RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2020.035

ScienceAsia 46 (2020): **280–287**

Novel SSR and SNP markers in *Viola yedoensis* Makino resistant to cadmium stress

Hua Peng[a,†], **Jian Gao**[b,c,†], **Chun Zhang**[d,*], **Lilan Ou**[d], **Mao Luo**[a,d,e,*]

[a] Department of Tourism and Cultural Industry, Sichuan Tourism College, Chengdu 610100 China
[b] Department of Life Sciences and Technology, Yangtze Normal University, Fuling 408100 China
[c] Center for Green Development and Collaborative Innovation in Wuling Mountain Region, Yangtze Normal University, Fuling 408100 China
[d] Department of Pharmacology, School of Pharmacy, Southwest Medical University, Luzhou 646000 China
[e] Collaborative Innovation Center for Prevention and Treatment of Cardiovascular Disease of Sichuan Province, Drug Discovery Reseach Center, Southwest Medical University, Luzhou 646000 China

[*]Corresponding authors, e-mail: zc83good@126.com, luomao20050908@163.com, Peng H. and Gao J. contributed equally to this article

Received 10 Jul 2019
Accepted 1 Apr 2020

ABSTRACT: *Viola yedoensis* Makino is a well-known native Chinese medicinal plant, which is reported to adapt to heavy metal pollution regions. In our preliminary studies, transcriptome sequencing was employed to identify global expression gene using a combined bioinformatics, and plenty of differential expression genes (DEGs) were also identified in *V. yedoensis* Makino responsive to cadmium (Cd) pollution. However, the relative molecular marker development was not expounded. In this study, we aimed to identify novel SSR (simple sequence repeat) and SNP (single-nucleotide polymorphism) markers in *V. yedoensis* Makino, as well as further study of the functional annotation, classification, and phylogenetic analysis of cyclotide family genes. Approximate 91.99 million high-quality clean reads were generated. 11176 simple sequence repeats (SSRs) were predicted in 9644 unigenes, of which, 1283 unigenes contained multiple SSRs. The SSRs consisted of 1732 dinucleotide motifs, 2733 trinucleotide motifs, 135 tetranucleotide motifs, 17 pentanucleotide motifs, and 13 hexanucleotide motifs. 163 motifs were detected among all of the SSRs identified. Moreover, 8250 SNP markers were detected in 1934 unigenes and 11328 SNP markers were detected in 2510 unigenes in VIYCd and VIYCK library respectively. Furthermore, 41 cyclotide family Unigenes in Violaceae were identified and categorized into six groups with the other species. This study firstly generated a valuable resource for SNP and SSR marker development studies of Violaceae transcriptome and thereby providing the basis for further studies on genomics and functional genomics in Violaceae.

KEYWORDS: *Viola yedoensis* Makino, SSRs, SNPs, cyclotide family

INTRODUCTION

Viola yedoensis Makino, an important native Chinese traditional medicine, is a small perennial plant with violet flowers distributed in China [1–4]. The dried whole plant (known as "Herba Violae") is an important constituent of the traditional medicine "Zi Hua Di Ding", which have been used traditionally to treat acute pyogenic infection, respiratory and inflammatory diseases, including bronchitis, hepatitis, swelling, boils and enteritis [1, 5]. Recent studies have revealed that cyclotides are a group of plant-derived peptides with unique structural properties, and almost all of cyclotides are isolated from Violaceae family [2]. Cyclotides have been reported to possess potent effect of anti-human immunodeficiency virus and weak antibacterial activity and unmanageable toxicity [3, 5].

In recent years, the high-throughput sequencing technology (second-generation sequencing technology) for large scale transcriptome sequencing has dramatically improved the efficiency of genome annotation and gene discovery of non-model species, especially for Violaceae family [6–8]. However, there are not much research on *V. yedoensis* Makino, especially not enough for genomic study and functional gene identification. Simple sequence repeat (SSR) and single-nucleotide polymorphism (SNP) markers have been frequently used in the study of genetic identification and fingerprint mapping, and can be easily detected by the high-throughput sequencing technology [6]. Currently only few SSR

and SNP markers were developed in *V. yedoensis* Makino. SSR has been widely used in the study of genetic identification and fingerprint mapping with the characteristics of high polymorphic information content, simple technology, and good reproducibility. SNP is the most abundant type of markers and is easy to develop and detect by high-throughput sequencing technology [6].

