Using Bioinpormcotics Methods to Develop EST-SSR Makers from Sheep's ESTs

Wenxiang Zhang, Zunbao Wang, Zongsheng Zhao, Xiancun Zeng,
Hongbin Wu and Peng Yu
College of Animal Science and Technology, Shihezi University,
Shihezi, 832003, Xinjiang P.R. China

Abstract: About 29685 ESTs of sheep's skin from NCBI were mined to develop EST-SSR markers. Totally, 1456 SSRs were identified from 1364 ESTs using SSRHunter soft tools. The frequency of EST-SSRs was 4.9%. The dinucleotide repeat motif was the most abundant SSR accounting for 32.14% followed by 28.23, 7.14, 25.21 and 7.28%, respectively for tri, hexa, penta and tetra-nucleotide repeats. Among the di-nucleotide repeats, AC/TG (58.33%) were the most abundant type. About 261 primer pairs were designed from microsatellite among which 62 pairs could show clear PCR products by electrophoresisand 5 of these primers revealed moderate to high Polymorphism Information Content (PIC) in Chinese Merino sheep. The mean PIC value was 0.6971, ranging from 0.5499-0.7440.

Key words: Sheep, EST, SSR, NCBI, genetic mapping, China

INTRODUCTION

Simple Sequence Repeats (SSRs) have increasingly become the marker of choice for population genetic analyses. Unfortunately, the development of traditional anonymous SSRs from genomic DNA is costly and time consuming (Ellis and Burke, 2007). These problems are further compounded by a paucity of resources in taxa that lack clear economic importance.

However, the advent of the genomics age has resulted in the production of vast amounts of publicly available DNA sequence data including large collections of Expressed Sequence Tags (ESTs) from a variety of different taxa.

Recent research has revealed that ESTs are a potentially rich source of SSRs that reveal polymorphisms not only within the source taxon but also in related taxa (Cordeiro *et al.*, 2001; Varshney *et al.*, 2002). This study is on the purpose of identifying EST-SSR markers and investigate the type and distribution of repeat motifs in the expressed sequence tags of sheep's skin. The results will facilitate the use of molecular markers in sheep breeding.

These EST-SSR markers can be used for further study such as in genetic mapping, identification of Quantitative Trait Loci (QTLs) and comparative genomics studies of sheep (Scott *et al.*, 2000; Rohrer *et al.*, 2002; Gupta *et al.*, 2003).

MATERIALS AND METHODS

ESTs mining and EST-SSR identification: ESTs of sheep's skin were directly downloaded from NCBI dbEST on December 21, 2008, total 3295. Using SSRHunter software and SSRIT to identify SSRs from these sheep ESTs. Theparam-eters for data mining were set as: 7 repetitions for di, 6 repetitions for tri, 5 repetitions for tetra and 4 repetitions for penta and hexa-nucleotides.

Development of EST-SSR primer pairs: For each microsatellite-containing EST, primers were designed using primer premier (VERSION 5.0) and Oligo (Version 6.7). The major parameters for primer design were set as follows: primer length 18-23 bp with 20 bp as the optimum; PCR product size 100-450 bp; optimum annealing temperature 50-63°C; GC content 40-60% with 50% as the optimum. The primers were synthesized by Bioasia Biotech, Shanghai, China.

DNA extraction, PCR amplification and electrophoresis: The study covered a total of 120 Chinese Merino sheep kept on a farm located in the northwastern part of xinjiang Province. Genomic DNA was isolated as per the method described by Sambrook with minor modifications (Sambrook and Russell, 2002). After checking the quality and quantity DNA was diluted to a final concentration of 50 ng μL^{-1} in water and stored at 4°C. PCR amplification was performed in a thermal cycler. The reaction mixture

---

Corresponding Author: Zunbao Wang, College of Animal Science and Technology, Shihezi University, Shihezi, 832003, Xinjiang
P.R. China

Table 1: Polymorphisic EST-SSR markers in this study

Locus	Primer (5-3)	Motif	TM	Product (bp)
SER006	F:CCTTCTGTTATTACTTGCTG;R:TAGTCATGGTGAGGGGAG	TG	50.3°C	249-261
SER009	F:CTGGGTGATTCTGGAGGT;R:CAATAGGTTTCTGTCGGTTA	TG	55.5°C	213-235
SER012	F:GTATTCCAGGGGACTTCA;R:TAGCAGGGACACTTCTTTT	TG	55.0°C	259-275
SER016	F:TGTAGGGTGGGATATGAGGT;R:AAGCCATGTGAGGAGCAG	GT	54.7°C	286-298
SER018	F:ATTGCTGGTTCTGTGGGT;R:TTGGGAGTCTGGTTGGTAT	TG	51.5°C	214-228

contained about 50 ng of template DNA, 1 μmol L^{-1} of each primer, 200 μmol L^{-1} of each dNTP, 1.5 mmol L^{-1} of MgCl2 and 1.5 U Taq (Sangon) with 1×PCR buffer in a total volume of 25 μL.

