Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2766097 C 2021/09/14

(11)(21) 2 766 097

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2010/06/30
(87) Date publication PCT/PCT Publication Date: 2011/01/06
(45) Date de délivrance/Issue Date: 2021/09/14
(85) Entrée phase nationale/National Entry: 2011/12/20
(86) N° demande PCT/PCT Application No.: IB 2010/001766
(87) N° publication PCT/PCT Publication No.: 2011/001286
(30) Priorité/Priority: 2009/06/30 (US61/221,813)

(51) Cl.Int./Int.Cl. *C12N 15/82* (2006.01), *C12N 15/00* (2006.01), *C12N 15/29* (2006.01), *C07K 14/415* (2006.01)

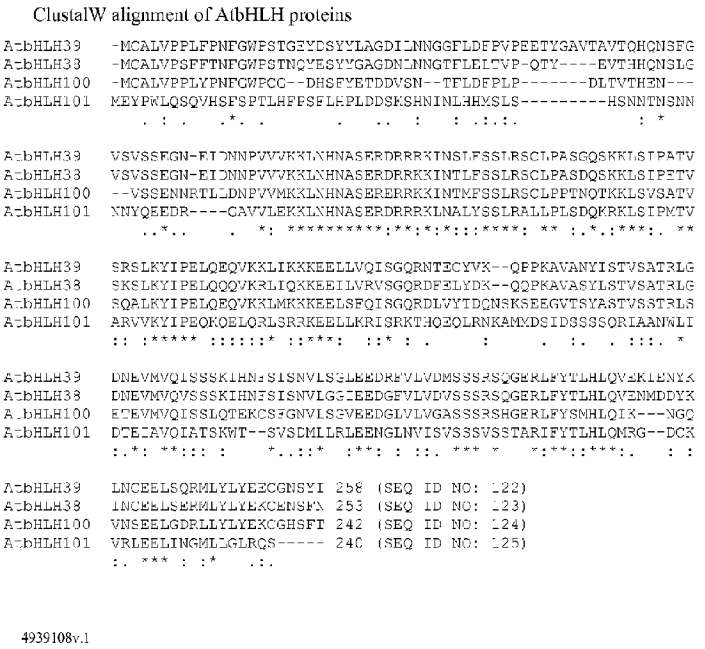
(72) Inventeurs/Inventors:
WAN, JIANGXIN, CA;
HUANG, YAFAN, CA

(73) Propriétaire/Owner:
PERFORMANCE PLANTS, INC., CA

(74) Agent: RIDOUT & MAYBEE LLP

(54) Titre : FACTEURS DE TRANSCRIPTION DU SOUS-GROUPE 1B A MOTIF BHLH QUI CONFERENT UNE TOLERANCE A LA CHALEUR

(54) Title: BHLH SUBGROUP 1B TRANSCRIPTION FACTORS THAT PROVIDE HEAT TOLERANCE



(57) Abrégé/Abstract:
The invention relates to methods of producing a desired phenotype in a plant by manipulation of gene expression within the plant. The method relates to means to increase expression level of a bHLH subgroup Ib gene expression or activity, wherein a desired phenotype such as increased heat tolerance relative to a wild type control plant following heat stress results in reduced flower abortion and increased yield. Included are plants produced by said methods. The invention also relates to nucleic acid sequences and constructs useful such methods and methods of generating and isolating plants having increased expression of a bHLH subgroup Ib expression or activity.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 January 2011 (06.01.2011)

PCT

(10) International Publication Number
WO 2011/001286 A3

(51) International Patent Classification:
C12N 15/82 (2006.01) *C12N 15/00* (2006.01)
A01H 5/00 (2006.01) *C12N 15/29* (2006.01)
A01H 5/10 (2006.01) *C07K 14/415* (2006.01)

(21) International Application Number:
PCT/IB2010/001766

(22) International Filing Date:
30 June 2010 (30.06.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/221,813 30 June 2009 (30.06.2009) US

(71) Applicant *(for all designated States except US)*: **PERFORMANCE PLANTS, INC.** [CA/CA]; 700 Gardiners Road, Kingston, Ontario K7M 3X9 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **WAN, Jiangxin** [CA/CA]; 376 Canton Road, R.R. #3, Bath, Ontario K0H 1G0 (CA). **HUANG, Yafan** [CA/CA]; 376 Canton Road, R.R. #3, Bath, Ontario K0H 1G0 (CA).

(74) Agent: **EVERITT, Peter, R.**; c/o Ridout & Maybee LLP, Suite 308, 1 City Centre Drive, Mississauga, Ontario L5B 1M2 (CA).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
17 March 2011

(54) Title: EXPRESSION OF TRANSCRIPTION REGULATORS THAT PROVIDE HEAT TOLERANCE

Figure 1

ClustalW alignment of AtbHLH proteins

```
AtbHLH39    -MCALVPPLFPNFGWPSTGEYDSYYLAGDILNNGGFLDFPVPEETYGAVTAVTQHQNSFG
AtbHLH38    -MCALVPSFFTNFGWPSTNQYESYYGAGDNLNNGTFLELTVP-QTY----EVTHHQNSLG
AtbHLH100   -MCALVPPLYPNFGWPCG--DHSFYETDDVSN--TFLDFPLP-------DLTVTHEN---
AtbHLH101   MEYPWLQSQVHSFSPTLHFPSFLHPLDDSKSHNINLHHMSLS---------HSNNTNSNN
               . : .   .*. .       .   .. :   : .:.:.           : *

AtbHLH39    VSVSSEGN-EIDNNPVVVKKLNHNASERDRRRKINSLFSSLRSCLPASGQSKKLSIPATV
AtbHLH38    VSVSSEGN-EIDNNPVVVKKLNHNASERDRRKKINTLFSSLRSCLPASDQSKKLSIPETV
AtbHLH100   --VSSENNRTLLDNPVVMKKLNHNASERERRKKINTMFSSLRSCLPPTNQTKKLSVSATV
AtbHLH101   NNYQEEDR----GAVVLEKKLNHNASERDRRRKLNALYSSLRALLPLSDQKRKLSIPMTV
               ..*..    .  *: **********:**:*:*:::****: ** :.*.:***:. **

AtbHLH39    SRSLKYIPELQEQVKKLIKKKEELLVQISGQRNTECYVK--QPPKAVANYISTVSATRLG
AtbHLH38    SKSLKYIPELQQQVKRLIQKKEEILVRVSGQRDFELYDK--QQPKAVASYLSTVSATRLG
AtbHLH100   SQALKYIPELQEQVKKLMKKKEELSFQISGQRDLVYTDQNSKSEEGVTSYASTVSSTRLS
AtbHLH101   ARVVKYIPEQKQELQRLSRRKEELLKRISRKTHQEQLRNKAMMDSIDSSSSQRIAANWLT
            :: :***** ::::::* ::***:  ::* : .     :     .  :.  . :::. *

AtbHLH39    DNEVMVQISSSKIHNFSISNVLSGLEEDRFVLVDMSSSRSQGERLFYTLHLQVEKIENYK
AtbHLH38    DNEVMVQVSSSKIHNFSISNVLGGIEEDGFVLVDVSSSRSQGERLFYTLHLQVENMDDYK
AtbHLH100   ETEVMVQISSLQTEKCSFGNVLSGVEEDGLVLVGASSSRSHGERLFYSMHLQIK---NGQ
AtbHLH101   DTEIAVQIATSKWT--SVSDMLLRLEENGLNVISVSSSVSSTARIFYTLHLQMRG--DCK
            :.*: **::: :    *..::*  :**: : ::. *** *   *:**::***:.   : :

AtbHLH39    LNCEELSQRMLYLYEECGNSYI 258 (SEQ ID NO: 122)
AtbHLH38    INCEELSERMLYLYEKCENSFN 253 (SEQ ID NO: 123)
AtbHLH100   VNSEELGDRLLYLYEKCGHSFT 242 (SEQ ID NO: 124)
AtbHLH101   VRLEELINGMLLGLRQS----- 240 (SEQ ID NO: 125)
            :. *** : :*   .:.
```

(57) Abstract: The invention relates to methods of producing a desired phenotype in a plant by manipulation of gene expression within the plant. The method relates to means to increase expression level of a bHLH subgroup Ib gene expression or activity, wherein a desired phenotype such as increased heat tolerance relative to a wild type control plant following heat stress results in reduced flower abortion and increased yield. Included are plants produced by said methods. The invention also relates to nucleic acid sequences and constructs useful such methods and methods of generating and isolating plants having increased expression of a bHLH subgroup Ib expression or activity.

WO 2011/001286 A3

bHLH SUBGROUP 1b TRANSCRIPTION FACTORS THAT PROVIDE HEAT TOLERANCE

REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application Serial No. 61/221,813, filed June 30, 2009.

FIELD OF THE INVENTION

The invention is in the field of plant molecular biology and relates to transgenic plants having novel phenotypes, methods of producing such plants and polynucleotides and polypeptides useful in such methods. More specifically, the invention relates to expression of a transcriptional regulator and transgenic plants having increased activity of the transcriptional regulator to produce a plant having a beneficial phenotype.

BACKGROUND OF THE INVENTION

Environmental stresses are responsible for significant yield increase in agricultural crops. The relationship between climate variation and production of corn and soybean throughout the United States for the period 1982-1998 was studied (Lobell and Asner, 2003) and found that even gradual temperature changes have a measurable impact on crop yield. In corn and soybean it has been estimated that yield is reduced by 17% per degree as the growth temperature rises above the season optimum. Both monocots and dicots are sensitive to heat stress, particularly during flowering and seed development which translates into a significant impact on seed yield (Young et al., 2004; Sato et al., 2002; Angadi et al., 2000; Carlson, 1990; Wahid, A., et al. 2007). In the field, heat stress is often accompanied with other environmental stresses as drought which further adds to the burden of plant productivity. Heat stress can have a myriad of cellular effects

on plants such as, altered membrane fluidity and permeability, protein aggregation and protein denaturation, the resulting cellular damage may result in deleterious changes in plant growth and development which impacts the ability to survive. It has been suggested that plants possess an inherent ability for basal and acquired thermotolerance and that a common heat response mechanisms is present in diverse plant species (Kapoor et al., 1990; Vierling, 1991; Flahaut et al., 1996; Burke et al., 2000; Hong and Vieling, 2000; Massie et al., 2003; Larkindale et al., 2005). A number of studies have been conducted to identify and characterize genes and pathways that are involved in plant thermotolerance. For example, heat shock transcription factors (HSF) and heat shock proteins (HSP) have received much attention to elucidate the roles and effects of these genes in response to heat stress as have plant growth hormones such as abscisic acid and ethylene. It is unclear how plants sense heat however, it is apparent that multiple signaling pathways and cellular components are involved (Larkindale et al. 2005, Plant Physiol. 138:882-897) and that cross talk of signaling pathways exists between environmental and nutritional stresses such as, heat shock stress, water stress/drought, cold stress, oxidative stress and heavy metal stress.

Transcription factors are DNA binding proteins that interact with specific promoter or enhancer sequences and alter the gene expression of the associated gene. Where the specific sequence that binds the transcription factor is associated with a suite of genes whole pathways can be coordinately regulated with various component genes being simultaneously up-regulated or down-regulated. A transcription factors may coordinately alter a suite of genes in response to a stimulus such as an environmental stress, nutritional status or pathogen attack, for example, or can be a component of a signaling pathway, such as a hormone signaling pathway for example. Transcription factors posses a modular structure and are classified primarily on the basis of the DNA binding domain. Some transcriptional regulators are participants in multiple signal cascades. The pathway and downstream genes regulated may vary depending on the presence or absence of other regulators and pathway components. The transcriptional regulators interact as a network whereby the outcome is dependent on a multitude of interacting factors.

Transcriptional activation is primarily mediated through transcription factors that interact with enhancer and promoter elements. Binding of transcription factors to such DNA elements constitutes a crucial step in transcriptional initiation. Each transcription factor binds to its

specific binding sequence in a promoter and activates expression of the linked coding region through interactions with coactivators and/or proteins that are a part of the transcription complex.

The transcription factor bHLH39 is a member of one of the largest families of transcription factors in *Arabidopsis thaliana* comprised of as many as 162 proteins (Heim et al., 2003, Toledo-Ortiz et al., 2003, Bailey et al., 2003). This family of proteins is distinguished from other transcription factors by its basic helix-loop-helix domain (bHLH). Studies have shown the basic region to be critical for DNA-binding, while the hydrophobic helix-loop-helix region is required for homodimer and heterodimer formation. bHLH proteins can have multiple binding partners, and consequently modulate the expression level of a different subset genes depending on its current partner (Zhang et al., 2003). Functional studies have implicated bHLH proteins in a range of cellular processes such as root epidermal cell fate determination (Tominaga, R., et al, 2007), anthocyanin production (Ramsay et al., 2003), and light signaling (Martinez-Garica et al., 2000) and iron uptake (Ling et al., 2002).

The basic region of bHLH proteins, consisting of approximately 15-17 residues, is responsible for binding to cis elements in promoters of target genes. Seventy-five percent of all *Arabidopsis* bHLH proteins are predicted to bind to the core motif known as the E-box (5'CANNTG-3') (Toledo-Ortiz et al., 2003). Specificity for the E-box can be predicted according to the presence of two critical residues, glutamic acid-85 and arginine-88. bHLH39, and its three closest *Arabidopsis* homologues, bhLH38, bHLH100, and bHLH101 all contain the critical glutamic acid and arginine, and therefore are predicted to bind to this motif. The type of E-box binding can be further divided according to the binding preference of two central nucleotides. bHLH39 is predicted to bind to the G-box motif CACGTG according to the presence of two residues, arginine-89 and histadine-81. Arginine-89 and Histadine-81 are responsible for contacting the two central "CG" nucleotides in the G-box and stabilizing the interacting. These two critical residues are also found in the three other *Arabidopsis* bHLH39 homologues. The four critical residues are conserved in all 95 homologues except in two cases, the *Cicer arietinum* homologue has an arginine in place of the histadine-81, and a *Vitis vinifera* homologue has a valine substituted for the glutamic acid-85 (Fig. 1).

Further evidence supporting the binding specificity of bHLH39 comes from reports of the rice ortholog OsIRO2 which binds to 5'-CACGTGG-3' (Ogo et al., 2006). As is the case with OsIRO2, residues outside of the core binding motif most likely also affect binding specificity.

The bHLH proteins bind as dimers to their DNA targets, and binding partner specificity is coded in the Helix-loop-helix domain. Evidence from the crystal structure of an intact human Max-DNA complex showed residue leucine 99 to be critical for dimer formation (Brownlie et al, 1997). This residue is conserved across all bHLH39 homologues except a *Vitis vinifera* homologue. Other residues that show over 95% conservation across all bHLH39 homologues included Arginine-100, Tyrosine-124, Isoleucine-125, and Proline-126. The requirement for a proline at position 126 appears to be specific to bHLH39, bHLH38, bHLH100, and bHLH101, suggesting this residue is important in facilitating dimerization specificity for this group.

According to structural similarities outside of the DNA-binding domain, bHLH39 is similar to 10 other bHLH proteins, which together form subgroup 1b (Heim et al., 2003). The bHLH39 and its closest *Arabidopsis* homologue, bHLH38, show 79% similarity and are located in tandem on the genome, suggesting a recent evolutionary duplication. Other members of the subgroup1b include bHLH100 and bHLH101. The percent similarity between AtbHLH39 and three other AtbHLH subgroup 1b members (according to Clustal W alignment) are shown below in Table A.

Table A: Percent similarity between AtbHLH39, AtbHLH38, AtbHLH100 and AtbHLH101

	AtbHLH39	AtbHLH38	AtbHLH100	AtbHLH101
AtbHLH39	100	79	60	31
AtbHLH38		100	57	32
AtbHLH100			100	29%
AtbHLH101				100

Both bHLH39 and bHLH38 were identified in a search for downstream targets of a DNA binding with one finger (Dof) transcription factor, ocs-element binding factor (OBP3) which is inducible by salicylic acid. In this case bHLH39 was named ORG3 and bHLH38 was named ORG2. The bHLH39 and bHLH38 were shown to have co-regulated and enhanced expression in OBP3 overexpression lines and being down-regulated in OBF3 loss-of-function lines (Kang et al., 2003). Subsequently, bHLH39 and bHLH38 were shown to be responsive to iron deficiency-mediated stress, along with their two closest homologues in bHLH subgroup 1b, bHLH100, and bHLH101. The role of bHLH39 and bHLH38 expression in relation to iron deficiency has been studied. A transcriptional regulator FIT (AtbHLH29) has been shown to interact with bHLH28

or bHLH39 to transcriptionally regulate the iron uptake genes FRO2 and IRT1. The overexpression of either AtbHLH38 or AtbHLH39 with FIT was found to alter the expression pattern of FRO2 and IRT1 to constitutive activation and result in iron deficiency tolerance (Yuan, et al., 2008). FRO2, a ferric chelate reductase, is responsible for the increase of iron in the soil, a process required to increase the solubility and bioavailability of iron. The iron is subsequently transported across the membrane by the iron transporter, IRT1. Ectopic expression of either AtbHLH38 or AtbHLH39 protein under the control of the 35S promoter in tobacco plants leads to the synthesis and excretion of riboflavin, a known defense mechanism to iron deficient conditions (Vorwieger et al., 2007).

In a transcription regulatory pathway it is common for a first transcription regulator to regulate a second transcription regulator, depending on the network of interacting factors a single transcriptional regulator can play roles in a variety of pathways resulting in a variety of physiological or biochemical outcomes. Such a relationship has been shown between some MYB and some bHLH proteins (Ramsay and Glover, 2005).

The MYB family of transcription factors is composed of at least 198 genes (Yanhui et al. 2006) and has been proposed to have regulatory functions in a wide array of processes ranging from growth and development to defense responses. Plant MYB proteins are classified based on the presence and number of imperfect MYB repeats each composed of about 52 amino acids. The MYB domain forms a helix-turn-helix conformation and represents the DNA binding domain. Three major groups of MYB proteins have been classified as R1R2R3-MYB, R2R3-MYB and MYB-related proteins.

The R2R3-MYB family of proteins in *Arabidopsis* consists of 125 proteins and is characterized by having a R2R3 DNA binding domain at their N-terminus (Kranz et al., 1998, and Stracke et al., 2001). These genes are involved in a number of biological processes including mediating hormone actions, secondary metabolism (Paz-Ares et al., 1987), control of cell morphogenesis (Oppenheimer et al., 1991), meristem, floral and seed development (Kirik et al., 1998, Schmitz et al., 2002) and response to various environmental factors (Kranz et al., 1998; Jin and Martin, 1999; Meissner et al, 1999).

SUMMARY OF THE INVENTION

This invention is bases upon the discovery that overexpression of a bHLH subgroup 1b gene transcriptional regulator results in a plant with an altered phenotype such for example increased heat stress tolerance, reduced flower abortion during heat stress, and increased yield relative to a wild type plant.

More specifically, the invention relates to the identification of a bHLH39 or bHLH101 as transcriptional regulators that when overexpressed will produce plants having a heat stress tolerance phenotype.

In one aspect the invention provides a method of producing a transgenic plant, by transforming a plant, a plant tissue culture, or a plant cell with a vector containing a nucleic acid construct that increases the expression or activity of a bHLH subgroup 1b gene to obtain a plant, tissue culture or a plant cell with increased bHLH subgroup 1b expression or activity and growing the plant or regenerating a plant from the plant tissue culture or plant cell, wherein a plant having increased heat stress tolerance relative to a wild type plant is produced.

Accordingly, the present invention provides a method of producing a plant having an improved property, wherein the method includes increasing the expression or activity of an endogenous bHLH subgroup 1b gene, wherein a plant is produced having an advantageous phenotype or improved property. In a particular embodiment, the present invention provides a method for producing plants having increased heat stress tolerance relative to a wild type plant, wherein the method includes include generation of transgenic plants and modification of plants genome using the methods described herein.

Heat stress tolerance refers to the ability of a plant to withstand the debilitating effects of heat which reduces yield of a wild type plant and out perform a wild type plant. As used herein, the term "increased heat stress tolerance" refers to a plant heat stress tolerance is greater as compared to the heat stress tolerance of a corresponding wild-type plant. For example, a plant having increased heat stress tolerance as compared to a wild-type plant may have 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60 % 70%, 75% or greater heat stress tolerance than the corresponding wild-type plant.

The methods of the invention involve increasing the activity of an endogenous or heterologous bHLH subgroup 1b gene by overexpression or promoter modification, wherein a

plant is produced having an advantageous phenotype or improved property, such as increased heat stress tolerance relative to a wild type plant. In one aspect, the invention provides a method of producing a plant having increased heat stress tolerance relative to a wild type plant, by introducing into a plant cell a nucleic acid construct that increases the expression or activity of bHLH subgroup 1b gene or protein. For example, a plant having increased heat stress tolerance relative to a wild type plant is produced by a) providing a nucleic acid construct containing a promoter operably linked to a nucleic acid construct that expresses bHLH subgroup 1b activity; b) inserting the nucleic construct into a vector; c) transforming a plant, tissue culture, or a plant cell with the vector to obtain a plant, tissue culture or a plant cell with increased bHLH subgroup 1b activity; d) growing the plant or regenerating a plant from the tissue culture or plant cell, wherein a plant having increased heat stress tolerance relative to a wild type plant is produced. The construct includes a promoter such as a constitutive promoter, a tissue specific promoter or an inducible promoter. Preferably, the tissue specific promoter is a root promoter. A preferable inducible promoter is a heat or drought inducible promoter.

Provided by the invention is a transgenic plant having an advantageous phenotype or improved property such as increased heat stress tolerance, produced by the methods described herein.

In another aspect the invention provides a plant having a non-naturally occurring mutation in a bHLH gene, wherein the plant has increased bHLH subgroup 1b expression or activity and the plant has increased heat stress tolerance relative to a wild type plant. Increased bHLH expression or activity refers to a 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, or 75-fold increase or greater, at the DNA, RNA or protein level of an bHLH gene as compared to wild-type bHLH, or a 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60 or 75 fold increase of bHLH activity as compared to wild-type bHLH activity.

A further aspect is a plant having an endogenous bHLH subgroup 1b gene that has an altered promoter sequence operably associated with it. Insertion of enhancer elements or promoter mutations that result in increased gene expression are envisioned by the present invention.

The invention further provides a transgenic seed produced by the transgenic plant(s) of the invention, wherein the seed produces plant having an advantageous phenotype or improved property such as for example, increased heat stress tolerance relative to a wild type plant.

In another embodiment, the invention provides nucleic acids for expression of nucleic acids in a plant cell to produce a transgenic plant having an advantageous phenotype or improved property such as increased heat stress tolerance relative to a wild type plant.

Exemplary sequences encoding a wild type bHLH gene or portion thereof that find use in aspects of the present invention are described in SEQ ID NO's: 1-17, 29-52, 79-83 and 89-97. Exemplary sequences encoding a bHLH39 gene are described in SEQ ID NO's:1-17. Exemplary sequences encoding a bHLH38 gene are described in SEQ ID NO's:29-52. Exemplary sequences encoding a bHLH101 gene are described in SEQ ID NO's:79-83. Exemplary sequences encoding a bHLH100 gene are described in SEQ ID NO's:89-97.The invention further provides compositions which contain the nucleic acids of the invention for expression in a plant cell to produce the transgenic plants described herein.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below.

In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from and are encompassed by the following detailed description and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a ClustalW alignment of various AtbLHL proteins, including AtbHLH39, AtbHLH38, AtbHLH100 and AtbHLH101.

DETAILED DESCRIPTION

The invention is based in part on the discovery of plants having an improved agronomic property, such as for example, increased heat stress tolerance relative to a wild-type control. More specifically, the invention is based upon the discovery that overexpression of a bHLH subgroup 1b transcriptional regulator confers on the plant an improved agronomic property, such as for example, increased heat stress tolerance relative to a wild-type control which can include reduced flower abortion and increased yield

The surprising result that overexpression of a bHLH39 subgroup 1b gene is sufficient to confer heat tolerance has been shown. According to microarray and EMSA analysis, bHLH39 is a downstream target of Myb68, another transcription factor which confers heat tolerance when overexpressed. Microarray analysis shows the gene expression of IRT1 is not significantly affected in the 35S-bHLH39, suggesting the heat tolerance conferred by bHLH39 occurs through a separate pathway from the iron deprivation response.

Accordingly the invention provides methods of enhancing (e.g, increasing) the heat stress tolerance of plants by increasing the expression or activity of a bHLH subgroup 1b gene or polypeptide. Methods to increase the expression or activity of a bHLH subgroup 1b gene are known in the art. For example, a plant having increased heat stress tolerance as compared to a wild-type (e.g. control) plant is produced by introducing to a plant cell a nucleic acid construct that increases the expression or activity of a bHLH subgroup 1b gene or polypeptide. The invention also includes the transgenic plants produced by the methods of the invention and the seeds produced by the transgenic plants that produce a plant having increased heat stress tolerance.

For convenience, before further description of the present invention, certain terms employed in the specification, examples and appended claims are defined herein. These

definitions should be read in light of the remainder of the disclosure and as understood by a person of ordinary skill in the art.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below.

In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

A "promoter sequence", or "promoter", means a nucleic acid sequence capable of inducing transcription of an operably linked gene sequence in a plant cell. Promoters include for example (but not limited to) constitutive promoters, tissue specific promoters such as a root promoter, an inducible promoters such as a drought inducible promoter, a heat inducible promoter or an endogenous promoters such as a promoter normally associated with a gene of interest., i.e. a bHLH subgroup 1b gene or cryptic or synthetic promoter sequences which are capable of directing expression of a gene in a plant cell but are not normally associated with an expressible gene.

