INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
17–1130/2018/20–3–617–620
DOI: 10.17957/IJAB/15.0531
http://www.fspublishers.org

Full Length Article

Whole Genomic EST-SSR Development Based on High-Throughput Transcript Sequencing in Proso Millet (*Panicum miliaceum*)

Yanmiao Jiang[1,2†], Haiquan Li[1,2†], Jianan Zhang[1,2], Jinying Xiang[1,2], Ruhong Cheng[1,2*] and Guoqing Liu[1,2*]
[1]*Institute of Millet Crops, Hebei Academy of Agricultural & Forestry Sciences, 050035 Hebei, China*
[2]*Key Laboratory of Minor Crops in Hebei, Shijiazhuang 050035 Hebei, China*
[*]For correspondence: rhcheng63@126.com; guoqingliu@hotmail.com
[†]Both authors contributed equally to this work

Abstract

Proso millet (*Panicum miliaceum* L.) with short growing period, is highly suited to drought and semi-arid climate, and thus possesses great potential for research. In this study, 8139 Expressed Sequence Tags-simple sequence repeat (EST-SSR) makers were developed from the transcript sequencing of Proso millet cultivar Yanshu 5. Tri-type was the major type, accounting for 83.65% of all the found SSR types, followed by di-type accounting for 11.1%. CCG/CGG was the most ample repeat-motif types in Proso millet. Twenty four primer pairs of these markers were synthesized for empirical validation and 21 primers obtained targeted fragments by PCR amplification. Three of the 21 markers were polymorphic among 20 Chinese Proso millet landraces which suggested that these EST-SSR markers can be used for genetic studies. Markers developed in this work can be used for agronomic trait location, linkage map construction and molecular breeding in Proso millet in the future. © 2018 Friends Science Publishers

Keywords: Proso millet; High-throughput sequencing; SSR development; Characterization

Introduction

Proso millet (*Panicum miliaceum*, 2n=4x=36), is one of the oldest C_4 plant with a growth period of about 60 to 100 days in summer and can be planted in a wide range of region (Lagler *et al*., 2005; Kalinova and Moudry, 2006). As an allotetraploid self-pollinated cereal crop with highest water-use efficiency, Proso millet is the best crop for rotation in dry land production system in the semi-arid high plains in the world. The grains are used for feeding bird and livestock in Europe and the United States, whereas for human food in countries like China, Russia and India (Kalinova and Moudry, 2006). It was one of the main food of human, however after 18th century, its importance slowly decreased along with the import of other crops such as potato, wheat, rice and so on.

The grains of Proso millet contain rich proteins, vitamins, mineral substances and the nutrition value are comparable or better than today's staple food. The ratio of protein, saccharide and lipids is very close to recommendations (Geervani and Eggum, 1989; Kalinova and Moudry, 2006).

As Proso millet is a minor crop worldwide right now, studies on its genetics and breeding are very limited. Some kinds of markers have been reported in Proso millet. AFLP and RAPD markers were used for Proso millet genetic diversity analysis (Mribu and Hilu, 1994; Karam *et al*., 2004, 2006).

Simple sequence repeat (SSR) markers are highly reproducible, multi-allelic, co-dominant, ample and evenly distributed in the genome and considered as the most informative and useful marker (Powell *et al*., 1996; Gupta and Varshney, 2000; Li *et al*., 2016). From barley, wheat, rice and oat, Hu *et al.* (2009) developed 46 SSRs markers for analyzing Proso millet genetic diversity. Cho *et al*. (2010) identified 25 Proso millet SSR markers from its BAC library. Rajput *et al*. (2014) got 339 SSR markers from Switchgrass, which can be used in Proso millet. Wang *et al*. (2014) developed 1210 SSR markers from the sequence of high-throughput sequencing and magnetic beads enriched library. Until now, whole genomic EST-SSR markers of Proso millet have not been reported yet.

Currently, next generation sequencing (NGS) technology has been used in whole genome re-sequencing and *de novo* sequencing, and have unique advantage in transcriptome analysis (Zhou *et al*., 2010). It is a cost-efficient way to use NGS to obtain amount of EST sequences in non-model plant because it will generate a resource of large full length sequences. Because it is genic and it may be associated with functional regions potentially, EST-SSR markers have advantage over the other SSR makers.

