We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,200
Open access books available

116,000
International authors and editors

125M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter 3

Rice Germplasm in Korea and Association Mapping

Aye Aye Khaing, Gang Li, Xiao Qiang Wang, Min Young Yoon, Soon Wook Kwon, Chang Yong Lee, Beom Seok Park and Yong Jin Park

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/56614

1. Introduction

Rice is a traditional staple food crop in Korea and many other countries. Although the center of rice origin is still unclear, it is believed to be introduced from China to the Korean Peninsula in the early Bronze Age via one of two possible routes—across the West Sea or along the northeastern seashore from China according to Hammer (2005) and Vavilov (1935). Rice germplasm has evolved through several millennia of cultivation and selection by our farming ancestors. An important consequence of the domestication of both plants and animals is a reduction of genetic variability (Hancock, 2004). Maintaining biodiversity is an important worldwide problem and different countries have various policies intended to preserve biodiversity. Because conservation of biodiversity and ecosystems is closely linked to the quality of human life, the preservation and improvement of ecosystems are problematic for agriculture.

Genetic diversity in a crop species is essential for sustained high productivity. Breeding efforts have been devoted to improving grain quality, yield potential, resistance to diseases and insect pests, and environmental stress tolerance. Progress in plant breeding requires a continuous supply of genes or gene-complexes. In this respect, the researcher is often handicapped by the limited availability of germplasm resources. The assembly of large varietal collections, systematic screening for desired traits and subsequent incorporation of the relevant genes into existing cultivars is imperative to meet these needs. The use of landrace varieties has increased in recent years. Wild rice accessions have contributed greatly to rice breeding as a source of resistance genes (e.g., *Xa21, BPH14, BPH15*) (Ronald et al., 1992; Song et al., 1995; Yang et al., 2004; Du et al., 2009; Hu et al., 2012). Much of the diversity in the rice gene pool is contained in gene banks around the world. Molecular biology has contributed significantly to an increased
understanding of many aspects of plant biology by generating technologies and methods of analysis that provide new approaches or supplement classical methods of analysis. Plant genetic resource scientists and other researchers are increasingly aware of the potential benefits of applying new technologies to germplasm conservation and research.

The integration of genetic data with molecular biotechnology will help breeders produce new rice varieties with the desired traits and make the conservation of rice genetic resources more efficient. Because of newly developed methods for association mapping of genes or QTLs related to desired traits, many genome-wide association analyses have been conducted in rice and resulted in valuable genome-wide association maps to describe the genetic architecture of complex traits. However, further efforts are needed to obtain more genomic information to fill in the gaps of our knowledge and meet the needs and challenges of rice breeding. This chapter will focus on the status of rice germplasm preservation activities, research programs, and outcomes of association mapping in rice in Korea.

2. Research on rice germplasm in Korea

The Ministry of Foreign Affairs and Trade (2009) had outlined eight major environmental issues as current threats to Korea: global warming, desertification, wildlife extinction, rain forest reduction, acid rain, depletion of the ozone layer, marine pollution, and air pollution. The rate of climate change is faster in Korea than the global average, leading to a rapid reduction in national biodiversity. One hundred and ninety families comprising 4000 species of vascular plants and ferns occur in Korea (Lee and Yim, 1978). Approximately 3700 different kinds of flowering plants are estimated to occur naturally (Chung, 1957; Lee, 1980). Four hundred and seven different endemic taxa in six genera are distributed throughout Korea (Lee, 1982). However, some plant species are on the verge of extinction because of pollution and a wide range of developmental activities during the last 20 years in Korea (Ministry of Environment, 1994), highlighting the importance of conservation efforts (Ahn et al., 1994). Conservation programs usually involve activities such as collection, characterization, evaluation, regeneration, documentation, and storage of each germplasm accession.

The National Biodiversity Strategy was implemented in 1997 to integrate and consolidate plans formulated by various ministries and government institutes, including Comprehensive Biological Resources Conservation Plans. The Rural Development Administration (RDA) Gene Bank is one of the institutions responsible for these plans. Rice research programs covering agronomic practices, physiology, post-harvest technology, grain quality evaluation, rice breeding and genetics, and biotechnology, are led by the National Institute of Crop Science (NICS) under the RDA. Other institutions affiliated with NICS carry out rice research programs to target specific problems in various regions of Korea. From 1980 to 1990, rice sheath blight (Acroclindirum oryzae) was the most destructive disease affecting production from damaging approximately 555,000 hectares of rice paddy fields in Korea. Furthermore, rice pests including brown plant hopper, white-backed plant hopper, and small brown plant hopper attacked 586,000 hectares of rice nationwide during the same period (NASTI, 1996). A continuous
Rice season normally begins in mid-April and ends in mid-October in Korea. The lowest temperature in both April and October is 13°C. Because of the unprecedented yield loss due to cold damage in 1980 (damage to 80% of total rice hectarage and a yield reduction of 3.9 tons per hectare), cultivation of high-yielding “Tongil-type” rice cultivars declined rapidly, and only high-yielding japonica cultivars have been grown since 1990. In 2010, 20 mid- to late-maturing japonica cultivars were grown on 891,493 hectares, accounting for 92.9% of the total rice production area (Kang and Kim, 2012). Large decrease of temperature also occurred in 1971 and 2003, causing damage to 17% and 20% of total rice hectarage, respectively. Preharvest sprouting may become a serious problem for rice production, as well as for other crops, because of the trend in recent years for frequent and unusually heavy rain at harvest time. Breeders are making efforts to address this problem.

