Asian Journal of Biochemistry

ISSN 1815-9923

Academic
Journals Inc.

www.academicjournals.com

Asian Journal of Biochemistry 1 (2): 194-210, 2006
ISSN 1815-9923

Genomics in Rice Improvement

[1]Parvez Sofi and [2]A.R. Trag
[1]Division of Plant Breeding and Genetics, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar 191 121, Srinagar
[2]Director Research, SKUAST-K, Shalimar 191 121, Srinagar, India

Abstract: Rice has been identified as a model crop for genomic research not only because of its economic importance but also desirable attributes of its genome. Great advances have been made in structural and functional gernomics analysis under IRGP. The working draft has been published in 2002 and a finished draft is due to be ready by 2005. The availability of high througput technologies has greatly facilitated the sequencing of rice genome. Advances in proteomics, gene chips and microarrays have facilitated functional validation of rice gemnome by analysing gene expression profiles. Due to great synteny in grass family in gene order and content, the results can successfully be applied to other members with complex genomes besides other orphan crops. The genomics research will usher a new era in rice improvement by facilating large scale gene discovery, development of transgenics and marker assisted selection besider unravelling many critical phenomenon such as gene expression, interaction and genetic basis of heterosis thus in coming years genomics is surely going to have a major impact on pace of rice improvement

Key words: Genomics, sequencing, rice, proeomics, gene chips, microarrays

Introduction

The Genomic Era in rice started with the development of first RFLP map in rice by McCouch *et al*. (1988). The human genome project accelerated the development and refinement of techniques of nucleic acid sequencing which made it possible to sequence entire genomes of organisms with greater accuracy and speed. For many experts, the sequencing of rice genome may have more societal and scientific ramifications than finishing the human genome map. Donald Kennedy, editor of Science says: over the next 20 years, the rice genome will make more of a difference to global health than human genome we published a year ago.

Genomics is the study of how genes and genetic information is organized within genomes (the entire genetic complement of an organism) and how this organization determines their function in a given organism-plant or animal. Every living organism possesses a genome that contains the biological information needed to construct and maintain a living example of that species. Thus genomic research in any species is mapping, sequencing and functional analysis of all the different coding and non-coding sequences (Ghosh *et al*., 2000). The basic activity of genomics can be broadly classified into three categories.

- Structural genomics: which studies the organization of genes in genomes and the structure of all proteins encoded by DNA sequences.

Corresponding Author: Parvez Sofi, Division of Plant Breeding and Genetics, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar 191 121, Srinagar, India

- Functional genomics: Which studies the function and interaction of genes because of such organization, their effect and expression in time and space and explains why and how a given gene behaves in a certain manner under certain situations. Functional genomics also studies protein functions of an organism (Proteomics).
- Comparative genomics: Which studies the synteny relationships between genomes of related organisms for possible gene exploration and transfer within related groups.

Genomics research is profoundly altering the way the biologists think about plant systems. It seeks to analyse how genes are structured, how they function and how they evolved (McCouch, 1998). The analysis is based on sophisticated computations of vast pool of genetic information generated by genomic research into easily accessible form (Bioinformatics). Genomics as science has the potential to unify disparate kingdoms, build bridges across ancient genetic voids and establish new channels across species-specific domains. In light of deteriorating biophysical resource base and consequently high pressure on existing natural resources for agriculture, genomic research is going to be the Thrust area of scientific investigation. The interpretations of insights provided by genomics will not only change our existing knowledge but also dictate the methods we employ for crop improvement especially in terms of their breeding value. Genomics put to practice has changed the whole spectrum of breeding methodologies. Till now we would study the phenotype and predict the genotype but now we are dissecting the genotype into the minutest possible details and predicting the phenotype of plants.

Rice as a Model Plant for Genomic Research

The suitability of an organism for efficient genomic studies depends on a number of factors such as genomic size (total genomic DNA content), ploidy level, chromosome number , total recombination distance in genome, nature and extent of repetitive DNA sequences, number of functional genes and flexibility of genetic manipulation through genetic transformation. Rice has been on the forefront of genomic research in cereal crops because of following characteristics:

- Small genomic size: (450 Mbp): Rice genome is three time that of *Arabidopsis* but only 1/5th of that maize though the number of chromosomes in rice (12) is greater than of maize (10). The size of genome has a great bearing on the nature of cloning vector to be used. Thus YAC (Yeast Artificial Chromosomes) and BAC (Bacterial Artificial Chromosomes) are ideal vectors because they stably handle inserts upto 200-2000 Kb and <300 Kb, respectively (Table 1).
- Relatively higher total recombination distance in genome (1700 cM) as against 1500 cM (maize) and Barley (1400 cM).
- Relatively lower proportion of kilobases per recombination unit (cM)-250 as against 1700 (maize) and 4000 (barley). This ratio is an indicator of how precisely a target gene can be located in terms of physical distance through linkage mapping techniques. The smaller ratio means more precision. This is all the more important in map-based cloning i.e., cloning genes based on their position on genome.

Table 1: Genome size in terms of chromosome number (1n), base pairs (Mb) and total recombinational map length (cM) of some plant sp.

Crop	1N	Mbp	cM	Kilo base/cM	%age repetitive DNA
Arabidopsis thaliane	5	100	500	200	-
Maize	10	2500	1500	1700	60
Tomato	12	950	1300	750	-
Wheat	21	1000	-	-	80
Rice	12	450	1700	250	58-66
Barley	7	5300	1400	4000	70

Table 2: Number of chains needed for a single genome and for 95% coverage using average size of inserts 700 Kb (YAC) and 200 Kb (BAC) in rice

	YAC (700 Kb)		BAC (200 Kb)	
Genome size	Single genome	95% coverage	Single genome	95% coverage
450 Mb (Appropriate)	643	1924	2250	6739

- Relatively lower amount of repetitive DNA. The compact nature of rice genome provides distinct advantage in gene isolation and genome sequencing as against other cereal crops.
- Diploid nature.
- Large number of EST (50000) database is available in rice facilitating a quick identification of genes of interest (Shimamoto and Kyozuka, 2002).
- Rice is amenable to manipulations in tissue culture techniques. Transformation protocols for development of transgenics have been well standardized.
- Several rice lines have been developed in which genes are tagged by insertion elements (Greco *et al*., 2001; Jeon and An. 2001) various such systems like Ac/Ds and Tos-17 have been used as tags in functional genomic studies in rice.
- Synteny relationships of rice genome with other cereals (Sorrells *et al*., 2003) provides for application of information to other crops as well.
- Large number of genetic stocks including mapping populations, introgression/substitution lines, genetic mutants, near isogenic lines have been accumulated in rice (Xu and Zhang, 2004).

