~~Correspondence~~ Association between phenotypic and genotypic distance matrices of advanced breeding ~~lines~~ lines of lowland rice from diverse parents ~~using Mantel test.~~

Abstract

Evaluation of advanced breeding lines developed from a set of diverse parents involved in different cross combination facilitates selection of promising genotypes, resource management and planning further breeding activities. ~~A set of~~ Totally ~~22~~, 22 advanced breeding lines of lowland rice in F_7 were evaluated based on agro-morphologically and ~~genotyped~~ genotypically with molecular markers. Here we report correspondence between different distance matrices subjected to correlation based on Mantel test. ~~An insignificant~~ Non significant (p-value > 0.05) correlation was observed between the matrices due to causal factors such as data type, markers used and genetic properties of lines as explained in the text. However, certain promising lines were selected from phenotypic performance as well as molecular fingerprints and grain quality evaluation relative to checks (CAUR1 and Shasarang) with specific context to north eastern hill region. ???

Keywords: Agro-morphology, molecular markers, Mantel test, distance matrix, rice

Introduction

Diversity and variation at species and inter/intra population levels are the inevitable yet explorable elements on which plant breeders primarily depend. They act as tools in designing a breeding program and selection is practiced by identifying diverse lines for many population and line improvement activities such as heterotic grouping (Silva et al. 2020), creation of general combining ability based synthetic populations (Goldringer et al., 2017), donor selections (Allier et al., 2020), narrowing down to best lines as promising varieties and creation of mini cores (Zhang et al., 2012) thus enhancing resource management and breeding activities through evaluation of population structure. Genetic variation is studied from several qualitative and quantitative traits detected from expressed phenotypic variation categorised as morphological technique, in-silico data (Carvalho et al., 2019) and further, from the biochemical and molecular techniques at protein level and DNA sequence level using breeder friendly molecular markers, respectively. Different categorical distance matrices generated from a set of lines under study are compared and related by testing linear independence of its elements by Mantel test (Mantel, 1967) when distance based hypothesis is formulated thus measuring the correlation between the two. The test was first pioneered to compare phenetic distances among local populations to geographic distances (Sokal, 1979). Correspondence level between morphological and molecular dissimilarity distance matrices in crop improvement programs were tested by many workers using Mantel test (Sun et al., 2011; Chung et al., 2011; Zhang et al., 2012; Ehinger et al., 2014; Kaviriri et al., 2020). Chung et al. (2011) reported less inflated type I error rate for Mantel test when comparing with multivariate distance matrix regression.

Advanced breeding lines in self pollinated crops like rice are the genotype material of later generations developed from crossing a set of favourable parental elite lines or landraces. The progeny in each generation undergoes selection for favourable trait expression as per breeding objective and harbour promising better genotype with respect to parents. Such advanced lines are well structured, share ancestry and capable of replacing parental lines in terms of farmer adoption owing to accumulation of favourable alleles after careful within line selections by breeders. However, agronomical and molecular diversity evaluation of these lines is requisite for further utilization as potential genotypes as starting material. ~~We report here~~

Therefore, the present study was to ~~the~~evaluate the relationship between distance matrices and diversity generated based on phonotypic ~~data~~ and ~~molecular~~ genotypic data ~~on advanced breeding lines with specific reference to Mantel test results.~~

Comment [G1]: Please add the scientific name of the crop

Comment [G2]: General comments
Abstract should consists the following points in a very short and precise way:
1-short definition of the crop
2-specific objective of the study
3-materials and methods (study area, design, number of treatments)
4-result obtained
5-conclussion and recommendation
✓ Please incorporate these all to make full of abstract

Comment [G3]: Make alphabetical order

Comment [G4]: First introduce the crop in very short way

Materials and Methods

Description of study area : the study was conducted at College of Post Graduate studies, CAU (I), Umiam, Meghalaya under rain fed lowland acidic and low available soil phosphorus (P) soil conditions without any management practices during 2017-18. Cropping season.

