Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2016, 3(3):663-668**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Analysis of Housekeeping and Tissue Specific ESTs for Inexact Microsatellites in *Capsicum* L.

Ayse Gul Ince, Mehmet Karaca

Department of Field Crops, Faculty of Agriculture, Akdeniz University, Turkey

Abstract A microsatellite is a tract of tandemly repetitive DNA in which certain DNA motifs are repeated. There are several types of microsatellites, which are also called simple sequence repeats (SSRs). The same type of repeating unit (motif) makes microsatellites perfect or exact while any mutations in the repeating unit make microsatellites imperfect or inexact microsatellites. In the present study the density differences of inexact microsatellites and motif types of housekeeping and tissue specific ESTs were studied in 7 pepper (*Capsicum* L.) tissues and development stages. Based on the analyses of 13,261 expressed sequence tags (ESTs) we observed that some tissues contained more inexact microsatellite containing ESTs while some other tissues contained less density of inexact microsatellites. Analyses also revealed that housekeeping genes and tissue specific ESTs contained statistically different amount of inexact microsatellites. These results indicated that microsatellites or SSRs are not just genetic markers used in markers assisted selection and fingerprinting studies, microsatellites may have an effect on gene expression and may play an important role in contributing to the different expression profiles between housekeeping and tissue-specific genes.

Keywords gene expression, housekeeping genes, development stages, microsatellites, SSR, tissue specific

1. Introduction

Expressed sequence tags (ESTs) are 300-700 bp short, single pass sequence reads derived from either the 5' or 3' end of a complementary DNA (cDNA) at a low cost. ESTs represent mRNAs expressed in a tissue at a development stage. ESTs are grouped into clusters and contigs based on sequence similarity and identity matches. Assembled ESTs are used for the discovery and characterization of candidate genes, investigation of alternative splicing, construction of expression maps, discrimination between genes exhibiting tissue or disease specific expression, genome annotation, gene structure prediction, development and characterization of microsatellites and single nucleotide polymorphisms (SNPs) and facilitate proteome analysis. A large amount of ESTs were generated in recent years with the development of next-generation sequencing approaches. Thus, ESTs are now much easily available for many plant species [1-2].

To date it is known that a significant portion of genomic DNA consists of repeating motifs of various sizes, varying from very small to very large nucleotides. Repeated DNA could be classified into two large families, tandem and dispersed repeats. Dispersed repeats contain transposons, tRNA genes, and gene paralogues. Tandem repeats contain gene tandems, ribosomal DNA repeat arrays, satellites, midisatellite, minisatellites, and microsatellites. Among the tandem repeats, microsatellite, also known as simple sequence repeats (SSRs) are abundant in structural regions such as telomeres, centromeres, histone binding areas and transcribed regions. Microsatellites have attracted the attention of researchers due to their extensive use in the construction of genetic maps, the association between repeat variations and genetic diseases, their practicability and ease of use in studies of population genetics, and for genotyping and paternity analysis. The presence of microsatellite to cDNAs accelerated the use of microsatellites in applied and theoretical genetics [3-9].

Microsatellites are DNA regions composed of small motifs of 1 to 6 nucleotides repeated in tandem, which are widespread in both eukaryotic and prokaryotic genomes. Microsatellites are classified according to the motif length (mono-, di-, tri-, tetra-, penta- or hexa-nucleotide repeats), the type of repeats (exact and inexact), exact compound and inexact compound (Fig. 1). In an exact microsatellite the repeat sequence is not interrupted by any base not belonging to the motif while in an inexact microsatellite there is a pairs of bases between the repeated motifs that does not match the motif sequence. In the case of an exact compound microsatellite there are at least two adjacent distinctive exact sequence-repeats. On the other hand, an inexact compound microsatellite there are at least two adjacent sequence repeats and at least one of which consist of inexact repeats [3, 8, 10-12].

