttgaaca aacaatctaa	78900
aacagaacaa attggtgaag gataaagtat tattattacc tcacccaaag cttcattaat	78960
ccccaaatct ttcacaaaac cacaagaaac aatatctgtc ccaaaatcag gatcaataat	79020
ctgagacaga gccttcaaca catcttttct tgatgtttga gcaacactct caccaacact	79080
actactagct gttttccaaa atgcataagt taaaaaatat cagaacttgc ttagtctcct	79140
tcgcgtataa cgtaacaat ttttcaggcg gagacaaagg aacaaacttt ataacttga	79200
gctgaagcag ctttagctac agagagatgc tgagagaccg gtttgaggat tcttgctctg	79260
gagattgaga taatggaagc ttgagaatgg agaaacttgt gagaagaag caatcttgtt	79320
gtggaattgc ttctcttttg ggtctgaatc tcgaaagaag gatgccgcaa cgactgtgga	79380
tgaagaagcg gcattgctac aagagagaga cgaagagagt ggcaaaaaaa actagcggtc	79440
tgtgataatt tgatcgctca actcttttaa agataaaact ttttcaaagc tttttattat	79500
ttttattatt attattcaat atatcataaa tttttacatt tgtaatcaga tattattgtt	79560
ttttttactg atgtgtgtca aattacatct aagataaatc aaaattaagt ttttactgat	79620
gtatgttgaa ttctacaaaa gtttcttttg ttaacacaaa atgttaaatg tttcgaaatt	79680
aagataaatc aaaattaagt aaattaaaac ttaaaaagac aggaaagcag atttgattga	79740
agtaaatca gatgattaaa atgttagatt tagagaaatt ttcaaaaaaa tagactaata	79800
gtagatctag caattaatca agattgttat tgcaccgttc taaaactaca aatcttaatc	79860
tatagttaat cgactctcgt tataattgac taactatgcc taacaatatt ggttttgtgt	79920
ctccatctaa aattactaga attaagcaag cattagaatt cagtttgata agtttaccta	79980
agcatctaaa ccgggttggc agtcacttag ttattaaggc tctcctaaca ataattccag	80040
taataaacgc atctcgattt ttacatatta acttataatc aaattttagt tcgttactct	80100
agaactagct ttaagaataa tcaaaggaga agaactctag ggataaatac taacagatta	80160
tcgatttacg atttcatcta aattttctaat gataaaactt aaaccaaga aggagattac	80220
tcagacataa ttaaagaaac ataaaacatg tttgaataag attacataaa gaaatcaaaa	80280
gtagaatgga gtttaaaaaa tatcttttct ctataggtat aaggttttga aattttctgaa	80340
gtagtatcca caaacctcaa agaatttgga gagcaaaaaa tgaaataaag ctt	80393

<210> SEQ ID NO 5

<211> LENGTH: 368

<212> TYPE: DNA

-continued

<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 5

```

tgggacaccg ttacacacg aacgtttact ctaatagagc atgtatgtat gattgtctaa    60
ttccagtgta tctggtcacg cttgttactg cgcatagccg acctagcgga accccggatt    120
ttgaacccgt catcttatca agactgattc tgcgccgacc tttgcgactc cacggagcac    180
aattctatgg tgctattgca atatatgcct gcatacacgt ctgcataatgc tggctctcatc    240
gcgttttggg ggtcttctca tcagatacct atccagaggc tggctgcccc attatgggtac    300
acaatccttt ttgaatttca agctgcagcc cgggccgtcg accacgcgtg cccttagttg    360
agtcgtat                                     368

```

<210> SEQ ID NO 6

<211> LENGTH: 703

<212> TYPE: DNA

<213> ORGANISM: Hordeum vulgare

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (2)..(2)

<223> OTHER INFORMATION: n is a, c, g, or t

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (7)..(7)

<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 6

```

gnacgnaaa gtcctttgac tggtttgatg gttctcctat tgacgaactt ttatgcaagg    60
taagggagat atatggcctg gacgagaaaa ctagtttccg caacgtcact atctcgttgg    120
aagggaggcc tcaaccttta tatcttgga ctgctactca aattggagtg atatcaactg    180
aggggatccc cagtttacca aaaatgctac tccctccaaa ttgtgccggg cttccgtcaa    240
tgtatattag agatcttctt cttaatcctc catcttttga tgttgccctc gcaattcaag    300
aggcttgtag gcttatgtgc agcataactt gttcaattcc agaatttacc tgcataccat    360
cagcgaagct tgtgaaacta cttgagtcga aagaggttaa tcacatcgaa tttttagaaa    420
taaaaaatgt ccttgacgag attatgttga tgaatggaat cactgagctt tcagctatcc    480
agaacaaatt gctcgaacct gcttcggtgg ttactggctt gaaagttgat gctgatatac    540
taattaaaga atgtagatct atttcgaaac gtataggtga agtgatatct ttagctggcg    600
aaagtgacca ggcaatatct tcacgcgaat atattcccaa ggagttcttc aatgatatgg    660
agtcaccttg gaaggggccg tgtgaaaagg gtccatgctg aag                                     703

```

<210> SEQ ID NO 7

<211> LENGTH: 232

<212> TYPE: PRT

<213> ORGANISM: Hordeum vulgare

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (2)..(2)

<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid

<400> SEQUENCE: 7

```

Thr Xaa Lys Ser Phe Asp Trp Phe Asp Gly Ser Pro Ile Asp Glu Leu
1           5           10          15
Leu Cys Lys Val Arg Glu Ile Tyr Gly Leu Asp Glu Lys Thr Ser Phe
20          25          30

```

-continued

Arg Asn Val Thr Ile Ser Leu Glu Gly Arg Pro Gln Pro Leu Tyr Leu
 35 40 45
 Gly Thr Ala Thr Gln Ile Gly Val Ile Ser Thr Glu Gly Ile Pro Ser
 50 55 60
 Leu Pro Lys Met Leu Leu Pro Pro Asn Cys Ala Gly Leu Pro Ser Met
 65 70 75 80
 Tyr Ile Arg Asp Leu Leu Leu Asn Pro Pro Ser Phe Asp Val Ala Ser
 85 90 95
 Ala Ile Gln Glu Ala Cys Arg Leu Met Cys Ser Ile Thr Cys Ser Ile
 100 105 110
 Pro Glu Phe Thr Cys Ile Pro Ser Ala Lys Leu Val Lys Leu Leu Glu
 115 120 125
 Ser Lys Glu Val Asn His Ile Glu Phe Cys Arg Ile Lys Asn Val Leu
 130 135 140
 Asp Glu Ile Met Leu Met Asn Gly Ile Thr Glu Leu Ser Ala Ile Gln
 145 150 155 160
 Asn Lys Leu Leu Glu Pro Ala Ser Val Val Thr Gly Leu Lys Val Asp
 165 170 175
 Ala Asp Ile Leu Ile Lys Glu Cys Arg Phe Ile Ser Lys Arg Ile Gly
 180 185 190
 Glu Val Ile Ser Leu Ala Gly Glu Ser Asp Gln Ala Ile Ser Ser Ser
 195 200 205
 Glu Tyr Ile Pro Lys Glu Phe Phe Asn Asp Met Glu Ser Ser Trp Lys
 210 215 220
 Gly Pro Cys Glu Lys Gly Pro Cys
 225 230

<210> SEQ ID NO 8
 <211> LENGTH: 540
 <212> TYPE: DNA
 <213> ORGANISM: Hordeum vulgare

<400> SEQUENCE: 8

```

ctagtgtaaa tggcggttg gttgataggc ctgatggtct gggaaatggg ttggaacctc      60
caacaggttc ttttgactg ctgcgaaagg atgtcgagag cattgttact gcgatatgcg      120
aagacaagct gttggacctg tacaacaaga gaagcatctc agagcagatt gaggtggtct      180
gtgtaactgt aggtgctagg gagcaaccgc caccttcaac cgttggcagg tccagcatct      240
atatcattat cagacgtgac aacaagctct atgttggaaca gacggatgat ctcgtagggc      300
gtcttggtgc tcatagatcc aaggaaggta tgcaagatgc cacaatatta tacatcgtgg      360
ttcctggcaa gagcgttgcg tgccaactgg agactcttct cataaatcag ctaccctcga      420
aaggttttaa gctcaccaac aaggcagatg gcaagcatcg gaactttggt atgtctgtaa      480
cctctggaga agccatggcc gcgcactgaa ctgccccact gaacatccag ttttaactcg      540
  
```

<210> SEQ ID NO 9
 <211> LENGTH: 168
 <212> TYPE: PRT
 <213> ORGANISM: Hordeum vulgare

<400> SEQUENCE: 9

Ser Val Asn Gly Gly Leu Val Asp Arg Pro Asp Gly Leu Gly Asn Gly
 1 5 10 15

-continued

Leu Glu Pro Pro Thr Gly Ser Phe Gly Leu Leu Arg Lys Asp Val Glu
 20 25 30
 Ser Ile Val Thr Ala Ile Cys Glu Asp Lys Leu Leu Asp Leu Tyr Asn
 35 40 45
 Lys Arg Ser Ile Ser Glu Gln Ile Glu Val Val Cys Val Thr Val Gly
 50 55 60
 Ala Arg Glu Gln Pro Pro Pro Ser Thr Val Gly Arg Ser Ser Ile Tyr
 65 70 75 80
 Ile Ile Ile Arg Arg Asp Asn Lys Leu Tyr Val Gly Gln Thr Asp Asp
 85 90 95
 Leu Val Gly Arg Leu Gly Ala His Arg Ser Lys Glu Gly Met Gln Asp
 100 105 110
 Ala Thr Ile Leu Tyr Ile Val Val Pro Gly Lys Ser Val Ala Cys Gln
 115 120 125
 Leu Glu Thr Leu Leu Ile Asn Gln Leu Pro Ser Lys Gly Phe Lys Leu
 130 135 140
 Thr Asn Lys Ala Asp Gly Lys His Arg Asn Phe Gly Met Ser Val Thr
 145 150 155 160
 Ser Gly Glu Ala Met Ala Ala His
 165

<210> SEQ ID NO 10
 <211> LENGTH: 540
 <212> TYPE: DNA
 <213> ORGANISM: Hordeum vulgare

<400> SEQUENCE: 10

```

ctagtgtaaa tggcggtctg gttgataggc ctgatggtct gggaaatggg ttggaacctc      60
caacaggttc ttttgactg ctgcgaaagg atgtcgagag cattgttact gcgatatgcg      120
aagacaagct gttggacctg tacaacaaga gaagcatctc agagcagatt gaggtggtct      180
gtgtaactgt aggtgctagg gagcaaccgc caccttcaac cgttggcagg tccagcatct      240
atatcattat cagacgtgac aacaagctct atgttggaaca gacggatgat ctcgtagggc      300
gtcttggtgc tcatagatcc aaggaaggtg tgcaagatgc cacaatatta tacatcgtgg      360
ttcctggcaa gagcgttgcg tgccaactgg agactcttct cataaatcag ctaccctcga      420
aagggtttta gctcaccaac aaggcagatg gcaagcatcg gaactttggt atgtctgtaa      480
cctctggaga agccatggcc gcgcactgaa ctgccccact gaacatccag ttttaactcg      540
  
```

<210> SEQ ID NO 11
 <211> LENGTH: 444
 <212> TYPE: DNA
 <213> ORGANISM: Zea mays

<400> SEQUENCE: 11

```

taattacttc cttagacaag ggaaatatta taactccctt ggcccctact atgcacaagg      60
ctagaccac tatcagttca acaaaactag ggcggcatgg tgtcagttag ctccgcctc      120
ctattgaata tccaatagca aaaagacctt cagctgacta gttccgtcga gtagcaactg      180
cctcgccaga gattcgagat ataccgaagt tcctgtgctt cccgtctgcc ttgttgatga      240
gcttgaagcc cctcgaaggg agctggttta tgagaagggt ttccagctgg caggcaacgc      300
tcttgccagg gaccaagacg tataataccg tagcgtcccg catgccttcc ttcatctgt      360
  
```


-continued

```

gggcgttcaa gcgccccaga agatcgtccg tctgtccaac atagagcctg ttgtcgcttc 420
tgataatcac gtagatgcta gatc 444

```

```

<210> SEQ ID NO 12
<211> LENGTH: 94
<212> TYPE: PRT
<213> ORGANISM: Zea mays

```

```

<400> SEQUENCE: 12

```

```

Ser Ser Ile Tyr Val Ile Ile Arg Ser Asp Asn Arg Leu Tyr Val Gly
1          5          10          15
Gln Thr Asp Asp Leu Leu Gly Arg Leu Asn Ala His Arg Ser Lys Glu
20          25          30
Gly Met Arg Asp Ala Thr Val Leu Tyr Val Leu Val Pro Gly Lys Ser
35          40          45
Val Ala Cys Gln Leu Glu Thr Leu Leu Ile Asn Gln Leu Pro Ser Arg
50          55          60
Gly Phe Lys Leu Ile Asn Lys Ala Asp Gly Lys His Arg Asn Phe Gly
65          70          75          80
Ile Ser Arg Ile Ser Gly Glu Ala Val Ala Thr Arg Arg Asn
85          90

```

```

<210> SEQ ID NO 13
<211> LENGTH: 338
<212> TYPE: DNA
<213> ORGANISM: Medicago truncatula

```

```

<400> SEQUENCE: 13

```

```

caatggtaat aattctaatt ggacacatca ttccgaaaag tttttatcaa caattttctca 60
ggaggggaatc tcttttagcta atccaattga agtttcacat aaggagggtg agagtgcctat 120
cactgtaatc tgccaagatt ttatagcgga actgcgaagg aaaaagatca catcataact 180
tatcaagata aagtgtttct taattggcac tagggaatgg ccacctccga tgactatatg 240
ctcttcaagt gtctactgta tgctcagacc agatcagaaa ctctacgtag gagagacgga 300
taatctcgag gatcgagttc gtgcacatcg atcgaaag 338

```

```

<210> SEQ ID NO 14
<211> LENGTH: 679
<212> TYPE: DNA
<213> ORGANISM: Allium cepa

```

```

<400> SEQUENCE: 14

```

```

ggaatcttca tggaaggcc gtgtgaagag gatacatgct gaggatgtgt ttgctgaagt 60
tgacaaagct gctcagtcct tgtctattac agttatggaa gactttgttc caatcgtttc 120
tagagtaaaa gcggttatgt cttctcttgg aggtccaaag ggtgaagtat gttatgctag 180
agaacatgaa gctgttttgt tcaaaggaaa gcgttttatg ccactctgttt gggctaatac 240
acctggggaa gagcagatca agaaacttaa acctgccttg gattcaaaag gaagaaaagt 300
cggagaggaa tggttcacia cgatcaatat tgagaatgca ttaactaggt atcatgaatc 360
tacggaaaag gcaagaatta aagttttgga cttattaaga gaactttctg gagaaatgca 420
ggctaaaatt aacatccttg tcttctcttc catgctgctt gtcatactta aatctctttt 480
tggccatggt agtgaaggta ggagaagagg atgggtgttt cctgacctgc acaattccca 540

```

-continued

```

aatcataagg ataatagttt ggacactggg aatgaaacac ttgagctaag agatttatca 600
cctttatggg ttgatgctgt gcaaggaagt gcaatggaaa atactgtcag aatgcattct 660
atgtttcttt tactgggcc 679

```

```

<210> SEQ ID NO 15
<211> LENGTH: 179
<212> TYPE: PRT
<213> ORGANISM: Allium cepa

```

```

<400> SEQUENCE: 15

```

```

Glu Ser Ser Trp Lys Gly Arg Val Lys Arg Ile His Ala Glu Asp Val
1          5          10          15
Phe Ala Glu Val Asp Lys Ala Ala Gln Ser Leu Ser Ile Thr Val Met
20          25          30
Glu Asp Phe Val Pro Ile Val Ser Arg Val Lys Ala Val Met Ser Ser
35          40          45
Leu Gly Gly Pro Lys Gly Glu Val Cys Tyr Ala Arg Glu His Glu Ala
50          55          60
Val Trp Phe Lys Gly Lys Arg Phe Met Pro Ser Val Trp Ala Asn Thr
65          70          75          80
Pro Gly Glu Glu Gln Ile Lys Lys Leu Lys Pro Ala Leu Asp Ser Lys
85          90          95
Gly Arg Lys Val Gly Glu Glu Trp Phe Thr Thr Ile Asn Ile Glu Asn
100         105         110
Ala Leu Thr Arg Tyr His Glu Ser Thr Glu Lys Ala Arg Ile Lys Val
115         120         125
Leu Asp Leu Leu Arg Glu Leu Ser Gly Glu Met Gln Ala Lys Ile Asn
130         135         140
Ile Leu Val Phe Ser Ser Met Leu Leu Val Ile Ser Lys Ser Leu Phe
145         150         155         160
Gly His Val Ser Glu Gly Arg Arg Arg Gly Trp Val Phe Pro Asp Leu
165         170         175

His Asn Ser

```

```

<210> SEQ ID NO 16
<211> LENGTH: 662
<212> TYPE: DNA
<213> ORGANISM: Citrus sinensis

```

```

<400> SEQUENCE: 16

```

```

attggtttga tgcagcagaa ggcagtgtctg tacataatac agttgatatg cagtcattat 60
ttctcctgac tgggtccaaat ggggggtgta aatctagttt acttagatca atttgtgctg 120
cttcgttact tggcatatgt ggtcttatgg tgcccgaga gtcagcctca attccttact 180
ttgatgctat catgcttcac atgaaatcct atgatagccc tgtgacggg aaaagctcat 240
ttcaggtatt ctggttccct gtactgaggt tgtaagtttg ctcatgccat gatagatcga 300
gcttagccat gatcttgtga ggcattgtag tagtaactgg tgcaggtag aatgttgag 360
tactacaatt tacacattgc acttcacctc tcattctcaa tctggtggaa aagcgtaatg 420
tattaatttt ctgtggatat tatatgtctg cattctctta atttcagtat ttgctgcaa 480
aggttatctc cattaagttg cacatgttgc tcagtacett aagtttttac ttgaacaag 540
caattttttg tatgttgtaa ttatcttcga taggagtggg atcaagtaat atgcaataa 600

```

-continued

ttccgtttta atggttcagg tagaaatgtc agaaatacgg tcaattgtca ctgcaaccac 660

tt 662

<210> SEQ ID NO 17

<211> LENGTH: 81

<212> TYPE: PRT

<213> ORGANISM: Citrus sinensis

<400> SEQUENCE: 17

Trp Phe Asp Ala Ala Glu Gly Ser Ala Val His Asn Thr Val Asp Met
1 5 10 15

Gln Ser Leu Phe Leu Leu Thr Gly Pro Asn Gly Gly Gly Lys Ser Ser
20 25 30

Leu Leu Arg Ser Ile Cys Ala Ala Ser Leu Leu Gly Ile Cys Gly Leu
35 40 45

Met Val Pro Ala Glu Ser Ala Ser Ile Pro Tyr Phe Asp Ala Ile Met
50 55 60

Leu His Met Lys Ser Tyr Asp Ser Pro Ala Asp Gly Lys Ser Ser Phe
65 70 75 80

Gln

<210> SEQ ID NO 18

<211> LENGTH: 600

<212> TYPE: DNA

<213> ORGANISM: Solanum tuberosum

<400> SEQUENCE: 18

gcacacagac actgtgtatt gtgcactgat atcgagcaat gtattgggtt acggcaaaaa 60

acgtcgccgt ttcagttccc cgttggcggt cactgtccct tttcctccgt ccaccacttc 120

gccggcggtt cttctctttc tctccacata ctctgtgccg agagcagata cgttgcttga 180

aggagcgcaa gttttttgcc acaacggcaa aaaaaaactc aaacaaccaa aaagtgttcc 240

agaggaaaa gactatgtta atattatgtg gtggaagag agaatggaat tcttgagaaa 300

gccttcttct gttctactgg ctaagaggct tacatattgt aacttgctgg ggttgatcc 360

gagtttgaga aatggaagtc ttaaagaggg aacacttaac tcggagatgt tgcgttcaa 420

gtcaaaattt cctcgtgaag tttgttctg tagagtaggt gatttttatg aagcaattgg 480

attcgatgct tgtattcttg tggaatatgc tggtttaaat ccatttggtg gcctgcgctc 540

agatagtata ccaaaagctg gttgtccagt tgtgaatcta agacagacgt tggatgatct 600

<210> SEQ ID NO 19

<211> LENGTH: 187

<212> TYPE: PRT

<213> ORGANISM: Solanum tuberosum

<400> SEQUENCE: 19

Met Tyr Trp Val Thr Ala Lys Asn Val Ala Val Ser Val Pro Arg Trp
1 5 10 15

Arg Ser Leu Ser Leu Phe Leu Arg Pro Pro Leu Arg Arg Arg Phe Phe
20 25 30

Ser Phe Ser Pro His Thr Leu Cys Arg Glu Gln Ile Arg Cys Leu Lys
35 40 45

Glu Arg Lys Phe Phe Ala Thr Thr Ala Lys Lys Lys Leu Lys Gln Pro

-continued

50					55					60					
Lys	Ser	Val	Pro	Glu	Glu	Lys	Asp	Tyr	Val	Asn	Ile	Met	Trp	Trp	Lys
65				70					75					80	
Glu	Arg	Met	Glu	Phe	Leu	Arg	Lys	Pro	Ser	Ser	Val	Leu	Leu	Ala	Lys
			85					90					95		
Arg	Leu	Thr	Tyr	Cys	Asn	Leu	Leu	Gly	Val	Asp	Pro	Ser	Leu	Arg	Asn
			100					105					110		
Gly	Ser	Leu	Lys	Glu	Gly	Thr	Leu	Asn	Ser	Glu	Met	Leu	Leu	Phe	Lys
		115					120					125			
Ser	Lys	Phe	Pro	Arg	Glu	Val	Leu	Phe	Cys	Arg	Val	Gly	Asp	Phe	Tyr
	130					135					140				
Glu	Ala	Ile	Gly	Phe	Asp	Ala	Cys	Ile	Leu	Val	Glu	Tyr	Ala	Gly	Leu
145				150					155					160	
Asn	Pro	Phe	Gly	Gly	Leu	Arg	Ser	Asp	Ser	Ile	Pro	Lys	Ala	Gly	Cys
			165					170					175		
Pro	Val	Val	Asn	Leu	Arg	Gln	Thr	Leu	Asp	Asp					
			180					185							

<210> SEQ ID NO 20

<211> LENGTH: 3396

<212> TYPE: DNA

<213> ORGANISM: Oryza sativa

<400> SEQUENCE: 20

```

atggccattc agcggctgct cgcgagctcg ctctgggccg ccacgccgcg gtggcttccc      60
gtcgccgcgc actcgtttct ccggcgccgc caccgccctc gctgctcccc gctccccgcg      120
ctgctattta acaggaggtc ctggtctaaa ccaaggaaag tctcacgaag catttccatt      180
gtgtctagga agatgaacaa acaaggagat ctctgtaatg aaggcatgct gccacatatt      240
ctgtggtgga aagagaaaat ggagagggtc aggaaccat catcaatgca attgactcag      300
agacttggtg attcaaatat tttaggattg gatccaactt taagaaatgg aagcttgaag      360
gatggaagcc tgaacacgga aatgttgcaa ttcaaatcga agtttcctcg tgaagttcta      420
ctttgcagag tgggagattt ctacgaggct gttgggtttg atgcatgtat cctgtggag      480
catgcaggct taaatccttt tggaggcttg cgttctgata gtattccaaa agctggatgt      540
ccagtcatga atttcgggca gacattggat gatttgactc gatgtggtta ctctgtgtgc      600
atagttgaag aaattcaagg cccaaccaa gctcgtgcta ggaaaggccg atttatttct      660
ggccatgcac atcctggtag tccttatgta tttggtcttg ctgaagtaga ccatgatgtt      720
gagttccctg atccaatgcc ttagttggg atttcacgat ctgcaaaagg ctattgcctg      780
atttctgtgc tagagacaat gaaaacatat tcagctgagg agggcttaac agaggaagca      840
gttgttacta agcttcgcat atgccgttat catcatctat accttcatag ttctttgagg      900
aacaattctt caggcacatc acgctgggga gaatttgccg aaggtgggct attgtgggga      960
gagtgcagtg gaaaatcttt tgagtggttt gatggtaatc ctattgaaga actgttatgc     1020
aaggtaaggg aaatatatgg gcttgaagag aagactgttt tccgtaatgt cagtgtctca     1080
ttggaaggga ggccccaacc ctgttatctt ggaacagcta ctcaaattgg ggtgatacca     1140
actgagggaa taccagttt gctaaaaatt gttctccctc caaacttttg tggccttcca     1200
tcattgtata ttagagatct tcttcttaac cctccatctt ttgatgttgc atcatcagtt     1260

