

ASSESSMENT OF GENETIC DIVERSITY IN RICE GERMPLASM COLLECTION USING A PANEL OF 50 STANDARD SSR MARKERS: VALIDATION AND APPLICATIONS

Phong Ngoc Hai Trieu¹, Nguyen Le Van¹ and Do Duc Tuyen^{2,3,*}

¹Center for Analyzing Soil, Water, and Crop Products, CuuLong Delta Rice Research Institute, Tan Thanh Commune, Thoi Lai District, Can Tho City, Vietnam;

²Scientific Management and International Cooperation Office, CuuLong Delta Rice Research Institute, Tan Thanh Commune, Thoi Lai District, Can Tho City, Vietnam;

³Genetics and Plant Breeding Department, CuuLong Delta Rice Research Institute, Tan Thanh Commune, Thoi Lai District, Can Tho City, Vietnam.

*Correspondence: ddtuyenvlt09@gmail.com (D. D. Tuyen).

ABSTRACT

Assessing genetic diversity and establishing a genetic database for rice germplasm collection is crucial to conserve and effectively utilize rice genetic resources. In addition to traditional agronomic traits, molecular markers offer distinct advantages in achieving this goal. This study genotyped 48 different rice varieties/lines using 50 SSR markers spread across 12 chromosomes of the rice genome. The results revealed that 30 SSR markers from the panel exhibited polymorphism, with the number of alleles ranging from 2 to 5. The analysis of these 30 SSR markers indicated significant differences among the lines/varieties, with a maximum similarity coefficient of 91%. To simplify and enhance the analysis, a subset of 24 SSR markers was identified, which showed highly similar results to the original set of markers, with a correlation coefficient of 97.7%. As a result, this set of 24 SSR markers derived from 30 polymorphic markers holds tremendous potential for analyzing genetic diversity on a large scale, constructing a comprehensive genetic database for the rice gene bank, and identifying distinct rice varieties/lines. These findings highlight the importance of utilizing molecular markers in conservation efforts and the sustainable utilization of rice genetic resources.

Keywords: Genetic diversity, molecular marker, SSR marker, rice variety.

INTRODUCTION

Rice (*Oryza sativa* L.) is one of the three main crops and provides food for over half of the world's population (FAO 2008; Kisk 2008). In Vietnam, rice plays a significant role in both the nation's economic growth and food security. The genetic diversity of Vietnamese rice germplasm is also highly valued (La et al. 2011), and many new varieties are developed each year. The diversity of Vietnamese rice germplasm has been useful for rice production and research but also has led to certain difficulties in management and exploitation. Evaluating genetic diversity and

building a database for rice genetic resources is meaningful for both theoretical and applied studies.

The panel of 50 standard SSR markers distributed across 12 chromosomes of the rice genome published by the International Rice Research Institute (IRRI) showed a high potential for detecting polymorphism in the rice population (http://gramene.org/markers/microsat/50_ssr.html). These markers have been utilized in many rice diversity studies. They have been used to distinguish rice varieties grown in Bangladesh (Rahman et al. 2009). Thirty-five

SSR markers were used to divide 23 hybrid rice varieties into 4 groups in another study in India (Nikam *et al.* 2016). Combining the markers in the standard set with others has also been used in many genetic diversity studies on rice (Yadav *et al.* 2013; Nikam *et al.* 2016; Doan *et al.* 2016). Notably, some markers were used to identify lines in the testing system, and also suggested using additional SSR markers for the DUS testing (Pourabed *et al.* 2015). Therefore, the standard set of markers has shown advantages in evaluating diversity as well as identifying rice lines/varieties.

This study aims to validate the efficiency of using the panel of 50 standard SSR markers for investigating the genetic diversity of rice varieties in Vietnam. The results of this study indicate a concise and efficient set of markers

for identifying rice lines/varieties and creating a rice genetic database for Genbank.

MATERIALS AND METHODS

Materials

Totally, forty-eight rice varieties from different origins comprising 29 local rice varieties obtained from the Genbank of Cuu Long Delta Rice Research Institute (CLRRI), 10 hybrid rice varieties developed by CLRRI, 5 hybrid rice varieties from external organizations, 2 rice varieties from other countries, and 2 rice varieties from the Central and Northern regions of Vietnam were examined (**Table 1**). The diversity of these materials is reflected in type and grains, as shown in **Figure 1**.

