BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net **Hypothesis**

Volume 8(3)

Mining and gene ontology based annotation of SSR markers from expressed sequence tags of *Humulus lupulus*

Swati Singh[1]*, Sanchita Gupta[2], Ashutosh Mani[1], Anoop Chaturvedi[3]

[1]Center of Bioinformatics,University of Allahabad,Allahabad, India-211002; [2]Biotechnology Division, Central Institute of Medicinal & Aromatic Plants (Council of Scientific & Industrial Research), Post Office CIMAP, Lucknow India; [3]Department of Statistics,University of Allahabad,Allahabad, India-211002. Swati Singh – Email: swatinatural@gmail.com; *Corresponding author

Received January 12, 2012; Accepted January 17, 2012; Published February 03, 2012

Abstract

Humulus lupulus is commonly known as hops, a member of the family moraceae. Currently many projects are underway leading to the accumulation of voluminous genomic and expressed sequence tag sequences in public databases. The genetically characterized domains in these databases are limited due to non-availability of reliable molecular markers. The large data of EST sequences are available in hops. The simple sequence repeat markers extracted from EST data are used as molecular markers for genetic characterization, in the present study. 25,495 EST sequences were examined and assembled to get full-length sequences. Maximum frequency distribution was shown by mononucleotide SSR motifs i.e. 60.44% in contig and 62.16% in singleton where as minimum frequency are observed for hexanucleotide SSR in contig (0.09%) and pentanucleotide SSR in singletons (0.12%). Maximum trinucleotide motifs code for Glutamic acid (GAA) while AT/TA were the most frequent repeat of dinucleotide SSRs. Flanking primer pairs were designed in-silico for the SSR containing sequences. Functional categorization of SSRs containing sequences was done through gene ontology terms like biological process, cellular component and molecular function.

Keywords: *Humulus lupulus*, expressed sequence tag, molecular markers, simple sequence repeats.

Background:

Hop (*Humulus lupulus*) is a medicinal plant, but its major profitable use is in flavoring of beer. This plant is dioecious (2n = 2x = 20) with two heteromorphic sex chromosomes, X and Y **[1, 2, 3].** The reproductive mode affects many aspects of breeding and crop management such as male and female reproductive organs are dimorphic **[4],** families are highly heterozygous, phenotypically variable and breeding while cultivar developments are accomplished by single mating followed by selection and fixation of favorable genotypes by various means of asexual reproduction. SSR discovery in hops has relied on the hybridization-based screening of genomic libraries by means of artificial repetitive sequences and sequencing of isolated clones in order to build up locus-specific primers previously **[5].** However, high-throughput sequencing results engender information on thousands of expressed sequence tags (ESTs) **[6].** Microsatellite or Simple Sequence Repeats (SSR) or Short tandem repeats (STR) are 1-6 bp tandemly repeated motifs present in both coding and non-coding regions of prokaryotic and eukaryotic genome. The prominent frequencies of length polymorphism associated with microsatellites provide the basis for development of a marker system that has extensive application in genetic research including studies of genetic variation, linkage mapping, gene tagging and evolution **[7].** SSRs are used extensively as molecular markers because of their multiallelic nature, co-dominant inheritance and relative abundance. The foremost annoyance of SSRs as markers has been their time consuming development in laboratory. However, with fast-paced boost of nucleic acid in recent years, it became realistic to screen for microsatellites in database for numerous plant species. Variations in SSR regions originate mostly from errors during the replication process, frequently

DNA polymerase slippage. These errors generate base pair insertions or deletions respectively. **[8, 9].** We have mined SSRs from EST of *Humulus lupulus* to get the SSR polymorphism. ESTs are short and single pass sequences read from mRNA (cDNA) **[10].** This represents a snapshot of genes expressed in a given tissue. The use of EST or cDNA-based SSRs has been reported for several species including grape **[11]**, sugarcane **[12]**, durum wheat **[13]**, rye **[14]**, medicinal plant like basil **[15]** and Periwinkle **[16].** There are various SSR identification softwares such as MISA, SSR Finder, SSRIT, TRF, TROLL and sputnik. We used MISA **[17]** to identify SSR. Different types of SSRs and their percentage distributions were examined. The forward and reverse primer pairs were designed from the flanking ends of SSRs. The functional annotation of these SSR containing sequence was done.

Methodology:

Sequence data source

There are 25,495 ESTs of *Humulus lupulus* present in dbEST at NCBI. The retrieved sequences were isolated from different plant tissues like leaves, stem, root, etc. There is a chance of occurrence of redundancy in the EST sequences. In order to remove the redundancy, CAP3 assembler **[18, 19]** was used for sequence assembly. The resulting non-redundant sequences are contigs and singletons.

Microsatellite Identification

SSR were detected using MIcroSAtellite identification tool (MISA) written in the Perl scripting language **[17, 20].** This tool analyzes microsatellite repeats in FASTA formatted contig and singlet files. EST derived SSRs were considered to contain repeat motifs ranging in length from 1 to 6 bp. The minimum numbers of repeats were 10 for mononucleotides, 6 for dinucleotides and 5 for trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides. The analysis of SSRs was done based on their types (mono to hexanucleotides), number of repeats, percentage frequency of occurrences of each SSR motif and their distribution in the sequence.

Gene Ontology Classification

SSR-ESTs sequences with significant matches to protein entries of Swiss Prot-Uniprot KB database were functionally classified. Characterization of SSR-ESTs performed through gene ontology terms using Amigo **[21, 22].** The ontology classification was done in terms of their biological process, molecular function and cellular component. This characterization has been based on analyzed SSR repeats.

Marker development

Primer pairs for the SSR containing sequences were designed using BatchPrimer3 software for developing microsatellite markers **[23].** The microsatellites containing contigs and singletons were used for designing primers pairs. Forward and reverse primer pairs were designed for marker development.

