Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2017, 4(11):236-246**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Descriptive Statistics and PIC Values of Genomic- and Transcriptomic-Microsatellites in Several Plant Species

Ayse Gul Ince[1], Mehmet Karaca[2], Emine Uygur Gocer[2], Adnan Aydin[2]

[1]Vocational School of Technical Sciences, Akdeniz University, 07070 Antalya, Turkey
[2]Field Crops Department, Faculty of Agriculture, Akdeniz University, 07070 Antalya, Turkey

Abstract Microsatellites, also known as simple sequence repeats (SSRs) are very powerful molecular markers due to their high level of polymorphism, co-dominant inheritance, reproducibility and multi-allelic nature. Alleles of genomic- and transcriptomic-microsatellites of cotton and maize as well as published microsatellite data from many other crops species were used to reveal level of polymorphism differences between genomic- and transcriptomic-microsatellites; relationship between number of alleles and polymorphism, between motif length and polymorphism, between number of repeats and polymorphism, and level of fixation index values of genomic- and transcriptomic-microsatellites based on the observed and expected heterozygosity values. Based on 477 microsatellite loci studied, results indicated that polymorphic information content (PIC) values of genomic-microsatellites were higher than transcriptomic-microsatellites. Polymorphism levels of microsatellites increased as the number of allele increased. Although it is expected that loci with more number of repeats could be more polymorphic, there were not clear relationships between the number of repeats and polymorphism, and between motif length and polymorphism. Expected and observed heterozygosity values of genomic- and transcriptomic-microsatellites were different. Results also revealed that genomic and transcriptomic-microsatellite markers produce more null alleles, and occurrences of homoplasy are more frequent among different species. Results indicated that functional markers developed from transcriptomic-microsatellites might not be suitable for phylogenetic studies but they could be very useful in agro-genomic research.

Keywords EST, functional markers, microsatellites, motif type, SSR, simple sequence repeats

Introduction

First termed by Litt and Lutty in 1989 [1], Simple Sequence Repeats (SSRs), or also widely called microsatellites, are tandemly repeated DNA sequences of mono-, di-, tri-, tetra-, penta- and hexa-nucleotide motif units distributed in plant and animal genomes [2-5]. Microsatellite DNA markers are obtained using polymerase chain reaction (PCR) in high stringent annealing conditions that permit the specific amplification of single loci [6]. Microsatellites are co-dominant, abundantly found throughout the genome, highly polymorphic, multi-allelic, reproducible and reliable markers. They are widely utilized in many applications including genetic diversity, genetic mapping, hybrid identification, population genetics, marker-assisted selection, and in many other agro-genomic studies [7-9].

Microsatellites or SSRs could be classified according to the type of tandemly repeated sequences they contained. Perfect microsatellites are consisted of perfect repetitions, e.g., $(CG)_{20}$, imperfect microsatellites are consisted of repeated sequences that are interrupted by different nucleotides, e.g., $(GT)_{12}GC(TT)_8$, compound or composite microsatellites are consisted of two or more different motifs in tandem, e.g., $(CT)_7(CG)_6$. The compound type microsatellites can also be perfect or imperfect. Microsatellites could be classified based on their location in organellar or nuclear genomes. Microsatellites can be present in organellar genomes, such as

chloroplasts and mitochondria, and nuclei. Also microsatellites are classified according to their discovery methods. Genomic-microsatellites are conventional types and they are obtained using genomic library preparation and identification methods. Next-generation sequencing technologies created another way of developing microsatellite markers without cloning and heavy manipulation tasks. RNA-sequencing (RNA-seq) based microsatellite markers are very similar to that of expressed sequence tag-microsatellites (EST-microsatellites). In last three-four years, EST-microsatellites have been replaced with the transcriptomic-microsatellites obtained from RNA-seq and genome sequencing studies [6,9,10]. Transcriptomic-microsatellites contain exonic regions of protein coding or non-protein coding regions of eukaryotic genes. As in EST-derived microsatellites, transcriptomic-microsatellites target transcribed regions (exons and untranslated regions) and have a higher rate of transferability across species than genomic-microsatellite markers [2,3,6,9-12].

There are several hypotheses to explain the occurrence and expansion of microsatellites in genomes. Among the hypotheses two are dominantly pronounced. *De novo* microsatellite hypothesis suggests that the birth of microsatellites is a consequence of the creation of a proto-microsatellite. Proto-microsatellite is a short region DNA consisting of 3 or 4 repeated units of simple sequences, which are defined as a scramble of repetitive motifs lacking a clear tandem arrangement. It is suggested that a proto-microsatellite is expanded by strand slippages of DNA polymerase during replication [13, 14]. The other hypothesis is the unequal crossing-over during meiosis. Replication slippage is a symmetrical process, where the same number of repeats are added and removed. This process inevitably leads to either the loss of microsatellites or the insertion of a high number of repeats (gain or loss of motifs). The misalignment that gives rise to mutations occurs between a newly synthesized DNA strand and its complementary template strand. The two strands dissociate and reanneal incorrectly, forming a loop, which is stable due to the repetitive nature of the sequence. If the loop is formed on the nascent strand, the resulting mutation will be a repeat length expansion, while loops on the template strand result in a reduction of the repeat length [13]. Unequal crossing-over during meiosis (recombination) is usually associated with the exchange of repeated units between homologous chromosomes, and therefore, plays a limited role in microsatellite mutation. However, this mechanism might be responsible for microsatellite multi-step mutations [15].

