

DOI: 10.5586/asbp.3510

Publication history

Received: 2015-11-11

Accepted: 2016-08-18

Published: 2016-09-26

Handling editor

Przemysław Wojtaszek, Faculty of Biology, Adam Mickiewicz University, Poland

Authors' contributions

JHC, MBL: conceived and designed the experiments; MBL, CCL, XD, XFW: performed the experiments; MBL, ZMY, YX, JJZ: analyzed the data; ZMS: collected plant material in the field; MBL, JJZ, XD: wrote the paper

Funding

This study was financially supported by the National Natural Science Foundation of China (grant No. 30900913 and 31471547), Hunan Provincial Natural Science Foundation of China (grant No. 13JJ4116), National Modern Agro-industry Technology Research System (grant No. CARS-19-E04), and National Key Technology Support Program (grant No. 2013BAD01B03-13).

Competing interests

No competing interests have been declared.

Copyright notice

© The Author(s) 2016. This is an Open Access article distributed under the terms of the [Creative Commons Attribution License](#), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

Citation

Luan MB, Yang ZM, Zhu JJ, Deng X, Liu CC, Wang XF, et al. Identification, evaluation, and application of the genomic-SSR loci in ramie. *Acta Soc Bot Pol.* 2016;85(3):3510. <http://dx.doi.org/10.5586/asbp.3510>

Digital signature

This PDF has been certified using digital signature with a trusted timestamp to assure its origin and integrity. A verification trust dialog appears on the PDF document when it is opened in a compatible PDF reader. Certificate properties provide further details such as certification time and a signing reason in case any alterations made to the final content. If the certificate is missing or invalid it is recommended to verify the article on the journal website.

ORIGINAL RESEARCH PAPER

Identification, evaluation, and application of the genomic-SSR loci in ramie

Ming-Bao Luan, Ze-Mao Yang, Juan-Juan Zhu, Xin Deng, Chen-Chen Liu, Xiao-Fei Wang, Ying Xu, Zhi-Min Sun, Jian-Hua Chen*

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences / Key laboratory of Stem-fiber Biomass and Engineering Microbiology, Ministry of Agriculture, 348 West Xianjiahu Road, Changsha, Hunan, China

* Corresponding author. Email: ramiere@126.com

Abstract

To provide a theoretical and practical foundation for ramie genetic analysis, simple sequence repeats (SSRs) were identified in the ramie genome and employed in this study. From the 115 369 sequences of a specific-locus amplified fragment library, a type of reduced representation library obtained by high-throughput sequencing, we identified 4774 sequences containing 5064 SSR motifs. SSRs of ramie included repeat motifs with lengths of 1 to 6 nucleotides, and the abundance of each motif type varied greatly. We found that mononucleotide, dinucleotide, and trinucleotide repeat motifs were the most prevalent (95.91%). A total of 98 distinct motif types were detected in the genomic-SSRs of ramie. Of them, The A/T mononucleotide motif was the most abundant, accounting for 41.45% of motifs, followed by AT/TA, accounting for 20.30%. The number of alleles per locus in 31 polymorphic microsatellite loci ranged from 2 to 7, and observed and expected heterozygosities ranged from 0.04 to 1.00 and 0.04 to 0.83, respectively. Furthermore, molecular identity cards (IDs) of the germplasms were constructed employing the ID Analysis 3.0 software. In the current study, the 26 germplasms of ramie can be distinguished by a combination of five SSR primers including Ibg5-5, Ibg3-210, Ibg1-11, Ibg6-468, and Ibg6-481. The allele polymorphisms produced by all SSR primers were used to analyze genetic relationships among the germplasms. The similarity coefficients ranged from 0.41 to 0.88. We found that these 26 germplasms were clustered into five categories using UPGMA, with poor correlation between germplasm and geographical distribution. Our study is the first large-scale SSR identification from ramie genomic sequences. We have further studied the SSR distribution pattern in the ramie genome, and proposed that it is possible to develop SSR loci from genomic data for population genetics studies, linkage mapping, quantitative trait locus mapping, cultivar fingerprinting, and as genetic diversity studies.

Keywords

ramie; genomic-SSR; SLAF

Introduction

Ramie [*Boehmeria nivea* (L.) Gaudich], also known as Chinese grass, is a specialty fiber crop that provides raw material for the textile industry. It is also known as the second most important fiber crop, after cotton, in China [1]. The bark (phloem) of the vegetative stalks is used principally for fabric production. To date, products such as shirts, underwear, and health-care socks have been developed and sold on the home market [2]. Its leaves are rich in crude protein, crude fiber, lysine, methionine,

carotenoids, vitamin B2, and calcium, and have a high yield of approximately 7.5 t dry weight ha⁻¹. Therefore, it is the best concentrated feed for various ruminant animals and poultry in southern China [3].

In order to improve quality and yield in crops, molecular markers have been widely used to study germplasm diversity [4,5]. To date, numerous molecular marker technologies continue to be developed and used in plant genetic analysis. A simple sequence repeat (SSR) is a DNA repeat consisting of 1–6 nucleotide repeat units [6]. SSRs have frequently been employed to analyze the genomes of plants due to their simple technology, site specificity, high reproducibility, relatively high polymorphism, codominant inheritance, and relative abundance in genomes [7]. In recent years, there have been many reports focused on the development and characterization of SSR markers in plants [8,9]. The development of SSR markers can include two types: one based on expressed sequence tags (ESTs), called EST-SSR [10,11], and another based on the genomic sequence, called genomic-SSR [12,13]. EST-SSR markers are derived from gene coding sequences and illustrate gene expression. This provides information that can help to directly identify functional genes [14]. Genomic-SSR comes from genomic sequences, which include both non-coding and coding sequences. Genetic polymorphism is significantly higher in genomic-SSR than in EST-SSR, but the development of genomic-SSR has a high cost and long development cycle compared to EST-SSR. However, genomic-SSR has great potential in the location of trait regulation factors, as many regulatory functions of non-coding sequences have been discovered [15].

