西北植物学报,2018,38(5):0850－0857
Acta Bot. Boreal. -Occident. Sin.

文章编号:1000-4025(2018)05-0850-08 **doi**:10.7606/j.issn.1000-4025.2018.05.0850

濒危植物羊踯躅全基因组 SSR 标记开发与鉴定研究

刘梦露,戴亮芳,程马龙,张　金,李海燕,谢建坤,罗向东*

(江西师范大学 生命科学学院,南昌 330022)

摘　要:该研究利用基于全基因组限制性酶切位点简化基因组测序技术(RAD-seq 技术),开发濒危植物羊踯躅(*Rhododendron molle* G. Don)全基因组 SSR 标记,并对 3 个群体共 63 份羊踯躅材料进行验证鉴定,为进一步研究羊踯躅的遗传多样性和群体遗传结构以及保护利用提供技术支持。结果显示:(1)羊踯躅基因组测序获得原始数据 7.653G bp,过滤后为 7.513G bp;经组装发现,羊踯躅 171.534 M bp 的基因组分布在 498 252 contigs 中。(2)通过 SSR 检测,在 11 961 SSR 位点中获得了 11 687 对 SSR 分子标记,并且二核苷酸为基序的重复类型最丰富,达 51.76%。(3)随机选取 128 对 SSR 标记在 6 个羊踯躅株系中进行 PCR 扩增,获得 20 对高多态性的 SSR 标记。(4)用所选的 20 对多态性 SSR 标记对 3 个群体共 63 份羊踯躅材料进行验证分析发现,这些多态性 SSR 标记位点的等位基因数为 4～16 个,期望杂合度(H_e)为 0.489～0.908。研究表明,羊踯躅的 SSR 丰度适中,且二核苷酸为羊踯躅中最丰富的重复序列,该实验进一步证明 RAD-seq 技术是一种经济有效的基因测序方法,实验中开发的 SSR 引物将有助于进一步研究羊踯躅和其他近缘种的群体结构和多样性。

关键词:濒危物种;羊踯躅;微卫星标记;多态性

中图分类号:Q789　　**文献标志码**:A

Genome-wide Identification of SSR Markers in Endangered Species *Rhododendron molle* G. Don

LIU Menglu, DAI Liangfang, CHENG Malong, ZHANG Jin,
LI Haiyan, XIE Jiankun, LUO Xiangdong*

(College of Life Science, Jiangxi Normal University, Nanchang 330022, China)

Abstract: The study set of simple sequence repeat (SSR) markers were developed using Restriction site Associated DNA (RAD) approach and Illumina paired-end sequencing to the whole genome of endangered plants *Rhododendron molle* G. Don. And it would provide technical support for further investigation of population structure and diversity of *R. molle* and other congener species from the evaluation of 63 individuals from three different populations. The results showed that: (1) the sequencing output generated a FASTQ file size of raw reads 7.653G bp, and the clean reads was to 7.513G bp. About 171.534 M bp of *R. molle* genome distributed over 498 252 contigs were obtained upon assembly of sequencing data. (2) After filtering and SSR detection, a final set of 11 687 simple sequence repeats with primers were obtained from 11 961 microsatellite loci. The di-nucleotides motif unit were the most abundant (51.76%) repeat sequences. (3) 128 microsatellite markers were selected randomly to detect the polymorphism in 6 different individuals of *R. molle*, from which 20 polymorphic primers were identified. (4) Then the characteristics

收稿日期:2018-01-19;**修改稿收到日期**:2018-05-06

基金项目:国家自然科学基金(31360327,1260255,31660384);江西省自然科学基金(20151BAB204008,20171BAB204021)

作者简介:刘梦露(1993－),女,在读硕士研究生,主要从事植物分子生物学研究。E-mail:497461823@qq.com

* 通信作者:罗向东,博士,教授,主要从事植物分子生物学研究。E-mail:xdluolf@163.com

and availability of polymorphic primers were further evaluated in 63 individuals from three different populations. The number of alleles per locus ranged from 4 to 16, expected heterozygosity (H_e) values ranged from 0.489 to 0.908, respectively. Our result showed that the abundance of SSR existed in *R. molle* was moderate and the di-nucleotides were the most abundan repeat sequences. It further confirms that the RAD-seq method is an efficient and cost-effective means for SSR discovery, and these polymorphic microsatellite markers developed in this study will be useful for further investigation of population structure and diversity of *R. molle* and other congener species.