Recent studies have shown that *V. yedoensis* Makino is almost normally distributed in a Lead (Pb)/ zinc (Zn)/cadmium (Cd) mine located in Sichuan Province, China. In our preliminary studies [9], Illumina paired-end sequencing technology was adopted to characterize the transcriptome of Violaceae through analysis of large-scale transcript sequences. The results would offer valuable sequence resource to Violaceae family and explore some unknown Cd-defensive strategies that possibly exist in the Viola species, as well as provide an efficient, inexpensive and reliable approach that can be readily adopted by researchers studying non-model organisms using transcriptome sequencing [9]. According to the results from the preliminary experiments, we characterize novel SSRs and SNPs of *V. yedoensis* Makino, and assess their functional annotation, classification, and phylogenetic analysis of cyclotide family genes.

MATERIALS AND METHODS

Sample preparation and transcriptome sequencing

V. yedoensis Makino were sown and transplanted in the nutrient solution (Table 1) as described previously. There were randomly divided to two groups, including control solution group (CK) and solution containing 2.5 mM $Cd(NO_3)_2$ group (Cd treatment). After 72 h, whole plants were collected for total RNA extraction using TRIzol kit (Invitrogen, USA), and then were quantified, purified, and pooled for subsequent Illumina sequencing. In our preliminary studies [9], the *de novo* transcriptome assembly for gene identification, analysis, functional annotation and classification were accessed. Based on this, in the present study, we determine to further identify novel SSR and SNP markers and cyclotide family unigenes in *V. yedoensis* Makino responsive to cadmium (Cd) pollution.

Table 1 The detailed components of the nutrient solution.

Component	Concentration
$Ca(NO_3)_2 \cdot 4H_2O$	4.0 mmol/l
KNO_3	6.0 mmol/l
NH_4NO_3	1.0 mmol/l
$Fe \cdot EDTA$	100 μmol/l
H_3BO_3	1.0 μmol/l
KH_2PO_4	1.0 μmol/l
$MnSO_4$	0.1 μmol/l
$CuSO_4 \cdot 7H_2O$	0.1 μmol/l
$MgSO_4 \cdot 7H_2O$	1.0 μmol/l
$(NH_4)_6Mo_7O_{24}(_4H_2O)$	0.016 μmol/l
$ZnSO_4 \cdot 7H_2O$	1.0 μmol/l

SSR prediction and SNP detection

Mining of the SSRs present in the assembled unigenes of Violaceae was performed as described previously [10]. Briefly, five categories of SSRs, including those with dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide motifs, were classified. For each category of SSR, the minimum number of contiguous repeat units is five. Statistical analysis was used to investigate the number of SSRs with each type of motif and the distribution of the lengths of repeat units. In addition, SNPs is also identified mentioned below: we firstly used Simple Object Access Protocol (SOAP) to map the reads sequences to the identified unigene, and put pair-end and single-end comparison results together, duplicated reads and multi-mapped reads were also filtered out, results were sorted according to the transcript and coordinate position, then we used SOAPsnp [11] to SNP calling for the sorted results, parameters were set as follows: "seed Length" 30, "minLength" 50, "minInsert" 100, "maxInsert" 1000, "misMatch" 3, "ASCII" "!". In order to get the high quality SNPs, the quantity and quality ($\geqslant 20$), the sequencing depth ($\geqslant 2$), and the SNP spacing ($\geqslant 5$) were also selected to filter out the results. SSRs predicted *in silico* were validated by PCR amplification. The products were separated on 1.5% agarose gels and visualized using Goldview Nucleic Acid Stain.

Phylogenetic analysis and RT-PCR validation of cyclotide family unigenes

Hidden Markov Model (HMM) profile of cyclotide family domain (PF03784) downloaded from Protein family (Pfam; http://pfam.xfam.org) was exploited for the identification of the cyclotide family genes from Violaceae assembly transcriptome with HMMER (v 2.3.2) [12]. Conserved sequences of cyclotide family was extracted from the HMM profile by the HMMER software [12], and then was adopted to query the Violaceae assembly transcriptome. Searching parameters were followings:

Table 2 The primers for identifing the cyclotide family unique genes.

