Bioinformation Volume 5

open access

www.bioinformation.net Issue 3

Hypothesis

Analysis of unigene derived microsatellite markers in family solanaceae

Sanchita Gupta[1*], Kumar Parijat Tripathi[2], Sudeep Roy[1], Ashok Sharma[1]

[1]Biotechnology Division, Central Institute of Medicinal & Aromatic Plants (Council of Scientific & Industrial Research), Post Office CIMAP, Lucknow India; [2]Institute for Evolution and Biodiversity, University of Münster, Germany; Sanchita Gupta -E-mail: 0804sanchita@gmail.com , *Corresponding author.

Received August 4, 2010; Accepted August 26, 2010; Published September 20, 2010

Abstract:
The family Solanaceae is the source of several economically important plants. The aim of this study was to trace and characterize simple sequence repeat (SSR) markers from unigene sequences of *Solanum lycopersicum*, an important member of family Solanaceae. 18,228 unigene sequences of *Solanum lycopersicum* was taken in order to develop SSR markers and analyzed for the *in-silico* design of PCR primers. A total of 12,090 (66.32 %) unigenes containing 17,524 SSRs (microsatellites) were identified. The average frequency of microsatellites in unigenes was one in every 1.3 kb of sequence. The analysis revealed that trinucleotide motifs, coding for Glutamic acid (GAA) and AT/TA were the most frequent repeat of dinucleotide SSRs. Flanking sequences of the SSRs generated 877 primers with forward and reverse strands. Functional categorization of SSRs containing unigenes was done through gene ontology terms like Biological process, Cellular component and Molecular function.

Keywords: Unigene, Simple Sequence Repeat (SSR), *Solanum lycopersicum*, primers, functional annotation, gene ontology.

Background:
Tomato (*Solanum lycopersicum*) fruit is an important source of antioxidant (mainly pigment) compounds, as well as lycopene, β-carotene, ascorbic acid and polyphenols **[1]**. Tomato ripening involves a number of physiological processes which include the visible breakdown of chlorophyll and build-up of carotenoids, with massive accumulation of antioxidant components such as lycopene and β-carotene **[2]** within the plastids. Microsatellites are abundant across prokaryotic and eukaryotic genomes **[3]**. Here microsatellites were extracted from unigene. The unigene database was used to identify molecular markers used in the identification of genes of different plants. Expressed Sequence Tag (EST) databases were also mined for SSR markers. These SSR motifs would serve as locus-specific markers. But a major disadvantage of EST-derived microsatellites is the sequence redundancy that yields multiple set of markers at the same locus. Unigene derived microsatellite (UGMS) markers have the advantage of assaying variation in the expressed component of genome with unique identity and positions **[4]**. Therefore the simple sequence repeat (SSR) or microsatellite marker is currently the preferred molecular marker due to its highly desirable properties **[5]** and can serve as efficient and cost effective alternative markers in such species **[6]**. Several studies to identify microsatellites in rice, barley, wheat, maize, soybean, grapevine, sunflower and *Brassica* sp. Have been reported **[7, 8].** *In-silico* microsatellite marker studies have also been done in commercially important medicinal and aromatic plants **[9, 10]**.

Microsatellites or simple sequence repeats (SSRs) are stretches of DNA consisting of tandemly repeated short units of 1-6 base pairs in length **[8, 11]**.They are ubiquitous in prokaryotes as well as eukaryotes and can be found both in coding and non-coding region **[12]**. The elevated frequency of length polymorphism associated with microsatellite provides the basis for the development of a marker system that has broad application in genetic research including studies of genetic variation, linkage mapping, gene tagging and evolution **[7]**. The microsatellite markers have become a valuable tool for genetic studies, as they are able to efficiently screen large population sizes **[13]**. The uniqueness and the value of microsatellites arise from their multiallelic nature, codominant transmission, ease of detection by PCR, relative abundance, extensive genome coverage **[14]**.

In this study, SSRs were mined from unigenes. Various types of SSRs and their percentage distributions were determined. SSR markers developed from *Solanum lycopersicum* unigene sequences are used as marker tags to other plants. The primer sequences are the complementary sequences of the flanking ends of a stretch of simple sequence repeats (SSRs). The functional perspectives of the SSRs suggest that microsatellites are more than mere repetitive sequences and their role have been attributed to many biological functions **[9]**.

Methodology:
Sequence data source: The assembled and functionally annotated sequences of ESTs i.e. unigene sequences of solanaceous plant *Solanum lycopersicum* were retrieved from unigene database of NCBI (ftp://ftp.ncbi.nih.gov/repository/UniGene/Solanum_lycopersicum/). There were 18,228 unigenes available in the database. These unigenes are non-redundant datasets used to identify microsatellites, primers and gene ontology characterization.

Microsatellite identification: The unigene sequences were mined for microsatellites using a program MISA (MIcro Satellite) identification tool **[15]** written in the Perl scripting language.This tool analyses microsatellite repeats in FASTA formatted unigene sequences. The minimum motif repeat size were set to 10 for mononucleotide, 6 for dinucleotide, 5 for trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide in locating the microsatellites. The analysis of SSRs was done on the basis of their types (mono-hexanucleotide), number of repeats, percentage frequency of occurrence of each SSR motif and their distribution in the sequence. The results were cross checked through CUGI's SSR server **[16]**.