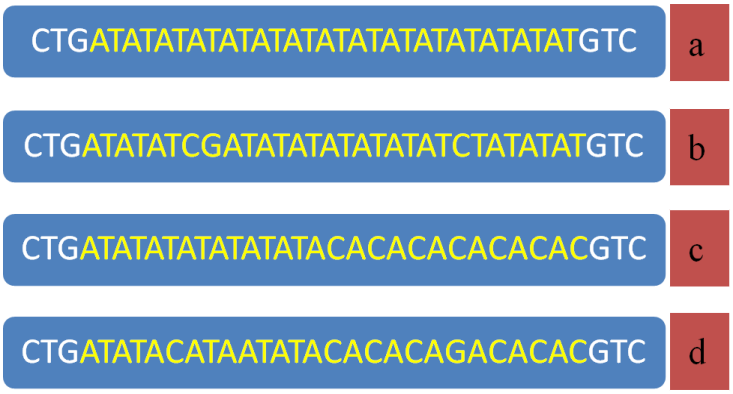

Figure 1. Exact, inexact, exact compound and inexact compound microsatellites. a: exact or perfect microsatellite repeats consisiting of $[AT]_{15}$*, b: inexact or imperfect microsatellite repeats consisiting of interrupted [AT] repeats, c: exact compound microsatellite consisiting of* $[TA]_7$ *and* $[CA]_6$ *repeats, d) inexact compound microsatellite consisiting of exact* $[CA]_6$ *repeats and interrupted [TA] repeats.*

Inexact microsatellite and inexact compound microsatellite repeats may originate from frequent point mutations in the exact microsatellite repeats. However, exact microsatellite and exact compound microsatellite may also originate from inexact microsatellites. New studies are required to test which of the mentioned scenarios are true. The inexact or imperfect microsatellite repeats may be considered as extensions in the definitions of exact tandem repeats during certain editing [13] or sequences with lesser degree of periodicity and spaced from another island showing periodic sequence [14]. Additionally, inexact motif patterns of microsatellites may also depend on the genomic context such as the particular location on a chromosome and functional potential of the transcribed product, as well as the effectiveness of mismatch repair enzymes. Moreover, inexact motif rates in microsatellites are also affected by stabilization patterns and potential secondary structures [10].

Microsatellite repeats can be found in housekeeping and tissue specific genes [15]. Housekeeping genes are those genes constitutively expressed in all tissues to maintain cellular functions. Previous research revealed that the expression level of the housekeeping genes varies among tissues and changes under certain circumstances. On the other hand tissue specific genes are those genes specifically expressed in a tissue or development stage [10]. Despite the most fundamental characteristics of housekeeping genes and tissue specific genes, no previous study has quantified inexact microsatellite densities of the housekeeping genes and the tissue specific genes in a variety of tissues. This study was undertaken to determine inexact microsatellite density differences in tissue and housekeeping genes. Also we investigate inexact microsatellite differences among anther, young fruits, flower bud, early root, hairy root, placenta and leaf tissues of *Capsicum annuum* L. cultivar Demre Sivrisi.

2. Materials and Methods

2.1. Expressed Sequence Tags (ESTs)

A total of 116535 *Capsicum annuum* L. ESTs from National Center for Biotechnology Information at http://www.ncbi.nlm.nih.gov/ were downloaded and analyzed. These ESTs consisted of 129,149,486 base pair nucleotide information.

2.2. Assignment of ESTs

Keyword Finder and OrgMiner [16] bioinformatic programs were used to obtain ESTs specific to each of anther, hairy root, early root, leaf, young fruit, placenta and flower bud library. In order to identify tissue specific (library specific) ESTs and housekeeping ESTs, a total of 23,098 ESTs containing 11.06 mega base nucleotides were assembled into contiguous sequences (contigs) using Sequencher software (Gene Codes, Ann Arbour, MI). Contigs assembly parameters were set to minimum overlap of 50 bases and 95% sequence homology match.

2.3. Microsatellite analyses

Microsatellites in each dataset were identified using the Tandem Repeats Analyzer 1.5 (TRA 1.5) program [3]. Microsatellites in the present study were considered sequences containing a minimum of 18, 9, 7, 5, 5 and 4 nucleotide inexact repeats for mono- di-, tri-, tetra-, penta- and hexa-nucleotides, respectively.

