Development of SSR Markers from Transcriptomes for Orinus (Poaceae), an Endemic of the Qinghai–Tibetan Plateau

Authors: Liu, Yu-Ping, Su, Xu, Luo, Wen-Chun, Lv, Ting, Chen, Ke-Long, et al.

DEVELOPMENT OF SSR MARKERS FROM TRANSCRIPTOMES FOR *ORINUS* (POACEAE), AN ENDEMIC OF THE QINGHAI–TIBETAN PLATEAU[1]

YU-PING LIU[2,3], XU SU[2,4,8], WEN-CHUN LUO[5], TING LV[2,3], KE-LONG CHEN[3], AJ HARRIS[6], AND SAYED AFZAL SHAH[7]

[2]Key Laboratory of Medicinal Plant and Animal Resources in the Qinghai–Tibetan Plateau, School of Geography and Life Science, Qinghai Normal University, Xining 810008, People's Republic of China; [3]Key Laboratory of Physical Geography and Environmental Process in Qinghai Province, School of Geography and Life Science, Qinghai Normal University, Xining 810008, People's Republic of China; [4]Key Laboratory of Education, Ministry of Environments and Resources in the Qinghai–Tibetan Plateau, School of Geography and Life Science, Qinghai Normal University, Xining 810008, People's Republic of China; [5]State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou 730000, People's Republic of China; [6]Department of Botany, National Museum of Natural History, Smithsonian Institution, P.O. Box 37012, Washington, D.C. 20013-7012 USA; and [7]Department of Plant Science, Quaid-i-Azam University, Islamabad 44000 Pakistan

• *Premise of the study:* Transcriptomes were used to develop microsatellite markers for the plant genus *Orinus* (Poaceae), which comprises three species of grasses (*O. thoroldii*, *O. kokonoricus*, and *O. intermedius*) that are widely distributed in the Qinghai–Tibetan Plateau.

• *Methods and Results:* Primer pairs were developed for 16 high-quality simple sequence repeats (SSRs) using transcriptomes. SSRs were amplified in 248 individuals representing the three species of *Orinus*; the number of alleles per locus ranged from one to seven, with an average of 2.6. The expected and observed heterozygosity per locus varied from 0.00 to 0.83 and from 0.00 to 1.00, respectively, with respective mean values of 0.32 and 0.34.

• *Conclusions:* These newly developed SSR markers will be valuable for evaluating the population genetic structure of *Orinus* throughout its range.

Key words: next-generation sequencing; *Orinus*; orthologous gene; Poaceae; simple sequence repeat (SSR) marker.

Simple sequence repeats (SSRs) have been widely used for DNA fingerprinting, molecular-assisted breeding, detecting gene locations, genetic diversity analyses, and evolutionary studies because they are codominant, are highly polymorphic, can be amplified repeatably, and provide many informative sites distributed throughout the genome (Agarwal et al., 2008; Izzah et al., 2014). SSRs that are developed from transcribed RNA sequences, known as expressed sequence tag SSRs (EST-SSRs), can be developed cheaply and efficiently using next-generation sequencing technology (Simon et al., 2009). Previously, microsatellite markers developed from transcriptomes have primarily been for woody and medicinal plants (e.g., Liu et al., 2014;

Kim et al., 2017). Few microsatellite markers have been developed to study endemic alpine grasses.

