# In Silico Identification of Rice Gene Homologues in Brachypodium, Sorghum and Maize: Insight into Development of Gene Specific Markers

**Corresponding Author:**
Dr. Rupesh K Deshmukh,
Researcher, School of Computational Sciences, Swami Ramanand Teerth Marathwada University - India

**Submitting Author:**
Mr. Krishnanand P Kulkarni,
Research Scholar, School of Life Sciences, SRTMU, Nanded - India

**Competing Interests:**

We do not have any competing interests.

**Additional Files:**

Supplementary_Material

# In Silico Identification of Rice Gene Homologues in Brachypodium, Sorghum and Maize: Insight into Development of Gene Specific Markers

**Author(s):** Kulkarni KP, Kulkarni SS, Gedda M, Bandevar M, Sonah H, Gacche RN, Deshmukh NK, Deshmukh RK

## Abstract

Rice, one of the important cereal species, is considered as a model for studying grass family. Whole rice genome is recently sequenced, fully annotated and hundreds of genes are characterized. Molecular resources in rice are ever increasing since the draft became publicly available and the information can be utilized to study other cereal genomes as well. This study presents a comprehensive identification of rice genes for important traits like biotic and abiotic stress tolerance and grain yield and design of gene-specific PCR-based markers. A total 925 functionally characterized genes for were identified through literature search and subjected for detection of their homologue counterparts in three different cereal species. The identified homologs found to be syntenic in their relationship with rice and appear to have shared a common ancestor. These selected rice genes were used to design gene-specific markers which have power of early prediction of the phenotype and can be exploited in marker assisted breeding and crop improvement programs. For designing gene-specific markers, sequences of homologous genes were mined for presence of microsatellite repeat motifs and primers were designed flanking the repeat motifs. We believe that these markers will find great usage in phenotyping for stress tolerance or any other trait of interest in related cereal crops and in selecting genotypes in breeding program aimed at transferring a particular gene in an elite genetic background.

## Introduction

Rice *(Oryza sativa L.)* is the most important cereal crop of the developing world feeding more than two billion people as a staple food. It has also emerged as a model crop for comparative genomic and molecular biological studies because of availability of its complete genome sequence (Bennetzen, 2007; Devos, 2010). Besides, re-sequencing efforts using gene-chip technology have added a certain value to the reference genome sequence by uncovering the wealth of genomic variation (Jackson et al., 2006). Together, sequencing and resequening data has been accumulating very fast thereby assisting in the identification of agronomically important genes in cereal crops like maize, wheat, oats, sorghum and millet. This enormous progress in rice genomics has made it possible structural and functional comparisons of genes involved in various biological and developmental processes in rice and other cereals (Shimamoto and Kyozuka, 2002). It has been observed through comparative mapping analysis of close related grass species that the few markers are very well conserved within large chromosomal segments of the members of grass family genomes (Devos and Gale, 1997; Gale and Devos, 1998). These conserved regions have been thought to be derived from a common ancestor and to be collinear (Keller and Feuillet, 2000). Colinearity has also been observed in more distant species, giving space for comparison of gene organization in grass genomes. Several researchers made comparative mapping studies using a defined set of markers or probes to evaluate degree of colinearity among different species. In an experiment, Varshney et al. mapped EST-SSRs to a rye genetic linkage map, which positioned in the expected orthologous region compared to their position in barley (Varshney et al., 2005). Such studies have shown the application of comparative mapping strategies in identifying useful genes and inferring the basic elements of genome evolution. Furthermore, these studies can guide in identifying the novel alleles of genes of interests among close as well as distant species so that the best alleles can be utilized in crop improvement programs.