2.4. Statistical analysis

Chi-square (χ^2) goodness-of-fit tests with 1 degree of freedom were applied to test whether inexact microsatellite densities were significantly different within and between datasets.

$$E_i = \frac{N}{L} * L_i$$

where E_i is the expected number of microsatellites in a dataset, N is the total number of microsatellites in the two different datasets, L is the total length in base pairs of the two datasets, and L_i is the length in base pairs of the dataset under investigation [15].

3. Results and Discussion

In the present study housekeeping and tissue specific ESTs were studied in a total of seven different tissues and development stages of pepper cultivar Demre Sivrisi (*Capsicum annuum* L.) and other species in the GenBank.". Using 510 ESTs belonging to anther it was found that anther ESTs (χ^2=45.89, P≤0.0001) contained lower densities of inexact microsatellites than they would be predicted purely on the grounds of base composition. Also ESTs belonging to placenta (2747 ESTs, χ^2=45.89, P≤0.0001) showed low densities of inexact microsatellites (Table 1). Early root (1460 ESTs, χ^2=10.15, P≤0.0005), flower bud (1919 ESTs, χ^2=37.54, P≤0.0001) and hairy root (1473 ESTs, χ^2=5.87, P≤0.0031) contained higher densities of inexact microsatellites than they would be predicted purely on the grounds of base composition (Table 1). On the other hand, leaf tissues (3147 ESTs) and young fruit tissues (1998 ESTs) contained inexact microsatellites as expected densities (Table 1). Previous studies have showed that exact microsatellite densities of plant and animal tissues differed [3, 12, 15, 17].

Inexact mono-nucleotide microsatellite differences among tissues varied. Analysis revealed that inexact mononucleotide densities of anther tissues (χ^2=41.65, P≤0.0001) and placenta tissues (χ^2=97.56, P≤0.0001) contained lower densities of inexact microsatellites than they would be predicted purely on the grounds of base composition. Early root (χ^2=16.43, P≤0.0005), flower bud (χ^2=57.40, P≤0.0001) and hairy root (χ^2=8.75, P≤0.0005) contained higher densities of inexact mono-nucleotide microsatellites than they would be predicted purely on the grounds of base composition (Table 1). However analysis revealed that leaf tissues and young fruit tissues contained inexact mono-nucleotide microsatellites as expected densities (Table 1). Differences in exact microsatellite and mono-nucleotide microsatellites were also previously reported in human [12], Arabidopsis [3], turmeric [4] and pepper [10].

Inexact di-nucleotide microsatellite densities of tissues and development stages differed greatly. Among the tissues and development stages two showed statistically significant differences while 5 had the same densities as they would be predicted purely on the grounds of base composition. Flower bud (χ^2=13.85, P≤0.0005) contained less amount of di-nucleotide repeats while leaf tissues (χ^2=19.44, P≤0.0005) contained more inexact di-nucleotides (Table 1). Inexact tri-nucleotide microsatellite density of young fruit (χ^2=6.27, P≤0.0031) was statistically significant (Table 1). Analysis revealed that none of the tissues and development stages studied in the present study contained significant density difference in tetra-, penta- and hexa-nucleotide microsatellites (Table 1).

Table 1: Inexact microsatellite densities among tissues and development stages of *C. annuum* L. cultivar Demre Sivrisi and other species in the GenBank

