Genetic Diversity Assessment of Four Rice Varieties using SNP Markers

Oumarou Souleymane, Batieno Teiyoué Benoit Joseph, Baboucarr Manneh, Kwadwo Ofori, and Eric Danquah

Abstract—Markers assisted selection is a tool for improving the speed and efficiency of crop improvement. The identification of SNPs within the target genomic region of a QTL is necessary for efficient breeding of quantitative traits such as salt tolerance. The study was conducted to characterize parental lines using SNP markers and to identify the polymorphic SNPs for salt tolerance in QTL study of offspring. Four rice lines were evaluated under five levels of salt conditions including the control. The experimental design was a split-plot with two replications. Leaf samples were collected from four rice lines at heading stage and sent to LGC Genomics laboratory for genotyping using rice SNP platform. DNA extraction and SNP genotyping were performed using an internal protocol. The phenotypic data recorded were the visual scoring, tiller number, green leaf number, panicle number, panicle weight and dry weight. Phenotypic data were analyzed using SAS 9.2 software. The genotypic data were analyzed using MEGA6 and GGT2 software. Significant diversity was revealed among lines concerning all the phenotypic traits. Significant diversity, divergence and substitution pattern were observed among lines. 200 markers out of 1896 were polymorphic and selected for the next step.

Index Terms—Rice; Lines; SNP Markers; Characterization.

I. INTRODUCTION

Two different categories of genetic markers are used in genetics and plant breeding. These include classical markers and DNA markers [25]. Based on different polymorphism detecting techniques (RFLP, AFLP, RAPD, SSR, SNP, etc.), DNA markers have been developed into many systems [2]. Over the past years, progresses of molecular markers, have extensively impacted the concept of conventional rice breeding [13]. Markers assisted selection is a way to improve the efficacy and speed of plant breeding programs [17]. It offers advantages to be growth stage independent, unaffected by environment effect, and no dominance influence. According to Syvanen [20] variation in DNA sequence is largely responsible of phenotypic differences between plant individuals or varieties.

AFLP markers for ion transport and selectivity were identified by Flower et al. [4] in rice. In studies using a salt tolerant variety (Pokkali) major gene has been mapped [9]. SaIT is thought to be this major gene [2]. On chromosome 1 was found by a major QTL for three mapped traits associated with salt tolerance [9]. Koyama et al. [10], Lin et al. and Takehisa et al. [21] have mapped QTLs for salt tolerance involved in high yield in saline environments [7]. Single nucleotide polymorphism (SNP) represents the bulk of natural genetic variation in organisms. It is a difference between two individuals based on single nucleotide [3]. SNP markers have become very attractive in genetic study and breeding [8], because, they are co-dominant, linked to genes, ultimately polymorphic and could be found in both intron and exon regions [14]. Cargill et al. [1], Kwon [12], Schafer [15], and Syvanen [20] found that the large number of SNPs in the genomes of individuals provides the basis for innovative approaches to QTL genomic mapping. Hence, SNPs may be important in the expression of traits that are controlled by a combination of multiple genes. Thus, the identification of SNPs within the target genomic region of a QTL is necessary for efficient breeding of quantitative traits such as salt tolerance. For this purpose, this study was carried out with following objectives:

- Assess genetic diversity of parental lines under salt stress.
- Characterize parental lines using SNP markers.
- Identify the polymorphic SNPs for salt tolerance QTL study in offspring.

II. MATERIAL AND METHODS

A. Material for phenotyping

A greenhouse pot experiment was conducted at the Regional Agronomic Research Centre (CERRA) of Kollo in 2011 wet season to determine the response of rice varieties to different levels of saline water at different growth stages. Kollo research station is geographically located in latitude 13°19’43” N, longitude 2° 19’16” E and 250 m altitude.

The soil collected from the rice salt free field was clay with dark color and characterized (Souleymane, 2014). The soil was clayey and was composed of 52.8% clay, 30.4% silt and 16.7% sand. The experimental soil was acidic (pH=3.9), the sodium adsorption ratio (SAR) and potassium adsorption ratio (PAR) were equal and very low (0.06). The exchangeable sodium percentage was about 6%. The electrical conductivity (0.12dS/cm), organic matter percentage (2.26%) and the cation exchange capacity were also low (10.16).

