The Japanese Rice Genome Research Program.

T Sasaki, M Yano, N Kurata, et al.

Genome Res. 1996 6: 661-666
Access the most recent version at doi:10.1101/gr.6.8.661

REVIEW

Around the Genomes

The Japanese Rice Genome Research Program

Takuji Sasaki,[1] Masahiro Yano, Nori Kurata, and Kimiko Yamamoto

Rice Genome Research Program, National Institute of Agrobiological Resources /Institute of the Society for Technoinnovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305, Japan

---

Through his experiments using garden peas, Mendel's recognition of the significance of the number 2 in inheritance, heralded the dawn of theoretical genetics. However, in agriculture, for centuries farmers had improved crops and vegetables by selection based on desired traits such as adaptability to the environment or quantity of product. They planted seeds from individuals of preferred traits and continued further selection to fix those characteristics. A number of plant varieties have been established in this way, and the method used by farmers was based on Mendelian principles. Even since the discovery of the mechanism of inheritance by Watson and Crick, breeding has been largely based on the crossing of favorable traits and selection after observation of phenotypes among the progeny. The main reason genetics has not always been used to improve agricultural products is that there has been a lack of information linking phenotype and genotype, particularly given the complicated nature of the inheritance of important agronomical traits.

Analysis of the human genome promises to help elucidate the characteristics of inherited diseases and to provide the possibility of overcoming them. It is equally rational that staple foods necessary to sustain human life should be chosen as targets of genome analysis, with the goal of accelerated production and improvement. Rice is one of the most important crops for people living in the Asian monsoon area. It has been estimated that ~50% of the human population depends on rice as their daily food (White 1994). Rice is cultivated under a wide range of environments, from arid highlands to flooded lowlands. The mechanisms by which a single species, *Oryza sativa*, can adapt to such wide conditions must be programmed into its genome, and an understanding of these mechanisms should contribute to an increase in rice production.

The compactness of the rice genome is a distinct advantage in gene isolation and genomic sequencing. Its 430-Mb genome is the smallest among the Gramineae and is only three times larger than that of the model plant *Arabidopsis*. An additional reason for studying the rice genome is its apparent synteny with the much larger, often polyploid genomes of other grasses used as crop plants, such as wheat, barley, and maize (Ahn and Tanksley 1993; Kurata et al. 1994a; Saghai Maroof et al. 1996). Thus the results of genomic analysis in rice could have direct relevance for many of the major food sources of human populations.

For these reasons, rice has been chosen as a target species for genome research by a number of groups. The International Rice Research Institute (IRRI) in the Philippines, collaborating with a group at Cornell University, published the first rice linkage map based on restriction fragment length polymorphism (RFLP) markers (McCouch et al. 1988), and the Cornell group subsequently developed another map with 726 DNA markers (Causse et al. 1994). There are six rice genome research groups in China; the central group, in Shanghai, recently reported construction of bacterial artificial chromosome (BAC) contigs covering 98% of the rice genome (Tao et al. 1995). A group in Korea that is carrying out random cDNA analysis of a rice panicle cDNA library is making the information from their project available from their World Wide Web (WWW) server (http://bioserver.myongji.ac.kr). In Japan, rice genome research was initiated by the Japanese Ministry of Agriculture, Forestry and Fisheries (MAFF). The Rice Genome Research Program (RGP) began in 1991 as a collaboration between the National Institute of Agrobiological Resources and the Society for Technoinnovation of Agriculture, Forestry, and Fisheries. The program has three main

---

[1] **Corresponding author.**
E-MAIL tsasaki@abr.affrc.go.jp; FAX 81 298 38 7468/2302.

6:661–666 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

research themes: genetic mapping, physical mapping, and random cDNA analysis. In this report we have described the current progress and future goals of the Japanese RGP, focusing on genetic mapping. RGP also exchanges information and materials with the other groups engaged in rice genome research, and work is under way to integrate the Cornell genetic maps with the RGP map.