Rice Genome Size

The estimates of genome size (total DNA content) and physical length of rice chromosomes are important considerations for sequencing rice genome (Table 2). The 2c (twice gametic) value for *O. sativa* ssp. *japonica* cv. Nipponbare ranges from 0.86-0.91 pg (Arumganathan and Earle, 1991). Assuming a mass of 650 D per base pair, the haploid genome of Nipponbare comes out to be 417 MB.

Chen *et al*. (2002) estimated thae size of gaps between base pairs using ratio of physical distance to local distance. When BAC contigs and gaps were totaled, the estimate for genome size came out to be 403 Mb. They also stated that physical map does not cover the nucleolar organisers on chromosome 9 nor does it include telomeres. Also gap estimates in centromere must be approximated because virtually no recombination takes place in centromere.

The current estimate of chromosome lengths are calculated assuming a genome size of 430 Mb and dividing it by each chromosomes fraction of total genetic distance as measured in Nipponbare × Kasalanth mapping populations. Presently IRGSP has almost completed chromosomes, 1, 4 and 10. The size of chromosome 1 is not precisely known because of many gaps including the centromeric region. The sequencing groups working on these chromosomes have put their lengths at 47, 36 and 24.5 Mb, respectively. However, from the integrated physical maps, the estimates are 44.3, 36.6 and 25.7 Mb, respectively.

Rice Genomic Research

In 1988, a research group of Cornell university for the first time used RFLP technique to construct a rice linkages map (McCouch *et al*., 1988) and subsequently developed a full linkage map with 726 DNA markers (Causse, 1994). This heralded the beginning of systematic research in rice genomic analysis. In Japan, another research group also developed a rice RFLP linkage map. To improve the understanding of rice genome, Japanese government started a seven year project on Rice Genome Research Programme (RGP) between 1991-1997, as a collaborative initiative of National Institute of Agro-biological Resources (NIAR) and Society for Techno-Innovation of Agriculture, Forestry and Fisheries (STAFF) under Japanese Ministry of Agriculture, Forestry and Fisheries. This project had four main objectives.

- c-DNA analysis
- Genetic mapping
- Physical mapping
- Bioinformatics

As a result, a large number of c-DNA's were isolated, analysed and sequenced and used as markers for constructing a linkages map and a physical map of Yeast Artificial Chromosomes (YAC). The genetic map based on morphological markers and linkage map based on RFLP markers were integrated and led to isolation of a number of important genes like *Xa21* which confers resistance against bacterial leaf blight\of rice using positional cloning (map based cloning).

The International Rice Genome Sequencing Project (IRGSP) was started in 1997 at an International Symposium on plant molecular biology in Singapore (Sasaki and Burr, 2000) scientists from many nations agreed to an international collaboration initiative for sequencing of rice genome. The representative from Japan , Korea, China, UK and USA met six months later in Tsukuba and established following guidelines.

- Use of single cultivar (Nipponbare), also called as GA3
- Sharing of material and database
- Clone-by-clone approach as used in human genome project
- Immediate sequence release

By now, IRGSP has evolved to include 11 countries with a working group having representative from different participating nations. The representatives from different countries in working group are (Eckardt, 2000).

- USA ---- Rod Wing (Clemenson University)
- Japan ---- Takiji Sasaki (Programme Director of RGP in Japan)
- Brazil ---- A.C. Oliveira (UFP)
- Canada ---- Thomes Bureau (Mc Gill University)
- India --- Akhilesh Tyagi (DU) and Naghendra Singh (IARI)

The details of participating nations and institutions with respective chromosomes assignments are given in following table

Institution		Chromosome assignments
Rice Genome Research Programme (Japan)	:	1,6,7,8
Korea Rice Genome Research Programme (Korea)	:	1
CCW (USA) Clemenson University Genome Institute (USA)		
Cold Spring Harbour Laboratory (USA).		
Washington University Genome Sequencing Center (USA)	:	3,10
TIGR (The Institute of Genome Research) USA.		
PGIR (Plant Genomic Initiative at Rutgers University) USA	:	10
University of Wisconsin USA	:	11
National Center for Gene Research, Chinese Academy of Sciences (China)	:	4
Indian Rice Genome Programme (University of Delhi)	:	11
Academia Sinca Plant Genome Center (Taiwan)	:	5
Genoscope (France)	:	12
Unwerridad Federal de Pelotas (Brazil)	:	12
Kesetsart University (Thialand)	:	9
McGill University (Canada)	:	9
John Innes Center (United Kingdom)	:	2

The rice genome sequencing project is being carried out as per the guidelines agreed upon by member countries. Large insert genomic libraries used as sequencing templates are constructed in BAC's or P-1 derived artificial chromosomes (PAC's). Though a YAC physical map covering 63% of rice genome was constructed by Saji *et al.* (2001) but the greater instability, high chimera frequency and problems in manipulation and purification make YAC's less than ideal choice for use as template for genome sequencing. The BACs and PACs have advantages of large insert and low copy number and are thus substrates of choice. The sequencing is done on BAC and PAC libraries derived from a single japonica variety Nipponbare which was agreed upon as a singe template throughout IRGSP. The primary reason for using Nipponbare cultivar is that more than 20000 EST's have been identified and physical map based on YAC that covers about 50% of geenome has been published. Using other cultivar was strongly discouraged because of non-distinction between genetic polymorphism and sequencing errors. However, China working on chromosome 4, uses an *indica* variety Guang Lu Ai 4 (Sasaki and Burr, 2000).

In order to accelerate and broaden the scope of rice genomics and identification and isolation of genes determining QTLs for growth and physiological traits in cultivars from *Indica*, Chinese scientists from 11 institutions led by Jun Yu of BG1 (Beijing Genomics Institute) in collaboration with University of Washington Genome Centre (UWGC) started whole genome shotgun approach in May 2000 and working draft was prepared on October 2002. They used the cultivar Liang-You-Pei-Jiu (LYP9) of super hybrid rice breed. The genome size was 466 Mb in size with an estimated 46022 to 55615 genes (Yu *et al*., 2002). Functional coverage in assembled sequence was 92%. Recently Tao *et al*. (2001) have also reported fingerprinting of 21087 BAC clones from *indica* rice cultivar Teqing. However, very little genetic information is available. Another sequencing endaevour was taken up by Syngeneta in Collaboration with Myriad Genetics Inc. led by Stephen Goff of Syngeneta using a *japonica* cultivar Nipponbare which is mainly cultivated in temeprate climates (Goff *et al*., 2002).

The Monsanto company independently sequenced rice genome (Barry, 2001). Monsanto has pledged to make these sequences available to the members of IRGSP and public researchers. Monsanto has sequenced 50895 BAC clones with Hind-III fingerprints and more than 13500 STC's. About 3400 BAC's were selected representing 393 Mb of rice genome. Monsanto has transferred these clones to RGP along with information on in silico physical map of 2000 clones and more than 125000 STC's (Eckardt, 2000).

