IJAS Iranian Journal of Applied Animal Science

Analyzing Simple Sequence Repeats Derived from Expressed Sequence Tags in Dromedary Camels

Research Article

A. Barazandeh[1*], M. Mokhtari[1], M. Moghbeli Damaneh[1] and Z. Roudbari[1]

[1] Department of Animal Science, Faculty of Agriculture, University of Jiroft, Jiroft, Iran

Received on: 11 Dec 2018
Revised on: 26 Mar 2019
Accepted on: 31 Mar 2019
Online Published on: Sep 2020

*Correspondence E-mail: Abarazandeh@ujiroft.ac.ir

Online version is available on: www.ijas.ir

ABSTRACT

The objective of the current research was to make a character description of simple sequence repeats (SSR) derived from expressed sequence TAGs (EST) markers of dromedary camels (EST-SSR markers) and to conduct a practical analysis of these sequences for their application in comparative genomics and molecular genetics studies. A complete of 862 SSRs were discovered from 17155 EST sequences using the SSR Locator software. 827 EST out of 17155 EST had SSRs, that 794 (96%), 31 (3.8%) and 2 (0.2%) of them contained 1, 2 and 3 SSRs, respectively. The dimeric motifs were the most abundant SSRs (38.86%), followed by 27.15%, 21.46%, 6.96%, and 5.57% for tri-, hexa-, tetra- and pentameric motifs. The most plentiful dimer, trimer, tetramer, pentamer and hexamer motif were AC/TG (54%), GCC/GGC (19.2%), TTTA (13.3%), AAAAG (10.4%) and AACCAC (67.6 %), respectively. BLASTX was used to examine the final non-redundant EST-SSRs. Almost all of EST-SSRs were found out to be protected in the macromolecule catabolic process and RNA processing and splicing. EST-SSR markers might be applied as a novel resource of useful markers in the biological survey. Also, these markers may be a valuable source for further molecular genetics and genomics research of camels and related species.

KEY WORDS camel, EST, gene ontology, genomic, molecular marker, SSR.

INTRODUCTION

Camels are important livestock animals for transport and also as supply sources of milk and meat. Nowadays, interests in camels breeding are promoting for human nutrition and production of modern therapeutics (Nguyen *et al.* 2000; Kastelic *et al.* 2009; Jirimutu *et al.* 2012; Wu *et al.* 2014; Altaher and Kandeel, 2015). Several DNA-based marker systems have been expanded, containing 'restriction fragment length polymorphism' (RFLP) (Saiki *et al.* 1985), 'random amplified polymorphic DNA' (RAPD) (Mbwana *et al.* 2006), 'inter-simple sequence repeats' (ISSRs) (Reddy *et al.* 2002), 'simple sequence repeats' (SSRs) (Peakall *et al.* 1998), 'amplified fragment length polymorphism' (AFLP) (Vekemans *et al.* 2002), and their variants, to monitoring genetic variability, genome study, molecular breeding and 'marker-assisted selection' (MAS) in varied species (Joshi *et al.* 2010; Bakhtiarizadeh *et al.* 2012; Asadi and Rashidi Monfared, 2014). In comparison with the other type of genetic markers, SSRs (microsatellites) are uniquely determined by co-dominant inheritance, multi-allelic matter, high reliability, frequent in the genome, high polymorphism, and high percentage of cross-species transferability (Yan *et al.* 2008; Kumar *et al.* 2015; Sadder *et al.* 2015; Nirapathpongporn *et al.* 2016; Du *et al.* 2017). Thus SSRs markers are applied for assessing the genetic variability, protection of species, genetic mapping, 'marker-assisted selection', and supplying a valuable tool for conducting a

connection among morphologic and genetic changes (Kim *et al.* 2008; Asadi and Rashidi Monfared, 2014; Wang *et al.* 2017; Cai *et al.* 2019). Meaningful advances have been made for developing more valuable approaches for achieving the new SSRs (Zane *et al.* 2002; Kim *et al.* 2008; Vieira *et al.* 2016; Wang *et al.* 2017; Taheri *et al.* 2018), but isolation of these markers remained costly, labour -intensive and time-consuming (Yan *et al.* 2008; Wang *et al.* 2015).

Expressed sequence tags (ESTs) are sequenced from segments of the coding regions of the genome under determined biological conditions (Ellis and Burke, 2007). ESTs can be extended from cDNA libraries to prepare an economical source of gene-based molecular markers (Mirkin, 2006; Kim *et al.* 2017).

The aggregation of many numbers of ESTs in a general database has led to the growth of a novel category of functional genomic markers called microsatellite markers derived from EST (EST-SSRs) that can be very quickly developed at a low cost, via data mining (Yan *et al.* 2008; Bakhtiarizadeh *et al.* 2011).

EST-SSRs markers have many advantages relative to other DNA-based genetic markers as well as identification of changes untranslated sequences (5′-UTR and 3′-UTR), introns and coding sequences, and also having a more suitable level of cross-species transferability as well a lot of conserved than SSRs markers (Li *et al.* 2004a; Guzinski *et al.* 2016). Due to this fact that EST-SSR genomic markers, connection with coding sequences might also lead to tagging genes directly for quantitative trait locus (QTL) mapping of vital traits (Asadi and Rashidi Monfared, 2014; Zhou *et al.* 2016). Presently, with the development of genomic data, especially ESTs, the employment of bioinformatics tools lead to an increase of the discernment of EST-SSR markers in many species such as shrimp (Pérez *et al.* 2005), zebrafish (Ju *et al.* 2005), cattle (Yan *et al.* 2008), sheep (Zhang *et al.* 2010), human cancer (Bakhtiarizadeh *et al.* 2011), chicken (Bakhtiarizadeh *et al.* 2012), fish (Zheng *et al.* 2014), quail (Bai *et al.* 2016), passerine bird (Khimoun *et al.* 2017), pond loach (Feng *et al.* 2018), *Ephedra sinica* (Jiao *et al.* 2019). The purpose of the current study was to specify cluster EST-SSRs markers in camel and the term enrichment analysis of them, to measure and compare the frequency and distribution of different kinds of EST-SSRs, and to extend EST-SSR markers as genetic and genomic tools in camel.

MATERIALS AND METHODS

Collecting EST sequences

Whole dromedary camel ESTs (17155) considered in the current investigation were obtained from the website (http://camel.kacst.edu.sa/) (Al-Swailem *et al.* 2010) and were saved as FASTA format. These authors have been used Nine inbred camels from three distinct breeds (black, white and brown coat color) and three age categories (young (0-6 months), adult (2-3 years), and aged (4-6 years)). They carried out RNA isolation of the nine camels. Samples were collected and pooled of eleven tissues body of the camel (liver, heart, stomach, pancreas, muscle, brain, kidney, lung, spleen, colon and genitals).

