Association Mapping of Agronomic, Grain Characteristics, Yield Components and Yield Traits Using an Elite Breeding Panel

Hasina Begum

Assistant Professor, City University, Faculty of Agriculture, Birulia, Savar, Dhaka, Bangladesh
Email: *hbhb55[at]gmail.com*

Abstract: *Genome - wide association studies (GWAS) based on linkage disequilibrium (LD) provide a promising tool for the detection of quantitative trait loci (QTL) underlying complex agronomic traits. A total of 28 QTLs were located within the vicinity of previously identified QTLs for all traits using mixed linear model analysis in the irrigated elite breeding lines in dry season, and 45 QTLs in wet season. The highest association was for QHd3a conferring days to flowering (P < 10 $^{-16}$) followed by known QTL for grain length breadth ratio (P < 10 $^{-7}$ and qgy10.1 QTL for grain yield per plot (P < 10 $^{-6}$). Most QTLs had small effects which is typical of most quantitative traits. Most of the QTLs identified are season specific. Some other novel QTL alleles were also identified in this study that may be useful for increasing the yield potential in rice. These potential QTLs for selected traits are of interest to breeder and need to be further validated. Elite breeding populations proved to be interesting material for identifying regions involved in the variation of important traits in rice. This was the first study in rice in which an elite breeding panel was used. Previously for AM in rice, panels have consisted of land races and traditional varieties. We confirmed some regions already observed to be involved in the genetic control of plant height, days to flowering, length breadth ratio, 1000 grain - weight, yield per plant, filled grains per plant and grain yield per plot variation, Moreover, we discovered new QTLs for traits investigated. Association mapping offers great potential to enhance the genetic improvement of rice. The use of high throughput and cost effective next generation sequencing techniques that may enable GWA studies to become a popular and routine approach in rice. Accounting for population structure remains a big limitation for association studies that requires careful choice of germplasm and the development of advanced statistical approaches.*

Keywords: GWAS, QTL, Rice, Breeding, Association Mapping

1. Introduction

GWAS have emerged as a powerful approach for identifying genes underlying quantitative traits at an unprecedented rate (The International HapMap Consortium, 2005, 2007; The Wellcome Trust Case Control Consortium, 2007; [1]. However, despite their promise, GWAS have largely not been applied to the dissection of complex traits in crops [2] [3] [4]. This is mainly due to the lack of effective genotyping techniques for plants and the limited resources for developing high - density haplotype maps like those seen in other well - developed systems, such as the human genome HapMap project (The International HapMap Consortium, 2005, 2007; The Wellcome Trust Case Control Consortium, 2007). Rice is an ideal candidate system for the application of GWAS because it is self - fertilizing and has a high - quality reference genome sequence (International Rice Genome Sequencing Project, 2005).

Association mapping, or LD mapping, has been used in a number of plant species in recent years ([5] [6] [7] [8] [9]. Association mapping has the potential of simultaneous discovery of gene loci responsible for multiple traits without the need to develop permanent segregating populations. A higher proportion of polymorphic molecular markers could provide better genome coverage than any bi - parental population. As association mapping exploits the historical recombination events that have occurred during establishment of the experimental population, higher mapping resolution could be obtained than that possible in small bi - parental experimental crosses [10]. This strategy has an advantage in the ability to detect the comparative effects of multiple alleles at each genetic locus that exists in crop germplasm.

Association mapping in plants of candidate genes and genome - wide association studies (GWAS) have successfully identified associations of marker alleles with traits (e. g., [5] [11] [12] [13] [14] [15] [16] [17]. In barley, GWAS have been conducted to detect QTL in elite germplasm for yield and agronomic traits [6] [18], *Fusarium* head blight resistance [19], winter hardiness [20], and growth habit and inflorescence type [21], agronomic traits and yield in rice [22] [23]. GWAS were conducted to detect QTL for spot blotch resistance in a wild barley collection [24]. In addition, GWAS was conducted as the starting point to identify markers associated with lateral floret fertility, which directly led to the isolation of the INTERMEDIUM - C gene [25].

For most crop plants, including rice, breeding lines selected at the end of a selection cycle represent a potentially useful experimental population for AM studies. These lines may be evaluated in a large number of locations, over two or more years, for many agronomic and quality traits. However, the genome - wide reduction of genetic diversity observed during wheat breeding [26] could reduce the efficiency of AM for some traits. Rice population mixing lines from several breeding programs, each led by independent breeders with their own germplasm, could increase the level of diversity for an efficient use of AM.