The information generated by IRGSP about rice genome using BAC's and PAC's is shown in Table 3. As of December 2002, the IRGSP has already published about 367 Mb of non-overlapping nucleotide sequences of rice genome in public database.

Table 3: Features of rice genome as per different versions of draft sequence (Goff *et al*., 2002; Yu *et al*., 2002)

Feature	Nipponbare (*japonica*)	93-11 (*indica*)
Gene size (Kb)	2.4	4.5
Gene density (Kb/gene)	5.7	4.5
No. of exons/gene	4.2	3.3
Exon length (bp)	296	201
No. of introns/gene	3.2	2.3
Intron length (bp)	371	356
Exon average GC content (%)	54.9	51.4
Intron average GC content (%)	38.9	37.0
Overall GC content	44.0	43.3
Physical distance/genetic distance (Kb/cM)	253	---
Estimated genome size (Mb)	420	466
Estimated gene number	3200-5000	46000-55000

Objectives of IRGSP

- Isolation, characterization and sequencing genes of biological importance
- Characterization and mapping of agronomically important genes with application to plant breeding.
- Constructing an ordered set of genomic (DNA) clones used to construct physical and genetic map of chromosomes.
- Elucidating the mechanism for coordinated and programmed gene expression.
- Studying the synteny relationships between related species and genera.

Methods Used in Genome Sequencing-Structural Genomics

The procedures used by IRGSP for sequencing rice genome are elucidated as follows (Barry, 2001; Chen *et al*., 2002)

BAC Library Construction and Fingerprinting

BAC libraries are constructed from HMW (High molecular weight) DNA from rice embedded in agarose gel obtained from 4-5 week old greenhouse grown seedlings (*O. sativa* spp. *japonica* cv. Nipponbare) (Peterson *et al*., 2000). The BAC vectors pBelo BAC 11 and pBAC Indigo are used to construct Hind-III and Eco-R1 libraries, respectively by partial digestion of DNA with Hind-III and Eco-R1 followed by ligation. The ligation reaction is transformed into *E. coli*, plated on selective media and arrayed. The Hind-III library consists of 36864 clones with an average insert size of 129 Kb whereas Eco-R1 library consists 55296 clones with average insert size of 121 Kb.

The BAC clones are fingerprinted using methods of Marra *et al*. (1997) wherein the purified BAC clone is completely digested by Hind-III. The fragments are run on a high resolution agarose gel and fingerprint of each clone is formulated by IMAGE software (Sultson *et al*., 1989) based on immigration distance of restriction fragments. The Hind-III fingerprint data is subjected to overlap analysis using fingerprinted contig software (FPC) developed by Soderlund *et al*. (2000), in order to analyze the unrecognized overlaps between contigs. A contig (BAC contig) is a contigous set of mutually overlapping BAC clones that has been anchored to a position along the length of a particular chromosome (Zhang and Wing, 1997).

Anchoring BAC contigs to Rice Genetic Map

This step involves correlation of genetic and physical maps and is done by DNA gel RFLP, BAC end sequence analysis and chromosome walking. It is accomplished by four methods

- Generation of DNA probes for DNA gel blot hybridization or overgo hybridization
- In silico hybridization using DNA chips to anchor physical contigs genetically. This involves end sequencing of every BAC clone insert in Hind-III and *Eco-R1* library.
- Chromosome walking with overgo primers to identify potentially overlapping and extending clones
- Integration of Monsanto draft rice genome data for tracing out associations between Monsanto Rice BAC clones and physical map through in silico searches. Straight selection is done to remove doubtful matches due to repetitive sequences.

Sequencing, Finishing and Annonation

In IRGSP, shotgun approach is used to sequence BAC/PAC clones wherein individual clones from a Sequence Ready Contig (SRC) are shattered by Sonication or Nebulisation and fragments sub-cloned to produce a shotgun library. The clones from shotgun library are sequenced at random to provide a desired coverage of total sequence.

After sequencing softwares such as PHRED and PHRAP are used to assemble the shotgun DNA sequence data. However, there are always gaps due to AT and GC rich repetitive DNA. The process of removing these gaps is called as Finishing. This is done by another software called as MUMmer.

Sequence annonations are done by a software package, Gene Terpret developed by Rabbit hutch Biotechnology Corp. Using this software researchers annonating four clones on top of short arm of chromosome 10 located 89 protein coding genes.

Assigning Function to Genes-Functional Genomics

One of the major implications of plant genomic research is that about 54% of genes in higher plants can be assigned some degree of function by comparing them with sequences of genes with known function. Though function of genes can be deciphered in forms of conversion of gene language to protein sequences but knowing the general function does not always provide an insight into specific role of that gene in an organism. Even though genes are associated with transcription and signal transduction, the mere knowledge about transcription factors encoded by genes does not provide details about the process controlled by those genes. Thus rice genome sequencing will invariably be followed by large scale functional genomics in which all of about 5000-60000 genes (Goff *et al*., 2002) making up rice genome will be assigned function on the basis of experimental evidences (Somerville and Somerville, 1999). Unlike structural genomics, there are no international collaborative efforts on rice functional genomics but Japan, China and Korea have started efforts in this direction. China started its China Rice Functional Genomics Programme (CRFGP) in 1999 (Xue *et al*., 2003). The basic mission of CRFGP is to identify the genes of agronomic importance. Till now they have focussed on development of tools for functional analysis of rice genome and identifying rice genes essential from breeding perspectives.

The process of assigning function to sequenced genes has got great impetus by Reverse Genetics. The technique of insertional mutagenesis will also speed up the process. This involves screening a large collection of insertional mutants for insertional inactivation of any gene using PCR primed with oligomcleotides based on sequences of target gene and insertional mutagen.

Functional analysis of rice genome by mutational studies is due to presence of repetitive DNA sequences. Almost 50% of rice genome is repetitive DNA. Because of tight linkage between repetitive DNA, sequences, it is very difficult to generate double mutants by genetic recombination. A possible solutions is use of homologous recombination to eliminate tandem genes simultaneously by gene replacement. Alternatively point mutations by RNA-DNA hybrids may be useful. These methods are however, only useful in case of clone by-clone approach and not suited to High thorough put approach because of low efficiency and necessity of regenerating plants from single cultivated plants.

