



Genetic diversity of local rice varieties (*Oryza sativa* L.) in Vietnam's Mekong Delta based on SSR markers and morphological characteristics

Tran Huu Phuc¹, Van Quoc Giang², Nguyen Van Manh², Huynh Ky^{2*}

¹Department of Crop Sciences, College of Agriculture, Can Tho University, Vietnam

²Department of Genetic and Plant Breeding, College of Agriculture, Can Tho University, Vietnam

*Corresponding author: hky@ctu.edu.vn

SUBMITTED 27 January 2021 REVISED 21 April 2021 ACCEPTED 22 April 2021

ABSTRACT Based on target traits, use of the genetic diversity of rice is beneficial for crop improvement. In this study, 41 rice varieties local to Vietnam's Mekong Delta were evaluated on the basis of 11 quantitative morphological traits, along with the assessment of genetic diversity according to 50 SSR markers. The actual yield had a significance level of 0.05, while plant height and panicles per square meter had a high significance level of 0.001. Cluster analysis based on 11 quantitative traits also revealed that two were the optimal number of clusters used in this study. The highest polymorphic information content (PIC) value obtained was for RM286 (0.49), with a range of 0.00 to 0.49 and an average PIC of 0.14. Both structure and phylogenetic tree analyses as inferred from 50 SSR markers by the unweighted pair-group method with arithmetic mean (UPGMA) also indicated that the 41 local rice varieties could be divided into two major groups. This study provides a useful information for Mot bui do cao CM, and Mot bui five varieties for improvements in the yield and intermediate amylose content of local rice-breeding programs in future, especially for the Mekong Delta region.

KEYWORDS Genetic diversity; local rice; SSR; Vietnam's Mekong delta

1. Introduction

Rice (*Oryza sativa* L.) is one of the most important crops worldwide and is the main food for more than half of the world's population (Islam et al. 2018). Vietnam has a long-standing rice civilization and a large diversity of landscapes that contribute a high genetic diversity to its rice seed bank and potential materials for further researching (Phung et al. 2014). Thus, Mekong River Delta is also affected by climate change scenarios leading to the sea-level fluctuations at the mouth of the Mekong Delta (Keskinen et al. 2010). The sea-level rise is forecasted to have a significant impact on rice paddy fields in particular areas of the Mekong River basin in case of the saline water intrusion (Yu et al. 2010). To solve the problem, several concentration efforts as well as breeding programs on improving the productivity and quality of rice varieties have been established in Vietnam by the government and scientists. The priority of breeding programs, the diversity of genotypes in the population should be studied, in which the more genetic diversity is the more successful breeding strategies. The diversity of gene pools is including agronomic traits, such as quantity and quality of seed production, tolerance to abiotic stress and resistance to biotic stress (Litrice and Violle 2015).

Based on the evaluation and agro-morphological prop-

erties of rice genetic diversity, plant breeders have obtained the initial information of rice cultivated varieties through mutation and hybridization (Roy and Sharma 2014). Compared with traditional approaches such as morpho-physiological trait evaluation, modern breeding techniques and molecular markers are powerful tools that help breeders assess the genetic variation among rice germplasms in an effective way (Thi et al. 2015). Recent studies have used SSR markers to assess the genetic diversity of rice varieties because of its stability, effectiveness, and high polymorphism (Hue et al. 2018; Nguyen et al. 2012). Besides, amylose content (AC) in rice is one of the most important targets to determine rice quality, which is influenced primarily by starch, a composition of amylose and amylopectin (Khoomtong and Noomhorm 2015). The percentage of amylose in starch indicates the rice's cooking properties so identification of amylose content will bring to plant breeders the information of rice itself, such as waxy (0%–5%), very low (5%–12%), low (12%–20%), intermediate (20%–25%), and high (25%–33%). As can be seen from previous reviews, modern breeding techniques have advantages that help plant breeders to overcome their problems. Recent studies have investigated the genetic diversity of cultivated rice varieties in Vietnam, such as upland rice (Nguyen et al. 2012), which

are mostly coming from Northern Delta Vietnam, or local-colored rice landraces, which are collected in the central and north of Vietnam 7 (Hue et al. 2018)). However, there is a lack of studies on the genetic diversity of local rice germplasms in the Mekong River Delta in the south of Vietnam, especially in the coastal region. By using a combined morphological-molecular analysis, along with statistical approaches, this study can perform a full analysis of the genetic variability of 41 local rice varieties in Vietnam's Mekong River delta. The study will contribute the useful information.