At present, there are less than 2000 EST-SSR primer pairs [16,17] and 18 genomic-SSR markers [18] developed in ramie. This is because traditional development of genomic-SSR primers requiring construction of genomic libraries, probe hybridization, cloning, and sequencing, is laborious, time-consuming, expensive, and inefficient. Thus, the development of ramie genomic-SSR markers has lagged behind that of EST-SSR markers. With the development of next generation sequencing technology, the cost of sequencing has been reduced, and a large number of DNA sequences can rapidly be obtained, which provides a good resource for the development of genomic-SSR markers. Thereby, development of genomic-SSR markers has become convenient, low-cost, and fast, leading to great advancements in many crops [19,20]. In this paper, high throughput sequencing technology was applied to detect genomic-SSRs in ramie for the first time, and the characteristics and regularity of the SSR sites were analyzed. Of these genomic-SSRs, 31 markers were developed, and the characterization and application of those markers is reported.

Material and methods

Plant material and DNA isolation

In this study, 26 ramie accessions (Tab. 1) growing in the Changsha (112°54' E, 28°18' N) were used. DNA was isolated from young leaves collected from each ramie accession using the Plant Genomic DNA Kit (DP305; Tiangen, Tiangen Biotech Co., LTD. Beijing, China). No specific permits were required for the described field studies, and the field studies did not involve endangered or protected species.

Specific-locus amplified fragment sequencing library construction

Specific-locus amplified fragment sequencing (SLAF-seq) is an efficient method of large-scale genotyping, which is based on reduced representation library (RRL) methods and high-throughput sequencing [21]. The following procedure was performed as described by Sun et al. [21] with modifications. Since the genome sequence of ramie has not been published, it is difficult to select an optimal enzyme to digest genomic DNA for a short fragment library. In this study, according to the hemp genome sequence [22], which is closely related to ramie, the RSI1 enzyme was chosen to digest the genomic DNA of ramie (germplasm codes 25th and 26th in Tab. 1). DNA fragments of

Tab. 1 Ramie germplasms and its ID.

Code	Germplasm	Origin	ID
1	Dayujiandaobai	Hubei Province	11111
2	Tongmuqingma	Guangxi Province	21212
3	Dayehongzhameng	Jiangxi Province	12322
4	Xiaoqinggan	Sichuan Province	23422
5	Lichuanhoupizhuma	Jiangxi Province	23423
6	Pingchangjima	Sichuan Province	24422
7	Gaoanma	Jiangxi Province	14532
8	Sima	Shannxi Province	14424
9	Anlongzhumaerhao	Sichuan Province	24222
10	Gegenma	Hunan Province	04223
11	Nanchenghoupizhuma	Jiangxi Province	24120
12	Zixima	Jiangxi Province	35223
13	Xunyanglvbaima	Shannxi Province	45125
14	Manniangzhuma	Hubei Province	51223
15	Leiyanghuangkema	Hunan Province	56325
16	Yinnisanhao	Indonesia	46435
17	Miaobazhuma	Sichuan Province	64443
18	Huangpinghuangganma	Guizhou Province	10425
19	Yizhangyuanma	Hunan Province	74215
20	Huangmazhu	Hainan Province	81632
21	Ziyangdayepao	Shannxi Province	71255
22	Nanchengbopizhuma	Jiangxi Province	83253
23	Enshiqingmaerhao	Hubei Province	74255
24	Yedouzi	Jiangxi Province	87263
25	Zhongzhuyihao	Hunan Province	14325
26	Hejiangqingma	Sichuan Province	15325

180–304 bp were then excised and diluted for pair-end sequencing on an Illumina High-seq 2500 sequencing platform (Illumina, Inc; San Diego, CA, USA) at Biomarker Technologies Corporation in Beijing. Real-time monitoring was performed for each cycle during sequencing. The ratio of high quality reads (quality scores above Q20), which indicates a 1% possibility of error and 99% confidence, and guanine–cytosine (GC) content were calculated for quality control.

SSR identification and SSR primer design

From the SLAF library of Zhongzhuyihao (code 25th in Tab. 1) and Hejiangqingma (code 26th in Tab. 1) genome, the simple sequence repeat identification tool (SSRIT; <http://archive.gramene.org/db/markers/ssr-tool>) was used to select sequences containing SSR motifs that contained clear repeated units. Sequences were selected that contained mono-, di-, tri-, tetra-, penta-, or hexanucleotide units repeated at least 15, 8, 5, 4, 3, and 3 times, respectively.

In addition, 43 SSR primers, identified with “Ibg” (Institute of Bast Fiber Crops, Chinese Academy of Agricultural Science, genome) as a short prefix, were designed with the Primer 5 software. The main parameters for primer design were set as below: primer length 18–24 bp with 20 bp as the optimum; PCR product size 80–200 bp; optimum annealing temperature 55°C; GC content 35–60% with 50% as the optimum. The primers were synthesized by Bioasia Biotech (Shanghai, China).

SSR evaluation

For 26 ramie varieties (Tab. 1, codes 1–26), PCR reactions were carried out in 10- μ L reaction volumes with 1 \times PCR buffer, 0.2 mM dNTP, 1 U Taq DNA polymerase (Tiangen), 0.5 μ M each primer and 0.5 μ L DNA under

the following PCR conditions: 5 min at 94°C; 30 cycles of 30 s at 95°C; 30 s at the primer-specific annealing temperature; 30 s at 72°C; and a final extension of 10 min at 72°C. The PCR products were separated on 8% polyacrylamide gels using electrophoresis, and silver staining was conducted according to Zhang et al. [23]. Clear bands from PCR products were genotyped.

Statistical analyses

Observed heterozygosity (HO), expected heterozygosity (HE) and significant deviations from Hardy–Weinberg equilibrium (HWE) were calculated using Popgen 1.31 [24]. The molecular identity cards (IDs) of ramie germplasms were constructed using ID Analysis 3.0 (software registration number 2007SR11870, North-east Agriculture University, China). The genetic similarity matrix was obtained using the SIMQUAL

sub-routine of the NTSYS-pc statistical software package [25] based on Jaccard's algorithms. The similarity coefficients were used to construct genetic distance phenograms by the SHAN method based on the unweighted pair-group method with arithmetic averages (UPGMA). The cluster analyses were performed using NTSYS-PC software version 2.1 [25].

