1

Introduction to Plant Biotechnology

Richard B. Flavell

Ceres Inc., Malibu, CA, USA

INTRODUCTION

Mankind uses a wide range of plant species for food, feed, fibre, energy, chemicals and drugs. The exploitation of plants for the benefit of mankind (plant biotechnology) has been developed as part of evolution of societies throughout the world. Plant biotechnology, in its broadest sense, is therefore pervasive in every society and in some applications is highly developed, having progressed via a huge number of technical and intellectual advances.

Plant breeding was practised in primitive forms in the early civilisations. Indeed, settlements were made possible by the purposeful growing of specific selected plants, i.e., agriculture for food, feed and fibre. These selected plants, 'land races', formed the basis of agriculture for thousands of years. Most of the plants used directly to produce agricultural products today have been selected and bred specifically for the purpose from their wild relatives. While slow progress in crop improvement via selection occurred at many places in the world throughout history, it was the birth of the laws of heredity and the science of genetics around 1900 that turned plant breeding into a science-based progressive technology.

Today's most advanced plant breeding involves the improvement of plants by selection of the most favourable combinations of (all) the alleles in the species via knowledge-based procedures and in some cases via the addition of new genes designed in the laboratory to confer specific properties. It depends on the integration of genetics, mathematics with statistics, chemistry, pathology, cytology, molecular biology and computational biology, together with considerable automation, data handling and analytical systems. It is the molecular biology, automation and data handling that have changed so much over the past 10–20 years. These technical advances have been especially extraordinary and have set the stage for developments that offer opportunities unparalleled in the history of man's exploitation of plants. The two chapters in this introductory volume provide overviews of firstly, the theory and practice of plant breeding, secondly, many of the applications of plant biotechnology in the provision of food, feed, fibre, chemicals, drugs and energy as part of, and in addition to, plant breeding.

The diversity in the plant kingdom is immense. While a large number of species have been exploited directly to sustain societies, the proportion used to provide food for most people is very small. Tree species are perennial sources of food such as fruits, as well as fibre for construction and paper, while medicines and drugs are extracted from organs of a very few species. The diversity in the plant kingdom is a big challenge for plant breeding and biotechnology—so many species need research and development to be deployed on them.

PLANT BREEDING

There are many products that illustrate the spectacular successes of plant breeding. For example, maize and wheat are essentially man-made

Handbook of Plant Biotechnology
Edited by Paul Christou and Harry Klee

crops. Their ancestral wild forms are considerably inferior for food production from many points of view. Today's tomatoes and potatoes have been bred from small toxic wild forms (e.g., Frary *et al.*, 2000). The 'Green Revolution' in wheat and rice was possible because of the recognition and incorporation of novel dwarfing mutant alleles into cultivators in the 1950s and 1960s.

Plant breeders base their goals and activities on knowledge of the consumer and farmers to be served, the environments in which the plants need to succeed and the biological and genetic features of the species. Yield in the various environments in which a crop is grown is a near-universal breeding objective. To optimise yield, it has been necessary to select for tolerance/resistance to pests and diseases, to abiotic stresses such as drought, salinity, heavy metals, cold or heat and a host of other factors including flowering time, seed quality, taste and industrial utilities. The genetic features of a species determine whether true breeding homozygous cultivars or F_1 hybrids are required. If true breeding homozygous cultivars are required then the breeding strategy has to involve creation of new combinations of genes, selection of preferred phenotypes and then production of homozygous lines. The selection can be early or late in the process. Homozygosity can also be created early via chromosome doubling in haploids or late. For outbreeding species the cultivar can be an improved population of genotypes or single or double cross hybrids. The various breeding strategies that are followed are based on a statistical understanding of quantitative genetics. Plant breeding and plant performance are based on all the genes in a plant. Thus, they have a very complex basis. This complexity has been modelled by quantitative geneticists (Mackay, 2001), but now the complexity is being defined in physical and functional detail by 'genomics'. Genomics is the term given to the generation and study of large quantities of data comprising chromosomal DNA sequences, genes, the time and place of expression of the genes, gene-phenotype associations, genetic and physical maps and the details of genetic variation.