Table 1. List of the rice genotypes used in the study.

No.	Name	Rice genotypes	Origin/Source
1	Acc.1013	Local rice	CLRRI
2	Acc.1029	Local rice	CLRRI
3	Acc.1079	Local rice	CLRRI
4	Acc.1192	Local rice	CLRRI
5	Acc.121	Local rice	CLRRI
6	Acc.1228	Local rice	CLRRI
7	Acc.1231	Local rice	CLRRI
8	Acc.1408	Local rice	CLRRI
9	Acc.1715	Local rice	CLRRI
10	Acc.1748	Local rice	Vietnam
11	Acc.2086	Local rice	CLRRI
12	Acc.2173	Local rice	CLRRI
13	Acc.2197	Local rice	CLRRI
14	Acc.2357	Local rice	CLRRI
15	Acc.237	Local rice	CLRRI
16	Acc.2608	Local rice	CLRRI
17	ACC.262	Local rice	CLRRI
18	Acc.2944	Local rice	CLRRI

No.	Name	Rice genotypes	Origin/Source
19	Acc.583	Local rice	CLRRI
20	Acc.672	Local rice	CLRRI
21	Acc2982	Local rice	CLRRI
22	OM108	OM6976/NPT8	CLRRI
23	OM3536	TD8/OM1738	CLRRI
24	OM3673	IR65418/OM6976	CLRRI
25	OM380	IR50404/OM5472	CLRRI
26	OM4218	OM2031/MTL250	CLRRI
27	OM4900	C53/Jasmine85	CLRRI
28	OM5451	Jasmine 85/OM2490	CLRRI
29	OM6976	IR68114/OM997//OM2718/ OM2886	CLRRI
30	OM8	OM18/BátTiên//IR64Saltol//2*Jasmi ne85/////OM5451/////AP	CLRRI
31	OM9	Bát Tiên/Jasmine 85//IR64 Saltol//OM5451/////AP/////OM7347	CLRRI
32	PL20S	-	Long Ho, Vinh Long
33	Sen-tim	Local rice	Vietnam
34	ST24	-	Center for Seeds and Seedlings, Soc Trang
35	Thai-Brown	-	Florida, US
36	TT	-	Thanh Tri, Ha Noi
37	VD20	Introduced rice variety	CLRRI
38	BT	-	Nghe An
39	DT8	BVN/OM4900	Southern Seed Corporation
40	GB.07	Local rice	Vietnam
41	Jasmine85	Selected from Thailand Jasmine	CLRRI
42	KDM105	-	Thailand
43	M26	Local rice	CLRRI
44	NDHD	Derived from LD2012 variety	Dinh An cooperative, Lap Vo, Dong Thap
45	Nep31	Local rice	CLRRI
46	Nep-than	Local rice	Dong Thap
47	NTCD1	Local rice	CLRRI
48	NTCD2	Local rice	CLRRI



Figure 1. Type and color grain of 48 rice lines/varieties.

The set of 50 SSR markers distributed on 12 chromosomes of the rice genome (**Table 2**) has been published by the International Rice Research Institute (IRRI), indicating a high potential for detecting polymorphism in the rice population

(http://gramene.org/markers/microsat/50_ssr.html). The primers of these markers were synthesized by Phu Sa Biochemical Co., Ltd (<http://www.phusabiochem.com/vi/.html>), and PCR reactions were optimized and carried out in the Central Laboratory, CLRRI.

Table 2. 50 SSR makers and the distribution in the rice genome.

No.	Maker	Chr.	Position (cM)	Forward	Reverse
1	RM495	1	2.8	aatccaaggtgcagagatgg	Caacgatgacgaacacaacc
2	RM1	1	29.7	gcgaaaacacaatgcaaaaa	Gcgttggttgacctgac
3	RM283	1	31.4	gtctacatgtacccttgttggg	Cggcatgagagtctgtgatg
4	RM259	1	54.2	tggagtttgagaggaggg	Cttgttgcattggtccatgt
5	RM312	1	71.6	gtatgcatatttgataagag	Aagtcaccgagtttaccttc
6	RM5	1	94.9	tgcaactctagctgctcga	Gcatccgatcttggatggg
7	RM237	1	115.2	caaatcccactgctgtcc	Tgggaagagagcactacagc
8	RM431	1	178.3	tcttgcgaactgaagagttg	Agagcaaaacctggttcac

