Identification of SNPs in RNA-seq data of two cultivars of *Glycine max* (soybean) differing in drought resistance

Ramon Oliveira Vidal\(^1\)\(^2\), Leandro Costa do Nascimento\(^1\), Jorge Maurício Costa Mondego\(^3\), Gonçalo Amarante Guimarães Pereira\(^1\) and Marcelo Falsarella Carazzolle\(^1\)\(^4\)

\(^1\)Laboratório de Genômica e Expressão, Universidade Estadual de Campinas, Campinas, SP, Brazil.
\(^2\)Laboratório Nacional de Biociências/Associação Brasileira de Tecnologia de Luz Sincrotron, Campinas, SP, Brazil.
\(^3\)Centro de Pesquisa e Desenvolvimento em Recursos Genéticos Vegetais, Instituto Agronômico de Campinas, Campinas, SP, Brazil.
\(^4\)Centro Nacional de Processamento de Alto Desempenho, Universidade Estadual de Campinas, Campinas, SP, Brazil.

**Abstract**

The legume *Glycine max* (soybean) plays an important economic role in the international commodities market, with a world production of almost 260 million tons for the 2009/2010 harvest. The increase in drought events in the last decade has caused production losses in recent harvests. This fact compels us to understand the drought tolerance mechanisms in soybean, taking into account its variability among commercial and developing cultivars. In order to identify single nucleotide polymorphisms (SNPs) in genes up-regulated during drought stress, we evaluated suppression subtractive libraries (SSH) from two contrasting cultivars upon water deprivation: sensitive (BR 16) and tolerant (Embrapa 48). A total of 2,222 soybean genes were up-regulated in both cultivars. Our method identified more than 6,000 SNPs in tolerant and sensitive Brazilian cultivars in those drought stress related genes. Among these SNPs, 165 (in 127 genes) are positioned at soybean chromosome ends, including transcription factors (MYB, WRKY) related to tolerance to abiotic stress.

**Key words:** single nucleotide polymorphisms, deep sequencing, drought resistance.

**Introduction**

Soybean (*Glycine max*) is a legume crop that plays an important economic role in the international market, with a world production of almost 260 million tons for the 2009/2010 harvest. Brazil ranks as the world’s second largest producer and exporter, with about 25% of the production.

Soybean production is influenced by weather oscillations, especially long periods of drought. In the Brazilian soybean culture, the frequency of drought events has increased in recent decades, probably associated with the weather changes in the world (Stokstad, 2004; Schiermeier, 2006). For example, in the states of the south of Brazil, responsible for 40% of domestic production, losses have been as much as 25% of production in recent harvests. During this time some drought tolerant and sensitive cultivars were isolated from these regions. An understanding of the molecular mechanisms governing such contrasting phenotypes upon water deprivation could provide insights for the creation of new cultivars and help in assisted selection.

Single nucleotide polymorphisms (SNPs) are single base differences between DNA sequences of individuals or lines. They can be assayed and exploited as high-throughput molecular markers. SNP markers have the potential for use in association genetics approaches (Cardon and Bell, 2001; Flint-Garcia et al., 2003). The recent availability of high-throughput DNA sequencing data has enabled studies based on highly informative SNPs. The evaluation of SNPs in large EST sequence data sets from agricultural crops has been employed to generate high-density genetic maps and identify variable genomic regions (Du et al., 2003; Choi et al., 2007; Novaes et al., 2009; Pindo et al., 2008; Duran et al., 2009). The scalability and availability of SNPs in highly automated genotyping assays has made this molecular marker the first choice in genetic linkage and association studies in a variety of species.

High quality reference genome sequences and resources used to perform low coverage resequencing by novel sequencing technologies such as 454 Life Sciences...
(Barbazuk et al., 2007), Illumina Solexa (Van Tassell et al., 2008), and SOLiD (Melum et al., 2010) on different individuals, cultivars or even species are prerequisites for the traditional method of whole genome SNP discovery. In this context, genomic sequences of different individuals are aligned to a reference genome and nucleotide variation is detected.

This study used high-throughput mRNA sequencing (RNA-seq) reads derived from suppression subtractive libraries (SSH) to identify SNPs in up-regulated genes from tolerant (Embrapa 48) and sensitive (BR 16) cultivars submitted to drought stress. Such SNPs can be useful for assisted selection of soybean varieties with higher drought tolerance.

Material and Methods

Construction of cDNA libraries and sequencing

Soybean genotype Embrapa 48 (tolerant) and BR 16 (sensitive) were analyzed under two experimental conditions: drought stress and normal irrigation (for further details see Rodrigues et al., 2012, this issue). Leaves and roots from stressed and control plants were collected at three different times (25-50, 75-100, 125-150 min). RNA was isolated from tissue samples and used to construct suppression subtractive hybridization (SSH) cDNA libraries (Rodrigues et al., 2012). These three libraries, enriched in genes up-regulated during drought stress, were sequenced using Illumina/Solexa technology. The reads corresponding to genes enriched in such libraries were used in SNP mapping (see below).

