Investigating Drought-Resistance Biomarkers in Indica Rice:

Computational Gene Expression Analysis Toward Enhanced Food Security

Tony Wang

Non-Trivial Fellow, Cohort II (Summer 2023)

Introduction

In the face of mounting challenges posed by the disruptive impacts of climate change, potential nuclear winters, and advanced global warfare, ensuring a consistent and secure food supply stands as a pivotal aspect of bolstering human civilization resilience and reducing human suffering. The urgency of this matter cannot be overstated, particularly given the potential ramifications for future populations' safety and longevity. The world is witnessing an escalation in climate volatility, underscoring the critical necessity of securing our food supply against stress-induced disruptions.

Currently, the biggest bottleneck in the pursuit of a robust food supply chain lies in the rapidly increasing prevalence of severe drought conditions, exacerbated by the changing climate patterns. These arid circumstances pose a grave threat to agricultural productivity, potentially sparking mass crop failures and devastating famines. This will especially affect traditionally underserved communities, which are more reliant on yearly crop yields. To this end, prioritizing the development of drought-resistant crop varieties is of critical importance.

Within the realm of genetic modification, the innovative precision of CRISPR-Cas9 technology has emerged as a potential tool for enhancing drought tolerance in crops. However, while the potential of CRISPR-Cas9 is undeniably profound, its implementation is marred by formidable challenges. While the technology offers targeted genome editing, the threat of off-target effects is severe, carrying the risk of unintended genetic alterations due to network effects. The intricate complexity of biological pathways further complicates matters, as gene editing endeavors could inadvertently disrupt vital genes or divert resources away from essential processes, thereby undermining overall growth and yield.

For example, one proposed CRISPR-Cas9 approach involves manipulating the abscisic acid (ABA) signaling pathway to regulate water stress responses. However, this is a complex task due to the intricate network of genes involved. Moreover, the intricate nature of the ABA signaling network could lead to unintended consequences or incomplete control over drought responses, demanding a deep understanding of plant physiology scientists still do not have. Thus, directly incorporating desiccation tolerance genes through CRISPR-Cas9 carries the very serious risk of disrupting the delicate evolutionarily perfected balance of growth and development, leading to abnormal phenotypes with unintended consequences.

Intriguingly, a paradigm shift is emerging in the form of gene expression analysis, an innovative new domain that has traditionally been concentrated within the realm of human genetics and disease detection. However, harnessing the power of gene expression analysis within the context of agriculture holds immense promise. Currently, this field is very understudied in agriculture, as gene expression is almost

always only done to study humans and animals. Thus, very few gene expression studies have been done specifically looking to optimize drought traits, especially in the rice strain selected, Indica rice.

A gene expression based approach entails identifying genetic biomarkers associated with drought resistance, then subsequently enabling a selective breeding strategy to gradually enhance the expression of drought-responsive genes. Identifying biomarkers for drought resistance can allow scientists to utilize selective breeding to gradually increase the overexpression of helpful drought-responsive genes. This method, while resource-intensive, circumvents the risk of unintended genetic anomalies and unanticipated long-term effects, providing a robust pathway towards sustainable enhancement of drought resilience. Moreover, the feasibility of this approach is accentuated by the availability of microarray tests for easy biomarker identification.

One of the most common grains that many global populations, especially those in underdeveloped areas, depend on is rice. Rice is a staple food for many populations; beyond its nutritional richness and extended shelf life, rice is a crucial caloric source for prolonged sustenance. Traditional rice varieties, like indigenous strains such as Dagad deshi in India, possess many desirable traits, including their unique responses to abiotic stresses. Drought or osmotic stress is one of the major abiotic stresses afflicting crop plants in India. The resilience exhibited by these varieties under drought and osmotic stress conditions has captivated the attention of plant breeders and biotechnologists alike, propelling intensive investigations into their underlying genetic mechanisms.

Methods

Dataset Selection

Data was retrieved and loaded through the Gene Expression Omnibus (GEO), a free online National Center for Biotechnology Information (NCBI) data repository containing an extensive array of publicly available gene expression datasets from a massive variety of worldwide biological studies. GSE41647 was deemed the best fit for this study. In this dataset, the transciptomes of two contrasting cultivars of Indica rice, Dagad deshi (tolerant) and IR20 (susceptible), were analyzed under control and stress conditions to elucidate the differences in their responses to drought stress using Affymetrix microarray platform. The Affymetrix microarray platform leverages chip-based technology reliant on oligonucleotide

probes to capture complementary DNA (cDNA) strands. The cDNA was synthesized from the prevalent mRNA molecules.

The dataset was described by the data providers as follows:

> Hydroponically grown seven-day-old seedlings of Dagad deshi and IR20 were subjected to drought stress by placing them on 3 mm Whatmann sheets under light for 3 h and 6 h at 28±1oC. For control samples, seedlings were kept in RGM (root growth medium) for 6 h at 28±1oC. RNA extracted from each sample was hybridized on rice Affymetrix microarrays. Three biological replicates of each sample were used for microarray analysis. Overall, eighteen samples were analyzed representing control, 3 h and 6 h of dehydration stress for each cultivar (Dagad deshi and IR20).

Differential Expression Analysis

GSE41647 was analyzed using the bioinformatics tool GEO2R. GEO2R is a software that processes gene expression values and outputs a table of differentially expressed genes between two user-defined groups (in this case, it was drought resistant and drought susceptible strains). Due to the relative scarcity of gene expression studies concerning indica rice, GEO2R predominantly returned locus location IDs in lieu of gene names. For the sake of clarity, these location IDs were referred to as genes throughout this paper.