Rice breeders see the development of genetically improved cultivars using modern breeding techniques as an efficient way to reduce the losses in rice production caused by biotic and abiotic stresses. Sequencing the rice genome for genotyping and developing marker-assisted selection (MAS) system have fast-tracked research efforts. In the past, most national programs gave a lower priority to collecting wild relatives of rice than to collecting rice cultivars. Wild rice resources are agronomically unattractive, relatively expensive to conserve, and difficult to use. However, wild rice germplasms are known to contain a broad array of useful genes (Hodgkin, 1991). The benefits for the landrace germplasms to be used in breeding new cultivars in response to climate and environment changes in Korea are resistances to diseases in order to maintain superior qualities suited to consumers’ preferences. Plant germplasm resource activities in Korea are performed by The Basic Conservation Programme for Nature and Environment (1994–2003) under the Ministry of Environment (NASTI, 1996).

The RDA Gene Bank conserves 24,673 rice accessions, including Korean landraces and wild types. Many gene banks are having financial difficulty to maintain germplasm collections due to a rapid increase of accession number. These problems may restrict a full exploitation, evaluation, and utilization of these accessions, thus managing such collections presents major challenges (Holden, 1984). The concept of a core collection for resolving these problems has received increased attention over the last few years. Germplasm sampling methods include sequential, stratified, biased (for example, by ecology or country), and random sampling. An understanding of factors underlying the traits being sought will help reduce the time required for identification of useful genes. For very rare traits, such as some associated with resistance to virus infection, searching among wild *Oryza* species and *O. glaberrima* may be most appropriate. Efficient methods for evaluation of germplasm to identify genes for crop improvement will promote the use of conserved germplasm. Frankel and Brown (1984) proposed the concept of a core set of lines to resolve such problems. A desire core set should include the maximum genetic diversity in a crop species including its wild relatives with minimum repetition and provide a manageable set of accessions to gene bank managers, plant breeders, and research scientists. Such a core collection would become the focus of the search for desirable new characteristics, detailed evaluation, and development of new techniques. An initial set of 4406 rice accessions was selected based on ecological types and accession passport information.
including their countries of origin. Using simple sequence repeat (SSR) genotype information, a final core set comprising the 166 conserved accessions currently used by the RDA was generated by a heuristic approach using the PowerCore software developed by Kim et al. (2007). Based on this resulted core set, some association mapping studies have been conducted and further researches are still being undertaken.

3. Association mapping in rice

Association mapping analyzes loci in diverse populations and associates them with both one another and with phenotypes. It is a powerful genetic mapping tool for crops and provides high-resolution, broad allele coverage, and cost-effective gene tagging for the evaluation of plant germplasm resources. Genetic mapping of QTL can be performed in two main ways (Ross-Ibarra et al., 2007): (1) Linkage-mapping as well as “gene tagging” using experimental populations (also referred to as “biparental” mapping populations) and (2) LD-mapping or “association mapping” using diverse lines from the natural populations or germplasm collections (Abdurakhmonov and Abdukarimov, 2008). LD mapping is based on identification of associations between phenotype and allele frequencies. The advantage of LD mapping for the breeder is that mapping and commercial variety development can be conducted simultaneously (Langridge and Chalmers, 2005). For phenotypes or traits that are governed by multiple genes or QTLs, diverse alleles or advantageous allele combinations should be mined by association mapping followed by gene-tagging efforts using biparental crosses.

The localization of alleles relies on creating a statistical association between markers and QTL alleles and on the efficacy of markers. For markers to be effective, they must be closely linked to the target locus and be able to detect polymorphisms in material likely to be used in a breeding program. Improvements in marker screening techniques have facilitated the tracking of genes (Subudhi et al., 2006). Isoenzyme and other protein-based marker systems had in wide long been used before DNA markers became popular (Langridge and Chalmers, 2005). Since then, a variety of DNA-based molecular markers have been developed, including restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs) (Williams et al., 1990), amplified fragment length polymorphisms (AFLPs) (Vos et al., 1995), SSRs (Litt and Luty, 1989), single-strand conformational polymorphisms (SSCPs), cleaved amplified polymorphic sequence (CAPS) markers (Konieczny and Asubel, 1993), sequence tagged sites (STTs) (Olson et al. 1989), sequence-characterized amplified region (SCAR) markers (Martin et al., 1991), and single nucleotide polymorphisms (SNPs) (Brookes, 1999). A total of 2740 SSRs were experimentally confirmed for rice in 2002 or approximately one SSR for every 157 kb (McCouch, 2002). The highly polymorphic nature of many microsatellites is of particular value (Banni et al., 2012, Yoon et al., 2012, Moe and Park, 2012, Zhao et al., 2012a; Khaing et al., 2013) for analysis of closely related genotypes or within narrowly adapted gene pools. Thus, the availability of a high-density SSR map is a valuable public resource for interpretation of the functional significance of the rapidly emerging rice genome sequence information.