Unique gene	Primer sequence	Size
comp109420_c0 Vya15	F:5′-TTACTACTCAAACCAAACCCTT-3′ R:5′-TTCGCTTCAGTTCAATCCTAC-3′	273bp
comp108309_c0 Vya10	F:5′-GTCATGTAACGATCAGT-3′ R:5′-GGAGATCTTTGTAGAAT-3′	343bp
comp109293_c0 Vya14	F:5′-TGCAGCTCCTTGTACTGA-3′ R:5′-CAAAAGCCCTTCTTCATA-3′	398bp
comp102243_c0 Vya1	F:5′-CGCAACAATAGATAGACA-3′ R:5′-TCACATCCAATACAACAG-3′	343bp
comp403622_c0 Vya26	F:5′- ACCAAAACCATCGTCTC-3′ R:5′-AGACAACAAGCGAAAC-3′	252bp
comp87919_c0 Vya37	F:5′-ACCAAAACCATCGTCTCAAACCCGG-3′ R:5′-CCGGGTTTGAGACGATGGTTTTGGT-3′	280bp
comp109209_c0 Vya13	F:5′-TGGAACGAAGGGAACA-3′ R:5′-AGACTTGCGTCGGTGG-3′	358bp
comp120436_c0 Vya21	F:5′-CATTACTACTCAAACCAAACCCTT-3′ R:5′-TTCGCTTCAGTTCAATCCTACCTC-3′	305bp
comp4648_c0 Vya27	F:5′-GTTTATCGACTTGTGCAG-3′ R:5′-ATGGAAAAGAAGAAGAAG-3′	388bp
comp104964_c0 Vya6	F:5′-ATCCAACTACCCAACA-3′ R:5′-GCCATGATTTAAGCGT-3′	394bp
comp99755_c1 Vya41	F:5′-GAGTTCTTGTAGCAAACT-3′ R:5′-AAGTCACATACCTTACCG-3′	419bp

BLASTp, E = 1e−10, and other parameters were defaulted. All non-redundant hits with expected values less than 1.0 were collected, and then the identified Violaceae cyclotide family unigenes were first aligned using CLUSTAL X version 2.0 [13]. Subsequently, based on the nucleotide sequence alignments, MEGA4.0 software was used to generate phylogenetic trees of Violaceae Unigenes using the neighbor-joining approach [14]. The phylogenetic trees of cyclotide family were furnished with 1000 pseudoreplicate bootstrap values at each node, with the similar cyclotide family from the *Melicytus ramiflorus* (Mra), *Viola baoshanensis* (Vba) and *Viola odorata* (Voa). The accession numbers for the sequences of cyclotide family is named Vya1-Vya41 in (Supplementary Table S1). In addition, 13 assembled unigenes were randomly selected from four cyclotide family groups for RT-PCR validation. Total RNA from *V. yedoensis* leaves was reverse-transcribed by using SuperScript III Reverse Transcriptase (Invitrogen) and oligo(dT)18. Forward (Fwd) and reverse (Rev) primers were designed using Primer3. The sequences of primers used for validation of assembled *V. yedoensis* cyclotide family unigenes were listed in Table 2.

Table 3 SSR analysis of the *V. yedoensis* transcriptome.

Unit size	No. of SSRs	Percent
1	6546	58.57
2	1732	15.50
3	2733	24.45
4	135	1.21
5	17	0.15
6	13	0.12

RESULTS AND DISCUSSION

SSR analysis of the Violaceae transcriptome

In our preliminary studies, the *de novo* transcriptome assembly for gene identification, analysis, functional annotation and classification were accessed [9]. However, the molecular marker development was not expounded. SSR markers, also known as microsatellites, are repeating DNA sequences of 2–6 base pairs, which are widely used as molecular markers for genetic mapping and to analyze species diversity. Approximate 91.99 million high-quality clean reads were generated. Then, 11176 SSRs were predicted in 9644 unigenes, of which, 1283 unigenes contained multiple SSRs. The SSRs included 1732 (15.5%) dinucleotide motifs, 2733 (24.45%) trinucleotide motifs, 135 (1.21%) tetranucleotide motifs, 17 (0.15%) pentanucleotide

Table 4 The distribution of the number of repeat units in all SSRs.