Results

SLAF-seq library construction

By Illumina sequencing, we obtained 3 937 489 reads (total 397 686 389 bp, average 80 bp) from Zhongzhuyihao and 3 604 166 reads (total 364 020 766 bp, average 80 bp) from Hejiangqingma. Each of the read fragments was considered one marker site. The reads of one marker site were clustered into one group by similarity clustering.

In general, each group included 1–4 high-depth fragments and the rest were low-depth fragments. The high-depth fragments were often regarded as potential genotypes, and low-depth fragments may have resulted from sequencing errors. To correct sequencing errors, mismatched base pairs were corrected by comparing the low-depth fragments with the reference sequence, obtained from the accurate high-depth fragments, within a group. Most SLAF markers were consistent for both samples. Some SLAF markers may not be identified in the various samples owing to errors as a result of sample digestion. In this study, 81 665 and 91 465 SLAF markers from two ramie accessions, Zhongzhuyihao and Hejiangqingma, respectively, were obtained, and a total of 115 369 SLAF markers for the two ramie accessions (Tab. 2) were used to identify SSR loci.

Tab. 2 SLAF markers for both samples.

Category	SLAF markers
Zhongzhuyihao	81 665
Hejiangqingma	91 465
Consistent for both samples	57 761
Total for both samples	115 369

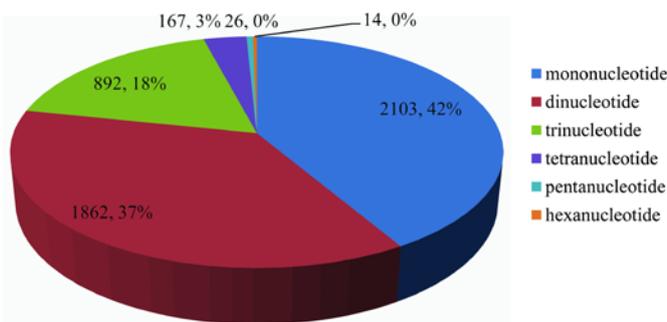


Fig. 1 Distribution of the SSR repeat motif. The graph was based on a total of $N = 5064$ SSRs detected in 115 369 SLAF markers of ramie genome.

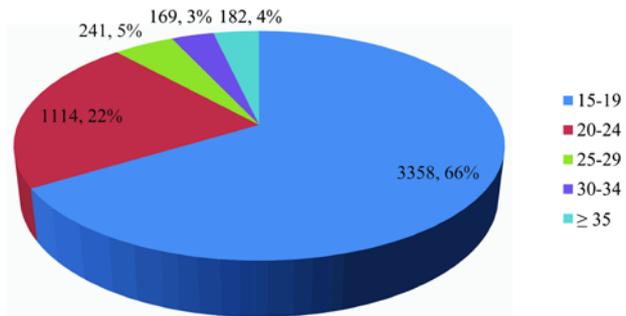


Fig. 2 Distribution of the length of repeats. The graph was based on a total of $N = 5064$ SSRs detected in 115 369 SLAF markers of ramie genome.

Characteristics of the genomic-SSR loci

From 115 369 sequences of the SLAF library, 4774 sequences containing SSR motifs (4.14%) were identified. The average distance between two SSRs was 3.65 kb. Of the 4774 sequences containing SSR motifs, there were 264 with two SSRs, 10 with three SSRs, and two with four SSRs. SSRs of ramie had a variety of repeat lengths such as mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats, but the frequencies of these SSR repeat types were different. Mononucleotide, dinucleotide, trinucleotide motifs were the most common motif type (95.91%), accounting for 41.53%, 36.77%, and 17.61%, respectively. The least frequent repeat motifs were tetranucleotide, pentanucleotide, and hexanucleotide repeats, only accounting for 4.09% (Fig. 1).

The length of SSR motifs ranged from 15 bp to 72 bp. The most abundant SSR motifs were those with lengths ranging from 15 bp to 19 bp, accounting for 66.31%, the next most abundant SSR motif length was in the range from 20 bp to 24 bp (22.00%), and the least abundant was from 30 bp to 34 bp in length (3.34%; Fig. 2).

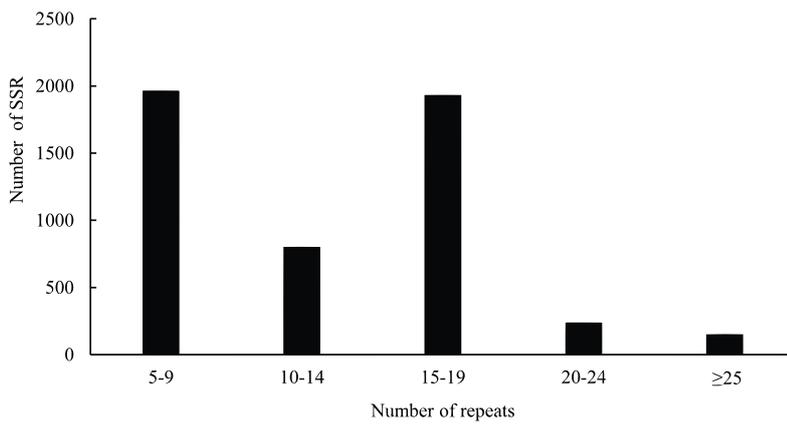


Fig. 3 Distribution of the number of repeats. The graph was based on a total of $N = 5064$ SSRs detected in 115 369 SLAF markers of ramie genome.

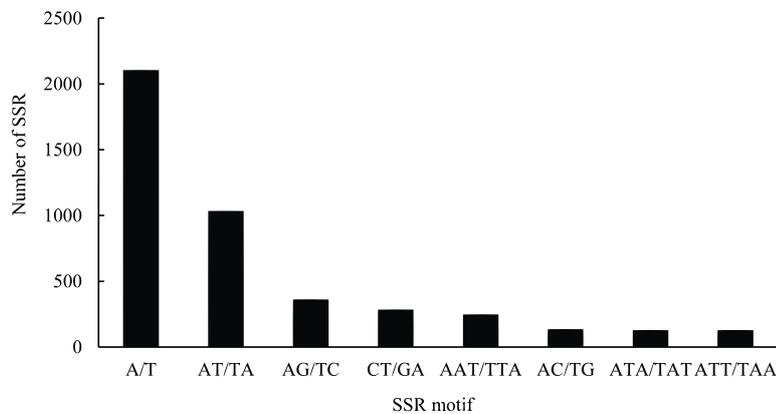


Fig. 4 The number of principal type in SSR motif.

