Identification of Candidate Genes for Rice Grain Aroma by Combining QTL Mapping and Transcriptome Profiling Approaches

V. Pachauri[1]*, V. Mishra[1], P. Mishra[1], A.K. Singh[2], S. Singh[2], R. Singh[3] and N.K. Singh[1]

[1]Rice Genome Laboratory, National Research Centre on Plant Biotechnology, Lal Bahadur Shastri Building,
Indian Agricultural Research Institute, New Delhi-110012, India
[2]Division of Genetics, Indian Agricultural Research Institute, New Delhi-110012, India
[3]DNA Fingerprinting Lab, National Bureau of Plant Genetic Resources, New Delhi-110012, India

The present report is in continuation to our earlier reports on the identification and fine mapping of three aroma QTLs in basmati rice using a bi-parental mapping population derived from a cross between Pusa 1121, a basmati rice variety, and Pusa 1342, a non-aromatic rice variety. We used a combination of genetic mapping and transcriptome profiling to narrow down the number of differentially expressed genes in rice to identify potential candidate genes for rice grain aroma. Highly aromatic and non-aromatic recombinant inbred lines (RILs) were identified through sensory analysis of mature milled grains. RILs with similar phenotypes were bulked together using bulk segregant analysis approach which drastically reduced the number of differentially expressed genes from 4016 to 1344. The transcriptome profiles generated were analyzed through Affymetrix rice genome array containing probe sets designed from all the predicted rice gene sequences. Microarray-based transcriptome profiling revealed one down-regulated gene co-located in QTL region *aro3.1* on chromosome 3, eight genes co-located in the *aro4.1* region on chromosome 4 and the *badh2* gene on chromosome 8 to be differentially expressed in the aromatic parent and aromatic bulk. These genes are the most suitable candidates for future validation and development of new molecular functional markers to facilitate marker assisted breeding.

Keywords: rice aroma, QTLs, bulk segregant analysis, transcriptome, microarray

Introduction

Identification of gene(s) underlying various quantitative traits is important for the development of functional markers that can be used in marker-assisted plant breeding. Map-based cloning approach has long been used for gene identification but is a time consuming and expensive tool (Salvi and Tuberosa 2005; Price 2006). In recent times, microarray based expression profiling of genes also known as transcriptome profiling has

* Corresponding author; E-mail: vinita.pachauri@nrcpb.org; Phone: +91-9990085561

been an important tool for gene identification. The availability of whole genome microarray chips of various organisms has paved the way for the effective study of complex traits, by assessing the expression of almost all the genes at a time (Kathiresan et al. 2006). Wayne and McIntyre (2002) were among the first who combined genetic mapping studies and expression profiling to narrow down the number of candidate genes for ovariole number in Drosophila. The only short coming with this technique is that a very high number of up- and down-regulated genes are identified which is basically due to cross-talks and overlapping pathways of genes involved in the quantitative traits. This is overcome by using bulk segregant analysis (BSA) approach, where transcriptome profiles of bulk of inbred lines having similar and extreme quality trait(s) are studied.

This helps in reducing background as well as the number of individual samples to be analyzed, hence, total time and cost involved in experimentation is also cut down by manyfolds. This approach has been used in animal system (Aitman et al. 1999) as well as crops including maize (Marino et al. 2009) and rice (Deshmukh et al. 2010; Pandit et al. 2010).

Since rice aroma, a polygenic quantitative trait with complex inheritance pattern is highly influenced by environment it is difficult to identify genes that undermine this trait (Pachauri et al. 2010). Initially, it was reported that rice aroma is controlled by a single dominant gene with a segregation ratio of 3:1 (Kadam and Patankar 1938). Subsequently, Jodon (1944) contradicted this observation and reported that rice aroma is controlled by a single recessive gene, *fgr*. Using a biparental population derived from a cross between Pusa 1121 and Pusa 1342, it was revealed that rice aroma is controlled by three Quantitative Trait Loci (QTLs), *aro3.1, aro4.1* and *aro8.1* located on short arm of chromosome 3 and long arms of chromosome 4 and 8, respectively (Amarawathi et al. 2008). The major aroma QTL *(aro8.1)* was identified on chromosome 8 with LOD score of 11.54 between SSR markers RM223 and RM80. This QTL explained 18.9% of the phenotypic variation for aroma. This QTL was mapped in the same region as that reported earlier by Ahn et al. (1992) and Lorieux et al. (1996). Later, studies by Bradbury et al. (2005), Wanchana et al. (2005), Chen et al. (2006) and Amarawathi et al. (2008) identified *badh2* (recessive allele) as a candidate gene for aroma on this chromosome, which codes for enzyme betaine aldehyde dehydrogenase (BADH). This enzyme is involved in the synthesis of glycinebetaine-a powerful osmoprotectant against salt and drought stress in a large number of species. Rice does not accumulate glycinebetaine but it has two functional genes coding for the BADH enzyme. Most of the aromatic rice varieties from different isozyme groups share the same 8 bp deletion in intron 7 of *badh2* gene (Bradbury et al. 2005) for which Amarawathi et al. (2008) designed a perfect gel based marker (nksbadh2) that discriminates between aromatic and non-aromatic varieties. The *BADH1* gene located on chromosome 4 having similar biochemical function has also been anticipated to have a contributory role in aroma expression in rice (Singh et al. 2010).