The term "expression cassette" means a vector construct wherein a gene or nucleic acid sequence is transcribed. Additionally, the expressed mRNA may be translated into a polypeptide.

The terms "expression" or "overexpression" are used interchangeably and mean the expression of a gene such that the transgene or operably linked gene is expressed. The total level of expression in a cell may be elevated relative to a corresponding wild-type cell.

The term "non-naturally occurring mutation" refers to any method that introduces mutations or genetic changes into a plant or plant population. For example, chemical mutagenesis such as ethane methyl sulfonate or methanesulfonic acid ethyl ester, fast neutron mutagenesis, DNA insertional means such as a T-DNA insertion or site directed mutagenesis methods. Also included are methods to induce genetic change such as meganuclease methods that are a particular class of "DNA scissors" they are capable of cutting a chromosome at a specific site in a living cell.

The term "heat stress" refers to a condition where plant growth or productivity is inhibited relative to a plant where heat is not a limiting factor.

The term "heat stress tolerance" refers to the ability of a plant to outperform a wildtype plant under heat stress conditions.

The term "drought stress" refers to a condition where plant growth or productivity is inhibited relative to a plant where water is not limiting. The term "water-stress" is used synonymously and interchangeably with the drought water stress.

The term "drought tolerance" refers to the ability of a plant to outperform a wildtype plant under drought stress conditions or water limited conditions or to use less water during grow and development relative to a wildtype plant.

The term "dry weight" means plant tissue that has been dried to remove the majority of the cellular water and is used synonymously and interchangeably with the term biomass.

The term "null" is defined as a segregated sibling of a transgenic line that has lost the inserted transgene and is therefore used as a control line.

A number of various standard abbreviations have been used throughout the disclosure, such as g, gram; WT, wild-type; DW, dry weight; WUE, water use efficiency; d, day.

The term "bHLH subgroup lb " means a bHLH39, bHLH38, bHLH100 or bHLH101 transcriptional regulator. In some cases the term bHLH is used to refer to the bHLH subgroup lb, as appropriate in the context.

The term "bHLH nucleic acid" refers to at least a portion of a bHLH nucleic acid. A portion is of at least 21 nucleotides in length with respect to a nucleic acid and a portion of a protein or polypeptide is at least 7 amino acids. The term "AtbHLH" refers to an *Arabidopsis thaliana* bHLH gene, the term "Bn bHLH" refers to a *Brassica napus* bHLH gene.

Determining homology between two or more sequences

To determine the percent homology between two amino acid sequences or between two nucleic acids, the sequences are aligned for optimal comparison purposes *(e.g.,* gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or

nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, Needleman and Wunsch (1970). Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the coding sequence portion of the DNA sequence shown in SEQ ID NO:1.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.,* A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.,* the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.,* the window size), and multiplying the result by 100 to yield the percentage of positive residues.

Increased Expression of bHLH subgroup 1b expression and activity

An aspect of the invention pertains to means and methods of increasing or overexpressing bHLH subgroup 1b gene expression and activity, resulting in an increase of bHLH subgroup 1b protein expression and activity. The term "bHLH expression or activity" embraces both these levels of increase. Increased bHLH expression or activity refers to a 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, or 75-fold increase or greater, at the DNA, RNA or protein level of an bHLH subgroup 1b gene as compared to wild-type bHLH, or a 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60 or 75 fold increase of bHLH subgroup 1b protein activity as compared to wild-type bHLH subgroup 1b activity.

Sequences encoding a bHLH subgroup 1b gene or portion thereof that are useful in preparing constructs for bHLH subgroup 1b expression include for example, SEQ ID NO's: 1-17, 29-52, 79-83 and 89-97.

Expression constructs are to provide overexpression either constitutively throughout the plant, or in specific tissues. Alternatively expression can be engineered to occur in response to a temporal, spatial or environmentally regulated stimulus.

Strategies of gene expression will be apparent to the skilled worker including those not discussed here and those developed in the future.

Identification of AtbHLH homologues

Homologues of *Arabidopsis thaliana* bHLH subgroup 1b (AtbHLH) were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*,. 1990 and Altschul *et al.*,. 1997). The tblastn or blastn sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff and Henikoff, 1992). The output of a BLAST report provides a score that takes into account the alignment of similar or identical residues and any gaps needed in order to align the sequences. The scoring matrix assigns a score for aligning any possible pair of sequences. The P values reflect how many times one expects to see a score occur by chance. Higher scores are preferred and a low threshold P value threshold is preferred. These are the sequence identity criteria. The tblastn sequence analysis program was used to query a polypeptide sequence against six-way translations of sequences in a nucleotide database. Hits with a P value less than -25, preferably less than -70, and more preferably less than -100, were identified as homologous sequences (exemplary selected sequence criteria). The blastn sequence analysis program was used to query a nucleotide

sequence against a nucleotide sequence database. In this case too, higher scores were preferred and a preferred threshold P value was less than -13, preferably less than -50, and more preferably less than -100.

A bHLH subgroup 1b gene can be isolated via standard PCR amplification techniques. Use of primers to conserved regions of a bHLH subgroup 1b gene and PCR amplification produces a fragment or full length copy of the desired gene. Template may be DNA, genomic or a cDNA library, or RNA or mRNA for use with reverse transcriptase PCR (RtPCR) techniques. Conserved regions can be identified using sequence comparison tools such as BLAST or CLUSTALW for example. Suitable primers have been used and described elsewhere in this application.

Alternatively, a fragment of a sequence from a bHLH subgroup 1b gene is ^{32}P-radiolabeled by random priming (Sambrook *et al.*, 1989) and used to screen a plant genomic library (the exemplary test polynucleotides) As an example, total plant DNA from *Arabidopsis thaliana*, *Nicotiana tabacum*, *Lycopersicon pimpinellifolium*, *Prunus avium*, *Prunus cerasus*, *Cucumis sativus*, or *Oryza sativa* are isolated according to Stockinger *et al.* (Stockinger *et al.*, 1996). Approximately 2 to 10 μg of each DNA sample are restriction digested, transferred to nylon membrane (Micron Separations, Westboro, Mass.) and hybridized. Hybridization conditions are: 42.degree. C. in 50% formamide, 5X SSC, 20 mM phosphate buffer 1X Denhardt's, 10% dextran sulfate, and 100 μg/ml herring sperm DNA. Four low stringency washes at RT in 2X SSC, 0.05% sodium sarcosyl and 0.02% sodium pyrophosphate are performed prior to high stringency washes at 55°C in 0.2.times.SSC, 0.05% sodium sarcosyl and 0.01% sodium pyrophosphate. High stringency washes are performed until no counts are detected in the washout according to Walling *et al.* (Walling *et al.*, 1988). Positive isolates are identified, purified and sequenced. Other methods are available for hybridization, for example the ExpressHyb ^{TM} hybridization solution available from Clontech.

bHLH Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a bHLH subgroup 1b protein, a bHLH subgroup 1b gene or genomic sequence or portions thereof and analogs or homologs thereof. As used herein the term expression vector includes vectors which are designed to provide transcription of the nucleic acid

sequence. Transcribed sequences may be designed to express the gene construct to increase the total expression or activity of an endogenous gene activity correlating to the transcribed sequence. The expressed sequence may be an endogenous bHLH subgroup 1b encoding protein or from a heterologous species.

The transcribed nucleic acid may be translated into a polypeptide or protein product. The polypeptide may be a non-full length, mutant or modified variant of the endogenous protein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors or plant transformation vectors, binary or otherwise, which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel (1990). Regulatory sequences include those that direct

constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences) or inducible promoters (*e.g.*, induced in response to abiotic factors such as environmental conditions, heat, drought, nutrient status or physiological status of the cell or biotic such as pathogen responsive). Examples of suitable promoters include for example constitutive promoters, ABA inducible promoters, tissue specific promoters and abiotic or biotic stress inducible promoters. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired as well as timing and location of expression, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, bHLH subgroup 1b proteins, mutant forms of bHLH subgroup 1b proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of bHLH subgroup 1b genes, bHLH subgroup 1b proteins, or portions thereof, in prokaryotic or eukaryotic cells. For example, bHLH subgroup 1b genes or bHLH subgroup 1b proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors),) yeast cells, plant cells or mammalian cells. Suitable host cells are discussed further in Goeddel (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

In one embodiment, a nucleic acid of the invention is expressed in plants cells using a plant expression vector. Examples of plant expression vectors systems include tumor inducing (Ti) plasmid or portion thereof found in *Agrobacterium*, cauliflower mosaic virus (CaMV) DNA and vectors such as pBI121.

For expression in plants, the recombinant expression cassette will contain in addition to the bHLH subgroup 1b nucleic acids, a promoter region that functions in a plant cell, a transcription initiation site (if the coding sequence to transcribed lacks one), and optionally a transcription termination/polyadenylation sequence. The termination/polyadenylation region may be obtained from the same gene as the promoter sequence or may be obtained from different genes. Unique restriction enzyme sites at the 5' and 3' ends of the cassette are typically included to allow for easy insertion into a pre-existing vector.

Examples of suitable promoters include promoters from plant viruses such as the 35S promoter from cauliflower mosaic virus (CaMV) (Odell *et al.*, 1985), promoters from genes such as rice actin (McElroy *et al.*, 1990), ubiquitin (Christensen *et al.*, 1992; pEMU (Last *et al.*, 1991), MAS (Velten *et al.*,1984), maize H3 histone (Lepetit *et al.*, 1992); and Atanassvoa *et al.*, 1992), the 5'- or 3'-promoter derived from T-DNA of *Agrobacterium tumefaciens*, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Pat. No. 5,683,439), the Nos promoter, the rubisco promoter, the GRP1-8 promoter, ALS promoter, (WO 96/30530), a synthetic promoter, such as Rsyn7, SCP and UCP promoters, ribulose-1,3-diphosphate carboxylase, fruit-specific promoters, heat shock promoters (HSP 81.1 or HSP18.2), seed-specific promoters, root specific promoters i.e. uclacyanin2 (UCC2, At2g44790) and other transcription initiation regions from various plant genes, for example, including the various opine initiation regions, such as for example, octopine, mannopine, and nopaline. In some cases a promoter associated with the gene of interest (e.g. bHLH) may be used to express a construct targeting the gene of interest, for example the native *At*bHLH promoter. Additional regulatory elements that may be connected to a bHLH subgroup 1b encoding nucleic acid sequence for expression in plant cells include terminators, polyadenylation sequences, and nucleic acid sequences encoding signal peptides that permit localization within a plant cell or secretion of the protein from the cell. Such regulatory elements and methods for adding or exchanging these elements with the regulatory elements of bHLH subgroup 1b gene are known and include, but are not limited to, 3' termination and/or polyadenylation regions such as those of the *Agrobacterium tumefaciens* nopaline synthase (nos) gene (Bevan *et al.*, 1983); the potato proteinase inhibitor II (PINII) gene (Keil *et al.*, 1986); and
An *et al.* (1989); and the CaMV 19S gene (Mogen *et al.*, 1990).

Plant signal sequences, including, but not limited to, signal-peptide encoding DNA/RNA sequences which target proteins to the extracellular matrix of the plant cell (Dratewka-Kos *et al.*, 1989) and the *Nicotiana plumbaginifolia* extension gene (De Loose *et al.*, 1991), or signal peptides which target proteins to the vacuole like the sweet potato sporamin gene (Matsuoka *et al.*, 1991) and the barley lectin gene (Wilkins *et al.*, 1990), or signals which cause proteins to be secreted such as that of PRIb (Lund *et al.*, 1992), or those which target proteins to the plastids such as that of rapeseed enoyl-ACP reductase (Verwoert *et al.*, 1994) are useful in the invention.

In another embodiment, the recombinant expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. For example, the promoter associated with a coding sequence identified in the TAIR data base as At2g44790 (P_{4790}) is a root specific promoter. Especially useful in connection with the nucleic acids of the present invention are expression systems which are operable in plants. These include systems which are under control of a tissue-specific promoter, as well as those which involve promoters that are operable in all plant tissues.

Organ-specific promoters are also well known. For example, the chalcone synthase-A gene (van der Meer *et al.*, 1990) or the dihydroflavonol-4-reductase (dfr) promoter (Elomaa *et al.*, 1998) direct expression in specific floral tissues. Also available are the patatin class I promoter is transcriptionally activated only in the potato tuber and can be used to target gene expression in the tuber (Bevan, 1986). Another potato-specific promoter is the granule-bound starch synthase (GBSS) promoter (Visser *et al.*, 1991).

Other organ-specific promoters appropriate for a desired target organ can be isolated using known procedures. These control sequences are generally associated with genes uniquely expressed in the desired organ. In a typical higher plant, each organ has thousands of mRNAs that are absent from other organ systems (reviewed in Goldberg, 1986).

The resulting expression system or cassette is ligated into or otherwise constructed to be included in a recombinant vector which is appropriate for plant transformation. The vector may also contain a selectable marker gene by which transformed plant cells can be identified in culture. The marker gene may encode antibiotic resistance. These markers include resistance to G418, hygromycin, bleomycin, kanamycin, and gentamicin. Alternatively the marker gene may encode a herbicide tolerance gene that provides tolerance to glufosinate or glyphosate type herbicides. After transforming the plant cells, those cells having the vector will be identified by their ability to grow on a medium containing the particular antibiotic or herbicide. Replication sequences, of bacterial or viral origin, are generally also included to allow the vector to be cloned in a bacterial or phage host, preferably a broad host range prokaryotic origin of replication is included. A selectable marker for bacteria should also be included to allow selection of bacterial cells bearing the desired construct. Suitable prokaryotic selectable markers also include resistance to antibiotics such as kanamycin or tetracycline.

Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art. For instance, in the case of *Agrobacterium* transformations, T-DNA sequences will also be included for subsequent transfer to plant chromosomes.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a polypeptide of the invention encoded in an open reading frame of a polynucleotide of the invention. Accordingly, the invention further provides methods for producing a polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

A number of cell types may act as suitable host cell for expression of a polypeptide encoded by an open reading frame in a polynucleotide of the invention. Plant host cells include, for example, plant cells that could function as suitable hosts for the expression of a polynucleotide of the invention include epidermal cells, mesophyll and other ground tissues, and vascular tissues in leaves, stems, floral organs, and roots from a variety of plant species, such as *Arabidopsis thaliana, Nicotiana tabacum, Brassica napus, Zea mays, Oryza sativa, Gossypium hirsutum and Glycine max.*

Transformed Plants Cells and Transgenic Plants

The invention includes a protoplast, plants cell, plant tissue and plant (*e.g.,* monocot or dicot) transformed with a bHLH subgroup 1b nucleic acid, a vector containing a bHLH subgroup 1b nucleic acid or an expression vector containing a bHLH subgroup 1b nucleic acid. As used herein, "plant" is meant to include not only a whole plant but also a portion thereof (*i.e.,* cells, and tissues, including for example, leaves, stems, shoots, roots, flowers, fruits and seeds).

The plant can be any plant type including, for example, species from the genera *Arabidopsis, Brassica, Oryza, Zea, Sorghum, Brachypodium, Miscanthus, Gossypium, Triticum, Glycine, Pisum, Phaseolus,, Lycopersicon, Trifolium, Cannabis, Cucurbita, Rosa, Vitis, Juglans, Fragaria, Lotus, Medicago, Onobrychis, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciahorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesis, Pelargonium, Panieum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Lolium, Avena, Hordeum, Secale, Picea, Caco*, and *Populus*.

The invention also includes cells, tissues, including for example, leaves, stems, shoots, roots, flowers, fruits and seeds and the progeny derived from the transformed plant.

Numerous methods for introducing foreign genes into plants are known and can be used to insert a gene into a plant host, including biological and physical plant transformation protocols (See, for example, Miki *et al.*, (1993) "Procedure for Introducing Foreign DNA into Plants", In: Methods in Plant Molecular Biology and Biotechnology, Glick and Thompson, eds., CRC Press, Inc., Boca Raton, pages 67-88; and Andrew Bent in, Clough SJ and Bent AF, (1998) "Floral dipping: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana")*. The methods chosen vary with the host plant, and include chemical transfection methods such as calcium phosphate, polyethylene glycol (PEG) transformation, microorganism-mediated gene transfer such as *Agrobacterium* (Horsch *et al.*, 1985), electroporation, protoplast transformation, micro-injection, flower dipping and biolistic bombardment.

***Agrobacterium*-Mediated Transformation**

The most widely utilized method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium tumefaciens* and *A. rhizogenes* which are plant pathogenic bacteria which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectfully, carry genes responsible for genetic transformation

of plants (See, for example, Kado, 1991). Descriptions of the *Agrobacterium* vector systems and methods for *Agrobacterium*-mediated gene transfer are provided in Gruber *et al.* (1993) and Moloney *et al.*, (1989).

Transgenic *Arabidopsis* plants can be produced easily by the method of dipping flowering plants into an *Agrobacterium* culture, based on the method of Andrew Bent in, Clough SJ and Bent AF, 1998. Floral dipping: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Wild type plants are grown until the plant has both developing flowers and open flowers. The plants are inverted for 1 minute into a solution of *Agrobacterium* culture carrying the appropriate gene construct. Plants are then left horizontal in a tray and kept covered for two days to maintain humidity and then righted and bagged to continue growth and seed development. Mature seed is bulk harvested.

Direct Gene Transfer

A generally applicable method of plant transformation is microprojectile-mediated transformation, where DNA is carried on the surface of microprojectiles measuring about 1 to 4 μm. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds of 300 to 600 m/s which is sufficient to penetrate the plant cell walls and membranes (Sanford *et al.*, 1993; Klein *et al.*, 1992).

Plant transformation can also be achieved by the Aerosol Beam Injector (ABI) method described in U.S. Pat. 5,240,842 and U.S. Pat. 6,809,232. Aerosol beam technology is used to accelerate wet or dry particles to speeds enabling the particles to penetrate living cells. Aerosol beam technology employs the jet expansion of an inert gas as it passes from a region of higher gas pressure to a region of lower gas pressure through a small orifice. The expanding gas accelerates aerosol droplets, containing nucleic acid molecules to be introduced into a cell or tissue. The accelerated particles are positioned to impact a preferred target, for example a plant cell. The particles are constructed as droplets of a sufficiently small size so that the cell survives the penetration. The transformed cell or tissue is grown to produce a plant by standard techniques known to those in the applicable art.

Regeneration of Transformants

The development or regeneration of plants from either single plant protoplasts or various explants is well known in the art (Weissbach and Weissbach, 1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those

individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a polypeptide of interest introduced by *Agrobacterium* from leaf explants can be achieved by methods well known in the art such as described (Horsch *et al.*, 1985). In this procedure, transformants are cultured in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant strain being transformed as described (Fraley *et al.*, 1983). In particular, U.S. Pat. No. 5,349,124
details the creation of genetically transformed lettuce cells and plants resulting therefrom which express hybrid crystal proteins conferring insecticidal activity against Lepidopteran larvae to such plants.

This procedure typically produces shoots within two to four months and those shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil or other media to allow the production of roots. These procedures vary depending upon the particular plant strain employed, such variations being well known in the art.

Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants, or pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important, preferably inbred lines. Conversely, pollen from plants of those important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

A preferred transgenic plant is an independent segregate and can transmit the bHLH subgroup 1b gene construct to its progeny. A more preferred transgenic plant is homozygous for the gene construct, and transmits that gene construct to all offspring on sexual mating. Seed from a transgenic plant may be grown in the field or greenhouse, and resulting sexually mature transgenic plants are self-pollinated to generate true breeding plants. The progeny from these plants become true breeding lines that are evaluated for increased expression of the bHLH subgroup 1b gene.

Method of Producing Transgenic Plants

Also included in the invention are methods of producing a transgenic plant having increased heat stress tolerance, reduced flower abortion, and increased yield relative to a wild type plant following a heat stress. The method includes introducing into one or more plant cells a compound that increases bHLH subgroup 1b expression or activity in the plant to generate a transgenic plant cell and regenerating a transgenic plant from the transgenic cell. The compound can be, *e.g.*, (i) a bHLH subgroup 1b polypeptide; (ii) a bHLH subgroup 1b nucleic acid, analog, homologue, orthologue, portion, variant or complement thereof; (iii) a nucleic acid that increases expression of a bHLH subgroup 1b nucleic acid. A nucleic acid that increases expression of a bHLH subgroup 1b nucleic acid may include promoters or enhancer elements. The bHLH subgroup 1b nucleic acid can be either endogenous or exogenous, for example an *Arabidoposis* bHLH subgroup 1b nucleic acid may be introduced into a *Brassica* or corn species. Preferably, the compound is a bHLH subgroup 1b nucleic acid sequence endogenous to the species being transformed. Alternatively, the compound is a bHLH subgroup 1b nucleic acid sequence exogenous to the species being transformed and having at least 70%, 75%, 80%, 85%, 90% or greater homology to the endogenous target sequence.

In various aspects the transgenic plant has an altered phenotype as compared to a wild type plant (*i.e.*, untransformed). By altered phenotype is meant that the plant has a one or more characteristic that is different from the wild type plant. For example, when the transgenic plant has been contacted with a compound that increases the expression or activity of a bHLH subgroup 1b nucleic acid, the plant has a phenotype such as increased heat stress tolerance, reduced flower abortion, and increased yield relative to a wild type plant following heat stress.

The plant can be any plant type including, for example, species from the genera *Arabidopsis, Brassica, Oryza, Zea, Sorghum, Brachypodium, Miscanthus, Gossypium, Triticum, Glycine, Pisum, Phaseolus,, Lycopersicon, Trifolium, Cannabis, Cucurbita, Rosa, Vitis, Juglans, Fragaria, Lotus, Medicago, Onobrychis, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciahorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesis, Pelargonium, Panieum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Lolium, Avena, Hordeum, Secale, Picea, Caco*, and *Populus*.

EXAMPLES

Example 1: Identification of Homologous genes

Blast searches of bHLH39 and its *Arabidopsis* homologues bHLH38, bHLH100, and bHLH101 were performed against NCBI's protein, nucleotide, and EST databases and TIGR's unigene database (1e-01). Genomic sequence databases was used for species whose complete sequence in known, such as Rice and Sorghum. To confirm a sequence was indeed an ortholog of the sequences of interest, all putative homologues were then blasted against the complete *Arabidopsis* protein database, and any sequence whose top *Arabidopsis* hit was not one of the four sequences of interest was filtered from any further analysis. The minimize redundancy between multiple hits, all EST sequence hits were assembled using the cap3 assembly program ensuring they have a minimum of 90% identity across at least 40 nucleotide with a maximum of 5 gaps. Open reading frames were determined using the EMBOSS program getorf.

Blast searches of AtbHLH39, AtbHLH38, AtbHLH100, AtbHLH101 retrieved 96 homologues sequences from 40 different species. Sequence conservation was limited to DNA binding and dimerization domains, with little to no conservation outside of this area. Homologues were discovered in both monocots and dicots suggesting an important functional requirement in plants. Homologues of AtbHLH39 in agronomical important species such as rice and corn were retrieved as well as species used in biofuels such as Brachypodium and Switchgrass.

According to blast searches against the Brachypodium genomic database, there are three homologues to this group of genes, and all three are most similar to AtbHLH38. There is EST evidence for two of three the homologues. Whether the third homologue is a pseudogene or codes for a functional protein is yet to be determined.

In two other monocot species, rice and Sorghum, there is only one homologue from this group, and in both species the homologue is closest to AtbHLH38. In both rice and Sorghum, the gene encodes at least two splice variants.

There are at least two homologues of this group in *Brassica napus*, one protein with close homology to AtbHLH39, and another with close homology to AtbhLH38. However, the BnbHLH39 homologue contains a nonsense mutation which results in the coding of a partial protein with no DNA binding or dimerization domain. This finding is confirmed with several EST's, and also occurs in *Brassica rapa*.

Example 2: Vector Construction

The binary vector pBI121 was optimized for transformation of *Arabidopsis* and different crops. The GUS gene in pBI121 was deleted by SmaII and EcolCR1 digestions and religation, resulting in vector pBI121ΔGUS. This vector was used to clone gene for overexpression. The C-terminal 1.1-Kb portion of the GUS gene was isolated from pBI121 as an EcoR V-Sac I fragment (positions 6613-7715 in pBI121) and cloned at the Sma I and Sac I sites in pBI121ΔGUS, resulting in pBI121tGUS (with N-terminal portion of the GUS gene deleted). The vector was used to make gene down-regulation Hairpin constructs with the partial GUS sequence as the loop or spacer.

The NPTII gene in the vector pBI121 and its derivatives contains a point mutation (G to T at position 3383 in pBI121, amino acid substitution E182D). The mutant enzyme showed several fold lower enzyme activity than its wild type (PNAS, 87:3435-3439, 1990). In order to improve transformation efficiency of different crops, the vectors pBIΔGUS and pBItGUS were restored with the WT-NPTII gene: a Nhe I-BstB I fragment (0.9 kb, positions 2715-3648) was replaced with a Nhe I-BstB I fragment of exactly the same sequence except the single nucleotide difference. The fragment was isolated by restriction digestion from plasmid pRD400 which contained the WT-NPTII gene (PNAS, 87:3435-3439, 1990; Gene, 122:383-384, 1992). The modified vectors were named pBI300ΔGUS and pBI300tGUS, respectively. The WT-NPTII gene was also isolated from pBI300ΔGUS as a Nhe I-Hind III fragment (2.2 kb) and cloned at the corresponding sites in pBI121. This generated vector pBI300GUS. To distinguish these vectors from others, pBI121-based binary vectors containing the WT-NPTII gene were designated pBI300 series.

