Rice genomics to understand rice plant as an assembly of genetic codes

Takuji Sasaki

Rice Genome Research Programme, National Institute of Agrobiological Sciences, 1-2, Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan

Rice genome research is a topic and an attractive theme in plant genomics. Because of its practical importance to feed about half of the world's population, and because of predicted unbalance of food supply and population increase in the near future, breeders, geneticists and molecular biologists in plant fields have strong concern in the results of rice genome research. In this review, recent research results of the rice genome are presented to offer information to develop ideas for understanding and improvement of rice plants.

RICE is one of the most important crops in the world that provides food for about half of the world population. In areas such as Asia, Africa and Latin America where the demand for rice is a top priority, the population is expected to increase 1.5-fold by 2025. However, the capability of land to sustain rice yield is almost saturated as a result of insufficient irrigation or limited adaptation of cultivated rice to prevailing environmental conditions. To overcome this situation, we must introduce new innovations or novel strategies for breeding new rice varieties. One approach is to select favourable progenies using reliable DNA markers or introduce favourable genes by transformation. These strategies can be realized only by understanding the concept of inheritance in rice plants based on indispensable information encoded in the DNA.

There exists a large number of local cultivars of *Oryza sativa* in which two major subspecies, japonica and indica varieties are known. Unlike in the case of humans, the long history of evolution of *Oryza* species generates much more divergence in nucleotide sequences. Hence, one specific rice variety must be chosen as a standard for a complete analysis of its entire genome. If not, the utilization of genetic information in understanding the rice plant would be limited because of the existence of polymorphism among alleles of many rice varieties. Therefore, the Japanese Rice Genome Research Programme (RGP) uses the japonica variety, Nipponbare as a resource for genetic and molecular analysis of the rice genome. Alternatively, for genetic dissection of the rice genome, a high frequency in polymorphism within the DNA sequences is favourable to generate many molecular genetic markers. For this purpose, the indica and japonica varieties are preferred as parents to produce progenies for linkage analysis. Once genetic markers that can unambiguously define the chromosomal regions are established, molecular dissection should be promptly performed.

The ultimate goal beyond dissecting the genome by genetic analysis is to decode all of the information for inheritance written in the DNA. The advancement of genome analysis supported by automated equipments was unexpected a decade ago. At present however, an automated capillary DNA sequencer can sequence daily a total of about 380 kb of template DNA with the aid of robotic machines for sample preparation. In addition, huge amounts of sequences are expediently accelerated through the development of innovative computer software and hardware systems. While whole genome sequencing programmes are costly endeavours, the information encoded in the sequences remains extremely valuable.

In this article, the processes involved in genome analysis and the results obtained from each step, and examples of their application are described.

Genetic analysis of the rice genome

Following the first successful application of the RFLP (Restriction Fragment Length Polymorphism) technique to tag the serious inheritable disease, Huntington disease[1,2] in the human genome, this procedure was subsequently used to analyse the genomes of other organisms. In rice, the first publication of a molecular genetic map using RFLP was in 1988 by McCouch *et al.*[3], and subsequently, more fine maps were released as described in Kurata *et al.*[4], Causse *et al.*[5], and Harushima *et al.*[6]. More recently, a sophisticated linkage map has been developed by Yano *et al.*[7] in which 3267 DNA markers were allocated and almost all of them were converted from RFLP to CAPS (Cleaved Amplified Polymorphic Sequence) markers. CAPS facilitates the identification of polymorphism within a PCR product using a restriction enzyme as an alternative to a laborious Southern hybridization tech-

e-mail: tsasaki@nias.affrc.go.jp

nique. The quality and number of markers on this map are undoubtedly the most outstanding among all organisms genetically analysed by DNA markers.

Other types of DNA polymorphism are also utilized to genetically anatomize rice chromosomes. The most commonly used are SSR (Simple Sequence Repeat), AFLP (Amplified Fragment Length Polymorphism), and RAPD (Random Amplified Polymorphic DNA). SSR is also a co-dominant marker like RFLP, but is much easier to detect by PCR and allows detection of more alleles than RFLP. The only significant limitation for SSR detection is cost performance. So far, SSR markers have been established by screening a genomic library with a synthetic DNA containing a target repeat, and then by designing PCR primers using flanking sequences to the repeat from the positive clones. As the genomic sequence of rice is gradually generated, more SSR markers should be easily found and widely used. Currently, there is one genetic map with 312 SSR markers available for rice[8].