2. Materials and Methods

For this study, we used 41 rice varieties collected from local areas in Vietnam's Mekong River Delta (Table 1). All experiments were conducted in three provinces (Ca Mau, Ben Tre, Bac Lieu, and Kien Giang) in the Mekong River Delta. Forty-one rice varieties were seeded directly paddy filed conditions with three replicates. Eleven morphological were evaluated following the standard system of IRRI (IRRI, 2013), and the young leaves at seedling stage were collected for DNA extraction.

The amylose content in rice varieties was estimated using the iodine colorimetric method (Khoomtong and Noomhorm 2015). The DNAs from young leaves of 41 rice varieties were extracted using the CTAB buffer (Kawata et al. 2003). For genetic diversity assessment, a panel of 50 standard SSR markers (Suppl. Table 1) were taken from rice genome databases (Youens-Clark et al. 2011). The obtained results were scored in binary data and determined by the distance matrix method to connect the relationships among individuals. The PIC (polymorphism information content) of the SSR markers was the mean of the PIC of each allele (Kempf et al. 2016) and the formula for co-dominant markers (Zargar et al. 2016) was used to calculate the PIC value as $PIC_i = 2f_i(1-f_i)$, where PIC_i is the polymorphism information content of allele i , f_i is the frequency of the amplified fragments, and $1-f_i$ is the frequency of non-amplified fragments. Moreover, statistical approaches such as PCA, ANOVA, and heatmap3 as an advanced heatmap for clustering (Zhao et al. 2014) were performed via RStudio version 3.5.1 (RStudio 2015). In addition, population structures of interested accessions were measured and visualized using Structure version 2.3.4 (Pritchard et al. 2000).

3. Results and Discussion

3.1. Diversity of phenotypic traits

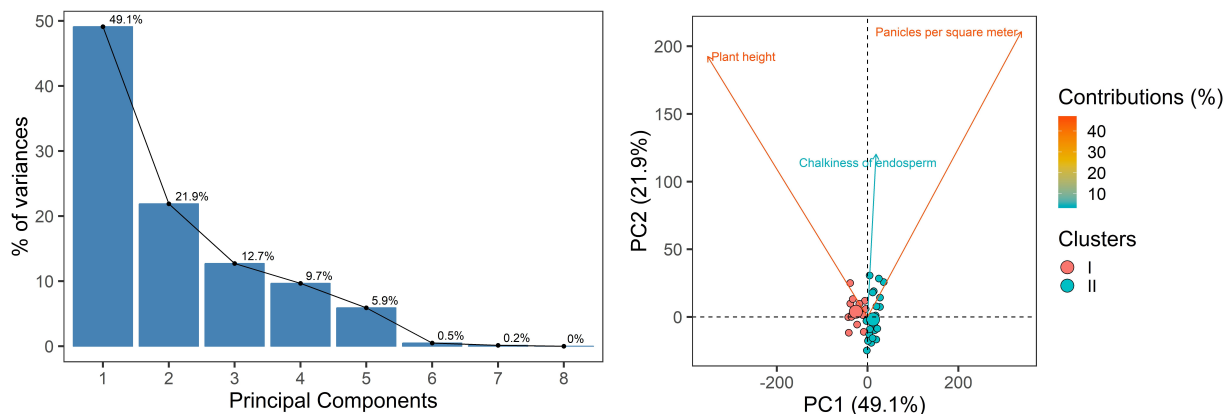
Using principal component analysis (Figure 1a and 1b), we observed that the first axis mainly accounted for the variation in the germplasm (49.1%), followed by the second axis (21.9%). This result indicated that plant height (group I) and panicles per square meter (group II) within the axes exhibit great influence on the phenotype of the population. Moreover, the first five components accounted for 99.30%

of the total variation, with components PCA3 (Chalkiness of endosperm), PCA4 (leaf length), and PCA5 (amylose content) contributing 12.7%, 9.7%, and 5.90%, respectively.