Tab. 3 Main repeat motif in the different type of repeat.

Type of repeat	Main repeat motif
Mononucleotide	A/T
Dinucleotide	AT/TA, AG/TC, CT/GA
Trinucleotide	AAT/TTA, ATA/TAT, ATT/TAA
Tetranucleotide	AAAT/TTTA, ATTT/TAAA
Pentanucleotide	AAAAT/TAAAA
Hexanucleotide	CTTTTT/GAAAAA

SSR primers validation

We developed 43 pairs of SSR primers using the DNA of 24 individual ramie germplasms (codes 1–24) as templates. The results were summarized in Tab. 4. Thirty-one of 139 polymorphic loci were successfully amplified. The number of alleles per locus ranged from 2 to 7, and observed and expected heterozygosities ranged from 0.04 to 1.00 and from 0.04 to 0.83, respectively. Ibg2-211 had the highest Shannon's information index (1.66), while Ibg3-216 and Ibg4-52 had the lowest Shannon's information index (0.10). Of the 31 loci, 15 displayed significant deviations from Hardy–Weinberg expectations ($p < 0.05$). Of the remaining 12 primer pairs, seven failed to amplify clear bands and five markers were monomorphic.

Construction of 26 germplasm identity cards

PCR amplification banding patterns of 26 ramie germplasms were obtained from amplification using 31 SSR primers expressed 1, 2, ..., n , respectively. Molecular identity cards (IDs) of the 26 germplasms were constructed using the ID Analysis 3.0 software.

The number of repeats in each SSR varied from 5 to 65. The majority of SSRs had 5–9 repeats, accounting for 38.70% of SSRs, and the next most common were: 15 to 19 repeats (38.07%), 10 to 14 (15.74%), 20 to 24 (4.60%), and more than 25 (2.88%; Fig. 3). The largest number of repeats for each repeat length was 65, 35, 20, 9, 6, and 9 for the mono-, di-, tri-, tetra-, penta-, and hexanucleotides, respectively.

A total of 98 repeat types were detected in the genomic-SSRs of ramie. There were 2, 6, 30, 30, 19, and 11 different types of repeat motifs within mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively. The A/T repeat was the most abundant, accounting for 41.45% of all motifs, AT/TA repeats accounted for 20.30%, and other repeats that represented more than 5% of motifs included AG/TC, CT/GA, AAT/TTA, AC/TG, ATA/TAT, and ATT/TAA (Fig. 4). Among the 2103 mononucleotide motifs, 2099 (99.81%) were composed of A/T repeats, 55.21% of dinucleotide motifs contained AT/TA repeats, and 30.28% of trinucleotide motifs contained AAT/TTA repeats. AAAT/TAAAA and ATTT/TAAAA repeats accounted for 36.53% of tetranucleotide motifs. The results (Tab. 3) suggested that the base composition of ramie repeat units showed a preference for A and T nucleotides, and the most common repeat units were abundant in A and/or T.

Tab. 4 Characterization of 31 microsatellite loci, described by locus name, forward (F) and reverse (R) primer sequences, repeat motif, optimal annealing temperatures (T_a), size of allele (Size), number of alleles (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E), Shannon's information index (SII), and significance of departure from Hardy-Weinberg equilibrium (HWE). Polymorphism statistics were performed on 24 individuals.

Locus	Repeat	Primer sequences (5'-3')	Size (bp)	T_a (°C)	N_A	H_O	H_E	SII	HWE p -value
Ibg1-11	(T) ₁₆	F: CGGTCTTAAATTCAGTC R: TAATCTCATAGGCACCAC	121	57	5	0.54	0.53	1.04	0.22
Ibg1-16	(T) ₁₈	F: TAAGGCAGACCAAATTAC R: ACCAAAATCAAGAGACTC	126	55	3	0.40	0.66	0.68	0.09
Ibg1-20	(T) ₁₆	F: TGCCAAACATAAAACAATG R: ACTCGAATGAATTTAFTCTC	118	54	4	0.44	0.63	1.11	0.02
Ibg1-27	(T) ₁₅	F: TTCGTTTCTATATGAATTGG R: ACGGTGATAGTGATAAC	115	54	3	0.55	0.45	0.78	0.27
Ibg1-31	(A) ₁₅	F: AGAAATTTAAAGCAIGAGAAC R: ACAATAGTACTGTGAAG	110	54	2	0.65	0.45	0.63	0.02
Ibg2-207	(AT) ₉	F: TTAATTAATCCCTCTTTTAGC R: GCTCTTCTTATAGTTC	123	53	6	0.43	0.71	1.41	0.00
Ibg2-210	(TA) ₈	F: ACTTTGTTAGACTGATTTATC R: TTCTCCACCTTAATTTCC	136	55	7	0.67	0.69	1.36	0.00
Ibg2-211	(AT) ₁₃	F: ACATAATGAAATAATCAGTAAC R: GGAAGAATAAACCCTCTAATTG	149	55	6	0.70	0.83	1.66	0.00
Ibg2-213	(TA) ₁₀	F: TTAGTCAAATCTTGCTATG R: GCATCTATCAGACATTAC	108	54	4	0.96	0.68	1.20	0.00
Ibg2-214	(AC) ₉	F: AAGTTAGGTTAAGGAGTTG R: GTAGCCACATACATAAATG	110	55	4	0.95	0.57	0.94	0.00
Ibg2-218	(TC) ₁₇	F: AATACCGAGAGCGGAAG R: AAACAAAACAACACTGACAAAC	163	55	3	0.78	0.60	0.98	0.10
Ibg3-206	(TTA) ₆	F: GGAAAGGCTAAAACCTTAC R: AAGATGGGTTAGTCAATC	102	54	2	0.05	0.05	0.50	1.00
Ibg3-209	(TAA) ₅	F: AGAAATAAAGTTTCATCAGTC R: GCTGTCATTAAGGATAATAAG	114	54	6	0.57	0.83	0.96	0.00