```

-continued

caagaggcctt gcaggcttat gggtagcata acttgctcga ttcctgaatt tacatgcata	1320
ccggcagcaa agccttgtaa attactcgag tcaaaagagg ttaatcacat cgaattttgt	1380
agaataaaga atgtcctcga tgagggtgtg ttcattggga gcaatgctga gctttctgct	1440
atcctgaata aattgcttga tcctgccgcc atagtactg ggttcaaagt tgaagccgat	1500
atactagtga atgaatgtag ctttatttca caacgtatag ctgaagtaat ctctttaggt	1560
ggtgaaagt accaggcaat aacttcactt gaatatattc cgaagaggtt cttcaatggt	1620
atggagtcatt cttggaagg acgtgtaaaa aggggtgatg ctgaagagga gttctcaa	1680
gttgatatag ctgctgaggc actgtcaaca gcggtcattg aagattttct gccaatatt	1740
tcaagagtaa aatctgtgat gtccctcaat ggaagttcga agggagaaat cagttatgca	1800
aaagagcatg aatctgtttg gtttaagggt aggcgattca caccaaatgt gtgggccaac	1860
actcctgggt aactacagat aaagcaattg aagcctgcaa ttgactcaa aggtagaaag	1920
gtcggagaag aatggttcac cactatcaaa gttgagaatg ctttaaccag gtaccatgaa	1980
gcttgtgata atgcaaaacg taaagtctct gagttgttga gaggactttc aagtgaattg	2040
caggacaaga ttaatgtcct tgtcttttgc tcaacgatgc tcatcataac aaaagcactt	2100
tttggtcatg ttagtgaagg acgaagaagg ggttggtgct ttcctactat atctcccttg	2160
tgtaaggata atgttacaga ggaaatctca agtgaaatgg aattgtcagg aacttttcct	2220
tactggcttg atactaacca agggaatgca atactgaatg atgtccatat gcactctttg	2280
tttattctta ctgggtccaa cgggtggtgt aaatccagta tgctgagatc agtctgtgct	2340
gtgcattac ttggaatatg tggcctgatg gtgccagctg cttcagctgt catccacat	2400
ttcgattcca tcatgctgca tatgaaagca tatgatagcc cagctgatgg taaaagtctg	2460
tttcagattg aaatgtcaga gatacgatct ttagtctgcc gagctacagc taggagtctt	2520
gttctaattg atgaaatatg taggggcaca gaaacagcaa aaggaacatg tatagctggt	2580
agcatcattg aaagactcga taatgttggc tgcattagca tcatatcaac tcatttgcat	2640
ggcatttttg accttccact gtcactccac aatactgatt tcaaagctat gggaaccgaa	2700
atcatcgata ggtgcattca gccaacatgg aaattaatgg atggcatctg tagagagagt	2760
cttgcttttc aaacagccag gaaagaagg atgcctgact tgataattag aagagctgag	2820
gaactatatt tggctatgag cacaacagc aagcatacat catcagctgt ccaccatgaa	2880
atatccatag ccaactctac tgtaaatagc ttggttgaga agcctaatta cctgagaaat	2940
ggactagagc ttcaatctgg ttccttcgga ttactaagaa aagaaattga gagtgttgtt	3000
accacaatat gcaagaagaa actgttggt ctctacaaca aaaggagcat ctcagaactg	3060
attgaggtgg tctgtgttgc tgtgggtgct agggagcaac cccaccttc aactgttggc	3120
aggtccagca ttatgtaat tatcagacgt gacagcaagc tctatattgg acagacggat	3180
gatcttggtg gtgcacttag tgctcacaga tcgaaggaag gtatgcagga tgccacgata	3240
ttatatattt tggtagctgg gaagagcatt goatgccaac tggaaactct tctcataaat	3300
cagctacctt tgaaagggtt caagctcatc aacaaggcag atggcaagca tcgaaatttc	3360
ggtatatctc ttgtcccagg agaggcaatt gccgca	3396

<210> SEQ ID NO 21

<211> LENGTH: 3396

<212> TYPE: DNA

-continued

<213> ORGANISM: *Oryza sativa*

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (1)..(3396)

<400> SEQUENCE: 21

atg gcc att cag cgg ctg ctc gcg agc tcg ctc gtg gcc gcc acg ccg	48
Met Ala Ile Gln Arg Leu Leu Ala Ser Ser Leu Val Ala Ala Thr Pro	
1 5 10 15	
cgg tgg ctt ccc gtc gcc gcc gac tcg ttt ctc cgg cgc cgc cac cgc	96
Arg Trp Leu Pro Val Ala Ala Asp Ser Phe Leu Arg Arg Arg His Arg	
20 25 30	
cct cgc tgc tcc ccg ctc ccc gcg ctg cta ttt aac agg agg tcc tgg	144
Pro Arg Cys Ser Pro Leu Pro Ala Leu Leu Phe Asn Arg Arg Ser Trp	
35 40 45	
tct aaa cca agg aaa gtc tca cga agc att tcc att gtg tct agg aag	192
Ser Lys Pro Arg Lys Val Ser Arg Ser Ile Ser Ile Val Ser Arg Lys	
50 55 60	
atg aac aaa caa gga gat ctc tgt aat gaa ggc atg ctg cca cat att	240
Met Asn Lys Gln Gly Asp Leu Cys Asn Glu Gly Met Leu Pro His Ile	
65 70 75 80	
ctg tgg tgg aaa gag aaa atg gag agg tgc agg aaa cca tca tca atg	288
Leu Trp Trp Lys Glu Lys Met Glu Arg Cys Arg Lys Pro Ser Ser Met	
85 90 95	
caa ttg act cag aga ctt gtg tat tca aat att tta gga ttg gat cca	336
Gln Leu Thr Gln Arg Leu Val Tyr Ser Asn Ile Leu Gly Leu Asp Pro	
100 105 110	
act tta aga aat gga agc ttg aag gat gga agc ctg aac acg gaa atg	384
Thr Leu Arg Asn Gly Ser Leu Lys Asp Gly Ser Leu Asn Thr Glu Met	
115 120 125	
ttg caa ttc aaa tcg aag ttt cct cgt gaa gtt cta ctt tgc aga gtg	432
Leu Phe Lys Ser Lys Phe Pro Arg Glu Val Leu Leu Cys Arg Val	
130 135 140	
gga gat ttc tac gag gct gtt ggg ttt gat gca tgt atc ctt gtg gag	480
Gly Asp Phe Tyr Glu Ala Val Gly Phe Asp Ala Cys Ile Leu Val Glu	
145 150 155 160	
cat gca ggc tta aat cct ttt gga ggc ttg cgt tct gat agt att cca	528
His Ala Gly Leu Asn Pro Phe Gly Gly Leu Arg Ser Asp Ser Ile Pro	
165 170 175	
aaa gct gga tgt cca gtc atg aat ttg cgg cag aca ttg gat gat ttg	576
Lys Ala Gly Cys Pro Val Met Asn Leu Arg Gln Thr Leu Asp Asp Leu	
180 185 190	
act cga tgt ggt tac tct gtg tgc ata gtt gaa gaa att caa ggc cca	624
Thr Arg Cys Gly Tyr Ser Val Cys Ile Val Glu Glu Ile Gln Gly Pro	
195 200 205	
acc caa gct cgt gct agg aaa ggc cga ttt att tct ggc cat gca cat	672
Thr Gln Ala Arg Ala Arg Lys Gly Arg Phe Ile Ser Gly His Ala His	
210 215 220	
cct ggt agt cct tat gta ttt ggt ctt gct gaa gta gac cat gat gtt	720
Pro Gly Ser Pro Tyr Val Phe Gly Leu Ala Glu Val Asp His Asp Val	
225 230 235 240	
gag ttc cct gat cca atg cct gta gtt ggg att tca cga tct gca aaa	768
Glu Phe Pro Asp Pro Met Pro Val Val Gly Ile Ser Arg Ser Ala Lys	
245 250 255	
ggc tat tgc ctg att tct gtg cta gag aca atg aaa aca tat tca gct	816
Gly Tyr Cys Leu Ile Ser Val Leu Glu Thr Met Lys Thr Tyr Ser Ala	
260 265 270	
gag gag ggc tta aca gag gaa gca gtt gtt act aag ctt cgc ata tgc	864

-continued

Glu	Glu	Gly	Leu	Thr	Glu	Glu	Ala	Val	Val	Thr	Lys	Leu	Arg	Ile	Cys	
		275					280					285				
cgt	tat	cat	cat	cta	tac	ctt	cat	agt	tct	ttg	agg	aac	aat	tct	tca	912
Arg	Tyr	His	His	Leu	Tyr	Leu	His	Ser	Ser	Leu	Arg	Asn	Asn	Ser	Ser	
	290					295				300						
ggc	aca	tca	cgc	tgg	gga	gaa	ttt	ggc	gaa	ggg	ggg	cta	ttg	tgg	gga	960
Gly	Thr	Ser	Arg	Trp	Gly	Glu	Phe	Gly	Glu	Gly	Gly	Leu	Leu	Trp	Gly	
305					310				315						320	
gag	tgc	agt	gga	aaa	tct	ttt	gag	tgg	ttt	gat	ggg	aat	cct	att	gaa	1008
Glu	Cys	Ser	Gly	Lys	Ser	Phe	Glu	Trp	Phe	Asp	Gly	Asn	Pro	Ile	Glu	
				325					330					335		
gaa	ctg	tta	tgc	aag	gta	agg	gaa	ata	tat	ggg	ctt	gaa	gag	aag	act	1056
Glu	Leu	Leu	Cys	Lys	Val	Arg	Glu	Ile	Tyr	Gly	Leu	Glu	Glu	Lys	Thr	
				340				345						350		
gtt	ttc	cgt	aat	gtc	agt	gtc	tca	ttg	gaa	ggg	agg	cct	caa	ccc	ttg	1104
Val	Phe	Arg	Asn	Val	Ser	Val	Ser	Leu	Glu	Gly	Arg	Pro	Gln	Pro	Leu	
		355					360					365				
tat	ctt	gga	aca	gct	act	caa	att	ggg	gtg	ata	cca	act	gag	gga	ata	1152
Tyr	Leu	Gly	Thr	Ala	Thr	Gln	Ile	Gly	Val	Ile	Pro	Thr	Glu	Gly	Ile	
	370					375					380					
ccc	agt	ttg	cta	aaa	att	gtt	ctc	cct	cca	aac	ttt	ggg	ggc	ctt	cca	1200
Pro	Ser	Leu	Leu	Lys	Ile	Val	Leu	Pro	Pro	Asn	Phe	Gly	Gly	Leu	Pro	
385					390					395					400	
tca	ttg	tat	att	aga	gat	ctt	ctt	ctt	aac	cct	cca	tct	ttt	gat	gtt	1248
Ser	Leu	Tyr	Ile	Arg	Asp	Leu	Leu	Leu	Asn	Pro	Pro	Ser	Phe	Asp	Val	
				405					410					415		
gca	tca	tca	gtt	caa	gag	gct	tgc	agg	ctt	atg	ggg	agc	ata	act	tgc	1296
Ala	Ser	Ser	Val	Gln	Glu	Ala	Cys	Arg	Leu	Met	Gly	Ser	Ile	Thr	Cys	
			420					425					430			
tcg	att	cct	gaa	ttt	aca	tgc	ata	ccg	gca	gca	aag	ctt	gtg	aaa	tta	1344
Ser	Ile	Pro	Glu	Phe	Thr	Cys	Ile	Pro	Ala	Ala	Lys	Leu	Val	Lys	Leu	
		435					440					445				
ctc	gag	tca	aaa	gag	gtt	aat	cac	atc	gaa	ttt	tgt	aga	ata	aag	aat	1392
Leu	Glu	Ser	Lys	Glu	Val	Asn	His	Ile	Glu	Phe	Cys	Arg	Ile	Lys	Asn	
		450				455					460					
gtc	ctc	gat	gag	gtg	ttg	ttc	atg	ggg	agc	aat	gct	gag	ctt	tct	gct	1440
Val	Leu	Asp	Glu	Val	Leu	Phe	Met	Gly	Ser	Asn	Ala	Glu	Leu	Ser	Ala	
465					470				475						480	
atc	ctg	aat	aaa	ttg	ctt	gat	cct	gcc	gcc	ata	gtt	act	ggg	ttc	aaa	1488
Ile	Leu	Asn	Lys	Leu	Leu	Asp	Pro	Ala	Ala	Ile	Val	Thr	Gly	Phe	Lys	
				485				490						495		
gtt	gaa	gcc	gat	ata	cta	gtg	aat	gaa	tgt	agc	ttt	att	tca	caa	cgt	1536
Val	Glu	Ala	Asp	Ile	Leu	Val	Asn	Glu	Cys	Ser	Phe	Ile	Ser	Gln	Arg	
			500					505					510			
ata	gct	gaa	gta	atc	tct	tta	ggg	ggg	gaa	agt	gac	cag	gca	ata	act	1584
Ile	Ala	Glu	Val	Ile	Ser	Leu	Gly	Gly	Glu	Ser	Asp	Gln	Ala	Ile	Thr	
		515					520					525				
tca	tct	gaa	tat	att	ccg	aaa	gag	ttc	ttc	aat	ggg	atg	gag	tca	tct	1632
Ser	Ser	Glu	Tyr	Ile	Pro	Lys	Glu	Phe	Phe	Asn	Gly	Met	Glu	Ser	Ser	
		530				535					540					
tgg	aag	gga	cgt	gta	aaa	agg	gtg	cat	gct	gaa	gag	gag	ttc	tca	aat	1680
Trp	Lys	Gly	Arg	Val	Lys	Arg	Val	His	Ala	Glu	Glu	Glu	Phe	Ser	Asn	
545					550					555					560	
gtt	gat	ata	gct	gct	gag	gca	ctg	tca	aca	gcg	gtc	att	gaa	gat	ttt	1728
Val	Asp	Ile	Ala	Ala	Glu	Ala	Leu	Ser	Thr	Ala	Val	Ile	Glu	Asp	Phe	
				565					570					575		
ctg	cca	att	att	tca	aga	gta	aaa	tct	gtg	atg	tcc	tca	aat	gga	agt	1776

-continued

Leu	Pro	Ile	Ile	Ser	Arg	Val	Lys	Ser	Val	Met	Ser	Ser	Asn	Gly	Ser	
		580						585					590			
tcg	aag	gga	gaa	atc	agt	tat	gca	aaa	gag	cat	gaa	tct	gtt	tgg	ttt	1824
Ser	Lys	Gly	Glu	Ile	Ser	Tyr	Ala	Lys	Glu	His	Glu	Ser	Val	Trp	Phe	
		595					600					605				
aaa	ggg	agg	cga	ttc	aca	cca	aat	gtg	tgg	gcc	aac	act	cct	ggt	gaa	1872
Lys	Gly	Arg	Arg	Phe	Thr	Pro	Asn	Val	Trp	Ala	Asn	Thr	Pro	Gly	Glu	
		610				615					620					
cta	cag	ata	aag	caa	ttg	aag	cct	gca	att	gac	tca	aaa	ggt	aga	aag	1920
Leu	Gln	Ile	Lys	Gln	Leu	Lys	Pro	Ala	Ile	Asp	Ser	Lys	Gly	Arg	Lys	
		625			630					635					640	
gtc	gga	gaa	gaa	tgg	ttc	acc	act	atc	aaa	gtt	gag	aat	gct	tta	acc	1968
Val	Gly	Glu	Glu	Trp	Phe	Thr	Thr	Ile	Lys	Val	Glu	Asn	Ala	Leu	Thr	
				645					650					655		
agg	tac	cat	gaa	gct	tgt	gat	aat	gca	aaa	cgt	aaa	gtt	ctt	gag	ttg	2016
Arg	Tyr	His	Glu	Ala	Cys	Asp	Asn	Ala	Lys	Arg	Lys	Val	Leu	Glu	Leu	
			660					665					670			
ttg	aga	gga	ctt	tca	agt	gaa	ttg	cag	gac	aag	att	aat	gtc	ctt	gtc	2064
Leu	Arg	Gly	Leu	Ser	Ser	Glu	Leu	Gln	Asp	Lys	Ile	Asn	Val	Leu	Val	
		675					680					685				
ttt	tgc	tca	acg	atg	ctc	atc	ata	aca	aaa	gca	ctt	ttt	ggt	cat	gtt	2112
Phe	Cys	Ser	Thr	Met	Leu	Ile	Ile	Thr	Lys	Ala	Leu	Phe	Gly	His	Val	
		690				695					700					
agt	gaa	gga	cga	aga	agg	ggt	tgg	gtg	ctt	cct	act	ata	tct	ccc	ttg	2160
Ser	Glu	Gly	Arg	Arg	Gly	Trp	Val	Leu	Pro	Thr	Ile	Ser	Pro	Leu		
		705			710				715					720		
tgt	aag	gat	aat	gtt	aca	gag	gaa	atc	tca	agt	gaa	atg	gaa	ttg	tca	2208
Cys	Lys	Asp	Asn	Val	Thr	Glu	Glu	Ile	Ser	Ser	Glu	Met	Glu	Leu	Ser	
				725					730					735		
gga	act	ttt	cct	tac	tgg	ctt	gat	act	aac	caa	ggg	aat	gca	ata	ctg	2256
Gly	Thr	Phe	Pro	Tyr	Trp	Leu	Asp	Thr	Asn	Gln	Gly	Asn	Ala	Ile	Leu	
			740					745					750			
aat	gat	gtc	cat	atg	cac	tct	ttg	ttt	att	ctt	act	ggt	cca	aac	ggt	2304
Asn	Asp	Val	His	Met	His	Ser	Leu	Phe	Ile	Leu	Thr	Gly	Pro	Asn	Gly	
		755					760					765				
ggt	ggt	aaa	tcc	agt	atg	ctg	aga	tca	gtc	tgt	gct	gct	gca	tta	ctt	2352
Gly	Gly	Lys	Ser	Ser	Met	Leu	Arg	Ser	Val	Cys	Ala	Ala	Ala	Leu	Leu	
		770				775				780						
gga	ata	tgt	ggc	ctg	atg	gtg	cca	gct	gct	tca	gct	gtc	atc	cca	cat	2400
Gly	Ile	Cys	Gly	Leu	Met	Val	Pro	Ala	Ala	Ser	Ala	Val	Ile	Pro	His	
		785			790					795				800		
ttc	gat	tcc	atc	atg	ctg	cat	atg	aaa	gca	tat	gat	agc	cca	gct	gat	2448
Phe	Asp	Ser	Ile	Met	Leu	His	Met	Lys	Ala	Tyr	Asp	Ser	Pro	Ala	Asp	
			805						810					815		
ggt	aaa	agt	tcg	ttt	cag	att	gaa	atg	tca	gag	ata	cga	tct	tta	gtc	2496
Gly	Lys	Ser	Ser	Phe	Gln	Ile	Glu	Met	Ser	Glu	Ile	Arg	Ser	Leu	Val	
		820					825						830			
tgc	cga	gct	aca	gct	agg	agt	ctt	gtt	cta	att	gat	gaa	ata	tgt	agg	2544
Cys	Arg	Ala	Thr	Ala	Arg	Ser	Leu	Val	Leu	Ile	Asp	Glu	Ile	Cys	Arg	
		835					840					845				
ggc	aca	gaa	aca	gca	aaa	gga	aca	tgt	ata	gct	ggt	agc	atc	att	gaa	2592
Gly	Thr	Glu	Thr	Ala	Lys	Gly	Thr	Cys	Ile	Ala	Gly	Ser	Ile	Ile	Glu	
		850				855					860					
aga	ctc	gat	aat	gtt	ggc	tgc	ata	ggc	atc	ata	tca	act	cat	ttg	cat	2640
Arg	Leu	Asp	Asn	Val	Gly	Cys	Ile	Gly	Ile	Ile	Ser	Thr	His	Leu	His	
		865			870				875					880		
ggc	att	ttt	gac	ctt	cca	ctg	tca	ctc	cac	aat	act	gat	ttc	aaa	gct	2688

-continued

Gly Ile Phe Asp Leu Pro Leu Ser Leu His Asn Thr Asp Phe Lys Ala	
885 890 895	
atg gga acc gaa atc atc gat agg tgc att cag cca aca tgg aaa tta	2736
Met Gly Thr Glu Ile Ile Asp Arg Cys Ile Gln Pro Thr Trp Lys Leu	
900 905 910	
atg gat ggc atc tgt aga gag agt ctt gct ttt caa aca gcc agg aaa	2784
Met Asp Gly Ile Cys Arg Glu Ser Leu Ala Phe Gln Thr Ala Arg Lys	
915 920 925	
gaa ggt atg cct gac ttg ata att aga aga gct gag gaa cta tat ttg	2832
Glu Gly Met Pro Asp Leu Ile Ile Arg Arg Ala Glu Glu Leu Tyr Leu	
930 935 940	
gct atg agc aca aac agc aag cat aca tca tca gct gtc cac cat gaa	2880
Ala Met Ser Thr Asn Ser Lys His Thr Ser Ser Ala Val His His Glu	
945 950 955 960	
ata tcc ata gcc aac tct act gta aat agc ttg gtt gag aag cct aat	2928
Ile Ser Ile Ala Asn Ser Thr Val Asn Ser Leu Val Glu Lys Pro Asn	
965 970 975	
tac ctg aga aat gga cta gag ctt caa tct ggt tcc ttc gga tta cta	2976
Tyr Leu Arg Asn Gly Leu Glu Leu Gln Ser Gly Ser Phe Gly Leu Leu	
980 985 990	
aga aaa gaa att gag agt gtt gtt acc aca ata tgc aag aag aaa ctg	3024
Arg Lys Glu Ile Glu Ser Val Val Thr Thr Ile Cys Lys Lys Lys Leu	
995 1000 1005	
ttg gat ctc tac aac aaa agg agc atc tca gaa ctg att gag gtg	3069
Leu Asp Leu Tyr Asn Lys Arg Ser Ile Ser Glu Leu Ile Glu Val	
1010 1015 1020	
gtc tgt gtt gct gtg ggt gct agg gag caa ccc cca cct tca act	3114
Val Cys Val Ala Val Gly Ala Arg Glu Gln Pro Pro Pro Ser Thr	
1025 1030 1035	
gtt ggc agg tcc agc att tat gta att atc aga cgt gac agc aag	3159
Val Gly Arg Ser Ser Ile Tyr Val Ile Ile Arg Arg Asp Ser Lys	
1040 1045 1050	
ctc tat att gga cag acg gat gat ctt gtg ggt cga ctt agt gct	3204
Leu Tyr Ile Gly Gln Thr Asp Asp Leu Val Gly Arg Leu Ser Ala	
1055 1060 1065	
cac aga tcg aag gaa ggt atg cag gat gcc acg ata tta tat att	3249
His Arg Ser Lys Glu Gly Met Gln Asp Ala Thr Ile Leu Tyr Ile	
1070 1075 1080	
ttg gta cct ggg aag agc att gca tgc caa ctg gaa act ctt ctc	3294
Leu Val Pro Gly Lys Ser Ile Ala Cys Gln Leu Glu Thr Leu Leu	
1085 1090 1095	
ata aat cag cta cct ttg aaa ggt ttc aag ctc atc aac aag gca	3339
Ile Asn Gln Leu Pro Leu Lys Gly Phe Lys Leu Ile Asn Lys Ala	
1100 1105 1110	
gat ggc aag cat cga aat ttc ggt ata tct ctt gtc cca gga gag	3384
Asp Gly Lys His Arg Asn Phe Gly Ile Ser Leu Val Pro Gly Glu	
1115 1120 1125	
gca att gcc gca	3396
Ala Ile Ala Ala	
1130	

<210> SEQ ID NO 22

<211> LENGTH: 1132

<212> TYPE: PRT

<213> ORGANISM: Oryza sativa

<400> SEQUENCE: 22

Met Ala Ile Gln Arg Leu Leu Ala Ser Ser Leu Val Ala Ala Thr Pro

-continued

1	5	10	15
Arg Trp Leu	Pro Val Ala Ala	Asp Ser Phe Leu	Arg Arg Arg His Arg
	20	25	30
Pro Arg Cys	Ser Pro Leu Pro	Ala Leu Leu Phe	Asn Arg Arg Ser Trp
	35	40	45
Ser Lys Pro	Arg Lys Val Ser	Arg Ser Ile Ser	Ile Val Ser Arg Lys
	50	55	60
Met Asn Lys	Gln Gly Asp Leu Cys	Asn Glu Gly Met	Leu Pro His Ile
	65	70	75
Leu Trp Trp	Lys Glu Lys Met	Glu Arg Cys Arg	Lys Pro Ser Ser Met
	85	90	95
Gln Leu Thr	Gln Arg Leu Val	Tyr Ser Asn Ile	Leu Gly Leu Asp Pro
	100	105	110
Thr Leu Arg	Asn Gly Ser Leu	Lys Asp Gly Ser	Leu Asn Thr Glu Met
	115	120	125
Leu Gln Phe	Lys Ser Lys Phe	Pro Arg Glu Val	Leu Leu Cys Arg Val
	130	135	140
Gly Asp Phe	Tyr Glu Ala Val	Gly Phe Asp Ala	Cys Ile Leu Val Glu
	145	150	155
His Ala Gly	Leu Asn Pro Phe	Gly Gly Leu Arg	Ser Asp Ser Ile Pro
	165	170	175
Lys Ala Gly	Cys Pro Val Met	Asn Leu Arg Gln	Thr Leu Asp Asp Leu
	180	185	190
Thr Arg Cys	Gly Tyr Ser Val	Cys Ile Val Glu	Glu Ile Gln Gly Pro
	195	200	205
Thr Gln Ala	Arg Ala Arg Lys	Gly Arg Phe Ile	Ser Gly His Ala His
	210	215	220
Pro Gly Ser	Pro Tyr Val Phe	Gly Leu Ala Glu	Val Asp His Asp Val
	225	230	235
Glu Phe Pro	Asp Pro Met Pro	Val Val Gly Ile	Ser Arg Ser Ala Lys
	245	250	255
Gly Tyr Cys	Leu Ile Ser Val	Leu Glu Thr Met	Lys Thr Tyr Ser Ala
	260	265	270
Glu Glu Gly	Leu Thr Glu Glu	Ala Val Val Thr	Lys Leu Arg Ile Cys
	275	280	285
Arg Tyr His	His Leu Tyr Leu	His Ser Ser Leu	Arg Asn Asn Ser Ser
	290	295	300
Gly Thr Ser	Arg Trp Gly Glu	Phe Gly Glu Gly	Leu Leu Trp Gly
	305	310	315
Glu Cys Ser	Gly Lys Ser Phe	Glu Trp Phe Asp	Gly Asn Pro Ile Glu
	325	330	335
Glu Leu Leu	Cys Lys Val Arg	Glu Ile Tyr Gly	Leu Glu Glu Lys Thr
	340	345	350
Val Phe Arg	Asn Val Ser Val	Ser Leu Glu Gly	Arg Pro Gln Pro Leu
	355	360	365
Tyr Leu Gly	Thr Ala Thr Gln	Ile Gly Val Ile	Pro Thr Glu Gly Ile
	370	375	380
Pro Ser Leu	Leu Lys Ile Val	Leu Pro Pro Asn	Phe Gly Gly Leu Pro
	385	390	395
Ser Leu Tyr	Ile Arg Asp Leu	Leu Leu Asn Pro	Pro Ser Phe Asp Val
	405	410	415