In Figure 1c, there are 13 rice varieties in group I, including Mong chim den, Nang cum 1, Tet ran, Ba bui 2, Nang quot bien, Bo liep 2, Thom lun mua, Nang thom, Doc Phung, Tai nguyen CL, Trang bo cau, Thom man, and Tai

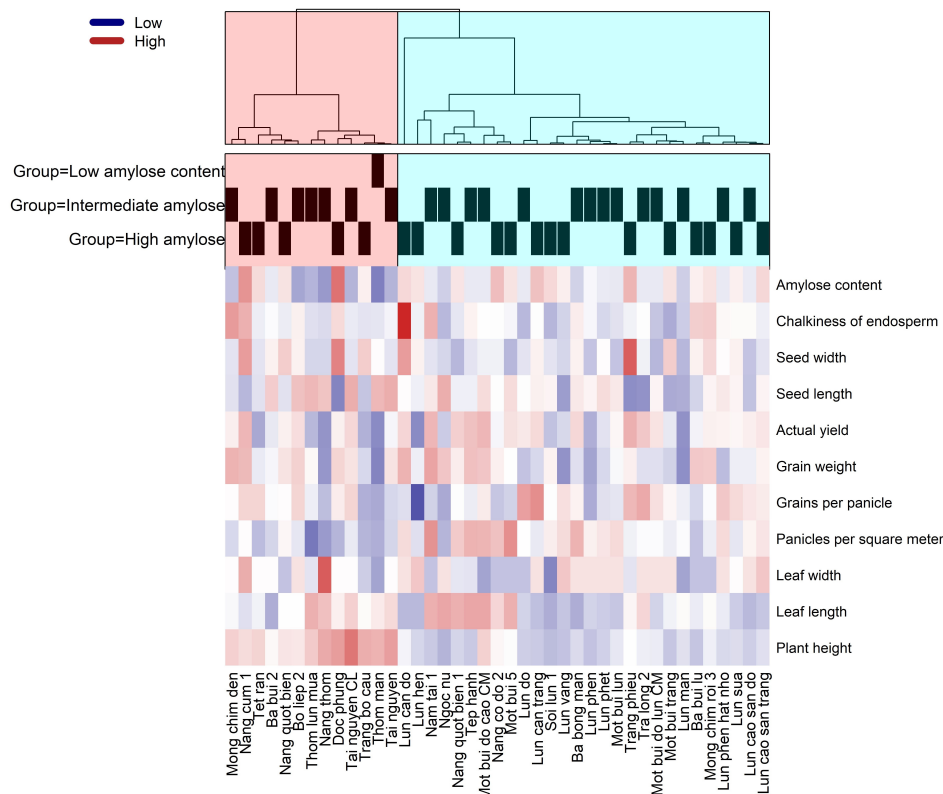
TABLE 1 List of 41 rice germplasm including geographic location.

Code	Variety	Location
V1	Doc phung	Ben Tre
V2	Lun can do	Ca Mau
V3	Lun can trang	Ca Mau
V4	Bo liep 2	Ca Mau
V5	Mot bui do lun CM	Ca Mau
V6	Mot bui lun	Ca Mau
V7	Ba bong man	Ca Mau
V8	Lun cao san do	Ca Mau
V9	Lun cao san trang	Ca Mau
V10	Tai nguyen CL	Ca Mau
V11	Nang co do 2	Ca Mau
V12	Tra long 2	Ca Mau
V13	Ba bui 2	Ca Mau
V14	Tep hanh	Ca Mau
V15	Nam tai 1	Ca Mau
V16	Mot bui 5	Kien Giang
V17	Mot bui do cao CM	Ca Mau
V18	Nang cum 1	Ca Mau
V19	Lun phen hat nho	Ca Mau
V20	Lun phet	Ca Mau
V21	Thom lun mua	Bac Lieu
V22	Lun phen	Ca Mau
V23	Lun hen	Ca Mau
V24	Lun vang	Ca Mau
V25	Lun sua	Ca Mau
V26	Mot bui trang	Ca Mau
V27	Mong chim den	Ca Mau
V28	Mong chim roi 3	Ca Mau
V29	Ba bui lun	Ca Mau
V30	Tai nguyen	Bac Lieu
V31	Trang bo cau	Bac Lieu
V32	Tet ran	Bac Lieu
V33	Lun man	Kien Giang
V34	Soi lun 1	Ca Mau
V35	Ngoc nu	Ca Mau
V36	Nang thom	Kien Giang
V37	Thom man	Kien Giang
V38	Nang quot bien 1	Ca Mau
V39	Trang phieu	Ca Mau
V40	Lun do	Ca Mau
V41	Nang quot bien	Ca Mau