No.	Maker	Chr.	Position (cM)	Forward	Reverse
9	RM154	2	4.8	accctctccgctcgcctctc	Ctctctctctcgacgcctcc
10	RM452	2	58.4	ctgatcgagagcgtaaggg	Gggatcaaaccacgtttctg
11	RM489	3	29.2	acttgagacgatcggacacc	Tcaccatggatgtgtcag
12	OSR13	3	53.1	catttgtcgtcacggagta	Agccacagcgcccatctctc
13	RM338	3	108.4	cacaggagcaggagaagagc	Ggcaaaccgatcactcagtc
14	RM55	3	168.2	ccgtcgcgtagtagagaag	Tcccggttatttaaggcg
15	RM514	3	216.4	agattgatctccattcccc	Cacgagcatattactagtgg
16	RM307	4	0	gtactaccgacctaccgttcac	Ctctatgcatgaactgctc
17	RM124	4	150.1	atcgtctcgttgctggctgctg	Catggatcaccgagctcccc
18	RM507	5	0	cttaagctccagccgaaatg	Ctaccctcatcctgcc
19	RM413	5	26.7	ggcgattcttgatgaagag	Tccccaccaatcttctctc
20	RM161	5	96.9	tgcagatgagaagcggcgctc	Tgtgtcatcagacggcctccg
21	RM178	5	118.8	tcgcgtgaaagataagcggcgc	Gatcaccgttcctccgctgc
22	RM334	5	141.8	gttcagtgttcagtccacc	Gactttgatcttgggtggacg
23	RM133	6	0	ttggattgtttgctggctcgc	Ggaacacggggtcggaagcgac
24	RM510	6	20.8	aaccggattagtttctcgcc	Tgaggacgacgagcagattc
25	RM454	6	99.3	ctcaagcttagctgctgctg	gtgatcagtccatcatagcg
26	RM162	6	108.3	gccagcaaaaccaggatccgg	caaggtcttgcggcttgcgg
27	RM125	7	24.8	atcagcagccatggcagcgacc	aggggatcatgtccgaagggc
28	RM11	7	47	tctctcttcccccgatc	atagcgggagaggttag
29	RM455	7	65.7	aacaaccaccacctgtctc	agaaggaaaagggtcctgatc
30	RM118	7	96.9	ccaatcggagccaccggagagc	cacatctccagcgacgccgag
31	RM408	8	0	caacgagctaactccgtcc	actgctactgggtagctgacc
32	RM152	8	9.4	gaaaccaccacactcaccg	ccgtagacctctgaagtag
33	RM25	8	52.2	ggaagaatgatctttcatgg	ctaccatcaaaaccaatgttc
34	RM44	8	60.9	acgggcaatccgaacaacc	tgggaaaacctaccctacc
35	RM284	8	83.7	atctctgatactccatccatcc	cctgtacgttgatccgaagc
36	RM433	8	116	tgcgctgaactaaacacagc	agacaaacctggccattcac
37	RM447	8	124.6	cccttgtgctgtctctctc	acgggcttctctctctc
38	RM316	9	1.8	ctagttgggcatac gatggc	acgcttatatgttacgtaac
39	RM105	9	32.1	gtcgtcgaccatcggagccac	tggtcgaggtggggatcgggtc

No.	Maker	Chr.	Position (cM)	Forward	Reverse
40	RM215	9	99.4	caaaatggagcagcaagagc	tgagcacctcctctctgtag
41	RM474	10	0	aagatgtacgggtggcattc	tatgagctggtagcaatgg
42	RM271	10	59.4	tcagatctacaattccatcc	tcggtagacctagagagcc
43	RM171	10	73	aacgcgaggacacgtacttac	acgagatacgtacgcctttg
44	RM484	10	97.3	tctcctcctcaccattgtc	tgctgccctctctctctctc
45	RM552	11	40.6	cgcagttgtggatttcagtg	tgctcaacgtttgactgtcc
46	RM536	11	55.1	tctctcctctgtttggctc	acacaccaacacgaccacac
47	RM287	11	68.6	ttcctgttaagagagaaatc	gtgtatttggtagaaagcaac
48	RM144	11	123.2	tgccctggcgcaaatttgatcc	gctagaggagatcagatggtagtgc
49	RM19	12	20.9	caaaaacagagcagatgac	ctcaagatggacccaaga
50	RM277	12	57.2	cggtcaaatcatcacctgac	caaggcttgaagggaag