Gene silencing approach based on double stranded RNA from bidirectional transcription of genes in transgenic plants may be useful for gene inactivation for high thorough put approach

Gene Chips and Mkroarrays

Gene chips and microarrays are one of the promising experimental approaches in functional genomics (Somerville and Somerville, 1999). Gene chips and microarray techniques involve placing of DNA sequences representing all genes of an organisms on a miniature solid support consisting of plastic slides, silicon wafers or nylon membrane and using them as substrates for hybridization to quantify the effect of gene in terms of its expression. A typical micro-array analysis involves exposure of an immobile phase that could be either PCR amplified genome sequences, C-DNA's or oligonucleotides concentrated within a solid background, to a mobile phase of flourescently labeled DNA probe. The resultant binding of c-DNA sequences is detected as a signal which is an index of gene expression. A single micro-array unit provides for expression profiling of a large number of genes from

different tissues or a single gene in different tissues (Joseph *et al*., 2002). The micro-array analysis serves two basic purposes; gene discovery and identification of gene clusters with similar expression profiles. Thus using global database of micro-array expression profiles, it is possible to group genes into function related clusters (Schenk and Libert, 2001). These could be later used as markers for cultivar recognition and identification of pathogen and pathogen isolates. This system can generate a database of quantitative information on degree of expression of genes in response to pathogens, pests, drought, cold, salinity, photoperiod and other environmental variations. It can provide information as to which genes respond to changes during developmental process such as germination and flowering, phytohomones, growth regulators, herbicides, pesticides etc. In effect such databases will provide insights into the pathways of genes which control complex responses, a phenomenon termed as Genome ecology, wherein genome is viewed as a whole and relationships of gene products with each other is considered in terms of relative levels of expression. Micro-arrays are of significant importance in this regard because they can be made in any organism in which complementary DNA's can be isolated. In microarray analysis, gene sequences need not to be known before hand but can be determined after, arrays have been analyzed for identification of a gene of interest.

The data base for DNA microarrays and gene chips from different experimental sources will create a powerful tool for assigning function to genes of otherwise unknown function. The principle of microarrays and gene chips essentially involves assigning to different genes based on similarities in patterns of expression in response to diverse stimuli, developmental or experimental conditions. Thus different unknown genes can be clustered in similar group and assigned hypothetical function, Though in case of plant genome analysis, work on microarrays is in its initial phase but its potential is surely undoubtful. The challenge however is databasing and integrating massive data inputs in this approach.

In contrast to microarrays which are produced by directly spotting DNA on matrix, gene chips are produced by synthesizing oligonucleotides on a solid support by photolithography. Thus arrays can be produced that contain several thousand oligonucleotides. With advancement and refinement of technique, gene chips database will be developed to represent the entire genome in all its complexities. Gene chips have been used in functional analysis of all genes in yeast genome. By hybridising yeast genome DNA to such chips, 3714 SNP's between two genotypes could be identified in single hybridization. However presently gene chips are not routinely used in plant functional genomics because of their high cost.

Plant Artificial Chromosomes (PLAC)

The gene-by-gene approach for transfer among plant species is an ideal choice when number of genes governing a trait is limited. In case of traits conditional by a large number of genes, simultaneous transfer of genes is desired. In case of seed storage proteins where large number of genes are involved, it may be necessary to manipulate dozens of seed protein genes in order to tailor the amino acid content of seeds. In principle it may be possible to identify genes for useful biological pathways or traits by fragmenting the donor genome into large pieces of about 50 genes segments and then introducing them in a donor plant like rice and testing for components of phenotype. This is possible when the introduced gene confers a dominant or semi-dominant phenotype such as presence or absence of enzymatic activity, disease reaction or modification of developmental processes. By introducing such 50 gene fragments, it would require only 500 transgenic plants to be assayed to explore the entire genome of a typical diploid angiosperm.

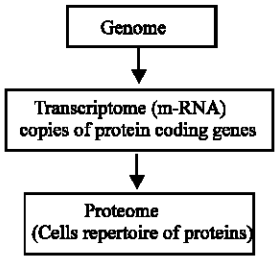
This analysis can be accomplished by making Plant Artificial Chromosomes (PLAC) libraries of plants like rice. In principle it is possible to make Plant Artificial Chromosomes (PLAC) libraries in yeast and then introducing them in a suitable host to analyse components of phenotype. By providing a defined chromosomal environment for cloned genes, the use of PLAC's, may enhance our ability to produce transgenic rice lines with defined levels of expression.

Proteomics in Rice Functional Genomics

Proteomics or proteome analysis refers to monitoring changes in protein expression in response to changing environments. The technique essentially involves extraction of proteins from rice tissues and separating them by two-dimensional PAGE. An image analyser is used to reveal the protein spots on PAGE gels stained by Coomassie Blue (Komatsu *et al*., 2003). The separated proteins are electroblotted into a polyvinylidene difluoride membrane and N-terminal amino acid sequences are determined using a protein sequence or mass spectrometry after enzyme digestion of proteins. Finally a data file on rice proteins can be constructed showing amnio acid sequence and sequence homologues. The two-dimensional gel electrophoresis can resolve several thousand protein species in a single slab gel but is very cucbersome in operation, has a poor dynamic range and is biased towards abundant and soluble proteins (Zhu *et al*., 2003).

Proteomics became a reality with development of protein electroblotting techniques which allow transfer of sample from the gel matrix into a support which is suitable for gas phase sequencing. Proteins can also be recognized by their amino acid composition, extra molecular weight determined by mass spectrometry or their partial amino acid sequence. Gel separated proteins can be rapidly identified by mass Spectrometry and if genomic information is available, such analysis can allow systematic identification of protein component of genome. It also allows for analysis of isoforms, secondary modifications of proteins like glycosylation, phosphorylation or proteolysis. Recently non-destructive methods such as Matrix-assisted Laser Desorption Ionization (MALDI) and Electrospray Ionisation (ESI) have made it possible to analyse biomolecules such as proteins. MALDi-Mass spectrometry method is a rapid method of identifying proteins when a fully decoded genome is available as the results can be conveniently compared with the predicted gene products of an organism. The ESI-MS system is advantageous because separation techniques such as liquid chromatography can be easily coupled with MS, which has proven to be an efficient alternative to 2D-technioques for proteins profiling. In fact LC-MS has been widely used for characterization of proteins. Due to the complexity of proteome, only a fraction of it can be characterized at a time. However, recently certain techniques have been developed for simplification of protein complex before separation and characterization, based on cysteine residues but such techniques suffer from one major limitation that a protein has to contain at least one cysteine residue in order to be detected (Zhu *et al*., 2003).

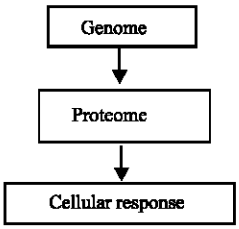
The information generated by proteome analysis helps in knowing protein sequence and sequence, homology and thus predicting function of unknown protein under different conditions. This in turn will help in their molecular cloning.



Komatsu *et al*. (2003) studied proteins from embryos and endosperm of *O. sativa* cv. Nipponbare and identified a number of proteins including a calcium biding protein that was similar to calreticulin of maize and a gibberlim-binding protein which was similar to ribulose-1,5-biphosphate carboxylase/oxygenase activase of barley. Both of them have role in signal transudation pathways. The former has role in plant cell regeneration while as latter mediates cellular response to gibberlic acid.