The experiment design was a standard split-plot with two replicates. The treatments were five salt levels: 10 dSm-1 and 12 dS m-1 according to Shannon et al. [16] method, 10 dSm-1 and 12 dS m-1 according to Gregorio’s method [5] and control, and 4 rice lines. The salt levels were assigned as the main plot factor and the treatments of lines were assigned as sub-plot factors. All factors were considered as fixed.

The plant material used in the genotyping was composed

Published on December 16, 2017.

O. Souleymane, B. T. B. Joseph, B. Manneh, K. Ofori, and E. Danquah are with the National Institute for Agricultural Research of Niger (INRAN), Niger. (e-mail: umarsou@gmail.com).

DOI: http://dx.doi.org/10.24018/ejers.2017.2.12.504
of four parents among which we had two salt tolerant (IRRI113 and NSIC RC106) varieties imported from IRRI and two local farmer preferred varieties (Gambiaka and IR1529).

The symptoms of salt stress for each plant were recorded 42 days after salt treatment according to the standard evaluation system used at International Rice Research Institute [6]. The phenotypic data recorded were tiller number, green leaf number, panicle number, panicle weight and dry weight.

B. Material genotyping

Leaf sampling was done at heading stage. The samples were taken from the young leaves and put into 96 well plates. The plates were then sufficiently sealed to prevent sample leakage and contamination. All DNA sample plates were clearly labeled. The physical plate name must match the name given to the plate in the submitted DNA sample plate file. The samples plates were then sent to LGC Genomics laboratory for DNA extraction and SNP genotyping using an internal protocol. One thousand eight hundred ninety-six (1896) SNPs markers were used to genotype the parental lines.

C. Data analysis

SAS 9.2 software was used for statistical analysis of phenotypic data. GLM Procedure for analysis of variance and LSD method was used for means separation.

Genotypic data were analyzed using MEGA6 and GGT2 software. The test of homogeneity of substitution patterns between sequences was performed using the following procedure: The probability of rejecting the null hypothesis that sequences have evolved with the same pattern of substitution, as judged from the extent of differences in base composition biases between sequences [11]. A Monte Carlo test was used to estimate the P-values [11]. Evolutionary analyses were conducted in MEGA6 [23].

For the maximum likelihood Estimate of substitution, each entry is the probability of substitution (r) from one base (row) to another base (column). Substitution pattern and rates were estimated under the model of Tamura and Nei [22]. For estimating ML (maximum likelihood) values, a tree topology was automatically computed. Relative values of instantaneous r were considered when evaluating them. For simplicity, sum of r values was made equal to 100, the nucleotide frequencies were A = 25.12%, T/U = 22.21%, C = 25.88%, and G = 26.79.

The estimates of Evolutionary Divergence between Sequences analysis were conducted using the Maximum Composite Likelihood.

III. RESULTS

A. Phenotypic diversity

The number of tiller significantly differed among lines. Differences also existed among treatments across lines (Appendix 1). Significant differences existed among treatments and also among lines in terms of green leaves number. The later trait was significantly influenced by the interaction between replication (block) and treatment. Lines were significantly and differently affected by salt treatments as far as panicle number is concerned. The panicle number was significantly affected by line treatment interaction. In terms of panicle weight some evidences showed that there was significant variability among lines as well as among treatments. An interaction also existed between salt level and lines for this trait. The analysis of variance also showed that biomass produced greatly varied among lines and also among treatments.

The visual score showed that the lines were tolerant (4.8) moderately tolerant (5.5), susceptible (7) and 9 highly susceptible (Table I). Lines means comparison for tiller number displayed 3 groups, the most performing being IRRI113. The second group was composed of NSIC RC10 and the third group was composed of Gambiaka and IR1529. IRRI113 and NSIC RC 106 had significant ability to keep more green leaves in salt condition. IR1529 and Gambiaka had poor performance of bearing functional leaves in salt condition. Thus significant diversity existed among lines concerning this trait. The panicle number differed significantly among lines across treatments. There was clear evidence that the best lines concerning panicle number were IRRI113 and NSIC RC 106. Lines means comparison (Table 2) also showed that there was strong evidence that IRRI113, and NSIC RC106, were lines that bore the heaviest panicles under salt stress.