Gene ontology classification: Though unigenes are functionally annotated sequences have known or putative function. The unigene sequences were characterized through gene ontology terms using Amigo (http://amigo.geneontology.org) **[17]**. The ontology classification was done in terms of their biological process, molecular function and cellular component. This characterization of unigenes has been done on the basis of analyzed SSR repeats.

Bioinformation Volume 5

open access

www.bioinformation.net Issue 3 **Hypothesis**

Marker development: For the development of microsatellite markers, we designed primer pairs for all the identified microsatellites **[18]**.The microsatellites (excluding monomers) containing unigenes were used for designing primers pairs. The primers were designed from the flanking sequences having microsatellite repeats using PRIMER3 software (Whitehead Institute, USA). Forward and reverse primer pairs were designed for marker development. The optimum and maximum primer sizes were set to 20 and 25 nucleotides, respectively. The GC % was set to 40.0 to 60.0 and the Tm value between 50^0C to 70^0C.

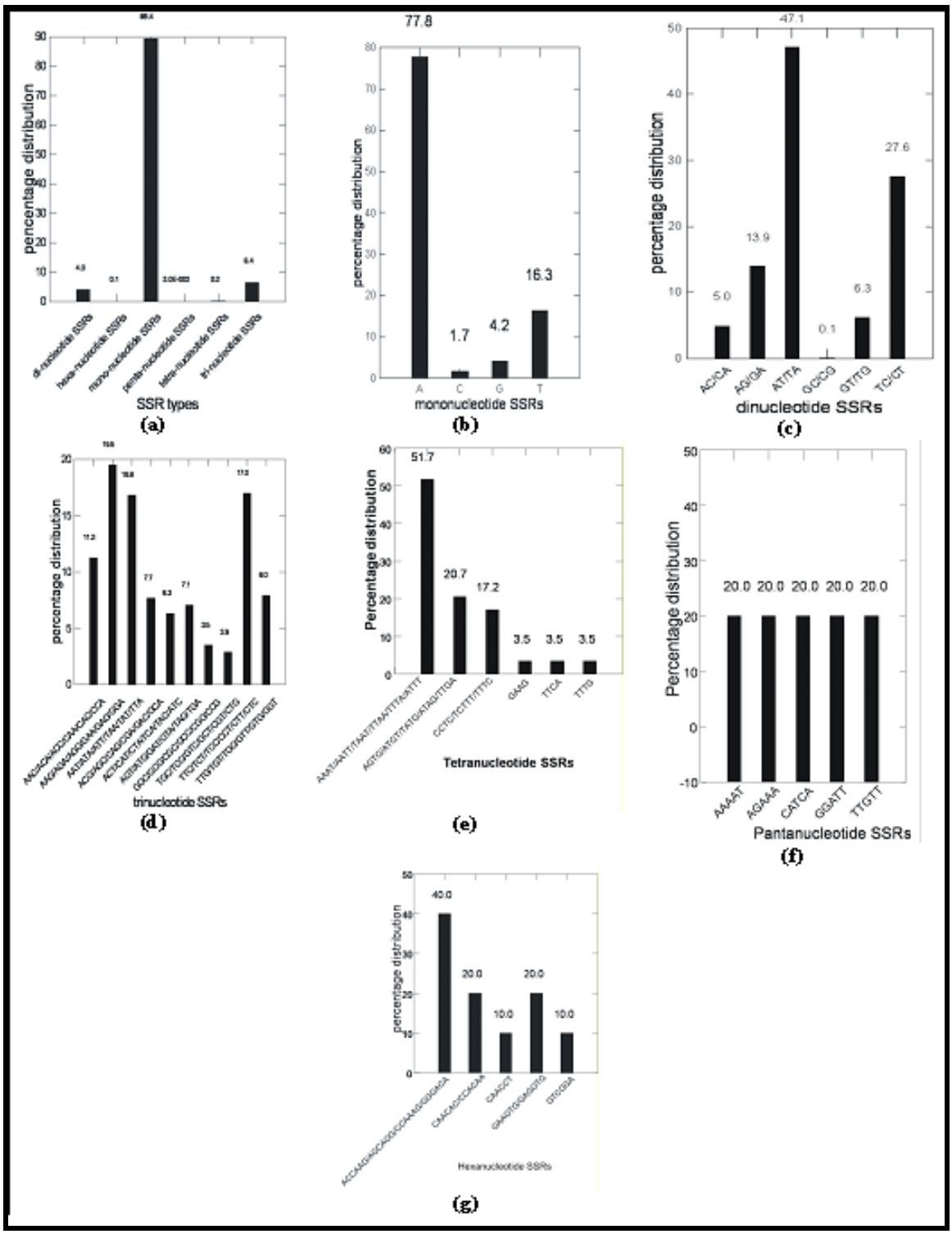

Fig. 1: a) Percentage distribution of different SSRs. b) Percentage distribution of mononucleotide SSRs. c) Percentage distribution of dinucleotide SSRs. d) Percentage distribution of trinucleotide SSRs. e) Percentage distribution of tetranucleotide SSRs. f) Percentage distribution of pentanucleotide SSRs. g) Percentage distribution of hexanucleotide SSRs.

