Oryza Vol. 55 (Special Issue) 2018 (18-23)
DOI 10.5958/2249-5266.2018.00002.4

Simultaneous mapping of multiple QTLs & complex traits in rice: Association mapping-indispensable for effective molecular breeding

SK Pradhan[1], E Pandit[1], SP Mohanty[1], RK Panda[3], D Panigrahi[3] and D Pani[2]

[1], ICAR-National Rice Research Institute, Cuttack, Odisha
[2]NBPGR Base Center, Cuttack, Odisha
[3]OUAT, Bhubaneswar, Odisha
**Corresponding author e-mail: pradhancrri@gmail.com*

ABSTRACT

Association mapping or linkage disequilibrium mapping is extensively used in dissecting the genetic nature of many human diseases and now extended to plant system for mapping the complex loci. Association Mapping is a method used to detect QTLs controlling traits based on correlating genotype with phenotype. The main advantage of association mapping is the use of natural variability, relatively less expensive and less time requiring for mapping. The lead lecture includes linkage disequilibrium (LD), its assessment, mapping and generalized scheme for association mapping involving basic steps in mapping a complex trait. The talk also includes the association mapping study on cold stress tolerance, high temperature stress tolerance, early seedling vigour, grain protein content and grain Fe and Zn content in rice performed at National Rice Research Institute, Cuttack, Odisha, India.

INTRODUCTION

Locating a gene on the chromosome is usually known as gene mapping or gene tagging or QTL mapping. Gene mapping is usually conducted by involving two types of mapping populations (1) employing the F_1 derived mapping populations commonly known as biparental mapping and (2) using the diverse germplasm lines from the natural populations or breeding lines called as linkage disequilibrium (LD) mapping or association mapping (AM). The genetic bases of crop plants are mostly determined through employing the biparental mapping procedure. However, association analysis is extensively used in dissecting the genetic nature of many human diseases. This approach has also been extended to plant system for mapping the complex loci. Association Mapping is a method used to detect QTLs controlling traits based on correlating genotype with phenotype (Neumann et al., 2011). Validation of previously reported QTLs and their locations on the chromosome are also done using association mapping. The basic principle of this mapping is on the existence of linkage disequilibrium (LD) which is the non-random association of alleles at different loci (Flint-Garcia et al., 2003). LD may continue generation after generation in a population under certain conditions.

The main advantage of association mapping is the use of natural variability available in the collected germplasms. This method allows utilization of conserved germplassm resources worldwide for various conserved traits. No need of development of expensive and much time requiring mapping populations. The mapping population may also include developed breeding lines, cultivars, landraces, wild rice representing a highly diverse panel for the studied trait. The studied panel population needs to be highly diverse for the maping trait to detect large number novel alleles controlling the trait. This association mapping also provides broader allele coverage as a high resolution gene mapping method as compared to the biparental mapping method. The wider coverage is possible here due to many recombination events from a large number of meiosis across the germplasm compared to few meiotic events in linkage mapping. However, biparental mapping should also be continued to detect major genes controlling traits in rice crop. Here, a comprehensive review on association mapping of complex traits in rice and more particularly work done on this aspect at National Rice Research Institute, Cuttack has been presented.

Linkage disequilibrium (LD)

Linkage disequilibrium is also known as allelic association, gametic phase disequilibrium and gametic disequilibrium. This may be described as nonrandom association of alleles at different loci. It is the correlation between polymorphisms obtained in a population due to mutation and recombination over years. However, in a random mating population with presence of independently inherited loci and in absence of selection, mutation, or migration, the loci will be in linkage equilibrium (LE). But, due to linkage, selection, genetic drift and admixture in a population, the loci will be in linkage disequilibrium. Linkage and LD are both different which is generally confused as similar phenomena. Linked loci refer to coinheritance of both the loci on a chromosome while LD gives correlation between allele in a population. For measuring the presence or absence Linkage disequilibrium or linkage equilibrium of two loci, commonly r^2 and D' are frequently used two statistics. Here frequency of each haplotype are used. The square of the correlation coefficient between the two loci is r^2 while D' is calculated from the haplotype frequency of the two loci. For example, a pair of loci with alleles A and a at locus one, and B and b at other locus, with allele frequencies of ρA, ρa, ρB, and ρb, respectively. The resulting haplotype frequencies are ρAB, ρAb, ρaB, and ρab. The basic component of all LD statistics is the difference between the observed and expected haplotype frequencies,

$$D_{ab} = (\rho AB - \rho A\rho B)$$

If two loci are in linkage equilibrium, then $D = 0$ and if the two loci are in linkage disequilibrium, then $D \neq 0$.