~~Plant materials~~ Breeding materials, experimental design and management:

~~The genotypes consisted of~~ Twenty two ~~22~~(22) advanced breeding lines of rice in their F_7 generation and 2 checks CAUR1 and Shasarang (Table 1). The lines were derived from 8 non aromatic parents consisting of elite varieties and local landraces. The material was sown in replicated block design ~~in experimental farm of College of Post Graduate studies, CAU (I), Umiam, Meghalaya under rainfed lowland acidic and low available soil phosphorus (P) soil conditions without any management practices during 2017-18.~~

Table I. List of advanced breeding lines against their parentage used in the study

Sl.No.	Advanced breeding line
1	CAUS101
2	CAUS102
3	CAUS103
4	CAUS104
5	CAUS105
6	CAUS106
7	CAUS107
8	CAUS108
9	CAUS109
10	CAUS110
11	CAUS111
12	CAUS112
13	CAUS113
14	CAUS114
15	CAUS115
16	CAUS116
17	CAUS117
18	CAUS118
19	CAUS119
20	CAUS120
21	CAUS121
22	CAUS122

Data collected

Phenotypic and genotypic data

Agro-morphological data was recorded for 16 qualitative and 18 quantitative traits on the lines including DUS traits. Further, the lines were genotyped with a panel of 30 random SSR markers (McCouch *et al.* 2002; http://www.gramene.org) and 24 random SNPs.

Comment [G5]: Please don't use abbreviation directly since it is difficult for the reader to capture easily

Comment [G6]: Incorporate the annual min and max temperature, soil type and relative humidity of the study area

Comment [G7]: Please Identify as standard and/or local check if used both don't say check

Comment [G8]: State the number of replication and types of design used during conducting the trial.
- ✓The trial conducted where? At laboratory or at field?
- ✓ what are spacing, seed rate , sowing method ,rows number and plot size you are used during conducting the trial
- ✓Please revise the methodology again especially list the procedure you are followed at laboratory in short and precise

Comment [G9]: What are those qualitative and quantitative data? Please list those data

Comment [G10]: It is better if omitted using abbreviation to easily grasp the idea by reader

Comment [G11]:

Comment [G12]:

Statistical analysis

Trait means, standard deviation, CV, maximum and minimum values as well as ANOVA were evaluated using MS-Excel 2016. Euclidean distance between lines based on phenotypic and molecular data were measured in R software v4.0.3 using package “stats”. Mantel test for distance matrices was performed with 1000 permutations using “ade4” package of R software v4.0.3 and GenAlEx 6.5 add-in of MS-Excel.

Comment [G13]: You evaluated these Statistical parameters by using MS-Excel 2016,R software and others but no result indicated in result and discussion part. For instance where is ANOVA?