Microsatellite mining

To find EST-SSRs has used the SSR Locator software (Maia *et al.* 2008). In this study, EST-SSRs have studied which their motifs consist of 2 to 6 nucleotides. Then, the minimum repeat pattern was selected as seven for dinucleotides, six for trinucleotide and five for other motifs including tetra-, penta-, and hexamers. All subsequent analyses were executed under R environment and Microsoft Excel. Graphs are also drawn by these softwares.

Primer design and functional annotation

Primers were designed by primer 3 in batch mode with the cooperation of the SSR Locator interface module for each EST-SSRs. For designing the primers, the sequences were considered that contained enough quantity of flanking sequence. The evaluation criteria were used: primer size 18-25 bp, with the optimum of 20 bp, primer annealing temperature 58-63 ˚C (optimum of 60 ˚C), primer GC content equal to 30%, with the optimum of 50% and product length 100-300 bp.

BLASTX (with an E value equal and/or less than 10^{-6}) were used to compare the genes containing SSRs with the non-redundant protein database for survey the function of these genes.

To identify over-represented gene ontology categories and the functional clustering of EST-SSRs were analyzed and success annotated to familiar proteins with the database for annotation visualization and integrated discovery (DAVID) bioinformatics tool (Huang *et al.* 2009). The background model with the default DAVID settings was applied to gene annotation of the whole genome.

RESULTS AND DISCUSSION

Screening of ESTs for SSRs`

Distribution of EST and EST-SSR for the camel are presented in Table 1. A complete of 862 SSRs were detected from 17155 EST sequences. From 17155 EST only 827 EST had SSRs, that 794 (96%), 31 (3.8%) and 2(0.2%) of them contained 1, 2 and 3 SSRs, respectively. The dimeric motifs were the most abundant SSRs (38.86%) in a camel, followed by 27.15%, 21.46%, 6.96%, and 5.57% for tri-, hexa-, tetra- and pentameric motifs (Figure1).

Table 1 Summary of mining expressed sequence TAGs (EST) and EST-SSRs distribution in camel

Parameter	Number
Total number of expressed sequence TAGs (ESTs)	17155
Total sequences containing simple sequence repeats (SSRs)	827
Sequences containing one SSRs	794
Sequences containing two SSRs	31
Sequences containing three SSRs	2
Total SSR-ESTs identified	862

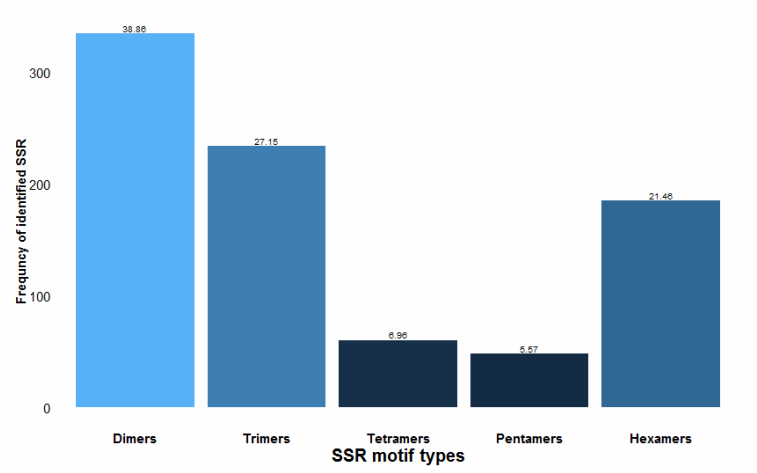

Figure 1 Frequency distribution of different microsatellite markers derived from EST (EST-SSRs) (2-6 motif unit) in camel
The numbers on the columns demonstrate the percentage of each EST-SSR

The repetitiveness of the various SSR derived from EST is shown in Figure 2 for every repeat number. The number of repeats ranged from 4 to 48. Hexamers of four repeats were the most prevalent (18.91%) and after that were trimers of six repeats (13.23%) and dimers of seven repeats (12.99%). The length of SSRs changed between 14 and 108 bp according to the length of the repeat motif (repeat number×motif length).

Distributions of camel SSRs with various repeat motifs

The observed frequencies of different repeat motifs containing the SSRs are presented in Figures 3-6. The recognized SSRs containing 4 kinds of dimer motifs, 24 kinds of the trimer, 38 kinds of the tetramer, 31 kinds of Pentamer and 47 kinds of hexamer motifs. The best frequent dimer motif was AC/TG (54%) and the AT/TA was the second plentiful kind (32.8%). Also, the GC/CG (1.2%) was the least frequent kind (Figure 3). The GCC/GGC (19.2%) was the most frequent trimer motif, followed by AGC/GCT (10.3%), CAG/CTG (9.8%) and CTC/GAG (9.8 %) (Figure 4). Most popular motifs between tetramers were TTTA (13.3%), TTTG (6.7%) and AAAC (6.7%) (Figure 5). The AAAAG (10.4%) and TTGTT (10.4%) were most popular motifs across pentamers (Figure 6). The most plentiful hexamer motif was AACCAC (67.6%), while other Hexamer motifs had almost identical frequencies.

Development of EST-SSR markers

To design pair primer all 827 sequences that included SSRs were used. 732 (88.51%) of them were ready to be accustomed to design primer pairs and 95 (11.49%) EST-SSR failed to have right flanking sequences for primers. Results of the virtual polymerase chain reaction (PCR) run shows that 597 of 732 primers made appropriate fragments.

Gene ontology analysis and annotation of EST-SSRs sequences

To examine the 827 sequences recognized as including SSRs was applied BLASTX. Incomplete annotation of camel genome caused, only 382 of 827 sequences were annotated. The GO enrichment analysis of sequences including EST-SSRs at all three levels of GO classification is shown in Table 2. The most of EST-SSRs were discover to be included in the macromolecule catabolic process and RNA processing and splicing and cellular homeostasis. Most of the EST-SSRs enriched to cellular components were dependent on the organelle, membrane-enclosed and nuclear lumen. The GO assignments for the molecular function displayed that superlative of the camel EST sequences including SSRs were involved in transcription regulator activity and RNA binding.

Functional annotation clustering determined 3 annotated classes associated with the detected genes ($P \leq 0.05$).