Here, we show the utility of SSR markers derived from transcriptome assemblies to detect genomic variation in the three species of *Orinus* Hitchc. (Poaceae), which are endemic grasses that occur at high elevations in xeric, alpine areas of the Qinghai–Tibetan Plateau (QTP) (Su et al., 2015). *Orinus* comprises *O. thoroldii* (Stapf ex Hemsl.) Bor, which is mainly distributed in the western QTP, *O. kokonoricus* (K. S. Hao) Tzvelev, which occurs in the eastern QTP, and *O. intermedius* X. Su & J. Quan Liu, which is a hybrid of *O. thoroldii* and *O. kokonoricus* and has a range in the southeastern QTP (Su et al., 2015, 2017). *Orinus* has ecological and conservation value for its role in soil stabilization in the QTP, especially due to its expansive root system (Su et al., 2013), and this genus may represent a good system for elucidating the timing and mechanisms of desertification in the QTP (Su et al., 2015). Therefore, developing SSR markers for *Orinus* is a first step toward utilizing the population history of its species to better understand the origins of the genus and to improve conservation efforts for desert habitats in the QTP. Here, we develop microsatellite markers in *Orinus* using transcriptomes obtained via Illumina paired-end sequencing.

TABLE 1. Characteristics of 16 polymorphic SSR markers developed in three species of *Orinus*.

Locus		Primer sequences (5'–3')	Repeat motif	Allele size range (bp)	T_a (°C)	GenBank accession no.	GenBank accession of best BLAST hit	Organism of best BLAST hit	E-value
Ori1	F:	AGAAGATCATTGCTTTCAACC	$(CT)_9$	136–171	59	KY852238	KP856157.1	*Quercus variabilis*	8.0×10^{-7}
	R:	TTTTCCATCGGCACTAGCT							
Ori3	F:	GGAAGGGAAAGACGAAGAAGC	$(GGA)_9$	160–178	60	KY705074	NM_001150415.1	*Zea mays*	3.0×10^{-12}
	R:	AATCGACTGCTGTATGCAAAGG							
Ori4	F:	AAGAAACAAAAAGTGGTGAGG	$(TGA)_9$	278–305	58	KY705075	EU976105.1	*Zea mays*	5.0×10^{-28}
	R:	TAAGGGTGTTTCTGCTGGCCTA							
Ori5	F:	GAGCTAAGCAGAGCACTCATCCT	$(GAA)_7$	235–250	59	KY705076	NM_001151471.1	*Zea mays*	4.0×10^{-36}
	R:	CTTAGCCTTCCCTTGCCGATA							
Ori12	F:	GGTATGGGTAGGACTGCAGCTTT	$(TAA)_{12}$	243–267	60	KY705077	XM_004975213.3	*Setaria italica*	1.0×10^{-29}
	R:	TAATACAAAACTTGGACAGCG							
Ori13	F:	CATTCTCCCATCTGCTCGTCTC	$(CCG)_6$	120–161	60	KY852239	XM_004973362.2	*Setaria italica*	1.0×10^{-10}
	R:	TTCCGAACGATTTGAGCG							
Ori14	F:	GTGTATATGAAACGGATGGAACAC	$(AT)_{10}$	155–204	58	KY852240	CR382128.1	*Yarrowia lipolytica*	5.0×10^{-3}
	R:	AATAAAGATGCATGTACTCGTCC							
Ori15	F:	GCAAAAGGCATAACCTAACCTAAAC	$(AAT)_{10}$	204–228	59	KY705078	AC116411.7	*Mus musculus*	3.0×10^{-12}
	R:	AGCATCCAATACAATACTCTTCGAC							
Ori17	F:	GGAGAGTTCAGACGGACAGACA	$(TCC)_{12}$	297–321	60	KY705079	NM_001175683.1	*Zea mays*	3.0×10^{-43}
	R:	ATCCCGACCACTACAGCCTT							
Ori21	F:	GTTTCGCCTGCGTCCTTG	$(CTC)_{12}$	249–270	60	KY705080	AC104200.12	*Mus musculus*	1.0×10^{-10}
	R:	AGTGGCATCCATCAAAACAAGA							
Ori31	F:	GCCAGCTGCTTCTTGCGAC	$(TCT)_7$	182–221	69	KY852241	XM_004967825.1	*Setaria italica*	1.0×10^{-81}
	R:	CTCGAGGAGGAAGAGGACGA							
Ori32	F:	AGCAAGCATACCTAATGTTTTG	$(TG)_{13}$	294–324	59	KY705081	XM_004975367.2	*Setaria italica*	2.0×10^{-21}
	R:	CACGGCGTTCATATTCGG							
Ori33	F:	TTCTTGACGAGCTTGACCCT	$(TCC)_6$	184–222	60	KY852242	XM_004982769.1	*Setaria italica*	8.0×10^{-83}
	R:	CGTCGTGCTCCAACTCCCT							
Ori36	F:	AGAAGGGTGGAGTCGATCATG	$(AG)_{10}$	205–215	59	KY705082	CP018161.1	*Oryza sativa*	2.0×10^{-14}
	R:	CAACAAGCAACACGATACTGATAGA							
Ori38	F:	GCATTCTGTCCAGTTTCAAGC	$(CT)_{21}$	154–192	59	KY705083	AY486591.1	*Hevea brasiliensis*	2.0×10^{-15}
	R:	ACTTGGCGCCATCTGTTT							
Ori40	F:	TCAGAGATTTGGTGTAAGTTGCTG	$(TC)_7$	141–188	60	KY852243	KF785779.1	*Nandina domestica*	5.0×10^{-3}
	R:	GCTTGCAAGAATCGAATTAGAGA							