Sequence comparison methods brought into play extensively since the confirmation that human and mouse genomes are 80% similar at genomic level. Data mining and sequence comparison method using both *indica* and *japonica* strains have helped in identifying several hundred Cyt P450 genes in rice and have been analyzed before the rice genome sequence made publicly available (Nelson et al., 2004). Several gene families have been identified in grasses through sequence comparisons and evolutionary relationships

have been established. Sequence comparisons have helped in identifying the putative homologs (*hd1, LFY, FTL1, etc*) and orthologs (*AGA-MOUS, MADS-Box genes etc*) of several important genes in rice (Kyozuka and Shimamoto, 2002; Kyozuka et al., 2000; Rao et al., 2008). Conversely, rice homologs have also been identified in other related cereals such as maize (*Zea mays*) and barley (*Hordeum vulgare*). Low Silicon Rice 2 (Lsi2)-like Si efflux transporters have been identified in maize and barley which may have different Si uptake system from that in rice (Mitani et al., 2009b). These results show that comparative analysis of expression patterns of genes in rice and *other crops may* suggest various similar to distinct roles played by those genes in respective plants. With the availability of huge sequence resources in cereal crops, comparative analysis may provide insights into the structural and functional details of various gene families and may find valuable in future research and breeding programs.

Few of the recent mapping experiments have revealed that the comparative studies might be more difficult for those genes like leaf resistance genes in wheat that evolve rapidly. Chromosome rearrangements like inversions, translocations and insertions may not specify the evolutionary relationships and may not help in defining the gene order (Keller and Feuillet, 2000). To circumvent such exceptions, it becomes necessary to use markers from coding part of genome and therefore it is essential to have a large number of markers from conserved regions of genomes. A limited gene transfer among grasses using sexual crosses has added a number of important traits to the modern cereal cultivars. However, the vast majority of genes and traits in single grass species have never been used in other species. Besides, key genes and quantitative trait loci (QTL) for several important traits, such as plant height, flowering time and shattering show orthologous relationships in barley, wheat, maize and rice. Gene-specific markers can be designed which can be used in phenotyping of plants for various agronomically important traits. In this report we have identified 925 rice homologs in three cereal crops viz. Brachypodium, Maize and Sorghum and gene-specific PCR-based markers were designed.

# Methods

Function annotated and well characterized rice genes were searched from literature and public databases. All the DNA sequences related to the gene loci were retrieved in FASTA format using the Batch Entrez retrieval system from GenBank and BLASTn

tool (http://www.ncbi.nlm.nih.gov/blast) against TIGR gene models to get locus identifier (http://rice.plantbiology.msu.edu/). The coding DNA sequence (CDS) corresponding to the selected rice genes were retrieved from TIGR. The CDS sequences of Sorghum, Maize, and Brachypodium were downloaded from Phytozome (http://www.phytozome.org/) and subsequently used for the development of local database named as 'Bramas'. Homologous genes in these three cereal species were searched using rice CDS as query sequences against Bramas by BLASTn with cutoff bit score >200 and e-value ofwww.blast2go.org/). The sequences were annotated for biological process, cellular component and molecular function and plotted using Web Gene Ontology Annotation (WEGO) (Ye et al., 2006).

**Comparative physical mapping**

The syntenic relationships among these cereal species for the homologous genes were verified manually in Microsoft Excel using the genomic location information of these genes. All the genes were mapped onto respective chromosomes using Mapchart 2.1 (http://www.biometris.wur.nl/uk/Software/MapChart/)

**Detection of microsatellite repeats motifs in homologous genes and primer design**

Mining of di-, tri-, tetra-, penta-, and hexanucleotide perfect microsatellite repeat motifs as well as non-interrupted and interrupted compound microsatellites from these homologous sequences was carried out using a Perl-based program, MISA (MIcroSAtellite identification tool; http://pgrc.ipk-gatersleben.de/misa). MISA is a freely downloadable Perl script available on internet. Primer pairs flanking the microsatellite repeat motifs were developed using Primer3 Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer 3plus.cgi) software tool. The parameters were changed according to own interest. The primer size parameter was changed to min-17, opt-21 and max-27. The GC% was changed to min-45%, max-65%. Then the SSRs were searched for both forward and reverse primers.

