Development of SSR Markers for a Tibetan Medicinal Plant, Lancea tibetica (Phrymaceae), Based on RAD Sequencing

Authors: Tian, Zunzhe, Zhang, Faqi, Liu, Hairui, Gao, Qingbo, and Chen, Shilong

Source: Applications in Plant Sciences, 4(11)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1600076

Applications in Plant Sciences 2016 4(11): 1600076

PRIMER NOTE

DEVELOPMENT OF SSR MARKERS FOR A TIBETAN MEDICINAL PLANT, *LANCEA TIBETICA* (PHRYMACEAE), BASED ON RAD SEQUENCING[1]

ZUNZHE TIAN[2,3], FAQI ZHANG[2,4], HAIRUI LIU[2,3], QINGBO GAO[2], AND SHILONG CHEN[2,4]

[2]Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Xining 810001, People's Republic of China; and [3]University of Chinese Academy of Sciences, Beijing 100039, People's Republic of China

- *Premise of the study: Lancea tibetica* (Phrymaceae), a Tibetan medicinal plant, is endemic to the Qinghai–Tibet Plateau. The over-exploitation of wild *L. tibetica* has led to the destruction of many populations. To enhance protection and management, biological research, especially population genetic studies, should be carried out on *L. tibetica*. Simple sequence repeat (SSR) markers of *L. tibetica* were developed to analyze population diversity.
- *Methods and Results:* Four thousand four hundred and forty-one SSR loci were identified for *L. tibetica* based on restriction-site associated DNA (RAD) sequencing on the Illumina HiSeq platform. One hundred SSR loci were arbitrarily selected for primer design, and 38 of them were successfully amplified. These markers were tested on 56 individuals from three populations of *L. tibetica*, and 10 markers displayed polymorphisms. The total number of alleles per locus ranged from three to eight, and observed and expected heterozygosities ranged from 0.200 to 1.000 and 0.683 to 0.879, respectively. We tested for cross-amplification of these 10 markers in the related species *L. hirsuta* and found that nine could be successfully amplified.
- *Conclusions:* The SSR markers characterized here are the first to be developed and tested in *L. tibetica*. They will be useful for future population genetic studies on *L. tibetica* and closely related species.

Key words: *Lancea tibetica*; Phrymaceae; population diversity; RAD sequencing; simple sequence repeat (SSR).

Lancea tibetica Hook. f. & Thomson (Phrymaceae) is an herb endemic to the Qinghai–Tibet Plateau. It usually grows in grasslands, sparse forests, or ravines at altitudes of 2000–4500 m. As a traditional Tibetan medicinal plant, it has been used in the treatment of leukemia, intestinal angina, heart disease, and cough (Hong et al., 1998). Investigations into the chemical constituents of *L. tibetica* have resulted in the isolation of phenylpropanoid glycosides and lignans, which contribute to the species' antioxidant effects (Song et al., 2011). To increase production of traditional medicine from this species, the harvest of wild populations has been greatly expanded. The serious depletion of *L. tibetica* through over-collecting has led to a need for proper management and a conservation plan to ensure its sustainable use into the future. A thorough study at the population level is required to evaluate the extent of remaining genetic resources and to inform management plans.

Simple sequence repeat (SSR) markers are widely used for population genetic studies due to their codominant nature, polymorphisms, and reproducibility (Litt and Luty, 1989). The development of SSR markers for *L. tibetica* will enable us to assess genetic diversity and contribute to a conservation strategy. The restriction-site associated DNA (RAD) method was proposed by Miller et al. (2007) as a reliable approach that reduces genome complexity. RAD sequencing has been successfully applied in many organisms, including crop species like barley (Chutimanitsakun et al., 2011). In this paper, we describe the process of isolation and characterization of 10 polymorphic SSR markers from *L. tibetica* based on RAD sequencing.

METHODS AND RESULTS

***Plant materials and DNA extraction*—**In total, 56 *L. tibetica* individuals from three natural populations (YD, QML, and MY) were sampled (Appendix 1). Fresh leaves were collected and dried in silica gel; the voucher specimens are deposited in the Herbarium of the Northwest Institute of Plateau Biology (HNWP), Chinese Academy of Sciences, Xining, Qinghai Province, China. Total genomic DNA was extracted from dried leaves with the cetyltrimethylammonium bromide (CTAB) method (Doyle, 1987).