In order to use Basta as a selection agent, a Basta resistance marker was subcloned into pBI121. A 1.3-kb Ase I fragment encompassing the Basta selection marker (35S-Bar-nosT) was amplified by PCR from vector pEGAD using a forward primer containing a Pme I site and a reverse primer containing a Hind III site. The fragment was cloned, between the sites of PmeI (position 2492) a III (position 4950) in pBI121, pBIΔGUS and pBItGUS. As a result the, the kanamycin selection marker (Pnos-NPTII-nosT) was replaced with the Basta selection marker in these vectors. To distinguish these vectors from others, pBI121-based binary vectors containing the Basta selection marker gene were designated pBI800 series.

pBI300 vectors contain the WT-NPTII gene driven by the nopaline synthase (nos) promoter. Since the promoter is not active in monocotyledon plants, it was necessary to replace it with a strong promoter for transformation of monocotyledon plants. For this purpose, the Brachyposium TIF1 gene (BdTIF1, also called BdGOS2) promoter was cloned into vectors of the pBI300 series to drive the WT-NPTII gene. The BdTIF1 promoter sequence (-1 through -2548 with respect to the ATG start codon, including the first exon, first intron and portion of the second exon) was amplified by PCR as a PmeI-NheI fragment and cloned at the corresponding sites in HSP81.1-AtMyb68-pBI300 (a pBI300-based vector containing the *Arabidopsis* heat shock protein 81.1 promoter and *Arabidopsis* Myb68 coding sequence, see below for details). As a result the TIF1p promoter was placed 5' upstream to the WT-NPTII gene with a small fragment of about 120 bp (including 65 bp of the nos promoter) between them. In order to eliminate the 120-bp fragment, the WT-NPTII coding sequence and flanking vector sequence was amplified as a Nhe I-Sal I fragment (2.1 Kb) and ligated to the plasmid digested with Nhe I and Sal I. The cloning resulted in vector HSP81.1-AtMyb68-pBI500. pBI121-based vectors containing the WT-NPTII gene driven by the BdTIF1 promoter were designated pBI500 series. The HSP81.1-AtMyb68 sequence was also replaced with the 35S promoter sequence as a Hind III-BamH I fragment isolated by restriction digestion from pBI300ΔGUS. This generated vector pBI500ΔGUS.

Example 3: bHLH subgroup b1 over-expression constructs

Construct 35S-AtbHLH39:

The AtbHLH39 (At3g56980) coding sequence was amplified by RT-PCR from *Arabidopsis* using forward primer BHLH039FW-XbaI and reverse primer BHLH039RV-BamH I. The PCR product (0.8 Kb) was cloned at the Xba I and BamH I site in the binary vector pBI300ΔGUS and pBI800ΔGUS, generating construct 35S-AtbHLH39-pBI300 and 35S-AtbHLH39-pBI800, respectively.

Construct 35S-AtbHLH101:

The AtbHLH101 (At5g04150) coding sequence was amplified by RT-PCR from *Arabidopsis* using forward primer BHLH101FW-XbaI and reverse primer BHLH101RV-BamH I. The PCR product (0.7 Kb) was cloned at the Xba I and BamH I site in the binary vector

pBI300ΔGUS and pBI800ΔGUS, generating construct 35S-AtbHLH101-pBI300 and 35S-AtbHLH101-pBI800, respectively.

Construct 35S-AtbHLH38:

Using the same strategy as described above, the AtbHLH38 (at3g56970) coding sequence is amplified by RT-PCR from *Arabidopsis* using a forward primer containing an Xba I site and reverse primer containing a Bam HI site, and cloned at the Xba I and BamH I site in the binary vector pBI300ΔGUS, generating construct 35S-AtbHLH38-pBI300.

Construct 35S-AtbHLH100:

Using the same strategy as described above, the AtbHLH100 (At2g41240) coding sequence is amplified by RT-PCR from *Arabidopsis* using a forward primer containing an Xba I site and reverse primer containing a Bam HI site, and cloned at the Xba I and BamH I site in the binary vector pBI300ΔGUS, generating construct 35S-AtbHLH100-pBI300.

Constructs HSP81.1-AtbHLH39:

Several steps were involved in the development of the construct. A first, promoter sequence (-401 to -1 with respect to the ATG start codon) of the *Arabidopsis* heat shock protein gene HSP81.1 (At5g52640) was PCR amplified with primers having Sal I and Xba I ends from *Arabidopsis* genomic DNA, and cloned at the same sites in pBI101, thereby replacing the 35S promoter. This vector was named pHSP81.1-GUS. Sequencing revealed point mutations of T to C at position -266 and C to T at -121 in the promoter. The GUS staining of *Arabidopsis* seedlings transformed with this construct showed the same heat induction expression profile as reported in the literature. Hence these mutations do not apparently affect the functionality of the promoter. Secondly, MCS2-oligo from New England Biolabs was annealed and ligated to pHSP81.1-GUS that had been digested with Xba I and Sma I. This resulted in vector pHSP81.1MCS-GUS. Thirdly, the GUS sequence was deleted by SmaI and EcolCRI digestion, and vector self-religation. This led to vector pHSP81.1MCSΔGUS. Fourthly, the AtMyb68 coding sequence (At5g65790) was cloned at the XbaI and BamH I sites for Myb68 overexpression. Fifthly, the mutant NPTII sequence was replaced with its WT sequence as described above, producing vector HSP81.1-AtMyb68-pBI300. Finally, the AtbHLH39 coding

sequence as an Xba I-BamH I fragment was cloned at the Xba I and BamH I site thereby replacing the AtMyb68 sequence, resulting in construct HSP81.1-AtbHLH39-pBI300. Likewise the Xba I-BamH I fragment of AtbHLH39 coding sequence replaced the AtMyb68 sequence in HSP81.1-AtMyb68-pBI500, resulting in the vector HSP81.1-AtbHLH39-pBI500.

Construct UCC2-AtBHLH39:

Arabidopsis uclacyanin2 gene (UCC2, At2g44790) is expressed at very high level in the roots. Its expression is detectable but at very low levels in other parts of the plant. Its cell-specific expression profile in the root is similar to that of AtbHLH39, *i.e.* predominantly expressed in endodermis and cortex and stele. The UCC2 gene promoter sequence (-1 through -1475 with respect to the ATG start codon) was amplified by PCR using forward primer containing a Sal I site (P790-Sal-F) and reverse primer containing an XbaI site (P790-Xb-R). The PCR product of the UCC2 promoter was cloned at the Sal I and Xba I sites in the vector HSP81.1-AtMyb68-pBI300 (see above), replacing the HSP81.1 promoter. The AtbHLH39 coding sequence was then cloned into this vector as an Xba I-BamH I fragment, replacing the AtMyb68 coding sequence. The resulting vector is named UCC2-AtbHLH39-pBI300.

Construct BdBS-AtbHLH39-pBI500:

Brachypodium biotin synthase (BdBS) gene promoter (-1 through -553 with respect to the ATG start codon) was PCR amplified as a Sal I-Xba I fragment and was cloned to substitute the HSP81.1 promoter in HSP81.1-AtbHLH39-pBI500. The resulting vector was named BdBS-AtbHLH39-pBI500.

Construct BdUCC-AtbHLH39-pBI500:

The closest Brachypodium homolog of the *Arabidopsis* uclacyanin2 gene was found in genomic sequence super_67. The sequence was 34% identical to *Arabidopsis* uclacyanin2 along aligned regions. The open reading frame was determined. A reverse blast of the translated protein sequence to the *Arabidopsis* TAIR8 proteins found it to be the most similar to *Arabidopsis* Uclacyanin1, a close homologue of uclacyanin2. In *Arabidopsis*, uclacyanin1 shares similar expression pattern to uclacyanin2 but with an overall weaker expression. The promoter sequence of the Brachypodium uclacyanin homolog (BdUCC) promoter (-22 through -

1405 relative to the ATG start codon) was amplified as a Sal I-Xba I fragment as was cloned to substitute the HSP81.1 promoter in HSP81.1-AtbHLH39-pBI500. The resulting vector was named BdUCC-AtbHLH39-pBI500.

Example 4: Cloning of the Brachypodium bHLH39

There are three Brachypodium bHLH genes with high homology to the *Arabidopsis* bHLH subgroup 1b comprising bHLH39, bHLH 38, bHLH 100 and bHLH 101. All three Brachypodium homologues are most closely related to AtbHLH38. There is strong EST evidence for homologue #1 (super_13.506): the 5' portion is identical to sequence in EST DV488230 while the 3' portion is identical to DV488393 which contains polyAs. However, DV488393 also contains sequence apparently belonging to an intron. Therefore, the exact open reading frame remains to be determined. The putative coding sequence (BdbHLH39H1) is cloned by RT-PCR using forward primer BdH1-Xb-F (containing an Xba I site) and reverse primer BdH1-Bm-R (containing a BamH I site), and cloned into the corresponding sites in a vector to make constructs HSP81.1-BdbHLH39H1-pBI500, BdBS-BdbHLH39-BI500 and BdUCC-BdbHLH39H1-pBI500.

Primers

Table 1: Cloning Primers

SEQ ID	Name	Sequence (5' to 3')	PCR product
106	BHLH039FW-XbaI	aaaTCTAGAATGTGTGCATTAGTACCTCCATTGTTTC	AtbHLH39 CDS, 0.8 Kb
107	BHLH039RV-BamHI	aaaGGATCCTCATATATATGAGTTTCCACATTCCTCA TAC	
108	BHLH101FW-XbaI	aaatCTAGAATGGAGTATCCATGGCTGCAGTCTC	AtbHLH101CDS, 0.7 Kb
109	BHLH101RV-BamHI	aaaGGATCCTTATGATTGGCGTAATCCCAAGAGC	
110	P790-Sal-F	acgtGTCGAC CTT AGC CAA TGG ATG AGG ATG	AtUCC2 promoter, 1.5 Kb
111	P790-Xb-R	acgtTCTAGA TTT TTG TTT ACT GTA GAA GAG	
112	BdGOS-Pm-F10	acgtGTTTAAAC GCA TAG ACT CTC AGC GGA GAG	BdTIF1 promoter, 2.5 Kb
113	BdGOS-Nh-R	acgtGCTAGC gaaaactcctggtgagagtgg	
114	NPTII-Nh-F	acgtGCTAGC atgattgaacaagatggattgcac	WT-NPTII and flanking sequence, 2.1 Kb
115	NPTII-Sal-R	acgtGTCGAC CTG CAG GCA TGC AAG CTT GG	

SEQ ID	Name	Sequence (5' to 3')	PCR product
116	BdBSp-Sal-F	acgtGTCGAC ctctggatgcctaaacaaacgac	BdBS promoter, 0.5 Kb
117	BdBSp-Xb-R	acgtTCTAGA ggcttttgtcggtcggcctg	
118	BdUCCp-Sa-F4	acgtGTCGAC GGA GGT GCA GTT TGC AGC AG	BdUCC promoter, 1.4 Kb
119	BdUCCp-Xb-R4	acgtTCTAGA TAT AGA GAG AGG GTG ATC AAC GA	
120	BdH1-Xb-F	acgtTCTAGA ATG GGG CAC AAG CAG CTG TTC	BdbHLH39 homolog, super_13.506 (0.7 Kb)
121	BdH1-Bm-R	acgtGGATCC TCA CTG ATG CAT ATG CAG TCC	

Example 5: Plant Transformations

The constructs described above have been or are transformed into *Arabidopsis* and *Brachypodium* as appropriate. Other species are transformed with an appropriate vector and transformed plants produced.

Table 2: Transformation of Constructs

CONSTRUCT	TARGET SPECIES	Transformed
35S-AtbHLH39-pBI300	Dicots i.e.Arabidopsis(At), Brassica(Bn)	At, Bn
35S-AtbHLH39-pBI800	Dicots i.e.Arabidopsis(At), Brassica(Bn)	
35S-AtbHLH101-pBI300	Dicots i.e.Arabidopsis(At), Brassica(Bn)	At
35S-AtbHLH101-pBI800	Dicots i.e.Arabidopsis(At), Brassica(Bn)	
35S-AtbHLH38-pBI300	Dicots i.e.Arabidopsis(At), Brassica(Bn)	
35S-AtbHLH100-pBI300	Dicots i.e.Arabidopsis(At), Brassica(Bn)	
HSP81.1-AtbHLH39-pBI300	Dicots i.e.Arabidopsis(At), Brassica(Bn)	At, Bn
HSP81.1-AtbHLH39-pBI500	Monocots, i.e. Brachypodium	
UCC2-AtbHLH39-pBI300	Dicots i.e.Arabidopsis(At), Brassica(Bn)	
HSP81.1-AtbHLH39-pBI500	Monocots, i.e. Brachypodium	
BdBS-AtbHLH39-pBI500	Monocots, i.e. Brachypodium	
HSP81.1-AtbHLH39-pBI500	Monocots, i.e. Brachypodium	
BdUCC-AtbHLH39-pBI500	Monocots, i.e. Brachypodium	
HSP81.1-BdbHLH39H1-pBI500	Monocots, i.e. Brachypodium	
BdBS-BdbHLH39-BI500	Monocots, i.e. Brachypodium	
BdUCC-BdbHLH39H1-pBI500	Monocots, i.e. Brachypodium	

Example 6: Methods of transformation

Arabidopsis transgenic plants were made by the method of dipping flowering plants into an *Agrobacterium* culture, based on the method of Andrew Bent in, Clough SJ and Bent AF, 1998. Floral dipping: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Wild type plants were grown under standard conditions with a 16 hour, 8 hour light to dark day cycle, until the plant has both developing flowers and open flowers. The plant was inverted for 2 minutes into a solution of *Agrobacterium* culture carrying the appropriate gene construct. Plants were then left horizontal in a tray and kept covered for two days to maintain humidity and then righted and bagged to continue growth and seed development. Mature seed was bulk harvested.

Transformed T1 plants were selected by germination and growth on MS plates containing 50 μg/ml kanamycin or an appropriate selection medium. Green, kanamycin resistant (Kan^R) seedlings were identified after 2 weeks growth and transplanted to soil. Plants were bagged to ensure self fertilization and the T2 seed of each plant harvested separately. During growth of T1 plants leaf samples were harvested, DNA extracted and Southern blot and PCR analysis performed.

T2 seeds were analyzed for Kan^R segregation. From those lines that showed a 3:1 resistant phenotype, surviving T2 plants were grown, bagged during seed set, and T3 seed harvested from each line. T3 seed was again used for Kan^R segregation analysis and those lines showing 100% Kan^R phenotype were selected as homozygous lines. Further molecular and physiological analysis was done using T3 seedlings.

Transgenic *Brassica napus, Glycine max* and *Zea maize* plants are produced using *Agrobacterium* mediated transformation of cotyledon petiole tissue. Seeds are sterilized as follows. Seeds are wetted with 95% ethanol for a short period of time such as 15 seconds. Approximately 30 ml of sterilizing solution I is added (70% Javex, 100μl Tween20) and left for approximately 15 minutes. Solution I is removed and replaced with 30 ml of solution II (0.25% mecuric chloride, 100μl Tween20) and incubated for about 10 minutes. Seeds are rinsed with at least 500 ml double distilled sterile water and stored in a sterile dish. Seeds are germinated on plates of $^1/_2$ MS medium, pH 5.8, supplemented with 1% sucrose and 0.7% agar. Fully expanded cotyledons are harvested and placed on Medium I (Murashige minimal organics (MMO), 3% sucrose, 4.5 mg/L benzyl adenine (BA), 0.7% phytoagar, pH5.8). An *Agrobacterium* culture

containing the nucleic acid construct of interest is grown for 2 days in AB Minimal media. The cotyledon explants are dipped such that only the cut portion of the petiole is contacted by the *Agrobacterium* solution. The explants are then embedded in Medium I and maintained for 5 days at 24°C, with 16,8 hr light dark cycles. Explants are transferred to Medium II (Medium I, 300 mg/L timentin,) for a further 7 days and then to Medium III (Medium II, 20 mg/L kanamycin). Any root or shoot tissue which has developed at this time is dissected away. Transfer explants to fresh plates of Medium III after 14 -21 days. When regenerated shoot tissue develops the regenerated tissue is transferred to Medium IV (MMO, 3% sucrose, 1.0% phytoagar, 300 mg/L timentin, 20 mg/L 20 mg/L kanamycin). Once healthy shoot tissue develops shoot tissue dissected from any callus tissue are dipped in 10X IBA and transferred to Medium V (Murashige and Skooge (MS), 3% sucrose, 0.2 mg/L indole butyric acid (IBA), 0.7% agar, 300 mg/L timentin, 20 mg/L 20 mg/L kanamycin) for rooting. Healthy plantlets are transferred to soil. The above method, with or without modifications, is suitable for the transformation of numerous plant species including *Glycine max*, *Zea maize* and cotton.

Transgenic *Glycine max*, *Zea maize* and cotton are produced using *Agrobacterium*-based methods which are known to one of skill in the art. Alternatively one can use a particle or non-particle biolistic bombardment transformation method. An example of non-particle biolistic transformation is given in U.S. Patent Application 20010026941. This method has been used to produce transgenic *Glycine max* and *Zea maize* plants. Viable plants are propagated and homozygous lines are generated. Plants are tested for the presence of drought tolerance, physiological and biochemical phenotypes as described elsewhere.

Transformation of plant tissue such as *Zea maize,* for example, are achieved by sonication of callus tissue culture. Callus tissue was produced as follows. Ears of corn were harvested18 days after silking and surface sterilized in 50% v/v bleach for 20 minutes followed by three washing with sterile distilled water. Immature embryos ranging in size from 2 to 4 mm were harvested from the kernels. Embryos were placed on $MSD_{1.5}$ medium (2% sucrose, 1X MS macronutrient and micronutrient salts, 1X MS vitamins, 1.5 mg / L 2,4-D, 0.8% agar, pH 5.8) scutellum side up. Embryos were incubated at 26-28 °C in the dark. Friable callus from 2 week old cultures were transferred to fresh $MSD_{1.5}$ medium and further incubated at 26-28 °C in the dark. Friable callus was subcultured to fresh $MSD_{1.5}$ medium every 21 days.

Transformation of callus tissue was performed as described below. The construct was introduced into GV3101 *Agrobacterium* by inoculation of a single colony of GV3101 *Agrobacterium* containing the HPR-GUS plasmid into 10 mL of LB amended with 150 µg/mL rifampicin, 100 µg/mL gentamycin sulfate, and 50 µg/mL kanamycin. The culture was allowed to grow overnight at 28°C with 200 rpm shaking. Corn callus was cut into pieces approximately 3-5 mm in size. The *Agrobacterium* culture was centrifuged at 1500 x g for 10 minutes and washed twice with 10 mL liquid $MSD_{1.5}$ liquid (2% sucrose, 1X MS macronutrient and micronutrient salts, 1X MS vitamins, 1.5 mg / L 2,4-D, pH 5.8). The bacteria was resuspended in liquid $MSD_{1.5}$ to an OD_{600nm} of 0.25 and

1 mL of diluted *Agrobacterium* or liquid $MSD_{1.5}$, for negative controls, was placed in 1.5 mL microfuge tubes containing four pieces of callus added to each tube. Callus and *Agrobacterium* culture was sonicated in a Branson 200 Ultrasonic Cleaner for 0, 3, 10, 30, 100, or 300 seconds with bacteria or 0 or 300 seconds without bacteria (in $MSD_{1.5}$ liquid alone). After sonication, the callus was blotted on sterile filter paper and placed on $MSD_{1.5}$A medium ($MSD_{1.5}$ solid medium amended with 100 µM acetosyringone). The co-cultivation period was 4 days in the dark at 28°C. Callus was rinsed in liquid $MSD_{1.5}$, blotted on sterile filter paper, and placed on $MSD_{1.5}$T medium ($MSD_{1.5}$ solid medium amended with 400 µg/mL Timentin) for 3 days in the dark at 28°C. Seven days after sonication, callus was added to 1 mL GUS staining solution (50 mM $NaPO_4$, pH 7.0, 0.1% Triton X-100, 1 mM EDTA, 2 mM DTT, 0.5 mg/mL X-GlcA) and left to incubate overnight at 37 °C. The staining solution was replaced with 1 mL fixation buffer (10% formaldehyde, 50% ethanol) and incubated for 30 minutes at room temperature. The fixation buffer was replaced with 80% ethanol and incubated for 1 hour at room temperature. The 80% ethanol was replaced with 100% ethanol and incubated for 1 hour at room temperature. The callus was assessed for blue staining, indicating GUS activity.

Other methods of plant transformation are used when appropriate and are commonly described and known in the art.

Example 7: Expression analysis

Total RNA was isolated from 22 transgenic 35S-AtbHLH39 lines and wild type *Arabidopsis* lines. Approximately 10 µµg of total RNA was loaded into each lane. The Northern was probed with radiolabeled HPR cDNA in ExpressHyb hybridization solution (Clontech) and

exposed using a phosphoimaging screen. For quantification blots were reprobed with tubulin, a constitutively expressed gene, for a comparative standard.

The expression level of AtbHLH39 was elevated 4 fold to 126 fold higher than that of the WT Columbia control. The best performing lines (line-34 and line- 97) based on heat and drought tolerance, had 50 and 40 fold increase in expression of AtbHLH39 respectively.

Example 8: Microarray and expression analysis

bHLH subgroup 1b includes AtbHLH38, AtbHLH39, AtbHLH100 and AtbHLH101. Microarray data of lines transformed with an overexpression construct to increase MYB68 expression or activity (35S-AtMYB68) showed that the RNA expression of AtbHLH39 and AtbHLH101 was increased 32 fold and 15 fold, respectively when AtMYB68 was over-expressed 37 fold. The AtbHLH38 and AtbHLH100 genes were not present on the microarray chip, so the expression of these closely related genes is yet to be determined. The four AtbHLH38, AtbHLH39, AtbHLH100, AtbHLH101 members have high sequence homology in group 1b, and may be at least partially functionally redundant.

AtbHLH39 expression is mainly co-localized with AtMYB68 in the pericycle of the root as is that of Myb68 (Birnbaum et al. 2003; Schmid et al. 2005). Analysis of the endogenous MYB68 promoter and the AtbHLH39 promoter demonstrates that these promoters express predominantly in root tissue.

Further microarray analysis, the gene expression of IRT1 is not significantly affected in the transgenic plants over-expressing AtbHLH39 (35S-AtbHLH39 construct) suggesting that the heat tolerance phenotype conferred by AtbHLH39 overexpression does not regulate IRT1 gene expression and that the reported involvement ofbHLH39 in iron deprivation responses occurs through a separate pathway.

Transcription characteristics of members of the bHLH subgroup 1b transcription factor gene family AtbHLH39, AtbHLH38, AtbHLH100, and AtbHLH101 were analyzed. The AtbHLH39 expresses predominantly in the root, whereas AtbHLH38, AtbHLH100, and AtbHLH101 have low expression in leaf, bud, flower and root tissues. The AtbHLH38 transcription is barely detectable in root and leaf, AtbHLH100 expresses in very low level in the bud and flower and AtbHLH101 also expresses at very low level in root and leaf.

Example 9: MYB68 binds to the bHLH39 promoter

Electrophoretic mobility shift assays (EMSA) were performed to determine whether MYB68 protein could bind to the promoter sequence of AtbHLH39. Four different segments of the promoter were labeled and tested: P1 (-1 to -224, with respect to the ATG start codon), P2 (-204 to -419), P3 (-401 to -623) and P4 (-601 to -788). The labeled probe DNA was incubated with purified MYB68 fusion protein. A DNA-protein complex was detected with probe P1 while other regions of the promoter showed no binding with MYB68. Formation of the complex P1/Myb68 was eliminated by the addition of cold competitor DNA with the same sequence as the probe, but not affected by cold competitors of P3. Moreover, no DNA-protein complex was formed when probe P1 was incubated with a truncated MYB68 protein (MYB68CD) in which the R2R3 DNA-binding domains was deleted. The data demonstrated that MYB68 protein bind specifically to the AtbHLH39 promoter in the region -1 to -224.

To further localize the MYB68 binding site, four double-stranded oligonucleotides covering the promoter region -1 to -232 were tested as competitors in binding assays: P11 (-1 to -61), P12 (-52 to-118), P13 (-109 to-178) and P14 (-169-to -232). Binding of MYB68 to the probe P1 was abolished by the competitor P12, but not affected by other fragments, indicating that the MYB68 binding site is within the sequence -52 to-118.

Example 10: Constitutive expression of At-bHLH39 in *Arabidopsis* results in reduced flower abortion following heat stress

An experiment was set up with 14 transgenic 35S-AtbHLH39 lines and a wild type (WT) control. Plants were grown in 2.25 inch pots under optimal conditions (22C, 18hr light of 200uE, 60% RH) in a growth chamber until three days post-appearance of the first flower. A heat stress treatment was applied by placing plants at 42C for 2 hours. One week following the stress period the plants were assessed for number of aborted flowers. The results are shown in Table 3. Ten of the transgenic lines had reduced flower abortion relative to WT controls. One of the lines (34-1) had statistically significantly less aborted flowers than the WT.

Table 3: Average flower abortion following 2 hr at 42C (n=6 to 11) ±SE

Entry	# of aborted flowers	Flower abortion as % of WT
97-7	2.5±0.5	44%
34-1	2.9±0.3	51%
46-10	3.2±0.3	56%
57-2	3.6±0.4	64%
1-7	3.8±0.2	67%
41-2	4.1±0.3	72%
38-10	4.5±0.7	79%
96-4	4.5±0.5	79%
12-4	4.7±0.5	82%
30-2	4.8±0.5	85%
94-10	5.1±0.9	90%
10-1	5.7±0.5	100%
52-4	6.2±0.7	108%
103-1	6.2±0.6	109%
WT	5.7±0.6	100%

Example 11: Constitutive expression of AtbHLH39 in *Arabidopsis* results in drought tolerance of the plants

Drought tolerance was assessed in six transgenic At-bHLH39 lines that also showed reduced flower abortion following heat stress. Plants were grown (5 per 3 inch pot, n=8) under optimal conditions in a growth chamber (22C, 18hr light of 200uE, 60% RH) until the first open flower. Drought treatment was applied by watering up all plants to the same saturated level. Further water was withheld. Plants were weighed daily to determine the daily water loss and all plants were harvested on day four of treatment, at which time all plants were visibly wilting. The water loss relative to final shoot biomass was calculated and is a representative indicator of drought tolerance. Data was normalized to WT which was set as 100% (Table 4). All six of the lines showed reduced water loss relative to shoot biomass, 3 of which were significant statistically. This is indicative of drought tolerant phenotype. Two of the best performing drought tolerant lines were also the best performing heat tolerant lines thereby indicating a close

link between the two traits: drought tolerance and heat tolerance as a result of constitutive expression of AtbHLH39 in *Arabidopsis*.