Microsatellite DNA markers are “ideal” genetic markers since they are very polymorphic and multi-allelic, they permit classification of individuals into more than two groups [3,16-18]. Most microsatellite markers are co-dominant to enable discrimination between heterozygous and homozygous state [2,19]. Microsatellites are not epistatic meaning that they have no inter-locus interactions and thus they do not interfere with other marker loci. Most microsatellites are abundant and distributed almost evenly over the entire genome. They are not pleiotropic and polymorphisms are not affected with the variations in environmental, plant developments and tissue and organ [6,9]. However, there are several limitations of microsatellite markers in genetic, phylogenetic and agro-genomic studies. Occurrence of null alleles in microsatellites is one of three main limitations. Null alleles are locus deletion in the annealing primer site preventing locus amplification. Presence of null alleles causes heterozygous identification impossible and leads to erroneous estimations of allele frequencies and segregation rates [20]. Homoplasy causes erroneous estimations of allele frequencies and segregation rates. Homoplastic alleles are those alleles that are identical in length (size) but not in sequence, or identical in length and sequence but with different evolutionary history [21]. Presence of linkage disequilibrium in microsatellites is the third limitation. Linkage disequilibrium causes deviations from the random association of alleles in a population, which may be primarily caused by population sub-structuring and high levels of in breeding [22].

Occurrences of null alleles, homoplasy and linkage disequilibrium in genomic- and transcriptomic-microsatellite markers may differ and their presence may limit the utilization of microsatellite markers. This study used genomic- and transcriptomic-microsatellites from cotton and maize along with 472 genomic- and transcriptomic-microsatellite loci published to reveal: (i) level of polymorphism between genomic-and transcriptomic-microsatellites; (ii) relationship between number of alleles and polymorphism; (iii) relationship between motif length and polymorphism; (iv) relationship between number of motif repeats and polymorphism, and (v) determine the fixation index values of genomic- and transcriptomic-microsatellites based on observed and expected heterozygosity.

Materials and Methods

Twenty cotton samples consisting *Gossypium hirsutum* L. and *G. herbaceum* L. cultivars, and 8 maize (*Zea mays* L.) commercial F_1 hybrid varieties were used as plant materials. Genomic DNA from single seed from each plant sample was extracted using a protocol described in Karaca et al. [23]. Four cotton transcriptomic and genomic primer pairs [17], and 4 maize transcriptomic and genomic primer pairs (https://www.maizegdb.org/data.center/ssr) were utilized in this study.

Polymerase chain reactions (PCRs) were carried out in a 25 µL reaction mixture containing 1X PCR assay buffer (50 mM KCl, 10 mM TRIS–Cl, 2.5 mM $MgCl_2$), 0.2 mM each of dNTPs, 0.5 µM each of forward and reverse primers, 1 unit of *Taq* DNA polymerase and 80-100 ng of genomic DNA template. PCR was carried out in a thermal cycler (GeneAmp PCR System 9700) with the following amplification profile: 5 min hold at 95 ^oC, followed by a 10 cycle pre-PCR consisting of 20 s at 95 ^oC for denaturation, 30 s at 60 ^oC for annealing, and 1 min at 72 ^oC for extension [24]. Annealing temperatures were reduced 0.5 ^oC each cycle. PCR was continued for 30 more cycles at a 55 ^oC annealing temperature with a final extension for 10 min at 72 ^oC [25, 26].

The amplified products were separated by electrophoresis in 3% high resolution agarose gels (Agarose SERVA) containing 0.05 µg/mL ethidium bromide. The size of the amplified fragments (alleles) was determined using DNA size standards. DNA fragments were visualized under UV light in a gel documentation system [25]. Amplified bands (alleles) were co-dominantly scored. Scores for microsatellite amplicons were given approximate base pairs. Descriptive statistics including frequency of allele, observed and expected heterozygosity and inbreeding coefficient or Fixation Index (F) were calculated with GenAlEx 6.5 [27, 28].

Frequency of allele (Fa):	$\frac{2N_{xx} + N_{xy}}{2N}$	Where: Nxx is the number of homozygotes for allele x (xx), and Nxy is the number of heterozygotes containing the x allele (y can be any other allele). N is the number of sample.
Observed heterozygosity (Ho):	$\frac{N_{xy}}{N}$	Where: Nxy is the number of heterozygotes, N is the number of sample.
Expected heterozygosity (He):	$= 1 - \sum_{i=1}^{n} p_i^2$	Where: pi is the allele frequency of the i-th allele and n is the total number of allele at a locus.
Inbreeding coefficient or Fixation index (F):	$\frac{He - Ho}{He}$	Where: He is expected heterozygosity and Ho is observed heterozygosity.

Polymorphism information content (PIC) values using the equation below were calculated using Excel software [29].

$$\boldsymbol{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2\, p_i^2 p_j^2 = 1 - \sum_{i=1}^{n} p_i^2 - (\sum_{i=1}^{n} p_i^2)^2 + \sum_{i=1}^{n} p_i^4,$$

where pi equals the frequency of the ith allele and pj the frequency of the (i+1) allele.

In addition to plant materials used in the present study, 472 loci of microsatellites from different plant species are used (Table 1). Plant materials used in the present study included genomic and EST-derived and transcriptomic-microsatellites from sage, oil palm, castor, *Allium* spp., *Garcinia* spp., chickpea, wild and cultivated barley, olive, *Philcoxia*, *Jatropa*, *Elymus*, *Zea* and *Zanthoxylum* samples.

Simple correlation analyses were performed between PIC and number of alleles (Na), between PIC and observed heterozygosity (Ho), between PIC and expected heterozygosity (He), to reveal any positive or negative relationships between these statistics. Finally, scatter pilots between PIC and repeating motifs and between PIC and motif length were drawn to better visualize the relationships.

