ONLINE RESOURCES

Abundance and characteristics of microsatellite markers in Gansu zokor (*Eospalax cansus*), a fossorial rodent endemic to the Loess plateau, China

JUNHU SU[1,2], LIMIN HUA[1,2], JING WANG[3], DIANNE M. GLEESON[4], YANMING WEI[1], ROBYN HOWITT[5] and WEIHONG JI[1,2,6*]

[1]*College of Grassland Science, Gansu Agricultural University, Key Laboratory of Grassland Ecosystem of Ministry of Education, Pratacultural Engineering Laboratory of Gansu Province, Lanzhou 730070, People's Republic of China*
[2]*Gansu Agricultural University-Massey University Research Centre for Grassland Biodiversity, Gansu Agricultural University, Lanzhou 730070, People's Republic of China*
[3]*Lanzhou Vocational Technology College, Lanzhou 730070, People's Republic of China*
[4]*Institute for Applied Ecology, University of Canberra, Bruce ACT 2601, Australia*
[5]*Ecological Genetics Laboratory, Landcare Research, Private Bag 92170, Auckland, New Zealand*
[6]*Institute of Natural Sciences, Massey University, North Shore Mail Centre 0632, Auckland, New Zealand*

[Su J., Hua L., Wang J., Gleeson D. M., Wei Y., Howitt R. and Ji W. 2014 Abundance and characteristics of microsatellite markers in Gansu zokor (*Eospalax cansus*), a fossorial rodent endemic to the Loess Plateau, China. *J. Genet.* **93**, e25–e28. Online only: http://www.ias.ac.in/jgenet/OnlineResources/93/e25.pdf]

Introduction

Microsatellites, also known as simple sequence repeats (SSRs) or short tandem repeats (STRs) are regions widely distributed in the genome that contain tandem repeats of short nucleotide sequences. Microsatellite motifs are conserved in species. Their abundance, codominance, robustness and ease of amplification by polymerase chain reaction (PCR) make them good genetic markers for a wide range of studies (Geleta and Grausgruber 2012). They have long been utilized for population level studies in the fields of genetics, ecology, evolution conservation and management (Selkoe and Toonen 2006). Unlike nonspecific markers, such as amplified fragment length polymorphism (AFLP), inter-simple sequence repeat (ISSR) and random amplified polymorphic DNA (RAPD), microsatellites are single-locus DNA markers that require specific primers for PCR amplifications (Zane *et al.* 2002) and, therefore, designing primers is an important step in their development.

Early, development of microsatellites used traditional cloning based methods involving repeat enrichment, cloning and Sanger sequencing. That was both costly and labour intensive process (Schoebel *et al.* 2013). Other early methods included screening the available sequence data of a species for microsatellite loci and cross-species amplification of microsatellite loci from closely-related species. However, sequence information for nonmodel species is often not enough (Jian *et al.* 2012) and the percentage of loci that can be successfully cross amplified may decrease with increasing genetic distance (Jarne and Lagoda 1996).

Recent advances in sequencing technology such as the 454 GS Junior (Roche, Clifton, USA) whole-genome sequencing are valuable and cost effective means of searching for microsatellite for studies on nonmodel organisms (Schoebel *et al.* 2013). These next-generation sequencing technologies offer the ability to generate millions of base pair reads that are suitable for multiple applications (Hudson 2008). However, development of microsatellite markers for a species requires prior knowledge of its genomic sequences, lack of which makes this technology very expensive and time-consuming (Yu *et al.* 2009; Geleta and Grausgruber 2012). The discovery of microsatellites in the genome of a species and information on the occurrence and frequency of different microsatellites is valuable not only for development of microsatellite markers in that species (Yu *et al.* 2009), but also to understand its population genetics and demographic history. Unfortunately, little is known about occurrence and frequency of different microsatellites in many animals because of the limited availability of their genomic sequence data.

Gansu zokor (*Eospalax cansus,* Mysopalacinae) is a typical subterranean rodent species endemic to the Loess plateau in China inhabiting prairie, forest, meadow and farmland habitats. Zokors play an important role in ecosystem

*For correspondence. E-mail: j.j.weihong@massey.ac.nz.