Tab. 4 Continued

Locus	Repeat	Primer sequences (5'-3')	Size (bp)	T _a (°C)	N _A	H _o	H _E	SII	HWE p-value
Ibg3-210	(TTA) ₅	F: AAACGAAATTTGTCCCTTTG R: TTATTGTCAATTACCATCTTG	139	54	4	0.57	0.59	1.10	0.35
Ibg3-212	(ATT) ₅	F: TAGTCCATAGCAGTAATTC R: TAGCGTCTTAGAGAGAG	145	56	4	0.57	0.51	0.95	0.58
Ibg3-214	(AAT) ₅	F: ATCATCTCCAGAAATTC R: GTAGCCACATACATAAATG	148	55	4	0.62	0.53	0.98	0.67
Ibg3-216	(TTA) ₅	F: TGGAGGTGAAAGATAATG R: ACACAATACAATTACCAATC	156	55	2	0.04	0.04	0.10	1.00
Ibg3-218	(ATA) ₅	F: ATGTATATGAAACCTAAACTG R: GATTTGAATTTCTCTTTTCG	93	54	5	0.36	0.33	0.70	1.00
Ibg4-41	(TTAT) ₅	F: AATATAGCATCAATGAGAAC R: GATAGTTAAGTAATTTGGAAATG	143	53	5	0.47	0.69	1.24	0.28
Ibg4-49	(TATT) ₆	F: GCTTATTACCTTGAACCTTG R: ATATATTGCCGTGTTTAG	128	53	4	0.33	0.62	1.07	0.04
Ibg4-50	(AATT) ₅	F: GACGAGGATGATTCATTTG R: TACGGATGTTACTAGGTG	151	57	3	1.00	0.53	0.79	0.00
Ibg4-52	(ATAA) ₅	F: ACTAATCATCCATTACATATTTG R: TTTAATGGTTCGTTTGAG	145	54	2	0.04	0.04	0.10	1.00
Ibg4-53	(AATT) ₅	F: TATATGTTTGTATACCTAACG R: ATGAAGTAGAACCCAAACC	141	55	2	0.36	0.47	0.66	0.27
Ibg4-57	(TTTA) ₅	F: AGATCAGTTCAACTACAC R: TTCTACAACACCCCAATAG	115	55	4	0.50	0.50	0.86	0.88
Ibg5-1	(CATGT) ₅	F: ACAAACACAAACATACAAC R: GCATATAGAATGAAATAGAGAAG	132	55	3	0.47	0.51	1.16	0.00
Ibg5-3	(AAAAAT) ₅	F: TTAAGTGTAGATTCAAAGC R: AATAAGTTCGTGTCIC	144	55	3	0.47	0.44	0.76	0.81
Ibg5-5	(AAACC) ₆	F: CATTCACTTCGTTCAAG R: GCATTGTATTGTGTATAGG	135	55	3	0.76	0.53	1.33	0.11

Locus	Repeat	Primer sequences (5'-3')	Size (bp)	T _a (°C)	N _A	H ₀	H _E	SII	HWE p-value
Ibg5-6	(TTATT)5	F: TATAAATAGTGTAACTTCCAC R: TGTTTTCTAATTTTAGTTCAC	108	54	4	0.23	0.45	0.86	0.04
Ibg6-468	(A)17	F: CTAATCATAATAATTTGTTCCAC R: TATAGGCTCAATAAATTG	96	53	4	0.71	0.72	1.29	0.00
Ibg6-477	(A)20(CA)8	F: TCCTAAATTAATAATACACTAAG R: TCCTAAGACTTCTTTTCC	109	55	4	0.68	0.65	1.17	0.01
Ibg6-481	(A)22(A)16	F: CAAATGCTGAAATAAAAATAAAG R: TTGTGATAACTGTCTCTTC	124	54	4	0.39	0.65	1.11	0.00

Tab. 4 Continued

The results showed that 17 unique band types were developed using 10 SSR primers. In the current study, the 26 germplasm resources of ramie could be distinguished by five SSR primers ($k = 5$) including Ibg5-5, Ibg3-210, Ibg1-11, Ibg6-468, and Ibg6-481. The 5-bit molecular ID is shown in Tab. 1.

Analysis of genetic diversity

All allele polymorphisms produced by the SSR primers were used to analyze genetic relationships among the 26 ramie germplasms. The similarity coefficient ranged from 0.41 to 0.88. The lowest similarity coefficient was 0.41 between code 2th and code 12th and 19th, while the highest similarity coefficient was 0.88 between code 25th and code 26th. The germplasms were clustered into five categories using UPGMA (Fig. 5). The second category contained 18 germplasms, while the remaining categories contained only two germplasms each. Germplasms from the same provinces did not always cluster into the same group, while those from the different provinces could be clustered into the same group. For example, Nanchenghoupizhuma (code 11th) and Nanchengbopizhuma (code 22th), which were both collected from the same region, did not cluster into the same group. Yizhangyuanma (code 19th), from the Hunan Province, had the large similarity coefficient with Enshiqingmaerhao (code 23rd), which was collected from the Hubei Province. Thus, the results suggested that there was no significant correlation between germplasm and geographical distribution.

Discussion

SSR loci and evolutionary relationship

Single nucleotide repeat motifs comprise the largest portion of ramie genomic-SSR loci, followed by dinucleotide and trinucleotide repeat motifs. This composition was similar to grapes (*Vitis vinifera* L.) [26], *Saccharomyces cerevisiae*, vertebrates, mammals, arthropods [27], rice blast fungus [28], and nine other kinds of fungi [29]. However, this was significantly different from *Caenorhabditis elegans* [27], sorghum (*Sorghum bicolor* L.) [30], soybean rhizobia [31], *Mesorhizobium loti* [31], and *Sinorhizobium meliloti* [31], in which the genomic-SSR loci are comprised of nucleotide repeat units ranging from 4 bp to 6 bp in length. Generally, a large number of short repeat motifs show a relatively high mutational frequency for the species [27], while a majority of long repeat motifs show low mutational frequency or a short evolutionary time for the species [31–33]. Many short repeat motifs existed in the ramie genome. This indicates that ramie may have experienced a long evolution time or a high mutational frequency.