Table 1: Source and types of microsatellite locus data used in the study

Type	# Loci	#Sample and type	Source	Reference
Genomic-microsatellites	9	50 sage samples from natural populations	*Salvia officinalis* L.	[30]
	8	15 oil palm varieties	*Elaeis guineensis* Jacq.	[31]
	8	38 castor genotypes	*Ricinus communis* L.	[32]
	22	32 date palm cultivar	*Phoenix dactylifera* L.	[33]
	11	60 samples from two *Allium* spp.	*Allium mongolicum* Regel	[34]
	23	30 *Garcinia* genotypes	*Garcinia gummi-gutta* (L.) Roxb.	[35]
	9	23 chickpea accessions	*Cicer arietinum* L.	[36]
	10	189 adult walnut trees	*Juglans regia* L.	[37]
	48	41 breadfruit samples	*Artocarpus altilis*, *A. camansi* and *A. heterophyllus*	[38]
	5	30 *Philcoxia* samples	*Philcoxia minensis* Taylor & V. C. Souza	[39]
Transcriptomic-microsatellites	20	96 wild and cultivated barley accessions	*Hordeum vulgare* L.	[40]
	46	24 olive varieties	*Olea europaea* L.	[41]
	49	96 wild and cultivated barley accessions	*Hordeum vulgare* L.	[40]
	28	22 sweet corn	*Zea mays* L.	[42]
	112	15 *Elymus sibiricus* accessions	*Elymus sibiricus* L.	[43]
	13	25 *Arnica* samples	*Arnica montana*	[44]
	17	370 *Citrus* accessions	*Citrus* spp.	[45]
	19	25 *Jatropa* accessions	*Jatropa curcas* L.	[46]
	15	102 pickly ash samples	*Zanthoxylum* spp.	[47]

Results and Discussion

This study consisted of laboratory experiments and *in silico* studies. In the laboratory studies two different plant species used were different in their ploidy levels. Maize is a diploid while cotton is an allotetraploid species. It is well known that tetraploid cotton species behaves as diploid species during meiosis [2]. In the present study, genomic- and transcriptomic-microsatellite markers from two species were scored as diploid using co-dominant scoring method [16, 28]. Among 4 cotton loci, two transcriptomic-microsatellites were polymorphic in twenty samples while genomic-microsatellites were non polymorphic. On the other hand, 4 maize microsatellite loci used three produced polymorphic markers within maize hybrid cultivars. One of the two transcriptomic-microsatellite loci was not polymorphic among maize hybrid cultivars. A total of 5 loci were polymorphic and their descriptive statistics were used in further studies.

There are two commonly used indices, or measures for the polymorphism degree definition or evaluation. One of the indices is heterozygosity usually referred as H, which could be divided into expected heterozygosity (He) and observed heterozygosity (Ho). In the present study, we compared Ho and He for genomic- and transcriptomic-microsatellites. Ho and He are used to calculate fixation index values. The other index is polymorphism information content (PIC) which shows discrimination power of markers usually generated with primer or primer pairs or restriction enzymes. PIC also indicates the population polymorphism depending on the number and frequency of the alleles [29].

Number of alleles per cotton and maize locus, observed heterozygosity, expected heterozygosity and PIC along with fixation index values were calculated using GenAlEx software version 6.5 [28]. The number of alleles for polymorphic loci ranged from 3 to 8, with an average value of 4.6 alleles per locus. Observed heterozygosity

values ranged from 0.118 to 0.437, with an average value of 0.235 alleles per locus. Expected heterozygosity values ranged from 0.213 to 0.532, with an average value of 0.423 alleles per locus. PIC values of cotton and maize loci ranged from 0.22 to 0.645, with an average of 0.388 alleles per locus. It was obvious that there were null alleles and homoplasy in cotton data due to presence of *G. herbaceum* samples. We also found that one of cotton transcriptomic-microsatellites were not at Hardy–Weinberg equilibrium [20-22] calculated using GenAlEx software version 6.5 [28].

Comparison of PIC values between genomic- and transcriptomic-microsatellite polymorphism levels in cotton and maize revealed that there were some levels of PIC value differences between the two sources of microsatellites. PIC values of transcriptomic-microsatellites were lower than that of genomic-microsatellites. However, the number of alleles and loci obtained from laboratory experiments were very small for a better scientific conclusion. Therefore, this study used additional alleles and loci from different plant species presented in Table 1.

The observed heterozygosity (Ho) is the part of heterozygous genes in a population. Ho is calculated for each locus as the total number of heterozygotes divided by the sample size. Among 477 loci (5 from laboratory experiments and 472 from *in silico* studies) used in this study, Ho values ranged from 0.11 to 0.60, with an average value of 0.416 alleles per locus. The highest Ho value was obtained from di-nucleotides of transcriptomic-microsatellites (0.60) while the lowest value of Ho (0.11) was obtained from penta-nucleotides of transcriptomic-microsatellites. Analyses based on 477 loci indicated that the values of Ho decreased as the motif length increased from di- to penta-nucleotides while Ho values of hexa-nucleotides from transcriptomic microsatellites were not paralleled with this conclusion (Table 2).

Expected heterozygosity (He) is often used to describe levels of genetic diversity because He depends solely on the number and frequencies of alleles. He is less sensitive to the sample size than the observed heterozygosity. This makes He more attractive in population studies. Among 477 loci used in this study, He values ranged from 0.36 to 0.74, with an average value of 0.57 alleles per locus. The highest He value of 0.74 was obtained from di-nucleotides of genomic-microsatellites. Lowest value (0.36) of He was obtained from penta-nucleotides of transcriptomic-microsatellites. Results indicated that He values decreased as the motif length increased from di- to penta-nucleotides. However, while He values of hexa-nucleotides from transcriptomic microsatellites were not paralleled with this conclusion (Table 2). Hexa-nucleotide motifs are related with the tri-nucleotide repeats and called three-fold repeats. Similar Ho and He descriptive statistics was obtained.