GENOMICS

Genomics has developed from technology developments in molecular biology especially in gene discovery and characterisation technologies whose development lies outside the plant kingdom. The process of gene purification via cloning came from bacterial genetics and the ability to splice novel DNA into replicating plasmids and to select and propagate the rare plasmid containing the desired gene. Efficient, high throughput sequencing of complete plant genomes, expressed sequence tags (ESTs) and full-length cDNAs, followed major innovations in automated sequencers and computing and ways of assembling and analysing the huge amounts of sequence data. This biotechnology research is now underpinned by a diverse array of industries supplying reagents and equipment. Entrepreneurs and venture capital investments have developed many such industries.

The extraordinary technical developments in plant molecular genetics, over the past 20 years, culminated at the end of 2000 in determination of almost all the complete nucleotide sequence of the nuclear genome of the plant *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative, 2000). So, for the first time in the history of plant biology it was possible to start analysing the genetic content of a species, predict all the genes to discover the functions of each of them and start the long haul of discovering how the set of 30 000 to 70 000 genes provides the developmental complexity of the organism and its differences from other species. The 'complete' sequences of two rice genomes were reported in 2002 (Goff *et al.*, 2002; Yu *et al.,* 2002) and several more complete plant genome sequences will emerge in the next few years (Chandler and Brendel, 2002; VanderBosch and Stacey, 2003). New information indicates that the genomes contain not only well-known kinds of genes but many sources of transcribed RNA that are involved in the post-transcriptional control of mRNA levels and thereby in regulating plant development (Kidner and Martienssen, 2003).

In the model species *Arabidopsis* and rice, programmes are very advanced to obtain mutations 'in every gene' for gaining knowledge about their function. In addition, in species favourable for transformation, genetic variants are being created by the design, construction and insertion of novel copies of each gene, expressed differently or encoding a novel protein variant. This also provides *in vivo* knowledge on the role of each gene and protein.

It has been known for over three decades that chromosomes contain many families of highly repeated sequences. Some are organised in

large arrays, especially around centromeres and telomeres. The molecular biology of centromere and telomere functions dependent upon these arrays is now becoming known. Other families of repeats have evolved from transposable elements that spread through chromosomes of populations either via movement of the DNA element itself or via transcription intermediates. The movement of such elements leads to considerable structural diversity of chromosomes during evolution and many mutations affecting gene function. These mutations lead to silencing, down- or up-regulation of gene expression.

The sequencing and mapping of these elements and genes along plant chromosomes, within and between species, is revealing the molecular basis of variation that has evolved during species divergence, that breeders have been working with (Tenaillon *et al.*, 2001) and the haplotypes that have been selected to create successful crops (Ching *et al.*, 2002). These comparisons are opening up the powerful subject of 'comparative genomics'. The most noteworthy discovery of comparative genomics for plant biotechnology to date is the finding that during the divergence of grasses, the order of genes along large chromosomal segments has been conserved (Gale and Devos, 1998). This conservation has occurred despite the accumulation and deletion of large quantities of repeated sequences and transposable elements between the genes. The practical significance of this conservation is that from the complete sequence of the rice genome, the position of related genes can be predicted in the chromosomes of maize, wheat, barley, millet, sugarcane and all other grasses, once the related chromosome segments have been aligned. This has greatly helped the production of genetic maps of all these species and hence, breeding strategies.

Knowledge of gene sequences allows protein product sequences to be predicted. Comparative protein analysis across the kingdom is helping to predict the functions of closely related proteins from any other organism once the function of a protein is discovered in one organism.