The QTL *aro3.1* on chromosome 3 was mapped in marker interval RM5474–RM282 with LOD score 3.20, explaining 10.3% of the phenotypic variation, whereas QTL *aro4.1* was located in the marker interval RM5633–RM273 on chromosome 4 with LOD score of 3.30, explaining 6.1% of phenotypic variation. The QTL *aro3.1* was later further fine mapped by Singh et al. (2007), to a further narrow region of 390 kb flanked by markers

CHR3_22 and CHR3_24. A significant LOD score of 2.92 was observed that explained 11% of the total phenotypic variation for aroma.

The possible involvement of other gene(s)/allele(s) responsible for grain aroma cannot be ruled out as some aromatic rice varieties lacking *badh2* gene have been reported (Fitz-gerald et al. 2008). Being a complex inherited trait, limited information is available about aroma genes other than *badh2,* so the present investigation was done to screen out genes co-located in the aroma QTL regions, with major focus on chromosomes 3 and 4. Our ear-lier reported data on the mapped QTLs for basmati grain aroma (Amarawathi et al. 2008; Singh et al. 2007) were taken as the basis for this further investigation. RILs with similar phenotype were pooled to create bulks which significantly reduced the number of samples to be analyzed and their transcriptome profiles were generated through microarray.

Materials and Methods

Plant material

A mapping population of 180 RILs was developed from a cross between rice varieties Pusa 1121 and Pusa 1342 by single seed descent method (Singh et al. 2002). Pusa 1121, a basmati long grain variety having the status of longest cooked grains in the world was taken as the female parent, while Pusa 1342, a non-aromatic variety was the male parent in the cross developed. The entire mapping population was grown at the experimental fields at Division of Genetics, Indian Agricultural Research Institute, New Delhi.

Phenotyping of RILs for grain aroma

Sensory analysis of milled rice grains was done according to the method given by Sood and Siddiq (1978). Scoring for grain aroma was done on an arbitrary scale of 0–3 for three consecutive years (2008, 2009, 2010); score 0 for non-aromatic types, score 1 for slightly aromatic types, score 2 for moderately aromatic types and score 3 for highly aromatic type RILs. Random blind checks were used to increase the efficacy of scoring. Average score of four biological replicates was recorded each year.

Genotyping and transcriptome profiling of RILs

DNA extraction and genotyping

DNA from parental lines and RILs was isolated from about 12–15-day-old seedlings by CTAB method of Murray and Thompson (1980). The parents and RILs were genotyped with the nearest linked markers from chromosome 3 (CHR3_22), chromosome 4 (RM5633) and gene based nksbadh2 marker from chromosome 8. PCR amplification was performed in a 10 µl volume containing 45 ng of template genomic DNA, 5 pmol (13 ng) each of forward and reverse primers, 0.2 mM dNTPs, 1X PCR buffer (10 mM Tris, pH 80, 50 mM KCl and 50 mM ammonium sulphate), 1.5 mM MgCl2 and 0.5 units of Taq DNA polymerase (all the reagents from Puregene Tauras Scientific). Template DNA was ini-tially denatured at 94°C for 5 min followed by 35 cycles of repeated PCR amplification

with the following parameters in a sequential order: denaturation at 94°C for 1 min, annealing at 55°C for 1 min, primer extension at 72°C for 30 sec with final extension at 72°C for 10 min. Amplified products were electrophoretically resolved on 4% Metaphor Agarose gels (Lonza, USA) prestained with ethidium bromide (0.1 mg/ml). The gels were photographed and visualized by transillumination under short-wave UV light by gel documentation system (AlphaImager, Flourochem 5500).

Transcriptome profiling of parents and RILs

Rice whole genome microarray chips were used to study the transcriptome profiles of the two parents as well as RIL bulks to identify differentially expressed genes between the groups. Four biological replicates were created for each of the four groups (aromatic parent, non-aromatic parent, aromatic RIL bulk and non-aromatic RIL bulk) making a total of sixteen samples.