<table>
<thead>
<tr>
<th>Traits</th>
<th>Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>visual scoring</td>
<td>4.8 5.5 7 9</td>
</tr>
<tr>
<td>Tiller number</td>
<td>14.7A 7.7B 4.2C 3.1C</td>
</tr>
<tr>
<td>Green leaves number</td>
<td>4.1A 4.1B 1.58B 1B</td>
</tr>
<tr>
<td>Panicle number</td>
<td>11.4A 4.3B 3.6C 1B</td>
</tr>
<tr>
<td>Panicle weight (g)</td>
<td>19.50 12.66 5.36 4.87</td>
</tr>
<tr>
<td>Dry weight (g)</td>
<td>37.72A 22.44Bc 17.37Bc 16.49Bc</td>
</tr>
</tbody>
</table>

NB: Lines followed by the same letter are not significantly different

B. Genotypic Statistics on individuals

The heterozygous frequencies were null meaning that all the parents were homozygous (Table II). All the lines had four alleles per locus. The ‘A’ allele was stable across lines with a frequency of about 23%. The allele B frequency ranged from 20.7 in IRRI113 genome to 22.4 in Gambiaka genome with an average of 21.8%. The following values: 25.1, 26.4, and 25.32 respectively were the minimum, maximum and average allele D frequency across lines. The most frequent allele’s average percentage ranged from 21.8 to 25.32% across lines. The average missed values percentage equaled to 4.62. The recombinant ranged from 1439 for IR1529 to 1485 for NSIC RC106. The average distance occupied one the chromosome was 446.27, 413.11, 467.59, 479.91cM by allele A, B, C, and D respectively.
Clear evidence of genetic diversity was shown by the linkage group (Fig. 1). This figure showed the allele frequencies in chromosome 1 of the four lines.

**C. Homogeneity of Substitution Patterns between Sequences**

The estimates of homogeneity of substitution per line are shown above the diagonal (Table III). This table showed that Gambiaka and IR1529 on one hand and Gambiaka and NSIC RC 106 on the other hand did not have sequences involved the same pattern of substitution (P < 0.05). This was also true for IR1529 and NSIC RC106. The estimates of the disparity index per line are shown for each sequence pair below the diagonal. The disparity index for sequences was high between the two local varieties (Gambiaka and IR1529) and between the local varieties and NSIC RC 106. The substitution pattern was not significant between the imported lines (NSIC RC106 and IRR1103) and between the local varieties and IRR1113.

**D. Maximum likelihood Estimate of substitution**

Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics (Table IV). A transition is a point mutation that changes a purine nucleotide to another purine (A ↔ G) or a pyrimidine nucleotide to another pyrimidine (C ↔ T). Approximately 68% of nucleotide polymorphisms (SNPs) were transitions and 32% were transversion. Transversional substitution is a change of purine nucleotide to pyrimidine or vice versa. The nucleotide frequencies were: A = 25.12%, T/U = 22.21%, C = 25.88%, and G = 26.79%.

**E. Estimates of net base composition disparity between sequences**

Disparity Index per line is shown for all sequence pairs in Table V. The difference in bases composition between IRR1113 and the other lines (IR1529, Gambiaka and NSIC RC106) were larger than can be expected based on evolutionary divergence between sequences and by chance alone (values greater than 0). So in terms of base composition IRR1113 was significantly different from the others.