(a)

(b)



(c)

FIGURE 1 Cluster analysis of 41 rice germplasms. (a) Proportion of information retained by each principal component (PCA1. Panicles per square meter; PCA2. Plant height; PCA3. Chalkiness of endosperm; PCA4. Leaf length; PCA5. Amylose content; PCA6. Grain per panicle; PCA7. Actual yield; PCA8. Grain weight) (b) Principal component analysis of 41 rice germplasm based on morphological traits. (c) Heatmap analysis of the distribution of 11 quantitative morphological traits among 2 groups.

nguyen. Group II has 28 rice varieties: Lun can do, Lun hen, Nam tai 1, Ngoc nu, Nang quot bien 1, Tep hanh, Mot bui do cao CM, Nang co do 2, Mot bui 5, Lun do, Lun cao san trang, Soi lun 1, Lun vang, Ba bong man, Lun phen, Lun phet, Mot bui lun, Trang phieu, Tra long 2, Mot bui do lun CM, Mot bui trang, Lun man, Ba bui lun, Mong chim roi 3, Lun phen hat nho, Lun sua, Lun cao san do, and Lun cao san trang. In group I, Thom man had the lowest amylose content. The heatmap obtained from cluster analysis (Figure 1c) showed that group I had 13 rice varieties which represented the plant height group, while group II tended to express mainly on panicles per square

meter. Thus, based on phenotypic traits, the analysis have been widely applied for studied genetic diversity (Liu et al. 2015; Veasey et al. 2008), germplasm GenBank material can be identified (Islam et al. 2018). Finding from this study have the potential for future use costal of Vietnam’s Mekong delta with higher yield rice molecular breeding program.

The frequency distribution of 11 quantitative morphological traits were divided among two groups obtained from cluster analysis (Figure 2). Only three agromorphological traits (actual yield, panicles per square meter, and plant height) demonstrated significant variation

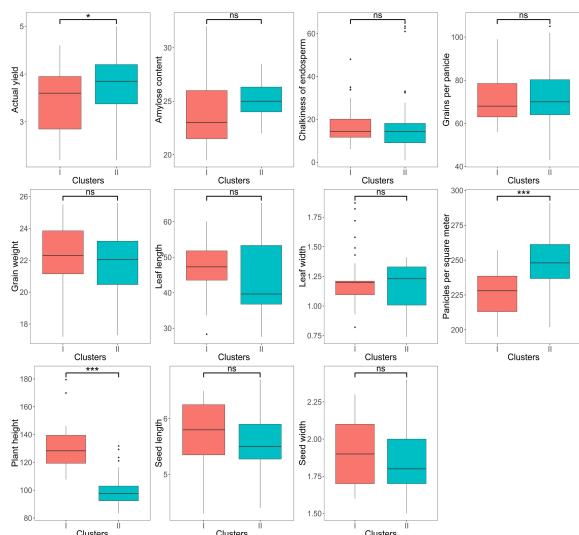


FIGURE 2 Distribution of 11 quantitative morphological traits in two groups derived from cluster analysis.