To cite this paper: Jiang, Y., H. Li, J. Zhang, J. Xiang, R. Cheng and G. Liu, 2018. Whole genomic EST-SSR development based on high-throughput transcript sequencing in proso millet (*Panicum miliaceum*). *Int. J. Agric. Biol*., 20: 617–620

In this study, EST-SSR makers have been developed from the transcript sequencing and characterized in Proso millet. The developed SSR markers will be helpful in the assessment of genetic diversity, linkage map construction, phylogenetic and population genetic relationship evaluation and for marker-assisted selection.

Materials and Methods

Plant Material

The leaves of Proso millet cultivars Yanshu5 seedling were collected and used for the isolation of RNA for RNAseq. Twenty Chinese Proso millet cultivars were used for the polymorphism test of the developed markers (Table 1).

RNA Preparation and Sequencing

Total RNA was extracted with TRIzol reagent (Tiangen Biotech Co., Ltd., Beijing, China) following the manufacturer's instructions. The sequencing was performed on Illumina Higseq 2000 platform by Shanghai Major Bio-Pharm Technology Corporation. The clean reads were assembled using Trinity software for *de novo* transcriptome assembly without a reference genome.

EST-SSR Identification and Validation

Software Batch Primer 3 (http://probes.pw.usda.gov/batchprimer3/) was used to recognize the common SSR loci repeat types and design corresponding primers (You *et al*., 2008). The minimum repeat numbers of Di-, Tri-, Tetra-, Penta- and Hexa-nucleotide were 6, 4, 3, 3, and 3, respectively. Parameters for designing primers were as follows: primer length between 19 to 25 nucleotides (20 optimum); the best annealing temperature was 58°C; GC content between 40 to 60% (50% optimum). A set of 24 primer pairs was synthesized for empirical testing. Genomic DNA was extracted from young leaves from a set of Chinese Proso millet cultivars using CTAB (Doyle and Doyle, 1987).

PCR reactions were performed in a total volume of 20 μL containing 1× PCR buffer (15 mM $MgCl_2$, 500 mM KCl, 100 mM Tris-HCl, pH 8.3), 200 μM dNTPs, 0.10 μM each primer and 1 μL of template DNA (approximately 30 ng/μL) and 1 unit of Tag DNA polymerase. PCR program was performed as follows: denaturation at 95°C for 4 min, then followed by 35 cycles of 40 sec at 95°C, 40 sec at 58°C 35 sec at 72°C, finally extension at 72°C for 10 min. The PCR products were separated on a 6% polyacrylamide gel followed by silver staining.

Results

Classification of Repeat Motif Types

We found a total of 8734 SSRs in contigs assembled from the RNA-seq. Among all the detected SSR types, the Tri-type comprised the greatest number (83.6%) of the total number. The Di-type took the second place (11.1%). The other types take relatively lower proportions (Fig. 1A). AG and CT claimed relatively higher proportion of the Di-type (Fig. 1B). The repeat motif of Tri-type is shown in Fig 1B. CCG and CGG accounted for most of the Tri-type.

Table 1: Cultivars used for polymorphism detection

No.	Cultivars	No.	Cultivars
1	Huangkemi	11	Jilinshuzi
2	Bamengmengmizi	12	Tongyuheishuzi
3	Linxishugu	13	Haixingshuzi
4	Qiemamizi2	14	Dongmanshuzi
5	Qiemamizi1	15	Hongmizi
6	Baodingshuzi	16	Bailishu
7	Keqidahuangmi	17	Hongruanmi
8	Huai137	18	Jinshu
9	Hongruanmi	19	Yanshu5
10	Taiyuan111	20	Zjkyeshengmizi