Gene identification from RNA-seq data

The reference transcriptome assembly was constructed using 1,276,813 soybean ESTs available at NCBI. The bdtrimmer software (Baudet and Dias, 2007) was used to exclude ribosomal, vector, low quality and short (less than 100 bp) sequences (using default parameters). The remaining sequences were assembled with the CAP3 program (Huang and Madan, 1999) using default parameters, generating 60,747 unigenes (30,809 contigs and 29,938 singlets) (Nascimento et al., 2012, this issue).

The RNA-seq reads from SSH libraries from tolerant and sensitive cultivars were submitted to quality filtration considering bases greater than Q20 and merged in one single file. The information about the cultivars was included on the read ID to facilitate SNP genotyping. Bowtie software (Langmead et al., 2009), considering default parameters (maximum of 2 mismatches), was used to map the reads against the reference transcriptome and the output file was saved in SAM format (Figure 1).

SNP detection

SNP detection was performed with the SAMtools pileup program (Li et al., 2009), which found the variations in the SAM file, followed by VarScan (Koboldt et al., 2009), which identifies and filters variants based on read counts, base quality and allele frequency (Figure 1).

We developed a Perl script to compare the filtered SNP lists generated by the pipeline described above for the two datasets. Putative SNPs were tagged if the reads involved were mapped unambiguously on the reference transcriptome and the minor allele appeared at least 10 times. The SNPs were discarded if the depth was less than 20 and the frequency of one allele was more than 80%. For each candidate SNP, the algorithm accessed the reads over that position in the SAM file, and observed if all variations occurred only between the two cultivars.

The unigenes identified through mapping of RNA-seq reads that presented polymorphic sites between both cultivars were annotated and grouped into GO terms using blast2go software (Conesa and Götz, 2008). In order to identify the chromosome position of polymorphic genes, the unigenes were mapped to the soybean genome assembly (Schmutz et al., 2010) using Exonerate software (Slater and Birney, 2005).

Results and Discussion

Alignment of RNA-seq reads to reference transcriptome

The soybean data available for bioinformatics analysis comprise the reference genome (Williams 82 cultivar; Schmutz et al., 2010), gene models and the EST assembly described in Material and Methods. We chose this assembly as reference to the mapping in order to avoid (1) splicing alignment problems and (2) the absence of the untranslated regions.
After evaluating the SSH libraries, a total of 12,285,871 reads of 45 bp and 30,326,963 reads of 76 bp were obtained for sensitive and tolerant cultivars, respectively. For the sensitive cultivar, 6,317,010 reads were aligned to 7,039 contigs and 2,659 singlets, providing an average depth of coverage of 651 reads by reference sequence. For the tolerant cultivar, 6,120,258 reads were aligned to 15,667 contigs and 6,284 singlets, providing an average depth of coverage of 279 reads by reference sequence. We searched for SNPs in contigs that mapped into both cultivars. After mapping, 7,897 contigs have reads assigned to both cultivars, and were used in SNP identification.

Polymorphism detection

The identification of sequence polymorphisms was performed using SAMtools pileup and VarScan software. We identified a total of 44,510 variations (in 11,000 unigenes) that come from SNPs within each cultivar, SNPs between the cultivars and reference assembly, and SNPs between the cultivars.

To identify inter-cultivar SNPs in the soybean reference sequences and to identify the SNPs between both cultivars (i.e. non-allelic SNPs), we used an in-house SNP filter developed to identify the positions of robust candidate sequence polymorphisms relative to the aligned Solexa reads from each cultivar. We applied another filter requiring at least 10 reads on each cultivar and a maximum of 80% difference between the major and minor allele.

The 6,698 putative polymorphisms were identified in over 2,222 transcripts in tolerant and sensitive cultivars (~3 SNPs/gene). The majority of these polymorphisms represent allelic SNPs (intra-cultivar SNPs) and are not useful as molecular markers for soybean breeding. Nevertheless, we found 165 putative SNPs between tolerant and sensitive cultivars in a total of 127 unigenes (Supplementary Material Table S1). Figure 2 summarizes the GO annotation. As expected, the GO analysis of these 127 genes revealed that many are related to stress response and other ontologies possibly related to stress. Among such genes are a series of signal transducers (calmodulin, ankyrin, GTP binding proteins) and transcription factors (i.e., WRKY, MYB, Zinc Finger, Homeodomain-ZIP), all of which were cataloged in a set of soybean transcription factors (TFs) database focusing on abiotic stress responsive transcription factors (Mochida et al., 2009). Among these soybean TFs, the WRKY and MYB genes were experimentally evaluated. Zhou et al. (2008) verified that soybean WRKY genes provide tolerance to abiotic stresses in transgenic Arabidopsis plants. Soybean MYB genes were considered part of the stress tolerance apparatus based on salt and freezing stress assays in transgenic plants (Liao et al., 2008). The presence of variability in genes that likely coordinate the first steps of stress signaling, thus controlling a series of protective proteins against drought effects, denote such genes as possible markers for assisted selection. We consider all of these polymorphisms as strong candidates for molecular markers for selective breeding.

