G

GENETIC MODIFICATION

Contents

Gene Cloning, General Principles

R M Twyman, University of York, York, UK

Introduction

The analysis of plant genes requires a method for cloning, i.e., the preparation of millions of copies of the same DNA sequence so that enough DNA is present for laboratory scale experiments. Cloning methods can be divided into two broad categories:

- Cell based cloning, in which the plant gene of interest is joined to a replicon (such as a plasmid or bacteriophage vector) and amplified by repeated replication cycles inside a particular host cell (usually the bacterium *Escherichia coli*).
- *In vitro* cloning, by the polymerase chain reaction (PCR), where DNA amplification is achieved using purified enzymes.

The two techniques are compared in **Figure 1**.

These routine procedures can be applied to any DNA sequence of interest. Therefore, the bottleneck in cloning is not the amplification process itself, but the isolation of the target gene from its source. Under most circumstances, the source is very complex, and the desired sequence is diluted in a background of many millions of nontarget sequences. There are two typical sources of DNA for cloning:

- Genomic DNA. Whole genomic DNA represents the entire plant genome (usually the nuclear genome – plastid genomes are considered separately). Genomic DNA isolated from any cell type or organ of an individual plant is generally the same. Each gene is equally represented. Regulatory sequences and introns are present. The gene density in some species is very low, and in such cases much of the genome may be taken up by intergenic spacer regions, transposable elements, and tandem repeat sequences. In corn (*Zea mays*; maize), for example, only 10% of the genome is represented by genes.
- Complementary DNA (cDNA). Complementary DNA is reverse transcribed from mRNA. It does not represent the entire genome, only the genes expressed in the cell type or organ from which the mRNA was isolated. This means that cDNA prepared from different tissues and organs contains different sequences, and different genes are not represented equally – strongly expressed genes will produce more transcripts and give rise to more cDNA copies. Regulatory elements and introns are not present. Splice variants are represented by different cDNAs.

Both cell based cloning and the PCR can be used to isolate specific genes from complex sources. PCR based approaches can be applied directly to the source DNA, while in the case of cell based cloning, the typical approach is to create a clone library. For genomic DNA, this first involves breaking the source DNA into manageable fragments. The fragmented genomic DNA or whole cDNA is then cloned nonselectively into vectors, with the aim of generating a collection of clones representing the entire source population of DNA molecules. While each plant gene

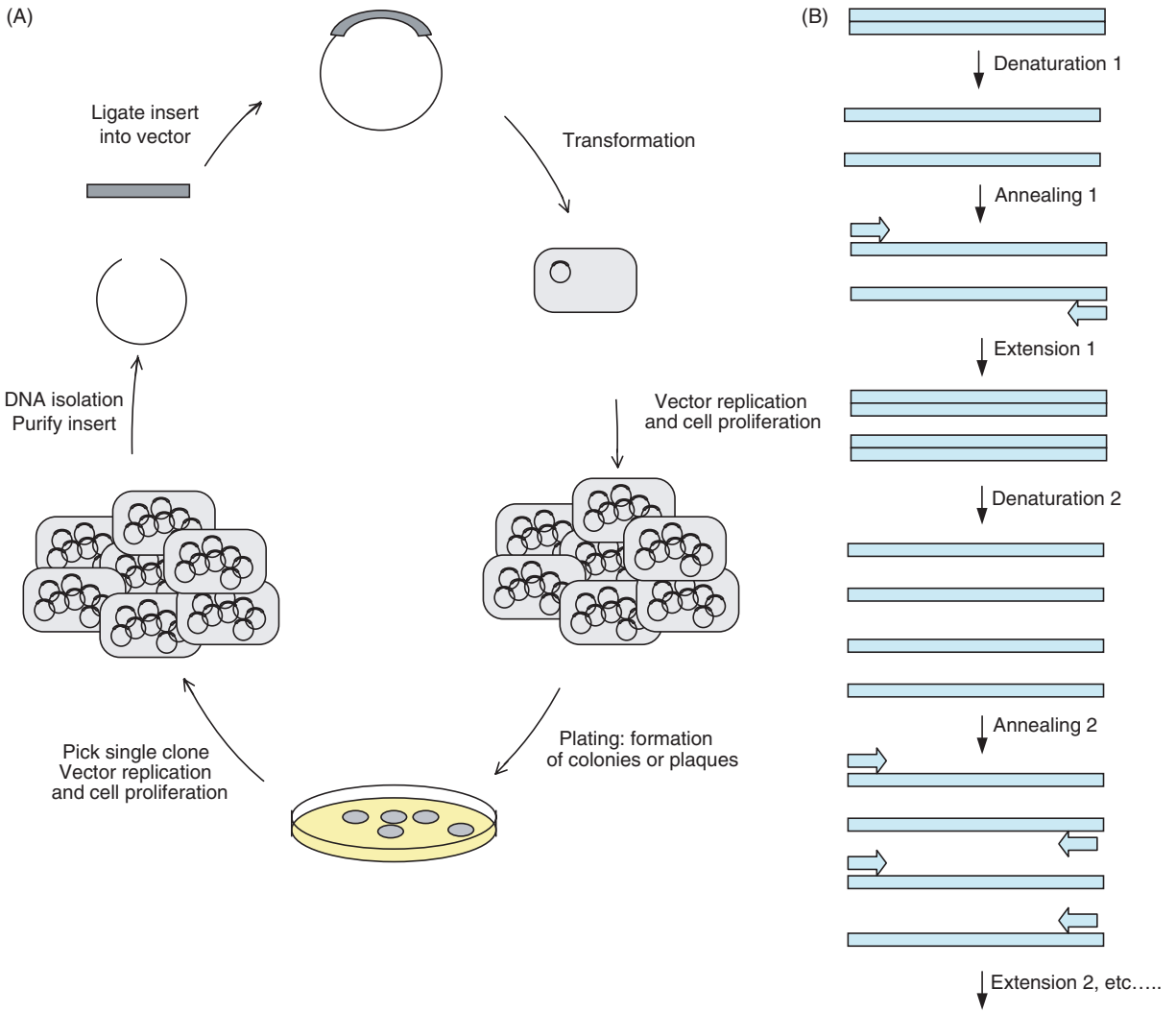

Figure 1 Principles of (A) cell based cloning and (B) the polymerase chain reaction.

is represented equally in genomic DNA, the random sampling involved in nonselective cloning means that genomic libraries are not equally representative of all genes. The DNA must be broken up randomly (to ensure no systematic exclusion of particular sequences), and enough clones must be generated to cover the genome several times (to avoid exclusion due to sampling errors). For example, to ensure a 95% probability of including every gene, there must be approximately fivefold coverage of the genome. The precise number of clones required depends on the genome size and the average insert size, which depends on the type of cloning vector used (**Table 1**).

The clone library must then be screened to identify the gene of interest. For genomic and cDNA libraries, a standard approach is to use hybridization or PCR assays to identify clones containing particular genes. Some knowledge of the target gene's sequence is generally required, as is the case for direct PCR amplification from the source DNA. In contrast, since cDNAs are discrete functional sequences, they can be expressed by including appropriate regulatory elements in the cloning vector. Expression libraries can then be produced, and these can be screened by exploiting properties of the gene product, rather than the gene itself.

Sequence-Dependent Cloning Strategies

Many gene isolation strategies are based on the identification of specific sequences by hybridization or PCR. In each case, the limitation is that there must already be some knowledge of the desired sequence before the screening process begins. Thus, sequence-dependent screening is like the chicken and the egg: without a clone, you cannot produce a probe (or primer set), and without a probe (or primer set) you cannot isolate a clone.

Table 1 Different types of vectors used in the cloning of plant genes

Type of vector	*Main uses*	*Maximum insert size*
E. coli plasmid	Subcloning, cDNA cloning, direct gene transfer to plants	10–15 kb
Binary vector	*Agrobacterium* mediated gene transfer to plants	20–40 kb
λ insertion	cDNA cloning	7–8 kb
M13, phagemids	Sequencing	5 kb
λ replacement	cDNA cloning, genomic cloning	18–20 kb
Cosmid, fosmid	Genomic cloning	40 kb
P1	Genomic cloning	125 kb
BAC	Genomic cloning, genome projects	300 kb
PAC	Genomic cloning, genome projects	300 kb
BIBAC/TAC	Transfer of large inserts to plants	300 kb
YAC	Genomic cloning, genome projects	500–1500 kb

Abbreviations: BAC, bacterial artificial chromosome; PAC, P1 artificial chromosome; BIBAC, binary BAC; TAC, transformation-competent artificial chromosome; YAC, yeast artificial chromosome.

Fortunately, there are several ways to generate probes/primers without knowing the full and precise sequence of the target gene:

- In some cases, it may be possible to obtain a partial sequence to use as a probe or for primer design. For example, expressed sequence tags (ESTs) generated by random sequencing of cDNA libraries can be used to isolate homologous full length cDNAs or genomic clones.
- Gene or cDNA sequences isolated from one species can often be used as probes to identify homologous genes in other species. Depending on the evolutionary conservation of the target gene, it may even be possible to use sequences from bacteria, yeast, or animals. For example, plant homeobox genes have been isolated using animal sequences as the basis for probe/primer design.
- Gene or cDNA sequences can be used to isolate homologous genes (paralogs) or partly homologous genes (domain sharing genes) from the same species.
- Sequences corresponding to a highly conserved domain can be used to isolate entire gene families. A good example in this case is the use of the first *Arabidopsis* MADS box genes to isolate related genes in *Arabidopsis* and other species.
- Probes or primer pairs can be designed and synthesized *in vitro* based on known amino acid sequences (**Figure 2**).

In the first strategy listed above, the partial sequence is identical to part of the target gene, and a suitable screening strategy would be to use high stringency hybridization or highly specific PCR primers. In all the others, however, there is likely to be a degree of disparity between the probe and its target(s). In such cases, hybridization assays are carried out with reduced stringency, which means a degree of mismatch can be tolerated between the probe and its complementary target. Similarly, PCR primers can be designed to match the most highly conserved regions within a family of sequences, and/or a degree of degeneracy can be built in to primer design. Degeneracy is always required where probes or primers are designed on the basis of amino acid sequences, since more than one codon is used to specify most amino acids. Protein sequences that contain tryptophan and methionine are useful for probe or primer design because these amino acids are each specified by a single codon, thus reducing the total amount of degeneracy required (**Figure 2**).

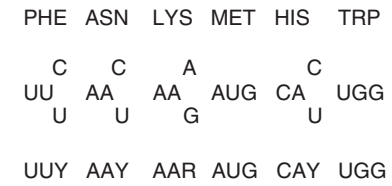

Figure 2 Designing a probe/primer based on a known amino acid sequence. The amino acid sequence shown on the top line can be encoded by any of 16 different nucleotide sequences as represented by the middle line, and is therefore said to show 16-fold degeneracy. The bottom line shows the nucleotide sequence with the standard degenerate representations (Y is any pyrimidine, R is any purine). Although primer combinations have been used successfully with up to 128-fold degeneracy, the total amount should be limited as much as possible. This is helped by choosing sequences containing the nondegenerate residues methionine and tryptophan.

Map Based Cloning Strategies

Map based cloning is another term for positional cloning, a strategy developed for the isolation of human disease genes in cases where no sequence or biochemical information is available. The strategy uses linkage analysis to define a candidate region of the genome flanked by the two closest genetic markers. A bridge of physical clones is then built across the interval using the genetic markers as a starting point (**Figure 3**). This technique, known as chromosome

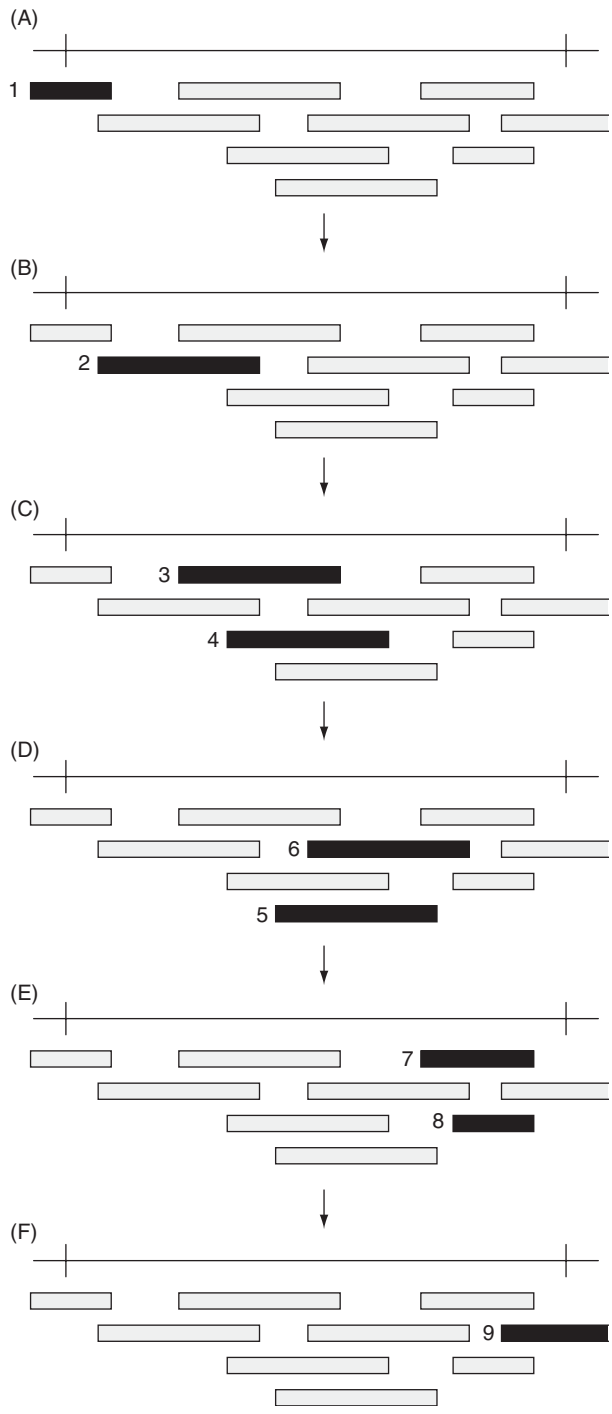

Figure 3 The principle of map-based (positional) cloning. The thin line represents a candidate interval of the genome defined by two genetic markers. (A) If the left flanking marker is used as a probe, genomic clone 1 is identified. (B) If clone 1 is used as a probe, overlapping clone 2 is identified. (C) Clone 2 identifies clones 3 and 4, as well as clone 1, which can be ignored. (D) Similarly, clone 4 identifies clones 5 and 6. (E) Clone 6 identifies clones 7 and 8, but note that clone 5 would identify clone 7 only, because it does not overlap with clone 8. (F) Clone 8 identifies clone 9, which also hybridizes to the right flanking marker. The genetic interval is now spanned by a bridge of physical clones.

walking, is laborious if large numbers of clones are involved. However, in plants it is often possible to identify markers that lie within a few hundred kilobases of the target gene, and with the availability of large insert vectors such as YACs and BACs (**Table 1**) the chromosome walks can be very short indeed. It is sometimes possible to identify a single clone spanning the candidate interval, whereupon chromosome walking becomes chromosome landing.

The isolation of the rice (*Oryza sativa*) *Hd1* gene provides a useful example of map-based cloning. Genetic mapping initially established a candidate region of 100 kb containing a quantitative trait locus (QTL) affecting sensitivity to day length. Physical mapping, using the flanking genetic markers as probes, identified a single BAC clone spanning the entire candidate interval. This BAC clone was used to establish a very fine genetic map based on CAPS markers (cleaved amplified polymorphic sequences). Further mapping then refined the interval to 12 kb, which contained two genes. One of these genes, *Hd1*, happened to be a rice homolog of the *Arabidopsis* gene *CONSTANS*, which regulates flowering time in response to day length. Prior to this map based cloning strategy, investigators had absolutely no idea of the sequence of the gene or its biochemical function – only its approximate location and its physiological effect.

Shortcuts to Map-Based Cloning

The genetic mapping of plant genes often requires laborious crosses between lines with alternative (mutant or wild-type) alleles at a given locus. Positional cloning on this basis is a labor-intensive and, sometimes, disappointing venture, which involves systematic library screening to identify contigs of overlapping clones spanning a given candidate interval. A useful short cut is to generate mutants by insertional mutagenesis, for this leaves a DNA tag within the interrupted gene. Several cloning strategies have been devised to isolate such genes, including the following:

- Genomic libraries can be produced from the tagged plant line and then screened with a probe (or PCR primers) corresponding to the insert. Genomic clones thus isolated generally contain flanking DNA sequences that can be used in wild-type libraries to isolate corresponding intact genes or cDNAs. It is important to make sure the insertion sequence is unique in the genome, or irrelevant flanking sequences could be obtained.
- Plasmid rescue. In this technique, the insertion element contains the origin of replication and

antibiotic resistance marker of a bacterial plasmid. Whole genomic DNA from the insertion line is simply digested with a restriction enzyme that does not cut within the insert, diluted, and circularized with DNA ligase. If this mixture of genomic DNA circles is used to transform *E. coli*, only bacteria containing the insert and its immediate flanking sequences will survive under antibiotic selection (**Figure 4**).

- Inverse PCR. In this technique, genomic DNA from the insertion line is digested with a restriction enzyme that does not cut within the insert, diluted, circularized with DNA ligase, and then amplified using PCR primers that anneal to the insert but face outward. This amplifies not the insert but the flanking sequences. These can be used as probes to isolate intact clones (**Figure 4**).
- Other PCR methods. A number of more sophisticated PCR approaches are available that do not require circularization of the genomic DNA prior to amplification. These include transposon display and amplification of insertion mutagenized sites (AIMS), both of which involve the digestion of genomic DNA with restriction enzymes, the addition of adaptor oligonucleotides, and several rounds of nested PCR. In transposon display, amplification is exponential and requires a transposon specific primer and an adaptor specific primer. In AIMS, only the transposon specific primer is used and amplification is linear. Another nested PCR technique is thermal asymmetric interlaced (TAIL) PCR, which involves the use of three nested transposon specific primers and a small arbitrary primer. As an alternative to

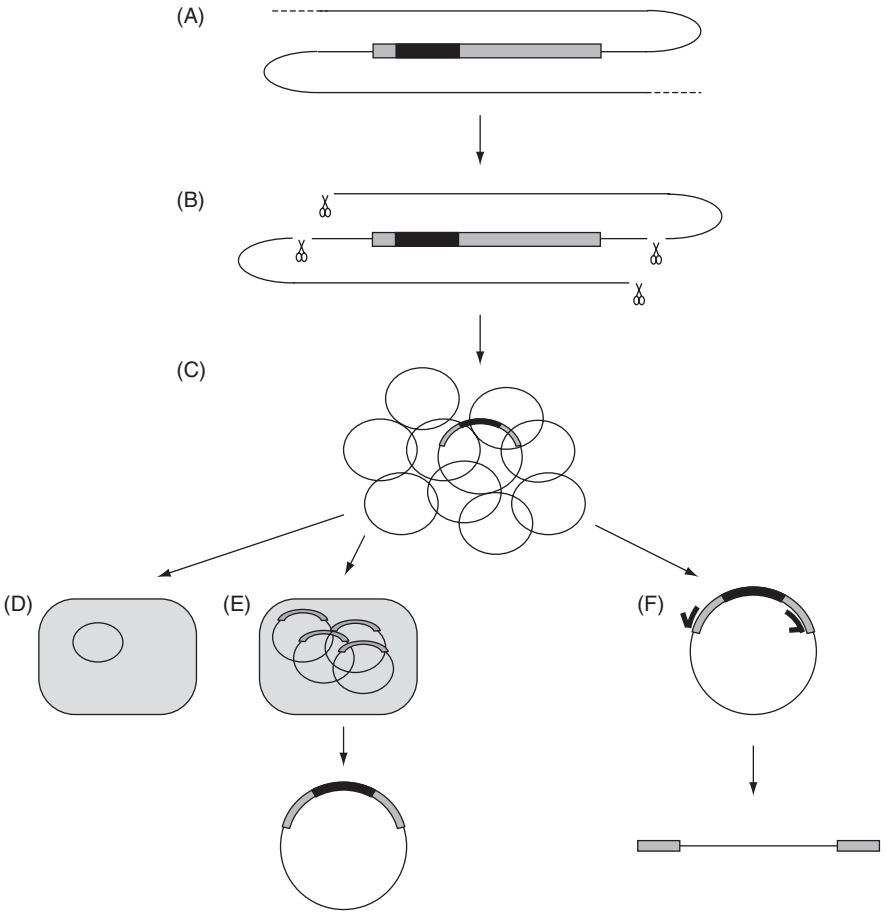

Figure 4 Cloning the flanking sequences adjacent to a transposon or T-DNA insertion element. (A) The element integrates into genomic DNA. (B) The genomic DNA is isolated and cleaved with restriction enzymes that do not cut within the element so that flanking sequences remain associated with it. The genomic DNA can then be used to construct a clone library, and the interrupted gene can be isolated by hybridization with a probe corresponding to the insertion element (not shown). (C) If the insert contains a plasmid origin of replication and selectable marker gene (black bar), recircularization of the genomic fragments with DNA ligase followed by random transformation of bacteria will produce (D) some bacterial cells containing random genomic fragments and (E) some carrying the insertion element and flanking sequences as a functional plasmid. Only bacteria containing the rescued plasmid will proliferate under selection, allowing the flanking sequences to be recovered. (F) Alternatively, the random circles of DNA can be amplified by PCR using primers that anneal to the insertion element but face outward, producing the flanking sequences as an amplification product.

genomic DNA, insertion events can be detected at the mRNA level by rapid amplification of cDNA ends (RACE).

Expression Cloning and Functional Cloning

Any cloning strategy that relies on the expression of the target gene is an expression cloning strategy. Included within this category is functional cloning, which refers to cloning strategies in which the biochemical function of the gene is exploited to identify it in a background of nontarget sequences. The most common expression cloning strategy is the immunological screening of cDNA libraries. This involves the use of labeled antibodies (or more commonly, antibody sandwich reactions) to isolate expression clones based on the structure of the corresponding protein. An antibody must be available, which generally means that the corresponding protein must already have been purified to some extent. Sometimes this approach is used because the alternative technique of degenerate probe design has failed, and it is often favored where an antibody is already available. In any case, a further requirement is that the protein folds properly when it is expressed, so that the epitope recognized by the antibody probe is retained.

For certain classes of protein, alternative probes may be used. For example, DNA-binding proteins can be isolated using labeled DNA probes corresponding to the protein's recognition site, a technique called southwestern screening due to its similarities to Southern blot and western blot methods. It is necessary to ensure that nonspecific DNA binding is blocked since many proteins bind to DNA, but have no particular sequence specificity. Similarly, northwestern screening involves the use of RNA probes to identify RNA binding proteins. Occasionally, protein ligands have been used to isolate interacting proteins from expression libraries. However, it is now much more common practise to use high-throughput screening techniques to identify cDNA clones corresponding to interacting proteins (see below).

Functional cloning strategies have been widely used in animals, but much more rarely in plants. Some functional cloning approaches include:

- Screening for the biochemical activity of a protein, e.g., enzyme activity, effect on plant growth.
- Functional complementation of mutant phenotypes in transformed plant cells or transgenic plants (**Figure 5A**). The development of high-capacity plant transformation vectors (binary bacterial artificial chromosome (BIBAC) and transformation-competent artificial chromosome

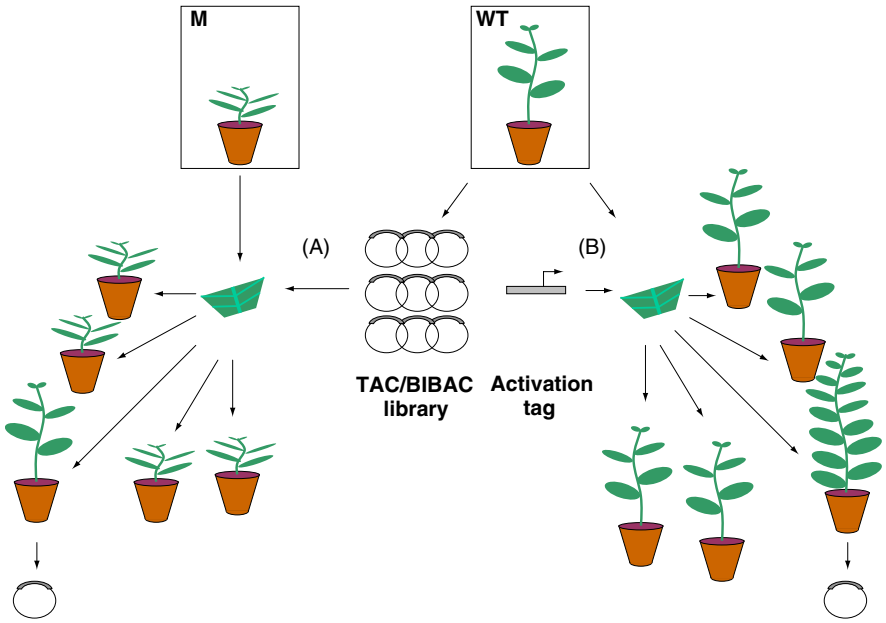

Figure 5 Functional cloning. (A) Complementation approach. A large insert library (BIBAC or TAC) is derived from a wild-type plant. The clones are systematically introduced into explants of a mutant plant and the transgenic clones regenerated. Any clone capable of rescuing the mutant phenotype of the plant must carry the corresponding gene. (B) Gain-of-function approach. Wild-type plants are transformed with an activation trap construct, essentially a T-DNA or transposon carrying a strong, outward-facing promoter. This integrates into the genome randomly and may activate adjacent genes. In some plants, a strong gain-of-function phenotype may be recognized. The activated gene can be cloned by the methods shown in **Figure 4**. M, mutant; WT, wild-type.