In total, 9204 genes were deemed statistically significant by GEO2R within this dataset. T-tests were then used to determine the statistical p values. GEO2R was also used to further calculate fold changes for each gene in the context of drought resistant vs drought susceptible values. The fold change is a ratio of the average expression value of a gene in one group divided by the average expression value in a different group. Fold change can be greater or less than one; some genes are overexpressed while others are underexpressed in drought resistant strains disease compared to normal controls. The fold change values were instrumental in characterizing the degree of gene expression alteration under varying stress conditions. GEO2R was also used to verify a normal distribution of gene expression values. It was confirmed that no outliers were present in the dataset.

After the lists of differentially expressed genes were obtained, they were processed in a flat-file sheet for further analysis. Differentially expressed genes with p values of greater than 0.05 were removed, and the remaining differentially expressed genes were sorted into two categories: overexpression and underexpression. A data filtration step ensued, and any locus IDs containing no corresponding gene title

were removed. Afterward, there were a culminated 5157 loci location IDs left over, each of which had a corresponding associated gene title.

Network Analyses

The finalized list of differentially expressed genes (DEGs) was analyzed in the Search Tool for the Retrieval of Interacting Proteins Database (STRING DB) to construct a comprehensive network of the protein-protein interactions of the DEG products. STRING scans numerous databases to create a graph of input genes and how their protein products interact. STRING also calculates statistically significant Gene Ontology processes and pathways. The genes were further filtered to only loci locations present in the String database. Solitary genes were hidden to improve visual clarity.

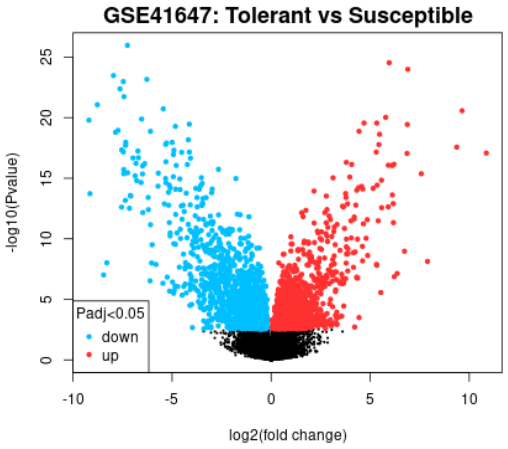
Python Analyses

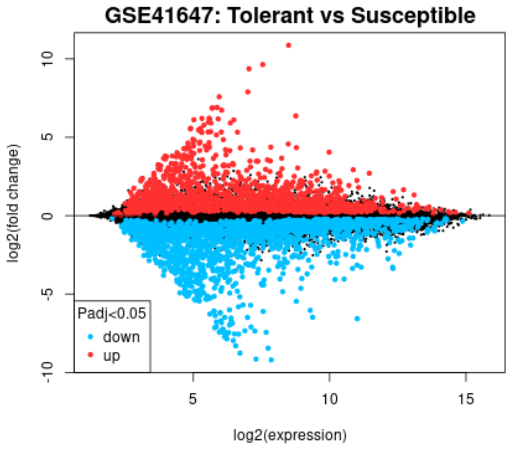
A Colab Notebook coding environemnt was set up, with the necessary MatPlotLib, NumPy, and Pandas libraries imported for creating plots/visualizations, performing numerical computations, and handling tabular data respectively. The GSE41647 flat-file data, having undergone filtration to retain loci locations with corresponding gene titles, was transformed into a Pandas dataframe by uploading it as a tab-separated values (tsv) file.

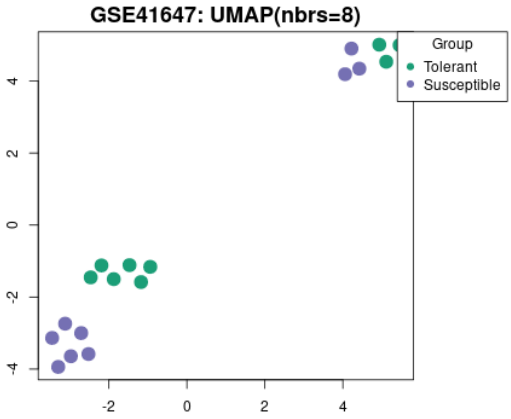
Subsequently, within the dataframe, genes meeting specific criteria were identified: genes with an adjusted p-value less than 0.01 and an absolute log2 fold change greater than 1 were classified as significantly differentially expressed. A volcano plot, a graphical representation of these differentially expressed genes, was generated, depicting two sets of scatter plots—one for significantly differentially expressed genes and another for those that were not significant. Horizontal lines were added at the negative log10 of the adjusted p-value threshold (0.01) to visualize the cutoff for statistical significance. Vertical lines were superimposed at log2 fold change thresholds (-1 and 1) to illustrate the fold change-based significance cutoff. Furthermore, a bar plot spotlighting the most significant genes within the filtered dataset was generated.

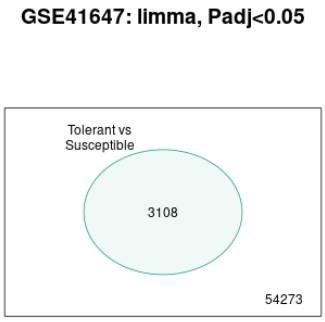
Results

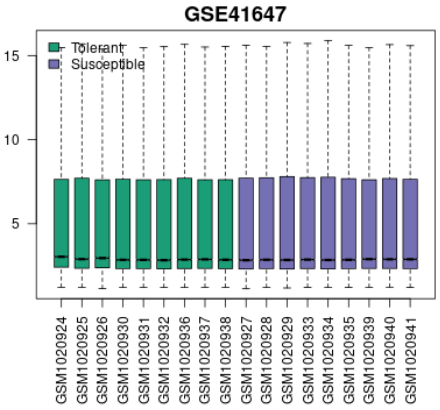
GEO2R Analyses results:

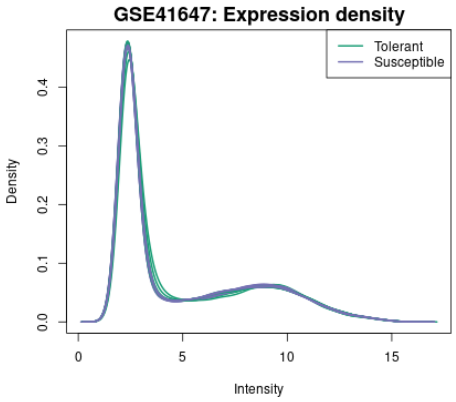


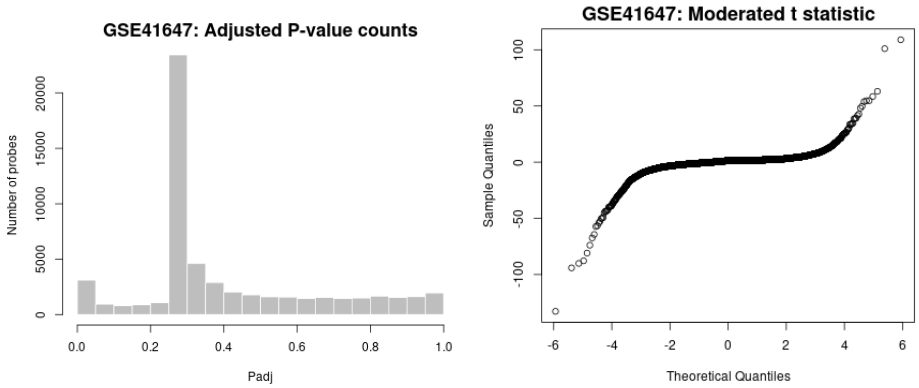


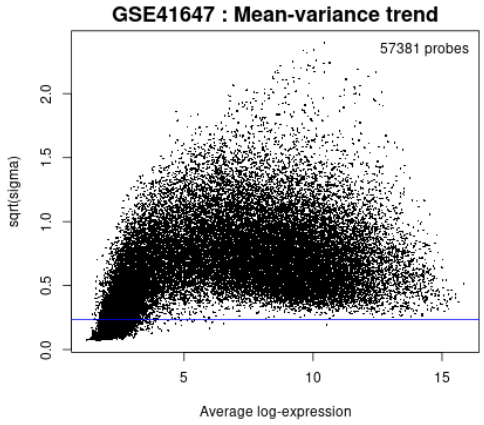




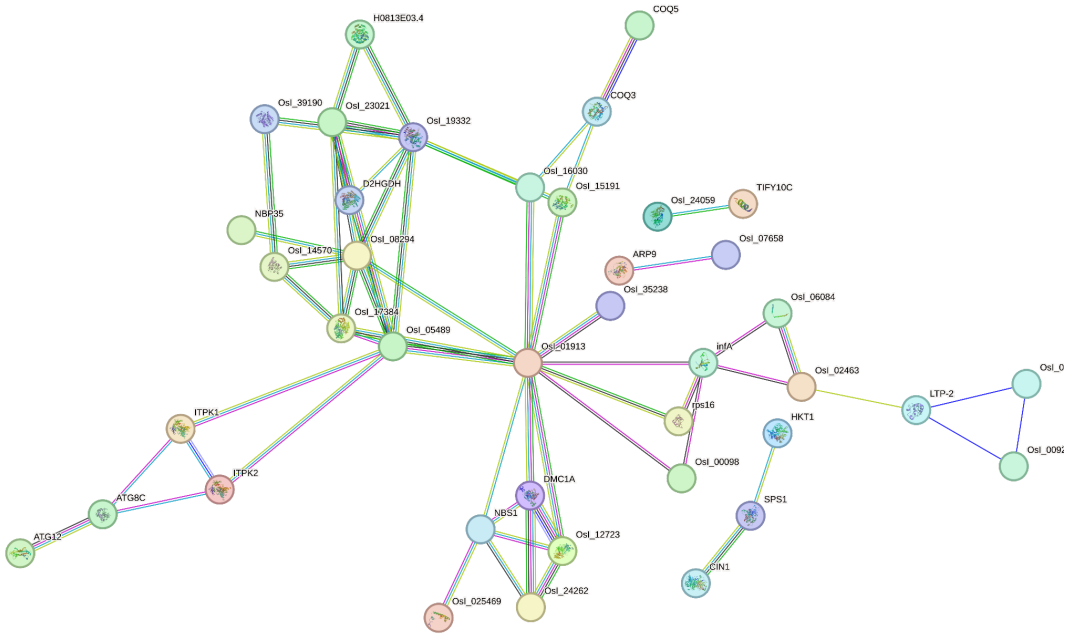








StringDB Analyses Results:



The analysis of differentially expressed genes (DEGs) led to the discovery of three pivotal hub genes—D2HGDH, DMC1A, and COQ3—each having known protein structures. This observation is significant as these genes not only represent potentially functional elements within the context of drought stress response but also offer tangible targets for further in-depth investigation in wetlab studies. The presence of established protein structures elevates the feasibility of dissecting the molecular mechanisms underlying their roles in conferring drought resistance.