-continued

Ala	Ser	Ser	Val	Gln	Glu	Ala	Cys	Arg	Leu	Met	Gly	Ser	Ile	Thr	Cys
			420					425					430		
Ser	Ile	Pro	Glu	Phe	Thr	Cys	Ile	Pro	Ala	Ala	Lys	Leu	Val	Lys	Leu
		435				440						445			
Leu	Glu	Ser	Lys	Glu	Val	Asn	His	Ile	Glu	Phe	Cys	Arg	Ile	Lys	Asn
	450					455					460				
Val	Leu	Asp	Glu	Val	Leu	Phe	Met	Gly	Ser	Asn	Ala	Glu	Leu	Ser	Ala
465				470						475					480
Ile	Leu	Asn	Lys	Leu	Leu	Asp	Pro	Ala	Ala	Ile	Val	Thr	Gly	Phe	Lys
			485					490						495	
Val	Glu	Ala	Asp	Ile	Leu	Val	Asn	Glu	Cys	Ser	Phe	Ile	Ser	Gln	Arg
		500						505					510		
Ile	Ala	Glu	Val	Ile	Ser	Leu	Gly	Gly	Glu	Ser	Asp	Gln	Ala	Ile	Thr
	515					520						525			
Ser	Ser	Glu	Tyr	Ile	Pro	Lys	Glu	Phe	Phe	Asn	Gly	Met	Glu	Ser	Ser
	530					535					540				
Trp	Lys	Gly	Arg	Val	Lys	Arg	Val	His	Ala	Glu	Glu	Glu	Phe	Ser	Asn
545					550					555					560
Val	Asp	Ile	Ala	Ala	Glu	Ala	Leu	Ser	Thr	Ala	Val	Ile	Glu	Asp	Phe
			565						570					575	
Leu	Pro	Ile	Ile	Ser	Arg	Val	Lys	Ser	Val	Met	Ser	Ser	Asn	Gly	Ser
		580						585					590		
Ser	Lys	Gly	Glu	Ile	Ser	Tyr	Ala	Lys	Glu	His	Glu	Ser	Val	Trp	Phe
	595						600					605			
Lys	Gly	Arg	Arg	Phe	Thr	Pro	Asn	Val	Trp	Ala	Asn	Thr	Pro	Gly	Glu
	610					615					620				
Leu	Gln	Ile	Lys	Gln	Leu	Lys	Pro	Ala	Ile	Asp	Ser	Lys	Gly	Arg	Lys
625				630						635					640
Val	Gly	Glu	Glu	Trp	Phe	Thr	Thr	Ile	Lys	Val	Glu	Asn	Ala	Leu	Thr
			645					650						655	
Arg	Tyr	His	Glu	Ala	Cys	Asp	Asn	Ala	Lys	Arg	Lys	Val	Leu	Glu	Leu
		660						665					670		
Leu	Arg	Gly	Leu	Ser	Ser	Glu	Leu	Gln	Asp	Lys	Ile	Asn	Val	Leu	Val
	675						680					685			
Phe	Cys	Ser	Thr	Met	Leu	Ile	Ile	Thr	Lys	Ala	Leu	Phe	Gly	His	Val
	690					695					700				
Ser	Glu	Gly	Arg	Arg	Arg	Gly	Trp	Val	Leu	Pro	Thr	Ile	Ser	Pro	Leu
705				710						715					720
Cys	Lys	Asp	Asn	Val	Thr	Glu	Glu	Ile	Ser	Ser	Glu	Met	Glu	Leu	Ser
			725						730					735	
Gly	Thr	Phe	Pro	Tyr	Trp	Leu	Asp	Thr	Asn	Gln	Gly	Asn	Ala	Ile	Leu
		740					745						750		
Asn	Asp	Val	His	Met	His	Ser	Leu	Phe	Ile	Leu	Thr	Gly	Pro	Asn	Gly
	755						760					765			
Gly	Gly	Lys	Ser	Ser	Met	Leu	Arg	Ser	Val	Cys	Ala	Ala	Ala	Leu	Leu
	770					775					780				
Gly	Ile	Cys	Gly	Leu	Met	Val	Pro	Ala	Ala	Ser	Ala	Val	Ile	Pro	His
785				790						795					800
Phe	Asp	Ser	Ile	Met	Leu	His	Met	Lys	Ala	Tyr	Asp	Ser	Pro	Ala	Asp
			805						810					815	

-continued

Gly Lys Ser Ser Phe Gln Ile Glu Met Ser Glu Ile Arg Ser Leu Val
 820 825 830
 Cys Arg Ala Thr Ala Arg Ser Leu Val Leu Ile Asp Glu Ile Cys Arg
 835 840 845
 Gly Thr Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly Ser Ile Ile Glu
 850 855 860
 Arg Leu Asp Asn Val Gly Cys Ile Gly Ile Ile Ser Thr His Leu His
 865 870 875 880
 Gly Ile Phe Asp Leu Pro Leu Ser Leu His Asn Thr Asp Phe Lys Ala
 885 890 895
 Met Gly Thr Glu Ile Ile Asp Arg Cys Ile Gln Pro Thr Trp Lys Leu
 900 905 910
 Met Asp Gly Ile Cys Arg Glu Ser Leu Ala Phe Gln Thr Ala Arg Lys
 915 920 925
 Glu Gly Met Pro Asp Leu Ile Ile Arg Arg Ala Glu Glu Leu Tyr Leu
 930 935 940
 Ala Met Ser Thr Asn Ser Lys His Thr Ser Ser Ala Val His His Glu
 945 950 955 960
 Ile Ser Ile Ala Asn Ser Thr Val Asn Ser Leu Val Glu Lys Pro Asn
 965 970 975
 Tyr Leu Arg Asn Gly Leu Glu Leu Gln Ser Gly Ser Phe Gly Leu Leu
 980 985 990
 Arg Lys Glu Ile Glu Ser Val Val Thr Thr Ile Cys Lys Lys Lys Leu
 995 1000 1005
 Leu Asp Leu Tyr Asn Lys Arg Ser Ile Ser Glu Leu Ile Glu Val
 1010 1015 1020
 Val Cys Val Ala Val Gly Ala Arg Glu Gln Pro Pro Pro Ser Thr
 1025 1030 1035
 Val Gly Arg Ser Ser Ile Tyr Val Ile Ile Arg Arg Asp Ser Lys
 1040 1045 1050
 Leu Tyr Ile Gly Gln Thr Asp Asp Leu Val Gly Arg Leu Ser Ala
 1055 1060 1065
 His Arg Ser Lys Glu Gly Met Gln Asp Ala Thr Ile Leu Tyr Ile
 1070 1075 1080
 Leu Val Pro Gly Lys Ser Ile Ala Cys Gln Leu Glu Thr Leu Leu
 1085 1090 1095
 Ile Asn Gln Leu Pro Leu Lys Gly Phe Lys Leu Ile Asn Lys Ala
 1100 1105 1110
 Asp Gly Lys His Arg Asn Phe Gly Ile Ser Leu Val Pro Gly Glu
 1115 1120 1125
 Ala Ile Ala Ala
 1130

<210> SEQ ID NO 23

<211> LENGTH: 433

<212> TYPE: DNA

<213> ORGANISM: Sorghum bicolor

<400> SEQUENCE: 23

```

aaggaaggca tgcaggatgc tacgatatta tacatcttgg ttcctggcaa gagcgttgcc      60
tgccagctgg aaacccttct cataaatcag cttccttcga ggggcttcaa gtcacatcaac    120
aaggcagacg gaaagcatag gaacttcggt atatctcgaa tctctggaga ggcaatcgcc    180

```

-continued

```

accagctaa actaatcagc taaagatcta atttagttag tcttgacgct agtgagtctc 240
atattgcac cttcatctct ttgtcttttg gctactcaat aggaggcagg aactaactga 300
caccatagtc cgccccaatt ttgtgagatg aattatcagt ggtgctaccc ttgtgcatag 360
taggggccta gggggcgatc ttcccttgtc taagcatgta gtacgggtgca aatgattagc 420
aatgcaatga cac 433

```

```

<210> SEQ ID NO 24
<211> LENGTH: 64
<212> TYPE: PRT
<213> ORGANISM: Sorghum bicolor

```

```

<400> SEQUENCE: 24

```

```

Lys Glu Gly Met Gln Asp Ala Thr Ile Leu Tyr Ile Leu Val Pro Gly
1      5      10      15
Lys Ser Val Ala Cys Gln Leu Glu Thr Leu Leu Ile Asn Gln Leu Pro
20     25     30
Ser Arg Gly Phe Lys Leu Ile Asn Lys Ala Asp Gly Lys His Arg Asn
35     40     45
Phe Gly Ile Ser Arg Ile Ser Gly Glu Ala Ile Ala Thr Gln Leu Asn
50     55     60

```

```

<210> SEQ ID NO 25
<211> LENGTH: 667
<212> TYPE: DNA
<213> ORGANISM: Sorghum bicolor

```

```

<400> SEQUENCE: 25

```

```

tggtaaatct actatgttgc gatcagctcg tgcagcttcg ctgcttggaa tatgtggcct 60
gatggtacct tcaacttcag ctgtaatccc gcattttgat tccattatgc tgcatatgaa 120
agcctacgat agccagccg atgggaaaag ttcatttcag attgaaatgt cggagatacg 180
tgcttttagtc agccgagcta ctgctaggag tcttgtcctg attggtgaaa tatgtagggg 240
cacagaaaact gcaaaaggaa cctgtattgc tggtagcatc atcgaaaggc tggataatgt 300
tggtgccta ggcacatat caactcacct gcattgggatt tttgacttgc ctctctcaact 360
cagcactact gatttcaaag ctatgggaac tgaagtggtc gacgggtgca ttcaccaac 420
atggaaactg atggatggca tctgtagaga aagccttgct tttcaaacag ccaggaggga 480
aggcatgcct gagttcataa tcagaagggc tgaggagcta tatttgacta tgagtacaaa 540
taacaagcag accgcatcaa tgggtccaaa tgagcctcgt aatgacagcc ccagtgtaaa 600
tggtctgggtt gagaagcctg aatatctgaa atacaggcta gaaattctgc ctggtacctt 660
tgagccg 667

```

```

<210> SEQ ID NO 26
<211> LENGTH: 222
<212> TYPE: PRT
<213> ORGANISM: Sorghum bicolor

```

```

<400> SEQUENCE: 26

```

```

Gly Lys Ser Thr Met Leu Arg Ser Val Cys Ala Ala Ser Leu Leu Gly
1      5      10      15
Ile Cys Gly Leu Met Val Pro Ser Thr Ser Ala Val Ile Pro His Phe
20     25     30

```

-continued

Asp Ser Ile Met Leu His Met Lys Ala Tyr Asp Ser Pro Ala Asp Gly
 35 40 45
 Lys Ser Ser Phe Gln Ile Glu Met Ser Glu Ile Arg Ala Leu Val Ser
 50 55 60
 Arg Ala Thr Ala Arg Ser Leu Val Leu Ile Gly Glu Ile Cys Arg Gly
 65 70 75 80
 Thr Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly Ser Ile Ile Glu Arg
 85 90 95
 Leu Asp Asn Val Gly Cys Leu Gly Ile Ile Ser Thr His Leu His Gly
 100 105 110
 Ile Phe Asp Leu Pro Leu Ser Leu Ser Thr Thr Asp Phe Lys Ala Met
 115 120 125
 Gly Thr Glu Val Val Asp Gly Cys Ile His Pro Thr Trp Lys Leu Met
 130 135 140
 Asp Gly Ile Cys Arg Glu Ser Leu Ala Phe Gln Thr Ala Arg Arg Glu
 145 150 155 160
 Gly Met Pro Glu Phe Ile Ile Arg Arg Ala Glu Glu Leu Tyr Leu Thr
 165 170 175
 Met Ser Thr Asn Asn Lys Gln Thr Ala Ser Met Val His Asn Glu Pro
 180 185 190
 Arg Asn Asp Ser Pro Ser Val Asn Gly Leu Val Glu Lys Pro Glu Tyr
 195 200 205
 Leu Lys Tyr Arg Leu Glu Ile Leu Pro Gly Thr Phe Glu Pro
 210 215 220

<210> SEQ ID NO 27
 <211> LENGTH: 351
 <212> TYPE: DNA
 <213> ORGANISM: Glycine max
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (89)..(91)
 <223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 27

```

ggaaatattt tgttacaatc ttgttacagc aaggaacaca aaaatttaaat agtgtgatct      60
ttgacatgtc ttccatataa agtcagtcnn ncttttgac caagttaggc ccaaattttt      120
tcatacaaga aatagaaaag aatgagaaag tacaaccac aagaattccg cctcaaggat      180
gtatgcaaaa ataagtaatg atattggcaa gtacgaagct tcgtaacaac tgcttcttct      240
gtcaagcaat cttcagaaga atatgtcttc atggtctcta gtaccatatt aatgcaataa      300
ccctcgcag aatgagatat tcctactaca ggcattgggt ctgcctcgtg c                351
  
```

<210> SEQ ID NO 28
 <211> LENGTH: 406
 <212> TYPE: DNA
 <213> ORGANISM: Glycine max

<400> SEQUENCE: 28

```

ggaattcggc acgaggctga gctcaatgaa atattgaaac atttaatcga gccacatgg      60
gtggcaactg ggtagaaat tgactttgaa accttggttg caggatgtga gatcgcatct      120
agtaagattg gtgaaatagt atctctggat gatgagaatg atcagaaaat caactcgttc      180
tcttttattc ctcacgaatt ttttgaggat atggagtcta aatggaaagg tcgaataaaa      240
  
```

-continued

agaatccaca tagatgatgt attcactgca gtggaaaaag cagctgaggc cttacatata	300
gcagtcactg aagattttgt tcctgtagtt gctagaataa aggcatttgt agcccctctc	360
ggaggtccta acggagaaat atcttatgct cgggagcaag aagcag	406

<210> SEQ ID NO 29

<211> LENGTH: 3393

<212> TYPE: DNA

<213> ORGANISM: Glycine max

<400> SEQUENCE: 29

atgtacaggg tagccacaag aaacgtcgcc gttttcttcc ctctgtgtgtg ttccctcgcg	60
cactacactc cttctctatt tcccatttcc acttcattcg ctccctctcg ttcccttaga	120
ataaatggat gtgtaaaaga tgtgtcgagt tatacggata agaaggtttc aagggggagt	180
agtagggcca ccaagaagcc caaaatacca aataacgttt tagatgataa agaccttcct	240
cacatactgt ggtggaagga gaggttgcaa atgtgcagaa agttttcaac tgtccagtta	300
attgaaagac ttgaattttc taatttgctt ggcctgaatt ccaacttgaa aaatggaagt	360
ctgaaggaag gaacactcaa ctgggaaatg ttgcaattca agtcaaaatt tccacgtcaa	420
gtattgcttt gcagagtgg ggaattctat gaagcttggg gaatagatgc ttgtattctt	480
gttgaatatg tgggtttaaa tcccattggt ggtctgcat cagatagtat cccaagagct	540
agttgtcctg tcgtgaatct tcggcagact ttagatgac tgacaacaaa tggttattca	600
gtgtgcattg tggaggaggc tcagggccca agtcaagctc gatccaggaa acgtcgcttt	660
atatctgggc atgctcatcc tggaaatccc tatgtatatg gacttgctac agttgatcat	720
gatcttaact ttccagaacc aatgcctgta gtaggaaat ctctattctgc gaggggttat	780
tgcatataa tgggtactaga gaccatgaag acatattctt ctgaagattg cttgacagaa	840
gaagcagttg ttacgaagct tcgtacttgc caatatcatt acttattttt gcatacatcc	900
ttgaggcgga attcttgttg aacctgcaac tggggagaat ttggtgaggg agggctatta	960
tggggagaat gtagttctag acattttgat tggtttgatg gcaaccctgt ctccgatctt	1020
ttggccaagg taaaggaact ttatagtatt gatgatgagg ttacctttcg gaacacaact	1080
gtgtcttcag gacatagggc tcgaccatta actcttgaa catctactca aattggtgcc	1140
attccaacag aaggaatacc ttctttgttg aaggttttac ttccatcaaa ttgcaatgga	1200
ttaccagtat tgtacataag ggaacttctt ttgaatctc cttcatatga gattgcatcc	1260
aaaaattcaag caacatgcaa acttatgagc agtgtaacgt gttcaattcc agaatttaca	1320
tgtgtttcgt cagcaaaagt tgtaaagcta cttgaatgga gggagggtcaa tcatatggaa	1380
ttttgtagaa taaagaatgt actggatgaa attttgacaga tgtatagtac ctctgagctc	1440
aatgaaatat tgaaacattt aatcgagccc acatgggttg caactgggtt agaaattgac	1500
tttgaacact tggttgcagg atgtgagatc gcacttagta agattggtga aatagtatct	1560
ctggatgatg agaatgatca gaaaatcaac tcgttctctt ttattcctca cgaatttttt	1620
gaggatatgg agtctaaatg gaaaggctga ataaaaagaa tccacataga tgatgtattc	1680
actgcagtgg aaaaagcagc tgaggcctta catatagcag tcaactgaaga ttttgttcct	1740
gttgtttcta gaataaaggc tattgtagcc cctctcgag gtcctaaggg agaaatatct	1800
tatgctcggg agcaagaagc agtttggttc aaaggcaaac gctttacacc gaatttgggtg	1860

-continued

```

gctggtagcc ctggagagga acaaattaaa cagcttaggc atgctttaga ttctaaaggt 1920
agaaaggtag gggaggaatg gtttaccaca ccaaaggctg aggctgcatt aacaaggtag 1980
catgaagcaa atgccaaggc aaaagaaaga gttttggaaa ttttaagggg actcgtgtgt 2040
gagttgcaat acagtataaa cattcttgtc ttttcttcca tgttgcttgt tattgcaaaa 2100
gctttatttg ctcatgcaag tgaagggaga agaaggagat gggcttttcc cacgcttgta 2160
gaatcccatg ggtttgagga tgtgaagtca ttggacaaaa cccatgggat gaagataagt 2220
ggtttattgc catattggtt ccacatagca gaagggtgtg tgcgtaatga tgttgatatg 2280
caatcattat ttctgttgac aggaccgaat ggtgggtggga aatcaagttt tcttaggtca 2340
atttgtgctg ctgcactact tgggatagt ggactcatgg ttctgcaga atcagcccta 2400
attccttatt ttgactccat cacgcttcat atgaagtc atgatatgccc agctgataaa 2460
aagagttcct ttcaggttga aatgtcagaa cttcgatcca tcattggcgg aacaaccaac 2520
aggagccttg tacttgttga tgaaatatgc cgaggaacag aaactgcaa agggacttgc 2580
attgctggta gcatcattga aacccttgat ggaattgggt gtctgggtat tgtatccact 2640
cacttgcatt gaatatattac ttgtccccta aacaaaaaaaa acactgtgca caaagcaatg 2700
ggcacaacat ccattgatgg acaaataatg cctacatgga agttgacaga tggagtttgt 2760
aaagaaagtc ttgcttttga aacggctaag agggaaggaa ttcttgagca tattgtttaga 2820
agagctgaat atctttatca gttggtttat gctaaggaaa tgctttttgc agaaaatttc 2880
ccaaatgaag aaaagtttcc tacctgcac aatgttaata atttgaatgg aacacatctt 2940
cattcaaaaa ggttcctatc aggagcta atcaatggaag ttttacgcga ggaagttgag 3000
agagctgtca ctgtgatttg ccaggatcat ataaaggacc taaaatgcaa aaagattgca 3060
ttggagctta ctgagataaa atgtctcata attggtacaa gggagctacc acctccatcg 3120
gttgtaggtt cttaagcgt ctatgtgatg ttcagaccag ataagaaact ctatgtagga 3180
gagactgatg atctcgaggg acgggtccga agacatcgat taaaggaagg aatgcattgat 3240
gcatcattcc tttattttct tgtcccagg aaaagccttg catgccaatt tgaatctctg 3300
ctcatcaacc aactttcttg tcaaggcttc caactgagca atatagctga tggtaaacad 3360
aggaattttg gcacttccaa cctgtatata taa 3393

```

```

<210> SEQ ID NO 30
<211> LENGTH: 3393
<212> TYPE: DNA
<213> ORGANISM: Glycine max
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(3393)

```

```

<400> SEQUENCE: 30

```

```

atg tac agg gta gcc aca aga aac gtc gcc gtt ttc ttc cct cgt tgc 48
Met Tyr Arg Val Ala Thr Arg Asn Val Ala Val Phe Phe Pro Arg Cys
1 5 10 15

tgt tcc ctg gcg cac tac act cct tct cta ttt ccc att ttc act tca 96
Cys Ser Leu Ala His Tyr Thr Pro Ser Leu Phe Pro Ile Phe Thr Ser
20 25 30

ttc gct ccc tct cgt ttc ctt aga ata aat gga tgt gta aag aat gtg 144
Phe Ala Pro Ser Arg Phe Leu Arg Ile Asn Gly Cys Val Lys Asn Val
35 40 45

```


-continued

tcg agt tat acg gat aag aag gtt tca agg ggg agt agt agg gcc acc	192
Ser Ser Tyr Thr Asp Lys Lys Val Ser Arg Gly Ser Ser Arg Ala Thr	
50 55 60	
aag aag ccc aaa ata cca aat aac gtt tta gat gat aaa gac ctt cct	240
Lys Lys Pro Lys Ile Pro Asn Asn Val Leu Asp Asp Lys Asp Leu Pro	
65 70 75 80	
cac ata ctg tgg tgg aag gag agg ttg caa atg tgc aga aag ttt tca	288
His Ile Leu Trp Trp Lys Glu Arg Leu Gln Met Cys Arg Lys Phe Ser	
85 90 95	
act gtc cag tta att gaa aga ctt gaa ttt tct aat ttg ctt ggc ctg	336
Thr Val Gln Leu Ile Glu Arg Leu Glu Phe Ser Asn Leu Leu Gly Leu	
100 105 110	
aat tcc aac ttg aaa aat gga agt ctg aag gaa gga aca ctc aac tgg	384
Asn Ser Asn Leu Lys Asn Gly Ser Leu Lys Glu Gly Thr Leu Asn Trp	
115 120 125	
gaa atg ttg caa ttc aag tca aaa ttt cca cgt caa gta ttg ctt tgc	432
Glu Met Leu Gln Phe Lys Ser Lys Phe Pro Arg Gln Val Leu Leu Cys	
130 135 140	
aga gtt ggg gaa ttc tat gaa gct tgg gga ata gat gct tgt att ctt	480
Arg Val Gly Glu Phe Tyr Glu Ala Trp Gly Ile Asp Ala Cys Ile Leu	
145 150 155 160	
gtt gaa tat gtg ggt tta aat ccc att ggt ggt ctg cga tca gat agt	528
Val Glu Tyr Val Gly Leu Asn Pro Ile Gly Gly Leu Arg Ser Asp Ser	
165 170 175	
atc cca aga gct agt tgt cct gtc gtg aat ctt cgg cag act tta gat	576
Ile Pro Arg Ala Ser Cys Pro Val Val Asn Leu Arg Gln Thr Leu Asp	
180 185 190	
gat ctg aca aca aat ggt tat tca gtg tgc att gtg gag gag gct cag	624
Asp Leu Thr Thr Asn Gly Tyr Ser Val Cys Ile Val Glu Glu Ala Gln	
195 200 205	
ggc cca agt caa gct cga tcc agg aaa cgt cgc ttt ata tct ggg cat	672
Gly Pro Ser Gln Ala Arg Ser Arg Lys Arg Arg Phe Ile Ser Gly His	
210 215 220	
gct cat cct gga aat ccc tat gta tat gga ctt gct aca gtt gat cat	720
Ala His Pro Gly Asn Pro Tyr Val Tyr Gly Leu Ala Thr Val Asp His	
225 230 235 240	
gat ctt aac ttt cca gaa cca atg cct gta gta gga ata tct cat tct	768
Asp Leu Asn Phe Pro Glu Pro Met Pro Val Val Gly Ile Ser His Ser	
245 250 255	
gcg agg ggt tat tgc att aat atg gta cta gag acc atg aag aca tat	816
Ala Arg Gly Tyr Cys Ile Asn Met Val Leu Glu Thr Met Lys Thr Tyr	
260 265 270	
tct tct gaa gat tgc ttg aca gaa gaa gca gtt gtt acg aag ctt cgt	864
Ser Ser Glu Asp Cys Leu Thr Glu Glu Ala Val Val Thr Lys Leu Arg	
275 280 285	
act tgc caa tat cat tac tta ttt ttg cat aca tcc ttg agg cgg aat	912
Thr Cys Gln Tyr His Tyr Leu Phe Leu His Thr Ser Leu Arg Arg Asn	
290 295 300	
tct tgt gga acc tgc aac tgg gga gaa ttt ggt gag gga ggg cta tta	960
Ser Cys Gly Thr Cys Asn Trp Gly Glu Phe Gly Glu Gly Gly Leu Leu	
305 310 315 320	
tgg gga gaa tgt agt tct aga cat ttt gat tgg ttt gat ggc aac cct	1008
Trp Gly Glu Cys Ser Ser Arg His Phe Asp Trp Phe Asp Gly Asn Pro	
325 330 335	
gtc tcc gat ctt ttg gcc aag gta aag gaa ctt tat agt att gat gat	1056
Val Ser Asp Leu Leu Ala Lys Val Lys Glu Leu Tyr Ser Ile Asp Asp	
340 345 350	