Values of Ho and He were used to calculate Fixation Index (F) values which ranged from -0.01 to 0.69. Generally F values close to zero are expected under random mating while F values close to +1 indicate inbreeding or undetected alleles. F values close to -1 indicate excess of heterozygosity due to negative assortative mating, or selection for heterozygotes [28]. Penta-nucleotides of transcriptomic-microsatellites produced the highest F value (0.69) followed with hexa-nucleotides of transcriptomic-microsatellites (0.61) and di-nucleotides of genomic-microsatellites (0.44). On the other hand, the lowest F value (-0.01) was found in di-nucleotides of transcriptomic-microsatellites, followed with tetra-nucleotides (0.08) and tri-nucleotides (0.12) of transcriptomic-microsatellites. F values of genomic- and transcriptomic-microsatellites were different. Results of present study indicated that alleles of genomic- and transcriptomic-microsatellites should be combined for population structure and characterization studies in order to obtain reliable characteristics of populations.

Polymorphism information content (PIC) values of genomic- and transcriptomic-microsatellites ranged from 0.41 to 0.67, with an average value of 0.53 per locus. In the study, microsatellites of tri-nucleotide repeats from genomic DNA showed the highest value of PIC (0.67). The PIC values of genomic-microsatellite markers (0.65±0.24) were generally higher than that of transcriptomic-microsatellite markers (0.48±0.16). Within the transcriptomic-microsatellites, di-nucleotide repeats showed the highest PIC value (0.50) while tri-nucleotide repeats of genomic-microsatellites showed the highest PIC value (0.67). This indicated that di-nucleotide transcriptomic-microsatellites and tri-nucleotides of genomic-microsatellitesare much powerful in genetic polymorphism studies.

Correlation analyses between descriptive statistical values, PIC and F values along with their averaged values and standard deviations were presented in Table 2. Compound type microsatellite from genomic and transcriptomic sources were excluded from analyses because the number of compound microsatellites of

genomic- and transcriptomic sources were quite different. When larger number of compound microsatellites from genomic- and transcriptomic-microsatellites are studied, much reliable results could be obtained.

Numbers of genomic-microsatellite motif length ranged from two to five. However, the number of observations in tetra- and penta-nucleotides was very low in genomic-microsatellites. There was only one observation for tetra-nucleotides and three observations for penta-nucleotides. The highest number of repeats were found in di-nucleotides (109) followed with tri-nucleotides (64). Unfortunately, the number of loci and motif lengths among genomic- and transcriptomic-microsatellite data were quite different. Transcriptomic-microsatellites contained di- to hexa-nucleotide motif lengths. Tri-nucleotide microsatellites were the most abundant (160) followed with di-nucleotides (70) while hexa-nucleotides were the lowest motif length (12) followed with penta-nucleotides in transcriptomic-microsatellites (Figure 1 and Table 2).

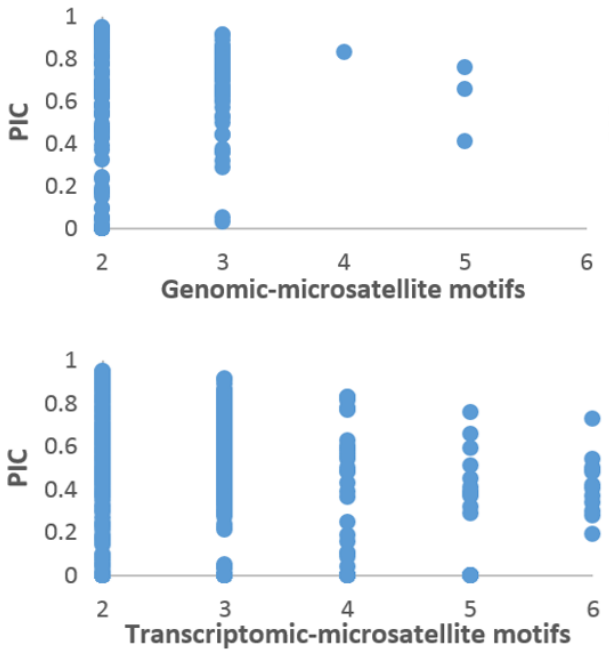

Figure 1: Scatter pilots between PIC and motifs in genomic- and transcriptomic-microsatellites

Table 2: Descriptive statistics, PIC, fixation index (F) and correlation analyses of genomic- and transcriptomic microsatellites

Type	Motif	N	Rpt (SD)	Mean Values (SD)				Fixation Index (F)	Correlations			
				Na	Ho	He	PIC		PIC-Rpt	PIC-Na	PIC-Ho	PIC-He
Genomic-microsatellites	Di-	109	14.13 (±6.49)	9.13 (±6.95)	0.41 (±0.29)	0.74 (±0.20)	0.63 (±0.25)	0.44	-0.23	0.67	-0.04	0.62
Genomic-microsatellites	Tri-	64	13.86 (±8.34)	8.03 (±4.57)	0.57 (±0.28)	0.70 (±0.17)	0.67 (±0.19)	0.18	0.41	0.61	0.06	0.80
Genomic-microsatellites	All	173	13.91 (±7.19)	8.65 (±6.12)	0.52 (±0.28)	0.73 (±0.20)	0.65 (±0.24)	0.28	0.01	0.64	0.00	0.64
Transcriptomic-microsatellites	Di-	70	12.41 (±7.08)	4.31 (±2.02)	0.60 (±0.38)	0.59 (±0.16)	0.50 (±0.18)	-0.01	0.10	0.76	0.28	0.85
Transcriptomic-microsatellites	Tri-	160	6.19 (±1.51)	4.27 (±1.56)	0.50 (±0.40)	0.57 (±0.13)	0.49 (±0.14)	0.12	0.20	0.70	0.23	0.91
Transcriptomic-microsatellites	Tetra-	39	6.90 (±3.12)	4.26 (±2.75)	0.38 (±0.27)	0.42 (±0.20)	0.48 (±0.25)	0.08	0.67	0.76	0.68	0.84
Transcriptomic-microsatellites	Penta-	18	4.11 (±0.32)	3.22 (±1.40)	0.11 (±0.10)	0.36 (±0.14)	0.41 (±0.09)	0.69	-0.45	0.73	0.11	0.77
Transcriptomic-microsatellites	Hexa-	12	5.08 (±1.93)	4.17 (±1.47)	0.17 (±0.13)	0.45 (±0.20)	0.42 (±0.14)	0.61	-0.04	0.82	-0.21	0.94
Transcriptomic-microsatellites	All	299	7.57 (±4.66)	4.21 (±1.86)	0.48 (±0.38)	0.54 (±0.16)	0.48 (±0.16)	0.12	0.18	0.73	0.32	0.84