High throughput SNP technology made association mapping feasible, routine techniques. For rice, extensive genomics resources are available to facilitate SNP discovery. In rice,

this technology is more advanced compared to other crops. High throughput molecular marker platforms provides good genome coverage (from hundreds to thousands) together with decreasing genotyping costs have encouraged plant geneticists to use naturally occurring variation for identifying genomic regions involved in complex traits [27]. Association genetics offers a potentially powerful approach to identify genetic variants which control complex traits and promises to overcome several of the issues hindering the adoption of QTL - MAS in breeding programmes [28]. One of the main advantages of association mapping is the panel or experimental population used: association panel's benefit from decades of recombination events accumulated in a heterogeneous genetic background. Hence, association studies promise high QTL mapping resolution over a wide range of genetic diversity, i. e. the results should be applicable to all the genetic backgrounds surveyed.

In rice, population structure and its effect on diversity and LD have been reported before. [29] detected five major groups from a diverse sample of 234 rice accessions including *indica*, *aus*, tropical *japonica*, temperate *japonica* and aromatic, and suggested that a higher degree of resolution of population structure is needed to effectively utilize LD for association mapping. The rice population was highly structured and significant LD surrounding the xa5 locus was observed between sites up to 100 kb apart [30]. [31] analyzed a 500 - kb region on chromosome 6 and found a 250 kb selective sweep at the *waxy* locus that led to elevated LD in that region. Although the level of LD may vary across the genome because of different recombination rates, selective pressures, etc., these studies seem to indicate that LD decays in rice at 1 cM or less (assuming an average of 250 kb/cM across the genome) [32]. [33] used unlinked SNPs to determine the amount of background LD in five 500 - kb regions of the rice genome in three major cultivated rice varieties (*indica*, tropical *japonica*, and temperate *japonica*) and in the wild ancestor of Asian rice, *Oryza rufipogon*, and found that the extent of LD is greatest in temperate *japonica* (approximately 500 kb or over), followed by tropical *japonica* (approximately 150 kb) and *indica* (approximately 75 kb), compared to LD in *O. rufipogon* which extends over a much short distance (40 kb). However, other studies using different rice accessions indicated that LD decays at 20–30 cM [32] [34]. These studies suggest that the extent of LD varies among different genomic regions [33], and different rice accessions studied [34].

2. Materials and Methods

Plant Materials
In dry and wet season we used 431 elite breeding lines for phenotyping. For the genotyping we used a population subset from the total entries and the number of entries was 325.

Genotyping by Sequencing
SNP genotyping was conducted in collaboration with Prof Susan McCouch's group using GBS approach in Cornell University. Genotyping by Sequencing was performed following the methods of [35]. Seeds of the lines were sent to the Cornell University.

Data Filtering

Association analyses were conducted using the Trait Analysis by Association Evolution, and Linkage (TASSEL) software version 4.1.34. The general linear model (GLM) and compressed Mixed Linear Model (MLM) were used for the association mapping. TASSEL software Version 4.1.34 [36] was used for all AM data analysis, including filtering by entries (taxa) or SNP markers.

We used the "Filter - > Sites" for filtering in TASSEL, and we calculated the "minimum counts" to calculate the call rates. For example, to calculate SNPs with at least 95% SNP data (i. e. only 5% missing data) for the 325 taxa, we calculated $0.95 * 325 = 309$ for the minimum count.

SNP data set for the population were filtered at a maximum count which accounted for sites where 75% of the lines have a call and a minimum frequency of 0.05 for the minor allele. The working genotype data set for GWAS were generated after imputation of missing genotypes, with a total of 64, 903 SNP sites for the population (minor allele frequency > 0.05).

In order to obtain a minimum minor allele count (MAC) of 5, we have used a MAF of 0.02. We determined this as follows: The total number of taxa = 325. In order to derive a suitable MAF, a minimum MAC (marker allele count) of 5 was used which provided a MAF=0.02. We filtered the SNPs at a call rate of 75% (i. e.244 out of 325 taxa), which means that the MAC should be 5/244, which corresponds to MAF = 0.02 (MAF is based on the number of lines with a SNP call, not the total number of taxa). After doing this, we get 64, 903 SNPs. Then we proceeded with MLM with this "final" dataset (325 taxa x 64, 903 SNPs).