Bioinformation Volume 5

www.bioinformation.net Issue 3

open access

Hypothesis

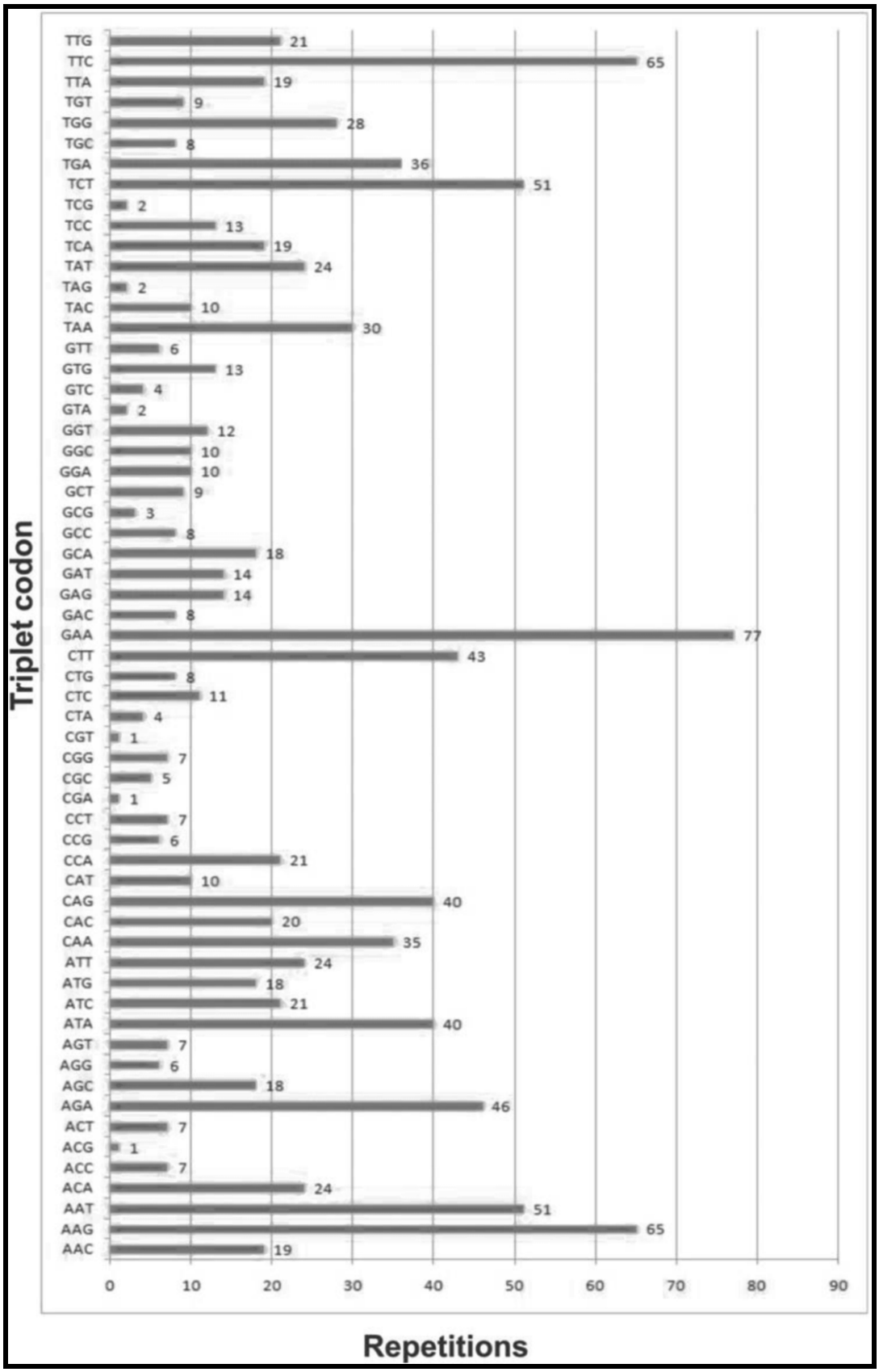

Figure 2: Frequency of distribution of triplet codons

Bioinformation Volume 5

www.bioinformation.net Issue 3

open access

Hypothesis

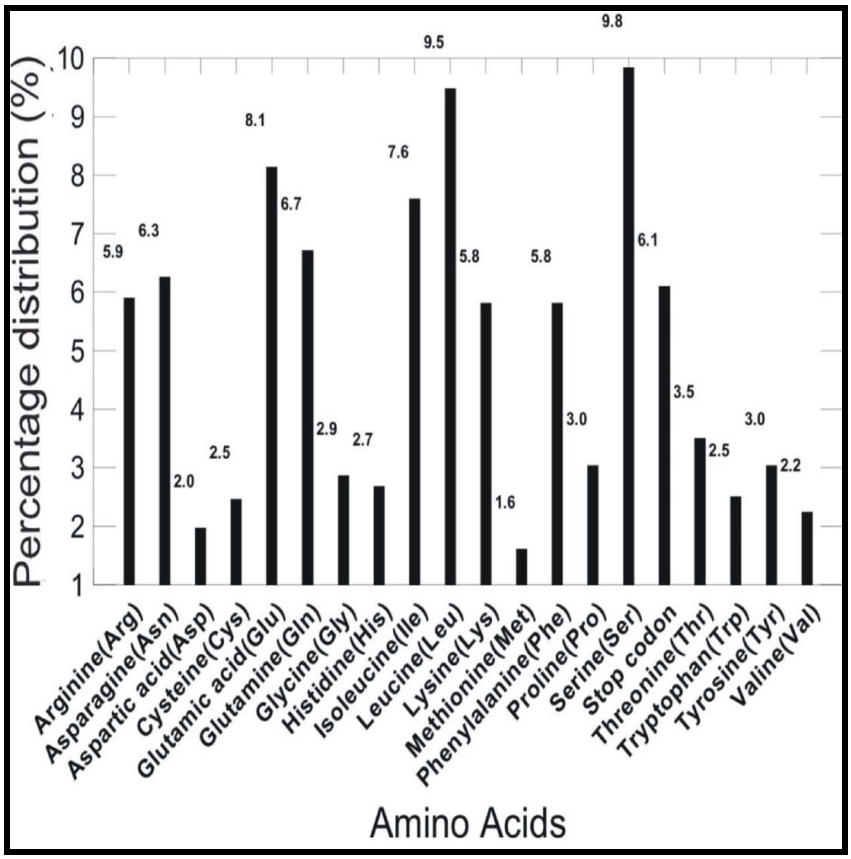

Figure 3: Percentage distributions of amino acids