(TAC) vectors) has greatly facilitated the isolation of plant genes by functional complementation in transgenic plants.

- Analysis of gain-of-function phenotypes in transformed plant cells or transgenic plants. This can be achieved by overexpression or ectopic expression. An interesting variation is the use of activation tag constructs, insertional vectors based on transposons or T-DNA that contain a strong, outward facing promoter, which boosts the expression of any adjacent gene (**Figure 5B**). Such constructs have been used, for example, to identify genes encoding transcription factors that coordinately regulate several genes of a metabolic pathway.

Difference Cloning

It is often desirable to identify cell-specific genes or genes that are expressed in particular states or in response to particular stimuli. Difference cloning describes any cloning strategy that identifies sequences represented in one source of DNA, but not another, or sequences that differ significantly in abundance between sources. In most cases, this means differentially expressed cDNAs, representing genes that are expressed in one sample, but inactive in another (e.g., two different tissues, two different developmental stages, stimulated vs. nonstimulated, healthy vs. diseased). There are various cell based and PCR based difference cloning approaches, but all of them follow one of two principles: enrichment or display.

Enrichment techniques work by removing sequences that are shared between two DNA sources, therefore leaving behind those sequences that are unique. One common approach is subtraction cloning (**Figure 6**), in which mRNA from one source (known as the driver) is used to subtract homologous cDNA sequences from a second source (known as the tester). This is often achieved by labeling the driver mRNA sample universally with biotin, and then adding it in great excess to the tester sample, which has been used for first strand cDNA synthesis. The single strands are allowed to anneal. Since the biotinylated mRNA is present in such great excess, most of the hybridized cDNA strands that represent shared genes will form duplexes with biotinylated mRNA, allowing them to be selectively removed by affinity to streptavidin. Several rounds of subtraction can be carried out to generate a first strand cDNA population that is highly enriched for clones unique to the tester. This can be used to generate a subtracted cDNA library. A reciprocal library can be generated by using the second sample as the driver and the first as the tester. A similar, PCR based approach is representational difference analysis, in which linkers are added to the tester population and mixed with a large excess of the driver population, which lacks the linkers. Most of the shared products form heteroduplexes in which only one strand has linkers, and therefore cannot be amplified using primers that anneal to the linkers. However, products unique to the tester population can only form homoduplexes in which both strands have linkers, and these can be successfully amplified.

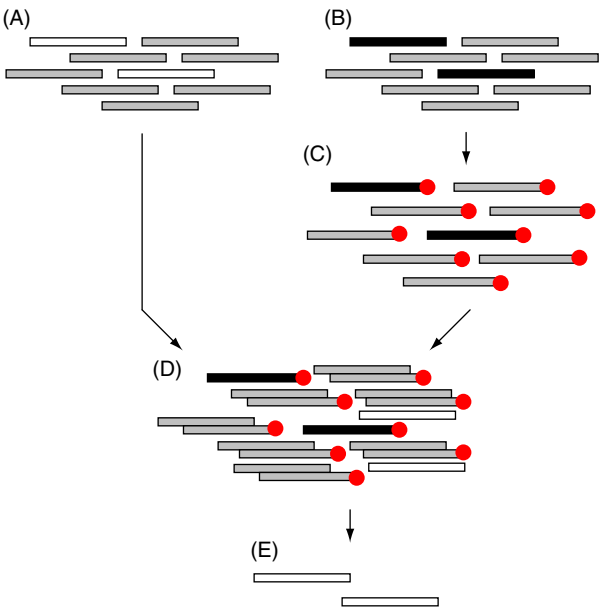

Figure 6 The principle of subtraction cloning. Two different mRNA samples are shown, representing the tester (A) and representing the driver (B) Transcripts unique to the tester population are white, while those unique to the driver population are black. Gray transcripts are common to both populations. (C) The mRNA from the driver population is universally labeled with biotin. (D) The mRNA from the tester population (A) is converted into cDNA and mixed with a large excess of labeled driver mRNA. The two populations hybridize. (E) The mixture is passed through a column of streptavidin beads. Most of the tester cDNAs that are common to both populations are sequestered into biotinylated duplexes and are retained on the column. cDNAs unique to the tester population are eluted and can be used to make a subtracted library.

Display methods work on the principle that all clones (or a representative subset thereof) are visually displayed, allowing differentially expressed sequences to be identified by eye. The most popular display method is differential display PCR, in which arbitrary PCR reactions are used to amplify pools of cDNA fragments equivalent to ESTs, and display them side-by-side on a sequencing gel. Bands that are present in one sample, but not the other, are selected for more detailed characterization, as these potentially correspond to differentially expressed genes.

Cloning Strategies: From Genes to Genomes

The cloning strategies discussed above have been developed to identify individual genes of interest with the assumption that the rest of the genome is an unknown void. However, genome projects are ongoing for a number of important plant species and two genome sequences, those of *Arabidopsis thaliana* and rice, have already been completed. With the sequence in hand, the next tasks are gene finding (structural annotation) and the elucidation of gene functions (functional annotation). Due to the large amounts of data arising from the genome projects, bioinformatics methods are employed for the heavy work of front-end annotation. The integration of database search methods and *ab initio* gene prediction can build up a virtual gene landscape, which can be refined by the assignment of known genes and ESTs onto the physical clone map. The increasing use of databases to share information between laboratories means that quite a lot can be learned about the function of a gene without doing any experiments – functional annotations based on sequence or structural homology, conserved expression patterns, mutant phenotypes, and protein interactions are available across many different species. Therefore, when map-based cloning reveals a number of possible genes in the region defined by linkage analysis, it is often possible to home in on candidate genes, which are the most likely to be involved in the process of interest. In species with well developed physical maps, it may even be possible to obtain the corresponding DNA clone from a central resource.

The isolation of the corn *Ivr2* gene provides an example of this candidate gene approach. Genetic mapping initially established a very large region containing a QTL affecting the response of corn plants to water stress. However, this interval was far too large to consider a search for individual genes. In a separate series of experiments, corn plants were subjected to water stress and cDNAs were isolated corresponding to genes expressed specifically under stress conditions. These cDNAs were assigned to the physical map and one of them mapped to the same locus as the previously identified QTL. In this case, the correct gene was identified from the many alternatives in the candidate interval because of its relevant expression profile.

The reductionist strategy of studying individual genes is being supplanted progressively by a new holistic approach that involves the study of whole genomes. At the beginning of this article, it was stated that the bottleneck in cloning is not amplifying the DNA sequence of interest, but isolating that sequence from the nonspecific background. We are rapidly approaching the day when this bottleneck will also be removed and all genes will potentially be available "off the shelf" from central genome project resources, at least for a range of model species and the commercially important crops. The new bottleneck will therefore not be a cloning strategy, but a strategy to link genes with their corresponding phenotypes. Genomics has provided us with a whole spectrum of high throughput functional analysis tools to make this possible, and this article closes with some examples:

- DNA microarrays. These are small devices with thousands of different DNA sequences immobilized in a grid. Genome-wide arrays are available for several microbes, and it is likely that an *Arabidopsis* genomic array will be available in the near future. They can be interrogated with complex hybridization probes, and differentially labeled cDNAs from two samples can be hybridized simultaneously to identify genes that are expressed more strongly in one sample than in another. Microarrays have been used to identify genes induced by drought stress, disease, oxidative stress, wounding, insect feeding, and many other processes in several species, but particularly in *Arabidopsis* and rice. In *Arabidopsis*, microarrays have been used for a methodical analysis of the molecular basis of systemic acquired resistance.
- Large scale EST sampling. This is an alternative to microarrays and involves the statistical analysis of EST data from different cDNA libraries to identify differentially expressed genes. It has been used to identify genes induced by salt stress in *Arabidopsis* and in several salt tolerant plants.
- Proteomics. The proteome is the entire collection of proteins in the cell. A high-resolution separation method, such as two-dimensional electrophoresis or high performance liquid chromatography, is used to resolve the thousands of proteins in a plant cell into individual fractions, which are identified by mass spectrometry. This approach has been used to characterize proteins involved in the response to a number of biotic and abiotic stresses, and to identify polymorphisms that can be used as genetic markers.
- Large-scale transposon and T-DNA mutagenesis programs have been established in rice, corn, and *Arabidopsis*. In the case of corn and *Arabidopsis*, databases of insertion flanking sequences have been established. Therefore, an investigator studying any particular gene can search the database for insertions and order seeds for the corresponding mutant plant line. Viral based gene silencing and RNA interference represent newer technologies

for systematic gene inactivation, which are more rapid than the production of mutants.

- Large-scale studies of protein interactions are possible using mass spectrometry to analyze protein complexes or library-based techniques such as phage display and the yeast two-hybrid system to probe interactions systematically. Interactions among the MADS box transcription factors that act combinatorially in the specification of cell fates within the developing floral organs have been studied in this way.

Summary

There are two established techniques for cloning DNA sequences – cell based molecular cloning and *in vitro* amplification by PCR. Cell based cloning procedures are particularly versatile because they can be used to identify genes based on the properties of the gene product, as well as the sequence itself. The current bottleneck in plant gene cloning is the isolation of specific genes from the large background of irrelevant sequences in genomic DNA or cDNA. Sequence based, map based, immunological, and functional approaches have been described, as well as strategies that help to enrich for differentially expressed genes. As more genomes are sequenced and annotated, this bottleneck will be removed and genes will be available to order. A whole-genome sequencing strategy has been used for the relatively small rice and *Arabidopsis* genomes, whereas for the larger genomes of crops such as wheat (*Triticum* spp.) and corn, the large excess of repetitive DNA may make a gene-focused strategy more applicable. In either case, the challenge ahead will not be cloning, but identifying associations between genes and their functions.

List of Technical Nomenclature

Ab initio gene prediction Identification of genes in genomic DNA by looking for gene-like features (promoters, splice sites) rather than relying on homology to known sequences.

Activation tag An insertion element that carries a strong, outward facing promoter, which can activate nearby genes.

Adaptor, linker A double-stranded oligonucleotide that fits over the sticky ends left by a restriction digest (adaptor) or blunt ends left by a restriction digest or PCR (linker), and provides a primer annealing site.

AIMS Amplification of insertion mutagenized sites. Nested PCR using a transposon specific primer for linear amplification.

Candidate gene approach Map-based cloning strategy in which the choice of genes within the candidate interval is assisted by functional information.

CAPS markers Polymorphic markers identified by PCR followed by the digestion of the PCR products with restriction enzymes.

Chromosome landing Situation where a single genomic clone contains both the target gene and the nearest genetic marker, avoiding the necessity for chromosome walking.

Chromosome walking Assembly of a physical map by identifying a series of overlapping clones.

Clone library A collection of DNA clones in vectors.

Codon The sequence of three nucleotides in a mRNA that specify an amino acid.

Degeneracy The situation in which an amino acid can be specified by two or more codons.

Differential display PCR Side-by-side display of equivalent RT-PCR products from two sources, showing differentially expressed genes.

Domain sharing genes Genes that can be aligned over part of their lengths because they share a particular domain, but may be completely unrelated elsewhere in the sequence. Such genes have arisen by domain shuffling during evolution.

Driver The source of DNA that is used to drive hybridization and therefore remove common sequences during subtractive cloning.

Ectopic Outside the normal spatial or temporal domain.

Epitope The specific region on a protein recognized by an antibody.

Expressed sequence tags (ESTs) Short cDNA fragments obtained by random sequencing of clones from cDNA libraries.

Expression qcloning Any cloning strategy that requires the expression of the target gene.

Expression library A clone library in which each cDNA insert is expressed.

Functional cloning An expression cloning strategy in which the biochemical, cellular, or biological function of the gene is exploited.

Functional complementation Restoration of the wild-type phenotype by supplying a mutant cell or organism with the missing gene product.

Gene families Homologous genes, often used to describe complete sets of paralogous genes in one genome.

Genetic mapping, linkage analysis Mapping based on recombination frequencies (frequency of meiotic crossovers) between genes and/or markers.

Genetic marker A polymorphic DNA sequence that can be used in genetic mapping.

Homeobox DNA sequence encoding the homeodomain, a DNA binding protein domain found in many transcription factors.

Homologous Sequences related by divergence from a common ancestor.

Host The cell used for cell based cloning.

Immunological screening Library screening with an antibody probe.

Insertional mutagenesis Mutation produced by the insertion of a transposon or a T-DNA sequence.

Inverse PCR PCR on a circularized template in which the primers face outward from an insertion sequence and amplify flanking DNA.

MADS box A DNA binding protein domain found in many transcription factors, including those controlling floral organ specification.

Map based cloning, positional cloning Gene cloning strategy dependent entirely on map position, with no biochemical, sequence, or functional information.

Mass spectrometry Technique for determining accurate molecular masses, which can be applied to the identification of proteins.

Molecular cloning, DNA amplification Production of many copies of the same DNA sequence.

Nested PCR Sequential PCRs in which one or both primers are represented by nested sets placed progressively further inward along the template to increase specificity.

Northwestern screening Detection of RNA binding proteins using an RNA probe.

Paralogous Homologous sequences in the same genome that have arisen by gene duplication and divergence.

Phage display Detection of interacting proteins by testing bait proteins with libraries of phage carrying different proteins on their surfaces, achieved by nonselective cloning into the phage coat protein gene.

Physical mapping Any mapping technique in which the distances between genes and other markers are measured in real physical distances (kilobases, megabases).

Plasmid rescue The inclusion of a plasmid origin of replication and antibiotic resistance marker in a transposon so that the flanking genomic region can be isolated and cloned as a bacterial plasmid.

Polymerase chain reaction Exponential *in vitro* cloning method.

Quantitative trait locus (QTL) Locus that contributes to a trait showing continuous variation.

RACE Rapid amplification of cDNA ends. A PCR technique for the cloning of full-length cDNA sequences.

Replicon A DNA element that can replicate autonomously.

RNA interference Posttranscriptional silencing induced by double-stranded RNA. Probably the underlying mechanism of viral gene silencing.

Southwestern screening Detection of DNA binding proteins using a DNA probe.

Splice variant Alternative transcript/cDNA produced by differential splicing of the original gene.

Stringency Conditions that determine how perfectly matched two complementary sequences need to be in order to anneal.

Subtraction cloning Construction of a library enriched for differentially expressed genes by removing cDNA clones common to two samples.

TAIL PCR Thermal asymmetric interlaced PCR. Similar to AIMS, but using a short arbitrary primer rather than an adaptor-specific primer.

Tester The source of DNA that has common sequences subtracted by the driver.

Transposon display Nested PCR using a transposon-specific primer and a primer annealing to an oligonucleotide adaptor.

Two-dimensional gel electrophoresis Protein separation technique in which proteins are fractionated according to their charge in the first dimension and then by their mass in the second dimension.

Vector A replicon used for cloning.

Viral gene silencing Deliberate posttranscriptional silencing of plant genes using viral vectors carrying plant transgenes.

Yeast two-hybrid system A high-throughput method for detecting protein interactions in which interacting fusion proteins assemble a functional transcription factor.

See also: **Crop Improvement**: Chromosome Engineering; Genetic Maps; Molecular Markers; Mutation Techniques; Plant Genomes. **Genetic Modification**: Transformation, General Principles.

Further Reading

Brown TA (2001) *Gene Cloning and DNA Analysis: An Introduction*, 4th edn. Oxford: Blackwell Science.

Jannink J-L, Bink MCAM, and Jansen RC (2001) Using complex plant pedigrees to map valuable genes. *Trends in Plant Science* 6: 337–342.

Maes T, De Keukeleire P, and Gerats T (1999) Plant tagnology. *Trends in Plant Science* 4: 90–96.

Primrose SB and Twyman RM (2002) *Principles of Genome Analysis and Genomics*, 3rd edn. Oxford: Blackwell Science.

Primrose SB, Twyman RM, and Old RW (2001) *Principles of Gene Manipulation*, 6th edn. Oxford: Blackwell Science.

Rabinowicz PD, McCombie WR, and Martinessen RA (2003) Gene enrichment in plant genomic shotgun libraries. *Current Opinion in Plant Biology* 6: 1–7.

Shibata D and Liu YG (2000) *Agrobacterium*-mediated plant transformation with large DNA fragments. *Trends in Plant Science* 5: 354–357.

Waugh R, Dear PH, Powell W, and Machray GC (2002) Physical education – new technologies for mapping plant genomes. *Trends in Plant Science* 7: 521–523.

Insertional and Transposon Mutagenesis

R M Twyman, University of York, York, UK
A Kohli, John Innes Centre, Norwich, UK

Introduction

Insertional Mutagenesis in Forward Genetics

The most direct way to establish the function of a plant gene is by mutation. Traditionally, gene–function relationships have been studied in a phenotype driven approach, sometimes described as forward genetics. Mutagenesis has been achieved by irradiating seeds or treating them with chemical mutagens, and then searching for interesting mutant phenotypes in subsequent generations. If a large enough population of seeds is mutagenized, it should be possible to recover mutants affecting any gene in the genome. Dominant mutations are often revealed in the first generation of plants, but recessive mutations require breeding to homozygosity. If such mutations are lethal, they can be maintained in heterozygous plants.

The disadvantage of radiation and chemical mutagens is that they tend to generate point mutations, which are difficult to map and make the process of gene isolation laborious and time consuming. Insertional mutagenesis is the use of DNA sequences as mutagens. Insertional mutagenesis is advantageous because genes tend to be completely disrupted, which favors the recovery of null mutants. Furthermore, the interrupted gene becomes tagged with the insertion, which can be used to clone flanking sequences, either by traditional means or strategies based on the polymerase chain reaction (PCR). The approach is generally known as gene tagging or signature tagged mutagenesis (STM) and is discussed in greater detail below.

Insertional Mutagenesis in Reverse Genetics

Reverse genetics is a gene-driven approach, in which a sequence with no known function is deliberately mutated allowing the resulting mutant phenotype to be studied. In certain lower plants, such as the moss *Physcomitrella patens*, it is possible to introduce exact changes into preselected genes by a process called gene targeting. This relies on homologous recombination between the target gene (in the genome) and a homologous sequence that has been mutated *in vitro* and introduced into the cell. Unfortunately, homologous recombination occurs at a very low frequency in higher plant cells. Therefore, although gene targeting has been achieved in *Arabidopsis thaliana* and in tobacco (*Nicotiana tabacum*) the process is so inefficient that it is unsuitable for routine application.

Insertional mutagenesis has recently become established as an alternative reverse genetics strategy in plants. Although it is not possible to systematically target specific genes for disruption, it is possible to saturate the genome with mutations and maintain a resource of seeds representing interruptions in every single gene. A researcher discovering a particular gene and wanting to know its function can then search for insertional mutants affecting the gene of interest, either by screening the mutant seed bank or by searching databases of flanking sequences. We consider the practicalities of this approach below.

Insertional Mutagenesis Systems

A number of different insertional mutagenesis systems have been developed for plants, but all are based either on transposons or the integration of T-DNA. Transposons are mobile genetic elements that jump

from site to site in the genome when supplied with the enzyme transposase. A number of transposons have been used for mutagenesis programs, including *Activator* (*Ac*), *Suppressor-mutator* (*Spm*) and *Mutator* (*Mu*) from corn (*Zea mays*; maize), *Tam3* from *Antirrhinum majus*, and *Tos17* from rice (*Oryza sativa*). Some of these transposons function only in their natural host plant, while others are promiscuous and can be used in a variety of heterologous species. The corn *Ac* element, for example, has been shown to function in *Arabidopsis*, tobacco, rice, potato (*Solanum tuberosum*), petunia (*Petunia hybrida*), tomato (*Lycopersicon esculentum*), and many other plants. T-DNA is not a transposon but is a short segment of DNA that is transferred from a bacterial plasmid to the plant genome when plants are infected by the soil pathogen *Agrobacterium tumefaciens*. T-DNA transfer by *A. tumefaciens* has been used extensively to generate transgenic plants, and, more recently, for insertional mutagenesis.

Properties of Transposons and T-DNA

Transposons and T-DNA each have their advantages and disadvantages as insertional mutagens (**Table 1**). T-DNA is a bacterial sequence, and therefore always provides a unique signature when introduced into the plant genome. However, not all plants are susceptible to *A. tumefaciens*, so transposons remain the only choice for insertional mutagens in many species. All plants contain some form of transposable element population, but this means that several copies of any given transposon may already be present in the genome making it difficult to identify novel insertion events. This problem can be overcome by introducing unique foreign sequences into a transposon that is going to be used as an insertional mutagen in its normal host species. Alternatively, a transposon from one species may be introduced into a heterologous plant. In this case, either the *Agrobacterium* T-DNA system or a direct transfer approach such as particle bombardment must be used to achieve initial gene transfer.

T-DNA and transposons also differ in their behavior. Once integrated, T-DNA is unable to jump ever again because it lacks the ability to self-mobilize. Therefore, insertions generated by T-DNA are usually stable and permanent. In contrast, if a transposase is present, transposons may continue to jump in subsequent generations of plants. The advantage of this continued activity is that large numbers of insertion events can be generated with a small number of starter lines. However, the disadvantage is that the first generation of insertion events may be unstable, since the transposons used as mutagens in plants tend to move by a cut-and-paste mechanism, which involves excision from the original insertion site and integration elsewhere. There is a number of strategies available to generate stable transposon insertions. The most common is to use a two-component transposon system, which comprises an autonomous (self-mobilizing) element and a non-autonomous derivative. For example, the corn transposon *Activator* (*Ac*) is autonomous because it encodes its own transposase, but shorter derivatives of *Ac* called *Dissociation* (*Ds*) lack the transposase gene. *Ds* elements can transpose if transposase is provided by a copy of *Ac* in the same genome, but not when *Ac* is absent. Therefore, if *Ac* and *Ds* are introduced into the same genome, both elements can be mobile, but if *Ac* is removed by crossing, the *Ds* insertions will be stable. Another strategy is to use a retrotransposon system, such as *Tos17* in rice. Retrotransposons mobilize at the RNA level and move by a replicative mechanism involving transcription of an integrated element and the insertion of a cDNA copy of the transcript at a different location. The original insertion remains intact during this process.

Both transposons and T-DNA appear to favor integration into gene-rich DNA, which may reflect some preference for active chromatin structure. However, while the integration of T-DNA is essentially random within the gene space of the plant genome, transposons all appear to show greater or lesser degrees of target site preference, often defined

Table 1 Comparison of transposon and T-DNA derived insertional mutagens

Transposons	*T-DNA*
Endogenous in all plants	Bacterial in origin but can be introduced into many plants
Ideally, transposon based insertion vectors need to be modified or introduced into heterologous species to distinguish them from endogenous elements	Bacterial sequence; provides unique sequence signature in all plants
Insertions may be unstable	Insertions generally stable
Small number of founder lines can generate multiple insertions	Large number of founder lines required
Integrates preferentially into gene rich DNA	Integrates preferentially into gene rich DNA
Shows target site preference and, in some cases, local reinsertion	Random integration within gene space
Generally clean insertions	Often generates complex insertions

by a weak consensus sequence at the precise integration site. Transposons tend to generate clean insertions that facilitate the isolation of flanking sequences. In contrast, T-DNA insertions are often messy. There may be two or more copies arranged as tandem or inverted repeats, and genomic DNA at the integration site is more likely to be rearranged. The isolation of flanking sequences can be difficult at such complex insertion sites.