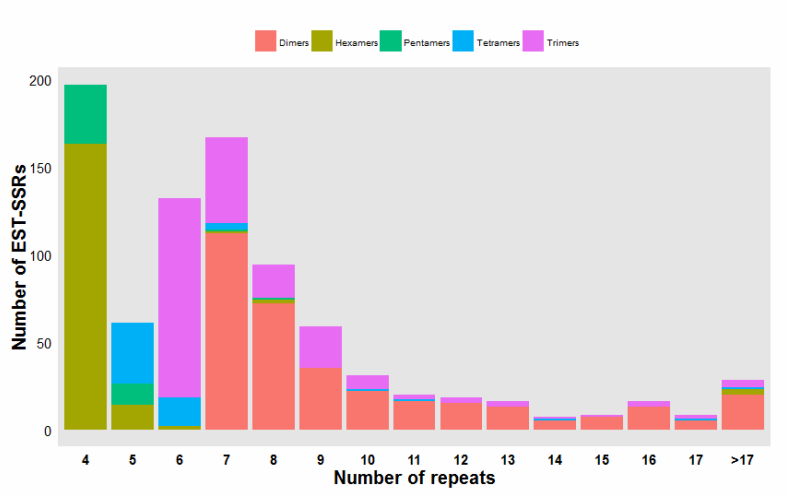

Figure 2 Frequency distribution of the microsatellite markers derived from EST (EST-SSRs) based on the number of repeats of the different SSR motif types

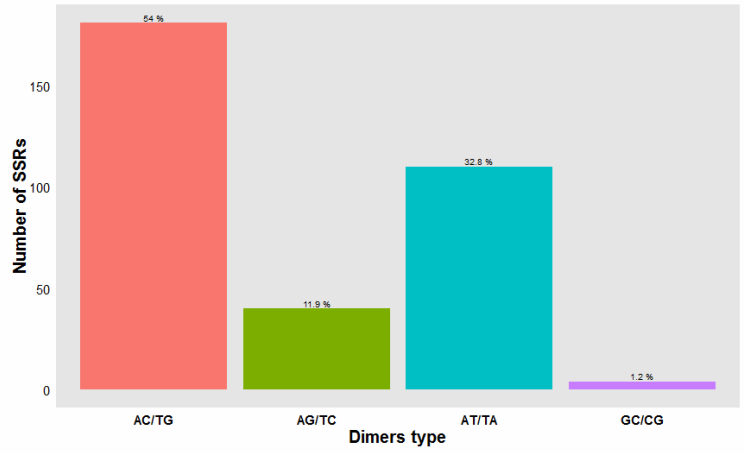

Figure 3 Frequency distribution for the 4 dimer motifs recognized in the camel sequence
The numbers on the columns demonstrate the percentage of these dimer motifs across all dimer types

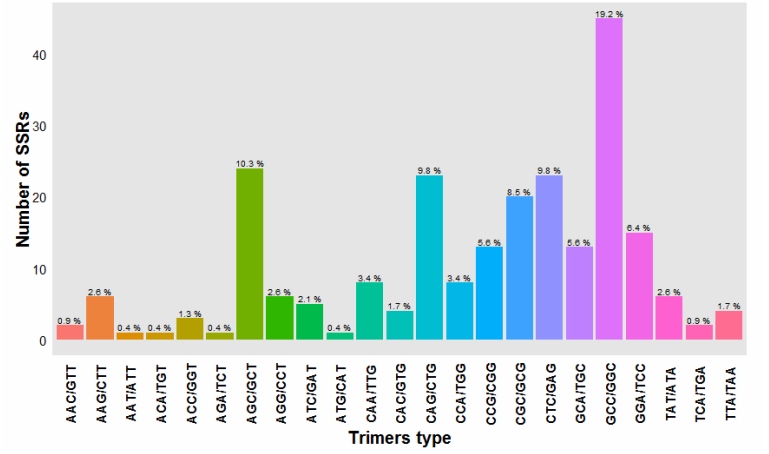

Figure 4 Frequency distribution for all 23 trimer motifs recognized in the camel sequence
The numbers on the columns demonstrate the percentage of these trimer motifs across all trimer types

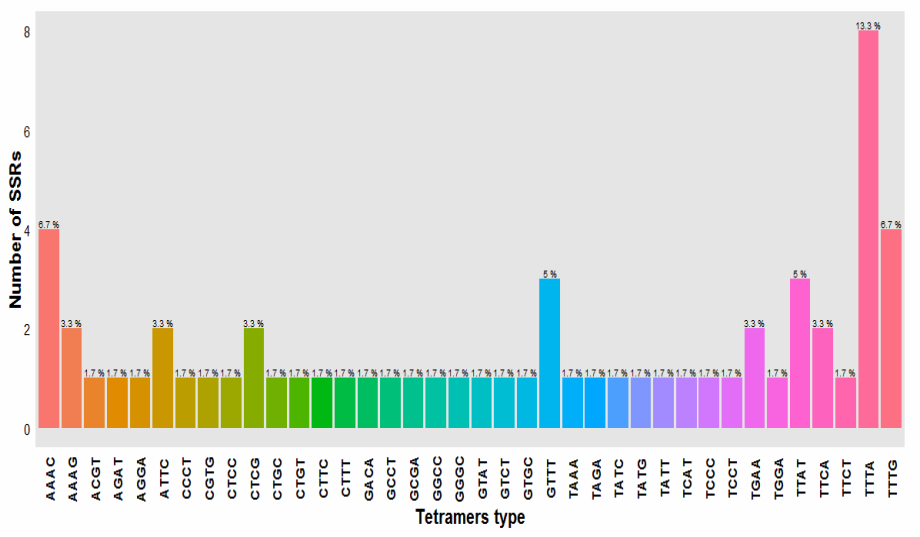

Figure 5 Frequency distribution for all tetramer motifs recognized in the camel sequence
The numbers on the columns demonstrate the percentage of them