***RAD library preparation and sequencing*—**We selected one individual from each of the populations and pooled them. Subsequently, the RAD library was constructed based on published methods (Barchi et al., 2011). The library was quantified with Qubit (Invitrogen, Eugene, Oregon, USA) and sequenced using the Illumina MiSeq platform (Illumina, San Diego, California, USA).

[1]Manuscript received 27 June 2016; revision accepted 14 September 2016.

The authors thank Richard Gornall for helpful comments on the manuscript and for helping to improve the English. This work was supported by the National Natural Science Foundation of China (31400322, 31270270), Applied Basic Research Programs of Qinghai Province (2016-ZJ-761), International Scientific and Technological Cooperation Projects of Qinghai Province (2014-HZ-812), the Chinese Academy of Sciences (CAS) Youth Innovation Promotion Association (2016378), CAS "Light of West China" Program, and Lancang Watershed Conservation Fund of Shanshui Conservation Center.

[4]Author for correspondence: fqzhang@nwipb.cas.cn; slchen@nwipb.cas.cn

doi:10.3732/apps.1600076

TABLE 1. Characteristics of 38 microsatellite loci developed in *Lancea tibetica.*

Locus	Primer sequences (5′–3′)	Repeat motif	Fragment size (bp)	T_a (°C)	GenBank accession no.
LT4	F: ATTGATTGATTCACGTTCCAAAT	$(TA)_6$	132	54	KU764519
	R: TGAAAATGAATAACTTGGGGATCT				
LT7	F: TTTGGAAAGCATGATCTACCACT	$(AAT)_5$	151	56	KU764520
	R: TTTCTGGACTGTTGTAATCTTGAAA				
LT9	F: GGATTTCTAAGTGCAATCCTCAA	$(GA)_7$	140	51	KU764521
	R: CATCACTCACCAAATGAAAGACA				
LT10	F: AATTGTTCCAGGTATGCAGTGTT	$(TG)_6$	155	51	KU764522
	R: CTATTCTGCAAGTTAATGCAGGG				
LT12	F: GTAGACATTTTTGCAGCACCTCT	$(CAT)_5$	150	51	KU764523
	R: ATGAGGACTCAAAGACAGCTCAG				
LT15	F: CTTATAACCTATCGTTCTCCGGC	$(AG)_7$	138	54	KU764524
	R: ATTTCGCTCTCTCTTTCACACAC				
LT16	F: TGTATTGTCAATGGAAGAGGCAT	$(AAG)_4$	155	51	KU764525
	R: GAATGAGATGCTCCACTAACCAC				
LT18	F: AACAAGTTTATGCAAGGAGGAGA	$(TCT)_5$	160	51	KU764526
	R: CCCAAGTCCCAAATGATATAAAA				
LT25	F: GATGCCAAGGAATTGTTATATGC	$(TA)_7$	153	51	KU764527
	R: TTTCTAGAAGTCGGAGCTGTCC				
LT28	F: AACAGCAATGGCAATATGGTATC	$(TA)_9$	158	56	KU764528
	R: AACTGTTCAAGTTGGCAAAACAT				
LT6	F: TCTATCGGTGCTAAAACACCTTC	$(GT)_{11}$	153	51	KX377923
	R: CTCATCCTCATCATACCGATCAT				
LT11	F: TTGCCCTTATGTTTATCAAGGAA	$(CT)_6$	144	51	KX377922
	R: CACAGAAGAAGGATGAGGAGAAA				
LT17	F: GACAGAACCCCTCTCTGAATCTT	$(TAC)_4$	152	51	KX377921
	R: GCGCCATAAGGTATAGCACTTC				
LT19	F: ATTACCAACTTTCAACCAAAGCA	$(CCT)_4$	159	51	KX377920
	R: GCTTGTTGTCTTCTTTCCCAATA				
LT26	F: TGAGCAGGTGCCTTTATTGTTAT	$(CT)_7$	149	58	KX377919
	R: TCAGCAGATCCTTATTATTTGTGC				
LT30	F: AGGTCAGGAACAACAATACCTCA	$(AG)_8$	158	49	KX377918
	R: CTATATATCTTGCTTGCAAATCCG				
LT40	F: TCTCTCTTTCTTCCCTCTCCATC	$(GA)_{10}$	148	55	KX377917
	R: AAATCAAGGAATCTGTGCAATGT				
LT45	F: GGAGAGGGAAAAGAAGAAGAAGA	$(GAA)_4$	78	53	KX377916
	R: TACCAATGTAGCCGGAAATAAAG				
LT49	F: AACGAAAATACTTTCCGTCTACAAA	$(AT)_8$	121	52	KX377915
	R: CTTGTTCTGGTCTGGTTTAAGGA				
LT60	F: CTATAAATACCTCCCTCCCCCTC	$(CT)_7$	127	51	KX377914
	R: GTTTACGAGCACTCCTAGGTGAA				
LT61	F: TGCCTATTCTTTACAAGAGCACA	$(AT)_6$	117	52	KX377913
	R: TTAATTGTAAATCGCAAAAACCC				
LT65	F: TAAATGGTTTGCATCTTGGAAAT	$(TAT)_5$	119	51	KX377912
	R: GCAAAAATAAGTTTAACCGCGTA				
LT66	F: TTTTGCTTTGTTGGATTCTTGAT	$(GAT)_4$	148	53	KX377911
	R: GCATCCTAAACTTACCGTTTTCA				
LT67	F: TTTTGCAGGTTTAAGACAAAGGA	$(GT)_7$	106	51	KX377910
	R: TACATCGACACTTTTCAATCCAA				
LT69	F: AGCGTAAGAAGATGATAGAAGGG	$(AAT)_4$	145	51	KX377909
	R: TGATCCTATTAGAGTTGCAAACG				
LT72	F: TCAAACAAGCATGGGAGTACTTT	$(AGA)_4$	140	52	KX377908
	R: TTTGAACGAATTAGAGGAGGACA				
LT75	F: ATACCAACCTGTGGCGTATATTG	$(AT)_6$	150	51	KX377907
	R: TGAAGATGTAAGAACAACCAGCA				
LT77	F: GATCATGTCCCATCAAATTCAAC	$(AT)_{14}$	157	51	KX377906
	R: TTGTGTTATCTCCTGCGGTACTC				
LT79	F: GAAGAGGTCAAGGCAAAGATACA	$(AG)_6$	144	51	KX377905
	R: TGAAATCGAGAATTGAAGAACAAA				
LT81	F: TAGAAAAGTGAGGAATGGGACAA	$(AT)_6$	157	50	KX377904
	R: TGGTTTAGGAAATTTAACGATTGA				
LT83	F: CATAATTTTGTGAGATCTTGGGC	$(TTG)_4$	145	52	KX377903
	R: AATTCTCCAAATGCAGATGATGT				
LT86	F: TACTTGCTCCCCAAGTCTTCATA	$(AG)_8$	135	51	KX377902
	R: CGAGTGTAAGGCGTTAGGAGATA				
LT87	F: AAGTACTCGAGAAGCAGGAGTCA	$(TTA)_4$	116	49	KX377901
	R: CCACCATAAAATCCTTTCCAAAT				
LT91	F: GTACTTAGCGTGGGACTTTGTTG	$(TGA)_7$	143	50	KX377900
	R: CCACATCATCATCAATTGCATAC				

