Abstract
Rice is the staple food for more than half of the world's population. Its rich genetic diversity in the form of thousands of land races and progenitors species, besides its economic significance, has aroused unending interest among scientists for several decades. Rice grain quality is based on a combination of subjective and objective factors and it is very difficult to define as it varies from region to region. The primary components of rice grain quality include appearance, eating, cooking, milling and nutritional qualities. The texture of cooked rice is largely dependent upon the amount of amylose in the grain. These factors influence how the rice grain cooked. Short, medium, and long grain cultivars are evaluated for amylose content and the temperature at which the starch gelatinizes to meet the criteria that are set for each grain type by the industry. Each of these components consists of attributes that are determined by the physico-chemical properties of the rice grain and the standards depend on the socio-cultural factors such as the history and traditions of the localities where rice is grown.

Yield enhancement is always a prime motto of rice breeding program. Rice grain yield is determined by several components like number of panicle per plant, grain number per panicle grain weight, panicle length, plant height and many more. Consequently the genetic study of these yield related traits has great importance to speed up yield improvement program. Most of yield related traits are inherited quantitatively and are known to be affected by environmental factors. Segregation analysis and manipulation based on the phenotype would be difficult for such quantitative traits. Identification of QTLs hopefully makes it easy and helps to track segregation as simpler as mendalian inheritance. The QTL analysis provides better understanding of QTL-environment, QTL-QTL interaction which would be beneficial for successful exploitation of QTLs. In addition, identification of QTLs is a valuable starting point for positional cloning of genes underlying quantitative phenotypes and for interpreting the molecular and biochemical mechanisms involve. During the past decade, many attempts have been made to characterize the QTLs for grain yield. However, the genes involved in these QTLs have not been identified and their chromosomal positions remain vague. The present study was focused on fine mapping of QTLs for yield contributing traits in rice using a NIL-F₂ population developed by the cross between RIL151 and RIL131 and the objective-wise summary of work done is given below:
Phenotyping, validation of gene based markers and allelic diversity for starch related properties in rice

Grain quality is an important consideration in rice breeding program. Preferences for rice grain performance and cooking quality vary among rice consumers living in different parts of the world or even among people living in different regions of the same country. The primary components of rice grain quality include appearance, eating, cooking, and milling quality, and nutritional qualities, which are determined by their physico-chemical properties and socio-cultural factors. Among grain quality characters, amylose content is one of the most important of all the cooking quality characteristics because it greatly affects the cooking and processing behavior of milled rice. Rice grain samples contain 80% starch, 12% water, 7.5% protein, and 0.5% ash (Chandler 1979). The starch content is affected to a large extent by climate, rice variety and the level of nitrogen applied. Rice that has just been harvested from the field contains 72% to 75% starch, 1.6% to 2.1% dextrin, 1.5% to 2.6% glucose and 0.3% to 0.5% sucrose (Grist 1986). Starch is composed of linear amylose and branched chains of amyllopectin, which are polymers of α D-glucose. Amylose content is the single most important characteristic for predicting the cooking and processing quality of rice (Juliano, 1985). Rice varieties vary in amount of amylose content, ranging from waxy types with essentially no amylose, to temperate japonica short- and medium-grain types with 16-18% amylose, to tropical japonica long grains with 21-24% amylose, and to many indica types with up to 28% amylose (Rutger and Mackill, 2000). Intermediate amylose content rice shows high volume expansion (not necessarily elongation) and high degree of flakiness. High amylose grains cook dry, are less tender, and become hard upon cooling. In contrast, low-amylose rice cooks moist and sticky. Intermediate amylose rice is preferred in most rice-growing areas of the world. On contrary high yielding cultivars and hybrids are frequently associated with poor cooking and eating qualities and thus are not preferred by producers or consumers. It is, therefore, understandable that emphasis on breeding for quality improvement has assumed great significance with the introduction of high yielding rice varieties and more so due to strong and varied consumer preferences (Khush et al., 1979).