Table 4: Water loss relative to shoot dry weight and drought tolerance in 35S-bHLH39 transgenic lines

Entry	Water lost in 3d/shoot DW d4	Drought tolerance (% of WT)	Shoot DW d4 (g)
34-1	125±3	128%	0.56±0.01
97-7	140±5	119%	0.51±0.02
94-10	150±5	113%	0.46±0.02
65-1	153±6	112%	0.47±0.02
30-2	162±4	106%	0.43±0.01
1-7	166±3	104%	0.43±0.01
WT	173±5	100%	0.40±0.01

Example 12: Constitutive expression of At-bHLH101 in *Arabidopsis* results in reduced flower abortion following heat stress

An experiment was set up with 17 transgenic 35S-AtbHLH101 lines and a WT control. Plants were grown in 2.25 inch pots under optimal conditions (22C, 18hr light of 200uE, 60% RH) in a growth chamber until three days post-appearance of the first flower. A heat stress treatment was applied by placing plants at 42C for 1.75 hours. One week following the stress period the plants were assessed for number of aborted flowers. The results are shown in Table 5 and show ten of the transgenic lines had at least 10% reduced flower abortion relative to WT controls and two lines had over 50% reduced flower abortion relative to WT controls.

Table 5: Flower abortion following heat stress (n=12)

Entry	# of aborted flowers	Flower abortion (% of WT)
61-1	1.3±0.3	41%
88-3	1.3±0.4	43%
18-1	2.0±0.5	65%
90-1	2.3±0.5	73%
15-2	2.5±0.5	81%
49-5	2.5±0.5	81%
19-4	2.7±0.4	86%
70-8	2.7±0.3	86%
75-2	2.7±0.5	86%
16-2	2.8±0.4	89%
39-11	2.8±0.4	92%
11-1	2.9±0.4	95%
81-3	3.0±0.4	97%
79-6	3.1±0.4	100%
97-2	3.1±0.4	100%
45-2	3.2±0.5	103%
9-1	3.2±0.4	103%
WT	3.1±0.3	100%

Example 13: Constitutive expression of AtbHLH101 in *Arabidopsis* results in drought tolerance of the plants

Drought tolerance was assessed in ten transgenic At-bHLH101 lines that also showed reduced flower abortion following heat stress. Plants were grown (5 per 3 inch pot) under optimal conditions in a growth chamber (22C, 18hr light of 200uE, 60% RH) until the first open flower. Drought treatment was applied then by watering up all plants to the same saturated level. Further water was withheld. Plants were weighed daily to determine the daily water loss and all plants were harvested on day four of treatment, at which time all plants were visibly wilting. The water loss relative to final shoot biomass was calculated and is a representative indicator of

drought tolerance. Data was normalized to WT which was set as 100% (Table 6). Eight of the ten lines examined showed some degree of drought tolerance and one line had a statistically significant 27% greater drought tolerance relative to WT. This line was also one of the best four performing heat tolerant lines as seen by reduced flower abortion.

Table 6: Water loss relative to shoot dry weight and drought tolerance in 35S-bHLH101 transgenic lines

entry	Water lost in 3d/shoot DW d4	Drought tolerance (% WT)	Shoot DW d4 (g)
90-1	147±5	127%	0.48±0.02
16-2	174±6	114%	0.41±0.02
61-1	176±5	113%	0.40±0.01
70-8	177±7	112%	0.41±0.02
6-5	181±3	110%	0.39±0.01
11-1	187±7	107%	0.38±0.02
45-2	189±8	106%	0.37±0.02
19-4	191±5	105%	0.37±0.01
79-6	203±7	100%	0.35±0.01
82-4	203±3	99%	0.34±0.01
WT	202±6	100%	0.34±0.01

Example 14: Constitutive expression of At-bHLH39 in *Arabidopsis* results in increased seed yield relative to a wild type control following heat stress

Plants are grown (3 per 3" pot) under optimal conditions in a growth chamber (22C, 18hr light of 200uE, 60% RH) until flowering. At flowering plants are split into two groups where the first group is exposed to heat stress (all plants flowered within a couple of days) and the second group is maintained under optimal conditions until maturity. The heat stress treatment consists of a daily exposure to 45C. Temperatures are ramped from 22 to 45C over a one hour period and maintained at 45C for a time period of 2hr to 3 hr. Daily heat stress treatments are applied for a period of 10 days. Following the heat stress treatments plants are returned to optimal conditions and grown to maturity. All plants are harvested at maturity and final seed yield per pot is

determined. By comparing the yield (as % of optimal) of transgenic plants to that of WT the degree of yield protection is calculated.

Table 7: SEQUENCE ID REFERENCE CHART

SPECIES	SEQ ID NO:	bHLH	Reference	Seq. type	Length
ARABIDOPSIS THALIANA	1	bHLH39	AT3G56980	nucleotide	777
AVENA SATIVA	2	bHLH39	CN817002	nucleotide	222
BRACHYPODIUM DISTACHYON	3	bHLH39	super_13.506_gen	nucleotide	2091
BRACHYPODIUM DISTACHYON	4	bHLH39	super_13.506_cds	nucleotide	729
BRASSICA NAPUS	5	bHLH39	EE515575	nucleotide	587
BRASSICA NAPUS	6	bHLH39	TC84782	nucleotide	595
BRASSICA NAPUS	7	bHLH39	TC88840	nucleotide	631
BRASSICA RAPA	8	bHLH39	Contig2	nucleotide	693
GLYCINE MAX	9	bHLH39	TC269627	nucleotide	723
HORDEUM VULGARE	10	bHLH39	AK251746	nucleotide	738
PANICUM VIRGATUM	11	bHLH39	Contig2	nucleotide	723
SOLANUM LYCOPERSICUM	12	bHLH39	DV105842	nucleotide	351
TRITICUM AESTIVUM	13	bHLH39	TC358765	nucleotide	714
TRITICUM AESTIVUM	14	bHLH39	TC343683	nucleotide	617
TRITICUM AESTIVUM	15	bHLH39	TC300244	nucleotide	434
TRITICUM AESTIVUM	16	bHLH39	CA618726	nucleotide	261
ZEA MAYS	17	bHLH39	TC429418	nucleotide	228
ARABIDOPSIS THALIANA	18	bHLH39	AT3G56980	protein	258
AVENA SATIVA	19	bHLH39	CN817002	protein	158
BRACHYPODIUM DISTACHYON	20	bHLH39	super_13.506_ORF	protein	242
GLYCINE MAX	21	bHLH39	TC269627	protein	241
HORDEUM VULGARE	22	bHLH39	AK251746	protein	246
PANICUM VIRGATUM	23	bHLH39	Contig2	protein	241
SOLANUM LYCOPERSICUM	24	bHLH39	DV105842	protein	117
TRITICUM AESTIVUM	25	bHLH39	TC358765	protein	238
TRITICUM AESTIVUM	26	bHLH39	TC343683	protein	206
TRITICUM AESTIVUM	27	bHLH39	TC300244	protein	144
TRITICUM AESTIVUM	28	bHLH39	CA618726	protein	87
ARABIDOPSIS THALIANA	29	bHLH38	AT3G56970	nucleotide	762
BRASSICA NAPUS	30	bHLH38	TC95626	nucleotide	753
BRASSICA OLERACEA	31	bHLH38	AM061155	nucleotide	765
BRASSICA RAPA	32	bHLH38	EX134222	nucleotide	435
CICER ARIETINUM	33	bHLH38	FE670123	nucleotide	300
HORDEUM VULGARE	34	bHLH38	TC164142	nucleotide	501
HORDEUM VULGARE	35	bHLH38	BAF30424.1	nucleotide	759
MEDICAGO TRUNCATULA	36	bHLH38	TC127269	nucleotide	767
MEDICAGO TRUNCATULA	37	bHLH38	TC115041	nucleotide	780
MEDICAGO TRUNCATULA	38	bHLH38	AJ496888	nucleotide	300
PANICUM VIRGATUM	39	bHLH38	Contig1	nucleotide	690
POPULUS	40	bHLH38	TC89850	nucleotide	549
POPULUS	41	bHLH38	EEF05011.1	nucleotide	795
POPULUS	42	bHLH38	EEE91492.1	nucleotide	474
RICINUS COMMUNIS	43	bHLH38	EEF30834	nucleotide	774

SPECIES	SEQ ID NO:	bHLH	Reference	Seq. type	Length
RICINUS COMMUNIS	44	bHLH38	EEF30835	nucleotide	555
SOLANUM LYCOPERSICUM	45	bHLH38	TC194645	nucleotide	717
SORGHUM BICOLOR	46	bHLH38	TC117663	nucleotide	642
TRITICUM AESTIVUM	47	bHLH38	TC337566	nucleotide	672
TRITICUM AESTIVUM	48	bHLH38	CA650144	nucleotide	390
TRITICUM AESTIVUM	49	bHLH38	CA502657	nucleotide	459
VIGNA UNGUICULATA	50	bHLH38	FF388259	nucleotide	732
VITIS VINIFERA	51	bHLH38	CAO17950.1	nucleotide	1563
VITIS VINIFERA	52	bHLH38	CAN79614	nucleotide	735
ARABIDOPSIS THALIANA	53	bHLH38	AT3G56970	protein	253
BRASSICA NAPUS	54	bHLH38	TC95626	protein	251
BRASSICA OLERACEA	55	bHLH38	AM061155	protein	255
BRASSICA RAPA	56	bHLH38	EX134222	protein	145
CICER ARIETINUM	57	bHLH38	FE670123	protein	100
HORDEUM VULGARE	58	bHLH38	TC164142	protein	167
HORDEUM VULGARE	59	bHLH38	BAF30424.1	protein	252
MEDICAGO TRUNCATULA	60	bHLH38	TC127269	protein	246
MEDICAGO TRUNCATULA	61	bHLH38	TC115041	protein	260
MEDICAGO TRUNCATULA	62	bHLH38	AJ496888	protein	100
ORYZA SATIVA	63	bHLH38	NP_001045424.1	protein	247
PANICUM VIRGATUM	64	bHLH38	Contig1	protein	230
POPULUS	65	bHLH38	TC89850	protein	183
POPULUS	66	bHLH38	EEF05011.1	protein	264
POPULUS	67	bHLH38	EEE91492.1	protein	158
RICINUS COMMUNIS	68	bHLH38	EEF30834.1	protein	257
RICINUS COMMUNIS	69	bHLH38	EEF30835.1	protein	184
SOLANUM LYCOPERSICUM	70	bHLH38	TC194645	protein	239
SORGHUM BICOLOR	71	bHLH38	TC117663	protein	214
TRITICUM AESTIVUM	72	bHLH38	TC337566	protein	224
TRITICUM AESTIVUM	73	bHLH38	CA650144	protein	130
TRITICUM AESTIVUM	74	bHLH38	CA502657	protein	153
VIGNA UNGUICULATA	75	bHLH38	FF388259	protein	244
VITIS VINIFERA	76	bHLH38	CAN64266.1	protein	245
VITIS VINIFERA	77	bHLH38	CAO17950.1	protein	520
VITIS VINIFERA	78	bHLH38	CAN79614.1	protein	244
ARABIDOPSIS THALIANA	79	bHLH101	AT5G04150	nucleotide	723
BRASSICA OLERACEA	80	bHLH101	AM060621	nucleotide	546
BRASSICA RAPA	81	bHLH101	Contig1	nucleotide	678
ORYZA SATIVA	82	bHLH101	CI296230	nucleotide	261
ORYZA SATIVA	83	bHLH101	TC345105	nucleotide	450
ARABIDOPSIS THALIANA	84	bHLH101	AT5G04150	protein	240
BRASSICA OLERACEA	85	bHLH101	AM060621	protein	182
BRASSICA RAPA	86	bHLH101	Contig1	protein	226
ORYZA SATIVA	87	bHLH101	CI296230	protein	87
ORYZA SATIVA	88	bHLH101	TC345105	protein	150
ARABIDOPSIS THALIANA	89	bHLH100_2	AT2G41240.2	nucleotide	726
ARABIDOPSIS THALIANA	90	bHLH100_1	AT2G41240.1	nucleotide	729

SPECIES	SEQ ID NO:	bHLH	Reference	Seq. type	Length
ORYZA SATIVA	91	bHLH100_1	TC340917	nucleotide	621
PANICUM VIRGATUM	92	bHLH100_1	FL920216	nucleotide	711
SORGHUM BICOLOR	93	bHLH100_1	TC113263	nucleotide	732
TRITICUM AESTIVUM	94	bHLH100_2	TC317240	nucleotide	417
TRITICUM AESTIVUM	95	bHLH100_1	TC303529	nucleotide	705
TRITICUM AESTIVUM	96	bHLH100_1	CD865039	nucleotide	691
ZEA MAYS	97	bHLH100_1	TC409749	nucleotide	465
ARABIDOPSIS THALIANA	98	bHLH100_2	AT2G41240.2	protein	241
ARABIDOPSIS THALIANA	99	bHLH100_1	AT2G41240.1	protein	242
ORYZA SATIVA	100	bHLH100_1	TC340917	protein	207
PANICUM VIRGATUM	101	bHLH100_1	FL920216	protein	237
SORGHUM BICOLOR	102	bHLH100_1	TC113263	protein	244
TRITICUM AESTIVUM	103	bHLH100_2	TC317240	protein	139
TRITICUM AESTIVUM	104	bHLH100_1	TC303529	protein	235
ZEA MAYS	105	bHLH100_1	TC409749	protein	155

Sequences

>SEQIDNO:1

ATGTGTGCATTAGTACCTCCATTGTTTCCAAACTTTGGGTGGCCATCAACGGGAGAGTACGACAGCTACTACCTCGC
CGGAGATATCCTCAACAACGGCGGGGTTTCTTGATTTTCCGGTACCGGAGGAGACTTATGGAGCTGTTACAGCGGTGA
CTCAACATCAGAATAGCTTTGGTGTTTCTGTTTCGTCGGAGGGAAATGAAATAGACAACAATCCGGTGGTCGTCAAG
AAGCTTAATCACAATGCTAGTGAGCGTGACCGTCGCAGGAAAATTAACTCTTTGTTCTCATCTCTCCGTTCATGTCT
TCCTGCCTCTGGCCAATCGAAGAAGCTAAGCATTCCTGCGACGGTTTCTCGAAGCTTGAAGTACATACCAGAGCTGC
AAGAGCAAGTGAAGAAGCTAATAAAAAAGAAGGAAGAGCTCTTGGTGCAAATTTCAGGTCAAAGAAACACTGAATGT
TACGTTAAGCAGCCACCAAAGGCCGTCGCGAATTATATCTCGACCGTTTCTGCGACTAGGCTTGGTGACAACGAAGT
GATGGTCCAAATCTCATCGTCCAAGATTCATAACTTTTCGATATCTAATGTTTTAAGTGGGTTAGAAGAAGATAGGT
TTGTTCTTGTGGACATGTCATCTTCAAGGTCTCAAGGAGAAAGGCTTTTCTACACTTTGCATTTACAAGTGGAGAAG
ATTGAAAATTACAAGCTGAATTGCGAAGAGTTAAGTCAGAGGATGTTGTACTTGTATGAGGAATGTGGAAACTCATA
TATATGA

>SEQIDNO:2

GCACGAGGCCTCCCTCCGCTCCCTCCTCCCCGACACCGATCACAGCAAGAAGCTGAGCATCCCCATCACTGTGACGC
GGGTGCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACACGCTGGAGAAGAAGAAGGAAGAGCTGACCCAGGCG
AACTGCAAACCAGGAGTTGTGGCCATGAAGGAGAACACGGCTCCGATCGTGTCCGCCACCTGCCTCGA

>SEQIDNO:3

ATGGGGCACAAGCAGCTGTTCGTGGACGACCCGTTCGCGAGCAGCATCTCGTCTCTGGAGGCGGAGGCCATCTTCTC
CGGCGCCGGCGGGCAGTGGCGCGCCGGCGGCGGCCTCGACGACCGTGACCTCTCCGCCATGCCGGCGGCGGCCAACA
CCTCGTCGGGCGGCTCCGGCTCTCCCGGCGGCGGCGGCAGGAAGATGAGCCACAACGCGTACGAGCGCGACCGCCGC
AAGCAGCTCAACGAGCTCTACTCCTCCCTCCGCTCCCTCCTCCCCGACGCCGACCACACCGTATGCAAATCAAATTG
AAGCCATAGATCATAATTTGATCCTGAATCCTGATGGATCTGGTGATGATTTGACTAATTGCAGAAGAAGCTGAGCA
TCCCGATCACAGTGTCGCGGGTGCTAAAGTACATCCCGGAGCTGCAGAAGGAAGTGGACGGGCTGGAGAGGAAGAAG
GAGGAGCTGACGCGCGCCAACTGCAAGCCCGGGGGTGATCGCCATGAAGGACCAGAACGTTGCTCCCGTCGTCTCCGC
GACCTGCCTCGACGACAAGGACATCATGGTTCAGGTCAGCTTGCTCAGCGGCATGGCGGCGGCGGCTCTGCCGATGT
CCACGTGCATAAAGATTCTGGAGAACGAAGGTCTTCGCCTCGTCAGCTCGTCCACTTCTGCCTTTGGGAACAGGACG
TTCTATAACCTCCATCTTCAGGTAATTGGTACATCTGTCTGCATGAAGCCTTAATTTCCTATTGGTAATTATCAATG
TCATCGATCCATGCTTGCTCGATTCATTTTGGCAAATTTGCCTCATACTTACTCGACATCGTAGTAGAAGAGAAGGA
AAAAAAAAAGCCCAGCATTTTTCTTTGAGAAACAAAGCACAACATTTTCTGCTCCCAAGATCCATTCCGAAAGCCGG
GTGCACAACACTACGAGTAAAAATCGTTTCCTCTTAGTTATATAACGTGTGGTACCTTCAACTTAAATGAGCATCAG
TTGATGCAGAGTGGTACCCCTGTTCGGAAAAGACTTCAACATGGCACCATCTTCACTAGAGGTCCCCCAGCCCCTAA
TTCAGACGAAGGCATCATTTTAGTCCATTAACCCCGGTAGGTTGAGATTACAAAGGCTAGATTTATATAACGGTTTT
GCTATTTCTAAGGCGACTGAGAAATAAGTATTTCTAACAACCATTTGATCCAATCACTGAGATTCGTGCAAGATCCA
TACATATACAAAGTATTAGAAAACCTTAATCGGATGGATATGATCGTTGACTGAGAAATAGACACTTCTTTTATGTA
AAGGGTGTTTACTATATATATGTGGGTTTTGATATACTCATATTGAATAGATTTGAAGATCATGTCAGGCCCTACTT
AGGGTAATCTGTTGTGAAATTAAACTGTGTTCATAGGGCAAAAACATTGTCTTATAAATCAGCACAAAATCAACGAA
TTGGGAGTTTTTACGTAACAAATATAAATTGTAGCAACACAAATTAATTGGCTACAATACAATCCAAGAACAACAAG
ACGAGTATACACGGCAAACGATCATGCATATGGTGAGTTGGTGACCAGATCACCCGCTCTACTAGATGCTCCTAGAT
GCATATGGTGCCATAATTTTAGGAAAACCGAACTGGAAGAAGCACCTAAAAAACGAGTATGAAATCATCAATCGACC
AAGATTAGAGGTCGGTCTCATCATCGACAACATGAACAATTAAGATGCATGCTGGAGAACAGATAATCTAACACAGC
CACAGGTTTATTACAAAAAGCTTAACAGAAACTTCGCTAAAGCAACCAAAGAATGAGAACAAAAAAATATATCTTCT
AATAACATGTGTGTGCTGTTGTATAGCATCTGAAGGCGTAATGCGAAACTCTAATTTATCTGAAGTATGTAGTGCTT
ATATGCTTATATAACATGTAAATAAGCAATATATATTCATTAATTTCATTTAATTTGTATACTGAAACAGAGAAACC
AGCGAACGATGAGCAAGGAGTGCCCAGCGTTCTGTGACGAGCTGGAGAAAGCCATCAAGAAAAAGGCAGGACTGCAT
ATGCATCAGTGA

>SEQIDNO:4

ATGGGGCACAAGCAGCTGTTCGTGGACGACCCGTTCGCGAGCAGCATCTCGTCTCTGGAGGCGGAGGCCATCTTCTC
CGGCGCCGGCGGGCAGTGGCGCGCCGGCGGCGGCCTCGACGACCGTGACCTCTCCGCCATGCCGGCGGCGGCCAACA

CCTCGTCGGGCGGCTCCGGCTCTCCCGGCGGCGGCGGCAGGAAGATGAGCCACAACGCGTACGAGCGCGACCGCCGC
AAGCAGCTCAACGAGCTCTACTCCTCCCTCCGCTCCCTCCTCCCCGACGCCGACCACACCAAGAAGCTGAGCATCCC
GATCACAGTGTCGCGGGTGCTAAAGTACATCCCGGAGCTGCAGAAGGAAGTGGACGGGCTGGAGAGGAAGAAGGAGG
AGCTGACGCGCGCCAACTGCAAGCCCGGGGTGATCGCCATGAAGGACCAGAACGTTGCTCCCGTCGTCTCCGCGACC
TGCCTCGACGACAAGGACATCATGGTTCAGGTCAGCTTGCTCAGCGGCATGGCGGCGGCGGCTCTGCCGATGTCCAC
GTGCATAAAGATTCTGGAGAACGAAGGTCTTCGCCTCGTCAGCTCGTCCACTTCTGCCTTTGGGAACAGGACGTTCT
ATAACCTCCATCTTCAGAGAAACCAGCGAACGATGAGCAAGGAGTGCCCAGCGTTCTGTGACGAGCTGGAGAAAGCC
ATCAAGAAAAAGGCAGGACTGCATATGCATCAGTGA

>SEQIDNO:5

AAAAAAAAATTAGAAAAGAAGAAAAGAGTTTATCGGGTCTCTCTCACGAGTCACGGCGTCGACGAAACTGGAGGTGA
AGCCATCTCCGATCGAGAGTCGCCGTCATCCCTATCCCACTCCTGCAATATCTACTCCGACGCTGTCGTCGTTAATT
TCGGCACCTCCTCGCCAAGTCGCCATCGTCCCTGTCGTGTGCGAGCTGTCACTGTCCATGCAACGAATAAATACTCT
GTGTCTGCATCGCTTTGACAGTGCTTGGTGCTTCTATGATATGTCCTCCGGTTTATGCTTTGCCCAGATTCGGAAGA
AGATGGGTTTGTTCTTGTTGATGTTTCATCTTCTAGGTCTCATGGAGAAAGGCTCGTCTACAGTTTGCATCTTCAAA
TGGGAAACATAAATAATCACGAGCTGACGTGCGAAGAGCTAAGCCAGAGAATGTATACTTGTATGAGGAATGCGGAA
ACTCGTTTAGATGATAATCTGTTCTTGTTTCTTTTAGTTATGTCATCTGTTTCTCAACATGTAACATTCATGTAGCC
GAGTTGTTTCGTTTATTTTTCTTGTCGAAATCAATGATCGATTATCGC

>SEQIDNO:6

GACGGCGTCGACGAAACTGGAGGTGAAGCCATCTCCGATCGAGAGTCGCCGTCATCCCTATCCCACTCCTGCAATAT
CTACTCCGACGCTGTCGTCGTTAATTTCGGCACCTCCTCGCCAAGTCGCCATCGTCCCTGTCGTGTAAGCCCTAGGT
CTGATTTGTCTTTCCTTTTAAATCGACGAAATTGAAATTGTGCGAGCTGTCACTGTCCATGCAACGAATAAATACTC
TGTGTCTGCATCGCTTTGACAGTGCTTGGTGCTTCTATGATATGTCCTCCGGTTTATGCTTTGCCCAGATTCGGAAG
AAGATGGGTTTGTTCTTGTTGATGTTTCATCTTCTAGGTCTCATGGAGAAAGGCTCGTCTACAGTTTGCATCTTCAA
ATGGGAAACATAAATAATCACGAGCTGACGTGCGAAGAGCTAAGCCAGAGAATGTATACTTGTATGAGGAATGCGGA
AACTCGTTTAGATGATAATCTGTTCTTGTTTCTTTTAGGTTATGTCATCTGTTTCTCAACATGTAACATTCATGTAG
CCGAGTTGTTTCGTTTATTTTCTTGTCGAAATTCAATGATCGATTATCACATTTAC