# Results and Discussion

Grasses has been the most studied family in plant kingdom. Recent progress in the high-throughput sequencing and genotyping technologies have helped in comparative analysis of genomes of related species. These revolutions facilitated the gene discovery and have helped in revealing associations between

genotype and phenotype. Because of large amount of sequence availability in almost every cereal crop, homolog-based gene identification has taken much interest of researchers of this generation. Genes having homologs in related species have been identified in different species such as Arabidopsis, rice, cotton, maize, barley and grapevine (Fulton et al., 2002; Mitani et al., 2009a; Park et al., 2010; Shelden et al., 2009). The genes with shared homology have advantage that those can be used for phenotyping of different traits in related crops. Keeping this in view, we identified rice homologs in brachypodium, maize and sorghum and designed Simple Sequence Repeat (SSR)-based gene-specific markers for them in the present study.

Well characterized rice genes were thoroughly searched in literature and found 925 genes which were used for finding their homologous counterparts. The gene length varied from 573bp to 5658bp with average length of 1529. Their annotation and functional characterization was done using an offline tool Blast2GO (www.blast2go.org/) to have an account of functional role of selected genes in various developmental and cellular mechanisms. Out of 925 selected rice genes, 534 (57.73%) had matches to genes or sequences of known function. These were assigned to the one of the three categories and plotted using WEGO (Illustration 1). A major proportion of the assigned loci represent genes that appear to be involved in several metabolic processes, such as energy-generating processes and the biosynthesis and deprivation of cellular building blocks. Thus, it is evident from this study that the selected genes signify various aspects of cellular metabolism and have remained highly conserved across plant species (Yang et al., 1999). We mapped these genes onto 12 rice chromosomes and found that they formed clusters in different regions of the chromosomes (Illustration 2). This was according to the expectations as there have been reports of genes present in clusters and gene-rich hotspots are have been documented in various cereal genomes. These sets of conserved sequences in a genome can find better application in comparative genome analyses and to provide information the evolutionary aspects of convergence or divergence of a species.

Both comparative mapping and sequence comparisons among the cereal crops have indicated that sequencing and functional analysis of the rice and other sequenced genomes will have a significant impact in terms of gene identification and crop improvement. Gene identification through map-based cloning approach has been difficult in cereals and

therefore we must look for other approaches that are quite feasible. Comparative mapping approaches can help to identify genes in organism where genomes sequence is available. However this would require large set of markers, especially those ones which are tightly linked to a phenotype of desired interest and which have been relatively stable in both sequence and copy number. Different dominant and codominant marker systems have been applied in mapping, marker assisted breeding and germplasm evaluation experiments (Xu et al., 2005). Thus it is necessary to have markers based on sequences conserved in most of the cereal crops. SSRs have been the preferred markers as it is easy to identify different repeat motifs in a complete as well as partial genome sequences and to design a marker flanking that repeat motif. Therefore we mined selected genes for SSR motifs and found all types of di-, tri-, tetra-, penta-, and hexanucleotide perfect microsatellite repeat motifs. Non-interrupted and interrupted compound microsatellites have also been observed, although in lesser numbers. Marker were designed flanking these repeat motifs and keeping their product size range in between 100-300bp (Supplementary file-1). Of about 89% of the repeat motifs were tri-nucleotide repeats and di-nucleotide repeats were least in numbers. Tri-nucleotide repeats have been found in very large numbers inside these genes because they have less selection pressure, compared to other nucleotide repeats. The positional relationships of these loci were evident when we analyzed the physical positions of these loci in all the species under consideration and found that these are syntenic to each other. Particulars of syntenic relevance of these loci for rice chromosome 1 with other systems are shown in the illustration 3. We believe that these genes may help in comparative genome analyses as well as in phenotyping for agronomically important traits.