The two measure of correlation coefficient of two loci, r^2 is estimated using the following formula

$$r^2 = (D_{ab}^{\ 2}) / \rho A \ \rho a \ \rho B \ \rho b$$

Steps in association mapping

A brief outline of the association mapping procedure is described so that the method can be utilized for mapping of complex loci (Fig.1)

Step 1: Preparation of a panel population from screened materials for a trait from a natural population or germplasm collection or breeding lines and cultivars with wide genetic diversity;

Step 2: Phenotyping for the trait of interest using the panel population preferably across environments in replications

Step 3: Genotyping the panel population individuals with available molecular markers

Step 4: Estimation of the extent of LD in the panel population using the molecular marker data

Step 5: Population structure and kinship analysis of the panel population and

Step 6: On the basis of LD and population structure information, correlation of phenotyping and genotyping data were used to know marker association which is close to the targeted trait of interest using suitable statistical software. Further, the candidate gene may be cloned and annotated for its biological function.

Two models are usually used in association analysis. The models are general linear model (GLM)

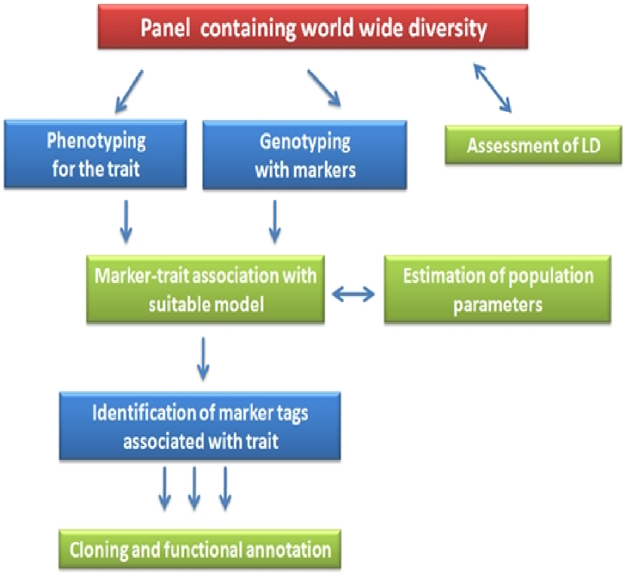

Fig. 1. A generalized scheme for association mapping involving basic steps in mapping a complex trait.

and the mixed linear model (MLM). In the GLM method, the marker-phenotype associations are detected by employing the population membership estimates of individual in the panel as covariates to control for population structure (Pritchard and Rosenberg, 1999), since population structure may cause spurious associations (Kang et al., 2008). MLM, additionally to the population structure, incorporates kinship in the association analysis allowing an improved control of type I and type II error rates over GLM due to relatedness and population structure (Yu et al., 2006). False discoveries rates are sometimes involved in association mappings. A false discovery refers to the situation when one concludes erroneously that a genomic region harbors a gene contributing to a quantitative trait (Sabatti, 2007). False discoveries are common in association studies due to the multiple hypotheses testing (Sabatti, 2007; Storey, 2003). For testing the association of false discoveries in association studies, many methods are employed. Bonferroni multiple correction test is a common conservative method to detect false association which provides a cut-off value based on the proposed threshold divided by number of tests. The Q value method proposed by Storey (2002) is also used in testing false associations in which the q-values are estimated from p-values.