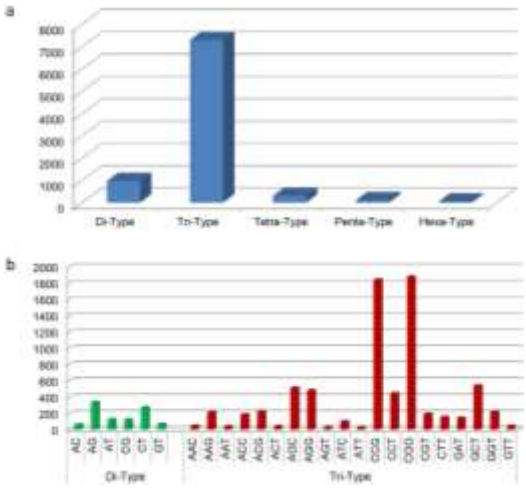

Fig. 1: Characterization of SSR mining (a) Distribution of different repeat type classes; (b) Frequency of classified repeat types

EST-SSR Identification and Validation

The contigs from transcriptome data were used for SSR identification and validation. After processed by Batch Primer3, a total of 8734 EST-SSR motif (di- to hexa-nucleotide) were identified from a total of 27784 contigs and 8139 primer pairs were designed (Supplementary Table S1). A sub-set of 24 primer pairs were randomly selected and synthesized for empirical validation (Table 2). Initial testing of the SSRs was performed on Yanshu5.

Twenty four primer pairs were synthesized to test the amplification efficiency of the primers. Five of them were Di-type, seven Tri-type, four Tetra-type, four Penta-type and four Hexa-type. Twenty one pairs amplified clear expected bands, accounting for 87.5% of all the designed primer pairs. About 7121 primer pairs could be used from this study expected by this probability (Fig. 2).

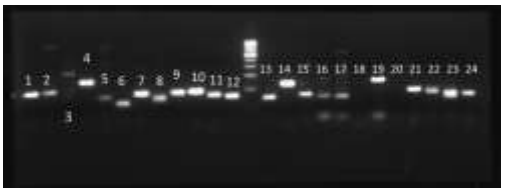

Fig. 2: Amplification efficiency of the primers tested in this study

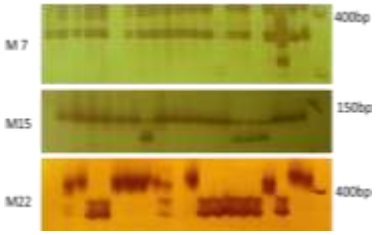

Fig. 3: Polymorphic markers among the rest cultivars

Table 2: Details of the SSR molecular markers validated in this study

Index	Contig	Repeat motif	Forward 5'-3'	Reverse 5'-3'	Tm℃	size
Pm8108	comp32882	GGCCGT	CTCTACGCACACAGATCAAAC	AGCATCATCAAGTCCATGTC	56	163
Pm8109	comp39156	CCCACG	GTTGCCCTTCTCCTCCTC	TCGTCTGAGAAAAGAAGTCAT	56	172
Pm8110	comp40188	CGCCAA	CTCTACCTTCCCCTGCTC	GAACTCGTCGTGCGTGAG	53	300
Pm8112	comp41557	TCTCCT	CTATCTAGCGGTGGTCTGTC	TAGAGCTACTTGTCGATCTGG	56	239
Pm8027	comp1674	CGCGG	GAGCAACTAGGTAGCAGTAGG	TCACTGCTCCTCAAGCAT	55	130
Pm8028	comp8402	AAAAC	ATTTACAACCACTCCACCATA	AATAATTGTTGATGCAAGAGC	56	100
Pm8029	comp16834	AGATT	GAAAAGTACGAACTGGTTCAA	ACTTTAGCACGAGGGTGAA	56	154
Pm8030	comp24776	GGTGT	ACATTTGTTTGTTTGGGAAG	AAGAAGAGATGGAGGAACAAG	56	130
Pm7688	comp4843	TGAT	TCATAACAGCATCTGAGTCAC	AACAGTGAACAACAGCTTGAT	56	162
Pm7689	comp7759	AGAT	ATACTCCTCTGATAGCGGTCT	GATCCTTTCATCTCCTATGCT	56	164
Pm7690	comp8223	CTGA	GGACTCGATCTTTCCTCTAGT	AGGGTATCTTTTCCCCTTTAT	56	146
Pm7691	comp17137	AACG	CTCATCAACTACCCATACCAA	GTCGACGACTTCCACAAG	56	140
Pm562	comp101	TCG	GAACCCGCTCAGCATAAG	GGAGGAGGAGGATGAGTTAAT	56	130
Pm563	comp101	TCC	GTAATTCCCTGCGATTAACT	AGTTCTGCATCTCAGGTTTC	56	224
Pm564	comp1123	GCG	GCATACTGCTGCAGATACATA	CGGCTCCTCTCTCTTCAG	56	144
Pm565	comp1123	GCC	AGGCCGCTGAAGAGAGAG	GATGCTCATCGCGAACTT	56	162
Pm566	comp1123	GGT	CCGCTGAAGAGAGAGGAG	GATGCTCATCGCGAACTT	56	159
Pm567	comp1860	CTT	TAGGTGTACCAAGCTCATAGG	CTCTTGTAGAGGACCAGGAAC	-	225
Pm568	comp1860	GGC	GGACGTGGACTTCTTCTTCT	CGTGGTAGAACACGAGGAG	56	245
Pm1	comp10243	CG	GCTCCATCCGTTCGAGTC	TAACAGCTAGATCAACCCAGA	-	205
Pm2	comp15084	AC	TGTTCAGTTTCTTCATGACCT	ATGGGATCGATTCATTCTAAG	56	163
Pm3	comp16582	TA	ACGCCTCTTCTTTTCTTTTT	ATATCATCAAATCGAAACGTG	56	149
Pm4	comp16585	AC	GACAGACACACGCATACATTA	GCTTTGTGACAGTTGTGGTTA	56	137
Pm5	comp16636	TC	CAAGTGATTTAACAAAGAAACG	TGCATCAATTCTTGAGGTAAT	56	141