Analysis Population Structure
Related individuals share both causal and non - causal alleles, and that LD between these sites can lead to synthetic associations, are actually a single problem, that of confounding due to genetic background [37]. A powerful method to account for this artifact was first developed in the field of animal breeding: mixed models that handle population structure by accounting for the amount of phenotypic covariance that is due to genetic relatedness (i. e. including relationship or kinship as a random term within the model). Since then, mixed models have been applied to GWAS [11] [38] [39] and can markedly reduce the number of false positive associations.

Principle component analysis was done using the TRANSFORM function in TASSEL. PCA is a statistical tool that transforms a set of correlated variables into a smaller number of uncorrelated variables called principal components (PCs). The first PC captures as much of the variation as possible, and the succeeding PCs account for a decreasing fraction of the remaining variance. PCs were generated using the filtered SNP data set with minor allele frequency > 0.05. The filtered genotype was then transformed using the default option of collapse of non - major alleles.

Kinship
To minimize spurious associations, we compared mixed linear model (MLM) with PCA plus kinship, and kinship

only. Both methods produced identical results. The MLM model with the kinship matrix was only selected for further analysis.

Association Analysis
Several statistical models were used to test associations between markers and traits. These models, described by [11], included a naïve GLM model with no correction for population structure, a Q - model with a Q - matrix and a PC - model with a PC - matrix as cofactors for correcting population structure, and a MLM model with a K - matrix (mixed model without an inferred population structure as cofactor), a Q–K - model and a PC–K - model (mixed models with inferred population structures as fixed effects). All analysis was performed using TASSEL 4.1.34. The p value distributions are shown with a Q–Q plot, which displays the observed p values against the cumulative P values in a negative log10 scale. Under the assumption that the set of genetic markers are unlinked to QTLs, the p values of the association tests are expected to have a uniform distribution, indicated by the diagonal line [38]. A large deviation from this null expectation implies that the statistical test may indicate spurious associations. The cumulative distribution of p values across the different models permitted identification of the most reliable models, which were subsequently used for all the traits of this investigation.

Mixed Linear Model Analysis
The mixed linear model (MLM) approach has been demonstrated to be an improved method of simultaneously accounting for population structure and multiple levels of relatedness among individuals [39]. However, with the MLM method, a large dataset requires more computation time. To reduce this computation time, [40] described a quicker: ''compressed MLM method'' which is implemented by TASSEL.

MLM uses both fixed and random effects which incorporates kinship among the individuals. For running the mixed linear model, a kinship matrix was generated using only SNP markers sites with no missing data. A united data file with the genotype and phenotype of the lines was created. The united file along with kinship matrix was used to analyze associations using MLM across 64, 903 SNPs. The compression level was set to optimum level [40] to reduce computation time. The MLM analysis gives 3 outputs the model statistics, model effects and the compression if applied. The model statistics gives the P - value from the F - distribution for which we used a cut - off $P < 0.0001$ and this output also gives the R^2 for the marker. The model effects indicate the allelic states and effect of the state.

Association mapping was performed using a significant level of $P < 10^{-4}$, which has been used in other GWAS experiments in rice [41]; [42]. Manhattan plots were used to graphically represent of QTLs for each trait. The MLM output from TASSEL was filtered using Excel. The most highly significant SNPs for each QTL peak are reported. For major QTL peaks, defined by those QTLs accounting for a relatively large proportion of the phenotypic variance (R^2) and defined by many significant SNPs, only the top 5 (five) SNPs (defined by the lowest P value) are represented.
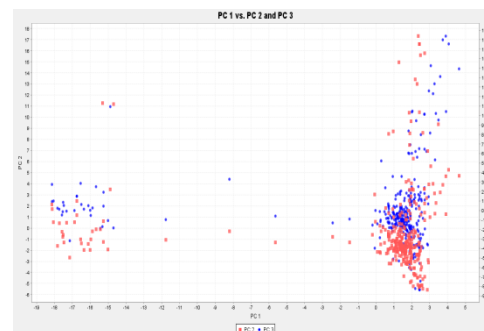
3. Results

Figure 1: Scatter plot matrix showing the first four principle components of the genotype dataset. PCA analysis was performed after removing the japonica outliers.