The next generation of genetic markers is based on SNPs, which provide an attractive tool for QTL mapping studies and marker-assisted selection in plant breeding programs (Mohler and
Schwarz, 2005). SNP discovery is performed primarily in silico or using new sequencing approaches (Henry and Edwards, 2009). Large-scale SNP analysis is now possible in plants using a range of platforms. The increasing ease of sequencing and automated genotyping has made association mapping in plants a more attractive option by altering the conventional plant genome mapping method, which involves linkage analysis in a segregating population. This trend is likely to continue as the sequencing of genomes increases. Recently, genome-wide association studies (GWAS) with SNP variants have been conducted using new sequencing platforms (Table 1).

3.1. International rice association-mapping activity

Genome mapping of rice was first attempted using linkage analysis of appearance or phenotype (Nagao and Takahashi, 1963). Nowadays, improvement of the linkage map has been achieved using isozymes (Nakagahra, 1977) and RFLPs and SSRs (McCouch et al., 1988; Saito et al., 1991; McCouch et al., 1991 and Yu et al., 1991, Tanksley et al., 1991, Causse et al., 1994; Kurata et al., 1994; Harushima et al., 1998; McCouch et al., 2002). Relatively few association-mapping studies in rice have been performed. Some rice association-mapping studies using various populations and molecular markers are summarized in Table 1 in which most are conducted using SSR markers.

Whole-genome resequencing is a promising strategy to identify the relationship between sequence variation and normal or mutant phenotypes. High-throughput genome resequencing - if accurate - has the advantage of allowing researchers to identify the specific nucleotides associated with a given phenotype, and allowing the effective detection and analysis of genetic variations important for molecular breeding. An important application of NGS is the resequencing of targeted regions for the identification of mutant alleles, and we believe that mapping by sequencing will become a centerpiece in efforts to discover the genes responsible for QTLs. Generally speaking, the availability of a wide range of low- and high-multiplex single nucleotide polymorphism (SNP) assay methods (sequencing accuracy and depth of coverage relies on the experimental design) makes SNPs an ideal marker option for QTL mapping, association analysis, MAS, and the construction of high-density genetic maps for fine mapping and cloning of agronomically important genes (McCouch et al., 2010).

SNP discovery by resequencing whole-genome or subgenome of sample materials is often among the first use of a reference genome sequence. For inbreeding species such as rice, lines to be resequenced are normally purified through 1 or 2 generations of inbreeding (via single seed descent). After a DNA sample is resequenced using NGS technology, SNPs can be identified by comparing the sequenced genome with a reference genome like Nipponbare for japonica rice. For example, using information on the features of the B73, Gore et al. (2009) targeted the gene fraction of the maize genome for resequencing in the founder inbred lines of the nested association mapping population. Two data sets comprising 3.3 million SNPs were used to produce a first haplotype map (“HapMap”) and to analyze the distribution of recombination and diversity along the maize chromosomes.

A suitable example is the construction of a comprehensive HapMap for rice that was used for the genome-wide associate study of 14 agronomic traits, such as heading date and tillering (Huang et al., 2010). The researchers made use of low-coverage (1-fold per rice line) sequence
data across lines, for a combined coverage of ~508-fold, and detected 3.6 million SNPs which can explain ~36% of the phenotypic variance for 14 agronomic traits. This work provided a new approach to low-fold sequence coverage, which can be used to detect not only SNPs but also more complex polymorphisms, and partially overcome the need for deeper sequence coverage (Clark, 2010). Further study was performed with the similar strategy for 950 worldwide rice varieties by Huang et al. (2012) and thirty-two novel loci associated with flowering time and ten grain-related traits were identified. Additionally, 40 cultivated accessions selected from the major groups of rice and 10 from their wild progenitors (O. rufipogon and O. nivara) were resequenced to >15X raw data coverage (Xu et al., 2012). After mapping the sequence read back to an IRGSP reference genome, the authors investigated the genome-wide variation pattern in a comparative analysis. The data revealed examples of structural variation in genomes and included 6.5 million high-quality SNPs after excluding sites with missing data in any accession. Using these population and SNP data, the authors also identified thousands of new rice genes and tracked down those with a significantly lower diversity in cultivated, but not wild rice. These variants represent a valuable resource for those interested in improving rice cultivars.