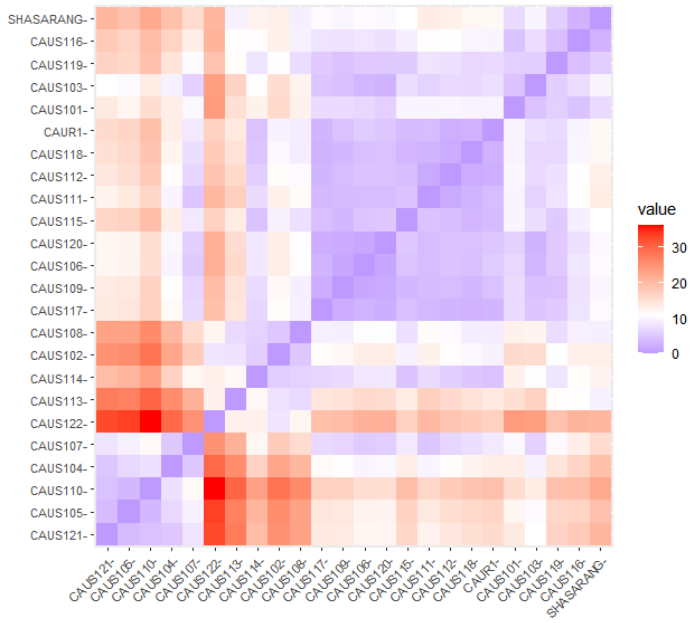
Results and Discussions

Dissimilarity distance based clustering grouped the advanced breeding lines into two and three major clusters based on phenotypic and genotypic data, respectively based on phenotypic and genotypic distance matrix (Fig. 1 and 2). One line was found to be an outlier in the clustering approach based on molecular distance matrix while no outliers were observed in grouping based on phenotypic data (data not shown). Molecular data based clustering grouped certain breeding lines with their either of parents, however, the same pattern was not observed in phenotypic data based clustering as phenotyping for quantitative traits involves environmental effects which could be corrected by multi location evaluation considering genotype-environment interaction. Euclidean distance evaluated based on phenotypic data ranged from 1.44 to 36.17 while the range of 1.03 to 7.34 was observed for genotypic data. ~~To test the association between phenotype and genotype of breeding lines evaluated from distance matrices, we used Mantel test.~~ The test result used by Mantel is shown in Fig. 2. There was ~~insignificant~~ negative and non significant correlation between dissimilarity matrices of genotypic and phenotypic distances with r(AB)= -0.177 (p-value > 0.05). This suggests ~~that molecular and phenotypic categorization are~~those molecular and phenotypic categorizations are not identical as observed in clustering dissimilarities. The scatterplot revealed weak R^2 value of 0.03 from regression of matrices.

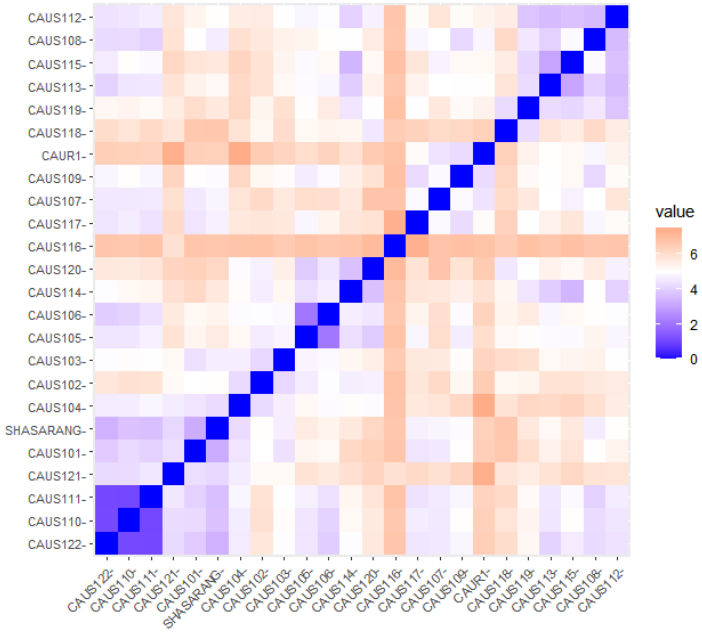
Comment [G14]: Which line? Point out. What is the secret behind?

Comment [G15]: Why

Comment [G16]: Please support your finding with other finding. What are your references ?

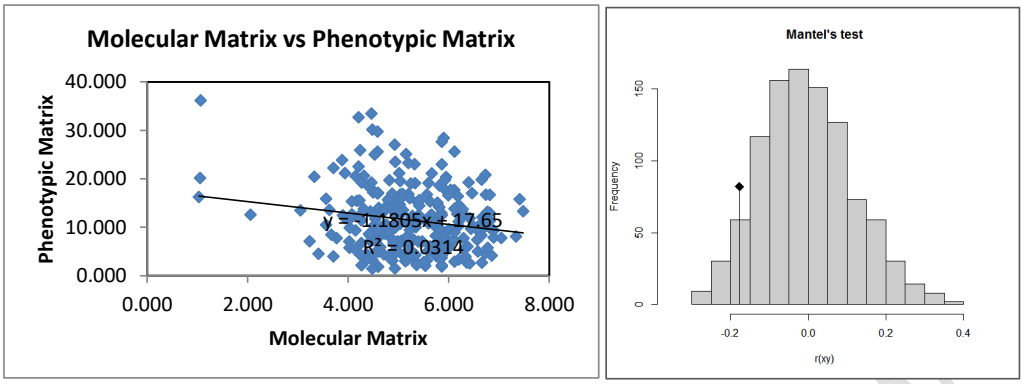


a) Phenotypic distance matrix



b) Genotypic distance matrix

Fig. 1. Euclidean distance dissimilarity matrix for a) phenotypic and b) genotypic data.



a) Scatter plot b) Mantel test histogram

Fig. 2. Representation of correlation analysis between dissimilarity matrices based on phenotypic data and genotypic molecular data. Point corresponds to correlation value.

