# 13

# The Hybrid Origin Hypothesis of Cultivated Rice (*Oryza sativa*) Explains Some of the Gaps in Its RFLP Maps and Suggests an Efficient Mapping Population for Useful Genes and QTLs

G. Second[1,2*], A. Parco[1] and A. Caiole[1]

1 *International Rice Research Institute (IRRI), Manila, Philippines*
2 *Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM)*

**Two hundred and four mapped RFLP markers were surveyed on 148 accessions representing the range of genetic diversity in cultivated rice. For 54 markers, no polymorphism was detected. Some of the monomorphic markers are clustered on specific chromosome segments and correspond to unpopulated intervals in RFLP maps derived from intraspecific populations. It is suggested that these chromosome segments have been homogenized during domestication. The hybrid origin hypothesis of rice is consistent with this result. This hypothesis also suggests a way to efficiently map useful genes and QTLs based on the constructing of sets of "contig isolines" representing the entire genome of *indica* rice in a *japonica* background, and *vice versa*.**

## INTRODUCTION

The recent development of genetic maps based on molecular markers, RFLP in particular, has shown that markers are not distributed at random along the chromosomes. Some chromosome segments are consistently less populated than would be expected if markers were placed at random along the chromosomes. Various explanations have been proposed to explain this phenomenon, including unequal distribution of crossing-over along the chromosomes and regions of high recombination (Gill 1995), scoring errors (Lincoln and Lander 1992), stretches of repeated sequences, and common ancestors in the pedigree of the mapping population (McCouch and Tanksley 1991). We report here the observation that certain segments of chromosomes are defined by monomorphic markers in intraspecific comparisons, though polymorphism is observed for the same markers in interspecific comparisons. For this reason, a map based on an intraspecific mapping population is likely to lack markers in these regions. We suggest that this explains some of the unpopulated intervals in RFLP genetic

* Present address: *Centro Nacional de Pesquisa de Recursos Genéticos e Biotechnologia (CENARGEN), SAIN Parque Rural, Brasilia, D. F. Brasil*

maps of rice and that the explanation is likely to be applicable to many other species as well. We also point out that considering the biphyletic origin of cultivated rice, the construction of "contig-isolines" designed to capture individual chromosomal segments of the *indica* genome in a *japonica* background, and *vice versa*, would provide an efficient population structure for mapping useful genes.

## Materials and Methods

One hundred and forty-eight accessions were chosen to represent the breadth of diversity of *Oryza sativa*, based on agroecological types, geographic origin, isozyme and RFLP analysis. One genotype for each accession was analyzed at 204 RFLP loci that had been previously mapped in a backcross population derived from an *O. sativa* × *O. longistaminata* cross (Causse *et al.* 1994). Sixty-five of the markers had been screened and first mapped in an $F_2$ population derived from a cross between two *O. sativa* cultivars (McCouch *et al.* 1988). In most cases, the restriction enzyme that had been used to map the probes in the interspecific cross (Causse *et al.* 1994) was also used in this polymorphism survey. The identity of the probe/enzyme combinations used to map was available and could be verified by comparison of the results of this study with the molecular weights of bands derived in the mapping work (Susan McCouch, personal communication). Details concerning the RFLP data set associated with this analysis is accessible in the Rice Genome Database ("RiceGenes") through the National Agricultural Library in Washington D. C., USA, or through Gopher (McCouch and Paul 1994).