## Conclusions

We have analyzed 925 rice genes for revealing the presence of their homologue counterparts in three different cereal crop systems. As the genomes of all these crops systems have been sequenced completely, it has made possible to analyze sequence similarity searches and thereby identifying conserved blocks in coding parts of genes. The conserved regions in genomes are assumed to have come from a common ancestor and they may have less sequence variation. Therefore markers or probes designed using such sequences can be used in other crops systems where their homologues are present. In summary, we

believe that the PCR-based gene-specific markers designed in this study can be utilized in various crop improvement programs.

# References

1. Bennetzen, J.L., 2007: Patterns in grass genome evolution. Curr Opin Plant Biol 10, 176-81.

2. Devos, K.M., 2010: Grass genome organization and evolution. Curr Opin Plant Biol 13, 139-45.

3. Devos, K.M., and M.D. Gale, 1997: Comparative genetics in the grasses. Plant Mol Biol 35, 3-15.

4. Fulton, T.M., R. Van der Hoeven, N.T. Eannetta, and S.D. Tanksley, 2002: Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14, 1457-67.

5. Gale, M.D., and K.M. Devos, 1998: Comparative genetics in the grasses. Proc Natl Acad Sci U S A 95, 1971-4.

6. Jackson, S., S. Rounsley, and M. Purugganan, 2006: Comparative sequencing of plant genomes: choices to make. Plant Cell 18, 1100-4.

7. Keller, B., and C. Feuillet, 2000: Colinearity and gene density in grass genomes. Trends Plant Sci 5, 246-51.

8. Kyozuka, J., and K. Shimamoto, 2002: Ectopic expression of OsMADS3, a rice ortholog of AGAMOUS, caused a homeotic transformation of lodicules to stamens in transgenic rice plants. Plant Cell Physiol 43, 130-5.

9. Kyozuka, J., T. Kobayashi, M. Morita, and K. Shimamoto, 2000: Spatially and temporally regulated expression of rice MADS box genes with similarity to Arabidopsis class A, B and C genes. Plant Cell Physiol 41, 710-8.

10. Mitani, N., N. Yamaji, and J.F. Ma, 2009a: Identification of maize silicon influx transporters. Plant Cell Physiol 50, 5-12.

11. Mitani, N., Y. Chiba, N. Yamaji, and J.F. Ma, 2009b: Identification and characterization of maize and barley Lsi2-like silicon efflux transporters reveals a distinct silicon uptake system from that in rice. Plant Cell 21, 2133-42.

12. Nelson, D.R., M.A. Schuler, S.M. Paquette, D. Werck-Reichhart, and S. Bak, 2004: Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiol 135, 756-72.

13. Park, W., B.E. Scheffler, P.J. Bauer, and B.T. Campbell, 2010: Identification of the family of aquaporin genes and their expression in upland cotton (Gossypium hirsutum L.). BMC Plant Biol 10, 142.

14. Rao, N.N., K. Prasad, P.R. Kumar, and U. Vijayraghavan, 2008: Distinct regulatory role for RFL, the rice LFY homolog, in determining flowering time and plant architecture. Proc Natl Acad Sci U S A 105, 3646-51.

15. Shelden, M.C., S.M. Howitt, B.N. Kaiser, and S.D. Tyerman, 2009: Identification and functional characterisation of aquaporins in the grapevine, Vitis vinifera functional plant biology 36, 1065–1078.

16. Shimamoto, K., and J. Kyozuka, 2002: Rice as a model for comparative genomics of plants. Annu Rev Plant Biol 53, 399-419.

17. Varshney, R., R. Sigmund, A. Borner, V. Korzun, N. Stein, M. Sorrells, P. Langridge, and A. Graner, 2005: Interspecific transferability and comparativemapping of barley EST-SSR markers in wheat, rye and rice. plant science 168, 195-202.

18. Xu, Y., S. McCouch, and Q. Zhang, 2005: How can we use genomics to improve cereals with rice as a reference genome? Plant Mol Biol 59, 7-26.