Repeat	5	6	7	8	9	10	11	12	All
AC	–	45	19	11	6	6	11		98
AG	–	131	64	40	22	18	12	1	288
AT	–	98	59	27	18	10	10		222
CA	–	35	18	16	7	6	3		85
CG	–	2	1						3
CT	–	105	44	24	21	15	4		213
GA	–	76	49	21	22	18	17	2	205
GC	–	8	1						9
GT	–	49	19	12	12	16	10		118
TA	–	91	35	37	21	19	5		208
TC	–	84	38	21	12	6	9		170
TG	–	44	33	15	7	5	9		113
AAC	27	8	11						46
AAG	51	25	15						91
AAT	26	11	10						47
ACA	31	5	2						39
ACC	23	10							33
ACG	1	1	1						3
ACT	6	2	1						9
AGA	34	21	11						66
AGC	59	16	3	1					79
AGG	46	18	9						73
AGT	12	2							14
ATA	15	5	8	1					29
ATC	38	13	5						56
ATG	30	15	4						49
ATT	23	9	5						37
CAA	45	11	7						63
CAC	29	11	3						43
CAG	59	23	7						89
CAT	26	12	3						41
CCA	35	16	4	1					56
CCG	10	3		1					14
CCT	41	19	2	1					63
CGA	5	4	1						10
CGC	7	4	2	1					14
CGG	7	5							12
CGT	5		1						6
CTA	5		2						7
CTC	48	15	5						68
CTG	42	9	5						56
CTT	32	14	6	1					53
GAA	59	23	13	1					96
GAC	4	3		1					8
GAG	77	36	10						124
GAT	36	13	9						58
GCA	41	21	11						73
GCC	5	1	2						8
GCG	6	2							8
GCT	42	16	8						66
GGA	80	18	5	2					105
GGC	22	7	2						31
GGT	51	13	7	3					74
GTA	6	5	1	1					13
GTC	8		1						9
GTG	35	12	5						52
GTT	28	7	10						45
TAA	12	7	11						30
TAC	13	2							15
TAG	6	4	1						11
TAT	20	6	5						31
TCA	44	6	5		1				56
TCC	45	22	5						72
TCG	10	3	2						15
TCT	36	17	7						60
TGA	38	17	4	1					60
TGC	47	19	4						70
TGG	45	19	6	1					71
TGT	15	6	5	1					27
TTA	24	8	4	1					37
TTC	40	13	9						62
TTG	58	14	8						80

Repeat	5	6	7	8	9	10	11	12	All
AAAC	2								2
AAAG	3								3
AAAT	2								2
AACC		1							1
AAGA	1	1							2
AATA	4								4
AATC	1								1
AATG	1								1
AATT	2								2
ACAA	1								1
ACAT	1								1
ACCA	1								1
ACTC	1								1
AGAA	1	1							2
AGGA	1								1
AGGG	1								1
ATAA	4	1							5
ATAC	4								4
ATCA		1							1
ATCC	1								1
ATGC	1								1
ATGG		1							1
ATGT	1								1
ATTA	4								4
ATTT	2								2
CAAA	2								2
CAAT	1								1
CACT	1	1							2
CATA	1								1
CATC	1								1
CCAA	1								1
CCAT		1							1
CCTC	1								1
CGTC	1								1
CTCA	1								1
CTCC	2								2
CTGA	1								1
CTTC	2	1							3
CTTT	1								1
GAAA	2	1							3
GAAG	4								4
GACA		1							1
GAGG	1								1
GATG	3	1							4
GCAT	1								1
GCCT	1								1
GTGA	1								1
GTTA	1								1
TAAA	3								3
TAAT	3	1							4
TACC	1								1
TATG	1								1
TATT	4								4
TCAA	1								1
TCAC	1								1
TCCA	1								1
TCCC	1								1
TCCG		1							1
TCCT	1								1
TCTT	3								3
TGCA	1								1
TGGA	2								2
TGGC	1								1
TGTA	3								3
TTAA	4								4
TTAT	1								1
TTCA	4								4
TTCC						1			1
TTCT	4								4
TTTA	4								4
TTTC	1								1
TTTG	5								5

