# Progress in Rice Genome Project and Cross-Species Implication (Synteny)

Takuji Sasaki[1]

## Abstract

The Rice Genome Research Program (RGP) was established in 1991 to analyze the genetic phenomena of rice as an assemblage of information originating from the rice genome and to use the results as tools for characterization of genes controlling an agronomically important trait and improvement of rice and other crops. To attain this goal, RGP covers five activities, namely, cDNA analysis, genetic mapping, physical mapping, DNA marker application and database management. Genetic map with 2,292 DNA markers is utilized for accurate tagging of phenotypical traits and gene isolation. Our linkage map was constructed based mainly on RFLP markers which are advantageous for direct use as probes for detecting polymorphism among other crop species. The cross-hybridization of a marker from rice or other crops to both species can precisely identify the corresponding loci in both genomes. This method could clarify the synteny among grass genomes. The genome size of rice is the smallest among important crops and the findings of conservation of gene ordering around a specific trait, such as disease resistance or stress tolerance, in crops other than rice could facilitate the molecular characterization of the trait using the materials and information obtained in rice.

## Introduction

Rice is one of the major staples in the world, especially in the Asian and African countries. Rice is cultivated under a wide range of environments, from arid highlands to flooded lowlands. The wide adaptability of cultivated rice, *Oryza sativa*, to such conditions is very important for maintaining a constant yield of rice. However, the population living there is still increasing and the earth surface is limited. To produce a sufficient amount of staple, more efficient cultivation of rice is required. For the development of new varieties that can grow under more severe conditions, exploitation of genetic resources and incorporation of favorable characters among them into cultivated rice are most promising. Many studies have been carried out to evaluate agronomically important traits among the species of the genus *Oryza* and to identify the location of their genes by linkage analysis after introgression by crossing. This procedure is the first major one to utilize the alien genes for cultivated rice in order to expand the cultivated area.

However, for an accurate and a comparable identification of genes, universally standardized markers are indispensable. DNA which is the element of the genome is

[1]National Institute of Agrobiological Resources, Kannondai 2-1-2, Tsukuba,Ibaraki, Japan

the only material to be chosen for this purpose. Genetic analysis of the rice genome by DNA markers produced several linkage maps (McCouch *et al.*, 1988; Kurata *et al.*, 1994a; Causse *et al.*, 1994). The next step is to utilize DNA markers on these maps based on genetic materials for segregation of target traits. Once the trait is accurately tagged by DNA marker(s), its presence or absence could be determined and time saved by setting PCR primers based on a sequence information of the DNA marker(s). Ultimately, we can expect to isolate a gene corresponding to a trait, though this process requires more effort and time than tagging. So far, only one gene corresponding to a phenotype has been isolated from rice, that is, *Xa21*, one of the bacterial blight disease resistance genes (Song *et al.*, 1995). The isolated gene or the genetic information obtained through the isolation procedure can be used for producing transgenic plants.

The results obtained in the rice genome analysis can be transferred to other crops, such as barley, wheat and maize, because the ordering of genes within these genomes is conserved (Moore *et al.*, 1995). This homologous relationship is now referred to as synteny. In addition, the genome size of rice is the smallest among cultivated grasses. Although macroscopic level synteny among grasses was observed and confirmed, micro scopic level analysis is needed if the tagging of genes from plants other than rice is planned based on the synteny between rice and other grasses.

In this report, the progress of the rice genome research program is described focusing on linkage analysis and some aspects about synteny based on a detailed genetic map of rice.

## Genetic analysis by DNA markers

The genetic map of RGP was constructed using 186 F2 plants from Nipponbare (japonica variety) and Kasalath (indica variety). Up to now, 2,292 polymorphic DNAs have been mapped along the total 1,530cM genetic distance (Fig.1). This map shows the highest density and quality among linkage maps ever published not only for rice but also for other species. The markers used consisted mainly of rice cDNAs, or the expressed parts of the genome, and the others of rice genomic DNAs, RAPDs, wheat DNAs, barley DNAs and maize DNAs. These markers are utilized for a precise tagging of phenotypical traits or for screening of YAC (yeast artificial chromosome) library to pick and align positive YAC clones along the genetic map. The information in this map is also useful for considering genetical events, such as recombination frequency, explicitly by molecular tools.