TABLE 1. Continued.

Locus	Primer sequences (5′–3′)	Repeat motif	Fragment size (bp)	T_a (°C)	GenBank accession no.
LT93	F: AGCCAGTCGTCTCATTACAAAAA R: TTCTGCAGAGACTGGATCTGAAT	$(CT)_6$	138	55	KX377899
LT95	F: CGCAGTAGCAGATAGTGAATGTG R: TCCTCAAAATCAATGTCAGTGTG	$(ATC)_5$	119	54	KX3779898
LT97	F: TCGGGTTTATGTCTTACACTTGAG R: AGATCCTTAATTTTTATGAGCAATCA	$(AAT)_4$	148	53	KX3779897
LT98	F: ACATTGAAGACTAAGACATGGCG R: GAGATACAAACCCTAACCCTCGT	$(GA)_6$	156	52	KX3779896

Note: T_a = annealing temperature.

Before doing any further analysis, quality control and filtering of raw data were performed to detect whether the raw reads were of high enough quality, following Zhang et al. (2014). After that, clean reads were clustered using CD-HIT-EST (Li and Godzik, 2006) and assembled de novo using VelvetOpt (Zerbino and Birney, 2008). Sequencing produced 2,764,204,500 bp of clean reads after quality control from 2,800,948,250 bp of raw reads. We obtained 1,417,277 cluster tags, but only 222,628 cut cluster tags. We obtained 401,203 high-quality contigs, with an average size of 265 bp (N50 = 361) through de novo assembly.