While the function of individual proteins can be assessed from *in vitro* studies or from comparative studies across the kingdom, the *in vivo* function in a plant needs to be assessed via genetics, assuming that there is genetic variation for the gene or variations can be created. Genetic analysis of a few crop plants and some tree species is becoming well developed. Such analyses have provided the chromosomal location of quantitative trait loci (QTLs) that contribute to key characteristics of plant size, shape, yield, chemical composition, disease resistance and tolerance to biotic stresses, etc. The analyses were painstaking till the last 10 years or so, because they were dependent on other phenotypically expressed genetic markers which are present on the same chromosome to assay linkage. With the deployment of restriction enzymes, PCR and DNA sequencing, variants can now be discovered rapidly for almost any chromosome sequence and hence detailed maps can now be readily constructed for any species. All genomes appear to contain a huge number of arrays of repeating nucleotides due to frequent errors in replication,deletion and expansion via unequal crossing over (Tautz, 1989). These 'minisatellites' provide prolific and easy-to-measure polymorphisms and have revolutionised the genetic mapping of genomes. Complete genome sequencing and genetic mapping, i.e. expanding knowledge about chromosome segments that contain genetic variants affecting vital characters leads us into an era of *in vivo* gene function analysis. High-throughput analyses of single nucleotide polymorphisms are also being brought on line to allow rapid analyses of haplotypes and mutations linked to QTLs (Rafalski, 2002). QA1

The combination of genomics, genetics and plant breeding is leading to an understanding of the genes (alleles) involved in specifying valuable attributes in crop species. The availability of molecular markers covering the entire genome of a crop is driving molecular marker-assisted breeding to make plant breeding more efficient. The principles and practice of efficient plant breeding, incorporating high-throughput genomics are now established. There is now a need to establish these procedures in more corporate and public sector breeding programmes and for more crops. Reductions in cost coming from new innovation and more efficient deployment should greatly help more widespread use of marker-assisted breeding.

TISSUE CULTURE

The culture of plant organs and cells is important in many areas of plant biotechnology. In plant

(cereal) breeding, haploid cells in anthers are routinely cultured for chromosome doubling, and then regenerated to produce doubled haploid homozygous plants (Forster and Thomas, 2003). In the process of creating transgenic plants, transformed cells are proliferated in culture before being induced to differentiate again into roots, shoots and whole plants.

The biochemistry of cells in culture is significantly different from that in differentiated cells and this has been exploited for looking for metabolites that are formed rarely in whole plants. Certain plant species and tissues can be induced to form single cell, or a few cell, cultures very rapidly and these are used in fermenters to facilitate the controlled, contained production of valuable molecules in the cells. While this technology has held much promise, it is being displaced in emphasis as transgenic plants are providing new ways of modifying pathways and boosting yields of desired metabolites. In some cases cells can be cultured in fermenters and then be provoked to form somatic embryos by hormone manipulation. The massive production of such embryos facilitates the clonal production of large numbers of identical seedlings (Guha and Maheswari, 1964).

Some species, including some agriculturally important species can be propagated asexually. Clonal propagation offers a way of multiplying complex genotypes (e.g., hybrids) and this has a major place in many industries, e.g., forestry, potato production, horticulture (e.g., strawberries) and floriculture.

TRANSGENIC PLANTS

In 1982, the first fertile transgenic plants were created and in the following decade or so, over 100 plant species were transformed worldwide. Today, two main procedures are used for introducing novel genes into plants. The most widely used procedure is that evolved in the soil-borne bacterium, *Agrobacterium tumefaciens*, to transfer a specific segment of one of its plasmids into plant cells. The other is biolistics—the shooting of DNA into cells. Following both methods, transformed cells are selected in culture and regenerated into whole plants. Both procedures result in one, or frequently more than one, copy of the novel gene being integrated into the plant chromosome. Although frequently the genes behave as regular Mendelian genes, they can be affected by the environment in which they reside and also become epigenetically modified with methyl groups. This can alter their activity. Thus, it is important that many transformants are made and sifted to find individuals in which the gene behaves stably as required in plant breeding and commerce.