The PCR program was: 5 min at 95°C followed by 35 cycles of 94°C for 45 sec, 45 sec at the annealing temperature of each primer pair (Table 1) and 60 sec at 72°C with a final extension at 72°C for 10 min. PCR products were visualized on silver-stained 10% polyacrylamide gels.

General statistical analyses: To analyze the variation of microsatellite loci, number of alleles at a certain locus (A) and allele size range (R) were determined by software Quantity One (Bio-Rad). For 5 microsatellite loci observed and expected heterozygosity estimates were calculated after Leveneand Nei as implemented in POPGENE software.

The observed and effective numbers of alleles were also calculated using POPGENE software. Allelic frequencies were utilized for the calculation of the Polymorphic Information Content (PIC) values.

RESULTS AND DISCUSSION

Searching for ESTs containing microsatellites: A total of 209 SSRs were identified from 196 EST sequences. The frequency of EST-SSRs in skin of sheep was 5.9%. The di-inucleotide repeat motifs were the most abundant SSRs in skin of sheep accounting for 71.8% followed by 22, 13, 7 and 4% for tri, hexa, penta and tetra-nucleotide repeats. Depending upon the length of the repeat unit itself (2-6 bp) the lengths of SSRs varied from 14-58 bp. The frequencies of EST-SSRs are shown in Fig. 1.

Frequencies of sheep's skin SSRs with different repeat motifs: Investigation of the distribution of SSR motifs can help gain insights into the composition of genome. The observed frequency of di-nucleotide repeat motifs comprising the SSRs is shown in Fig. 2.

Among the di-nucleotide repeats, AC/TG was the most abundant type accounting for 67% of all di-nucleotide repeats found in the sheep's skin ESTs. AG/TC was the second abundant type, accounting for 19%. The GC/CG motif was the least abundant type only accounting for 1%.
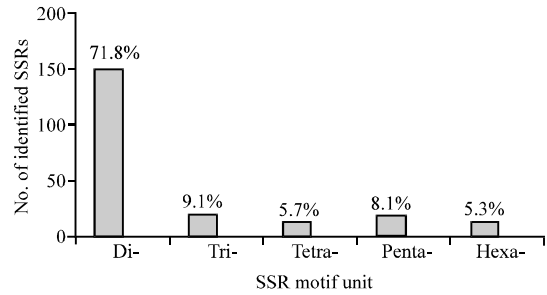

Fig. 1: Frequency distribution of different repeat types (2-6 motif unit) microsatellite identified in EST of sheep's skin. The numbers on the bars indicate the percentage of each repeat type microsatellites in total number
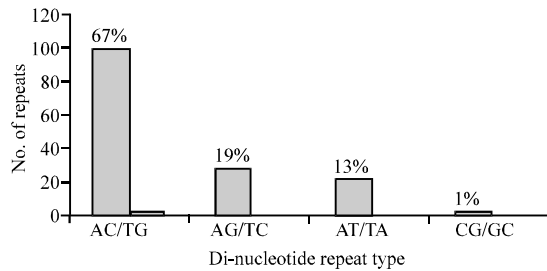

Fig. 2: Frequency distribution of 4 di-nucleotide repeat type in EST of sheep's skin. The numbers on the bars indicate the percentage of the 4 di-nucleotide repeat type in all di-nucleotide repeat types

EST-SSR locus polymorphism: About 22 primer pairs were used in PCR amplification, 18 pairs could show clear PCR products by electrophoresis and 5 pairs were polymorphic in Chinese Merino sheep (Table 1 and Fig. 3 and 4).

Genetic data analysis: Microsatellite allele frequency profile in Chinese Merino sheep population is shown in Table 2. A total of 25 alleles were detected and the amplification fragments ranged from 213-286 bp. The number of alleles observed across the microsatellite loci varied from 3 (SER006) to 7 (SER012) with an average allele number of 4.8/locus. The observed number of alleles across the loci was more than the effective number of alleles (2.6422-4.8485) as expected. The PIC (Polymorphic Information Content) showed that most of the loci were

Table 2: Allele frequency of population

Allele	SER006	SER009	SER011	SER012	SER016	SER018
Allele A	0.2083	0.1000	0.2333	0.0833	0.2000	0.1417
Allele B	0.5000	0.1500	0.2917	0.0667	0.2500	0.1750
Allele C	0.2917	0.2833	0.3167	0.1250	0.2167	0.2416
Allele D	-	0.2750	0.1583	0.0500	0.1916	0.2667
Allele E	-	0.1917	-	0.0167	0.1417	0.1750
Allele F	-	-	-	0.3916	-	-
Allele G	-	-	-	0.2667	-	-