## Results

### *Non random distribution of polymorphic markers*

The probes were characterized as "monomorphic" when a single band was observed at the mapped locus for all 148 accessions and "polymorphic" when the size (molecular weight) of the DNA fragment differed among accessions. Markers were classified in two groups, those that were known to be polymorphic between cultivated varieties based on previous map construction (KP) (McCouch *et al.* 1988), and those for which information was not available

TABLE 13.1 Percentage of probes revealing RFLP according to their classification

| Probe | Polymorphic | Monomorphic | Total | % Monomorphic |
|---|---|---|---|---|
| **KP**†1 | | | | |
| RG < 600 | 51 | 14 | 65 | 21.5 |
| Various others | 4 | 0 | 4 | 0 |
| Subtotal | 55 | 14 | 69 | 20.3 |
| **KNP**†2 | | | | |
| RG > 600 | 15 | 5 | 20 | 25.0 |
| RZ | 59 | 24 | 83 | 28.9 |
| CDO | 21 | 11 | 32 | 34.4 |
| Subtotal | 95 | 40 | 135 | 29.6 |
| Total | 150 | 54 | 204 | 26.4 |

†1 Probes selected for polymorphism between two cultivated varieties
†2 Probes not selected for polymorphism between two cultivated varieties

(KNP). They were further classified as genomic (RG) or cDNA probes (CDO and RZ). Among the KP probes were two known genes, namely *waxy* and 45S rDNA, and two genomic probes from the *japonica* variety Nipponbare (Npb173 and Npb24) (Saito *et al.* 1991). Table 13.1 shows the proportion of monomorphic and polymorphic probes in each group. There is a striking difference in the percentage of monomorphic probes between the KP (20.3%) and the KNP (29.6%) groups that is not explained by the origin of the probes, whether genomic or cDNA.

The distribution of the monomorphic probes on the genetic map according to their classification as KP or KNP is shown in Fig. 13.1. While monomorphic KPs are scattered, with no more than two probes being contiguous, many of the monomorphic KNPs are found in cluster of four to five contiguous probes. The theoretical probability ($P$) of finding probes with no polymorphism in clusters of $n$ probes in a random distribution is given by the following formula, where $m$ is the proportion of monomorphic probes:

$$P = n(1-m)^2 m^n$$

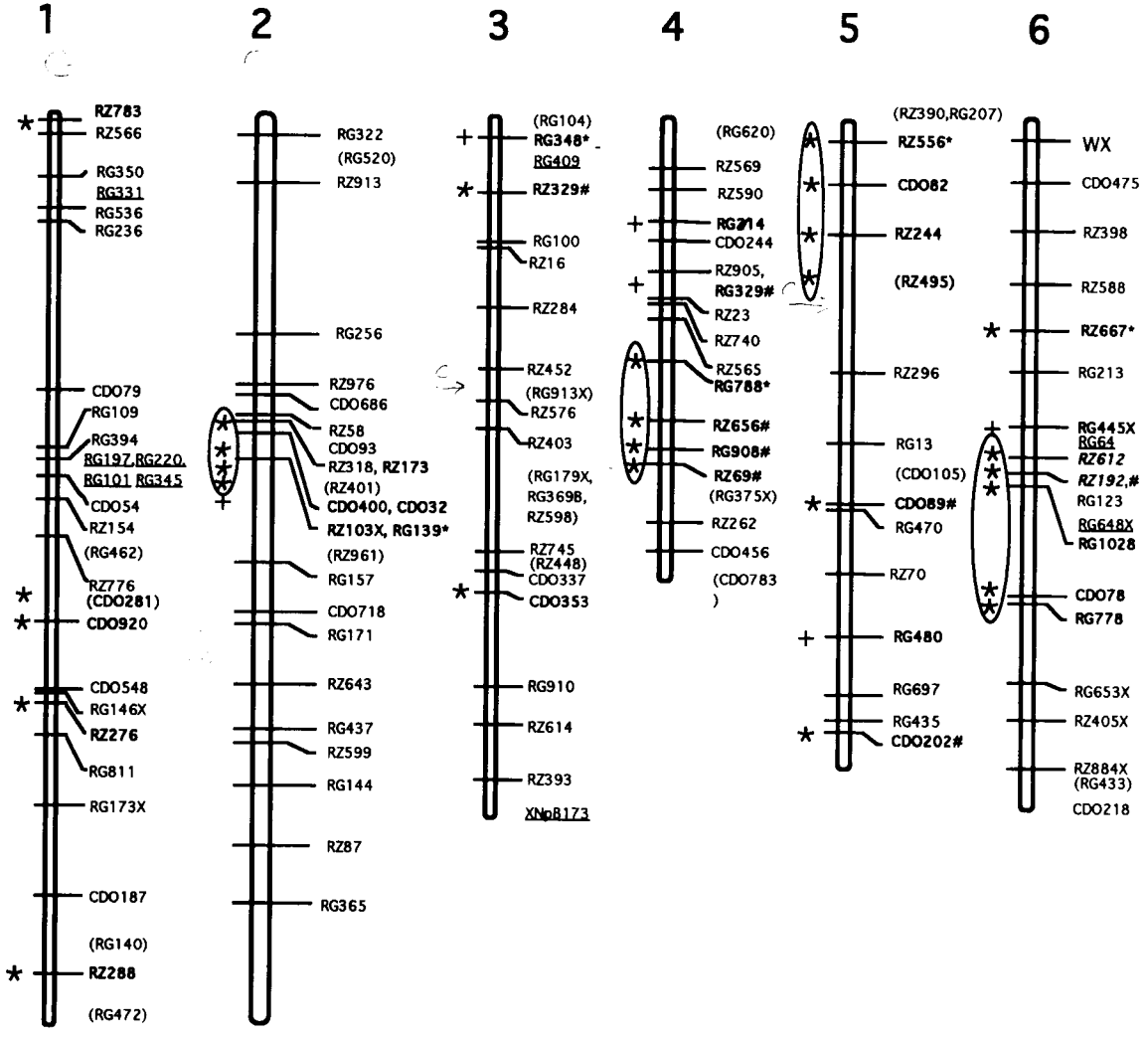
Table 13.2 shows the observed and expected numbers of monomorphic probes found in clusters of various sizes in the KP and KNP groups of probes. The distribution of monomorphic KP is not different from an expected random distribution, while that of the monomorphic KNP appears to be composed of two subgroups. One comprises 18 probes that appear to be scattered at random over the genome and the other comprises 22 probes that are in clusters of four to five probes.

