Assessment of genetic diversity among basmati and non-basmati aromatic rices of India using SSR markers

A. K. P. Sivaranjani, Manish K. Pandey, I. Sudharshan, G. Ram Kumar, M. Sheshu Madhav, R. M. Sundaram, G. S. Varapradas and N. Shobha Rani*

Crop Improvement Section, Directorate of Rice Research, Rajendranagar, Hyderabad 500 030, India

Understanding the genetic diversity among aromatic rice local land races would be useful for their effective utilization in breeding programmes aimed at rice grain quality improvement. The study was conducted to analyse the genetic diversity among 16 aromatic short grain local land races collected from various parts of India and 30 basmati local land races collected from the traditional basmati growing areas, using 26 hyper-variable simple sequence repeat markers spanning the rice genome. Both the aromatic groups could be discriminated by either single (RM28102) or combination of two (RM577 + RM30) markers. Based on the similarity co-efficient values, genotypes were classified into two major clusters with 70% dissimilarity revealing presence of high diversity. The information generated from the present study would be useful in genetic purity assessment, varietal identification for plant variety protection, mapping of genes and background selection for recipient genome recovery in marker-aided breeding programmes aimed at aromatic rice improvement.

Keywords: Basmati rice, molecular marker diversity, SSR markers, varietal identification.

Rice is one of the most important cereal crops being cultivated for over 10,000 years in the world and it is the main source of energy for more than half of the world population. In India, rice is a major food crop supplying 30% of the calorie requirement to the Indian population and a major portion of the rural diet is exclusively rice based. Aromatic rices are a specialty with unique cooking qualities and possess a distinct aroma. India has rich genetic diversity with respect to aromatic rices. Among the aromatic rices, basmati rice occupies a special position in the world rice markets and has become increasingly popular in the Middle East, USA, Europe and even in non-traditional rice growing countries such as Australia. High quality traditional basmati rice varieties command premium price, which is generally three times higher than that of non-basmati rices in the Indian as well as international markets. Basmati rice originated in the foothills of Himalayas and spread to north west Indian states of the Indian subcontinent. Hundreds of locally adapted land races of basmati and indigenous non-basmati aromatic short grain rices are grown in the Indian subcontinent, which have evolved as a result of both natural and human selection. Some of these indigenous short grain rices surpass the long grain basmati with respect to many grain and cooking quality traits. Though local aromatic rices have a limited market base, their prices are often equal to or even higher than basmati in the native areas of cultivation where farmers and consumers know their qualities. They are also adapted to local conditions in areas where basmati may not be grown and have very strong aroma under the prevailing warmer climate during grain maturity period.

A greater appreciation of the genetic diversity contained in the aromatic rice gene pool is necessary for classification, proper maintenance, conservation of these prized rice varieties and their effective utilization in the breeding programmes. Although rice germplasm characterization and diversity analysis has been done by several workers, limited information is available on the aromatic rices, particularly land races of basmati and non-basmati types. Such studies have especially become essential for initiating genetic improvement efforts of short grain aromatic rices, as meagre attention has been paid for their improvement till now and there is an imminent threat of this category of rices being replaced by high-yielding semi-dwarf varieties leading to loss of such a valuable bioresource.

Assessment of genetic diversity is an integral part of plant breeding as exploitation of the genetic diversity helps the plant breeder to develop new varieties. Selection of diverse parents with desirable qualities can be done based on analysis using morphological, biochemical and/or DNA markers. Though morphological characterization is an important methodology for studying diversity, it is cumbersome, whereas there are limitations in using biochemical markers as they are limited in number and are predisposed by environmental factors or the developmental stage of the plant. DNA markers are the most widely used and are predominant due to their abundance and repeatability; they remain unaffected across different stages, seasons, locations and agronomic practices. Of the several classes of DNA markers available, microsatellite or simple sequence repeat (SSR) markers are considered as most amenable for several applications including genetic diversity studies due to their multiallelic nature, high reproducibility, co-dominant inheritance, abundance and extensive genome coverage. A carefully chosen set of SSR markers providing genome-wide coverage will facilitate an unbiased assay of genetic diversity, thus giving a robust, unambiguous molecular description of rice cultivars. SSR markers have been effectively used to study genetic diversity among closely related rice cultivars and distantly related genotypes. However, only a few efforts have been made to

*For correspondence. (e-mail: n_shobha@yahoo.com)
characterize aromatic rice varieties including basmati varieties\textsuperscript{19–24}. Further an array of land races of basmati and short grain rices have not been much addressed in terms of molecular characterization. Screening of these local land races with microsatellite markers would create a valuable database for varietal identification and will serve as a resource for marker assisted selection\textsuperscript{25}. Hence, the present study was carried out with objectives to analyse the structure of genetic diversity of Indian aromatic rice gene pool using SSR markers and identify marker(s), which can be used to establish the varietal identity of the elite aromatic rice genotypes.