among 41 rice germplasms. Although group I showed higher plant height distribution, group II showed a higher number of both panicles per square meter and actual yield. Yield is typically the most important trait for breeding purposes. Overall, actual yield, panicles per square meter, and plant height were 2.43–4.57 ton/ha, 203–280 m², 87.47–156 cm, respectively. Thus, this finding will help for improvement of rice yield in Vietnam's Mekong delta based on group of rice varieties with high in actual yield, panicles per square meter. In addition, low and intermediate amylose content is one of the characters that rice breeder select for the demand of the market as high quality of rice (Islam et al. 2018)

3.2. SSR polymorphism among 41 local rice varieties

To obtain highly informative SSR markers, 50 SSR primers were tested in this study. Following ΔK , the PIC value was confirmed based on a variation between 0.0 and 0.5, and above 0.5 were effective and informative markers. In this study, PIC values absorbed from 50 SSR analysis mostly ranged from 0.00 to 0.49. The effective and informative primers were RM211, RM6329, RM127, RM253, RM24330, RM271, and RM286. The lowest PIC values were scored by 16 SSR primers (RM431, RM283, RM452, RM338, RM161, RM507, RM334, RM178, RM17749, RM162, RM455, RM125, RM11, RM536, RM144, and RM277).

According to the results, as shown in Figure 3, the similarity coefficient Nei and Li (Nei 1973) of relationships in this study ranged from 0.78 to 0.96. Besides, 41 rice germplasms were officially divided into two main clusters. The first cluster contained 39 germplasms (Doc phung, Lun can do, Lun can trang, Bo liep 2, Mot bui do lun CM, Mot bui lun, Ba bong man, Lun cao san do, Lun cao san trang, Tai nguyen CL, Nang co do 2, Tra long 2, Ba bui 2, Tep hanh, Nam tai 1, Mot bui 5, Mot bui do cao CM, Nang cum 1, Lun phen hat nho, Lun phet, Thom lun mua,

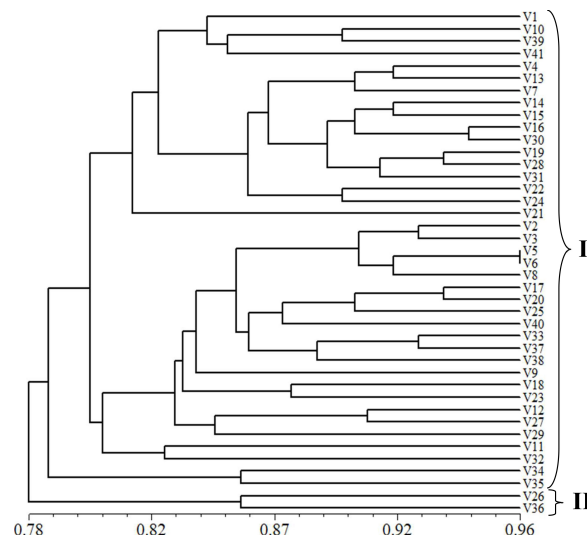


FIGURE 3 The dendrogram illustrates the similarity coefficient Nei and Li of 41 rice germplasms obtained from SSR markers based on UPGMA method.

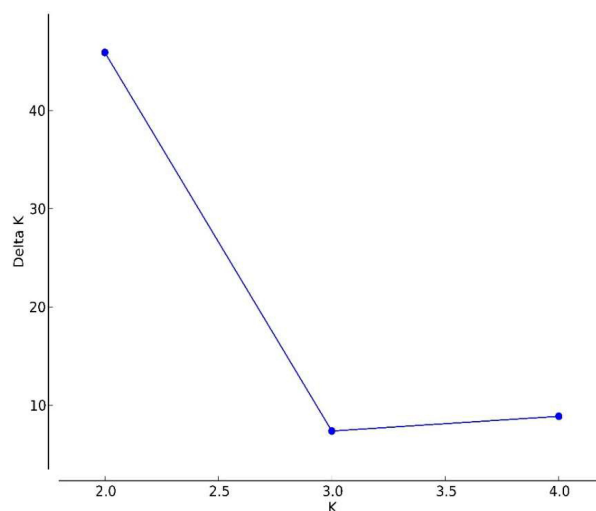


FIGURE 4 Estimation of population using LnP(D) derived ΔK for K from 1 to 4.