Linkage diseqilibrium mapping in rice

LD mapping has been extensively used animal system while its use is increasing in crop plants. As this mapping approach can map many QTLs simultaneously with less resource and time using fixed lines and landraces with the existing natural variability for a trait, its use in rice mapping is rapidly increasing (Pradhan et al., 2016; Pandit et al., 2017). Many traits in rice have already been mapped in rice using association mapping. The yield trait is a complex in nature and mapped by Agrama et al., 2007; Zhao et al., 2011 and Huang et al., 2012. The seedling stage cold tolerance is governed by many genes and QTLs. The various QTLs controlling this complex trait have been mapped through association mapping by Zhang et al., 2012 and Pandit et al. 2017. Cold tolerance at germination and booting stages are mapped by Pan et al., 2015). High temperature stress tolerance association analysis was reported by Pradhan et al., 2016. QTLs controlling seedling vigour in rice following LD mapping was reported by Anandan et al., 2016. The drought tolerance in rice is a complex trait and the QTLs through LD mapping are reported by Muthukumar et al., 2015 and Swamy et al., 2017. The genetic nature of grain quality traits in rice are reported by Zhao et al., 2013 and for mineral element in grain by Huang et al., 2015. The marker-trait association analysis work conducted at national Rice research Institute, Cuttack are presented in brief.

Association mapping for seedling stage chilling stress in *indica* rice

The association mapping for seedling stage cold tolerance in rice by Pandit et al., 2017 detected multiple QTLs for seedling stage cold tolerance. They studied a panel population representing 304 sortlisted germplasm lines for seedling stage chilling tolerance in indica rice. Six phenotypic classes were observed in response to six low temperature regimes stress exposure under control phenotyping facility. The panel containing sixty six genotypes representing all phenotypic classes was used for genetic diversity, population structure and association mapping for the trait using 60 trait linked markers. They detected a moderate level of genetic diversity was detected in the panel population for the trait. Deviation of Hardy-Weinberg's expectation was detected in the in the studied population using Wright's F statistic. The panel showed 30 percent variation among population and 70% among individuals. The entire population was categorized into three sub-populations through STRUCTURE analysis and the analysis revealed tolerance for the trait had a common primary ancestor for each sub-population with few admix individuals. The panel population showed the presence of many QTLs in the individuals for the trait representing like genome-wide expression of the trait. The significantly associated markers like RM1347, RM328, RM152, RM341, RM50, RM2634, RM4112, RM5310, RM7179, RM3701, RM104, RM9, RM1211, RM245, RM3602, RM493, RM1335, RM282 and RM5704 were significantly associated at chilling stress of 80C to 40C for 7 to 21 days duration. Thus, the primers linked to the seedling stage cold tolerance QTLs namely *qCTS9, qCTS-2, qCTS6.1, qSCT2, qSCT11, qSCT1a, qCTS-3.1, qCTS11.1, qCTS12.1, qCTS-1b* and *CTB2* need to be pyramided for development of strongly chilling tolerant variety.

Molecular marker-trait association analysis for

high temperature stress tolerance in rice

The association analysis study in rice by Pradhan et al., 2016 reported multiple QTLs for high temperature stress tolerance in rice. They studied a set of breeding lines and landraces representing 240 germplasm lines. Based on spikelet fertility per cent under high temperature, tolerant genotypes were broadly classified into four classes. Genetic diversity indicated a moderate level of genetic base of the population for the trait studied. Presence of LD was confirmed with deviation of Hardy-Weinberg expectation in the population. The molecular variance revealed 25 percent variation between population, 61 percent among individuals and 14 percent within individuals in the set. The STRUCTURE analysis categorized the entire population into three sub-populations and revealed that each sub-population had a common primary ancestor. The composition of materials in the panel showed the presence of many QTLs representing the entire genome for the expression of tolerance. The strongly associated marker RM547 tagged with spikelet fertility under stress and the markers like RM228, RM205, RM247, RM242, INDEL3 and RM314 indirectly controlling the high temperature stress tolerance were detected through both mixed linear model and general linear model TASSEL analysis. These markers can be deployed as a resource for marker-assisted breeding program of high temperature stress tolerance.