RNA isolation, cDNA synthesis and GeneChip hybridization

Fresh leaf samples of RILs and parents were harvested from the fields after around 60–75 days of transplanting, a stage when seed setting had already initiated. For aromatic and non-aromatic bulks, 100 mg leaf sample from each RIL was taken while for parental lines, 600 mg leaf samples were taken. Whole RNA was isolated using Qiagen RNeasy Kit following the manufacturer's instruction manual. cDNA synthesis and chip hybridization was done according to instructions from the Affymetrix GeneChip expression technical manual. cDNA was synthesized from 8 μg crude total RNA using Superscript double-stranded cDNA synthesis kit and poly(T) nucleotide primers containing sequences recognized by T7 RNA polymerase. cRNA was tagged with biotin using the IVT labelling kit provided by Affymetrix. Ten μg of fragmented cRNA was hybridized, washed and stained with streptavidin-phycoerythrin in the Affymetrix fluidics station 450. The probe Gene-Chips were then scanned in a GeneChip Scanner 3000.

Affymetrix Rice GeneChip array and probe annotation

The Affymetrix rice genome array contained probe sets designed from 48564 *japonica* and 10260 *indica* gene sequences. Sequence information for this array was obtained from National Centre for Biotechnology Information (NCBI) UniGene build number 52 (http://www.ncbi.nlm.nih.gov/UniGene), GenBank mRNAs, and 59,712 gene predictions from TIGR's osa1, version 2.0. Gene models that had any indication of transposable elements were removed from the list of TIGR genes. The array is represents about 46,000 distinct rice genes of which about 26,000 are 3_ anchored Unigene ESTs and mRNA clusters, including known rice full-length cDNA clones, and 19,431 are solely from the TIGR gene predictions.

Statistical analysis of microarray data

The array data was analyzed using GeneChip Operating System (GCOS 1.2) and GeneSpring software version GX 9.0.6. A default target intensity value (TGT) setting of n = 500 was used while scaling factor for the arrays ranged from 3.1–8.5. The detection calls

(present, absent or marginal) for the probe sets were made by GCOS. Robust multichip average (RMA) algorithm of GeneSpring was used for normalization of probe levels and only those probes with expression levels above the background levels with statistical significance ($P < 0.01$) were recorded as "present". For false discovery rate (FDR), the Benjamini and Hochberg (1995) algorithm with a cutoff setting of <2% was applied. Fold changes >2 were considered as significant. To evaluate the differences in gene expression between Pusa 1121 and Pusa 1342, the non-aromatic parent (Pusa 1342) was treated as the control.

Validation of differential gene expression by quantitative real-time PCR (qRT-PCR)

Expression levels of the ten genes co-located in the aroma QTL regions on chromosomes 3, 4 and 8 were also analyzed using qRT-PCR for validation of results obtained through microarray experiments. Sequence of genes was obtained from TIGR rice database (http://rice.plantbiology.msu.edu), while primers were designed from the exonic sequences using IDT SciTools software (http://eu.idtdna.com/Scitools/Applications/RealTimePCR) (Table 1). For normalization, eukaryotic elongation factor 1-alpha (eEF-1α) was used as internal control. SuperScript III Platinum SYBER Green One-Step qRT-PCR Kit provided by Invitrogen was used according to the standard instruction manual. RNA was isolated from the same seedlings at similar stage as used for microarray analysis. The PCR conditions were: 50°C for 1 h followed by an initial denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 30 s, annealing at 55–60°C for 1 min and extension at 72°C for 1 min. The reactions were performed in Stratagene Mx3000p while the data was analyzed by Mxpro-QPCR software (Stratagene).

Results

Phenotyping of parents and RILs for grain aroma

The two parents differed greatly for aroma; Pusa 1121 had a high aroma content (given a score of 3 on an arbitrary scale of 0–3) whereas Pusa 1342 was non-aromatic with a sensory score of 0. The mode of three years was taken as the final score, on the basis of which, 30 (16.67%) RILs were non-aromatic while 76 (42.22%) were mildly aromatic, 67 (37.22%) were moderately aromatic and 7 (3.85%) were highly aromatic. Earlier reports suggested the involvement of only one recessive gene for aroma located on chromosome 8 (Ahn et al. 1992; Bradbury et al. 2005; Wanchana et al. 2005).

   On the other hand studies by Pinson (1994) and Lorieux et al. (1996) have indicated involvement of multiple genes for rice aroma. The present investigation also suggested the involvement of three or more genes as only a very small proportion of RILs (3.85%) retained the original high aroma of Pusa 1121 (the expected proportion of genes in RILs will be $1/2^n$; where n is the number of genes). Based on the scores of three consecutive years, six highly aromatic RILs and six non-aromatic RILs were identified and used for expression profiling.