They also constructed c-DNA's for both of these proteins and isolated positive clones by southern blotting and later on sequenced one full length c-DNA insert for calreticulin (CRO-1) an two full length c-DNA inserts for Rubisco-Activase (Oscra-A1 and Oscra-A2).

Proteomic analyses have potential to study the proteins present or induced in rice plant under a variety of conditions. The knowledge about tissue specificity, developmental stage and physiological conditions will help identify the role of these proteins. Besides helping in sequence homology, it will also help in molecular clonning of such proteins. The importance of proteome analysis is all the more significant from the fact that it is the central link between genome and cell. On one hand it is culmination of genome expression and on other hand it heralds beginning of biochemical activities which constitute cellular responses (Brown, 1999).



Understanding Gene Systeny-Comparative Genomics

The similarities in the organization and function of higher plant genomes reflect common descent from an ancestor. Comparative studies throw light on mode, mechanism and speed associated with plant evolution and also provide conduits by which genomic information from one plant taxon can be used to make predictions about others (Paterson, 1999). This forms the basic premise of comparative genomics.

Comparative genomics offers unprecedented opportunities to link disparate biological kingdoms (McCouch, 1998). Cross-species, cross-genera and cross-kingdom comparisons will not only hold key to understanding of gene structure and function but also how changes in DNA have resulted in biological diversity. In this connection grasses (Poaceae) offer a well characterized system for development of tools to facilitate comparative genetic interpretations among members of diverse and evolutionary successful family. The discovery that gene content and gene order (genome colinearity) have been maintained within poaceae family for some 60 million years of evolution has further enhanced the eligibility of grass family for such comparative studies.

Rice is an excellent and ideal choice for comparative studies because of several desirable characteristic of its genome already described. Since cereals are of immense economic importance to mankind and human food chain, developments in this area of genomics will surely reflect in improved productivity of our food systems. Rice genomic tools and knowledge of rice genomic analysis are already being applied on other major cereals like wheat, maize, sorghum, barley etc. More importantly many aspects of rice genomics can be applied to certain Orphan Crops i.e., minor grasses which themselves do not fit for such studies.

The basic idea behind comparative genomics in rice is to advantage of faster speed of gene discovery and location in a model crop like rice and transferring the information to other crops which have large complex genomes and are polyploids. The cereals including wheat, rye, oats, barley have large complex genomes (10 pg/2C nucleus) in comparison to rice (3.3 pg). The large complex genome which is a result of high proportion of repetitive sequences is a major obstacle in speedy progress in systematic sequencing (Lewng and An, 2004). The establishment of conservation of gene order and content in grasses (grass alleles) is a major encouragement for using rice as a reference plant to decrease

the expenditure and labor involved in individual crop sequencing (Freeling, 2001). In fact Ahn and Tanksley (1992) first demonstrated that rice and maize share conserved linkage. The subsequent studies in comparative mapping of grass species revealed significant colinearity at macro level even though variations in colinearity do occur at micro level due to inversions mutations, translocations, deletions and duplications accumulated over millions of years in course of evolution.

Comparative genomics encompasses cross genome comparisons of genome organization and function to estimate similarity (Synteny) of biological organisation (Sorrels *et al*., 2003). Organismal evolution often provides threads of continuity that allow comparative analysis of biological systems to link genes, proteins and traits across species and genera. These patterns of relatedness can refine our existing knowledge about similarity among related groups. This can lead to new elucidations, hypothesis and predictions. Comparative genomics aims at creating such knowledge data bases with following objectives.

- To compare the organisation of related genomes and infer the basic processes of genome evolution.
- To transfer the information from model plants to related organisms
- To integrate the information on gene location and gene expression across species.

Since 1990, extensive species specific maps were developed for several members of poaceae family including rice (*Oryza*), wheat (*Triticum*), Maize (*Zea*), barley (*Hordium*), rye (*Secale*), Oats (*Avena*), Sorghum (*Sorghum*), pearlmillet (*Pennesitum*) and Sugarcane (*Saccharium*). Comparative mapping among these genera by using a common set of c-DNA clones suggest that some degree some genes are arranged in identical order along large tracts of chromosomes. This strengthens the concept that these crops contain essentially the same set of genes with observed rearrangement due to translocation and other mutational process during course of evolution, but undoubtly posses the same ancestral set of genes.

The present information in comparative genomic studies was developed by using a set of anchor CDNA probes developed by Van Deynze *et al*. (1998) at Cornell University (Gene Bank Accession Number : AA231638, AA231938). Linkages among anchored loci helped identify the homologous (not homologous) regions of distantly related genomes and provides instantaneous coordination of mapping results. These anchored probes consist of 152 c-DNA probes selected from rice (67), wheat (1) barley (21) and oats (63), c-DNA libraries that met the following criteria.

- Hybridization with majority of grass species, surveyed by southern analysis.
- Appeared to be low or single copy in rice.
- Provide optimum genome coverage when mapped onto existing linkages maps of rice, maize, barley, wheat and oats.

Comparative Mapping of Gramineae (Poaceae)

Though Poaceae family diverged over 65 billion years ago, comparative studies have revealed a high level of gene order conservation at macro level (Hulbert *et al*., 1990; Kurata *et al*., 1994; Gale and Devos, 1998). In case of domesticated grasses the conserved linkage blocks and their relationship with rice have led to hypothesis about basic organization of ancestral grass genome (Wilson *et al*., 1999). The conservation of gene content and order at megabase level (Macro colinerity) is critical for utilization of model species for positional gene cloning (Tanksley *et al*., 1995) development of molecular markers and for identifying syntenic regions in model species, which might contain genes of interest. Rice is a valuable point of comparison for such comparative studies because of its genome characteristics (Goff *et al*., 2002). Wheat on the other hand has a genome size 40 times larger than rice,

is a polyploid (2n - 6x : AABBDD genome), has 25-30% gene duplication (Akhinov *et al*., 2003) and over 80% repetitive DNA and is thus not so amenable to genomic studies. Thus crop species with not so ideal genome characteristics can benefit from comparative genomic studies. Similarly the results can be applied in case of economically less important grasses which do not warrant such expensive research.

The great amount of macro colimearity among grasses is providing compelling evidence for studying grasses as a synthetic genome with conservation of gene order along genomic regions spanning over megabases. However, despite such macrocolinearity, the practical applications of such synteny are limited on a comparative basis possibly because genomic information is not comprehensively available in grasses other than rice. Besides the colinearity observed at recombinational map level does not always correspond to local genome structure level (Feuillet and Keller, 2002).

The colinearity at the level of small genomic regions smaller than few cM has been found to be incomplete among cereals (Dubcovsky *et al*., 2001). Such microcolinearity has been observed in closely related species for genes of a same gene family (paralogs) or genes from a common ancestral gene (orthologs) which are often analysed and cloned in multiple species (Xu and Zhang, 2004)

The information from such comparative genomics studies from cereals is important for map based cloning of genes in cereals with large and complex genomes for improvement of cereals. However, such an application requires detailed understanding of evolution of cereal genome.