Genetic Mapping

A high-density linkage map of the rice genome has been constructed based on RFLP markers. The parental lines for the cross were determined by choosing combinations of rice varieties that showed a high frequency of RFLPs and yielded fertile F_1 seeds in sufficient amounts to enable an F_2 population to be generated. The combination of japonica variety Nipponbare and indica variety Kasalath satisfied these requirements. After extraction of DNA from the green leaves of each progeny plant, eight different restriction enzymes were used to cleave the DNAs. RFLPs were detected on Southern blots probed mainly by rice cDNA clones, using the enhanced chemiluminescence system (ECL) of Amersham. Because of the small size of the rice genome, 2 μg of rice DNA loaded in each electrophoresis well gave sufficiently strong luminescence with the ECL system. In addition, as removal of previously used probes is not necessary in ECL, we were able to hybridize a single filter >30 times to make a high-density linkage map with nearly 2000 DNA markers from a single F_2 population of 186 plants.

In Figure 1, the skeleton of the resulting linkage map, which evolved from the previous map published in 1994 (Kurata et al. 1994b), is shown. The details of this map will be published soon after rechecking the locus assignment of each RFLP. On this new map, we have increased the number of markers by ~600, using mainly rice cDNAs. We have focused on rice cDNAs because a clear Southern pattern and polymorphism is usually obtained, and because loci can often be identified by performing sequence similarity searches of publicly available data bases. For example, putative gene identifications for RFLP markers on a part of chromosome 6 are shown in Figure 2. Of the 24 loci shown, 16 show significant similarity to genes previously identified in other species, including other types of plants.

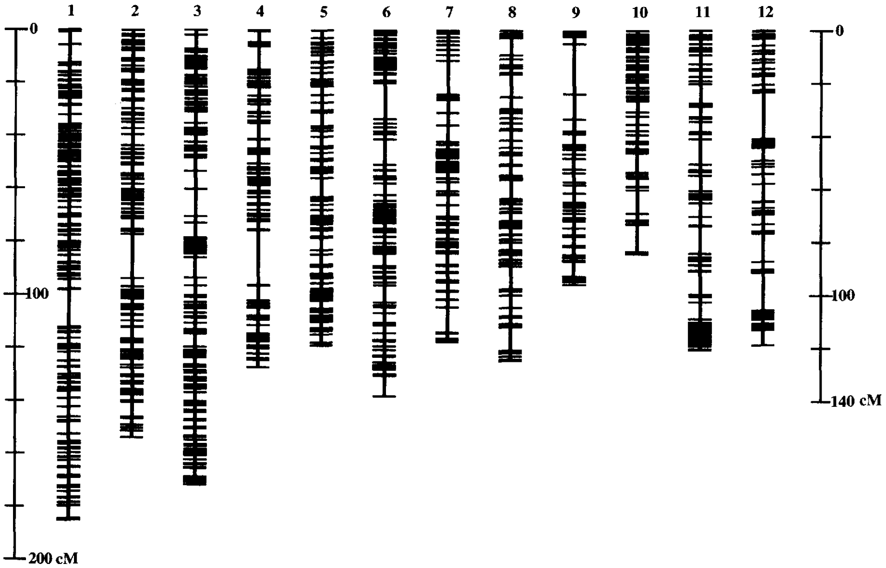
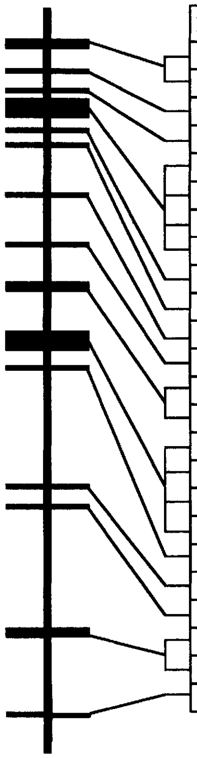


Figure 1 Rice genetic linkage map with 1904 markers covering a total of 1556 cM of genetic distance. The horizontal bars indicate loci. The DNA markers are derived from rice cDNAs, rice genomic DNAs, *Not*I-linking clones, YAC end clones, RAPD markers, STS markers, telomere-associated clones, wheat DNA clones, and barley DNA clones.

