
Recent Progress in Gene Isolation and Gene Mapping: Implications for Agricultural Biotechnology

Richard B. Flavell
John Innes Centre
Norwich Research Park
Norwich, UK

The improvement of plants and animals for specific purposes has been at the heart of economic prosperity and stability of societies throughout history. It remains so today for all societies, although the economic importance of domesticated crops and animals differs widely between countries. The basis of improving organisms for specific purposes is changing combinations of genes. It always has been that way and always will be, because the properties and performance of one individual of a species versus another, in a given environment, is closely related to the genetic differences between them. "Classical" plant and animal breeders have achieved major improvements by selecting parents, making crosses and selecting progeny that serve human beings much better than wild strains in domesticated agriculture. However, relatively few species have been subjected to intense artificial selection. Many improvements have been made by the incorporation of a few genes, and sometimes a relatively rare version of a single gene, into already successful plant cultivars. Most advances have resulted from recombining versions of many genes into new and better combinations.

In spite of substantial progress achieved by breeders in the past, our crops and animals still do not serve us optimally because the right genes and gene combinations have not been found. Furthermore, our needs and rural, economic and political environments keep changing. Preferences for land usage and local demands for food, fiber and feed also rarely remain constant. Thus, the objectives for breeders keep changing and changes are expected to continue.

Our crop plants and forests regularly suffer in yield and quality due to environmental stresses of flooding, drought, heat and cold. They also suffer from pests and diseases, sometimes to dramatic extents. Often, yields are sustainable only with inputs of artificial fertilizer, fungicides and insecticides that may not always be available or are considered undesirable because of their perceived effects on the environment or on human health. There is concern over undesirable linkages between human health and possible "toxins" in food. Furthermore, improvements in appearance of food, taste, cooking properties and other quality traits are continually being demanded. Improved plants and trees as raw materials for industrial processes are also needed around the world.

Improvements in all these vital characteristics can be made by utilizing genes in new ways and in new combinations. I would maintain it is a very high priority to continue to do so. Societies that fail to adapt crops and animals to its needs and to changing environments will certainly fail. In this short paper I highlight some of the progress and issues of plant biotechnology that are emerging from discoveries of genes and gene mapping. Plant biotechnology is emphasized because this is my own area of specialty. Nevertheless, most of the salient conclusions are relevant to farm animal biotechnology.

Gene Mapping

Gene mapping, that is the determination of the localization of genes in a genome, is now an integrated part of the process of improving the genetic constitution of organisms to serve humans better. We need gene mapping to:

- Know how traits, simple or complex, are genetically determined;
- Make selection of gene combinations more efficient in breeding programs;
- Assess genetic diversity within a species at defined loci; and
- Identify the role of isolated genes in specifying phenotypic traits.

We have vast stores of genes and variants of genes in the organisms of the earth. It is essential that we maintain good representative samples of gene variants and combinations of genes in seed and embryo banks. Plant breeders have utilized some of these genes over the centuries without, of course, knowing of their existence as chemical, functional units; as genes. Before the recent phase of molecular biology, we knew of the existence of a gene only when two individuals had different variants of the gene and the alternative form could be identified in the progeny of a sexual cross and mapped to the same chromosomal position. This was the genetics that Mendel opened up for us.

To map one gene locus on a chromosome a second identifiable gene is needed, because gene mapping involves localizing one gene with respect to another. Major problems in producing maps have been a shortage of genes easily identified from their effects on the properties of an organism and the shortage of reference, identifiable marker genes in the same individuals.

Nevertheless, in a few plant species that have been the subject of much attention by geneticists (maize and pea) genetic maps displaying many localized genes have been produced. Now, thanks to the introduction of the techniques of molecular genetics, the opportunities are very different. First, it is not necessary to have variation in the effects of a gene to identify it – it can be identified by its chemical characteristics. Second, molecular reference markers that map all over the genome are readily created for many species. These detect variation in DNA sequences that are so abundant that it usually exists in all individuals. These two advances have revolutionized the making of genetic maps.