*Additional survey of RFLP with several enzymes*

Twenty-two probes that did not show RFLP with the enzyme used to map them from segregations in the interspecific population were hybridized on the same DNAs restricted with a different enzyme. Among them, six belonged to the KP group and RFLP was indeed observed with the enzyme that had been used to originally map them. For the 16 other probes that belonged to the KNP group, only two showed RFLP (RG788 with the enzyme *Dra*I and RZ556 with the enzyme *Sca*I), although in total 27 probe/enzyme combinations were survey. It was thus concluded that these probes show a relatively small RFLP among cultivated varieties, irrespective of the enzyme used.

TABLE 13.2 Expected at random and observed numbers of monomorphic probes found in clusters of $n$ contiguous probes in KP and KNP groups

| Group | $n$ | Expected | Observed |
|---|---|---|---|
| **KP** | 1 | 8.90 | 8 |
| **KP** | 2 | 3.61 | 6 |
| **KP** | >2 | 1.50 | 0 |
| **KNP** | 1 | 19.80 | 16 |
| **KNP** | 2 | 11.72 | 2 |
| **KNP** | 3 | 5.21 | 0 |
| **KNP** | 4 | 2.05 | 12 |
| **KNP** | 5 | 0.76 | 10 |
| **KNP** | >5 | 0.27 | 0 |