Methods

The total DNA of 48 rice lines/varieties was extracted from leaves at the seedling stage using TPS buffer (GSL-IRRI, 2017) (1M KCl, 1M Tris-pH 8, 0.5M EDTA-pH 8). PCR reactions were carried out with a volume of 10 μ L, including 25-30 ng of DNA template, 10 pM of forward and reverse primers, 10 mM dNTPs, 10X PCR buffer, and 0.5 μ L Taq polymerase. The thermal cycle consisted of denaturation at 94°C for 5 minutes, followed by 35 cycles of denaturation at 94°C for 1 minute, annealing at 55°C for 1 minute, extension at 72°C for 1 minute, and a final extension at 72°C for 7 minutes. The PCR products were run on a 3% agarose gel in 1X TAE buffer, stained with SafeView, and then observed. The banding patterns were recorded on the Uvidoc HD6 system. The polymorphic information content (PIC) was calculated using the formula from PowerMarker V3.25 software (Liu and Muse 2004), and the phylogenetic trees were

constructed using the NTSYS 2.1 software (Rohlf 2000).

RESULTS AND DISCUSSION

Evaluation of the polymorphism of 50 SSR molecular markers on 48 rice lines/varieties

There were 47/50 SSR markers of the panel presented PCR products, of which 30 markers were found to be polymorphic. The number of alleles ranged from 2 to 5, with RM307 being the most polymorphic marker with 5 alleles, RM152 with 4 alleles, 5 markers with 3 alleles (RM474, RM124, RM215, RM133, and RM316), and the remaining markers with 2 alleles (**Table 3**). Therefore, 17 monomorphic markers were unable to differentiate between the rice lines/varieties in this study. Most polymorphic markers showed clear differences among these rice lines/varieties (**Figures 2 and 3**).

$$PIC = 1 - \sum_{i=1}^n p_i^2$$

Table 3. The diversity coefficient and PIC values of the SSR markers used in this analysis.

No	Marker	Allele frequency	No. of allele	Diversity value	PIC
1	RM307	0.58	5	0.61	0.58
2	RM152	0.52	4	0.61	0.55
3	RM474	0.56	3	0.51	0.39
4	RM287	0.52	2	0.50	0.37
5	OSR13	0.63	2	0.46	0.36
6	RM25	0.64	2	0.46	0.36
7	RM552	0.64	2	0.46	0.35
8	RM154	0.70	2	0.42	0.33
9	RM19	0.77	2	0.35	0.29
10	RM124	0.80	3	0.34	0.32
11	RM118	0.78	2	0.34	0.28
12	RM510	0.79	2	0.33	0.28
13	RM334	0.80	2	0.32	0.27
14	RM215	0.81	3	0.32	0.3
15	RM433	0.80	2	0.32	0.27
16	RM44	0.81	2	0.30	0.26
17	RM413	0.82	2	0.29	0.25
18	RM161	0.84	2	0.27	0.23
19	RM514	0.84	2	0.27	0.23
20	RM284	0.84	2	0.26	0.23
21	RM11	0.86	2	0.23	0.21
22	RM133	0.89	3	0.21	0.19
23	RM162	0.89	2	0.20	0.18
24	RM489	0.91	2	0.17	0.15
25	RM259	0.92	2	0.15	0.14
26	RM237	0.94	2	0.12	0.11
27	RM454	0.96	2	0.08	0.08
28	RM316	0.96	3	0.08	0.08
29	RM283	0.98	2	0.04	0.04
30	RM125	0.98	2	0.04	0.04
Average		0.79	2.3	0.30	0.26