Maps of maize, soybean, wheat, rice and pine, for example, have been created in the last few years with many hundreds or thousands of markers. It is now possible to generate a comprehensive molecular marker map from scratch in less than one year, if plants with segregating genetic variation are available. The commonly used molecular markers are:

1. Short fragments of DNA (probes) that hybridize to one or only a few sites per genome and are used to find fragments in total DNA cleaved by restriction endonucleases. Variation between individuals in endonuclease recognition sites leads to fragments of differing length (Restriction Fragment Length Polymorphism, RFLP) which are easily recognized using the DNA probe after fractionation of the DNA fragments into different size classes.
2. Fragments of DNA in one or two copies per genome that terminate in specific short sequences that are recognized *in vitro* by DNA primers that can initiate DNA synthesis and consequently be amplified in the polymerase chain reaction (PCR). Variations in the primer binding sequences or in the length of DNA between the primer binding sites give rise to differences in the product of the PCR.
3. Fragments of DNA recognized by a combination of: 1. and 2. above (Amplified Fragment Length Polymorphism, AFLP).

Geneticists and breeders are now in a phase of rapidly expanding genetic maps using molecular DNA markers and genes specifying easily scorable properties in the phenotype or in the test tube.

Where DNA sequence or other easily scored markers are present all over the genome it is much easier to discover the genetic components (QTLs) of genetically complex traits and map their location reasonably accurately. This opportunity is providing geneticists and breeders with the means of locating very important genes that were previously unrecognizable.

IDENTIFICATION OF GENES VIA LARGE-SCALE GENOME PROJECTS

The identification of genes and their mapping is growing by the application of DNA sequencing techniques. Gene products are being sequenced as well as the genes themselves. Messenger RNAs, present wherever genes are expressed in

cells, can readily be copied in vitro into DNA using the enzymes reverse transcriptase and DNA polymerase. The DNA copies can be amplified in bacteria, isolated and readily sequenced. Parts of such DNA copies have been sequenced from some 18,000 *Arabidopsis* and 20,000 rice genes, for example. Assuming plants contain 20,000 to 30,000 different sorts of genes, clearly most genes of these species will be characterized by sequencing in the relatively near future. Once the sequences are in databases, they can be compared with all other genes similarly sequenced in any other species. Now that the functions of so many genes and parts of genes are becoming known in bacteria, yeast, *Drosophila*, *C. elegans* and mammals, clues to the function of an unknown gene can be readily obtained. Thus, plant gene identification is now being aided by the molecular genetics of many species across the kingdoms — a very significant change for plant scientists.

To get a more complete understanding of a gene's function it is useful to discover if it maps on a chromosome at a site known to influence a specific character. Also, it is desirable to insert the gene into the species to inactivate the resident gene and then observe the change to the organism. This is readily done in many plant species, even though it is relatively time consuming. Alternatively, if some gene sequence is known, it is possible to screen large populations of plants to find individuals where a roving transposable element has been inserted into or near the gene of interest and see what effects the mutation, insertion (and inactivation), has on the individual.

The application of these approaches to *Arabidopsis* and rice, as representatives of the dicot and monocot groups of plants, is providing molecular markers and tools for exploring all plant genomes. These tools have greatly expanded our knowledge of genes and ways of mapping genes and ascribing functions to them.

Another approach to identifying and mapping genes is to sequence long chromosomal regions, discover the genes from the sequence, scrutinize each sequence against other known genes, and investigate the effects of mutations in the gene via experimentation or by discovering if the gene co-localizes with known mutations. An international program has been established to sequence the complete *Arabidopsis* genome (except perhaps the repetitious telomeric and entromeric regions) by 2004. A major initiative to sequence the rice genome is being undertaken by the Japanese. These major plant genome sequencing projects sit alongside programs to sequence the human genome, and those of several bacteria, yeast and *C. elegans*. Thus, in a decade or so nearly all genes in some species will have been recognized and it will be possible to investigate the role of every gene in an organism. However, there will be the need to investigate the variation within key species, the allelic variation, because improvement depends upon utilizing specific alleles. These are not necessarily easy to construct in vitro from other alleles without knowledge of the subtle links between structure and function, especially for regulatory genes.

Specific genes can be isolated from banks of clones of plant mRNAs or chromosomal sequences if a means of recognizing a specific gene in the banks is available. The bank can be screened using a DNA fragment with the same or similar sequence or an antibody against the gene product. Single gene sequences can also be amplified from mixtures of mRNA or chromosomal sequences by the polymerase chain reaction (PCR) using short DNA sequence primers known to be specific for the desired gene. Often, however, the sequence of the desired gene is unknown, if a gene is mapped relative to nearby known molecular markers, the gene can be isolated by first isolating the chromosomal DNA fragments containing the marker and then finding the neighboring fragments by sequence homology between overlapping fragments and by "walking" along overlapping fragments until the desired gene is reached and identified. This route to genes is an important reason for developing the means of mapping genes on chromosomes in fine detail.