Polymorphism Detection of the EST-SSR Makers

We used 20 Chinese Proso millet cultivars including landrace and modern cultivars for the polymorphism detection (Table 1). Three of the 21 pairs of primers were detected to be polymorphic among the 20 cultivars. Marker 7 was Penta-type, marker 15 was Tri-type and marker 21 was Di-type (Fig. 3). All three polymorphic markers amplified two types of band patterns.

Discussion

More and more attentions were paid to Proso millet due to its many unique characteristics. Thus use and conservation of Proso millet germplasm, especially for illustrating the unique characters and application in breeding requires more information on its genetic diversity (Hu *et al*., 2008; Cho *et al*., 2010; DvořÁková *et al*., 2015). So far, SSR markers are one of the most convenient DNA markers to study crop genetic diversity. Although some SSR markers have been developed from Proso millet (Hu *et al*., 2009; Rajput *et al*., 2014; 2016), the genome-wide range of EST-SSR markers has not been reported yet. In this study, we developed 8139 EST-SSR markers from the transcript contigs. In the marker validation experiments 21 pairs showed clear expected bands, accounting for 87.5% of all the primer pairs designed. This is an effective way to develop SSR markers. Three pairs were diverse in the 20 Proso millet cultivars, which showed that those EST-SSR markers can be used for analysis the diversity in the Proso millet germplasms.

The previous reports showed that the AT/TA repeat type was the most-common SSR type in plants and CT/GA repeat type take the second place (He *et al*., 2003; Hsu *et al*., 2011; Singh *et al*., 2012). In a diversity study using SSR markers in Proso millet, the most-common type of SSR repeat was the CT/GA repeat (32.6%), followed by the AC/TG repeat (26.1%), and the type of AT/TA was only15.2% (Hu *et al*., 2009). Among the 254 amplified SSR markers derived from switchgrass, the majority (87%) amplified dinucleotide repeats and AG/GA was the amplest repeat type in Proso millet (Rajput *et al*., 2016). In another study on Proso millet, the predominance repeat motifs were Tri-nucleotide SSR (53.02%) and the frequency of Di-nucleotide SSR was 37.21% (Cho *et al*., 2010). However, in several plant species, including Arabidopsis (Morgante *et al*., 2002), rice (Goff *et al*., 2005), sesame (Dixit *et al*., 2005),

Medicago Truncatula (Mun *et al*., 2006), Mung Bean (Gwag *et al*., 2006), Foxtail Millet (Jia *et al*., 2009), and garlic (Ma *et al*., 2009), it has been shown that Tri-nucleotide repeats are the most frequent SSR type. In this study, SSR Tri-nucleotide repeat motif was the most frequent ones accounting for 83.6% of all types, as reported for cucumber (Guo *et al*., 2010) and sweet potato (Wang *et al*., 2010).