Key words: endangered species; *Rhododendron molle* G. Don; microsatellites; polymorphic

Rhododendron molle G. Don (Ericaceae) is one of *Rhododendron* species native to China and Japan. With flourishing branches and gorgeous flowers, it is one of elite parents for new varieties improvement in *Rhododendron*[1]. Moreover, the plant of *R. molle* is rich in ericol-in, rhodofoxin and rhodojaponin-Ⅲ. Rhodojaponin-Ⅲ is one of grayanane diterpenoids, possesses significant blood pressure lowering and heart rate slowing effects[2]. Both of ancient and modern medicine studies suggested that different organs and tissues of *R. molle* (including flower, fruit, root and leaf ect.) could be used as analgesics and insecticides[2-4]. In recent years, the new research on the pharmacological effects of *R. molle* has extended to anti-schistosome[5]. Therefore, *R. molle* is a high valuable medicinal and attractive ornamental plants.

R. molle was widely distributed across central and south China until the 1980s. As a result of overexploitation, habitat loss, weak fertility and lower ability of natural regeneration, *R. molle* has dramatically decreased in recent years. It has suffered severe habitat fragmentation and population loss. At present, *R. molle* exists in a restricted number of populations with small population sizes. The remnant populations are exposed to higher risk of critically endangered or extinction[6]. Understanding the processes that determine population genetic structure and diversity is very important for the conservation and management of threatened populations, which is necessary to employ more powerful DNA markers[7]. Simple sequence repeats (SSR) or microsatellites marker is a co-dominant marker, which is regarded as one of the most effective markers to investigate the genetic diversity and population genetic structure[8].

Here, we report a set of novel polymorphic microsatellites markers for endangered *R. molle* by using Restriction site Associated DNA approach and Illumina paired-end sequencing (RAD-seq), which will provide a useful tool for studying population genetics and evolutionary history as well as developing a conservation strategy for this species.

1 Materials and methods

1.1 Plant material and DNA extraction

The young leaves of *R. molle* were collected from three locations of China, Yongxiu County (29°07′N, 115°43′E, Jiangxi Province), Panan County (28°52′N, 120°26′E, Zhejiang Province) and Jinzhai County (31°23′N, 115°46′E, Anhui Province). The young leaf tissues were sampled and flash frozen in liquid nitrogen and stored at −80 ℃ until DNA extraction. DNA was extracted from the young leaf tissues following modified CTAB method[9]. The quality and concentration of the DNA were examined by electrophoresis on 1% agarose gels with D2000 DNA marker.

1.2 Methods

1.2.1 RAD library construction and Illumina sequencing The genomic restriction site associated DNA sequences (RAD-seq) libraries were created by Novogene company (Novogene, China) as described by Baird *et al.*[10]. Genomic DNA of *R. molle* was digested with the restriction endonuclease (*Eco*R I), sheared randomly. Illumina P_1 and P_2 adaptors were ligated after end repair and addition of single 'A' base. After ligation, 1.0×SPRI (<100 bp) cleanup was performed and the library was amplified for enrichment of adapter ligated fragments. Then a single library with an insert size range of 300−700 bp was prepared and sequenced

from both ends with 100 bp read lengths in one lane of an Illumina sequencer on the Illumina HiSeq™2000/Miseq™. The raw data from the images acquired after sequencing was transformed by base calling into raw reads and stored in FASTQ format. Raw reads were subjected to quality check by using *SeQC* v2. 1 (Genotypic Proprietary Tool)[11]. Clean reads were obtained after sequencing error rate distribution checks, GC content distribution checks and sequencing data filtering.

1.2.2 Assembly of RAD-seq and SSR markers discovery A total of 7.513G bp clean reads was partially assembled using VelvetOptimiser to obtain a final assembly sequence[12]. The filtered assembled contigs were used as a source of sequence information for SSRs screening by SR search software with 2 and 6 as the minimum and maximum length of the repeating unit, respectively. The minimum distance between two SSR sequence is 12 bp and SSR upstream and downstream sequence length of 100 bp. SSR primers were designed using Primer 3.0[13]. And the major parameters for designing the primers were: (1) primer sequence length from 20 to 28 nucleotides, with 24 as the optimum, (2) annealing temperature from 60 ℃ to 65 ℃, an optimum annealing temperature of 63 ℃, and (3) a pair of primers annealing temperature difference of the maximum of 1 ℃.