Many transposons, including *Ac*, also demonstrate a phenomenon known as local reinsertion, where transposition preferentially occurs to linked sites. This can be useful if the aim is to saturate a particular region of the genome with mutations, but it can make genome-wide saturation mutagenesis difficult. One way to approach this problem is to generate founder lines with regularly spaced *Ac* insertions. Another way to avoid local reinsertion is to use one of several available counterselection strategies, i.e., strategies that select against local reinsertion, to avoid transposition to linked sites. Alternatively, a different transposon system should be used. The corn transposon *Mu*, for example, does not demonstrate preferential local reinsertion. Local reinsertion may also depend to a certain extent on the genetic background. *Ac*, for example, transposes preferentially to linked sites in corn and *Arabidopsis*, but not in rice.

Identifying Interrupted Genes

Forward Genetics – Isolation of Flanking Sequences

Any phenotype generated by a basic transposon or T-DNA insertion is likely to occur because the insertion element has disrupted the corresponding gene, which means that the gene's sequence will flank the insertion site. Insertional mutagenesis, therefore, provides a short cut to isolating the gene in question without recourse to linkage mapping and the assembly of large clone contigs. Conceptually, the simplest method of gene isolation is to generate a genomic library from the mutated plant, and isolate clones containing the insertion element (and several kilobases of flanking DNA) by hybridization with the cloned transposon or T-DNA. However, there is a number of even more convenient methods, which allow the flanking sequences to be isolated in a matter of hours without the need for library preparation.

Plasmid rescue In the simple plasmid rescue technique, the T-DNA or transposon is modified to contain the origin of replication from a bacterial plasmid and an antibiotic resistance marker functional in *Escherichia coli*. Genomic DNA is isolated from a particular mutant plant of interest and cleaved with a restriction enzyme that does not cut in the insertion element or cuts only once, but leaves the origin and selectable marker gene intact. The DNA is then diluted and DNA ligase is added, promoting self-circularization. This mixture of random genomic circles is then used to transform *E. coli*, and bacteria are cultured under antibiotic selection. Only the genomic fragment containing the plasmid sequences will be maintained as a plasmid in the bacterial culture (**Figure 1**).

PCR based approaches The first PCR based technique developed for flanking sequence isolation was inverse PCR. Genomic DNA from a particular mutant plant line is isolated, cleaved with a restriction enzyme as discussed above for plasmid rescue, and self-circularized. PCR amplification is then carried out using primers that specifically anneal to the insertion element (transposon or T-DNA) but point outward rather than inward. In a circular DNA molecule, this will result in the amplification of sequences flanking the insert (**Figure 2**).

A number of more sophisticated PCR approaches is available that do not require circularization of the genomic DNA prior to amplification. These include transposon display and amplification of insertion mutagenized sites (AIMS), both of which involve the digestion of genomic DNA with restriction enzymes, the addition of adaptor oligonucleotides, and several rounds of nested PCR. In transposon display, amplification is exponential and requires a transposon specific primer and an adaptor specific primer. In AIMS, only the transposon specific primer is used and amplification is linear. Another nested PCR technique is thermal asymmetric interlaced (TAIL) PCR, which involves the use of three nested transposon specific primers and a small arbitrary primer. As an alternative to genomic DNA, insertion events can be detected at the mRNA level by rapid amplification of cDNA ends (RACE).

Reverse Genetics – Identifying Insertions in Specific Genes

Two general strategies are available for the identification of inserts in specific genes, one based on high-throughput sequencing and one based on gene-by-gene PCR analysis. In both strategies, the sequences or PCR data must be correlated with a seed bank or collection of plant lines.

The first strategy is the systematic sequencing of all flanking sequences, and their deposition into a

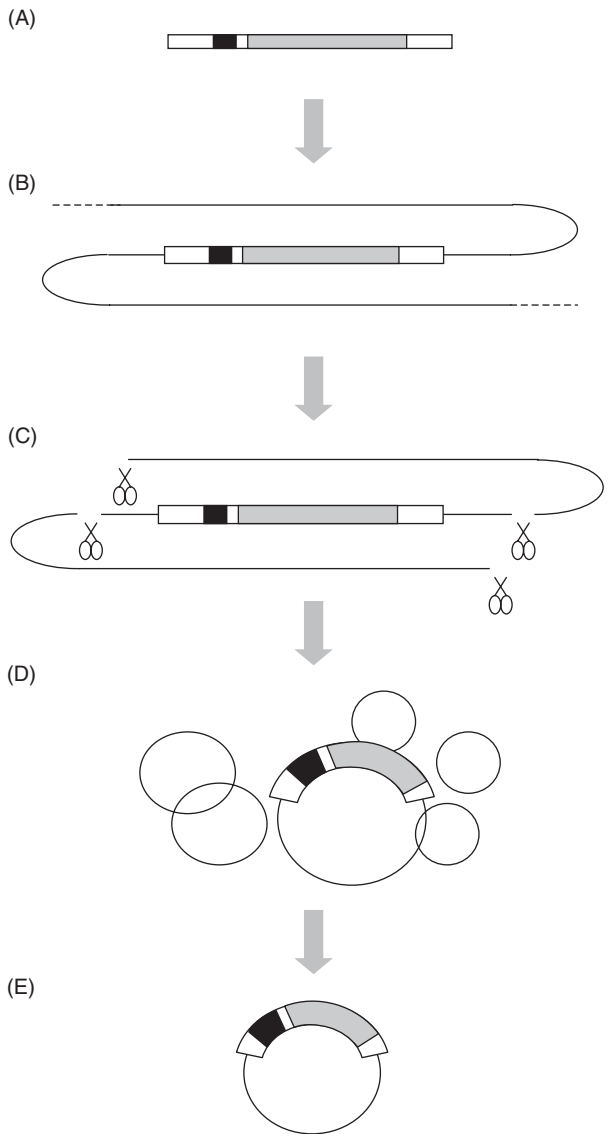

Figure 1 Principle of plasmid rescue technique. (A) Insertion construct contains plasmid origin of replication (black) and bacterial antibiotic resistance marker (gray). (B) The construct inserts into genomic DNA. (C) Genomic DNA is cleaved with restriction enzymes that do not cut in the construct. The construct is liberated with flanking genomic sequences. (D) The cleaved DNA is diluted and ligated to form circles. One of the circles forms a functional plasmid. (E) Transformation of *E. coli* followed by antibiotic selection allows the plasmid, and the genomic sequences it contains, to be cloned.

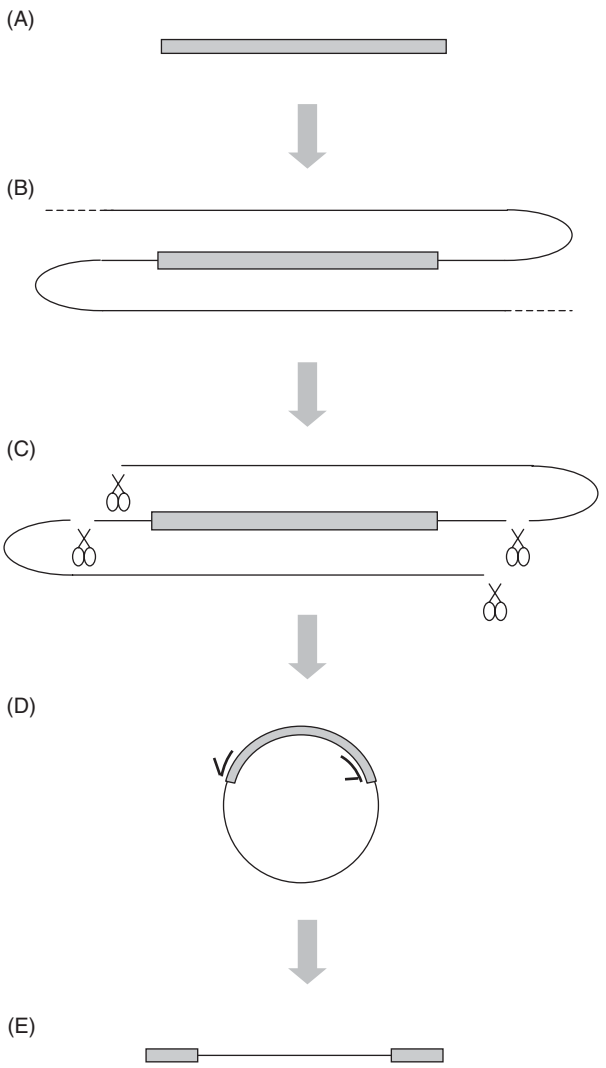

Figure 2 Principle of inverse PCR technique. (A) The insertion construct is shown as a gray box. (B) The construct inserts into genomic DNA. (C) Genomic DNA is cleaved with restriction enzymes that do not cut in the construct. The construct is liberated with flanking genomic sequences. (D) The cleaved DNA is diluted and ligated to form circles. The circle containing the insertion construct is recognized by insertion-specific primers. (E) Amplification using this primer pair allows the flanking sequences to be cloned.

database that can be searched by researchers working on particular genes. Ideally, it should be possible to search the database over the Internet using bioinformatic tools such as BLAST or FASTA. For example, several genome-wide transposon mutagenesis projects have been carried out in *Arabidopsis*, and many of the resulting mutant lines are maintained at the Nottingham *Arabidopsis* Stock Centre in the UK and/or the US *Arabidopsis* Biological Resource Center in Columbus, Ohio. The flanking sequences have been annotated, and the results are available on searchable databases allowing individual researchers to identify inserts in specific genes and obtain the corresponding plants. Another 130 000 plant lines, with T-DNA inserts and associated flanking sequence data, are maintained at the University of Wisconsin *Arabidopsis* Knockout Facility.

The alternative strategy is the maintenance of a genomic DNA bank that can be screened by PCR. Screening is achieved simply by providing one PCR

primer representing the gene of interest and one primer representing the insertion element. Typically, DNA samples from seeds are pooled in a hierarchical manner, so that individual PCRs do not have to be carried out on thousands of DNA samples. Such screening facilities are available, for example, at the Pioneer Hi-Bred Trait Utility System for Corn (TUSC) project, and the Maize Targeted Mutagenesis project (MTM). In each case, PCR primers facing away from the transposon are used in combination with a gene specific primer to identify insertions into specific genes (**Figure 3**). The Maize Gene Discovery Project uses a similar approach on a library of plasmids obtained by plasmid rescue with a modified *Mu* transposon called *RescueMu*.

Specialized Insertion Constructs

Another advantage of insertional mutagenesis over traditional random mutagenesis is that the insertion construct can be modified to provide information about the interrupted gene, or even to modify genome structure and activity. Several different types of specialized insertion vector are discussed below.

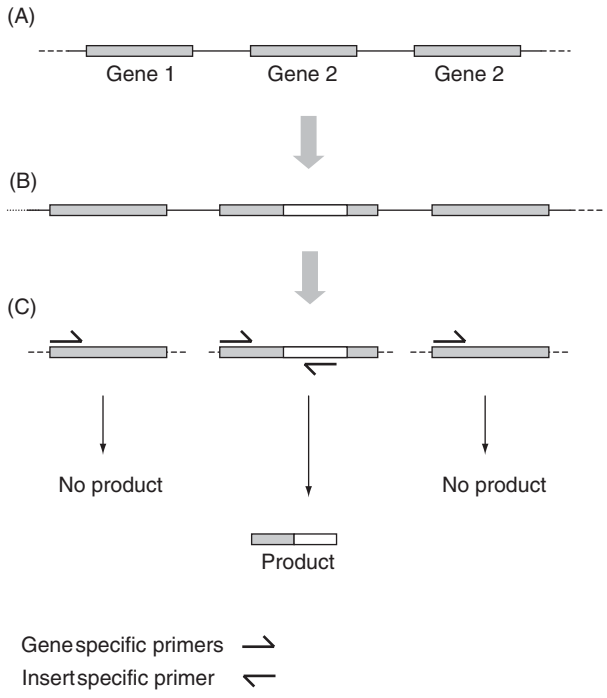

Figure 3 Identification of insertions in specific genes by PCR. (A) A section of genomic DNA is shown containing three genes. (B) An insertion construct (white box) interrupts gene 2. (C) Individual PCRs carried out using one gene specific primer and one insert specific primer generate a product only if the plant contains an insert in the gene being tested.

Enhancer Traps

An enhancer trap is an insertional construct that contains a weak promoter upstream of a reporter gene, such as *gusA* (**Figure 4A**). The promoter is too weak to generate anything other than minimal expression on its own, but if the construct inserts in the vicinity of an endogenous enhancer, this may activate the gene and generate a strong and specific expression pattern. Enhancer traps based on T-DNA vectors have been used in plants, but their usefulness for gene isolation is limited since enhancers may act over distances of hundreds of kilobases.

Gene Traps

A gene trap is an insertional construct containing a reporter gene or even a selectable marker gene that is activated only when the construct integrates within a

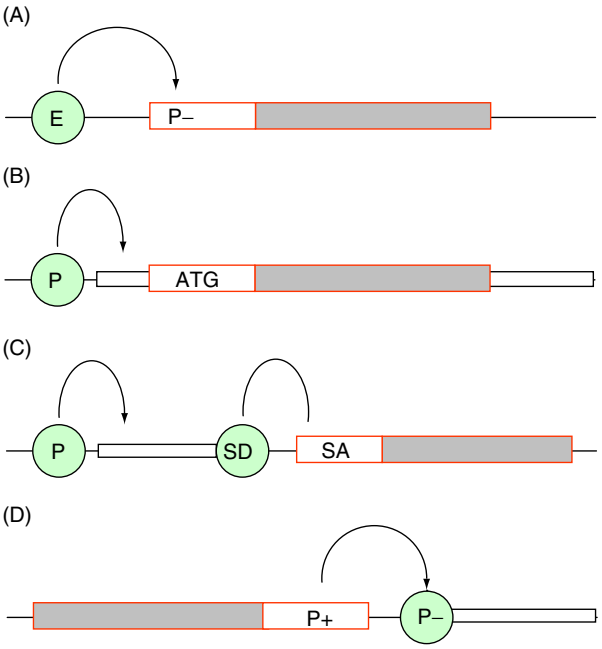

Figure 4 Specialized insertion elements. In each case the insertion element is shown as a wide box, and plant sequences that respond to the insertion are shown as circles. (A) Enhancer trap. The reporter gene (gray box) is driven by a weak promoter (P –). If it inserts in the vicinity of a local enhancer (E) the reporter gene is activated. (B) Promoter trap. The reporter gene (gray box) has no promoter, just an initiation codon (ATG). The reporter can only be activated if it inserts within a plant gene (narrow white box) with an active promoter (P). (C) Splice trap. The reporter gene (gray box) is downstream of a splice acceptor site (SA). This can only be activated if it inserts within a gene and downstream of a splice donor site (SD), enabling it to be expressed as a transcriptional fusion. The gene must also have an active promoter (P). (D) Activation tag. This construct has an outward facing strong promoter (P +). If it inserts adjacent to a gene with a weaker or more restricted promoter (P –) the gene may be upregulated or expressed ectopically.

gene. There are several different types, but all work on the basis that marker gene expression is dependent somehow on the gene's potential or actual activity. The simplest constructs are known as promoter traps and comprise a completely promoterless marker gene, which is activated only if the construct inserts downstream of an active plant promoter (**Figure 4B**).

More sophisticated splice traps (**Figure 4C**) contain an intron and splice acceptor site before the marker gene. In this case, marker gene expression is dependent not only on an active promoter, but also on intron splicing and the inclusion of the marker gene as a transcriptional fusion. Gene traps of this nature have been widely used in plants, and several functional genomics programs have been initiated. For example, a T-DNA gene trap vector incorporating the screenable marker *gusA* has been used to produce 22 000 rice mutants. Over 5000 tagged lines were analyzed for reporter gene expression in leaves and roots, 7000 lines for expression in flowers, and 2000 for expression in seeds. Overall, about 2% of the lines showed marker gene activity, in many cases restricted to highly specific cell types or tissues.

A disadvantage of the current gene trap systems used in plants is that they are dependent on gene activity. Integration into a gene that is not expressed when the plants are analyzed will not yield a reporter phenotype. In the future, this problem could be addressed by the use of more sophisticated vectors in which expression of the marker is driven by its own promoter, but is dependent on the gene for polyadenylation.

Activation Tag Constructs

An activation tag carries a strong, outward facing promoter (**Figure 4D**). Unlike other insertion constructs, which tend to generate loss-of-function phenotypes by gene disruption, the principle of the activation tag is to generate gain-of-function phenotypes by promoting overexpression or ectopic expression of adjacent genes. This strategy has been used, for example, to identify genes regulating alkaloid biosynthesis in the medicinal plant *Catharanthus roseus*. A T-DNA construct containing the strong and constitutive CaMV35S promoter adjacent to the right border was introduced into the plant, and the resulting population of transformants selected for increases in tryptophan decarboxylase activity. This enzyme is a key early component of the alkaloid biosynthesis pathway, and its increased expression is necessary for any increase in downstream alkaloid production. Using this method, a novel transcription factor called ORCA3 was identified.

Monitoring Excision Activity

A transposon construct, in which the transposon is sited between a promoter and a reporter gene, allows excision events to be monitored directly. Where transposition has not occurred, the reporter gene remains silent. However, upon excision of the transposon, the promoter and reporter gene are brought into juxtaposition, and the reporter is activated. This direct screening method can be used in conjunction with PCR based tests to characterize excision activity. An example of green fluorescent protein used as an excision marker for *Ac* in rice is shown in **Figure 5**.

Chromosome Engineering

Transposon mutagenesis can be used to introduce more striking changes to the plant genome. Large chromosomal deletions have been introduced into the *Arabidopsis* genome by introducing a DNA construct containing a *Ds* transposon and two *loxP* sites, which are recognized by the Cre site specific recombinase. The construct also contains a reporter gene and promoter arranged to monitor for excision as discussed above. The design of the construct is shown in **Figure 6**. The critical feature is that one of the *loxP* sites lies within the *Ds* element and one lies without. Integration is random, but excision of the *Ds* element results in preferential reinsertion at a linked site, and carries the *loxP* site with it. In this manner, a large section of genomic DNA is delimited by two *loxP* sites. If Cre recombinase is then supplied, the two *loxP* sites recombine resulting in the deletion of the intervening genomic DNA.

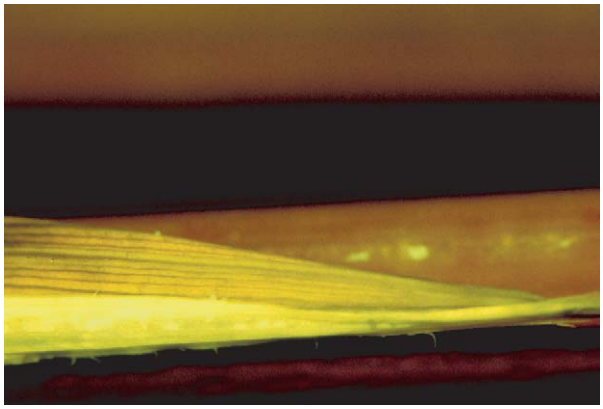

Figure 5 Green fluorescent protein used as an excision marker in rice. Rice plants were transformed with a construct in which the *Ac* element is interposed between the *gfp* gene and its promoter. The observed mosaic pattern reflects late excision events affecting restricted clones of cells.

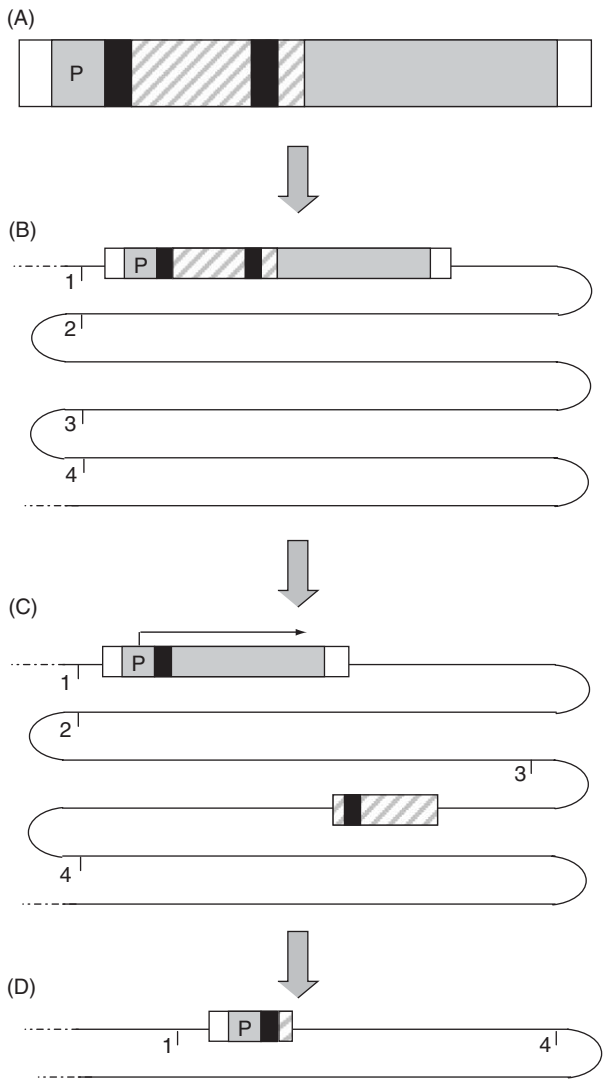

Figure 6 Chromosome engineering with *Ds*/Cre-*loxP* vectors. (A) Construct design. The gray box represents a reporter gene (P is its promoter). Between the promoter and reporter gene is a *Ds* element (hatched box) with one 5′ *loxP* site and one internal *loxP* site (black boxes). (B) The construct inserts into genomic DNA. The ticks and numbers represent physical markers along the chromosome. (C) The *Ds* element excises from the construct and reinserts 10–50 kb away, carrying the *loxP* site with it. The reporter gene is activated (arrow). (D) If Cre recombinase is made available, site specific recombination occurs between the *loxP* sites and deletes a large segment of genomic DNA, as shown by the loss of markers 2 and 3.

Conclusions

Insertional mutagenesis using either T-DNA or transposons is being developed into an increasingly sophisticated palette of tools for functional genomics, providing the means to generate genome-wide libraries of mutants that allow the interrupted genes to be identified rapidly. In combination with other genomic resources, such as genome sequences and EST collections, this provides an efficient route to high-throughput functional annotation. While basic insertion constructs generate loss-of-function phenotypes, additional refinements show how the interrupted genes are expressed, and may even produce gain-of-function phenotypes resulting from overexpression and ectopic expression. The use of such transposon or T-DNA libraries in concert with DNA arrays, expression proteomics, protein structural analysis, and protein interaction mapping provides an unprecedented integrated approach to the functional analysis of plants.

List of Technical Nomenclature

Activation tag	An insertional DNA element containing a strong, outward facing promoter. The function of such an element is to activate neighboring genes, causing dominant overexpression or ectopic expression phenotypes.
Adaptor oligonucleotide	A double-stranded oligonucleotide that fits over the sticky ends left by a restriction digest and provides a primer annealing site.
AIMS	Amplification of insertion mutagenized sites. Nested PCR using a transposon-specific primer for linear amplification.
Autonomous transposon	A transposon encoding its own transposase and able to self-mobilize.
Clean insertion	Insertion of one copy of a DNA element with minimal disturbance to surrounding DNA.
Complex insertion	Insertion of two or more copies of a DNA element and concomitant rearrangement of the surrounding DNA.
Cre recombinase	An enzyme from bacteriophage P1 that catalyzes recombination between short DNA sequences known as *loxP*.
Cut-and-paste	The movement of transposons by excision from the original site and insertion at a new site. Does not increase transposon number unless transposition occurs from replicated to nonreplicated DNA.
Enhancer trap	An insertional element containing a reporter gene downstream of a weak promoter, such as a naïve TATA box. When this inserts in the vicinity of an enhancer, the reporter gene is expressed with the same pattern as the gene normally activated by that enhancer.
Gene space	Gene rich regions of the plant genome.

Gene tagging Tagging an interrupted gene with the insertional DNA element so that it can be identified and cloned.

Gene targeting Introducing a specific change into a preselected gene by homologous recombination.

Gene trap An insertional element containing a reporter gene whose activity is dependent on insertion within a gene.

Homologous recombination Recombination occurring between two similar DNA sequences carried out by enzymes provided by the host plant.

Insertional mutagenesis Mutagenesis by gene disruption with a DNA sequence.

Inverse PCR PCR on a circularized template in which the primers face outward from an insertion sequence and amplify flanking DNA.

Local reinsertion The tendency of active transposons to transpose to sites linked to their origin.