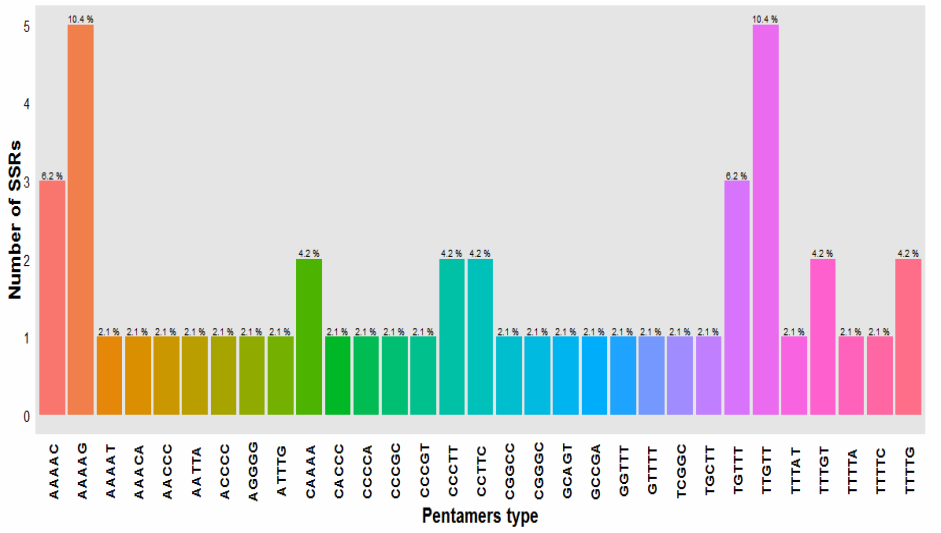

Figure 6 Frequency distribution for all pentamer motifs recognized in the camel sequence
The numbers on the columns demonstrate the percentage of them

Two-dimensional heat maps of clusters used to detection of similarities and dissimilarity of annotations among the gene group members. Cluster 1 had the highest enrichment score (5.23) and included 41 genes (Figure 7). Cluster 2 included 28 genes with an enrichment score of 2.84 (Figure 8). For developing the available camel SSR markers, the database containing 17155 ESTs was systematically searched for microsatellite motifs.

The outcomes clearly demonstrate that a useful source for mining SSRs are camel ESTs. It had been shown that the quantity of EST-SSRs was 4.0%.

This EST-SSR frequency was similar to cattle (4%) (Yan *et al.* 2008). Microsatellite-containing ESTs varies between vertebrate and ranged from 2% to 15% (Slate *et al.* 2007; Zhang *et al.* 2010; Bakhtiarizadeh *et al.* 2012; Nirapathpongporn *et al.* 2016; Zhou *et al.* 2016; Feng *et al.* 2018). These difference in the quantity of EST-SSRs perhaps influenced by redundancy, identification criteria of SSR, databases size and mining tools (Yan *et al.* 2008; Zhou *et al.* 2016). In present study, the dimeric motifs were the best frequent SSRs (38.86%) in camel that was in agreement with several other animal and chicken species (Yan *et al.* 2008; Bakhtiarizadeh *et al.* 2012; Abe and Gemmell, 2014; Sadder *et al.* 2015) but was dissimilar to some crop species (Varshney *et al.* 2002; Durand *et al.* 2010; Joshi *et al.* 2010; Zhou *et al.* 2016; Wang *et al.* 2017) and *Misgurnus anguillicaudatus* (Jiao *et al.* 2019) that trimeric motifs were plentiful.

Table 2 The GO enrichment analysis of sequences containing EST-SSRs at three levels of GO category

Category	Term	Group	Percent	P-value
Biological process	'GO:0000398'	'nuclear mRNA splicing, via spliceosome'	0.06	8.54E-04
	'GO:0000375'	'RNA splicing, via transesterification reactions'	0.06	8.54E-04
	'GO:0006457'	'protein folding'	0.06	0.001982
	'GO:0008380'	'RNA splicing'	0.07	0.002176
	'GO:0016071'	'mRNA metabolic process'	0.08	0.003957
	'GO:0006027'	'glycosaminoglycan catabolic process'	0.02	0.008316
	'GO:0002252'	'immune effector process'	0.04	0.011007
	'GO:0006397'	'mRNA processing'	0.06	0.015172
	'GO:0019725'	'cellular homeostasis'	0.08	0.018303
	'GO:0006026'	'aminoglycan catabolic process'	0.02	0.018337
	'GO:0043488'	'regulation of mRNA stability'	0.02	0.020041
	'GO:0006396'	'RNA processing'	0.09	0.020987
	'GO:0016052'	'carbohydrate catabolic process'	0.04	0.023605
	'GO:0043487'	'regulation of RNA stability'	0.02	0.023637
	'GO:0009057'	'macromolecule catabolic process'	0.11	0.023956
	'GO:0000272'	'polysaccharide catabolic process'	0.02	0.031547
	'GO:0034976'	'response to endoplasmic reticulum stress'	0.02	0.045048
	'GO:0051726'	'regulation of cell cycle'	0.06	0.048085
	'GO:0022417'	'protein maturation by protein folding'	0.01	0.048918
Cellular component	GO:0043233	organelle lumen'	24.55	7.83E-07
	'GO:0070013	intracellular organelle lumen'	23.95	1.22E-06
	'GO:0031974	membrane-enclosed lumen'	24.55	1.31E-06
	'GO:0030530	heterogeneous nuclear ribonucleoprotein complex'	2.99	2.13E-05
	'GO:0005654	nucleoplasm'	14.37	2.21E-05
	'GO:0031981	nuclear lumen'	19.16	3.51E-05
	'GO:0044451	nucleoplasm part'	10.78	4.18E-05
	'GO:0030529	ribonucleoprotein complex'	8.98	6.91E-04
	'GO:0005730	nucleolus'	8.98	0.010729
	'GO:0042470	melanosome'	2.99	0.012499
	'GO:0048770	pigment granule'	2.99	0.012499
	'GO:0005783	endoplasmic reticulum'	10.78	0.016242
	'GO:0000323	lytic vacuole'	4.19	0.020065
	'GO:0005764	lysosome'	4.19	0.020065
	'GO:0005829	cytosol'	13.17	0.025243
	'GO:0016604	nuclear body'	3.59	0.027647
	'GO:0005773	vacuole'	4.19	0.04242
	'GO:0005681	spliceosome'	2.99	0.0446
	GO:0005788	endoplasmic reticulum lumen'	2.40	0.047065
Molecular function	GO:0003723	'RNA binding'	0.36	4.85E-05
	GO:0050733	'RS domain binding'	0.05	0.001324
	GO:0004352	'glutamate dehydrogenase activity'	0.04	0.019011
	GO:0070728	'leucine binding'	0.04	0.019011
	GO:0016639	'oxidoreductase activity	0.04	0.019011
	GO:0008486	'diphosphoinositol-polyphosphate diphosphatase activity'	0.04	0.037664
	GO:0030528	'transcription regulator activity'	0.40	0.049479