Lun phen, Lun hen, Lun vang, Lun sua, Mong chim den, Mong chim roi 3, Ba bui lun, Tai nguyen, Trang bo cau, Tet ran, Lun man, Soi lun 1, Ngoc nu, Thom man, Nang quot bien 1, Trang phieu, Lun do, and Nang quot bien), and the second cluster had two varieties (Mot bui trang and Nang Thom). In cluster I, only Mot bui do lun CM and Mot bui lun shared a similar coefficient relationship of 96%, which indicates that these two varieties are closely related to each other. They may be shared the same ancestor with different name because its came from Ca Mau province (Table 1). Regarding to Jaccard's coefficient (SJ), the Nei and Li coefficient varies only by the double weight given to the frequency of bands in each of the two studied genotypes (Duarte et al. 1999; Mohammadi and Prasanna 2003), so the Nei and Li coefficient is better matched to the type of analysis described in this study. Since phenotypic traits are based on genotypic traits, this analysis is highly desir-

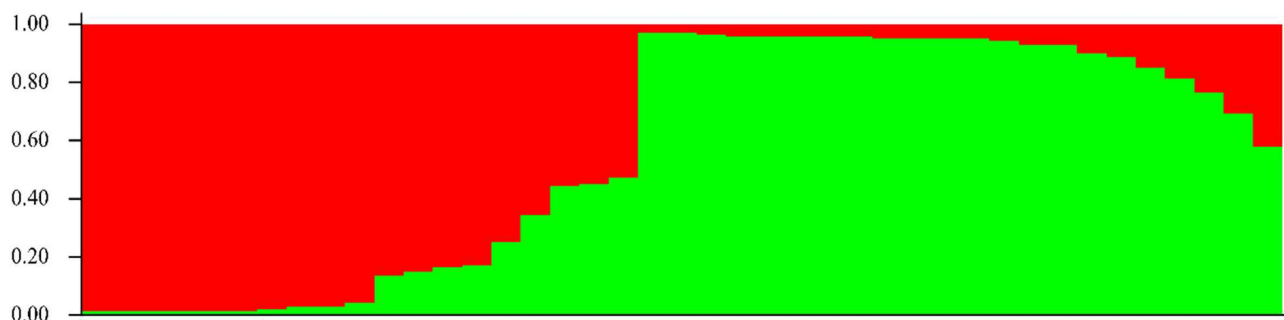


FIGURE 5 Population structure of 41 genotypes using 50 SSR markers was determined using STRUCTURE 2.3.4 software.

able, valuable, and used for selection.

3.3. Population structure model-based approach

A model-based structure analysis was also carried out to observe the number of populations that may be generated from 41 genotypes using 50 SSR markers (Figure 5). The $\ln P(D)$ as well as Evanno's ΔK values identified two genetically distinct groups ($K = 2$) (Figure 4). Structure simulations were carried out via varying K from 1 to 4 with 10 runs for each K using all 41 genotypes and obtaining the highest likelihood at $K = 2$. Therefore, two populations were obtained with slight mixing in some of the genotypes, as represented in Figure 4. Additionally, this population structure analysis (Figure 5) confirmed the grouping of genotypes, as observed by UPGMA cluster analysis (Figure 3).

Thus, the effectiveness of any crop breeding program depends on the amount of genetic diversity within the desired improvement characteristics and the degree to which these characteristics are inherited (Adjebeng-Danquah et al. 2020; Patel et al. 2014; Ravi et al. 2003). From this point of view, the study of genetic variation is very useful for the pre-breeding program in order to choose the parental line with different desirable characteristics. Yet, the application of molecular marker has been widely used to study genetic diversity, in which these molecular markers were linked to genomic region that presented agronomic traits, but these molecular characteristics may not impact by environments. Structure analysis of agromorphological characteristics divided all varieties into two classes in this study, such as amylose content and plant height (Figure 5). There were major variations in three agromorphological characteristics, in particular panicles per square meter characteristics. This trait is one of the production components, so these varieties with high number of panicles per square meter may be used for high production breeding program. From molecular analysis using 50 SSR markers, two main clusters were finally created. In this study, at $K = 2$, all 41 varieties divided into two populations, indicating genetic differentiation in the overall varieties. The findings of this study provide a useful information for further work in the effort to increase the yields of 41 local rice varieties in Vietnam's Mekong River Delta.