Note: T_a = optimal annealing temperature.

METHODS AND RESULTS

Twenty-five to 30 individuals were collected from three populations each of the three *Orinus* species (248 total collections) (Appendix 1). In addition, representative individuals of *O. thoroldii* and *O. kokonoricus* were collected (Appendix 1), from which fresh leaves were obtained; these were immediately frozen in liquid nitrogen in the field and later stored at −80°C prior to RNA extraction. For all collections, voucher specimens were deposited at the Herbarium of the Northwest Plateau Institute of Biology (HNWP) (Appendix 1).

Total RNA and genomic DNA were extracted using a cetyltrimethylammonium bromide (CTAB) procedure (Ghangal et al., 2009). The RNA was quantified and its quality assessed using NanoDrop 2000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). The total RNA samples were purified to remove poly(A) tags using approximately 5 µg of RNAs and (dT)-conjugated beads (Life Technologies, Carlsbad, California, USA), and purified RNAs were divided into 200-bp fragments using divalent cations at 75°C. The first strand of cDNA was synthesized with reverse transcriptase and random hexamer primers, and the second strand was synthesized by RNase H (Invitrogen, Ghent, Belgium) and *Taq* DNA polymerase I (New England BioLabs, Ipswich, Massachusetts, USA). Finally, the cDNAs representing the transcriptome were sequenced on an Illumina (Solexa) Genome Analyzer II (Illumina, San Diego, California, USA). All sequence information has been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (Bioproject no. PRJNA385721).

SSR primers were designed using the transcriptomes of *O. thoroldii* and *O. kokonoricus*. To accomplish this, 79,111,257 and 67,617,602 raw reads were first filtered from *O. thoroldii* and *O. kokonoricus*. As a result, 60,124,556 and 50,037,026 high-quality reads, respectively, were obtained for de novo assembly. De novo assembly was performed in Trinity 2.2 with default parameters (Grabherr et al., 2011). From Trinity, 23,029 high-quality contigs were obtained from *O. thoroldii* and 24,086 from *O. kokonoricus*, representing total lengths of 20,808,832 and 22,281,570 bp, respectively, with an average size of 903 bp (N50 = 1188 bp) and 925 bp (N50 = 1203 bp). The assembly of each species was used to map the reads of the other using Bowtie 2 (Langmead and Salzberg, 2012) to identify an orthologous set of genes. Within the orthologous set, we searched for candidate SSRs using MISA 4.0 (http://pgrc.ipk-gatersleben.de/misa) and also identified single-nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms (indels) using SAMtools 1.4 (Li et al., 2009). Indels were selected as SSR candidates only if the indels appeared informative between the two species and could be confidently aligned at their 3′ and 5′ ends. In total, 58 polymorphic candidate loci were recovered from *O. thoroldii* and 52 from *O. kokonoricus*. For each locus, primers were designed in Primer3 (Rozen and Skaletsky, 2000), and primers with binding sites containing SNPs were rejected. Thus, a total of 50 primers suitable for both species of *Orinus* were found. Ten individuals were selected from each of three species of *Orinus* to test the 50 primers. Amplification was performed using a standard 25-µL PCR reaction containing 1.00 µL of template DNA, 0.20 mM MgCl$_2$, 0.25 mM dNTPs, 2.00 µM of each primer, 2.50 µL of 10× PCR buffer, 0.25 µL of *Taq* DNA polymerase (5 U·µL^{-1}; TaKaRa Biotechnology Co., Dalian, Liaoning, China), and distilled water up to the final volume. The amplification was carried out under the following thermocycling protocol: enzyme initiation at 94°C for 5 min; followed by 36 cycles of denaturation at 94°C for 50 s, annealing at 49–58°C for 50 s (Table 1), and 72°C for 1 min; and a final extension at 72°C for 10 min. Amplification success was determined by viewing the results in a 1.5% (w/v) agarose gel. Sixteen primers showed successful amplification in all three *Orinus* species, and all of them possessed clear polymorphisms (Table 1).