Preferences in terms of the processing, cooking, and eating qualities of rice differ globally. Plant breeders are attempting to fulfill consumer demand for rice varieties with specific qualities. The major components of rice grain quality include appearance, milling, cooking, eating, and nutritional aspects. The chemical composition of rice grain is important because of its relationship with eating quality of rice. Amylose content is one of the most important traits that determine cooking quality, which is controlled by a major locus waxy (Wx) on chromosome 6 (Wang et al., 1992; He et al., 1999; Tan et al., 1999). Genes associated with amylose content, such as starch synthase IIa (SSIIa) and Wx, are of particular interest. Sano et al. (1986) identified two alleles of the Wx locus using RFLP markers that correspond to the indica and japonica subspecies. Most grain quality mapping studies have used the O. sativa germplasm (He et al., 1999; Tan et al., 1999, 2000, 2001; Zhou et al., 2003). Borba et al. (2010) conducted association mapping study on yield traits and grain quality traits including amylose content, and the significant association detected between amylose content and RM190 was in agreement with previous QTL analyses. Zhao et al. (2011) identified 44,100 SNP variants across 413 diverse rice accessions collected from 82 countries and observed GWAS for six categories of traits covering morphology related traits; yield-related traits; seed and grain morphology related traits; stress-related phenotypes; cooking, eating and nutritional-quality-related traits; and plant development, represented by flowering time. This study demonstrated that different traits have different genetic architectures.

Olsen and Purugganan (2002) elucidated the origin and evolution of glutinous rice based on the haplotype of the Wx gene. By using dCAPS markers, waxy mutations and waxy rice cultivation were shown to have occurred predominantly in the japonica line during the evolution of domestic rice cultivation (Yamanaka et al., 2004). Genetic polymorphisms of starch-synthesis genes have been demonstrated to be associated with starch physicochemical properties using molecular markers such as SSRs, SNPs, and STSs. These markers can be extremely useful in marker-assisted breeding (Bao et al., 2002; Bao et al., 2006a). SSIIa was identified as the major gene responsible for determination of gelatinization temperature (GT). Among four SNPs in the SSIIa gene, some that cause amino acid substitutions have been
identified. The GC/TT SNP is strongly associated with GT (Bao et al., 2006b; Nakamura et al., 2005; Umemoto and Aoki, 2005; Waters et al., 2006).

Rice nutritional quality is another important factor for consumer acceptance. In developing countries where rice is the main food, its nutrient content makes a significant contribution to the intake of some essential nutrients. Interest in natural antioxidants in rice is growing due to their role in preventing oxidative stress-related diseases (Aguilar-Garcia et al., 2007; Willcox et al., 2004; Zhang et al., 2013). Rice contains antioxidant compounds such as γ-oryzanol, tocols, and polyphenols, which are associated with a reduced risk of developing chronic diseases such as cardiovascular disease, type 2 diabetes, and some cancers (Liu, 2007; Tan et al., 2001; Toyokuni et al., 2002). Pigments and flavonoids in colored rice are positively correlated with the antioxidant capacity (Xia et al., 2003; Yawadio et al., 2007). Association mapping of pigments and flavonoid contents was carried out in brown rice using SSR markers. Significant correlations between phytochemical content and marker loci were found and markers associated with multi-phenotypic traits such as grain color, phenolic content and antioxidant capacity were identified (Shao et al., 2011).

The amino acid composition of rice grain is an important characteristic related to nutrient quality. Environmental conditions, potash, and nitrogen dramatically influence the amino acid and protein contents of rice (Wu et al., 2004). Few reports of mapping of QTLs for the contents of protein and amino acids in rice grain have been published. Twelve main effect QTLs (M-QTLs) were identified for 10 components of amino acid content in milled rice. Most of the main effect QTLs for amino acid content tended to co-localize within the genome (Lu et al., 2009).