Tissues	# Base	# EST	# InExact-SSR		Mono-		Di-		Tri-		Tetra-		Penta-		Hexa-	
			O	E	O	E	O	E	O	E	O	E	O	E	O	E
Anther	352370	510	54	129	44	110	3	7	6	8	0	0.6	0	0.5	1	3.2
Other Tissues	8051052	12751	3033	2958	2588	2521	158	154	187	185	13	12.4	11	10.5	76	73.8
χ^2			45.895***		41.653***		2.175		0.565		0.569		0.481		1.606	
Total	8403422	13261	3087		2632		161		193		13		11		77	
Early Root	896544	1460	384	329	345	280	17	17	16	21	1	1.4	2	1.2	3	8.2
Other Tissues	7506878	11801	2703	2858	2287	2351	144	144	177	172	12	11.6	9	9.8	74	68.8
χ^2			10.153**		16.430***		0.002		1.146		0.121		0.651		3.706	
Total	8403422	13261	3087		2632		161		193		13		11		77	
Flower Bud	1239354	1919	576	455	526	388	7	24	27	28	2	1.9	1	1.6	13	11.4
Other Tissues	7164068	11342	2511	2632	2106	2244	154	137	166	165	11	11.1	10	9.4	64	65.6
χ^2			37.549***		57.404***		13.851***		0.088		0.004		0.280		0.279	
Total	8403422	13261	3087		2632		161		193		13		11		77	
Young Fruit	1246773	1998	477	458	398	391	22	24	41	29	2	1.9	0	1.6	14	11.4
Other Tissues	7156649	11263	2610	2628	2234	2242	139	137	152	164	11	11.1	11	9.4	63	65.6
χ^2			0.925		0.169		0.175		6.270*		0.003		1.916		0.682	
Total	8403422	13261	3087		2632		161		193		13		11		77	
Hairy Root	930283	1473	384	342	339	291	14	17.8	20	21.4	2	1.4	1	1.2	8	8.5
Other Tissues	7473139	11788	2703	2745	2293	2341	147	143.2	173	171.6	11	11.6	10	9.8	69	68.5
χ^2			5.877*		8.755**		0.922		0.098		0.246		0.044		0.036	
Total	8403422	13261	3087		2632		161		193		13		11		77	
Placenta	1614401	2754	406	593	306	506	33	30.9	40	37.1	5	2.5	4	2.1	18	14.8
Other Tissues	6789021	10507	2681	2494	2326	2126	128	130.1	153	155.9	8	10.5	7	8.9	59	62.2
χ^2			73.026***		97.567***		0.172		0.285		3.104		2.085		0.861	
Total	8403422	13261	3087		2632		161		193		13		11		77	
Leaf	2123697	3147	806	780	674	665	65	40.7	43	48.8	1	3.3	3	2.8	20	19.5
Other Tissues	6279725	10114	2281	2307	1958	1967	96	120.3	150	144.2	12	9.7	8	8.2	57	57.5
χ^2			1.147		0.157		19.441***		0.915		0.915		0.023		0.020	
Total	8403422	13261	3087		2632		161		193		13		11		77	

O: observed, E: expected, *: $P \leq 0.0031$, **: $P \leq 0.0005$ *** $P \leq 0.0001$

Based on the analysis of contigs ESTs specifically found in a particular tissues were called tissue specific ESTs. Analysis revealed that 375 ESTs were anther specific, 710 ESTs young root specific, 1124 ESTs flower bud specific, 1081 ESTs young fruit specific, 760 ESTs hairy root specific, 1727 ESTs placenta specific and 2128 ESTs leaf specific. These ESTs were considered tissue specific (TS) ESTs. On the other hand we considered those ESTs found in more than one tissue as housekeeping ESTs. For instance hairy root tissue contained 760 ESTs that were specifically found in hairy root while 713 ESTs not only found in hairy root but also were present in other tissues. Thus 713 ESTs were considered as housekeeping ESTs and inexact microsatellite density differences between a tissue or development stage and the rest of the other tissues and development stages were determined (Table 2).

Statistically significant inexact microsatellite differences between housekeeping and tissue specific ESTs were observed in anther and placenta tissues. Anther (χ^2=20.93, P≤0.0001) and placenta tissues (χ^2=55.13, P≤0.0005) contained less amount of inexact microsatellite densities than they would be predicted purely on the grounds of base composition (Table 2). In these tissues inexact microsatellites were very low while housekeeping ESTs had higher densities of microsatellites indicating that microsatellites play important role in gene expression.

Inexact mono-nucleotide microsatellite densities between tissue specific and housekeeping ESTs were statistically different in anther, young fruit and placenta tissues (Table 2). Among the tissues, young root tissue had statistically significant inexact di-nucleotide microsatellite density between housekeeping and tissue specific ESTs. Hairy root tissue (χ^2=4.14, P≤0.0031) was the only tissue with penta-nucleotide microsatellite density differences among the tissues and development stages studied in the present study (Table 2). Other types of inexact microsatellites of such as tri-, tetra- and hexa-nucleotides contained repeat densities as they would be predicted purely on the grounds of base composition in tissue specific and housekeeping ESTs.