Subsequently, the 16 primers were used to carry out fluorescence-based genotyping for all 248 sampled individuals of *Orinus*. Fluorescence-based genotyping was performed using a modification of the method presented in Hayden et al. (2008). In brief, the forward primers for the 16 SSRs were labeled with 6-FAM fluorescent tags (Applied Biosystems, Foster City, California, USA) and PCR reactions were performed as described above. The labeled products were detected on an ABI 3730XL sequencer with the GeneScan 500 LIZ Size Standard (Applied Biosystems). The profiles of the amplified loci were examined using GeneMapper 3.7 (Applied Biosystems), and peaks were scored manually by visual inspection. Each SSR marker was characterized by calculating three measures of genetic diversity in GENEPOP 4.2 (Rousset, 2008): number of alleles per locus, observed heterozygosity, and expected heterozygosity (Table 2). By these measures, the markers were highly polymorphic, with the number of alleles ranging from one to seven within this genus, with an average of 2.6 alleles per locus. The expected and observed heterozygosity ranged from 0.00 to 0.83 and 0.00 to 1.00, respectively, with respective mean values of 0.32 and 0.34 (Table 2).

TABLE 2. Genetic diversity statistics for each sampled population of the three *Orinus* species based on 16 pairs of SSR primers.[a]

| | O. thoroldii | | | | | | | | | O. kokonoricus | | | | | | | | | | | | | | O. intermedius | | | | | |
| | Ge'er (N = 30) | | | Renbu (N = 28) | | | Zhongba (N = 28) | | | Gonghe (N = 27) | | | Nangqian (N = 25) | | | Bianba (N = 26) | | | Aba (N = 30) | | | Rangtang (N = 26) | | | Mangkang (N = 28) | | |
Locus	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]	A	H_o	H_e[b]
Ori1	1	0.000	0.000	3	0.600	0.540	6	0.533	0.584	4	0.571	0.498	1	0.000	0.000	3	0.200	0.416	2	0.100	0.089	1	0.000	0.000	2	1.000	0.500
Ori3	2	0.143	0.133	2	0.111	0.278	2	0.400	0.320	1	0.000	0.000	2	0.125	0.117	2	0.250	0.219	1	0.000	0.000	2	0.500	0.375	2	0.143	0.133
Ori4	2	0.333	0.278	2	0.222	0.346	2	0.600	0.420	2	0.429	0.337	1	0.000	0.000	3	1.000	0.580	2	0.500	0.375	2	0.500	0.375	3	0.143	0.255**
Ori5	3	0.667	0.486	4	0.778	0.543	4	0.600	0.665	2	0.143	0.133	1	0.000	0.000	2	0.400	0.320	3	0.333	0.569	1	0.000	0.000	2	0.143	0.133
Ori12	3	0.333	0.292	2	0.111	0.105	1	0.000	0.320**	2	0.286	0.490	1	0.000	0.564*	3	0.800	0.660	2	0.500	0.375	2	0.250	0.219	2	0.143	0.133
Ori13	2	0.400	0.320	2	0.111	0.105	2	0.000	0.000	2	0.200	0.480	3	0.333	0.287	3	0.467	0.520	2	0.400	0.320	1	0.000	0.000	3	0.500	0.457