Marker-trait association analysis for early seedling vigor in rice

The association analysis study for early seedling vigour by Anandanan et al., 2016 revealed 16 and 10 SSRs significant associations with the complex trait through general linear model (GLM) and mixed linear model (MLM) approaches, respectively. They studied for their morphological and physiological responses in the field under direct seeded aerobic situation. A panel population containing 96 rice which represents 629 rice genotypes were phenotyped and genotyped using 39 polymorphic SSRs markers. Structure analysis categorized the panel population into two distinct populations, whereas unrooted tree grouped the genotypes into three clusters. Both model based and structure based approach had clearly distinguished the early vigor genotypes from nonearly vigor genotypes. Marker alleles on chromosome 2 were associated with shoot dry weight on 28 days after sowing (DAS), vigor index on 14 and 28 DAS. Improvement in the rate of seedling growth will be useful for identifying rice genotypes for direct seeded conditions through marker-assisted selection. SSR genomic region RM230 and RM125 mapped on chromosome 8 and 7 were significantly associated with QTL controlling leaf width variation. The RM230 and RM3839 marker allele was significantly associated with root length. Marker alleles RM13 on chromosome 1 and RM16 on chromosome 3 were positively associated with shoot length. Allele RM224 located on chromosome 11 was significantly associated with shoot dry. RM341 coincided with QTLs for vigor index on chromosome 2 at 14 and 28 DAS and showed a stronger association detected by both GLM and MLM methods.

Association mapping for grain protein content in rice

The grain protein content association mapping was reported by Pradhan et al., 2018a. Protein malnutrition can easily be removed through biofortification. Relevant QTLs need to be identified for this.Hence, association mapping study was taken up for precision improvement of protein content through breeding program. A panel population possessing biofortified genotypes and shortlisted landraces were studied for genetic diversity, population structure and association mapping for grain protein content in rice. Wide genotypic variation for grain protein content was found across locations. AMOVA and structure analysis showed linkage disequilibrium for grain protein content with continuance of LD. The analysis showed 15% of the variation among populations, 73% among individuals while 12% within individuals in the panel population. In structure analysis at 2nd peak for ΔK at K=4 classified the panel population into four subgroups with each group clusterring with developing center and protein content as a subpopulation. The analysis also revealed a common primary ancestor for all the subpopulations. The marker-trait association study showed multiple QTLs controlling protein content in rice grain. Marker-trait association study using 92 molecular markers detected 4 strongly associated and consistent QTLs viz., RM6712, RM85, RM407 and RM300 in both the years for grain protein content in rice. Two new QTLs namely *qPC2.2* and *qPC3.2* were detected for controlling the grain protein content. Two previous reported QTLs *viz., qPC-3* and *qProt-8* were validated for use in breeding program. Two very close markers RM407 and RM152 near to protein controlling QTL, *qProt8* were detected for marker-assisted selection. The strongly associated markers with grain protein controlling QTLs namely *qPC2.2, qPC4.2, qPC3.2, qPC-3* and *qProt8* will be useful for their pyramiding in high yielding varieties for developing protein rich rice.

Association mapping of grain Fe and Zn content in rice

The mapping for grain Fe and Zn content through association mapping was performed by Pradhan et al., 2018b. The common milled rice is usually low in iron and zinc content. Consumption of polished rice is the main reason of Fe and Zn deficiency Indian population. A panel was constituted containing biofortified genotypes, some shortlisted landraces and check varieties for the the association mapping study. Phenotyping results indicated a wide variation in grain Fe and Zn content in the studied lines. The study detected 7 QTLs for grain Fe and 9 for grain Zn content to be strongly associated through generalized linear model (GLM) and mixed linear model (MLM). New QTLs namely *qFe3.2, qFe6.1, qFe7.2, qFe8.2* for grain Fe and *qZn2.2, qZn5.2, qZn11.2 and qZn12.3* for grain Zn content were identified. However, 8 more QTLs *qFe 3.3, qFe5.2, qZn2.3, qZn4.1, qZn9.1, qZn11.3, qZn12.2* and *qZn12.3* were identified for Fe-Zn content by considering one model only. Most of the Fe and Zn QTLs were detected to be co-localized in the same chromosome. The strongly associated markers with high grain Fe-Zn content QTLs namely *qFe1.1, qFe5.2 and qFe6, qFe3.2, qFe6.1, qFe7.2, qFe8.2, qZn3.2, qZn6.1, qZn6* and *qZn3.1, qZn2.2, qZn5.2, qZn11.2 and qZn12.3* will be useful for their pyramiding in high yielding varieties for developing Fe-Zn rich rice.