Conclusion

Proso millet has a short growing-period, high water use efficiency and is highly suited to a semi-arid climate. However, it is one of the least-studied species of crops of family Poaceae, and its genomic resources are very limited. In this study, 8139 SSR markers were developed and 3 of 24 makers have diversity in the empirical validation. A large number of makers can be characterized and used to construct linkage map, assess phylogenetic and population genetic relationships and for marker-assisted selection. Consequently, it will promote the genetic study of Proso millet.

Acknowledgments

This work was financially supported by China Agricultural Research System (CARS-06-13.5-A10) and a Special Fund for Agricultural Science, Technology and Innovation in Hebei, China to GL (201603286).

References

Cho, Y., J. Chung, G. Lee, K. Ma, A. Dixit, J. Gwag and Y. Park, 2010. Development and characterization of twenty-five new polymorphic microsatellite markers in proso millet (*Panicum miliaceum* L.). *Gen. Genomics*, 32: 267–273

Dixit, A., M. Jin, J. Chung, J. Yu, H. Chung, K. Ma, Y. Park and E. Cho, 2005. Development of polymorphic microsatellite markers in sesame (*Sesamum indicum* L.). *Mol. Ecol. Notes*, 5: 736–738

Doyle, J. and J.L. Doyle, 1987. Genomic plant DNA preparation from fresh tissue CTAB method. *Phytochem. Bull*., 19: 11–15

DvořÁková, Z., P.H. Cepková, D. Janovská, I. Viehmannová, E. Svobodová, E.F. Cusimamani and L. Milella, 2015.Comparative analysis of genetic diversity of 8 millet genera revealed by ISSR markers. *Emirates J. Food Agric.*, 27: 617–628

Geervani, P. and B.O. Eggum, 1989. Nutrient composition and protein quality of minor millets. *Plant Foods Hum. Nutr*., 39: 201–208

Goff, S.A., R. Darrell, L. Tien-Hung, P. Gernot, W. Ronglin, D. Molly, G. Jane, S. Allen, O. Paul and V. Hemant, 2005. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 296: 92–100

Guo, S., Y. Zheng, J.G. Joung, S. Liu and Z. Zhang, 2010. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics*, 11: 384

Gupta, P.K. and R.K. Varshney, 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica*, 113: 163–185

Gwag, J.G., W.C. Jong, K.C. Hun, H.L. Jeong, H.M.A. Kyung, D. Anupam, J.P. Yong, G.C. Eun, S.K. Tae and H.L. Suk, 2006. Characterization of new microsatellite markers in mung bean, *Vigna radiata* (L.). *Mol. Ecol. Notes*, 6: 1132–1134

He, C., V. Poysa and K. Yu, 2003. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. *Theor. Appl. Genet.*, 106: 363–373

Hsu, C.C., Y.L. Chung, T.C. Chen, Y.L. Lee, Y.T. Kuo, W.C. Tsai, Y.Y. Hsiao, Y.W. Chen, W.L. Wu and H.H. Chenm, 2011. An overview of the *Phalaenopsis* orchid genome through BAC end sequence analysis. *BMC Plant Biol*., 11: 3

Hu, X., J. Wang, P. Lu and H. Zhang, 2009. Assessment of genetic diversity in broomcorn millet (*Panicum miliaceum* L.) using SSR markers. *J. Genet. Genomics*, 36: 491–500

Hu, Y.G., J. Zhu, F. Liu, Z. Zhang, Y. Chai and W.N. Song, 2008. Genetic diversity among Chinese landraces and cultivars of broomcorn millet (*Panicum miliaceum*) revealed by the polymerase chain reaction. *Ann. Appl. Biol*., 3: 357–364

Jia, X., Z. Zhang, Y. Liu, C. Zhang, Y. Shi, Y. Song, T. Wang and Y. Li, 2009. Development and genetic mapping of SSR markers in foxtail millet [*Setaria italica* (L.) P. *Beauv*.]. *Theor. Appl. Genet.*, 118: 821–829