Preliminary PCA was performed after GBS data was generated (n = 368). A total of 31 taxa were omitted as these taxa represented outliers with probably many japonica introgressions. The irrigated breeding lines are predominantly indica. It was very clear from the PCA that about 13 entries were not in the same sub population as the majority of entries and these were readily identified using only the first two eigenvectors. When these were pulled out and the remaining entries were re - analyzed using PCA again, another group of 18 entries was identified as a separate sub population, but it was much less distinct that the first group. Another 13 samples were removed due to low SNP calls (probably due to row template quality). PCA was performed again after this analysis and cluster analysis to confirm population structure using TASSEL (Fig 1)
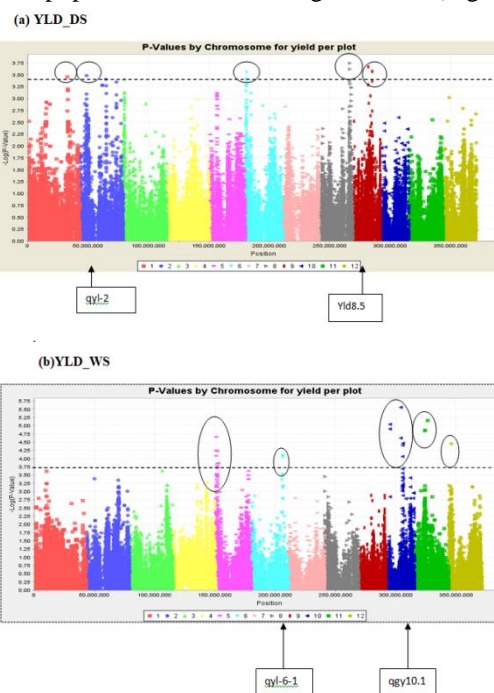

Figure 2a and 2b: Comparison of yield QTLs between dry and wet seasons. Circles indicate the significant QTLs. Previously identified yield QTLs are also indicated with arrows.

4. Discussion

QTLs with small effects, which is consistent with the genetic control of many quantitative traits. Only flowering date had large QTL with phenotypic effect (25 - 30%). For both the small and large effect QTL, we identified previously detected QTL.

There were considerable differences in the QTL regions between dry and wet seasons. This is also consistent in the combine analysis. So, this is normal to get different QTLs across seasons. For most of the traits, QTLs were not detected in the same genomic location across seasons (Fig 2a and 2b).

Flowering, grain length and grain length breadth ratio were consistent across seasons as they had significant SNPs across both seasons. Most of the QTLs were season specific. Based on environment and physical data for rice growing season, there are marked differences which are very obvious to field breeders or researcher. These are clearly demonstrate the genetic control of the traits studies are dependent on environmental factors, and highlight the need for conducting over multiple seasons of data for field experiments.

Some previously detected QTLs or known genes co - localized with QTL detected in this study. Comparative mapping of these previously detected QTLs could enable the QTL region to be delimited within a narrow region, and even locate the actual genes controlling the QTL. There are several bioinformatics methods (e. g. C Map) which could be used to compare QTLs between independent experiments. However many other QTL represent putative novel QTL and those QTLs may be of interest to breeders and geneticists.
Data shown that common QTLs are detected in yield with flowering time and filled grain per plant across different chromosome. Data also shown that common QTL detected in yield with flowering time, flag leaf width and 1000 GW. Data shown that the suggested peak in dry season which are not in significant level (as our cut off P < 0.0001). But we found some previously detected known QTLs. Data also shows that there are many significant QTL across different chromosomes and some previously detected known QTLs were also detected.

Furthermore, these QTLs detected in this study represent QTLs that are highly relevant in irrigated breeding materials and therefore are of great interest to rice breeders.

Yield is the ultimate goal for molecular breeding for breeders. There are few reports for marker assisted breeding for yield. Previous work focused on yield components only and there are very few reports on this topic. The QTLs detected for yield in dry and wet seasons were clearly season specific and there were no significant QTLs detected across both seasons. For the dry season no significant QTLs detected for yield per plot. However they have three suggestive QTL or possively several suggestive QTL on chromosome 1, 2, 6, 8 and 9. Further research is required to validate the QTLs across seasons and years. In other words compare QTL results across different dry and different wet seasons.

