

This paper was presented at a colloquium entitled “Protecting Our Food Supply: The Value of Plant Genome Initiatives,” organized by Michael Freeling and Ronald L. Phillips, held June 2–5, 1997, sponsored by the National Academy of Sciences at the Arnold and Mabel Beckman Center in Irvine, CA.

The rice genome project in Japan

TAKUJI SASAKI*

Rice Genome Research Program, National Institute of Agrobiological Resources, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305, Japan

ABSTRACT Since 1991, the Rice Genome Research Program in Japan has carried out rice genomics, such as large-scale cDNA analysis, construction of a fine-scale restriction fragment length polymorphism map, and physical mapping of the rice genome with yeast artificial chromosome clones. These studies have made a great impact on research into grass genomes and made rice a model plant for other cereal crop research. Starting in 1998, the Rice Genome Research Program will step into a new stage of genomics—that of genome sequencing. This project eventually should reveal all of the genomic sequence information in the rice plant and be an indispensable aid in understanding the genomics of other grass species.

Rice is a staple food for about half of the world's people and is mainly harvested and consumed in Asia and Africa, where the population is expected to double during the next 50 years. However, increasing rice production is becoming more difficult because of disease, drought, and salinization. We need to develop improved rice plants that are tolerant to such stresses. To achieve this tolerance, innovative tools are needed to identify the genetic information hidden in rice DNA, which controls all of the characteristics of rice plants.

So far, plant genomics is not sufficiently advanced to even understand the molecular genetic mechanisms responsible for the famous phenomena observed by Gregor Mendel. Everybody learns Mendel's law of inheritance in their high school days, but even about 150 years after his establishment of modern genetics, no detailed information is available about factors such as what controls the height of garden beans. But, after Watson and Crick (1), we have a great deal of information about DNA and are only now entering the age of genomics—the age of understanding the assembled genetic information derived from the genome.

With regard to the understanding of rice, the Japanese government started its 7-year Rice Genome Research Program (RGP) in 1991. Although this program ultimately aims to clarify the entire rice genome sequence, facilitate achievement of this goal, and increase the genetic tools available for a wide range of applications, the RGP is composed of four main groups: cDNA analysis, genetic mapping, physical mapping, and informatics (2).

The strategy of cDNA analysis in the RGP is random cloning and partial sequencing. This method is a quick and easy way to clone many genes expressed in rice. The sequences obtained are used (i) as tags for genomic regions of expressed genes (expressed sequence tags, ESTs); (ii) to clarify the gene expression profile in various tissues at different growing stages; and (iii) to explore the functions of gene products by searching public databases for similarities.

So far, we mainly have sequenced 400–500 5'-terminal bases of about 36,000 cDNA clones from 15 main cDNA libraries such as green and etiolated seedlings, young roots, panicles at the flowering stage, and calluses cultured with 2,4-dichlorophenoxyacetic acid (3). By using the FASTA algorithm, a similarity search against the Protein Identification Resources database found significantly matched sequences among about 25% of the rice sequences. After surveying by the Institute for Genomic Research assembler, redundancy among the 36,000 sequences was estimated to be about 59%. This estimate means that about one-third to one-half of the total expected rice genes already may be isolated.

Having such a large number of ESTs is a powerful tool for genetic and physical identification of expressed gene locations, for example as restriction fragment length polymorphism (RFLP) markers for linkage analysis or hinge markers for yeast artificial chromosome (YAC) contigs. Although RFLP analysis is, in some sense, a time-consuming and high-cost method of detecting polymorphism in the genome, the accuracy and reproducibility of linkage analyses based on RFLP is higher than those based on such methods as randomly amplified polymorphic DNA. Our latest RFLP linkage map, constructed by using 186 F2 plants derived from the japonica variety Nipponbare and the indica variety Kasalath is composed of about 2,300 DNA markers with a total genetic distance of 1,550 cM for 12 linkage groups (4). Nearly 70% of the markers are rice ESTs, and the remainder are clones derived from rice genomic DNA and genomic DNA or cDNAs from other cereals. On this map, the location of the centromere on each linkage group was assigned by use of markers that were judged, by using secondary and telotrisomics developed at the International Rice Research Institute, to flank the centromere (5). Coincidentally, this mapping revealed that the chromosomal orientation of each linkage group was with the short arm at the top and the long arm at the bottom. The position of the centromere clearly suggested that the genetic recombination at each centromere is very low and, also, that this region is flanked by regions of high recombination frequency. This finding of the meaning of marker-dense regions and gap regions by use of only linkage analysis is a remarkable result. This finding should improve our understanding of the molecular mechanisms of meiotic events, leading to effective introgression of desirable genes by recombination.