Kalinova, J. and J. Moudry, 2006. Content and quality of protein in proso millet (*Panicum miliaceum* L.) varieties. *Plant Food. Hum. Nutr*., 61: 43–47

Karam, D., P. Westra, S.J. Nissen, S.M. Ward and J. Figueiredo, 2004. Genetic diversity among proso millet (*Panicum miliaceum*) biotypes assessed by AFLP technique. *Planta Daninha*, 22: 167–174

Karam, D., P. Westra, S.J. Niessen, S.M. Ward and J.E.F. Figueiredo, 2006. Assessment of silver-stained AFLP markers for studying DNA polymorphism in proso millet (*Panicum miliaceum* L.). *Braz. J. Bot*., 29: 609–615

Lagler, R., G. Gyulai, M. Humphreys, Z. Szabo, L. Horvath, A. Bittsanszky, J. Kiss, L. Holly and L. Heszky, 2005. Morphological and molecular analysis of common millet (*P. miliaceum*) cultivars compared to an aDNA sample from the 15th century (Hungary). *Euphytica*, 146: 77–85

Li, M., Y. Yang, F. Feng, X. Li, L. Gu, Y. Tian, J. Wang, J. Wang, Z. Li, A. Chen, X. Chen and Z. Zhang, 2016. SSR development and utilization with *Rehmannia glutinosa* transcriptome. *Int. J. Agric. Biol*., 18: 589–600

Ma, K., J. Kwag, W. Zhao, A. Dixit, G. Lee, H. Kim, I. Chung, N. Kim, J. Lee, J. Ji, T. Kim and Y. Park, 2009. Isolation and characteristics of eight novel polymorphic microsatellite loci from the genome of garlic (*Allium sativum* L.). *Sci. Hortic*., 122: 355–361

Morgante, M., M. Hanafey and W. Powell, 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.*, 30: 194–200

Mribu, H.K. and K.W. Hilu, 1994. Detection of interspecific and intraspecific variation in *Panicum* millets through random amplified polymorphic DNA. *Theor. Appl. Genet*., 88: 412–416

Mun, J.H., D.J. Kim, H.K. Choi, J. Gish, F. Debelle, J. Mudge, R. Denny, G. Endre, O. Saurat and A.M. Dudez, 2006. Distribution of microsatellites in the genome of *Medicago Truncatula*: a resource of genetic markers that integrate genetic and physical maps. *Genetics*, 172: 2541–2555

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey and A. Rafalski, 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite)markers for germplasm analysis. *Mol. Breed*., 2: 225–238

Rajput, S.G., T. Plyler-Harveson and D.K. Santra, 2014. Development and Characterization of SSR Markers in Proso Millet Based on Switchgrass Genomics. *Amer. J. Plant Sci*., 1: 175–186

Rajput, S.G., D.K. Santra and J. Schnable, 2016. Mapping QTLs for morpho-agronomic traits in proso millet (*Panicum miliaceum* L.). *Mol. Breed*., 36: 1–18

Singh, S., S. Gupta, A. Mani and A. Chaturvedi, 2012. Mining and gene ontology based annotation of SSR markers from expressed sequence tags of *Humulus Lupulus*. *Bioinformation*, 8: 114–122

Wang, Y., M. Liu, P. Lu, Z. Qiao, T. Yang, H. Li and X. Cui, 2014. Screening SSR primer for construction of genetic linkage maps of proso millet. *Crops*, 4: 32–38

Wang, Z., B. Fang, J. Chen, X. Zhang, Z. Luo, L. Huang, X. Chen and Y. Li, 2010. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics*, 11: 726

You, F.M., N. Huo, Y.Q. Gu, M.C. Luo, Y. Ma, D. Hane, G.R. Lazo, J. Dvorak and O.D. Angerson, 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinform.*, 9: 253

Zhou, X., L. Ren, Q. Meng, Y. Li, Y. Yu and J. Yu, 2010. The next-generation sequencing technology and application. *Protein Cell*, 1: 520–536

(Received 10 October 2017; Accepted 22 November 2017)