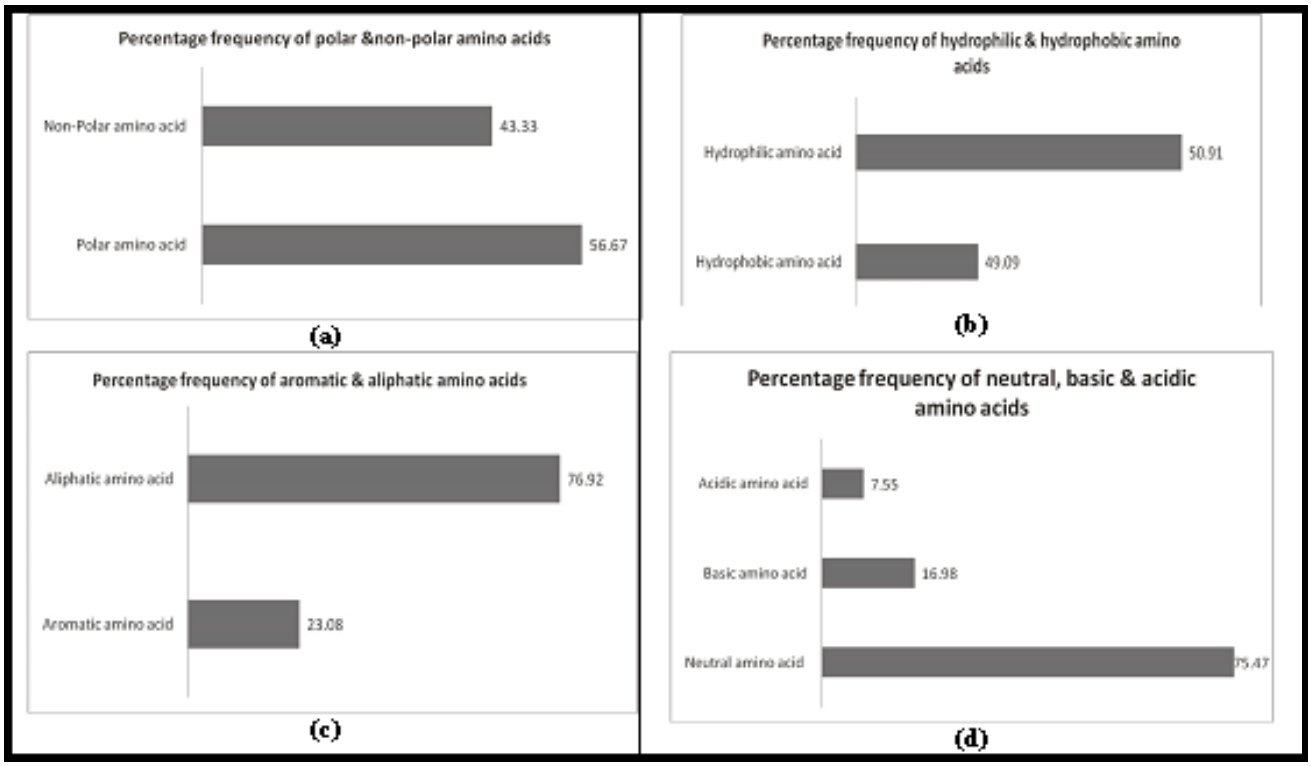

Figure 4: a Percentage frequency of polar & non-polar amino acids. **b)** Percentage frequency of hydrophilic & hydrophobic amino acids. c. Percentage frequency of aromatic & aliphatic amino acids. d. Percentage frequency of neutral, basic & acidic amino acids.