Discussion:

ESTs are often represented by redundant cDNA sequences making them difficult to analyze effectively for SSRs. To eliminate the redundancy in sequences CAP3 program was used. The identification of overlapping sequences to generate contig and singleton sequences was done. 88.84% of ESTs forming contigs indicate that the majority of the ESTs have overlapping sequences while only 27.47% sequences were unique and have no corresponding overlapping sequences. The reduction of redundancy was found to be a sizeable proportion that has reduced 61.39% that means that number of ESTs prior to SSR analysis **Table 1 (see supplementary material).** The study of occurrence of different types of SSR repeats revealed that percentage distribution of mononucleotide SSRs is 60.44% in contigs and 62.16% in singletons followed by dinucleotide SSRs, 21.48% in contig and 19.9% in singleton **(Figure 1a, 2a).**

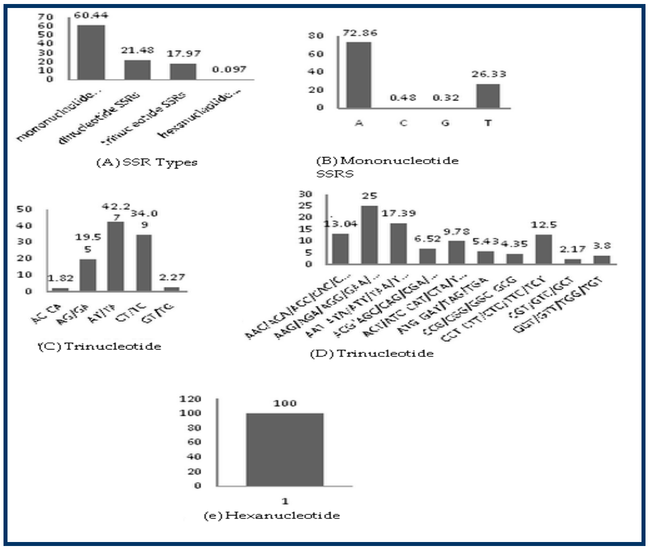

Figure 1: **(A)** Percentage distribution of different SSRs; **(B)** Percentage distribution of mononucleotide SSRs; **(C)** Percentage distribution of drinucleotide; **(D)** Percentage distribution of rinucleotide SSRs; **(E)** Percentage distribution of hexanucleotide SSRs

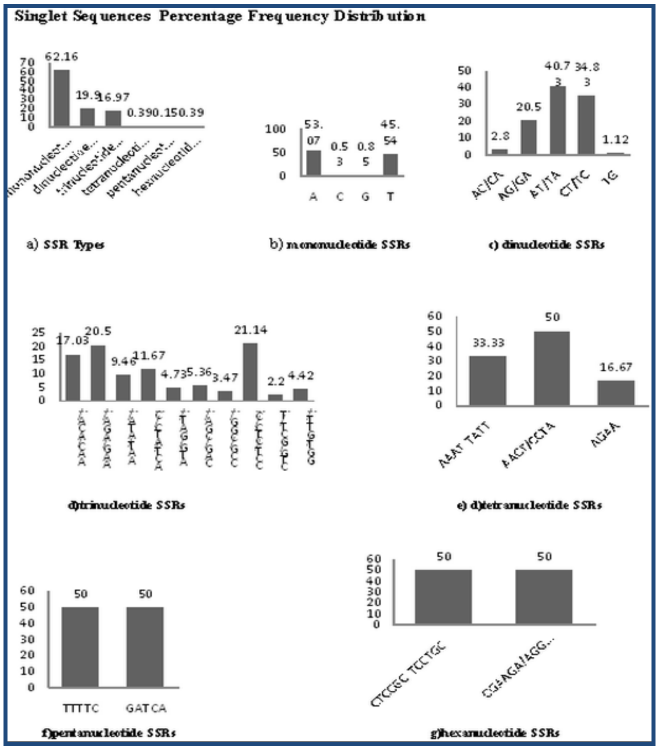

Figure 2: **(a)** Percentage distribution of different SSRs; **(b)** Percentage distribution of mononucleotide SSRs; **(c)** Percentage

distribution of drinucleotide SSRs; **(d)** Percentage distribution of trinucleotide SSRs; **(e)** Percentage distribution of tetranucleotide SSRs; **(f)** Percentage distribution of pentanucleotide SSRs; **(g)** Percentage distribution of hexanucleotide SSRs.