For the identification of genes corresponding to traits, near-isogenic lines or inbred lines are needed to correlate the location of traits and DNA markers. When the major genes are the target for tagging, the evaluation of phenotype is clear and the utilization of DNA markers is not essential. However, in the case of the traits relative to the contribution of multiple genes, DNA markers can show their superiority to the previous markers, such as isozymes, for detecting traits based on the contribution of each gene. Such traits, occasionally called quantitative traits, are mostly important for agronomy and must be analyzed exactly for genetical and molecular aspects. In

RGP, interval mapping of QTL is performed for 11 traits observed in parents for mapping population, such as heading date, culm length, panicle length and seed width, with 857 loci on our genetic map (Yano *et al.*, 1995). For example, 5 QTLs were identified for heading time and the accuracy of some of the loci was revealed by segregation analysis using near-isogenic lines as for the loci obtained by backcrossing.

The success of gene identification by map-based cloning largely depends on the accurate linkage analysis of trait by DNA markers. At present, no standardized population size has been determined for identifying a sufficient number of recombinations between target trait and DNA markers to minimize the number of candidate cosmid clones. Usually, more than 1,000 plants showing target phenotype from self-pollinated near-isogenic lines is required. For such a minute analysis, additional DNA markers should be produced by screening a cDNA library by candidate YACs (Yoshimura *et al.*, 1996) or by subcloning DNA fragments produced by digestion of candidate YACs with restriction enzymes (Monna *et al.*, 1997). In RGP, the identification of the *Xa-1* gene, *Pi-b* gene and one major gene for heading time is challenged. The final identification of candidate gene as a true one requires the change of characters by transformation.

## Alignment of YAC clones along the genetic map

In RGP, a YAC genomic library from the rice variety Nipponbare was constructed in which about 7,000 clones with an average insert size of 350kb are involved (Umehara *et al.*, 1995). The conversion of the genetic map to a physical map is crucial for translating genetical events into molecular language. To achieve this objective, DNA markers on the genetic map are used as probes for screening the YAC clones dotted on a filter with high density. For the assignment of the accurate location of positive YACs on the map, Southern hybridization is required. Up to now, 1,245 probes were used and a total of 5,183 positive YACs were recognized. After assigning the location of each YAC, overlapping of YACs by neighboring DNA markers on the map was identified. The total number of contiguous YACs with multi-markers was 169 and that of YACs without overlapping 350, respectively (Table 1). By using a larger number of markers, more overlapping of YACs is expected to be confirmed.

## Capture of expressed rice genes

Random cloning and partial sequencing of rice cDNAs are performed to verify the signature of expressed genes of rice. Up to now, about 30,000 cDNA clones were analyzed (Table 2). Using FASTA algorithm, about 25% of their translated sequences were found to be similar to those registered in one of the public databases, PIR. Redundancy among analyzed cDNAs is checked by terminal overlapping and is estimated at about 50%. We have estimated the gene number of rice of 20,000-40,000 and 40-75% of them seem to be captured. These cDNAs have been used as DNA markers on the genetic map and the function of the gene product could be assigned by a probability of 20-30%.

## Synteny among grass genomes

DNA markers on our genetic map are extensively used for mapping of wheat (Kurata *et al.*, 1994b, Fig.2), barley (collaboration with Dr.A.Kleinhofs and Dr.A.Grane r), maize (collaboration with Dr.M.McMullen) and foxtail millet (collaboration with Dr. M.Gale). DNA markers of wheat, barley and maize were used as probes for rice genetic mapping. This cross-hybridization enables to clarify the corresponding loci among these genomes. The conservation of ordering of the corresponding loci was clearly observed as blocks and this homologous relation is now referred to as synteny. The synteny shown in Fig.2 suggested that it may be possible to characterize genes common to rice and wheat. To achieve this objective, a more detailed analysis of synteny around a target trait is required. For example, to isolate a stem rust disease resistance gene (*Rpg1*) in barley, rice YACs corresponding to the genomic region of *Rpg1* were extensively analyzed for their ordering (Kilian *et al.*, 1995).

Through synteny analysis, discrete differences in recombination frequency among grasses were recognized. As shown in Fig.3, comparison of the genetic distance of syntenic chromosomes between rice and barley showed that about half of the distance of rice chromosome 1 around the centromere corresponds to 10% of the distance of barley chromosome 3 around also the centromere. Also the whole rice chromosome 10 corresponds to only about 20% of the genetic distance of wheat chromosome 1. Although, at present, the difference in physical distance has not been determined for these regions, such information about the difference in recombination frequency is important for understanding the genetics and for application of DNA markers to interspecies usage.