Bioinformation Volume 5
open access
www.bioinformation.net Issue 3 Hypothesis

Discussion:
18,228 unigene sequences were reported in *Solanum lycopersicum*. The unigene derived microsatellite (UGMS) markers have the advantage of unique identity and positions in the transcribed regions of the genome. With the availability of large unigene databases, it is now possible to systematically search for microsatellites in the unigenes **[19]**. The SSR repeat motifs were analyzed from these unigenes. Out of 18.228 unigene sequences (22.2 Mb), 12,090 showed the presence of simple sequence repeats (SSRs) suggesting that 66.32 % of unigenes contained SSRs. The average density of SSRs was 1(one) SSR per 1.3 kb of unigene sequence screened.

The study of occurrences of different types of SSR repeats revealed that the percentage of occurrence of mono-nucleotide SSR (89.36 %) was the highest followed by tri-nucleotide SSR (6.38 %) **(Figure 1a).** Among mononucleotide repeats, polyA/polyT repeats were predominant while polyC/polyG repeats were rare **(Figure 1b)**. As reported earlier, A-T repeat motifs are the most abundant type of SSRs in plants **[14]**. All dinucleotide repeat combinations excluding homomeric dinucleotides can be grouped into six unique classes, namely, (AG)n, (AT)n, (AC)n, (GT)n, (TC)n and (GC)n. It is evident that AT/TA dinucleotide repeats were more frequent, followed by TC/CT and AG/GA combinations. It is also important to note that GC/CG combination was the least frequent **(Figure 1c)**.

Among 10 unique trinucleotide repeat classes, AAG/AGA/AGG/GAA/GAG/GGA (19.5%) was the most frequent. The lowest frequency of trinucleotides was observed with TGC/TCG/GTC/CGT/CTG (2.86%) (Figure 1d). Similarly, AAAT/AATT/TAAT/TTAA/TTTA/ATTT motif showed maximum frequency within tetranucleotide (51.72%) followed by AGTG/ATGT/TATG/ATAG/TTGA (20.69 %). This also depicts the abundance of adenine and thymine dominance over other counterpart nucleotides. TTCA, TTTG and GAAG (3.45%) showed the least occurrence **(Figure 1e)**.

Four different pentanucleotide SSR motifs AAAAT, AGAAA, CATCA, GGATT, and TTGTT are present in equal proportions. Their contribution with respect to the motifs stands for 20% each, respectively **(Figure 1f)**. In hexanucleotide SSR motifs ACCAAG/AGCAGG/CCAAAG/GGGACA (40%) showed maximum distribution **(Figure 1g)**.

Codon Repetitions
The trinucleotide SSRs are triplet codon that code for a particular amino acid. It was observed that out of all triplet codons GAA (encoding Glutamic acid) repetitions are predominant followed by AAG (encoding Lysine) and TTC (encoding Phenylalanine) repeats. The triplet codons form an open reading frame (ORF) translated to proteins **(Figure 2)**.

Amino acid distribution
The trinucleotide microsatellites code for 21 types of amino acids that includes stop codon. It was observed that out of all coded amino acids Serine (Ser) demonstrated the highest percentage of occurrence followed by Leucine (Leu). In *Solanum lycopersicum* serine occur in Serine/threonine-protein phosphatase 5 and is involved in biological process such as intracellular signaling cascade, lipid metabolic process, protein amino acid dephosphorylation. Serine protease show biological process such as negative regulation of catalytic activity, proteolysis. Methionine (Met) and Aspartic acid (Asp) showed the least occurrence **(Figure 3)**.

The analysis of data revealed that the majority of amino acids were polar (56.67%) in nature (**Figure 4a**). The hydrophilic (50.91 %) amino acids occurred than more frequently hydrophobic (49.09 %) (**Figure 4b**). Similarly, frequency of aliphatic amino acids (76.92 %) were more than aromatic amino acids (23.08 %) (**Figure 4c**). The distribution study of chemical nature of amino acids gives an insight that neutral amino acids occurred more frequently with 75.47 % occurrence in comparison to basic and acidic amino acids 16.98 % and 7.55 % participation (**Figure 4d**).
Figure. 4: a Percentage frequency of polar & non-polar amino acids. **b**. Percentage frequency of hydrophilic & hydrophobic amino acids. **c**. Percentage frequency of aromatic & aliphatic amino acids**. d**. Percentage frequency of neutral, basic & acidic amino acids.