Table 3: Genetic parameters of 5 microsatellite loci in Chinese Merino sheep

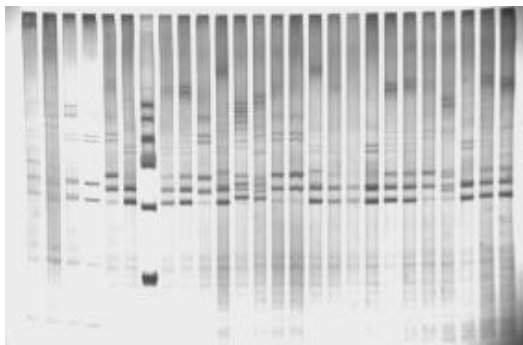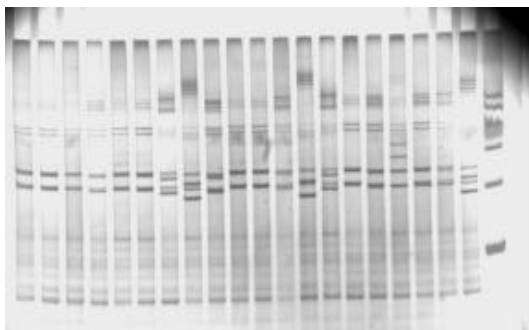| Parameters	SER006	SER009	SER011	SER012	SER016	SER018	Mean	STDEV
PIC	0.5499	0.7383	0.6859	0.7100	0.7607	0.7555	0.7001	0.0835
He	0.4833	0.6833	0.6500	0.7667	0.5333	0.6000	0.6194	0.1754
Ne	2.6422	4.4417	3.7756	3.9323	4.8485	4.7431	4.0639	0.8036

Fig. 3: Electrophoresis result type

Fig. 4: Electrophoresis result type

highly informative indicating polymorphism across the loci with an overall mean of 0.7001±0.0835 (Table 3). The average observed heterozygosity was less than the expected (Table 3). The average expected gene diversity within the population ranged from 0.4833 (SER007) to 0.7667 (SER012) with an over all mean of 0.6194.

The rapid and inexpensive development of SSRs from Expressed Sequence Tag (EST) databases has been shown to be a feasible option for obtaining high-quality nuclear markers. Moreover, the National Center for Biotechnology Information (NCBI) EST database contains an ever-increasing number of these single-pass cDNA sequences, meaning that the resources necessary for the efficient development of large numbers of so-called EST-SSRs already exist for a wide variety of taxa. Yin *et al.* (2005) analysised the repeated sequence of the skin of goat and the result showed that almost >10 copise is keratin and the zipper structure of amino acids is likely the most abundant district of SSR. So we chose the skin's EST to develop EST-SSR marker. The result showed that the percentage of sequences containing SSR in the skin's EST database was 6.3% which was relatively abundant.

Di-nucleotide SSRs are mostly abundant in sheep's skin ESTs registered in GenBank. Some similar results were also reported in many animal's researches, different from plants (Thiel *et al.*, 2003; Yue *et al.*, 2004; Yan *et al.*, 2008). The duplication of AT/CG is the most affluence. A total of 22 primer pairs were designed from microsatellite, among which 18 pairs could show clear PCR products by electrophoresis and 5 microsatellite site are high polymorphism. They may be expected as the markers of wool trait locus, researchers look forward to contributing to the later relative research such as in genetic mapping and identification of Quantitative Trait Loci (QTLs).

EST is a part of functional gene. The better explanation can be made by its difference of appearance. So we will use 5 polymorphismy microsatellite site to find the difference between the shag sheep (duolang sheep, half-fuzz sheep (romney marsh sheep) and fuzz sheep (Chinese merino sheep) in latter research.

CONCLUSION

The result showed that these EST-SSR markers can be used for further study such as in genetic mapping, identification of Quantitative Trait Loci (QTLs) and comparative genomics studies of sheep.

REFERENCES

Cordeiro, G.M., R. Casu, C.L. McIntyre, J.M. Manners and R.J. Henry, 2001. Microsatellites markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci., 160: 1115-1123.

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

Gupta, P.K., S. Rustgi, S. Sharma, R. Singh, N. Kumar and H.S. Balayan, 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol. Genet. Genomics, 270: 315-323.

Rohrer, G.A., S.C. Fahrenkrug, D. Nonneman, N. Tao and W.C. Warren, 2002. Mapping microsatellite markers Identified in porcine EST sequences. Anim. Genet., 33: 372-376.

Sambrook, J. and D.W. Russell, 2002. Molecular Cloning: A Laboratory Manual. Science Press, Beijing, pp: 463-470.

Scott, K.D., P. Eggler, G. Seaton, M. Rossetto, E.M. Ablett, L.S. Lee and R.J. Henry, 2000. Analysis of SSRs derived from grape ESTs. Theor. Applied Genet., 100: 723-726.

Thiel, T., W. Michalek, R.K. Varshney and A. Graner, 2003. Exploiting EST databases for the developpment and characterization of gene-derived SSR-markers in barley. Theor. Applied Genet., 106: 411-422.

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

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

Yin, J., J. Li, Y. Zhang, C. Li and Z. Guo, 2005. Study on gene expression of skin from adult inner mongol ia cashmere goat. Acta Vet. Zootechnica Sinica, 36: 996-1000.

Yue, G.H., M.Y. Ho, L. Orban and H. Komen, 2004. Microsatellites wit hingenes and ESTs of common carp and their applicability in silver crucian carp. Aquaculture, 234: 85-98.