5. Summary and Conclusion

A total of 28 QTLs were located within the vicinity of previously identified QTLs for all traits using MLM analysis in the irrigated elite breeding lines in dry season, and 45 QTLs in wet season. The highest association was for $QHd3a$ conferring days to flowering ($P < 10^{-16}$) followed by known QTL for grain length breadth ratio ($P < 10^{-7}$ and $qgy10.1$ QTL for grain yield per plot ($P < 10^{-6}$). These potential QTLs for selected traits are of interest to breeder and need to be further evaluated. They may represent novel QTL alleles that may be useful for increasing the yield potential in rice. Although the resolution varied among loci, mostly due to LD, the resolution was less than 3 cM or ~750 kb. However, the peak signals of the GWAS - identified loci often appeared near (but not within) the known genes. For the known QTL with large effects, as in the case of days to flowering, the distance from the observed peak to the $QHd3a$ locus was in the exact same position. Similar results were found for length breadth ratio and yield per plot. With this resolution, no further QTL mapping or fine - mapping is needed because the SNP is already very near to the gene. Association mapping enabled the identification of QTL with better resolution.

Elite breeding populations proved to be interesting material for identifying regions involved in the variation of important traits in rice. This was the first study in rice in which an elite breeding panel was used. Previously for AM in rice, panels have consisted of land races and traditional varieties. In bred wheat, a panel comprising elite breeding lines was also used for test weight, grain yield and heading date. We confirmed some regions already observed to be involved in the genetic control of plant height, days to flowering, LBR, 1000GW, YPP, FGP and YLD variation, including one region of plant height on chromosome 8, covering gene $ph8$, flowering gene in chromosome 3 included $QHd3a$, 1000GW gene included tgw6 gene in chromosome 6, $gw5$ gene in chromosome 5; YPP gene $yd6$ in chromosome 6; FGP gene $QNFGP - 1 - 1$ gene in chromosome 1; and YLD gene $yld8.5$ in chromosome 8, $qgy10.1$ gene in chromosome 10 and $qYl - 6 - 1$ gene in chromosome 6, those become likely candidate genes. Moreover, we discovered new QTLs involved in those traits.

To mitigate the shortcomings of GWAS in inbreeding crops, future association studies might implement novel strategies such as joint linkage and LD mapping which were already successfully applied in various species. However the experimental design needs to be more powerful to improve the detection accuracy and reduce the genotype x environment interaction. This experiment will also allow testing several models of genomic selection with one or two training populations and one or two validation populations.

Association mapping offers great potential to enhance the genetic improvement of rice. The use of high throughput and cost effective next generation sequencing techniques that may enable GWA studies to become a popular and routine approach in rice. Accounting for population structure remains a big limitation for association studies that requires careful choice of germplasm and the development of advanced statistical approaches. In addition, as the size of

populations and the density of marker screening rapidly increase, so does the probability of detecting non - linked (false) associations. These issues reinforce the need to independently validate candidate genes and/or markers in diverse genetic backgrounds (independent populations) to eliminate false positives. A simple way to validate QTLs is to use/produce biparental mapping populations and use linkage based QTL mapping. The SNPs associated with the QTLs could be directly used to confirm marker –trait association. There is also a need for large - scale cost - effective precision phenotyping, which remains a major logistical challenge and bottleneck to the development of molecular genetics research and breeding programs. Nevertheless, significant progress is being made in facilitating technologies for such phenotyping. Finally, there is undoubtedly an urgent need to bridge the gap between genomics researchers and molecular breeders in developed and developing countries, and particularly to share new knowledge faster and to enable gains in genetic improvement to catch up with those in the leading rice producing countries. If successful, millions of dollars of genomics research investment may finally benefit the poorest people in the world.

References

[1] Altshuler, D., M. J. Daly and E. S. Lander.2008. Genetic mapping in human disease. Science.322: 881–888.

[2] Nordborg, M. and D. Weigel.2008. Next - generation genetics in plants. Nature.456: 720–723.

[3] Atwell S, 2010. Genome - wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature.465: 627–631.

[4] Gore, M. A., J. M. Chia, R. J. Elshire, Q. Sun, E. S. Ersoz, E. T Al.2009. A firstgeneration haplotype map of maize. Science.326: 1115–1117.

[5] Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen and Es. Buckler.2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet.28: 286–289.

[6] Kraakman, A. T. W., R. E. Niks, V. A. N. Denberg, P. Stam and F. A. Van Eeuwijk.2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics.168: 435–446.

[7] Aranzana, M. J., S. Kim, K. Zhao, E. Bakker, M. Horton, K. Jakob, C. Lister, J. Molitor, C. Shindo, C. Tang, C. Toomajian, B. Traw, Zheng H, J. Bergelson, C. Dean, P. Marjoram, and M. Nordborg.2005. Genome - wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1: 531–539.