The commercial plantings of transgenic crops today include soybean, corn, canola, cotton and papaya. Acreages have increased each year since their introduction in 1996. They have been greatest in the USA where they were first produced but are now being adopted in all continents including developing countries, but not in the EU. The soybeans have been engineered to be resistant to the herbicide glyphosate, easing weed control, while the cotton and corn cultivars carry a novel gene that encodes a protein that is toxic to specific insect larvae, thus protecting the crops from damage by these pests. This protection provides considerable financial savings due to the reduced use of insecticides, which are also a health hazard. The introduction of these transgenic crop cultivars on this scale is the most rapid adoption of cultivars ever and is witness to the extraordinary improvements that can be made using carefully selected transgenes.

Before transgenic products can be released into commerce they need to be deregulated by government agencies or their allowed use defined. The process of gaining the information required by the government agencies can be very time consuming and relatively expensive. This makes the cost of developing transgenic products for commerce greater than for non-transgenics. Thus, it is currently recognised that transgenic agricultural products need to be substantially better than existing cultivars before they become financially worthwhile to the producer.

Governments have drawn up the guidelines for allowing the release of transgenic plants (genetically modified organisms, GMOs) into the environment and into food/feed chains, based upon many criteria. The reasons for treating them as distinct from non-GMO products are based on the assumptions that it is possible to create products by foreign gene insertion that are toxic to people or the environment, and that gene flow from transgenic crops to wild relatives and non-transgenic crops via pollination

may have consequences that are unwanted in the short- or long-term. Different societies and governments differ in their views on the acceptability of GMOs in food and local environments. This is currently causing considerable confusion for the international food/feed producers, food processing industries and consumers. For the past few years there has been a moratorium on the growing of such products in the EU. While the USA, Argentina, Australia, Canada and China, for example, have embraced specific transgenic plants, the EU and other countries are being extremely cautious. The EU has decreed that all products from transgenic plants are labelled as such. The consequential effects of the EU stance on world trade issues are large. The EU decision to label the products of GMOs is leading to the requirement to identify the existence of transgenes in all products and trace the transgenes from seed to product. The required assay systems that need to be deployed at many steps in the chain are providing new opportunities for biotechnology companies that provide such detection kits and services.

MACROMOLECULES AS RAW MATERIALS AND HIGH VALUE PRODUCTS

The principal molecules in plant organs harvested for food, animal feed or industrial uses are carbohydrates, proteins and oils but there are also a host of other molecules in lower concentration. Plant biotechnology research has focused on these valuable molecules not only for their value in the food and feed industries but also for the huge range of other industries that incorporate these molecules into their products. Sometimes the crops containing these molecules are produced specifically for the industry in question and sometimes they are bought as commodity byproducts of the agricultural/food/feed processing industries.

Starches in grains are composed of amylose and amylopectin chains of varying lengths and branches. The properties of the starch are determined, in part, by the chain lengths and degrees of amylose branching. Manipulation of the levels and kinds of the enzymes involved in starch biosynthesis create starches with different properties suitable for different industrial purposes. This has been a very active field of R&D, especially in corn and potatoes, the major sources of industrial starch.

The oils from oilseeds are used widely in industry and are also a major source of oils for food products and cooking. The length of the carbon chain and the extent of unsaturation in the oil are key parameters in determining the properties of the oils. Reductions in C18 polyunsaturates in consumed products are desirable for health reasons. By adding different oil biosynthesis enzymes, or reducing the levels of existing ones, major changes in the fatty acid profile can be made. For example, by adding a thioesterase from the bay tree to canola, it was possible to increase the concentration of lauric acid from 0 to 40% of the total fatty acids in seeds. Manipulation of fatty acid profiles has been a significant R&D activity.