Characteristics of the genomic-SSR loci

The average distance between two genomic-SSRs is 3.65 kb in ramie, which is higher than in trees such as grape (*V. vinifera* L.; 2.99 kb) [26], apple (*Malus domestica* Borkh.; 3.22 kb) [19], and Chinese fir [*Cunninghamia lanceolata* (Lamb.) Hook.; 0.96 kb] [34], and less than in theophytes such as foxtail millet [*Setaria italica* (L.) Beauv.; 7.36 kb] [35] and flax (*Linum usitatissimum* L.; 28.7 kb) [36]. Single nucleotide repeat motifs of ramie genomic-SSRs were mostly composed of A/T repeats, accounting for 99.8%. The main types of dinucleotide repeat motifs were

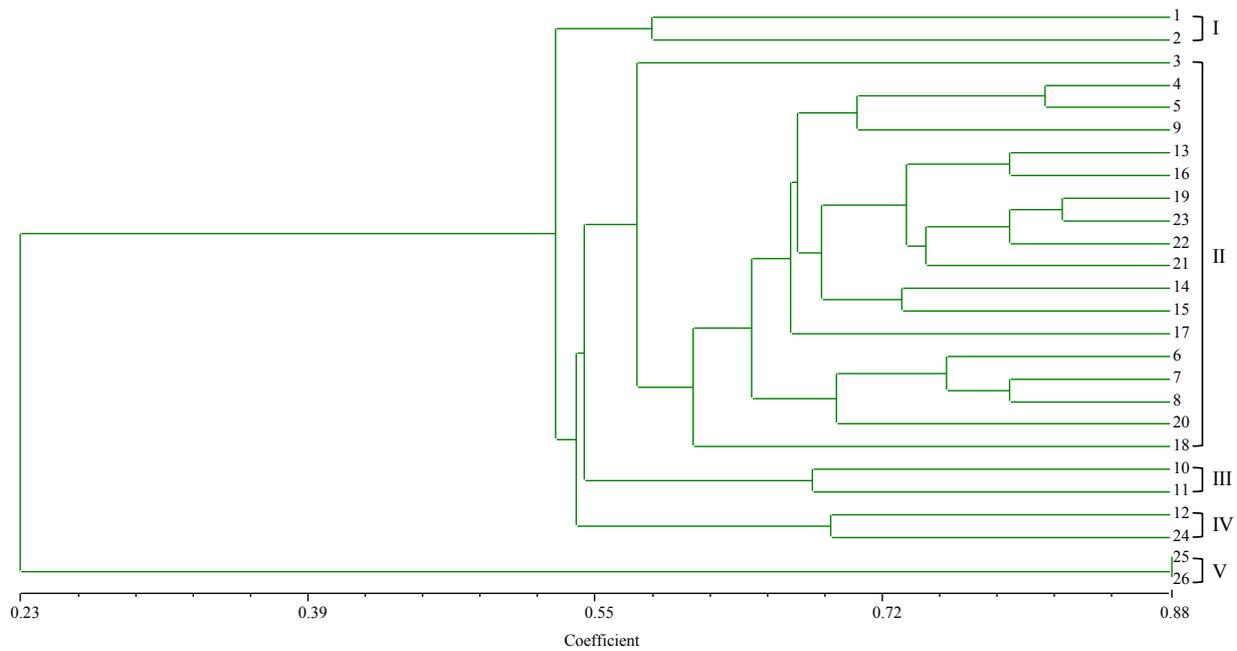


Fig. 5 UPGMA dendrogram of 26 ramie varieties based on genomic-SSR marker polymorphism. The 1 scale is genetic similarity coefficient of 0.014.

AT/TA repeats (53%). Trinucleotide repeat motifs mainly contained two A or two T nucleotides (80.27%). This indicated that mono-, di-, and trinucleotide repeat motifs were rich in A and T, consistent with the results obtained by Tóth et al. [27], who had analyzed eukaryotic genome SSR sites. The likely reason is that methylation of C bases had converted them to T [37]. The base composition of SSR loci showed that the most common type of trinucleotide repeat motifs were CCG/CGG in monocots such as rice (*Oryza sativa* L.), maize (*Zea mays* L.), and wheat (*Triticum aestivum* L.). However, the content of CCG/CGG in trinucleotide repeat motifs was less in dicots such as Arabidopsis [*Arabidopsis thaliana* (L.) Heynh.] and soybean [*Glycine max* (L.) Merrill.] [38]. These results may be explained by the higher GC content of trinucleotide repeat motifs in monocots [39] and codon usage bias between the different species [40].

Comparison of genomic-SSR and EST-SSR in ramie

The density of genomic-SSRs and EST-SSRs are about 160.4 SSR/Mb (not containing mononucleotide) and 51.8 SSR/Mb [17] in ramie, respectively. The SSR density detected from the genome is higher than that from ESTs, consistent with previous studies. For example, 85.4% of the SSRs were located in intergenic regions in *Populus*, 10.7% in introns, and only 2.7% and 1.2% located in exons and UTRs, respectively [41]. The SSR density of the cucumber (*Cucumis sativus* L.) genome (551.9 SSR/Mb) is also higher than that of its ESTs (458.7 SSR/Mb) [42]. For ramie, mononucleotides, dinucleotides, and trinucleotides are the most abundant types of repeat sequences both in genomic-SSRs and EST-SSRs. A large number of short repeated motifs were detected in genomic and EST sequences of ramie. The primary repeating motifs of genomic-SSRs were A/T, AT/TA, AG/TC, CT/GA, and AAT/TTA, while those of EST-SSR were AG/CT, AAG/CTT, AT/TA and AAC/GTT [17]. TA/AT and AG/CT were the most common dinucleotide motifs of genomic-SSRs and EST-SSRs, respectively, while TTA/AAT and AAG/CTT were the most common trinucleotide motifs of genomic-SSRs and EST-SSRs, respectively. These results were the same as those in some reports of dicots [42]. AG/CT was the most common motifs of dinucleotide in EST-SSR, while it also had a high position in genomic-SSR. AAG/CTT was the most common motif of trinucleotides in EST-SSR, while it ranked fourth among 30 kinds of trinucleotide motifs in genomic-SSR. The likely reason was that most EST-SSRs

were also included in the genomic-SSR. The differences in repeat motifs between genomic-SSR and EST-SSR may reveal the characteristics of SSRs in non-coding sequences.