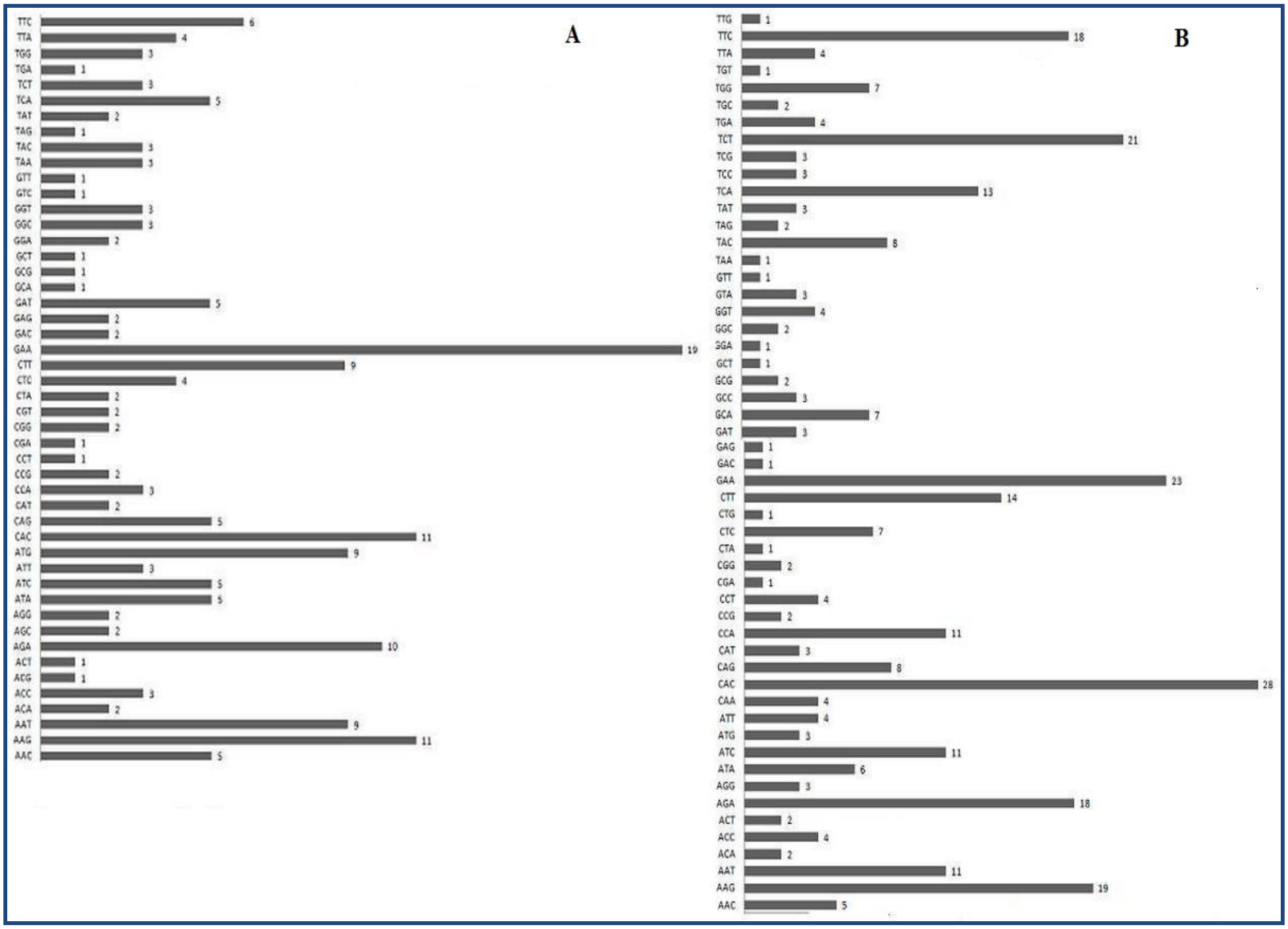

Figure 3: Frequency of distribution of contig triplet codon **(A)** Repetition of contig sequence codon; **(B)** Repetition of singlet sequence codon.

Among mononucleotide repeats, polyA/polyT repeats were predominant while polyC/polyG repeats were rare **(Figure 1b, 2b).** A-T repeat motifs are the most abundant type of SSRs in plants **[22].** All dinucleotide repeat combinations excluding homomeric dinucleotides can be grouped into classes namely, (AG)n, (AT)n, (AC)n, (GT)n and (TC)n. It is evident that AT/TA dinucleotide repeats were more frequent followed by CT/TC and AG/GA **(Figure 1c, 2c).**

Among 10 unique trinucleotide repeat classes, AAG/AGA/AGG/GAA/GAG/GGA (Contig-25% and singlet - 20.50) was the most frequent. The lowest frequency of trinucleotides was observed with TGC/TCG/GTC/CGT/CTG (contig-2.17% and singlet-2.2%) **(Figure 1d, 2d).** No tetranucleotide, pentanucleotide microsatellite was observed in contig sequences however these can be seen in singleton sequences Frequency of AACT/CCTA (50%) was maximum followed by AAAT/TATT and least frequency depicted in AGAA (**Figure 2e).** Only two pentanucleotides repeats were found there contribution was 50% **(Figure 2f**).In hexanucleotide SSR motif CCGCCT depicted in contig and in singlet CTCCGC/TCCTGC and CGAAGA/AGGAGC both are seen with frequency of 50% (Figure 2g).

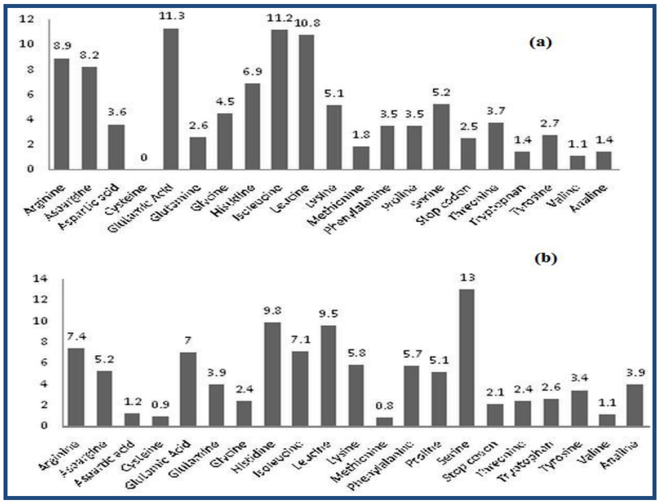

Figure 4(a): Percentage distribution of amino acids of contig sequences; **(b)** percentage distribution of amino acids of singlet sequences.

Codon Repetitions

The trinucleotide SSRs are triplet codon that code for a particular amino acid. It was observed that out of all triplet codons of contig sequences, GAA (encoding Glutamic Acid) repetitions are predominant followed by AAG (encoding Lysine) and CAC (Histidine), while in singleton sequences CAC is predominated followed by GAA. The triplet codon forms an open reading frame (ORF) translated to proteins **(Figure 3a, 3b)**

Amino acid distribution

The trinucleotide microsatellite codes for 21 types of amino acids, which includes stop codon. It was observed that out of all coded amino acid in contigs sequences Asparagine and Glutamic acid demonstrated the highest percentage of occurrence followed by Isoleucine and Leucine **(Figure 4a).** Where as in Singleton sequences Serine demonstrated the highest percentage followed by Histidine **(Figure 4b).**

In *Humulus lupulus* serine occurs in Serine/threonine-protein kinase and is involved in protein phosphorylation, positive regulation of DNA and Serine acetyltransferase is involved in cellular response to sulfate starvation. In contigs sequences valine showed the least occurrence and in singlets methionine and Cysteine were lowest in frequency.

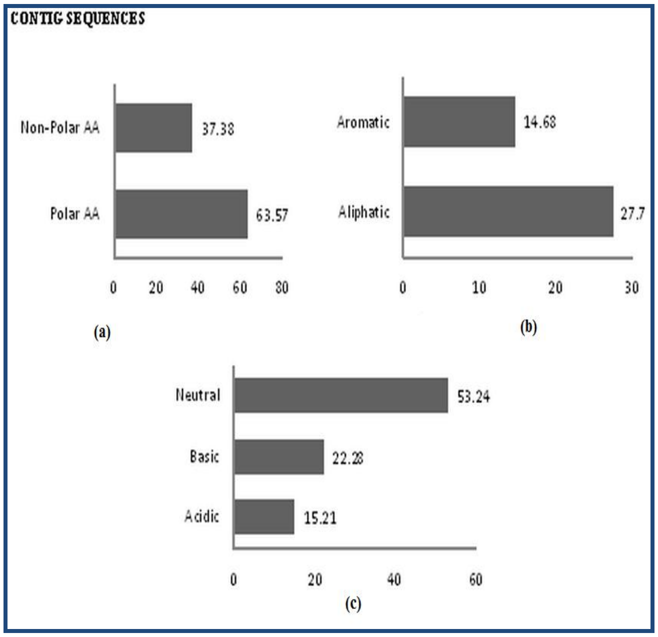

Figure 5: (a) Percentage frequency of polar & non-polar amino acids; **(b)** Percentage frequency of hydrop hilic & hydrophobic amino acids; **(C)** Percentage frequency of aromatic & aliphatic amino acids