The available genetic maps were constructed using progenies such as F_2 population or recombinant inbred (RI) lines derived from varieties that are genetically far apart, such as japonica and indica varieties as parents. This combination is expected to generate more polymorphisms than that between the same subspecies. The frequency of polymorphism observed within the cross of japonica varieties decreased to 10–30% compared to that obtained between crossing japonica and indica rice varieties[9]. It is important to consider these findings in the application of DNA markers to breeding as selection markers. In almost all breeding efforts, local varieties from similar ancestors are crossed to preserve the genetic background of established elite varieties. To overcome this situation, more sensitive techniques to detect polymorphism based on single nucleotide replacement must be developed using extensive genome sequence information.

The markers on molecular genetic maps are indispensable to assign any polymorphism in DNA sequences or phenotypes on each of the 12 rice chromosomes. Gene isolation based on forward genetics, or map-based cloning is primarily based on accurate tagging of phenotype by DNA markers and assignment of DNA fragment for genome sequencing on the corresponding original chromosomal position. Therefore, every available genetic marker is necessary for precise dissection of the rice genome. The total number of co-dominant markers such as RFLP, SSR or CAPS so far published is about 5000. This means that the average marker density within the rice genome is one marker in every 80 kb. Considering the low gene density in the centromere and telomere regions, this number is thought to satisfy the demand for a preliminary survey of linked markers to a targeted phenotype. For example, at RGP the first rough mapping of the targeted trait *Xa1*, the resistance gene to bacterial blight and *Pib*, the resistance gene to rice blast, was performed using a segregation population with about 100 plants. By this mapping, the target phenotypes were tagged within less than 10 cM with markers published in 1998 by Harushima *et al.*[6]. In this case, the average marker density is one marker in every 170 kb. For refinement of tagging, a scaled-up population size and generation of new markers using contiguous large-sized DNA fragments covering and/or flanking the target phenotype are required. The genome size of rice is relatively small, and still more than one gene is expected even within several ten thousand nucleotides. Therefore, fine mapping is preferred to transformation with ambiguous fragment that might bring more than two genes.

The most prominent application of DNA markers is for detailed mapping of quantitative trait loci (QTL). Since a QTL is controlled by an interactive association of multigene effects, the resulting phenotype is characterized by continuous variation. This makes it difficult to accurately evaluate QTL as a genetic phenomenon and complicates verification of its occurrence. However, the development of DNA markers enables a precise map localization of QTL by combining the mapping population derived from the substitution lines which carry chromosomal segment of one parental line with the background of another parent. For example, in the case of the cross between japonica variety, Nipponbare (N) and indica variety Kasalath (K), a total of 110 lines are established by backcrossing N to F_2 or F_3 population. These lines cover almost all of the 12 rice chromosomes each with one specific segment of K chromosome (M. Yano, pers. commun.). This population is extensively used to identify each locus of QTL regarding heading date of rice[10–12]. Using the F_2 population of the N/K cross, so far at least 8 loci were identified and each of them was isolated as a single Mendelian factor using these substitution lines. Hd1, the major QTL of heading date was recently isolated and variation of this allele among rice varieties was clarified[13]. This is a promising step that may provide a clue in the identification of genes involved in QTL and elucidation of biochemical and physiological events associated in multi-gene interaction.

For introgression of preferable disease resistance genes (*R*-genes) to an elite variety, gene pyramiding was performed[14]. This strategy is superior to introgression of one major *R*-gene, because its breakdown can be suppressed by interaction of several *R*-genes with similar but not the same specificity to the pathogen. For this strategy to succeed, its evaluation must be done only with DNA markers after the trial of introgression, mainly because resistance to a pathogen is not necessarily a clear indication of accurate introgression of the target gene, and this process requires a long time. Pyramiding of the four resistance genes to bacterial blight, *Xa4*, *xa5*, *xa13* and *Xa21*, was performed using this strategy[15]. These four genes were chosen for their mutual coverage of specificity to races of *Xanthomonas oryzae* pv. *oryzae* known in Eastern Asia. The resultant new pyramid lines gene-

rally showed a higher level of resistance and/or wider spectrum of resistance to the pathogen than was expected from the parental behaviour.