The 46 aromatic accessions selected for the study consisted of 16 aromatic short grain local land races (ASGLLR) and 30 basmati local land races (BLLR) collected from different states of India and from the traditional basmati growing areas respectively (Table 1). The seeds of the rice varieties were soaked in petri plates on moistened blotting paper at room temperature for germination for about 10 days and then the leaf samples were collected from 7 to 10-day-old seedlings for genomic DNA extraction and molecular marker analysis.

Twenty six highly polymorphic SSR markers covering all 12 chromosomes were used to study molecular diversity among the 46 aromatic genotypes (Table 2). Total genomic DNA was extracted from leaf samples using the procedure described in ref. 26 and used for polymerase chain reaction (PCR) analysis using selected SSR markers.

### Table 1. Local land races of basmati and aromatic short grain rices used in the present study

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Number</th>
<th>Land races</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aromatic short grain local land</td>
<td>16</td>
<td>Tanrubhug, Amrubhug, Shyamjira, Badshaha, RARU-3014, Ganjeikaili, Neelabati, Kalajeera, Bishrubhug, Lectimachi, Sheeltalki, Tulaighanti, Tulasikanthi, RB-2816, Kanakjeer-B, Kanakjeer</td>
</tr>
</tbody>
</table>

### Table 2. Selected markers from different chromosomes and their PIC values

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chromosome</th>
<th>No. of alleles</th>
<th>Polymorphic alleles</th>
<th>Physical position (bp)</th>
<th>Polymorphic information</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM 14</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>48,525,000</td>
<td>0.61</td>
</tr>
<tr>
<td>RM 577</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>8,375,101</td>
<td>0.58</td>
</tr>
<tr>
<td>RM 166</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>34,346,727</td>
<td>0.6</td>
</tr>
<tr>
<td>RM 6</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>29,579,845</td>
<td>0.72</td>
</tr>
<tr>
<td>RM 251</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>19,775,000</td>
<td>0.60</td>
</tr>
<tr>
<td>RM 282</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>12,387,497</td>
<td>0.38</td>
</tr>
<tr>
<td>RM 16294</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>911,630</td>
<td>0.69</td>
</tr>
<tr>
<td>RM16303</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1,097,445</td>
<td>0.6</td>
</tr>
<tr>
<td>RM 16913</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>9,631,036</td>
<td>0.65</td>
</tr>
<tr>
<td>RM 17630</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4,122</td>
<td>0.44</td>
</tr>
<tr>
<td>RM 13</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>7,150,000</td>
<td>0.49</td>
</tr>
<tr>
<td>RM 289</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>7,787,118</td>
<td>0.57</td>
</tr>
<tr>
<td>RM 3</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>26,869,497</td>
<td>0.47</td>
</tr>
<tr>
<td>RM 30</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>27,252,300</td>
<td>0.51</td>
</tr>
<tr>
<td>RM 533</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>17,511,283</td>
<td>0.54</td>
</tr>
<tr>
<td>RM 478</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>25,948,820</td>
<td>0.48</td>
</tr>
<tr>
<td>RM 42</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>19,964,404</td>
<td>0.17</td>
</tr>
<tr>
<td>RM 38</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2,109,496</td>
<td>0.66</td>
</tr>
<tr>
<td>RM 24017</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>957,256</td>
<td>0.49</td>
</tr>
<tr>
<td>RM 24654</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>20,197,402</td>
<td>0.57</td>
</tr>
<tr>
<td>RM 258</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>17,756,102</td>
<td>0.42</td>
</tr>
<tr>
<td>RM 224</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>3,025,000</td>
<td>0.72</td>
</tr>
<tr>
<td>RM 206</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>21,797,485</td>
<td>0.45</td>
</tr>
<tr>
<td>RM 287</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>16,730,739</td>
<td>0.62</td>
</tr>
<tr>
<td>RM 28102</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>15,907,555</td>
<td>0.79</td>
</tr>
<tr>
<td>RM 27507</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>45,523</td>
<td>0.31</td>
</tr>
</tbody>
</table>
(Table 2). The PCR mixture contained 50 ng template DNA, 5 pmoles of each primer (Integrated DNA Technologies, USA), 0.05 mM dNTPs (MBI Fermentas, Lithuania), 1× PCR buffer (10 mM Tris, pH 8.4, 50 mM KCl, 1.8 mM MgCl2; Sanmar Fine Chemicals, India) and 1 U of Taq DNA polymerase (Sanmar Fine Chemicals, India) in a reaction volume of 10 μl. Template DNA was initially denatured at 94°C for 5 min followed by 35 cycles of PCR amplification with the following parameters: a 30 s denaturation at 94°C, a 30 s annealing at 55°C and 1–2 min of primer extension at 72°C. A final extension was done at 72°C for 7 min. PCR amplified products were electrophoretically resolved in 3% MetaPhor®/Agarose gels (Lonza, USA) in 0.5× TBE buffer at 100 V for 3.5 h in Hoeffer Super Submarine Electrophoresis unit (GE Biosciences, USA). The gels were stained in ethidium bromide (10 mg/ml) and placed over a UV-Trans illuminator and documented in an Alpha Imager gel documentation system (Alpha Innotech, USA).