Although many QTL analyses and genetic mapping studies of grain quality have been conducted, association-mapping studies of biotic and abiotic traits in rice are few in number. The genes or QTLs related to these traits are complex. Genetic mapping, including association mapping and linkage mapping, are useful methods of identifying alleles for these traits. As shown in Table 1, most association-mapping studies focused on morphological and agronomic characteristics. Four studies were related to grain and eating quality and only one addressed disease resistance and aluminum tolerance. Biotic and abiotic stress-tolerance traits remain to be explored by association mapping.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Number of accessions and population type</th>
<th>Number and types of markers used</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virk et al., 1996</td>
<td>200 rice accessions</td>
<td>7 RAPD</td>
<td>Ten morphological traits; culm number, culm length, culm diameter, grain length, grain width, leaf length, leaf width, days to 50% flowering, panicle length and seedling height</td>
</tr>
<tr>
<td>Zhang et al., 2005</td>
<td>218 inbred lines, worldwide germplasms</td>
<td>60 SSR, 114 RFLPs</td>
<td>Multiple agronomic traits such as plant height, heading date, flag leaf length and width, tiller number, stem diameter, panicle length, grain length and width, grain length/width ratio, grain thickness, 1000-grain weight</td>
</tr>
<tr>
<td>Reference</td>
<td>Number and types of markers used</td>
<td>Traits</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Iwata et al., 2007</td>
<td>332 rice accessions 179 RFLPs</td>
<td>Size and shape of milled rice grains</td>
<td></td>
</tr>
<tr>
<td>Agrama et al., 2007</td>
<td>183 rice accessions 123 SSRs</td>
<td>Grain length and width, grain length/width ratio, 100 grain weight, grain thickness</td>
<td></td>
</tr>
<tr>
<td>Yan et al., 2009</td>
<td>90 accessions 108 SSRs + 1 indel</td>
<td>Single, dual and total stigma exertions and spikelet characteristics</td>
<td></td>
</tr>
<tr>
<td>Wen et al., 2009</td>
<td>170 rice accessions 126 SSRs, 6 indels</td>
<td>Heading date, plant height, panicle length</td>
<td></td>
</tr>
<tr>
<td>Borba et al., 2010</td>
<td>242 inbred lines, worldwide germplasms 86 SSRs</td>
<td>Yield, amylose content, head-milled rice</td>
<td></td>
</tr>
<tr>
<td>Huang et al., 2010</td>
<td>517 landraces including japonica and indica ~3.6 million SNPs</td>
<td>Fourteen agronomic traits</td>
<td></td>
</tr>
<tr>
<td>Iwata et al., 2010</td>
<td>332 rice accessions 179 RFLPs</td>
<td>Grain shape variation</td>
<td></td>
</tr>
<tr>
<td>Jin et al., 2010</td>
<td>416 rice accessions 100 SSRs</td>
<td>Grain color</td>
<td></td>
</tr>
<tr>
<td>Ordonez Jr. et al., 2010</td>
<td>192 elite rice breeding lines and tropical japonica germplasm base 97 SSRs</td>
<td>Grain quality and flowering time</td>
<td></td>
</tr>
<tr>
<td>Zhao et al., 2010</td>
<td>395 diverse O. sativa accessions 1,536 SNPs</td>
<td>Amylose content</td>
<td></td>
</tr>
<tr>
<td>Famoso et al., 2011</td>
<td>373 diverse O. sativa accessions 36,901 SNPs</td>
<td>Al tolerance</td>
<td></td>
</tr>
<tr>
<td>Hu et al., 2011</td>
<td>303 O. sativa accessions 24 SSRs</td>
<td>Awnness</td>
<td></td>
</tr>
<tr>
<td>Zhao et al., 2011</td>
<td>413 diverse accessions of O. sativa 44,100 SNPs</td>
<td>Thirty-four traits of agronomic characteristics, cooking and eating quality, disease resistance</td>
<td></td>
</tr>
<tr>
<td>Bryant et al., 2011</td>
<td>174 accessions 156 SSRs</td>
<td>Silica concentration in rice hulls</td>
<td></td>
</tr>
<tr>
<td>Li et al., 2011</td>
<td>217 accessions 154 SSRs and 1 indel</td>
<td>Yield and yield components among 14 traits</td>
<td></td>
</tr>
<tr>
<td>Lou et al., 2011</td>
<td>48 accessions 218 markers (SSRs + indels)</td>
<td>Grain metabolites</td>
<td></td>
</tr>
<tr>
<td>Shao et al., 2011</td>
<td>416 rice accessions including 361 white rice, 50 red rice, and 6 black rice 100 SSRs</td>
<td>Color parameters of brown rice grain, phenolic content, flavonoid content and antioxidant activity</td>
<td></td>
</tr>
<tr>
<td>Zhang et al., 2011</td>
<td>A core collection consisting of 150 rice varieties 274 SSRs</td>
<td>Six morphological traits: glume hair, phenol reaction, length of 1st-2nd rachis internode, glume color at heading, leaf hair, and grain length/width</td>
<td></td>
</tr>
<tr>
<td>Zhou et al., 2012</td>
<td>128 japonica rice varieties 152 SSRs</td>
<td>Eleven quantitative traits of agronomic and economic importance</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. International association-mapping studies of various traits using various markers in rice.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Number of accessions and population type</th>
<th>Number and types of markers used</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang et al., 2012</td>
<td>950 worldwide rice cultivars</td>
<td>1,345,417 SNPs</td>
<td>Flowering time and grain yield traits</td>
</tr>
<tr>
<td>Jia et al., 2012</td>
<td>217 entries</td>
<td>154 SSR markers and 1 indel marker</td>
<td>Sheath blight resistance</td>
</tr>
<tr>
<td>Li et al., 2012</td>
<td>203 accessions</td>
<td>154 SSRs and 1 indel</td>
<td>Harvest index and related components among 14 traits</td>
</tr>
<tr>
<td>Clark et al., 2013</td>
<td>233 rice (Oryza sativa) accessions</td>
<td>36,901 SNPs</td>
<td>Root growth and development</td>
</tr>
</tbody>
</table>

3.2. Association mapping of rice in Korea

To identify useful alleles from a representative core set of rice lines for transferring into elite lines, an allele-mining set of 166 accessions (Zhao et al., 2010) was successfully developed from 4046 rice accessions which were selected from 10368 accessions in the Korea RDA Gene Bank by 39 phenotype traits (Chung and Park, 2009), through a modified heuristic algorithm approach based on 15 SSR markers using the PowerCore software (Kim et al., 2007). Chung et al. also employed the PowerCore software of Kim et al. to develop the first preliminary core set by phenotypes. The gene diversity and population structure (Q) were analyzed using PowerMarker 3.25 (Liu and Muse, 2005) and Structure 2.2 (Evanno et al., 2005) based on 170 SSR markers. Analysis of these data identified the major substructure groups when the number of populations was set at four (Fig. 1).