Mapping SNPs in soybean chromosomes

Putative SNPs detected in uniquely mapped reference sequence unigenes were plotted along the soybean chromosomes. Alignment with the soybean genome showed that the genes with these identified putative SNPs were not distributed uniformly across the genome. Instead, they are more prominent at the chromosome ends (Figure 3). Wu et al. (2010) detected that many SNPs were clustered in gene-rich, high-recombination euchromatic regions in soybean chromosomes. This positional trait of SNPs may be a consequence of intense recombination/mutation events that are crucial for increased variability in autogamous species such as soybean.

Conclusion

The identification of SNPs in contrasting cultivars for drought stress may be useful to breeders in Marker Assisted Selection (MAS) or even in Genome Wide Selection (GWAS) to identify SNPs contrasting drought resistant cultivars. These SNPs can be used to develop molecular markers for assisted selection.
(GWS). They can be added to the upcoming markers derived from high-throughput gene sequencing. We believe that the presence of SNPS in transcription factors is outstanding. Such proteins could be controlling a series of genes responsive to stress, making them good candidates for transgenesis and as starting points to understand drought tolerance mechanisms in soybean.

Acknowledgments

The authors would like to thank all researchers of the Brazilian Soybean Genome Consortium (GENOSOJA) involved in the generation of data used in this paper, and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil) for financial support.

References

Barbazuk WB, Emrich SJ, Chen HD, Li L and Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910-918.

Baudet C and Dias Z (2007) New EST trimming procedure applied to SUCEST sequences. In: Proceedings of the Second Brazilian Conference on Advances in Bioinformatics and Computational Biology. Springer-Verlag, Berlin, pp 57-68.

Cardon LR and Bell JI (2001) Association study designs for complex diseases. Nat Rev Genet 2:91-99.

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS et al. (2007) A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685-96.

Conesa A and Götz S. (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:e619832.

Du CF, Liu HM, Li RZ, Li PB and Ren ZQ (2003) Application of single nucleotide polymorphism in crop genetics and improvement. Yi Chuan 25:735-739.

Duran C, Appleby N, Clark T, Wood D, Imelfort M, Batley J and Edwards D (2009) AutoSNPdb: An annotated single nucleotide polymorphism database for crop plants. Nucleic Acids Res 37:D951-D953.

Flint-Garcia SA, Thornsberry JM and Buckler ES (2003) Structure of linkage disequilibria in plants. Annu Rev Plant Biol 54:357-374.

Huang X and Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868-877.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK and Ding L (2009) VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25:2283-2285.

Langmead B, Trapnell C, Pop M and Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078-2079.

Liao Y, Zou HF, Wang HW, Zhang WK, Ma B, Zhang JS and Chen SY (2008) Soybean GmMYB76, GmMYB92, and GmMYB177 genes confer stress tolerance in transgenic Arabidopsis plants. Cell Res 18:1047-1060.

Melum E, May S, Schillhabel MB, Thomsen I, Karlsen TH, Rosenstiel P, Schreiber S and Franke A (2010) SNP discovery performance of two second-generation sequencing platforms in the NOD2 gene region. Hum Mutat 31:875-885.

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K and Tran LS (2009) In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. DNA Res 16:353-369.

Nascimento LC, Costa GGL, Binneck E, Pereira GAG and Carazolle MF (2012) A web-based bioinformatics interface applied to Genosoja Project: Databases and pipelines. Genet Mol Biol 35(suppl 1):203-211.

Novaes E, Drost DR, Farmerie WG, Pappas Jr GJ, Grattapaglia D, Sederoff RR and Kirst M (2009) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. BMC Genomics 9:e132.

Pindo M, Vezzulli S, Coppola G, Cartwright DA, Zharkikh A, Velasco R and Troggio M (2008) SNP high-throughput screening in grapevine using the SNPlex genotyping system. BMC Plant Biol 8:e18.

Schiermeier Q (2006) The costs of global warming. Nature 439:374-375.

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

Slater G and Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:e31.

Stokstad E (2004) Global survey documents puzzling decline of amphibians. Science 306:391.

Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC and Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Meth 5:247-252.

Wu X, Ren C, Joshi T, Vuong T, Xu D and Nguyen HT (2010) SNP discovery by high-throughput sequencing in soybean. BMC Genomics 11:469.

Zhou QY, Tian AG, Zou HF, Xie ZM, Lei G, Huang J, Wang CM, Wang HW, Zhang JS and Chen SY (2008) Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54 confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. Plant Biotechnol J 6:486-503.

Internet Resources

Soybean transcription factors database, http://soybeantfdb.psc.riken.jp (October 10, 2011).

Supplementary Material

The following online material is available for this article:

Table S1 - All genes with intra-cultivar SNP annotations.