Subsequently, we identified the SSR repeats from the assembled contigs using Trimmomatic version 0.32 (Bolger et al., 2014) and set the parameters for detection of di-, tri-, tetra-, penta-, and hexanucleotide motifs with flanking regions in SSR pipeline version 0.951 (Miller et al., 2013). A total of 4441 perfect SSR repeats from the assembled contigs were obtained in the study. Among them, the numbers of di-, tri-, tetra-, penta-, and hexanucleotide repeats were 2026, 2081, 220, 73, and 41, respectively.

***SSR primer design and genetic diversity analysis*—**SSR primers were designed using Primer3web (Untergasser et al., 2012) for the SSR sequences. Primers were designed according to the following criteria: amplified regions within a size range of 100–200 bp, primer annealing temperature range 55.0–62.0°C, and GC content range 45–60%. Different repeat motifs of SSR sequences were arbitrarily selected to design primers to obtain 100 pairs of qualified SSR primers. PCRs were performed with all 56 samples, with a 30-µL reaction mixture: 20–30 ng of template DNA, 5 µL 10× PCR buffer (15 mM $MgCl_2$), 1.5 µL of each primer (5 pM), 1.0 µL *Taq* DNA polymerase (TaKaRa Biotechnology Co., Dalian, China), 0.5 µL dNTP mix (10 mM), and supplemented with ddH_2O. The PCR program included the following steps: 94°C for 5 min, one cycle; 94°C for 35 s at the appropriate annealing temperatures (annealing temperatures for each specific primer pairs are given in Table 1) for 35 s; 72°C for 30 s, 35 cycles; 72°C for 10 min, one cycle. PCR products were visualized on 1.0% agarose gels with ethidium bromide. Of the 100 pairs of SSR primers tested, 38 amplified successfully (Table 1). These 38 primer pairs were used for PCR amplification in all 56 samples to detect polymorphism. PCR conditions are the same as those described above. PCR products were applied on agarose and then separated on 12% w/v nondenaturing polyacrylamide gels (PAGE) following Wang et al. (2014), with DL500 DNA Marker (TaKaRa Biotechnology Co.). We calculated the inbreeding coefficient (F_{IS}), total number of alleles per locus (A), observed heterozygosity (H_o), expected heterozygosity (H_e), null allele frequency (r), and deviations from Hardy–Weinberg equilibrium (HWE) using GENEPOP version 4.4 (Rousset, 2008).

After PAGE analysis, 10 pairs of SSR primers were found to be highly polymorphic among the three populations of *L. tibetica*; the other 28 showed no significant difference. A ranged from three to eight. H_o and H_e ranged from 0.200 to 1.000 and from 0.683 to 0.879 (Table 2), respectively, which indicates that genetic diversity in this species is relatively high. Additionally, r ranged from 0.000 to 0.307. Some loci (LT25 in population YD, LT4 in population QML, LT7 and LT9 in population MY) showed a significant departure from HWE, which could be caused by the presence of null alleles (Chapuis and Estoup, 2007).

There are just two species in the genus *Lancea* Hook. f. & Thomson, *L. tibetica* and *L. hirsuta* Bonati. We tested cross-amplification in *L. hirsuta* for all of the polymorphic primers developed for *L. tibetica*. *Lancea hirsuta* is distributed in northwestern Sichuan and northwestern Yunnan, China. We sampled five individuals from Xinduqiao (voucher no. Zhang2015569; geographic coordinates: 30°04′N, 101°29′E; altitude: 3496 m), Sichuan, China. All of the polymorphic primers were successfully amplified in *L. hirsuta* with the same PCR conditions used for *L. tibetica*, except for marker LT28.

CONCLUSIONS

In this study, we present the first report on *L. tibetica* SSR marker development based on RAD sequences. A total of 4441 SSR markers were identified at the genome-wide level. The 10 SSR loci that displayed polymorphisms among *L. tibetica* populations also have the potential to be useful for population genetic studies on the closely related *L. hirsuta*.