A method of developing ramie genomic-SSR molecular markers

Traditionally, development of SSR molecular marker was based on library filtering and magnetic bead enrichment. This was a time-consuming and expensive process. In addition, limited SSR molecular markers could be developed by traditional methods, and their repeated units were mainly two or three nucleotides, which restricted the application of SSR markers. Until now, fewer than 2000 SSRs were developed in ramie, mainly based on magnetic bead enrichment and an EST library. Previously, there were only 18 pairs of published ramie genomic-SSR molecular markers. With the development of next-generation sequencing technology and bioinformatics, molecular markers can be developed from short sequencing reads at high-throughput, with low cost, and with a lower sequencing error rate. In this study, SLAF-seq technology was used to develop SSR molecular markers for the first time in ramie. In previous research, SLAF-seq technology was used to discover SNP markers in some crops [43,44]. This study expanded the application range for SLAF-seq technology and provided a novel method for developing SSR markers in ramie. Although there were some shortcomings, such as difficult primer design due to short sequences, SLAF-seq provides a faster and higher-throughput method to develop SSR molecular markers.

Application of genomic-SSR

Numerous studies related to SSR effects illustrated their functional importance in gene expression, when SSRs are distributed throughout the genome in both coding and non-coding regions [39,40,45]. The present usage of SSR focuses on the following aspects: (i) functional genetic analysis of important agronomic traits, (ii) QTL fine mapping, (iii) genetic structure and differentiation of population studies, and (iv) fast and accurate detection of genetic linkage. Furthermore, using SSRs as molecular markers could contribute to candidate gene prediction, identification, annotation and functional analysis. In the current study, identification cards and genetic relationships of 26 ramie germplasms were constructed and analyzed based on 31 pairs of genome SSR molecular markers. The results showed that 10 primers amplified 17 unique bands, and using five primers could differentiate the 26 ramie germplasms. This indicated that genomic-SSR had high specificity and a strong ability to identify germplasms. Furthermore, we compared the current data with the previous data of EST-SSR (19 germplasms, 31 pairs of genomic-SSR primers, and 27 pairs of EST-SSR primers). The results showed that Shannon information index values were higher in genomic-SSRs than in EST-SSRs. It indicated that genomic-SSRs had higher genetic polymorphism in comparison with EST-SSRs, and were more suitable for genetic analysis.

References

1. Xiong HP. Ma lei zuo wu yu zhong xue [Bast-fiber crops breeding]. Beijing: China Agricultural Science and Technology Press; 2008.
2. Luan MB, Zou ZZ, Zhu JJ. Development of a core collection for ramie by heuristic search based on SSR markers. *Biotechnol Biotechnol Equip.* 2014;28:798–804. <http://dx.doi.org/10.1080/13102818.2014.953768>
3. Zhu TT, Yu CM, Wang YZ. Preliminary evaluation on nutritional value for ramie [*Boehmeria nivea* (L.) Gaud.] *Zhongzhu No. 1 and Zhongzhu No. 2. Plant Fiber Science in China.* 2014;36(3):113–121.

4. Kaur S, Panesar PS, Bera MB, Kaur V. Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. *Crit Rev Food Sci Nutr*. 2015;55:41–49. <http://dx.doi.org/10.1080/10408398.2011.646363>
5. Zhang DD, Hua YP, Wang XH, Zhao H, Shi L. A high-density genetic map identifies a novel major QTL for boron efficiency in oilseed rape (*Brassica napus* L.). *PLoS One*. 2014;9:e112089. <http://dx.doi.org/10.1371/journal.pone.0112089>
6. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. *Euphytica*. 2011;177:309–334. <http://dx.doi.org/10.1007/s10681-010-0286-9>
7. Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci*. 1996;1:215–222. [http://dx.doi.org/10.1016/S1360-1385\(96\)86898-0](http://dx.doi.org/10.1016/S1360-1385(96)86898-0)
8. Zhang XW, Ye ZW, Wang TK, Xiong HR, Yuan XL. Characterization of the global transcriptome for cotton (*Gossypium hirsutum* L.) anther and development of SSR marker. *Gene*. 2014;551:206–213. <http://dx.doi.org/10.1016/j.gene.2014.08.058>
9. Singh R, Kumar S, Kashyap PL, Srivastava AK, Mishra S. Identification and characterization of microsatellite from *Alternaria brassicicola* to assess cross-species transferability and utility as a diagnostic marker. *Mol Biotechnol*. 2014;56:1049–1059. <http://dx.doi.org/10.1007/s12033-014-9784-7>
10. Wang XL, Xu F, Li L, Zhang YP, Ding YD. Development of 39 novel polymorphic microsatellite markers for the giant mottled eel *Anguilla marmorata* and cross-amplification in other eel species. *Conserv Genet Resour*. 2014;6:865–871. <http://dx.doi.org/10.1007/s12686-014-0227-3>
11. Noakes AG, Best T, Staton ME, Koch J, Romero-Severson J. Cross amplification of 15 EST-SSR markers in the genus *Fraxinus*. *Conserv Genet Resour*. 2014;6:969–970. <http://dx.doi.org/10.1007/s12686-014-0260-2>
12. Zhang Y, Dai SL, Hong Y, Song XB. Application of genomic SSR locus polymorphisms on the identification and classification of chrysanthemum cultivars in China. *PLoS One*. 2014;9:e104856. <http://dx.doi.org/10.1371/journal.pone.0104856>
13. Lei YT, Zhao YY, Yu F, Li Y, Dou QW. Development and characterization of 53 polymorphic genomic-SSR markers in Siberian wildrye (*Elymus sibiricus* L.). *Conserv Genet Resour*. 2014;6:861–864. <http://dx.doi.org/10.1007/s12686-014-0225-5>
14. Luo R, Wu WL, Zhang Y, Li YH. SSR marker and its application to crop genetics and breeding. *Genomics Appl Biol*. 2010;29:137–143.
15. Liu G, Zhang DQ, Xie YJ, Gu ZJ, Zhang HY. Rapid screening and transferability analysis of genomic-SSR and EST-SSR primers in eucalypt. *Scientia Silvae Sinicae*. 2013;49:127–133.
16. Chen JH, Luan MB, Song SF. Isolation and characterization of EST-SSRs in the ramie. *Afr J Microbiol Res*. 2011;5:3504–3508.
17. Liu TM, Zhu SY, Fu LL. Development and characterization of 1827 expressed sequence tag-derived simple sequence repeat markers in ramie (*Boehmeria nivea* L. Gaud.). *PLoS One*. 2013;8:e60346. <http://dx.doi.org/10.1371/journal.pone.0060346>
18. Jiang YB, J YC, Zhou JL. Isolation and characterization of microsatellites from ramie [*Boehmeria nivea* (L.) Gaud.]. *Acta Agronomica Sinica*. 2007;33:158–162.
19. Guan L, Zhang Z, Wang XW, Xue HB. Evaluation and application of the SSR loci in apple genome. *Scientia Agricultura Sinica*. 2011;44:4415–4428.
20. Ibrahim C, Visam G, Jens A, Sami D, Anne F. Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing. *Mol Breed*. 2014;34:323–334. <http://dx.doi.org/10.1007/s11032-014-0036-0>
21. Sun XW, Liu DY, Zhang XF, Li WB, Liu H. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One*. 2013;8:e58700. <http://dx.doi.org/10.1371/journal.pone.0058700>
22. van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol*. 2011;12:R102. <http://dx.doi.org/10.1186/gb-2011-12-10-r102>
23. Zhang J, Wu YT, Guo WZ. Fast screening of microsatellite markers in cotton with page/silver staining. *Acta Gossypii Sinica*. 2000;12:267–269.
24. Yeh FC, Yang RC, Boyle T. POPGENE version 1.31. Quick user guide. Edmonton, AB: University of Alberta and Center for International Forestry Research; 1999.