CONCLUSION

Association mapping approach is much useful in dissecting complex traits in rice. Many QTLs can be mapped simultaneously and reported QTLs can also be validated. This approach is also useful for identifying potential donor for a trait. Utilization of natural variability in the process makes it economic in terms of resource deployment. Therefore, association mapping may play an important role in identifying complex loci in rice, further leading to its functional characterization and evaluating donor lines for molecular breeding programme.

REFERENCE

Agrama HA, Eizenga GC, Yan W (2007). Association mapping of yield and its components in rice cultivars. Mol Breeding 19(4):341-356 DOI: 10.1007/s11032-006-9066-6

Anandan A, Anumalla M, Pradhan SK, Ali J (2016). Population structure, diversity and trait association analysis in rice (*Oryza sativa* L.) germplasm for Early Seedling Vigor (ESV) using trait linked SSR markers. PLoS ONE. 11:e0152406. doi: 10.1371/journal.pone.0152406

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, et al. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. 44: 32-39. doi: 10.1038/ng.1018

Huang Y, Sun C, Min J, Chen Y, Tong C and Bao J (2015). Association mapping of quantitative trait loci for mineral element contents in whole grain rice (Oryza sativa L.). J. Agric. Food Chem. 63: 10885-10892. doi: 10.1021/acs.jafc.5b04932

Flint-Garcia SA, Thornsberry JM and Buckler ES (2003). Structure of linkage disequilibrium in plants, Annual Review of Plant Biology 54: 357-374

Muthukumar C, Subathra T, Aiswarya, Gayathri V and Babu RC (2015). Comparative genome-wide association studies for plant production traits under drought in diverse rice (*Oryza sativa* L.) lines using SNP and SSR markers. Current Science 109 (1):139-147

Pan Y, Zhang H, Zhang D, Li J, Xiong H, Yu J et al. (2015). Genetic analysis of cold tolerance at the germination and booting stages in rice by association mapping. PLoS ONE. 10:e0120590. doi: 10.1371/journal.pone.0120590

Pandit E, Tasleem S, Barik SR, Mohanty DP, Nayak DK, Mohanty SP, Das S and Pradhan SK (2017). Genome-wide association mapping reveals multiple qtls governing tolerance response for seedling stage chilling stress in indica rice. Front. plant sci. 2017; 8:552. doi: 10.3389/fpls.2017.00552

Pradhan SK, Barik SR, Sahoo A, Mohapatra S, Nayak DK and Mahender A (2016). Population structure, genetic diversity and molecular marker-trait association analysis for high temperature stress tolerance in rice. PLoS ONE. 11:e0160027. doi: 10.1371/journal.pone.0160027

Pritchard JK and Rosenberg NA (1999). Use of unlinked genetic markers to detect population stratification in association studies. Am. J. Hum. Genet. 65: 220-28

Swamy BPM, Shamsudin NAA, Rahman SNA, Mauleon R, Ratnam W, Cruz MTS and Kumar A (2017). Association mapping of yield and yield-related traits under reproductive stage drought stress in rice (oryza sativa l.). Rice. 10: 21. doi: 10.1186/s12284-017-0161-6

Pradhan SK, Pandit E, Ghritlahre SK, Chatopadhyay K, Chakraborti M, Chandra R , Kar MK , Singh S , Reddy JN (2018a). Association mapping reveals multiple QTLs for grain protein content in rice useful for biofortification (unpublished).

Pradhan SK, Pandit E, Naveenkumar R, Ghritlahre SK, Nayak, DK, Barik SR, Mohanty DP, Mohanty SP, Rao S , Rabindra Babu V, Singh S , Patnaik SSC and Reddy JN (2018b). Genome-wide association mapping of QTLs useful for biofortification of grain Fe and Zn content in rice (unpublished).

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat. Commun. 2:467. doi: 10.1038/ncomms1467

Zhang Y, Zou M and De T (2012). Association analysis of rice cold tolerance at tillering stage with SSR markers in japonica cultivars in Northeast China. Chin. J. Rice Sci 26: 423-430.

Zhao WG, Jong WC, Soon WK, Jeong HL, Kyung HM and Yong JP (2013). Association analysis of physicochemical traits on eating quality in rice (*Oryza sativa* L.). Euphytica 191: 9-21