TABLE 2. Results of initial primer screening of 10 polymorphic loci in three *Lancea tibetica* populations.[a]

	Population YD (N = 17)					Population QML (N = 20)					Population MY (N = 19)				
Locus	A	H_e	H_o	r	F_{IS}	A	H_e	H_o	r	F_{IS}	A	H_e	H_o	r	F_{IS}
LT4	6	0.800	0.529	0.131	0.346	3	0.683*	0.200	0.307	0.713	5	0.758	0.474	0.141	0.382
LT7	7	0.838	0.706	0.054	0.162	3	0.719	0.600	0.160	0.169	5	0.807*	0.579	0.116	0.288
LT9	7	0.848	0.706	0.062	0.172	5	0.745	0.800	0.024	–0.076	8	0.865*	0.789	0.110	0.089
LT10	7	0.816	0.882	0.038	–0.084	7	0.814	0.950	0.000	–0.173	7	0.859	1.000	0.000	–0.169
LT12	5	0.779	0.765	0.000	0.019	5	0.768	0.900	0.000	–0.177	6	0.831	0.895	0.000	–0.079
LT15	5	0.779	0.882	0.000	–0.137	6	0.814	0.750	0.028	0.081	4	0.708	0.842	0.012	–0.195
LT16	8	0.850	0.765	0.199	0.103	8	0.879	0.850	0.030	0.034	6	0.812	0.632	0.095	0.223
LT18	6	0.820	0.706	0.058	0.143	7	0.821	0.850	0.000	–0.035	5	0.700	0.895	0.000	–0.289
LT25	7	0.872*	0.824	0.165	0.057	4	0.781	0.850	0.034	–0.091	6	0.797	0.737	0.046	0.077
LT28	6	0.804	0.824	0.000	–0.025	6	0.833	0.850	0.036	–0.020	6	0.835	0.895	0.000	–0.074

Note: A = total number of alleles per locus; F_{IS} = inbreeding coefficient; H_e = expected heterozygosity; H_o = observed heterozygosity; N = number of individuals sampled; r = null allele frequency.

[a] Population and voucher information are provided in Appendix 1.

* Significant departure from HWE at $P < 0.01$.

LITERATURE CITED

Barchi, L., S. Lanteri, E. Portis, A. Acquadro, G. Valè, L. Toppino, and G. L. Rotino. 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics* 12: 304.

Bolger, A., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* doi:10.1093/bioinformatics/btu170.

Chapuis, M. P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.

Chutimanitsakun, Y., R. W. Nipper, A. Cuesta-Marcos, L. Cistué, A. Corey, T. Filichkina, E. A. Johnson, and P. M. Hayes. 2011. Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. *BMC Genomics* 12: 4.

Doyle, J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Hong, D. Y., H. B. Yang, C. L. Jin, and H. H. Noel. 1998. Scrophulariaceae. *In* Z. Y. Wu and P. H. Raven [eds.], Flora of China, vol. 18. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Li, W., and A. Godzik. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)* 22: 1658–1659.

Litt, M., and J. A. Luty. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44: 397–401.

Miller, M. P., B. J. Knaus, T. D. Mullins, and S. M. Haig. 2013. SSR pipeline: A bioinformatic infrastructure for identifying microsatellites from paired-end Illumina high-throughput DNA sequencing data. *Journal of Heredity* 104: 881–885.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240–248.

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

Song, Z. H., Y. H. Wang, Z. Z. Qian, T. J. Smillie, and I. A. Khan. 2011. Quantitative determination of 10 phenylpropanoid and lignan compounds in *Lancea tibetica* by high-performance liquid chromatography with UV detection. *Planta Medica* 77: 1562–1566.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3: New capabilities and interfaces. *Nucleic Acids Research* 40: e115.

Wang, L., J. Xu, T. Xia, H. Zhang, D. Liu, and Y. Shen. 2014. Population structure and linkage disequilibrium in six-rowed barley landraces from the Qinghai-Tibetan Plateau. *Crop Science* 54: 2011–2022.

Zerbino, D., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.

Zhang, F. Q., Q. B. Gao, G. Khan, K. M. Luo, and S. L. Chen. 2014. Comparative transcriptome analysis of aboveground and underground tissues of *Rhodiola algida*, an important ethno-medicinal herb endemic to the Qinghai-Tibetan Plateau. *Gene* 553: 90–97.

Appendix 1. Locality information for *Lancea tibetica* populations in the study.

Population code	Location	*N*	Voucher no.[a]	Geographic coordinates	Altitude (m)
YD	Yadong, Xizang, China	17	Chen2014498	27°47′26″N, 99°08′52″E	4350
QML	Qumalai, Qinghai, China	20	Chen2014684	33°58′03″N, 96°34′39″E	4570
MY	Menyuan, Qinghai, China	19	Zhang2014341	37°51′00″N, 101°04′51″E	3636

Note: *N* = number of individuals sampled.

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