Application of rice markers to other cereals

It is believed that cereal plants have a common ancestral plant and diverged some 60 million years ago just after the extinction of dinosaurs[16]. Rice plant is thought to diverge from this ancestral plant by speciation resulting from chromosomal rearrangement[17]. Chromosomal shuffling which occurred during this process should be evidenced by cross hybridization of RFLP markers derived from cDNAs, which reflect nucleotide sequences within expressed genes that are relatively highly conserved during an evolutionary process. First, RFLP markers derived from rice and maize cDNAs were exchanged for cross-hybridization, and it was revealed that their mapped positions in the counterpart plant genome were regionally aligned in the same order as in the original plant[18]. This co-linearity is called synteny and has been clarified among cereal crops such as wheat[19], foxtail millet[20], and sorghum[21] based on rice RFLP markers. These results were summarized with the comparative mapping results between other cereals and the well-known circles of cereal genomes putting rice, with the smallest genome size among them, at their common centre[22]. These circles clearly show the intimate relationships among cereal genomes and provide a concrete evidence of the existence of an ancestral plant species among cereal crops. In addition, this relationship offers the opportunity for rice to be the model or reference plant among cereals to understand common genome structures and to identify any existing orthologous genes.

The first trial to obtain disease resistance gene (*R*-gene) using a syntenic relationship was performed on barley stem rust resistance gene, *Rpg1* (ref. 23). A detailed micro-level analysis of this region by additional DNA markers generated from barley cDNAs revealed that the co-linearity of these markers between rice and barley was not exactly conserved. Next, more extensive mapping of PCR products of a sequence characteristic for *R*-gene, nucleotide binding site–leucine-rich repeat (NBS–LRR) was performed in rice, barley, and foxtail millet[24]. This study clearly revealed the independent localization of *R*-gene homologues among these three cereal genomes. These two results suggest the inability of macro-level synteny among cereals to predict the location of disease resistance genes. Furthermore, this fact indicates that the integration of disease resistance genes in cereal genomes might have occurred very recently after divergence of each species. Plant–pathogen interaction is grouped into two categories, resistance and indifference. In general, each cereal species has its incompatible pathogenic micro-organism; for example, *Magnaporthe grisea* causes blast only on rice and *Puccinia graminis* f. sp. *tritici* causes rust in barley. Independent phenotypes may correspond to the observed independent localization of resistance genes among each species.

In addition to *R*-genes, syntenic structure among cereal species may help in the identification of orthologous genes involved in ancestral and common characteristics such as plant height or heading date. The leading study in the isolation of such genes was performed using a combination of genomic information from *Arabidopsis*, rice, and wheat[25]. The targeted phenotype was a gene conferring dwarfism in wheat, *Rht1*, which was used to breed a variety propagated by the so-called Green Revolution programme. The gene causing dwarfism insensitive to gibberellin is observed among many plant species. In *Arabidopsis*, the *GAI* gene shows reduced response to gibberellin, and its nucleotide sequence has a corresponding homologous gene among rice ESTs. This was used as a probe to screen wheat cDNA library and the resultant clone was proven to correlate with *Rht1* locus. In addition, the orthologous gene in maize, *d8*, was isolated using the same probe. The localization of phenotype dwarfism from *Rht1* and *d8* is similar to the syntenic genome structure common to both wheat and maize. In rice, no dwarf mutant is known to exist in the corresponding loci. Recently however, the *slender* mutant showing taller plant height than the wild type was mapped in the corresponding loci of *Rht1* and *d8*. The Rht1 and d8 protein is presumably a member of the nuclear transcription factor, and it may be plausible that the opposite effect can be induced by mutation.

The above facts demonstrate the importance of careful inspection of synteny at microlevel and genetic mapping of corresponding genes to confirm synteny. Also, we must recognize that similar yet unidentical phenotypes may be conferred by orthologous genes. So far however, not enough data are available to conclude whether synteny can validly identify orthologous genes demonstrating similar physiological functions among cereal crops. The important point is for rice researchers to definitely clarify the function of a target gene in rice and then to use this information in investigating the occurrence of presumably a corresponding gene in other cereal crops.