Table 2: Inexact microsatellite densities between tissue specific and housekeeping genes in *C. annuum* L. cultivar Demre Sivrisi and other species in the GenBank

Tissues	# Base	#EST	# InExact SSRs		Motifs											
					Mono-		Di-		Tri-		Tetra-		Penta-		Hexa-	
			O	E	O	E	O	E	O	E	O	E	O	E	O	E
Anther (TS)	223169	375	18	34.2	13	27.9	3	1.9	2	3.8	0	0	0	0	0	0.6
Anther (HK)	129201	135	36	19.8	31	16.1	0	1.1	4	2.2	0	0	0	0	1	0.4
χ^2			20.93***		21.63***		1.737		2.325						1.727	
Total	352370	510	54		44		3		6						1	
Young Root (TS)	282194	710	122	120.9	116	108.6	1	5.4	2	5.1	1	0.3	1	0.63	1	0.9
Young Root (HK)	614350	750	262	263.1	229	236.4	16	11.6	14	10.9	0	0.7	1	1.37	2	2.1
χ^2			0.016		0.738		5.163*		2.671		2.177		0.318		0.005	
Total	896544	1460	384		345		17		16		1		2		3	
Flower Bud (TS)	559390	1124	237	259.9	216	237.4	1	3.2	15	12.2	1	0.9	0	0.4	4	5.9
Flower Bud (HK)	679964	795	339	316.1	310	288.6	6	3.8	12	14.8	1	1.1	1	0.6	9	7.1
χ^2			3.703		3.520		2.690		1.183		0.019		0.823		1.083	
Total	1239354	1919	576		526		7		27		2		1		13	
Young Fruit (TS)	509974	1081	179	195.1	141	162.8	10	9.0	21	16.8	2	0.8	0	0	5	5.7
Young Fruit (HK)	736799	917	298	281.9	257	235.2	12	13.0	20	24.2	0	1.2	0	0	9	8.3
χ^2			2.251		4.938*		0.189		1.805		2.890				0.160	
Total	1246773	1998	477		398		22		41		2				14	
Hairy Root (TS)	333591	760	125	137.7	105	121.6	5	5.1	10	7.2	0	0.7	0	0.4	5	2.9
Hairy Root (HK)	596692	713	259	246.3	234	217.4	9	8.9	10	12.8	2	1.3	1	0.6	3	5.1
χ^2			1.176		2.262		0.112		1.163		2.612		4.148*		1.650	
Total	930283	1473	384		339		14		20		2		1		8	
Placenta (TS)	790475	1727	124	198.79	77	149.8	15	16.2	17	19.6	4	2.5	3	1.9	8	8.8
Placenta (HK)	823926	1027	282	207.21	229	156.2	18	16.8	23	20.4	1	2.5	1	2.1	10	9.2
χ^2			55.138***		69.365***		0.163		0.669		1.927		1.085		0.147	
Total	1614401	2754	406		306		33		40		5		4		18	
Leaf (TS)	1229959	2128	452	466.80	371	390.4	45	37.7	19	24.9	1	0.6	2	1.7	14	11.6
Leaf (HK)	893738	1019	354	339.20	303	283.6	20	27.3	24	18.1	0	0.4	1	1.3	6	8.4
χ^2			1.115		2.280		3.414		3.326		0.727		0.094		1.198	
Total	2123697	3147	806		674		65		43		1		3		20	

TS: tissue specific, HS: housekeeping, O: observed, E: expected, $P \leq 0.0031$, **: $P \leq 0.0005$ *** $P \leq 0.0001$

4. Conclusion

Inexact microsatellite densities among some tissues and development stages significantly differed. ESTs in some tissues contained more inexact microsatellites than they would be predicted purely on the grounds of base composition. This indicated that inexact microsatellites play some important roles in tissue differentiation. Present study revealed that inexact microsatellite densities of tissue specific and housekeeping genes significantly differed. This indicated that inexact microsatellites play some important roles in tissue specific and housekeeping gene expression.

Acknowledgements

This research was supported in part by the Scientific and Technological Research Council and the Scientific Research Projects Coordination Unit of Akdeniz University.