Genes are also being isolated in maize, Antirrhinum, petunia and *Arabidopsis* following their inactivation by the insertion of a "jumping gene" into them. Purification of the jumping gene fragment from its new location leads to co-purification of the inactivated gene.

Mapping of Genes Based on Conservation of Gene Order During Evolution

Recognition of a gene in one species can help recognition of the equivalent gene in another species. This can be through similarity in sequence where this is known, but plant breeders and geneticists wish to select parents and analyze genetic variation where the varying sequences are unknown. Now, from recent knowledge of genetic maps, it is possible to predict the location of genes in one species from their location in related species. This is because the sequence and linear order of the genes along chromosomal segments are conserved during evolution. Thus, once the related chromosomal segments are identified (this is relatively easily done using conserved molecular markers) genes known on a segment of one species can be expected to occur in a similar location in the related chromosome segments of other species. This discovery is of profound significance for plant breeding and genetics. The genetics of related species can be combined into a common pool of knowledge. The genetics, physiology and biochemistry knowledge gained about one species can be used to aid the genetics and breeding of its relatives. Therefore, the knowledge of genetics, physiology and biochemistry of rice can be transferred to maize, wheat, etc., and vice versa in all combinations. This discovery will serve to unite plant genetics and plant geneticists whether studying a model species or a crop.

It is clear that genes, gene products and/or segments of genes are conserved across the kingdoms and strongly within the dicots and monocots. Thus, information and genes isolated from model species such as *Arabidopsis* or

rice are extremely useful for isolating and mapping genes and discovering the function of genes in crop plants. The case for a role for research into model species to underpin crop improvement is now proven. This gives crop improvement research a very different profile and has substantial implications for research funding bodies.

Impact of "Gene Mapping" on "Agricultural" Science

We are already witnessing the impact from the incorporation of the new gene mapping into agricultural sciences. Plant breeding companies are now able to learn which chromosome segments they have selected in successful improvements over the past decades, which segments are associated with what traits, which segments are desirable, and which carry undesirable genes. It is possible to conclude where recombination sites are localized, what genetic linkages are readily broken, and which are very difficult to break. The introduction of known desirable segments into elite cultivars by backcrossing is now achieved much more rapidly and precisely using molecular diagnostics for the segment.

As breeding companies recognize the segments they wish to preserve or substitute it will be possible to define the desired genotypes by a collection of markers. The assembly of genotypes by molecular markers will be followed by computers. The whole process of breeding could be monitored by computers with little need for field trials, especially in the early generations of genotype assembly. This approach is dependent upon markers being measured rapidly, efficiently and cheaply. Automation of such procedures is being tackled by the large breeding companies, but it has yet to be proven as cost effective throughout large breeding programs. The DNA "fingerprinting" of elite germplasm by molecular markers will lead to companies being able to recognize their specially selected gene combinations in the breeding material of other companies. Legal suits have already emerged from companies marketing the germplasm of others.

A big impact of mapping will be as part of the large program of discovery of new genes. As the determinants of key traits, including those determined by many genes, become known at the molecular level, new variants will be recognized and followed as DNA molecules, not just as chromosome segments in breeding programs. When specific genes from any organism are redesigned for use in plants and inserted into crops by transformation techniques it is possible to create variation in plants at will and evaluate it. The impact of being able to improve any attribute by adding a limitless supply of genes from the genome projects of species from all the kingdoms and their variants is going to be enormous. New ways of achieving improved plants will be devised and these plants will serve new customers, markets and industries. The possibilities will be limited by our imagination and no longer by the gene pools within the species. Agricultural science will become much richer.

The issues created by the new gene technology are very substantial and rapidly evolving. The details vary considerably from crop to crop and society to society. Therefore, the future is difficult to predict with precision. One can be certain, however, for the foreseeable future the application of gene technology to crop improvement will be a topic for debate. Many views abound on the ethics of manipulating genes, but many of them are based on scientific ignorance of the processes involved and of the past history of crop improvement. This, however, does not make the views any less important. Some people insist that they wish to know if "foreign" genes created by a scientist have been inserted into their food. They, therefore, seek to have all food with additional genes labeled as such. This creates huge problems of how to create meaningful labels and how specific variants are kept separate in distribution systems, especially in commodity crops. Food safety is a very important issue and anything that could undermine food safety carries much risk in the minds of the public. Assessment of what is healthy and safe is very expensive and long-term. The issue is made more complex by specific food constituents having adverse effects on some individuals but not others.