The analysis of data revealed that the majority of amino acids were polar in nature, both in contig (63.57%) and singleton sequences (61.47%) and frequency of non polar contigs and singleton sequences are 37.38% ,36.87% respectively(Figure 5a,6a). Similarly, frequency of occurance of aliphatic amino acids in contig and singleton sequences 27.7% and 23.01% were more than aromatic amino acids in contig and singleton (14.68% and 20.8%) **(Figure 5b, 6b).** The distribution study of chemical nature of amino acids gives an insight that neutral amino acid occurred more frequently than with 53.24 in contigs sequences and 56.73 in singleton sequences in comparison with basic amino acids (contig-22.28%,singleton-23.01%) and acidic amino acids (contigs-15.21%, singleton-8.83%) **(Figure 5c, 6c).**

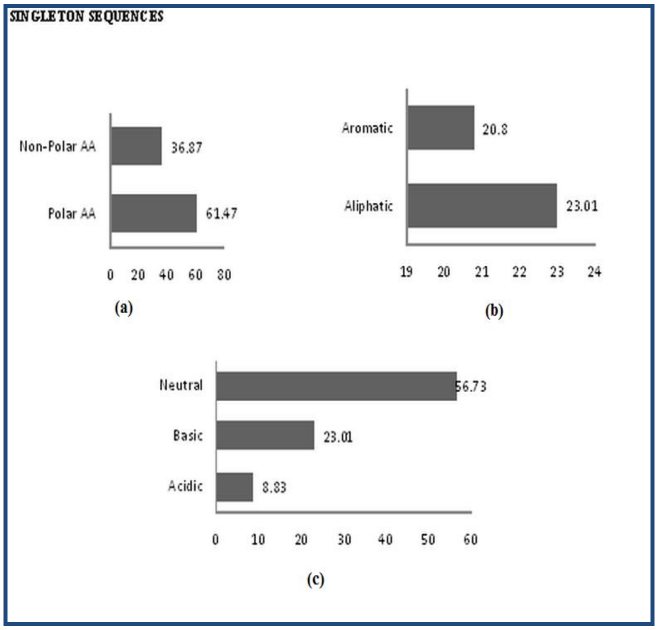

Figure 6: a) Percentage frequency of polar & non-polar amino acids; **b)** Percentage frequency of hydrophilic & hydrophilic amino acids; **c)** Percentage frequency of aromatic & aliphatic amino acids.

Gene Ontology Classification

Gene ontology based functional annotation of SSR- ESTs was performed through BLASTx using NCBI database. BLAST best hit were retained meeting the following criteria: E-value < 1e-4, and similarity >=70%. The most significant matches for the SSR-ESTs with unique SSR motif were considered. Out of 2651 unique SSR-ESTs, 835 had significant matches to proteins. Functional annotation of these SSR-ESTs was performed using Amigo. Gene ontology for the corresponding SSR's was determined on the basis of sequence, domain and motif similarity **Table 2 and 3 (see supplementary material).** A biological process is a series of events accompolished by one or more ordered assemblies of molecular functions. In a gamut of biological process corresponding to SSR-ESTs, the most frequent was Ribosome Biogenesis (32SSR-EST) followed by translation, protein folding, embroyo development in seed dormancy,ATP synthesis coupled proton transport, brassinosteroid biosynthetic process, regulation of transcription, DNA-dependent, protein phosphorylation, Photosynthesis.Molecular Function describes activities, such as catalytic or binding activities,that occur at the molecular level.In a gamut of molecular function, the most frequent was Structural constitue of ribosome (95 SSR-ESTs), protein binding, DNA binding, ATP binding, electron carrier activity, hydrogen ion transporting, ATP synthase activity, rotational mechanism, metal ion binding, lipid binding, sequence specific DNA binding. A cellular component is a component of a cell, but with the provision that it is a part of some larger object; this may be an anatomical structure or a gene product group. In a gamut of cellular components housing putative proteins,the most frequent was chloroplast (151SSR-ESTs) followed by cytosol (94 SSR-ESTs), Cell wall (66 SSR-ESTs), Cytoplasm, Nucleus, Plasma Membrane.

Primer designing

The primer designing have been done for PCR amplification of the desired microsatellites using BatchPrimer3.0 software. Out of 829 SSR-ESTs with significant matches, primers were designed for 268 SSR-EST contigs and 373 SSR-EST singletons. It was observed that forward and reverse primer pairs were obtained from mononucleotide SSRs (264) followed by trinucleotide (226). Hence a total of 641 SSR primer pairs were designed.

Conclusion:

Microsatellites serve for divergent roles in the field of plant genomics.EST database provide a valuable resource for the development of microsatellite markers, which are associated with transcribed genes. Simple Sequence Repeats are an important class of molecular markers for genomics and plant breeding applications due to their abundance, hyper variability, and suitability for high-throughput analysis, high polymorphism and transportability. Development of SSR markers from the EST database saves both cost and time, once sufficient amounts of EST sequence are available. Computational Approaches have been used here to mine ever increasing EST sequences in public databases. The publicly available collections of 25,495 expressed sequence tags (ESTs) from *Humulus lupulus* have been assembled and clustered using CAP3 assembly program. Assembly of EST sequences resulted in 9844 non-redundant EST sequences which were reported to have 2955 EST-SSRs. Among all the percentage frequency of mono-nucleotide SSRs is maximum and hexanucleotide has minimum frequency. Functional annotation of 2651 SSR-EST was performed and 835 have significant matches.829 SSR-ESTs were subject for primer designing which yielded a total of 641 primer set for *Humulus lupulus* that can be applied in studies of genetic variation, linkage mapping and comparative genomics.

Acknowledgements:

First Author is thankful to University Grants Commission, New Delhi for a research fellowship.