## References

1) Causse,M.A., Fulton,T.M., Cho,Y.G., Ahn,S.N., Chunwongse,J., Wu,K., Xiao,J., Yu,Z. Ronald,P.C., Harrington,S.E., Second,G., McCouch,S.R. and Tanksley,S.D. (1994):Saturated molecular map of the rice genome based on an interspecific back cross population. Genetics **138**: 1251-1274.

2) Kilian,A., Kudrna,D.A., Kleinhofs,A., Yano,M., Kurata,N., Steffenson,B. and Sasaki,T. (1995): Rice-barley synteny and its application to saturation mapping of the barley Rpg1 region. Nucleic Acids Res. **23**: 2729-2733.

3) Kurata,N., Nagamura,Y., Yamamoto,K., Harushima,Y., Sue,N.,Wu,J., Antonio, B.A., Shomura,A., Shimizu,T., Lin,S-Y., Inoue,T.,Fukuda,A., Shimano,T., Kuboki, Y., Toyama,T., Miyamoto,Y., Kirihara,T., Hayasaka,K., Miyao,A., Monna,L., Zhong,H.S., Tamura,Y., Wang,Z-X., Momma,T.,Umehara,Y., Yano,M., Sasaki,T. and Minobe,Y. (1994a): A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genetics,8 : 365-372.

4) Kurata,N., Moore,G., Nagamura,Y., Foote,T., Yano,M., Minobe,Y. and Gale,M. (1994b): Conservation of genome structure between rice and wheat. Bio/Technology **12**: 276-278.

5) McCouch,S.R., Kochert,G., Yu,Z.H., Wang,Z.Y., Khush,G.S., Coffman, W.R. and

Tanksley,S.D. (1988): Molecular mapping of rice chromosomes. Theor.Appl.Genet. **76**, 815-829.

6) Monna,L., Miyao,A., Zhing,H.S., Yano,M., Iwamoto,M., Umehara,Y., Kurata,N., Hayasaka,H. and Sasaki,T. (1997): Saturation mapping with subcloning of YACs: DNA markers production targeting the rice blast disease resistance gene, Pi-b. Theor.Appl.Genet. (**94**, 170-176).
7) Moore,G., Devos,K.M., Wang,Z. and Gale,M.D. (1995): Grasses, line up and form of circle. Current Biology **5** : 737-739.
8) Singh,K., Ishii,T., Parco,A., Huang,N., Brar,D.S. and Khush,G.S. (1996): Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). Proc.Natl.Acad.Sci.USA **93** : 6163-6168.
9) Song,W.-Y., Wang,G.-L., Chen,L.-L., Kim,H.-S., Pi,L.-Y., Holsten,T., Gardner,J., Wang,B., Zhai,W.-X., Zhu,L.-H., Fauquet,C. and Ronald,P. (1995): A receptor kinase- like protein encoded by the rice disease resistance gene, Xa21. Science **270** :1804- 1806.
10) Umehara,Y., Inagaki,A., Tanoue,H., Yasukochi,Y., Nagamura,Y.,Saji,S., Otsuki,Y., Fujimura,T., Kurata,N. and Minobe,Y. (1995): Construction and characterization of rice YAC library for physical mapping. Molecular Breeding **1** : 79-89.
11) Yano,M., Nagamura,Y., Kurata,N. and Sasaki,T. (1995): Genetic linkage map of DNA markers and its application to genetic analysis in rice. Gamma Field Symposia **34**, (PP.17-30. Institute of Radiation Breeding, NIAR, MAFF, Japan).
12) Yoshimura,S., Yamanouchi,U., Katayose,Y., Kurata,N., Yamamoto,K., Sasaki,T. (1996): Isolation and sequence analysis of candidate cDNA of Xa-1, the bacterial blight resistance gene in rice (in Japanese). Breeding Sci **46**(Suppl.1) 16.

**Table 1** Number of sequenced clones

| cDNA Library | Analyzed clones | Hit clones[1](%)[2] | Submitted clones[3] |
|---|---|---|---|
| Callus | | | |
| BA[4] treatment | 2157 | 563 (26.1) | 608 |
| Heat[5] shock treatmennt | 3015 | 642 (21.3) | 0 |
| Growth phase[6] | 2492 | 745 (29.9) | 2448 |
| GA[7] treatment | 2975 | 803 (27.0) | 0 |
| Root | 1965 | 517 (26.3) | 1849 |
| Green shoot | 4891 | 1264 (25.8) | 3431 |
| Etiolated shoot | 4522 | 980 (21.7) | 2654 |
| Flowering panicle | 1742 | 390 (22.4) | 0 |
| Ripening panicle | 1702 | 618 (36.3) | 0 |
| Young panicle | 940 | 168 (17.8) | 0 |
| Young panicle 2 | 935 | 175 (18.7) | 0 |
| Others | 1700 | 439 (25.8) | 0 |
| Total | 29018 | 7304 (25.2) | 10990 |