Table 1. Details of primers used for qRT-PCR analysis of candidate aroma genes co-located in reported aroma QTLs identified through transcriptome profiling of bulked RILs from Pusa 1121 and Pusa 1342

Gene ID	Locus ID	Forward sequence	Reverse sequence	Product size (bp)
Os03g0327600	LOC_Os03g21040	GCACAGCAACAGCATCAAG	ACTCGTCCTGGTATCCGG	143
Os04g0352400	LOC_Os04g28420	GGGATTTGAAGGACAATGCTG	GAAGGAGCTGTCATACTCGATG	146
Os04g0434800	LOC_Os04g35520	TGGAGCACATACACTTGGAAG	CATTCAGCTGTCCACGATTG	117
Os04g0401700	LOC_Os04g32920	CCTCTCCCTCATCCTCTACAC	CATCCTGATCTTGGCGTACC	139
Os04g0438300	LOC_Os04g35760	TTTTCAGGCTCATCTCCTACG	TCAGTGACTCAAATGGCGTC	150
Os04g0445700	LOC_Os04g36800	GGGTGAGGCTGATGTTATCG	ATCTCGCTCTTTGTCCCATG	145
Os04g0468600	LOC_Os04g39320	ATGGAGCAGGTGTGGAATC	CTTCTCGGTGAGGTCGTTG	98
Os04g0474800	LOC_Os04g39880	AATAACTCTGGTCTCGCACTG	CCAATCCTCTCATGCTTAGGG	146
Os04g0469700	LOC_Os04g39430	CAGGTATGGGAGGGTGTTTAAG	GCCCAGAATGCCATGAATTG	145
Os08g0424500	LOC_Os08g32870	CCTATCGGTGTAGTTGGGTTG	GCAAGCTCCAAACAAGTCAC	143
eEF-1α	AK061464*	TTTCACTCTTGGAGTGAAGCAGAT	GACTTCCTTCACGATTTCATCGTAA	105

* NCBI accession number

Table 2. Frequency of alleles contributed by aromatic parent Pusa 1121 of nearest linked flanking markers from the three reported aroma QTLs, one each on chromosomes 3, 4 and 8 and their association with sensory aroma scores

Sensory aroma score	Nature of aroma	No. of RILs	CHR3_22 (aro3.1)	RM5633 (aro4.1)	*badh2* (aro8.1)
0	No aroma	30	8/28 (28.5%)	9/29 (31%)	16/23 (69.6%)
1	Mild aroma	76	23/66 (34.8%)	33/75 (44%)	46/61 (75.4%)
2	Moderate aroma	67	26/56 (46.4%)	19/65 (29.2%)	40/50 (80%)
3	High aroma	7	6/7 (85.7%)	5/6 (83.3%)	6/7 (85.7%)
	Total	180	63/157 (40.1%)*	66/175 (37.7%)*	108/141 (76.6%)*

* excluding RILs with heterozygote and missing genotype data

Segregation of RILs for linked markers on aroma QTLs

Association between sensory aroma scores of RILs and their allelic composition was studied at the three reported aroma QTLs. The nearest linked markers from chromosomes 3 and 4 and the *badh2* gene locus from chromosome 8 were observed for the frequency of alleles contributed by the two diverse parents. All the markers segregated in the expected 1:1 ratio at a cutoff P value of 0.01. Segregation pattern of *badh2* alleles in the RILs was heavily distorted in favour of Pusa 1121 allele. More than 75% of the RILs with moderate or mild aroma possessed the Pusa 1121 allele. However, 69.6% of non-aromatic RILs were also found to possess the Pusa 1121 allele; this again re-enforces the hypothesis that *badh2* gene alone is not responsible for the overall aroma of rice (Table 2).

Differentially expressed genes between parents and RILs

To identify genes responsible for rice grain aroma, the observations of QTL mapping and fine mapping were used to analyze the differential expression of genes in highly aromatic and non-aromatic RIL bulks. The aromatic nature of parents and RILs was identified by sensory analysis of mature milled grains and then six highly aromatic and six non-aromatic RILs were pooled together to create aromatic and non-aromatic bulks, respectively. RNA was extracted from whole seedlings of parental lines and RILs for cRNA synthesis and hybridization with the Affymetrix Rice GeneChip carrying probes for 59,712 genes. Treating the expression level in non-aromatic parent, Pusa 1342 as base, 2273 gene probes were over expressed and 1743 gene probes were under expressed in the aromatic parent, Pusa 1121. Similarly, treating the expression levels in non-aromatic bulk as base, the corresponding number of over expressed and under expressed gene probes in the aromatic bulk came down to 931 and 804, respectively.