19. Yang, Y.W., K.N. Lai, P.Y. Tai, and W.H. Li, 1999: Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other lineages. Journal of Molecular Evolution 48, 597-604.

20. Ye, J., L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, J. Wang, S. Li, R. Li, and L. Bolund, 2006: WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34, W293-7.

# Illustrations

## Illustration 1

Gene Ontology (GO) categories of 925 rice genes after BLASTx search: Histogram presentation of the GO annotation was generated automatically by the web histogram tool WEGO
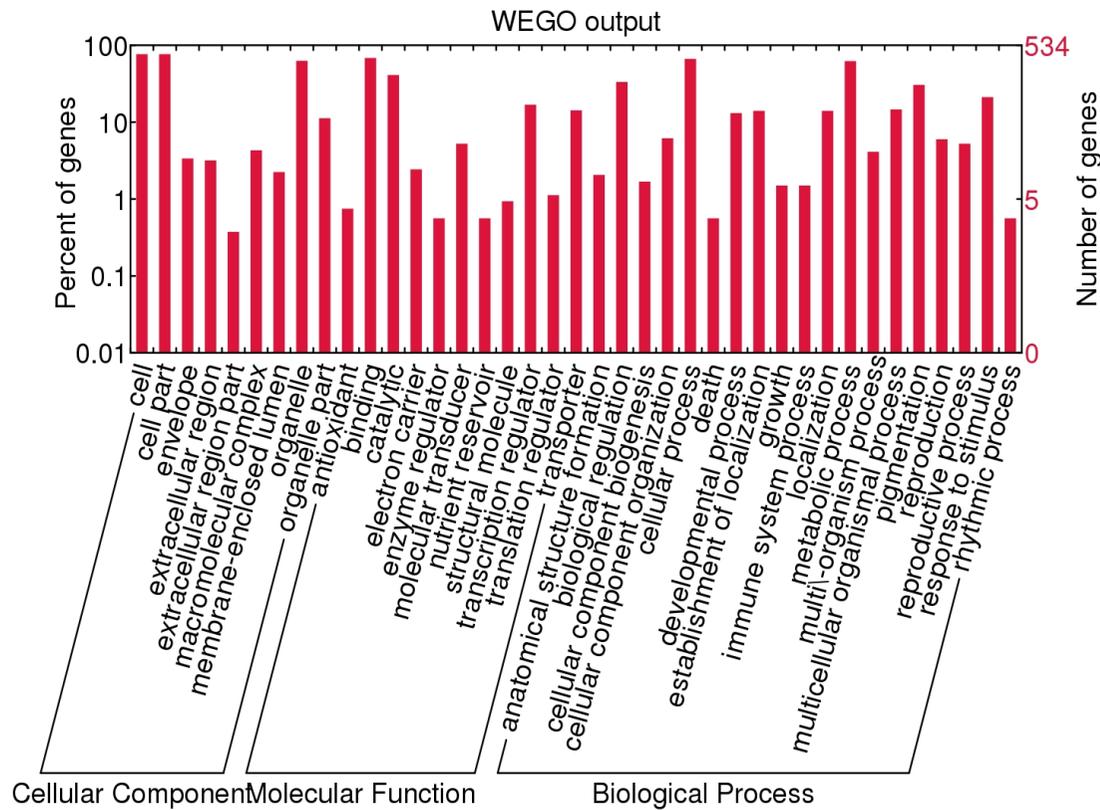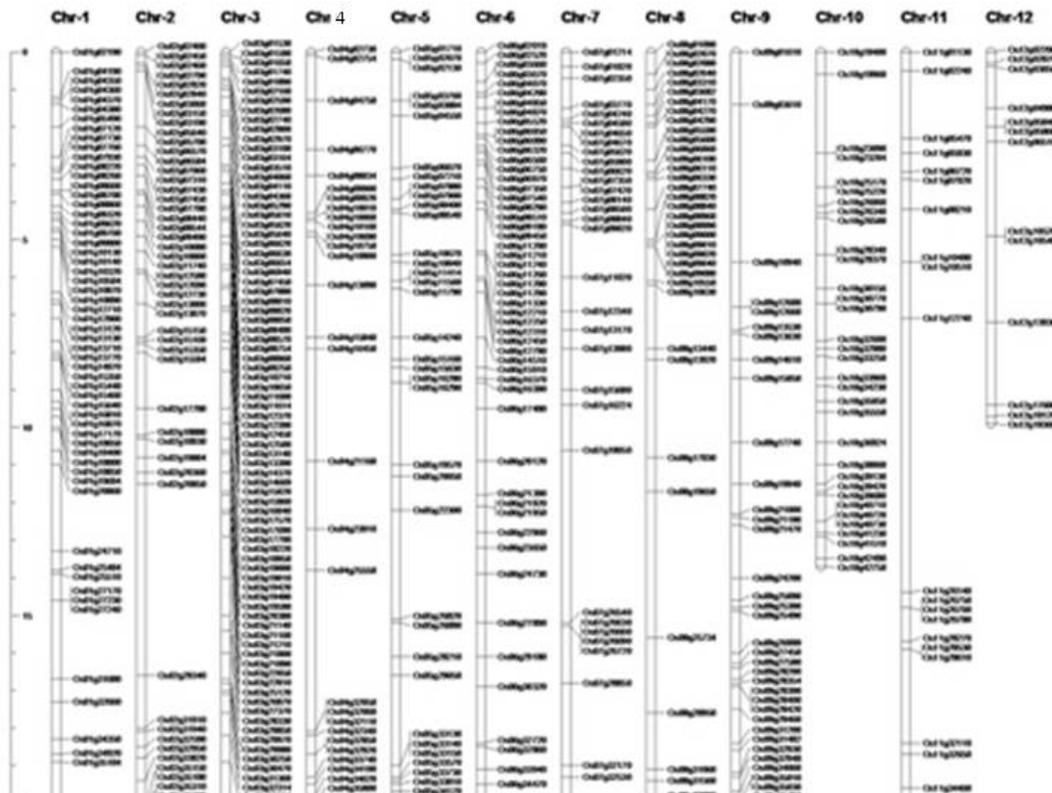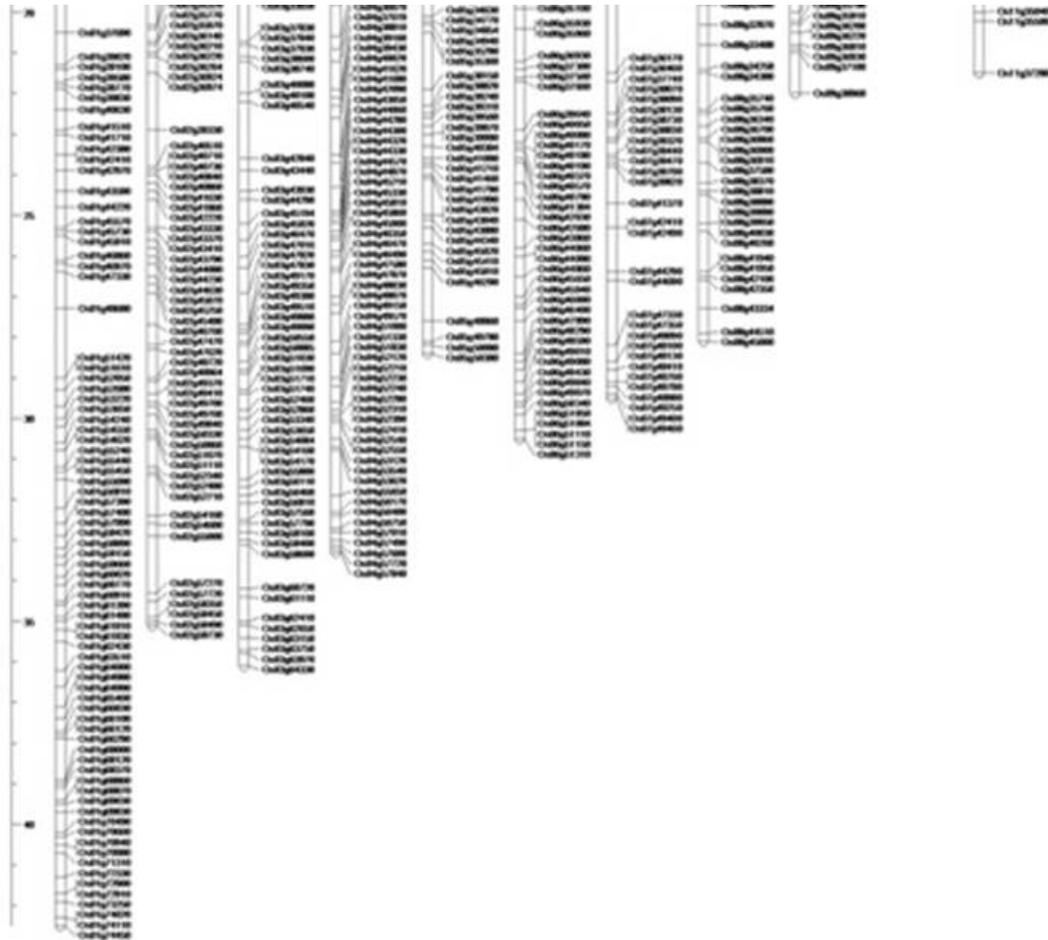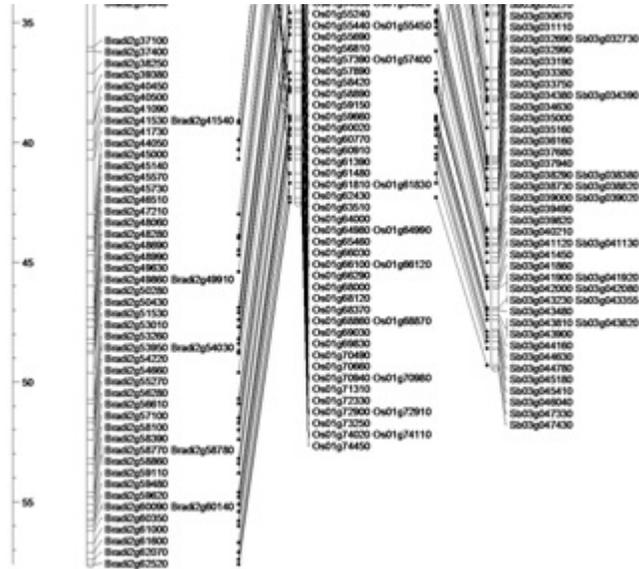
# Illustration 2

A map showing physical positions of the selected rice genes on 12 rice chromosomes

## Illustration 3

Comparative physical mapping of rice chromosome1, with chromosome2 of Brachypodium and Chromosome3 of Sorghum, based on syntenic loci.

## Disclaimer

This article has been downloaded from WebmedCentral. With our unique author driven post publication peer review, contents posted on this web portal do not undergo any prepublication peer or editorial review. It is completely the responsibility of the authors to ensure not only scientific and ethical standards of the manuscript but also its grammatical accuracy. Authors must ensure that they obtain all the necessary permissions before submitting any information that requires obtaining a consent or approval from a third party. Authors should also ensure not to submit any information which they do not have the copyright of or of which they have transferred the copyrights to a third party.

Contents on WebmedCentral are purely for biomedical researchers and scientists. They are not meant to cater to the needs of an individual patient. The web portal or any content(s) therein is neither designed to support, nor replace, the relationship that exists between a patient/site visitor and his/her physician. Your use of the WebmedCentral site and its contents is entirely at your own risk. We do not take any responsibility for any harm that you may suffer or inflict on a third person by following the contents of this website.