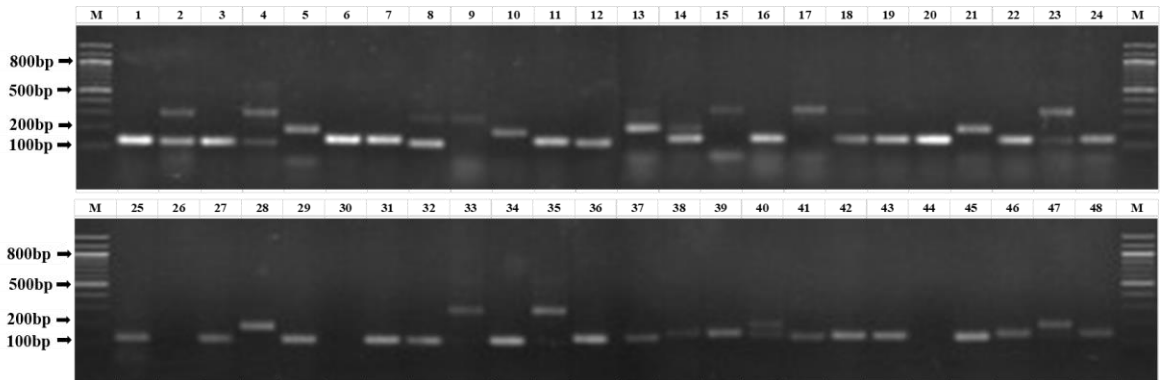


Figure 2. Electrophoresis analysis of the amplified PCR products of 48 rice lines/varieties using RM307.

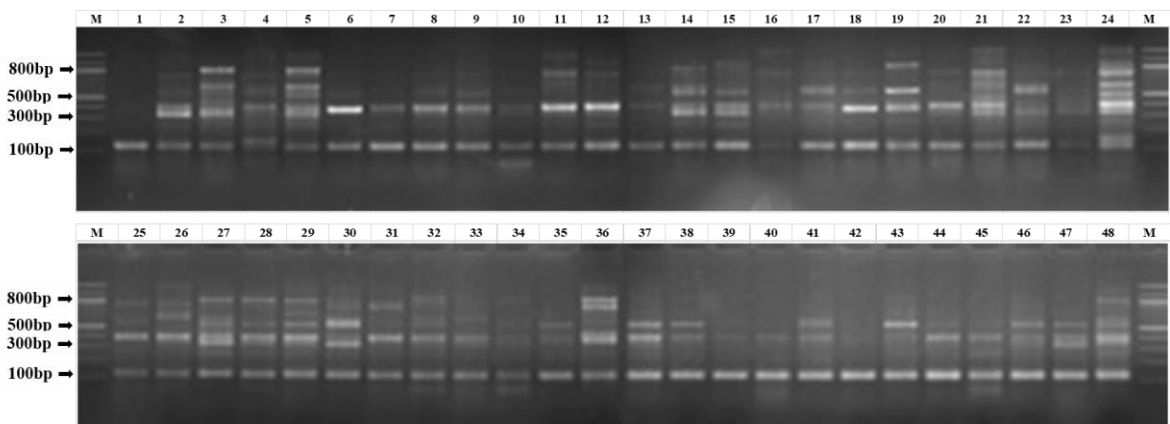


Figure 3. Electrophoresis analysis of the amplified PCR products of 48 rice lines/varieties using RM125.

In previous studies that used a set of 50 markers, some markers had higher allele numbers than what was found in this study. For example, RM154 had 24 alleles detected in one study and 7 alleles in another study, while RM413 had 7 alleles detected (Ali *et al.* 2011). This difference may be due to the fact that the lines/varieties in this study were mostly collected from Vietnam and were derived from a limited number of rice lines/varieties.

Analyzing the genetic diversity of 48 rice lines/varieties by using polymorphic SSR markers

To analyze the diversity of the 48 rice

lines/varieties, 30 polymorphic markers that showed PIC coefficients ranging from 0.04 (RM125) to 0.58 (RM307) with an average PIC coefficient of 0.26, were used. These markers were also used to cluster the lines/varieties genetically, and the results showed that the lines/varieties were divided into groups from a Jaccard similarity coefficient of 0.43 (**Figure 4A**). The highest correlation was found between M26 and Acc.2357, with a similarity coefficient of 0.91. Therefore, using 30 polymorphic SSR markers, all lines/varieties in this study were distinguished.