>SEQIDNO:7

GACGACAAAAAAAAAAATTAGAAAAGAAGAAAAGAGTTTATCGGGTCTCTCTCACGAGTCACGGCGTCGACGAAACT
GGAGGTGAAGCCATCTCCGATCGAGAGTCGCCGTCATCCCTATCCCACTCCTGCAATATCTACTCCGACGCTGTCGT
CGTTAATTTCGGCACCTCCTCGCCAAGTCGCCATCGTCCCTGTCGTGTAAGCCCTAGGTGCGAGCTGTCACTGTCCA
TGCAACGAATAAATACTCTGTGTCTGCATCGCTTTGACAGTGCTTGGTGCTTCTATGATATGTCCTCCGGTTTATGC
TTTGCCCAGATTCGGAAGAAGATGGGTTTGTTCTTGTTGATGTTTCATCTTCTAGGTCTCATGGAGAAAGGCTCGTC
TACAGTTTGCATCTTCAAATGGGAAACATAAATAATCACGAGCTGACGTGCGAAGAGCTAAGCCAGAGAATGTATAC
TTGTATGAGGAATGCGGAAACTCGTTTAGATGATAATCTGTTCTTGTTTCTTTTAGTTATGTCATCTGTTTCTCAAC
ATGTAACATTCATGTAGCCGAGTTGTTTCGTTTATTTTCTTGTCGAAATCAATGATCGATTATCACATTTACATGCT
ACTGAATATTGACTG

>SEQIDNO:8

ACGACAAAAAAAAATTAGAAAAGAAGAAAAGAGTTTATCGGGTCTCTCTCACGAGTCACGGCGTCGACGAAACTGGA
GGTGAAGCCATCTCCGATCGAGAGTCGCCGTCATCCCTATCCCACTCCTGCAATATCTACTCTGACGCTGTCGTCGT
TAATTTCGGCACCTCCTCGCCAAGTCGCCATCGTCCCTGTCGTGTAAGCCCTAGGTCTGATTTGTCTTTCCTTTTAA
ATCGACGAAATTGAAATTGTGCGAGCTGTCACTGTCCATGCAACGAATAAATACTCTGTGTCTGCATCGCTTTGACA
GTGCGTGGTGCTTCTATGATATGTCCTCTGGTTTATGCTTTGCCCAGATTCGGAAGAAGATGGGTTTGTTCTTTGTT
GATGTTTCATCTTCTAGGTCTCATGGAGAAAGGCTCGTCTACAGTTTGCATCTTCAAATGGGAAACATAAATAATCA
CGAGCTGACGTGCGAAGAGCTAAGCCAGAGAATGTATACCTGTATGAGGAATGCGGAAACTCGTTTAGATGATAATC
TGTTCTTGTTTCTTTTTGTTATGTCGTCTGTTTCTCAACATGTAACATTCATGTAGCCGAGTTGTTTCGTTTATTTT
CTTGTCGAAATCAATGATCGATTATCACATTTACATGCTACTGAATATTGACTGATTACTATGAAATTCACTAATAT

>SEQIDNO:9

ATGGTTGCTTTGTTTTCCCCTCCGGTGTTCTCAACCAAGGGATGGTTCTTAGAAGAAGAGCCATTAAGCTATGATGT
GTCTTCAGATTACTCATTTCCCTATCAATTTTTTGCACCACAGACACAGATTGAACTTGAAATAGAAAGGTCCTCTG
CACCATCCCCTGAAGACCCTGCCATGGTCAAAAAGCTTAGCCACAACGCTAGTGAACGTGATCGCCGCAAGAAGGTT
AATGACTTGGTTTCTTCACTTCGTTCACTTCTTCCTGGGCCAGATCAAACGAAAAAAATGAGCATTCCAGCTACAGT
TTCGCGAGTTTTAAAATACATACCTGAGTTACAACATCAAGTGCAAGCACTAACTAAGAAAAAAGAGGAGCTTCTGT
GCAGAATTTCAAAAAATCTCAAAGGAGATTCGGTGAACAAAGAATCTCAAAGGAGAATTTCCCATCACAATTCTGAT
TTTGCTGTTTCAACTAGTAGGCTCAACGATTGTGAAGCTGTTGTTCACATTTCCTCTTATGAGGCTCACAAGGCTCC
ACTATCCGACATCTTGCAATGTTTAGAAAATAATGGCCTTTATTTGCTAAATGCTTCTTCCTCTGAAACTTTTGGAG
GAAGGGTCTTCTACAACTTGCATTTCCAGGTGGAAAAAACTCATAGATTAGAGTCCGAAATTCTAACTGAGAAGCTT
TTGTCAATATATGAGAAGCAAAGGATTTTC

>SEQIDNO:10

ATGGGGCACCAGACCCAGATGTTCGACGACCCGTTCGCGAGCAGTATGTCGTCCCTGGACGCAGACATCTTCTCCGT
CGCCGGCGGCCTCCACCCATCGCAGTGGCCGGGACTCGACCACGACGTCTCGCTGGCGCCGGCTGCCAACAACGGCA
CCTCCTCCGGCGGCTACGGCTCCCCCGGGGGCGGCGATGGCTCGGGCTCCCACCGCAAGATCAGCCACAACGCGTAC
GAGCGCGACCGCCGCAAGCAGCTCAACGAGCTCTACTCCGACCTCCGCTCCCTCCTCCCCGACTCCGATCACACCAA
GAAGCTGAGCATTCCGATCACGGTGTCGCGGGTGCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACGGACTGG
AGAAGAAGAAGGAGGAGCTTACGAGGGCCAGCTGCAAGCCAGGCGTATTGACCATGAAGGAGAACACGGTCCCGATC
GTGTCCGCCACCTGCCTCGACGAAAGGGAGATCATGGTCCAGGTTAGCTTGGTGAGCACCATGGCCGGAGCTCTGCC
CATGTCCAAGTGCATCAAAGTGCTGGAGAACGAAGGCCTCCGCCTCATCAGCTCGTCCACTTCTGCTTTCCAGAACA
GGACGTTCTATAGCCTCCATCTTCAGAGAACCCAACGGACGATGAGCAAAGAGTGTCCGGCATTTTGTGAAGAACTG
GAGAATGCCCTGACGCAGAAGGCGGGACTACGTCTACATCACCAG

>SEQIDNO:11

ATGACCCCTTCTCGAGCAGCATCTCGTCGCTGGAGGCGGACATCTTCTCCGCCGGCGGCCAGCGGCGCGTCGCCGCC
GTGGCCGGACCTCGAACTCGACCTCGACCTCGACGACGACGACATCCACGACCTCTCCGCGCCGGCGGCCAACGCCA
CCTCCTCAGGAGGCTATGGCTCGGGCGGAGGCTCCGGCGGCTCCCACAGGAAGCTCAGCCACAACGCGTACGAGCGC
GACCGCCGGAAGCAGCTCAACGAGCTCTACTCCTCGCTCCGCTCCCTCCTCCCCGACGCTGACCACACTAAGAAGCT
GAGCATCCCCACCACGGTCTCCCGAGTTCTCAAGTACATCCCCGAGCTGCAGAAGCAGGTGGACAACCTGGAGAGGA
GGAAGAAGGAGCTGACGAACGCCAACTGCAAACCAGGAGTTCTGAAGACGAGCCAGATTGTAACTCCCATTGTTTCT
GCTACCTGCCTCAACGATACGGAGATCATGGTTCAGGTCAGCCTGCAGAGCAATGTGGCTGCCACAAGTCTTCCTCT
GTCCAAGTGCATAAAAGTGCTGGAGAACGAAGGCCTTCACCTGATTAGTTCATCAACTTACTCCACCTTCGACAACA
GGACATTCTATAGCCTCCATCTTCAGAGAAGTCAAAGAACGATGAAGGAGGAGTGCCCAGCATTCTGCGATGAACTG
GAGAGGATTATCAAGAAGAAAGCAGGGGCG

>SEQIDNO:12

ATGTTAGCCATTTCTTCTTCTTCTCCTCCTTTATTTTCTACTACTACTAATAATTTTGGTTGGCTTTTGGAAGATCT
TATAAGCCATGAATTAACAAATAGTGGAGAAACTTCAAATTCATCTCAAAAAAGCCTTCAACATTGTGATTCAAATA
AATTTGATCAAATTATTATCAACAGTGGTGATCAGTATCAACCTGATCAGACGGTTAAGAAGCTTAATCATAACGCA
AGTGAACGTGACCGTAGAAAGAAAATCAACAGCTTATATTCTTCTCTTCGTTCTTTACTACCTCCTTCTGATCATAC
GAAAAAGCTAAGCATTCCATCAACAGTATCAAGAATTCTAAAG

>SEQIDNO:13

ATGGGGCACCAGCACCAGATGTTCAACGACCCCTTCGCGAGCAGCATGTCGTCACTGGAGGAAGACATGTTCTCCGG
TGCCGGAGGCTACCACCACCTCACGCCGTCCATGCAGTGGCCGGGCTTGGATAACGACATACCGTCGGCGCCGGCTG
CCAACAACGCCACCTCCTCCGGTGGCTCTGGATCACACCGCAAGATGAGTCACAACGCGTACGAACGTGACCGCCGC
AAGCAGCTCAACGAGCAATATTCCTCCCTCCGCTCCCTCCTCCCCGACGATGATCACACCAAGAAGATGAGCATTCC
GACCACGGTGTCGCGGGTGATCAACTACATCCCGGAGCTGCAGAAGGAGGTAGACCGCCTGGAGAAGAAGAAGGAGG
AGCTGAGGCGGGGCAGCTGCGAGCAAGGCGCCATGAGGCAGAACACGGCCCCGATCGTGTCCGCCACCTGCCTCGAC
GACAGGGAGATCATGGTCCAGGTCAGCCTGGTGAGCACCATGGCGGGAGCTCTGCCCATGTCCAAGTGCATCAAGGT

GCTGGAGAACCAAGGCCTTCGCCTCATAAATTCCTCGACTTCCGCGTTTCAGAACAGGACGTTCTACAGCCTCCATC
TTCAGAGAACCCAACGGACAATGAGCAAGGAGGGCCAAACATTTTGTAACGAATTGGAGAACGCTGTGAAGCAAAAG
GCGGGACTACATCTACATCAT

>SEQIDNO:14

ATGGGGCACCAGCACCAGATGTTCGAAGACCCGTTCGCGAGCAGCATATCGTCGCTGGAGGCCGAGATATTCTCCGT
CGCCGGCGGCCACCACCATACGCAGTGGCCGGGCCTCGACCACGACATCCCGCTGGCCCCGGCTGCCAATAACGGCA
CCTCCTCCGGCGGCTACGGCTCCCCCGGGGGCGGCGATGGCTCGGGCTCCCATCGCAAGATCAGCCACAACGCCTAC
GAGCGCGACCGCCGCAAGCAGCTCAACGAGCTCTACTCCGACCTCCGCTCCCTCCTCCCCGACACCGATCACACGAA
GAAGCTTAGCATTCCGATCACGGTGTCGCGGGTGCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACGGCCTGG
AGAAGAAGAAGGAGGAGCTGACGCGCGCCAGCTGCAAGCCCGGCGTGCTGACCATGAAGGGGGACACGGCTCCGATC
GTGTCCNGCCACTGCCTCGACGACAGGGAGATCATGGTCCANNGTCAGCTGGTGAGCACCATGGGCGGAGTCTGCCA
TGTCAAGTGCTCAGGTGCTGAGACGAAGGCTCGGCTCATAGTCGTCACTCCGGTTCAGACAGACTCTATATTCATCT
C

>SEQIDNO:15

ATGGGGCACCAGCACCAGATGTTCGAAGACCCGTTCGCGAGCAGCATATCGTCGCTGGAGGCGGACATCTTCTCCGT
CGCCGCCGGCCACCACCATCCGCAGTGGCCGGGCCTCGACCACGACGTCCCGTTGGCGCCGGCTGCCAACAACGGCA
CATCCTCCGGCGGCTACGGCTCCCCCGGTGGCGGCGACGGCTCGGGCTCCCACCGCAAGATCAGCCACAACGCGTAC
GAGCGCGACCGCCGCAAGCAGCTCAACGAGCTCTACTCCGACCTCCGCTCCCTCCTCCCCGACACCGATCACACGAA
GAAACTGAGCATTCCGATCACGGTGTCGCGGGTGCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACGGACTGG
AGAAGAAGAAGGAGGAGCTGACGCGCGCCAACTGCAGCCCGGCGTGCTG

>SEQIDNO:16

AACAACGCCACCTCCTCCGGCGGCTCTGGATCACACCGAAAGATGAGTCACAACGCGTACGAGCGTGACCGCCGCAA
GCAGCTCAACGAGCAATATTCCTCCCTCCGCTCCCTCCTCCCCGATGACGACCACAATAAGAAGATGAGCATTCCGA
CCACGGTGTCGCGGGTGATCAAGTACATCCCGGAGCTGCAGAAGGAGGTAGACGGCCTGGAGAAGAAGAAGGAGGAG
CTGAGGCGGGCCAGCTACGAGCAAGCGCCA

>SEQIDNO:17

CGTGTCCCGAGTTCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACAACCTGGAGAGGAGGAAGAAGGAGCTGA
CCAACGCCAACTGCAAGCCGGGAGTTCTGAAAACCACCAAGGCCGTAACTCCCATTGTTTCTGCTACCTGCCTCAAC
GACACGGAGATCATGGTTCAGGTCAGCCTGCACAGCGATGTGGCCGCCACAGCTCTCCCTCTCTCCAAGTGCAT

>SEQIDNO:18

MCALVPPLFPNFGWPSTGEYDSYYLAGDILNNGGFLDFPVPEETYGAVTAVTQHQNSFGVSVSSEGNEIDNNPVVVK
KLNHNASERDRRRKINSLFSSLRSCLPASGQSKKLSIPATVSRSLKYIPELQEQVKKLIKKKEELLVQISGQRNTEC
YVKQPPKAVANYISTVSATRLGDNEVMVQISSSKIHNFSISNVLSGLEEDRFVLVDMSSSRSQGERLFYTLHLQVEK
IENYKLNCEELSQRMLYLYEECGNSYI

>SEQIDNO:19

RHEASLRSLLPDTDHSKKLSIPITVTRVLKYIPELQKQVDTLEKKKEELTQANCKPGVVAMKENTAPIVSATCLDDR
DIMVQVSLLSNMAGALPVSKCIKVLENEGLRLVSSSTSAFQNKTFYSLHVQRTQRTISKVCPAFCDELENAIKRAGM
RLQQ

>SEQIDNO:20

MGHKQLFVDDPFASSISSLEAEAIFSGAGGQWRAGGGLDDRDLSAMPAAANTSSGGSGSPGGGGRKMSHNAYERDRR
KQLNELYSSLRSLLPDADHTKKLSIPITVSRVLKYIPELQKEVDGLERKKEELTRANCKPGVIAMKDQNVAPVVSAT

CLDDKDIMVQVSLLSGMAAAALPMSTCIKILENEGLRLVSSSTSAFGNRTFYNLHLQRNQRTMSKECPAFCDELEKA
IKKKAGLHMHQ

>SEQIDNO:21

MVALFSPPVFSTKGWFLEEEPLSYDVSSDYSFPYQFFAPQTQIELEIERSSAPSPEDPAMVKKLSHNASERDRRKKV
NDLVSSLRSLLPGPDQTKKMSIPATVSRVLKYIPELQHQVQALTKKKEELLCRISKNLKGDSVNKESQRRISHHNSD
FAVSTSRLNDCEAVVHISSYEAHKAPLSDILQCLENNGLYLLNASSSETFGGRVFYNLHFQVEKTHRLESEILTEKL
LSIYEKQRIF

>SEQIDNO:22

MGHQTQMFDDPFASSMSSLDADIFSVAGGLHPSQWPGLDHDVSLAPAANNGTSSGGYGSPGGGDGSGSHRKISHNAY
ERDRRKQLNELYSDLRSLLPDSDHTKKLSIPITVSRVLKYIPELQKQVDGLEKKKEELTRASCKPGVLTMKENTVPI
VSATCLDEREIMVQVSLVSTMAGALPMSKCIKVLENEGLRLISSSTSAFQNRTFYSLHLQRTQRTMSKECPAFCEEL
ENALTQKAGLRLHHQ

>SEQIDNO:23

MTPSRAASRRWRRTSSPPAASGASPPWPDLELDLDLDDDDIHDLSAPAANATSSGGYGSGGGSGGSHRKLSHNAYER
DRRKQLNELYSSLRSLLPDADHTKKLSIPTTVSRVLKYIPELQKQVDNLERRKKELTNANCKPGVLKTSQIVTPIVS
ATCLNDTEIMVQVSLQSNVAATSLPLSKCIKVLENEGLHLISSSTYSTFDNRTFYSLHLQRSQRTMKEECPAFCDEL
ERIIKKKAGA

>SEQIDNO:24

MLAISSSSPPLFSTTTNNFGWLLEDLISHELTNSGETSNSSQKSLQHCDSNKFDQIIINSGDQYQPDQTVKKLNHNA
SERDRRKKINSLYSSLRSLLPPSDHTKKLSIPSTVSRILK

>SEQIDNO:25

MGHQHQMFNDPFASSMSSLEEDMFSGAGGYHHLTPSMQWPGLDNDIPSAPAANNATSSGGSGSHRKMSHNAYERDRR
KQLNEQYSSLRSLLPDDDHTKKMSIPTTVSRVINYIPELQKEVDRLEKKKEELRRGSCEQGAMRQNTAPIVSATCLD
DREIMVQVSLVSTMAGALPMSKCIKVLENQGLRLINSSTSAFQNRTFYSLHLQRTQRTMSKEGQTFCNELENAVKQK
AGLHLHH

>SEQIDNO:26

MGHQHQMFEDPFASSISSLEAEIFSVAGGHHHTQWPGLDHDIPLAPAANNGTSSGGYGSPGGGDGSGSHRKISHNAY
ERDRRKQLNELYSDLRSLLPDTDHTKKLSIPITVSRVLKYIPELQKQVDGLEKKKEELTRASCKPGVLTMKGDTAPI
VSXHCLDDREIMVXXQLVSTMGGVCHVKCSGAETKARLIVVTPVQTDSIFIS

>SEQIDNO:27

MGHQHQMFEDPFASSISSLEADIFSVAAGHHHPQWPGLDHDVPLAPAANNGTSSGGYGSPGGGDGSGSHRKISHNAY
ERDRRKQLNELYSDLRSLLPDTDHTKKLSIPITVSRVLKYIPELQKQVDGLEKKKEELTRANCSPAC

>SEQIDNO:28

NNATSSGGSGSHRKMSHNAYERDRRKQLNEQYSSLRSLLPDDDHNKKMSIPTTVSRVIKYIPELQKEVDGLEKKKEE
LRRASYEQAP

>SEQIDNO:29

ATGTGTGCATTAGTCCCTTCATTTTTCACAAACTTCGGTTGGCCGTCAACGAATCAATACGAAAGCTATTACGGTGC
CGGAGATAACCTAAATAACGGCACATTTCTTGAATTGACGGTACCACAGACTTATGAAGTGACTCATCATCAGAATA
GCTTGGGAGTATCTGTTTCGTCAGAAGGAAATGAGATAGACAACAATCCGGTTGTGGTCAAGAAGCTTAATCACAAT

GCTAGTGAACGTGACCGACGCAAGAAGATCAACACTTTGTTCTCATCTCTCCGTTCATGTCTTCCAGCTTCTGATCA
ATCGAAGAAGCTAAGTATTCCTGAAACGGTTTCAAAGAGCTTAAAGTACATACCAGAGCTGCAACAGCAAGTGAAGA
GGCTAATACAAAAGAAGGAAGAAATTTTGGTACGAGTATCGGGTCAAAGAGACTTTGAGCTTTACGATAAGCAGCAA
CCAAAGGCGGTCGCGAGTTATCTCTCAACGGTTTCTGCCACTAGGCTTGGTGACAACGAAGTGATGGTCCAAGTCTC
ATCGTCCAAGATTCATAACTTTTCGATATCAAATGTGTTGGGTGGGATAGAAGAAGATGGGTTTGTTCTTGTGGATG
TTTCATCATCAAGATCTCAAGGAGAGAGGCTCTTCTACACTTTGCATCTTCAAGTGGAGAATATGGATGATTACAAG
ATTAATTGCGAAGAATTAAGTGAAAGGATGTTGTACTTGTACGAGAAATGTGAAAACTCGTTTAACTAG

>SEQIDNO:30

ATGTGTGCATTAGTCCCTCCATTGTTCCCAAACTTTGGGTGGCCGTCGACAGGAGAGTACGAGAGTAACTACCTGGC
CGGAGTGAACCTCGAGGACTTTACGTTTCTTGATTTTCCGGCACCAGAGACATATGGAGTGGAACATCATCAGGAGA
TTCAGGAAATGTTGGGGGTCTCTGTTCCGTCCGAGGGGAATGGAGTTGTAACCAAGAAGCTTAATCACAATGCTAGT
GAGCGTGACCGTCGCAAGAAGATCAACTCTTTGTTCTCGTCTCTCCGTTCATGTCTCCCAGCTTCTGATCAAACGAA
GAAGCTAAGTATTCCTCAGACGGTTTCTCGGAGCTTGAAGTACATTCCAGAGCTGCAAGAGCAAGTGAAGAAGCTAA
TACAAAAGAAGGAAGAACTCTTGGTGCGAGTATCAGGTCAAAGAGCCATTGAACATTATGTTGAGCCGCAGCCAAAG
GCCGTTGCACGTTACGTCTCGACCATTTCTGCGACTAAGCTTGGAGACAACGAAGTGCTGGTCCAAATCTCATCGTC
CAAGAATCATAACTTTTCGATATCTAATGTGTTGAGTGGGTTAGAAGAAGATGGGTTTGTTCTTGTTGATGTTTCAT
CTTCCAGGTATCATGGAAAATGGCTCTTCTACTCTTTGCATCTTCAAATGGGAAATAAAGATAATCACAAACTGAAG
TGCGAAGAGCTAAGCCAGAGAATTTTGTACTTGTATGAGGAATGTGAAAACTCATTTAGA

>SEQIDNO:31

ATGTGTGCATTAGTCCCTCCACTGTTCCCCGACTTTGGGTGGCCGTCGACGGCAGGTTACGAGAGCTACTACCTCGG
CGGAGAAAACCTCAACAACGACATGTTTCTTGATTTTCCGGTTGTGGAAACTTATGGAGTATTGGCTCATCATCAGA
ACAGCTTAGGAGTTTCTGTTTCGTCGGAGGGAAATGGAATAGACAACAACCCGGTTGTTAAAAAGAAGCTTAATCAC
AATGCTAGTGAGCGTGACCGTCGCAAGAAGATCAACTCTTTGTTTGCATCTCTCCGCTCATGTCTTCCAACCTCAGA
TCAATCGAAAAAGCTAAGCATTTCAGCCACCGTTTCACGAAGCTTGAAGTACATACCAGAGTTGCAAGAGCAAGTGA
AGAAGTTATTACAAAAGAAGGAAGAACTCTTGGTTCGAGTATCAGGTCAACGAGACATTGAACTTTACGTTAAGCCA
CAACCAAAGGCAATTGCAAGTTATGTCTCCACTGTTTCCGCGACTAGGCTTGGAGACAACGAAGTGATGGTCCAAAT
CTCATCATCCAAGATTCATAACTTCTCGATATCTAAAGTGTTAACTGGGTTAGAAGAAGATGGTTTTGTTCTTGTGG
ATGTTTCATCTTCAAGGTTTCAAGGGGAAAGGCTTTTCTACACTTTGCATCTTCAAGTAGAAAATATGGATGATCAT
TACAAAATGAATTGCGAAGAGTTAAGTGAAAGGATGTTGTACTTGTACGAGGAATGTGAAAATTNNNTTAGG

>SEQIDNO:32

ATTCCTCAGACGGTTTCTCGGAGCTTGAAGTACATACCAGAGCTACAAGAGCAAGTGAAGAAGCTAATACAAAAGAA
GGAAGAACTCTTGGTGCGAGTATCAGGTCAAAGAGACATTGAACATTACGTTGAGCCGCACCCAAAGGCCGTTGCAC
GTTACGTCTCGACCATTTCTGCGACTAAGCTTGGAGACAACGAAGTGATGGTCCAAATCTCATCGTCCAAGAATCAT
AACTTTTCGATATCTAATGTGTTGAGTGGGTTAGAAGAAGATGGGTTTGTTCTTGTTGATGTTTCATCTTCAAGGTC
TCATGGAGAAAGGCTCTTCTACACTTTGCATCTTCAAATGGGAAATAAAGATGATTACAAACTGACATGCGAAGAGC
TACGCCAGAGAATGTTATACTTGTATGAGGAATGTGGAAACTCGTTTAGA

>SEQIDNO:33

ATTCCAACTCCCTCACAAGCCACAAGCAGCGACCTTAGCATGGTCAAGAAACTTATCCGCAATGCTAGTGAACGAGA
TCGCCGCAAGAAAATCAATACTTTGTATTCTTCACTTCGTTCACTTCTTCCTGTGGCAGAACAGATGAAGAAGTTGA
GCAATCCGGCAACAATTTCACGAGTCCTAAAGTACATACGTGAGTTACAGAAGCAGGTAGAAGGACTACTTACGAGA
AAGGAGGCGATTTTATTGAAACTATCTCCAGAAGTAGATGAGGTGAAGAGTAAAGAATCTGAGAGGAAG

>SEQIDNO:34

AGCGACCGCCGCAAGCAGCTCAACGAGCAATATTCCTCCCTCCGCTCCCTCCTCCCCGATGACGATCACAATAAGAA
GATGAGCATTCCGACCACGGTGTCGCGGGTGATCAAGTACATCCCGGAGCTACAGAAGGAGGTCGACGGTCTGGAGA
AGAAGAAGGAGGAGCTCAGGCGAGCTAGCAGCGAGCAAGGCGTGCTGACTATGAGGCAGAACACGGCTCCTGTCGTC
TCCGCCACCTGCCTCGACGACAGGGAAATCATGGTCCAGGTCAGTCTGGTGAGCACCATGGCCGCAGCTCTGCCCAT
GTCCAAGTGCATCAAGGTGCTGGAGAACGAAGGCCTTCGCCTCATAAATTCCTCGACTTCCGCGTTCCAGAACAGGA