Association mapping was conducted on this core set of lines over the past 2 years (as shown in Table 2). Zhao et al. (2012b) analyzed 130 accessions from the core set using 170 SSR markers for association analysis of physicochemical traits related to eating quality. Linkage disequilibrium (LD) patterns and distributions are of fundamental importance for genome-wide mapping associations. The mean $r^2$ value for all intrachromosomal loci pairs was 0.0940. LD between linked markers decreased with distance. Marker–trait associations were investigated using the unified mixed-model approach considering both Q and kinship (K). In total, 101 marker–trait associations ($P < 0.05$) were identified using 52 SSR markers covering 12 chromosomes (Fig. 2.). Although direct comparisons of the chromosomal locations of marker–trait associations with previously reported QTLs are difficult because different materials and mapping molecular markers were used, most marker–trait associations were located in regions containing QTLs associated with a given trait. Indeed, some were located in similar or proximal regions related to starch synthesis. The new markers related to eating quality will facilitate the understanding of QTLs and marker-assisted selection (Zhao et al., 2012b).
Figure 1. Values of $\Delta K$, with its modal value used to detect the true $K$ of four groups ($K = 4$). For each $K$ value, five independent runs (blue diamonds) were considered and averaged over the replicates (Zhao et al., 2012b).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Numbers of lines and population type</th>
<th>Number and types of markers used</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhao et al., 2009</td>
<td>84 accessions from land race core set</td>
<td>25 SSRs</td>
<td>16 amino acids</td>
</tr>
<tr>
<td>Zhao et al., 2012</td>
<td>130 accessions from core set</td>
<td>170 SSRs</td>
<td>Eating quality</td>
</tr>
<tr>
<td>Lu et al., 2012a</td>
<td>104 accessions from core set</td>
<td>86 SNPs and indels</td>
<td>Amylose content, RVA</td>
</tr>
<tr>
<td>Lu et al., 2012b</td>
<td>107 accessions from core set</td>
<td>83 SNPs, indels, and SSRs</td>
<td>Amylose content, RVA</td>
</tr>
</tbody>
</table>

Table 2. Rice association-mapping studies for various traits and marker types in Korea.

Association analysis of candidate genes has been used to trace the origin of agronomically important traits. Lu et al. (2012a) used the rice core lines for association-mapping to investigate the relationship between sequence variations from parts of 10 SSRGs and the amylose content (AC) and rapid viscosity analysis (RVA) profiles. Eighty-six sequence variations were found in 10 sequenced amplicons including 79 SNPs, six insertion-deletions (indels), and one polymorphic SSR. Among them, 61 variations were exon-based, of which 41 should lead to amino acid changes. The association mapping results showed a sum of four significant associations between three phenotypic indices and three sequence variations. An ADP-glucose pyrophosphorylase small subunit 1 (OsAGPS1) SNP (A to G) was significantly associated with increased AC ($P < 0.001, r^2 = 15.6\%$) while a 12-bp deletion of an AGPase large subunit 4 (OsAGPL4) (Table 3) was significantly related to decreased breakdown viscosity ($P < 0.001, r^2 = 16.6\%$) in both general linear model (GLM) and mixed linear model (MLM) (Lu et al., 2012a). One SNP with a g/c transversion at the 63rd nucleotide downstream of the OsBEIIb gene termination codon on rice chromosome 2 was significantly associated with multiple trait
indices in both the GLM and MLM, including the final viscosity \( (P < 0.001, r^2 = 23.87\%) \), in both 2009 and 2010, and AC \( (P < 0.01, r^2 = 11.25\%) \) and trough viscosity \( (P < 0.01, r^2 = 20.43) \) in 2010 (Table 4). This study provided a new perspective on the use of allele mining in breeding strategies based on marker-assisted selection (Lu et al., 2012b).