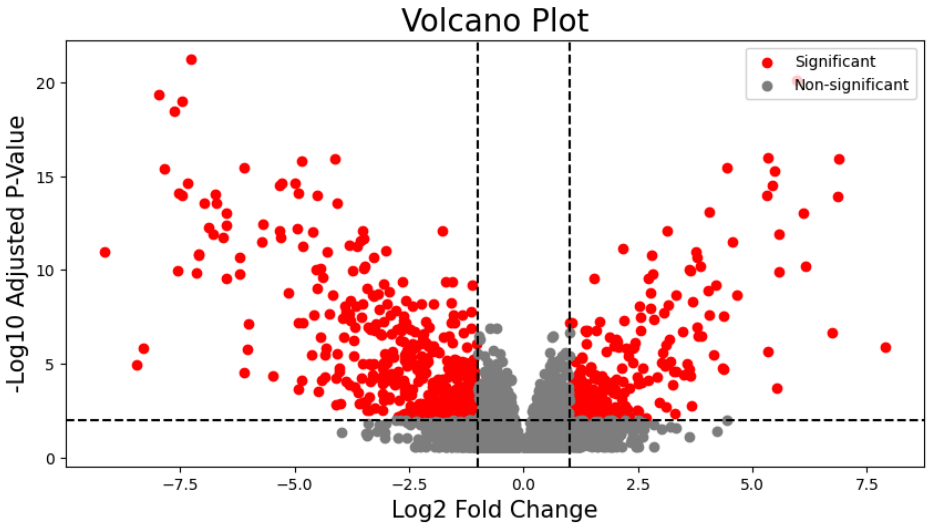
ChatGPT descriptions of the biological functions of the three biomarkers:

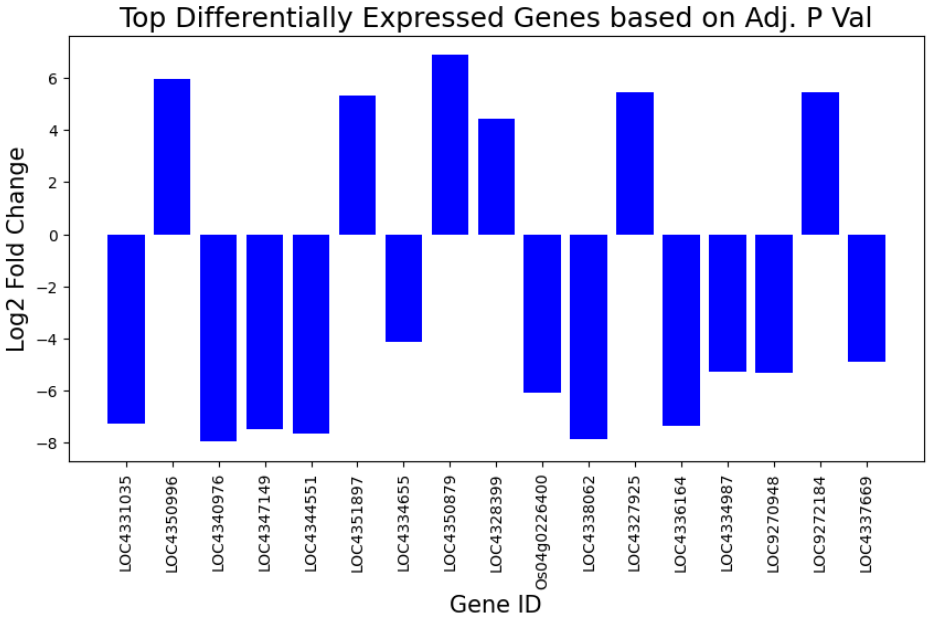
> D2HGDH (D-2-hydroxyglutarate dehydrogenase): D2HGDH emerges as a crucial hub gene, and its possession of a known protein structure instills confidence in its amenability to rigorous study. Previous research has linked D2HGDH to metabolic processes related to D-2-hydroxyglutarate metabolism, a metabolite implicated in various physiological and pathological states. The presence of D2HGDH in the network suggests its potential involvement in modulating responses to drought stress, possibly by influencing metabolic homeostasis under adverse conditions.
>
> DMC1A (Disrupted Meiotic cDNA 1A): DMC1A, an additional hub gene with a confirmed protein structure, beckons for closer scrutiny due to its established relevance to meiotic

recombination processes. Previous studies have illuminated its role in promoting genetic diversity through the homologous recombination of DNA during meiosis. The inclusion of DMC1A in the network underscores its putative involvement in orchestrating genetic adaptations that could underlie drought resistance mechanisms.

COQ3 (4-hydroxybenzoate polyprenyltransferase): COQ3, a third vital hub gene with a known protein structure, holds promise as a key participant in biological pathways crucial to plant resilience. Its implication in the biosynthesis of coenzyme Q, an essential component of the electron transport chain, underscores its potential relevance to energy metabolism and stress response. The identification of COQ3 as a hub gene accentuates its prospective role in sustaining energy homeostasis during drought conditions.

The identification of D2HGDH, DMC1A, and COQ3 as critical hub genes with known protein structures is very important for deciphering the molecular intricacies of drought resistance. The confirmation of these genes' protein structures gives them with heightened scientific tractability, enabling focused empirical investigations into their roles and mechanisms of action. The StringDB network, together with the GEO2R and Python analyses, collectively bridge theoretical understanding with empirical observations. This multifaceted approach substantiates the potential functional relevance of DEGs and sets the stage for further probing their roles in conferring drought resilience.





Discussion

This project is a comprehensive exploration of drought resistance mechanisms in rice, a crucial study given the escalating global challenges posed by climate change and other environmental stressors. Leveraging bioinformatics tools, empirical analyses, and computational methods, this research unveils

critical insights into the genetic underpinnings of drought response, while also identifying tangible targets for further investigation.

The identification of D2HGDH, DMC1A, and COQ3 as central hub genes with known protein structures represents a cornerstone of this study. These genes not only lend themselves to potential functional dissection but also offer a bridge between theoretical conjecture and empirical investigation. D2HGDH, associated with D-2-hydroxyglutarate metabolism, is poised to play a role in metabolic adjustments during drought stress. DMC1A, a facilitator of meiotic recombination, carries the promise of influencing genetic diversity adaptations under adverse conditions. COQ3's involvement in coenzyme Q biosynthesis implies its relevance to energy metabolism and stress responses. These findings collectively underscore the potential multi-dimensional roles these genes play in orchestrating drought resistance strategies in rice.