Keywords. Zokor; subterranean rodents; microsatellites; genomic sequencing.

functions such as nutrient cycling, soil structure and vegetation composition and are considered to be ecosystem engineers (Zhang *et al.* 2003). They are also regarded as an important pest, destroying crops, competing with livestock, and causing soil erosion (Wang *et al.* 1993). Large scale culling of zokors using toxins has shown to have had significant negative impacts on biodiversity and potentially on other ecosystem aspects such as soil structure, nutrient and community structure (Singleton *et al.* 1999). Ecologically-based approaches were proposed as a potentially more efficient and environmentally friendly management (Singleton *et al.* 1999). Both ecological and behavioural knowledge are required to design such management regimes. However, the fossorial lifestyle of zokors makes it difficult to conduct ecological and behavioural studies (Norris *et al.* 2004) and existing information about this rodent, including life history and ecology are very limited. Consequently, molecular techniques become important tools for studying such aspects, and to date mitochondrial DNA (Su *et al.* 2014) and RAPDs (Tsvirka *et al.* 2011) have been used. Microsatellite markers have not previously been developed for any zokor species. To screen microsatellite loci for further ecological and behavioural study of zokors, we obtained 82.4 Mb (millions of base pairs) of sequence from the zokor genome through shotgun sequencing. In this study, we report the mining and characterization of SSRs in a Gansu zokor genome. Such information will be useful for further comparative studies to understand the function, evolution and mechanisms of microsatellites in subterranean rodents.

Materials and methods

Sample collection

One population of 30 *E. cansus* were collected from Tianzhu, Gansu province, NW China, during July 2012. All specimens were identified to species based on external characteristics and skull morphology (Allen 1940). Tail tissue samples were preserved in 95% ethanol for later DNA extraction.

DNA extraction, PCR amplification and sequencing

Two individuals' total genomic DNA was extracted using a Sangon Biotech (Shanghai, China) DNA Tissue Kit (animal no. SK8221) following the manufacturer's protocol. Shotgun sequencing of the two individuals DNA PCR products was performed on a 454 GS Junior (Roche Clifyon) at EcoGene® (Land-Care Research, Auckland, New Zealand).

Data analysis

The software Msatcommander 1.0.8 for Mac (Faircloth 2008) was used to screen raw data for microsatellite motifs (Schoebel *et al.* 2013). Separate searches were made for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide motifs with 14, seven, five, four, three and three repeats or more respectively, these searches resulted in seven output files per nucleotide repeat motifs. These output files for each nucleotide repeat motif were then combined into one dataset in Microsoft Excel (ver. 14.2.2). From each dataset, the number of motif types, lengths and variation, and frequency were calculated. We also calculated and compared the proportions of different repeat motifs and different lengths for each type of motif.

Results and discussion

Sequencing from the 454 GS Junior System platform yielded 186,832 reads (82,387,680 total bases) that ranged from 40 to 692 bp with the mean read length of 479 bp. The approximate size of the rodent genome is 2700 Mb (millions base pairs) based on the size of two related species *Mus musculus* and *Rattus norvegicus* (Rodentia). Allowing for sequence read overlap and some repeat sequences, our sequencing achieved coverage of up to ~3.1% (82.4 Mb) of the genome.

According to the size of repeat units, SSRs are classified into mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats (Tóth *et al.* 2000). Among the total SSRs, the amount of dinucleotide repeats is the most abundant accounting for 60.6% of total reads, followed by tetranucleotide (16.3%), pentanucleotide (7.3%), trinucleotide (6.6%), mononucleotide (5.2%) and hexanucleotide repeats (4.0%), respectively.

The number of repeat types is varied. Four types of mononucleotide repeat motifs were all recorded. Six, 20, 49, 80 and 88 types of dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide motifs were found, respectively. The highest repeat number is 32 (table 1). On an average there is one SSR per 5498.7 bp. The frequency of each repeat motif is shown in table 1.

Based on repeat architecture, SSRs are classified as perfect, imperfect or compound (Chambers and MacAvoy 2000). In a perfect SSR, the repeat motifs are arranged in a head to tail manner without any interruption by any base, while in an imperfect SSR there could be an insertion, deletion or substitution of bases in the repeated motifs. In the case of a compound SSRs the sequence contains two adjacent distinct SSRs separated by none to any number of base pairs. In this study, we have focussed on perfect and imperfect SSRs.

Table 1. The number, proportion and number of repeats for each type of SSRs in *E. cansus.*

Repeat types	Number of SSRs	Proportion	Number of repeats
Mononucleotide	779	5.2	14–28
Dinucleotide	9077	60.6	7–32
Trinucleotide	993	6.6	5–27
Tetranucleotide	2438	16.3	4–15
Pentanucleotide	1095	7.3	3–15
Hexanucleotide	601	4.0	3–8
Total	14983	100	3-32