Because the mapped cDNAs were generally obtained from tissues of plants grown under normal physiological conditions, constitutively expressed genes that play key roles in important physiological traits, not only in the parental strains but also in other cultivars, may be included among them (see below).

Utility of the Genetic Map

Variability in quantitative traits, such as plant height, seed weight, and flowering time is the net result of mutiple independently segregating genetic loci, or quantitative trait loci (QTL), that interact with each other and the environment. Hillel et al. (1990) demonstrated the use of DNA polymorphisms to identify gene introgression by backcrossing; this strategy can be used to identify individual genes that contribute to a quantitative trait. To evaluate the contribution of each candirice chromosome 6



Locus name	distance (cM)	accession number	protein name (sequence ID in ENTREZ)	species
C1496	10.2	D15906	proteasome chain 7 (gi:464445)	rat
C1084	10.2	D15683	none	
C688	10.5	D15472	early nodulin 93 (gi:218262)	soybean
R1952	10.8	D24457	none	
R2291	11.1	D24638	protein kinase(gi:862342)	tobacco
C1368	11.1	D28225	none	
R2750	11.1	D24908	cds, unknown function (gi:1204225)	S.pombe
R1962	11.1	D28309	none	
R2749	11.4	D24907	cysteine synthase (gi:218277)	spinach
S10283B	11.7	D45967	aminolevulinate dehydratase(gi:1041423)	barley
C764	12.5	D15525	proteasome (gi:1262338)	human
C425B	13.4	D15300	histone H3.3-like protein (gi:16324)	Arabidopsis
C1032	13.8	D15657	floral homeotic protein AGL3 (gi:113513)	Arabidopsis
R2634	13.8	D24835	none	
C226A	14.3	D15187	T11G6.8 (gi:1184598)	C.elegans
C191B	14.3	D15165	none	
S2069	14.3	D40238	none	
R845	14.3	D28293	cystathionine beta-lyase (gi:704397)	Arabidopsis
S1434	14.6	D39822	pyrophosphate-energized inorganic pyrophosphatase (gi:285638)	barley
C498	16.3	D15349	ribosomal protein L39 (gi:1177369)	maize
S1568	16.6	D39908	CP-MII.2=serine carboxypeptidase(gi:619351)	barley
R1966	18.8	D24462	sucrose synthase 2 (gi:20095)	rice
R1014	18.8	D24051	nascent polypeptide associated complex alpha chain (gi:556642)	human
R1954	19.9	D24459	none	

Figure 2 A part of the enlarged rice RFLP linkage map of chromosome 6 with putative protein names for each locus. BLASTX was used to carry out a similarity search on the nonredundant protein data base in National Center for Biotechnology Information (NCBI) on May 15, 1996. The name of the protein that is ranked at the top of the list of similarity for each cDNA is shown, together with its sequence identification in the ENTREZ data base released by NCBI, if it showed a similarity score of >100. (None) No siginificant similarity was observed for the cDNA.

date area in the genome, near-isogenic lines can be constructed effectively and efficiently, using a saturated genetic linkage map (Yano et al. 1996). In the RGP, this strategy is being applied to identify QTL for the heading time of rice. Five loci have been identified and tagged tightly by RFLP markers (T. Yamamoto, Y. Kuboki, S.-Y. Lin, T. Shimano, A. Showmurd, M. Sato, M. Yano, and T. Sasaski, unpubl.). One of these loci turned out to be tightly linked to a gene that shows sequence similarity to the *CONSTANS* gene, a zinc finger motif gene that controls flowering time in *Arabidopsis* (Putterhill et al. 1995). This example further demonstrates the usefulness of cDNAs in genetic analysis. To date, ~30,000 cDNA clones have been analyzed as part of the RGP.