1.2.3 Identification of polymorphism SSR and data analysis 128 pairs of primers selected randomly from 11 687 pairs of primers were synthesized by Sangon Company (Shanghai). Polymerase chain reaction (PCR) amplifications in six different geographical *R. molle* individuals were performed to identify polymorphism microsatellite loci. Each PCR reaction consisted of 4 μL of ultrapure water, 7.5 μL of 2× Taq PCR Green Mix, 2 μL each of forward and reverse primers (0.5 μmol · L^{-1} 5 pmol), and 1.5 μL of template genomic DNA in a final reaction volume of 15 μL. Annealing temperature gradient from 55 ℃ to 60 ℃ was set for finding optimum annealing temperature. The thermal cycles for the SSR PCR runs were programmed for an initial denaturation at 94 ℃ for 3 min; followed by 34 cycles of 94 ℃ for 45 s, 56 ℃ for 45 s, 72 ℃ for 1 min and finally, 72 ℃ for a final extension of 7 min. The PCR product was then separated on an 8% non-denaturing polyacrylamide gel and visualized by silver staining. 20 bp DNA ladder was used as standard for scoring. The characteristics and availability of polymorphic primers were further evaluated in 63 individuals from three different populations. The evaluation parameters include the number of alleles per locus (A), observed heterozygosity (H_o), expected heterozygosity (H_e) as well as deviations from Hardy-Weinberg equilibrium (HWE), which were calculated by the GenAlEx 6.1[14]. Null allele frequency was checked by Micro-Checker 2.2[15].

2 Results and analysis

2.1 RAD-seq for *R. molle*

A FASTQ format file size of 7.653G bp raw reads was generated after Illumina sequencing (Table 1). After trimming adapter sequences, containing 'N' and low quality bases, the initial clean data (7.513G bp) were obtained. The error rate of clean reads was about 0.04%. The Q_{20} and Q_{30} were to 93.73% and 88.33%, respectively (Table 1). The percentage of GC content observed in *R. molle* was about 39.69% and RAD-Tag digestion ratio is to 98.48% (Table 1). Clean reads were assembled *de novo* to serve as framework reference genome. About 171.534 Mb of *R. molle* genome distributed over 498 252 contigs were obtained upon assembly of sequencing data. The contig lengths ranged from 100 to 952 bp with an N_{50} length of 470 bp (Table 1, Fig. 1). These results suggested that the RAD-seq sequencing was successful, which was attributed to the sufficient samples data, normal GC distribution and high sequencing quality in this study.

2.2 SSR discovery and evaluation for *R. molle*

After several filtering criteria, 349 912 of 498 252 contigs were selected to detect SSR loci, since the contigs of less than 212 bp were filtered. Approximately 3.27% of all assembled contigs (11 457 out of 349 912) contained at least one detectable SSR

locus and the total number of microsatellite loci was 11 961. Finally, a set of 11 687 simple sequence repeats with primers were obtained from 11 961 microsatellite loci by Primer 3.0 software. We found that the proportion of five different SSR units was significant different. Different repeat units occurred at numbers of di-, tri-, tetra-, penta- and hexa-repeats were 6 049 (51.76%), 4408 (37.72%), 695 (5.94%), 369 (3.16%), 166 (1.42%), respectively (Table 2).

Table 1　Details of the raw reads' quality for RAD-seq in *R. molle*

Feature	Value
Raw Base/bp	7.653G
Clean Base/bp	7.513G
Effective Rate/%	98.16
Error Rate/%	0.04
Q_{20}/%	93.73
Q_{30}/%	88.33
N_{50} length/bp	470
GC Content/%	39.69
Digestion ratio/%	98.48
Average depth/X	26.16
Coverage at least 1X/%	93.70
Coverage at least 4X/%	78.71

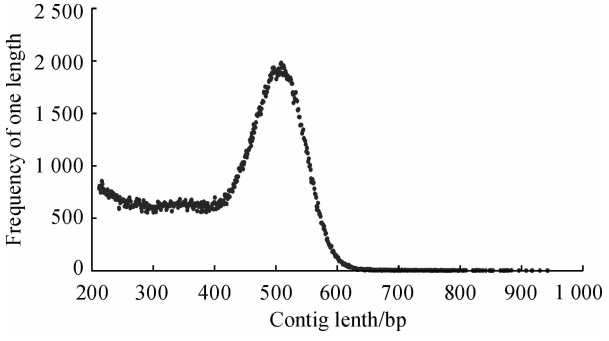


Fig. 1　Distribution of contig length (>212 bp)