1
RZ783
RZ566
RG350
RG331
RG536
RG236
CDO79
RG109
RG394
RG19Z,RG220,
RG101 RG345
CDO54
RZ154
(RG462)
RZ776
(CDO281)
CDO920
CDO548
RG146X
RZ276
RG811
RG173X
CDO187
(RG140)
RZ288
(RG472)
2
RG322
(RG520)
RZ913
RG256
RZ976
CDO686
RZ58
CDO93
RZ318, RZ173
(RZ401)
CDO400, CDO32
RZ103X, RG139*
(RZ961)
RG157
CDO718
RG171
RZ643
RG437
RZ599
RG144
RZ87
RG365
3
(RG104)
RG348*
RG409
RZ329#
RG100
RZ16
RZ284
RZ452
(RG913X)
RZ576
RZ403
(RG179X,
RG369B,
RZ598)
RZ745
(RZ448)
CDO337
CDO353
RG910
RZ614
RZ393
XNpB173
4
(RG620)
RZ569
RZ590
RG214
CDO244
RZ905,
RG329#
RZ23
RZ740
RZ565
RG788*
RZ656#
RG908#
RZ69#
(RG375X)
RZ262
CDO456
(CDO783
)
5
(RZ390,RG207)
RZ556*
CDO82
RZ244
(RZ495)
RZ296
RG13
(CDO105)
CDO89#
RG470
RZ70
RG480
RG697
RG435
CDO202#
6
WX
CDO475
RZ398
RZ588
RZ667*
RG213
RG445X
RG64
RZ612
RZ192,#
RG123
RG648X
RG1028
CDO78
RG778
RG653X
RZ405X
RZ884X
(RG433)
CDO218