Ori14	2	1.000	0.500	2	0.000	0.444	2	0.000	0.000	2	0.100	0.092	1	0.000	0.000	2	0.200	0.320	1	0.000	0.000	2	0.067	0.180	6	0.067	0.447
Ori15	1	0.000	0.000	2	0.222	0.198	2	0.100	0.095	1	0.000	0.000	2	0.000	0.320	2	0.200	0.180	2	0.500	0.375	3	0.000	0.625*	6	0.857	0.776
Ori17	3	0.333	0.569	2	0.111	0.105	2	0.300	0.455	2	0.000	0.000	2	0.000	0.124	3	0.800	0.540	1	0.000	0.000	2	0.250	0.219	2	0.286	0.245
Ori21	2	0.500	0.486	4	0.667	0.623	2	0.400	0.320	2	0.429	0.337	4	0.625	0.578	2	1.000	0.500*	3	1.000	0.500	4	0.750	0.656	4	0.286	0.622
Ori31	3	0.500	0.467	4	0.300	0.610	2	0.100	0.095	4	0.571	0.497	6	0.067	0.447	2	0.100	0.095	3	0.500	0.580	3	0.500	0.395	3	0.500	0.457
Ori32	2	0.500	0.375	2	0.111	0.105	1	0.000	0.000	2	0.286	0.490	7	0.750	0.828*	3	0.600	0.580	1	0.000	0.000	2	0.067	0.064	2	0.000	0.245**
Ori33	2	0.100	0.095	4	0.500	0.665	2	0.100	0.095	2	0.200	0.480	5	0.643	0.676	2	0.267	0.320	1	0.000	0.000	2	0.000	0.124	3	0.067	0.127
Ori36	2	0.667	0.444	2	0.111	0.105	2	0.100	0.095	2	0.143	0.133	5	0.500	0.375	1	0.000	0.444	1	0.000	0.000	2	0.000	0.375*	2	0.000	0.245**
Ori38	3	0.833	0.667	6	0.667	0.784*	5	0.700	0.775	5	0.714	0.622	5	0.500	0.500	2	0.800	0.480	3	1.000	0.625	4	0.750	0.656	6	1.000	0.776
Ori40	1	0.000	0.000	4	0.571	0.497	2	0.400	0.320	4	0.267	0.333	1	0.000	0.000	2	0.111	0.105	1	0.000	0.000	2	0.067	0.064	4	0.200	0.218

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity.
[a] Locality and voucher information for the populations are provided in Appendix 1.
[b] Significant deviation from Hardy–Weinberg equilibrium after correction for multiple tests (*$P < 0.05$, **$P < 0.01$).

CONCLUSIONS

We have developed 16 polymorphic SSR markers from two cDNA libraries for investigating population structure in *Orinus*. These markers amplified easily and showed considerable polymorphisms for 248 individuals from three populations each of the three species of *Orinus*. These markers represent valuable new tools that will facilitate the development of *Orinus* as a model for understanding the origins and phylogeographic processes of the alpine desert of the QTP.

LITERATURE CITED

Agarwal, M., N. Shrivastava, and H. Padh. 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27: 617–631.

Ghangal, R., S. Raghuvanshi, and P. C. Sharma. 2009. Isolation of good quality RNA from a medicinal plant sea buckthorn, rich in secondary metabolites. *Plant Physiology and Biochemistry* 47: 1113–1115.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Hayden, M. J., T. M. Nguyen, A. Waterman, G. L. McMichael, and K. J. Chalmers. 2008. Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Molecular Breeding* 21: 271–281.