Repeat	5	6	7	8	9	10	11	12	All
AAAGC			1						1
AAGCC	1								1
AATAA	1								1
AATCA	1								1
AGAAT	1								1
AGCTG	1								1
AGGCT	1								1
ATCAC	1								1
CCATC	1								1
CGATG	1								1
CGGAA		1							1
GAAAC	1								1
GAAGA			1						1
GGTTG	1								1
TGACC	1								1
TGAGC	1								1
TTGGT	1								1
ACGCCA		1							1
ACGGCT		1							1
ATCCCT	1								1
CACAGT		1							1
CAGGAA	1								1
CCCGGA		1							1
CCGGGA		1							1
GGGAGA		1							1
GTAGGG		1							1
GTGAGC	1								1
TGGTTA	1								1
TTCTCC							1		1
TTGACA		1							1

motifs, and 13 (0.12%) hexanucleotide motifs (Table 3). Furthermore, 163 motifs were detected among all of the SSRs identified. The most abundant SSR motif was GA/AG/TA/AT (9, 23 SSRs), followed by AAG/AGA/TTC/TCT/CTT/GAA/GAG (7, 22 SSRs), and then AC/CA/TC/CT (566 SSRs). The distribution of the number of repeat units in all SSRs was also investigated (Table 4). The result revealed that most SSRs contained fewer than 12 repeat sequences, and no SSRs with more than 20 repeat sequences were observed. Here, we predicted 6295 SSRs from the assembled unigene library of Violaceae. These SSRs will likely be of value for genetic analyses of species in the genus Violaceae and other related non-model plants. We selected primer S1 and primer S2 to check the PCR amplification and polymorphism of *V. yedoensis* samples, for example, 11 *V. yedoensis* samples with the SSR primers developed in this study were detected, and gel pictures of results were shown in Fig. 1. Our data suggested that SSRs locus predicted *in silico* could be very useful resources for SSR marker development of *V. yedoensis*.

SNP analysis of the Violaceae transcriptome

Furthermore, application of SNP marker technology has potential to improve plant genetics, and facilitate faster release of plant hybrids to the market place. The SNP usually occurs in non-coding regions more frequently than in coding regions or, in general, where natural selection is acting and fixating the allele of the SNP that constitutes the most favorable genetic adaptation [15]. Other factors, like genetic recombination and mutation rate, can also determine SNP density [16]. Here, in total, 8250 SNP markers were detected in 1934 unigenes and 11328 SNP markers were detected in 2510 unigenes in VIYCd and VIYCK library, respectively. Moreover, 83 homogenous SNPs were detected in 76 unigenes and 55 SNP markers were detected in 51 unigenes in two libraries, respectively (Fig. 2, Supplementary Tables S2 and S3).

Phylogenetic analysis of cyclotide family genes

Most interestingly, 41 cyclotide family genes were identified and then annotated based on assembly transcriptome, and 11 cyclotide family unique randomly selected from six groups (Fig. 3) were validated by RT-PCR. Of these, all of them could successfully yield amplicons, thereby provided evidence for the quality validation of our assembled cyclotide family Unigenes. In addition, Gene Ontology (GO) annotation showed that cyclotide family genes were involved in defense response and DNA repair/DNA metabolic process. Based on the Phylogenetic analysis of cyclotide family unigenes, we classified these cyclotide family into 6 groups, of which, comp109420 (Vya15) contained MADS-box domain by KOG classify which is clustered in group V (Fig. 4, Supplementary Table S1). It is interesting that group V was only found in our research.

Cyclotides are a novel family of plant-derived defense peptides that are biosynthetically produced via the processing of cyclotide precursor (CP) proteins containing one, two or three cyclotide domains. Cyclotide sequence analyses indicated that the cDNA clones encoded a variety of Möbius and bracelet cyclotides, which were involved in the known bioactivities of cyclotides, and also might play a previously unreported role in mediating the metal tolerance of *V. baoshanensis*. In addition, Cd-responsive gene Vb40 was identified, which could deduce a novel cysteine-rich mini-protein and could improve copper (Cu) tolerance in hosted yeast, indicating that these species-specific genes possibly functioned in *V. baoshanensis* heavy metal (Cd) tolerance [17]. In this study, 41 cyclotide family unigenes in Violaceae were annotated based on six known databases and categorized into six groups. Based on GO annotation, we demonstrated that these cyclotide family genes were probably involved in mediating the metal tolerance in Violaceae.