The constructed StringDB network visually captures the web of interactions among DEGs, potentially uncovering hidden relationships and emergent pathways pivotal for drought adaptation. This holistic view provides a scaffold upon which to construct hypotheses about functional relationships and regulatory dynamics. The GEO2R and Python analyses also provide valuable insights into the extent of differential expression and facilitating insightful visualizations of significance thresholds. These analyses not only substantiate the relevance of identified genes but also offer a platform for data interpretation and hypothesis generation.

The broader implications of this study extend beyond the confines of rice drought resistance research. The methods employed here—ranging from gene expression analysis and network construction to computational exploration—represent a blueprint for unraveling intricate biological phenomena. The convergence of multi-disciplinary tools in this research underscores the necessity of a holistic approach when addressing complex challenges.

In conclusion, this study analyzed the genetic landscape of drought resistance in rice, culminating in the identification of pivotal hub genes with known protein structures and a comprehensive understanding of their potential roles. This study offers insights that extend beyond rice to the broader realm of plant biology. As humanity grapples with ever-increasing environmental uncertainties, studies of this nature contribute to the arsenal of solutions necessary for securing future food supplies and building resilient civilizations.

Supplemental

Gene analyses from StringDB:

Your Current Organism:
Oryza sativa Indica
NCBI taxonomy Id: 39946
Other names: Indian rice, Indica rice, O. sativa Indica Group, Oryza sativa (indica cultivar-group), Oryza sativa (indica group), Oryza sativa Indica Group, Oryza sativa subsp. indica, Oryza sativa subsp. indica Kato, Oryza sp. Poi-6, long-grained rice

:

IBH1	*Transcription factor IBH1; Atypical and probable non DNA-binding bHLH transcription factor that acts as negative regulator of cell elongation and plant development. Binds the transcription factor ILI1 and forms a heterodimer of antagonistic bHLH transcription factors that function downstream of BZR1 to mediate brassinosteroid regulation of cell elongation and lamina inclination (By similarity). (202 aa)*
ACO1	*1-aminocyclopropane-1-carboxylate oxidase 1; Belongs to the iron/ascorbate-dependent oxidoreductase family. (322 aa)*
CHS1	*Chalcone synthase 1; The primary product of this enzyme is 4,2',4',6'- tetrahydroxychalcone (also termed naringenin-chalcone or chalcone) which can under specific conditions spontaneously isomerize into naringenin; Belongs to the thiolase-like superfamily. Chalcone/stilbene synthases family. (398 aa)*
OsI_12723	*DNA repair protein RAD51 homolog; Binds to single and double-stranded DNA and exhibits DNA- dependent ATPase activity. Unwinds duplex DNA. Component of the meiotic recombination pathway. Seems to play a role in mediating chromosome homology search, chromosome pairing and synapsis at early stages and probably chromosome crossing-over at later stages in meiosis. Probably is involved in the repair of meiotic double strand breaks (DBSs) and in homologous recombination; Belongs to the RecA family. RAD51 subfamily. (339 aa)*
STLP1	*Sialyltransferase-like protein 1; Possesses sialyltransferase-like activity in vitro. Transfers sialic acid to the oligosaccharide Gal-beta-1,3-GalNAc and to glycoproteins such as asialofetuin, alpha-1-acid glycoprotein (NeuAc- alpha-2,3-Gal-beta-1,3-GalNAc-) and andasialo-alpha-1-acid glycoprotein. The transferred sialic acid is linked to galactose of Gal-beta-1,3-GalNAc through alpha-2,6-linkage. Belongs to the glycosyltransferase 29 family. (393 aa)*

OsI_19806	*Protein N-terminal glutamine amidohydrolase; Mediates the side-chain deamidation of N-terminal glutamine residues to glutamate, an important step in N-end rule pathway of protein degradation. Conversion of the resulting N-terminal glutamine to glutamate renders the protein susceptible to arginylation, polyubiquitination and degradation as specified by the N-end rule. Does not act on substrates with internal or C-terminal glutamine and does not act on non-glutamine residues in any position. (231 aa)*
OsI_24059	*Molybdenum cofactor sulfurase; Sulfurates the molybdenum cofactor. Sulfation of molybdenum is essential for xanthine dehydrogenase (XDH) and aldehyde oxidase (ADO) enzymes in which molybdenum cofactor is liganded by 1 oxygen and 1 sulfur atom in active form. (824 aa)*
HKT1	*Cation transporter HKT1; Seems to be involved in regulation of K(+)/Na(+) homeostasis. Seems to act as a high-affinity sodium transporter, which mediates increased sodium uptake in roots under potassium deficiency and contributes to sodium accumulation and salt toxicity. May also act as a potassium transporter. Belongs to the TrkH potassium transport family. HKT (TC 2.A.38.3) subfamily. (530 aa)*
GSTU1	*Probable glutathione S-transferase GSTU1; Conjugation of reduced glutathione to a wide number of exogenous and endogenous hydrophobic electrophiles; Belongs to the GST superfamily. Tau family. (231 aa)*
DMC1A	*Meiotic recombination protein DMC1 homolog A; Recombinase that may participate in meiotic recombination, specifically in homologous strand assimilation, which is required for the resolution of meiotic double-strand breaks (Probable). Exhibits DNA-dependent ATPase activity when bound to single-stranded DNA (ssDNA). Mediates renaturation of homologous complementary strands as well as assimilation of single strands into homologous supercoiled duplexes leading to D-loop formation. Binds circular single-stranded DNA (ssDNA) and circular double-stranded DNA (dsDNA) in vitro. Catalyzes DNA ho [...] (344 aa)*
OsI_35550	*Methylthioribose-1-phosphate isomerase; Catalyzes the interconversion of methylthioribose-1-phosphate (MTR-1-P) into methylthioribulose-1-phosphate (MTRu-1-P). Belongs to the eIF-2B alpha/beta/delta subunits family. MtnA subfamily. (374 aa)*
ITPK2	*Inositol-tetrakisphosphate 1-kinase 2; Kinase that can phosphorylate various inositol polyphosphate such as Ins(3,4,5,6)P4 or Ins(1,3,4)P3 and participates in phytic acid biosynthesis in developing seeds. Phytic acid is the primary storage form of phosphorus in cereal grains and other plant seeds. Belongs to the ITPK1 family. (349 aa)*