DNA markers on a fine genetic map may be used to accurately assign the genotypes of siblings that were obtained by backcrossing to establish a near isogenic line for a specific trait. The advantage of using DNA markers becomes apparent when quantitative trait loci are anatomized to clarify the contribution of each locus to the total quantity. For example,

Abbreviations: RGP, Rice Genome Research Program; EST, expressed sequence tag; YAC, yeast artificial chromosome; RFLP, restriction fragment length polymorphism.

*To whom reprint requests should be addressed. e-mail: tsasaki@abr.affrc.go.jp.

near isogenic lines each carrying five quantitative trait loci as a single Mendelian factor for heading time have been established by an accurate genotyping of candidate progenies (6). Because RFLP is not convenient for surveying many plants in a short time, PCR primers using the sequence information of corresponding RFLP markers should be established.

The most important uses of RFLP markers on a linkage map are as probes for screening a YAC library to convert genetic phenomena to physical architecture. Our YAC library from one of the parents for linkage analysis, Nipponbare, was composed of about 7,000 clones with an average insert size of 350 kb (7). The total insert size of this library is 5.5 equivalents of the rice genome (430 Mb). Colony hybridization of YACs was performed by RFLP markers on a filter dotted with 1,536 YAC clones over about 100 cm². The positions of the positive YACs on the linkage map were confirmed by further Southern hybridization using the same restriction enzyme to detect the polymorphism in Nipponbare as was used for the linkage analysis. Sequence-tagged site markers could identify positive YACs by a three-dimensional pooling method. By using both methods, 2,600 independent YACs have been identified so far, with about 1,300 markers aligned along the linkage map (8–15). On the basis of genetic distance, we estimated the coverage of each chromosome by YACs to be about 50%. Increasing the markers on the latest map for screening, and by using YAC end clones as markers to investigate overlapping of adjacent YACs, the coverage could be increased to about 70% in the case of chromosome 6. For the remaining chromosomes, the same strategy is expected to yield the same increase in coverage. In addition, mapping ESTs to YACs can effectively identify overlapping YACs without linkage analysis. For this purpose, nucleotide sequences of the 3'-untranslated region, which is specific for each gene, is to be used to design PCR primers. At present, this third method of YAC assignment is being widely promoted not only for physical mapping, but also to get information on where within the genome each specific EST is derived. The latter information will be important to identify the exact position of the expressed genomic region when we begin the rice genome sequencing.

For map-based cloning of genes corresponding to a trait, both DNA markers and YACs are required, as well as a segregating population for the target trait. We applied our map-based cloning tools to isolate a rice blight disease-resistance gene, *Xa1*. After obtaining a YAC clone by DNA markers tagging *Xa1*, this YAC was used to screen a cDNA library constructed by using leaves of a resistant variety that had been inoculated with the pathogen. Positive cDNAs were used for RFLP markers for additional fine mapping, and one of the cosegregating cDNAs showed sequence characteristics of other disease-resistance genes in plants. This candidate gene was confirmed to be the actual *Xa1* gene by transforming rice plants susceptible to the rice blight disease caused by race 1 of *Xanthomonas oryzae* pv. *oryzae* to resistance, with a cosmid clone carrying this gene. The *Xa1* gene is revealed as a member of a resistance gene family with nucleotide binding sites and leucine-rich regions (16).