Molecular analysis of the rice genome structure

All of the information on rice heredity is encoded in an array of four nucleotides A, G, T and C. To obtain this information in the shortest possible time frame, we must exert much effort as in the human genome project which has been tackled by a public international collaboration. As rice is one of the most important cereal crops in the world, essential genomic information should be made available to the public. In this context, the International Rice Genome Sequencing Project (IRGSP) was organized

in February 1998 to complete the rice genome sequence following certain standardized guidelines[26]. The details of this collaboration can be found in the website: http://demeter.bio.bnl.gov/Guidelines.html. Some of the defining policies that have been established for the organization include the use of a common cultivar as template, accuracy of final sequence, data release policy and chromosomal sharing. There are 11 countries that currently participate in the IRGSP and hold biannual meetings to discuss scientific and political issues on rice genome sequencing.

The sequencing strategy at IRGSP is based upon BAC/PAC contigs constructed by combination of fingerprints, end-sequences of BACs/PACs (STCs) and PCR from primers generated by mapped ESTs. The anchor points of BACs/PACs on the contigs were assigned using genetic markers generated by RFLP analysis. IRGSP utilizes data from four public genomic libraries, including CUGI *Hin*dIII-BAC, CUGI *Eco*RI-BAC, RGP *Sau*3 AI-PAC, and RGP *Mbo*I-BAC. In addition, a private company, Monsanto, has provided selected clones from its *Hin*dIII-BAC library. All the libraries were generated from the DNA of the japonica rice cultivar, Nipponbare. Although a physical map covering all of the rice genome has not been completed, a nearly completed physical map for chromosome 1 is now available. Physical mapping of chromosomes 3, 6, and 10 is also underway and the remaining eight rice chromosomes will soon be initiated. Within the rice genome, there exist many similar sequences such as transposons or simple sequence repeats that interfere in the accurate assignment of PACs/BACs to its original genome position. In some cases, genetic mapping is required to resolve ambiguities in assigning the locus.

The shotgun method is used for sequencing selected BACs/PACs, with primarily dye-terminator sequencing reaction by dideoxy nucleotide method. At RGP, usually 1000 subclones each of 2 kb and 5 kb libraries are prepared as plasmids for each PAC/BAC clone and sequenced for both strands. The average size of a PAC/BAC clone and the detectable length of one sequencing trial are 150 kb and 500 bp respectively. This means that about 10 times coverage (10 ×) of the BAC/PAC template sequence is attained, which can almost complete the sequence, except when a complicated portion of the sequence is present. To align shotgun sequences consistently from one end of a BAC/PAC clone to another, a computer program such as the phred/phrap/consed[27] is used. This program also gives a score of confidentiality to each nucleotide by sequencing trace files. If gaps remain, a subclone bridging the gaps must first be located and then sequenced after other sequencing reaction methods are used. Sequencing using several methods is recommended because in general, gaps are rich in unusual nucleotide composition or could transform into a stable secondary structure. This gap-filling process is tedious and requires trained and talented personnel to accomplish the procedure.

The final step in genome sequencing is annotation of sequence such as gene prediction or promoter site assignment using a computer software to predict target areas based on statistical analysis. For gene prediction, several software based on Hidden Markov Model (HMM) have been developed using actual data on codon usage for each amino acid from cDNA sequence, particularly a full length cDNA. So far, the prediction program RiceHMM is the only one for rice genome (http://rgp.dna.affrc.go.jp/Analysis.html) that is based on the prediction program, GenScan[28]. Several prediction tools have also been developed such as the GRAIL[29], GeneMark[30], and FGENESH[31], which use similar statistical methods but have not yet been trained for rice statistical data. In addition, other types of software to locate splice sites of exon/intron or initiation and stop codon, and long terminal repeat of transposable elements are used. The annotation must also include the similarity search results against known genes or gene products deposited in a public database such as GenBank. At present, deposited gene sequence data have enormously increased in view of the progress made in major genome sequencing projects, including human, bacteria, nematode, fly and *Arabidopsis thaliana*. Similarity search is usually performed using BLAST (Basic Local Alignment Search Tool) for both nucleotide and amino acid sequences[32,33].