Conditions on how genetically manipulated foods and non-food crops will be introduced in the U.S., Europe and other countries have been agreed upon. They are probably acceptable to most citizens, but not to all. There are significant groups of people in certain societies that are hostile to the technology. These groups, aided by sections of the media, may make it difficult for supermarkets and restaurants to sell genetically engineered products. At present, retailers are cautious about marketing such products, requiring evidence of public confidence before proceeding.

A major issue that will impede international trading of genetically engineered products is the lack of harmony between legislation and licenses governing the sale of such items in different countries. There is, therefore, an urgent need for international agreements on legislation and licensing.

Plant improvement has always provoked product displacement and changes in land use, producing economic and social consequences that benefit some societies and/or individuals and harm others. The new genetics will no doubt lead to new uses of plants as raw materials that will change farming patterns. Production of novel oils in temperate crops may substitute for imports of tropical crop oils, for example. Pharmaceutical proteins could be produced in plants instead of bacteria or animal cells and novel starches could be produced in potato instead of maize.

Plant improvement will become much more complex — technically and intellectually. It will, therefore, become more expensive. Plant breeding in western countries is likely to be increasingly dominated by the multinational companies. There are concerns that this may decrease choice for farmer and consumer.

One of the major concerns of the adoption of transgenic plants is their potential effects on the environment. If foreign genes are spread via pollen into populations of wild relatives and accumulate, the natural populations could be genetically contaminated forever. In this area there are ethical as well as scientific issues. It is not easy to predict gene flow from cultivated crops to wild relatives or the ecological consequences of such gene flow. The principles are not different from the present growing of highly selected agricultural variants in places where wild relatives also grow. However, the specific outcomes may differ on a case by case basis and are not easy to predict. Plant species have always evolved by changes in gene frequency and plant breeders and agriculture has inevitably influenced ecosystems dramatically. Will the continual and ever-increasing use of novel genes designed by researchers provide disruption at too great a rate? Or will rates of change be influenced little by changing the gene pools of our crop plants?

If any environmental damages were to result from the use of novel genes who will be responsible? Who will pay? These legal issues are likely to influence the extent to which improved transgenic crops are used enthusiastically.

The use of purified genes in crop improvement is a patentable activity. This is making a major change in academic institutions and industry. It will have a major effect on how such genes and improvements are used in commerce to benefit the consumer. If patents are to be valuable then they need to be filed before publication. This means that academic institutions and industries need to recognize the value of a gene long before full information on their practical potential is available and more than 10 years before a product is likely to be realized. This scenario creates problems for academic institutions and industry.

Academic institutions have to worry about when and what to publish, when and how to interact with industry, and on what terms. Industry is challenged to make decisions about future technology that is unproven or not reduced to practice. Industries argue for exclusive licenses that can cause great problems for the academic research scientist because it may be necessary to work with other companies in the future to progress research involving the licensed materials. Many scientists and academics are concerned that with patenting of the principal sources of added-value in the cultivars of the future they will compromise their public-sector status and become increasingly motivated by monetary rewards rather than scholarship. Many inventors and academic institutions wish to see their discoveries benefiting a large community. It will be important to assess whether industries holding exclusive licenses will make arrangements with breeders worldwide to move advantages to many societies or will be very restrictive. Many people are concerned about whether patents will deny developing countries the advantages of biotechnology or whether such societies will use them avidly and risk that no legal issues of any consequence will result.

If plants, in contrast to genes, are allowed to be patented this is likely to inhibit the flow of elite germplasm around the world. Free flow of germplasm has been very important in the history of plant improvement.

It is not possible for any individual or group to predict or control all these issues. Societies must monitor the issues and develop laws and case history to ensure that the production and exploitation of gene technology realizes its potential, globally, to help the health and wealth of all societies.

Concluding Remarks

The advances in gene technology and genetic mapping are driving plant and animal improvement into a new era. It is an era that will be very exciting and productive but will also have argument, economic change and some social chaos. This is probably inevitable given the way genes and gene combinations can be recognized, redesigned and exploited as never before in the history of plant breeding. The industry will change, as will the role of research in public sector institutions underpinning the industry. There is tremendous scope for the new genetics to influence agricultural research – it has already begun – but consumer acceptability, investment needs, and laws will have a greater influence than ever before. The way forward is complex. But hopefully the needs to provide food and materials in a sustainable way will always be at the top of the agenda.