Conservation Genet Resour
DOI 10.1007/s12686-013-0033-3

MICROSATELLITE LETTERS

Rapid development of microsatellite markers for the critically endangered Saiga (*Saiga tatarica*) using Illumina® Miseq next generation sequencing technology

Carsten Nowak · Steffen Zuther · Sergey V. Leontyev · Jutta Geismar

Received: 24 July 2013 / Accepted: 13 August 2013
© Springer Science+Business Media Dordrecht 2013

Abstract We developed 17 variable microsatellite markers for the threatened Saiga antelope. Using one Illumina Miseq lane resulted in 105,948 unique fragments containing a microsatellite motif. Eighty-one ordered primer pairs resulted in 26 analyzable fragments, of which 17 markers showed variability in at least one population from Kazakhstan. Number of alleles ranged from 2 to 11 and values of heterozygosity varied from 0.08 to 0.91 (H_O) and 0.08 to 0.88 (H_E). The markers are currently used to delineate conservation units and to help understanding annual migration dynamics in this species.

Keywords Short tandem repeats · Next generation sequencing · Kazakhstan · Antelopes · Antilopinae

The Saiga antilope *Saiga tatarica* (Linnaeus, 1766) was historically distributed over the vast steppe plains of Eurasia, until massive anthropogenic persecution repeatedly led to the near-extinction of this species within the past century. While some of the relict populations of the western subspecies (*S. t. tatarica*) currently increase in size across their mostly Kazakh distribution range, a lack of knowledge concerning its annual migration dynamics and population connectivity requires the use of high-resolution genetic markers in order to assist conservation management for this emblematic species.

Genomic DNA from liver samples of two Saiga specimens from the Ural population was extracted using the Blood and Tissue Kit (Qiagen). 5 µg of pooled DNA was used for library preparation and shotgun sequencing by Genterprise (Mainz, Germany) following the protocol of the manufacturer (Illumina, San Diego, CA). Samples where loaded on one lane of an Illumina® MiSeq250; yielding 19,879,594 sequences in one read (both ends). Paired fragments of both reads were aligned with Cope v 1.1.3 (Liu et al. 2012), with a lower limit of fragment overlapping = 1 in order to increase the number of aligned sequences. 6,245,870 sequences were filtered for microsatellite motifs using the software QDD2 beta (Meglecz et al. 2010) and subsequently aligned. 225,276 sequences contained at least one microsatellite (105,948 of these being unique). Primer pairs for 81,886 fragments were developed using standard approaches and a subsample of suggested sequences was manually controlled for overlapping results with PALfinder v 0.2.3 (Castoe et al. 2012), to double-check for specific and unique flanking regions. After manual check of primer positions 81 primer pairs were ordered, of which 45 amplified a fragment of the target length checked by standard gel electrophoresis. PCR was performed in a total volume of 10 µl with 3.4–3.8 µl template DNA, 100–150 nM of each primer, and 5 µl HotStarTaq MasterMix Kit (Qiagen, Hilden). Cycling

Electronic supplementary material The online version of this article (doi:10.1007/s12686-013-0033-3) contains supplementary material, which is available to authorized users.

C. Nowak (✉) · J. Geismar
Conservation Genetics Group, Senckenberg Research Institutes and Natural History Museums, Clamecystrasse 12, Gelnhausen, Germany
e-mail: Carsten.Nowak@senckenberg.de; cnowak@senckenberg.de

C. Nowak · J. Geismar
Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

S. Zuther · S. V. Leontyev
Association for the Conservation of Biodiversity of Kazakhstan (ACBK), Beybitshilik St., 18, 406 Office, Astana 010000, Kazakhstan

Table 1 Characteristics of the newly developed microsatellite markers for the Saiga