These results are then integrated to predict a gene within a genome sequence. In RGP, an automated system named Rice GAAS (Genome Automated Annotation System) has been developed for gene prediction. In Rice GAAS several programs are integrated, including three coding region prediction programs, one splice site prediction program, four similarity search programs, one tRNA-gene prediction program, two repetitive DNA analysis programs, one regulatory element scan program, one protein localization site prediction program, and one secondary structure prediction program for membrane protein. This program is publicly available via http://rgp.dna.affrc.go.jp/Analysis.html and can be used for gene prediction within an anonymous rice nucleotide sequence. A similar system is working at TIGR (The Institute for Genomic Research) with a combination of several prediction programs similar to those used in RGP. At TIGR, in addition to gene prediction, the annotated genes are linked further with their orthologs found in plant species to show alignment among orthologs using a program named TOGA (http://www.tigr.org/tdb/toga.shtml). This information must be useful for effective utilization of sequenced and annotated results for functional genomics. After the completion of the genome sequence of *A. thaliana*, its predicted gene often shows significant similarity with a rice predicted gene for its translated amino acid sequence. Although not all of the functions of the predicted gene in *Arabidopsis* have been elucidated,

such information is important to deduce and clarify the functions of the corresponding rice gene and to know the divergence and/or conservation of gene functions during a long history of evolution of monocot and dicot plants. So far, no clear synteny between rice and *Arabidopsis* has been demonstrated. A direct comparison between the two species will be pivotal in understanding plant evolution. The important point for rice breeding is how to find specific genes from the rice genome sequence. Through this effort, we must obtain much more fundamental genome information to understand the concepts of heredity and physiology in plants.

Gene disruption and loss of function analysis of rice gene

To elucidate gene function, several strategies are used to disrupt genes by physical, chemical or biological methods. Among them, the biological method is feasible to identify the genomic position of disrupted genes if the genome sequence is available and if the disruption is caused by a sequence-known factor. To satisfy these conditions, a transposable element such as Ac/Ds from maize has been used to disrupt rice genes[34]. However, because of instability after transposition, low efficiency in transposition or requirement of closed area for cultivation of recombinant plants, transposon from alien plants is not adequate for rice plant. Search trials for endogenous transposons in rice resulted in the discovery of *Tos17* that overcomes several problems described above. This endogenous retrotransposon can be activated by cell culture and after five months in culture, regenerated plants can produce about five new copies of *Tos17* in addition to the two original copies. These new copies are dispersed everywhere in the rice genome and based upon their nature as a retroelement, copies are stably inserted into a new genomic position[35].

Rice plants with genes disrupted by *Tos17* could be analysed by genetics and reverse-genetics approaches. If an abnormal appearance is recognized in a regenerated rice plant, it is expected that the gene related to the changed appearance is disrupted or tagged by *Tos17*. After identifying the *Tos17* that is linked to the phenotype, gene isolation using *Tos17* as a marker can be initiated. So far, genes related to viviparous, narrow leaf, semi-dwarf or brittle culm have been isolated (H. Hirochika, pers. commun.). On the one hand, reverse genetics method is advantageous for gene-disrupted rice lines by *Tos17*. DNAs from gene-disrupted lines are isolated and used as template for PCR primed by combination of *Tos17* and targeted gene-specific sequences. If *Tos17* is inserted into the targeted gene, PCR products must appear only in such lines. The relationship of the disrupted gene and the phenotype unravels the gene function. The first gene identified by this method was rice homeobox gene, *OSH15* that affects the architecture of internodes resulting in dwarfism[36]. To facilitate detection of disrupted genes, flanking sequences to the inserted *Tos17* are amplified by TAIL (Thermal Asymmetric Interlaced)[37] or Suppression PCR[38] and then the product is sequenced[39]. The sequence data are summarized as a database for searching against rice ESTs to identify the disrupted exon or against the rice genome sequence to identify the location of the disrupted sequence. If each data from the disrupted rice plant are linked to a phenotype, gene function could be easily clarified. The number of the analysed flanking sequence is now about 8000.