GAM1	*Transcription factor GAMYB; Transcriptional activator of gibberellin-dependent alpha- amylase expression in aleurone cells. Involved in pollen and floral organs development. May bind to the 5'-TAACAAA-3' box of alpha-amylase promoter. (553 aa)*
SPX4	*SPX domain-containing protein 4. (320 aa)*
ARP9	*Actin-related protein 9; Belongs to the actin family. ARP8 subfamily. (586 aa)*
OsI_025469	*Probable histone H2A.2; Core component of nucleosome. Nucleosomes wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template. Histones thereby play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. DNA accessibility is regulated via a complex set of post-translational modifications of histones, also called histone code, and nucleosome remodeling. (135 aa)*
OsI_01913	*CASP-like protein 1E1; Belongs to the Casparian strip membrane proteins (CASP) family. (215 aa)*
RR22	*Two-component response regulator ORR22; Transcriptional activator that binds specific DNA sequence. Functions as a response regulator involved in His-to-Asp phosphorelay signal transduction system. Phosphorylation of the Asp residue in the receiver domain activates the ability of the protein to promote the transcription of target genes. May directly activate some type-A response regulators in response to cytokinins. (696 aa)*
OsI_06217	*Non-structural maintenance of chromosomes element 4; Component of the SMC5-SMC6 complex, that promotes sister chromatid alignment after DNA damage and facilitates double-stranded DNA breaks (DSBs) repair via homologous recombination between sister chromatids. (360 aa)*
TIFY10C	*Protein TIFY 10c; Repressor of jasmonate (JA) responses. Acts as a repressor of JA-induced resistance to the bacterial blight pathogen Xanthomonas oryzae pv. oryzae (Xoo). Regulates JA-induced accumulation of linalool at the transcriptional level of linalool synthase gene LIS. Linalool is important for resistance to bacterial blight pathogen Xoo. (232 aa)*
OsI_02463	*Protein transport protein Sec61 subunit beta; Necessary for protein translocation in the endoplasmic reticulum. (80 aa)*
HOX17	*Homeobox-leucine zipper protein HOX17; Probable transcription factor. (247 aa)*
ITPK1	*Inositol-tetrakisphosphate 1-kinase 1; Kinase that can phosphorylate various inositol polyphosphate such as Ins(3,4,5,6)P4 or Ins(1,3,4)P3 and participates in phytic acid biosynthesis in developing seeds. Phytic acid is*

	the primary storage form of phosphorus in cereal grains and other plant seeds. Belongs to the ITPK1 family. (354 aa)
gstu4	*Glutathione S-transferase. (233 aa)*
Osl_25391	*Eukaryotic translation initiation factor 3 subunit E; Component of the eukaryotic translation initiation factor 3 (eIF-3) complex, which is involved in protein synthesis of a specialized repertoire of mRNAs and, together with other initiation factors, stimulates binding of mRNA and methionyl-tRNAi to the 40S ribosome. The eIF-3 complex specifically targets and initiates translation of a subset of mRNAs involved in cell proliferation. (439 aa)*
PFP-ALPHA	*Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha; Regulatory subunit of pyrophosphate--fructose 6-phosphate 1- phosphotransferase. (617 aa)*
Osl_09866	*HVA22-like protein. (288 aa)*
Osl_08294	*Glutamate dehydrogenase; Belongs to the Glu/Leu/Phe/Val dehydrogenases family. (410 aa)*
Osl_24262	*DNA repair protein REV1; Deoxycytidyl transferase involved in DNA repair. Transfers a dCMP residue from dCTP to the 3'-end of a DNA primer in a template- dependent reaction. May assist in the first step in the bypass of abasic lesions by the insertion of a nucleotide opposite the lesion. Required for normal induction of mutations by physical and chemical agents; Belongs to the DNA polymerase type-Y family. (1100 aa)*
HOX19	*Homeobox-leucine zipper protein HOX19; Probable transcription factor; Belongs to the HD-ZIP homeobox family. Class II subfamily. (292 aa)*
MADS26	*MADS-box transcription factor 26; Probable transcription factor. (222 aa)*
rps16	*30S ribosomal protein S16, chloroplastic. (85 aa)*
RH10	*DEAD-box ATP-dependent RNA helicase 10; Has ATP-dependent RNA helicase activity in vitro. Acts as a thermosensitive RNA chaperone required for normal processing of pre- rRNA intermediates. Required for normal cell division at high temperatures. Required for a primary metabolism adaptation to high temperatures to support thermotolerant growth by regulating gene expression. Partially rescues the yeast rRNA helicase RRP3 mutant which has repressed cell proliferation at 38 degrees Celsius. (472 aa)*