A total of 481 types of repeat motifs were identified among the 11 687 (SSRs length≥12 bp) (Table 2). Of the different perfect SSR motif units, the most type number was penta-nucleotide (32.43%), followed by tetra-nucleotide (26.82%), hexa-nucleotide (25.79%), tri-nucleotide (12.47%) and di-nucleotide (2.49%). Further analysis indicated that the number of different repeat motifs in different repeat units was distributed unevenly. The $(GA)_n$ repeat motif was the most abundant type in di-nucleotide repeat motifs. The four other main unit type were the $(AAG)_n$ in tri-nucleotide, $(AAAT)_n$ in tetra-nucleotide, and $(AAAAG)_n$ in penta-nucleotide, $(TCTGAG)_n$/$(AAAAAG)_n$ in hexa-nucleotide repeat motifs, and occurred at frequencies of 19.42%, 5.20%, 12.09%, 4.07% and 3.01%, respectively.

2.3　Characteristics of polymorphic microsatellites markers

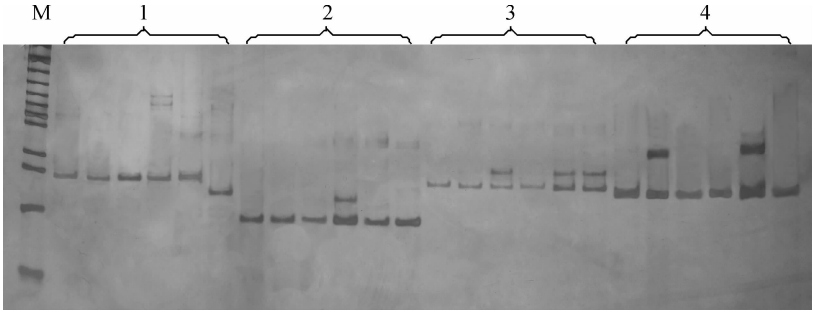
128 microsatellite markers were selected randomly to detect the polymorphism among six *R. molle* individuals from three different provinces, of which 20 pairs of primer present high polymorphism. To test and characterize the genetic variability of *R. molle* at the 20 loci, 63 individuals were collected from three locations of China (Yongxiu County, Panan County and Jinzhai County). All the SSR bands were clear and distinct (Fig. 2).

The characteristics of the polymorphic SSR markers of *R. molle* were shown in Table 3. The number of *A* varied from 4 to 16 with an average of 9.2 alleles per locus, H_o and H_e values ranged from 0.050 to 1.000 and 0.489 to 0.908 (Table 3). Seven loci (HD26, HD34, HD35, HD55, HD62, HD88, and HD100) showed a significant

Table 2　Number and frequency of five perfect SSR repeat motifs in *R. molle*

SSR motif unit	No. of SSR	The most abundant repeat motif		No. of the repeat motif type
		The most abundant	No. and frequency*	
Di-	6049	$(GA)_n$	1 175 (19.42%)	12
Tri-	4408	$(AAG)_n$	229 (5.20%)	60
Tetra-	695	$(AAAT)_n$	84 (12.09%)	129
Penta-	369	$(AAAAG)_n$	15 (4.07%)	156
Hexa-	166	$(TCTGAG)_n$/$(AAAAAG)_n$	5 (3.01%)	124
Total	11 687	—	—	481

* Note: The data before brackets was the number of the most abundant repeat motif, and in the brackets was the frequency of the most abundant repeat motif



M. 20 bp DNA ladder; 1—4. Amplification products of six different geographical *R. molle* individuals using four different primers (HD24, HD25, HD35 and HD37, respectively)

Fig. 2 Amplification products in some individuals of *R. molle* by polymorphic primers

departure from HWE ($P < 0.05$) probably due to heterozygote deficiency in all populations. The rest of thirteen loci were not deviated from HWE in at least one population. All loci were successfully amplified for three populations of *R. molle*, which suggested that these genomic-SSR markers will be valuable for genetic research in *R. molle*, such as population genetic structure analysis, diversity analysis and conservation etc.