Refrences:

[1] Winge *et al. Compt Rend Trav Lab Carlsberg* 1923 **15**: 1
[2] Derenne *et al. Bull Inst Agro de Gembloux* 1954 **22**: 18
[3] Jacobsen *et al. Hereditas* 1957 **43**: 357
[4] Carter KW *et al. J Fla Med Assoc* 1983 **70**: 774
[5] Queller et al. Trends Ecol Evol 1993 8: 285
[6] Nagel *et al. Plant Cell* 2008 **20**: 186
[7] Kantety RV *et al. Plant Molecular Biology*. 2002 **48**: 501
[8] Ravi R & Iyer, *et al. The journal of biological chemistry* 2000 2174
[9] Victoria *et al. BMC Plant Biology*. 2011 **11**: 15
[10] Adams MD *et al*. Science 1991 **252**: 1651 [PMID: 2047873]
[11] Scott KD *et al. Theor Appl Genet* 2000 **100**: 723
[12] Cordeiro G *et al. Plant Sci* 2001 **160**: 1115
[13] Eujayl I *et al. Theor Appl Genet* 2002 **104**: 399 [PMID: 12582712]
[14] Hackauf B *et al. Plant Breed* 2002 **121**: 17
[15] Gupta S *et al. Plant Omics J*. 2010 **3**: 121
[16] Joshi RK *et al Bioinformation* 2011 **5**: 378
[17] Thiel T *et al. Theor Appl Genet.* 2003 **106**: 411 [PMID: 12589540]
[18] http://www.genome.clemson.edu/resources/online_tools/cap3
[19] Ramalakshmi OI *et al. Bioinformation* 2010 **5**: 240
[20] http://pgrc.ipk-gatersleben.de/misa/misa.html
[21] Carbon S *et al. Bioinformatics*. 2009 **25**: 288
[22] Sanchita *et al. Bioinformation* 2010 **5**: 113
[23] http://probes.pw.usda.gov/cgi bin/batchprimer3/batchprimer3.cgi

Edited by P Kangueane
Citation: Singh *et al*. Bioinformation 8(3): 114-122 (2012)

Supplementary material:

Table 1: Reduction in redundancy

Total no. of ESTs	No. of ESTs forming contigs (%)	No. of contigs	No. of singletons (%)	No. of assembled sequences	Reduction in Redundancy (%)
25,495	22652 (88.84%)	2843	7001 (27.46%)	9844	61.39%

Table 2: Gene ontology based functional annotation and classification of Contigs-SSRs of *Humulus Lupulus*.

Gene ontology (Biological process)	SSRESTs (numbers)	Gene ontology (Molecular function)	SSR-ESTS (number)	GeneOntology(Cellular Component)	SSRESTs (numbers)
Abscisic acid biosynthetic process	1	Acetate-coa ligase activity	1	Actin cytoskeleton	1
Abscisic acid mediated signalling pathway	2	Acetolactate synthase activity	1	Anchored to membrane	1
Acetyl-coa biosynthetic process from pyruvate	2	Acid phosphatase activity	1	Apoplast	18
Actin filament depolymerization	1	Acyl carrier activity	2	ATP binding cassette transporter	1
Actin nucleation	1	Acyl-coa dehydrogenase activity	1	Catalytic activity	1
Actin polymerization	1	Adenylylsulfate kinase activity	3	Cell plate	1
Activation of protein kinase C activity by GPCR	1	Alanine-glyoxylate transaminase activity	1	Cell surface	1
Aerobic respiration	3	Alcohol dehydrogenase(nad) activity	1	Cell wall	43
Aging	2	Allene-oxide cyclase activity	1	Central vacuole	2
Amino acid transport	2	Alpha-amylase activity	2	Chloroplast	73
Ammonia assimilation cycle	3	Alpha-galactosidase activity	1	Chloroplast envelope	2
Anther dehiscence	2	Amino acid transmembrane transporter	3	Chloroplast photosystem II	5
Anthocyanin metabolic process	1	Amino acid binding	2	Chloroplast thykaloid membrane	6
Anti-apoptosis	1	Ammonia-lysate activity	1	CUL4 RING ubiquitin ligase complex	1
Arginine process to glutamate	1	Anchored to membrane	1	Cytoplasm	12
Aromatic amino acid family biosynthetic process	2	Antioxidant activity	1	Cytoplasmic large ribosomal unit	6
Asparagine biosynthetic process	2	Aspargine synthase	1	Cytosol	56
ATP biosynthetic process	1	Atp binding	28	Cytosolic large ribosomal unit	10
ATP synthesis coupled proton transport	14	Atpase activity	9	Cytosolic small ribosomal subunit	8
Auxin mediated signalling	3	Beta-amylase activity	1	Early endosome membrane	1
Auxin polar transport	2	Beta-galactosidase activity	1	Endomembrane system	33
Base-excision repair	1	Binding	6	Endoplasmic reticulum	10
Brassinosteroid biosynthetic process	11	Caffeate o-methyltransferase activity	2	Extracellular region	1
Calcium mediated signalling	6	Calcium channel activity	1	Extrinisic to membrane	1
Carbohydrate biosynthetic process	1	Calcium ion binding	5	Extrinsic to vaculor membrane	1
Carbohydrate metabolic process	6	Calmodulin binding	5	Integral to membrane	3
Carboxylic acid metabolic process	1	Carbohydrate tranmembrane transporter	1	Intracellular	2
Carotenoid biosynthetic process	3	Carbon sulfur lyase activity	2	Intrinsic to endoplasmic reticulum	1
Cell death	1	Carboxypeptidase activity	1	Membrane	8
Cell proliferation	1	Catalytic activity	6	Mitochondrial respiratory complex	7
Cell redox homeostatis	7	Cellulose synthase activity	2	Mitochondrion	8
Cell wall loosening	1	Chalcone isomerase activity	2	Nuclear envelope	1
Cell wall modification	3	Chaperone binding	1	Nuclear speck	2
Cell wall organization	2	Chitin binding	1	Nucleic acid binding	1
Cell wall thickening	1	Chitinase activity	1	Nucleolus	3
Cellular copper ion homeostatis	1	Chloroallyl aldehyde dehydrogenase	2	Nucleosome	1
Cellular iron homeostatsis	1	Chlorophyll binding	1	Nucleus	15
Cellular metabolic process	1	Chromatin dna binding	1	Peroxisome	1
Cellular respiration	1	Cobalt ion binding	1	Plasma membrane	2
Cellular response to cold	2	Conjugate hydrolase activity	1	Plastid large ribosomal unit	1
Cellular response to ethylene stimulus	1	Copper chaperone activity	4	SCAR complex	1
Cellular response to fatty acid	1	Copper ion binding	9	Vacuole	3
Cellular response to nitrogen starvation	1	Cyclic nucleotide binding	2		
Cellular response to phosphate starvation	1	Cyclin-dependent protein kinase reg.	1		
Cellular response to selenium ion	3	Cystathionine beta-lyase activity	1		
Cellular response to water deprivation	2	Cysteine synthase activity	1		
Chlorophyll biosynthetic process	2	Cysteine-type endopeptidase activity	2		
Chloroplast thylakoid membrane	1	Cytochrome-b5 reductase activity	2		
Cinnamic acid biosynthetic process	3	Diacylglycerol kinase activity	1		
Copper ion transport	1	Dihydrolate reductase activity	1		
Cytokinesis	1	Dna binding	15		
De-etiolation	1	Dna photolyase activity	1		
Defense response	3	Dna-directed rna polymerase activity	2		
Defense response to bacterium	4	Double-stranded rna binding	4		
Development process	2	Electron carrier activity	5		
DNA duplex unwinding	1	Endopeptidase inhibitor activity	1		
DNA mediated transformation	1	Enzyme inhibitor activity	1		
DNA repair	1	Epoxide hydrolase activity	1		
Double-strand break repair	1	Fatty-acyl-coa binding	1		
Embroyo development in seed dormancy	14	Fatty-acyl-coa reductase activity	1		
Endocytic recycling	1	Fk506 binding	1		
Entrainment of circadian clock	1	Flavin adenine dinucleoide binding	1		
Ethylene biosynthetic process	1	Glutamate synthase(nadh) activity	2		
Extracellular transport	1	Glutathione transferase activity	1		
Fatty acid beta-oxidation	1	Glycerol-3-phosphate o-acytransferase	1		
Fatty acid biosynthetic process	4	Gtp binding	8		
Fatty acid metabolic process	1	Heme binding	6		
Fatty acid catabolic process	1	Histidine-trna ligase activity	1		