Nested PCR Sequential PCRs in which one or both primers are represented by nested sets placed progressively further inward along the template to increase specificity.

Non-autonomous transposon A derivative of a full-length transposon lacking transposase, but able to mobilize in *trans* if transposase is supplied by an autonomous element in the same genome.

Plasmid rescue The inclusion of a plasmid origin or replication and antibiotic resistance marker in a transposon, so that the flanking genomic region can be isolated and cloned as a bacterial plasmid.

Promoter trap A gene trap activated by insertion downstream of an active promoter.

RACE Rapid amplification of cDNA ends. A PCR technique for the cloning of full length cDNA sequences.

Replicative transposition Transposition in which a copy of the element is left behind at the original site. Leads to increases in transposon number.

Retrotransposon A DNA element that mobilizes using an RNA intermediate.

Signature tagged mutagenesis The same as gene tagging. The insertional DNA element is the "signature" that tags the interrupted gene.

Splice trap A gene trap containing a splice acceptor site activated by incorporation into a transcriptional fusion product.

Starter line The first generation of plants carrying a newly introduced transposon or T-DNA sequence.

TAIL PCR Thermal asymmetric interlaced PCR. Similar to AIMS but using a short arbitrary primer rather than an adaptor specific primer.

Target site preference The tendency for transposons to preferentially integrate at certain target sites, often defined by a weak consensus sequence.

Transgenic plant A plant carrying extra (usually foreign) DNA in every cell.

T-DNA Transferred DNA, the segment of DNA on the Ti-plasmid of *A. tumefaciens*, which is transferred to the plant genome.

Transposase The DNA cutting and joining enzyme that allows transposons to move.

Transposon A DNA sequence that can jump from site to site in the genome.

Transposon display Nested PCR using a transposon-specific primer and a primer annealing to an oligonucleotide adaptor.

Two-component transposon system A system comprising an autonomous transposon and a defective transposon.

See also: **Crop Improvement**: Mutation Techniques. **Genetic Modification**: Gene Cloning, General Principles; Transformation in Dicotyledons; Transformation in Monocotyledons; Transformation in Plastids; Transformation, General Principles. **Growth and Development**: Control of Gene Expression, Post Transcriptional Regulation; Control of Gene Expression, Regulation of Transcription.

Further Reading

Azpiroz-Leehan R and Feldmann KA (1997) T-DNA insertion mutagenesis in *A. thaliana*: going back and forth. *Trends in Genetics* 13: 146–152.

Coelho PSR, Kumar A, and Snyder M (2000) Genome-wide mutant collections: toolboxes for functional genomics. *Current Opinion in Microbiology* 3: 309–315.

Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA, and Hamer JE (2001) Recent advances in large-scale transposon mutagenesis. *Current Opinion in Chemical Biology* 5: 67–73.

Jeon J-S, Lee S, Jung K-H, *et al.* (2000) T-DNA insertional mutagenesis for functional genomics in rice. *Plant Journal* 22: 561–570.

Kempin SA, Liljegre SJ, Block LM, *et al.* (1997) Targeted disruption in *Arabidopsis*. *Nature* 389: 802–803.

Krysan PJ, Young JC, and Sussman MR (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 11: 2283–2290.
Maes T, De Keukeleire P, and Gerats T (1999) Plant tagnology. *Trends in Plant Science* 4: 90–96.
Maize Targeted Mutagenesis Project (MTM). http://mtm.cshl.org.
Nottingham *Aribidopsis* Stock Centre. http://nasc.nott.ac.uk.
Osbourne BI and Baker B (1995) Movers and shakers: maize transposons as tools for analyzing other plant genomes. *Current Opinion in Cell Biology* 7: 406–413.
Parinov S and Sundaresan V (2001) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome-sequencing project. *Current Opinion in Biotechnology* 11: 157–161.
Parinov S, Sevugan M, Ye D, *et al.* (1999) Analysis of flanking sequences from *Dissociation* insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell* 11: 2263–2270.
Ramachandran S and Sundaresan V (2001) Transposons as tools for functional genomics. *Plant Physiology and Biochemistry* 39: 243–252.
The Maize Gene Discovery Project. http://www.zmdb.iastate.edu/zmdb/education/overview.php.
Tissier A, Marillonnet S, Klimyuk V, *et al.* (1999) Multiple independent defective *Suppressor-mutator* transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell* 11: 1841–1852.
US *Arabidopsis* Biological Resource Center, Columbus, Ohio. http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm.
Walbot V (2000) Saturation mutagenesis using maize transposons. *Current Opinion in Plant Biology* 3: 103–107.
Weigel D, Ahn JH, Blazquez MA, *et al.* (2000) Activation tagging in *Arabidopsis*. *Plant Physiology* 122: 1003–1013.

Transformation, General Principles

H D Jones, Rothamsted Research, Harpenden, UK

Introduction and Historical Perspective

Over a period of two decades, research in plant genetic transformation has progressed from proof-of-concept to a mature, platform technology in plant science. With the increasing availability of tissue-specific, developmentally regulated and constitutive promoters, reporter genes and techniques such as insertional mutagenesis and gene tagging, it offers many opportunities for both basic research in model species and for the modification of economically important traits in crops.

Genetic transformation involves the insertion of nucleic acid molecules, usually in the form of an expression cassette, into plant cells. These molecules are integrated, via illegitimate recombination by the cell's repair machinery, into the nuclear genome and can be either maintained in culture as a cell line, or regenerated into adult plants and transmitted via gametes into subsequent generations. Alternatively, the introduced DNA may be intended for short-term "transient" expression studies. Specialized transformation procedures involving homologous recombination can be used for targeting genes into the genome of plastids.

In the 1970s, progress in plant transformation lagged behind the successes in bacterial, yeast and animal systems. The early attempts to transfer DNA to plant cells targeted protoplasts with *Agrobacterium*, polyethylene glycol (PEG) or electroporation delivery methods. The first reports that foreign genes could be inserted and expressed in cultured plant cells, principally tobacco and petunia, came in the early 1980s. The development of protoplast-to-plant regeneration systems allowed the production of adult transformed plants that transmitted the inserted DNA into the next generation. By the mid-1980s, strategies for *Agrobacterium*-mediated T-DNA delivery and direct gene transfer for several species were available. Since then, many of the major model plant species, as well as human food, animal feed, fiber and ornamental crops, have been transformed by one route or another. This section describes the general principles of genetic transformation and should be read in conjunction with the other sections in this text that focus on specific applications, ethics, etc.

DNA Delivery and Integration

There are two fundamental ways to deliver DNA into the nucleus of target plant cells, either by "direct gene transfer" or by utilizing the capacity of the soil bacterium *Agrobacterium tumifaciens*, the causative agent of crown gall disease. Direct methods utilize a physical or chemical stimulus to deliver a naked, double-stranded DNA molecule into the target cell nucleus. In contrast, *Agrobacterium*-mediated gene transfer involves the transportation of single-stranded T-DNA, with associated virulence (vir) proteins, from bacterial to plant cell via specialized intercellular channels. Once in the plant cell nucleus, the T-DNA is made double-stranded after the displacement of bound vir proteins. The principles underlying these methods are discussed below. Regardless of the method used to deliver DNA into

the nucleus, the process of transgene integration into the plant genome appears to be similar. Molecular analysis of integration sites suggests that transgenes insert via double-stranded illegitimate recombination, utilizing the plant's own DNA repair machinery, at a single (or sometimes more than one) locus. The locus may contain one or multiple, concatameric copies of the transgene, which may have undergone rearrangements and/or may have generated short lengths of "filler" DNA homologous to flanking plant genomic DNA. *In situ* hybridization data suggest that physical transgene integration occurs at random within and between plant chromosomes; however, other analyses demonstrate a preference for actively transcribed regions in the plant genome with the possibility that native plant genes are disrupted in the process.

Direct Gene Transfer

A range of direct gene transfer methods have been used to generate transgenic plants, including microinjection, vortexing with silicon carbide whiskers, electroporation, particle bombardment, laser-mediated uptake, PEG-mediated uptake, etc. Each method has been used with particular species, explant sources and regeneration protocols, with varying degrees of efficiency and reproducibility. The most widely used and versatile direct delivery method is particle bombardment (or biolistics), particularly in cereal crops where *Agrobacterium* delivery has proved difficult. This procedure involves accelerating submicron gold or tungsten particles (microcarriers) coated with DNA into regenerable plant tissues. Parameters considered to have a marked effect on successful biolistic-mediated DNA delivery include (i) preparation of the microcarrier suspension, (ii) preparation and precipitation of DNA onto microcarriers, (iii) speed and spread of particles when they hit the explant, (iv) the choice of callus or regenerable plant tissue to be bombarded, and (v) the use of preculture or osmotic shock treatments. Hand-held and bench-top biolistic devices are used extensively for both stable transformation and for studies involving transient gene expression.

Agrobacterium-Mediated DNA Delivery

It was first suspected that *Agrobacterium tumifaciens* was the causative agent for crown gall disease in the early 1900s; however, it was not until the 1970s that the large (Ti) plasmid actually responsible for tumor induction was identified. Another key discovery 10 years later, that only the T-DNA portion of the Ti plasmid actually integrates into the plant genome, led to the development of disarmed (i.e. nononcogenic) *Agrobacterium* vector systems for transformation. Nowadays, the most widely used vectors are based on binary systems in which an expression cassette containing the gene of interest is cloned between the left and right border sequences of an artificial T-DNA. A plasmid containing this T-DNA is then transformed into competent cells of a suitable *Agrobacterium* strain already containing a disarmed "helper" Ti plasmid supplying the *vir* genes. Finally, *Agrobacterium* cells are co-incubated with regenerable plant tissues such as leaf disks, cotyledons, stem segments or callus suspension cultures to allow the bacterial cells to attach and form a transporter complex which mediates T-DNA import. This step can be aided by physical stimuli such as wounding, application of vacuum, or the addition of phenolic inducers such as acetosyringone.

In Planta Transformation versus Regeneration via Tissue Culture

Assuming DNA has been successfully delivered and integrated into the genome of a target plant cell, there remains a significant challenge to recover whole, fertile, nonchimeric plants. Most transformation systems require a tissue culture stage, which is labour- and materials-intensive and increases the chance of genetic instability due to unpredictable somaclonal variation. There has been considerable interest in various *in planta* transformation methods because of their simplicity and lack of any tissue-culture requirement. The basic principles of the two approaches are described below.

In Planta Transformation

This concept depends on the ability to introduce naked or T-DNA into gametes, pregametic tissue or zygotes around the time of fertilization. Attempts at injecting, electroporating or bombarding DNA into pollen, floral tissues or germline meristems of various species have proved to be generally unreproducible. However, one notable exception is the success of *Agrobacterium*-mediated germline transformation of *Arabidopsis*. This is now the universal method of choice for *Arabidopsis* transformation, and involves applying *Agrobacterium* directly to the mature flowers with the aid of a surfactant or vacuum. Seeds from the treated plants are germinated on selective media, and although efficiencies are low (less than 1%), the large number of seeds obtained, and ease of screening, make this a reproducible, facile method.

Regeneration via Tissue Culture

For species other than *Arabidopsis*, successful transformation depends on the remarkable plastidity and totipotency of plant cells to recover adult plants from transformed somatic cells. There are two principal routes by which plants can be recovered via tissue culture: somatic embryogenesis and organogenesis. Some species, such as sugar beet (*Beta vulgaris*), can be regenerated via either route, but with most species one prevails as the method of choice. Somatic embryogenesis is a nonsexual propagation process where somatic cells differentiate into embryo-like structures with shoot and root meristems. With appropriate phytohormones and other culture medium additions, somatic embryos can be "germinated" and give rise to viable adult plants. In most species, somatic embryos are derived indirectly from callus, cells, but in a few cases, somatic embryos can be produced directly with no callus phase, for example, from citrus (*Citrus* spp.) ovaries or cassava (*Manihot* spp.) leaves. Organogenesis refers to the ability of some plant tissues (e.g., hypocotyls, cotyledons, leaf bases or callus derived from them) to reorganize into shoot meristems, which can subsequently be rooted and generate complete plants. If successful rooting is not possible, then grafting onto a root-stock is possible. Selection pressure is applied at appropriate stages to inhibit the growth of nontransformed tissue. **Figure 1** summarizes the main tissue culture routes through which plants can be recovered, and the stages suitable for DNA delivery.

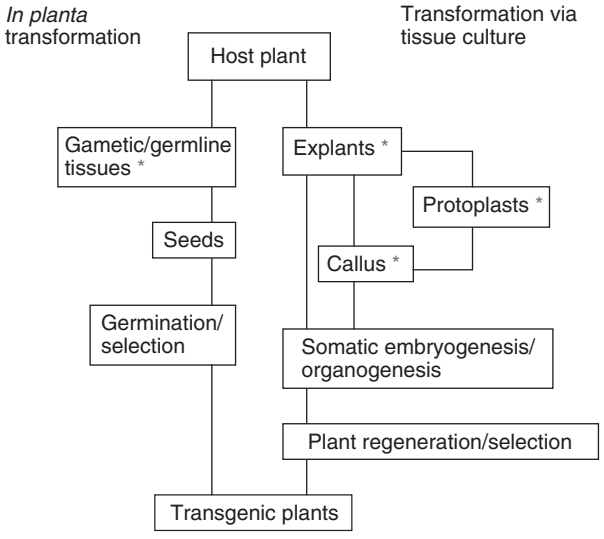

Figure 1 Principal stages in genetic transformation of plants. Asterisks denote tissues suitable for DNA delivery. The left-hand route denotes *in-plant*a transformation, and the right-hand route, transformation via tissue culture.

Occasionally, genetic instability may result in somatic cells as a consequence of cell/tissue culture, referred to as "somaclonal variation." This has been identified in many early tissue culture protocols, and has been widely exploited (e.g. in potato (*Solanum tuberosum*)) as a source of variation for breeding material. Factors that can promote genetic instability in micropropagated plant cells, such as high levels of 2,4-D and long culture periods, are now actively minimized in transformation methods.

Selection

Selectable marker genes are a vital part of most transformation protocols. They are delivered alongside the gene of interest, either on the same plasmid or on a separate plasmid. A wide range of selectable marker regimes is available and is particularly important in species where transformation efficiencies are low. Selectable marker genes can be categorized into those based on resistance genes that confer the ability to grow in the presence of toxic compounds such as antibiotics or herbicides which kill or otherwise compromise untransformed tissue (negative selection). Alternatively, a range of positive selection systems are available which provide transformed tissues with an enhanced ability to utilize, for example, an unusual carbohydrate or amino acid supply and thus enrich the culture for transformed tissue expressing the marker gene. Concerns about the environmental spread of selectable markers, particularly antibiotic resistance genes, from field-planted GM plants prompted the development of approaches to eliminate selectable markers. One concept is to co-transform with a gene of interest and a marker gene at nonlinked loci, thus the primary transformant contains both genes, but genetic segregation in the progeny plants allows identification of individuals lacking the marker but possessing the gene of interest. This concept can be augmented to include co-integrated marker and trait genes by incorporating a transposition system in the vector design. Thus, the marker gene can be separated from the gene of interest by transposition in later generations and then removed by segregation. Yet other approaches, such as *cre/lox* or *flp/frt*, use excision/recombination to physically remove selected sequences after transformation.

Predictability of Transgene Inheritance and Expression

In most cases, the transgene insertion is inherited by subsequent generations as a dominant trait in

predictable Mendelian ratios, depending on locus number and linkage (3:1 for a single locus, 15:1 for two unlinked loci, etc.). However, distorted inheritance ratios have also been observed, which may persist or revert to Mendelian in subsequent generations. Possible causes of distorted segregation patterns include chimerism in the primary transformant, counterselection of the chromosome containing the transgene, or disruption of plant genes involved in DNA replication, gametogenesis, fertilization, etc. In addition, variation in levels of gene expression can occur between individuals transformed with a given expression cassette. Factors that can influence this include copy number, integration pattern, and interactions between foreign and native DNA that vary depending upon the precise point of insertion (so-called "position effect"). The traditional strategy to overcome nonMendelian inheritance ratios and unpredictable levels of expression has been to create a large number of individual transgenic events, and then to select suitable individuals for further study. However, there is also considerable interest in modifications that increase the predictability of transgene expression. For instance, targeted homologous recombination in the nuclear genome has been demonstrated at low frequency in some species, and the use of matrix attachment regions to insulate the transgene may modulate variability of expression.

Applications of Transformation for Basic Research and Crop Biotechnology

The *Arabidopsis* genome has been completely sequenced, but analysis of the resulting DNA and protein sequences has failed to identify the function of most of the genes. The challenge of functional genomics is to ascribe functions to the vast amount of sequence data becoming available. Genetic transformation can be used for functional genomics research, both in forward genetics studies (to go from phenotype to sequence) and reverse genetics studies (from sequence to phenotype). Depending on the scientific rationale, a wide range of strategies involving transformation can be used, including (i) the overexpression or silencing of native or heterologous genes, (ii) the use of reporter genes under the control of specific promoters, either in stable or transient expression, and (iii) the generation of a randomly tagged population for gene or promoter identification. In addition to basic research, transformation is also a platform technology for the introduction of novel traits and improvement of commercial cultivars. As an adjunct to conventional plant breeding, transformation has the potential to increase crop yield, reduce chemical inputs, or provide the means by which pharmaceuticals or nutraceuticals can be made in "plant factories." The principle of these strategies is described below and summarized in **Figure 2**.

Overexpression/Gene Silencing

The most common application of genetic transformation is for the functional analysis of a cloned gene. This can be achieved by the overexpression of the sense strand to generate more protein, and/or down-regulation via, for instance, antisense RNA expression, co-suppression or other gene silencing mechanisms. Both these approaches can be "constitutive," or targeted to specific tissues or developmental stages by the careful choice of promoter (see below). The choice of background species or wild-type/mutant genotype provides another layer of experimental choice.

Targeting Gene Expression

Many biological processes are controlled, spatially and temporally, at the level of transcription. The core promoter, along with enhancer elements and other *cis*-acting regulatory sequences, are nontranscribed DNA sequences that form part of the mechanisms that regulate this gene expression. The choice of promoter is critical for efficient transgene expression, and the scope of experimental design is improved by the availability of a well-characterized "catalog" of promoters that can be used to "drive" target gene expression in particular organs or developmental stages of the species in question. The availability of well-defined inducible, tissue-specific or developmentally regulated promoters has been a bottleneck in the design of transformation experiments in the past, but screens of promoter-tagged populations have begun to identify more. The use of promoter::reporter gene fusions in transgenic plants continues to be a powerful tool in the study of regulated gene expression.

Reporter Genes

Reporter genes encode products that can readily be detected in transformed tissue, and are a powerful tool for the investigation of *cis*- or *trans*-acting gene regulatory factors in both transient and stable transformation experiments. In an ideal reporter system the gene product should be nontoxic, and the assay should be facile, *in situ*, nondestructive and quantifiable. Unfortunately, no single gene meets all of these requirements, each having specific strengths and limitations. Examples of commonly used reporter systems include GUS, GFP, luciferase and anthocyanin, and depending on the purpose of the

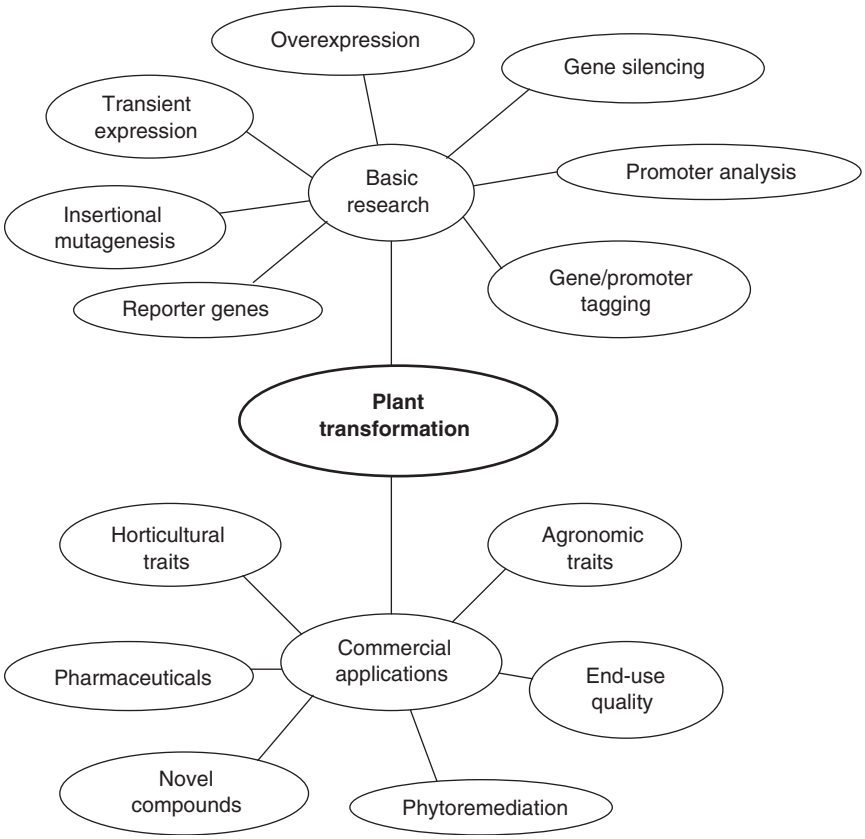

Figure 2 Applications of plant transformation: strategies for basic research and target traits for commercial applications.

experiment, these may be used as a transcription tag, a transcriptional fusion or a translational fusion. Reporter genes have been successfully used to characterize native and heterologous gene expression as well as protein trafficking.

Insertional Mutagenesis

The disruption of a random gene by the insertion of a known segment of DNA, and the subsequent observation of a mutant phenotype, has been used to great effect in several plant species. The inserted sequence not only knocks out or modifies the expression of the host gene, but also provides a known "tag" in the gene that facilitates the cloning and sequencing of flanking sequences, and therefore the identification of the disrupted gene. In species where the generation of large numbers of primary transformants is facile, particularly *Arabidopsis*, direct T-DNA insertions can be used. T-DNAs are thought to insert randomly within the genome, and it has been estimated that 100 000 independently transformed lines would be more than enough to saturate the *Arabidopsis* genome. Large T-DNA collections have been generated and systematic "reverse genetic" screens attempted. In other species where transformation is more difficult, transposons can be utilized. Transposons are short mobile genetic elements that can jump from one locus to another within the plant genome. The most commonly used transposons originally came from corn (*Zea mays*; maize), e.g. *Ac-Ds* (Activator-Dissociation), although other species, such as petunia (*Petunia* spp.) and snapdragon (*Antirrhinum* spp.), also possess several endogenous transposons. The excisable element can, for example, be engineered to flank a conventional expression cassette containing a scorable or screenable marker, and used to create a large number of tagged knockouts in a heterologous plant host by transformation and transposition.

Gene/Promoter Tagging

A major limitation of insertional mutagenesis is that it fails to generate a phenotype in the many genetically redundant (although functional) genes that exist in any organism. In addition, there is considerable interest in isolating regulatory sequences with specific expression patterns for use in studies involving heterologous tissue-specific or developmentally regulated overexpression. Various alternative approaches to highlight such genes and promoters are

enhancer traps, promoter traps, gene traps and activation traps. These methods rely on either the activation of introduced marker genes by native *cis*-acting regulatory sequences, or the activation of native genes by introduced regulatory sequences. An elegant extension of the enhancer trap approach is a two-component system incorporating GAL4 fused to a minimal promoter which, in turn, drives GFP expression under the control of the GAL4 activating sequence. Work in *Arabidopsis* has shown that lines with highly specific patterns of GPF expression can be obtained, and that other genes fused to the GAL4 activating sequence show the same patterns when transformed into these lines.

Transient Expression

The production of stably transformed plants is not a trivial process and some biological questions can be usefully addressed by using short-lived "transient expression" from noninherited, extrachromosomal foreign DNA sequences. Transient expression typically lasts for 1–7 days and can be induced from conventional transformation vectors or from viral vectors that are capable of systemically spreading through the plant after mechanical inoculation. It can be used as a rapid screen for assessing promoter activity, or comparing the effects of induced point mutations in a gene, etc. Facile cloning methods to generate viral vectors make large-scale screening of whole cDNA libraries possible.

See also: **Ethics and Biosafety**: Development and Commercialization of Genetically Modified Plants; Ethics of Genetically Modified Crops. **Genetic Modification**: Transformation in Plastids. **Tissue Culture**: General Principles. **Tissue Culture and Plant Breeding**: Somaclonal Variation.