DHA (docosahexenoic acid) and EPA (eicosapentenoic acid) are fatty acids that enhance human health. They are present in fish and microalgae but absent in plants. The transfer of the genes from fish and microalgae into an oilseed plant to make an alternative, cheap source of these fatty acids for human consumption is an attractive proposition. Many chemicals derived from petroleum are used in adhesives, paints, detergents, lubricants, nylon, cosmetics, etc. The opportunities of producing specific desirable forms in plants via agriculture increases with the ability to design transgenics.

It may become economically worthwhile making plastics such as PHA (polyhydroxyalkanoate) and PLAs (polylactides) in plants by adding the relevant microbial genes that direct synthesis of these polymers. Many products are now nutritionally enhanced by the addition of vitamins and antioxidants. By adding or manipulating genes encoding the relevant biosynthesis pathways it should be possible to create plants that are nutritionally enhanced endogenously and so be considered as "functional' foods or 'nutraceuticals'. The most well-known example today is perhaps 'golden rice', that contains genes encoding three enzymes that lead to accumulation of provitamin A, a molecule that can relieve blindness in children whose major diet is rice based.

Plant proteins in cereal seeds that are used as major sources of animal feed are often low in essential amino acids such as lysine, methionine and cysteine. Biotechnological manipulations have led to increase in these essential amino acids. The pharmaceutical industry also focuses on producing antibodies and antigens that are

proteins (Chargelegue *et al.*, 2001; Cramer *et al.*, 1998). These may be produced cost effectively in animal or microbial cells. However, where the amounts required are extremely high, plants may be much more economically suitable production vehicles (Ma *et al.,* 1995). There is therefore now a lot of activity in producing large quantities of therapeutic proteins in transgenic plants. Associated with these initiatives is discussion about the desirability of producing pharmacologically active molecules in crop plants also used for food/feed. The probability that mistakes will happen in keeping the seed lots separate is too high, some argue.

Transgenic plants may be good production vehicles for industrial enzymes and other proteins not known to be pharmacologically active (Hood *et al.*, 1999). Thus, several companies are inserting genes for enzymes into plants, growing and harvesting the plants and purifying the required enzyme. This is a growth opportunity in plant biotechnology. The promise of massive production of specific proteins in plants has given rise to a plethora of ingenious promoters, expression vectors and ways of exploring amplification of gene templates via viral vectors in order to boost production of the desired proteins.

METABOLITES

Many important ingredients in food and feed are metabolites. Plant biotechnology is increasingly focusing on defining the genes and enzymes responsible for the production of these metabolites, and in particular, how production is regulated. Many flavonoids, terpenes, alkaloids and vitamins are made in only specialised cells or in response to biotic or abiotic stresses and so understanding and manipulating the complex pathways of 'secondary' metabolism is an exciting opportunity. Already many of the genes specifying pathways have been characterised and specific plants making larger amounts of nutritionally important metabolites such as lycopene and vitamin E have been created.

Many of these secondary metabolites also have effects as antagonists or agonists on other cellular molecules. They can thus be a source of fungicides, insecticides or herbicides in agriculture if they kill or inhibit growth of the target organisms but not the crop. Thus, the agrochemical industries have sought extensively to find such potentially valuable agrochemicals from plants. Where the secondary metabolites agonise or antagonise targets in man, they can be a source of drugs, e.g., aspirin, taxol, codeine and vincristine. Thus many pharmaceutical and small biotechnology companies have undertaken large screening programmes involving tens of thousand of species looking for plant metabolites that interact with known target molecules or processes in animal cells or in bacteria or fungi that are harmful to humans. The discovery of a valuable new drug is rare. This is because of many reasons. One of them is that many of the plant tissues used in the screens have not been optimised for the levels of secondary metabolites likely to serve as drugs. Here is an opportunity for plant biotechnology to serve the pharmaceutical industry better. Also, even when a potential drug has been characterised, producing enough of it from natural sources may be problematic or prohibitively expensive. Here genetic or environmental manipulation of the metabolic pathway to produce much more of the desired product can be extremely valuable.