Another important use of high-quality linkage maps for rice is to analyze the syntenic relationship between rice and other crops. Over many years of plant breeding research, detailed genetic analyses of phenotypes have been carried out in maize, barley, wheat, and other crops. However, the large genome sizes of these plants present a considerable obstacle to obtaining molecular information about agriculturally important phenotypes. If the genomic structure of rice is homologous to those of other grasses and they all have similar numbers of genes, the small genome of rice will make it a useful model for gene characterization. For example, stem rust disease caused by interaction between *Puccinia graminis* and a corresponding resistance gene is observed in barley and not in rice. However, a detailed analysis of rice/barley synteny around the resistance gene has allowed the region containing the gene to be narrowed to ~30 kb (Kilian et al. 1995). Even though this *Puccinia* fungus is not of concern in rice, rice yeast artificial chromosomes (YACs) hybridized to cDNAs synthesized from barley carrying the resistance gene. This observation might provide a clue to understanding how the interaction between host and pathogen can be so specific even though the homologous genes exist in both sensitive and resistant host species, and in resistant nonhost species.

As well as disease-resistant gene homologs, genes contributing to QTL are also expected to be found in syntenic regions of different grass species (Paterson et al. 1995). The location of QTL for heading time in rice, maize, and barley is well explained by the synteny of genomic structure among these species.

Conversion of a Genetic Map to a Physical Map

To convert the genetic map to a one of aligned nucleotide sequences, an array of YAC clones is being constructed. Our YAC library, constructed from one parent (Nipponbare) of the F_2 population used to make the linkage map, contains 7000 clones with an average insert length of 350 kb, corresponding to 5.5 genome equivalents (Umehara et al. 1995). Using a high-density filter of these YACs, colony hybridization has been carried out for each RFLP marker on the linkage map. In the case of sequence-tagged site (STS) markers, the conventional three-dimensional pooling method allowed YACs to act as template DNAs for PCR. To identify the exact location of positive YACs on the linkage map, Southern blot analysis is required after digestion by the corresponding restriction enzyme. To date, a total of ~1000 DNA markers have been used to screen the YAC library, and approximately half of the rice genome is estimated to be covered by YAC clones.

Although this contiguous "chromosome landing" strategy (Tanksley et al. 1995) does not necessarily identify overlaps between neighboring YAC clones, it is nonetheless useful for making a direct comparison between the genetic and physical maps, aiding map-based cloning of genes. In addition, the physical distance corresponding to a gap or a clustered region observed in a genetic map could be clarified and could help in identifying the molecular features associated with differences in recombination probability (Fig. 3).

Data Storage and Data Base Development

Because the data elements produced by the RGP have tight links to one another, we have developed an in-house relational data base, RiceBase3, which uses Sybase software operating on a Unix machine and consists of ~100 linked data tables. For easier detection of relationships among frequently used data, we have also constructed another type of data base, called RiceDB, which is based on the structure of ACeDB, the nematode genome data base. At present, RiceDB is available only by anonymous ftp from the WWW server at the address http://bank.dna.affrc.go.jp.

We have published 10,990 partial nucleotide sequences of cDNAs through DDBJ/EMBL/GenBank. These cDNAs are available from the DNA Bank of MAFF; their names and the Southern hybridization patterns of DNA markers on our published map (Kurata et al. 1994) are available at the WWW address given above.

Future Plans

By genetic and structural analysis of the rice genome, the RGP aims to identify and characterize the genes controlling agronomically important traits in rice. To achieve this goal it will be essential to combine information from the RFLP linkage map, the physical map, and the catalog of expressed genes. One immediate priority is to accelerate construction of the physical map by screening the YAC library with >1000 DNA markers and by identifying the overlaps between neighboring YAC clones. cDNAs that cannot be assigned to the genetic map because of nonpolymorphism will be located on the partially ordered YACs by Southern hybridization. In some cases, this may reveal overlaps between neighboring YACs.

For rapid identification of the genes controlling many traits, mutant lines precisely tagged with our DNA markers are required. This work should yield information that will be useful in marker-assisted selection schemes.