Further Reading

Barcelo P, Rasco-Gaunt S, Thorpe, C, and Lazzeri PA (2001) Transformation and gene expression. *Advances in Botanical Research Incorporating Advances in Plant Pathology* 34: 59–126.

Galun E and Breiman A (1997) *Transgenic Plants*. London: Imperial College Press London.

Gelvin SB and Schilperoort RA (eds) (2000) *Plant Molecular Biology Manual*. Dordrecht: Kluwer Academic Publishers.

Hall RD (ed.) (1999) *Plant Cell Culture Protocols*. Totowa, NJ: Humana Press.

Hansen G and Wright MS (1999) Recent advances in the transformation of plants. *Trends in Plant Science* 4: 226–231.

Lindsey K (ed.) (1998) *Transgenic Plant Research*. New York: Harwood Academic.

Potrykus I and Spangenberg G (eds) (1995) *Gene Transfer to Plants*. Berlin: Springer.

Reynolds PHS (ed.) (1999) *Inducible Gene Expression in Plants*. Wallingford: CABI Publishing.

Springer PS (2000) Gene traps: tools for plant development and genomics. *Plant Cell* 12: 1007–1020.

Zupan J, Muth TR, Draper O, and Zambryski P (2000) The transfer of DNA from *Agrobacterium tumifaciens* into plants: a feast of fundamental insights. *Plant Journal* 23: 11–28.

Transformation in Dicotyledons

D J Murphy, University of Glamorgan, Cardiff, UK

The production of transgenic crops relies on three distinct technologies:

1. The isolation and manipulation of small sequences of DNA (often called gene cloning).
2. The insertion of such exogenous DNA sequences into plant tissues (i.e., transformation).
3. The selection of transformed cells or tissues and their regeneration into adult plants capable of further propagation.

Early attempts to transform plants with exogenous DNA focused on the integration of naked DNA, normally in the form of plasmids, into isolated plant cells. Typically, the cells were pretreated with cell wall digesting enzymes to create protoplasts before being shocked, e.g., by electrical treatments or reagents like EDTA, in order to allow the DNA to enter transient pores in their plasma membranes. However, these methods proved to be relatively inefficient and unreliable. The real breakthrough in plant transformation technology came with the introduction, in the early 1980s, of *Agrobacterium* vector systems. The great advantage of *Agrobacterium* vectors was that they not only delivered DNA into plant cells at reasonably high efficiency, but they also facilitated the incorporation of the exogenous DNA into the genome of the host plant cell.

For most of the first decade or so of plant transformation research and development, applications tended to be limited to dicotyledenous crops. There were two main reasons for this; firstly, the natural host range of the common *Agrobacterium* vectors was restricted to certain dicotyledons; and, secondly, the easiest plants to regenerate from transformed cells of tissue slices were also dicotyledons. Hence, much of the early work on plant transformation was concentrated on Solanaceous

crops such as tobacco (*Nicotiana tabacum*), tomato (*Lycopersicum esculentum*), and potato (*Solanum tuberosum*). Later, in the 1980s and early 1990s, other important dicotyledonous crops, especially many of the brassicas, were successfully transformed, although their transformation is often highly genotype dependent. In practice, this means that one particular variety of a brassica species may be used for the initial transformation – in the case of oilseed rape (*Brassica napus*; canola), this is often the variety "Westar" – but the transformed plants then have to be backcrossed into more agronomically suitable elite breeding lines for commercial production. Hence, the process can be slower and more costly than if the elite lines themselves could be used for direct transformation.

Nowadays, the use of improved strains of *Agrobacterium* and the refinement of direct DNA delivery via projectile mediated techniques (biolistics) means that virtually all dicotyledonous species can be transformed with reasonable efficiency. Nevertheless, significant challenges remain before the transformation of dicotyledons is optimized, with regard to ease, efficiency and predictability. Transformation of plants is still fairly crude and inefficient compared with, for example, the transformation of microbes or animals. Transgenes in plants insert more or less randomly into the genome, the number of copies inserted can range from one to a dozen or more, and the introduced DNA sequences are frequently modified and/or fragmented in the process. For example, pieces of DNA may lose part of one or both of their end regions, or they may be cleaved at several points and the resulting fragments inserted into different parts of the genome. Both the position of insertion in the genome and the copy number of a transgene can significantly affect its expression in the resulting plant. The presence of multiple copies of a transgene can result in instability of its expression. Finally, transgenes are normally inserted into the recipient genome as part of a multigene construct that also contains regulatory elements and a selectable marker, often an antibiotic or herbicide resistance gene. There is already public concern about the use of antibiotic resistance markers, and many researchers and breeders are also concerned that, as transgenic crops become more widespread, they will inevitably be subject to many further rounds of transformation as additional genes are inserted in order to keep improving the crop. The accumulation of many different selectable marker genes would soon present problems as breeders ran out of new benign markers for further rounds of transformation. The continued use of the same marker can lead to cosuppression and the loss of expression of the transgene itself, and also of genes with related sequences.

Although there are important challenges in optimizing transformation technologies, there has been much progress in addressing these challenges in research laboratories around the world. These include the development of excisable markers, the use of plastid transformation to achieve precise transgene insertion, and the use of novel selection systems that do not involve antibiotic resistance genes. In the next decade, dicotyledon transformation should become much more reliable, as well as cheaper and easier. This may facilitate its uptake in developing countries where, due to its cost and complexity, the current transformation technology remains largely out of reach of their scientists.

See also: **Genetic Modification**: Gene Cloning, General Principles; Transformation in Monocotyledons; Transformation in Plastids; Transformation, General Principles.

Further Reading

Bent AF (2000) *Arabidopsis* in planta transformation. Uses, mechanisms, and prospects for transformation of other species. *Plant Physiology* 124: 1540–1547.

Birch RG (1997) Plant transformation: problems and strategies for practical application. *Annual Review of Plant Physiology and Plant Molecular Biology* 48: 297–326.

Crispeels MJ and Sadava DE (2003) *Plants, Genes and Crop Biotechnology*. Sudbury, MA: Jones & Bartlett.

Daniell H, Muthukumar B, and Lee SB (2001) Marker free transgenic plants: engineering the chloroplast genome without the use of antibiotic selection. *Current Genetics* 39: 109–116.

De la Riva GA, González-Cabrera J, and Vázquez-Padrón R (1998) *Agrobacterium tumefaciens*: a natural tool for plant transformation. *Electronic Journal of Biotechnology* 1: no. 3: http://www.ejb.org/content/vol1/issue3/full/1/#51

Newell CA (2000) Plant transformation technology: Developments and applications. *Molecular Biotechnology* 16: 3–65.

Transformation in Monocotyledons

P A Lazzeri and P Barcelo, DuPont Wheat Transformation Laboratory, Barcelona, Spain

Introduction

Many of the world's most important crop plants are monocotyledonous. These include temperate and

tropical cereals, which are the single crop group most exploited for human and livestock nutrition, and other major groups of crops, such as forage and turf grasses, palms, and vegetable and ornamental bulb species. Thus, there are many opportunities for applying transformation technology in monocotyledon crop improvement, both as a basic research technique and as a tool in monocotyledon crop breeding.

The first demonstrations of plant transformation were made in the early to mid 1980s using dicotyledon species and soon after this attention turned to the major cereal crops such as corn (*Zea mays*; maize), wheat (*Triticum* spp.), and rice (*Oryza sativa*). However, it quickly became apparent that the early transformation techniques, which were based either on *Agrobacterium* as a vector or on DNA uptake into protoplasts, worked well in dicotyledon species such as tobacco (*Nicotiana tabacum*) and *Petunia*, but did not function in monocotyledon species. Monocotyledons, and the cereals in particular, were much more difficult to culture and regenerate *in vitro* and were not naturally susceptible to infection by *Agrobacterium*. Thus began more than a decade of research aimed at the development of efficient transformation technology applicable in monocotyledon crop species. This work, which focused on overcoming the recalcitrance of cereal species to transformation, resulted in the invention of a range of novel tissue culture and gene transfer methodology.

Owing to the great economic importance of the major monocotyledon crops, much of the research on transformation technology has been done in industry, although academic scientists have also made major contributions. The field has progressed very quickly, so that the time span between the development of new transformation methods and their application to produce genetically modified (GM) commercial varieties has only been a few years. Today, routine transformation methods are available for all the major cereals and grasses and most of the other important monocotyledon crops can be transformed, although bulb species such as onion (*Allium cepa*) and garlic (*Allium sativum*) are still recalcitrant and further development in the technology is required.

In this article, the transformation procedures commonly used for monocotyledons are introduced; the target tissues, culture and selection systems, and the gene transfer technologies available are considered. The vectors used as vehicles for DNA delivery and the way in which transgenes are integrated into the host genome, are expressed, and inherited are examined. Finally, the status of application of transformation in monocotyledon crops and the prospects for future developments in this field are discussed.

Transformation Target Tissues and Culture Systems

The basis of all transformation systems is the integration of a new gene into the genome of a recipient cell, followed by the recovery of a whole plant from the transformed cell. Under normal circumstances, there are a limited number of cells in a plant that can give rise to new plants. These are the germline cells in meristems, which form floral organs, or shoot meristems, which may be induced to root and thus produce new plantlets. Efforts have been made to develop transformation systems that deliver genes into germline cells in whole plants, but until the very recent development of inflorescence infiltration techniques there had been little success (see section below). Therefore, the majority of established transformation techniques target cells in cultured tissues, which are more accessible to gene delivery.

There are clear differences between the response to tissue culture of monocotyledon and dicotyledon plants. In dicotyledon species, it is generally possible to take explants from a range of different organs, such as leaves, stems, or floral parts, and to put these into culture and regenerate plants. In contrast, in monocotyledons, a much more limited range of tissues may be used to establish cultures and regenerate new plants. With few exceptions, these explants are taken from immature plant parts whose cells are either still dividing or have recently ceased division. The explants that are most frequently used for monocotyledon transformation systems are immature embryos and inflorescences, immature leaf tissue, isolated shoot meristems and cultured anthers or microspores (see **Table 1**). The use of immature tissues as transformation targets has the disadvantage that a constant supply of donor plants must be grown in order to supply explants. In addition, the physiological status of the donor plants has a major effect on the response of explants in culture and to achieve efficient and reliable transformation it is necessary to grow these plants under controlled environment conditions and free from pathogens. This, together with laborious manual isolation of explants, makes the production of target tissue costly. To avoid these constraints, transformation techniques targeting cultures derived from mature seeds have recently been developed for rice and there is interest in applying this simple approach in other monocotyledon species.

Table 1 Target tissues and culture systems used for monocotyledon transformation

Target tissue	*Advantages*	*Disadvantages*
Microspore	May target haploid cells, many target cells available	Inefficient transformation and regeneration
Immature embryo	Good target tissue, efficient plant regeneration	Need constant supply of donor plants
Immature inflorescence	Good source of target tissue	Regeneration less efficient than from embryos
Seedling leaf base	Good source of target tissue from young seedlings	Low efficiency transformation and regeneration
Shoot meristem	Good source of target tissue from young seedlings	Low efficiency transformation and regeneration
Embryogenic callus	Good target tissue, callus lines may be maintained	Difficult to establish lines with good regeneration capacity
Cell suspension	Renewable tissue source, good target for transformation	Regeneration efficiency often low, abnormal regenerants
Protoplast	High efficiency transformation possible	Regeneration often very difficult, abnormal regenerants common

A further aspect of the choice of target tissue for transformation is that it is possible to deliver the transforming DNA into either primary or secondary cultures. Primary cultures are the original explants isolated from the donor plant, which may be targeted immediately after isolation, or cultured for one to several days before transformation. Secondary cultures are tissues derived from proliferation and growth of the original explant, which have been subcultured one or more times. Examples of secondary cultures are callus lines, propagated meristem cultures, cell suspension cultures, and suspension cell protoplasts. Secondary cultures have the advantages that they can provide large numbers of target cells from the proliferation of the original explant tissue and it may be possible to visually select tissue types or cell lines that are amenable to transformation and regeneration. However, in many monocotyledons it is difficult to establish secondary cultures that retain a good capacity for plant regeneration. This problem is most acute in long-term secondary cultures, which typically accumulate genetic aberrations such as chromosome number changes, so that plants regenerated from them are abnormal.

Gene Delivery Methods

There are two systems employed for transformation of plants: the use of *Agrobacterium* as a biological gene transfer vector, or the use of physical or chemical methods to effect DNA delivery, termed direct gene transfer (DGT) techniques. Most monocotyledon species are naturally recalcitrant to infection by *Agrobacterium* and early research attempting to transform monocotyledons using techniques developed in dicotyledon species showed little success. As a result, emphasis was given to using DGT techniques for these species.

The transformation techniques applied in monocotyledons are evaluated in **Table 2**. As the table shows, a number of DGT techniques have been used to achieve transient gene expression or occasional stable transformation in monocotyledons, but only four techniques are used routinely for the production of transgenic plants. These techniques are protoplast transformation, tissue electroporation, silicon carbide fiber vortexing, and particle bombardment.

Protoplast Transformation

This was the method used to generate the first transgenic monocotyledon plants. It is based on the use of chemical or electrical stimuli to cause the uptake of DNA into protoplasts derived from suspension cultures. While the technique can give high transformation efficiencies, it has the disadvantages that suspension cultures are difficult to produce in most monocotyledons, regeneration from protoplast cultures is poor and regenerants may be variant, and protoplast transformants often have complex integration patterns, leading to unreliable transgene expression (see section below).

Tissue Electroporation

This transformation method uses high-voltage electrical pulses to create temporary pores in the cell membranes of tissue pieces suspended in a buffer containing DNA. This allows DNA uptake into the cells, which are subsequently cultured and regenerated into plants. The method has been shown to function in several monocotyledon species, but has only been used for reproducible production of transgenic plants in rice and corn.

Silicon Carbide Fiber Vortexing

In this technique, plant tissue and microscopic silicon carbide fibers are vortexed together in a buffer

Table 2 Gene transfer techniques and evaluation of their use in monocotyledons

Technique	*Evaluation*
Imbibition of tissues in DNA solutions, injection of DNA into plant organs	Occasional transient gene expression, no confirmed stable transformation
Ultrasound treatment of tissues in DNA-containing buffer solution	Transient gene expression, stable transformation not confirmed
Laser-mediated puncture of plants cells in DNA-containing buffer solution	Transient gene expression, stably transformed callus but no transgenic plant
Microinjection of cells with DNA solution	Transient gene expression, transgenic plants produced in limited species, but technique technically very difficult
Vortexing of tissues with silicon carbide fibers in DNA-containing buffer	Transgenic plants produced from suspension cells in a few species, but not suited to multicellular explants
DNA uptake into protoplasts	Transgenic plants produced in cereals, forage grasses, and other selected monocotyledon species. Technically demanding, plants recovered may be variant
Electroporation of cells or tissues	Transgenic plants produced in several species, but with limited transformation efficiency. Applicable to particular cell/tissue types only
Particle bombardment	DNA delivery to all cell types in all plants. The standard transformation technique for most monocotyledon species; widely applied in commercial GM crop breeding
Agrobacterium transformation	Applied in the major cereal species; use being extended to other monocotyledons, but requires specific culture types. Can be highly efficient in amenable varieties. Used in commercial GM crop breeding

containing DNA. The cells are punctured by the fibers allowing DNA uptake and transformation. The technique is most efficient when applied to suspension culture cells, which have thin walls and present a high surface area for targeting, and is less suitable for transforming multicellular explants. It has been applied in monocotyledons, such as corn and rice in which regenerable suspensions are available.

Particle Bombardment

The development of particle bombardment technology first made the routine transformation of monocotyledon species possible. This is still the method most widely used for monocotyledon GM, although in cereals such as rice, corn, and barley (*Hordeum vulgare*), *Agrobacterium* techniques are being used increasingly. In particle bombardment, dense microscopic particles, usually of gold, have DNA precipitated onto them and are then accelerated into plant cells. DNA is subsequently released from the particles, allowing transformation to take place. A number of different particle acceleration systems have been developed, using explosive discharge, air, gas, or electrostatic propulsion, but the standard device used in most laboratories is the PDS-1000He particle gun, which uses helium gas as propellant (see **Figure 1**). A major advantage of particle bombardment for gene delivery is that it is essentially species and tissue independent, although the recovery of

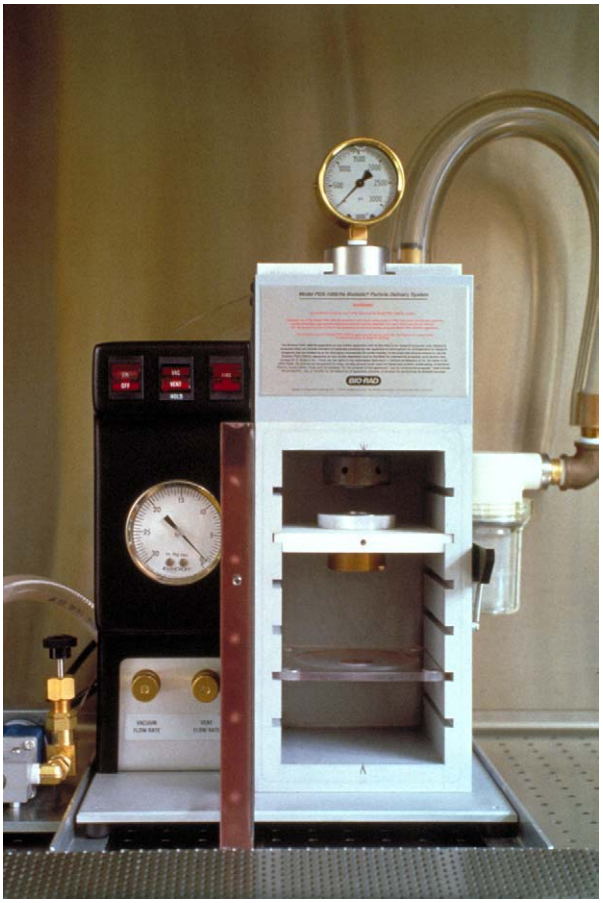

Figure 1 The PDS 1000 helium-powered particle gun, the instrument most commonly used for DGT transformation of monocotyledons. Photo courtesy of Huw Jones.

transgenic plants obviously requires the ability to culture and regenerate the targeted cells.

Agrobacterium Transformation

After more than a decade of research, the breakthrough in the development of viable *Agrobacterium*-mediated transformation in monocotyledons came when it was shown in 1994 by Hiei and colleagues that japonica rice callus cultures could be transformed at high efficiency using bacterial strains containing highly active virulence genes. This research, and following studies, demonstrated that although monocotyledon cells do not exhibit the wound response processes associated with the *Agrobacterium* interaction in dicotyledons, cultured monocotyledon cells are susceptible to infection and the gene transfer mechanism functions as in normal host plants. Following the *Agrobacterium*-mediated transformation of rice, the technique has been applied in several other cereals, including corn, barley, wheat, and sorghum (*Sorghum* spp.) It seems that when the correct type of culture system can be developed, which is typically an actively developing embryogenic callus culture, and these cultures are exposed to *Agrobacteria* containing appropriate virulence functions, transformation is possible in a number of monocotyledon species. Under optimal conditions, *Agrobacterium* transformation in monocotyledons can be highly effective and up to 25–50% of treated calluses may yield transgenic plants in amenable varieties of rice and corn. Although the mechanisms of gene delivery by *Agrobacterium* and in particle bombardment are clearly quite different, it is notable that the same monocotyledon species and varieties are either amenable or recalcitrant to transformation by either method; for example, japonica rice varieties are good subjects, while bulbaceous monocotyledons, such as onion or tulip (*Tulipa* spp.), are difficult targets. This suggests that DNA delivery is not the only limiting factor and that the capability for manipulation and regeneration *in vitro* is of primary importance. In this context, the recent demonstration that rice may be transformed by *Agrobacterium* infiltration into developing inflorescences (panicles) of intact plants, a technique first developed in *Arabidopsis*, is a significant advance, as it avoids the need for tissue culture.

A general feature of *Agrobacterium* transformation, in comparison with DGT techniques, is that transgene insertions tend to be simple, with low transgene copy numbers. This is a positive characteristic for the reliability of transgene expression and for handling transgenic lines in breeding programs. In consequence, there is general interest in moving toward *Agrobacterium* techniques for commercial transformation of monocotyledon crop species.

Transformation Vectors for Gene Delivery

The primary function of a plant transformation vector is to deliver genes into a recipient genome. The "active components" of a transformation vector are similar, regardless of the transformation method (i.e., *Agrobacterium* or DGT), and are promoter, structural gene, and terminator sequences. The promoter is responsible for controlling expression of the gene, the structural gene codes for the new protein/enzyme, and the terminator sequence terminates transcription of the transgene unit. The promoter–gene–terminator unit is cloned in this order into a DGT or *Agrobacterium* plasmid DNA molecule, whose features and functions are detailed below (**Figure 2** shows the structures of typical transformation vectors).

Vectors for DGT transformation are based on circular DNA plasmids and have three major features. The first feature is a sequence that allows the insertion (cloning) of the promoter–gene–terminator sequences described above. This feature is termed a multiple cloning site, which includes target sites for several "rare-cutter" restriction enzymes ideal for cutting and pasting desired DNA sequences into the site. The second and third features of the vector allow it to be propagated (amplified) to provide the relatively large quantities of DNA needed for DGT transformation procedures (for example, a single particle gun discharge may use 0.3 μg of DNA and an experiment may comprise 20 "shots"). For plasmid amplification, an origin of replication of the plasmid in the host bacterial cells is required. The third feature is a bacterial selectable marker gene, to ensure only bacteria containing the transformation vector are proliferated. The most common marker genes confer resistance to antibiotics such as ampicillin. There is concern about the incorporation of antibiotic marker genes in DGT vectors as they will subsequently be present in the genome of transgenic plants. It has been hypothesized (but not demonstrated) that horizontal gene transfer might occur, leading to their appearance in microorganisms in the wider environment. This has motivated changes in the design of transformation vectors. One strategy substitutes antibiotic resistance genes for other markers such as genes complementing a metabolic mutation in the bacterial genome, so that only bacterial cells containing the complementary gene are able to grow on minimal culture medium. A second strategy uses restriction enzymes to isolate the functional promoter–gene–terminator fragment from

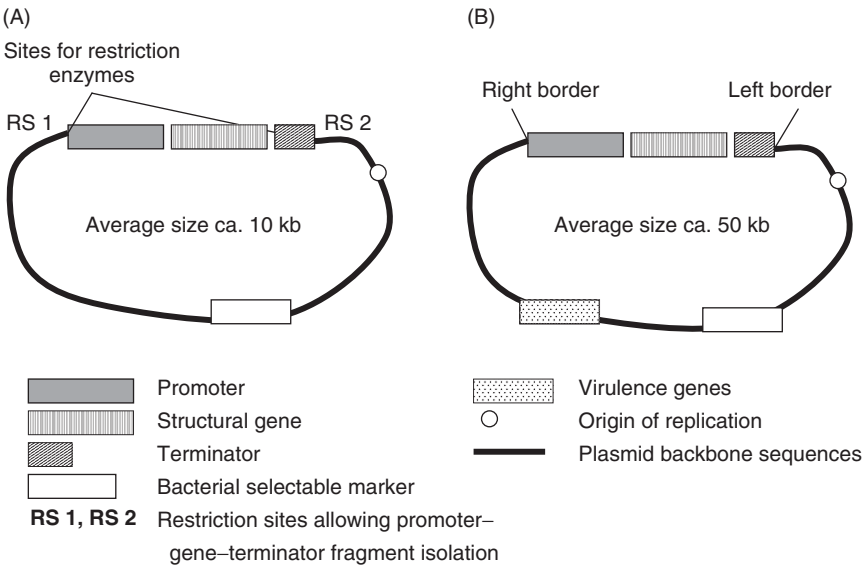

Figure 2 Structure of typical vectors used for (A) direct gene transfer (DGT) or (B) *Agrobacterium* transformation of monocotyledons.

isolated transformation plasmid DNA. The isolated fragment preparation is then used for transformation. This strategy has the advantage that strictly defined sequences are integrated into the target genome, avoiding all nonessential DNA.