1) The FASTA algorithm was used for PIR similarity search and an optimized score of at least 200 was required for putative assignment.
2) (Hit clones / Analyzed clones) x 100
3) Submitted to DDBJ.
4) B-A: 6-benzyladenine
5) This callus was exposed to 37℃ for 3.5 hr after incubation at 25℃ for 12 days.
6) Callus grown in medium with 2,4-D (2,4-dichlorophenoxy acetic acid)
7) GA: gibberellin $GA_3$

**Table 2** Summary of screening of YAC clones

| Chromosome | Number of DNA markers used | Number of positive YACs | Number of contiguous YACs | |
|---|---|---|---|---|
| | | | Multi-markers | Single-marker |
| Chr. 1 | 174 | 635 | 20 | 31 |
| Chr. 2 | 133 | 578 | 21 | 26 |
| Chr. 3 | 147 | 580 | 20 | 34 |
| Chr. 4 | 114 | 634 | 10 | 37 |
| Chr. 5 | 105 | 570 | 30 | 30 |
| Chr. 6 | 146 | 619 | 21 | 22 |
| Chr. 7 | 108 | 526 | 13 | 24 |
| Chr. 8 | 78 | 227 | 6 | 47 |
| Chr. 9 | 60 | 254 | 7 | 19 |
| Chr. 10 | 57 | 157 | 6 | 30 |
| Chr. 11 | 85 | 305 | 11 | 22 |
| Chr. 12 | 54 | 146 | 4 | 28 |
| Total | 1,245 | 5,183 | 169 | 350 |

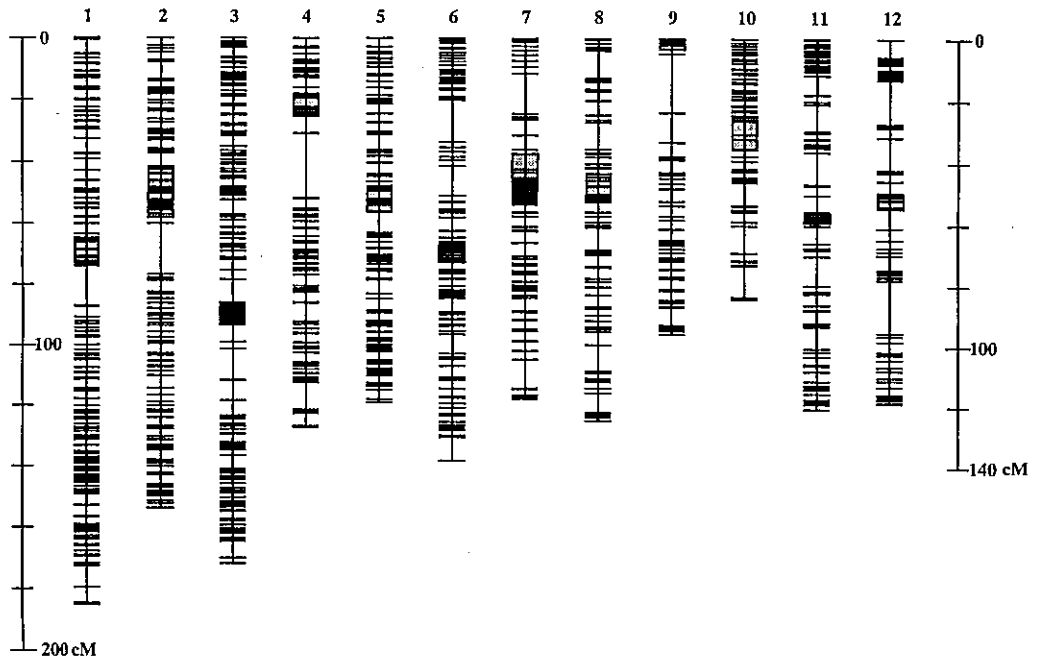


**Fig.1** Rice genetic map based on 186 F2 plants derived by crossing of Nipponbare (japonica) and Kasalath (indica)

A total of 2,292 markers were mapped along the genetic distance of 1,530cM. The shadowed area on each chromosome indicates the area sandwiched by the DNA markers flanking the centromeres (Singh *et al.*, 1996).