3 Discussion

Due to the advantage of high abundance, random distribution within the genome, high polymorphism information and co-dominant inheritance, SSR markers have been widely used for construction of genetic maps, diversity analysis and genetic structure[8]. There are a number of traditional ways to develop SSR, including microsatellite enrichment, selectively amplified microsatellite polymorphic loci (SAMPL), and EST SSRs or genomic microsatellites according to public DNA databases[16]. All of these ways, however, should obtain the detailed sequence information about both sides of the repeat sequence and then design primers, which are costly and time-consuming. Recently, RAD-seq was developed and regarded as a promising and powerful tool to acquire the genomic and transcriptomic data for non-model organisms and non-sequenced genomes, which has been used for genome assembly and SSR marker development in a variety of species[17]. In the present study, 11 687 microsatellites were developed using RAD-seq. Then 20 polymorphic microsatellite loci from 128 microsatellites were identified. Compared with previous studies, the RAD-seq for SSR development for *R. molle* without an available reference genome shows more time-saving, high efficient and high throughput[18].

In this study, the RAD-seq data was reliably called for 93.70% (sequence coverage at least 1X) and average depth of coverage of 26.16X. It allowed the discovery of a set of 11 961 perfect SSR loci of ≥12 bp length in *R. molle*. About 3.3% contigs in *R. molle* possessed at least one SSR, which was lower than previously reported SSR prevalence in the ESTs of pigeonpea (7.6%), wheat (7.4%) and flax (3.5%), higher than barley (2.8%)[19-22]. The SSR frequency also depended on the parameters used in detecting SSR markers, such as the repeat length and number of repeat unit thresholds. The abundance of SSR (SSR/kbp) in *R. molle* was 14.3 compared to 20.0 in cotton[17], 14.0 in *Arabidopsis*, 8.4 in pigeonpea and 5.4 in wheat[19-20]. Our result showed the abundance of SSR existed in *R. molle* was moderate.

SSRs are microsatellites with 2—6 nucleotides tandem repeats which are distributed randomly throughout the genome of all the eukaryotes[11]. In the present study, we found that the proportion of different SSR units was significant different. The di-nucleotides were the most abundant (51.76%) repeat sequences, which was agree with other species, for example, di-nucleotides account for 48 to 67% of repeat sequences in plant libraries[23], e.g. pigeonpea (60.41%)[19], *P. alata* (62.5%)[24]. In ESTs of grape and soybean, however, the tri-nucleotides SSR locus was the most abundant class[25]. Yu *et al.*[26] reported that the di-nucleotide repeats were

Table 3 Characteristics of 20 SSR primer pairs detected polymorphism in three populations of *R. molle* G. Don (N= 21 for each population)