Vectors for *Agrobacterium* transformation have a more complex structure than DGT plasmids, as apart from the basic transgene they also contain sequences involved in the mechanism of transfer from the bacterial cell to the plant cell and in integration into the plant genome. One feature of *Agrobacterium* transformation vectors (T-DNA vectors) is a pair of short palindromic sequences ($\sim$25 bp), the so-called "right" and "left" borders, between which the promoter–gene–terminator sequences described above can be cloned. The vector then operates by a set of *Agrobacterium* genes called virulence genes, recognizing the T-DNA borders, cutting the sequence at these points and packing, transferring, and integrating the DNA sequences cloned between them into the recipient plant genome. The virulence genes can be located either on the same T-DNA plasmid or on another plasmid molecule present in the bacterium. Monocotyledon species have been considered recalcitrant to *Agrobacterium* transformation and this only became routine when "supervirulent" plasmids containing extra copies of the virulence genes (e.g., *vir*G, *vir*C and *vir*B) were used for transformation. A last important difference between DGT and *Agrobacterium* transformation vectors is that the latter have a much larger "payload" of DNA, which means that they are more suitable for the transfer of multiple genes in a single cassette.

Selection Systems

Most plant transformation systems target cells randomly and result in the production of a few transgenic cells in a population of nontransformed cells. In order to identify these cells and recover transgenic plants from them, there is the need for a selection system. This can function at the level of visual selection, in which a gene whose expression gives a visual phenotype, may be used to track transformed cells as they divide and regenerate (**Figure 3** illustrates the use of the beta-glucuronidase (GUS) marker gene to visualize transgene expression). However, in most cases, it is necessary for the transformed cells to show preferential growth over nontransformed cells, as it is otherwise impractical to recover transformants from an overwhelming excess of nontransgenic cells. This is particularly important in monocotyledon transformation systems, which tend to show relatively low transformation efficiencies.

The most widespread selection strategy is the use of a resistance (marker) gene, which gives transformed cells the ability to grow in the presence of a toxic selection agent, while wild-type cells are inhibited or killed. The first selection genes applied in monocotyledon transformation were antibiotic resistance genes, used in combination with antibiotics such as kanamycin or hygromycin (see **Table 3** for examples of selection systems used in monocotyledon transformation). However, these genes were not fully satisfactory, as monocotyledons were found to be relatively resistant to most antibiotic selection agents and there were concerns over the release of antibiotic resistance genes from GM cereal

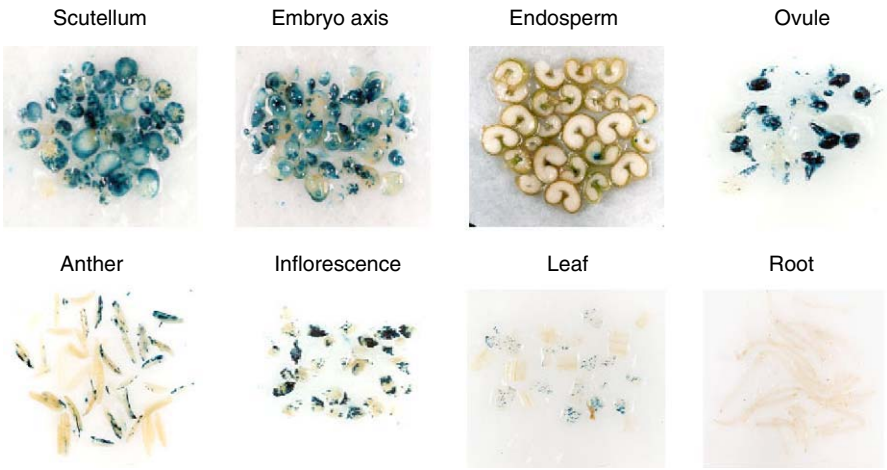

Figure 3 Expression of the GUS (beta-glucuronidase) marker gene in transgenic wheat tissues. Blue coloration indicates expression of the GUS transgene.

Table 3 Examples of selection systems used in monocotyledon transformation

System type	*Marker gene*	*Selection agent*
Antibiotic	*neo*/*nptII* (neomycin phosphotransferase)	Kanamycin, geneticin, paromomycin
Cytotoxin	*hpt* (hygromycin phosphotransferase)	Hygromycin
Cytotoxin	Mutant *dhfr* (dihydrofoliate reductase)	Methotrexate
Antibiotic	*aadA* (spectinomycin resistance)	Spectinomycin
Herbicide	*bar*/*PAT* (phosphinothricin acetyl transferase)	Phosphinothricin (PPT), glufosinate (Basta^{TM}/ Liberty^{TM}), bialaphos
Herbicide	*EPSPS* (enolpyruvate shikimate phosphate synthase) + *GOX* (glyphosate oxidoreductase)	Glyphosate (Roundup)
Herbicide	*ALS* (acetolactate synthase)	Sulfonylurea herbicides (e.g., Rimsulfuron, Nicosulfuron)
Positive selection	*manA* (phosphomannose isomerase)	Mannose
Positive selection	*xylA* (xylose isomerase)	Xylose

crops. In consequence, attention moved to herbicide resistance genes, in particular, the *bar* or *PAT* gene, encoding the enzyme phosphinothricin acetyl transferase. This gives resistance to phosphinothricin (PPT) and related compounds, the active ingredient in herbicides such as Basta^{TM} (Liberty^{TM}) and bialaphos. Other herbicide marker genes that have been used with success in monocotyledons are *EPSPS* + *GOX* and *ALS*, giving resistance to glyphosate (Roundup) and sulfonylurea herbicides, respectively (see **Table 3**).

A more recent development in selection technology is the use of positive marker genes, which give transformed cells a metabolic and hence growth advantage over nontransgenic cells. A system that has been used with success in several monocotyledon species, including wheat and corn, is the phosphomannose isomerase gene, which allows transgenic cells to use mannose as carbon source. This, and similar positive selection genes (see **Table 3**) have the advantage for practical application in GM agriculture that they may be viewed as being more acceptable in terms of potential health and environmental impact than antibiotic or herbicide resistance markers.

Transgene Integration, Expression and Stability

Following transgene delivery, two further steps are required: first, the foreign DNA has to be introduced into the plant cell nucleus; and, second, the DNA must be integrated into the plant genome. The major difference between the two transformation methods lies in the way in which foreign DNA is prepared and introduced into the nucleus. In the case of DGT methods, foreign DNA is introduced into the nucleus as a naked DNA molecule. In contrast, in *Agrobacterium*-mediated gene transfer, a T-complex, consisting of a T-DNA covered with virulence proteins, is produced, which is transported from the bacterium

into the plant cell through a channel assembled by other virulence proteins. The T-complex is then targeted to the nucleus.

Independent of the DNA delivery method, transgenes are usually integrated into the plant genome by illegitimate recombination at one or two independent locations (loci) in the genome. At the transgenic locus, transgene units (promoter–gene–terminator) can integrate completely and properly with no truncations or can be integrated as rearranged/truncated insertions. Low number and simple transgene insertions are most desirable for subsequently predictable transgene expression, as complex patterns of integration tend to lead to transgene instability and silencing.

The number of insertions of the transgene and the complexity of these insertions can be highly variable between transformation methods, experiments, and among plant species. In general, protoplast-based methods tend to produce transformants with larger numbers of insertions than other DGT methods, such as particle bombardment and also *Agrobacterium*-mediated transformation, which tends to produce the lowest insertion numbers. Considering transgene expression, once transgenes are integrated into the genome, the single most important factor controlling and regulating transgene expression is the promoter sequence controlling the transgene. These DNA sequences are responsible for the level as well as for the spatial and temporal expression of a given gene and are crucial for expressing genes in the desired manner.

There are three major classes of promoters currently used in monocotyledon transformation; those able to drive expression constitutively, those driving expression specifically in certain tissues and at certain developmental stages, and those only able to drive expression under inductive conditions. Constitutive promoters are mainly used to drive expression of selectable marker genes for the identification of transgenic tissues *in vitro*; this is generally achieved by using promoters with strong and ubiquitous activity. The most important promoter for monocotyledon transformation has been the corn ubiquitin promoter, which functions well across a range of species. This promoter, in common with several other highly active promoters used in monocotyledons, contains a 5′ intron sequence. This sequence is an integral part of the promoter unit, since its removal greatly reduces the activity of the promoter.

Tissue-specific and developmentally regulated promoters allow the expression of transgenes only in specific tissues or under certain developmental conditions, leaving the rest of the plant unmodified by transgene expression. For engineering monocotyledon crops, there are many circumstances in which such type of regulation is required; e.g., genes involved in grain quality may only need to be expressed in endosperm cells. In cereals, in particular, promoters have been characterized for a number of tissues, such as endosperm, embryo, anther tapetum, meristem, mesophyll, and phloem.

A third class of promoters is those that are only expressed under specific induction conditions. To date, relatively few promoters in this category have been characterized in monocotyledons, but sequences regulated by the action of stresses, such as pathogens or wounding, or by the action of chemicals, e.g., ethanol- and safener-induced promoters, are available.

The stability with which transgenes are inherited and expressed over generations is of central importance in plant transformation. While transgene instability has been described in many plants, including monocotyledon species, a general rule is that lines containing simple, low copy number transgene insertions show relatively stable and predictable transgene expression. Poor or unstable expression is, in many cases, related to plants that do not contain an intact functional copy of the transgene, or which contain multiple and complex insertion patterns. While the underlying mechanisms are still not fully understood, inactivation of the transgenes often involves methylation or the inactivation of plant defense mechanisms, which degrade DNA sequences identified as "foreign." In this context, heterologous sequences from distant species may be more likely not to function properly in the recipient genome than sequences from related (i.e., other monocotyledon) plants.

One aspect of transgene stability in transgenic crops concerns the potential for unwanted transfer of transgene sequences from one crop variety into another variety or related species via cross-pollination, or into another plant or microorganism by a horizontal gene transfer mechanism. Discussion has focused on the problems that could arise from the spread of antibiotic or herbicide resistance genes, and a number of approaches, which have been termed "clean transformation" methods, have been proposed to address the problem. One technique is the use of recombinase systems such as Cre-Lox or Flp-Frt, incorporated into the transformation vector. On activation, the recombinases allow unwanted sequences, such as marker genes, to be excised from the transgenic plant. An alternative approach involves the targeting of the plant transformation vector to the chloroplast, where it is integrated at a specific site by a homologous recombination mechanism. In this

case, pollen transfer of transgene sequences is greatly reduced, as plastid transmission to pollen is minimal in most plants. These techniques have been shown to work at reasonable efficiency in several dicotyledon species and have also been demonstrated in monocotyledons. Unfortunately, the present levels of efficiency achieved in monocotyledons would make their routine use a major hindrance to applied transformation programs. Currently, the more pragmatic solution is the use of marker genes, which are judged to have neutral environmental or health impact (see "Selection Systems" above), delivered in transformation vectors containing only essential sequences. In the future, however, as gene excision systems become more sophisticated and efficient, it is likely that these will become standard tools in monocotyledon transformation to ensure that transgenic crops contain only those sequences needed for expression of the desired new trait.

Applications

Considering that the transformation of monocotyledons is a relatively young technology, dating from the end of the 1980s, it has already had considerable impact in both basic and applied research and in commercial breeding of monocotyledon crops. In the field of basic research, transformation technology allows the investigation of the function of endogenous (native) genes by either up- or downregulating their expression and then examining the effect on the plant phenotype. The extreme level of downregulation, when a gene is silenced completely, is the transgenic equivalent of a classical loss-of-function mutation, such as those produced artificially by mutagenic chemicals or irradiation. In the transgenic case, however, the mutation can be directed at a specific target gene and there is the possibility of controlling the level of reduction in activity. In contrast, other mutagenic treatments are essentially random in terms of target and severity. In monocotyledons, transgenic approaches to gene overexpression and downregulation are increasingly used to understand metabolic processes. They can be applied in any case in which the gene for a target enzyme has been isolated. The approach has been used with success to study carbohydrate (starch) and amino acid biosynthesis in species such as corn and rice, by manipulating the expression of a series of synthetic enzymes. The power of this approach is such that in those species that are readily transformable, it is becoming standard practice to consider a knockout experiment to determine the function of any new gene that is isolated. A second basic research application of monocotyledon transformation is the testing of genes from other plant species to determine their function in monocotyledons. This approach is today providing important information on plant evolution and speciation and shows how essentially the same sets of genes are able in different plants to produce an enormous variety of differing structures and growth habits.

The main application of monocotyledon transformation is, however, its use in applied research and direct crop improvement by GM. In this application, transformation is used to create novel genetic variation, as the result of transgene expression, which can then be used to breed improved GM varieties. As in basic research applications, the targets for manipulation may be either endogenous or heterologous genes. Useful mutants can be produced by downregulation approaches, to reduce levels of unwanted molecules such as allergens or toxins, or to remove the activity of particular enzymes. Alternatively, the activity of desirable enzymes and the levels of valuable metabolites may be increased by overexpression. Transformation may also be used to express heterologous genes, resulting in the production of novel molecules in the monocotyledon crop. The first generation of GM crop varieties was produced using the latter approach, by expressing genes conferring resistance to normally non-selective herbicides (e.g., RoundupReady crops resistant to glyphosate or LibertyLink crops resistant to glufosinate) and creating insect-resistant varieties by expression of the *Bacillus thiuringiensis* (Bt) toxin gene.

These first commercial products have demonstrated the enormous potential of transformation technology to deliver monocotyledon crops with valuable new traits. As gene discovery and transformation research progresses, we can expect to see GM monocotyledons with a much broader range of modifications, including pest and disease resistance, increased tolerance to abiotic stresses, such as drought and salt, and with modified architecture, flowering, and productivity.

List of Technical Nomenclature

Direct gene transfer	The transformation method in which gene transfer is effected by a non-biological stimulus.
Explant	The plant tissue isolated or excised from the parent organ and used to initiate a tissue culture.
Protoplast	The plant cell after the removal of the cell wall (normally by enzymatic digestion).

Regeneration	The formation of a new plant by cultured plant tissue.
Stable transformation	The transformation in which the transferred transgene is stably integrated into the new host genome.
Suspension culture	Plant cells grown in shaken liquid culture.
Transgene	A gene artificially inserted (transformed) into a host genome.
Transient gene expression	The temporary expression of a transgene in a cell without stable integration into the host genome.

See also: **Crop Improvement**: Marker Assisted Selection; Molecular Markers. **Genetic Modification**: Gene Cloning, General Principles; Insertional and Transposon Mutagenesis; Transgene Stability and Inheritance. **Growth and Development**: Control of Gene Expression, Regulation of Transcription. **Tissue Culture**: Clonal Propagation, *In Vitro*.

Further Reading

Abelson PH and Hines PJ (1999) The plant revolution. *Science* 285: 367–389.

Barcelo P, Rasco-Gaunt S, Thorpe C, and Lazzeri PA (2001) Biotechnology of cereals: transformation and gene expression. *Advances in Botanical Research* 34: 59–128.

Birch RG (1997) Plant transformation: problems and strategies for practical application. *Annual Reviews of Plant Physiology and Molecular Biology* 48: 297–326.

Finer JJ, Finer KR, and Ponappa T (1999) Particle bombardment mediated transformation. *Current Topics in Microbiology and Immunology: Plant Biotechnology* 240: 60–80.

Hansen G and Chilton MD (1999) Lessons in gene transfer to plants by a gifted microbe. *Current Topics in Microbiology and Immunology: Plant Biotechnology* 240: 21–57.

Hansen G and Wright MS (1999) Recent advances in the transformation of plants. *Trends in Plant Science* 4: 226–231.

Komari T, Hiei Y, Ishida Y, Kumashiro T, and Kubo T (1998) Advances in cereal gene transfer. *Current Opinion in Plant Biology* 1: 161–165.

Lazzeri PA and Shewry PR (1993) Biotechnology of cereals. *Biotechnology and Genetic Engineering Reviews* 11: 79–146.

Lindsey K (ed.) (1998) *Transgenic Plant Research*. Amsterdam: Harwood Academic.

Wilmink A and Dons JJM (1993) Selective agents and marker genes for use in transformation of monocotyledonous plants. *Plant Molecular Biology Reporter* 11: 165–185.

Transformation in Plastids

P Maliga, Rutgers University, Piscataway, NJ, USA

Introduction

DNA in a higher plant cell is found in the nucleus, plastids, and mitochondria. Most efforts to improve crops focus on engineering the nuclear genome which encodes $>25\,000$ genes. The plastid genome (ptDNA) of higher plants encodes only ~ 120 genes most of which are photosynthetic genes and genes required for plastid maintenance. In addition, $>10\%$ of nuclear genes target proteins to plastids. These proteins complement plastid-encoded genes required for plastid maintenance and photosynthesis, and are involved in starch metabolism, lipid and amino acid biosynthesis, and nitrogen assimilation. Plastid transformation in a higher plant, tobacco (*Nicotiana tabacum*), was reported in 1990. Research in tobacco was inspired by the pioneering work on plastid transformation in *Chlamydomonas reinhardtii*, a unicellular alga, described in 1988. Transformation of the plastid genome was 7 years delayed relative to nuclear gene transformation (1983). Transformation of the mitochondrial genome in higher plants has yet to be accomplished.

More than 10 years after the initial breakthrough plastid transformation in higher plants is still routinely obtained only in tobacco. However, efforts are under way to extend plastid transformation to all major agronomic crops. Readily obtainable high protein levels, feasibility of expressing several proteins from polycistronic mRNAs, and lack of pollen transmission make plastids an attractive alternative to transgene expression in the nucleus. Progress has been made to engineer plastid-based herbicide resistance and insect resistance traits and to express recombinant proteins for molecular farming applications. The plastid engineering toolkit is complete with an efficient marker gene excision system to obtain marker-free transplastomic plants.

The Plastid Genetic System

There are several plastid types in a plant cell. Small rapidly dividing proplastids ($0.5\,\mu m$ to $1\,\mu m$) are found in meristematic cells; green photosynthetic chloroplasts ($\sim 10\,\mu m$) are present in leaves; yellow (lemon) or red (tomato) chromoplasts are characteristic of fruits; starch-containing amyloplasts are found in tubers and oil-containing elaioplasts in seed storage tissue. All plastid types have the same genome that is represented as a double-stranded circular DNA

120–180 kb in size (**Figure 1**). Staining of ptDNA revealed that the genome copies are clustered in plastid nucleoids, a location where proteins anchor ptDNA to membranes. Proplastids contain only one to two nucleoids. Each nucleoid contains ~10 ptDNA. Since meristematic cells contain 10–14 proplastids, the total number of ptDNA copies in a meristematic cell is 100 to 200. Leaf cells contain ~100 chloroplasts; each chloroplast contains ~10 (7–14) nucleoids, therefore the total number of ptDNA copies is ~10 000 per leaf cell. Given the large number of copies it is not surprising that ptDNA in a leaf cell can make up 10% of total cellular DNA. The challenge of plastid transformation is to uniformly alter all genome copies, the condition of obtaining a genetically stable plant.

Approaches to Stable Transformation of the Plastid Genome

Stable transformation of the plastid genome in higher plants is obtained by incorporation of the transforming DNA in the plastid genome where it is replicated with the rest of the genome. Plastid transformation vectors are *Escherichia coli* plasmid derivatives, which carry sequences (1–2 kb) to direct integration at specific sites in the plastid genome. The plastid vector is propagated in *E. coli*, and then introduced into plastids where the marker gene and the gene of interest integrate by two homologous recombination events in the targeted region of the plastid genome (**Figure 2**). The *E. coli* vector part is not inserted in the plastid genome. Since it lacks a plastid replication origin, it is subsequently lost. Plastid targeting sequences do not have special properties other than being homologous and may come from any part of the plastid genome. By now insertion of foreign DNA has been accomplished in at least 20 locations in the ptDNA (**Figure 1**). Stable integration of foreign DNA by homologous targeting is the general approach used for plastid transformation in higher plants.

An alternative to stable integration of transforming DNA in the plastid genome has been incorporation of plastid replication origins in plastid

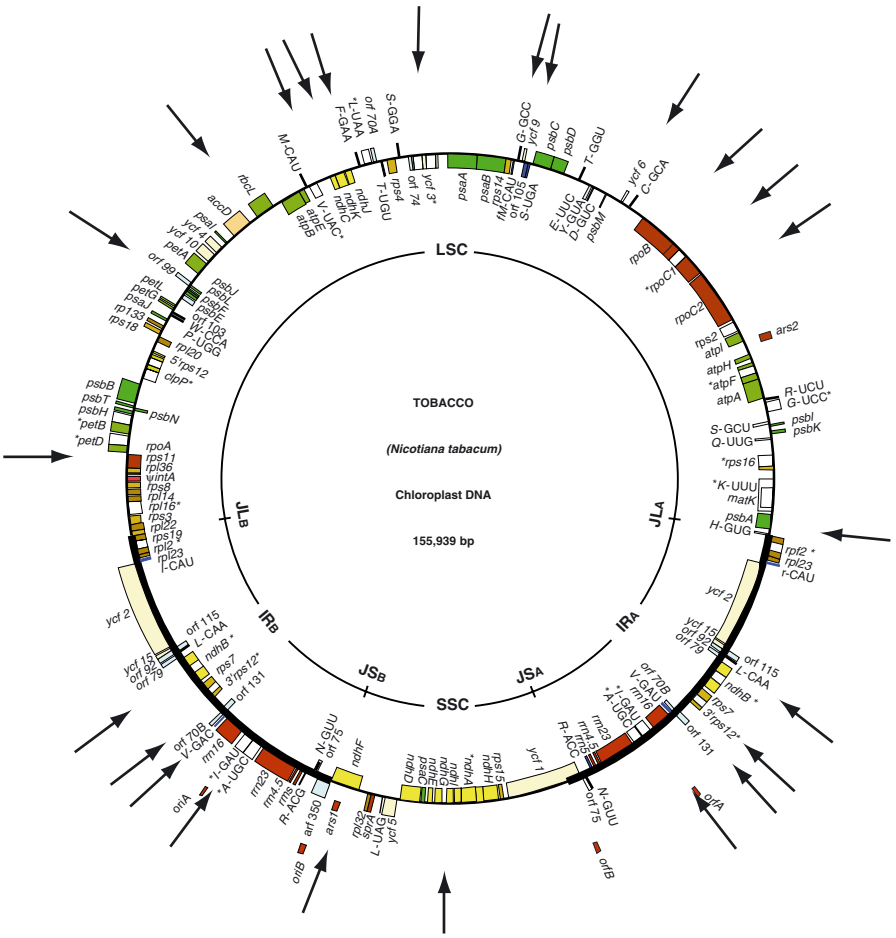

Figure 1 The map of the tobacco plastid genome. Arrows point at sites where foreign DNA has been inserted in the plastid genome. Modified from unpublished tobacco plastid genome map, kindly provided by M. Sugiura, with permission.

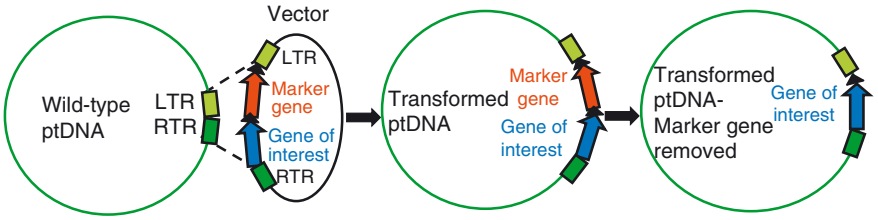

Figure 2 Plastid transformation involves incorporation of foreign DNA by two homologous recombination events via the left (LTR) and right (RTR) targeting sequence and removal of the marker gene using the CRE-*lox* site-specific recombination system. Reproduced with permission from Maliga P (2003) Progress towards commercialization of plastid transformation technology. *Trends in Biotechnology* 21: 20–28.

transformation vectors, which therefore can be independently maintained as episomal elements. Such plasmids are termed shuttle vectors as they are maintained in *E. coli* as plasmids by replication using the ColE1 replication origin and in plastids using a plastid *ori* sequence. These first-generation shuttle vectors are not practical because they are rapidly lost in the absence of selection. However, stable forms may be obtained in the future by improving vector design.

Methods for Introduction of Transforming DNA

The most commonly used approach for plastid transformation is biolistic DNA delivery. The transforming DNA is precipitated on the surface of microscopic ($\sim$1 μm) tungsten or gold particles. The $\sim$100 000 particles delivered in a charge are then accelerated in a suitable biolistic device to a speed such that the heavy DNA-coated metal particles penetrate the cell wall and settle in the plant tissue. In case of a tobacco leaf, the penetrated layer is at least four to five cells deep. The transforming DNA dissolves inside the chloroplasts from particles lodged inside the chloroplast or comes in contact with ptDNA by entering the chloroplast through holes punched by the metal particles. Each particle carries about 20 to 50 copies of the plastid vector DNA. Targets for biolistic transformation can be chloroplasts in intact tobacco leaf tissue or proplastids in cultured rice (*Oryza sativa*) embryogenic cells.