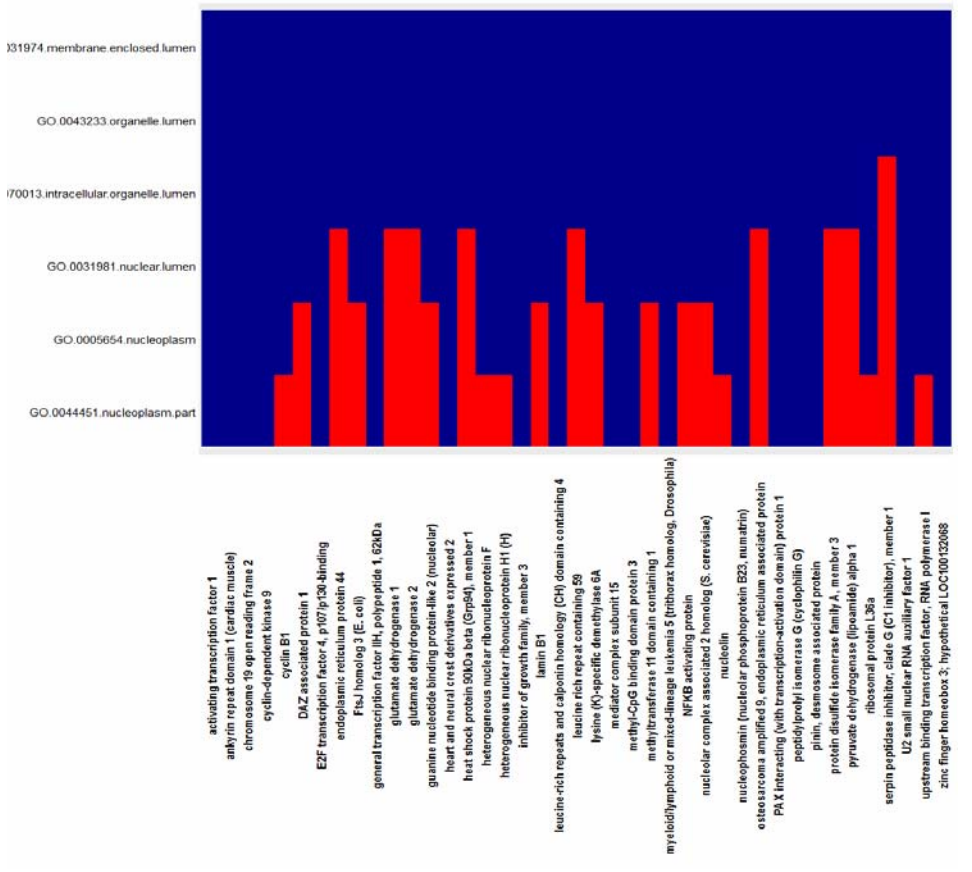

Figure 7 Two-dimensional gene annotation heat map for cluster 1
This cluster contains 41 genes with an average enrichment score of 5.23
The blue area of the heat map demonstrates common annotations and the red areas demonstrate differences in annotations

The trimeric motifs were the second most frequent repeats (27.15%), followed by 21.46%, 6.96%, and 5.57% for hexa-, tetra- and pentameric motifs. The frequency of hexamer repeats was in agreement with cattle (13%) (Yan *et al.* 2008) but was the difference from chickens (less than 1%) (Bakhtiarizadeh *et al.* 2012).

The quantity of the various SSRs motifs for every repeat number indicated that smaller repeat motifs are major between the identified SSRs. Amazingly, the occurrence of the repeat unit decrease with enlarging the length of them. This may be distinguished by the very fact that longer repeats motifs have higher mutation rates and therefore are less stable (Toth *et al.* 2000). The AC/TG was the best frequent kind of Dimeric motifs (54%) in the current investigation. The second frequency was AT/TA (32.8%) and The GC/CG motif has the lowest frequency (1.2%). This pattern of dimeric SSRs was similar to what had been found in alpaca (Reed and Chaves, 2008), cattle (Yan *et al.* 2008), sheep (Zhang *et al.* 2010), zebrafish (Ju *et al.* 2005) but different from that in grass(AG/TC) (Wang *et al.* 2017), rubber tree (AG/TC), (Nirapathpongporn *et al.* 2016), mint (AG/TC) (Kumar *et al.* 2015) chicken (AT/TA) (Bakhtiarizadeh *et al.* 2012), turmeric (AG/TC) (Joshi *et al.* 2010), hops (AT/TA) (Singh *et al.* 2012). This pattern could also be dependent on higher frequencies of confident amino acids in some species and various frequency of dimeric motif in different regions of genomes (Toth *et al.* 2000). The most plentiful trimer motif was GCC/GGC (19.2%), followed by AGC/GCT (10.3%). These results were in agreement with cattle (Yan *et al.* 2008) and other investigation in the animal species (Li *et al.* 2004b) which AGC and GCC were the most frequent, but were in dis-agreement with catfish (Serapion *et al.* 2004), zebrafish (Ju *et al.* 2005) and *Misgurnus anguillicaudatus* (Jiao *et al.* 2019) that the AAT/TAA repeat was the best frequent and trimer motifs made up only Gs and / or Cs nucleotides are infrequent. Also, the present result is similar to some of the plant species that GCC/GGC was the best frequent motif (Qin *et al.* 2015; Zhou *et al.* 2016; Wang *et al.* 2017). In the chicken, CAG was the most abundant trimeric repeat motif (Bakhtiarizadeh *et al.* 2012), although this trimer was one of abundant trimer in this study, but not in the first rank.