OsI_17384	*Probable gamma-aminobutyrate transaminase 2, mitochondrial; Transaminase that degrades gamma-amino butyric acid (GABA). Belongs to the class-III pyridoxal-phosphate-dependent aminotransferase family. (497 aa)*
OsI_29578	*Protein transport protein SEC23; Component of the coat protein complex II (COPII) which promotes the formation of transport vesicles from the endoplasmic reticulum (ER). The coat has two main functions, the physical deformation of the endoplasmic reticulum membrane into vesicles and the selection of cargo molecules; Belongs to the SEC23/SEC24 family. SEC23 subfamily. (367 aa)*
OsI_14570	*Arginase 1, mitochondrial; Catalyzes the hydrolysis of L-arginine to urea and L- ornithine. The latter can be utilized in the urea cycle or as a precursor for the synthesis of both polyamines and proline. Belongs to the arginase family. (340 aa)*
PIMT2	*Protein-L-isoaspartate O-methyltransferase. (286 aa)*
OsI_14861	*Senescence-specific cysteine protease SAG39; Cysteine protease that may have a developmental senescence specific cell death function during apoptosis, heavy metal detoxification, and hypersensitive response. Belongs to the peptidase C1 family. (339 aa)*
NBP35	*Cytosolic Fe-S cluster assembly factor NBP35; Component of the cytosolic iron-sulfur (Fe-S) protein assembly (CIA) machinery. Required for maturation of extramitochondrial Fe-S proteins. Functions as Fe-S scaffold, mediating the de novo assembly of an Fe-S cluster and its transfer to target apoproteins. Essential for embryo development. (355 aa)*
SWEET 1B	*Bidirectional sugar transporter SWEET1b; Mediates both low-affinity uptake and efflux of sugar across the plasma membrane. (261 aa)*
OEP80	*Outer envelope protein 80, chloroplastic; Belongs to the OEP80 (TC 1.B.33.2) family. (665 aa)*
ATG12	*Ubiquitin-like protein ATG12; Ubiquitin-like protein involved in cytoplasm to vacuole transport (Cvt) and autophagy vesicles formation. Conjugation with ATG5 through a ubiquitin-like conjugating system is essential for its function. ATG12/ATG5 conjugate has an essential role in plant nutrient recycling (By similarity); Belongs to the ATG12 family. (93 aa)*
OsI_15191	*Lipase; Belongs to the AB hydrolase superfamily. Lipase family. (428 aa)*

OsI_00098	*Peptidyl-prolyl cis-trans isomerase; PPIases accelerate the folding of proteins. It catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides; Belongs to the cyclophilin-type PPIase family. (295 aa)*
GRP-1	*Putative glycine-rich cell wall structural protein 1; Responsible for plasticity of the cell wall. (165 aa)*
BRE1B	*E3 ubiquitin-protein ligase BRE1-like 2; E3 ubiquitin-protein ligase that monoubiquitinates H2B to form H2BK143ub1. H2BK143ub1 gives a specific tag for epigenetic transcriptional activation and is also prerequisite for H3K4me and maybe H3K79me. It thereby plays a central role in histone code and gene regulation. Forms a ubiquitin ligase complex in cooperation with the E2 enzyme UBC2/RAD6. (844 aa)*
OsI_05489	*L-lactate dehydrogenase; Belongs to the LDH/MDH superfamily. (353 aa)*
OsI_23021	*Formate dehydrogenase, mitochondrial; Catalyzes the NAD(+)-dependent oxidation of formate to carbon dioxide. Involved in the cell stress response; Belongs to the D-isomer specific 2-hydroxyacid dehydrogenase family. FDH subfamily. (378 aa)*
ATG8C	*Autophagy-related protein 8C; Ubiquitin-like modifier involved in autophagosomes formation. May mediate the delivery of the autophagosomes to the vacuole via the microtubule cytoskeleton; Belongs to the ATG8 family. (120 aa)*
COQ5	*2-methoxy-6-polyprenyl-1,4-benzoquinol methylase, mitochondrial; Methyltransferase required for the conversion of 2- polyprenyl-6-methoxy-1,4-benzoquinol (DDMQH2) to 2-polyprenyl-3-methyl-6-methoxy-1,4-benzoquinol (DMQH2). (294 aa)*
H0813E03.4	*Probable aldo-keto reductase 3; Belongs to the aldo/keto reductase family. (355 aa)*
PSK2	*Phytosulfokine-alpha; Promotes plant cell differentiation, organogenesis and somatic embryogenesis as well as cell proliferation; Belongs to the phytosulfokine family. (119 aa)*
OsI_06084	*Protein transport protein sec61 gamma subunit. (69 aa)*
PIR7B	*Esterase PIR7B; Belongs to the AB hydrolase superfamily. (268 aa)*
OsI_03978	*Katanin p80 WD40 repeat-containing subunit B1 homolog; May participate in a complex which severs microtubules in an ATP-dependent manner. Microtubule severing may promote rapid reorganization of cellular microtubule arrays; Belongs to the WD repeat KATNB1 family. (843 aa)*