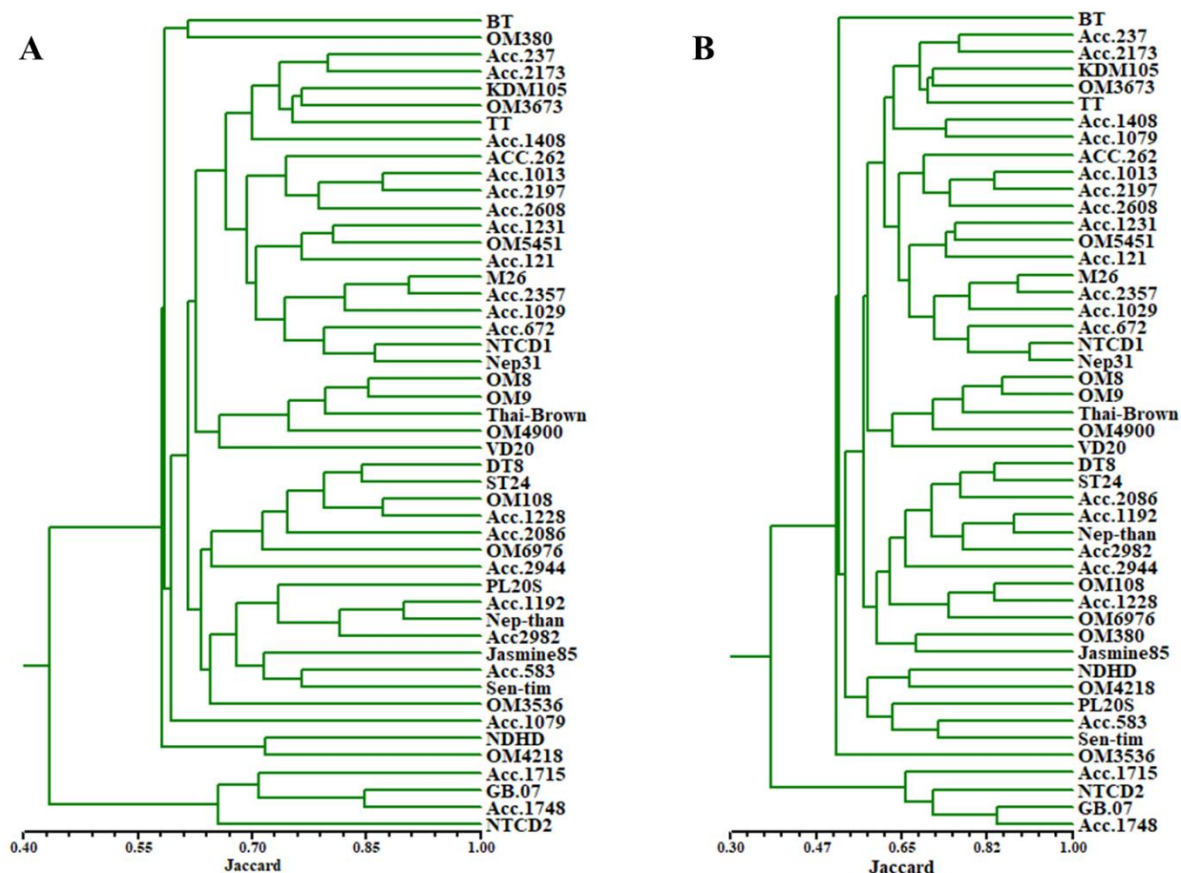


Figure 4. (A) Phylogenetic tree of 48 rice line/varieties based on the set of 30 SSR markers; (B) Phylogenetic tree of 48 rice line/varieties based on the set of 24 SSR markers.

Some studies have also used genetic clustering analysis to divide subgroups, such as the division of three groups based on the analysis of 12 SSR markers (Pourabed et al. 2015) or the identification of two main groups and four subgroups based on the analysis of 35 SSR markers across 23 rice varieties (Nikam et al. 2016).