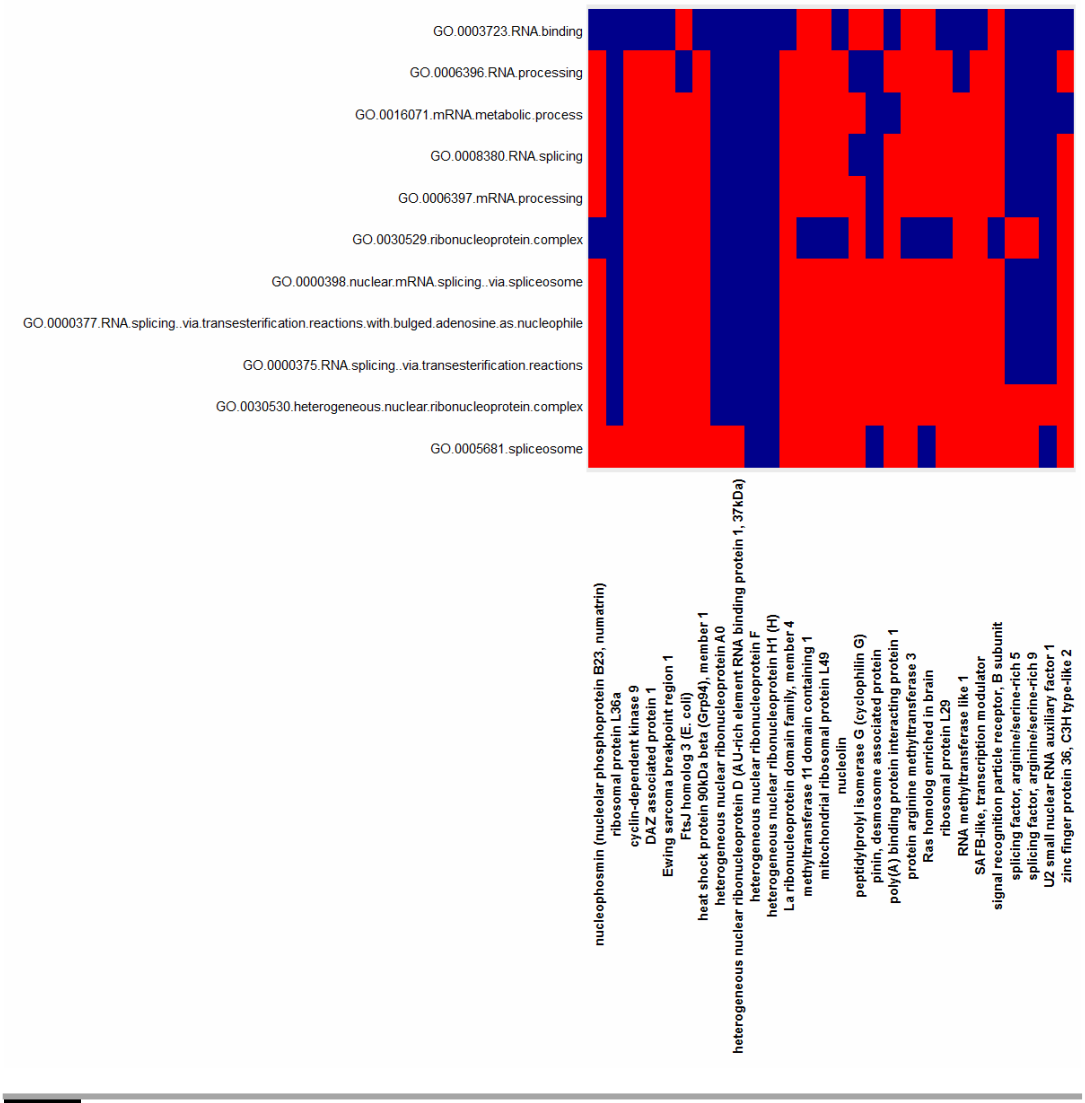

Figure 8 Two-dimensional gene annotation heat map for cluster 2
This cluster contains 28 genes with an average enrichment score of 2.84
The blue area of the heat map demonstrates common annotations and the red areas demonstrate differences in annotations

Dissimilar the delivery of the trimer motifs, the AT-rich tetramer, and pentamer motifs were the most abundant kind of camel EST-SSRs (Figures 5-6). Moreover, entirely composition of SSRs in camel coding regions is comparable to that in vertebrates and demonstrate that G/C repeats are less frequent than A/T repeats for these regions. Finally, current results obviously show that the major microsatellite types are taxon-dependent. In the study of Toth *et al.* (2000) reported that 'strand-slippage theories' alone cannot present SSRs distribution in the whole genome, also, enzymes and various proteins associated with different aspects of DNA-proceeding (such as replication and repair) and 'Chromatin remodeling' could be to blame for the taxon-specificity of SSRs frequency.

Annotating the sequences containing SSRs provides favorable conditions to inspect the functional variability of the various proteins (Bakhtiarizadeh *et al.* 2012). In general, GO is a helpful tool to unify the representation of gene and gene product features across all species (Consortium, 2008). Table 2 indicated the top-level (P≤0.05) GO terms at three levels along with the gene groups associated with the GO term. At the GO biological process related to gene list, 19 of the 42 assignments were significant. 19 of the 28 hits were meaningful in the cellular component and for molecular function 7 of the 15 assignments were significant. The results of the GO enrichment analysis showed that categories associated with gene expressions were significantly enriched that was in accordance with previous studies

(Bakhtiarizadeh *et al.* 2012; Zhou *et al.* 2016). This suggests that EST-SSRs may play functional roles in the regulation of gene. The functional categories of genes based on GO term showed that one cluster of 3 had more than 40 genes (Figure 7), indicating that these genes were categorized in the same functional group and also in the organelle, membrane-enclosed and nuclear lumen and nucleoplasm of cellular component clusters. Additionally, there are significant genes in the lists that are associated with dryland adaptations, containing fat and water metabolism, responses to aridity and heat stress (Al-Swailem *et al.* 2010; Jirimutu *et al.* 2012; Wu *et al.* 2014). The new evidence of this study shows that the genomic distribution of SSRs is non-random, likely due to their roles in the regulation of gene activity.

CONCLUSION

Nowadays, the extension of functional molecular markers like EST-SSRs is a very important and key goal for animal breeding. Especially, in marker-assisted selection programs. In this study, to develop the useful camel EST-SSR markers, the database including 17155 ESTs was systematically searched for microsatellite motifs. Our results clearly revealed that camel ESTs are a valuable resource for mining SSR markers. Finally, EST-SSRs recognized in this research are a helpful resource of camel genomic markers that can be proved and applied in various population genetic experiments in dromedary camel. Likewise, it is obvious that the number of EST-SSRs is not high, but this condition will be significantly modified with the utilization of next-generation sequencing data.

ACKNOWLEDGEMENT

Our special thanks to University of Jiroft for providing facilities support for this study.

REFERENCES

Abe H. and Gemmell N.J. (2014). Abundance, arrangement, and function of sequence motifs in the chicken promoters. *BMC Genom*. **15,** 900-912.