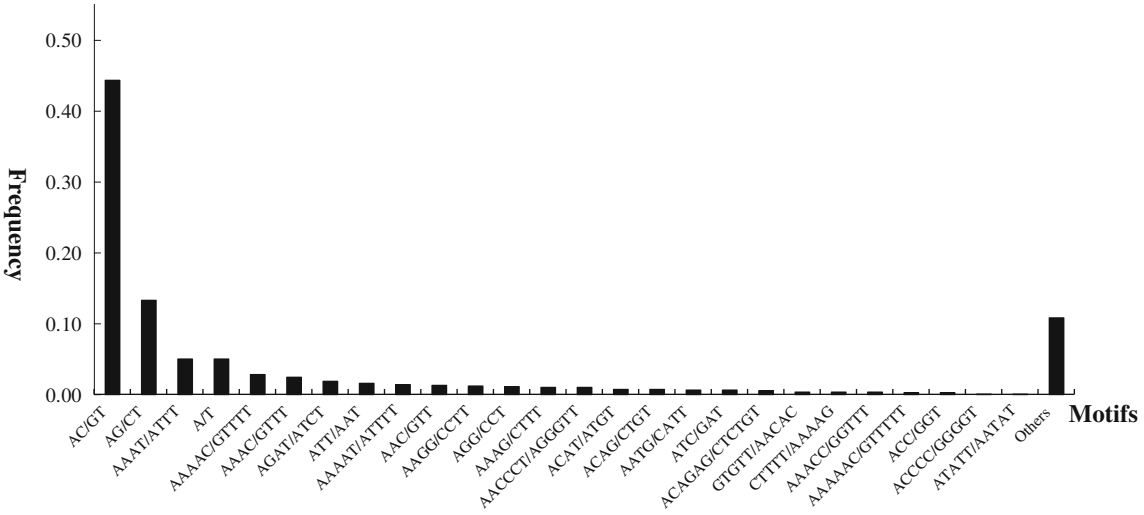

Figure 1. Proportion of different types of SSR motifs.

A total of 14,983 SSRs were identified from 82.4 Mb genome sequences, among which perfect SSR accounted for 92.29% (13828).

Among the mononucleotide motifs, the most abundant was A and T repeats, accounting for 97.2%. The most frequent dinucleotide motif was AC (3748). Among the six repeat motifs of dinucleotide, the motif GT showed a striking dominance followed by AG, while GC was completely absent. Similar distributions have been recorded in most animal species (Ju *et al.* 2005). Of the trinucleotide motifs, ATT motif was predominant in the genome followed by three other abundant motifs in descending order AGG, AAC and AAT. These four types of motifs accounted for 46.8 % of all trinucleotide motifs. The other 16 repeat motifs accounted for 53.2%. Of the 26 tetranucleotide motifs, the numbers of motif AAAT (381) were similar to motif of ATTT (380), followed by AAAC (179), GTTT (192), AGAT (136) and CTTT (122). These six types of motifs accounted for 57.0%. Of the 88 pentanucleotide motifs, GTTTT was dominant, accounting for 22.5% (246), followed by AAAAC (171, 15.6%) and AAAAT (103, 9.4%). There are 85 other motif types accounting for 52.5%. Among the 80 types of hexanucleotide repeats, AACCCT was dominant, accounting for 16.81%, and other 79 types of motifs were rarely found in all the sequences. The order of the dominant motifs is shown in figure 1.

The variation in length of an SSR motif was also considered for its usefulness as a marker (Ju *et al.* 2005). We chose 14 bp as a minimum length, with 14 bp being the dominant length of the Gansu zokor microsatellites, accounting for 22%. Among which, the dinucleotide motifs (1685) were the most abundant, followed by trinucleotide motifs (1319). The amount of the motifs with length over 21 bp was 38%, among which, 22 and 24 bp were dominant. The length distribution and frequency of microsatellites are shown in figure 2.

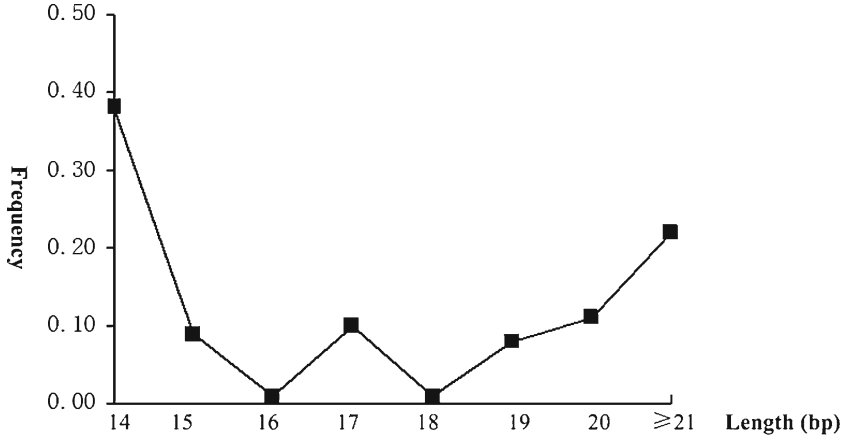

Figure 2. Length distribution and frequency of microsatellites.