Locus	Accession no.	Primer sequence (5′–3′)	Repeat motif	Range	N_A	H_O/H_E	Dropout	False	PCR success
STa02[M4]	KF559258	F: (6FAM-)TGGGATCTTCCTGGATCAGA	$(AC)_9$	149–151	2	0.16/0.15	0.21	0	0.96
		R: GCAAATTGTGAAAGGTACTCCG			(2/2)	0.40/0.32			
STa03[M1]	KF559259	F: (NED-)GTGCCTTGTTAGGCTCCTGG	$(AC)_8$	87–91	2	0.16/0.15	0	0	0.99
		R: TTTCTCTGCTCCAGAATGACG			(2/1)	0/0			
STa04[M6]	KF559260	F: (VIC-)TCTAAATACTTCCCGAGTGCG	$(AC)_7$	101	1	0/0	0	0	0.9
		R: GAACAGCTGTGCTGATTTCAGAT				0/0			
STa05[M4]	KF559261	F: (6FAM-)CTTTGGAAGGACAACTGGGC	$(AC)_7$	195	1	0/0	0	0	0.99
		R: CAGCTCCTTCACACCTGGTC				0/0			
STa06[M5]	KF559262	F: (6FAM-)GTGACCAGGACACAGGATGG	$(AG)_7$	119	1	0/0	0	0	0.96
		R: GGACCTCAGCAACATGTCTGTA				0/0			
STa07 [M2]	KF559263	F: (VIC-)TTTGGCACACTGAGGTCTGA	$(AC)_7$	119	1	0/0	0.46	0	0.97
		R: TCTTATATGAACACATGCGGG				0/0			
STa10[M5]	KF559264	F: (VIC-)AACTCATTCAATGCATTTCTGG	$(AC)_6$	120–124	2	0.84/0.49***	0.05	0.05	0.98
		R: GCAGTGAGTAGAAGCTGGTGC			(2/2)	0.44/0.47			
STa14[M2]	KF559265	F: (PET-)GGACCCAGAAACCACTCCTT	$(AAAT)_8$	102–110	3	0.60/0.53	0.07	0	0.95
		R: CTCTTGCCCTTGCACAACAT			(3/3)	0.52/0.53			
STa16[M6]	KF559266	F: (PET-)CTGAGTGCTCTGTTGCATGG	$(AAGT)_5$	109–113	2	0.04/0.04	0	0	0.99
		R: CTTGTGGTCTCCACAGGCTT			(2/1)	0/0			
STa20[M3]	KF559267	F: (NED-)CATTCCTGTGGACATGGGTT	$(AAAT)_8$	123–139	5	0.28/0.32	0.18	0.03	0.96
		R: CTGTCACCTCACATGGCAAA			(5/3)	0.36/0.37			
STa26[M1]	KF559268	F: (PET-)GGTGGTGAAGAGGCAGTGAC	$(AGGT)_9$	155–179	4	0.64/0.55	0.06	0.05	0.98
		R: AAGGAGGATGAACGGTGGAT			(4/2)	0.56/0.49			
STa30[M6]	KF559269	F: (PET-)GGGCATCCTTTGAGTCACTG	$(AAAT)_6$	181–193	4	0.20/0.25	0.24	0	0.97
		R: CCATTGTGGTCTGGTCAGGT			(4/2)	0.04/0.04			
STa39[M4]	KF559270	F: (PET-)GCATAAGCTCTTCCCTGGGT	$(AGAT)_8$	177–201	7	0.84/0.79	0.09	0	0.97
		R: GTCTGTCTGCCTGCCTGTCT			(7/5)	0.8/0.69			
STa41[M2]	KF559271	F: (PET-)CAGTGACGCACAAACCACAT	$(AC)_{18}$	137–157	11	0.92/0.85	0.05	0.03	0.95
		R: ACCTGGGAGGAGAGTAACGG			(10/7)	0.74/0.73			
STa43[M3]	KF559272	F: (6FAM-)TGAGGGTCCTGGAGATTGAA	$(AC)_{13}$	97–117	12	0.80/0.88	0.15	0	0.95
		R: TCTGCAGAAGGATCAGGGAG			(11/6)	0.75/0.70			
STa44[M3]	KF559273	F: (PET-)TCCTTTCGTTGGTTTCCTCA	$(AC)_{12}$	196–200	3	0.24/0.33	0.04	0.01	0.58
		R: GCTATAGTCCATGGCGTTGC			(3/3)	0.21/0.19			
STa47[M6]	KF559274	F: (6FAM-)TCTGTGCTGACTTGCTCTCG	$(AC)_{10}$	121–125	3	0.48/0.47	0.13	0.04	0.97
		R: CACAAATTTGCCATGGGTTT			(3/3)	0.56/0.61			