This material is available as part of the online article from http://www.scielo.br/gmb.
Table S1. All genes with SNPs intra-cultivars annotations.

| Seq. Name   | Seq. Description            | Seq. Length | min. eValue | GOs                                                                 | Enzyme Codes |
|-------------|----------------------------|-------------|-------------|----------------------------------------------------------------------|--------------|
| Contig11685 | 3-dehydroquinate synthase   | 1152        | 3.92E-166   | P:biosynthetic process; P:cellular amino acid and derivative metabolic process; F:catalytic activity; C:plastid | EC:4.2.3.4   |
| Contig123   | 40s ribosomal protein s11   | 819         | 3.31E-77    | C:ribosome; F:structural molecule activity; C:plastid; P:translation; F:RNA binding | EC:3.6.5.3   |
| Contig11958 | 40s ribosomal protein s15a  | 740         | 2.03E-64    | F:structural molecule activity; C:plasma membrane; C:cell wall; C:membrane; C:vacuole; C:ribosome; C:cytosol; P:translation; C:mitochondrion | EC:3.6.5.3   |
| Contig23481 | 40s ribosomal protein s29   | 613         | 6.92E-29    | C:ribosome; F:structural molecule activity; P:translation           | EC:3.6.5.3   |
| SJ07-E1-S10-274-C02-UC.F | 60s ribosomal protein l19 | 604         | 8.15E-11    | C:mitochondrion                                                      |              |
| Contig5872  | 60s ribosomal protein l44   | 795         | 8.09E-57    | C:ribosome; F:structural molecule activity; P:translation; C:mitochondrion | EC:3.6.5.3   |
| Contig22380 | acyl acp-thioesterase       | 686         | 6.17E-46    | P:biosynthetic process; P:cellular process; P:lipid metabolic process; F:hydrolase activity; F:transferase activity; C:plastid | EC:3.1.2.14  |
| SJ01-E1-L06-022-G03-UC.F | acyl- oxidase  | 555         | 1.42E-98    | F:catalytic activity; F:transporter activity; C:peroxisome; P:response to biotic stimulus; P:response to external stimulus; P:response to stress; P:biosynthetic process; P:cellular process; P:lipid metabolic process; P:catabolic process; P:multicellular organismal development; F:nucleotide binding; F:biological_process | EC:1.3.99.3; EC:1.9.3.1; EC:1.3.3.6 |
| Contig19038 | af149277_1peroxidase 1 precursor | 1085      | 4.58E-121   | F:binding; F:catalytic activity                                     |              |
| Contig14738 | akin gamma                  | 1877        | 0           | F:kinase activity; C:cell wall; P:metabolic process; P:cellular process |              |
| SJ17-E1-R04-027-H06-UC.F | al-induced protein  | 479         | 3.57E-51    | P:response to endogenous stimulus; C:plasma membrane; P:biological_process |              |
| Contig985   | allene oxide synthase       | 786         | 5.05E-104   | F:catalytic activity; F:binding; F:molecular_function               | EC:4.2.1.92  |
| Contig23356 | alpha tubulin 1             | 1759        | 0           | C:cytoskeleton; P:cellular component organization; P:cellular process; P:nucleobase, nucleoside, nucleotide and nucleic acid metabolic process; P:catabolic process; F:hydrolase activity; | EC:3.6.5.1; EC:3.6.5.2; EC:3.6.5.3; |
| Contig   | Gene Description                          | Length | K-value | Description                                                                 |
|---------|-------------------------------------------|--------|---------|-----------------------------------------------------------------------------|
| Contig27771 | alpha tubulin 1                          | 1661   | 0       | Structural molecule activity; Cytoplasm; Carbohydrate binding; Nucleotide binding |
| Contig1153 | aluminum-induced protein                  | 980    | 1.64E-111 | Response to endogenous stimulus; Plasma membrane; Biological process         |
| Contig25238 | ankyrin repeat-containing protein         | 1060   | 4.81E-59  | Membrane; Response to stress; Response to abiotic stimulus; Cell; Transport; Binding; Thylakoid; Nucleus; Cytosol; Lipid binding |
| Contig22377 | annexin 2                                | 875    | 7.41E-110 | Response to stress; Response to abiotic stimulus; Cell; Transport; Cellular process; Transporter activity |
| Contig3332 | aquaporin-like transmembrane channel protein | 1335   | 1.79E-152 | Membrane; Response to stress; Plasma membrane; Transport; Cellular process; Transporter activity |
| Contig1476 | at1g19180 t29m8_5                         | 846    | 2.06E-132 | Response to stress; Response to biotic stimulus; Multicellular organismal development; Response to external stimulus; Response to endogenous stimulus; Signal transduction; Protein binding; Flower development; Nucleus |
| Contig7639 | at1g47330 t3f24_2                         | 763    | 5.05E-106 | Cytoplasm; Plasma membrane                                                |
| Contig4713 | at5g18670 t1a4_50                        | 2035   | 0       | Carbohydrate metabolic process; Catabolic process; Hydrolase activity; Binding |
| Contig19095 | auxin efflux carrier component            | 1844   | 0       | Transport; Cellular process; Anatomical structure morphogenesis; Multicellular organismal development; Plasma membrane; Membrane; Post-embryonic development; Response to abiotic stimulus; Flower development; Reproduction; Protein binding; Transporter activity; Tropism; Embryonic development; Cytoplasm; Cell |
| Contig13059 | auxin response factor 9                  | 1160   | 5.28E-86 | Binding; Response to endogenous stimulus; Transcription                   |
| Contig4535 | beta-galactosidase                        | 263    | 4.16E-47 | Carbohydrate metabolic process; Cell wall; Plastid                         |