References

[1]. Somme, L., Raabova J., Jacquemart, A.L. and Raspe O. (2012). Development and multiplexing of microsatellite markers using pyrosequencing in the clonal plant *Comarum palustre* (Rosaceae), *Molecular Ecology Resources*, *12*: 91-97.

[2]. Goodwin, S., John, D., McPherson, W. and McCombie R. (2016). Coming of age: ten years of next-generation sequencing technologies, *Nature Reviews Genetics*, *17:* 333-351.

[3]. Bilgen, B., Karaca, M., Onus, A.N., and Ince, A.G. (2004). A software program combining sequence motif searches with keywords for finding repeats containing DNA sequences, *Bioinformatics*, *20*: 3379-3386.

[4]. Ince, A.G., Karaca, M. and Onus, A.N. (2009). Differential expression of microsatellites in leaves and rhizomes of Turmeric (*Curcuma longa* Linn.), *Planta Medica*, *75*: 929-929.

[5]. Sharma, P.C., Grover, A. and Kahl, G. (2007). Mining microsatellites in eukaryotic genomes, *Trends in Biotechnology*, *25*: 490-498.

[6]. Guo, W.J., Ling, J. and Li, P. (2009). Consensus features of microsatellite distribution: Microsatellite contents are universally correlated with recombination rates and are preferentially depressed by centromeres in multicellular eukaryotic genomes, *Genomics*, *93*: 323-331.

[7]. Asari, N.S., Ramaswamy, M., Subbian, E.A., Palliyarakkal, M.K., Malhotra, S.K. and Karun, A. (2014). Standalone EST microsatellite mining and analysis tool (SEMAT): for automated EST-SSR analysis in plants, *Tree Genetics and Genomes*, *10*: 1755-1757.

[8]. Ince, A.G. and Karaca, M. (2015). E-microsatellite markers for some naturally occurring *Salvia* species in the Mediterranean region, *Turkish Journal of Biology*, *39*: 69-77.

[9]. Bolton, K.A., Holliday, E.G., Attia, J., Bowden, N.A., Avery-Kiejda, K.A. and Scott, R.J. (2016). A novel polymorphic repeat in the upstream regulatory region of the estrogen-induced gene EIG121 is not associated with the risk of developing breast or endometrial cancer, *BMC Research Notes*, *9:* 287-287.

[10]. Ince, A.G. (2010). *In vitro* and *in silico* approaches to determine the tissue/organ specific microsatellite DNA gene contents in *Capsicum* cDNA libraries, Ph.D. thesis, Akdeniz University, 2010, Antalya, Turkey.

[11]. Miah, G., Mohd, Y.R., Mohd, R.I., Adam, B.P., Harun, A.R., Nurul, K.I. and Mohammad A.L. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance, *International Journal of Molecular Sciences*, *14*: 22499-22528.

[12]. Karaca, M., Bilgen, M., Onus, A.N., Ince, A.G. and Elmasulu, S.Y. (2005). Exact Tandem Repeats Analyzer (e-TRA) for DNA sequence mining, *Journal of Genetics*, *84:* 49-54.

[13]. Volfovsky, N., Haas, B.J. and Salzberg S.L (2001). A clustering method for repeat analysis in DNA sequences, *Genome Biology*, *2*, RESEARCH0027.

[14]. Fischetti, V.A., Landau, G.M., Sellers, P.H. and Schmidt, J.P. (1993). Identifying periodic occurrences of a template with applications to protein structure, *Information Processing Letters*, *45:* 11-18.

[15]. Lawson, M.J. and Zhang, L. (2008). Housekeeping and tissue-specific genes differ in simple sequence repeats in the 50-UTR region, *Gene*, *407:* 54-62.

[16]. Ince, A.G., Karaca, M., Bilgen, M. and Onus, A.N. (2008). Digital differential display tools for mining microsatellite containing organism, organ and tissue, *Plant Cell, Tissue and Organ Culture*, *94:* 281-290.

[17]. Ince, A.G., Karaca, M. and Onus, A.N. (2011). Exact microsatellite density differences among *Capsicum* tissues and development stages, *Journal of Agricultural Sciences*, *17*: 291-299.