Co-localization of differentially expressed genes in QTL regions

A total of ten differentially expressed genes between the aromatic and non-aromatic RIL bulks were co-located in the three QTL regions mapped earlier by our research group. The random distribution of 10 genes in the whole rice genome of 389 Mbp predicts one gene per 38.9 Mbp of the present genetic map. Apart from the already reported *badh2* gene on chromosome 8, one differentially expressed gene was mapped on QTL *aro3.1* on chromosome 3 between markers CHR3_22 and CHR3_24, while eight genes were located on *aro4.1* on chromosome 4 between markers RM5633 and RM273. This cluster of eight genes hints at the possible presence of a yet unmapped QTL within this region. We observed a QTL with LOD score 2.02 in this region using composite interval mapping (CIM) function of Zmapqtl Model 6 of WinQTL Cartographer version 2.5. This QTL cannot be claimed to be significant as the LOD score was on the borderline of significance. However, this QTL region is co-localized with the *BADH1* gene which is anticipated to have a significant role in the expression of grain aroma.

S. No.	Affymetrix probe Id	Gene ID	Locus ID	Fold change value		Annotated gene function
				Microarray	qRT-PCR	
aro3.1						
1	Os.12186.1.S1_at*	Os03g0327600	LOC_Os03g21040	(–) 2.94	(–) 0.58	Hydroxyproline-rich glycoprotein family, stress responsive protein
aro4.1						
2	Os.25639.1.S1_at*	Os04g0352400	LOC_Os04g28420	(–) 17.87	(–) 6.18	FK506-binding protein with peptidyl-prolyl cis-trans isomerase activity
3	Os.26502.1.S1_a_at*	Os04g0434800	LOC_Os04g35520	(–) 2.66	(–) 1.11	L-ascorbate peroxidase activity
4	Os.11262.2.S1_x_at*	Os04g0401700	LOC_Os04g32920	(+) 2.19	(+) 0.94	HAK1 protein with potassium ion transmembrane transporter activity
5	Os.52307.1.S1_at*	Os04g0438300	LOC_Os04g35760	(+) 2.31	(+) 0.67	DUF150 family, ribosomal small subunit biogenesis
6	Os.47381.2.S1_at	Os04g0445700	LOC_Os04g36800	(+) 2.04	(+) 1.05	3-oxoacyl-(acyl-carrier protein) synthase family protein. Transferase activity, transferring acyl groups other than amino-acyl groups
7	Os.11920.1.S1_s_at	Os04g0468600	LOC_Os04g39320	(+) 2.12	(+) 0.53	Heavy Metal Associated (HMA) domain containing family II, Expressed protein
8	Os.22910.1.S1_at*	Os04g0474800	LOC_Os04g39880	(+) 10.70	(+) 3.59	BGLU-12 protein with glucan 1,3-beta-glucosidase activity
9	Os.52691.1.S1_at*	Os04g0469700	LOC_Os04g39430	(+) 5.98	(+) 1.13	Cytochrome P450 family, hormone synthesis, etc.
aro8.1						
10	Os.15306.1.S1_a_at*	Os08g0424500	LOC_Os08g32870	(–) 7.78	(–) 0.56	*badh2* gene having oxidoreductase activity

* genes commonly expressed in parents as well as RIL bulks

Annotation of differentially expressed genes co-located in QTL regions

Functional annotation of the ten differentially expressed genes co-located in the three aroma QTLs was analyzed in silico. The *badh2* gene (LOC_Os08g32870) on chromosome 8 was found to be down regulated in the aromatic parent and bulk. Among the other nine differentially expressed genes co-located in the grain aroma QTLs, a stress responsive protein (LOC_Os03g21040) in QTL *aro3.1* was constitutively down regulated in the aromatic parent and bulk. In the aroma QTL *aro4.1,* a FK506-binding protein with peptidyl-prolyl cis-trans isomerase activity (LOC_Os04g28420) and another protein with L-ascorbate peroxidase activity (LOC_Os04g35520) were down regulated in the aromatic parent and bulk, while six other genes were up regulated. These were, (i) an HAK1 protein with potassium ion transmembrane transporter activity (LOC_Os04g32920), (ii) a ribosomal protein with small subunit biogenesis (LOC_Os04g35760), (iii) a synthase protein transferring acyl groups other than amino-acyl groups (LOC_Os04g36800), (iv) an expressed protein belonging to heavy metal associated domain containing family 2 (LOC_Os04g39320), (v) a BGLU-12 protein with glucan 1,3-beta-glucosidase activity (LOC_Os04g39880), and (vi) an expressed protein belonging to cytochrome P450 family (LOC_Os04g39430). Differential expression of all these genes was also validated by qRT-PCR; the results of two experiments were consistent except that the magnitude of differential expression was comparatively lower in case of RT-PCR (Table 3).