Motif: the length of repeats unit, Rpt: repeat numbers, N: Number of individuals analyzed, Na: number of alleles, He: expected heterozygosity; H: heterozygosity, Ho: observed heterozygosity, PIC: polymorphism information content, SD: standard deviation

Results revealed that there existed an inverse relationship between motif length and PIC with a negative correlation value of -0.174. Also it is noteworthy that the relationship between motif length and PIC might be microsatellite class specific as correlation was more consistent for some microsatellite types (Table 2). However, larger data set with well-distributed motif lengths are required to confirm and verify the observed negative correlation between motif length and PIC values.

In the present study, we also studied relationship between PIC and number of motif repeats (Figure 2, Table 2). Motif repeats of transcriptomic-microsatellites mainly consisted of repeating motifs less than 12 while genomic-microsatellites contained less than 20 repeating motifs. It appeared that transcriptomic-microsatellites with higher number of repeats have lower level of PIC values. More than half of transcriptomic-microsatellites with 5 to 15 repeats had PIC value of 0.5. On the other hand, majority of genomic-microsatellites with 10 to 20 repeats had PIC value of 0.5 (Figure 2, Table 2).

Based on the combined genomic- and transcriptomic-microsatellite data the scatter plots between PIC values and number of repeats (repeating motifs) indicate that there existed no strong correlation between the PIC and number of repeats (Figure 2). Also overall correlation between the repeats and PIC values was weak with correlation value of 0.244. However, we detected a moderate correlation (0.67) between number of repeats and PIC in transcriptomic-microsatellites consisting of tetra-nucleotide repeats (Table 2).

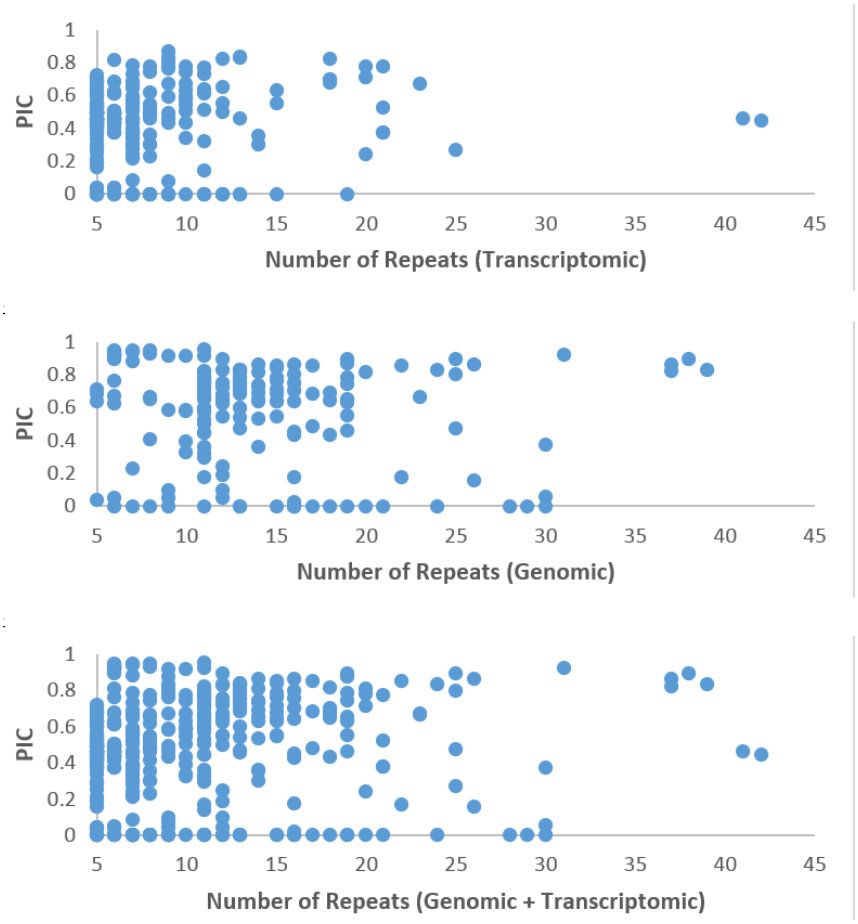

Figure 2: Scatter pilots between PIC and number of repeats in genomic-, transcriptomic-microsatellites and combined data

The number of microsatellite alleles in genomic- and transcriptomic-microsatellites was different. To understand the possible relationship between polymorphism of genomic- and transcriptomic-microsatellite markers, scatter

plots were made between number of alleles detected and the PIC values. The scatter plot between number of alleles (3-10) and PIC showed the widest variations. This indicated that there was a low level of correlation between PIC values and 3-10 alleles. As presented in Figure 3 and Table 2 we found that there were considerable degree of relationship between the allele number greater than 10 and PIC values. These findings indicated that markers with higher number of alleles potentially produce higher level of PIC.

Correlation analyses between expected heterozygosity and PIC revealed close relationships between these two entities. In the present study, the highest correlation value between He and PIC was found in hexa-nucleotides of transcriptomic-microsatellites while the lowest correlation between He and PIC was found in di-nucleotides of genomic-microsatellites. In general, higher correlation values between expected heterozygosity and polymorphism were found in transcriptomic-microsatellites.