CCTTCTATAGCCTCCATCTTCAGAGAACCCAACGAACAATGAGCAAGGAGGGCCAAACATTTTGTAACGAACTGGAG
AACGCCGTGAAGCAAAAGGCAGGACTGCATCTGCATCAT

>SEQIDNO:35

ATGGGGCACCAGACCCAGATGTTCGACGACCCGTTCGCGAGCAGTATGTCGTCCCTGGACGCAGACATCTTCTCCGT
CGCCGGCGGCCTCCACCCATCGCAGTGGCCGGGACTCGACCACGACGTCTCGCTGGCGCCGGCTGCCAACAACGGCA
CCTCCTCCGGCGGCTACGGCTCCCCCGGGGGCGGCGATGGCTCGGGCTCCCACCGCAAGATCAGCCACAACGCGTAC
GAGCGCGACCGCCGCAAGCAGCTCAACGAGCTCTACTCCGACCTCCGCTCCCTCCTCCCCGACTCCGATCACACCAA
GAAGCTGAGCATTCCGATCACGGTGTCGCGCGTGCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACGGACTGG
AGAAGAAGAAGGAGGAGCTTACGCGGGCCAGCTGCAAGCCAGGCGTATTGACCATGAAGGAGAACACGGTCCCGATC
GTGTCCGCCACCTGCCTCGACGAAAGGGAGATCATGGTCCAGGTTAGCTTGGTGAGCACCATGGCCGGAGCTCTGCC
CATGTCCAAGCGCATCAAAGTGCTGGAGAACGAAGGCCTCCGCCTCATCAGCTCGTCCACTTCTGCTTTCCAGAACA
GGACGTTCTATAGCCTCCATCTTCAGAGAACCCAACGGACGATGAGCAAAGAGTGTCCGGCATTTTGTGAAGAACTG
GAGAATGCCCTGACGCAGAAGGCGGGGTACGTCTACATCACCAGTAGATTATATGTAGCAGAATAA

>SEQIDNO:36

ATGTTAGCGATATCTCCTCCTATGTTTTCAACAATTGGATGGCCCTTTGAGGAGCCTTTAAGCCATAACCAGCATCA
GAATTCATTCTACAAAGACACTGTTGATCAATTATTTAATTTTCATGATCAAGTTGAGGCAGAAATTAATTCAACAG
ATCCCTCACAATCCACAAGCAGTGACCTTAGCATGGTCAAGAAGCTTGTTCATAATGCCAGTGAACGCGATCGCCGC
AAGAAGATCAATAATTTGTATTCATCACTTCGATCACTCCTTCCTGTTTCTGATCAAATGAAATTAAGCATTCCGGG
AACAATTTCTAGAGTCCTGAAATACATACCTGAATTACAGAATCAAGTAGAGGGACTAATTAAGAGAAAGGATGAGA
TCTTATTGGGACTTTCTCCACAAGTAGAAGAGTTTATTCTAAGCAAAGAATCTCAAAGGAAGAAGCATAGTTACAAC
TCTGGTTTTGTAGTTTCAAGTAGTAGGCTCAATGATAGTGAAATTACCATTCAGATTTCATGTTACACTGTCCAAAA
GATTCCACTTTCTGAGATCTTGATTTGTTTGGAAAATGATGGCCTTTTGCTGCTTAATGTTTCTTCATCAAAGACCT
TTGGAGGGAGGGTCTTCTATAATTTGCATTTCCAGGTGGATAAAACACAGATATTAGAATCTCATATTCTAAATGAG
AAGCTCTTATCAATAATGGAGAAGGAAGGAGAGTTTTTAAAACAATAATTAAAGTTTAGGATTTGGCTCTTTAA

>SEQIDNO:37

ATGGTTGCATTCTGCCCACCTCAGTTCTCATACTCAAACATGGGATGGCTCTTAGAGGAGTTAGAGCCAGAGTCCTT
AATTAGTCATAAAGAGAAGAACTATGCATCTTTAGAGTACTCGTTACCGTATCATCAATTCTCTTCACCAAAGGAAC
ATGTTGAAATTGAAAGGCCACCATCCCCTAAACTTATGGCCAAGAAACTTAACCACAATGCTAGTGAACGTGATCGC
CGCAAGAAGATTAATAGCTTGATTTCTTCACTTCGTTCACTTCTTCCCGGTGAAGATCAAACGAAAAAAATGAGCAT
TCCGGTAACAATTTCACGTGTCTTAAAATACATCCCTGATTTACAAAAGCAGGTGCAAGGACTTACCAAGAAAAAAG
AAGAGCTTCTATCAAGAATTTCTCATCGACAAGAATATGCAGTTAACAAAGAATCACAAAGGAAGAAAATTCCAAAT
TACAATTCTGCTTTTGTAGTTTCAACAAGTAGGCTTAATGATACTGAGCTTGTTATTCATATTTCGTCTTATGAGGC
CAACAAGATTCCTCTATCTGAGATCTTGATGTGTTTAGAAAATAATGGTCTTCTTCTACTTAACTCTTCTTCTTCTA
AAACCTTTGGAGGGAGGCTCTTCTATAACTTGCATTTTCAGGTGGATAAAACTCAAAGATATGAGTGTGATGATCTG
ATTCAAAAGCTTTCTTCAATATATGAGAAGCAGCAAAATAATCATTTGGGCACTATGGATCAAACGATCAATAGTGG
TCTGATATAT

>SEQIDNO:38

ATGTTAGCGATATCTCCTCCTATGTTTTCAACAATTGGATGGCCCTTTGAGGAGCCTTTAAGCCATAACCAGCATCA
GAATTCATTCTACAAAGACACTGTTGATCAATTATTTAATTTTCATGATCAAGTTGAGGCAGAAATTAATTCAACAG
ATCCCTCACAATCCACAAGCAGTGACCTTAGCATGGTCAAGAAGCTTGTTCATTATGCCAGTGAACGCGATCGCCGC
AAGAAGATCAATAATTTGTATTCATCACTTCGATCACTCCTTCCTGTTTCTGATCAAATGGTACTTAAT

>SEQIDNO:39

ATGGAGCACCAGCTGTTCGATGACCCCTTCTCTAGCAGCATCTCGTCGCTGGAGGCGGACATCTTCTCCGCCGGCGG
CCAGCTGCCGTCGCCGCCGTGGCCGGACCTCGACCTCGACCTCGACGACGACGACATCCACGACCTCTCCGCGCCGA
CCGGCAACCCCACCTCCTCAGGAGGCTATGGCTCGGGCGGAGGCTCCGGAGGCTCCCACAGGAAGCACAGCCACAAC
GCGTACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTCTACTCCTCGCTCCGCTCCCTCCTCCCCGACGCTGACCA
CACTAAGAAGCTGAGCATCCCCACCACGGTCTCCCGAGTTCTCAAGTACATCCCCGAGCTGCAGAAGCAGGTGGACA

ACCTGGAGAGGAGGAAGAAAGAGCTGACGAACGCCAACTGCAAACCAGGAGTTCTGAACACGAGCCAGATTGTAACT
CCCATTGTTTCTGCTACTTGCCTCAACGATACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAACGTGGCTGCCAC
AAGTCTTCCTCTGTCCAAGTGCATAAAAGTGATGGAGAATGAAGGCCTTCACCTAATTAGTTCATCAACTTACTCCA
CCTTCGACAACAGGACATTCTATAGCCTCCATGTTCAGAGAAGTCAAAGAACGATGAAAGAGGAGTGCCCAGCA

>SEQIDNO:40

ATGGTCAAGAAACTTAGCCACAACGCTAATGAACGTGACCGTCGCAAGAAGATTAAAAGTTTGTATTCTTCACTTCG
TTCACTTCTCCCAGCAGCAGATCAAATGAAGAAATTAAGCGTGCCGGCCACTGTTTCACGTGCGCTTAAGTACCTAC
CAGAGCTTCAACAGCAAGTGGAGAGACTGGTTCAAAGAAAGGAGGAGCTTTTATCAAAGTTATCAAAGCAAGGTGGT
ATAATTCATCAAGAAAATCAAAGAAATGACACCGTGTATAGCTCTTTATCATCGGTATCGGCAAGCCAGCTTAGTGA
TAGAGAAGTTGTCGTTCATATTTCCACTTACAAGAACCATAAAAGTCCATTATCAGAAATCTTGCTCACCTTAGAGG
AAGATGGACTTGTTCTAAAAAACTCTTCTTCCTTTGAGTCATTTGGGGACAGGGTCTTCTATAATTTACATCTTCAG
GTCATGGAAGGAACTTACACATTGGATAGTGAGGCCATGAGGGCGAAGCTTGTGTCTTTATCTGTAAAGAGGGAATC
ATCGTCTCTA

>SEQIDNO:41

ATGTTAGAAGAATTATCTCCCATCAGTTTGTTCTCAACATTTGGATGGCCCTTGGAGGAAGCCATAAGCCATGAACA
GCACTACAGCTTTAGAGATGGTGAAACTCCAGAGTCATTTACTCACTTCCCTCCATCTCAGCCAGATGTAAGACAGC
TTGATCGCTCCACATCATTCACGGCCCACAGTGGAAGCGGTGACCCTAGCATGGCTAAGAAGCTTAACCACAACGCT
AGCGAACGTGACCGTCGCAAAAAGATCAACAGTTTGTATTCTTCACTCCGTTCACTACTTCCTGCAGCCGATCAAAG
GAAGAAATTAAGCATACCGTATACAGTTTCACGTGTGCTTGTATACATACCAAAACTTCAACAACAAGTGGAGAGAC
TGATTCAAAGGAAGGAGGAGCTTCTATCGAAGTTATCTAGGCAAGCTGACGATTTAACTCATCAAGAAAATCAAAGA
AAAGGCACCATGTATAGCTCTTTATCATCGGTATCGGCGAGCCGGCTCAGTGACAGGGAAGTTGTCATTCATATCTC
AACTAACAAGCTCCATAGAAGTTCATTATCAGAAATCTTGGTTAATTTAGAGGAGGCTGGACTTCTTCTACTAAATT
CTTCTTCCTTCGAGTCCTTTGGAGGCAGAGTCTTCTATAATTTACACCTTCAGGCCATGGAAGGAACTTACACAGTA
GAGTGCGAGGCCTTGAATGAGAGGCTTGTGTCCTTGTGCGAGAAGAGGGAGTCATTGTTTCCATTAAATTCAAGTTC
TCCATATTCTAACTGTGTATTCTAG

>SEQIDNO:42

GATCCTAACATGGTTAAGAAGCTTAACCACAACGCTAGCGAACGTGATCGTCGCAAGAAGATCAACAGTTTGTATTC
TTCACTCCGTTCACTTCTTCCAGCTTCCGATGGAATGAAGAAATTAAGCATACCGTCCACAATTTCACGTGTGCTTA
AGTACATACCAGAACTTCAACAGCAAGTGGAGAGACAGATCCAAAGGAAGGAGGAGCTTCTATCAAATCTATCTCGG
CAAGATGATTTAATTCATCAAGAAAATCAAAGAAAAGACACCATGTATAGCTCTTTATCATCGGTATCGGCAAGCCG
GCTTGGTGATAGAGAAGTTGTCGTTCAAATTTCCACTTGCAAGGTCCTTAAAAGCCCAATATCAGAAATCTTGCTTA
ATTTAGAGGAAAATGGACTTGTTCTAATAAATTCTTCTTCCTTTGAGTCCTTTGGAGGCAACGTCTTCTACCATTTA
CATCTTCAGGTA

>SEQIDNO:43

ATGTTAGCATTATCTCCTCCTGTATTTCCAACACCTGAATGGCCCTTAGAGGACCCCTTAGGCATTGACCAAATCTC
CTACTTCTGTAGAGAAACTCAGCCTGCTACTGCTGCTTTTCTTCCATCTTATCAGCAAGAGTTATTATTATTAGAGC
TTGATCATCAACAATCCACATCTTTCACAGCCTATAATAGCAGTGGTGGTGACGCTAACGATATGGTGAAGAAGCTT
AATCATAATGCAAGCGAACGTGATCGTCGCAAGAAGATGAACACCCTCTATTCTTCCCTCCGATCACTATTTCCGGC
CGCCGATGAAATGAAGAAGCTGAGTATACCTGCCACAATTTCGAGGGTGTTGAAGTACATACCAGAACTACAAGAAC
AGTTAGAGAGATTGGTCCAAAGGAAGGAAGAGATTTTGCTAAGAATATCTAAGCAAAATCATATTGTTAATCCCCAA
ATAAACCAAAGAAAAGGCACTTCTCACAGCAGTTTATCAGTAGTATCAGCTAATCAAATTAGTGACAAAGAAGCCAT
TATTCAAATTTCTACGTACAGTAATACTATCCATACAAGTCCACTATCAGAAATCTTGCTTCTTTTGGAGGAGGAAG
GCCTTCTTTTGATTAATTCTTCTTCCGCTGAATCCTTTGGTGGCAGGGTCTTCAACAATTTACATGTTCAGGTTGAT
GATACTTATACATTGGAATGTGATGCTTTAAGTGAGAAGCTTGCATCTCTGTATGCCAAGAGGGACGGGCTGTTCCC
ATGA

>SEQIDNO:44

ATGGTCAAGAAGCTTAATCATAATGCAAGCGAAAGGGATCGCCGCAAGAAGATGAACACTCTCTATTCTTCCCTCCG
ATCACTTCTTCCGGCCTCCGATCAAATGAAGAAGCTGAGCATACCTGCCACAATTTCCAGGGTGTTGAAGTACATAC
CAGAACTACAACAACAATTGGAGAGATTCGTCCAAAGGAAAGAAGAACTATTACTGAGAATATCTAAGCAGAATCAT
ATTATTAATCCCCAAATAAACCAAAGAAAAGGCACTACTCACAGCACCTTATCAGTAGTATCAGCTAATCAAATTAG
TGACAAAGAAGTTGTTATTCAAGTTTCTACTTACAATAATACTATCCATACAAGTCCATTATCAGAAATCTTGCTTC
TTCTGGAGGAGGAAGGCCTTCTTCTGATTAATTCTTCCTCCTTTGAGTCCTTTGGAGGCAGGGTCTTCTACAATTTA
CATCTTCAGGTTGATGGAACTTATATATTGGAGTGTGATGCTTTAAGCGAGAAGCTTGCAGCTTTATATGAGAGAGA
CGGGTTATTTCCATGA

>SEQIDNO:45

ACAACGATAATAACTACGCCTCAATTTCAAACTGATCAGAATAACAAGTTGTTTGAAGGTTTACGTGCCGATAATAC
TATTGATTTACCTTCATCTCATCATTATCAACAACAATGTTTGAAAGGAAGTGAGTTTGATGTTGATGAGTTAGGGG
TAGAAAGGTCATTAATGGAGAAGAAGCTAAATCATAATGCAAGTGAACGTAATAGAAGGAAGAAGATGAATTTTCTT
TATTCAACTCTTCGTTCTTTGCTTCCTCCTCCTACTAATAAACATCAAAAGAAAAAATTAAGCTTTCCAGCAACAGT
ATCATATGTACAAGAATACATCCCAGAGTTGAAGAAAGAAATAGAGAGGCTAAGCAAAACAAAAGATTTGCTTTTAT
CAAAGAAATCAAATTATTCATTACTCAAAATTGATGATAATAATAAGAGAAAATTAATTATTGGTGGAACTTCTTGT
AATTCTTCAACAACATCAATTTGTGCAAGTCAACTAAGTAATTCACAAGTTTTGGTACAAATTTCAACAACTCAAGA
AAATAATTTTCCAATTTCACAAGTATTTGCAAGTGTAGAGGAAGATGGATTAATTTTGCTAAATGCATCATCCTTTA
AATCTTTTGGAGACAAGATTTTTCACAGCTTGCATTTTCAGATGCAAGGACCAATTGAAATGGACATTCAGGTTTTG
AAGACTAAGCTTTTAGTAATGTGT

>SEQIDNO:46

ATGGACCATCAGCTGTTCGACGACCCCTTCGGGAGCAGCATCTCGTCGCTGGAGGCGGACATCTTCTCCGCCGGCGG
CGGCGGACAGCTGCCGTCGCCGCCGTGGCCGGACCTCGACCTCGACGACGACTACGACATACACGACCTCTCCGCGC
CGGCCGCCAACGCCGCCACCTCCTCGGGAGGAGGCTATGGCTCCGGCGGCTCCGGCAGGAAGCTCAGCCACAACGCA
TACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTCTACTCCTCGCTCCGATCCCTCCTCCCGGACGCTGATCACAC
TAAGAAGCTGAGCATCCCCACCACCGTGTCCCGAGTTCTCAACACCAAAGAGATCGTAACTCCCATTGTTTCTGCTA
CTTGCCTTAACGACACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAATGTGGCCGCCACAGCTCTCCCTCTCTCC
AAGTGCATAAAGGTGCTAGAAAACGAAGGCCTTCTCCTCGTCAGCTCATCAACCTACTCCACCTTCGAGAACAAGAC
ATTCTATAGCCTCCATCTTCAGAGAAGTCAAAGAACGATGAAGGAGCAGTGCCCAGGATTCTGCGACGAACTGGAGA
AGATCGTCAGGAAGAAAGCAGGGGCG

>SEQIDNO:47

ATGGAGCATCAACTATTCGACGACGCCGTCCCGAGCAGCATGATCTGGCCGTTAGAGGCAGAAAACGGTTTCACCGA
CGAGCTGCCGTCTTTGCAGTTACCGGACGTGGACCTTGACTTCGACATCCACGAGTTCTCCGCACCGGCAACGGCAC
CGGCGAAAGCGGCCTCCTCGGGTGGCTCCGGATTGGTTGGTTCCGGTTCAGGATCGCATAAGAAGCTCAACCACAAC
GCGTACGAGCGCGACCGGCGGACGCAGCTCAATCAGCTCTACTCGACTCTCCGTTCTCTCATCCCCAACGCAGATCA
CACAAAGAAGCTGAGCATTCCGACGACGGTGTGTCAGGTCCTCGACTACATACCCAAGCTGCAGAAGCAGGTCGAGG
ATCTCAAGAAGAAGAAACAGGAGCTCAGTACAGCCAAATGCAGAGAAAGACTGCAGCGCGTCAAGGACAACACATGC
CGTATTGTTTCTGCCACTCCTCTCGATGGCAACGAAATCATGGTCCAGGTTAGCCTGCTGAGCAACATGGCTGCAAG
TCTTCCTCTATCCAAGTGCATAAACGTATTTGAGAACAAAGGGCTTCACCTCATCAGTTCATCGACTTTCTCCACCG
AGGTCAATAGAACATTTTACAGCTTCCACTTTGAGGTACGTTTTTACATGCGCCCT

>SEQIDNO:48

GGCGATGGCTCGGGCTCCCATCGCAAGATCAGCCACAACGCCTACGAGCGCGACCGCCGCAAGCAGCTCAACGAGCT
CTACTCCGACCTCCGCTCCCTCCTCCCCGACACCGATCACACGAAGAAGCTTAGCATTCCGATCACGGTGTCGCGGG
TGCTCAAGTACATCCCGGAGCTGCAGAAGCAGTGGGCTTGNAGAAANAAAGAAGGAGGAGNTGACGCGCGCCAACTG
CAACCCGGNGTGNTGACCATGAAGGGGAACACGGTCCGATTGTTCCGCCACCTGCCTCGACGACAGGGANATTATGG
TCCAAGTCAACCTGGTGAGCACATGGCCGGANTNTGCCCATTTCAAATGCATCAAAGTGCTGGAAAACAANNGCTCC
GGTCA

>SEQIDNO:49

ATGGAGCAGCTGTTCGTCGACGACCCAGCCTTCGCGAGCAGCATGTCGTCGCTTGAGGCGGACATCTTCTCCGGCGC
CGGCCAGCTGCCGTCCTCGCCGTGGCTGGACCTAGACCTCGACGACGATGTCCAAGACCTCTCCATGGCGCCGACGA
CGGCGAACGCGGTGTCCTCCGGCTACGGCTCCGGCGGATCCGGCTCCCACAGGAAGCTCAGCCACAACGCCTACGAG
CGCGACCGCCGGAAGCAGCTCAACGAGCTCTACTCCTCCCTCCGCGCTCTCCTCCCCGACGCCGATCACACTAAGAA
GCTGAGCATCCCGACGACGGTGTCTCGCGTGCTCAAATACATACCCGAGCTGCAGAAGCAGGTGGAGAATCTGGAGA
GGAAGAAGAAGGAGCTGACGACGACGAGCACCACCAACTGCCAACCAAGAGTGTTGGGGAGCCAGCTGATGAGC

>SEQIDNO:50

ATGGTCGCATTGTTCTCTCCTCCTCTCTTCTCAACCAAAGGGTGGCTCTTAGAGGAGGAGCCATTCGGCTATAATAA
TACCCATAATCTCTCCTACAAAGATGATGCGTCTTCTCAGTACTCCTTTCCCTATCAATTTTATTCACCACAGACAC
AGATTGAGGTTGAAATTGAAAGGTCCACTGCACCATCCTCTGACCCTGCCATGGTCAAGAAACTTAGCCACAATGCT
AGCGAACGTGATCGCCGCAAGAAGGTCAACAACTTGGTTTCTTCACTTCGCTCACTTCTTCCAATGGCAGATCAAAC
GAAAAAAATGAGCATTCCTGCAACAGTTTCCAGAGTGTTGAAATACATACCTGAACTACAACAGCAAGTGCAAGCAC
TAACAAAGAGAAAAGAGGAGCTTCTGTGCAGAATTTCTCGGCAATTGCAAGGAGAAGCAGTGAACAAAGAATCTCAG
AGAAAAATTTCCCATCACAACTCTTCTTTTGTTGTCTCAACGACTAGGCTTAACGATTGTGAAGCTGTAGTTCACAT
TTCATCTCATGAGACACACAAGGCTCCACTATCAGAGATTCTGCAGTGCTTAGAAAATGATGGCCTTTTTCTGCTAC
ATGCTTCTTCCTCAGAAACCTTTGGAGGAAGGTTCTTCTACAATTTGCATTTTCACGTGGAGAAAACTGATAGATTA
GAGACCGAGATTTTAACTGAGAAGCTTTTACCAATATAT

>SEQIDNO:51

ATGTTAGCATTCTCCCCTCCATTGTTTCCAACCCTTGGATGGCCCTTGGAGGATCCCATAAGCCATGCACAGAACTA
CATATATGGAGAAACAGAAACTTCAGAATCGTTTCTTCACTTGCCCTCATCTCAGCCACAAGTGGAACTCAATTGCT
CCACCCCATATGCAGCAGTTAGTGGTAATCCCACGATGGTTAAGAAACTTAACCACAACGTCAGTGTGCGGGATCGT
CGGAAGAAGATCAACAGCTTGTACTCCTCTCTGCGTTCACTACTTCCATCAGCTGATCAAGTGAAGAAATTAAGCAT
TCCTTCGACAGTTTCATGTGTGCTAAAATACATACCAGAGCTGCAACGGCAAGTGGAGAGACTGATCCAAAAGAAAG
AAGAGTTTTTATCAAAGATTTCTAGGGAAGGAGATCTAATTCACCTAGAAAATCAAAGAAATGGCACACTTGGAAGC
TCTTTATCTGCTGTTTCAGCAAGAAGGCTTAGTGACAGGGAAATTGTGGTTCAGATATCCACATTTAAGGTCCATGA
GAGTCCACTTTCTGAGGTTTTGTTAAATTTGGAGGAGGATGGGCTTCTTGTAATCAATGCATCATCTTTTGAGTCCT
TTGGAGGGAGGGTCTTCTACAACTTACATCTTCAGGTTGAAGGAACTCAAGGAATGGAGTCGACCGCCCACTTGGAG
ATGACAAGAACTCTAAAAAACAAACACATGAATATATTGGTGATTCATATGGACTTTCCTCCTTTTTTTCTTAAGAT
GTTCCTGATCTTTACAAGAGTATTTACCAATCATATATCAACTTCCTACCAATGCTATGTTGGTAAATTAATTATCA
TTCTACTGCTACATGTAATTAAGAAAGAAAATTTTGAAACTTCTAAACATCAATTGGCCAGCGCCGAACCAATAATT
GTAGCAGGTCAAGCTGCCAAAAAATTGGATGATGAGCTTCTCTTGGAAACCAAGATGAAAACTGAAGGGATGGGAGT
ACTGGAAACACCAATATTGATTAAAGCAAAGAATGGTACCAAAGAAATGGTGGAGAGAATCCTTGATCTTTACCCCA
TGGCAATTCATGACATAGACTCCAACAAGAAGAATATAGTGCTATTGGCGGTGGAGAATAGGCACCCCCATGTGTAT
GAGCTCTTCCTGAAGAGAAATATTGTGAAAGATAGTGTATTTGGTGCAGTTGATAATAAAGGCAACAGTGCATTGCA
TCTGGCTGCCATGTTTGCAGATTATCGGCCTTGGGTCACTCCTGGTGTTGCATTGCAAATGCAATGGGAAGTCAAAT
GGTATGAGTATGTGAAGAAGTCCATGCCACCAAATTTCTTCCGTTTCCACAACAATGAAAACAAGTCTACAAAGCAG
ATTTTCACCCGTGAACACAGAGATCTGGTGCAAAAGGGTGGGCAATGGCTAAATAACACAGCCACCTCATGCTCGTT
GGTAGTAACACTCATTGCAACAGTTGCCTTCGCCACATCAACTGCTGTACCGGGCGGCACCAAGGAGGGGACTGATT
CATGTCCTCTCAATGGTCCCTAA