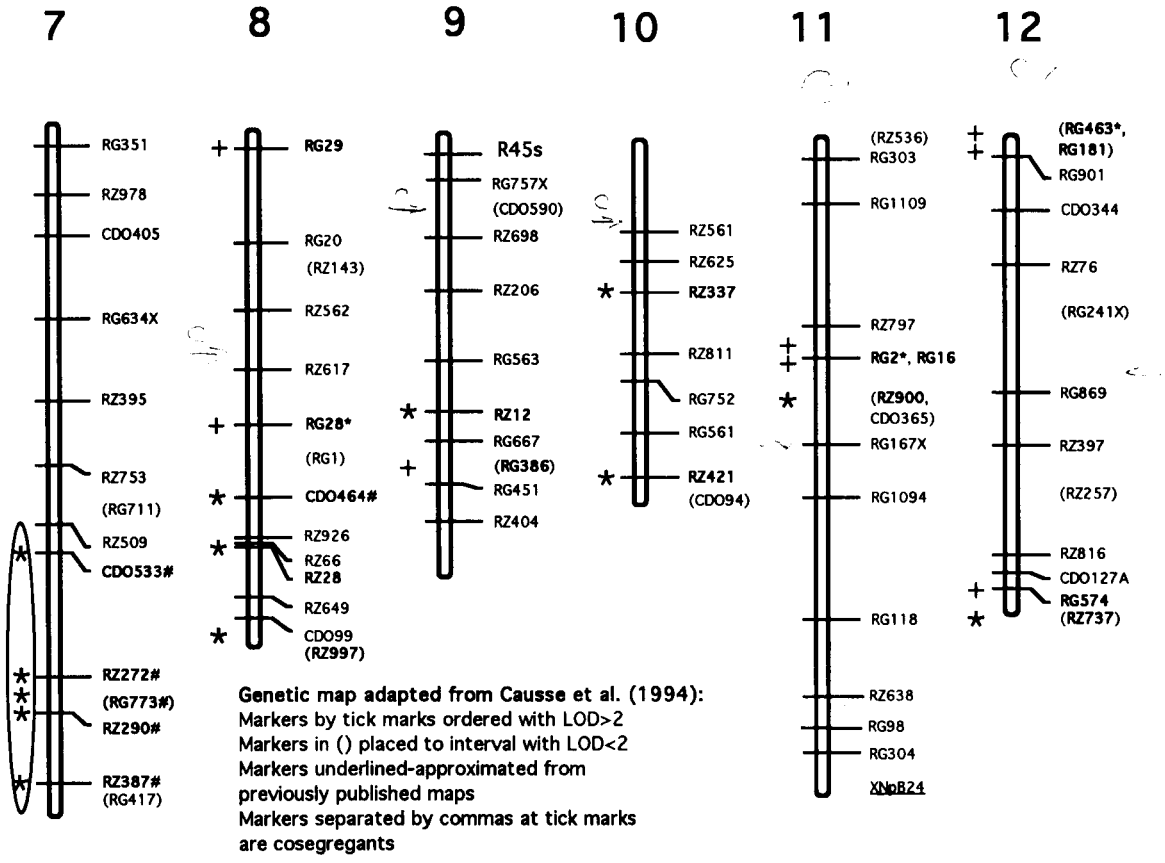



Fig. 13.1 Genetic map of the probes utilized
Bold characters indicate the probe did not revealed polymorphism in our set of cultivars, with the enzyme used to map it. Placed on the left of the chromosomes, + or ★ indicate such probe had been, or had not been, respectively, selected for polymorphism between two cultivated varieties prior to its mapping. Recognized clusters of such probes are encircled. ★ or # indicate additional restriction enzyme(s) did, or did not, respectively, revealed polymorphism. Italics indicate some doubt on probe identity.

## DISCUSSION

It was expected that probes selected for polymorphism between two cultivated varieties previously known to detect polymorphism (KP) would show a higher degree of polymorphism in other cultivated varieties than probes that were not (KNP). However, the non random distribution of monomorphic KNP probes is of interest. These probes tend to cluster in chromosome segments, and the hypothesis of a lower degree of RFLP polymorphism of these segments was confirmed by testing additional restriction enzymes and finding little increase in the frequency of polymorphism. Apart from the clusters of monomorphic probes, only 18 of 135 (13%) probes that had not been previously tested for polymorphism in *O. sativa* were monomorphic (based on a single probe/enzyme combination) in this study.

Additional evidence for the hypothesis that some chromosome segments are predictably less polymorphic for RFLP probes than other segments comes from a comparison of two high density rice RFLP genetic maps, one developed in Tsukuba in Japan from an intraspecific *O. sativa* population, and the other being the interspecific population developed at Cornell University in the USA. A formal alignment of these two maps has been published (Xiao *et al.* 1992) and further comparison is possible from more advanced mapping (Nagamura *et al.* 1993). Alignment of the two maps for chromosome 2 shows that there is a cluster of monomorphic probes on this chromosome which corresponds to a gap in the intraspecific map ("intra-map") but that this region is well populated with markers in the map developed from the interspecific cross ("inter-map"). This gap remained in the more saturated version of the intra-map (Nagamura *et al.* 1993). A similar observation can be made on chromosome 4. On chromosomes 5 and 6, the clusters indicated in Fig. 13.1 appear to correspond to areas which are less densely covered with markers in the intra-map. The most striking example is provided by chromosome 7. It is the only chromosome which is shorter in the intra-map (Xiao *et al.* 1992) as compared with the inter-map. Alignment of the two maps clearly shows that the cluster of markers on chromosome 7 corresponds to a segment that is missing in the intra-map. The more recent version of the intra-map (Nagamura *et al.* 1993) shows a chromosome 7 longer than in the inter-map but with an additional segment poorly covered with markers (a segment of nearly 50 cM is marked only by two RAPDs). This supports the existence of specific "monomorphic" chromosome segments within the species *O. sativa* and explains some of the gaps, or areas poorly covered by markers, in maps derived from intraspecific crosses.

The clusters of probes which are monomorphic in *O. sativa* revealed polymorphism between *O. sativa* and *O. longistaminata*, and this made it possible to map them (Causse *et al.* 1994). These probes also show polymorphism between the two cultivated rice species, *O. sativa* and *O. glaberrima* (unpublished results). It is thus likely that the chromosome segments they mark have been homogenized by common descent. Because this homogenization concerns the entire species, it is likely to have occurred during a bottle-neck in the effective size of the reproductive population for the *O. sativa* species.