[8] Mazzucato, A., R. Papa, E. Bitocchi, P. Mosconi, L. Nanni, V. Negri, Mepicarella Me, F. Siligato, G. P. Soressi, B. Tiranti and F. Veronesi.2008. Genetic diversity, structure and marker - trait associations in a collection of Italian tomato (Solanum lycopersicum L.) landraces. Theor Appl Genet.116: 657–669.

[9] Zhu, C., M. Gore, E. Buckler and J. Yu.2008. Status and prospects of association mapping in plants. The Plant Genome.1 (1): 5 - 20.

[10] Flint - Garcia, Sa., A. C. Thuillet, J. Yu, G. Pressoir, S. M. Romero, S. E. Mitchell, et al.2005. Maize association population: a high - resolution platform for quantitative trait locus dissection. Plant J.44: 1054–64.

[11] Zhao, K., M. J. Aranzana, S. Kim, C. Lister, C. Schindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram and M. Nordborg.2007. An Arabidopsis example of association mapping in structured samples. Plos Genet.3: e4.

[12] Belo, A., P. Zheng, S. Luck, B. Shen, Dj. Meyer, B. Li, S. Tingey, and A. Rafalski.2008. Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Mol. Genet. Genomics 279: 1–10.

[13] Wang J., P. E. Mcclean, R. Lee, R. J. Goos and T. Helms.2008. Association mapping of iron deficiency chlorosis loci in soybean (Glycine max L. Merr.) advanced breeding lines. Theor Appl Genet.116: 777–787.

[14] Breseghello, F. and M. E. Sorrells.2006b. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172: 1165 - 1177.

[15] Pressoir, G., P. J. Brown, W. Zhu, N. Upadyayula, T. Rocheford, E. S. Buckler and S. Kresovich.2009. "Natural variation in maize architecture is mediated by allelic differences at the PINOID co - ortholog barren inflorescence2. ". The Plant journal: for cell and molecular biology 58 (4): 618–28.

[16] Wilson, L. M., S. R. Whitt, A. M. Ibanez, T. R. Rocheford, M. M. Goodman and E. S. Buckler.2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16: 2719–2733.

[17] Cockram, J., J. White, D. Zuluaga, D. Smith, J. Comadran, M. Macaulay, Z. Luo, M. Kearsey, P. Werner, D. Harrap, C. Tapsell, H. Liu, P. Hedley, N. Stein, D. Schulte, B. Steuernagel, D. Marshall, W. Thomas, L. Ramsay, I. Mackay, D. Balding, Agoueb Consortium. R. Waugh and D. O'sullivan.2010. Genome - wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proceedings of the National Academy of Sciences of the United States of America.107 (50): 21611 - 21616.

[18] Kraakman, Atw., F. Marti´Nez, B. Mussiraliev, F. Van Eeuwijk and R. Niks.2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Mol Breed.17: 41–58.

[19] Massman, J., B. Cooper, R. Horsley, S. Neate, R. Dill - Macky, S. Chao, Y. Dong, P. Schwarz, G. S. Muehlbauer and K. P. Smith.2011. Genome - wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. Mol Breed.27: 439–454.

[20] Von Zitzewitz, J., F. Condon, A. Corey, A. Cuesta - Marcos, T. Filichkina, K. Haggard, S. P. Fisk, K. P. Smith, G. J. Muehlbauer, I. Karsai and P. M. Hayes.2011. The genetics of winter hardiness in barley: perspectives from genome - wide association mapping. The Plant Genome 4: 76–91.

[21] Cuesta - Marcos, A., P. Szucs, T. J. Close, T. Filichkin, G. J. Muehlbauer, K. P. Smith and Pm. Hayes.2010. Genome - wide SNPs and resequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure. BMC Genomics 11: 707.

[22] Huang, X., X. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, C. Zhu, T. Lu, Z. Zhang, M. Li, D. Fan, Y. Guo, A. Wang, L. Wang, L. Deng, W. Li, Y. Lu, Q. Weng, K. Liu, T. Huang, T. Zhou, Y. Jing, W. Li, Z. Lin, Es. Buckler, Q. Qian, Q. Zhang, J. Li and B. Han.2010. Genome wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42: 961–967.

[23] Huang. X., Y. Zhao, X. Wei, C. Li, A. Wang Et al.2011. Genome - wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44: 32 - 39.