During starch biosynthesis in the grain filling step, several genes and gene isoforms interact and affect starch quantity and composition (Bhave et al., 1990, James et al., 2003, Preiss 1997, Sano 1984, Schultz and Juvik 2004, Wang et al., 1995). As for the sequence difference of genes associated with starch physicochemical properties, the Granule bound starch synthase 1
(GBSS-I) gene harbors two alleles, \( Wx^a \) and \( Wx^b \), in non-waxy rice cultivars. \( Wx^a \) has allele (AGGT) or \( Wx^b \) allele (AGTT) in the first intron affecting splicing (Cai et al., 1998, Hirano and Sano 1998, Issiki et al., 1998). The \( Wx^a \) allele contributes to a higher level of waxy protein than the \( Wx^b \) allele, and thus results in high amylose content in the grain. The \( Wx^a \) allele is predominant in the indica subspecies, while the \( Wx^b \) allele is predominant in the japonica subspecies. The SNPs in the two different exon regions have amino acid substitutions (Larkin and Park 2003) and the 23 bp insertion in the exon region (Wanchana et al., 2003). The starch synthase IIa (SSIIa) gene has seven SNPs and contiguous SNPs (GC/TT) associating with gelatinization temperature and pasting temperature, respectively (Bao et al., 2006b). The starch branching enzyme 1 (SBE1) and starch branching enzyme 3 (SBE3) have the insertion/deletion associations with the starch characteristics (Larkin et al., 2003, Bao et al., 2006a).

However, the information regarding allelic variation at \( Wx \) locus in Indian rice germplasm is lacking. Therefore, in the present study an attempt was made to look at the allelic diversity at the \( Wx \) locus in Indian rice germplasm. In order to study the allelic variation at \( Wx \) locus, validation of six different gene based CAPS and dCAPS markers in a set of 96 diverse Indian rice germplasm lines showing variation for amylose content and alkali spreading value was undertaken. On the basis of alkali spreading values (ASV) and amylose content, genotypes were grouped into low, intermediate and high categories. Six different starch related gene based markers including Granule bound starch synthase-I (GBSS-I), ADP-glucose pyrophosphorylase larger unit (ADPase-L), ADP-glucose pyrophosphorylase small unit (ADPase-S), Putative isoamyloase-type starch debranching enzyme (SDE), Putative 1,4-alpha-glucanbranching enzyme (BE), Putative GIGANTEA (GI) were used for validation in the set of 96 genotypes. The amplified products were digested using the respective restriction enzymes. GBSS-I-1804 amplified 291 bp fragments in all the 96 lines. After digestion with Nhel 291 bp fragment got digested into 264+27 bp fragments. All the varieties in which digestion occurred had amylose content in the range of 19-22% and all those varieties in which digestion did not take place amylose content was 12-20%. One of the SNP of Granule-bound starch synthase I (GBSS-I-2309) showed clear differentiation of the japonica varieties from the rest of the group. GBSS-I-2309 amplified 295 bp fragments in all the germplasm lines. After digestion with Nhel 295 bp fragment got digested into 244+51 bp fragments. All the varieties that showed digestion suggested T allele at this locus while the japonicas did not get digested suggested C allele at the
locus, clearly differentiating the *japonicas* from the rest of the group. Another locus, ADP-glucose pyrophosphorylase (*AGPase-S*) that participates in ADP-glucose biosynthesis, and is the substrate of amylose and amylopectin had been evaluated as the rate limiting step in starch biosynthetic pathway. SNP of ADP-glucose pyrophosphorylase (*AGPase-S-2843*) clearly inferred that at this particular locus there was no digestion among all the *japonica* varieties, suggesting a conserved genomic region, while all other cultivars showed digestion. Out of all the 22 CAPS and dCAPS two markers, *GBSS I-2309* and *AGPase-S-2843* were found to be highly associated and clearly differentiated the *japonicas* from the germplasm lines having intermediate or high amylose contents, based on their genotypic profile.

For the association test between the nucleotide polymorphism and apparent amylose content DARwin 5 software was used, UPGMA algorithm grouped 96 varieties into 3 distinct groups varying in apparent amylose content. Group I contained the maximum number of the accessions with the average amylose content ranging from 14-20%. Group II was the second largest group with 15 accessions showing amylose content ranging from 21-26% and Group III had the accessions with the lowest amylose contents ranging from 10-13% with few outliers. In cluster analysis the varieties bred at one centre fell in single group suggesting a common pedigree. The statistical values for marker application results using 96 accessions gave a mean major allele frequency (MAF) of 0.66 (0.52–0.92) and polymorphism information content (PIC) of 0.35 (0.20–0.40). The association analysis showed that the SNP of Granule-bound starch synthase I (*GBSS I*) and ADP-glucose pyrophosphorylase small subunit (*ADPase-S*) genes were highly associated with apparent amylose content variation than the others. The $r^2$ value of the two most associated markers ranged from 17 to 54.1% and from 21.3 to 64.5%, respectively (P<0.0001). The highest value of $r^2$ was at GBSS I- 2309-T/C ($r^2=0.54$). The high association at *AGPase-S-2843-C/T* ($r^2=0.65$), whose gene is concerned to ADP-glucose biosynthesis (Preiss 1997) was observed.