Al-Swailem A.M., Shehata M.M., Abu-Duhier F.M., Al-Yamani E.J., Al-Busadah K.A., Al-Arawi M.S., Al-Khider A.Y., Al-Muhaimeed A.N., Al-Qahtani F.H., Manee M.M., Al-Shomrani B.M., Al-Qhtani S.M., Al-Harthi A.S., Akdemir K.C., Inan M.S. and Otu H.H. (2010). Sequencing, analysis, and annotation of expressed sequence tags for *Camelus dromedarius*. *PLoS One*. **5,** e10720.

Altaher Y. and Kandeel M. (2015). Molecular analysis of some camel cytochrome P450 enzymes reveals lower evolution and drug-binding properties. *J. Biomol. Struct. Dyn*. **33,** 1-10.

Asadi A.A. and Rashidi Monfared S. (2014). Characterization of EST-SSR markers in durum wheat EST library and functional analysis of SSR-containing EST fragments. *Mol. Genet. Genom*. **289,** 625-640.

Bai J.Y., Pang Y.Z., Qi Y.X., Zhang X.H. and Yun X.Y. (2016). Development and application of EST-SSR markers in quails. Rev. Bras. Ciênc. Avíc. **18,** 27-32.

Bakhtiarizadeh M.R., Ebrahimi M. and Ebrahimie E. (2011). Discovery of EST-SSRs in lung cancer: Tagged ESTs with SSRs lead to differential amino acid and protein expression patterns in cancerous tissues. *PLoS One*. **6,** e27118.

Bakhtiarizadeh M.R., Arefnejad B., Ebrahimie E. and Ebrahimi M. (2012). Application of functional genomic information to develop efficient EST-SSRs for the chicken (*Gallus gallus*). *Genet. Mol. Res*. **11,** 1558-1574.

Cai K., Zhu L., Zhang K., Li L., Zhao Z. and Zeng W. (2019). Development and characterization of EST-SSR markers from RNA-Seq data in *Phyllostachys violascens*. *Front. Plant Sci*. **10,** 1-9.

Consortium G.O. (2008). The gene ontology project in 2008. *Nucleic Acids Res*. **36,** 440-444.

Du J., Zhang Z., Zhang H. and Junhong T. (2017). EST-SSR marker development and transcriptome sequencing analysis of different tissues of Korean pine. *Biotechnol. Biotechnol. Equip*. **31,** 679-689.

Durand J., Bodénès C., Chancerel E., Frigerio J.M., Vendramin G., Sebastiani F., Buonamici A., Gailing O., Koelewijn H.P., Villani F., Mattioni C., Cherubini M., Goicoechea P.G., Herrán A., Ikaran Z., Cabané C., Ueno S., Alberto F., Dumoulin P.Y., Guichoux E., de Daruvar A., Kremer A. and Plomion C. (2010). A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genom*. **11,** 570-581.

Ellis J.R. and Burke J.M. (2007). EST-SSRs as a resource for population genetic analyses. *Heredity*. **99,** 125-132.

Feng B., Yi S. V., Zhang M. and Zhou X. (2018). Development of novel EST-SSR markers for ploidy identification based on de novo transcriptome assembly for *Misgurnus anguillicaudatus*. *PLoS One*. **13,** 1-15.

Guzinski J., Mauger S., Cock J.M. and Valero M. (2016). Characterization of newly developed expressed sequence tag-derived microsatellite markers revealed low genetic diversity within and low connectivity between European *Saccharina latissima* populations. *J. Appl. Phycol*. **28,** 3057-3070.

Huang D.W., Sherman B.T. and Lempicki R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc*. **4,** 44-57.

Jiao S., Sun Y., Zhang D., Gao Q., Jin Y., Liu H., Ma Y., Yang Y., Porth I. and Mao J. (2019). Development of novel EST- -SSR markers for Ephedraceae (*Ephedra sinica*) by transcriptome database mining. *Appl. Plant Sci*. **7,** 3-7.

Jirimutu Wang Z., Ding G., Chen G., Sun Y., Sun Z., Zhang H., Wang L., Hasi S., Zhang Y., Li J., Shi Y., Xu Z., He C., Yu S., Li S., Zhang W., Batmunkh M., Ts B., Narenbatu U., Bat-Ireedui S., Gao H., Baysgalan B., Li Q., Jia Z., Turigenbayila S., Narenmanduhu W.Z., Wang J., Pan L., Chen Y., Ganerdene Y., Dabxilt E., Altansha A., Liu T., Cao M., Aruuntsever B., Hosblig H.F., Zha-ti A., Zheng G., Qiu F., Sun Z., Zhao L., Zhao W., Liu B., Li C., Chen Y., Tang X., Guo C., Liu W.,

Ming L., Temuulen C.A., Li Y., Gao J., Li J., Wurentaodi N.S., Sun T., Zhai Z., Zhang M., Chen C., Baldan T., Bayaer T., Li Y. and Meng H. (2012). Genome sequences of wild and domestic bactrian camels. *Nat. Commun*. **3,** 1202-1209.

Joshi R.K., Kuanar A., Mohanty S., Subudhi E. and Nayak S. (2010). Mining and characterization of EST derived microsatellites in *Curcuma longa*. *Bioinformation*. **5,** 128-131.

Ju Z., Wells M.C., Martinez A., Hazlewood L. and Walter R.B. (2005). An *in silico* mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, *Fundulus*, and *Xiphophorus*. *In Silico Biol*. **5,** 439-463.

Kastelic D., Frkovic-Grazio S., Baty D., Truan G., Komel R. and Pompon D. (2009). A single-step procedure of recombinant library construction for the selection of efficiently produced llama VH binders directed against cancer markers. *J. Immunol. Methods*. **350,** 54-62.

Khimoun A., Ollivier A., Faivre B. and Garnier S. (2017). Level of genetic differentiation affects relative performances of expressed sequence tag and genomic SSRs. *Mol. Ecol. Resour*. **17,** 893-903.

Kim B.Y., ParK H.S., Lee J.H., KwaK M. and Kim aNd Y. (2017). Development of microsatellite markers baseD on expresseD sequence tags in AspArAgus cochinchinensis (asparagaceae). *Appl. Plant Sci*. **5,** 1-5.

Kim K.S., Ratcliffe S.T., French B.W., Liu L. and Sappington T.W. (2008). The utility of EST-derived SSRs as population genetics markers in a beetle. *J. Hered*. **99,** 112-124.

Kumar B., Kumar U. and Yadav H.K. (2015). Identification of EST–SSRs and molecular diversity analysis in Mentha piperita. *Crop J*. **3,** 335-342.

Li B., Xia Q., Lu C., Zhou Z. and Xiang Z. (2004a). Analysis on frequency and density of microsatellites in coding sequences of several eukaryotic genomes. *Genom. Proteom. Bioinf*. **2,** 24-31.