Synteny Between Rice and Wheat

Sorrells *et al*. (2003) performed comparative DNA sequence analysis of rice and wheat genomes by comparing 4485 deletion-mapped wheat EST's against 2251 rice BAC/PAC's using NCBI-BLAST. The structural relationships between the genomes indicated that for most individual rice chromosomes there is preponderance of wheat genes from one or two wheat homoeologous group. Wheat EST's matching sequences on rice chromosome R.1 are largely from wheat chromosome W3, whereas R2 and R3 are generally related to W6 and W4. R4 and R7 are related to W2 and R5 and R10 to W1 and R6 and R8 to W7. However, the centromere locations in rice did not correspond well with wheat. The prominent features of rice-wheat genome comparisons were grouped into four categories.

A: Regions of conserved gene content with one wheat genome location.
B: Regions of conserved gene content with multiple wheat genome locations.
C: Poorly conserved regions with one wheat genome location.
D: Poorly conserved regions with multiple wheat genome locations.

The category A-regions are prominent in all rice chromosomes, B-regions are less prominent and localized. C-regions are more common in centromeric regions of R1, R2 , R3 and long arms of R2 and R8 and D-regions are widespread and apparent in short arms of R3, R6 and long arms of R3, R4 and R10.

The study however found numerous discontinuties in gene order of wheat and rice. They concluded that grass genomes are labile, rapidly evolving entities and that structural and functional relationships are too complex.

Synteny between Rice and Sorghum

The comparative mapping of rice probes on sorghum have revealed a high degree of colinearity and conservation of gene order (Paterson *et al*., 1995). The results of these comparative studies revealed that most of the markers on rice chromosome R1 are found on sorghum chromosome S7. Similarly the colinearity of other chromosomes is as under

R8=S1; R4=S2; R3=S3; R2=S4; R5 and R7=S6; R6=S8; R12=S9 and R11=S10.

These results suggested that there is extensive conservation of gene order in rice and sorghum genomes although phylogenetically they are distantly related. The practical implications of such conservation is in map based cloning of specific genes.

Application of Genomics in Rice Breeding

The availability of complete genome sequence of rice is a landmark achievement and should stimulate large scale gene discovery in rice and related crops, through comparative mapping. The discovery of genes governing economically important traits can be used in Marker Assisted Selection (MAS), improving rice through transgenic approach and discovery of new beneficial alleles in germplasm (allelic mining). Knowledge of function of all plant genes coupled with development of tools for modifying and interrogating genomes will lead to robust genetic engineering experimentation in which rational changes can be designed for plant improvement.

Marker Assisted Selection

Perhaps the most significant advance in practical utilization of rice genomic research is Marker Assisted Selection (MAS). MAS is useful especially in case of selection independent of environments, selection without test crossing, selection without field work, selection at early stage, and selection for multiple genes/traits. The markers can be detected by relatively low cost gel systems. SSR's (Micro -satelites) have been of special attention due to their abundance and polymorphism. The sequence published by Goff *et al*. (2002) indicated that on an average one SSR comprising of at least 8 repeats of 2-4 bp was found after every 8 kb in a total of 48351 kb genome. For plant breeders, information on economically important genes especially those conditioned by polygenes with similar phenotype like disease resistance is of prime consideration and utility. In fact major advances in MAS have been made in resistance breeding in rice especially blast and blight. Molecular mapping now allows rapid analysis about a gene being completely new or located in vicinity of an already described gene. The blast resistance genes are now being systematically identified and characterized.

MAS in plant breeding is especially useful in gene pyramiding and selection for traits which are otherwise difficult to screen (Marker Assisted Back-crossing). Alien introgression can be made more precise in terms of minimizing linkage drag by selecting against donor markers. Using MAS, several resistance genes for blight and blast have been pyramided in agronomically superior cultivars.

MAS using gene sequence identified through functional genomics can be a powerful tool for rice improvement through conventional hybridization and selection strategies (Mackill, 2003).

Transgenics

Advances in rice transformation can be integrated with genomics to achieve results similar to marker assisted backcrossing. In rice this approach is currently being employed for bacterial blight resistance gene *Xa21*, the first resistance gene to be cloned in rice (Zhang *et al*., 1998). Though transformation approach can be fast and more accurate. There are obvious constraints in form of undesirable genetic alterations, gene silencing and biosafety compliance. However, transgenics are very useful in case of multiple gene traits which are difficult to be transferred through conventional breeding. When a number of genes are cloned, they can be put together in a single gene construct and introduced together. The advantage with this approach is that such genes will cosegregate in subsequent manipulations by conventional breeding. They can also be introduced into other varieties by crossing and will behave as a single factor.

For efficient transgenic approach, a robust transformation technology and a controllable expression system are pre-requisite. This has been amply demonstrated in case of waxy gene in rice which reduces amylose content in grains. This trait has been transferred in elite restorer lines in China. More importantly, efficient transformation protocols have been developed in recalcitrant cultivated *indica* lines and the expression level has been found to be stable over generations (Liu, 2000).

Gene Expression and Transcription Factors

The expression of genes can be regulated by changing the promoter or by use of anti-sense constructs. A gene *pdc-1* codes for pyruvate decarboxylase which is associated with submergence tolerance. Quimio *et al*. (2000) achieved over-expression of *pdc-1* which resulted in higher submergence tolerance. Similarly, the expression of *waxy* gene was down regulated by using anti-sense construct to reduce amylose content (Liu, 2000).

Gene expression in plants is regulated by a number of factors which are genetically governed. Transcription factors are important regulators of plant development and response to different conditions by influencing expression of genes. These transcription factors mainly regulate expression of downstream genes which may be involved in response to stress conditions (Chen *et al*., 2002).

At present, using genomic analysis, expression profiling of about 400 transcription factors using c-DNA chips has been done which have revealed several factors associated with gene activation, developmental processes and cellular response to different conditions. In principle a c-DNA fused with a binding domian could encode a transcription factor if the fusion is capable of inactivating a reporter gene in yeast. Various transcription factors including GAI4, MADS, AP2, bZIP,MYB Zu jinger, ERF, NAM, AUX/IAA, WRKY, TFB3, GRAS, HSF, Tubby, BRCT, Fungal TF, SBP, TCP< jimonjiand HB have been identified and characterized functionally (Kikuchi *et al*., 2003).

Osnato *et al*. (2001) have cloned and characterized a transcriptional factor MYB4 which is involved in cold tolerance in rice. The trascriptional activity of MYB4 was also tested on several cold tolerance genes such as *paf93* and *pt 59* of barley and *des90* of tomato. MYB4 encodes a transcription factor which is 257 amino acids long and upon over-expression confers better cold tolerance by preventing photooxidative damage at 2°C and other cellular membrane damages in chilling such as ion leakage.