An alternative approach for DNA delivery utilizes polyethylene glycol (PEG) treatment to facilitate introduction of the transforming DNA into protoplasts. PEG treatment was originally developed for the transformation of the nucleus. For plastid transformation, protoplasts are obtained by enzymatic removal of the cell wall, and are then suspended in a DNA solution and treated with PEG. PEG treatment triggers shrinking of the protoplasts accompanied by engulfing some of the cell membrane and results in eventual disintegration of protoplasts. The PEG treatment, however, is terminated before the membrane is irreversibly damaged and the protoplasts are nursed back to health. The transforming DNA apparently finds its way into chloroplasts in a process not fully understood yielding transplastomic clones. Isolation and handling of protoplasts requires advanced cell-culture skills, which is the reason why the technically less demanding biolistic delivery to plastids in intact tissues is the commonly used method.

Another approach for the delivery of DNA into plastids is microinjection. Successful DNA delivery by microinjection was shown by reporter gene expression. However, utility of microinjection remains to be shown for stable transformation of the plastid genome.

Genetic Markers for Identification of Transplastomic Clones

Plastid transformation in higher plants was made feasible by marker genes that allowed selective amplification of transformed plastid genomes. First transplastomic tobacco plants were obtained by selection for spectinomycin resistance. Utility of spectinomycin resistance as a plastid marker has been known through the study of plastid-encoded spectinomycin resistance mutants. Spectinomycin blocks shoot regeneration from tobacco leaf sections and formation of green callus due to inhibition of plastid protein synthesis. Spectinomycin (and streptomycin) resistant mutants were identified by their ability to regenerate shoots and form green callus on selective spectinomcyin medium (**Figure 3**). Spectinomycin resistance is due to mutations in the plastid 16S rRNA that prevents spectinomycin binding. Initial plastid transformation vectors utilized plastid targeting sequences that contained mutant plastid *rrn16* genes as markers. The mutant *rrn16* genes were rapidly abandoned in favor of the 100 times more efficient *aadA* marker gene, a bacterial gene that also confers spectinomycin/streptomycin

Figure 3 Selection of transplastomic clones by spectinomycin resistance. (A) Spectinomycin inhibits callus formation, greening, and shoot regeneration from tobacco leaf segments on shoot regeneration medium. Transplastomic clones are resistant to spectinomycin and are identified as green shoots (or calli). (B) The shoots are chimeric, visualized by accumulation of green fluorescent protein in transplastomic sectors. Shoots regenerated from the transformed sector are homoplastomic. Reproduced with permission from Maliga P (2003) Progress towards commercialization of plastid transformation technology. *Trends in Biotechnology* 21: 20–28.

resistance. Spectinomycin resistant clones transformed with *aadA* are selected by the same procedure as the lines transformed with mutant *rrn16* genes (**Figure 3**). An alternative marker for plastid transformation is kanamycin resistance conferred by the *neo* or *aphA-6* genes, which encode different neomycin phosphotransferases. The enzyme encoded by *neo* has narrow substrate specificity and confers resistance only to kanamycin. The *aphA-6* gene product confers resistance to multiple drugs.

A potentially useful, new positive selection scheme involves identification of transplastomic tobacco lines by resistance to betaine aldehyde (BA). Resistance to BA is conferred by expression of betaine aldehyde dehydrogenase (BADH) in plastids. BADH is the product of a plant nuclear gene. Identification of transplastomic clones apparently relies on a phenotype unique to plastid expression. The applicability of the selection protocol is yet to be confirmed.

Not every marker gene is suitable for selective amplification of the rare, initial transplastome although resistance may be obtained when most plastid genomes carry the transgene. The *bar* gene is an excellent selectable marker for nuclear gene transformation conferring resistance to phosphinothricin (PPT) herbicides. Expression of the *bar* gene in plastids confers PPT resistance when it is introduced by selection for a linked *aadA* (spectinomycin resistance) gene. However, *bar* was not suitable for direct selection of transplastomic lines even if expressed at a high level (>7% TSP (total soluble cellular protein)). Thus, it appears that subcellular localization of the gene encoding the detoxifying enzyme is critical when directly selecting for herbicide resistance and initial, low-level expression from a few transformed plastid genome copies is insufficient to protect the cell from PPT.

A negative plastid selection scheme has been developed based on expressing the bacterial enzyme cytosine deaminase. Cytosine deaminase catalyzes deamination of cytosine to uracil. 5-fluorocytosine (5FC) is toxic for cytosine deaminase-expressing cells, since the enzyme converts 5-fluorocytosine to the toxic 5-fluorouracil. Plant cells lack cytosine deaminase but become sensitive to 5FC when the enzyme is expressed in plastids. Resistance to 5FC was utilized as a marker to identify cells in which the gene encoding cytosine deaminase was removed by the CRE-*lox* site-specific recombinase.

In addition to the positive and negative selection schemes, visual markers are also available in plastids. β-Glucuronidase (GUS) enzymatic activity can be visualized by histochemical staining and the green fluorescent protein (GFP) can be visualized in UV light. A gene encoding GFP was fused with *aadA* and the fusion gene was used as visual and selective (spectinomycin resistance) bifunctional marker.

Plastid Transformation Vectors

There are at least as many plastid transformation vectors (>20) as insertion sites targeting insertions at unique locations in the plastid genome (**Figure 1**). However, there are only two vector families that have been refined to provide cloning convenience by combining a plastid targeting sequence, a selective marker, and a multiple cloning site. Both of these utilize *aadA* as selective marker (**Figure 4**). The *trnV/rps12* intergenic region as insertion site has the advantage that there is no read-through transcription from the direction of the *rrn* operon. The *trnV/rps12* intergenic region is in the inverted repeat region of the plastid genome. Genes at this site are present in two copies per genome.

Vector development is critical for extending the technology to new crops, and for new applications. In the near future new vector families are expected to appear that will utilize kanamycin resistance as a marker gene and/or will have recognition sequences for site-specific recombinases to allow efficient removal of marker genes.

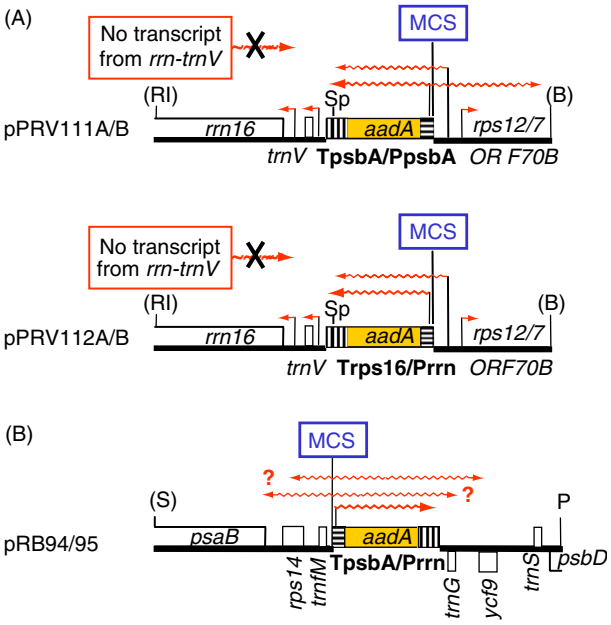

Figure 4 Plastid transformation vectors. (A) The plastid repeat vector (pPRV) family (GenBank accession numbers U12809-U12815) (Zoubenko (1994) #900). Cloning convenience is ensured by alternative expression signals for *aadA* in the pPRV111 and pPRV112 series, and the pUC multiple cloning sites (MCS) available in both orientations (A or B). There is no read-through transcription of transgenes in MCS from the *rrn* operon. (B) The pRB94/pRB95 vectors (EMBL Accession numbers AJ312392, AJ312393). Vectors differ with respect to the orientation of the Bluescript MCS. No information is published on read-through transcription at the MCS. Reproduced with permission from Maliga P (2003) Progress towards commercialization of plastid transformation technology. *Trends in Biotechnology* 21: 20–28.

Status of Implementing Plastid Transformation in Crops

Plastid transformation has been most successful in the solanaceous crops tobacco, potato (*Solanum tuberosum*), and tomato (*Lycopersicon esculentum*). Plastid transformation in potato has opened up the possibility to engineer tuber-specific plastid gene expression. Transformation of plastids in tomato is the starting-point for attempts for fruit-specific expression of recombinant proteins. Tomato fruits are good candidates for delivery of oral vaccines. Plastid transformation was also feasible but inefficient in species that belong to the mustard (Brassicaceae) family, including *Arabidopsis thaliana*, *Brassica napus*, and *Lesquerella fendleri*. Out of the cereals, heteroplastomic plants have been obtained in rice. Plastid transformation in each of these species has some unique features. Plastid transformation of tobacco is discussed below as it is the best characterized system.

Sterile leaves of culture-grown tobacco plants are bombarded with vector DNA-coated tungsten particles. The leaves are dissected into small squares ($1\,cm^2$) and cultured on nonselective shoot induction medium to induce cell division. After 2 days, the leaves are transferred onto spectinomycin-containing ($500\,mg\ l^{-1}$) shoot induction medium. Nontransformed cells proliferate slowly and form white callus on the selective medium. Cells carrying transformed ptDNA copies regenerate green shoots in 3–12 weeks. These shoots are chimeric and consist of sectors with transformed and nontransformed plastids (**Figure 3B**). Shoots regenerated from the transformed sectors yield genetically stable, homoplastomic plants. Starting with vector DNA, it takes about 3–5 months to obtain genetically stable transplastomic plants. Removal of the marker gene may extend this time period by about one month (see below).

Plastid transformation efficiency is measured as transplastomic clone per bombarded sample. In tobacco, one leaf (5–7 cm) constitutes one sample. Plastid transformation efficiencies are difficult to reproduce: the efficiency is typically in the range between 0.5 and 5.0 transplastomic clones per bombarded sample.

Mechanisms Yielding Genetically Stable Transplastomic Plants

Plastid transformation involves transformation of one or a few ptDNA copies, then gradual elimination of the nontransformed copies by dilution on a selective medium. Achieving the homoplastomic state in cultured tobacco cells takes ~20 cell divisions. Homoplastomic cells derive from heteroplastomic cells because somatic cell division does not involve exact duplication of the cytoplasm and because genomes of the organelles that carry the selective marker are preferentially replicated on a selective medium in tissue culture. Formation of homoplastomic cells is accelerated by the ptDNA being clustered in nucleoids and that the nucleoids themselves are being localized in independently segregating plastids. Shoot regeneration from cultured tobacco leaf segments involves formation of meristematic cells with concomitant reduction in plastid number from ~100 chloroplasts in leaves to ~12 proplastids in meristematic cells and concomitant reduction in nucleoid number from 7 to 14 (in chloroplasts) to 1 to 2 (in proplastids), then rebuilding the organelle and genome copy number in leaf cells on a selective medium. Plastid division during the chloroplast to proplastids transition, in the absence of an increase in nucleoid number, rapidly yields plastids which carry the nucleoid that

came in contact with the transforming DNA (**Figure 5A**; nucleoid #1). Replication and segregation of transformed genome copies eventually yield proplastids that are homoplastomic for the transgenic ptDNA (**Figure 5A**; event #1a). Reduction of plastid number in meristematic cells, coupled with random segregation of plastids (only some of the progeny cells carry transformed plastids), accelerates elimination of plastids which do not carry transforming DNA (**Figure 5B**; progeny cell #1). DNA replication and segregation at the nucleoid and organellar levels yield the homoplastomic sectors that are visible in the leaves regenerated on spectinomycin medium (**Figure 3B**). Shoots regenerated from these sectors are typically homoplastomic and carry only transformed genome copies.

The observed reduction of ptDNA copy and organelle number in tissue culture cells is not universally applicable to all species. It is known, for example, that embryogenic rice tissue culture cells contain as many plastid genome copies as leaf cells. Thus, obtaining the homoplastomic state after bombardment of rice embryogenic cells is likely to take more cell generations than in tobacco.

Engineering of Plastid Genes for High-Level Expression

To facilitate expression of foreign genes in chloroplasts, modular cassettes have been designed with suitable restriction sites at the boundaries of the cassettes (**Figure 6**). The 5′ PL cassette (PL, promoter and leader) includes the promoter and translation control sequences. The translation control sequences can be the mRNA 5′-UTR (untranslated region) (**Figure 6A**), or the 5′-TCR (translation control region) that includes the 5′-UTR and an N-terminal segment of the coding region (**Figure 6B**). The mRNA 5′-UTR includes a stem–loop structure required for mRNA stability and sequences that facilitate loading of mRNAs onto ribosomes. The 3′ T-cassette encodes the mRNA 3′-UTR, which also includes a stem–loop structure. The 3′-UTR functions as an inefficient terminator of transcription and is required for mRNA stability.

High-level expression of foreign genes is dependent on transcription from a strong promoter. In most cases the strongest plastid promoter, driving expression of the plastid rRNA operon, is used. Targets for engineering are the 5′-UTR and coding region N-terminus because dramatic (10 000-fold) changes in protein accumulation have been obtained by choosing alternative 5′-UTRs. Success of improving protein expression by engineering the 5′-UTR is underlining the importance of posttranscriptional regulation in plastid gene expression.

The need for synthetic genes is a concern when expressing a protein in a heterologous host. Infrequent use of specific codons in the host organism normally indicates limiting supply of cognate tRNAs. Although the tobacco plastid genome is relatively AT-rich, there is no example for extremely low codon usage frequencies. Indeed, codon optimization in tobacco plastids yielded only two- to threefold increases in protein levels, much less increase than the 10 000-fold change obtained by engineering translation control signals.

Premature transcription termination and degradation of mRNAs was a problem when expressing high-AT bacterial genes in the yeast and plant nuclei. Examples are the *Bacillus thuringiensis* insecticidal protein gene and the gene encoding the tetanus toxin C-terminal fragment. The same coding sequences yielded stable mRNAs and high levels of protein expression in plastids indicating that plastids are a versatile system for expression of foreign genes of diverse origin.

Potential and Limits of Transplastomic Technology

It may be more attractive to incorporate transgenes in the plastid genome than in the nucleus. Given the size variability of plastid genomes it is assumed, that the plastid genome can be enlarged by about 50 kb. Dependent on the size of the genes, 50 kb of foreign DNA may encode 20 to 30 genes. The largest inserts published thus far are about 5 kb in size, encoding a *Bacillus thuringiensis* insecticidal protein gene and a selectable marker gene. The genes have been inserted in the inverted repeat region, therefore the plastid genome expressing the two genes is increased by a total of ~10 kb. Some of the plastid genes are expressed in operons therefore expression of enzymes in the same metabolic pathways may be practical from polycistronic mRNAs. The largest protein expressed, Cry1A(c), is 133 kDa in size. Protein levels may be readily obtained in the 5–25% range.

Localization of a transgene in the plastid genome may not always be advantageous. No plastid-encoded protein is known to be exported, a limitation of plastid expression when subcellular localization other than the plastid is desired. Plastid proteins may be modified posttranslationally. Glycosylation, however, is a protein modification that is known not to occur in plastids. Also, little is known about tissue-specific expression of plastid genes other than

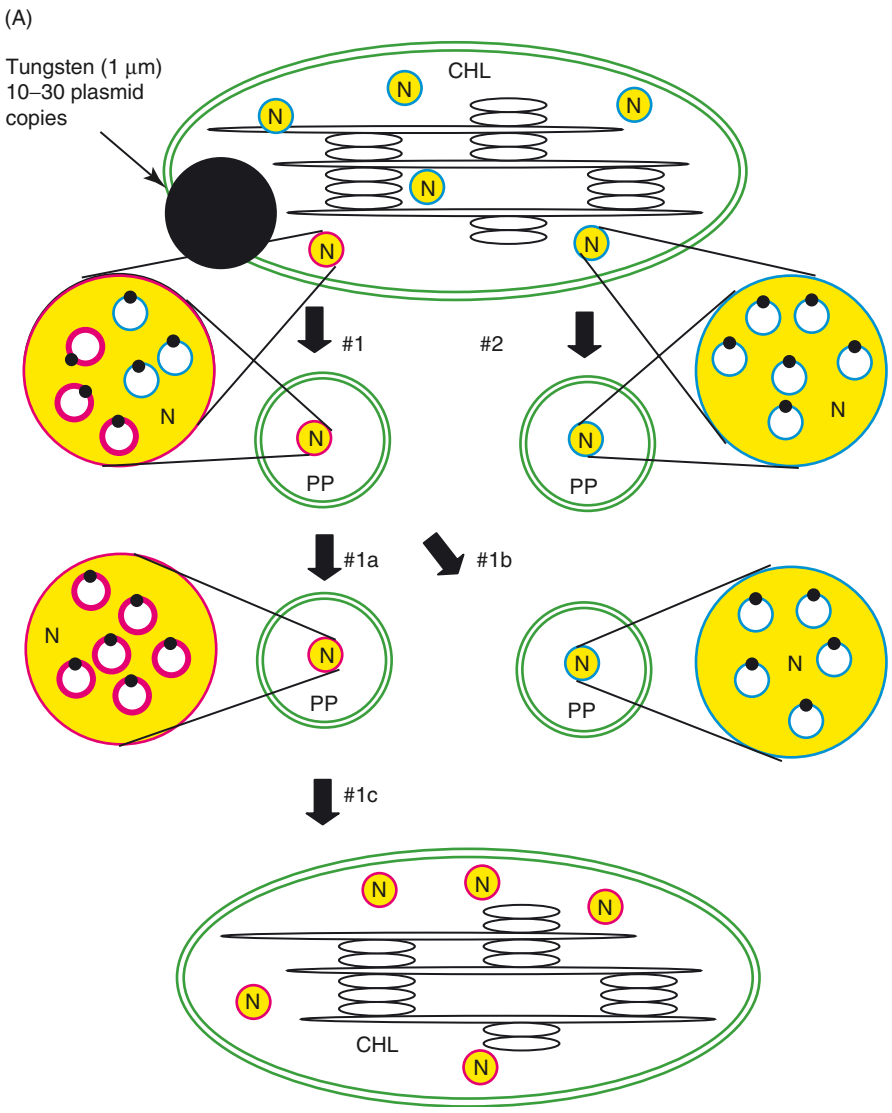

Figure 5 Chloroplasts (CHL) to proplastid (PP) dedifferentiation accelerates formation of homoplastomic cells during shoot regeneration. (A) Nucleoids are the units of segregation during the chloroplast to proplastids transition. Transformed and nontransformed nucleoids (N) are marked as gray and open circles, respectively. In enlarged nucleoids, transformed and nontransformed ptDNA is depicted as gray and open circles, respectively; ptDNA are anchored to membranes by proteins (black dots). Nucleoid #1 is heteroplastomic, and is the progenitor of homoplastomic transgenic (#1a) and wild-type (#1b) proplastids. Homoplastomic proplastid differentiates into chloroplast (#1c; bottom). Wild-type proplastid (#2) is antibiotic sensitive and divides slowly. (B) Reduction in plastid number during chloroplasts (~100; elongated forms) to proplastid (~12; circles) transition and lack of exact duplication of the cytoplasm accelerates formation of homoplastomic cells. On top is shown leaf mesophyll cell with one transformed chloroplast (unshaded) and nucleus (Nu). Meristematic cell #1 is heteroplastomic. Cleavage of the cytoplasm yields one meristematic cell with transformed chloroplasts only (#1A) and one with wild-type plastids (#1b). Meristematic cell with transformed plastids is the progenitor of homoplastomic mesophyll cells (1c). Meristematic cells with wild-type plastids divide slower and are eventually diluted out.

expression in chloroplasts. Information on plastid gene expression in nongreen tissue is needed to aid the design of tissue-specific expression in fruits, tubers, epidermis cells, etc. Expression tools borrowed from prokaryotic organisms may fill the need. Nuclear control of plastid gene expression was reported by expressing a phage RNA polymerase in the plant nucleus that was driving transgene expression from a phage promoter in plastids.

Marker Gene Elimination Systems

Marker genes are essential to obtain transplastomic plants. However, as soon as plastid transformation is accomplished, removal of the marker genes is highly desirable. There is only a small number of efficient plastid marker genes; removal of the marker allows a second round of engineering using the same marker gene. Also, release of crops with selectable marker

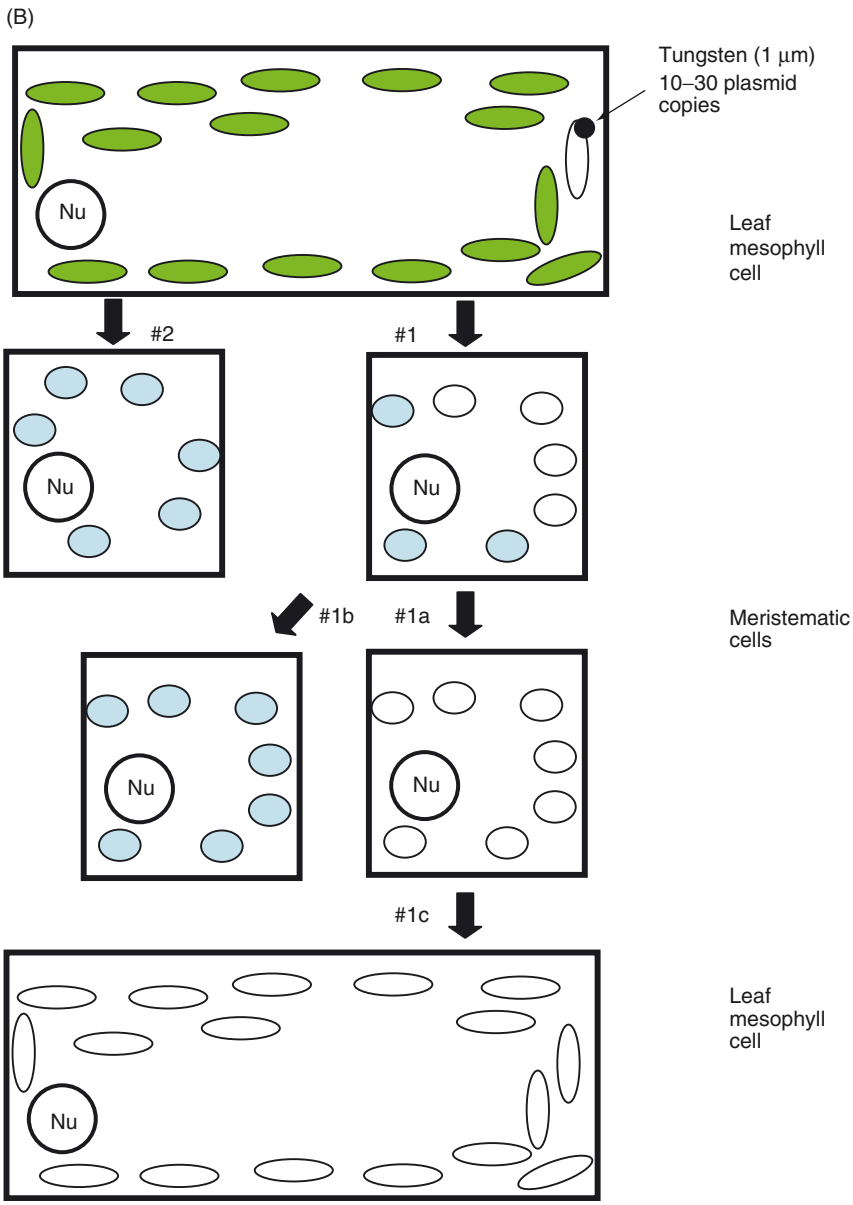

Figure 5 Continued.

genes is undesirable due to regulatory concerns. Furthermore, the marker genes may be expressed at a high level, imposing an unnecessary metabolic burden on the crop.

Multiple approaches have been developed for the removal of marker genes from the plastid genome. The most efficient system uses the P1 phage CRE-*loxP* site-specific recombination system (**Figure 7**). According to the CRE-*loxP* scheme, the marker gene and the gene of interest are introduced into the plastid genome in the absence of CRE activity. The marker gene is flanked by two directly oriented *lox* sites. When elimination of the marker gene is required, a gene encoding a plastid-targeted CRE site-specific recombinase is introduced into the nucleus which, subsequent to its import in plastids, excises sequences between the *loxP* sites. In tobacco, removal of the marker gene extends the time needed to obtain marker-free plants only by 1 month.