CONCLUSION

In our preliminary studies, we determined the *de novo* transcriptome assembly for gene identification, analysis, functional annotation and classification using Illumina paired-end sequencing technology for Violaceae transcriptome *de novo* sequencing and assembly without reference genome, and plenty of differential gene expression levels were measured in *V. yedoensis* Makino responsive to cadmium (Cd) pollution. However, the relative molecular marker development was not expounded. Based on this, in the present study, we determine to identify novel SSR and SNP markers in *V. yedoensis* Makino, and also study the functional annotation, classification, and phylogenetic analysis of cyclotide family genes. Taken together, such large number of sequences and deep depth of coverage can provide sufficient transcriptomic sequence information for discovering novel genes. Additionally, thousands of SSR markers produced in this study will enable genetic linkage mapping construction and gene-based association studies. The results demonstrated that Illumina paired end sequencing can be used as a fast

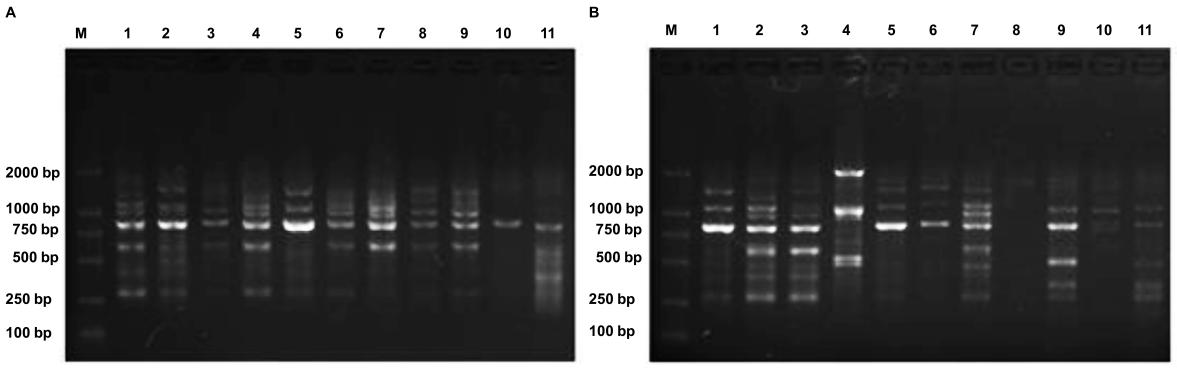

Fig. 1 PCR amplification of 11 *V. yedoensis* samples with the SSR primers developed in this study. (A) Primer S1 amplification results. (B) Primer S2 amplification results. Lane 1, GuLin of Luzhou in Sichuan; Lane 2, XuYong of Luzhou in Sichuan; Lane 3, NaXi of Luzhou in Sichuan; Lane 4, BanQiao of Zigong in Sichuan; Lane 5, BeiBei of Chongqing; Lane 6, HeChuan of Chongqing; Lane 7, YongChuan of Chongqing; Lane 8, WenHai of Lijiang in Yunnan; Lane 9, WeiYuan of Neijiang in Sichuan; Lane 10, XinWen of Yibing in Sichuan; Lane 11, GongXian of Yibing in Sichuan; Lane M, molecular marker DL2000.

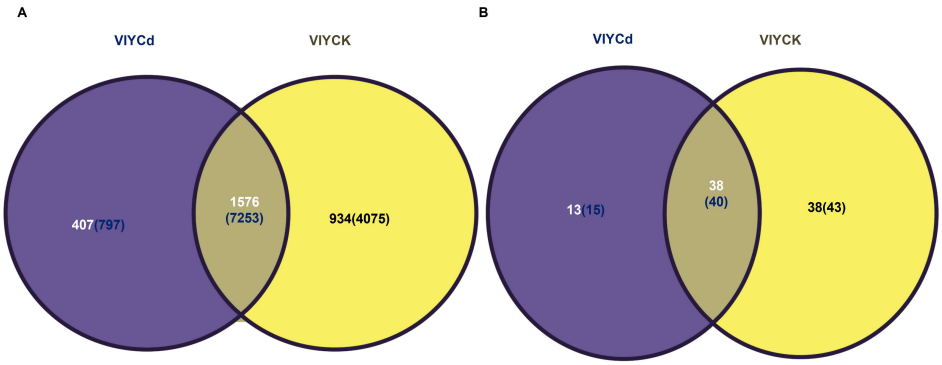

Fig. 2 The total and homogenous SNPs identified in two libraries. (A) Total SNPs were identified in two libraries. Violet and yellow color showed SNPs in CK and Cd libraries, respectively, 13927 SNPs were common distributed in both libraries. (B) Homogenous SNPs were identified in two libraries. Violet and yellow color showed homogenous SNPs in CK and Cd libraries, respectively, 321 homogenous SNPs were common distributed in both libraries.