-continued

gag gtt acc ttt cgg aac aca act gtg tct tca gga cat agg gct cga	1104
Glu Val Thr Phe Arg Asn Thr Thr Val Ser Ser Gly His Arg Ala Arg	
355 360 365	
cca tta act ctt gga aca tct act caa att ggt gcc att cca aca gaa	1152
Pro Leu Thr Leu Gly Thr Ser Thr Gln Ile Gly Ala Ile Pro Thr Glu	
370 375 380	
gga ata cct tct ttg ttg aag gtt tta ctt cca tca aat tgc aat gga	1200
Gly Ile Pro Ser Leu Leu Lys Val Leu Leu Pro Ser Asn Cys Asn Gly	
385 390 395 400	
tta cca gta ttg tac ata agg gaa ctt ctt ttg aat cct cct tca tat	1248
Leu Pro Val Leu Tyr Ile Arg Glu Leu Leu Leu Asn Pro Pro Ser Tyr	
405 410 415	
gag att gca tcc aaa att caa gca aca tgc aaa ctt atg agc agt gta	1296
Glu Ile Ala Ser Lys Ile Gln Ala Thr Cys Lys Leu Met Ser Ser Val	
420 425 430	
acg tgt tca att cca gaa ttt aca tgt gtt tcg tca gca aag ctt gta	1344
Thr Cys Ser Ile Pro Glu Phe Thr Cys Val Ser Ser Ala Lys Leu Val	
435 440 445	
aag cta ctt gaa tgg agg gag gtc aat cat atg gaa ttt tgt aga ata	1392
Lys Leu Leu Glu Trp Arg Glu Val Asn His Met Glu Phe Cys Arg Ile	
450 455 460	
aag aat gta ctg gat gaa att ttg cag atg tat agt acc tct gag ctc	1440
Lys Asn Val Leu Asp Glu Ile Leu Gln Met Tyr Ser Thr Ser Glu Leu	
465 470 475 480	
aat gaa ata ttg aaa cat tta atc gag ccc aca tgg gtg gca act ggg	1488
Asn Glu Ile Leu Lys His Leu Ile Glu Pro Thr Trp Val Ala Thr Gly	
485 490 495	
tta gaa att gac ttt gaa acc ttg gtt gca gga tgt gag atc gca tct	1536
Leu Glu Ile Asp Phe Glu Thr Leu Val Ala Gly Cys Glu Ile Ala Ser	
500 505 510	
agt aag att ggt gaa ata gta tct ctg gat gat gag aat gat cag aaa	1584
Ser Lys Ile Gly Glu Ile Val Ser Leu Asp Asp Glu Asn Asp Gln Lys	
515 520 525	
atc aac tcg ttc tct ttt att cct cac gaa ttt ttt gag gat atg gag	1632
Ile Asn Ser Phe Ser Phe Ile Pro His Glu Phe Phe Glu Asp Met Glu	
530 535 540	
tct aaa tgg aaa ggt cga ata aaa aga atc cac ata gat gat gta ttc	1680
Ser Lys Trp Lys Gly Arg Ile Lys Arg Ile His Ile Asp Asp Val Phe	
545 550 555 560	
act gca gtg gaa aaa gca gct gag gcc tta cat ata gca gtc act gaa	1728
Thr Ala Val Glu Lys Ala Ala Glu Ala Leu His Ile Ala Val Thr Glu	
565 570 575	
gat ttt gtt cct gtt gtt tct aga ata aag gct att gta gcc cct ctc	1776
Asp Phe Val Pro Val Val Ser Arg Ile Lys Ala Ile Val Ala Pro Leu	
580 585 590	
gga ggt cct aag gga gaa ata tct tat gct cgg gag caa gaa gca gtt	1824
Gly Gly Pro Lys Gly Glu Ile Ser Tyr Ala Arg Glu Gln Glu Ala Val	
595 600 605	
tgg ttc aaa ggc aaa cgc ttt aca ccg aat ttg tgg gct ggt agc cct	1872
Trp Phe Lys Gly Lys Arg Phe Thr Pro Asn Leu Trp Ala Gly Ser Pro	
610 615 620	
gga gag gaa caa att aaa cag ctt agg cat gct tta gat tct aaa ggt	1920
Gly Glu Glu Gln Ile Lys Gln Leu Arg His Ala Leu Asp Ser Lys Gly	
625 630 635 640	
aga aag gta ggg gag gaa tgg ttt acc aca cca aag gtc gag gct gca	1968
Arg Lys Val Gly Glu Glu Trp Phe Thr Pro Lys Val Glu Ala Ala	
645 650 655	

-continued

tta aca agg tac cat gaa gca aat gcc aag gca aaa gaa aga gtt ttg	2016
Leu Thr Arg Tyr His Glu Ala Asn Ala Lys Ala Lys Glu Arg Val Leu	
660 665 670	
gaa att tta agg gga ctc gct gct gag ttg caa tac agt ata aac att	2064
Glu Ile Leu Arg Gly Leu Ala Glu Leu Gln Tyr Ser Ile Asn Ile	
675 680 685	
ctt gtc ttt tct tcc atg ttg ctt gtt att gcc aaa gct tta ttt gct	2112
Leu Val Phe Ser Ser Met Leu Leu Val Ile Ala Lys Ala Leu Phe Ala	
690 695 700	
cat gca agt gaa ggg aga aga agg aga tgg gtc ttt ccc acg ctt gta	2160
His Ala Ser Glu Gly Arg Arg Arg Arg Trp Val Phe Pro Thr Leu Val	
705 710 715 720	
gaa tcc cat ggg ttt gag gat gtg aag tca ttg gac aaa acc cat ggg	2208
Glu Ser His Gly Phe Glu Asp Val Lys Ser Leu Asp Lys Thr His Gly	
725 730 735	
atg aag ata agt ggt tta ttg cca tat tgg ttc cac ata gca gaa ggt	2256
Met Lys Ile Ser Gly Leu Leu Pro Tyr Trp Phe His Ile Ala Glu Gly	
740 745 750	
gtt gtg cgt aat gat gtt gat atg caa tca tta ttt ctg ttg aca gga	2304
Val Val Arg Asn Asp Val Asp Met Gln Ser Leu Phe Leu Leu Thr Gly	
755 760 765	
ccg aat ggt ggt ggg aaa tca agt ttt ctt agg tca att tgt gct gct	2352
Pro Asn Gly Gly Gly Lys Ser Ser Phe Leu Arg Ser Ile Cys Ala Ala	
770 775 780	
gca cta ctt ggg ata tgt gga ctc atg gtt cct gca gaa tca gcc cta	2400
Ala Leu Leu Gly Ile Cys Gly Leu Met Val Pro Ala Glu Ser Ala Leu	
785 790 795 800	
att cct tat ttt gac tcc atc acg ctt cat atg aag tca tat gat agt	2448
Ile Pro Tyr Phe Asp Ser Ile Thr Leu His Met Lys Ser Tyr Asp Ser	
805 810 815	
cca gct gat aaa aag agt tcc ttt cag gtt gaa atg tca gaa ctt cga	2496
Pro Ala Asp Lys Lys Ser Ser Phe Gln Val Glu Met Ser Glu Leu Arg	
820 825 830	
tcc atc att ggc gga aca acc aac agg agc ctt gta ctt gtt gat gaa	2544
Ser Ile Ile Gly Gly Thr Thr Asn Arg Ser Leu Val Leu Val Asp Glu	
835 840 845	
ata tgc cga gga aca gaa act gca aaa ggg act tgc att gct ggt agc	2592
Ile Cys Arg Gly Thr Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly Ser	
850 855 860	
atc att gaa acc ctt gat gga att ggg tgt ctg ggt att gta tcc act	2640
Ile Ile Glu Thr Leu Asp Gly Ile Gly Cys Leu Gly Ile Val Ser Thr	
865 870 875 880	
cac ttg cat gga ata ttt act ttg ccc cta aac aaa aaa aac act gtg	2688
His Leu His Gly Ile Phe Thr Leu Pro Leu Asn Lys Lys Asn Thr Val	
885 890 895	
cac aaa gca atg ggc aca aca tcc att gat gga caa ata atg cct aca	2736
His Lys Ala Met Gly Thr Thr Ser Ile Asp Gly Gln Ile Met Pro Thr	
900 905 910	
tgg aag ttg aca gat gga gtt tgt aaa gaa agt ctt gct ttt gaa acg	2784
Trp Lys Leu Thr Asp Gly Val Cys Lys Glu Ser Leu Ala Phe Glu Thr	
915 920 925	
gct aag agg gaa gga att cct gag cat att gtt aga aga gct gaa tat	2832
Ala Lys Arg Glu Gly Ile Pro Glu His Ile Val Arg Arg Ala Glu Tyr	
930 935 940	
ctt tat cag ttg gtt tat gct aag gaa atg ctt ttt gca gaa aat ttc	2880
Leu Tyr Gln Leu Val Tyr Ala Lys Glu Met Leu Phe Ala Glu Asn Phe	
945 950 955 960	

-continued

cca aat gaa gaa aag ttt tct acc tgc atc aat gtt aat aat ttg aat	2928
Pro Asn Glu Glu Lys Phe Ser Thr Cys Ile Asn Val Asn Asn Leu Asn	
965 970 975	
gga aca cat ctt cat tca aaa agg ttc cta tca gga gct aat caa atg	2976
Gly Thr His Leu His Ser Lys Arg Phe Leu Ser Gly Ala Asn Gln Met	
980 985 990	
gaa gtt tta cgc gag gaa gtt gag aga gct gtc act gtg att tgc cag	3024
Glu Val Leu Arg Glu Glu Val Glu Arg Ala Val Thr Val Ile Cys Gln	
995 1000 1005	
gat cat ata aag gac cta aaa tgc aaa aag att gca ttg gag ctt	3069
Asp His Ile Lys Asp Leu Lys Cys Lys Lys Ile Ala Leu Glu Leu	
1010 1015 1020	
act gag ata aaa tgt ctc ata att ggt aca agg gag cta cca cct	3114
Thr Glu Ile Lys Cys Leu Ile Ile Gly Thr Arg Glu Leu Pro Pro	
1025 1030 1035	
cca tcg gtt gta ggt tct tca agc gtc tat gtg atg ttc aga cca	3159
Pro Ser Val Val Gly Ser Ser Ser Val Tyr Val Met Phe Arg Pro	
1040 1045 1050	
gat aag aaa ctc tat gta gga gag act gat gat ctc gag gga cgg	3204
Asp Lys Lys Leu Tyr Val Gly Glu Thr Asp Asp Leu Glu Gly Arg	
1055 1060 1065	
gtc cga aga cat cga tta aag gaa gga atg cat gat gca tca ttc	3249
Val Arg Arg His Arg Leu Lys Glu Gly Met His Asp Ala Ser Phe	
1070 1075 1080	
ctt tat ttt ctt gtc cca ggt aaa agc ttg gca tgc caa ttt gaa	3294
Leu Tyr Phe Leu Val Pro Gly Lys Ser Leu Ala Cys Gln Phe Glu	
1085 1090 1095	
tct ctg ctc atc aac caa ctt tct ggt caa ggc ttc caa ctg agc	3339
Ser Leu Leu Ile Asn Gln Leu Ser Gly Gln Gly Phe Gln Leu Ser	
1100 1105 1110	
aat ata gct gat ggt aaa cat agg aat ttt ggc act tcc aac ctg	3384
Asn Ile Ala Asp Gly Lys His Arg Asn Phe Gly Thr Ser Asn Leu	
1115 1120 1125	
tat aca taa	3393
Tyr Thr	
1130	

<210> SEQ ID NO 31

<211> LENGTH: 1130

<212> TYPE: PRT

<213> ORGANISM: Glycine max

<400> SEQUENCE: 31

Met Tyr Arg Val Ala Thr Arg Asn Val Ala Val Phe Phe Pro Arg Cys	
1 5 10 15	
Cys Ser Leu Ala His Tyr Thr Pro Ser Leu Phe Pro Ile Phe Thr Ser	
20 25 30	
Phe Ala Pro Ser Arg Phe Leu Arg Ile Asn Gly Cys Val Lys Asn Val	
35 40 45	
Ser Ser Tyr Thr Asp Lys Lys Val Ser Arg Gly Ser Ser Arg Ala Thr	
50 55 60	
Lys Lys Pro Lys Ile Pro Asn Asn Val Leu Asp Asp Lys Asp Leu Pro	
65 70 75 80	
His Ile Leu Trp Trp Lys Glu Arg Leu Gln Met Cys Arg Lys Phe Ser	
85 90 95	
Thr Val Gln Leu Ile Glu Arg Leu Glu Phe Ser Asn Leu Leu Gly Leu	
100 105 110	

-continued

Asn	Ser	Asn	Leu	Lys	Asn	Gly	Ser	Leu	Lys	Glu	Gly	Thr	Leu	Asn	Trp
		115					120					125			
Glu	Met	Leu	Gln	Phe	Lys	Ser	Lys	Phe	Pro	Arg	Gln	Val	Leu	Leu	Cys
	130					135					140				
Arg	Val	Gly	Glu	Phe	Tyr	Glu	Ala	Trp	Gly	Ile	Asp	Ala	Cys	Ile	Leu
	145				150					155					160
Val	Glu	Tyr	Val	Gly	Leu	Asn	Pro	Ile	Gly	Gly	Leu	Arg	Ser	Asp	Ser
				165					170					175	
Ile	Pro	Arg	Ala	Ser	Cys	Pro	Val	Val	Asn	Leu	Arg	Gln	Thr	Leu	Asp
			180					185					190		
Asp	Leu	Thr	Thr	Asn	Gly	Tyr	Ser	Val	Cys	Ile	Val	Glu	Glu	Ala	Gln
	195						200					205			
Gly	Pro	Ser	Gln	Ala	Arg	Ser	Arg	Lys	Arg	Arg	Phe	Ile	Ser	Gly	His
	210					215					220				
Ala	His	Pro	Gly	Asn	Pro	Tyr	Val	Tyr	Gly	Leu	Ala	Thr	Val	Asp	His
	225				230					235					240
Asp	Leu	Asn	Phe	Pro	Glu	Pro	Met	Pro	Val	Val	Gly	Ile	Ser	His	Ser
			245						250					255	
Ala	Arg	Gly	Tyr	Cys	Ile	Asn	Met	Val	Leu	Glu	Thr	Met	Lys	Thr	Tyr
		260						265					270		
Ser	Ser	Glu	Asp	Cys	Leu	Thr	Glu	Glu	Ala	Val	Val	Thr	Lys	Leu	Arg
	275						280					285			
Thr	Cys	Gln	Tyr	His	Tyr	Leu	Phe	Leu	His	Thr	Ser	Leu	Arg	Arg	Asn
	290					295					300				
Ser	Cys	Gly	Thr	Cys	Asn	Trp	Gly	Glu	Phe	Gly	Glu	Gly	Gly	Leu	Leu
	305				310					315					320
Trp	Gly	Glu	Cys	Ser	Ser	Arg	His	Phe	Asp	Trp	Phe	Asp	Gly	Asn	Pro
			325						330					335	
Val	Ser	Asp	Leu	Leu	Ala	Lys	Val	Lys	Glu	Leu	Tyr	Ser	Ile	Asp	Asp
		340						345					350		
Glu	Val	Thr	Phe	Arg	Asn	Thr	Thr	Val	Ser	Ser	Gly	His	Arg	Ala	Arg
	355						360					365			
Pro	Leu	Thr	Leu	Gly	Thr	Ser	Thr	Gln	Ile	Gly	Ala	Ile	Pro	Thr	Glu
	370					375					380				
Gly	Ile	Pro	Ser	Leu	Leu	Lys	Val	Leu	Leu	Pro	Ser	Asn	Cys	Asn	Gly
	385				390					395					400
Leu	Pro	Val	Leu	Tyr	Ile	Arg	Glu	Leu	Leu	Leu	Asn	Pro	Pro	Ser	Tyr
			405					410						415	
Glu	Ile	Ala	Ser	Lys	Ile	Gln	Ala	Thr	Cys	Lys	Leu	Met	Ser	Ser	Val
		420						425					430		
Thr	Cys	Ser	Ile	Pro	Glu	Phe	Thr	Cys	Val	Ser	Ser	Ala	Lys	Leu	Val
	435						440					445			
Lys	Leu	Glu	Trp	Arg	Glu	Val	Asn	His	Met	Glu	Phe	Cys	Arg	Ile	
	450					455				460					
Lys	Asn	Val	Leu	Asp	Glu	Ile	Leu	Gln	Met	Tyr	Ser	Thr	Ser	Glu	Leu
	465				470					475					480
Asn	Glu	Ile	Leu	Lys	His	Leu	Ile	Glu	Pro	Thr	Trp	Val	Ala	Thr	Gly
			485					490						495	
Leu	Glu	Ile	Asp	Phe	Glu	Thr	Leu	Val	Ala	Gly	Cys	Glu	Ile	Ala	Ser
			500					505						510	

-continued

Ser Lys Ile Gly Glu Ile Val Ser Leu Asp Asp Glu Asn Asp Gln Lys	515	520	525	
Ile Asn Ser Phe Ser Phe Ile Pro His Glu Phe Phe Glu Asp Met Glu	530	535	540	
Ser Lys Trp Lys Gly Arg Ile Lys Arg Ile His Ile Asp Asp Val Phe	545	550	555	560
Thr Ala Val Glu Lys Ala Ala Glu Ala Leu His Ile Ala Val Thr Glu	565	570	575	
Asp Phe Val Pro Val Val Ser Arg Ile Lys Ala Ile Val Ala Pro Leu	580	585	590	
Gly Gly Pro Lys Gly Glu Ile Ser Tyr Ala Arg Glu Gln Glu Ala Val	595	600	605	
Trp Phe Lys Gly Lys Arg Phe Thr Pro Asn Leu Trp Ala Gly Ser Pro	610	615	620	
Gly Glu Glu Gln Ile Lys Gln Leu Arg His Ala Leu Asp Ser Lys Gly	625	630	635	640
Arg Lys Val Gly Glu Glu Trp Phe Thr Thr Pro Lys Val Glu Ala Ala	645	650	655	
Leu Thr Arg Tyr His Glu Ala Asn Ala Lys Ala Lys Glu Arg Val Leu	660	665	670	
Glu Ile Leu Arg Gly Leu Ala Ala Glu Leu Gln Tyr Ser Ile Asn Ile	675	680	685	
Leu Val Phe Ser Ser Met Leu Leu Val Ile Ala Lys Ala Leu Phe Ala	690	695	700	
His Ala Ser Glu Gly Arg Arg Arg Arg Trp Val Phe Pro Thr Leu Val	705	710	715	720
Glu Ser His Gly Phe Glu Asp Val Lys Ser Leu Asp Lys Thr His Gly	725	730	735	
Met Lys Ile Ser Gly Leu Leu Pro Tyr Trp Phe His Ile Ala Glu Gly	740	745	750	
Val Val Arg Asn Asp Val Asp Met Gln Ser Leu Phe Leu Leu Thr Gly	755	760	765	
Pro Asn Gly Gly Gly Lys Ser Ser Phe Leu Arg Ser Ile Cys Ala Ala	770	775	780	
Ala Leu Leu Gly Ile Cys Gly Leu Met Val Pro Ala Glu Ser Ala Leu	785	790	795	800
Ile Pro Tyr Phe Asp Ser Ile Thr Leu His Met Lys Ser Tyr Asp Ser	805	810	815	
Pro Ala Asp Lys Lys Ser Ser Phe Gln Val Glu Met Ser Glu Leu Arg	820	825	830	
Ser Ile Ile Gly Gly Thr Thr Asn Arg Ser Leu Val Leu Val Asp Glu	835	840	845	
Ile Cys Arg Gly Thr Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly Ser	850	855	860	
Ile Ile Glu Thr Leu Asp Gly Ile Gly Cys Leu Gly Ile Val Ser Thr	865	870	875	880
His Leu His Gly Ile Phe Thr Leu Pro Leu Asn Lys Lys Asn Thr Val	885	890	895	
His Lys Ala Met Gly Thr Thr Ser Ile Asp Gly Gln Ile Met Pro Thr	900	905	910	
Trp Lys Leu Thr Asp Gly Val Cys Lys Glu Ser Leu Ala Phe Glu Thr				

-continued

915	920	925
Ala Lys Arg Glu Gly Ile Pro Glu His Ile Val Arg Arg Ala Glu Tyr 930 935 940		
Leu Tyr Gln Leu Val Tyr Ala Lys Glu Met Leu Phe Ala Glu Asn Phe 945 950 955 960		
Pro Asn Glu Glu Lys Phe Ser Thr Cys Ile Asn Val Asn Asn Leu Asn 965 970 975		
Gly Thr His Leu His Ser Lys Arg Phe Leu Ser Gly Ala Asn Gln Met 980 985 990		
Glu Val Leu Arg Glu Glu Val Glu Arg Ala Val Thr Val Ile Cys Gln 995 1000 1005		
Asp His Ile Lys Asp Leu Lys Cys Lys Lys Ile Ala Leu Glu Leu 1010 1015 1020		
Thr Glu Ile Lys Cys Leu Ile Ile Gly Thr Arg Glu Leu Pro Pro 1025 1030 1035		
Pro Ser Val Val Gly Ser Ser Ser Val Tyr Val Met Phe Arg Pro 1040 1045 1050		
Asp Lys Lys Leu Tyr Val Gly Glu Thr Asp Asp Leu Glu Gly Arg 1055 1060 1065		
Val Arg Arg His Arg Leu Lys Glu Gly Met His Asp Ala Ser Phe 1070 1075 1080		
Leu Tyr Phe Leu Val Pro Gly Lys Ser Leu Ala Cys Gln Phe Glu 1085 1090 1095		
Ser Leu Leu Ile Asn Gln Leu Ser Gly Gln Gly Phe Gln Leu Ser 1100 1105 1110		
Asn Ile Ala Asp Gly Lys His Arg Asn Phe Gly Thr Ser Asn Leu 1115 1120 1125		
Tyr Thr 1130		

<210> SEQ ID NO 32
 <211> LENGTH: 757
 <212> TYPE: DNA
 <213> ORGANISM: Saccharum officinarum
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (512)..(512)
 <223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 32

ccgcctctct cgccccccac ttcccacgcc ccacgccgcc tcccattcca gttccagcgt	60
ggacgcgacg ccggcgcgga gacgcggcgt ctccaagcac tagccccctg ttgttcttcc	120
gcgcggcgcg gccggcgcca tgcacgggt gctcgtgagc tcgctcgtgg ccgccacgcc	180
gcggtggctc cccctcgccg actccatcct ccggcgccgc cgcccgcgct gctccctct	240
tcccatgctg ctattcgacc ggaggacttg gtccaagcca aggaaggtct cagcaggcat	300
ttcagtgga tctaggaaag ctaacaaaca gggagaatat tgtgatgaaa gcatgctatc	360
tcatatcatg tggtggaag agaaaatgga gaagtgcaga aaaccatcat ctgtacagtt	420
gactcagagg cttgtgtatt cgaatatatt agggttggat ccgaatttaa gaaatggaag	480
cttgaaagat ggaacctga acatggagat tntgctatit aaatcaaaat ttcctcgtga	540
ggttctactt tgcagaaaca tgcaggctta aattctcttt ggagggttgc gttctgacag	600

-continued

```

aattcctaaa gctgggtgtc cagccggaat ttacggagac attggatgag ttgactcgat   660
gtgggaattc tgtgtgcaaa gtgaagaaat tacaggccga cccaagccct gccccggaaa   720
gtcgattaat tctgggcattg cccatcctgg agcccta                               757

```

```

<210> SEQ ID NO 33
<211> LENGTH: 139
<212> TYPE: PRT
<213> ORGANISM: Saccharum officinarum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (125)..(125)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid

```

```

<400> SEQUENCE: 33

```

```

Met His Arg Val Leu Val Ser Ser Leu Val Ala Ala Thr Pro Arg Trp
 1             5             10             15
Leu Pro Leu Ala Asp Ser Ile Leu Arg Arg Arg Arg Pro Arg Cys Ser
          20             25             30
Pro Leu Pro Met Leu Leu Phe Asp Arg Arg Thr Trp Ser Lys Pro Arg
          35             40             45
Lys Val Ser Arg Gly Ile Ser Val Ala Ser Arg Lys Ala Asn Lys Gln
          50             55             60
Gly Glu Tyr Cys Asp Glu Ser Met Leu Ser His Ile Met Trp Trp Lys
65             70             75             80
Glu Lys Met Glu Lys Cys Arg Lys Pro Ser Ser Val Gln Leu Thr Gln
          85             90             95
Arg Leu Val Tyr Ser Asn Ile Leu Gly Leu Asp Pro Asn Leu Arg Asn
          100            105            110
Gly Ser Leu Lys Asp Gly Thr Leu Asn Met Glu Ile Xaa Leu Phe Lys
          115            120            125
Ser Lys Phe Pro Arg Glu Val Leu Leu Cys Arg
          130            135

```

```

<210> SEQ ID NO 34
<211> LENGTH: 504
<212> TYPE: DNA
<213> ORGANISM: Saccharum officinarum

```

```

<400> SEQUENCE: 34

```

```

cacgtacctg tcctgaattc cccgaccgac ccatgcgtga gaacaagctt taattaaaac   60
atacctaagt atcttctggg gtgcgcttca cgcccacaga ggggaggaag gcatgcaaga   120
tgctaccacc ctatacatct tggttcctgg caagagcggt gcctgccagc tagaaaccct   180
tctcataaat cagcttcctt ctgagggcct caagctcatc aacaaggtag acggaaagca   240
taggaacttc ggtatatctt gaatctctgg agaggcaatt gctactcaac taaactaatc   300
acgtgaagat ctaatttagc tagacgacac tagtgagtct cattttggct actcaatagg   360
aggcaggagc taactgacac catgccgccc caatattggt gaactgatag cggagctagc   420
cttgaccata atacgggcat ctttttctcg tctaatgatg tagtacaatg caaatgatta   480
gcaatgcaat gacactcggt gtgc                                           504

```

```

<210> SEQ ID NO 35
<211> LENGTH: 72
<212> TYPE: PRT
<213> ORGANISM: Saccharum officinarum

```


-continued

<400> SEQUENCE: 35

Gly Arg Leu His Ala His Arg Gly Glu Glu Gly Met Gln Asp Ala Thr
 1 5 10 15
 Thr Leu Tyr Ile Leu Val Pro Gly Lys Ser Val Ala Cys Gln Leu Glu
 20 25 30
 Thr Leu Leu Ile Asn Gln Leu Pro Ser Glu Gly Phe Lys Leu Ile Asn
 35 40 45
 Lys Val Asp Gly Lys His Arg Asn Phe Gly Ile Phe Arg Ile Ser Gly
 50 55 60
 Glu Ala Ile Ala Thr Gln Leu Asn
 65 70

<210> SEQ ID NO 36

<211> LENGTH: 671

<212> TYPE: DNA

<213> ORGANISM: *Nicotiana tabacum*

<400> SEQUENCE: 36

aacaattcctt agccttctat gcttcagttt gtaaatgcta ctgttgagat tttttgttgt 60
 ctatttcacag ctgggtcaagt tgcttgagtt gagggaggca aatcatgtag agttctgcaa 120
 aataaagaat gtggtcgtatg aaatactgca gatgtacaga aattcagagc ttcgtgctat 180
 tttagagtca gctgatggat cctacttggg tggcaaccgg gttaaaagtc gattttgata 240
 ctctagttaa tgaatgtggg gagatttctg gtagaatcag tgaaataata tctgtacatg 300
 gtgaaagtga tcaaaagata agtccctatc ctatcatccc aaatgatttt tttgaagata 360
 tggagtcgcc atggaaagggt cgtgtcaaga ggatccattt ggaggaagca tatgcagaag 420
 tagacaaggc tgcagatgct ttatctttgg ctgtgagtct ctttttattt atcttcaaca 480
 atcctaataga tttacaagtt gtgcactgt gtgcgcttta atactcttctc attagctaag 540
 atatacatatt gctgtaaagg cagtcagctt ttcaacgtcc agtaaaagct ttttgataaa 600
 tccagtaata ttatctagga atttactgat cgatgaacaa ttttggggta atcgatagac 660
 aaataaaciaa g 671