Discussion

Identification of candidate genes related to various traits like grain features and cooking qualities, drought and salinity tolerance, etc. in crops has been a matter of serious investigation for last many decades. In many crops, QTL intervals related to different quality traits have been identified through different mapping studies using a variety of mapping populations like F_2, RILs, NILs, etc. Even after genetic fine mapping, the number of genes in a 1 cM QTL interval may range from 25–30 (IRGSP 2005). With such a large number of genes it becomes extremely difficult to predict which or how many of these might be actually responsible for the functional polymorphism of the trait under observation. Combining conventional genetic fine mapping techniques with expression profiling has emerged as a new and powerful tool to identify genes underlying various traits in a variety of organisms like maize, rice, drosophila, etc.

Three QTLs for rice grain aroma have been mapped and fine mapped, one each on the short arm of chromosome 3 *(aro3.1),* long arm of chromosome 4 *(aro4.1)* and long arm of chromosome 8 *(aro8.1).* Though a recessive allele of *BADH2* gene *(badh2)* is said to be the major gene responsible for rice grain aroma, the contributory role of two QTLs has also been proven. With this information available, locating differentially expressed genes co-localized in these QTLs was easier. The same RIL population which was used to identify and fine map these QTLs was used for generating the transcriptome profiles. RIL population has several advantages over F_2 or F_3 populations in mapping of QTLs controlling grain quality. First, the homozygous genotype of each RIL avoids the difficulty associated

with triploid endosperm and dosage effects. This simplifies QTL mapping of grain quality free from the maternal effects. RILs also make it possible to replicate the phenotypic measurements, so that even minor genes controlling grain quality can be detected. In addition, each RIL can supply enough genetically identical seeds for grain quality analysis requiring large sample size. Highly aromatic and non-aromatic RILs were identified on the basis of repeated sensory analysis of milled fully mature rice grains. Despite being subjective and non-quantitative in nature, sensory analysis is the most economical, convenient and user friendly qualitative method available for primary screening of rice samples for the presence and nature of their aroma (popcorn-like, fruity, nutty, etc.). The 180 RILs ranged from non-aromatic to highly aromatic types. Even within the aromatic types, the RILs differed in the intensity of aroma which is due to qualitative and quantitative differences in the nature and concentration of different compounds. Variations in some RILs were observed over a period of three years, which is due to various environmental factors like temperature and thermal variance, humidity and day/night durations.

RILs of similar phenotypes were bulked together using bulk segregant analysis approach which significantly reduced the number of samples to be analyzed to four; aromatic parent, non-aromatic parent, aromatic bulk and non-aromatic bulk. This approach drastically reduced the number of differentially expressed genes to about threefold from 4016 to a low of 1344. Due to the availability of whole genome rice gene chip arrays, it was easy to study the expression of all the genes that were commonly up or down regulated in the RIL bulks. Ten of these genes were located in the mapped QTL intervals for grain aroma on chromosomes 3, 4 and 8. The functional annotation of all these genes was searched in silico (Table 3). Of the three mapped aroma QTLs, *aro8.1* is the most significant and is represented by a non-functional (recessive) allele of the *BADH2* gene coding for enzyme betaine aldehyde dehydrogenase. In the present study, the *badh2* gene (LOC_Os08g32870) was found to be down regulated in both aromatic parent as well as aromatic bulk. This gene has an oxido-reductase activity, acting on the aldehyde or oxogroup of donors with NAD or NADP as acceptor. It is responsible for the synthesis of 2-acetyl-1-pyrroline (2-AP), the major rice aroma compound.