Rice gene Analogues

The comparative genomics has established the genomic relationship between different members of Poaceae. The genomes of different cereals have been found to have a conserved gene content and order with minor translocational and mutational variations (Goff *et al*., 2002). Genomics thus helps in identification of different versions of rice genes from other species and genera. Such analysis will help in identification of orthologous genes in other cereals with similar sequence and function to those in rice but with markedly different phenotype. Such genes can be introduced into rice to produce novel phenotypes. This is especially important in case of resistance breeding where transfer of genes from related species is quite easy.

A very important consideration in this regard is that since gene order is highly conserved among cereals and there is great synteny, the genes transferred from one cereal to another will do well because of similar genetic background.

Germplasm Mining

Rice genome is believed to contain some 5000-6000 genes. The structural and functional analysis of such a large number of genes is a long term endavour. But recent advances in high throughput gene expression studies involving microarray and gene chips will go a long way in identifying genes and assigning function to them. These studies will help in functional analysis of rice genome in spatial and temporal aspects, i.e., the conditions under which various genes are expressed or are knocked out of function or their function is altered. This will assist in establishing biochemical pathways vis-à-vis gene expression. The practical aspects for a plant breeder would be to look for allelic variants (Gene tilling) in rice germplasm which confer superior phenotype. Such an approach will be quite useful in backcrossing for alein introgression from exotic cultivars.

Genomics and Hybrid Rice

Hybrid rice breeding has been a driving force for rice improvement with substantial increase in grain yield especially in China. One of the major limitations is narrow genetic base of restorer lines which decrease possibility of selecting optimal parental lines for hybrid combinations. It is very tedious to transfer restorer genes in elite genetic background by conventional breeding. This can be overcome by transgenic approach using several restorer genes like Rf-1, Rf-2 and Rf-3 for WA (wild abortive) system. Recently some genes conditioning TGMS have been identified which will help in popularizing two-line approach of hybrid rice production.

Poor grain quality sometimes is a major concern in hybrid rice breeding. Genomics helps in identifying and cloning such genes which affect grain quality in rice. Chinese Rice Functional Genomis Programme has objectively targeted such genes and has identified and cloned several genes by map-based cloning. Continued success of our endeavour to exploit heetrosis will depend on our ability to apply genome manipulations. The primary areas of application of genomics for hybrid rice will be.

- Characterization of genetic variability in rice gene pool
- Understanding of genetic basis of heterosis by molecular dissection of phenomenon
- Identification of heterotic chromosomal blocks (HCB's) which confer hybrid advantage
- Prediction of heterosis using molecular markers.

Genomics and Rice Ideotypes

Ideotype breeding in rice aims at arriving at a predetermined phenottypic expression for designing high yielding cultivars with good grain quality. Such an approach can be greatly facilitated by genomics by discovery of genes conditioning plant growth and developmental traits such as tillering, fertility and plant architecture.

Less unproductive tillers or Monoculm plant has been an important consideration in rice ideotype. A rice gene called Monoculm-1 (MOC-1) has been recently cloned. The MOC-1 mutants are usually monoculm without tillers due to defect in formation of tiller buds. MOC-1 encodes a putative nuclear protein mainly expressed in auxiliary buds and promotes their outgrowth.

Recently another gene governing feeding quality of rice culm BC-1 (Brittle culm-1) has been cloned by map-based cloning. The BC-1 gene reduces cellulose content of culm which is a desirable nutritive change because ruminants do not possess any cellulase enzymes in their rumens.

Conclusions

Rice is unique in the sense that it is not only the model cereal for genomic research but also an important staple food. The global resources can be coherently integrated to achieve goals. The information generated by genomics will not only help in improvement of rice but also other cereals because of great synteny in gene order. Genomics offers tremendous opportunities to scientists for innovation and also policy makers to put in place a coherent international effort in this regard. The population of rice eaters is anticipated to double upto 2025, thus we need to produce more rice within the limitations of declining resource base. The understanding rice genome, besides being a major landmark in plant biotechnology will bring us one step further towards our goal of producing more rice with less land, less water and less chemicals. Thus the future of genomics in rice improvement is undoubtedly promising.

References

Ahn, S. and S.D. Tanksley, 1992. Comparative linkage maps of rice and maize. Proceedings of National Academy of Sciences, USA., 90: 7980-7984.

Akhinov, E.D., J.A. Goodyear, S. Geng, L. Qi, B. Exhalier, B.S. Gill, G. Lazo, S. Chao, O.D. Anderson and A.M. Linkiewicz, 2003. The organisation and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res., 5: 753-763.

Arumganathan, K. and E.D. Earle, 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep., 3: 208-218.

Barry, G., 2001. The use of Monsanto draft rice genome sequence in research. Plant Physiol., 125: 1164-1165.

Brown, T.A., 2002. Genomes. John Wiley and Sons Inc. USA., pp: 572.

Causse, M.A., 1994. Saturated molecular map of rice genome based on an Inter-specific backcross population. Genetics, 138: 1251-1274.

Chen, M., G. Presting, W.B. Barbazuk, J.L. Goicoecha, B. Blackman and G. Feng *et al*., 2002. An integrated physical and genetic map of rice genome. Plant Cell, 14: 537-545.

Dubcousky, J., W. Ramakrishna and P. San Miguel, 2001. Comparative sequence analysis of collinear barley and rice BAC's. Plant Physiol., 123: 1342-1353.

Eckardt, N.A., 2000. Sequencing the rice genome. Plant Cell, 12: 2011-2017.0

Feuillet, C. and B. Keller, 2002. High gene density is conserved at syntenic loci of small and large grass genomes. Proceedings of National Academy of Science. USA., 96: 8265-8270.

Freeling, M., 2001. Grasses as a single genetic system: Reassessment. Plant Physiol., 15: 1191-1197.

Gale, M.D. and K.M. Devos, 1998. Comparative genetics in grasses. Proceedings of National Academy of Science, USA., 95: 1971-1974.

Ghosh, S.K., K. Virupakshappa and M.S. Singh, 2000. Genomics and food security. The Business Line. Nov. 29, 2000.

Goff, S.A., D. Ricke, T.H. Lan, G. Presting, R. Wang, M. Dunn and J. Glazebrook *et al*., 2002. A draft sequence of rice genome (*Oryza sativa* L. sp. *japonica)*. Science, 296: 92-100.

Greco, R., P.B.F. Ouwerkerk, C. Sallaud, A. Kohli and L. Colombo, 2001. Transposon insertional mutagenesis in rice. Plant Physiol., 125: 1175-1177.

Hulbert, S.H., T.E. Richter, J.D. Axtell and J.L. Bennetzen, 1990. Genetic mapping and characterization of sorghum and related crops by maize DNA probes. Proceedings of National Academy of Science, USA., 87: 4251-4255.