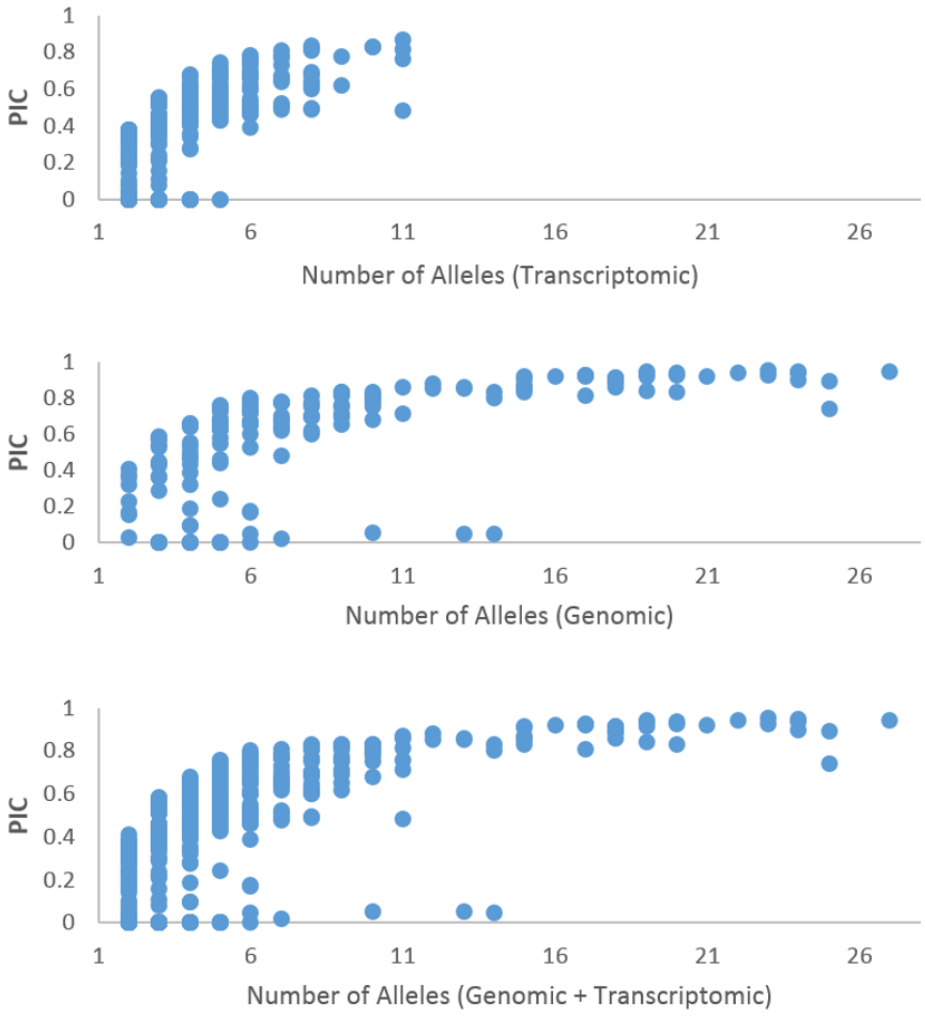

Figure 3: Scatter pilots between PIC and number of alleles in genomic-, transcriptomic-microsatellites and combined data

Correlation analyses between observed heterozygosity and PIC indicated that these two entities were not related with an exception of tetra-nucleotides of transcriptomic-microsatellites. Observed heterozygosity values and PIC values were correlated in tetra-nucleotides of transcriptomic-microsatellites. Low correlation values between observed heterozygosity and PIC values indicated that observed heterozygosity was dependent upon sample size and could not be used in identification of informative microsatellite markers.

4. Conclusions

Present study used 5 loci from cotton and maize and 472 microsatellite loci from other plant species presented in Table 1. Results clearly indicated that polymorphic information content values of genomic-microsatellites were

higher than transcriptomic-microsatellites. Analysis revealed that PIC values of genomic- and transcriptomic-microsatellites increased as the number of allele increased. The number of alleles was higher in genomic-microsatellites than transcriptomic-microsatellites. Although it is expected that loci with longer motifs and more number of repeats are much more likely to be more variable than shorter motif and fewer number of repeats due to faster mutation rate, we could not confirm any relationship between numbers of repeats and polymorphism, and between the motif length and polymorphism.

Results clearly showed that there were differences between expected heterozygosity and observed heterozygosity values of genomic- and transcriptomic-microsatellites. Based on five loci used in cotton and maize, results also revealed that genomic- and transcriptomic-microsatellite markers produce more null alleles. We noted that cotton and maize differed in the level of homoplasy. Based on the descriptive statistics we suggested that functional markers obtained from transcriptomic-microsatellites might not be suitable for phylogenetic and evolutions studies but they could be very useful in genetic fingerprinting and agro-genomic studies.

References

[1]. Litt, M., & Luty, J. A. (1989). A Hypervariable Microsatellite Revealed by *In Vitro* Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. *The American Journal of Human Genetics*, 44: 397-401.

[2]. Karaca, M. (2001). Characterization of Cynodon spp. and Gossypium spp. Genomes Using Molecular and Cytological Techniques. Ph.D. Dissertation [DAI, 62, no. 05B (2001): p. 2119 ISBN: 0-493-26105-2]. Mississippi State University, Mississippi State, MS.

[3]. Karaca, M., Bilgen, M., Onus, A. N., Ince, A. G., & Elmasulu, S. Y. (2005). Exact Tandem Reoeats Analyzer (E-TRA): A New Program for DNA Sequence Mining. *Journal of Genetics*, 84: 49-54.

[4]. Ince, A. G., Karaca, M., & Onus A. N. (2010). Differential Expression Patterns of Genes Containing Microsatellites in *Capsicum annuum* L. *Molecular Breeding*, 25: 645–658.