**F. Estimates of divergence between sequences**

The numbers of base substitutions per line from between sequences are shown in Table IV. Significant divergence between sequences was noted among all the lines.
TABLE VI: EVOLUTIONARY DIVERGENCE BETWEEN SEQUENCES

<table>
<thead>
<tr>
<th></th>
<th>Gambiaka</th>
<th>IR1529</th>
<th>IRRI113</th>
<th>NSIC RC106</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambiaka</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR1529</td>
<td>0.32</td>
<td>0.32</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>IRRI113</td>
<td>0.39</td>
<td>0.31</td>
<td>0.31</td>
<td>0.34</td>
</tr>
<tr>
<td>NSIC RC106</td>
<td>0.35</td>
<td>0.34</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>

G. Estimates of base composition difference between sequences

The difference in base composition bias per line is shown in Table VII. Note that even when the substitution patterns are homogeneous among lineages, the compositional distance will correlate with the number of differences between sequences.

TABLE VII: BASE COMPOSITION DIFFERENCE BETWEEN SEQUENCES

<table>
<thead>
<tr>
<th></th>
<th>Gambiaka</th>
<th>IR1529</th>
<th>IRRI113</th>
<th>NSIC RC106</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambiaka</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR1529</td>
<td>0.10</td>
<td>0.06</td>
<td>0.02</td>
<td>0.35</td>
</tr>
<tr>
<td>IRRI113</td>
<td>0.66</td>
<td>0.37</td>
<td>0.37</td>
<td>0.35</td>
</tr>
<tr>
<td>NSIC RC106</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
<td>0.35</td>
</tr>
</tbody>
</table>

H. Polymorphic SNPs distribution across rice genome

A list of 200 Polymorphic SNPs was selected. The SNPs were approximately evenly distributed (Fig. 2) across the rice genome. There were a good number of polymorphic SNP markers in the chromosomes 1 to 11, but only 6 SNPs were polymorphic in the Chromosome 12. Thus, the number of SNP markers ranged from 6 to 19 per chromosome.

The average distance of SNPs in chromosome 1 was 93.21 cM with a maximum of 186 cM and a minimum of 0.87 cM. In the chromosome 2, the average distance was 62.89 cM with a maximum of 161 and a minimum of 4.08. In chromosome 3 the mean distance was 51.05 with a maximum of 160 and a minimum of 0.89 cM. The distance ranged from 1.78 to 115 cM in chromosome 4, from 6.35 to 107 cM in chromosome 5 and from 0.62 to 116 cM in chromosome 6.

The average distance of the 17 SNPs in chromosome 7 was 47.91 cM with a maximum of 113 cM and a minimum of 0.34 cM. In chromosome 8 the markers were distributed with a mean distance of 57.75 cM a maximum of 121.7 cM and a minimum of 5.25 cM. The SNP marker the most distanced in chromosome 9 was at 92.8 cM where the nearest was at 1.31 cM, the average being 52.67 cM. In chromosome 10 SNPs distances ranged from 1.09 cM to 83.2 cM. The average distance of the 6 SNP markers was 55.5 cM with a maximum of 100 and a minimum 43.6 cM.

IV. DISCUSSION

Results revealed significant diversity among lines concerning all the parameters measured. The tiller number, green leaf number, panicle number, panicle weight and dry weight differed significantly among lines within and across treatments this confirmed the study of Souleymane et al. [18]. Phenotypic diversity was also confirmed by the genotypic data.

The heterozygous frequency was zero for all the lines. This means that the lines used in crossing were highly homozygous with 100% homozygosity. Thus, these lines were pure fixed lines. The most frequent allele differed from line to line except for the allele A which was slightly stable across lines. These differences might be due to genotypic diversity of the 4 parents.

The results showed significant difference in substitution pattern between the two farmer preferred varieties: IR1529 and Gambiaka, and between the local varieties and NSIC RC106. This was confirmed by a high disparity index. This might also be due to the genetic diversity of these lines (IR1529, Gambiaka and NSIC RC 106). But our method failed to show significant difference of substitution pattern and high disparity between IRRI113 and the 3 others lines. The results showed that 68% of the substitutions were transition and 32% were transversion. This confirmed the results of Tamura et al. [23] who showed that about two third (2/3) of the substitutions were transitions. According to Steven [19] because of "wobble" phenomenon, transitions rarely bring about amino acid substitution. Hence, Steven
In terms of base composition IRRI113 was significantly different from the others lines (IR1529, Gambiaka, NSIC RC106). Thus, the only one that was not significantly different from the others in terms of substitution pattern was in terms of base composition. This means that all the lines were genetically different either in terms of base composition or in terms of substitution pattern or both. The results showed also that sequences diverge significantly among all the lines. This corroborated the above results.