CLOSING PERSPECTIVE

Plant biotechnology, like all aspects of the life sciences, has changed very rapidly over the past 20 years. The new ways of analysing genes, molecules and processes constitute new platforms, together with established plant breeding systems, for discovery, product design and making novel products. These products will be improved sources of food, feed, fibre, energy, chemicals or drugs. The knowledge of the ways plants develop their architecture, survive abiotic and biotic stresses, utilise fertiliser, water and CO_2, and make complex secondary metabolites and macromolecules will increase much more rapidly than over previous decades. Most importantly the genes and alleles that programme these attributes will become known rapidly from the combination of genomics, genetics and plant breeding for the species intensively studied. The poorly understood species will remain a challenge, but comparative molecular genetics will enable them to be investigated much more rapidly than before.

There is a new vast array of genes and promoters available from genome sequencing in all kingdoms.

This toolbox provides a limitless number of options to explore the improvement of plants for commercial and social benefits via the addition of transgenes.

From standpoint of the research one can expect that the properties of plants will change considerably over the coming decades, given the new opportunities to reprogramme them genetically. However, which products will emerge depends on which forms are economically and environmentally sound and socially acceptable. The accompanying chapters summarise much of the technologies, progress and future perspectives for plant biotechnology. They illustrate that the principles of the new plant biotechnology are now well established, but of course, the technologies will continue to change due to innovations and new examples will come and go as options are evaluated scientifically, economically and socially.

REFERENCES

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

Chandler VL, Brendel V (2002). The maize genome sequencing project. *Plant Physiol* **130**: 1594–7.

Chargelegue D, Obregon P, Drake PMW (2001). Transgenic plants for vaccine production: expectations and limitations. *Trends Plant Sci* **6**: 495–6.

Ching A, Caldwell KS, Jung J, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* **3**: 19–32.

Cramer CL, Boothe JG, Oishi KK (1998). Transgenic plants for therapeutic proteins: linking upstream and downstream strategies. In Hammond J, McGarvery P, Yusibov V (eds), *Current Topics in Microbiology and Immunology. Plant Biotechnology. New Products and Applications.* Springer: Berlin; pp 95–118.

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knapp E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000). *fw2.2:* a quantitative trait locus to the evolution of tomato fruit size. *Science* **298**: 85–8.

Gale MD, Devos KM (1998). Comparative genetics in the grasses. *Proc Natl Acad Sci USA* **95**: 1971–4.

Goff SA *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92–100. **QA2**

Guha S, Maheswari SC (1964). *In vitro* production of embryos from anthers of datura. *Nature* **204**: 497.

Hood EE, Kusnadi A, Nikolov Z, Howard J (1999). Molecular farming of industrial proteins from transgenic maize. In Shahidi F, Kolodziejczyk P, Whitaker JR, Munguia AL, Fuller G (eds), *Chemicals via Higher Plant Bioengineering.* Kluwer/Plenum: New York; pp 127–48.

Kidner CA, Martienssen RA (2003). Macro effects of micro RNAs in plants. *Trends Genetics* **19**: 13–6.

Ma J, Hiatt A, Hein M, Vine N, Wang F, Stablia P, van Dolleweerd C, Mostov K, Lehner T (1995). Generation and assembly of secretory antibodies in plants. *Science* **268**: 716–9.

Rafalski AJ (2002). Applications of single nucleotide polymorphism in crop genetics. *Curr Opin Plant Biol* **5**: 94–100.

Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA makers. *Nucleic Acids Res* **17**: 6463–71.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA* **98**: 9161–6.

VanderBosch KA, Stacey G (2003). Summaries of legume genomics projects from around the globe. Community resources for crops and models. *Plant Physiol.* **131**: 840–65.

Yu J *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**: 79–92. **QA3**

QUERIES TO THE AUTHOR

QA1. Please check whether the meaning of this sentence is retained after editing.
QA2. Please provide names of all authors for this reference. (Goff SA *et al.* (2002))
QA3. Please provide names of all authors for this reference. (Yu J *et al.* (2002))