Zokors are typical subterranean rodents with their fossorial lifestyle making ecological and behavioural studies difficult. Our results provide baseline information for the screening of microsatellites, which is now underway. This will enable detailed population genetic, ecology and behavioural studies to be conducted on this species and other zokor species.

Acknowledgements

Funding for this project was provided by the Ministry of Agriculture Public Benefit Research Foundation of China (grant no. 201203041), and the Science and Research Project of the Education Department of Gansu Province, China (grant no. 1102-02). We thank Duckchul Park, Diana Prada and Julia Allwood for laboratory assistance.

References

Allen G. M. 1940 *The mammals of China and Mongolia*. In Natural history of Central Asia (ed. W. Granger), Vol. 11, Part 2, pp. 913–937. American Museum of Natural History, New York, USA.

Chambers G.K. and MacAvoy E.S. 2000 Microsatellites: consensus and controversy. *Comp. Biochem. Physiol. B* **126**, 455–476.

Faircloth B.C. 2008 MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* **8**, 92–94.

Geleta N. and Grausgruber H. 2012 Classifying Ethiopan Tetraploid wheat (*Triticum turgidum* L.) landraces by combined analysis of molecular and phenotypic data. *Sci. Technol. Arts Res. J.* **1**, 1–9.

Hudson S.E. 2008 Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* **8**, 3–17.

Jarne P. and Lagoda P.J.L. 1996 Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* **11**, 424–429.

Jian Z.H., Liu X.S., Hu J.B., Chen Y.H. and Feng J.C. 2012 Mining microsatellite markers from public expressed sequence tag sequences for genetic diversity analysis in pomegranate. *J. Genet.* **91**, 353–358.

Ju Z., Wells M.C., Kazianis S., Rains J.D. 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.

Schoebel C.N., Brodbeck S., Buehler D., Cornejo C., Gajurel J., Hartikainen H. *et al.* 2013 Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. *J. Evol. Biol.* **26**, 600–611.

Norris R., Zhou K.Y., Zhou C.Q., Yang G., Kipatrick C.W. and Honeycutt R.L. 2004 The phylogenetic position of the zokors (Myospalacinae) and comments on the families of muroids (Rodentia). *Mol. Phylogenet. Evol.* **31**, 972–978.

Selkoe K.A. and Toonen R.J. 2006 Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* **9**, 615–629.

Singleton G.R., Hinds L.A., Leirs H. and Zhang Z.B. 1999 *Ecologically-based management of rodent*. Australian Centre for International Agricultural Research, Canberra, Australia.

Su J.H., Ji W.H., Wang J., Gleeson D.M., Zhou J.W., Hua L.M. and Wei Y.M. 2014 Phylogenetic relationships of extant zokors (Mysopalacidae) (Rodentia: Muridae) inferred from mitochondrial DNA sequences. *Mitochondrial DNA* **25**, 135–141.

Tóth G., Gáspári Z. and Jurka J. 2000 Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**, 967–981.

Tsvirka M.V., Pavlenko M.V. and Korablev V.P. 2011 Genetic diversity and phylogenetic relationships in the zokor subfamily myospalacinae (Rodentia, Muridae) inferred from RAPD-PCR. *Russ. J. Genet.* **47**, 205–215.

Wang Q.Y., Bian J.H. and Shi Y.Z. 1993 Influence of plateau zokors on the vegetation and soil nutrients in alpine meadow. *Acta Theriol. Sin.* **13**, 31–37.

Yu J.W., Dixit A., Ma K.H., Chung J.W. and Park Y.J. 2009 A study on relative abundance, composition and length variation of microsatellites in 18 underutilized crop species. *Genet. Resour. Crop. Evol.* **56**, 237–246.

Zhang Y.M., Zhang Z.B. and Liu J.K. 2003 Burrowing rodents as ecosystem engineers: the ecology and management of plateau zokors Myospalax fontanierii in alpine meadow ecosystems on the Tibetan Plateau. *Mammal Rev.* **33**, 284–294.

Zane L., Bargelloni L. and Patarnello T. 2002 Strategies for microsatellite isolation: a review. *Mol. Ecol.* **11**, 1–16.

Received 13 June 2013, in revised form 21 November 2013; accepted 22 November 2013
Published on the Web: 23 April 2014