Qualitative multistage traits that depict an array of characters were converted into binary characters based on the variations present. Only the clear and unambiguous bands of SSR markers were scored. Markers were scored for the presence and absence of the corresponding band among the genotypes. The score 1 and 0 indicates the presence and absence of the bands respectively. A data matrix comprising 1 and 0 was formed depending upon the character and this data matrix was subjected to further analysis.

To measure the informativeness of the markers, the polymorphism information content (PIC) for each SSR marker was calculated according to the formula:

\[
\text{PIC} = 1 - \sum p_i^2 - \sum p_i^2 p_j^2, 
\]

where \(i\) is the total number of alleles detected for a SSR marker and \(p_i\) the frequency of the \(i\)th allele in the set of 46 genotypes investigated and \(j = i + 1\). This formula gives us an indication of how many alleles a certain marker has, and how much these alleles divide evenly. The PIC value was calculated using the online software Polymorphism Information Content Calculator available at [www.agri.hujj.ac](http://www.agri.hujj.ac).

The binary data matrix was subjected to cluster analysis. Sequential agglomerative hierarchical non-overslapping (SAHN) clustering was performed on squared Euclidean distance matrix and similarity matrix using Jacqard’s coefficient utilizing the unweighted pair group method with arithmetic averages (UPGMA) method. Data analysis was done using the software NTSYSpc version 2.02 (ref. 27).

All the markers considered for the present study were observed to be polymorphic, amplifying a total of 61 allelic fragments. Seven markers (RM6, RM166, RM251, RM24654, RM287, RM28102 and RM224) amplified 21 alleles (three bands each). The PIC values of the markers ranged from 0.170 (RM42) to 0.729 (RM28102) with an average of 0.53 per marker (Table 2). Similar PIC values have been reported in earlier studies19–22,24 while studying genetic diversity among aromatic rice genotypes. However, higher average PIC values of 0.62–0.73 have also been reported11,12,15,17,26–30. The higher values reported in these studies may be due to use of genotypes with wide origin which are distantly related, whereas the present study was carried out among closely related aromatic indica genotypes.

One of the most important applications of molecular diversity studies is to identify a marker which can differentiate a genotype from the remaining genotypes but this is difficult to achieve in a closely related set of genotypes. Interestingly, most of the genotypes considered in the present study could be differentiated from each other by using a single marker or marker combinations. RM27507 located on chromosome 12 differentiated ASGLLRs from most of the BLRRs (except Basmati 334, HBC 30, HBC 85, Taranbhog, Amrutbhog, Shyamjera and Badshaha; Figure 1). A combination of two markers (RM577, RM30) located on chromosomes 1 and 6 also clearly differentiated both the groups. Nine genotypes, viz. Basmati Mehtra, Basmati 427, Shyamjira, Basmati Bahar, Basmati 62, Basmati Surkh 89-15, Basmati 334, Amrutbhog and RB2816 can be differentiated with the markers RM251, RM42, RM42, RM3, RM14, RM287, RM577, RM166 and RM24654 respectively (Figure 2). Seventeen genotypes can be distinguished by using a combination of two markers (Table 3). These genotypes include Basmati Sathi, Basmati 443 (using the marker combination RM17630 + RM 3), Basmati Kamon (RM6 + RM478), Basmati 122 (RM27507 + RM478), Basmati Sufaid (RM206 + RM258), HBC30 (RM30 + RM258), HBC85 (RM16303 + RM289), Basmati Jamuna (RM16303 + RM30), Badshaha (RM16913 + RM224), RAU3014 (RM577 + RM3), Ganjeikalli (RM16294 + RM24564), Neelabati (RM166 + RM6), Bishnubhog (RM166 + RM251), Kanakjeer-B (RM577 + RM3) and Kanakjeer (RM289 + RM38). More than 50% of genotypes used in the study were successfully differentiated from the rest of
Figure 2. Dendrogram showing genetic relationship among the rice genotypes under study.