Flavonoid biosynthetic process	5
Gene silencing	2
Glycolipid transport	1
Heat acclimation	1
Heme biosynthetic process	1
Histidine biosynthetic process	1
Histone deacetylation	1
Histone H2B ubiquitnation	1
Hyperosmotic salinity response	5
Intracellular Protein Transport	1
Iron-sulfur cluster assembly	1
Lateral root development	1
Lignin biosynthetic process	3
Lipid storage	4
Lipid transport	5
L-methionine salvage from methylthioadenosine	2
L-phenylalanine biosynthetic process	2
Malate metabolic process	1
Male gamete generation	2
Male meiosis	2
Metabolic process	4
Methionine biosynthetic process	2
Microgametogenesis	1
Mrna transcription	1
Multicellular organismal development	3
Negative regulation of flower development	1
Negative regulation of transcription,DNA-dependent	1
Negative regulation of plant-type hypersensitive resp.	1
Nitrate assimilation	1
Nuclear mrna splicing,via spliceosome	3
Nucleosome assembly	3
Nuclear-transcribed mrna poly(A) tail shortening	1
Nucleocytoplasmic transport	1
Organ growth	1
Oxidation-reduction process	14
Pentose-phosphate shunt	1
Peptidyl-proline modification	1
Phenylalanine biosynthetic process	1
Phosphatidyl Glycerol biosynthetic process	1
Phosphatidylcholine biosynthetic process	3
Photosynthesis	9
Photosystem II stabilization	2
Pollen tube growth	1
Pollen development	1
Protein folding	15
Protein O- linked glycosylation	1
Protein phosphorylation	5
Protein targeting to vacuole	1
Protein transport	1
Protein ubiquitination involved in ubiquitin-dependent	1
Proteolysis	3
Proton transport	2
Regulation of transcription,DNA-dependent	2
Regulation of cell cycle	1
Regulation of programmed cell death	1
Respiratory electron transport chain	1
Response to auxin stimulus	1
Response to cadmium ion	4
Response to gibberellin stimulus	1
Response to high light intensity	1
Response to hypoxia	1
Response to karrikin	4
Response to metal ion	9
Response to nematode	2
Response to other organism	2
Response to salicylic acid stimulus	1
Response to salt stress	1
Response to stress	1
Response to water deprivation	4
Ribosome biogenesis	32
RNA metabolic process	1
Starch biosynthetic process	1
Sterol biosynthetic process	1
Transcription,DNA dependent	1
Translation	25
Transmembrane transport	2
Transport	3
Ubiquitin-dependent protein catabolic process	3
Vesicle-mediated transport	1
10-formyltetrahydrlformate biosynthetic process	1

Hydrogen ion transporting atp synthase	4
Hydrolase activity	1
Inorganic anion transporter activity	1
Intracellular cyclic nucleotide activated	1
Inositol monophosphatase activity	1
Isopentenyl-diphosphate delta-isomerase	1
Kinase activity	4
Lipase activity	1
Lipid binding	5
Lipoxygenase activity	1
L-malate dehydrogenase activity	1
Long-chain fatty acid-coa ligase activity	1
L-tyrosine:2-oxoglutarate amino-transf.	1
Manganese ion binding	6
Map kinase activity	1
Metal ion binding	5
Metalloendopeptidase activity	1
Mitochondrial respiratory chain complex	2
Mrna binding	1
Nucleic acid binding	4
Nucleocytoplasmic transporter activity	1
Nucleotide binding	1
Nucleotide-sugar transmembrane transp.	1
Oxidoreductase activity	6
Pantothenate kinase activity	1
Peptidyl-prolyl cis trans isomerase	1
Phenylalanine ammonia-lyase activity	1
Phosphatidylinositol-4,5-biophosphate bi	2
Phosphoglycerate dehydrogenase activity	1
Phosphoric diester hydrolase activity	1
Polygalacturonase activity	1
Polynucleotide 5'-hydroxyl-kinase	1
P-p-bond-hydrolysis-driven protein	1
Polygalacturonate 4-alpha-galacturosyl-tr	1
Potassium ion transmembrane	1
Proline dehydrogenase activity	1
Protein binding	33
Protein disulfide oxidoreductase activity	3
Protein disulfide isomerase activity	1
Protein homodimerization activity	1
Protein transporter activity	2
Protein serine/threonine phosphatase	1
Purine base transmembrane transporter	2
Rrna-binding	7
Sequence specific dna binding	21
Serine o-acetyltransferase activity	1
Serine racemase activity	1
Serine-type endopeptidase inhibitor	2
Sequence-specific signalling pathway	1
Signal transducer activity	2
Small conjugating proten ligase activity	2
Snorna binding	1
Steroid binding	1
Strictosidine synthase activity	1
Structural constitute of ribosome	35
Succinate dehydrogenase (ubiquinone)	1
Sugar:hydrogen symporter activity	1
Transferase activity	3
Translation initiation factor activity	1
Transmembrane transporter activity	2
Triglyceride lipase activity	1
Triose-phosphate isomerase activity	1
Ubiquinol-cytochrome-c reductase	1
Ubiquitin binding	1
Ubiquitin protein ligase activity	1
Uracil phosphoribosyltransferase activity	1
Vacuole	1
Water channel activity	1
Zinc ion binding	6
1,3-beta-d-glucan synthase activity	1
1-acylglycerol-3-phosphate o-acyltran	1
9-cis-epoxycarotenoid dioxgenase activity	1
Inorganic anion transporter activity	1
Intracellular cyclic nucleotide activated	1
Inositol monophosphatase activity	1
Isopentenyl-diphosphate delta-isomerase	1
Kinase activity	4
Lipase activity	1
Lipid binding	5
Lipoxygenase activity	1