Locus name 1)	Primer Sequence 2) (5′ to 3′)	Repeat motif	Locus size range/bp	Jinzhai			Panan			Yongxiu		
				A	*Ho*	*He*	*A*	*Ho*	*He*	*A*	*Ho*	*He*
HD24	F: GGGGGACCATTTATCAGTAGAATTG R: GCAATGATGAAATCAGTGGCATC	$(AT)_8$	136-182	9	0.750	0.820	16	0.571	0.895	10	0.381	0.837
HD25	F: GAAGATGGCTACTCTCTGGTCGTC R: ACCCAAATCACTCTGAAACCCTCT	$(AC)_{11}$	116-142	8	0.905	0.837	10	0.810	0.863	15	0.952	0.893
HD26 *	F: CCAAGTGCCATTCTTCACATTGTA R: CTCGATGAGAATCAACTTCGGAAT	$(TGA)_6$	126-162	6	0.286	0.722	3	0.190	0.643	6	0.050	0.759
HD34 *	F: TATTTGTGTGTGGTTGATCTTGGG R: CTTGGTGCACTGTAATCTCTCCCT	$(GA)_{10}$	90-142	14	0.368	0.873	10	0.350	0.864	13	0.714	0.885
HD35 *	F: AAAAGAAAGTGCTGCATTAGTGGC R: TAAGCGTGACATTTCCGTGTATTG	$(GCA)_5$	130-158	7	0.286	0.769	9	0.300	0.831	8	0.619	0.827
HD37	F: TCTAGCTCTCCTATCGGCAAAAGA R: CAAATGGTGATCCAACAACTCAAA	$(CT)_{10}$	137-172	4	0.429	0.643	8	0.714	0.698	6	0.476	0.745
HD39	F: GAGCCCTTTATATTCCTCAGCCAT R: TGGCTTGAAACTTCTGAATCCACT	$(AG)_{12}$	96-136	5	0.333	0.719	8	0.667	0.800	10	0.905	0.853
HD43	F: GCTCTCAGTTGTTTGGCTCATTTT R:CAGCCATGAGTTAGGACATATCACG	$(AG)_{10}$	110-145	11	0.714	0.881	8	0.286	0.810	14	0.524	0.908
HD55 *	F: TTTTTCGTTTCCCGCCTATTACTT R: GTGCTCATTTTTCATAGGCTGGTT	$(CTG)_5$	126-152	11	0.190	0.859	10	0.150	0.850	7	0.429	0.786
HD62 *	F:GGTGCATTCGTGTATGCTTGATTA R:ACCAGCACCCCTCCTATACCTAAG	$(AC)_{14}$	100-155	11	0.524	0.844	15	0.381	0.889	9	0.333	0.810
HD63	F:TCTTCCCCATATATTGCTCCAAGA R:TGGGGTTAGGAAAGACAAGTACCA	$(AC)_{14}$	113-140	15	0.619	0.868	8	0.286	0.787	9	0.190	0.846
HD64	F:GAAACAAACCATAACCGTAACCGA R:TTGATTTCTTCACACTTTTCTATTCGC	$(AC)_{14}$	124-165	9	1.000	0.825	13	0.600	0.894	9	0.333	0.823
HD74	F:CGTAAGATGACCCAAACATCGAAT R:GTTAGGTGGATTTGCGTTGTCTCT	$(GA)_{15}$	116-180	8	0.381	0.735	8	0.333	0.770	15	0.952	0.853
HD80	F:AAATGGCTTCATCTCGAGTCTCTG R:AACTCCATGTTGAAATGAGCACAA	$(CAA)_6$	130-158	8	0.524	0.735	5	0.619	0.663	4	0.429	0.489
HD83	F:ATTGACCCTAGATCCACTCCCTG R:AAGAAAACAACGCAAACACACAAA	$(AAG)_6$	110-142	5	0.286	0.737	6	0.571	0.717	7	0.095	0.734
HD88 *	F:TGGGTTTGTGGTTGATTGTGTTAG R:GACCCATTTCCCAAAAACTCTTCT	$(AAG)_7$	125-152	9	0.316	0.864	12	0.143	0.883	9	0.238	0.819
HD97	F:TCCTTACCGGAACTCTTGAATCTG R:ATCGGAGAATCGTCGAGATTGATA	$(AG)_{16}$	98-132	7	0.550	0.784	5	0.381	0.766	5	0.667	0.736
HD100 *	F:GTGGACTGGACTGGACTGGTTACT R:GGAAAGGTGGATTGTGGATTATGA	$(CCT)_9$	134-164	4	0.143	0.481	8	0.500	0.815	7	0.381	0.735
HD103	F:GCATCTTCTGAAGGCAAAAGGTAA R:CTAATTTCCATACCCTGTCTCCCC	$(AG)_{16}$	138-166	6	0.524	0.677	7	0.526	0.699	6	0.850	0.765
HD109	F:TTCTACCCTGGTTTTGAGATTGGA R:CAACACGATGGTAGCCTTTTTAGG	$(GA)_{15}$	114-140	3	0.800	0.535	5	0.476	0.607	4	0.800	0.696

Note: 1) * means the SSR significant deviation from Hardy-Weinberg equilibrium in three populations ($P<0.05$); 2) F. forward primer; R. reverse primer; *A*. number of alleles; H_o. observed heterozygosity; H_e. expected heterozygoisty

mainly in the noncoding regions, whereas a portion of tri-nucleotide repeats were in the coding region. In this study, RAD-seq is characterized high throughput and whole-genome scanning, which resulted in RAD-seq tags include large amounts of information representing noncoding regions and the di-nucleotides were the most abundant (51.76%) repeat sequences in *R. molle*.

Out of all the repeat motifs from 11 687 primer pairs, $(GA)_n$(10.05%) was the most abundant di-nucleotide repeat in the samples, which was in agreement with the previous results obtained in jute[8], barley, maize, rice, sorghum, and wheat[27]. But Tan *et al*.[28] reported that TG/CA (9.2%) was the most common repeat motif for bermudagrasses. In *R. molle*, CTT/AAG, TCT/AGA and AAT/ATT accounted for 4.31%, 4.15% and 4.04% of tri-nucleotide repeats. Similarly, in jute the most abundant trimeric motifs were AAG/CTT, AAC/GTT, TCT/AGA[8], and in *P. alata*, AAG/CTT, AGA/TCT and GAA/TTC were the most abundant motifs[24]. Analysis of the most abundant di-nucleotide and tri-nucleotide repeat, we finnd that most of these repeats are rich in A/T which is prone to mutation during the SSR evolution[29]. SSR plays an important role in species formation due to the effect to regulate of gene expression[29]. An evolutionary view, natural selection of species and organism adapt to the environment result in the difference of repetitive sequences. Higher the evolution degree of biological was, more the repetitive sequences accounted for the proportion of the genome[30]. Therefore, some kind of SSR, such as GA, TG or CTT/AAG, AGA/TCT, their abundance in a species perhaps due to natural selection and the evolution of the species. In addition, it is likely that these types of repeats have already been used to design probes to capture DNA fragments containing microsatellites.