The hypothesis of a hybrid origin of cultivated rice as depicted in Fig. 13.2 can explain this situation. Following hybridization between two incipient cultivars that were derived from divergent lineages of wild rice (or between an incipient cultivar and a divergent wild species of rice), it is hypothesized that male sterility (as well as other factors) encouraged backcrossing from which an introgression resulted. The homogenization of a chromosome fragment may have come about when one of the parental forms disappeared after selection or drift from the

population. Assuming that such introgressive hybridization often takes place during the domestication of cultivated species, we predict that many of the maps derived from an intraspecific domesticated population will have gaps that would be more easily filled from a true interspecific population (probably also the case for *Arabidopsis thaliana*). The use of markers with a higher level of allelic diversity (*i.e.*, for neutral mutations, higher mutation rates) could also overcome this problem.

The analysis of the structure of RFLP revealed by "polymorphic" probes in the present study (Second *et al.*, in preparation) indicates that extensive introgressive hybridization between the two dominant types of cultivated rice, the *indica* and *japonica* subspecies, is the origin of the diversity of cultivated rice. In a pair wise comparison of two cultivars, a patchy degree of polymorphism along the chromosomes, resulting in gaps in a genetic map, is due not only to the segments that are relatively monomorphic within the species but also from the segments that are shared specifically by the two varieties.

The hypothesis of a hybrid origin of cultivated rice is reinforced by the results of this study. We would like to point out that this hypothesis also suggests an efficient way to map a large proportion of useful genes and QTLs in rice based on a single cross. Considering that rice has essentially two wild progenitors, and that successive bottle-necks in the reproductive population size during the domestication process are likely, the allelic diversity derived from each ancestor is low. This means that there is essentially one allele from each progenitor (at

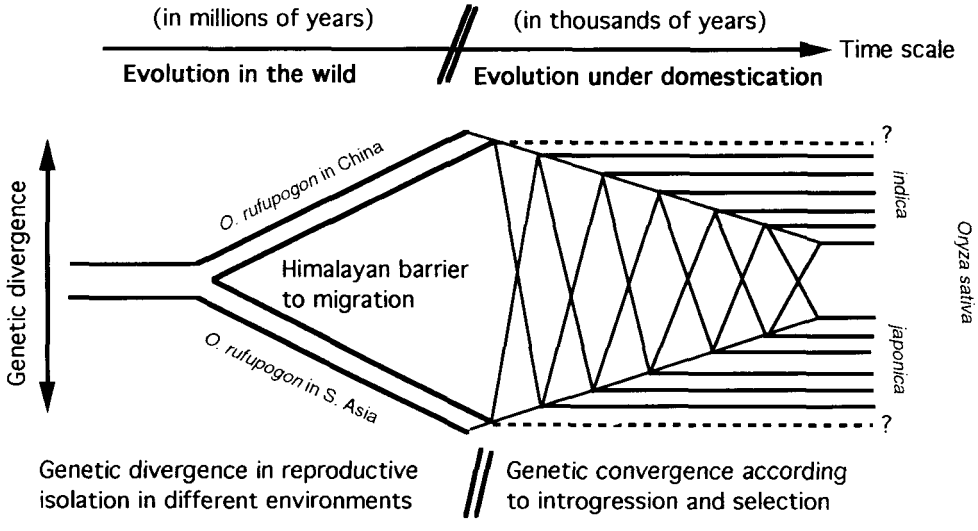



Fig. 13.2 Phylogenetic relationships in *O. sativa* according to the hypothesis of its hybrid origin (adapted from Second 1992). A vertical axis represents genetic divergence and a horizontal axis represents time with a one thousand fold difference in scale, whether before domestication (estimated time two millions years) and after domestication (estimated time at ten thousands years ago). The allelic diversity of the wild ancestors (thick branches) was observed reduced in the cultivars (thin horizontal branches ) as explained by the founder effect during the domestication process. Reticulation is meant to represent recurrent reciprocal introgression in the course of domestication between the *indica* and *japonica* subspecies. Dashed lines with question marks indicate the original *indica* and *japonica* domesticates. As suggested by the currently reported homogeneous chromosome segments, at least one of them was lost in the process.