Profiling of gene expression

The methods described above can only identify a single target gene and clarify its expression profile such as by Northern blotting analysis. However, it is plausible that many genes must be expressed in rice plant in a concerted manner at any stage and situation in its life cycle. With the progress of chip technology, it is now possible to monitor a genome-wide change in gene expression[40]. To realize this objective, genomic DNA or cDNA clones with their analysed sequences are fixed on the surface of a glass plate to be monitored by fluorescent dye-labelled tester transcript extracted from rice plant after perturbation of culturing condition. If hybridization occurs, we can know the sequence characteristics of hybridized transcripts relevantly expressed with the perturbation. Because of many genes changing expression levels by perturbation, it is not easy to identify specific and key gene(s) responsible for perturbation. To narrow the candidate genes, other types of experiment confirming the results of monitoring must be required.

The accuracy of a gene expression profile system depends upon the following factors: (1) the degree of saturation with actual gene numbers of cDNA clones that are used for microarray; (2) dynamic range of fluorescence from hybridized probe and target; (3) reproducibility of the hybridization data. The first requirement is the most basic and significant, but the actual gene number in rice is still based on speculations. This means that microarray users must make an effort to collect as many cDNAs as possible, and recognize the limit of cDNA collections.

Direct transfer of genes to rice plant

One application of the molecular analysis of the rice genome is the direct transfer of an identified preferred gene to a rice variety lacking it. In 1994, the utility of *Agrobacterium* in rice had been shown to be also effective for rice plant[41] as in the case of dicot plants. In addition to this development, new promoters other than the 35S promoter of cauliflower mosaic virus have been identified. For example, the promoter of rice storage protein, glutelin, could be effectively used in seeds

possessing an alien gene expression system. So far, genes from species other than rice are mainly used in rice transformation. For example, the three genes involved in vitamin A biosynthesis from bacteria and daffodil were used to make the so-called 'Golden Rice'[42]. This type of development should be ascertained by persistently finding many useful genes in rice and by understanding genomic spots where transgenes are easily accommodated. These processes can be elucidated from the information obtained from the genome sequence. There are enormous arguments concerning the use of genetically modified organisms (GMOs). In this regard, the public should be educated on safety issues for human consumption, and a balanced viewpoint regarding the importance of technology to ensure a sustained source of staple in the future must be seriously considered.

Concluding remarks

Molecular genetic analysis undoubtedly will make a great contribution in understanding heredity as well as the biological and physiological phenomena accompanying gene action on heredity in rice. This knowledge must be used for both basic and applied plant sciences to improve rice plant, particularly its grain quality, to ensure a steady supply for human populations who depend upon it as a major source of staple. Genome analysis and the relatively new field of research on rice genomics are briefly reviewed in this article. I believe that the highlights of my discussion in this area will assist the scientific community in understanding the current status and direction of research priorities in rice genome analysis. While these findings may be useful in the study of other important cereal crops, their direct or indirect application in rice production is undoubtedly a major accomplishment in plant biology. As for proteome analysis, no prominent findings have been obtained in rice as those currently used in human genomics, such as genetically altered drugs against certain diseases. Although proteome analysis is still an untested strategy in rice breeding, this approach is also expected to enhance our understanding of some areas in basic research such as the interaction of signal transduction as a defence reaction against a devastating pathogen. The progress in genomics is fast and innovative, rendering the discovery of a wide array of novel methods in genomic research. Their ultimate adaptation as new knowledge and tools for research will pave the way for the much needed improvement in production of rice and other plant species.