25. Rohlf FJ. NTSYSpc. Numerical Taxonomy and Multivariate Analysis System. Version 2.1. New York, NY: Applied Biostatistics Inc., Exeter Software; 2000.
26. Cai B, Li CH, Yao QH, Zhou J, Tao JM, Zhang Z. Analysis of SSRs in grape genome and development of SSR database. *Journal of Nanjing Agricultural University*. 2009;32:28–32.
27. Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res*. 2000;10:967–981. <http://dx.doi.org/10.1101/gr.10.7.967>
28. Li CY, Li JB, Zhou XG, Dong AR, Xu MH. Frequency and distribution of microsatellites in the whole genome of rice blast fungus, *Magnaporthe grisea*. *Chinese Journal of Rice Science*. 2004;18:269–273.
29. Haraoglu H, Lee CM, Meyer W. Survey of simple sequence repeats in complete fungal genomes. *Mol Biol Evol*. 2005;22:639–649. <http://dx.doi.org/10.1093/molbev/msi057>
30. Lu JB, Li JQ, Lu J, Zhan QW. Design of SSR primers and verification of e-PCR in non-coding regions of sorghum genome. *Seed*. 2010;29:1–6.
31. Gao YM, Han YQ, Tang H, Sun DM, Wang YJ, Wang WD. Analysis of simple sequence repeats in rhizobium genome. *Scientia Agricultura Sinica*. 2008;41:2992–2998.
32. Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol*. 1997;17:2851–2858. <http://dx.doi.org/10.1128/MCB.17.5.2851>
33. Harry B, Schlötterer C. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide under representation. *Genetics*. 2000;155:1213–1220.
34. Xu Y, Chen JH, Li Y. Development of EST-SSR and genomic-SSR in Chinese fir. *Journal of Nanjing Forestry University*. 2014;38:9–14.
35. Zhang H, Wang XM, Wang DJ. Survey of SSRs in foxtail millet genome and development of SSR markers. *Mol Plant Breed*. 2013;11:30–36.
36. Braulio J, Rodrigo AC, Gabriela AA. Identifying novel polymorphic microsatellites from cultivated flax (*Linum usitatissimum* L.) following data mining. *Plant Mol Biol Report*. 2011;29:753–759. <http://dx.doi.org/10.1007/s11105-010-0270-5>
37. Schorderet DF, Gartler SM. Analysis of CpG suppression in methylated and nonmethylated species. *Proc Natl Acad Sci USA*. 1992;89:957–961. <http://dx.doi.org/10.1073/pnas.89.3.957>
38. Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with non-repetitive DNA in plant genome. *Nat Genet*. 2002;30:194–200. <http://dx.doi.org/10.1038/ng822>
39. Lee SB, Kaitanis C, Jansen RK, Hosterler JB, Tallon LJ, Town CD, et al. The complete chloroplast genome sequence of *Gossypium hisutum*: organization and phylogenetic relationships to other angiosperms. *BMC Genomics*. 2006;7:61–72. <http://dx.doi.org/10.1186/1471-2164-7-61>
40. Shang MZ, Liu F, Hua JP, Wang KB. Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*. *Scientia Agricultura Sinica*. 2011;44:245–253.
41. Yin TM, Zhang XY, Gunter LE. Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the *nisqually-1* genome. *New Phytol*. 2009;181:498–503. <http://dx.doi.org/10.1111/j.1469-8137.2008.02663.x>
42. Cavagnaro PF, Senalik DA, Yang LM. Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics*. 2010;11:569. <http://dx.doi.org/10.1186/1471-2164-11-569>
43. Qi Z, Huang L, Zhu R, Xin D, Liu C. A high-density genetic map for soybean based on specific length amplified fragment sequencing. *PLoS One*. 2014;9:e104871. <http://dx.doi.org/10.1371/journal.pone.0104871>
44. Li B, Tian L, Zhang JY. Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max*. *BMC Genomics*. 2014;15:1086. <http://dx.doi.org/10.1186/1471-2164-15-1086>
45. Sangwan I, Obrian MR. Identification of a soybean protein that interacts with GAGA element dinucleotide repeat DNA. *Plant Physiol*. 2002;129:1788–1794. <http://dx.doi.org/10.1104/pp.002618>