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
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-μL reaction volumes with 1× PCR buffer, 0.2 mM dNTP, 1 U Taq DNA polymerase (Tiangen), 0.5 μM each primer and 0.5 μL 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
Results

SLAF-seq library construction

By Illumina sequencing, we obtained 3,937,489 reads (total 3,976,863 bp, average 80 bp) from Zhongzhuyihao and 3,604,166 reads (total 3,640,020 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.

Characteristics of the genomic-SSR loci

From 115,369 sequences of the SLAF library, 4,774 sequences containing SSR motifs (4.14%) were identified. The average distance between two SSRs was 3.65 kb. Of the 4,774 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).

Tab. 2 SLAF markers for both samples.

<table>
<thead>
<tr>
<th>Category</th>
<th>SLAF markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhongzhuyihao</td>
<td>81,665</td>
</tr>
<tr>
<td>Hejiangqingma</td>
<td>91,465</td>
</tr>
<tr>
<td>Consistent for both samples</td>
<td>57,761</td>
</tr>
<tr>
<td>Total for both samples</td>
<td>115,369</td>
</tr>
</tbody>
</table>

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.
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 41.43% 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/TTTA and ATTT/TAAA 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.

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.
<table>
<thead>
<tr>
<th>Locus</th>
<th>Repeat</th>
<th>Primer sequences (5′-3′)</th>
<th>Size (bp)</th>
<th>$T_a$ (°C)</th>
<th>$N_A$</th>
<th>$H_o$</th>
<th>$H_e$</th>
<th>SII</th>
<th>HWE $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibg1-11</td>
<td>(T)16</td>
<td>F: GGCTCTTAATTGCAGTC  R: TAAATCTCATAGGGCGAC</td>
<td>121</td>
<td>57</td>
<td>5</td>
<td>0.54</td>
<td>0.53</td>
<td>1.04</td>
<td>0.22</td>
</tr>
<tr>
<td>Ibg1-16</td>
<td>(T)18</td>
<td>F: TAAGGCAGACAAATTAC  R: ACCAAATCAAGAGACTC</td>
<td>126</td>
<td>55</td>
<td>3</td>
<td>0.40</td>
<td>0.66</td>
<td>0.68</td>
<td>0.09</td>
</tr>
<tr>
<td>Ibg1-20</td>
<td>(T)16</td>
<td>F: TGGCAACATAAAGCAATG  R: ACTGGAATGCATTATCTC</td>
<td>118</td>
<td>54</td>
<td>4</td>
<td>0.44</td>
<td>0.63</td>
<td>1.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Ibg1-27</td>
<td>(T)15</td>
<td>F: TTGGTTCTATATGAAATGG  R: ACCGIGATAGGTATAAC</td>
<td>115</td>
<td>54</td>
<td>3</td>
<td>0.55</td>
<td>0.45</td>
<td>0.78</td>
<td>0.27</td>
</tr>
<tr>
<td>Ibg1-31</td>
<td>(A)15</td>
<td>F: AGAATTTAAAGCATGAGAC  R: ACATAGTGACCTGAGAG</td>
<td>110</td>
<td>54</td>
<td>2</td>
<td>0.65</td>
<td>0.45</td>
<td>0.63</td>
<td>0.02</td>
</tr>
<tr>
<td>Ibg2-207</td>
<td>(AT)9</td>
<td>F: TTAATTACCTCTTTTACG  R: GCITCTCTCITATAGTTT</td>
<td>123</td>
<td>53</td>
<td>6</td>
<td>0.43</td>
<td>0.71</td>
<td>1.41</td>
<td>0.00</td>
</tr>
<tr>
<td>Ibg2-210</td>
<td>(TA)8</td>
<td>F: ACITTTGTTAGACTTTTTATAC  R: TTCTCCACCTTAATTCC</td>
<td>136</td>
<td>55</td>
<td>7</td>
<td>0.67</td>
<td>0.69</td>
<td>1.36</td>
<td>0.00</td>
</tr>
<tr>
<td>Ibg2-211</td>
<td>(AT)13</td>
<td>F: ACATAGGAAGATAATGCTAACC  R: GGGAAGATATAACCTTCTTTG</td>
<td>146</td>
<td>55</td>
<td>6</td>
<td>0.70</td>
<td>0.83</td>
<td>1.66</td>
<td>0.00</td>
</tr>
<tr>
<td>Ibg2-213</td>
<td>(TA)10</td>
<td>F: TTAGTCAACTCTGTGCTATG  R: GCATCTATCTACACATTAC</td>
<td>108</td>
<td>54</td>
<td>4</td>
<td>0.96</td>
<td>0.68</td>
<td>1.20</td>
<td>0.00</td>
</tr>
<tr>
<td>Ibg2-214</td>
<td>(AC)9</td>
<td>F: AGGTTAGGTTAACAGGAGTTG  R: GTAGCCACATACATAATG</td>
<td>110</td>
<td>55</td>
<td>4</td>
<td>0.95</td>
<td>0.57</td>
<td>0.94</td>
<td>0.00</td>
</tr>
<tr>
<td>Ibg2-218</td>
<td>(TC)17</td>
<td>F: AATACCGAGAGGAGCAAG  R: AAACAAAACCAACTGACAAAC</td>
<td>163</td>
<td>55</td>
<td>3</td>
<td>0.78</td>
<td>0.60</td>
<td>0.98</td>
<td>0.10</td>
</tr>
<tr>