Jeon, J. and G. An, 2001. Gene tagging in rice. A high throughput system for functional genomics. Plant Sci., 161: 211-219.

Joseph, B., A. Shrinivasan and G. Kumaramanickavel, 2002. Microarays-chipping in genomics. Ind. J. Biotechnol., 1: 245-254.

Kikuchi, S., K. Satoh, T. Nagata, N. Kawagashira, K. Doi and N. Kishimoto *et al*., 2003. Collection, mapping and annonation of over 28000 c-DNA clones from *japonica* rice. Science, 301: 376-379.

Komatsu, S., H. Konishu, S. Shen and G. Yang, 2003. Rice Proteomics: A step towards functional analysis of the genome. Mol. Cell. Proteomics, 2: 2-10.

Kurata, N., G. Moore, Y. Nagamura, T. Foote, M. Yano, Y. Minobe and M. Gale, 1994. Conseration of genomic structures between rice and wheat. Biotechnology, 12: 276-278.

Leung, H. and G. An, 2004. Rice functional genomics: Large scale gene discovery and application in crop improvement. Adv. Agron., 82: 55-111.

Liu, Q., 2000. Stable inheritance of the antisnese waxy gene in transgenic rice with reduced amylose level and improved quality. Trans. Res., 12: 71-82.

Mackill, D.J., 2003. Application of genomics to rice breeding. IRRI Newslett., 281.

Marra, M., K. Dewar, P. Dunn, J.R Ecker, S. Fisher, S. Kloska, H. Leharch, R. Martienssen, S. Meier-Ewert and T. Altman, 1997. High throughput fingerprint analysis of large insert clones: contig construction and selection of clones for DNA sequencing. Genome Res., 7: 1072-1084.

McCouch, S.R., G. Kochert, Z.B. Yu, Z.Y. Wang, G.S. Khush, W.R. Coffman and S.D. Tanksley, 1988. Molecular Mapping of rice chromosomes. Theor. Applied Gene., 76: 815-829.

Mc Couch, S.R., 1998. Towards a plant genomic initiative: Thought on the value of cross species and cross genera comparison in the grasses. Proceedings of National Academy of Science USA., 95: 1983-1985.

Osnato, M., M. Bracale, F. Locatelli, C. Vannini, E. Magnani and Coraggiov, 2001. MYB4: A rice transcriptional factor involved incold tolerance. Oral Abstract. Proceedings of XLV Congress of Italian Society of Agricultural Genetics. 26-29 September, 2001.

Paterson, A.H., 1999. Comparative structural and functional analysis of plant genomes. Abstracts of Rice Genome Forum VII, Tsukuba, Japan. Feb. 10, 1999.

Peterson, D.G., J.P. Tomkins, D.A. Frisich, R.A. Wing and A.H. Paterson, 2000. Construction of bacterial artificial chromosome (BAC) libraries : An illustrated manual. J. Agric. Genomics 5. (www. Nigr. Org/research/jag).

Saji, S., Y. Umehara, B. Antonio, H. Yamane, H. Tanone, T. Baba, H. Aoki, N. Ishige, J.Z. Wu, K. Koike, T. Matsumoto and T. Sasaki, 2001. A physical map with yeast artificial chromosome (YAC) covering 63% of 12 rice chromosomes. Genome, 44: 32-37.

Sasaki, T. and B. Burr, 2000. International Rice Genome Sequencing Project: The effect to completely sequence rice genome. Curr. Opin. Plant Biol., 3: 138-141.

Schenk, P.M. and P.R. Libert, 2001. DNA microarrays: New tools in agricultural biotechnology. AgBiotech Net Reviews. Abstract No. ABN-078, pp: 46.

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

Soderlund, C., S. Humphray, A. Dunham and L. French, 2000. Contigs built with fingerprints: Markers and FPC V4.7. Genome Res., 10: 1772-1787.

Somerville, C. and S. Somerville, 1999. Plant functional genomics. Science, 285: 381-382.

Sorrells, M.E., M. La Rota, Bermidez- C.E. Kandianis, R.E. Greene and R. Kantety *et al*., 2003. Comparative DNA sequence analsyis of wheat and rice genomics. Genome Res., 13: 1818-1827.

Sultson, J., F. Mallet, R. Durbin and T. Horsnell, 1989. Image analysis of restriction enzyme fingerprint autoradiograms. Comp. Appl. Bio. Sci., 13: 101-106.

Tanksley, S.D., M.W. Gomal and G.B. Martin, 1995. Chromosome landing: A parasigm for map based cloning in plants with large genomes. Trends in Genetics, 11: 63-68.

Tao, Q., Y.L. Chang, J.Z. Wang, H.M. Chen, Islam M.N. Faridi, C. Scheuring, B. Wang, D.M. Stelly and H.B. Zhang, 2001. Bacterial Artificial chromosome (BAC) based physical map of rice genome constructed by restriction fingerprint analysis. Genetics, 158: 1711-1724.

Van Deynze, A.E., M.E. Sorrells, W.D. Park, N.M. Ayers, H. Fu, S.W. Cartinhove, E. Paul and S.R. McCouch, 1998. Anchor probes for comparative mapping of grass genera. Theor. Applied Gene., 97: 356-369.

Wilson, W.A., S.E. Harrington, L. Woodman, M. Lee, M.E. Sorrells and S.R. McCouch, 1999. Can we infer genome structure of progenitor maize through comparative analysis of rice, maize and domesticated panicoides. Genetics, 153: 453-473.

Xue, Y. and Z. Xu, 2000. An introduction to China rice functional genomics programme. Fundamental Genomics, 9: 161-169.

Xue, Y., J. Li and Z. Xu, 2003. Recent highlights of the China Rice Functional Genomics Programme. Trends in Genetics, 18: 390-394.

Xu, Y. and Q. Zhang, 2004. Rice genome: Implications for breeding rice and other cereals. Paper presented at 4th International Crop Science Congress, 2004.

Yu, J., S. Hu, J. Wang, G. Wong, S. Li and B. Liu *et al*., 2002. A draft sequence of rice genome. (*Oryza sativa* L. ssp. indica). Science, 296: 79-92.

Zhang, H.B. and R.A. Wing, 1997. Physical mapping of rice genome with Bacterial Artificial Chromosome. Plant Mol. Biol., 35: 115-127.

Zhang, S., W. Song, L. Chen, D. Ruan, N. Taylor, P. Ronald, R. Beachy and C. Fauquet, 1998. Transgenic *Elite indica* rice varieties resistant to *Xanthomonas oryzae*. Mol. Breed., 4: 551-558.

Zhu, H., M. Bilgin and M. Snyder, 2003. Proteomics. Ann. Rev. Biochem., 72: 783-812.