In addition to this, amplification was checked among 96 germplasm lines using a *Wx* microsatellite marker, Oligo 484 and 485. The *Wx* microsatellite was polymorphic enough to distinguish most rice cultivars in different amylose classes. The gels were scored and different alleles were detected. The amplified products ranged from 103 to 127 bp in length differentiated the lines broadly into two categories varying in their amylose contents. In the first group genotypes having intermediate amylose content (21-26%) similar to Pusa Basmati 1121
amplified a fragment of ~107 bp while those belonging to low amylose (10-14%) group, similar to Taipae-309, a fragment of ~120 bp was amplified, suggesting the difference in the CT repeats in the 5'-untranslated region of the Wx gene.

In conclusion, this study validated new SNP marker systems for six putative genes involved in starch biosynthesis in rice. These new SNP markers can be used for the fast and cost-effective genetic characterization of diverse rice germplasm and can also be utilized for marker-assisted selections, linkage, and association studies in rice.

**Development, phenotypic and molecular characterization of F$_2$ derived population for fine mapping of a major QTL (qSPP4-I) for yield components in NPT derived population in rice**

As the world's population continues to grow rapidly and becomes more demanding, the pressure on resources is increasing, whilst climate change poses further challenges. The balance between the supply and demand of the major food crops is fragile, fueling concerns for long-term global food security. The need to accelerate plant breeding for increased yield potential and better adaptation to drought and other abiotic stresses is an issue of increasing urgency. In order to meet the growing demand of the ever increasing population, rice productivity must increase for which rice varieties with higher yield potential are needed. Rice yield is the final product of the manifestation of several interrelated components, like number of tillers per plant spikelets per panicle and grain weight and is therefore a very complex quantitative trait. With the whole genome sequencing, rice has become the 'model monocot plant' for molecular genomics research. It benefits from having the smallest genome of the major cereals, dense genetic maps and relative ease of genetic transformation (Sasaki et al., 2000). Saturated genetic linkage map of rice using different types of markers namely, RFLP and SSR, has been developed, which has been further enriched with SNP markers, with the availability of whole genome sequence. Utilizing these markers, a large number of genes of agronomic importance have been mapped and many genes/QTLs have been eventually cloned. Recent progress in the generation of high-density molecular genetic map in rice has made it possible to map individual genes associated with the complex traits like yield. Though hundreds of QTLs have been identified during the last decade for yield per se, only a few QTLs for grain number/panicle, test grain weight, panicle number per plant, grain size and grain filling duration (Bai et al., 2012) could be cloned and
functionally characterized, mostly in *japonica* background. These includes important yield component traits such as grain number (*Gn1a, Ghd7, DEP1* and *WFP*), grain weight (*GS3* and *GW2*), grain size (*GS3* and *GW5*) grain filling (*GIFI*) and panicle number (*DEP1* and *WFP*). Among all the yield and yield related traits, grain number per-panicle is one of the most important components for rice yield. Spikelets on the primary and secondary branches determine the grain number per-panicle in rice. A total number of 315 QTLs for grain number have been earlier reported on almost all the 12 rice chromosomes, but most of them have low LOD values and were not evaluated across the environments. Among all the QTLs reported, QTL on chromosome 1 (*Gnl*) have been fine mapped and the gene have been cloned (Ashikari *et al.*, 2005).

The are two major areas where progress has been achieved for rice improvement i.e. the use of molecular markers for identifying and incorporating favorable genes within the rice species, and the use of transgenic technologies for incorporation of traits for herbicide tolerance, biotic stress resistance, abiotic stress resistance, and nutritional value into rice (Coffman *et al.*, 2004, Leung and An 2004). These scientific advances have resulted in increased understanding and characterization of various genes at the molecular level that are associated with traits important to plant breeders. These traits are considered primary targets for molecular marker-aided selection (MAS) as breeding for them using conventional techniques often proved to be difficult.