The genotypes indicating the importance and relevance of DNA fingerprinting. These genotype specific markers may be helpful in establishing varietal identity and for checking adulteration among basmati varieties\(^{19,20,22,23}\) and can also play an important role in generating additional information to supplement the morphological descriptors recommended by the International Union for the Protection of New Varieties of Plants (UPOV). In a recent study, it has been shown that SSR markers near rice genes discriminated closely related rice varieties\(^{11}\).

A cluster analysis using UPAGMA based on similarity coefficients was done to resolve the phylogenetic relationships among the different aromatic rice genotypes considered for the present study. The analysis classified aromatic rice genotypes into two major clusters (I&II) with a dissimilarity value of 70% between these two clusters (Figure 3). The first cluster consists of 27 genotypes (16 ASGLRs and 11 BLLRs) and the second cluster consisted of the remaining 19 genotypes (19 BLLRs). The first cluster was again sub-divided into five subclusters IA, IB, IC, ID, and IE consisting of six, nine, four, three and four genotypes respectively. Of these, the subcluster IA included all six BLLRS (Basmati Mehtra, Basmati 134, Basmati 443, Basmati Sathi, Basmati 397, Basmati 213) with 74% overall similarity within the subcluster and 38% dissimilarity with other subclusters. The subcluster IB consisted of nine accessions which were observed to possess 28% of similarity within the subcluster and 38% dissimilarity with other subclusters. The subcluster IC consisted of four accessions with 74% of overall similarity within the subcluster but having 38% dissimilarity with other subclusters, whereas the subcluster-ID consisted of three accessions with 72% overall similarity within the group and the subcluster IE
Table 3. Genotypes which can be identified with a single simple sequence repeat marker or with a combination of two markers

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Distinguishing marker</th>
<th>Chromosome</th>
<th>Based on allele variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basmati Mehtra</td>
<td>RM 251</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Basmati 427, Shyamjeera</td>
<td>RM 42</td>
<td>8</td>
<td>2, 1</td>
</tr>
<tr>
<td>Basmati Bahar</td>
<td>RM 3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Basmati 62</td>
<td>RM 14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Basmati Surch-89-15</td>
<td>RM 287</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Basmati 334</td>
<td>RM 577</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Amrutbhog</td>
<td>RM 166</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>RB 2816</td>
<td>RM 24654</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Basmati Sathi, Basmati 443</td>
<td>RM 17630, RM 3</td>
<td>4, 6</td>
<td>2, 1</td>
</tr>
<tr>
<td>Basmati Kamon</td>
<td>RM 6, RM 478</td>
<td>2, 7</td>
<td>1, 2</td>
</tr>
<tr>
<td>Basmati 122</td>
<td>RM 16303, RM 478</td>
<td>12, 7</td>
<td>1, 1</td>
</tr>
<tr>
<td>Basmati Sufaid</td>
<td>RM 206, RM 258</td>
<td>11, 10</td>
<td>3, 1</td>
</tr>
<tr>
<td>HBC 30</td>
<td>RM 30, RM 258</td>
<td>6, 10</td>
<td>2, 1</td>
</tr>
<tr>
<td>HBC 85</td>
<td>RM 16303, RM 289</td>
<td>4, 12</td>
<td>2, 1</td>
</tr>
<tr>
<td>Basmati Jamuna</td>
<td>RM 16303, RM 30</td>
<td>4, 6</td>
<td>2, 2</td>
</tr>
<tr>
<td>Badshaha</td>
<td>RM 16913, RM 224</td>
<td>4, 11</td>
<td>1, 1</td>
</tr>
<tr>
<td>RAU 3014</td>
<td>RM 577, RM 3</td>
<td>1, 6</td>
<td>1, 1</td>
</tr>
<tr>
<td>Ganjelaelli</td>
<td>RM 16294, RM 24654</td>
<td>4, 9</td>
<td>1, 3</td>
</tr>
<tr>
<td>Neelabati, Kalajeera</td>
<td>RM 166, RM 6</td>
<td>2, 2</td>
<td>2, 1</td>
</tr>
<tr>
<td>Bishnubhog</td>
<td>RM 166, RM 251</td>
<td>2, 3</td>
<td>5, 2</td>
</tr>
<tr>
<td>Lectimachi</td>
<td>RM 3, RM 38</td>
<td>6, 8</td>
<td>2, 1</td>
</tr>
<tr>
<td>Kanakjeer B</td>
<td>RM 577, RM 3</td>
<td>1, 6</td>
<td>1, 1</td>
</tr>
<tr>
<td>Kanakjeer</td>
<td>RM 289, RM 38</td>
<td>5, 8</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