Recently, a program to sequence the entire genome of *Arabidopsis* was announced (Goodman et al. 1995). A future goal will be to initiate a similar program for rice, to yield both comparative and independent genomic information to that for *Arabidopsis*. However, to realize this idea, a well-established automated system for rapid and reliable genomic sequencing will be required, as well as a higher-capacity computer system for summarizing and editing the sequence information. This sort of computer system is completely different from that maintained in the present RGP. Nevertheless, we have obtained par-

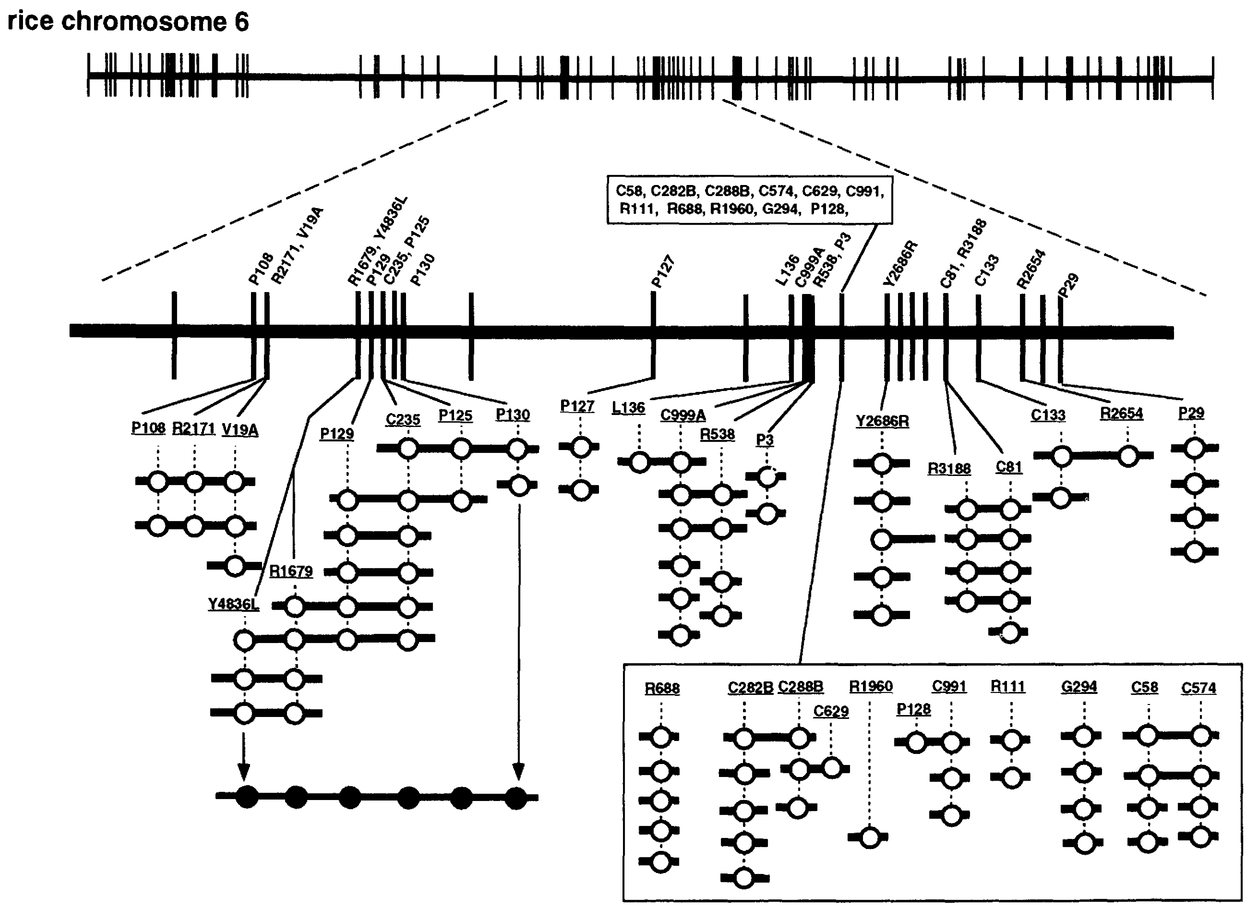


Figure 3 An example of the alignment of YAC clones along the linkage map of chromosome 6. Bars with open circles indicate YAC clones. From Y4836L to P130, the nine YAC clones that overlap are represented as a contig shown here as a black array. Eleven DNA markers that cosegregated at one locus shown in a box above the enlarged linkage map hybridized to >20 YAC clones that do not necessarily overlap. The physical distance corresponding to this genetic locus is estimated to be >1 Mb.

tially ordered YACs and identified several interesting regions for sequencing. New, rapid sequencing technology will be applied to small-scale sequencing of genomic fragments such as cosmids.

The present 7-year program will end in March 1998. During the remaining 2 years, we expect to complete the preparation of the basic tools for rice genome research. A second-phase rice genome program will be needed to establish effective applications of the information and materials obtained by the current RGP.

ACKNOWLEDGMENTS

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

REFERENCES

Ahn, S. and S.D. Tanksley. 1993. Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci.* **90:** 7980–7984.

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

Goodman, H.M., J.R. Ecker, and C. Dean. 1995. The genome of Arabidopsis thaliana. *Proc. Natl. Acad. Sci.* **92:** 10831–10835.