The 200 polymorphic SNPs markers were evenly distributed across rice genome. However, chromosomes did not have the same number of polymorphic SNPs. Rice chromosomes have different genome structure and characteristic. According to Wu et al. [24] and Zhang et al. [26] the size of rice chromosomes varies from 45Mb to 24Mb and the gene density ranges from 8.9 to 11.6. According to Guo-Liang [8] SNPs might be present within coding sequences of genes, non-coding regions of genes or in the intergenic regions between genes at different frequencies in different chromosome regions.

V. CONCLUSION

Molecular characterization of parental lines allowed us to know the homozygosity level. Thus, all the lines were 100% homozygous and genetically diverse. The substitution pattern and disparity index between sequences showed differences between local varieties and one exotic parental line. However, divergence between sequences and base composition difference between sequences gave evidence of genetic diversity of the 4 parents. Two hundred out of 1896 SNPs were polymorphic and were distributed evenly across the rice genome. The SNPs number and position on chromosomes differed from one chromosome to another. The two hundred polymorphic SNPs will be used to line the rice generations derived from the cross of those parents to find QTL for salt tolerance. This will facilitate the selection of salt tolerance lines.

**APPENDIX**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>TN</th>
<th>GLN</th>
<th>PN</th>
<th>PW</th>
<th>DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td>1</td>
<td>364.59</td>
<td>15.68</td>
<td>450</td>
<td>415.64</td>
<td>4433.05</td>
</tr>
<tr>
<td>Treatment</td>
<td>4</td>
<td>7093.36***</td>
<td>680.4***</td>
<td>1728.35**</td>
<td>3000.44**</td>
<td>5276.94**</td>
</tr>
<tr>
<td>Replication*Treatment</td>
<td>4</td>
<td>54.05***</td>
<td>51.22**</td>
<td>170.85**</td>
<td>191**</td>
<td>1452.59**</td>
</tr>
<tr>
<td>Line</td>
<td>3</td>
<td>1039.0***</td>
<td>261.85**</td>
<td>4682.70**</td>
<td>6350.37**</td>
<td>36080.92**</td>
</tr>
<tr>
<td>Treatment*Line</td>
<td>12</td>
<td>1453.31***</td>
<td>236.32**</td>
<td>5685.45*</td>
<td>7667.09*</td>
<td>46630.85**</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>1691.31</td>
<td>220.10</td>
<td>3894.15</td>
<td>5391.37</td>
<td>32169.40</td>
</tr>
</tbody>
</table>

* *, **, and *** significant at 0.05, 0.01 and 0.001 probability level respectively. TN= Tillers number, GLN = green leaves number, PN= panicle number, DW = dry weight.

**ACKNOWLEDGMENT**

We acknowledge the Generation Challenge Programme (GCP) for supporting genotyping cost and IRRI for providing salt tolerant germplasm.

**REFERENCES**


DOI: http://dx.doi.org/10.24018/ejers.2017.2.12.504


[26] Zhang, Y, Y Huang, L Zhang, Y Li, T Lu, Y Lu, Q Feng, Q Zhao, Z Cheng, Y Xue, RA Wing, and B Han (2004) Structural features of the rice chromosome 4 centromere. Nucl. Acids Res. 32 2023-2030.

The Author was born at Karta/Niger 01/01/1977. Hold a PhD in plant breeding at West Africa Centre for Crop Improvement (WACCI/University of Ghana, MSc in plant modeling at Université Cheikh Anta Diop in Dakar/Senegal, MSc in crop production at Université Abdou Moumouni de Niamey/Niger. Oumarou is an ASSOCIATE RESEARCHER at National Institute for Agricultural Research of Niger (INRAN) at Regional centre for Agricultural research of Kollo. He is interested in development of rice improved varieties tolerant to abiotic stress. His has published many articles including:


Dr. Souleymane is also a lecturer at Universities of Tahoua and Say in Niger.