Table 3. Association between sequence variations and phenotype

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Nucleotide variations</th>
<th>AAc</th>
<th>Trait</th>
<th>P_GLM</th>
<th>P_MLM</th>
<th>Rsq_Marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>OsAGP1J</td>
<td>1st exon 106–117th</td>
<td>del(TCCGCCCGCCGCC) del(SAAA)</td>
<td>AAc</td>
<td>BDV</td>
<td>9.99E−04</td>
<td>2.64E−05</td>
<td>0.1656</td>
</tr>
<tr>
<td>OsAGP1J</td>
<td>1st intron 36th</td>
<td>del(gegptgep)</td>
<td>AAc</td>
<td>BDV</td>
<td>9.99E−04</td>
<td>2.64E−05</td>
<td>0.1656</td>
</tr>
<tr>
<td>OsAGP1J</td>
<td>3rd exon 133rd</td>
<td>AAT→GAT</td>
<td>AAc</td>
<td>AC</td>
<td>9.99E−04</td>
<td>2.03E−04</td>
<td>0.1555</td>
</tr>
</tbody>
</table>

AAC, amino acid changes; P_GLM, adjusted P-values with 1000 permutations; P_MLM, P-values significant in the FDR test; amino acid code: S, serine; A, alanine; N, asparagine; D, aspartic acid; BDV, breakdown viscosity; AC, amylase content; FV, final viscosity (Lu and Park, 2012a).
P_GLM: adjusted P-values with 10,000 permutations in GLM; P_MLM: nominal P-values in MLM; Q value: adjusted nominal P-value in MLM by false discovery rate; AC: amylose content; PV: peak viscosity; TV: trough viscosity; FV: final viscosity (Lu and Park, 2012b).

Table 4. Associations between sequence variations and eating quality indicators.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>SNP</th>
<th>Trait</th>
<th>Year</th>
<th>P_GLM</th>
<th>P_MLM</th>
<th>Qvalue</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>OsBEIIb</td>
<td>3end63rd</td>
<td>g/c</td>
<td>AC</td>
<td>2010</td>
<td>2.10E-03</td>
<td>2.19E-04</td>
<td>9.64E-03</td>
<td>0.1125</td>
</tr>
<tr>
<td>PV</td>
<td>2010</td>
<td>3.10E-03</td>
<td>5.82E-04</td>
<td>2.56E-02</td>
<td>0.1590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV</td>
<td>2009</td>
<td>2.00E-04</td>
<td>7.41E-04</td>
<td>1.12E-02</td>
<td>0.1893</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>1.00E-04</td>
<td>9.76E-05</td>
<td>3.14E-03</td>
<td>0.2043</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FV</td>
<td>2009</td>
<td>1.00E-04</td>
<td>4.77E-05</td>
<td>1.31E-03</td>
<td>0.2267</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>1.00E-04</td>
<td>3.85E-06</td>
<td>1.69E-04</td>
<td>0.2387</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OsBE</td>
<td>-224th</td>
<td>c/t</td>
<td>TV</td>
<td>2010</td>
<td>3.60E-03</td>
<td>7.10E-03</td>
<td>3.80E-02</td>
<td>0.1784</td>
</tr>
<tr>
<td>OsSSIIb</td>
<td>-1042nd</td>
<td>t/g</td>
<td>TV</td>
<td>2010</td>
<td>ns</td>
<td>5.90E-03</td>
<td>3.80E-02</td>
<td>0.0986</td>
</tr>
<tr>
<td></td>
<td>-984th</td>
<td>g/t</td>
<td>TV</td>
<td>2010</td>
<td>ns</td>
<td>6.90E-03</td>
<td>3.80E-02</td>
<td>0.0885</td>
</tr>
<tr>
<td></td>
<td>-957th</td>
<td>c/t</td>
<td>TV</td>
<td>2010</td>
<td>ns</td>
<td>6.90E-03</td>
<td>3.80E-02</td>
<td>0.0885</td>
</tr>
<tr>
<td></td>
<td>-782nd</td>
<td>a/g</td>
<td>TV</td>
<td>2010</td>
<td>ns</td>
<td>6.90E-03</td>
<td>3.80E-02</td>
<td>0.0885</td>
</tr>
</tbody>
</table>

P_GLM: adjusted P-values with 10,000 permutations in GLM; P_MLM: nominal P-values in MLM; Q value: adjusted nominal P-value in MLM by false discovery rate; AC: amylose content; PV: peak viscosity; TV: trough viscosity; FV: final viscosity (Lu and Park, 2012b).

Zhao et al. (2009) evaluated the contents of 16 amino acids in brown rice by genotyping using 25 SSR markers. A total of 42 marker-trait associations for amino acid content covering three chromosomes (P <0.05) were identified by the MLM model (Fig. 4), which accounted for more than 40% of the total variation (Zhao et al., 2009). In our research group, association mapping of rice traits related to cold-stress tolerance during germination, preharvest sprouting resistance, salt tolerance, blast disease resistance, and grain physicochemical properties are undertaken using SSRs and SNP variants on advanced resequencing platforms.

In conclusion, association mapping is a promising approach to overcoming the limitations of conventional linkage mapping in plant breeding. Recent research has demonstrated the significant potential of LD-based association mapping of physicochemical traits and other important agronomic traits in rice accessions using SSR/SNP markers. This type of mapping could be a useful alternative to linkage mapping for the detection of marker-trait associations, and lead to implementation of marker-assisted selection in rice breeding programs.