<td>Ibg3-206</td>
<td>(TTA)6</td>
<td>F: GGAAGGGCTAATACCTAC  R: AAGATGGTTAGCTACTTC</td>
<td>102</td>
<td>54</td>
<td>2</td>
<td>0.05</td>
<td>0.05</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Ibg3-209</td>
<td>(TAA)5</td>
<td>F: AGAAATAAGTGCATCGTC  R: GCTGTCACTAGGATATAAG</td>
<td>114</td>
<td>54</td>
<td>6</td>
<td>0.57</td>
<td>0.83</td>
<td>0.96</td>
<td>0.00</td>
</tr>
<tr>
<td>Locus</td>
<td>Repeat</td>
<td>Primer sequences (5’–3’)</td>
<td>Size (bp)</td>
<td>T&lt;sub&gt;s&lt;/sub&gt; (°C)</td>
<td>N&lt;sub&gt;A&lt;/sub&gt;</td>
<td>H&lt;sub&gt;o&lt;/sub&gt;</td>
<td>H&lt;sub&gt;e&lt;/sub&gt;</td>
<td>SII</td>
<td>HWE p-value</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>--------------------------</td>
<td>-----------</td>
<td>------------------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td>----</td>
<td>-------------</td>
</tr>
<tr>
<td>Ibg3-210</td>
<td>(TTA)5</td>
<td>F: AACAAGATTTTGTCCITTTG&lt;br&gt;R: TTTATGCACATACACTCTTG</td>
<td>139</td>
<td>54</td>
<td>4</td>
<td>0.57</td>
<td>0.59</td>
<td>1.10</td>
<td>0.35</td>
</tr>
<tr>
<td>Ibg3-212</td>
<td>(ATT)5</td>
<td>F: TAGTCCATAGCAGTAAATCC&lt;br&gt;R: TAGCCGTCTTGAAGAGAG</td>
<td>145</td>
<td>56</td>
<td>4</td>
<td>0.57</td>
<td>0.51</td>
<td>0.95</td>
<td>0.58</td>
</tr>
<tr>
<td>Ibg3-214</td>
<td>(AAT)5</td>
<td>F: ATCACTCOCAGAATTTC&lt;br&gt;R: GTAGCCACATACATAATG</td>
<td>148</td>
<td>55</td>
<td>4</td>
<td>0.62</td>
<td>0.53</td>
<td>0.98</td>
<td>0.67</td>
</tr>
<tr>
<td>Ibg3-216</td>
<td>(TTA)5</td>
<td>F: TGGAGGTGAAGATAATG&lt;br&gt;R: ACACAATAACATACACATC</td>
<td>156</td>
<td>55</td>
<td>2</td>
<td>0.04</td>
<td>0.04</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Ibg3-218</td>
<td>(ATA)5</td>
<td>F: ATGTATATGTAAACTAACAG&lt;br&gt;R: GATTGATATACATACACAT</td>
<td>93</td>
<td>54</td>
<td>5</td>
<td>0.36</td>
<td>0.33</td>
<td>0.70</td>
<td>1.00</td>
</tr>
<tr>
<td>Ibg4-41</td>
<td>(TTAT)5</td>
<td>F: AATAAGCTCAATGAGAAC&lt;br&gt;R: GATGTAAGTAGAATTCG</td>
<td>143</td>
<td>53</td>
<td>5</td>
<td>0.47</td>
<td>0.69</td>
<td>1.24</td>
<td>0.28</td>
</tr>
<tr>
<td>Ibg4-49</td>
<td>(TATT)6</td>
<td>F: GCTATTACCTTTGAACCTT&lt;br&gt;R: ATATATGCTTTGTTTAG</td>
<td>128</td>
<td>53</td>
<td>4</td>
<td>0.33</td>
<td>0.62</td>
<td>1.07</td>
<td>0.04</td>
</tr>
<tr>
<td>Ibg4-50</td>
<td>(ATT)5</td>
<td>F: ACCAGGAGTATCATITG&lt;br&gt;R: TATCGATGTCGATGATGT</td>
<td>151</td>
<td>57</td>
<td>3</td>
<td>1.00</td>
<td>0.53</td>
<td>0.79</td>
<td>0.00</td>
</tr>
<tr>
<td>Ibg4-52</td>
<td>(ATAA)5</td>
<td>F: ACTAATCATTACATATATG&lt;br&gt;R: TTTATTGCTGTTGCTTGAG</td>
<td>145</td>
<td>54</td>
<td>2</td>
<td>0.04</td>
<td>0.04</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Ibg4-53</td>
<td>(ATT)5</td>
<td>F: TATATGTTTTGATTACCTAAC&lt;br&gt;R: ATGAAATGAACCAACAAC</td>
<td>141</td>
<td>55</td>
<td>2</td>
<td>0.36</td>
<td>0.47</td>
<td>0.66</td>
<td>0.27</td>
</tr>
<tr>
<td>Ibg4-57</td>
<td>(TTTA)5</td>
<td>F: AGATCGTTCACTACAAC&lt;br&gt;R: TTACACACCACCATAG</td>
<td>115</td>
<td>55</td>
<td>4</td>
<td>0.50</td>
<td>0.50</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>Ibg5-1</td>
<td>(CATGT)5</td>
<td>F: ACAAACAAACACATACAC&lt;br&gt;R: GCATAGAATGATAGAGAG</td>
<td>132</td>
<td>55</td>
<td>3</td>
<td>0.47</td>
<td>0.51</td>
<td>1.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Ibg5-3</td>
<td>(AAAAT)5</td>
<td>F: TTAAGTTGATTTCAAG&lt;br&gt;R: ATTATGTTGCTC</td>
<td>144</td>
<td>55</td>
<td>3</td>
<td>0.47</td>
<td>0.44</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td>Ibg5-5</td>
<td>(AAACC)6</td>
<td>F: CATTCTCCTTTTCA&lt;br&gt;R: GCATTGTATTGTTTAGG</td>
<td>135</td>
<td>55</td>
<td>3</td>
<td>0.76</td>
<td>0.53</td>
<td>1.33</td>
<td>0.11</td>
</tr>
</tbody>
</table>
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 the phytoes 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