<210> SEQ ID NO 37

<211> LENGTH: 488

<212> TYPE: DNA

<213> ORGANISM: *Lycopersicon esculentum*

<400> SEQUENCE: 37

gtttggtgaa ggtggacttt tgtggggaga atgtaatgct agacagcagg aatgggttga 60
 tggcaatcct atcgatgagc ttttgttcaa ggtaaaagag ctttatggtc tcaatgatga 120
 cattccattc agaaatgtca ctgttggttc agaaaatagg ccccgctcctt tacaccttgg 180
 aactgccaca caaatgtgtg ctattccaac cgaagggatt ccatgtttgt taaagggtgtt 240
 gcttcctcct cattgcagtg gtctaccagt cctgtatatt agggatcttc ttttaaatcc 300
 accaccctat gagatttctt cagacattca agaggcatgc agacttatga tgagtgtcac 360
 atgttcaatt cctgatttta cctgtatttc atctgcaaag ctggtcaagc tgcttgagtt 420
 gaggagggca aatcacgttg agttctgcaa aataaagagc atgggtcgaag agatactgca 480
 gttgtata 488

-continued

<210> SEQ ID NO 38
<211> LENGTH: 3373
<212> TYPE: DNA
<213> ORGANISM: Lycopersicon esculentum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (689)..(689)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 38

atgtattggg ttacggcaaa aaacgtcgtc gtttcagttc cccgttggcg ttcactgtcc	60
cttttctctc gtccaccact tcgcccgcgt ttcttatctt tctctccaca tactctgtgc	120
cgagagcaga tacgttgcgt gaaggagcgg aagttttttg ccacaacggc aaaaaaactc	180
aaacaaccaa aaagtattcc agaggaaaaa gactatgtta atattatgtg gtggaaagag	240
agaatggaat tcttgagaaa gccttcttcc gctcttctgg ctaagaggct tacatattgt	300
aacttgctgg gtgtggatcc gagtttgaga aatggaagtc ttaagagggt aacacttaac	360
tcggagatgt tgcagttcaa gtcaaaattt ccacgtgaag ttttgctctg tagagtaggt	420
gatttttatg aagctattgg attcgaatgc tgtattcttg tggaatatgc tggtttaa	480
ccatttgggt gcctgcactc agatagtata ccaaaagctg gttgtccagt tgtgaatcta	540
agacagacgc ttgatgatct cacacgtaat ggtttctctg tgtgcgtcgt ggagggaagt	600
cagggtccaa ctcaagctcg tgctcgttaag agtcgattta tatcagggca tgcacatcca	660
ggcagtcctt atgttttttg ccttggtgna gatgatcaag atcttgattt tccagaacca	720
atgcctgttg ttggaatata ccgttcagcg aaggggtatt gcattatctc tgtttacgag	780
actatgaaga ctactctgt ggaagatggc ctaactgaag aagccgtagt caccaaactt	840
cgtacttgct gatgccatca tttttttttg cataattcat tgaagaacaa ttcctcagga	900
acatcgcggt ggggagagtt tgggtgaagg ggacttttgt ggggagaatg taatgctaga	960
cagcaggaat ggttggatgg caatcctatc gatgagcttt tgttcaaggt aaaagagctt	1020
tatggtctca atgatgacat tccattcaga aatgtcactg ttgtttcaga aaataggccc	1080
cgtcctttac accttggaac tgccacacaa attggtgcta ttccaaccga agggattcca	1140
tgtttgttaa aggtgttgct tcctcctcat tgcagtggtc taccagtcct gtatattagg	1200
gatcttcttt taaatccacc agcctatgag atttcttcag acatacaaga ggcatgcaga	1260
cttatgatga gtgtcacatg ttcaattcct gattttacct gtatttcacg tgcaaagctg	1320
gtcaagctgc ttgagttgag ggaggcaaat cacgttgagt tctgcaaaat aaagagcatg	1380
gtcgaagaga tactgcagtt gtatagaaat tcagagcttc gtgctatwgt agagttactg	1440
atggatccta cttgggtggc aactgggttg aaagttgatt ttgatacact agtaaatgaa	1500
tgtggaaga tttcttgtag aatcagtgaa ataatatccg tacatggtga aaatgatcaa	1560
aagattagtt cctatcctat catcccaaat gatttctttg aagatatgga gttgttgtgg	1620
aaaggccgtg tcaagaggat ccatttggag gaagcatatg cagaagtaga aaaggctgcg	1680
gatgccttat ctttagccat aacagaagat ttcctaccta ttatttcaag aataagggcc	1740
acgatggccc cacttgaggg aactaaaggg gagattttgt atgcccgta gcatggagct	1800
gtatggttta agggaaagag atttgtacca actgtttggg ctggaaccgc tggagaagaa	1860
caaattaagc aactcagacc tgctctagat tcaaagggga agaaggttgg agaagaatgg	1920

-continued

```

ttcactacaa tgagggtgga agatgcaata gctaggtatc acgaggcaag tgctagggca 1980
aagtcaaggg tcttggaatt gctaagggga ctttcttctg aattactatc taagatcaat 2040
atccttatct ttgcactctgt cttgaatgtg atagcaaaat cattattttc tcatgtgagt 2100
gaaggaagaa gaagaaattg gattttccca acaatcacac aatttaacaa atgtcaggac 2160
acagaggcac ttaattggaac tgatggaatg aagataattg gtctatctcc ttattgggtt 2220
gatgcagcac gagggactgg tgtacaggat acagtagata tgcagtccat gtttctttta 2280
acagggccaa atgggtggggg caaatcaagc ttgctgcgtt cgttggtgtc agctgcattg 2340
ctaggaatgt gtgggttcat ggttccagct gaatcagctg tcattcctca ttttgactca 2400
attatgctgc atatgaaatc atatgatagt cctgttgatg gaaaaagttc atttcagatt 2460
gaaatgtctg aaattcggtc tctgattact ggtgccactt caagaagtct tgtacttata 2520
gatgaaatat gtcgaggaac agaaacagca aaaggacat gtattgctgg aagtgtcata 2580
gaaaccctgg acgaaattgg ctgtttggga attgtatcaa cccacttgca tggaatattt 2640
gatttaccac tgaaatcaa gaagaccgtg tataaagcaa tgggagctga atatgttgac 2700
ggccaaccaa taccaacttg gaaactcatt gatgggatct gtaaagagag tctagcattt 2760
gaaacagctc agagagaagg aattccagaa atattaatcc aaagagcaga agaattgtat 2820
aattcagctt acgggaatca gataccaagg aagatagacc aaataagacc tcttcgttca 2880
gatattgacc tcaatagcac agataacagt totgaccaat taaatggtag aagacaaata 2940
gctttggatt ctgacacaaa gttaatgcat cgaatgggaa tttcaagcaa gaaacttgaa 3000
gatgctatct gtcttatctg tgagaagaag ttaattgagc tgtataaaat gaaaaatccg 3060
tcagaaatgc caatggtgaa ttgcgttctt attgctgcca gggacagcc ggctccatca 3120
acaattgggt cttcaagtgt ctatataatg ctaagacctg acaaaaagtt gtatgttgga 3180
cagactgatg atcttgaggg cagagtacgt gctcatcgct tgaaggaggg aatggaaaac 3240
gcgtcattcc tatatttctt agtctctggc aagagcatcg cctgccaatt ggaaactctt 3300
ctaataaatc aacttcctaa tcatggtttt cagctaacaa acgttgctga tggtaagcat 3360
cgtaattttg gca 3373

```

```

<210> SEQ ID NO 39
<211> LENGTH: 3373
<212> TYPE: DNA
<213> ORGANISM: Lycopersicon esculentum
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(3372)
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (689)..(689)
<223> OTHER INFORMATION: n is a, c, g, or t

```

```

<400> SEQUENCE: 39

```

```

atg tat tgg gtt acg gca aaa aac gtc gtc gtt tca gtt ccc cgt tgg 48
Met Tyr Trp Val Thr Ala Lys Asn Val Val Val Ser Val Pro Arg Trp
1 5 10 15

cgt tca ctg tcc ctt ttc ctc cgt cca cca ctt cgc cgg cgt ttc tta 96
Arg Ser Leu Ser Leu Phe Leu Arg Pro Pro Leu Arg Arg Arg Phe Leu
20 25 30

tct ttc tct cca cat act ctg tgc cga gag cag ata cgt tgc gtg aag 144
Ser Phe Ser Pro His Thr Leu Cys Arg Glu Gln Ile Arg Cys Val Lys
35 40 45

```

-continued

gag cgg aag ttt ttt gcc aca acg gca aaa aaa ctc aaa caa cca aaa	192
Glu Arg Lys Phe Phe Ala Thr Thr Ala Lys Lys Leu Lys Gln Pro Lys	
50 55 60	
agt att cca gag gaa aaa gac tat gtt aat att atg tgg tgg aaa gag	240
Ser Ile Pro Glu Glu Lys Asp Tyr Val Asn Ile Met Trp Trp Lys Glu	
65 70 75 80	
aga atg gaa ttc ttg aga aag cct tct tcc gct ctt ctg gct aag agg	288
Arg Met Glu Phe Leu Arg Lys Pro Ser Ser Ala Leu Leu Ala Lys Arg	
85 90 95	
ctt aca tat tgt aac ttg ctg ggt gtg gat ccg agt ttg aga aat gga	336
Leu Thr Tyr Cys Asn Leu Leu Gly Val Asp Pro Ser Leu Arg Asn Gly	
100 105 110	
agt ctt aaa gag gga aca ctt aac tcg gag atg ttg cag ttc aag tca	384
Ser Leu Lys Glu Gly Thr Leu Asn Ser Glu Met Leu Gln Phe Lys Ser	
115 120 125	
aaa ttt cca cgt gaa gtt ttg ctc tgt aga gta ggt gat ttt tat gaa	432
Lys Phe Pro Arg Glu Val Leu Leu Cys Arg Val Gly Asp Phe Tyr Glu	
130 135 140	
gct att gga ttc gat gct tgt att ctt gtg gaa tat gct ggt tta aat	480
Ala Ile Gly Phe Asp Ala Cys Ile Leu Val Glu Tyr Ala Gly Leu Asn	
145 150 155 160	
cca ttt ggt ggc ctg cac tca gat agt ata cca aaa gct ggt tgt cca	528
Pro Phe Gly Gly Leu His Ser Asp Ser Ile Pro Lys Ala Gly Cys Pro	
165 170 175	
gtt gtg aat cta aga cag acg ctt gat gat ctc aca cgt aat ggt ttc	576
Val Val Asn Leu Arg Gln Thr Leu Asp Asp Leu Thr Arg Asn Gly Phe	
180 185 190	
tct gtg tgc gtc gtg gag gaa gtt cag ggt cca act caa gct cgt gct	624
Ser Val Cys Val Val Glu Glu Val Gln Gly Pro Thr Gln Ala Arg Ala	
195 200 205	
cgt aag agt cga ttt ata tca ggg cat gca cat cca ggc agt ccc tat	672
Arg Lys Ser Arg Phe Ile Ser Gly His Ala His Pro Gly Ser Pro Tyr	
210 215 220	
gtt ttt ggc ctt gtt gna gat gat caa gat ctt gat ttt cca gaa cca	720
Val Phe Gly Leu Val Xaa Asp Asp Gln Asp Leu Asp Phe Pro Glu Pro	
225 230 235 240	
atg cct gtt gtt gga ata tcc cgt tca gcg aag ggg tat tgc att atc	768
Met Pro Val Val Gly Ile Ser Arg Ser Ala Lys Gly Tyr Cys Ile Ile	
245 250 255	
tct gtt tac gag act atg aag act tac tct gtg gaa gat ggc cta act	816
Ser Val Tyr Glu Thr Met Lys Thr Tyr Ser Val Glu Asp Gly Leu Thr	
260 265 270	
gaa gaa gcc gta gtc acc aaa ctt cgt act tgt cga tgc cat cat ttt	864
Glu Glu Ala Val Val Thr Lys Leu Arg Thr Cys Arg Cys His His Phe	
275 280 285	
ttt ttg cat aat tca ttg aag aac aat tcc tca gga aca tcg cgt tgg	912
Phe Leu His Asn Ser Leu Lys Asn Asn Ser Ser Gly Thr Ser Arg Trp	
290 295 300	
gga gag ttt ggt gaa ggt gga ctt ttg tgg gga gaa tgt aat gct aga	960
Gly Glu Phe Gly Glu Gly Gly Leu Leu Trp Gly Glu Cys Asn Ala Arg	
305 310 315 320	
cag cag gaa tgg ttg gat ggc aat cct atc gat gag ctt ttg ttc aag	1008
Gln Gln Glu Trp Leu Asp Gly Asn Pro Ile Asp Glu Leu Leu Phe Lys	
325 330 335	
gta aaa gag ctt tat ggt ctc aat gat gac att cca ttc aga aat gtc	1056
Val Lys Glu Leu Tyr Gly Leu Asn Asp Asp Ile Pro Phe Arg Asn Val	
340 345 350	

-continued

act gtt gtt tca gaa aat agg ccc cgt cct tta cac ctt gga act gcc	1104
Thr Val Val Ser Glu Asn Arg Pro Arg Pro Leu His Leu Gly Thr Ala	
355 360 365	
aca caa att ggt gct att cca acc gaa ggg att cca tgt ttg tta aag	1152
Thr Gln Ile Gly Ala Ile Pro Thr Glu Gly Ile Pro Cys Leu Leu Lys	
370 375 380	
gtg ttg ctt cct cct cat tgc agt ggt cta cca gtc ctg tat att agg	1200
Val Leu Leu Pro Pro His Cys Ser Gly Leu Pro Val Leu Tyr Ile Arg	
385 390 395 400	
gat ctt ctt tta aat cca cca gcc tat gag att tct tca gac ata caa	1248
Asp Leu Leu Leu Asn Pro Pro Ala Tyr Glu Ile Ser Ser Asp Ile Gln	
405 410 415	
gag gca tgc aga ctt atg atg agt gtc aca tgt tca att cct gat ttt	1296
Glu Ala Cys Arg Leu Met Met Ser Val Thr Cys Ser Ile Pro Asp Phe	
420 425 430	
acc tgt att tca tct gca aag ctg gtc aag ctg ctt gag ttg agg gag	1344
Thr Cys Ile Ser Ser Ala Lys Leu Val Lys Leu Leu Glu Leu Arg Glu	
435 440 445	
gca aat cac gtt gag ttc tgc aaa ata aag agc atg gtc gaa gag ata	1392
Ala Asn His Val Glu Phe Cys Lys Ile Lys Ser Met Val Glu Glu Ile	
450 455 460	
ctg cag ttg tat aga aat tca gag ctt cgt gct atw gta gag tta ctg	1440
Leu Gln Leu Tyr Arg Asn Ser Glu Leu Arg Ala Xaa Val Glu Leu Leu	
465 470 475 480	
atg gat cct act tgg gtg gca act ggg ttg aaa gtt gat ttt gat aca	1488
Met Asp Pro Thr Trp Val Ala Thr Gly Leu Lys Val Asp Phe Asp Thr	
485 490 495	
cta gta aat gaa tgt gga aag att tct tgt aga atc agt gaa ata ata	1536
Leu Val Asn Glu Cys Gly Lys Ile Ser Cys Arg Ile Ser Glu Ile Ile	
500 505 510	
tcc gta cat ggt gaa aat gat caa aag att agt tcc tat cct atc atc	1584
Ser Val His Gly Glu Asn Asp Gln Lys Ile Ser Ser Tyr Pro Ile Ile	
515 520 525	
cca aat gat ttc ttt gaa gat atg gag ttg ttg tgg aaa ggc cgt gtc	1632
Pro Asn Asp Phe Phe Glu Asp Met Glu Leu Leu Trp Lys Gly Arg Val	
530 535 540	
aag agg atc cat ttg gag gaa gca tat gca gaa gta gaa aag gct gcg	1680
Lys Arg Ile His Leu Glu Glu Ala Tyr Ala Glu Val Glu Lys Ala Ala	
545 550 555 560	
gat gct tta tct tta gcc ata aca gaa gat ttc cta cct att att tca	1728
Asp Ala Leu Ser Leu Ala Ile Thr Glu Asp Phe Leu Pro Ile Ile Ser	
565 570 575	
aga ata agg gcc acg atg gcc cca ctt gga gga act aaa ggg gag att	1776
Arg Ile Arg Ala Thr Met Ala Pro Leu Gly Gly Thr Lys Gly Glu Ile	
580 585 590	
ttg tat gcc cgt gag cat gga gct gta tgg ttt aag gga aag aga ttt	1824
Leu Tyr Ala Arg Glu His Gly Ala Val Trp Phe Lys Gly Lys Arg Phe	
595 600 605	
gta cca act gtt tgg gct gga acc gct gga gaa gaa caa att aag caa	1872
Val Pro Thr Val Trp Ala Gly Thr Ala Gly Glu Glu Gln Ile Lys Gln	
610 615 620	
ctc aga cct gct cta gat tca aag ggg aag aag gtt gga gaa gaa tgg	1920
Leu Arg Pro Ala Leu Asp Ser Lys Gly Lys Val Gly Glu Glu Trp	
625 630 635 640	
ttc act aca atg agg gtg gaa gat gca ata gct agg tat cac gag gca	1968
Phe Thr Thr Met Arg Val Glu Asp Ala Ile Ala Arg Tyr His Glu Ala	
645 650 655	

-continued

agt gct agg gca aag tca agg gtc ttg gaa ttg cta agg gga ctt tct Ser Ala Arg Ala Lys Ser Arg Val Leu Glu Leu Leu Arg Gly Leu Ser 660 665 670	2016
tct gaa tta cta tct aag atc aat atc ctt atc ttt gca tct gtc ttg Ser Glu Leu Leu Ser Lys Ile Asn Ile Leu Ile Phe Ala Ser Val Leu 675 680 685	2064
aat gtg ata gca aaa tca tta ttt tct cat gtg agt gaa gga aga aga Asn Val Ile Ala Lys Ser Leu Phe Ser His Val Ser Glu Gly Arg Arg 690 695 700	2112
aga aat tgg att ttc cca aca atc aca caa ttt aac aaa tgt cag gac Arg Asn Trp Ile Phe Pro Thr Ile Thr Gln Phe Asn Lys Cys Gln Asp 705 710 715 720	2160
aca gag gca ctt aat gga act gat gga atg aag ata att ggt cta tct Thr Glu Ala Leu Asn Gly Thr Asp Gly Met Lys Ile Ile Gly Leu Ser 725 730 735	2208
cct tat tgg ttt gat gca gca cga ggg act ggt gta cag gat aca gta Pro Tyr Trp Phe Asp Ala Ala Arg Gly Thr Gly Val Gln Asp Thr Val 740 745 750	2256
gat atg cag tcc atg ttt ctt tta aca ggt cca aat ggt ggg ggc aaa Asp Met Gln Ser Met Phe Leu Thr Gly Pro Asn Gly Gly Gly Lys 755 760 765	2304
tca agc ttg ctg cgt tcg ttg tgt gca gct gca ttg cta gga atg tgt Ser Ser Leu Leu Arg Ser Leu Cys Ala Ala Ala Leu Leu Gly Met Cys 770 775 780	2352
ggg ttc atg gtt cca gct gaa tca gct gtc att cct cat ttt gac tca Gly Phe Met Val Pro Ala Glu Ser Ala Val Ile Pro His Phe Asp Ser 785 790 795 800	2400
att atg ctg cat atg aaa tca tat gat agt cct gtt gat gga aaa agt Ile Met Leu His Met Lys Ser Tyr Asp Ser Pro Val Asp Gly Lys Ser 805 810 815	2448
tca ttt cag att gaa atg tct gaa att cgg tct ctg att act ggt gcc Ser Phe Gln Ile Glu Met Ser Glu Ile Arg Ser Leu Ile Thr Gly Ala 820 825 830	2496
act tca aga agt ctt gta ctt ata gat gaa ata tgt cga gga aca gaa Thr Ser Arg Ser Leu Val Leu Ile Asp Glu Ile Cys Arg Gly Thr Glu 835 840 845	2544
aca gca aaa ggg aca tgt att gct gga agt gtc ata gaa acc ctg gac Thr Ala Lys Gly Thr Cys Ile Ala Gly Ser Val Ile Glu Thr Leu Asp 850 855 860	2592
gaa att ggc tgt ttg gga att gta tca acc cac ttg cat gga ata ttt Glu Ile Gly Cys Leu Gly Ile Val Ser Thr His Leu His Gly Ile Phe 865 870 875 880	2640
gat tta ccc ctg aaa atc aag aag acc gtg tat aaa gca atg gga gct Asp Leu Pro Leu Lys Ile Lys Lys Thr Val Tyr Lys Ala Met Gly Ala 885 890 895	2688
gaa tat gtt gac ggt caa cca ata cca act tgg aaa ctc att gat ggg Glu Tyr Val Asp Gly Gln Pro Ile Pro Thr Trp Lys Leu Ile Asp Gly 900 905 910	2736
atc tgt aaa gag agt cta gca ttt gaa aca gct cag aga gaa gga att Ile Cys Lys Glu Ser Leu Ala Phe Glu Thr Ala Gln Arg Glu Gly Ile 915 920 925	2784
cca gaa ata tta atc caa aga gca gaa gaa ttg tat aat tca gct tac Pro Glu Ile Leu Ile Gln Arg Ala Glu Glu Leu Tyr Asn Ser Ala Tyr 930 935 940	2832
ggg aat cag ata cca agg aag ata gac caa ata aga cct ctt cgt tca Gly Asn Gln Ile Pro Arg Lys Ile Asp Gln Ile Arg Pro Leu Arg Ser 945 950 955 960	2880

-continued

gat att gac ctc aat agc aca gat aac agt tct gac caa tta aat ggt Asp Ile Asp Leu Asn Ser Thr Asp Asn Ser Ser Asp Gln Leu Asn Gly 965 970 975	2928
aca aga caa ata gct ttg gat tct agc aca aag tta atg cat cga atg Thr Arg Gln Ile Ala Leu Asp Ser Ser Thr Lys Leu Met His Arg Met 980 985 990	2976
gga att tca agc aag aaa ctt gaa gat gct atc tgt ctt atc tgt gag Gly Ile Ser Ser Lys Lys Leu Glu Asp Ala Ile Cys Leu Ile Cys Glu 995 1000 1005	3024
aag aag tta att gag ctg tat aaa atg aaa aat ccg tca gaa atg Lys Lys Leu Ile Glu Leu Tyr Lys Met Lys Asn Pro Ser Glu Met 1010 1015 1020	3069
cca atg gtg aat tgc gtt ctt att gct gcc agg gaa cag ccg gct Pro Met Val Asn Cys Val Leu Ile Ala Ala Arg Glu Gln Pro Ala 1025 1030 1035	3114
cca tca aca att ggt gct tca agt gtc tat ata atg cta aga cct Pro Ser Thr Ile Gly Ala Ser Ser Val Tyr Ile Met Leu Arg Pro 1040 1045 1050	3159
gac aaa aag ttg tat gtt gga cag act gat gat ctt gag ggc aga Asp Lys Lys Leu Tyr Val Gly Gln Thr Asp Asp Leu Glu Gly Arg 1055 1060 1065	3204
gta cgt gct cat cgc ttg aag gag gga atg gaa aac gcg tca ttc Val Arg Ala His Arg Leu Lys Glu Gly Met Glu Asn Ala Ser Phe 1070 1075 1080	3249
cta tat ttc tta gtc tct ggc aag agc atc gcc tgc caa ttg gaa Leu Tyr Phe Leu Val Ser Gly Lys Ser Ile Ala Cys Gln Leu Glu 1085 1090 1095	3294
act ctt cta ata aat caa ctt cct aat cat ggt ttt cag cta aca Thr Leu Leu Ile Asn Gln Leu Pro Asn His Gly Phe Gln Leu Thr 1100 1105 1110	3339
aac gtt gct gat ggt aag cat cgt aat ttt ggc a Asn Val Ala Asp Gly Lys His Arg Asn Phe Gly 1115 1120	3373

<210> SEQ ID NO 40
 <211> LENGTH: 1124
 <212> TYPE: PRT
 <213> ORGANISM: Lycopersicon esculentum
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (230)..(230)
 <223> OTHER INFORMATION: The 'Xaa' at location 230 stands for Glu, Gly,
 Ala, or Val.
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (476)..(476)
 <223> OTHER INFORMATION: The 'Xaa' at location 476 stands for Ile.
 <400> SEQUENCE: 40

Met Tyr Trp Val Thr Ala Lys Asn Val Val Val Ser Val Pro Arg Trp 1 5 10 15
Arg Ser Leu Ser Leu Phe Leu Arg Pro Pro Leu Arg Arg Arg Phe Leu 20 25 30
Ser Phe Ser Pro His Thr Leu Cys Arg Glu Gln Ile Arg Cys Val Lys 35 40 45
Glu Arg Lys Phe Phe Ala Thr Thr Ala Lys Lys Leu Lys Gln Pro Lys 50 55 60
Ser Ile Pro Glu Glu Lys Asp Tyr Val Asn Ile Met Trp Trp Lys Glu 65 70 75 80