Hillel, J., T. Schaap, A. Haberfeld, A.J. Jeffreys, Y. Plotzky, A. Cahaner, and U. Lavi. 1990. DNA fingerprints applied to gene introgression in breeding programs. *Genetics* **124:** 783–789.

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

Kurata, N., G. Moore, Y. Nagamura, T. Foote, M. Yano, Y. Minobe, and M. Gale. 1994a. *Bio/Technology* **12:** 276–278.

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

McCouch, S.R., G. Kochert, Z.H. Yu, Z.Y. Wang, G.S. Khush, W.R. Coffman, and S.D. Tanksley. 1988. Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* **76:** 815–829.

Paterson, A.H., Y.-R. Lin, Z. Li, K.F. Schertz, J.F. Doebley, S.R.M. Pinson, S.-C. Liu, J.W. Stansel, and J.E. Irvine. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Nature* **269:** 1714–1718.

Putterill, J., F. Robson, K. Lee, R. Simon, and G. Coupland. 1995. The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80:** 847–857.

Saghai Maroof, M.A., G.P. Yang, R.M. Biyashev, P.J. Maughan, and Q. Zhang. 1996. Analysis of the barley and rice genomes by comparative RFLP linkage mapping. *Theor. Appl. Genet.* **92:** 541–551.

Tanksley, S.D., M.W. Ganal, and G.B. Martin. 1995. Chromosome landing: A paradigm for map-based gene cloning in plants with large genomes. *Trends Genet.* **11:** 63–68.

Tao, Q., Y. Qian, H. Zhao, S. Yu, L. Qiu, B. Wu, J. Zhu, D. Yu, X. Liu, and G. Hong. 1995. Construction of Oryza sativa genome contigs by finger-print strategy. *Cell Res.* **5:** 263–271.

Umehara, Y., A. Inagaki, H. Tanoue, Y. Yasukochi, Y. Nagamura, S. Saji, Y. Ohtsuki, T. Fujimura, N. Kurata, and Y. Minobe. 1995. Construction and characterization of a rice YAC library for physical mapping. *Mol. Breeding* **1:** 79–89.

White, P.T. 1994. Rice: The essential harvest. *Natl. Geogr.* **185:** 48–79.

Yano, M., Y. Nagamura, N. Kurata, T. Sasaki, and Y. Minobe. 1996. Toward the construction of an integrated genetic and physical map of rice. In *Genomes of plants and animals. Twenty-first Stadler Genetics Symposium* (ed. J.P. Gustafson and R.B. Flavell), pp. 57–72. Plenum Publishing, New York, NY.