least for a large majority of genes and for adaptive mutations). The adaptive genetic diversity comes primarily from recombination between loci (we have verified that recombination between one *indica* and one *japonica* parent does generate diversity in recombinant inbreds that approximates the diversity of the whole species for several characters — amylose content, architecture of the plant, tolerance to the tungo virus etc.). It is thus possible to generate a large proportion of the diversity of the species that is related to the two wild progenitors, from a single cross involving one *indica* and one *japonica* parent. Based on the development of "contig-isolines" this diversity can be compartmentalized. Mapped markers can be used to prepare isolines having one fragment derived from an *indica* parent in a *japonica* background and the reciprocal situation where one *japonica* fragment exists in an *indica* background. With a map of about 2,000 cM, 100 isolines with overlapping introgressed fragments of 20 to 30 cM long could be developed which represent the *indica* genome in a *japonica* background and *vice versa*, with another 100 isolines for the *japonica* genome. Isolines offer many advantages including true breeding and permanent propagation, as well as limiting the influence of the genetic background. $F_2$s derived from crosses with the recurrent parents could be used to confirm the precise location of the fragments responsible for observed differences between the lines.

*Acknowledgments*: We thank Dr. Susan McCouch for kindly editing the manuscript.

## REFERENCES

Causse MA, Fulton TM, Yong GC, Sang NA, Chunwongse J, Wu K, Xiao JH, Yu ZH, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994): Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138: 1251–1274.

Gill BS (1995): Molecular Cytogenetics: An intergrated approach for evolutionary, structural and functional analysis of plant chromosomes. *In*: Tunewaki K (ed) (1995) *Plant Genome and Plastome: Their Structure and Evolution*. Kodansha, Tokyo. pp. 23–31.

Lincoln SE, Lander ES (1992): Systematic detection of errors in genetic linkage data. *Genomics* 14: 604–610.

McCouch SR, Paul E (1994): Rice genes, an international genome database and bulletin board for rice. *DNA Link* 3: 40–41.

McCouch SR, Tanksley SD (1991): Development and use of restriction fragment length polymorphism in rice breeding and genetics. *In*: Toenniessen G, Khush G (eds) *Rice Biotechnology*. CAB International, Tucson, Arizona, USA. pp. 109–133.

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988): Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* 76: 815–829.

Nagamura Y, Yamamoto K, Harushima Y, Wu JZ, Antonio BA, Sue N, Shomura A, Lin SY, Miyamoto Y, Toyama T, Kirihara T, Shimizu T, Wang ZX, Tamura Y, Ashikawa I, Yano M, Kurata N (1993): A high-density STS and EST linkage map of rice. *Rice Genome* 2(1): 3.

Saito A, Yano M, Kishimoto N, Nakagahra M, Yoshimura A, Saito S, Kuhara K, Ukai Y, Kawase M, Nagamine T, Yoshimura S, Ideta O, Ohsawa R, Hayano Y, Iwata N, Sugiura M (1991): Linkage map of restriction fragment length polymorphism loci in rice. *Japan. J. Breed.* 41: 665–670.

Second G (1982): Origin of the genetic diversity of cultivated rice (*Oryza* spp.), study of the polymorphism scored at 40 isozyme loci. *Jpn. J. Genet.* 57: 25–57.

Xiao J, Fulton T, McCouch S, Tanksley S, Kishimoto N, Ohsawa R, Ukai Y, Saito A (1992): Progress in integration of the molecular maps of rice. *Rice Genet. Newsletter* 9: 124–128.

# Plant Genome and Plastome:

## Their Structure and Evolution

*In Commemoration of Professor Hitoshi Kihara's Centennial*

Edited by
Koichiro Tsunewaki,
Professor Emeritus, *Kyoto University* and
Professor, *Fukui Prefectural University, Japan*

KODANSHA SCIENTIFIC LTD.
Tokyo