4. Future directions

4.1. Genomics and GWAS in germplasm research

With the development of next- and third-generation sequencing technologies, the whole genomes of individual rice accessions can now be sequenced with less than $ 1000 (US dollar).
Also, new efficient genotyping technologies, such as RADs (Restriction Associated DNAs) (Baird et al., 2008) and GBS (Genotyping-by-Sequencing) allow the generation of genotyping data for up to 40,000 genes at low cost in few days.

Natural alleles and alleles obtained from artificially mutagenized populations provide an important resource for crop breeding. By using all available alleles and detailed phenotypic data from core sets of rice lines, new genes and useful traits can be identified. Molecular tags for useful traits developed using GWASs based on genotypic and phenotypic information can be used to track target traits during segregation of populations in rice breeding (Figure 4).

To identify new alleles from a representative core set of rice lines and transfer them into elite lines, we finally selected 166 from ~25,000 accessions in the RDA Gene Bank. We completed whole-genome resequencing of 84 core accessions with 7x coverage in 2012. We plan to resequence the whole genomes of the remaining 82 core accessions in addition to 84 bred varieties from a validation set in 2013. We are currently undertaking the phenotyping of the core accessions for agronomic traits, and chemical composition for the GWAS analysis with the resequence information. We are also planning to improve the software algorithm for the association analysis to increase the ability to identify alleles from the core set of lines using whole genomic SNP or indel genotype data and phenotypic information. More precise characterization of rice traits that confer resistance to stress from climate change is required to

**Figure 3.** Three regions of putative marker–trait associations on three chromosomes (3, 7, and 8) for amino acid content in brown rice. Genetic distances are indicated in cM on the left of each map and the corresponding marker names are indicated on the right. Ala, alanine; Arg, arginine; Asp, aspartic acid; Glu, glutamine; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Tyr, tyrosine; Val, valine (Zhao et al., 2009).
screen useful alleles using GWASs. Using whole-genome genotype information, we are able to develop large numbers of molecular tags across 12 different rice linkage groups based on their contributions to specific phenotypes.

4.2. Strategy for identification of major and minor QTLs for molecular breeding

The core accessions are highly diverse with many traits useful for rice breeding. Upon selection of an accession with a desirable trait, bi-parental mapping populations will be developed using two japonica varieties (Shindongjinbyeo and Junambyeo) and one indica variety (Hanareumbyeo). Major QTLs will be surveyed with an F8 RIL-segregating population using whole-genome resequencing of 96 samples for first mapping, and then, we can resequence this target region using the expanded 3000 to 5000 samples for fine mapping till the targeted gene can be cloned. We expect that all major QTLs will contribute more than 10% to target traits. To identify minor QTLs that contribute less than 10% to a target trait, BC4F1 population will be first developed, and then, selfing will be done till BC4F6. The recurrent parent will be an elite line for the purposes of QTL mapping and for transferring target traits into the elite lines. Mapping of minor QTLs will be performed using a BC4F6 segregating population (as shown in Fig. 5).
Natural variation results from the expression of different alleles during evolution. As a result of the contributions to farmers over the past ~8000 years, many important traits have been accumulated in the natural germplasm collections currently maintained in seed banks. Whole genome resequencing allows efficient identification of unused alleles from conserved germplasm. We are at present developing a platform for allele mining in rice breeding systems using GWAS approaches and diverse germplasm accessions with the support of the Next-Generation BioGreen 21 Program (No.PJ009099) from Rural Development Administration, Republic of Korea. We believe our effort will facilitate the molecular breeding of rice.

Figure 5. Strategies for identification of major and minor QTLs in rice from selected accessions carrying useful traits through GWAS. The major QTLs will be localized and tagged by molecular markers in the F₈ generation. Minor QTLs will be localized using a BC₂F₈ population.
Author details

Aye Aye Khaing¹, Gang Li¹, Xiao Qiang Wang¹, Min Young Yoon¹, Soon Wook Kwon², Chang Yong Lee³, Beom Seok Park⁴ and Yong Jin Park¹,⁵*

*Address all correspondence to: yjpark@kongju.ac.kr

1 Department of Plant Resources, College of Industrial Sciences, Kongju National University, Yesan, Republic of Korea

2 Department of Plant Bioscience, College of Natural Resources & Life Science, Pusan National University, Milyang, Republic of Korea

3 Department of Industrial and Systems Engineering, Kongju National University, Kongju, Republic of Korea

4 Department of Agricultural Bio-resources, National Academy of Agricultural Science (NAAS), Rural Development Administration (RDA), Suwon, Republic of Korea

5 Legume Bio-Resource Center of Green Manure (LBRCGM), Kongju National University, Yesan, Republic of Korea

References


[53] Lu FH and Park YJ (2012a) Sequence variations in OsAGPase significantly associated with amylose content and viscosity properties in rice (Oryza sativa L.). Genet Res (Camb) 94: 179–189


cessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol. 30: 105-111