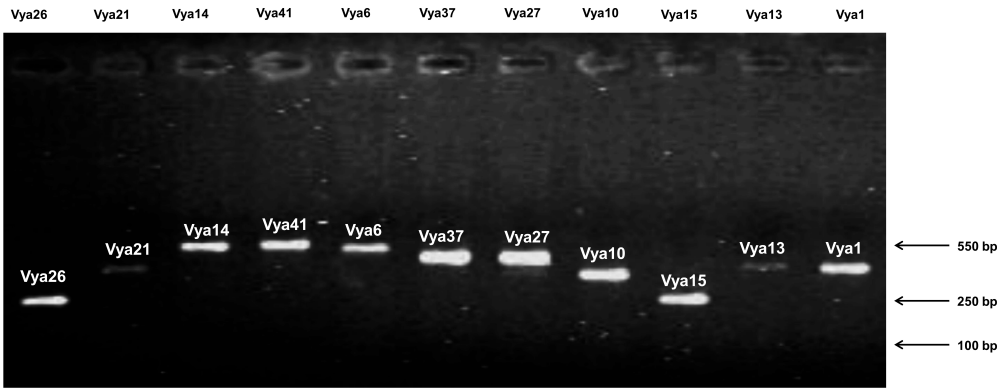

Fig. 3 Cyclotide family unique randomly selected from six groups validated by RT-PCR. The CDS length is in range from 252bp to 419bp in these 11 Unigenes.

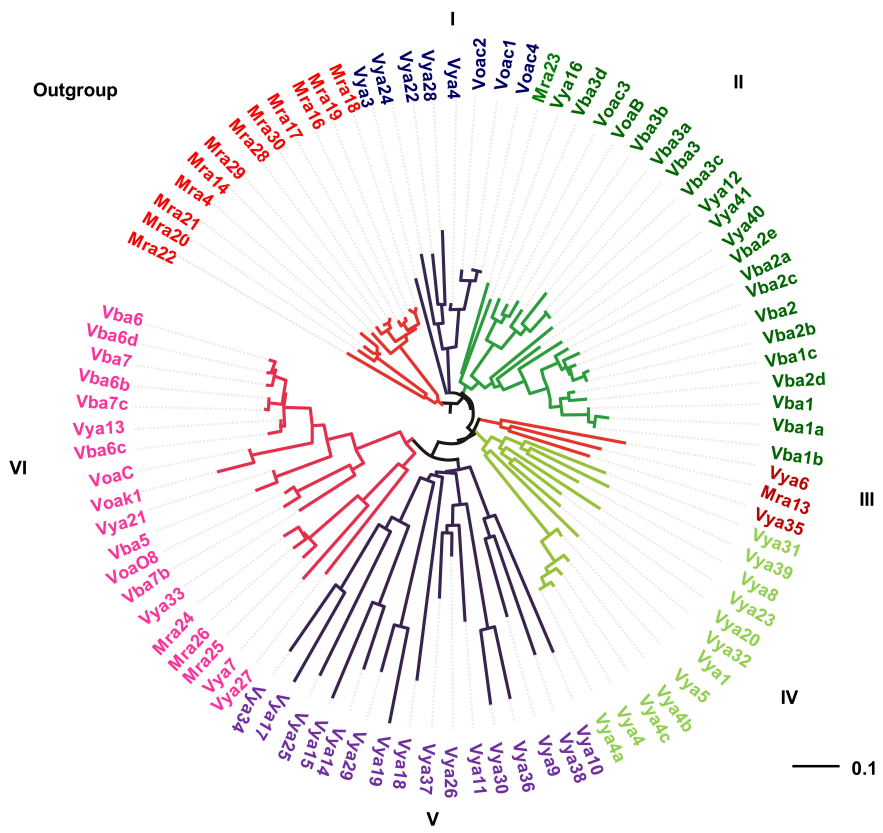

Fig. 4 The phylogenetic tree analysis of cyclotide family Unigenes based on transcriptome assembly and annotation. The phylogenetic tree of cyclotide family Unigenes in multiple plants were constructed using the neighbor-joining method, including *Melicytus ramiflorus* (Mra), *Viola baoshanensis* (Vba) and *Viola odorata* (Voa).

and cost-effective approach to the gene discovery and molecular marker development for non-model organism, especially those with large genome.