Several causes contribute to lack of significant correspondence between matrices. Firstly, the nature of data is a major cause. Phenotype is a combination of expression attributed to genotype and environment under the multiple gene control while the molecular data was collected based on random SSRs and SNPs available from non coding regions of DNA. Association of two such distance matrices would be insignificant. Secondly, more number of markers should be included for closely related well structured study material used here to enhance probability of finding linked markers to the quantitative traits which would possibly provide significant Mantel test result or atleast less increase in p-value from 0.05. Thirdly, the advanced breeding lines used here have a locally popular and adapted maternal parent common in most of the cross combinations, causing observation of only 2 major groups in phenotype based clustering approach. Therefore, the genotypes similar to used here are to be evaluated for more phenotypic traits to identify more diversity among the lines. However, certain promising lines such as CAUS103, CAUS104, CAUS105 and CAUS107 were identified with specific reference to molecular fingerprinting, higher phenotypic expression than checks, performance under biotic stress and better grain quality relative to consumer and geographic context of north eastern hill region.

Comment [G17]: Your results and discussions:
✓Are not elaborated very well
✓Are not supported by other finding

Conclusion and recommendation

~~The present report focused on correspondence between agro-morphological and molecular data based dissimilarity distance matrices evaluated on a set of advanced breeding lines of lowland rice. Significant Mantel test correlation if observed would be surprising and meant that data was incorrect as limited number of molecular markers were used on the lines that are well structured and related. However, that is not true~~ . The present study was focused on association between agro-morphological and molecular data based dissimilarity distance matrices evaluated on twenty two advanced breeding lines of lowland rice. The study revealed that, non significant differences were observed based on the association between agro-morphological and molecular of the evaluated materials due to the limitation number of molecular markers used at the time of evaluation and other factors. However some promising lines such as CAUS103, CAUS104, CAUS105 and CAUS107 were identified. Generally, if the study was conducted with the full methodology, there may be strong correlation among the studied materials in relation to agro-morphological and molecular data and provide basic information for further breeding activities for crop improvement.

References

Allier A, Teyssèdre S, Lehermeier C, Moreau L, Charcosset A. Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genomics. 2020; 21: 349. https://doi.org/10.1186/s12864-020-6756-0

Carvalho YGS, Vitorino LC, de Souza UJB, Bessa LA. Recent trends in research on the genetic diversity of plants: implications for conservation. Diversity. 2019; 11:62

Chung D, Zhang Q, Kraja AT, Borecki IB. Province MA. Distance-based phenotypic association analysis of DNA sequence data. BMC Proceedings. 2011; 5(Suppl 9):S54.

Ehinger M, Mohr TJ, Starcevich JB, Sachs JL, Porter SS, Simms EL. Specialization-generalization trade-off in a Bradyrhizobium symbiosis with wild legume hosts. BMC Ecology. 2014; 14:8.

Goldringer I, Serpolay E, Rey F, Costanzo A. Varieties and populations for on-farm participatory plant breeding. DIVERSIFOOD Innovation Factsheet. 2017; http://www.diversifood.eu/wpcontent/uploads/2018/06/Diversifood_innovation_factsheet2_VarietiesPopulations.pdf (accessed on 2 January 2020).

Kaviriri DK, Zhang Q, Zhang X, Jiang L, Zhang J, Wang J, Khasa DP, You X, Zhao X. Phenotypic variability and genetic diversity in a *Pinus koraiensis* clonal trial in northeastern China. Genes. 2020; 11:673.

Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Research. 1967; 27:209–220.

Silva KJ, Pastina MM, Guimarães CT, Magalhães JV, Pimentel LD, Schaffert RE, Pinto MO, Souza VF, Bernardino KC, Silva MJ, Borém AM. Genetic diversity and heterotic grouping of sorghum lines using SNP markers. Scientia Agricola. 2020; 78(6), e20200039. Epub October 16, 2020.https://dx.doi.org/10.1590/1678-992x-2020-0039

Sokal RR. Testing statistical significance of geographic variation patterns. Systematic Zoology. 1979 28: 227–232.

Sun YV, Zhao W, Shedden KA, Kardia LA. Identification of genes associated with complex traits by testing the genetic dissimilarity between individuals. BMC Proceedings. 2011; 5(Suppl 9):S120

Ward JH. Hierarchical grouping to optimize an objective function. J. American Stat. Assoc. 1963; 58: 236-244.

Zhang Y, Zhang X, Zhuo C, Wang L, Wei W, Li D. Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. BMC Genetics. 2012; 13:102.

Comment [G18]:
- ✓Make sure that the references in the text should be exists in the references
- ✓Please follow the guide line and instructions of international journal of Environment and climate change