Applications of Plastid Transformation

Applications in Basic Science

Plastid transformation in basic science is widely utilized to understand plastid gene function and gene regulation. By now knockout mutants have been described for more than 17 plastid genes. Study of these mutants helped to identify the genes' function. For example, study of *ndh* knockout plants

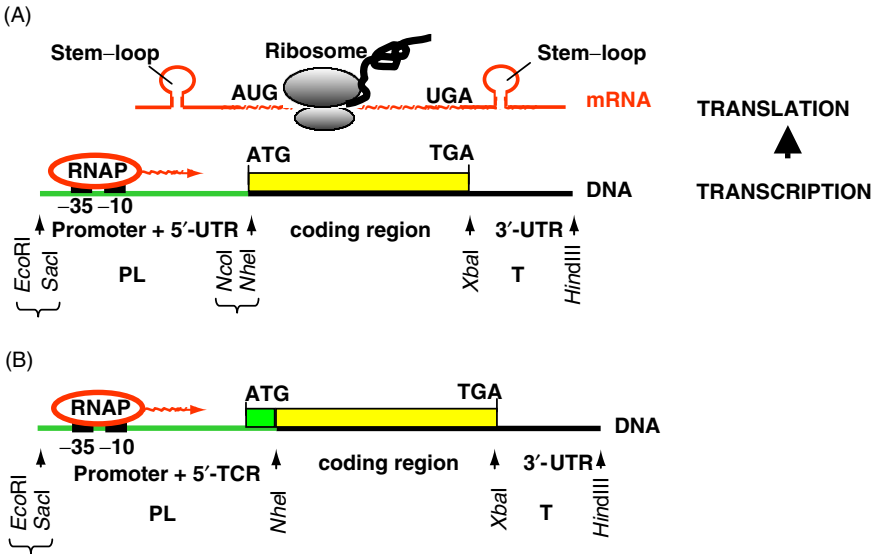

Figure 6 Modular design of plastid promoter-leader (PL) and terminator (T) cassettes. PL cassettes encode a promoter and (A) 5′-untranslated region (5′-UTR) or (B) a 5′-translation control region (5′-TCR, 5′-UTR plus a segment of the coding region). Proteins are expressed from *Nhe*I/*Xba*I or *Nco*I/*Xba*I fragments. Stem–loop structures formed by nucleotide pairing in the 5′-UTR and 3′-UTR and the AUG translation initiation and UGA stop codons are marked. Reproduced with permission from Maliga P (2002) Engineering the plastid genome of higher plants. *Current Opinion in Plant Biology* 5: 164–172.

yielded the first insights into the function of the "chlororespiratory pathway." Deletion of the plastid-encoded RNA polymerase subunits revealed promoters for a nuclear-encoded, plastid targeted second plastid RNA polymerase and shed light on the division of labor between the two transcription machineries. Plastid transgenes were utilized to identify plastid promoter elements and sequences regulating mRNA translation. Study of transgenes incorporating segments of edited genes yielded new information about the functional significance of posttranscriptional C-to-U conversion that affects only 20 to 31 specific C nucleotides in 19 transcripts. Plastid transformation is also utilized to explore the feasibility of improving photosynthetic efficiency.

Applications in Agriculture and Molecular Farming

Currently, no transplastomic plant is approved for field release. This will change as soon as plastid transformation is implemented in major agronomic crops as groundwork has been laid to engineer commercially important traits. Progress has been made to understand engineering resistance to herbicides with glyphosate and phosphinothricin (Bialaphos, Liberty) as active compounds. Also, *B. thuringiensis* insecticidal protein genes were expressed at a high level in chloroplasts, allowing implementation of a high-dose strategy. A second area with good potential for applications of plastid transformation is expression of recombinant proteins for molecular farming. The examples include expression of antibodies, vaccines, and industrial enzymes.

Biosafety and Transplastomic Plants

In general, transplastomic plants are considered to be the safest form of GM crops for transgene containment as in most crops plastids are not transmitted by pollen. In contrast, if a transgene is incorporated in the nucleus, each pollen carries the transgene. Examples for strict maternal inheritance of plastids are corn (*Zea mays*, maize), wheat (*Triticum aestivum*), rice, cotton (*Gossypium* spp.), soybean (*Glycine max*), tobacco, tomato, oilseed rape (*Brassica napus*), and sunflower (*Helianthus annuus*). Examples for biparental inheritance of plastids are potato and alfalfa (*Medicago sativa*, lucerne).

Gene flow should be considered within a species and between species, by hybridization with weedy relatives. Plastid localization eliminates outgoing transgene flow via pollen. However, opportunities still exist for the much less frequent hybridization at the field's edge, or by hybridization due to fertilization with pollen from outside the field. Thus, the possibility for hybridization with weedy relatives should be considered. Danger is presented only if the hybrids of the transplastomic crop and the weedy relative are viable. An alternative source of plastid genome transfer to weeds could be extremely rare, paternal transmission of plastids detected under laboratory conditions in plants with an alien cytoplasm or in interspecific crosses. Such plastids,

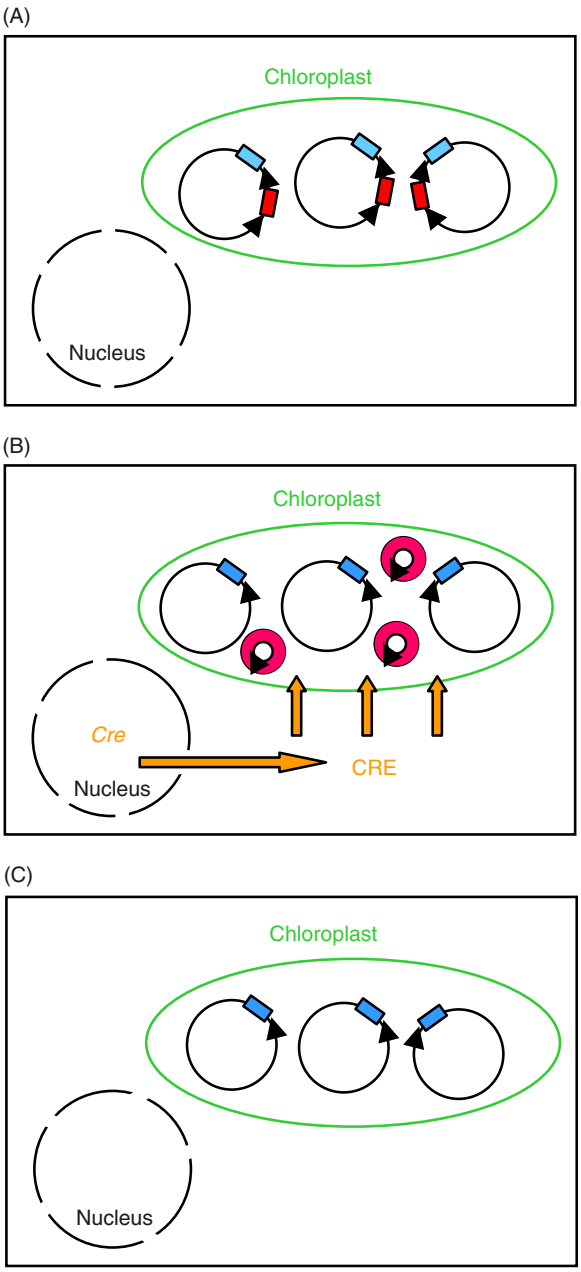

Figure 7 Removal of marker genes by the CRE-*lox* site-specific recombination system. (A) Plant cell with transformed plastid genome. Marker gene flanked by *lox* sites. The gene of interest is represented by light gray box. (B) The *Cre* gene is introduced into the nucleus. The plastid-targeted CRE enters the organelle and simultaneously excises the marker genes from all genome copies. Note both products, the circularized marker gene, and modified plastid genome. (C) The marker gene is rapidly lost in somatic cells. The nuclear *Cre* gene segregates out in the seed progeny. Only one *lox* site and the gene of interest are present in the plastid genome.

even if transfer occurs, are very unlikely to get established in the recipient weed in the light of difficulties to obtain homoplastomic plants. Nevertheless, the possibility of plastid genome transfer from future transplastomic crops to weedy relatives and their potential impact should be evaluated.

List of Technical Nomenclature

Heteroplasmic	Refers to the presence of different cytoplasmic genetic determinants determined by segregation in the seed progeny. Term was commonly used in the past when plastid or mitochondrial localization of a maternally inherited trait was not known.
Heteroplastomic	Refers to the presence of different plastid genomes.
Homoplasmic	Refers to lack of segregation of a maternally inherited trait. Term was commonly used in the past when plastid or mitochondrial localization of a maternally inherited trait was not known.
Homoplastomic	Refers to a uniform population of plastid genomes.
Integrating plastid vector	Integrates in plastid genome via targeting sequences. Episomal element (plasmid) in *E. coli*.
Plastid shuttle vector	Episomal element in plastids and *E. coli*.
Plastome	The genome of plastids.
Transformation of plastids	Stable transformation of plastids that yields nonsegregating seed progeny.
Transient gene expression in plastids	Expression of genes in plastids from introduced DNA that is not stably maintained.
Transplastome	Transformed plastid genome.
Transplastomic	Pertaining to the transformed plastid genome.

See also: **Crop Improvement**: Alien Cytoplasms. **Ethics and Biosafety**: Development and Commercialization of Genetically Modified Plants. **Genetic Modification, Applications**: Cell Factories; Molecular Farming; Plantibodies. **Genetic Modification**: Transformation in Dicotyledons; Transformation in Monocotyledons. **Growth and Development**: Control of Gene Expression, Regulation of Transcription; Field Crops. **Pests**: Genetic Modification of Pest Resistance, Insect Pests. **Weeds**: Herbicide Resistance.

Further Reading

Bendich AJ (1987) Why do chloroplasts and mitochondria contain so many copies of their genome. *Bioessays* 6: 279–282.

Bock R (2001) Transgenic plastids in basic research and plant biotechnology. *Journal of Molecular Biology* 312: 425–438.

Heifetz PB (2000) Genetic engineering of the chloroplast. *Biochimie* 82: 655–666.

Koop HU, Steinmüller K, Wagner H, *et al.* (1996) Integration of foreign sequences into the tobacco plastome via peg-mediated protoplast tranformation. *Planta* 199: 193–201.

Kuriowa T (1991) The replication, differentiation, and inheritance of plastids with emphasis on the concept of organelle nuclei. *International Review of Cytology* 128: 1–62.

Maliga P (2002) Engineering the plastid genome of higher plants. *Current Opinion in Plant Biology* 5: 164–172.

Maliga P (2003) Progress towards commercialization of plastid transformation technology. *Trends in Biotechnology* 21: 20–28.

Pyke KA (1999) Plastid division and development. *Plant Cell* 11: 549–556.

Staub JM (2002) Expression of recombinant proteins via the plastid genome. In: Parekh SR and Vinci VA (eds) *Handbook of Industrial Cell Culture: Mammalian, Microbial and Plant Cells*, pp. 261–280. Totowa: Humana Press Inc.

Sugiura M (1992) The chloroplast genome. *Plant Molecular Biology* 19: 149–168.

Thomas MR and Rose RJ (1983) Plastid number and plastid ultrastructural changes associated with tobacco mesophyll protoplast culture and plant regeneration. *Planta* 158: 329–338.

Whitney SM and Andrews TJ (2001) Plastome-encoded bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) supports photsynthesis and growth of tobacco. *Proceedings of the National Academy of Sciences of the USA* 98: 14738–14743.

Transgene Stability and Inheritance

H D Jones, Rothamsted Research, Harpenden, UK
M Cannell, Syngenta, Jealotts Hill International Research Centre, Bracknell, UK

Introduction

The improvement of crop performance by genetic engineering requires that inserted genes are expressed at the desired level, in the intended tissue, at the required stage of development, and/or following exposure to a specific stimulus. The ideal outcome of transgenic overexpression is that, for a given promoter/gene cassette, the number of transgene copies would correlate with the level of protein expression and this would be identical between independent lines. In addition, the transgenes would be faithfully inherited accordingly to Mendelian principles, giving rise to the desired phenotype in the descendants and derivatives of the primary transformant. However, since the first reports of plant transformation in the early 1980s, it has become obvious that not every transformation event with a particular gene construct expresses or transmits its transgene(s) to its progeny in the intended (or even a predictable) way. Different expression levels are often found between different events with the same transgene. Some of the possible causes of this variation are copy number, ploidy, and so-called "position effects." In addition, various forms of gene silencing, unintended insertional mutagenesis, somaclonal variation, and/or unexpected transgene inheritance patterns can be found. The molecular basis for much of this variation is poorly understood and, as a consequence, crop genetic engineers generally have to produce large numbers of transformants in order to identify the individuals with the appropriate number of transgene copies that exhibit and predictably inherit the required phenotype. This article will cover some aspects of current thinking on transgene stability and inheritance, the factors influencing them, and strategies employed to avoid instability.

Transgene Integration and Inheritance

The method of transgene delivery (most commonly *Agrobacterium*-mediated or by direct gene transfer such as particle bombardment) can impact on the complexity of transgene integration and inheritance. There are reports indicating that *Agrobacterium*-mediated transformation results in fewer transgene copies with a more simple integration pattern than direct methods. Both methods generally result in plants in which integration occurred at a single genetic locus and although single or multiple copies of the transgene may be present within a single locus, this locus segregates and is inherited as a single dominant allele, at a Mendelian ratio of 3:1. Inheritance studies indicate that the primary transformed plant possesses the transgene in the hemizygous state (the term hemizygous is used here in preference to heterozygous because there is no corresponding allele present on the homolog of the transformed chromosome (see **Figure 1**)).

Occasionally, plants might be recovered in which integration occurred at more than one genetic locus,

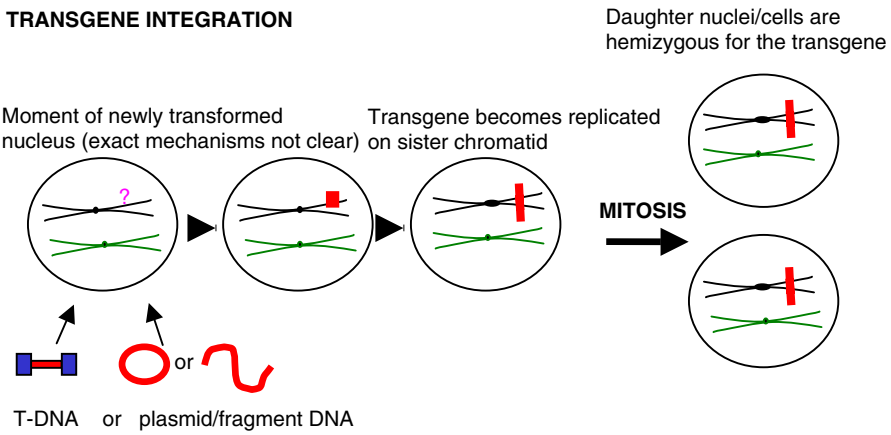

Figure 1 Transforming DNA is integrated at one or more loci and becomes replicated by mitosis resulting in daughter cells that are hemizygous for the transgene.

leading to independent segregation of two or more traits. This situation can sometimes be used to facilitate the removal-by-segregation of unwanted selectable marker genes. The relative spatial organization of the selectable and trait genes on the DNA vector is an important consideration. For example, different transgene sequences may be delivered simultaneously, either on whole plasmids or as linear fragments of DNA, and the selectable marker and trait genes may reside either on the same or on separate plasmid/fragment molecules. If they are on the same molecule prior to delivery (this approach is sometimes called cointegration), clearly it is unlikely that they will separate by segregation. However, if they are delivered on different molecules (called cotransformation), a proportion of transformants with the marker and trait genes at separate loci may be recovered. Traditionally *Agrobacterium* delivers a single transfer (T-DNA) fragment, flanked by left and right border sequences. Such an arrangement leads to the insertion of only one type of T-DNA-bounded transgene, resulting in transgenes being inseparable by segregation. Relatively recently, however, vector systems have been developed in which two different T-DNA fragments may be inserted during the same transformation experiment. As with cotransformation in direct gene delivery methods, these may generate transformants in which the marker and trait genes segregate independently.

Integration at a Molecular Level

The exact order and nature of the molecular events leading to the stable integration of a transgene into the host plant chromosome are not clear. However, it is likely that with the exception of a few specific applications and regardless of the method of DNA-delivery, genetic transformation involves double-stranded illegitimate recombination at one or multiple loci, utilizing the cell's nuclear repair machinery. The locus may contain one, or multiple, copies of the transgene, which may have undergone rearrangements and/or may have generated short lengths of "filler" DNA homologous to flanking plant genomic DNA. The moment of integration most likely coincides with a DNA metabolic event such as replication or transcription. Some *in situ* hybridization data suggest that physical transgene integration occurs at random within and between plant chromosomes; however, other analyses demonstrate a preference for distal chromosomal locations and actively transcribed regions in the plant genome with the possibility that native plant genes are disrupted in the process.

Variability in Transgene Expression

It is common to find differences in transgene expression between independent transgenic events with the same construct. There has been much discussion as to the basis of this variation. It is likely that the major contribution is from line to line variation in gene silencing modulated by factors such as copy number, transgene rearrangements, and position effect. The physical location of a transgene within the genome may modulate an increase in expression when inserted near to native enhancer elements or decrease it when in regions of condensed chromatin. Other factors known to have a direct effect on transgene expression are gene dosage via homo/heterozygosity and ploidy.

Gene Silencing

Gene silencing is a mechanism that downregulates or degrades specific genes and can propagate this effect to homologous sequences. It was once thought to be a quirk of the transformation process but it is now recognized as part of the plant's innate gene regulatory mechanisms. Transgene silencing probably reflects a manifestation of a natural "plant defense system" against invasive nucleic acids and also an inherent genome maintenance and regulatory system. The evolutionary drivers for its occurrence were most likely pathogenic attack (exogenous) and disruptive mobile DNA elements such as transposons and retrotransposons (endogenous). The combination of this surveillance system for nonself DNA and the fact that the most commonly used sequences in transgene cassettes are widely heterologous with respect to the host must certainly contribute to transgene silencing as a widespread phenomenon. It can occur naturally or be triggered by transgenes or viruses and involves some degree of DNA or RNA sequence homology. Gene silencing can operate at transcriptional and posttranscriptional levels and our understanding of the precise molecular components involved has increased enormously over the last 2 or 3 years. Silencing effects can be triggered by transgenes and by viral infection. Various approaches have been used in attempts to intentionally induce selective gene silencing in transgenic organisms, including antisense RNAs, chimeric RNA/DNA oligonucleotides, double-stranded RNAs, cosuppression, various virus-induced silencing methods, etc., but precise downregulation of targeted genes can be unpredictable and unintentional silencing of native or introduced genes in transformed plants may be observed. Additionally, transgene-mediated silencing does not necessarily occur in the primary transformant but can be triggered during subsequent clonal propagation, self-fertilization, or hybridization.

Mechanisms of Gene Silencing

One of the main mediators of transcriptional transgene silencing in response to genome perturbation is DNA methylation. Although the precise signals and order of events leading to *de novo* methylation of transgenes are not fully understood, it is clear that DNA–DNA or RNA–DNA interactions are crucially involved and that shared homology within integrated copies of transgenes can induce methylation. In some cases, it is directed specifically at the promoter region, and affects transcription probably by attracting factors that are able to form chromatin structures that are transcription repressive. These effects are most commonly seen within an integration event, where end-to-end insertional repeats or rearrangements may lead to the formation of secondary structures via DNA–DNA pairing and induce nucleotide methylation or chromatin condensation. Transcriptional gene silencing has also been observed in *trans*, where an insertion at one locus leads to silencing of other transgene insertions or native genes at distant genetic locations. Post-transcriptional gene silencing involves the formation of double-stranded RNA, which leads to sequence-specific RNA degradation. This type of gene silencing is also termed cosuppression because it has been observed in plants where sense-overexpression sometimes leads to homology-dependent downregulation of native and transgenes. However, these are rare events with the vast majority of sense transformation events giving rise to conventional overexpression phenotypes.

Chimerism

With the exception of germline transformation, transgenes are delivered into cultured protoplasts or explants of dividing plant tissue and it is assumed that different cells become transformed independently from one another. Many of the individual transformed cells (and nontransformed cells) have the potential to divide, somatically differentiate, and regenerate into a whole plant, and tissue culture conditions are manipulated to encourage this process. To allow transformed cells to grow preferentially, a selectable marker gene is delivered along with the trait gene, and the appropriate positive or negative selection pressure is applied during the plant culture phase. Tissue culture and selection procedures are generally intended to allow only whole plants that are derived from one original transformed cell to regenerate. However, some poorly optimized systems can give rise to plants that survive selection but are derived from a single cell that was not transformed. In some cases, these may be nontransgenic plants that simply escaped selection. In other cases, the transformation may have occurred in a cell or cells that were derived from the original progenitor cell and from which subsequent divisions lead to an unequal distribution of transformed and nontransformed tissue within the same whole plant. Transgenic sectors develop that may or may not incorporate gametic tissue. Such plants are chimeras and their existence is sometimes an explanation of poor or zero transgene inheritance.

Strategies to Avoid Instability

With our current state of knowledge it is impossible to predict the expression profile and inheritance pattern of every *de novo* transgenic event. However, strategies are emerging that are likely to increase the proportion of those displaying stable, expected phenotypes. These include low number of transgene copies with no rearrangements, low sequence homology between different introduced sequences and to native sequences, no viral/pathogenic sequences in the introduced DNA, short tissue-culture phase with low auxin concentrations to reduce somaclonal variation, and selection regimes that reduce the likelihood of chimeras. In addition, there is evidence that the inclusion of intronic or matrix attachment region (MAR) sequences into vector designs influences expression stability.

Introns

The inclusion of introns between promoter and transgene coding region has been shown to positively influence expression stability (intron-mediated enhancement, IME). For example, the most commonly used promoter for high-level constitutive transgene expression in dicotyledons is the 35S promoter of the cauliflower mosaic virus (CaMV35S). However, this promoter works poorly in monocotyledons unless linked to an intron such as the first one from the corn (*Zea mays*; maize) *adh* gene (*adh*1 intron). Similarly, the widely used monocotyledon promoter of the corn ubiquitin gene drives only weak expression unless coupled to the first intron of the gene. IME is often dependent on which promoter and/or gene is used and the precise mechanism of action is not fully understood. Many introns have been shown to increase expression levels and, interestingly, IME appears to occur at higher levels in monocotyledons than dicotyledons. From this knowledge, one would intuitively expect the presence of introns within transgene coding sequences to also have a positive effect on expression stability. However, a systematic comparison of expression levels from cDNAs vs genomic transgenes has yet to be carried out. The reason for the widespread use of cDNA sequences as transgenes rather than genomic clones is twofold. Foremost is the greater availability of cDNA sequences from ESTs (expressed sequence tags) and gene cloning projects and, secondly, it is generally easier to subclone them into cassettes containing different promoters and other DNA elements for expression analysis. Thus, although cDNAs have greater flexibility in everyday use than their genomic counterparts, it is possible that their widespread use has contributed to some of the transgene expression problems observed to date.

Matrix Attachment Regions

Matrix attachment regions (MARs) are A/T-rich DNA sequences, which bind to the protein nuclear matrix. They are frequently located in gene-rich areas associated with transcribed units and there is evidence that they can reduce or eliminate some forms of gene silencing. Genes flanked by MAR sequences have been shown to exhibit less variation in expression among independent transgenic plants or cell culture lines. In particular, MARs appear to prevent silencing caused by multiple transgene copies *in cis*. There is conjecture as to the precise role of MAR sequences in transgene cassettes, but current theories suggest that they may act as boundary elements facilitating the formation of decondensed chromatin loops that in turn enhance transgene transcription and reduce some forms of gene silencing.

Conclusions

The long-term expression profile of any particular transgene is influenced by several factors and our current understanding of these various factors is incomplete. There are, however, strategies that can be adopted to increase the chances of predictable long-term transgene expression and these will be refined with improved understanding of gene silencing, in particular.

See also: **Crop Improvement**: Mutation Techniques. **Tissue Culture and Plant Breeding**: Somatic Hybridization.

Further Reading

Grierson D, Lycett GW, and Tucker GA (1996) *Mechanism and Applications of Gene Silencing*. Nottingham: Nottingham University Press.

Hannon GJ (2002) RNA interference. *Nature* 418: 244–251.

Lindsey K (ed.) (1998) *Transgenic Plant Research*. Amsterdam: Harwood Academic.

Matzke MA and Matzke AJM (2000) Plant gene silencing (special issue). *Plant Molecular Biology* 43: 1–298.

Reznikoff W and Gold L (1986) *Maximising gene expression*. Burlington, MA: Butterworths.

Vance V and Vaucheret H (2001) RNA silencing in plants – defence and counter-defence. *Science* 292: 2277–2280.

Vaucheret H and Fagard M (2001) Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends in Genetics* 17: 29–35.