<table>
<thead>
<tr>
<th>Locus</th>
<th>Repeat</th>
<th>Primer sequences (5’–3’)</th>
<th>Size (bp)</th>
<th>Ta (°C)</th>
<th>HWE p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibg5-6</td>
<td>(TTAT)5</td>
<td>F: TATAAATAGTGTAACTTCCAC R: TGTTTTCTAATTTTAGTTCAC</td>
<td>108</td>
<td>54</td>
<td>0.04</td>
</tr>
<tr>
<td>Ibg6-468</td>
<td>(A)17</td>
<td>F: CTAATCATAATAATTTGTTCAC R: TATAGGCTCTCAATAATTG</td>
<td>96</td>
<td>53</td>
<td>1.00</td>
</tr>
<tr>
<td>Ibg6-477</td>
<td>(CA)8</td>
<td>F: TCCTAAATTACTAATACACTA R: TCCTAAGACTTCTTTTCC</td>
<td>109</td>
<td>55</td>
<td>1.00</td>
</tr>
<tr>
<td>Ibg6-481</td>
<td>(A)22 (A)16</td>
<td>F: CAAATGCTGAAATAAAAATAAG R: TTGTGATAACTGTCCTTC</td>
<td>124</td>
<td>54</td>
<td>1.00</td>
</tr>
</tbody>
</table>
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

![Fig. 5](image_url) UPGMA dendrogram of 26 ramie varieties based on genomic-SSR marker polymorphism. The 1 scale is genetic similarity coefficient of 0.014.
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 markers 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