In the context of Indian rice germplasm, limited studies have been made with respect to quantitative traits such as yield and it’s components. However, most of the studies on QTL mapping for grain yield and its component traits are based on mapping population developed from intervarital crosses within sub species *indica* or *japonica*. Very few attempt, have been made to develop mapping population from interspecific crosses or utilizing new plant types (NPTs) of ricer derived from *indica x japonica* crosses. In the earlier study done in our lab, a RIL population using single seed descent method was developed by crossing a new plant type rice line, Pusa 1266 with highly-adapted *indica* rice varieties, Jaya for mapping of different yield and yield related traits. A marker interval RM3276-RM5709 on long arm of chromosome 4 influencing as many as four traits i.e., Spikelets per panicle, filled grains per panicle, panicles per plant and spikelet setting density were identified with a LOD scores of 13.84, 15.27, 16.26 and 7.11 respectively (Marathi *et al.*, 2012).
Four QTLs governing four important yield related traits (spikelets per panicle, filled grains per panicle, number of panicles per plant and spikelet setting density) were fine mapped using a F$_2$ population of 2994 plants, developed by crossing two RILs (RIL151 and RIL131) contrasting for the traits. Phenotyping showed transgressive segregation for all the four traits suggesting high proportion of additive gene action. High significant positive correlation was exhibited between SPP and FGP (0.93), SSD (0.96). The significant genetic correlation between the traits in segregating population is indicative of the genetic linkage or pleiotropy of the underlying gene(s). PPP and SSD showed comparatively weak correlation (0.06) depicting yield component compensation mechanism that exists in a typical plant system. Existence of significant correlation between majorities of character pair observed in F$_2$ population indicated presence of sufficient genetic variation for all these traits and their suitability as mapping population. Linear regression of F$_3$ on F$_2$ for 324 recombinants revealed high significant $b$ value of 0.55, 0.54, 0.59 and 0.65 with respect to the traits SPP, FGP, PPP and SSD respectively. Linear regression is a measure of heritability of the trait and high regression coefficient ($b$) explains that the grain and panicle traits under study in the population are under genetic control and are stably inherited.

Fine mapping of the interval was done using the two flanking markers and the internal markers on 324 recombinants, identified using the flanking markers. Total 5 markers were added in the QTL interval of the revised map with total seventeen markers for the whole chromosome 4. Additional QTL information showed new QTLs between markers nkssr4_19 and RM17486, with significant phenotypic variation and LOD values. Candidate gene search showed a number of genes in the region. The 0.03 Mbp qSPP4-1 region of rice chromosome 4 between markers nksssr04-11 and RM17470 showed 6 predicted gene loci and have the evidence for expression using the FGENESH 2.6 software (www.softberry.com) (TIGR, version 6.1, http://rice.plantbiology.msu.edu/). The gene, LOC_Os04g51792, has been associated with the cytokinin response in roots and leaves and iron and phosphorus interactions in rice seedlings. According to the earlier reports QTL (Gnl) on rice chromosome 1 has been cloned and has been associated with the cytokinin oxidase activity and is a negative regulator of cytokinin in the inflorescence primordia leading to the low grain number in the panicle (Ashikari et al., 2005). Another locus, LOC_Os04g51800, has been known as MYB protein, transcription factor play a key role in plant development, such as secondary metabolism, hormone signal transduction,
disease resistance, cell shape, control of cell morphogenesis, floral and seed development. (Du et al., 2009). Over all, these genes could be associated with the traits after the transcriptome profiling studies.

Conclusion:

1. A set of 96 diverse germplasm lines were well characterized for amylose content and alkali spreading value.
2. Validation study with the new SNP marker systems for six putative genes involved in starch biosynthesis using a diverse set 96 Indian rice genotypes showed significant associations of these markers with the respective traits in germplasm. Validated markers could be utilized for in fast and cost-effective genetic characterization of diverse rice germplasm and can also be utilized for marker-assisted selections, linkage, and association studies.
3. Association of the phenotype and genotype grouped all the lines into three major groups varying in their amylose content.
4. A set of 2994 F2 population developed by crossing RIL151 with RIL131 was used as a fine mapping of major QTLs \( qSPP4.1, qFGP4.2, qPPP4.2 \) and \( qSSD4.2 \) for yield and yield related traits (spikelets per panicle, filled grains per panicle, spikelet setting density and number of panicles per plant).
5. The 1.38 Mb gap was narrowed down to 0.03 Mb, depicting six genes that have the evidence of expression.
6. Two gene loci, LOC\_Os04g51792 and LOC\_Os04g51800 could be associated with the yield related traits after the transcriptome profiling studies.

Overall, the study has helped in validation of marker-trait association with amylose content in a set of diverse rice germplasm. Validated markers and identified parents can be used in marker assisted selection. Further, a QTL region harboring four QTLs for different traits was fine mapped and leading to identification of 6 candidate gens. Of these two genes have been identified as putative candidate genes and can be associated with yield and related traits.