Although map-based cloning is undoubtedly a very strategic way to isolate a gene, many ESTs may be used for quick cloning by searching for rice homologues among known genes in other species, such as *Arabidopsis*. Once an EST with a sequence homologous to a known gene is found it needs to be mapped and its loci compared with that known to convey a phenotypic trait. However, the conventional rice map does not correlate

well with DNA markers, and this cloning is only virtual. In addition, the criterion of homology is ambiguous and, in some instances an EST once thought to be a true homologue has turned out to be only a member of a family of the true gene. Nevertheless, virtual cloning is only available by using the many ESTs and fine RFLP maps produced by genome research. Before converting virtual cloning to real cloning, a genetically established segregating population carrying the target trait is needed. To facilitate the cloning of biologically and agronomically important genes, RFLP maps and conventional maps need to be integrated.

The next step of genome analysis is genome sequencing. Rice is thought to be a model plant for all of the cereal crops because of its small genome size and synteny with other grass species. As mentioned above, the basic tools and knowledge for genome sequencing of rice are now sufficiently established. The RGP will enter the next 7-year program in April 1998, and systematic genome sequencing then should start (17). Hopefully, the efforts of the RGP will stimulate rice genome sequencing in the worldwide consortium, and knowledge of the rice genome sequence will be used to improve not only rice, but all cereal crops. Our DNA materials, scientific information, and general information are available through the following worldwide web pages: <http://www.dna.affrc.go.jp:83/index.html>, <http://www.dna.affrc.go.jp:84/index.html>, and <http://www.staff.or.jp/>.

This work was supported by funds from the Japanese Ministry of Agriculture, Forestry, and Fisheries (MAFF) and the Japan Racing Association (JRA).

1. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171**, 737–738.
2. Sasaki, T., Yano, M., Kurata, N. & Yamamoto, K. (1996) *Genome Res.* **6**, 661–666.
3. Yamamoto, K. & Sasaki, T. (1997) *Plant Mol. Biol.* **35**, 135–144.
4. Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., et al. (1998) *Genetics* **148**, 1–16.
5. Singh, K., Ishii, T., Parco, A., Huang, N., Brar, D. S. & Khush, G. S. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6163–6168.
6. Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y. & Sasaki, T. (1997) *Theor. Appl. Genet.* **95**, 1025–1032.
7. Umehara, Y., Inagaki, A., Tanoue, H., Yasukochi, Y., Nagamura, Y., Saji, S., Otsuki, Y., Fujimura, T., Kurata, N. & Minobe, Y. (1995) *Mol. Breeding* **1**, 79–89.
8. Umehara, Y., Tanoue, H., Kurata, N., Ashikawa, I., Minobe, Y. & Sasaki, T. (1996) *Genome Res.* **6**, 935–942.
9. Wang, Z.-X., Itonuma, A., Umehara, Y., Van Houten, W., Ashikawa, I., Minobe, Y., Kurata, N. & Sasaki, T. (1996) *DNA Res.* **3**, 291–296.
10. Saji, S., Umehara, Y., Kurata, N., Ashikawa, I. & Sasaki, T. (1996) *DNA Res.* **3**, 297–302.
11. Antonio, B. A., Emoto, M., Wu, J., Umehara, Y., Kurata, N. & Sasaki, T. (1996) *DNA Res.* **3**, 393–400.
12. Shimokawa, T., Kurata, N., Wu, J., Umehara, Y., Ashikawa, I. & Sasaki, T. (1996) *DNA Res.* **3**, 401–406.
13. Koike, K., Yoshino, K., Sue, N., Umehara, Y., Ashikawa, I., Kurata, N. & Sasaki, T. (1997) *DNA Res.* **4**, 27–33.
14. Umehara, Y., Kurata, N., Ashikawa, I. & Sasaki, T. (1997) *DNA Res.* **4**, 127–131.
15. Tanoue, H., Shimokawa, T., Wu, J., Sue, N., Umehara, Y., Ashikawa, I., Kurata, N. & Sasaki, T. (1997) *DNA Res.* **4**, 133–140.
16. Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.-X., Kono, I., Kurata, N., Yano, M., Iwata, N. & Sasaki, T. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 1663–1668.
17. Sasaki, T. (1997) *Rice Genome* **6**, 1.