-continued

Arg	Met	Glu	Phe	Leu	Arg	Lys	Pro	Ser	Ser	Ala	Leu	Leu	Ala	Lys	Arg
				85					90					95	
Leu	Thr	Tyr	Cys	Asn	Leu	Leu	Gly	Val	Asp	Pro	Ser	Leu	Arg	Asn	Gly
			100					105					110		
Ser	Leu	Lys	Glu	Gly	Thr	Leu	Asn	Ser	Glu	Met	Leu	Gln	Phe	Lys	Ser
		115					120					125			
Lys	Phe	Pro	Arg	Glu	Val	Leu	Leu	Cys	Arg	Val	Gly	Asp	Phe	Tyr	Glu
	130					135					140				
Ala	Ile	Gly	Phe	Asp	Ala	Cys	Ile	Leu	Val	Glu	Tyr	Ala	Gly	Leu	Asn
145					150					155					160
Pro	Phe	Gly	Gly	Leu	His	Ser	Asp	Ser	Ile	Pro	Lys	Ala	Gly	Cys	Pro
				165					170					175	
Val	Val	Asn	Leu	Arg	Gln	Thr	Leu	Asp	Asp	Leu	Thr	Arg	Asn	Gly	Phe
		180						185					190		
Ser	Val	Cys	Val	Val	Glu	Glu	Val	Gln	Gly	Pro	Thr	Gln	Ala	Arg	Ala
		195					200					205			
Arg	Lys	Ser	Arg	Phe	Ile	Ser	Gly	His	Ala	His	Pro	Gly	Ser	Pro	Tyr
	210					215					220				
Val	Phe	Gly	Leu	Val	Xaa	Asp	Asp	Gln	Asp	Leu	Asp	Phe	Pro	Glu	Pro
225					230					235					240
Met	Pro	Val	Val	Gly	Ile	Ser	Arg	Ser	Ala	Lys	Gly	Tyr	Cys	Ile	Ile
				245					250					255	
Ser	Val	Tyr	Glu	Thr	Met	Lys	Thr	Tyr	Ser	Val	Glu	Asp	Gly	Leu	Thr
			260					265					270		
Glu	Glu	Ala	Val	Val	Thr	Lys	Leu	Arg	Thr	Cys	Arg	Cys	His	His	Phe
		275					280					285			
Phe	Leu	His	Asn	Ser	Leu	Lys	Asn	Asn	Ser	Ser	Gly	Thr	Ser	Arg	Trp
	290					295					300				
Gly	Glu	Phe	Gly	Glu	Gly	Gly	Leu	Leu	Trp	Gly	Glu	Cys	Asn	Ala	Arg
305					310					315					320
Gln	Gln	Glu	Trp	Leu	Asp	Gly	Asn	Pro	Ile	Asp	Glu	Leu	Leu	Phe	Lys
				325				330						335	
Val	Lys	Glu	Leu	Tyr	Gly	Leu	Asn	Asp	Asp	Ile	Pro	Phe	Arg	Asn	Val
			340					345					350		
Thr	Val	Val	Ser	Glu	Asn	Arg	Pro	Arg	Pro	Leu	His	Leu	Gly	Thr	Ala
		355					360					365			
Thr	Gln	Ile	Gly	Ala	Ile	Pro	Thr	Glu	Gly	Ile	Pro	Cys	Leu	Leu	Lys
	370					375					380				
Val	Leu	Leu	Pro	Pro	His	Cys	Ser	Gly	Leu	Pro	Val	Leu	Tyr	Ile	Arg
385					390					395					400
Asp	Leu	Leu	Leu	Asn	Pro	Pro	Ala	Tyr	Glu	Ile	Ser	Ser	Asp	Ile	Gln
				405					410					415	
Glu	Ala	Cys	Arg	Leu	Met	Met	Ser	Val	Thr	Cys	Ser	Ile	Pro	Asp	Phe
			420					425					430		
Thr	Cys	Ile	Ser	Ser	Ala	Lys	Leu	Val	Lys	Leu	Leu	Glu	Leu	Arg	Glu
		435					440					445			
Ala	Asn	His	Val	Glu	Phe	Cys	Lys	Ile	Lys	Ser	Met	Val	Glu	Glu	Ile
	450					455					460				
Leu	Gln	Leu	Tyr	Arg	Asn	Ser	Glu	Leu	Arg	Ala	Xaa	Val	Glu	Leu	Leu
465					470					475					480

-continued

Met	Asp	Pro	Thr	Trp	Val	Ala	Thr	Gly	Leu	Lys	Val	Asp	Phe	Asp	Thr	485	490	495
Leu	Val	Asn	Glu	Cys	Gly	Lys	Ile	Ser	Cys	Arg	Ile	Ser	Glu	Ile	Ile	500	505	510
Ser	Val	His	Gly	Glu	Asn	Asp	Gln	Lys	Ile	Ser	Ser	Tyr	Pro	Ile	Ile	515	520	525
Pro	Asn	Asp	Phe	Phe	Glu	Asp	Met	Glu	Leu	Leu	Trp	Lys	Gly	Arg	Val	530	535	540
Lys	Arg	Ile	His	Leu	Glu	Glu	Ala	Tyr	Ala	Glu	Val	Glu	Lys	Ala	Ala	545	550	555
Asp	Ala	Leu	Ser	Leu	Ala	Ile	Thr	Glu	Asp	Phe	Leu	Pro	Ile	Ile	Ser	565	570	575
Arg	Ile	Arg	Ala	Thr	Met	Ala	Pro	Leu	Gly	Gly	Thr	Lys	Gly	Glu	Ile	580	585	590
Leu	Tyr	Ala	Arg	Glu	His	Gly	Ala	Val	Trp	Phe	Lys	Gly	Lys	Arg	Phe	595	600	605
Val	Pro	Thr	Val	Trp	Ala	Gly	Thr	Ala	Gly	Glu	Glu	Gln	Ile	Lys	Gln	610	615	620
Leu	Arg	Pro	Ala	Leu	Asp	Ser	Lys	Gly	Lys	Lys	Val	Gly	Glu	Glu	Trp	625	630	635
Phe	Thr	Thr	Met	Arg	Val	Glu	Asp	Ala	Ile	Ala	Arg	Tyr	His	Glu	Ala	645	650	655
Ser	Ala	Arg	Ala	Lys	Ser	Arg	Val	Leu	Glu	Leu	Leu	Arg	Gly	Leu	Ser	660	665	670
Ser	Glu	Leu	Leu	Ser	Lys	Ile	Asn	Ile	Leu	Ile	Phe	Ala	Ser	Val	Leu	675	680	685
Asn	Val	Ile	Ala	Lys	Ser	Leu	Phe	Ser	His	Val	Ser	Glu	Gly	Arg	Arg	690	695	700
Arg	Asn	Trp	Ile	Phe	Pro	Thr	Ile	Thr	Gln	Phe	Asn	Lys	Cys	Gln	Asp	705	710	715
Thr	Glu	Ala	Leu	Asn	Gly	Thr	Asp	Gly	Met	Lys	Ile	Ile	Gly	Leu	Ser	725	730	735
Pro	Tyr	Trp	Phe	Asp	Ala	Ala	Arg	Gly	Thr	Gly	Val	Gln	Asp	Thr	Val	740	745	750
Asp	Met	Gln	Ser	Met	Phe	Leu	Leu	Thr	Gly	Pro	Asn	Gly	Gly	Gly	Lys	755	760	765
Ser	Ser	Leu	Leu	Arg	Ser	Leu	Cys	Ala	Ala	Ala	Leu	Leu	Gly	Met	Cys	770	775	780
Gly	Phe	Met	Val	Pro	Ala	Glu	Ser	Ala	Val	Ile	Pro	His	Phe	Asp	Ser	785	790	795
Ile	Met	Leu	His	Met	Lys	Ser	Tyr	Asp	Ser	Pro	Val	Asp	Gly	Lys	Ser	805	810	815
Ser	Phe	Gln	Ile	Glu	Met	Ser	Glu	Ile	Arg	Ser	Leu	Ile	Thr	Gly	Ala	820	825	830
Thr	Ser	Arg	Ser	Leu	Val	Leu	Ile	Asp	Glu	Ile	Cys	Arg	Gly	Thr	Glu	835	840	845
Thr	Ala	Lys	Gly	Thr	Cys	Ile	Ala	Gly	Ser	Val	Ile	Glu	Thr	Leu	Asp	850	855	860
Glu	Ile	Gly	Cys	Leu	Gly	Ile	Val	Ser	Thr	His	Leu	His	Gly	Ile	Phe	865	870	875
Asp	Leu	Pro	Leu	Lys	Ile	Lys	Lys	Thr	Val	Tyr	Lys	Ala	Met	Gly	Ala			

-continued

885										890										895											
Glu	Tyr	Val	Asp	Gly	Gln	Pro	Ile	Pro	Thr	Trp	Lys	Leu	Ile	Asp	Gly																
			900						905						910																
Ile	Cys	Lys	Glu	Ser	Leu	Ala	Phe	Glu	Thr	Ala	Gln	Arg	Glu	Gly	Ile																
		915					920							925																	
Pro	Glu	Ile	Leu	Ile	Gln	Arg	Ala	Glu	Glu	Leu	Tyr	Asn	Ser	Ala	Tyr																
		930				935								940																	
Gly	Asn	Gln	Ile	Pro	Arg	Lys	Ile	Asp	Gln	Ile	Arg	Pro	Leu	Arg	Ser																
		945			950				955					960																	
Asp	Ile	Asp	Leu	Asn	Ser	Thr	Asp	Asn	Ser	Ser	Asp	Gln	Leu	Asn	Gly																
				965					970					975																	
Thr	Arg	Gln	Ile	Ala	Leu	Asp	Ser	Ser	Thr	Lys	Leu	Met	His	Arg	Met																
			980				985							990																	
Gly	Ile	Ser	Ser	Lys	Lys	Leu	Glu	Asp	Ala	Ile	Cys	Leu	Ile	Cys	Glu																
		995					1000							1005																	
Lys	Lys	Leu	Ile	Glu	Leu	Tyr	Lys	Met	Lys	Asn	Pro	Ser	Glu	Met																	
		1010				1015								1020																	
Pro	Met	Val	Asn	Cys	Val	Leu	Ile	Ala	Ala	Arg	Glu	Gln	Pro	Ala																	
		1025				1030								1035																	
Pro	Ser	Thr	Ile	Gly	Ala	Ser	Ser	Val	Tyr	Ile	Met	Leu	Arg	Pro																	
		1040				1045								1050																	
Asp	Lys	Lys	Leu	Tyr	Val	Gly	Gln	Thr	Asp	Asp	Leu	Glu	Gly	Arg																	
		1055				1060								1065																	
Val	Arg	Ala	His	Arg	Leu	Lys	Glu	Gly	Met	Glu	Asn	Ala	Ser	Phe																	
		1070				1075								1080																	
Leu	Tyr	Phe	Leu	Val	Ser	Gly	Lys	Ser	Ile	Ala	Cys	Gln	Leu	Glu																	
		1085				1090								1095																	
Thr	Leu	Leu	Ile	Asn	Gln	Leu	Pro	Asn	His	Gly	Phe	Gln	Leu	Thr																	
		1100				1105								1110																	
Asn	Val	Ala	Asp	Gly	Lys	His	Arg	Asn	Phe	Gly																					
		1115				1120																									

<210> SEQ ID NO 41

<211> LENGTH: 622

<212> TYPE: DNA

<213> ORGANISM: Triticum aestivum

<400> SEQUENCE: 41

```

cctactacga acatagctag gccatatgac caatcagaca aaattggggt ggaaaacatg      60
gtatcagtta gctcctgcct cctataagcc aaaaaaacag ataaggaaat caaagatgaa    120
gctccactcc cctttggcct ctacgagtta aaactggatg ttcagtgggt cagttcagtg    180
tgcagccatg gcttctccag aggttacaga cataccaaag ttccgatgct tgccatctgc    240
cttgttgggt agcttaaaac ctttcgtggg tagctgattt atgagaagag tctccagttg    300
gcaggcaaca ctcttgccag gaacaatgat gtataatatt gtggcatcct gcataccttc    360
cttcgatcta tgagcaccaa gacggccac aagatcatcc gtctgtccaa catagagctt    420
gttgtcagct ctgatgatga tatagatgct ggacctccca acagttgaag gtggcggttg    480
ctccctagca cctacagtaa cgcagaccac ctcaaccagt tctgagatgc ttctcttggt    540
gtagagatcc aacagtttat ctttgcatat tgtggtaaca atgctctcga catcctttgg    600
cagcagtcga gtagcacctg ac                                           622

```

-continued

<210> SEQ ID NO 42
 <211> LENGTH: 148
 <212> TYPE: PRT
 <213> ORGANISM: Triticum aestivum

<400> SEQUENCE: 42

```
Ser Gly Ala Thr Gly Leu Leu Pro Lys Asp Val Glu Ser Ile Val Thr
1           5           10           15
Thr Ile Cys Lys Asp Lys Leu Leu Asp Leu Tyr Asn Lys Arg Ser Ile
20           25           30
Ser Glu Leu Val Glu Val Val Cys Val Thr Val Gly Ala Arg Glu Gln
35           40           45
Pro Pro Pro Ser Thr Val Gly Arg Ser Ser Ile Tyr Ile Ile Ile Arg
50           55           60
Arg Asp Asn Lys Leu Tyr Val Gly Gln Thr Asp Asp Leu Val Gly Arg
65           70           75           80
Leu Gly Ala His Arg Ser Lys Glu Gly Met Gln Asp Ala Thr Ile Leu
85           90           95
Tyr Ile Ile Val Pro Gly Lys Ser Val Ala Cys Gln Leu Glu Thr Leu
100          105          110
Leu Ile Asn Gln Leu Pro Thr Lys Gly Phe Lys Leu Thr Asn Lys Ala
115          120          125
Asp Gly Lys His Arg Asn Phe Gly Met Ser Val Thr Ser Gly Glu Ala
130          135          140
Met Ala Ala His
145
```

<210> SEQ ID NO 43
 <211> LENGTH: 523
 <212> TYPE: DNA
 <213> ORGANISM: Zinnia elegans

<400> SEQUENCE: 43

```
ggagtcttcg tggaagaatc gtgttaagaa gattcattta aaagaagctt atgaagaagt      60
ggataaggca gctgaagcct tacccttagc tgtaacggag gattttcttc ctataatttg      120
tagaataaaa gctaccacag caccacttgg aggacaaaaa ggggaaattt tgtatgttcg      180
ggaacacaaa gctatatggt tcaagggcaa acgttttgta ccaaccatag gggctaatac      240
gcctgtagaa aagcaaatta aacaacttaa gccctctgta gattcaaagg gtagaaaagt      300
tggagaggaa tggtttacca caagtaaagt ggaggatgca ctctcaaggt accatgaagc      360
tggtgcacaaa gcgaagtcca tgggtgttaga gttattgagg ggactgtctg ctgaattgca      420
agctgaaatt aatgttctcg tgtttgcctc catgttgctt attatcgcaa aggcattggt      480
tgctcatgtg aggtattcta tatctgaatt ttttgaccgt tgt                        523
```

<210> SEQ ID NO 44
 <211> LENGTH: 174
 <212> TYPE: PRT
 <213> ORGANISM: Zinnia elegans

<400> SEQUENCE: 44

```
Glu Ser Ser Trp Lys Asn Arg Val Lys Lys Ile His Leu Lys Glu Ala
1           5           10           15
```

-continued

Tyr Glu Glu Val Asp Lys Ala Ala Glu Ala Leu Ser Leu Ala Val Thr
 20 25 30
 Glu Asp Phe Leu Pro Ile Ile Cys Arg Ile Lys Ala Thr Thr Ala Pro
 35 40 45
 Leu Gly Gly Pro Lys Gly Glu Ile Leu Tyr Val Arg Glu His Lys Ala
 50 55 60
 Ile Trp Phe Lys Gly Lys Arg Phe Val Pro Thr Ile Gly Ala Asn Thr
 65 70 75 80
 Pro Val Glu Lys Gln Ile Lys Gln Leu Lys Pro Ser Val Asp Ser Lys
 85 90 95
 Gly Arg Lys Val Gly Glu Glu Trp Phe Thr Thr Ser Lys Val Glu Asp
 100 105 110
 Ala Leu Ser Arg Tyr His Glu Ala Gly Ala Lys Ala Lys Ser Met Val
 115 120 125
 Leu Glu Leu Leu Arg Gly Leu Ser Ala Glu Leu Gln Ala Glu Ile Asn
 130 135 140
 Val Leu Val Phe Ala Ser Met Leu Leu Ile Ile Ala Lys Ala Leu Phe
 145 150 155 160
 Ala His Val Arg Tyr Ser Ile Ser Glu Phe Phe Asp Arg Cys
 165 170

<210> SEQ ID NO 45

<211> LENGTH: 3381

<212> TYPE: DNA

<213> ORGANISM: Phaseolus vulgaris

<400> SEQUENCE: 45

```

atgtacaggg cagttaccag aaacgtcgcc gttttcctgc ctcgttgccg ctctctctcg      60
cacttctctc attcgtatt tcccttcttc atttcatccc ttccctctcg ctctctctcg      120
ataaatggac gtgtcaagaa tgtatcaact tatatggata ataacagggt ttcaagggga      180
agtagtagga ccaccaagaa gccaaaagta ccaaataatg ttttagatga caaagatctt      240
cctcacatat cgtggtggaa ggagaggttg caaatgtgca aaaagttttc gactgtccag      300
ctaattcaaa ggcttgaatt ttctaatttg cttggtcttg attccaaatt gaaaaatgga      360
agtgtgaagg aaggaacact caactgggaa atgttgcagt tcaagtcaaa atttccacgt      420
caagtattac tctgcagagt aggggaattc tatgaagcat ggggaataga tgcttgtgtt      480
ctagttgaat atgctggttt aaatccctgt ggtggtctcc aatcagatag tgttccaagg      540
gctggttgtc ctggttgtaa tcttcgacag actttagatg atctgaccca aaatggttat      600
tcagtgtgca tcattgagga agttcagggc ccaactcaag ctcgatccag gaaacgccgc      660
tttatatctg ggcattctca tcctggaaat ccctatgtat atggacttgc tgcagttgat      720
catgatctta actttcctga gccaatgcct gtaataggaa tatctcattc tgcgaggggc      780
tattgcatta acatggtgct agagactatg aaaacatact cttatgaaga ttgcttgaca      840
gaggaagcaa ttgtgacaaa gcttcgtact tgtcaatatc atcacttatt cttgcataca      900
tctttgacgc aggattcttg tggcaccagc aaatggggag aattcgggtga ggggggtctc      960
ttatggggag aatgtagtgc tagacatttt gaatgggttg atggcagccc tctctctgat      1020
ctcttggtca aggtaaagga gctttatggt cttgatgatg aggttacttt tcgaaacaca      1080
accgtatctt cgagacatag ggctcgacct ttaacccttg gaacatctac tcaaattggt      1140

```

-continued

gccattcata	cggaaggaat	accttctttg	ttaaaggtct	tactttcacc	aagttgcaat	1200
ggattaccgg	ttctgtatat	aaggaatctt	ctcttgaatc	ctccttctta	tgagatcgca	1260
tccaaaaattc	aggaaacatg	caaaccttatg	agcagtttaa	cgtgctcaat	tccagaattt	1320
acgtgtgttt	cttcagcaaa	gcttgtaaag	ctacttgagt	ggagggaggt	caaccatatg	1380
gaattttgta	gaataaagaa	tgtgcttgat	gagattttgc	atatgtacaa	aacctctgag	1440
ctcaatgaaa	tattgaaaaa	tttaattgat	ccaacatggg	cgacaactgg	gttagacatc	1500
gactttgaaa	cactggtttc	tggtgtgtaa	gttgcatcta	gtaagatcag	tgaataatc	1560
tctctggatg	gtgggaatga	tcagaaaatc	aactctttat	ctattattcc	ttatgaattt	1620
tttgaagata	cggagtctaa	atggaaaggt	cgaataaaaa	gagtcocat	agatgaggtg	1680
ttacagcag	tgcaaaaagc	agctgaggtc	ttgcacatag	ctgtcactga	agattttgtt	1740
cctgttgttt	ctagagtaaa	ggctactata	gccccacttg	gaggtcctag	gggagaaatt	1800
tcttatgctc	gtgagcatga	ggcagtttgg	ttcagaggca	aacgctttac	gccgagtttg	1860
tggtctggta	gccctgggga	ggaacaaatt	aaacagctta	ggcatgcttt	agattctaaa	1920
ggtaaaagg	taggggagga	atggtttact	acaccgaagg	ttgaggctgc	attaacaagg	1980
taccatgaag	caaatgccaa	ggcaacagaa	cgaagtttgg	aaattttaag	ggaactcgct	2040
actgaattgc	attacagtat	aaacattctt	gtcttttcat	ccacgttgct	tgttattacc	2100
aaagctttat	tcgctcatgc	aagtgaagg	agaagaagga	gatgggtttt	tccaacactt	2160
gcagaatcga	atgggtttga	ggatgtgaaa	tcttcggaca	aatccatgg	gatgaagata	2220
gttggtttag	caccttattg	gttcacata	gcagaagga	ttgtgcgtaa	tgatgttgat	2280
atgcaatcat	tatttctttt	gacaggacca	aatggtggtg	ggaatcaag	ttacttcgt	2340
tcaatttggt	ctgccgcat	acttggtata	tgtgggtcca	tggttcctgc	agaatctgcc	2400
gtgatccctt	attttgactc	catcacgctt	catatgaagt	cgtatgatag	tccagctgat	2460
aaaaagagtt	cctttcaggt	ggaaatgtca	gaacttagat	ccatcattgg	cggaaaccacc	2520
aaaaggagcc	ttgtacttgt	tgatgaaatt	tgccgaggaa	cagaaactgc	aaaagggact	2580
tgtattgctg	gtagtatcat	tgaaactcta	gaaagaattg	gttgtctggg	tgttgtgtcc	2640
actcacttgc	atggaatatt	tactttgccc	ctcaacatca	aaagcactgt	gcacaaagca	2700
atgggcacaa	cgtgcattga	tggacaaata	cttcctacat	ggaagctgac	agatggagtc	2760
tgtaaagaaa	gtcttgcttt	tgaaactgcc	attagggag	gaattcctga	gcctattata	2820
agaagagctg	aatgtcttta	taagtcagtt	tatgcagagg	aaaatttccc	aatgaagag	2880
aagttttcta	cttgcaacaa	tttgaataat	ttgaatacaa	caagtcttta	ttctaaagg	2940
ttcttatcag	gagctaatca	aatggaaggt	tttcgccagg	aagttgaaag	agctattact	3000
gtgatatgcc	aggattatat	aatggaacgg	aaaaacaaaa	agattgcatt	ggagcttcct	3060
gagataaaat	gtctccta	cggtgaagag	gagcagccac	ctccatctgt	tgtaggttct	3120
tcaagcgtct	atgtgatttt	cacgccagat	aagaaactct	acgtaggaga	gacggatgat	3180
ctagagggcc	gggttcgaag	acatagattg	aaagaaggta	tgatgaagc	atcatttctt	3240
tattttcttg	ttccgggaaa	aagcttgga	tgccaatttg	aatctctgct	catcaaccag	3300
ctttctagtc	aaggcttcca	actgagcaac	atggctgatg	gtaaacatag	gaattttggc	3360
acttccaacc	tctatgcata	a				3381

-continued

```

<210> SEQ ID NO 46
<211> LENGTH: 3381
<212> TYPE: DNA
<213> ORGANISM: Phaseolus vulgaris
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(3381)

<400> SEQUENCE: 46

atg tac agg gca gtt acc aga aac gtc gcc gtt ttc ctg cct cgt tgc      48
Met Tyr Arg Ala Val Thr Arg Asn Val Ala Val Phe Leu Pro Arg Cys
1          5          10          15

cgc tct ctc tcg cac ttc tct cat tcg cta ttt ccc ttc ttc att tca      96
Arg Ser Leu Ser His Phe Ser His Ser Leu Phe Pro Phe Phe Ile Ser
          20          25          30

tcc ctt ccc tct cgc ttc ctt cga ata aat gga cgt gtc aag aat gta     144
Ser Leu Pro Ser Arg Phe Leu Arg Ile Asn Gly Arg Val Lys Asn Val
          35          40          45

tca act tat atg gat aat aac agg gtt tca agg gga agt agt agg acc     192
Ser Thr Tyr Met Asp Asn Asn Arg Val Ser Arg Gly Ser Ser Arg Thr
          50          55          60

acc aag aag cca aaa gta cca aat aat gtt tta gat gac aaa gat ctt     240
Thr Lys Lys Pro Lys Val Pro Asn Asn Val Leu Asp Asp Lys Asp Leu
65          70          75          80

cct cac ata tcg tgg tgg aag gag agg ttg caa atg tgc aaa aag ttt     288
Pro His Ile Ser Trp Trp Lys Glu Arg Leu Gln Met Cys Lys Lys Phe
          85          90          95

tcg act gtc cag cta att caa agg ctt gaa ttt tct aat ttg ctt ggt     336
Ser Thr Val Gln Leu Ile Gln Arg Leu Glu Phe Ser Asn Leu Leu Gly
          100          105          110

ctg gat tcc aaa ttg aaa aat gga agt gtg aag gaa gga aca ctc aac     384
Leu Asp Ser Lys Leu Lys Asn Gly Ser Val Lys Glu Gly Thr Leu Asn
          115          120          125

tgg gaa atg ttg cag ttc aag tca aaa ttt cca cgt caa gta tta ctc     432
Trp Glu Met Leu Gln Phe Lys Ser Lys Phe Pro Arg Gln Val Leu Leu
          130          135          140

tgc aga gta ggg gaa ttc tat gaa gca tgg gga ata gat gct tgt gtt     480
Cys Arg Val Gly Glu Phe Tyr Glu Ala Trp Gly Ile Asp Ala Cys Val
          145          150          155          160

cta gtt gaa tat gct ggt tta aat ccc tgt ggt ggt ctc caa tca gat     528
Leu Val Glu Tyr Ala Gly Leu Asn Pro Cys Gly Gly Leu Gln Ser Asp
          165          170          175

agt gtt cca agg gct ggt tgt cct gtt gtg aat ctt cga cag act tta     576
Ser Val Pro Arg Ala Gly Cys Pro Val Val Asn Leu Arg Gln Thr Leu
          180          185          190

gat gat ctg acc caa aat ggt tat tca gtg tgc atc att gag gaa gtt     624
Asp Asp Leu Thr Gln Asn Gly Tyr Ser Val Cys Ile Ile Glu Glu Val
          195          200          205

cag ggc cca act caa gct cga tcc agg aaa cgc cgc ttt ata tct ggg     672
Gln Gly Pro Thr Gln Ala Arg Ser Arg Lys Arg Arg Phe Ile Ser Gly
          210          215          220

cat gct cat cct gga aat ccc tat gta tat gga ctt gct gca gtt gat     720
His Ala His Pro Gly Asn Pro Tyr Val Tyr Gly Leu Ala Ala Val Asp
          225          230          235          240

cat gat ctt aac ttt cct gag cca atg cct gta ata gga ata tct cat     768
His Asp Leu Asn Phe Pro Glu Pro Met Pro Val Ile Gly Ile Ser His
          245          250          255

tct gcg agg ggc tat tgc att aac atg gtg cta gag act atg aaa aca     816