| Contig       | Protein Name                        | Length | P-value       | EC Number |
|-------------|-------------------------------------|--------|---------------|-----------|
| Contig16059 | calcium-dependent protein           | 1079   | 1.16E-100     | 2.7.11.17 |
| SJ10-E1-R05-029-D10-UC.F | calcium-dependent protein kinase | 578    | 1.74E-89      | 2.7.11.17 |
| Contig6964  | callose synthase 10                 | 1067   | 5.93E-174     | 2.4.1.34  |
| Contig2415  | calmodulin                          | 899    | 1.21E-78      | 1.3.1.74  |
| Contig22225 | calmodulin-binding transcription activator | | 764 | 1.18E-70 |
| Contig23083 | carbonic anhydrase                  | 873    | 3.59E-72      | 4.2.1.1   |
| Contig2904  | catalase                            | 1811   | 0             | 1.11.16   |
| Contig15505 | chloroplast protease                | 1222   | 3.38E-179     | 3.4.24.0  |
| Contig22316 | chromatin remodeling complex subunit | | 1092 | 3.29E-42 |
| Contig7885  | coatamer gamma                      | 994    | 2.35E-131     | -         |
| Contig3263  | cysteine proteinase inhibitor       | 803    | 6.87E-96      | -         |
| Contig   | Description                                      | Accession | FDR  | EC Numbers                                      |
|---------|--------------------------------------------------|-----------|------|------------------------------------------------|
| 10810  | cysteine synthase                                | 668       | 4.42E-46 | EC:2.5.1.47                                     |
| 3637   | double wrky type transfactor                     | 1556      | 4.24E-180 |
| 13855  | ferredoxin-nadp+ reductase                       | 1492      | 0     | EC:1.6.99.1; EC:1.18.1.2                        |
| 29868  | gcpe protein                                     | 2290      | 0     | EC:1.17.4.3                                     |
| 22053  | gdp dissociation inhibitor                       | 249       | 1.83E-31 | -                                                |
| 14301  | germin-like protein                              | 688       | 1.97E-116 | -                                                |
| 14559  | global transcription factor group                | 1876      | 1.51E-89 | -                                                |
| 19241  | glycerol 3-phosphate permease                    | 988       | 4.64E-154 | -                                                |
| 28897  | gtp-binding protein                              | 962       | 6.66E-118 | -                                                |
| 12976  | high affinity potassium transporter              | 537       | 3.85E-71 | -                                                |
| 11277  | isoflavone synthase 1                            | 1773      | 0     | -                                                |
| 29114  | light-harvesting complex ii protein lhcb3        | 1157      | 2.40E-147 | -                                                |
| 10765  | lipoxygenase                                     | 2939      | 0     | EC:1.13.11.1                                     |
| Contig   | Description                                      | Length | E-value  | Functions                                                                                                         | EC Numbers                              |
|---------|--------------------------------------------------|--------|----------|-------------------------------------------------------------------------------------------------------------------|-----------------------------------------|
| 11027   | Lipoxygenase                                     | 1103   | 6.35E-126| Cytoplasm; Biosynthetic process; Cellular process; Lipid metabolic process; Binding; Catalytic activity; Metabolic process | EC:1.13.11.1                            |
| 5611    | Lipoxygenase l-5                                | 2187   | 0        | Cytoplasm; Biosynthetic process; Cellular process; Lipid metabolic process; Binding; Catalytic activity; Metabolic process | EC:1.13.11.2                            |
| 1867    | lrr receptor protein                             | 1027   | 1.26E-117| Receptor activity; Cell wall; Plasma membrane; Protein modification process; Nucleotide binding; Membrane; Kinase activity; Signal transduction; Metabolic process; Catalytic activity | EC:2.7.11.0; EC:1.3.1.74               |
| 1074    | mee14 (maternal effect embryo arrest 14)         | 1209   | 3.18E-113| Reproduction; Post-embryonic development; Embryonic development; Response to stress; Response to biotic stimulus; Plastid |                                          |
| 11066   | myo-inisitol oxygenase                           | 1308   | 5.41E-154| Cytoplasm; Binding; Carbohydrate metabolic process; Cellular process; Catabolic process; Anatomical structure morphogenesis; Catalytic activity; Metabolic process | EC:1.13.99.1                            |
| 7836    | n chain improved model of plant photosystem i    | 965    | 1.67E-76 | Plastid; Thylakoid; Generation of precursor metabolites and energy; Photosynthesis; Membrane; Protein binding       |                                          |
| 26849   | Na+ H+ antiporter                                | 541    | 1.64E-85 | Transport activity; Transport; Cellular process; Biological_process; Membrane                                      |                                          |
| 22303   | Nucleobase ascorbate transporter                 | 1980   | 0        | Membrane; Vacuole; Cell wall; Transport; Cellular process; Transporter activity                                   |                                          |