>SEQIDNO:52

ATGTTAGCATTCTCCCCTCCATTGTTTCCAACCCTTGGATGGCCCTTGGAGGATCCCATAAGCCATGCACAGAACTA
CATATATGGAGAAACAGAAACTTCAGAATCGTTTCTTCACTTGTCCTCATCTCAGCCACAAGTGGAACTCAATTGCT
CCACCCCATCTGCAGCAGTTAGTGGTAATCCCACGATGGTTAAGAAACTTAACCACAACGCCAGTGAGCGGGATCGT
CGGAAGAAGATCAACAGCTTGTACTCCTCTATGCGTTCACTACTTCCATCAGCTGATCAAGNGAAGAAATTAAGCAT
TCCTTCGACAGTTTCACGTGTGCTAAAATACATACCAGAACTGCAACGACAAGTGGAGAGATTGATTCAAAAGAAAG
AAGAGTTTTTATCAAAGATTTGTAGGGAAGGAGATCCAATTCACCTAGAAAATCAAAGAAATGGCACACTTGGAAGC
TCTTTATCTGCTGTTTCAGCAAGAAGGCTTAGTGACAGGGAAATTGTGGTTCAGATATCCACATTTAATGTCCATGA
GAGTCCTCTTTCTGAGGTTTTGTTAAATTTGGAGGAGGATGGGCTTCTTGTAATCAATGCATCATCTTTTGAGTCCT

TTGGAGGGAGGGTCTTCTACAACTTACATCTTCAGGTTGAAGGAACTCAAGGAATGGAGTGTGAGTTGTTGAGCGAG
AAGCTACTTTCATTGTGTGAAAGGAGAGAGGCTTTTCCATGA

>SEQIDNO:53

MCALVPSFFTNFGWPSTNQYESYYGAGDNLNNGTFLELTVPQTYEVTHHQNSLGVSVSSEGNEIDNNPVVVKKLNHN
ASERDRRKKINTLFSSLRSCLPASDQSKKLSIPETVSKSLKYIPELQQQVKRLIQKKEEILVRVSGQRDFELYDKQQ
PKAVASYLSTVSATRLGDNEVMVQVSSSKIHNFSISNVLGGIEEDGFVLVDVSSSRSQGERLFYTLHLQVENMDDYK
INCEELSERMLYLYEKCENSFN

>SEQIDNO:54

MCALVPPLFPNFGWPSTGEYESNYLAGVNLEDFTFLDFPAPETYGVEHHQEIQEMLGVSVPSEGNGVVTKKLNHNAS
ERDRRKKINSLFSSLRSCLPASDQTKKLSIPQTVSRSLKYIPELQEQVKKLIQKKEELLVRVSGQRAIEHYVEPQPK
AVARYVSTISATKLGDNEVLVQISSSKNHNFSISNVLSGLEEDGFVLVDVSSSRYHGKWLFYSLHLQMGNKDNHKLK
CEELSQRILYLYEECENSFR

>SEQIDNO:55

MCALVPPLFPDFGWPSTAGYESYYLGGENLNNDMFLDFPVVETYGVLAHHQNSLGVSVSSEGNGIDNNPVVKKKLNH
NASERDRRKKINSLFASLRSCLPTSDQSKKLSISATVSRSLKYIPELQEQVKKLLQKKEELLVRVSGQRDIELYVKP
QPKAIASYVSTVSATRLGDNEVMVQISSSKIHNFSISKVLTGLEEDGFVLVDVSSSRFQGERLFYTLHLQVENMDDH
YKMNCEELSERMLYLYEECENXXR

>SEQIDNO:56

IPQTVSRSLKYIPELQEQVKKLIQKKEELLVRVSGQRDIEHYVEPHPKAVARYVSTISATKLGDNEVMVQISSSKNH
NFSISNVLSGLEEDGFVLVDVSSSRSHGERLFYTLHLQMGNKDDYKLTCEELRQRMLYLYEECGNSFR

>SEQIDNO:57

IPTPSQATSSDLSMVKKLIRNASERDRRKKINTLYSSLRSLLPVAEQMKKLSNPATISRVLKYIRELQKQVEGLLIR
KEAILLKLSPEVDEVKSKESERK

>SEQIDNO:58

SDRRKQLNEQYSSLRSLLPDDDHNKKMSIPTTVSRVIKYIPELQKEVDGLEKKKEELRRASSEQGVLTMRQNTAPVV
SATCLDDREIMVQVSLVSTMAAALPMSKCIKVLENEGLRLINSSTSAFQNRTFYSLHLQRTQRTMSKEGQTFCNELE
NAVKQKAGLHLHH

>SEQIDNO:59

MGHQTQMFDDPFASSMSSLDADIFSVAGGLHPSQWPGLDHDVSLAPAANNGTSSGGYGSPGGGDGSGSHRKISHNAY
ERDRRKQLNELYSDLRSLLPDSDHTKKLSIPITVSRVLKYIPELQKQVDGLEKKKEELTRASCKPGVLTMKENTVPI
VSATCLDEREIMVQVSLVSTMAGALPMSKRIKVLENEGLRLISSSTSAFQNRTFYSLHLQRTQRTMSKECPAFCEEL
ENALTQKAGYVYITSRLYVAE

>SEQIDNO:60

MLAISPPMFSTIGWPFEEPLSHNQHQNSFYKDTVDQLFNFHDQVEAEINSTDPSQSTSSDLSMVKKLVHNASERDRR
KKINNLYSSLRSLLPVSDQMKLSIPGTISRVLKYIPELQNQVEGLIKRKDEILLGLSPQVEEFILSKESQRKKHSYN
SGFVVSSSRLNDSEITIQISCYTVQKIPLSEILICLENDGLLLLNVSSSKTFGGRVFYNLHFQVDKTQILESHILNE
KLLSIMEKEGEFLKQ

>SEQIDNO:61

MVAFCPPQFSYSNMGWLLEELEPESLISHKEKNYASLEYSLPYHQFSSPKEHVEIERPPSPKLMAKKLNHNASERDR
RKKINSLISSLRSLLPGEDQTKKMSIPVTISRVLKYIPDLQKQVQGLTKKKEELLSRISHRQEYAVNKESQRKKIPN
YNSAFVVSTSRLNDTELVIHISSYEANKIPLSEILMCLENNGLLLLNSSSSKTFGGRLFYNLHFQVDKTQRYECDDL
IQKLSSIYEKQQNNHLGTMDQTINSGLIY

>SEQIDNO:62

MLAISPPMFSTIGWPFEEPLSHNQHQNSFYKDTVDQLFNFHDQVEAEINSTDPSQSTSSDLSMVKKLVHYASERDRR
KKINNLYSSLRSLLPVSDQMVLN

>SEQIDNO:63

MEQLFVDDPAFASSMSSLEADIFSGAGQLPSSPWLDLDLDDDVQDLSMAPTTANAVSSGYGSGGSGSHRKLSHNAYE
RDRRKQLNELYSSLRALLPDADHTKLSIPTTVSRVLKYIPELQKQVENLERKKKELTTTSTTNCKPGVLGSQLMSEG
MAPIVSATCINDMEIMVQVSLLSNVAGSVLPLSKCIKVLENEGLHFISSSTSSGFGNRTFYSIHLQRSEGTINEECP
AFCERLEKVVRNKAKL

>SEQIDNO:64

MEHQLFDDPFSSSISSLEADIFSAGGQLPSPPWPDLDLDLDDDDIHDLSAPTGNPTSSGGYGSGGGSGGSHRKHSHN
AYERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTVSRVLKYIPELQKQVDNLERRKKELTNANCKPGVLNTSQIVT
PIVSATCLNDTEIMVQVSLHSNVAATSLPLSKCIKVMENEGLHLISSSTYSTFDNRTFYSLHVQRSQRTMKEECPA

>SEQIDNO:65

MVKKLSHNANERDRRKKIKSLYSSLRSLLPAADQMKKLSVPATVSRALKYLPELQQQVERLVQRKEELLSKLSKQGG
IIHQENQRNDTVYSSLSSVSASQLSDREVVVHISTYKNHKSPLSEILLTLEEDGLVLKNSSSFESFGDRVFYNLHLQ
VMEGTYTLDSEAMRAKLVSLSVKRESSSL

>SEQIDNO:66

MLEELSPISLFSTFGWPLEEAISHEQHYSFRDGETPESFTHFPPSQPDVRQLDRSTSFTAHSGSGDPSMAKKLNHNA
SERDRRKKINSLYSSLRSLLPAADQRKKLSIPYTVSRVLVYIPKLQQQVERLIQRKEELLSKLSRQADDLTHQENQR
KGTMYSSLSSVSASRLSDREVVIHISTNKLHRSSLSEILVNLEEAGLLLLNSSSFESFGGRVFYNLHLQAMEGTYTV
ECEALNERLVSLCEKRESLFPLNSSSPYSNCVF

>SEQIDNO:67

DPNMVKKLNHNASERDRRKKINSLYSSLRSLLPASDGMKKLSIPSTISRVLKYIPELQQQVERQIQRKEELLSNLSR
QDDLIHQENQRKDTMYSSLSSVSASRLGDREVVVQISTCKVLKSPISEILLNLEENGLVLINSSSFESFGGNVFYHL
HLQV

>SEQIDNO:68

MLALSPPVFPTPEWPLEDPLGIDQISYFCRETQPATAAFLPSYQQELLLLELDHQQSTSFTAYNSSGGDANDMVKKL
NHNASERDRRKKMNTLYSSLRSLFPAADEMKKLSIPATISRVLKYIPELQEQLERLVQRKEEILLRISKQNHIVNPQ
INQRKGTSHSSLSVVSANQISDKEAIIQISTYSNTIHTSPLSEILLLLEEEGLLLINSSSAESFGGRVFNNLHVQVD
DTYTLECDALSEKLASLYAKRDGLFP

>SEQIDNO:69

MVKKLNHNASERDRRKKMNTLYSSLRSLLPASDQMKKLSIPATISRVLKYIPELQQQLERFVQRKEELLLRISKQNH
IINPQINQRKGTTHSTLSVVSANQISDKEVVIQVSTYNNTIHTSPLSEILLLLEEEGLLLINSSSFESFGGRVFYNL
HLQVDGTYILECDALSEKLAALYERDGLFP

>SEQIDNO:70

TTIITTPQFQTDQNNKLFEGLRADNTIDLPSSHHYQQQCLKGSEFDVDELGVERSLMEKKLNHNASERNRRKKMNFL
YSTLRSLLPPPTNKHQKKKLSFPATVSYVQEYIPELKKEIERLSKTKDLLLSKKSNYSLLKIDDNNKRKLIIGGTSC
NSSTTSICASQLSNSQVLVQISTTQENNFPISQVFASVEEDGLILLNASSFKSFGDKIFHSLHFQMQGPIEMDIQVL
KTKLLVMC

>SEQIDNO:71

MDHQLFDDPFGSSISSLEADIFSAGGGGQLPSPPWPDLDLDDDYDIHDLSAPAANAATSSGGGYGSGGSGRKLSHNA
YERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTVSRVLNTKEIVTPIVSATCLNDTEIMVQVSLHSNVAATALPLS
KCIKVLENEGLLLVSSSTYSTFENKTFYSLHLQRSQRTMKEQCPGFCDELEKIVRKKAGA

>SEQIDNO:72

MEHQLFDDAVPSSMIWPLEAENGFTDELPSLQLPDVDLDFDIHEFSAPATAPAKAASSGGSGLVGSGSGSHKKLNHN
AYERDRRTQLNQLYSTLRSLIPNADHTKKLSIPTTVCQVLDYIPKLQKQVEDLKKKKQELSTAKCRERLQRVKDNTC
RIVSATPLDGNEIMVQVSLLSNMAASLPLSKCINVFENKGLHLISSSTFSTEVNRTFYSFHFEVRFYMRP

>SEQIDNO:73

GDGSGSHRKISHNAYERDRRKQLNELYSDLRSLLPDTDHTKKLSIPITVSRVLKYIPELQKQWAXRXKEGGXDARQL
QPGVXTMKGNTVRLFRHLPRRQGXYGPSQPGEHMAGXCPFQMHQSAGKQXLRS

>SEQIDNO:74

MEQLFVDDPAFASSMSSLEADIFSGAGQLPSSPWLDLDLDDDVQDLSMAPTTANAVSSGYGSGGSGSHRKLSHNAYE
RDRRKQLNELYSSLRALLPDADHTKKLSIPTTVSRVLKYIPELQKQVENLERKKKELTTTSTTNCQPRVLGSQLMS

>SEQIDNO:75

MVALFSPPLFSTKGWLLEEEPFGYNNTHNLSYKDDASSQYSFPYQFYSPQTQIEVEIERSTAPSSDPAMVKKLSHNA
SERDRRKKVNNLVSSLRSLLPMADQTKKMSIPATVSRVLKYIPELQQQVQALTKRKEELLCRISRQLQGEAVNKESQ
RKISHHNSSFVVSTTRLNDCEAVVHISSHETHKAPLSEILQCLENDGLFLLHASSSETFGGRFFYNLHFHVEKTDRL
ETEILTEKLLPIY

>SEQIDNO:76

MLAFSPPLFSTFGWPWEDPXSHEQNYIYQETEASESFLHLPSSEPQAELNYSTPSAAVSGNPTMVKKLNHNASERDR
RKKINSLYSSLRSLLPAADQAKKLSIPSTVSRVLKYIPELQKQVERLIQKKEELLSKISRQGDIIHQEKQRKATLAS
SLSAVSANRLSDREIVVQISTFKVHESPLSEVLLNLEEDGLLVINASSFESFGGRVFYNLHLQVEGTHRMECEVLSE
KLLSLCEKRRDAFP

>SEQIDNO:77

MLAFSPPLFPTLGWPLEDPISHAQNYIYGETETSESFLHLPSSQPQVELNCSTPYAAVSGNPTMVKKLNHNVSVRDR
RKKINSLYSSLRSLLPSADQVKKLSIPSTVSCVLKYIPELQRQVERLIQKKEEFLSKISREGDLIHLENQRNGTLGS
SLSAVSARRLSDREIVVQISTFKVHESPLSEVLLNLEEDGLLVINASSFESFGGRVFYNLHLQVEGTQGMESTAHLE
MTRTLKNKHMNILVIHMDFPPFFLKMFLIFTRVFTNHISTSYQCYVGKLIIILLLHVIKKENFETSKHQLASAEPII
VAGQAAKKLDDELLLETKMKTEGMGVLETPILIKAKNGTKEMVERILDLYPMAIHDIDSNKKNIVLLAVENRHPHVY
ELFLKRNIVKDSVFGAVDNKGNSALHLAAMFADYRPWVTPGVALQMQWEVKWYEYVKKSMPPNFFRFHNNENKSTKQ
IFTREHRDLVQKGGQWLNNTATSCSLVVTLIATVAFATSTAVPGGTKEGTDSCPLNGP

>SEQIDNO:78

MLAFSPPLFPTLGWPLEDPISHAQNYIYGETETSESFLHLSSSQPQVELNCSTPSAAVSGNPTMVKKLNHNASERDR
RKKINSLYSSMRSLLPSADQXKKLSIPSTVSRVLKYIPELQRQVERLIQKKEEFLSKICREGDPIHLENQRNGTLGS
SLSAVSARRLSDREIVVQISTFNVHESPLSEVLLNLEEDGLLVINASSFESFGGRVFYNLHLQVEGTQGMECELLSE
KLLSLCERREAFP

>SEQIDNO:79
ATGGAGTATCCATGGCTGCAGTCTCAAGTTCATTCCTTTTCACCTACTCTCCATTTTCCTTCCTTCCTTCATCCTTT
AGATGATTCCAAGAGCCATAACATCAATCTTCATCATATGAGTCTTAGTCACAGCAATAATACTAACAGTAACAATA
ACAATTATCAAGAAGAAGATCGAGGAGCGGTGGTTTTGGAGAAGAAACTGAATCACAACGCAAGCGAACGAGACCGC
CGTAGAAAACTTAACGCCTTGTACTCTTCACTTCGTGCTCTCTTGCCTCTTTCTGATCAAAAGAGGAAGCTGAGCAT
TCCTATGACGGTAGCGAGAGTAGTGAAATACATACCAGAGCAGAAGCAAGAACTTCAACGTTTGTCTCGGAGAAAAG
AAGAGCTCTTGAAGAGGATCTCGAGAAAAACTCACCAAGAGCAGCTGAGAAACAAAGCAATGATGGACTCAATAGAT
TCTTCTTCCTCTCAACGGATCGCAGCAAATTGGCTCACTGACACAGAGATTGCTGTCCAGATTGCTACGTCGAAATG
GACATCTGTTTCAGACATGTTGCTTAGGTTAGAAGAAAACGGGCTTAATGTCATAAGCGTCTCTTCTTCCGTTTCTT
CCACCGCAAGGATCTTCTACACTCTACATCTTCAGATGAGAGGAGATTGCAAAGTGAGACTGGAGGAACTCATCAAT
GGTATGCTCTTGGGATTACGCCAATCATAA

>SEQIDNO:80
ATGTGTGCCTTAACACCAATGTTTCCAAGTAACCAACAAGAATGGTACTCTACTTCAACAATGGAGTATCCATGGCT
TGATTCCTTCTCTCCTACTCTCCCTTCTTCTCTTTATCCTTCTTTCGACCAACTAGATGAATTCAAGAGCTATAACA
TCAATCTTCTTCCTCATCATATGAATCTTGCTGACATAAATGGTACTAACAATGATCAAGAAGAACATCAAGGATCG
GTTTTGGAAAAGAAACTGAATCACAACGCAAGTGAACGCGACCGCCGTAGAAAGCTAAACGCCTTATACGCTTCACT
TCGTGCTCTCTTGCCTCCTTCTGATCAAAAGAGAAAGTTGAGCATTCCAAAGACCATAGCGGGAGTGGTGAAGTATA
TACCAGAGCAGAAGCAAGAACTTCAACGTTTGTCTAGGAGGAAAGAAGAGCTTATGAAGAGAATCTCCAATAAGACA
GAGACTTTGAATCATCAACAAGAACAGCTGAGAAATAGAGCATTAATGATGGAGTCAATAGATTCTTCTTCACAAAA
GATCGCT

>SEQIDNO:81
ATGGAGTATCCATGGCTTGATTCCTTCTCTCCTACTCTCCCTTCTTCTCTTTATCCTTCTTTCGACCAACTAGATGA
ATTCAAGAACTATAACATCAATCTTCTTCCTCATCATATGAATCTTGCCGACATAAATGGTACTAACAATACCAGTA
ACAATGATCAAGAAGAACATCAAGGATCGGTTTTGGAAAAGAAACTGAATCACAACGCAAGTGAACGCGACCGCCGT
AGAAAGCTAAACGCCTTATACGCTTCACTTCGTGCTCTCTTGCCTCCTTCTGATCAAAAGTCGGCGAATCAGAGAAA
GTTGAGCATTCCAAAGACCGTAGCGGGAGTGGTGAAGTATATACCAGAGCAGAAGCAAGAACTTCAACGTTTGTATA
GGAGGAAAGAAGAGCTTATGAAGAGGATCTCCAATAAGATAGAGACTTTGAATCATCAACAAGAACAGCTGAGAAAT
AGAGCATTAATGATGGAGTCAATAGATTCTTCTTCACAAAAGATCGCTGCAAATTGGATCACCAACACAGAAATAGC
TGTCCAGATTGCTACATGGAAATGGACATCTATCTCAGACATGTTGCTTAGGTTAGAAGAAAACGGGCTTAATGTCA
TAAGCGTCTCTTCTTCGGTTTCTTCCACCGCAAGGATCTTCTACACACTGCATCTTCAGATG

>SEQIDNO:82
GCTTTCTCTTTCAGCTCGATCGATCCACAGCTCAACGAGCTCTACTCCTCCCTCCGCGCTCTCCTCCCCGACGCCGA
TCACACTAAGAAGCTGAGCATCCCGACGACGGTGTCTCGCGTGCTCAAGTACATACCCGAGCTGCAGAAGCAGGTGG
AGAATCTGGAGAGGAAGAAGAAGGAGCTGACGACGACGAGCACCACCAACTGCAAACCAGGAGTGTTGGGGAGCCAG
CTGATGAGCGAGGGCATGGCTCCCATCGTT

>SEQIDNO:83
ATGGAGCAGCTGTTCGTCGACGACCCAGCCTTCGCGAGCAGCATGTCGTCGCTTGAGGCGGACATCTTCTCCGGCGC
CGGCCAGCTGCCGTCCTCGCCGTGGCTGGACCTAGACCTCGACGACGATGTCCAAGACCTCTCCATGGCGCCGACGA
CGGCGAACGCGGTGTCCTCCGGCTACGGCTTCGGCGGATCCGGCTCCCACAGGAAGCTCAGCCACAACGCCTACGAG
CGCGACCGCCGGAAGCAGCTCAACGAGCTCTACTCCTCCCTCCGCGCTCTCCTCCCCGACGCCGATCACACTAAGAA
ACTGAGCATTTCGACGAACGTGTCCTGCGTGGTTCAGTACATAACCGAACCTGCAGAAACAAGTGGAGAATATGGAG
AAGAAAAAAAAGAGCTGACGACGACGAGCACCACCAACTGTCAACCCCAAGATGTGGGTAGAAGC

>SEQIDNO:84
MEYPWLQSQVHSFSPTLHFPSFLHPLDDSKSHNINLHHMSLSHSNNTNSNNNNYQEEDRGAVVLEKKLNHNASERDR
RRKLNALYSSLRALLPLSDQKRKLSIPMTVARVVKYIPEQKQELQRLSRRKEELLKRISRKTHQEQLRNKAMMDSID
SSSSQRIAANWLTDTEIAVQIATSKWTSVSDMLLRLEENGLNVISVSSSVSSTARIFYTLHLQMRGDCKVRLEELIN
GMLLGLRQS

>SEQIDNO:85
MCALTPMFPSNQQEWYSTSTMEYPWLDSFSPTLPSSLYPSFDQLDEFKSYNINLLPHHMNLADINGTNNDQEEHQGS
VLEKKLNHNASERDRRRKLNALYASLRALLPPSDQKRKLSIPKTIAGVVKYIPEQKQELQRLSRRKEELMKRISNKT
ETLNHQQEQLRNRALMMESIDSSSQKIA

>SEQIDNO:86
MEYPWLDSFSPTLPSSLYPSFDQLDEFKNYNINLLPHHMNLADINGTNNTSNNDQEEHQGSVLEKKLNHNASERDRR
RKLNALYASLRALLPPSDQKSANQRKLSIPKTVAGVVKYIPEQKQELQRLYRRKEELMKRISNKIETLNHQQEQLRN
RALMMESIDSSSQKIAANWITNTEIAVQIATWKWTSISDMLLRLEENGLNVISVSSSVSSTARIFYTLHLQM

>SEQIDNO:87
AFSFSSIDPQLNELYSSLRALLPDADHTKKLSIPTTVSRVLKYIPELQKQVENLERKKKELTTTSTTNCKPGVLGSQ
LMSEGMAPIV

>SEQIDNO:88
MEQLFVDDPAFASSMSSLEADIFSGAGQLPSSPWLDLDLDDDVQDLSMAPTTANAVSSGYGFGGSGSHRKLSHNAYE
RDRRKQLNELYSSLRALLPDADHTKKLSISTNVSCVVQYITEPAETSGEYGEEKKELTTTSTTNCQPQDVGRS

>SEQIDNO:89
ATGTGTGCACTTGTCCCTCCATTATATCCCAATTTCGGCTGGCCTTGCGGAGATCATAGCTTCTATGAAACCGACGA
CGTATCCAACACGTTTCTTGATTTTCCGTTGCCGGACTTGACGGTGACTCATGAGAATGTGTCGTCTGAGAATAACA
GAACATTACTAGACAATCCCGTGGTGATGAAGAAGCTTAATCACAACGCGAGTGAACGTGAGCGTCGCAAGAAGATC
AACACAATGTTCTCATCTCTTCGTTCTTGTCTTCCTCCCACCAATCAAACGAAGTTAAGTGTTTCGGCAACAGTTTC
ACAAGCATTGAAGTACATACCAGAGCTGCAAGAGCAAGTTAAAAAGCTCATGAAGAAGAAAGAAGAGCTCTCGTTTC
AAATTTCGGGTCAAAGAGATCTCGTTTACACCGACCAAAACAGTAAGTCAGAGGAAGGGGTTACAAGCTATGCGTCG
ACAGTTTCTTCGACTAGGCTCAGTGAGACTGAAGTGATGGTCCAAATTTCATCGTTACAGACTGAAAAATGTTCGTT
TGGGAATGTCTTGAGTGGTGTAGAAGAAGATGGGTTGGTTCTTGTGGGTGCTTCATCTTCAAGGTCTCATGGAGAGC
GACTCTTTTACTCTATGCATCTTCAGATAAAAAATGGCCAGGTGAATTCCGAAGAATTAGGTGATAGATTGTTGTAC
TTGTACGAGAAATGTGGACACTCGTTTACATGA