When we consider the mutation caused due to change in the last nucleotide of triplet codon, one amino acid changes into other. Percent frequency of occurrence of mutation was also tabulated. The analysis of the data revealed that mutation of Stop codon to Tyrosine and vice versa was the highest **(Table 1 see supplementary material).**

Study of functional protein of unigene having polymorphic condition
Microsatellites present in the transcribed regions of the genome have the potential to reveal functional diversity [4]. SSR motifs showing variation in their tandem repeats, multiple occurrences within same unigene and replacement of SSR motifs give a special account of polymorphism. Due to the polymorphic study there was change in the function of corresponding proteins. The SSR motifs could be utilized as molecular markers according to gene or protein of interest. Out of whole unigene sequences, 3 unigenes have dinucleotide SSR **(Table 2 see supplementary material)** and 8 unigenes having trinucleotides SSR repeats showed change in the protein function. The change in coding amino acids were also listed with corresponding trinucleotide SSRs **(Table 3 see supplementary material)**. Some amino acids such as Threonine, Lysine and Glutamine are responsible for homeodomain protein. Cysteine is for zinc-finger like protein. The findings support the theory that SSRs are randomly distributed in genomes and generally showed direct or indirect role in protein regulation **[20]**.

Change in gene ontology characterization has also been studied with reference to changes in functional proteins due to polymorphism of dinucleotide and trinucleotide SSRs **(Table 4 see supplementary material)**. It was observed that the proteins involved in polymorphism belong to transport, ovule development, metabolic process, defense response, regulation of transcription and biosynthetic process of biological category. The cellular components for these proteins were mostly cytoplasm, plastids, plasma membrane and mitochondrion. These proteins were DNA, GTP protein and zinc ion binding type and some have transcription factor activity, trypsin inhibitor activity and defensin protein in their molecular function.

Change in physicochemical properties of amino acids in trinucleotide polymorphism
The function of protein changes due to the change in amino acids repeat encoded by SSR repeats. The physicochemical properties of amino acids also vary due to conversion of amino acids **(Table 5 see supplementary material).** It was observed that in almost all the cases the changeover was from hydrophilic amino acids to hydrophobic amino acids. Two unigenes were found in which hydrophilic amino acid transformation remains the same property i.e. Glutamine and Glutamic acid converts to lysine which is hydrophilic in nature.

Gene ontology classification
Unigenes with corresponding SSRs were assigned GO terms of biological process, cellular component and molecular function **(Table 6 see supplementary material).** It was observed that some unigenes have known function but some show similarity with proteins of other plants. Some show no function or are putative. It was observed that most of the SSR containing unigenes belong to transport category of biological process. This indicates the development of SSR markers for proteins and enzymes involved in transport of lipid, potassium ion, and phytohormones in the metabolic pathway. Group of proteins were reported which played enzymatic roles in folic acid derivative, gibberellins and malate pathways. Identified transcription factors such as Homeodomain/Homeobox protein, auxin response factor 8 and NAC domain associated with these SSR have been identified that have role in gene regulation process. There were other proteins identified as regulatory gene category such as Myb-related transcription factor protein, glycine rich protein and homeodomain protein. Superoxide dismutase [SOD], carbonic anhydrase and telomere binding protein, CTR1 like kinase protein, WRKY transcription factor 2 and Isochorismate synthase are the prominent proteins that figured out under different heads of defense and stress response. Different class of enzymes

Bioinformation Volume 5

open access

www.bioinformation.net Issue 3 **Hypothesis**

like transferase, oxidase, oxidoreductase, synthase, kinase, aminopeptidase and most importantly fructofuranosidase and nitrilase involved in different synthetic pathways are also of worth mention to the findings. Beside this, single SSR corresponding to enzymes with catalytic role in chloroplast only was reported. Unigenes having SSRs corresponding to cell metabolism showed their localization in endoplasmic reticulum, cytosol, plastid, and chloroplast pertaining to different functional pathway was identified. It was also revealed that predicted unigenes possess some important role in following molecular functions i.e., regulation of transcription and translation, phosphorylation, transport regulation, signaling, defense response, stress and biosynthesis of pigments.

Primer designing

The primer designing have been done for PCR amplification of the desired microsatellites using Primer 3.0 software. The primers flanking the microsatellites repeat-motifs could be designed for 877 (47.02%) of the 1,865 microsatellites in the analyzed plant species **[18]**. It was observed that the forward and reverse primers obtained from trinucleotide SSRs were (628) maximum followed by dinucleotide (227). The primers from tetranucleotide and hexanucleotide SSRs were 14 and 8.

Conclusions:

This study is a step forward towards the utilization of in-silico approaches to analyze microsatellites (SSRs) from unigene sequences of plants. Unigene database provide a valuable resource for the development of SSR markers which are associated with transcribed genes. Our study revealed the analysis of microsatellites in UniGenes of *Solanum lycopersicum*. UGMS markers identified and characterized in this study provided insight about the abundance and distribution of SSR in the expressed organeller genome of *Solanum lycopersicum*. SSR markers are very informative because they show co-dominancy and highly polymorphic. SSRs markers are highly mutable loci, can be used for scanning of genome by genotyping and a particular region can be identified in the genome.

The development of SSR markers from unigene sequences saves both cost and time, once sufficient amount of data is available. It is also interesting to note the mutational changes among trinucleotides SSRs due to change in last triplet codon. This helps in deducing markers of own choice to study. Out of the unigenes investigated, some possessed more than one SSRs. Flipping of one SSR to another led to the change in function of protein. Microsatellites have proved very useful as molecular markers in diverse areas of genetic research including genome characterization and mapping.