Li Y.C., Korol A.B., Fahima T. and Nevo E. (2004b). Microsatellites within genes: Structure, function, and evolution. *Mol. Biol. Evol*. **21,** 991-1007.

Maia L.C.D., Palmieri D.A., Souza V.Q.D, Kopp M.M., Carvalho F.I.F.D. and Costa de Oliveira A. (2008). SSR locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int. J. Plant Genom*. **2008,** 412696.

Mbwana J., Bölin I., Lyamuya E., Mhalu F. and Lagergård T. (2006). Molecular characterization of Haemophilus ducreyi isolates from different geographical locations. *J. Clin. Microbiol*. **44,** 132-137.

Mirkin S.M. (2006). DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol*. **16,** 351-358.

Nguyen V.K., Hamers R., Wyns L. and Muyldermans S. (2000). Camel heavy-chain antibodies: Diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J*. **19,** 921-30.

Nirapathpongporn K., Kongsawadworakul P., Viboonjun U., Teerawattanasuk K., Chrestin H., Segiun M., Clément-Dement A. and Narangajavana J. (2016). Development and mapping of functional expressed sequence tag-derived simple sequence repeat markers in a rubber tree. *Mol. Breed*. **36,** 39-45.

Peakall R., Gilmore S., Keys W., Morgante M. and Rafalski A. (1998). Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Mol. Biol. Evol*. **15,** 1275-1287.

Pérez F., Ortiz J., Zhinaula M., Gonzabay C., Calderón J. and Volckaert F.A.M.J. (2005). Development of EST-SSR markers by data mining in three species of shrimp: Litopenaeus vannamei, Litopenaeus stylirostris, and Trachypenaeus birdy. *Mar. Biotechnol*. **7,** 554-569.

Qin Z., Wang Y., Wang Q., Li A., Hou F. and Zhang L. (2015). Evolution analysis of simple sequence repeats in plant genome. *PLoS One*. **10,** e0144108.

Reddy M.P., Sarla N. and Siddiq E.A. (2002). Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. *Euphytica*. **128,** 9-17.

Reed K.M. and Chaves L.D. (2008). Simple sequence repeats for genetic studies of Alpaca. *Anim. Biotechnol*. **19,** 243-309.

Sadder M., Migdadi H., Al-haidary A. and Okab A. (2015). Identification of simple sequence repeat markers in the dromedary (*Camelus dromedarius*) genome by next-generation sequencing. **39,** 218-228.

Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A. and Arnheim N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. **230,** 1350-1354.

Serapion J., Kucuktas H., Feng J. and Liu Z. (2004). Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar. Biotechnol*. **6,** 364-377.

Singh S., Gupta S., Mani A. and Chaturvedi A. (2012). Mining and gene ontology based annotation of SSR markers from expressed sequence tags of Humulus lupulus. *Bioinformation*. **8,** 114-22.

Slate J., Hale M.C. and Birkhead T.R. (2007). Simple sequence repeats in zebra finch (*Taeniopygia guttata*) expressed sequence tags: A new resource for evolutionary genetic studies of passerines. *BMC Genom*. **8,** 52-60.

Taheri S., Abdullah T.L., Yusop M.R., Hanafi M.M., Sahebi M., Azizi P. and Shamshiri R.R. (2018). Mining and development of novel SSR markers using next generation sequencing (NGS). *Molecules*. **23,** 399-406.

Toth G., Gaspari Z. and Jurka J. (2000). Microsatellites in different eukaryotic genomes : Survey and analysis. Genome Res. **10,** 967-981.

Varshney R.K., Thiel T., Stein N., Langridge P. and Graner A. (2002). In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell. Mol. Biol. Lett*. **7,** 537-546.

Vekemans X., Beauwens T., Lemaire M. and Roldan Ruiz I. (2002). Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol. Ecol*. **11,** 139-151.

Vieira M.L.C., Santini L., Diniz A.L. and Munhoz C. de F. (2016). Microsatellite markers : What they mean and why they are so useful. *Genet. Mol. Biol*. **39,** 312-328.

Wang B.H., Rong P., Cai X.X., Wang W., Zhu X.Y., Chen C.J.,

Xu Y.Y., Huang X.J., Zhuang Z.M. and Wang C.B. (2015). Development of EST-SSR markers related to disease resistance and their application in genetic diversity and evolution analysis in Gossypium. *Genet. Mol. Res*. **14,** 10630-10644.

Wang P., Yang L., Zhang E., Qin Z., Wang H., Liao Y., Wang X. and Gao L. (2017). Characterization and development of EST-SSR markers from a cold-stressed transcriptome of centipedegrass by illumina paired-end sequencing. *Plant Mol. Biol. Report*. **35,** 215-223.

Wu H., Guang X., Al-Fageeh M.B., Cao J., Pan S., Zhou H., Zhang L., Abutarboush M.H., Xing Y., Xie Z., Alshanqeeti A.S., Zhang Y., Yao Q., Al-Shomrani B.M., Zhang D., Li J., Manee M.M., Yang Z., Yang L., Liu Y., Zhang J., Altammami M.A., Wang S., Yu L., Zhang W., Liu S., Ba L., Liu C., Yang X., Meng F., Wang S., Li L., Li E., Li X., Wu K., Zhang S., Wang J., Yin Y., Yang H., Al-Swailem A.M. and Wang J. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun*. **5,** 5188-5199.

Yan Q., Zhang Y., Li H., Wei C., Niu L. and Guan S. (2008). Identification of microsatellites in cattle unigenes. *J. Genet. Genom*. **35,** 261-266.

Zane L., Patarnello T., Bargelloni L. and Patarnello T. (2002). Strategies for microsatellite isolation: A review. *Mol. Ecol*. **11,** 1-16.

Zhang W., Wang Z., Zhao Z., Zeng X., Wu H. and Yu P. (2010). Using bioinformcotics methods to develop EST-SSR makers from sheep's ESTs. *J. Anim. Vet. Adv*. **9,** 2759-2762.

Zheng X., Kuang Y., Lü W., Cao D. and Sun X. (2014). Transcriptome-derived EST–SSR markers and their correlations with growth traits in crucian carp *Carassius auratus*. *Fish Sci*. **80,** 977-984.

Zhou Q., Luo D., Ma L., Xie W., Wang Y., Wang Y. and Liu Z. (2016). Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus*) using Illumina sequencing. *Sci. Rep*. **6,** 20549-20559.