Table 3: Gene ontology based functional annotation and classification of Singlet-SSRs of *Humulus Lupulus*

Gene ontology(Biological process)	SSR-ESTs (numbers)	Gene ontology(Molecular function)	SSR-ESTS (numbers)	Gene ontology (Cellular component)	SSR-ESTs (numbers)
Abscisic acid mediated signalling pathway	3	Acetate-coa ligase activity	1	Anchored to membrane	4
Acetate metabolic process	1	Acetolactate synthase activity	1	Apoplast	14
Actin cytoskeleton organization	3	Acid phosphatase activity	1	Axon	1
Activation of protein kinase C	1	Acyl carrier activity	2	CCAAT-binding factor complex	1
Aerenchyma formation	1	Acyl-coa dehydrogenase activity	1	Cell surface	1
Aerobic respiration	1	Adenylylsulfate kinase activity	3	Cell wall	23
Aging	4	Alanine-glyoxylate transaminase activity	1	Cellular bud neck	1
Aluminium ion transport	2	Alcohol dehydrogenase(nad) activity	1	Chloroplast	78
Amino acid catabolic process	1	Allene-oxide cyclase activity	1	Chloroplast photosystem II	6
Amino acid import	1	Alpha-amylase activity	2	Chloroplast stroma	1
Amino acid tansport	1	Alpha-galactosidase activity	1	Chloroplast stromal thylakoid	1
Ammonia assimilation cycle	2	Amino acid transmembrane transporter	3	Chloroplast thylakoid lumen	2
Anther dehiscence	2	Amino acid binding	2	Chromosome	1
Anti-apoptosis	1	Ammonia-lysate activity	1	Clathrin vesicle coat	1
Aspargine biosynthetic process	1	Anchored to membrane	1	CUL4 RING ubiquitin ligase complex	2
ATP synthesis coupled proton	4	Antioxidant activity	1	Cytoplasm	21
Auxin mediated signaling pathway	3	Aspargine synthase	1	Cytosol	38
Auxin polar transport	2	Atp binding	28	Cytosolic large ribosomal subunit	6
Biosynthetic process	4	Atpase activity	9	Cytosolic small ribosomal subunit	5
Branched chain family amino acid biosynthetic	1	Beta-amylase activity	1	DNA-directed RNA polymerase II,core	1
Brassinosteroid biosynthetic process	5	Beta-galactosidase activity	1	Endosome membrane	4
Brassinosteroid mediated signalling pathway	3	Binding	6	Extracellular space	3
Cadmium ion transport	1	Caffeate o-methyltransferase activity	2	FACT complex	1
Calcium-mediated signalling	5	Calcium channel activity	1	Golgi membrane	4
Carbohydrate ,metabolic process	1	Calcium ion binding	5	Integral to membrane	4
Carbohydrate biosynthetic process	1	Calmodulin binding	5	Intracelluar	4
Carbohydrate metabolic process	7	Carbohydrate tranmembrane transporter	1	Intracellular cyclic nucleotide activated	1
Carbohydrate transport	2	Carbon sulfur lyase activity	2	Membrane	11
Carbon fixation	1	Carboxypeptidase activity	1	Mitochondrial matrix	2
Carboxylic acid metabolic process	1	Catalytic activity	6	Mitochondrial respiratory chain comp.	5
Carotene catabolic process	1	Cellulose synthase activity	2	Mitochondrion	12
Cation transport	1	Chalcone isomerase activity	2	Nuclear speck	2
Cell death	1	Chaperone binding	1	Nucleus	35
Cell fate specification	2	Chitin binding	1	Peroxisome	1
Cell fate specification	1	Chitinase activity	1	Plant-type cell wall	3
Cell redox homeostasis	3	Chloroallyl aldehyde dehydrogenase	2	Plasma membrane	21
Cell tip growth	1	Chlorophyll binding	1	Plasmodesma	4
Cell wall modification	1	Chromatin dna binding	1	Plastid	1
Cell wall thickening	3	Cobalt ion binding	1	Serine/threonine phosphatase complex	3
Cellular iron ion homeostatis	7	Conjugate hydrolase activity	1	Ubiquitin-protein ligase activity	2
Cellular localization	1	Copper chaperone activity	4	Vacuole membrane	2
Cellular metabolic process	3	Copper ion binding	9	1,3-beta-D-glucan synthase complex	1
Cellular respiration	1	Cyclic nucleotide binding	2		
Cellular response to cold	1	Cyclin-dependent protein kinase regulator	1		
Cellular response to ethylene stimulus	5	Cystathionine beta-lyase activity	1		
Cellular response to hydrogen peroxide	1	Cysteine synthase activity	1		
Cellular response to iron ion	1	Cysteine-type endopeptidase activity	2		
Cellular response to nitrogen starvation	1	Cytochrome-b5 reductase activity	2		
Cellular response to phosphate starvation	1	Diacylglycerol kinase activity	1		
Cellular respose to sulfate starvation	1	Dihydrolate reductase activity	1		
Chlorophyll biosynthetic process	1	Dna binding	15		
Chloroplast organization	3	Dna photolyase activity	1		
Cholestrol metabolic process	1	Dna-directed rna polymerase activity	2		
Coenzyme A biosynthetic process	1	Double-stranded rna binding	4		
Cold acclimation	1	Electron carrier activity	5		
Cotyledon development	2	Endopeptidase inhibitor activity	1		
Cyanide catabolic process	2	Enzyme inhibitor activity	1		
Cytochrome complex assembly	1	Epoxide hydrolase activity	1		
Cytochrome complex assembly	1	Fatty-acyl-coa binding	1		
Cytokinin transport	1	Fatty-acyl-coa reductase activity	1		
De-etiolation	1	Fk506 binding	1		
Defense response	2	Flavin adenine dinucleoide binding	1		
Defense response to bacterium	8	Glutamate synthase(nadh) activity	2		
Defense response to fungus	2	Glutathione transferase activity	1		
Defense response to insect	1	Glycerol-3-phosphate o-acytransferase	1		
DNA duplex unwinding	1	Gtp binding	8		
DNA repair	2	Heme binding	6		
DNA replication	1	Histidine-trna ligase activity	1		
DNA topological change	1	Hydrogen ion transporting atp synthase	4		
Electron transport chain	1	Hydrolase activity	1		
Embroyo development	5	Inorganic anion transporter activity	1		
Embroyo development ending in seed dormancy	6	Intracellular cyclic nucleotide activated	1		
Embryo sac development	1	Inositol monophosphatase activity	1		
Endocytic recyling	2	Isopentenyl-diphosphate delta-isomerase	1		
Entrainment of circadian clock	1	Kinase activity	4		
Establishment of cell polarity	5	Lipase activity	1		
Ethylene mediated signalling pathway	1	Lipid binding	5		
Far-red light signalling pathway	1	Lipoxygenase activity	1		
Fatty acid biosynthetic process	3	L-malate dehydrogenase activity	1		
Fatty acid metabolic process	1	Long-chain fatty acid-coa ligase activity	1		
Fatty-acyl-coa biosynthetic process	1	L-tyrosine:2-oxoglutarate aminotransferase	1		
Flavonoid biosynthetic process	5	Manganese ion binding	6		
G1 phase of mitotic cell cycle	1	Map kinase activity	1		
Growth	1	Metal ion binding	5		
Heat acclimation	1	Metalloendopeptidase activity	1		
Histydyl-trna aminoacylation	1	Mitochondrial respiratory chain complex	2		
Hyperosmotic salinity response	8	Mrna binding	1		
Inositol triphosphate metabolic process	2	Nucleic acid binding	4		
Intracellular protein transport	2	Nucleocytoplasmic transporter activity	1		