To our knowledge, this is the first report of RAD-seq to develop more than ten thousand of SSR markers of *R. molle* which will be highly valuable in the diversity and genetic structure analysis. The 20 primer pairs will be used in future investigation of spatial genetic structure, comparative levels of genetic variation within and between populations. We expect that these markers could be utilized to facilitate efficient conservation and management strategies for this medicinal and ornamental plant. On the other hand, these developed SSR primers can be used as universal primers of the related species[31].

In this article, a dataset of 498 252 contigs were derived from 7.513G bp clean reads of *R. molle* using next-generation DNA sequencing technique. A comprehensive set of 11 687 genomic-SSR markers was developed. 128 microsatellite markers were selected randomly to detect the polymorphism. 20 out of 128 pairs of polymorphic primers were identified in 63 individuals from three different populations. The number of alleles per locus ranged from 4 to 16, expected heterozygosity (H_e) values ranged from 0.489 to 0.908, respectively. Our results further confirm that the RAD-seq method is an efficient and cost-effective means for SSR discovery. And these polymorphic microsatellite makers developed in this study will be useful for further investigation of population structure and diversity of *R. molle* and other congener species.

References:

[1] ZHONG G H, LIU J X, WENG Q F, *et al*. Laboratory and field evaluations of rhodojaponin-III against the imported cabbage worm *Pieris rapae* L. (Lepidoptera: Pieridae) [J]. *Pest Management Science*, 2006,**62**: 976-981.

[2] CHENG D M, HU M Y. Research advances in the research on *Rhododendron molle* G. Don [J]. *Natural Product Research and Development*, 1999,**11**: 109-113.

[3] BAO G H, WANG L Q, CHENG K F, *et al*. Diterpenoid and phenolic glycosides from the roots of *Rhododendron molle* [J]. *Planta Medica*, 2003,**69**: 434-439.

[4] ZOU H Y, LUO J, XU D R, *et al*. Tandem solid-phase extraction followed by HPLC-ESI/QTOF/MS/MS for rapid screening and structural identification of trace diterpenoids in flowers of *Rhododendron molle* [J]. *Phytochemical Analysis*, 2014,**25**: 255-265.

[5] 谷旭田, 赵红梅, 陈洪亮. 羊踯躅等 6 种中药对钉螺与血吸虫毛蚴杀灭效果观察[J]. 中兽医医药杂志. 2008,**3**: 20-22.

GU X T, ZHAO H M, CHEN H L. Study on the effects of six kinds of Chinese herbs on killing *Oncomelania* and blood fluke miracidium [J]. *Journal of Traditional Chinese Veterinary Medicine*, 2008,**3**: 20-22.

[6] 程马龙，戴亮芳，罗向东，等. 濒危羊踯躅自然居群的遗传多样性和遗传结构分析[J]. 西北植物学报，2016，**36**(3)：674-680.
CHENG M L, DAI L F, LUO X D, *et al*. Genetic diversity and genetic structure analysis on natural populations of endangered *Rhododendron molle* G. Don [J]. *Acta Botanica Boreail-Occidentalia Sinica*, 2016,**36**(3): 674-680

[7] TAN X X, LI Y, GE X J. Development and characterization of eight polymorphic microsatellites for *Rhododendron simsii* Planch (Ericaceae) [J]. *Conservation Genetics*, 2009, **10**: 1 553-1 555.

[8] DAS M, BANERJEE S, DHARIWAL R, *et al*. Development of SSR markers and construction of a linkage map in jute [J]. *Journal of Genetics*, 2012,**91**: 21-31.

[9] DOYLE J J. DNA protocols for plants-CTAB total DNA isolation. In: G. M. HEWITT and A. JOHNSTON, ed. *Molecular Techniques in Taxonomy* [M]. Berlin: Springer-Verlag. 1991, 283-293.