>SEQIDNO:90
ATGTGTGCACTTGTCCCTCCATTATATCCCAATTTCGGCTGGCCTTGCGGAGATCATAGCTTCTATGAAACCGACGA
CGTATCCAACACGTTTCTTGATTTTCCGTTGCCGGACTTGACGGTGACTCATGAGAATGTGTCGTCTGAGAATAACA
GAACATTACTAGACAATCCCGTGGTGATGAAGAAGCTTAATCACAACGCGAGTGAACGTGAGCGTCGCAAGAAGATC
AACACAATGTTCTCATCTCTTCGTTCTTGTCTTCCTCCCACCAATCAAACGAAGAAGTTAAGTGTTTCGGCAACAGT
TTCACAAGCATTGAAGTACATACCAGAGCTGCAAGAGCAAGTTAAAAAGCTCATGAAGAAGAAAGAAGAGCTCTCGT
TTCAAATTTCGGGTCAAAGAGATCTCGTTTACACCGACCAAAACAGTAAGTCAGAGGAAGGGGTTACAAGCTATGCG
TCGACAGTTTCTTCGACTAGGCTCAGTGAGACTGAAGTGATGGTCCAAATTTCATCGTTACAGACTGAAAAATGTTC
GTTTGGGAATGTCTTGAGTGGTGTAGAAGAAGATGGGTTGGTTCTTGTGGGTGCTTCATCTTCAAGGTCTCATGGAG
AGCGACTCTTTTACTCTATGCATCTTCAGATAAAAAATGGCCAGGTGAATTCCGAAGAATTAGGTGATAGATTGTTG
TACTTGTACGAGAAATGTGGACACTCGTTTACATGA

>SEQIDNO:91
CCACGTCGTCTGCGCGGCCGACTCCTTCTACGTCGGCCTCCCGATCCCGGTGGTGTCCGCCGGCGAGGAGCTGATGG
CGGGGCGAACCTCATCCACAACGCCTACGAGCGCGACCGCCGGAAGCAGCGCAACGAGCTCTACTCCTCCCTCCGCG
CTCTCCTCCCCGACGCCGATCACACTAAGAAGCTGAGCATCCCGACGACGGTGTCTCGCGTGCTCAAGTACATACCC
GAGCTGCAGAAGCAGGTGGAGAATCTGGAGAGGAAGAAGAAGGAGCTGACGACGACGAGCACCACCAACTGCAAACC
AGGAGTGTTGGGGAGCCAGCTGATGAGCGAGGGCATGGCTCCCATCGTTTCGGCTACCTGCATCAATGACATGGAGA
TCATGGTTCAGGTCAGCTTGTTGAGCAATGTGGCGGGTTCAGTTCTTCCTCTCTCCAAGTGTATCAAAGTACTGGAG
AACGAAGGTCTTCACTTCATCAGTTCATCGACTTCCTCCGGATTTGGGAACAGGACATTCTACAGTATCCATCTTCA

GAGAAGTGAAGGAACGATCAACGAGGAGTGCCCAGCATTTTGTGAAAGGTTGGAGAAAGTCGTCAGGAACAAAGCAA
AGCTT

>SEQIDNO:92
ATGGAGCACCAGCTGTTCGATGACCCCTTCTCTAGCAGCATCTCGTCGCTGGAGGCGGACATCTTCTCCGCCGGCGG
CCAGCTGCCGTCGCCGCCGTGGCCGGACCTCGACCTCGACCTCGACGACGACGACGGCATCCACGACCTCTCCGCGC
CGGCCGGCAACCCCACCTCTTCAGGAGGCTATGGCTCGGGCGGAGGCTCCCACAGGAAGATCAGCCACAACGCGTAC
GAGCGTGACCGCCGGAAGCAGCTCAACGAGCTCTACTCCTCGCTCCGCTCCCTCCTCCCCGACGCTGACCACACTAA
GAAGCTGAGCATCCCCACCACGGTCTCCCGAGTTCTCAAGTACATCCCCGAGCTGCAGAAGCAGGTGGACAACCTGG
AGAGGAGGAAGAAGGAGCTGACGAACGCCAACTGCAAACCAGGAGTTCTGAACACGAGCCAGATTGTAACTCCCATT
GTTTCTGCTACTTGCCTCAACGATACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAACGTGGCTGCCACAAGTCT
TCCTCTGTCCAAGTGCATAAAAGTGATGGAGAACGAAGGCCTTCACCTAATTAGTTCATCAACTTACTCCACCTTCG
ACAACAGGACATTCTATAGCCTCCATGTTCAGAGAAGTCAAAGAACGATGAAGGAGGAGTGCCCAGCATTCTGCGAT
GAACTGGAGAGGATTATC

>SEQIDNO:93
ATGGACCATCAGCTGTTCGACGACCCCTTCGGGAGCAGCATCTCGTCGCTGGAGGCGGACATCTTCTCCGCCGGCGG
CGGCGGACAGCTGCCGTCGCCGCCGTGGCCGGACCTCGACCTCGACGACGACTACGACATACACGACCTCTCCGCGC
CGGCCGCCAACGCCGCCACCTCCTCGGGCGGCGGCTATGGCTCCGGCGGCTCCGGCAGGAAGCTCAGCCACAACGCA
TACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTCTACTCCTCGCTCCGATCCCTCCTCCCGGACGCTGATCACAC
TAAGAAGCTGAGCATCCCCACCACCGTGTCCCGAGTTCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGATAACC
TGGAGAGGAGGAAGAAGGAGCTGACCAACGCCAACTGCAAGCCGGGAGTTCTCAACACCAAAGAGATCGTAACTCCC
ATTGTTTCTGCTACTTGCCTTAACGACACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAATGTGGCCGCCACAGC
TCTCCCTCTCTCCAAGTGCATAAAGGTGCTAGAAAACGAAGGCCTTCTCCTCGTCAGCTCATCAACCTACTCCACCT
TCGAGAACAAGACATTCTATAGCCTCCATCTTCAGAGAAGTCAAAGAACGATGAAGGAGCAGTGCCCAGGATTCTGC
GACGAACTGGAGAAGATCGTCAGGAAGAAAGCAGGGGCG

>SEQIDNO:94
GCCTCGTGCCGGCGGGTGCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACGGACTGGAGAAGAAGAAGGAGGA
GCTGACGCGCGCCAACTGCAAGCCCGGCGTGCTGACCATGAAGGAGAACATGGCTCCGATCGTGTCCGCCACCTGCC
TCGATGACAGAGAAATCATGGTCCAGGTCAGCCTGGTGAGCACCATGGCCGGAGTTCTGCCCATGTCCAAGTGCATC
AAGGTGCTGGAGAACGAAGGCCTACGCCTCATCAGCTCGTCCACTTCCGCGTTTCACAACAGGACGTTCTATAGCCT
CCATCTTCAGAGAACCCAACGGACGATGAGCAAGGAGTGTCCGGCATTTTGTGAAGAACTGGAGAACGCCCTGACGC
AAAAGGCAGGACTACGTCTACATCACCACCAG

>SEQIDNO:95
ATGGATCACCAGCTGTACGGCGACCCCTCCGCGAGCAGCTTCTCTCCGCTGGAGGCACAGATCTTCTCCGGCCAGCT
GCCGCCGTCGTCAACGCCATGGCCAAATCTCGACGTTGACCTCGCCCTGGACCTCGACGTTCTCGAGGATGACATCG
TCCGGGAGCTCTCTGCTGGCACAGTGGCAAACGCGGCATCGTCAGGTTCCGGCTCCGGCGCCCACAAGAAGCTCAGC
CACAACGCGTACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTATACCTCTCGCTCCGTTCTCTCCTCCCGGACGC
CGACCACACCAAGAAGCTGAGTATTCCGACGACGGTGTGTCGAGCGCTCAAGTACATCCCCGAGCTGCAGAAACAGG
TCGAGAATCTGGAGAAGAAGAAAGAGAAACTGGCTAGTGCCAACTGCAAACCAGGGGTACTGAGCGTGACCGGCAGC
ATAGCTCCAACTGTGTCCGCTACTTGCCTCAACCACAAGGAAATCATGGTTCAGATTAGCTTGCTGAGAGATACAGA
TGCTTCTACAGCTCTACCTCTTTCCAAGTGTATAAATGTACTGGAGAACGAAGGACTTCAGCTCATCAGTTCATCGA
CTTCCTCCACCTTTGGGAACAAAACGTTCTATAACCTCCATCTTCAGAGAAGTCAAGGAGCCACTAAACATGGAGTG
CCCATCGTTTTG

>SEQIDNO:96
CGCAGATCTTCTCCAGCCAGCTGCCGCCGTCACCGCCGTGGCCGAATCTCGATGTTGACGTTGACCTGGACCTCGAC
GTTCTTGAGGACGACGTCGTCCGCGAACTCTCAGGGAGGCCGGCAAACGCGGCATCGTCAGGCTCCGGCTCCGGCGG
CCCCGGCTCCCACAAGAAGCTCAGTCACAACGCGTACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTCTACCTCT
CACTCCGTTCTCTCCTGCCGGACGCCGACCACACTAAGAAGCTGAGTATTCCGACGATGGTGTGTCGAGCGCTCAAG
TACATCCCGAGCTGCAGAAACAGGTCGAGAATCTGGAGAAGAAGAAAGAGAAACTTGCTAGTTCCAACTGCAAACCA
GAGGTACTGAGCGCAAGCGGCAGCATAGCTCTAACTGTGTCCGCTACTTGCCTCAACGACAAGGAAATCATGGTTCA
GATTAGCTTGCTGAGACATACGGATGCTGCTACAGCTCTACCTCTTTCCAAGTGTATAAATGTACTGGAGAACGAAG
GACTTGAGCTCGTCAGTTCATCGACTTCCTGCACCTTTGGGAACAAAATGTTCTATAACCTCCATCTTCAGAGAAGT
CAAGGAGCGCTAACATGGGAGTGTCCATCCTTCTGTGACAAATTGGAACAAGCAATCAGGAAAACAGCAGGATTA

>SEQIDNO:97
CGACGGAAGCAGCTCAACGACCTTTACTCCTCGCTCCGCTCCCTCCTCCCGGACGCTGACCACACCAAGAAGCTGAG
CATCCCCACCACCGTGTCCCGAGTCCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACAACCTGGAGAGGAGGA
AGCGGGAGCTGACCAACGCCAACTGCAAGCCGGGAGTTCTCAACACCAGCGAGATCGTAACTACTCCCATTGTTTCT
GCTACTTGCCTCAACGACACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAATGTGGCAGCCACGGCTCTCCCTCT
CTCCAAGTGCATAAAGGTGCTGGAGGACGCAGGCCTTCACCTCATCAGCTCATCAACCTACTCCACCTTTGGGAACA
AGACATTCTATAGCCTCCATCTTCAGGTGTGCATGCATGTTCATTCAATGGTTCCTGCCGTTTCCTTCAATTTTTTT
ATC

>SEQIDNO:98
MCALVPPLYPNFGWPCGDHSFYETDDVSNTFLDFPLPDLTVTHENVSSENNRTLLDNPVVMKKLNHNASERERRKKI
NTMFSSLRSCLPPTNQTKLSVSATVSQALKYIPELQEQVKKLMKKKEELSFQISGQRDLVYTDQNSKSEEGVTSYAS
TVSSTRLSETEVMVQISSLQTEKCSFGNVLSGVEEDGLVLVGASSSRSHGERLFYSMHLQIKNGQVNSEELGDRLLY
LYEKCGHSFT

>SEQIDNO:99
MCALVPPLYPNFGWPCGDHSFYETDDVSNTFLDFPLPDLTVTHENVSSENNRTLLDNPVVMKKLNHNASERERRKKI
NTMFSSLRSCLPPTNQTKKLSVSATVSQALKYIPELQEQVKKLMKKKEELSFQISGQRDLVYTDQNSKSEEGVTSYA
STVSSTRLSETEVMVQISSLQTEKCSFGNVLSGVEEDGLVLVGASSSRSHGERLFYSMHLQIKNGQVNSEELGDRLL
YLYEKCGHSFT

>SEQIDNO:100
PRRLRGRLLLRRPPDPGGVRRRGADGGANLIHNAYERDRRKQRNELYSSLRALLPDADHTKKLSIPTTVSRVLKYIP
ELQKQVENLERKKKELTTTSTTNCKPGVLGSQLMSEGMAPIVSATCINDMEIMVQVSLLSNVAGSVLPLSKCIKVLE
NEGLHFISSSTSSGFGNRTFYSIHLQRSEGTINEECPAFCERLEKVVRNKAKL

>SEQIDNO:101
MEHQLFDDPFSSSISSLEADIFSAGGQLPSPPWPDLDLDLDDDDGIHDLSAPAGNPTSSGGYGSGGGSHRKISHNAY
ERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTVSRVLKYIPELQKQVDNLERRKKELTNANCKPGVLNTSQIVTPI
VSATCLNDTEIMVQVSLHSNVAATSLPLSKCIKVMENEGLHLISSSTYSTFDNRTFYSLHVQRSQRTMKEECPAFCD
ELERII

>SEQIDNO:102
MDHQLFDDPFGSSISSLEADIFSAGGGGQLPSPPWPDLDLDDDYDIHDLSAPAANAATSSGGGYGSGGSGRKLSHNA
YERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTVSRVLKYIPELQKQVDNLERRKKELTNANCKPGVLNTKEIVTP
IVSATCLNDTEIMVQVSLHSNVAATALPLSKCIKVLENEGLLLVSSSTYSTFENKTFYSLHLQRSQRTMKEQCPGFC
DELEKIVRKKAGA

>SEQIDNO:103
ASCRRVLKYIPELQKQVDGLEKKKEELTRANCKPGVLTMKENMAPIVSATCLDDREIMVQVSLVSTMAGVLPMSKCI
KVLENEGLRLISSSTSAFHNRTFYSLHLQRTQRTMSKECPAFCEELENALTQKAGLRLHHHQ

>SEQIDNO:104
MDHQLYGDPSASSFSPLEAQIFSGQLPPSSTPWPNLDVDLALDLDVLEDDIVRELSAGTVANAASSGSGSGAHKKLS
HNAYERDRRKQLNELYLSLRSLLPDADHTKKLSIPTTVCRALKYIPELQKQVENLEKKKEKLASANCKPGVLSVTGS
IAPTVSATCLNHKEIMVQISLLRDTDASTALPLSKCINVLENEGLQLISSSTSSTFGNKTFYNLHLQRSQGATKHGV
PIVL

>SEQIDNO:105
RRKQLNDLYSSLRSLLPDADHTKKLSIPTTVSRVLKYIPELQKQVDNLERRKRELTNANCKPGVLNTSEIVTTPIVS
ATCLNDTEIMVQVSLHSNVAATALPLSKCIKVLEDAGLHLISSSTYSTFGNKTFYSLHLQVCMHVHSMVPAVSFNFF
I

Literature

1. Bailey, P.C., Martin C., Toledo-Ortiz ,G., Quail ,P.H., Huq, E., Heim, M.A., Jakoby, M., Weber, M., Weisshaar, B.(2003). Update on the Basic Helix-Loop-Helix Transcription Factor Gene Family in *Arabidopsis thaliana*. Plant Cell **15**(11), 2497-2502.

2. Brownlie, P., Ceska, T.A., Lamers, M., Romier, C., Stier, G., Teo, H., Suck, D. (1997). The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. Structure **5**, 509-520.

3. Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. **20**, 735-747.

4. Kang, H.-G., Foley, R.C., Onate-Sanchez, L., Lin, C., Singh, K.B. (2003). Target genes for OBP3, a Dof transcription factor, include novel basic helix-loop-helix domain proteins inducible by salicylic acid. Plant J **35**, 362-372.

5. Ling H.Q., Bauer P., Bereczky Z., Keller B., Ganal M. (2002). The tomato fer gene encoding a bHLH protein controls iron-uptake responses in roots. Proc. Natl. Acad. Sci. USA 2002; **99**:13938-13943

6. Martinez-Garcia, J.F., Huq, E., Quail, P.H. (2000). Direct Targeting of Light Signals to a Promoter Element-Bound Transcription Factor. Science **288**, 859-863.

7. Ogo, Y., Itai, R.N., Nakanishi, H., Inoue, H., Kobayashi, T., Suzuki, M., Takhashi, M., Mori, S., Nishizawa, N.K. Isolation and characterization of IRO2, a novel iron-regulated bHLH transcription factor in graminaceous plants. J. Exp. Bot. **57**, 2867-2878.

8. Ramsay, N.A., Walker, A.R., Mooney, M., Gray, J.C. (2003). Two basic-helix-loop-helix genes (*MYC-146* and *GL3*) from *Arabidopsis* can activate anthocyanin biosynthesis in a white-flowered *Matthiola incana* mutant. Plant Mol. Biol. **52**, 679-687.

9. Toledo-Ortiz, G., Huq, E., and Quail, P.H. (2003). The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell **15**(8), 1749-1770.

10. Tominaga, R., Iwata, M., Okada, K.,, Wada, T. (2007). Functional Analysis of the Epidermal-Specific MYB Genes *CAPRICE* and *WEREWOLF* in *Arabidopsis*. Plant Cell **19**(7), 264-2277.

11. Vorwieger, A., Gryczka, C., Czihal, A., Douchkov, D., Tiedemann, J., Mock, H.-P., Jakoby, M., Weisshaar, B., Saalback, I., Baumlein, H. (2007). Iron assimilation and transcription factor controlled synthesis of riboflavin in plants. Planta **226**, 147-158.

12. Wang H.-Y., Klatte, M., Jakoby, M., Baumlein, H., Weisshaar, B., Bauer, P. (2007) Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in *Arabidopsis thaliana*. Planta **226**, 897-908.

13. Yuan, Y., Wu, H., Wang, N., Li, J., Zhao, W., Du, J., Wang, D, Ling, H.-Q. (2008). FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostatis in *Arabidopsis*.

14. Zhang, F., Gonzalez, A., Zhao, M., Payne, T., Lloyd, A. (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. Development **130**, 4859-4869.

We Claim:

1. A method of producing a transgenic plant having increased heat stress tolerance relative to a wild type control plant, comprising:

transforming a plant, a plant tissue culture, or a plant cell with a vector comprising a nucleic acid construct that comprises a nucleic acid encoding a basic helix-loop-helix (bHLH) subgroup 1b polypeptide selected from the group consisting of bHLH38, bHLH39, bHLH100 and bHLH101 to obtain a transformed plant, a transformed plant tissue culture, or a transformed plant cell with increased bHLH subgroup 1b gene expression or activity as compared to a wild type plant of the same species not transformed with said vector, and

growing said transformed plant or regenerating a plant from said transformed plant tissue culture or transformed plant cell, wherein the nucleic acid is selected from SEQ ID Nos: 1-17, 29-52, 79-83, and 89-97.

2. The method of claim 1, wherein said nucleic acid construct comprises a constitutive promoter, an inducible promoter or a tissue specific promoter.

3. The method of claim 2, wherein said tissue specific promoter is a root promoter.

4. The method of claim 1, wherein said bHLH subgroup 1b polypeptide is bHLH38, and wherein said nucleic acid is selected from SEQ ID Nos: 29-52.

5. The method of claim 1, wherein said bHLH subgroup 1b polypeptide is bHLH39, and wherein said nucleic acid is selected from SEQ ID Nos: 1-17.

6. The method of claim 1, wherein said bHLH subgroup 1b polypeptide is bHLH100, and wherein said nucleic acid is selected from SEQ ID Nos: 89-97.

7. The method of claim 1, wherein said bHLH subgroup 1b polypeptide is bHLH101, and wherein said nucleic acid is selected from SEQ ID Nos: 79-83.

Date Reçue/Date Received 2020-08-07

Figures

Figure 1

ClustalW alignment of AtbHLH proteins

```
AtbHLH39    -MCALVPPLFPNFGWPSTGEYDSYYLAGDILNNGGFLDFPVPEETYGAVTAVTQHQNSFG
AtbHLH38    -MCALVPSFFTNFGWPSTNQYESYYGAGDNLNNGTFLELTVP-QTY----EVTHHQNSLG
AtbHLH100   -MCALVPPLYPNFGWPCG--DHSFYETDDVSN--TFLDFPLP-------DLTVTHEN---
AtbHLH101   MEYPWLQSQVHSFSPTLHFPSFLHPLDDSKSHNINLHHMSLS---------HSNNTNSNN
               . : .   .*. .       .   ..  :   : .:.:.            : *

AtbHLH39    VSVSSEGN-EIDNNPVVVKKLNHNASERDRRRKINSLFSSLRSCLPASGQSKKLSIPATV
AtbHLH38    VSVSSEGN-EIDNNPVVVKKLNHNASERDRRKKINTLFSSLRSCLPASDQSKKLSIPETV
AtbHLH100   --VSSENNRTLLDNPVVMKKLNHNASERERRKKINTMFSSLRSCLPPTNQTKKLSVSATV
AtbHLH101   NNYQEEDR----GAVVLEKKLNHNASERDRRRKLNALYSSLRALLPLSDQKRKLSIPMTV
               ..*..    .  *: **********:**:*:*:::****: ** :.*.:***:. **

AtbHLH39    SRSLKYIPELQEQVKKLIKKKEELLVQISGQRNTECYVK--QPPKAVANYISTVSATRLG
AtbHLH38    SKSLKYIPELQQQVKRLIQKKEEILVRVSGQRDFELYDK--QQPKAVASYLSTVSATRLG
AtbHLH100   SQALKYIPELQEQVKKLMKKKEELSFQISGQRDLVYTDQNSKSEEGVTSYASTVSSTRLS
AtbHLH101   ARVVKYIPEQKQELQRLSRRKEELLKRISRKTHQEQLRNKAMMDSIDSSSSQRIAANWLT
            :: :***** ::::::* ::***:  ::* : .     :     .  :.  . :::. *

AtbHLH39    DNEVMVQISSSKIHNFSISNVLSGLEEDRFVLVDMSSSRSQGERLFYTLHLQVEKIENYK
AtbHLH38    DNEVMVQVSSSKIHNFSISNVLGGIEEDGFVLVDVSSSRSQGERLFYTLHLQVENMDDYK
AtbHLH100   ETEVMVQISSLQTEKCSFGNVLSGVEEDGLVLVGASSSRSHGERLFYSMHLQIK---NGQ
AtbHLH101   DTEIAVQIATSKWT--SVSDMLLRLEENGLNVISVSSSVSSTARIFYTLHLQMRG--DCK
            :.*: **::: :    *..::*  :**: : ::. *** *   *:**::***:.   : :

AtbHLH39    LNCEELSQRMLYLYEECGNSYI 258 (SEQ ID NO: 122)
AtbHLH38    INCEELSERMLYLYEKCENSFN 253 (SEQ ID NO: 123)
AtbHLH100   VNSEELGDRLLYLYEKCGHSFT 242 (SEQ ID NO: 124)
AtbHLH101   VRLEELINGMLLGLRQS----- 240 (SEQ ID NO: 125)
            :. *** : :*   .:.
```

ClustalW alignment of AtbHLH proteins

```
AtbHLH39    -MCALVPPLFPNFGWPSTGEYDSYYLAGDILNNGGFLDFPVPEETYGAVTAVTQHQNSFG
AtbHLH38    -MCALVPSFFTNFGWPSTNQYESYYGAGDNLNNGTFLELTVP-QTY----EVTHHQNSLG
AtbHLH100   -MCALVPPLYPNFGWPCG--DHSFYETDDVSN--TFLDFPLP-------DLTVTHEN---
AtbHLH101   MEYPWLQSQVHSFSPTLHFPSFLHPLDDSKSHNINLHHMSLS---------HSNNTNSNN
                . : .   .*. .       .   ..  :   : .:.:.            : *

AtbHLH39    VSVSSEGN-EIDNNPVVVKKLNHNASERDRRRKINSLFSSLRSCLPASGQSKKLSIPATV
AtbHLH38    VSVSSEGN-EIDNNPVVVKKLNHNASERDRRKKINTLFSSLRSCLPASDQSKKLSIPETV
AtbHLH100   --VSSENNRTLLDNPVVMKKLNHNASERERRKKINTMFSSLRSCLPPTNQTKKLSVSATV
AtbHLH101   NNYQEEDR----GAVVLEKKLNHNASERDRRRKLNALYSSLRALLPLSDQKRKLSIPMTV
               ..*..    .  *: **********:**:*:*::****: ** :.*.:***:. **

AtbHLH39    SRSLKYIPELQEQVKKLIKKKEELLVQISGQRNTECYVK--QPPKAVANYISTVSATRLG
AtbHLH38    SKSLKYIPELQQQVKRLIQKKEEILVRVSGQRDFELYDK--QQPKAVASYLSTVSATRLG
AtbHLH100   SQALKYIPELQEQVKKLMKKKEELSFQISGQRDLVYTDQNSKSEEGVTSYASTVSSTRLS
AtbHLH101   ARVVKYIPEQKQELQRLSRRKEELLKRISRKTHQEQLRNKAMMDSIDSSSSQRIAANWLT
            :: :***** ::::::* ::***:  ::* : .    :    .  :.  . :::. *

AtbHLH39    DNEVMVQISSSKIHNFSISNVLSGLEEDRFVLVDMSSSRSQGERLFYTLHLQVEKIENYK
AtbHLH38    DNEVMVQVSSSKIHNFSISNVLGGIEEDGFVLVDVSSSRSQGERLFYTLHLQVENMDDYK
AtbHLH100   ETEVMVQISSLQTEKCSFGNVLSGVEEDGLVLVGASSSRSHGERLFYSMHLQIK---NGQ
AtbHLH101   DTEIAVQIATSKWT--SVSDMLLRLEENGLNVISVSSSVSSTARIFYTLHLQMRG--DCK
            :.*: **::: :    *..::*  :**: : ::. *** *   *:**::***:.   : :

AtbHLH39    LNCEELSQRMLYLYEECGNSYI 258 (SEQ ID NO: 122)
AtbHLH38    INCEELSERMLYLYEKCENSFN 253 (SEQ ID NO: 123)
AtbHLH100   VNSEELGDRLLYLYEKCGHSFT 242 (SEQ ID NO: 124)
AtbHLH101   VRLEELINGMLLGLRQS----- 240 (SEQ ID NO: 125)
            :. *** : :*   .:.
```