| 21879   | Nucleosome chromatin assembly factor group       | 1839   | 0        | Intracellular; Anatomical structure morphogenesis; Cell differentiation; Multicellular organismal development; Cell cycle; Flower development; Regulation of gene expression, epigenetic; Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process; Nucleus; Reproduction; Post-embryonic development; Cellular component organization; Cellular process; Protein binding; Transferase activity; Plastid; Embryonic development | EC:2.3.1.48                            |
| 2668    | Omega-3 fatty acid desaturase                    | 1766   | 0        | Biosynthetic process; Cellular process; Lipid metabolic process; Catalytic activity; Membrane; Plastid; Metabolic process |                                          |
| 3844    | Organic anion transporter                        | 1609   | 1.88E-162| Transport activity; Membrane                                         |                                          |
| Contig     | Description                                           | Accession | Score     | Fscore | Functions                                                                 | EC          |
|-----------|-------------------------------------------------------|-----------|-----------|--------|---------------------------------------------------------------------------|-------------|
| 8110      | peptide methionine sulfoxide reductase                |           | 857       | 1.11E-102 | C:cytosol; F:catalytic activity; P:metabolic process; P:protein modification process; P:response to stress | 1.8.4.11    |
| 26213     | phosphate transporter                                  |           | 755       | 2.46E-81  | P:transport; C:plasma membrane; P:cellular process; F:transporter activity; C:membrane |             |
| 11742     | phosphoribulokinase precursor                         |           | 1871      | 0        | P:photosynthetic process; P:carbohydrate metabolic process; P:response to biotic stimulus; P:response to abiotic stimulus; C:plastid; F:nucleotide binding; F:protein binding; F:kinase activity; C:membrane; C:thylakoid; P:metabolic process; P:cellular process; C:extracellular region | 2.7.1.19    |
| 27945     | plasma membrane intrinsic protein                     |           | 1146      | 2.45E-144 | C:membrane; C:plasma membrane; P:transport; P:cellular process; F:transporter activity |             |
| SJ18-P1-S12-007-U51-UC.F | PREDICTED: hypothetical protein [Vitis vinifera] |           | 243       | 2.59E-09  | -                                                                          |             |
| 1857      | Ubiquinone biosynthesis                               |           | 1271      | 7.11E-95  | C:plasma membrane; C:endoplasmic reticulum                               |             |
| 3159      | WD-repeat protein putative phi-1-like phosphate-induced protein |           | 803       | 6.40E-110 | F:nucleotide binding; F:protein binding                                  |             |
| 3879      | putative phi-1-like phosphate-induced protein         |           | 1211      | 1.07E-177 | C:cell wall; P:response to endogenous stimulus                           |             |
| 5355      | NADH-ubiquinone oxidoreductase 24 kD subunit          |           | 1092      | 9.30E-146 | P:response to stress; F:nucleotide binding; F:binding; F:catalytic activity; C:membrane; C:mitochondrion; P:generation of precursor metabolites and energy | 1.6.5.3     |
| 5389      | kelch repeat-containing F-box family protein          |           | 761       | 9.51E-81  | F:molecular_function; P:biological_process                                |             |
| 6081      | -                                                     |           | 1027      | 4.22E-97  | P:multicellular organismal development; P:development; P:transcription; P:anatomical structure morphogenesis; F:DNA binding; F:transcription factor activity; P:flower development; C:nucleus |             |
| 6275      | class III HD-Zip protein 8                            |           | 2215      | 0        | -                                                                          |             |
| 7590      | heat shock protein binding protein                    |           | 692       | 1.34E-88  | F:binding; P:cellular homeostasis; P:response to stress; P:response to abiotic stimulus; P:response to endogenous stimulus; P:signal transduction |             |
| 8634      | Clp protease 2 proteolytic subunit                    |           | 369       | 1.11E-15  | C:plastid; F:hydrolase activity; C:thylakoid; P:protein metabolic process; P:catabolic process | 3.4.21.0    |
| Contig   | Description                                                                 | Start | End   | E-value   | Functions and Locations                                                                 |
|----------|-----------------------------------------------------------------------------|-------|-------|-----------|----------------------------------------------------------------------------------------|
| Contig9422 | -                                                                           | 945   | 1.62E-68 | F:molecular_function; P:biological_process                                            |
| Contig9779 | ubiquitin-protein ligase                                                    | 756   | 8.57E-66 | P:response to stress; P:response to abiotic stimulus; P:protein modification process; P:biological_process; P:metabolic process; F:catalytic activity; C:cytosol |