4. Conclusions

Variation of agro-morphological characteristics was informative in order to distinguish among the population, in particular the number of square meter panicles. This is valuable trait for contributing to rice yield that have been recorded in 28 varieties, especially these two varieties Mot bui bo cao CM, and Mot bui 5 that they were higher in number of square meter panicles and intermediate amylose content. In addition, the use of the regular panel SSR marker has been given with adequate depth of detail and may assess the genetic diversity of the Mekong Delta population. It is very important of Can Tho University rice varieties collection with rich genetic resource that will help to improve in yield for Mekong Delta rice production.

Acknowledgments

This study was funded in part by the Can Tho University Improvement Project VN14-P6 supported by a Japanese ODA loan. Thank for ProofreadingServices.com for proofreading and editing for this manuscript.

Authors' contributions

THP, HK designed the study. THP, VQG, NVM carried out the laboratory work. VQG, HK analyzed the data. HK, VQG wrote the manuscript. All authors read and approved the final version of the manuscript

Competing interests

The author declare that they have no competing interest.

References

Adjebeng-Danquah J, Manu-Aduening J, Asante IK, Agyare RY, Gracen V, Offei SK. 2020. Genetic diversity and population structure analysis of Ghanaian and exotic cassava accessions using simple sequence repeat (SSR) markers. *Heliyon*. 6(1). doi:10.1016/j.heliyon.2019.e03154.