Supplementary data: Supplementary Tables S1, S2 and S3 are available upon request from the authors.

S1. Assembly and annotation of cyclotide genes in *V. yedoensis* based on the database.

S2. Total SNPs identified in unigenes of two libraries.

S3. Homogenous SNPs identified in unigenes of two libraries.

Acknowledgements: This work was supported by the National Natural Science Foundation of China (81800434, 81001700), Grant of Sichuan Province Science and Technology Agency Grant (14JC0137, 2018JY017 and 2019YJ0487), Grant of Chengdu Municipal Science and Technology Bureau of Sichuan Province (2019-YFYF-00167-SN), and the Foundation of Luzhou Municipal Science and Technology Bureau of Sichuan Province (2017LZXNYD-T05, 2016LZXNYD-J24 and 2019LZXNYDJ49).

REFERENCES

1. Li W, Xie JY, Li H, Zhang YY, Cao J, Cheng ZH, Chen DF (2012) *Viola yedoensis* liposoluble fraction ameliorates lipopolysaccharide-induced acute lung injury in mice. *Am J Chin Med* **40**, 1007–1018.
2. Xie C, Veitch NC, Houghton PJ, Simmonds MS (2003) Flavone C-glycosides from *Viola yedoensis* MAKINO. *Chem Pharm Bull (Tokyo)* **51**, 1204–1207.
3. Wang CK, Colgrave ML, Gustafson KR, Ireland DC, Goransson U, Craik DJ (2008) Anti-HIV cyclotides from the Chinese medicinal herb *Viola yedoensis*. *J Nat Prod* **71**, 47–52.
4. Zeng HR, Wang B, Zhao Z, Zhang Q, Liang MY, Yao YQ, Bian K, Zhang WR (2016) Effects of *Viola yedoensis* Makino anti-itching compound on degranulation and cytokine generation in RBL-2H3 mast cells. *J Ethnopharmacol* **189**, 132–138.
5. Jeong YH, Oh YC, Cho WK, Shin H, Lee KY, Ma JY (2016) Anti-inflammatory effects of *Viola yedoensis* and the application of cell extraction methods for investigating bioactive constituents in macrophages. *BMC Complement Altern Med* **16**, ID 180.

6. Zou D, Chen X, Zou D (2013) Sequencing, *de novo* assembly, annotation and SSR and SNP detection of sabaigrass (*Eulaliopsis binata*) transcriptome. *Genomics* **102**, 57–62.
7. Mora-Ortiz M, Swain MT, Vickers MJ, Hegarty MJ, Kelly R, Smith LM, Skøt L (2016) De-novo transcriptome assembly for gene identification, analysis, annotation, and molecular marker discovery in *Onobrychis viciifolia*. *BMC Genomics* **17**, 756.
8. Liang M, Yang X, Li H, Su S, Yi H, Chai L, Deng X (2015) *De novo* transcriptome assembly of pummelo and molecular marker development. *PLoS One* **10**, e0120615.
9. Gao J, Luo M, Zhu Y, He Y, Wang Q, Zhang C (2015) Transcriptome sequencing and differential gene expression analysis in *Viola yedoensis* Makino (Fam. Violaceae) responsive to cadmium (Cd) pollution. *Biochem Biophys Res Commun* **459**, 60–65.
10. Zhang J, Liang S, Duan J, Wang J, Chen S, Cheng Z, Zhang Q, Liang X, et al (2012) *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC Genomics* **13**, ID 90.
11. Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, Wang J (2009) SNP detection for massively parallel whole-genome resequencing. *Genome Res* **19**, 1124–1132.
12. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* **14**, 755–763.
13. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
14. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.
15. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. *Nat Genet* **40**, 340–345.
16. Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* **5**, 94–100.
17. Zhang J, Hu M, Li JT, Guan JP, Yang B, Shu WS, Liao B (2009) A transcriptional profile of metallophyte *Viola baoshanensis* involved in general and species-specific cadmium-defense mechanisms. *J Plant Physiol* **166**, 862–870.