[5]. Ashley, M. V. (2010). Plant Parentage, Pollination, and Dispersal: How DNA Microsatellites Have Altered the and Scape. *Critical Reviews in Plant Sciences*, 29: 148-161.

[6]. El Mujtar, V., Gallo, L., Lang, T., Garnier-Gere, P. (2014). Development of Genomic Resources for Nothofagus Species Using Next-Generation Sequencing Data. *Molecular Ecology Resources*, 14: 1281-1295.

[7]. Ince, A. G., Karaca, M., Bilgen, M., & 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.

[8]. Azpilicueta, M. M., Gallo, L. A., van Zonneveld, M., Thomas, E., Moreno, C., & Marchelli, P. (2013). Management of Nothofagus Genetic Resources: Definition of Genetic Zones Based on a Combination of Nuclear and Chloroplast Marker Data. *Forest Ecology and Management*, 302: 414-424.

[9]. Dufresnes, C., Brelsford, A., Beziers, P., & Perrin, N. (2014). Stronger Transferability but Lower Variability in Transcriptomic- than in Anonymous Microsatellites: Evidence From Hylid Frogs. *Molecular Ecology Resources*, 14: 716-725.

[10]. Vukosavljev, M., Esselink, G. D., van 't Westende, W. P. C., Cox, P., Visser, R. G. F., Arens, P., & Smulders, M. J. M. (2015). Efficient Development of Highly Polymorphic Microsatellite Markers Based on Polymorphic Repeats in Transcriptome Sequences of Multiple İndividuals. *Molecular Ecology Resources*, 15: 17-27.

[11]. Marchelli, P., Gallo, L., Scholz, S., & Ziegenhagen, B. (1998). Chloroplast DNA Markers Reveal a Geographical Divide Across Argentinean Southern Beech *Nothofagus nervosa* (Phil.) Dim. et Mil. Distribution Area. *TAG Theoretical and Applied Genetics*, 97: 642-646.

[12]. Karaca, M., Ince, A. G., Aydin, A., & Ay, S. T. (2013) Cross-Genera Transferable e-microsatellite Markers for 12 Genera of *Lamiaceae* family. *Journal of Science Food and Agriculture*, 93: 1869-1879.

[13]. Vieira, M. L. C., Santini, L., Diniz, A. L., &Munhoz, C. F. (2016). Microsatellite Markers: What They Mean and Why They are so useful. *Genetics and Molecular Biology*, 39: 312-328.

[14]. Bhargava, A., & Fuentes, F. F. (2010). Mutational Dynamics of Microsatellites. *Molecular Biotechnology*, 44: 250-266.
[15]. Grover A., & Sharma, P. C. (2011). Is Spatial Occurrence of Microsatellites in the Genome a Determinant of Their Function and Dynamics Contributing to Genome Evolution? *Current Science*, 100: 859-869.
[16]. Ince, A. G., Karaca, M., & Onus, A. N. (2010). Genetic Relationships Within and Between Capsicum Species. *Biochemical Genetics*, 48: 83-95.
[17]. Karaca, M., & Ince, A. G. (2011). New Non-Redundant Microsatellite and CAPS-Microsatellite Markers for Cotton (*Gossypium* L.). *Turkish Journal of Field Crops*, 16: 172-178.
[18]. Ince, A. G., Karaca, M., & Elmasulu, S. Y. (2014). New Microsatellite and CAPS-Microsatellite Markers for Clarifying Taxonomic and Phylogenetic Relationships Within *Origanum* L. *Molecular Breeding*, 344: 643-654.
[19]. Ince, A. G., Karaca, M., & Onus A.N. (2010). CAPS-Microsatellites: Use of CAPS Method to Convert Non-Polymorphic Microsatellites into Useful Markers. *Molecular Breeding*, 25: 491-499.
[20]. Chapuis, M. P., & Estoup, A. (2007). Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution*, 24: 621-31.
[21]. Dickey, A. M., Hall, P. M., Shatters Jr, R. G., & Mckenzie, C. L. (2013). Evolution and Homoplasy at the Bem6 Microsatellite Locus in Three Sweet Potato Whitefly *(Bemisia tabaci*) Cryptic Species. *BMC Research Notes*, 6: 24.
[22]. Kiselyova, T. Y., Kantanen, J., Vorobyov, N. I., Podoba, B. E., & Terletsky, V. P. (2014). Linkage Disequilibrium Analysis for Microsatellite Loci in Six Cattle Breeds. *Russian Journal of Genetics*, 50: 464-473.
[23]. Karaca, M., Ince, A. G., Elmasulu, S. Y., Onus, A. N., & Turgut, K. (2005). Coisolation of Genomic and Organelle DNAs from 15 Genera and 31 Species of Plants. *Analytical Biochemistry*, 343: 353-355.
[24]. Ince, A. G., & Karaca, M. (2009). The MAGi RNA Extraction Method: Highly Efficient and Simple Procedure for Fresh and Dry Plant Tissues. *Journal of the Science of Food and Agriculture*, 89: 168-176.
[25]. Karaca, M., Ince, A. G., Ay, S. T., Turgut, K., & Onus, A. N. (2008). PCR-RFLP and DAMD-PCR Genotyping for *Salvia* Species. *Journal of Science Food and Agriculture*, 88: 2508-2516.
[26]. Ince, A. G., & Karaca, M. (2011). Genetic Variation in Common Bean Landraces Efficiently Revealed by Td-DAMD-PCR Markers. *Plant Omics Journal*, 4: 220-227.
[27]. Peakall, R., & Smouse, P. E. (2006). GenAlEx 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Molecular Ecology Notes*, 6: 288-295.
[28]. Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research–an Update. *Bioinformatics*, 28: 2537-2539.
[29]. Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics*, 32: 314-331.
[30]. Radosavljevic, I., Satovic, Z., Jakse, J., Javornik, B., Greguras, D., Jug-Dujakovic, M., & Liber, Z. (2012). Development of New Microsatellite Markers for *Salvia officinalis* L. and Its Potential Use in Conservation-Genetic Studies of Narrow Endemic *Salvia brachyodon* Vandas. *International Journal of Molecular Science*, 13: 12082-12093.
[31]. Okoye, M. N., Uguru, M. I., Bakoume, C., Singh, R., & Okwuagwu, C. O. (2016). Assessment of Genetic Diversity of NIFOR Oil Palm Main Breeding Parent Genotypes Using Microsatellite Markers. *American Journal of Plant Sciences*, 7: 218-237.
[32]. Bajay, M. M., Pinheiro, J. B., Batista, C. E. A., Nobrega, M. B. M., & Zucchi, M. I. (2009). Development and Characterization of Microsatellite Markers for Castor (*Ricinus communis* L.), an Important Oleaginous Species for Biodiesel Production. *Conservation Genetics Resources*, 1: 237-239.
[33]. Al-Faifi, S. A., Migdadi, H. M., Algamdi, S. S., Khan, M. A., Ammar, M. H., Al-Obeed, R. S., Al-Thamra, M. I., EL-Harty, E. H., & Jakse, J. (2016). Development, Characterization and Use of