Development of a panel of optimal markers for analysis of diversity and identification of rice varieties

Based on the values of PIC, diversity value, and distribution of SSR markers in the genome, a small set of 24 SSR markers from 30 polymorphic markers with the highest PIC values and diversity values were selected for the

analysis of genetic diversity among 48 rice lines/varieties. The results of genetic clustering were shown in **Figure 4B**, in which the lines/varieties were grouped at a Jaccard similarity coefficient of 0.38, with the two most similar lines with a coefficient of 0.91. These results also indicate significant differences among the rice lines/varieties, as seen in the analysis of a larger set of more than 30 polymorphic SSR markers. This was also demonstrated by the high correlation coefficient (Matrix correlation) of 0.98 between the two phylogenetic trees (**Figure 5**). Therefore, this small set of 24 SSR markers from the total set of 50 SSR markers could be utilized to analyze genetic diversity and identify different rice varieties in the future.

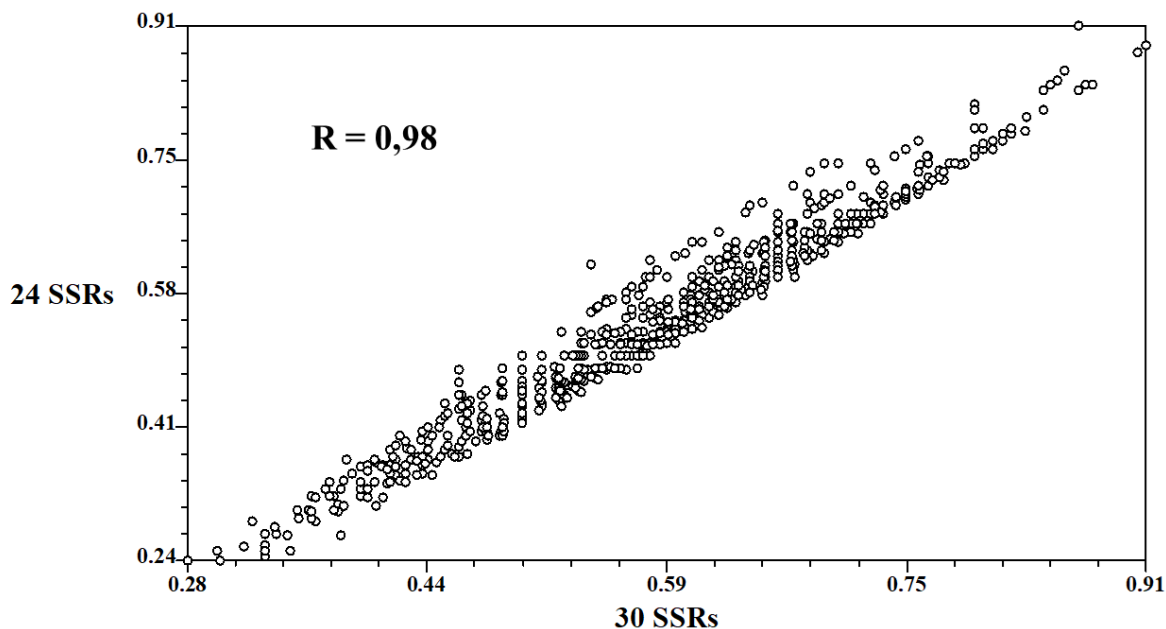


Figure 5. Correlation between phylogenetic trees based on the set of 30 SSR markers and phylogenetic trees based on the set of 24 SSR markers.

CONCLUSIONS

In this study, the use of 50 polymorphic markers was effective in distinguishing rice lines and varieties with various origins. A subset of 24 markers was identified as highly promising for future breeding programs to analyze the genetic diversity of the rice varieties in the CLRRI's Genbank, and to establish a rice variety database. Additionally, these markers may be utilized in the identification of newly developed rice varieties by CLRRI.

COMPETING INTERESTS

The authors declare they have no conflict of interest, financial or otherwise.

AUTHOR'S INFORMATION AND CONTRIBUTIONS

Phong Ngoc Hai Trieu carried out experiment and wrote the draft version of the manuscript. Nguyen Le Van carried out experiment, reviewed and edited the manuscript. Do Duc Tuyen designed the study, analyzed data, reviewed and edited the manuscript. All authors read and approved the final manuscript.