The QTL *aro3.1* is known to be a minor QTL for grain aroma sharing a complementary role with the other QTL on chromosome 4 (*aro4.1*). Earlier Singh et al. (2007) reported 51 genes in this QTL region flanked by the markers CHR3_22 and CHR3_24. However by bulk segregant analysis approach, a single gene (LOC_Os03g21040) was found to be constitutively down regulated in both aromatic parent as well as aromatic bulk. This gene belongs to the hydroxyproline-rich glycoprotein family and is expressed under abiotic stress. Eight differentially expressed genes were found to be co-located in the 10.8 Mb interval of *aro4.1*, flanked by markers RM5633 and RM273. The down regulated LOC_Os04g28420 gene is a FK506 binding protein (also known as FKBP5) with peptidyl-prolyl cis-trans isomerase activity. The protein encoded by this gene is a member of the immunophilin protein family, which plays a key role in basic cellular processes involving protein folding and trafficking. The second gene with down regulated expression LOC_Os04g35520 belongs to the L-ascorbate peroxidase subfamily 7 and has heme and metal ion binding activity. It is known to play an important role in ROS detoxification/removal of hydrogen

peroxide (Senadheera et al. 2012). This gene also has its orthologs in *Arabidopsis thaliana*. This protein is expressed in the aerial vegetative parts and reproductive tissues. It binds one cation per subunit; probably K^+, but might also be Ca^{2+}. Six other genes identified in this QTL region were found to be up-regulated in the aromatic parent and bulk. Among these, LOC_Os04g32920, a transmembrane protein belongs to the HAK1 protein family having similarity with K^+ uptake permease (KUP) transporter proteins. LOC_Os04g35760 is a ribosomal biogenesis protein showing homology to the DUF150 family. The DUF150 proteins are present in large number of species from archaea to human but their exact function is not known. Another gene LOC_Os04g36800 belongs to 3-oxoacyl-(acyl-carrier protein) synthase family. This transferase gene is capable of transferring acyl groups other than amino-acyl groups and catalyzes the condensation reaction of fatty acid biosynthesis. The oxido-reductase gene LOC_Os04g39880 belongs to 2OG-Fe(II) oxygenase family having beta-glucosidase activity (a BGLU-12 protein). It catalyzes an oxidation-reduction (redox) reaction in which hydrogen or electrons are transferred from 2-oxoglutarate (2OG) and one other donor and one atom of oxygen is incorporated into each donor. The gene LOC_Os04g39430 belonging to the cytochrome P_{450} family is involved in plant hormone synthesis, phytoalexin synthesis, flower petal biosynthesis and herbicide degradation. Such proteins are known to work by activating molecular oxygen by inserting one of its atoms into the substrate and reducing the other to form water. The sixth gene LOC_Os04g39320 belongs to heavy metal associated (HMA) domain containing family II. The HMA domain contains two conserved cysteines that interact selectively and non-covalently with any metal ion.

The rationale of the present investigation was to complement QTL mapping studies by transcriptome profiling of bulked RILs. The purpose of transcriptome profiling was to normalize the irrelevant background differential expression of genes between the aromatic and non-aromatic parent while retaining the differential expression of genes relevant for grain aroma in rice. A drastic reduction in the number of differentially expressed genes was observed using the bulk segregant approach, ten of which were co-located in the three aroma QTLs. A detailed insight into these genes will be of great importance in unravelling the complex mystery of rice aroma and also in elucidating the role of the minor aroma QTLs. Further validation of these differentially expressed genes, particularly in QTL *aro4.1* on chromosome 4, by genetic fine mapping and transformation techniques will lead to the development of new gene-based polymorphic functional markers for applications in various molecular breeding programs.

Data Submissions

Microarray data generated from this study has been deposited at NCBI Gene Expression Omnibus with accession number GSE47812 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47812).

Acknowledgements

References

Ahn, S.N., Bollich, C.N., Tanksley, S.D. 1992. RFLP tagging of a gene for aroma in rice. Theor. Appl. Genet. **84**:825–828.

Aitman, T.J., Glazier, A.M., Wallace, C.A., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., Al-Majali, K.M., Trembling, P.M., Mann, C.J., Shoulders, C.C., Graf, D., St. Lezin, E., Kurtz, T.W., Kren, V., Pravenec, M., Ibrahimi, A., Abumrad, N.A., Stanton, L.W., Scott, J. 1999. Identification of Cd36 fat as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. Nat. Genet. **21**:6–83.

Amarawathi, Y., Singh, R., Singh, A.K., Singh, V.P., Mohapatra, T., Sharma, T.R., Singh, N.K. 2008. Mapping of quantitative trait loci for basmati quality traits in rice. Mol. Breed. **21**:49–65.

Benjamini, Y., Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. of the Royal Statistical Soc., Series B **57**:289–300.

Bradbury, L.M., Fitzgerald, T.L., Henry, R.J., Jin, Q., Waters, D.L. 2005. The gene for fragrance in rice. Plant Biotech. J. **3**:363–370.

Chen, S., Wu, J., Yang, Y., Shi, W., Xu, M. 2006. The *fgr* gene responsible for rice fragrance was restricted within 69 kb. Plant Sci. **171**:505–514.