- Botstein D, White RL, Skolnick M, Davis RW. 2014. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet.* 32(3):314–331. doi:10.17348/era.9.0.151-162.
- Duarte JM, Dos Santos JB, Melo LC. 1999. Comparison of similarity coefficients based on RAPD markers in the common bean. *Genet Mol Biol.* 22(3):427–432. doi:10.1590/S1415-47571999000300024.
- Hue HT, Nghia LT, Minh HT, Anh LH, Trang LTT, Khanh TD. 2018. Evaluation of Genetic Diversity of Local-Colored Rice Landraces Using SSR Markers. *Intl Lett Nat Sci.* 67:24–34. doi:10.18052/www.scipress.com/ilns.67.24.
- Islam MZ, Khalequzzaman M, Bashar MK, Ivy NA, Mian MA, Pittendrih BR, Haque MM, Ali MP. 2018. Variability Assessment of Aromatic Rice Germplasm by Pheno-Genomic traits and Population Structure Analysis. *Sci Rep.* 8(1). doi:10.1038/s41598-018-28001-z.
- Kawata M, Matsumura Y, Oikawa T, Kimizu M, Fukumoto F, Kuroda S. 2003. Analysis of DNA extraction buffer components from plant tissue by polymerase chain reaction. *Anal Biochem.* 318(2):314–317. doi:10.1016/S0003-2697(03)00240-9.
- Kempf K, Mora-Ortiz M, Smith LM, Kölliker R, Skøt L. 2016. Characterization of novel SSR markers in diverse sainfoin (*Onobrychis viciifolia*) germplasm. *BMC Genet.* 17(1). doi:10.1186/s12863-016-0431-0.
- Keskinen M, Chinvano S, Kumm M, Nuorteva P, Snidvongs A, Varis O, Västilä K. 2010. Climate change and water resources in the lower Mekong River Basin: Putting adaptation into the context. *J Water Clim Change.* 1(2):103–117. doi:10.2166/wcc.2010.009.
- Khoontong A, Noomhorm A. 2015. Development of a Simple Portable Amylose Content Meter for Rapid Determination of Amylose Content in Milled Rice. *Food Bioprocess Technol.* 8(9):1938–1946. doi:10.1007/s11947-015-1550-8.
- Litrico I, Violle C. 2015. Diversity in Plant Breeding: A New Conceptual Framework. doi:10.1016/j.tplants.2015.07.007.
- Liu W, Shahid MQ, Bai L, Lu Z, Chen Y, Jiang L, Diao M, Liu X, Lu Y. 2015. Evaluation of genetic diversity and development of a core collection of wild rice (*Oryza rufipogon* Griff.) populations in China. *PLoS ONE.* 10(12). doi:10.1371/journal.pone.0145990.
- Mohammadi SA, Prasanna BM. 2003. Analysis of genetic diversity in crop plants - Salient statistical tools and considerations. doi:10.2135/cropsci2003.1235.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA.* 70(12 (I)):3321–3323. doi:10.1073/pnas.70.12.3321.
- Nguyen TT, Nguyen NM, Hoang LH, Furuya N, Tsuchiya K. 2012. Genetic diversity in Vietnamese upland rice germplasm revealed by SSR markers. *Journal of the Faculty of Agriculture, Kyushu University* 57(2):383–391. doi:10.5109/25195.
- Patel S, Ravikiran R, Chakraborty S, Macwana S, Sasidharan N, Trivedi R, Aher B. 2014. Genetic diversity analysis of colored and white rice genotypes using Microsatellite (SSR) and Insertion-Deletion (INDEL) markers. *Emir J Food Agric.* 26(6):497–507. doi:10.9755/ejfa.v26i6.15722.
- Phung NTP, Mai CD, Mournet P, Frouin J, Droc G, Ta NK, Jouannic S, Lê LT, Do VN, Gantet P, Courtois B. 2014. Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. *BMC Plant Biol.* 14(1). doi:10.1186/s12870-014-0371-7.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics.* 155(2):945–959. doi:10.1093/genetics/155.2.945.
- Ravi M, Geethanjali S, Sameeyafarheen F, Maheswaran M. 2003. Molecular Marker based Genetic Diversity Analysis in Rice (*Oryza sativa* L.) using RAPD and SSR markers. *Euphytica.* 133(2):243–252. doi:10.1023/A:1025513111279.
- Roy SC, Sharma BD. 2014. Assessment of genetic diversity in rice [*Oryza sativa* L.] germplasm based on agro-morphology traits and zinc-iron content for crop improvement. *Physiol Mol Biol Plants.* 20(2):209–224. doi:10.1007/s12298-014-0221-y.
- RStudio. 2015. RStudio: Integrated Development for R.
- Thi H, Vu H, Thi H, Nguyen T, Khanh T, Huu T, Nakamura C, Khuat T, Chiharu. 2015. Biotechnology and Biotechnological Equipment Genetic diversity of Vietnamese lowland rice germplasms as revealed by SSR markers in relation to seedling vigour under submergence. *Biotechnol Biotechnol Equip.* 30. doi:10.1080/13102818.2015.1085330.
- Veasey EA, Da Silva EF, Schammas EA, Oliveira GCX, Ando A. 2008. Morphoagronomic genetic diversity in American wild rice species. *Braz Arch Biol Technol.* 51(1):95–104. doi:10.1590/s1516-89132008000100012.
- Youens-Clark K, Buckler E, Casstevens T, Chen C, DeClerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, et al. 2011. Gramene database in 2010: Updates and extensions. *Nucleic Acids Res.* 39(SUPPL. 1):D1085–D1094. doi:10.1093/nar/gkq1148.
- Yu B, Zhu T, Breisinger C, Hai NM. 2010. Impacts of Climate Change on Agriculture and Policy Options for Adaptation The Case of Vietnam. *IFPRI Discussion (August):*1–23.
- Zargar SM, Farhat S, Mahajan R, Bhakhri A, Sharma A. 2016. Unraveling the efficiency of RAPD and SSR markers in diversity analysis and population structure estimation in common bean. *Saudi J Biol Sci.* 23(1):139–149. doi:10.1016/j.sjbs.2014.11.011.
- Zhao S, Guo Y, Sheng Q, Shyr Y. 2014. Advanced Heat Map and Clustering Analysis Using Heatmap3. *BioMed Res Int.* 2014. doi:10.1155/2014/986048.