REFERENCES

- Ali ML, McClung AM, Jia MH, Kimball JA, McCouch SR, and Eizeng GC (2011) A rice diversity panel evaluated for genetic and agro-morphological diversity between subpopulations and its geographic distribution. *Crop Sci* 51(5):2021-2035.
- Doan TQ, Nguyen TH, Vu TTH, and Tran VQ (2016) Analysis of genetic diversity in local sticky rice based on morphological characteristics and SSR markers. *Vietnam Journal of Agricultural Sciences*, 14(4):527-538 (In Vietnamese).
- Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11-15.
- FAO (2008) Food outlook. Global market analysis. November 2008. pp 14-27. Available at <http://www.fao.org/docrep/011/ai474e/ai474e05.htm> (verified 25 May 2011). FAO, Rome, Italy.

- Kish S (2008) Determining rice gene function. Available at http://www.csrees.usda.gov/newsroom/impact/2009/nri/02171_rice_gene.html (verified 25 May 2011). USDA-CSREES, Washington, DC.
- La TN, Tran DS, Le KT, Luu QH, Vu LC, Vu VT, and Hoang TH (2011) Vietnam plant resources: Achievements and plan of plant resources conservation for sustainable agricultural development and food Security. *Journal of Vietnam Agricultural Science and Technology*, 2(14):3-10 (In Vietnamese).
- Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics Applications Note* 21(9):2128-2129.
- Nikam DP, Bhave SG, Sawardekar SV, Gokhale NB, Patil DM, and Mhatre NK (2016) DNA fingerprinting of rice varieties using SSR markers. *J Agric Res Technol* 41(1):142-147.
- Pourabed E, Noushabadi MRJ, Jamali SH, Alipour NM, Zareyan A, and Leila SL (2015) Identification and DUS testing of rice varieties through microsatellite markers. *Int J Plant Genomics* 2015:1-7. DOI: 10.1155/2015/965073.
- Rahman MS, Molla MdR, Alam MdS. and Rahman L (2009) DNA fingerprinting of rice (*Oryza sativa* L.) cultivars using microsatellite markers. *Aust J Crop Sci* 3(3):122-128.
- Rohlf FJ (2000) NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.1. Exeter Software, Setauket, New York 2000.
- Yadav S, Singh A, Singh MR, Goel N, Vinod KK, Mohapatra T, and Singh AK (2013) Assessment of genetic diversity in Indian rice germplasm (*Oryza sativa* L.): use of random versus trait-linked microsatellite markers. *J Genet* 92(3):546-557.

ĐÁNH GIÁ KHẢ NĂNG ỨNG DỤNG BỘ CHỈ THỊ PHÂN TỬ CHUẨN TRONG NGHIÊN CỨU ĐA DẠNG VÀ XÁC ĐỊNH CÁC GIỐNG LÚA

Đánh giá đa dạng di truyền và xây dựng dữ liệu di truyền cho các dòng/giống lúa được xem là việc cần thiết nhằm quản lý và khai thác nguồn gen cây lúa được tốt hơn. Ngoài các chỉ tiêu nông học việc sử dụng các chỉ thị phân tử đã thể hiện tính ưu việt cho mục đích này. Thí nghiệm được thực hiện đối với 48 dòng/giống lúa có nguồn gốc khác nhau và trên cơ sở phân tích 50 chỉ thị phân tử SSR chuẩn phân bố trên 12 nhiễm sắc thể của genome cây lúa. Kết quả cho thấy 30 chỉ thị phân tử SSR thể hiện đa hình với số allele thấp nhất là 2 và cao nhất là 5. Các dòng/giống lúa thể hiện sự khác biệt khá lớn trên cơ sở phân tích 30 chỉ thị phân tử SSR, trong đó các dòng/giống có hệ số tương đồng cao nhất là 91%. Trong số 30 chỉ thị phân tử SSR, một bộ nhỏ 24 chỉ thị phân tử SSR được chọn lọc và cũng cho kết quả tương tự với hệ số tương quan giữ hai bộ chỉ thị phân tử là 97,7%. Như vậy, bộ chỉ thị gồm 24 chỉ thị SSR rất triển vọng trong phân tích đa dạng di truyền trên quy mô lớn, xây dựng cơ sở dữ liệu cho ngân hàng gen cây lúa hay xác định các dòng/giống lúa.

Từ khóa: *Chỉ thị phân tử, chỉ thị SSR, dòng/giống lúa, đa dạng di truyền.*