```

-continued

Ser	Ala	Arg	Gly	Tyr	Cys	Ile	Asn	Met	Val	Leu	Glu	Thr	Met	Lys	Thr	
			260					265					270			
tac	tct	tat	gaa	gat	tgc	ttg	aca	gag	gaa	gca	att	gtg	aca	aag	ctt	864
Tyr	Ser	Tyr	Glu	Asp	Cys	Leu	Thr	Glu	Glu	Ala	Ile	Val	Thr	Lys	Leu	
		275					280					285				
cgt	act	tgt	caa	tat	cat	cac	tta	ttc	ttg	cat	aca	tct	ttg	acg	cag	912
Arg	Thr	Cys	Gln	Tyr	His	His	Leu	Phe	Leu	His	Thr	Ser	Leu	Thr	Gln	
		290				295					300					
gat	tct	tgt	ggc	acc	agc	aaa	tggt	gga	gaa	ttc	ggg	gag	ggg	ggg	ctc	960
Asp	Ser	Cys	Gly	Thr	Ser	Lys	Trp	Gly	Glu	Phe	Gly	Glu	Gly	Gly	Leu	
305						310				315					320	
tta	tggt	gga	gaa	tgt	agt	tct	aga	cat	ttt	gaa	tggt	ttt	gat	ggc	agc	1008
Leu	Trp	Gly	Glu	Cys	Ser	Ser	Arg	His	Phe	Glu	Trp	Phe	Asp	Gly	Ser	
				325					330					335		
cct	ctc	tct	gat	ctc	ttg	gtc	aag	gta	aag	gag	ctt	tat	ggg	ctt	gat	1056
Pro	Leu	Ser	Asp	Leu	Leu	Val	Lys	Val	Lys	Glu	Leu	Tyr	Gly	Leu	Asp	
			340					345					350			
gat	gag	gtt	act	ttt	cga	aac	aca	acc	gta	tct	tcg	aga	cat	agg	gct	1104
Asp	Glu	Val	Thr	Phe	Arg	Asn	Thr	Thr	Val	Ser	Ser	Arg	His	Arg	Ala	
		355					360					365				
cga	cct	tta	acc	ctt	gga	aca	tct	act	caa	att	ggg	gcc	att	cat	acg	1152
Arg	Pro	Leu	Thr	Leu	Gly	Thr	Ser	Thr	Gln	Ile	Gly	Ala	Ile	His	Thr	
	370					375					380					
gaa	gga	ata	cct	tct	ttg	tta	aag	gtc	tta	ctt	tca	cca	agt	tgc	aat	1200
Glu	Gly	Ile	Pro	Ser	Leu	Leu	Lys	Val	Leu	Ser	Pro	Ser	Ser	Cys	Asn	
385					390					395					400	
gga	tta	ccg	gtt	ctg	tat	ata	agg	aat	ctt	ctc	ttg	aat	cct	cct	tct	1248
Gly	Leu	Pro	Val	Leu	Tyr	Ile	Arg	Asn	Leu	Leu	Leu	Asn	Pro	Pro	Ser	
			405						410					415		
tat	gag	atc	gca	tcc	aaa	att	cag	gaa	aca	tgc	aaa	ctt	atg	agc	agt	1296
Tyr	Glu	Ile	Ala	Ser	Lys	Ile	Gln	Glu	Thr	Cys	Lys	Leu	Met	Ser	Ser	
		420						425					430			
tta	acg	tgc	tca	att	cca	gaa	ttt	acg	tgt	gtt	tct	tca	gca	aag	ctt	1344
Leu	Thr	Cys	Ser	Ile	Pro	Glu	Phe	Thr	Cys	Val	Ser	Ser	Ala	Lys	Leu	
		435				440						445				
gta	aag	cta	ctt	gag	tggt	agg	gag	gtc	aac	cat	atg	gaa	ttt	tgt	aga	1392
Val	Lys	Leu	Leu	Glu	Trp	Arg	Glu	Val	Asn	His	Met	Glu	Phe	Cys	Arg	
		450				455					460					
ata	aag	aat	gtg	ctt	gat	gag	att	ttg	cat	atg	tac	aaa	acc	tct	gag	1440
Ile	Lys	Asn	Val	Leu	Asp	Glu	Ile	Leu	His	Met	Tyr	Lys	Thr	Ser	Glu	
465					470					475					480	
ctc	aat	gaa	ata	ttg	aaa	aat	tta	att	gat	cca	aca	tggt	gcg	aca	act	1488
Leu	Asn	Glu	Ile	Leu	Lys	Asn	Leu	Ile	Asp	Pro	Thr	Trp	Ala	Thr	Thr	
				485					490					495		
ggg	tta	gac	atc	gac	ttt	gaa	aca	ctg	gtt	tct	gga	tgt	gaa	gtt	gca	1536
Gly	Leu	Asp	Ile	Asp	Phe	Glu	Thr	Leu	Val	Ser	Gly	Cys	Glu	Val	Ala	
		500						505					510			
tct	agt	aag	atc	agt	gaa	ata	atc	tct	ctg	gat	ggg	ggg	aat	gat	cag	1584
Ser	Ser	Lys	Ile	Ser	Glu	Ile	Ile	Ser	Leu	Asp	Gly	Gly	Asn	Asp	Gln	
		515					520					525				
aaa	atc	aac	tct	tta	tct	att	att	cct	tat	gaa	ttt	ttt	gaa	gat	acg	1632
Lys	Ile	Asn	Ser	Leu	Ser	Ile	Ile	Pro	Tyr	Glu	Phe	Phe	Glu	Asp	Thr	
		530				535					540					
gag	tct	aaa	tggt	aaa	ggg	cga	ata	aaa	aga	gtc	cat	ata	gat	gag	gtg	1680
Glu	Ser	Lys	Trp	Lys	Gly	Arg	Ile	Lys	Arg	Val	His	Ile	Asp	Glu	Val	
545					550					555					560	
ttt	aca	gca	gtg	caa	aaa	gca	gct	gag	gtc	ttg	cac	ata	gct	gtc	act	1728

-continued

Phe Thr Ala Val Gln Lys Ala Ala Glu Val Leu His Ile Ala Val Thr	
565 570 575	
gaa gat ttt gtt cct gtt gtt tct aga gta aag gct act ata gcc cca	1776
Glu Asp Phe Val Pro Val Val Ser Arg Val Lys Ala Thr Ile Ala Pro	
580 585 590	
ctt gga ggt cct agg gga gaa att tct tat gct cgt gag cat gag gca	1824
Leu Gly Gly Pro Arg Gly Glu Ile Ser Tyr Ala Arg Glu His Glu Ala	
595 600 605	
gtt tgg ttc aga ggc aaa cgc ttt acg ccg agt ttg tgg tct ggt agc	1872
Val Trp Phe Arg Gly Lys Arg Phe Thr Pro Ser Leu Trp Ser Gly Ser	
610 615 620	
cct ggg gag gaa caa att aaa cag ctt agg cat gct tta gat tct aaa	1920
Pro Gly Glu Glu Gln Ile Lys Gln Leu Arg His Ala Leu Asp Ser Lys	
625 630 635 640	
ggt aaa agg gta ggg gag gaa tgg ttt act aca ccg aag gtt gag gct	1968
Gly Lys Arg Val Gly Glu Glu Trp Phe Thr Thr Pro Lys Val Glu Ala	
645 650 655	
gca tta aca agg tac cat gaa gca aat gcc aag gca aca gaa cga gtt	2016
Ala Leu Thr Arg Tyr His Glu Ala Asn Ala Lys Ala Thr Glu Arg Val	
660 665 670	
ttg gaa att tta agg gaa ctc gct act gaa ttg cat tac agt ata aac	2064
Leu Glu Ile Leu Arg Glu Leu Ala Thr Glu Leu His Tyr Ser Ile Asn	
675 680 685	
att ctt gtc ttt tca tcc acg ttg ctt gtt att acc aaa gct tta ttc	2112
Ile Leu Val Phe Ser Ser Thr Leu Leu Val Ile Thr Lys Ala Leu Phe	
690 695 700	
gct cat gca agt gaa ggg aga aga agg aga tgg gtt ttt cca aca ctt	2160
Ala His Ala Ser Glu Gly Arg Arg Arg Arg Trp Val Phe Pro Thr Leu	
705 710 715 720	
gca gaa tcg aat ggg ttt gag gat gtg aaa tct tcg gac aaa atc cat	2208
Ala Glu Ser Asn Gly Phe Glu Asp Val Lys Ser Ser Asp Lys Ile His	
725 730 735	
ggg atg aag ata gtt ggt tta gca cct tat tgg ttc cac ata gca gaa	2256
Gly Met Lys Ile Val Gly Leu Ala Pro Tyr Trp Phe His Ile Ala Glu	
740 745 750	
ggt att gtg cgt aat gat gtt gat atg caa tca tta ttt ctt ttg aca	2304
Gly Ile Val Arg Asn Asp Val Asp Met Gln Ser Leu Phe Leu Leu Thr	
755 760 765	
gga cca aat ggt ggt ggg aaa tca agt tta ctt cgt tca att tgt gct	2352
Gly Pro Asn Gly Gly Gly Lys Ser Ser Leu Leu Arg Ser Ile Cys Ala	
770 775 780	
gcc gca tta ctt ggt ata tgt ggg ctc atg gtt cct gca gaa tct gcc	2400
Ala Ala Leu Leu Gly Ile Cys Gly Leu Met Val Pro Ala Glu Ser Ala	
785 790 795 800	
gtg att cct tat ttt gac tcc atc acg ctt cat atg aag tcg tat gat	2448
Val Ile Pro Tyr Phe Asp Ser Ile Thr Leu His Met Lys Ser Tyr Asp	
805 810 815	
agt cca gct gat aaa aag agt tcc ttt cag gtg gaa atg tca gaa ctt	2496
Ser Pro Ala Asp Lys Lys Ser Ser Phe Gln Val Glu Met Ser Glu Leu	
820 825 830	
aga tcc atc att ggc gga acc acc aaa agg agc ctt gta ctt gtt gat	2544
Arg Ser Ile Gly Gly Thr Lys Arg Ser Leu Val Leu Val Asp	
835 840 845	
gaa att tgc cga gga aca gaa act gca aaa ggg act tgt att gct ggt	2592
Glu Ile Cys Arg Gly Thr Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly	
850 855 860	
agt atc att gaa act cta gaa aga att ggt tgt ctg ggt gtt gtg tcc	2640

-continued

Ser Ile Ile Glu Thr Leu Glu Arg Ile Gly Cys Leu Gly Val Val Ser	
865 870 875 880	
act cac ttg cat gga ata ttt act ttg ccc ctc aac atc aaa agc act	2688
Thr His Leu His Gly Ile Phe Thr Leu Pro Leu Asn Ile Lys Ser Thr	
885 890 895	
gtg cac aaa gca atg ggc aca acg tgc att gat gga caa ata ctt cct	2736
Val His Lys Ala Met Gly Thr Thr Cys Ile Asp Gly Gln Ile Leu Pro	
900 905 910	
aca tgg aag ctg aca gat gga gtc tgt aaa gaa agt ctt gct ttt gaa	2784
Thr Trp Lys Leu Thr Asp Gly Val Cys Lys Glu Ser Leu Ala Phe Glu	
915 920 925	
act gcc att agg gaa gga att cct gag cct att ata aga aga gct gaa	2832
Thr Ala Ile Arg Glu Gly Ile Pro Glu Pro Ile Ile Arg Arg Ala Glu	
930 935 940	
tgt ctt tat aag tca gtt tat gca gag gaa aat ttc cca aat gaa gag	2880
Cys Leu Tyr Lys Ser Val Tyr Ala Glu Glu Asn Phe Pro Asn Glu Glu	
945 950 955 960	
aag ttt tct act tgc aac aat ttg aat aat ttg aat aca aca agt ctt	2928
Lys Phe Ser Thr Cys Asn Asn Leu Asn Asn Leu Asn Thr Thr Ser Leu	
965 970 975	
tat tct aaa ggg ttc tta tca gga gct aat caa atg gaa ggt ttt cgc	2976
Tyr Ser Lys Gly Phe Leu Ser Gly Ala Asn Gln Met Glu Gly Phe Arg	
980 985 990	
cag gaa gtt gaa aga gct att act gtg ata tgc cag gat tat ata atg	3024
Gln Glu Val Glu Arg Ala Ile Thr Val Ile Cys Gln Asp Tyr Ile Met	
995 1000 1005	
gaa cgg aaa aac aaa aag att gca ttg gag ctt cct gag ata aaa	3069
Glu Arg Lys Asn Lys Lys Ile Ala Leu Glu Leu Pro Glu Ile Lys	
1010 1015 1020	
tgt ctc cta atc ggt aag agg gag cag cca cct cca tct gtt gta	3114
Cys Leu Leu Ile Gly Lys Arg Glu Gln Pro Pro Pro Ser Val Val	
1025 1030 1035	
ggt tct tca agc gtc tat gtg att ttc acg cca gat aag aaa ctc	3159
Gly Ser Ser Ser Val Tyr Val Ile Phe Thr Pro Asp Lys Lys Leu	
1040 1045 1050	
tac gta gga gag acg gat gat cta gag ggc cgg gtt cga aga cat	3204
Tyr Val Gly Glu Thr Asp Asp Leu Glu Gly Arg Val Arg Arg His	
1055 1060 1065	
aga ttg aaa gaa ggt atg gat gaa gca tca ttt ctt tat ttt ctt	3249
Arg Leu Lys Glu Gly Met Asp Glu Ala Ser Phe Leu Tyr Phe Leu	
1070 1075 1080	
gtt ccg gga aaa agc ttg gca tgc caa ttt gaa tct ctg ctc atc	3294
Val Pro Gly Lys Ser Leu Ala Cys Gln Phe Glu Ser Leu Leu Ile	
1085 1090 1095	
aac cag ctt tct agt caa ggc ttc caa ctg agc aac atg gct gat	3339
Asn Gln Leu Ser Ser Gln Gly Phe Gln Leu Ser Asn Met Ala Asp	
1100 1105 1110	
ggt aaa cat agg aat ttt ggc act tcc aac ctc tat gca taa	3381
Gly Lys His Arg Asn Phe Gly Thr Ser Asn Leu Tyr Ala	
1115 1120 1125	

<210> SEQ ID NO 47

<211> LENGTH: 1126

<212> TYPE: PRT

<213> ORGANISM: Phaseolus vulgaris

<400> SEQUENCE: 47

Met Tyr Arg Ala Val Thr Arg Asn Val Ala Val Phe Leu Pro Arg Cys

-continued

1	5	10	15
Arg Ser Leu	Ser His Phe	Ser His Ser	Leu Phe Pro Phe Phe Ile Ser
	20	25	30
Ser Leu Pro	Ser Arg Phe	Leu Arg Ile	Asn Gly Arg Val Lys Asn Val
35		40	45
Ser Thr Tyr	Met Asp Asn	Asn Arg Val	Ser Arg Gly Ser Ser Arg Thr
50		55	60
Thr Lys Lys	Pro Lys Val	Pro Asn Asn	Val Leu Asp Asp Lys Asp Leu
65		70	75
Pro His Ile	Ser Trp Trp	Lys Glu Arg	Leu Gln Met Cys Lys Lys Phe
	85	90	95
Ser Thr Val	Gln Leu Ile	Gln Arg Leu	Glu Phe Ser Asn Leu Leu Gly
	100	105	110
Leu Asp Ser	Lys Leu Lys	Asn Gly Ser	Val Lys Glu Gly Thr Leu Asn
115		120	125
Trp Glu Met	Leu Gln Phe	Lys Ser Lys	Phe Pro Arg Gln Val Leu Leu
130		135	140
Cys Arg Val	Gly Glu Phe	Tyr Glu Ala	Trp Gly Ile Asp Ala Cys Val
145		150	155
Leu Val Glu	Tyr Ala Gly	Leu Asn Pro	Cys Gly Gly Leu Gln Ser Asp
	165	170	175
Ser Val Pro	Arg Ala Gly	Cys Pro Val	Val Asn Leu Arg Gln Thr Leu
	180	185	190
Asp Asp Leu	Thr Gln Asn	Gly Tyr Ser	Val Cys Ile Ile Glu Glu Val
195		200	205
Gln Gly Pro	Thr Gln Ala	Arg Ser Arg	Lys Arg Arg Phe Ile Ser Gly
210		215	220
His Ala His	Pro Gly Asn	Pro Tyr Val	Tyr Gly Leu Ala Ala Val Asp
225		230	235
His Asp Leu	Asn Phe Pro	Glu Pro Met	Pro Val Ile Gly Ile Ser His
	245	250	255
Ser Ala Arg	Gly Tyr Cys	Ile Asn Met	Val Leu Glu Thr Met Lys Thr
	260	265	270
Tyr Ser Tyr	Glu Asp Cys	Leu Thr Glu	Glu Ala Ile Val Thr Lys Leu
275		280	285
Arg Thr Cys	Gln Tyr His	His Leu Phe	Leu His Thr Ser Leu Thr Gln
290		295	300
Asp Ser Cys	Gly Thr Ser	Lys Trp Gly	Glu Phe Gly Glu Gly Gly Leu
305		310	315
Leu Trp Gly	Glu Cys Ser	Ser Arg His	Phe Glu Trp Phe Asp Gly Ser
	325	330	335
Pro Leu Ser	Asp Leu Leu	Val Lys Val	Lys Glu Leu Tyr Gly Leu Asp
	340	345	350
Asp Glu Val	Thr Phe Arg	Asn Thr Thr	Val Ser Ser Arg His Arg Ala
	355	360	365
Arg Pro Leu	Thr Leu Gly	Thr Ser Thr	Gln Ile Gly Ala Ile His Thr
370		375	380
Glu Gly Ile	Pro Ser Leu	Leu Lys Val	Leu Leu Ser Pro Ser Cys Asn
385		390	395
Gly Leu Pro	Val Leu Tyr	Ile Arg Asn	Leu Leu Leu Asn Pro Pro Ser
	405	410	415

-continued

Tyr Glu Ile Ala Ser Lys Ile Gln Glu Thr Cys Lys Leu Met Ser Ser
 420 425 430
 Leu Thr Cys Ser Ile Pro Glu Phe Thr Cys Val Ser Ser Ala Lys Leu
 435 440 445
 Val Lys Leu Leu Glu Trp Arg Glu Val Asn His Met Glu Phe Cys Arg
 450 455 460
 Ile Lys Asn Val Leu Asp Glu Ile Leu His Met Tyr Lys Thr Ser Glu
 465 470 475 480
 Leu Asn Glu Ile Leu Lys Asn Leu Ile Asp Pro Thr Trp Ala Thr Thr
 485 490 495
 Gly Leu Asp Ile Asp Phe Glu Thr Leu Val Ser Gly Cys Glu Val Ala
 500 505 510
 Ser Ser Lys Ile Ser Glu Ile Ile Ser Leu Asp Gly Gly Asn Asp Gln
 515 520 525
 Lys Ile Asn Ser Leu Ser Ile Ile Pro Tyr Glu Phe Phe Glu Asp Thr
 530 535 540
 Glu Ser Lys Trp Lys Gly Arg Ile Lys Arg Val His Ile Asp Glu Val
 545 550 555 560
 Phe Thr Ala Val Gln Lys Ala Ala Glu Val Leu His Ile Ala Val Thr
 565 570 575
 Glu Asp Phe Val Pro Val Val Ser Arg Val Lys Ala Thr Ile Ala Pro
 580 585 590
 Leu Gly Gly Pro Arg Gly Glu Ile Ser Tyr Ala Arg Glu His Glu Ala
 595 600 605
 Val Trp Phe Arg Gly Lys Arg Phe Thr Pro Ser Leu Trp Ser Gly Ser
 610 615 620
 Pro Gly Glu Glu Gln Ile Lys Gln Leu Arg His Ala Leu Asp Ser Lys
 625 630 635 640
 Gly Lys Arg Val Gly Glu Glu Trp Phe Thr Thr Pro Lys Val Glu Ala
 645 650 655
 Ala Leu Thr Arg Tyr His Glu Ala Asn Ala Lys Ala Thr Glu Arg Val
 660 665 670
 Leu Glu Ile Leu Arg Glu Leu Ala Thr Glu Leu His Tyr Ser Ile Asn
 675 680 685
 Ile Leu Val Phe Ser Ser Thr Leu Leu Val Ile Thr Lys Ala Leu Phe
 690 695 700
 Ala His Ala Ser Glu Gly Arg Arg Arg Arg Trp Val Phe Pro Thr Leu
 705 710 715 720
 Ala Glu Ser Asn Gly Phe Glu Asp Val Lys Ser Ser Asp Lys Ile His
 725 730 735
 Gly Met Lys Ile Val Gly Leu Ala Pro Tyr Trp Phe His Ile Ala Glu
 740 745 750
 Gly Ile Val Arg Asn Asp Val Asp Met Gln Ser Leu Phe Leu Leu Thr
 755 760 765
 Gly Pro Asn Gly Gly Gly Lys Ser Ser Leu Leu Arg Ser Ile Cys Ala
 770 775 780
 Ala Ala Leu Leu Gly Ile Cys Gly Leu Met Val Pro Ala Glu Ser Ala
 785 790 795 800
 Val Ile Pro Tyr Phe Asp Ser Ile Thr Leu His Met Lys Ser Tyr Asp
 805 810 815

-continued

Ser Pro Ala Asp Lys Lys Ser Ser Phe Gln Val Glu Met Ser Glu Leu
 820 825 830
 Arg Ser Ile Ile Gly Gly Thr Thr Lys Arg Ser Leu Val Leu Val Asp
 835 840 845
 Glu Ile Cys Arg Gly Thr Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly
 850 855 860
 Ser Ile Ile Glu Thr Leu Glu Arg Ile Gly Cys Leu Gly Val Val Ser
 865 870 875 880
 Thr His Leu His Gly Ile Phe Thr Leu Pro Leu Asn Ile Lys Ser Thr
 885 890 895
 Val His Lys Ala Met Gly Thr Thr Cys Ile Asp Gly Gln Ile Leu Pro
 900 905 910
 Thr Trp Lys Leu Thr Asp Gly Val Cys Lys Glu Ser Leu Ala Phe Glu
 915 920 925
 Thr Ala Ile Arg Glu Gly Ile Pro Glu Pro Ile Ile Arg Arg Ala Glu
 930 935 940
 Cys Leu Tyr Lys Ser Val Tyr Ala Glu Glu Asn Phe Pro Asn Glu Glu
 945 950 955 960
 Lys Phe Ser Thr Cys Asn Asn Leu Asn Asn Leu Asn Thr Thr Ser Leu
 965 970 975
 Tyr Ser Lys Gly Phe Leu Ser Gly Ala Asn Gln Met Glu Gly Phe Arg
 980 985 990
 Gln Glu Val Glu Arg Ala Ile Thr Val Ile Cys Gln Asp Tyr Ile Met
 995 1000 1005
 Glu Arg Lys Asn Lys Lys Ile Ala Leu Glu Leu Pro Glu Ile Lys
 1010 1015 1020
 Cys Leu Leu Ile Gly Lys Arg Glu Gln Pro Pro Pro Ser Val Val
 1025 1030 1035
 Gly Ser Ser Ser Val Tyr Val Ile Phe Thr Pro Asp Lys Lys Leu
 1040 1045 1050
 Tyr Val Gly Glu Thr Asp Asp Leu Glu Gly Arg Val Arg Arg His
 1055 1060 1065
 Arg Leu Lys Glu Gly Met Asp Glu Ala Ser Phe Leu Tyr Phe Leu
 1070 1075 1080
 Val Pro Gly Lys Ser Leu Ala Cys Gln Phe Glu Ser Leu Leu Ile
 1085 1090 1095
 Asn Gln Leu Ser Ser Gln Gly Phe Gln Leu Ser Asn Met Ala Asp
 1100 1105 1110
 Gly Lys His Arg Asn Phe Gly Thr Ser Asn Leu Tyr Ala
 1115 1120 1125

<210> SEQ ID NO 48
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: primer

<400> SEQUENCE: 48

ggccatgggtg tgaattgcat agtcgtcg

28

<210> SEQ ID NO 49
 <211> LENGTH: 28
 <212> TYPE: DNA

-continued

<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 49

ggccatggaa acatcacttg acgtcttc

28

<210> SEQ ID NO 50
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Arabadopsis thaliana

<400> SEQUENCE: 50

agtggttggt tgggt

15

<210> SEQ ID NO 51
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Arabadopsis thaliana

<400> SEQUENCE: 51

agtggttatt tgggt

15

<210> SEQ ID NO 52
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Arabadopsis thaliana

<400> SEQUENCE: 52

gatgttgtag tttaa

15

<210> SEQ ID NO 53
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Arabadopsis thaliana

<400> SEQUENCE: 53

gatgttgtag tttaa

15

<210> SEQ ID NO 54
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Arabadopsis thaliana

<400> SEQUENCE: 54

tactcagaga ttg

13

<210> SEQ ID NO 55
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Arabadopsis thaliana

<400> SEQUENCE: 55

tactcaaaga ttg

13

<210> SEQ ID NO 56
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 56

-continued

Leu Leu Phe Tyr Arg Met Gly Asp Phe Tyr Glu Leu Phe Tyr Asp Asp
1 5 10 15

Ala

<210> SEQ ID NO 57
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 57

Val Val Leu Thr Gln Met Gly Ser Phe Tyr Glu Leu Tyr Phe Glu Gln
1 5 10 15

Ala

<210> SEQ ID NO 58
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: *Arabidopsis thaliana*

<400> SEQUENCE: 58

Val Val Phe Phe Lys Met Ala Lys Phe Tyr Glu Leu Phe Glu Met Asp
1 5 10 15

Ala

<210> SEQ ID NO 59
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: *Arabidopsis thaliana*