This study demonstrated the utility of computational approaches for mining SSRs from ever increasing repertoire of publicly available plant unigene sequences present in different databases. Computational approaches provide an attractive alternative way to conventional laboratory methods for rapid and economical development of SSR markers by utilizing freely available genomic sequences in public databases.

Acknowledgement:

Funding by Department of Biotechnology, New Delhi under BTIS network programme is gratefully acknowledged. The help of Dr. Feroz Khan, Scientist, CIMAP, at various stages of manuscript preparation is also acknowledged.

References:

[1] A Lai *et al. Postharvest Biology and Technology*. **43:** 335 (2007).
[2] D Laval-Martin *et al., Phytochemistry*, **14:** 2357 (1975).
[3] P Rajendrakumar *et al. In Silico Biology*, **8** (2008).
[4] SK Parida *et al. Theor Appl Genet*. **112:** 808–817 (2006).
[5] C He *et al. Theoretical and Applied Genetics* **106:** 363 (2003).
[6] R Sharma *et al. BMC Plant Biology*. **9:** 53 (2009).[PMID: 19426565]
[7] RV Kantety *et al. Plant Molecular Biology*. **48:** 501 (2002).
[8] RK Varshney *et al. Cellular & Molecular Biology Letters*, **7**: 537 (2002).
[9] KP Tripathi *et al. Online Journal of Bioinformatics*. **9**: 78 (2008).
[10] KP Tripathi *et al. Plant omics journal*. **2**(6): 228 (2009).
[11] T Thiel *et al. Theor Appl Genet*. **106**: 411 (2003).
[12] S Jung *et al. Functional & Integrative Genomics*. **5:** 136 (2005).
[13] DJ Mullan *et al. Genome*. **48:** 811 (2005).
[14] W Powell *et al. Trends Plant Sci*. **1**: 215 (1996).
[15] MISA- (http://pgrc.ipk-gatersleben.de/misa/).
[16] CUGI'S ssr server- http://www.genome.clemson.edu/cgi-bin/cugi_ssr.
[17] S Carbon *et al. Bioinformatics*. **25**: 288 (2009).
[18] Primer3 - http://frodo.wi. mit.edu/primer3/input.htm.
[19] SK Parida *et al. Genome*. **53**: 55 (2010).
[20] M Michele *et al. Nature Genetics*. **30**: 194 (2002).

Edited by P. Kangueane

Citation: Gupta *et al.* Bioinformation 5(3): 113-121 (2010)

Bioinformation Volume 5

open access

www.bioinformation.net Issue 3 **Hypothesis**

Supplementary material:

Table 1: Percentage frequency of amino acids due to mutation

Mutation in amino acid	Percentage frequency of occurrence
Asparagine <->Lysine	7.14 %
Arginine <-> Serine	14.28 %
Isoleucine <-> Methionine	10.72 %
Glutamine <-> Histidine	14.28 %
Glutamic acid <-> Aspartic acid	14.28 %
Stop codon <-> Tyrosine	17.88 %
Stop codon <-> Cysteine	7.14 %
Tryptophan <-> Cysteine	7.14 %
Leucine <-> Phenylalanine	7.14 %

Table 2: Polymorphic study of dinucleotide SSRs

Sr. no.	Unigene id	Di nucleotide SSR	Change in functional protein
1.	gnl|UG|Les#S38752764	(TC)7 (GA)10	Jasmonic acid 1 Homeobox-leucine zipper protein HAT7 DNA-binding protein
2.	gnl|UG|Les#S24583765	(TC)6 (AT)16	protease inhibitor defensin protein gamma-thionin/defensin
3.	gnl|UG|Les#S38753241	(TC)6(TA)6 (AT)8	ARA6; GTP binding small GTPase Rab5b RAB5B

Table 3: Polymorphic study of trinucleotide SSRs

Sr. no.	Unigene id	Tri nucleotide SSR	Change in protein	Change in amino acid
1.	gnl|UG|Les#S19872558	(ACA)5	Bell-like homeodomain protein 2	Threonine
		(TGG)5	BEL5 protein	Tryptophan
2.	gnl|UG|Les#S15898250	(AAG)5	Bell-like homeodomain protein 3	Lysine
		(TGG)5	BEL1-related homeotic protein 14	Tryptophan
3.	gnl|UG|Les#S19872593	(CAC)5	Clone 132063F, mRNA sequence	Histidine
		(CCA)5	leucine-rich repeat/extensin	Proline
4.	gnl|UG|Les#S19871641	(CAG)7	Clone 133664F, mRNA sequence	Glutamine
		(AAG)6	ATP binding / ATP-dependent helicase/ nucleic acid binding	Lysine
5.	gnl|UG|Les#S38752473	(AAG)6	Transcribed locus	Lysine
		(CCG)5	proline-rich family protein	Proline
6.	gnl|UG|Les#S38752397	(GAG)5	unknown protein	Glutamic Acid
		(AAG)6	Kynurenine 3-monooxygenase and related flavoprotein monooxygenases RAP domain protein	Lysine
7.	gnl|UG|Les#S38752238	(AAC)5	predicted protein	Threonine
		(TGC)7	zinc finger-like protein	Cysteine
8.	gnl|UG|Les#S5294089	(ATA)6	Transcribed locus	Isoleucine
		(CAA)6	BEL1-like homeodomain transcription factor	Glutamine