1. Botstein, D., White, R. L., Skolnick, M. and Davis, R. W., *Am. J. Hum. Genet.*, 1980, **32**, 314–331.
2. Gusella, J. F. *et al.*, *Nature*, 1983, **306**, 234–238.
3. McCouch, S. R., Kochert, G., Yu, Z. H., Wang, Z. Y., Khush, G. S., Coffman, W. R. and Tanksley, S. D., *Theor. Appl. Genet.*, 1988, **76**, 815–829.
4. Kurata, N. *et al.*, *Nature Genet.*, 1994, **8**, 365–372.
5. Causse, M. A. *et al.*, *Genetics*, 1994, **138**, 1251–1274.
6. Harushima, Y. *et al.*, *ibid*, 1998, **148**, 479–494.
7. http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html.
8. Temnykh, S. *et al.*, *Theor. Appl. Genet.*, 2000, **100**, 697–712.
9. Kono, I., Takeuchi, Y., Shimano, T., Sasaki, T. and Yano, M., *Breed. Res.* (*in Japanese*), 2000, **2**, 197–203.
10. Yano, M. and Sasaki, T., *Plant Mol. Biol.*, 1997, **35**, 145–153.
11. Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y. and Sasaki, T., *Theor. Appl. Genet.*, 1997, **95**,1025–1032.
12. Yamamoto, T., Kuboki, Y., Lin, S.-Y., Sasaki, T. and Yano, M., *ibid*, 1998, **97**, 37–44.
13. Yano, M. *et al.*, *Plant Cell*, 2001, **12**, 2473–2484.
14. Tanksley, S. D., Young, N. D., Paterson, A. H. and Bonierbale, M. W., *Biotechnology*, 1989, **7**, 257–264.
15. Huang, N. *et al.*, *Theor. Appl. Genet.*, 1997, **95**, 313–320.
16. Moore, G., Foote, T., Helentjaris, T., Devos, K., Kurata, N. and Gale, M., *Trends Genet.*, 1995, **11**, 81–82.
17. Moore, G., Aragon-Alcaide, L., Roberts, M., Reader, S., Miller, T. and Foote, T., *Plant Mol. Biol.*, 1997, **35**, 17–23.
18. Ahn, S. and Tanksley, S. D., *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 7980–7984.
19. Kurata, N., Moore, G., Nagamura, Y., Foote, T., Yano, M., Minobe, Y. and Gale, M., *Bio/Technology*, 1994, **12**, 276–278.
20. Devos, K. M., Wang, Z. M., Beales, J., Sasaki, T. and Gale, M. D., *Theor. Appl. Genet.*, 1998, **96**, 63–68.
21. Nagamura, Y., Tanaka, T., Nozawa, H., Kaidai, H., Kasuga, S. and Sasaki, T., in Proceedings of the International Workshop on Utilization of Transgenic Plant and Genome Analysis in Forage Crops, National Grassland Research Institute, Japan, 1998, pp. 97–103.
22. Devos, K. M. and Gale, M. D., *Plant Mol. Biol.*, 1997, **35**, 3–15.
23. Kilian, A., Kudrna, N., Kleinhofs, A., Yano, M., Kurata, N., Stefenson, B. and Sasaki, T., *Nucleic Acids Res.*, 1995, **23**, 2729–2733.
24. Leister, D. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 370–375.
25. Peng, J. *et al.*, *Nature*, 1999, **400**, 256–261.
26. Sasaki, T. and Burr, B., *Curr. Opin. Plant Biol.*, 2000, **3**, 138–141.
27. http://genome.Washington.edu/consed/consed.html.
28. Burge, C. and Karlin, S., *J. Mol. Biol.*, 1997, **268**, 78–94.
29. Uberbacher, E. C. and Mural, R. J., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 11261–11265.
30. Lukashin, A. V. and Borodovsky, M., *Nucleic Acids Res.*, 1998, **26**, 1107–1115.
31. Salamov, A. and Solovyev, V., *Genome Res.*, 2000, **10**, 516–522.
32. Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. and Lipman, D. J., *J. Mol. Biol.*, 1990, **215**, 403–410.
33. Altschul, S. F., Madden, T. L., Scaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J., *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
34. Izawa, T. *et al.*, *Plant Mol. Biol.*, 1997, **35**, 219–229.
35. Hirochika, H., *ibid*, 1997, **35**, 231–240.
36. Sato, Y., Sentoku, N., Miura, Y., Hirochika, H., Kitano, H. and Matsuoka, M., *EMBO J.*, 1999, **18**, 992–1002.
37. Liu, Y.-G. and Whittier, R. F., *Genomics*, 1995, **25**, 674–681.
38. Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A. and Lukyanov, S. A., *Nucleic Acids Res.*, 1995, **23**, 1087–1088.
39. Miyao, A., Yamazaki, M. and Hirochika, H., *Plant Biotechnol.*, 1998, **15**, 253–256.
40. Richmond, T. and Somerville, S., *Curr. Opin. Plant Biol.*, 2000, **3**, 108–116.
41. Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T., *Plant J.*, 1994, **6**, 271–282.
42. Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P. and Potrykus, I., *Science*, 2000, **287**, 303–305.

ACKNOWLEDGEMENT. I thank Dr B. A. Antonio from the STAFF Institute for critically reviewing the manuscript.