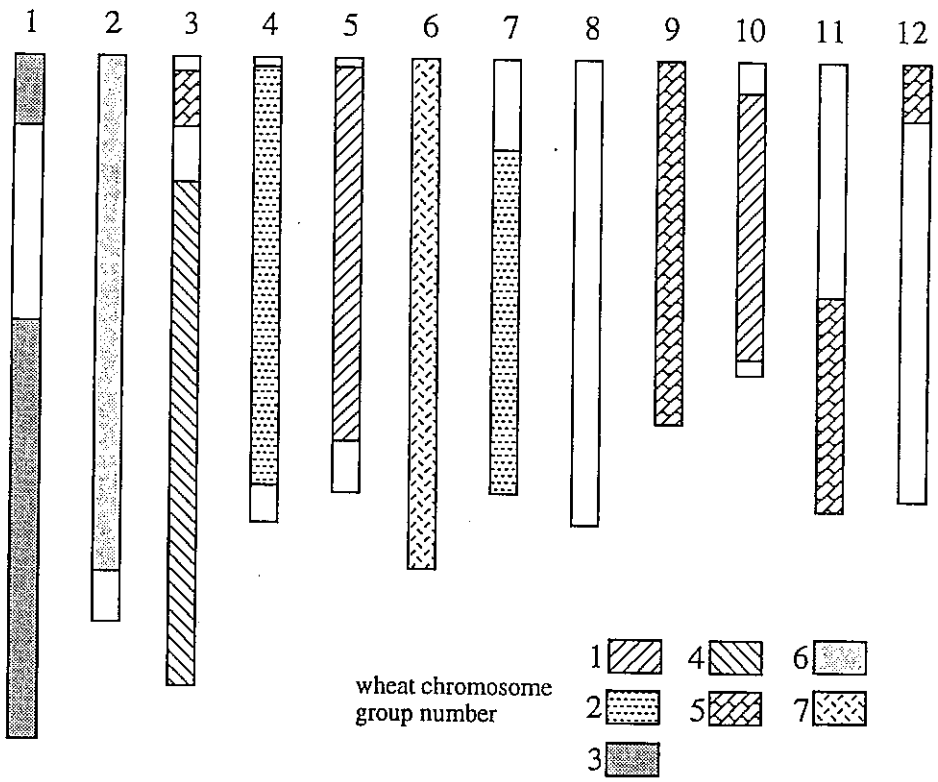


**Fig.2** Graphical drawing of rice-wheat synteny based on rice genetic map

The seven wheat linkage groups are indicated based on different patterns. The white regions did not show a clear conservation of marker ordering between the two species.

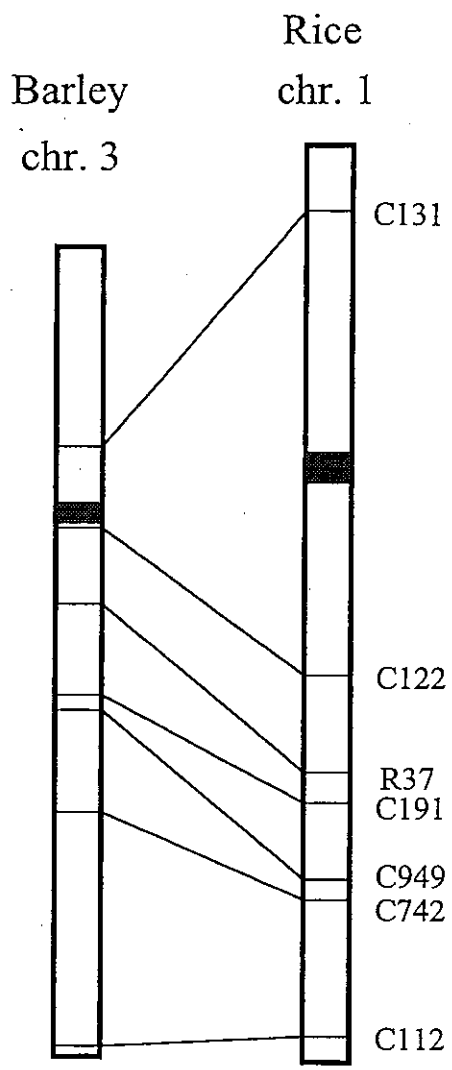


**Fig.3** Comparison of the genetic distances between syntenic chromosomes between rice chromosome 1 and barley chromosome 3

Rice RFLP markers are indicated on the right side of rice chromosome and the lines connecting two chromosomes indicate the corresponding loci. The original figure was given by Dr.A.Graner, Institute for Resistance Genetics, Germany.