Bioinformation Volume 5
www.bioinformation.net Issue 3
open access
Hypothesis

Table 4: Changes in gene ontology

Sr. no.	Change in protein	Change in gene ontology		
		Biological process	Cellular copmponent	Molecular function
1.	Jasmonic acid 1	jasmonic acid biosynthetic process	plastid	allene-oxide cyclase activity
				12-oxophytodienoate reductase activity
	Homeobox-leucine zipper protein HAT7	regulation of transcription, DNA-dependent	plasma membrane nucleus	protein binding
				DNA binding
2.	protease inhibitor	negative regulation of proteolysis	outer membrane-bounded periplasmic space	serine-type endopeptidase inhibitor activity
				trypsin inhibitor activity
	defensin protein	defense response	cell wall	
3.	GTP binding	intracellular protein transport	chloroplast	GTP binding
	small GTPase Rab5b	protein transport	plasma membrane	GTP binding
4.	Bell-like homeodomain protein 2		plastid	protein binding
	BEL5 protein		cytoplasmic membrane-bounded vesicle	
5.	Bell-like homeodomain protein 3		plastid	protein binding
	BEL1-related homeotic protein 14	ovule development	cytosol	DNA binding
			nucleus	protein binding
				transcription factor activity
6.	Clone 132063F, mRNA sequence			
	leucine-rich repeat/extensin	cell morphogenesis involved in differentiation	plant-type cell wall	protein binding
7.	Clone 133664F, mRNA sequence			
	ATP binding / ATP-dependent helicase/ nucleic acid binding	cellular iron ion homeostasis	chloroplast	ATP binding
8.	Transcribed locus			
	proline-rich family protein		plastid	
9.	unknown protein			
	Kynurenine 3-monooxygenase	NAD metabolic process	mitochondrial outer membrane	kynurenine 3-monooxygenase activity
10.	predicted protein			
	zinc finger-like protein		mitochondrion	zinc ion binding
				DNA binding
11.	Transcribed locus			
	BEL1-like homeodomain transcription factor	ovule development	cytosol	transcription factor activity

Table 5: Change in characteristics of amino acids in trinucleotide polymorphism

Sr. no.	Change in amino acid	Change in characteristics	
1	Threonine -> Tryptophan	Polar hydrophilic neutral	Non polar hydrophobic neutral
2	Lysine -> Tryptophan	Polar hydrophilic charged (+)	Non polar hydrophobic neutral
3	Histidine -> Proline	Polar hydrophilic charged (+)	Non polar hydrophobic neutral
4	Glutamine -> Lysine	Polar hydrophilic neutral	Polar hydrophilic charged (+)
5	Lysine -> Proline	Polar hydrophilic charged (+)	Non polar hydrophobic neutral
6	Glutamic Acid -> Lysine	Polar hydrophilic charged (-)	Polar hydrophilic charged (+)
7	Threonine -> Cysteine	Polar hydrophilic neutral	Polar hydrophobic neutral
8	Isoleucine -> Glutamine	Non polar hydrophobic neutral	Polar hydrophilic neutral

Bioinformation Volume 5
www.bioinformation.net Issue 3
open access
Hypothesis

Table 6: Gene ontology based functional annotation and classification of SSRs of *Solanum lycopersicum*

Sr. No.	Gene Ontology	Unigenes
1. Biological process		
1.	Metabolism	4
2.	Transport	8
3.	Cell cycle	1
4.	Glucose starvation	2
5.	Regulation of transcription	5
6.	Regulation of translation	1
7.	Phosphorylation	2
8.	Biosynthesis	2
9.	Signaling	1
10.	Catabolic process	1
11.	Stress response	7
12.	Auxin stimulus	2
13.	Protein folding	1
14.	Protein binding	2
15.	Cell wall organization	3
16.	Actin synthesis	1
17.	Anthocyanin biosynthesis	1
2. Cellular component		
1.	Mitochondrion	2
2.	Plasma membrane	2
3.	Endoplasmic reticulum membrane	4
4.	Cytoplasm	7
5.	Contractile Vacuole	1
6.	Nucleus	6
7.	Cytoplasmic membrane bounded vesicle	5
8.	Chloroplast	2
9.	Plastid	6
11.	Extracellular region	1
12.	Cytoskeleton	1
13.	Integral to membrane	2
3. Molecular function		
1.	Lipid binding	1
2.	Potassium channel	1
3.	Protein binding	8
4.	Oxidase activity	2
5.	Synthase	2
6.	Calcium ion binding	1
7.	Kinase	3
8.	Aminopeptidase activity	2
9.	Catalytic activity	1
10.	Transcription factor activity	4
11.	Transporter	1
12.	Reductase	1
13.	RNA binding	2
14	Actin binding	1
15.	Transferase	1
16.	Dehydratase	1
17.	DNA binding	3
18.	Phosphate binding	1
19.	Oxidoreductase	2
20.	Chromatin binding	1
21.	Fructofuranosidase activity	1
22.	Nitrilase activity	1
23.	ATP, GTP binding	2
24.	Antioxidant activity	1