[10] BAIRD N A, ETTER P D, ATWOOD T S, *et al*. Rapid SNP discovery and genetic mapping using sequenced RAD markers [J]. *PLoS One*, 2008,**3**: 3 376-3 376.

[11] MUDALKAR S, GOLLA R, GHATTY S, *et al*. De novo transcriptome analysis of an imminent biofuel crop, *Camelina sativa* L. using Illumina GAIIX sequencing platformand identification of SSR markers [J]. *Plant Molecular Biology*, 2014,**84**: 159-171.

[12] LI H, DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2009,**25**: 1 754-1 760.

[13] ZERBINO D R, BIRNEY E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs [J]. *Genome research*, 2008,**18**: 821-829.

[14] PEAKALL R, SMOUSE P E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research [J]. *Molecular Ecology Notes*, 2006,**6**: 288-295.

[15] OOSTERHOUT O C, HUTCHINSON W F, WILLS D P M, *et al*. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data [J]. *Molecular Ecology Notes*, 2004,**4**: 535-538.

[16] VARSHNEY R K, GRANER A, SORRELLS M E. Genic microsatellite markers in plants: features and applications [J]. *Trend in Biotechnology*, 2005,**23**: 48-55.

[17] WU K, LIU H Y, YANG M M, *et al*. High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (*Sesamum indicum* L.) based on RAD-Seq techonology [J]. *BMC Plant Biology*, 2014,**14**: 1-14.

[18] ABDELKRIM J, ROBERTSON B, STANTON J A, *et al*. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing [J]. *Bio Techniques*, 2009,**46**: 185-192.

[19] DUTTA S, KUMAWAT G, SINGH B P, *et al*. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh] [J]. *BMC Plant Biology*, 2011,**11**: 1-13.

[20] PENG J H, LAPITAN N L. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers [J]. *Functional & Integrative Genomics*, 2005,**5**: 80-96.

[21] CLOUTIER S, NIU Z, DATLA R, *et al*. Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.) [J]. *Theoretical & Applied Genetics*, 2009,**119**: 53-63.

[22] VARSHNEY R K, GROSSE I, HAHNEL U, *et al*. Genetic mapping and BAC assignment of EST-derived SSR markers shows nonuniform distribution of genes in the barley genome [J]. *Theoretical and Applied Genetics*, 2006,**113**: 239-250.

[23] LI Y C, KOROL A B, FAHIMA T, *et al*. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review [J]. *Molecular Ecology*, 2002,**11**: 2 453-2 465.

[24] PENHA H A, PEREIRA G D S, ZUCCHI M I, *et al*. Development of microsatellite markers in sweet passion fruit, and identification of length and conformation polymorphisms within repeat sequences [J]. *Plant Breeding*, 2013,**132**: 731-735.

[25] HISANO H, SATO S, ISOBE S, *et al*. Characterization of the soybean genome using EST-derived microsatellite markers [J]. *DNA Research*, 2007,**14**: 271-281.

[26] YU J K, DAKE T M, SINGH S, *et al*. Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat [J]. *Genome*, 2004,**47**: 805-818.

[27] KANTETY R V, ROTA M L, MATHEWS D E, *et al*. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. *Plant Molecular Biology*, 2002,**48**: 501-510.

[28] TAN C C, WU Y Q, CHARLES M, *et al*. Development of simple sequence repeat markers for bermudagrass from its expressed sequence tag sequences and preexisting sorghum SSR markers [J]. *Molecular Breeding*, 2012,**29**: 23-30.

[29] KATTI M V, RANJEKAR P K, GUPTA V S. Differential distribution of simple sequence repeats in eukaryotic genome sequences [J]. *Molecular Biology and Evolution*, 2001,**18**: 1 161-1 167.

[30] 郭瑞星，刘小红，荣廷昭，等. 植物 SSR 标记的发展及其在遗传育种中的应用[J]. 玉米科学，2010,**45**：457-459.
GUO R X, LIU X H, RONG T Z, *et al*. Development of SSR maker and its application in plant genetics and breeding [J]. *Journal of Maize Sciences*, 2005,**13**: 8-11.

[31] QIN Z C, YANG J B. Development and characterization of 15 microsatellite loci for *Rhododendron delavayi* Franch. (Ericaceae) [J]. *Hortscience*, 2010,**45**: 457-459.

(编辑：宋亚珍)