| Contig10455 | -                                                                           | 621   | 9.79E-63 | P:protein metabolic process; P:cellular process; F:protein binding; C:membrane; C:plastid; C:thylakoid |
| Contig10755 | aluminum induced protein with YGL and LRDR motifs epsilon-adaptin           | 832   | 5.78E-101 | C:cytosol; C:nucleus; C:plasma membrane                                               |
| Contig11725 | Phosphatidic acid phosphatase (PAP2) family protein aldehyde dehydrogenase | 1434  | 5.19E-153 | F:catalytic activity; C:cytoplasm; P:metabolic process; P:cellular process            |
| Contig12024 | Phosphatidic acid phosphatase (PAP2) family protein aldehyde dehydrogenase | 1434  | 5.19E-153 | F:catalytic activity; C:cytoplasm; P:metabolic process; P:cellular process            |
| Contig12307 | -                                                                           | 924   | 9.04E-133 | -                                                                                     |
| Contig14052 | -                                                                           | 712   | 4.18E-48 | C:mitochondrion; F:binding                                                            |
| Contig17122 | protein kinase                                                               | 829   | 1.68E-84 | F:kinase activity; P:protein modification process; F:nucleotide binding; C:plastid |
| Contig19813 | -                                                                           | 924   | 9.04E-133 | -                                                                                     |
| Contig20875 | phosphoribosylanthranilate transferase                                      | 1370  | 2.40E-160 | F:molecular_function; P:biological_process; C:cellular_component                      |
| Contig21517 | -                                                                           | 574   | 2.69E-34 | C:membrane; C:vacuole; C:mitochondrion                                               |
| Contig26353 | SWAP (Suppressor-of-White-APricot)/surp domain-containing protein / ubiquitin family protein | 932   | 7.21E-94 | F:RNA binding; P:nucleobase, nucleoside, nucleotide and nucleic acid metabolic process |
| Contig28841 | MYB transcription factor MYB122                                             | 590   | 1.65E-82 | F:DNA binding; P:transcription; C:nucleus                                            |
| Contig29674 | Avr9/Cf-9 rapidly elicited protein 20                                        | 465   | 2.52E-52 | P:protein metabolic process; P:cellular process; F:binding; F:protein binding; P:response to endogenous stimulus; C:nucleus; C:cytosol |
| Contig29829 | plastid transcriptionally                                                    | 1875  | 3.93E-130 | C:plastid; F:binding; P:metabolic process; F:catalytic activity;                     |
| Contig ID | Description | Accession | E-value | Process | Location | EC Number |
|----------|-------------|-----------|---------|---------|----------|-----------|
| SJ18-P1-S12-232-U15-UC.F | Active 16 aconitate hydratase, cytoplasmic / citrate hydro-lyase | | 5.73E-25 | P:response to stress; C:cytosol; F:binding; C:cell wall; C:mitochondrion; P:response to abiotic stimulus; C:membrane; C:vacuole; F:nucleotide binding; P:response to endogenous stimulus; F:catalytic activity; P:metabolic process; P:cellular process; C:plastid; P:biological_process | C:membrane; C:thylakoid | EC:4.2.1.3 |
| SJ18-P1-S12-286-X20-UC.F | Coatamer gamma subunit | | 7.49E-33 | F:protein binding; P:transport; P:cellular process; C:plastid; C:membrane; C:Golgi apparatus; F:structural molecule activity | - |
| SJ01-E1-F05-036-F07-UC.F | Cell division protein kinase | | 4.75E-56 | C:plasma membrane; P:cellular process; F:kinase activity | EC:2.7.11.0 |
| SJ07-E1-S10-199-A01-UC.F | Protein binding | | 9.81E-09 | F:zinc ion binding | - |
| Contig27355 | Protein binding | | 0 | F:protein kinase activity; F:ATP binding; F:transferase activity; P:protein amino acid phosphorylation; F:RNA polymerase II carboxy-terminal domain kinase activity; F:protein serine/threonine kinase activity; F:kinase activity; C:cytosol; F:protein binding | - |
| Contig832 | Protein kinase | | 2.72E-12 | F:protein kinase activity; F:ATP binding; F:transferase activity; P:protein amino acid phosphorylation; F:RNA polymerase II carboxy-terminal domain kinase activity; F:protein serine/threonine kinase activity; F:kinase activity; C:cytosol; F:protein binding | - |
| SJ08-E1-S11-008-F11-UC.F | Protein phosphatase 2c | | 8.79E-21 | P:protein modification process; F:binding; F:hydrolase activity; C:cell | - |
| Contig3754 | Pspbp2_tobame: full=oxygen-evolving enhancer protein 2-chloroplastic short=oe2 ame: full=23 kda subunit of oxygen evolving system of photosystem ii ame: full=23 kda thylakoid membrane protein ame: full=oec 23 kda subunit flags: precursor | | 3.93E-38 | C:membrane; F:binding; P:photosynthesis; C:plastid; C:thylakoid | - |
| Contig1650 | Receptor serine threonine | | 5.86E-86 | F:receptor activity; F:kinase activity; P:protein modification process; F:nucleotide binding | EC:2.7.10.2; EC:2.7.11.0 |
| Contig3679 | Receptor-like protein | | 7.74E-95 | C:cell; P:protein modification process; F:nucleotide binding | EC:2.7.11.0 |
Contig14231 ribbon protein 1178 5.87E-149 F:kinase activity; P:signal transduction; P:transduction; P:multicellular organismal development; F:binding; C:nucleus; F:structural molecule activity; P:anatomical structure morphogenesis; P:transport; P:cellular process; P:cellular component organization; C:plastid; C:vacuole; C:ribosome; C:cytosol; C:plasma membrane; F:RNA binding activity