Genomic SSR Markers for Assessment of Genetic Diversity in Some Saudi Date Palm (*Phoenix dactylifera* L.) Cultivars. *Electronic Journal of Biotechnology*, 21: 18-25.

[34]. Zhang, T-Y., Chen H-K., & Zhang, C-R. (2014). Polymorphic Microsatellite Markers for *Allium mongolicum* Regel (*Amaryllidaceae*). *Genes & Genetic Systems*, 89: 133-136.

[35]. Ravishankar, K.V., Vasudeva, R., Hemanth, B., Sandya, B. S., Sthapit, B. R., Parthasarathy, V. A., & Rao, V. R. (2017). Isolation and Characterization of Microsatellite Markers in Garcinia gummi-gutta by Next-Generation Sequencing and Cross-Species Amplification. *Journal of Genetics*, 96: 2013-2018.

[36]. Torutaeva, E., Asanaliev, A., Prieto-Linde, M. L., Zborowska, A., Ortiz, R., Bryngelsson, T., & Garkava-Gustavsson, L. (2014). Evaluation of Microsatellite-Based Genetic Diversity, Protein and Mineral Content in Chickpea Accessions Grown in Kyrgyzstan. *Hereditas*, 151: 81-90.

[37]. Ebrahimi, A., Zarei, A., Lawson, S., Woeste, K. E., & Smulders, M. J. M. (2016). Genetic Diversity and Genetic Structure of Persian Walnut (*Juglans regia*) Accessions from 14 European, African, and Asian Countries Using SSR Markers. *Tree Genetics & Genomes*, 12: 114.

[38]. Bellis, F. D., Malapa, R., Kagy, V., Lebegin, S., Billot, C., & Labouisse, J. P. (2016). New Development and Validation of 50 SSR Markers in Breadfruit (*Artocarpus altilis*, *Moraceae*) by Next-Generation Sequencing. *Applications in Plant Sciences*, 4: 1600021.

[39]. Scatigna, A. V., Oliveira, F. A., Mantello, C. C., Francisco, P. M., Souza, A. P., & Simoes, A. O. (2015). Microsatellite Markers for Studies with the Carnivorous Plant *Philcoxia minensis* (*Plantaginaceae*). *Applications in Plant Sciences*, 3: 1500035.

[40]. Zhang, M., Mao, W., Zhang, G., & Wu, F. (2014). Development and Characterization of Polymorphic EST-SSR and Genomic SSR Markers for Tibetan Annual Wild Barley. *PLOS ONE*, journal.pone: 0094881.t004.

[41]. Arbeiter, A. B., Hladnik, M., Jakse, J., & Bandelj, D. (2016). Identification and Validation of Novel EST-SSR Markers in Olives. *Scientia Agricola*, 74: 215-225.

[42]. Lopes A. D., Scapim, C. A., Machado, M. F. P. S., Mangolin, C. A., Silva1, T. A., Cantagali1, L. B., Teixeira, F. F., & Mora, F. (2015). Genetic Diversity Assessed by Microsatellite Markers in Sweet Corn Cultivars. *Scientia Agricola*, 72: 513-519.

[43]. Zhou, Q., Luo, D., Ma, L., Xie, W., Wang, Y., Wang, Y., & Liu, Z. (2016). Development and Cross-Species Transferability of EST-SSR Markers in Siberian wildrye (*Elymus sibiricus* L.) Using Illumina sequencing. *Scientific Reports*, 6: 1-10.

[44]. Duwe, V. K., Ismail, S.A., Buser, A., Sossai, E., Borsch, T., & Muller, L. A. H. (2015). Fourteen Polymorphic Microsatellite Markers for the Threatened *Arnica montana* (*Asteraceae*). *Applications in Plant Sciences*, 3: 1400091.

[45]. Cristofani-Yaly, M., Novelli, V. M., Bastianel, M., & Machado, M. A. (2011). Transferability and Level of Heterozygosity of Microsatellite Markers in Citrus Species. *Plant Molecular Biology Reporter*, 29: 418-423.

[46]. Yadav, H. K., Ranjan, A., Asif, M. H., Mantri, S., Sawant, S. V., & Tuli, R. (2011). EST-Derived SSR Markers in *Jatropha curcas* L.: Development, Characterization, Polymorphism, and Transferability Across the Species/Genera. *Tree Genetics & Genomes*, 7: 207-219.

[47]. Kim, Y. M., Jo, A., Jeong, J. H., Kwon, Y. R., & Kim, H. B. (2017). Development and Characterization of Microsatellite Primers for *Zanthoxylum schinifolium* (*Rutaceae*). *Applications in Plant Sciences*, 5: 1600145.