Ion transport	1
Jasmonic acid biosynthetic process	2
Lewis a epitope biosynthetic process	1
Lignin biosynthetic process	3
Lipid metabolic process	2
Lipid storage	3
Lipid transport	2
L-methionine biosynthetic process	1
L-serine biosynthetic process	1
Male gamete generation	3
Meiotic chromosome segregration	1
Metabolic process	6
Mrna modification	1
Multicellular organismal development	5
Negative regulation of abscisic acid mediated	1
Negative regulation of transcription,DNA	1
Nuclear mrna splicing,via spliceosome	3
Nucleocytoplasmic transport	2
Nucleotide-excision repair	1
Oxidation-reduction process	9
Peptidyl-histidine phosphorylation	1
Peptidyl-proline modification	1
Phosphatidyglycerol biosynthetic process	1
Photoinhibition	1
Photoreactive repair	1
Photorespiration	1
Photosynthesis	11
Plant-type cell wall loosening	1
Plant-type hypersensitive response	2
Pollen germination	3
Pollen sperm cell differentiation	1
Porphyrin-containing compound metabolic process	1
Post-translational protein modification	2
Potassium ion transport	3
Production of ta-sirnas involved in RNA	1
Protein folding	12
Protein import into chloroplast thykaloid	1
Protein phosphorylation	14
Protein targeting to mitochondrion	1
Protein targeting to vacuole	1
Protein transport	1
Protein ubiquitination involved in ubiquitin-	5
Proteolysis	3
Purine base transport	1
Regulation of abscisic acid mediated signalling	1
Regulation of cell cycle	1
Regulation of double fertilization forming a zygote	1
Regulation of transcription,DNA-dependent	13
Regulation to translation	2
Response to abscisic acid stimulus	2
Response to auxin stimulus	4
Response to brassinosteroid stimulus	1
Response to cadmium ion	1
Response to cold	1
Response to endoplasmic reticulum stress	1
Response to far red light	1
Response to jasmonic acid	2
Response to karrikin	4
Response to metal ion	2
Response to oxidative stress	1
Response to phosphate starvation	1
Response to salt stress	3
Response to stress	3
Response to water deprivation	1
Respose to nematode	1
Ribosome biogenesis	14
RNA methylation	1
RNA splicing	2
Serine family amino acid metabolic process	1
Small gtpase mediated signal transduction	2
Sterol biosynthetic process	1
Sulfate assimilation	1
Transcription,DNA-dependent	3
Translation	14
Transmembrane protein	2
Transport	1
Tricarboxylic acid cycle	2
Trichome branching	1
Trna splicing	1
Ubiquitin-dependent protein catabolic process	1
Vegetative to reproductive phase transition	1
(1->3)-beta-D-glucan biosynthetic process	1
10-formyltetrahydrofolate biosynthetic process	1
Nucleotide binding	1
Nucleotide-sugar transmembrane	1
Oxidoreductase activity	6
Pantothenate kinase activity	1
Peptidyl-prolyl cis trans isomerase activity	1
Phenylalanine ammonia-lyase activity	1
Phosphatidylinositol-4,5-biophosphate	2
Phosphoglycerate dehydrogenase activity	1
Phosphoric diester hydrolase activity	1
Polygalacturonase activity	1
Polynucleotide 5'-hydroxyl-kinase activity	1
P-p-bond-hydrolysis-driven protein	1
Polygalacturonate 4-alpha- related	1
Potassium ion transmembrane transporter	1
Proline dehydrogenase activity	1
Protein binding	33
Protein disulfide oxidoreductase activity	3
Protein disulfide isomerase activity	1
Protein homodimerization activity	1
Protein transporter activity	2
Protein serine/threonine phosphatase	1
Purine base transmembrane transporter	2
Rrna-binding	7
Sequence specific dna binding	21
Serine o-acetyltransferase activity	1
Serine racemase activity	1
Serine-type endopeptidase inhibitor	2
Sequence-specific signalling pathway	1
Signal transducer activity	2
Small conjugating proten ligase activity	2
Snorna binding	1
Steroid binding	1
Strictosidine synthase activity	1
Structural constitute of ribosome	35
Succinate dehydrogenase (ubiquinone)	1
Sugar:hydrogen symporter activity	1
Transferase activity	3
Translation initiation factor activity	1
Transmembrane transporter activity	2
Triglyceride lipase activity	1
Triose-phosphate isomerase activity	1
Ubiquinol-cytochrome-c reductase	1
Ubiquitin binding	1
Ubiquitin protein ligase activity	1
Uracil phosphoribosyltransferase activity	1
Vacuole	1
Water channel activity	1
Zinc ion binding	6
1,3-beta-d-glucan synthase activity	1
1-acylglycerol-3-phosphate o-	1
9-cis-epoxycarotenoid dioxgenase activity	1