Table 1 continued

Locus	Accession no.	Primer sequence (5′–3′)	Repeat motif	Range	N_A	H_O/H_E	Dropout	False	PCR success
STa55[M4]	KF559275	F: (VIC-)CCGATCTGAGGTTCAACAGG	$(AC)_8$	125–129	3	0.12/0.11	0.32	0	0.69
		R: AAAGGCTGTCCAAGGGACTC			(3/3)	0.27/0.24			
STa56[M6]	KF559276	F: (NED-)AAAGGGCCCTCTTCTCCTTT	$(AC)_8$	110–112	2	0.16/0.21	0	0	0.85
		R: GGCCTTACCTTCACCTTGCT			(2/2)	0.08/0.08			
STa59[M3]	KF559277	F: (VIC-)TTGAGACCTGACCACAGGGT	$(AC)_7$	129	1	0/0	0	0	0.99
		R: ACTCAATTCATCCTTCGGCA				0/0			
STa61[M5]	KF559278	F: (PET-)CCAGGGAAGTCTAGCTGGGT	$(AC)_7$	135	1	0/0	0	0	0.95
		R: TGGAAGGTGGCTGTGAAGAT				0/0			
STa69[M4]	KF559279	F: (NED-)CAGTTCTTTGCTGTAGAACTTAGGC	$(AAC)_6$	122	1	0/0	0	0	0.81
		R: TGCAGATGTGTAGTGTAAGCTATGG				0/0			
STa70[M1]	KF559280	F: (6FAM-)GTTCCAGGTTGGGAACACAG	$(AG)_6$	130	1	0/0	0	0	0.97
		R: CCCATGGTATTGAAGCCAAA				0/0			
STa71[M3]	KF559281	F: (VIC-)CTCCCAACCTCTCCTCCAGT	$(AC)_6$	186	1	0/0	0	0	0.65
		R: TAGGGTTGAACCAGCTTGGA				0/0			
STa72[M1]	KF559282	F: (NED-)TGAGGGATAAGGTGCCAAGA	$(ACGG)_8$	156–160	2	0.56/0.44	0.11	0.04	0.93
		R: TTCGTGGGAGCCAGTGATTA			(2/2)	0.36/0.34			
STa80[M4]	KF559283	F: (PET-)GGGAAGGAGGGCTGTCTTTA	$(AC)_9$	120–124	3	0.56/0.40	0.10	0.04	0.99
		R: ATTGCTAGCTCTGTGGGCTG			(2/3)	0.48/0.45			
Mean							0.10	0.01	0.92

Calculations are based on 50 individuals (25 from Betpak-Dala, first/above-listed values; and 25 from Ural). Given are locus names, assignment to the respective multiplex PCR reaction (M1-6), Genbank accession numbers, primer sequences, repeat motifs, allele ranges, number of alleles, observed (H_O) and expected (H_E) heterozygosities, dropout rates, rates of false alleles and PCR success rates

conditions consisted of a 15 min initial denaturation step at 95 °C, 34 amplification cycles following a touchdown protocol with 94 °C for 30 s, annealing step for 90 s, 72 °C for 60 s, and a final elongation step at 72 °C for 30 min. Annealing temperature was set to 60 °C for 4 cycles, 58 and 54 °C for 5 cycles, and 50 °C for 25 cycles. Two specimens for each primer pair were chosen for Sanger sequencing, resulting in 29 polymorphic primer pairs. Multiplex PCR was performed on blood samples (FTA cards) from 25 individuals each from two populations (Betpak-Dala and Ural).

Twenty-six markers resulted in scorable fragments, of which 17 loci showed considerable genetic variation (H_E = >0.1 in at least one of the populations), while nine loci were completely monomorphic among the tested samples (Table 1). Allelic variation was moderate at most loci, with only three loci exceeding five alleles summarized over both populations (STa39 = 7, STa41 = 11, STa43 = 12 alleles). Genotyping success was 92 % across samples and loci, which is a relatively high value given the relatively low quality of many blood samples used. Error rates were moderate at all loci but STa07 and STa55, which showed increased rates of allelic dropout.

The Illumina® MiSeq-approach resulted in a plethora of microsatellite-containing fragments, of which we extracted a robust set of 17 variable markers. Compared to previous approaches in our laboratory using classical enrichment protocols or 454 NGS approaches for microsatellite development (e.g. Geismar et al. 2011; Geismar and Nowak 2013), this approach had the best cost-value ratio. The markers are currently used to assess fine-scale genetic patterns in this species.

Acknowledgments We thank Genterprise company (Mainz) and the BiK-F laboratory centre (Frankfurt) for sequencing service. Annika Tiesmeyer and Jenny Wertheimer kindly assisted in data processing. Main funding comes from the UK Darwin Initiative, the Hessian Initiative for the Development of Scientific-Economic Excellence (LOEWE), and the Frankfurt Zoological Society.

References

Castoe TA, Poole AW, de Koning APJ et al (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. PLoS ONE 7:e30953. doi:10.1371/journal.pone.0030953

Geismar J, Nowak C (2013) Isolation and characterisation of new microsatellite markers for the stonefly *Brachyptera braueri* comparing a traditional approach with high throughput 454 sequencing. Conserv Genet Resour 5(2):413–416. doi:10.1007/s12686-012-9816-1

Geismar J, Sauer J, Haase P, Nowak C (2011) New microsatellite markers for the assessment of fine-scale dispersal patterns in the endangered montane caddisfly *Drusus discolor*. Conserv Genet Resour 3(4):605–607. doi:10.1007/s12686-011-9414-7

Liu B, Yuan J, Yiu S-M et al (2012) COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. Bioinformatics 28(22):2870–2874. doi:10.1093/bioinformatics/bts563

Meglecz E, Costedoat C, Dubut V et al (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics 26(3):403–404. doi:10.1093/bioinformatics/btp670