Figure 3. SSR marker RM287 can differentiate Basmati Surch 89-15 (lane 19) from the remaining genotypes. 1, Sheetal-kali (ASGLLR); 2, RAU 3014 (ASGLLR); 3, Shyamjira (ASGLLR); 4, Kanakjeer-B (ASGLLR); 5, Lectimachi (ASGLLR); 6, Kanakjeer (ASGLLR); 7, Badshaha (ASGLLR); 8, Basmati 397 (BLLR); 9, Basmati 5828 (BLLR); 10, Basmati 106-12 (BLLR); 11, HBC 46 (BLLR); 12, HBC 85 (BLLR); 13, HBC 30 (BLLR); 14, HBC 45 (BLLR); 15, HBC 46 (BLLR); 16, Basmati 242 (BLLR); 17, Basmati 122 (BLLR); 18, Basmati 124-10 (BLLR); 19, Basmati Surch 89-15 (BLLR); 20, Basmati 62 (BLLR); 21, Basmati 443 (BLLR); 22, Basmati 6129 (BLLR); 23, Basmati 208 (BLLR), and 24, Basmati 410 (BLLR).

The moderate to high degree of dissimilarity among the aromatic accessions observed in the present study exemplifies the high level of diversity at molecular level among the set of genotypes used for the study and indicates the possibility of using these genotypes in breeding programmes targeted at improvement of elite aromatic rice varieties. The marker RM251 showed amplification of a distinct fragment (i.e. allele 3) in Basmati 6129, Basmati Kamon, Basmati Kota, whereas the marker RM38 displayed the amplification if allele 2 in HBC 45, HBC 46 and HBC 85. Basmati Kamon and Basmati Kota clustered together with a similarity coefficient of 80% and both can be differentiated by using markers RM6 and RM78. RM258 showed amplification of allele 2 in Basmati Kamon and Basmati Kota and RM6 showed the presence of allele 3 in Basmati 124-10, Tulasi and Kanakjeer-B. The marker RM27507 showed the presence of allele 2 in Badshah, Basmati Sathi, Basmati 62. Basmati 242, Bishnubhog, RB 2816 and Kanakjeer. The genotypes Basmati 62 and Basmati 242 were clustered together, with similarity co-efficient of 68%. The separation of ASGLLRs from BLLRs as inferred from molecular marker assay in the present study corresponded well with their geographic distribution as well as grain characteristics and the similarity differentiation of short and long slender genotypes has been reported earlier.

The present study revealed a moderate to high level of diversity at molecular level among the ASGLLRs and BLLRs and a limited number of markers efficiently
grouped genotypes with similar origin, morphology and quality features. A single SSR marker or a combination of two markers could efficiently differentiate all the ASG (LLRs and BLLRs) and these markers/marker combinations may be useful in breeding programmes targeted at improvement of aromatic varieties in terms of identification of diverse donor parents, purity testing and plant variety protection. The results further highlighted the potential of application of SSR-marker based assay for studying genetic diversity at molecular level among very closely related rice genotypes.


ACKNOWLEDGEMENTS. We thank the Department of Biotechnology (DBT) for financial support. We also thank Project Director, Directorate of Rice Research for his guidance.

Received 5 October 2009; accepted 15 June 2010