<400> SEQUENCE: 59

Val Leu Leu Cys Arg Val Gly Glu Phe Tyr Glu Ala Ile Gly Ile Asp
1 5 10 15

Ala

<210> SEQ ID NO 60
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: consensus

<400> SEQUENCE: 60

Leu Leu Phe Tyr Arg Met Gly Asp Phe Tyr Glu Leu Phe Tyr Asp Asp
1 5 10 15

Ala

<210> SEQ ID NO 61
<211> LENGTH: 173
<212> TYPE: PRT
<213> ORGANISM: *Escherichia coli*

<400> SEQUENCE: 61

Asp Lys Pro Gly Ile Arg Ile Thr Glu Gly Arg His Pro Val Val Glu
1 5 10 15

Gln Val Leu Asn Glu Pro Phe Ile Ala Asn Pro Leu Asn Asn Ser Pro
20 25 30

Gln Arg Arg Met Leu Ile Ile Thr Gly Pro Asn Met Gly Gly Lys Ser
35 40 45

-continued

Thr Tyr Met Arg Gln Thr Ala Leu Ile Ala Leu Met Ala Tyr Ile Gly
 50 55 60
 Ser Tyr Val Pro Ala Gln Lys Val Glu Ile Gly Pro Ile Asp Arg Ile
 65 70 75 80
 Phe Thr Arg Val Gly Ala Ala Asp Asp Leu Ala Ser Gly Arg Ser Thr
 85 90 95
 Phe Met Val Glu Met Thr Glu Thr Ala Asn Ile Leu His Asn Ala Thr
 100 105 110
 Glu Tyr Ser Leu Val Leu Met Asp Glu Ile Gly Arg Gly Thr Ser Thr
 115 120 125
 Tyr Asp Gly Leu Ser Leu Ala Trp Cys Ala Glu Asn Leu Ala Asn Lys
 130 135 140
 Ile Lys Ala Leu Thr Leu Phe Ala Thr His Tyr Phe Glu Leu Thr Gln
 145 150 155 160
 Leu Pro Glu Lys Met Glu Gly Glx Val Ala Asn Val His
 165 170

<210> SEQ ID NO 62
 <211> LENGTH: 177
 <212> TYPE: PRT
 <213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 62

Glu Ser Asn Lys Leu Glu Val Val Asn Gly Arg His Leu Met Val Glu
 1 5 10 15
 Glu Gly Leu Ser Ala Arg Ser Leu Glu Thr Phe Thr Ala Asn Asn Cys
 20 25 30
 Glu Leu Ala Lys Asp Asn Leu Trp Val Ile Thr Gly Pro Asn Met Gly
 35 40 45
 Gly Lys Ser Thr Phe Leu Arg Gln Asn Ala Ile Ile Val Ile Leu Ala
 50 55 60
 Gln Ile Gly Cys Phe Val Pro Cys Ser Lys Ala Arg Val Gly Ile Val
 65 70 75 80
 Asp Lys Leu Phe Ser Arg Val Gly Ser Ala Asp Asp Leu Tyr Asn Glu
 85 90 95
 Met Ser Thr Phe Met Val Glx Glu Met Ile Glu Thr Ser Phe Ile Leu
 100 105 110
 Gln Gly Ala Thr Glu Arg Ser Leu Ala Ile Leu Asp Glu Ile Gly Arg
 115 120 125
 Gly Thr Ser Gly Lys Glu Gly Ile Ser Ile Ala Tyr Ala Thr Leu Lys
 130 135 140
 Tyr Leu Leu Glu Asn Asn Gln Cys Arg Thr Leu Phe Ala Thr His Phe
 145 150 155 160
 Gly Gln Glu Leu Lys Gln Ile Asp Asn Lys Cys Ser Lys Gly Met Ser
 165 170 175

Glu

<210> SEQ ID NO 63
 <211> LENGTH: 177
 <212> TYPE: PRT
 <213> ORGANISM: *Arabidopsis thaliana*

<400> SEQUENCE: 63

Gly Val Pro His Leu Ser Ala Thr Gly Leu Gly His Pro Val Leu Arg

-continued

1	5	10	15
Gly Asp Ser	Leu Gly Arg Gly Ser Phe Val Pro Asn Asn Val Lys Ile		
	20	25	30
Gly Gly Ala	Glu Lys Ala Ser Phe Ile Leu Leu Thr Gly Pro Asn Met		
	35	40	45
Gly Gly Lys	Ser Thr Leu Leu Arg Gln Val Cys Leu Ala Val Ile Leu		
	50	55	60
Ala Gln Ile	Gly Ala Asp Val Pro Ala Glu Thr Phe Glu Val Ser Pro		
	65	70	75
Val Asp Lys	Ile Cys Val Arg Met Gly Ala Lys Asp His Ile Met Ala		
	85	90	95
Gly Gln Ser	Thr Phe Leu Thr Glu Leu Ser Glu Thr Ala Val Met Leu		
	100	105	110
Thr Ser Ala	Thr Arg Asn Ser Leu Val Val Leu Asp Glu Leu Gly Arg		
	115	120	125
Gly Thr Ala	Thr Ser Asp Gly Gln Ala Ile Ala Glu Ser Val Leu Glu		
	130	135	140
His Phe Ile	Glu Lys Val Gln Cys Arg Gly Phe Phe Ser Thr His Tyr		
	145	150	155
His Arg Leu	Ser Val Asp Tyr Gln Thr Asn Pro Lys Val Ser Leu Cys		
	165	170	175

His

<210> SEQ ID NO 64

<211> LENGTH: 177

<212> TYPE: PRT

<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 64

Leu Asp Glu	Gly Ala Lys Pro Leu Asp Gly Ala Ser Arg Met Lys Leu
1	5 10 15
Thr Gly Leu	Ser Pro Tyr Trp Phe Asp Val Ser Ser Gly Thr Ala Val
	20 25 30
His Asn Thr	Val Asp Met Gln Ser Leu Phe Leu Leu Thr Gly Pro Asn
	35 40 45
Gly Gly Gly	Lys Ser Ser Leu Leu Arg Ser Ile Cys Ala Ala Ala Leu
	50 55 60
Leu Gly Ile	Ser Gly Leu Met Val Pro Ala Glu Ser Ala Cys Ile Pro
	65 70 75 80
His Phe Asp	Ser Ile Met Leu His Met Lys Ser Tyr Asp Ser Pro Val
	85 90 95
Asp Gly Lys	Ser Ser Phe Gln Val Glu Met Ser Glu Ile Arg Ser Ile
	100 105 110
Val Ser Gln	Ala Thr Ser Arg Ser Leu Val Leu Ile Asp Glu Ile Cys
	115 120 125
Arg Gly Thr	Glu Thr Ala Lys Gly Thr Cys Ile Ala Gly Ser Val Val
	130 135 140
Glu Ser Leu	Asp Thr Ser Gly Cys Leu Gly Ile Val Ser Thr His Leu
	145 150 155 160
His Gly Ile	Phe Ser Leu Pro Leu Thr Ala Lys Asn Ile Thr Tyr Lys
	165 170 175

Ala

-continued

<210> SEQ ID NO 65
<211> LENGTH: 1558
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: consensus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (421)..(422)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (427)..(427)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (435)..(435)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (440)..(443)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (445)..(446)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (450)..(451)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (453)..(453)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (455)..(456)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (460)..(460)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (462)..(462)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (466)..(466)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (468)..(468)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (470)..(471)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (476)..(477)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (479)..(480)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (482)..(483)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (485)..(486)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid

-continued

```
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (492)..(492)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (494)..(495)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (514)..(514)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (525)..(525)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (528)..(528)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (641)..(641)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (660)..(660)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (697)..(697)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (751)..(751)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (765)..(765)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (787)..(788)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (820)..(820)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (822)..(822)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (876)..(876)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (896)..(896)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (900)..(900)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (904)..(904)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (909)..(909)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (932)..(933)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (936)..(936)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (949)..(949)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (956)..(958)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (961)..(961)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (970)..(970)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (983)..(983)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (986)..(986)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1079)..(1079)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1088)..(1088)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1093)..(1093)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1108)..(1108)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1144)..(1149)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1152)..(1152)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1157)..(1157)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1159)..(1159)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1175)..(1175)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1177)..(1177)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1180)..(1180)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1261)..(1261)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1296)..(1296)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1316)..(1316)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1329)..(1329)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1362)..(1362)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1369)..(1369)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1372)..(1372)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1377)..(1388)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1391)..(1391)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1393)..(1394)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1398)..(1398)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1401)..(1401)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1406)..(1419)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1421)..(1421)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1423)..(1423)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1438)..(1438)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1445)..(1446)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1449)..(1451)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1454)..(1454)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1457)..(1457)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
```

-continued

```

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1481)..(1481)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1509)..(1509)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1536)..(1537)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1557)..(1557)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid

```

```

<400> SEQUENCE: 65

```

```

Ser Tyr Ile Arg Lys Arg Ser Ser Lys Lys Leu Lys Pro Val Leu Asp
1          5          10          15
Asp Lys Asp Leu Pro His Ile Leu Trp Trp Lys Glu Arg Leu Gln Cys
20         25         30
Arg Lys Pro Ser Thr Val Gln Leu Ile Arg Leu Tyr Ser Asn Leu Leu
35         40         45
Gly Leu Asp Pro Ser Leu Arg Asn Gly Ser Leu Lys Glu Gly Thr Leu
50         55         60
Asn Trp Glu Met Leu Gln Phe Lys Ser Lys Phe Pro Arg Glu Val Leu
65         70         75         80
Leu Cys Arg Val Gly Glu Phe Tyr Glu Ala Ile Gly Ile Asp Ala Cys
85         90         95
Ile Leu Val Glu Tyr Ala Gly Leu Asn Pro Phe Gly Gly Leu Arg Ser
100        105        110
Asp Ser Ile Pro Lys Ala Gly Cys Pro Val Val Asn Leu Arg Gln Thr
115        120        125
Leu Asp Asp Leu Thr Arg Asn Gly Tyr Ser Val Cys Ile Val Glu Glu
130        135        140
Val Gln Gly Pro Thr Gln Ala Arg Ser Arg Lys Arg Phe Ile Ser Gly
145        150        155        160
His Ala His Pro Gly Ser Pro Tyr Val Tyr Gly Leu Ala Val Asp His
165        170        175
Asp Leu Asp Phe Pro Glu Pro Met Pro Val Val Gly Ile Ser Arg Ser
180        185        190
Ala Arg Gly Tyr Cys Ile Ile Ser Val Leu Glu Thr Met Lys Thr Tyr
195        200        205
Ser Glu Asp Gly Leu Thr Glu Glu Ala Val Val Thr Lys Leu Arg Thr
210        215        220
Cys Arg Tyr His His Leu Phe Leu His Thr Ser Leu Arg Asn Asn Ser
225        230        235        240
Ser Gly Thr Ser Arg Trp Gly Glu Phe Gly Glu Gly Gly Leu Leu Trp
245        250        255
Gly Glu Cys Ser Ser Arg Phe Glu Trp Phe Asp Gly Asn Pro Ile Ser
260        265        270
Glu Leu Leu Lys Val Lys Glu Leu Tyr Gly Leu Asp Asp Glu Val Thr
275        280        285
Phe Arg Asn Val Thr Val Ser Ser Arg Pro Arg Pro Leu His Leu Gly
290        295        300

```

-continued

Thr	Ala	Thr	Gln	Ile	Gly	Ala	Ile	Pro	Thr	Glu	Gly	Ile	Pro	Ser	Leu
305					310					315					320
Leu	Lys	Val	Leu	Leu	Pro	Pro	Cys	Gly	Leu	Pro	Val	Leu	Tyr	Ile	Arg
				325					330					335	
Asp	Leu	Leu	Leu	Asn	Pro	Pro	Ser	Tyr	Glu	Ile	Ala	Ser	Lys	Ile	Gln
			340					345					350		
Glu	Thr	Cys	Lys	Leu	Met	Ser	Ser	Val	Thr	Cys	Ser	Ile	Pro	Glu	Phe
		355					360					365			
Thr	Cys	Val	Ser	Ser	Ala	Lys	Leu	Val	Lys	Leu	Leu	Glu	Arg	Glu	Val
		370				375						380			
Asn	His	Ile	Glu	Phe	Cys	Arg	Ile	Lys	Asn	Val	Leu	Asp	Glu	Ile	Leu
385					390					395					400
Met	Tyr	Arg	Ser	Glu	Leu	Glu	Ile	Leu	Lys	Leu	Ile	Asp	Pro	Thr	Trp
				405					410					415	
Val	Ala	Thr	Gly	Xaa	Xaa	Met	Tyr	Arg	Val	Xaa	Thr	Arg	Asn	Val	Val
			420					425					430		
Val	Ser	Xaa	Pro	Arg	Trp	Arg	Xaa	Xaa	Xaa	Xaa	Phe	Xaa	Xaa	Ser	Ser
		435					440					445			
Phe	Xaa	Xaa	Phe	Xaa	Ser	Xaa	Xaa	Pro	Ser	Arg	Xaa	Leu	Xaa	Ile	Asn
		450				455					460				
Gly	Xaa	Val	Xaa	Asn	Xaa	Xaa	Ser	Tyr	Ile	Arg	Xaa	Xaa	Lys	Xaa	Xaa
465				470						475					480
Arg	Xaa	Xaa	Ser	Xaa	Xaa	Ser	Lys	Lys	Leu	Lys	Xaa	Pro	Xaa	Xaa	Val
			485						490					495	
Leu	Asp	Asp	Lys	Asp	Leu	Pro	His	Ile	Leu	Trp	Trp	Lys	Glu	Arg	Leu
			500					505					510		
Gln	Xaa	Cys	Arg	Lys	Pro	Ser	Thr	Val	Gln	Leu	Ile	Xaa	Arg	Leu	Xaa
		515					520					525			
Tyr	Ser	Asn	Leu	Leu	Gly	Leu	Asp	Pro	Ser	Leu	Arg	Asn	Gly	Ser	Leu
		530				535					540				
Lys	Glu	Gly	Thr	Leu	Asn	Trp	Glu	Met	Leu	Gln	Phe	Lys	Ser	Lys	Phe
545					550					555					560
Pro	Arg	Glu	Val	Leu	Leu	Cys	Arg	Val	Gly	Glu	Phe	Tyr	Glu	Ala	Ile
			565						570					575	
Gly	Ile	Asp	Ala	Cys	Ile	Leu	Val	Glu	Tyr	Ala	Gly	Leu	Asn	Pro	Phe
			580					585					590		
Gly	Gly	Leu	Arg	Ser	Asp	Ser	Ile	Pro	Lys	Ala	Gly	Cys	Pro	Val	Val
		595					600					605			
Asn	Leu	Arg	Gln	Thr	Leu	Asp	Asp	Leu	Thr	Arg	Asn	Gly	Tyr	Ser	Val
		610				615					620				
Cys	Ile	Val	Glu	Glu	Val	Gln	Gly	Pro	Thr	Gln	Ala	Arg	Ser	Arg	Lys
625					630					635					640
Xaa	Arg	Phe	Ile	Ser	Gly	His	Ala	His	Pro	Gly	Ser	Pro	Tyr	Val	Tyr
			645					650						655	
Gly	Leu	Ala	Xaa	Val	Asp	His	Asp	Leu	Asp	Phe	Pro	Glu	Pro	Met	Pro
			660				665						670		
Val	Val	Gly	Ile	Ser	Arg	Ser	Ala	Arg	Gly	Tyr	Cys	Ile	Ile	Ser	Val
		675					680					685			
Leu	Glu	Thr	Met	Lys	Thr	Tyr	Ser	Xaa	Glu	Asp	Gly	Leu	Thr	Glu	Glu
	690					695					700				
Ala	Val	Val	Thr	Lys	Leu	Arg	Thr	Cys	Arg	Tyr	His	His	Leu	Phe	Leu

-continued

705	710	715	720
His Thr Ser Leu Arg 725	Asn Asn Ser Ser 730	Gly Thr Ser Arg Trp 735	Gly Glu
Phe Gly Glu Gly Gly 740	Leu Leu Trp Gly 745	Glu Cys Ser Ser Arg 750	Xaa Phe
Glu Trp Phe Asp Gly 755	Asn Pro Ile Ser 760	Glu Leu Leu Xaa 765	Lys Val Lys
Glu Leu Tyr Gly Leu 770	Asp Asp Glu Val Thr 775	Phe Arg Asn Val Thr 780	Val
Ser Ser Xaa Xaa Arg 785	Pro Arg Pro Leu His 790	Leu Gly Thr Ala Thr 795	Gln 800
Ile Gly Ala Ile 805	Pro Thr Glu Gly Ile 810	Ser Leu Leu Lys Val 815	Leu
Leu Pro Pro Xaa Cys 820	Xaa Gly Leu Pro Val 825	Leu Tyr Ile Arg Asp 830	Leu
Leu Leu Asn Pro Pro 835	Ser Tyr Glu Ile Ala 840	Ser Lys Ile Gln Glu Thr 845	
Cys Lys Leu Met Ser 850	Ser Val Thr Cys Ser 855	Ile Pro Glu Phe Thr Cys 860	
Val Ser Ser Ala Lys 865	Leu Val Lys Leu Leu 870	Glu Xaa Arg Glu Val Asn 875	880
His Ile Glu Phe Cys 885	Arg Ile Lys Asn Val 890	Leu Asp Glu Ile Leu Xaa 895	
Met Tyr Arg Xaa Ser 900	Glu Leu Xaa Glu Ile 905	Leu Lys Xaa Leu Ile Asp 910	
Pro Thr Trp Val Ala 915	Thr Gly Leu Lys Ile 920	Asp Phe Asp Thr Leu Val 925	
Asn Glu Cys Xaa Xaa 930	Ala Ser Xaa Lys Ile 935	Ser Glu Ile Ile Ser Leu 940	
Asp Gly Glu Asn Xaa 945	Asp Gln Lys Ile Ser 950	Ser Xaa Xaa Xaa Ile Pro 955	960
Xaa Glu Phe Phe Glu 965	Asp Met Glu Ser Xaa 970	Trp Lys Gly Arg Val Lys 975	
Arg Ile His Ile Glu 980	Glu Xaa Phe Thr Xaa 985	Val Glu Lys Ala Ala Glu 990	
Ala Leu Ser Ile Ala 995	Val Thr Glu Asp Phe 1000	Leu Pro Ile Ile Ser Arg 1005	
Ile Lys Ala Thr Met 1010	Ala Pro Leu Gly Gly 1015	Pro Lys Gly Glu Ile 1020	
Ser Tyr Ala Arg Glu 1025	His Glu Ala Val Trp 1030	Phe Lys Gly Lys Arg 1035	
Phe Thr Pro Ser Leu 1040	Trp Ala Gly Thr Pro 1045	Gly Glu Glu Gln Ile 1050	
Lys Gln Leu Arg Pro 1055	Ala Leu Asp Ser Lys 1060	Gly Lys Lys Val Gly 1065	
Glu Glu Trp Phe Thr 1070	Thr Pro Lys Val Glu 1075	Xaa Ala Leu Thr Arg 1080	
Tyr His Glu Ala Xaa 1085	Ala Lys Xaa Arg Val 1090	Leu Glu Leu 1095	
Leu Arg Gly Leu Ser 1100	Ser Glu Leu Gln Xaa 1105	Lys Ile Asn Ile Leu 1110	

-continued

Val Phe Ala Ser Met Leu Leu	Val Ile Thr Lys Ala	Leu Phe Ala
1115	1120	1125
His Ala Ser Glu Gly Arg Arg	Arg Arg Trp Val Phe	Pro Thr Leu
1130	1135	1140
Xaa Xaa Xaa Xaa Xaa Glu	Asp Xaa Lys Ser Leu	Asp Xaa Thr
1145	1150	1155
Xaa Gly Met Lys Ile Ser Gly	Leu Ser Pro Tyr Trp	Phe Asp Ile
1160	1165	1170
Ala Xaa Gly Xaa Ala Val Xaa	Asn Asp Val Asp Met	Gln Ser Leu
1175	1180	1185
Phe Leu Leu Thr Gly Pro Asn	Gly Gly Gly Lys Ser	Ser Leu Leu
1190	1195	1200
Arg Ser Ile Cys Ala Ala Ala	Leu Leu Gly Ile Cys	Gly Leu Met
1205	1210	1215
Val Pro Ala Glu Ser Ala Val	Ile Pro His Phe Asp	Ser Ile Met
1220	1225	1230
Leu His Met Lys Ser Tyr Asp	Ser Pro Ala Asp Gly	Lys Ser Ser
1235	1240	1245
Phe Gln Val Glu Met Ser Glu	Ile Arg Ser Ile Ile	Xaa Gly Ala
1250	1255	1260
Thr Ser Arg Ser Leu Val Leu	Ile Asp Glu Ile Cys	Arg Gly Thr
1265	1270	1275
Glu Thr Ala Lys Gly Thr Cys	Ile Ala Gly Ser Ile	Ile Glu Thr
1280	1285	1290
Leu Asp Xaa Ile Gly Cys Leu	Gly Ile Val Ser Thr	His Leu His
1295	1300	1305
Gly Ile Phe Thr Leu Pro Leu	Xaa Ile Lys Asn Thr	Val His Lys
1310	1315	1320
Ala Met Gly Thr Glu Xaa Ile	Asp Gly Gln Ile Ile	Pro Thr Trp
1325	1330	1335
Lys Leu Thr Asp Gly Val Cys	Lys Glu Ser Leu Ala	Phe Glu Thr
1340	1345	1350
Ala Lys Arg Glu Gly Ile Pro	Glu Xaa Ile Ile Arg	Arg Ala Glu
1355	1360	1365
Xaa Leu Tyr Xaa Ser Val Tyr	Ala Xaa Xaa Xaa Xaa	Xaa Xaa Xaa
1370	1375	1380
Xaa Xaa Xaa Xaa Xaa Glu Lys	Xaa Ser Xaa Xaa Ile	Asn Ile Xaa
1385	1390	1395
Asn Leu Xaa Thr Thr Ser Leu	Xaa Xaa Xaa Xaa Xaa	Xaa Xaa Xaa
1400	1405	1410
Xaa Xaa Xaa Xaa Xaa Ala	Xaa Met Xaa Ile Leu	Arg Lys Glu
1415	1420	1425
Leu Glu Arg Ala Ile Thr Val	Ile Cys Xaa Lys Lys	Ile Ile Glu
1430	1435	1440
Leu Xaa Xaa Lys Lys Xaa Xaa	Xaa Glu Leu Xaa Glu	Ile Xaa Cys
1445	1450	1455

-continued

Leu	Leu	Ile	Gly	Ala	Arg	Glu	Gln	Pro	Pro	Pro	Ser	Thr	Val	Gly
1460						1465					1470			
Ser	Ser	Ser	Val	Tyr	Val	Met	Xaa	Arg	Pro	Asp	Lys	Lys	Leu	Tyr
1475						1480					1485			
Val	Gly	Gln	Thr	Asp	Asp	Leu	Glu	Gly	Arg	Val	Arg	Ala	His	Arg
1490						1495					1500			
Leu	Lys	Glu	Gly	Met	Xaa	Asp	Ala	Ser	Phe	Leu	Tyr	Phe	Leu	Val
1505						1510					1515			
Pro	Gly	Lys	Ser	Ile	Ala	Cys	Gln	Leu	Glu	Thr	Leu	Leu	Ile	Asn
1520						1525					1530			
Gln	Leu	Xaa	Xaa	Gln	Gly	Phe	Gln	Leu	Ser	Asn	Ile	Ala	Asp	Gly
1535						1540					1545			
Lys	His	Arg	Asn	Phe	Gly	Thr	Ser	Xaa	Leu					
1550						1555								

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

- (a) a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and SEQ ID NO:45;
- (b) a nucleic acid molecule comprising at least a portion of any of said nucleic acid molecules of (a);
- (c) a complement of a of a nucleic acid molecule of (a) or (b); and
- (d) a nucleic acid molecule comprising an allelic variant of a nucleic acid molecule comprising any of said nucleic acid sequences.

2. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule is a plant nucleic acid molecule.

3. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule is selected from the group consisting of *Arabidopsis*, *Oryza*, *Glycine*, *Hordeum*, *Zea*, *Medicago*, *Allium*, *Citrus*, *Solanum*, *Sorghum*, *Saccharum*, *Nicotiana*, *Lycopersicon*, *Triticum*, *Zinnia*, and *Phaseolus* nucleic acid molecules.

4. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule is selected from the group consisting of: a nucleic acid molecule comprising a nucleic acid sequence that encodes a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47, and SEQ ID NO:65; and a nucleic acid molecule comprising an allelic variant of a nucleic acid molecule encoding a protein having any of said amino acid sequences.

5. An isolated protein encoded by a plant MSH1 nucleic acid molecule that hybridizes to the complement of a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and SEQ ID NO:45 under stringent hybridization conditions.

6. An isolated protein comprising a plant MSH1 protein.

7. The protein of claim 5, wherein said protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47 and SEQ ID NO:65.

8. The protein of claim 5, wherein said protein comprises at least a portion of an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:47 and SEQ ID NO:65.

9. A method to identify a compound capable of inhibiting MSH1 activity of a plant, said method comprising:

- (a) contacting an isolated plant MSH1 nucleic acid molecule selected from the group consisting of SEQ ID NO:1, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:41, SEQ ID NO:43, and SEQ ID NO:45 with a putative inhibitory compound which, in the absence of said compound, said plant MSH1

nucleic acid molecule has the activity of suppressing ectopic recombination; and

- (b) determining if said putative inhibitory compound inhibits said activity.

10. The method of claim 9, wherein the putative inhibitory compound is a RNA molecule suspected of having RNAi activity.

11. A compound identified by the method of claim 9.

12. A method for identification of plant mutants arising from mitochondrial ectopic recombination comprising

- (a) providing a plant,

- (b) suppressing expression of an MSH1-homologous gene in the plant, and

- (c) detecting an aberrant phenotype,

whereby a plant mutant is identified.

13. The method of claim 12, wherein said suppressing expression of an MSH1-homologous gene in said plant comprises contacting said plant with an compound identified by the method of claim 9.

14. The method of claim 12, wherein said aberrant phenotype is cytoplasmic male sterility.

15. A plant mutant identified by the method of claim 12.

* * * * *