Izzah, N. K., J. Lee, M. Jayakodi, S. Perumal, M. Jin, B. S. Park, K. Ahn, et al. 2014. Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. *capitata* L.) and construction of an EST-based genetic map. *BMC Genomics* 15: 149.

Kim, B. Y., H. S. Park, S. Kim, and Y. D. Kim. 2017. Development of microsatellite markers for *Viscum coloratum* (Santalaceae) and their application to wild populations. *Applications in Plant Sciences* 5: 1600102.

Langmead, B., and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Liu, B. B., T. Bin, H. Ma, Z. Q. Lu, Q. Qiu, K. S. Mao, and J. Q. Liu. 2014. Development and characterization of EST-SSR markers in *Ostryopsis* (Betulaceae). *Applications in Plant Sciences* 2: 1300062.

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

Simon, S. A., J. Zhai, R. S. Nandety, K. P. McCormick, J. Zeng, D. Mejia, and B. C. Meyers. 2009. Short-read sequencing technologies for transcriptional analyses. *Annual Review of Plant Biology* 60: 305–333.

Su, X., W. Yue, and J. Q. Liu. 2013. Germplasm collection and preservation of *Orinus* (Poaceae) in the Qinghai-Tibet Plateau. *Plant Diversity and Resources* 35: 343–347.

Su, X., G. L. Wu, L. L. Li, and J. Q. Liu. 2015. Species delimitation in plants using the Qinghai-Tibetan Plateau endemic *Orinus* (Poaceae: Tridentinae) as an example. *Annals of Botany* 116: 35–48.

Su, X., Y. P. Liu, G. L. Wu, W. C. Luo, and J. Q. Liu. 2017. A taxonomic revision of *Orinus* (Poaceae) with a new species, *O. intermedius*, from the Qinghai-Tibet Plateau. *Novon* 25: 206–213.

Appendix 1. Locality information for the populations of *Orinus* used in this study.

Species	N	Population	Geographic coordinates	Altitude (m)	Voucher no.[a]
O. thoroldii	30	Ge'er, Xizang, China	31°36′13.7″N, 80°22′15.6″E	4570	*X. Su 2011055*
	28	Renbu, Xizang, China	29°18′0.4″N, 89°46′7.3″E	3767	*X. Su 2011018*
	28	Zhongba, Xizang, China	29°41′7.9″N, 84°8′48.1″E	4563	*X. Su 2011044*
	1	Gongga, Xizang, China	29°0′27.0″N, 85°26′48.8″E	4687	*X. Su 2011078*[b]
O. kokonoricus	27	Gonghe, Qinghai, China	36°11′3.0″N, 100°59′16.9″E	2826	*X. Su 2012040*
	25	Nangqian, Qinghai, China	32°32′50.6″N, 96°11′45.2″E	4119	*X. Su 2011074*
	26	Bianba, Xizang, China	30°58′40.3″N, 94°43′35.3″E	3597	*X. Su 2013083*
	1	Gonghe, Qinghai, China	36°21′26.3″N, 100°43′5.8″E	3130	*X. Su 2013008*[b]
O. intermedius	30	Aba, Sichuan, China	32°45′26.7″N, 102°33′3.8″E	3319	*X. Su 2012003*
	26	Rangtang, Sichuan, China	31°46′16.2″N, 100°58′57.1″E	3478	*X. Su 2012007*
	28	Mangkang, Xizang, China	29°32′27.2″N, 98°15′3.3″E	3507	*X. Su 2012016*

Note: N = number of individuals sampled.

[a] All voucher specimens were deposited at the Herbarium of the Northwest Plateau Institute of Biology (HNWP), Chinese Academy of Sciences, Xining, Qinghai Province, China.

[b] These representative individuals of *Orinus thoroldii* and *O. kokonoricus* were only used for RNA extraction.