Deshmukh, R., Singh, A., Jain, N., Anand, S., Gachche, R., Singh, A., Gaikwad, K., Sharma, T., Mohapatra, T., Singh, N. 2010. Identification of candidate genes for grain number in rice (*Oryza sativa* L.) Funct. Integr. Genomics **10**:339–347.

Fitzgerald, M.A., Hamilton, N.R.S., Calingacion, M.N., Verhoeven, H.A., Butardo, V.M. 2008. Is there a second fragrance gene in rice? Plant Biotechnol. J. **6**:416–423.

IRGSP 2005. The map based sequence of the rice genome. Nature **436**:793–800.

Jodon, N.E. 1944. The inheritance of flower fragrance and other characters in rice. J. Amer. Soc. Agron. **36**:844–848.

Kadam, B.S., Patankar, V.K. 1938. Inheritance of aroma in rice. Chron. Bot. **4**:32.

Kathiresan, A., Lafitte, H.R., Chen, J., Mansueto, L., Bruskiewich, R., Bennett, J. 2006. Gene expression microarray and their application in drought stress research. Field Crops Res. **97**:101–110.

Lorieux, M., Petrov, M., Huang, N., Guiderdoni, E., Ghesquière, A. 1996. Aroma in rice: Genetic analysis of quantitative trait. Theor. Appl. Genet. **93**:1145–1151.

Marino, R., Ponnaiah, M., Krajewski, P., Frova, C., Gianfranceschi, L., Pe, E.M., Gorla, S.M. 2009. Addressing drought tolerance in maize by transcriptional profiling and mapping. Mol. Genet. Genomics **218**:163–179.

Murray, M.G., Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acid Res. **8**:4321–4325.

Pachauri, V., Singh, M.K., Singh, A.K., Singh, S., Shakeel, N.A., Singh, V.P., Singh, N.K. 2010. Origin and genetic diversity of aromatic rice varieties, molecular breeding and chemical and genetic basis of rice aroma. J. Pl. Biochem. Biotechnol. **19**:127–143.

Pandit, A., Rai, V., Bal, S., Sinha, S., Kumar, V., Chauhan, M., Gautam, R.K., Singh, R., Sharma, P.C., Singh, A.K., Gaikwad, K., Sharma, T.R., Mohapatra, T., Singh, N.K. 2010. Combining QTL mapping and transcriptome profiling of bulked RILs for identification of functional polymorphism for salt tolerance genes in rice (*Oryza sativa* L.) Mol. Genet. Genomics **284**:121–136.

Pinson, S.R.M. 1994. Inheritance of aroma in six rice cultivars. Crop Sci. **34**:1151–1157.

Price, A. 2006. Believe it or not, QTLs are accurate! Trends Plant Sci. **11**:213–216.

Salvi, S., Tuberosa, R. 2005. To clone or not to clone plant QTLs: Present and future challenges. Trends Plant Sci. **10**:297–304.

Senadheera, P., Tirimanne, S., Maathuis, F.J.M. 2012. Long term salinity stress reveals cultivar specific differences in root oxidative stress response. Rice Sci. **19**:212–221.

Singh, V.P., Singh, A.K., Atwal, S.S., Joseph, M., Mohapatra, T. 2002. Pusa 1121: A rice line with exceptionally high cooked kernel elongation and basmati quality. Int. Rice Res. Notes **27**:25–26.

Singh, R., Singh, A.K., Sharma, T.R., Singh, A., Singh, N.K. 2007. Fine mapping of aroma QTLs in Basmati rice *(Oryza sativa)* on chromosomes 3, 4 and 8. J. Pl. Biochem. Biotechnol. **16**:75–82.

Singh, A., Singh, P.K., Singh, R., Pandit, A., Mahato, A.K., Gupta, D.K., Tyagi, K., Singh, A.K., Singh, N.K, Sharma, T.R. 2010. SNP haplotypes of the BADH1 gene and their association with aroma in rice (*Oryza sativa* L.). Mol. Breed. **26**:325–338.

Sood, B.C., Sidiq, E.A. 1978. A rapid technique for scent determination in rice. Indian J. Genet. Pl. Breed. **38**:268–271.

Wanchana, S., Kamolsukyunyong, W., Ruengphayak, S., Toojinda, T., Tragoonrung, S., Vanavichit, A. 2005. A rapid construction of a physical contig across a 4.5cM region for rice grain aroma facilitates marker enrichment for positional cloning. Sci. Asia **31**:299–306.

Wayne, M.L., McIntyre, L.M. 2002. Combining mapping and arraying: An approach to candidate gene identification. Proc. Natl Acad. Sci. USA **99**:14903–14906.