OsI_00923	*Non-specific lipid-transfer protein; Plant non-specific lipid-transfer proteins transfer phospholipids as well as galactolipids across membranes. May play a role in wax or cutin deposition in the cell walls of expanding epidermal cells and certain secretory tissues. (123 aa)*
infA	*Translation initiation factor IF-1, chloroplastic; One of the essential components for the initiation of protein synthesis. Stabilizes the binding of IF-2 and IF-3 on the 30S subunit to which N-formylmethionyl-tRNA(fMet) subsequently binds. Helps modulate mRNA selection, yielding the 30S pre-initiation complex (PIC). Upon addition of the 50S ribosomal subunit IF-1, IF-2 and IF-3 are released leaving the mature 70S translation initiation complex. (107 aa)*
OsI_16030	*Protein phosphatase methylesterase 1; Demethylates proteins that have been reversibly carboxymethylated. (320 aa)*
GH3.8	*Probable indole-3-acetic acid-amido synthetase GH3.8; May catalyze the synthesis of indole-3-acetic acid (IAA)- amino acid conjugates, providing a mechanism for the plant to cope with the presence of excess auxin; Belongs to the IAA-amido conjugating enzyme family. (605 aa)*
LAC19	*Putative laccase-19; Lignin degradation and detoxification of lignin-derived products; Belongs to the multicopper oxidase family. (590 aa)*
OsI_07776	*Aldose 1-epimerase; Converts alpha-aldose to the beta-anomer. (378 aa)*
OsI_07760	*Ubiquitin carboxyl-terminal hydrolase; Recognizes and hydrolyzes the peptide bond at the C-terminal Gly of ubiquitin. Involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. Belongs to the peptidase C19 family. (415 aa)*
OsI_001009	*Salt stress root protein RS1. (204 aa)*
OsI_01675	*PEROXIDASE_4 domain-containing protein; Belongs to the peroxidase family. (168 aa)*
LTP-2	*Non-specific lipid-transfer protein 2; Transfer lipids across membranes. May play a role in plant defense or in the biosynthesis of cuticle layers. (96 aa)*
OsI_36717	*Barwin domain-containing protein. (146 aa)*
RR6	*Two-component response regulator ORR6; Functions as response regulator involved in His-to-Asp phosphorelay signal transduction system. Phosphorylation of the Asp residue in the receiver domain activates the ability of the protein to promote the transcription of target genes. Type-A response regulators*

	seem to act as negative regulators of the cytokinin signaling. Belongs to the ARR family. Type-A subfamily. (170 aa)
CIN1	*Beta-fructofuranosidase, insoluble isoenzyme 1; May play a role in sucrose partitioning during seed development and in stress response. (577 aa)*
COQ3	*Ubiquinone biosynthesis O-methyltransferase, mitochondrial; O-methyltransferase that catalyzes the 2 O-methylation steps in the ubiquinone biosynthetic pathway; Belongs to the class I-like SAM-binding methyltransferase superfamily. UbiG/COQ3 family. (327 aa)*
NBS1	*Nijmegen breakage syndrome 1 protein; Component of the MRE11-RAD50-NBN complex (MRN complex) which plays a critical role in the cellular response to DNA damage and the maintenance of chromosome integrity. The complex may be involved in double-strand break (DSB) repair, DNA recombination, maintenance of telomere integrity, and cell cycle checkpoint control. Functions also in the very early stages of meiosis. (560 aa)*
OsI_22455	*Acylphosphatase. (140 aa)*
SPL11	*Protein spotted leaf 11; Defense related protein that negatively regulates programmed cell death. In vitro, possesses E3 ubiquitin ligase activity. (694 aa)*
OsI_04666	*Stearoyl-[acyl-carrier-protein] 9-desaturase 1, chloroplastic; Converts stearoyl-ACP to oleoyl-ACP by introduction of a cis double bond between carbons 9 and 10 of the acyl chain. Belongs to the fatty acid desaturase type 2 family. (381 aa)*
YAB5	*Protein YABBY 5; May be involved in leaf cell growth and differentiation, rather than abaxial cell fate determination; Belongs to the YABBY family. (266 aa)*
OsI_03981	*SAND domain-containing protein. (235 aa)*
OsI_39190	*Cysteine synthase; Belongs to the cysteine synthase/cystathionine beta- synthase family. (408 aa)*
SWEET 6B	*Bidirectional sugar transporter SWEET6b; Mediates both low-affinity uptake and efflux of sugar across the plasma membrane. (254 aa)*
D2HGDH	*Probable D-2-hydroxyglutarate dehydrogenase, mitochondrial; Catalyzes the oxidation of D-2-hydroxyglutarate to alpha- ketoglutarate. (559 aa)*

HOX29	*Homeobox-leucine zipper protein HOX29; Probable transcription factor. (861 aa)*
OsI_05585	*Remorin_C domain-containing protein. (510 aa)*
OsI_07658	*AIG1-type G domain-containing protein. (330 aa)*
SPS1	*Probable sucrose-phosphate synthase 1; Plays a role in photosynthetic sucrose synthesis by catalyzing the rate-limiting step of sucrose biosynthesis from UDP- glucose and fructose- 6-phosphate. Involved in the regulation of carbon partitioning in the leaves of plants. May regulate the synthesis of sucrose and therefore play a major role as a limiting factor in the export of photoassimilates out of the leaf. Plays a role for sucrose availability that is essential for plant growth and fiber elongation (By similarity); Belongs to the glycosyltransferase 1 family. (1084 aa)*
OsI_19332	*Lactoylglutathione lyase; Catalyzes the conversion of hemimercaptal, formed from methylglyoxal and glutathione, to S-lactoylglutathione. (237 aa)*
OsI_35238	*Cytochrome b-c1 complex subunit 6; This is a component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is part of the mitochondrial respiratory chain. This protein may mediate formation of the complex between cytochromes c and c1. Belongs to the UQCRH/QCR6 family. (88 aa)*
OsI_03088	*Phospholipase A1-II 3; Acylhydrolase that catalyzes the hydrolysis of phospholipids at the sn-1 position; Belongs to the AB hydrolase superfamily. Lipase family. (420 aa)*
MIF1	*Mini zinc finger protein 1; Inhibits zinc finger homeodomain (ZHD) transcription factors, by interacting with them to prevent both their nuclear localization and their DNA-binding properties. (105 aa)*
PRR37	*Two-component response regulator-like PRR37; Controls photoperiodic flowering response. Seems to be one of the component of the circadian clock. Expression of several members of the ARR-like family is controlled by circadian rhythm. The particular coordinated sequential expression of PRR73, PRR37, PRR95, PRR59 and PPR1 result to circadian waves that may be at the basis of the endogenous circadian clock (By similarity). (742 aa)*