Contig4137 ribosomal protein s14 842 5.06E-60 C:ribosome; F:structural molecule activity; P:translation

Contig11808 s-adenosylmethionine decarboxylase 1953 0 P:biosynthetic process; P:cellular amino acid and derivative metabolic process; F:catalytic activity

Contig24388 serine threonine protein phosphatase 2a regulatory subunit 846 1.41E-134 F:enzyme regulator activity; F:binding; C:plasma membrane; C:intracellular; C:cell wall; P:metabolic process; P:cellular process

Contig24411 serine-threonine protein plant-1 1761 2.90E-95 C:plastid

Contig8686 sin3 histone deacetylase complex 1236 2.85E-109 C:nucleus; P:transcription

Contig24008 small gtp-binding protein 1132 1.06E-99 F:nucleotide binding; C:plasma membrane; F:hydrolyase activity; P:signal transduction; P:transport

Contig5904 stem-specific protein 655 1.11E-50 C:cytosol; C:nucleus; C:plasma membrane

Contig14015 sucrose transport protein suc2 2027 0 P:pollination; P:cellular process; P:transport; P:response to biotic stimulus; C:vacuole; C:plasma membrane; F:transporter activity; C:membrane; P:reproduction; P:post-embryonic development; C:cytoplasm; P:translation; F:transcription regulator activity; P:cellular component organization; P:transcription; F:protein binding; F:translation factor activity, nucleic acid binding; C:nucleus; F:binding

Contig27892 transcription initiation factor iib 1279 4.12E-167 C:cytoplasm; P:translation; F:transcription regulator activity; P:cellular component organization; P:transcription; F:protein binding; F:translation factor activity, nucleic acid binding; C:nucleus; F:binding

Contig25975 transitional endoplasmic reticulum 259 6.70E-29 P:cell cycle; F:hydrolyase activity; F:nucleotide binding; C:plasma membrane

Contig9825 triosephosphate 1249 3.61E-160 P:reproduction; P:biosynthetic process; P:carbohydrate metabolism
| Contig | Gene Name | Length | Score  |
|--------|-----------|--------|--------|
| 14497  | Tubulin alpha-3 | 949    | 3.69E-121 |
| 416    | Ubiquitin-conjugating enzyme e2 i | 930    | 1.09E-93  |
| 28259  | UDP-D-Aploise UDP-D-Xylose synthase 1 | 1769   | 0        |
| SJ06-E1-S01-008-D01-UC.F | UDP-glucuronic acid decarboxylase 2 | 502    | 4.91E-85  |
| 22609  | Uncharacterized protein | 635    | 1.12E-24  |
| 21502  | Vacuolar ATP synthase subunit | 1215   | 1.06E-124 |
| 28734  | Vacuolar sorting protein 4b | 636    | 1.49E-101 |
| 21850  | Virion binding | 719    | 1.62E-47  |
| 23300  | WD40-repeats protein | 1309   | 8.76E-104 |
| 5209   | WRKY transcription | 870    | 1.77E-77  |

**Gene Ontology Terms:**
- P:photosynthesis; P:multicellular organismal development; P:mitochondrion; P:cellular process; P:lipid metabolic process; P:plastid; P:catabolic process; P:cellular component organization; C:thylakoid; F:catalytic activity; F:protein binding; C:extracellular region
- EC:3.6.5.1; EC:3.6.5.2; EC:3.6.5.3; EC:3.6.5.4
- C:cytosol; C:cell wall; C:cytoskeleton; P:cellular component organization; P:cellular process; P:nucleobase, nucleoside, nucleotide and nucleic acid metabolic process; P:catabolic process; P:response to abiotic stimulus; F:hydrolase activity; F:structural molecule activity; P:biological_process; F:nucleotide binding; C:plasma membrane
- EC:6.3.2.19
- F:nucleotide binding; C:cytoplasm; C:extracellular region; F:catalytic activity; P:biosynthetic process; P:nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
- EC:4.1.1.35
- C:membrane; C:Golgi apparatus; P:biosynthetic process; P:carbohydrate metabolic process; P:nucleobase, nucleoside, nucleotide and nucleic acid metabolic process; F:catalytic activity; P:cellular process; C:plasma membrane; C:vacuole; F:binding
- EC:4.1.1.35; EC:4.2.1.46
- F:molecular_function; P:biological_process; C:cellular_component
- EC:3.6.3.14
- P:transport; P:biosynthetic process; P:generation of precursor metabolites and energy; P:nucleobase, nucleoside, nucleotide and nucleic acid metabolic process; F:transporter activity; F:nucleotide binding; F:hydrolase activity; C:membrane
- EC:3.6.4.3
- C:endosome; P:anatomical structure morphogenesis; P:cellular component organization; P:cell differentiation; P:multicellular organismal development; P:cellular process; P:transport; F:nucleotide binding; F:hydrolase activity; C:nucleus
- EC:3.6.3.14
- F:binding
- C:intracellular; F:nucleotide binding
- P:transcription, DNA-dependent; F:sequence-specific DNA binding; P:regulation of transcription; F:transcription factor activity
| Contig   | Description                      | Length | Exp. | Functions                                                                 |
|---------|----------------------------------|--------|------|---------------------------------------------------------------------------|
| Contig18831 | wrky transcription               | 1909   | 0    | F:DNA binding; P:biological_process; F:protein binding; F:transcription factor activity; P:transcription; P:response to extracellular stimulus; P:cell communication; P:response to stress |
| Contig21742 | yth domain-containing zinc finger protein | 3028   | 0    | C:cytoplasm; C:nucleus; F:protein binding P:flower development; P:transcription; F:DNA binding; F:protein binding; C:nucleus |
| Contig23318 | zinc finger protein              | 1360   | 4.58E-119 |                                                                                     |