[24] Roy, J. K., K. P. Smith, G. J. Muehlbauer, S. Chao, T. J. Close and B. J. Steffenson.2010. Association mapping of spot blotch resistance in wild barley. Mol Breed.26: 243–256.

[25] Ramsay, L., J. Comadran, A. Druka, D. F. Marshall, W. T. B. Thomas, M. Macaulay, K. Mackenzie, C. Simpson, J. Fuller, N. Bonar, P. M. Hayes, U. Lundqvist, J. D. Franckowiak, T. J. Close, G. J. Muehlbauer and R. Waugh.2011. INTERMEDIUM - C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. Nat Genet 43: 169–172.

[26] Fu, Y. B. and D. J. Somers.2009. Genome - wide reduction of genetic diversity in wheat breeding. Crop Sci.49: 161–168.

[27] Close, T. J., P. R. Bhat, S. Lonardi, Y. Wu and N. Rostoks. et al.2009. Development and implementation of high - throughput SNP genotyping in barley. BMC Genom 10: 582.

[28] Mackay, I. and W. Powell.2007. Methods for linkage disequilibrium mapping in crops. Trends Plant Science.12: 57–63.

[29] Garris, A. J., T. H. Tai, J. Coburn, S. Kresovich and Sr. Mccouch.2005. Genetic structure and diversity in Oryza sativa L. Genetics.169: 1631–1638.

[30] Garris, A. J., S. R. Mccouch and S. Kresovich.2003. Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (Oryza sativa L). Genetics165: 759–769.

[31] Olsen, K. M., A. L. Caicedo, N. Polato, A. Mcclung, S. Mccouch and D. Purugganan.2006. Selection under domestication: evidence for a sweep in the rice Waxy genomic region. Genetics.173: 975–983.

[32] Agrama, H. A., G. C. Eizenga and W. Yan.2007. Association mapping of yield and its components in rice cultivars. Mol. Breed.19: 341– 356.

[33] Mather, K. A., A. L. Caicedo, N. R. Polato, K. M. Olsen, S. Mccouch and M. D. Purugganan.2007. The extent of linkage disequilibrium in rice (Oryza sativa L.). Genetics.177: 2223–2232.

[34] AGRAMA, H. A., and G. C. EIZENGA.2008. Molecular diversity and genome - wide linkage disequilibrium patterns in a worldwide collection of Oryza sativa and its wild relatives. Euphytica.160: 339–355.

[35] Elshire, R., J. Glaubitz, Q. Sun, J. Poland, K. Kawamoto, E. Buckler and S. Mitchell.2011. A robust, simple genotyping - by - sequencing (GBS) approach for high diversity species. Plos one.6 (5): e19379.

[36] Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss and E. S. Buckler.2007. TASSEL: Soft ware for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635.

[37] Vilhjalmsson, B. J. and M. Nordborg.2012. The nature of confounding in genome - wide association studies. Nat Rev Genet.14 (1): 1–2.

[38] Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly and E. Eskin.2008. Efficient control of population structure in model organism association mapping. Genetics, 178 (3): 1709–1723.

[39] Yu, J., G. Pressoir, Wh. Briggs, I. Vroh Bi, M. Yamasaki, Jf. Doebley, Md. Mcmullen, Bs. Gaut, Dm. Nielsen, Jb. Holland, Et al.2006a. A unified mixed - model method for association mapping that accounts for multiple levels ofrelatedness. Nat Genet, 38 (2): 203–208.

[40] Zhang, Z., E. Ersoz, L. Chao - Qiang, Rj. Todhunter, K. Tiwari, A. Gore, Pj. Bradbury, J. Yu, D. A. Arnett, J. M. Ordovas and E. S. Buckler.2010. Adaptation of mixed linear model for genome - wide association studies. Nat Genet.42: 355–360.

[41] Bandillo, N., R. Chitra, Mp Andrea, Sma Lynn, It. Lombina, D - Ec. Jade, T. Chih - Wei, M. Susan, et al.2013. Multi - parent advanced generation inter - cross (MAGIC) population in rice; progress and potential for genetics research and breeding. Rice.6 (1): 1 - 15.

[42] Courtois, B., A. Audebert, A. Dardou, S. Roques, T. Ghneim - Herrera, et al.2013. Genome - Wide Association Mapping of Root Traits in a Japonica Rice Panel. PLoS ONE 8 (11): e78037.